THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

THEORY-ENHANCED AUTOMATION OF THE DIGITAL PUBLICS' RELATIONSHIP ASSESSMENTS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

DOCTOR OF PHILOSOPHY

By

HYE LIM LEE

Norman, Oklahoma

2023

THEORY-ENHANCED AUTOMATION OF THE DIGITAL PUBLICS' RELATIONSHIP ASSESSMENTS

A DISSERTATION APPROVED FOR THE

GAYLORD COLLEGE OF JOURNALISM AND MASS COMMUNICATION

BY THE COMMITTEE CONSISTING OF

Dr. Jeong-Nam Kim, Chair

Dr. Robert Kerr

Dr. Xiaochen Angela Zhang

Dr. Ji Hwan Park

Dr. Yun Jang

Abstract

The current dissertation aims to develop a Machine Learning (ML) method for automating the assessment of digital public relations by incorporating the Organization-Public Relationship Assessment (OPRA) developed from the public relations theory. The study targets customers/consumers and employees. For methods, Natural Language Processing (NLP) techniques, specifically text-embedding and classification, are used to analyze the crawled data and three survey data. The results demonstrate that TF-IDF, BERT embedding, and the SVM classification model perform best. The case study outcomes using TripAdvisor and Glassdoor review data validate the previous results. This dissertation project can serve as a pioneering effort to enhance the theoretical foundation of most current data analytics tools in public relations.

Table of Contents

Theory-Enhanced Automation of the Digital Publics' Relationship Assessments

Introduction

Public relations research has undergone significant transformation in recent years due to the emergence of new technological advancements (e.g., Galloway & Swiatek, 2018; Panda et al., 2019; Santa Soriano & Valdés, 2021; Wang et al., 2021). Such technologies include machine learning (ML). While ML has been widely used in various fields, public relations research has not yet fully utilized their potential (Panda et al., 2019). In this dissertation, I will argue that public relations research has not fully embraced the ML method, despite its significance in predicting the communication behaviors of the digital publics.

ML has a lot to offer to public relations research. Public relations study is about understanding and predicting the behaviors, attitudes, and opinions of the public (Baskin et al., 1997; Cutlip et al., 1979, 2013; Grunig & Hunt, 1984; Lattimore et al., 2019; Ledingham, 2003). Public relations researchers rely on various research methods, such as surveys, focus groups, and content analysis. However, these methods are often limited in terms of their scope, depth, and accuracy in handling digital datasets. In contrast, ML can analyze vast amounts of data and identify patterns and trends that would otherwise go unnoticed. For example, sentiment analysis (Medhat et al., 2014) can help public relations researchers understand the tone and context of social media conversations about a particular topic, issue, or organization (e.g., Tam & Kim, 2019). Additionally, public relations researchers can gain deeper insights into the attitudes and opinions of the public, which can inform communication strategies and tactics (e.g., Galloway & Swiatek, 2018; Panda et al., 2019; Santa Soriano & Valdés, 2021; Wang et al., 2021). ML can help them identify the most effective messaging and channels for reaching different segments of the public. They can also be used to identify potential crises or issues before they become public knowledge, allowing public relations professionals to address them proactively (Grunig, 2009; Grunig et al., 2021).

However, one of the main challenges in developing effective data science applications in the public relations field is the need for more expertise and experience in data science among practitioners

(Janssen, 2022; Penn, n.d.). In addition, currently, available commercial online listening tools primarily focus on social media marketing, which does not adequately address the critical role of public relations in building and nurturing relationships between organizations and their publics.

ML, based on mathematical and statistical models, is often applied in practice to solve real-world problems. For example, ML algorithms are used in various fields, from medicine to finance, to make predictions and inform decision-making. However, ML studies often lack theories and concepts to interpret the data (Davenport & Ronanki, 2018; Kitchin, 2014, 2017). Although the field of public relations has developed valuable concepts and theories, such as organization-public relationship assessment (OPRA, Huang, 2001), these theories and models are not extensively utilized in ML research. To fully leverage the power of ML, researchers and practitioners need to adopt appropriate theories and concepts that can help them make sense of the complex and often opaque algorithms underlying these technologies. In a similar vein, Kitchin (2017) argues that data science needs paradigms and frameworks to understand the complicated results of large-scale data analysis. In conclusion, a solid theoretical foundation allows researchers to explain how and why their models work.

By incorporating ML into their research, public relations professionals and researchers can enhance their ability to understand and predict public behavior, ultimately leading to more effective communication and public engagement. Developing theories and concepts for ML research in public relations is necessary for a deeper understanding of the application of these technologies in both research and practice. With this background, the current dissertation project proposes public relations research model that uses text-embedding techniques and supervised ML and leverages previous public relations concepts, such as OPRA.

Literature Review

Understanding Digital Communication in Public Relations

The rise of digital technology has enabled individuals to express their opinions to a global audience. This expansion creates several issues. First, novel problems such as polarization, disinformation, misinformation, fake news, and conspiracy theories have emerged (Sunstein, 2017). It has become easy and inexpensive to spread unverified and unfiltered information about various issues and organizations. The speed at which information spreads is sometimes immeasurable, surpassing the capacity of human agency to keep up with its rapidity (Boyd, 2014).

Second, digital communication generates vast amounts of data, commonly called "big data." The volume and variety of this data present new challenges for public relations practitioners. It is more than just a problem in terms of problematic communication. Public relations professionals previously monitored and scanned the environment using traditional social science methodologies, such as surveys, focus groups, and content analysis. However, due to their resource-intensive nature, these methods may be less effective in handling the enormous amount and variability of data (Shirky, 2010). ML presents a potential solution with their ability to promptly process digital communication data (Chen et al., 2012; Yang et al., 2011). However, the fundamentals of the digital publics' causes still need to be fully comprehended. It remains necessary to gain a better understanding of the thoughts and behaviors of the digital publics.

In this context, public relations add a new dimension to these issues, necessitating advanced approaches to understanding digital communication behaviors. Digital communication has unique characteristics that differentiate it from traditional communication methods. It empowers individuals to engage in public discourse with global citizens, making it a powerful tool for public participation (Castells, 2007, 2010, 2012; Papacharissi, 2010; Rheingold, 2008; Valkenburg, 2017). When individuals participate in digital communication, they often have a genuine interest in the issues, which results in more authentic and accurate data. However, the increased volume, speed, and variety of information generated by the digital publics pose challenges in terms of managing and analyzing this vast amount of data. This is where ML come into play. By leveraging these technologies, researchers can analyze large volumes of data and uncover patterns, trends, and correlations that may not be apparent using traditional research methods. This can enhance our understanding of problematic communication issues and enable the development of effective strategies to address them.

Approaches To Studying the Digital Publics

With the significant potential of digital data, public relations can benefit greatly from data analytics-driven research, particularly in grasping the perceptions and attitudes of the digital publics more accurately (Grunig, 2009; Grunig et al., 2021). The public relations industry arduously develops online listening tools to understand the digital publics better. Several widely used online listening prototypes in the industry employ data analytics (e.g., Brandwatch, Meltwater, Brand24, etc.). For example, Brandwatch offers a user-friendly dashboard for monitoring selected queries on the web, providing updates on topics, sentiments, and user engagement. The coverage includes online news, blogs, forums, reviews, and social media platforms, such as Reddit, Twitter, Tumblr, and YouTube. The platform also offers visualization options like word clouds and topic wheels.

According to Brinker (2020), over 8,000 online listening tools were reportedly available for use in the public relations industry in 2020. While the industry is interested in designating the communication behaviors of the digital publics, previous approaches have lacked theoretical foundations and relied on descriptive and exploratory analysis due to the limitations of large-scale data analysis. (For instance, although sentiment analysis has predicted 28 emotions (e.g., through Google's GoEmotions), it only offers limited insights into executable actions based on public sentiment (e.g., consumers' sentiments toward a brand). These functions merely provide *descriptive presentations* of available data, akin to traditional news-clipping tasks. Merely counting volumes of mentions, identifying topics, or sorting sentiments as positive or negative rarely captures the complexity of the digital publics' mindset.

Instead of merely clipping media report coverages, public relations should adopt a management role in developing strategic communication programs for organizations (Grunig et al., 1992; see Chapter 12). Stack (2011) defines public relations as "a management function that researches an organization and its publics to establish mutually beneficial relationships through communication" (p. 22). Similarly, Cutlip et al. (1994) state that "public relations is the management function that identifies, establishes, and maintains mutually beneficial relationships between an organization and the various publics on whom its success or failure depends" (p. 2). Public relations should adopt scientific and rigorous research methodologies to fulfill this management role. Although existing online listening tools surpass mass-media-relations-focused approaches, they do not provide substantial in-depth insights to practitioners and researchers. If public relations only use this level of technology to understand the digital publics, it will put the whole industry at risk because automation technology can replace many of the public relations roles.

For instance, with the advancements in ML-powered chatbots and automated content generators, such as ChatGPT, monitoring news and generating press releases or articles is now possible without extensive human intervention. Additionally, ML-driven analytics platforms have the capability to process and interpret data, providing actionable insights without the need for human analysts. These developments have raised concerns within the public relations industry, as professionals may face the risk of marginalization or potential job loss due to the emergence of ML systems. Numerous news reports have highlighted the potential threat to communication jobs, including those in advertising and marketing. Bernard Marr, in his article for Forbes (2023), warns about the capabilities of ChatGPT and how it could potentially endanger customer communication jobs. Similarly, public relations professionals cannot ignore these changes and must fully embrace and utilize ML methods for effective management and strategic goal achievement. Conclusively, public relations should design the ML approach extensively using ML methods to maximize the pre-existing human intelligence in the field.

The Issue of The Atheoretical ML Method

It is undeniable that ML has gained popularity in analyzing complex data, including in the field of public relations. However, there are concerns regarding the lack of theoretical grounding for ML methods in social science research. Criticisms of ML methods often revolve around the absence of a clear theoretical framework and meta-learning ability (Lake et al., 2017; Markus, 2008).

While ML is effective in data analysis, it frequently falls short in providing explanations for patterns or relationships, hindering theory development. Data analytics is driven by large amounts of digital data, often disregarding theoretical or conceptual frameworks and relying solely on a data-driven and inductive approach. This perspective disregards the importance of theory. Wise and Schaffer (2015) highlight the challenges of analyzing data using ML, such as determining which variables to focus on and interpreting numerous micro-results. Large-scale data analysis remains limited to descriptive and predictive findings without a solid theoretical foundation.

ML methods tend to prioritize prediction over explanation, lacking interpretative power (Alfuasian et al., 2021). While prediction is useful in practical applications, it is insufficient for theoretical development. Understanding causal relationships between variables, which necessitates a deeper comprehension of underlying mechanisms, is crucial for theory building. Furthermore, ML methods may encounter issues like overfitting, where algorithms become overly tailored to training data, resulting in inaccurate predictions for new data. This can lead to spurious correlations and a lack of generalizability, further undermining theory development. Marcus (2018) also acknowledges the limitations of deep learning, including its inability to reason abstractly, lack of flexibility in handling unexpected situations, and inadequate transparency for safety-critical applications. Marcus suggests a hybrid approach that combines deep learning with symbolic reasoning and rule-based systems to overcome these limitations.

Rudin (2019) argues against using "black box" ML models in high-stakes decision-making contexts and claims that the interpretable models that illustrate their decisions are preferable. Black box models, which are challenging to understand and produce biased outcomes, can be replaced by interpretable models that foster transparency in decision-making processes. Rudin argues the importance of developing interpretable ML methods, including model-agnostic approaches that can be applied to any model. Interpretability is especially critical in decision-making, where trust and fairness are of utmost importance.

Further, some researchers have recently called for integrating ML methods with existing theoretical frameworks to address these concerns — the issue of the atheoretical ML method. There is an approach known as "explainable artificial intelligence (AI)," (e.g., Gunning et al., 2019; Holzinger et al., 2022), which combines AI with causal inference models to deliver transparent and interpretable results. As such, ML should be used to complement traditional research methods rather than replace them. ML can generate hypotheses and identify patterns, which can be further explored and tested using traditional research approaches. While ML methods have the potential to offer valuable insights into complex data, they must be firmly grounded in theoretical frameworks to contribute fully to theory development. Integrating ML with existing theoretical frameworks and utilizing it in conjunction with traditional research methods can lead to the development of more effective theoretical models that explain the mechanisms behind complex phenomena.

Regardless of these continuous efforts, a higher level of interpretability is required when researchers use ML methods. The stakes are high, particularly when social scientists, who are not fully data-oriented, pursue the methods. Hence, I tried to configure the integrative approach using a theoretically guided ML method to improve the understanding of the publics' perceptions of relationships with organizations.

Generic Approach

Three points can summarize the previous section. First, the rise of digital technology has posed challenges for public relations in the realm of digital communication. The sheer volume and variety of data generated by digital communication require public relations practitioners to leverage ML for processing and analysis. However, there is a need for a deeper understanding of the thoughts and behaviors of the digital publics to tackle these challenges effectively.

Second, public relations has turned to data-analytics-driven research, particularly through online listening tools, to gain insights into the perceptions and attitudes of the digital publics. While these tools provide valuable information, they often lack theoretical grounding and rely on descriptive analysis, limiting their effectiveness in developing strategic communication programs. This raises concerns about potential job displacement in the public relations industry. Thus, professionals should view these technologies as opportunities to enhance their strategic management capabilities.

Third, critics argue that ML methods often lack theoretical grounding, focusing more on prediction than explanation. To overcome these limitations, researchers suggest integrating ML methods with existing theoretical frameworks and combining them with traditional research approaches. This approach, called "explainable ML," aims to generate transparent and interpretable results by using ML to generate hypotheses and patterns that can be further explored through traditional research methods. Integrating ML with existing theories can improve the development of theoretical models that explain complex phenomena.

With this background, the current dissertation project proposes a model that adopts the explainable ML approach, combining theoretical concepts from public relations with supervised ML methods. The objective is to facilitate a better understanding of the communication behaviors of the digital publics, assist organizations in developing strategic management programs, and effectively leverage the integration of human intelligence and AI. This chapter briefly introduces the project's main concept, methodology, and final product.

Concept: Organization-Public Relationship Assessment

In this project, I will focus on the OPRA, as it is one of the most extensively studied subjects in the public relations field. Building and maintaining relationships is the ultimate goal of public relations, as highlighted by Cutlip et al. (1994). The concept of relationship building has been the foundation of public relations for decades, with agencies like Edelman using it to report on trust through their *Trust Barometer* for nearly 20 years.

The significance of the relationship concept in public relations research can be examined from various perspectives. Firstly, it is crucial to establish trust and credibility between organizations and their publics. Research has demonstrated that organizations investing in strong relationships with stakeholders are likelier to enjoy a positive reputation, increased loyalty, and long-term success (e.g., Grunig, 1997; Ledingham & Bruning, 2000). Relationship-building is particularly vital for crisis management in public relations. During a crisis, having a solid relationship with the public can help minimize damage to an organization's reputation. Effective crisis management enables organizations to overcome negative perceptions and rebuild stakeholder trust (e.g., Antonacopoulou & Sheaffer, 2014; Coombs, 2021).

Furthermore, the relationship concept in public relations research is important as it contributes to increased organizational effectiveness (Grunig & Dozier, 2003; Gruing & Grunig, 2000). By cultivating relationships with stakeholders, organizations can gather valuable feedback, which can be used to enhance products or services, develop better communication strategies, and identify potential issues before they escalate into crises. In this way, the relationship concept enables organizations to be more responsive and adaptable to their environment (Grunig & Grunig, 1992). The relationship concept in public relations research is also significant because it shifts communication focus from one-way to two-way. Traditionally, organizations have relied on one-way communication, broadcasting messages without actively seeking feedback or engaging in dialogue (Kent & Taylor, 2002; Wilcox et al., 2020). However, the relationship concept highlights the importance of two-way communication, where organizations aim to foster a conversation with their publics rather than merely disseminating information.

The relationship concept is a fundamental aspect of public relations research. Therefore, this project uses it as a key concept to study through advanced methods to expand knowledge.

Organization-Public Relationship Assessment (OPRA)

I have chosen the Organization-Public Relationship Assessment (OPRA) among the many relationship concepts in public relations studies. It is a theoretical framework that systematically evaluates the relationship between an organization and its publics. The OPRA model was developed by James E. Grunig and Todd T. Hunt in their book *Managing Public Relations* in 1984. It has been extensively used by public relations scholars and practitioners and is regarded as one of the field's most influential and useful models (Cheng, 2008; Huang & Zhang, 2013).

OPRA presumes that an organization's success is intertwined with its relationships with its publics. A positive relationship nurtures trust, loyalty, and support. It results in great benefits for the organization's reputation, financial performance, and other important outcomes (DeMiglio & Embry, 2006; Ledingham & Brunig, 1998; Shen, 2016). OPRA provides a framework to evaluate the quality of these relationships and identify improvement areas. The model consists of four key components: *control mutuality, commitment, trust,* and *satisfaction*. *Control mutuality* is the extent to which both parties in the relationship have equivalent power levels. *Commitment* means the level of attachment and dedication that the publics perceive in the organization's efforts of relationship-building and maintenance. *Trust* indicates whether an organization is perceived as having competence, dependability, and integrity by the publics. Finally, *satisfaction* refers to the extent to which the organization meets the needs and expectations of its publics.

OPRA can provide a clear and systematic way to evaluate the effectiveness of public relations efforts. Organizations can comprehensively view their relationship with the public and identify improvement areas by measuring the key components of control mutuality, commitment, trust, and satisfaction. This can help them develop more targeted and effective communication strategies, build stronger relationships with their stakeholders, and ultimately achieve their organizational goals. Another strength of OPRA is its versatility. It has been applied to many different kinds of organizational contexts and publics, from small businesses to international corporations and local to global levels. OPRA is a useful tool for public relations experts and scholars who need to gauge the effectiveness of their communication efforts for relationship management in various contexts.

OPRA is typically conducted through surveys or other research methods that gather data from both the organization and its publics. The collected data is then analyzed to identify areas of strength and weakness in the relationship and develop strategies to improve it. However, a shift in method in the digital communication context is now called for. Instead of the traditional methods to evaluate OPRA, I propose adopting ML for explaining and predicting OPRA.

Method: Supervised Machine Learning (ML)

ML is a subfield of AI that uses algorithms and statistical models to enable computer systems to learn from data and improve performance on a specific task (Bishop, 2006). There are several types of ML, including *supervised, unsupervised, semi-supervised,* and *reinforcement learning.*

Supervised learning includes training a model on labeled data (often by human coders) to predict a target variable. Linear regression, logistic regression, and decision trees (DTs) are examples of supervised learning algorithms (Bengio et al., 2018). On the other hand, unsupervised learning involves training a model on unlabeled data. The common purpose of this method is to discover similar patterns in the given data. Clustering, principal component analysis (PCA), and autoencoders are often employed for this method (Bishop, 2006). Semi-supervised learning is a mix of supervised and unsupervised learning. The model utilizes a small amount of labeled data and a large amount of unlabeled data. This model is highly recommended when acquiring labeled data is expensive or time-consuming (Chapelle et al., 2006). Reinforcement learning is a model used to make decisions based on feedback from the environment. The model goes through a trial-and-error process, taking rewards for making correct decisions and penalties for incorrect decisions (Sutton & Barto, 2018).

I used supervised learning with the text-embedding technique in this study, mainly due to a shortage of datasets and predictions.

Natural Language Processing (NLP)

This project processes text data by adopting NLP, which is a type of computational language processing algorithm to facilitate the analysis of text data. In communication research studying the digital publics, NLP has gained popularity because of the need to process large-scale data from social media, such as Twitter and Facebook. This method offers researchers great insights into social media users' emotions, opinions, and sentiments, as well a glimpse of the publics' attitudes and perceptions. It can also be easily combined with actual communication behaviors, such as how they interact with each other and with the content on the platforms.

A specific exemplary application of this method is sentiment analysis, which is used to analyze the emotional content of text data. It can be useful for understanding the attitudes and opinions of the digital publics toward certain topics or issues. Another application is topic modeling, which is a method used to identify patterns and themes in large amounts of text data. This can be useful for identifying the key issues and concerns of the digital publics and how they change over time. There have not been many NLP and ML methods applied to public relations research yet. But, some public relations studies have used sentiment analysis. For example, the study conducted by Ji et al. (2019) examines the functional and emotional traits of corporate social media message strategies using Facebook data from S&P 500 companies. The study revealed that emotional traits have a greater impact on engagement outcomes compared to functional traits and that the combined effects of emotional and functional traits differ depending on the specific outcomes being examined. Zhao et al. (2020) explored how the public responds to organizational crisis communication strategies in the context of social-mediated crises with sentiment analysis. They found that the preferred responses from the public were those that involved organizations expressing empathy, demonstrating a commitment to learning and improvement, and effectively communicating their messages.

NLP is widely employed in various communication studies beyond the field of public relations. In a study conducted by Dang-Xuan et al. (2013), NLP methods were utilized to examine the sentiments expressed by Twitter users during the 2012 U.S. presidential election. The findings revealed a generally negative sentiment among Twitter users toward both candidates. This research showcases the potential of NLP in gaining insights into the attitudes and opinions of the digital publics toward political candidates. Similarly, in another study by Bastos and Mercea (2018), NLP techniques were applied to analyze the language utilized by Twitter users during the Brexit referendum in the UK. The researchers discovered that Twitter served as a platform for both sides of the debate to disseminate information but with distinct language patterns observed between the camps. This study underscored the capacity of NLP to comprehend the language and communication dynamics among the digital publics.

The shortcoming of previous research utilizing NLP in public relations and communication studies is that they have often relied solely on Twitter posts due to accessibility. However, in today's digital landscape, we can access vast text data from various sources, such as Google Maps reviews and YouTube comments. It is necessary to broaden the range of data sources to analyze public perceptions of organizations' endeavors to foster relationships. In this study, I utilize various datasets, including reviews from platforms such as Amazon, TripAdvisor, and Glassdoor. This approach enhances the external validity of the model and its overall usability.

Study Design

The study went through five phases. Detailed explanations about the data collection and methodologies are presented in the Method section.

Phase 1: Data Collection (Survey and Scraping)

Standard question items from OPRA (Huang, 2000) were used to collect survey responses containing open-ended comments. Multiple coders coded the comments to create labeled data. Additionally, actual communication data from users was scraped for learning. Data collection targeted two types of publics: customers/consumers and employees.

Phase 2: Classification Based on ML Modeling for Both Studies

Several classifiers and text-embedding techniques were applied to the datasets for analysis in order to build the initial ML model of concepts.

Phase 3: Theoretical Interpretation of Classification Outcomes

The model's outcomes were interpreted based on conceptual and theoretical backgrounds during these phases.

Phase 4: Confirmation of the Model

The model(s) was(were) modified accordingly, and the final model(s) for testing was(were) confirmed.

Phase 5: Testing Model Performance with Actual Data - Case Study

The confirmed model(s) was(were) tested using actual data obtained from the web, specifically TripAdvisor and Glassdoor reviews.

Phase 6: Building a Deep Learning Model if There Are More Data Points

The ultimate goal of this study is to develop a deep learning model to enhance prediction accuracy and create a dashboard that users can easily use. However, building a deep learning model requires many data points, necessitating additional data collection through multiple surveys. Considering the limited budget and time, the study initially focuses on the simple ML model phase and continues to collect more data to pursue deep learning modeling in the future. Therefore, Phase 6 should be regarded as a potential future study. If the deep learning modeling phase is completed, the study anticipates the development of an application that provides visualization and deeper insights into digital text data.

Method[1]

The entire process of data collection and analysis is presented in Figure 1. Firstly, I collected two types of data: directly crawled data from the website and survey data. Then, the second human-coding process was done with scrapped data, and it went through data preprocessing, text embedding, and classifications, which is a supervised ML process.

[Insert Figure 1 here]

The project identified three studies based on the types of publics. Study 1 involved consumers and customers, Study 2 involved employees, and Study 3 combined the two studies' target publics. Once classifications were done, the models were evaluated based on accuracy, and the chosen models were applied to a case study with the new dataset covering two different types of publics.

After going through the interpretation and evaluation of the outcomes, I tentatively chose the final model predicting accurate evaluation of OPRA.

Data Collection

When studying the digital publics, it is important to select specific types of the public to focus on. My research examined the relationships between organizations and their customers/consumers and the employees' experience. It is important to note that many different types of publics could be studied, each with their own unique characteristics.

While customers and consumers have a more casual relationship with companies, their experiences are still significant. They interact with products and services regularly but are only sometimes loyal to a single brand. With so many options in the market, they may easily switch to alternatives if dissatisfied.

[1] Acknowledgment: This project's analysis process has been assisted by the expertise of Dr. Jang Yun's lab (Sejong University, South Korea). In particular, Nayoung, Koo, Changyoung, Jeong, and Hyein, Hong did the programming.

Employees, on the other hand, have a much deeper relationship with their workplace. Their experiences are all-encompassing and often rooted in their everyday work. It is not a decision they can easily reverse, as their job is often a key aspect of their life.

Although I focused mainly on customers and consumers in this study due to data accessibility issues, I also examined employees' experiences as a supplementary aspect. By considering these two groups, I intended to better understand the digital publics and their interactions with companies.

Customer Data

Scraped Data (Amazon.com). I scraped the reviews from Amazon.com in June 2021. The study crawled review comments with 1-star and 5-star ratings. The target products for scraping reviews were electronic products, such as Redragon Keyboard, Samsung EvoPlus SSD card, Apple iPad, and AirPods. The total number of reviews from Amazon.com was about 13,503. The average number of reviews per product was around 3,376.

Human Coding Process. No values were assigned to scraped data from Amazon, so it underwent human coding. In August 2022, two researchers—my advisor and myself—conducted the first coding. We used a coding system where 0 represented "No" and 1 represented "Yes," with an additional indicator of -1 if the response tone was negative. However, in the analysis, only 0 or 1 was used. So I converted every -1 to 0 if it was applied during later analysis. In September 2022, the second coding was carried out by two research assistants—one master's and one Ph.D. student—both of whom were native English speakers. The coding system used in this phase involved assigning a value of 0 for "No" and 1 for "Yes," without any additional tone indicators.

In the coding process, a few problems arose, such as different coding patterns and coding schemes. These issues made it difficult to work together, as their coding results may not be easily understandable by others. Additionally, an unbalanced ratio of 0s and 1s could cause data analysis and modeling issues. Fortunately, there were some solutions to these problems. For instance, emotional analysis (see Emotional Labeling section) showed that every label had similar thought patterns despite using different coding patterns and schemes. Recoding to follow a standardized pattern and conducting ad hoc sentiment analysis helped ensure that the coding was more uniform and easier to understand. Another solution to fix this problem was to use sampling to create a more even distribution of 0s and 1s in datasets. It helped improve data analysis and modeling accuracy, as an unbalanced ratio could skew results and create bias. I could produce more accurate results by using these solutions.

Survey Data. The first survey was conducted in October 2021, and seven University of Oklahoma students, who took a communication management class in the fall semester of 2021, participated with extra points on their grades. They were asked to write comments from experience about the companies and brands they thought were the worst or the best. The comments they left were the units of analysis ($N = 32$). At the analysis stage, the best comments were coded as 1 and the worst as 0 (see the questionnaires in Appendix A-1).

Additionally, I collected Amazon Mechanical Turk (MTurk) survey responses. Two surveys used MTurk panels. The first MTurk survey, conducted in May 2022, asked about the participants' most and least favorite companies as consumers. Then the respondents were asked to answer questions about their perceived relationships with the organizations using a five-point Likert scale (1 = strongly disagree, 5 = strongly agree). Afterward, the participants were asked to write the reasoning behind their answers with at least 40 words or 150 characters. The total number of participants was 235 out of 606 after data cleaning ($N = 235$). As the analysis focused on classifying binary groups, the Likert-scale answers from this survey data were recoded into binary values at the analysis stage: 1 (strongly disagree), 2 (disagree), and 3 (neither disagree nor agree) were recoded as 0, and 4 (agree) and 5 (strongly agree) were recoded as 1 (see the questionnaires in Appendix A-2).

The second MTurk survey, following the same structure as the first one, was conducted in October 2022. The only difference was that the respondents answered whether they thought the companies, brands, and organizations tried to enhance trust, satisfaction, and commitment and control mutuality in their customer relationships (yes = 1, no = 0). As the answers were binary, there was no additional recoding process. After cleaning, the total number of participants was 26 out of 50 ($N = 26$; see the questionnaires in Appendix A-3-a).

Employee Data

Employee data also utilized the scraped data and survey data responses. First, data were collected from Glassdoor.com, which is the most popular platform for current/previous employees to comment on their workplace experiences. I manually crawled 30 reviews from Google, Meta, and Twitter in July 2022, when giant tech companies announced widespread layoffs. In particular, reviews for Meta and Twitter were scraped before and after the layoff timing. There were 60 reviews for each. The total number of reviews used for analysis from the scraped data was 90. The sample size was much smaller than the Amazon data, so it was manually coded as 0 or 1.

The survey data for analyzing the employee public was collected through a third survey. Like the customer data, after cleaning, the total number of participants was 26 out of 50 (*N* = 26; see the questionnaires in Appendix A-3-b).

Pilot Testing: Emotion Labeling

I tried to find a useful analytical approach to develop an algorithm for OPRA. As a pilot test, I connected it to the emotional dictionary with the OPRA labels since there are no models available for analyzing the relationship concepts. Therefore, I first tried to use emotion datasets as training datasets. The dataset used for the emotional analysis of sentences is GoEmotions (Demszky et al., 2020), one of the latest and most diverse emotion lexicons. This dataset is a Bidirectional Encoder Representations from Transformers (BERT)-based model of Google emotion data based on analysis of Reddit posts. This data provides 28 emotions (*neutral, admiration, approval, annoyance, gratitude, disapproval, amusement, curiosity, love, optimism, disappointment, joy, realization, anger, sadness, confusion, caring, excitement, surprise, disgust, desire, fear, remorse, embarrassment, nervousness, pride, relief, grief*). The embedded words in the phrases are the basis for judgments about emotions and human-coded concepts from the survey and comment datasets. The example below presents how the study tries to connect the emotion labels from GoEmotion and human-coded OPRA labels (i.e., supervised ML).

The data analysis went through two phases. Firstly, the first mTurk survey and Amazon review datasets were analyzed using several classifiers to predict the emotions of texts. The following classifiers were used: *Naive Bayes (NB)* (e.g., Hristea, 2012), *Random Forest (RF)* (Brieman, 2001)*, Linear Regression (LR)* (e.g., Hastie et al., 2009), *Support Vector Machine (SVM)* (e.g., Suykens & Vandewalle, 1999), *Convolutional Neural Networks (CNN*) (e.g., Krizhevsky, et al., 2017), *and CNN–Long Short-Term Memory (CNN–LSTM)* (e.g., Shi et al., 2015). The entire outcomes are presented in Appendix B.

NB is based on Bayes' theorem for classification or prediction tasks. It is "naive" because it presumes that all the features in the classification are independent of each other, which is often not the case. The algorithm works by computing the probability of each class for a given input based on the probability of each feature in the input belonging to each class. It then chooses the class with the highest probability as the predicted class for the input. LR is a popular supervised learning algorithm used in ML for predicting continuous numerical outputs. It involves modeling the relationship between a dependent variable and one or more independent variables using a linear equation.

CNN is an algorithm that is particularly effective for analysis tasks. It is derived from the structure and function of the visual cortex in the brain containing neurons specialized for detecting different visual characteristics. CNN–LSTM is also a type of deep learning architecture like CNN, but it combines the capabilities of CNN and long short-term memory (LSTM) networks. It is widely used for sequential data analysis tasks, such as video processing, natural language processing, and time-series analysis. In this architecture, the input data is first processed by a CNN, extracting the features from the data hierarchically. The output of the CNN is then conveyed to an LSTM network, where the researchers can temporally model the data and learn long-term patterns. The LSTM network can have information over a longer period of time, which is useful for speech recognition.

Based on the results, the best performers were CNN and CNN–LSTM (both with 50% accuracy). Feature importance is a value that numerically represents the importance of variables used to predict the target value. The results for trust using *DT, Extra Tree (ET), RF, and Extra Tree Ensemble (ETE)* showed that the prediction accuracy of the trust label was the same in all models (82%). As a result of the analysis of trust, the prediction accuracy of the control mutuality value was the same in all models.

The main takeaway is that both data results show high importance for some of the 28 emotions when predicting OPRA values. Additionally, some emotions with a value of 0 were consistently present in all four tree models, indicating that these emotions do not predict any concepts of OPRA. This result suggests that OPRA values can be predicted using only a few emotions among the 28 available, and these patterns were similar across all coders. The pilot testing of emotion labeling results suggested that although there were differences in the explicit human coding patterns, there was some level of similarity. Furthermore, this testing indicated that a direct and simple approach, such as text embedding and classification, may be more effective in automatically evaluating the OPRA of digital text data.

Analysis

Data Preprocessing

Data preprocessing is an essential step in ML that involves transforming raw data into a clean and structured format suitable for analysis (Garcia et al., 2015). It encompasses various tasks, including deleting unnecessary words, cleaning the data, normalizing it, scaling features, and handling missing values. This step is imperative to secure the quality and accuracy of the model's outcomes. By adopting proper techniques depending on the dataset's characteristics, data preprocessing plays a significant role in strengthening the performance of ML models, finally leading to the best possible results. In this paper, I used the nine preprocessing techniques given below:

1. *Convert text to lowercase:* All words in the text were converted to lowercase to avoid distinguishing between the same word in different cases. For example, "Samsung" and "samsung" would be recognized as the same word.
2. *Remove URLs, numbers, whitespaces, and accents:* These elements were unnecessary for text analysis and could be removed to simplify the data.
3. *Named-entity recognition (NER):* This technique was used to handle named entities, such as a person's name and company names. For instance, "New York" would be recognized as a single entity "New-York".

4. *Expand contractions:* Contractions like "don't" were expanded to their full form, such as "do not", to make them more recognizable to the ML model.
5. *Expand acronyms:* Acronyms, such as "ASAP," were expanded to their original words, such as "as soon as possible," so that the model can better understand them.
6. *Stopwords:* Commonly used words, such as "the" and "and," were removed as they did not carry much meaning in the text. I also reviewed the stopwords list and applied the updated version of the list to preprocessing.
7. *Tokenization:* It was the process of breaking down the text into smaller units, such as words, phrases, and sentences, to assign meaning more effectively. For example, "never give up" would be split into "never", "give", and "up".
8. *Stemming:* This technique involved reducing words to their stem, which is the base form of the word. For instance, "historical" and "history" would both be stemmed to "histor".
9. *Lemmatization:* This technique is used to reduce the different forms of a word to a single form, but it is more precisely than stemming. For example, "historical" and "history" would both be lemmatized to "history."

The filenames used in the analysis were labeled to indicate whether specific techniques had been applied (0: not performed, 1: performed). Each filename consisted of six digits (e.g., 000-0-00), with each digit representing a different processing step. The first digit indicated whether "Named Entity Recognition" processing had been performed. The second digit indicated whether "Expansion of Contractions" processing had been performed. The third digit indicated whether "Expansion of Abbreviations" processing had been performed. The fourth digit indicated whether "Stopword Removal" processing had been performed. The fifth digit indicated whether "Stemming" processing had been performed. The sixth digit indicated whether "Lemmatization" processing had been performed (0: not performed, 1: performed). Therefore, the preprocessing datasets were labeled with numbers ranging from 0 to 7. Data 0 corresponded to 000-0-00, data 1 corresponded to 000-0-11, data 2 corresponded to 000-1-

00, data 3 corresponded to 000-1-11, data 4 corresponded to 111-0-00, data 5 corresponded to 111-0-11, data 6 corresponded to 111-1-00, and data 7 corresponded to 111-1-11.

Text Embedding

Text embedding is a widely used technique in ML that involves converting text data into numerical vector representations, enabling computers to process and understand text (Mikolov et al., 2013). This technique has greatly impacted NLP tasks, such as sentiment analysis, text classification, and information retrieval. I used *Term Frequency-Inverse Document Frequency (TF-IDF), Word2Vec, GloVe,* and *BERT,* which are among the most commonly employed text-embedding techniques.

TF-IDF is a statistical method that assesses the importance of each word in a document (Spärck, 1972). It assigns weights to words based on their frequency within the document and their inverse frequency across the entire corpus. Word2Vec (Mikolov et al., 2013) and GloVe (Pennington et al., 2014) are neural network-based models that learn word embeddings by analyzing word co-occurrences in extensive datasets. Word2Vec represents each word as a high-dimensional vector, capturing semantic relationships between words. BERT is a pre-trained language model, which employs a transformer-based neural network to generate context. It is renowned for its outstanding performance in various NLP tasks and has been widely adopted across industries and applications (Devlin et al., 2018).

Each text-embedding method has its strengths and weaknesses. For instance, TF-IDF is simple and efficient but may not capture fine-grained semantic relationships between words. Word2Vec and GloVe, on the other hand, are more sophisticated and can capture semantic relationships effectively, but they require substantial amounts of training data. BERT, being even more sophisticated, can generate contextualized embeddings, but it demands even larger training data and computational resources.

This study explores TF-IDF, Word2Vec, GloVe, and BERT to determine the optimal text-embedding technique for predicting labels from online and survey text data. The study employs the generated embeddings as input features for the ML model.

Classification Models

ML classification methods are widely used in various applications, including image recognition, NLP, and fraud detection. For this research, I used SVM, LR, ET, RF, and K-Nearest Neighbors (KNN). SVM and LR are known as the best-performing models, which are similar to deep learning. ET and RF are DT-based models, known for enabling progressive accuracy improvement. KNN is the simplest model that allows easy interpretations.

SVM is a supervised ML method that can be used for both binary and multi-class classification (Suykens & Vandewalle, 1999). It aims to find the best hyperplane that separates the classes. The hyperplane that maximizes the margin between the classes is chosen as the decision boundary. SVM has been shown to perform well in many applications, including text classification and image recognition. LR is a popular ML method used for binary classification problems (Hosmer Jr. et al., 2013). It models the probability of the output variable given the input features. The logistic function maps the input features to a probability value between 0 and 1.

RF is another ensemble learning method that combines multiple DTs to improve classification performance (Breiman, 2001). The key idea of RF is to create a diverse set of DTs by randomly selecting features and thresholds for each node in the tree. The final prediction is made by aggregating the predictions of all the DTs. ET is an ensemble learning method that combines multiple DTs to improve classification performance (Geurts et al., 2006). It is similar to RF but uses a different method to create the DTs. ET randomly selects features and thresholds for each node in the tree, which leads to a more diverse set of DTs.

KNN is a simple and effective ML method for classification problems (Altman, 1992). The idea behind KNN is to find the K nearest neighbors to a given data point in the feature space and then assign the class label that is most common among those neighbors.

I chose these classification models because SVM and LR are similar to deep learning methods. Further, ensemble models (ET, RF) can improve performance using several models. Lastly, KNN is easy to interpret and has been widely used in different fields.

Training and Testing Dataset

In ML, dividing the dataset into training and testing sets is an assessment step for the performance of the given model. The training set is used to train the model, and the testing set evaluates how accurately the model can predict new data. The common ratio of data splitting is 4:1 (training vs. testing) (Joseph, 2022). There are other ratios as well, including 70:30, 60:40, and even 50:50. The ratio often relies on some factors, such as the dataset's complexity or size. In addition, several factors must be considered in terms of the technique selected for data splitting (e.g., stratified sampling) to construct training and testing datasets. These factors include the size of the dataset, model complexity, and the research problem. For this particular study, I only utilized the simple split method. Given the small dataset size, a ratio of 2.3:1 was used, with approximately 70% of the data utilized for training and the remaining 30% for testing. Note that I did not use a validation dataset because of the limited samples. There are some techniques for overcoming this issue, such as bootstrapping, but it was not conducted, so I consider it as one of the limitations.

Accuracy

Accuracy is a metric used to evaluate the performance of a classification model. It measures the percentage of correctly predicted instances out of all instances in the dataset. The formula for calculating accuracy is as follows (see Table 1):

Accuracy = (Number of Correct Predictions) / (Total Number of Predictions)

[Insert Table 1 here]

For example, if a model correctly predicts 90 out of 100 instances, the accuracy is 90%. It is important to note that accuracy alone may not always be a sufficient metric for evaluating a model's performance, especially in cases where the dataset is imbalanced, or there are multiple classes.

Other evaluation metrics, such as *precision, recall,* and *the F1 score*, may also be necessary. Precision measures the proportion of true positives (correctly predicted positive samples) among all the samples that the model predicted as positive. It is defined as the ratio of true positives to the sum of true and false positives. Similarly, recall measures the proportion of true positives among all the actual positive samples in the dataset. It is defined as the ratio of true positives to the sum of true positives and false negatives. The F1 score is the mean of precision and recall, a balanced measure of the model's overall performance. It is defined as the weighted average of precision and recall. Support is the number of samples in each class used to calculate precision, recall, and F1 score. I utilized accuracy in this study to assess both positives and negatives. However, incorporating additional metrics may be more beneficial, which will be further elaborated on in the Limitations section. Moreover, due to the absence of the validation dataset, I could not include *hyperparameter tuning*. This was primarily due to the limited dataset, which will be discussed in greater detail during the analysis and discussion.

Developmental Study

The developmental study aimed to evaluate the performances of models using several datasets. I targeted two types of publics: consumers/customers and employees. The consumer and customer data (Study 1) was easy to access and collect, so the current project first focused more on the consumers and customers as foundational work for building the system. In this stage, four text embedding techniques and six classifiers were applied to four labels using three types of data (scraped, survey, and combined), resulting in a total of 288 outcomes for each study. The results are reported based on the study, data type, and label.

Study 1: Consumers and Customers

Dataset 1: Amazon Data

The training data consisted of 197 samples, and the testing data consisted of 50 samples for all the labels, resulting in a total of $N = 247$ data points.

Control Mutuality. The best model of the control mutuality label was the KNN using TF-IDF text embedding of the preprocessing data 7. The training accuracy was .92, and the testing accuracy was .92. The other datasets, including the data 0, 1, and 4 with the TF-IDF-KNN models, also showed good performance: .93 (training) and .94 (testing). The other best model of control mutuality was the KNN using BERT text embedding of the dataset applied for the preprocessing datasets 1, 5, and 7. The accuracy of training data was as follows: .92 (data 1), .93 (data 5), and .91 (data 7), while the accuracy of the testing data was .92 (data 1) and .90 (data 5 and 7). GloVe-KNN and Word2Vec-KNN showed .90 testing accuracy with .90 training accuracy for data 6.

[Insert Table 2 here]

Commitment. The best model of the commitment label was also the KNN using TF-IDF text embedding of the dataset for all of the datasets, as evidenced by the control mutuality results. The accuracy of the training data was .89, while the accuracy of the testing data was .90 (data 3 and 7) and .88 (data 0, 1, 2, 4, 5, and 6). There were many well-performing models, including the BERT-SVM, the BERT-KNN, the GloVe-SVM, the GloVe-LR, the GloVe-LR, the TF-IDF-SVM, the TF-IDF-LR, the Word2Vec-SVM, and the Word2Vec-LR. The accuracy level of the testing data was .84, while the accuracy level of the training data varied.

[Insert Table 3 here]

Satisfaction. The best model of the satisfaction label was the KNN using the BERT text embedding of the preprocessing data 3. The accuracy of the training data was .80, and the accuracy of the testing data was .76. The second best model was the Word2Vec-KNN of data 1 and 5. The accuracy level of the testing was .84 for both, whereas the accuracy level of the training was .80 (data 1) and .78 (data 5).

[Insert Table 4 here]

Trust. The best model of the trust label was the KNN using the BERT text embedding of the preprocessing data 3, similar to the satisfaction label. The accuracy of training was .74, and the accuracy of testing was .72. The second-best model was the BERT-KNN of the data 7. The accuracy level of the testing was .76 and the training was .70.

[Insert Table 5 here]

Summary. The results of the Amazon dataset indicated that the KNN model showed the best performance for labels with limited data points. For control mutuality and commitment, the TF-IDF embedding performed best and showed a much higher testing accuracy level (.88 and .92) as compared to satisfaction (.76) and trust (.72) with BERT embedding.

[Insert Table 6 here]

Dataset 2: Survey Data

Training and testing data sizes varied by labels because there were some missing data in each survey dataset.

Control Mutuality. The best model of the control mutuality label was the SVM using the TF-IDF text embedding of the preprocessing data 3, 4, 5, and 6. The accuracy of training (N = 836) was .93, and the accuracy of testing (N = 210) was .85 (Total N = 1,046). The other datasets, including data 0 and 7 with the TF-IDF-SVM models, also showed good performance: .93 (training) and .84 (testing). In addition, the TF-IDF-LR, the TF-IDF-KNN, and the BERT-SVM also performed well.

[Insert Table 7 here]

Commitment. The best model of the commitment label was the SVM using the TF-IDF text embedding of the preprocessing data 1. The accuracy of training was .92 (N = 834), and the accuracy of testing (N = 209) was .88 (Total N = 1,043). Data 3 with the TF-IDF-SVM models also performed well: .93 (training) and .87 (testing). Additionally, the TF-IDF-LR, TF-IDF-KNN, and BERT-SVM performed well.

[Insert Table 8 here]

Satisfaction. The best model of the satisfaction label was the SVM using the TF-IDF text embedding of the preprocessing data 4 and 6. The accuracy of training was .94 (N = 772). and the accuracy of testing (N = 194) was .85 (Total N = 966). Data 3 and 7 with the TF-IDF-SVM models also showed good performance: .94 (training) and .84 (testing). In addition, the TF-IDF-LR exhibited good performance.

[Insert Table 9 here]

Trust. The best model of the trust label was the SVM using the TF-IDF text embedding of the preprocessing data 3 and 4. The accuracy of training was .92 (N = 961), and the accuracy of testing (N = 241) was .86 (Total N = 1,202). The rest of the datasets with the TF-IDF-SVM models also performed well: the training accuracy was .85, while the testing accuracy varied. Besides them, the BERT-SVM, the BERT-LR, the GloVe-LR, the TF-IDF-LR, and the TF-IDF-KNN performed well.

[Insert Table 10 here]

Summary. Unlike the results of the Amazon dataset, the results from the three surveys showed consistency in the best-performing models. The TF-IDF with the SVM embedding exhibited the best outputs for all labels. The testing accuracy levels were not very different among the four labels; all of them were higher than .85.

[Insert Table 11 here]

Dataset 3: Amazon Data and Survey Data

The training data and testing data sizes varied by labels because there were some missing data in each survey dataset.

Control Mutuality. The best model of the control mutuality label was the LR using the TF-IDF text embedding of the preprocessing data 5 and 7. The accuracy of training (*N* = 1,038) was .93, and the accuracy of testing (*N* = 260) was .83 (Total *N* = 1,298). The other datasets, including data 1, 3, and 4 with the TF-IDF-LR models, also showed good performance: .93 (training) and .82 (testing). The TF-SVM of data 7 showed excellent performance: .93 (training) and .92 (testing). Additionally, the BERT-LR and the BERT-SVM showed good performances.

[Insert Table 12 here]

Commitment. The best models of the commitment label were the SVM using the TF-IDF text embedding of the preprocessing data 2, 4, and 6 and the LR using the TF-IDF text embedding of the data 2. The accuracy of training (*N* = 1,200) was .91 or .92, and the accuracy of testing (*N* = 301) was .83 (Total *N*= 1,501). The other datasets, including data 3 and 5 with the TF-IDF-SVM models, also showed good performance: .82 (testing). The TF-IDF-LR of the data 0, 3, 4, 5, 6, and 7 showed good performance, with .84 testing accuracy. The training accuracy levels varied depending on the preprocessed dataset. Besides these, the BERT-SVM also showed good performance.

[Insert Table 13 here]

Satisfaction. The best model of the satisfaction label was the SVM using the TF-IDF text embedding of the preprocessing data 1, 3, and 7. The accuracy of training (N = 970) was .93, and the accuracy of testing (N = 243) was .85 (Total N = 1,213). The other datasets, including data 2, 4, 5, and 6 with the TF-IDF-SVM models, also showed a good performance with a .84 testing accuracy. Further, the TF-IDF-LR of data 1, 3, and 7 also exhibited excellent performance with a .84 testing accuracy. The training accuracy levels varied depending on the preprocessed dataset. The BERT-SVM also showed a good performance.

[Insert Table 14 here]

Trust. Like the satisfaction label, the best model of the trust label was the SVM using the TF-IDF text embedding of the preprocessing data 1 and 5. The accuracy of training was .92 (N = 1,160), and the accuracy of testing (N = 290) was .80 (Total N = 1,450). The rest of the datasets with the TF-IDF-SVM models also performed well, with a .79 testing accuracy; The training accuracy levels varied. The TF-IDF-LR also showed good performance.

[Insert Table 15 here]

Summary. The results of the merged customer datasets (Amazon dataset and three survey datasets) showed different patterns between control mutuality and commitment versus satisfaction and trust in terms of the best-performing models. For the first group, the TF-IDF with the LR model had the best outcomes; for the second group, the TF-IDF with the SVM model showed the best outcomes. The testing accuracy levels decreased slightly. The lowest level was .80 for the trust label. Notably, the gap between training and testing accuracy levels increased from the previous dataset. The differences were larger than .10 for all four labels. When considering the balance, most of the BERT-SVM model results for every label had the smallest gap, although the testing accuracy in the trust label data was quite low (< .07).

[Insert Table 16 here]

Study 2: Employees

Dataset: Survey Data and Glassdoor Data

The training data size was 168, and the testing data size was 43 for all the labels (Total $N = 211$).

Control Mutuality. The best model of the control mutuality label was the KNN using the Word2Vec text embedding of the preprocessing data 5. The accuracy of training was .79, and the accuracy of testing was .75. The other datasets, including data 1 and 7 with the Word2Vec-KNN models, also showed good performance: .81 (training) and .73 (testing).

[Insert Table 17 here]

Commitment. The best model of the commitment label was the KNN using the TF-IDF text embedding of the preprocessing data 3. The accuracy of training was .85, and the accuracy of testing was .77. The other datasets, including data 1, 5, and 7 with the TF-IDF-KNN models, also showed good performance: .85 (training) and .74 (testing).

[Insert Table 18 here]

Satisfaction. The best model of the satisfaction label was the KNN using the BERT text embedding of the preprocessing data 1, 5, and 7. The accuracy of training was .93 (data 1 and 5) and .95 (data 7). The accuracy of testing was .88. The second-best performing cases were from the TF-IDF-SVM model. The testing accuracy was .86 (data 4, 6, and 7) and .88 (data 5). The other noticeable results were the BERT-KNN, the GloVe-SVM, and the TF-IDF-KNN outcomes.

[Insert Table 19 here]

Trust. The best model of the trust label was the SVM using the TF-IDF text embedding of the preprocessing data 1, 5, and 7. The training accuracy was .91 (data 1) and .92 (data 5 and 7). The testing accuracy was .81. The rest of the data with the TF-IDF-SVM showed good performances: the testing accuracy was .79, and the training accuracy varied based on the data. The outcomes from the TF-IDF-KNN and the Word2Vec-KNN showed good performances.

[Insert Table 20 here]

Summary. The employee dataset showed the most inconsistent patterns. In the control mutuality label results, there were few significant outcomes, except the Word2Vec-KNN models of the data 5. However, the overall testing accuracy of Word2Vec-KNN models was also lower than the commitment, satisfaction, and trust best outcomes (.75). The most popular embedding was the TF-IDF for the other three labels. Besides the trust label, the KNN appeared to be the most frequent in the best-performing outcomes. This inconsistency will be discussed further in the Discussion section.

[Insert Table 21 here]

Study 3: Customers and Employees

The training and testing data sizes varied by labels because there were some missing data in each survey dataset.

Control Mutuality. The best model of the control mutuality label was the LR using the TF-IDF text embedding of the preprocessing data 3 and 5. The accuracy of training (*N* = 1,208) was .2, and the accuracy of testing (*N* = 302) was .86 (Total *N*= 1,510). The other datasets, including data 0, 1, 2, and 5 with the TF-IDF-LR models, also performed well: .92 or .93 (training) and .85 (testing). The TF-IDF-SVM of data 2 showed the second-best performance: .92 (training) and .84 (testing). The BERT-SVM and the TF-IDF-SVM of the other datasets also performed well.

[Insert Table 22 here]

Commitment. The best model of the commitment label was the LR using the TF-IDF text embedding of the preprocessing data 6 and 7. The accuracy of training (*N* = 1,031) was .92, and the accuracy of testing (*N* = 258) was .85 (Total *N* = 1,289). The other datasets, including data 1, 3, 4, and 5 with the TF-IDF-LR models, also showed good performance, with a .84 testing accuracy. The TF-SVM of data 0, 2, 6, and 7 also showed excellent performances, with a .84 testing accuracy. The testing accuracy levels varied in both cases depending on the preprocessed datasets. Besides, the BERT-SVM showed good performance.

[Insert Table 23 here]

Satisfaction. The best model of the satisfaction label was the SVM using the TF-IDF text embedding of preprocessing data 1 and 7. The accuracy of training (*N* = 1,139) was .91, and the accuracy of testing (*N* = 285) was .80 (Total *N* = 1,424). The other datasets, including data 3 and 5 with the TF-IDF-SVM models, also performed well: .91 or .92 (training) and .82 (testing). The second-best performance was the TF-IDF-LR from data 0, 1, and 3: .89 (training) and .81 (testing). In addition to these, the BERT-SVM showed good performance. The TF-IDF-ET showed good testing accuracy levels, but training accuracies were 1.00, which indicated that they were overfitting. Hence, it was not considered a good-performing model.

[Insert Table 24 here]

Trust. Like the satisfaction label, the best model of the trust label was the SVM using the TF-IDF text embedding of preprocessing data 7. The accuracy of training (*N* = 1,406) was .91, and the accuracy of testing (*N* = 352) was .81 (Total *N*= 1,758). The other datasets, including data 4, 5, and 6 with the TF-IDF + SVM models, also showed good performance: .91 or .92 (training) and .80 (testing). The second-best performing model was the TF-IDF-LR from data 4: .87 (training) and .79 (testing).

[Insert Table 25 here]

Summary. When all of the available datasets were merged, the best-performing text embedding technique was the TF-IDF. It showed great results when combined with the SVM model, except for control mutuality. Among the results, the testing accuracy of control mutuality based on the TF-IDF-LR combination was the highest, while the testing accuracy of the trust label with the TF-IDF-SVM combination was the lowest (.81).

[Insert Table 26 here]

Summary of Study 1 to Study 3 Results

In summary, the studies showed that GloVe was ineffective in text embedding. The most successful embedding method was TF-IDF, followed closely by BERT. It is worth noting that Study 2 had the most variation in results. Based on the studies, possible theories for the results were formulated. TF-IDF was found to be the best-performing method, which could be attributed to the smaller dataset.

However, the results suggested that a bigger data size was necessary for accurate results. Another possible reason for the success of TF-IDF could be the sparsity of the data, which means that there were many distinct terms with limited instances or occurrences. This is also connected to the data size issue. The current method is shallow, so it does not apply any hyperparameter tuning processes. Despite these factors, BERT was also found to be one of the best performers. This suggests that BERT can perform well if more data is available, as it generally outperforms traditional embeddings. Moreover, BERT's ability to comprehend sentence nuances makes it a good choice for relationship assessment, which requires a thorough understanding of context.

[Insert Table 27 here]

DT-based models like ET or RF were ineffective for classification in any of the studies. SVM was the most frequently successful classification model, followed by KNN. Interestingly, the combined datasets in studies, such as Study 1 (customers and consumers) with data 3 (all data) and Study 3 (all publics), showed that LR performed well. The results showed that SVM and KNN were the best-performing methods for embedding because the current study had small or medium-sized datasets. SVM is particularly effective for binary classifications, such as yes or no, while KNN is known to handle data with outliers well. In this study, there were inconsistencies among coders, which led to higher instances of outliers. Additionally, RF generally performed better than both SVM and KNN, especially when dealing with larger samples and features but it did not because of the limited data size.

[Insert Table 28 here]

Summary of Accuracy Results

As the results from Study 1 suggest, there were lower accuracy levels in satisfaction and trust. Conversely, Study 2 displayed lower control mutuality and commitment accuracy levels. It is worth noting that both studies had a small sample size of around 200, which may reflect the difference in the relationship being studied. Interestingly, satisfaction and trust were more pronounced in customer relationships, while control mutuality and commitment were more pronounced and complex in employee relationships (see Figure 2).

[Insert Figure 2 here]

Model Comparison

The top-performing models across Study 1 to Study 3 were BERT-KNN, TF-IDF-SVM, and TF-IDF-KNN. However, it is important to note that having high training accuracy does not guarantee model balance. Interestingly, in many cases, the model with the smallest discrepancy between training and testing accuracy was BERT-SVM. Both training and testing accuracy levels exceeded .70.

Figures 3 to 7 provide visualizations of the training and testing accuracy levels and their differences for each study. Notably, the KNN models often displayed substantial disparities between training and testing accuracy, raising concerns about stability. On the other hand, the SVM models exhibited relatively smaller differences. Furthermore, TF-IDF models tended to yield higher testing accuracy compared to BERT embedding. However, it also resulted in larger gaps between training and testing accuracy levels, which could lead to less stable or inaccurate predictions when applied to new datasets.

Therefore, I opted only to use SVM models for the case studies. Although TF-IDF embedding may have its challenges, it was included in the case studies because TF-IDF-SVM consistently achieved the highest testing accuracy in most scenarios. In conclusion, BERT-SVM and TF-IDF-SVM were tested in the case studies to determine the optimal model.

[Insert Figure 3-7 here]

Case Study

In the developmental study, I discovered that the unbalanced ratio of 0s and 1s in satisfaction and trust was causing a problem. To solve this issue, I saved and used the models with an even distribution of 0s and 1s in a case study for satisfaction and trust. This approach helped improve the accuracy levels and ensured a fair representation of satisfaction and trust in the study.

[Insert Table 29 here]

Data

To validate the two recommended models, namely (1) BERT-SVM and (2) TF-IDF-SVM, I utilized two distinct datasets accessible to the public. For the customer data, I employed the TripAdvisor users' hotel review data obtained from Kaggle (https://www.kaggle.com/datasets/andrewmvd/trip-advisor-hotel-reviews, *N* = 20,491).

Regarding the employee data, I gathered reviews from Glassdoor.com, focusing on different companies. However, I could only retrieve information from 5 pages for each company due to Glassdoor's policy limitations on web scraping. I collected employee reviews for 10 companies, including *Hulu, Netflix, Uber, Lyft, Amazon, Glassdoor, Indeed, GrubHub, DoorDash*, and *Yelp* (*N* = 899).

Results

To validate the accuracy of the OPRA label prediction model, a case study was conducted using new data. The initial model's prediction accuracy for OPRA labels was found to be approximately 80%. This case study assessed whether the accuracy remained consistent when applied to new data. In other words, the objective was to evaluate the model's prediction accuracy on this new dataset. To achieve this, a sample of 50 comments was taken from each new dataset, and a manual check was performed for any incorrect classifications.

The case study consisted of two stages. Firstly, I selected the preprocessing data type, the classification model, and the text-embedding technique. In this case study, I chose preprocessing data 7, encompassing all preprocessing techniques with SVM, and utilized two different embeddings: (a) BERT and (b) TF-IDF. Secondly, I saved the model as pickle files to predict the labels of the new datasets.

The trust and satisfaction labels exhibited poor performance due to a higher occurrence of 1s compared to 0s. The ratio of 1s in the entire dataset was more than 60% greater than the ratio of 0s for both labels. For instance, in the Study 3 dataset, the ratio of 1s was approximately 68% for both the trust and satisfaction labels. Consequently, I created models for these two labels to balance the number of 1s and 0s and utilized them for the case study with the new datasets.

Application 1: BERT-SVM

Customer Data (TripAdvisor). The results showed that the trust label had the highest correct identification (84%), while the control mutuality label had the lowest accuracy (72%). The commitment and satisfaction labels fell in between, with 78% and 74% a correct identification, respectively (see the complete details of outcomes in Appendix C-1).

[Insert Table 30 here]

Employee Data (Glassdoor). The results indicated that the trust label achieved the highest accuracy (92%), while the satisfaction label had the lowest correct identification (82%). The control mutuality and commitment labels performed well, with 90% and 86% correct identification , respectively (see the complete details of outcomes in Appendix D-1).

[Insert Table 31 here]

Application 2: TF-IDF - SVM

Customer Data (TripAdvisor). The results revealed that the satisfaction label had the highest accuracy (68%), while the trust label had the lowest accuracy (44%). The control mutuality and commitment labels fell in between, with 60% and 50% accuracy, respectively (see the complete details of outcomes in Appendix C-2).

[Insert Table 32 here]

Employee Data (Glassdoor). The results showed that the control mutuality label achieved the highest accuracy (70%). However, the remaining labels demonstrated poor performance: trust (64%), satisfaction (60%), and commitment (52%) (see the complete details of outcomes in Appendix D-2).

[Insert Table 33 here]

Summary

Compared to the BERT-SVM combination, the TF-IDF-SVM combination exhibited lower accuracy levels. Furthermore, the customer data, particularly with longer texts, showed inferior prediction performance compared to the employee data. Some inconsistencies in the longer text could not be fully captured, particularly in TF-IDF-SVM. Nevertheless, the case study using the BERT-SVM showed fairly good predictions, meaning that the BERT-SVM model would perform better with more datasets.

In summary, the case study validated the accuracy of the OPRA label prediction model when the appropriate model was selected. Utilizing new data sources had minimal impact on the model's prediction accuracy, highlighting its moderate level of robustness and generalizability. However, further investigation with various datasets is necessary after more learning.

Discussion

Theoretical Implications

In recent public relations scholarship, some new theories and concepts, such as dialogue theory (e.g., Kent & Talyor, 2002), have been raised and become popular. These new theories and concepts frequently tackle the outdatedness of the classical theories, but I found through this project that when exploring the opinion of the digital publics, it is essential to look beyond the newest theories and public relations strategies. OPRA is an original effort to fundamentally contemplate the ideal relationship goal that the organization should pursue. Revisiting these concepts can incorporate them into the new research methods and better amplify the voices of the publics because the digital space is more participative. Thus, I believe, the main contribution of the current work is reinviting the classical concepts in a contemporary context, resulting in a more inclusive, advanced, and accurate approach to comprehending the digital publics (Grunig, 2023).

Next, with the ML results, we can more deeply rethink the distinctive dissimilarities of the nature of relationships between diverse communities and the publics. For example, control mutuality and commitment in employees' evaluations should be monitored for a longer time than the customers and consumers. In previous research, surveys or focus group research assessed the employees' perspectives in relationships but were not efficient timewise. Besides employees, the types of publics taking more time to establish those two factors such as voters, residents, and fans, should be monitored by having a proposer time base.

Finally, there could be some concern that these ML methods can invalidate the classical concepts' arguments and propositions, as we have not used the publics' actual voices quite often in the past. ML can negate the theoretical prediction from public relations scholars if it accumulates more empirical outcomes. However, as human nature cannot be dramatically changed, the relationship does. Therefore, this concern should rather be used to motivate further research that integrates new methods to enrich our previous understanding of the publics' relationships with the organization. The ML method could be helpful when it comes to discerning what has been right or wrong in our comprehension of relationships in the public relations field.

Methodological Implications

In communication research, survey methods have been widely used to gain insights into the publics' opinions and attitudes. However, this traditional method faces methodological limits as our communication environment becomes increasingly digitalized. With the explosive growth of digital interactions, we are now faced with massive volumes and incredible velocity of communication that are difficult to capture and analyze through traditional survey methods. To address this challenge, a new method has been developed that utilizes the digital publics' communicated content.

As this dissertation proposes, this method can offer several advantages over traditional survey methods, including high ecological validity, natural interactions, and actual communicated digital content. It also enables the evaluation of dynamic processes that were previously difficult to study. Despite the availability of many social media analytics tools in the market, most being atheoretical leads to a divorce between theory and methods. Therefore, a theory-loaded automated evaluation of the publics' minds must ensure these tools are grounded in sound theoretical foundations. This pathfinding approach is crucial to enhancing the theoretical foundation of most current data analytics tools and can ultimately lead to more effective communication research.

The proposed model in this study can be a powerful tool that enables social researchers and corporate managers to analyze massive amounts of digital publics' comments quickly and effectively. For example, Starbucks had a crisis because of a labor issue. Complaints were filed with the National Labor

Relations Board (NLRB) in 2022 by workers in Buffalo, New York, who claimed Starbucks unlawfully terminated them for their involvement in the organization of unions. The workers alleged that these terminations were retaliatory measures to discourage unionization (Sainato, 2022). This crisis ruined Starbucks' reputation, and many social media posts blamed the company. In response to this crisis, the company had to sort through over two million digital texts to identify the public's perception of the company's relationship. It was a massive task, but the current project's suggested tool can significantly reduce the data, making it more manageable and useful for analysis.

The study's outcome also has practical applications in assessing employee relations. By classifying comments based on control mutuality and commitment, organizations can gain valuable insights into employee attitudes and identify areas that require attention. NLP and topic modeling techniques can extract themes and keywords related to relational causes and consequences, providing management with valuable intelligence to guide strategic actions. Traditionally, this type of analysis would take days or weeks to complete, but with this method, it can be done in just minutes. This system bridges the gap between theoretical foundations and data analytics, allowing for a more automated and efficient assessment of relationships with the digital publics. Ultimately, this approach provides organizations with the tools to make informed decisions and proactively improve stakeholder relationships.

However, this study ironically presents that ML could be imperfect. ML necessitates the coding of labeled data by human coders. The evaluation and interpretation of results are only possible with human experts. While it can certainly be useful tools, they are not a replacement for human intelligence. This project proves that humans are still an integral part of the process. Rather than seeing ML as a threat to intellectualism, we should view it as a way to complement and enhance human intelligence. It can create a powerful and effective system that benefits everyone involved.

In conclusion, I want to highlight that this study is not solely focused on improving the accuracy of predictions. The real value of this research lies in providing tools that empower the digital publics' voices. Unlike existing tools that disregard the values of public opinion, this study aims to gain a deeper and more refined understanding of their thoughts and minds. This approach highlights the significance of social listening and the inclusivity of public opinion in creating a truly effective system. By incorporating the voices of the publics, we can create a system that benefits everyone involved and truly represents the diverse perspectives of our society.

Limitations

The current study in public relations research has certain limitations due to ML being a novel approach. Firstly, the data used in the study had various formats, including scraped data and survey data with different response styles. This diversity in data formats could create challenges for effective ML. Moreover, the data quality could be compromised, as human coders coded the scraped data, while the survey data consisted of open comments based on respondent ratings.

Secondly, there was an issue due to the datasets being small to medium-sized. For instance, Study 2 on employees (with around N = 200) produced poorer results, which raised concerns about generalizability. Moreover, there are distinct relationship dynamics with organizations that could be contributing to the issue. Because of the limited sample size, the training and testing dataset ratio was too small, and there was no validation dataset splitting. In the future, if the sample size cannot be larger, other techniques, such as bootstrapping, can be applied to overcome these challenges and ensure better accuracy levels. Moreover, to enhance the utility, future research could include a diverse range of publics, including employees and other relevant groups.

Thirdly, concerns can be expressed about the practicality of binary outputs in real-world applications. A higher number of evaluations would likely be more meaningful in assessing relationships, as it can capture the dynamics of the relationship rather than simply positive or negative tones of emotions in sentiment analysis. Despite this, it is important to note that this study can still adequately capture the relationship dynamics even with four binary weights, resulting in 16 combinations. Additionally, organizations often need to pay close attention to negative assessments.; future studies could employ interval scales and differentiate between positive and negative evaluations.

Finally, when evaluating imbalanced datasets, it is important to reconsider the metrics used for evaluation. Most labels in these datasets were often imbalanced, leading to inaccurate evaluations. Different metrics such as precision, recall, F-1 score, and receiver operating characteristic (ROC) curve can be used to provide a more accurate evaluation to address this issue. By using these metrics, we can better capture the nuances of the dataset and ensure that our evaluations are more meaningful.

Briefly, the current studies using ML in public relations research face limitations related to the data format, dataset size, and the use of binary outputs. Future research should address these limitations by considering diverse data sources, including employees and other publics, using interval scales, and differentiating between positive and negative evaluations.

Future Research

Firstly, future studies should collect more data. The results indicate that the OPRA features require more input. However, future research could consider using interval (or continuous) scales because they can capture a variety of relationship assessments. In this regard, it would be necessary to differentiate between positive and negative evaluations (e.g., trust vs. distrust, satisfaction vs. dissatisfaction).

There are some effective solutions to this problem. For example, GitHub and crowdsourcing platforms can be utilized to collect data with a clear description of the project's purpose and theoretical background. It can be applied to the human-coding process. Furthermore, generative AI, such as ChatGPT, can be actively adopted to replace the time-consuming and costly human coding process.

Lastly, more case studies with different datasets and correction work with multiple coders should be done to upgrade learning, as the case study was conducted only by a single person (myself). If possible, future studies could include a more diverse range of publics, including voters, fans, and other groups, to study more dynamic and unique relationships.

Vision: Theory-Enhanced Intelligence Model of Automated Relationship Assessment (TIARA)

The ultimate goal of the current project is to develop an automated analytical model named the Theory-enhanced Intelligence Model of Automated Relationship Assessment (TIARA) that will capture the digital publics' voice regarding its relationships with organizations. This model will serve as foundational work for building the entire system. TIARA will leverage theory-enhanced automation to provide situational awareness of digital publics' relationships. This model is expected to enable social researchers and corporate managers to effectively handle the vast amount, speed, and often meaningless nature of big data derived from public comments in the digital space. TIARA will employ NLP, topic modeling, and keyword extraction methods to extract themes and keywords related to relational causes and consequences, thereby offering management intelligence for organizations.

This project holds significant methodological and practical implications. Utilizing the direct voices of the digital publics will increase the ecological validity of the data and overcome the limitations of survey methods in studying the digital publics. Moreover, the developed method will enhance the theoretical foundation in data analytics within the public relations field, which often needs to be improved in traditional ML approaches. The emergence of digital communication has necessitated advanced approaches to comprehending digital communication behaviors, particularly in the context of relationships between organizations and the publics. TIARA will serve as a system that empowers social researchers and corporate managers to effectively navigate the vast size, rapid pace, and often ambiguous nature of big data from public comments in the digital realm, providing valuable management intelligence for organizations.

As Figure 8 suggests, I designed the following steps to enable the development of the TIARA system. After collecting sufficient data, I will go through more corrections and move to the deep-learning phase. TIARA should also contain visualization to help users better understand the results. This project is part of a long-term collaboration. This initial finding can be a foundation for future public relations scholarship and practice.

[Insert Figure 8 here]

Applications

I offer several possible study subjects by taking advantage of TIARA. In recent years, text analysis has gained popularity as a method for measuring relationships in different communication contexts. This approach has been applied to political, environmental, employee, and fandom communication areas. For example, YouTube comments can be analyzed to assess voters' relationships with political candidates, while comments on Regulations.gov can measure stakeholders' relationships with agencies in environmental communication. Glassdoor and Indeed comments can be used to evaluate employees' relationships with their workplaces in employee communication.

Text analysis is also applicable in the entertainment and sports industry to assess the relationships between artists/sports stars and their fans. For instance, fandoms of famous football players like Messi and Ronaldo can be compared, or the fandoms of K-pop groups like BTS and Blackpink can be analyzed. This approach can also be used in cultural diplomacy to compare the fandoms of cities like Seattle, Madrid, and Seoul using Twitter/Instagram posts and comments.

Furthermore, Tian et al. (2023) recently employed an ML approach to study corporate sustainability. They emphasized the importance of sustainable development and corporate sustainability in promoting the sustainable growth of the economy and society. Their study focused on quantifying the disclosure of corporate sustainability information in China using text mining techniques and constructing an environmental characteristics dictionary. The constructed dataset was validated and found to be valid and reliable, enabling investigation into the progress and impacts of sustainable corporate development.

These examples demonstrate how using ML and theoretical concepts of public relations scholarship can effectively measure relationships between entities in various communication contexts. It allows for insights into public opinions, stakeholder engagement, and the assessment of sustainable practices.

Conclusion

The traditional survey method faces challenges due to the massive volumes and velocity of communication in today's increasingly digital environment. To address this, a new method has been developed that utilizes the digital publics' communications to gain insights into attitudes and opinions. This approach can be utilized as an automated tool that allows social researchers and corporate managers to quickly and effectively analyze massive amounts of comments from the digital publics. It provides practical applications in assessing employee relations and identifying areas that require attention. However, the current studies on this method have limitations, including data quality, small dataset size, and limited output usefulness. Future research can expand the application of this approach to different communication contexts and expand the analysis of the disclosure of corporate sustainability information.

References

Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, *46*(3), 175-185. https://doi.org/10.1080/00031305.1992.10475879

Alufaisan, Y., Marusich, L. R., Bakdash, J. Z., Zhou, Y., & Kantarcioglu, M. (2021, May). Does explainable artificial intelligence improve human decision-making?. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 35, No. 8, pp. 6618-6626). https://doi.org/10.1609/aaai.v35i8.16819

Amara, M., Ghedira, K., Zidi, K., & Zidi, S. (2015, November). A comparative study of multi-class support vector machine methods for Arabic characters recognition. In *2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA)* (pp. 1-6). IEEE. https://doi.org/10.1109/AICCSA.2015.7507146

Antonacopoulou, E. P., & Sheaffer, Z. (2014). Learning in crisis: Rethinking the relationship between organizational learning and crisis management. *Journal of Management Inquiry*, *23*(1), 5-21. https://doi-org.ezproxy.lib.ou.edu/10.1177/1056492612472730

Baskin, O. W., Aronoff, C. E., & Lattimore, D. (1997). *Public relations: The profession and the practice*. McGraw-Hill Humanities, Social Sciences & World Languages.

Bastos, M., & Mercea, D. (2018). Parametrizing Brexit: mapping Twitter political space to parliamentary constituencies. *Information, Communication & Society*, *21*(7), 921-939. https://doi.org/10.1080/1369118X.2018.1433224

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, *35*(8), 1798-1828. https://doi.org/10.1109/TPAMI.2013.50

Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4, No. 4, p. 738). Springer.

Boyd, D. (2014). *It's complicated: The social lives of networked teens*. Yale University Press.

Breiman, L. (2001). Random forests. *Machine learning, 45*(1), 5-32. https://doi.org/10.1023/A:1010933404324

Bruning, S. D., DeMiglio, P. A., & Embry, K. (2006). Mutual benefit as outcome indicator: Factors influencing perceptions of benefit in organization–public relationships. *Public Relations Review*, 32(1), 33–40. https://doi.org/10.1016/j.pubrev.2005.10.005

Cameron, G. T., Cropp, F., & Reber, B. H. (2001). Getting past platitudes: Factors limiting accommodation in public relations. *Journal of Communication Management, 5*(3), 242-261. https://doi.org/10.1108/13632540110806802

Castells, M. (2007). Communication, power and counter-power in the network society. *International Journal of Communication*, *1*(1), 29-45. https://doi.org/10.1016/j.hrmr.2012.09.004

Castells, M. (2010). Communication power: Mass communication, mass self-communication, and power relationships in the network society. *Media and Society*, *25*(5), 3-17. https://doi.org/10.1177/016344371002500501

Castells, M. (2012). *Networks of outrage and hope: Social movements in the Internet age*. John Wiley & Sons.

Chapelle, O., Scholkopf, B., & Zien, A. (2009). Semi-supervised learning (chapelle, o. et al., eds.; 2006) [book reviews]. *IEEE Transactions on Neural Networks*, *20*(3), 542-542. https://doi.org/10.1109/TNN.2009.2015974

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS quarterly*, 1165-1188. https://doi.org/10.2307/41703503

Cheng, Y. (2018). Looking back, moving forward: A review and reflection of the organization-public relationship (OPR) research. *Public Relations Review*, *44*(1), 120-130. https://doi.org/10.1016/j.pubrev.2017.10.003

Coombs, W. T. (2021). *Ongoing crisis communication: Planning, managing, and responding*. Sage publications.

Cutlip, S. M., Center, A. H., & Broom, G. M. (1979, 1994, 2013). *Effective public relations 7th edition.* Prentice Hall.

Dang-Xuan, L., Stieglitz, S., Wladarsch, J., & Neuberger, C. (2013). An investigation of influentials and the role of sentiment in political communication on Twitter during election periods. *Information, communication & society*, *16*(5), 795-825. https://doi.org/10.1080/1369118X.2013.783608

Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review*, *96*(1), 108-116. https://hbr.org/webinar/2018/02/artificial-intelligence-for-the-real-world

Demszky, D., Movshovitz-Attias, D., Ko, J., Cowen, A., Nemade, G., & Ravi, S. (2020). GoEmotions: A dataset of fine-grained emotions. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics.* https://doi.org/10.48550/arXiv.2005.00547

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Kent, M. L., & Taylor, M. (2002). Toward a dialogic theory of public relations. *Public relations review*, *28*(1), 21-37. https://doi.org/10.1016/S0363-8111(02)00108-X

Galloway, C., & Swiatek, L. (2018). Public relations and artificial intelligence: It's not (just) about robots. *Public Relations Review*, *44*(5), 734-740. https://doi.org/10.1016/j.pubrev.2018.10.008

García, S., Luengo, J., & Herrera, F. (2015). *Data preprocessing in data mining*. Springer.

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, *63*, 3-42. https://doi.org/10.1007/s10994-006-6226-1

Grunig, J. E. (1992). *Excellence in public relations and communication management*. Routledge.

Grunig, J. E. (1997). A situational theory of publics: Conceptual history, recent challenges and new research. In D. Moss, T. MacManus, & D. Verčič. (Eds). *Public relations research: An international perspective* (pp. 3-48). International Thomson Business Press.

Grunig, J. E. (2023). Public relations, social inclusion, and social exclusion. *Journalism & Communication Monographs, 25*(2), 90–108.

Grunig, J. E. (2009). Paradigms of global public relations in an age of digitalisation. *PRism, 6(2*), 1-19. https://www.prismjournal.org/uploads/1/2/5/6/125661607/v6-no2-a1.pdf

Grunig, J. E., & Dozier, D. M. (2003). *Excellent public relations and effective organizations: A study of communication management in three countries*. Routledge.

Grunig, J. E., & Grunig, L. A. (2000). Public relations in strategic management and strategic management of public relations: Theory and evidence from the IABC Excellence project. *Journalism studies*, *1*(2), 303-321. https://doi.org/10.1080/14616700050028271

Grunig, J. E., & Grunig, L. A. (2001). *Guidelines for formative and evaluative research in public affairs: A report for the Department of Energy Office of Science.* University of Maryland Department of Communication. http://www.instituteforpr.org/wp-content/uploads/Guidelines-for-Formative-and-Evaluative-Research-in-Public-Affairs.pdf

Grunig, J. E., & Grunig, L. A. (2013). Models of public relations and communication. In L. A. Grunig & J. Grunig (Eds.), *Excellence in public relations and communication management* (pp. 285-325). Routledge.

Grunig, J. E., & Hunt, T. (1984). *Managing public relations.* Holt, Rinehart and Winston.

Grunig, J. E., Kim, J.–N., & Lee, H. (2021). Paradigms of public relations public relations in an age of digitalization. *Hongbohak- Yeongu, 25*(1), 1-30. http://doi.org/10.15814/jpr.2021.25.1.1

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G. Z. (2019). XAI—Explainable artificial intelligence. *Science robotics*, *4*(37), eaay7120. http://doi.org/10.1126/scirobotics.aay7120

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer.

Holzinger, A., Biemann, C., Pattichis, C. S., & Kell, D. B. (2017). What do we need to build explainable ML systems for the medical domain?. *arXiv preprint arXiv:1712.09923*.

Holzinger, A., Saranti, A., Molnar, C., Biecek, P., & Samek, W. (2022, April). Explainable ML methods-a brief overview. In *ML-Beyond Explainable ML: International Workshop, Held in Conjunction* *with ICML 2020, July 18, 2020, Vienna, Austria, Revised and Extended Papers* (pp. 13-38). Cham: Springer International Publishing.

Hon, L. & Grunig, J. E. (1999). *Guidelines for Measuring Relationships in Public Relations.* IPR. https://www.instituteforpr.org/wp-content/uploads/Guidelines_Measuring_Relationships.pdf

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression*. John Wiley & Sons.

Hristea, F. T. (2012). *The Naïve Bayes model for unsupervised word sense disambiguation: aspects concerning feature selection*. Springer Science & Business Media. https://doi.org/10.48550/arXiv.2005.00547

Huang, Y. H. (2001). OPRA: A cross-cultural, multiple-item scale for measuring organization-public relationships. *Journal of Public Relations Research*, *13*(1), 61-90. https://doi.org/10.1207/S1532754XJPRR1301_4

Huang, Y. H. C., & Zhang, Y. (2013). Revisiting organization–public relations research over the past decade: Theoretical concepts, measures, methodologies and challenges. *Public Relations Review*, *39*(1), 85-87. https://doi.org/10.1016/j.pubrev.2012.10.001

Janssen, L. (2022l, Jan 22). Data scientists: Who they are and what they do. *Forbes.* https://www.forbes.com/sites/louiscolumbus/2022/01/22/data-scientists-who-they-are-what-they-do/?sh=7c2daa2e2f63

Ji, Y. G., Chen, Z. F., Tao, W., & Li, Z. C. (2019). Functional and emotional traits of corporate social media message strategies: Behavioral insights from S&P 500 Facebook data. *Public relations review*, *45*(1), 88-103. https://doi.org/10.1016/j.pubrev.2018.12.001

Joseph, V. R. (2022). Optimal ratio for data splitting. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, *15*(4), 531-538. https://doi-org.ezproxy.lib.ou.edu/10.1002/sam.11583

Kent, M. L., & Taylor, M. (2002). Toward a dialogic theory of public relations. *Public Relations Review*, *28*(1), 21-37. https://doi.org/10.1016/S0363-8111(02)00108-X

Ki, E. J., & Hon, L. (2009). Causal linkages between relationship cultivation strategies and relationship quality outcomes. *International Journal of Strategic Communication, 3*(4), 242-263. https://doi.org/10.1080/15531180903218630

Kim, J.-N., & Rhee, Y. (2011). Strategic thinking about employee communication behavior (ECB) in public relations: Testing the models of megaphoning and scouting effects in Korea. *Journal of Public Relations Research*, *23*(3), 243-268. https://doi.org/10.1080/1062726X.2011.582204

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big data & Society*, *1*(1), 2053951714528481. https://doi.org/10.1177/2053951714528481

Kitchin, R. (2015). Thinking critically about and researching algorithms. *Information, Communication & Society*, *20*(1), 14-29. https://doi.org/10.1080/1369118X.2016.1154087

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84-90. https://doi.org/10.1145/3065386

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and brain sciences*, *40*, e253. https://doi.org/10.1017/S0140525X16001837

Ledingham, J. A. (2003). Explicating relationship management as a general theory of public relations. *Journal of Public Relations Research*, *15*(2), 181-198. https://doi.org/10.1207/S1532754XJPRR1502_4

Ledingham, J. A., & Bruning, S. D. (1998). Relationship management in public relations: Dimensions of an organization-public relationship. *Public relations review*, *24*(1), 55-65. https://doi.org/10.1016/S0363-8111(98)80020-9

Ledingham, J. A., & Bruning, S. D. (2000). *Public relations as relationship management: A relational approach to the study and practice of public relations*. Routledge.

Lee, Y., & Kim, J.-N. (2021). On evolving nature of relationship by perspective mutuality: reconceptualizing relationship typology between organization and its publics. *Journalism & Mass Communication Quarterly*, *98(1), 148-178*. https://doi.org/10.1177/1077699020949255

Marcus, G. (2018). Deep learning: A critical appraisal. *arXiv preprint arXiv:1801.00631*.

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, *5*(4), 1093-1113. https://doi.org/10.1016/j.asej.2014.04.011

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Panda, G., Upadhyay, A. K., & Khandelwal, K. (2019). Artificial intelligence: A strategic disruption in public relations. *Journal of Creative Communications*, *14*(3), 196-213. https://doi.org/10.1177/0973258619866585

Papacharissi, Z. (Ed.). (2010). *A networked self: Identity, community, and culture on social network sites*. Routledge.

Penn, J. (n.d.). *From novice to practitioner: A critical appraisal of social media measurement for public relations practice*. Unpublished manuscript. University of Greenwich, London.

Pennington, J., Socher, R., & Manning, C. D. (2014, October). GloVe: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532-1543).

Rhee, Y., & Moon, B. (2009). Organizational culture and strategic communication practice: Testing the competing values model (CVM) and employee communication strategies (ECS) model in Korea. *International Journal of Strategic Communication*, *3*(1), 52-67. https://doi.org/10.1080/15531180802608386

Rheingold, H. (2008). Using participatory media and public voice to encourage civic engagement. In Davidson, C. N., & Goldberg, D. T. (Eds.), *The Future of Learning Institutions in a Digital Age* (pp. 99-108). The MIT Press.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206-215. https://doi.org/10.1038/s42256-019-0048-x

Sainato, M. (2022, Oct 16). More workers say Starbucks retaliated for union efforts: 'I lost everything when they fired me'. *The Guardians*. https://www.theguardian.com/us-news/2022/oct/16/fired-workers-starbucks-retaliation-union-organising-nlrb

Santa Soriano, A., & Valdés, R. M. T. (2021). Engaging universe 4.0: The case for forming a public relations-strategic intelligence hybrid. *Public Relations Review*, *47*(2), 102035. https://doi.org/10.1016/j.pubrev.2021.102035

Seltzer, T., Gardner, E., Bichard, S., & Callison, C. (2012). public relations in the ER: Managing internal organization–public relationships in a hospital emergency department. *Public Relations Review*, *38*(1), 128-136. https://doi.org/10.1016/j.pubrev.2011.12.002

Shen, H. (2017). Refining organization–public relationship quality measurement in student and employee samples. *Journalism & Mass Communication Quarterly*, *94*(4), 994-1010. https://doi.org/10.1177/1077699016674186

Shi, X., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., & Woo, W. C. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In Advances in neural information processing systems (pp. 802-810). https://doi.org/10.48550/arXiv.1506.04214

Shirky, C. (2010). *Cognitive surplus: Creativity and generosity in a connected age*. Penguin UK.

Sparck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, *28*(1), 11-21. https://doi.org/10.1108/eb026526

Stacks, D. (2011). *Primer of public relations research* (2^{nd}). Guilford.

Sunstein, C. R. (2018). *# Republic: Divided democracy in the age of social media*. Princeton University Press.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Suykens, J. A., & Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural Processing Letters*, *9*, 293-300. https://doi.org/10.1023/A:1018628609742

Tian, J., Cheng, Q., Xue, R., Han, Y., & Shan, Y. (2023). A dataset on corporate sustainability disclosure. Scientific Data, 10(1), 182. https://doi.org/10.1038/s41597-023-02093-3

Valkenburg, P. M. (2017). Understanding self-effects in social media. *Human Communication Research*, *43*(4), 477-490. https://doi.org/10.1111/hcre.12113

Wang, Y., Cheng, Y., & Sun, J. (2021). When public relations meets social media: A systematic review of social media related public relations research from 2006 to 2020. *Public Relations Review*, *47*(4), 102081. https://doi.org/10.1016/j.pubrev.2021.102081

Wilcox, D. L., Ault, P. H., & Agee, W. K. (2006). *Public relations: Strategies and tactics.* Longman Pub Group.

Wise, A. F., & Shaffer, D. W. (2015). Why theory matters more than ever in the age of big data. *Journal of Learning Analytics*, *2*(2), 5-13. https://doi.org/10.18608/jla.2015.22.2

Xie, Z., & Mu, F. (2019, July). Distributed representation of words in cause and effect spaces. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 33, No. 01, pp. 7330-7337). https://doi.org/10.1609/aaai.v33i01.33017330

Yang, J., & Leskovec, J. (2011, February). Patterns of temporal variation in online media. In *Proceedings of the fourth ACM international conference on Web search and data mining* (pp. 177-186). https://doi.org/10.1145/1935826.1935863

Zhao, X., Zhan, M., & Ma, L. (2020). How publics react to situational and renewing organizational responses across crises: Examining SCCT and DOR in social-mediated crises. *Public Relations Review*, *46*(4), 101944. https://doi.org/10.1016/j.pubrev.2020.101944

Table 1.
Accuracy calculation

		Actual	
		Positive (value=1)	Negative (value=0)
Predicted	Positive (value=1)	**True Positive (TP)**	False Positive (FP)
	Negative (value=0)	False Negative (FN)	**True Negative (TN)**

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$

Table 2.

Results (Control Mutuality, Study1- Amazon data)

Pre-processing data 0					Pre-processing data 1					Pre-processing data 2					Pre-processing data 3				
emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff
BERT	SVM	0.86	0.94	-0.08	BERT	SVM	0.86	0.94	-0.08	BERT	SVM	0.86	0.94	-0.08	BERT	SVM	0.86	0.94	-0.08
BERT	RF	1.00	0.94	0.06	BERT	RF	1.00	0.94	0.06	BERT	RF	1.00	0.94	0.06	BERT	RF	1.00	0.94	0.06
BERT	LR	0.98	0.92	0.06	BERT	LR	1.00	0.92	0.08	BERT	LR	0.97	0.94	0.03	BERT	LR	1.00	0.94	0.06
BERT	KNN	0.87	0.94	-0.07	**BERT**	**KNN**	**0.92**	**0.90**	**0.02**	BERT	KNN	0.90	0.92	-0.02	BERT	KNN	0.90	0.92	-0.02
BERT	EXT	1.00	0.94	0.06	BERT	EXT	1.00	0.96	0.04	BERT	EXT	1.00	0.94	0.06	BERT	EXT	1.00	0.94	0.06
Glove	SVM	0.86	0.94	-0.08	Glove	SVM	0.86	0.94	-0.08	Glove	SVM	0.86	0.94	-0.08	Glove	SVM	0.86	0.94	-0.08
Glove	RF	1.00	0.94	0.06	Glove	RF	1.00	0.94	0.06	Glove	RF	1.00	0.92	0.08	Glove	RF	1.00	0.96	0.04
Glove	LR	0.86	0.94	-0.08	Glove	LR	0.86	0.94	-0.08	Glove	LR	0.86	0.94	-0.08	Glove	LR	0.86	0.94	-0.08
Glove	KNN	0.89	0.94	-0.05	Glove	KNN	0.88	0.92	-0.04	Glove	KNN	0.90	0.88	0.02	Glove	KNN	0.89	0.82	0.07
Glove	EXT	1.00	0.94	0.06	Glove	EXT	1.00	0.96	0.04	Glove	EXT	1.00	0.94	0.06	Glove	EXT	1.00	0.94	0.06
TF-IDF	SVM	0.93	0.94	-0.01	TF-IDF	SVM	0.92	0.94	-0.02	TF-IDF	SVM	0.92	0.94	-0.02	TF-IDF	SVM	0.92	0.94	-0.02
TF-IDF	RF	1.00	0.94	0.06	TF-IDF	RF	1.00	0.94	0.06	TF-IDF	RF	1.00	0.94	0.06	TF-IDF	RF	1.00	0.94	0.06
TF-IDF	LR	0.86	0.94	-0.08	TF-IDF	LR	0.86	0.94	-0.08	TF-IDF	LR	0.86	0.94	-0.08	TF-IDF	LR	0.86	0.94	-0.08
TF-IDF	**KNN**	**0.93**	**0.94**	**-0.01**	**TF-IDF**	**KNN**	**0.93**	**0.94**	**-0.01**	**TF-IDF**	**KNN**	**0.93**	**0.90**	**0.03**	**TF-IDF**	**KNN**	**0.92**	**0.92**	**0.00**
TF-IDF	EXT	1.00	0.94	0.06	TF-IDF	EXT	1.00	0.94	0.06	TF-IDF	EXT	1.00	0.94	0.06	TF-IDF	EXT	1.00	0.94	0.06
Word2Vec	SVM	0.86	0.94	-0.08	Word2Vec	SVM	0.86	0.94	-0.08	Word2Vec	SVM	0.86	0.94	-0.08	Word2Vec	SVM	0.86	0.94	-0.08
Word2Vec	RF	1.00	0.94	0.06	Word2Vec	RF	1.00	0.94	0.06	Word2Vec	RF	1.00	0.94	0.06	Word2Vec	RF	1.00	0.94	0.06
Word2Vec	LR	0.86	0.94	-0.08	Word2Vec	LR	0.86	0.94	-0.08	Word2Vec	LR	0.86	0.94	-0.08	Word2Vec	LR	0.86	0.94	-0.08
Word2Vec	KNN	0.88	0.94	-0.06	Word2Vec	KNN	0.88	0.90	-0.02	Word2Vec	KNN	0.92	0.88	0.04	Word2Vec	KNN	0.88	0.94	-0.06
Word2Vec	EXT	1.00	0.94	0.06	Word2Vec	EXT	1.00	0.94	0.06	Word2Vec	EXT	1.00	0.94	0.06	Word2Vec	EXT	1.00	0.94	0.06

Pre-processing data 4					Pre-processing data 5					Pre-processing data 6					Pre-processing data 7				
emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff
BERT	SVM	0.86	0.94	-0.08	BERT	SVM	0.86	0.94	-0.08	BERT	SVM	0.86	0.94	-0.08	BERT	SVM	0.86	0.94	-0.08
BERT	RF	1.00	0.94	0.06	BERT	RF	1.00	0.96	0.04	BERT	RF	1.00	0.94	0.06	BERT	RF	1.00	0.94	0.06
BERT	LR	0.98	0.84	0.14	BERT	LR	1.00	0.92	0.08	BERT	LR	0.98	0.94	0.04	BERT	LR	1.00	0.94	0.06
BERT	KNN	0.89	0.94	-0.05	**BERT**	**KNN**	**0.93**	**0.90**	**0.03**	BERT	KNN	0.88	0.90	-0.02	**BERT**	**KNN**	**0.91**	**0.90**	**0.01**
BERT	EXT	1.00	0.94	0.06	BERT	EXT	1.00	0.94	0.06	BERT	EXT	1.00	0.94	0.06	BERT	EXT	1.00	0.94	0.06
Glove	SVM	0.86	0.94	-0.08	Glove	SVM	0.86	0.94	-0.08	Glove	SVM	0.86	0.94	-0.08	Glove	SVM	0.86	0.94	-0.08
Glove	RF	1.00	0.94	0.06	Glove	RF	1.00	0.94	0.06	Glove	RF	1.00	0.94	0.06	Glove	RF	1.00	0.96	0.04
Glove	LR	0.86	0.94	-0.08	Glove	LR	0.86	0.94	-0.08	Glove	LR	0.86	0.94	-0.08	Glove	LR	0.86	0.94	-0.08
Glove	KNN	0.89	0.92	-0.03	Glove	KNN	0.88	0.92	-0.04	**Glove**	**KNN**	**0.90**	**0.90**	**0.00**	**Glove**	**KNN**	**0.88**	**0.84**	**0.04**
Glove	EXT	1.00	0.94	0.06	Glove	EXT	1.00	0.96	0.04	Glove	EXT	1.00	0.94	0.06	Glove	EXT	1.00	0.94	0.06
TF-IDF	SVM	0.93	0.94	-0.01	TF-IDF	SVM	0.92	0.94	-0.02	TF-IDF	SVM	0.92	0.94	-0.02	TF-IDF	SVM	0.92	0.94	-0.02
TF-IDF	RF	1.00	0.94	0.06	TF-IDF	RF	1.00	0.94	0.06	TF-IDF	RF	1.00	0.94	0.06	TF-IDF	RF	1.00	0.94	0.06
TF-IDF	LR	0.86	0.94	-0.08	TF-IDF	LR	0.86	0.94	-0.08	TF-IDF	LR	0.86	0.94	-0.08	TF-IDF	LR	0.86	0.94	-0.08
TF-IDF	**KNN**	**0.93**	**0.94**	**-0.01**	TF-IDF	KNN	0.92	0.94	-0.02	**TF-IDF**	**KNN**	**0.92**	**0.90**	**0.02**	**TF-IDF**	**KNN**	**0.92**	**0.92**	**0.00**
TF-IDF	EXT	1.00	0.94	0.06	TF-IDF	EXT	1.00	0.94	0.06	TF-IDF	EXT	1.00	0.94	0.06	TF-IDF	EXT	1.00	0.94	0.06
Word2Vec	SVM	0.86	0.94	-0.08	Word2Vec	SVM	0.86	0.94	-0.08	Word2Vec	SVM	0.86	0.94	-0.08	Word2Vec	SVM	0.86	0.94	-0.08
Word2Vec	RF	1.00	0.94	0.06	Word2Vec	RF	1.00	0.94	0.06	Word2Vec	RF	1.00	0.94	0.06	Word2Vec	RF	1.00	0.94	0.06
Word2Vec	LR	0.86	0.94	-0.08	Word2Vec	LR	0.86	0.94	-0.08	Word2Vec	LR	0.86	0.94	-0.08	Word2Vec	LR	0.86	0.94	-0.08
Word2Vec	KNN	0.90	0.94	-0.04	Word2Vec	KNN	0.89	0.92	-0.03	**Word2Vec**	**KNN**	**0.90**	**0.90**	**0.00**	Word2Vec	KNN	0.88	0.94	-0.06
Word2Vec	EXT	1.00	0.94	0.06	Word2Vec	EXT	1.00	0.94	0.06	Word2Vec	EXT	1.00	0.94	0.06	Word2Vec	EXT	1.00	0.94	0.06

Note: The cell(s) highlighted in yellow with red colored text indicate the best performing case(s). The cell(s) highlighted in orange indicate the second best performing case(s).

Table 3.

Results (Commitment, Study1- Amazon data)

Pre-processing data 0					Pre-processing data 1					Pre-processing data 2					Pre-processing data 3				
emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff
BERT	SVM	0.84	0.84	0.00	BERT	SVM	0.84	0.84	0.00	BERT	SVM	0.84	0.84	0.00	BERT	SVM	0.84	0.84	0.00
BERT	RF	1.00	0.84	0.16	BERT	RF	1.00	0.84	0.16	BERT	RF	1.00	0.82	0.18	BERT	RF	1.00	0.84	0.16
BERT	LR	0.97	0.86	0.11	BERT	LR	1.00	0.86	0.14	BERT	LR	0.98	0.86	0.12	BERT	LR	1.00	0.88	0.12
BERT	KNN	0.85	0.80	0.05	BERT	KNN	0.90	0.86	0.04	BERT	KNN	0.88	0.80	0.08	BERT	KNN	0.86	0.80	0.06
BERT	EXT	1.00	0.84	0.16	BERT	EXT	1.00	0.84	0.16	BERT	EXT	1.00	0.84	0.16	BERT	EXT	1.00	0.84	0.16
Glove	SVM	0.84	0.84	0.00	Glove	SVM	0.84	0.84	0.00	Glove	SVM	0.84	0.84	0.00	Glove	SVM	0.84	0.84	0.00
Glove	RF	1.00	0.86	0.14	Glove	RF	1.00	0.84	0.16	Glove	RF	1.00	0.86	0.14	Glove	RF	1.00	0.84	0.16
Glove	LR	0.84	0.84	0.00	Glove	LR	0.84	0.84	0.00	Glove	LR	0.84	0.84	0.00	Glove	LR	0.84	0.84	0.00
Glove	KNN	0.90	0.84	0.06	Glove	KNN	0.86	0.82	0.04	Glove	KNN	0.88	0.78	0.10	Glove	KNN	0.87	0.74	0.13
Glove	EXT	1.00	0.86	0.14	Glove	EXT	1.00	0.84	0.16	Glove	EXT	1.00	0.84	0.16	Glove	EXT	1.00	0.84	0.16
TF-IDF	SVM	0.89	0.84	0.05	TF-IDF	SVM	0.90	0.84	0.06	TF-IDF	SVM	0.89	0.84	0.05	TF-IDF	SVM	0.89	0.84	0.05
TF-IDF	RF	1.00	0.84	0.16	TF-IDF	RF	1.00	0.84	0.16	TF-IDF	RF	1.00	0.84	0.16	TF-IDF	RF	1.00	0.84	0.16
TF-IDF	LR	0.84	0.84	0.00	TF-IDF	LR	0.84	0.84	0.00	TF-IDF	LR	0.84	0.84	0.00	TF-IDF	LR	0.84	0.84	0.00
TF-IDF	KNN	0.89	0.88	0.01	TF-IDF	KNN	0.89	0.88	0.01	TF-IDF	KNN	0.89	0.88	0.01	TF-IDF	KNN	0.89	0.90	-0.01
TF-IDF	EXT	1.00	0.84	0.16	TF-IDF	EXT	1.00	0.84	0.16	TF-IDF	EXT	1.00	0.84	0.16	TF-IDF	EXT	1.00	0.84	0.16
Word2Vec	SVM	0.84	0.84	0.00	Word2Vec	SVM	0.84	0.84	0.00	Word2Vec	SVM	0.84	0.84	0.00	Word2Vec	SVM	0.84	0.84	0.00
Word2Vec	RF	1.00	0.84	0.16	Word2Vec	RF	0.99	0.84	0.15	Word2Vec	RF	1.00	0.84	0.16	Word2Vec	RF	0.99	0.84	0.15
Word2Vec	LR	0.84	0.84	0.00	Word2Vec	LR	0.84	0.84	0.00	Word2Vec	LR	0.84	0.84	0.00	Word2Vec	LR	0.84	0.84	0.00
Word2Vec	KNN	0.89	0.82	0.07	Word2Vec	KNN	0.88	0.80	0.08	Word2Vec	KNN	0.91	0.78	0.13	Word2Vec	KNN	0.88	0.82	0.06
Word2Vec	EXT	1.00	0.84	0.16	Word2Vec	EXT	0.99	0.84	0.15	Word2Vec	EXT	1.00	0.84	0.16	Word2Vec	EXT	0.99	0.84	0.15
Pre-processing data 4					Pre-processing data 5					Pre-processing data 6					Pre-processing data 7				
emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff
BERT	SVM	0.84	0.84	0.00	BERT	SVM	0.84	0.84	0.00	BERT	SVM	0.84	0.84	0.00	BERT	SVM	0.84	0.84	0.00
BERT	RF	1.00	0.84	0.16	BERT	RF	1.00	0.84	0.16	BERT	RF	1.00	0.84	0.16	BERT	RF	1.00	0.84	0.16
BERT	LR	0.96	0.80	0.16	BERT	LR	1.00	0.86	0.14	BERT	LR	0.98	0.86	0.12	BERT	LR	1.00	0.86	0.14
BERT	KNN	0.85	0.82	0.03	BERT	KNN	0.90	0.90	0.00	BERT	KNN	0.87	0.80	0.07	BERT	KNN	0.89	0.78	0.11
BERT	EXT	1.00	0.84	0.16	BERT	EXT	1.00	0.86	0.14	BERT	EXT	1.00	0.84	0.16	BERT	EXT	1.00	0.84	0.16
Glove	SVM	0.84	0.84	0.00	Glove	SVM	0.84	0.84	0.00	Glove	SVM	0.84	0.84	0.00	Glove	SVM	0.84	0.84	0.00
Glove	RF	1.00	0.84	0.16	Glove	RF	1.00	0.84	0.16	Glove	RF	1.00	0.84	0.16	Glove	RF	1.00	0.84	0.16
Glove	LR	0.84	0.84	0.00	Glove	LR	0.84	0.84	0.00	Glove	LR	0.84	0.84	0.00	Glove	LR	0.84	0.84	0.00
Glove	KNN	0.89	0.84	0.05	Glove	KNN	0.87	0.82	0.05	Glove	KNN	0.87	0.80	0.07	Glove	KNN	0.87	0.78	0.09
Glove	EXT	1.00	0.86	0.14	Glove	EXT	1.00	0.86	0.14	Glove	EXT	1.00	0.84	0.16	Glove	EXT	1.00	0.84	0.16
TF-IDF	SVM	0.89	0.84	0.05	TF-IDF	SVM	0.90	0.84	0.06	TF-IDF	SVM	0.89	0.84	0.05	TF-IDF	SVM	0.88	0.84	0.04
TF-IDF	RF	1.00	0.84	0.16	TF-IDF	RF	1.00	0.84	0.16	TF-IDF	RF	1.00	0.84	0.16	TF-IDF	RF	1.00	0.84	0.16
TF-IDF	LR	0.84	0.84	0.00	TF-IDF	LR	0.84	0.84	0.00	TF-IDF	LR	0.84	0.84	0.00	TF-IDF	LR	0.84	0.84	0.00
TF-IDF	KNN	0.89	0.88	0.01	TF-IDF	KNN	0.89	0.88	0.01	TF-IDF	KNN	0.89	0.88	0.01	TF-IDF	KNN	0.89	0.90	-0.01
TF-IDF	EXT	1.00	0.84	0.16	TF-IDF	EXT	1.00	0.84	0.16	TF-IDF	EXT	1.00	0.84	0.16	TF-IDF	EXT	1.00	0.84	0.16
Word2Vec	SVM	0.84	0.82	0.02	Word2Vec	SVM	0.84	0.84	0.00	Word2Vec	SVM	0.84	0.84	0.00	Word2Vec	SVM	0.84	0.84	0.00
Word2Vec	RF	1.00	0.84	0.16	Word2Vec	RF	0.99	0.84	0.15	Word2Vec	RF	1.00	0.84	0.16	Word2Vec	RF	0.99	0.84	0.15
Word2Vec	LR	0.84	0.84	0.00	Word2Vec	LR	0.84	0.84	0.00	Word2Vec	LR	0.84	0.84	0.00	Word2Vec	LR	0.84	0.84	0.00
Word2Vec	KNN	0.90	0.84	0.06	Word2Vec	KNN	0.88	0.80	0.08	Word2Vec	KNN	0.91	0.80	0.11	Word2Vec	KNN	0.88	0.80	0.08
Word2Vec	EXT	1.00	0.84	0.16	Word2Vec	EXT	0.99	0.84	0.15	Word2Vec	EXT	1.00	0.84	0.16	Word2Vec	EXT	0.99	0.84	0.15

Note: The cell(s) highlighted in yellow with red colored text indicate the best performing case(s). The cell(s) highlighted in orange indicate the second best performing case(s).

Table 4.

Results (Satisfaction, Study1- Amazon data)

Pre-processing data 0					Pre-processing data 1					Pre-processing data 2					Pre-processing data 3				
emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff
BERT	SVM	0.68	0.62	0.06	BERT	SVM	0.77	0.68	0.09	BERT	SVM	0.68	0.54	0.14	BERT	SVM	0.79	0.66	0.13
BERT	RF	1.00	0.60	0.40	BERT	RF	1.00	0.70	0.30	BERT	RF	1.00	0.62	0.38	BERT	RF	1.00	0.70	0.30
BERT	LR	0.97	0.60	0.37	BERT	LR	1.00	0.68	0.32	BERT	LR	0.96	0.64	0.32	BERT	LR	1.00	0.60	0.40
BERT	KNN	0.82	0.56	0.26	BERT	KNN	0.81	0.60	0.21	BERT	KNN	0.84	0.66	0.18	**BERT**	**KNN**	**0.80**	**0.76**	**0.04**
BERT	EXT	1.00	0.58	0.42	BERT	EXT	1.00	0.64	0.36	BERT	EXT	1.00	0.68	0.32	BERT	EXT	1.00	0.68	0.32
Glove	SVM	0.51	0.44	0.07	Glove	SVM	0.51	0.44	0.07	Glove	SVM	0.51	0.44	0.07	Glove	SVM	0.51	0.44	0.07
Glove	RF	1.00	0.76	0.24	Glove	RF	1.00	0.78	0.22	Glove	RF	1.00	0.74	0.26	Glove	RF	1.00	0.70	0.30
Glove	LR	0.69	0.78	-0.09	Glove	LR	0.68	0.84	-0.16	Glove	LR	0.67	0.70	-0.03	Glove	LR	0.69	0.82	-0.13
Glove	KNN	0.81	0.66	0.15	Glove	KNN	0.80	0.68	0.12	Glove	KNN	0.82	0.62	0.20	Glove	KNN	0.82	0.68	0.14
Glove	EXT	1.00	0.80	0.20	Glove	EXT	1.00	0.76	0.24	Glove	EXT	1.00	0.76	0.24	Glove	EXT	1.00	0.76	0.24
TF-IDF	SVM	0.97	0.78	0.19	TF-IDF	SVM	0.97	0.74	0.23	TF-IDF	SVM	0.97	0.76	0.21	TF-IDF	SVM	0.96	0.76	0.20
TF-IDF	RF	1.00	0.74	0.26	TF-IDF	RF	1.00	0.76	0.24	TF-IDF	RF	1.00	0.76	0.24	TF-IDF	RF	1.00	0.74	0.26
TF-IDF	LR	0.98	0.78	0.20	TF-IDF	LR	0.97	0.76	0.21	TF-IDF	LR	0.98	0.82	0.16	TF-IDF	LR	0.97	0.76	0.21
TF-IDF	KNN	0.88	0.68	0.20	TF-IDF	KNN	0.87	0.68	0.19	TF-IDF	KNN	0.88	0.66	0.22	TF-IDF	KNN	0.88	0.68	0.20
TF-IDF	EXT	1.00	0.82	0.18	TF-IDF	EXT	1.00	0.80	0.20	TF-IDF	EXT	1.00	0.84	0.16	TF-IDF	EXT	1.00	0.82	0.18
Word2Vec	SVM	0.51	0.42	0.09	Word2Vec	SVM	0.51	0.44	0.07	Word2Vec	SVM	0.52	0.42	0.10	Word2Vec	SVM	0.51	0.44	0.07
Word2Vec	RF	1.00	0.76	0.24	Word2Vec	RF	1.00	0.80	0.20	Word2Vec	RF	1.00	0.72	0.28	Word2Vec	RF	1.00	0.74	0.26
Word2Vec	LR	0.62	0.62	0.00	Word2Vec	LR	0.66	0.66	0.00	Word2Vec	LR	0.63	0.62	0.01	Word2Vec	LR	0.68	0.70	-0.02
Word2Vec	KNN	0.78	0.68	0.10	**Word2Vec**	**KNN**	**0.80**	**0.74**	**0.06**	Word2Vec	KNN	0.86	0.66	0.20	Word2Vec	KNN	0.77	0.62	0.15
Word2Vec	EXT	1.00	0.72	0.28	Word2Vec	EXT	1.00	0.70	0.30	Word2Vec	EXT	1.00	0.76	0.24	Word2Vec	EXT	1.00	0.70	0.30
Pre-processing data 4					Pre-processing data 5					Pre-processing data 6					Pre-processing data 7				
emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff
BERT	SVM	0.63	0.70	-0.07	**BERT**	**SVM**	**0.75**	**0.72**	**0.03**	BERT	SVM	0.63	0.76	-0.13	BERT	SVM	0.79	0.66	0.13
BERT	RF	1.00	0.56	0.44	BERT	RF	1.00	0.66	0.34	BERT	RF	1.00	0.62	0.38	BERT	RF	1.00	0.70	0.30
BERT	LR	0.95	0.66	0.29	BERT	LR	1.00	0.64	0.36	BERT	LR	0.96	0.68	0.28	BERT	LR	1.00	0.56	0.44
BERT	KNN	0.82	0.52	0.30	BERT	KNN	0.81	0.68	0.13	BERT	KNN	0.83	0.58	0.25	**BERT**	**KNN**	**0.81**	**0.72**	**0.09**
BERT	EXT	1.00	0.64	0.36	BERT	EXT	1.00	0.64	0.36	BERT	EXT	1.00	0.68	0.32	BERT	EXT	1.00	0.62	0.38
Glove	SVM	0.51	0.44	0.07	Glove	SVM	0.51	0.44	0.07	Glove	SVM	0.51	0.44	0.07	Glove	SVM	0.51	0.44	0.07
Glove	RF	1.00	0.72	0.28	Glove	RF	1.00	0.76	0.24	Glove	RF	1.00	0.68	0.32	Glove	RF	1.00	0.74	0.26
Glove	LR	0.69	0.74	-0.05	Glove	LR	0.69	0.84	-0.15	Glove	LR	0.67	0.76	-0.09	Glove	LR	0.72	0.80	-0.08
Glove	KNN	0.80	0.62	0.18	Glove	KNN	0.80	0.66	0.14	Glove	KNN	0.82	0.64	0.18	Glove	KNN	0.81	0.68	0.13
Glove	EXT	1.00	0.74	0.26	Glove	EXT	1.00	0.76	0.24	Glove	EXT	1.00	0.76	0.24	Glove	EXT	1.00	0.68	0.32
TF-IDF	SVM	0.97	0.76	0.21	TF-IDF	SVM	0.97	0.76	0.21	TF-IDF	SVM	0.97	0.78	0.19	TF-IDF	SVM	0.97	0.76	0.21
TF-IDF	RF	1.00	0.76	0.24	TF-IDF	RF	1.00	0.76	0.24	TF-IDF	RF	1.00	0.82	0.18	TF-IDF	RF	1.00	0.78	0.22
TF-IDF	LR	0.98	0.80	0.18	TF-IDF	LR	0.97	0.78	0.19	TF-IDF	LR	0.98	0.82	0.16	TF-IDF	LR	0.97	0.76	0.21
TF-IDF	KNN	0.90	0.68	0.22	TF-IDF	KNN	0.86	0.70	0.16	TF-IDF	KNN	0.89	0.68	0.21	TF-IDF	KNN	0.89	0.68	0.21
TF-IDF	EXT	1.00	0.82	0.18	TF-IDF	EXT	1.00	0.80	0.20	TF-IDF	EXT	1.00	0.82	0.18	TF-IDF	EXT	1.00	0.86	0.14
Word2Vec	SVM	0.51	0.42	0.09	Word2Vec	SVM	0.51	0.44	0.07	Word2Vec	SVM	0.52	0.42	0.10	Word2Vec	SVM	0.51	0.44	0.07
Word2Vec	RF	1.00	0.78	0.22	Word2Vec	RF	1.00	0.76	0.24	Word2Vec	RF	1.00	0.80	0.20	Word2Vec	RF	1.00	0.76	0.24
Word2Vec	LR	0.62	0.60	0.02	Word2Vec	LR	0.67	0.66	0.01	Word2Vec	LR	0.63	0.62	0.01	Word2Vec	LR	0.70	0.68	0.02
Word2Vec	KNN	0.80	0.64	0.16	**Word2Vec**	**KNN**	**0.78**	**0.74**	**0.04**	Word2Vec	KNN	0.84	0.72	0.12	Word2Vec	KNN	0.77	0.64	0.13
Word2Vec	EXT	1.00	0.76	0.24	Word2Vec	EXT	1.00	0.76	0.24	Word2Vec	EXT	1.00	0.74	0.26	Word2Vec	EXT	1.00	0.76	0.24

Note: The cell(s) highlighted in yellow with red colored text indicate the best performing case(s). The cell(s) highlighted in orange indicate the second best performing case(s).

Table 5.

Results (Trust,, Study1- Amazon data)

Pre-processing data 0					Pre-processing data 1					Pre-processing data 2					Pre-processing data 3				
emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff
BERT	SVM	0.62	0.66	-0.04	BERT	SVM	0.68	0.70	-0.02	BERT	SVM	0.62	0.66	-0.04	BERT	SVM	0.68	0.72	-0.04
BERT	RF	1.00	0.70	0.30	BERT	RF	1.00	0.70	0.30	BERT	RF	1.00	0.70	0.30	BERT	RF	1.00	0.68	0.32
BERT	LR	0.96	0.66	0.30	BERT	LR	1.00	0.60	0.40	BERT	LR	0.97	0.66	0.31	BERT	LR	1.00	0.68	0.32
BERT	KNN	0.78	0.62	0.16	BERT	KNN	0.82	0.62	0.20	BERT	KNN	0.82	0.68	0.14	**BERT**	**KNN**	**0.74**	**0.72**	**0.02**
BERT	EXT	1.00	0.72	0.28	BERT	EXT	1.00	0.66	0.34	BERT	EXT	1.00	0.68	0.32	BERT	EXT	1.00	0.70	0.30
Glove	SVM	0.62	0.66	-0.04	Glove	SVM	0.62	0.66	-0.04	Glove	SVM	0.62	0.66	-0.04	Glove	SVM	0.62	0.66	-0.04
Glove	RF	1.00	0.66	0.34	Glove	RF	1.00	0.72	0.28	Glove	RF	1.00	0.70	0.30	Glove	RF	1.00	0.78	0.22
Glove	LR	0.69	0.70	-0.01	Glove	LR	0.68	0.70	-0.02	Glove	LR	0.67	0.72	-0.05	Glove	LR	0.68	0.72	-0.04
Glove	KNN	0.85	0.60	0.25	Glove	KNN	0.83	0.62	0.21	Glove	KNN	0.84	0.54	0.30	Glove	KNN	0.85	0.56	0.29
Glove	EXT	1.00	0.66	0.34	Glove	EXT	1.00	0.70	0.30	Glove	EXT	1.00	0.66	0.34	Glove	EXT	1.00	0.70	0.30
TF-IDF	SVM	0.95	0.74	0.21	TF-IDF	SVM	0.94	0.72	0.22	TF-IDF	SVM	0.96	0.72	0.24	TF-IDF	SVM	0.95	0.74	0.21
TF-IDF	RF	1.00	0.76	0.24	TF-IDF	RF	1.00	0.76	0.24	TF-IDF	RF	1.00	0.76	0.24	TF-IDF	RF	1.00	0.74	0.26
TF-IDF	LR	0.88	0.70	0.18	TF-IDF	LR	0.89	0.72	0.17	TF-IDF	LR	0.88	0.70	0.18	TF-IDF	LR	0.89	0.72	0.17
TF-IDF	KNN	0.88	0.68	0.20	TF-IDF	KNN	0.87	0.68	0.19	TF-IDF	KNN	0.88	0.66	0.22	TF-IDF	KNN	0.88	0.68	0.20
TF-IDF	EXT	1.00	0.82	0.18	TF-IDF	EXT	1.00	0.80	0.20	TF-IDF	EXT	1.00	0.84	0.16	TF-IDF	EXT	1.00	0.82	0.18
Word2Vec	SVM	0.63	0.66	-0.03	Word2Vec	SVM	0.63	0.66	-0.03	Word2Vec	SVM	0.63	0.66	-0.03	Word2Vec	SVM	0.63	0.66	-0.03
Word2Vec	RF	1.00	0.62	0.38	Word2Vec	RF	1.00	0.66	0.34	Word2Vec	RF	1.00	0.66	0.34	Word2Vec	RF	1.00	0.68	0.32
Word2Vec	LR	0.63	0.68	-0.05	Word2Vec	LR	0.65	0.70	-0.05	Word2Vec	LR	0.63	0.68	-0.05	Word2Vec	LR	0.65	0.70	-0.05
Word2Vec	KNN	0.78	0.60	0.18	Word2Vec	KNN	0.80	0.66	0.14	Word2Vec	KNN	0.85	0.54	0.31	Word2Vec	KNN	0.77	0.66	0.11
Word2Vec	EXT	1.00	0.64	0.36	Word2Vec	EXT	1.00	0.70	0.30	Word2Vec	EXT	1.00	0.66	0.34	Word2Vec	EXT	1.00	0.66	0.34
Pre-processing data 4					Pre-processing data 5					Pre-processing data 6					Pre-processing data 7				
emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff
BERT	SVM	0.62	0.66	-0.04	BERT	SVM	0.69	0.72	-0.03	BERT	SVM	0.62	0.66	-0.04	BERT	SVM	0.68	0.72	-0.04
BERT	RF	1.00	0.70	0.30	BERT	RF	1.00	0.68	0.32	BERT	RF	1.00	0.72	0.28	BERT	RF	1.00	0.70	0.30
BERT	LR	0.96	0.64	0.32	BERT	LR	1.00	0.58	0.42	BERT	LR	0.96	0.66	0.30	BERT	LR	1.00	0.62	0.38
BERT	KNN	0.82	0.58	0.24	BERT	KNN	0.80	0.68	0.12	BERT	KNN	0.82	0.58	0.24	**BERT**	**KNN**	**0.76**	**0.70**	**0.06**
BERT	EXT	1.00	0.72	0.28	BERT	EXT	1.00	0.68	0.32	BERT	EXT	1.00	0.72	0.28	BERT	EXT	1.00	0.70	0.30
Glove	SVM	0.62	0.66	-0.04	Glove	SVM	0.62	0.66	-0.04	Glove	SVM	0.62	0.66	-0.04	Glove	SVM	0.62	0.66	-0.04
Glove	RF	1.00	0.66	0.34	Glove	RF	1.00	0.74	0.26	Glove	RF	1.00	0.64	0.36	Glove	RF	1.00	0.72	0.28
Glove	LR	0.67	0.70	-0.03	Glove	LR	0.67	0.70	-0.03	Glove	LR	0.68	0.72	-0.04	Glove	LR	0.68	0.72	-0.04
Glove	KNN	0.86	0.58	0.28	Glove	KNN	0.83	0.62	0.21	Glove	KNN	0.84	0.60	0.24	Glove	KNN	0.85	0.60	0.25
Glove	EXT	1.00	0.70	0.30	Glove	EXT	1.00	0.72	0.28	Glove	EXT	1.00	0.68	0.32	Glove	EXT	1.00	0.68	0.32
TF-IDF	SVM	0.95	0.74	0.21	TF-IDF	SVM	0.94	0.70	0.24	TF-IDF	SVM	0.96	0.72	0.24	TF-IDF	SVM	0.95	0.72	0.23
TF-IDF	RF	1.00	0.72	0.28	TF-IDF	RF	1.00	0.76	0.24	TF-IDF	RF	1.00	0.80	0.20	TF-IDF	RF	1.00	0.80	0.20
TF-IDF	LR	0.90	0.70	0.20	TF-IDF	LR	0.90	0.72	0.18	TF-IDF	LR	0.89	0.70	0.19	TF-IDF	LR	0.90	0.72	0.18
TF-IDF	KNN	0.90	0.68	0.22	TF-IDF	KNN	0.86	0.70	0.16	TF-IDF	KNN	0.89	0.68	0.21	TF-IDF	KNN	0.89	0.68	0.21
TF-IDF	EXT	1.00	0.82	0.18	TF-IDF	EXT	1.00	0.80	0.20	TF-IDF	EXT	1.00	0.82	0.18	TF-IDF	EXT	1.00	0.86	0.14
Word2Vec	SVM	0.63	0.66	-0.03	Word2Vec	SVM	0.63	0.66	-0.03	Word2Vec	SVM	0.63	0.66	-0.03	Word2Vec	SVM	0.63	0.66	-0.03
Word2Vec	RF	1.00	0.64	0.36	Word2Vec	RF	1.00	0.62	0.38	Word2Vec	RF	1.00	0.64	0.36	Word2Vec	RF	1.00	0.64	0.36
Word2Vec	LR	0.63	0.68	-0.05	Word2Vec	LR	0.66	0.70	-0.04	Word2Vec	LR	0.63	0.68	-0.05	Word2Vec	LR	0.66	0.70	-0.04
Word2Vec	KNN	0.82	0.64	0.18	Word2Vec	KNN	0.81	0.64	0.17	Word2Vec	KNN	0.86	0.64	0.22	Word2Vec	KNN	0.79	0.68	0.11
Word2Vec	EXT	1.00	0.70	0.30	Word2Vec	EXT	1.00	0.68	0.32	Word2Vec	EXT	1.00	0.66	0.34	Word2Vec	EXT	1.00	0.70	0.30

Note: The cell(s) highlighted in yellow with red colored text indicate the best performing case(s). The cell(s) highlighted in orange indicate the second best performing case(s).

Table 6.

A summary of the Amazon data results (Study 1)

Control Mutuality

Pre-processing data 0				
emb	model	cmt_train	cmt_test	diff
TF-IDF	**KNN**	**0.89**	**0.88**	**0.01**
Pre-processing data 1				
TF-IDF	**KNN**	**0.89**	**0.88**	**0.01**
Pre-processing data 2				
TF-IDF	**KNN**	**0.89**	**0.88**	**0.01**
Pre-processing data 4				
TF-IDF	**KNN**	**0.89**	**0.88**	**0.01**
Pre-processing data 5				
TF-IDF	**KNN**	**0.89**	**0.88**	**0.01**
Pre-processing data 6				
TF-IDF	**KNN**	**0.89**	**0.88**	**0.01**

Commitment

Pre-processing data 3				
emb	model	cmt_train	cmt_test	diff
TF-IDF	**KNN**	**0.92**	**0.92**	**0.00**
Pre-processing data 7				
TF-IDF	**KNN**	**0.92**	**0.92**	**0.00**

Satisfaction

Pre-processing data 3				
emb	model	train	test	diff
BERT	**KNN**	**0.80**	**0.76**	**0.04**

Trust

Pre-processing data 3				
emb	model	train	test	diff
BERT	**KNN**	**0.74**	**0.72**	**0.02**

Table 7.

Results (Control Mutuality, Study1- Survey data)

Pre-processing data 0					Pre-processing data 1					Pre-processing data 2					Pre-processing data 3				
emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff
BERT	SVM	0.61	0.65	-0.04	**BERT**	**SVM**	**0.77**	**0.73**	**0.04**	BERT	SVM	0.60	0.64	-0.04	**BERT**	**SVM**	**0.79**	**0.72**	**0.07**
BERT	RF	1.00	0.71	0.29	BERT	RF	1.00	0.71	0.29	BERT	RF	1.00	0.68	0.32	BERT	RF	1.00	0.71	0.29
BERT	LR	0.88	0.74	0.14	BERT	LR	1.00	0.73	0.27	BERT	LR	0.89	0.70	0.19	BERT	LR	1.00	0.70	0.30
BERT	KNN	0.82	0.63	0.19	BERT	KNN	0.86	0.69	0.17	BERT	KNN	0.82	0.66	0.16	BERT	KNN	0.84	0.7	0.14
BERT	EXT	1.00	0.72	0.28	BERT	EXT	1.00	0.73	0.27	BERT	EXT	1.00	0.67	0.33	BERT	EXT	1.00	0.72	0.28
Glove	SVM	0.60	0.64	-0.04	Glove	SVM	0.60	0.64	-0.04	Glove	SVM	0.60	0.64	-0.04	Glove	SVM	0.60	0.64	-0.04
Glove	RF	1.00	0.71	0.29	Glove	RF	1.00	0.66	0.34	Glove	RF	1.00	0.62	0.38	Glove	RF	1.00	0.65	0.35
Glove	LR	0.66	0.60	0.06	Glove	LR	0.67	0.70	-0.03	Glove	LR	0.65	0.64	0.01	Glove	LR	0.66	0.69	-0.03
Glove	KNN	0.81	0.62	0.19	Glove	KNN	0.81	0.59	0.22	Glove	KNN	0.82	0.62	0.20	Glove	KNN	0.81	0.57	0.24
Glove	EXT	1.00	0.67	0.33	Glove	EXT	1.00	0.64	0.36	Glove	EXT	1.00	0.66	0.34	Glove	EXT	1.00	0.62	0.38
TF-IDF	**SVM**	**0.93**	**0.84**	**0.09**	**TF-IDF**	**SVM**	**0.92**	**0.83**	**0.09**	**TF-IDF**	**SVM**	**0.93**	**0.85**	**0.08**	**TF-IDF**	**SVM**	**0.93**	**0.83**	**0.10**
TF-IDF	RF	1.00	0.78	0.22	TF-IDF	RF	1.00	0.81	0.19	TF-IDF	RF	1.00	0.82	0.18	TF-IDF	RF	1.00	0.80	0.20
TF-IDF	**LR**	**0.91**	**0.82**	**0.09**	**TF-IDF**	**LR**	**0.91**	**0.80**	**0.11**	**TF-IDF**	**LR**	**0.91**	**0.81**	**0.10**	**TF-IDF**	**LR**	**0.91**	**0.80**	**0.11**
TF-IDF	**KNN**	**0.90**	**0.79**	**0.11**	**TF-IDF**	**KNN**	**0.90**	**0.77**	**0.13**	**TF-IDF**	**KNN**	**0.90**	**0.80**	**0.10**	**TF-IDF**	**KNN**	**0.90**	**0.79**	**0.11**
TF-IDF	EXT	1.00	0.83	0.17	TF-IDF	EXT	1.00	0.81	0.19	TF-IDF	EXT	1.00	0.81	0.19	TF-IDF	EXT	1.00	0.83	0.17
Word2Vec	SVM	0.60	0.64	-0.04	Word2Vec	SVM	0.60	0.64	-0.04	Word2Vec	SVM	0.60	0.64	-0.04	Word2Vec	SVM	0.60	0.64	-0.04
Word2Vec	RF	1.00	0.66	0.34	Word2Vec	RF	1.00	0.65	0.35	Word2Vec	RF	1.00	0.64	0.36	Word2Vec	RF	1.00	0.67	0.33
Word2Vec	LR	0.61	0.64	-0.03	Word2Vec	LR	0.62	0.64	-0.02	Word2Vec	LR	0.61	0.64	-0.03	Word2Vec	LR	0.62	0.65	-0.03
Word2Vec	KNN	0.80	0.57	0.23	Word2Vec	KNN	0.81	0.56	0.25	Word2Vec	KNN	0.79	0.59	0.20	Word2Vec	KNN	0.80	0.60	0.20
Word2Vec	EXT	1.00	0.65	0.35	Word2Vec	EXT	1.00	0.65	0.35	Word2Vec	EXT	1.00	0.67	0.33	Word2Vec	EXT	1.00	0.65	0.35

Pre-processing data 4					Pre-processing data 5					Pre-processing data 6					Pre-processing data 7				
emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff
BERT	SVM	0.60	0.64	-0.04	**BERT**	**SVM**	**0.78**	**0.77**	**0.01**	BERT	SVM	0.60	0.64	-0.04	**BERT**	**SVM**	**0.79**	**0.77**	**0.02**
BERT	RF	1.00	0.69	0.31	BERT	RF	1.00	0.74	0.26	BERT	RF	1.00	0.70	0.30	BERT	RF	1.00	0.73	0.27
BERT	LR	0.88	0.73	0.15	BERT	LR	1.00	0.74	0.26	BERT	LR	0.89	0.74	0.15	BERT	LR	1.00	0.70	0.30
BERT	KNN	0.83	0.60	0.23	BERT	KNN	0.86	0.67	0.19	BERT	KNN	0.82	0.67	0.15	BERT	KNN	0.86	0.73	0.13
BERT	EXT	1.00	0.69	0.31	BERT	EXT	1.00	0.70	0.30	BERT	EXT	1.00	0.68	0.32	BERT	EXT	1.00	0.72	0.28
Glove	SVM	0.60	0.64	-0.04	Glove	SVM	0.60	0.64	-0.04	Glove	SVM	0.60	0.64	-0.04	Glove	SVM	0.60	0.64	-0.04
Glove	RF	1.00	0.68	0.32	Glove	RF	1.00	0.62	0.38	Glove	RF	1.00	0.66	0.34	Glove	RF	1.00	0.65	0.35
Glove	LR	0.66	0.63	0.03	Glove	LR	0.67	0.68	-0.01	Glove	LR	0.66	0.65	0.01	Glove	LR	0.68	0.69	-0.01
Glove	KNN	0.81	0.62	0.19	Glove	KNN	0.80	0.60	0.20	Glove	KNN	0.83	0.64	0.19	Glove	KNN	0.81	0.59	0.22
Glove	EXT	1.00	0.68	0.32	Glove	EXT	1.00	0.67	0.33	Glove	EXT	1.00	0.67	0.33	Glove	EXT	1.00	0.69	0.31
TF-IDF	**SVM**	**0.93**	**0.85**	**0.08**	**TF-IDF**	**SVM**	**0.93**	**0.85**	**0.08**	**TF-IDF**	**SVM**	**0.93**	**0.85**	**0.08**	**TF-IDF**	**SVM**	**0.93**	**0.84**	**0.09**
TF-IDF	RF	1.00	0.83	0.17	TF-IDF	RF	1.00	0.82	0.18	TF-IDF	RF	1.00	0.81	0.19	TF-IDF	RF	1.00	0.81	0.19
TF-IDF	**LR**	**0.90**	**0.81**	**0.09**	**TF-IDF**	**LR**	**0.91**	**0.79**	**0.12**	**TF-IDF**	**LR**	**0.91**	**0.82**	**0.09**	**TF-IDF**	**LR**	**0.91**	**0.79**	**0.12**
TF-IDF	**KNN**	**0.90**	**0.80**	**0.10**	**TF-IDF**	**KNN**	**0.91**	**0.76**	**0.15**	**TF-IDF**	**KNN**	**0.90**	**0.80**	**0.10**	**TF-IDF**	**KNN**	**0.90**	**0.79**	**0.11**
TF-IDF	EXT	1.00	0.82	0.18	TF-IDF	EXT	1.00	0.81	0.19	TF-IDF	EXT	1.00	0.81	0.19	TF-IDF	EXT	1.00	0.81	0.19
Word2Vec	SVM	0.60	0.64	-0.04	Word2Vec	SVM	0.60	0.64	-0.04	Word2Vec	SVM	0.60	0.64	-0.04	Word2Vec	SVM	0.60	0.64	-0.04
Word2Vec	RF	1.00	0.66	0.34	Word2Vec	RF	1.00	0.63	0.37	Word2Vec	RF	1.00	0.68	0.32	Word2Vec	RF	1.00	0.68	0.32
Word2Vec	LR	0.61	0.64	-0.03	Word2Vec	LR	0.62	0.64	-0.02	Word2Vec	LR	0.61	0.64	-0.03	Word2Vec	LR	0.63	0.65	-0.02
Word2Vec	KNN	0.81	0.57	0.24	Word2Vec	KNN	0.81	0.56	0.25	Word2Vec	KNN	0.81	0.59	0.22	Word2Vec	KNN	0.79	0.59	0.20
Word2Vec	EXT	1.00	0.67	0.33	Word2Vec	EXT	1.00	0.60	0.40	Word2Vec	EXT	1.00	0.68	0.32	Word2Vec	EXT	1.00	0.68	0.32

Note: The cell(s) highlighted in yellow with red colored text indicate the best performing case(s). The cell(s) highlighted in orange indicate the second best performing case(s).

Table 8.

Results (Commitment, Study1- Survey data)

Pre-processing data 0					Pre-processing data 1					Pre-processing data 2					Pre-processing data 3				
emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff
BERT	SVM	0.67	0.66	0.01	**BERT**	**SVM**	**0.83**	**0.74**	**0.09**	BERT	SVM	0.67	0.66	0.01	**BERT**	**SVM**	**0.80**	**0.77**	**0.03**
BERT	RF	1.00	0.67	0.33	BERT	RF	1.00	0.71	0.29	BERT	RF	1.00	0.70	0.30	BERT	RF	1.00	0.72	0.28
BERT	LR	0.90	0.73	0.17	BERT	LR	1.00	0.78	0.22	BERT	LR	0.90	0.73	0.17	BERT	LR	1.00	0.78	0.22
BERT	KNN	0.85	0.69	0.16	BERT	KNN	0.88	0.70	0.18	BERT	KNN	0.85	0.67	0.18	BERT	KNN	0.86	0.76	0.10
BERT	EXT	1.00	0.71	0.29	BERT	EXT	1.00	0.70	0.30	BERT	EXT	1.00	0.71	0.29	BERT	EXT	1.00	0.75	0.25
Glove	SVM	0.67	0.66	0.01	Glove	SVM	0.67	0.66	0.01	Glove	SVM	0.67	0.66	0.01	Glove	SVM	0.67	0.66	0.01
Glove	RF	1.00	0.67	0.33	Glove	RF	1.00	0.68	0.32	Glove	RF	1.00	0.68	0.32	Glove	RF	1.00	0.67	0.33
Glove	LR	0.69	0.67	0.02	Glove	LR	0.70	0.66	0.04	Glove	LR	0.69	0.67	0.02	Glove	LR	0.69	0.67	0.02
Glove	KNN	0.79	0.64	0.15	Glove	KNN	0.81	0.63	0.18	Glove	KNN	0.81	0.58	0.23	Glove	KNN	0.83	0.57	0.26
Glove	EXT	1.00	0.69	0.31	Glove	EXT	1.00	0.68	0.32	Glove	EXT	1.00	0.68	0.32	Glove	EXT	1.00	0.68	0.32
TF-IDF	**SVM**	**0.93**	**0.82**	**0.11**	**TF-IDF**	**SVM**	**0.92**	**0.88**	**0.04**	**TF-IDF**	**SVM**	**0.93**	**0.83**	**0.10**	**TF-IDF**	**SVM**	**0.93**	**0.87**	**0.06**
TF-IDF	RF	1.00	0.80	0.20	TF-IDF	RF	1.00	0.79	0.21	TF-IDF	RF	1.00	0.80	0.20	TF-IDF	RF	1.00	0.81	0.19
TF-IDF	**LR**	**0.89**	**0.79**	**0.10**	**TF-IDF**	**LR**	**0.90**	**0.82**	**0.08**	**TF-IDF**	**LR**	**0.90**	**0.81**	**0.09**	**TF-IDF**	**LR**	**0.90**	**0.81**	**0.09**
TF-IDF	KNN	0.92	0.83	0.09	**TF-IDF**	**KNN**	**0.91**	**0.83**	**0.08**	**TF-IDF**	**KNN**	**0.90**	**0.83**	**0.07**	**TF-IDF**	**KNN**	**0.89**	**0.84**	**0.05**
TF-IDF	EXT	1.00	0.79	0.21	TF-IDF	EXT	1.00	0.82	0.18	TF-IDF	EXT	1.00	0.77	0.23	TF-IDF	EXT	1.00	0.81	0.19
Word2Vec	SVM	0.66	0.65	0.01	Word2Vec	SVM	0.67	0.66	0.01	Word2Vec	SVM	0.66	0.65	0.01	Word2Vec	SVM	0.67	0.66	0.01
Word2Vec	RF	1.00	0.70	0.30	Word2Vec	RF	1.00	0.72	0.28	Word2Vec	RF	1.00	0.69	0.31	Word2Vec	RF	1.00	0.68	0.32
Word2Vec	LR	0.67	0.66	0.01	Word2Vec	LR	0.67	0.66	0.01	Word2Vec	LR	0.67	0.66	0.01	Word2Vec	LR	0.67	0.66	0.01
Word2Vec	KNN	0.81	0.63	0.18	Word2Vec	KNN	0.82	0.60	0.22	Word2Vec	KNN	0.83	0.59	0.24	Word2Vec	KNN	0.81	0.63	0.18
Word2Vec	EXT	1.00	0.67	0.33	Word2Vec	EXT	1.00	0.67	0.33	Word2Vec	EXT	1.00	0.70	0.30	Word2Vec	EXT	1.00	0.71	0.29
Pre-processing data 4					Pre-processing data 5					Pre-processing data 6					Pre-processing data 7				
emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff
BERT	SVM	0.67	0.66	0.01	**BERT**	**SVM**	**0.83**	**0.79**	**0.04**	BERT	SVM	0.67	0.66	0.01	**BERT**	**SVM**	**0.82**	**0.78**	**0.04**
BERT	RF	1.00	0.71	0.29	BERT	RF	1.00	0.70	0.30	BERT	RF	1.00	0.71	0.29	BERT	RF	1.00	0.74	0.26
BERT	LR	0.90	0.77	0.13	BERT	LR	1.00	0.80	0.20	BERT	LR	0.91	0.75	0.16	BERT	LR	1.00	0.80	0.20
BERT	KNN	0.84	0.69	0.15	BERT	KNN	0.87	0.70	0.17	BERT	KNN	0.86	0.68	0.18	BERT	KNN	0.87	0.73	0.14
BERT	EXT	1.00	0.71	0.29	BERT	EXT	1.00	0.71	0.29	BERT	EXT	1.00	0.68	0.32	BERT	EXT	1.00	0.72	0.28
Glove	SVM	0.67	0.66	0.01	Glove	SVM	0.67	0.66	0.01	Glove	SVM	0.67	0.66	0.01	Glove	SVM	0.67	0.66	0.01
Glove	RF	1.00	0.68	0.32	Glove	RF	1.00	0.68	0.32	Glove	RF	1.00	0.67	0.33	Glove	RF	1.00	0.69	0.31
Glove	LR	0.69	0.67	0.02	Glove	LR	0.70	0.67	0.03	Glove	LR	0.70	0.67	0.03	Glove	LR	0.70	0.67	0.03
Glove	KNN	0.79	0.64	0.15	Glove	KNN	0.82	0.63	0.19	Glove	KNN	0.80	0.58	0.22	Glove	KNN	0.82	0.58	0.24
Glove	EXT	1.00	0.71	0.29	Glove	EXT	1.00	0.67	0.33	Glove	EXT	1.00	0.70	0.30	Glove	EXT	1.00	0.69	0.31
TF-IDF	**SVM**	**0.94**	**0.83**	**0.11**	**TF-IDF**	**SVM**	**0.93**	**0.85**	**0.08**	**TF-IDF**	**SVM**	**0.94**	**0.83**	**0.11**	**TF-IDF**	**SVM**	**0.93**	**0.84**	**0.09**
TF-IDF	RF	1.00	0.80	0.20	TF-IDF	RF	1.00	0.78	0.22	TF-IDF	RF	1.00	0.79	0.21	TF-IDF	RF	1.00	0.79	0.21
TF-IDF	**LR**	**0.89**	**0.77**	**0.12**	**TF-IDF**	**LR**	**0.89**	**0.79**	**0.10**	**TF-IDF**	**LR**	**0.89**	**0.78**	**0.11**	**TF-IDF**	**LR**	**0.89**	**0.79**	**0.10**
TF-IDF	**KNN**	**0.92**	**0.82**	**0.10**	**TF-IDF**	**KNN**	**0.91**	**0.83**	**0.08**	**TF-IDF**	**KNN**	**0.91**	**0.84**	**0.07**	**TF-IDF**	**KNN**	**0.91**	**0.86**	**0.05**
TF-IDF	EXT	1.00	0.80	0.20	TF-IDF	EXT	1.00	0.81	0.19	TF-IDF	EXT	1.00	0.79	0.21	TF-IDF	EXT	1.00	0.80	0.20
Word2Vec	SVM	0.66	0.65	0.01	Word2Vec	SVM	0.67	0.66	0.01	Word2Vec	SVM	0.66	0.65	0.01	Word2Vec	SVM	0.67	0.66	0.01
Word2Vec	RF	1.00	0.68	0.32	Word2Vec	RF	1.00	0.71	0.29	Word2Vec	RF	1.00	0.67	0.33	Word2Vec	RF	1.00	0.68	0.32
Word2Vec	LR	0.67	0.66	0.01	Word2Vec	LR	0.67	0.66	0.01	Word2Vec	LR	0.67	0.66	0.01	Word2Vec	LR	0.67	0.66	0.01
Word2Vec	KNN	0.81	0.63	0.18	Word2Vec	KNN	0.82	0.60	0.22	Word2Vec	KNN	0.82	0.62	0.20	Word2Vec	KNN	0.82	0.64	0.18
Word2Vec	EXT	1.00	0.71	0.29	Word2Vec	EXT	1.00	0.69	0.31	Word2Vec	EXT	1.00	0.69	0.31	Word2Vec	EXT	1.00	0.69	0.31

Note: The cell(s) highlighted in yellow with red colored text indicate the best performing case(s). The cell(s) highlighted in orange indicate the second best performing case(s).

Table 9.

Results (Satisfaction, Study1- Survey data)

Pre-processing data 0					Pre-processing data 1					Pre-processing data 2					Pre-processing data 3				
emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff
BERT	SVM	0.69	0.65	0.04	**BERT**	**SVM**	**0.78**	**0.73**	**0.05**	BERT	SVM	0.69	0.65	0.04	**BERT**	**SVM**	**0.79**	**0.71**	**0.08**
BERT	RF	1.00	0.70	0.30	BERT	RF	1.00	0.74	0.26	BERT	RF	1.00	0.73	0.27	BERT	RF	1.00	0.73	0.27
BERT	LR	0.90	0.76	0.14	BERT	LR	0.99	0.79	0.20	BERT	LR	0.91	0.74	0.17	BERT	LR	1.00	0.76	0.24
BERT	KNN	0.84	0.68	0.16	BERT	KNN	0.86	0.74	0.12	BERT	KNN	0.84	0.68	0.16	BERT	KNN	0.86	0.72	0.14
BERT	EXT	1.00	0.69	0.31	BERT	EXT	1.00	0.71	0.29	BERT	EXT	1.00	0.74	0.26	BERT	EXT	1.00	0.72	0.28
Glove	SVM	0.69	0.65	0.04	Glove	SVM	0.69	0.65	0.04	Glove	SVM	0.69	0.65	0.04	Glove	SVM	0.69	0.65	0.04
Glove	RF	1.00	0.66	0.34	Glove	RF	1.00	0.66	0.34	Glove	RF	1.00	0.66	0.34	Glove	RF	1.00	0.65	0.35
Glove	LR	0.70	0.66	0.04	Glove	LR	0.70	0.66	0.04	Glove	LR	0.70	0.66	0.04	Glove	LR	0.70	0.67	0.03
Glove	KNN	0.82	0.64	0.18	Glove	KNN	0.81	0.64	0.17	Glove	KNN	0.80	0.63	0.17	Glove	KNN	0.79	0.64	0.15
Glove	EXT	1.00	0.68	0.32	Glove	EXT	1.00	0.68	0.32	Glove	EXT	1.00	0.66	0.34	Glove	EXT	1.00	0.68	0.32
TF-IDF	**SVM**	**0.94**	**0.82**	**0.12**	**TF-IDF**	**SVM**	**0.94**	**0.83**	**0.11**	**TF-IDF**	**SVM**	**0.95**	**0.83**	**0.12**	**TF-IDF**	**SVM**	**0.94**	**0.84**	**0.10**
TF-IDF	RF	1.00	0.80	0.20	TF-IDF	RF	1.00	0.80	0.20	TF-IDF	RF	1.00	0.80	0.20	TF-IDF	RF	1.00	0.80	0.20
TF-IDF	**LR**	**0.88**	**0.77**	**0.11**	**TF-IDF**	**LR**	**0.89**	**0.78**	**0.11**	**TF-IDF**	**LR**	**0.89**	**0.78**	**0.11**	**TF-IDF**	**LR**	**0.89**	**0.78**	**0.11**
TF-IDF	KNN	0.91	0.80	0.11	TF-IDF	KNN	0.91	0.78	0.13	TF-IDF	KNN	0.91	0.80	0.11	TF-IDF	KNN	0.91	0.79	0.12
TF-IDF	EXT	1.00	0.79	0.21	TF-IDF	EXT	1.00	0.82	0.18	TF-IDF	EXT	1.00	0.80	0.20	TF-IDF	EXT	1.00	0.82	0.18
Word2Vec	SVM	0.69	0.65	0.04	Word2Vec	SVM	0.69	0.65	0.04	Word2Vec	SVM	0.69	0.65	0.04	Word2Vec	SVM	0.69	0.65	0.04
Word2Vec	RF	1.00	0.66	0.34	Word2Vec	RF	1.00	0.68	0.32	Word2Vec	RF	1.00	0.69	0.31	Word2Vec	RF	1.00	0.66	0.34
Word2Vec	LR				Word2Vec	LR	0.69	0.65	0.04	Word2Vec	LR	0.69	0.65	0.04	Word2Vec	LR	0.69	0.65	0.04
Word2Vec	KNN	0.80	0.64	0.16	Word2Vec	KNN	0.82	0.61	0.21	Word2Vec	KNN	0.80	0.61	0.19	Word2Vec	KNN	0.81	0.65	0.16
Word2Vec	EXT	1.00	0.66	0.34	Word2Vec	EXT	1.00	0.67	0.33	Word2Vec	EXT	1.00	0.68	0.32	Word2Vec	EXT	1.00	0.64	0.36
Pre-processing data 4					Pre-processing data 5					Pre-processing data 6					Pre-processing data 7				
emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff
BERT	SVM	0.69	0.65	0.04	**BERT**	**SVM**	**0.78**	**0.71**	**0.07**	BERT	SVM	0.69	0.65	0.04	**BERT**	**SVM**	**0.78**	**0.72**	**0.06**
BERT	RF	1.00	0.67	0.33	BERT	RF	1.00	0.72	0.28	BERT	RF	1.00	0.74	0.26	BERT	RF	1.00	0.73	0.27
BERT	LR	0.90	0.78	0.12	BERT	LR	0.99	0.79	0.20	BERT	LR	0.91	0.78	0.13	BERT	LR	1.00	0.75	0.25
BERT	KNN	0.84	0.65	0.19	BERT	KNN	0.86	0.72	0.14	BERT	KNN	0.85	0.66	0.19	BERT	KNN	0.86	0.71	0.15
BERT	EXT	1.00	0.68	0.32	BERT	EXT	1.00	0.72	0.28	BERT	EXT	1.00	0.74	0.26	BERT	EXT	1.00	0.72	0.28
Glove	SVM	0.69	0.65	0.04	Glove	SVM	0.69	0.65	0.04	Glove	SVM	0.69	0.65	0.04	Glove	SVM	0.69	0.65	0.04
Glove	RF	1.00	0.68	0.32	Glove	RF	1.00	0.66	0.34	Glove	RF	1.00	0.68	0.32	Glove	RF	1.00	0.68	0.32
Glove	LR	0.70	0.66	0.04	Glove	LR	0.70	0.67	0.03	Glove	LR	0.70	0.66	0.04	Glove	LR	0.70	0.68	0.02
Glove	KNN	0.82	0.65	0.17	Glove	KNN	0.81	0.63	0.18	Glove	KNN	0.80	0.62	0.18	Glove	KNN	0.80	0.63	0.17
Glove	EXT	1.00	0.66	0.34	Glove	EXT	1.00	0.68	0.32	Glove	EXT	1.00	0.65	0.35	Glove	EXT	1.00	0.68	0.32
TF-IDF	**SVM**	**0.94**	**0.85**	**0.09**	**TF-IDF**	**SVM**	**0.94**	**0.83**	**0.11**	**TF-IDF**	**SVM**	**0.94**	**0.85**	**0.09**	**TF-IDF**	**SVM**	**0.94**	**0.84**	**0.10**
TF-IDF	RF	1.00	0.79	0.21	TF-IDF	RF	1.00	0.80	0.20	TF-IDF	RF	1.00	0.80	0.20	TF-IDF	RF	1.00	0.79	0.21
TF-IDF	**LR**	**0.88**	**0.77**	**0.11**	**TF-IDF**	**LR**	**0.89**	**0.78**	**0.11**	**TF-IDF**	**LR**	**0.89**	**0.78**	**0.11**	**TF-IDF**	**LR**	**0.89**	**0.79**	**0.10**
TF-IDF	KNN	0.92	0.84	0.08	TF-IDF	KNN	0.91	0.81	0.10	TF-IDF	KNN	0.92	0.80	0.12	TF-IDF	KNN	0.91	0.81	0.10
TF-IDF	EXT	1.00	0.81	0.19	TF-IDF	EXT	1.00	0.82	0.18	TF-IDF	EXT	1.00	0.81	0.19	TF-IDF	EXT	1.00	0.82	0.18
Word2Vec	SVM	0.69	0.65	0.04	Word2Vec	SVM	0.69	0.65	0.04	Word2Vec	SVM	0.69	0.65	0.04	Word2Vec	SVM	0.69	0.65	0.04
Word2Vec	RF	1.00	0.69	0.31	Word2Vec	RF	1.00	0.68	0.32	Word2Vec	RF	1.00	0.70	0.30	Word2Vec	RF	1.00	0.70	0.30
Word2Vec	LR	0.69	0.65	0.04	Word2Vec	LR	0.69	0.65	0.04	Word2Vec	LR	0.69	0.65	0.04	Word2Vec	LR	0.69	0.65	0.04
Word2Vec	KNN	0.80	0.66	0.14	Word2Vec	KNN	0.80	0.61	0.19	Word2Vec	KNN	0.80	0.63	0.17	Word2Vec	KNN	0.81	0.64	0.17
Word2Vec	EXT	1.00	0.69	0.31	Word2Vec	EXT	1.00	0.69	0.31	Word2Vec	EXT	1.00	0.66	0.34	Word2Vec	EXT	1.00	0.65	0.35

Note: The cell(s) highlighted in yellow with red colored text indicate the best performing case(s). The cell(s) highlighted in orange indicate the second best performing case(s).

Table 10.

Results (Trust, Study1- Survey data)

Pre-processing data 0					Pre-processing data 1					Pre-processing data 2					Pre-processing data 3				
emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff
BERT	SVM	0.71	0.72	-0.01	**BERT**	**SVM**	**0.78**	**0.77**	**0.01**	BERT	SVM	0.71	0.72	-0.01	BERT	SVM	0.74	0.78	-0.04
BERT	RF	1.00	0.75	0.25	BERT	RF	1.00	0.76	0.24	BERT	RF	1.00	0.75	0.25	BERT	RF	1.00	0.76	0.24
BERT	**LR**	**0.87**	**0.83**	**0.04**	BERT	LR	0.99	0.79	0.20	**BERT**	**LR**	**0.87**	**0.82**	**0.05**	BERT	LR	1.00	0.84	0.16
BERT	KNN	0.84	0.66	0.18	BERT	KNN	0.86	0.71	0.15	BERT	KNN	0.85	0.68	0.17	BERT	KNN	0.86	0.77	0.09
BERT	EXT	1.00	0.75	0.25	BERT	EXT	1.00	0.74	0.26	BERT	EXT	1.00	0.76	0.24	BERT	EXT	1.00	0.76	0.24
Glove	SVM	0.71	0.72	-0.01	Glove	SVM	0.71	0.72	-0.01	Glove	SVM	0.71	0.72	-0.01	Glove	SVM	0.71	0.72	-0.01
Glove	RF	1.00	0.73	0.27	Glove	RF	1.00	0.74	0.26	Glove	RF	1.00	0.72	0.28	Glove	RF	1.00	0.73	0.27
Glove	**LR**	**0.71**	**0.72**	**-0.01**	**Glove**	**LR**	**0.72**	**0.72**	**0.00**	**Glove**	**LR**	**0.71**	**0.72**	**-0.01**	**Glove**	**LR**	**0.71**	**0.72**	**-0.01**
Glove	KNN	0.82	0.64	0.18	Glove	KNN	0.83	0.67	0.16	Glove	KNN	0.82	0.64	0.18	Glove	KNN	0.81	0.67	0.14
Glove	EXT	1.00	0.73	0.27	Glove	EXT	1.00	0.73	0.27	Glove	EXT	1.00	0.73	0.27	Glove	EXT	1.00	0.72	0.28
TF-IDF	**SVM**	**0.93**	**0.85**	**0.08**	**TF-IDF**	**SVM**	**0.92**	**0.85**	**0.07**	**TF-IDF**	**SVM**	**0.93**	**0.85**	**0.08**	**TF-IDF**	**SVM**	**0.92**	**0.86**	**0.06**
TF-IDF	RF	1.00	0.77	0.23	TF-IDF	RF	1.00	0.79	0.21	TF-IDF	RF	1.00	0.77	0.23	TF-IDF	RF	1.00	0.78	0.22
TF-IDF	**LR**	**0.86**	**0.79**	**0.07**	**TF-IDF**	**LR**	**0.86**	**0.80**	**0.06**	**TF-IDF**	**LR**	**0.86**	**0.78**	**0.08**	**TF-IDF**	**LR**	**0.87**	**0.80**	**0.07**
TF-IDF	**KNN**	**0.91**	**0.83**	**0.08**	**TF-IDF**	**KNN**	**0.90**	**0.80**	**0.10**	**TF-IDF**	**KNN**	**0.90**	**0.82**	**0.08**	**TF-IDF**	**KNN**	**0.90**	**0.79**	**0.11**
TF-IDF	EXT	1.00	0.81	0.19	TF-IDF	EXT	1.00	0.81	0.19	TF-IDF	EXT	1.00	0.81	0.19	TF-IDF	EXT	1.00	0.78	0.22
Word2Vec	SVM	0.70	0.73	-0.03	Word2Vec	SVM	0.70	0.72	-0.02	Word2Vec	SVM	0.71	0.73	-0.02	Word2Vec	SVM	0.71	0.72	-0.01
Word2Vec	RF	1.00	0.73	0.27	Word2Vec	RF	1.00	0.74	0.26	Word2Vec	RF	1.00	0.73	0.27	Word2Vec	RF	1.00	0.71	0.29
Word2Vec	LR	0.71	0.72	-0.01	Word2Vec	LR	0.71	0.72	-0.01	Word2Vec	LR	0.71	0.72	-0.01	Word2Vec	LR	0.71	0.72	-0.01
Word2Vec	KNN	0.84	0.66	0.18	Word2Vec	KNN	0.83	0.69	0.14	Word2Vec	KNN	0.84	0.68	0.16	Word2Vec	KNN	0.83	0.68	0.15
Word2Vec	EXT	1.00	0.75	0.25	Word2Vec	EXT	1.00	0.73	0.27	Word2Vec	EXT	1.00	0.73	0.27	Word2Vec	EXT	1.00	0.71	0.29

Pre-processing data 4					Pre-processing data 5					Pre-processing data 6					Pre-processing data 7				
emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff
BERT	**SVM**	**0.71**	**0.72**	**-0.01**	**BERT**	**SVM**	**0.76**	**0.75**	**0.01**	**BERT**	**SVM**	**0.71**	**0.72**	**-0.01**	BERT	SVM	0.75	0.79	-0.04
BERT	RF	1.00	0.75	0.25	BERT	RF	1.00	0.74	0.26	BERT	RF	1.00	0.73	0.27	BERT	RF	1.00	0.77	0.23
BERT	**LR**	**0.88**	**0.80**	**0.08**	BERT	LR	0.99	0.81	0.18	**BERT**	**LR**	**0.88**	**0.83**	**0.05**	BERT	LR	0.99	0.84	0.15
BERT	KNN	0.85	0.65	0.20	BERT	KNN	0.87	0.73	0.14	BERT	KNN	0.84	0.70	0.14	BERT	KNN	0.86	0.77	0.09
BERT	EXT	1.00	0.75	0.25	BERT	EXT	1.00	0.76	0.24	BERT	EXT	1.00	0.74	0.26	BERT	EXT	1.00	0.78	0.22
Glove	SVM	0.71	0.72	-0.01	Glove	SVM	0.71	0.72	-0.01	Glove	SVM	0.71	0.72	-0.01	Glove	SVM	0.71	0.72	-0.01
Glove	RF	1.00	0.73	0.27	Glove	RF	1.00	0.73	0.27	Glove	RF	1.00	0.71	0.29	Glove	RF	1.00	0.71	0.29
Glove	**LR**	**0.71**	**0.72**	**-0.01**	**Glove**	**LR**	**0.72**	**0.72**	**0.00**	**Glove**	**LR**	**0.71**	**0.72**	**-0.01**	**Glove**	**LR**	**0.72**	**0.72**	**0.00**
Glove	KNN	0.83	0.64	0.19	Glove	KNN	0.82	0.70	0.12	Glove	KNN	0.81	0.63	0.18	Glove	KNN	0.81	0.66	0.15
Glove	EXT	1.00	0.76	0.24	Glove	EXT	1.00	0.73	0.27	Glove	EXT	1.00	0.73	0.27	Glove	EXT	1.00	0.71	0.29
TF-IDF	**SVM**	**0.92**	**0.86**	**0.06**	**TF-IDF**	**SVM**	**0.92**	**0.85**	**0.07**	**TF-IDF**	**SVM**	**0.93**	**0.85**	**0.08**	**TF-IDF**	**SVM**	**0.92**	**0.85**	**0.07**
TF-IDF	RF	1.00	0.78	0.22	TF-IDF	RF	1.00	0.80	0.20	TF-IDF	RF	1.00	0.79	0.21	TF-IDF	RF	1.00	0.79	0.21
TF-IDF	**LR**	**0.87**	**0.82**	**0.05**	**TF-IDF**	**LR**	**0.87**	**0.82**	**0.05**	**TF-IDF**	**LR**	**0.87**	**0.82**	**0.05**	**TF-IDF**	**LR**	0.87	0.81	0.06
TF-IDF	**KNN**	**0.91**	**0.81**	**0.10**	**TF-IDF**	**KNN**	**0.90**	**0.80**	**0.10**	**TF-IDF**	**KNN**	**0.91**	**0.83**	**0.08**	**TF-IDF**	**KNN**	**0.90**	**0.79**	**0.11**
TF-IDF	EXT	1.00	0.81	0.19	TF-IDF	EXT	1.00	0.81	0.19	TF-IDF	EXT	1.00	0.80	0.20	TF-IDF	EXT	1.00	0.81	0.19
Word2Vec	SVM	0.71	0.73	-0.02	Word2Vec	SVM	0.71	0.72	-0.01	Word2Vec	SVM	0.71	0.73	-0.02	Word2Vec	SVM	0.71	0.72	-0.01
Word2Vec	RF	1.00	0.73	0.27	Word2Vec	RF	1.00	0.73	0.27	Word2Vec	RF	1.00	0.74	0.26	Word2Vec	RF	1.00	0.71	0.29
Word2Vec	LR	0.71	0.72	-0.01	Word2Vec	LR	0.71	0.72	-0.01	Word2Vec	LR	0.71	0.72	-0.01	Word2Vec	LR	0.71	0.72	-0.01
Word2Vec	KNN	0.83	0.66	0.17	Word2Vec	KNN	0.83	0.71	0.12	Word2Vec	KNN	0.85	0.68	0.17	Word2Vec	KNN	0.82	0.65	0.17
Word2Vec	EXT	1.00	0.75	0.25	Word2Vec	EXT	1.00	0.75	0.25	Word2Vec	EXT	1.00	0.73	0.27	Word2Vec	EXT	1.00	0.73	0.27

Note: The cell(s) highlighted in yellow with red colored text indicate the best performing case(s). The cell(s) highlighted in orange indicate the second best performing case(s).

Table 11.
A Summary of three survey data results (Study 1)

Control Mutuality

Pre-processing data 2				
emb	model	train	test	diff
TF-IDF	**SVM**	**0.93**	**0.85**	**0.08**

Commitment

Pre-processing data 1				
emb	model	train	test	diff
TF-IDF	**SVM**	**0.92**	**0.88**	**0.04**
Pre-processing data 4				
TF-IDF	**SVM**	**0.93**	**0.85**	**0.08**
Pre-processing data 5				
TF-IDF	**SVM**	**0.93**	**0.85**	**0.08**
Pre-processing data 6				
TF-IDF	**SVM**	**0.93**	**0.85**	**0.08**

Satisfaction

Pre-processing data 4				
emb	model	train	test	diff
TF-IDF	**SVM**	**0.94**	**0.85**	**0.09**

Trust

Pre-processing data 3				
emb	model	train	test	diff
TF-IDF	**SVM**	**0.92**	**0.86**	**0.06**
TF-IDF	**SVM**	**0.92**	**0.86**	**0.06**

Table 12.

Results (Control Mutuality, Study1- Amazon + Survey data)

Pre-processing data 0					Pre-processing data 1					Pre-processing data 2					Pre-processing data 3				
emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff
BERT	SVM	0.64	0.60	0.04	**BERT**	**SVM**	0.76	0.68	0.08	BERT	SVM	0.63	0.57	0.06	**BERT**	**SVM**	**0.77**	**0.73**	**0.04**
BERT	RF	1.00	0.67	0.33	BERT	RF	1.00	0.70	0.30	BERT	RF	1.00	0.65	0.35	BERT	RF	1.00	0.71	0.29
BERT	**LR**	**0.86**	**0.75**	**0.11**	BERT	LR	0.98	0.76	0.22	BERT	LR	0.87	0.71	0.16	BERT	LR	0.98	0.71	0.27
BERT	KNN	0.81	0.65	0.16	BERT	KNN	0.85	0.67	0.18	BERT	KNN	0.82	0.63	0.19	BERT	KNN	0.83	0.70	0.13
BERT	EXT	1.00	0.69	0.31	BERT	EXT	1.00	0.68	0.32	BERT	EXT	1.00	0.68	0.32	BERT	EXT	1.00	0.77	0.23
Glove	SVM	0.52	0.50	0.02	Glove	SVM	0.52	0.50	0.02	Glove	SVM	0.52	0.50	0.02	Glove	SVM	0.52	0.50	0.02
Glove	RF	1.00	0.68	0.32	Glove	RF	1.00	0.65	0.35	Glove	RF	1.00	0.67	0.33	Glove	RF	1.00	0.65	0.35
Glove	LR	0.66	0.65	0.01	Glove	LR	0.65	0.65	0.00	Glove	LR	0.66	0.65	0.01	Glove	LR	0.64	0.65	-0.01
Glove	KNN	0.81	0.58	0.23	Glove	KNN	0.79	0.53	0.26	Glove	KNN	0.80	0.64	0.16	Glove	KNN	0.79	0.58	0.21
Glove	EXT	1.00	0.66	0.34	Glove	EXT	1.00	0.67	0.33	Glove	EXT	1.00	0.67	0.33	Glove	EXT	1.00	0.63	0.37
TF-IDF	**SVM**	**0.94**	**0.80**	**0.14**	**TF-IDF**	**SVM**	**0.93**	**0.82**	**0.11**	**TF-IDF**	**SVM**	**0.93**	**0.80**	**0.13**	**TF-IDF**	**SVM**	**0.92**	**0.83**	**0.09**
TF-IDF	RF	1.00	0.81	0.19	TF-IDF	RF	1.00	0.83	0.17	TF-IDF	RF	1.00	0.81	0.19	TF-IDF	RF	1.00	0.82	0.18
TF-IDF	**LR**	**0.93**	**0.81**	**0.12**	**TF-IDF**	**LR**	**0.93**	**0.82**	**0.11**	**TF-IDF**	**LR**	**0.93**	**0.81**	**0.12**	**TF-IDF**	**LR**	**0.93**	**0.82**	**0.11**
TF-IDF	KNN	0.89	0.73	0.16	TF-IDF	KNN	0.88	0.74	0.14	TF-IDF	KNN	0.88	0.73	0.15	TF-IDF	KNN	0.88	0.73	0.15
TF-IDF	EXT	1.00	0.82	0.18	TF-IDF	EXT	1.00	0.82	0.18	TF-IDF	EXT	1.00	0.82	0.18	TF-IDF	EXT	1.00	0.83	0.17
Word2Vec	SVM	0.54	0.50	0.04	Word2Vec	SVM	0.53	0.50	0.03	Word2Vec	SVM	0.53	0.50	0.03	Word2Vec	SVM	0.53	0.50	0.03
Word2Vec	RF	1.00	0.63	0.37	Word2Vec	RF	1.00	0.66	0.34	Word2Vec	RF	1.00	0.67	0.33	Word2Vec	RF	1.00	0.66	0.34
Word2Vec	LR	0.62	0.60	0.02	Word2Vec	LR	0.64	0.63	0.01	Word2Vec	LR	0.63	0.64	-0.01	Word2Vec	LR	0.64	0.61	0.03
Word2Vec	KNN	0.81	0.61	0.20	Word2Vec	KNN	0.83	0.52	0.31	Word2Vec	KNN	0.79	0.61	0.18	Word2Vec	KNN	0.80	0.54	0.26
Word2Vec	EXT	1.00	0.69	0.31	Word2Vec	EXT	1.00	0.65	0.35	Word2Vec	EXT	1.00	0.69	0.31	Word2Vec	EXT	1.00	0.63	0.37
Pre-processing data 4					Pre-processing data 5					Pre-processing data 6					Pre-processing data 7				
emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff
BERT	SVM	0.65	0.62	0.03	**BERT**	**SVM**	**0.75**	**0.70**	**0.05**	BERT	SVM	0.66	0.58	0.08	**BERT**	**SVM**	**0.77**	**0.72**	**0.05**
BERT	RF	1.00	0.68	0.32	BERT	RF	1.00	0.67	0.33	BERT	RF	1.00	0.65	0.35	BERT	RF	1.00	0.72	0.28
BERT	**LR**	**0.85**	**0.74**	**0.11**	BERT	LR	0.98	0.73	0.25	**BERT**	**LR**	**0.86**	**0.74**	**0.12**	BERT	LR	0.98	0.69	0.29
BERT	KNN	0.82	0.63	0.19	BERT	KNN	0.85	0.68	0.17	BERT	KNN	0.81	0.60	0.21	BERT	KNN	0.83	0.70	0.13
BERT	EXT	1.00	0.68	0.32	BERT	EXT	1.00	0.67	0.33	BERT	EXT	1.00	0.68	0.32	BERT	EXT	1.00	0.73	0.27
Glove	SVM	0.52	0.50	0.02	Glove	SVM	0.52	0.50	0.02	Glove	SVM	0.52	0.50	0.02	Glove	SVM	0.52	0.50	0.02
Glove	RF	1.00	0.63	0.37	Glove	RF	1.00	0.65	0.35	Glove	RF	1.00	0.67	0.33	Glove	RF	1.00	0.68	0.32
Glove	LR	0.66	0.65	0.01	Glove	LR	0.65	0.66	-0.01	Glove	LR	0.67	0.67	0.00	Glove	LR	0.64	0.68	-0.04
Glove	KNN	0.81	0.62	0.19	Glove	KNN	0.79	0.55	0.24	Glove	KNN	0.80	0.65	0.15	Glove	KNN	0.79	0.55	0.24
Glove	EXT	1.00	0.65	0.35	Glove	EXT	1.00	0.66	0.34	Glove	EXT	1.00	0.69	0.31	Glove	EXT	1.00	0.62	0.38
TF-IDF	**SVM**	**0.93**	**0.79**	**0.14**	**TF-IDF**	**SVM**	**0.93**	**0.81**	**0.12**	**TF-IDF**	**SVM**	**0.94**	**0.79**	**0.15**	**TF-IDF**	**SVM**	**0.93**	**0.82**	**0.11**
TF-IDF	RF	1.00	0.84	0.16	TF-IDF	RF	1.00	0.83	0.17	TF-IDF	RF	1.00	0.82	0.18	TF-IDF	RF	1.00	0.81	0.19
TF-IDF	**LR**	**0.93**	**0.82**	**0.11**	**TF-IDF**	**LR**	**0.93**	**0.83**	**0.10**	**TF-IDF**	**LR**	**0.93**	**0.81**	**0.12**	**TF-IDF**	**LR**	**0.93**	**0.83**	**0.10**
TF-IDF	KNN	0.88	0.72	0.16	TF-IDF	KNN	0.88	0.75	0.13	TF-IDF	KNN	0.88	0.73	0.15	TF-IDF	KNN	0.88	0.73	0.15
TF-IDF	EXT	1.00	0.82	0.18	TF-IDF	EXT	1.00	0.83	0.17	TF-IDF	EXT	1.00	0.82	0.18	TF-IDF	EXT	1.00	0.83	0.17
Word2Vec	SVM	0.54	0.50	0.04	Word2Vec	SVM	0.53	0.50	0.03	Word2Vec	SVM	0.53	0.50	0.03	Word2Vec	SVM	0.53	0.50	0.03
Word2Vec	RF	1.00	0.66	0.34	Word2Vec	RF	1.00	0.61	0.39	Word2Vec	RF	1.00	0.69	0.31	Word2Vec	RF	1.00	0.65	0.35
Word2Vec	LR	0.62	0.61	0.01	Word2Vec	LR	0.65	0.64	0.01	Word2Vec	LR	0.63	0.62	0.01	Word2Vec	LR	0.65	0.63	0.02
Word2Vec	KNN	0.81	0.61	0.20	Word2Vec	KNN	0.84	0.54	0.30	Word2Vec	KNN	0.80	0.63	0.17	Word2Vec	KNN	0.81	0.55	0.26
Word2Vec	EXT	1.00	0.67	0.33	Word2Vec	EXT	1.00	0.63	0.37	Word2Vec	EXT	1.00	0.70	0.30	Word2Vec	EXT	1.00	0.67	0.33

Note: The cell(s) highlighted in yellow with red colored text indicate the best performing case(s). The cell(s) highlighted in orange indicate the second best performing case(s).

Table 13.

Results (Commitment, Study1- Amazon + Survey data)

Pre-processing data 0					Pre-processing data 1					Pre-processing data 2					Pre-processing data 3				
emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff
BERT	SVM	0.65	0.67	-0.02	**BERT**	**SVM**	**0.76**	**0.74**	**0.02**	BERT	SVM	0.59	0.64	-0.05	**BERT**	**SVM**	**0.78**	**0.74**	**0.04**
BERT	RF	1.00	0.68	0.32	BERT	RF	1.00	0.77	0.23	BERT	RF	1.00	0.68	0.32	BERT	RF	1.00	0.71	0.29
BERT	LR	0.88	0.76	0.12	BERT	LR	0.99	0.75	0.24	BERT	LR	0.89	0.68	0.21	BERT	LR	0.99	0.72	0.27
BERT	KNN	0.82	0.62	0.20	BERT	KNN	0.84	0.69	0.15	BERT	KNN	0.84	0.63	0.21	BERT	KNN	0.82	0.72	0.10
BERT	EXT	1.00	0.67	0.33	BERT	EXT	1.00	0.74	0.26	BERT	EXT	1.00	0.66	0.34	BERT	EXT	1.00	0.76	0.24
Glove	SVM	0.56	0.62	-0.06	Glove	SVM	0.56	0.61	-0.05	Glove	SVM	0.56	0.62	-0.06	Glove	SVM	0.56	0.61	-0.05
Glove	RF	1.00	0.70	0.30	Glove	RF	1.00	0.66	0.34	Glove	RF	1.00	0.67	0.33	Glove	RF	1.00	0.70	0.30
Glove	LR	0.68	0.66	0.02	Glove	LR	0.66	0.63	0.03	Glove	LR	0.67	0.66	0.01	Glove	LR	0.65	0.63	0.02
Glove	KNN	0.80	0.65	0.15	Glove	KNN	0.77	0.67	0.10	Glove	KNN	0.81	0.62	0.19	Glove	KNN	0.79	0.67	0.12
Glove	EXT	1.00	0.70	0.30	Glove	EXT	1.00	0.68	0.32	Glove	EXT	1.00	0.70	0.30	Glove	EXT	1.00	0.69	0.31
TF-IDF	**SVM**	**0.94**	**0.84**	**0.10**	**TF-IDF**	**SVM**	**0.93**	**0.83**	**0.10**	**TF-IDF**	**SVM**	**0.93**	**0.84**	**0.09**	**TF-IDF**	**SVM**	**0.93**	**0.84**	**0.09**
TF-IDF	RF	1.00	0.81	0.19	TF-IDF	RF	1.00	0.83	0.17	TF-IDF	RF	1.00	0.82	0.18	TF-IDF	RF	1.00	0.81	0.19
TF-IDF	**LR**	**0.93**	**0.83**	**0.10**	**TF-IDF**	**LR**	**0.92**	**0.84**	**0.08**	**TF-IDF**	**LR**	**0.93**	**0.83**	**0.10**	**TF-IDF**	**LR**	**0.92**	**0.84**	**0.08**
TF-IDF	KNN	0.90	0.81	0.09	TF-IDF	KNN	0.89	0.79	0.10	TF-IDF	KNN	0.90	0.83	0.07	TF-IDF	KNN	0.89	0.79	0.10
TF-IDF	EXT	1.00	0.82	0.18	TF-IDF	EXT	1.00	0.81	0.19	TF-IDF	EXT	1.00	0.83	0.17	TF-IDF	EXT	1.00	0.81	0.19
Word2Vec	SVM	0.57	0.62	-0.05	Word2Vec	SVM	0.56	0.61	-0.05	Word2Vec	SVM	0.56	0.62	-0.06	Word2Vec	SVM	0.56	0.61	-0.05
Word2Vec	RF	1.00	0.65	0.35	Word2Vec	RF	1.00	0.68	0.32	Word2Vec	RF	1.00	0.65	0.35	Word2Vec	RF	1.00	0.68	0.32
Word2Vec	LR	0.60	0.66	-0.06	Word2Vec	LR	0.63	0.61	0.02	Word2Vec	LR	0.60	0.66	-0.06	Word2Vec	LR	0.62	0.60	0.02
Word2Vec	KNN	0.79	0.64	0.15	Word2Vec	KNN	0.77	0.66	0.11	Word2Vec	KNN	0.80	0.63	0.17	Word2Vec	KNN	0.79	0.69	0.10
Word2Vec	EXT	1.00	0.64	0.36	Word2Vec	EXT	1.00	0.69	0.31	Word2Vec	EXT	1.00	0.67	0.33	Word2Vec	EXT	1.00	0.67	0.33
Pre-processing data 4					Pre-processing data 5					Pre-processing data 6					Pre-processing data 7				
emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff
BERT	SVM	0.60	0.64	-0.04	**BERT**	**SVM**	**0.75**	**0.76**	**-0.01**	BERT	SVM	0.59	0.64	-0.05	**BERT**	**SVM**	**0.78**	**0.76**	**0.02**
BERT	RF	1.00	0.72	0.28	BERT	RF	1.00	0.78	0.22	BERT	RF	1.00	0.69	0.31	BERT	RF	1.00	0.75	0.25
BERT	LR	0.89	0.76	0.13	BERT	LR	0.99	0.76	0.23	BERT	LR	0.90	0.75	0.15	BERT	LR	0.99	0.73	0.26
BERT	KNN	0.82	0.63	0.19	BERT	KNN	0.86	0.71	0.15	BERT	KNN	0.84	0.66	0.18	BERT	KNN	0.83	0.68	0.15
BERT	EXT	1.00	0.67	0.33	BERT	EXT	1.00	0.74	0.26	BERT	EXT	1.00	0.72	0.28	BERT	EXT	1.00	0.77	0.23
Glove	SVM	0.56	0.62	-0.06	Glove	SVM	0.56	0.61	-0.05	Glove	SVM	0.56	0.62	-0.06	Glove	SVM	0.56	0.61	-0.05
Glove	RF	1.00	0.67	0.33	Glove	RF	1.00	0.69	0.31	Glove	RF	1.00	0.69	0.31	Glove	RF	1.00	0.69	0.31
Glove	LR	0.68	0.65	0.03	Glove	LR	0.66	0.64	0.02	Glove	LR	0.67	0.67	0.00	Glove	LR	0.66	0.63	0.03
Glove	KNN	0.80	0.67	0.13	Glove	KNN	0.78	0.66	0.12	Glove	KNN	0.80	0.62	0.18	Glove	KNN	0.79	0.66	0.13
Glove	EXT	1.00	0.69	0.31	Glove	EXT	1.00	0.67	0.33	Glove	EXT	1.00	0.72	0.28	Glove	EXT	1.00	0.69	0.31
TF-IDF	**SVM**	**0.93**	**0.84**	**0.09**	**TF-IDF**	**SVM**	**0.92**	**0.83**	**0.09**	**TF-IDF**	**SVM**	**0.94**	**0.84**	**0.10**	**TF-IDF**	**SVM**	**0.93**	**0.84**	**0.09**
TF-IDF	RF	1.00	0.82	0.18	TF-IDF	RF	1.00	0.83	0.17	TF-IDF	RF	1.00	0.83	0.17	TF-IDF	RF	1.00	0.84	0.16
TF-IDF	**LR**	**0.92**	**0.84**	**0.08**	**TF-IDF**	**LR**	**0.92**	**0.84**	**0.08**	**TF-IDF**	**LR**	**0.92**	**0.85**	**0.07**	**TF-IDF**	**LR**	**0.92**	**0.85**	**0.07**
TF-IDF	KNN	0.91	0.82	0.09	TF-IDF	KNN	0.90	0.79	0.11	TF-IDF	KNN	0.90	0.83	0.07	TF-IDF	KNN	0.90	0.78	0.12
TF-IDF	EXT	1.00	0.83	0.17	TF-IDF	EXT	1.00	0.82	0.18	TF-IDF	EXT	1.00	0.82	0.18	TF-IDF	EXT	1.00	0.83	0.17
Word2Vec	SVM	0.56	0.62	-0.06	Word2Vec	SVM	0.56	0.61	-0.05	Word2Vec	SVM	0.56	0.61	-0.05	Word2Vec	SVM	0.56	0.61	-0.05
Word2Vec	RF	1.00	0.66	0.34	Word2Vec	RF	1.00	0.68	0.32	Word2Vec	RF	1.00	0.68	0.32	Word2Vec	RF	1.00	0.68	0.32
Word2Vec	LR	0.60	0.65	-0.05	Word2Vec	LR	0.63	0.62	0.01	Word2Vec	LR	0.61	0.65	-0.04	Word2Vec	LR	0.63	0.62	0.01
Word2Vec	KNN	0.79	0.62	0.17	Word2Vec	KNN	0.79	0.66	0.13	Word2Vec	KNN	0.80	0.61	0.19	Word2Vec	KNN	0.80	0.70	0.10
Word2Vec	EXT	1.00	0.69	0.31	Word2Vec	EXT	1.00	0.68	0.32	Word2Vec	EXT	1.00	0.70	0.30	Word2Vec	EXT	1.00	0.69	0.31

Note: The cell(s) highlighted in yellow with red colored text indicate the best performing case(s). The cell(s) highlighted in orange indicate the second best performing case(s).

Table 14.

Results (Satisfaction, Study1- Amazon+ Survey data)

Pre-processing data 0					Pre-processing data 1					Pre-processing data 2					Pre-processing data 3				
emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff
BERT	SVM	0.64	0.66	-0.02	**BERT**	**SVM**	**0.79**	**0.76**	**0.03**	BERT	SVM	0.64	0.66	-0.02	**BERT**	**SVM**	**0.79**	**0.77**	**0.02**
BERT	RF	0.99	0.68	0.31	BERT	RF	1.00	0.73	0.27	BERT	RF	1.00	0.72	0.28	BERT	RF	1.00	0.77	0.23
BERT	LR	0.89	0.74	0.15	BERT	LR	0.98	0.77	0.21	BERT	LR	0.89	0.72	0.17	BERT	LR	0.99	0.75	0.24
BERT	KNN	0.83	0.67	0.16	BERT	KNN	0.84	0.67	0.17	BERT	KNN	0.84	0.61	0.23	BERT	KNN	0.84	0.70	0.14
BERT	EXT	0.99	0.71	0.28	BERT	EXT	1.00	0.73	0.27	BERT	EXT	1.00	0.72	0.28	BERT	EXT	1.00	0.74	0.26
Glove	SVM	0.64	0.66	-0.02	Glove	SVM	0.64	0.66	-0.02	Glove	SVM	0.64	0.66	-0.02	Glove	SVM	0.64	0.66	-0.02
Glove	RF	1.00	0.68	0.32	Glove	RF	1.00	0.70	0.30	Glove	RF	1.00	0.69	0.31	Glove	RF	1.00	0.71	0.29
Glove	LR	0.67	0.70	-0.03	Glove	LR	0.68	0.70	-0.02	Glove	LR	0.66	0.69	-0.03	Glove	LR	0.67	0.70	-0.03
Glove	KNN	0.82	0.61	0.21	Glove	KNN	0.81	0.59	0.22	Glove	KNN	0.79	0.63	0.16	Glove	KNN	0.79	0.63	0.16
Glove	EXT	1.00	0.72	0.28	Glove	EXT	1.00	0.65	0.35	Glove	EXT	1.00	0.71	0.29	Glove	EXT	1.00	0.70	0.30
TF-IDF	**SVM**	**0.94**	**0.83**	**0.11**	**TF-IDF**	**SVM**	**0.93**	**0.85**	**0.08**	**TF-IDF**	**SVM**	**0.94**	**0.84**	**0.10**	**TF-IDF**	**SVM**	**0.93**	**0.85**	**0.08**
TF-IDF	RF	1.00	0.81	0.19	TF-IDF	RF	1.00	0.84	0.16	TF-IDF	RF	1.00	0.82	0.18	TF-IDF	RF	1.00	0.85	0.15
TF-IDF	**LR**	**0.89**	**0.82**	**0.07**	**TF-IDF**	**LR**	**0.90**	**0.84**	**0.06**	**TF-IDF**	**LR**	**0.89**	**0.82**	**0.07**	**TF-IDF**	**LR**	**0.89**	**0.84**	**0.05**
TF-IDF	KNN	0.88	0.77	0.11	TF-IDF	KNN	0.88	0.80	0.08	TF-IDF	KNN	0.88	0.79	0.09	TF-IDF	KNN	0.89	0.78	0.11
TF-IDF	EXT	1.00	0.82	0.18	TF-IDF	EXT	1.00	0.85	0.15	TF-IDF	EXT	1.00	0.83	0.17	TF-IDF	EXT	1.00	0.84	0.16
Word2Vec	SVM	0.63	0.66	-0.03	Word2Vec	SVM	0.64	0.66	-0.02	Word2Vec	SVM	0.63	0.66	-0.03	Word2Vec	SVM	0.64	0.66	-0.02
Word2Vec	RF	1.00	0.68	0.32	Word2Vec	RF	1.00	0.70	0.30	Word2Vec	RF	1.00	0.69	0.31	Word2Vec	RF	1.00	0.72	0.28
Word2Vec	LR	0.64	0.67	-0.03	Word2Vec	LR	0.65	0.66	-0.01	Word2Vec	LR	0.64	0.67	-0.03	Word2Vec	LR	0.65	0.67	-0.02
Word2Vec	KNN	0.80	0.64	0.16	Word2Vec	KNN	0.81	0.59	0.22	Word2Vec	KNN	0.79	0.63	0.16	Word2Vec	KNN	0.82	0.61	0.21
Word2Vec	EXT	1.00	0.67	0.33	Word2Vec	EXT	1.00	0.66	0.34	Word2Vec	EXT	1.00	0.68	0.32	Word2Vec	EXT	1.00	0.71	0.29
Pre-processing data 4					Pre-processing data 5					Pre-processing data 6					Pre-processing data 7				
emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff
BERT	SVM	0.64	0.66	-0.02	**BERT**	**SVM**	**0.77**	**0.74**	**0.03**	BERT	SVM	0.64	0.66	-0.02	**BERT**	**SVM**	**0.80**	**0.76**	**0.04**
BERT	RF	0.99	0.69	0.30	BERT	RF	1.00	0.75	0.25	BERT	RF	1.00	0.70	0.30	BERT	RF	1.00	0.76	0.24
BERT	LR	0.90	0.76	0.14	BERT	LR	0.98	0.77	0.21	BERT	LR	0.90	0.73	0.17	BERT	LR	0.99	0.76	0.23
BERT	KNN	0.83	0.63	0.20	BERT	KNN	0.84	0.64	0.20	BERT	KNN	0.82	0.61	0.21	BERT	KNN	0.83	0.70	0.13
BERT	EXT	0.99	0.70	0.29	BERT	EXT	1.00	0.74	0.26	BERT	EXT	1.00	0.70	0.30	BERT	EXT	1.00	0.72	0.28
Glove	SVM	0.64	0.66	-0.02	Glove	SVM	0.64	0.66	-0.02	Glove	SVM	0.64	0.66	-0.02	Glove	SVM	0.64	0.66	-0.02
Glove	RF	1.00	0.68	0.32	Glove	RF	1.00	0.70	0.30	Glove	RF	1.00	0.71	0.29	Glove	RF	1.00	0.72	0.28
Glove	LR	0.66	0.69	-0.03	Glove	LR	0.68	0.71	-0.03	Glove	LR	0.67	0.69	-0.02	Glove	LR	0.68	0.72	-0.04
Glove	KNN	0.81	0.60	0.21	Glove	KNN	0.81	0.59	0.22	Glove	KNN	0.79	0.63	0.16	Glove	KNN	0.79	0.65	0.14
Glove	EXT	1.00	0.70	0.30	Glove	EXT	1.00	0.69	0.31	Glove	EXT	1.00	0.69	0.31	Glove	EXT	1.00	0.68	0.32
TF-IDF	**SVM**	**0.94**	**0.84**	**0.10**	**TF-IDF**	**SVM**	**0.93**	**0.84**	**0.09**	**TF-IDF**	**SVM**	**0.94**	**0.84**	**0.10**	**TF-IDF**	**SVM**	**0.93**	**0.85**	**0.08**
TF-IDF	RF	1.00	0.84	0.16	TF-IDF	RF	1.00	0.86	0.14	TF-IDF	RF	1.00	0.84	0.16	TF-IDF	RF	1.00	0.83	0.17
TF-IDF	**LR**	**0.90**	**0.82**	**0.08**	**TF-IDF**	**LR**	**0.90**	**0.83**	**0.07**	**TF-IDF**	**LR**	**0.89**	**0.83**	**0.06**	**TF-IDF**	**LR**	**0.89**	**0.84**	**0.05**
TF-IDF	KNN	0.88	0.78	0.10	TF-IDF	KNN	0.89	0.79	0.10	TF-IDF	KNN	0.89	0.79	0.10	TF-IDF	KNN	0.89	0.78	0.11
TF-IDF	EXT	1.00	0.83	0.17	TF-IDF	EXT	1.00	0.84	0.16	TF-IDF	EXT	1.00	0.82	0.18	TF-IDF	EXT	1.00	0.86	0.14
Word2Vec	SVM	0.63	0.66	-0.03	Word2Vec	SVM	0.64	0.66	-0.02	Word2Vec	SVM	0.63	0.66	-0.03	Word2Vec	SVM	0.64	0.66	-0.02
Word2Vec	RF	1.00	0.68	0.32	Word2Vec	RF	1.00	0.68	0.32	Word2Vec	RF	1.00	0.70	0.30	Word2Vec	RF	1.00	0.74	0.26
Word2Vec	LR	0.64	0.67	-0.03	Word2Vec	LR	0.65	0.66	-0.01	Word2Vec	LR	0.65	0.67	-0.02	Word2Vec	LR	0.65	0.68	-0.03
Word2Vec	KNN	0.80	0.63	0.17	Word2Vec	KNN	0.81	0.57	0.24	Word2Vec	KNN	0.79	0.65	0.14	Word2Vec	KNN	0.81	0.63	0.18
Word2Vec	EXT	1.00	0.69	0.31	Word2Vec	EXT	1.00	0.67	0.33	Word2Vec	EXT	1.00	0.68	0.32	Word2Vec	EXT	1.00	0.71	0.29

Note: The cell(s) highlighted in yellow with red colored text indicate the best performing case(s). The cell(s) highlighted in orange indicate the second best performing case(s).

Table 15.

Results (Trust, Study1- Amazon + Survey data)

Pre-processing data 0					Pre-processing data 1					Pre-processing data 2					Pre-processing data 3				
emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff
BERT	SVM	0.66	0.63	0.03	BERT	SVM	0.77	0.69	0.08	BERT	SVM	0.66	0.63	0.03	BERT	SVM	0.75	0.66	0.09
BERT	RF	1.00	0.66	0.34	BERT	RF	1.00	0.68	0.32	BERT	RF	1.00	0.67	0.33	BERT	RF	1.00	0.65	0.35
BERT	LR	0.86	0.74	0.12	BERT	LR	0.99	0.72	0.27	BERT	LR	0.86	0.70	0.16	BERT	LR	0.99	0.72	0.27
BERT	KNN	0.81	0.62	0.19	BERT	KNN	0.83	0.69	0.14	BERT	KNN	0.84	0.64	0.20	BERT	KNN	0.84	0.65	0.19
BERT	EXT	1.00	0.69	0.31	BERT	EXT	1.00	0.67	0.33	BERT	EXT	1.00	0.67	0.33	BERT	EXT	1.00	0.66	0.34
Glove	SVM	0.66	0.63	0.03	Glove	SVM	0.66	0.63	0.03	Glove	SVM	0.66	0.63	0.03	Glove	SVM	0.66	0.63	0.03
Glove	RF	1.00	0.70	0.30	Glove	RF	1.00	0.69	0.31	Glove	RF	1.00	0.68	0.32	Glove	RF	1.00	0.70	0.30
Glove	LR	0.68	0.66	0.02	Glove	LR	0.69	0.67	0.02	Glove	LR	0.67	0.66	0.01	Glove	LR	0.68	0.63	0.05
Glove	KNN	0.80	0.62	0.18	Glove	KNN	0.80	0.64	0.16	Glove	KNN	0.80	0.60	0.20	Glove	KNN	0.80	0.61	0.19
Glove	EXT	1.00	0.72	0.28	Glove	EXT	1.00	0.71	0.29	Glove	EXT	1.00	0.70	0.30	Glove	EXT	1.00	0.66	0.34
TF-IDF	**SVM**	**0.93**	**0.79**	**0.14**	**TF-IDF**	**SVM**	**0.92**	**0.80**	**0.12**	**TF-IDF**	**SVM**	**0.93**	**0.80**	**0.13**	**TF-IDF**	**SVM**	**0.92**	**0.79**	**0.13**
TF-IDF	RF	1.00	0.76	0.24	TF-IDF	RF	1.00	0.74	0.26	TF-IDF	RF	1.00	0.76	0.24	TF-IDF	RF	1.00	0.76	0.24
TF-IDF	**LR**	**0.89**	**0.76**	**0.13**	**TF-IDF**	**LR**	**0.88**	**0.77**	**0.11**	**TF-IDF**	**LR**	**0.89**	**0.75**	**0.14**	**TF-IDF**	**LR**	**0.89**	**0.77**	**0.12**
TF-IDF	KNN	0.87	0.72	0.15	TF-IDF	KNN	0.88	0.73	0.15	TF-IDF	KNN	0.88	0.70	0.18	TF-IDF	KNN	0.89	0.72	0.17
TF-IDF	EXT	1.00	0.78	0.22	TF-IDF	EXT	1.00	0.78	0.22	TF-IDF	EXT	1.00	0.75	0.25	TF-IDF	EXT	1.00	0.78	0.22
Word2Vec	SVM	0.65	0.63	0.02	Word2Vec	SVM	0.66	0.63	0.03	Word2Vec	SVM	0.65	0.63	0.02	Word2Vec	SVM	0.66	0.63	0.03
Word2Vec	RF	1.00	0.70	0.30	Word2Vec	RF	1.00	0.67	0.33	Word2Vec	RF	1.00	0.66	0.34	Word2Vec	RF	1.00	0.69	0.31
Word2Vec	LR	0.66	0.63	0.03	Word2Vec	LR	0.66	0.63	0.03	Word2Vec	LR	0.66	0.63	0.03	Word2Vec	LR	0.66	0.63	0.03
Word2Vec	KNN	0.81	0.60	0.21	Word2Vec	KNN	0.82	0.64	0.18	Word2Vec	KNN	0.81	0.61	0.20	Word2Vec	KNN	0.81	0.63	0.18
Word2Vec	EXT	1.00	0.69	0.31	Word2Vec	EXT	1.00	0.68	0.32	Word2Vec	EXT	1.00	0.68	0.32	Word2Vec	EXT	1.00	0.66	0.34
Pre-processing data 4					Pre-processing data 5					Pre-processing data 6					Pre-processing data 7				
emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff
BERT	SVM	0.66	0.63	0.03	BERT	SVM	0.76	0.68	0.08	BERT	SVM	0.66	0.63	0.03	**BERT**	**SVM**	**0.77**	**0.70**	**0.07**
BERT	RF	1.00	0.70	0.30	BERT	RF	1.00	0.67	0.33	BERT	RF	1.00	0.66	0.34	BERT	RF	1.00	0.64	0.36
BERT	LR	0.86	0.72	0.14	BERT	LR	0.98	0.71	0.27	BERT	LR	0.87	0.72	0.15	BERT	LR	0.99	0.74	0.25
BERT	KNN	0.82	0.65	0.17	BERT	KNN	0.84	0.66	0.18	BERT	KNN	0.83	0.66	0.17	BERT	KNN	0.85	0.66	0.19
BERT	EXT	1.00	0.68	0.32	BERT	EXT	1.00	0.65	0.35	BERT	EXT	1.00	0.68	0.32	BERT	EXT	1.00	0.66	0.34
Glove	SVM	0.66	0.63	0.03	Glove	SVM	0.66	0.63	0.03	Glove	SVM	0.66	0.63	0.03	Glove	SVM	0.66	0.63	0.03
Glove	RF	1.00	0.70	0.30	Glove	RF	1.00	0.69	0.31	Glove	RF	1.00	0.69	0.31	Glove	RF	1.00	0.67	0.33
Glove	LR	0.67	0.65	0.02	Glove	LR	0.69	0.65	0.04	Glove	LR	0.67	0.65	0.02	Glove	LR	0.69	0.63	0.06
Glove	KNN	0.80	0.62	0.18	Glove	KNN	0.80	0.64	0.16	Glove	KNN	0.81	0.61	0.20	Glove	KNN	0.80	0.63	0.17
Glove	EXT	1.00	0.68	0.32	Glove	EXT	1.00	0.71	0.29	Glove	EXT	1.00	0.70	0.30	Glove	EXT	1.00	0.68	0.32
TF-IDF	**SVM**	**0.93**	**0.79**	**0.14**	**TF-IDF**	**SVM**	**0.91**	**0.80**	**0.11**	**TF-IDF**	**SVM**	**0.93**	**0.79**	**0.14**	**TF-IDF**	**SVM**	**0.92**	**0.79**	**0.13**
TF-IDF	RF	1.00	0.76	0.24	TF-IDF	RF	1.00	0.75	0.25	TF-IDF	RF	1.00	0.76	0.24	TF-IDF	RF	1.00	0.77	0.23
TF-IDF	**LR**	**0.88**	**0.76**	**0.12**	**TF-IDF**	**LR**	**0.88**	**0.77**	**0.11**	**TF-IDF**	**LR**	**0.88**	**0.75**	**0.13**	**TF-IDF**	**LR**	**0.89**	**0.77**	**0.12**
TF-IDF	KNN	0.88	0.73	0.15	TF-IDF	KNN	0.88	0.73	0.15	TF-IDF	KNN	0.88	0.72	0.16	TF-IDF	KNN	0.89	0.73	0.16
TF-IDF	EXT	1.00	0.78	0.22	TF-IDF	EXT	1.00	0.78	0.22	TF-IDF	EXT	1.00	0.75	0.25	TF-IDF	EXT	1.00	0.78	0.22
Word2Vec	SVM	0.65	0.63	0.02	Word2Vec	SVM	0.66	0.63	0.03	Word2Vec	SVM	0.65	0.64	0.01	Word2Vec	SVM	0.66	0.63	0.03
Word2Vec	RF	1.00	0.68	0.32	Word2Vec	RF	1.00	0.68	0.32	Word2Vec	RF	1.00	0.68	0.32	Word2Vec	RF	1.00	0.66	0.34
Word2Vec	LR	0.66	0.63	0.03	Word2Vec	LR	0.66	0.63	0.03	Word2Vec	LR	0.66	0.63	0.03	Word2Vec	LR	0.66	0.63	0.03
Word2Vec	KNN	0.81	0.63	0.18	Word2Vec	KNN	0.81	0.64	0.17	Word2Vec	KNN	0.82	0.61	0.21	Word2Vec	KNN	0.81	0.60	0.21
Word2Vec	EXT	1.00	0.69	0.31	Word2Vec	EXT	1.00	0.70	0.30	Word2Vec	EXT	1.00	0.66	0.34	Word2Vec	EXT	1.00	0.68	0.32

Note: The cell(s) highlighted in yellow with red colored text indicate the best performing case(s). The cell(s) highlighted in orange indicate the second best performing case(s).

Table 16.

A summary of all the customer data results (Study 1)

Control Mutuality

Pre-processing data 5				
emb	model	train	test	diff
TF-IDF	**LR**	**0.93**	**0.83**	**0.10**
Pre-processing data 7				
TF-IDF	**LR**	**0.93**	**0.83**	**0.10**

Commitment

Pre-processing data 6				
emb	model	train	test	diff
TF-IDF	**LR**	**0.92**	**0.85**	**0.07**
Pre-processing data 7				
TF-IDF	**LR**	**0.92**	**0.85**	**0.07**

Satisfaction

Pre-processing data 1				
emb	model	train	test	diff
TF-IDF	**SVM**	**0.93**	**0.85**	**0.08**
Pre-processing data 3				
TF-IDF	**SVM**	**0.93**	**0.85**	**0.08**
Pre-processing data 7				
TF-IDF	**SVM**	**0.93**	**0.85**	**0.08**

Trust

Pre-processing data 1				
emb	model	train	test	diff
TF-IDF	**SVM**	**0.92**	**0.80**	**0.12**
Pre-processing data 5				
TF-IDF	**SVM**	**0.91**	**0.80**	**0.11**

Table 17.

Results (Control Mutuality, Study2- Survey data)

Pre-processing data 0					Pre-processing data 1					Pre-processing data 2					Pre-processing data 3				
emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff
BERT	SVM	0.61	0.37	0.24	BERT	SVM	0.74	0.51	0.23	BERT	SVM	0.61	0.37	0.24	BERT	SVM	0.77	0.49	0.28
BERT	RF	1.00	0.49	0.51	BERT	RF	1.00	0.51	0.49	BERT	RF	1.00	0.47	0.53	BERT	RF	1.00	0.51	0.49
BERT	LR	0.95	0.67	0.28	BERT	LR	1.00	0.74	0.26	BERT	LR	0.96	0.53	0.43	BERT	LR	1.00	0.70	0.30
BERT	KNN	0.79	0.49	0.30	BERT	KNN	0.83	0.65	0.18	BERT	KNN	0.74	0.49	0.25	BERT	KNN	0.85	0.60	0.25
BERT	EXT	1.00	0.42	0.58	BERT	EXT	1.00	0.51	0.49	BERT	EXT	1.00	0.47	0.53	BERT	EXT	1.00	0.58	0.42
Glove	SVM	0.61	0.37	0.24	Glove	SVM	0.61	0.37	0.24	Glove	SVM	0.61	0.37	0.24	Glove	SVM	0.61	0.37	0.24
Glove	RF	1.00	0.47	0.53	Glove	RF	1.00	0.44	0.56	Glove	RF	1.00	0.44	0.56	Glove	RF	1.00	0.47	0.53
Glove	LR	0.63	0.42	0.21	Glove	LR	0.65	0.44	0.21	Glove	LR	0.64	0.40	0.24	Glove	LR	0.65	0.42	0.23
Glove	KNN	0.81	0.49	0.32	Glove	KNN	0.79	0.53	0.26	Glove	KNN	0.83	0.44	0.39	Glove	KNN	0.80	0.51	0.29
Glove	EXT	1.00	0.49	0.51	Glove	EXT	1.00	0.44	0.56	Glove	EXT	1.00	0.44	0.56	Glove	EXT	1.00	0.44	0.56
TF-IDF	SVM	0.95	0.47	0.48	TF-IDF	SVM	0.93	0.58	0.35	TF-IDF	SVM	0.95	0.51	0.44	TF-IDF	SVM	0.94	0.58	0.36
TF-IDF	RF	1.00	0.49	0.51	TF-IDF	RF	1.00	0.42	0.58	TF-IDF	RF	1.00	0.44	0.56	TF-IDF	RF	1.00	0.42	0.58
TF-IDF	LR	0.90	0.42	0.48	TF-IDF	LR	0.88	0.42	0.46	TF-IDF	LR	0.90	0.42	0.48	TF-IDF	LR	0.88	0.42	0.46
TF-IDF	KNN	0.84	0.60	0.24	TF-IDF	KNN	0.81	0.58	0.23	TF-IDF	KNN	0.82	0.56	0.26	TF-IDF	KNN	0.84	0.58	0.26
TF-IDF	EXT	1.00	0.44	0.56	TF-IDF	EXT	1.00	0.49	0.51	TF-IDF	EXT	1.00	0.49	0.51	TF-IDF	EXT	1.00	0.47	0.53
Word2Vec	SVM	0.77	0.81	-0.04	Word2Vec	SVM	0.77	0.81	-0.04	Word2Vec	SVM	0.77	0.81	-0.04	Word2Vec	SVM	0.77	0.81	-0.04
Word2Vec	RF	1.00	0.76	0.24	Word2Vec	RF	1.00	0.79	0.21	Word2Vec	RF	1.00	0.75	0.25	Word2Vec	RF	1.00	0.79	0.21
Word2Vec	LR	0.77	0.81	-0.04	Word2Vec	LR	0.77	0.81	-0.04	Word2Vec	LR	0.77	0.81	-0.04	Word2Vec	LR	0.77	0.81	-0.04
Word2Vec	KNN	0.83	0.62	0.21	**Word2Vec**	**KNN**	**0.80**	**0.73**	**0.07**	Word2Vec	KNN	0.79	0.68	0.11	Word2Vec	KNN	0.79	0.68	0.11
Word2Vec	EXT	1.00	0.78	0.22	Word2Vec	EXT	1.00	0.81	0.19	Word2Vec	EXT	1.00	0.76	0.24	Word2Vec	EXT	1.00	0.78	0.22

Pre-processing data 4					Pre-processing data 5					Pre-processing data 6					Pre-processing data 7				
emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff
BERT	SVM	0.61	0.37	0.24	BERT	SVM	0.75	0.49	0.26	BERT	SVM	0.61	0.37	0.24	BERT	SVM	0.77	0.49	0.28
BERT	RF	1.00	0.51	0.49	BERT	RF	1.00	0.56	0.44	BERT	RF	1.00	0.44	0.56	BERT	RF	1.00	0.51	0.49
BERT	LR	0.95	0.70	0.25	BERT	LR	1.00	0.72	0.28	BERT	LR	0.96	0.53	0.43	BERT	LR	1.00	0.70	0.30
BERT	KNN	0.78	0.49	0.29	BERT	KNN	0.85	0.67	0.18	BERT	KNN	0.73	0.49	0.24	BERT	KNN	0.85	0.60	0.25
BERT	EXT	1.00	0.47	0.53	BERT	EXT	1.00	0.51	0.49	BERT	EXT	1.00	0.53	0.47	BERT	EXT	1.00	0.60	0.40
Glove	SVM	0.61	0.37	0.24	Glove	SVM	0.61	0.37	0.24	Glove	SVM	0.61	0.37	0.24	Glove	SVM	0.61	0.37	0.24
Glove	RF	1.00	0.49	0.51	Glove	RF	1.00	0.42	0.58	Glove	RF	1.00	0.47	0.53	Glove	RF	1.00	0.51	0.49
Glove	LR	0.63	0.42	0.21	Glove	LR	0.65	0.42	0.23	Glove	LR	0.65	0.40	0.25	Glove	LR	0.65	0.42	0.23
Glove	KNN	0.82	0.53	0.29	Glove	KNN	0.79	0.53	0.26	Glove	KNN	0.83	0.42	0.41	Glove	KNN	0.79	0.51	0.28
Glove	EXT	1.00	0.51	0.49	Glove	EXT	1.00	0.47	0.53	Glove	EXT	1.00	0.44	0.56	Glove	EXT	1.00	0.47	0.53
TF-IDF	SVM	0.94	0.49	0.45	TF-IDF	SVM	0.93	0.58	0.35	TF-IDF	SVM	0.95	0.53	0.42	TF-IDF	SVM	0.95	0.58	0.37
TF-IDF	RF	1.00	0.44	0.56	TF-IDF	RF	1.00	0.47	0.53	TF-IDF	RF	1.00	0.40	0.60	TF-IDF	RF	1.00	0.42	0.58
TF-IDF	LR	0.89	0.44	0.45	TF-IDF	LR	0.89	0.40	0.49	TF-IDF	LR	0.90	0.44	0.46	TF-IDF	LR	0.89	0.40	0.49
TF-IDF	KNN	0.83	0.53	0.30	TF-IDF	KNN	0.82	0.56	0.26	TF-IDF	KNN	0.82	0.51	0.31	TF-IDF	KNN	0.85	0.56	0.29
TF-IDF	EXT	1.00	0.42	0.58	TF-IDF	EXT	1.00	0.53	0.47	TF-IDF	EXT	1.00	0.49	0.51	TF-IDF	EXT	1.00	0.49	0.51
Word2Vec	SVM	0.77	0.81	-0.04	Word2Vec	SVM	0.77	0.81	-0.04	Word2Vec	SVM	0.77	0.81	-0.04	Word2Vec	SVM	0.77	0.81	-0.04
Word2Vec	RF	1.00	0.76	0.24	Word2Vec	RF	1.00	0.79	0.21	Word2Vec	RF	1.00	0.83	0.17	Word2Vec	RF	1.00	0.73	0.27
Word2Vec	LR	0.77	0.81	-0.04	Word2Vec	LR	0.77	0.81	-0.04	Word2Vec	LR	0.77	0.81	-0.04	Word2Vec	LR	0.77	0.81	-0.04
Word2Vec	KNN	0.83	0.60	0.23	**Word2Vec**	**KNN**	**0.79**	**0.75**	**0.04**	Word2Vec	KNN	0.80	0.65	0.15	**Word2Vec**	**KNN**	**0.81**	**0.73**	**0.08**
Word2Vec	EXT	1.00	0.79	0.21	Word2Vec	EXT	1.00	0.78	0.22	Word2Vec	EXT	1.00	0.78	0.22	Word2Vec	EXT	1.00	0.79	0.21

Note: The cell(s) highlighted in yellow with red colored text indicate the best performing case(s). The cell(s) highlighted in orange indicate the second best performing case(s).

Table 18.

Results (Commitment, Study2- Survey data)

Pre-processing data 0					Pre-processing data 1					Pre-processing data 2					Pre-processing data 3				
emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff
BERT	SVM	0.62	0.58	0.04	BERT	SVM	0.76	0.65	0.11	BERT	SVM	0.62	0.58	0.04	BERT	SVM	0.77	0.63	0.14
BERT	RF	1.00	0.60	0.40	BERT	RF	1.00	0.65	0.35	BERT	RF	1.00	0.56	0.44	BERT	RF	1.00	0.67	0.33
BERT	LR	0.96	0.74	0.22	BERT	LR	1.00	0.77	0.23	BERT	LR	0.95	0.65	0.30	BERT	LR	1.00	0.74	0.26
BERT	KNN	0.83	0.40	0.43	BERT	KNN	0.85	0.67	0.18	BERT	KNN	0.80	0.53	0.27	BERT	KNN	0.87	0.60	0.27
BERT	EXT	1.00	0.58	0.42	BERT	EXT	1.00	0.65	0.35	BERT	EXT	1.00	0.60	0.40	BERT	EXT	1.00	0.63	0.37
Glove	SVM	0.62	0.58	0.04	Glove	SVM	0.62	0.58	0.04	Glove	SVM	0.62	0.58	0.04	Glove	SVM	0.62	0.58	0.04
Glove	RF	1.00	0.60	0.40	Glove	RF	1.00	0.49	0.51	Glove	RF	1.00	0.58	0.42	Glove	RF	1.00	0.49	0.51
Glove	LR	0.65	0.58	0.07	Glove	LR	0.68	0.58	0.10	Glove	LR	0.66	0.58	0.08	Glove	LR	0.68	0.58	0.10
Glove	KNN	0.80	0.56	0.24	Glove	KNN	0.83	0.49	0.34	Glove	KNN	0.79	0.60	0.19	Glove	KNN	0.80	0.49	0.31
Glove	EXT	1.00	0.60	0.40	Glove	EXT	1.00	0.58	0.42	Glove	EXT	1.00	0.60	0.40	Glove	EXT	1.00	0.56	0.44
TF-IDF	SVM	0.95	0.72	0.23	TF-IDF	SVM	0.94	0.72	0.22	TF-IDF	SVM	0.95	0.74	0.21	TF-IDF	SVM	0.95	0.74	0.21
TF-IDF	RF	1.00	0.58	0.42	TF-IDF	RF	1.00	0.56	0.44	TF-IDF	RF	1.00	0.65	0.35	TF-IDF	RF	1.00	0.65	0.35
TF-IDF	LR	0.86	0.58	0.28	TF-IDF	LR	0.88	0.63	0.25	TF-IDF	LR	0.86	0.56	0.30	TF-IDF	LR	0.88	0.60	0.28
TF-IDF	KNN	0.85	0.67	0.18	**TF-IDF**	**KNN**	**0.84**	**0.74**	**0.10**	TF-IDF	KNN	0.85	0.67	0.18	**TF-IDF**	**KNN**	**0.85**	**0.77**	**0.08**
TF-IDF	EXT	1.00	0.72	0.28	TF-IDF	EXT	1.00	0.70	0.30	TF-IDF	EXT	1.00	0.67	0.33	TF-IDF	EXT	1.00	0.70	0.30
Word2Vec	SVM	0.62	0.58	0.04	Word2Vec	SVM	0.62	0.58	0.04	Word2Vec	SVM	0.62	0.58	0.04	Word2Vec	SVM	0.62	0.58	0.04
Word2Vec	RF	1.00	0.53	0.47	Word2Vec	RF	1.00	0.58	0.42	Word2Vec	RF	1.00	0.63	0.37	Word2Vec	RF	1.00	0.49	0.51
Word2Vec	LR	0.62	0.58	0.04	Word2Vec	LR	0.62	0.58	0.04	Word2Vec	LR	0.62	0.58	0.04	Word2Vec	LR	0.62	0.58	0.04
Word2Vec	KNN	0.79	0.58	0.21	Word2Vec	KNN	0.77	0.60	0.17	Word2Vec	KNN	0.76	0.56	0.20	Word2Vec	KNN	0.79	0.63	0.16
Word2Vec	EXT	1.00	0.53	0.47	Word2Vec	EXT	1.00	0.60	0.40	Word2Vec	EXT	1.00	0.60	0.40	Word2Vec	EXT	1.00	0.60	0.40

Pre-processing data 4					Pre-processing data 5					Pre-processing data 6					Pre-processing data 7				
emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff
BERT	SVM	0.62	0.58	0.04	BERT	SVM	0.75	0.63	0.12	BERT	SVM	0.62	0.58	0.04	BERT	SVM	0.74	0.65	0.09
BERT	RF	1.00	0.58	0.42	BERT	RF	1.00	0.70	0.30	BERT	RF	1.00	0.63	0.37	BERT	RF	1.00	0.63	0.37
BERT	LR	0.96	0.72	0.24	BERT	LR	1.00	0.74	0.26	BERT	LR	0.96	0.70	0.26	BERT	LR	1.00	0.74	0.26
BERT	KNN	0.84	0.44	0.40	BERT	KNN	0.84	0.63	0.21	BERT	KNN	0.79	0.56	0.23	BERT	KNN	0.88	0.63	0.25
BERT	EXT	1.00	0.60	0.40	BERT	EXT	1.00	0.65	0.35	BERT	EXT	1.00	0.58	0.42	BERT	EXT	1.00	0.60	0.40
Glove	SVM	0.62	0.58	0.04	Glove	SVM	0.62	0.58	0.04	Glove	SVM	0.62	0.58	0.04	Glove	SVM	0.62	0.58	0.04
Glove	RF	1.00	0.58	0.42	Glove	RF	1.00	0.42	0.58	Glove	RF	1.00	0.58	0.42	Glove	RF	1.00	0.58	0.42
Glove	LR	0.65	0.58	0.07	Glove	LR	0.68	0.58	0.10	Glove	LR	0.66	0.58	0.08	Glove	LR	0.68	0.58	0.10
Glove	KNN	0.78	0.56	0.22	Glove	KNN	0.83	0.51	0.32	Glove	KNN	0.78	0.58	0.20	Glove	KNN	0.81	0.49	0.32
Glove	EXT	1.00	0.63	0.37	Glove	EXT	1.00	0.53	0.47	Glove	EXT	1.00	0.58	0.42	Glove	EXT	1.00	0.58	0.42
TF-IDF	SVM	0.96	0.72	0.24	TF-IDF	SVM	0.95	0.72	0.23	TF-IDF	SVM	0.95	0.74	0.21	TF-IDF	SVM	0.95	0.74	0.21
TF-IDF	RF	1.00	0.56	0.44	TF-IDF	RF	1.00	0.56	0.44	TF-IDF	RF	1.00	0.60	0.40	TF-IDF	RF	1.00	0.58	0.42
TF-IDF	LR	0.87	0.60	0.27	TF-IDF	LR	0.88	0.63	0.25	TF-IDF	LR	0.86	0.58	0.28	TF-IDF	LR	0.87	0.60	0.27
TF-IDF	KNN	0.87	0.67	0.20	**TF-IDF**	**KNN**	**0.85**	**0.74**	**0.11**	TF-IDF	KNN	0.87	0.67	0.20	**TF-IDF**	**KNN**	**0.85**	**0.74**	**0.11**
TF-IDF	EXT	1.00	0.70	0.30	TF-IDF	EXT	1.00	0.70	0.30	TF-IDF	EXT	1.00	0.65	0.35	TF-IDF	EXT	1.00	0.65	0.35
Word2Vec	SVM	0.62	0.58	0.04	Word2Vec	SVM	0.62	0.58	0.04	Word2Vec	SVM	0.62	0.58	0.04	Word2Vec	SVM	0.62	0.58	0.04
Word2Vec	RF	1.00	0.49	0.51	Word2Vec	RF	1.00	0.53	0.47	Word2Vec	RF	1.00	0.60	0.40	Word2Vec	RF	1.00	0.56	0.44
Word2Vec	LR	0.62	0.58	0.04	Word2Vec	LR	0.62	0.58	0.04	Word2Vec	LR	0.62	0.58	0.04	Word2Vec	LR	0.62	0.58	0.04
Word2Vec	KNN	0.80	0.56	0.24	Word2Vec	KNN	0.75	0.60	0.15	Word2Vec	KNN	0.75	0.58	0.17	Word2Vec	KNN	0.80	0.63	0.17
Word2Vec	EXT	1.00	0.53	0.47	Word2Vec	EXT	1.00	0.58	0.42	Word2Vec	EXT	1.00	0.63	0.37	Word2Vec	EXT	1.00	0.58	0.42

Note: The cell(s) highlighted in yellow with red colored text indicate the best performing case(s). The cell(s) highlighted in orange indicate the second best performing case(s).

Table 19.

Results (Satisfaction, Study2- Survey data)

Pre-processing data 0					Pre-processing data 1					Pre-processing data 2					Pre-processing data 3				
emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff
BERT	SVM	0.92	0.84	0.08	BERT	SVM	0.92	0.84	0.08	BERT	SVM	0.92	0.84	0.08	BERT	SVM	0.92	0.84	0.08
BERT	RF	1.00	0.84	0.16	BERT	RF	1.00	0.84	0.16	BERT	RF	1.00	0.84	0.16	BERT	RF	1.00	0.84	0.16
BERT	LR	0.98	0.84	0.14	BERT	LR	1.00	0.88	0.12	BERT	LR	0.98	0.84	0.14	BERT	LR	1.00	0.88	0.12
BERT	KNN	0.92	0.84	0.08	BERT	KNN	0.93	0.88	0.05	BERT	KNN	0.93	0.79	0.14	BERT	KNN	0.95	0.88	0.07
BERT	EXT	1.00	0.84	0.16	BERT	EXT	1.00	0.84	0.16	BERT	EXT	1.00	0.84	0.16	BERT	EXT	1.00	0.84	0.16
Glove	SVM	0.92	0.84	0.08	Glove	SVM	0.92	0.84	0.08	Glove	SVM	0.92	0.84	0.08	Glove	SVM	0.92	0.84	0.08
Glove	RF	1.00	0.84	0.16	Glove	RF	1.00	0.86	0.14	Glove	RF	1.00	0.84	0.16	Glove	RF	1.00	0.86	0.14
Glove	LR	0.92	0.84	0.08	Glove	LR	0.92	0.84	0.08	Glove	LR	0.92	0.84	0.08	Glove	LR	0.92	0.84	0.08
Glove	KNN	0.92	0.79	0.13	Glove	KNN	0.93	0.84	0.09	Glove	KNN	0.92	0.81	0.11	Glove	KNN	0.91	0.79	0.12
Glove	EXT	1.00	0.84	0.16	Glove	EXT	1.00	0.84	0.16	Glove	EXT	1.00	0.84	0.16	Glove	EXT	1.00	0.84	0.16
TF-IDF	SVM	0.92	0.84	0.08	TF-IDF	SVM	0.92	0.84	0.08	TF-IDF	SVM	0.92	0.84	0.08	TF-IDF	SVM	0.92	0.84	0.08
TF-IDF	RF	1.00	0.84	0.16	TF-IDF	RF	1.00	0.84	0.16	TF-IDF	RF	1.00	0.84	0.16	TF-IDF	RF	1.00	0.84	0.16
TF-IDF	LR	0.92	0.84	0.08	TF-IDF	LR	0.92	0.84	0.08	TF-IDF	LR	0.92	0.84	0.08	TF-IDF	LR	0.92	0.84	0.08
TF-IDF	KNN	0.96	0.81	0.15	TF-IDF	KNN	0.96	0.86	0.10	TF-IDF	KNN	0.96	0.84	0.12	TF-IDF	KNN	0.96	0.84	0.12
TF-IDF	EXT	1.00	0.84	0.16	TF-IDF	EXT	1.00	0.84	0.16	TF-IDF	EXT	1.00	0.84	0.16	TF-IDF	EXT	1.00	0.84	0.16
Word2Vec	SVM	0.92	0.84	0.08	Word2Vec	SVM	0.92	0.84	0.08	Word2Vec	SVM	0.92	0.84	0.08	Word2Vec	SVM	0.92	0.84	0.08
Word2Vec	RF	1.00	0.84	0.16	Word2Vec	RF	1.00	0.81	0.19	Word2Vec	RF	1.00	0.84	0.16	Word2Vec	RF	1.00	0.84	0.16
Word2Vec	LR	0.92	0.84	0.08	Word2Vec	LR	0.92	0.84	0.08	Word2Vec	LR	0.92	0.84	0.08	Word2Vec	LR	0.92	0.84	0.08
Word2Vec	KNN	0.94	0.79	0.15	Word2Vec	KNN	0.95	0.81	0.14	Word2Vec	KNN	0.93	0.79	0.14	Word2Vec	KNN	0.95	0.81	0.14
Word2Vec	EXT	1.00	0.84	0.16	Word2Vec	EXT	1.00	0.84	0.16	Word2Vec	EXT	1.00	0.84	0.16	Word2Vec	EXT	1.00	0.81	0.19

Pre-processing data 4					Pre-processing data 5					Pre-processing data 6					Pre-processing data 7				
emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff
BERT	SVM	0.92	0.84	0.08	BERT	SVM	0.92	0.84	0.08	BERT	SVM	0.92	0.84	0.08	BERT	SVM	0.92	0.84	0.08
BERT	RF	1.00	0.84	0.16	BERT	RF	1.00	0.84	0.16	BERT	RF	1.00	0.84	0.16	BERT	RF	1.00	0.84	0.16
BERT	LR	0.98	0.84	0.14	BERT	LR	1.00	0.88	0.12	BERT	LR	0.98	0.84	0.14	BERT	LR	1.00	0.88	0.12
BERT	KNN	0.92	0.84	0.08	BERT	KNN	0.93	0.88	0.05	BERT	KNN	0.93	0.79	0.14	BERT	KNN	0.95	0.88	0.07
BERT	EXT	1.00	0.84	0.16	BERT	EXT	1.00	0.84	0.16	BERT	EXT	1.00	0.84	0.16	BERT	EXT	1.00	0.84	0.16
Glove	SVM	0.92	0.84	0.08	Glove	SVM	0.92	0.84	0.08	Glove	SVM	0.92	0.84	0.08	Glove	SVM	0.92	0.84	0.08
Glove	RF	1.00	0.86	0.14	Glove	RF	1.00	0.86	0.14	Glove	RF	1.00	0.84	0.16	Glove	RF	1.00	0.86	0.14
Glove	LR	0.92	0.84	0.08	Glove	LR	0.92	0.84	0.08	Glove	LR	0.92	0.84	0.08	Glove	LR	0.92	0.84	0.08
Glove	KNN	0.93	0.79	0.14	Glove	KNN	0.92	0.81	0.11	Glove	KNN	0.93	0.81	0.12	Glove	KNN	0.91	0.77	0.14
Glove	EXT	1.00	0.84	0.16	Glove	EXT	1.00	0.84	0.16	Glove	EXT	1.00	0.84	0.16	Glove	EXT	1.00	0.84	0.16
TF-IDF	SVM	0.92	0.84	0.08	TF-IDF	SVM	0.92	0.84	0.08	TF-IDF	SVM	0.92	0.84	0.08	TF-IDF	SVM	0.92	0.84	0.08
TF-IDF	RF	1.00	0.84	0.16	TF-IDF	RF	1.00	0.84	0.16	TF-IDF	RF	1.00	0.84	0.16	TF-IDF	RF	1.00	0.84	0.16
TF-IDF	LR	0.92	0.84	0.08	TF-IDF	LR	0.92	0.84	0.08	TF-IDF	LR	0.92	0.84	0.08	TF-IDF	LR	0.92	0.84	0.08
TF-IDF	KNN	0.96	0.86	0.10	TF-IDF	KNN	0.96	0.88	0.08	TF-IDF	KNN	0.96	0.86	0.10	TF-IDF	KNN	0.96	0.86	0.10
TF-IDF	EXT	1.00	0.84	0.16	TF-IDF	EXT	1.00	0.84	0.16	TF-IDF	EXT	1.00	0.84	0.16	TF-IDF	EXT	1.00	0.84	0.16
Word2Vec	SVM	0.92	0.84	0.08	Word2Vec	SVM	0.92	0.84	0.08	Word2Vec	SVM	0.92	0.84	0.08	Word2Vec	SVM	0.92	0.84	0.08
Word2Vec	RF	1.00	0.84	0.16	Word2Vec	RF	1.00	0.81	0.19	Word2Vec	RF	1.00	0.81	0.19	Word2Vec	RF	1.00	0.84	0.16
Word2Vec	LR	0.92	0.84	0.08	Word2Vec	LR	0.92	0.84	0.08	Word2Vec	LR	0.92	0.84	0.08	Word2Vec	LR	0.92	0.84	0.08
Word2Vec	KNN	0.93	0.79	0.14	Word2Vec	KNN	0.95	0.81	0.14	Word2Vec	KNN	0.93	0.77	0.16	Word2Vec	KNN	0.95	0.81	0.14
Word2Vec	EXT	1.00	0.84	0.16	Word2Vec	EXT	1.00	0.84	0.16	Word2Vec	EXT	1.00	0.84	0.16	Word2Vec	EXT	1.00	0.84	0.16

Note: The cell(s) highlighted in yellow with red colored text indicate the best performing case(s). The cell(s) highlighted in orange indicate the second best performing case(s).

Table 20.

Results (Trust, Study2- Survey data)

Pre-processing data 0				
emb	model	train	test	diff
BERT	SVM	0.77	0.81	-0.04
BERT	RF	1.00	0.75	0.25
BERT	LR	0.96	0.75	0.21
BERT	KNN	0.87	0.71	0.16
BERT	EXT	1.00	0.75	0.25
Glove	SVM	0.77	0.81	-0.04
Glove	RF	1.00	0.75	0.25
Glove	LR	0.77	0.81	-0.04
Glove	KNN	0.85	0.65	0.20
Glove	EXT	1.00	0.76	0.24
TF-IDF	**SVM**	**0.92**	**0.79**	**0.13**
TF-IDF	RF	1.00	0.81	0.19
TF-IDF	LR	0.78	0.81	-0.03
TF-IDF	**KNN**	**0.86**	**0.78**	**0.08**
TF-IDF	EXT	1.00	0.81	0.19
Word2Vec	SVM	0.77	0.81	-0.04
Word2Vec	RF	1.00	0.76	0.24
Word2Vec	LR	0.77	0.81	-0.04
Word2Vec	KNN	0.83	0.62	0.21
Word2Vec	EXT	1.00	0.78	0.22

Pre-processing data 1				
emb	model	train	test	diff
BERT	SVM	0.77	0.81	-0.04
BERT	RF	1.00	0.81	0.19
BERT	LR	1.00	0.81	0.19
BERT	KNN	0.86	0.71	0.15
BERT	EXT	1.00	0.81	0.19
Glove	SVM	0.77	0.81	-0.04
Glove	RF	1.00	0.78	0.22
Glove	LR	0.78	0.81	-0.03
Glove	KNN	0.82	0.70	0.12
Glove	EXT	1.00	0.79	0.21
TF-IDF	**SVM**	**0.91**	**0.81**	**0.10**
TF-IDF	RF	1.00	0.81	0.19
TF-IDF	LR	0.78	0.81	-0.03
TF-IDF	KNN	0.88	0.73	0.15
TF-IDF	EXT	1.00	0.83	0.17
Word2Vec	SVM	0.77	0.81	-0.04
Word2Vec	RF	1.00	0.79	0.21
Word2Vec	LR	0.77	0.81	-0.04
Word2Vec	**KNN**	**0.80**	**0.73**	**0.07**
Word2Vec	EXT	1.00	0.81	0.19

Pre-processing data 2				
emb	model	train	test	diff
BERT	SVM	0.77	0.81	-0.04
BERT	RF	1.00	0.78	0.22
BERT	LR	0.96	0.78	0.18
BERT	KNN	0.85	0.60	0.25
BERT	EXT	1.00	0.78	0.22
Glove	SVM	0.77	0.81	-0.04
Glove	RF	1.00	0.75	0.25
Glove	LR	0.77	0.81	-0.04
Glove	KNN	0.82	0.67	0.15
Glove	EXT	1.00	0.73	0.27
TF-IDF	**SVM**	**0.92**	**0.79**	**0.13**
TF-IDF	RF	1.00	0.81	0.19
TF-IDF	LR	0.78	0.81	-0.03
TF-IDF	KNN	0.88	0.75	0.13
TF-IDF	EXT	1.00	0.81	0.19
Word2Vec	SVM	0.77	0.81	-0.04
Word2Vec	RF	1.00	0.75	0.25
Word2Vec	LR	0.77	0.81	-0.04
Word2Vec	KNN	0.79	0.68	0.11
Word2Vec	EXT	1.00	0.76	0.24

Pre-processing data 3				
emb	model	train	test	diff
BERT	SVM	0.79	0.81	-0.02
BERT	RF	1.00	0.81	0.19
BERT	LR	1.00	0.81	0.19
BERT	KNN	0.86	0.73	0.13
BERT	EXT	1.00	0.81	0.19
Glove	SVM	0.77	0.81	-0.04
Glove	RF	1.00	0.78	0.22
Glove	LR	0.78	0.81	-0.03
Glove	KNN	0.83	0.71	0.12
Glove	EXT	1.00	0.83	0.17
TF-IDF	**SVM**	**0.91**	**0.79**	**0.12**
TF-IDF	RF	1.00	0.81	0.19
TF-IDF	LR	0.78	0.81	-0.03
TF-IDF	**KNN**	0.88	0.73	0.15
TF-IDF	EXT	1.00	0.81	0.19
Word2Vec	SVM	0.77	0.81	-0.04
Word2Vec	RF	1.00	0.79	0.21
Word2Vec	LR	0.77	0.81	-0.04
Word2Vec	KNN	0.79	0.68	0.11
Word2Vec	EXT	1.00	0.78	0.22

Pre-processing data 4				
emb	model	train	test	diff
BERT	SVM	0.77	0.81	-0.04
BERT	RF	1.00	0.78	0.22
BERT	LR	0.96	0.78	0.18
BERT	KNN	0.87	0.71	0.16
BERT	EXT	1.00	0.78	0.22
Glove	SVM	0.77	0.81	-0.04
Glove	RF	1.00	0.76	0.24
Glove	LR	0.77	0.81	-0.04
Glove	KNN	0.84	0.65	0.19
Glove	EXT	1.00	0.73	0.27
TF-IDF	**SVM**	**0.92**	**0.79**	**0.13**
TF-IDF	RF	1.00	0.79	0.21
TF-IDF	LR	0.79	0.81	-0.02
TF-IDF	**KNN**	**0.87**	**0.78**	**0.09**
TF-IDF	EXT	1.00	0.81	0.19
Word2Vec	SVM	0.77	0.81	-0.04
Word2Vec	RF	1.00	0.76	0.24
Word2Vec	LR	0.77	0.81	-0.04
Word2Vec	KNN	0.83	0.60	0.23
Word2Vec	EXT	1.00	0.79	0.21

Pre-processing data 5				
emb	model	train	test	diff
BERT	SVM	0.77	0.81	-0.04
BERT	RF	1.00	0.81	0.19
BERT	LR	1.00	0.79	0.21
BERT	KNN	0.87	0.75	0.12
BERT	EXT	1.00	0.81	0.19
Glove	SVM	0.77	0.81	-0.04
Glove	RF	1.00	0.81	0.19
Glove	LR	0.78	0.81	-0.03
Glove	KNN	0.83	0.65	0.18
Glove	EXT	1.00	0.78	0.22
TF-IDF	**SVM**	**0.92**	**0.81**	**0.11**
TF-IDF	RF	1.00	0.81	0.19
TF-IDF	LR	0.79	0.81	-0.02
TF-IDF	**KNN**	**0.88**	**0.76**	**0.12**
TF-IDF	EXT	1.00	0.81	0.19
Word2Vec	SVM	0.77	0.81	-0.04
Word2Vec	RF	1.00	0.79	0.21
Word2Vec	LR	0.77	0.81	-0.04
Word2Vec	**KNN**	**0.79**	**0.75**	**0.04**
Word2Vec	EXT	1.00	0.78	0.22

Pre-processing data 6				
emb	model	train	test	diff
BERT	SVM	0.77	0.81	-0.04
BERT	RF	1.00	0.78	0.22
BERT	LR	0.97	0.76	0.21
BERT	KNN	0.86	0.57	0.29
BERT	EXT	1.00	0.78	0.22
Glove	SVM	0.77	0.81	-0.04
Glove	RF	1.00	0.75	0.25
Glove	LR	0.77	0.81	-0.04
Glove	KNN	0.81	0.68	0.13
Glove	EXT	1.00	0.76	0.24
TF-IDF	**SVM**	**0.93**	**0.79**	**0.14**
TF-IDF	RF	1.00	0.81	0.19
TF-IDF	LR	0.79	0.81	-0.02
TF-IDF	**KNN**	**0.88**	**0.78**	**0.10**
TF-IDF	EXT	1.00	0.79	0.21
Word2Vec	SVM	0.77	0.81	-0.04
Word2Vec	RF	1.00	0.83	0.17
Word2Vec	LR	0.77	0.81	-0.04
Word2Vec	KNN	0.80	0.65	0.15
Word2Vec	EXT	1.00	0.78	0.22

Pre-processing data 7				
emb	model	train	test	diff
BERT	SVM	0.78	0.81	-0.03
BERT	RF	1.00	0.81	0.19
BERT	LR	1.00	0.79	0.21
BERT	KNN	0.85	0.76	0.09
BERT	EXT	1.00	0.81	0.19
Glove	SVM	0.77	0.81	-0.04
Glove	RF	1.00	0.78	0.22
Glove	LR	0.78	0.81	-0.03
Glove	KNN	0.81	0.67	0.14
Glove	EXT	1.00	0.79	0.21
TF-IDF	**SVM**	**0.92**	**0.81**	**0.11**
TF-IDF	RF	1.00	0.81	0.19
TF-IDF	LR	0.79	0.81	-0.02
TF-IDF	**KNN**	**0.88**	**0.75**	**0.13**
TF-IDF	EXT	1.00	0.79	0.21
Word2Vec	SVM	0.77	0.81	-0.04
Word2Vec	RF	1.00	0.73	0.27
Word2Vec	LR	0.77	0.81	-0.04
Word2Vec	**KNN**	**0.81**	**0.73**	**0.08**
Word2Vec	EXT	1.00	0.79	0.21

Note: The cell(s) highlighted in yellow with red colored text indicate the best performing case(s). The cell(s) highlighted in orange indicate the second best performing case(s).

Table 21.
A summary of the employee data results (Study 2)

Control Mutuality

Pre-processing data 5				
emb	model	train	test	diff
Word2Vec	**KNN**	**0.79**	**0.75**	**0.04**

Commitment

Pre-processing data 3				
emb	model	train	test	diff
TF-IDF	**KNN**	**0.85**	**0.77**	**0.08**

Satisfaction

Pre-processing data 1				
emb	model	train	test	diff
BERT	**KNN**	**0.93**	**0.88**	**0.05**
Pre-processing data 5				
BERT	**KNN**	**0.93**	**0.88**	**0.05**
TF-IDF	**KNN**	**0.96**	**0.88**	**0.08**

Trust

Pre-processing data 1				
emb	model	train	test	diff
TF-IDF	**SVM**	**0.91**	**0.81**	**0.10**
Pre-processing data 5				
TF-IDF	**SVM**	**0.92**	**0.81**	**0.11**
Pre-processing data 7				
TF-IDF	**SVM**	**0.92**	**0.81**	**0.11**

Table 22.

Results (Control Mutuality, Study3- All data)

Pre-processing data 0					Pre-processing data 1					Pre-processing data 2					Pre-processing data 3				
emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff
BERT	SVM	0.63	0.68	-0.05	**BERT**	**SVM**	**0.75**	**0.70**	**0.05**	BERT	SVM	0.62	0.67	-0.05	**BERT**	**SVM**	**0.74**	**0.73**	**0.01**
BERT	RF	1.00	0.70	0.30	BERT	RF	1.00	0.67	0.33	BERT	RF	1.00	0.70	0.30	BERT	RF	1.00	0.67	0.33
BERT	LR	0.84	0.71	0.13	BERT	LR	0.97	0.68	0.29	BERT	LR	0.84	0.70	0.14	BERT	LR	0.98	0.72	0.26
BERT	KNN	0.80	0.64	0.16	BERT	KNN	0.83	0.67	0.16	BERT	KNN	0.80	0.61	0.19	BERT	KNN	0.83	0.69	0.14
BERT	EXT	1.00	0.67	0.33	BERT	EXT	1.00	0.71	0.29	BERT	EXT	1.00	0.66	0.34	BERT	EXT	1.00	0.70	0.30
Glove	SVM	0.52	0.56	-0.04	Glove	SVM	0.51	0.56	-0.05	Glove	SVM	0.52	0.56	-0.04	Glove	SVM	0.51	0.56	-0.05
Glove	RF	1.00	0.68	0.32	Glove	RF	1.00	0.61	0.39	Glove	RF	1.00	0.66	0.34	Glove	RF	1.00	0.65	0.35
Glove	LR	0.66	0.69	-0.03	Glove	LR	0.65	0.65	0.00	Glove	LR	0.66	0.69	-0.03	Glove	LR	0.62	0.66	-0.04
Glove	KNN	0.81	0.60	0.21	Glove	KNN	0.78	0.60	0.18	Glove	KNN	0.80	0.63	0.17	Glove	KNN	0.80	0.62	0.18
Glove	EXT	1.00	0.68	0.32	Glove	EXT	1.00	0.64	0.36	Glove	EXT	1.00	0.65	0.35	Glove	EXT	1.00	0.65	0.35
TF-IDF	**SVM**	**0.93**	**0.81**	**0.12**	**TF-IDF**	**SVM**	**0.91**	**0.83**	**0.08**	**TF-IDF**	**SVM**	**0.92**	**0.84**	**0.08**	**TF-IDF**	**SVM**	**0.91**	**0.83**	**0.08**
TF-IDF	RF	1.00	0.81	0.19	TF-IDF	RF	1.00	0.82	0.18	TF-IDF	RF	1.00	0.81	0.19	TF-IDF	RF	1.00	0.83	0.17
TF-IDF	**LR**	**0.92**	**0.85**	**0.07**	**TF-IDF**	**LR**	**0.92**	**0.85**	**0.07**	**TF-IDF**	**LR**	**0.93**	**0.85**	**0.08**	**TF-IDF**	**LR**	**0.92**	**0.86**	**0.06**
TF-IDF	KNN	0.88	0.71	0.17	TF-IDF	KNN	0.88	0.71	0.17	TF-IDF	KNN	0.88	0.70	0.18	TF-IDF	KNN	0.88	0.69	0.19
TF-IDF	EXT	1.00	0.84	0.16	TF-IDF	EXT	1.00	0.82	0.18	TF-IDF	EXT	1.00	0.81	0.19	TF-IDF	EXT	1.00	0.83	0.17
Word2Vec	SVM	0.53	0.56	-0.03	Word2Vec	SVM	0.52	0.57	-0.05	Word2Vec	SVM	0.52	0.58	-0.06	Word2Vec	SVM	0.52	0.57	-0.05
Word2Vec	RF	1.00	0.68	0.32	Word2Vec	RF	1.00	0.64	0.36	Word2Vec	RF	1.00	0.69	0.31	Word2Vec	RF	1.00	0.64	0.36
Word2Vec	LR	0.61	0.68	-0.07	Word2Vec	LR	0.63	0.66	-0.03	Word2Vec	LR	0.62	0.69	-0.07	Word2Vec	LR	0.63	0.66	-0.03
Word2Vec	KNN	0.78	0.60	0.18	Word2Vec	KNN	0.80	0.60	0.20	Word2Vec	KNN	0.79	0.61	0.18	Word2Vec	KNN	0.78	0.60	0.18
Word2Vec	EXT	1.00	0.68	0.32	Word2Vec	EXT	1.00	0.70	0.30	Word2Vec	EXT	1.00	0.67	0.33	Word2Vec	EXT	1.00	0.66	0.34
Pre-processing data 4					Pre-processing data 5					Pre-processing data 6					Pre-processing data 7				
emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff
BERT	SVM	0.66	0.70	-0.04	BERT	SVM	0.76	0.65	0.11	BERT	SVM	0.65	0.69	-0.04	**BERT**	**SVM**	**0.78**	**0.73**	**0.05**
BERT	RF	1.00	0.69	0.31	BERT	RF	1.00	0.70	0.30	BERT	RF	1.00	0.71	0.29	BERT	RF	1.00	0.72	0.28
BERT	LR	0.84	0.71	0.13	BERT	LR	0.97	0.67	0.30	BERT	LR	0.85	0.72	0.13	BERT	LR	0.98	0.72	0.26
BERT	KNN	0.79	0.69	0.10	BERT	KNN	0.83	0.68	0.15	BERT	KNN	0.80	0.63	0.17	BERT	KNN	0.84	0.68	0.16
BERT	EXT	1.00	0.71	0.29	BERT	EXT	1.00	0.70	0.30	BERT	EXT	1.00	0.71	0.29	BERT	EXT	1.00	0.70	0.30
Glove	SVM	0.52	0.56	-0.04	Glove	SVM	0.51	0.56	-0.05	Glove	SVM	0.52	0.56	-0.04	Glove	SVM	0.51	0.56	-0.05
Glove	RF	1.00	0.68	0.32	Glove	RF	1.00	0.64	0.36	Glove	RF	1.00	0.68	0.32	Glove	RF	1.00	0.63	0.37
Glove	LR	0.66	0.67	-0.01	Glove	LR	0.64	0.65	-0.01	Glove	LR	0.67	0.70	-0.03	Glove	LR	0.64	0.66	-0.02
Glove	KNN	0.81	0.58	0.23	Glove	KNN	0.78	0.60	0.18	Glove	KNN	0.80	0.62	0.18	Glove	KNN	0.79	0.63	0.16
Glove	EXT	1.00	0.68	0.32	Glove	EXT	1.00	0.68	0.32	Glove	EXT	1.00	0.65	0.35	Glove	EXT	1.00	0.64	0.36
TF-IDF	**SVM**	**0.93**	**0.81**	**0.12**	**TF-IDF**	**SVM**	**0.92**	**0.83**	**0.09**	**TF-IDF**	**SVM**	**0.93**	**0.83**	**0.10**	**TF-IDF**	**SVM**	**0.91**	**0.83**	**0.08**
TF-IDF	RF	1.00	0.84	0.16	TF-IDF	RF	1.00	0.82	0.18	TF-IDF	RF	1.00	0.83	0.17	TF-IDF	RF	1.00	0.81	0.19
TF-IDF	**LR**	**0.92**	**0.84**	**0.08**	**TF-IDF**	**LR**	**0.92**	**0.85**	**0.07**	**TF-IDF**	**LR**	**0.92**	**0.84**	**0.08**	**TF-IDF**	**LR**	**0.92**	**0.86**	**0.06**
TF-IDF	KNN	0.88	0.72	0.16	TF-IDF	KNN	0.88	0.71	0.17	TF-IDF	KNN	0.88	0.70	0.18	TF-IDF	KNN	0.88	0.70	0.18
TF-IDF	EXT	1.00	0.82	0.18	TF-IDF	EXT	1.00	0.82	0.18	TF-IDF	EXT	1.00	0.83	0.17	TF-IDF	EXT	1.00	0.83	0.17
Word2Vec	SVM	0.53	0.57	-0.04	Word2Vec	SVM	0.52	0.57	-0.05	Word2Vec	SVM	0.52	0.58	-0.06	Word2Vec	SVM	0.52	0.57	-0.05
Word2Vec	RF	1.00	0.67	0.33	Word2Vec	RF	1.00	0.65	0.35	Word2Vec	RF	1.00	0.67	0.33	Word2Vec	RF	1.00	0.65	0.35
Word2Vec	LR	0.62	0.69	-0.07	Word2Vec	LR	0.63	0.66	-0.03	Word2Vec	LR	0.62	0.69	-0.07	Word2Vec	LR	0.64	0.66	-0.02
Word2Vec	KNN	0.79	0.66	0.13	Word2Vec	KNN	0.78	0.63	0.15	Word2Vec	KNN	0.79	0.62	0.17	Word2Vec	KNN	0.78	0.61	0.17
Word2Vec	EXT	1.00	0.68	0.32	Word2Vec	EXT	1.00	0.67	0.33	Word2Vec	EXT	1.00	0.68	0.32	Word2Vec	EXT	1.00	0.68	0.32

Note: The cell(s) highlighted in yellow with red colored text indicate the best performing case(s). The cell(s) highlighted in orange indicate the second best performing case(s).

Table 23.

Results (Commitment, Study3- All data)

Pre-processing data 0				
emb	model	train	test	diff
BERT	SVM	0.61	0.57	0.04
BERT	RF	1.00	0.71	0.29
BERT	**LR**	**0.86**	**0.79**	**0.07**
BERT	KNN	0.81	0.68	0.13
BERT	EXT	1.00	0.72	0.28
Glove	SVM	0.58	0.54	0.04
Glove	RF	1.00	0.67	0.33
Glove	LR	0.67	0.63	0.04
Glove	KNN	0.80	0.61	0.19
Glove	EXT	1.00	0.66	0.34
TF-IDF	**SVM**	**0.91**	**0.81**	**0.10**
TF-IDF	RF	1.00	0.79	0.21
TF-IDF	**LR**	**0.91**	**0.82**	**0.09**
TF-IDF	KNN	0.89	0.74	0.15
TF-IDF	EXT	1.00	0.81	0.19
Word2Vec	SVM	0.58	0.54	0.04
Word2Vec	RF	1.00	0.67	0.33
Word2Vec	LR	0.59	0.56	0.03
Word2Vec	KNN	0.79	0.60	0.19
Word2Vec	EXT	1.00	0.68	0.32

Pre-processing data 1				
emb	model	train	test	diff
BERT	**SVM**	0.76	0.74	0.02
BERT	RF	1.00	0.72	0.28
BERT	**LR**	0.98	0.76	0.22
BERT	KNN	0.85	0.67	0.18
BERT	EXT	1.00	0.70	0.30
Glove	SVM	0.58	0.54	0.04
Glove	RF	1.00	0.67	0.33
Glove	LR	0.65	0.60	0.05
Glove	KNN	0.80	0.60	0.20
Glove	EXT	1.00	0.66	0.34
TF-IDF	**SVM**	**0.91**	**0.81**	**0.10**
TF-IDF	RF	1.00	0.80	0.20
TF-IDF	**LR**	**0.90**	**0.81**	**0.09**
TF-IDF	KNN	0.88	0.74	0.14
TF-IDF	EXT	1.00	0.81	0.19
Word2Vec	SVM	0.58	0.54	0.04
Word2Vec	RF	1.00	0.66	0.34
Word2Vec	LR	0.62	0.59	0.03
Word2Vec	KNN	0.79	0.64	0.15
Word2Vec	EXT	1.00	0.65	0.35

Pre-processing data 2				
emb	model	train	test	diff
BERT	SVM	0.60	0.56	0.04
BERT	RF	1.00	0.69	0.31
BERT	**LR**	**0.86**	**0.76**	**0.10**
BERT	KNN	0.83	0.69	0.14
BERT	EXT	1.00	0.68	0.32
Glove	SVM	0.58	0.54	0.04
Glove	RF	1.00	0.66	0.34
Glove	LR	0.67	0.61	0.06
Glove	KNN	0.78	0.59	0.19
Glove	EXT	1.00	0.67	0.33
TF-IDF	**SVM**	**0.91**	**0.83**	**0.08**
TF-IDF	RF	1.00	0.80	0.20
TF-IDF	**LR**	**0.91**	**0.83**	**0.08**
TF-IDF	KNN	0.90	0.76	0.14
TF-IDF	EXT	1.00	0.79	0.21
Word2Vec	SVM	0.58	0.56	0.02
Word2Vec	RF	1.00	0.66	0.34
Word2Vec	LR	0.60	0.57	0.03
Word2Vec	KNN	0.80	0.61	0.19
Word2Vec	EXT	1.00	0.67	0.33

Pre-processing data 3				
emb	model	train	test	diff
BERT	**SVM**	**0.77**	**0.73**	**0.04**
BERT	RF	1.00	0.69	0.31
BERT	**LR**	0.98	0.75	0.23
BERT	KNN	0.84	0.66	0.18
BERT	EXT	1.00	0.69	0.31
Glove	SVM	0.58	0.54	0.04
Glove	RF	1.00	0.64	0.36
Glove	LR	0.66	0.60	0.06
Glove	KNN	0.81	0.61	0.20
Glove	EXT	1.00	0.64	0.36
TF-IDF	**SVM**	**0.91**	**0.82**	**0.09**
TF-IDF	RF	1.00	0.79	0.21
TF-IDF	**LR**	**0.90**	**0.82**	**0.08**
TF-IDF	KNN	0.88	0.73	0.15
TF-IDF	EXT	1.00	0.80	0.20
Word2Vec	SVM	0.58	0.54	0.04
Word2Vec	RF	1.00	0.65	0.35
Word2Vec	LR	0.62	0.56	0.06
Word2Vec	KNN	0.80	0.61	0.19
Word2Vec	EXT	1.00	0.66	0.34

Pre-processing data 4				
emb	model	train	test	diff
BERT	SVM	0.60	0.57	0.03
BERT	RF	1.00	0.70	0.30
BERT	**LR**	**0.86**	**0.78**	**0.08**
BERT	KNN	0.81	0.69	0.12
BERT	EXT	1.00	0.72	0.28
Glove	SVM	0.58	0.54	0.04
Glove	RF	1.00	0.67	0.33
Glove	LR	0.67	0.62	0.05
Glove	KNN	0.79	0.61	0.18
Glove	EXT	1.00	0.68	0.32
TF-IDF	**SVM**	**0.92**	**0.83**	**0.09**
TF-IDF	RF	1.00	0.80	0.20
TF-IDF	**LR**	**0.90**	**0.82**	**0.08**
TF-IDF	KNN	0.89	0.76	0.13
TF-IDF	EXT	1.00	0.80	0.20
Word2Vec	SVM	0.58	0.56	0.02
Word2Vec	RF	1.00	0.69	0.31
Word2Vec	LR	0.59	0.56	0.03
Word2Vec	KNN	0.79	0.60	0.19
Word2Vec	EXT	1.00	0.66	0.34

Pre-processing data 5				
emb	model	train	test	diff
BERT	**SVM**	**0.76**	**0.74**	**0.02**
BERT	RF	1.00	0.73	0.27
BERT	**LR**	0.98	0.76	0.22
BERT	KNN	0.85	0.69	0.16
BERT	EXT	1.00	0.72	0.28
Glove	SVM	0.58	0.54	0.04
Glove	RF	1.00	0.67	0.33
Glove	LR	0.66	0.60	0.06
Glove	KNN	0.81	0.60	0.21
Glove	EXT	1.00	0.67	0.33
TF-IDF	**SVM**	**0.92**	**0.82**	**0.10**
TF-IDF	RF	1.00	0.82	0.18
TF-IDF	**LR**	**0.91**	**0.82**	**0.09**
TF-IDF	KNN	0.88	0.76	0.12
TF-IDF	EXT	1.00	0.81	0.19
Word2Vec	SVM	0.58	0.54	0.04
Word2Vec	RF	1.00	0.67	0.33
Word2Vec	LR	0.63	0.59	0.04
Word2Vec	KNN	0.80	0.62	0.18
Word2Vec	EXT	1.00	0.66	0.34

Pre-processing data 6				
emb	model	train	test	diff
BERT	SVM	0.61	0.57	0.04
BERT	RF	1.00	0.73	0.27
BERT	**LR**	**0.86**	**0.76**	**0.10**
BERT	KNN	0.82	0.68	0.14
BERT	EXT	1.00	0.72	0.28
Glove	SVM	0.58	0.54	0.04
Glove	RF	1.00	0.70	0.30
Glove	LR	0.67	0.61	0.06
Glove	KNN	0.78	0.63	0.15
Glove	EXT	1.00	0.67	0.33
TF-IDF	**SVM**	**0.92**	**0.83**	**0.09**
TF-IDF	RF	1.00	0.80	0.20
TF-IDF	**LR**	**0.90**	**0.82**	**0.08**
TF-IDF	KNN	0.89	0.78	0.11
TF-IDF	EXT	1.00	0.80	0.20
Word2Vec	SVM	0.58	0.56	0.02
Word2Vec	RF	1.00	0.67	0.33
Word2Vec	LR	0.60	0.58	0.02
Word2Vec	KNN	0.80	0.62	0.18
Word2Vec	EXT	1.00	0.67	0.33

Pre-processing data 7				
emb	model	train	test	diff
BERT	**SVM**	**0.77**	**0.74**	**0.03**
BERT	RF	1.00	0.70	0.30
BERT	**LR**	0.98	0.76	0.22
BERT	KNN	0.85	0.68	0.17
BERT	EXT	1.00	0.71	0.29
Glove	SVM	0.58	0.54	0.04
Glove	RF	1.00	0.65	0.35
Glove	LR	0.66	0.59	0.07
Glove	KNN	0.81	0.60	0.21
Glove	EXT	1.00	0.67	0.33
TF-IDF	**SVM**	**0.91**	**0.81**	**0.10**
TF-IDF	RF	1.00	0.80	0.20
TF-IDF	**LR**	**0.91**	**0.82**	**0.09**
TF-IDF	KNN	0.88	0.74	0.14
TF-IDF	EXT	1.00	0.81	0.19
Word2Vec	SVM	0.58	0.54	0.04
Word2Vec	RF	1.00	0.66	0.34
Word2Vec	LR	0.63	0.58	0.05
Word2Vec	KNN	0.81	0.61	0.20
Word2Vec	EXT	1.00	0.66	0.34

Note: The cell(s) highlighted in yellow with red colored text indicate the best performing case(s). The cell(s) highlighted in orange indicate the second best performing case(s).

Table 24.

Results (Satisfaction, Study3- All data)

Pre-processing data 0					Pre-processing data 1					Pre-processing data 2					Pre-processing data 3				
emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff
BERT	SVM	0.69	0.67	0.02	**BERT**	**SVM**	**0.76**	**0.75**	**0.01**	BERT	SVM	0.69	0.67	0.02	**BERT**	**SVM**	**0.77**	**0.74**	**0.03**
BERT	RF	1.00	0.70	0.30	BERT	RF	1.00	0.73	0.27	BERT	RF	1.00	0.69	0.31	BERT	RF	1.00	0.74	0.26
BERT	LR	0.88	0.76	0.12	BERT	LR	0.99	0.79	0.20	BERT	LR	0.88	0.75	0.13	BERT	LR	0.99	0.75	0.24
BERT	KNN	0.84	0.67	0.17	BERT	KNN	0.85	0.71	0.14	BERT	KNN	0.84	0.67	0.17	BERT	KNN	0.84	0.72	0.12
BERT	EXT	1.00	0.73	0.27	BERT	EXT	1.00	0.74	0.26	BERT	EXT	1.00	0.73	0.27	BERT	EXT	1.00	0.73	0.27
Glove	SVM	0.69	0.67	0.02	Glove	SVM	0.68	0.67	0.01	Glove	SVM	0.69	0.67	0.02	Glove	SVM	0.69	0.67	0.02
Glove	RF	1.00	0.71	0.29	Glove	RF	1.00	0.71	0.29	Glove	RF	1.00	0.71	0.29	Glove	RF	1.00	0.70	0.30
Glove	LR	0.71	0.68	0.03	Glove	LR	0.71	0.69	0.02	Glove	LR	0.70	0.68	0.02	Glove	LR	0.70	0.68	0.02
Glove	KNN	0.82	0.66	0.16	Glove	KNN	0.83	0.60	0.23	Glove	KNN	0.81	0.64	0.17	Glove	KNN	0.82	0.65	0.17
Glove	EXT	1.00	0.70	0.30	Glove	EXT	1.00	0.70	0.30	Glove	EXT	1.00	0.69	0.31	Glove	EXT	1.00	0.69	0.31
TF-IDF	**SVM**	**0.93**	**0.81**	**0.12**	**TF-IDF**	**SVM**	**0.91**	**0.83**	**0.08**	**TF-IDF**	**SVM**	**0.93**	**0.83**	**0.10**	**TF-IDF**	**SVM**	**0.92**	**0.82**	**0.10**
TF-IDF	RF	1.00	0.82	0.18	TF-IDF	RF	1.00	0.80	0.20	**TF-IDF**	**RF**	1.00	0.81	0.19	**TF-IDF**	**RF**	1.00	0.82	0.18
TF-IDF	**LR**	**0.89**	**0.81**	**0.08**	**TF-IDF**	**LR**	**0.89**	**0.81**	**0.08**	**TF-IDF**	**LR**	**0.89**	**0.80**	**0.09**	**TF-IDF**	**LR**	**0.89**	**0.81**	**0.08**
TF-IDF	KNN	0.89	0.77	0.12	TF-IDF	KNN	0.90	0.75	0.15	TF-IDF	KNN	0.90	0.78	0.12	TF-IDF	KNN	0.90	0.76	0.14
TF-IDF	EXT	1.00	0.81	0.19	TF-IDF	EXT	1.00	0.82	0.18	TF-IDF	EXT	1.00	0.80	0.20	TF-IDF	EXT	1.00	0.82	0.18
Word2Vec	SVM	0.65	0.65	0.00	Word2Vec	SVM	0.68	0.68	0.00	Word2Vec	SVM	0.66	0.66	0.00	Word2Vec	SVM	0.68	0.68	0.00
Word2Vec	RF	1.00	0.69	0.31	Word2Vec	RF	1.00	0.69	0.31	Word2Vec	RF	1.00	0.67	0.33	Word2Vec	RF	1.00	0.67	0.33
Word2Vec	LR	0.69	0.68	0.01	Word2Vec	LR	0.69	0.68	0.01	Word2Vec	LR	0.69	0.68	0.01	Word2Vec	LR	0.69	0.68	0.01
Word2Vec	KNN	0.81	0.62	0.19	Word2Vec	KNN	0.82	0.64	0.18	Word2Vec	KNN	0.81	0.61	0.20	Word2Vec	KNN	0.82	0.63	0.19
Word2Vec	EXT	1.00	0.66	0.34	Word2Vec	EXT	1.00	0.68	0.32	Word2Vec	EXT	1.00	0.68	0.32	Word2Vec	EXT	1.00	0.66	0.34
Pre-processing data 4					Pre-processing data 5					Pre-processing data 6					Pre-processing data 7				
emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff
BERT	SVM	0.69	0.67	0.02	**BERT**	**SVM**	**0.75**	**0.75**	**0.00**	BERT	SVM	0.69	0.67	0.02	**BERT**	**SVM**	**0.78**	**0.74**	**0.04**
BERT	RF	1.00	0.72	0.28	BERT	RF	1.00	0.71	0.29	BERT	RF	1.00	0.71	0.29	BERT	RF	1.00	0.74	0.26
BERT	LR	0.88	0.76	0.12	BERT	LR	0.99	0.76	0.23	BERT	LR	0.89	0.76	0.13	BERT	LR	0.99	0.76	0.23
BERT	KNN	0.83	0.64	0.19	BERT	KNN	0.85	0.71	0.14	BERT	KNN	0.86	0.64	0.22	BERT	KNN	0.84	0.69	0.15
BERT	EXT	1.00	0.72	0.28	BERT	EXT	1.00	0.72	0.28	BERT	EXT	1.00	0.71	0.29	BERT	EXT	1.00	0.74	0.26
Glove	SVM	0.69	0.67	0.02	Glove	SVM	0.68	0.67	0.01	Glove	SVM	0.69	0.67	0.02	Glove	SVM	0.69	0.67	0.02
Glove	RF	1.00	0.69	0.31	Glove	RF	1.00	0.70	0.30	Glove	RF	1.00	0.74	0.26	Glove	RF	1.00	0.71	0.29
Glove	LR	0.70	0.68	0.02	Glove	LR	0.71	0.69	0.02	Glove	LR	0.70	0.68	0.02	Glove	LR	0.70	0.68	0.02
Glove	KNN	0.82	0.66	0.16	Glove	KNN	0.82	0.61	0.21	Glove	KNN	0.80	0.67	0.13	Glove	KNN	0.81	0.65	0.16
Glove	EXT	1.00	0.72	0.28	Glove	EXT	1.00	0.71	0.29	Glove	EXT	1.00	0.72	0.28	Glove	EXT	1.00	0.68	0.32
TF-IDF	**SVM**	**0.93**	**0.81**	**0.12**	**TF-IDF**	**SVM**	**0.91**	**0.82**	**0.09**	**TF-IDF**	**SVM**	**0.93**	**0.81**	**0.12**	**TF-IDF**	**SVM**	**0.91**	**0.83**	**0.08**
TF-IDF	RF	1.00	0.80	0.20	TF-IDF	RF	1.00	0.84	0.16	TF-IDF	RF	1.00	0.82	0.18	TF-IDF	RF	1.00	0.82	0.18
TF-IDF	**LR**	**0.88**	**0.79**	**0.09**	**TF-IDF**	**LR**	**0.88**	**0.80**	**0.08**	**TF-IDF**	**LR**	**0.88**	**0.80**	**0.08**	**TF-IDF**	**LR**	**0.88**	**0.80**	**0.08**
TF-IDF	KNN	0.90	0.77	0.13	TF-IDF	KNN	0.90	0.76	0.14	TF-IDF	KNN	0.90	0.78	0.12	TF-IDF	KNN	0.91	0.77	0.14
TF-IDF	EXT	1.00	0.84	0.16	TF-IDF	EXT	1.00	0.85	0.15	TF-IDF	EXT	1.00	0.81	0.19	TF-IDF	EXT	1.00	0.83	0.17
Word2Vec	SVM	0.65	0.66	-0.01	Word2Vec	SVM	0.68	0.68	0.00	Word2Vec	SVM	0.66	0.66	0.00	Word2Vec	SVM	0.68	0.68	0.00
Word2Vec	RF	1.00	0.68	0.32	Word2Vec	RF	1.00	0.69	0.31	Word2Vec	RF	1.00	0.68	0.32	Word2Vec	RF	1.00	0.69	0.31
Word2Vec	LR	0.69	0.68	0.01	Word2Vec	LR	0.69	0.68	0.01	Word2Vec	LR	0.69	0.67	0.02	Word2Vec	LR	0.69	0.68	0.01
Word2Vec	KNN	0.81	0.62	0.19	Word2Vec	KNN	0.82	0.62	0.20	Word2Vec	KNN	0.80	0.63	0.17	Word2Vec	KNN	0.81	0.63	0.18
Word2Vec	EXT	1.00	0.67	0.33	Word2Vec	EXT	1.00	0.68	0.32	Word2Vec	EXT	1.00	0.69	0.31	Word2Vec	EXT	1.00	0.69	0.31

Note: The cell(s) highlighted in yellow with red colored text indicate the best performing case(s). The cell(s) highlighted in orange indicate the second best performing case(s).

Table 25.

Results (Trust, Study3- All data)

Pre-processing data 0					Pre-processing data 1					Pre-processing data 2					Pre-processing data 3				
emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff
BERT	SVM	0.68	0.66	0.02	**BERT**	**SVM**	**0.74**	**0.70**	**0.04**	BERT	SVM	0.68	0.66	0.02	**BERT**	**SVM**	**0.73**	**0.68**	**0.05**
BERT	RF	1.00	0.69	0.31	BERT	RF	1.00	0.68	0.32	BERT	RF	1.00	0.68	0.32	BERT	RF	1.00	0.68	0.32
BERT	LR	0.85	0.73	0.12	BERT	LR	0.98	0.73	0.25	BERT	LR	0.86	0.73	0.13	BERT	LR	0.97	0.74	0.23
BERT	KNN	0.82	0.65	0.17	BERT	KNN	0.85	0.72	0.13	BERT	KNN	0.84	0.66	0.18	BERT	KNN	0.85	0.70	0.15
BERT	EXT	1.00	0.69	0.31	BERT	EXT	1.00	0.70	0.30	BERT	EXT	1.00	0.67	0.33	BERT	EXT	1.00	0.71	0.29
Glove	SVM	0.68	0.66	0.02	Glove	SVM	0.68	0.66	0.02	Glove	SVM	0.68	0.66	0.02	Glove	SVM	0.68	0.66	0.02
Glove	RF	1.00	0.68	0.32	Glove	RF	1.00	0.70	0.30	Glove	RF	1.00	0.70	0.30	Glove	RF	1.00	0.72	0.28
Glove	LR	0.69	0.68	0.01	**Glove**	**LR**	**0.70**	**0.70**	**0.00**	Glove	LR	0.68	0.67	0.01	Glove	LR	0.69	0.68	0.01
Glove	KNN	0.81	0.67	0.14	Glove	KNN	0.79	0.65	0.14	Glove	KNN	0.81	0.64	0.17	Glove	KNN	0.80	0.63	0.17
Glove	EXT	1.00	0.72	0.28	Glove	EXT	1.00	0.71	0.29	Glove	EXT	1.00	0.72	0.28	Glove	EXT	1.00	0.71	0.29
TF-IDF	**SVM**	**0.92**	**0.78**	**0.14**	**TF-IDF**	**SVM**	**0.91**	**0.78**	**0.13**	**TF-IDF**	**SVM**	**0.92**	**0.78**	**0.14**	**TF-IDF**	**SVM**	**0.91**	**0.77**	**0.14**
TF-IDF	RF	1.00	0.74	0.26	TF-IDF	RF	1.00	0.75	0.25	TF-IDF	RF	1.00	0.75	0.25	TF-IDF	RF	1.00	0.75	0.25
TF-IDF	**LR**	**0.88**	**0.78**	**0.10**	**TF-IDF**	**LR**	**0.88**	**0.78**	**0.10**	**TF-IDF**	**LR**	**0.87**	**0.77**	**0.10**	**TF-IDF**	**LR**	**0.88**	**0.78**	**0.10**
TF-IDF	KNN	0.87	0.75	0.12	TF-IDF	KNN	0.87	0.74	0.13	TF-IDF	KNN	0.87	0.75	0.12	TF-IDF	KNN	0.87	0.74	0.13
TF-IDF	EXT	1.00	0.76	0.24	TF-IDF	EXT	1.00	0.77	0.23	TF-IDF	EXT	1.00	0.78	0.22	TF-IDF	EXT	1.00	0.77	0.23
Word2Vec	SVM	0.64	0.63	0.01	Word2Vec	SVM	0.68	0.66	0.02	Word2Vec	SVM	0.66	0.64	0.02	Word2Vec	SVM	0.68	0.66	0.02
Word2Vec	RF	1.00	0.69	0.31	Word2Vec	RF	1.00	0.70	0.30	Word2Vec	RF	1.00	0.67	0.33	Word2Vec	RF	1.00	0.67	0.33
Word2Vec	LR	0.68	0.66	0.02	Word2Vec	LR	0.68	0.67	0.01	Word2Vec	LR	0.68	0.66	0.02	Word2Vec	LR	0.68	0.67	0.01
Word2Vec	KNN	0.81	0.64	0.17	Word2Vec	KNN	0.81	0.67	0.14	Word2Vec	KNN	0.80	0.61	0.19	Word2Vec	KNN	0.81	0.63	0.18
Word2Vec	EXT	1.00	0.68	0.32	Word2Vec	EXT	1.00	0.70	0.30	Word2Vec	EXT	1.00	0.68	0.32	Word2Vec	EXT	1.00	0.70	0.30
Pre-processing data 4					Pre-processing data 5					Pre-processing data 6					Pre-processing data 7				
emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff	emb	model	train	test	diff
BERT	SVM	0.68	0.66	0.02	**BERT**	**SVM**	**0.74**	**0.71**	**0.03**	BERT	SVM	0.68	0.66	0.02	**BERT**	**SVM**	**0.75**	**0.70**	**0.05**
BERT	RF	1.00	0.69	0.31	BERT	RF	1.00	0.71	0.29	BERT	RF	1.00	0.66	0.34	BERT	RF	1.00	0.70	0.30
BERT	LR	0.85	0.72	0.13	BERT	LR	0.97	0.73	0.24	BERT	LR	0.87	0.76	0.11	BERT	LR	0.98	0.70	0.28
BERT	KNN	0.81	0.68	0.13	BERT	KNN	0.85	0.68	0.17	BERT	KNN	0.83	0.66	0.17	BERT	KNN	0.86	0.70	0.16
BERT	EXT	1.00	0.72	0.28	BERT	EXT	1.00	0.69	0.31	BERT	EXT	1.00	0.68	0.32	BERT	EXT	1.00	0.70	0.30
Glove	SVM	0.68	0.66	0.02	Glove	SVM	0.68	0.66	0.02	Glove	SVM	0.68	0.66	0.02	Glove	SVM	0.68	0.66	0.02
Glove	RF	1.00	0.69	0.31	Glove	RF	1.00	0.71	0.29	Glove	RF	1.00	0.72	0.28	Glove	RF	1.00	0.70	0.30
Glove	LR	0.69	0.67	0.02	Glove	LR	0.70	0.69	0.01	Glove	LR	0.68	0.67	0.01	Glove	LR	0.69	0.69	0.00
Glove	KNN	0.80	0.65	0.15	Glove	KNN	0.79	0.66	0.13	Glove	KNN	0.81	0.67	0.14	Glove	KNN	0.81	0.64	0.17
Glove	EXT	1.00	0.71	0.29	Glove	EXT	1.00	0.71	0.29	Glove	EXT	1.00	0.69	0.31	Glove	EXT	1.00	0.70	0.30
TF-IDF	**SVM**	**0.92**	**0.80**	**0.12**	**TF-IDF**	**SVM**	**0.91**	**0.80**	**0.11**	**TF-IDF**	**SVM**	**0.92**	**0.80**	**0.12**	**TF-IDF**	**SVM**	**0.91**	**0.81**	**0.10**
TF-IDF	RF	1.00	0.75	0.25	TF-IDF	RF	1.00	0.75	0.25	TF-IDF	RF	1.00	0.74	0.26	TF-IDF	RF	1.00	0.76	0.24
TF-IDF	**LR**	**0.87**	**0.79**	**0.08**	**TF-IDF**	**LR**	**0.87**	**0.78**	**0.09**	**TF-IDF**	**LR**	**0.87**	**0.78**	**0.09**	**TF-IDF**	**LR**	**0.87**	**0.78**	**0.09**
TF-IDF	KNN	0.87	0.75	0.12	TF-IDF	KNN	0.87	0.75	0.12	TF-IDF	KNN	0.87	0.75	0.12	TF-IDF	KNN	0.87	0.75	0.12
TF-IDF	EXT	1.00	0.76	0.24	TF-IDF	EXT	1.00	0.76	0.24	TF-IDF	EXT	1.00	0.78	0.22	TF-IDF	EXT	1.00	0.79	0.21
Word2Vec	SVM	0.65	0.64	0.01	Word2Vec	SVM	0.68	0.66	0.02	Word2Vec	SVM	0.66	0.64	0.02	Word2Vec	SVM	0.68	0.66	0.02
Word2Vec	RF	1.00	0.69	0.31	Word2Vec	RF	1.00	0.69	0.31	Word2Vec	RF	1.00	0.68	0.32	Word2Vec	RF	1.00	0.70	0.30
Word2Vec	LR	0.68	0.66	0.02	Word2Vec	LR	0.68	0.67	0.01	Word2Vec	LR	0.68	0.66	0.02	Word2Vec	LR	0.68	0.67	0.01
Word2Vec	KNN	0.81	0.64	0.17	Word2Vec	KNN	0.81	0.66	0.15	Word2Vec	KNN	0.80	0.64	0.16	Word2Vec	KNN	0.81	0.65	0.16
Word2Vec	EXT	1.00	0.71	0.29	Word2Vec	EXT	1.00	0.69	0.31	Word2Vec	EXT	1.00	0.69	0.31	Word2Vec	EXT	1.00	0.73	0.27

Note: The cell(s) highlighted in yellow with red colored text indicate the best performing case(s). The cell(s) highlighted in orange indicate the second best performing case(s).

Table 26.
A summary of all data results (Study 3)

Control Mutuality

Pre-processing data 3				
emb	model	train	test	diff
TF-IDF	**LR**	**0.92**	**0.86**	**0.06**
Pre-processing data 7				
TF-IDF	**LR**	**0.92**	**0.86**	**0.06**

Commitment

Pre-processing data 2				
emb	model	train	test	diff
TF-IDF	**SVM**	**0.91**	**0.83**	**0.08**
TF-IDF	**LR**	**0.91**	**0.83**	**0.08**
Pre-processing data 6				
TF-IDF	**SVM**	**0.92**	**0.83**	**0.09**
Pre-processing data 4				
TF-IDF	**SVM**	**0.92**	**0.83**	**0.09**

Satisfaction

Pre-processing data 1				
emb	model	train	test	diff
TF-IDF	**SVM**	**0.91**	**0.83**	**0.08**
Pre-processing data 7				
TF-IDF	**SVM**	**0.91**	**0.83**	**0.08**

Trust

Pre-processing data 7				
emb	model	train	test	diff
TF-IDF	**SVM**	**0.91**	**0.81**	**0.10**

Table 27.
Results summary of text-embedding techniques

Target Audience	Study 1: Customer and Consumers			Study 2: Employees	Study 3: All (Study 1 + Study 2)
Method	Crawled	Survey	Crawled + Survey	Crawled + Survey	Crawled + Survey
Input	Human Coding (Multiple coders)	People answer	Human coding + People Answer	Human coding (single coder) + People Answer	Human coding + People Answer
N	247	996~1,201	1,213~1,450	211	1,424~1,758
Control Mutuality	TF-IDF	TF-IDF	TF-IDF	Word2Vec	TF-IDF
Commitment	TF-IDF	TF-IDF	TF-IDF	TF-IDF	TF-IDF
Satisfaction	BERT	TF-IDF	TF-IDF	TF-IDF/BERT	TF-IDF
Trust	BERT	TF-IDF	TF-IDF	TF-IDF	TF-IDF

Table 28.
Results summary of classifications

Target Audience	**Study 1: Customer and Consumers**			**Study 2: Employees**	**Study 3: All (Study 1 + Study 2)**
Method	Crawled	Survey	Crawled + Survey	Crawled + Survey	Crawled + Survey
Input	Human Coding (Multiple coders)	People answer	Human coding + People Answer	Human coding (single coder) + People Answer	Human coding + People Answer
N	247	996~1,201	1,213~1,450	211	1,424~1,758
Control Mutuality	KNN	SVM	LR	KNN	LR
Commitment	KNN	SVM	LR	KNN	SVM/LR
Satisfaction	KNN	SVM	SVM	KNN	SVM
Trust	KNN	SVM	SVM	SVM	SVM

Table 29.
Ratio of Study 1 to Study 3

Control Mutuality					
	Study 1 + Data 1	Study 1 + Data 2	Study 1 + Data 3	Study 2	Study 3
Proportion of ones	12.10%	61.00%	51.80%	55.90%	52.40%
Proportion of zeros	87.90%	39.00%	48.20%	44.10%	47.60%
Commitment					
	Study 1 + Data 1	Study 1 + Data 2	Study 1 + Data 3	Study 2	Study 3
Proportion of ones	16.20%	66.50%	56.90%	61.10%	61.10%
Proportion of zeros	83.80%	33.50%	43.10%	38.90%	38.90%
Satisfaction					
	Study 1 + Data 1	Study 1 + Data 2	Study 1 + Data 3	Study 2	Study 3
Proportion of ones	49.4%	68.40%	64.60%	90.00%	68.30%
Proportion of zeros	50.6%	31.60%	35.40%	10.00%	31.70%
Trust					
	Study 1 + Data 1	Study 1 + Data 2	Study 1 + Data 3	Study 2	Study 3
Proportion of ones	36.80%	71.10%	65.30%	71.10%	67.50%
Proportion of zeros	63.20%	28.90%	34.70%	28.90%	32.50%

Table 30.

Case study results (BERT-SVM, Customer data: TripAdvisor reviews)

Note. The red colored numbers indicate wrong identifications.

text	control mutuality	commitment	trust	satisfaction
friendly teenage son stayed hotel casci september, experience confirmed excellent reviews trip advisor based decision book hotel, staff friendly helpful especially son accidently left treasured nba singlet room, quick phone ascertained handed housekeeper posted australia cost, stopover florence days later enroute venice allowed quick taxi ride train station pick personally, hotel casci excellent central position room impecibly clean quiet faced street breakfast fine great coffee offers hot chocolate son, staff gallery bookings recommendations tours, economical hotel choice iwith real personal touch not hesitate recommend,	1	1	1	1
enjoyable stay enjoyable stay hotel convenient location wonderful room perfectly soundproof considering work renovations close wonderful breakfast, got room average 200 euros includ, bkfast little bit expensive end happy splurged having seen pictures hotels similar category, u want really special advise u splurge bellavista room floor big balcony wonderful panoramic views florence sunlit rainshower bathroom, think 315 euros, fault hotel staff courteous professional not friendly, lots smiles hotel experience.without hesitation stay, note hotel renovating adding rooms good noise insolation outside did ask work not started till 9am, sensitive u ask room away work renvoation.ps u love shopping brand names ask tour mall prada outlet worth especially shoes,	1	0	0	1
great hotel business travellers couples travel beijing stayed hotels years great, attentive service clean comfortable rooms friendly staff good set restaurants, lots really good restaurants close local flavor especially hot pots peking duck, concierge knowledgeable helpful, terrific beds pillows great bath robes, good value money,	1	1	1	1

beautiful hotel destroyed terrible service, stay started problems desk, 20 minutes 3:00. told check time 3:00 lunch come 3:00. returned lunch given room told luggage room, no concierge took room took 20 minutes room located far away, course luggage not, went desk told did not want room, said okay not room 6:00. sat lobby travel clothes waited room, concierage told tell hotel 9:00 morning, n't tell 2 hours sat lobby, wanted try sell time share, got introduction hotel, 6:30 finally got new room luggage, day totally wasted, turned room utility closet did not sleep night heard pipes night, went lobby morning asked room changed, said n't hotel 100 booked no rooms, hotel actually 25-30 booked, finally gave room concierge did time showed room, room 3rd floor night door roof blowing opening closing, went lobby complain gave 100 booked speech, head guest services come not believe bad, called maintaince 15 minutes door closed, end problem, massages spa droped 250.00, massages good, day took massage workshop zen pavillian, told spa use pool steam room sauna, no showed did wrong, fabulous cabanas outside massage single couple, no offered instead massages upstairs listened drilling doing downstairs construction spa, bad did not offer beautiful cabanas, food average- enjoyed japanese asian best, outdoor restaurants did not open rained nanos did not open day- told hotel not 2 outdoor buffets open, hmmmm not change room 100 booked restaurants did not open did not people, not impressed market grill food laid day 11:00 6:00. atmosphere bummer met lobby restaurants pool nothing similar complaints, shame hotel beautiful grounds gardens really nice, service terrible food just fair rooms musty nightly entertainment weak, overall recommend resort return, travelled 30 years great places, real disappointment felt like waste money, way did tip including real nice tip desk arrive did not help, trip bummer,	0	1	0	1

sun sand smiles people, want 5 star pay 4 star resort, stayed hotel weeks april 12 april 26. flew air transit newfoundland canada, reading review wondering resort like.it beautful place visit.the problem resorts people stay, remember going country golden rule, want you.respect hotel..the rooms clean, bed comfortable clean, seen bug hotel, love clean room sheets got more.the cleaning lady cleaned room nice thanked day dollar inportant me.you leave gifts pillow not, remember people 30 t0 40 dollars weeks work hard heart.. look evey day say hola country respect food.. plenty food eat not hungry, book 3 different restaurants check, reviews people said got sick food, maybee 12 people went didnt, fussy come food safety.i eat hot food hot cold food cold, problem, wait staff dining room best.for 2 weeks cleaned table got need drink meal problem, canada ate restaurant tip 5 10 dollars time eat depending cost meal.i tipped 1 dollar meal beaches best place feet, travel alot happy here.you walk direct meet lots people walk direct meet, easy beach, quite place lie sun tan places booze inclued people noisey drunk, room night time12 1..2..3..or 4..am good time voice noise carries people sleeping bar..the drinks really good sandys bar..say hello sandy elizabeth rest bar staff, lot people staying hotel wants served first.. holidays relex mix best drinks, best vacations took time enjoy vacation people helped make happen dominican people staff hotel best respect little enjoy holidayall couples respected here..very gay friendly,	0	1	1	1
fantastic stay pulitzer wife recently stayed pulitzer nights stop honeymoon plesantly surprised especially reading mixed reviews ta, hotel located great town block tram stop anne frank house, upgraded large beautiful canal view room, bed fantastic especially 10 hour flight la, staff extremely helpful professional, used starwood points room ca n't comment value relative hotels area loved stay unique comfortable property return,	1	1	1	1
great value location stayed separate times week march flights burbank airport, hotel convienent airport 5 miles away local sights visiting hollywood area universal warner brother studios hollywood bowl la brea tar pits, breakfast free parking plus, bathroom cleaned bit better visit rest place good, visits manager left basket goodies cookies candy water bottles greatly appreciated long travels lots la traffic jams, faced freeway times did n't traffic noise intrusive, pool small not reason, visited hilton property street friends staying fancy/nice attractive pool scene considering mandatory parking no breakfast provided- holiday inn express better value site seeing not planning alot time hang hotel,	0	0	0	0

lovely hotel lovely area husband stayed nights honeymoon wonderful experience, room small appointed clean, breakfast courtyard delightful, walter sandro helpful making recommendations restaurants called reserve table popular restaurant, neighborhood major tourist areas comfortable walking distance sights, bars restaurants area wonderful local places i.e, no tourist menus highly recommend,	1	1	1	1
fantastic hotel fantastic country, stayed 3 days singapore route bali honeymoon wish stayed longer, hotel lovely staff helpul pool area great not second feel middle city, stayed garden wing room good size balcony over-looking pool area, ate restaurant blu nights great bit pricey worth, really nice bar area champange cocktails views bar restaurant amazing, breakfast line really not fault place choice different types food seen believed buffet style turnaround food quick nothing gets cold, thought eat curry breakfast, singapore.. say definitely hotel stay doubt, p.s, area singapore called clarke quay really cool bars/restaurants/clubs defintely worth visit, did n't place night bit shame,	1	1	1	1
great location hotel nice suprise, rooms bathrooms clean tastefully decorated decent size bidet not older tired look expected, took public bus airport stop pizzale roma minute walk bridges little odd suitcases lots people doing no ramps bridges deal stairs cheap commute 3gbp unlike believe water shuttle, conveniently located close stores restaurants walk st. marks square, hotel arrange complimentary water shuttle glass factory no sales pressure watch glass blowing leave like, breakfast morning sitting outside canal lovely self serve continental breakfast adequate, opposite hotel beautiful church interior impressive 2 mins check, outside hotel gondala, allowed drink wine bottles seating area canal instead bar service, staff wonderful helpful, not hesitate stay,	1	1	1	1
loved secrets boyfriend went apple vacations secrets excellence punta cana, absolutely beautiful, left reading alot reviews kindof nervous, not sure people giving resort bad reviews, people friendly little spanish helps, simple things like hello good night thank, wonderful time grounds absolutely beautiful food drinks good, went excursions outside resort, highly recommend outback safari tour guides great learn dominic, highly recommend resort,	1	1	0	1

location great hotel n't, went away barcelona week partner, impressions hotel looks good, bus airport drops road realised saw past walked 15mins booked double room got twin room, went reception ask moved surprise birthday trip partner said busy n't, partner fine ok.staff ok mornings cleaner just walk rooms knocking got annoying 3 days 8-9a.m, mentioned reception bothered again.the hotel noisy light sleeper n't stay luckily read previous reviews took ear plugs didnt help.breakfast not great dishes picked food stuck not impressed, happen week there.location good, prefer not stay city slightly hotel convienant, partner decided barcelona stay little bit closer city night times make sure didnt metro late area hotel did bit dead.we loved barcelona sadly hotel not, def, barcelona soon felt week not,	0	0	0	0
great budget hotel hotel layne excellent budget hotel highly recommend, close heart san francisco costs little, run hospitable family patels possible make guests welcome, rooms clean comfortable far superior hotels higher price brackets.stay,	1	1	1	1
march 23-30 great vacation great time despite worries resort, went idea just relazing not worrying minor things, days left resort major, did change things work, got little tired temp, buffet set beach end walking punta cana princess lunch, highly recommend want change lunch menu.go platinum, say platinum juan v. know, great guy, saying told website know reading, hey juan thanks presidente, took advantage platinum ordered room service times, especially breakfast n't feel like getting going breakfast, breakfast patio awesome.there huge langauge barrier country, did notice staff especially friendly spoke good english, gals went did speak spanish extremely helpful, highly suggest pick basics spanish.food good complaint warmer, did n't just food.beach..one word awesome, jamaica twice just does not compare, really feel like walking powder, nice walk ocean not step soft sand, ocean quite bit rougher used riot play, definately, having restaurants bars beach big difference just, say probably not bring young children resort, resorts area cater alot younger kids, good teenage age not good littler one's.if questions feel free email,	0	1	1	0

mom loved hotel mother 28 yo just london week visit family spent lot time researching affordable/nice hotels online, sumner fabulous hotel actually looks like pictures website s boasts, free internet lobby generally available wanted liked room*** clean size expected not hotel room big european hotel room size definitely room bathroom appropriately sized clean overhead shower option opposed hand held, a/c control nice window, bed comfortable flat screen tv, no safe room hotel leave passports, small frig room room leave items e.g. mom kept bottle wine, location great ran morning hyde park area restaurants local eateries close major sites hyde park marylebone buckingham mile kensington just park tube lines/bus stops close, suggest getting local street map hotel does not pretty sure took tube walked switch tube lines lot walked eliminated changes, hotel staff helpful looked byo places neighborhood dinner, breakfast totally convenient coffee delicious, mom not big morning eaters ca n't comment eggs bananas cereals perfect mom nice bread yogurt apple/orange selection, overall great stay, suggest requesting room not floor breakfast imagine loud breakfast hours breakfast smells tend waft, room not near breakfast ca n't say sure,	0	1	0	1
welcome way station lobby spacious inviting, especially appreciated guest pc internet access, lobby staff friendly efficient, room comfortable beds nicely appointed well-maintained.the location close airport terminal free shuttle service great.the reason rate average n't premium experience money just fine,	1	0	0	1
rooms clean air conditioner worked great rooms nice inside air conditioner worked great room small frig micro-wave coffe maker, definently stay,	1	1	0	0
stunning stunning views did lot research choose hotel 4 nights hong kong way home month new zealand, came regretted, views hotel lobby baror harbourside room simply best hk hotel isliterally edge water, uninterupted view everchanging traffic water, important, hk really busy urban environment know, staying simply exit hotel walk waterfront away constant bustle, breakfast great food nothing speak hey hk restaurants,	1	1	1	1
not nothing, maybe no thing perfect hotel paris price demand little gem gets pretty close, wanting value money, does n't amenities big boys does n't charge dollar, website book special rates 72 hours stay, breakfast acceptable rooms clean comfortable not large, located 10 minutes eiffel tower sparkling window arrived late night, good restaurants nearby close city tour bus routes, went paris stay,	0	0	0	1

casablanca stay sam stayed hotel october 7th 11th excellent stay, booked directly hotel web site no problems reservations communications prior stay.this trip nyc did n't know expect delighted hotel love casablanaca theme decor staff friendly helpful room bigger exepected fantastic bathroom shower location 100yards time sqaure, cleanliness public areas room excellent.the hotel oasis calm right times square.the room rate expensive opinoion worth definately stay, people mentioned busy area no problems noise.breakfast afternoon snacks great.thoroughly recommend,	0	1	1	1
good not great stayed f1 gp great location rest city half way sites business distrcit shopping clarke/boat quay lobby nice check quick staff friendly room ok good size bathroom bit stingy towels balconey nice touch think rooms pool area pleasant little airflow boiling hot sing ink bar little disappointing bit clubby clientle breakfast good exec floor room dining good portions little miserable overall staff excellent little pricey reasonable value money told,	0	1	1	1
great location adequate well-priced rooms central florence stayed triple room beacci tornabuoni 2 nights friends, biggest plus excellent location historical district, 5 min, walk duomo piazza di signoria uffizi gallery ponte vecchio 10 min, main train station taxi 9 euros, tornabuoni major upmarket shopping street good local restaurants easy reach, great price 135e large-sized triple room off-season, room 2 no view fine, not expecting luxury rarely room sleep, room clean plenty space store luggage small fridge/minibar small tv mainly italian channels adequate bathroom shower/tub combo bidet, breakfast fine hard-boiled eggs breads cheese ham yogurt fresh fruits no custom-made items, short perfect hotel main decent clean place central location, want luxury look,	1	1	1	1

exec level average airport hotel used city break boston 4 nights, booked exec level room 10th floor 199 plus various taxes, 10th floor certainly re-fit new furniture decorations, overall scheme pleasing, exec level access lounge free continental breakfast hot snacks evening, note no free alcoholic drinks massachusetts state law.pros easy 5 min walk termial catch silver line downtown 10 min journey 10 min walk terminal e. shuttle runs fequently not bothered walk want blue line subway.cons stayed hotel 5 times nice hotel felt experience little disappointing, check n't 4.00pm arrived international flight 2pm not room ready baggage man surprised wait checked bags wait, think address quite easily no reception bothered, did check no apology lengthy wait 3 hours rooms spacious comfortable no complaints, no noise airport neighbours quiet no disturbing noise, housekeeping bit hit miss dust clearly visible arrived gradually disappeared dueing stay, bedding towels changed everyday, bed comfortable exec lounge nice facility breakfast changes cereals fruit bagels/muffins/pastries cold boiled eggs, really expect little variation, n't cost hot eggs bacon menu hilton hampton inn manage, said fresh good quality staying 1 2 nights ok. longer send looking alternatives.staff polite little remote no complaints, overall worth money given accommodations facilities compared downtown boston prices easy downtown want, little bit investment outstanding,	1	1	1	1
bad hotel bad hotel horrible arrived friday june 20th florence boiling weather hot humid.in room air conditioned did n't work broken did n't sleep, day repaired apparatus air not cool, paied 450,00 4 night room kind basement flat not clean.breakfast not good sub-bran product, calculate 4 star hotel ex ambasciatori price.the advantages internet free closeness central station.i want suggest tipical tuscan restaurant good price ristorante fagioli 50122 firenze fi 47/r corso dei tintori tel 39 055 244285. closure august saturday sunday,	1	0	0	0
fantastic little hotel booked hotel reviews trip advisor.i not disappointed engish speaking reception staff clean small rooms cleaned day towels replaced everyday free electronic room safe.did not try breakfast not breakfast person.good air rooms essential road outside noisy.only mins walk fron gare nord.2 mins notre dame llorette metro station.would use planning year,	1	1	1	1

good hotel care booking fira palace occasions thoroughly enjoyed stays, rooms large good condition, hotel good location montjuic park quite peaceful hectic day town easy reach centre metro.there things look buffet breakfast comes little expensive 15 euros head no simple coffee pastry option 10am, secondly launch reduced rates web site, staying room 131euros neighbor paying 97 euros, thirdly booked year asked superior room, arrival placed turned standard room, having knew standard room contacted reception, checked room rate confirmed paying superior room price, challenging did finally admit given standard room.so looks like case beware,	1	1	1	1
lovely oasis calm stayed hotel le lavoisier weeks ago night way provence, lovely experience, fabienne feel welcome particularly helpful recommending restaurant dinner, room quiet overlooked street, definitely return trips,	1	1	1	1
fabulous boutique hotel stayed nights recent trip london, stayed company hotels prior trip charlotte street hotel loved wanted closer soho, treat, facilities top-notch hip decor great restaurant/bar buzzy energy late-night honor-system snack bar ground floor sitting rooms/libraries, room surprisingly large surprisingly quiet, hotel gym surprisingly good top-of-the-line equipment, hotel epicenter soho film industry sports private screening room sunday afternoon film club week, overheard film-industry conversations bar bumped couple script-toting people elevator, disappointment strange wifi connect wifi signal desk code appears internet browser connected strangely complicated, wifi kind steep charge 30p minute maximum 20 gbp day, free place like no,	1	1	1	1
stay away hotel located strange location, it__√á_√©_ not close, hotel does provide shuttle local restaurants, desk completely unhelpful needs training, hotel staff acted like control, air-conditioning room did not work, asked new room said sorry, told checking did room sold-out hotel, beds old uncomfortable, breakfast area not large hotel no place sit, plus 9am, won__√á_√©_ stay hotel,	0	0	0	0
besy hotel amsterdam, amsterdam times different degrees luck various hotels sampled, nadia far away nicest hotel stayed city, booked reading reviews, fantastic value money cosy friendly centrally located, rooms tiny like cabins need, staff nicest people hope meet nothing trouble, stairs overwhelming open door receptionist peering miles away floor instructed leave bags hall staff quickly brougt rooms, notice complaining having wait room ready wait staff escort dining room offer free hot cold drinks hardly hardship, breakfast simple continental spread lovely raisin toast croissants cheese great coffee, dining rom really cosy like rest hotel.will definitely staying,	1	1	1	1

california business added personal days visit favorite city, castle inn great location away chaos union square disneyland fisherman wharf close catch cable car walk bit, particularly sold castle inn included on-site parking, end spending 25/day deal hastle getting car ability hop car distant destinations great plus.the rooms modest comfortable clean, area safe quiet, varsa concierge terrific, great knowledge city helpful just generally charming, believe continental breakfast included opted block polk st. great coffee peet, forget starbuck motel not fancy great deal, hope come family soon, huge suite accomodate large family, just hint-i think inn directly better rate online deals-ask varsa,	0	0	1	0
okay hotel commonwealth decent place, convenient fenway park view freeway disappointment, decor little weird,	0	1	1	0
great hotel florence greatful hotel disaster lodging rome.this place surprisingly nice clean newly refurbished rooms new bathroom fixtures.i really enjoyed stay concur positive reviews, return italy think like station florence hotel faenza make day trips italy location.finding hotel train station challenge stick maps ask, italians friendly eager help,	1	1	1	1
good location not hotel hotel location best thing, carpets room common areas stained worn, charge internet access, booked night priceline hotwire, night charged additional 9.96 amenities fee additional 15 room tax, not big deal fees unexpected planning book hotwire priceline aware, check extremely helpful staff places eat, good restaurants bellhops desk staff, lot not good overpriced restaurants nearby really important ask just read posted menus did night regret, not particular review want rental car charlie car rental best rates really friendly helpful staff, pick airport hotel return, airport location open 24 hours day downtown location just open day, nice new cars reasonable rates, need car visit rain forest, sure tour park ranger, free informative,	1	0	0	1
great stay husband spent 4 nights hotel, wonderful stay room 24th floor, great view comfortable bed.this hotel close dadeland mall miami metrorail.we access concierge lounge serves continental breakfast daily hors d'oeuvres afternoon.all great stay, defintely stay,	1	1	1	1
n't hesitate soon can-it wonderful resort wife 2 daughters 5 3 years old best vacations, planning come punta cana stay wonderful resort, non-expensive people friendly rooms comfortable food awesome options lots pools specially beach fun miss that.kids friendly kids pool facilities recomend resort, went punta cana stayed resort reading lot bad comments property decided, believe mad right selection price	0	1	1	1

paid 500 4 including meals room tips want.ocanview rooms comfortable train end enjoy resort.5 days 4 nights package looking decent nice resort,				
great hotel writing review evening staying palace hotel, great hotel, rooms tidy nice character, staff extremely helpful receptionists spent half hour phone calls place trying tokyo making feel unwelcome time, breakfast nice hotel shopping arcade restaurants fitness centre, close jr tokyo metro otemachi stations place tokyo, course imperial palace stone throw away, course using complementary internet connection room write review try massage place basement floor, massage absolutely heavenly great way tired feeling jetlag,	1	1	1	1
lovely hotel summer traveled europe month stayed wonderful hotels voted hotel santa maria novella best hotels, hotel santa maria novella beautiful, entire staff friendly extremely accommodating, rooms elegant, truly feel vacation, centrally located, train station literally street, best walk florence main attractions, given train station just steps hotel decided day trip pisa awesome addition itinerary, stayed nights hotel santa maria novella, breakfast staff warm friendly opinion just wanted wonderful stay, just lovely, happened st. john day religious holiday, coolest thing roof returning dinner enjoying fireworks, far places eat hotel staff best restaurant recommendations.we definitely return hotel santa maria novella,	1	1	1	1
needs work 6 stayed fenice palace june 17-19. booked hotel great location note not dissapointed, close duomo ponte vecchio florence great walking city anyway.i say street noise second floor rooms sleep impossible air conditioning weak noisy not acceptable prices paid, bathroom room 209 think basic miserable shower hardly room turn, know europe shower stall better flimsy curtain sticks allows water drench floor time bathe.staff generally pleasant tolerant weak attempts italian, check check painless efficient, just wish complaints ac met empathy.i think probably better hotels available trip italia no idea, travel wonderful city, stay however.ciao,	1	1	0	0
hardly better, excellent hotel, location ideal 10 min walk forbidden city, staff helpful organised tours great wall, restaurants jncluidng 24h grand cafe class added convenience, staff helpful spoke english, rooms clean quiet serviced really picky say air heating slightly efficient, definitively recommended definitively choice went beijing, note booked hotel website direct cheapest way book room,	0	0	0	0

did n't want leave stayed uma ubud week october absolutely loved, researched tripadvisor impressed consistently high reviews, wanted laidback luxury good facilities 25m pool gym yoga, certainly n't feel restricted resort not resort really like boutique hotel fabulous eating options close, favourites naughty nuris great ribs cooked roadside bbq ibu okra spelling wrong best suckling pig sit floor eat, ate mozaic half liked bit bland, make sure reserve table ouside garden really disappointing elevator type music dreadful, uma really perfect welcome pick-up airport cold towels winner fantastic friendly staff, got really ill able supply medicine, luckily knocked 24 hours, terrace room loved great terrace chilling love way staff light outdoor candles night, bathroom perfect spacious, spa nice really strongly recommend walking minutes road direction ubud bali botanical spa treatments fraction price surroundings far beautiful balinese, discovered second day sorry day, envy uma flash, spend second week friends villa canggu fabulous, difficult not enjoy bali,	0	0	0	0
westin madrid, worn decent just returned 4 night stay using starwood points room property madrid, platinum starwood member try use points nicer properties expensive especially european locations, liked best property location excellent, prado door modern museum art square, starbucks located block decent coffee pastry, spanish version dennys called vips food cheaper restaurants not interesting just filling.what did not like hotel gouge, small bottle water 7, rollaway 50 euros night cup coffee hotel 5, stay away overpriced not interesting breakfast buffet 22 euros, concierge lousy restaurant recommendations tapas food not terrible quality high price, point returning let know poor recommendation thanked day returned reason overheard giving recommendation customer, typically means getting sort kickback steer certain places eat, avoid concierge directions, rooms walls heard everyone__√á_√©_ door open close room, superior room spacious bit worn, so-so, rate room 500 euros night fled night using points stay did not mind, decent hotel stay use points absurdly overpriced prices charge accommodations, queen size bed seen better day springs tended squeak got bed, faced square kind cool constant street noise meant kept window closed.in end fine using points hotel not worth charging room,	1	1	1	1

worst hotel experience booked nonsmoking room online weeks advance stay crowne plaza downtown seattle, arrival desk staff asked consider smoking room, completely unacceptable family clearly stated no not consider smoking room mainly concern infant daughter health, particular staff member went speak quietly desk staff member agreed particular room, entered room smell cigarette smoke apparent not imagine desk staff intentionally smoking room informing, days later discovered ashtray matches checked desk discover fact smoking room intentionally not kept reserved nonsmoking room, clothes belongings baby reeked cigarette smoke.in addition appalling lack concern health wishes room not serviced night stay, asked desk manager room not serviced checked housekeeping informed not disturb tag door did not disturb, not case tag inside door entire time anxiously awaiting staff garbage dirty diapers replace towels, desk manager offered pathetic compensation complimentary hotel breakfast family exposed carcinogenic cigarette fumes days not having rooms cleaned, definitely stay crowne plaza hotel,	1	0	0	0
aka hh campomanes wonderful experience lovely contemporary hotel right heart central madrid, street opera metro stop near plaza isabell ii, great location, walking distance teatro real puerto del sol, calle arenal great shopping street neat nightclubs, rooms great individual a/c 2 comfy twin beds, no noise whatsever, rooms include minibar staellite tv internet access planned, bathrooms extremely clean hot running water, housekeeping turns sheets toiletries daily, staff friendly speak good english, good info local events tourist stops help buy tix operas shows flamencos etc.breakfast buffet included 4 types cereals fruit juice/milk/water sliced cakes bread sliced meats turkey ham cheese fresh fruits yogurt criossants, breakfast open 8am-10:30am good stint keeps tucked 2pm madrid lunchtime, stayed 4 nights christmas break loved, comes price 99 euro night not cheapest place, definitely really good value money, gladly recommend, tip want laundry coin operated laundromat lavamat calle la cruz close puerto del sol, hotel dry-clean clothes not worth price, not tell coin op laundry obvious reasons,	1	1	1	1
castle inn great value booked castle inn direct got informative e-mail set standard 4 night stay, room standard motel room clean decoration good equipped microwave fridge/freezer, lady desk gave lots information not checked, helpful stay, overall great value agood place stay staying time san francisco,	1	1	0	0

worth, overall good stay, good points rooms cleaning continental breakfasts location swimming pool holiday inn door reception, poor points lifts slow coffee facilities room, coffee percolator en suite, coffee/milk/sugar maybe 2 cups, overcome tea bags milk sugar breakfast, not coffee rooms right no homeless seen area, excellent location fishermans transport parts city.overall happy,	1	0	0	1
good resort just came week vacation caribe club princess punta cana, sure value, beach gorgeous complex just big pools clean beautiful staff friendly eager good service, caribe complex tropical princess enjoy infrastructure hotels, pools 7 remember correctly la carte restaurants big buffet restaurant long beach, food good buffet, wide selection food, did n't enjoy el pilon dominican la carte restaurant just matter taste, rooms ok. little little ants surprise discover big cockroach bathroom, crawled bathroom drain, stinky smell present entrance building, n't know came gross, great resort apart little problems, grounds beautiful beach fabulous staff great pools excellent food good, drink pool bar talk francisco real great guy,	1	1	1	1
stay hotel ny stayed couple nights hotel, right heart times square, previous review said rooms small immaculately clean serviced quite, little local knowledge ask room no noise street unlike hotels stayed previously.the staff helpful immpecable manners worked good nature.the glass wine provided lounge stayed 3 hours offer welcome, crackers cheese grapes complimented choice red white fizzy wine, oh coffee tea cookies available 24hrs.we stay no hesitation recommending hotel.kr,	0	0	0	0
highly recommend hotel stayed superior king room view overlooking boylston st. historic hotel bay area block away amtrak station subway station nearby, lenox fully restored beautiful lobby plush furnishings, room spacious appointed quality furnishings walk wardrobe attractive bathroom quality fittings accessories, king bed extremely comfortable, hotel quality restaurant separate bar irish pub hotel, hotel numerous restaurants bars boylston nearby newbury sts, staff helpful friendly, booked hotel historic hotels america website advance purchase rate, make point return lenox visiting boston future, highly recommended,	1	1	1	1

good hotel, stayed 3 nights january, room warm windows opened ventilation, small kitchen great boiling water tea storing beer things fridge gas stove oven immaculate, fancied cook, room big decor good bit dodgy places not spoiled, bathroom nice water pressure shower bit toooo, mini bar big tv loads amenities business people cd player built alarm clock fax printer internet access, stayed 32nd floor reasonable view just faciniating watching new york hotel location n't bad prepared walking did n't bother think best way explore healthy, blockheads mexican restaurant block right hotel great place youngsters llike drunk eat nice food, come staff friendly helpful did n't bother, barking dog good food drinks smokers like n't bad going outside sneaky cigarette temperature, great, 6 stars,	1	1	1	1

Table 31.
Case study results (BERT-SVM, Employee data: Glassdoor reviews)

Note. The red colored numbers indicate wrong identifications.

text	control mutuality	commitment	trust	satisfaction
Culture , good people, team oriented	1	0	0	1
Remote opportunities for people if they really want that. My non-mgmt coworkers were great people.	0	0	0	1
Cons: Wear and tear on your vehicle: DoorDash drivers use their own cars to make deliveries, which can put a lot of wear and tear on your vehicle. This means you may need to pay for repairs or maintenance more frequently. Risk of accidents: Driving around all day can be dangerous, especially in busy areas or during bad weather. Inconsistent earnings: The amount you can earn with DoorDash can vary depending on the number of deliveries you make and the tips you receive. This can make it difficult to rely on DoorDash as a steady source of income.	1	1	0	0
This is a very toxic work environment. Leadership is a joke - seriously. I think they all just chill and grocery shop during the day then complain that they have too many people to Manager (usually 4-5 ppl). The bar is incredibly low for leadership at lyft. I've never seen anything like it. Nothing is expected of them and when cuts come - they'll cut the entire workforce (which has happened) before they let the lousy managers go. They are NOT subject matter experts and are not expected to know anything.	0	0	0	0
Smart colleagues Half day Fridays Reasonable compensation	1	1	0	1
great place unlimited pto scope for growth	1	1	1	1
1. Once you learn everything this job is fairly easy.	0	1	0	0
cons	0	0	0	0
Great benefits. Training is well planned and the managers work with you when it comes to learning the product. Days can be a bit challenging when it comes to making call 80 dials minimum plus 2 hours of talk time are the minimum.	1	1	1	1
questionable decisions by management make it hard to want to stay	1	0	0	0
-Open PTO -Performance Based Bonuses - Work-life balance	1	0	1	1
-Decent pay for an entry level position. -Good opportunity to learn the business	1	1	0	0

Hybrid work from home schedule Catered lunches Open PTO	1	0	0	0
great wlb great community and culture	1	1	1	1
Really smart people, a lot of opportunity for growth, always encouraged to be innovative, think big, and create something new. Competitive salary and benefits with other major tech companies. 100% self motivating work environment. No dress code and 4 legged friends are welcome.	0	1	1	0
Changes come quickly and organizational changes aren't always very transparent, so I've found I need to be very flexible and go with the flow	1	0	0	0
Lack of transparency, and goals not attainable	0	0	0	0
Easy. Flexible schedule. Independent. Relaxing	0	0	1	1
Work - life balance & salary	0	0	0	0
Toxic work environment within customer support team, constant leadership and org changes, unstable US culture from multiple layoffs, limited growth in the US due to restructuring talent to Mexico City	0	0	0	0
Every growing young business has their challenges, this one is no different, despite the size of the business.	1	1	1	0
Flexibility, involved in the community, instant payouts, self-employed, work at your own pace/risk	1	0	0	1
Promotions are somewhat political, oriented to how visible you are to upper management. Many moving updates across products can be difficult to keep up with - nothing incredibly out of the usual among similar tech companies. Confidence in C-Suite to do the right thing.	1	1	0	0
friendship and good environment with members	1	1	1	1
Instant access to pay and cashout	1	1	1	1
Best delivery experience I've had!	1	1	1	1
Great community Great communication Great productivity	1	1	1	1
Miscommunication. No one is on the same page in terms of what information is needed, leaving confusion for what process to take or information to give to viewers calling. Horrible Bonuses.	0	0	0	0
None that I can think of.	0	0	0	0
I genuinely love working here. I have been in tech for over 15 years and can confidently say this is the best of the best. I have worked here now for almost a year and have had the same experience consistently. I feel supported, I am not burnt out, and mangers lead with EMPATHY.	1	1	0	0

- Culture is like nothing I've experienced before. Co-workers truly assume good intentions, I haven't encountered any snarky or pretentious people. If you make a mistake nobody belittles you for it. - The mission is really important. Hiring seems broken everywhere, and Glassdoor is genuinely trying to fix it. -Benefits are quite good, with great health insurance family plans and wellness options like Classpass - Work life balance actually means something. I generally start and end my day at reasonable times and rarely need to work beyond that. Unlimited PTO and ability to actually use it is also great -Transparent leadership	1	1	1	0
- Amazing people to work with. Seriously, they're the smartest, most humble and passionate people you'll come across. Just about everyone believes in what we're doing. - Glassdoor genuinely care about the whole person. This is the first employer I've had that I haven't had anxiety about having to leave to take care of my child, or take care of myself when I'm sick. It's like they really understand that there is more to employee lives than Glassdoor work. - No boys club or high school style politics here. You do your work well and contribute to the team/company? You move up. Backstabbing or any other oppressive types of behavior are unacceptable here. This place will literally change how you feel about work.	1	1	1	0
Culture Management Leadership Reviews People	1	1	0	1
Good start up company out of college	1	1	1	1
Need to work during peak meal hours	0	0	0	0
The tech stack is improving but there are still legacy apps to deal with. It felt the compensation wasn't very competitive	1	1	1	0
Self-driven, tech based position with flexible hours and ability to make your own paycheck.	1	0	0	1
Great company that wants to help you succeed	1	1	1	1
Ok pay better than most of the other delivery apps i used	1	1	1	0
Fast learning and new technology	1	0	1	0
I'm starting to yhink there isn't one other than you can make some money on the fly other than that can not think of one	0	1	0	0
Processes and tools change frequently which makes it tough	0	0	0	0
Still runs like a start-up and processes are not in place to streamline and scale up as efficiently as they should be at this stage. I feel like I'm doing the job of 3 people.	0	0	0	0
Having someone in your car. Miles on your car.	1	1	1	1
Ever changing quota that could be surprising then difficult to manage	0	0	0	1

Get paid immediately. Can work on your own schedule. Nice people.	1	0	0	0
As the culture transitioned to becoming more corporate: wage gaps, not family friendly, poor management, employees not valued, minimal opportunity for promotion.	1	1	1	1
Very long hours for not enough pay	0	0	0	0
Door Dash is a wonderful company to work for. I love many things about being a Dasher. Mainly I love the service I provide for individuals who may be disabled, busy, at work, or just relaxing. I truly take pride in being there for others by lifting burdens when I can. Being a Dashing is also a wonderful experience, because it allows me to interact with so many different individuals each day, which gives me the opportunity to put smiles on many faces daily. It's like I'm going on an adventure each day. For example, I took a few days off to take care of my car, and my new challenge for these next 3 days are to see how close I can get to my weekly financial goal in just 3 days as opposed to 7. Another aspect I love, is the flexibility. I'm building a nonprofit organization currently and mental, physical and spiritual wellness are very important to me in my personal life. Door Dash allows me the flexibility to make a schedule that works best for me, one that enables me to take care of my living expenses, work in my nonprofit and stay mentally, physically and spiritually well. There are many more reasons why I love being a Dasher, however I'm working on over communicating 😂😂😂😂	1	1	1	1
at 12.50 an hour you'd be better off getting a job at McDonalds. You can always do Uber after your shift if you need to make extra.	0	0	0	0

Table 32.
Case study results (TF-IDF-SVM, Customer data: TripAdvisor reviews)

Note. The red colored numbers indicate wrong identifications.

text	control mutuality	commitment	trust	satisfaction
friendly teenage son stayed hotel casci september, experience confirmed excellent reviews trip advisor based decision book hotel, staff friendly helpful especially son accidently left treasured nba singlet room, quick phone ascertained handed housekeeper posted australia cost, stopover florence days later enroute venice allowed quick taxi ride train station pick personally, hotel casci excellent central position room impecibly clean quiet faced street breakfast fine great coffee offers hot chocolate son, staff gallery bookings recommendations tours, economical hotel choice iwith real personal touch not hesitate recommend,	0	0	0	1
enjoyable stay enjoyable stay hotel convenient location wonderful room perfectly soundproof considering work renovations close wonderful breakfast, got room average 200 euros includ, bkfast little bit expensive end happy splurged having seen pictures hotels similar category, u want really special advise u splurge bellavista room floor big balcony wonderful panoramic views florence sunlit rainshower bathroom, think 315 euros, fault hotel staff courteous professional not friendly, lots smiles hotel experience.without hesitation stay, note hotel renovating adding rooms good noise insolation outside did ask work not started till 9am, sensitive u ask room away work renvoation.ps u love shopping brand names ask tour mall prada outlet worth especially shoes,	0	0	0	0
great hotel business travellers couples travel beijing stayed hotels years great, attentive service clean comfortable rooms friendly staff good set restaurants, lots really good restaurants close local flavor especially hot pots peking duck, concierge knowledgeable helpful, terrific beds pillows great bath robes, good value money,	1	0	0	1

beautiful hotel destroyed terrible service, stay started problems desk, 20 minutes 3:00. told check time 3:00 lunch come 3:00. returned lunch given room told luggage room, no concierge took room took 20 minutes room located far away, course luggage not, went desk told did not want room, said okay not room 6:00. sat lobby travel clothes waited room, concierage told tell hotel 9:00 morning, n't tell 2 hours sat lobby, wanted try sell time share, got introduction hotel, 6:30 finally got new room luggage, day totally wasted, turned room utility closet did not sleep night heard pipes night, went lobby morning asked room changed, said n't hotel 100 booked no rooms, hotel actually 25-30 booked, finally gave room concierge did time showed room, room 3rd floor night door roof blowing opening closing, went lobby complain gave 100 booked speech, head guest services come not believe bad, called maintaince 15 minutes door closed, end problem, massages spa droped 250.00, massages good, day took massage workshop zen pavillian, told spa use pool steam room sauna, no showed did wrong, fabulous cabanas outside massage single couple, no offered instead massages upstairs listened drilling doing downstairs construction spa, bad did not offer beautiful cabanas, food average- enjoyed japanese asian best, outdoor restaurants did not open rained nanos did not open day- told hotel not 2 outdoor buffets open, hmmmm not change room 100 booked restaurants did not open did not people, not impressed market grill food laid day 11:00 6:00. atmosphere bummer met lobby restaurants pool nothing similar complaints, shame hotel beautiful grounds gardens really nice, service terrible food just fair rooms musty nightly entertainment weak, overall recommend resort return, travelled 30 years great places, real disappointment felt like waste money, way did tip including real nice tip desk arrive did not help, trip bummer,	1	0	1	1

sun sand smiles people, want 5 star pay 4 star resort, stayed hotel weeks april 12 april 26. flew air transit newfoundland canada, reading review wondering resort like.it beautful place visit.the problem resorts people stay, remember going country golden rule, want you.respect hotel..the rooms clean, bed comfortable clean, seen bug hotel, love clean room sheets got more.the cleaning lady cleaned room nice thanked day dollar inportant me.you leave gifts pillow not, remember people 30 t0 40 dollars weeks work hard heart.. look evey day say hola country respect food.. plenty food eat not hungry, book 3 different restaurants check, reviews people said got sick food, maybee 12 people went didnt, fussy come food safety.i eat hot food hot cold food cold, problem, wait staff dining room best.for 2 weeks cleaned table got need drink meal problem, canada ate restaurant tip 5 10 dollars time eat depending cost meal.i tipped 1 dollar meal beaches best place feet, travel alot happy here.you walk direct meet lots people walk direct meet, easy beach, quite place lie sun tan places booze inclued people noisey drunk, room night time12 1..2..3..or 4..am good time voice noise carries people sleeping bar..the drinks really good sandys bar..say hello sandy elizabeth rest bar staff, lot people staying hotel wants served first.. holidays relex mix best drinks, best vacations took time enjoy vacation people helped make happen dominican people staff hotel best respect little enjoy holidayall couples respected here..very gay friendly,	1	0	0	1
fantastic stay pulitzer wife recently stayed pulitzer nights stop honeymoon plesantly surprised especially reading mixed reviews ta, hotel located great town block tram stop anne frank house, upgraded large beautiful canal view room, bed fantastic especially 10 hour flight la, staff extremely helpful professional, used starwood points room ca n't comment value relative hotels area loved stay unique comfortable property return,	1	0	0	1
great value location stayed separate times week march flights burbank airport, hotel convienent airport 5 miles away local sights visiting hollywood area universal warner brother studios hollywood bowl la brea tar pits, breakfast free parking plus, bathroom cleaned bit better visit rest place good, visits manager left basket goodies cookies candy water bottles greatly appreciated long travels lots la traffic jams, faced freeway times did n't traffic noise intrusive, pool small not reason, visited hilton property street friends staying fancy/nice attractive pool scene considering mandatory parking no breakfast provided- holiday inn express better value site seeing not planning alot time hang hotel,	1	1	1	1

lovely hotel lovely area husband stayed nights honeymoon wonderful experience, room small appointed clean, breakfast courtyard delightful, walter sandro helpful making recommendations restaurants called reserve table popular restaurant, neighborhood major tourist areas comfortable walking distance sights, bars restaurants area wonderful local places i.e, no tourist menus highly recommend,	1	1	0	1
fantastic hotel fantastic country, stayed 3 days singapore route bali honeymoon wish stayed longer, hotel lovely staff helpul pool area great not second feel middle city, stayed garden wing room good size balcony over-looking pool area, ate restaurant blu nights great bit pricey worth, really nice bar area champange cocktails views bar restaurant amazing, breakfast line really not fault place choice different types food seen believed buffet style turnaround food quick nothing gets cold, thought eat curry breakfast, singapore.. say definitely hotel stay doubt, p.s, area singapore called clarke quay really cool bars/restaurants/clubs defintely worth visit, did n't place night bit shame,	1	1	1	1
great location hotel nice suprise, rooms bathrooms clean tastefully decorated decent size bidet not older tired look expected, took public bus airport stop pizzale roma minute walk bridges little odd suitcases lots people doing no ramps bridges deal stairs cheap commute 3gbp unlike believe water shuttle, conveniently located close stores restaurants walk st. marks square, hotel arrange complimentary water shuttle glass factory no sales pressure watch glass blowing leave like, breakfast morning sitting outside canal lovely self serve continental breakfast adequate, opposite hotel beautiful church interior impressive 2 mins check, outside hotel gondala, allowed drink wine bottles seating area canal instead bar service, staff wonderful helpful, not hesitate stay,	0	0	1	1
loved secrets boyfriend went apple vacations secrets excellence punta cana, absolutely beautiful, left reading alot reviews kindof nervous, not sure people giving resort bad reviews, people friendly little spanish helps, simple things like hello good night thank, wonderful time grounds absolutely beautiful food drinks good, went excursions outside resort, highly recommend outback safari tour guides great learn dominic, highly recommend resort,	1	1	1	1

location great hotel n't, went away barcelona week partner, impressions hotel looks good, bus airport drops road realised saw past walked 15mins booked double room got twin room, went reception ask moved surprise birthday trip partner said busy n't, partner fine ok.staff ok mornings cleaner just walk rooms knocking got annoying 3 days 8-9a.m, mentioned reception bothered again.the hotel noisy light sleeper n't stay luckily read previous reviews took ear plugs didnt help.breakfast not great dishes picked food stuck not impressed, happen week there.location good, prefer not stay city slightly hotel convienant, partner decided barcelona stay little bit closer city night times make sure didnt metro late area hotel did bit dead.we loved barcelona sadly hotel not, def, barcelona soon felt week not,	1	0	1	1
great budget hotel hotel layne excellent budget hotel highly recommend, close heart san francisco costs little, run hospitable family patels possible make guests welcome, rooms clean comfortable far superior hotels higher price brackets.stay,	0	0	0	1
march 23-30 great vacation great time despite worries resort, went idea just relazing not worrying minor things, days left resort major, did change things work, got little tired temp, buffet set beach end walking punta cana princess lunch, highly recommend want change lunch menu.go platinum, say platinum juan v. know, great guy, saying told website know reading, hey juan thanks presidente, took advantage platinum ordered room service times, especially breakfast n't feel like getting going breakfast, breakfast patio awesome.there huge langauge barrier country, did notice staff especially friendly spoke good english, gals went did speak spanish extremely helpful, highly suggest pick basics spanish.food good complaint warmer, did n't just food.beach..one word awesome, jamaica twice just does not compare, really feel like walking powder, nice walk ocean not step soft sand, ocean quite bit rougher used riot play, definately, having restaurants bars beach big difference just, say probably not bring young children resort, resorts area cater alot younger kids, good teenage age not good littler one's.if questions feel free email,	0	0	0	1

mom loved hotel mother 28 yo just london week visit family spent lot time researching affordable/nice hotels online, sumner fabulous hotel actually looks like pictures website s boasts, free internet lobby generally available wanted liked room*** clean size expected not hotel room big european hotel room size definitely room bathroom appropriately sized clean overhead shower option opposed hand held, a/c control nice window, bed comfortable flat screen tv, no safe room hotel leave passports, small frig room room leave items e.g. mom kept bottle wine, location great ran morning hyde park area restaurants local eateries close major sites hyde park marylebone buckingham mile kensington just park tube lines/bus stops close, suggest getting local street map hotel does not pretty sure took tube walked switch tube lines lot walked eliminated changes, hotel staff helpful looked byo places neighborhood dinner, breakfast totally convenient coffee delicious, mom not big morning eaters ca n't comment eggs bananas cereals perfect mom nice bread yogurt apple/orange selection, overall great stay, suggest requesting room not floor breakfast imagine loud breakfast hours breakfast smells tend waft, room not near breakfast ca n't say sure,	1	1	0	1
welcome way station lobby spacious inviting, especially appreciated guest pc internet access, lobby staff friendly efficient, room comfortable beds nicely appointed well-maintained.the location close airport terminal free shuttle service great.the reason rate average n't premium experience money just fine,	0	0	0	1
rooms clean air conditioner worked great rooms nice inside air conditioner worked great room small frig micro-wave coffe maker, definently stay,	0	0	0	1
stunning stunning views did lot research choose hotel 4 nights hong kong way home month new zealand, came regretted, views hotel lobby baror harbourside room simply best hk hotel isliterally edge water, uninterupted view everchanging traffic water, important, hk really busy urban environment know, staying simply exit hotel walk waterfront away constant bustle, breakfast great food nothing speak hey hk restaurants,	1	0	1	1
not nothing, maybe no thing perfect hotel paris price demand little gem gets pretty close, wanting value money, does n't amenities big boys does n't charge dollar, website book special rates 72 hours stay, breakfast acceptable rooms clean comfortable not large, located 10 minutes eiffel tower sparkling window arrived late night, good restaurants nearby close city tour bus routes, went paris stay,	1	1	1	1

casablanca stay sam stayed hotel october 7th 11th excellent stay, booked directly hotel web site no problems reservations communications prior stay.this trip nyc did n't know expect delighted hotel love casablanaca theme decor staff friendly helpful room bigger exepected fantastic bathroom shower location 100yards time sqaure, cleanliness public areas room excellent.the hotel oasis calm right times square.the room rate expensive opinoion worth definately stay, people mentioned busy area no problems noise.breakfast afternoon snacks great.thoroughly recommend,	0	0	0	1
good not great stayed f1 gp great location rest city half way sites business distrcit shopping clarke/boat quay lobby nice check quick staff friendly room ok good size bathroom bit stingy towels balconey nice touch think rooms pool area pleasant little airflow boiling hot sing ink bar little disappointing bit clubby clientle breakfast good exec floor room dining good portions little miserable overall staff excellent little pricey reasonable value money told,	1	0	1	1
great location adequate well-priced rooms central florence stayed triple room beacci tornabuoni 2 nights friends, biggest plus excellent location historical district, 5 min, walk duomo piazza di signoria uffizi gallery ponte vecchio 10 min, main train station taxi 9 euros, tornabuoni major upmarket shopping street good local restaurants easy reach, great price 135e large-sized triple room off-season, room 2 no view fine, not expecting luxury rarely room sleep, room clean plenty space store luggage small fridge/minibar small tv mainly italian channels adequate bathroom shower/tub combo bidet, breakfast fine hard-boiled eggs breads cheese ham yogurt fresh fruits no custom-made items, short perfect hotel main decent clean place central location, want luxury look,	1	1	1	1

exec level average airport hotel used city break boston 4 nights, booked exec level room 10th floor 199 plus various taxes, 10th floor certainly re-fit new furniture decorations, overall scheme pleasing, exec level access lounge free continental breakfast hot snacks evening, note no free alcoholic drinks massachusetts state law.pros easy 5 min walk termial catch silver line downtown 10 min journey 10 min walk terminal e. shuttle runs fequently not bothered walk want blue line subway.cons stayed hotel 5 times nice hotel felt experience little disappointing, check n't 4.00pm arrived international flight 2pm not room ready baggage man surprised wait checked bags wait, think address quite easily no reception bothered, did check no apology lengthy wait 3 hours rooms spacious comfortable no complaints, no noise airport neighbours quiet no disturbing noise, housekeeping bit hit miss dust clearly visible arrived gradually disappeared dueing stay, bedding towels changed everyday, bed comfortable exec lounge nice facility breakfast changes cereals fruit bagels/muffins/pastries cold boiled eggs, really expect little variation, n't cost hot eggs bacon menu hilton hampton inn manage, said fresh good quality staying 1 2 nights ok. longer send looking alternatives.staff polite little remote no complaints, overall worth money given accommodations facilities compared downtown boston prices easy downtown want, little bit investment outstanding,	1	0	1	1
bad hotel bad hotel horrible arrived friday june 20th florence boiling weather hot humid.in room air conditioned did n't work broken did n't sleep, day repaired apparatus air not cool, paied 450,00 4 night room kind basement flat not clean.breakfast not good sub-bran product, calculate 4 star hotel ex ambasciatori price.the advantages internet free closeness central station.i want suggest tipical tuscan restaurant good price ristorante fagioli 50122 firenze fi 47/r corso dei tintori tel 39 055 244285. closure august saturday sunday,	1	1	1	1
fantastic little hotel booked hotel reviews trip advisor.i not disappointed engish speaking reception staff clean small rooms cleaned day towels replaced everyday free electronic room safe.did not try breakfast not breakfast person.good air rooms essential road outside noisy.only mins walk fron gare nord.2 mins notre dame llorette metro station.would use planning year,	1	1	1	1

good hotel care booking fira palace occasions thoroughly enjoyed stays, rooms large good condition, hotel good location montjuic park quite peaceful hectic day town easy reach centre metro.there things look buffet breakfast comes little expensive 15 euros head no simple coffee pastry option 10am, secondly launch reduced rates web site, staying room 131euros neighbor paying 97 euros, thirdly booked year asked superior room, arrival placed turned standard room, having knew standard room contacted reception, checked room rate confirmed paying superior room price, challenging did finally admit given standard room.so looks like case beware,	1	0	1	1
lovely oasis calm stayed hotel le lavoisier weeks ago night way provence, lovely experience, fabienne feel welcome particularly helpful recommending restaurant dinner, room quiet overlooked street, definitely return trips,	1	1	1	1
fabulous boutique hotel stayed nights recent trip london, stayed company hotels prior trip charlotte street hotel loved wanted closer soho, treat, facilities top-notch hip decor great restaurant/bar buzzy energy late-night honor-system snack bar ground floor sitting rooms/libraries, room surprisingly large surprisingly quiet, hotel gym surprisingly good top-of-the-line equipment, hotel epicenter soho film industry sports private screening room sunday afternoon film club week, overheard film-industry conversations bar bumped couple script-toting people elevator, disappointment strange wifi connect wifi signal desk code appears internet browser connected strangely complicated, wifi kind steep charge 30p minute maximum 20 gbp day, free place like no,	0	0	0	0
stay away hotel located strange location, it__√á_√©_ not close, hotel does provide shuttle local restaurants, desk completely unhelpful needs training, hotel staff acted like control, air-conditioning room did not work, asked new room said sorry, told checking did room sold-out hotel, beds old uncomfortable, breakfast area not large hotel no place sit, plus 9am, won__√á_√©_ stay hotel,	0	0	1	1
besy hotel amsterdam, amsterdam times different degrees luck various hotels sampled, nadia far away nicest hotel stayed city, booked reading reviews, fantastic value money cosy friendly centrally located, rooms tiny like cabins need, staff nicest people hope meet nothing trouble, stairs overwhelming open door receptionist peering miles away floor instructed leave bags hall staff quickly brougt rooms, notice complaining having wait room ready wait staff escort dining room offer free hot cold drinks hardly hardship, breakfast simple continental spread lovely raisin toast croissants cheese great coffee, dining rom really cosy like rest hotel.will definitely staying,	0	1	1	1

california business added personal days visit favorite city, castle inn great location away chaos union square disneyland fisherman wharf close catch cable car walk bit, particularly sold castle inn included on-site parking, end spending 25/day deal hastle getting car ability hop car distant destinations great plus.the rooms modest comfortable clean, area safe quiet, varsa concierge terrific, great knowledge city helpful just generally charming, believe continental breakfast included opted block polk st. great coffee peet, forget starbuck motel not fancy great deal, hope come family soon, huge suite accomodate large family, just hint-i think inn directly better rate online deals-ask varsa,	1	1	1	1
okay hotel commonwealth decent place, convenient fenway park view freeway disappointment, decor little weird,	1	1	1	1
great hotel florence greatful hotel disaster lodging rome.this place surprisingly nice clean newly refurbished rooms new bathroom fixtures.i really enjoyed stay concur positive reviews, return italy think like station florence hotel faenza make day trips italy location.finding hotel train station challenge stick maps ask, italians friendly eager help,	1	1	1	0
good location not hotel hotel location best thing, carpets room common areas stained worn, charge internet access, booked night priceline hotwire, night charged additional 9.96 amenities fee additional 15 room tax, not big deal fees unexpected planning book hotwire priceline aware, check extremely helpful staff places eat, good restaurants bellhops desk staff, lot not good overpriced restaurants nearby really important ask just read posted menus did night regret, not particular review want rental car charlie car rental best rates really friendly helpful staff, pick airport hotel return, airport location open 24 hours day downtown location just open day, nice new cars reasonable rates, need car visit rain forest, sure tour park ranger, free informative,	0	1	0	0
great stay husband spent 4 nights hotel, wonderful stay room 24th floor, great view comfortable bed.this hotel close dadeland mall miami metrorail.we access concierge lounge serves continental breakfast daily hors d'oeuvres afternoon.all great stay, defintely stay,	1	0	1	1
n't hesitate soon can-it wonderful resort wife 2 daughters 5 3 years old best vacations, planning come punta cana stay wonderful resort, non-expensive people friendly rooms comfortable food awesome options lots pools specially beach fun miss that.kids friendly kids pool facilities recomend resort, went punta cana stayed resort reading lot bad comments property decided, believe mad right selection price	0	0	1	1

paid 500 4 including meals room tips want.ocanview rooms comfortable train end enjoy resort.5 days 4 nights package looking decent nice resort,				
great hotel writing review evening staying palace hotel, great hotel, rooms tidy nice character, staff extremely helpful receptionists spent half hour phone calls place trying tokyo making feel unwelcome time, breakfast nice hotel shopping arcade restaurants fitness centre, close jr tokyo metro otemachi stations place tokyo, course imperial palace stone throw away, course using complementary internet connection room write review try massage place basement floor, massage absolutely heavenly great way tired feeling jetlag,	0	0	0	1
lovely hotel summer traveled europe month stayed wonderful hotels voted hotel santa maria novella best hotels, hotel santa maria novella beautiful, entire staff friendly extremely accommodating, rooms elegant, truly feel vacation, centrally located, train station literally street, best walk florence main attractions, given train station just steps hotel decided day trip pisa awesome addition itinerary, stayed nights hotel santa maria novella, breakfast staff warm friendly opinion just wanted wonderful stay, just lovely, happened st. john day religious holiday, coolest thing roof returning dinner enjoying fireworks, far places eat hotel staff best restaurant recommendations.we definitely return hotel santa maria novella,	0	0	0	0
needs work 6 stayed fenice palace june 17-19. booked hotel great location note not dissapointed, close duomo ponte vecchio florence great walking city anyway.i say street noise second floor rooms sleep impossible air conditioning weak noisy not acceptable prices paid, bathroom room 209 think basic miserable shower hardly room turn, know europe shower stall better flimsy curtain sticks allows water drench floor time bathe.staff generally pleasant tolerant weak attempts italian, check check painless efficient, just wish complaints ac met empathy.i think probably better hotels available trip italia no idea, travel wonderful city, stay however.ciao,	1	1	1	1
hardly better, excellent hotel, location ideal 10 min walk forbidden city, staff helpful organised tours great wall, restaurants jncluidng 24h grand cafe class added convenience, staff helpful spoke english, rooms clean quiet serviced really picky say air heating slightly efficient, definitively recommended definitively choice went beijing, note booked hotel website direct cheapest way book room,	1	1	1	1

did n't want leave stayed uma ubud week october absolutely loved, researched tripadvisor impressed consistently high reviews, wanted laidback luxury good facilities 25m pool gym yoga, certainly n't feel restricted resort not resort really like boutique hotel fabulous eating options close, favourites naughty nuris great ribs cooked roadside bbq ibu okra spelling wrong best suckling pig sit floor eat, ate mozaic half liked bit bland, make sure reserve table ouside garden really disappointing elevator type music dreadful, uma really perfect welcome pick-up airport cold towels winner fantastic friendly staff, got really ill able supply medicine, luckily knocked 24 hours, terrace room loved great terrace chilling love way staff light outdoor candles night, bathroom perfect spacious, spa nice really strongly recommend walking minutes road direction ubud bali botanical spa treatments fraction price surroundings far beautiful balinese, discovered second day sorry day, envy uma flash, spend second week friends villa canggu fabulous, difficult not enjoy bali,	0	0	0	0
westin madrid, worn decent just returned 4 night stay using starwood points room property madrid, platinum starwood member try use points nicer properties expensive especially european locations, liked best property location excellent, prado door modern museum art square, starbucks located block decent coffee pastry, spanish version dennys called vips food cheaper restaurants not interesting just filling.what did not like hotel gouge, small bottle water 7, rollaway 50 euros night cup coffee hotel 5, stay away overpriced not interesting breakfast buffet 22 euros, concierge lousy restaurant recommendations tapas food not terrible quality high price, point returning let know poor recommendation thanked day returned reason overheard giving recommendation customer, typically means getting sort kickback steer certain places eat, avoid concierge directions, rooms walls heard everyone__√á_√©_ door open close room, superior room spacious bit worn, so-so, rate room 500 euros night fled night using points stay did not mind, decent hotel stay use points absurdly overpriced prices charge accommodations, queen size bed seen better day springs tended squeak got bed, faced square kind cool constant street noise meant kept window closed.in end fine using points hotel not worth charging room,	0	1	1	1

worst hotel experience booked nonsmoking room online weeks advance stay crowne plaza downtown seattle, arrival desk staff asked consider smoking room, completely unacceptable family clearly stated no not consider smoking room mainly concern infant daughter health, particular staff member went speak quietly desk staff member agreed particular room, entered room smell cigarette smoke apparent not imagine desk staff intentionally smoking room informing, days later discovered ashtray matches checked desk discover fact smoking room intentionally not kept reserved nonsmoking room, clothes belongings baby reeked cigarette smoke.in addition appalling lack concern health wishes room not serviced night stay, asked desk manager room not serviced checked housekeeping informed not disturb tag door did not disturb, not case tag inside door entire time anxiously awaiting staff garbage dirty diapers replace towels, desk manager offered pathetic compensation complimentary hotel breakfast family exposed carcinogenic cigarette fumes days not having rooms cleaned, definitely stay crowne plaza hotel,	0	0	0	0
aka hh campomanes wonderful experience lovely contemporary hotel right heart central madrid, street opera metro stop near plaza isabell ii, great location, walking distance teatro real puerto del sol, calle arenal great shopping street neat nightclubs, rooms great individual a/c 2 comfy twin beds, no noise whatsever, rooms include minibar staellite tv internet access planned, bathrooms extremely clean hot running water, housekeeping turns sheets toiletries daily, staff friendly speak good english, good info local events tourist stops help buy tix operas shows flamencos etc.breakfast buffet included 4 types cereals fruit juice/milk/water sliced cakes bread sliced meats turkey ham cheese fresh fruits yogurt criossants, breakfast open 8am-10:30am good stint keeps tucked 2pm madrid lunchtime, stayed 4 nights christmas break loved, comes price 99 euro night not cheapest place, definitely really good value money, gladly recommend, tip want laundry coin operated laundromat lavamat calle la cruz close puerto del sol, hotel dry-clean clothes not worth price, not tell coin op laundry obvious reasons,	0	1	0	1
castle inn great value booked castle inn direct got informative e-mail set standard 4 night stay, room standard motel room clean decoration good equipped microwave fridge/freezer, lady desk gave lots information not checked, helpful stay, overall great value agood place stay staying time san francisco,	1	1	0	1

worth, overall good stay, good points rooms cleaning continental breakfasts location swimming pool holiday inn door reception, poor points lifts slow coffee facilities room, coffee percolator en suite, coffee/milk/sugar maybe 2 cups, overcome tea bags milk sugar breakfast, not coffee rooms right no homeless seen area, excellent location fishermans transport parts city.overall happy,	0	0	0	0
good resort just came week vacation caribe club princess punta cana, sure value, beach gorgeous complex just big pools clean beautiful staff friendly eager good service, caribe complex tropical princess enjoy infrastructure hotels, pools 7 remember correctly la carte restaurants big buffet restaurant long beach, food good buffet, wide selection food, did n't enjoy el pilon dominican la carte restaurant just matter taste, rooms ok. little little ants surprise discover big cockroach bathroom, crawled bathroom drain, stinky smell present entrance building, n't know came gross, great resort apart little problems, grounds beautiful beach fabulous staff great pools excellent food good, drink pool bar talk francisco real great guy,	0	0	0	0
stay hotel ny stayed couple nights hotel, right heart times square, previous review said rooms small immaculately clean serviced quite, little local knowledge ask room no noise street unlike hotels stayed previously.the staff helpful immpecable manners worked good nature.the glass wine provided lounge stayed 3 hours offer welcome, crackers cheese grapes complimented choice red white fizzy wine, oh coffee tea cookies available 24hrs.we stay no hesitation recommending hotel.kr,	1	0	0	0
highly recommend hotel stayed superior king room view overlooking boylston st. historic hotel bay area block away amtrak station subway station nearby, lenox fully restored beautiful lobby plush furnishings, room spacious appointed quality furnishings walk wardrobe attractive bathroom quality fittings accessories, king bed extremely comfortable, hotel quality restaurant separate bar irish pub hotel, hotel numerous restaurants bars boylston nearby newbury sts, staff helpful friendly, booked hotel historic hotels america website advance purchase rate, make point return lenox visiting boston future, highly recommended,	1	1	1	1

good hotel, stayed 3 nights january, room warm windows opened ventilation, small kitchen great boiling water tea storing beer things fridge gas stove oven immaculate, fancied cook, room big decor good bit dodgy places not spoiled, bathroom nice water pressure shower bit toooo, mini bar big tv loads amenities business people cd player built alarm clock fax printer internet access, stayed 32nd floor reasonable view just faciniating watching new york hotel location n't bad prepared walking did n't bother think best way explore healthy, blockheads mexican restaurant block right hotel great place youngsters llike drunk eat nice food, come staff friendly helpful did n't bother, barking dog good food drinks smokers like n't bad going outside sneaky cigarette temperature, great, 6 stars,	1	1	0	1

Table 33.

Case study results (TF-IDF-SVM, Employee data: Glassdoor reviews)

Note. The red colored numbers indicate wrong identifications.

text	control mutuality	commitment	trust	satisfaction
Culture , good people, team oriented	0	0	1	1
Remote opportunities for people if they really want that. My non-mgmt coworkers were great people.	1	0	1	0
Cons: Wear and tear on your vehicle: DoorDash drivers use their own cars to make deliveries, which can put a lot of wear and tear on your vehicle. This means you may need to pay for repairs or maintenance more frequently. Risk of accidents: Driving around all day can be dangerous, especially in busy areas or during bad weather. Inconsistent earnings: The amount you can earn with DoorDash can vary depending on the number of deliveries you make and the tips you receive. This can make it difficult to rely on DoorDash as a steady source of income.	0	0	0	0
This is a very toxic work environment. Leadership is a joke - seriously. I think they all just chill and grocery shop during the day then complain that they have too many people to Manager (usually 4-5 ppl). The bar is incredibly low for leadership at lyft. I've never seen anything like it. Nothing is expected of them and when cuts come - they'll cut the entire workforce (which has happened) before they let the lousy managers go. They are NOT subject matter experts and are not expected to know anything.	1	1	1	0
Smart colleagues Half day Fridays Reasonable compensation	0	1	0	0
great place unlimited pto scope for growth	1	1	0	0
1. Once you learn everything this job is fairly easy.	0	1	1	0
cons	1	0	1	1
Great benefits. Training is well planned and the managers work with you when it comes to learning the product. Days can be a bit challenging when it comes to making call 80 dials minimum plus 2 hours of talk time are the minimum.	0	1	0	0
questionable decisions by management make it hard to want to stay	1	1	0	0
-Open PTO -Performance Based Bonuses - Work-life balance	1	1	1	1
-Decent pay for an entry level position. -Good opportunity to learn the business	1	1	0	1

Hybrid work from home schedule Catered lunches Open PTO	1	0	0	0
great wlb great community and culture	1	0	1	1
Really smart people, a lot of opportunity for growth, always encouraged to be innovative, think big, and create something new. Competitive salary and benefits with other major tech companies. 100% self motivating work environment. No dress code and 4 legged friends are welcome.	0	0	0	1
Changes come quickly and organizational changes aren't always very transparent, so I've found I need to be very flexible and go with the flow	1	0	1	1
Lack of transparency, and goals not attainable	0	0	0	0
Easy. Flexible schedule. Independent. Relaxing	0	1	0	0
Work - life balance & salary	0	1	0	1
Toxic work environment within customer support team, constant leadership and org changes, unstable US culture from multiple layoffs, limited growth in the US due to restructuring talent to Mexico City	0	0	0	1
Every growing young business has their challenges, this one is no different, despite the size of the business.	0	0	0	0
Flexibility, involved in the community, instant payouts, self-employed, work at your own pace/risk	0	0	0	0
Promotions are somewhat political, oriented to how visible you are to upper management. Many moving updates across products can be difficult to keep up with - nothing incredibly out of the usual among similar tech companies. Confidence in C-Suite to do the right thing.	0	0	1	0
friendship and good environment with members	1	0	1	1
Instant access to pay and cashout	1	0	0	1
Best delivery experience I've had!	1	0	1	1
Great community Great communication Great productivity	1	1	1	1
Miscommunication. No one is on the same page in terms of what information is needed, leaving confusion for what process to take or information to give to viewers calling. Horrible Bonuses.	1	1	1	0
None that I can think of.	1	0	1	1
I genuinely love working here. I have been in tech for over 15 years and can confidently say this is the best of the best. I have worked here now for almost a year and have had the same experience consistently. I feel supported, I am not burnt out, and mangers lead with EMPATHY.	0	0	0	0

- Culture is like nothing I've experienced before. Co-workers truly assume good intentions, I haven't encountered any snarky or pretentious people. If you make a mistake nobody belittles you for it. - The mission is really important. Hiring seems broken everywhere, and Glassdoor is genuinely trying to fix it. -Benefits are quite good, with great health insurance family plans and wellness options like Classpass - Work life balance actually means something. I generally start and end my day at reasonable times and rarely need to work beyond that. Unlimited PTO and ability to actually use it is also great -Transparent leadership	1	0	1	1
- Amazing people to work with. Seriously, they're the smartest, most humble and passionate people you'll come across. Just about everyone believes in what we're doing. - Glassdoor genuinely care about the whole person. This is the first employer I've had that I haven't had anxiety about having to leave to take care of my child, or take care of myself when I'm sick. It's like they really understand that there is more to employee lives than Glassdoor work. - No boys club or high school style politics here. You do your work well and contribute to the team/company? You move up. Backstabbing or any other oppressive types of behavior are unacceptable here. This place will literally change how you feel about work.	1	0	0	0
Culture Management Leadership Reviews People	0	0	0	1
Good start up company out of college	0	0	0	0
Need to work during peak meal hours	1	0	1	1
The tech stack is improving but there are still legacy apps to deal with. It felt the compensation wasn't very competitive	1	1	1	1
Self-driven, tech based position with flexible hours and ability to make your own paycheck.	0	0	1	1
Great company that wants to help you succeed	0	0	0	0
Ok pay better than most of the other delivery apps i used	1	0	1	0
Fast learning and new technology	1	0	1	1
I'm starting to yhink there isn't one other than you can make some money on the fly other than that can not think of one	0	1	0	0
Processes and tools change frequently which makes it tough	1	0	1	1
Still runs like a start-up and processes are not in place to streamline and scale up as efficiently as they should be at this stage. I feel like I'm doing the job of 3 people.	0	0	0	0
Having someone in your car. Miles on your car.	1	1	1	1
Ever changing quota that could be surprising then difficult to manage	0	1	1	0

Get paid immediately. Can work on your own schedule. Nice people.	0	0	0	0
As the culture transitioned to becoming more corporate: wage gaps, not family friendly, poor management, employees not valued, minimal opportunity for promotion.	0	0	0	0
Very long hours for not enough pay	1	1	1	1
Door Dash is a wonderful company to work for. I love many things about being a Dasher. Mainly I love the service I provide for individuals who may be disabled, busy, at work, or just relaxing. I truly take pride in being there for others by lifting burdens when I can. Being a Dashing is also a wonderful experience, because it allows me to interact with so many different individuals each day, which gives me the opportunity to put smiles on many faces daily. It's like I'm going on an adventure each day. For example, I took a few days off to take care of my car, and my new challenge for these next 3 days are to see how close I can get to my weekly financial goal in just 3 days as opposed to 7. Another aspect I love, is the flexibility. I'm building a nonprofit organization currently and mental, physical and spiritual wellness are very important to me in my personal life. Door Dash allows me the flexibility to make a schedule that works best for me, one that enables me to take care of my living expenses, work in my nonprofit and stay mentally, physically and spiritually well. There are many more reasons why I love being a Dasher, however I'm working on over communicating 😂😂😂😂	0	0	0	0
at 12.50 an hour you'd be better off getting a job at McDonalds. You can always do Uber after your shift if you need to make extra.	1	1	0	0

Figure 1.
Study Design

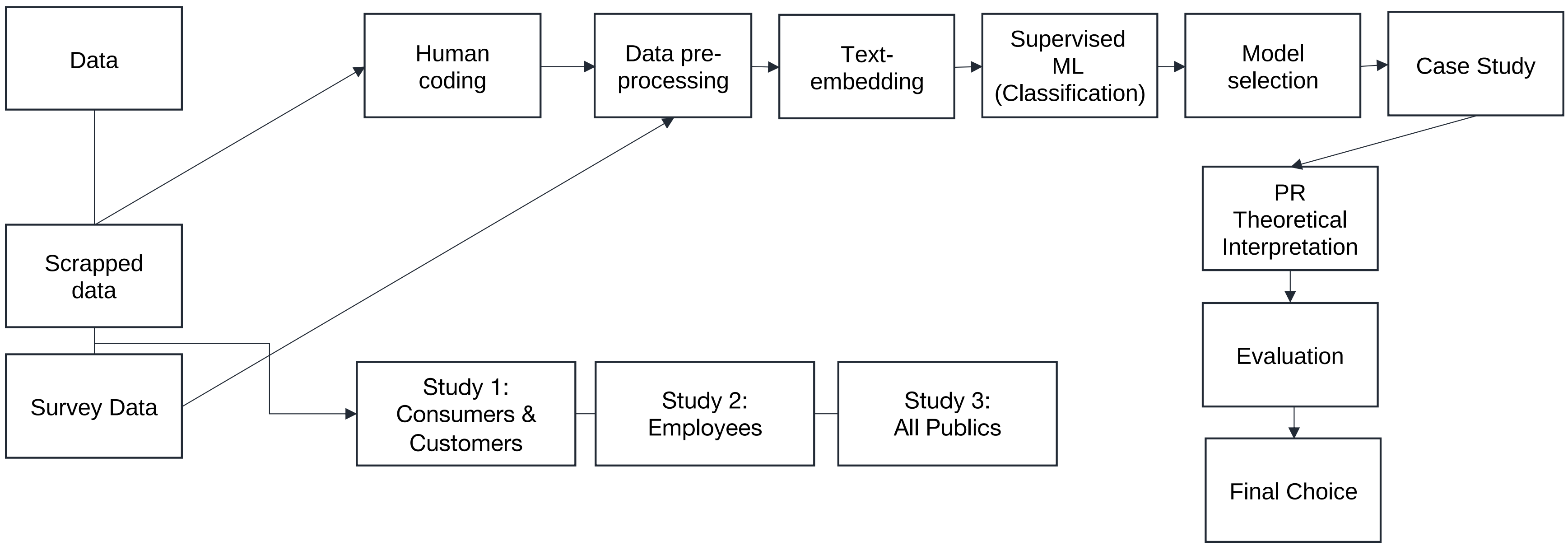


Figure 2.
Accuracy Results Summary

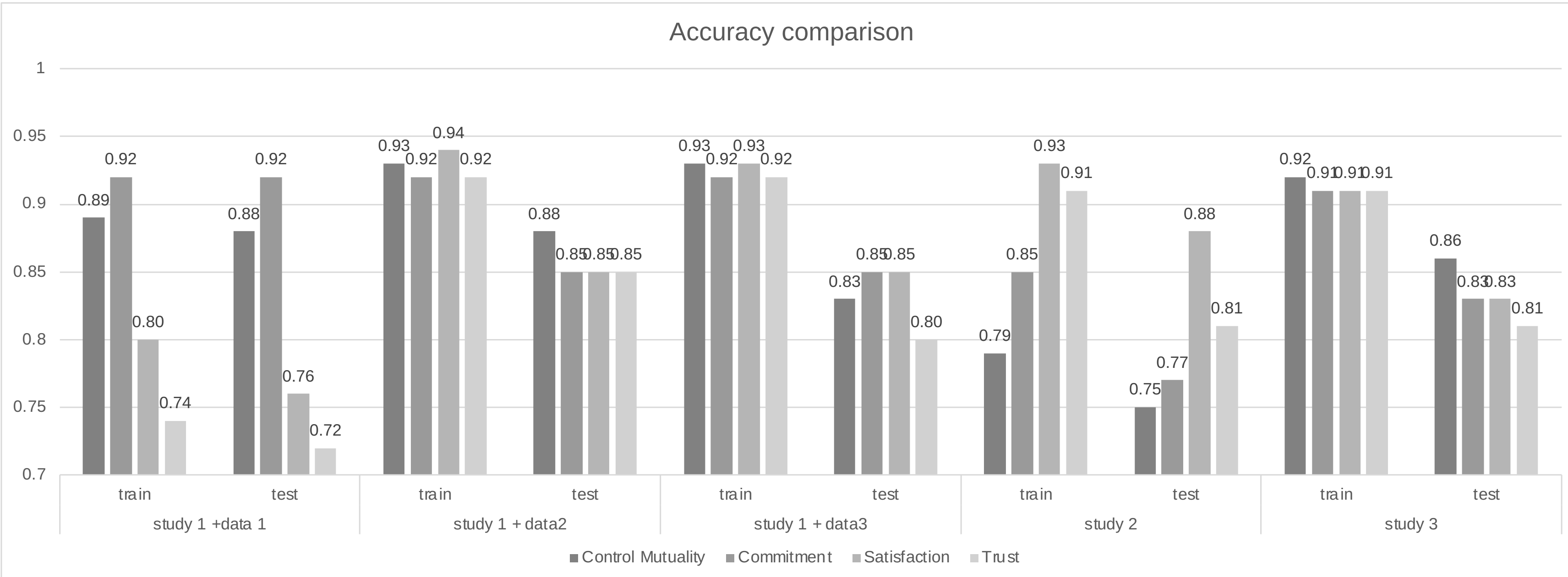


Figure 3.

BERT vs. TF-IDF & SVM vs. KNN (Study1- Amazon data)

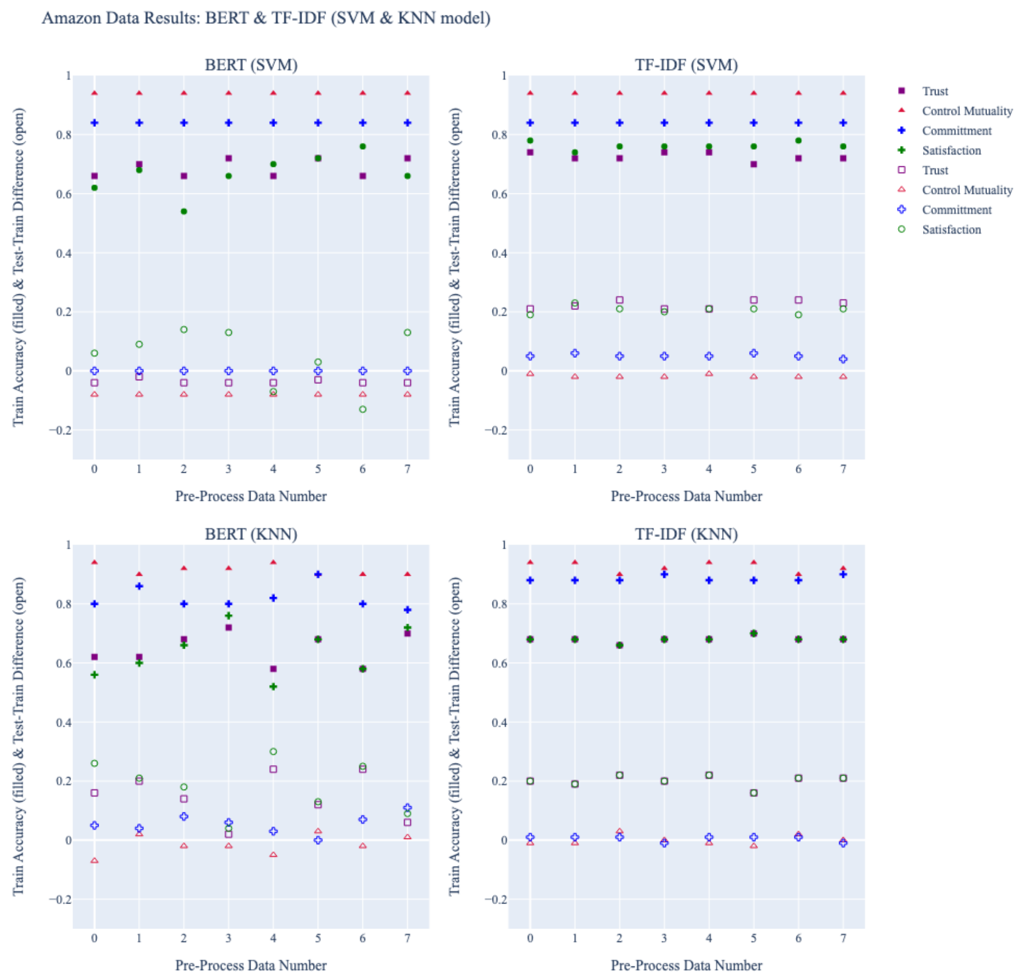


Figure 4.

BERT vs. TF-IDF & SVM vs. KNN (Study1- Survey data)

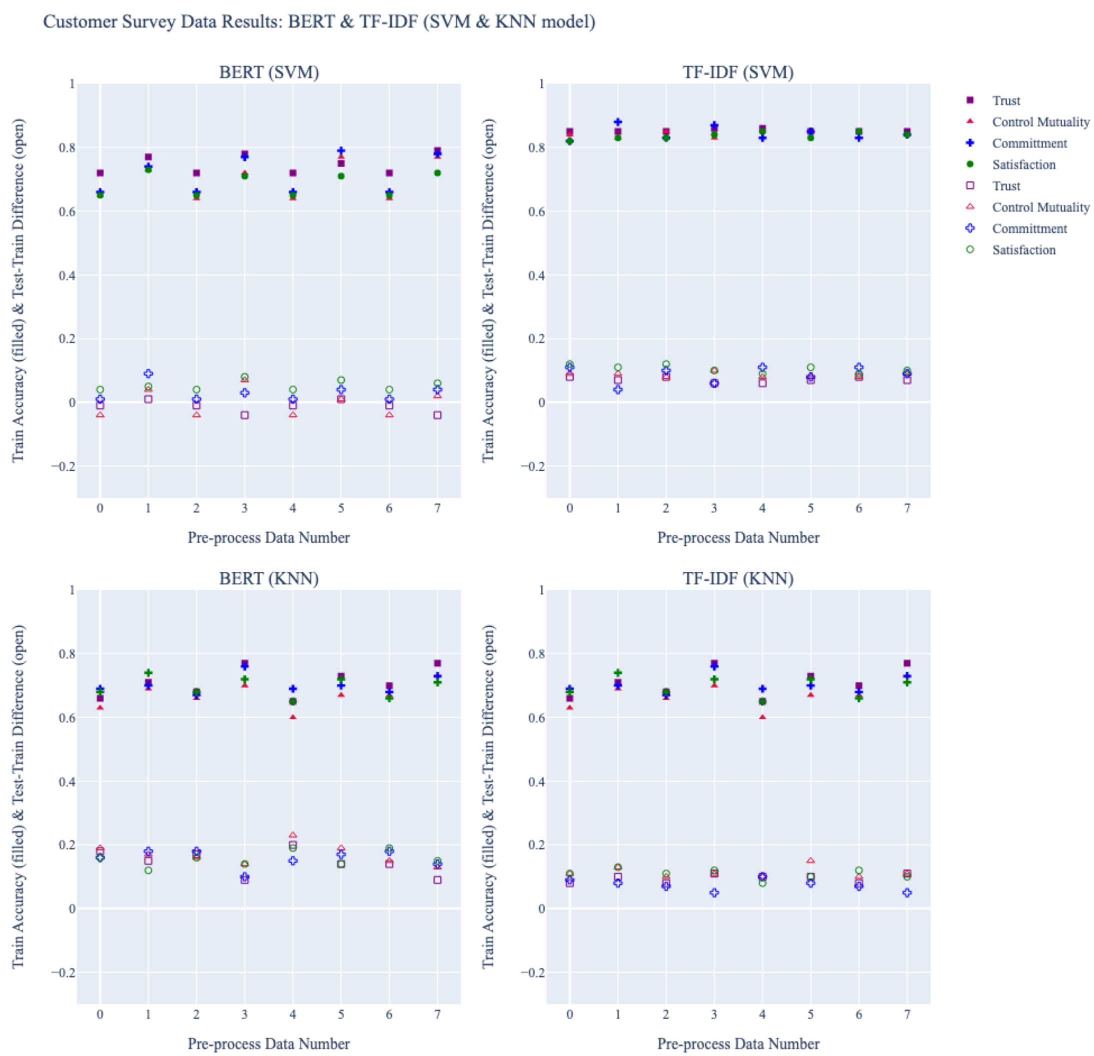


Figure 5.

BERT vs. TF-IDF & SVM vs. KNN (Study1- Amazon + Survey data)

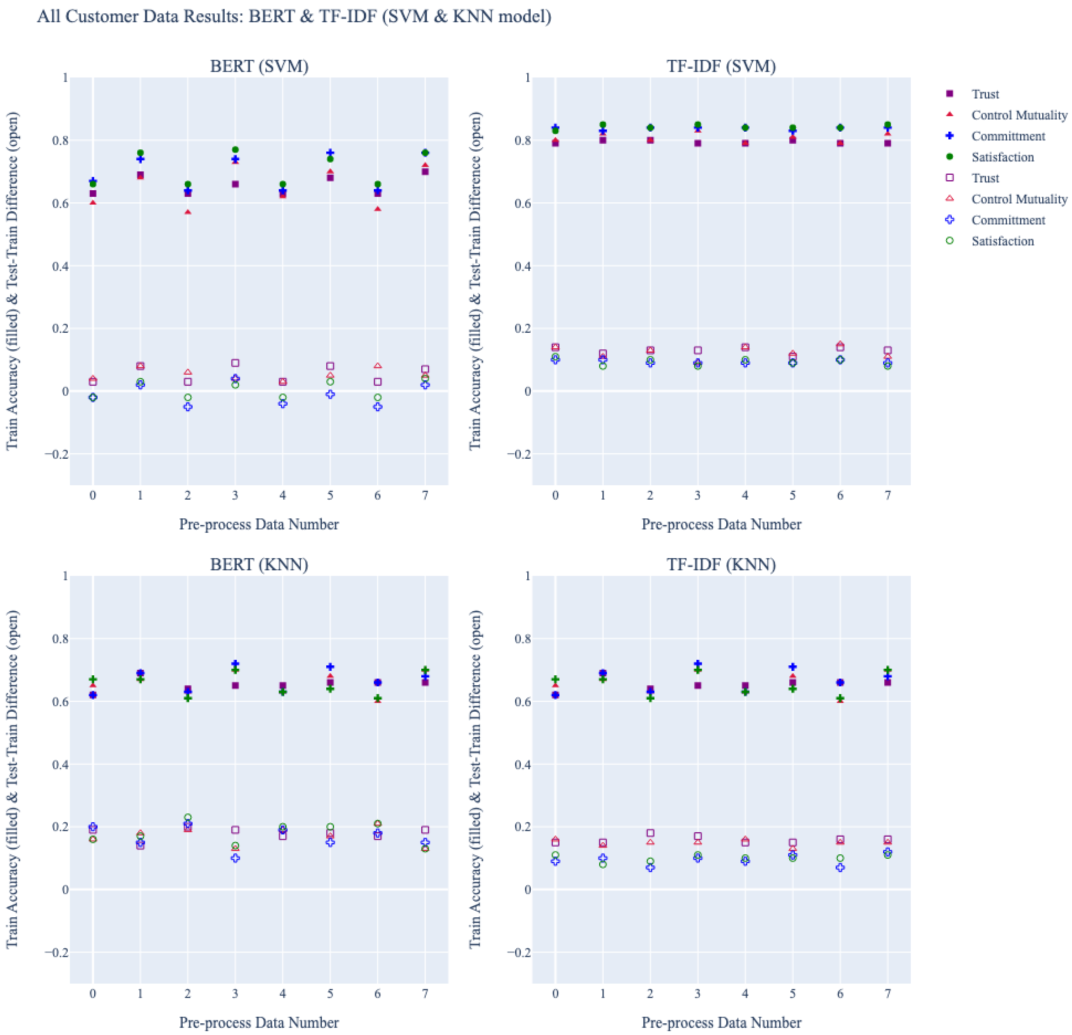


Figure 6.

BERT vs. TF-IDF & SVM vs. KNN (Study2- Survey data)

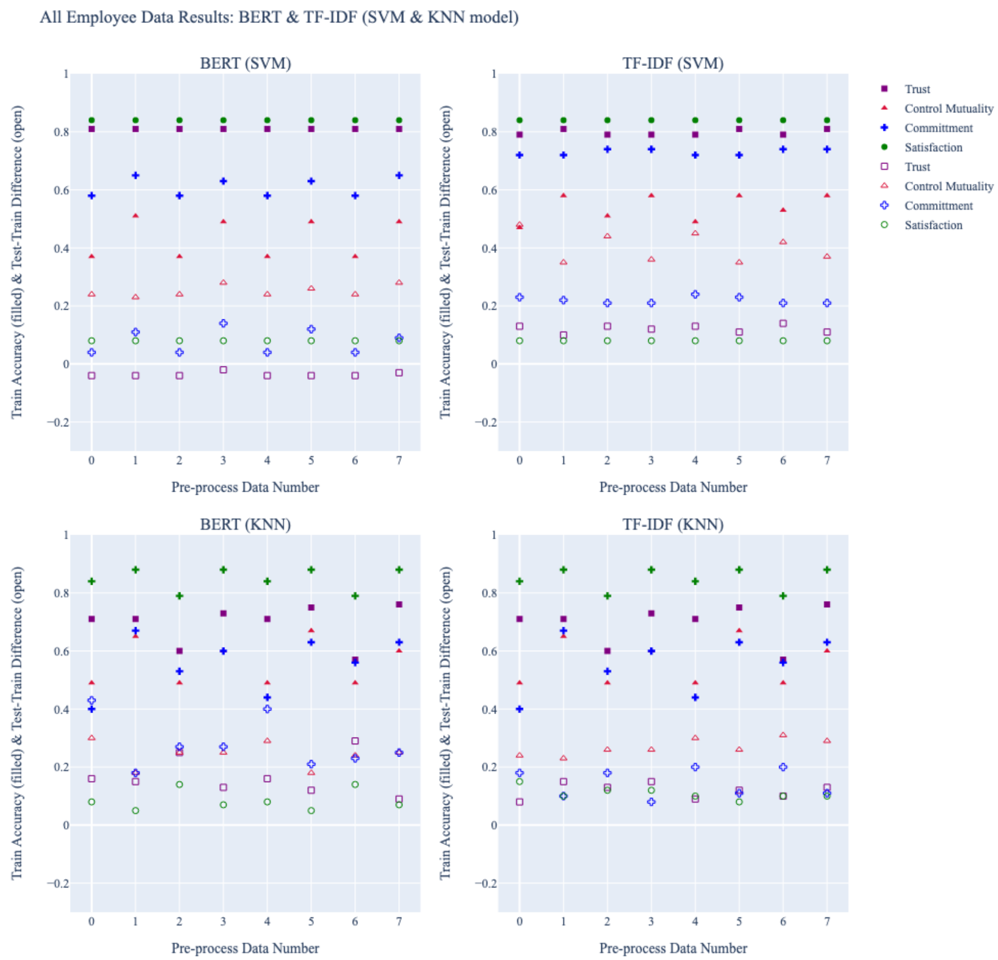


Figure 7.

BERT vs. TF-IDF & SVM vs. KNN (Study2- All data)

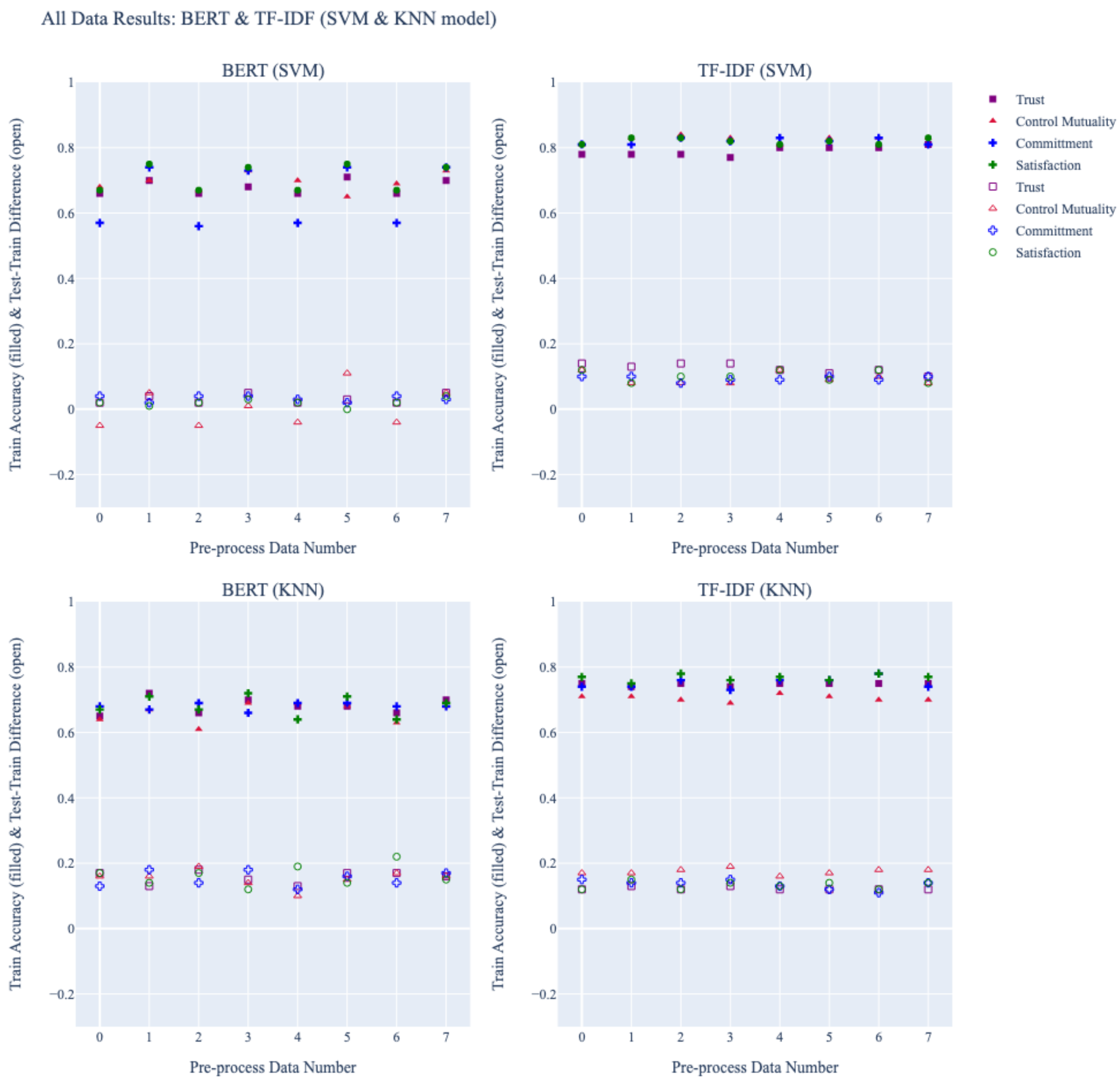
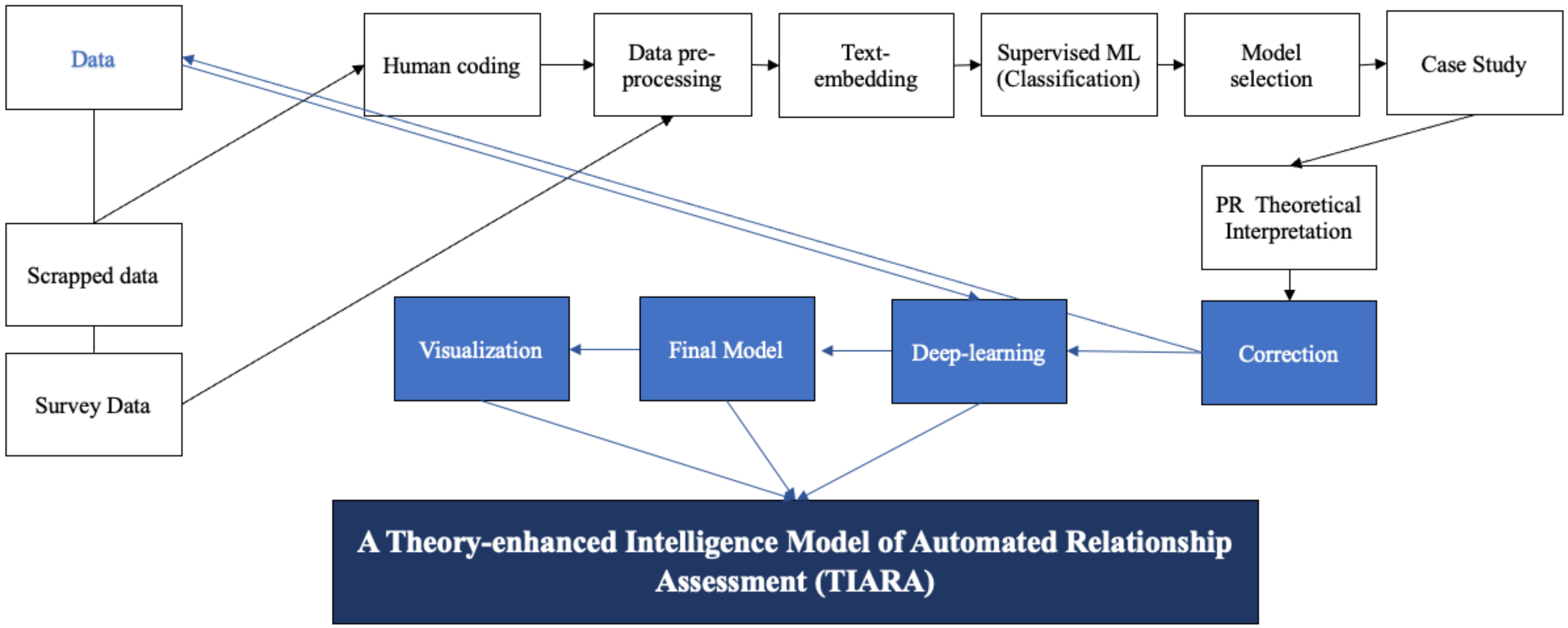


Figure 8.
Next steps for the TIARA



Appendix A-1. Survey Questionnaire (Survey 1, October 2021)
Note. The original questionnaires are available online:
https://www.dropbox.com/s/ou28ovxqrj4xgc6/Survey%201%20Qs.pdf?dl=0

Appendix A-2. Survey Questionnaire (Survey 2, May 2022)
Note. The original questionnaires are available online:
https://www.dropbox.com/s/ou28ovxqrj4xgc6/Survey%201%20Qs.pdf?dl=0

Appendix A-3-a. Survey Questionnaire (Survey 3 -Customer, October 2023)
Note. The original questionnaires are available online:
- Customer (Best): https://www.dropbox.com/s/byilmjj7iiwxxc7/Survey%203%20Qs_Customer_Best.pdf?dl=0
- Customer (Worst): https://www.dropbox.com/s/bgutfehwfjadivf/Survey%203%20Qs_Customer_Worst.pdf?dl=0

Appendix A-3-b. Survey Questionnaire (Survey 3 -Employee October 2023)
Note. The original questionnaires are available online:
- Employee (Best): https://www.dropbox.com/s/f3p8hjkj3rfjodn/Survey%203%20Qs_Employee%20Best%20.pdf?dl=0
- Employee (Worst): https://www.dropbox.com/s/1e9yd8ets2blsa2/Survey%203%20Qs_Employee%20Worst.pdf?dl=0

Appendix B. Emotional labeling outcomes
Note. The original 13,053 outcomes are available online:
- Positive: https://docs.google.com/spreadsheets/d/1zx4HDP4isvr6QwOXE5Y6zWSFi9ina9-z9CLvnYNH5eo/edit?usp=sharing
- Negative: https://docs.google.com/spreadsheets/d/1Vcu8y_MGcpS_EsRjfqfuNwhRGToSEV9PltEiZeGF3sQ/edit?usp=sharing

Appendix C-1. Case study outcomes – BERT-SVM model: TripAdvisor (samples)

Note. The original 20,491 outcomes are available online:
https://docs.google.com/spreadsheets/d/1Y7u5NL8pmYTla-BHAIaXV5fE0CHt61_5OQ-CR-haaOY/edit?usp=sharing

text	control mutuality	commitment	trust	satisfaction
nice hotel expensive parking got good deal stay hotel anniversary, arrived late evening took advice previous reviews did valet parking, check quick easy, little disappointed non-existent view room room clean nice size, bed comfortable woke stiff neck high pillows, not soundproof like heard music room night morning loud bangs doors opening closing hear people talking hallway, maybe just noisy neighbors, aveda bath products nice, did not goldfish stay nice touch taken advantage staying longer, location great walking distance shopping, overall nice experience having pay 40 parking night,	1	1	1	1
ok nothing special charge diamond member hilton decided chain shot 20th anniversary seattle, start booked suite paid extra website description not, suite bedroom bathroom standard hotel room, took printed reservation desk showed said things like tv couch ect desk clerk told oh mixed suites description kimpton website sorry free breakfast, got kidding, embassy suits sitting room bathroom bedroom unlike kimpton calls suite, 5 day stay offer correct false advertising, send kimpton preferred guest website email asking failure provide suite advertised website reservation description furnished hard copy reservation printout website desk manager duty did not reply solution, send email trip guest survey did not follow email mail, guess tell concerned guest.the staff ranged indifferent not helpful, asked desk good breakfast spots neighborhood hood told no hotels, gee best breakfast spots seattle 1/2 block away convenient hotel does not know exist, arrived late night 11 pm inside run bellman busy chating cell phone help bags.prior arrival emailed hotel inform 20th anniversary half really picky wanted make sure good, got nice email saying like deliver bottle champagne chocolate covered strawberries room arrival celebrate, told needed foam pillows, arrival no champagne strawberries no foam pillows great room view alley high rise building good not better housekeeping staff cleaner room property, impressed left morning shopping room got short trips 2 hours, beds comfortable.not good ac-heat control 4 x 4 inch screen bring green shine directly	0	0	1	0

eyes night, light sensitive tape controls.this not 4 start hotel clean business hotel super high rates, better chain hotels seattle,				
nice rooms not 4* experience hotel monaco seattle good hotel n't 4* level.positives large bathroom mediterranean suite comfortable bed pillowsattentive housekeeping staffnegatives ac unit malfunctioned stay desk disorganized, missed 3 separate wakeup calls, concierge busy hard touch, did n't provide guidance special requests.tv hard use ipod sound dock suite non functioning. decided book mediterranean suite 3 night weekend stay 1st choice rest party filled, comparison w spent 45 night larger square footage room great soaking tub whirlpool jets nice shower.before stay hotel arrange car service price 53 tip reasonable driver waiting arrival.checkin easy downside room picked 2 person jacuzi tub no bath accessories salts bubble bath did n't stay, night got 12/1a checked voucher bottle champagne nice gesture fish waiting room, impression room huge open space felt room big, tv far away bed chore change channel, ipod dock broken disappointing.in morning way asked desk check thermostat said 65f 74 2 degrees warm try cover face night bright blue light kept, got room night no, 1st drop desk, called maintainence came look thermostat told play settings happy digital box wo n't work, asked wakeup 10am morning did n't happen, called later 6pm nap wakeup forgot, 10am wakeup morning yep forgotten.the bathroom facilities great room surprised room sold whirlpool bath tub n't bath amenities, great relax water jets going,	1	1	1	1
unique, great stay, wonderful time hotel monaco, location excellent short stroll main downtown shopping area, pet friendly room showed no signs animal hair smells, monaco suite sleeping area big striped curtains pulled closed nice touch felt cosy, goldfish named brandi enjoyed, did n't partake free wine coffee/tea service lobby thought great feature, great staff friendly, free wireless internet hotel worked suite 2 laptops, decor lovely eclectic mix pattens color palatte, animal print bathrobes feel like rock stars, nice did n't look like sterile chain hotel hotel personality excellent stay,	1	1	1	1

Appendix C-2. Case study outcomes – TF-IDF-SVM model: TripAdvisor (samples)

Note. The original 20,491 outcomes are available online:
https://docs.google.com/spreadsheets/d/1fbFBWxGJuWgA7QVxJHjkcgDEqMU8m2LDUiHpK7_sg1s/edit?usp=sharing

text	control mutuality	commitment	trust	satisfaction
nice hotel expensive parking got good deal stay hotel anniversary, arrived late evening took advice previous reviews did valet parking, check quick easy, little disappointed non-existent view room room clean nice size, bed comfortable woke stiff neck high pillows, not soundproof like heard music room night morning loud bangs doors opening closing hear people talking hallway, maybe just noisy neighbors, aveda bath products nice, did not goldfish stay nice touch taken advantage staying longer, location great walking distance shopping, overall nice experience having pay 40 parking night,	0	0	0	1
ok nothing special charge diamond member hilton decided chain shot 20th anniversary seattle, start booked suite paid extra website description not, suite bedroom bathroom standard hotel room, took printed reservation desk showed said things like tv couch ect desk clerk told oh mixed suites description kimpton website sorry free breakfast, got kidding, embassy suits sitting room bathroom bedroom unlike kimpton calls suite, 5 day stay offer correct false advertising, send kimpton preferred guest website email asking failure provide suite advertised website reservation description furnished hard copy reservation printout website desk manager duty did not reply solution, send email trip guest survey did not follow email mail, guess tell concerned guest.the staff ranged indifferent not helpful, asked desk good breakfast spots neighborhood hood told no hotels, gee best breakfast spots seattle 1/2 block away convenient hotel does not know exist, arrived late night 11 pm inside run bellman busy chating cell phone help bags.prior arrival emailed hotel inform 20th anniversary half really picky wanted make sure good, got nice email saying like deliver bottle champagne chocolate covered strawberries room arrival celebrate, told needed foam pillows, arrival no champagne strawberries no foam pillows great room view alley high rise building good not better housekeeping staff cleaner room property, impressed left morning shopping room got short trips 2 hours, beds comfortable.not good ac-heat control 4 x 4 inch screen bring green shine directly	0	0	0	0

eyes night, light sensitive tape controls.this not 4 start hotel clean business hotel super high rates, better chain hotels seattle,				
nice rooms not 4* experience hotel monaco seattle good hotel n't 4* level.positives large bathroom mediterranean suite comfortable bed pillowsattentive housekeeping staffnegatives ac unit malfunctioned stay desk disorganized, missed 3 separate wakeup calls, concierge busy hard touch, did n't provide guidance special requests.tv hard use ipod sound dock suite non functioning. decided book mediterranean suite 3 night weekend stay 1st choice rest party filled, comparison w spent 45 night larger square footage room great soaking tub whirlpool jets nice shower.before stay hotel arrange car service price 53 tip reasonable driver waiting arrival.checkin easy downside room picked 2 person jacuzi tub no bath accessories salts bubble bath did n't stay, night got 12/1a checked voucher bottle champagne nice gesture fish waiting room, impression room huge open space felt room big, tv far away bed chore change channel, ipod dock broken disappointing.in morning way asked desk check thermostat said 65f 74 2 degrees warm try cover face night bright blue light kept, got room night no, 1st drop desk, called maintainence came look thermostat told play settings happy digital box wo n't work, asked wakeup 10am morning did n't happen, called later 6pm nap wakeup forgot, 10am wakeup morning yep forgotten.the bathroom facilities great room surprised room sold whirlpool bath tub n't bath amenities, great relax water jets going,	0	1	0	1
unique, great stay, wonderful time hotel monaco, location excellent short stroll main downtown shopping area, pet friendly room showed no signs animal hair smells, monaco suite sleeping area big striped curtains pulled closed nice touch felt cosy, goldfish named brandi enjoyed, did n't partake free wine coffee/tea service lobby thought great feature, great staff friendly, free wireless internet hotel worked suite 2 laptops, decor lovely eclectic mix pattens color palatte, animal print bathrobes feel like rock stars, nice did n't look like sterile chain hotel hotel personality excellent stay,	0	1	1	1

Appendix D-1. Case study outcomes – TF-IDF-SVM model: TripAdvisor (samples)

Note: The original 899 outcomes are available online:
https://docs.google.com/spreadsheets/d/1s8zfPrED8XxkmO71er_LkRnDsgv_nzb4VROIs3l_GF4/edit?usp=sharing

text	control mutuality	commitment	trust	satisfaction
I've worked at several tech companies as a PM, both large and small, and Uber has by far been my favorite. It's a great blend of getting the job security, pay, and benefits of an established tech company with the fast pace and innovative focus that you'd more so expect at a startup. If you want to hone your craft as a PM and get things done, Uber is the place for you!!!	1	1	0	0
I worked part time as a 5 star Uber driver to make extra money in my spare time. I like driving and I am a people person.	1	1	0	0
I love being able to work around truly passionate people who are ready to change the world. The culture is great, free snacks and food is big plus. I had the opportunity to take ownership of projects within my first month	1	1	0	1
Flexibility is great, and they do reasonable effort to make the app function well with frequent new features than lyft	0	1	1	1
Great compensation, benefits, perks, e.g., free food. Great brand with global recognition. Lots of smart and capable people to work with.	1	1	1	1
Flexible schedule Opportunity to earn extra income Easy to get started No passengers Additional perks	1	0	0	0
Easy Cash - Easy Application Process - Easy To Use	0	0	0	0
Meet people from different walks of life. Solid promotions, job helped me learn my way around the city.	1	0	0	1
You get to go a lot of places around the city and meet a lot of interesting people. I only work in the day so I can't comment on the night crowd.	1	1	0	0

Benefits are amazing Work life balance is highly encouraged and supported Highly intelligent individuals lift you up DEI is also one of the best I have seen in a company, especially this big Salary is at par with tech companies I am passionate about our values	1	1	1	1
They’ve helped me a llt	1	1	1	0
Decent place to transition to a corporate role and learn the basics Flexible WFH policy	1	0	0	1
Earn money on your own schedule	0	0	0	0
You make your own schedule, so you can work anytime/anywhere in the United States	0	1	0	0
You get boost times to earn extra money. You get stats based on previous 4 weeks what your day will look like and at what times are the busiest. Pay is decent.	1	0	0	0
The company's values are strong and HR works hard to make the environment positive, but it also depends on each individual being aware of their role in this commitment.	1	1	1	0
The job offers self scheduling.	1	1	1	0
- Benefits - job opportunities - culture	1	1	0	0
Good benefits Laid back dress code Inclusive hiring	1	1	1	1
They have great health benefits, if you got in early you were lucky enough to get stock in Uber. PTO is very flexible specially if you have kids. Overall their benefits in the COE are amazing! They have programs and courses available to advance skills. They fully pay for your college education at Arizona state.	0	1	1	0
Working at Uber is amazing, you get to choose your own hours and have a lot of control in how much you work and therefor make.	0	1	0	0
You can deliver whenever Good tips	1	1	0	0
Lots of perks and mgmt is experienced	0	1	1	1
Great pay, Great experience with business and customer service.	1	1	1	1
Benefits Hybrid model Lunches Learning opportunities	1	1	1	1
Tons of change for the better since the earlier days.	0	1	1	1
Good place to work Good place to work	0	1	0	0

Appendix D-2. Case study outcomes – TF-IDF-SVM model: Glassdoor (samples)

Note: The original 899 outcomes are available online:
https://docs.google.com/spreadsheets/d/1qgibUVqY2xKqgM0tjXBjhShc7qox2DyaHzCnxEboOc0/edit?usp=sharing

text	control mutuality	commitment	trust	satisfaction
I've worked at several tech companies as a PM, both large and small, and Uber has by far been my favorite. It's a great blend of getting the job security, pay, and benefits of an established tech company with the fast pace and innovative focus that you'd more so expect at a startup. If you want to hone your craft as a PM and get things done, Uber is the place for you!!!	1	1	1	1
I worked part time as a 5 star Uber driver to make extra money in my spare time. I like driving and I am a people person.	0	1	0	1
I love being able to work around truly passionate people who are ready to change the world. The culture is great, free snacks and food is big plus. I had the opportunity to take ownership of projects within my first month	1	1	1	1
Flexibility is great, and they do reasonable effort to make the app function well with frequent new features than lyft	1	1	1	1
Great compensation, benefits, perks, e.g., free food. Great brand with global recognition. Lots of smart and capable people to work with.	1	1	1	1
Flexible schedule Opportunity to earn extra income Easy to get started No passengers Additional perks	0	0	0	0
Easy Cash - Easy Application Process - Easy To Use	0	0	1	0
Meet people from different walks of life. Solid promotions, job helped me learn my way around the city.	1	1	1	1
You get to go a lot of places around the city and meet a lot of interesting people. I only work in the day so I can't comment on the night crowd.	0	1	0	0

Benefits are amazing Work life balance is highly encouraged and supported Highly intelligent individuals lift you up DEI is also one of the best I have seen in a company, especially this big Salary is at par with tech companies I am passionate about our values	1	1	1	0
They've helped me a llt	1	0	1	1
Decent place to transition to a corporate role and learn the basics Flexible WFH policy	1	1	0	0
Earn money on your own schedule	0	1	0	0
You make your own schedule, so you can work anytime/anywhere in the United States	0	1	1	0
You get boost times to earn extra money. You get stats based on previous 4 weeks what your day will look like and at what times are the busiest. Pay is decent.	0	1	0	0
The company's values are strong and HR works hard to make the environment positive, but it also depends on each individual being aware of their role in this commitment.	1	1	1	1
The job offers self scheduling.	1	1	1	0
- Benefits - job opportunities - culture	0	1	0	0
Good benefits Laid back dress code Inclusive hiring	0	1	1	1
They have great health benefits, if you got in early you were lucky enough to get stock in Uber. PTO is very flexible specially if you have kids. Overall their benefits in the COE are amazing! They have programs and courses available to advance skills. They fully pay for your college education at Arizona state.	1	1	0	0
Working at Uber is amazing, you get to choose your own hours and have a lot of control in how much you work and therefor make.	0	1	0	0
You can deliver whenever Good tips	1	1	1	0
Lots of perks and mgmt is experienced	1	1	1	1
Great pay, Great experience with business and customer service.	1	1	1	1
Benefits Hybrid model Lunches Learning opportunities	1	1	0	1
Tons of change for the better since the earlier days.	1	1	1	0

Good place to work Good place to work	1	1	0	0