DEVELOPING PREDICTIVE MODELS FOR FUEL CONSUMPTION AND MAINTENANCE COST USING EQUIPMENT FLEET DATA

By

ARPAN CHAUDHARY

Bachelor of Technology in Civil Engineering

Galgotias University

Greater Noida, India

2015

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE
July, 2019

DEVELOPING PREDICTIVE MODELS FOR FUEL

CONSUMPTION AND MAINTENANCE COST

USING EQUIPMENT FLEET DATA

Thesis Approved:

Dr. Yongwei Shan

Thesis Adviser

Dr. Gouranga Banik

Dr. Joshua Li

ACKNOWLEDGMENTS

First of all, I would like to thank my family and friends for their love and support throughout my life. Secondly, I would like to thank the Oklahoma Department of Transportation for providing me the data to conduct this research. Lastly, I would like to thank my advisor, Dr. Yongwei Shan, for all his guidance and support throughout the work.

Name: ARPAN CHAUDHARY

Date of Degree: JULY, 2019

Title of Study: DEVELOPING PREDICTIVE MODELS FOR FUEL CONSUMPTION AND MAINTENANCE COST USING EQUIPMENT FLEET DATA

Major Field: CIVIL ENGINEERING

Abstract: The state departments of transportation (DOTs) possess a big construction equipment fleet that is engaged in various highway maintenance and repair activities all over the state. Fleet managers are called upon to give the budget estimates required to keep the equipment functioning throughout the year. These decisions are not easy to make as DOTs manage a big equipment fleet from pickup cabs to big size motor graders etc. This decision-making process could be improved by employing DATA MINING techniques on the equipment management data available with the DOTs. This study utilized the construction equipment data provided by the Oklahoma Department of Transportation (ODOT) and applied Multiple Linear Regression (MLR) to develop predictive models for fuel consumption and maintenance cost. The dataset was divided into two parts based on the operational charge type, i.e. equipment charged for operation by dollar/hour and equipment charged for operation by dollar/mile. In a total of the data from 2000 pieces of equipment was analyzed in this research. Four best models were selected based on the smallest average squared error (ASE) value. Apart from operational data, the model development utilized information such as equipment purchase price, age, and specified useful life of the equipment. Fuel consumption could be predicted based on yearly hours worked by the equipment or yearly miles driven. Other input variables used are current odometer value of the equipment, fuel consumption in gallons, age of the equipment, purchase price of the equipment and class code id. Maintenance cost model development used cumulative work hours and cumulative miles driven recorded in the span of the year 2011 to 2017 by Oklahoma DOT. Other input variables used are the age of the equipment, the purchase price of the equipment, current odometer value, the useful life of the equipment assigned by the manufacturer and class code ID. The coefficients of variables obtained from the MLR test are explained to see the impact on fuel consumption and maintenance cost. Finally, the model was validated by utilizing the 30% validation data and yielded good prediction results. The predictive models will help state DOTs better budget equipment operational budget as well as facilitate the equipment rental rate update process. Furthermore, future recommendations are stated at the end of the last chapter so that this study could be taken forward.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1. **Background**

Construction Equipment is one of the most capital intensive long term investments (Rise 2016). State departments of transportation (DOTs) utilize a variety of vehicles for construction as well as highway maintenance operations. DOTs make expenditures to acquire vehicles and administer the operational use of the equipment. They also perform routine and preventive maintenance tasks and repair damages on the equipment (NCHRP 2018). For the approval of the funding and budgets for the DOT's, agency's past performance, objectives and goals, and proposed activities are legislatively reviewed (Rall 2016). The performance achievement of objectives could be improved with a better equipment management plan and tools to achieve efficient equipment management system. Equipment decision making could be improved by using historically acquired data and utilizing the current operational data to make sound equipment decision and improve the productivity of the machine in the fleet. Hence, it is the sole responsibility for the equipment managers to utilize the machine properly and match their capacities to specific project requirements (Manikandan 2018).

The knowledge of equipment economics is critical for the equipment management from the perspective of the budget for equipment fleet within state DOTs. The availability of the construction equipment, the ownership and operating costs relates to the economics of construction equipment (Kannan 2011). Operating costs consist of maintenance and repair costs and fueling costs. These costs not only depend on the utilization of machine but also on the strategies that are built to support the equipment throughout its life (Hall 2013). Hence, the operational analysis of the equipment fleet is essential in order to make sound economic decisions.

Operating costs are tracked down by the accounting team and then undergo the calculation process (Mitchell Jr 1998). The use of technical software and advanced machinery records the relevant equipment data and is being used to improve the functionality, operation, maintenance, and management of the construction equipment (Monnot and Williams 2010). Development of predictive models using the equipment data such as maintenance data, fueling data, work orders, initial purchase price, odometer readings, and present age of the equipment for a large equipment fleet with the DOTs will not only help to improve the economic decision making for the equipment but will also allow the managers to use this knowledge in updating operational rental rates regularly and operating cost forecasting.

1.2. Problem statement

DOTs are engaged in the tasks of heavy civil maintenance and construction throughout the year. They lack in effective asset management (FHWA 2019), and fleet managers need resources for effective equipment decisions which are not only financial decisions but also an operational decision (TRB 2019). Operating costs are the costs that are constantly paid for running the equipment, and they could be of different magnitude depending upon the equipment type and its condition. The repair or replacement decisions are a consideration for DOT fleets as a method to manage fleet expenditures, and they must include both maintenance and fueling consumption in decision-making criteria. It would not be ideal to solely predict maintenance, repair, and fueling costs based on experience especially when a big fleet is in consideration. The equipment in the modern era is equipped with the technology that could collect volumes of data regarding every aspect of equipment's operation (Monnot and Williams 2010). Some studies have been done in the past using the equipment data aiming toward the development of statistical models to make predictions. Although different studies contributed towards construction equipment management, still most of the equipment decisions within state DOTs are made based on the experience of fleet managers and operating costs are estimated using available rate schedule published by FEMA (2017) or Cost Recovery Rental Rate Blue Book (EquipmentWatch 2019). However, with the adoption of computerized equipment fleet management systems, fleet managers could calculate these costs associated with the equipment by using the data that is available to them. Better predictions or forecasting could be achieved if equipment managers have the right model developed out of their own field data.

1.3. Objectives

This research was conducted using the equipment database provided by the Oklahoma Department of Transportation as it manages a big equipment fleet and has a big volume of data in its database. The main objectives to achieve in this research are:

1. Developing models to predict the annual fuel consumption per equipment type using regression analysis.
2. Developing models to predict cumulative maintenance cost associated with the equipment using regression analysis.

1.4. Scope

In order to achieve the above-mentioned tasks, the field data is a prerequisite. The data was studied and processed in order to perform the required analysis to develop the model. The equipment under study includes pick-up trucks, motor graders, front end loaders, wheel tractors, power sweepers, large-sized trucks, etc. The equipment was divided into two categories based on their operating cost charge type i.e. equipment charged by dollar/hour and by dollar/mile. A total of 2000 pieces of equipment was used in the research and grouped together by class types. The similar type of pieces of equipment was categorized by unique class code numbers that provide the description to the equipment. In total 47 different class codes were used to group the pieces of equipment.

1.5. Organization of thesis

In order to explain the importance of equipment's maintenance and operational costs and fulfill the objectives of this research, this thesis is structured in five chapters. Chapter One gives an introduction to the research. It contains a problem statement, research objective, and scope of work. Chapter Two talks about different equipment management practices used by state DOTs. The chapter discusses previous research on using statistical analysis on equipment maintenance and fueling data to develop predictive models. Various studies have been summarized in the development of fuel consumption models using equipment data. The use of equipment operational data to predict equipment failures, the residual value of the equipment, forecasting maintenance, etc. is explained briefly in the chapter. Chapter Three describes the research methodology. The chapter discusses the various data analysis steps taken to get the results. Chapter Four discusses the results of regression analysis and the best models developed for the prediction of fuel consumption and maintenance costs. Chapter Five contains the conclusion, limitations, and recommendations for further research that could be done in this field.

CHAPTER II

LITERATURE REVIEW

This chapter discusses the utilization of equipment data for different research work done in the past. In the beginning, various models developed for fuel consumption and prediction of maintenance cost are discussed. This chapter also talks about the equipment management practices adopted by the Department of Transportation of the United States. The use of equipment data by various studies is reviewed in this chapter. The techniques used by other researchers such as linear and multiple regression, time series analysis, decision tree, etc. to develop statistical models are discussed in this chapter.

2.1. Fuel consumption models

Consumption of fuel is an important factor in determining the operating cost of the equipment. Some studies have been done in the past to develop fuel consumption models for different research objectives. A study used the Nebraska Tractor Test laboratory NTTL data to compare the results with ASAE Standards (Grisso et al. 2004). The researchers found a 4.8% decrease in average annual specific volumetric fuel consumption, and equations were developed for the diesel tractor engines. The models were based on the information about engine and chassis configuration, tractor weight during testing, and unballasted weight. A later study was done by Grisso et al. (2010) to develop the factsheet for the prediction of the fuel consumption during full and partial loads by using the field data and Nebraska Tractor Test laboratory (NTTL) data.

The developed models were based on engine load, engine speed, engine power, and fuel consumption.

Another study was done using the data collected at the Oak Ridge National Laboratory (ORNL) by driving the test vehicle in the fields and collecting the data (Ahn et al. 2002). The authors developed a microscopic fuel consumption and emission model using regression analysis as their approach. Although the model provides better results for only light-duty vehicles and if only, the vehicle's emission is consistent with the ORNL emission data. Another research was conducted to estimate the operating cost, fuel consumption, and pollutant emissions using aaSIDRA intersection analysis and an aaMOTION trip simulator (Akcelik and Besley 2003). The study used on-road vehicle parameters, traffic and road parameters, and cost parameters to develop the model. The developed models gave good results but the researcher did not consider construction vehicles. There was another research with the focus on estimating the fuel use and emission rates of non-road diesel construction equipment such as backhoe, bulldozer, excavator, motor grader, off-road truck, and wheel loader. The analysis was done by performing representative duty cycles on the field equipment data (Lewis 2009). The fuel use was estimated by Multiple Linear Regressions. However, the research was limited to particular types of off-road construction equipment and the average fuel use was predicted based on the emission of gases. Engine Modal analysis was used to conduct this research to finally estimate the emission rate of the gases by the equipment.

Hence, the field data available for these kinds of studies are important to develop predictive and estimating models. Few authors have recorded equipment data in varied conditions (Abolhasani et al. 2008; Rasdorf et al. 2010) to help the researchers collect the field data. The more the availability of real-time field data, the more accurate will be the analysis results. Also, in order to make better estimations and predictions, if yearly consumption fuel is taken into consideration, the organizations would be able to budget the costs more precisely.

2.2. Maintenance and repair cost estimation

Maintenance and repair costs are one of the most discussed topics in construction equipment management. Morris (1988) did an *exploratory analysis* on farm tractors' data to estimate the maintenance and repair cost. The research focused primarily on hours of operation and cumulative maintenance and repair costs. As a result, the linear and quadratic models were developed for different scenarios. In another study, a *Markov model* was proposed for a machine that continuously operates and deteriorates during its service life (Sim and Endrenyi 1993). The model incorporated deterioration and Poisson failures, minimal repair, periodic minimal maintenance, and major maintenance after a given number of minimal maintenances. Another study discussed *Condition Based Modelling (CBM)* that could be used to improve equipment reliability at reduced costs (Alaswad and Xiang 2017). The paper provided models for cost minimization, reliability maximization, and different models for CBM processes. This paper discussed different models that could be used for formulating maintenance policies. Another study used *Genetic Algorithms* (GA) technique to develop a model for preventive maintenance planning (Lapa et al. 2006). Probabilistic modeling was discussed to calculate the reliability of a component under a given maintenance policy. The study provided good results for preventive maintenance policy which provides low costs with a high level of reliability. The study was conducted on the nuclear power plant reactors but is relatable with the construction equipment management strategies. However, the costs would not be that low if the main goal is to privilege reliability. Another research used *Cumulative Cost Model* to forecast equipment repair costs (Mitchell et al. 2010). The study utilized two methodologies *Life to Date* (LTD) repair costs and the *Period-Cost-Based* (PCB) model on the data recorded from 155-wheel loaders. The cumulative equipment repair cost is described by a second-order polynomial model relating cumulative work hours of use to cumulative costs. Although LTD gave superior results over PCB results, PCB method is preferred when the data in hand is not consistent. However, after the literature review, it is felt that method of forecasting and development of predictive models could be created more specifically if different kinds of heavy equipment are considered for the analysis instead of similar kinds of equipment. This will make prediction more realistic as DOTs possess different kinds of equipment in their inventory.

2.3. DOTs equipment management techniques

Asset Management is a systematic approach of maintaining, upgrading, and operating physical assets cost-effectively by combining engineering principles with sound business practices and economic theory (Pantelias 2005). Equipment management is one of the major aspects of asset management that requires strong decision making based on equipment performance and project needs. DOTs are addressing this need by using third-party software to record the equipment data or have developed their own tools to upgrade the equipment management system.

The DOTs have accepted the need for transition in equipment management tactics and they have started using a more advanced system to manage their equipment fleet**. Illinois DOT** (IDOT) replaced its old spreadsheet and manual paperwork method with *AgileAssets maintenance and fleet management* systems. AgileAssets provides data collection facility and plans future integrations to link the fueling and Automatic Vehicle Locator (AVL) systems (IDOT 2019). Currently, **Ohio DOT** (ODOT) is using geospatial tools to support asset management*. Linear Referencing System (LRS)* allows the State to collect and integrate the data to support back end management (ODOT 2012). **Michigan DOT** (MDOT) uses *Geographic Information Systems* (GIS) to manage the equipment data to support decision-making (MDOT 2017). **Minnesota DOT** (MnDOT) equipment and vehicles fleet are managed by utilizing M5, an equipment management system from Asset Works (MnDOT 2016). Table 1 summarizes the similar software services that different DOTs are using across the United States to better their equipment fleet management practices (AgileAssets 2018; FHWA 2018; PDRMA 2018; Roadsoft 2019; Scopatz et al. 2014).

Table 1: Summary of equipment management systems used by DOT's

Software	Developer/year	Description	DOT client
Fleet and equipment manager of AgileAssets®	AgileAssets	• Estimate the depreciation, LCC, and replacement of equipment • Fuel, inventory and repair management • Record the history of vehicle usage, maintenance, labor requirements, and used costs for parts, etc.	Oklahoma DOT, Illinois DOT, Colorado DOT
Real Cost	Federal Highway Administration (FHWA)	• Perform life-cycle cost analysis (LCCA) • Estimate equipment cost, service life • Compare life-cycle costs between alternatives	Virginia DOT
RTA Fleet Management	Ron Turley/1979	• Track the equipment, performance, vehicle use, and labor. • Determine the maintenance, repair necessary time	Minnesota DOT
Roadsoft	Michigan Technological University	• Roadway asset management • Collect the roadway and traffic data such as features of roadway or roadside, traffic operations, and crashes, etc.	Michigan DOT
FlletFocus	AssetWorks/1984	• Equipment life cycle management (budgeting, acquisition, capital improvement, campaigns, and disposal management) • Track various functions of vehicles and equipment • Estimate repair, preventive maintenance, operating cost of vehicles, equipment	New Jersey DOT, New York DOT, Ohio DOT, Oregon DOT, Virginia DOT, Washington DOT

2.4. Previous works utilizing equipment operational data

Equipment management data have been used in past decades to help equipment managers make equipment decisions such as preventive maintenance, fuel consumption, equipment replacement, etc. So far the chapter discussed the studies on the development of fuel consumption models and utilization of equipment maintenance and repair data to develop different maintenance cost, estimation models. The improvement in DOTs equipment management system to run a more reliable fleet is provided with the current software and tools that they have acquired. This section provides a brief discussion about different research conducted on equipment operational data. Table 2 summarizes the different works conducted by different authors by using operational data of the equipment and enlists the major findings of those researches.

Table 2: Summary of different works undertaken on equipment operational data

Author/s	Objective	The data	Main findings
Gunnar Lucko	To develop a mathematical tool for the prediction of residual value for the selected group of heavy construction equipment. Multiple linear regression analysis of 28 datasets was carried out (Lucko 2003)	Total of 11 different types of equipment with 35,542 entries was used. 28 categories were selected by size as measured by horsepower, operating weight, or bucket volume.	Three best algebraic models were selected from 11 models that gave the best results for the analysis. Within the same equipment type, the loss in RVP for higher machines as compared to the smaller machines over the same period of time.
Hon-lun Yip, Hongqin Fan, Yat-hung Chiang	Prediction of the cost of construction equipment by the application of general regression neural network (GRNN) models and conventional Box-Jenkins time series models (Yip et al. 2014)	The data was taken from the contractor for dump truck and wheel loader. The data provided information about monthly maintenance costs and fuel consumption of the equipment.	Forecasting results can be used for making better equipment management decisions. Results can be used for setting an accurate rate of charge on equipment use. The global trend of maintenance cost change was modeled, change patterns of maintenance cost were modeled.
Zane W. Mitchell, Jr.	Development of a regression model to represent repair costs in terms of the machine age in cumulative hours of use (Mitchell Jr 1998)	The research used the field data of 270 heavy construction machines. Cumulative work hours and cumulative repair costs were in the data set for two different types of machines.	Final selected model out of 15 models: $CCI = 1 + \beta 1x + \beta 2x2 + \varepsilon$ The equation can be used directly to estimate average to date, average incremental, or average period repair costs. A relationship to show how repair costs accumulate as machine ages.

Author/s	Objective	The data	Main findings
Seung C. Ok, Sunil K. Sinha	Construction equipment productivity estimation was done for dozer operation using an Artificial Neural Network Model on recorded equipment data (Ok and Sinha 2006)	The data was compiled from different projects to estimate dozer productivity. The parameters were different types of dozer, blades, soil types, weather conditions, dozing grades and distances.	The developed model explained the dozer equipment productivity estimation with seven independent factors. The research demonstrated that the artificial neural network model can be used for estimation of equipment productivity.
Neelima Suresh J, Sahimol Eldhose	Equipment productivity forecasting model for multi-story building construction through regression analysis. (J Suresh 2018)	The model was developed using 50 sets of data collected from multi-storeyed residential projects. Dataset included the records about maintenance costs, equipment specifications, working cycle, handling of equipment and age of equipment.	Equipment productivity forecasting model was developed with the coefficient of determination of the model equal to 0.533. The developed model could be used for high level of quality and cost-effectiveness in the projects.
John C. Hildreth	Effect of period length on forecasting maintenance and repair costs for heavy equipment by the period cost methodology (Hildreth 2018)	The data was collected from a fleet of excavators. Operating weight and maintenance records of 21 machines were used for the study.	The longer the data recording period with more number of the equipment is, the better will be the forecasting results.
Hongqin Fan	Prediction of equipment failure using data mining and statistical analysis (Fan 2012)	The data was taken from a contractor having equipment odometer and hour meter readings, equipment downtime and uptime, equipment repair details, working hours and work locations.	While two different statistical models were analyzed for the research with pros and cons for both, the researcher found that the Power Law Models could be generated with fewer data as compared to the Time Series Models. However, the Time Series Model was able to detect changes of failure patterns better than the Power Law Model, and

Author/s	Objective	The data	Main findings
			accuracy of time series was found to be better than the Power Law Models.
Qing Fan, Hongqin Fan	Using Time Series Modelling to perform the Reliability Analysis and Failure Prediction of construction equipment (Fan and Fan 2015)	A large amount of data about maintenance, repair, and equipment failure records was taken from the contractors for analysis and predictions.	Using the time series modeling, the authors predicted the number of failures per interval and compared them with the actual number of failures. A forecast was done for Minimum Time Between Failures using the two factors Time Between Failures and Time To Repair.

2.5 Knowledge gap

In past years, the researchers have been using equipment operational data to better the equipment management practices to make equipment decisions such as replacement, updating rental rates, maintenance prediction, predicting equipment reliability, economic life, etc. This chapter discussed various statistical modeling techniques used for developing predictive and forecasting models to better the equipment management program. However, various factors like unavailability of data, different geographical conditions of the job sites, and lack of technology are keeping this topic active among the researchers to develop tools that could justify the equipment decision making with strong background proof.

Although there has been various research done in the area of fuel consumption and maintenance and repair prediction, the studies are limited. For examples, previous studies focused on a few pieces of particular vehicles, such as farm tractors, buses, wheel loaders, but an amalgamation of equipment used in state DOTs has rarely been the focus. This study will take fueling costs and maintenance and repair cost into account. Additionally, the study will use the data recorded from different types of equipment having the average age range of two to twenty years. The equipment data was obtained from Oklahoma DOT; therefore, the research will provide a model that may allow DOTs (assuming all DOTs are utilizing almost similar kind of equipment fleet) to predict the fuel consumption and maintenance cost of their construction equipment fleet.

CHAPTER III

METHODOLOGY

In this section, the data source and research parameters are introduced. The data source, data preparation, data processing using MySQL Workbench tool, data mining technique, and development of models using SAS Enterprise Miner and other techniques implemented in this study are discussed.

3.1. Data Source

The raw data was obtained from the Oklahoma Department of Transportation in the form of EXCEL spreadsheets. The spreadsheets contained the equipment operational as well as purchase information. They provided a classification based on equipment class code, which is a specific integer number given to a similar category of the equipment. The class code data table contains the information about equipment's average purchase price, average salvage value and distinguished the equipment based on the size, equipment type, and operational charge rate. For the final results of this research, all the individual equipment will be aggregated to their respective class code number only. The data description table for the class types and class code numbers is provided in table 3 and table 4, which summarizes the equipment size and type description.

Table 3: Description of equipment charged by dollar/mile used for analysis

Equipment Class Code Id	Equipment Size Description	Equipment Type Description
5085	Four-Door Sedan-Full Size	Auto - Factory Color
5086	Four-Door Sedan-Mid Size	Auto - Factory Color
5089	Four-Door Sedan-Mid Size	Auto - White Color
5090	Four-Door Sedan-Compact	Auto - White Color
5385	1/2 Ton Fleet side	Pickup
5386	3/4 Ton Fleet side	Pickup
5388	3/4 Ton	Pickup
5392	1 Ton Fleet side	Pickup
5393	1 Ton W/O Bed	Pickup
5394	1 Ton, Dual Rear	Pickup
5395	Full-size	Pickup
5398	10,000 G.V.W	Pickup
5399	15,000 G.V.W	Pickup
5401	4900 G.V.W	Van-Mini
5402	3/4 Ton - Window	Van
5404	3/4 Ton Station Wagon	Carryall/Suburban
5407	8500 G.V.W	Van
5418	2 Ton W/Steel Flat Bed(86-B-2)	Truck - Maintenance
5419	2 Ton W/Steel Flat Bed(86-B-6)	Truck - Maintenance
5420	24000 G.V.W - Diesel	Truck
5421	2-1/2 Ton W/Winch, Flat Bed,	Truck
5425	2-1/2 Ton	Truck
5427	3 Ton W/10 Yd Dump Bed 86-B-10	Truck
5428	3 Ton - Diesel - Haul	Truck - Tractor
5429	3 Ton Diesel	Truck - Diesel-Haul
5430	41000 G.V.W - Diesel	Truck

Equipment Class Code Id	Equipment Size Description	Equipment Type Description
5431	25000 G.V.W -Mid Range	Truck
5433	27,500 G.V.W -Mid Range	Truck
5434	24000 G.V.W -Diesel	Truck
5435	41000 G.V.W -Diesel	Truck
5441		46,000 G.V.W Diesel
5442	3/4 Ton	Crew Cab Pickup

Table 4: Description of equipment charged by dollar/hour used for analysis

Equipment Class Code Id	Equipment Size Description	Equipment Type Description
5120	55 Net H.P.	Backhoe-Loader-Tractor Unit
5121	80 Net H.P.	Backhoe-Loader-Tractor Unit
5123	92 Net H.P.	Backhoe-Loader-Tractor Unit
5189	Self-Propelled	Power Sweeper
5191	Diesel Powered	Street Vacuum Sweeper
5236	125 H.P.	Motor Grader
5237	150 H.P.	Motor Grader
5238	150 H.P.	Motor Grader
5355	2 Yd.	Front End Loader
5357	1/3 Cu. Yd. Cap.	Skid Steer Loader
5360	5000 Lbs.	Fork Lift
5362	10,000 Lb. Pneumatic	Fork Lift
5371	70 H.P.	Wheel Tractor
5375	85 H.P. Diesel	Wheel Tractor
5378	400 Cc Gas Engine	All-Terrain Vehicle

The data set is divided into two categories based on the equipment operational charge type for the sake of the research objective. The two categories of equipment are as follows:

1. Equipment charged by dollar/mile – Equipment such as trucks, pick-up trucks, and cabs, pick-up vans, etc.
2. Equipment charged by dollar/hour – Heavy civil equipment such as motor graders, front end loaders, etc.

The reason behind the data split into these two categories is that the DOTs have a big inventory and the equipment have varied tasks. The trucks and other similar equipment are engaged more in daily work as compared to the other heavy civil equipment like Motor Graders, Front End Loaders, etc. Hence, it is required to have different models for each category as the work engagement is different for both equipment types. Both categories have different types of machines purchased in different years and having all the data records such as maintenance and repair records, fueling records, work order of the equipment (hours worked or miles driven), purchase value, odometer readings, etc.

The inventory data obtained from ODOT had the list of equipment purchased since the 1970s. Therefore, some equipment was already inactive and had been sold and some are still active with some brand new pieces purchased after 2016 or so. Hence, inactive equipment was ignored from the dataset as they are not of interest to the research objectives. Moreover, since the data was provided by the ODOT, the author just focused on the class types of the equipment that are important to the ODOT, this was another important criterion for taking these particular equipment class into consideration. Therefore, in the end, there were 1,190 pieces of active equipment used for the analysis in the ODOT equipment inventory that is charged by dollar/mile and 827 pieces of equipment that are charged by dollar/hour. The research took into account individual equipment for the calculation process.

3.2. Data Preparation

The EXCEL spreadsheets were exported from the AgileAssets equipment inventory database. The data was available in separate EXCEL sheets for different types of records. All the different equipment records were first brought down at the individual equipment level. To fulfill the research objectives, the analysis was done with individual equipment and putting them in their class code category to use the final developed model. This was required so that the predictions could be made on different equipment category as different equipment type has different fuel consumption and a different maintenance expense. In order to facilitate this task, the data tables were imported to a unified visual tool for database management, MySQL Workbench, to query the information needed for this research and compile the data together in a single spreadsheet. The data was arranged in a way to get the different equipment records such as yearly fuel consumption in gallons, yearly maintenance and repair costs, equipment purchase price, equipment purchase date, ODOT's specified useful life of equipment, current odometer values, yearly work hours, yearly miles driven, and equipment class code numbers for each individual piece of equipment. Different MySQL queries were created to align the dataset as required for the research objective. The queries created for data preparation are provided in the appendices. For the prediction of fuel consumption for the equipment charged by dollar/hour, the data was organized with yearly fuel consumption with respect to the yearly hours worked. Wherever, for the equipment charged by dollar/mile, the data was organized with the yearly fuel consumption with respect to the yearly miles driven of the equipment. Table 5 summarizes the variables used for the model development for both cases.

Table 5: Variable description for predicting fuel consumption

Variable Name	Description	Variable Type
Equipment charged by dollar/hour rates		
Fuel Quantity	Predicted fuel amount	Target Variable
Original_Value	Equipment purchase price	Input Variable
Yearly_Hours	Yearly Hours worked by equipment	Input Variable
Class_Code_ID	Equipment category number	Input Variable
Equipment charged by dollar/mile rates		
Fuel Quantity	Predicted fuel amount	Target Variable
Original_Value	Equipment purchase price	Input Variable
Yearly_Miles	Yearly miles are driven	Input Variable
Age	Age of the equipment	Input Variable
Class_Code_ID	Equipment category number	Input Variable
Current_Odometer	At present odometer reading	Input Variable

Similarly, the dataset was prepared for the prediction of maintenance cost for the equipment charged by dollar/hour and dollar/mile rates. The data was organized by keeping yearly worked hours, yearly miles were driven and yearly maintenance and repair cost. The other variables that were used for both the cases are summarized in table 6.

Table 6: Variable description for predicting maintenance costs

Variable Name	Description	Variable Type
Equipment charged by dollar/hour rates		
Total Maintenance	Predicted maintenance cost	Target Variable
Original_Value	Equipment purchase price	Input Variable
Total_Hours	Cumulative Hours worked by equipment	Input Variable
Class_Code_ID	Equipment category number	Input Variable
Age	Age of the equipment	Input Variable
Current_Odometer	At present odometer reading	Input Variable
Useful_Life_eq	The expected work life of the equipment	Input Variable
Equipment charged by dollar/mile rates		
Total Maintenance	Predicted maintenance cost	Target Variable
Original_Value	Equipment purchase price	Input Variable
Age	Age of the equipment	Input Variable
Class_Code_ID	Equipment category number	Input Variable
CURRENT_ODOMETER	At present odometer reading	Input Variable

3.3. Data Mining

Data Mining (DM) is a technique that provides the platform to analyze the bulk of raw data, extract the data patterns and then converts the data into actionable information (Leventhal 2010). Data Mining is sometimes misunderstood to be similar to statistical analysis (SA), but there is a big difference in both the techniques, in fact, statistical analysis could be referred as a component of data mining. For instance, the SA analyzes a fairly small database as compared to the DM databases that have thousands of variables in it. That means statistics is about quantifying the data whereas, DM is about understanding, modeling, evaluation, and deployment of big databases. Another difference is the type of data, DM takes into account numeric as well as non-numeric data (EDUCBA 2018), whereas, SA can work only on numeric and clean data. In the end, the task of both techniques is to find solutions to the existing problem using available data and implementing similar techniques. Various DM software is available these days but we used Statistical Analysis System (SAS) Enterprise Miner for the objectives of this study. Since the data was available in a big amount and after combining two or more tables using MySQL, the database became more complex. The author applied the Data Mining technology discussed by (Leventhal 2010) to achieve the goals of this study. Data Mining is implemented when dealing with a certain kind of business problem that has a specific model development goal. For this study, the goal was to predict fuel consumption and maintenance cost. The data mining is used for data collection, data preparation to deploy the models and assessing results. Since the target variable is continuous in nature i.e. fuel consumption and maintenance cost, the average square error (ASE) was used to assess the model performance. Accordingly, the model with the minimum average square error was chosen as the final model.

3.4. Addressing Data Problems

SAS Enterprise Miner is used to manage the big dataset by executing the data mining process to develop descriptive and predictive models with high accuracy. The trends and anomalies can be searched using highly interactive and visualization tools using SAS. No data is initially ever ‘clean’. The need and requirements of the research determine the manipulations that are supposed to be done on the raw dataset. The dataset required for the study was prepared from the raw dataset using the MySQL server. The data cleaning and manipulations were done using SAS Enterprise Miner. The dataset had duplicate variables and outliers for equipment operational data. The following sections discuss the solutions to these problems.

The high number of outliers introduces bias in the study. The skewed data is improved either by using a logical range of values and assigning any outlier the upper bound of the range or categorizing the data based on the distribution. For the objective of fuel consumption modeling, the author removed the outliers of the dataset that comprises of the data from 2010 and 2018 as it was negligible but hampering the patterns. For the maintenance cost modeling, the author used cumulative maintenance, odometer reading, and cumulative hours of operation because the maintenance of previous years is most likely to affect future maintenance costs. Since ODOT started to record data using AgileAssets after 2010, the equipment purchased before 2010 was not considered in maintenance cost modeling as they had insufficient operational records and the SAS algorithms could not handle them. For the input variable 'age' of the equipment, it was calculated by subtracting purchase date of the equipment from 2018 i.e. the year till when equipment is still active. Multiple regression was performed in SAS Enterprise to run all the different algorithms and select the best fit models for the objective of this research.

3.5. Developing Models

For the purpose of predictive modeling in SAS Enterprise Miner, the prepared dataset is split into Model (70%) and Holdout (30%) samples. Data Partition node is used to divide the data into training and validation sets, 70% were used for training and 30% for validation. Then the data treatment was performed as discussed in section 3.4 so that the models can be created from the prepared dataset. Since the skewed data was resolved, the other steps followed are the analysis of variable distribution, imputing missing values, and decreasing the number of unique levels for the categorical variable. Fortunately, the data was not that dirty so these analyses did not ask for any major change or assumption in the data for modeling. Figure 1 and figure 2 presents a flow chart for the creation of models for both the research objectives and comparison of the models to select the best fit models.

Multiple Regression (MR) analysis was chosen to create the models which are the most important component of this research. Multiple regression is fairly robust and may work well for predictions even when some assumptions are violated. MR model are mathematically expressed in the form of $Y = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots + \beta_n X_n$, where Y is the target variable (refer Table 2 & Table 3 for target variable), X are the input variable and β is the coefficient. Another assumption is that X variables are linearly independent i.e. it is not possible to express any X as a linear combination of the other X's and X variables are linearly related to the Y variable.

Finally, SAS EM allows the user to compare different models through the comparison node which enables the user to compare the performance of the competing models using different benchmark criteria. Since the objective of this study was to predict the numerical variable i.e. fuel consumption in gallons and maintenance cost in dollars, therefore Average Squared Error (ASE) is used as the selection criterion. ASE is calculated by dividing the sum of squared errors (SSE) with the number of observations (SAS 2019). According to the ASE criterion, the model with the least value is considered to be the best model as compared to the model with a higher ASE value because the model with the lesser value is less biased as compared to the other models. As a result, four best models were selected for the research objectives to predict the fuel consumption and maintenance cost for both the categories, equipment charged by dollar per hour, and equipment charged by dollar per mile. The figures depicting model formation as discussed in this section are shown below (Figure 1 and Figure 2).

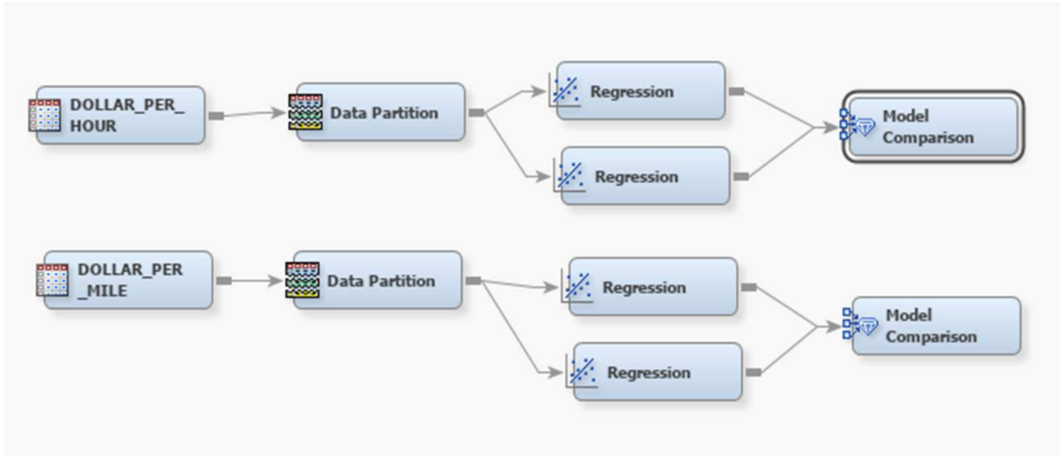

Figure 1: Fuel consumption prediction model flow chart

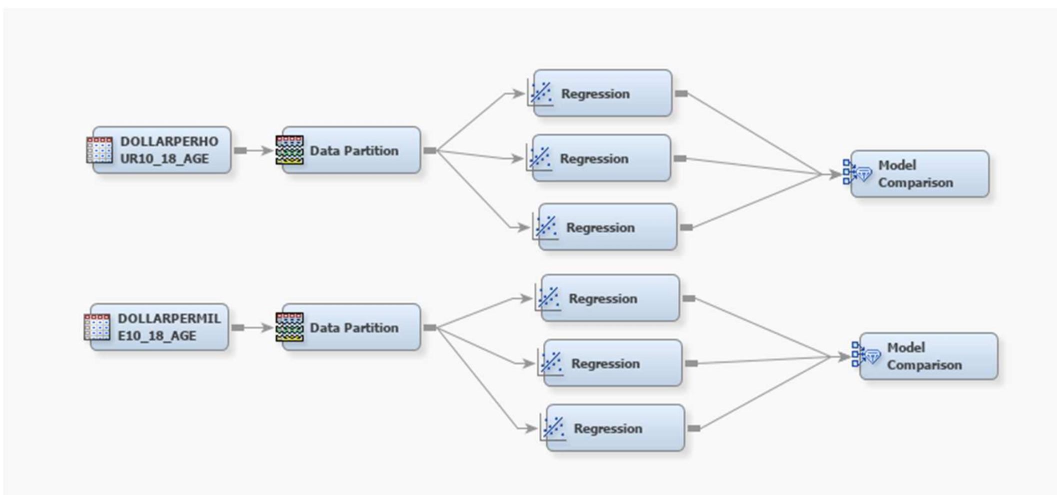

Figure 2: Maintenance cost prediction model flow chart

CHAPTER IV

RESULTS AND DISCUSSION

Multiple Linear Regression (MLR) results of fuel consumption and maintenance cost predictive models for both the equipment categories, i.e. equipment charged by dollar/hour and equipment charged by dollar/mile are present in this section. Table 7 provides a brief description of the input variable used to develop the predictive models. This chapter contains the best four models, their discussion and the model validation.

Table 7: Description of variables in the models for future calculations

Input Variables	Description
ORIGINAL_VALUE	Purchase price of the equipment
Yearly_hours	Yearly hours worked by the equipment
_CLASS_CODE_ID	Put integer value 1, it includes a number of similar kind of equipment.
YEARLY_MILES	Yearly miles are driven
Age	The current age of the equipment
CURRENT_ODOMETER	Current odometer value of the equipment
Useful_life_eq	Probable life of equipment given by the manufacturer

4.1. Descriptive Statistics of the data

The data description is necessary in order to provide some information about the data on which the models were developed in this study. This section discusses the input variable's mean, standard deviation, missing values in the data, minimum and maximum values of the input variables. Table 8 and Table 9 provide the exploratory descriptive statistics about the input variables. For all the continuous variable current_odometer, original_value, etc., there was no missing value in the dataset. The average mean current odometer value for equipment charged by dollar/hour was significantly lower as compared to the equipment charged by dollar/mile. Therefore, the average yearly fuel quantity was higher for equipment charged by dollar/mile as compared to the equipment charged by dollar/hour. Maximum and minimum columns show the maximum and minimum values of the input variables respectively. The average original value (purchase price) of the equipment charged by dollar/hour is higher as compared to the equipment charged by dollar/mile. It is self-explanatory as equipment like motor graders, loaders, etc. are expensive as compared to the pick-up trucks and other equipment charged by dollar/mile. Standard deviation is the deviation from the mean value.

Table 8 – Descriptive statistics of fuel consumption model

Equipment charged for operation by dollar/hour					
Variable	**Mean**	**Standard Deviation**	**Missing**	**Minimum**	**Maximum**
CURRENT_ODODMETER	2,580.275	1,746.318	0	9	12,879
ORIGINAL_VALUE	60,709.16	37,758.98	0	0	188,213
Useful_life_eq	10.96462	1.973163	0	10	15
Yearly_Fuel_Quantity	540.7342	407.3998	0	0	3152
Yearly_hours	228.4833	151.334	0	1	925
Equipment charged for operation by dollar/mile					
Variable	**Mean**	**Standard Deviation**	**Missing**	**Minimum**	**Maximum**
CURRENT_ODOMETER	126,124.6	70,684.61	0	-18,187	415,496
ORIGINAL_VALUE	52,514.32	30,160.44	0	0	244,308

YEARLY_MILES	11,391.94	9,397.003	0	4	91,319
Total_Fuel_Quantity	1,570.232	1,256.1	0	6	9,911
Useful_life_eq	12.96673	2.456225	0	10	15
Year	2014.779	2.32112	0	2010	2018

Table 9 – Descriptive statistics of maintenance cost model

Equipment charged for operation by dollar/hour					
Variable	**Mean**	**Standard Deviation**	**Missing**	**Minimum**	**Maximum**
CURRENT_ODOMETER	1,415.327	837.2747	0	9	5,541
Cumulative_cost_2011-2017	10,244.64	7,473.411	0	0	48,092.12
Cumulative_hours_2011-2017	1,369.166	823.4376	0	0	4,912
ORIGINAL_VALUE	71,991.41	38,383.93	0	0	174,494
Useful_life_eq	10.94104	1.956613	0	10	15
Equipment charged for operation by dollar/mile					
CURRENT_ODOMETER	75,964.79	44,121.37	0	160	235,511
Cumulative_cost_2011-2017	9,895.364	10,754.65	0	0	65,962.63
Cumulative_miles_2011-2017	63,212.04	45,861.47	0	0	219,275
ORIGINAL_VALUE	54,160.31	33,851.03	0	0	132,112
Useful_life_eq	11.94715	2.439802	0	10	15

Before moving forward in discussing the results obtained from the regressions, Equation 1 represents the MLR equation. This equation will be used to represent the model in algebraic equation.:

$\mathbf{Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots\ldots.. + \beta_n X_n}$ ……………………….Equation 1

Y = Fuel Consumption – for fuel consumption models

Y = Maintenance cost – for maintenance cost models

$X_1, X_2, X_3, \ldots\ldots X_n$ = Input Variables

$\beta_0, \beta_1, \beta_2, \beta_3, \ldots\ldots \beta_n$ = Coefficients

4.2. Model Development in SAS® Enterprise Miner

To develop the model, different tests were performed to select the best fit model which is statistically significant in predicting the target variable. The analysis tests are discussed briefly in this section to give an overview of how the best predictive model is selected. Analysis of Variance (ANOVA) was performed in order to determine the statistical significance of the model. P-Value < 0.05 (Alpha Value) specified that the model was statistically significant and confirmed that the independent variables reliably predicted the dependent variable. Degree of Freedom (DF) is associated with the source of the variance. CLASS_CODE is the categorical value in the model and hence the degree of freedom was calculated by subtracting 1 from the number of predictors (number of predictors – 1). This is the reason that the developed model contains multiple class codes. Now, the efficiency test was required to see how good the model is to make good predictions. The test is called the model fit statistics test, which defines the efficiency of the model. The concerned value was adjusted R-Square for this research objective as it explains how much variability can be explained by the model (UCLA 2018). R-Square and adjusted R-Square value show what percentage of variation is in the data is due to the independent variable but adjusted R square is preferred when more than one independent variable is used to develop the model (Derby 2015). Table 10 contains the adjusted R-Square values for all four models developed in this study. Analysis of the effects test was run in order to show the hypothesis tests for each of the variables in the model individually (UCLA 2019). It was run to check how the variables in the model significantly improve the model fit. The selection of the best fit model was based on Average Squared Error (ASE). ASE is the sum of squared errors (SSE) divided by the number of observations. For fuel prediction, only one regression each was done; hence, there was no need for model comparison. Whereas, for maintenance cost prediction three separate regression models were compared. The model with the lowest ASE value was chosen to be the best fit model since smaller values are preferred (SAS 2019). Finally, the best fit models were generated consisting of all the input variables with respective coefficients, which are the values for the regression parameters for predicting the dependent variable from the independent variables (UCLA 2018).

Table 10 – R-Square and Adjusted R-Square values of the models

MODEL	R-Square	Adjusted R-Square
Fuel consumption for dollar/hour equipment	0.7701	0.7690
Fuel consumption for dollar/mile equipment	0.7851	0.7835
Maintenance Cost for dollar/hour equipment	0.6182	0.5999
Maintenance cost for dollar/mile equipment	0.4864	0.4246

4.3. Fuel Consumption predictive model

The MLR approach successfully provided the models for predicting fuel consumption for both the equipment categories. For both models, all the parameters had p values < 0.0001. For the equipment charged for operation by dollar per hour, the selected model consists of the following effects: Intercept, ORIGINAL_VALUE, Yearly_hours, _CLASS_CODE. For the equipment charged for operation by dollar per mile, the effects are Intercept, CLASS_CODE_ID, CURRENT_ODOMETER, ORIGINAL_VALUE, YEARLY_MILES, age.

Fuel consumption predictive model for the equipment charged by dollar/hour:

The final model is developed after getting the analysis of maximum likelihood estimates. Table 11 is the final result of MLR in which the standard estimates (coefficients) and parameters of the regression are available. Error is the estimate of how scattered the data is. The larger the error, the less reliable the parameter is (Derby 2015). The t value and Pr > |t| are the values important to statisticians and these values test the hypothesis that the parameter is actually equal to zero.

Table 11: Analysis of maximum likelihood estimates

Parameter	DF	Standard Estimate	Error	t Value	Pr > |t|
Intercept	1	-47.5636	17.8232	-2.67	0.0077
ORIGINAL_VALUE	1	0.00110	0.000231	4.74	< .0001
Yearly_hours	1	1.9703	0.0257	76.54	< .0001
_CLASS_CODE 5120	1	-62.7692	20.7491	-3.03	0.0025
_CLASS_CODE 5121	1	-92.1433	25.6897	-3.59	0.0003
_CLASS_CODE 5123	1	-132.5	24.1089	-5.50	< .0001
_CLASS_CODE 5189	1	-56.0303	17.0892	-3.28	0.0011
_CLASS_CODE 5191	1	53.9059	76.3294	0.71	0.4801
_CLASS_CODE 5236	1	62.4718	43.3611	1.44	0.1497
_CLASS_CODE 5237	1	260.2	29.1582	8.92	< .0001
_CLASS_CODE 5238	1	75.4559	19.7692	3.82	0.0001
_CLASS_CODE 5355	1	-70.8765	15.3202	-4.63	< .0001
_CLASS_CODE 5357	1	-97.7508	20.3136	-4.81	< .0001
_CLASS_CODE 5360	1	-64.3132	39.9107	-1.61	0.1072
_CLASS_CODE 5362	1	71.7761	105.6	0.68	0.4968
_CLASS_CODE 5371	1	-17.2746	54.3593	-0.32	0.7507
_CLASS_CODE 5375	1	179.5	14.5039	12.38	< .0001
_CLASS_CODE 5378	1	-70.5771	47.7145	-1.69	0.0908

Based on the variable coefficients from the above table, the model of the fuel prediction model for equipment charged by dollar/hour was created. Equation 2 represents the developed model equation

-47.5636(Intercept)…………………………………………………….Equation 2

*+ ORIGINAL_VALUE *(.00110)*
*+ Yearly_hours * (1.9703)*
*+ _CLASS_CODE 5120 * (-62.7692)*
*+ _CLASS_CODE 5121 * (-92.1433)*
*+ _CLASS_CODE 5123 *(-132.5)*
*+ _CLASS_CODE 5189 * (-56.0303)*
*+ _CLASS_CODE 5191 *(53.9059)*
*+ _CLASS_CODE 5236 * (62.4718)*
*+ _CLASS_CODE 5237 * (260.2)*
*+ _CLASS_CODE 5238 * (75.4559)*
*+ _CLASS_CODE 5355 * (-70.8765)*
*+ _CLASS_CODE 5357 *(-97.7508)*
*+ _CLASS_CODE 5360 *(-64.3132)*
*+ _CLASS_CODE 5362 * (71.7761)*
*+ _CLASS_CODE 5371 * (-17.2746)*
*+ _CLASS_CODE 5375 * (179.5)*
*+ _CLASS_CODE 5378 *(-70.5771)*

Model Interpretation: 0.00110 is the coefficient of the original value. If all other variables are kept constant and increase the original value of the equipment by $10,000, the annual fuel consumption will increase by 10 gallons. Coefficient of yearly hours is 1.9703, which means if the yearly work hours are increased by 300 work hours, then the annual fuel consumption will increase by 591.09 gallons. For the categorical variable i.e. Class code, the highest coefficient is for 5237 and lowest for 5123. This means these two categories of equipment will consume maximum and minimum fuel respectively in the list of equipment charged by dollar/hour for operation if all other variables have the same input value.

Fuel consumption predictive model for the equipment charged by dollar/mile:

Similarly, as discussed for table 11, the same process was followed for the development of the predictive model for equipment charged by dollar/mile. Table 12 represents the analysis of maximum likelihood estimates for the current model development.

Table 12: Analysis of maximum likelihood estimates

Parameter	DF	Standard Estimate	Error	t Value	Pr > |t|
Intercept	1	8.5127	53.8440	0.16	0.8744
CURRENT_ODOMETER	1	0.00115	0.000158	7.24	< .0001
ORIGINAL_VALUE	1	0.00614	0.000632	9.71	< .0001
YEARLY_MILES	1	0.1075	0.00108	99.99	< .0001
age	1	-18.2660	2.3112	-7.90	< .0001
CLASS_CODE_ID 5085	1	752.4	330.6	2.28	0.0229
CLASS_CODE_ID 5086	1	-683.8	156.4	-4.37	< .0001
CLASS_CODE_ID 5089	1	-411.9	204.8	-2.01	0.0443
CLASS_CODE_ID 5090	1	-364.9	204.8	-1.78	0.0748
CLASS_CODE_ID 5385	1	-619.0	44.6906	-13.85	< .0001
CLASS_CODE_ID 5386	1	-525.5	83.5221	-6.29	< .0001
CLASS_CODE_ID 5388	1	-819.1	256.7	-3.19	0.0014
CLASS_CODE_ID 5392	1	-191.3	569.8	-0.34	0.7371
CLASS_CODE_ID 5393	1	-172.2	77.9077	-2.21	0.0272
CLASS_CODE_ID 5394	1	-213.8	53.6270	-3.99	< .0001
CLASS_CODE_ID 5395	1	-780.9	64.1338	-12.18	< .0001
CLASS_CODE_ID 5398	1	-332.0	256.7	-1.29	0.1960
CLASS_CODE_ID 5399	1	59.6537	123.3	0.48	0.6284
CLASS_CODE_ID 5401	1	-100.9	115.4	-0.87	0.3821
CLASS_CODE_ID 5402	1	-39.3478	131.9	-0.30	0.7655
CLASS_CODE_ID 5404	1	99.8479	403.6	0.25	0.8046
CLASS_CODE_ID 5407	1	-46.6372	68.0153	-0.69	0.4929
CLASS_CODE_ID 5418	1	247.9	234.8	1.06	0.2912
CLASS_CODE_ID 5419	1	232.5	161.0	1.44	0.1488
CLASS_CODE_ID 5420	1	164.2	62.4919	2.63	0.0086
CLASS_CODE_ID 5421	1	142.1	148.1	0.96	0.3374

Parameter	DF	Standard Estimate	Error	t Value	Pr > |t|
CLASS_CODE_ID 5425	1	436.7	77.0814	5.67	< .0001
CLASS_CODE_ID 5427	1	148.1	217.6	0.68	0.4962
CLASS_CODE_ID 5428	1	701.7	96.0165	7.31	< .0001
CLASS_CODE_ID 5429	1	403.4	99.1518	4.07	< .0001
CLASS_CODE_ID 5430	1	337.3	55.9980	6.02	< .0001
CLASS_CODE_ID 5431	1	339.5	60.5616	5.61	< .0001
CLASS_CODE_ID 5433	1	634.4	105.2	6.03	< .0001
CLASS_CODE_ID 5434	1	136.4	41.6222	3.28	0.0011
CLASS_CODE_ID 5435	1	429.5	43.1101	9.96	< .0001
CLASS_CODE_ID 5441	1	846.0	129.2	6.55	< .0001
CLASS_CODE_ID 5442	1	-596.3	43.6056	-13.68	< .0001

Based on the variable coefficients from the above table, the model of the fuel prediction model for equipment charged by dollar/hour was created. Equation 3 represents the developed model equation:

8.5127 (intercept) +……………………………………………………………………Equation 3
*Age * (-18.2660) +*
*YEARLY_MILES * (0.1075) +*
*ORIGINAL_VALUE * (.00614) +*
CURRENT_ODOMETER (.00115) +*
*CLASS_CODE_ID 5085 * (752.4) +*
*CLASS_CODE_ID 5086 * (-638.8) +*
*CLASS_CODE_ID 5089 * (-411.9) +*
*CLASS_CODE_ID 5090 * (-364.9) +*
*CLASS_CODE_ID 5385 * (-619) +*
*CLASS_CODE_ID 5386 * (-525.5) +*
*CLASS_CODE_ID 5388 * (-819.1) +*
*CLASS_CODE_ID 5392 * (-191.3) +*
*CLASS_CODE_ID 5393 * (-172.2) +*
*CLASS_CODE_ID 5394 * (-213.8) +*
*CLASS_CODE_ID 5395 * (-780.9) +*
*CLASS_CODE_ID 5398 * (-332) +*

*CLASS_CODE_ID 5399 * (59.6537) +*
*CLASS_CODE_ID 5401 * (-100.9) +*
*CLASS_CODE_ID 5402 * (-39.3478) +*
*CLASS_CODE_ID 5404 * (99.8479) +*
*CLASS_CODE_ID 5407 * (-46.6372) +*
*CLASS_CODE_ID 5418 * (247.9) +*
*CLASS_CODE_ID 5419 * (232.5) +*
*CLASS_CODE_ID 5420 * (164.2) +*
*CLASS_CODE_ID 5421 * (142.1) +*
*CLASS_CODE_ID 5425 * (436.7) +*
*CLASS_CODE_ID 5427 * (148.1) +*
*CLASS_CODE_ID 5428 * (701.7) +*
*CLASS_CODE_ID 5429 * (403.4) +*
*CLASS_CODE_ID 5430 * (337.3) +*
*CLASS_CODE_ID 5431 * (339.5) +*
*CLASS_CODE_ID 5433 * (634.4) +*
*CLASS_CODE_ID 5434 * (136.4) +*
*CLASS_CODE_ID 5435 * (429.5) +*
*CLASS_CODE_ID 5441 * (846) +*
*CLASS_CODE_ID 5442 * (-596.3)*

Model Interpretation: The coefficient of the current odometer, 0.00115, conveys that if the odometer value is increased by 30000 miles and other input variables are kept constant, the fuel consumption will increase by 34.5 gallons. If the equipment is driven additional 5000 miles yearly, then the fuel consumption will increase by 537.5 gallons.

4.4. Maintenance cost predictive model

The MLR approach successfully provided the models for predicting maintenance cost for both the equipment categories after running complicated regression. For both the models, all the parameters had p values < 0.0001. For the equipment charged for operation with dollar per hour, the selected model consists of the following variables: EQUIPMENT_CLASS_CODE_ID, CURRENT_ODOMETER* CURRENT_ODOMETER* CURRENT_ODOMETER, CURRENT_ODOMETER* CURRENT_ODOMETER*Useful_life_eq, CURRENT_ODOMETER * CURRENT_ODOMETER *age, CURRENT_ODOMETER*Useful_life_eq*age, and ORIGINAL_VALUE*ORIGINAL_VALUE*age. For the equipment charged for operation by dollar per mile, the effects are: Intercept, EQUIPMENT_CLASS_CODE_ID, age, age*age, CURRENT_ODOMETER*CURRENT_ODOMETER*age, CURRENT_ODOMETER*ORIGINAL_VALUE*age, age*age*age

Maintenance cost predictive model for the equipment charged for operation by dollar/hour: Table 13 is the analysis of maximum likelihood estimates for maintenance cost prediction model for the equipment charged by dollar/hour. As discussed in the last section, the parameters and standard estimates were used to develop the predictive model for this case.

Table 13: Analysis of maximum likelihood estimates

Parameter	**DF**	**Standard Estimate**	**Error**	**t Value**	**Pr > |t|**
Intercept	1	3417.1	1105.7	3.09	0.0022
CURRENT_ODOMETER*CURRENT_ODOMETER*CURRENT_ODOMETER	1	1.3186	3.5247	3.74	0.0002
CURRENT_ODOMETER*CURRENT_ODOMETER*Useful_life_eq	1	-0.00025	0.000082	-3.02	0.0027
CURRENT_ODOMETER*CURRENT_ODOMETER*age	1	-0.00080	0.000173	-4.63	<.0001
CURRENT_ODOMETER*Useful_life_eq*age	1	0.2577	0.0383	6.72	<.0001
ORIGINAL_VALUE*ORIGINAL_VALUE*age	1	-1.117	3.1818	-3.48	0.0006
EQUIPMENT_CLASS_CODE_ID 5121	1	-2788.3	2185.8	-1.28	0.2031
EQUIPMENT_CLASS_CODE_ID 5123	1	-2581.6	1206.9	-2.14	0.0333
EQUIPMENT_CLASS_CODE_ID 5189	1	-1691.9	1325.3	-1.28	0.2027
EQUIPMENT_CLASS_CODE_ID 5237	1	14100.3	2953.7	4.77	<.0001

Parameter	DF	Standard Estimate	Error	t Value	Pr > |t|
EQUIPMENT_CLASS_CODE_ID 5238	1	3729.5	2416.4	1.54	0.1238
EQUIPMENT_CLASS_CODE_ID 5355	1	-1947.3	821.8	-2.37	0.0185
EQUIPMENT_CLASS_CODE_ID 5357	1	-1175.3	1313.6	-0.89	0.3717
EQUIPMENT_CLASS_CODE_ID 5360	1	-3356.9	3065.4	-1.10	0.2744
EQUIPMENT_CLASS_CODE_ID 5375	1	-1911.4	1112.9	-1.72	0.0869

Based on the variable coefficients from the above table, the model of the maintenance cost prediction model for equipment charged by dollar/hour was created. Equation 4 represents the developed model equation:

3417.1 (intercept) +………………………………………………………………. Equation 4
*EQUIPMENT_CLASS_CODE_ID 5121 * (-2788.3) +*
*EQUIPMENT_CLASS_CODE_ID 5123 * (-2581.6) +*
*EQUIPMENT_CLASS_CODE_ID 5189 * (-1691.9) +*
*EQUIPMENT_CLASS_CODE_ID 5237 * (14100.3) +*
*EQUIPMENT_CLASS_CODE_ID 5238 * (3729.5) +*
*EQUIPMENT_CLASS_CODE_ID 5355 * (-1947.3) +*
*EQUIPMENT_CLASS_CODE_ID 5357 * (-1175.3) +*
*EQUIPMENT_CLASS_CODE_ID 5360 * (-3356.9) +*
*EQUIPMENT_CLASS_CODE_ID 5375 * (-1911.4) +*
*CURRENT_ODOMETER*CURRENT_ODOMETER* CURRENT_ODOMETER * (1.318E-6) +*
*CURRENT_ODOMETER*CURRENT_ODOMETER*Useful_life_eq * (-0.00025) +*
*CURRENT_ODOMETER*CURRENT_ODOMETER*age * (-0.00080) +*
CURRENT_ODOMETER Useful_life_eq *age * (0.2577)*
*ORIGINAL_VALUE*ORIGINAL_VALUE*age* (-1.11E-7)*

Model Interpretation: The developed model is not linear. Hence, the interpretation of this model is complicated because the target variable is not directly dependent on individual input variable, and polynomial degree inputs define the relationship with the target. Therefore, coefficients cannot be interpreted based on individual variable interpretation.

Maintenance cost predictive model for the equipment charged for operation by dollar/mile: Following the similar process for the fourth and last predictive model of this research, table 20 was obtained giving the results for the analysis of maximum likelihood estimates. The errors came out to be big numbers for this model and that explains the weakness of the final model in this case.

Table 14: Analysis of maximum likelihood estimates

Parameter	**DF**	**Standard Estimate**	**Error**	**t Value**	**Pr > |t|**
Intercept	1	42546	8758.6	4.86	<.0001
Age	1	-27977.2	5928.5	-4.72	<.0001
CURRENT_ODOMETER*CURRENT_ODOMETER*age	1	8.214E-8	1.134E-8	7.24	<.0001
age*age	1	6797.3	1233.4	5.51	<.0001
CURRENT_ODOMETER*ORIGINAL_VALUE*age	1	-2.78E-7	3.057-8	-9.09	<.0001
Age*age*age	1	-457.5	80.4431	-5.69	<.0001
EQUIPMENT_CLASS_CODE_ID 5085	1	-7795.5	7642.3	-1.02	0.3082
EQUIPMENT_CLASS_CODE_ID 5086	1	39985	7601.1	5.26	<.0001
EQUIPMENT_CLASS_CODE_ID 5089	1	1102.8	3914.6	0.28	0.7783
EQUIPMENT_CLASS_CODE_ID 5090	1	966.8	4464.3	0.22	0.8286
EQUIPMENT_CLASS_CODE_ID 5385	1	-2335.7	1327.7	-1.76	0.0792
EQUIPMENT_CLASS_CODE_ID 5386	1	4278.2	3954.4	1.08	0.2799
EQUIPMENT_CLASS_CODE_ID 5393	1	-13391.5	5494.1	-2.44	0.0152
EQUIPMENT_CLASS_CODE_ID 5394	1	-2197.4	1981.3	-1.11	0.2680
EQUIPMENT_CLASS_CODE_ID 5395	1	2516.3	1997.0	1.26	0.2083
EQUIPMENT_CLASS_CODE_ID 5399	1	-15484.2	5422.9	-2.86	0.0045
EQUIPMENT_CLASS_CODE_ID 5401	1	-6417	7550.3	-0.85	0.3958

Parameter	DF	Standard Estimate	Error	t Value	Pr > |t|
EQUIPMENT_CLASS_CODE_ID 5407	1	-1275.4	4480.6	-0.28	0.7760
EQUIPMENT_CLASS_CODE_ID 5418	1	-4730.4	7649.9	-0.62	0.5366
EQUIPMENT_CLASS_CODE_ID 5419	1	-5336.9	3518.9	-1.52	0.1300
EQUIPMENT_CLASS_CODE_ID 5420	1	747.1	5440.9	0.14	0.8908
EQUIPMENT_CLASS_CODE_ID 5428	1	6079	2998.9	2.03	0.0432
EQUIPMENT_CLASS_CODE_ID 5429	1	20515.4	5510.6	3.72	0.0002
EQUIPMENT_CLASS_CODE_ID 5431	1	253.3	5397.4	0.05	0.9626
EQUIPMENT_CLASS_CODE_ID 5434	1	-3502.7	1789.8	-1.96	0.0509
EQUIPMENT_CLASS_CODE_ID 5435	1	-918.8	1297.7	-0.71	0.4793
EQUIPMENT_CLASS_CODE_ID 5441	1	-2680.5	7559.3	-0.35	0.7231
EQUIPMENT_CLASS_CODE_ID 5442	1	-6954.6	1408.1	-4.94	<.0001

Based on the variable coefficients from the above table, the model of the maintenance cost prediction model for equipment charged by dollar/miler was created. Equation 4 represents the developed model equation:

*42546.4 (intercept) +………………………………………………………………......*Equation 5
age (-27877.2) +
*age *age (6797.3) +*
*CURRENT_ODOMETER*CURRENT_ODOMETER*age (8.214E-8) +*
*CURRENT_ODOMETER*ORIGINAL_VALUE*age (-2.78E-7) +*
*age * age *age (-457.5) +*
*EQUIPMENT_CLASS_CODE_ID 5085 * (-7795.5) +*
*EQUIPMENT_CLASS_CODE_ID 5086 * (39985.8) +*
*EQUIPMENT_CLASS_CODE_ID 5089 * (1102.8) +*
*EQUIPMENT_CLASS_CODE_ID 5090 * (966.8) +*
*EQUIPMENT_CLASS_CODE_ID 5385 * (-2335.7) +*
*EQUIPMENT_CLASS_CODE_ID 5386 * (4278.2) +*
*EQUIPMENT_CLASS_CODE_ID 5393 * (-13391.5) +*
*EQUIPMENT_CLASS_CODE_ID 5394 * (-2197.4) +*
*EQUIPMENT_CLASS_CODE_ID 5395 * (2516.3) +*
*EQUIPMENT_CLASS_CODE_ID 5399 * (-15484.2) +*
*EQUIPMENT_CLASS_CODE_ID 5401 * (-6417.6) +*
*EQUIPMENT_CLASS_CODE_ID 5407 * (-1275.4) +*
*EQUIPMENT_CLASS_CODE_ID 5418 * (-4730.4) +*
*EQUIPMENT_CLASS_CODE_ID 5419 * (-5336.9) +*
*EQUIPMENT_CLASS_CODE_ID 5420 * (474.1) +*

*EQUIPMENT_CLASS_CODE_ID 5428 * (6079) +*
*EQUIPMENT_CLASS_CODE_ID 5429 * (20515.4) +*
*EQUIPMENT_CLASS_CODE_ID 5431 * (253.3) +*
*EQUIPMENT_CLASS_CODE_ID 5434 * (-3502.7) +*
*EQUIPMENT_CLASS_CODE_ID 5435 * (-918.8) +*
*EQUIPMENT_CLASS_CODE_ID 5441 * (-2680.5) +*
*EQUIPMENT_CLASS_CODE_ID 5442 * (-6956.4)*

Model Interpretation: Likewise, the previous maintenance cost model for equipment charged by dollar/hour, the developed model is not linear. Hence, the interpretation of this model is complicated because the target variable is not directly dependent on individual input variable, and polynomial degree inputs define the relationship with the target. Therefore, coefficients cannot be interpreted based on individual variable interpretation.

4.5. Model Validation

This section provides a graphical summary of the performance of the model. This graphical summary is also called a score ranking matrix that shows how well the model is working on the dataset. If the gap between the mean predicted line and mean target line is small, then the model performs well in prediction. The lines indicate the prediction line and a target line. For the first two cases of fuel consumption the target and predicted are close to each other, hence the model performance can be observed as good. While for the maintenance models the performance cannot be commented well as they did not perform that well as fuel consumptions model. Still, the methodology could be used to enhance their performance by having data for a longer period of time.

The x-axis (depth) in the graphs displays the selected percentile values of the predicted probability groups. The group with the highest predicted probability has the lowest depth. Therefore, the group with depth 100 has the lowest predicted probability (Support 2019). The Y-axis denotes the value of target variable i.e. fuel consumption and maintenance cost. The decreasing order of the target value is the trend shown in the graph and at depth equal to 100, the value of the mean predicted and mean target could be either negative or zero.

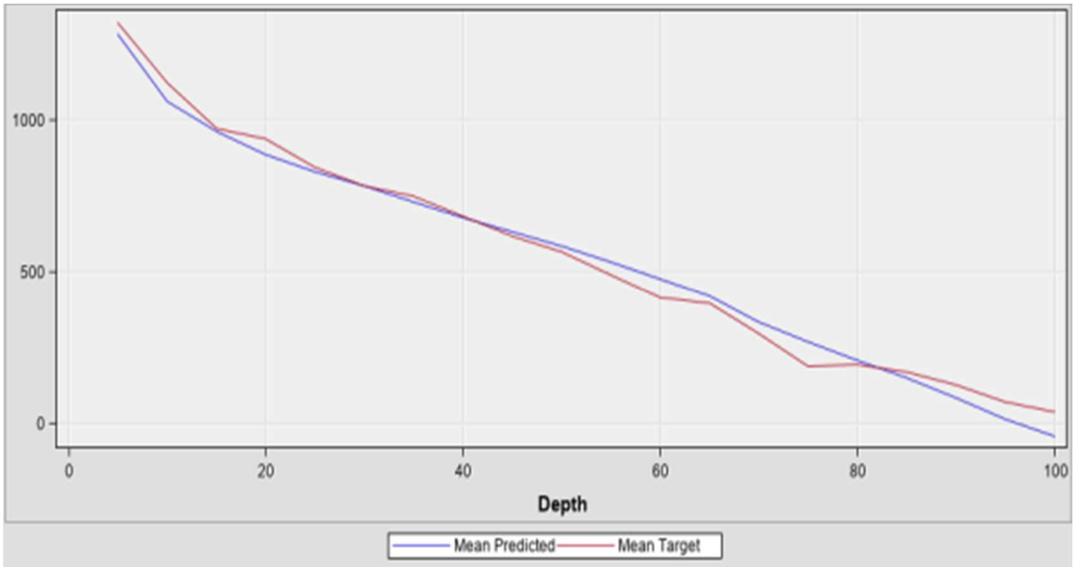

FIGURE 3: Validation data model for fuel consumption for equipment charged by dollar per hour

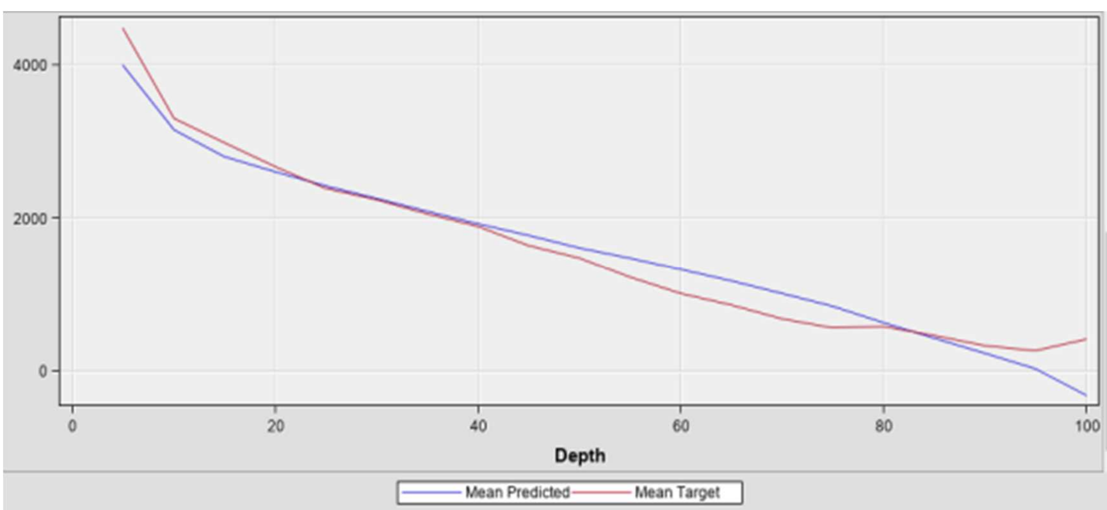

FIGURE 4: Validation data model for fuel consumption for equipment charged by dollar per mile

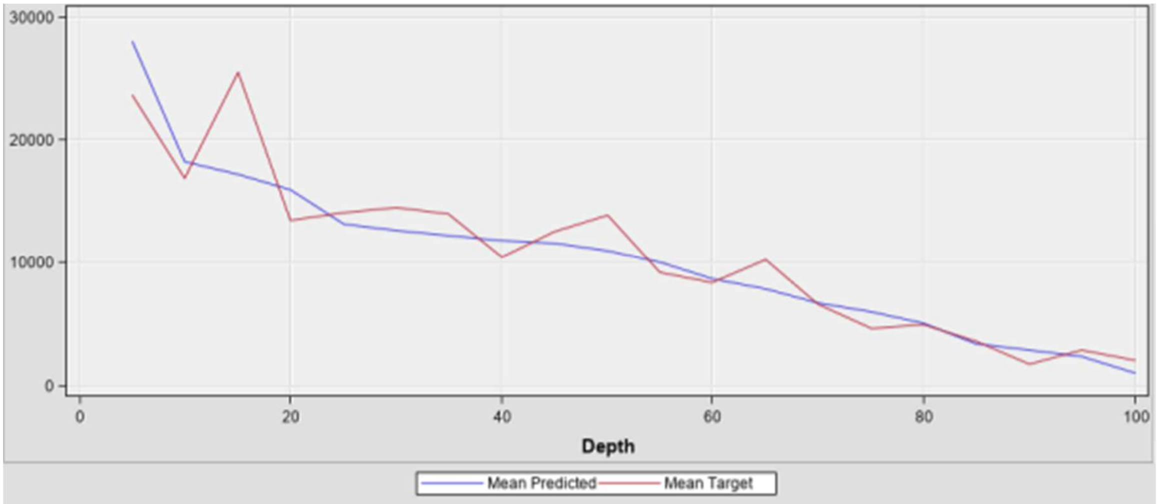

FIGURE 5: Validation data model for maintenance cost for equipment charged by dollar per hour

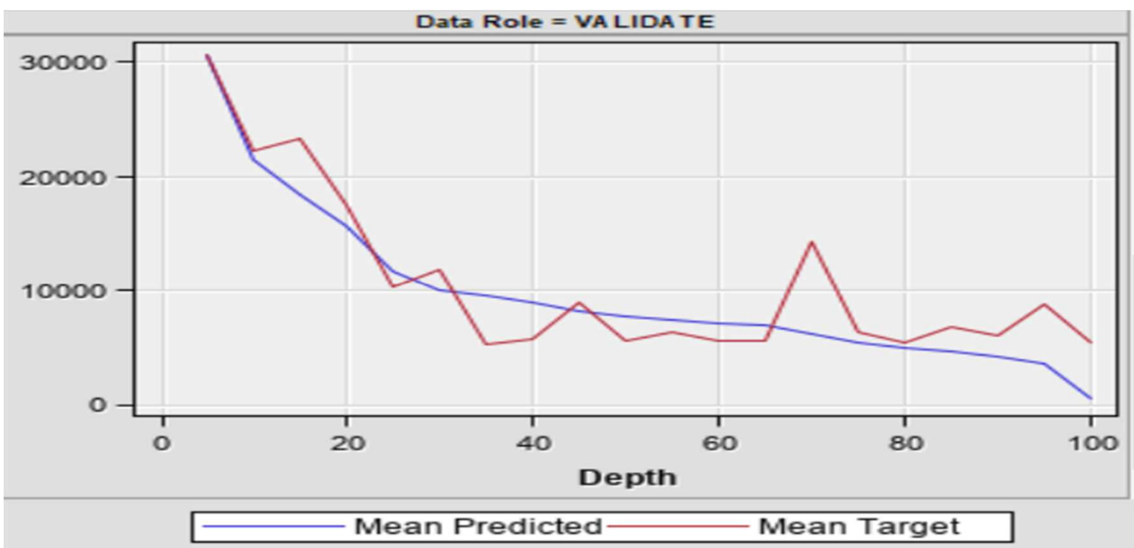

FIGURE 6: Validation data model for maintenance cost for equipment charged by dollar per mile

4.6. Summary of the results

In multiple linear regression, the coefficient of determination i.e. R^2 is the common measure to comment how well the model predicts the target or dependent variable (Harel 2009), where $R^2 = 1$ represents a perfect model fit and $R^2 = 0$ represents no linear fit. The range of 0-0.25,0.25-0.50 and 0.50-1.00 donate a weak, moderate and strong model respectively (Hair et al. 2013; Latan and Ramli 2013). The models developed for this research objective have the Adjusted R square value of 0.76, 0.78, 0.59 and 0.42 respectively. Hence, the developed fuel consumption models can make close to realistic predictions if implemented on the right category of the data for individual equipment for the total expected life of the equipment. In case of maintenance cost models, the model performance could be made better with more data in hand.

CHAPTER V

CONCLUSION AND RECOMMENDATION

5.1. Contribution of the research

This research is conducted to provide predictive tools that could be used by the state department of transportation of the United States. DOTs have to manage all the federal and state construction activities that engage a big equipment fleet. They need statistical tools developed from the equipment field data to predict various equipment related decisions and project efficient budgets. The predictive models developed in this research accounts for the prediction of fuel consumption and maintenance cost of the equipment. The output of this research could be taken into account for budget estimation, rental rate calculations, and equipment maintenance related decisions.

5.1. Research Findings

The research utilized the data analytics to study, prepare and test the data. For the development of the required models, different input variables were selected from the database and statistical analysis was performed in the SAS Enterprise Miner. The data for the research was obtained from the Oklahoma DOT and which is why the data was divided into two sets, equipment charged by dollar/mile and equipment charged by dollar/hour as practiced by ODOT to charge the equipment for operation. Hence, the developed models could be used by different DOTs as they are the target audience for this research. Since the data was The different sets of input variables were used to predict the target variable for both cases. The following predictors were used to develop the models in this research:

1. Good predictors to forecast fuel consumption: The Purchase price of the equipment, Yearly hours worked by the equipment, present age of the equipment and current odometer reading of the equipment.
2. Good predictors to forecast maintenance cost: Current odometer reading of the equipment, the useful life of the equipment specified by the manufacturer, present age of the equipment, the purchase price of the equipment.

The validation model graphs show a good alignment of the target and prediction line to each other in the fuel consumption modeling case. To, predict the target variable using the models developed in this study, the user needs to have all these input variables listed above to make the predictions. The efficiency or prediction accuracy of the model is determined by the adjusted R-Square value. The predictive accuracy of the fuel consumption models is significantly better as compared to the maintenance cost models. According to the author, the reason being that the dataset used to develop fuel consumption model contains a big unit of equipment purchased since the 1990s whereas, the maintenance cost model was developed using the equipment data purchased since 2010. Also, fuel consumption model took into account yearly data records, i.e. yearly hours worked, yearly miles driven, and yearly fuel consumption.

5.2. Limitations

The major limitation of this research is that the dataset did not allow the author to distinguish between preventive maintenance and equipment repairs. It could have allowed the author to develop a model that could predict the model considering both the factors separately and make better predictions. Another limitation was the time since the data was available. Oklahoma DOT started to record the data since November 2010, hence the work orders and maintenance and fuel records before that were not available for the equipment fleet. Lack of information about the equipment inventory such as engine size, engine specifications, etc. was another drawback that limited the scope of this research. If there were parameters to further distinguish the equipment, it would have allowed the author to explore and develop specific predictive models for the equipment that are mainly used by the DOT's.

5.3. Future recommendations

Since some DOTs are still using spreadsheets and manually entering data in the data systems for the equipment fleet management, while some DOT's have already started using third party software services or developed the data collection techniques to record the data and utilize it for equipment management. Therefore, there are a few recommendations for the future research objectives after reviewing previous works and working on the research objectives for this study:

1. The model can be validated with other state DOTs data to verify its generalizability. However, the approach devised by this research should be applicable to any state DOTs, as the equipment data have been used provided by ODOT. Other State DOTs can develop their own models that fit their own need.

2. Perform a similar study to develop maintenance cost model using parameters like engine size, number of axles, etc. to further distinguish equipment types. This will broaden the maintenance cost prediction with a focus on particular equipment category.

3. Conduct a study with the focus on distinguishing equipment based on the type of fuel consumed and developing predictive models for a particular type of fuel consumption.

4. Develop separate maintenance cost predictive models for preventive and scheduled maintenance, and repairs and breakdowns. The expenditures and frequency of both the maintenance tasks are different and better economic decisions could be made if using separate models.

REFERENCES

Abolhasani, S., Frey, H. C., Kim, K., Rasdorf, W., Lewis, P., and Pang, S.-h. (2008). "Real-world in-use activity, fuel use, and emissions for nonroad construction vehicles: a case study for excavators." *Journal of the Air & Waste Management Association*, 58(8), 1033-1046.

AgileAssets (2018). "Resources by topic: CLIENT." <https://www.agileassets.com/topic/client/>.

Ahn, K., Rakha, H., Trani, A., and Van Aerde, M. (2002). "Estimating vehicle fuel consumption and emissions based on instantaneous speed and acceleration levels." *Journal of transportation engineering*, 128(2), 182-190.

Akcelik, R., and Besley, M. "Operating cost, fuel consumption, and emission models in aaSIDRA and aaMOTION." *Proc., 25th conference of Australian institutes of transport research (CAITR 2003)*, University of South Australia Adelaide, Australia, 1-15.

Alaswad, S., and Xiang, Y. (2017). "A review on condition-based maintenance optimization models for stochastically deteriorating system." *Reliability Engineering & System Safety*, 157, 54-63.

Bayzid, S. M., Mohamed, Y., and Al-Hussein, M. (2016). "Prediction of maintenance cost for road construction equipment: a case study." *Canadian Journal of Civil Engineering*, 43(5), 480-492.

Derby, N. (2015). "Getting Correct Results From PROC REG." *Seattle, WA*, <https://www.sas.com/content/dam/SAS/en_ca/User%20Group%20Presentations/Regina-User-Group/Derby-PROCREG(Paper)-March2015.pdf>.

EDUCBA (2018). "DATA MINING vs STATISTICS." <https://www.educba.com/data-mining-vs-statistics/>.

EquipmentWatch (2019). "Cost recovery rental rate blue book." <https://equipmentwatch.com/blue-book-cost-recovery/>.

Fan, H. (2012). "Data mining and statistical analysis of construction equipment failure." *Gerontechnology*.

Fan, Q., and Fan, H. (2015). "Reliability analysis and failure prediction of construction equipment with time series models." *Journal of Advanced Management Science Vol*, 3(3).

FHWA, U. S. D. (2018). "Office of research, development and technology." <https://highways.dot.gov/turner-fairbank-highway-research-center/offices/office-research-development-and-technology>.

FHWA, U. S. D. (2019). "Asset Management Overview." <https://www.fhwa.dot.gov/asset/if08008/amo_06.cfm>.

Grisso, R. D., Kocher, M. F., and Vaughan, D. H. (2004). "Predicting tractor fuel consumption." *Applied Engineering in Agriculture*, 20(5), 553.

Grisso, R. D., Perumpral, J. V., Vaughan, D. H., Roberson, G. T., and Pitman, R. M. (2010). "Predicting tractor diesel fuel consumption."

Hair, J. F., Ringle, C. M., and Sarstedt, M. (2013). "Partial least squares structural equation modeling: Rigorous applications, better results and higher acceptance." *Long range planning*, 46(1-2), 1-12.

Hall, B. (2013). "Optimum equipment management through life cycle costing." <https://www.rpmglobal.com/wp-content/uploads/mp/files/resources/files/rpm-perspectives-2013-119.pdf>.

Harel, O. (2009). "The estimation of R 2 and adjusted R 2 in incomplete data sets using multiple imputation." *Journal of Applied Statistics*, 36(10), 1109-1118.

Hildreth, J. C. (2018). "Effect of Period Length on Forecasting Maintenance and Repair Costs for Heavy Equipment by the Period Cost Based Methodology." *International Journal of Construction Education and Research*, 1-13.

IDOT, A. (2019). "ILLINOIS DOT INTEGRATE MAINTENANCE AND FLEET MANAGEMENT SYSTEMS." *Illinois*, <https://www.agileassets.com/blog/illinois-dot-integrate-maintenance-and-fleet-management-systems/>.

J Suresh, N. E., Sahimol (2018). "Equipment Productivity Forecasting Model for Multi-storey Building Construction Through Regression Analysis." *IOSR Journal of Engineering*, 08(3).

Kannan, G. (2011). "Field studies in construction equipment economics and productivity." *Journal of construction engineering and management*, 137(10), 823-828.

Lapa, C. M. F., Pereira, C. M. N., and de Barros, M. P. (2006). "A model for preventive maintenance planning by genetic algorithms based in cost and reliability." *Reliability Engineering & System Safety*, 91(2), 233-240.

Latan, H., and Ramli, N. A. (2013). "The results of partial least squares-structural equation modelling analyses (PLS-SEM)." *Available at SSRN 2364191*.

Leventhal, B. (2010). "An introduction to data mining and other techniques for advanced analytics." *Journal of Direct, Data and Digital Marketing Practice*, 12(2), 137-153.

Lewis, M. P. (2009). "Estimating fuel use and emission rates of nonroad diesel construction equipment performing representative duty cycles."

Lucko, G. (2003). "A Statistical Analysis and Model of the Residual Value of Different Types of Heavy Construction Equipment." Virginia Polytechnic Institute and State University.

Manikandan, M. A., M. ; Pazhani, K.C. (2018). "A Study and analysis of construction equipment management used in cinstruction projects for improving productivity." *Chennai, India*, 05(03).

MDOT (2017). "MICHIGAN DEPARTMENT OF TRANSPORTATION asset management." <https://www.michigan.gov/documents/mdot/Asset_Mgmt_Overview_605863_7.pdf>.

Mitchell Jr, Z. W. (1998). "A STATISTICAL ANALYSIS OF CONSTRUCTION EQUIPMENT REPAIR COSTS USING HELD DATA & THE CUMULATION COST MODEL."

Mitchell, Z., Hildreth, J., and Vorster, M. (2010). "Using the cumulative cost model to forecast equipment repair costs: Two different methodologies." *Journal of Construction Engineering and Management*, 137(10), 817-822.

MnDOT (2016). "Maintenance Manual." <http://www.dot.state.mn.us/maintenance/pdf/manual/ch4.pdf>.

Monnot, J. M., and Williams, R. C. (2010). "Construction equipment telematics." *Journal of Construction Engineering and Management*, 137(10), 793-796.

Morris, J. (1988). "Estimation of tractor repair and maintenance costs." *Journal of Agricultural Engineering Research (UK)*.

NCHRP (2018). "Guide to Calculating Ownership and Operating Costs of Department of Transportation Vehicle and Equipment Fleet." <https://apps.trb.org/cmsfeed/TRBNetProjectDisplay.asp?ProjectID=4367>.

ODOT (2012). "GIS in TRANSPORTATION - ODOT Asset management Geospatial tools." *Ohio*, <https://www.gis.fhwa.dot.gov/newsletters/Newsletter_Spring2012.pdf>.

Ok, S. C., and Sinha, S. K. (2006). "Construction equipment productivity estimation using artificial neural network model." *Construction Management and Economics*, 24(10), 1029-1044.

Pantelias, A. (2005). "Asset Management Data Collection For Supporting Decision Processes." Citeseer.

PDRMA (2018). "AssetWORKS: Proprty tracking, accounting and reporting." <https://pdrma.org/member_coverages/property_and_casualty/solutions/asset_works_software>.

Rall, J. (2016). "A 50-State Review of State Legislatures and Departments of Transportation." *Roseland, Indiana*, <http://www.financingtransportation.org/pdf/50_state_review_nov16.pdf>.

Rasdorf, W., Frey, C., Lewis, P., Kim, K., Pang, S.-H., and Abolhassani, S. (2010). "Field procedures for real-world measurements of emissions from diesel construction vehicles." *Journal of Infrastructure Systems*, 16(3), 216-225.

Rise, M. (2016). "What is construction equipment management?" *India*, <https://www.mahindraconstructionequipment.com/blog/what-is-construction-equipment-management>.

Roadsoft (2019). "Roadway Asset Management: Software Suite." <https://www.roadsoft.org/about>.

SAS (2019). "SAS® Visual Analytics 7.3: User's Guide." <http://support.sas.com/documentation/cdl/en/vaug/68648/PDF/default/vaug.pdf>.

Scopatz, R., Zhou, Y., Wojtowicz, A., Carter, D., Smith, S., and Harrison, P. (2014). "Michigan Roadsoft: Integration of State and Local Safety Data."

Sim, S., and Endrenyi, J. (1993). "A failure-repair model with minimal and major maintenance." *IEEE Transactions on Reliability*, 42(1), 134-140.

Support, S. (2019). "Gains and Lift plots for binary-response models." <http://support.sas.com/kb/41/683.html>.

TRB (2019). "Guideline for Decision Making for Repair vs. Replacement of Highway Maintenance Equipment." <https://apps.trb.org/cmsfeed/TRBNetProjectDisplay.asp?ProjectID=4570>.

UCLA (2019). "LOGIT REGRESSION ", <https://stats.idre.ucla.edu/sas/dae/logit-regression/>.

UCLA, i. (2018). "Regression Analysis | SAS Annotated Output." <https://stats.idre.ucla.edu/sas/output/regression-analysis/>.

Yip, H.-l., Fan, H., and Chiang, Y.-h. (2014). "Predicting the maintenance cost of construction equipment: Comparison between general regression neural network and Box–Jenkins time series models." *Automation in Construction*, 38, 30-38.

APPENDICES

//MySQL Queries created for Data Preparation

1.

```
Create table test. maintenance_cost as
select equipment_id,year,sum(completed_cost)
from test.`setup_project_data`
where year between 2010 and 2018
group by equipment_id,year
order by equipment_id,year;
```

2.

```
Create table test.fueling_cost as
SELECT EQUIPMENT_ID,year,sum(FUEL_AMOUNT),sum(FUEL_COST)
from test.`equipment_fueling_data`
where year between 2010 and 2018
group by equipment_id,year
order by equipment_id,year;
```

3.

```
Create table test.maintenance_count as
SELECT t.EQUIPMENT_ID, t.Year, count(t.DATE_COMPLETED) as Maintenance_count
FROM
test.`setup_project_data` as t
group by t.EQUIPMENT_ID, t.Year;
```

4.

Aligning maintenance cost to dollar per hour

```
SELECT * FROM test.maintenance_cost  a
```

RIGHT join test.dollar_hour_data_2010_to_2018 b

on b.EQUIPMENT_ID = a.EQUIPMENT_ID;

5.

*SELECT * FROM test.maintenance_cost a*

RIGHT join test.dollar_mile_data_2010_to_2018 b

on b.EQUIPMENT_ID = a.EQUIPMENT_ID;

6.

*SELECT * FROM test.fueling_cost a*

RIGHT join test.dollar_hour_data_2010_to_2018 b

on b.EQUIPMENT_ID = a.EQUIPMENT_ID;

7.

*SELECT * FROM test.fueling_cost a*

RIGHT join test.dollar_mile_data_2010_to_2018 b

on b.EQUIPMENT_ID = a.EQUIPMENT_ID;

8.

*SELECT * FROM test.maintenance_count a*

RIGHT join test.dollar_hour_data_2010_to_2018 b

on b.EQUIPMENT_ID = a.EQUIPMENT_ID;

9.

*SELECT * FROM test.maintenance_count a*

RIGHT join test.dollar_mile_data_2010_to_2018 b

on b.EQUIPMENT_ID = a.EQUIPMENT_ID;

10.

*SELECT * FROM test.working_hours a*

RIGHT join test.dollar_mile_data_2010_to_2018 b

on b.EQUIPMENT_ID = a.EQUIPMENT_ID;

11.

```
SELECT * FROM test.working_hours  a
RIGHT join test.dollar_hour_data_2010_to_2018 b
on b.EQUIPMENT_ID = a.EQUIPMENT_ID;
```

VITA

Arpan Chaudhary

Candidate for the Degree of

Master of Science

Thesis: DEVELOPING PREDICTIVE MODELS FOR FUEL CONSUMPTION AND MAINTENANCE COST USING EQUIPMENT FLEET OPERATIONAL DATA

Major Field: Civil Engineering

Biographical:

Education:

Completed the requirements for the Master of Science in Civil Engineering at Oklahoma State University, Stillwater, Oklahoma in July 2019.

Completed the requirements for the Bachelor of Science in Civil Engineering at Galgotias University, Greater Noida, U.P./India in 2015.

Experience:

Graduate Teaching/Research Assistant at Oklahoma State University, Stillwater, Oklahoma, USA from August 2017 to May 2019.

Field Engineer summer internship, Kleinfelder, Tulsa, Oklahoma, USA from June 2018 to August 2018.

Field Engineer, Nigrah Developers Pvt. Ltd., India, from September 2015 to October 2016.