BOTSWANA INTERNATIONAL UNIVERSITY
OF SCIENCE & TECHNOLOGY

PREDICTION OF FUEL CONSUMPTION OF HAULAGE TRUCKS IN OPEN PIT MINES

BY

GOMOLEMO KAONE TADUBANA

STUDENT NUMBER: 12000987

MEng MINING ENGINEERING

A THESIS SUBMITTED TO THE FACULTY OF ENGINEERING AND TECHNOLOGY IN FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF MASTERS OF ENGINEERING IN MINING ENGINEERING OF BIUST

MARCH 2021

BOTSWANA INTERNATIONAL UNIVERSITY OF SCIENCE & TECHNOLOGY

FACULTY OF ENGINEERING AND TECHNOLOGY

DEPARTMENT OF MINING AND GEOLOGICAL ENGINEERING

A THESIS REPORT ENTITLED:

PREDICTION OF FUEL CONSUMPTION OF HAULAGE TRUCKS IN OPEN PIT MINES

BY

GOMOLEMO KAONE TADUBANA

STUDENT NUMBER: 12000987

tadubana.gomolemo@studentmail.biust.ac.bw

SUBMITTED IN FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF MASTERS OF ENGINEERING IN MINING ENGINEERING

THESIS SUPERVISORS:

Prof. Raymond S. Suglo

Dr. Boyce Sigweni

March 2021

DECLARATION AND COPYRIGHT

I, **Gomolemo Kaone Tadubana** declare that this thesis is my own work. It is being submitted for the degree of Master of Engineering in Mining Engineering in the Botswana International University of Science and Technology (BIUST), Palapye. It has not been submitted for any degree or examination in any other University.

Signed by:



Gomolemo Kaone Tadubana (ID: 12000987), On the **17th** day of **March** Year **2021**

The Disseration is a copyright material protected under the Berne Convention, the copyright Act of 1999 and other international and national enactments, in that behalf, on intellectual property. It must not be reproduced by any means , in full or in part , except for short extracts in fair dealing; for researcher private study, critical scholarly review or discourse with an acknowledgement, without the written permission of the Office of the DVCAA, on behalf of both the author and the BIUST.

CERTIFICATION

The undersigned certifies that they have read and hereby recommends for acceptance by the Faculty of Engineering and Technology a dissertation titled **Prediction of Fuel Consuption of Haulage Trucks in Open Pit Mining,** in fulfilment of the requirement for the degree of Masters of Engineering in Mining Engineering of the BIUST.

Prof. Raymond S. Suglo
(Supervisor)
Date: 23/03/2021

Dr Boyce Sigweni
(Co – Supervisor)
Date: 23/03/2021

ABSTRACT

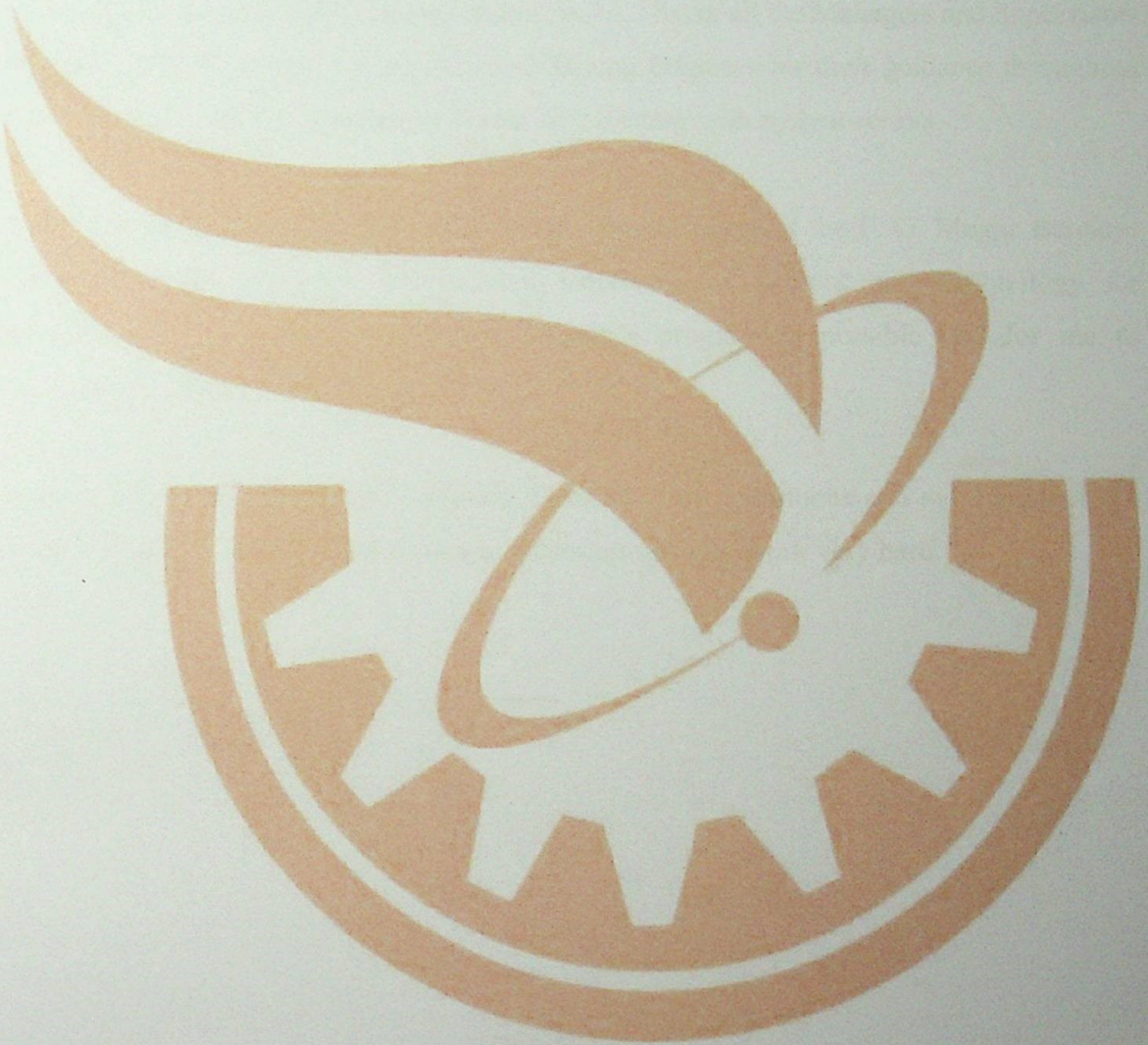
Fuel consumption by haulage trucks is one of the costliest components in open-pit mining. Reducing fuel consumption by trucks could lead to large savings in materials handling costs. This study sought to investigate the factors that affect fuel consumption by haul trucks and to propose an algorithm for predicting fuel consumption by trucks in open-pit mines. Case-Based Reasoning (CBR) methods namely case-based reasoning using forward sequential selection (CBR-FSS), traditional CBR and Naïve techniques were used to predict fuel consumption using datasets from Komatsu and Caterpillar trucks operating at Orapa Mine.

The results show that the CBR method can be used to predict fuel consumption by trucks in open-pit mines; the predicted values of fuel consumption using the CBR-FSS technique gave better results in terms of much lower absolute residual values on all the datasets used, gave higher standardised accuracy values and effect sizes than the other prediction techniques. Thus, the CBR-FSS can be used as a mine planning tool to predict fuel consumption using available data from the previous trips.

It is recommended that mine managers use the CBR method in predicting fuel consumption by haulage trucks to reduce fuel costs and identify trucks with high fuel consumption for early maintenance work to be done on them. Finally, larger datasets from other mines with effects of weather should be used in future studies to improve the accuracy of the predictions and make the findings more applicable to all conditions in open-pit mines.

DEDICATION

This thesis is dedicated to the memory of my daughter Wenzi Loapi Mapolelo-Tadubana. You will always be remembered.



ACKNOWLEDGEMENT

This thesis would not have been possible without the help of my Supervisor, Prof. Raymond Suglo and co-Supervisor, Dr. Boyce Sigweni, as they provided unconditional assistance, patience and excellent working condition throughout this research.

I also extend my sincere appreciation to the management of Debswana Orapa Mine for allowing me to conduct this research at their mine. I thank all the Managers and Supervisors especially Mr. Kgwangu Sithelo, a Senior Mining Engineer for their guidance throughout the mine visits and Mr. Tommasso Turchi for assisting with python scripts.

I also extend my deepest appreciation to the administrators of the F. G. Mogae Barclays Scholarship and Botswana International University of Science and Technology for providing the funds which made my Master's programme possible and for me to successfully complete this thesis.

Lastly, I thank my mother, Ms. Florida Tadubana, for continuous and solid support. To Pulane Mapolelo, thank you for always encouraging me to work very hard.

TABLE OF CONTENTS

Contents	Page
DECLARATION	**ii**
CERTIFICATION	**iii**
ABSTRACT	**iv**
DEDICATION	**v**
ACKNOWLEDGEMENT	**vi**
CHAPTER 1 INTRODUCTION	**1**
1 BACKGROUND	1
1.1 Problem Statement	2
1.2 Research Rationale	2
1.3 Research Objectives	3
1.4 Research Questions	3
1.5 Expected Outcomes	4
1.6 Methods Used	4
1.7 Scope of Work	6
1.8 Thesis Organisation	6
CHAPTER 2 GENERAL INFORMATION ABOUT AREA	**7**
2 HISTORY OF MINE	7
2.1 Geology of Mine	8
2.2 Orapa Mine's Value Chain	8
2.3 Haulage System at Orapa Mine	10
2.4 Summary	11
CHAPTER 3 LITERATURE REVIEW	**12**
3 INTRODUCTION	12
3.1 Haulage Truck Fuel Consumption Parameters	12
3.2 Introduction to Machine Learning	21
3.3 Analogy Based Software Effort Estimation	31
3.4 Methods of Accuracy Estimation and Validation Techniques	35

3.5 Summary 44

CHAPTER 4 DATA COLLECTION AND ANALYSIS 44

4 INTRODUCTION 44

4.1 Data Collection 44

4.2 Results and Analysis 48

4.3 Discussions 59

CHAPTER 5 CONCLUSIONS AND RECOMMENDATIONS 61

5.1 Conclusions 61

5.2 Recommendations 61

REFERENCES 62

APPENDICES 74

APPENDIX 1 SAMPLE OF CAT TRUCK #29 DATASET 75

APPENDIX 1 SAMPLE OF CAT TRUCK #29 DATASET 76

APPENDIX 3 RANDOM SAMPLE OF ACTUAL DATASET FOR CAT TRUCK #3 (AFTERNOON SHIFT) 77

APPENDIX 4 RANDOM SAMPLE OF ACTUAL DATASET FOR CAT TRUCK #3 (NIGHT SHIFT) 78

APPENDIX 5 RANDOM SAMPLE OF ACTUAL DATASET FOR KOMATSU 830E TRUCK #1 79

APPENDIX 6 RANDOM SAMPLE OF ACTUAL DATASET FOR CAT 793 TRUCK #2 82

LIST OF FIGURES

Figure	Title	Page
Figure 1	Location of Orapa Diamond Mine	7
Figure 2	Layout of Orapa Diamond Mine Pit	9
Figure 3	Sketch of Key Factors that Affect Haul Truck Performance	13
Figure 4	Features Influencing Driving Behaviour	14
Figure 5	Effects of Rolling Resistance on Production and Fuel Cost	15
Figure 6	Truck Cycle Time	17
Figure 7	Average Tonnage vs. Litres per Tonne	17
Figure 8	Case Based Reasoning Flowchart	25
Figure 9	Filter Methods	27
Figure 10	Wrapper Method	28
Figure 11	Wrapper Approach to Feature Weight Set Selection	30
Figure 12	Schematic of ANGEL	32
Figure 13	*k*-Nearest Neighbour	34
Figure 14	Validation Data Set using Growing Window	36
Figure 15	Evaluation Data Set using Moving Window	37
Figure 16	Python Code for Creating Windows	47
Figure 17	Python Code for Random Guessing Prediction	48
Figure 18	Distribution of Residuals for CAT Truck #29 Dataset	50
Figure 19	Distribution of Residuals for CAT 793 Truck #3 Morning Shift Dataset	52
Figure 20	Distribution of Residuals for CAT 793 Truck #3 Afternoon Shift Dataset	52
Figure 21	Distribution of Residuals for CAT 793 Truck #3 Night Shift Dataset	53
Figure 22	Boxplots of Distribution of Residuals for Komatsu 830E Truck #1 Dataset	56
Figure 23	Boxplots showing Distribution of Residuals for CAT 793 Truck #2 Dataset	57
Figure 24	Absolute Residuals for CAT 793 Truck #2 Dataset	59

LIST OF TABLES

Table	Title	Page
Table 1	Mining and Off-Highway Trucks Engine Load Factors (LF)	21
Table 2	Advantages and Disadvantages of Feature Selection Methods	30
Table 3	Classes of Effect Size	42
Table 4	Selected Datasets for Scenario 2	46
Table 5	Selected Datasets for Scenario 3	46
Table 6	ArchANGEL Setup	48
Table 7	Random Sample (10%) Actual Fuel Used, Predicted Fuel Used and Absolute Errors for CAT Truck #29 Dataset	49
Table 8	Mean Absolute Residual CAT Truck #29 Dataset	50
Table 9	Standardised Accuracy and Effect Size CAT Truck #29 Dataset	51
Table 10	Mean Absolute Residuals for Shift Prediction for CAT 793 Truck #3 Datasets	52
Table 11	SA and Effect Size for CAT 793 Truck #3 Morning Shift Dataset	54
Table 12	SA and Effect Size for CAT 793 Truck #3 Afternoon Shift Dataset	54
Table 13	SA and Effect Size for CAT 793 Truck #3 Night Shift Dataset	55
Table 14	Mean Absolute Residual of Komatsu 830E #1 and CAT 793 #2 Datasets	56
Table 15	SA and Effect Size for Komatsu 830E Truck #1 Dataset	57
Table 16	SA and Effect Size for CAT 793 Truck #2 Dataset	58

CHAPTER 1

INTRODUCTION

1 BACKGROUND

Open-pit mining is a process of extracting useful minerals from near-surface deposits by excavation or cut made at the surface (Wetherelt and van der Wielen, 2011; Mero *et al.*, 2017). Open-pit mining methods generally employ a conventional mining cycle of operations to extract the minerals. These include drilling and blasting for rock fragmentation and, loading and hauling for materials handling.

In any mine today, managing the operations cautiously results in the profitability of a mine (Coronado, 2014). Research shows that truck haulage accounts for a significant amount of production costs (≈ 26%). In addition, together with shovels, they constitute up to 40% of the total cost (Bishop, 1968; Michaelson, 1979; Sullivan, 1990). Fuel consumption is one of the primary operating costs in shovel-truck operations, with fuel cost representing a significant component of materials handling cost (Dumakor *et al.*, 2017). Hence it is necessary to reduce fuel consumption by haulage trucks to minimise mining costs. The need for cost-saving has driven both mine operators and governments to conduct multiple research studies on how to reduce energy consumption in mining (Soofastaei, 2016).

Fuel efficiency is the relationship between fuel consumption and machine productivity. It is also defined as the ratio of the amount of useful work done to the energy input (Awuah-Offei, 2016). It requires measuring both the amount of fuel consumed and productivity. Fuel efficiency is more important due to the increasing cost of fuel prices on the world market. In mining operations, fuel is used by heavy machinery in the excavation and handling of materials. Excavator-truck combination is the most common materials handling method in most open-pit mines worldwide. The trucks consume a lot of fuel. This research attempts to investigate and predict the fuel consumption of haulage trucks in open-pit mines with reference to Debswana Orapa Mine in Botswana.

In open-pit mines, dump trucks are used along with excavators to handle the ore and waste from the production site to either the processing plant or waste dumps (Sahoo *et al.*, 2010).

The fuel consumption of dump trucks is affected by the distance and nature of the haul roads, payload, speed, rolling resistance and topography of the area. Other factors include empty and gross weight, power, acceleration, idle time, aerodynamics, tyre quality, wheel alignment and inflation pressure of tyres, fuel quality, the operator's efficiency, outside temperature, weather and quality of a truck's maintenance program (Kecojevic and Komljenovic, 2010).

Most of these parameters can be controlled by the mine operators to some degree. Adequate management of these factors may significantly reduce the fuel consumption by haulage trucks while providing the required truck efficiency, without important investments or operational changes (Kecojevic and Komljenovic, 2010).

1.1 Problem Statement

Truck haulage constitutes a major portion of the total cost in open pit mining operations due to rising fuel prices (Soofastaei, 2016). High fuel consumption by large haulage trucks mainly occurs during the transportation of ore and waste and when the trucks are idling. During idling, there is no production by both the truck and the operator, yet the engine is kept running.

Predicting fuel consumption of haulage trucks may assist with easy identification of a truck with irregular fuel consumption in mine planning. It would therefore be easy to recognize the cause of abnormal fuel consumption that may be due to driver behavior, changes in road conditions, etc. This will help the management to think ahead and come up with corrective measures to reduce fuel consumption.Therefore, the prediction of fuel consumption by haulage trucks could lead to a significant reduction in haulage cost.

1.2 Research Rationale

This research work will assist:

i. The management of mines with a system that can be used to predict fuel consumption.
ii. In planning for how much fuel is going to be required and hence identify the haulage truck with abnormal fuel consumption.

1.3 Research Objectives

The objectives of this research are to:

i. Investigate the factors that affect the fuel consumption of haul trucks in open-pit mines.
ii. Propose and implement an algorithm for predicting fuel consumption by haul trucks in open-pit mines.
iii. Evaluate the system against benchmark in terms of accuracy and performance.

1.4 Research Questions (RQ)

There are several research issues that need to be addressed to understand the problems of fuel consumption by haulage trucks in open-pit mines and prediction techniques. This research tries to find answers to the following research questions:

RQ1: What are the parameters influencing fuel consumption by haulage trucks in open-pit mines?

Rationale: To determine which parameters are more detrimental to fuel consumption and to identify the controllable and uncontrollable parameters.

Methods: Systematic literature review.

RQ2: What is the mechanism used by Orapa Mine for controlling/ de-queuing trucks during the mining operations?

Rationale: To determine if their queuing theory selects the best route based on fuel consumption, idling, or/and travel times.

Methods: Interviews, Questionnaire and Ethnography.

RQ3: What are the techniques or methods that can be used to predict fuel consumption by haulage trucks in open-pit mines?

Rationale: To predict fuel consumption and to find a technique that mine managers can understand easily.

Method: Systematic literature review.

RQ4: How does the selected prediction technique compare with other methods?

Rationale: To compare different techniques in order to determine the simplest technique for mine managers.

Methods: Different validation methods, effect size and practical significance.

1.5 Expected Outcomes

The expected outcomes of this research are to be able to predict fuel consumption by haulage trucks using the developed prediction models. A well-functioning model for predicting the fuel consumption of haulage trucks in a mine could be used to determine haul routes that will minimise fuel consumption. This could be useful because reducing fuel consumption means reducing operating costs and reducing the environmental impact of dust and exhaust fumes in the operations. The model could also be used to detect and identify haulage trucks that consume a higher than normal amount of fuel. This will enable the early identification and correction of possible defects in the haulage trucks.

1.6 Methods Used

This section outlines all methods used to answer the research questions. A description of how fuel consumption by haulage trucks data is extracted at Orapa Mine is provided and the method for data modelling is discussed.

Survey: In order to determine the variables influencing fuel consumption by haulage trucks in open-pit mines and the techniques used for predicting fuel consumption by haulage trucks, a comprehensive literature review was conducted on similar topics.

The ethnography of the Orapa Mine pit and the time studies of the various activities in the haulage cycle by dump trucks were conducted. The trucks were observed and timed from the loading points through haulage segments in the pit to dumping points than back to loading points.

Interviews: One key objective was to determine the system Orapa Mine uses to track the fuel consumption by haulage trucks. During interviews with the personnel responsible for the fleet management system, it was determined that Orapa Mine uses a Wencomine fleet management system to track the equipment used in mining and haulage operations on the mine.

It was found that the fleet management system does all processes without much manipulation from the personnel (e.g. route allocation, report generation, etc.). However, during route allocation, the fleet management system does not consider the fuel consumed

by the haulage trucks along the various routes. It is configured to find which route will result in the shortest travel times.

Data was gathered from Orapa Mine through reports generated by the service providers (Komatsu and Barloworld Equipment). The data is on the fuel consumed by the various mining equipment and also captures 11 parameters (e.g. payload, cycle time) on a trip basis from Barloworld which is responsible for the CAT 789 and CAT 793 trucks which are of 190 ton and 240 ton respectively.

The data gathered from Komatsu is on the fuel consumption and 37 parameters (e.g. payload, speed, cycle time, etc.) on a trip basis. Komatsu is responsible for the Komatsu 730 and Komatsu 830 trucks which have 190 tons and 240 tons capacities respectively.

Experimental procedure: The steps taken to determine if Case-Based Reasoning (CBR) can be used to predict fuel consumption by haulage trucks are outlined. It is worthy of note that the available records in the literature show that the CBR technique has never been used in the mining industry. The CBR technique uses old knowledge to solve and predict solutions to new problems. To determine if CBR can be used to predict fuel consumption in open-pit mines the following steps were followed:

Case base construction: A case base containing the past trips by a haulage truck is constructed for each scenario. The case base contains trips together with their features and their specifications. These factors are the main contributors to fuel consumption and feature specifications which are the specific values of features that a trip achieved for a certain amount of fuel consumed.

To use case-based reasoning (CBR), a case base (dataset) is needed. Research shows that the CBR technique has not been used to predict fuel consumption by haulage trucks in open-pit mines. Therefore, the dataset does not exist. Therefore, a data set is to be formed which entails the dependent variables, the variable to be predicted (i.e. fuel consumption) and a set of independent variables (i.e. parameters used to predict the dependent variables). Literature was used to come up with independent variables and data obtained from Orapa Mine was used.

Prediction: Prediction was done using ArchANGEL which incorporates CBR shell. This research work followed the Shepperd-MacDonell method (SMM) employed by Ayoung *et al.* (2015) when predicting the Design Reality Gap (DRG) scores for telecentres for evaluating the prediction methods.

1.7 Scope of Work

This research only considered the application of CBR prediction techniques using data collected from Orapa Mine. Mines operate in different environments with geological, climatic and operational conditions. Accordingly, the results obtained by using CBR methods in one mine may be different from those of other mines.

Another limitation of this research is that data on the weather could not be obtained from the weather data provider. Thus, the results obtained from using CBR to predict fuel consumption by haul trucks in Orapa Mine did not include the effects of weather.

1.8 Thesis Organisation

Chapter 1 has the background, statement of the problem, objectives of the research, outcomes of research, etc. Chapter 2 has the general information about Orapa Mine and outlines the current system used in tracking the fuel consumption of haulage trucks at Orapa Mine. Chapter 3 has the literature review which includes parameters influencing fuel consumption by haulage trucks and the techniques used in predicting fuel consumption by haulage trucks in mines. Chapter 4 has the data collection and analysis while Chapter 5 contains the Conclusions and Recommendations arising from the work.

CHAPTER 2

GENERAL INFORMATION ABOUT AREA

2 HISTORY OF MINE

Orapa Mine is the oldest operating mine in Botswana and the largest open-pit diamond mine in the world. The diamond deposits in Orapa were discovered by a team of De Beers geologists in 1967. The mine was officially opened in July 1971 by the then President of Botswana, His Excellency Sir Seretse Khama (Anon., 2018a). The mine is located in central Botswana 240 km west of Francistown as shown in Figure 1.

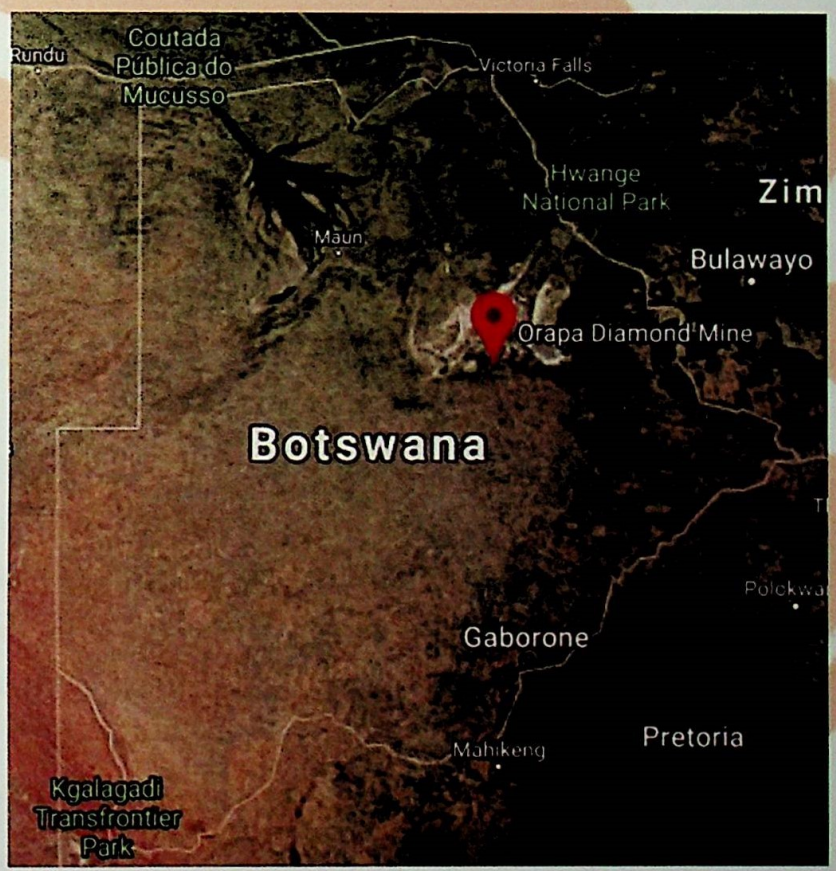


Figure 1 Location of Orapa Diamond Mine (Anon., 2018b)

Orapa Mine is owned by Debswana Diamond Company (Pty) Limited, a collaboration between the Botswana government and De Beers Group of Companies. The company was incorporated on June 23, 1969, under the original name of De Beers Botswana Mining Company (Pty) Limited. The company changed its name on March 23, 1992, to Debswana Diamond Company (Pty) Limited. Debswana is the world's leading diamond producer by

value and Botswana's largest private-sector employer, with over 5,000 employees and over 6,000 contractors. Orapa pit is currently at a depth of 250 m and is estimated to reach 450 m by 2026.

Debswana operates four diamond mines namely Orapa, Letlhakane, Damtshaa and Jwaneng Mines (Anon., 2018c). Through these mines, Debswana has made a tremendous contribution to Botswana's economic growth through diamond revenue, transforming Botswana from an agriculture-based economy in the 1960s into a country that has consistently displayed one of the world's highest economic growth rates.

2.1 Geology of Mine

The Cretaceous Orapa A/K1 is exposed as a bilobate kimberlite pipe with preserved crater lithofacies. The pipe is currently the biggest of all known kimberlite bodies within the Orapa Kimberlite Cluster (Maphane, 2017). Orapa's kimberlite occupies an area of 118 km^2 on the surface and is part of the Karoo Super-group lying above Archaean granite and it consists of Stormberg-aged basalt lavas. The surface intersection of the Orapa kimberlite pipe is ovoid in form, 1,500 m long and 1,000 m wide (Anon., 2018a). A number of these kimberlite bodies are currently being economically exploited (e.g. Orapa A/K1, Damtshaa B/K9 and B/K11 and Letlhakane D/K1 and D/K2) or at advanced evaluation stage (Maphane, 2017). The Orapa kimberlite has the world's second-largest diamond-producing kimberlite pipe, which is made up of two separate pipes (or intrusions) that coalesced near the surface about 93 million years ago. Orapa Mine is projected to generate about 153 million carats out of 205 million tons of mineral over its 15-year life. (Anon., 2018a). Figure 2 shows the current layout of the Orapa Diamond Mine.

2.2 Orapa Mine's Value Chain

The value chain of the mine includes all the processes that gradually add value from mining to sales to the diamond resource. Additional value is added as the diamond resources pass through each step. These phases include (Anon., 2018d):

1. Mining: Drilling and blasting operations are used to break down millions of tons of ore and waste. Large shovels then load the kimberlite (ore) and waste rock into trucks and carry them to the primary crusher and waste dumps, respectively.

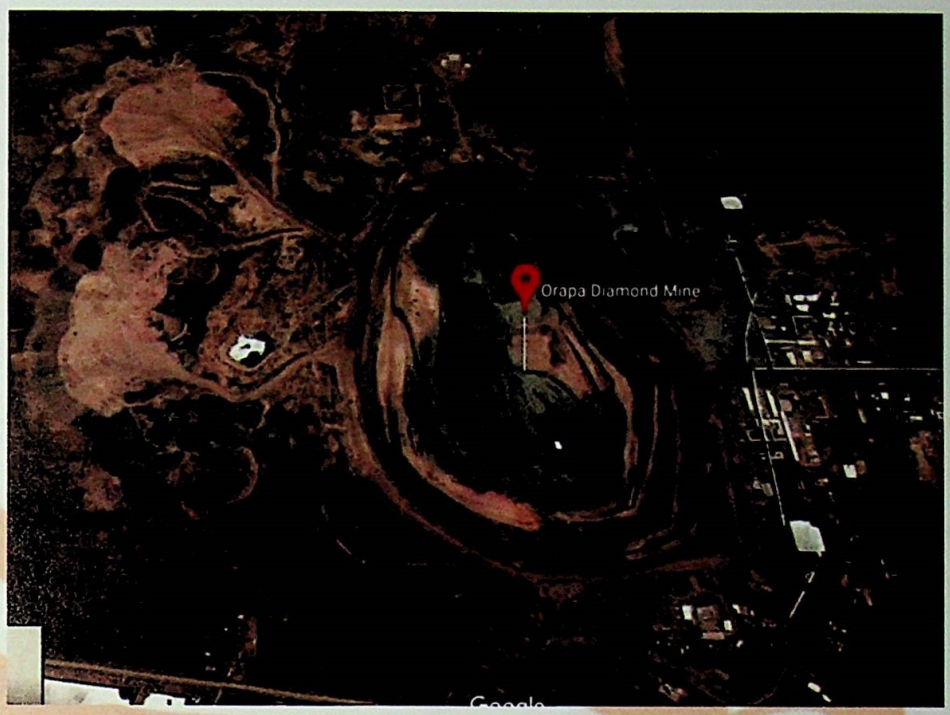


Figure 2 Layout of Orapa Diamond Mine Pit (Anon., 2018b)

2. Crushing: Ore is broken down through different stages of crushing and screening into controllable sizes to the size of the rock fed to the processing plant and release diamonds from the host kimberlite rock.
3. Treatment: Diamonds are denser than the rock that contains them. In a method known as Dense Media Separation (DMS), heavy minerals, including diamonds, are segregated from light minerals until the material has been reduced to the right sizes and diamonds are liberated. To float light minerals and isolate them from heavy minerals, a Dense Medium Separation technique known as ferrosilicon-powdered iron and silicone in water is used. DMS is one of several pre-concentration methods used for early rejection of waste from run- of- mine ores running at relatively coarse particle sizes before additional milling and beneficiation.
4. Recrush: The remaining material is forwarded to the Recrush Plant from the DMS Plant. The ore is recrushed using a more advanced process. This plant helps release the smaller non-liberated diamonds from the initial treatment process. The process yields a concentrate.
5. Recovery: Diamonds have many remarkable properties that can be utilised to recover them from the concentrate prepared source. They emit light under X-rays that allow detection and separation of them (Moore, 2009). Diamonds repel water and are drawn

to grease. The diamonds bind to the grease when the concentrate is mixed with water and passed over a grease belt. Diamonds fluoresce under laser sorters that are used to further concentrate it within the processing stream.

6. Sorting and Sales: Sorting, washing, packing and weighing of diamonds takes place in an automated Fully Integrated Sorthouse (FISH). FISH maximises the protection of diamonds while improving the sorting of diamonds using laser technology. The diamonds are then sold to the DTC Botswana in Gaborone. At DTC Botswana, the diamonds are sorted into more than 12,000 different categories before being sold to De Beers by the DTC Board, who in turn sells them to their clients in Gaborone. The rough diamonds then find their way to cutting and polishing centres where they are sold to manufacturers producing diamond jewellery for the leading consumer markets.

2.3 Haulage System at Orapa Mine

Orapa Mine uses a fleet management system called Wencomine provided by Wenco International Mining Systems for its mining fleet. Wenco International Mining Systems provides GPS-based fleet management systems for open-pit mines worldwide.

Wencomine Fleet Management System (FMS) meets the open pit needs of the mine by allocating vehicles, managing payloads, streamlining haul cycles and monitoring other mining operational details. Wenco has a tight integration that includes fuel level sensors, tyre pressure monitoring systems and payload managers that help to equip dispatch with the requisite control, expertise and understanding to minimize costs, increase safety and profit sustainability across the operation (Wells, 2018). Wenco also helps the mine to control off-routes trucks, engine overspeeds and payloads. Every truck, shovel and auxiliary unit is completely controlled through the Wencomine drag-and-drop interface. Switching tasks, reading reports and responding to warnings from the original equipment manufacturer is as simple as clicking on a button (Wells, 2019a). Wencomine supports decision taking that resets the mine performance. With Wencomine it is possible to track key performance indicators (KPIs) such as waiting time, cycle time, payloads, availability and quality of grade.

Wencomine entails a dynamic dispatch that operates in two stages. First, to generate a schedule for trucks and shovels, it runs the parameters of the mine through the dispatching

algorithm. Then it sends the task to the fleet to execute the plan. The service draws on the mine's data stored in the Wencomine database. Therefore, plans often adhere to specific requirements. Dynamic dispatch chooses the arrangement that shrinks haul cycle time, reduces the number of truck needs, shortens waiting times and produces more material. Dispatchers see a mine projection and change according to need (Wells, 2019b).

Wencomine is equipped with advanced GPS-based navigation called PitNav or turn by turn navigation which searches for route congestion, blast delays and other obstacles to travel. PitNav then redirects the haulage trucks to alternative routes conserving hundreds of hours. PitNav navigates around hazardous sites then excludes equipment from these specific GPS coordinates, keeping operators and equipment safe. PitNav appears on the haul truck's onboard touchscreens. Operators follow the responsive turn indicators until they arrive at the assigned destination (shovels, dumps and tie-down locations). Direction updates are in real-time and allow the equipment to travel with great efficiency (Wells, 2019c).

2.4 Summary

This section gave the background of the study area Orapa mine and further addresses research question 2 which seeks to determine the mechanism used by Orapa Mine for controlling haulage trucks during operations. Wencomine systems, a GPS-based fleet management system, is used. The system selects the best routes for haulage trucks based on travel time not considering the amount of fuel that is going to be consumed.

CHAPTER 3

LITERATURE REVIEW

3 INTRODUCTION

According to Siami-Irdemoosa and Dindarloo (2015) and Rodovalho *et al.* (2016), the main haulage equipment used in open pit mines are dump trucks and they constitute a sizeable proportion of both capital and operating costs. Fuel consumption in surface mines accounts for about 30% of overall energy costs. Research by Coffey (2016) shows that an increase in engine speed reduces efficiency, whereas an increase in torque increases engine efficiency. Coffey (2016) further note that if fuel consumption is a function of driving resistance and engine efficiency then it is necessary to recognise that driving resistance is a function of road conditions and driver behaviour while engine speed is a function of vehicle speed and gear position.

3.1 Haulage Truck Fuel Consumption Parameters

The findings from various researchers show that the factors affecting fuel economy can be categorised into four main groups which are vehicle performance, road traffic conditions, environmental conditions and driver operating behaviour. However, Lin *et al.* (2011) and Awuah-Offei *et al.* (2012) have grouped the factors affecting fuel consumption into three main categories namely operator's practice, operating conditions and equipment and road conditions and driving behaviour Their studies also suggest that fuel consumption is not related to the vehicle's design, manufacturing and assembly. A paper on fuel consumption and carbon dioxide emissions from passenger cars in Europe in the laboratory and real-world emissions categorised the factors affecting fuel consumption into three groups as (Fontaras *et al.*, 2017):

i. Vehicle characteristics and systems-related factors. These factors are based on the key energy consumption contributors, which determine the fuel consumption.
ii. Environmental and traffic conditions. This includes factors such as weather conditions, the morphology of the road and traffic conditions.
iii. Vehicle driver-related factors such as driving style and vehicle maintenance.

A study by Soofastaei *et al.* (2016) into reducing fuel consumption of haul trucks in surface mines using artificial intelligence models found that the factors that contribute to the fuel consumption by haul trucks include haul road properties and conditions, condition of trucks, truck speed, rolling resistance and environmental conditions (see Figure 3).

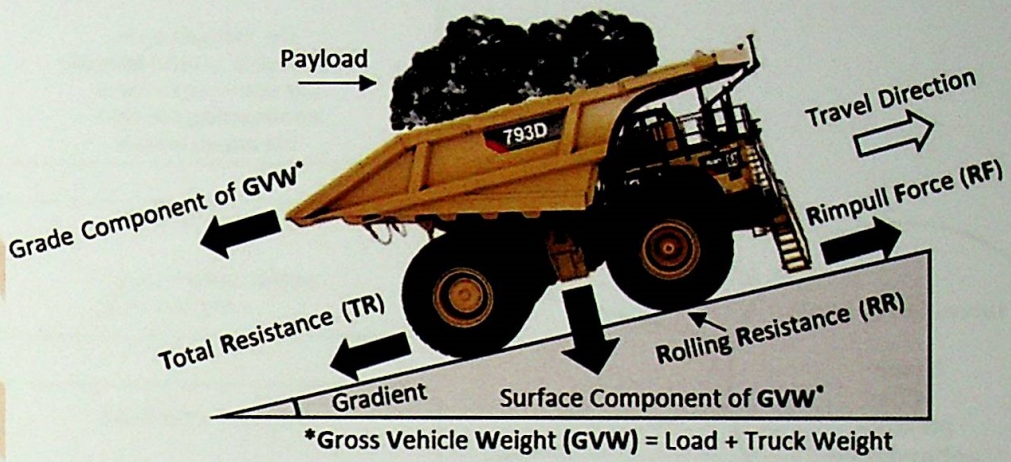


Figure 3 Sketch of Key Factors that Affect Haul Truck Performance (Soofastaei *et al.*, 2016)

3.1.1 Driving Style

Several studies have been conducted on the effects of driving style on fuel consumption. Driving behaviour refers to a driver's personal driving style and is characterised by rapid and average speed, acceleration and gear selection (Brundell-Freij and Ericsson, 2005). A report by Rodovalho *et al.* (2016) also demonstrates how driving style is affected by the performance of the truck operator, which varies under conditions of acceleration, braking, cornering, variable speed and manoeuvring. A study by Carpatorea *et al.* (2016) found that around 30% of the operational cost for the fleet is as a result of the driver. Therefore, having efficient drivers is important to cut down fuel consumption in haulage operations. According to Ishibashi *et al.* (2007), many parameters such as driving environment, traffic situation, driving context, driver's condition and characteristics affect the driver's behaviour. The driver's characteristics include gender and age, driving experience and habits such as years of working experience at the particular mine site, involvement in traffic accidents, driving frequency and annual mileage. A study by Zacharof *et al.* (2016) on the factors affecting fuel consumption and carbon dioxide emissions by passenger cars concluded that aggressive driving increases fuel consumption and this is mainly characterised by high acceleration and high speed. A study by Rolim *et al.* (2017) on urban

bus operators in Lisbon, Portugal, showed that driving behaviour can be influenced by the different factors shown in Figure 4.

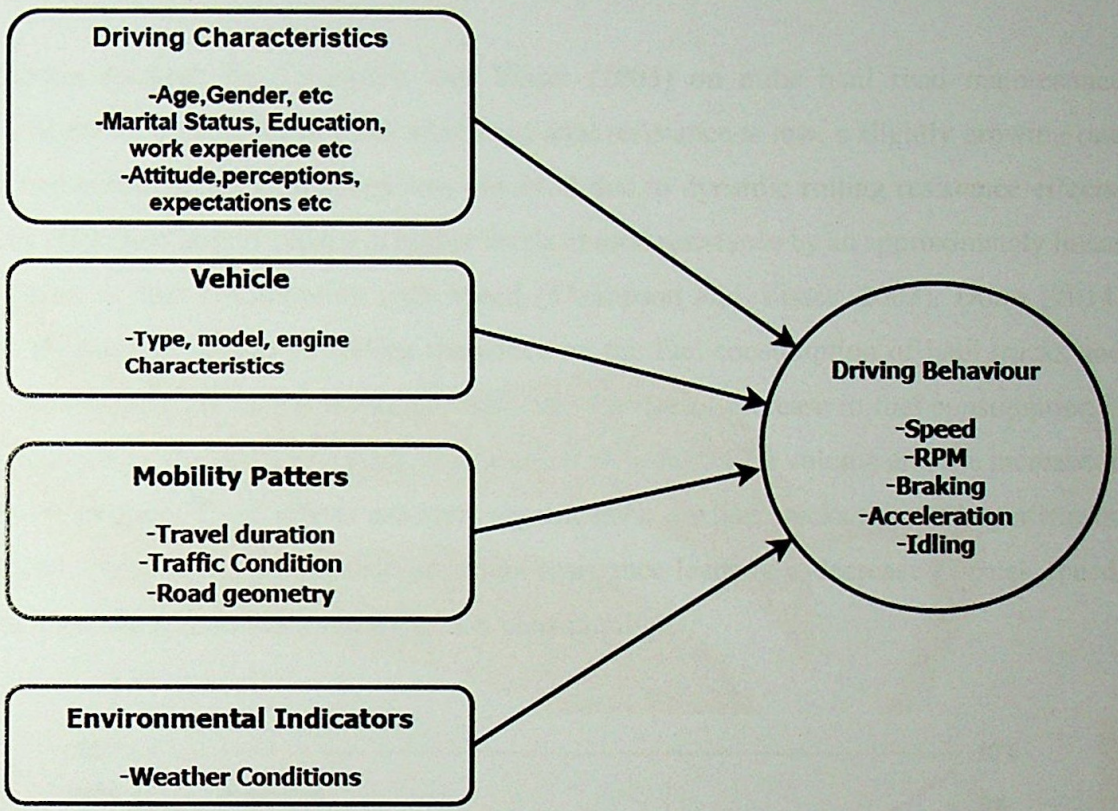


Figure 4 Features Influencing Driving Behaviour (Rolim *et al.*, 2017)

3.1.2 Rolling Resistance

Rolling resistance (RR) is defined as a measure of the force required to overcome the retarding effect between the tyres and road (Adair *et al.*, 2015). It occurs when the vehicle is moving on the road's surface and is due to the interaction between the tyres and road material (Wicaksana *et al.*, 2011). Rolling resistance is a parameter that affects fuel consumption and for a vehicle to accelerate or maintain speed, it needs to be overcome. The work by Paterlini *et al.* (2015) shows that rolling resistance accounts for 10 - 13% of fuel consumption of the vehicle. According to Adair *et al.* (2015) and Soofastaei *et al.* (2017), rolling resistance is a significant factor affecting productivity, cycle times, fuel consumption, maintenance and safety of haul truck operations in open-pit mines. It is also known that haul road maintenance, tyre pressure and truck speed have a great influence on rolling resistance (Adair *et al.*, 2015). Haul road maintenance is used to repair defects and other road surface issues on mine haul roads. Maintenance intervals are determined by

various aspects including mine schedules, financial requirements and constraints, availability of personnel and equipment, weather conditions and the quality/condition of the haul roads (Thompson and Visser, 2003).

Another research by Thompson and Visser (2003) on mine haul road maintenance management systems shows that where the total resistance is low, a slightly growing rate of fuel consumption with speed was observed due to dynamic rolling resistance effects. This effect was largely hidden at higher levels of total resistance by an approximately linear increase in fuel consumption with speed (Thompson and Visser, 2003). Dotto (2014) investigated the impact of rolling resistance on the fuel consumption of haul trucks and found that when it comes to rolling resistance, the rate of increase in fuel consumption is influenced by the surface wearing course material used, traffic volume and the increase in total resistance. These effects are more significant in smaller trucks than in larger trucks. Figure 5 shows that an increase in rolling resistance leads to a decrease in truck speed, lower production and an increase in fuel consumption.

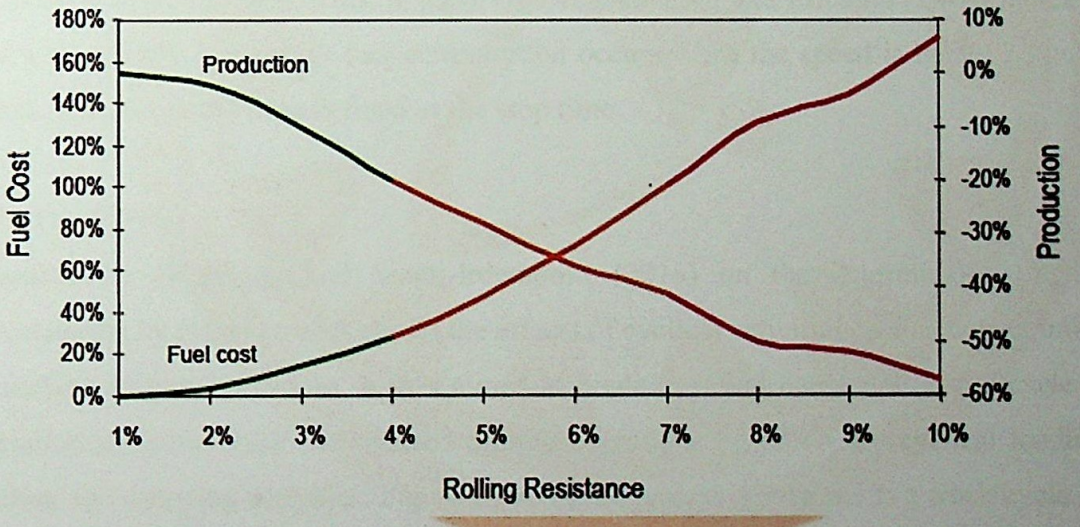


Figure 5 Effects of Rolling Resistance on Production and Fuel Cost (Holman, 2006)

The Total Resistance (TR) is equal to the sum of the Rolling Resistance (RR) and the Grade Resistance (GR) when the truck is moving up a haul road (Soofastaei *et al.*, 2016). Theoretically, total resistance can be calculated using equation (1):

$$TT = RR + GR \quad (1)$$

Grade Resistance (GR) is the amount of force required to overcome gravity (Johansson, 2016). It is measured as a percentage and calculated as the ratio between the rise of the road and the horizontal length.

3.1.3 Speed

A paper by Richardson and McIver (2015) on improving mine haul road surfaces and ways to reduce haul truck fuel efficiency showed that fuel consumption by trucks is dictated by the conditions of the haul roads as transmission shift is minimised by smooth constant grade, which allows the driver to maintain high speed and allowing constant braking efforts on returns which reduce fuel consumption. On the other hand, poorly designed or maintained roads lead to dramatic cost increases due to higher fuel consumption. Another study by Dey *et al.* (2017) reveals attempts by drivers to run the trucks on flatter surface roads under steady-state conditions. However, observations show that a few drivers change the speed with frequent braking under poor road conditions resulting in varying speeds. It is also shown that fuel consumption increases due to variation of speeds on the haul roads and the speeds between 20 - 30 km/hr give the best fuel economy and safe driving conditions along the haul roads. A paper by Brundell-Freij and Ericsson (2005) indicates that a significant increase in fuel consumption occurs when the speed is below 2 km/hr, which Ericsson (2001) has defined as the stop time.

3.1.4 Cycle Time

Research by Dindarloo and Siami-Irdemoosa (2016) on the determinants of fuel consumption by mining trucks shows the effects of cyclical activities on fuel consumption in surface mining operations. It was aimed at predicting fuel consumption per cycle of operation of dump trucks at selected mining operations based on the cyclical loading, hauling and dumping activities. Figure 6 shows the typical operations in a truck cycle. In materials handling, each component of the cycle is associated with a specific portion of the total fuel consumed per cycle. The functional relationship between the cycle components and the amount of fuel consumed per cycle determines the proportion of each component to the total consumption of fuel by truck. Work by Soofastaei *et al.* (2016b) also shows that the level of fuel consumption in surface mines is a function of typical operations in the truck cycle.

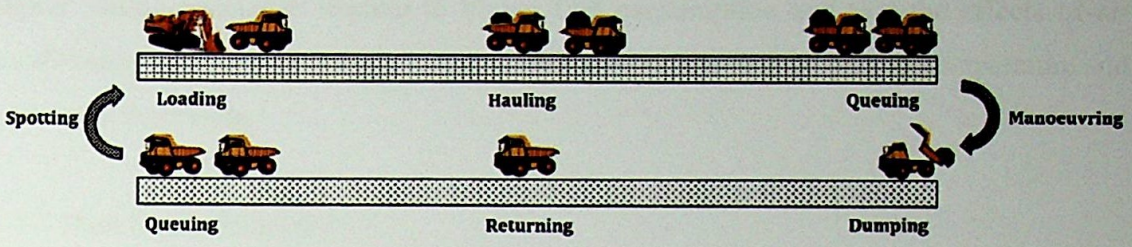


Figure 6 Truck Cycle Time (Soofastaei *et al.*, 2016b)

3.1.5 Payload

A paper by Soofastaei (2016) shows that variance in the payload of haul trucks affect their fuel consumption. It also indicates different parameters that affect the loading of trucks include blast design, degree of fragmentation of the muck pile, operator's efficiency and mine planning. It also states that optimal payload must be strived for by the operator during the loading of haul trucks. Another research on factors affecting fuel economy by Smith (2012) suggests that a heavily loaded truck will consume more fuel than a lightly loaded truck. However, a study by Motlogelwa and Minnitt (2013) showed that fuel consumption in litres per ton decreases with increasing tonnage (see Figure 7).



Figure 7 Average Tonnage vs. Litres per Tonne (Motlogelwa and Minnitt, 2013)

3.1.6 Weather Conditions

Weather conditions also have a significant contribution to the fuel consumption of haulage trucks in mines. The findings of Fontaras *et al.* (2017) show that in winter, tyres exhibit

higher rolling resistance leading to higher fuel consumption and that the effects of air conditioning on fuel consumption depend primarily on the desired interior temperature and ambient temperature.

3.1.7 Haul Road Geometry

A road network for mines typically consists of 10 - 25 km of road segments with variable traffic volumes and quality of construction materials. If the road surface is not regularly repaired, there will be rutting and haulage segments where vehicles have to slow down their speeds to deal with adverse conditions (Kaufman and Ault, 1977). Holman (2006) concluded that poorly maintained haul roads raise operating costs in terms of maintenance and fuel costs.

3.1.8 Tyre Condition

Tyre costs account for 18% - 22% of the total operating cost of a haulage fleet (Knights and Boerner, 2001; Meech and Parreira, 2013). The findings from Thompson (1979) show that fuel consumption is affected by tyres because of the force and the work required from the engine to propel the vehicle. Furthermore, when a truck travels on level road energy is dissipated in the truck's tyres or in overcoming the aerodynamic drag of the truck.

3.1.9 Idling Time

Truck idling refers to running a vehicle without moving the vehicle (Anon., 2015). During idling, the engine often continues to partially burn fuel, resulting in the build-up of residues on engine parts, inefficient fuel consumption and mechanical degradation. (Rahman *et al.*, 2013). Idling-affected components include engine gas reticulation valves and diesel fluid filter. When grounded, the damage caused by idling can result in costly repairs and the mine loses returns on its investment. Brady and Van Order (2012) found that truck idling is a great source of air pollution, higher operating costs and greater fuel consumption. The work by Motlogelwa and Minnitt (2013) shows that as a result of queuing, Uitvlugt mine in Guyana loses a lot of diesel when the haul trucks are left idling for long periods.

Ore and waste loading are conducted simultaneously in multi-bench pits at several different locations and often in several different pits. Different sizes of excavators are used to load materials in surface mines into trucks. Hauling of materials from production sites to the dumping sites must be done by means of a network of roads of varying lengths and grades.

Haul roads can be highly complex, cover wide areas of land and pass through extreme changes in elevation. Excavator load times depend on excavator capacity, digging conditions, truck capacity and the operator's efficiency. Queues may also form at the excavator locations when the individual shovels are allocated too many trucks (Necmettin, 2004). Production scheduling can reduce idle times of haulage trucks and excavators (Darling, 2011). Rahman *et al.* (2013) state that truck drivers idle their engines to maintain cab comfort to obtain the power needed for accessories such as air conditioners, heaters, refrigerators and lights. The idling of engines increases their fuel consumption.

A paper by Lutsey *et al.* (2004) states that when idling, the amount of fuel consumed while the vehicle is idling can be estimated by the engine speed. Engine speed (rpm) has a substantial effect on the fuel consumption of heavy-duty trucks that idle. A study was undertaken on reducing stoppage time for haul trucks. The analysis included different operating modes of a haul truck like fuel consumption, vehicle speeds, haul distances and systems. The study showed that trucks spend more time at stopping points while empty than when loaded (Hargroves *et al.*, 2014).

3.1.10 Queueing Theory

Queueing theory deals with waiting of humans, units, or entities at a given location for a service to be done to them. Queueing is very common in many fields such as petrol stations, banks and computer systems. Queueing theory is a field that applies probability to queueing problems in traffic engineering, telecommunication and operations research (Sztrik, 2012). One way of choosing a fleet of trucks involves applying the queuing principle to the haul cycle. Queuing theory has been developed to model systems that serve randomly occurring demands and predict the behaviour of those systems (May, 2013).

A queuing system is one in which customers arrive for the service, wait for the service if it is not available immediately and switch to the next server or exit the system after (Gross *et al.*, 2008). For the prediction of a haul truck's idle time, haulage trucks are the customers in the queuing system as they might wait to be serviced at the loading and dumping locations.

3.1.11 Engine Load Factor

The engine load factor is defined as a portion of the rated engine power that is utilised during the work process (Klanfar *et al.*, 2016). It is also defined as the ratio of average load to the maximum load in an operating cycle (Kecojevic and Komljenovic, 2010). Simonsen (2012) researched the engine load factor and states that the engine load factor also influences the fuel consumption of haul trucks. The findings of Kecojevic and Komljenovic (2011) confirm that fuel consumption is strongly related to the engine load factor.

Trani *et al.* (2016) show that an analysis of on-site fuel consumption can be carried out by characterising fuel equipment and load factors. Runge (1998) suggested equation (2) for calculating fuel consumption (FC):

$$FC = 0.3 \times LF \times P \tag{2}$$

where LF = Engine load factor and P = Power

Equation (2) was modified to give equation (3).

$$FC = \left[\frac{SFC}{FD}\right] \times LF \times P \tag{3}$$

where: SFC = engine specific fuel consumption at full power (0.213 - 0.268 kg/kw.hr)
FD = fuel density (0.85 kg/l for diesel).

The values of LF for mining and off-highway trucks are presented in Table 1.

The engine power (P) of a truck can be determined using equation (4):

$$P = \frac{RF \times S}{3.6} \tag{4}$$

where RF is Rimpull Force

Table 1 Mining and Off-Highway Trucks Engine Load Factors (LF)

Operating Condition	LF (%)	Description
Low	20 – 30	Continuous operation at an average gross weight less than recommended. Excellent haul roads. No overloading, low load factor
Medium	30 – 40	Continuous operation at an average gross weight approaching recommended. Minimal overloading, good haul roads, moderate load factor
High	40 – 50	Continuous operation at or above the maximum recommended gross weight. Overloading, poor haul roads, high load factor.

Source: Anon. (2017)

3.2 Introduction to Machine Learning

Machine learning is mostly used in data analysis for prediction by constructing models and algorithms to produce reliable and valid results and decisions (Angra and Ahuja, 2017). In machine learning, feature subset selection is the most important task (Min *et al.*, 2014). Machine learning is closely linked to computational statistics that focus on computer-based predictions. This has a close association with mathematical optimisation that provides the field with tools, theory and applications (Ongsulee, 2017). In machine learning, the learning algorithms are given data that represents a task so that they can learn from it and predict a solution (Sigweni, 2016). The categories of learning are supervised learning and unsupervised learning.

Unsupervised learning: Unsupervised learning is having input data (x) and no corresponding output variables. The objective of unsupervised learning is to model the hidden structure or distribution of data in order to learn more about the data (Brownlee, 2016). In unsupervised learning, a machine is trained using unclassified or unlabelled information and allowing the algorithm to process that information without guidance. The machine's job is to group unsorted information according to similarities, trends and differences without prior data training (Bansal, 2014).

Supervised learning: This is done using the knowledge of what the output values should be. The goal is to learn a function that, the given data and desired outputs, best approximates the relationship between input and output are observable in data (Soni, 2018). Supervised

learning has input variables (x) and output variables (Y) and an algorithm is used to learn from input to output the mapping function (Brownlee, 2016). The aim is to estimate the mapping function so well that you can predict the output variable (Y) when having new input data (x) using equation (5).

$$Y = f(x) \quad (5)$$

It is called supervised learning because the process of learning an algorithm from the training dataset can be considered as a teacher overseeing the learning process. The correct answer is known and the algorithm repetitively makes predictions about the training data and the teacher corrects it. Learning stops when the algorithm hits an appropriate level of success (Brownlee, 2016).

Supervised learning is done in the context of classification (mapping input to output labels) or the output variable is a category and regression (mapping input to a continuous output) or the output variable is a real value (Brownlee, 2016; Soni, 2018). Supervised learning is frequently used in applications where historical knowledge can be used to predict possible future events (e.g. fuel consumption for the next trip using data from the previous trip). Some of the approaches and algorithms proposed for supervised learning that has been used in the prediction of the consumption of fuel by haulage trucks are discussed in the next sections.

3.2.1 Fuel Consumption Prediction Techniques

This subsection outlines the techniques used by previous researchers to predict fuel consumption by haulage trucks in open-pit mines. This study aims at providing a mine planning tool that can be used to reduce high fuel consumption by haulage trucks in open pit mining. These techniques include Artificial Neural Networks (ANNs), Multiple Integer Linear Regression (MILP), Multiple Linear Regression (MLR), Auto-Regressive Integrated Moving Average (ARIMA) and Case-based Reasoning (CBR).

Artificial Neural Networks (ANNs)

ANNs have been used to predict fuel consumption (Siami-Irdemoosa and Dindarloo, 2015; Soofastaei, 2018). ANNs are intelligence methods which through a fitness function can simulate the effect of multiple variables on one primary factor.

ANNs can simulate the effect of multiple variables on one primary factor by a fitness function. ANNs are trained by the values of the connections between parameters in such a way that an input leads to the specific target output. The pros of ANN are that the hidden and strong non-linear dependencies can be found after learning from examples and the cons are that the prediction error cannot be generally estimated (Obitko, 1999) and ANN can learn the dependency that is valid in a certain period only.

Multiple Integer Linear Regression (MILP)

Integrated analysis is the use of models that concurrently optimise changing decisions from separate parts of a production system (Patterson *et al.*, 2016). The disadvantages of MILP methods are that linear programming does not consider the effect of time and uncertainty. Parameters appearing in the model are assumed to be constant which is not the case in real-life settings.

Multiple Linear Regression (MLR)

Kenton (2019) defined MLR as a statistical technique which uses several explanatory variables to predict variable out of response. MLR attempts to model a linear relationship between independent variables and dependent variables. The advantage of MLR is that it is extremely simple, but its main disadvantage is that it assumes that there is a straight-line relationship between variables which is sometimes incorrect.

Auto-Regressive Integrated Moving Average (ARIMA)

It is a statistical analysis tool that predicts future patterns using time series data. ARIMA model is a type of regression analysis that assesses the intensity of dependent characteristics in relation to other changing characteristics (Chen, 2019). Dindarloo and Siami-Irdemoosa (2016) predicted fuel consumption of haulage trucks based on their cyclic activities (e.g. loading time, hauling time when loaded, dumping time, etc.) using the ARIMA model. The entire modelling procedure must be repeated as there is no automatic updating with ARIMA as new data becomes available (Nanda, 1988).

However, Case-based reasoning (CBR) is preferred as the best technique to be used to predict fuel consumption by haulage trucks in open-pit mines. CBR was selected because:

- It is the simplest method that can be easily understood by mine managers and its domain does not need to be understood completely by the mine managers.

- CBR uses data from the previous haulage trips to predict the fuel consumption of the next trip.
- CBR system learns by acquiring new trips through use. Hence, decision making through the CBR system is dynamic.
- This system can be used even when the dataset is small and the maintenance of a CBR system is easy.

This method according to research, was used to predict fuel consumption for light driving vehicles (Rios-Torres *et al.* 2018). However, it was not used in the context of this research to predict fuel consumption by haulage trucks in open-pit mines

Case-Based Reasoning (CBR)

Case-based reasoning is the use of old knowledge to solve and predict solutions to new problems. Badamich (2015) states that in CBR it is assumed that similar situations have similar solutions. A paper by Rios-Torres *et al.* (2018) shows that driving style was predicted using case-based reasoning which in turn reduced fuel consumption by 12% in an urban area for light driving vehicles. The findings of Monfet *et al.* (2014) show that CBR can be used to predict energy demand in a building operation system demand. The advantage of CBR is that solutions are proposed quickly as a derivation from scratch is avoided. Domains do not need to be completely understood and it highlights the important parameters. The disadvantage of CBR is that retrieval/adaptation knowledge is still needed (Badamich, 2015).

Research shows that CBR was successfully used to find a solution similar to the new problem (i.e. in planning household activities, CBR can be used by remembering what worked and what did not work previously and use that to plan new household activities) as explained by Kolodner (1992) and Gupta (1994). Aamodt and Plaza (1994) state that after getting a prediction each time a new problem has been solved, the new solution is retained and made available for future identical problems. For a prediction to take place the cases must be similar and closely related to the new problem (Poole and Mackworth, 2010). Ayoung *et al.* (2015) state that CBR is different from other artificial intelligence techniques as it is not model-based.

The components of the CBR cycle are as illustrated in Figure 8 (Aamodt and Plaza, 1994; Pal and Shiu, 2004):

- Retrieving: Most similar cases to new problems are retrieved with the aid of a similarity function from the case base.
- Re-use the cases by integrating the solutions from the cases retrieved.
- Revising: Determining if the proposed solution is successful with the objective to solve the initial problem.
- Retaining the new solution once it is validated or verified.
- Once the new case is confirmed and validated it is incorporated into the existing case base (Kocsis *et al.*, 2014).

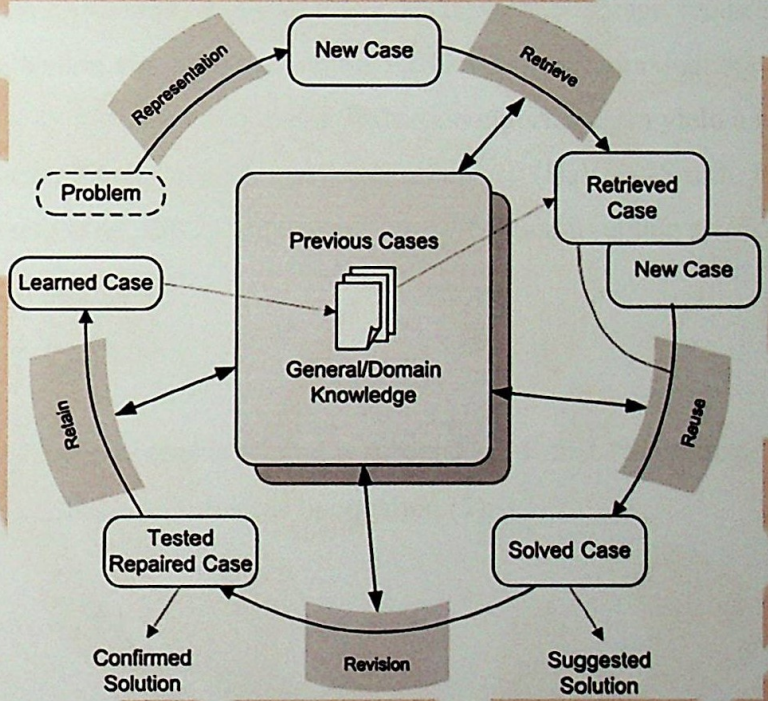


Figure 8 Case-Based Reasoning Flowchart (Kocsis *et al.*, 2014)

Using CBR, a case base data set needs to be created. A data set consists of a dependent variable (variables to be predicted) and independent variables. In this case, fuel consumption parameters are used to predict the dependent variables (Ayoung *et al.*, 2015). Regrettably, findings by some researchers have shown that for the task of prediction, CBR is susceptible to features that are misleading or irrelevant. Methods that eliminate or reduce the effects of irrelevant or misleading features are therefore required. Thus, CBR systems will benefit from the optimisation of feature sets by selecting only the features that benefit prediction

(Sigweni *et al.*, 2017). The method includes feature subset selection which is explained in detail in the next subsection.

Feature subset selection: Feature selection is a method widely used in machine learning (Min and Fangfang, 2010). Feature selection is the automatic selection of data attributes provided that are most relevant to the predictive modelling problem being worked on (Brownlee, 2014). In feature selection or sub-sets selection process a sub-set of features is chosen from the available data to apply a learning algorithm (Kumari and Swarnkar, 2011). It involves the determination of the optimum subset of features that would provide the most accurate prediction (Ayoung *et al.*, 2015). Hall and Smith, (1997) state that feature subset selection method enhances the performance of learning algorithms by reducing the hypothesis search space and, in some cases, reducing the storage requirements. A good feature subset selection algorithm can remove irrelevant and redundant features effectively and consider feature interaction so that an induction algorithm can yield a better prediction accuracy with more concise models and better efficiency (Hall and Smith, 1997; Zheng and Kwoh, 2011; Wang *et al.*, 2013). Equation (6) may be used to obtain a set of features:

$$F = \{f_1, \ldots\ldots f_i, \ldots\ldots\ldots f_n\} \quad (6)$$

The feature selection problem is to find a subset $F' \subseteq F$ that "maximizes" the ability to classify patterns or finding a subset as in equation (7):

$$F' = \{f_1, \ldots\ldots f_i, \ldots\ldots\ldots f_n\} \quad (7)$$

Where $\boldsymbol{n}$ optimises the objective function.

The objective is to determine the optimal subset of feature subset that maximises the scoring function S(i). Optimal feature subset is a subset of $\boldsymbol{F}$ such that the accuracy of the hypothesis is maximal (Guyon and Elisseeff, 2003).

Methods of feature subset selection

Feature subset selection can be categorised into three groups as Filter, Wrapper and Embedded Methods.

Filter methods: Filter methods assess the significance of the feature by examining the intrinsic properties of the data independent of an algorithm (Sutha and Tamilselvi, 2015). All features are scored and ranked according to certain statistical criteria. The lowest ranking features are discarded whereas the highest-ranking features are selected (Kumari and Swarnkar, 2011). The filter method may rank individual parameters or evaluate all feature subsets and feature filtering measures based on information, distance, consistency, similarity and statistical measurement (Jovic *et al.*, 2015). Filter methods fully disregard the effects of the weight of the selected features on the output of the estimation algorithm and the method is fast and independent of the classifiers (Sigweni, 2016). Filter methods do not use a learning algorithm to evaluate a subset of features, since they use statistical measures to rank features available (El Aboudi and Benhlima, 2016). Figure 9 is a schematic diagram of filter methods.

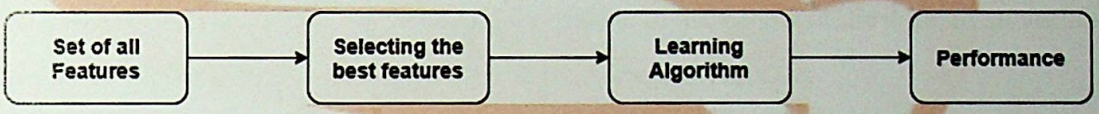


Figure 9 Filter Methods (Dhote *et al.*, 2015)

Variable ranking: The filter technique uses variable ranking which is a process of ordering features by values and by a scoring function that usually measures feature relevance (Kaushik, 2016). The selection of a subset of input variables is often based on a ranking being built to order the variables according to a given specific criterion. The goal is then to linearize the search and estimate the quality of the subsets that contain the top-ranked variables (Quevedo *et al.*, 2007).

The aim is to identify the important parameters influencing fuel consumption by haulage trucks. Suppose Y is the fuel consumed by haulage trucks, independent variables X_1, ..., X_p contains a set of random variables which potentially influence Y and $S_i = (Y_i X_i, \ldots\ldots\ldots\ldots, X_{ip})$, $i = 1, \ldots\ldots\ldots$, n, independent copies of $S = (Y, X_1, \ldots\ldots\ldots, X_p)$ where p could be very large. It is often assumed that there are many variables having no impact on the response of Y (Baranowski *et al.*, 2018).

The variable ranking makes use of the scoring function S_i, taking into account a set of m variables $\{f_k, y_k\}(k = 1, \ldots, m)$ consisting of n input variables $f_{k,i}(i = 1, \ldots, m)$ and one

output variable, y_k. The scoring function is computed from values $f_{k,i}$ *and* $y_k (k = 1, \ldots, m)$ (Guyon and Elisseeff, 2003).

Pearson's correlation: Pearson's correlation is a measure of linear dependence that can be used to calculate dependencies between quantitative variables f and y for n cases in a data set utilising the means of f and y (de Siqueira Santos *et al.*, 2014). A simple method for feature selection using variable ranking is to select the k highest-ranked features according to S. Equation (8) is used in ranking variables in the filter method:

$$r(f_i, y) = \frac{\sum_{k=1}^{m}(f_{k,i} - \bar{f}_i)(y_k - \bar{y})}{\sqrt{\sum_{k=1}^{m}(f_{k,i} - \bar{f}_i)^2 \sum_{i=1}^{n}(y_k - \bar{y})^2}} \quad (8)$$

where: r^2-value measures the percentage of variation in the values of the dependent variable that can be explained by the variation in the independent variable.

The r^2- values varies from 0 to 1, where the correlation is 1 for a perfect linear dependency and 0 if the random variables are not dependent. Where the linear dependency is imperfect, $0 < r^2 < 1$ (de Siqueira Santos *et al.*, 2014).

Wrapper method: The wrapper method embed the model hypothesis search within the feature subset search. In the Wrapper approach, there exists a feature subset selection algorithm as a wrapper around the induction algorithm. Using the induction algorithm, the feature subset selection algorithm conducts a search for a good subset as part of the evaluation function of the subsets of features (Kohavi and John, 1997). In this method, a search procedure is described in the space of possible subsets of features and various subsets of features are generated and evaluated as shown in Figure 10.

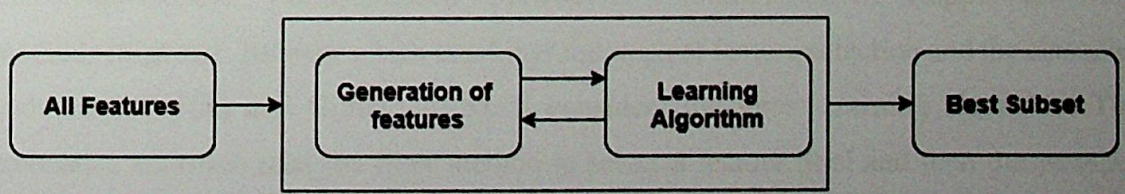


Figure 10 Wrapper Method (Kohavi and John, 1997; El Aboudi and Benhlima, 2016)

Evaluation of any given subset of features is obtained through training and testing of a specific classification model, execution of this approach is adapted to a specific classification algorithm (Kumari and Swarnkar, 2011).

Wrapper methods employ a learning algorithm that is adopted to evaluate the accuracy of the performance of a candidate feature subset which leads to better results (El Aboudi and Benhlima, 2016). A search algorithm in searching the space of all feature subsets is 'wrapped' around the classification model. A heuristic search model is used as the space of feature subsets grows exponentially (Kumari and Swarnkar, 2011).

In the Wrapper approach, the dataset is usually separated into training data and test data. A good subset is then searched using the feature subset selection algorithm which uses the induction algorithm itself as part of the feature subset evaluation function. The feature subsets with the best evaluation are the final subsets which are run on an induction algorithm. The resultant classifier is then evaluated on the test data (Sigweni, 2016).

Sigweni (2016) shows that a typical Wrapper model will perform the following steps, given a predefined induction algorithm:

1. Searching a subset of features.
2. Evaluating the selected subset of features by the performance of the induction algorithm.
3. Repeating steps 1 and 2 until the desired evaluation accuracy is achieved.

The main idea behind the Wrapper approach is shown in Figure 11. In the Wrapper method, several search strategies can be used to search for subsets of features including heuristic search, hill-climbing, best-first and greedy search algorithms.

Embedded method: The embedded approach is a combination of wrapper and filter methods (Sigweni, 2016) or which combines the steps of feature selection and the classifier construction (Hira and Gillies, 2015). It considers the given learning algorithm. The embedded approach uses the filter method to select a feature pool and then the wrapper method is used to find the optimum subset of features from the selected feature pool (Imani *et al.*, 2013). The embedded method uses a search procedure that is inferred in the classifier (Ramirez-Gallego *et al.*, 2018).

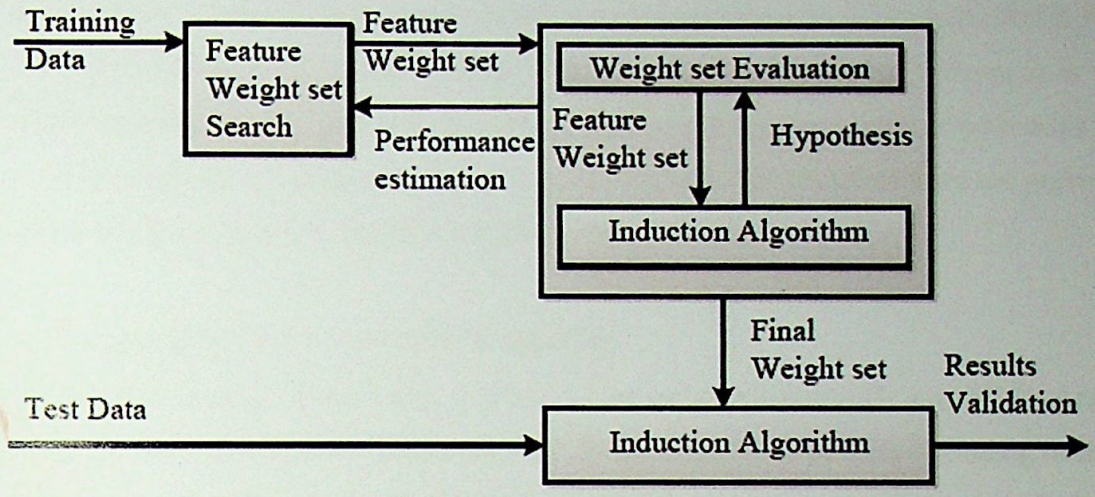


Figure 11 Wrapper Approach to Feature Weight Set Selection (Sigweni, 2016)

The advantages and disadvantages of feature selection methods (i.e. Filter, Wrapper and Embedded methods) are summarised in Table 2.

Table 2 Advantages and Disadvantages of Feature Selection Methods

Feature Selection Methods	Advantages	Disadvantages	Source
Filter	• Fast and scalable • Simple computation • Independent of the classification algorithm • Better computational difficulty than the wrapper approach	• Ignores the effect of the selected feature on classifier • Ignores feature dependency • Unclear cut-off rank to select true features	Kumari & Swarnkar (2011); Hira & Gillies, (2015)
Wrapper	• Higher classification accuracy • Interaction with the classifier • Considers the feature dependencies. • Good classification accuracy. • Minimises computational cost.	• Higher risk of overfitting • Break down with very large numbers of features in some cases • Longer runtime • Classifier dependent selection	Kumari & Swarnkar (2011) Jovic *et al.* (2015)
Embedded	• Interaction with the classifier. • Consider the feature dependencies. • Better runtime than wrapper methods.	• Classifier dependent selection • Computationally intensive.	Hira & Gillies (2015)

Case-based reasoning (CBR) is preferred as the best technique to be used to predict fuel consumption by haulage trucks in open-pit mines. CBR was selected because it is the simplest method that can be easily understood by mine managers and its domain does not need to be understood completely by the mine managers. CBR uses data from the previous haulage trips to predict the fuel consumption of the next trip.

3.3 Analogy Based Software Effort Estimation

Estimation by analogy is basically a form of case-based reasoning (Bajta, 2015). Azzeh and Nassif (2015) defined estimation by analogy as a simplified process of finding nearest analogies based on the notion of retrieval by similarity.

3.3.1 ArchANGEL Analogy Based System

The basic approach to case-based prediction is incorporated in a CBR shell entitled ArchANGEL that allows cases (usually projects) for effort prediction (Kirsopp and Shepperd, 2003). ArchANGEL applies the same principle as the CBR reasoning approach Therefore, it has the same objective which is automating the prediction process (Sigweni, 2016). ArchANGEL supports several validation techniques, search heuristic algorithms to perform feature and case selection on datasets and uses a user-friendly graphical user interface (GUI) as its primary user. ArchANGEL allows the estimator specified features to be used when it searches for source analogies. The built-in search algorithms consider all possible combinations of the feature subset and select the one that minimises the prediction metric to determine the best subset (Keung, 2009).

Estimation using analogical reasoning is matching one or more projects from a historical base case to a new project for which an estimate is required which is done by ArchANGEL (Schofield, 1998). Projects/trips are characterised by several descriptor parameters used to measure the similarity between trips. Once similar completed trips have been found, the known effort values of these trips can be used to form an estimate for the new trip.

ArchANGEL is a system based on the *k*-Nearest Neighbour Approach (*k*-NN) method for estimating project effort analogy. It divides the estimation process into 3 components as shown in Figure 12 (Sigweni, 2016):

i. Template is like a data dictionary so the system can have a shell architecture and is not limited to a specific dataset or features.

ii. Information captured in a template build trip case base. It is the storage for trip data.

iii. Using source trips as the trip's base case, allows the user to generate estimates for target trips.

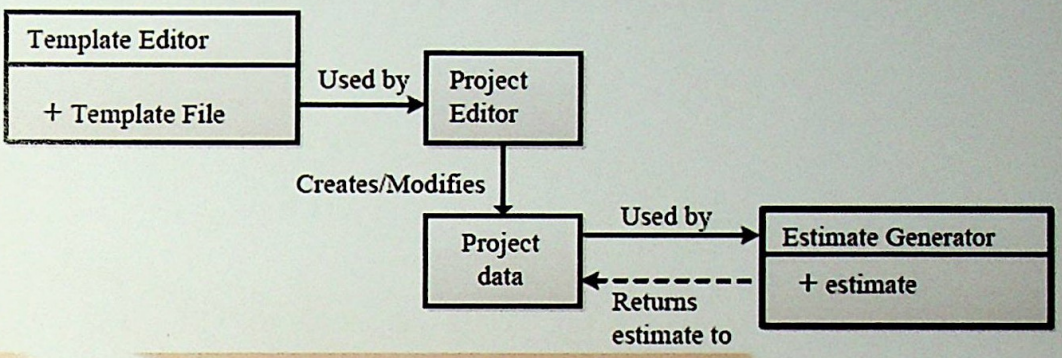


Figure 12 Schematic of ANGEL (Sigweni, 2016)

Figure 12 shows the communication between key components of the ANGEL tool. A template editor is used to model the environment in which data will be collected, generating a template before the projects are stored. The template is then saved as a text file. The project case editor then uses the template file to construct a project case base.

Before projects are stored, a template editor is used to outline the environment where data will be collected, creating a template. Afterward, the template is saved as a text file. The template file is then used by the project case editor to create a project case base. The estimate generator then uses the project data to generate a prediction for any new trips/projects (Sigweni, 2016).

Although many approaches have been used to predict fuel consumption by haulage trucks, feature subset selection and case-based reasoning will be employed together with ArchANGEL in this research.

Similarity measure: Measuring similarity is a vital part of analogy-based software effort estimation. For calculating the degree of similarity between software projects a similarity function is used. Similarity plays an important part in identifying the project that is closest to the estimated project (Sigweni, 2016).

The similarity between the target trips and all possible sources comparable to the case base is measured by the Euclidean distance metric, which is the measure of the true straight-line distance between two points in Euclidean space (Anon., 2019). According to Schofield (1998) the method used to measure similarity is the nearest neighbour algorithm which is used in the estimation by analogy involves measuring Euclidean distance in n-dimensional space as given in equation (9):

$$ED(x, y) = \sqrt{\sum_{j=1}^{i=n}(x_i - y_j)} \quad (9)$$

where: n is the number of parameters, x and y are two trip cases

Euclidean distance equation (10) measures the distances for two projects or trips:

$$D(t, t') = \sqrt{\sum_{i=1}^{n} W_1 Dis(f_1 - f_i)}$$

$$Dis(f_1 - f_i') = \begin{cases} (f_1 - f_i') \text{ if } f_1 \text{ and } f_i' \text{ are numerical or ordinal} \\ 1 \quad \text{if } f_1 \text{ and } f_i' \text{ are nominal and } f_i = f_i' \\ 0 \quad \text{if } f_1 \text{ and } f_i' \text{ are nominal and } f_i \neq f_i' \end{cases} \quad (10)$$

where: t, t′ denote the projects or trips

f_i and f_i'denote the i^{th} feature value of the corresponding trips = 1 is the i^{th} feature is selected = 0 ➡ the i^{th} feature is not selected

n is the total number of features

Dis = feature dissimilarity

k-Nearest Neighbour approach: The *k*-Nearest Neighbour approach (*k*-NN based approach) is an economical, automated tool to generate a ranking of projects by formulating a relation based or comparing it to *k*- similar cases on its training dataset (Klair and Kaur, 2012; Kafaf *et al.*, 2017). Sigweni (2016) defined ***k*** as the number of nearest neighbours used as donor projects/trips to predict the effort of the target project/trip. Therefore, the *k*-NN system influences the accuracy of the estimation. Li *et al.* (2009) also defined ***k*** as the number of most similar projects that will be used to generate the estimate.

For *k*-NN to classify a given new case, the method identifies the ***k*** most similar training case (neighbours) to the new case, then assigns the new case to the class containing most of its neighbours (Kassambara, 2018). The neighbours act as donors to the new case. Figure 13 shows the example Sigweni (2016) used to illustrate how the *k-NN* based approach works.

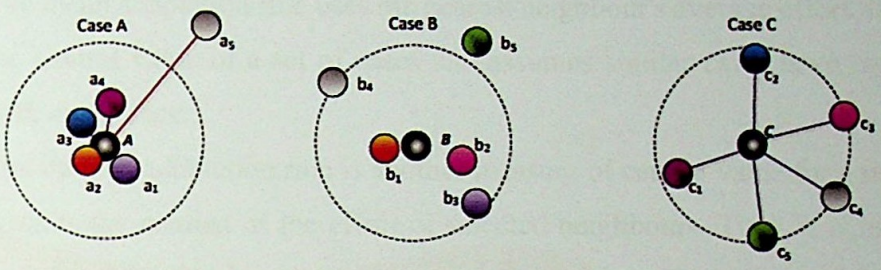


Figure 13 *k*-Nearest Neighbour (Sigweni, 2016)

In Case A: The nearest neighbours to A is a_1 to a_3, k = 3 may yield the best results, a_5 is not the nearest neighbour to case A and k = 5 may not yield the best results. This is due to a_5 being required to make a number of analogies equal 5 to estimate case A. Donor a_5 is clearly not similar to a_1 ... a_4. Case B is a situation in which either k = 2 or k = 4 will yield similar results. Therefore, it shows that there are situations in which different ***k*** values would suit the same case. Case C: k = 5 can be perfect since all donors are almost equidistant from C. This k-NN approach is referred to as the fixed k-NN approach.

Dynamic k-NN approach: In the dynamic k-NN approach various ***ks*** are tested and then the best ***k*** is selected (Wu *et al.*, 2010). Keung (2008) shows a procedure where an optimal k_i value for each project case x_i is simulated and used in the prediction dynamically. Given ***n*** project cases (x_1, x_2, x_3, ..., x_n) the training set and for each case x_i in turn:

i. Euclidean distance between x_i and all other project cases x_i is computed and ranked by their similarity.
ii. Prediction is then simulated with k = 1, k = 2, ..., k = n for case x_i. The mean residual error is used in this case as an evaluation criterion.
iii. This completes the simulation for the case x_1.
iv. Steps (i) to (iii) are repeated for the next project case x_{i+1}.

After selecting the ***k*** nearest neighbours, a case adaptation strategy is then applied to the selected donor cases to obtain a prediction of effort for the new case being predicted. Pal and Shiu (2004) state that the adaptation rules are mechanisms used to derive a new estimate to lessen the differences between donor cases and the new case being predicted. Various techniques of adaptation have been proposed, such as:

i. The mean adaptation rule uses the nearest neighbour's average effort. It expresses the central value of a set of cases and assumes similar cases have equal effects with each other.
ii. The median adaptation rule is another measure of central value for a set of cases. It takes the median of the effort of selected neighbours. Thus, it is more robust than the mean as ***k*** increases (Angelis and Stamelos, 2004; Sigweni, 2016).
iii. The inverse distance weighted mean allows neighbouring cases to have different effects on the prediction. Therefore, cases more similar to the new case being predicted have more effect on the prediction.

3.4 Methods of Accuracy Estimation and Validation Techniques

This section outlines the methods of accuracy estimation and the evaluation methods used in this research work.

3.4.1 Methods of Accuracy Estimation

In creating an effort estimation model, it is important to use a window of training data. The work by Amasaki and Lokan (2014) shows that windowing is one of the effective methods of improving accuracy estimation. The strategies in the windowing approach include:

1. Growing approach

In the growing window, trips are added to the data set as time passes. The real-life setting is reflected here, as completed projects predict the new one. The data set looks like a growing window and sequentially increases the size of the training set and all past projects are retained, so the data set grows as more projects finish (Amasaki and Lokan, 2014). In Figure 14, T_1 represents the earliest trip (i.e. completed first). In terms of Analogy Based Estimation, e.g. starting with 3 trips. Window 1 would be used to predict the fuel used for Trip 4 (T_4) and when T_4 is completed and its amount of fuel used is established, it is then added to the case base (Window 2) and available to be used to predict the fuel used for T_5

(Sigweni *et al.*, 2017). This shows that the window is growing. Hence T_{n-1} trips could be used to predict trip T_n.

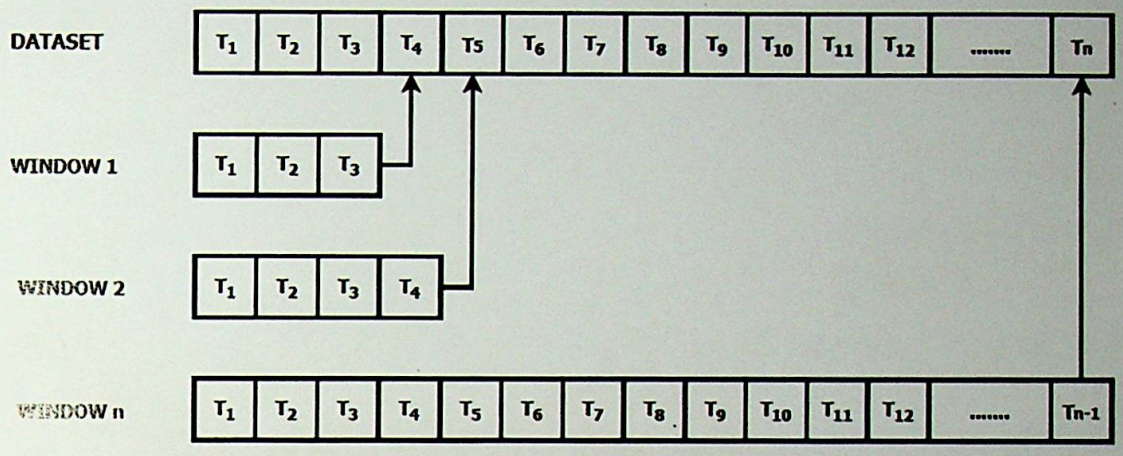


Figure 14 Validation Data Set using Growing Window (adopted from Sigweni *et al.*, 2016)

2. *Moving approach*

New trips are entered into the data set in the moving window as time passes and old trips cease to be considered as the window moves past them (Lokan and Mendes, 2009). This happens if one assumes that recent trips reflect what is happening currently in task and practice, while past trips become less relevant over time. In a moving window, the size of the window remains fixed. Figure 15 shows the evaluation data set using the moving window approach.

In a moving window for each new trip T_n to be estimated, a model should be created based on a window containing trip T_{n-t} to T_{n-1}, where ***t*** is the window size. This approach creates windows based on the number of previous cases to be included. It can also be referred to as the fixed size approach. In Figure 15, T_1 represents the earliest project. In terms of analogy based estimation, Window 1 would be used to predict the effort for case T_4 and when T_4 is completed and its effort now established it is then added to the case base (Window 2) and it is available to be used to predict the effort for case T_5. T_1 is now considered less relevant and it is excluded as shown in Window 2 thus maintaining 3 trips per window (Sigweni, 2016).

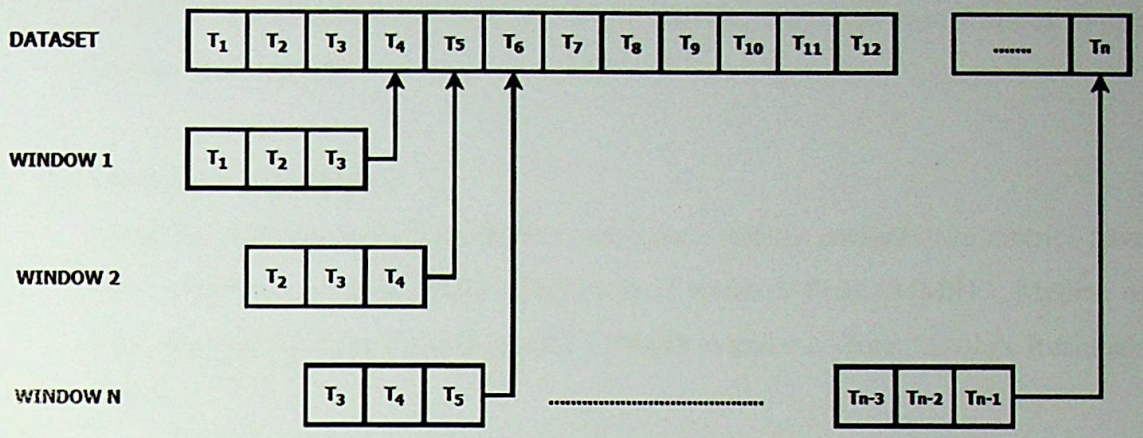


Figure 15 Evaluation Data Set using Moving Window (adopted from Sigweni, 2016)

3.4.2 Evaluation Method

The quality of prediction results plays a significant part in analysis and decision making and accuracy is a significant part of results quality. Therefore, it is necessary to evaluate the accuracy of the data (Zhang *et al.*, 2014). The accuracy from various estimation techniques is used as a starting point for comparison in analysing the value of CBR in estimating the fuel consumption by haulage trucks. Shepperd-MacDonell method (SMM) is employed to reduce the irregularities between validated research results and also providing a basis for understanding the results with a specific emphasis on continuous prediction systems (Shepperd and Macdonell, 2012).

Evaluating an estimation system is based upon Standardised Accuracy (SA), an unbiased statistic, calculation of effect sizes and testing the likelihood of the results relative to the baseline technique of random predictions (guessing). According to the Shepperd and MacDonell method, an estimation system, E_i, is assumed which is to be validated over a dataset D using some accuracy statistic S as per the validation scheme V (Shepperd and MacDonell, 2012). The SMM method can be utilised in comparing competing estimation systems. For example, given prospective estimation systems E_1 and E_2 and an accuracy statistic, S, the following steps must be taken:

1. Performance against random guessing

   A baseline of random guessing is established to find out if the suggested method performs better than random guessing. It is expected that any theoretically good system should execute better than random guessing over time. If that is not the case,

then it is assumed that the indicator is not utilising the target case features in any useful way or at all.

2. Testing for Significance

   To assess the accuracy of prediction techniques various performance metrics have been used those include: Mean Magnitude of Relative Error (MMRE), Median of Magnitude of Relative Error (MdMRE), PRED(p) and the Mean Absolute Residuals (MAR).

   MMRE is a commonly used measure for assessing the performance of prediction systems regardless of their flaws. The basic metric in MMRE is Magnitude of Relative Error (MRE) which is defined in equation (11):

$$\text{MRE} = \frac{|x_i - y_i|}{x_i} \qquad (11)$$

   MMRE is given by equation (12):

$$\text{MMRE} = \frac{|x_i - y_i|}{x_i n} \qquad (12)$$

Where: x_i is the i^{th} value predicted, y_i is the estimate, $|x_i - y_i|$ represents i^{th} the residual and n is the number of trips in the dataset. Low MMRE values suggest a low estimation error. Sadly, it has been shown that MMRE is a biased estimator of the central tendency of the residuals of a prediction system because it is an asymmetric measure (Foss *et al,* 2003; Kitchenham *et al,* 2001; Myrtveit *et al,* 2005). MMRE will be biased towards under-estimating prediction systems, resulting in over-optimism (Jorgensen and Molokken-Ostvold, 2006) since MMRE penalises overestimation more than underestimation.

MdMRE is calculated using the median of all the MREs i.e. MdMRE ids the median (MRE). Although MdMRE has similar patterns to MMRE, it is more likely to select the true model, particularly in cases of underestimation, as it is less sensitive to extreme outliers (Foss *et al,* 2003).

PRED(p) is the percentage of all predicted cost values that are within p% of the actual cost, p = 25% is the widely used value. PRED(p) is calculated as in equation (13):

$$\text{PRED(p)} = \frac{l}{n} \tag{13}$$

Where n is the number of all trips and l is the number of trips whose MRE $\leq p$-value. This evaluation method is not appropriate for comparing competing predicting models, as it does not differentiate between system models that fall outside the p% range (Li *et al*, 2009).

Due to the above-mentioned issues for MMRE, MdMRE and PRED(p), Shepperd and MacDonell, (2012) proposed a new framework for evaluating prediction systems based upon the unbiased statistic (e.g. Standardised accuracy, testing the result likelihood relative to the baseline technique of random guessing and calculation of effect sizes. The evaluation of these statistics is based upon Mean Absolute Residual (MAR) obtained using equation (14):

$$\text{MAR} = \frac{\sum_i^n (|x_i - y_i|)}{n} \tag{14}$$

where: x_i is the i^{th} value predicted, y_i is the estimate, $|x_i - y_i|$ represents i^{th} the residual and n is the number of trips in the dataset

MAR measures the average magnitude of the errors in a set of predictions without taking into consideration their direction. It measures accuracy for continuous variables (Pandey and Subbiah, 2018). It is used as the accuracy statistic S for continuous estimation systems. In contrast to Mean Magnitude of Relative Errors, MAR does not depend on proportions. For that reason, it is unbiased. Unfortunately, comparisons in MAR across the datasets cannot be made because residuals are not standardised. As a result of this, it is hard to interpret. In that case, Shepperd and McDonell (2012) presented a standardised accuracy measure (SA), where accuracy is measured as the MAR relative to random guessing E_0. Hence the SA for estimating E_i is given by equation (15):

$$\text{SA}_{E_1} = 1 - \frac{\text{MAR}_{E_i}}{\text{MAR}_{E_0}} \times 100 \tag{15}$$

where: MAR_{E_0}= mean of a large number (normally 1000) runs of random guessing.

SA = ratio representing how much better the estimation system E_i is than random guessing E_o. If the SA value is near zero it would be discouraging and a negative value would be more regrettable.

3. Effect Size

Effect size is a simpler way of revealing the difference between the two groups (Kim, 2015). Therefore, it can be said to be a true measure of the actual difference. MacLeod (2019) states that the larger the effect size the stronger the relationship between the two variables. There are three types of effect size which are:

i). Glass's Delta: It is a measure of effect size. Glass's delta uses the control group's standard deviation. Glass's delta is used when the standard deviation is significantly different between groups (Glen, 2016). Assessing the effect size. a standardised measure by Glass Δ as in (Shepperd and MacDonell, 2012) is used shown in equation (16):

$$\Delta = \frac{MAR_{E_i} - MAR_{E_0}}{\sigma_{E_0}} \qquad (16)$$

E_0 represents the sample standard deviation of the random guessing methodology.

Glass Δ will give two things: It standardises the contrast between the two estimation frameworks and afterward it contextualises the distinction as far as the measure of variety in the two measures of **S**.

ii). Cohen's ***d***: It is a measure of effect size. It is used when the standard deviations are not significantly different between groups and when the sample is greater than 20 (Glen, 2016).

iii). Hedges' ***g***: It is also a measure of effect size used when the standard deviations are not significantly different between groups and the sample size is small (< 20) (Ellis, 2010 and Glen, 2016).

"Significance" is commonly used in statistics to denote "statistical significance". The statistical significance is calculated as a '***p-value***,' the likelihood that a difference of at least the same size would have occurred by chance, irrespective of whether the two populations differed. The ***p-value*** will usually be determined from a 'Z-test' for differences between the means of two groups. By convention, if $p < 0.05$ (i.e. below 5%), the difference is taken to be large enough to be 'significant'. If not, then it is 'not significant' (Coe, 2002).

To determine which method to use the following steps are followed:

1. A significance level of (5%) is used.
2. The standard deviation of the control and experimental groups is determined. Then the standard error (***sd***) is determined using equation (17):

$$sd = \sqrt{\frac{S_1}{N_1} + \frac{S_2}{N_2}} \tag{17}$$

where: S_1 and S_2 are the standard deviations of the control and experimental groups respectively whereas N_1 and N_2 are sample sizes of the control and experimental groups respectively.

3. Z-value is then calculated using equation (18).

$$Z = \frac{\mu_1 - \mu_2}{sd} \tag{18}$$

Where μ_1 and μ_2 are the mean for control and experimental groups respectively and ***sd*** is the standard error or variance between the two groups.

4. A Z-score – This is a distribution table used to determine the significance between the two groups.

The interpretation of the standardised effect size is in terms of classes suggested by Coe (2002) as small (≈ 0.2), medium (≈ 0.5) and large (≈ 0.8). It is scale-free as shown in Table 3. The Δ has a unit of a standard deviation, so the effect is a decrease in the MAR of ***n*** fuel consumed per trip.

Table 3 Classes of Effect Size

Relative Size	Effect Size	Percent of Control Group (< Mean of Experimental Group)
Small	0.2	58
Medium	0.5	69
Large	0.8	79

Source: Adapted from Coe (2002) and McLeod (2019)

Repeated measure design: The repeated measure design was opted for in this work because it takes measurements on the same subject over time and under different conditions. It is characterised by having more than one measurement of at least a given variable for each subject. The repeated measure design also reduces the variances of estimates of trips by a haulage truck which allows fewer statistical interferences to be made with fewer trips. It also allows more experiments to be completed more quickly as fewer trips need to be trained to complete an entire experiment.

The CBR system is implemented using ArchANGEL to store and make calculations using functions embedded in the ArchANGEL system. Three estimation techniques were opted for in the prediction of fuel consumption by haulage trucks in open-pit mines using CBR methods. These techniques were utilised by Sigweni and Shepperd, (2015) in their blind analysis research and they constitute a range of potential methodologies that can be used for Case-based fuel consumption estimation. These include:

1. Naïve technique where a sample mean is used to estimate new fuel consumed per trip (Shepperd and MacDonell, 2012; Sigweni and Shepperd, 2015). This technique is included to determine the degree to which the other techniques offer any value and researchers have demonstrated the dangers of not using the proper benchmark.
2. Case-Based Reasoning (CBR) technique: This technique uses all parameters equally weighted to predict fuel consumption by haulage trucks. This method might be considered as a baseline technique and it has been in application since the 1990s (Shepperd and Schofield, 1997).
3. Case-Based Reasoning technique employing Forward Sequential Selection (CBR-FSS): This technique eliminates all irrelevant parameters in the prediction of fuel consumption as compared to CBR which uses all parameters equally weighted.

Studies show that CBR-FSS is a successful improvement over the CBR technique (Sigweni and Shepperd, 2014; Ayoung *et al.*, 2015).

3.5 Summary

This chapter addressed research questions 1, 3 and 4. Firstly, the factors that influence fuel consumption by haulage trucks were outlined (e.g. cycle time, payload, weather conditions). Secondly, prediction techniques that have been used by other researchers to predict fuel consumption by haulage trucks in open-pit mines were studied. Also, the CBR technique was selected as the best technique for this research work as it is the simplest method that can be understood by mine managers. Finally, the evaluation methods were outlined. However, MAR, standardised accuracy and effect sizes were selected as the best evaluation methods for determining the accuracy of the selected methods for predicting fuel consumption.

CHAPTER 4

DATA COLLECTION AND ANALYSIS

4 INTRODUCTION

This chapter outlines how the data was acquired from Orapa Diamond Mine and \how the collected information is used to predict fuel consumption by haulage trucks using the prediction methods. The qualities of each prediction method in predicting fuel consumption are outlined.

4.1 Data Collection

Data was collected from Orapa Mine through reports generated by the service providers (Komatsu and Barloworld Equipment) on fuel consumption as well as 11 other parameters (e.g. payload, cycle time, etc.) on a trip by trip basis. Barloworld is responsible for the maintenance and operation of CAT 789 and CAT 793 trucks which have 190 ton and 240-ton capacities respectively.

The data gathered from Komatsu has fuel consumption as well as 37 other parameters (e.g. payload, speed, cycle time, etc.) on a trip by trip basis. Komatsu is responsible for Komatsu (730 and 830) trucks with 190-ton and 240-ton capacities respectively.

4.1.1 Data Description

Orapa Mine's haulage operation delivers ore and waste materials from the pit to the dumps (crushers, stockpiles and waste dumps). Each haulage truck driver is assigned trips to set locations using dynamic dispatching in the Wencomine system. The trucks are equipped with an on-board screen where each driver inputs his/her details at the beginning of the shift. The on-board screens are where all the information is relayed to the driver from the vehicle's monitoring system.

Orapa Mine currently uses Wencomine as the vehicle monitoring system which runs the mine's operational parameters through its dispatching algorithm to produce a schedule for the haulage trucks and shovels. The system assigns trips according to the best possible travel paths using travel times. The system calculates the travel time considering the ramps,

slopes and switchbacks. The best possible travel path can be the most fuel-consuming route as the system does not consider the amount of fuel consumed when selecting the best possible path.

The dispatch system consists of infrared beacons which are communication points mounted on each segment of the haul routes and at other strategic locations in the pit. The programming algorithm assigns trucks to different locations throughout the shift using these beacons.

The Wencomine system stores data in the form of reports. The reports were extracted together with those from Barloworld and Komatsu for use in this work. The reports contain several variables influencing fuel consumption by haulage trucks and the amount of fuel consumed. The variables show the performance of each haulage truck per trip.

4.1.2 Data Preparation

The data collected from Orapa Mine was divided into datasets according to scenarios in subsequent sections for analysing each scenario.

Scenarios for analysis of fuel consumption by haulage trucks

This section discusses three different scenarios used to analyse the prediction of fuel consumption by haulage trucks in open-pit mines.

Scenario 1: Using CBR to predict fuel consumption by haulage trucks

Firstly, it needs to be determined if CBR can be used to predict the consumption of fuel by haulage trucks. According to research, CBR has not been used to predict fuel consumption by haulage trucks in open-pit mines in the context of this research. To determine whether CBR can be used to predict fuel consumption by haulage trucks, a CAT Truck #29 dataset comprising of 161 cases and 15 features in Orapa Mine was used in the analysis in Scenario 1.

Scenario 2: Using CBR to predict fuel consumption during shifts

This scenario is to determine if CBR can be used to predict fuel consumption by haulage trucks during the three shifts worked at Orapa Mine and how CBR performs during each shift. Orapa Mine runs three shifts - Morning, Afternoon and Night shifts. Each shift lasts 8

hours. The morning shift is from 06:00 hrs - 14:00 hrs, the Afternoon shift is from 14:00 hrs - 22:00 hrs while the Nightshift is from 22:00 hrs - 06:00 hrs. Night shift work is usually associated with fatigue from the lack of sleep of most workers during daytime which eventually causes workers to doze off during the night shifts. As a result, production values from night shifts are usually less than those in the other shifts. This research seeks to determine how fuel consumption is affected during the three shifts and the factors that might be contributing to the differences in fuel consumption. The details of the dataset used from Debswana Orapa Mine are contained in Appendix 1. These were divided according to shifts for analysis as Scenario 2 (see Table 4).

Table 4 Selected Datasets for Scenario 2

Dataset	No. of Cases	No. of Features
CAT Truck #3 - Morning shift	230	15
CAT Truck #3 - Afternoon shift	215	15
CAT Truck #3 - Night shift	203	15

Scenario 3: CBR for predicting fuel consumption among different haulage trucks available at Orapa Mine

Scenario 3 was used to check whether CBR can be used to predict fuel consumption by the different capacities of haulage trucks used at Orapa Mine.

Data acquired from Orapa Mine shows that there are two types of trucks with carrying capacities of 190 tons and 240 tons trucks used by both Caterpillar and Komatsu trucks in the materials handling operations. This research sought to compare the fuel consumption of the two types of trucks and identify the root causes of any differences in fuel consumption.

The datasets presented in Table 5 were used to determine whether CBR can be used to predict fuel consumption by haulage trucks of different capacities at Orapa Mine.

Table 5 Selected Datasets for Scenario 3

Dataset	No of Cases	No of Features
Komatsu 830E Truck #1	514	39
CAT 793 Truck #2	637	15

Windowing: Firstly, after data preparation, the data was divided into windows as explained in the methodology section using python code as shown in Figure 16. The growing window method was selected for use in this research as all the previous trips are used to predict fuel consumption during the next trips.

```
import csv
from dateutil.relativedelta import relativedelta
from datetime import datetime
import math

with open ('CATtrucksPayloadTruck 3.csv', 'r') as Trucks:
    head = ""
    cases = {}
    for line in Trucks.readlines():
        case = line.split(",")
        enddate = None
        try:
            enddate = datetime.strptime(case[1],"%m/%d/%Y %H:%M:%S")
        except ValueError:
            head = line
            continue
        cases[case[0]] = enddate, line
    windows = { }
    for tnum, tcase in cases.items():
        windows[tnum] = []
        for cnum, case in cases.items():
            if case[0] < tcase[0]:
                windows[tnum].append(case[1])
        windows[tnum].append(tcase[1])
    for cnum, case in windows.items():
        if len(case) > 3:
            wfile = open ('TRUCK3/window' + cnum + '.csv', 'w')
            wfile.write(head)
            for c in case:
                wfile.write(c)
```

Figure 16 Python Code for Creating Windows

4.1.3 Prediction Using ArchANGEL System

ArchANGEL was then set up for prediction using the feature subset strategies summarised in Table 6.

Table 6 ArchANGEL Setup

Field	Option
Target Feature	Fuel Used
Number of Donors (k)	2
Performance Indicator	Sum/r
Selection Strategy	FSS or CBR

The fuel consumed by haul rucks during all target trips in the evaluation datasets was predicted using the FSS and CBR techniques as the feature subset strategy and the sum of absolute residuals for these predictions were used as the accuracy indicator for each prediction technique. The prediction for random guessing was done using the Python code in Figure 17 where the input parameters are the actual fuel consumed and the output parameters are the standard deviation (sd), mean and the mean absolute residuals/error (mae or MAR) values.

```
for (l in 1:1000) {
    actual <- c()
    predicted <- sample(actual, replace=F)
    print(as.interger(mae(error)))
    sd(predicted)

    mean <- function(error)
{
    mean(abs(error))
}
```

Figure 17 Python Code for Random Guessing Prediction

4.2 Results and Analysis

In this section, the mean absolute error was used to determine the accuracy of the CBR methods used in predicting fuel consumption by haulage trucks and in determining whether the CBR technique can be used in predicting fuel consumption under different scenarios in open pit mining with particular reference to Orapa Mine.

Scenario 1: Using CBR to predict fuel consumption by haulage trucks

The CAT Truck #29 dataset was used to predict fuel consumption using three CBR techniques namely CBR-FSS, CBR and Naïve techniques. The results in Table 7 show the random sample (10%) of the actual fuel consumed by Truck #29 against the predicted values

of fuel consumed using the three techniques together with the absolute errors. In analysing the results, it is expected that a good prediction technique should have small absolute errors close to zero. Table 7 shows that the absolute errors from CBR-FSS are lower than those of the CBR and Naïve techniques. This shows that CBR-FSS is a better prediction technique in fuel consumption than the CBR and Naïve techniques.

Table 7 Random Sample (10%) of Actual Fuel Used, Predicted Fuel Used and Absolute Errors for CAT Truck #29 Dataset

Trip No.	Actual Fuel Consumed (litres) (x)	Predicted Fuel in litres (y)			Absolute Residuals/Errors (|x-y|)		
		CBR-FSS	CBR	Naïve	CBR-FSS	CBR	Naïve
5	92.74	92.07	89.03	85.17	0.68	3.71	7.58
20	87.06	86.12	99.38	85.17	0.95	12.31	1.90
21	86.12	86.59	93.21	85.17	0.47	7.09	0.95
29	21.77	20.50	39.74	85.17	1.26	17.97	63.40
47	90.85	90.69	84.23	85.17	0.16	6.63	5.68
49	88.96	89.21	99.33	85.17	0.25	10.38	3.79
62	101.26	101.23	111.20	85.17	0.03	9.94	16.09
70	86.12	85.65	88.96	85.17	0.47	2.84	0.95
74	88.01	88.64	156.22	85.17	0.63	68.21	2.85
107	29.34	28.39	64.16	85.17	0.95	34.83	55.83
114	88.01	88.48	118.07	85.17	0.47	30.05	2.85
120	90.85	90.85	99.79	85.17	0.00	8.94	5.68
134	102.21	103.15	97.98	85.17	0.95	4.22	17.04
138	34.07	33.83	83.48	85.17	0.24	49.42	51.10
139	35.02	35.02	35.96	85.17	0.00	0.95	50.15
142	35.02	35.96	142.66	85.17	0.95	107.64	50.15

Figure 18 is a boxplot showing the distribution of the residuals of the prediction techniques. It shows that the residuals from the Naïve technique were normally distributed while those of the CBR-FSS and CBR techniques were positively skewed. From Figure 18 the CBR-FSS technique has absolute residual values that are much closer to zero than the CBR and Naïve techniques. This indicates that CBR-FSS is a better prediction technique.

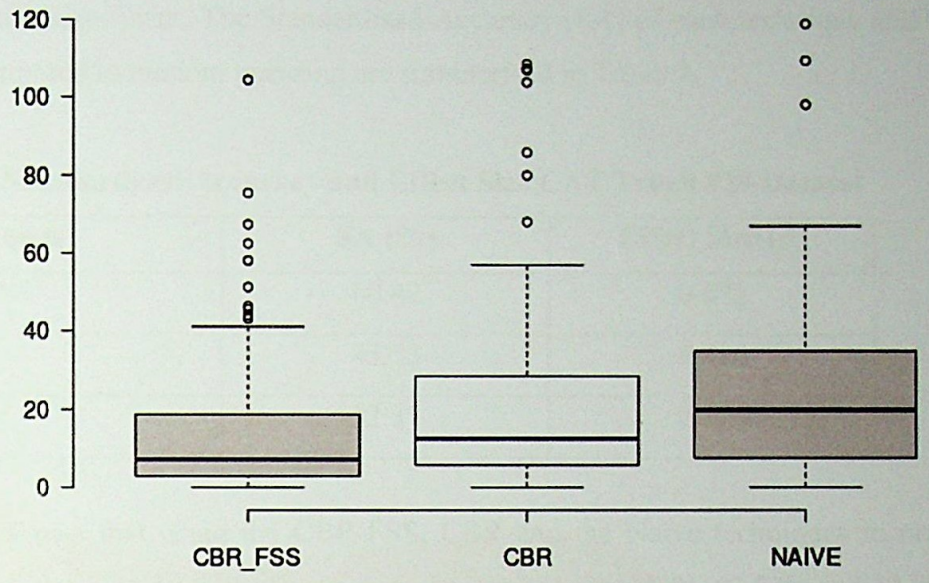


Figure 18 Distribution of Residuals for CAT Truck #29 Dataset

The results were analysed based on absolute residuals as in the Shepperd-McDonell method under Scenario 1. There was no need to look at the context variables because a repeated measure design was used. The three techniques CBR-FSS, CBR and Naïve were applied together with random guessing. Table 8 summarises the values of the mean absolute residuals obtained from the three predicting techniques and that of random guessing using the Python code for random guessing prediction in Figure 17. The MAR values were used to compute the standardised accuracy and effect sizes.

Table 8 Mean Absolute Residual CAT Truck #29 Dataset

Technique	MAR
CBR-FSS	14.24
CBR	20.24
Naïve	24.37
Random Guessing	35.97

Equations (17) and (18) and the Z-score distribution table were used to determine the statistical significance of the three prediction techniques against random guessing. The p-value was determined to be 0.0005 which is less than the significance level of 0.05. Therefore, the standard deviation between each technique against random guessing is

statistically significant. The Standardised Accuracy (SA) of each technique and the effect sizes compared to random guessing are summarised in Table 9.

Table 9 Standardised Accuracy and Effect Size CAT Truck #29 Dataset

Techniques	SA (%)	Effect Size (Δ)
CBR-FSS	60.42	0.672
CBR	43.73	0.486
Naïve	32.37	0.360

Table 9 shows that using the CBR-FSS, CBR and the Naïve techniques to predict fuel consumption by CAT Truck #29 per trip, the results are 60.42%, 43.73% and 32.37% better than random guessing respectively. Accordingly, the effect sizes for CBR-FSS, CBR and Naïve techniques are 0.672, 0.486 and 0.360 respectively. This means that the mean of random guessing is at the 76th percentile of the CBR-FSS technique, 69th percentile of the CBR technique and 66th percentile of the Naïve technique. Thus, using random guessing to predict fuel consumed on each trip by Truck #29 would be 76% higher than the values obtained by the CBR-FSS method, 69% higher than the fuel consumption predicted by the CBR technique and 66% of the fuel predicted by the Naïve technique. Considering the effect size, the value obtained by the CBR-FSS technique can be considered as of medium effect size while that for the CBR and Naïve technique may be considered as small.

Scenario 2: CBR Prediction of fuel consumption for during shifts

This section presents the predictions made by the three techniques using CAT Truck #3 dataset over the 3 shifts run in Orapa Mine. The random sample (10%) of the actual against predicted values and the absolute residuals are presented in Appendices 2 to 4.

The mean absolute error results obtained are summarised in Table 10 while Figures 19 to 21 are boxplots of the absolute residual distributions for each technique during the morning, afternoon and night shifts respectively.

Table 10 Mean Absolute Residuals for Shifts Prediction for CAT 793 Truck #3 Datasets

Technique	CAT 793 Truck #3 Morning Shift	CAT 793 Truck #3 Afternoon Shift	CAT 793 Truck #3 Night Shift
CBR-FSS	2.34	2.70	2.54
CBR	4.10	4.52	4.09
Naïve	4.89	12.25	6.00
Random Guessing	6.18	14.93	7.36

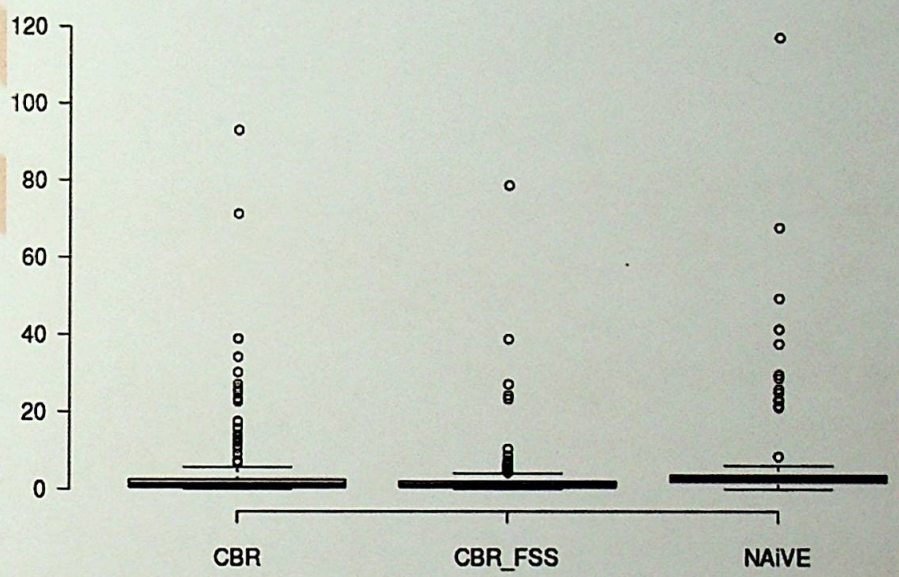


Figure 19 Distribution of Residuals for CAT 793 Truck #3 Morning Shift Dataset

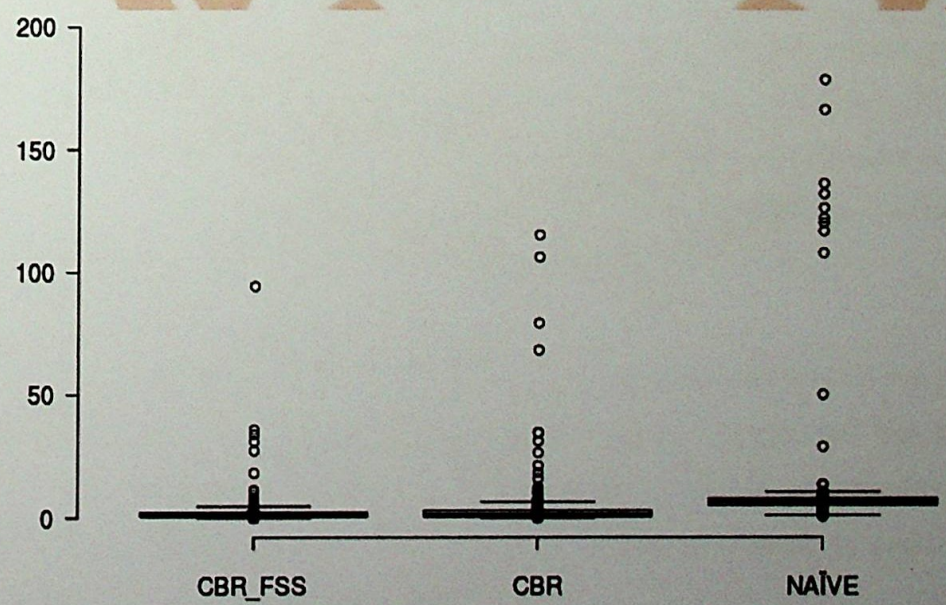


Figure 20 Distribution of Residuals for CAT 793 Truck #3 Afternoon Shift Dataset

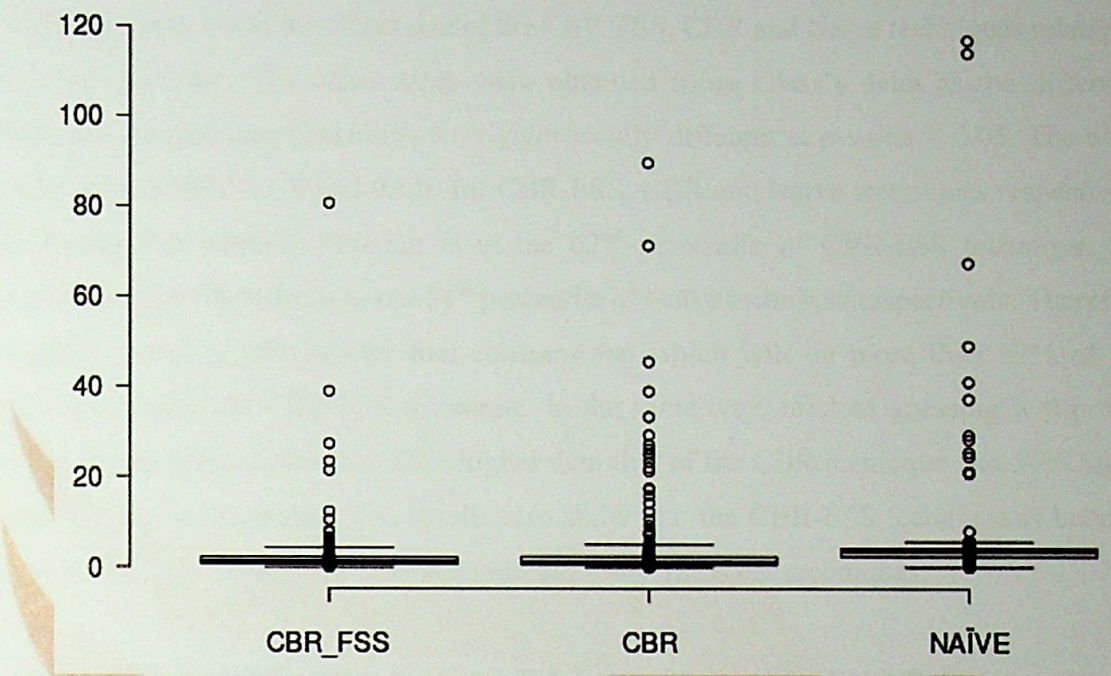


Figure 21 Distribution of Residuals for CAT 793 Truck #3 Night Shift Dataset

From Figure 19, the residuals from the Naïve technique are normally distributed while those of the CBR-FSS and CBR techniques are positively skewed. The CBR-FSS technique has a smaller median (as displayed by the black line in the whiskers) of 1.15 than the CBR and Naïve techniques whose medians are 1.18 and 2.84 respectively. Therefore, the CBR-FSS technique gives a better prediction of fuel consumed by haulage trucks as compared to other techniques in the morning shifts.

From Figure 20, the residuals from the Naïve technique are negatively skewed whereas the CBR-FSS and CBR techniques are positively skewed. The FSS technique has a smaller median of 1.24 than the Naïve and CBR-FSS techniques which have medians of 1.42 and 7.08 respectively. It shows that the CBR-FSS is better in predicting fuel consumption by haulage trucks than the Naïve and CBR techniques in the afternoon shifts.

Figure 21 shows that the residuals from CBR-FSS and CBR techniques are positively skewed as compared to the Naïve technique which is negatively skewed. The CBR-FSS technique has a median of 1.18 while the CBR and Naïve techniques have medians of 1.95 and 3.08 respectively indicating that the CBR-FSS technique is better in predicting fuel consumption by haulage trucks during night shifts.

Table 11 summarises the effect size of the CBR-FSS, CBR and Naïve techniques relative to random guessing. The effect sizes were obtained using Glass's delta as the difference between the standard deviations were significantly different at p-value < 0.05. The effect sizes were 0.382, 0.178 and 0.110 for CBR-FSS, CBR and Naïve techniques respectively. It means that random guessing is at the 62nd percentile of CBR-FSS technique, 58th percentile of CBR technique and 54th percentile of Naïve technique respectively. Therefore, random guessing will predict fuel consumption which will be more than 62% of that predicted using the CBR-FSS technique. In the same way, random guessing will predict average fuel consumptions of 58% higher than that of the CBR technique and 54% higher than the Naïve technique. The results also show that the CBR-FSS technique is better in predicting fuel consumption by haulage trucks than the other techniques.

Table 11 SA and Effect Size for CAT 793 Truck #3 Morning Shift Dataset

Technique	SA (%)	Effect Size (Δ)
CBR-FSS	62.01	0.328
CBR	33.58	0.178
Naïve	20.83	0.110

The SA values for the afternoon shift are summarised in Table 12 for each technique. compared to random guessing, the SA values are 81.91%, 69.77% and 17.99% for CBR-FSS, CBR and Naïve techniques respectively. The results for the Afternoon shift also show that the CBR-FSS technique is better at estimating fuel consumption by haulage trucks as it has a higher SA value. It is also noted that using the sample mean of the Naïve technique is 17.99% and it is a better fuel consumption predicting technique than random guessing.

Table 12 SA and Effect Size for CAT 793 Truck #3 Afternoon Shift Dataset

Technique	SA (%)	Effect Size (Δ)
CBR-FSS	81.91	0.422
CBR	69.77	0.360
Naïve	17.99	0.093

The effect sizes of 0.422, 0.360 and 0.093 respectively for CBR-FSS, CBR and Naïve techniques show that a trip from random guessing would predict fuel consumptions that are ≥ 66% of the fuel predicted using CBR-FSS, ≥ 62% CBR technique and ≥ 54% of the fuel predicted by Naïve technique. These effect sizes were computed using Glass's delta after it was determined that the standard deviation between the CBR prediction techniques and random guessing had a statistically significant p-value of 0.0001. The effect sizes in Table 12 also prove that the CBR-FSS technique is a better predictor of fuel consumption levels by trucks than the other techniques and random guessing.

Table 13 summarises the SA values of 65.45% for CBR-FSS, 44.39% for CBR and 18.51% for Naïve techniques compared to random guessing. The results show that the CBR-FSS technique is a better estimator because of its higher SA value. The standard deviations of all 3 techniques against random guessing were determined to be statistically significant i.e. the p-value < 0.05 (significance level). Hence, Glass's delta was used to determine the effect sizes.

Table 13 SA and Effect Size for CAT 793 Truck #3 Night Shift Dataset

Technique	SA (%)	Effect Size (Δ)
CBR-FSS	65.45	0.330
CBR	44.39	0.224
Naïve	18.51	0.094

From Table 13, an effect size of 0.330 means that random guessing is at the 62nd percentile of CBR-FSS. Thus, using random guessing, the average fuel consumed would be predicted to be more than 62% of the fuel predicted using CBR-FSS. In the same way, using random guessing, the average fuel consumed would be predicted to be ≥ 58% of the fuel predicted using CBR and ≥ 54% of the fuel consumed as predicted using the Naïve technique.

Scenario 3: CBR prediction of fuel consumption of different trucks (Komatsu 830E Truck #1 and CAT 793 Truck #2)

This section presents the results of the predictions made using the CAT 793 Truck #2 dataset and the Komatsu 830E Truck #3 dataset. A random sample of the actual data and the

predicted fuel used are presented in Appendices 5 and 6. The results of Scenario 3 were also based on absolute residuals which led to the calculation of SA values and effect sizes.

The mean absolute residuals obtained using CBR-FSS, CBR, Naïve technique and random guessing prediction techniques are summarised in Table 14.

Table 14 Mean Absolute Residual of Komatsu 830E #1 and CAT 793 #2 Datasets

Technique	Komatsu 830E Truck #1	CAT 793 Truck #2
CBR-FSS	0.49	3.16
CBR	10.61	5.54
Naïve	16.62	11.08
Random Guessing	28.17	14.07

Figure 22 shows the residual distribution for each prediction technique for the Komatsu 830E #1 truck dataset. The plots of the Naïve and CBR techniques are positively skewed while that of the CBR-FSS technique is normally distributed. The plot of CBR-FSS data shows that the absolute errors are close to zero which is a desirable condition for good prediction techniques. The CBR-FSS technique also has a smaller median than the other techniques. Furthermore, the absolute residual of the CBR-FSS technique is closer to zero than those of CBR and Naïve techniques. This is an indication that CBR-FSS is a better prediction method than the other techniques.

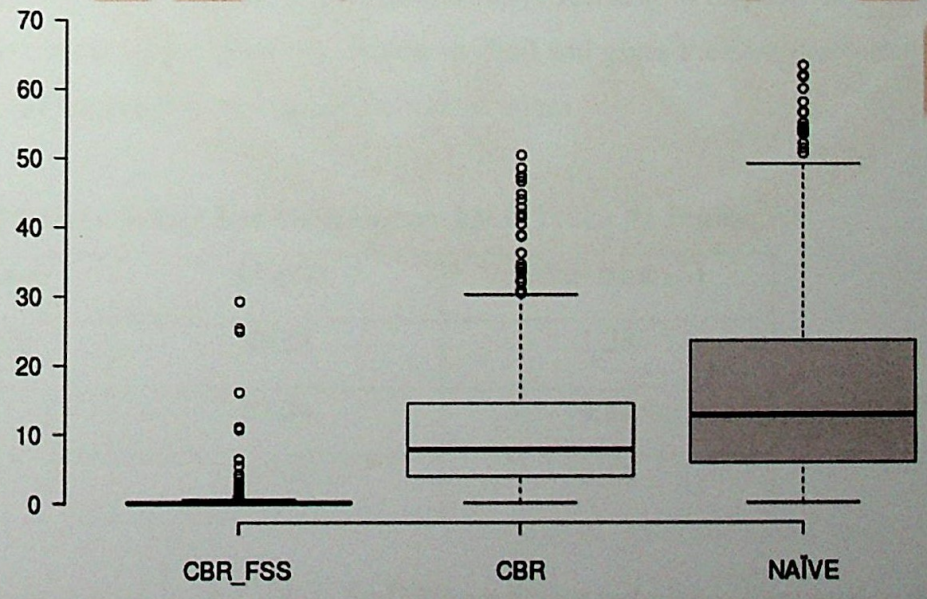


Figure 22 Boxplots of Distribution of Residuals for Komatsu 830E Truck #1 Dataset

Figure 23 shows the distribution of residuals for the prediction techniques for Caterpillar trucks. The plots of Naïve and CBR techniques are positively skewed while that of the CBR-FSS technique is normally distributed. The CBR-FSS has a smaller median than the other techniques which indicates that it is a better prediction method than the other techniques.

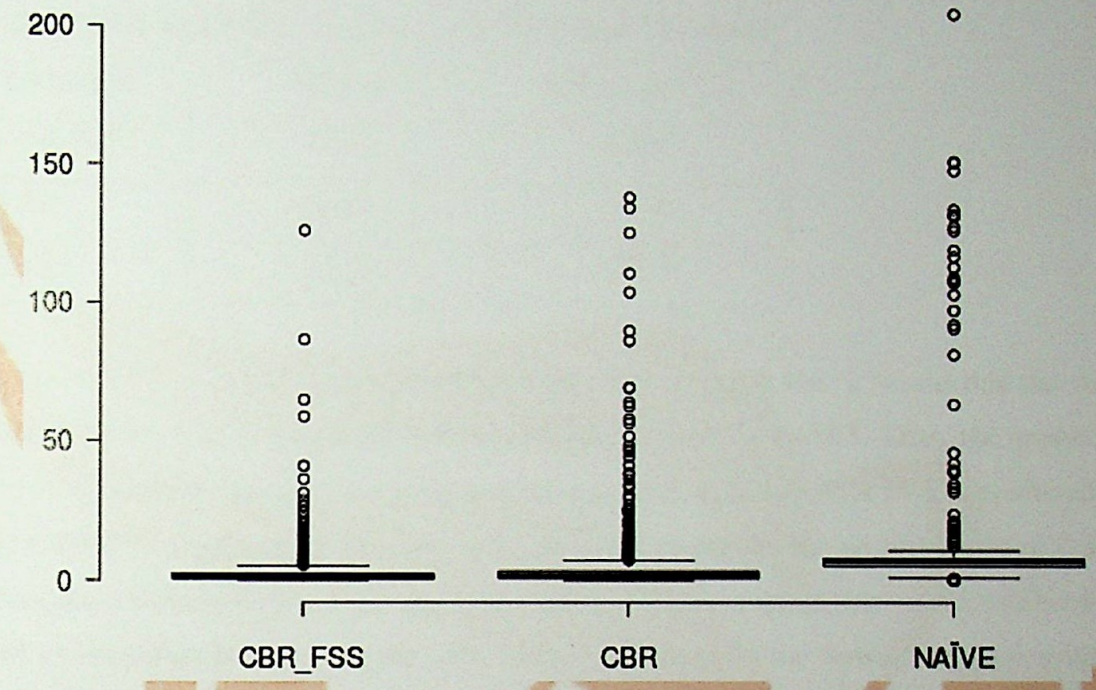


Figure 23 Boxplots showing Distribution of Residuals for CAT 793 Truck #2 Dataset

Table 15 shows the SA values for CBR prediction techniques compared to the random guessing technique that was obtained using the results from Table 14. They are 98.28%, 62.34% and 40.99% for CBR-FSS, CBRand Naïve techniques respectively. These results show that CBR-FSS is a better estimation method and gives the best prediction results in terms of fuel consumption by haulage trucks in mines.

Table 15 SA and Effect Size for Komatsu 830E Truck #1 Dataset

Technique	SA (%)	Effect Size (Δ)
CBR-FSS	98.28	1.28
CBR	62.34	0.81
Naïve	40.99	0.53

Table 16 summarises the SA values for CBR prediction techniques compared to the random guessing technique obtained using the results in Table 14. They are 77.54%, 60.62% and

21.25% for CBR-FSS, CBRand Naïve techniques respectively. The results also show that the CBR-FSS technique is the best estimation method as it gives the best prediction results in fuel consumption by haul trucks.

Table 16 SA and Effect Size for CAT 793 Truck #2 Dataset

Technique	SA (%)	Effect Size (Δ)
CBR-FSS	77.54	0.435
CBR	60.62	0.340
Naïve	21.25	0.119

From Table 16, CBR-FSS techniques have an effect size of 0.435. It means that the value obtained from random guessing is at the 66th percentile of CBR-FSS. Thus, the prediction of fuel consumed during a trip using random guessing will be ≥ 66% of that predicted by the CBR-FSS technique. In the same way, the CBR technique has an effect size of 0.340. Thus, the random guessing technique will predict fuel consumption that would ≥ 62% of the fuel consumption predicted by the CBR technique. Lastly for the Naïve technique, with an effect size of 0.119, the mean of the random guessing technique is at the 54th percentile of the traditional Naïve technique. Thus, the predicted fuel consumption for a trip using random guessing with would be ≥ 54% of the fuel consumption as predicted by the Naïve technique. Thus, using effect sizes as the criterion for comparison of prediction accuracy, the CBR-FSS is a better predictor than the other techniques.

The datasets used in the analysis in Scenario 3 (Komatsu Truck #3 and CAT Truck #2) were larger (514 and 505 cases respectively) than those used in Scenario 2 (CAT Truck #3_Morning, CAT Truck #3_Afternoon, CAT Truck #3_Night). Large datasets usually have larger SA values than smaller datasets (see Tables 8 to 15). Therefore, the larger the dataset the higher the accuracy of the prediction technique.

Figure 24 shows the absolute residuals of each prediction technique. It can be observed from Figure 24 that as the dataset grows, the magnitude of absolute residuals/errors decreases for each prediction technique. For example, for 100 trips the absolute errors are 4, 8 and 15 for the CBR-FSS, CBRand Naïve techniques respectively whereas for 500 trips the absolute residuals are 2.5, 5 and 10 for CBR-FSS, CBRand Naïve techniques respectively. Thus, the

accuracy of the prediction technique improves with the growth of the datasets. Figure 23 also shows that the CBR-FSS technique has the highest prediction accuracy than the other prediction technique used in this study as it has smaller errors in prediction than the other techniques irrespective of the number of trips used in the data set.

However, Figure 24 also shows that the values of the predicted absolute residuals tend to flatten out between 250 and 400 trips. This could be because either some important features were no longer important, or the new cases predicted had no similar cases to the previous cases used in the predictions.

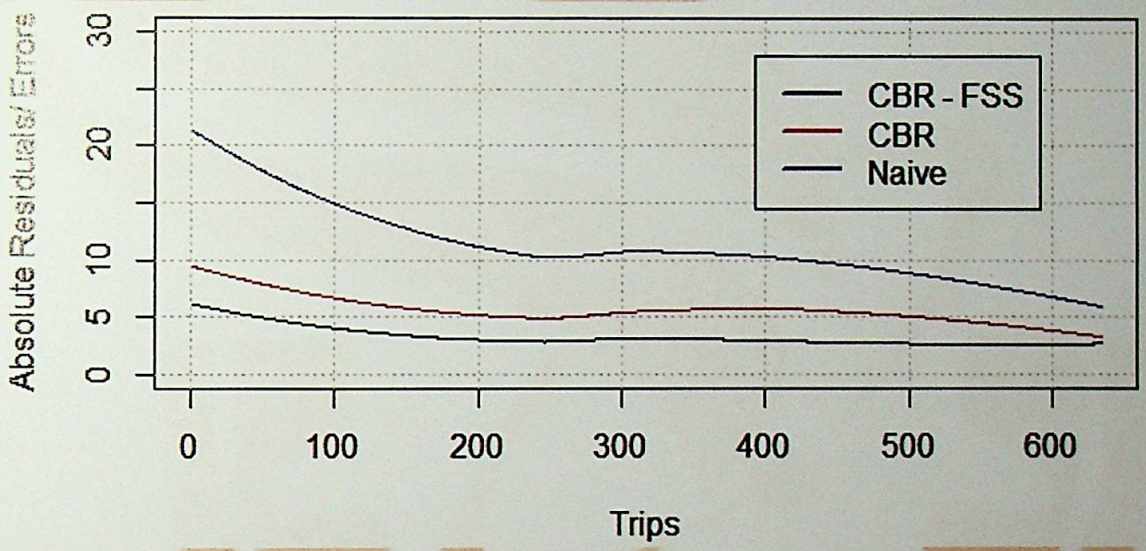


Figure 24 Absolute Residuals for CAT 793 Truck#2 Dataset

4.3 Discussions

The results based on Scenario 1 show that the CBR method can be used to predict fuel consumption by haulage trucks in mines. This is a significant finding as the technique has never been applied in predicting fuel consumption in mining.

The values of absolute residuals/errors obtained in this work show that the case-based reasoning using forward sequential selection (CBR-FSS) technique makes better predictions than the traditional CBR and the Naïve techniques in all the scenarios. The results also show that the CBR-FSS technique predicts fuel consumption values that are much closer to the actual values consumed by haul trucks than the other prediction techniques used in this study.

The results also show that there is a significant difference between the values predicted by the CBR-FSS and the other techniques in terms of standard accuracy (SA) and effect sizes. The CBR-FSS technique performed better than CBR and Naïve techniques as it consistently predicted higher SA values and effect sizes using all the datasets in this research.



CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

The objectives of this research work was to propose and implement a system for predicting fuel consumption by haulage trucks in open pit mines. Thereafter, evaluate the system against benchmarks in terms of accuracy and performance. Therefore from the analysis in this study, it is concluded that:

i) The case-based reasoning (CBR) method can be effectively used to predict the fuel consumption of haulage trucks in open-pit mines.

ii) The predicted values of fuel consumption using the CBR-FSS technique gave much lower absolute error/residual values than those of other prediction techniques on all the datasets of CAT 789 Truck #29, CAT 793 Truck #3 (Morning, Afternoon and Night shifts), Komatsu 830E Truck #1 and CAT 793 Truck #2 trucks.

iii) Prediction of fuel consumption using Case-based reasoning using forward sequential selection (CBR-FSS) technique gave higher standardised accuracies (SA values) and effect sizes than the other prediction techniques when evaluated against random guessing technique. This shows that the CBR-FSS technique can be used as a mine planning tool to predict fuel consumption using data from the previous trips.

5.2 Recommendations

From the results of this work, it is recommended that:

i. Mine managers can use the CBR method in predicting fuel consumption by haulage trucks in open-pit mines and it will assist them in identifying trucks with unusually high fuel consumption for early maintenance work to be done on them.

ii. In future studies, larger datasets from other mines that have effects of weather should be used to improve the accuracy of the predictions and make the findings more generalised or applicable to all conditions in open-pit mines.

iii. The fuel prediction system should be applied to ther mining scenarios such as maintenance or underground mining prediction

REFERENCES

Aamodt, A. and Plaza, E. (1994), "Case based Reasoning: Foundational Issues, Methodological Variationsand System Approaches", *AI Communications: The European Journal on Artificial Intelligence*, IOS, Vol. 7, No. 1, pp. 39 - 59.

Adair, L., Soofastaei, A., Aminossadati, S. M., Kizil, M. S. and Knights, P. (2015), "Rolling Resistance in Haul Truck Operations", *Australian Mining*, Vol. 107, No. 6, pp. 1 - 7.

Amasaki, S. and Lokan, C. (2014), "The Effect of Moving Windows on Software Effort Estimation: Comparative Study with CART", *6th International Workshop on Empirical Software Engineering in Practice*, IEEE, Osaka, Japan, pp. 1 - 6.

Amasaki, S. and Lokan, C. (2015), "On the Effectiveness of Weighted Moving Windows: Experiment on Linear Regression Based Software Effort Estimation", *Journal of Software: Evolution and Process*, Wiley Online Library, Vol. 27 No. 7, pp. 488 - 507.

Angelis, L. and Stamelos, I. (2000), "A Simulation Tool for Efficient Analogy-Based Cost Estimation", *Empirical Software Engineering*, Vol 5, N0 1, pp. 35 - 68.

Angra, S. and Ahuja, S. (2017), "Machine Learning and Its Applications: A Review", *2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC)*, IEEE, Chirala, India, pp. 57 - 60.

Anon. (2015), "The Real Effects of Engine Idle Time", *Teletrac Navman*, www.telematrac.com. (Accessed: September 20, 2018).

Anon. (2017), *Caterpillar Performance Handbook*, 47th ed., Catepilar Inc, Peoria III, pp. 1014.

Anon. (2018a), "Orapa Diamond Mine - Mining Technology", www.mining-technology.com/projects/orapa-diamond-mine-botswana. (Accessed: October 26, 2018).

Anon. (2018b), "Orapa Diamond Mine", www.google.com/maps/search. (Accessed: August 10, 2018).

Anon. (2018c), "Debswana Orapa Mine", www.debswana.com/Operations/Pages/Orapa-Mine. (Accessed: November 16, 2018).

Anon. (2018d), "Debswana Value Chain", www.debswana.com/About-Us/Pages/Value-Chain. (Accessed: November 16, 2018).

Anon. (2019), "n-Dimensional Euclidean Distance", www.hlab.stanford.edu/brian/euclidean_distance_in.html. (Accessed: February 18,

2019).

Awuah-Offei, K. (2016), "Energy Efficiency in Mining: A Review with Emphasis on the Role of Operators in Loading and Hauling Operations", *Journal of Cleaner Production*, Elsevier, Vol. 117, pp. 89 - 97.

Awuah-Offei, K., Osei, B. and Askari-Nasab, H. (2012), "Improving Truck-Shovel Energy Efficiency through Discrete Event Modeling", *SME Annual Meeting and Exhibit 2012, SME 2012, Meeting Preprints*, Society for Mining, Metallurgy and Exploration Inc. (SME), Seattle, Washington, pp. 333 - 339.

Ayoung, A. D., Sigweni, B. and Abbott, P. (2015), "Case-based Reasoning System for Predicting the Sustainability of a Telecentre", *10th International Conference for Internet Technology and Secured Transactions (ICITST)*, IEEE, London, UK, pp. 125 - 130.

Azzeh, M. and Nassif, A. B. (2015), "Analogy-Based Effort Estimation: A New Method to Discover Set of Analogies from Dataset Characteristics", *IET Software* , Vol. 9, pp. 39 - 50.

Badamich, A. (2015), "Case Based Reasoning Case Study: Housing Price", *Engineering*, pp. 1 - 19.

Bajta, M. (2015), "Analogy-Based Software Development Effort Estimation in Global Software Development", *2015 IEEE 10th International Conference on Global Software Engineering Workshops*, IEEE, Ciudad Real, Spain, pp. 51 - 54.

Bansal, S. (2014), "Supervised and Unsupervised Learning - GeeksforGeeks", www.geeksforgeeks.org/supervised-unsupervised-learning. (Accessed: November 8, 2018).

Baranowski, R., Chen, Y. and Fryzlewicz, P. (2018), "Ranking-Based Variable Selection for High-Dimensional Data", Houghton Street, London, pp. 1 - 32.

Bishop, T. S. (1968), "Trucks", in Chapter 9.2 of *Surface Mining*, 1st ed., (E. P. Pfleider, sen. ed.), SME - AIME, New York, pp. 553 - 583.

Brady, S. and Van Order, D. (2012), *Truck Idling Scoping Study*, McLean, Virginia, pp. 1 - 36.

Brownlee, J. (2014), "An Introduction to Feature Selection", www.machinelearningmastery.com/an-introduction-to-feature-selection. (Accessed: October 26, 2018).

Brownlee, J. (2016), "Supervised and Unsupervised Machine Learning Algorithms", www.machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms. (Accessed: November 8, 2018).

Brundell-Freij, K. and Ericsson, E. (2005), "Influence of Street Characteristics, Driver Category and Car Performance on Urban Driving Patterns", *Transportation Research Part D: Transport and Environment*, Pergamon, Vol. 10, No. 3, pp. 213 - 229.

Carpatorea, I., Nowaczyk, S., Rognvaldsson, T., Elmer, M. and Lodin, J. (2016), "Learning of Aggregate Features for Comparing Drivers Based on Naturalistic Data", *15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, IEEE, Anaheim, CA, USA, pp. 1067 - 1072.

Cheetham, B. and Goebel, K. (2004), "Case-Based Reasoning Watson Chapter 1-4", *University of Washington*, pp. 1 - 12.

Chen, J. (2019), "Autoregressive Integrated Moving Average (ARIMA)", www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp. (Accessed: May 29, 2019).

Coe, R. (2002), "It's the Effect Size, Stupid: What Effect Size is and Why it is Important", www.leeds.ac.uk/educol/documents/00002182.htm. (Accessed: February 20, 2020).

Coffey, J. P. (2016), "Mine Haul Road Rolling Resistance: Influences and Impacts", Curtin University, pp. 1 - 441.

Coronado, P. V. (2014), "Optimization of the Haulage Cycle Model for Open Pit Mining Using a Discrete-Event Simulator and a Context-Based Alert System", *University of Arizona*, pp. 13 - 157.

Dey, S., Mandal, S. K. and Bhar, C. (2017), "Analysis of Factors, which Influence the Cycle Time of Dumpers of Open Cast Coal Mines to Improve Production", *Advances in Modelling and Analysis A*, Vol. 78, No. 3, pp. 289 - 302.

Dhote, Y., Agrawal, S. and Deen, A. J. (2015), "A Survey on Feature Selection Techniques for Internet Traffic Classification", *International Conference on Computational Intelligence and Communication Networks (CICN)*, IEEE, Jabalpur, India, pp. 1375 - 1380.

Dindarloo, S. R. and Siami-Irdemoosa, E. (2016), "Determinants of Fuel Consumption in Mining Trucks", *Energy*, Pergamon, Vol. 112, pp. 232 - 240.

Dotto, M. S. (2014), "Impact of Hauler Scale in Mine Planning", University of Alberta, pp. 1 - 92.

Dumakor, N. K., Temeng, V. A. and Bansah, K. J. (2017), "Optimising Shovel-Truck Fuel Consumption using Stochastic Simulation", *Ghana Mining Journal*, University of Mines and Technology, Vol. 17, No. 2, pp. 39 - 49.

El Aboudi, N. and Benhlima, L. (2016), "Review on Wrapper Feature Selection

Approaches", *International Conference on Engineering and MIS (ICEMIS)*, IEEE, Agadir, Morocco, pp. 1 - 5.

Ellis, P. D. (2010), *The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysisand the Interpretation of Research Results,* Paul D. Ellis - Google Books, 1st edition, Cambridge University Press, New York, USA, 192pp.

Ericsson, E. (2001), "Independent Driving Pattern Factors and their Influence on Fuel-Use and Exhaust Emission Factors", *Transportation Research Part D: Transport and Environment*, Pergamon, Vol. 6, No. 5, pp. 325 - 345.

Fontaras, G., Zacharof, N.-G. and Ciuffo, B. (2017), "Fuel Consumption and CO_2 Emissions from Passenger Cars in Europe – Laboratory versus Real-World Emissions", *Progress in Energy and Combustion Science*, Pergamon, Vol. 60, pp. 97 - 131.

Foss, T., Strensrnd, B., Kitchenham, B. and Myrtveit, I. (2003), "A Simulation Study of the Model Evaluation Criterion MMRE", *Software Engineering,* IEEE Transactions. Vol. 29 (11), pp. 985 -995.

Glen, S. (2016), "Cohen's D: Definition, Examples, Formulas - Statistics How To", www.statisticshowto.datasciencecentral.com/cohens-d. (Accessed: February 18, 2020).

Gross, D., Shortle, J. F., Thompson, J. M. and Harris, C. M. (2008), *Fundamentals of Queueing Theory*, (4th ed.), John Wiley and Sons, Incorporated, Newark, pp. 1 - 509.

Gupta, U. G. (1994), "How Case-Based Reasoning Solves New Problems", *Interfaces*, INFORMS, Vol. 24, No. 6, pp. 110 - 119.

Guyon, I. and Elisseeff, A. (2003), "An Introduction to Variable and Feature Selection", *Journal of Machine Learning Research*, Vol. 3 (7-8), pp. 1157 - 1182.

Hall, M. A. and Smith, L. A. (1997), "Feature Subset Selection: A Correlation Based Filter Approach", *International Conference on Neural Information Processing and Intelligent Information Systems*, Springer, Berlin, pp. 855 - 858.

Hargroves, K., Gockowiak, K., McKeague, F. and Desha, C. (2014), "An Overview of Energy Efficiency Opportunities in Mining and Metallurgy Engineering - Google Search", Canberra, pp. 1 - 25.

Hira, Z. M. and Gillies, D. F. (2015), "A Review of Feature Selection and Feature Extraction Methods Applied on Microarray Data", *Advances in Bioinformatics*, pp. 1 - 13.

Holman, P. (2006), "Caterpillar Haul Road Design and Management", *www.directminingservices.com*, CAT Global Mining, St. Charles, IL, pp. 1 - 103.

Imani, M. B., Keyvanpour, M. R. and Azmi, R. (2013), "A Novel Embedded Feature

Selection Method: A Comparative Study in the Application of Text Categorization", *Applied Artificial Intelligence*, Taylor and Francis Group, Vol. 27, No. 5, pp. 408 - 427.

Ishibashi, M., Okuwa, M., Doi, S. and Akamatsu, M. (2007), "Indices for Characterising Driving Style and their Relevance to Car Following Behaviour", *SICE Annual Conference 2007*, IEEE, Takamatsu, Japan, pp. 1132 - 1137.

Johansson, J. (2016), "What is Grade Resistance?", *Johansson and Associates, LLC*, www.josephjohansson.com/truck-science/grade-resistance. (Accessed: October 29, 2018).

Jorgensen, M. and Molokken-Ostvold, K., (2006) "How Large are Software Cost Overruns? A Review of the 1994 Chaos Report". *Information and Software Technology*. Vol. 48(4), pp. 297 -301.

Jovic, A., Brkic, K. and Bogunovic, N. (2015), "A Review of Feature Selection Methods with Applications", *38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, IEEE, Opatija, Croatia, pp. 1200 - 1205.

Kafaf, D. AL, Kim, D.-K. and Lu, L. (2017), "B-kNN to Improve the Efficiency of kNN", *Proceedings of the 6th International Conference on Data Science, Technology and Applications (DATA 2017)*, Science and Technology Publications, Madrid, Spain, pp. 126 - 132.

Kassambara, A. (2018), "KNN: K-Nearest Neighbors Essentials", *Statistical Machine Learning Essentials*, www.sthda.com. (Accessed: March 13, 2019).

Kaufman, W. W. and Ault, J. C. (1977), *Design of Surface Mine Haulage Roads - A Manual*, No. BuMines-IC-8758, Washington, DC, United States, pp. 1 - 50.

Kaushik, S. (2016), "Feature Selection Methods with Example (Variable Selection Methods)", *Analytics Vidhya*, www.analyticsvidhya.com/introduction-to-feature-selection-methods-with-an-example-or-how-to-select-the-right-variables. (Accessed: November 16, 2018).

Kecojevic, V. and Komljenovic, D. (2010), "Haul Truck Fuel Consumption and CO_2 Emission under Various Engine Load Conditions", *Society for Mining, Metallurgy and Exploration Inc.*, Vol. 62, No. 12, pp. 44 - 48.

Kecojevic, V. and Komljenovic, D. (2011), "Impact of Bulldozer's Engine Load Factor on Fuel Consumption, CO_2 Emission and Cost", *American Journal of Environmental Sciences*, Vol. 7, No. 2, pp. 125 - 131.

Kenton, W. (2019), "Multiple Linear Regression - MLR Definition",

www.investopedia.com/terms/m/mlr.asp. (Accessed: May 18, 2019).

Keung, J. (2009), "Software Development Cost Estimation Using Analogy: A Review", *2009 Australian Software Engineering Conference*, IEEE, pp. 327 - 336.

Keung, J. W. (2008), "Theoretical Maximum Prediction Accuracy for Analogy-Based Software Cost Estimation", *15th Asia-Pacific Software Engineering Conference*, IEEE, Beijing, China, pp. 495 - 502.

Kim, H.-Y. (2015), *Statistical Notes for Clinical Researchers: Effect Size Open Lecture on Statistics*, Seongbukgu, Seoul, Korea, https://doi.org/10.5395/rde.2015.40.4.328. (Accessed: February 20, 2020).

Kirsopp, C. and Shepperd, M. (2003), "Case and Feature Subset Selection in Case-Based Software Project Effort Prediction", *Research and Development in Intelligent Systems XIX*, (Bramer, M., Preece, A. and Coenen, F., eds.), Springer, London, pp. 61 - 74.

Kitchenham, B., Pickard, L., MacDonell, S. and Shepperd, M. (2001), "What Accuracy Statistics Really Measure Software Estimation", *Software IEEE Proceedings*. Vol. 148, pp. 81 - 85.

Klair, A. S. and Kaur, R. P. (2012), "Software Effort Estimation using k-Nearest Neighbour (kNN)", *International Conference on Artificial Intelligence and Image Processing (ICAIIP'2012)*, Dubai, United Arab Emirates, pp. 216 - 218.

Klanfar, M., Korman, T. and Kujundžić, T. (2016), "Fuel Consumption and Engine Load Factors of Equipment in Quarrying of Crushed Stone", *Tehnicki Vjesnik - Technical Gazette*, Strojarski fakultet u Slavonskom Brodu, Elektrotehnički fakultet u Osijeku, Građevinski fakultet u Osijeku, Vol. 23, No. 1, pp. 163 - 169.

Knights, P. and Boerner, A. (2001), *Statistical Correlation of Off-Highway Tire Failures with Openpit Haulage Routes*, No. 99-345, Santiago, Chile, pp. 51 - 56.

Kocsis, T., Negny, S., Floquet, P., Meyer, X. and Rév, E. (2014), "Case-Based Reasoning System for Mathematical Modelling Options and Resolution Methods for Production Scheduling Problems: Case Representation, Acquisition and Retrieval", *Computers and Industrial Engineering*, Pergamon, Vol. 77, pp. 46 - 64.

Kohavi, R. and John, G. H. (1997), "Wrappers for Feature Subset Selection", *Artificial Intelligence*, Elsevier, Vol. 97, No. 1 - 2, pp. 273 - 324.

Kolodner, J. L. (1992), "An Introduction to Case-Based Reasoning", *Artificial Intelligence Review*, Kluwer Academic Publishers, Vol. 6, No. 1, pp. 3 - 34.

Kumari, B. and Swarnkar, T. (2011), "Filter versus Wrapper Feature Subset Selection in Large Dimensionality Micro Array: A Review", *International Journal of Computer*

Science and Information Technologies, Vol. 2, No. 3, pp. 1048 - 1053.

Li, Y. F., Xie, M. and Goh, T. N. (2009), "A Study of Mutual Information Based Feature Selection for Case-Based Reasoning in Software Cost Estimation", *Expert Systems with Applications*, Vol. 36, pp. 5921 - 5931.

Lin, H., Cui, X., Yu, Q. and Yang, S. (2011), "Experimental Study on Diesel Vehicle's Fuel Consumption Feature while Coasting on Level Road", *International Conference on Computer Science and Information Engineering CSIE 2011*, Springer, Zhengzhou, China, pp. 264 - 270.

Lokan, C. and Mendes, E. (2009), "Applying Moving Windows to Software Effort Estimation", *3rd International Symposium on Empirical Software Engineering and Measurement, IEEE*, Lake Buena Vista, FL, USA, pp. 111 - 122.

Lutsey, N., Brodrick, C.-J., Sperling, D. and Oglesby, C. (2004), "Heavy-Duty Truck Idling Characteristics: Results from a Nationwide Truck Survey", *Transportation Research Record: Journal of the Transportation Research Board*, Transportation Research Board of the National Academies, Vol. 1880, pp. 29 - 38.

Maphane, K. (2017), "Evolution of the Orapa A/K1 Geology Model - Insights from Analysis of Multi-Disciplinary Datasets", *11th International Kimberlite Conference*, Gaborone, Botswana, pp. 1 - 3.

May, M. A. (2013), "Applications of Queuing Theory for Open-Pit Truck/Shovel Haulage Systems", Virginia Tech, pp. 1 - 79.

McLeod, S. (2019), "What Does Effect Size Tell You? Simply Psychology", www.simplypsychology.org/effect-size.html. (Accessed: February 20, 2020).

Meech, J. and Parreira, J. (2013), "Predicting Wear and Temperature of Autonomous Haulage Truck Tires", *16th IFAC Symposium on Automation in Mining, Mineraland Metal Processing*, The University of British Columbia, Department of Mining Engineering, San Diego, California, USA, pp. 142 - 147.

Mero, J. L., Clark, G. B. and Hustrulid, W. A. (2017), "Mining - Britannica.com", *Encyclopedia Britannica*, www.britannica.com/technology/mining. (Accessed: June 2, 2018).

Michaelson, S. D. (1979), "Open Pit Mining - Past, Present and Future", *Open Pit Mine Planning and Design*, Port City Press, Baltimore, pp. 3 - 16.

Min, F., Hu, Q. and Zhu, W. (2014), "Feature Selection with Test Cost Constraint", *International Journal of Approximate Reasoning, Elsevier*, Vol. 55, No. 1, pp. 167 - 179.

Min, H. and Fangfang, W. (2010), "Filter-Wrapper Hybrid Method on Feature Selection", *Second WRI Global Congress on Intelligent Systems, IEEE*, Wuhan, China, pp. 98 - 101.

Monfet, D., Corsi, M., Choinière, D. and Arkhipova, E. (2014), "Development of an Energy Prediction Tool for Commercial Buildings Using Case-Based Reasoning", *Energy and Buildings*, Elsevier Ltd., Vol. 81, No. 1, pp. 152 - 160.

Moore, M. (2009), "Imaging Diamond with X-Rays", *Journal of Physics: Condensed Matter*, IOP Publishing, Vol. 21, No. 36, p. 1 - 15.

Motlogelwa, O. G. and Minnitt, R. C. A. (2013), "Optimization of diesel usage at Uitvlugt Mine", *Journal of the Southern African Institute of Mining and Metallurgy*, South African Institute of Mining and Metallurgy, Vol. 113, No. 4, pp. 345 - 349.

Myrtveit, I., Strensrund, E. and Shepperd, M. (2005), "Reliability and Validity in Comparative Studies of Software Prediction Models", *Software Engineering,* IEEE Transactions. Vol. 31(5), pp. 380 -391.

Nanda, S. (1988), "III Box-Jenkins Methods 1. Pros and Cons of ARIMA Forecasting", www.ecostat.unical.it/tarsitano/didattica/SeStoCor. (Accessed: May 26, 2019).

Necmettin, Ç, Ç. (2004), "Open-Pit Truck and Shovel Haulage System Simulation", *Graduate School of Natural and Applied Sciences of Middle East Technical University*, Ankara, Turkey, pp. 1 - 133.

Obitko, M. (1999), "Introduction - Prediction Using Neural Networks", www.obitko.com/tutorials/neural-network-prediction. (Accessed: May 18, 2019).

Ongsulee, P. (2017), "Artificial Intelligence, Machine Learningand Deep Learning", *15th International Conference on ICT and Knowledge Engineering (ICT&KE)*, IEEE, Bangkok, Thailand, pp. 1 - 6.

Pal, S. K. and Shiu, S. C. (2004), *Foundations of Soft Case-Based Reasoning*. John Wiley and Sons, Inc, Hoboken, New Jersey, Vol. 8, pp. 136 - 160.

Pandey, M. K. and Subbiah, K. (2018), "Performance Analysis of Time Series Forecasting Using Machine Learning Algorithms for Prediction of Ebola Casualties", *International Conference on Application of Computing and Communication Technologies*, Vol. 899, Springer Verlag, pp. 320 - 334.

Paterlini, G., Yucel, S., Lucking, M. M. and Magnuson, J. (2015), *Improving Fuel Efficiency by Measuring Rolling Resistance*, No. MN/RC 2015-39, St. Paul, Minnesota, pp. 1 - 2.

Patterson, S. R., Kozan, E. and Hyland, P. (2016), "An Integrated Model of an Open Pit Coal Mine: Improving Energy Efficiency Decisions", *International Journal of*

Production Research, Vol. 54, No. 14, pp. 4213 - 4227.

Poole, D. and Mackworth, A. (2010), "Artificial Intelligence - Foundations of Computational Agents - 7.6 Case-Based Reasoning", www. artint.info/html/ArtInt_190.html. (Accessed: September 26, 2018).

Quevedo, J. R., Bahamonde, A. and Luaces, O. (2007), "A Simple and Efficient Method for Variable Ranking According to their Usefulness for Learning", *Computational Statistics and Data Analysis*, North-Holland, Vol. 52, No. 1, pp. 578 - 595.

Rahman, S. M. A., Masjuki, H. H., Kalam, M. A., Abedin, M. J., Sanjid, A. and Sajjad, H. (2013), "Impact of Idling on Fuel Consumption and Exhaust Emissions and Available Idle-Reduction Technologies for Diesel Vehicles - A Review", *Energy Conversion and Management*, Pergamon, Vol. 74, pp. 171 - 182.

Ramirez-Gallego, S., Mourino-Talin, H., Martinez-Rego, D., Bolon-Canedo, V., Benitez, J. M., Alonso-Betanzos, A. and Herrera, F. (2018), "An Information Theory-Based Feature Selection Framework for Big Data Under Apache Spark", *IEEE Transactions on Systems, Manand Cybernetics: Systems*, Vol. 48, No. 9, pp. 1441 - 1453.

Richardson, S. and McIver, J. (2015), "Improving Mine Haul Road Roughness to Reduce Haul Truck Fuel Consumption", *18th Asia Pacific Automotive Engineering Conference*, Melbourne, Australia., pp. 1 - 6.

Rios-Torres, J., Liu, J. and Khattak, A. (2018), "Fuel Consumption for Various Driving Styles in Conventional and Hybrid Electric Vehicles: Integrating Driving Cycle Predictions with Fuel Consumption Optimization", *International Journal of Sustainable Transportation*, Taylorand Francis, pp. 1 - 15.

Rodovalho, E. da C., Lima, H. M. and de Tomi, G. (2016), "New Approach for Reduction of Diesel Consumption by Comparing Different Mining Haulage Configurations", *Journal of Environmental Management*, Academic Press, Vol. 172, pp. 177 - 185.

Rolim, C., Baptista, P., Duarte, G., Farias, T. and Pereira, J. (2017), "Real-Time Feedback Impacts on Eco-Driving Behavior and Influential Variables in Fuel Consumption in a Lisbon Urban Bus Operator", *IEEE Transactions on Intelligent Transportation Systems*, Vol. 18, No. 11, pp. 3061 - 3071.

Runge, I. C. (1998), "Mining Economics and Strategy", *Society for Mining, Metallurgy and Exploration Inc.*, Englewood, Colorado, pp. 103 - 127.

Sahoo, L. K., Bandyopadhyay, S. and Banerjee, R. (2010), "Energy Performance of Dump Trucks in Opencast Mine", Department of Energy Science and Engineering, *Proceedings of the 23rd International Conference on Efficiency, Cost, Optimization,*

Simulationand Environmental Impact of Energy Systems, ECOS 2010, CreateSpace, Lausanne, Switzerland, pp. 1899 - 1906.

Schofield, C. (1998), "An Empirical Investigation into Software Effort Estimation by Analogy", *Bournemouth University*, pp. 1 - 133.

Shepperd, M. and Macdonell, S. G. (2012), "Evaluating Prediction Systems in Software Project Estimation", *Information and Software Technology*, Vol. 54, No. 8, pp. 820 - 827.

Shepperd, M. and Schofield, C. (1997), "Estimating Software Project Effort using Analogies", *IEEE Transactions on Software Engineering*, Vol. 23, No. 11, pp. 736 - 743.

Siami-Irdemoosa, E. and Dindarloo, S. R. (2015), "Prediction of Fuel Consumption of Mining Dump Trucks: A Neural Networks Approach", *Applied Energy*, Elsevier, Vol. 151, pp. 77 - 84.

Sigweni, B. B. (2016), "An Investigation of Feature Weighting Algorithms and Validation Techniques Using Blind Analysis for Analogy-Based Estimation", *Brunel University* London, pp. 1 - 139.

Sigweni, B., Mangwala, M. and Ayoung, D. A. (2017), "Factor Subset Selection for Predicting Sustainability of a Telecentre Using Case-Based Reasoning", *2017 IEEE AFRICON*, IEEE, Capetown, South Africa, pp. 536 - 541.

Sigweni, B. and Shepperd, M. (2015), "Using Blind Analysis for Software Engineering Experiments", *Proceedings of the 19th International Conference on Evaluation and Assessment in Software Engineering - EASE 2015*, ACM Press, Nanjing, China, pp. 1 - 6.

Sigweni, B., Shepperd, M. and Turchi, T. (2016), "Realistic Assessment of Software Effort Estimation Models", *Proceedings of the 20th International Conference on Evaluation and Assessment in Software Engineering - EASE'16*, ACM Press, New York, USA, pp. 1 - 6.

Simonsen, A. M. (2012), *Fuel Consumption in Heavy-Duty Vehicles*, No. 10/2012, Norway, pp. 1 - 9.

de Siqueira Santos, S., Takahashi, D. Y., Nakata, A. and Fujita, A. (2014), "A Comparative Study of Statistical Methods Used to Identify Dependencies Between Gene Expression Signals", *Briefings in Bioinformatics*, Vol. 15, No. 6, pp. 906 - 918.

Soni, D. (2018), "Supervised vs. Unsupervised Learning", www. towardsdatascience.com/supervised-vs-unsupervised-learning. (Accessed: November 6, 2018).

Soofastaei, A. (2016), "Development of an Advanced Data Analytics Model to Improve

the Energy Efficiency of Haul Trucks in Surface Mines", *The University of Queensland*, pp. 1 - 163.

Soofastaei, A. (2018), "The Application of Artificial Intelligence to Reduce Greenhouse Gas Emissions in the Mining Industry", *Green and Emerging Technologies to Help the Environment*, pp. 1 - 18.

Soofastaei, A., Aminossadati, S., Kizil, M. and Knights, P. (2016a), "Reducing Fuel Consumption of Haul Trucks in Surface Mines Using Artificial Intelligence Models", *Coal Operators' Conference, Mining Engineering*, University of Wollongong, Wollongong, pp. 477 - 489.

Soofastaei, A., Aminossadati, S. M., Kizil, M. S. and Knights, P. (2016b), "The Effect of Average Truck Speed on Fuel Consumption in Surface Mining", *Mining Journal*, Vol. 4, No. 2, pp. 92 - 94.

Soofastaei, A., Karimpour, E., Knights, P. and Kizil, M. (2017), "Energy-Efficient Loading and Hauling Operations.", in K. Awuah-Offei (ed.), *Energy Efficiency in the Minerals Industry, Green Energy and Technology*, Springer International Publishing AG 2018, pp. 121 - 146.

Sullivan, T. W. (1990), "New Technology and Economies of Scale in Shovel-Truck Sizing", *Second International Symposium on Mine Planning and Equipment Selection*, Calgary, Canada, A. A. Balkema, Rotterdam, (R. K. Singhal and M. Vavra, eds.), pp. 3 - 6.

Sutha, K. and Tamilselvi, J. (2015), "A Review of Feature Selection Algorithms for Data Mining Techniques", *International Journal of Computer Science and Engineering*, Vol. 7, No. 6, pp. 63 - 67.

Sztrik, J. (2012), *Basic Queueing Theory*, (J. Bíró and Z. Heszberger, eds.), Vol. 193, *University of Debrecen, Faculty of Informatics*, Debrecen, Hungary, pp. 1 - 193.

Smith, G. (2012), "Factors Affecting Truck Fuel Economy", *The Goodyear Tire and Rubber Company*, pp. 63 - 79.

Thompson, G. (1979), "Fuel Economy Effects of Tire, No. SDSB 70-13", *National Service Centre for Environmental Publications (NSCEP)*, pp. 1 - 11.

Thompson, R. and Visser, A. (2003), "Mine Haul Road Maintenance Management Systems", *Journal of the South African Institute,* Vol. 103, No. 5, pp. 303 - 312.

Trani, M. L., Bossi, B., Gangolells, M. and Casals, M. (2016), "Predicting Fuel Energy Consumption during Earthworks", *Journal of Cleaner Production*, Vol. 112, No. 5, pp. 3798 - 3809.

Wang, G., Song, Q., Sun, H., Zhang, X., Xu, B. and Zhou, Y. (2013), "A Feature Subset Selection Algorithm Automatic Recommendation Method", *Journal of Artificial Intelligence Research*, Vol. 47, pp. 1 - 34.

Wells, D. (2018), "Partner Interfaces - Wenco Mining Systems", www.wencomine.com/blog/mine-dispatch/partner-interfaces. (Accessed: April 8, 2019).

Wells, D. (2019a), "Mining Fleet Wencomine Fleet Management System - Wenco Mining Systems", www.wencomine.com/products/wenco-fleet-management-system. (Accessed: April 8, 2019).

Wells, D. (2019b), "Mining Fleet Dynamic Dispatch - Wenco Mining Systems", www.wencomine.com/products/dynamic-dispatch. (Accessed: April 8, 2019).

Wells, D. (2019c), "PitNav turn-by-turn GPS Navigation for Mine Equipment", www.wencomine.com/products/pitnav-mine-gps-navigation. (Accessed: April 8, 2019).

Wetherelt, A. and van der Wielen, P. K. (2011), "Surface Extraction", in Chapter 10 of *SME Mining Engineering Handbook*, 3rd edition, (P. Darling, sen. ed.), Society for Mining, Metallurgy and Exploration Inc, Englewood, Colorado, pp. 857 - 1031.

Wicaksana, Y., Widodo, N., Kramadibrata, S., Wattimena, R. K., Sigalingging, J. W. and Rizky, M. (2011), "Determining Rolling Resistance Coefficient on Hauling Road Using Dump-truck in Open Pit Coal Mine", *International Symposium on Earth Science and Technology*, pp. 1 - 6.

Wu, J., Cai, Z. and Gao, Z. (2010), "Dynamic K-Nearest-Neighbor with Distance and Attribute Weighted for Classification", *International Conference on Electronics and Information Engineering*, IEEE, Kyoto, Japan, pp. 356 - 360.

Zacharof, N. G., Fontaras, G., Ciuffo, B., Siakmakis, S., Anagnostopoulos, K., Marotta, A., Pavlovic, Jand Codrea, C. (2016), "Review of in Use Factors Affecting the Fuel Consumption and CO_2 Emissions of Passenger Cars", *European Commission*, Luxembourg, pp. 63 - 66.

Zheng, Y. and Kwoh, C. K. (2011), "A Feature Subset Selection Method Based on High-Dimensional Mutual Information", *Entropy, Molecular Diversity Preservation International*, Vol. 13, No. 4, pp. 860 - 901.

Zhang, Y., Wang, H., Yang, Z. and Li, J. (2014), "Relative Accuracy Evaluation", *Plos One Journal*, Vol. 9(8), pp. 1 - 11.

APPENDICES

APPENDIX 1

SAMPLE OF CAT TRUCK #29 DATASET

Case ID	Cycle Time	Net Payload (tonne)	Empty Haul Distance (km)	Travel Time Empty (s)	Empty Stop Time (s)	Max Speed Empty (km/h)	Loading Time (s)	Loaded Haul Distance (km)	Travel Time Loaded (s)	Loaded Stop Time (s)	Max Speed Loaded (km/h)	Dumping Time (s)	Total Distance (km)	Fuel consumed (litres)
T1	01:19	208.5	3.86	565	104	43.2	231	5.48	1041	7	36.9	45	9.34	135.41
T2	01:52	211.3	4.98	577	75	43.2	298	3.7	949	134	26.5	48	8.68	125.84
T3	02:27	210.8	3.61	450	840	43.3	167	3.69	933	116	26.8	59	7.3	105.83
T4	03:09	216.4	3.62	516	506	43.2	171	3.7	920	278	30	58	7.32	106.12
T5	03:50	217.4	5.19	1009	325	41.2	240	5.28	1197	130	36.4	50	10.47	151.79
T6	04:55	212.6	4.32	664	111	40.3	170	4.39	1087	32	28.6	57	8.71	126.28
T7	05:30	214.5	8.9	1364	920	39.8	183	4.44	1018	72	36.3	211	13.34	193.40
T8	06:43	214.8	4.34	674	82	34.8	168	4.18	1045	11	29.4	69	8.52	123.52
T9	07:17	222.6	4.11	655	40	37.7	190	5.43	1358	582	36.6	52	9.54	138.31
T10	08:05	218.1	4.99	658	556	42.8	162	4.42	967	14	36.4	136	9.41	136.43
T11	08:46	217	4.34	511	219	40.4	184	4.41	924	7	36.6	55	8.75	126.86
T12	09:18	205	4.35	543	228	43.2	182	4.4	906	14	37.7	167	8.75	126.86

APPENDIX 1

SAMPLE OF CAT TRUCK #29 DATASET

Trip Number	Actual Fuel Used (Litres)	Predicted Fuel Used (Litres)			Absolute Residuals/ Errors		
		CBR–FSS	CBR	Naïve	CBR–FSS	CBR	Naïve
6.00	15.62	15.62	16.79	18.24	0.00	1.18	2.62
7.00	16.56	16.80	16.56	18.24	0.24	0.00	1.67
11.00	15.14	15.14	17.64	18.24	0.00	2.50	3.09
12.00	15.62	16.09	16.32	18.24	0.47	0.71	2.62
14.00	16.09	16.49	17.26	18.24	0.40	1.17	2.15
16.00	16.56	16.53	15.85	18.24	0.03	0.71	1.67
17.00	16.09	16.53	16.80	18.24	0.44	0.71	2.15
18.00	16.09	15.62	16.33	18.24	0.47	0.24	2.15
28.00	16.09	16.32	15.14	18.24	0.23	0.95	2.15
40.00	15.14	14.91	22.37	18.24	0.24	7.23	3.09
50.00	16.09	15.88	14.67	18.24	0.21	1.42	2.15
54.00	16.09	15.84	33.71	18.24	0.25	17.62	2.15
55.00	15.62	15.85	15.39	18.24	0.23	0.22	2.62
56.00	16.09	16.05	15.37	18.24	0.04	0.71	2.15
58.00	17.51	17.51	28.84	18.24	0.00	11.33	0.73
61.00	15.62	15.90	15.85	18.24	0.28	0.24	2.62
66.00	17.51	17.98	17.03	18.24	0.47	0.47	0.73
68.00	19.40	19.72	16.58	18.24	0.32	2.82	1.16
69.00	15.62	15.91	15.14	18.24	0.29	0.47	2.62
71.00	17.51	17.60	15.85	18.24	0.09	1.66	0.73
80.00	15.62	15.99	16.80	18.24	0.37	1.18	2.62
87.00	14.67	14.43	14.67	18.24	0.24	0.00	3.57
91.00	14.67	14.35	15.14	18.24	0.32	0.47	3.57

APPENDIX 3

RANDOM SAMPLE OF ACTUAL DATASET FOR CAT TRUCK #3 (AFTERNOON SHIFT)

Trip Number	Actual Fuel Used (Litres)	Predicted Fuel Used (Litres)			Absolute Residuals/ Errors		
		CBR–FSS	CBR	Naïve	CBR–FSS	CBR	Naïve
1	14.20	14.20	13.53	22.69	0.00	0.66	8.50
4	14.20	14.20	15.18	22.69	0.00	0.98	8.50
21	15.62	15.56	14.67	22.69	0.05	0.95	7.08
45	14.67	14.67	14.89	22.69	0.00	0.23	8.03
56	14.67	14.58	13.27	22.69	0.09	1.40	8.03
69	14.67	14.62	15.14	22.69	0.05	0.47	8.03
85	19.40	19.49	14.90	22.69	0.09	4.50	3.29
87	16.56	16.56	14.18	22.69	0.00	2.38	6.13
88	16.56	16.56	15.38	22.69	0.00	1.18	6.13
92	16.56	16.59	15.61	22.69	0.03	0.95	6.13
117	16.09	16.20	16.33	22.69	0.11	0.24	6.61
124	17.51	17.66	15.15	22.69	0.15	2.36	5.19
130	15.14	15.14	16.08	22.69	0.00	0.94	7.55
154	16.09	16.21	15.14	22.69	0.12	0.95	6.61
155	18.93	18.93	16.56	22.69	0.00	2.37	3.77
165	14.67	14.67	14.19	22.69	0.00	0.48	8.03
166	16.56	16.45	17.15	22.69	0.12	0.59	6.13
167	16.09	16.05	15.85	22.69	0.04	0.24	6.61
169	15.14	15.14	15.14	22.69	0.00	0.00	7.55
198	16.56	16.61	18.22	22.69	0.04	1.66	6.13
210	15.62	15.53	18.45	22.69	0.09	2.83	7.08

APPENDIX 4

RANDOM SAMPLE OF ACTUAL DATASET FOR CAT TRUCK #3 (NIGHT SHIFT)

Trip Number	Actual Fuel Used (Litres)	Predicted Fuel Used (Litres)			Absolute Residuals/ Errors		
		CBR–FSS	CBR	Naïve	CBR–FSS	CBR	Naïve
3	15.62	15.62	16.79	18.47	0.00	1.18	2.86
4	16.56	16.80	16.56	18.47	0.24	0.00	1.91
8	15.14	15.14	17.64	18.47	0.00	2.50	3.33
13	16.56	16.53	15.85	18.47	0.03	0.71	1.91
25	16.09	16.32	15.14	18.47	0.23	0.95	2.38
47	16.09	15.88	14.67	18.47	0.21	1.42	2.38
73	17.03	16.80	13.98	18.47	0.24	3.05	1.44
92	15.14	15.26	13.03	18.47	0.12	2.11	3.33
108	15.62	15.38	13.25	18.47	0.24	2.37	2.86
132	19.87	19.64	18.93	18.47	0.24	0.95	1.40
137	15.62	15.68	16.33	18.47	0.07	0.71	2.86
144	15.62	15.52	14.43	18.47	0.10	1.18	2.86
146	15.62	15.54	14.22	18.47	0.08	1.40	2.86
147	15.62	15.54	13.50	18.47	0.08	2.12	2.86
179	15.14	15.38	14.43	18.47	0.24	0.71	3.33
181	15.14	15.37	14.43	18.47	0.23	0.71	3.33
189	15.14	15.31	14.15	18.47	0.17	0.99	3.33
190	15.14	15.31	15.94	18.47	0.17	0.79	3.33
192	15.14	15.31	15.14	18.47	0.17	0.00	3.33
195	15.14	15.35	15.85	18.47	0.21	0.71	3.33

APPENDIX 5

RANDOM SAMPLE OF ACTUAL DATASET FOR KOMATSU 830E TRUCK #1

Trip Number	Actual Fuel Consumed (Litres)	Predicted Fuel Used (Litres)			Absolute Residuals/ Errors		
		CBR–FSS	CBR	Naïve	CBR–FSS	CBR	Naïve
18	134.07	134.07	130.26	135.07	0.00	3.81	1.01
24	124.65	124.65	122.48	135.07	0.00	2.18	10.42
26	124.51	124.65	121.57	135.07	0.14	2.94	10.56
28	146.43	146.43	127.28	135.07	0.00	19.16	11.36
30	123.67	123.67	137.18	135.07	0.00	13.51	11.41
31	123.67	123.67	124.09	135.07	0.00	0.42	11.41
47	141.09	141.09	123.05	135.07	0.00	18.05	6.02
48	140.95	140.95	134.51	135.07	0.00	6.44	5.88
49	140.95	140.95	125.18	135.07	0.00	15.78	5.88
52	177.35	177.35	130.81	135.07	0.00	46.54	42.28
53	145.73	145.73	124.41	135.07	0.00	21.32	10.66
55	145.87	145.73	135.27	135.07	0.14	10.60	10.80
86	137.86	137.93	130.00	135.07	0.07	7.86	2.79
87	136.04	136.04	135.23	135.07	0.00	0.81	0.96
95	144.05	144.05	142.06	135.07	0.00	1.99	8.97
100	140.39	140.39	143.07	135.07	0.00	2.68	5.32
119	146.58	146.53	144.37	135.07	0.05	2.20	11.50
122	133.93	134.00	132.12	135.07	0.07	1.80	1.15
138	140.81	140.88	132.74	135.07	0.07	8.07	5.74
149	141.80	141.84	140.58	135.07	0.05	1.22	6.72
152	118.33	118.26	130.08	135.07	0.07	11.76	16.75
156	138.00	138.07	137.93	135.07	0.07	0.07	2.93
157	149.10	149.15	138.14	135.07	0.05	10.97	14.03
160	146.86	146.86	136.86	135.07	0.00	10.00	11.78
162	139.13	139.06	131.84	135.07	0.07	7.29	4.05

181	141.52	141.59	139.20	135.07	0.07	2.31	6.44
182	124.79	124.75	135.67	135.07	0.05	10.88	10.28
203	125.78	125.78	132.05	135.07	0.00	6.28	9.30
206	136.88	136.83	155.14	135.07	0.05	18.26	1.80
220	134.35	134.21	127.81	135.07	0.14	6.53	0.73
222	139.27	139.27	143.91	135.07	0.00	4.64	4.19
223	120.30	120.30	132.42	135.07	0.00	12.12	14.78
238	125.35	125.35	131.45	135.07	0.00	6.10	9.72
239	122.12	122.12	130.09	135.07	0.00	7.97	12.95
240	133.37	133.49	142.32	135.07	0.12	8.96	1.71
241	145.45	145.59	148.91	135.07	0.14	3.46	10.38
242	148.54	148.63	157.20	135.07	0.08	8.65	13.47
243	151.21	151.28	144.40	135.07	0.07	6.81	16.14
267	133.51	133.58	138.56	135.07	0.07	5.06	1.57
291	89.94	89.94	134.60	135.07	0.00	44.66	45.13
298	120.58	120.51	121.09	135.07	0.07	0.52	14.50
300	113.83	113.97	129.15	135.07	0.14	15.31	21.24
302	119.31	119.31	132.82	135.07	0.00	13.51	15.76
303	179.18	179.30	173.95	135.07	0.12	5.23	44.10
304	135.89	135.85	138.48	135.07	0.05	2.58	0.82
329	116.36	116.36	126.04	135.07	0.00	9.68	18.71
336	147.70	147.84	140.77	135.07	0.14	6.93	12.63
384	149.95	149.85	129.64	135.07	0.09	20.31	14.87
387	143.20	143.25	143.98	135.07	0.05	0.78	8.13
392	117.91	117.91	128.63	135.07	0.00	10.72	17.17
395	132.38	132.52	131.11	135.07	0.14	1.27	2.69
398	123.53	123.67	138.82	135.07	0.14	15.29	11.55
420	145.59	145.66	130.99	135.07	0.07	14.60	10.52
428	137.16	137.21	118.97	135.07	0.05	18.19	2.09
431	123.25	123.25	116.17	135.07	0.00	7.08	11.83
469	108.49	108.63	109.59	135.07	0.14	1.10	26.58

470	124.09	124.23	131.95	135.07	0.14	7.86	10.98
486	108.35	108.43	105.30	135.07	0.08	3.05	26.72
493	179.46	179.60	137.61	135.07	0.14	41.85	44.39
500	130.27	130.27	143.64	135.07	0.00	13.37	4.80



APPENDIX 6

RANDOM SAMPLE OF ACTUAL DATASET FOR CAT 793 TRUCK #2

Trip Number	Actual Fuel Used (Litres)	Predicted Fuel Used (Litres)			Absolute Residuals/Errors		
		CBR–FSS	CBR	Naïve	CBR–FSS	CBR	Naïve
5	16.09	15.93	16.11	23.08	0.16	0.02	6.99
11	16.09	16.01	17.27	23.08	0.08	1.18	8.41
16	17.03	17.11	19.40	23.08	0.08	2.37	3.20
25	14.67	14.67	15.11	23.08	0.00	0.45	8.88
68	15.14	15.14	15.14	23.08	0.00	0.00	0.37
87	16.56	16.38	19.38	23.08	0.18	2.82	3.20
101	16.09	16.15	15.61	23.08	0.07	0.47	8.41
112	16.09	16.07	15.85	23.08	0.02	0.24	6.52
117	14.67	14.67	14.91	23.08	0.00	0.24	8.41
119	16.09	16.02	15.82	23.08	0.07	0.27	6.52
125	16.09	16.07	15.38	23.08	0.02	0.71	7.46
128	18.93	19.08	15.84	23.08	0.16	3.09	6.99
135	14.67	14.67	13.96	23.08	0.00	0.71	6.52
159	16.56	16.63	15.83	23.08	0.06	0.73	7.46
163	16.56	16.62	20.11	23.08	0.06	3.55	107.99
168	18.45	18.44	15.39	23.08	0.01	3.06	8.41
202	16.09	15.99	16.56	23.08	0.09	0.47	7.94
207	44.01	44.08	14.67	23.08	0.08	29.34	6.99
211	15.14	15.14	15.14	23.08	0.00	0.00	5.10
227	17.03	17.19	14.67	23.08	0.16	2.36	8.88
230	16.09	15.97	15.59	23.08	0.12	0.49	6.04
237	16.09	16.04	16.09	23.08	0.05	0.00	4.15
246	16.09	16.04	16.09	23.08	0.05	0.00	7.94
247	16.09	16.04	14.65	23.08	0.05	1.43	7.46

248	16.09	16.04	14.43	23.08	0.05	1.66	7.46
261	20.35	20.35	18.23	23.08	0.00	2.12	4.62
274	17.98	18.04	15.61	23.08	0.06	2.37	134.02
278	18.45	18.64	20.58	23.08	0.19	2.13	5.10
281	15.62	15.62	15.38	23.08	0.00	0.23	4.62
295	16.09	16.09	17.74	23.08	0.00	1.66	6.52
297	17.03	17.02	19.87	23.08	0.01	2.84	1.78
304	20.35	20.38	20.82	23.08	0.04	0.47	6.52
305	16.09	16.10	16.80	23.08	0.01	0.71	4.15
308	15.14	15.14	15.33	23.08	0.00	0.19	6.99
332	20.35	20.21	19.39	23.08	0.14	0.96	127.87
339	17.98	17.98	16.56	23.08	0.00	1.42	6.52
346	18.45	18.34	17.51	23.08	0.12	0.95	38.91
354	17.03	17.00	16.09	23.08	0.03	0.95	7.46
372	16.09	16.09	17.98	23.08	0.00	1.89	3.68
388	14.67	14.67	17.73	23.08	0.00	3.06	1.05
392	15.62	15.62	16.56	23.08	0.00	0.95	9.36
394	16.56	16.64	16.32	23.08	0.08	0.24	7.94
396	15.62	15.46	17.98	23.08	0.16	2.37	8.41
409	15.62	15.58	15.61	23.08	0.04	0.00	2.26
411	15.62	15.49	17.97	23.08	0.13	2.36	7.94
432	17.51	17.63	21.53	23.08	0.12	4.02	3.20
463	17.03	16.89	17.03	23.08	0.14	0.00	4.15
465	17.03	16.89	16.32	23.08	0.14	0.71	7.46
483	15.14	15.14	16.09	23.08	0.00	0.95	6.52
485	15.62	15.46	15.62	23.08	0.16	0.00	7.94
495	15.62	15.46	17.74	23.08	0.16	2.13	7.94
508	16.09	16.27	14.67	23.08	0.19	1.42	6.52
519	16.56	16.56	19.11	23.08	0.00	2.55	204.99
523	15.14	15.14	14.44	23.08	0.00	0.70	5.57

539	17.03	17.18	20.08	23.08	0.14	3.05	6.52
551	17.03	17.17	18.69	23.08	0.13	1.66	6.52
561	17.03	17.23	16.32	23.08	0.19	0.71	7.46
573	17.51	17.61	16.09	23.08	0.10	1.42	3.20
574	15.14	15.14	18.07	23.08	0.00	2.93	5.10
594	15.14	15.14	14.92	23.08	0.00	0.23	6.52
603	17.51	17.51	17.03	23.08	0.00	0.47	4.15
618	17.03	17.03	16.32	23.08	0.00	0.71	6.52
629	17.03	17.23	17.98	23.08	0.20	0.95	6.99