CONSTRUCTION EQUIPMENT FUEL CONSUMPTION DURING IDLING

Characterization using multivariate data analysis at Volvo CE

MUJTABA HASSANI

School of Business, Society and Engineering
Course: Master of science thesis work
Course code: ERA 402
Credits: 30 credits
Program: Master of Science degree in industrial engineering and management with focus on energy system

Supervisor: Maher Azaza
Examiner: Jan Skvaril
Customer: Volvo CE
Date: 2020-06-13
Email:
mhi15002@student.mdh.se

ABSTRACT

Human activities have increased the concentration of CO_2 into the atmosphere, thus it has caused global warming. Construction equipment are semi-stationary machines and spend at least 30% of its life time during idling. The majority of the construction equipment is diesel powered and emits toxic emission into the environment. In this work, the idling will be investigated through adopting several statistical regressions models to quantify the fuel consumption of construction equipment during idling. The regression models which are studied in this work: Multivariate Linear Regression (ML-R), Support Vector Machine Regression (SVM-R), Gaussian Process regression (GP-R), Artificial Neural Network (ANN), Partial Least Square Regression (PLS-R) and Principal Components Regression (PC-R). Findings show that pre-processing has a significant impact on the goodness of the prediction of the explanatory data analysis in this field. Moreover, through mean centering and application of the max-min scaling feature, the accuracy of models increased remarkably. ANN and GP-R had the highest accuracy (99%), PLS-R was the third accurate model (98% accuracy), ML-R was the fourth-best model (97% accuracy), SVM-R was the fifth-best (73% accuracy) and the lowest accuracy was recorded for PC-R (83% accuracy). The second part of this project estimated the CO_2 emission based on the fuel used and by adopting the NONROAD2008 model.

Keywords: Idling condition, environmental effect, diesel fuel, machine learning, multivariate data analysis, partial least square regression, support vector machine regression, principal component analysis, principal component regression, correlation coefficient matrix, artificial neural network, exhaust emission reduction techniques, global warming, emission regulation, CO_2 estimation techniques, Gaussian process regression

PREFACE

"If the teacher is wise, he does not bide you to enter the house of his wisdom. But leads you to the threshold of your own mind"

K. Gibran

To my beautiful daughter Zara

This thesis work is performed for a degree of Master of Science at Mälardalen University in city of Västerås, Sweden. The project was undertaken on behalf of Volvo Construction Equipment product support and manufacturer of construction equipment, at city of Eskilstuna.

I wish to express my sincere gratitude and appreciation to my supervisor Mr. Maher Azaza, associate senior lecturer and my examiner Mr. Jan Skvaril, associate senior lecturer at Mälardalen University who lightened my way through the work with their precious guidance and knowledge.

I would like to appreciate Uptime Center at Volvo Construction Equipment Eskilstuna, especially my supervisor Mr. Robert Sundkvist, head of Uptime Center for his support and instructions. I will also thank Mr. Leif Waad, product specialist at Uptime Center for his kind support and cooperation.

Västerås, June 2020

Mujtaba Hassani

CONTENT

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

Abbreviation	Description
AMR	Allied Market Research
ANN	Artificial Neural Network
AQG	Air Quality Guidelines
CE	Construction Equipment
CFR	Code of Federal Regulations
CH_4	Methane
CO_2	Carbon Dioxide
DOC	Diesel Oxidation Control
DPF	Diesel Particulate Filter
EEA	European Environment Agency
EEA	European Environment Agency

EMEA	Europe, Middle East and Africa
EMEA	Middle East, and Africa
EPA	Environmental Protection Agency
GP-R	Gaussian Process Regression
GP-RS	General Packet Radio Service
GP-RS	General Packet Radio Service
GWP	Greenhouse Gas Potential
HC	Hydrocarbon
ICCT	International Council on Clean Transportation
LSTMN	long short-term memory network
ML-R	Multivariate Linear Regression
MVDA	Multivariate Data Analysis
N_2O	Nitrous Oxide
NO_2	Nitrogen Dioxide
NRMM	Non-Road Mobile Machinery
NSOF	Non-Soluble Organic Fraction
NSR	NO_x Storage Reduction
PC	Principal Component
PCA	Principal Component Analysis
PC-R	Principal Component Regression
PDS	Pre-Diagnostic System
PDS	Pre-Diagnostic System
PEMS	Portable Emissions Measurement Systems
PLS-R	Partial Least Square Regression
PM	Particulate Matter
PPB	Parts Per Billion
PPMV	Parts Per Million Per Volume
RPM	Revolutions Per Minute
SAT	Satellites
SCR	Selective Catalytic Reduction
SOF	Soluble Organic Fraction
SVM-R	Support Vector Machine Regression
US EPA	United States Environmental Protection Agency
UV	Ultraviolet
WHO	World Health Organization

DEFINITIONS

Definition	Description
Equipment age	The number of year that the equipment has purchased
Engine tiers	Emission standard, manufacture of diesel engine must meet the performance level after a specified date
PEMS system	Measurement system that identifies the exhaust gas character
NONROAD2008	Exhaust gas estimation method developed by International Plant Protection Convention
Non-road/off-road	Refers to engines used in other than a motor vehicle
40 Code of Federal Regulations (CFR)	United states exhaust emission standards authority

1 INTRODUCTION

Air pollution causes lots of damage to the human body, plants, lakes, and animals (EPA, 2017). World Health Organization (WHO) estimates that this phenomenon causes about 4.3 million deaths each year, leaving most people suffer from stroke and heart-condition (World Health Organization, 2020). Saxena, et al. (2019), explains that air pollution sources can be classified into four categories: energy sector, industrial sector, agriculture and waste. Energy sector generates the largest proportion of air pollution emissions. This sector contains the transport sector, which further consists of sub-categories such as on-road and off-road vehicles.

According to Environmental Protection Agency (2017), on-road refers to vehicles used for transport of passengers or goods such as heavy-duty, light-duty and trucks. The term off-roads refers to vehicles and engines used in construction, agriculture and other purposes such as aircraft, construction equipment and vessel. Masters (2011), describes that construction equipment is used for various purposes such as construction industry, building roads, building bridges, etc. According to The Constructor (2020), there is a wide span of different types of construction equipment in different sizes. Nevertheless, the size and type of equipment depends on the application areas and the size of the projects. Fu (2013), describes that wheel loaders and excavators are some common construction equipment. Fan (2017), and Matthews, Rutty, & Andrey (2017), state that the most of this construction equipment are powered by diesel. Toth (2016) and Sajjad, et al. (2013), states that construction equipment averagely consumes 38 liters of diesel per hour and emits 87.400 g/h carbon dioxide (CO_2) into environment. Dybdahl, et al. (2004) states that particulate matter (PM) is a product of the combustions process of the diesel engines which considers toxic and carcinogenic. Sajjad et al. (2013); and Frey, Rasdorf, & Lewis, 2010; Heidari & Marr, (2015), observes that most of construction equipment (CE) are considered as stationary machines and spend a significant time during the idle-condition under its daily operation. There are several studies that believe the idle-condition of the engine as an issue that needs to be minimized. Sajjad et al. (2013) points out that the idle-condition of an engine has a negative impact on the environment, economy, and air quality. Furthermore, Brodrick, Dwyer, Farshchi, Harris, & King JR (2011) states that idle condition of the engines has negative impact on the operator as well as the engine its self. However, Sajjad et al. (2013), spots that idle duration of the engine is a function of ambient temperature. However, there is a significant gap in idle condition of the engine in cold versus warm ambient temperature. According to IEA (2005), geographical locations with cold ambient temperature refer to places where the ambient temperature is equal to or lower than 10 degrees Celsius during six months of a year. Warm ambient temperature refers to places where the ambient temperature is 25 degrees Celsius or higher during six months of a year.

Perozzi, Mattetti, Molari, & Sereni (2016), explains that there is a lack of data in the literature that monitors the idling condition of the CE. Furthermore, it is a difficult task to measure the duration of idle-condition of the construction equipment. Manley & Peters (2012), confirms the complexity of fuel consumption and exhaust emissions of construction equipment during idling. It proposes more studies needs to be done in this field in order to increase the knowledge about the factors that affect the fuel consumption and CO_2 emission. Fan (2017), suggests to measure the fuel consumption of the machinery of a construction site by weekly data and Perozzi, Mattetti, Molari, & Sereni (2016), proposes to maintain specific sensors to measure the duration of idling and its fuel consumption. Volvo CE has over 88.541 machines worldwide and owns 9.8% of the global market (Confidential, 2019). The machinery is equipped with several specific sensors and telematics system that measure and monitor the results (Volvo CE, 2020). Fan (2017), suggests to consider the CE machines as a black box and develop statistical models to explore the factors that impacts the fuel consumption of the machinery. By this, valuable information about machinery itself will be obtained and it will also increase the knowledge about CE machines.

1.1 Background

This section presents the theoretical background of this thesis work and the section will introduce general information about the back-bone of this study. Nevertheless, the section ends up with the presentation of the aim and delimitation of the study.

1.1.1 Global warming potential and environment

According to Erin (2019), the essence behind the rise in the global average temperature is the increase in consumption of fossil fuels. During the burning process of fossil fuels, greenhouse gases such as CO_2 are released into the atmosphere. Erin (2019) and Buchanan (2016) explain that demand of energy from fossil fuel such as coal, oil and gas has been increased during recent years. Buchanan (2016), explain that the energy sector is the main source of a large proportion of greenhouse gases and CO_2 emissions. Notwithstanding, a quarter of global emissions emanate from the transport sector. This has led to an increased greenhouse effect, which in turn has led to global warming. CO_2 is a greenhouse gas that allows short-wavelength solar rays to pass through the atmosphere but prevents long-wavelength radiation from leaving the Earth's atmosphere, this is called the "greenhouse effect". Buchanan (2016) mention that the greenhouse effect is necessary and without this phenomenon, the average soil surface temperature would be -18 degrees Celsius. United States Environmental Protection Agency (2020) states that natural greenhouse gases are comprised of other gases too. This mixture consists of; 81% CO_2, 10% CH_4, 7% N_2O and 3% fluorinated gases. However, recently the concentration of CO_2 from 400 parts per million per volume (PPMV) has increased substantially by 280 PPMV during the industrial period. Another greenhouse gas that has increased remarkably is the amount of CH_4 in the atmosphere. Currently, the CH_4 concentration is 2000 ppb (parts per billion), which has increased by 300% from the period before the industrialization (580-783 ppb). Schuerger, Moores, Clausen, Nadine, & Britt (2012), explains that CH_4 formation changes in the atmosphere, which affects its impact and durability. Based on this fact, it is difficult to estimate the lifetime of methane in CO_2 equivalence. Nevertheless, Erin (2019), claims that CH_4 has a lifetime of around 10-15 year in the atmosphere. Schuerger et al (2012), explains that CH_4 can be destroyed in interaction with ultraviolet (UV) rays.

Wei, Zhu, Shu, Tan, & Wang (2012), explain that to analyze the significance different gases in atmosphere, it is necessary to measure every gas by the same unit which is greenhouse gas potential (GWP). Buchanan (2016), indicates that GWP describes the ability of a greenhouse gas to contribute to the greenhouse effect and global warming. However, Erin (2019) observes that the European Union (EU) has the vision to reduce the greenhouse gases by 20% until 2020 by having 1990 as a reference year, through switching to renewable energy sources and increasing energy efficiency. Table 1 show the GWP of some common gases in the atmosphere.

Table 1 Global warming potential. Source: (Naturvårdsverket, 2019)

Greenhouse gas	GWP
CO_2	1
CH_4	25
NO_2	298

1.1.2 EPA and EU non-road emissions regulations

The European Commission has formed and established a series of emission standards for non-road diesel engines in recent decades. European Commission (2014), states that the term non-road mobile machinery (NRMM) refers to "small gardening and handheld equipment (lawnmowers, chain saws), construction machinery (excavators, loaders, bulldozers,) agricultural & farming machinery (harvesters, cultivators,); even railcars, locomotives and inland waterway vessels". Table 2 presents the regulation process concerning its adoption date and directive. The European Commission (2018), states that these regulations are intended to protect the health of EU citizens, improve air quality and protect the environment. The international council on clean transportation (2016), indicates that every EU-state regulatory authority most apply those directives both on internal and external markets for new machines. The European Commission (2014), describes that each directive consists of two parts, the first part describes how and when each directive should be applied and the second part focuses on the development and implementation of technical specifications. Furthermore, it also introduces the circumstances for the forthcoming directives. The European Commission (2018), states that regulations impose emission restrictions on non-road machines for different engine-power (see Table 2). Non-road manufacturers are forced to meet these requirements for new machines to enter the European market. Table 2 is adopted based on the information from the international council on clean transportation (2016).

Table 2 EU Exhaust emission regulations, *source: Directives on emission from non-road mobile machinery*

Directive	Adoption date	Progress
97/68/EC	December 1997	Implementation of Stage I & II emission standards regarding exhaust emission discrimination for diesel engines with a horsepower span of 37-560 kW
2002/88/EC	December 2002	Improvement and extend of the

2014/26/EC	April 2004	Manifestation of Stage IIIA, IIIB and IV emission standards. With contrast that regulations expand the range of the horsepower to 19-560 kW
2006/105/EC	November 2006	Modification of previous directive regarding concerns on the approval standard numbering system
2010/26/EC	March 2010	Modification of directive regarding to IIIB and IV emissions standards
2011/88/EU	November 2011	Modification of directive regarding IIIB engines
2012/46/EU	December 2012	Adjustment of previous directive and additional of reflection on technical process on emission measurement.

The international council on clean transportation (2016) states that each directive is stricter compared to the previous version, and also includes a wider range of equipment. Figure 1 presents the impact of emissions standards on HC, NO_x and PM emissions. The vertical axis presents the substances in grams per kilowatt-hour and the horizontal axis presents the European emission standards. The figure is adopted based on the information from the international council on clean transportation (2016).

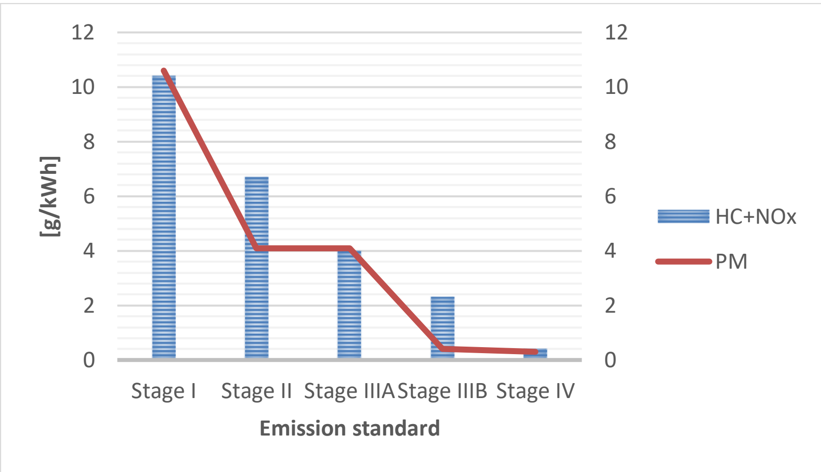


Figure 1 Emission standard's impact on substantial reductions of HC, NOx & PM for diesel engines. source (ICCT, 2016)

The International council on clean transportation (2016), states that each directive is stricter and imposes tougher limits on emissions of PM, HC and NO_x. The European Commission has upgraded the Stage IV standard and proposed Stage V. However, Stage V targets a broader range of diesel engine power (19 <kW <560) and limits PM and NOx + HC emissions by 97 and 94%, respectively.

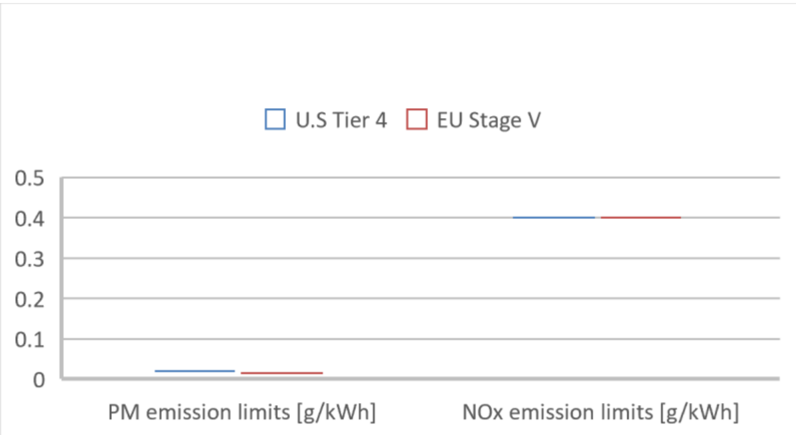
Hooftman, Messagie, Mierlo, & Coosemans (2018), observe that the European Environment Agency (EEA) annually publishes a report that measures the air quality level of members. However, the report considers the WHO Air Quality Guidelines (AQG) to ensure the health of European citizen. The report also considers the ambient air quality directives and examines the progress of the measures. Further, the authors analyze the results of the reports in 2000 to 2014 and find that thanks to the strenuous European Ambient Air Quality Directives, the air quality is acceptable and the directives have had a positive effect on air quality.

The environmental protection agency is the United States regulatory authority that controls and establishes inventory and regulations against techniques and operations with the intention of securing air quality standards (EPA, 2019). The environmental protection agency (2016), notes that in the United States, regulations governing the restriction of substances such as NO_x, CO, HC and NMHC for off-road diesel engines are handled in accordance with the procedures of 40 Code of Federal Regulations (CFR). Nevertheless, PM restrictions are enforced with regard to California Regulations. Table 3 describes the various regulations and their adoption time. Table 3 is developed based on the information from environmental protection agency (2016).

Table 3 EPA exhaust emission regulation

Standard	Adoption date	Progress
Tier 1	1998-2000	Limitation of substantial reductions of NOx & PM for diesel engines regarding to engines power in range of 8-560 kW.
Tier 2	2001-2006	Limitation of substantial reduction of NOx & PM for diesel engines regarding to engines power in range of 8-560 kW
Tier 3	2006-2008	Limitation of substantial reduction of NOx & PM for diesel engines regarding to engines power in range of 37-560 kW
Tier 4	2008-2015	Delamination of substantial reductions of NOx & PM for diesel engines regarding to engines power in range of 56-560 kW

Milieu Ltd (2004), describes that it's difficult to compare the American regulatory standards process with the European one, as each region limits different substances regarding to different geographical conditions. However, the international council on clean transportation (2016) states that the latest European regulation standard (Stage V) for new machines places a higher limit on the emission of PM compared to the American emission standard (Tier 4). Figure 2 shows that for NO_x emissions both make the same requirements but for the emissions of PM stage V is roughly strict.



Source ICCT 2016

Figure 2 Comparison of US Tier 4 standard versus EU Stage V

Naturvårdsverket (2018), states that since 2008 European Commission has regulated CO_2 emissions for manufactures of new passenger cars and light trucks. The EU target imposes environmental requirements on the manufacturer of the maintained vehicle type and that CO_2 emissions must not exceed 95- and 147-grams CO_2 per kilometer for passenger cars and light trucks, respectively. Naturvårdsverket (2018), asserts that in 2014 the European Commission took the initiative and introduced strategic work that through successive steps, increases the knowledge regarding CO_2 emissions from buses and heavy trucks. However, due to a lack of information on fuel consumption and CO_2 emissions of this on-road transport type, no measurements or reports were presented.

Naturvårdsverket (2018), declares that in 2017 the European Commission introduced a simulation program (Vehicle Energy Consumption Calculation Tool, VECT) to calculate new heavy vehicle fuel consumption and CO_2 emissions. Nonetheless, EU has not taken the CO_2 emission from construction equipment or construction sites into account. Moreover, Fan (2017), states that the life-time of construction equipment is long and regulations in CE-industry impose requirements for new machines without taking the CO_2 emission into account, therefore several actions are needed in this area to reduce the environmental cost.

1.1.3 Construction equipment industry

Allied market research (2020) and Market Report (2020), mentions that construction equipment is also known as heavy equipment. Market Report (2020), mentions that earlier, the CE industry faced an economic slowdown due to the world's economic condition. Nevertheless, there is some uncertainty regarding forecasting the market size of the construction equipment industry from a global perspective. This might be since the manufactures are not willing to share such an information. Although, according to Allied Market Research (AMR), the market size of the construction equipment industry is expected to be $288.8 billion by 2020. However, the analyses depict the market has grown by 9.2% during 2016-2020. Though, Market Report (2020), is expecting that the market has the potential to grow steadily by 4.5% until 2025. The market growth is suspected to be different in different regions, China is expected to have the highest growth by 7.2%, the market size in the US and EU is supposed to grow by 3.6%, the author's prediction model doesn't state in percentage how the CE industry will change in emerging countries such as Asia-Pacific, Latin America, and the Middle East. Though, Market Report (2020), states that they have the potential to grow and shape the market. Following the peak growth in EU construction output of 4.1% in 2017, the rate of growth slowed down to 2.8% in 2018. This trend has continued in the beginning of 2019 and was forecasted to have a growth rest of the year.

According to the Market Report (2020) and Allied Market Research (2020) the following list is the key players of the CE market:

- Volvo AB
- Caterpillar Inc.
- Komatsu Ltd
- Doosan Heavy industries & Construction Co. Ltd.
- Hitachi Construction Machinery Co. Ltd.
- J.C. Bamford Excavators limited
- Kobe Steel Ltd
- Liebherr Group
- Atlas Copco AB

- CNH Industrial N.V.

The client of this thesis work is Volvo CE Eskilstuna. Volvo CE is a part of the Volvo Group and environmental care is its core values (Swecon, 2020). Volvo CE's intention is to reduce the environmental impacts of its products and demands developing an algorithm to identify the fuel consumption and CO_2 emissions with a focus on idle-condition. Volvo CE launched headquarters of Uptime Center in Eskilstuna, Sweden on the 17th of December 2018. Uptime Center's intention is to support the entire Volvo dealer network in Europe, the Middle East, and Africa (EMEA) region to keep customers' machines running which happens to be Volvo CE vision (Bast, 2019).

Uptime Center is a unit that manages the input data coming from the construction equipment. It monitors how the customer treats machines and how the equipment performs from a time and space point of view. Furthermore, it is worthy to mention that Uptime Center doesn't have knowledge about the operator identity. Generally, the aftermarket product's support can take place through two different approaches: Proactive and Reactive product support.

Reactive support is the traditional customer support manner, where the end customer comes in contact with a Volvo dealer and describes the problem. Based on the problem description, the dealer coordinates how to tackle the issue to keep the machine uptime (rolling). Proactive product support defines as a unit (Uptime Center in this case), based on the machine's performance information, predicts the possible errors that may occur for the machine and thereby inform the end customer through the Volvo dealer. The reactive approach predicts the problem before it takes place and brings time reduction in entire product support process thus it identifies which component is needed and which kinds of mechanics are most sufficient to tackle the issue.

Uptime Center deals continuously with Active care which consists of two parts: monitoring and weekly report. The monitoring process goes through the CareTrack, which is the terminology that Volvo CE uses for its telematics system. The CareTrack information is upgraded every five minutes automatically. Figure 3 shows how the information flow of Volvo CE's telematics system.

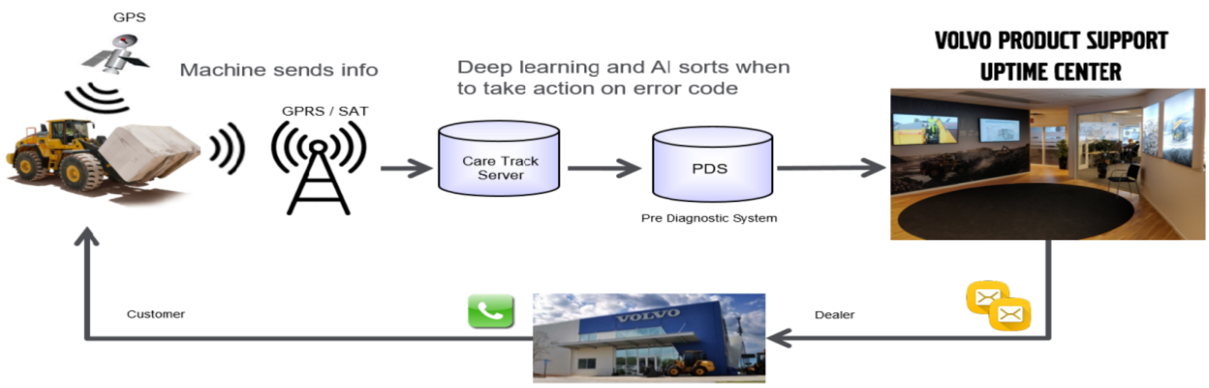


Figure 3 Volvo CE telematics system, Source: (Volvo Construction Equipment, 2019)

The machine sends information through General Packet Radio Service (GP-RS) or Satellites (SAT). All the information saves into CareTrack server, which is the system's data warehouse. The data warehouse is getting analyzed by the Pre-Diagnostic System (PDS), which is based on advanced Deep Learning techniques where different variables are trained to identify pattern recognition of the data warehouse. However, PDS send the results to Volvo product support (Uptime Center). However, the received information will be analyzed further by product specialists and identifies the error codes regarding the performance of the machine to obtain an overview of the machine. Furthermore, product specialists analyze the behavior of the operator machines as well. The behavior of operator machines measured based on engine speed and engine status. Further, the results of this procedure compiles into a report and sent to the closest dealer to the customer in the EMEA region. The Volvo dealer checks the report and informs the customer.

Moreover, a weekly report about the performance of the machinery is conducted and sent to the fleet manager in order to present an overall assessment. The weekly report contains overall information about the machine status and also proposal parts that may improve fuel consumption and performance of the machine.

1.1.4 Exhaust emissions from construction equipment

Kubsh (2017), mentions that agricultural and construction equipment are the key sources of air pollution in many countries. Menon & Dallmann (2016), claims that construction equipment is a key source of air pollution in the regions of United States and European Union. Further, according to Kubsh (2017), the equipment is mostly diesel-powered and emits harmful particles into the environment during the combustion process. Menon & Dallmann (2016), states that NO_x and PM are among the important pollutants that has raised lots of concern. Figure 4 displays the NO_x contribution from different sources in regions United States and the EU. On-road vehicles such as light and heavy machines are still the dominant source of air pollution compared to other sources. The figure is created based on the information which is presented in Kubsh (2017) report.

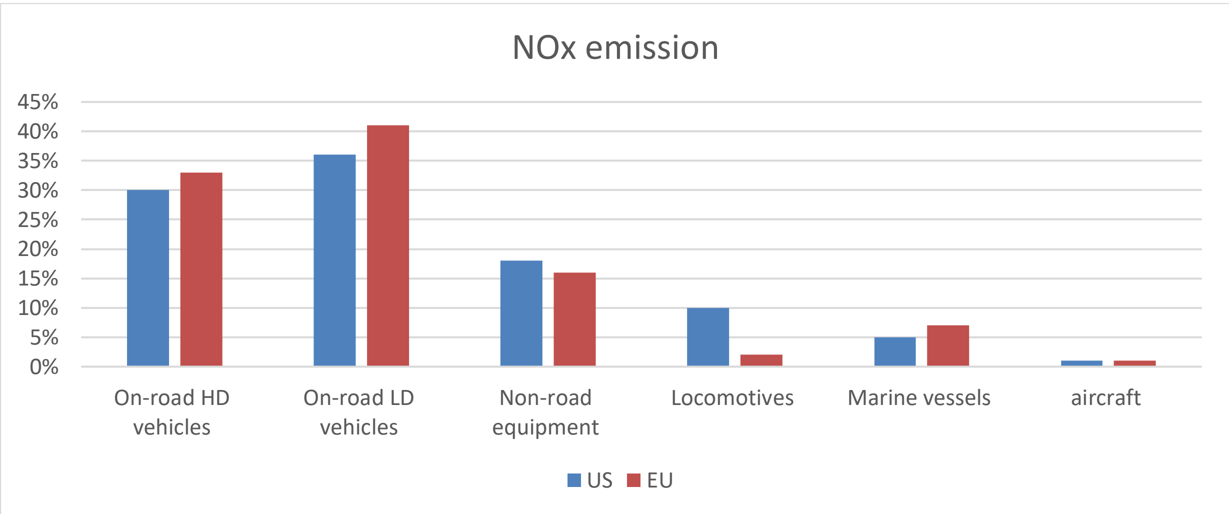


Figure 4 NOx emission per source in region US and EU. Source: Kubsh (2017)

Figure 5 displays the PM contribution from different sources in regions United States and EU. On-road vehicles such as light and heavy machines are still the dominant source of air pollution compared to other sources. The figure is created based on the information which is presented in Kubsh (2017) report.

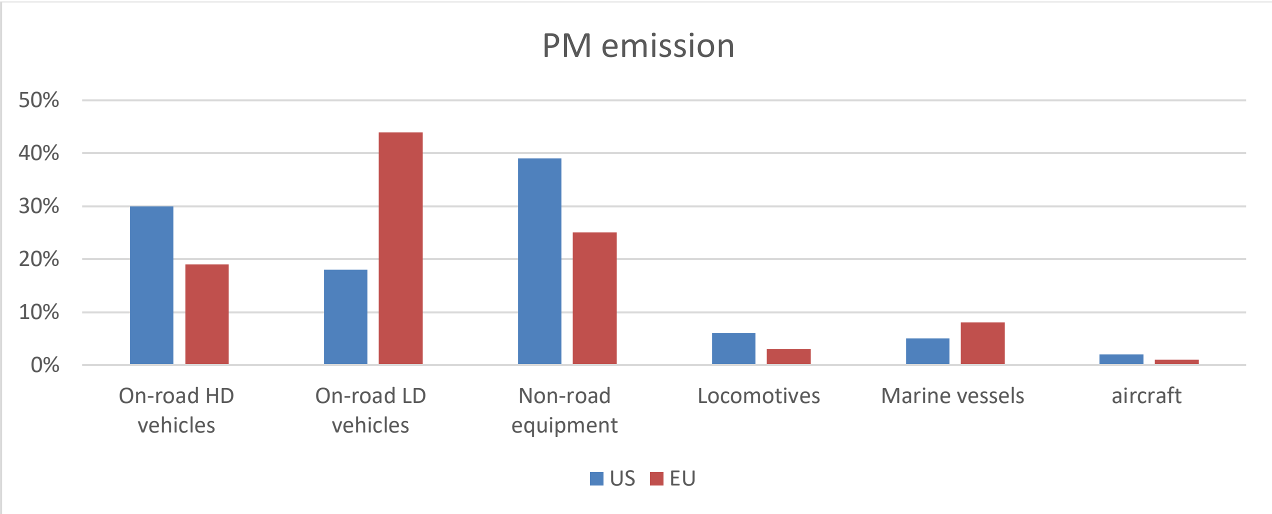


Figure 5 PM emission per source in region US and EU. Source: Kubsh (2017)

Naturvårdsverket (2018), states that there are some uncertainties regarding the estimation of emissions from CE machines in Sweden. This may partially depend on lack of fuel consumption reports in these sectors and partially because of the estimation tools itself. Moreover, the estimation devices based on stationary diesel engines, consequently needs to upgrade in order to present a more realistic profile about characterizes and the overall environmental impact of exhaust emission. The Swedish environmental protection agency monitors exhaust emissions from different sectors such as transport and industry etc. (Jonsson & Bondemark, 2017). Naturvårdsverket (2018), believes that a large proportion of construction equipment are diesel-powered combustion engines. Hence, it emits greenhouse gases into the atmosphere during its operation. Adriansson (2019), states that the largest application area of construction equipment in Sweden is in the industry- and construction sectors. However, CE machines contribute to 40% of the total greenhouse gas emissions from those sectors. According to Jonsson & Bondemark (2017), emissions from CE machines have increased by 25% compared to 1990. However, the increased amount of emission corresponds to a quarter of a million tons of CO_2 equivalent. Naturvårdsverket (2018), observes that during 2016 CE machines emitted 3.5 million tons of greenhouse gases which corresponds to 20% of total emissions from the transport sector. Adriansson (2019), analyzes the characterization greenhouse gas emission from CE machines and concludes that CO_2 comprises the largest part, but although other substances exist too.

According to Naturvårdsverket (2018), among CE machines, tractors emit the biggest amount of CO_2 in Sweden. Further, wheel loaders, excavators, snowmobiles/quad bikes, mining/Tipp trucks and riding mowers are other types of CE machines that contribute to 70% of total CO_2 in Sweden.

Jonsson & Bondemark (2017), mention that since 1990 Sweden's territorial emissions have decreased by 25 %. However, during the same period emissions from CE machines have moved out in the opposite direction and increased by 13%. Further, Naturvårdsverket (2018), observes that the Swedish transport administration in accordance with EU emission regulation has introduced Steg regulations69. Furthermore, this requirement means that CE machines that operate in the urbanized areas such as Malmo, Stockholm, and Gothenburg should not be older than 12 years (since the machines have purchased) and the machines should meet Step IV or later emission standards.

1.1.5 Previous studies in this field

Abolhasani, et al. (2012); Frey, Rasdorf, & Lewis (2010) and Heidari & Marr (2015), analyzed the accuracy of NONROAD2008 estimation method by using portable emissions measurement system (PEMS). PEMS is a device that is used to measure the exhaust emissions in real-time and NONROAD2008 is an estimation technique that is been developed by EPA. However, the authors explain that NONROAD2008 is based on a steady-state engine dynamometer test. It uses an uninstalled stationary engine in a laboratory environment and does not represent the exhaust emissions from a real-world perspective. Therefore, this area requires more studies to quantify the fuel consumption and exhaust emission rates of non-road equipment.

In an explanatory data analysis performed by Frey, Rasdorf, & Lewis (2010), explores the strength of predict variables such as; engine tire, engine size, engine model year and engine load with respect to fuel consumption and exhaust emission profile as respond variables. The authors put the generation of gases such as; NO_x, HC, CO and PM into focus. Further, the exhaust emissions are measured by applying PEMS system to evaluate the accuracy of the NONROAD2008 estimation model.

Abolhasani, et al. (2012), measures the emission profile of the CE in a similar manner as Frey, Rasdorf, & Lewis (2010). What is distinguishable is that Abolhasani, et.al (2012), focuses only on excavators and analyzes the impact of the duty cycle on fuel consumption versus emissions profile of the construction equipment. Heidari & Marr (2015), emphasizes the importance of quantification of the exhaust emission from construction equipment. Therefore, the authors validate and compare the NONROAD2008 and OFFROAD2011 methods with each other. The results explicated that there is a large gap in the emission level between estimation methods (NONROAD2008 & OFFROAD2011) versus real-world. However, the result of the experiment also showed that the gap was more substantial especially regarding equipment that meets higher Tire (Tire 3 & Tire 4) emission standards (Abolhasani, et. al., 2012; Frey, Rasdorf, & Lewis, 2010; Heidari & Marr, 2015). Abolhasani, et al. (2012); and Frey, Rasdorf, & Lewis, (2010); and Heidari & Marr, (2015), propose that the exhaust emission estimation models needs to be upgraded, specifically for the equipment that has a higher emission standard as they release a lower amount of emissions to the atmosphere.

Furthermore, some authors emphasize the importance of the idle-condition from a fuel economy and exhaust emission perspective. Perozzi, et al. (2016), observe that construction equipment is a stationary machine and it is a difficult task to measure the idle-condition of construction equipment in contrast with the on-road vehicle. Hence, in the case of the on-road vehicle, the idling duration can be determined by considering the gear position (if the gear is in the neutral mode, this means that the machine is idling). In the case of off-road vehicles, a complex study is required to develop an approach to estimate the idle duration. For this purpose, an embedded sensor system is needed to fulfill and tackle this issue.

Matthews, et al. (2017), analyzes the idling duration with regards to the ambient temperature, operator behavior and fuel consumption during both winter and summertime. The results indicate that the idling occurs mostly because of the operator's personal attitude. Further, the duration of idling is seasonal, and its portion is higher during the wintertime. Nevertheless, the authors acknowledged the inaccuracy of the data, partly because of poor maintenance of the device that measures the machine behavior and partly because of the delimitation of the device's memory capacity.

Brodrick, et al. (2011) and Fan (2017), studied the strength of different factors during idling such as engine load, engine speed and production year of the engine and their impacts on fuel consumption as well as exhaust emission of the equipment. Fan (2017), only analyzed 14-wheel loader from different manufacturers operating in different job sites. The data collection was excuted manullay on a weekly basis by installation of specific sensors in the gear-box. By this manner, the number and duration of idlings were recorded for every wheel loader. Thus, the authors acknowledge about doubts that may rise about the accuracy of the data. Therefore, they suggest a similar study should be done using data from Telematic systems. Brodrick, Dwyer, Farshchi, Harris, & King Jr (2011) perform a similar study as Fan (2017), but focusing on excavators instead.

Until this point, few studies have been executed on one single equipment model that simultaneously focuses on CO_2 emission and enjoy data from telematics systems. Fan (2017), Perozzi, et al. (2016); Brodrick, Dwyer, et al. (2011); and Frey, Rasdorf, & Lewis (2010) are united that there is very little real-world data available that presents the idle duration of off-road machines. They propose more studies should be done to identify the factors that impact the fuel consumption of diesel engines, specifically construction equipment. Additionally, they highlight that such studies provide an overall profile of all the components and their impact on fuel consumption. Frey, Rasdorf, & Lewis (2010) also emphasize the importance of such information and add that it helps to assess the environmental impact of CE activities.

1.2 Purpose/Aim

The purpose of this degree project is to develop a statistical model that explores the strength of contributing factors that impact the fuel consumption during idling. The outcome of the work will be to optimize emission control technologies to maximize the utilization degree of CO_2 emission reductions as well as the fuel economy of the construction equipment. The model will support the decision-maker or fleet manager to identify equipment’s components which contributes during the idling to quickly decide the most appropriate CO_2 emission technology to maximize the overall benefit.

1.3 Research questions

1. What data analysis techniques are suitable to estimate CO_2 emissions from CE?
2. How accurate the different prediction models are to estimate the fuel consumption of CE?
3. Which factors contribute highly to fuel consumption of CE during idling?

1.4 Scope of study and limitations

Since the construction equipment types vary a lot, a single model of excavator (EC480E) has been chosen. The machine selection is been executed with taking the client's feedback into consideration. However, the chosen model is an equipment with high fuel consumption and high idling duration. The data extraction is been executed from Volvo CE's telematic system which represents the equipment's performance in Africa, Middle-east and Europe region. During the combustion process of diesels engines several particles is released into the environment, but this experiment is focused on estimation of CO_2 emission during idling condition of the equipment. This degree project is been executed anonymously and do not take the operators identity and operator skills into account.

2 LITERATURE STUDY

This section is conducted to cover and present the theoretical framework of this study. The section starts by identifying the idle condition of the engine and its reasons. Further, the section explains the factors that contribute to idling.

2.1 Idle-condition of the equipment

Several studies assume the idle-condition of the engine as an issue that should be taken care of. Sajjad et al. (2013), remarks that the idle-condition of an engine harms the environment, economy, and health. Raczon (2016), confirms that idle-condition is an issue and needs to be minimized to reduce fuel consumption in construction equipment and increase the fuel efficiency.

Perozzi, Michele, Molari, & Eugenio (2016), analyzes the idle condition of tractors. However, idling is an engine condition in which the engine is on without purpose and the machinery is not in movement. Furthermore, the idle span of tractors can be classified as follows:

1. Short term: headland turns
2. Medium-term: job activity transformation
3. Long-term: comfort purpose or keep the engine at optimal condition especially in cold region.

Matthews, Rutty, & Andrey (2017), perform a similar study for heavy trucks. However, the authors define idle-condition a situation in which the engine is on and the vehicle is not in motion. Furthermore, the idling reasons for trucks can be categorized as following:

1. Startup the engine (waiting period to reach the optimal condition of the engine)
2. Waiting period because of drivers' personal purpose
3. Waiting period because of traffic conditions (red lights, heavy traffic, etc.)

Brodrick, Dwyer, Farshchi, Harris, & King Jr (2011), observe that idling costs about $838 million and 2 trillion gallons of diesel fuel. Moreover, it releases hazardous substances into the atmosphere and damages the engine. Matthews, Ruth, & Andrey (2017), indicate that during idling, engine generates vibration and noises that harms the driver's health.

Sajjad et al. (2013), analyzing the impacts of idling of passenger cars regarding fuel consumption and exhaust emissions. The analysis states that, in average, a diesel vehicle consumes 7-liter fuel/h and emits 16 kg/h CO_2 during idling. The analysis confirms that the emission content has a positive correlation with the ambient temperature. In cold temperature conditions, fuel consumption is higher compared to warm temperature conditions. According to International Energy Agency (2005), geographical locations with cold ambient temperatures refer to places where the ambient temperature is equal to or lower than 10 degrees Celsius for six months of a year. Places with a warm ambient temperature refer to locations where the ambient temperature is 25 degrees Celsius or higher within six months during a year.

Perozzi, Michele, Molari, & Eugenio (2016), states that off-road machines are stationary, therefore it's a difficult task to allocate the idling duration. However, the authors propose that by installing different specific sensors on the engine it's possible to monitor the idling pattern of the machine. Chrentienne (2014), states that identifying the factors that lead to idle-condition of the machinery is significant. Moreover, by developing mathematical tools regarding optimization of planning and scheduling it can be minimized. Fan (2017), explains that optimization of planning and scheduling of the construction site to reduce the idling condition brings only 10% reduction in fuel consumption, which is almost impossible.

2.2 Factors that impact the fuel consumption and exhaust emission of CE

Fan (2017), indicates that there are many factors that influence the fuel consumption and exhaust emission of the construction equipment. Nevertheless, those factors can be classified into four main categories. Figure 6 is adopted based on the information from the author's paper.

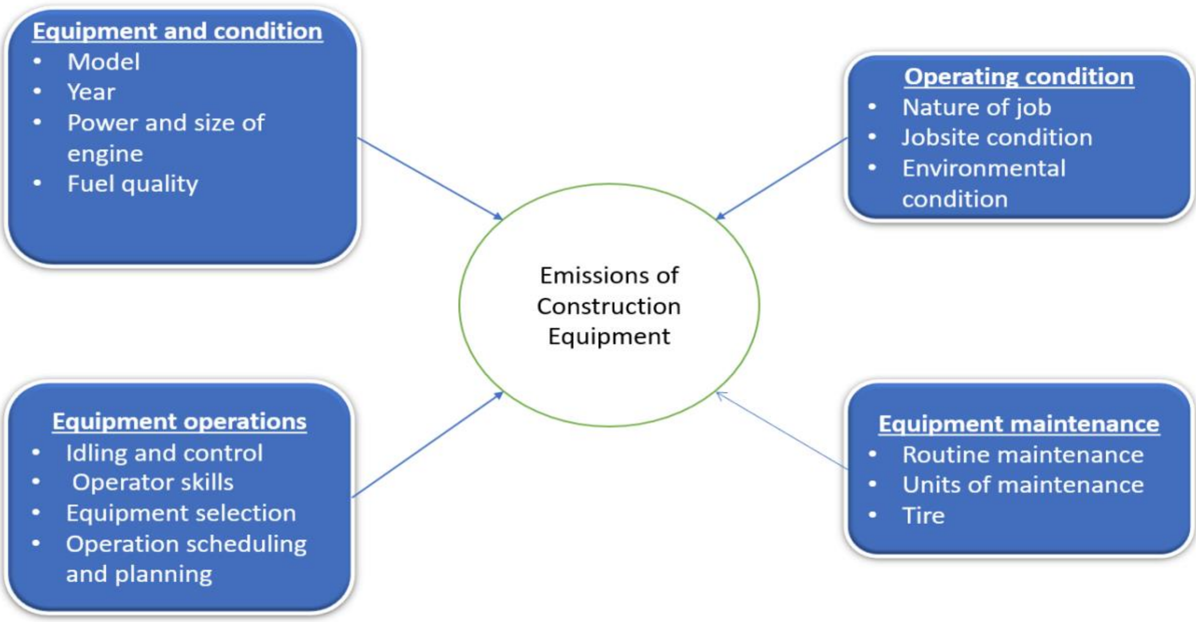


Figure 6 Impacting factors on CE: s exhaust emissions

2.2.1 Equipment and condition

Fan (2017), describes that engine manufacturing year, horsepower, manufacturer of the engine and engine size are the main factors affecting the emission rate of construction equipment during operation. Sandanayake, Zhang, & Setunge (2015), observe that the performance of the CE machines deteriorates by the time. Consequently, the rate of fuel consumption and exhaust emissions of the machinery increase. Fan (2017), states that by the implementing manufacture's instruction manual, the contractor can control this issue within the addition of maintenance rate.

Zietsman & Perkinson (2005), observes that the design of diesel engines is not similar, which impact the fuel consumption and emission rate of the engine. This fact is especially true in comparison to new CE machines versus old models. Hence, the new engines used to meet the regulations standards such as Tire 4 or Stage V (check section 1.1.3). Fan (2017), observes that the rate of the machinery exchange to the new generation is very slow because of the large investment cost of the CE machines.

2.2.2 Equipment maintenance

Fan (2017), observes that maintenance has a strong impact on the performance and fuel consumption of the construction equipment. Likewise, every machine has a specific manual and maintenance routine that needs to be followed properly. Waad (2020), highlights that fuel quality is not the same worldwide, not the condition of the machinery either. Hence, manufactures have developed different instruction manuals for different machines. Fan (2017),observes that poor maintenance strategy can cause higher repair cost which will have a negative impact on fuel consumption and exhaust emission as well.

Shukla, Gupta, & Agarwal (2018) and Resitoglu, Altinisik, & Keskin (2014), observe that after-treatment of exhaust emissions from diesel engines requires timely replacement of filters before it's saturated. Hence, any default in the after-treatment system of diesel engines will negatively affect the performance of the machinery. Therefore, it changes the fuel consumption rate and exhaust emission character of diesel engines. Fan (2017), states that timely replacement of tires can minimize fuel consumption and increase the productivity of the machinery.

2.2.3 Equipment operations

Fan (2017) and United States (1994) observe that one of the biggest constructor struggles is to determine the right machine for the right job. Hence, every job activity has its own requirements to have higher productive usage. However, if the machinery overmatches or annihilates the job-site it will change the fuel consumption rate negatively. United States (1994), highlights the importance of fleet management of construction equipment to reduce the fuel consumption rate and increase the productivity of the job-site.

Waad (2020) claims that in fleet management, operator skills have a significant impact on fuel consumption and emission character of CE machines at the job-sites. Fan (2017), states that a skilled operator performs the tasks by consuming less amount of fuel. United States (1994), states that the development of the operator's competence will improve the fuel economy of the CE machines at construction sites.

Furthermore, Volvo CE (2012) performs a case study to measure the impact of operator skills on fuel consumption and its productivity at the job site. However, 80 operators have been classified into four different groups as follows:

1) novice (operator with 2-10 hours operation experience)
2) occasional (operators with machine knowledge but low operation experience)
3) test operator (Volvo CE employee that have good knowledge operation skills but not professional)
4) professional operator (professional operator and professional knowledge about the machinery). However, all the operators have performed the same task during the same job condition and the results have been ranked. The results establish that there is a huge gap regarding fuel

> consumption and job productivity in between those four groups. Moreover, from a productivity point of view, there is a distinguished gap of 700% between professional operators versus novice operators. However, from a fuel economic point of view, this gape is almost 200%.

Fan (2017), confirms the importance of operator competence in fuel economy of CE machines. Further, skilled operators perform the tasks with a low amount of idling time, they have a better knowledge of how to take care of the machinery and identify the problem of the machine promptly. Nonetheless, skilled operators are expensive to hire and less competent operators can achieve similar results as well.

2.2.4 Operation condition

Fan (2017), states that the application area of construction equipment is as follows: digging, loading, hauling, backlighting, compaction, lifting etc. However, the mentioned job activity has a different job condition that impacts load condition and engine status. Brodrick, Dwyer, Farshchi, Harris, & King JR (2011), highlights the significance of diesel engines transient characteristics in fuel consumption and exhaust emission. However, an internal diesel engine consumes 0.6 gallons/h (2.27 liter/h) at 800 RPM (revolutions per minute) during idling. Afterwards, it increases to 2.25 gallons/h (8.52 liter/h) at 1200 RPM during the same status of the engine.

Matthews, Rutty, & Andrey (2017), observes that some papers establish that altitude of the construction site impacts the fuel consumption of the diesel engine and others do not. Fan (2017), states that diesel engines consume more fuel at high altitudes due to underperformance of the engine which leads to higher exhaust emission. Also, Ashrafur, et al. (2013) highlights the importance of the ambient temperature in fuel consumption of diesel engines. However, diesel engines that operate at the colder temperature consumes more fuel due to engine startup to achieve the efficient working condition.

2.3 Exhaust emissions and pollution reduction techniques

With increasing environmental awareness, many countries have formed strict emission control regulations on diesel engine exhaust (Na & Wei, 2011). Diesel emissions mainly consist of two main components carbonaceous material and soluble organic fraction; besides it also contains other components in smaller sizes such as sulfate and ash. (Shukla, Gupta, & Agarwal, 2018; Na & Wei, 2011; Isermann, 2014). McCormick, Ross, & Graboski (1997), claim that the emission of NO_x and diesel particles can head to harmful effects on human health. Graver, Frey & Hu (2016), emphasize that due to negative effects from the emissions of diesel engines, studies are started around the world to control and improve combustion strategies. Na & Wei (2011) observe that there is no singular technology that is able to reduce all components of exhaust emission, therefore a combination of different techniques is needed to solve the pollution problem. Colin & Allan (2016), observe that technology development in engine emissions can be classified into two categories: active combustion strategy and passive combustion strategy.

According to Shi, Cui, Deng, Peng, & Chen (2006), this categorization is based on the stage of the pollution formation during the combustion process of a diesel engine. An active combustion strategy approach refers to reduce the pollution in the combustion chamber, versus the passive combustion strategy approach refers to reduce emissions before pollutants are released into the atmosphere.

2.3.1 Active combustion strategy

In the 1970s, the war between Israel and Arab countries led to an oil crisis. As a result, Arab countries started to nationalize their oil reserves. Ten years later, the Iraq-Iran war also affected the fuel economy worldwide (Paul, 2014). According to A. J. (2016), the unstable fuel economy market shaped the engine production line of the vehicles manufacturers and motivated them to switch to diesel engines. Paul (2014), describes that during the 1990s, the USA and EU governments built strict restrictions on PM and NO_x, which urged engine manufacturers to develop new engines with lower emissions.

NO_x particles are harmful for human health; therefore, NO_x reduction has been out of the interest of many researchers especially in the field of diesel engines (Cerit & Buyukkaya, 2008). It is worthy to mention, some in-cylinder treatment techniques that reduce the NO_x content of the combustion process of a diesel engine are; injection timing retard, cooling the temperature and exhaust gas recirculation (EGR).

Many researchers believe that the injection timing technique is essential for the minimization of the net emission of a diesel engine. According to Cerit & Buyukkaya (2008), in addition to the minimization of the net emissions of a diesel engine, in this context, fuel economy should also be in the center of attention. However, Zehng, et al. (2014) studies injection strategies and their impact on the emission character with respect to four different fuels (pure diesel, Blends of gasoline, or / n-butanol). The analysis shows that reducing the injection time reduces the emission level of NO_x. Cerit & Buyukkaya (2008), states that delaying of injection time leads to a minimization of the peak cylinder pressure and temperature in the cylinder. This reduces the NO_x formation significantly during the combustion process.

Isermann (2014), emphasizes the importance of air temperature in pollution formation. However, by decreasing the air temperature, the air's heat capacity will increase in the cylinder. As a result, the peak pressure and temperature decrease in the cylinder, which also reduces NO_x emissions.

Wei, et. al (2012), observe that EGR is one the most effective in-cylinder techniques that have been developed to restrict NO_x accumulation. Na & Wei (2011), states that EGR is conformed for gasoline engines but it works with the same effect for diesel engines. Nonetheless, in the presence of EGR the combustion temperature is dropped which in turn leads to reduction of NO_x emissions. Shukla, Gupta, & Agarwal (2018), observe that automobile manufacturers have modified and improved the EGR technology for modern diesel engines. This technique is proficient to narrow NO_x formation by diminishing oxygen levels and decreasing combustion temperatures (Wei, Zhu, Shu, Tan, & Wang, 2012; Gupta, & Agarwal, 2018). Wei, Zhu, Shu, Tan, & Wang (2012), describe that by combining EGR with other emission reductions techniques, the net NO_x reduction of in-cylinder treatment will be more comprehensive.

Recently, diesel particles have been the target of many regulations. Diesel particles can be classified into two categories, Soluble Organic Fraction (SOF) and Non-Soluble Organic Fraction (Mohakumar & Senthikumar, 2017). Agarwal, Gupta, Shukla, & Dhar (2015) emphasize the importance of engine load in particle formation. According to Mohakumar & Senthikumar (2017), diesel particles are in the range of size from 7.5 to 1.0 microns and highlighting the importance of the particles size for human health. However, smaller particles are intended to be more dangerous for human body. Agarwal, Gupta, Shukla, & Dhar (2015), observe that fuel is of great importance in particle formation context. Due fuel with low hydrocarbon and high oxygen content contributes with smaller particle emission. However, one of the most common techniques that reduce the particulate matter formation is air management. Cerit & Buyukkaya (2008), emphasize the importance of air in the pollution formation of combustion engines. Zheng, et al. (2014), observe that particle formation is entirely dependent on how the air flows and mixes inside the engine chamber. LGuarieiro, Souza, & Torres (2009), observe that the design of a diesel engine has a decisive factor, in terms of air handling and particle emission reduction.

2.3.2 Passive combustion strategy

The most effective technique that is sufficient to reduce NO_x emissions is EGR, but using EGR causes increase in particle emissions (Shukla , Gupta, & Agarwal , 2018). It was mentioned in earlier sections that some in-cylinder technologies could affect the emissions of diesel engines. However, Kulkarni & Mohanta (2010) and Resitoglu, Altinisik, & Keskin (2015), observe that in order to meet emission standards, it is not enough to only implement in-cylinder techniques; after-treatment techniques are also necessary to satisfy the requirements of the regulations. However, some common after-treatment techniques that reduces the formation of NO_x are; NO_x storage-reduction (NSR) catalysts and selective catalytic reduction (SCR). However, Shukla, Gupta, & Agarwal, (2018), speaks that NSR reduces NO_x progressive in two steps. Firstly, substances are collected and secondly, released and reduced during engine processing by the extension of an additional hydrocarbon. Further, Shukla, Gupta, & Agarwal (2018) claim that SCR technology has been used since the 1980s in various industries in pollution control purpose. This technique mainly consists of ammonia as the main substance in a water solution. SCR consists of three steps: hydroxylation catalyst section, SCR catalyst section, and oxidation catalyst section.

Two after-treatment techniques that prevent the particular emission into environment are diesel oxidation control and diesel particulate filter. However, diesel oxidation control (DOC) mainly consists of a ceramic monolith and besides diesel particles also combats CO and HC emissions. However, the degree of efficiency of the DOC is dependent on the exhaust gas temperature, where maximum particle removal takes place at exhaust temperature ranges between 200-350 degrees Celsius (Shukla , Gupta, & Agarwal , 2018).

Diesel particulate filter (DPF) is a flittering process consisting of two steps of filtering and regeneration. During the filtration process, large amounts of soot are continuously collected, which leads to the pressure difference across the filter and affects engine performance. When the filter is fully saturated then the removal process should occur, which is called the regeneration process (Resitoglu, Altinisik, & Keskin, 2014).

2.3.3 CO2 neutralization approaches

Paul (2014); and Shi, Cui, Deng, Peng, & Chen (2006); and Colin & Allan (2016); and Graver, Frey & Hu (2016) state that since the 1970s most of the regulations have been focused to delaminate the amount of on PM and NO_x emissions. Na & Wei (2011) asserts that the concentration of CO_2 has increased on average by 1-2 PPM per year. Zheng, et al. (2014), illustrate that CO_2 is the primary product of a diesel engine and is a greenhouse gas. Agarwal, Gupta, Shukla, & Dhar (2015) observe that diesel engines release large amounts of CO_2 and steam into the atmosphere which are considered to be unregulated. Johnson (2008) states that the presence of strict emissions regulations that focus on CO_2 emissions could reduce CO_2 emissions and improve fuel economy. According to Agarwal, Gupta, Shukla, & Dhar (2015), regulation of CO_2 is in the process as well on account of its negative effects on the environment.

Jonsson & Bondemark (2017), states that there is some potential technical solution that may reduce the CO_2 emission from the construction equipment such as hybridization and electrification of the machinery. However, there are some fully developed hybrid excavators and electrified excavators on the market. Fan (2017), describes that by having the lifetime and initial investment cost of construction equipment, more actions are needed in this field to improve the fuel efficiency and reduce the environmental costs. Jonsson & Bondemark (2017) comes with the issue that, operational safety is critical for all of the constructors. Conversely, the contractor is avoiding to invest in new technology. Hence, makes it harder for greener products to manifest itself in the market.

Deng, et al. (2011), denies the character of hydrogen engines during idle as well as lean condition (machine is in motion). The result is based on a simulation program that shows that the combustion in the presence of low hydrogen leads to less CO_2 emissions during both idle-lean-condition and the engine efficiency is higher in presence of CO_2 low hydrogen. Lee, Woo & Park (2017), denies a similar study for a rail diesel engine, the results emphasize the importance of the pilot-timing as a significant factor in fuel consumption and exhaust emissions of the rail diesel engine.

2.4 Machine learning

This section is conducted to presents the concept of machine learning and its application area. It will also present some general techniques in this field.

2.4.1 Basics of machine learning

Machine learning is a part of artificial algorithm and it describes the driver behind a data set. Ayyadevara (2018), describes that machine learning is a technique that uses a set of data to explore its condition and recognizes the input data's patterns and its correlation to each other. This technique adjusts the data to minimize the error rate between the input data and create a model. Shalev-Shwartz (2014), describes that there are two different algorithms in developing a machine learning model: supervised and unsupervised learning. Supervised learning distinguish itself from the other algorithm through presents of output variable which is called response variable.

Brendan (2018), states the difference between unsupervised learning and supervised algorithm is the absence of label. However, an unsupervised algorithm uses a set of data that is not been labeled in contrast the input data for a supervised algorithm is labeled. Bironneau & Coleman (2019), state that the problem area of a supervised algorithm is to conclude the best calibration between the input and output data based on labeled variables. Figure 7 displays the field of machine learning and is created from information from Bironneau & Coleman (2019).

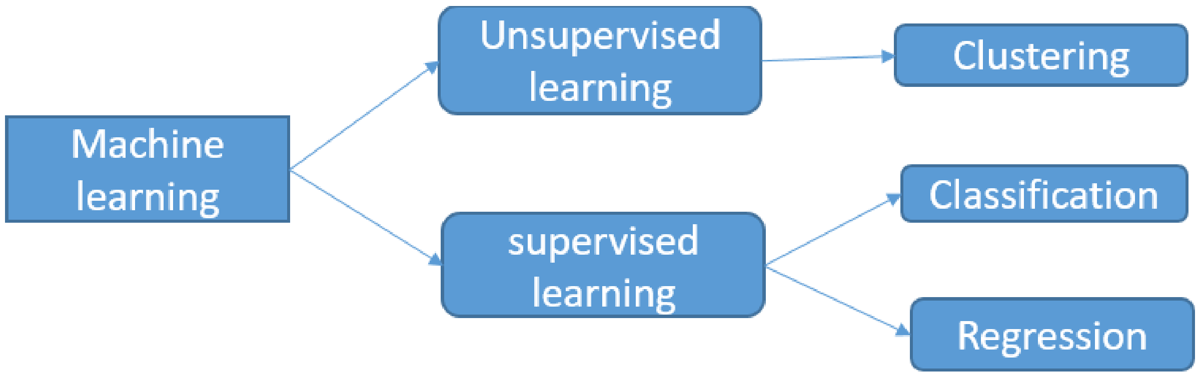


Figure 7 Machine learning techniques

According to Brendan (2018), unsupervised algorithm can further be classified into clustering. However, clustering uses the input data and classifies it into different groups. The most commercial clustering techniques are; K-Means, k-medoids fuzzy C-means, hierarchical, Gaussian mixture, Neural Networks and hidden Markova model (Mathworks, 2016). Brendan (2018), observes that supervised learning algorithm can further be classified into classification and regression. Bironneau & Coleman (2019) states that through the classification, the model take the unknown input data and based on the information attemps to match it with a known output sample. The most common classification techniques are: Support vector machine, discriminant analysis, Naïve bayes and nearest neighbor. Bironneau & Coleman (2019) observes that regression techniques can use eigher known or unkown sample to label it. The most commercial techniques in the field of regression are: Linear regression, support vector machine, ensemble method, partial least square and neural networks (Mathworks, 2016).

Further, Bironneau & Coleman (2019) and Eriksson, et al (2013) observes that principal components analysis (PCA) can be used in both fields of supervised and unsupervised machine learning, PCA is a technique which maps the structure of the dataset. More information about this will be presented in section 2.4.3.1.

However, Ayyadevara (2018) mentions that development process of a supervised model takes place during two phases: training and validation phase. During the training phase the variables are processed to predict the response variables with high degree of accuracy. However, during the validation phase the trained model get tested with new input data for some to outline the accuracy of the model with respect to the reference variables. This phase represents the model's accuracy and performance of the model. Shalev and Shwartz (2014) states that it is important that during the testing phase new input data uses, the purpose of this phase is to observe the model's correctness when it comes to new data. Moreover, if the machine learning algorithm has high accuracy in the first phase but low accuracy in the second phase, the model is considered to be overfitted.

2.4.2 Quantitative vs qualitative method

Quantitative methods quantify the properties of the examined material based on the input information. In this case, the properties are numeric for instance fuel consumption, temperature and moister. Qualitative method refers to non-numeric properties of the examined materiel for instance, color, engine type and machine type. However, in the case of quantitative method linear techniques are the most common methods that are used in this field. These methods describe linear relationships between response variables versus predict variables.

Multivariate data analysis (MVDA), is a technology that expresses response variables as a linear function of predicted variables. Eriksson, et al (2013) observer that MVDA is meant multiple linear regression (ML-R), but it also can be handled with projection techniques. The mentioned technique represents the variables in K-dimensional space regarding their projection. Based on the observed information coordinate the variables into a smaller dimension plane or hyperplane. The issue regarding projection techniques is the model will transmit in the presence of many predictor variables that have not the effect of the response variable. However, to avoid this problem pattern recognition techniques should apply to the input data. Eriksson et al (2013) observed that pattern recognition is a synonym for ML-R in chemical engineering and chemistry. Furthermore, when we know something about the problem by the implementation of pattern recognition techniques it will provide an overview of the problem.

2.4.2.1. Pre-work the data

Eriksson, et al (2013), Bironneau & Coleman (2019) and Mathworks (2016), mention that if the data contians several number of variables that contributes similar information to the model, the model can be overfitted. Howere, to avoiding the models to be overfitted some approches are needed to be considerd. The following section represents some common pre-processing techniques.

2.4.2.2. Correlation coefficient

Mathworks (2016) states that in MVDA especially in field of regression it's very important that there is a relationship among to predict variables and response variable. However, one manner to obtain this level of understanding is to check the correlation coefficient among predict versus response variables. Correlation coefficient is statistical tools which measure the strength of relationship between two variables. The following equation describes how to calculate correlation coefficient:

$$\rho_{xy} = \frac{Cov(x, y)}{\sigma_x \sigma_y} \qquad \text{Equation 1}$$

Correlation coefficient is defined in the range of [-1, 1]. Moreover, 1 represent a perfect correlation within two variables which means that every positive increase in one variable will lead to increase to the other variable. Furthermore, -1 represent a perfect negative correlation within two variables which means that increase of one variable will decrease the other variable. Correlation coefficient at zero or close to zero means that there is no relationship within the variables and variables are independent of each other.

2.4.2.3. Feature scaling

Eriksson, et al (2013), observe that the input data usually have substantially different numerical ranges. Dass (2012) and Eriksson et al (2013), indicate that pattern recognition techniques are based on variables' projection on each other. However, if this variance is large within the variables, the variables with a large variance will dominate the variables with less variance. Therefore, the input data should be normalized.

Eriksson, Byrne, Johansson, Trygg, & Vikström (2013), Dass (2012), states that projection methods are sensitive to the scale of the variables. One approach to obtain a better overview about the dataset to understand if feature scaling is needed or not is implementation of Euclidian distance. However, Euclidian distance generates straight distance between two variables in the k-space. The following equation express the Euclidian formula:

$$d = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \qquad \text{Equation 2}$$

Where:

$d \coloneqq distance\ between\ two\ variables$

$x_j\ \&\ y_i \coloneqq is\ the\ coordinates\ of\ the\ scatter\ point$

Table 4 has been extracted from the raw dataset in order to show the variance between the variables of this work. However, as it is shown in the table the variance of the duration of idle time is higher compared to the number of engine shutdown (when the engine is turned off by the operator) at different engine mode.

Table 4 Variance of different numerical ranges

time in I2	Time in I2	number of engine shutdown at <600 rpm	number of engine shutdown at 600<800 rpm
37602	190093	2	0
293372	738566	22	7
1996526	7401094	58	639
2868249	1394161	11	729
4414415	3760613	13	166
4485708	2612493	62	270
3633943	240402	1	256
1268504	3675828	7	544
3297508	1006560	64	1381
6288007	4729979	87	242
2027888	2145995	3	516
2559737	7693052	18	1058

Description: 1 the unit of each column is time in seconds

There are many ways to normalize a dataset, Unit variance (UV) and Max-min rescaling are the most common scaling features (Eriksson, et al 2013; Jeng & Chen, 2016).

The equation below expresses the unit variance (UV) procedure:

$$UV = \frac{1}{s} * X_i$$ Equation 3

Where s refers to standard deviation which can be calculated by the expression below:

$$s = \sqrt{\frac{\sum_{i=1}^{N}(X_i - \bar{X})^2}{n-1}}$$ Equation 4

Where:

- s: *Standard devition*
- $\bar{X}$: *mean*

Max-min rescaling approach can be express by the following equation:

$$X_i = \frac{X_i - X_{min}}{X_{max} - X_{min}}$$ Equation 5

Where:

X_{max} and X_{min} ≔ are the maxmimum and minimum of the varible i

Figure 8 shows the impact of rescaling feature. As it is shown in the figure below the dataset contain similar information in the both sides. However, the variance in between variables are lower after rescaling with respect to non-scaling.

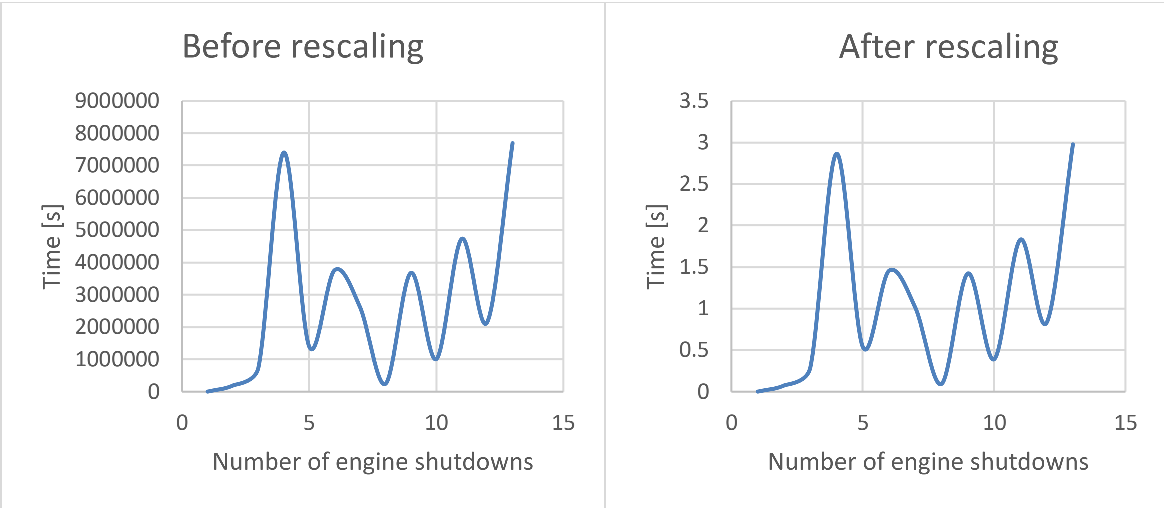


Figure 8 Before and after rescaling of the dataset (UV-rescaling)

2.4.2.4. *Mean centering*

Eriksson, et al (2013), state that mean centering is the next step of the pre-processing of the data set. Mean centering determines the average value of the variable and is extracted from the data.

The equation below expresses the mean centering procedure:

$$X = X_i - \bar{X}$$

Equation 6

Figure 9 demonstrates a graphical procedure of mean centering. In the left first and second principle component have been derived in the k-space without mean centering and in the right, similar condition is valid with presence of mean centering. The figure is created by the information presented in Eriksson, et al (2013) book.

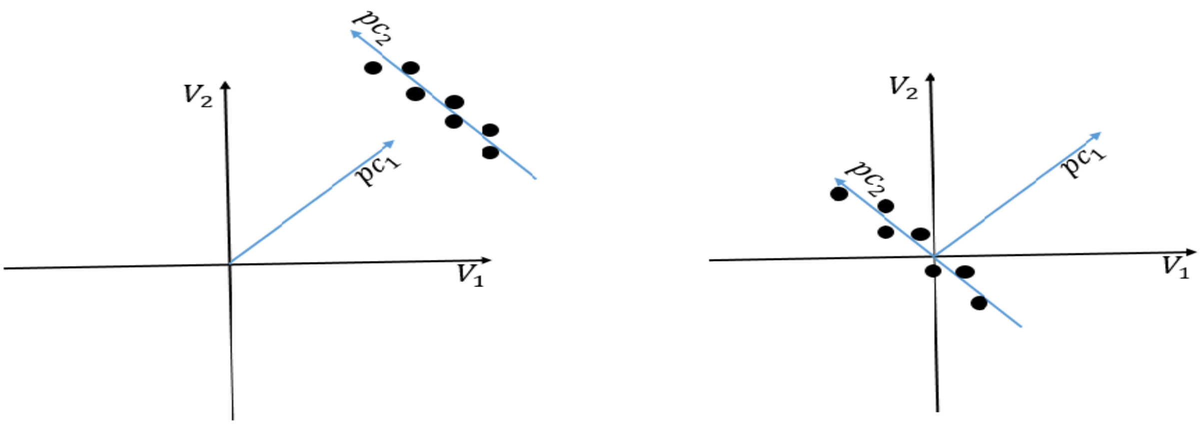


Figure 9 Mean centering

Mean centering of a dataset stabilizes the data and mean-centering procedure is not time consuming (Kamyabi, 2009).

2.4.3 Training projection model

This section presents principal components analysis and partial least square regression in detail by presenting geometric interpretation of each technique.

2.4.3.1. Principle component analysis

Chen (2016), states that principal component analysis (PCA) is used in both qualitative and quantitative data analysis. Eriksson et. al (2013) state that PCA generates a summary of the data and gives an overview of the variables. Furthermore, PCA measures and presents how variables are related to each other, which variables contribute the same information to the model and which variables generate unique information to the model. Dass (2012) and Razmkhah, Abrishamchi, & Torkian (2010), observe that PCA is also known as Karhunnen-Loeave expansion. However, Chen (2016) states that PCA is a variable reduction technique, this procedure is used to avoid non-homogeneity in sampling data and identifies temporal variation and also extracts the most important parameter.

Moreover, in explanatory data analysis, every variable defines an axis, as a result, develops a multidimensional variable space. Conversely, every observation defines a point at the multidimensional variable space. PCA by considering the maximum variation of the dataset finds its direction and derives a line that passes the center of the dataset. However, this line calls the first principle components as it visualized on the left-hand side of the figure below. Regularly a single principle component does not describe the data variability. Thus, the repetition of the projection process is needed to obtain a better overview, as a result of this procedure, the second principal component will be derived in the remaining space. This process can be continued as much as the number of variables. However, principal components are orthogonal to each other and contain information about the dataset. Figure 10 visualizes first and second principal components in the space.

Application of PCA will reduce dimensionality and as a result, a new plan will be developed. This has been shown in the figure below, every observation in the PCA plan is called the score plot and every observation have a score on each PC axis. The figure is created based on the information from Eriksson, et al (2013) book.

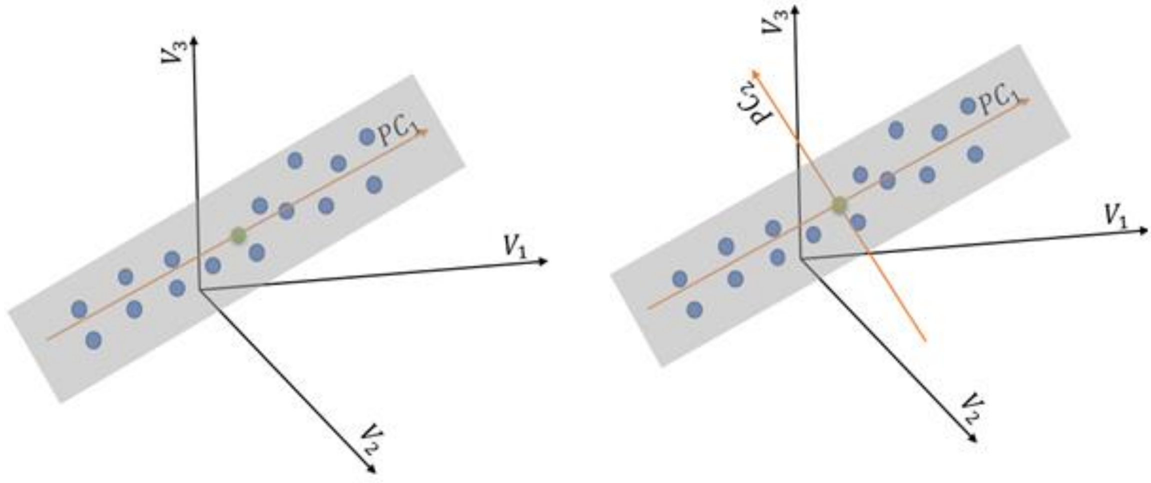


Figure 10 Plan built by principal components

Figure 11 represents the coordinate system that is developed along with first and second principle components. Principle components are orthogonal to each other as it been shown in the figure. The green scatter point indicates the mean value of the samples, the figure is created based on the information in the Eriksson, et al (2013) book.

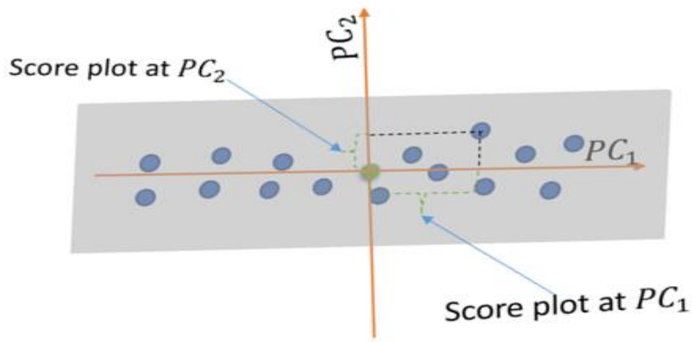


Figure 11 Principal components coordinate system

The position of the samples can be described with respect to principle components vectors. This process can be continued for third, fourth and so on, principle components.

2.4.3.2. Partial least square

Chen (2016) states that in chemometrics, Partial least square (PLS) is a popular method and many properties of this model are known. PLS constructs factors with reference information and presents it in PLS regression (PLS-R). Eriksson, et. al (2013) state that PLS stands for "projections to latent structures by means of partial least squares". However, the accuracy of PLS decreases when there are unrelated variables available. Therefore, the variables are required to be processed in order to increase the accuracy. Chen (2016), suggests that in the presence of many variables, either PLS-Discriminant Analysis (PLS-DA) or PCA should be applied to identify the important variables. However, the problem of PLS-DA in contrast with PCA is less accurate, therefore PCA is proposed during the classification of the dataset.

Moreover, Eriksson, et al. (2013) state that PLS associates the x variables with Y response variables and thereby structures the PLS model. Figure 12 shows the PLS terminology, the matrix in the left represents the x variables (predictor) and the matrix in the right shows the Y response variable. The figure is adopted by the information from Eriksson, et al (2013) book.

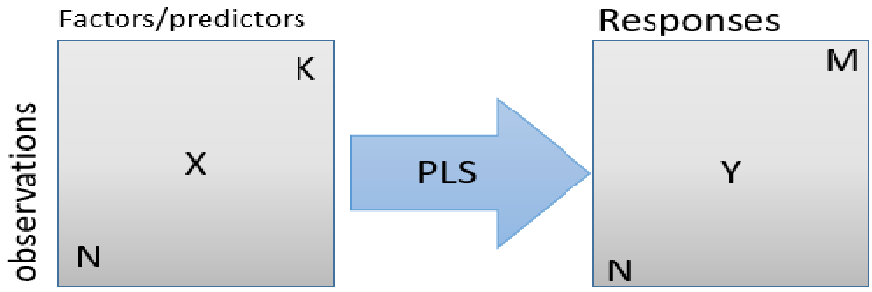


Figure 12 PLS terminology

Eriksson, et. al (2013) observe that there are many discussions regarding the number of response variables. Furthermore, the authors summarize that this is due to the correlation between the data. If the input data is positive correlated so by having many response variables it can lead result to better model. If there are correlation among input data, employing many response variables can help to achieve a better model. PLS perform the projection process for both of the predictors as well as the response variable in order to maximize the covariance between predictor and response variable. Figure 13 shows how this process is performed. The figure is created based on the information in the Eriksson, et al (2013) book.

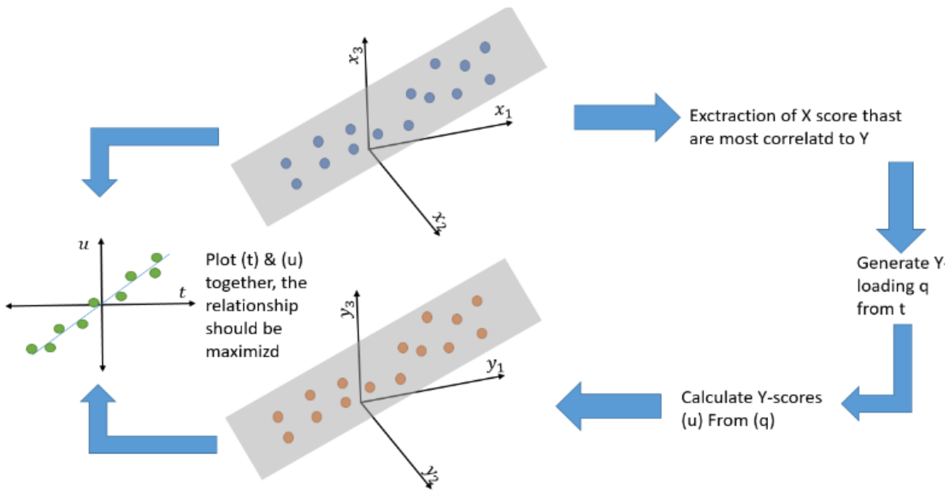


Figure 13 Projection to latent structures

3 METHOD

This section presents the method used to conduct this experiment. At an early stage of this work, a state of art was developed to obtain a general information about the topic as well as to increase the level of understanding in the problem area.

3.1 Raw data and chemometrics

Raw data set was collected from the Volvo CE's data warehouse. PCA analysis was adopted in order to understand the structure of variables as well as to find the relationship between them.

3.1.1 Samples

Samples used in this experiment is a data set of 362 observations and 36 variables have been extracted from the Volvo CE's telematics systems data warehouse. All the variables are presented in seconds except the response variable (see 3.2). In this section, a description of all variables is presented.

3.1.1.1. Response variable

The data set describes the fuel consumption during idling by the following variables:

- Fuel used at idle 1 mode [hz]
- Fuel used at idle 2 mode [hz]
- Duration of machine at idle 1 mode [s]
- Duration of machine at idle 2 mode [s]

3.1.1.2. Predictors

The data set describes the following parameters that impact the fuel consumption during idling:

- Hydraulic oil temperature in degree Celsius (duration of the engine at different hydraulic oil temperature spans in the range of [-20,120] seventeen different temperature span) [s]
- Machine Utilization (idle mode, work mode) [s]
- Air conditioner (auto and manual mode) [s]
- Idle time before engine shutdown [s]
- Number of idle times before engine shutdown (fourteen different RPM spans in range of [0,3000] RPM [s]
- Engine coolant temperature (six different temperature spans in range of [0,180] degree Celsius) [s]

3.1.2 Pre-processing of data

Section 2.4.2 highlights the significance of pre-processing of the raw data. In this study the following pre-processing features is adopted and different models are developed:

- Correlation coefficient matrix
- Scaling feature (Max-min & unit variance normalization)
- Mean centering

Correlation coefficient between predictors and response variable is calculated to gain a further understanding of the dataset. Two different normalization methods are used and similar models are developed to find a proper one. However, max-min normalization method finds the maximum value of the variable and normalize others correspond to it and unit variance method normalize the variable from the mean. Further, mean centering method finds the mean of each variable and moves it towards zero.

3.1.3 Training

This section presents the different regression models that is been developed in this work. PCA is developed to obtain the structure of the input data before training the model. However, regression models can be classified into two different categories: classic method of statistics and projection method of statistics.

3.1.3.1. Classic method of statistics

Multivariate linear regression is a common method that is been used in this field (Brodrick, Dwyer, Farshchi, Harris, & King JR, 2011; Fan, 2017). Further, according to Eriksson, et. al (2013) before training a multivariate linear regression, the following assumptions is needed to be considered:

- Predict variables are independent
- Predict variables are exact
- Errors are randomly distributed
- The number of observations is higher than the number of variables

3.1.3.2. Projection method of statistics

Gopal, Shetty, & Ramya (2014); and Eriksson, et. al (2013) states that the following projection models are the most common in explanatory data analysis:

- SVM-R
- PC-R
- PLS-R
- GP-R

Eriksson, et. al (2013) states before training a projection model, the following assumptions is needed to be considered:

- Predict variables are not independent
- Predict variables may have errors
- Errors may be structed
- The number of variables are higher than the number of observations

3.1.4 Performance of the model

Root mean square errors (RMSE) and standard correlation coefficient (R_2) are the most popular techniques to evaluate the accuracy of a regression model. However, RMSE describes the standard deviation between predicted versus reference response variables. RMSE expressed by the following equation:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(\bar{y}_i - y_i)^2}{N}} \qquad \text{Equation 7}$$

Where:

$\bar{y}_i$: Predict value

y_i: Reference value

N: Total number of predicted sample point

R_2 is measured on the scale of zero to one and describes the accuracy of the model corresponding to regression line. The R_2 can be expressed by the equation below:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(\bar{y}_i - \widetilde{y}_i)^2}{\sum_{i=1}^{N}(y_i - \widetilde{y}_i)^2} \qquad \text{Equation 8}$$

Where:

$\widetilde{y}_i$: Mean reference value

$\bar{y}_i$: Predict value

y_i: Reference value

N: Total number of predicted sample point

3.2 Flowchart of the model

In this section, an overall workflow of the data analysis of this paper is summarized. Figure 14 presents the progress of developing the regression models.

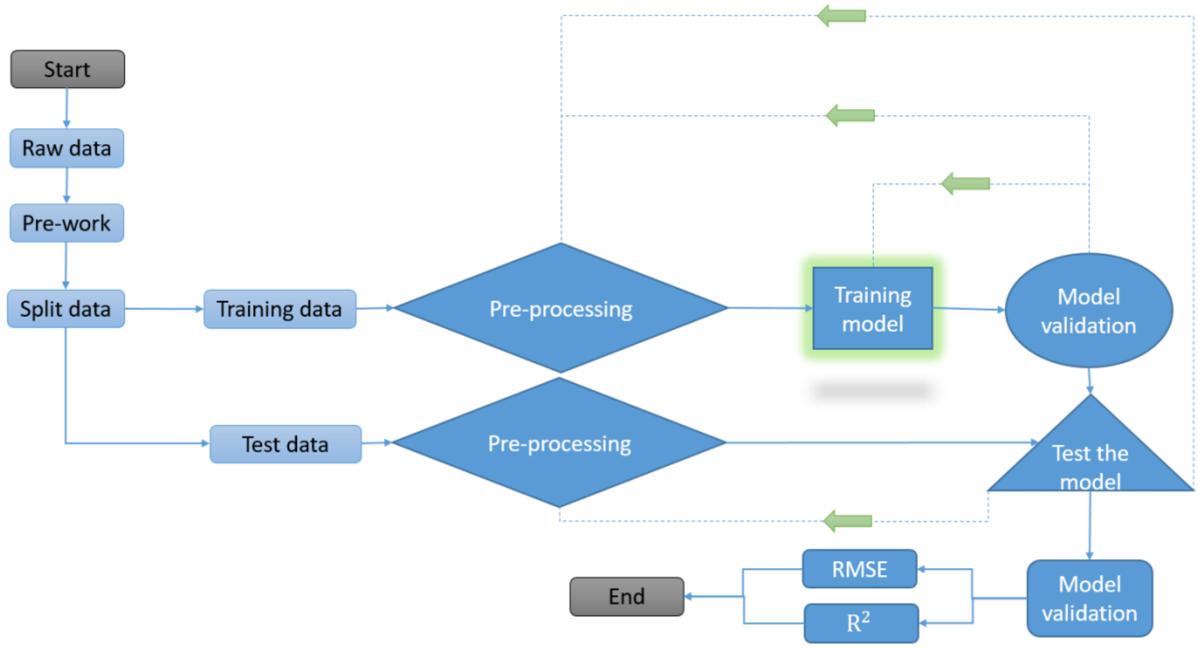


Figure 14 Workflow of the model development

The pre-work section indicates the data treatment of the dataset before proceeding to train the dataset. The missing data was identified and removed. There are four variables that represent the idle condition of the equipment, two of them is representing the idle duration and the others are representing the fuel that used during that period of time. Hence, the predictors are not representing which variable corresponds to which idle mode. The following equation is needed to be developed to describe the idling as a single response variable:

$$total\ duration_{idle} = duration\ at\ idle\ mode\ 1 + \ duration\ at\ idle\ mode\ 2 \quad \text{Equation 9} \quad [s]$$

Fuel used during idling is calculated by the equation below:

$$total\ fuel\ used_{idle} = fuel\ used\ at\ idle\ mode\ 1 + fuel\ used\ at\ idle\ mode\ 2 \quad \text{Equation 10} \quad [Hz]$$

The unit of used fuel is in hertz (Hz) and every Hz is 0.472 L fuel. However, fuel consumption per second is been calculated as follow:

$$Fuel\ consumption_{idle} = \frac{(\ total\ fuel\ used\ in\ \times 0.472)}{(total\ duration_{idle})} \quad \left[{}^{liter}/_{\text{seconds}}\right] \quad \text{Equation 11}$$

The proportion of engine coolant temperature, hydraulic oil temperature and air conditioner corresponds to idling are not given. Therefore, those variables are been estimated by following expression:

$$Engine\ coolant\ temperature_{idle} = \frac{Operation\ at\ idle\ mode}{Total\ operation} \times Engine\ coolant\ temperature\ [\text{s}] \quad \text{Equation 12}$$

$$Hydraulic\ oil\ temperature\ _{idle} = \frac{Operation\ at\ idle\ mode}{Total\ operation} \times hydraulic\ oil\ temperature\ [s] \quad \text{Equation 13}$$

$$Air\ conditioner\ _{idle} = \frac{Operation\ at\ idle\ mode}{Total\ operation} \times Air\ conditioner\ [\text{s}] \quad \text{Equation 14}$$

The raw data was randomly split into training (80%) and testing (20%) samples. The splitting process has been made using a function in MATLAB. The dashed line in Figure 14 indicates iteration loops, the number of iterations is dependent on the accuracy of the model, iterations stops when the accuracy reaches as high as possible. After the model is trained, it is validated through cross-validation, it indicates that the model is validated by same samples as it was trained. However, when the desired accuracy is been achieved, the model is validated by employing new samples.

3.3 Artificial neural network

Figure 15 presents the training process of artificial neural network (ANN). Firstly, the weight of each variable is initialized. Later, the error at each node is calculated and based on the response variable, the weight is adjusted. Step 2-4 is repeated in order to minimize the error. In the same way, step 2-5 is repeated until the error is as less as possible. In this case, this particular step will iterate 100 thousand times.

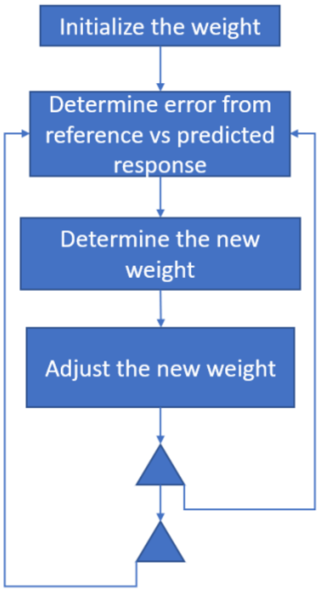


Figure 15 Work flow of training a neural network

More information about ANN and its structure will be presented later in this report (see 4.3).

3.4 Quantification of CO_2 emission

Quantification CO_2 emission of construction equipment generally can be executed through three different manners: estimation based on fuel used, estimation based on EPA emission factor for off-road machines and portable emissions measurement (PEMS). However, PEMS is the most accurate manner to estimate the exhaust gas emission. PEMS is a vehicle emissions testing device that measures the emissions on a timely basis. Nevertheless, due to unavailability of sources in this work, the application of PEMS is not feasible. Therefore, other approaches should take into consideration to perform this section.

Another method that is common in this field is the estimation of exhaust emission by using an equation developed by EPA. This equation is used for any off-road combustion vehicle. The outcome of the equation is the CO_2 emission factor in [g/hp×hr], hp indicates the horsepower and hr indicates operation hour at that specific horsepower. Further, to convert this factor into ton CO_2 emission, the following step is performed:

$$\left[\frac{g\ CO_2}{hp \times hr}\right] \times 0.000001 \rightarrow \left[\frac{ton\ CO_2}{hp \times hr}\right] \qquad \text{Equation 15}$$

Thus, the dataset contains two different variables that describe the idle condition of the engine (idle 1 & idle 2), the main difference between idle 1 and idle 2 mode is that they belong to different engine speeds. Idle 2 occurs in revolutions per minute (RPM) range of [0, 800] and the RPM range of idle 1 is [800, 1000]. To perform the next step of CO_2 emission estimation, the following assumption is needed to be considered:

Idle 2: all the duration of this mode is executed at 1000 RPM

Idle 1: all the duration of this mode is executed at 800 RPM

Which gives for different idle modes:

Idle 2

$$\left[\frac{ton\ CO_2}{hp * hr}\right] \times hp[horsepower\ regarding\ to\ 1000\ RMP] \times hr\ [total\ duration\ at\ this\ engine\ mode] \rightarrow\ ton\ CO_2 \qquad \text{Equation 16}$$

Idle 1

$$\left[\frac{ton\ CO_2}{hp * hr}\right] \times hp[horsepower\ regarding\ to\ 800\ RMP] \times hr\ [total\ duration\ at\ this\ engine\ mode] \rightarrow\ ton\ CO_2 \qquad \text{Equation 17}$$

Horsepower is determined as a function of engine speed and engine torque by the expression blow:

$$hp = \frac{RPM \times T}{5252} \qquad \text{Equation 18}$$

Where:

RPM: revolutions per minute

T: torque

Torque is equal to 649.2 and is obtained from Volvo CE's engine-pump-control documents regarding to model EC480E.

However, the other common estimation method for exhaust gas emissions is based on the amount of fuel used. Juhrich (2016) says that the composition of diesel fuel is seasonal which means that the carbon content is varied during summer versus wintertime. Therefore, the estimation of CO_2 emission, assumes that the carbon content is constant and does not consider the variability of the composition of diesel fuel during the year.

Further, the dataset is not a real time domain and does not represent the excavator's seasonal operation hours. The dataset has also been collected from a wide geographical area such as Europe, Middle East and Africa. Thus, the fuel quality varies and it is a difficult task to evaluate the composition of diesel fuel in every region. However, Integrated Pollution Prevention and Control (IPPC) proposes 2.64 [kg CO_2/liter fuel] emission factor. However, the CO_2 emission are been estimated by the following equation:

$$fuel\ consumption_{idle} = total\ fuel\ used \times 0.472 \times 2.64 \quad \left[{}^{kg\ CO_2}/_{liter\ fuel}\right] \qquad \text{Equation 19}$$

3.5 Software

Modelling, calculation and pre-working of the data set will take place in MATLAB environment. MATLAB is a programming language for data analysis, mathematic, graphics and other engineering proposes (MATLAB, 2020).

4 CURRENT STUDY

In this section, the material used to obtain results in this study will be described. It starts with outline of the correlation coefficients between the predictor variable and response variable. Afterwards, it continues with training and validating different models based on different pre-processing approaches.

4.1 Correlation coefficient matrix

Multiple linear regression and projection techniques are very sensitive and it can get overfitted if the predictors contain missing values or the predictors are not 100% relevant to the response variable (Eriksson, Byrne , Johansson , Trygg, & Vikström, 2013). However, as it was mentioned earlier in the method section, the dataset contains four variables that describe the idling of the equipment: engine duration at idle mode 1 & 2 and fuel used at idle mode 1 & 2. The response variable has been pre-worked through Equation 9 and Equation 10 in order to convert those parameters to one-unit variable. This section is conducted to discover the correlation coefficient value that corresponds to response variable and make sure that predictors are relevant to the response variable.

4.1.1 Correlation coefficient matrix among Idle 1 versus idle 2 mode

Engine speed is the distinguish link between idle 1 and idle 2 mode of the equipment. Idle 1 represents the engine mode at engine speed rate of [800,1000] RPM and idle 2 represents engine performance at [0,800] RPM. Table 5 shows the correlation coefficient for the fuel used and duration for different idle modes.

Table 5 Correlation between idle duration and fuel consumption

	Duration at Idle 2	**Duration at Idle 1**
Duration at Idle 2	1,00	
Duration at Idle 1	0,45	1,00
Fuel used Idle 2	0,99	0,46
Fuel used Idle 1	0,46	1,00

Description: 2 Time Idle 1& 2 represents the engine duration in that engine mode and fuel consumption Idle 2 & Idle 1 represents the fuel consumption at that specific engine mode

Table 5 shows that there is a strong correlation coefficient among to fuel used at idle 1 corresponds to fuel used at this engine mode.

Table 5 shows that there is a strong correlation coefficient between fuel used at idle 1 and duration at idle 1. In the same manner, there is also a strong correlation coefficient between fuel used idle 2 and duration at idle 2. It is worthy to remind, the more equipment operation at the idle mode, the more fuel the equipment consumes. A similar observation has been made for idle 2 mode of the equipment. However, the correlation coefficient among idle 1 to idle 2 is light positive. The positive correlation, in this case, is because of the functionality of the idle 2.

Idle 2 is the engine's auto idle mode, the engine turns to idle 2 automatically in order to improve the fuel economy.

Table 6 shows the correlation coefficient between fuel used for idle 1 and 2 at different hydraulic oil temperature ranges.

Table 6 Correlation coefficient between fuel used and hydraulic oil temperature

	Fuel used Idle 2	**Fuel used idle 1**
fuel used I2	1,00	
fuel used I1	0,45	1,00
H.T [-40]	0	0
H. T [-40, -30]	0,01	-0,01
H. T [-30, -20]	0,01	0,04
H. T [-20, -10]	0,13	0,06
H.T [-10,0]	0,23	0,17
H.T [0,30]	0,73	0,45
H.T [30,40]	0,76	0,53
H.T [40,60]	0,63	0,79
H.T [60,70]	0,39	0,58
H.T [70,80]	0,17	0,29
H.T [80,85]	0,05	0,11
H.T [85,90]	0,02	0,06
H.T [90,95]	0,00	0,00
H.T [95,100]	-0,02	-0,05
H.T [100,110]	-0,03	-0,07
H.T [110,120]	-0,02	-0,03
H.T [120]	-0,01	-0,02

Description: 3 Hydraulic oil Temperature (H.T) at [,] given temperature, fuel idle (I) 1 & 2 represents fuel consumption in that specific engine mode

Table 5 shows that a numeric change in the fuel used positively impacts the engine duration at idle 1 versus idle 2. Based on that information and simplicity purpose, only the fuel used correlation coefficient corresponding to the predictor will be calculated. Furthermore, it is known that a direct numerical change on the fuel consumption will indirectly impact the idle duration at idle mode 1 and 2.

However, Table 6 shows that the majority of the temperature ranges have a neutral correlation coefficient corresponds to fuel used at idle 1 and idle 2. Despite the hydraulic oil temperature duration at the temperature range of [-10, 80] degree Celsius (see blue colored cells in the table), there is not enough evidence to state that there is a causal correlation as well. Hence, hydraulic oil temperature is calculated by Equation 13 and this variable has been estimated based on some assumption. Moreover, it would be beneficial to take the ambient temperature into consideration to increase the knowledge regarding the causality of hydraulic oil temperature. Nevertheless, because of the Volvo group's protection policy, such information is not open for the public. Evaluating the impact of ambient temperature is an important variable to be considered especially in the regions with the extreme weather condition.

Table 7 shows the correlation coefficient between different engine coolant temperatures at different ranges in degrees Celsius corresponding to fuel used at different idle modes.

Table 7 Correlation coefficient among fuel used and engine coolant temperature

	Fuel used I2	Fuel used I1
Fuel used I2	1,000	
Fuel used I1	0,451	1,000
E_C_T [0,75]	0,642	0,565
E_C_T [75,85]	0,653	0,834
E_C_T [85,95]	0,372	0,289
E_C_T [95,105]	0,416	0,087
E_C_T [105,108]	0,286	0,011
E_C_T [180]	0,269	0,031

Description: 4 E_C_T Engine Coolant Temperature at [,] given temperature interval, fuel idle (I) 1 & 2 represents fuel consumption in that engine mode

According to Table 7, the engine coolant temperature at the temperature range of [0,85] degree Celsius is positively correlated to the response variable (see blue colored cells). It indicates that a numerical change in properties of the engine will impact the fuel used in the same direction. The other temperature ranges of engine coolant are lightly positive correlated.

However, a similar assumption as hydraulic oil temperature was made to estimate the duration of engine coolant temperature at idle mode. Due to a lack of information about the equipment, it is a difficult task to explain if the colored temperature ranges have a causal relation to the response variable.

Table 8 shows the correlation coefficient between numbers of engine shutdown at different engine speed corresponding to the fuel used at different idle mode. However, predictors in this field presents a poor operator behavior.

Table 8 Correlation coefficient between fuel used at idle mode and number of engine shutdown at different engine speed.

	Fuel used I2	Fuel used I1
Fuel used I2	1,000	
Fuel used I1	0,451	1,000
N_E_S_at [600] RPM	0,077	0,259
N_E_S_at [600,800] RPM	0,426	0,248
N_E_S_at [800,1000] RPM	0,477	0,555
N_E_S_at [1200,1400] RPM	0,313	0,452
N_E_S_at [1200,1400] RPM	0,501	0,330
N_E_S_at [1400,1600] RPM	0,387	0,234
N_E_S_at [1600,1800] RPM	0,204	0,202
N_E_S_at [1800,2000] RPM	0,104	0,088
N_E_S_at [2000,2200] RPM	0	0
N_E_S_at [2200,2400] RPM	0	0
N_E_S_at [2400,2600] RPM	0	0
N_E_S_at [2600,2800] RPM	0	0

N_E_S_at [2800,3000] RPM	0	0
N_E_S_at [3000] RPM	0	0

Description: Number of Engine Shutdown (N_E_S) at [,] given engine speed intervals, fuel idle (I) 1 & 2 represents fuel consumption in that engine mode

According to Table 8, the number of engine shutdown at different engine speed ranges [800, 1400] have a stronger correlation to the response variable in contrast to other predictors in this field. The number of engine shutdown at the engine speed higher than 2000 RPM is almost zero. Zero correlation coefficient in this context doesn't mean that the predictor at 2000 RPM has a lower impact on the response variable. Notwithstanding, the correlation coefficient is zero because of the low variance among those variables. However, the correlation coefficient at higher engine speed has a stronger correlation to the fuel used at idle 1 and the correlation coefficient at lower engine speed has a stronger correlation to the fuel used at idle 2. The reason behind this incident is that idle 1 represents idling at higher engine speed which is causal to the response variable as well. Hence, the engine's operation at higher engine speed is using higher fuel.

Table 9 shows the correlation coefficient for the duration of the engine in seconds before the engine shut down at different engine speeds corresponding to the response variable. The predictors in this field are appearing poor due to poor operator behavior.

Table 9 Correlation coefficient among fuel used at idle mode versus duration of engine before shutdown at different engine speed

	Fuel used I2	**Fuel used I1**
Fuel used I2	1,000	
Fuel used I1	0,451	1,000
T_ [3]s_at [600] RPM	0,068	0,258
T_ [3]s_at [600,800] RPM	0,488	0,279
T_ [3]s_at [800,1000] RPM	0,493	0,321
T_ [3,10] s_at [600] RPM	0,107	0,103
T_ [3,10] s_at [600,800] RPM	0,301	0,178
T_ [3,10] s_at [800,1000] RPM	0,455	0,457
T_ [10,60] s_at [600] RPM	0,063	0,049
T_ [10,60] s_at [600,800] RPM	0,312	0,177
T_ [10,60] s_at [800,1000] RPM	0,304	0,470
T_ [60,180] s_at [600] RPM	0,075	0,075
T_ [60,180] s_at [600,800] RPM	0,234	0,165
T_ [60,180] s_at [800,1000] RPM	0,224	0,444
T_ [180] s_at [600] RPM	0,016	-0,010
T_ [180] s_at [600,800] RPM	0,242	0,217
T_ [180] s_at [800,1000] RPM	0,319	0,446

Description: Time (T) at duration interval [] in seconds before the engine shutdowns at [] engine speed interval

The majority of the predictors in Table 9 have a correlation coefficient lower than 0.5 which indicates that the contribution of the predictors in this field has low impact on the response variables. However, by observing the table from top to bottom, it will be understood that at some place (see magenta colored cells), the magnitude of the positive correlation coefficient at idle 2 is higher than fuel used in idle 1 mode. By considering this fact, at that specific engine speed [800,1000] RPM, there is no causal relation among those variables because idle 2 presents engine speed at [0,800] RPM.

Table 10 shows the correlation coefficient of air conditioner system corresponding to fuel used at different idle mode.

Table 10 Correlation coefficient within idle duration and AC system at auto versus manual mode

	Fuel I2	Fuel I1
Fuel I2	1,000	
Fuel I1	0,451	1,000
AC_Aut_mode	0,516	0,616
AC_Man_mode	0,360	0,471

Description: Air conditioner (AC) at Auto mode & Manual mode

Table 10 shows that the correlation coefficient of the AC system at auto mode has a higher magnitude compared to the other manual mode. However, duration of AC system is been estimated based on Equation 14 and there is a poor knowledge about the how the system is built to explain if there is a causal relation between fuel used and duration of AC system or not. One theory might be that, the operator keeps the engine at higher engine speed to keep the cabin at desired temperature.

4.1.2 Correlation coefficient matrix among idle as a single variable

In this section, Equation 11 has been applied and fuel used at idle mode has been calculated as a single variable. Table 11 shows the correlation coefficient concerning hydraulic oil temperature and fuel used at the different temperature range in degree Celsius.

Table 11 Correlation coefficient among idle fuel used and hydraulic oil temperature

	Fuel used idle
Fuel used idle	1,000
H. T [-20, -10]	0,120
H.T [-10,0]	0,241
H.T [0,30]	0,716
H.T [30,40]	0,779
H.T [40,60]	0,814
H.T [60,70]	0,548
H.T [70,80]	0,258

Description: Hydraulic oil Temperature (H.T) in degree Celsius at [,] given temperature interval

According to Table 11, the strength of the correlation coefficient at higher oil temperature is higher compared to lower temperature, especially at the temperature range of [0,40] degree Celsius.

Table 12 shows the correlation coefficient between fuel used and different engine coolant temperature.

Table 12 Correlation coefficient among fuel used during idling and hydraulic oil temperature

	Fuel used idle
Fuel used idle	1
E_C_T [0,75]	0,712
E_C_T [75,85]	0,850
E_C_T [85,95]	0,394
E_C_T [95,105]	0,325
E_C_T [105,108]	0,200
E_C_T [180]	0,198

Description: E_C_T Engine Coolant Temperature in degree Celsius at [,] given temperature interval

According to Table 12, the variables at in the field of engine coolant temperature have a higher correlation coefficient at temperature range [0,85] degree Celsius. However, the duration of the engine coolant temperature has been estimated by Equation 12. As a result of knowledge lack about the nature of the equipment, it is not possible to state that there is causal relation among to the variables in this field as well.

Table 13 shows the correlation coefficient between fuel used at idle mode and duration of engine before shutdowns at different engine speeds.

Table 13 Correlation coefficient within duration of idle condition of the machinery at different RPM

	Fuel used idle
Fuel used idle	1,000
T_ [3]s_at [600] RPM	0,172
T_ [3]s_at [600,800] RPM	0,468
T_ [3]s_at [800,1000] RPM	0,492
T_ [3,10] s_at [600,800] RPM	0,292
T_ [3,10] s_at [800,1000] RPM	0,533
T_ [10,60] s_at [600,800] RPM	0,299
T_ [10,60] s_at [800,1000] RPM	0,436
T_ [60,180] s_at [600,800] RPM	0,240
T_ [60,180] s_at [800,1000] RPM	0,368
T_ [180] s_at [600,800] RPM	0,270
T_ [180] s_at [800,1000] RPM	0,434

Description: Time (T) at duration interval [] in seconds before the engine shutdowns at [] engine speed interval

According to Table 13, the strength of the correlation coefficient at higher engine speed is higher compared to the others. This indicates the contribution of those variables are higher to fuel consumption compared to others too. However, it is reasonable that correlation coefficient of the variables in this filed at higher engine speed contributes more to fuel consumption.

Table 14 shows the correlation coefficient of air conditioner corresponding to fuel used during idling.

Table 14 Correlation coefficient within idle condition versus AC system of the machine

	Fuel used idle
Fuel used idle	1,000
AC_Aut_mode	0,652
AC_Man_mode	0,475

Description: Air conditioner (AC) at auto mode & manual mode during idling

According to Table 14 the magnitude of the correlation coefficient of the equipment's AC system at auto mode is higher compared to manual mode. Duration of AC system corresponds to idling is been estimated by Equation 14.

4.2 Statistical techniques

This section in is conducted to throw light on the statistical techniques used for this work. The section is divided into three subsections. The first subsection presents the result of principal component analysis, the other sections are presenting the result of classic statistics as well as projection techniques.

It should be pointed out that for understanding the level of correctness, no pre-processing stage was applied to the model developing process. Later, different pre-processing methods were employed to develop different models such as ML-R, SVM-R, PC-R and PLS-R. The results will be presented in section 5.

4.2.1 Principal component analysis

In this section, the structure of the predictors has been analyzed through a seven PCs principal component analysis technique. Though, the section contains three subsections where each subsection represents different pre-processing approaches.

4.2.1.1. Pre-processed with mean centering

Figure 16 depicts that the percentage that could be explained with different principal components. The horizontal axis of Figure 16 displays the principal components and the vertical axis is represents the percentage contribution that can get explained by the PCs. According to Figure 16 the first three PCs can explain the majority (almost 99%) of the variance among to the predictors of the variance among to the predict variables. The blue line represents the calibration which indicates the training samples and the red line represents the validation which indicates the test samples.

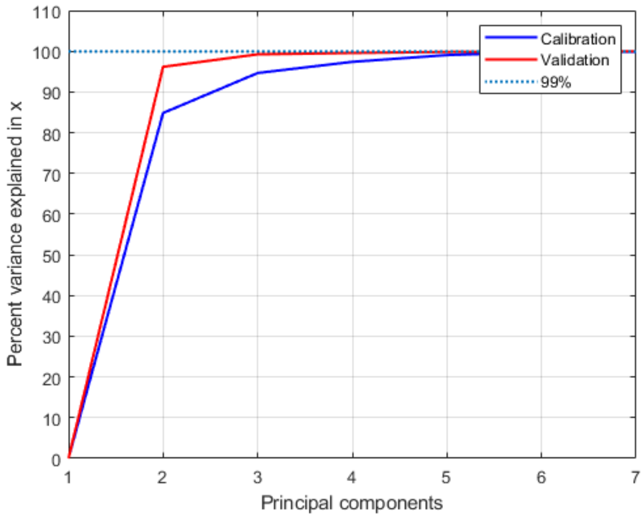


Figure 16 Percent variance explained by different principal components (mean centering)

Figure 17 shows the relationship between the variables along to first versus second principal components. However, the variables that are classified into different categories based on the relationship within the variables. As Figure 17 is showing the variables are located in different quadrant of the coordinate plan, the variables that are located on the same quadrant (same group) contribute similar information to the response variable. This indicates that those variables have the tendency to change in the same manner due to change in numerical value. However, the variables that are located at the invers side of the coordinates system are negative correlated. Variables that are located close to the origin has a lower strength to impact the response variable. Figure 17 contains an inner cycle and outer cycle, the variables that are located inside the inner cycle have a correlation coefficient lower than 0.5. Variables that are located between the inner and outer cycle have a correlation coefficient higher than 0.5 which means that those variables have a stronger impact on the response variable.

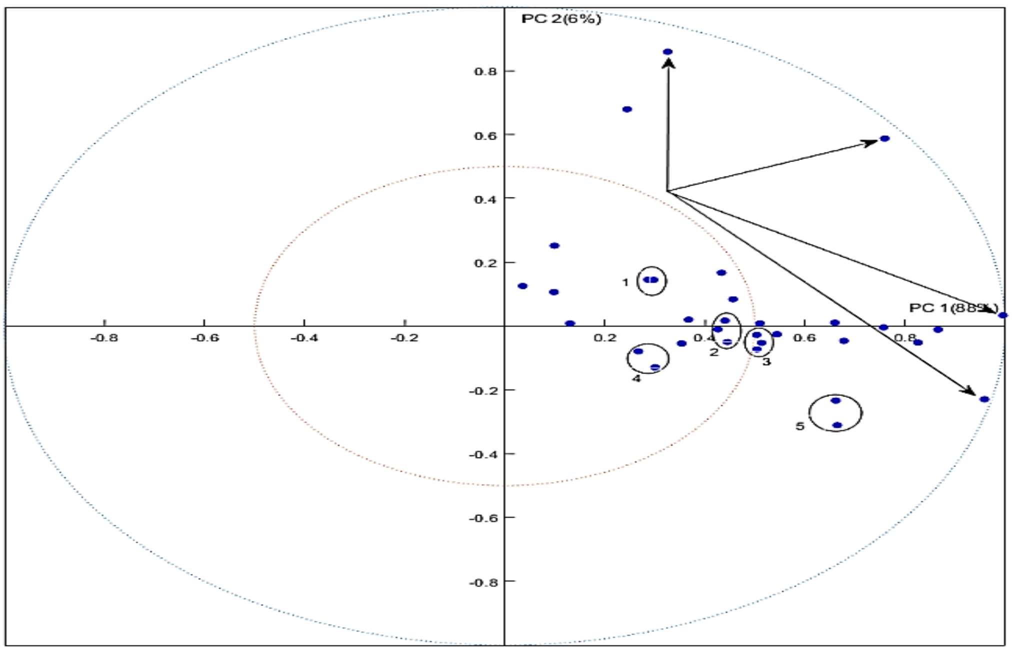


Figure 17 Correlation coefficient among to variables based on mean centering

1. Three seconds duration at [0,600] RPM and number of engine shutdown at [0,600] RPM.
2. [3,10] seconds duration at [600,800] RPM, number of engine shutdown at [1000,1200] RPM and [10,60] second duration at [600,800] RPM
3. Number of engine shutdown at [1200,1400] RPM, 3 seconds duration at engine speed of [600,800] RPM and number of engine shutdown at [600,800] RPM
4. Hydraulic oil temperature at [-10,0] and number of engine shutdown at [1600,1800] RPM
5. Hydraulic oil temperature at temperature span of [0,30] degree Celsius and hydraulic oil temperature at temperature span of [30,40] degree Celsius
6. From up to down: hydraulic oil temperature at [70,80], hydraulic oil temperature at [60,70], engine coolant temperature at [75,85] degree Celsius, hydraulic oil temperature at [40,60]

4.2.1.2. *Pre-processed with mean & max-min scaling feature*

Figure 18 represents how well each PC can explain the variance among predictors. The horizontal axis displays the principal components and the vertical axis represents contribution of each PC to explain the variance among predictors.

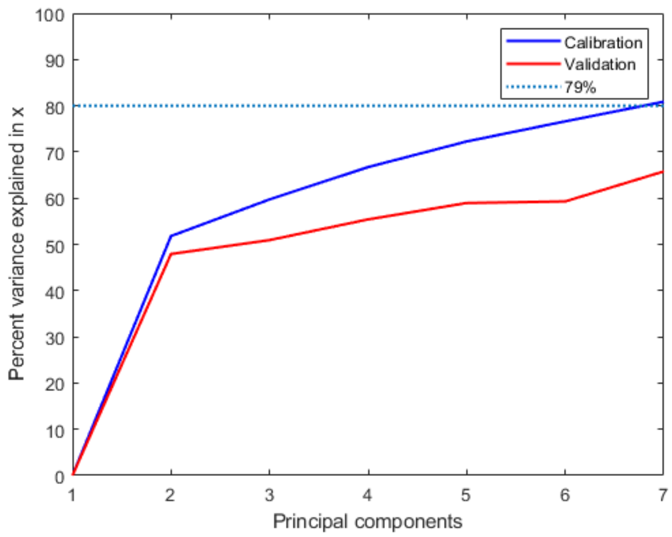


Figure 18 Explained variance of PCA

Figure 18 indicates that these seven principal components can describe only 79% of the predictors and 15% of the variance within the predictors can't be described. However, first principal component can describe 52% of the variance. The blue line indicates that how well the model can explain the train samples and the red line represents the test samples. The gap between calibration and validation is quite small between first and second PC but this gap is increasing along to other PCs. This might be because of high variation among to the predictor.

Figure 19 shows the relationship between 34 variables within first and second principal components. However, the variables that are classified to same group (same quadrant of PCs coordinate plan) contribute similar information to the model. Increase or decrease in numerical value will change the values of those variables in the similar manner. Variables that are located at the inverse side of the coordinate plan have tendency to change in the opposite direction within numerical change in properties.

However, as it mentioned earlier in this paper the outcome of correlation coefficient is [-1, 1]. Figure 19 contains an inner cycle (radius=1) and outer cycle (radius=0.5). The variables that are located between the inner and outer cycle has a significant impact on the response variables. Hence, in this case those variables have a correlation coefficient higher than 0.5. The horizontal axis is representing the first principal components and the vertical axis is representing the second principal component.

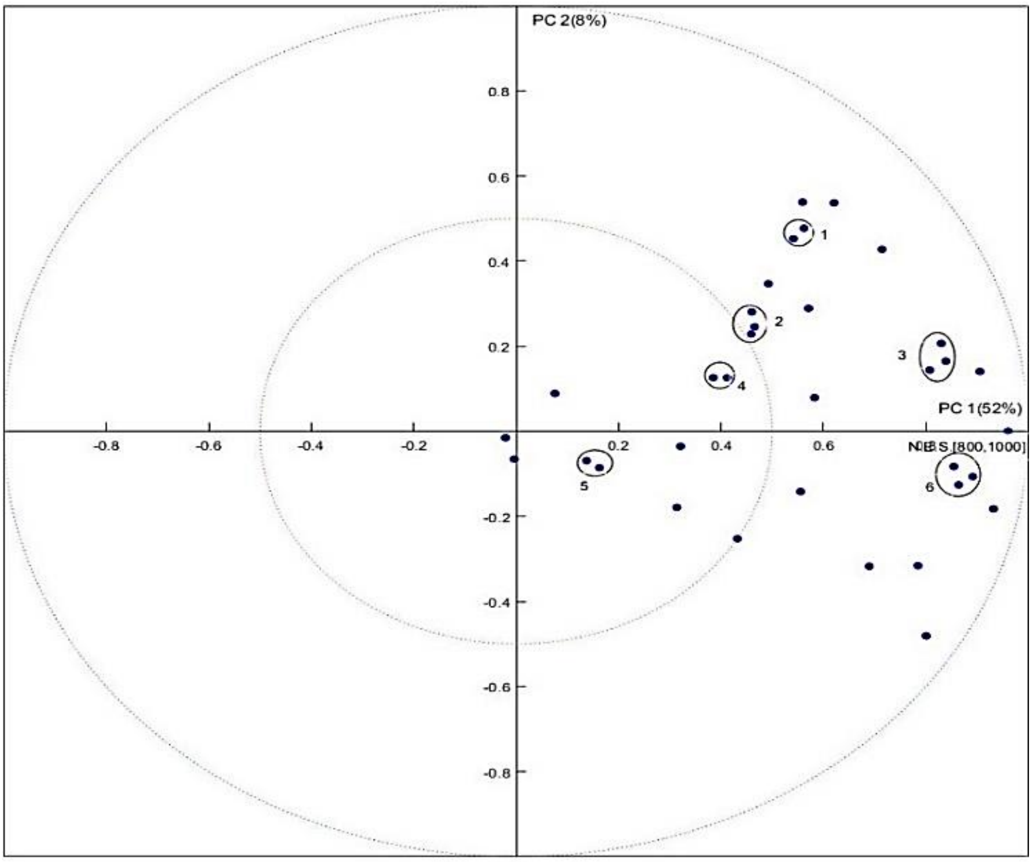


Figure 19 Correlation coefficient among to variables

1. Engine duration before the engine shutdown at the time span of [3,10] & [10,60] seconds at [600,800] RPM range
2. Number of engine shutdown at [0,600] RPM, 3 seconds duration at 600 RPM and Number of engine shutdown at [1200,1400] RPM
3. Hydraulic oil temperature at temperature range of [0, 30] degree Celsius, [3, 10] seconds duration at [800, 1000] RPM and hydraulic oil temperature at [30, 40] degree Celsius
4. Hydraulic oil temperature at [-10, 0] and [60,180] seconds duration at [600,800] RPM
5. Hydraulic oil temperature at [-20, -10] degree Celsius and number of engine shutdown at [1600,1800] RPM
6. Hydraulic oil temperature at [40, 60] degree Celsius and [10, 60] seconds duration at [800, 1000] RPM

4.2.1.3. *Predictor with significant correlation coefficient*

In this section, the predictors based on the information from section 4.1.2 has been chosen and a PCA is being trained and validated. The predictors are pre-processed through mean-centering and the results are presented. Figure 20 presents the percentage variance that could get explained by different PCs. The horizontal axis represents the principal components and the vertical axis is representing the percent variance that can get explained. The figure is demonstrating that the PC can describe 100% of the variance in the predictors. According to Figure 20, only three first PCs are needed to explain the variance of the predictors and the other PCs do not contribute unique information. However, the first principal component can describe almost 91% of the predictors and the rest of the information can be gained with considering the third and second PCs.

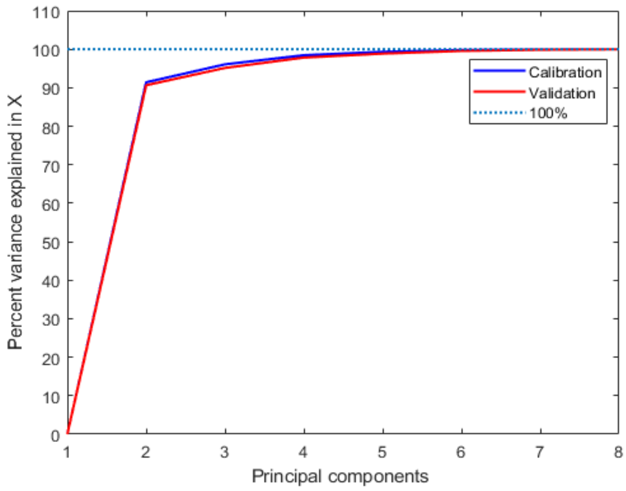


Figure 20 Explained variance in x with different PCs

Figure 20 shows that the majority of the variance among the predictors and the response variable can be explained perfectly. This can be observed by considering the position of calibration versus the validation line. As it is shown in the figure calibration and validation is located on each other. This indicates that the variance can be explained perfectly for both train-test-samples.

Figure 21 represents the relationship among to the predictors corresponding to first PCs. The figure is monitoring a set of scatter points which is labelled with different numbers. Each scatter point is representing a variable, on the other side of the figure every variable is been decreased in detail.

However, as it shown in the figure variables 1 & 2 are located closest to the origin and inside the inner cycle of the figure for PC1, PC2 and PC3. This indicates at those variables have a low strength to impact the target variables. As it shown in the figure some of the variables are located close to the boundary of the outer cycle. This indicates that those variables are impacting the target variable strongly.

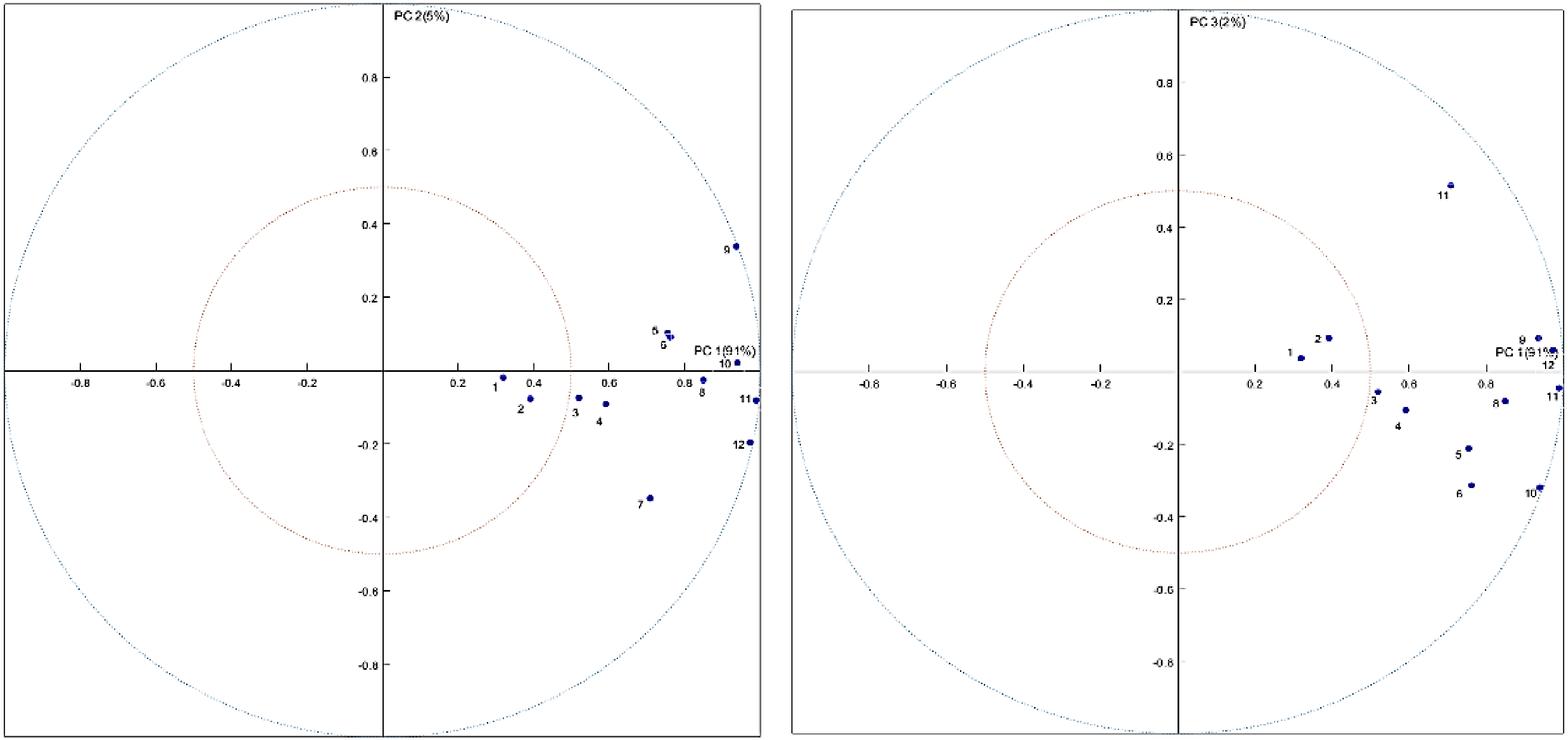


Figure 21 Relationship between the variables along to first, second and third variable

1. Duration of the engine at [60,180] seconds at engine speed of [600, 800]
2. Duration of the engine at [60,80] seconds at engine speed of [600, 800]
3. Duration of the engine at [10,60] seconds at engine speed of [800,1000]
4. Duration of the engine at [3,10] seconds at engine speed of [800,1000]
5. Duration of hydraulic oil temperature at [0,30] degree Celsius
6. Duration of hydraulic oil temperature at [30,70] degree Celsius
7. Duration of hydraulic oil temperature at [60,70] degree Celsius
8. Duration of engine coolant temperature at temperature range [0,75] degree Celsius
9. Duration of air conditioner in auto mode
10. Duration of hydraulic oil temperature at [40,60] degree Celsius
11. Duration of engine coolant temperature at temperature range [75,85] degree Celsius

4.2.2 Classic statistic

Multivariate linear regression is known as a classic statistic model. This section presents the different ML-R models which is been developed through different pre-processing approaches.

4.2.2.1. Pre-processed with mean centering

Figure 22 shows the accuracy of the ML-R which is been pre-processed through mean-centering. The horizontal axis is showing the reference samples and the vertical axis is representing the samples that are been predicted by the model. According to Figure 22 the majority of the test samples are located close to the prefect prediction line which indicates that the model can predict the response variable properly.

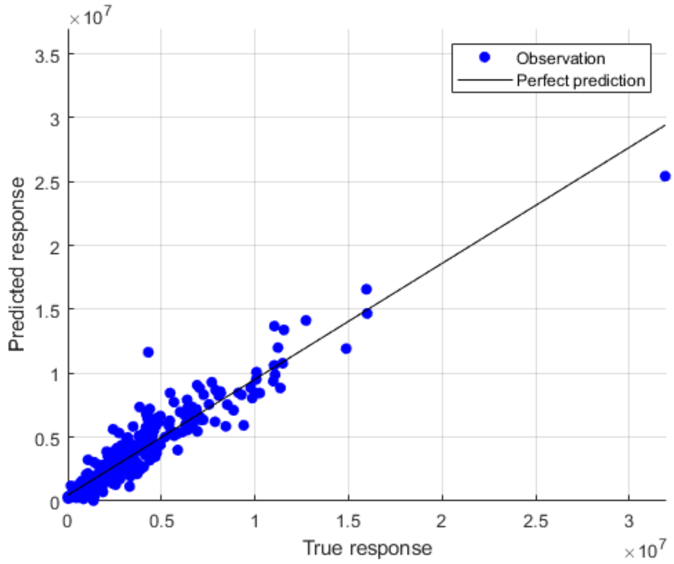


Figure 22 ML-R accuracy based on mean centering

Table 15 is showing the external validation parameters of the ML-R. The R_2 indicates that the ML-R model can explain almost 88% of the variance in fuel consumption per seconds during the idling. This is a reasonable prediction proportion for this kind of data. RMSE is the other validation parameter which is representing the distance between the predicted samples and the reference sample. However, an accurate model has a RMSE zero or close to zero but in this case its far from zero.

Table 15 Validation parameters of PC-R based on mean centering

	RMSE	R^2
Calibration	1.307e+05	0.87
Validation	6.688e+05	0.88

4.2.2.2. Pre-processed with mean centering & max-min scaling feature

Figure 23 displays the goodness of predication of MVR which is been pre-processed through mean centering and max-min scaling feature. The horizontal axis shows the reference value of the samples and the vertical axis shows the value that is been predicted by the model.

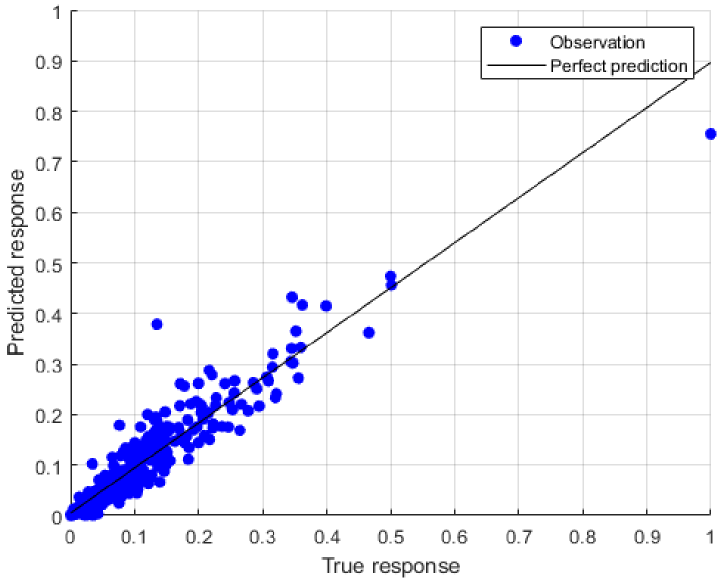


Figure 23 ML-R accuracy based on mean centering and MAX-MIN scaling feature

According to Figure 23, the variance of the samples close to the origin is higher, and the majority of the samples close to the origin is been predicted more properly.

Table 16 presents the external validation parameters of the model. The calibration has a lower accuracy compared to validation parameters. This might be because of higher variance among the samples but the validation parameters are indicating that the model can predict the response variable properly. However, in this case the RMSE has been improved remarkably.

Table 16 Validation factors of ML-R based on mean centering and max-min scaling feature

	RMSE	R^2
Calibration	0.0411	0.863
Validation	0.0200	0.98

4.2.2.3. Predictor with significant correlation coefficient

Figure 24 is representing ML-R which been pre-processed through mean centering. The predictor in this case has been chosen based on its correlation coefficient (see section 4.1) in order to reduce the predictors that do not contribute to the model. The vertical axis is representing the samples which is been predicted by the model and the horizontal axis is representing the reference samples.

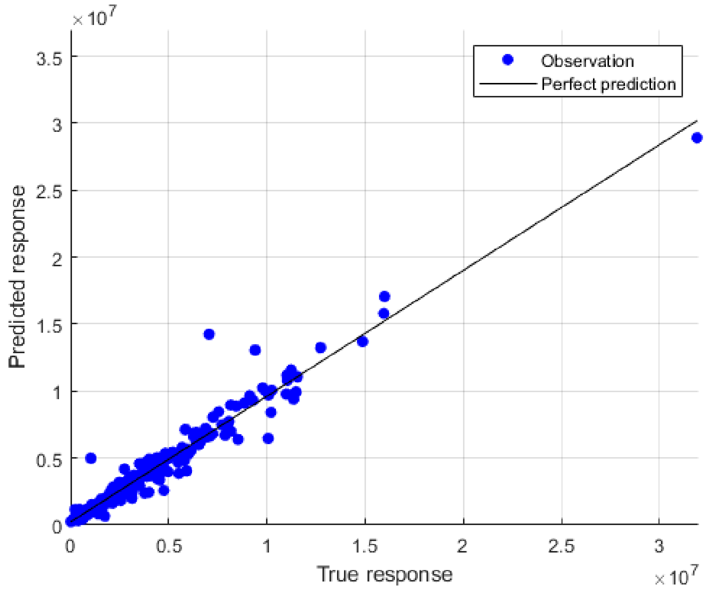


Figure 24 ML-R accuracy based on mean centering

Table 17 presents the external validation parameters of the ML-R model. The R_2 indicates that the trained ML-R model can describe almost 99% of the variance in the target variable. Despite this high accuracy in the R_2, the RMSE is indicating that there is a high residual along to predict versus response variables.

Table 17 Validation factors of ML-R based on mean centering

	RMSE	R^2
Calibration	8.051e+05	0.94
Validation	5.969e+05	0.99

4.2.3 Projection technique

This section is going to present the results of some common projection techniques, which is been adopted on the data set to predict the response variables. However, the models that will be presented are SVM-R, PLS-R, and PC-R.

4.2.3.1. Partial least square regression

In this section, PLS-R with seven factors has been trained on the predictors to predict the response variable. The models are been developed through different pre-processing approaches and the result is presented.

4.2.3.2. Pre-processing with mean centering

Figure 25 displays the goodness of contribution to how well each factor is able to explain the variance between the response variable. According to Figure 25, almost 83% of the variance among response variable can be explained by the first three factors.

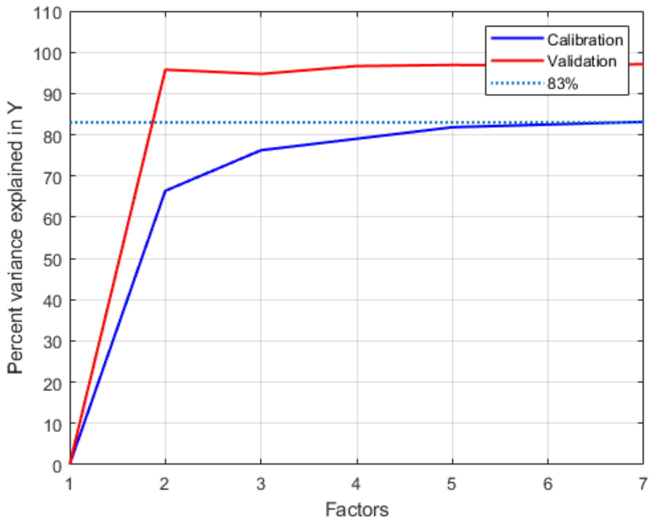


Figure 25 Percent variance explained by different factors of PLS-R (mean centering)

The blue line represents the calibration and stands for training samples and the red line represents the validation and stands for test samples. Figure 25 is showing that the majority of the factors can't explain the variance of the response variable. This can be observed from the slope of the calibration versus validation line where its zero.

Figure 26 displays goodness of prediction of PLS-R through mean centering. The horizontal axis of the figure shows the reference samples and the vertical axis shows the samples that is predicted by the model.

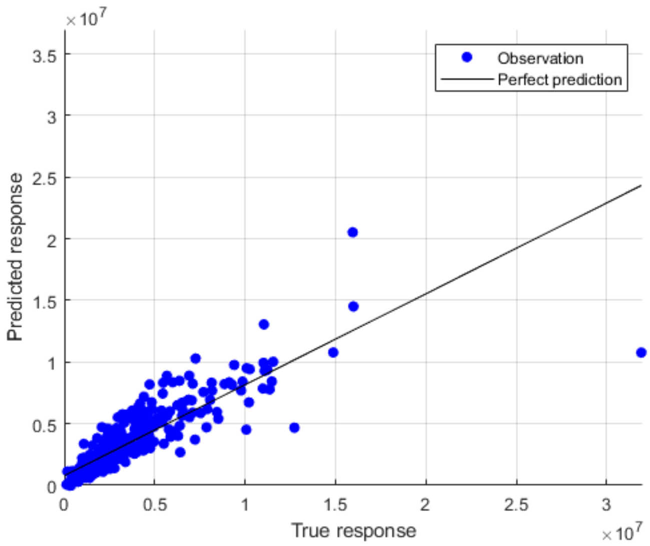


Figure 26 PLS-R accuracy based on mean centering

Table 18 displays the external validation parameters of the PLS-R. The R2 indicates that the PLS-R model can explain almost 66% of the variance of the response variable for calibration and 95 % during the external validation. However, the R_2 for the external validation is quite high but the RMSE for this kind of pre-processing is very poor.

Table 18 Validation parameters of PLS-R based on mean centering

	RMSE	R^2
Calibration	1.922 e+05	0.66
Validation	6.863 e+05	0.95

Figure 27 displays the regression coefficient of the variables corresponding to different factors. The horizontal axis is representing the predictor and the vertical axis is representing the regression coefficient. However, as mentioned earlier with considering Figure 25, only the first three factors can explain the variance of the response variable.

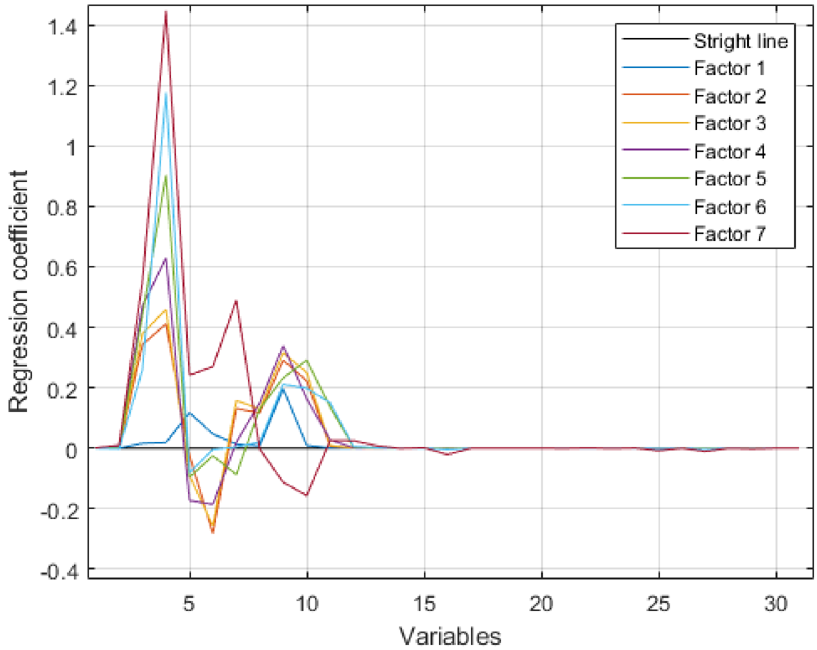


Figure 27 PLS-R important variables based on mean centering

Figure 27 depicts that the majority of the variables are located at the horizontal axis which indicates that those variables have a regression coefficient of almost zero. Barely the first 13 variables have a non-zero regression coefficient.

4.2.3.3. Pre-processed with mean centering & max-min scaling feature

Figure 28 presents an overview of the model's factors to outline the contribution of each factor that explains the variance of the response variable. The blue line indicates the calibration which is representing the training samples and the red line indicate the calibration which is representing the test samples. The horizontal axis is representing the factors and the vertical axis is representing the percentage that can be explained by different factors.

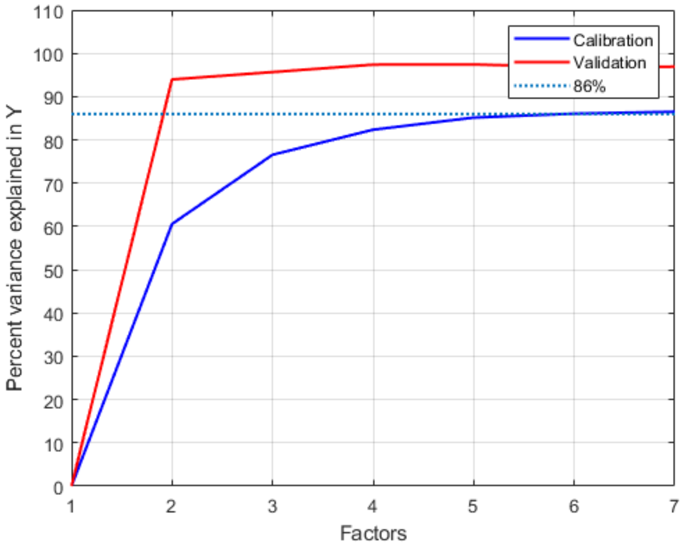


Figure 28 Explained variance of PLS-R based on mean centering and max-min scaling

Figure 28 shows that, all seven factors can explain at least 86% of the train samples. There is a gap between calibration and validation which might be because of high variance between train and test samples.

Figure 29 represents the accuracy of the PLS-R model based on this kind of pre-processing. The vertical axis is representing the predicted variables that is been executed by the model and the horizontal axis is representing the reference value of those variables.

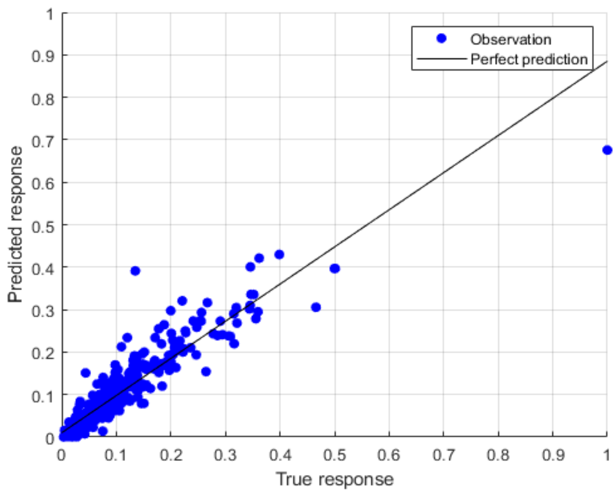


Figure 29 PLS-R accuracy based on mean centering and MAX-MIN scaling feature

According to Figure 29, the majority of the samples are located on the span of [0 - 0.6] of both axes. However, the majority of the observations are located close to the perfect prediction line which indicates the model can predict the response variable properly.

Table 19 presents the external validation parameters of the model. The calibration has a higher accuracy compared to validation parameters. This might be because of the fact that more samples were used for training purposes and less samples were used for testing purpose.

Table 19 Validation factors of PLS based on mean centering and max-min scaling feature

	RMSE	R^2
Calibration	0.026	0.8975
Validation	0.029	0.8680

Figure 30 displays the important variables of the PLS-R model regarding to the different factors. The horizontal axis showing the predict variables and the vertical axis is showing the regression coefficients that corresponds to predictor.

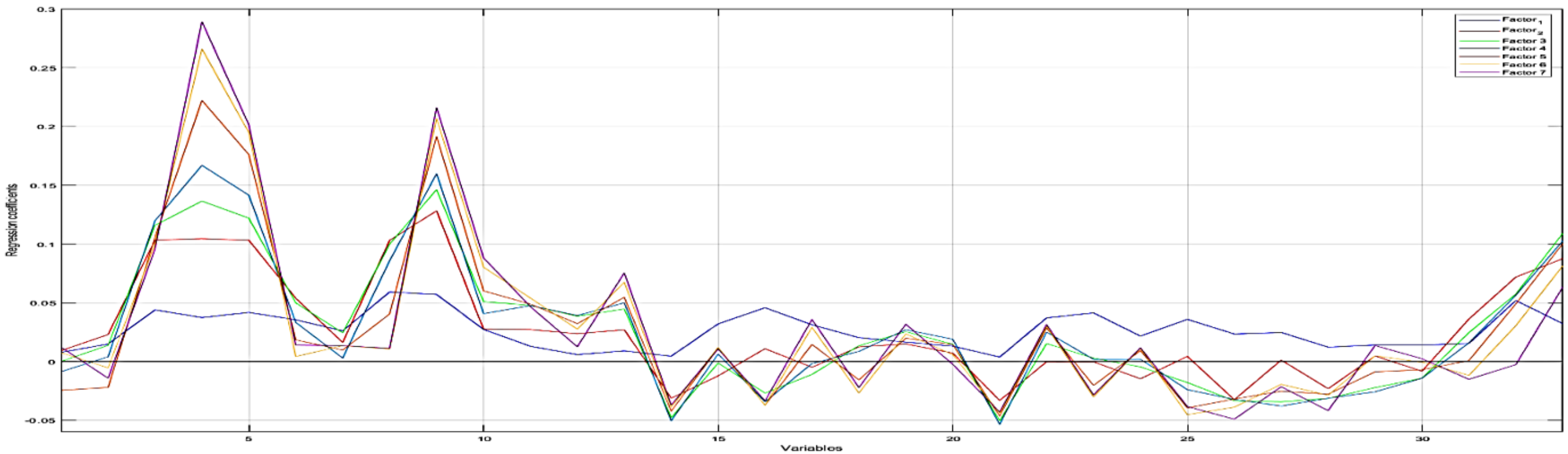


Figure 30 PLS-R important variables based on mean centering and max-min scaling feature. Y-axis and X-axis represents regression coefficient and variables respectively

4.2.3.4. Predictor with significant correlation coefficient

Figure 31 is showing how each factor can describe the variance of the target variable. The dashed line indicates that all of the factors can describe 95% of the variance in the response variable. However, according to Figure 31, the first factor can explain the change in response variable by almost 85% for calibration and 98% for the validation. The second factor is the next best factor and the slope of the other factors is almost zero which indicates that those factors do not contribute to the explanation of the model

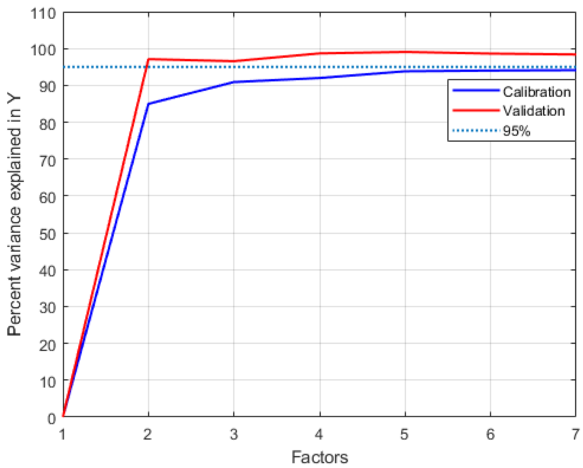


Figure 31 Explained variance of PLS-R based on mean centering

Figure 32 is displaying the goodness of prediction in the PLS-R model. However, by studying the figure, it is observable that residual between reference samples and predict samples is very small. Hence, the majority of the predicted variables are located very close to the perfect prediction line but the model contains some level of noise as well. This can be observed from the samples that are located far from the perfect prediction line.

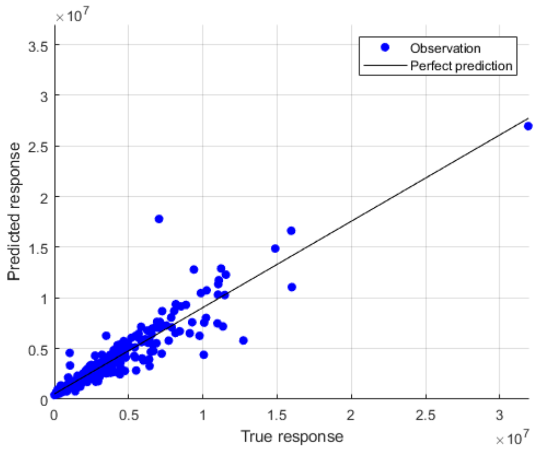


Figure 32 PLS-R accuracy based on mean centering

Table 20 is representing the external validation parameters of the PLS-R model. However, R_2 is indicating that the model can describe 97% of the variance of the response variable. In contrast, RMSE is which the other external validation factor of the model is showing very high value.

Table 20 Validation factors of PLS-R based on mean centering

	RMSE	R^2
Calibration	1.284e+05	0.84
Validation	7.622+05	0.97

Figure 33 presents the weight of the different variables that corresponds to different factors. However, as discussed earlier in this section, only first and second factors can describe the variance in the response variable and the other factors are not contributing so much information to the model.

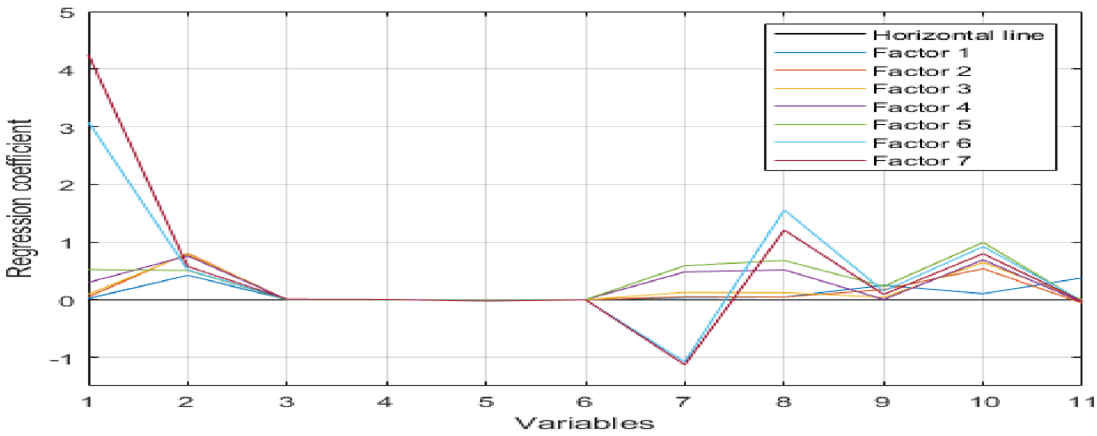


Figure 33 PLS-R important variables based on mean centering

Furthermore, Figure 33 shows that the variables in the range of [3-6] are located at the horizontal axis. Those variables can be observed from the figure (see variables 1-4), that have a low correlation coefficient correspond to target variable.

4.2.3.5. *Principal components regression*

In this section, a PC-R model with seven principal components is adopted on the predictor as a function of fuel consumption during the idling. The models are developed with applying different pre-processing approaches and the result is presented.

4.2.3.6. *Pre-processed with mean centering*

Figure 34 displays the percentage of variance in the response variable that could be explained by different principal components. Though, the figure shows that the first PC (blue one) can explain the majority (almost 68%) of the variance in response variable. The second PC is parallel to the horizontal axis that indicates that this particular PC does not contribute to explanation of the variance on the response variable.

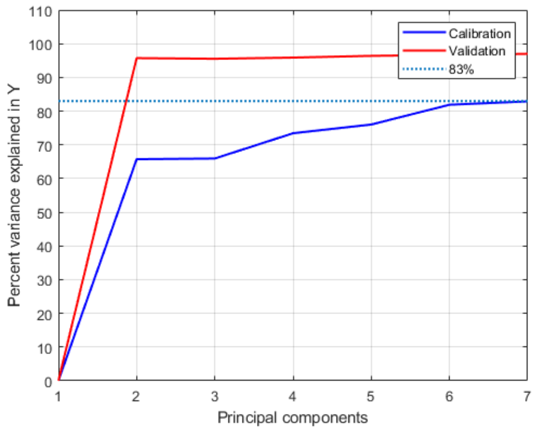


Figure 34 Principal component that explain the variance of the predictors

As it is shown in Figure 34, there is a big gap between result of calibration and validation, this might be because of high variance among the predictors. However, the first PC of the calibration sample can almost explain 95% of the variance among the target variable.

Figure 35 displays the goodness in predication of PC-R. The horizontal axis is showing the reference value of the samples and the vertical axis is showing the predicted samples which are estimated by the model.

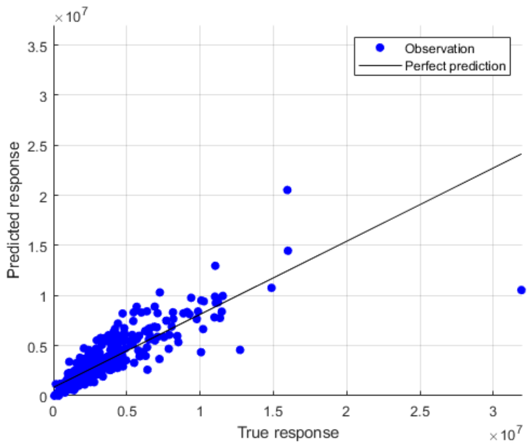


Figure 35 PC-R accuracy based on mean centering

Table 21 displays the external validation parameters of the regression model. The R_2 indicates that the PC-R model can explain almost 65% of the variance in response variable for calibrating, and 95% for the validation samples. RMSE which represents the distance between the predicted and reference samples indicates a big noise among the samples. RMSE should be as close as possible to zero to indicate that the model has a high accuracy.

Table 21 Validation parameters of PC-R based on mean centering

	RMSE	R^2
Calibration	1.4307e+06	0.657
Validation	6.0472e+05	0.957

4.2.3.7. *Pre-processed with mean centering & max-min scaling feature*

Figure 36 displays how each principle component can explain the variance of the response variable. The first sex PCs of the calibrating samples can describe 75% of the target variable. In contrast, validation samples show a proper result due to higher accuracy.

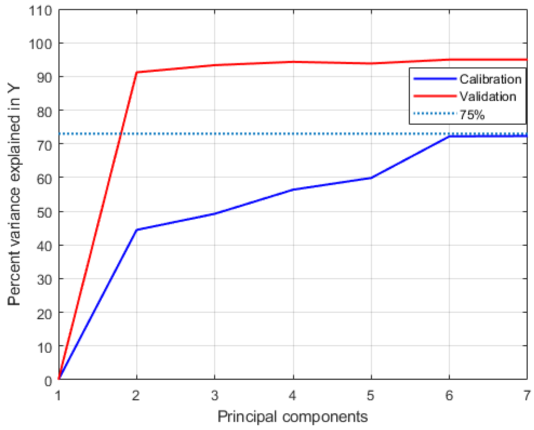


Figure 36 PC-R principal components based on mean centering and max-min scaling feature

Figure 37 shows the goodness of the prediction of PC-R. The horizontal axis shows the reference samples and the vertical axis shows the predicted samples. The black line indicates a perfect prediction, samples which are located close to the regression line have been predicted properly by the model. However, it is observable from the figure that the model contains some degree of noise due to high residual between the perfect prediction line and samples.

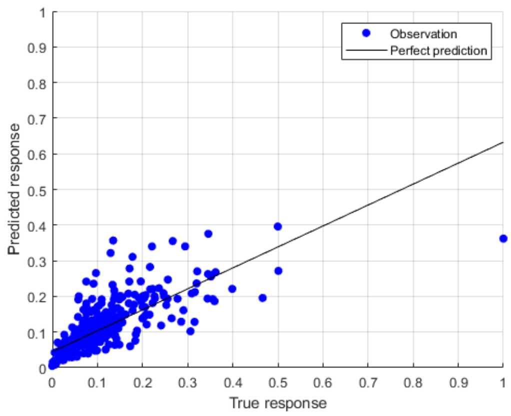


Figure 37 PC-R accuracy based on mean centering and max-min scaling feature

Table 22 is represents the external validation parameters of the model. In this case the RMSE value is improved remarkably by pre-processing the data set. The validation parameters show that the model contains a higher level of the accuracy as R_2 is equal to 93% and the RMSE has become much smaller compared to the non-scaled case.

Table 22 Validation factors of PC-R based on mean centering and max-min scaling feature

	RMSE	R^2
Calibration	0.0756	0.492
Validation	0.0269	0.933

4.2.3.8. *Predictor with significant correlation coefficient*

Figure 38 displays the PC's contribution to describe the variance of the response variable and the dashed line indicates that the model can describe almost 95% of the variance in the response variable.

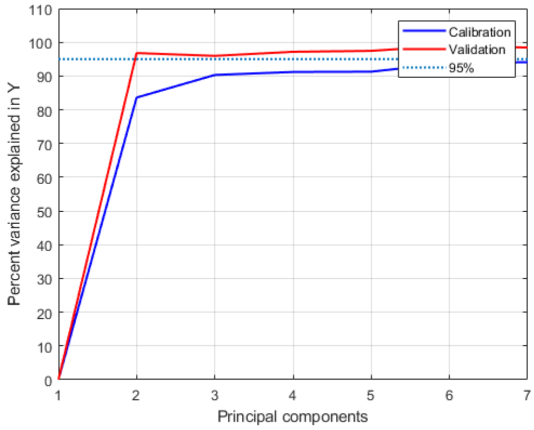


Figure 38 PC-R accuracy based on mean centering

Figure 39 shows the PC-R's goodness in prediction of the model, the majority of the samples have been predicted properly because they are located close to the perfect prediction line. The model contains some level of noise as well. This can be observed from the long distance between some samples to the prediction line.

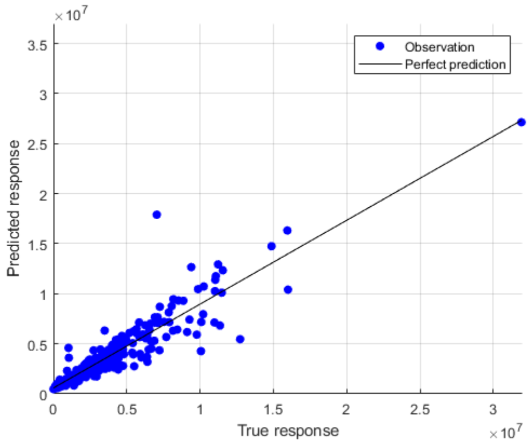


Figure 39 PC-R accuracy based on mean centering

Table 23 presents the external validation parameters of the model. The R_2 is indicating that the model can describe almost 96% of the variance in the response variable regarding to the test samples. However, RMSE which indicates the distance between the predicted and reference samples is very poor and needs to be improved.

Table 23 Validation factors of PC-R based on mean centering

	RMSE	R^2
Calibration	1.063e+05	0.83
Validation	8.210e+05	0.968

4.2.3.9. *Support vector machine regression*

In this section, SVM has been trained on the predictors as a function of fuel consumption per seconds during the idling. Different models are trained and evaluated based on different pre-processing approaches.

4.2.3.10. *Pre-processed with mean centering*

Figure 40 displays the goodness in predication of the SVM-R based on mean centering in the pre-processing stage. The horizontal axis of the figure shows the reference samples and the vertical axis shows the predicted variables. In accordance to Figure 40, the model can predict the test samples properly because test samples are located close to the perfect prediction line.

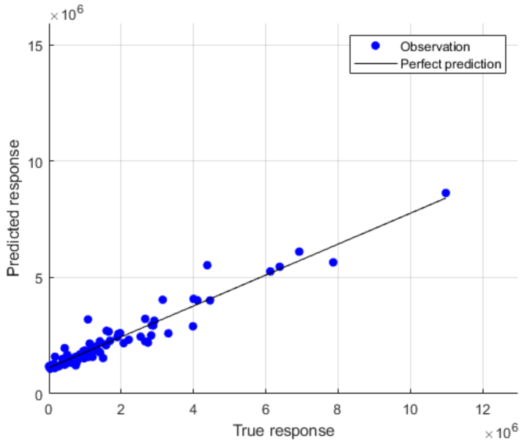


Figure 40 SVM accuracy based on mean centering

Table 24 shows the external validation parameters of the SVM-R model. The R_2 indicates that the PC-R model can explain almost 74% of the variance in response variable for test samples and 84% of the variance of the training samples. However, the RMSE values both for the validation and calibration indicates that the models contain high degree of noise. In contrast the R_2 values indicates that the model performance is sufficient in this field.

Table 24 Validation parameters of PC-R based on mean centering

	RMSE	R^2
Calibration	1.575e+05	0.846
Validation	1.813e+05	0.739

4.2.3.11. *Pre-processed with mean centering & max-min scaling feature*

Figure 41 presents the goodness of predication of the model which is mean centered and normalized through max-min scaling feature. The horizontal axis displays the reference samples and the vertical axis shows the output which is predicted by the model.

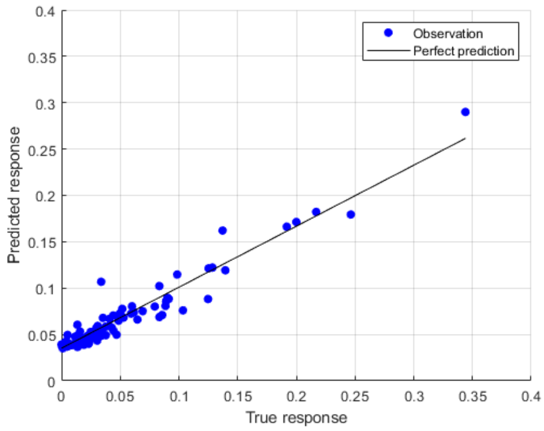


Figure 41 SVM accuracy based on mean centering and max-min scaling feature in pre-processing

Table 25 presents the external validation parameters of the model. According to Table 25 RMSE values have been improved remarkably compared to the previous case (non-scaling). However, the R_2 is the same as the previous case.

Table 25 Validation factors of SVM based on mean centering and max-min scaling feature

	RMSE	R^2
Calibration	0.0458	0.8498
Validation	0.0584	0.7316

4.2.3.12. Predictor with significant correlation coefficient

Figure 42 displays the goodness of predication in the model by a demonstration of the reference samples versus predicted samples. According to the figure, the model is well trained and can predict the test samples properly.

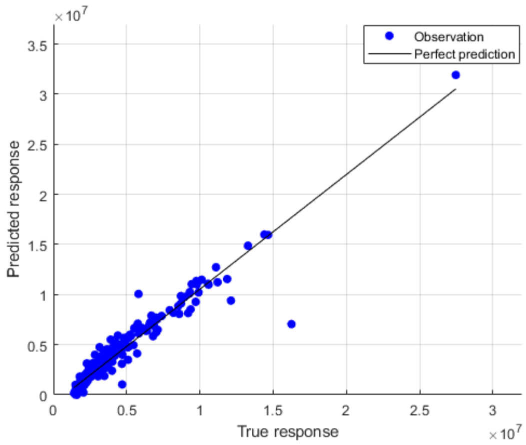


Figure 42 SVM-R accuracy of the predictor with high correlation coefficient

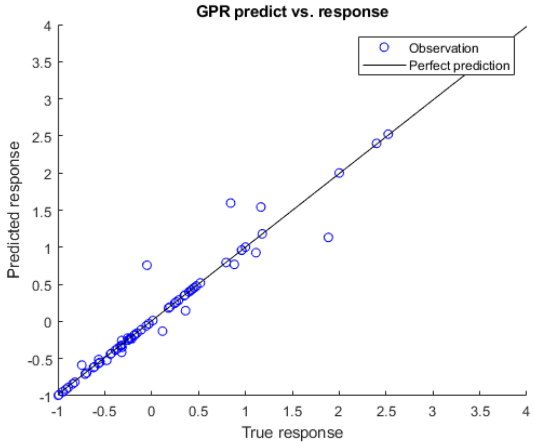
Table 26 represents the external validation parameters of the SVM-R. R_2 is indicating that the model can describe almost 93% of the variance in the response variable regarding to train samples and 81% of the variance regarding to test samples. In spite of high R_2 values, RMSE indicates that the model contains high degree of noise.

Table 26 Validation factors of SVM-R based on mean centering

	RMSE	R^2
Calibration	1.064e+06	0.93
Validation	1.684e+06	0.81

4.2.4 Gaussian process regression

This section presents the result of Gaussian Process Regression (GP-R) which is based on exponential method. GP-R is a machine learning technique that can be trained through four methods: Rational Quadratic GP-R, Squared Exponential GP-R, Matern 5/2 GP-R and Exponential GP-R. However, among those techniques of GP-R, the exponential was the most accurate algorithm. Figure 43 are displaying the accuracy of the GP-R model based on different pre-processing approaches, the figure in the left-hand side is been mean centered and normalized before the model is been trained. Nevertheless, the right-hand side is displaying the GP-R model which is been pre-processed only through mean centered.



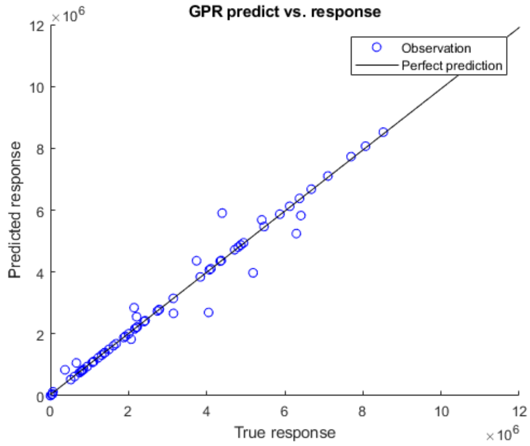


Figure 43 Left: GP-R accuracy mean centered and normalized. Right: GP-R accuracy mean centered

Table 27 represents the external validation parameters of the GP-R model, the upper part of the table shows the result of the GP-R model which is pre-processed through mean centering. However, the lower part of the model presents the GP-R model which is pre-processed through mean centering and normalization (max-min scaling).

Table 27 Validation parameters of GP-R

	RMSE	R_2
Calibration	1.2648e+06	0.82
Validation	3.4930e+05	0.98
Mean centered and normalized		
Calibration	0.51	0.76
Validation	0.17	0.98

4.3 Artificial neural network

In the previous sections, different statistical models (classics and projection) was trained to predict the response variable. In this section, variables will be trained to perform the prediction through a feedforward neural network with 10 hidden layers. Figure 44 represents a simple layer neural network.

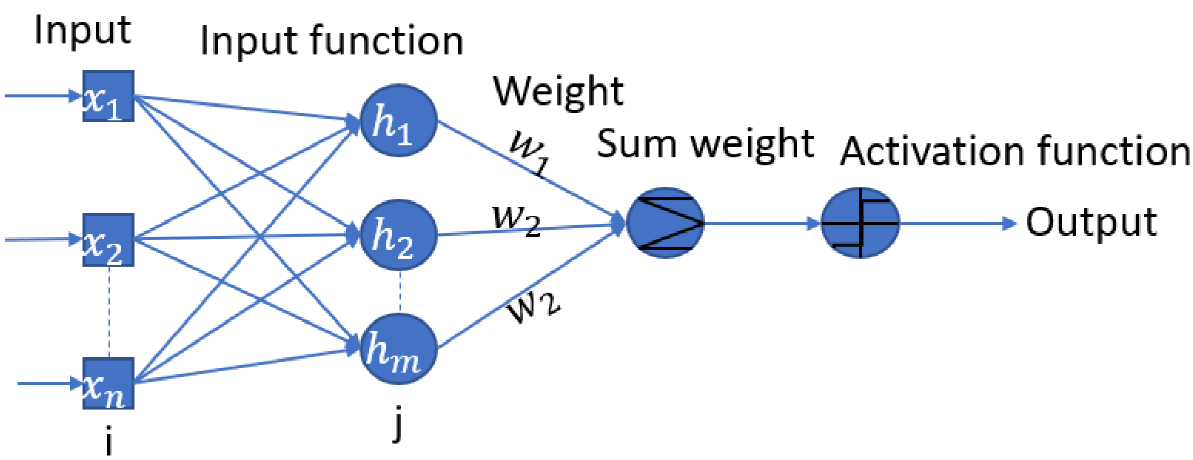


Figure 44 Terminology of simple layer neural network

Information exchange among different nodes can be expressed by the following equation:

$$h(x_i)_1 = (w_1 \times x_1) + (w_2 \times x_2) + (w_i \times x_i) \qquad \text{Equation 20}$$

Where,

W: weight and X: predictor

Activation function which gives the output of the network is expressed by the following equation:

$$y = \varphi(h_1) \qquad \text{Equation 21}$$

Error at each node can be expressed by the following equation:

$$e_i \triangleq d_i - y_i \qquad \text{Equation 22}$$

Where,

e_i: $error$

y_i: $predicted\ output$

d_i: $reference\ output$

After determination of the error at each node, the weight of each variable will be modified by the following expression:

$$w_{i,j} \leftarrow w_{i,j} + \alpha\, e_i x_i \qquad \text{Equation 23}$$

Where,

α: $learning\ rate$

In order to optimize the recalculation of the new weight, the Sigmoid function will be used which is expressed by following equation:

$$\varphi(x) = \frac{1}{1 + e^{-x}} \qquad \text{Equation 24}$$

The second optimization step is to modify Equation 23 by replacing error with delta which is expressed by following equation:

$$w_{i,j} \leftarrow w_{i,j} + \alpha\, \delta_i x_i \qquad \text{Equation 25}$$

Delta is the derivate of the activation function with respect to x, and is expressed by following equation:

$$\delta_i = \varphi' \frac{d}{dx}(h_i) \quad \text{(Activation function)} \qquad \text{Equation 26}$$

Simplification of equation 26 gives:

$$\delta_i = \varphi(x)(1 - \varphi(x)) \times e_i \qquad \text{Equation 27}$$

4.4 Long short-term memory network

In this section, long short-term memory network (LSTMN) method with 100 hidden layers is used to train the predictors to forecast the response variable. Table 28 shows the properties of the LSTMN model.

Table 28 The properties of LSTMN model

GradientDecayFactor	0.0100
SquaredGradientDecayFactor	0.9990
Epsilon	1.0000e-08
InitialLearnRate	1.0000e-04
LearnRateScheduleSettings	[1×1 struct]
L2Regularization	1.0000e-04
GradientThresholdMethod	l2norm
Gradient Threshold	Inf
MaxEpochs	1000
Minibatch Size	128
Verbose	1
Verbose Frequency	50
Validation Frequency	50
Shuffle	'once'
Execution Environment	'auto'
Sequence Length	'longest'
SequencePaddingValue	0
DispatchInBackground	0

Table 29 presents the performance and internal validation parameters of the LSTMN model.

Table 29 Performance and internal validation of LSTMN

Epoch	Iteration	Time Elapsed (hh:mm:ss)	Mini-batch RMSE	Mini-batch Loss	Base Learning Rate
1	1	00:00:00	4.61e+06	1.1e+13	1.0000e-04
50	50	00:00:04	4.61e+06	1.1e+13	1.0000e-04
100	100	00:00:08	4.61e+06	1.1e+13	1.0000e-04
150	150	00:00:12	4.61e+06	1.1e+13	1.0000e-04
200	200	00:00:15	4.61e+06	1.1e+13	1.0000e-04
250	250	00:00:19	4.61e+06	1.1e+13	1.0000e-04
300	300	00:00:22	4.61e+06	1.1e+13	1.0000e-04
350	350	00:00:26	4.61e+06	1.1e+13	1.0000e-04
400	400	00:00:30	4.61e+06	1.1e+13	1.0000e-04
450	450	00:00:33	4.61e+06	1.1e+13	1.0000e-04
500	500	00:00:37	4.61e+06	1.1e+13	1.0000e-04
550	550	00:00:41	4.61e+06	1.1e+13	1.0000e-04
600	600	00:00:44	4.61e+06	1.1e+13	1.0000e-04
650	650	00:00:48	4.61e+06	1.1e+13	1.0000e-04
700	700	00:00:52	4.61e+06	1.1e+13	1.0000e-04
750	750	00:00:55	4.61e+06	1.1e+13	1.0000e-04
800	800	00:00:59	4.61e+06	1.1e+13	1.0000e-04
850	850	00:01:03	4.61e+06	1.1e+13	1.0000e-04
900	900	00:01:07	4.61e+06	1.1e+13	1.0000e-04
950	950	00:01:10	4.61e+06	1.1e+13	1.0000e-04
1000	1000	00:01:14	4.61e+06	1.1e+13	1.0000e-04

Figure 45 displays the errors between the true response variables and predicted responses. The horizontal axis shows the test samples and vertical axis shows the error between true response and predicted response variables.

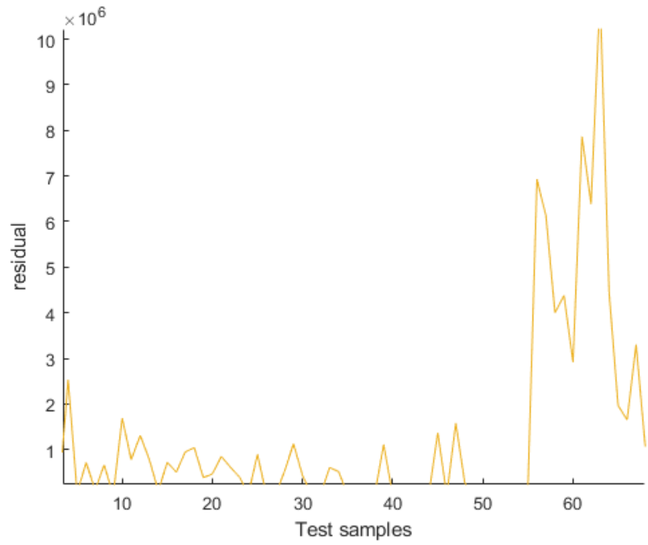


Figure 45 Residual between true response and predicted response

Figure 46 shows the goodness in prediction of the LSTMN model, where each observation indicates a test sample. The horizontal axis of the figure indicates the location of the true responses and the vertical axis shows the response variables which is predicted by the LSTMN.

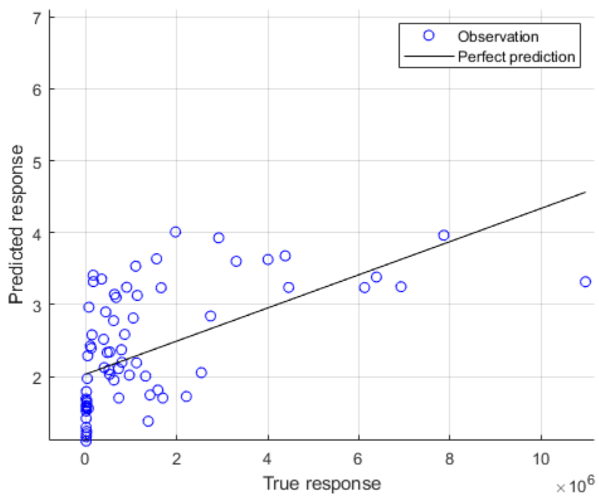


Figure 46 Accuracy of LSTMN model

4.5 Carbon dioxide emission estimation

This section is conducted to estimate the CO_2 emission from the excavator in EMEA region. The CO_2 emission is estimated through two different approaches: fuel used and NONROAD 2008 equation.

4.5.1 Emission estimation based on NONROAD2008

The emission factor for CO_2 is based on the brake-specific fuel consumption (BSFC) (Sandanayake, Zhang, & Setunge, 2019; Fan, 2017). It has been developed by EPA and is based on steady-state engine dynamometer tests in a laboratory environment. The paper is available at "Exhaust and Crankcase emission Factors for Non-road Engine Modeling-Compression-ignition". The equation that expresses the emission factor of CO_2 is as following:

$$EF_{CO_2} = (BSFC \times 453{,}6 - EF_{HC}) \times 0{,}861 \times \left(\frac{44}{12}\right) \qquad \text{Equation 28}$$

Where:

$$EF_{CO_2} \coloneqq Estimation\ factor\ for\ CO_2\ in\ \left[\frac{g}{(hP \times hr)}\right]$$

$$BSFC \coloneqq brake - specific\ fuel\ consumption\ in\ \left[\frac{1b}{(hp \times hr)}\right]$$

$$453.6 \coloneqq the\ conversion\ number\ (from\ pound\ to\ grams)$$

$$EF_{HC} \coloneqq emission\ factor\ for\ HC\ in\ \left[\frac{g}{(hp \times hr)}\right]$$

$$0.861 \coloneqq carbon\ fraction\ for\ diesel\ fuel$$

$$\left(\frac{44}{12}\right) \coloneqq mass\ ratio\ of\ CO2\ molecular\ to\ Carbon\ molecular$$

The following parameters are collected from the "Exhaust and Crankcase emission Factors for Non-road Engine Modeling-Compression-ignition" for EC480:

HC=0.1669

BSFC=0.367

Table 30 displays the CO_2 estimation of 362 excavators which is calculated through the EPA emission factor. However, Table 30 represents the total duration of the excavator in two different idle mode and total generation of CO_2 into the environment.

Table 30 Carbon dioxide emission estimation from EPA emission factor

	IDLE2	IDLE1
Total (hr)	183 859	142 852
Average (hr)	507	394
CO_2 total [ton]	5 695	5 775
CO_2 average [ton]	15,73	15
Total idle emission [ton]	11 470	
Average idle emission [ton]	31	

4.5.2 Emission estimation based on fuel used

Table 31 displays the CO_2 emission of 362 excavators which is been calculated through the fuel used. However, it presents the total fuel consumption of the excavator in both idle modes.

Table 31 carbon dioxide emission estimation during idling from fuel consumption

	Idle 2	Idle 1
Total fuel consumption (liter)	1 142 245	892 852
Average fuel consumption (liter)	6 685	5 225
CO_2 emission total (ton)	3 015	2 357
CO_2 emission average (ton)	17	13
CO_2 emission during idling (ton)	5 372	
CO_2 emission average (ton)	31	

5 RESULTS

In this section the result from different models will be presented. It starts by presenting the accuracy of the different models. However, the most accurate model will be collected to outline the weight of the predictor on the response variable.

5.1 Impact of pre-processing on the accuracy of the models

Table 32 presents the validation parameters of the ML-R, PLS-R, SVM and PC-R. It is also presenting the impact of pre-processing of the raw data on the result of the models.

Table 32 Impact of pre-processing on the accuracy of the models

Model	R-square		RMSE	
	Calibration	Validation	Calibration	Validation
Non-pre-processed				
ML-R	failed	failed	failed	failed
PLS-R	failed	failed	failed	failed
SVM-R	failed	failed	failed	failed
PC-R	failed	failed	failed	failed
Mean centering				
ML-R	0.87	0.98	1.307e+05	6.688e+05
PLS-R	0.66	0.95	1.922e+05	6.863e+05
SVM-R	0.846	0.739	1.575e+05	1.813e+05
PC-R	0.657	0.957	1.431e+06	6.047e+05
Mean centering + UV Scaling				
ML-R	0.59	0.52	0.0565	0.0601
PLS-R	0.54	0.50	0.0523	0.0531
SVM-R	0.64	0.60	0.0528	0.0542
PC-R	0.46	0.36	0.0565	0.0603
Mean centering + max-min scaling feature				
ML-R	0.86	0.98	0.041	0.020
PLS-R	0.89	0.97	0.026	0.029
SVM-R	0.84	0.73	0.045	0.058
PC-R	0.49	0.83	0.075	0.026
Predictor with high correlation coefficient (mean center)				
ML-R	0.94	0.99	8.051e+05	5.969e+05
PLS-R	0.84	0.97	1.284e+05	7.622e+05
SVM-R	0.93	0.81	1.064e+06	1.684e+06
PC-R	0.83	0.96	1.063e+05	8.210e+05

5.2 Impact of pre-processing Gaussian process regression model

Table 33 presents the validation parameters of the GP-R model, it is also showing the impact of pre-processing on the accuracy of the model.

Table 33 Impact of pre-processing on the accuracy of GP-R

	RMSE	R2
Mean centered		
Calibration	1.2648e+06	0.82
Validation	3.4930e+05	0.98
Mean centered and normalized		
Calibration	0.51	0.76
Validation	0.17	0.98

5.3 Accuracy of neural network and LSTMN

Table 34 presents the validation parameters of the ANN model, it is also showing the impact of pre-processing on the accuracy of the model.

Table 34 Accuracy of neural network

	Scaled & mean center	non-pre-processed
R-square	0.99	0.98
RMSE	0.45	2.857e+06

Table 35 is conducted to present the external validation parameters of the LSTMN model.

Table 35 External validation parameters of LSTMN

RMSE	**R^2**
2.457e+06	0

5.4 CO_2 emission from the equipment

Table 36 presents the estimation of CO_2 emission based on NONROAD2008 and fuel used.

Table 36 Compression of CO_2 estimation techniques

		IDLE2	IDLE1
NONROAD2008	Total (hr)	183 859	142 852
	Average (hr)	507	394
	CO_2 total [ton]	5 695	5 775

	CO_2 average [ton]	15,73	15,95
	Total idle emission [ton]	11 470	
	average idle emission [ton]	31	
Fuel used		IDLE 2	IDLE 1
	Total fuel consumption (liter)	1 142 245	892 852
	Average fuel consumption (liter)	6 685	5 225
	CO_2 emission total (ton)	3 015	2 357
	CO_2 emission average (ton)	17	13
	CO_2 emission during idling (ton)	5 372	
	CO_2 emission average (ton)	31	

5.5 Weight of predictors

In this section, the most accurate models are collected to outline the weight of the predictors corresponding to different factors. However, because PLS-R had the highest accuracy compared to other models, the important variables that impact the response variable will be outlined.

5.5.1 *Partial least square regression*

Figure 47 presents the weight of variables that contribute to the response variable (fuel consumption) corresponding to first factor. The horizontal axis displays the variables from highest to lowest corresponding to regression coefficient property. The vertical axis displays the regression coefficient values.

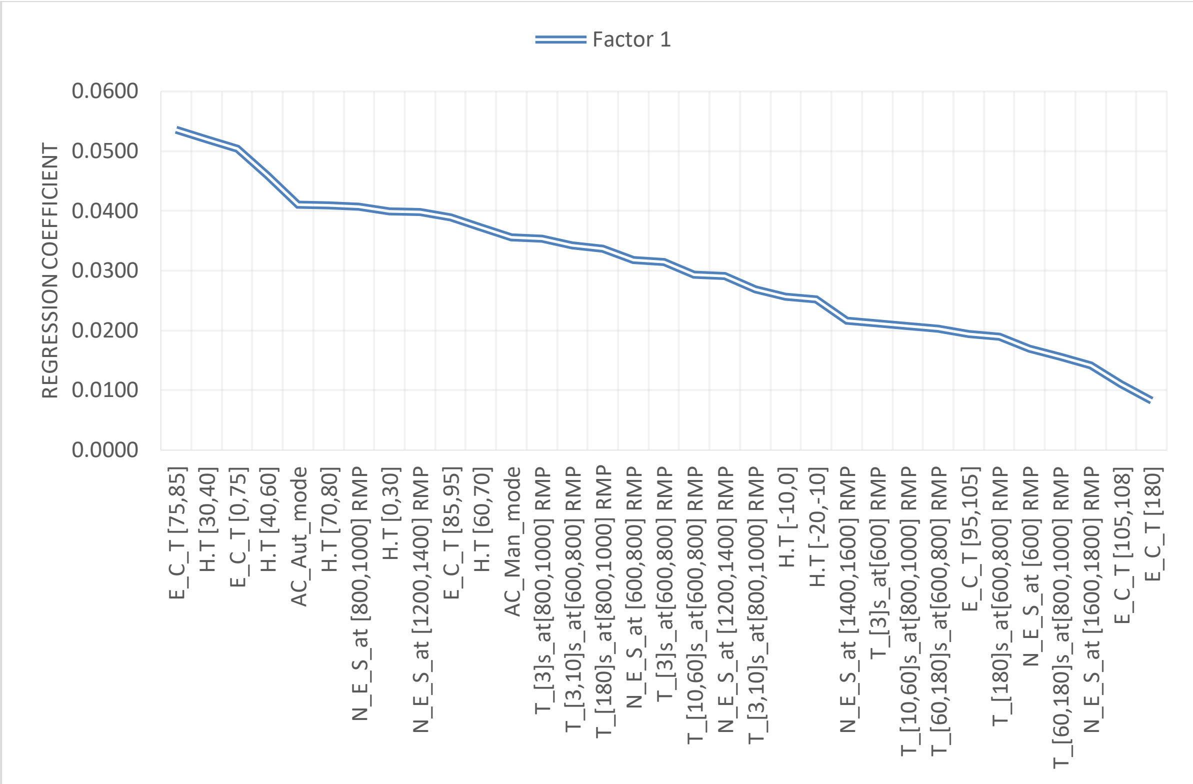


Figure 47 Important variables corresponding to first factor

Figure 48 presents the weight of the predictors corresponding to second factor of PLS-R model.

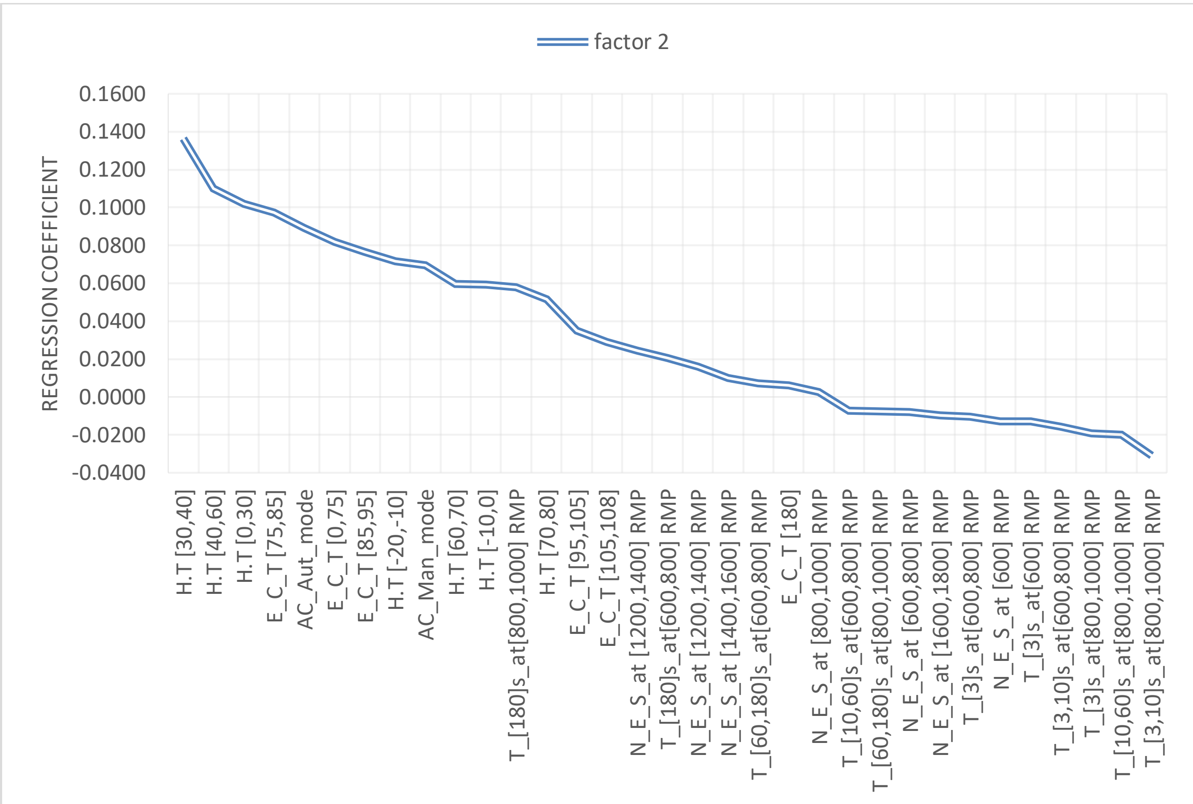


Figure 48 Important variables corresponding to second factor

Figure 49 present the weight of the predictors corresponds to third factor of the PLS-R model.

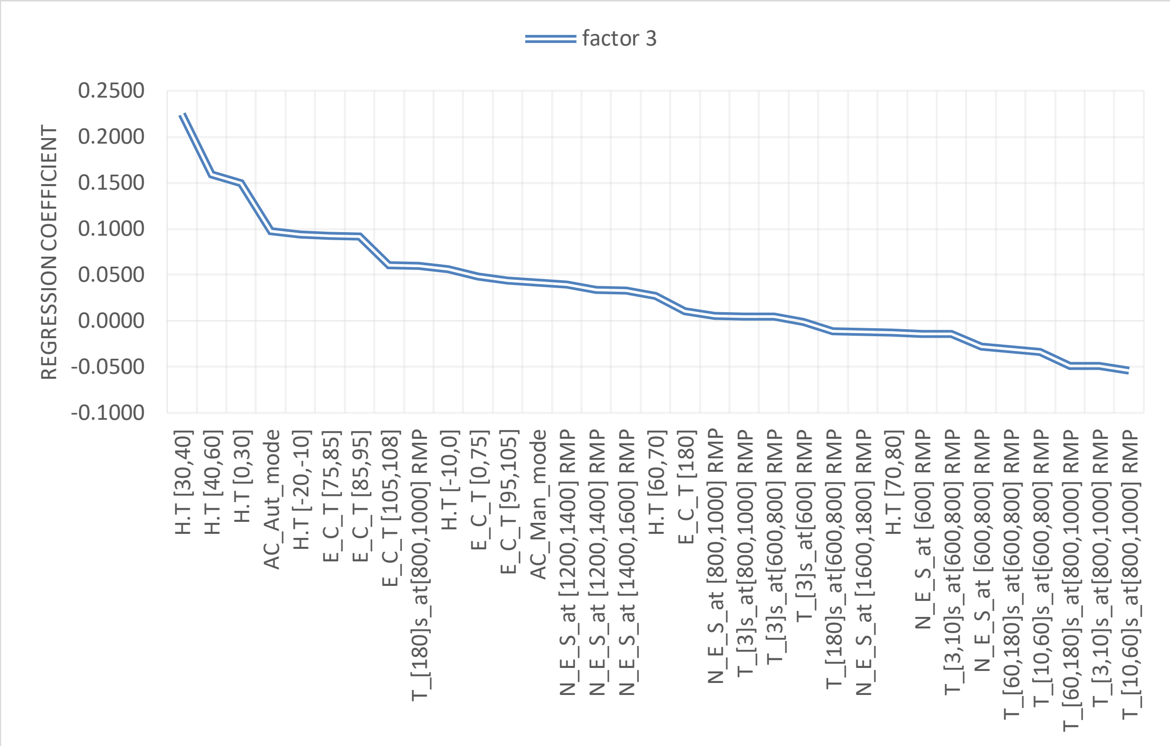


Figure 49 Important variables corresponding to third factor

5.5.2 Partial least square regression (predictor with high correlation coefficient)

This section illustrates the weight of the predictors that belong to the most accurate model in this field. Figure 50 displays the weight of the predictors which contribute to the response variable corresponding to the first factor, the horizontal axis is ranked from highest to lowest.

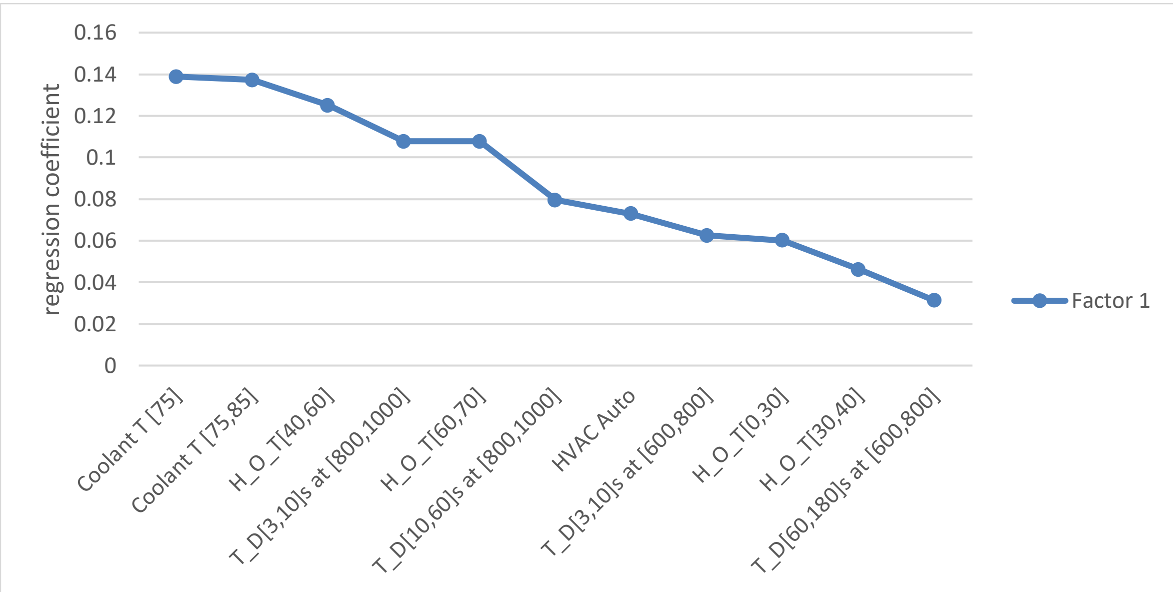


Figure 50 Important variables corresponding to first factor

Figure 51 presents the weight of the predictors corresponding to second factor.

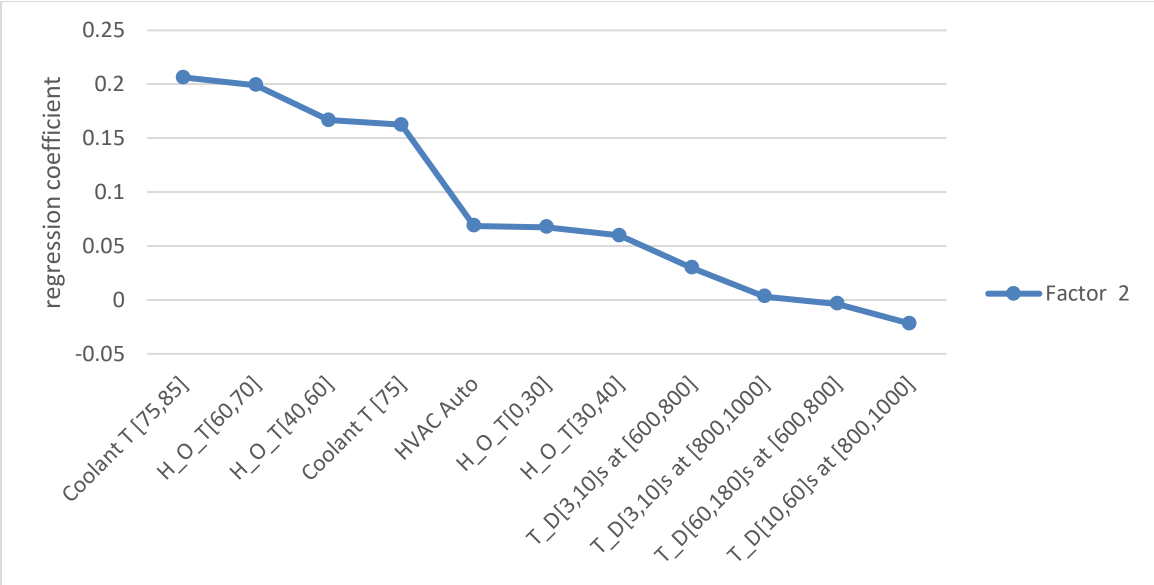


Figure 51 Important variables corresponding to second factor

Figure 52 present the weight of the predictors corresponding to third factor of the PLS-R model.

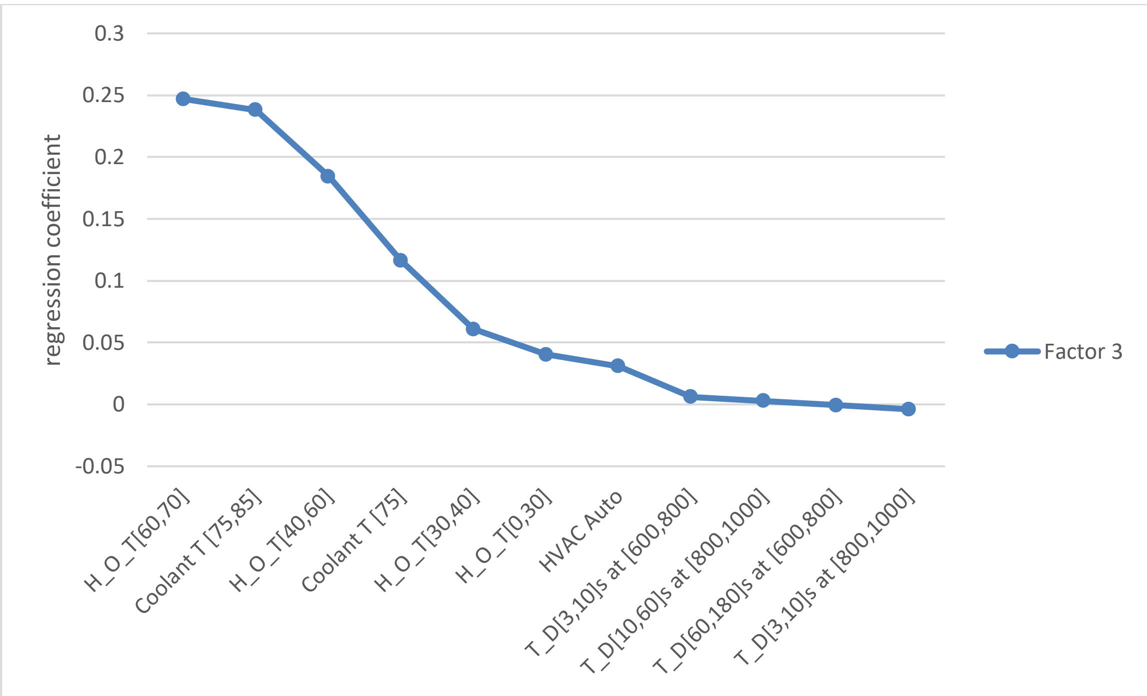


Figure 52 Important variables corresponding to third factor

6 DISCUSSION

This section is conducted to discuss the achieved results in this study. It starts by discussing the accuracy of the models and ends by a reflection on the social economic and environmental aspects of this particular thesis work.

6.1 Performance of models

The result of this work indicates that the pre-processing phase of this experiment has a high impact on the accuracy of the models. This incident observed especially when similar models developed by skipping the pre-processing part. However, this might depend on the high variance among the variables. Figure 53 illustrates the contribution of the samples to the model, the upper part of the model displays the none pre-processed case and the lower part of the figure shows the samples which are pre-processed through mean centering and max-min scaling feature. However, each color presents a sample in the figure and the data set contains in total 362 samples. The lower part is more colorful compared to the upper part. This indicates that the pattern recognition techniques which were developed with pre-processing, contain more samples to predict the response variable.

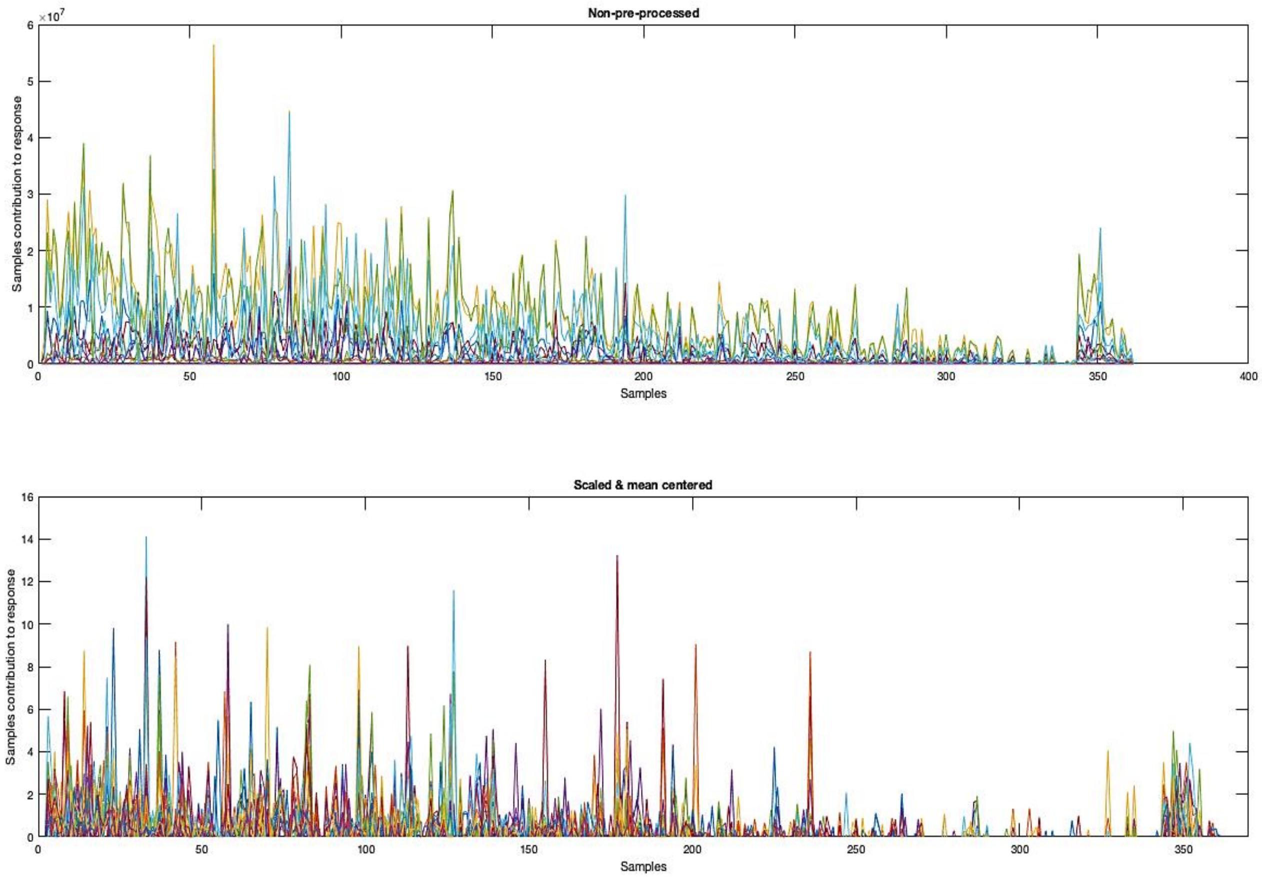


Figure 53 Samples contribution to the model

Furthermore, by adopting the first pre-processing (mean-centering) the result of different models improved remarkably. However, the issue that was observed in this case was that the model could not explain the weight of all predictors. This can be observed in Figure 54, the upper part of the figure presents the weight of the predictors that are normalized and mean-centered at the pre-processing phase, the lower part of the figure is illustrates the weight of the predictors which are only mean-centered at the pre-processing phase. However, according to Figure 54, the majority of the predictor's weight in the non-normalized case are located at the horizontal axis which indicates that those variables do not contribute to the response variable. However, the upper part of the figure which is normalized illustrates that most of the predictors are non-zero.

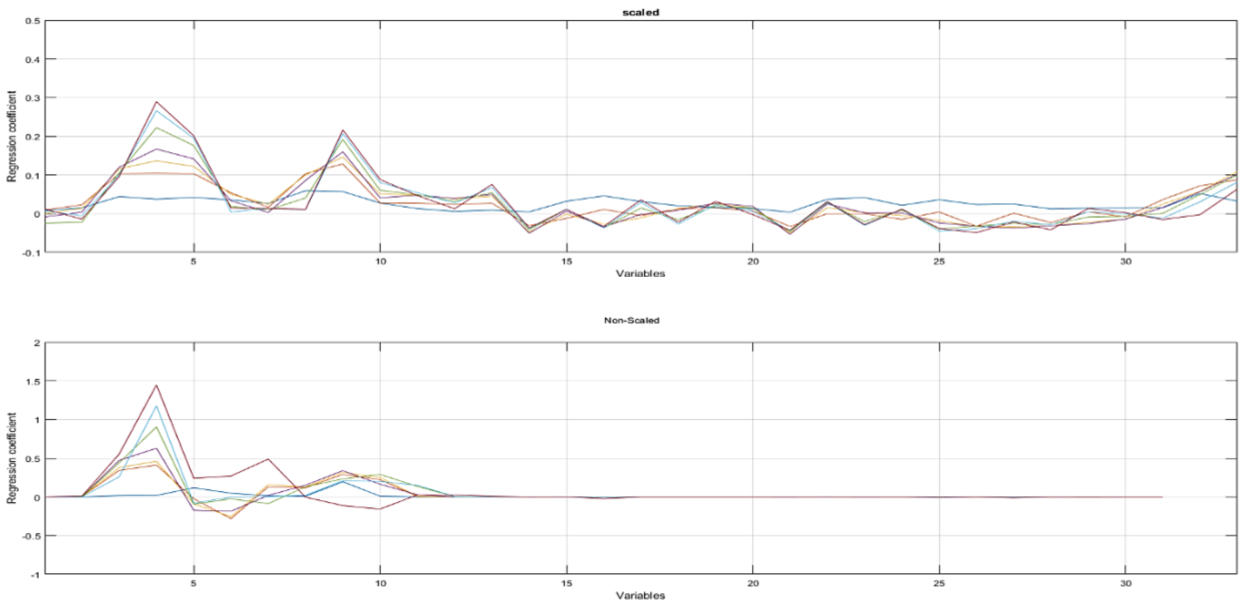


Figure 54 Impact of the normalization on the PLS-R model

A similar observation was detected even in the case of predictors with high correlation coefficient property. Figure 55 displays the weight of the variables that correspond to different factors of the PLS-R model. However, the upper part of the figure shows the weight of the predictors which are normalized and the lower part illustrates the non-normalized cased.

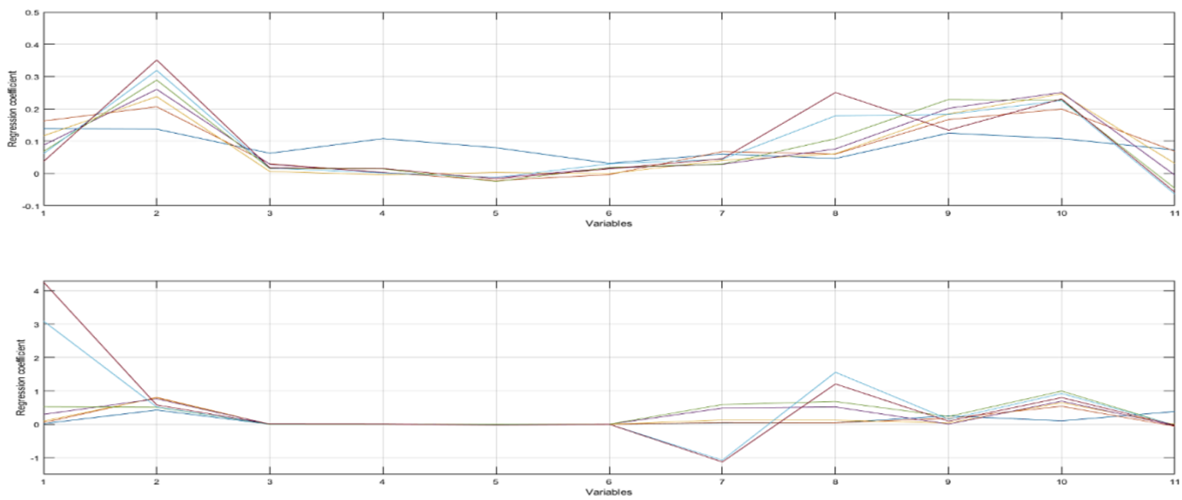


Figure 55 Impact of normalization on the PLS-R model

Pre-processing has a major impact on the accuracy of projection pattern recognition techniques. However, normalization of the data set is a necessary step that is needed to be applied on the data set of this field in order to obtain models that can explain the majority of the variance and response variable. Furthermore, there is another advantage that the normalization approach brings into the model, the RMSE's property becomes remarkably smaller. RMSE is a validation parameter that is used to assess and evaluate the accuracy of the models, it indicates the distance between predict variables and reference variables. The RMSE is very big in the case of non-normalization and becomes proper when the data set is normalized.

In the case of classic statistics, the impact of pre-processing can be observed from the t-values and p-values of the ML-R. However, the ML-R is developed with 95% confident intervals. In the case of non-scaling, the majority of the predictors have a p-value higher than 5%. It should be lower than 5% to have an accurate ML-R model.

In the case of neural network pre-processing is not a necessary approach to take into consideration. Hence in the case of the neural network, the variables are trained as pattern recognition techniques to learn the equipment behavior. However, the result of the neural network indicates that the validation parameters become more accurate especially RMSE by pre-processing the raw data. This incident might be because of high variance among the samples which are already discovered by the different projection and machine learning techniques. The variance among predictors was observed firstly during the model development. The first attempt was to train a model with 10 hidden layers but the model gave no result and demanded more samples. However, the procedure of reducing hidden layers was continued until a proper neural network was trained.

The result of this work indicates that deep learning has the lowest accuracy compared to classic statistics, projection, machine learning and artificial neural network. However, deep learning is a modern and accurate technique. The reason behind the bad prediction of this model is that it is a sample demanding method and requires at least 2000 samples. As it was mentioned earlier, there are high variance between both samples and variables in this work. Therefore, more samples are needed to train a model which contains high level of accuracy.

It is obvious that the accuracy of the calibration should be higher than accuracy of the validation because in the case of calibration the trained model is tested to estimate 5% of the training samples. However, PLS model is a projection technique and it follows the pattern recognition information. This incident (low accuracy of calibration) can be explained by considering that those test samples (5%) might belong to outliers or have a different pattern compared to most of the training samples.

In the case of projection technique, the model should be developed on the samples that contain most important information about the dataset. One of the biggest challenges of model development is to identify the outlier of the dataset. PCA is a powerful tool and most common technique that is applied in this field to identify the principal samples and outliers. During the model development, it was observed that by reducing the outliers, the accuracy of the model was decreased which indicates that the dataset contained some degree of outliers that different patterns compared to most of samples but still the outliers contained important information about pattern of the dataset.

6.2 Reflection on the weight of the predictors

Fan (2017), performs a similar study with difference that the author explores the strength of explanatory variables such as maker, equipment age, and horsepower. However, the data source was collected manually by a weekly data collection method, which contains some level of inaccuracy that is confirmed by the author. The author explains that this is because of the poor maintenance of the sensors. In this work, the multivariate data analyses are based on the data which comes from a telematics systems data warehouse. Fan (2017), explains further that the data from a telematics system is more accurate.

Peruzzi et. al (2016) estimates that the idling duration covers almost 20% of off-road equipment lifetime. The data set of this work indicates that the idling duration for excavator EC480 covers on average 31% of the equipment lifetime. Idling duration of the equipment has a negative impact on the fuel economy and emits tons of exhaust emission into environment.

Peruzzi et. al (2016) and Pekula, et al (2003), state that ambient air temperature has a significant impact on the equipment's duration at idle-condition. The weight of predictors indicates the importance of ambient temperature due to important variables of the most accurate models shows hydraulic oil temperature and engine coolant temperature at lower temperature range have more contribution to fuel consumption.

Perozzi et al (2016) and Matthews et al (2017), classifies the idling into different categories based on the magnitude of idle duration. However, the data set of this particular investigation does not have a real-world time domain which makes it difficult to classify the idling of the equipment into different categories. The duration of idling for this type of equipment covers 31% of its lifetime, by taking this statement into account it can be assumed that a higher ratio of the idling duration belongs to long and medium groups; the reason behind this might be that the operator keeps the idling duration for comfort purpose inside the cabin. The medium-term refers to the idling of the equipment e.g. adjustment of job activities.

Government of Canada (2015), states that by increasing public education in the field of fuel consumption and exhaust emission during the idling, the fuel economy improves and the exhaust emission reduces remarkably. One way to take advantage from this statement is to improve the operator's skills or constructor firms hire only experienced operators. The issue with this approach is that hiring experienced operators is an expensive option and small contractors don't have the financial ability for it. It is also notable that teaching operators is also costly (Fan, 2017). Nevertheless, Volvo CE's telematics system is already sending a weekly report about how every machine was used during the week to the fleet manager. The fleet manager should take responsibility and inform the operators through how the operator's action is impacting the environment. According to Government of Canada (2015) reducing the idle duration by 2-3 minutes daily, brings $630 million saving a year in fuel cost and corresponding exhaust emissions to the atmosphere.

6.3 Reflection on the social economic and environmental aspects of the work

In this section, an overview of the yielded results for equipment and component impact on the fuel consumption during idling will be presented. Both the fleet manager and the manufacturing line can get benefit of the result of this work in several ways. The manufactures can assess and evaluate the predictors with the highest weight and outline the causal correlation that corresponds to fuel consumption. Based on the conclusion, proceed to redesign the engine in such manner that efficiency of the predictor's with highest weight improvs.

As mentioned in 1.1.3, Volvo CE's telematics system is sending a weekly report about the performance of the equipment to the fleet manager in the EMEA region. It is possible to train a model for automatic regression into the Volvo CE's telematic system which predicts the fuel consumption by adjustment of operator behavior. However, fuel cost is a known phenomenon in ownership of construction equipment such as automated system that predicts the fuel consumption. By applying automated system into Volvo CE's telematics system, the fleet managers will be well aware of the fuel consumption during idling. Later, strategies about operator behavior can be decided accordingly. The saved costs in this field can be allocated to e.g. the ownership of equipment or to the project. However, there are some uncertainties about the CO_2 emission from construction equipment, an automated regression model in a similar manner can support authorities to estimate the exhaust emission at the national level.

6.4 CO_2 estimation

In this section, CO_2 emission from the excavator (EC480E) during idling is been estimated through two different approaches. The Environmental Protective Energy Agency (EPA) has developed an equation for CO_2 emission estimation for off-road equipment. It is based on steady-state engine dynamometer tests in the laboratory environment. Further, the second estimation approach is based on fuel consumption during the idling which takes the CO_2 emission factor which is recommended by International Plant Protection Convention (IPPC). However, the estimation shows that the CO_2 emission estimation based on EPA equation estimates a higher emission compared to estimation from fuel consumption. Thus, the fuel quality is not similar worldwide, the EPA exhaust emission technique is a proper manner to estimate the exhaust emission. The EPA equation does not take the fuel consumption into consideration, it takes the operation hour at a certain horsepower into account.

The excavator EC480E matches the latest version of emission requirement (Europe stage V), the problem with the EPA equation is that the parameters of the EPA equation are needed to be up to date, because it does not represent the CO_2 emission from EC480E properly.

6.5 Gender equity in equipment driving

Quinlan & Venderburg (2017), specify that gender and sex do not refer to the same topic. The term sex is used to describe the biology of an individual which refers to females or males. In contrast, the terms of gender are representing the sociocultural perspective. Sawyer (2012), observes that the sociocultural perspective analyzes groups and individuals together, based on both anthropological and psychological aspects. However, anthropological science provides multidisciplinary research on the nature of human and human-behavior. In contrast, psychological science investigates and presents studies on the nature of behavior and mind (Jonasson , Lauring, & Bjerregaard, 2012). Lane (2019), says that a "gender lens" is a fundamental approach that needs to be considered everywhere for human and economic development. During the particular period of time when" state-of-art" of this specific work was composed, it was observed that the majority of web-pages and books are barely using men pictures on the cover side to advertise their products. The phenomenon was especially observed when the objective of the research was to gain a better understanding of "construction-operator" or "construction-job-sites". This indicates that the target labor-gender in field of construction-job or equipment operator are mostly men, which might be decided by the decision-makers. However, this will impact the sociocultural construction of the gender division of labor in this spot. The majority of the product specialists at Uptime center Volvo CE, confirms that a higher distribution of labor in the field of operators are men. Quinlan & Venderburg (2017), observes that women are the key to future growth and a powerful driver of the economy in emerging business activities around the world. Hruby (2018), states that by having women in different positions brings positive financial support for the business and gives social advantages at the job environment. To move beyond this state, Quinlan & Venderburg (2017), observe that education is the key solution and more educated female means more women entering the labor market.

7 CONCLUSIONS

The result of this thesis work convinced that pre-processing of the data set has a significant impact on the accuracy of the models, among to pre-processing approaches mean-centering and max-min scaling feature improvs the forecasting degree of the models remarkably in the field of this data set. However, there are high variance among samples as well as variables of the data set because samples represent performance of the equipment in a wide geographical region (EMEA). By assessing and evaluation of the validation parameters of the models that are developed in this work, it can be concluded that PLS-R has the highest accuracy in comparison with other models. Models made by ANN and GP-R have higher coefficient of determination in compression to PLS-R. Nevertheless, the PLS-R model contains a RMSE closer to zero in contrast with ANN.

Based on the result of this work, it can be concluded that through an explanatory data analysis the strength of different factors that contributes to fuel consumption during the idling can be explored. It also increases the knowledge about the performance of the equipment during the idling and thus, gives a better overview. However, engine coolant temperature at the temperature range of [75,85] degrees Celsius have the highest weight in this field. Hydraulic oil temperature at temperature range of [30,40] degrees Celsius is the next weighted predictor. Engine coolant temperature at the temperature range of [0,75] degrees Celsius is the third predict variable and hydraulic oil temperature at temperature range of [40,60] degrees Celsius is the fourth weighted variable and so on (see section 5.5).

By considering the variance of fuel quality worldwide, estimation of exhaust emission based on NONROAD2008 is a proper method. Hence, estimation procedure of NONROAD2008 method is based on the equipment operating hours engine load and engine speed.

8 SUGGESTIONS FOR FURTHER WORK

Operation skill is not considered in this experiment, the literature confirms that operator skills have an important impact on the fuel consumption. There is high regulation imposed by Volvo Group on operator identity, it would be valuable to perform similar study with taking operator skills into consideration. This could be performed in such manner that contacts should be established with fleet managers and operators who voluntarily are willing to participate in this kind of study.

In this study, a single equipment model is investigated in EMEA region. The result showed that there is high variance among variables and observations. In order to tackle this issue, the data from entire excavator models should be extracted from Volvo CE's telematic system data warehouse. By acquiring such a huge data set and by applying deep learning methods such as long-short-term-memory-network, a more precise model can be developed.

9 REFERENCES

A.J., F. (2016). *When Trucks Stoprunning Energy and the Future of Transportation* . Cham: Springer: Imprint: Springer .

Abolhasani, S., Frey , C. H., Kangwook, K., Lewis , P., & Pang , S. (2012). Real-world In-in-Use Activity, fuel use and Emissions for Nonroad Construction vehicles: A case study for Excavators. pp. Vol.58(8),pp.1033-1046.

Adriansson, E. (2019, December 12). *Växthusgaser, utsläpp från arbetsmaskiner*. Retrieved from Utsläppen från arbetsmaskiner har varierat under de senaste decennierna, men åtgärder krävs för minskningar framöver.: http://www.naturvardsverket.se/Sa-mar-miljon/Statistik-A-O/Vaxthusgaser-utslapp-fran-arbetsmaskiner/

Agarwal, A. K., Gupta, T., Shukla, P. C., & Dhar, A. (2015). Prticulate emissions from biodiesel fuelled CI engines . *Energy Conversion and Management* , Vol. (94), pp-311-330.

Allied Market Research. (2020, March 03). Retrieved from Heavy Construction Equipment Industry Overview: https://www.alliedmarketresearch.com/heavy-construction-equipment-market

Ashrafur , R. S., Majuki, H. H., Kalam , M., Abedin , M. J., Sanjid , A., & 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* , Vol. 74,pp.171-182.

Ayyadevara , K. A. (2018). *Pro Machine Learning Algorithms, A hand on approach to Implementin Algorithms in Python and R*. Berkeley: Imprint: Apress.

Bast, A. (2019, April 26). *Construction Equipment-Europe*. Retrieved from VOLVO CE LAUNCHES EMEA UPTIME CENTER: https://www.volvoce.com/europe/en/about-us/news/2019/volvo-ce-launches-emea-uptime-center/

Bironneau, M., & Coleman, T. (2019). *Machine Learning with Go Quick start Guide* . MUMBAI: Packt Publishing .

Brendan, T. (2018). *data science*. Cambridge: MA : The MIT Press.

Brodrick, C. J., Dwyer, H. A., Farshchi, M., Harris, B. D., & King JR, F. G. (2011). Effects of Engine Speed and Accessory Load on Idling Emissions from Heavy-Duty Diesel Truck Engines.

Brodrick, C. J., Emilio, A. L., Burke, F. A., Farshchi, M., Li, L., & Deaton, M. (2004). Effect of Vehicle Operations. Weight, and Accessory Use on Emissions from a Modern Heavy-Duty Diesel Truck.

Buchanan, S. (2016). *Global warming* . California : Huntington Beach.

Cerit , M., & Buyukkaya, E. (2008). Experimental study of NOx eimmisions and injection timing of low heat rejection diesel engine . *Inernational Journal Of Thermal Sciences* , Vol.47, pp.1096-1106.

Chen, Y. (2016). Reference-related component analysis: A new method inheriting the advantages of PLS and PCA for separating interesting information and reducing data dimension . *Chemometrics and Intelligent Laboratory Systems* , Vol.156,pp.196-202.

Chretienne, P. (2014). On scheduling with the non-idling constraint. *Scholarly Journals*, vol.(12),Iss. 2, pp.101-121.

Colin, F. R., & Allan , K. (2016). *Internal combustion engines: applied thermosiences* . England: Wiley : Chichester, 3rd edition .

Collier, S., Ruehl, C., Yoon, S., Boriboonsomsin, K., Durbin, D. T., Scora, G., . . . Herner, J. (2019). Impact of Heavy-Duty Diesel Truck Activity on Fuel Consumption and Its Implication for the Reduction of Greenhouse Gas Emissions.

Confidential. (2019). *VCE Total Market and share (top 15 countries).* Eskilstuna: Commercial Managment.

CONSTRUCTION EQUOPMENT. (2010, September 28). *Volvo L220F Hybrid wheel Loader* . Retrieved from https://www.constructionequipment.com/volvo-l220f-hybrid-wheel-loader

Dass, R. (2012). pattern Recongintion Techniques: A review . *International Journal of computer Science and Telecommunications* , Vol.3,issue 8, pp,25-29.

Deng, J., Ma, F., Li, S., He, Y., Wang, M., Jiang , L., & Zhao, S. (2011). Experimental study on combustion and emission characteristics of a hydrogen-enriched compressed natural gas engine under idling conition. *International journal of hydrogen energy*, vol.(36)13150-13157.

Dybdahl, M., Risom , L., Bornholdt, J., Autrup , H., Loft , S., & Wallin, H. (2004). Inflammatory and genotoxic of diesel particles in vitro and in vivo. *Mutation Research*, vol. (562), pp.119-131.

EPA . (2020, April 10). *Greenhouse Gas Emissions* . Retrieved from Overview of Greenhouse Gases : https://www.epa.gov/ghgemissions/overview-greenhouse-gases

EPA. (2016). Nonroad Compression-Ignition Engines: Exhaust Emission Standards . *Office of Transportation and Air Quality* , EPA-420-B-16-022 .

EPA. (2017, April 2017). *Emission Standards Reference Guide*. Retrieved from Basic information about the Emission standarsds Reference Guide for On-road and Nonroad Vehicles and Engines: https://www.epa.gov/emission-standards-reference-guide/basic-information-about-emission-standards-reference-guide-road

EPA. (2018, MAY 10). *Regulations for Emissions From Vehicles and Engines*. Retrieved from https://www.epa.gov/regulations-emissions-vehicles-and-engines/regulations-emissions-heavy-equipment-compression#regulations

EPA. (2019, Septermber 6). *Laws & Regulations*. Retrieved from Regulations: https://www.epa.gov/laws-regulations/regulations

Eriksson, L., Byrne , T., Johansson , E., Trygg, J., & Vikström, C. (2013). *Multi- and Megavariate Data Analysis Basic Principles and Applications*. Malmo, Sweden: UMETRICS AN MKS COMPANY.

Erin, M. (2019). *Global warming* . New York: Cavendish Square .

European Commission. (2014, September 25). *Cutting emissions and cutting red tape: a new regulation for off-road engines*. Retrieved from https://ec.europa.eu/commission/presscorner/detail/en/IP_14_1044

European Commission. (2014). *Cutting emissions and cutting red tape: a new regulation for off-road engines*. Brussels: European Commission (press release).

European Commission. (2018, November 27). *Internal market, industry, Entrepreneurship and SMEs*. Retrieved from Non-Road Mobile Machinery (NRMM): https://ec.europa.eu/growth/sectors/automotive/environment-protection/non-road-mobile-machinery_en

Fan, H. (2017). A Critical Review and Analysis of Construction Equipment Emission Factors. *Procedia Engineering*, Vol.196,pp.351-358.

Ferguson , C. R., & Allan, K. (2016). *Internal combustion engines: applied thermoscienes (3rd ed.)*. England: Chichester: Wiley .

Fu, J. (2013). *Simulation of earthmoving Operations* . Stockholm : KTH Royal Institute of Technology.

Gaines, L. L., Hatman Brodrick, C.-j., & Solomon, M. (2009). Energy Use and Emissions of Idling-Reduction Options for Heavy-Duty Diesel Trucks.

Gopal , B., Shetty , A., & Ramya , B. (2014). Prediction of the presence of topsoil nitrogen from spaceborne hyperspectral data . *Geocarto international* , Vol.30(1), pp. 82-92.

Government of Canada . (2015, 12 29). *Natural Resources Canada* . Retrieved from Idling Wastes Fuel and Money : https://www.nrcan.gc.ca/energy/efficiency/communities-infrastructure/transportation/idling/4459

Graver , B. M., Frey , C. H., & Hu, J. (2016). Effect of Biodiesel Fuels on Real-world Emissions of Passenger Locomotives . *nvironmental Science & Technology* , vol 50(21),pp.12030-12039.

Hooftman, N., Messagie, M., Mierlo , V. J., & Coosemans , T. (2018). A Revieq of the European passenger car regulations- Real driving emissions vs local air quality . *Renewable and Sustainable Energy Reviews* , Vol. 86, pp. 1-21.

Hruby, A. (2018). *Driving gender equity in African business* . London: FT.cp,.

ICCT. (2016). *EUROPEAN STAGE V NON-ROAD EMISSIONS STANDARS.* Berlin: INTERNATIONAL COUNCIL ON CLEAN TRANSPORTATION.

IEA. (2005). *Making cars more fuel efficient.* Paris: International Energy Agency.

Jeng, R. H., & Chen , W. S. (2016). Two Feature-Level Fusion Methods with Feature Scaling and Hashing for Multimodal Biometrics. *Journal of IEtE Technical Review* , VOL. 34,pp.91-101.

Jonasson , C., Lauring, J., & Bjerregaard, T. (2012). *Communication and power: Psychological and anthripological perspectives on interaction in an intercultural organizational setting.* Scopus (Elsevier B.V).

Jonsson, L., & Bondemark, A. (2017). *FOSSILFRIHET FÖR ARBETSMASKINER.* Stockholm: WSP.

Joseph Van den Berg, A. (1999). TRUCKSTOP ELECTRIFICATION: REDUCING CO2 EMISSIONS FROM MOBILE SOURCES WHILE THEY ARE STATIONARY.

Juhrich, K. (2016). *CO2 Emission Factors for Fossil Fuels.* German Environment Agency (UBA).

Kamyabi, M. A. (2009). Simultaneous Spectrophotometric Determination of Paracetamol and p-Aminophenol by Using Mean Centering of Ratio Kinetic Profiles. *Journal of the Chinese Chemical Society* , vol.56,pp.142-149.

Kim, B. S., Jang , W. S., & Lee, D. E. (2014). Analysis of the CO2 Emission Characteristics of Earthwork Equipment . *Civil Engineering* , 19(1):1-9.

Kubsh, J. (2017). *Managing emissions from non-road vehicles* . Washington DC: The international council on Clean Transportation .

Kulkarni, M., & Mohanta , L. (2010). Diesel Exhaust Aftertreatment System Packaging and Flow Optimization on a Heavy-Duty Diesel Engine Powered Vehicle . *SAE International*, Vol.3, Issue 1, pp.143-155.

Lane , R. (2019). Karja Iversen: driving fore in women's health and gender equity. *Department of Error*, Vol.(393), Issue 10190,pp. 22-28.

Lee, K., Woo, S., Kim, Y., & Park, S. (2017). Optimization and calibration strategy using design of experiment for a diesel engine. *Applies Thermal Engineering* , Vol.(123), pp.917-928.

Lewis, P., Rasdorf, W., Frey, C. H., & Leming, M. (2012). Effects of Engine Idling on National Ambient Air Quality Standards Criteria Pollutant Emissions from Nonroad Diesel Construction Equipment.

LGuarieiro, L. L., Souza, A. F., & Torres, E. A. (2009). Emission profile of 18 carbonyl compounds, CO, CO2, and NOx emitted by a diesel engine fuelled with diesel and ternary blends containing diesel, ethanol and biodiesel or vegetable oils. *Atmospheric Environment*, Vol.43,pp.2754-2761.

Lim, H. (2002). Study of Exhaust Emissions from Idling Heavy-Duty Trucks and Commercially Available Idle-Reducing Devices.

Manley, D. K., & Peters, V. A. (2012). An examination of fuel consumption trends in construction projects . *Energy Policy*, Vol.(50),pp.496-506.

Market Report. (2020, February 13). *Global Construction Equipment Industry*. Retrieved from https://www.reportbuyer.com/product/1053026/global-construction-equipment-industry.html

Masters, N. R. (2011). *Heavy equipment Operator*. Ann Arbor, Michigan: Cherry Lake Publishing Skapad.

Mathworks. (2016). *Introducing machine learning*. Matlab and Simulink.

MATLAB. (2020, February 7). *Documentation* . Retrieved from MATLAB: https://se.mathworks.com/help/matlab/

McCormick , R. L., Ross, J. D., & Graboski, M. S. (1997). Effect of Several Oxygenates on Regulated Emissions from Heavy-Duty Diesel Engine. *nvironmental Science & Technology* , Vol.31(4),pp:1144-1150.

Menon , A., & Dallmann, T. (2016). *TECHNOLOGY PATHWAYS FOR DIESEL ENGINES USED IN NON-ROAD VEHICLES AND EQUIPMEENT*. Washington DC: International Council on Clean Transportation.

Milieu Ltd. (2004). *COMPARISON OF THE EU AND US AIR QUALITY STANDARDS & PLANNING REQUIREMNTS*. Danish National Environmental Research Institue.

Milosavljevic, B., Pesic, R., & Dasic, P. (2015). Binary Logistic Regression Modeling of Idle CO Emissions in Order to Estimate Predictors Influences in Old Vehicle Park.

Mohakumar , S., & Senthikumar, P. (2017). Particulate matter formation and its control methodologies for diesel engine: A comprehensive review . *Renewable and Sustainable Energy Reviews* , Vol.(80),pp.1227-1238.

Na, L., & Wei, W. (2011). Study on Exhaust Emission and Its Control Techniques in Diesel Engine. *Fourth International Conference on Intelligent Computation Technology and Automation* (pp. pp.710-713). Shaanxi, China : IEE, Computer Society.

Naturvårdsverket. (2018). *Arbetsmaskiners klimat- och luftutsläpp, Redovisning av regeringsuppdrag om karläggnin och förslag för minskade utsläpp.* Stockholm: Naturvårdsverket.

Naturvårdsverket. (2019, December 12). *Växthusgaser, utsläpp från arbetsmaskiner.* Retrieved from http://www.naturvardsverket.se/Sa-mar-miljon/Statistik-A-O/Vaxthusgaser-utslapp-fran-arbetsmaskiner/?visuallyDisabledSeries=81fcdf66610cbd63

Paul, R. 1.-a. (2014). *Automotive fuels reference book.* Wattendale, Pennsylvania : SAE International .

Pekula, N., Kurtiz, B., Hearne, J., Marchese, J. A., & Hesketh, R. P. (2003). The Effect of Ambient Temperature, Humidity, and Engine Speed on Idling Emissions from Heavy-Duty Diesel Trucks.

Perozzi , D., Mattetti, M., Molari, G., & Sereni, E. (2016). Methodology to analys farm tractor idling time . *BIOSYSTEMS ENGINEERING* , Vol. 148. pp.81-89.

Quinlan, J., & Vanderburg, J. (2017). *Gender lens investing: uncovering opportuities for growth, returns, and impact* . New Jersey : Wiley .

Razmkhah , H., Abrishamchi, A., & Torkian , A. (2010). Evaluation of spatial and temporal variation in water quality by patternrecognition techniques: A case study on Jajrood River (Tehran, Iran). *Journal of Environmental Management* , vol.91,pp-852-860.

Resitoglu, I. A., Altinisik, K., & Keskin, A. (2014). The pollutant emissions from diesel-engine vehicles and exhaust aftertreatment systems . *Clean technology environmental policy*, Vol.15,pp.15-273.

Sajjad, H., Sanjid, A., Abedin , M. J., Kalam, M. A., Masjuki, H. H., & Rahman, A. S. (2013). Impact of idling on fuel consumption and exhaust emissions and available idle-reduction technologies for diesel vehicles-A review . *Energy Conversion and Management* , vol.(74), pp 171-182.

Sandanayake , M., Zhang, G., & Setunge , S. (2015). *Environmental Emissions of Construction Equipment Usage in Pile Foundation Constrction Process -A case study.* Berlin Heidelberg : Springer- Verlag .

Sandanayake , M., Zhang, G., & Setunge, S. (2019). Environmental Emissions of Construction Equipment Usage in Pile Foundation Construction Process -A Case Study . *Cleaner Productin* , Vol.211,pp.1517-1526.

Saxena, P., Gupta, T. K., Naik, V., Awasthi, A., Bansal, S., Chakraborty , M., . . . Goswami, V. (2019). *Air Pollution: Sources, Impacts and Controls* . England ; Boston: Wallingford, Oxfordshire, , Massachusetts : CABI.

Schuerger , A., Moores , J. E., Clausen , C. A., Nadine , B., & Britt , D. T. (2012). Methane from UV-irradiated carbonaceous chorndrites under simulated Martian conditions . *Journal of geophysical research* , Vol.117(8), pp.1-19.

Sharma , N., Agarwal , A. K., Estwood, P., Gupta , R., & Singh, A. P. (2018). *Air Pollution and Control.* Singapore : Springer Singapor: Imprint: Springer.

Shi , L., Cui , Y., Deng, K., Peng , H., & Chen , Y. (2006). Study of low emission homogeneous charge compression ignition (HCCI) engine using combined internal and external exhaust gas rexirculation (EGR). *Energy*, vol. 31, pp. 2665-2676.

Shukla , P. C., Gupta, T., & Agarwal , A. K. (2018). *Techniques to Control Emissions from a Diesel Engine* . Singapore : Springer Nature .

Swecon. (2020, January 30). *Swecon Anläggningsmaskiner*. Retrieved from https://www.volvoce.com/sverige/sv-se/swecon/about-us/about-volvo-ce/

Sweyer , K. (2012). Extending Sociocultural theory to group creativity . *Vocations and Learning*, Vol.(5), pp. 59-75.

The Constructor. (2020, January 26). *16 Types of Heavy Equipment Used in Construction*. Retrieved from https://theconstructor.org/construction/heavy-construction-equipment-types/26305/

Toth , H. (2015). *A Career As a Heavy Equipment Operator* . New York City : Rosen Publishing Group.

Toth, H. (2016). *A career as a heavy equipment operator* . New York: Rosen Publishing .

United States . (1994). *Utilization of engineer construction equipment.* Washington D.C: Headquarters, Dept. of the Army.

USCENSUSBUREAU. (2002). *Industry general summary: 2002*. Washington D.C: U.S. Department of Commerce .

Waad, L. (2020, March 1). Fuel consumption and exhuast emission from construction equipment. (M. Hassani, Interviewer)

Wei, H., Zhu, T., Shu, G., Tan , L., & Wang , Y. (2012). Gasoline engine exhaust gas recirculation -A review. *Applied Energy*, Vol. 99, pp.534-544.

Volkov, N., & Chainikov, D. (2017). Reduction of Harmful Emissions During Start and Warming Up of the Engine.

Volvo CE. (2012). Volvo CE Pursues Advanced Control Systems to Boost Machine Productivity, Economy. *E&MJ engineering & Mining* , Vol. 213(5),pp.148(1).

Volvo CE. (2020, february 9). *Construction equipment in Norh America* . Retrieved from Activecare direct : https://www.volvoce.com/united-states/en-us/services/volvo-services/uptime-services/activecare-direct/

Volvo Construction Equipment (building Tomorrow). (2019, Jone 13). *Volvo construction equipment* . Retrieved from Volvo Construction Equipment Corporate Presentation : https://volvogroup.sharepoint.com/:p:/r/sites/cs-volvo-construction-equipment/_layouts/15/Doc.aspx?sourcedoc=%7B4581D763-158F-4770-8F5A-E0C040F50CD4%7D&file=Volvo%20Construction%20Equipment%20Corporate%20Presentation%202019.pptx&action=edit&mobileredirect

World Health Organization. (2020, January 21). *Indoor air pullution*. Retrieved from https://www.who.int/news-room/q-a-detail/indoor-air-pollution

Zheng, Z., Yue, L., Zhu, Y., Zhong, X., Yao, M., & Zhong, X. (2014). Effect of Two-stage injection on combustion and emissions under high EGR rate on a diesel engine b fueling blends of diesel/gasoline, Diesel/n-butanol, diesel/gasoline/n-butanol and Pure diesel. *Energy Conversion and Mangement*, Vol. 90, pp. 1-11.

Zietsman , J., & Perkinson, D. G. (2005). Estimating Extended Idling Emissions of Heavy-Duty Diesel Trucks in Texas. *Energy and Environmental Concerns* , Issue 1941,pp.34-42.

MÄLARDALEN UNIVERSITY
SWEDEN

P.O. Box 883, SE-721 23 Västerås, Sweden **Phone: +46 21 101 300**
P.O. Box 325, SE-631 05 Eskilstuna, Sweden **Phone: +46 16 153 600**
E-mail: info@mdh.se **Webb:** www.mdh.se