An Intelligent Model for Fleet Management by Use of Sensor Enabled Tags Integrated With GPRS Technology

Esther Wambui Nganga[1] George Okeyo[2] and Stephen Kimani[3]

Research Scholar [1], Professor [2] [3]

[1]P.O. Box 434-00600

Nairobi, Kenya

[2-3] Jomo Kenyatta University of Agriculture and Technology

School of Computing and Information Technology

P.O. Box 62000-00200

Nairobi, Kenya

_____

ABSTRACT

Fleet management is used to optimize transport functions and is implemented using the integration of technologies. This study seeks to address the issue of fuel adulteration. The high influx of vehicles has resulted in high consumption of fuel where fraudulent dealers exploit consumers by selling adulterated fuel. As a result; adulterated fuel leads to rapid wear and tear of parts, engine failures and complete breakdown leading to high cost of maintenance through frequent repairs and replacement of vehicle engines. Other consequences of fuel adulteration include the release of toxic gases into the air which are hazardous to the ecosystem and also a health hazard and loss of revenue for the government. Methods used previously to check for fuel adulteration include chemical methods; laboratory-based methods that examine chemical properties of the fuel, mechanical methods such as the use of sensors which involve installation of testing gadgets and mobile testing kits that can be used during fuelling. Presently the world is moving towards knowledge-based applications where there is a huge consumption of data; data from different sources make up big data that is analyzed to identify trends and make predictions. This study proposes the use of supervised machine learning that analyses vehicle data and is able to predict the number of vehicles using adulterated fuel based on the performance of a vehicle that is measured parameters by some onboard parameters. An experiment was performed using three classifiers and Random Forest classifier gave the best results examined by confusion matrix.
Key Words: Adulteration, Fleet Management, Random Forest classifier, Supervised Machine Learning.

_____

1. INTRODUCTION

Fleet management is a function which allows companies which rely on transportation in business to remove or minimize the risks associated with vehicle investment, improving efficiency, productivity and reducing their overall transportation and staff costs, providing 100% compliance with government legislation and much more [12]. An increased operational cost is one of the challenges in fleet management caused by rapid wear and tear, engine malfunction and eventual complete breakdown of the vehicle. This is largely caused by use of adulterated fuel. Adulteration is a worldwide problem however developing countries are worst hit by the practice because of the increase in motor vehicles; which is characteristic of a growing economy.

The main adulterant used in blending petrol and diesel is kerosene. This is largely so in Kenya because the government stopped taxing kerosene, in a bid to cushion underprivileged Kenyan households from high cost of living. However, years later, it is apparent that unscrupulous petrol dealers are the real beneficiaries of the no-tax policy on kerosene. The Petroleum Institute of East Africa (PIEA) said in the local dairies that adulteration of petrol and diesel with kerosene creates unfair competition for

players engaged in legitimate trade, with the country losing Sh3bn million because of tax evasion. The chairperson said unsuspecting motorists who buy adulterated petrol and diesel incur high expenses repairing engine parts or purchase new engines because of damage caused by fuel not complying with product standards. Adulterated fuel results in vehicles producing heavy smoke before the engines knock.

In a bid to protect the public, the government, through the Energy Regulatory Commission (ERC), issued a notice to all motorists in the country warning them off a scheme that can destroy their vehicles. The ERC asked drivers to be on alert over the presence of contaminated fuel in the Kenyan market following numerous complaints from consumers. The body advised the public to obtain a receipt for those paying for fuel in cash and to retain transaction messages for those using mobile money platforms. Drivers were also urged to report immediately in the event they realize signs of fuel contamination by their vehicles such as loss of engine power, jerking and producing excessive smoke after refueling at any station. Another measure by the government to curb the illicit trade was to increase tax on kerosene however the trade has continued to thrive. The government introduced test kits that could be used by oil firms to test presence of kerosene in diesel and petrol however were later discovered to be faulty as they did not detect presence of kerosene leaving the public exposed to fraudulent dealers.

Currently, there are other measures practiced all over the world to curb the vice. Such other methods include simulation of data from fuel samples, density tests of fuel, optical sensors, image processing, IR sensors, and computational techniques among others.

This research proposes an intelligent model that will be used to detect instances of fuel adulteration based on a vehicles performance. It is evident that adulterated fuel causes a vehicle to lose it engine power and also the contaminants lower the combustion rate of the fuel causing increased consumption in order to move. A combination of vehicle parameters that indicate possible use of adulterated fuel are factored in the model in order to project fuel problems. Such a model will be most beneficial to fleet operators as the vehicles make up the largest share of investment. In view of this, they are worst affected by the practice as consequences of fuel adulteration can immobilize vehicles. For the purpose of this study, four vehicle parameters have been considered; speed, gas mileage, knocking tendency and fuel quality. The parameters are retrieved from both on-board and off-board facilities. On-board includes all real time data (speed and gas mileage) precisely indicating performance of a vehicle at any given time. However we cannot rely fully on on-board data as the transmission units are limited in computational power and storage capacity. Off-board data is stored in resources with larger computational power and have access to historical data (knocking tendency and quality of fuel). This is important because on-board data may have challenges attributed to mechanical operations therefore the history in the off-board data will help determine whether an alarm will be classified as a true or false alarm. The off-board data can be clustered and labeled with failure root cause to form a knowledge base of operating modes linked to failures and normal conditions, and this will enable a data driven approach which is more precise in prediction.

After getting the parameters, machine learning will be used to establish the pattern in that set of data then use the knowledge to predict instances of similar nature. This model will predict vehicles suspect of using adulterated fuel and the information is vital in recommending preventive maintenance instead of reactive maintenance which is costly. The information can also be disseminated to relevant authorities for prosecution purposes or knowledgebase for the government; to enable efficiency in planning since fuel adulteration has a ripple effect on tax revenue, environment and health.

The rest of the paper is organized as follows: section II contains literature work from other authors, section III has discussion on the proposed model that includes the concept framework and classifier algorithm, section IV contains a summary of the dataset used for performing the experiment, experiment results and discussion of the results and lastly section V contains conclusion and future work.

2. RELATED WORKS

Fleet management involves activities that optimize functionality of fleet vehicles for maximum returns. Core activities include tracking, safety management and fuel management. For tracking purposes, vehicle status information is collected and sent to control center from where a vehicle is monitored. A model proposed by author [11] where IR sensors picked interruption signals and sent to the microcontroller which issued a message about the place of the vehicle to the car owner or authorized person. The controller also issued signals to the engine motor where speed gradually decreased to the off state and after that all the doors locked. Similarly author [6] proposed remote operation of a vehicle's ignition system and doors by invoking an SMS from the tracking system. Authors [14] proposed a tracking model that would be used to secure cargo by use of RFID card locks and biometric sensors to prevent pilferage and minimize human interference with the cargo while GPS system was used to monitor the vehicles to ensure safe arrival of cargo.

Safety management involves detecting accidents and this significantly reduces the response time to scenes of accidents. Author [13] proposed a system to detect accidents in significantly less time and sent the basic information to first aid centre within a few seconds covering geographical coordinates, the time and angle (measured by sensors) in which a vehicle accident had occurred. The alert message was sent to the central emergency dispatch server which informed ambulances near to that location to help in saving the valuable lives. A Switch was also provided in order to terminate sending of a message in rare case where there are no casualties; this saved the precious time of the ambulance. A real time collision detection model was proposed by reference [8] where vibration signals from an accelerometer were used to detect collision in real time by activating the GPS-GSM module to send coordinates of scene of accident to the operator and help was sent appropriately.

Fuel management serves to monitor and optimize costs associated with fuel. Core competencies in fuel management include quantity and quality of fuel. Considerable research has been done on fuel consumption/quantity.

Reference [4] proposed an intelligent vehicle monitoring system where fuel monitor sensor was used to monitor the fuel level; where, when and how much fuel was filled into the tank, and also showed the remaining content of the fuel in the tank and produced a warning signal to driver and owner if the fuel level was very low. The data was transferred to the cloud server, through GSM network, from where an authorized user of the tracking system could invoke the data for interrogation. Reference [1] also implemented a reed switch for fuel monitoring that worked by sensing the amount of fuel filled in the vehicle and amount of fuel consumed. The record was then stored in the system memory and GPS technology was used to track the vehicle.

Intelligent models have also been used in the area of fault detection; where data about performance is collected over a period of time and used to train a model that is used for future predictions of similar faults. This facilitates preventive maintenance than reactive maintenance which is costly. Reference [2] used machine learning to predict fuel consumption in heavy vehicles. Data used in this study included road, vehicle, driver and weather characteristics. The models used were comparable in performance and found out that important features that contribute to fuel consumption are road slope, vehicle speed and weight. Reference in [3] established that Support Vector Machine model of fuel consumption exposed comparatively better correlation than the other statistical fuel consumption models during an experiment to find out the consumption of passenger vehicles from data collected in real time

Reference [5] proposed Intelligent Driver System (IDS) that utilized vehicle performance data combined with GPS information on fixed routes to incrementally build a model of the historically most fuel efficient driving behavior. During driving, the calculated optimal velocity for specific location was compared to the current vehicle state and a fuzzy logic proportional derivative (PD) controller was used to compute the optimal control action. The control action was then projected to the drivers via a specialized human machine interface (HMI) or used directly as a predictive cruise control to achieve overall fuel economy improvements. Reference [16] used ML to model and predicts fuel consumption of a long distance public bus. They used a subset of available data for multi-variate time series analysis. They compared the predictive ability of three ML techniques in predicting the fuel consumption of the bus, given all available parameters as a time series. Based on the analysis, it was concluded that the random forest technique produced a more accurate prediction compared to both the gradient boosting and neural networks. Authors in [9] and [10] investigated supervised and unsupervised machine learning methods that used on-board data, off-board data and both to develop models that would detect faults earlier on before the scheduled maintenance. This saved on costs and increased the uptime of a vehicle. In their study on Analysis of Truck Compressor Failures Based on Logged Vehicle Data, they concluded that given the complexities of retrieving on-board data, off-board data could be best utilized to create a knowledge base that would be used to train prediction models for fault detection.

From the related works in fuel management, the issue of fuel consumption/quantity has been considerably addressed using machine learning; similarly, studies have been done to predict faults therefore this study proposes use of machine learning to predict quality of fuel; to be able to identify if a vehicle is running on adulterated fuel based on off-board data.

3. FLEET MANAGEMENT MODEL

3.1 Proposed model

The concept is based on an integrated technology environment where sensors capture vehicle status in the transmission unit and the data relayed to control centre via GPRS technology from where processing takes place. The proposed model is a supervised machine learning model that is used to predict vehicles running on adulterated fuel. With supervised learning, input variables (parameters) are identified then an algorithm is used to map them to an output variable. In this study, the model was created using off-board data; it contains on-board vehicle parameters that were recorded prior to vehicle malfunction and also maintenance and repair parameters that indicate the cause and effect, this therefore forms a knowledgebase from where inference can be made. The records contain text values that describe the value for each parameter hence use of classification algorithms. For purposes of the experiment in this study, we use four parameters; speed, fuel mileage, knocking tendency and fuel quality. In this case the input

variables are speed, fuel mileage, knocking tendency and output variable is fuel quality. Based on the dataset extracted from the knowledgebase, the algorithm is able to learn the trend of the vehicle parameters; the pattern of speed, fuel consumption and presence of knocking and is able to predict instances of vehicles that have used adulterated fuel. Following consequences of fuel adulteration on vehicle engines as discussed by reference [7], the predictions are important for planning for preventive maintenance of vehicles before the engine breaks down.
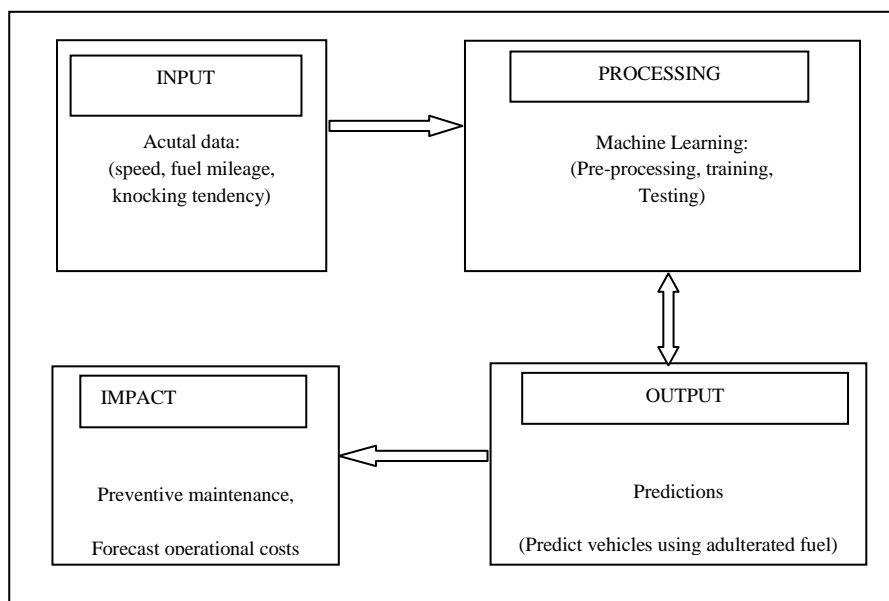

Figure 1: Conceptual Framework

3.2 supervised learning methods

The following is an overview of the supervised learning methods used for the experiment in this study.

- **Naïve Bayes (N.B)**

This classifier uses the Bayes rule of conditional probability and makes use of all the data features and analyses them individually on the assumption that they are equally important and independent of each other. It has the advantage of simplicity and quick convergence, however, it cannot learn about the interactions and relationships between the features in each sample [17].

- **Multi-Layer Perceptron (MLP)**

MLP is a feed forward artificial neural network that consists multi layers (usually 3) of neurons, each neuron is considered a processing unit that is activated using an activation function. MLP is a supervised machine learning method in which the network is trained using a labeled training data set; a trained MLP will be able to map a set of input data into a set of outputs [17].

- **Random Forest (R.F)**

RF is a decision tree based classification algorithm that is suitable for large data sets; it constructs a set of decision trees at training phase such that each tree operates on a predefined number of attributes chosen randomly. The classification process takes place by a majority vote of the results from each individual tree. Random Forest is trained on different parts of the training data set and aims at solving the problem of over fitting that is usually faced when using decision trees [17].
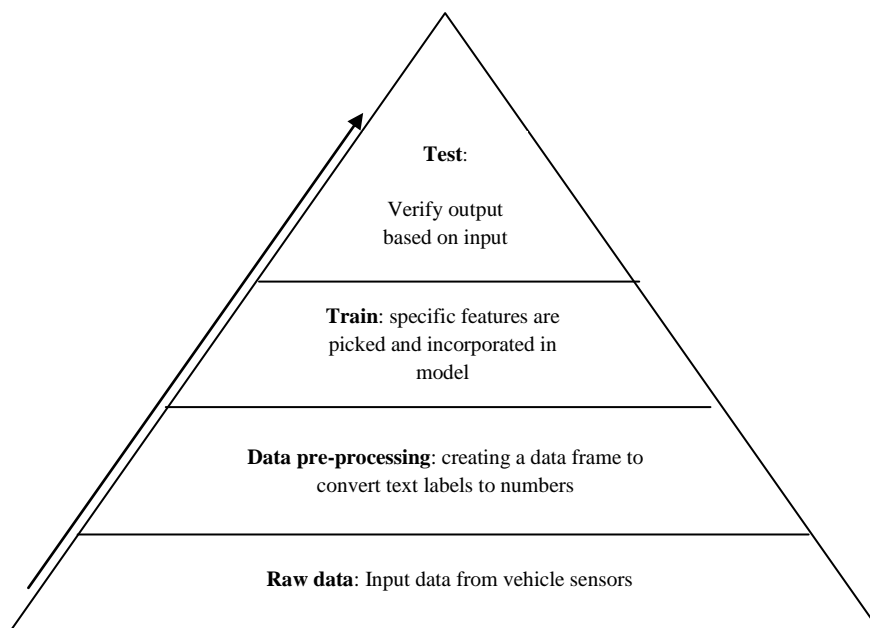
3.3 Model implementation

Figure 2: Hierarchy of processing model classifier

The hierarchy of activities in the supervised processing model are represented in four levels where output from one level forms the foundation in the following level. Level one contains the actual data; a dataset that is made up of on-board vehicle parameters collected in real time and off-board parameters that are characteristics of the vehicle. In level two, the data is pre-processed; involves converting the raw data into a form that will be used to train the classification model (RF) and splitting the data into training and testing data. In this model, we used 60/40 split because of the small dataset. Table 1 displays an extract of the raw data used in the model and table 2 displays the pre-processed data.

Table 1: Extract of raw data

speed	mileage	knocking
low	low	yes
average	average	yes
low	low	no
low	low	yes
high	high	no

Table 2 : Extract of pre-processed data

average	high	low	Average	high	low	no	yes
0	0	1	0	0	1	0	1
1	0	0	0	1	0	0	1

average	high	low	Average	high	low	no	yes
0	0	1	1	0	0	1	0
0	0	1	0	1	0	0	1
0	1	0	0	0	1	1	0

Level three involved training the model; features are picked from the training set and incorporated into the model. In level four, we tested the model on the test data (40%). We used Python programming language that uses Scikit-learn library for machine learning in this study. Figure 3 presents pseudocode for classification algorithm used in this model.

```
Input: dataset in excel
Output: prediction report

//Import dataset
Read data from an external file; data = pd.read_excel('name_of_file.xlsx')
//View data
data.head() will display the first five rows to confirm the dataset
//Start pre-processing using one hot encoding
Covert data for the features in text labels to numerical values (x variable) and assign class column to y
variable
one_hot = pd.get_dummies(X[feature])
//Import train test split to split the dataset into two portions (training, testing)
Split the data into two: x_train and y_train for training the model, y_train and y_test for testing
model, set the size of the test data
X_test and X_train.shape  //To verify the sizes of training and testing portions
//Create model
Import model_selection
model_name=Classifier_nameClassifier()
//Train the model
Pass the training data to the model; model_name.fit(X_train,y_train)
//Use model for prediction
Pass the test data to the model; model_predictions = model_name.predict(X_test)
//Evaluate accuracy of the model
model_accuracy = accuracy_score(model_predictions,y_test)
model_report = classification_report(model_predictions,y_test)
```
Figure 3: Pseudo code for processing model classifier

3.4 Performance Metrics
We conducted the experiment using different classification models to observe the best performing. To evaluate the models, we used confusion matrix that is used to describe a classifier's performance. The following are the terms that are used in the confusion matrix:

- True positive (TP): correct positive prediction
- False positive (FP): incorrect positive prediction
- True negative (TN): correct negative prediction
- False negative (FN): incorrect negative prediction

Metrics:

- Accuracy – to get the percentage of correct predictions
  (TP+TN)/TOTAL

- Precision – when a positive value is predicted, how often is the prediction correct
  TP/(TP+FP)

- Sensitivity/Recall –how often a prediction is correct when the actual value is positive
  TP/(TP+FN)

- F1 score – the weighted average of precision and sensitivity
  2TP/(2TP+FP+FN)

- Error rate/misclassification rate – how often is it wrong
  1-Accuracy

4. EXPERIMENT, RESULTS AND DISCUSSION

4.1 Experimental setup

The parameters were represented using text labels; speed and gas mileage were represented by low, average and high values, knocking tendency was represented by either yes or no and quality was represented by good or poor. We started by identifying the predictor and outcome variables; the predictor (X) values were speed, gas mileage and knocking tendency while the output value (Y) was fuel quality. We used Python programming language which is high level, object oriented and with simple syntax to use [15]. The code was written in Jupyter notebook. In addition, we used Scikit-learn; a library that contains algorithms for supervised and unsupervised machine learning. The model building started by importing the dataset followed by pre-processing the data as it contained categorical data. One-hot encoding was used to convert the categorical variables into numbers. This was followed by splitting the dataset into training and testing sets; the model's split for first data set was 60/40 where training set contained 39 records and testing contained 26 records and 70/30 for second set with 94 for training and 41 for testing. Both splits were valid as the training portions were larger than testing and a representative of the sample data. Classifier algorithms were imported from the Scikit library and trained using the training set, then used for prediction by passing the testing set to the models. Table 3 gives a summary of data collected for the experiment.

4.2 Dataset

We used secondary data that was obtained from two sources; first data set was from a fleet company that operates public service vehicles and second from an auto garage that deals with repairs and maintenance of vehicles. We used ms excel to extract the parameters of interest; speed, gas mileage, knocking tendency and quality of fuel. The entire data set had 200 records vehicles of the same make, 65 and 135 for first and second set respectively

Table 3: Summary of data set 1   **Table 4: Summary of data set 2**

Parameter \\ Value	Speed	Gas Mileage
Low	24	11
Average	20	18
High	21	36
	Knocking tendency	
Yes	24	
No	41	
	Quality of fuel	
Good	46	
Poor	19	

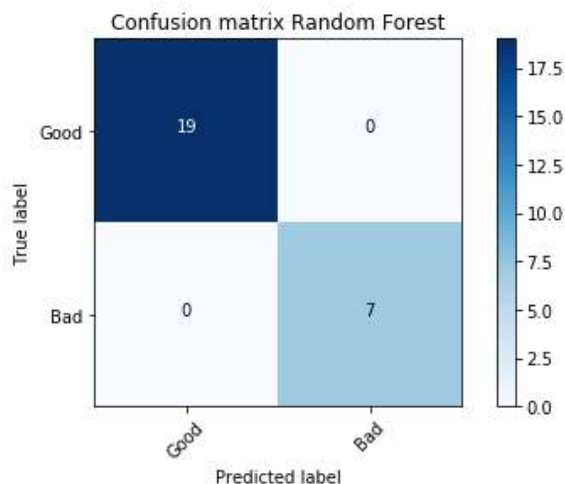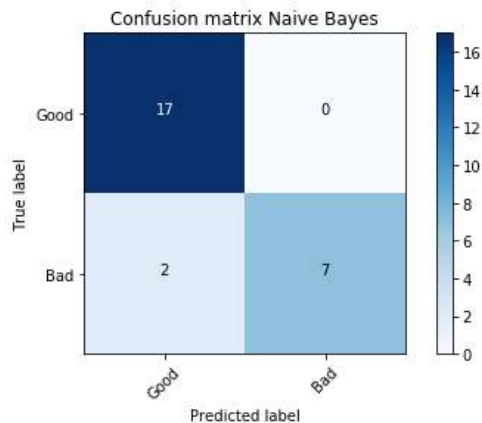| Parameter \\ Value	Speed	Gas Mileage
Low	63	61
Average	39	39
High	33	84
	Knocking tendency	
Yes	64	
No	71	
	Quality of fuel	
Good	79	
Poor	56	

4.3   Experiment results

Figure 4: Random Forest

Figure 5: Naïve Bayes

Figure 6: Multi-Layer Percetpron

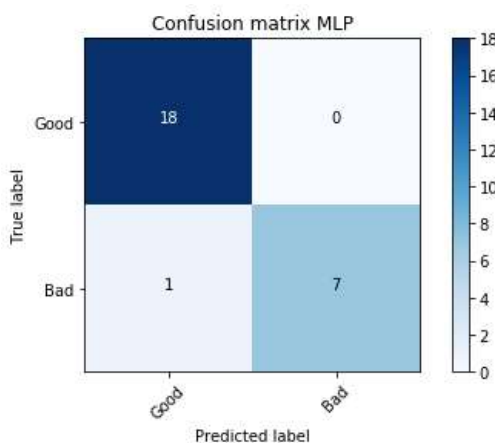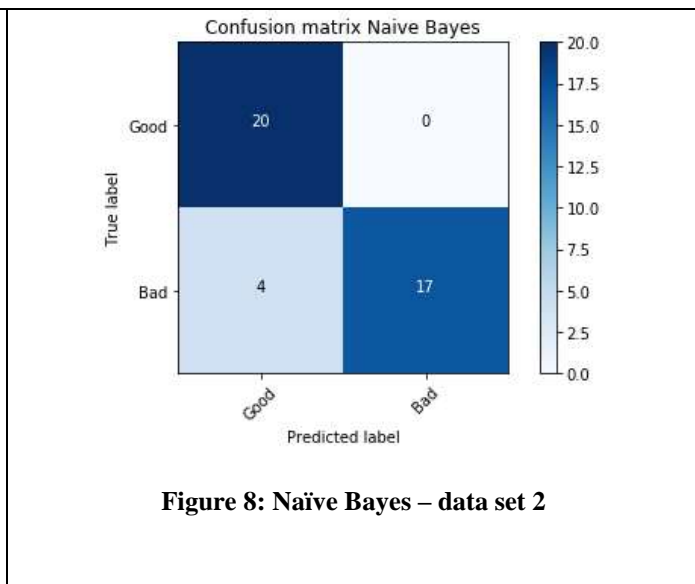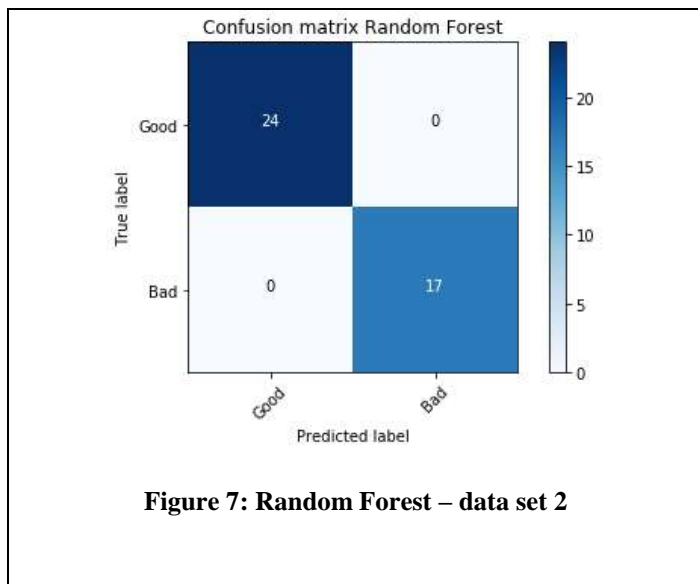| **Figure 7: Random Forest – data set 2**	**Figure 8: Naïve Bayes – data set 2**

Figure 4, 5 and 6 displays confusion matrix that describes the performance of the three classification models in predicting number of vehicles with poor (bad) quality of fuel for data set 1, figure 7 and 8 represent data set 2.

4.4 Discussion

Table 3: Summary of metrics on poor (bad) quality of fuel

Metric / Classifier	Data set 1				
	Accuracy	Precision	Recall	F1-Score	Support
Random Forest	1.0	1.00	1.00	1.00	7
Naïve Bayes	0.92	1.00	0.78	0.88	9
Multi-Layer Perceptron	0.96	1.00	0.88	0.93	8
	Data set 2				
Random Forest	1.0	1.00	1.00	1.00	17
Naïve Bayes	0.90	1.00	0.81	0.89	21

From the experiment, all classifiers performed well but Random Forest gave the best results with precision and recall of 1.00 for both data sets; it means that every prediction retrieved was relevant and that the prediction was complete.

In data set 1, Naïve Bayes and Multi-layer Perceptron had a high precision and low recall, the compromise results to use of F1-score which gives a balance between the two; in this case, Multi-layer perceptron performed better that Naïve Bayes. However, both scores were above 0.5 hence are at acceptable level. In data set 2, Multi-layer Perceptron performed as well as Random Forest with accuracy and F1-score of 1.00 while Naïve Bayes had a lower accuracy of 0.90 and F1-score of 0.89.

Therefore we concluded that Random Forest undoubtedly remains a best performer because of its accuracy void of basic mistakes common to other methods; evidence based on the two data sets.

5. CONCLUSION AND FUTURE WORK

The future work could be geared towards inclusion of other parameters that are important in evaluating the performance of a vehicle; most importantly to use on-board data to precisely predict fuel adulteration and other faults for preventive maintenance.

REFERENCES

[1]. H. Saghaei, "Design and Implementation of a Fleet Management System Using Novel GPS/GLONASS Tracker and Web-Based Software", arXiv: 1610.02667 [cs.CY], 2016

[2]. R. Ramani, S Selvaraju, S Valarmathy and P Niranjan, "Bank Locker Security System based on RFID and GSM Technology", International Journal of Computer Applications 57(18):15-20, November 2012.

[3]. M. Mukhtar, "GPS based Advanced Vehicle Tracking and Vehicle Control System", International Journal of Intelligent Systems Technologies and Applications 03(01):1-12 · February 2015

[4]. S. Singh, F. Morab, S. Thazeen and M. Saqhib, " Tracking And Checking Cargo Containers Pilferage Using Electronic Lock", International Journal of Technical Research and Applications e -ISSN: 2320-8163, Volume 2, Issue 5 (Sep -Oct 2014), PP. 113-116

[5]. K. Sawant, I. Bhole, P. Kokane, P. Doiphode, Y. Thorat, "Accident Alert and Vehicle Tracking System", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 4, Issue 5, May 2016

[6]. A. Pai, V. Vernekar, G. Kudchadkar, S. Arsekar, K. Tanna, R. Rebello, M. Desai, " Real Time Collision Detection and Fleet Management System", ICT and Critical Infrastructure: Proceedings of the 48th Annual Convention of Computer Society of India- Vol I pp 671-678, 2014

[7]. D. Jose, S. Prasad, V. G. Sridhar V. G, "Intelligent Vehicle Monitoring Using Global Positioning System and Cloud Computing", 2nd International Symposium on Big Data and Cloud Computing , ScienceDirect, Procedia Computer Science 50 (2015) 440 – 446

[8]. S. Aher, R. D. Kokate R. D, " Fuel Monitoring and Vehicle Tracking", International Journal of Engineering and Innovative Technology (IJEIT) Volume 1, Issue 3, March 2012

[9]. Almer H, "Machine learning and statistical analysis in fuel consumption prediction for heavy vehicles". KTH computer science and communication, 2015, Second Level, Sweden

[10]. Capraz A. G, Özel P, Şevkli M, Beyca O. F, "Fuel Consumption Models Applied to Automobiles Using Real-time Data: A Comparison of Statistical Models". The 6th International Conference on Sustainable Energy Information Technology. Procedia Computer Science 83 (2016) 774 – 781 774 – 781

[11]. Linda O, Manic M, "Improving Vehicle Fleet Fuel Economy via Learning Fuel-Efficient Driving Behaviors". 2012 5th International Conference on Human System Interactions

[12]. Wickramanayake S, Bandara D, "Fuel consumption prediction of fleet vehicles using Machine Learning: A comparative tudy".Conference: 2016 Moratuwa Engineering Research. DOI: 10.1109/MERCon.2016.7480121

[13]. Prytz R (2014). *Machine Learning Methods for Vehicle Predictive Maintenance Using Off-Board and On-Board Data*

[14]. Prytz R et al (2014). *Analysis of Truck Compressor Failures Based on Logged Vehicle Data*

[15]. O. Obodeh, N.C Akhere, "Experimental study on the effects of kerosene-doped gasoline on gasoline-powered engine performance characteristics", Journal of Petroleum and Gas Engineering Vol. 1(2),pp. 37-40, June 2010

[16]. Yasin A, Abuhasan A (2016). "An Intelligent Classification Model for Phishing Email Detection", International Journal of Network Security & Its Applications (IJNSA) Vol.8, No.4, July 2016

[17]. Srinath K. R," Python –The Fastest Growing Programming Language". International Research Journal of Engineering and Technology (IRJET) Volume: 04 Issue: 12 | Dec-2017