IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Flagging Vehicle Failures With Low Engine Oil Pressure Using Rule Based Algorithm

[1]Poonam Narkhede, [2]Priyanka Bamne, [3]Vandan Tiwari

[1]Student, [2]Assistant Professor, [3]Head of the Department
[1]Department of Computer Engineering,
[1]Shri Govindram Seksaria Institute of Technology and Science, Indore, India

Abstract: Failed engines in commercial vehicles result in loss of resources, time and money for the fleet operators and safety concerns for vehicle operators. Since the first observations in the 1900s, low engine oil pressure has been the most frequent cause of engine failure [1]. With time, indicators and the sensors have been added to check oil levels and communicate abnormal engine oil pressure to the vehicle operators via the engine controls units. With the advent of advance sensors in the 2100s, this data began to be shared on central servers for tracking and maintaining vehicles remotely. Using data analytic and model-based predictions have attempted to make necessary interventions to reduce downtime by conducting preventive maintenance. In this research, the implementation begins with rule-based labeling time series-data from lab tests and field observations. Next, to apply a model-based approach, parameters like engine oil pressure, engine speed, vehicle speed, and engine coolant temperature were identified as parameters indicating engine status. We attempt to use Different Machine learning models to classify the engine failed vehicle due to engine oil pressure.

Index Terms **- Oil pressure, Engine failure, Rule based, commercial vehicles, anomaly detection.**

I. Introduction

Commercial vehicles play a crucial role in the transportation of goods and services. Oil Pressure is one of the most crucial factors in efficient working of commercial vehicles.[1] Oil pressure is defined as the pressure of oil that is being pumped through the engine, which is critical for the proper functioning of the engine and the longevity of its components. The oil pressure is responsible for lubricating the engine components, and if it falls below the recommended level, it could lead to serve engine damage, eventually resulting in engine failure. On the other hand, if the oil pressure is too high, it may cause excessive oil consumption, engine damage, and reduced fuel efficiency [2].

II. Background

In the 1980s and 1990s, researchers began to focus on developing diagnostics tools and the techniques that could predict engine oil pressure failure before they occur. This included the use of sensors and monitoring devices [3]. Research on engine failure prediction had begun to shift towards more quantitative methods, such as using mathematical models to predict the likelihood of engine failure based on oil pressure levels [4]. These models were based on statistical techniques, such as regression analysis, and were used to identify the critical oil pressure levels that were associated with increased risk of engine failure. In the 2000s, researchers began to explore the use of advanced data analysis techniques, such as machine learning and artificial intelligence, to improve the accuracy and reliability of engine oil pressure failure prediction [5]. These studies have focused on developing models that can analyze large amount of data from a variety of sources, such as engine sensors, telematics systems, and maintenance records, to identify patterns and trends that indicates an increased risk of engine oil pressure failure. In the 2000sand 2010s, research on engine failure prediction continued to evolve, with a focus on developing more advanced mathematical models and incorporating data from other vehicle systems, such as the engine failure, such as those caused by wear and tear, overheating, and fuel system problem [4].

III. EXISTING WORK

Today, many commercial vehicle manufacturers vehicle manufacturer and fleet operators use advanced engine oil pressure monitoring systems to predict and prevent engine oil pressure failures. These systems use a combination of sensors, data analysis and predictive algorithms [5] to identify potentials issues and alert the operators to take preventive action before a failure occurs. In recent years, research on engine failure prediction has also focused on developing advanced diagnostic tools and systems that can monitor the health of an engine in real-time. These systems use sensors and other technologies to monitor oil pressure levels and other key engine parameters and provide early warning of potential problems before they lead to engine failure. Overall, the research on engine failure prediction due to engine oil pressure in commercial vehicles has been ongoing for several decades, with a focus on identifying more advanced mathematical models and diagnostics tools for predicting and preventing engine failure. The use of rule-based prediction models for engine failure due to oil pressure in time series data collected for commercial vehicles has been widely studied in the field of predictive maintenance [4]. There are various methods for predicting engine failure due to oil pressure, including statistical methods, machine learning algorithms, and rule-based systems. Rule based systems are preferred in the case of commercial vehicles due to their interpretation ability and ease of implementation. These systems use a set of predefined rules that are based on expert knowledge to make predictions. The rules are usually expressed in the form" if-then" statements, where the if" if" part specifies the conditions that need to be satisfied and the" then" part specifies the action to be taken.

IV. METHODS

Traditionally, engine failure in vehicles has been established by designing rules based on empirical observations. However, the purpose of this paper is to explore the use of machine learning approaches to predict engine failure [6]. Instead of relying solely on predefined rules, machine learning algorithms analyze the collected telematics data and indicators patterns that may be associated with engines failure [6]. By leveraging these machine learning techniques, the researchers aim to develop predictive models that can detect signs of potential engine failure in advance. In this research we have analyzed telematic data collected from commercial vehicle engines operating in field. By utilizing machine learning algorithms, the models can learn from the vast amount of data collected from telematics devices [6], identifying complex relationship and patterns that may not be evident through manual rule design alone [1]. This paper likely describes different machine learning approaches that were employed and evaluated to predict engine failure. These approaches may include technique such as classification algorithms, regression models, or other methods suitable for analyzing the telematics data and predicting engine failure based on the identified patterns. Overall, the papers represent an attempt to leverage machine learning to enhance the prediction of engine failure due to low oil pressure in commercial vehicles, moving beyond traditional rule-based methods and utilizing the rich data provided by telematics devices.

V. DATA DESCRIPTION

	count	mean	std	min	max
Var_1	26814	1219.8403	397.39673	0	3301.5
Var_2	26814	39.063325	22.216423	0	138.82
Var_3	26814	309.6205	105.66274	0	748
Var_4	26814	83.991124	9.4477303	0	100
Var_5	26814	91.656784	135.226	0	591.25
Var_6	26814	97.89416	2.7316322	66.5	103
Var_7	26814	416.97339	725.07051	15.25	1774.9688
Var_8	26814	75.062691	101.31235	-40	215
Label	26814	1.0040277	0.0633378	1	2

Fig 1. Data Description

VI. Model Approach

4.1 Correlations

By examining the correlation map strongly related variables or variables that exhibit interesting patterns are identified [7]. Variables that have a strong correlation with the target variables are highlighted. Also, correlation maps are used to detect multicollinearity by identifying a high correlation between two variables. Addressing multicollinearity is essential for building a robust and interpret-able model. Next, redundant or highly correlated features that can be identified need to be removed so as to reduce the dimensions. By understanding the relationships between variables, variable inclusion or exclusion can be decided. Variables that have the most impact and how changes in one variable may affect others is studied. This information is very crucial for making data-driven decisions and formulating strategies based on the insights gained from the correlation map. Following this process, a new rule to predict engine failure is determined. It was observed that while speed of engine, speed of vehicle, oil pressure, and coolant temperature are strongly correlated to engine failure; intake air temperature and CatlistupstrmTemp are also strongly correlated with engine failure.

4.2 ML Classifier

Algorithms like K-Nearest Neighbours, Support Vector Machines [8], Decision Tree, Random Forest and Ada Boost classifiers [9] were applied and compared in this supervised machine learning approach. The data set used for this approach was from real time engine condition data collected from multiple sensors, at per minute frequency. The primary focus of ML Classifiers is to improve the accuracy of event recognition and reduce the occurrence of false positive. To evaluate the effectiveness if ML classifiers, following matrices [10] are used:

1) TP Rate:

$$TPR = TP / (TP + FN)$$

2) FP Rate:

$$FPR = FP / (FP + TN)$$

3) TN Rate:

$$TNR = TN / (TN + FP)$$

4) FN Rate:

$$FNR = FN / (FN + TP)$$

5) ACC:

$$ACC = (TP + TN) / (TP + FP + TN + FN)$$

6) Precision:

$$Precision = TP / (TP + FP)$$

7) Recall:

$$Recall = TP / (TP + FN)$$

• AUC: Area Under ROC Curve In the context of evaluating the model, the following definitions are relevant:

• TP: Instances where abnormal events are correctly identified.

• FP: Instances where normal conditions are mistakenly judged as abnormal events.

• TN: Instances where normal conditions are accurately identified.

• FN: Instances where abnormal events are incorrectly classified as normal.

• AUC: quantifies the performance of a model by measuring the area under the ROC curve.

By analyzing the ROC curve and calculating the AUC value, we can intuitively gauge the model's performance. The AUC value ranges between 0.1 and 1. Larger AUC value signifies a better-performing model in terms of accurately identifying abnormal events, minimizing false positive, accurately identifying normal conditions, and reducing false negatives [11]. By using the above matrices, the following results are produced using telematics data.

Label	KNN	SVM	DT	RF	AdB
True Negative	15126	15129	14956	15129	15119
False Positive	3	0	173	0	10
False Negative	182	183	146	183	176
True Positive	1	0	37	0	7
Precision	0.988111	0.988049	0.990332	0.988049	0.988493
Recall	0.999802	1	0.988565	1	0.999339
F1-score	0.993922	0.993988	0.989448	0.993988	0.993886

Fig 2. Confusion Matrix of ML Classifiers

4.3 PCA

The data set contains a large number of features or variables, this leads to confusion, false representation and computational inefficiency. It may also lead to model over-fitting. The approach of PCA helps in dimensional reduction of the input features by transforming them into a new set of orthogonal variables. It enables the visualization of high-dimensional data in a lower-dimensional space. It achieves data compression by transforming data into a reduced set of principal components that capture the most significant variation, thereby saving memory and computational resources. With PCA, we can effectively identify and remove multi collinearity in data sets containing correlated variables. By transforming the features into orthogonal components, we eliminate redundancy and improve model stability. The number of principle components were limited to five.

Label	KNN	SVM	DT	RF	AdB
True Negative	18163	18165	17921	18165	18158
False Positive	2	0	244	0	7
False Negative	209	210	206	210	210
True Positive	1	0	4	0	0
Precision	0.988624	0.988571	0.988636	0.988571	0.988567
Recall	0.99989	1	0.986568	1	0.999615
F1-score	0.994225	0.994253	0.987601	0.994253	0.99406

Fig 3. Confusion Matrix of ML Classifiers Using PCA

The result of this approach is as same as the above ML Classifiers (ref Fig.3). There is no difference between in precision, recall and F1-score using PCA and without PCA. (Ref Fig.2 and Fig.3)

4.4 Isolation Forest

Isolation Forest is a Unsupervised Machine Learning algorithm specifically designed for anomaly detection [10]. It is particularly effective in identifying outliers or anomalies in data sets. This approach focuses on isolating anomalies and is especially suitable for low engine oil pressure because anomalies in oil pressure often have different patterns or characteristics compared to normal oil pressure.

Since engine failure due to low engine oil pressure can be considered anomalous to the regular engine functioning, anomaly detection approach was applied using isolation forest [10]. A threshold of 5 percent to define outliers and engine oil pressure below 150 kpa was used as threshold.

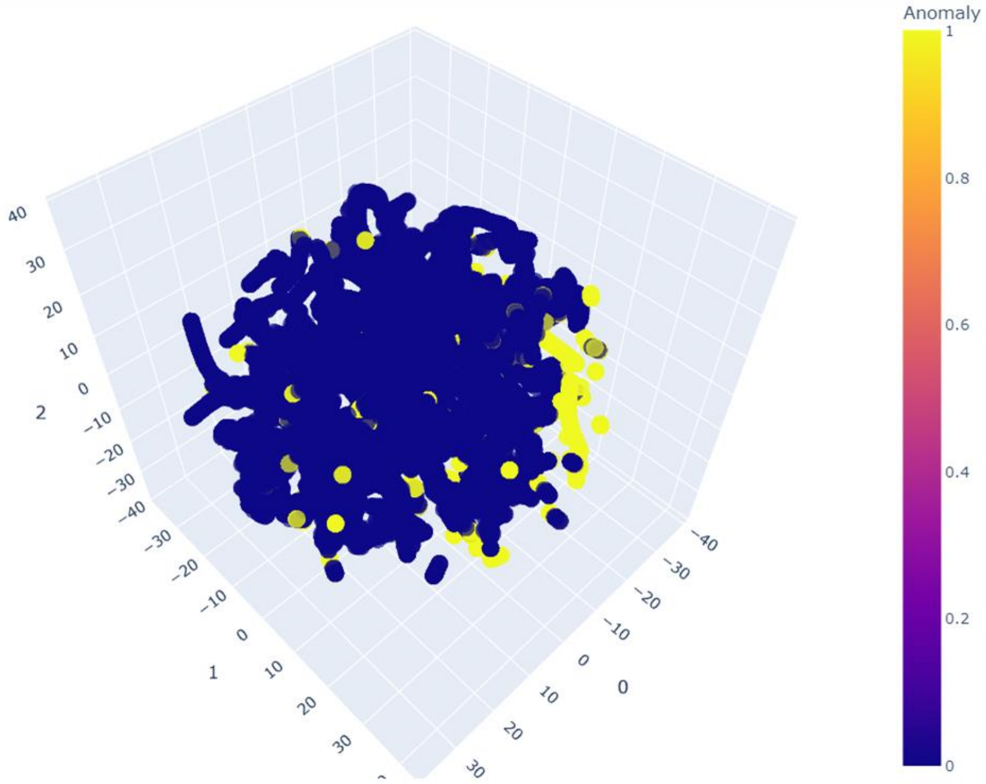

Fig 4. TSNE Graphical Representation of Anomalies

The above TSNE graph visualize the data points in lower dimensional space, typically 3D as per the given rules. It identifies the clusters or patterns of normal data points. Normal instances tend to be grouped closely together in the lower dimensional space, forming distinct clusters.

But Anomalies, might appear as isolated or distinct points in above TSNE graphs, separate from normal clusters. These outliers can be easily identified by their noticeable deviation from the majority of data points.

```
array([[0.65212897, 0.34787103],
       [0.64817247, 0.35182753],
       [0.53909238, 0.46090762],
       [0.59425749, 0.40574251],
       [0.59231485, 0.40768515],
       [0.50235924, 0.49764076],
       [0.60853742, 0.39146258],
       [0.5773792 , 0.4226208 ],
       [0.49279111, 0.50720889],
       [0.4265964 , 0.5734036 ],
       [0.21032117, 0.78967883]])
```

Fig 5. Degree of Anomality

The above figure shows that the higher scores indicate a higher likelihood of being an anomaly, while lower scores suggest a higher likelihood of being a normal data point. To interpret this result of Isolation Forest anomaly detection, we have set a threshold on anomaly scores.

4.5 Multimodal

By combining multiple machine learning models, accuracy to identify anomalous engine failure events due to low oil engine oil pressure can be improved [11, 12]. First, Multiple Linear Regression model is used to create linear regression models, each tailored to a specific subset of data. In the context of low engine oil pressure detection, different types of oil pressure values may exhibit distinct patterns or behaviors. By building separate linear regression models for different oil pressure categories, it becomes possible to capture the specific characteristics and relationship within each subset of data.

Once the relevant features are selected and separated, linear regression models are trained for each low engine oil pressure category. The models can be used to predict the likelihood of low oil pressure for new incoming events. The models produce a low oil pressure score or probability for each transaction, indicating the likelihood of low oil pressure activity based on the learned pattern.

After obtaining the scores, isolation forest model excels to identify outliers or anomalies within data set.

The anomaly score obtained from isolation forest can be compared against a predefined threshold to classify the low oil pressure as normal or abnormal. The threshold can be based on desired balance between false positive and false negative, considering the risk tolerance of low engine oil pressure detection system [13].

By combining the outputs of multi model linear regression and isolation forest, we leverage the strength of both techniques. Linear regression models capture the specific low oil pressure patterns within each category, while isolation forest identifies unusual patterns that may not be captured but linear regression models. The ensemble approach optimizes anomaly detection, improving the overall accuracy and coverage of the oil pressure detection system [14].

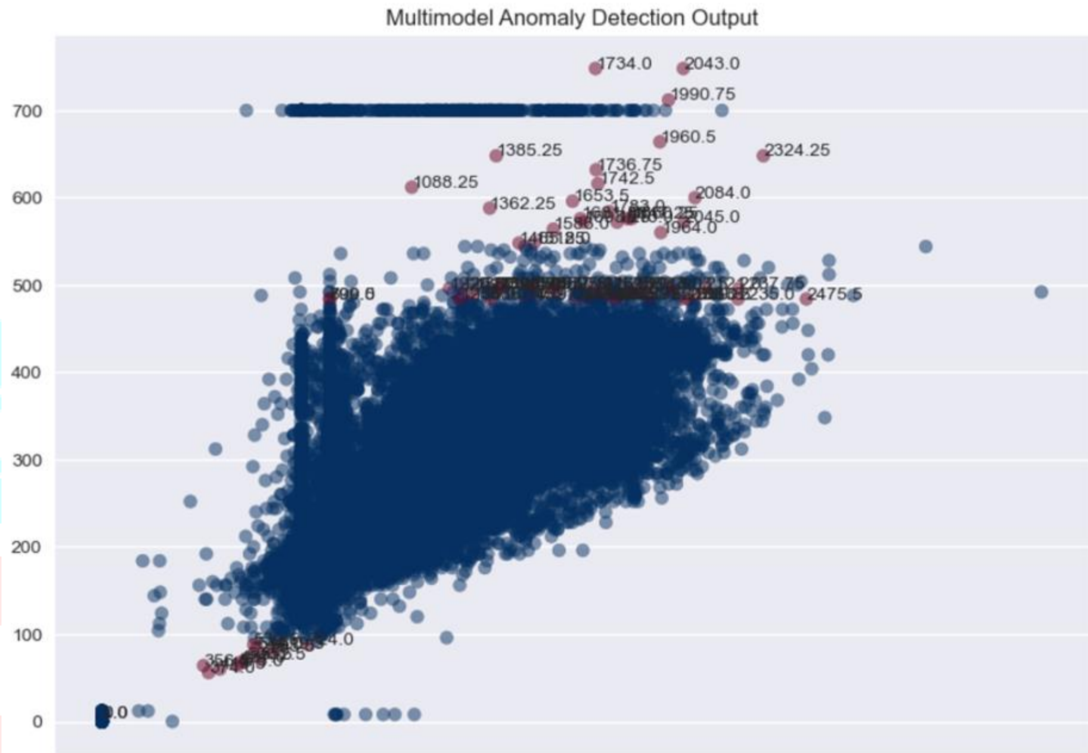

Fig 6. Graphical Representation of Anomalies Using Multimodal

VII. Results

In the mentioned approaches, various parameters such as oil pressure, engine speed, vehicle speed, coolant temperature, IST date time, and intake air temperature were utilized. Surprisingly, when comparing the results obtained using all these parameters with those obtained using fewer parameters, they turned out to be quite similar, except for the Multi model approach.

The Multi model approach stood out as it successfully identified anomalies, as shown in Figure 6. These identified anomalies were later cross validated manually using the dataset. However, it was noticed that the absence of oil level data limited the accuracy of the confusion matrix results.

In simpler terms, the different methods used similar sets of parameters, but the Multi model approach was unique in pinpointing anomalies. Nevertheless, the absence of oil level data hindered achieving even more precise results, and including this data could potentially enhance the accuracy of the analysis.

VIII. Conclusion

In conclusion, based on the comparison of multiple approaches, the statement suggest that the Isolation Forest algorithm along with linear regression i.e.Multi model, stands out as the most effective methods for engine failure prediction using anomaly detection. Its ability to detect anomalies without the need for labeled training data, scalability, and robustness against outlier make it a reliable choice for identifying potential engine failures and enabling proactive maintenance.

IX. FUTURE WORK

Multi-modal approaches considering seasonal effects, global positions, terrains and ambient environments should also be explored to increase model-based prediction further.

References

[1] Roger Moment, \emph{"Oil Pressure Diagnostics Using Gauge Readings to Diagnose your Engine."}

[2] Troubleshooting manual by Laycoming

[3] Fleming W.I., \emph{"Overview of Automotive Sensors."}, IEEE Sensors. Journal, Vol I, No. 4, December 2001.

[4] Vehicle Diagnostics - The Whole Story

[5] Song You, Mark Krage and Laci Jalics, \emph{"Overview of Remote Diagnosis and Maintenance for Automotive Systems."} SAE World Congress, Detroit, Michigan, April 2005.

[6] Siddharth H.D'Silva, \emph{"Diagnostics based on the Statistical Correlation of Sensors"} SAE International 11 Aug 2018.

[7] Jang, G. B., Cho, S. B. (2019). \emph{"Anomaly detection of 2.4l diesel engine using one-class svm with variational autoencoder"}. In N. S. Clements, B. Zhang, A. Saxena (Eds.), Proceedings of the Annual Conference of the Prognostics and Health Management Society, PHM (1 ed.). (Proceedings of the Annual Conference of the Prognostics and Health Management Society, PHM; Vol. 11, No. 1). Prognostics and Health Management Society. https://doi.org/10.36001/phmconf.2019.v11i1.804

[8] M.A. Taie, M. diab, and M. Elhelw., \emph{"Remote Prognosis, diagnosis and maintenance for automotive architecture based on least squares support vector machine and multiple classifiers"}, Proceedings of the 2012 4th International Congress on Untra Modern Telecommunication and Control Systems, ICUMT 2012 pp. 128-134, Russia, October 2012. DOI10.1109/ICUMT.2012.6459652

[9] Liu Liu and Min Xiao (2022). \emph{"A Study of an Engine Anomaly Detection Model IForest-ADOA."} (Journal of Physics: Conference Series. 2171 0120756. DOI 10.1088/1742-6596/2171/1/012076

[10] E. Bussey, . H. ., G. Nenadic, N. ., A. Ardis, P. ., G. Thurston, . M. . (2014). \emph{"Case Study: Models for Detecting Low Oil Pressure Anomalies on Commercial Vehicles."} Annual Conference of the PHM Society, 6(1). https://doi.org/10.36001/phmconf.2014.v6i1.2416

[11] Mylaraswamy, D. (2005). \emph{"Anomaly Detection in a Multi-engine Aircraft."} In: Pal, S.K., Bandyopadhyay, S., Biswas, S. (eds) Pattern Recognition and Machine Intelligence. PReMI 2005. Lecture Notes in Computer Science, vol 3776. Springer, Berlin, Heidelberg. https://doi.org/10.1007/115903162

[12] Mahdi Rezapour, \emph{"Anomaly Detection using Unsupervised Methods: Credit Card Fraud Case Study"} International Journal of Advanced Computer Science and Applications(IJACSA), 10(11), 2019. http://dx.doi.org/10.14569/IJACSA.2019.0101101

[13] Bopche B.N. (2021). \emph{"A Review Paper on Optimization of Engine oil Pressure After Overhauling in HINO 6DTI Engine."} (International Journal of Research in Engineering and Science) Volume 9 Issue 3 2021 PP. 66-70 www.ijres.org

LIST OF ABBREVIATION

ML	Machine Learning
KNN	K-Nearest Neighbors
SVM	Support Vector Machine
DT	Decision Tree
RF	Random Forest
AdB	AdaBoost
PCA	Pricniple Component Analysis
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
AUC	Area Under the Curve
ROC	Receiver Operating Characteristics