BALTIC JOURNAL OF LAW & POLITICS
A Journal of Vytautas Magnus University
VOLUME 15, NUMBER 4 (2022)
ISSN 2029-0454

sciendo

Cite: *Baltic Journal of Law & Politics* 15:4 (2022): 325-332
DOI: 10.2478/bjlp-2022-004034

Improved Detection of Truck Failure Due to Air Pressure System by Novel Xgboost Algorithm over Decision Tree Algorithm

Ch. Dheeraj
Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India, Pincode:602105

T.P. Anithaashri
Project Guide, Corresponding Author, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Science, Saveetha University, Chennai, Tamilnadu, India, Pincode:602105

Received: August 8, 2022; reviews: 2; accepted: November 29, 2022.

Abstract

Aim: To improve the detection of a truck failure due to an Air Pressure System (APS) using Machine learning algorithms and improve the accuracy of APS truck failure using the Novel XGBoost algorithm. **Materials and Methods**: The Novel XGBoost algorithm is used for the prediction of truck failure due to the Air Pressure System. The dataset has a total sample size of 10 for each group and SPSS package utilized for the performance analysis of accuracy in detecting the air pressure system. To improve the accuracy to detect truck failure due to an Air Pressure System using Novel XGBoost is proposed and compared with the decision tree algorithm. **Results and Discussion**: Test results prove that Novel XGBoost has an average accuracy of 98.24% which is better than the Decision Tree has an average accuracy of 96.62%. The analysis of the training dataset and testing dataset has been performed successfully using SPSS and acquired 98.24% accuracy for predicting the truck failure due to the Air pressure system. With the level of significance ($p<0.05$) the resultant data depicts the reliability of independent sample tests. **Conclusion**: The overall process of prediction of accuracy using the Novel XGBoost algorithm gave significantly better results compared to the Decision Tree algorithm.

Keywords

Truck Failure, Air Pressure System, Decision Tree, Novel XGBoost, Machine Learning, Deep Learning

INTRODUCTION

The APS is a critical component of a heavy-duty vehicle (United States. Department of Transportation 2000) that uses compressed air (Richards 1895) to force a piston to provide pressure to the brake pads, slowing the vehicle down. The benefits of using an APS (Ramaratham 2008) instead of a hydraulic system are the easy availability and long-term sustainability of natural air. In the industrial sector, heavy vehicles are the dominant model of transportation. They are the most adaptable and cost-effective means of transportation, and they are used in every industry on a daily basis. A thorough maintenance program (Baird 1943; *Trucks: Light, Medium & Heavy Duty Truck*

 ISSN 2029-0454
2022

Maintenance 1987) is essential for avoiding unwelcome breakdowns and thus saving money and effort. In this context, it is critical that all vehicle components be maintained on a regular basis. The research work carried out by the authors (Gondek, Hafner, and Sampson 2016) using a feature engineering and random forest algorithm has made me carry out this research. In this paper (Gondek, Hafner, and Sampson 2016) performed dimensionality reduction to increase the accuracy of the model and performed feature engineering and some other techniques for improving the accuracy. Truck APS failure prediction by these authors (Komatineni Sri Sai Nikhil, Erla Vinay Kumar) had done a good approach using algorithms like SVM, Novel XGBoost, and Random forest after applying a SMOTE technique. The Air Pressure System is one such crucial component (APS). The APS produces pressured air that is used for a variety of activities such as braking, gear shifting, and so on, making it a critical component to maintain.

The existing system (Lokesh et al. 2020) deals with only accuracy as metrics; it is not in the case of imbalanced data. The proposed method (Fernández et al. 2018) has been included to overcome imbalanced data problems.The given data contains failure of components related to the Air Pressure System and Failure of components not related to the Air Pressure System. To handle the missing values in the data please refer to this paper for imputation of missing values (Rafsunjani et al. 2019) of Air Pressure System failure prediction. Efficient work in increasing the accuracy of the model by applying many techniques using SMOTE. There are many steps included in the training of models such as feature engineering (Geigle, Mei, and Zhai 2018) and feature selection. The Novel XGBoost method (Brownlee 2016) is more general for simpler and complex methods. In such cases, the approach with Novel XGBoost plays a crucial role compared to other algorithms. Deep learning algorithms can be used for existing methods when our data is huge. Some of the deep learning algorithms are the artificial neural network, and convolutional neural network.To evaluate the Machine Learning classification (Hung, Jiang, and Wang 2020) refer to this existing proposed method. The existing approach is taken from a research paper named prediction of failures in the Air Pressure System of Scania trucks using random forest and feature engineering (Gondek, Hafner, and Sampson 2016).

Previously our team has a rich experience in working on various research projects across multiple disciplines (Venu and Appavu 2021; Gudipaneni et al. 2020; Sivasamy, Venugopal, and Espinoza-González 2020; Sathish et al. 2020; Reddy et al. 2020; Sathish and Karthick 2020; Benin et al. 2020; Nalini, Selvaraj, and Kumar 2020). The research gap in the existing system implemented through Machine Learning approaches is well researched, however, outliers handling, skew correction, a huge percentage of missing values, applying deep learning algorithms, including more parameters when performing hyperparameter tuning and some machine learning preprocessing techniques are some of the challenges yet to be resolved. Therefore, our proposed aim is to address the APS Failure prediction with high accuracy for the Novel XGBoost Algorithm. This is further compared with the Decision Tree for better performance.

MATERIALS AND METHODS

The software tool used for detecting the accuracy score is using the pycharm. Hardware configuration was AMD RYZEN 5 (2.10 GHz) processor with 8GB ram and 64 bit OS, x64 based processor system. The software configuration was Windows 11 professional. The data was preprocessed after performing many steps such as removing noise data, feature engineering, and feature selection.

The Novel XGBoost and Decision Tree algorithms were evaluated with mean accuracy along with F1_score with respect to training labels and testing records. Novel XGBoost method was iterated a various number of times with a sample size of 6000 calculated from CSV file and SPSS analysis is carried out with a level of significance $p<0.05$.

 ISSN 2029-0454

Novel XGBoost Algorithm

Novel XGBoost is a Decision Tree-based ensemble Machine Learning algorithm that uses a gradient boosting framework. Small to medium structured data Novel XGBoost is mostly preferred. A wide range of applications: can be used to solve regression and classification. It runs smoothly on Windows, Linux, and os x. It supports all major programming languages like C, python, java, C++. Novel XGBoost approaches the process of sequential tree building using parallelized implementation. Algorithms have been designed to make efficient use of hardware resources.

Pseudocode for Novel XGBoost

```
from  XGBoost import XGBCLASSIFIER
from sklearn.metrics import accuracy_score,f1_score
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler
from imblearn.oversampling import SMOTE
XGB=XGBCLASSIFIER()
X_train,X_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=0)
smote=SMOTE()
X_train_smote,y_train_smote=smote.fit_resample(X_train,y_train)
XGB.fit(X_train_smote,y_train_smote)
prediction=XGB.predict(X_test)
accuracy_score(y_test,prediction)
f1_score(y_test,prediction)
```

Decision Tree

A Decision Tree is a flowchart-like structure in which each internal node represents a test on a feature, each leaf node represents a class label and branches represent conjunctions of features that lead to those class labels. A Decision tree is one of the predictive modeling approaches used in statistics, data mining, and Machine Learning. Decision Trees are constructed via an algorithmic approach that identifies ways to split a dataset based on different conditions. It is one of the most widely used and practical methods for supervised learning. Decision Trees are a non-parametric supervised learning method used for both classification and regression tasks.

Pseudocode for Decision Tree

```
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score,f1_score
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler
from imblearn.oversampling import SMOTE
DT=DecisionTreeClassifier()
X_train,X_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=0)
smote=SMOTE()
X_train_smote,y_train_smote=smote.fit_resample(X_train,y_train)
DT.fit(X_train_smote,y_train_smote)
prediction=DT.predict(X_test)
accuracy_score(y_test,prediction)
f1_score(y_test,prediction)
```

Statistical Analysis

The statistical software used for performing analysis in IBM SPSS version 21.0. IBM SPSS is a statistical software tool used for the analysis of data. The datasets are normalized and then the data is converted into arrays. The number of clusters needed is visualized and analyzed and the existing algorithms are obtained. The independent variables are tyre_pressure, and air_compressor and the dependent variable is a failure. It helped to get the improved accuracy with increased efficiency in terms of taking less time by using the Novel XGBoost algorithm for the Air Pressure System. To check the data and accuracy reliability SPSS is used with a default alpha value of 0.05.

RESULTS

In the innovative Air Pressure System failure research it is proved that the Novel XGBoost algorithm appears to have better accuracy than the Decision Tree algorithm. Statistical analysis is done for comparing Novel XGBoost and decision algorithms for the SPSS tool. Parameters taken for compression of both algorithms are accurate. Finally, descriptive statistics were applied to the dataset in SPSS. From the group statistics mentioned above. The mean standard deviation and standard error mean are compared for the experimental algorithm Novel XGBoost and Decision algorithm with the significance value (<0.05).

Table 1 represents the comparison of accuracy in truck failure due to APS by Novel XGBoost and Decision Tree algorithms, by iterating for various numbers of times.

Table 2 represents the sample size(N=10), Mean, Standard deviation and Standard error mean are classified based on accuracy and loss of data. The accuracy of 98.24% of Novel XGBoost is significantly higher compared to the Decision Tree algorithm.

Table 3 represents the significance of the data and standard error difference, where the significance of Novel XGBoost and Decision Tree method with confidence interval as 95% and level of significance <0.05

Fig. 1**,** represents the mean accuracy between the two algorithms. The Novel XGBoost algorithm appears to produce consistent results with standard deviation.

DISCUSSION

In this study of the Air Pressure Failure system, Novel XGBoost has higher accuracy of 98% compared to the Decision Tree algorithm of 96%. Novel XGBoost has better significance (p<0.05) than Decision Tree and while using the independent sample t-tests. Similar work has been done by authors (Oliver Sampson, Daniel Hafner, Christopher Gondek) for similar concepts that we used in research. Most of the drawbacks of other research papers are they have only included mean accuracy scores (Wiers and de Kok 2017; Rafsunjani et al. 2019). They have not included the f1 score metric (Fernández et al. 2018) but we used it in research in order to get better performance of the model irrespective of accuracy score as a metric. In order to overcome the above drawbacks in the proposed system, further research should also consider other than Machine Learning algorithms (Kononenko and Kukar 2007) such as deep learning algorithms to ascertain more efficient ways to perform better for larger datasets (Fernández et al. 2018; Brownlee 2016; Leskovec, Rajaraman, and Ullman 2014).

It is recommended that research should be carried out on other parameters that can improve the accuracy of prediction. limitations of our proposed method like skew correction, and a huge percentage of missing values. It can be addressed in the future scope. The results of the proposed algorithm are better in both experimental and statistical analysis. The future scope is deep learning methods may be used to solve this particular problem and we can evaluate the neural network using our performance metric, various other imputation methods can be used such as the Soft-Impute algorithm. Deep learning algorithms like artificial neural networks can be applied to existing systems to overcome drawbacks. Novel XGBoost is good in all aspects compared to the Decision Tree. Novel

XGBoost std deviation error is lower than Decision Tree and many others. Hence, Novel XGBoost (Fernández et al. 2018; Brownlee 2016) seems to have a good accuracy rate compared to the Decision Tree performed in SPSS.

CONCLUSION

In the research the Novel XGBoost algorithm seems to appear with a better accuracy percentage (98.24%) in detecting the failure of the Air Pressure System. SPSS is used to depict the accuracy of 98.24% using Novel XGBoost is more efficient than Decision Tree accuracy of 96.62%.

DECLARATIONS

Conflict of Interests

No conflict of interest in this manuscript.

Author Contribution

Author CDK was involved in data collection, data analysis, and manuscript writing. Author TPA was involved in conceptualization, data validation, and critical review of the manuscript

Acknowledgement

The authors would like to express their gratitude towards the Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Funding

We thank the following organizations for providing financial support that enabled us to complete this study:

1. ReadMind Technologies Pvt Ltd,Chennai
2. Saveetha School of Engineering
3. Saveetha university
4. Saveetha Institute of Medical and Technical Sciences.

REFERENCES

Baird, John N. 1943. "Heavy Duty Trailer Maintenance." *SAE Technical Paper Series*. https://doi.org/10.4271/430097.

Benin, S. R., S. Kannan, Renjin J. Bright, and A. Jacob Moses. 2020. "A Review on Mechanical Characterization of Polymer Matrix Composites & Its Effects Reinforced with Various Natural Fibres." *Materials Today: Proceedings* 33 (January): 798–805.

Brownlee, Jason. 2016. *XGBoost With Python: Gradient Boosted Trees with XGBoost and Scikit-Learn*. Machine Learning Mastery.

Fernández, Alberto, Salvador García, Mikel Galar, Ronaldo C. Prati, Bartosz Krawczyk, and Francisco Herrera. 2018. *Learning from Imbalanced Data Sets*. Springer.

Geigle, Chase, Qiaozhu Mei, and Chengxiang Zhai. 2018. "Feature Engineering for Text Data." *Feature Engineering for Machine Learning and Data Analytics*. https://doi.org/10.1201/9781315181080-2.

Gondek, Christopher, Daniel Hafner, and Oliver R. Sampson. 2016. "Prediction of Failures in the Air Pressure System of Scania Trucks Using a Random Forest and Feature Engineering." *Lecture Notes in Computer Science*. https://doi.org/10.1007/978-3-319-46349-0_36.

Gudipaneni, Ravi Kumar, Mohammad Khursheed Alam, Santosh R. Patil, and Mohmed Isaqali Karobari. 2020. "Measurement of the Maximum Occlusal Bite Force and Its Relation to the Caries Spectrum of First Permanent Molars in Early Permanent Dentition." *The Journal of Clinical Pediatric Dentistry* 44 (6): 423–28.

Hung, Che-Yu, Bernard C. Jiang, and Chien-Chih Wang. 2020. "Evaluating Machine Learning Classification Using Sorted Missing Percentage Technique Based on Missing Data." *Applied Sciences*. https://doi.org/10.3390/app10144920.

Kononenko, Igor, and Matjaz Kukar. 2007. *Machine Learning and Data Mining*. Elsevier.

Leskovec, Jure, Anand Rajaraman, and Jeffrey David Ullman. 2014. *Mining of Massive Datasets*. Cambridge University Press.

Lokesh, Yangalasetty, Komatineni Sri Sai Nikhil, Erla Vinay Kumar, and B. Gopesh Krishna Mohan. 2020. "Truck APS Failure Detection Using Machine Learning." *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*. https://doi.org/10.1109/iciccs48265.2020.9121019.

Nalini, Devarajan, Jayaraman Selvaraj, and Ganesan Senthil Kumar. 2020. "Herbal Nutraceuticals: Safe and Potent Therapeutics to Battle Tumor Hypoxia." *Journal of Cancer Research and Clinical Oncology* 146 (1): 1–18.

Rafsunjani, Siam, Department of Computer Science, Faculty of Information Technology, American International University-Bangladesh, Rifat Sultana Safa, Abdullah Al Imran, Shamsur Rahim, and Dip Nandi. 2019. "An Empirical Comparison of Missing Value Imputation Techniques on APS Failure Prediction." *International Journal of Information Technology and Computer Science*. https://doi.org/10.5815/ijitcs.2019.02.03.

Ramaratham, Srivatsan. 2008. *A Mathematical Model for Air Brake Systems in the Presence of Leaks*.

Reddy, Poornima, Jogikalmat Krithikadatta, Valarmathi Srinivasan, Sandhya Raghu, and Natanasabapathy Velumurugan. 2020. "Dental Caries Profile and Associated Risk Factors Among Adolescent School Children in an Urban South-Indian City." *Oral Health & Preventive Dentistry* 18 (1): 379–86.

Richards, Frank. 1895. *Compressed Air: Practical Information Upon Air-Compression and the Transmission and Application of Compressed Air*.

Sathish, T., and S. Karthick. 2020. "Gravity Die Casting Based Analysis of Aluminum Alloy with AC4B Nano-Composite." *Materials Today: Proceedings* 33 (January): 2555–58.

Sathish, T., D. Bala Subramanian, R. Saravanan, and V. Dhinakaran. 2020. "Experimental Investigation of Temperature Variation on Flat Plate Collector by Using Silicon Carbide as a Nanofluid." In *PROCEEDINGS OF INTERNATIONAL CONFERENCE ON RECENT TRENDS IN MECHANICAL AND MATERIALS ENGINEERING: ICRTMME 2019*. AIP Publishing. https://doi.org/10.1063/5.0024965.

Sivasamy, Ramesh, Potu Venugopal, and Rodrigo Espinoza-González. 2020. "Structure, Electronic Structure, Optical and Magnetic Studies of Double Perovskite Gd2MnFeO6 Nanoparticles: First Principle and Experimental Studies." *Materials Today Communications* 25 (December): 101603.

Trucks: Light, Medium & Heavy Duty Truck Maintenance. 1987.

United States. Department of Transportation. 2000. *Medium- and Heavy-Duty Vehicle R&D Strategic Plan*.

Venu, Harish, and Prabhu Appavu. 2021. "Experimental Studies on the Influence of Zirconium Nanoparticle on Biodiesel–diesel Fuel Blend in CI Engine." *International Journal of Ambient Energy* 42 (14): 1588–94.

Wiers, Vincent C. S., and A. (ton) G. de Kok. 2017. *Designing, Selecting, Implementing and Using APS Systems*. Springer.

Tables and Figures

Table 1. Comparison between Novel XGBoost Classifier and Decision Tree Classifier with N=10 samples of the dataset with the highest performance of 98.42 and 96.40.

No of iterations	Novel XGBoost Classifier accuracy in %	Decision Tree Classifier accuracy in %
1	99.42	97.90
2	99.12	97.72
3	98.97	97.51
4	98.63	97.10
5	98.34	96.54
6	98.12	96.34
7	98.24	96.61
8	97.54	95.89
9	97.31	95.60
10	97.10	95.34

Table 2. Group statistics results (Mean of Novel XGBoost algorithm 98.24 appears to more compared to Decision Tree algorithm 96.62) and Standard Error Mean for Novel XGBoost is .25111 and Decision Tree is .28697

	Groups	N	Mean	Std.Deviation	Std.Error Mean
Accuracy	Novel XGBoost	10	98.2410	.79408	.25111
	Decision Tree	10	96.6170	.90749	.28697

Table 3 Independent Sample T-test Result is applied for dataset fixing confidence interval as 95% and level of significance as (<0.05) (Novel XGBoost appears to perform significantly better than Decision Tree algorithm)

	Levene's test for equality of variances	T-test `for Equality of Means	95% confidence interval of the difference

ISSN 2029-0454
2022

		F	Sig	t	df	Sig(2 tailed)	Mean diff	Std. error	Lower	Upper
Accuracy	**Equal variances assumed**	2.321	.0378	12.259	18	.001	4.62400	.38130	4.82286	5.42514
	Equal Variances not assumed			12.25	17.68	.001	4.62400	.38133	4.82185	5.42615

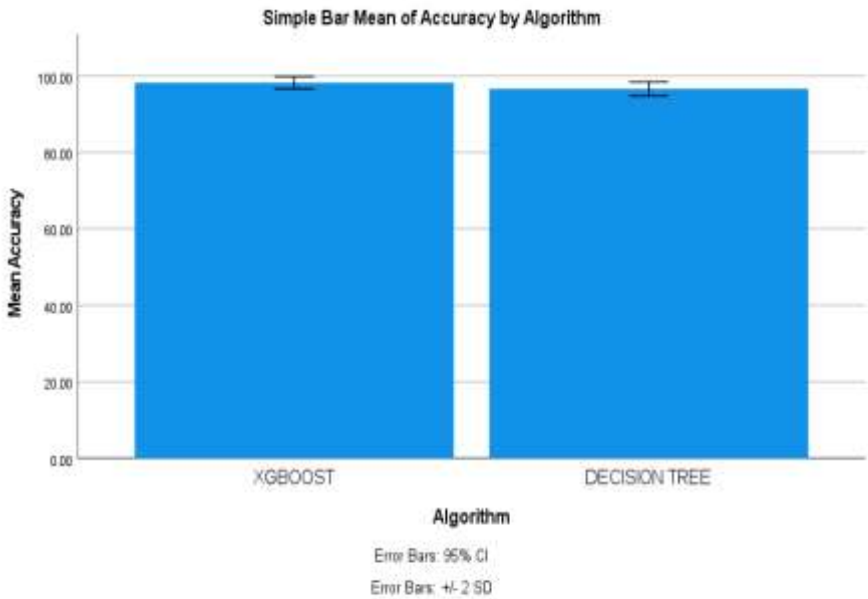

Fig.1. Comparison of mean accuracy between Novel XGBoost over Decision Tree where the former is better than the latter with a 2% increase. The X-axis gives the algorithms and the Y-axis Mean accuracy of prediction.