A MULTILAYER PERCEPTRON APPROACH FOR FAILURE PREDICTION IN THE AIR PRESSURE SYSTEM OF HEAVY TRUCKS

Felipe Cavalcanti de Barros Felipe, Isis Didier Lins, Marcio José das Chagas Moura

Centro de Estudos e Ensaios em Risco e Modelagem Ambiental (CEERMA) - UFPE

ABSTRACT

The maintenance costs of heavy vehicles have great relevance in the controllable costs since they are less vulnerable to external factors. Among the main components of total maintenance cost are the cost of inspections and the cost of unavailability of vehicles. Different maintenance policies might be implemented to reduce those costs. A possible maintenance policy is to carry out inspection and maintenance of vehicles based on failure prediction using data from sensors and predictive models. In this study, a failure prediction model using a multilayer perceptron and the combination of Borderline SMOTE and ADASYN was applied to the heavy truck sensor database made public by Scania in a Kaggle competition in 2016 to minimize maintenance cost. The model was able to reduce the total cost when compared to other models presented in the literature, obtaining a balanced accuracy of 95% and an area under the Receiver Operating Curve of 0.994.

1. INTRODUCTION

Maintenance costs for trucks represent a significant part of the expenses of logistics companies. However, compared to other main expenses, such as fuel and payroll, it becomes the main cost in terms of controllability since the others are subject to external factors (volatility in the price of oil, taxes, and others) [1] To slow down the depreciation of vehicles, guarantee greater reliability and availability of the fleet, it is ideal that carriers have a fleet maintenance plan. In some cases, the maintenance costs contribute to as much as 50% of the total transportation costs.

In general, maintenance costs and fleet unavailability are due to unnecessary inspections and failure to detect vehicle failures. Frequent inspections and component replacement might promote a high level of reliability. Still, this practice would increase the costs of inspection and reduce the overall availability of the fleet, given the frequent interruptions. Consequently, companies must detect failures without excessive maintenance and avoid unexpected failures and unnecessary inspection breaks [2].

Maintenance plans can be segmented into reactive (maintenance after failure), preventive (routine maintenance), or active (preventive and predictive or proactive repairs, which consist of maintenance plans where inspection and fault prediction activities are applied) [3]. A Prognostics and Health Management (PHM) program might be used to reduce total expenses by optimizing maintenance costs. PHM is a maintenance program that includes failure detection, diagnostic and prognostic of the status of equipment and systems via data (such as sensors measures) collection, preprocessing, and posterior prediction of status using a defined methodology [4]. When effectively used, the status prediction allows maintenance operators to intervene before failure and minimize interruption, corrective maintenance, and unavailability costs [5].

Problems of this nature might be modeled as classification problems (operational or failure). When sufficient data is available, Machine Learning models can be applied to predict the class of systems or equipment. However, as most systems are operational most of the time, the training dataset might be unbalanced in relation to the occurrence of each class, with the “failure” class being a minority. Therefore, the training database can induce the model during training to always foresee new equipment as operational since the accuracy would remain high, making the model accurate but useless. The generation of synthetic data from the minority class through oversampling algorithms (e.g., SMOTE [6], ADASYN [7]) may get around this problem.

In this context, the present study seeks to develop a new and more effective methodology for predicting the status of trucks using an open-source dataset comprising sensory data from Scania’s truck air pressure system, based on a multilayer perceptron and the combination of two different oversampling techniques at once. The proposed model also implements data preprocessing, and feature engineering algorithms. Later, the cost reduction based on the method will be evaluated and the results discussed and compared with the literature and with naive approaches to the problem.

2. DESCRIPTION

The dataset used comes from a proprietary database of Scania, one of the largest global manufacturers of heavy trucks. It was made available during a public Kaggle competition promoted by Scania in 2016, to predict whether the trucks are in operational or failure status based on 170 unknown and numerical explanatory variables. The dataset has 76,000 observations, segmented into 60,000 in the training dataset and 16,000 in the test dataset.

The main metric used to evaluate the performance of models is the total maintenance cost which is the sum of the cost of unidentified failures ($500 per unidentified failure) and the cost of unnecessary inspections due to false failure predictions ($10 per unnecessary inspection).

	Predicted label	
True label	**Negative**	**Positive**
Negative		False positive = $10
Positive	False negative = $500	

Table 1 - Cost per misprediction.

[8] evaluated the performance of Naive Bayes, Logistic Regression, Random Forest, Support-Vector Machine, and k-Nearest Neighbour methods after rebalancing the classes using SMOTE. The minimum total cost achieved was $15,940 when testing in the complete test dataset with 16,000 observations, using the logistic regression model.

[9] applied a Random Forest without any imbalanced learning method, but did not use the test dataset to evaluate its performance. Instead, the model was applied in the training dataset resulting in a total cost of around $ 36,000 or $ 0.6 per truck.

[10] compared four different classifiers: Support-Vector Machine, Cat-Boost, XG-Boost and Random Forest while also rebalancing the classes using SMOTE. However, only 10,000 observations from the test dataset were used in the testing step, instead of the total 16,000. Random Forest outperformed the other three classifiers with a total cost of $ 8,390, which averages at $ 0.839 per truck.

After training Logistic Regression, Support-Vector Machine, k-Nearest Neighbours, and Random Forest classifiers, [11] employed a ten-fold Cross-Validation step. The imbalance was dealt with the application of weighted metrics in the training steps, without any application of rebalancing methods. Random Forest classifier outperformed the other classifiers, achieving a total cost of $40,570, although the number of observations used in the testing step is not clear.

3. THEORETICAL FOUNDATIONS

3.1 Prognostics and Health Management (PHM)

One way to define a maintenance policy is through failure detection, diagnosis, prognosis, and monitoring of systems health, also known as Prognostics and Health Management. PHM starts with the analysis of data provided by sensors that monitor the physical characteristics of the asset (such as vibration, temperature, electric current, among others) and then preprocesses the database to separate signals from noise.

Then, based on the data collected, an information extraction model is used to explain the changes in the asset's condition to predict the asset's future state. The time from now until failure, or remaining useful life (RUL), is related to the current state of degradation, present environment, and other external and internal factors [4].

Once we estimate the current asset status and RUL, maintenance policies can be optimized. Thus, the overall cost of maintenance - mainly corrective - and the unavailability of the asset is reduced [5].

3.2 Multilayer perceptron (MLP)

Within the Machine Learning area, there is a group of models that seeks to learn a hierarchy of characteristics from the input data by building a deep architecture. We classify these as Artificial Neural Networks. These methods automatically learn characteristics at different levels, which increases the method's ability to extract complex information directly from the database, even without the help of humans.

The standard architecture of an Artificial Neural Network model is given by the occurrence of an input layer, several hidden layers and an output layer, simulating a Neural Network. Each layer may vary according to parameters such as the percentage of dropout, number of nodes, activation function, etc [12].

The learning of the model is driven by the oscillation of the weights associated with each node of each layer of the neural network. Using the stochastic gradient descent method, the weights of each node are updated at each iteration, to reduce minimize the loss function. The method seeks, in a stochastic way, to estimate the derivatives of the loss function in relation to the weights to know if the weights should increase or decrease in each iteration [13]. Multilayer Perceptrons are Artificial Neural Networks with an input layer, at least one hidden layer and an output layer where, in each node, an activation function (e.g., Sigmoid, Swish) is applied.

3.3 Imbalanced learning

For classification problems in reliability engineering datasets where we have previous data on assets classified by their operational status (operational or failure), it is common to observe an imbalance in the number of observations of "operational" and "failure" labels in the dataset. Therefore, traditional performance metrics, such as accuracy, end up losing practical sense, since a model that predicts that assets are always operational would achieve a high accuracy, although it is not of any use.

To avoid class imbalance, there are two main options: oversampling the minority class or undersampling the majority class [14]. However, when undersampling, information relevant to the model might be lost. Therefore, it those cases, an oversampling step may be the best option [15].

One of the traditional ways to carry out an oversampling is through bootstrapping: replicating the data of the existing minority class several times. However, smarter oversampling methods can be useful to generate observations that add information to the model.

3.4 Synthetic Minority Oversampling Technique (SMOTE)

The most traditional method of oversampling besides bootstrapping is SMOTE, or Synthetic Minority Oversampling Technique. The method consists of randomly choosing points from the minority class and, for each of these points, searching for its k-closest neighbors belonging to the minority class. Then, linear combinations of these points are generated [6]. Thus, the synthetic points generated are not exactly equal to any of the original points and can reduce the concentration of majority class points in a single region in the space of the attributes.

A concern when using SMOTE is that it might generate connections between outlier points that are closer to the majority class and the points of the minority class. Thus, improvements in this method could be made to better choose the points that will be used in the generation of synthetic observations.

3.5 Borderline SMOTE

One way to improve SMOTE is by using Borderline SMOTE. Borderline SMOTE applies a subclassification within the minority class: points that have their k-closest neighbors belonging to the majority class are considered noise and discarded, points that have k-closest neighbors belonging to both classes are considered border points and points that have only k-closest neighbors belonging to the minority class are purely minority [16].

Finally, Borderline SMOTE generates synthetic observations through the border points, since they are the ones that have the greatest potential to generate significant information for the model in distinguishing points belonging to the minority class and the majority class.

3.6 Adaptative Synthetic Sampling Method (ADASYN)

Another adaptation of SMOTE is the Adaptive Synthetic Sampling Method (ADASYN). Similarly, ADASYN prioritizes points of the minority class that have neighbors belonging to the majority class to generate synthetic data from them. Initially, ADASYN calculates the degree of impurity for each observation of the minority class as the percentage of k-nearest neighbors belonging to the majority class.

The method then normalizes the degree of impurity of all observations so that the sum of the normalized degrees of impurity equals one. Subsequently, the normalized degree of impurity of each point is multiplied by the number of points you want to generate synthetically to find how many points will be generated by this point. This way points closer to the majority class will generate more synthetic data than points closer to the minority class [7].

4. METHODOLOGY

The training dataset contains a significant amount of missing data. Some of the attributes contain up to 82% of missing values. Three steps were taken to reduce the impact of this problem in the preprocessing step. Firstly, all observations with more than 80% of the information missing were discarded. Secondly, all attributes with more than 50% of the information missing were discarded. Finally, to fill in the remaining blank information, the Multiple Imputation by Chained Equations (MICE) method was used with linear regressions. MICE imputes the mean at every missing value in the dataset; sets the mean imputations back to missing for one variable; regresses the missing values using the other independent variables; replaces the missing values with the regressed values and finally repeats the same steps for the other variables [17]. Subsequently, the now complete preprocessed trained dataset and the test dataset were standardized using MinMaxScaler. This way, the standardized datasets contain only values between 0 and 1 in the feature space.

As only 1.7% of the training dataset belonged to the minority class (failure), it became necessary to generate synthetic data from the minority class to prevent the model from predicting only the majority label during the training period, since this would result in a great accuracy but no practical usage. So, a combination of the oversampling methods ADASYN and Borderline SMOTE was used. Firstly, ADASYN was applied in the original training dataset to generate some of the synthetic observations required. Later, Borderline SMOTE was also applied in the original training dataset to generate more synthetic observations. Then, both new datasets were merged and the final training dataset was a combination of the original training dataset and the synthetic observations generated by ADASYN and Borderline SMOTE.

Lastly, aiming to reduce the total number of features to remove noise from the dataset, the last preprocessing step consisted in the application the Principal Component Analysis (PCA) algorithm. PCA starts by standardizing all of the variables; constructs a covariance matrix and computes the eigenvectors and eigenvalues of the covariance matrix to determine new variables that are constructed as linear combinations of the initial variables (which are called principal components) in a way that these new variables are uncorrelated and synthetize most of the information that was present in the initial variables [18]. Only the main features which explained most (in this case, 85%) of the variance of the data were kept.

Due to the large number of observations and features, the proposed methodology for the problem was creating a multilayer perceptron with three hidden layers for the classification of trucks. The first hidden layer has six nodes and a Rectified Linear Unit activation function. The second and third hidden layers have ten nodes and a Swish activation function. In all hidden layers, a dropout of 20% was used during the training stage. Later, because it is a classification problem, the output layer uses the sigmoid as an activation function. After training the model on the training dataset, the multilayer perceptron was tested on the test dataset.

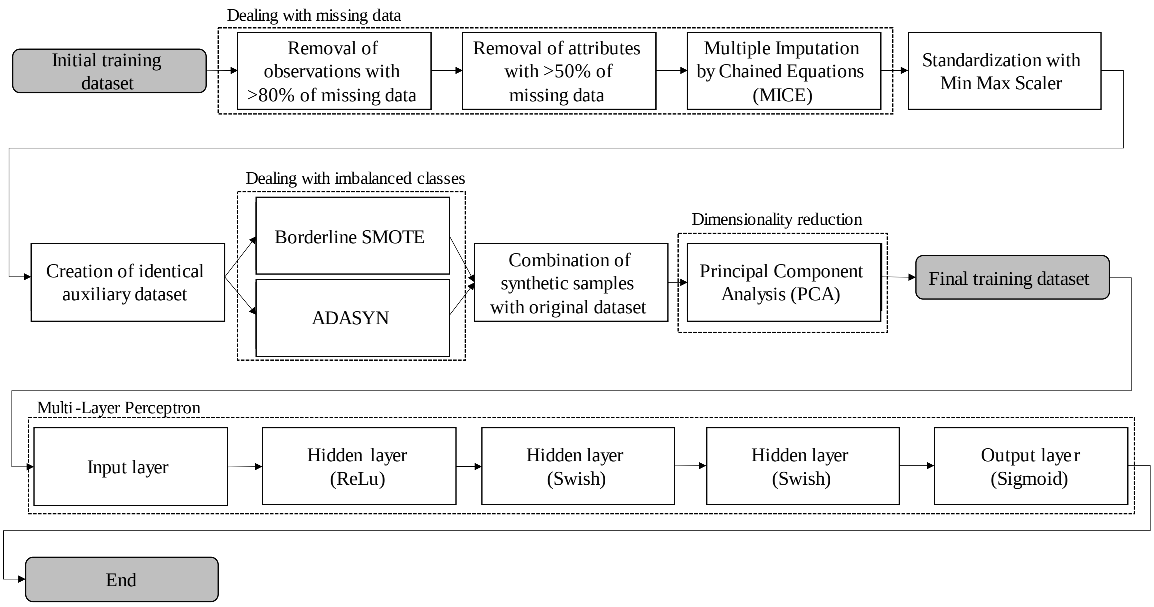

Figure 1 - Complete training process.

5. RESULTS

For this problem, we know that the cost of a false positive diagnosis is $10, while the cost of a false negative diagnosis is $500. The result found after implementing the Multilayer Perceptron model to the test dataset, which contains 16,000 observations, was a total cost of $14,110, with an accuracy of 95% and an area under the Receiver-Operating Curve (ROC) of 0.994. The confusion matrix and the ROC / AUC curve are shown below.

	Predicted label	
True label	**Negative**	**Positive**
Negative	True negatives = 14814	False positives = 811
Positive	False negatives = 12	True positives = 363

Table 2 - Confusion matrix.

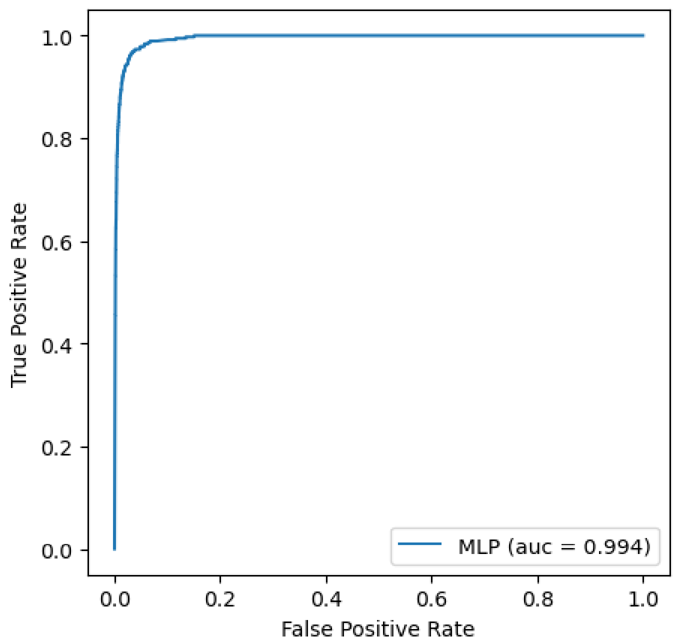

Figure 2 - Receiver-Operating Curve (ROC) and area under the ROC.

6. CONCLUSION

A naive maintenance policy that did not inspect any heavy truck would generate a total cost of $ 187,500, given by the cost of not inspecting the 375 faulty trucks ($500 per faulty truck not inspected). Another maintenance approach, which consists of inspecting all vehicles, would generate a total cost of $156,250 given by the cost of inspecting the 15,625 trucks unnecessarily in operational condition ($10 per operational truck unnecessarily inspected).

It is noted, then, that the proposed methodology manages to reduce the total cost by approximately 91% against a maintenance program that inspects all vehicles and by approximately 92.5% against a maintenance program that does not inspect any vehicle at all. This methodology also achieves a lower total cost compared to those proposed in the works presented in the literature review, which follow an identical testing process.

We can conclude that the proposed methodology based on PHM using a multilayer perceptron and the combination of Borderline SMOTE and ADASYN is much more efficient than a corrective maintenance policy or an inspection policy for all trucks in the fleet and allows a lower total maintenance cost than the methods available in the literature which follow a similar testing process. None of the methods analyzed applied Borderline SMOTE, ADASYN or multilayer perceptrons, although some use traditional SMOTE to rebalance the training dataset.

7. REFERENCES

[1] VICENTE MACIÁN, BERNARDO TORMOS, JESUS HERRERO, Maintenance management balanced scorecard approach for urban transport fleets, Eksploatacja I Niezawodnosc - Maintenance and Reliability, (2019).

[2] SERGII VORONOV, ERIK FRISK, MATTIAS KRYSANDER, Data-Driven Battery Lifetime Prediction and Confidence Estimation for Heavy-Duty Trucks, IEEE Transactions on Reliability, (2018).

[3] STANISLAW LEGUTKO, Development trends in machines operation maintenance, Polish Maintenance Society, (2009).

[4] MICHAEL PECHT, Prognostics and health management of electronics, Encycl. Struct. Heal. Monit., John Wiley & Sons, (2009).

[5] CAIO BEZERRA SOUTO MAIOR, MARCIO DAS CHAGAS MOURA, ISIS DIDIER LINS, ENRIQUE LOPEZ DROGUETT, HELDER HENRIQUE LIMA DINIZ, Remaining Useful Life Estimation by Empirical Mode Decomposition and Support Vector Machine, IEE Latin America Transactions, (2016).

[6] NITESH CHAWLA, KEVIN BOWYER, LAWRENCE HALL, PHILIP KEGELMEYER, SMOTE: Synthetic Minority Over-sampling Technique, Journal of Artificial inteligence, (2002).

[7] HAIBO HE, YANG BAI, EDUARDO GARCIA, SHUTAO LI, ADASYN: Adaptive synthetic sampling approach for imbalanced learning, IEEE International Joint Conference on Neural Networks, (2008).

[8] SUSHIL RAWAT, Predict component failure related with Air Pressure System at Scania Trucks using various machine learning methods, (2020).

[9] CHRISTOPHER GONDEK, DANIEL HAFNER, OLIVER SAMPSON, Prediction of Failures in the Air Pressure System of Scania Trucks Using a Random Forest and Feature Engineering, (2016).

[10] YANGALASETTY LAKESH, KOMATIMENI SRI SAI NIKHIL, ERLA VINAY KUMAR, B GOPESH KRISHNA MONAN, Truck APS Failure Detection using Machine Learning, Proceedings of the International Conference on Intelligent Computing and Control Systems, (2020).

[11] CAMILA FERREIRA COSTA, MARIO NASCIMENTO, IDA 2016 Industrial Challenge: Using Machine Learning for Predicting Failures, Department of Computing Science, University of Alberta, (2016).

[12] YOSHINOBU TAMURA, SHIGERU YAMADA, Reliability Analysis Based on Deep Learning for Fault Big Data on Bug Tracking System, 5th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), (2016).

[13] MATTHEW GARDNER, STEPHEN DORLING, Artificial Neural Networks (the Multilayer Perceptron), A review of applications in the atmosferic sciences, (1998).

[14] HAIBO HE, YING MA, Imbalanced Learning: Foundations, Algorithms and Applications. The Institute of Electrical and Electronics Engineers, (2013).

[15] BEE WAH YAP, KHATIJAHHUSNA ABD RANI, HEZLIN ARYANI ABD RAHMAN, SIMON FONG, ZURAIDA KHAIRUDIN, NIK NIK ABDULLAH, An Application of Oversampling, Undersampling, Bagging and Boosting in Handling Imbalanced Datasets, Proceedings of the First International Conference on Advanced Data and Information Engineering (2013).

[16] HUI HAN, WEN-YUAN WANG, BING-HUAN MAO, Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning, International Conference on Intelligent Computing 2005: Advances in Intelligent Computing, (2005).

[17] MELISSA AZUR, ELIZABETH STUART, CONSTANTINE FRANGAKIS, PHILIP LEAF, Multiple imputation by chained equations: what is it and how does it work?, International Journal of Methods in Psychiatric Research, (2011).

[18] SVANTE WOLD, KIM ESBENSEN, PAUL GELADI, Principal component analysis, Chemometrics and Intelligent Laboratory Systems, (1987).