Predictive Analytics in Utility Vehicle Maintenance

Jürgen Prinzbach, Stephan Trahasch
Electrical Engineering and Information Technology
Offenburg University of Applied Sciences
Offenburg, Germany
email: {juergen.prinzbach, stephan.trahasch}@hs-offenburg.de

***Abstract*—In public transportation, the motor pool often consists of various different vehicles bought over a duration of many years. Sometimes, they even differ within one batch bought at the same time. This poses a considerable challenge in the storage and allocation of spare parts, especially in the event of damage to a vehicle. Correctly assigning these parts before the vehicle reaches the workshop could significantly reduce both the downtime and, therefore, the actual costs for companies. In order to achieve this, the current software uses a simple probability calculation. To improve the performance, the data of specific companies was analysed, preprocessed and used with several modelling techniques to classify and, therefore, predict the spare parts to be used in the event of a faulty vehicle. We summarize our experience running through the steps of the Cross Industry Standard Process for Data Mining and compare the performance to the previously used probability. Gradient Boosting Trees turned out to be the best modeling technique for this special case.**

Keywords-maintenance; utility vehicle; spare parts; data analysis; predictive analytics.

I. INTRODUCTION

For service providers in the field of public transportation or waste disposal in particular, it is important to be able to optimally manage their own vehicle fleet in the area of maintenance and to minimize downtimes as much as possible. Unfortunately, the vehicles purchased over the years sometimes differ so much even within a single batch that many different spare parts have to be kept in stock. In addition to the enormous storage costs, this makes the assignment of a new part to a faulty vehicle more difficult than one would hope for [1]. Therefore, software is used in the areas of fleet management, workshop and logistics, to help traffic companies meet these challenges. However, cases still exist when a defect reported by a driver is checked upon arrival at the depot of the company by a shop assistant. In the worst case, when creating a repair order, the mechanic realizes that not all spare parts needed are in stock, which means that the downtime of the vehicle will be extended by the respective delivery time taking at least six to eight hours, even with special express delivery, depending on the industry and supplier. If the workshop manager were to receive a well-founded proposal on the material to be installed ahead of time, it could be ready at the point of entry to the workshop and both the time and cost could be reduced enormously.

For that reason, this work focuses on analysing and processing the data of several traffic companies ranging from the vehicle data to the respective repair processes. In particular, data quality should be taken into account, as the data basis of the software could be used by the respective company in the most diverse ways. The knowledge gained from this should make it possible, based on the Cross Industry Standard Process for Data Mining [2], to improve an already implemented software by creating and evaluating different models for the prediction of the corresponding spare parts.

This paper is organized as follows. Section 2 puts this paper in the context of related works, whereas Section 3 explains the current implementation of the mentioned probability in the application and the data this is based upon. Data analysis and preprocessing are explained in Section 4, whereas Section 5 describes the actual modelling. Section 6 summarizes the evaluation criteria and results, and the final section concludes this paper.

II. RELATED WORK

In the field of maintaining machines or plants, predictive maintenance is often used when talking about data analytics. This usually means the prediction of faults or failures of said machines in order to avoid larger failures through planned repairs or servicing. As described in [3], this is about the observation of the current state of the machine in the execution of its tasks. The bottom line is therefore the evaluation of log-based sensor data and the possible prediction of failures. Another elaboration [4] also attempts to improve their maintenance planning by detecting error signatures in environment variables in significant data sets containing machine records. Even though this paper deals with the avoidance of vehicle failures, such a preventive approach is currently not possible, which is partly due to the fact that the vehicle manufacturers do not make the data available during operation.

Rather, one could think about it as using predictive analytics techniques as a kind of "management tool" to reduce the planned and unplanned downtime of the respective machine [5] – in this case the vehicles. Detecting the correct and needed spare parts before the vehicle arrives in the depot could at least partially eliminate unnecessary activities, such as inspecting the vehicle or adjusting incorrect parts, thereby dramatically reducing the overall cost of the vehicle. In the optimal case, for example, the repair

 ISBN: 978-1-61208-681-1

could be planned so that it lies between two uses of the vehicle. This would make the vehicle practically not fail.

Breaking down the required task down to its core one realizes that it is ultimately about the classification of the respective spare part based on relevant attributes of the existing data sets. Which attributes in addition to the error message of the driver or vehicle are relevant or which algorithms are suitable in this case for determining the parts is therefore part of this work. In addition to Support Vector Machines (SVM) or simple decision trees, Gradient Boosted Trees could also be an option. A future relevant approach could also be "Gradient Boosted Decision Tables" using a novel method of storing the decision tables and a more efficient prediction algorithm [6].

III. CURRENT IMPLEMENTATION

In the area of public transportation, the software used here offers extensive functions for the administration and support of buses and their maintenance. It has a modular structure and supports a large number of vehicle types and their technical infrastructure. Among other things, vehicles can be planned, timetables managed, defects recorded and spare parts ordered. The last two points belong to the process of maintenance, which is triggered in the event of a fault on the vehicle. As soon as a defect is created in the system, possible spare parts are displayed with the respective usage probabilities. However, this is only possible with correspondingly good data and with reference to the vehicle and the work to be performed.

Due to the individual adaptability of the software, the various supported business areas as well as the high degrees of freedom in the administration of the data by the users, it may be difficult to obtain sufficient data. Furthermore, the number of processes, after which meaningful suggestions for spare parts can be generated, increases due to the variety of different vehicles of each company. However, if all prerequisites are met, it is possible to confirm everyday knowledge and gain new insights with the calculation of the probability of using specific parts. This already implemented probability is calculated from the ratio of the number of processes executed using a particular spare part to the total number of executions of that process. In this way, one obtains a simpler way of calculating the conditional probability of using a material, assuming that a particular process is applied to a defect. However, the probability also always depends on the particular vehicle, which – in simple terms – is defined by its brand and model. Thus, formula (1) can be used to calculate the probability of using a replacement part, where the individual components can be formalized as such that I_v represents the parts used and O_v the individual processes:

$$P(I_v \mid O_v) = P(I_v \cap O_v) / P(O_v) \quad (1)$$

This could lead to a result like the one shown in Table I. So, because of the probability in this particular case one would probably order item 1536 for the corresponding process and vehicle.

TABLE I. CALCULATION OF THE PROBABILITY OF THE USAGE OF ONE PARTICULAR ITEM FOR EACH PROCESS AND VEHICLE

Item	Process	Vehicle	P [%]
1536	82-1203	EVO-O530-BJ08	47.15
1531	82-1203	EVO-O530-BJ08	29.27
1539	82-1203	EVO-O530-BJ08	13.01
1537	82-1203	EVO-O530-BJ08	2.85
1529	82-1203	EVO-O530-BJ08	1.22

IV. DATA FOUNDATION

The required data is stored in a relational database management system. On this basis, the attributes needed to calculate the explained probability are simply merged via joins in a view. However, there is the question of how much the results can be trusted and business decisions to be made on that basis. On the one hand, some users may sometimes make very far-reaching changes to the data; on the other hand, they must also be appropriately maintained, and the processes carefully recorded. Here, one can probably assume that given freedoms are often exploited, which may corrupt the data quality and thus the results. In addition, it turns out that the recalculation and update of the probability is not always enabled for all processes. It should also be noted that the probability of use is based on purely historical observations and that no model for future events is included or can be derived.

Moreover, direct feedback on errors is just as impossible as basic testing of the quality of the process in the event of emerging defects. For the practical application of the method, with a few exceptions, it is still necessary to have a person with relevant specialist knowledge. So important decisions should not depend on this calculation – but it can help in assessing the situation at hand. To improve this situation, predictive analytics methods have been tested and their results analysed in further sections.

A. Data Understanding

In order to better understand the vehicle and deficiency data needed to predict spare parts and thus create different models, it is first necessary to understand the context of the data by generating it. Furthermore, the quality of the preliminary data has to be considered more closely so that the attributes used in the modelling can be selected. After that the data may be preprocessed for further usage. In this work, R [7] and R Studio [8] have been used with various packages, such as "caret" [9], "ROSE" [10] or "doParallel" [11] for all analysis and modelling work.

B. Data Analysis

For further analysis of the data, database backups are used of two companies who use the same software in different ways and to varying degrees. On closer inspection, the big difference between the existing data records has become clear. The first database (DB1) has more than 25 times as many lines with 862,350 defect entries as the second database (DB2), which is also reflected in the number of different attributes.

 ISBN: 978-1-61208-681-1

TABLE II. DISTRIBUTIONS AND CHARACTERISTICS OF SELECTED ATTRIBUTES

Attribute	Range	Mean	Median	Skew	Deviation
ManufacturerTypeKey	325	100.07	59	1.01	76.06
Manufacturer	44	29.33	34	-0.35	6.55
Model	103	60.54	73	-0.88	18.98
Process	1317	510.25	485	0.27	403.10
Fault	368	161.83	178	0	107.75
Material	1109	499.90	410	0.39	311.03

Thus, the first company with 57 vehicle manufacturers and more than 140 models has almost 30 times as many different vehicles in use as the other one. However, this stark difference or this high number of different manufacturers seems to be exceedingly unrealistic, since there are not so many brands in the area of buses in the local market. This could either be a mixture of different categories, such as passenger cars or some data may not have been recorded correctly. It is also noticeable that the granularity of work processes and defects differ greatly. For example, with 584 to 369, DB1 has more than 1.5 times more defects and 2,567 to 1,234 more than twice as many processes than DB2. These observations can also be demonstrated in the usable spare parts. While a larger number of different vehicles can be expected to have an increasing number of different replacement parts, the differences in granularity present show how fundamentally different the two companies deal with defects and workflows. Fig. 1 illustrates these observations by the different occurrences of the key attributes "ManufacturerTypeKey", an artificial primary key, which is composed among other things of the two attributes "Manufacturer" and "Model" which are also shown. Furthermore, the attributes "Process", "Fault" and "Material", which corresponds to the spare parts, are displayed. Note that the illustration assumes a minimum occurrence of defects and attributes of 50 each. Even though the two most common deficiencies in DB1 have been removed, as explained in the preprocessing section, it promises significantly better results.

Therefore, further investigations are being concentrated on this database. For example, the attributes "performance" and "weight" have a proportion of missing values (NAs)

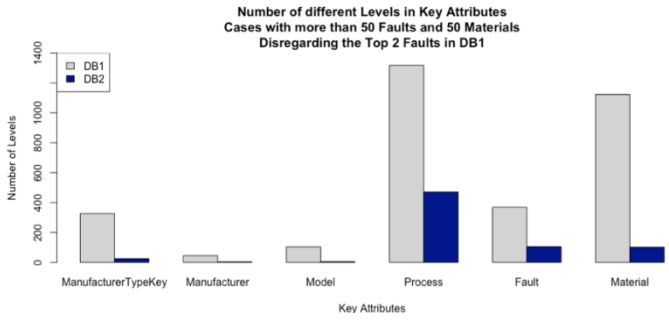

Figure 1. Number of different characteristics in key attributes with a minimum occurrence of the deficiencies and attributes of 50 each and the exclusion of the two most prevalent deficiencies in DB1

between 70% and 80%, which makes them completely useless due to the lack of possibilities for recalculation. For the other attributes, on the other hand, the number of NAs is so small that the respective rows could simply be removed. Thus, after the clean-up of the missing values, a number of 861,026 supposedly usable rows are obtained. However, when looking at the three most common deficiencies, it turns out that this unfortunately is not the case. For example, one can see from the descriptions "additional work and maintenance" and "lack, please more specific" of the fault types "Z1111" and "Z9999" that in the first case simple maintenance work has been carried out. Thus, there was no defect of the vehicle. In the second case, the clerk simply did not know what was really broken. So, both manifestations should not be used in the given context, as this could distort the result in case of doubt. This means that with 485,489 lines that make up these two most common deficiencies, nearly 60% of the total database is unusable for the process of learning which spare parts to use in which situation.

For further analysis, the distributions and characteristics of individual attributes of the resulting data set can now be considered. This information is presented in Table II, noting that only those datasets have been used in which both the feature of the spare part and the defect have occurred at least 50 times. Furthermore, the attributes have been numbered prior to the calculations. It can be seen that, for example, the months in which a deficiency occurred are distributed fairly evenly, whereas the affected vehicle models appear to be affected very differently. Therefore, if necessary, the data should be normalized in preprocessing.

Calculating a correlation matrix and looking at it by using a heat map, the correlation of 0.27 shows that the manufacturer-type key seems to have some connection with the target class of parts, while the day or month when the deficiency was reported appears to be completely insignificant (0.01 to 0.02). Whether this is actually the case will be demonstrated by the various experiments in creating the models. What seems logical, however, are the obvious links between the manufacturer and the model (0.71) or the age and miles driven (0.28).

C. Data Preprocessing

In the following, the activities performed in the field of data preparation are explained. Here, not only separately performed steps are mentioned, but also those which have been run through during the model creation with the help of the respective packages. First, approximately 15 attributes like "weight" and "performance" that have been found to be unhelpful during each experiment have been removed

 ISBN: 978-1-61208-681-1

because, for example, they contain too many missing values [12]. Following this, the errors for non-specific problems "Z1111" and "Z9999" were removed due to their low information content.

After this clean-up, some transformations of the data and generation of new attributes from existing columns were performed. This includes, for example, the generation of the age in years, which is derived from the first registration and the current date. The age of the vehicles can therefore be very important, as some parts become more susceptible to defects over time. This also means that certain repairs and thus also certain spare parts, for example, are needed after 5 years, rather than after just one year.

Furthermore, from the date of the defect notification, the corresponding month was extracted to obtain a seasonal component. From this, temporal correlations of potential failures of the heating can be obtained, which are more likely to occur in winter than in summer. In addition to this characteristic, the mileage since the last inspection has been roughly obtained from the kilometers travelled so far. This was only possible because the vehicles in the public transport industry are always maintained every 30,000 km. After calculating this information, the kilometer-based attributes are categorized according to the experiment. For example, the total mileage of the vehicles could be divided into 5,000 km classes.

In the next step, the attributes were converted to numerical factors and those columns were removed that had become unnecessary by generating additional information. Theoretically, a scaling or transformation at this point would be useful. However, a number of experiments have shown that performing these activities manually produces worse results than running them by the respective package just prior to the modelling. Now the data was prepared in such a way that further experiments could be carried out on the basis of it. Other possible steps at this point included both splitting the data into training and test data by a fixed percentage or performing a Principal Component Analysis (PCA). For example, the former would have specified that 70% of the data would be used to learn a model, while 30% would be used for later evaluation [12]. The PCA tries to further reduce the number of currently 11 attributes at this time by calculating artificial properties to reduce complexity while maintaining the same quality [12]. Furthermore, PCA offers other benefits, such as decorrelating the attributes. Ultimately, however, all attempts at optimization were doomed to failure, as the various deficiencies and materials occur in very different frequencies, which can be seen, for example, in the respective skewness in Table 2. Thus, there was a bias in the direction of the most prevalent manifestations, which will be shown by the experiments presented in the next section.

V. MODELLING

At the beginning, we performed experiments with various algorithms and various combinations of attributes and split ratios of the training and test data. For the latter, 70:30 and 80:20 were first investigated, while Naive Bayes [12] and Support Vector Machines (linear, radial, and polynomial) were mainly used with their default settings. It quickly became clear that a holistic prediction of the many different, very unevenly distributed target classes of the attribute "material" is not possible. For this reason, according to the number of different spare parts, we have to generate databases with all data records but binary target classes. Each database therefore stands for a single spare part and its use, which is why the target attribute "material" only indicates whether or not it is used – in other words, a "yes" or a "no". This created significantly better results. However, in some cases a few positive cases were faced with some 10,000 negative cases, which meant that some materials could be predicted extremely well and others extremely badly. Therefore, we tried to approximate the uneven classes with the help of packages like ROSE and thus to improve the results, which finally succeeded. We were also able to largely confirm the results of the correlation matrix for the individual attributes with some experiments. However, there was also one or the other surprise. While the matrix did not see any correlation with the day the defect was reported, this property proved helpful in determining the required spare part. In order to give a small but concrete overview of the modellings carried out, three of them are described below. First, however, we explain how the individual parameters of the respective packages were determined. For the modelling itself, mainly the Caret package [8] with different algorithms was used.

A. Parameter Settings

To determine the best possible parameters, models were created for 10 to 20 previously randomly selected spare parts and the respective results compared. Through this reduction, the calculation time could be minimized. However, with a more powerful production system, integration into the actual modelling process would be desirable. Finally, the following steps for parameter determination were carried out – here exemplified at the *k*-fold cross validation:

1. Definition of the possible values for the tuning parameters.
2. Execution of the modelling process including resampling of the data and prediction of the respective spare parts using the test data.
3. Creation of an evaluation matrix for all results meaning that the results of the respective predictions have been collected in a confusion matrix and the sensitivities have been read out.
4. Determination of the final tuning parameters by ordering the sensitivities in descending order of magnitude and frequency.

B. Naive Bayes vs. GBM and C5.0

After the initial experiments, it turned out that the generated data sets with binary target classes using Naive Bayes provided the best results so far. In this series of experiments, tree-based models, such as Gradient Boosting Trees (GBM) [12] or C5.0 Trees were tested. GBM should hereby maximize the Receiver Operating Characteristic (ROC), while C5.0 used a cost function to try and improve the results by increasing the cost of incorrect predictions.

 ISBN: 978-1-61208-681-1

Furthermore, it should be noted that only parts that were used more than 1000 times were evaluated, resulting in 84 models. Added to this is the restriction to defects that occurred more than 50 times. Finally, after some experiments a split ratio of 75:25 was calculated as the mean of the previous experiments.

C. Gradient Boosting Trees

With the aforementioned experiment, we found that Gradient Boosting Trees in our context enable the better models, which is why they were used as a priority thereafter. Added to this was the described determination of optimal parameters, which should further improve the results. Here are some of the parameters and options used in the Caret Package:

- Scaling and centering of the data
- Repeated *k*-fold cross validation with 5 folds and 2 repeats
- Between 400 and 500 trees at a depth of 7

However, at this time it was first noticed that the most prevalent shortage, which accounts for more than one-third of the data, is for maintenance and remanufacturing only and, therefore, does not represent a defect. For this reason, these samples did not contribute to the determination of the target class and were therefore not considered. The remaining shortcomings and materials have now been assumed to occur at least 100 times, ultimately using just over 200,000 samples and creating 788 models.

D. Gradient Boosting Trees without the two most common defects

In this experiment, only models were created using Gradient Boosting Trees. However, only those records were used that do not represent the two most common shortcomings "Z1111" and "Z9999". In addition, both the respective defects and the target class should total at least 50 times in the data, leaving 231,363 lines remaining. Following this, binary training and test data with a ratio of 75:25 were generated for each of the 1110 spare parts. In the modelling itself, the following parameters were used:

- Preprocessing:
  - Center and scale
  - Principal Component Analysis
- Train Control:
  - Repeated cross validation with 6 folds and no repetition
- Grid Settings:
  - 700 trees with a depth of 13
  - Shrinkage of 0.1

It should be noted at this point that this experiment was performed once with and once without the information of the work process. This is because the usage probability used so far includes this, while in the future it will work without this information. The tests carried out thus permit estimates of the quality of the individual models in both cases.

VI. Evaluation

After the experiments presented in the previous section and the training of different models for the classification of spare parts, the criteria for determining the model quality and the performance are explained below. Afterwards, the results are presented and conclusions drawn for future applications.

A. Underlying Criteria

In principle, the probability of using spare parts already implemented sets the standard for all new processes. Furthermore, it is especially important for companies to recognize the cases in which a spare part is really needed. This means that it is far less dramatic to get a material out of the warehouse for repair or to order and then not need it, as if the vehicle is already in the workshop and it is found that parts are missing. On the one hand, one can conclude that some of the known quality measures should be weighted more heavily than others, on the other hand, the relevant measures for the used probabilities must be calculated. The latter is relatively easy since it already covers or predicts the positive cases. This also coincides with the requirement to determine really needed parts.

Thus, the evaluation strategy is quite simple: For the predictive algorithms, the known criteria listed below are used, with the ultimate focus being on sensitivity and the possible comparison with the probabilities. For these measurements, a 2x2 confusion matrix is first calculated, which makes it possible to compare the actual classes to test data predicted with the respective model. This simple matrix is usable because the data has been converted into binary sets as described. From this, mainly criteria like sensitivity, accuracy and others were calculated [13].

Although other measures such as False Positive Rate or Positive Predictive Value have been calculated, they will not be listed here due to the lack of relevance to the results.

B. Results

Despite the poor accuracy of 20 to 30% achieved in first experiments with Naive Bayes, this algorithm is used again and again as a comparison. Looking at the average values for sensitivity, specificity and accuracy (see Fig. 2), the three algorithms compared here seem to work similarly well. Furthermore, it can be seen that the optimization towards the spare parts actually needed has an effect and, therefore, the

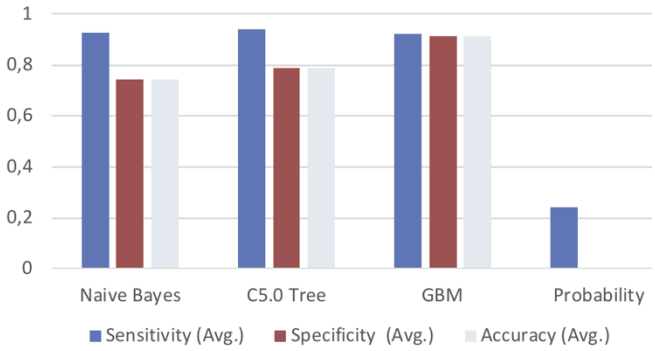

Figure 2. Average modelling results by criteria

 ISBN: 978-1-61208-681-1

true positive rate (TPR) is higher than the other values. The Gradient Boosting Trees, with the two most common defects removed, are an exception. With more than 90% on all criteria they provide very good results. Comparing these with the average sensitivity of the currently implemented probability reveals its seemingly blatant weaknesses. This also confirms the assumption that the modeling techniques of predictive analytics should provide better results. However, this diagram does not disclose some important information. First and foremost, only the average of the respective quality measure is indicated across all models and thus across all spare parts. This means that outliers, i.e., binary models for the respective spare parts that deliver bad classifications, are not recognizable in Section V (D). Another important point is that the probability is not calculated in all cases, which makes a scientifically sound comparison almost impossible. This is because the function may have been disabled due to other deficiency evaluations or set calculation limits. Nonetheless, these benchmarks can be seen as indicative, suggesting that better ways could be found to provide automatic suggestions for replacement parts to be installed in the event of vehicle defects.

VII. CONCLUSIONS

First, it must be noted that despite the problems during the experiments, it is in principle possible to predict the required spare part with predictive analytics in case of a defect in a vehicle. Based on the underlying criteria, this also worked better than the currently implemented probability.

However, in order to make an actual recommendation and to be able to compensate for variations in the quality of the forecast, a few points should be noted. First, care should be taken to improve the quality of the data. For this purpose, it would be useful to standardize the basic vehicle data across all companies using the software and at least to explain the information of the vehicle registration certificate to mandatory information. Furthermore, a uniform catalog of shortcomings should be drawn up in cooperation with the customer in order to avoid, for example, different granularities in case of defects. This would allow more attributes or even more databases from multiple customers to be used to create the models, which should allow them to be more accurate and less subject to fluctuations. Whether Gradient Boosting Trees still deliver the best results after that will have to be reevaluated. However, it may also be beneficial to use the probability calculation to validate the results of the models if the results are not too bad, which would be the case for accuracies of less than 60%.

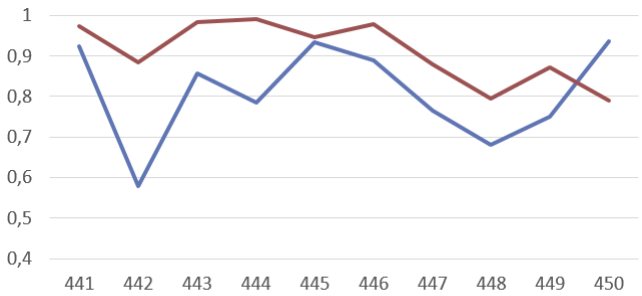

Figure 3. Results from ten randomly selected models created with GBM

These improvements will be addressed in future activities in this area and an integrated service in the cloud for companies in public transportation will be created, which then stores information about the work processes in case of emerging defects and can create models on the common data. Then, it should also be able to answer inquiries about new processes and make suggestions or make predictions about spare parts. This work has thus paved the way for far-reaching improvements to the repair of utility vehicles.

ACKNOWLEDGMENT

This work was supported by COS GmbH in Oberkirch (Germany) and the Federal Ministry of Education and Research.

REFERENCES

[1] J. Prinzbach, Predictive Analytics in der Instandhaltung, Master Thesis, Offenburg University of Applied Sciences, 2017.

[2] P. Chapman et al., CRISP-DM 1.0: Step-by-step data mining guide, 2000.

[3] L. Spendla, M. Kebisek, P. Tanuska, and L. Hrcka, “Concept of Predictive Maintenance of Production Systems in Accordance with Industry 4.0,” IEEE 15th International Symposium on Applied Machine Intelligence and Informatics (SAMI), 2017 2017 Jan 26 IEEE Press, Jan. 2017.

[4] B. Cline, R. S. Niculescu, D. Huffman, and B. Deckel, Predictive Maintenance Applications for Machine Learning, 2017 Annual Reliability and Maintainability Symposium (RAMS), IEEE Press, Jan. 2017.

[5] R. K. Mobley, *An introduction to predictive maintenance*, Butterworth-Heinemann, Amsterdam, New York, 2002.

[6] Y. Lou and M. Obukhov, “BDT: Gradient Boosted Decision Tables for High Accuracy and Scoring Efficiency,” Proc. 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '17, ACM Press, Aug. 2017.

[7] The R Foundation, "The R Project", [Online] Available from https://www.r-project.org [retrieved: Oct. 2018]

[8] RStudio, "RStudio" [Online] Available from https://www.rstudio.com [retrieved: Oct. 2018]

[9] The R Foundation, "caret", [Online] Available from https://cran.r-project.org/package=caret [retrieved: Oct. 2018]

[10] The R Foundation, "ROSE", [Online] Available from https://cran.r-project.org/package=ROSE [retrieved: Oct. 2018]

[11] The R Foundation, "doParallel", [Online] Available from https://cran.r-project.org/package=doParallel [retrieved: Oct. 2018]

[12] M. Kuhn and K. Johnson, Applied predictive modeling, New York: Springer, 2016.

[13] J. Han, M. Kamber, and J. Pei, Data mining: Concepts and techniques, Amsterdam: Elsevier/Morgan Kaufmann, 2012.

 ISBN: 978-1-61208-681-1