

Feature Analysis of Predictive Maintenance Models

Zhaoan Wang

Abstract—Research in predictive maintenance modeling has improved in the recent years to predict failures and needed maintenance with high accuracy, saving cost and improving manufacturing efficiency. However, classic prediction models provide little valuable insight towards the most important features contributing to the failure. By analyzing and quantifying feature importance in predictive maintenance models, cost saving can be optimized based on business goals. First, multiple classifiers are evaluated with cross-validation to predict the multi-class of failures. Second, predictive performance with features provided by different feature selection algorithms are further analyzed. Third, features selected by different algorithms are ranked and combined based on their predictive power. Finally, linear explainer SHAP (SHapley Additive exPlanations) is applied to interpret classifier behavior and provide further insight towards the specific roles of features in both local predictions and global model behavior. The results of the experiments suggest that certain features play dominant roles in predictive models while others have significantly less impact on the overall performance. Moreover, for multi-class prediction of machine failures, the most important features vary with type of machine failures. The results may lead to improved productivity and cost saving by prioritizing sensor deployment, data collection, and data processing of more important features over less importance features.

Keywords—Automated supply chain, intelligent manufacturing, predictive maintenance machine learning, feature engineering, model interpretation.

I. INTRODUCTION

ORGANIZATIONS have made major investments in recent years to modernize supply chains with forecasting models for customer demands, predictive maintenance for manufacturing, and inventory management. The COVID-19 pandemic caused unprecedented disruption of the supply chain. Automation of the supply chains and especially automated manufacturing has played an increasing role in the post-pandemic age.

Robotics incorporated with remote operations and sensor technology has an increasingly large role to play in intelligent manufacturing. In addition to improving productivity and product quality, manufacturing automation plays a crucial role in meeting global demands for essential products during challenging times such as COVID-19.

Predictive maintenance is performed based on an estimate of the health status of the manufacturing equipment. It allows for advance detection of pending failures and enables timely intervention before the occurrence of failures. Ran et al. [1] provide a comprehensive literature review on predictive maintenance with emphasis on system architectures, purposes

and approaches. Susto et al. [2] proposed train multiple classification modules to provide different performance tradeoffs in terms of frequency of unexpected breaks for optimal maintenance decision.

Most machine learning research in predictive maintenance focuses on experimenting with various machine learning models and evaluating their performance [2]. However, feature selection and analysis of feature importance are not well researched in the state-of-art literature. More importantly, the interpretation of predictive models and the analysis of features' influence on the prediction results deserve more attention: the former answers the question about why the prediction was made and the latter addresses the question about which features have the most influence on such decision. This study attempts to shed some light on these two questions.

The remainder of this paper consists of five sections. The second section describes the dataset used in the machine learning experiments. In the third section, four feature selection algorithms from different categories are applied to select top features that contribute most to the prediction. In the fourth section, the features selected from the four feature selection algorithms are aggregated to obtain the most important features across the four feature selection algorithms. In the fifth section, the most important features are analyzed and validated using SHAP [10]. Finally, the conclusions and future work are discussed.

II. DATASET

The original simulated data come from the three sources: timed telemetry data about machine conditions, machine technical parameters, and machine maintenance and failure records [3]. It has 9 features: time, machine id, machine age, error id, maintenance id, and machine parameters including voltage, rotation, pressure, and vibration. The data were first processed by calculating the median value and standard deviation values of machine parameters every three hours and every 24 hours. The time of the last maintenance of a machine was then obtained through the maintenance records consisting of the times and the types of maintenance. At the end, the data samples of every three hours and the total number of various errors in each time period are merged to form the dataset used in the machine learning experiments described in this paper. The details of the dataset creation process are described in [4] and [5]. The preprocessed dataset consists of 29 features and 291,668 samples with 4 types of failures. The preprocessed dataset is available as described in [6].

The class labels of the dataset are highly imbalanced, with all failure samples composing only approximately 2.28% of the entire dataset. Moreover, some technical information such

Zhaoan Wang is with the Engineering Technology Division, College of Engineering Wayne State University, Detroit, MI, USA (e-mail: gz8644@wayne.edu).

as the types of the machine and types of machine failures is anonymized. Thus, it is difficult to validate some of the important features and results of predictive models with actual industrial data.

III. FEATURE SELECTION

Four classifiers from Scikit package were used in the experiments: Support Vector Machine (SVM), Gradient Boosting Decision Tree (GBDT), Random Forest (RF), and ADABOOST. Among these classifiers, ADABOOST achieved the highest prediction power measured by F1-macro as shown in Fig. 1. As the dataset is highly imbalanced the F1 micro-average is preferable over micro-average strategy.

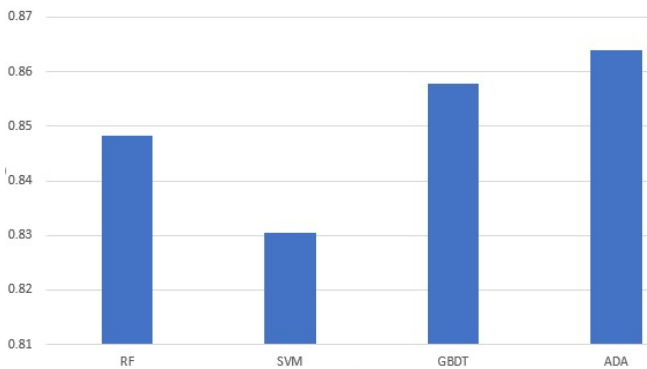


Fig. 1 Comparison of F1 macro scores over 4 classes with 10-fold stratified cross validation using SVM, GBDT, RF, and ADABOOST and all features

There are three categories of feature selection algorithm in feature engineering [7], [8]:

- Filter-based algorithm: Features are selected according to some univariate metric. For example, Minimum redundancy feature selection (mRmR) tends [8] to select features with a high correlation with the class and a low correlation between themselves while Chi-Square adopts a statistical approach to measuring the difference between the expected frequencies and the observed frequencies for two events.
- Embedded algorithms: Feature selection is integrated into the prediction algorithm. As an example, GINI index is a widely adopted algorithm for tree-based predictions based on measurement of the purity of the features with respect to the class. The purity represents the discrimination level of a feature to distinguish between the possible classes. Classification and Regression Trees (CART) is another popular algorithm for three-based prediction.
- Wrapper-based algorithms search heuristically for the optimal subset of features using a specific classification algorithm. Recursive Feature Elimination (RFE) recursively removes features, builds a model using the remaining attributes and calculates model accuracy.

GINI, Chi-Square and RFE algorithms were selected from each of the above three categories, respectively. Their implementations in Python Scikit package were employed to

obtain the most important 16 predictive features as listed in Table I.

TABLE I
 TOP 16 FEATURES SELECTED USING GINI, CHI-SQUARE AND RFE

GINI	Chi-Square	RFE
sinclastcomp2	rotatemean_24hrs	sinclastcomp3
sinclastcomp4	voltstd	sinclastcomp1
rotatemean_24hrs	pressuremean_24hrs	voltmean_24hrs
sinclastcomp3	sinclastcomp1	pressuremean_24hrs
model	vibrationmean_24hrs	vibrationmean_24hrs
pressuremean_24hrs	vibrationmean	sinclastcomp4
voltmean_24hrs	voltmean_24hrs	rotatemean_24hrs
sinclastcomp1	sinclastcomp2	age
vibrationmean_24hrs	sinclastcomp3	model
rotatemean	voltmean	error2count
age	sinclastcomp4	error5count
pressuremean	rotatesd_24hrs	error3count
rotatesd_24hrs	pressuremean	error1count
pressuresd_24hrs	model	error4count
vibrationsd_24hrs	rotatemean	sinclastcomp2
vibrationsd	age	pressuremean

Each of the three feature sets listed in the columns of Table I is provided to ADABOOST to predict machine failures using 10-fold stratified cross validation. The F1 macro scores are shown in Fig. 2 as results of cross validation.

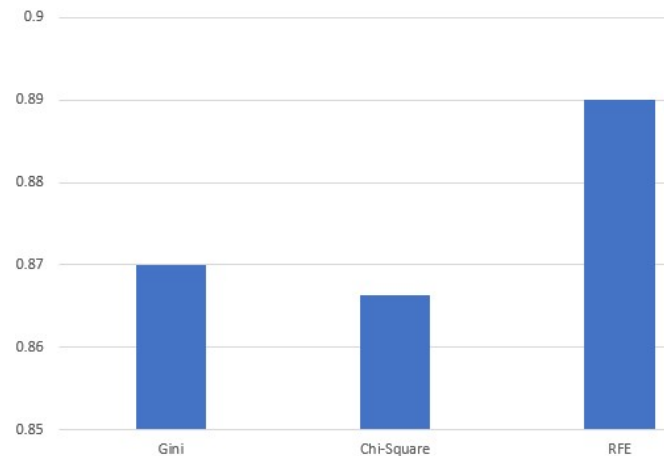


Fig. 2 Comparison of F1 macro scores over 4 classes with 10-fold stratified cross validation using ADA model and the most important 16 features selected by GINI, Chi-Square and RFE

IV. COMBINATION OF FEATURES SELECTED FROM MULTIPLE ALGORITHMS

Table I shows the most important features selected by each feature selection algorithm and highlights the differences particularly between features selected by GINI and RFE. Such disparity is not unexpected as each feature selection algorithm has its weaknesses and strengths that affect the feature selection accuracy. While embedded and wrapper methods are biased to the underlying classifier, filter-based methods are independent of any learning algorithm. However, a filter method may miss a feature that is not important by itself but

very important when combined with other features.

To minimize the bias introduced by a specific feature selection algorithm, 16 most important features selected from each algorithm are aggregated. The combination is performed in two steps. First, each feature in a column in Table I is assigned a ranking value from 1 to 16 with the value 16 assigned to the first feature in the column. Second, for each feature in the Table I, its ranking value in the three columns is summed to obtain a combined ranking value. If the feature is not present in a column, its ranking value is 0. Features are sorted according to their combined ranking values and normalized, shown in descending order in Table II.

TABLE II
 TOP 10 COMBINED FROM FEATURES SELECTED USING GINI, CHI-SQUARE
 AND RFE

Combined top 10 features	Normalized feature importance weight
rotatemean_24hrs	0.83
pressuremean_24hrs	0.79
sincelastcomp1	0.77
sincelastcomp3	0.77
voltmean_24hrs	0.71
vibrationmean_24hrs	0.67
sincelastcomp4	0.67
sincelastcomp2	0.56
model	0.48
age	0.33

To validate that the aggregated features are less bias than the features selected by each of the three individual algorithms, 10 combined features and the top 10 features from each algorithm are provided to ADABOOST model with 10-fold cross validation, respectively. As shown in Fig. 3, the combined features achieved the optimal performance compared to each individual algorithm.

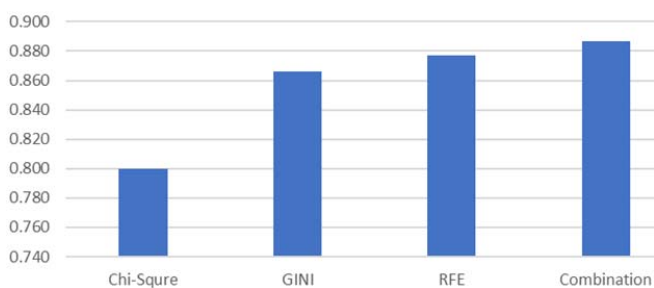


Fig. 3 Comparison of F1 macro average scores over 4 classes with 10-fold cross validation ADA model using the 10 combined features and the 10 most important features selected by GINI, Chi-Square and RFE, respectively

The 10 features shown in Table II represent the relative importance of each feature in the test dataset as a whole and provide a general comparison of the extent to which each feature in the dataset impacts prediction. This suggests that the mean feature values of rotation speed and pressure in 24 hours have the most predictive power in total for four classes in the model. However, Table II does not reveal the features that contribute the most to predicting a specific type of machine failure, as the values are averaged over the four failure classes.

V. PREDICTION MODEL INTERPRETATION

SHAP [10] is applied on the ADABOOST prediction results using the 10 top features listed in Table II for the following three purposes:

- to validate the classifier's decisions and interpret the results of the prediction and provide human-friendly feature explanations; and
- to cross-examine the most important features combined from the top features selected from GINI, Chi-Square and RFE as shown in Table I; and
- to identify the unique features for a specific type of machine failure.

The dataset consists of 4 types of machine failures. To identify features that are specific to a type of failures, One-vs-Rest (OVR) strategy is employed to transform a multi-class problem into binary one.

For a given type of machine failure, the dataset is transformed into positive samples containing the type of failure as the label and all other samples as negatives. The ADABOOST binary classifier is then applied to the transformed dataset for the failure types 1-4, respectively. Next, the SHAP values are obtained using SHAP packager [9] based on the results of classifications of failure types 1-4.

The results of SHAP values of each classification are illustrated in Figs. 4 (a)-(d). Each subfigure lists the features having the most influence on prediction of a specific type machine failures with the most important feature listed at the top. The red horizontal bar illustrates the relative influence level of the corresponding feature in predicting a type of machine failures while the blue bars show the influence of the same feature in predicting of the rest cases including the non-failure case.

The values corresponding to features at the bottom of each figure show the SHAP values. Features with larger absolute SHAP values correspond to more important features.

VI. CONCLUSION AND FUTURE WORK

Our study suggests that features combined from the results of multiple feature selection algorithms in three different categories provide better view on important features with less bias, compared with features selected using a single feature selection algorithm.

The results of the study also indicate the existence of unique features that play dominant roles in the prediction of a specific type of machine failures. As an example, the average pressure in 24 hours is the most important feature in predicting machine failures type 2 and type 3 as shown in Figs. 2 (b) and (c), respectively. However, mean pressure has much less predictive power for failures type 1 and 4. Interestingly, the machine model is the second most important feature for both failures type 2 and type 3 predictions. It might further suggest that failure types 2 and 3 are associated with specific machine models. However, such suggestions are inconclusive due to lack of machine model description in the dataset.

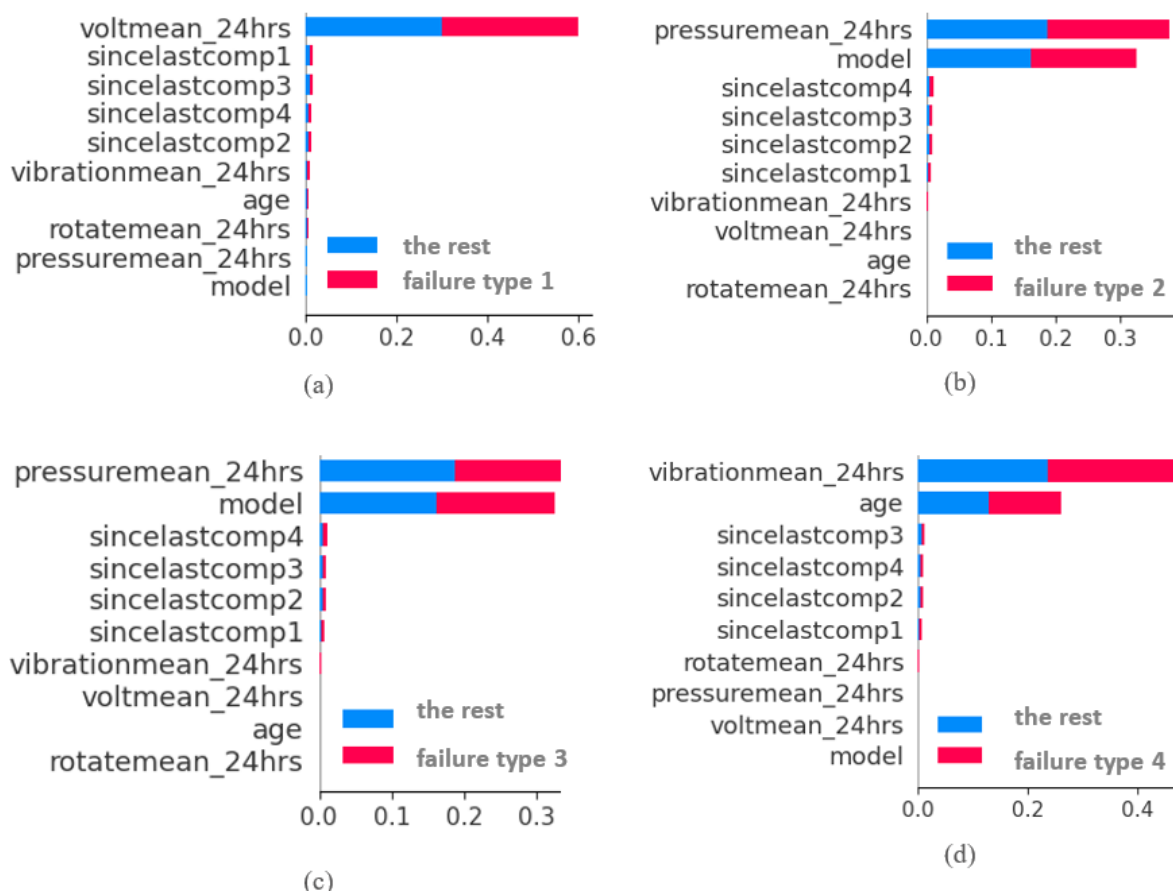


Fig. 4 Feature impact on machine failure prediction: failure type 1 (a), failure type 2 (b), failure type 3 (c), and failure type 4 (d)

The results of the study may help businesses save costs and improve productivity by prioritizing monitoring the important features over less important features. If the features from sensor data play a dominant role in failure prediction, more sensors can be deployed to collect data related to more the important features. In addition, the data associated with more important features can be processed faster and used for classifications in order to predict the features earlier.

A future improvement is to use other data sets, especially the data with descriptive information about the machine, model, and failures to valid the approach proposed in this study. With the availability of such information, the results of this study can be further interpreted.

REFERENCES

- [1] Y. Y. Ran, X. Zhou, P.F. Lin, Y. G. Wen, and R. L. Deng, "A Survey of Predictive Maintenance: Systems, Purposes and Approaches", IEEE Communications Surveys & Tutorials, NOV. 2019.
- [2] G. A. Susto, A. Schirru, S. Pampuri, S. McLoone and A. Beghi, "Machine Learning for Predictive Maintenance: A Multiple Classifier Approach," in IEEE Transactions on Industrial Informatics, vol. 11, no. 3, pp. 812-820, June 2015, doi: 10.1109/TII.2014.2349359.
- [3] Azure AI guide for predictive maintenance solutions. URL: <https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/predictive-maintenance-playbook>.
- [4] Data preparation for predictive maintenance, [https://docs.microsoft.com/en-us/azure/machinelearning/team-data-science-process/predictive-maintenance-playbook#data-preparation-for-](https://docs.microsoft.com/en-us/azure/machinelearning/team-data-science-process/predictive-maintenance-playbook#data-preparation-for-predictivemaintenance)

- [predictivemaintenance.](https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/predictivemaintenance-playbook)
- [5] <https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/predictivemaintenance-playbook>.
- [6] Source: <https://gallery.azure.ai/Experiment/Predictive-Maintenance-Modelling-Guide-Data-Sets-1>.
- [7] I. A. Guyon and Elisseeff, "An introduction to variable and feature selection," J. Mach. Learn. Res., vol. 3, pp. 1157-1182, 2013.
- [8] J. Li, K. Cheng, S. Wang, F. Morstatter, R. Trevino, J. Tang and H. Liu, "Feature Selection: A Data Perspective," ACM Computing Surveys, 50(6), pp.1-45, January 2018.
- [9] <https://github.com/slundberg/shap>.
- [10] S. M. Lundberg and S.I. Lee. "A Unified Approach to Interpreting Model Predictions", Advances in Neural Information Processing Systems (2017), pp. 4768-4777.