Predictive Maintenance System for Production Lines in Manufacturing

Zeynep ACAR, Betül ÇELİK
Bilgisayar Mühendisliği Bölümü
Yıldız Teknik Üniversitesi, 34220 Istanbul, Türkiye
{zeynep.acar1, betul.celik2}@yildiz.edu.tr

***Özetçe* —Bu çalışmada gerçek zamanlı üretilen sensör verileri kullanılarak üretim hatları için makine öğrenmesi temelli kestirimci bakım modeli geliştirilmiştir. Makine öğrenimi algoritmaları kullanılarak arıza gerçekleşmeden önce ne zaman gerçekleşeceğine dair bir tahminde bulunulması amaçlanmıştır. Sonuç olarak önceden gerekli önlemlerin alınarak üretim hattının duraksamdan işleyişe devam etmesine katkı sağlanmıştır.**

Boyut indirgeme ve özellik seçimi uygulanarak makine öğrenimi modelleri bu veriler üzerinde eğitilmiştir. Model sonuçları çeşitli metrikler kullanılarak karşılaştırılmıştır. Elde edilen değerlendirmeler sonucunda bir ensemble algoritması olan Random Forest ve boosting yöntemi olan XGBoost algoritmalarının yüksek performans sergilediği ortaya çıkmıştır.

Anahtar Kelimeler—Makine öğrenimi algoritmaları, önleyici bakım sistemi, özellik seçimi, boyut indirgeme, regresyon

***Abstract*—In this study, a machine learning-based predictive maintenance model was developed for production lines by using real-time sensor data. By using machine learning algorithms, it is aimed to make an estimation of when the failure will occur before it happens. As a result, the necessary precautions were taken in advance and it contributed to the continuation of the production line without stopping.**

Machine learning models are trained on this data by applying dimension reduction and feature selection. Model results were compared using various metrics. As a result of the evaluations, it has been revealed that Random Forest, which is an ensemble algorithm, and XGBoost, which is a boosting method, exhibit high performance.

Keywords—Machine learning, predictive maintenance system, feature selection, dimension reduction, regression

I. INTRODUCTION

Thanks to the digital developments in today's world, Big Data and Artificial Intelligence (AI) are shaping our daily lives and increasing the efficiency of business processes in many areas. Technological and scientific advancements, such as IoT and AI, form the basis of the fourth industrial revolution. During this period, AI and Big Data are changing the way modern people live and work by offering solutions to complex problems. Firms constantly need to increase their performance due to the intense competition in their respective markets. In this regard, companies that can adapt to the technological developments of the period are one step ahead in this competitive environment[1].

A. Data Collection and Preventive Maintenance

Today, data is collected from various sources in diverse ways, leading to the transformation of the collected data into more complex structures, thereby making data analysis challenging. The results of this analysis are crucial for making informed decisions on significant matters such as real-time operational management and risk and error detection. IoT devices play a pivotal role in facilitating the connection and exchange of data among production systems.

The production environment relies on the seamless operation of its equipment. Any malfunction in a component or subsystem can lead to a complete halt in the production line, resulting in significant financial losses for the company. While it may not be feasible to detect all errors in advance, a majority of the errors that cause disruptions can be predicted. This enables proactive measures to be taken, eliminating the error without interrupting production and preventing potential financial losses.

Predictive maintenance (PdM), also known as predictive maintenance, has gained widespread adoption to reduce maintenance costs and ensure sustainable operations. The primary objective of predictive maintenance is to anticipate the occurrence of faults to proactively perform preventive maintenance. Data-driven AI applications utilizing data collected from IoT devices can significantly contribute to preventive maintenance.

B. Purpose of This Study

The objective of this study is to develop a realistic preventive maintenance system that accurately predicts potential failures in production lines, employing machine learning methods. To identify the most suitable model for this problem, multiple algorithms were thoroughly examined and compared using a real-world dataset obtained from a company.

This report begins by reviewing existing studies on similar topics in the literature. The feasibility section outlines the project's stages and progression. The system analysis defines the project's objectives and requirements. The system design section introduces the methods employed, and the application section includes the results obtained from these methods. In the subsequent stages, the results will be compared, and a performance analysis will be conducted to determine the most effective algorithms based on the outcomes.

II. Related Work

A. Deep Learning for Improved System Remaining Life Prediction

In the study conducted by Zhang et al., a deep learning-based approach was developed for monitoring production system performance and predicting degradation. Specifically, the Long Short-Term Memory (LSTM) network, a deep architecture capable of capturing temporal patterns in time series data, was investigated to track system state degradation. The evaluation was conducted on NASA's C-MAPSS engine dataset. The results demonstrated that the deep LSTM network outperformed other commonly used machine learning techniques such as SVR, deep CNN, and deep RNN. Future work will focus on exploring the broader applicability of deep LSTM compared to other machine learning techniques and developing more efficient methods for network parameter tuning[2].

B. Statistical Approach to Predictive Maintenance

The objective of the study titled "ARIMA Model-based Real-Time Trend Analysis for Predictive Maintenance" by Francis et al. is to collect real-time sensor values and provide online access to manufacturers to enable continuous monitoring of the train's health condition, thereby enhancing train operations. Trend analysis is conducted to identify the occurrence of errors and their specific locations. For trend analysis and forecasting, the study employs the ARIMA Model, which is a statistical analysis method. However, when comparing ARIMA methods with machine learning techniques, it becomes apparent that ARIMA is not as successful or flexible[3].

C. Preventive Maintenance of the System with Artificial Intelligence Techniques

The study titled "Predictive Maintenance of Machine Tool System Using Artificial Intelligence Techniques Applied to Machine Condition Data" by Lee et al. introduces the design and optimization of artificial intelligence-based algorithms for monitoring two critical machine tools. The approach relies on a data-driven modeling methodology. The two components under monitoring are the cutting tool and the spindle motor. The algorithms are utilized for monitoring tool wear and bearing defects. To enhance the recognition capabilities of the AI model, proper processing of the raw fabrication dataset and selection of meaningful features from the extensive dataset are essential. The results demonstrate that the algorithms applied to features extracted from experimental data effectively monitor and assess the deterioration state of the instruments[4].

III. Preliminary Preparation

The primary requirement for an IoT application is to be effective and scalable, enabling the utilization of continuous and real-time data flow from sensors deployed in the factory for decision-making processes. The IoT system should incorporate a framework that allows seamless integration with external systems. By incorporating an integrated data flow forecasting model, a Predictive Maintenance (PdM) system can detect system failures, generate alarms based on predefined rules, execute commands in production systems, and send real-time warning messages.

A. Data Set

The dataset utilized in this project comprises 26,431 rows and 19 features, all consisting of numeric values. Prior to analysis, the dataset underwent normalization and standardization processes, ensuring the absence of missing, incorrect, or inconsistent values. The project's objective is to generate estimation results using 18 of these features and compare them with the actual values for analysis. Given the nature of the predictive maintenance model, the dataset may exhibit instability. However, this does not pose a problem as regression analysis is employed, and no classification is performed. No additional measures have been taken to balance the dataset.

IV. Implementing PdM in Dataset

A. Correlation Analysis

Data discovery is one of the most important parts of this work. Correlation analysis shows the direction (positive or negative) and strength of the relationship between different variables. In this project, correlation analysis was used to examine the relationships between the data collected from the sensors and the remaining life before failure, which was taken into account in the modeling task.

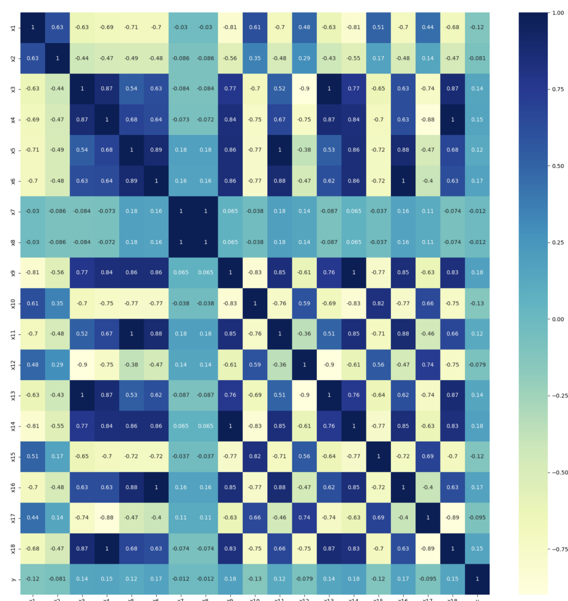

Figure 1 Predictive Maintenance System Architecture

In Figure 2, the relationship between all the features in the dataset is visualized using the correlation table. It has been observed that there is a high correlation between certain features. For instance, upon examining the correlation table, it is evident that features x3, x4, x5, x6, x13, and x14 exhibit a strong correlation. As features with high correlation can be represented in a more concise manner, they were considered for size reduction during the feature selection phase.

B. Dataset Analysis

The PCA algorithm is often used to reduce high-dimensional spaces to smaller sizes without much loss. In our study, we used this method to reduce the high-dimensional data we have. According to the PCA results, %95 of the variance in the data set can be represented by 6 principal components.

According to the results, it was observed that x1, x2, x7, x8 and x12 were more dominant in general, although the features that contributed more variance for the principal components varied.

Then, in order to find out which features are of higher importance, we generally preferred the Random Forest Regressor algorithm, which performs well in regression problems and can be easily interpreted.

The algorithm result is shown in figure 2. According to this chart, we selected the top 9 most important features, x1, x2, x5, x6, x7, x8, x11, x12, x13. Compared with the PCA results, the features that contribute high variance to the principal components in the PCA results are also among these 9 features.

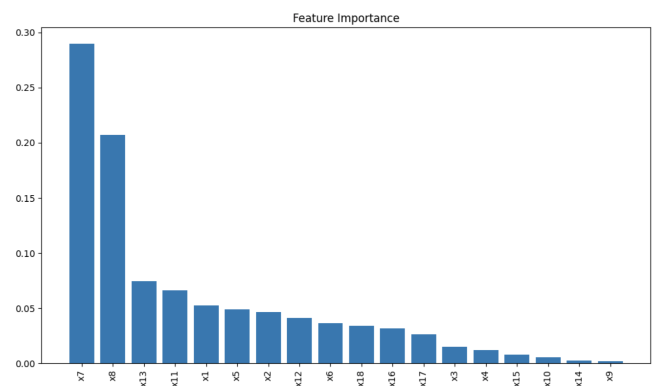

Figure 2 Feature Importances

The features determined in the data set were used to train Random Forest, Gradient Boosting, XGBoosting, Decision Tree and MLP Regressor models. And as a result of Random Search, the best performing parameters were determined for the models.

V. Model Results

There are various evaluation metrics for the model's outputs on the dataset, as well as for machine builds augmentation. Mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE), R-square (R2) and mean absolute percent error (MAPE) metrics are commonly used for this purpose in regression problems. These metrics were used because this study deals with a regression problem.

Looking at Table 1, as a result of the evaluations, it is seen that Decision Tree with 0.60 R2 value and MLP algorithms with 0.72 R2 value outperform other algorithms. It is seen that the best results belong to Random Forest and XGBoost algorithms with an R2 value of 0.98. When the MAPE values were compared, the Random Forest algorithm gave a better result, albeit with a small margin.

Table 1 Random Forest ile En İyi Sonuçlar

Algoritmalar	R^2	MSE	MAE	MAPE	RMSE
Random Forest	0.98	146.80	53.20	3.26	21552.27
XGBoost	0.98	154.29	59.17	3.28	23808.40
Gradient Boosting	0.97	177.60	91.85	4.74	31541.93
Decision Tree	0.72	580.44	372.84	20.66	336914.71
MLP Regressor	0.60	701.57	531.32	27.35	492209.00

Comparison of the estimation results resulting from the applied models with the actual values is visualized in Figure 3. Looking at Table 1, it is expected that the differences between the forecast results and the actual values in the MLP Regressor and Decision Tree plots are larger. It is seen that the other model results are compatible with the values in the table.

VI. Conclusion

In this project, it is aimed to develop a machine learning-based predictive maintenance model for production lines. During the project process, sensor data from a private company was used to make the necessary improvements and analyzes.

Initially, the data set was analyzed and it was determined that there were no missing values or outliers. In addition, since normalization operations were performed on the data set beforehand, no further processing was required for this purpose. Afterwards, the correlation matrix of the data set was created and it was examined which features had a high correlation with each other.

After the analyzes were made, methods were used to reduce the size of the data set. First of all, with principal component analysis, it was determined that the %95 variance could be expressed with 6 principal components and information was obtained about which features contributed more to this variance. Following this process, Random Forest algorithm was applied for feature selection. The 9 most important features determined according to the model result were selected.

Using the selected features, the data set was divided into two sets as training and testing. Afterwards, 5 different machine learning algorithms were applied to train the data. Random Search and Grid Search algorithms were evaluated to determine the optimum values of the parameters to be used in the models. Since the Grid Search algorithm takes too long to work, it has been decided to apply the Random Search algorithm.

The applied models give the estimated available time remaining before the failure as output. R^2, MAE, MAPE, MSE, RMSE metrics used in regression problems were used to evaluate and compare the performance of these results. Considering the evaluation results, it was observed that

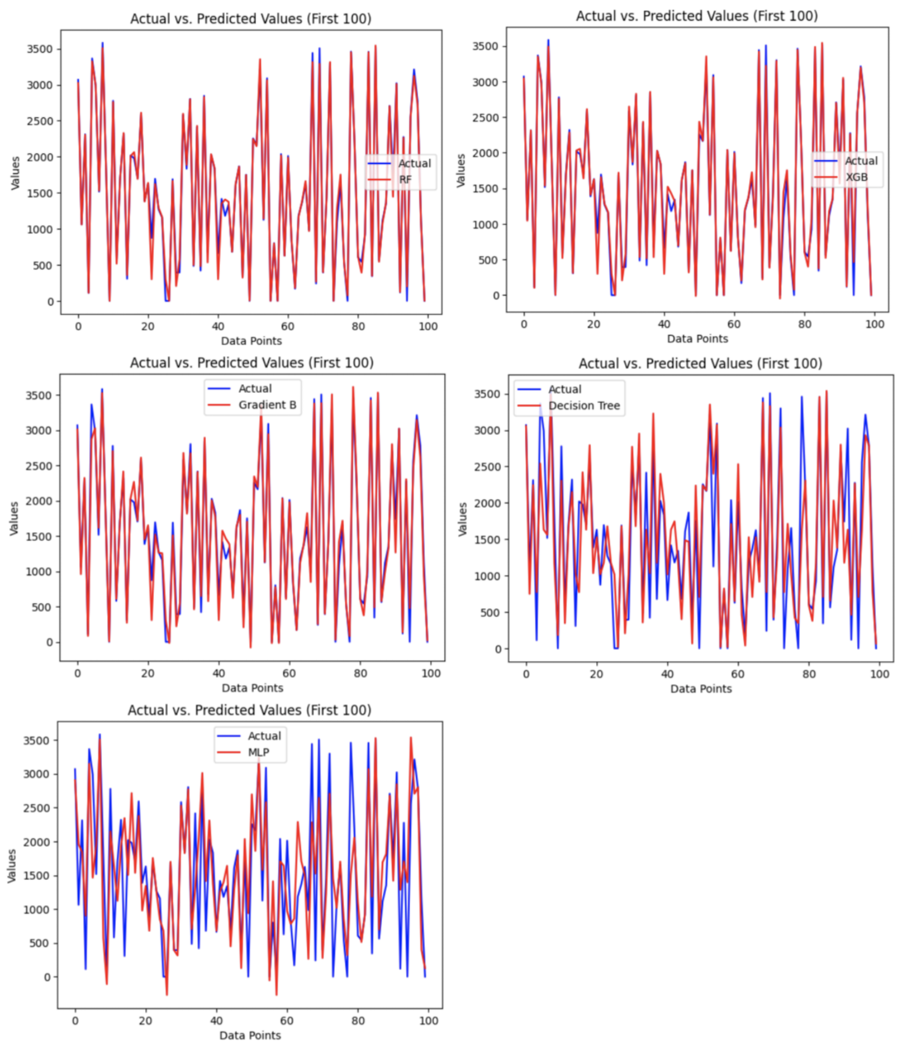

Figure 3 Algoritma Sonuç Karşılaştırmaları

while the success rate of MLP and Decision Tree algorithms was quite low, Random Forest and XGBoost algorithms achieved a higher success.

As a result, it has been determined that Random Forest and XGBoost algorithms are better with 0.98 R^2 success to determine the estimated remaining available time before failure. In this study, only one company's data was used, and studies can be carried out on different data sets in order to better evaluate the performance of the developed model. Although the developed model performs quite well on the data set used, the results can be compared by training different algorithms.

References

[1] S. Ayvaz and K. Alpay, "Predictive maintenance system for production lines in manufacturing: A machine learning approach using iot data in real-time," *Expert Systems with Applications*, vol. 173, p. 114598, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417421000397

[2] J. Zhang, P. Wang, R. Yan, and R. X. Gao, "Deep learning for improved system remaining life prediction," *Procedia CIRP*, vol. 72, pp. 1033–1038, 2018, 51st CIRP Conference on Manufacturing Systems. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2212827118304335

[3] F. Francis and M. Mohan, "Arima model based real time trend analysis for predictive maintenance," in *2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA)*, 2019, pp. 735–739.

[4] W. J. Lee, H. Wu, H. Yun, H. Kim, M. B. Jun, and J. W. Sutherland, "Predictive maintenance of machine tool systems using artificial intelligence techniques applied to machine condition data," *Procedia CIRP*, vol. 80, pp. 506–511, 2019, 26th CIRP Conference on Life Cycle Engineering (LCE) Purdue University, West Lafayette, IN, USA May 7-9, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2212827118312988