RES MILITARIS
REVUE EUROPEENNE D ETUDES EUROPEAN JOURNAL OF MILITARY STUDIES
Social Science Journal

Real-time Anomaly Detection and Predictive Maintenance in Metro Train APU Compressors: A Data-driven Approach

Swathi[1], Anand Gangula[2], Maruth Reddy Vootukuri[2], Rakesh Manchiryala[2]

[1]Assistant Professor, [2]UG Scholar, [1,2]Department of Computer Science and Engineering

[1,2]Kommuri Pratap Reddy Institute of Technology, Ghatkesar, Hyderabad, Telangana.

ABSTRACT

Metro rail systems play a crucial role in contemporary urban transportation networks. Ensuring the dependable functioning of auxiliary power units (APU) is essential for the overall efficiency and security of metro trains. Utilizing anomaly detection in APU compressors can effectively mitigate failures and reduce downtime, hence improving the efficiency and dependability of metro services. Anomaly detection in industrial settings often utilizes rule-based systems or threshold-based alarms as conventional solutions. Although these systems may have a certain level of effectiveness, they may not accurately detect minor irregularities or adjust well to changing operational circumstances. The main objective is to create a system that can consistently monitor APU compressors and identify any irregularities in their functioning. Therefore, this work seeks to transform maintenance methods in metro systems by utilizing sophisticated data analytics and machine learning approaches. This research aims to create a system that can independently and accurately detect anomalies by collecting and analyzing real-time operating data from APU compressors. By incorporating machine learning algorithms, it becomes possible to detect intricate patterns that may indicate prospective problems. This allows for prompt interventions to avert breakdowns and assure the continuous functioning of metro train systems. This technological progress shows significant potential for improving the safety, effectiveness, and dependability of urban transportation systems.

Keywords: APU compressors, anomaly detection, predictive maintenance, machine learning.

1. INTRODUCTION

Transportation vehicles can be maintained using many strategies, such as preventative maintenance, corrective maintenance, and condition-based maintenance. Preventive maintenance involves conducting frequent inspections according to a predetermined schedule, during which equipment is either replaced or repaired. This form of maintenance results in inefficient allocation of resources towards repairing or replacing equipment that is still functional in order to prevent unexpected breakdowns, as well as the loss of time spent addressing emergencies and analyzing issues. Corrective maintenance [2] involves waiting for a breakdown to happen before repairing the equipment. Condition-based maintenance is a maintenance method that assesses the current state of a system in order to determine the necessary maintenance tasks. Predictive maintenance (PdM) is a technique that relies on data analysis tools to evaluate historical and real-time data from different components of a system. Its purpose is to identify abnormalities and potential equipment flaws, allowing for timely repairs to prevent system failures. In recent years, machine learning techniques, particularly deep learning, have been proposed for predictive maintenance (PdM). Deep learning technologies, such as Deep Neural Network, Recurrent Neural Network, Convolution Neural Network, and Long Short-Term Memory, can anticipate the likelihood of equipment failure by autonomously analyzing historical data of the system. Sparse autoencoders are highly effective deep neural networks that have been effectively utilized for the purpose of failure detection. Autoencoders, as a type of machine learning model that does not require labeled data, have the ability to autonomously acquire features from data that does not have predefined categories. A recent analysis of literature reveals that PdM methods can be categorized into three primary classifications: model-based, knowledge-based, and data-driven approaches. Data-driven predictive maintenance (PdM) methodologies [4] identify faults and irregularities by examining the data obtained from various sensors in real-time. The data-driven algorithms efficiently integrate a substantial volume of real-time data from sensors to forecast and identify failures. These algorithms have garnered significant interest in contemporary industrial systems, as evidenced by references [5]–[8]. This research presents the implementation of a data-driven Prognostics and Health Management (PdM) framework using a sparse autoencoder. The framework aims to detect and predict faults in the air production unit (APU) system of a train in Metro do Porto. This system is essential and in high demand for the vehicle's operation. Its breakdown, without any backup, leads to the immediate need for vehicle repair. This has a significant impact not only on the running corporation, but primarily on the customers who witness their expectations of trust in transportation being undermined. The objective is to detect and differentiate between typical and atypical patterns in the data flow derived from a collection of sensors integrated into the APU system during train operation. The goal is to utilize unsupervised methods grounded in deep learning to forecast the progression of a failure.

1.1 Objective

The primary objective is to develop a robust and reliable system for real-time anomaly detection in metro train APU compressors, enhancing the safety, reliability, and efficiency of metro rail operations. The system should be capable of accurately identifying abnormal behaviors or patterns indicative of potential faults or malfunctions, thereby facilitating proactive maintenance and minimizing service disruptions.

1.2 Research Motivation

- Safety: Ensuring the safety of passengers and personnel aboard metro trains is paramount. Anomaly detection in APU compressors can help prevent catastrophic failures and accidents.
- Operational Efficiency: Detecting anomalies early allows for proactive maintenance, reducing the risk of unexpected breakdowns and minimizing service disruptions. This contributes to smoother operations and improved passenger experience.
- Cost Savings: Timely detection and mitigation of anomalies can prevent costly repairs and downtime, leading to significant cost savings for metro train operators.
- Data-driven Decision Making: Leveraging operational data for anomaly detection enables data-driven decision-making processes, optimizing maintenance schedules and resource allocation.

1.3 Problem Statement

Metro train APU (Auxiliary Power Unit) compressors play a crucial role in providing auxiliary power to various systems within the train, including air conditioning, lighting, and ventilation. Anomalies or malfunctions in these compressors can lead to system failures, service disruptions, and safety hazards for passengers and personnel onboard. The problem statement involves developing a system for real-time anomaly detection in metro train APU compressors using operational data. The system aims to detect abnormal patterns or behaviors indicative of potential faults or malfunctions in the compressors, allowing for timely intervention and maintenance to ensure the safety, reliability, and efficiency of metro train operations.

1.4 Applications

- Transportation Industry: The project's application lies primarily in the transportation sector, particularly in metro rail systems. Anomaly detection in APU compressors ensures the reliability and safety of metro train operations.

- Predictive Maintenance: The techniques developed in this project can be applied to various other domains beyond metro trains, such as industrial machinery, aerospace, and automotive industries, for predictive maintenance and condition monitoring.
- Smart Cities: Anomaly detection in critical infrastructure components like APU compressors contributes to the development of smart cities by enhancing the efficiency and reliability of public transportation systems.
- Data Analytics and Machine Learning: The project serves as a practical example of applying data analytics and machine learning techniques to solve real-world problems, fostering advancements in these fields and promoting interdisciplinary research.

2. LITERATURE SURVEY

Failures can be detected by finding patterns in data that do not correspond to normal system behavior and which represent anomalies. In the past decades, several anomaly detection approaches have been proposed for of failure prediction or early failure detection, e.g., [9], [10]. More specifically, and regarding railway industry, two recent literature surveys on the work related to different PdM methods can be found in [11] and [12]. Among all the proposed PdM methods for the railway industry, we are interested in data-driven based on learning methods. A recent work in [13] explores data-driven PdM based on anomaly and novelty detection implemented to predict failure in the automatic door system of the train and prevents the spread of breakdown in the system. The authors developed and implemented four common learning algorithms for anomaly detection. Moreover, the results show that a low-pass filter can significantly reduce the number of false alarms. Lee [14] used a logistic regression classifier to model the compressor behavior used for air leakage detection by anomaly detection in a train's braking pipes. Also, a density-based clustering method with a dynamic density threshold was used to distinguish anomalies from outliers and detect anomalies based on the severity degree. More recently, Chen et al. [15] focus specifically on predicting compressor failures using a recurrent neural network using Long Short-Term Memory (LSTM) architecture. The authors compared their method and a random forest method where the experimental results show that predictions by LSTM stay significantly more stable over time, while in terms of AUC score random forest slightly outperforms the LSTM. Most recently, Barros et al. [16] developed a real-time data analysis of the sensors installed on APUs that detects anomalies. They also provided rules based on peak frequency analysis. They considered definition of normal and abnormal behavior of sensors data which can be used for APU failure detection.

3. PROPOSED SYSTEM

The proposed methodology aims to develop a system for real-time anomaly detection in the APU (Auxiliary Power Unit) compressors of metro trains using operational data. Anomaly detection helps in identifying unusual patterns or events that deviate from normal behavior, which could indicate potential faults or malfunctions in the compressors.

Key Components:

- Graphical User Interface (GUI): The project includes a GUI built using the Tkinter library in Python. The GUI provides an intuitive interface for users to interact with different components of the system.
- Data Preprocessing: Upon uploading the dataset, the system preprocesses the data to prepare it for training the anomaly detection models. Preprocessing steps may include handling missing values, converting timestamps to date components, and splitting the data into training and testing sets.
- Machine Learning Models: The project utilizes machine learning models for anomaly detection. Specifically, it implements

- Custom k-Nearest Neighbors (KNN) Classifier: A kNN classifier is trained using the operational data to classify instances as normal or anomalous.
- Ensemble Model (Random Forest Classifier + Multi-layer Perceptron): An ensemble model is constructed by combining predictions from Random Forest Classifier (RFC) and Multi-layer Perceptron (MLP) Classifier. This ensemble approach aims to improve anomaly detection accuracy.
- Performance Evaluation: After training the models, the system evaluates their performance using various metrics such as precision, recall, F1-score, and accuracy. Confusion matrices are also generated to visualize the model's performance in classifying anomalies.
- Prediction on Test Data: Users can upload a test dataset to the system for anomaly prediction. The trained models are then used to predict anomalies in the test data, and the results are displayed to the user.
- Graphical Visualization: The system provides graphical visualization of classifier performance through bar graphs. This allows users to compare the performance of different classifiers based on evaluation metrics.

Overall, the project provides a comprehensive solution for real-time anomaly detection in metro train APU compressors, leveraging machine learning techniques and graphical visualization to analyze operational data and identify potential faults or anomalies in the compressors. It offers a user-friendly interface for dataset handling, model training, evaluation, and prediction, facilitating efficient monitoring and maintenance of metro train systems.

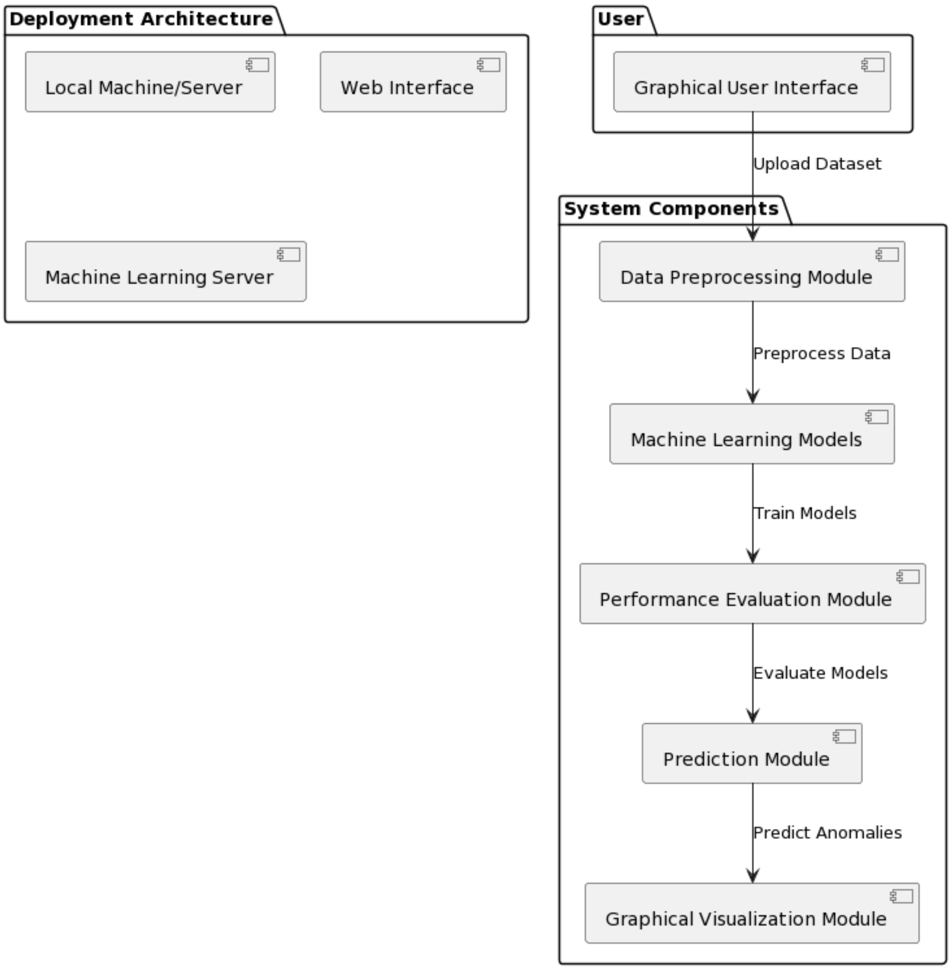

Figure 1: Architectural block diagram of proposed system.

In this project for real-time anomaly detection in metro train APU compressors, the ensemble model refers to a machine learning approach that combines the predictions of multiple individual classifiers to improve overall performance. The ensemble model enhances the anomaly detection capability by leveraging the complementary strengths of Random Forest and MLP Classifiers, ultimately improving the reliability and effectiveness of the anomaly detection system for metro train APU compressors.

Ensemble Model Components

- Random Forest Classifier (RFC): Random Forest is a popular ensemble learning method that builds a multitude of decision trees during training and outputs the class that is the mode of the classes (classification) of the individual trees. In this project, the Random Forest Classifier is trained on the preprocessed data to detect anomalies in the APU compressors.
- Multi-layer Perceptron (MLP) Classifier: MLP is a type of artificial neural network that consists of multiple layers of nodes (neurons), each connected to the next layer. In this project, the MLP Classifier is another individual classifier trained on the preprocessed data to detect anomalies.

Ensemble Model Strategy

The ensemble model combines the predictions from the Random Forest Classifier and the MLP Classifier using a simple averaging ensemble approach. The steps involved in this strategy are:

- Training Phase: Both Random Forest and MLP Classifiers are trained on the preprocessed data independently.
- Prediction Phase: After training, both classifiers make predictions on the test dataset separately.
- Combining Predictions: The ensemble model then combines the predictions from both classifiers using a simple averaging approach. For each instance in the test dataset, the ensemble model takes the majority vote or average prediction from both Random Forest and MLP Classifiers.
- Result: The combined predictions form the final output of the ensemble model, which is then used for anomaly detection.

Benefits of Ensemble Model

- Improved Accuracy: Ensemble models often achieve higher accuracy than individual classifiers by leveraging the strengths of multiple models.
- Robustness: Ensemble models are less prone to overfitting and generalization errors compared to single models.
- Versatility: By combining different types of classifiers, ensemble models can capture diverse patterns in the data, leading to better performance.

4. RESULTS AND DISCUSSION

Dataset Description

The dataset is system and its process where various measurements are recorded over time.

- timestamp: This column represents the time at which the measurements were recorded. It serves as a chronological reference for the dataset.
- TP2 and TP3: These columns represent measurements or parameters labeled as "TP2" and "TP3." The specific meaning of these parameters would depend on the context of the system.
- H1: This column represents a measurement or parameter labeled as "H1." The specific meaning of this parameter would depend on the context of the system.

- DV_pressure: This column represents a measurement of pressure, related to a device labeled "DV."
- Reservoirs: This column contain information related to reservoirs in the system. It represent their status or relevant measurements.
- Oil_temperature: This column represents the temperature of the oil in the system.
- Motor_current: This column represents the current flowing through a motor in the system.
- COMP: This column represent a measurement labeled as "COMP." The specific meaning would depend on the context of the system.
- DV_eletric: This column likely represents an electrical parameter related to a device labeled "DV."
- Towers: This column contain information related to towers in the system, potentially representing their status and relevant measurements.
- MPG: This column represent a measurement or parameter labeled as "MPG."
- LPS: This column represents a measurement or parameter labeled as "LPS."
- Pressure_switch: This column may represent the status or measurement of a pressure switch in the system.
- Oil_level: This column represents the level of oil in the system.
- Caudal_impulses: This column represent measurements or impulses related to the flow rate in the system.
- Anomaly: This column is a binary indicator (0 or 1) representing whether an anomaly is detected in the system based on the recorded measurements.

Results Description

Figure 2 displays the confusion matrix generated for the KNN classifier. A confusion matrix provides a tabular representation of model predictions versus actual class labels, helping to assess the classifier's performance in terms of true positives, true negatives, false positives, and false negatives.

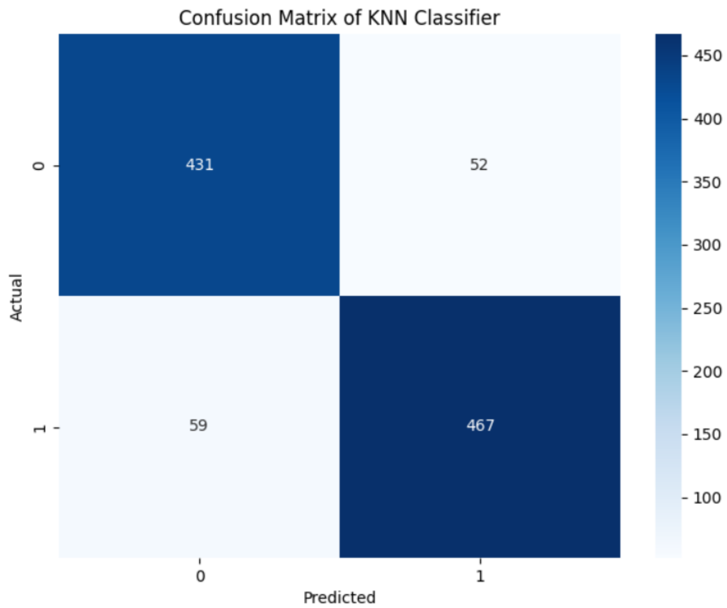

Figure 2: Displays the KNN model confusion matrix.

Figure 3 depicts the confusion matrix for the proposed ensemble model. It allows for a visual assessment of how well the ensemble model performs in classifying anomalies compared to the individual classifiers. Figure 4 presents a graphical comparison of the performance metrics obtained from different classifiers or models. It provides an easy-to-understand visualization for users to compare the effectiveness of various algorithms in detecting anomalies in metro train APU compressors.

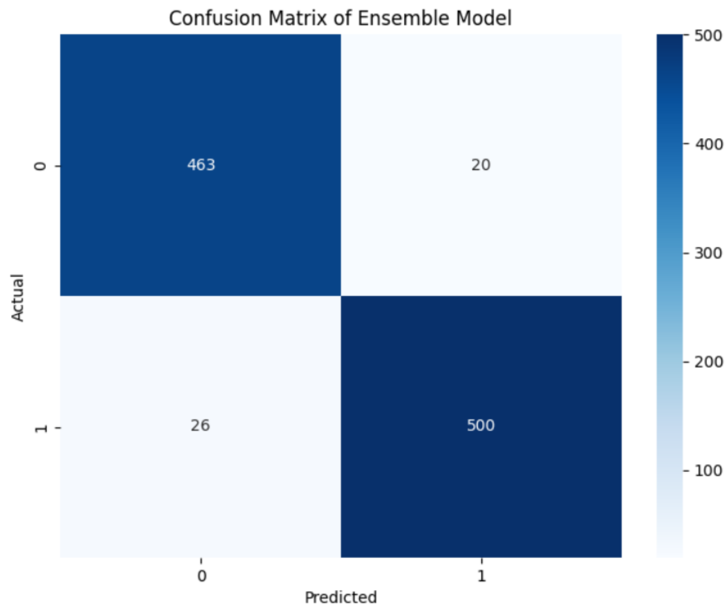

Figure 3: Displays the proposed ensemble model confusion matrix.

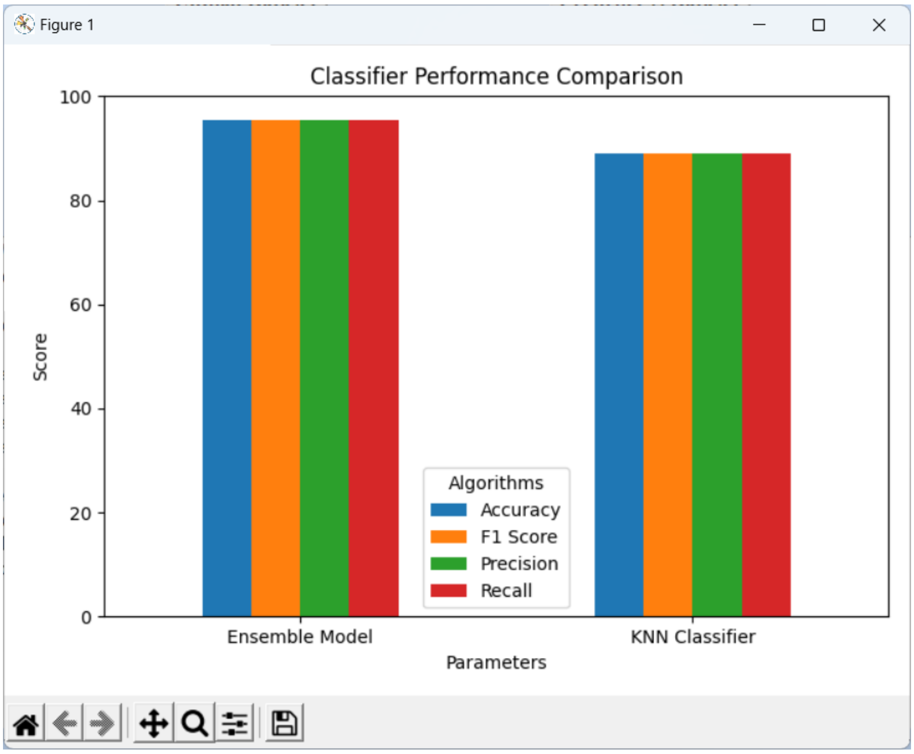

Figure 4: Plot shows the performance Comparison Graph.

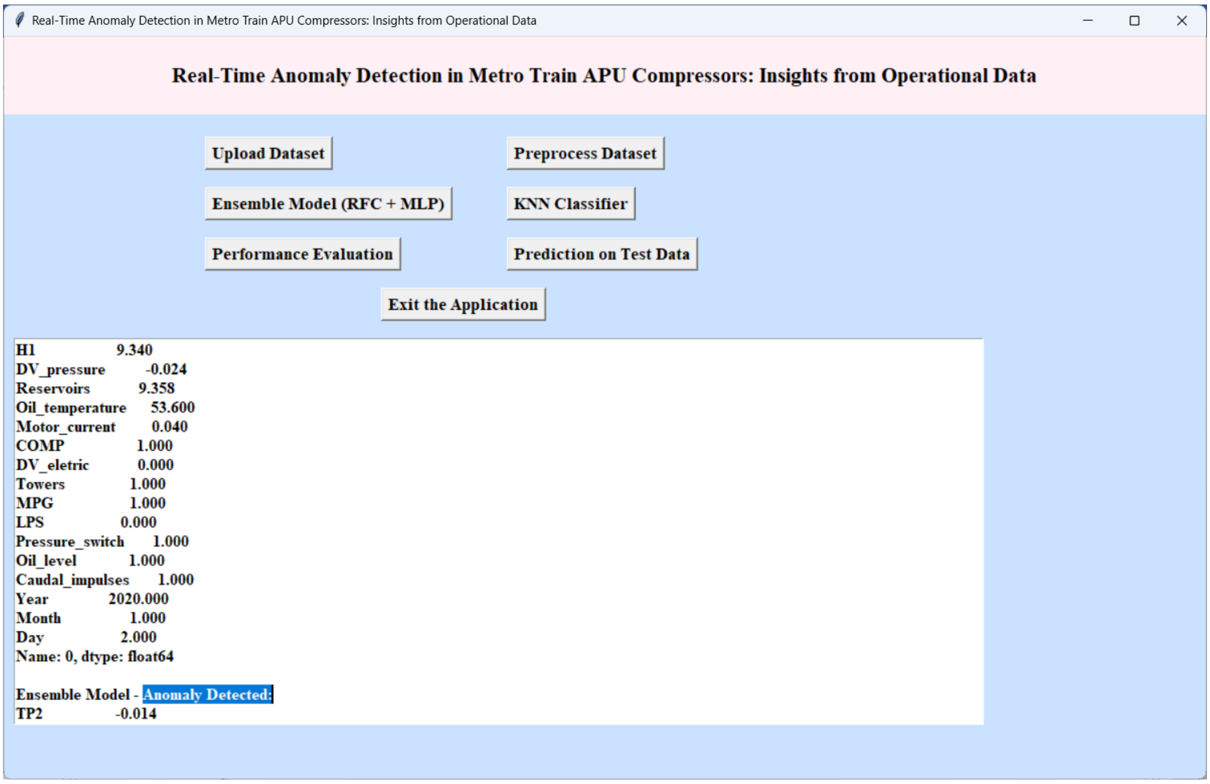

Figure 5: Prediction on new test data.

Figure 5 illustrates the process of making predictions on new test data using the trained models. It displays the input features of the test data, along with the model's prediction of whether each instance represents an anomaly or normal behavior.

Table 1: Performance comparison of quality metrics obtained using KNN and ensemble model.

Model	KNN	Ensemble model
Accuracy (%)	88.99	95.44
Precision (%)	88.96	95.41
Recall (%)	89.008	95.45
F1-score (%)	88.984	95.43

For the KNN model:

- The Accuracy is 88.99, indicating the accuracy between the actual and predicted values
- The Precision is 88.96, suggesting that, on average Precision between the actual and predicted values.
- The Recall is 89.008, suggesting that, on average Recall between the actual and predicted values.
- The F1-score is 88.984, representing the average F1-score between the actual and predicted values.

For the Random Forest model:

- The Accuracy is 95.44, indicating the accuracy between the actual and predicted values.

- The Precision is 95.41, suggesting that, on average Precision between the actual and predicted values.
- The Recall is 95.45, suggesting that, on average Recall between the actual and predicted values.
- The F1-score is 95.43, representing the average F1-score between the actual and predicted values.

5. CONCLUSION

The project on real-time anomaly detection in metro train APU compressors represents a significant advancement in maintenance practices for metro systems. The reliable operation of APU is crucial for the overall performance and safety of metro trains, making the detection of anomalies in APU compressors a critical aspect of ensuring efficient and uninterrupted metro services. The conventional methods of anomaly detection in industrial settings have limitations in capturing subtle deviations and adapting to evolving operating conditions. This research addresses these challenges by leveraging advanced data analytics and machine learning techniques for real-time anomaly detection. By collecting and analyzing real-time operational data from APU compressors, the project aims to develop a system that can autonomously and accurately detect anomalies. The integration of machine learning algorithms enhances the capability to identify complex patterns indicative of potential issues, enabling timely interventions to prevent failures and minimize downtime. This innovation holds great promise for revolutionizing maintenance practices in metro systems, contributing to the safety, efficiency, and reliability of urban transportation networks.

REFERENCES

[1] G. Budai, D. Huisman, and R. Dekker, “Scheduling preventive railway maintenance activities,” Journal of the Operational Research Society, vol. 57, no. 9, pp. 1035–1044, 2006.
[2] C. Stenstrom, P. Norrbin, A. Parida, and U. Kumar, “Preventive and ¨ corrective maintenance–cost comparison and cost–benefit analysis,” Structure and Infrastructure Engineering, vol. 12, no. 5, pp. 603–617, 2016.
[3] J. Yan, Y. Meng, L. Lu, and L. Li, “Industrial big data in an industry 4.0 environment: Challenges, schemes, and applications for pre dictive maintenance,” IEEE Access, vol. 5, pp. 23 484–23 491, 2017.
[4] W. Zhang, D. Yang, and H. Wang, “Data-driven methods for predictive maintenance of industrial equipment: a survey,” IEEE Systems Journal, vol. 13, no. 3, pp. 2213–2227, 2019.
[5] K. Wang, “Intelligent predictive maintenance (ipdm) system–industry 4.0 scenario,” WIT Transactions on Engineering Sciences, vol. 113, pp. 259–268, 2016.
[6] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, “Deep learning and its applications to machine health monitoring,” Mechanical Systems and Signal Processing, vol. 115, pp. 213–237, 2019.
[7] G. Li, C. Deng, J. Wu, X. Xu, X. Shao, and Y. Wang, “Sensor data-driven bearing fault diagnosis based on deep convolutional neural networks and s-transform,” Sensors, vol. 19, no. 12, p. 2750, 2019.
[8] Y. Jiang and S. Yin, “Recursive total principle component regression based fault detection and its application to vehicular cyber-physical systems,” IEEE Transactions on Industrial Informatics, vol. 14, no. 4, pp. 1415–1423, 2017.
[9] V. Chandola, A. Banerjee, and V. Kumar, “Anomaly detection: A survey,” ACM computing surveys (CSUR), vol. 41, no. 3, pp. 1–58, 2009.
[10] R. Chalapathy and S. Chawla, “Deep learning for anomaly detection: A survey,” arXiv preprint arXiv:1901.03407, 2019.

[11] E. Fumeo, L. Oneto, and D. Anguita, “Condition based maintenance in railway transportation systems based on big data streaming analysis,” Procedia Computer Science, vol. 53, pp. 437–446, 2015.

[12] P. C. L. Gerum, A. Altay, and M. Baykal-Gursoy, “Data-driven predictive ¨ maintenance scheduling policies for railways,” Transportation Research Part C: Emerging Technologies, vol. 107, pp. 137–154, 2019.

[13] R. P. Ribeiro, P. Pereira, and J. Gama, “Sequential anomalies: a study in the railway industry,” Machine Learning, vol. 105, no. 1, pp. 127–153, 2016.

[14] W.-j. Lee, “Anomaly detection and severity prediction of air leakage in train braking pipes,” International Journal of Prognostics and Health Management, vol. 21, 2017.

[15] K. Chen, S. Pashami, Y. Fan, and S. Nowaczyk, “Predicting air compressor failures using long short term memory networks,” in EPIA Conference on Artificial Intelligence. Springer, 2019, pp. 596–609.

[16] Barros, M., Veloso, B., Pereira, P.M., Ribeiro, R.P., Gama, J. (2020). Failure Detection of an Air Production Unit in Operational Context. In: Gama, J., et al. IoT Streams for Data-Driven Predictive Maintenance and IoT, Edge, and Mobile for Embedded Machine Learning. ITEM IoT Streams 2020 2020. Communications in Computer and Information Science, vol 1325. Springer, Cham. https://doi.org/10.1007/978-3-030-66770-2_5