Vibration Analytics for Smart Facilities

by

CHUNG King To

Submitted in partial fulfillment of the requirements for the degree of

Bachelor of Science (Honours)

in Business Computing and Data Analytics

Hong Kong Baptist University

April, 2024

ABSTRACT

This comprehensive study presents a holistic approach to predictive maintenance through vibration analytics, specifically targeting motors within water supply systems. Leveraging a dataset spanning 10 months, our methodology encompassed a multifaceted data handling system that included exploratory data analysis, deep learning techniques, and the development of a dynamic data visualization dashboard. The project effectively addressed common challenges in vibration-based predictive maintenance, such as data anomalies and computational constraints, by implementing sophisticated models and techniques like the sliding window and recurrent neural networks. The outcome is a robust system capable of continuous motor health monitoring, operational efficiency evaluation, and actionable insights for maintenance optimization. This initiative underscores the critical role of advanced analytics in predictive maintenance and sets a precedent for future enhancements in smart infrastructure management.

Table of Contents

1. Introduction

1.1 Background

The fusion of technology with industry has ushered us into the age of intelligent facilities, where integrated systems and smart components are the norms. In this landscape, motors are the unsung heroes, powering critical systems from manufacturing lines to water supply networks. Their uninterrupted operation is pivotal for the smooth execution of processes within these facilities. Within the realm of smart facility management, Vibration Analytics emerges as a crucial tool. This sophisticated technology not only ensures the peak performance of motors but also safeguards their longevity through continuous monitoring and data-driven insights.

1.2 Objective

Leveraging the advancements in smart technology, our project is committed to advancing the field of predictive maintenance with a focus on Vibration Analytics for motors in water supply systems. These motors are crucial for the seamless operation of urban infrastructure, analogous to the heart within the human body. With an understanding of the severe implications that motor failures can have, our approach involves the detailed analysis of a 10-month high-frequency vibration time series dataset. Our aim is to continuously monitor the health of these motors, evaluate their operational efficiency, and formulate strategies for optimization based on thorough vibration analysis. Embracing a proactive maintenance strategy will not only improve motor performance but also significantly reduce downtime and operational costs. The project seeks to provide actionable insights that lead to operational improvements and maintenance optimizations, ultimately prolonging the life of motors, enhancing system efficiency, and ensuring the reliability and sustainability of the water supply system, thereby guaranteeing continuous service and safeguarding public health.

Despite the widespread acknowledgment of the benefits of using vibration signals for monitoring the health of smart motor devices, several challenges persist. Firstly, the issue of missing values and records due to sensor anomalies poses a significant hurdle. Secondly, there's a delicate balance to be struck among computational costs, hysteresis effects, and the inherent complexity of the systems. Lastly, the collection of failure-state samples involves considerable expense, making it a formidable challenge to overcome. Our project intends to address these common difficulties by implementing innovative solutions and methodologies to improve the reliability and effectiveness of vibration-based predictive maintenance.

2. Literature Review

The seamless functioning of motors is pivotal in a plethora of industrial and commercial settings, where their failure can precipitate substantial operational downtime and consequent financial losses. Presently, three distinct maintenance strategies are prevalently adopted to manage motor health across its lifecycle: usage until failure, routine preventive maintenance, and condition-based maintenance (CBM). CBM relies heavily on the adept identification of motor faults to maintain optimal health (Ribeiro et al., 2021). A critical component in this detection mechanism is vibration analysis. This analytical approach, as supported by De Melo et al. (2022), involves scrutinizing the motor's vibration profile to discern various types of faults. This comprehensive literature review delves deeper into the domain of motor fault detection with a specific emphasis on the application of vibration analysis as an essential diagnostic instrument. Recent progress in this field has brought forth an amalgamation of intricate data mining methodologies, sophisticated computational models, and cutting-edge hardware. These advancements collectively work towards augmenting the accuracy and dependability of fault detection processes in motors, thereby enhancing the effectiveness of CBM strategies.

In the realm of motor operation, certain variables have proven to be particularly indicative of the motor’s condition. Ramtekkar et al. (2023) highlighted that the vibration signals measured along the X, Y, and Z axes (Axial, Radial, Tangential) can reveal significant insights about the motor's operational status. Consistently monitoring these vibrations in critical areas such as bearings, motor feet, and housing is crucial for preventing severe defects. Echoing this sentiment, Del Rosso et al. (2021) reinforce the importance of utilizing accelerometers in the axial and radial directions to capture comprehensive data about the motor's condition. They further advocate that statistical indexes like peak-to-peak and skewness, particularly in the axial direction, are effective in categorizing data into healthy and faulty clusters. In healthy motors, one would expect to observe symmetric probability density functions (PDFs) and minimal variance in the peak-to-peak values of the signal, providing a clear demarcation from faulty motors (Del Rosso et al., 2021). Such detailed analysis not only enhances the understanding of motor conditions but also paves the way for more precise maintenance interventions.

In the intricate process of vibration analytics, various classification models have been identified as effective tools in distinguishing between healthy and faulty motor conditions. Sharma & Jia (2021) emphasize the efficacy of models such as Support Vector Machine (SVM), Random Forest (RF), Sparse Representation Classifier (SRC), k-Nearest Neighbors (KNN), and Back Propagation Neural Network (BPNN) in this context. Notably, the RF model demonstrated superior performance in terms of accuracy, precision, and recall values in their research. De Melo et al. (2022) also validate the usefulness of SVM and KNN models in similar studies. Another notable approach is the Sliding Window Technique, recommended by Bagheri (2018 as cited in Henríquez et al., 2014), which is instrumental in classifying time series data. This technique aids in identifying recurring patterns within the dataset, thereby facilitating a deeper understanding of the data's structure. A critical aspect of these models is their reliance on labeled data, encompassing both healthy and faulty examples, to conduct effective supervised learning. However, a challenge often encountered in vibration analytics is the scarcity of faulty data, which can impede the learning process (Henríquez et al., 2014, as cited in Henríquez et al., 2014).

This literature review synthesizes key findings from various studies on motor fault detection using vibration analysis, underscoring the vital role of analytical models in diagnosing motor health. The integration of different variables, particularly vibration signals along multiple axes, offers a nuanced understanding of motor conditions (Ramtekkar et al., 2023; Del Rosso et al., 2021). Advanced classification models like SVM, RF, SRC, KNN, and BPNN have shown promising results in differentiating between healthy and faulty motors, with specific models exhibiting superior capabilities in certain aspects (Sharma & Jia, 2021; De Melo et al., 2022). The Sliding Window Technique further enriches the analysis by identifying patterns in time series data, enhancing the predictive accuracy (Bagheri, 2018 as cited in Henríquez et al., 2014). However, the efficacy of these models is often challenged by the limited availability of faulty data, emphasizing the need for comprehensive datasets for more robust machine learning outcomes (Henríquez et al., 2014, as cited in Henríquez et al., 2014). Overall, the advancements in vibration analytics and machine learning herald a new era of efficient and reliable motor fault detection, crucial for minimizing downtime and ensuring uninterrupted industrial and commercial operations.

3. Method

3.1 Proposed Method

To continuously monitor motor health, assess efficiency, and provide optimization suggestions and eventually enhance the water supply system's overall reliability and sustainability while reducing the cost of operation and maintenance, this project will develop a full data handling, analyzing and report system.

3.2 Data Source

The data set used in this project is collected at Tamar Central Government Office Sea Water Pump Room with smart sensor from 8th November,2022 to 11th August,2023. There are total of 7 set of pump and motor installed in the pump room. The project will be using data of Motor 2 as pilot to examinate the methodologies.

3.3 Data acquisition and processing approaches

In the data collection and preprocessing phase, the primary task involves obtaining and structuring a 10-month dataset of vibration time series data acquired from sensors installed on motors. To ensure data uniformity and accessibility, a data pipeline will be designed to systematically transform the raw sensor data into a consistent format, subsequently storing the transformed data in a database for efficient retrieval and analysis. Within this pipeline, data preprocessing tasks will be addressed, such as handling missing values and implementing noise reduction techniques, it is to enhance data quality and reliability, thus setting the stage for robust and accurate motor vibration analytics.

3.4 Data storage

Following the preprocessing phase, the data will be temporarily housed in a CSV file to streamline subsequent analyses. Upon the completion of model development, evaluation, and refinement, this processed dataset will be funneled into the models for predictive analytics and computations. The outcomes of these predictions and computations will then be securely archived in a MySQL database, primed for data visualization purposes. This database will adhere to the star schema configuration, as illustrated in Figure 1, ensuring an organized and efficient data structure for easy access and analysis.

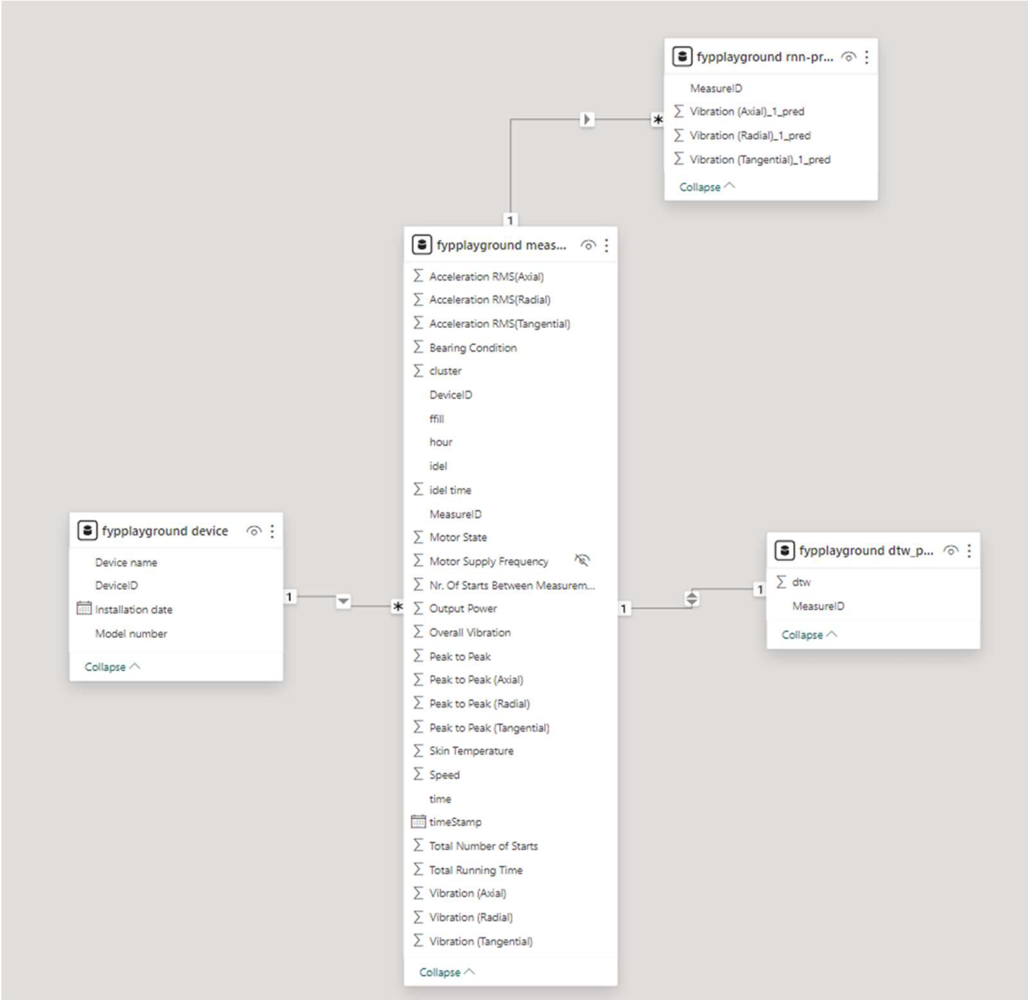

Figure 1: Database design

3.5 Data analysis approaches

Due to the complexity and the nature of the motor data set, it is difficult to construct a predictive model to directly predict the probability of fault. Therefore, other method must been taken. As the current data set of motor contain no data recorded during imnormal situation of motors. The usual approach of supervised learning on directly predicting fault could not be apply to the dataset. The methodology proposed will center on leveraging current operational data from a motor to develop a predictive model that functions as a classifier. This model is designed to assess the motor's operational status by analyzing real-time data against established patterns and thresholds derived from historical data. Through this approach, it aims to identify potential warning signs or confirm normal operation, enabling proactive maintenance and ensuring optimal performance of the motor.

- Exploratory Data Analysis (EDA)

In the realm of exploratory data analysis (EDA), an initial exploration of the dataset will be carried out to gather insights into its underlying characteristics. Through a combination of statistical and visual techniques, key data statistics, trends, and patterns will be vividly visualized and analyzed. This process will provide help in the identification of potential issues with motors, laying the foundation for more in-depth analysis and targeted actions to ensure motor health and performance optimization.

- Frequency domain analysis

In the context of frequency domain analysis, algorithms for transforming time domain data into the frequency domain will be investigated, for example, the Fast Fourier Transform. This enables the examination of the frequency spectrum to pinpoint prominent vibration frequencies and detect any potential anomalies in motor vibrations.

- Feature Extraction

Within the domain of feature extraction, pertinent attributes will be extracted from the frequency domain data to facilitate health assessment of motors. These attributes encompass characteristics such as amplitude, frequency components, and various statistical measures, all of which play an important role in the comprehensive analysis of motor vibrations for diagnostic purposes.

3.6 Peak-to-Peak Value Analysis

The Peak-to-Peak Value Analysis involves examining the range between the maximum and minimum values in the current data set and comparing these to the corresponding range in historical data. Minimal variance in the peak-to-peak values of the signal, providing a clear demarcation from faulty motors (Del Rosso et al., 2021). By setting thresholds based on historical min-max ranges, this method allows for the classification of new data points. If a new data point's peak-to-peak value lies outside the historical range, it is classified as a 'warning', indicating a potential anomaly or deviation from normal operation. Otherwise, it is classified as 'normal', suggesting that the motor is operating within expected parameters.

3.7 Sliding Window Technique

The Sliding Window Technique is employed to measure the similarity between the current operational data of the motor and its historical data using Dynamic Time Warping (DTW). This technique involves moving a 'window' across the data series to compare different segments over time, focusing on the alignment and distance between these segments. If the distance calculated for a new record exceeds the maximum distance observed in the historical data, the motor’s status is classified as a 'warning'. This method is particularly effective in identifying significant deviations over time, highlighting potential issues that may not be immediately apparent.

3.8 Health Assessment Model

Multiple machine learning or statistical models will be used to construct a health assessment model to perform health assessments on motors. This model will incorporate techniques including anomaly detection to effectively evaluate the condition of the motors. By utilizing this model, potential issues can be identified and flagged, contributing to the overall health assessment and maintenance of the motors. Recurrent neural network

3.9 Deep learning

Incorporating Recurrent Neural Networks (RNNs) into deep learning strategy leverages their potent capability to process sequential data, making them ideal for analyzing the intricate time-series data generated by motor vibrations. The core strength of RNNs lies in their architecture, which allows for the propagation of information across time steps, enabling the network to make informed predictions based on historical data. This project will harness the power of RNNs to construct a dynamic model that continually adapts to new data, enhancing its predictive accuracy over time. By systematically identifying patterns and anomalies in vibration data, the RNN model serves as a critical tool in the preemptive identification of potential faults, thereby facilitating timely interventions and maintenance actions.

3.10 Data visualization

- Dashboard Development

A web-based dashboard will be designed to provide a comprehensive overview of motor health status in real-time or near-real-time. This dynamic dashboard will present visualizations of crucial metrics, allowing users to monitor motor performance efficiently. It will also feature an alert system that promptly notifies users of any detected anomalies, ensuring timely attention to potential issues and facilitating informed decision-making for motor maintenance and optimization.

4. Project Scope

4.1 Literature Review

Through the review of the literature, academic publications concerning vibration analytics on motors will be investigated. The primary objectives are to identify potential determinizing features and examine the methodologies employed in the realm of vibration analytics on motors. During the thorough examination of relevant literature, insights into the key factors influencing motor vibrations will be gained and effective approaches for conducting vibration analysis on motors will be discerned.

4.2 Documentation and Reporting

Throughout the whole project, comprehensive documentation will be maintained, including data preprocessing steps, analysis techniques, model development, and dashboard design. Project's findings, outcomes, insights and recommendations will be summarized in the final report.

4.3 Dashboard Design

The dashboard will be user-centric, providing clear visualizations of motor performance data like speed, operational duration, and efficiency metrics such as missing records and idling percentages. It will enable users to select specific or multiple days for data analysis, enhancing the understanding of motor conditions. Predictive model outcomes will be integrated, offering straightforward interpretations of motor health predictions. This design aims to facilitate informed decision-making by making complex data accessible and actionable.

4.4 Decision support model

Decision support models will incorporate a robust blend of statistical and mathematical methodologies, alongside advanced deep learning techniques, to analyze and interpret motor data comprehensively. These models will be meticulously optimized for individual devices, ensuring a tailored approach to assessing each motor's health state. This customization will significantly enhance the precision and relevance of the insights provided, enabling more accurate predictions of device performance and potential maintenance needs. The integration of diverse analytical techniques ensures a comprehensive evaluation, elevating the overall effectiveness of the decision support framework in guiding maintenance strategies and operational optimizations.

5. Results

5.1 Data analysis

- Understanding the data and Data Cleaning

There are total of 7 set of motor data and 7 set of pump data through the time period of 8th November,2022 to 11th August,2023. The project will be first focusing on motor 2 dataset. After primary transforming the data into consistent format, there are total of 12196 rows of data with 19 columns. The data type and the non-null values of each column are listed in Figure 2. The percentage of missing value is calculated in table 2.

```
#   Column                               Non-Null Count  Dtype
--  ------                               --------------  -----
0   Vibration (Axial)                    6251 non-null   float64
1   Vibration (Radial)                   6251 non-null   float64
2   Vibration (Tangential)               6251 non-null   float64
3   Bearing Condition                    2473 non-null   float64
4   Output Power                         6237 non-null   float64
5   Motor Supply Frequency               6251 non-null   float64
6   Nr. Of Starts Between Measurements   6251 non-null   float64
7   Total Running Time                   6251 non-null   float64
8   Overall Vibration                    6251 non-null   float64
9   Peak to Peak                         6110 non-null   float64
10  Peak to Peak (Axial)                 6251 non-null   float64
11  Peak to Peak (Radial)                6251 non-null   float64
12  Peak to Peak (Tangential)            6251 non-null   float64
13  Skin Temperature                     6251 non-null   float64
14  Speed                                6251 non-null   float64
15  Total Number of Starts               6251 non-null   float64
16  Acceleration RMS(Axial)              6251 non-null   float64
17  Acceleration RMS(Radial)             6251 non-null   float64
18  Acceleration RMS(Tangential)         6251 non-null   float64
```

Figure 2: Data information of moto2 dataset

	Number of nan row	Percentage
Bearing Condition	9723	79.7%
Peak to Peak	6086	49.9%
Output Power	5959	48.9%
Peak to Peak (Radial)	5945	48.7%
Acceleration RMS (Radial)	5945	48.7%
Acceleration RMS (Axial)	5945	48.7%
Total Number of Starts	5945	48.7%
Speed	5945	48.7%
Skin Temperature	5945	48.7%
Peak to Peak (Tangential)	5945	48.7%
Vibration (Axial)	5945	48.7%
Peak to Peak (Axial)	5945	48.7%
Vibration (Radial)	5945	48.7%
Overall Vibration	5945	48.7%
Total Running Time	5945	48.7%

Number of Starts Between Measurements	5945	48.7%
Motor Supply Frequency	5945	48.7%
Vibration (Tangential)	5945	48.7%
Acceleration RMS (Tangential)	5945	48.7%

Table 2: Number of nan row and percentage in motor 2 data frame

From the table, most of the columns contains 48.9% of missing value, Peak to Peak column contain 49.9% of missing value, and Bearing Condition contain the most missing value, high up to 79.7% of missing value. It is unusual to such large amounts of missing value. To gain deeper insights into the situation, a scatter plot was generated using the Overall Vibration and Speed columns. Notably, the first occurrence of Not a Number (NaN) values in both columns was observed on 2023-02-28 at 14:43:00. Subsequently, a distinct pattern emerged, with one NaN record appearing every two hours. This pattern is visually represented in Figure 3.

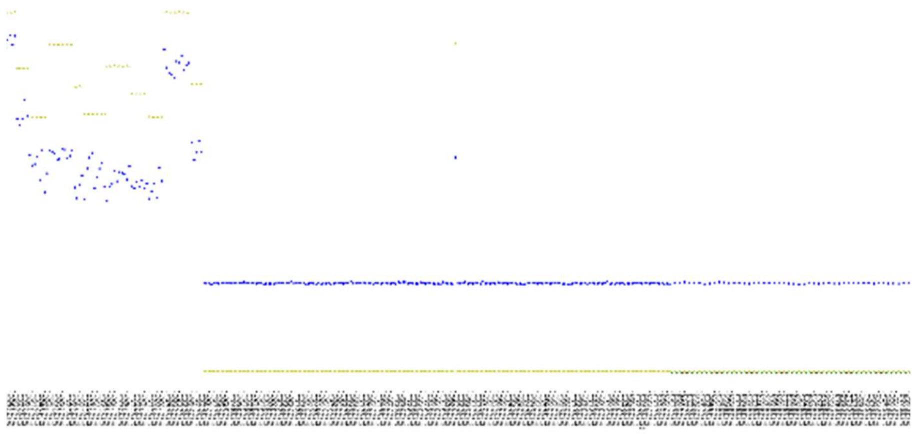

Figure 3: Scatter Plot on Overall Vibration and Speed

Such pattern is repeated through the rest of the dataset, it might be caused by the sensor setting error or related networking issue. However, it is not the key areas of the studies. To further transform the data frame to better fit the use of the project. The data had been filtered by the Speed column by only keeping the row which the motor was in active state with condition of Speed value over 0. The filter further reduces the size of the data set to 2651 rows which is 21.7% of the whole dataset. However, performing such filtering might also remove the missing rows between records. It might cause the time series became noncontinuous. Therefore, further action must be taken to reconstruct the time series. Missing rows between record had been search and reinsert into the data set. The data type and the non-null values of each columns are listed in Figure 4. The percentage of missing value is calculated in table 3.

```
 #   Column                              Non-Null Count  Dtype
---  ------                              --------------  -----
 0   Vibration (Axial)                   2651 non-null   float64
 1   Vibration (Radial)                  2651 non-null   float64
 2   Vibration (Tangential)              2651 non-null   float64
 3   Bearing Condition                   2472 non-null   float64
 4   Output Power                        2637 non-null   float64
 5   Motor Supply Frequency              2651 non-null   float64
 6   Nr. Of Starts Between Measurements  2651 non-null   float64
 7   Total Running Time                  2651 non-null   float64
 8   Overall Vibration                   2651 non-null   float64
 9   Peak to Peak                        2590 non-null   float64
 10  Peak to Peak (Axial)                2651 non-null   float64
 11  Peak to Peak (Radial)               2651 non-null   float64
 12  Peak to Peak (Tangential)           2651 non-null   float64
 13  Skin Temperature                    2651 non-null   float64
 14  Speed                               2651 non-null   float64
 15  Total Number of Starts              2651 non-null   float64
 16  Acceleration RMS(Axial)             2651 non-null   float64
 17  Acceleration RMS(Radial)            2651 non-null   float64
 18  Acceleration RMS(Tangential)        2651 non-null   float64
 19  Motor State                         2651 non-null   object
 20  idel time                           2651 non-null   float64
 21  time                                2767 non-null   datetime64[ns]
dtypes: datetime64[ns](1), float64(20), object(1)
```

Figure 4: Data information of filtered moto2 dataset

	Number of nan row	Percentage
Bearing Condition	295	2.42%
Peak to Peak	177	1.45%
Output Power	130	1.07%
Vibration (Axial)	116	0.95%
Skin Temperature	116	0.95%
Ideal time	116	0.95%
Motor State	116	0.95%
Acceleration RMS (Tangential)	116	0.95%
Acceleration RMS (Radial)	116	0.95%
Acceleration RMS (Axial)	116	0.95%
Total Number of Starts	116	0.95%
Speed	116	0.95%
Peak to Peak (Radial)	116	0.95%
Peak to Peak (Tangential)	116	0.95%
Vibration (Radial)	116	0.95%
Peak to Peak (Axial)	116	0.95%
Overall Vibration	116	0.95%
Total Running Time	116	0.95%
No. of Starts Between Measurements	116	0.95%
Motor Supply Frequency	116	0.95%
Vibration (Tangential)	116	0.95%

Table 3: Number of nan row and percentage in filtered motor 2 data frame

From the table, the column with the most missing value is the Bearing Condition of 2.42% missing, followed by Peak-to-Peak, Output Power and another column. None of the column had more than 2.5% of missing values. Forward filling was being applied to handle the missing values.

- Exploratory Data Analysis (EDA)

Bases on the finding during literature review, key variables of analytics, vibration signal and accelerometers in the axial and radial could be helpful in identifying motor's operational status (Del Rosso et al. 2021). The key variables had been selected out form the dataset alone with Speed and Output power. Figure 5 is a correlation heatmap that represents the relationship between different measured variables in a motor vibration dataset. The color scale on the right indicates the correlation coefficient values, where 1 is a perfect positive correlation (indicated by dark red), 0 represents no correlation (white), and -1 indicates a perfect negative correlation (dark blue). Each cell in the heatmap shows the correlation coefficient between the variables at the intersection of the corresponding row and column. High positive values suggest that as one variable increases, the other tends to increase as well, while high negative values indicate an inverse relationship.

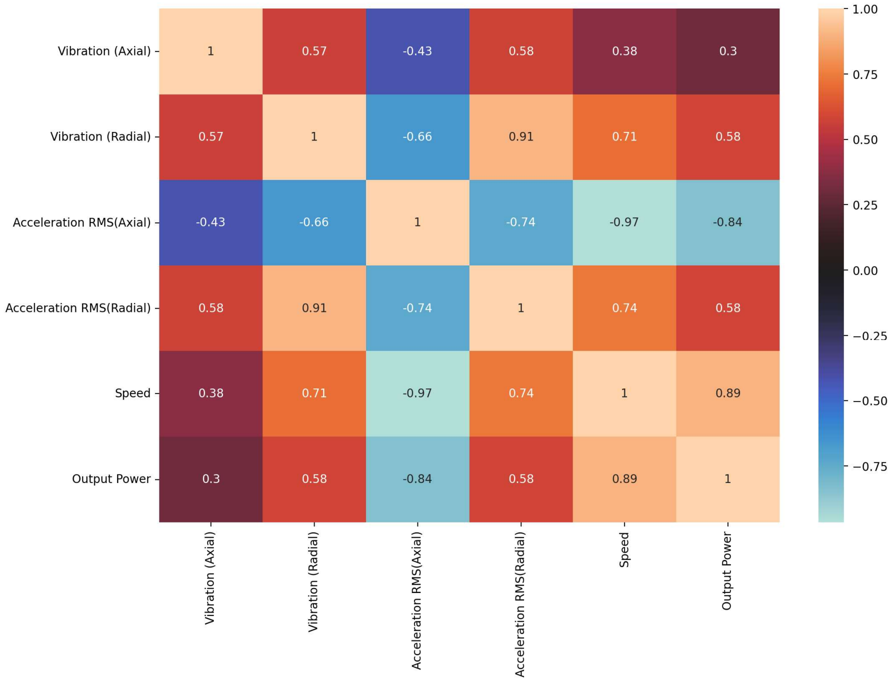

Figure 4: Heatmap of correlation of selected variables

From Figure 5, a strong negative correlation between Acceleration RMS (Axial) and Speed of the correlation can be observed as -0.97. It means when the speed is higher, the axial acceleration RMS will be lower. Figure 6 support such findings. From figure 5, there is a strong correlation between Acceleration RMS (Radial) and Vibration (Radial) with correlation as 0.91. It means when the Acceleration RMS (Radial) is higher, the radial vibration will be higher. Figure 7 support such findings.

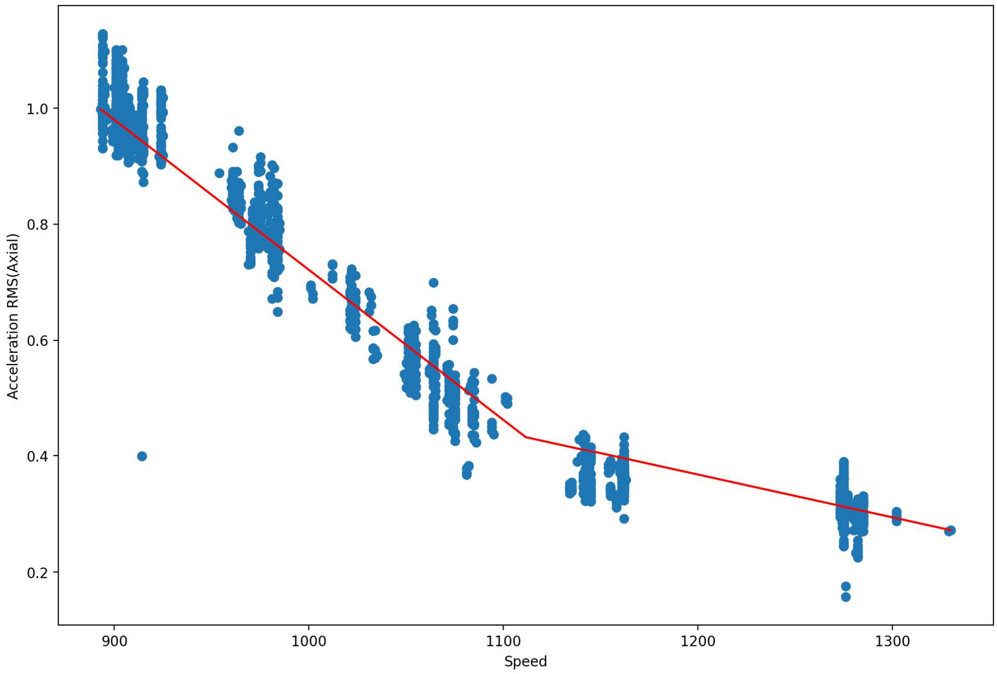

Figure 6: Scatter plot of between 'Acceleration RMS(Axial)' and 'Speed'

Figure 6 is a scatter plot showing the relationship between 'Acceleration RMS (Axial)' and 'Speed'. Each dot represents an observation from the dataset. The general trend is indicated by a fitted line, which appears to show a negative relationship, as speed increases, the RMS of axial acceleration decreases. The density of the points varies, but they appear to follow the trend line closely, suggesting a strong correlation between these two variables.

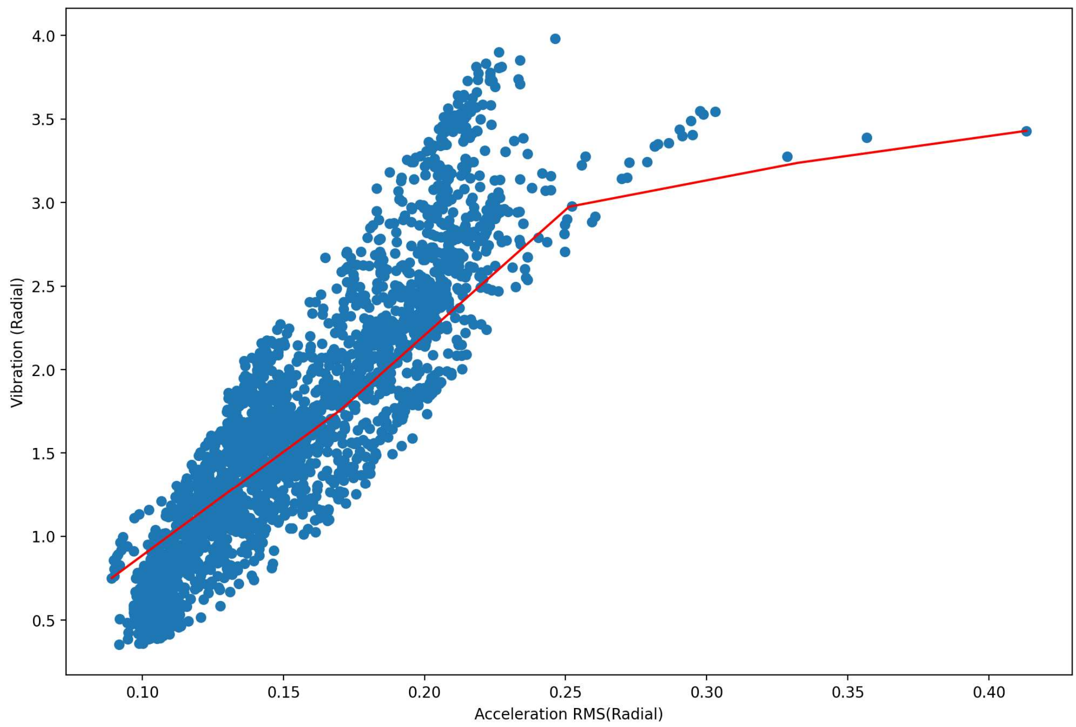

Figure 7: Scatter plot of between 'Acceleration RMS(Radial)' and ‘Vibration (Radial)’

Figure 7 is a scatter plot visualizing the relationship between 'Vibration (Radial)' and 'Acceleration RMS (Radial)'. The data points are plotted as blue dots, and a red fitted line indicates the trend in the dataset. It appears there is a positive relationship between the two variables; as the RMS of radial acceleration increases, the radial vibration also tends to increase. The distribution of data points forms a distinct elongated cluster, showing a concentration of observations along the fitted line, which suggests a strong linear correlation between these variables. The increasing trend is consistent throughout the range of the data, which might indicate that radial acceleration is a good predictor of radial vibration in this context.

Among the selected columns of ‘Vibration (Axial)’, ‘Vibration (Radial)’, ‘Acceleration RMS(Axial)’, ‘Acceleration RMS(Radial)’, ‘Speed’, ‘Output Power’, correlation consisted between columns. Variables ‘Acceleration RMS(Axial)’ and ‘Speed’ appeared to be the strongest negative correlation with the value of -0.97. The moving trend between two variables appear to be opposite. Variables 'Acceleration RMS(Radial)' and ‘Vibration (Radial)’ appear to be the strongest positive correlation with the value of 0.91. The moving trend of two variables appear to be at same direction. Knowing the correlation between variables, it identifies the proper predictor of different variables. By knowing the relationship between variables, predicted model could be constructed between variables. For example, ‘Speed’ might be useful to predict the value of ‘Acceleration RMS(Axial)’.

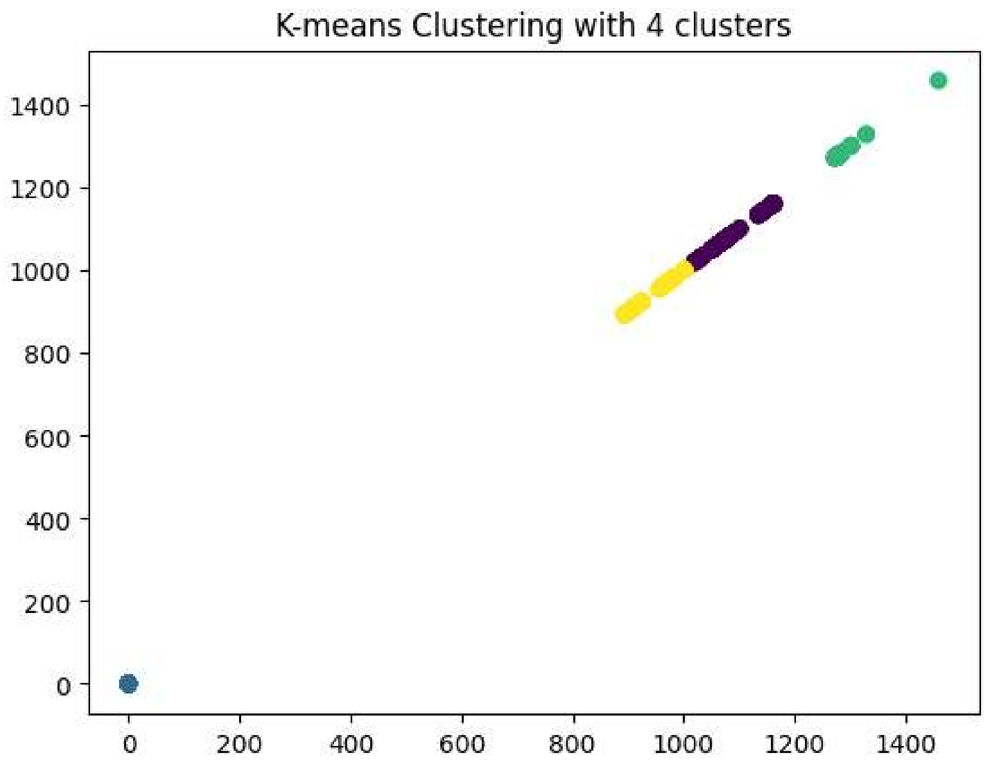

Figure 8: Scatter plot of clustering with 4 cluster

To gain insights into the operational dynamics of the machinery, categorizing the operating speeds into distinct segments is imperative. Engaging the expertise of seasoned technical and operational personnel from the facilities could provide a pragmatic framework for determining the optimal number of categories. Nonetheless, leveraging a machine learning strategy presents a viable alternative. The application of k-means clustering, experimenting with k values ranging from 1 to 6 and evaluating metrics such as the silhouette score, inertia, and the Davies Bouldin score, facilitates the discernment of the most appropriate cluster count. As depicted in Figure 8, the analysis delineates four as the optimal cluster count for Motor 2's data. In alignment with the clustering outcomes, operating speeds below 893 should be classified under cluster 0; speeds ranging from 894 to 1020 under cluster 1; those from 1021 to 1271 under cluster 2; and speeds exceeding 1271 under cluster 3. These clusters are indicative of the operational modes: off, low, medium, and high speed, respectively.

5.2 Peak-to-Peak Value Analysis

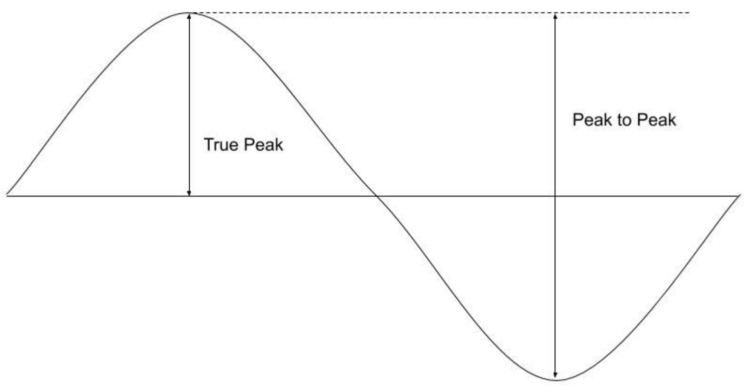

Figure 8: Idealized Waveform

Figure 9 represents an idealized waveform, commonly found in theoretical or simplified representations of vibrational signals. The term "Peak-to-Peak" (P-P) refers to the difference between the maximum positive amplitude and the maximum negative amplitude of a signal (Chandel, 2023).

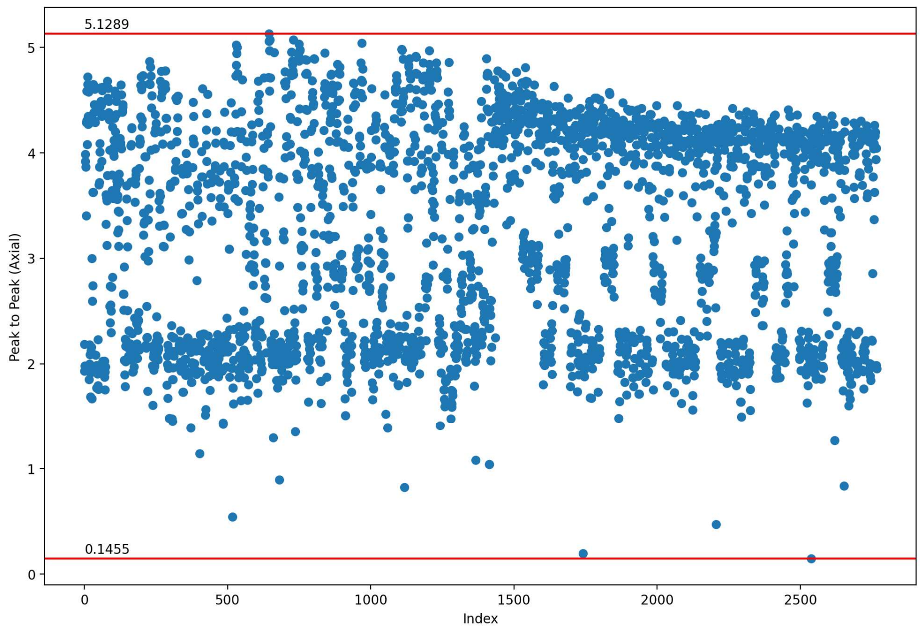

Figure 10: Scatter Plot of Peak-to-Peak (Axial)

Figure 10 is a scatter plot displaying a series of peak-to-peak measurements at hour level from an axial vibrational signal. The Y axis represents the peak-to-peak values of the vibrational signal for each corresponding index. A clear distinction is made between the highest recorded value, noted at approximately 5.1289, and the lowest, at around 0.1455. These values are marked by horizontal lines.

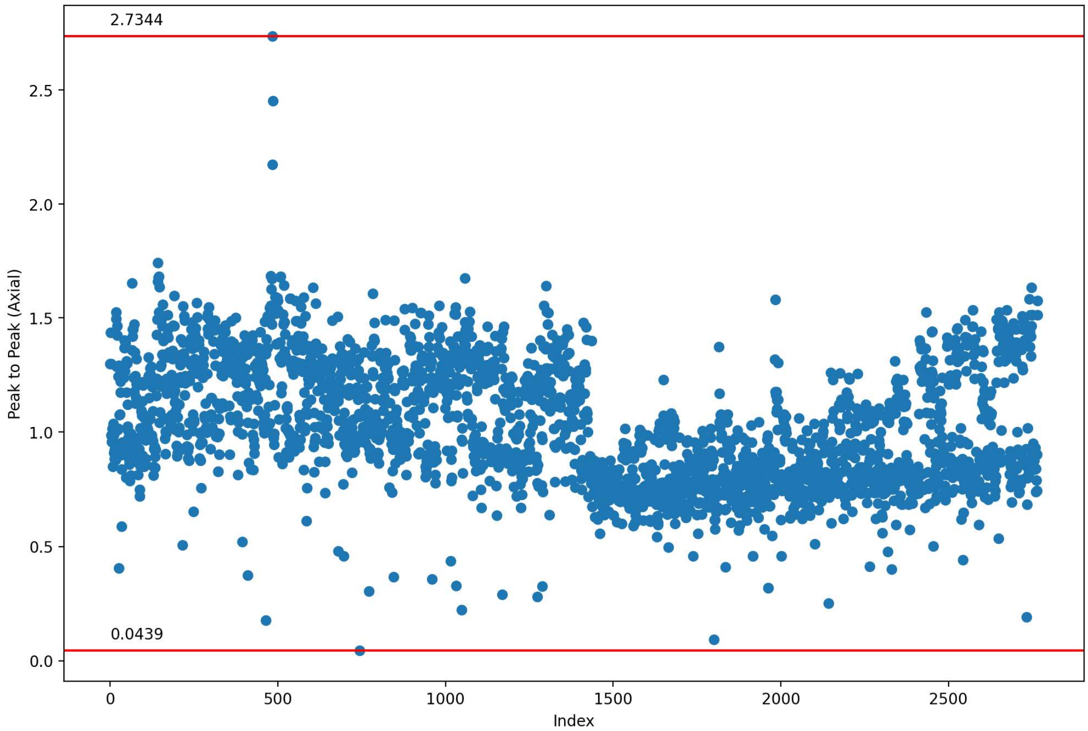

Figure 11: Scatter Plot of Peak-to-Peak (Radial)

Figure 11 is a scatter plot that showcases a collection of peak-to-peak values of radial vibrational signa. The Y axis corresponds to the signal's peak-to-peak values at each index. In this plot, the peak values are marked by horizontal lines, with the upper limit at approximately 2.7344 and the lower boundary at roughly 0.0439.

Peak-to-Peak value analytics could be used for setting an acceptable range of vibration value during the operation of the motor. If a peak-to-peak signal was being detected during operation, it could be label as warning to notify the technicians. By construction such decision rule, it could help to identify abnormal operation efficiently to better perform CBM on the motor.

5.3 Sliding Window Technique

The Sliding Window Technique, as recommended by Bagheri (2018) and referenced in the work of Henríquez et al. (2014), plays a pivotal role in the classification of time series data. This project will utilize a method known as Dynamic Time Warping (DTW) to assess the similarities between a target pattern and corresponding patterns from historical data of equal length. This process involves systematically comparing the target pattern against all possible matching-length patterns in the historical dataset, calculating the DTW distance for each comparison.

A practical example of this technique is illustrated in Figure 12, where the window comprising rows 1 to 13 of the dataset serves as the target pattern. The algorithm searches the entire dataset to find the most similar pattern, which, in this instance, is identified as rows 675 to 687. The calculated DTW distance between these two windows is 0.10225, indicating their degree of similarity.

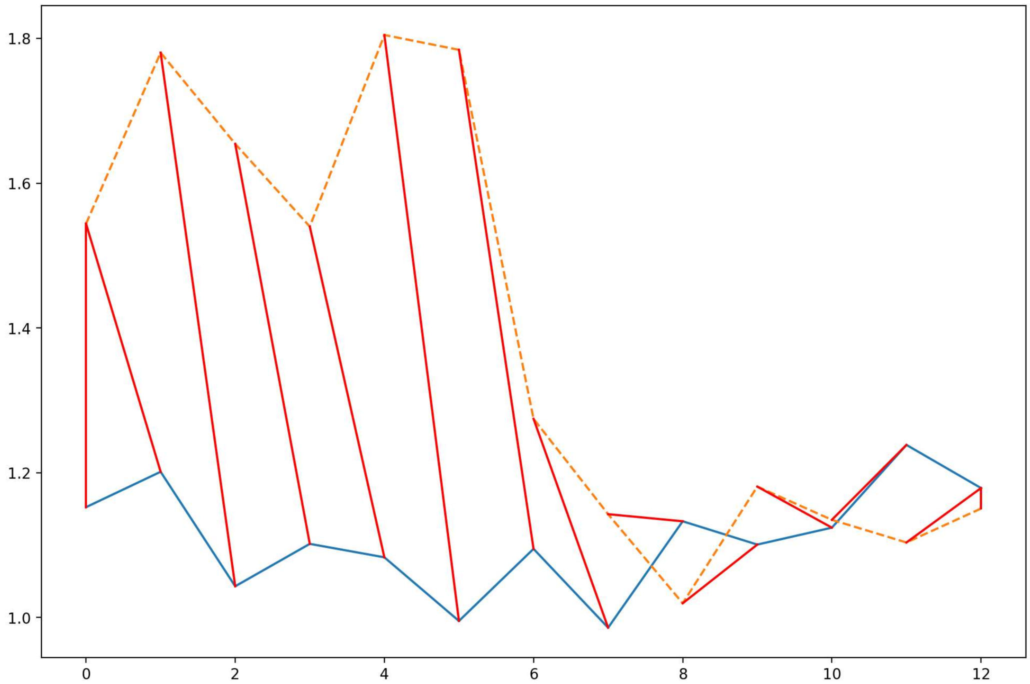

Figure 12: DTW between Rows 1 to 13 and Rows 675 to 687

This method is particularly useful for identifying recurring patterns within the dataset, which can be instrumental for technicians in spotting unusual vibration patterns. Given the absence of fault data for the motor, the emergence of an unexpected vibration pattern could signal a potential fault. However, the effectiveness of the sliding window approach hinges on determining an optimized threshold value to enhance accuracy. The determination of this threshold value is currently in progress, underscoring the ongoing refinement of this technique to achieve precise and reliable results.

5.4 Deep learning

Data selection prioritized active states, specifically those instances where the speed exceeded zero. The dataset was subsequently divided using a train-test split, allocating 80% of days for training purposes and 20% of days for testing.

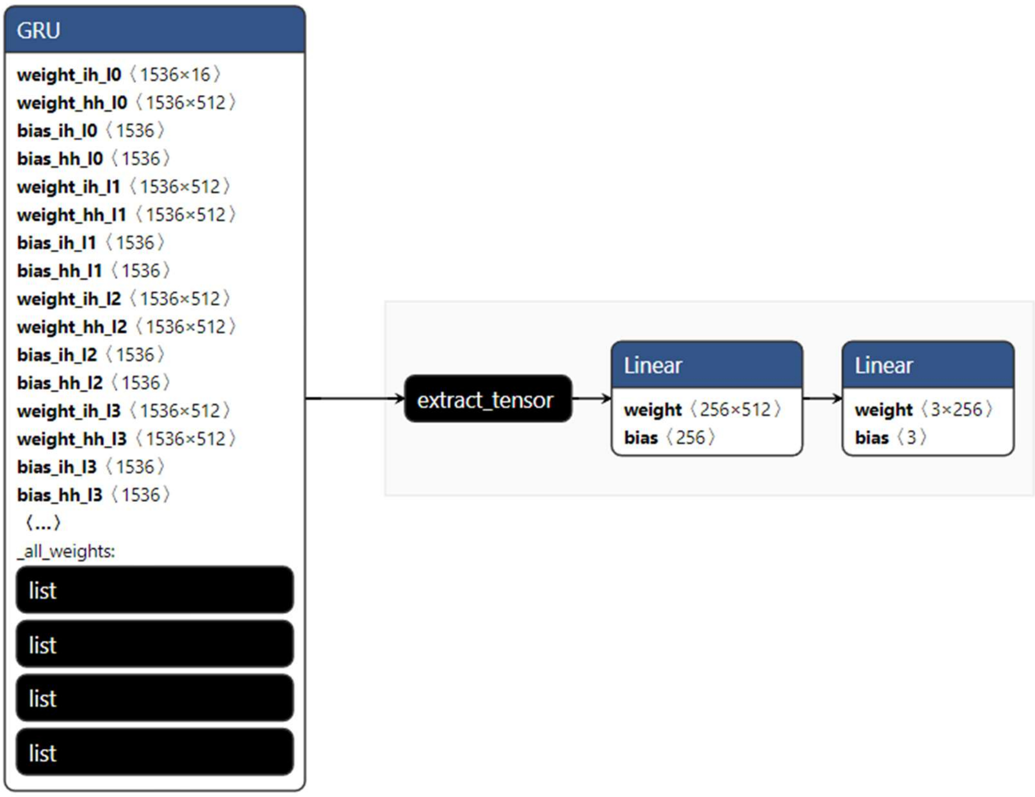

Figure 13: RNN model visualization

The model employed a Gated Recurrent Unit (GRU) architecture, this recurrent neural network model is adept at processing sequential data to estimate triaxial vibration values—axial, radial, and tangential. Configured to process inputs with 16 features, the model generates a 512-dimensional hidden state output from the initial GRU layer. To bolster the model's proficiency in pattern recognition across sequential data, the GRU layer is replicated fourfold, with a dropout rate of 0.1 applied to all but the final layer to minimize overfitting by selectively deactivating 10% of the outputs.

The sequence data, once processed through GRU layers, is funneled through a extra tensor module designed to extract the final time step's hidden state. This conversion from a sequence to a fixed-size vector ensures that the model's focus is trained on the latest and often most critical information in the sequence.

Subsequent to the extraction process, the model's architecture incorporates two fully connected linear layers. These sequentially reduce the hidden state dimensions from 512 to 256 and finally to a trio of output values.

For assessing model accuracy, the Mean Squared Error Loss function is utilized, evaluating the average of the squared differences between predicted and actual values. The Adam optimizer, with its learning rate set at 0.001, dynamically adjusts learning rates for each parameter based on the gradient's first and second moments.

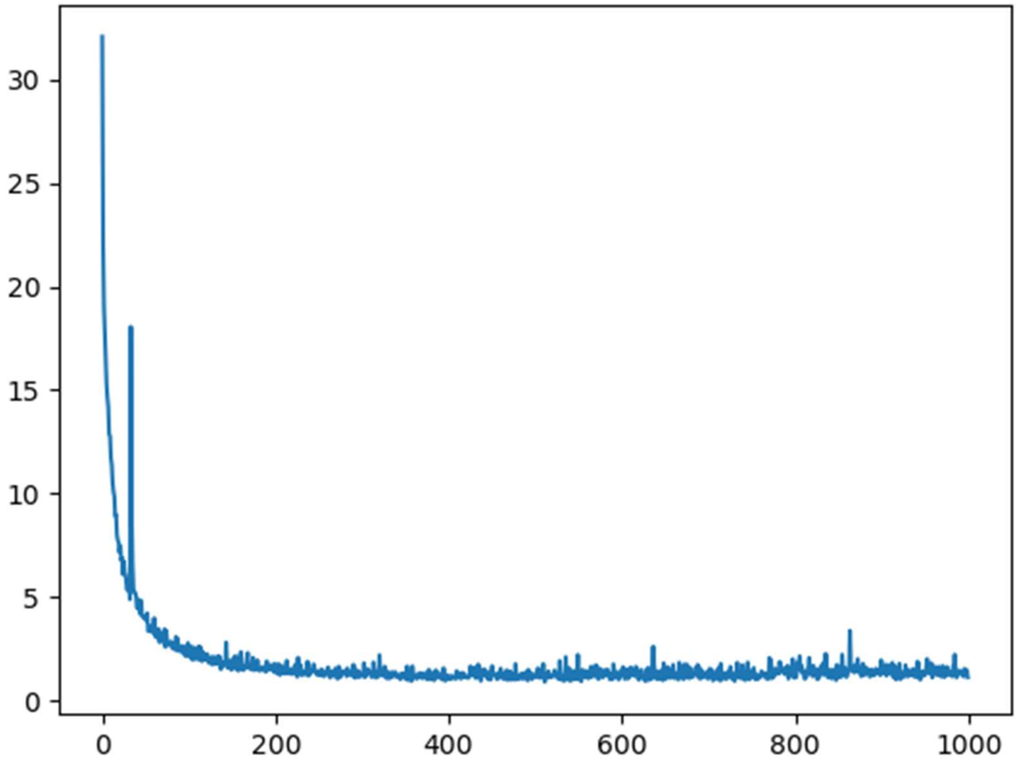

Figure 14: Loss curve of RNN model after 1000 epochs

Figure 14 delineates the model's loss curve across the initial 1000 epochs. Initially, the model's loss was significantly high but quickly descended. Post-100 epochs, the loss plateaued in the single digits, eventually stabilizing at approximately 0.9 after 1000 epochs.

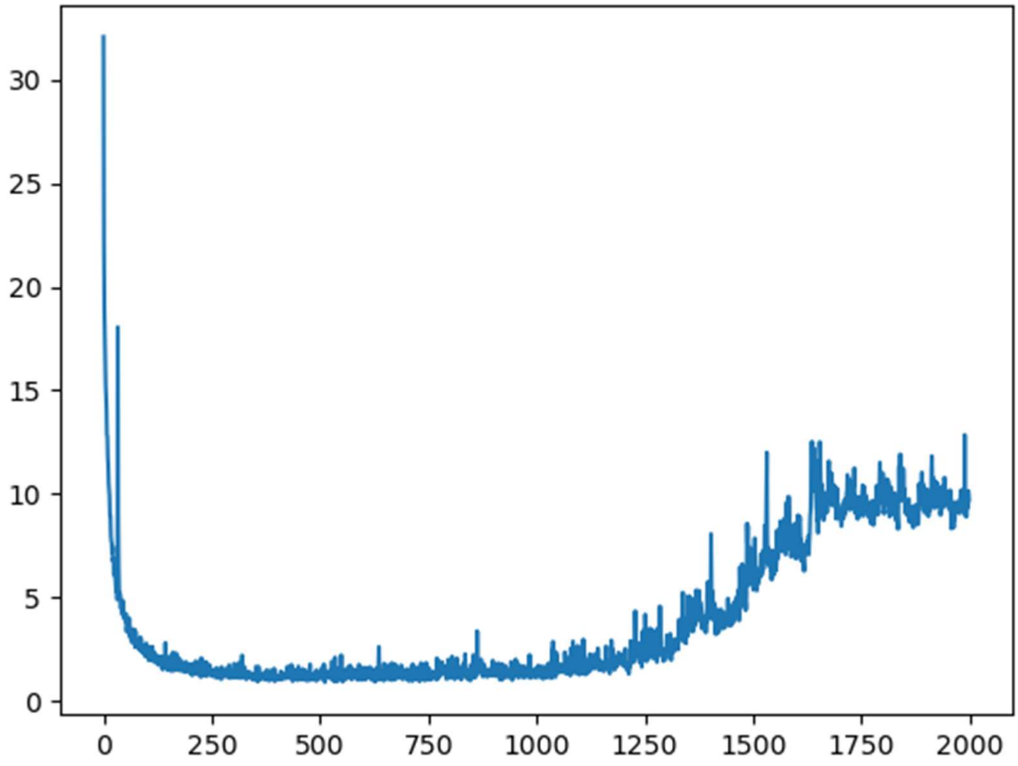

Figure 15: Loss curve of RNN model after 2000 epochs

Extending the training epochs beyond the initial 1000 brought an unexpected turn, asindicated in Figure 15. The loss which had previously stabilized, started to climb. It revealed signs of overfitting. This was evident as the model began to learn not only the pattern but also the noise within the dataset.

To combat this, a checkpointing system was implemented, saving the model only when it achieved a lower running loss. This strategy was proven effective, as shown in the loss curve.

Ultimately, the model's final loss was recorded at 0.901, which is considered to be within a satisfactory range. This performance benchmark suggests that the model possesses a reliable predictive ability. It can estimate the vibrational values in three dimensions based on the given features, demonstrating its practical utility in real-world applications.

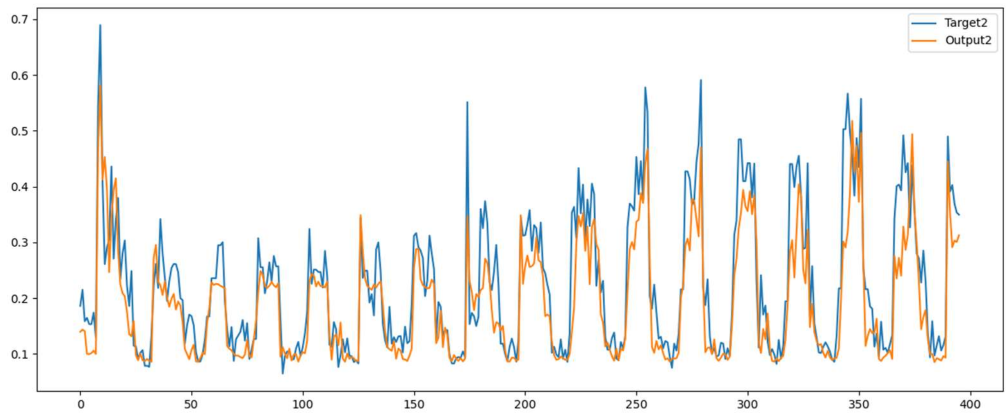

Figure 16a: Prediction and Ground true of radial vibration

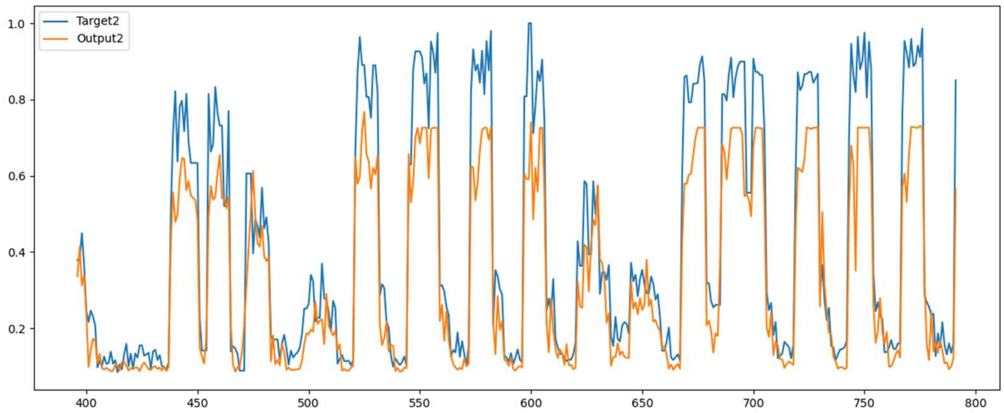

Figure 16b: Prediction and Ground true of radial vibration

Figure 16a and 16b plot the prediction of radial vibration from the model and the ground truth of the testing days. The blue line representing the prediction vlalue from the RNN model and the orange line representing the ground true from the dataset.

In figure 16a, the radial vibration signal ranging from approximately 0.1 to 0.7. It record the vlaue between the first sequence to 400 seqquencecs. The two lines follow a similar pattern, indicating that the model's predictions closely align with the actual target values, although there are some discrepancies where the orange line deviates from the blue line.

In figure 16b, ther radial vibration signal ranging from approximately 0.1 to 1, indicating higher vibration values than in the figure 14a. It record the vlaue between the 401 sequence to 800 seqquencecs. The model's output largely follows the trend of the target values, capturing the general pattern quite well.

5.5 Dashboard Design

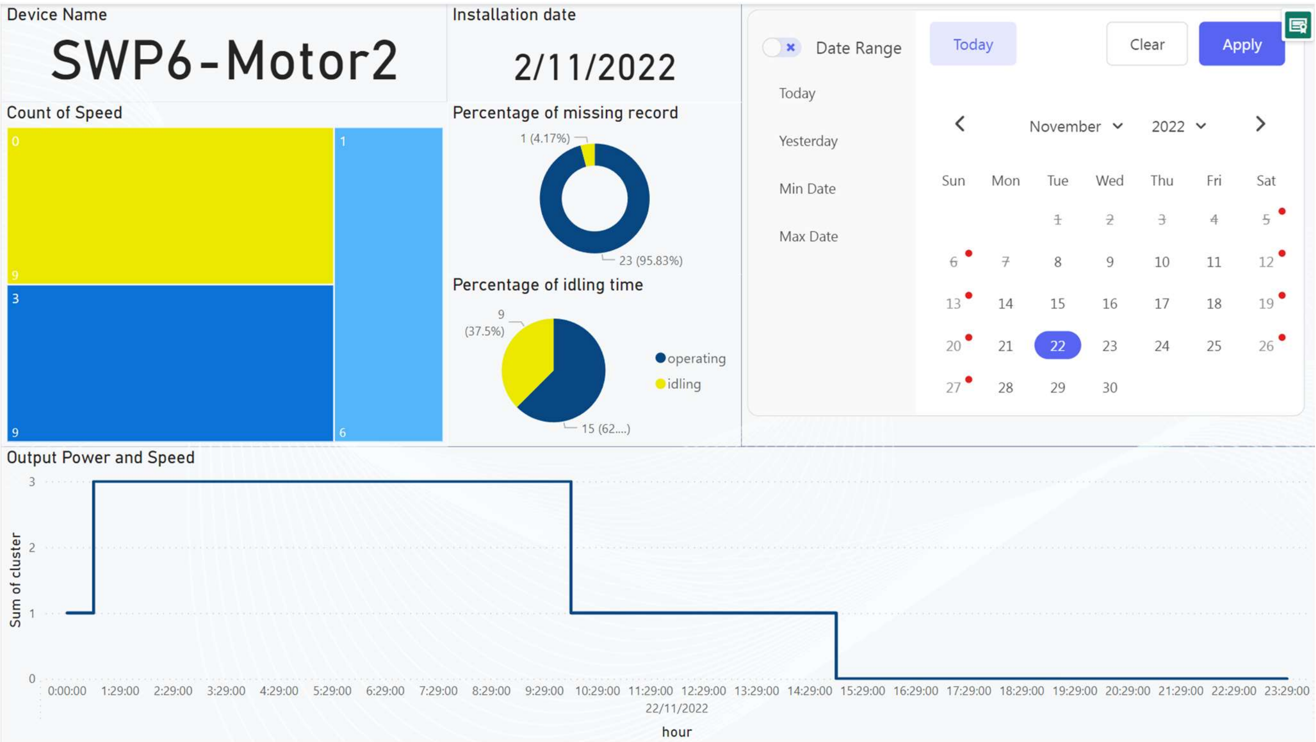

Figure 17: Dashboard Page1

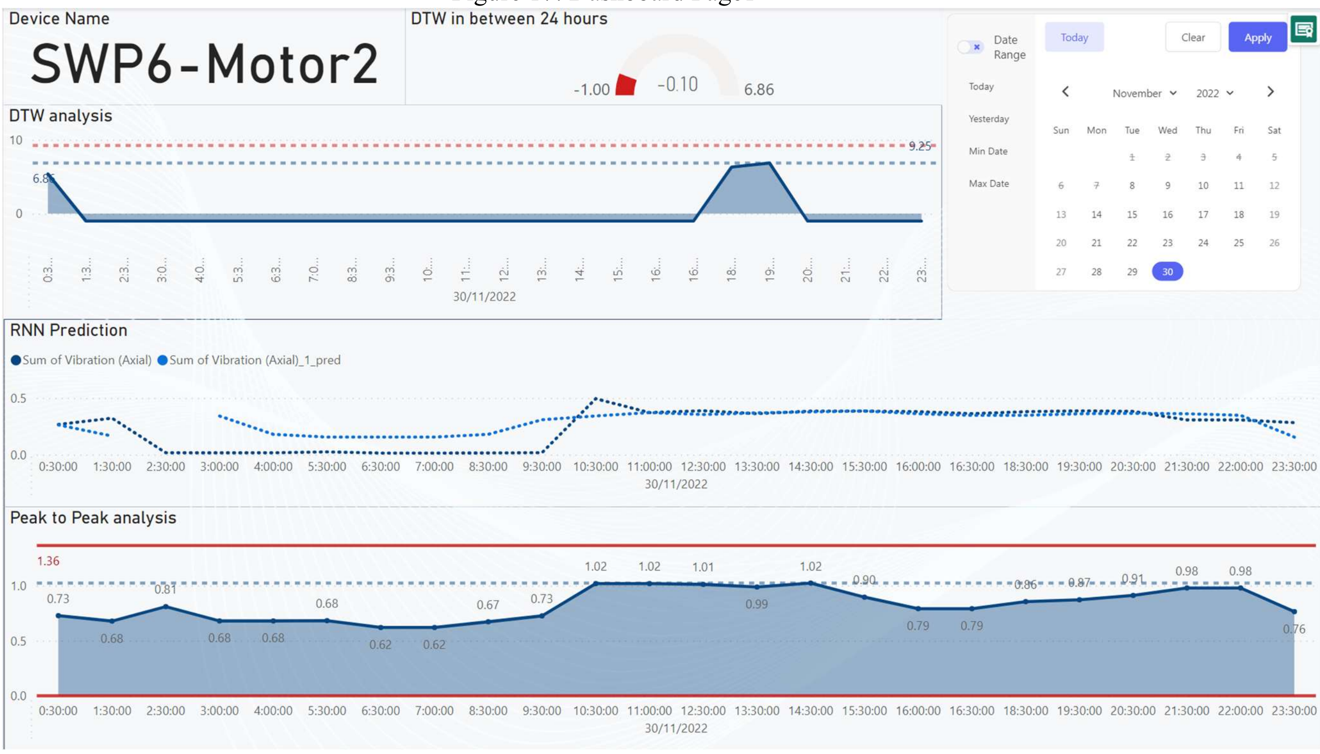

Figure 18: Dashboard Page2

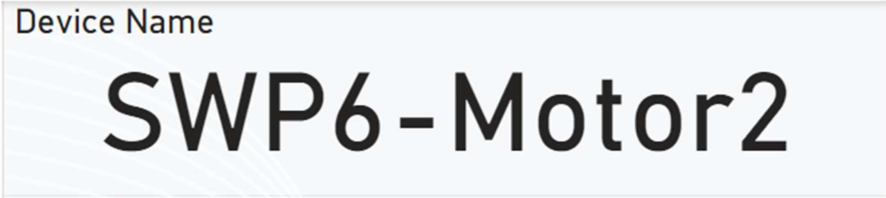

Figure 17a: Dashboard (Device name)

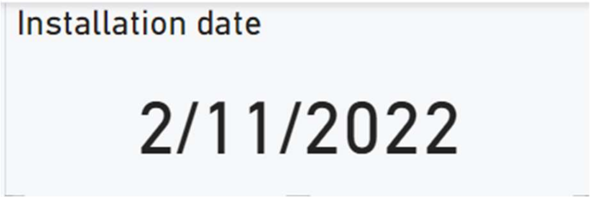

Figure 17b: Dashboard (Device installation date)

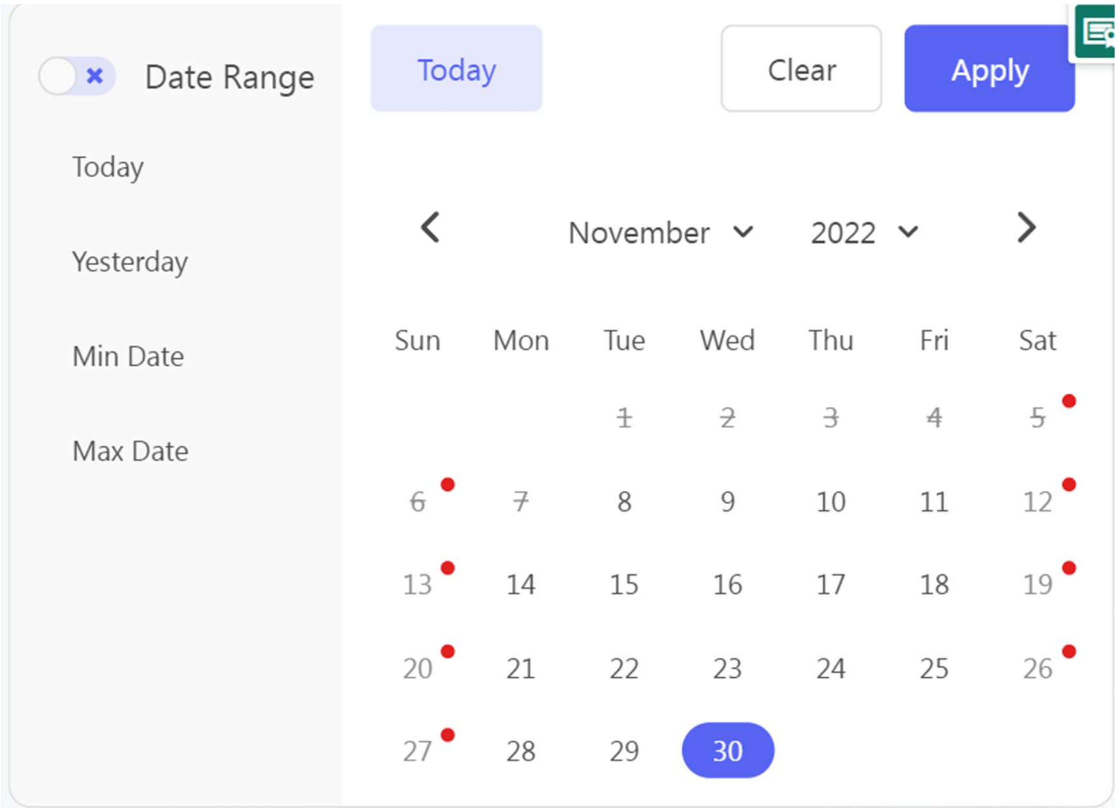

Figure 17c: Dashboard (calendar slicer)

The calendar slicer featured in Figure 17c enhances user interaction by enabling the selection of specific dates or date ranges for display on the dashboard. This slicer dynamically adjusts the data presented in all charts across pages 1 and 2, ensuring a tailored viewing experience.

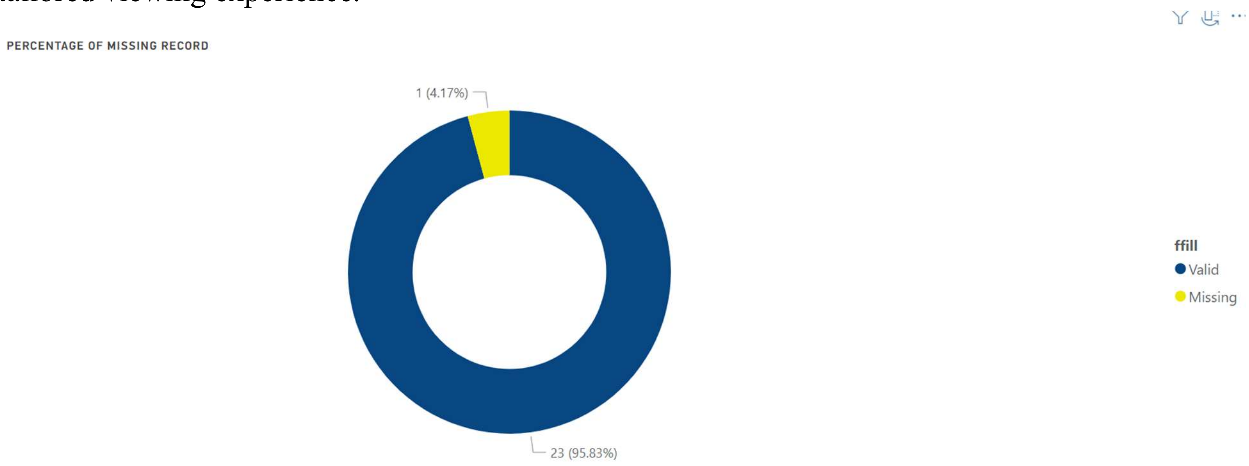

Figure 17d: Dashboard (Donut chart showing percentage of Missing Record)

In Figure 17d, a donut chart vividly illustrates the proportion of missing versus valid records for a chosen timeframe. The chart uses yellow to denote missing records and blue for valid ones, with detailed percentages and record counts for each category clearly displayed.

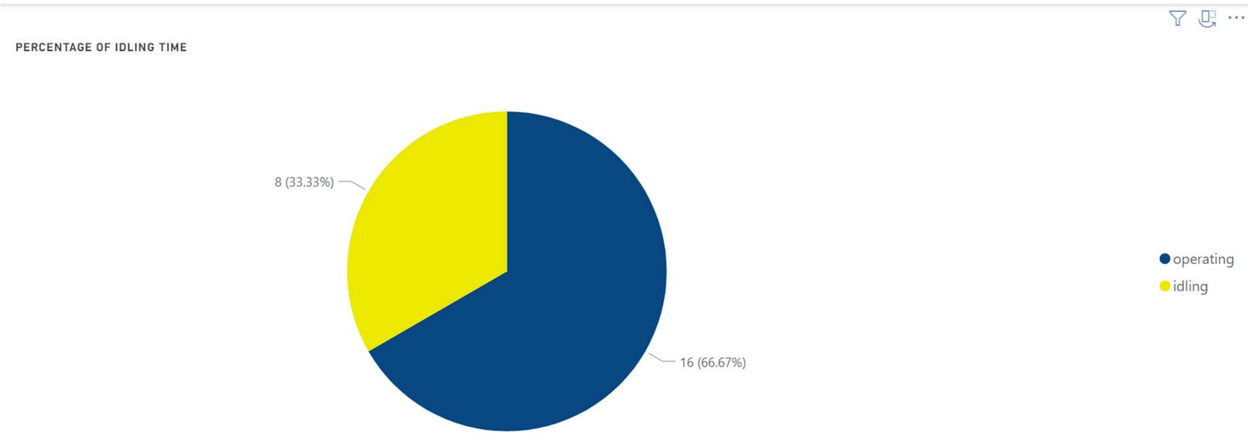

Figure 17e: Dashboard (Pie chart showing percentage of Idling hour)

Figure 17e introduces a pie chart that quantifies the idle hours as a percentage of total operational time within the selected timeframe. The chart contrasts idling hours in yellow against active operating hours in blue, providing a visual breakdown along with precise percentages and record numbers.

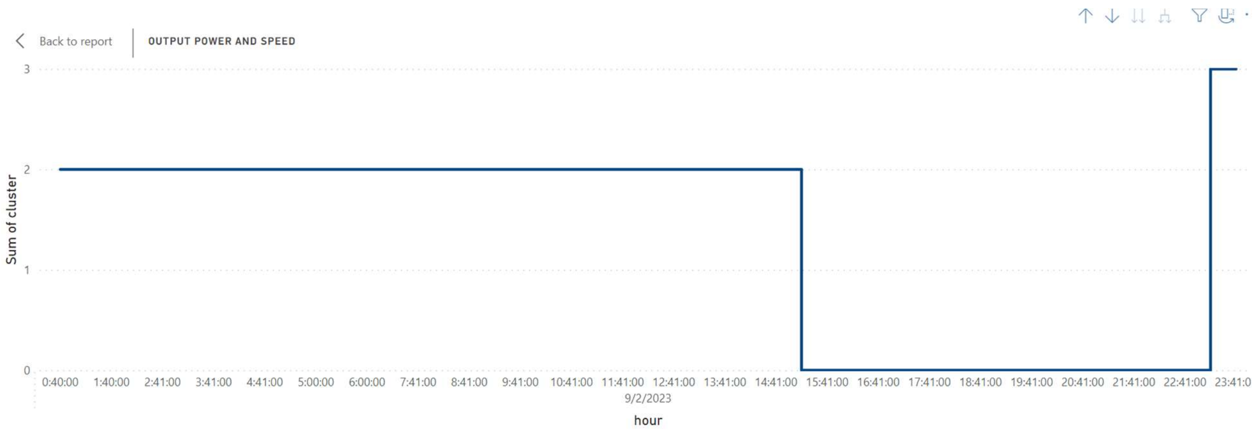

Figure 17f: Dashboard (Line chart showing the operation speed)

The line chart in Figure 17f delineates the operational speed of the device on the selected date, categorizing the data into four distinct groups labeled 0 through 3, which correspond to off, low, medium, and high speeds respectively.

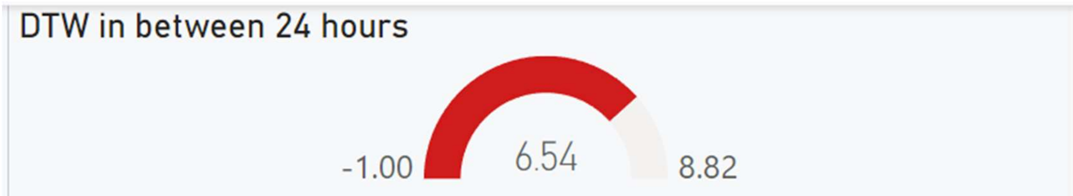

Figure 18a: Dashboard (Gauge Chart for average DTW score)

Figure 18a showcases a gauge chart that presents the average Dynamic Time Warping (DTW) score for the chosen date. Uncomputable DTW scores are indicated with a -1. The chart marks the minimum and maximum values on the left and right, respectively, with the average DTW score centered for quick reference..

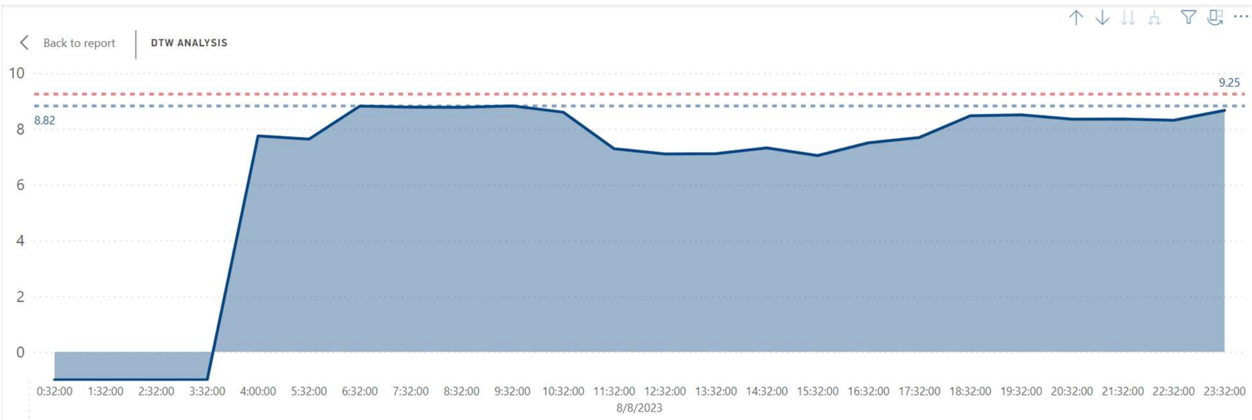

Figure 18b: Dashboard (Area Chart showing DTW score)

An area chart in Figure 18b breaks down the DTW scores by hour for the selected date, providing a granular view of the data. In instances where the DTW score cannot be calculated, a -1 is used to denote these gaps.

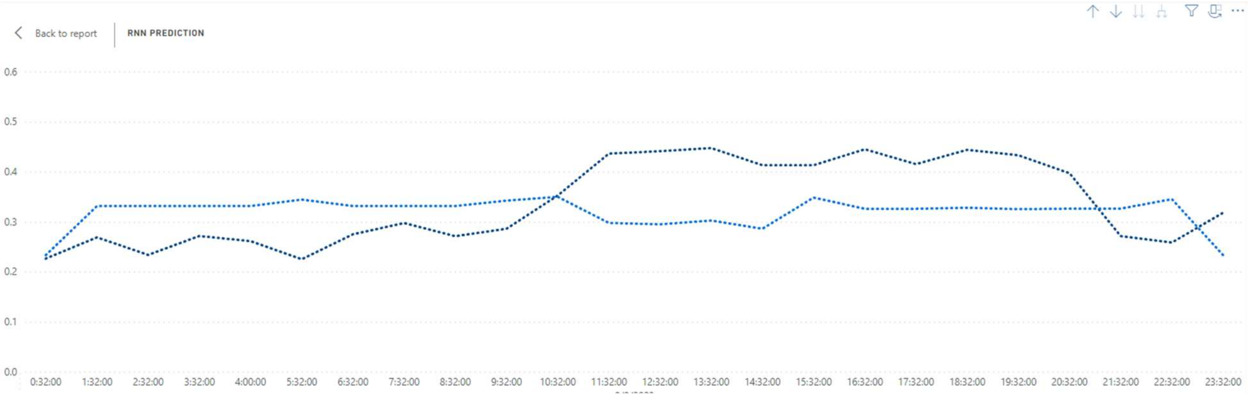

Figure 18c: Dashboard (Line Chart show RNN prediction and ground true)

Figure 18c features a line chart that compares the RNN model's predictions against the actual data, with the model's forecasts shown in light blue and the true values in deep blue. This visualization focuses on the axial vibration data, offering insights into the model's predictive accuracy.

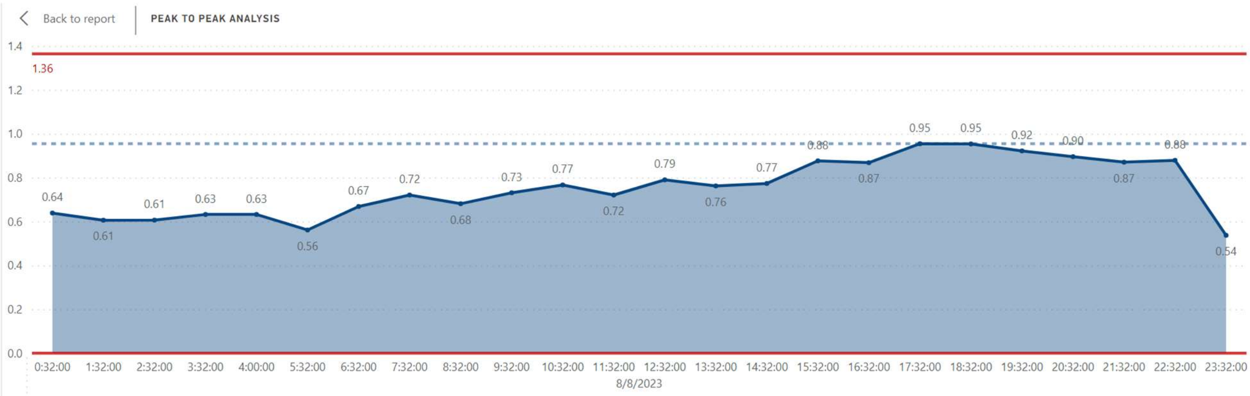

Figure 18d: Dashboard (Area chart for Peak-to-peak value analysis)

Figure 18d's area chart focuses on peak-to-peak value analysis, plotting these values hourly with a blue line and dot. The chart highlights the maximum value observed in the selected period with a dotted blue line, while a red line establishes the operational thresholds for the device, facilitating an immediate assessment of deviations from the norm.

6. Discussions

6.1 Difficulties

When it comes to datasets pertaining to mechanical systems or engineering, a limited understanding of mechanical principles can serve as a considerable obstacle. Accurate data interpretation and problem resolution in these fields necessitate a thorough grasp of the foundational mechanical concepts. A methodical review of existing literature stands out as a practical method to surmount this barrier. Engaging with scholarly articles and studies allows for the acquisition of vital knowledge on mechanical theories and applications, thereby filling in the knowledge gaps and facilitating more informed analyses and decisions within mechanical or engineering endeavors.

In the realm of data analysis, inconsistent sampling times pose a notable challenge, rendering the identification of missing records a daunting task. Such irregularities can introduce gaps that disrupt the continuity of the dataset, potentially skewing the accuracy of further analyses. Moreover, the dataset is significantly compromised by a high prevalence of invalid records, initially encompassing 12,196 rows and eventually dwindling to 6,514 rows. This considerable amount of flawed data undermines the dataset's integrity and complicates the processes of data handling and analysis.

The application of the DTW technique in data analysis demands significant computational resources, including extensive RAM and CPU capabilities, especially when managing sizable datasets or aiming for real-time analysis. In such scenarios, the utilization of the Department of Computer Science's research machines offers a strategic resolution to the computational demands imposed by DTW. This approach ensures the availability of adequate processing power to carry out efficient data analysis with DTW, surpassing the constraints posed by standard computational tools.

Furthermore, the absence of continuous data during motor operation presents a considerable challenge in training RNN, which rely on sequential data to forecast future states. The implementation of dynamic sequencing provides a creative resolution to this issue. By dynamically generating sequences from the available data, it's feasible to emulate scenarios of continuous operation. This strategy enables the effective training of RNN models, despite the intermittent nature of the data, ensuring that the models can assimilate the operational dynamics of the motors and accurately predict their forthcoming states.

6.2 Limitations

- Peak-to-Peak Value Analysis

In the Peak-to-Peak Value Analysis, the method's primary limitation lies in its inability to predict faults effectively. Minor misalignments or damages are challenging to detect directly in the signal, which means that by the time a peak-to-peak value signals an issue, the situation might already be dire. This limitation suggests that while peak-to-peak value analysis can be a useful tool for identifying significant problems, it might not be the most reliable method for early fault detection or for situations where small discrepancies need to be identified promptly.

- Dynamic Time Warping

The DTW technique, while powerful for pattern recognition and sequence alignment, comes with its own set of challenges. The foremost issue is its high computational demand; DTW requires considerable time to regenerate pattern sequences and calculate similarities across all sequences. This computational burden only grows as more data is collected, exacerbating the situation since calculations are constrained to CPU processing. Furthermore, DTW suffers from a long reaction time, needing continuous data over 12 hours to construct a single valid window. This means any interruption can lead to significant delays, requiring another 12-hour period to establish a new window, thus potentially missing critical early warning signs.

- Deep learning

The use of deep learning, specifically RNN, in this project introduces a different kind of limitation. While RNNs can be highly effective for sequential data analysis, their complexity can obscure the interpretability of the results. The visualization of warning signals is not straightforward, heavily relying on technicians' continuous observation and monitoring. This limitation highlights the need for enhanced interpretative tools or methodologies that can make the warning signals more apparent and actionable to non-expert users.

- Data dashboard

The dashboard designed to visualize and interact with the project's data exhibits several shortcomings. Primarily, the reliance on Microsoft software, Power Bi, for visualization purposes might not offer the most user-friendly or intuitive interface for all users. The dashboard's lack of mobile device support, particularly the calendar slicer's incompatibility with online web and mobile views, further restricts accessibility and convenience. Moreover, the absence of practical feedback from partners and users indicates that the dashboard may not be fully optimized for its intended audience, suggesting a gap in user-centric design and functionality.

6.3 Future Development

To address the limitations identified in the Peak-to-Peak Value Analysis, the integration of machine learning techniques can significantly enhance the method's predictive capabilities, enabling the system to forecast trends in peak-to-peak values and potentially identify issues before they escalate. This approach leverages historical data to improve fault detection and predict future trends, making the system more proactive rather than reactive.

For the Dynamic Time Warping (DTW) method, a strategic replacement of the standard DTW algorithm with more efficient variants like soft-DTW can offer significant benefits. Soft-DTW not only enhances pattern recognition capabilities but also supports GPU acceleration, thereby optimizing computational efficiency. Furthermore, the adoption of parallel processing and GPU acceleration can dramatically decrease the processing time required for data analysis, making the system more efficient and capable of handling larger datasets without significant delays.

In the realm of Deep Learning, replacing the current GRU architecture with LSTM could offer better performance due to LSTM's enhanced ability to capture long-term dependencies in data sequences. Moreover, the introduction of feedback loops, where technicians' insights and corrections inform continuous model training, can significantly improve the model's predictive accuracy and reliability. This iterative process ensures that the model evolves and adapts to new patterns or anomalies, maintaining its effectiveness over time.

Finally, for the Dashboard, setting up a cloud server for local deployment, combined with a user-centered optimization of the dashboard design, can ensure that the visualization is both intuitive and accessible. This should be informed by user feedback to meet the specific needs and preferences of its audience. Implementing a responsive design and ensuring cross-platform compatibility are critical steps toward making sure the dashboard's features, including the calendar slider, are fully functional and user-friendly across all devices, thereby enhancing overall user accessibility and satisfaction.

7. Conclusions

In conclusion, the project successfully developed and implemented a comprehensive vibration analytics system for motor health monitoring in water supply systems. Through the use of advanced data acquisition, processing techniques, and deep learning models, the project was able to continuously monitor the health of motors, evaluate their operational efficiency, and formulate optimization strategies. The exploratory data analysis and feature extraction methods provided significant insights into the condition of the motors, enabling early detection of potential issues and facilitating proactive maintenance strategies.

The peak-to-peak value analysis and sliding window technique further enhanced the system's ability to detect anomalies and assess motor health accurately. The development of a user-friendly dashboard for data visualization and the deep learning model utilizing recurrent neural networks for predicting motor conditions were key achievements of the project.

Despite facing challenges such as dealing with large amounts of missing or invalid data, computational limitations, and the need for continuous data for effective model training, the project overcame these obstacles through innovative solutions and optimizations. The limitations identified, such as the peak-to-peak value analysis's inability to detect minor faults effectively and the computational demands of the DTW technique, provide areas for future development. Enhancements such as integrating machine learning techniques for trend forecasting, adopting more efficient DTW variants, and optimizing the deep learning model could further improve the system's performance and reliability.

Overall, the project has laid a foundation for the advancement of predictive maintenance in the context of water supply systems, contributing to operational improvements, cost reductions, and the prolongation of motor life. The insights gained and the methodologies developed have implications for the broader field of smart facility management, highlighting the potential for technology-driven solutions to enhance system efficiency and sustainability.

REFERENCES

1. Al-Raheem, K. F., Roy, A., Ramachandran, K. P., Harrison, D. K., & Grainger, S. (2009). Rolling element bearing faults diagnosis based on autocorrelation of optimized: wavelet de-noising technique. *The International Journal of Advanced Manufacturing Technology*, *40*(3–4), 393–402. https://doi.org/10.1007/s00170-007-1330-3

2. Chandel, A. (2023, October 17). What are the Overall Vibration measurements and their use in Predictive Maintenance? - SenseGrow. *AI Software & IoT Sensors for Predictive Maintenance*. https://www.sensegrow.com/blog/iot-solutions/overall-vibration-measurements#:~:text=Peak%20to%20Peak%20is%20the,one%20extreme%20to%20the%20other.

3. Del Rosso, V., Andreucci, A., Boria, S., Corradini, M. L., & Ranalli, A. (2021). Mechanical fault detection for induction motors based on vibration analysis: a case study. *47th Annual Conference of the IEEE Industrial Electronics Society*. https://doi.org/10.1109/iecon48115.2021.9589189

4. De Melo, Á. S., Rocha, O. P., & Santos, A. Á. B. (2022). Computational model for electrical motors condition analysis and monitoring. *Journal of Bioengineering and Technology Apllied for Health*, *5*(2), 100–105. https://doi.org/10.34178/jbth.v5i2.206

5. Sharma, A., & Jia, Z. (2021). Review on engine vibration fault analysis based on data mining. *Journal of Vibroengineering (Spausdinta)*, *23*(6), 1433–1445. https://doi.org/10.21595/jve.2021.21928

6. Da Costa-Luis, C. (n.d.). *tqdm documentation*. https://tqdm.github.io/

7. *datetime — Basic date and time types*. (n.d.). Python Documentation. https://docs.python.org/3/library/datetime.html

8. Jayaswal, P., Wadhwani, A. K., & Mulchandani, K. B. (2008). Machine Fault Signature Analysis. *International Journal of Rotating Machinery*, *2008*, 1–10. https://doi.org/10.1155/2008/583982

9. *multiprocessing — Process-based parallelism*. (n.d.). Python Documentation. https://docs.python.org/3/library/multiprocessing.html

10. *Matplotlib documentation — Matplotlib 3.8.2 documentation*. (n.d.). https://matplotlib.org/stable/index.html

11. *NumPy documentation — NumPy v1.26 Manual*. (n.d.). https://numpy.org/doc/stable/

12. *Overview - YData Profiling*. (n.d.). https://docs.profiling.ydata.ai/latest/getting-started/concepts/

13. *pandas documentation — pandas 2.1.4 documentation*. (n.d.). https://pandas.pydata.org/docs/

14. Ribeiro, R. F., Areias, I. a. D. S., & Gomes, G. F. (2021). Fault detection and diagnosis using vibration signal analysis in frequency domain for electric motors considering different real fault types. *Sensor Review*, *41*(3), 311–319. https://doi.org/10.1108/sr-02-2021-0052

15. Ramtekkar, P., Bhute, Y., Chauhan, M., Khandelwal, V., & Deshmukh, S. (2023). Advanced Vibration Analyzer of Motor. *2023 IEEE 3rd International Conference on Technology, Engineering, Management for Societal Impact Using Marketing, Entrepreneurship and Talent (TEMSMET)*. https://doi.org/10.1109/temsmet56707.2023.10149993

16. *seaborn: statistical data visualization — seaborn 0.13.0 documentation*. (n.d.). https://seaborn.pydata.org/

17. *scikit-learn: machine learning in Python — scikit-learn 1.3.2 documentation*. (n.d.). https://scikit-learn.org/stable/index.html

18. *SciPy documentation — SciPy v1.11.4 Manual*. (n.d.). https://docs.scipy.org/doc//scipy/index.html

19. *tslearn's documentation — tslearn 0.6.3 documentation*. (n.d.). https://tslearn.readthedocs.io/en/stable/index.html

20. Sampaio, G. S., De Aguiar Vallim Filho, A. R., Da Silva, L. S., & Da Silva, L. A. (2019). Prediction of motor failure time using an artificial neural network. *Sensors (Basel)*, *19*(19), 4342. https://doi.org/10.3390/s19194342

21. Mishra, K., Jigyasu, R., Singh, S., & Chikkam, S. (2023). Motor Bearing Fault Prediction Using Artificial Intelligence Techniques. *IEEE Xplore*. https://doi.org/10.1109/icmoce57812.2023.10165790

22. Borrè, A., Seman, L. O., Camponogara, E., Stefenon, S. F., Mariani, V. C., & Coelho, L. D. S. (2023). Machine fault detection using a hybrid CNN-LSTM Attention-Based model. *Sensors (Basel)*, *23*(9), 4512. https://doi.org/10.3390/s23094512

23. *PyTorch documentation — PyTorch 2.2 documentation*. (n.d.). https://pytorch.org/docs/stable/index.html

APPENDICES

1. Analytical Environments and Tools

	Tools	Description
OS	Windows 11	/
Language	Python 3.10.13	/
	Conda 20.1	/
Code editor	Visual Studio Code	/
Code storage	Github	/
Libraries	Numpy	It provides a high-performance multidimensional array object and tools for working with these arrays.
	Pandas	Pandas is a fast, powerful, flexible, and easy-to-use open-source data analysis and manipulation tool, built on top of the Python programming language.
	Matplotlib	Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python.
	Seaborn	Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.
	Datetime	The datetime module in Python supplies classes for manipulating dates and times. It offers various functions and constants for all date and time manipulation needs.
	Ts-learn	Tslearn is a Python package designed for machine learning with time series data. It is built upon other popular libraries like scikit-learn, numpy, and scipy, providing specialized tools and functionalities for time series analysis. This makes it a valuable resource for tasks that involve temporal data.
	Scikit-learn	Scikit-learn is a machine learning library in Python that offers simple and efficient tools for predictive data analysis. It is accessible and reusable in various contexts, built on top of libraries like NumPy, SciPy, and matplotlib. Scikit-learn is known for its versatility in implementing numerous machine learning algorithms for

		classification, regression, clustering, and more.
	Tqdm	Tqdm is a Python library that provides fast, extensible progress bars for loops and other iterable objects. It is designed to be simple to use and highly customizable, adding immediate feedback on the progress of Python functions.
	Multiprocessing	The multiprocessing module in Python supports spawning processes using an API similar to the threading module. It provides local and remote concurrency, effectively sidestepping the Global Interpreter Lock by using subprocesses instead of threads.
	Scipy	SciPy is an open-source Python library used for scientific and technical computing. It includes modules for optimization, integration, interpolation, eigenvalue problems, algebraic equations, differential equations, and others. SciPy builds on NumPy, and its array operation is the core part of SciPy.
	Ydata_profiling	YData Profiling provides comprehensive profiling for various types of data such as tabular, time-series, text, and image data.
	Torch	Torch is an efficient and flexible open-source machine learning library, focused on deep learning. Built on Lua, it supports diverse neural network architectures and optimization algorithms, making it ideal for rapid prototyping and complex model development.
Visualization tool	Power Bi Desktop	Power BI Desktop is an interactive data visualization and analytics tool from Microsoft. It allows users to connect to a wide array of data sources, transform data, and create rich, interactive reports and dashboards.
Database	MySQL 8.0	MySQL is an open-source relational database management system. It facilitates efficient data storage, retrieval, and manipulation, making it a cornerstone for dynamic websites and applications.

Table 1: Used Tools and Environments