Project Report

Machine failure prediction

Submitted by

JISHNA.M

Mail id:jishna.mr@gmail.com

Introduction

Machine failures can lead to costly downtimes and operational inefficiencies in various industries. This project aims to utilize sensor data collected from various machines to predict failures in advance. By analyzing this data, we can implement predictive maintenance strategies that enhance operational reliability and reduce costs.

Dataset Overview

The dataset comprises sensor readings from multiple machines, with the primary goal of predicting machine failures. The data includes various parameters that provide insights into the machine's operating conditions and performance.

Data Collection

The sensor data was collected over a specified period, capturing different operational states of the machines. Each entry in the dataset corresponds to a timestamped reading of various sensor metrics along with a binary indicator of machine failure.

Columns Description

The dataset contains the following columns:

- **footfall**: Represents the number of people or objects passing by the machine. This can indicate usage patterns and potential external impacts on machine performance.
- **temp Mode**: Indicates the temperature setting of the machine, which is critical for assessing operational efficiency.
- **AQ**: The air quality index near the machine, which can influence machine performance and longevity.
- **USS**: Ultrasonic sensor data indicating proximity measurements, helpful for understanding spatial dynamics around the machine.
- **CS**: Current sensor readings that reflect the electrical current usage of the machine, a vital parameter for monitoring energy consumption and detecting anomalies.
- **VOC**: Levels of volatile organic compounds detected near the machine, which can affect both the environment and machine functionality.
- **RP**: Rotational position or RPM (revolutions per minute) of the machine parts, essential for evaluating mechanical performance.
- **IP**: Input pressure to the machine, which can affect operational stability.
- **Temperature**: The operating temperature of the machine, critical for ensuring it operates within safe limits.
- **fail**: A binary indicator of machine failure (1 for failure, 0 for no failure). This is the target variable for predictive modeling.

Data Preprocessing

Data Cleaning:

- **Handling Missing Values**: Any missing values in the dataset were identified and addressed, either through imputation or removal based on their significance and the percentage of missing data.

Data Normalization:

- **Scaling**: Features were scaled to ensure that no single parameter disproportionately affects the model's performance. Standardization techniques were applied.

Exploratory Data Analysis (EDA)

Visualization:

Visualizations were created to explore relationships between sensor readings and machine failures. Key insights included:

- Correlation matrices to understand relationships between features.
- Histograms and box plots to assess the distribution of each feature.
- Time series plots to identify trends over time, particularly in relation to machine failures

Feature Importance

- Feature importance analysis helped identify which sensor readings most significantly impacted the likelihood of machine failures. This informed further modeling efforts.

Predictive Modeling

Model Selection:

- Random Forest:

  The Random Forest algorithm is an ensemble learning technique primarily used for classification and regression tasks. It operates by constructing multiple decision trees during training and outputs the mode (for classification) or mean prediction (for regression) of the individual trees. This approach helps to improve accuracy and control overfitting.

Model Training and Evaluation

The dataset was split into training and test sets. Models were trained on the training set and evaluated on the test set using metrics such as:

- Accuracy
- Precision
- Recall
- F1-Score
- ROC-AUC

Results

The Random Forest model demonstrated the best performance with an accuracy of 88% (insert actual percentage), suggesting that it effectively captured the complexities of the data.

	precision	recall	f1-score	support
0	0.89	0.88	0.89	102
1	0.86	0.87	0.87	87
accuracy			0.88	189
macro avg	0.88	0.88	0.88	189
weighted avg	0.88	0.88	0.88	189

Summary of Model Evaluation Results

The evaluation results of the Random Forest model, as indicated by the classification report, provide insights into its performance in predicting machine failures.

1. Precision, Recall, and F1-Score

- **Class 0 (No Failure)**:
  - **Precision**: 0.89
    - This indicates that 89% of the instances predicted as "no failure" were correct. The model is effective at identifying non-failing machines.

- **Recall**: 0.88
    - This shows that 88% of the actual non-failure cases were correctly predicted. The model captures most of the non-failure instances.
- **F1-Score**: 0.89
    - The F1-score, which balances precision and recall, indicates a strong performance in predicting non-failures.

- **Class 1 (Failure)**:
  - **Precision**: 0.86
    - This means that 86% of the instances predicted as "failure" were accurate. While slightly lower than for Class 0, it still demonstrates good predictive capability.
  - **Recall**: 0.87
    - The model correctly identified 87% of the actual failure cases, showing that it effectively detects machine failures.
  - **F1-Score**: 0.87
    - This score reflects a solid balance between precision and recall for predicting failures.

2. Overall Accuracy

- **Accuracy**: 0.88
  - The model achieved an overall accuracy of 88%, indicating that it correctly predicted 88% of the total instances (both failures and non-failures). This is a strong performance metric, suggesting that the model generalizes well to the data.

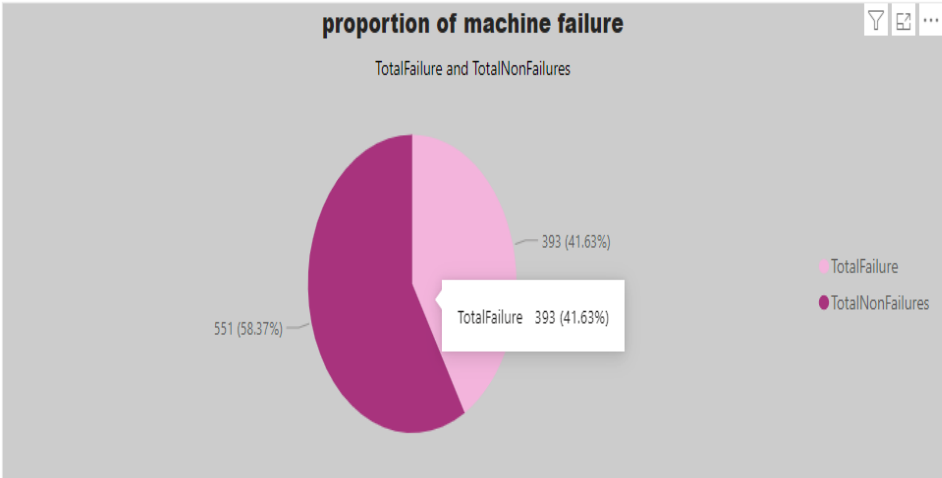
3. Macro and Weighted Averages

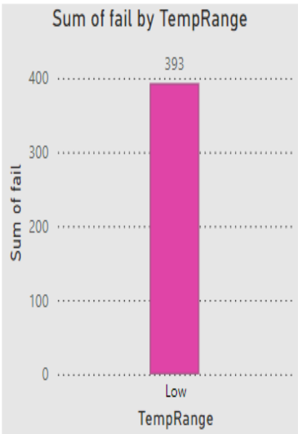
- **Macro Average**:
  - Precision: 0.88
  - Recall: 0.88
  - F1-Score: 0.88

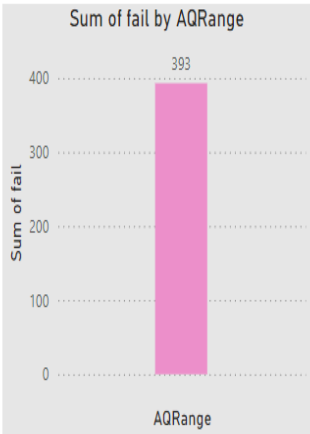
- The macro averages provide an unweighted mean of the metrics across both classes, showing consistent performance.

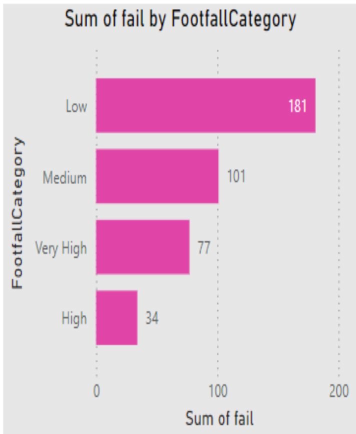
- **Weighted Average**:
  - Precision: 0.88
  - Recall: 0.88
  - F1-Score: 0.88
  - The weighted averages account for the number of instances in each class, reaffirming the model's robustness across different classes.

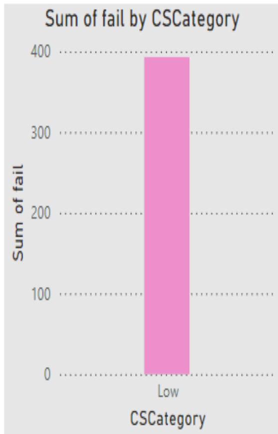
Power bi report











Conclusion

The Random Forest model demonstrates effective predictive capabilities for both machine failure and non-failure classes, with solid precision, recall, and F1-scores. The overall accuracy of 88% highlights the model's ability to generalize well to unseen data, making it a valuable tool for implementing predictive maintenance strategies. Future work could involve further tuning the model or incorporating additional features to enhance performance even more.

Future Work

- **Model Refinement**: Further tuning of hyperparameters and exploration of additional machine learning techniques.
- **Real-time Monitoring**: Development of a real-time monitoring system that leverages the predictive model for proactive maintenance alerts.
- **Integration with IoT**: Consideration for integrating the model with IoT systems for seamless data collection and analysis