Predictive Maintenance - Machine Breakdown Prediction
Final Report

Submitted by: Bharadwaj Phani Datta Mahanthi

Institution: KnowledgeHut Upgrad

Date: September 05, 2024

Abstract

Predictive maintenance plays a crucial role in minimizing machine downtime by predicting failures and allowing for proactive maintenance. This report outlines the development of a predictive maintenance system aimed at identifying anomalies leading to machine breakdowns. The project involved the use of a Random Forest Classifier and extensive feature engineering to achieve an accuracy of 99.96%. Future work includes exploring deep learning models and real-time data integration.

Introduction

Predictive maintenance involves using data-driven techniques to predict when machinery or equipment will fail, allowing for maintenance to be performed just in time. This prevents unexpected breakdowns and costly downtimes. The goal of this project was to build a system that can predict machine anomalies using historical data, thus preventing breakdowns. The dataset used in this project consisted of 18,000+ rows, capturing various features that represented the machine's condition.

1. Design Choices and Performance Evaluation

Model Selection

A Random Forest Classifier was selected for this project due to its robustness, interpretability, and capability to model complex, non-linear relationships. It also performs well with datasets that contain both continuous and categorical features.

Preprocessing

Missing values were handled using mean imputation for continuous variables and mode imputation for categorical variables. Outliers were analyzed using box plots, but due to the Random Forest's robustness, no scaling or removal was necessary. Standard scaling was applied to the features to ensure they are on the same scale.

Feature Engineering

Feature engineering was critical in improving model performance. Interaction terms and polynomial features were created to model non-linear relationships. Log transformations were applied to handle skewness in the data.

Hyperparameter Tuning

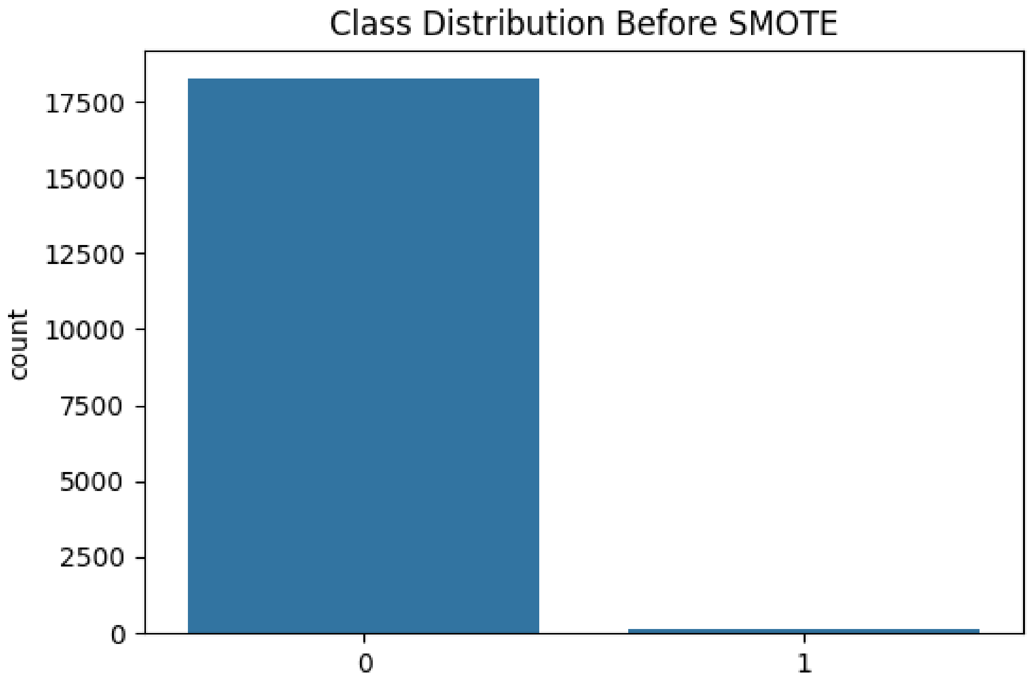
RandomizedSearchCV was used to tune hyperparameters such as 'n_estimators', 'max_depth', 'min_samples_split', and 'min_samples_leaf'. The optimal parameters chosen for the model are:

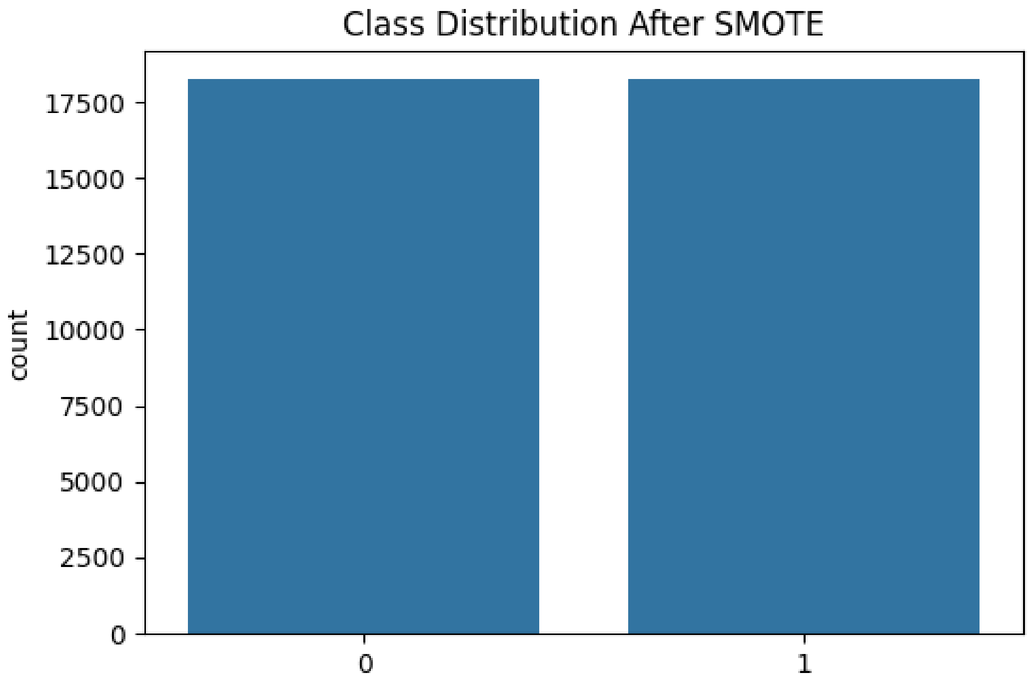
- n_estimators: 200
- max_depth: 15
- min_samples_split: 2
- min_samples_leaf: 1
- max_features: 'log2'
- bootstrap: True

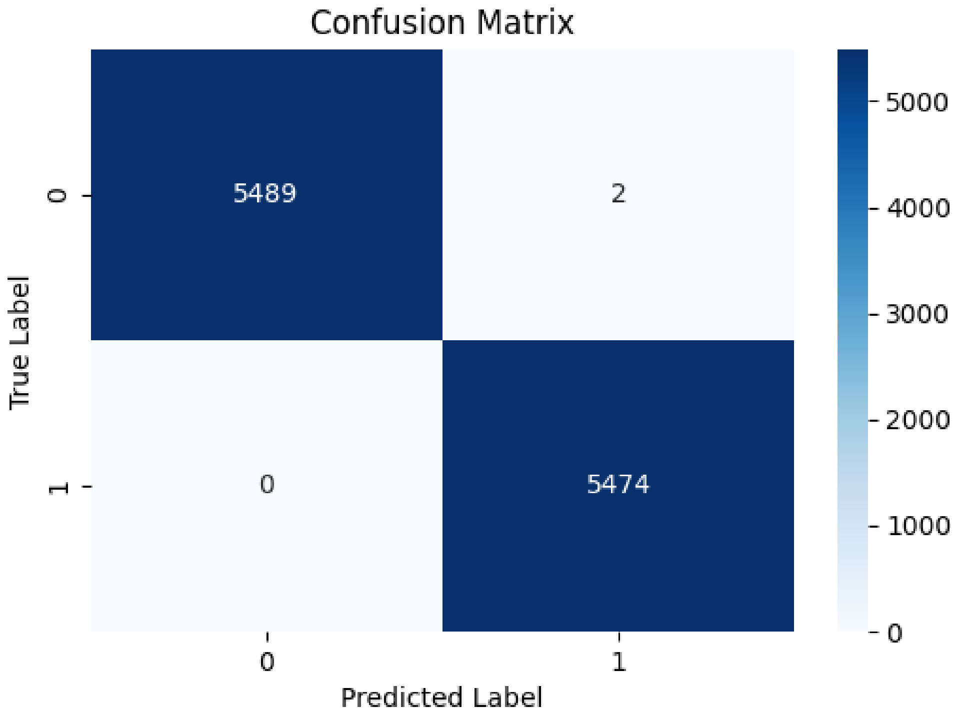
Model Evaluation

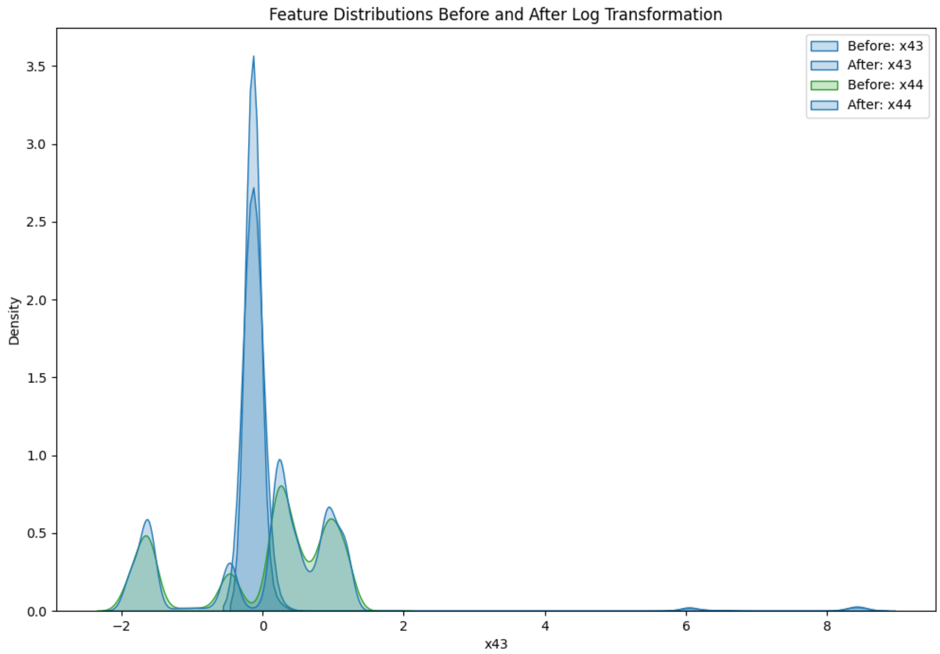
The model was evaluated using a 70-30 train-test split. The Random Forest model achieved an accuracy of 99.96%, an F1-score of 99.96%, a precision of 99.93%, and a recall of 100%. These metrics indicate that the model performed exceptionally well in predicting anomalies with minimal false positives or negatives.

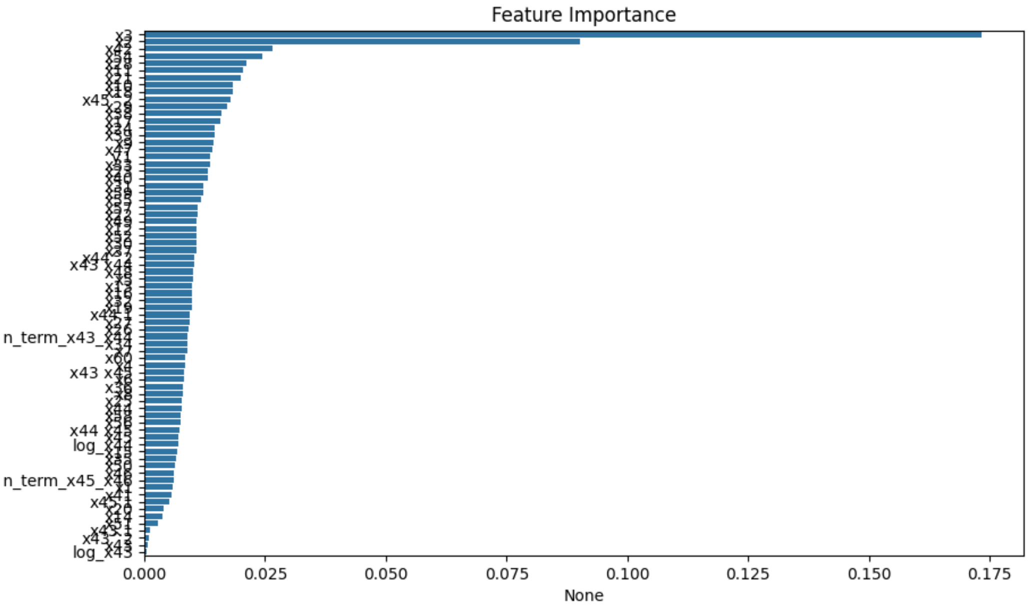
Evaluation Visualizations

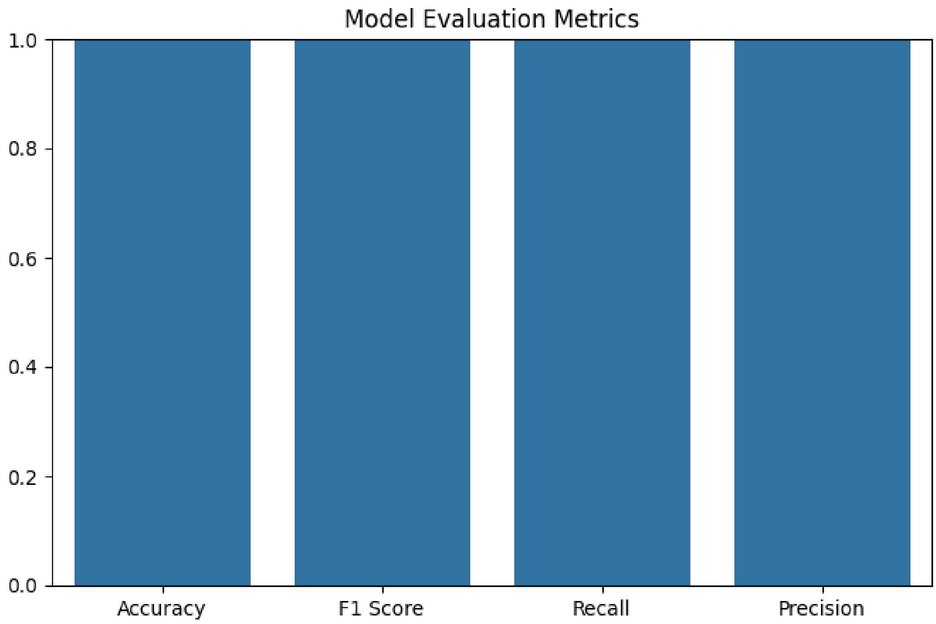


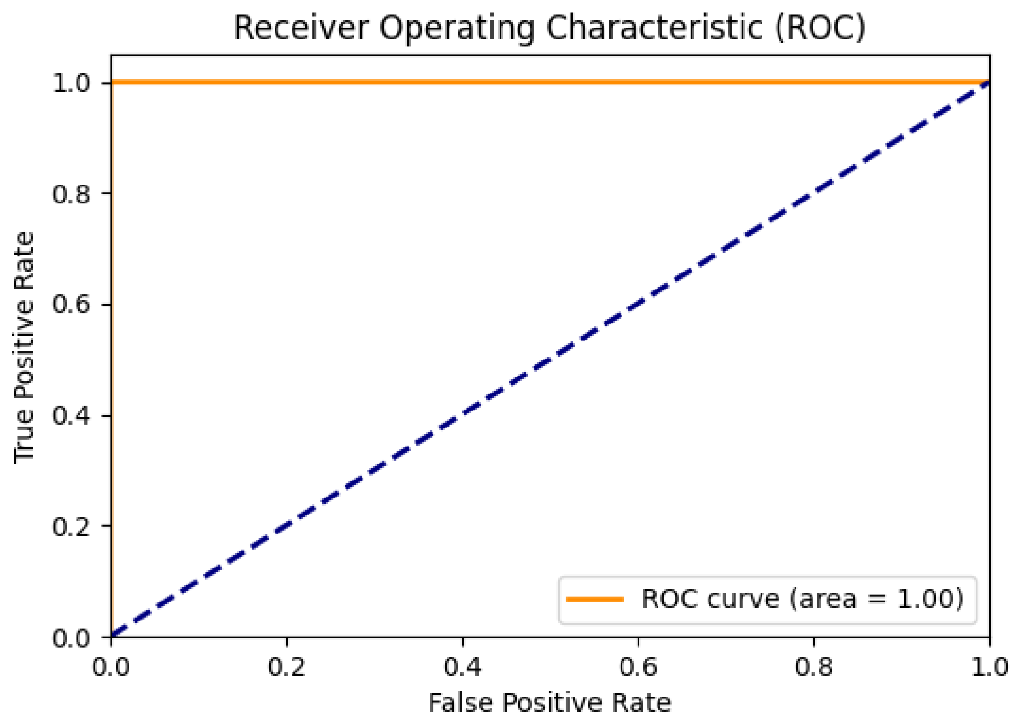



Confusion Matrix
True Label
Predicted Label
5489
2
0
5474
0
1
0
1
5000
4000
3000
2000
1000
0


Feature Distributions Before and After Log Transformation
Before: x43
After: x43
Before: x44
After: x44
Density
x43


Feature Importance
None
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175


Model Evaluation Metrics
1.0
0.8
0.6
0.4
0.2
0.0
Accuracy
F1 Score
Recall
Precision



Conclusion

This project successfully demonstrated the capability of a Random Forest Classifier to predict machine breakdowns with an impressive accuracy of 99.96%. The robust preprocessing, feature engineering, and hyperparameter tuning processes all contributed to the model's performance. Future work could involve deep learning models and real-time data integration to further enhance the system's predictive power.