Weekly Progress Report

Name: Nishant Mall

Domain: Data Science and Machine Learning

Date of submission: 27 Sept 2024

Week Ending: 04

I. Overview

This week, I continued to focus on the two projects: Predicting agriculture crop production in India and Predicting maintenance of gearboxes using vibration sensors. I delved deeper into model selection and evaluation, and explored techniques for handling data imbalance and improving model interpretability.

II. Achievements

1. **Prediction of Agriculture Crop Production in India:**
    - **Regression Modeling:** Employed regression models, including linear regression and multiple linear regression, to establish relationships between crop production and various factors such as rainfall, temperature, and soil conditions.
    - **Model Evaluation:** Evaluated the performance of regression models using metrics like R-squared, mean squared error (MSE), and root mean squared error (RMSE).
2. **Gearbox Predictive Maintenance:**
    - **Random Forest:** Utilized random forest, a powerful ensemble learning technique, to classify gearbox health based on vibration sensor data.
    - **Feature Importance:** Determined the importance of different features in the gearbox predictive maintenance model using techniques like permutation importance.
    - **Hyperparameter Tuning:** Optimized the hyperparameters of the random forest model to improve its performance.

III. Challenges

1. **Data Quality:** Encountered issues with missing values and inconsistencies in the agricultural dataset.
    - **Imputation Techniques:** Implemented statistical imputation techniques, such as mean, median, or mode imputation, to handle missing values.
    - **Outlier Detection:** Utilized statistical methods, like Z-scores to identify and address outliers in the data.

2. **Data Imbalance:** The gearbox predictive maintenance dataset may be imbalanced, with fewer instances of gearbox failures compared to normal operations.
    - **Overfitting and Underfitting:** Applied techniques like Overfitting (e.g., SMOTE) and underfitting to balance the class distribution.
    - **Cost-Sensitive Learning:** Considered adjusting the class weights in the learning algorithm to account for the imbalance.

IV. Learning Resources

- **Machine Learning Algorithms:** I explored various machine learning algorithms, including linear regression, decision trees, random forests, and support vector machines, to identify suitable models for both projects.
- **Model Evaluation Metrics:** I studied different metrics for evaluating model performance, such as mean squared error (MSE), root mean squared error (RMSE), R-squared, and accuracy, precision, recall, and F1-score.

V. Additional Comments

Working on these projects has been a rewarding experience. I have gained valuable insights into data science and machine learning techniques. It's been fun to explore different algorithms and approaches to solve real-world problems. I am looking forward to continuing my work on these projects and applying my knowledge to future endeavours.