Weekly Progress Report

Name: Nishant Mall
Domain: Data Science and Machine Learning
Date of submission: 22 Sept 2024

Week Ending: 03

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:

- **Statistical Analysis:** Conducted a comprehensive statistical analysis to identify significant correlations and dependencies between variables.
- **Probabilistic Modeling:** Explored probabilistic models, such as Bayesian networks, to capture uncertainty and dependencies in the data.
- **Feature Engineering:** Created new features based on statistical measures, including standard deviations, percentiles, and z-scores, to improve model performance.

2. Gearbox Predictive Maintenance:

- **Time Series Analysis:** Utilized time series analysis techniques, like autocorrelation and partial autocorrelation functions, to identify trends and patterns in the vibration data.
- **Statistical Hypothesis Testing:** Applied statistical hypothesis testing to assess the significance of differences between normal and abnormal gearbox conditions.
- **Probabilistic Inference:** Employed probabilistic inference methods, such as hidden Markov models, to model the dynamic behavior of gearbox health.

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 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. Next Week's Goals

- **Agriculture Crop Production Prediction:**
    - Fine-tune the selected models using hyperparameter tuning techniques.
    - Evaluate the models on a holdout dataset to assess their generalization performance.
- **Gearbox Predictive Maintenance:**
    - Implement techniques to handle data imbalance and improve model accuracy, such as Overfitting or Underfitting.
    - Explore methods for improving model interpretability, such as feature importance analysis and SHAP values.

VI. Additional Comments

I am making steady progress on both projects and am excited to see how the models perform in real-world scenarios. I am also eager to explore techniques for deploying these models and integrating them into practical applications.