Predicting Weightlifting Exercise Quality Using Accelerometer Data

Introduction

This project aims to predict the quality of weightlifting exercises using accelerometer data collected from the belt, forearm, arm, and dumbbell of six participants. The goal is to classify movements into one of five predefined categories (represented by the "classe" variable) using machine learning techniques.

Data Collection and Preparation

The dataset consists of sensor measurements recorded while participants performed exercises correctly and incorrectly. The training data was loaded and cleaned by removing columns with too many missing values, unnecessary identifiers, and features with near-zero variance. This preprocessing ensured that only relevant and informative features were used for model training.

Exploratory Data Analysis

To understand the relationships between features, a correlation analysis was performed. Highly correlated features were identified to prevent redundancy in the model. Additionally, class distributions were checked to ensure balanced representation in the dataset.

Model Training

A Random Forest model was selected for its robustness and accuracy in classification tasks. The dataset was split into training and validation sets, with 70% of the data used for training and the remaining 30% for validation. The model was trained on the processed dataset using 100 decision trees.

Model Evaluation

After training, the model was tested on the validation set. The confusion matrix revealed high accuracy in predicting the correct class of exercises, indicating that the model effectively captured the patterns within the data.

Prediction on Test Data

Once the model was validated, it was applied to the test dataset to classify 20 new cases. The predictions were saved for submission in the required format.

Conclusion

This project successfully developed a machine learning model to classify weightlifting exercises based on accelerometer data. The Random Forest model demonstrated strong performance, making it a reliable choice for similar classification tasks. Future improvements could involve hyperparameter tuning or exploring deep learning techniques for enhanced accuracy.