ML Project Weekly Report

Course name: Machine learning (CSE 523)
Week: Week-2 report
Group Name: Thunder
Instructor's name: Prof. Mehul S Raval

Project 9: Data-driven imputation scheme for human-subject based dataset

Project Progress:

Managing Missing Values: Reflecting on Strategies and Techniques in Data Preprocessing:

- Systematic examination of the multi-modal dataset of Division I basketball players to identify missing data across all features involving a comprehensive review of each feature to gauge the extent of missing information.
- Analysis of the potential reasons behind the missing values for devising suitable strategies for addressing them.
- Reflection on various approaches to handle missing values through a combination of imputation methods and the potential removal of incomplete records.
- Consideration of potential data imputation techniques, such as mean, median, and mode imputation, along with methods like MICE for effectively handling missing data.

Challenges faced:

Systematic examination revealed variable missingness patterns across features, confounding the choice of appropriate imputation algorithms. Understanding the fundamental causes of missing data, such as athlete neglect or gadget malfunction, was critical for developing successful solutions.

Next Steps:

Our following approach will take a sequential approach to data analysis, beginning with data imputation to manage missing values and assure dataset completeness. We will perform a feature sensitivity analysis to determine the importance of different features in our model. By carefully applying these processes, we want to optimize our dataset for analysis, obtain insights on feature relevance, and improve our model's performance.

Conclusion:

Our data preprocessing method indicated difficulties in handling missing values in our Division I basketball player dataset. We realize the need to analyse missing data patterns to identify appropriate imputation algorithms methodically. Next, we will prioritize data imputation to ensure dataset completeness and perform feature sensitivity analysis to refine our model.