

ML Project Weekly Report

Course name: Machine learning (CSE 523)
Week: Week-1 report
Group Name: Thunder
Instructor's name: Prof. Mehul S Raval

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

Project Overview:

The project strives to handle missing values in a multi-modal dataset of Division I basketball players, containing factors such as sleep patterns, training details, cardiac rhythm patterns, emotional-mental states, game scores, weekly readiness scores, and jump data (RSImod). The dataset contains missing values because of the possibilities of athlete negligence or device malfunction, hindering comprehensive analysis and insights into player performance and well-being.

Objective:

Creating a data-driven imputation strategy to efficiently fill in missing values in the Division I basketball dataset is the primary goal of this research. Throughout the project, our plan calls for feature sensitivity analysis to find the essential features for imputation to be conducted, along with model development for accurate implementation, comparison with state-of-the-art imputers, and validation of imputed values.

Project Progress:

- Preliminary data exploration to capture the dataset's structure and characteristics.
- Identification of missing values across diverse features and evaluation of their impact on data reliability.
- Exploration of potential techniques for feature sensitivity analysis to pinpoint significant features for imputation.
- Initiation of literature review to understand existing imputation methods and relevant studies in sports analytics and missing data imputation.

Next Steps:

- Complete feature sensitivity analysis to identify the most influential features for imputation.
- Begin model development for missing data imputation, considering techniques such as regression, decision trees, or neural networks.
- Explore state-of-the-art imputation methods, including Multiple Imputation by Chained Equations (MICE), for comparison purposes.
- Prepare for validation and evaluation of imputation models using appropriate metrics and cross-validation techniques.

Conclusion:

The project has made initial progress in understanding the dataset and identifying areas with missing data. Moving forward, the focus will be on analysing feature sensitivity, constructing models for

predicting missing values, and comparing them with existing imputation methods. Rectifying missing values will enhance the dataset's utility, enabling in-depth exploration of Division I basketball player performance and well-being. Additionally, data imputation techniques, feature extraction, classification, and comparison with the MICE technique will be included in the analysis