

ML Project Weekly Report

Course Name: Machine Learning (CSE 523)

Week Number: 4 (25th February - 2nd March 2024)

Group Name: White

Instructor's Name: Mehul S Raval

Project no.: 7

Project 7 Athlete Statistics Visualization and Prediction

Introduction

During the fourth week, our efforts were directed towards exploring various framework for athletic performance prediction. Our focus remained on understanding the complexities of our dataset and identifying suitable methods for feature engineering and model selection. But we mainly concentrated on the exploration of two areas: dimensionality reduction and model selection.

Weekly Activities

1. Feature Engineering Exploration:

We dedicated this week to exploring various feature engineering techniques aimed at condensing the feature space while retaining significant variance. Our exploration included techniques such as Principal Component Analysis (PCA), K-means Clustering, Correlation Matrix analysis, and Weighted Rank-Sum (Hard Voting).

2. Model Selection:

We researched and evaluated various machine learning algorithms suitable for predicting performance scores based on the reduced set of features:

- **Linear Regression:** This widely understood algorithm establishes a linear relationship between features and performance scores. We considered its interpretability, allowing us to understand how each feature contributes to the prediction.

- **Random Forest:** This ensemble method combines multiple decision trees, leading to more robust and accurate predictions compared to a single tree. We were interested in its ability to manage data uncertainties.
- **XGBoost:** This powerful ensemble method is known for its efficiency and accuracy in diverse prediction tasks. We explored its potential to handle the complex relationships between features and performance.

Challenges Faced

1. Selecting the Optimal Approach:

Identifying the most suitable dimensionality reduction technique and machine learning algorithm proved challenging. We meticulously evaluated each approach based on factors like information preservation, interpretability, and accuracy.

2. Balancing Accuracy and Interpretability:

Striking a balance between achieving high prediction accuracy and ensuring model interpretability was another key challenge. This balance is crucial for understanding how the model arrives at its predictions and gaining valuable insights for athletes and coaches.

Learnings

- We learned the importance of considering model selection criteria like interpretability and accuracy when building performance prediction models.
- We actively explored resources like online tutorials and research papers to enhance our understanding of these critical concepts.
- Through our exploration of model selection approaches, we learned about the strengths and limitations of Linear Regression, Random Forest, XGBoost, and SVM models in handling uncertainties and categorical outcomes within our dataset.

Conclusion

The fourth week of our project marked significant progress in our methodological framework for predictive analysis. Despite facing challenges, our exploration of feature engineering and model selection has provided valuable insights that will guide our future steps. As we move forward, we aim to refine our data processing techniques, train and evaluate our models, and construct an informative visualization dashboard.