



ML Project Weekly Report

Course Name: Machine Learning (CSE 523)

Week Number: 7 (24rd March - 30th March 2024)

Group Name: White

Instructor's Name: Mehul S Raval

Project no.: 7

Project 7 Athlete Statistics Visualization and Prediction

Introduction

- Continuing from the progress made in data pre-processing methods last week, our objective for this week is to further enhance the balance within the dataset, particularly between the minority and majority classes.

Weekly Activities

1. Data Balance

- Researched about Synthetic Minority Over-sampling Technique (SMOTE) to address class imbalance.
- Conducted trials for implementing SMOTE in Python.

2. Dashboard Study

- Explored the dashboard provided in the research paper for visual representation of athlete data
- Discussed the layout and functionality of the dashboard.

Challenges Faced

- One significant challenge encountered was with the implementation of SMOTE. While it effectively generates new instances of the minority class, it also introduces noise due to increased probability of class overlap.

Learnings

- Learned from the research paper about the application of Edited Nearest Neighbor (ENN) technique on the oversampled dataset.
- ENN aims to mitigate the noise introduced by SMOTE by removing majority class instances whose neighbors belong to a different class, identified using the K-Nearest Neighbor (KNN) algorithm.

Conclusion

- Throughout this week, we have diligently explored optimal preprocessing methods under the guidance of relevant research papers. As we proceed, we plan to refine our techniques by incorporating alternative preprocessing approaches, ensuring the generation of a clean and informative dataset for our machine learning project.