Date: 9th Mar 2024

**CSE 523: Machine Learning**

**Weekly Report 1**

**Project 7: Athlete Statistics Visualization and Prediction**

**Team: Byte Busters**

|  |  |
| --- | --- |
| **Enrolment No** | **Name** |
| AU2140060 | Deval Darji |
| AU2140085 | Krishn Patel |
| AU2140103 | Devanshi Shah |
| AU2140178 | Vedant Patel |

As mentioned previously, we had been split up into two teams, data team and approach team.   
For this week we tried to work on codes on the approaches finalized.

**Data Team:**

As stated in the previous report, we have decided to go for the KNN method for data imputation, in this method we first calculate the euclidean distance between the missing value instance and all the other instances, then we compare it and identify the nearest neighbors (2,3,4,etc), after that we take their mean/median and replace the missing value.

We decided to go with this method because we wanted the missing data to be as accurate and close to their similar data as well

We then decided to build a code for the same, this is the first draft and is a basic code used for KNN imputation

A screenshot of a computer

Description automatically generated

**Approach Team:**

With the beginning of PCA in the class, we wanted to measure our pros and cons for both PCA and factor analysis.

We finalized on factor analysis as we wanted to reduce the features on the basis of their underlying latent factors as PCA is just a mathematical approach with having the new variables just a linear combination of others and not taking the factors into consideration.

After doing these thorough research and analysis, we proceeded with the seal of approval and then started writing the code for the same.

A screenshot of a very rough draft has been attached as well, work and testing is to be done on the same

A screenshot of a computer

Description automatically generated