



Ahmedabad
University

CSE - 523 Machine Learning

Weekly Report-3

Project title: Athlete performance in collegiate basketball: Predicting match line-up and RSImod

Submitted to faculty: Mehul Raval

Date of Submission: 9-02-2024

Enrolment No.	Student Name	Programme
AU2140031	Priyanshu Thakkar	BTech
AU2140001	Adnan Kadiwala	BTech
AU2140158	Milee Bajpai	BTech
AU2140017	Vandit Shah	BTech

- **Overview of the Problem Statement:**

We looked at the various ways and the research paper to create a data-driven approach that will determine the best possible team's lineup for a collegiate basketball game weekly and also forecast the Relative Strength Index (RSImod) for upcoming weeks. This method is going to use player statistics, team performance metrics, and historical data for coaches, but they going to make more sensible decisions based on the information gathered. This in turn, will boost team performance and increase their chances to win.

- **Literature Survey Summary:**

We did the survey, focusing on a paper titled "A holistic approach to performance prediction in collegiate athletics", where the author explores performance forecasting in collegiate athletics through applying the Multiple Imputation by Chained Equation (MICE) technique, feature importance analysis that is based on Random Forest (RF) models and XGBoost, correlation, and analysis for models development.

Progress Made:

- ✓ Data Imputation

We opted to employ the Multiple Imputation by Chained Equation (MICE) technique to address this issue and prevent any potential bias in prediction outcomes. This technique is renowned for its accuracy in imputing missing values, as it relies on conditional modeling of the missing feature concerning the other available features.

```
[ ] from sklearn import linear_model
%matplotlib inline
from sklearn.model_selection import train_test_split
from sklearn.impute import SimpleImputer
from sklearn.ensemble import RandomForestClassifier
from sklearn import linear_model
from sklearn.experimental import enable_iterative_imputer
from sklearn.impute import IterativeImputer
from sklearn.ensemble import RandomForestRegressor

[ ] imp = IterativeImputer(max_iter=10, random_state=0, estimator=RandomForestRegressor(n_estimators=10, random_state=0))
imp.fit(df)
df=pd.DataFrame(imp.fit_transform(df),columns=df.columns)

/usr/local/lib/python3.10/dist-packages/sklearn/impute/_iterative.py:785: ConvergenceWarning: [IterativeImputer] Early stopping criterion not reached.
```

Data Imputation:

- We have gone through the code of data imputation used in the base paper. Now we are looking for some new methods for data imputation other than the base paper which we can use.

- References:
Taber, C. B., Sharma, S., Raval, M. S., Senbel, S., Keefe, A., Shah, J., Patterson, E., Nolan, J. K., Artan, N. S., & Kaya, T. (2024). A holistic approach to performance prediction in collegiate athletics: player, team, and conference perspectives. *Scientific Reports*, 14(1).
<https://doi.org/10.1038/s41598-024-51658-8>