Ahmedabad University

CSE - 523 Machine Learning

Weekly Report-2

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 Collection and Preprocessing: We got the dataset and now looking into the imputation technique and how we can handle missing values using the MICE technique as stated in the paper and how other techniques can be applied and which could provide a better imputation technique

RSImod Prediction: Looked into the code of the base paper on how they implemented the forecast of RSImod using XGBoost, after analyzing feature importance using Random Forest, XGBoost, and correlation methods

RSI Prediction

```python
from sklearn.ensemble import GradientBoostingClassifier
import xgboost as xgb
from sklearn.metrics import auc, accuracy_score, confusion_matrix, mean_squared_error ,f1_score ,classification_report
from sklearn.model_selection import cross_val_score, GridSearchCV, KFold, RandomizedSearchCV
from sklearn.model_selection import StratifiedKFold
from sklearn.model_selection import RepeatedKFold
from statistics import mean, stdev
skf = RepeatedKFold(n_splits=5, n_repeats=10, random_state=1)
```

```python
target_names = ['0', '1', '2','3']
```

```python
def classifierModelEvalution(x_record,y_record,method=""):
    accuracies = []
    f1_result_score = []
    class_accuracy = []
    for train_index, test_index in skf.split(x_record, y_record):
        x_train_fold, x_test_fold = x_record[train_index], x_record[test_index]
        y_train_fold, y_test_fold = y_record[train_index], y_record[test_index]
        clf = xgb.XGBClassifier(base_score=0.5, booster='gbtree', colsample_bylevel=1,
                colsample_bynode=1, colsample_bytree=0.6, gamma=1,
                learning_rate=0.1, max_delta_step=0, max_depth=5,
```

- Now working and learning on predicting the RSImod using a time series analysis approach using N weeks data for predicting the (N+1)th week. We are trying to utilise XGBoost to forecast RSImod for the subsequent week by aggregating independent feature readings over the window and training on RSImod for the next week.

- 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