

CS322:Big Data

Final Class Project Report

**Project (FPL Analytics / YACS coding): FPL Analytics**  **Date: 01-12-2020**

|  |  |  |  |
| --- | --- | --- | --- |
| SNo | Name | SRN | Class/Section |
| 1 | Archana J | PES1201801306 | I |
| 2 | Swati Naik | PES1201801489 | I |
| 3 | Srujan KN | PES1201801829 | I |
| 4 | Sumukha MK | PES1201801995 | I |

## Introduction

Aim of our project is to analyse various events that happen in a football league (Fantasy Premier League) and come up with necessary results .These events are associated to many metrics such as players’ free kick effectiveness, pass accuracy , duel effectiveness , etc. We use streaming data for the analysis.

## Related work

We referred to:

* PESU ACADEMY
* https://spark.apache.org/

## Design

The matches and the events data were streamed through port 6100, code for which was provided and we used Streaming Spark to read the streamed data. The players and teams data were stored in hdfs and taken as command line arguments while running the code. We calculated all the metrics such as free kick effectiveness, pass accuracy,duel effectiveness, own goal, shots on target , etc for each player using the streamed data. For clustering , we used k-Means to predict the approximate rating and chemistry coefficient of players who have played less than 5 matches. Coming to regression , we used polynomial regression for analysis of the data.

## Results

By extracting the data from the streamed data, we calculated results for the below metrics:

1. Pass Accuracy
2. Duel Effectiveness
3. Free Kick Effectiveness
4. Shots on Target
5. Foul Loss
6. Own Goal

After computing the results for pass accuracy, duel effectiveness , free kick effectiveness , shots on target, we observed that output values were bounded between 0 and 1 for each of the specified metrics.

At the end of every match,we even computed player contribution, player rating, player performance , chemistry for each player. Further we even built player profiles which consists of all the necessary attributes of a player along with number of own goals,fouls ,pass accuracy and shots on target.We made use of the machine learning models like Kmeans for the clustering where it has been done based on the profile of the player and clusterEvaluation to evaluate the model.

## Problems

Main problem faced initially was to extract data from the streamed data for further analysis. Since we are not reading from a stored dataset, which is simpler, extracting data values for metric calculation was a challenge. Proper usage of mllib for clustering and prediction also involved decent effort.

## Conclusion

Real Time Analysis of streaming data is one of the main learning aspect in this project . Practical use of RDDs and application of Mllib are also few key take-aways.

## EVALUATIONS:

|  |  |  |  |
| --- | --- | --- | --- |
| SNo | Name | SRN | Contribution (Individual) |
| 1 | ARCHANA J | PES1201801306 |  |
| 2 | SWATI NAIK | PES1201801489 |  |
| 3 | SRUJAN KN | PES120181829 |  |
| 4 | SUMUKHA MK | PES1201801995 |  |

## (Leave this for the faculty)

|  |  |  |  |
| --- | --- | --- | --- |
| Date | Evaluator | Comments | Score |
|  |  |  |  |

## CHECKLIST:

|  |  |  |
| --- | --- | --- |
| SNo | Item | Status |
| 1. | Source code documented |  |
| 2. | Source code uploaded to GitHub – (access link for the same, to be added in status 🡪) |  |
| 3. | Instructions for building and running the code. Your code must be usable out of the box. |  |