Project MoneyBall

Nikhil Raikar
Vamshi Bhushanabiona
Cheryl M Jacob
Pooja Ajit

Project Overview

Money is a very important aspect in almost every professional sport. For example, many smaller city market teams, must spend their money wisely to ensure the best outcome; whereas, a larger city market team has more income that is expendable. This money spending process originates during the Major League Baseball player draft held each June. The draft process involves fifty rounds of selections by all thirty teams. Each team gathers their general managers, scouts, and professional consultants to decide which players should be drafted. The higher the draftee the more valuable he is believed to the team. Therefore, the procedure to decide which players should be selected earliest is very important.

Requirements

Dataset	**http://seanlahman.com/baseball-archive/statistics**
Tools	SAP Hana Database Sap Predictive Analysis Python (For statistical computing) Javascript,HTML SAPUI5 (UI Framework)
Scrum	GitLab- https://gitlab.com/nikhilraikar88/Project_MoneyBall

Goals:

With the help of Sap Hana and Sap predictive analysis our goal with this project has been to use the data mining methods and techniques to predict and analyse the ability,skill and commitment level of a baseball player.

We implement various predictive analysis methods on the dataset by the following mining methods

- Classification:

  We use classification as one of the predictive methods to classify the data. One such reason being we can make a subset of categories by iterating through the data which will inturn give us a better picture with the player statistics being concerned.

- Regression: This can be used both as descriptive and predictive analysis. We use it as one such predictive methods because it models past relationship between the attributes of the dataset and predicts the future which in turn can help us get some good results out of our dataset.

- Clustering :

  We use clustering also as one of our predictive analysis so that we can use clustering for grouping of a particular set of objects based on their characteristics, aggregating them according to their similarities.

- Correlation:

  We use correlation of the data as one of our descriptive analysis methods.We mainly focus on finding the depending attributes or values which can help us further in predictive analysis.

Milestones

- Milestone 1
  - Project Plan
  - Data cleaning ,Reduce data complexity
  - Data Transformation
  - Preliminary analysis
  - Schema for the data has to be created.
  - Methods of predictive analysis has to be researched.
- Milestone 2
  - loading of data into sap hana.
  - Functional and Technical Design Documents.
  - Implementation of the Algorithm(method) 1 of predictive analysis.
- Milestone 3
  - Analysis of the result.
  - Error Handling for Algorithm 1.
  - Research on Algorithm 2 of predictive analysis.
- Milestone 4
  - Implementation of Algorithm 2 of predictive analysis.
  - Analysis of the result obtained.
  - Error Handling for Algorithm 2
- Milestone 5
  - Comparison of the two algorithms.
  - Errors,issues faced and resolved.

- Milestone 6

  - Documentation of the tasks.
  - How the system can be improved.
  - Future work and references.

Project Workflow:

Milestone 1	Data Cleaning and transformation Due Date: 22 Dec 2018
Milestone 2	Implementation Method 1
Milestone 3	Analysis and Further implementation
Milestone 4	Implementation Method 2
Milestone 5	Business Processes
Milestone 6	Improvement and Further additions