

CS513A-KDD PROJECT PROPOSAL: NBA Career Longevity

Project Group No: 14

Problem Statement:

In North America, there is a professional basketball league called the National Basketball Association (NBA). One of the four main professional sports leagues in the United States and Canada, the league has 30 teams (29 in the United States and 1 in Canada). The top men's professional basketball league in the world is this one.

For every athlete in any sport, including NBA rookies, career length is determined by a variety of factors. Factors such as games played, game count, and other data of the player throughout the game.

Given a set of features, we will predict if the player will be able to last 5 years in the league or not.

Dataset:

There are 19 characteristics in the dataset and 1 target variable. The target variable can have two values: 0 or 1. 0 indicates that the player did not endure five years, and 1 indicates that the player lasted five years or longer. Almost all of the features will be used during implementation.

Source of Dataset: <https://www.kaggle.com/competitions/iust-nba-rookies/data>

Implementation Strategy and Algorithms Used:

We've chosen to nine distinct models into action and compare them across four different group members. We picked a few models from our course as well as a few from outside.

The models that we chose are as follows:

1. K - Nearest Neighbour
2. Logistic Regression
3. Gaussian Naive Bayes
4. Decision Tree
5. Support Vector Machine
6. Random Forest
7. Random Forest with Grid Search
8. Xgboost
9. Xgboost with Grid Search

Model metrics and Evaluation: We will be using Confusion matrix, F1, recall, precision to measure the accuracy of our models.

Team Members Names: Group 14

1. Akshay Pradeep Patade
2. Deep Deven Manek
3. Feneel Doshi
4. Kevin Christie Dsa