Project Proposal
(Tennis Players)

Abstract

The Goal of this project it to use classification models to cover the basic Jupyter EDA and Visualizations. the basic goal of the analysis is to obviously looking at the players which cannot actually be ignored if we talk about this sport!

“Must Won”

“Must Lose”

"Who has played the most"

"Who has played the least"

and to see how amazingly well they have playing in the sport.

Design

In this project and based on data from ATP, I try to answer the question statistically.

- How many players have played the most matches?
- How many won?
- How much did he lose?
- How many set finished the match?
- Other questions

Data Description

In these datasets there are individual csv files for ATP tournament from 2015 to 2017 contains 6,351 rows and 50 columns.

Dataset legend

All the match statistics are in absolute number format, you can convert to percentages using the total point number

```
ace = absolute number of aces
df = number of double faults
svpt = total serve points
1stin = 1st serve in
1st won = points won on 1st serve
2ndwon = points won on 2nd serve
SvGms = serve games
bpSaved = break point saved
bpFaced = break point faced
```

Algorithms

Classification model: to predict logistic regression.

Tools

- Numpy and Pandas for data manipulation
- Scikit-learn for modeling
- Matplotlib and Seaborn for plotting
- Tableau for interactive visualizations

MVP

The goal of this project is to better understand who’s the player has been steadily improving over the 3 years.