Project Description

Problem Definition: The aim of our project is to develop a tool which can be used to view and analyze Tennis Data for the past few years. The data set contains information regarding all the matches taking place in ATP events all year round in detail along with the betting odds provided by different betting sites in the form of an excel sheet. We provide an interface to deduct the following things from the data set:

1. Tournament Analysis

* Exciting Matches
* Upsets at Each Tournament.

1. Player Odds

* Analysis of All players based on chosen betting site
* Find the odds data for a single player
* Top ten players based on profitability for that year

Software Description: We have written a Java program to access the excel file and establish the required analysis based on certain well defined conditions. We define the following classifications as follows:

1. Exciting Matches: We define the closeness of a match by whether the match went to the deciding set. The matches that went to the deciding set in general are more exciting than those which were ended in straight sets. Analysis was done to find such matches from the data set.
2. Upsets at Each Tournament: An upset can be loosely be described as when a player is beaten by someone who is much below him in ATP rankings. We have set a threshold that we deemed fit for being an upset and analysis was done.
3. Player Odds: The analysis of betting odds of each player gives us an idea of which player to bet on and consistency of each player over that calendar year.

* For all players we calculate the profit factor by the following formula.
* For victory we add the odds for that match based on the chosen site.
* For loss we subtract one as we assume that we bet one dollar on each match.

1. Based on this formula we calculate the profit factor for players based on user’s choice. It can be all players, a particular player or Top ten players for that year.
2. While entering the player name, the user need not enter the exact name. Even a part of the name can be entered, we give suggestions of the possible players who played in the calendar year with a name similar to that.

Results:

We have produced results in data format as well as in graphical format. In graphical format results have been produced for exciting tournaments and tournaments with upsets. All the graphs have been included in the zip file which can be reproduced after running the code. Once graphical format has been chosen as an output the program automatically produces an image of the graph in the directory of the project. User has an option to choose between graphical and data analysis.

Reproducibility:

Steps:

1. Import the project into an IDE such as eclipse
2. Add dependencies into the build path
3. Run the code like any other Java Application
4. Enter inputs as and when the code asks for it.

We have included all the dependencies such as Apache POI and JFreeChart in the zip file.

Following is the list of technologies used:

1. Apache POI
2. JDK 8 (Anything less then that is incompatible)
3. JAVA SE 1.8 (Anything less then that is incompatible)
4. JFreeChart
5. Tested on MAC OSX and Windows
6. Dataset is in the form of excel file

Dependencies have to be included in the build path of the project after importing it. Dependencies are in .jar format and are included in the zip file. They have to be added externally.

Datasets used are ATP Tennis data sets and covers span of 3 years and are in excel format.