

Assignment 2: Path to Victory (US Open Men - Game Predictor)

To run: Open Assign2-YatharthSharma.html, Please use Firefox, leave the json file in the same directory

Data Analysis:

- 1) Handling missing data: The missing data was completed by using the mean value of player's attribute (For example: to fill FastServe for roger federer, mean of all other roger federer matches was taken).
- 2) Some basic data analysis was done on excel/R by creating charts and using summary function in R to look for interesting pattern. But simple pattern were found
- 3) Since, this time we had a lot of data, I placed more focus of analysing data in order to predict the outcome based on player's stats.

Data Mining approach:

- 1) A **decision tree classification model** was applied for every player.
- 2) Output were 2 classes - win or lost (this was done for individual player)
- 3) Input was selected based on the above data analysis result and intuition of what the user would want to select the attribute as input.
- 4) Players who never won or lost were removed from the data set.
- 5) The decision tree was created for each a every player with 2 classes and was saved in <player_name>.json file (It created 350 json file, I tried to merge them but the file size cause delay in the visualization).
- 6) **Main Idea** : Predicted the outcome given player's past records.

Visualization Design:

The best way to visualization a decision tree classifier was to create an **interactive tree** itself which would map the path to victory/loss based on the input selected.

- The node of the tree (d3 circles) represents the splitting attribute at that level.
- The leaf of the tree (d3 path) represents whether the player will win or lose.
- Input was given with the help of a d3 slider and the player was selected using drop down menu.

Interactions and how to reach the result:

- The first interaction starts by user selecting a player [by default a player has been given]
- The selection of a player generates/regenerated the decision tree with the non leaf node colored according to the input parameters and the leaf node colored as {green: win; red:lost}
- User can provide the input using the slider presented above the tree.
- Sliding the input parameters caused 2 things [Extra focus was given here so that this interaction is simple]
 - Enlarges the nodes of the tree and colors the path that tree follows according to the input provided, thereby displaying if the player won provided that input.
 - Displaying the current value of the input above.
- Selecting a different player from the drop down menu created a different tree for the new player.
- A reset button was also provided to reset the input to their initial values.

Aim of the visualization: What if scenarios

Path to Victory is an interactive prediction tool with strong data analysis as its base (decision tree classifier). This tool can be used by the users to evaluate/predict the output of the match based on player's current statistics and can be used by the player/coach to access what input parameters should be focussed more in order to win the game.

My visualization uses user-driven narrative to trace out the path of victory or loss depending upon the values selected.

The main aim of this visualization was to provide users with a way by which they can model what-if scenarios as in: "How can Nadal win a match even if he has less than 40% firstServe wins",

"Given Dokovic's current input parameters, what can he do to win the match"

Moreover, by looking at the tree nodes, it can be easily seen which attributes affects a player the most.