

# Project 1 Proposal

## **What is the problem you want to solve?**

The problem I am trying to solve is to answer tennis related questions. The tennis betting website (<https://www.goranivanisevic.com/tennis-betting>) provides affecting and 10 tips for tennis betting. One question is how tight were the matches the player had? The website suggests that if the players had long, tight matches, we need to take into account the possible fatigue and lack of concentration they may experience in their next match. Another affecting factor or question is: do left hand players have better chance to win? I am going to leverage several stats and probability methods at the beginning, and then apply the machine learning algorithm(s) will be learnt in unit 10.

## **Who is your client and why do they care about this problem? In other words, what will your client do or decide based on your analysis that they wouldn't have done otherwise?**

My client are tennis fans, and/or any individuals who have interest in knowing tennis match factors and possibly take the analysis results into consideration when they try to predict the game results.

## **What data are you using? How will you acquire the data?**

The data used for this project comes from Github. I will be manually downloading the file, extract the files, and then load, parse, and clean the data in python. The files I will use is the subset of this download which are total 20 ATP match data files from year 2000 to 2019.

Site: [https://github.com/JeffSackmann/tennis\\_atp](https://github.com/JeffSackmann/tennis_atp)

Files: atp\_matches\_2000.csv to atp\_matches\_2019.csv

## **Briefly outline how you'll solve this problem. Your approach may change later, but this is a good first step to get you thinking about a method and solution.**

I will first download all file. I will then do Data Wrangling, Data Story, EDA, and test different machine learning algorithms.

## **What are your deliverables? Typically, this includes code, a paper, or a slide deck.**

My deliverables will be a jupyter notebook and slide deck that will be published to my Github account. I will include all my findings and all documented python code.