Predicting Tennis Match Outcomes with First Set Data

Author: Jason Zhang
Cell/Text: (860)-806-5148
Email: zhang.jas@northeastern.edu

Timeline: October 2021 - December 2021

Goal:

The goal of this project is to predict the outcome of tennis matches based solely on data from the first set of the match. Upon completion, this project can have applications in a variety of situations such as providing an informational advantage to those betting on matches, for players and their coaches to determine which improvements to their games can be made to maximize win potential against opponents, and much more.

Data science can be employed to solve this problem through its branches of machine learning classification. By processing the raw csv data into engineered features, the cleaned and normalized data can then be input into machine learning algorithms. Due to the numerical and structured nature of tennis's scoring system, the quantified data that can be produced is particularly suited for working with machine learning algorithms.[1]

The Problem:

The outcomes of tennis matches are notoriously difficult to predict, due to the volatile nature of the sport: changes in momentum, effects from the audience, and a variety of other factors all contribute to its unpredictability. In this project, I will be tackling this historical challenge, by generating an algorithm that predicts the outcome of tennis matches solely based upon play-by-play data from the first set. To do this, I plan on studying and normalizing tennis match data to discover particular features, and to engineer features of my own, which will be fed into a machine learning algorithm to have the match outcomes predicted. I will be looking to discover which variables are most indicative of the match outcomes, and to attempt to create new variables from the existing ones which will further aid the machine learning models. I also plan on testing the data with a variety of different models to determine which one produces the most accurate results.

[1] https://en.wikipedia.org/wiki/Tennis_scoring_system