

Hey!

My name is Sari Farah and this is my capstone project submission for the Data Science course at BrainStation.

The focus of this project was to predict the result of a shot in a football match. Did the shot result in a goal or not?

The motivation behind my question was "What defines a good shot location in a football match?" As a data scientist, this was an amazing opportunity to use the data I found to come up with an answer using machine learning techniques.

The dataset used for modeling was scraped from understat.com. It is a football statistics website used to follow the top 5 leagues (English premier league, Spanish La Liga, Italian Serie A, German Bundesliga, and French League 1). The data I collected spanned from the 2015/2016 – 2019/2020 seasons.

There are 3 notebooks uploaded for this project:

1. Datal Loading
2. EDA (Data Cleaning, Feature engineering, Data Visualization)
3. Modelling and Evaluation
4. Appendix

Additionally, the 'data' folder contains the files required to run these notebooks. The data is divided into two csv files, one for the original raw data and the other for the cleaned data used for the modelling step.

If you have any trouble accessing any of the files within this folder, please contact me at sarinrfarah@gmail.com.

To upload the data directly from the web scraper please follow the instructions [here](#).