

# Capstone II Project Proposal

---

## Problem Statement(s):

1. How can we build an algorithm/model that will output the result of a match (win/loss) using individual player performance features while maintaining a minimum of 90% accuracy?
2. How can we build an algorithm/model that will output which maps will be selected for match play (b01 → map1, b03 → map1, map2, map3) between any two teams using map picking trend features while maintaining a minimum of 90% accuracy?
3. How can we build an algorithm/model that will output the country any player is from using individual player performance features while maintaining a minimum of 90% accuracy?

## Context:

Counter-Strike (CS) is an online competitive first-person shooter (FPS) video game series that had its first iteration of the game released in 2000. It didn't take long for the game to gain immense popularity worldwide, and the franchise played a large role in establishing, developing, and growing the Esports industry over the past 22 years. CSGO (Counter-Strike Global Offensive) is the franchise's latest iteration of the game and was released in 2012. It is largely considered to be one of the titans (if not the titan) of the Esports industry and there are no signs of this changing any time soon.

Professional CSGO teams can benefit from deep-diving into data just like all professional sports teams can and do (NFL, NBA, MLB, etc.). We can build useful models that help predict how matches will go and what implications certain features of match data have on different elements of the game. This can vary from predicting match winners using individual player performance to predicting which maps will be selected to play given trends in map vetoes and picks by specific teams. There are many different elements in the data to look closer at to derive useful information that can be turned into actionable recommendations that will help a team perform at a higher level.

## Criteria for Success:

1. Correctly identifying the match winner (*match\_winner* in results.csv) at least 90% of the time upon entering the player's name in the model.
2. Correctly identify the map(s) selected to be played (for b01 → *left\_over* in picks.csv and for b03 → *t1\_picked*, *t2\_picked*, *left\_over* in picks.csv) at least 90% of the time upon entering the two teams playing each other in any given match in the model.
3. Correctly identifying the country that any given player is from (*country* in players.csv) at least 90% of the time upon entering individual player statistics such as kills, deaths, adr, etc. (whatever features end up being most useful in making accurate predictions).

## Scope of Solution Space:

The focus will be put on building the most accurate models possible so that any predictive insights produced are reliable and useful.

The data we will use contains professional match data from November 2015 to March 2020.

## Constraints:

If our best model cannot support 90% accuracy feasibly we will have to adjust our goals to align with what our data realistically tells us.

## Stakeholders:

Primary: CSGO team owners, CSGO team coaches, CSGO team players  
Secondary: CSGO match betters (gambling)

**Data Sources:**

Data source link: <https://www.kaggle.com/datasets/mateusdmachado/csgo-professional-matches>

All data was scraped from HLTV.org, the leading CSGO site in the world that contains all match data from the many years of competing at the professional level. The data consists of 4 separate tables all sharing a column *match\_id*: *economy.csv*, *picks.csv*, *players.csv*, and *results.csv*. We will most likely use all of the tables except *economy.csv*, unless later deemed useful.

**Deliverables:**

1. Notebooks for all of the steps throughout the project.
2. A slide deck going over the process and key findings.
3. A project report going over the process and key findings.