

# 1 Description

When playing online video games, the identity of the player behind the keyboard is often unknown, even if their account name in game remains the same. This can lead to issues such as account selling and MMR (Matchmaking rating) boosting, where an account's MMR is raised for a price or the account is sold to other players. This project looks specifically at the problem of player prediction in the game DOTA 2, a popular, free multiplayer online video game by Valve Corporation. Replays of all DOTA games are available publicly through Valve's Steam API, which allows full replay `.dem` files to be downloaded. Community tools have also been built around these replays to allow for data gathering.

The goal of this project is to use machine learning models to predict and identify the player behind the keyboard, based on their in-game actions such as mouse movement, item and hero selection. Replays of the same player will be parsed in order to extract data for the models and the different features used for player prediction evaluated both independently and combined to determine which feature or combination of features are most effective in predicting the player in DOTA.

Extensions to this project include...

Player identification is useful for detecting players using alternate accounts - tracking statistics for pro players - detect previously banned players - anonymity in online games leading to antisocial behaviour due to lack of consequences > easily make new accounts to continue behaviour despite bans, especially in free to play games - smurfing ranked games affecting players at lower skill level - elo boosting and selling high ranked accounts (hard to detect)

- use case of 'stolen' accounts, where player behaviour suddenly changes to be very different - 'identity theft' in online amateur tournaments

Being able to predict who the player is without game knowledge allows this approach to be applicable to other games.

Additionally, the ability to identify a player can lead to further prediction of their future actions based on their play styles and strategies, which can help build sophisticated game AI.

The goal is to first predict using different features independently, then combine them - ideally less game specific knowledge in prediction means those features could carry over to other games -

Specific player on a specific hero - what one player may do different compared to another player on the same hero

## 2 Objectives

### 2.1 Primary objectives

#### 2.1.1 Parse DOTA 2 replays to generate data such as mouse movements, gameplay statistics and strategies of the player

Replay files come in the form of binary `.dem` files, which are meant to be replayed directly using the game. Community tools have been developed to parse replays, but only acts as a library for developers to access all the statistics and events that happened in a game. Further processing must be done with these tools to extract the exact data needed for the machine learning models.

**2.1.2 Binary classification of a single player using different features independently on the same hero**

**2.1.3 Multiclass classification of players**

**2.2 Secondary objectives**

**2.2.1 Identification of patterns/strategies from a player's replays**

**2.2.2 Prediction of player actions**

**3 Ethics**

**4 Resources**

This project can be completed using equipment already provided by the school and therefore does not require any special resources.