PUBG Finish Placement Prediction

CSI 4106 - AI Project
November 6, 2018
Professor: Dr. Barrière
Project Coordinator: Yazan Otoum
Group number: 31 (English section), 23 (French Section)

Group members:
 English Section: Patrick Langis - 8196917 & Felix Singerman - 7970742
 French Section : Diego Pontones - 8281209
Project type: 6. Participation in an international evaluation

Background:

We decided to go with topic 6, *Participation in an international evaluation* using Kaggle. Kaggle, as a data science competition platform, is a great way to train and challenge data scientists internationally whilst contributing to real-life problems. We chose to use the Player Unknown's Battlegrounds (PUBG) finish placement prediction competition. PUBG is a battle-royale style video game for PC, Xbox One and Mobile devices. "100 players are dropped onto an island empty-handed and must explore, scavenge, and eliminate other players until only one is left standing, all while the play zone continues to shrink" (Kaggle, 2018). PUBG, together with Kaggle has given over 65,000 games worth of players data, and we are tasked to predict the final placement from in-game stats and initial player ratings. We will discover some of the best strategies in order to win in PUBG and what makes players successful based on the data given. This is a fairly standard task as a data scientist and we will be utilizing several AI techniques such as machine learning.

We believe that having a goal, and data to work with, we can perform some meaningful research on a purposeful topic. To have a good structure for artificial intelligence research, it is essential to have a clean set of data. We will be applying several of the topics we learned in our lectures such as machine learning (Regression, SVM, Decision trees) and possibly neural networks.

Proposed Work:

The work will be evenly divided amongst the group members. We will each have a part in coding and writing the report. As Diego is in the French section and Patrick and Felix are in the English section, two reports will be presented, one for each section.

We will be following the OSEMN pipeline for the coding portion of the task:

- O - Obtaining the data
  - This is mostly already completed as Kaggle provides the dataset. We will just need to download the dataset and sync it with our notebook.
- S - Scrubbing/cleaning the data
  - We estimate that this will take up at least 50% of our time as data tends to be messy and we would like to do some feature engineering.
- E - Exploring and Visualizing the data
  - We will show details and gain insights on the data using matplotlib, Seaborn, and other data visualization tools. We predict that this will take 15% of our time.
- M - Modelling the data
  - We predict this will take up close to 30% of our time as we try to build and optimize our models to predict the data. We will try several different models such as SVM, Regression etc.
- N - Interpreting the data
  - We will explain and demonstrate our results to the reader. We predict this will take 5% of our time.

We will use the common data science libraries provided in Python, many of which we have seen in class such as Pandas, scikit-learn, numpy, etc. We will submit the code as a Jupyter Notebook and use Google Colab to collaboratively work on the notebook. We will also use GitHub for version control and in order to host the code for any reader to be able to run and test the code.

Once we have completed the coding portion of the task, we will then collaboratively write a report outlining our project, its relation to AI, our algorithm, it's implementation with instruction on how to run the code, and a result analysis.

References

1. Kaggle, PUBG Finish Placement Prediction. (2018, October 4). Retrieved November 5, 2018, from https://www.kaggle.com/c/pubg-finish-placement-prediction