Dota 2 Performance and Burnout Tracker

GitHub Repo

1. Project Definition

1.1 Problem Statement

Dota 2 has an esports scene like most popular games do and most players are concerned over their stats such as kills to death ratio, matches won, and win streaks. However, the Dota API does not have a burnout feature and many players are not aware of how helpful this feature can be for both their performance and behavioral health. Burnout is a difficult feature to quantify, but my model aims to visualize burnout patterns by detecting unusual dips in performance by analyzing kill to death average data and other anomaly detection. If more modern APIs could implement this feature and create an indicator of whether or not a player is fatigued, it can help players solidify better behavioral patterns and understand which setup works best for them.

1.2 Connection to Course Material

This project relates to the course material because it utilizes many methods of data management and simulates data while structuring it in a SQL database. The Dota 2 model has a main program that calls the source files and runs analytics via imported Python libraries like Numpy, Pandas, Matplotlib, and scikit-learn. Since Dota API and Kaggle had many players that had identical performances I created an option where the user can simulate data log and populate the relational schema using SQLite. In order to determine burnout my model evaluated the player's performances using R2 scores, MAE, and z-score.

2. Novelty and Importance

2.1 Importance of the Project

Burnout from games is something that is not talked about more in the esports scene and is largely contributed to major game companies not creating a feature in their public APIs. Players with unhealthy behavioral patterns may not notice it themselves and continue to play and wonder why they are losing. A potential reason why no gaming company has created a feature like this may simply be because it is hard to quantify fatigue in a virtual activity. Considering esports is a growing profession, creating a burnout feature for popular games like Dota 2 is significant.

2.2 Excitement and Relevance

This model is exciting for me because gaming is personal hobby of mine. As technology is getting more advanced, the audience for competitive esports has also grown. With this in mind, more people are becoming competitive esports players and it made me question the health factor of these players. Sitting behind a monitor all day playing a game can create unhealthy behaviors if not addressed and acknowledged.

By building this model, it became a way to bridge gaming and data science and develop a model that analyze performance and determine when a player is going through a burnout based on anomalies and unusual patterns.

2.3 Review of Related Work

It seems that burnout is a feature mentioned in physical sports, but it is almost never talked about in the gaming scene. Although traditional sports are more physically demanding and can wear down the body easily, doing any activity for long periods of time can lead to damaging behavior and health. Unfortunately, gaming data is entirely consisted of kill to death ratio, win to loss ratio, MMR, and rankings. My project tries to expand these stats to not just virtual performance, but also the player's health.

3. Progress and Contribution

3.1 Data Utilization

My project gives the user two options. The user can type 1 to use simulated player data or 2 to pull real Dota 2 player logs from a Kaggle dataset. I initially pulled data from the official Dota 2 public API, but I noticed that the player performances were too identical for my model to detect any burnout.

Specifically, my model pulled players.csv and match.csv from the Kaggle dataset. I noticed that some of the players took longer than others to experience the burnout.

3.2 Models, Techniques, and Algorithms

I used z-score anomaly to make my model detect burnout. I initially thought to use a burnout range of anything less than 1.5 games lost, but I realized after running the model numerous times that it was too large of a range for the model to detect any true burnouts. I changed it to < -0.5 and that is when the model started to pick up true burnouts in the simulated database.

In addition, my model uses linear regression model by comparing timestamp versus the kill to death average per Dota 2 player. This was how my model was able to calculate the R2 score and MAE for the player's performances. Additionally, I chose 5 random matches out of 75 for the player's burnout to occur.

burnout_matches = random.sample(range(10, num_matches),5)

3.3 Experimental Design

I initially thought I could implement 3 options to my model offering simulated data, Dota Open API, and the kaggle dataset. However, after trying to pull data from the Dota Open API I noticed that there was either authentication problems or the player logs were missing crucial information form my model to detect a burnout. Therefore, my initial experimental design didn't quite work out, but I pivoted to just simulated data and the public kaggle dataset which had all the information I needed for the player logs.

```
Choose the data source you want to use for Dota 2 player logs:
1) Simulated match data using python script
2) Kaggle dataset - may take a few seconds
```

3.4 Key Results / Evaluation

Simulated Data:

I. Detected Burnout Flags: 1324 out of 3750

II. R2: 0.00113

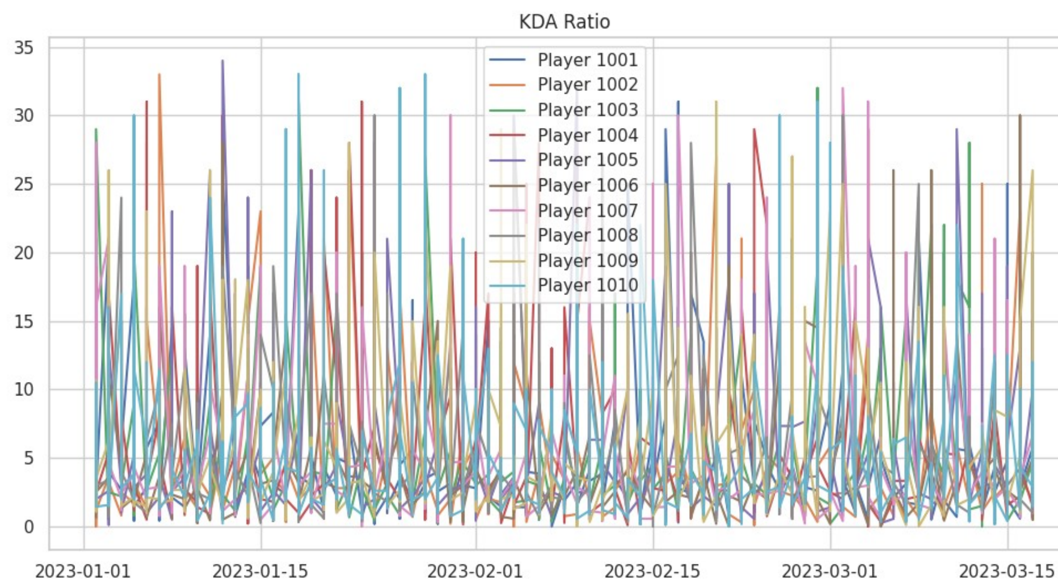
III. MAE: 3.70105

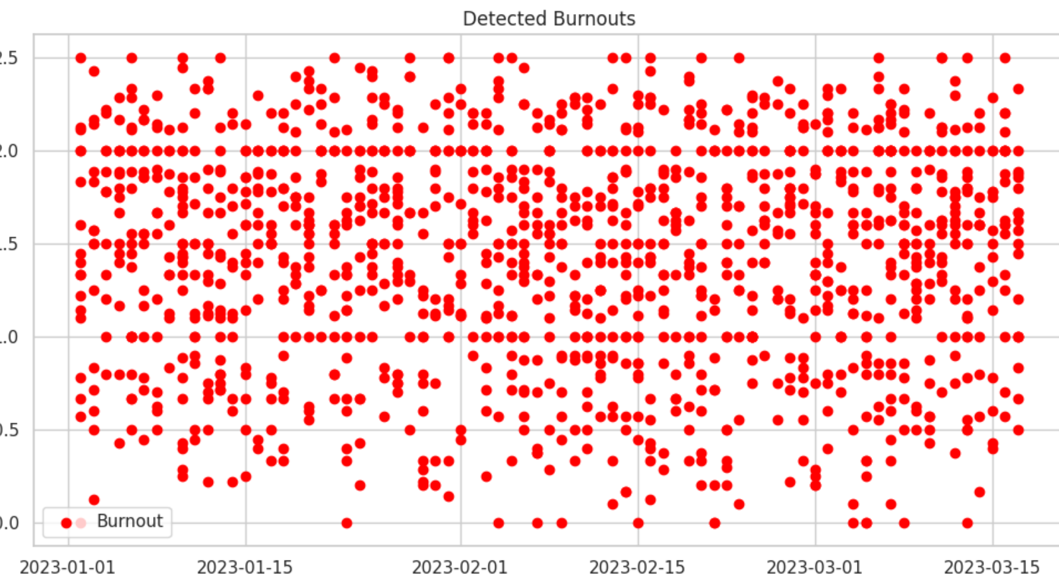
IV. Precision: 0.13369

V. Recall: 0.708

VI. F1 Score: 0.22490

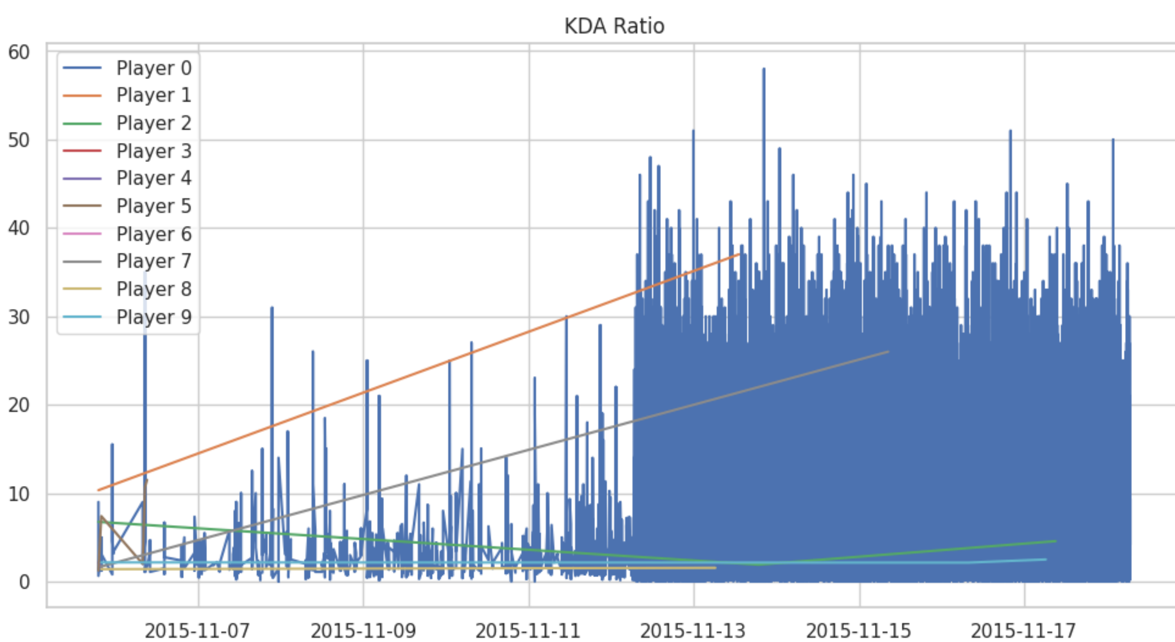
```
Choose the data source you want to use for Dota 2 player logs:
1) Simulated match data using python script
2) Kaggle dataset - may take a few seconds
Enter 1 or 2: 1
Using simulated Dota 2 data
Finished simulating Dota 2 match data for all players. Let's analyze now.
Database setup has been completed
Preprocessed and cleaned data. The CSV has been saved.
Running model
R2: 0.0011317867542881554 MAE: 3.7010503915481237
Precision: 0.1336858006042296
Recall: 0.708
F1: 0.22490470139771285
Burnout flags: 1324 out of 3750
```

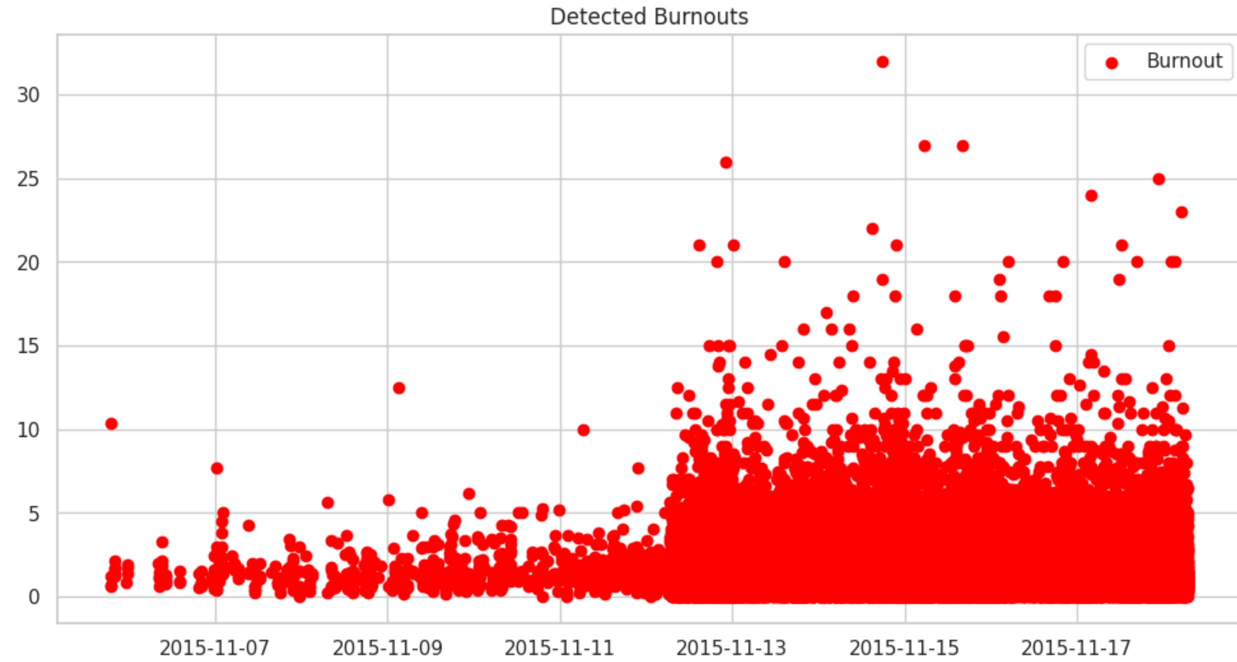




Kaggle Dataset:

```
R2: 1.0883956366392056e-05 MAE: 2.551620695330429
No ground truth
```





3.6 Advantages and Limitations

Advantages of my model include offline usage. The use of SQLite also helps the user see the database management step by step.

Some limitations to my model is that it relies a little on simulated data so if you want entirely real data, it can lack in that department. My model also doesn't have specific hero data for each player log. This could further explain what is going on with a player's performance and explain if they are using a character too much.

4. Changes After Proposal

4.1 Differences from Proposal

Originally, the project aimed to incorporate OpenDota API, but after failed burnout detections I pivoted to just using simulated data and data from kaggle. I also simplified from features such as binge playing and the length of how long a player played during the day. My model more so focuses on KDA ratios and timestamps for its visuals. The model mainly analyzed kills, deaths, and assists from the datasets.

5. Conclusion and Future Work

5.1 Summary of Contributions

I pulled real Dota 2 data from the kaggle dataset, cleaned the data, and used z-score to detect any burnouts. I evaluated the burnout predictions with true burnouts using MAE, R2 score, precision, and recall. I also created visualizations for the burnouts, giving the use a .png of both the KDA over time and burnouts.

5.2 Future Directions

In the future I would like to actually pull data from the OpenDota API. I couldn't do it this time around because it required me to handle authentication tokens from Dota, hourly limits of matches, and some missing information of the matches. Lastly, I would like to add hero statistics for each Dota player so it could be more specific and tell if a player is experiencing fatigue from playing the game or just a specific character.