**Steam Popularity Prediction Report**

By Spencer Sexton, Raunak More, and Sam Brickett

**Abstract:**

Any person who considers themselves a “PC gamer,” has likely heard of the Steam platform. It is owned by the multi-billion-dollar cooperation Valve and is the premiere distributor of digital games. However, for someone who is new to the platform or simply wants to play a new game, what is the best way for them to find a popular game relevant to them? This is where machine learning analysis comes in. The predictive model we created is capable of taking in tens of thousands of games on Steam and using them to predict what upcoming games will be popular.

**Introduction:**

Steam and its developers, Valve, both had very humble beginnings. It started in 1996, when two Microsoft employees Gabe Newell and Mike Harrington decided to start their own company. They both had experience working with games before, due to working on a port of Doom in 1995 for Microsoft. Thus, they founded Valve in Kirkland, Washington, and promptly began buying smaller developers and their companies. Together, they published titles such as Team Fortress and Counter-Strike, which were originally just mods of Quake and Source engine respectively. Mike Harrington left the company in 2000, due to his newly gained fortune from these titles.

In 2002, Valve introduced Steam, a software platform intended to allow customers to purchase, download, and update their games. It quickly grew as Valve began to allow other developers to list their games on Steam. Today, Steam offers over ten thousand games from publishers all around the world. It also offers chat features, support for controllers, and multiplayer functionality across many different platforms. It has become synonymous with online gaming due to its massive market share in digital game retail.

With such a wide array of games on Steam, Valve needed to introduce an effective way to sort through and organize its catalog. As such, they introduced the concept of tags. Whenever a publisher lists their game on Steam, they have the option to add any number of tags to their game. Any person who views the game on the Steam store also has the option to add any tags they think apply to the game as well. These tags include things such as: adventure, action, singleplayer, fantasy, and more. They encompass both genres the game is part of and features the game offers. Each game will then show the tags that are most often associated with that game. With this system, it becomes much easier for prospective players to find games that suit their interests.

Another feature that Steam offers its player base is user reviews. Anybody who owns a game is allowed to leave a review on that game. They choose to give it a thumbs up or a thumbs down and can write anything they want associated with that review. This allows Steam users who don’t own a specific game to see if other players like it before they buy. Steam will also calculate the total percentage of positive reviews to negative reviews and display it as a percentage.

**Motivation:**

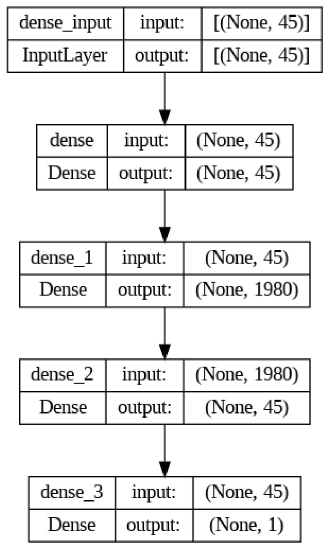
With so many different games on Steam and many more coming each month, we wondered what would be an effective way to predict the popularity of a game. We determined that the user review percentage is a good metric to evaluate whether a game is liked or not. However, we want to predict a game’s popularity before it even comes out, before other users get a chance to review the game. This is where our predictive analysis comes into play.

If we have the tags associated with the game, then we have a good idea of what the game is about. So, if adventure games are generally popular, it’s more likely that a game coming out with the “adventure” tag will also be more popular.

As researchers in this topic, it’s very interesting to see which types of games are more popular than others. It’s possible that certain tag combinations simply lead to better games than others, and we would like to determine what those combinations are. Not only will this assist in the future development of games, but it will also guide us on what types of games we should be looking at.

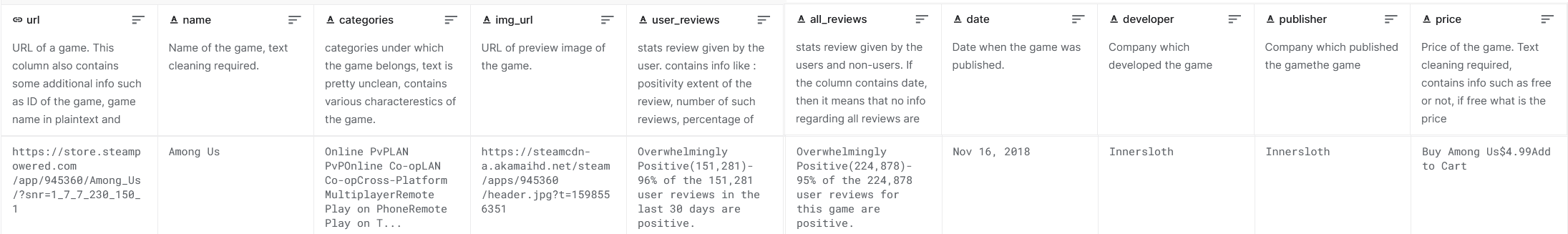
**Proposed Methodology:**

Our model will take in the tags from a large database of eighty thousand Steam games. It will then use the average user reviews on those games to analyze whether a game is popular or not. Using this data, we can then predict the average user reviews of any future game based upon that game’s tags.



*Figure 1: Structure of the machine learning model used in the analysis*

Most of the data in every datapoint (as shown in Figure 2) was filtered out, leaving behind only the categories and user\_reviews columns. Both of these columns were then preprocessed to remove the extraneous text, leaving behind a list of tags and a rating percentage (expressed as a decimal) for each game. After extracting tags for all the games used from the dataset and filtering out tags considered unrelated to the popularity of a game (e.g. EULA, or End User License Agreement), there were 45 unique tags for our vocabulary vector. We then generator input vectors for each game to be input into the 45 input neurons.

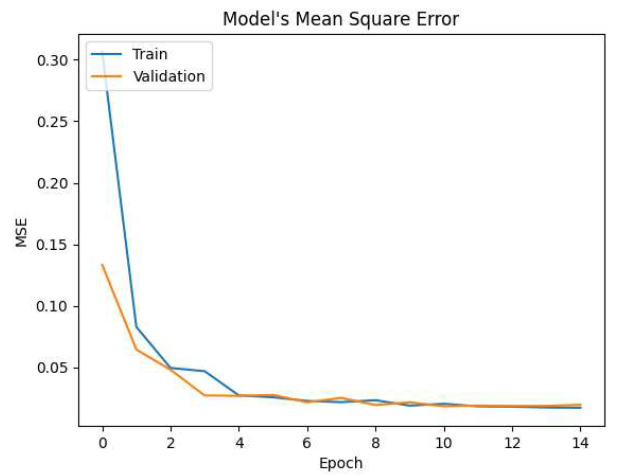
*Figure 2: Example datapoint used in this analysis*

**Source Code:**

We used Google Colab to work on the code for this project, found [here](https://drive.google.com/file/d/1iDgdp-NkyxRTjjtpDG9O4OxL7DmDTyko/view?usp=sharing). The completed runtime in pdf format can be found [here](https://drive.google.com/file/d/1nKWNjowEK6PIYR7HS3AEblXVqevX2gpt/view?usp=sharing).

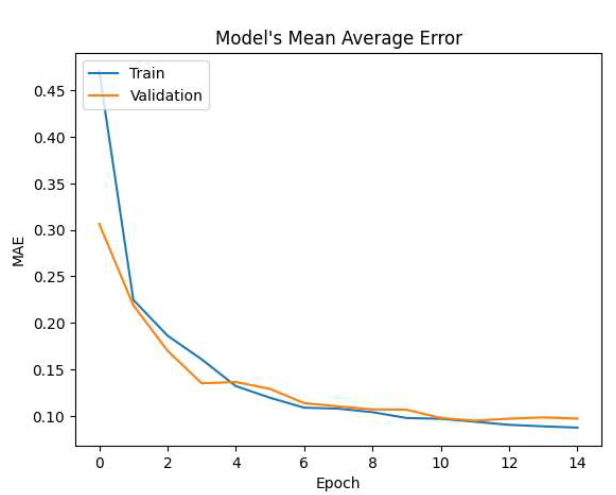
**Results:**

The mean squared error loss function was the basis for evaluating the model’s success. At the model’s completion, the final mean squared error was 0.0195. This minimal amount of error shows the strength of the model’s functionality.



*Figure 3: Plot of the model’s Mean Square Error across epochs.*

In the end, the model was able to predict the popularity of a game based upon its tags with a mean average error of .09. This was higher than desired, but we believe the reason is due to the data used rather than the model itself. We weren’t able to find a comprehensive dataset that combined both user reviews and tags. Any dataset that had the proper tags was too small and too uniform in tag distributions to effectively analyze meaningful differences between tags and groupings of tags. When we transitioned to this dataset with more data, we conceded the fact that the tags we were analyzing were categorical information about the game, such as singleplayer or multiplayer rather than genre. We believe this was the main issue with our model’s ability to more accurately predict the percentage of users that gave positive reviews and would be the first improvement should this experiment be replicated. Further potential changes to the model might include using additional data such as developer and price of the game in addition to the tags, though using the developer data would only be effective when predicting games created by developers that have already released games that have been reviewed.



*Figure 4: Plot of the model’s Mean Average Error across epochs.*

**Conclusion:**

In the end, we were able to achieve the intended goal: training a Deep Learning Model that predicts the popularity of games on Steam based on their associated tags. Not only was this objective achieved, but the final model also had minimal loss. However, there were some issues with our dataset that probably had a negative impact on the accuracy of our model. There weren’t any datasets that contained all of our desired data. A possible follow-up to this project would be to keep the same source code but replace the data with data pulled by us. Our accuracy would likely be much higher if we scraped the data off Steam ourselves. Strategies like this are outside the scope of this project, and so would be a logical next step if it were to be continued.

**References:**

Deepan.N. “80000 Steam Games Dataset.” Kaggle, 27 Dec. 2020, https://www.kaggle.com

/datasets/deepann/80000-steam-games-dataset.

Khan, Zak. “What Is Steam: The Popular Gaming Platform Explained.” Android Authority, 20

Apr. 2023, https://www.androidauthority.com/what-is-steam-3254981/.

“Valve History: Founding, Timeline, and Milestones.” Zippia, 9 Sept. 2022,

https://www.zippia.com/valve-careers-173484/history/.

Clement, J. “Steam Annual Game Releases 2022.” *Statista*, 20 Feb. 2023,

https://www.statista.com/statistics/552623/number-games-released-steam/.