Machine Learning to Predict Video Game Character Genders Based on Quantitative Stats and Attributes

Jack Schwanewede and **Harsh Desai**
{jaschwanewede,hadesai}@davidson.edu
Davidson College
Davidson, NC 28035
U.S.A.

Abstract

Video games are often seen as an ethereal space, a medium where anyone can be or do anything, without conforming to the science, physics, or social structures of the real world. However, just as Machine Learning models are not completely "objective" or free from bias, video games are created by developers who may embed inequalities in game narratives, character design, and the in-game statistics and mechanics. In particular, a character's gender may be highly correlated with certain stats or roles. Our project aims to test this hypothesis as a Machine learning classification problem. Here the target variable is the label character of the gender and the features include quantitative stats (health, defense, strength, etc), and other factors such as their role and difficulty, etc. In order to achieve this task, we used Machine Learning to train classification models using several techniques, including Random Forest Decision Trees, K-Nearest Neighbors, and Support Vector Machines. Our best model, a SVM with RBF kernel, had an accuracy of 63% and a weighted F1 score of 0.64. While our results were moderately accurate at predicting video game character genders, they were less accurate in predicting underrepresented female characters compared to male characters. Still, this paper takes a novel approach towards gender analysis in video games, and highlights the possibilities of embedded biases in video game development and character design.

1 Introduction

Video games have been a topic of interest to scholars from several disciplines, such as in media studies, Psychology, Sociology, Computer Graphics and Computer Science, and more. Media and film scholars have analyzed how narratives can be told through gameplay and other mediums, and Sociologists examine communities and social elements from online gaming. While video games allow players to assume new identities, the gaming sphere is a stereotypically masculine space that often mirrors social structures of the real world. Video game developers may design and create characters with different stats, abilities, and roles depending on the character demographics, such as gender. In this paper, we utilize a Machine Learning approach to classifying character genders from 5 Multiplayer Online Battleground Arena (MOBA) video games. We built a dataset using character stats and attributes, such as their Health Points, attack types, and more, as detailed in our Data Exploration section. We then use this data to train several classification models, such as Random Forest, K-Nearest Neighbors, and Support Vector Machines, to predict gender. Our goal is to see whether our classification models can distinguish character genders, highlighting how video game characters may have noticeable and predictable differences based on their gender.

2 Background

Existing work analyzing gender in video games has done so largely in looking at either gaming communities or the character content found in video games, such as qualitative analysis. Still, several studies have pointed out gender disparities or differences in video game worlds, characters, and the industry. Miller analyzed gender roles in video games and advertisement in magazines, and found that male characters were more likely to be playable and the heroes of the game, whereas female characters were often sidelined and support characters (Miller and Summers 2007). Men also were displayed as more muscular and powerful, while women were more attractive with revealing clothing. More recent work analyzed job and occupational roles in the fictional world of Azeroth, the world in the video game *World of Warcraft*. In this study, Sengun et al. found that more male non-playable characters (NPCs) made up the blue collar workforce, with jobs such as mining and blacksmithing (Sengun et al. 2021). In fact, almost every occupational field was predominately male, with exceptions of archaeology and herbalism having more female NPCs. These NPCs are not chosen by players, but instead designed by the developers of the game. Lastly, in looking at the gaming industry, Bailey et al. recorded the gender composition of development teams of the top selling video games from 2001-2017. For all of these games, the main playable character was male, and in 2001 and 2017 and found no examples where a woman was the main playable character (Bailey, Miyata, and Yoshida 2021). In fact, there were no playable female characters in 69% of the selected games. In the development teams themselves, there were gender differences in the types of roles, such as creative or leadership. *Final Fantasy XV* (2015) had a 13% female development team, but there were significant differences between categories like audio design (36% female), compared to game design (11% female) and programming (0% female).

Overall, several works identify different aspects of gender disparities or differences in video games. However, there is little work done to predict the gender of fictional video game characters based on stats. This project serves as a novel approach towards analyzing gender in video games, instead testing whether Machine Learning classification algorithms can pick up on subtle design differences of characters and accurately predict character gender, and highlighting how there may be embedded gender differences in game and character design.

Data Exploration

In this section, we detail the methods that were used to build and clean the dataset, as well as explaining the features.

First, five Multiplayer Online Battleground Arena (MOBA) games were selected based on the top watched games on Twitch, an online streaming platform, from January to August of 2025. Games considered for analysis, in order of viewership from high to low: *League of Legends*, *Dota 2*, *Mobile Legends: Bang Bang*, *Deadlock*, *Brawl Stars*, *Smite 2*, *Eternal Return*, *Heroes of the Storm*, and *Pokémon Unite*. In order to fit the criteria, the games had to have characters that were predominantly humanoid, and have gameplay that is 'characteristic' of a MOBA. For this reason, the final selection of games were *League*, *Dota*, *Mobile Legends*, *Smite*, and *Heroes*. The first *Smite* game was selected in place of *Smite 2*, as the sequel is currently in an open beta phase and the full game has not been released.

Data was scraped online in early August from several fandom wikis containing updated character stats and attributes, using Python and BeautifulSoup html parsers (of Legends Wiki 2025; Liquipedia 2025; Fandom.com 2025c; 2025b; 2025a). This resulted in five raw datasets, with each character having a column with their respective stats from the online sources. In order to classify gender, student researchers consulted the official websites of each game to read the pronouns used to describe each character. Characters with descriptions, or abilities, using he/him/his pronouns were classified as men (0), and characters with she/her/hers pronouns were classified as women (2). Characters not having deliberate pronouns or using 'it' were classified as 'other' (1).

After looking at the raw datasets, there were several steps taken in order to clean and standardize the data between games. Since each MOBA has slightly different mechanisms, some stats and gameplay elements function differently, so we had to be deliberate in how we matched different features between games and merged them together. We removed features that were unique to one game and could not be related to an equivalent feature from other game, such as Attack Backswing from *Dota 2*, or Unit Radius from *League*. The datasets were merged together with the following features:

- **Name:** The name of the character.
- **Game:** The game the character is from. One-hot encoded to five separate binary features (League, Dota, Mobile Legends, Heroes, Smite).
- **HP and HP/lvl:** Health points of the character, and health points gained per level up.
- **HP Regen HP Regen/lvl:** Health regeneration rate of the character, and health regeneration rate gained per level up.
- **Mana and Mana/lvl:** Mana points of the character, and mana points gained per level up.
- **Mana Regen and Mana Regen/lvl:** Mana regeneration rate, and regeneration rate gained per level up.
- **Armor and Armor/lvl:** Physical defense/resistance, and physical defense gained per level.
- **Magic Res and MR/lvl:** Magical defense/resistance, and resistance gained per level.
- **Attack Damage and AD/lvl:** Strength/damage of each auto attack, damage gain per level.
- **Attack Damage and AD/lvl:** Strength/damage of each auto attack, damage gain per level.
- **Attack Damage and AD/lvl:** Strength/damage of each auto attack, damage gain per level.
- **Base Attack Speed and Attack Speed/lvl:** the amount of attacks per second or attack speed stat, and speed gained per level.
- **Attack Range:** Distance/range of the auto attack.
- **Move Speed:** Movement speed of the character.
- **Price Free/Price Paid:** Price to purchase/unlock the character. Free refers to the price using free currency (rewards for gameplay), paid refers to the price using paid currency (purchasable in microtransactions).
- **Difficulty:** Difficulty of the character's gameplay and toolkit.
- **Melee/Ranged:** Two binary features, whether the character uses melee or ranged attacks. A character can have both melee and ranged attacks.
- **Mana Resource:** Binary feature, whether the character uses Mana as a resource for their attacks.
- **Support:** Binary feature, whether the character has a support role.

In total, there are 30 features, as well as the name of the character, and their gender. Once these features were established, we had to clean the entire dataset to check for null values and fill in missing columns. Some sources that were scraped had incomplete data for some features that required additional searching and verification, or more creative interpretations to fill in the data. For example, in *Dota 2*, some level up stats (HP/lvl, HP regen/lvl, mana/lvl, etc.) were not on the Wiki page. Instead, we found that Dota growth stats were dependent on a character's Agility, Intelligence, or Strength stats, which were collected in the raw datasets. By applying the growth formulas (found online), we were able to derive several features to fill in the data. Another example is that characters in *Heroes of the Storm* do not have a defense stat, meaning the armor and magic resistance features were zeroes for the most part. However, a few characters had mechanics that reduced incoming damage by a certain percent, which was added to the data. When we encountered an issue with the data, such as a missing or inconsistent value, we repeated the basic procedure of looking up details about the game's mechanism for that stat, and used reputable sources to fill in the data to be as accurate as possible.

In addition, we did some feature engineering to create the melee/ranged, mana resource, and support binary variables. These were based on some features that existed in the raw datasets, such as Resource Type or Role, that were too complicated to collapse. Roles between the games differed greatly, yet is an important feature we wanted to capture, so we created a Support binary variable, since all games have characters that can play Support roles. Also, when merging the datasets, we made note of the different scales in stats between games. In *League*, character HP stats range around 500-700, whereas in *Mobile Legends* the range is around 2500-3000. There are different scales between the games for several of the features, which we take into consideration in our experiments section.

The end goal of this project is to predict the target variable of Gender. For simplicity, we are only predicting whether the character is a man (0) or woman (1). In total, there were 648 character examples in the initial dataset. However, there were 8 characters in the list that had either multiple genders (more than 1 character), or no gender assigned (such as an entity). These non-gendered characters were Fiddlesticks, Kindred, Io, Jakiro, Phoenix, Puck, Techies, and Jawhead, and they were removed from analysis. This resulted in 640 total characters, with 169 from *League of Legends*, 121 from *Dota 2*, 130 from *Smite*, 128 from *Mobile Legends: Bang Bang*, and 92 from *Heroes of the Storm*. The full distribution can be viewed in Figure 1.

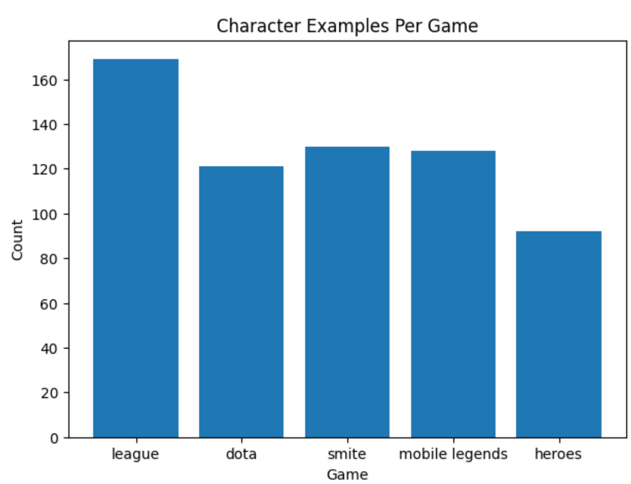

Figure 1: Distribution of characters per game

While there is a relatively equal split between the games, their is a slight imbalance in the gender split. Of the 640 examples, 425 (66.4%) are men, in comparison to 215 being women (33.6%), as shown in Figure 2. There are more male characters than female characters in all five games, most notably in Dota 2, where 97 of the 121 characters are male (80.2%). On the other hand, the most balanced gender split was in League of Legends, with 101 of 169 being men (59.8%) and 68 being women (40.2%).

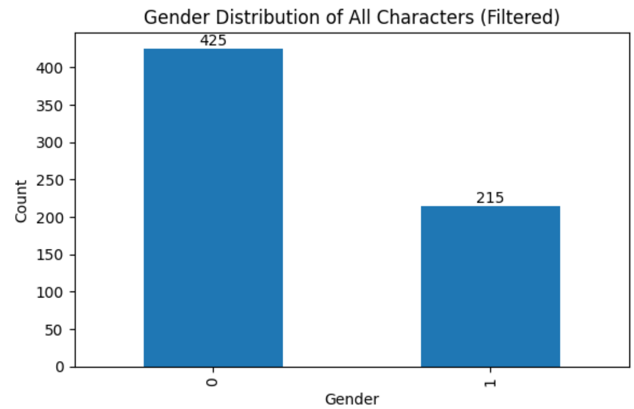

Figure 2: Distribution of the gender target feature.

3 Experiments

In this section, we outline our experiments for our data and models.

Initial models

Random Forests - Initial

For the initial Random Forest experiment, we trained a `RandomForestClassifier` consisting of 200 decision trees. The model was configured with no explicit maximum depth, while restricting tree growth using a minimum of five samples required to split an internal node and a minimum of two samples per leaf. To account for class imbalance, balanced class weights were applied during training. Prior to fitting the model, all features were scaled using Min–Max scaling to ensure a consistent feature range. The model was then trained on the training split and evaluated on the held-out test set using predicted class labels

KNN

For the K-Nearest Neighbors experiment, we trained a KNN classifier with $k = 7$ using distance-based weighting, where closer neighbors have a stronger influence on the predicted class. The Minkowski distance metric was used to measure similarity between samples. Because KNN is sensitive to feature scale, the data was scaled prior to training, and the model was trained on the scaled training set. Predictions were then generated on the scaled test set for evaluation.

SVM Linear

For the initial Support Vector Machine experiment, we trained a linear SVM using the `LinearSVC` implementation. The regularization parameter was set to $C = 1.0$, and balanced class weights were applied to address class imbalance. Prior to training, all features were scaled using Min–Max scaling to ensure comparable feature ranges, which is particularly important for margin-based classifiers. The model was trained on the scaled training set, and predictions were generated on the scaled test set for evaluation.

Final models

Random Forests - Final
For the final Random Forest experiment, we trained a more complex ensemble model and performed systematic hyperparameter tuning using grid search. The model was evaluated over a parameter grid that varied the number of trees, maximum depth, minimum samples required for splits and leaves, feature selection strategy, and class weighting. Five-fold stratified cross-validation was used to preserve class proportions across folds, and macro-averaged F1 score was selected as the scoring metric during cross-validation. Prior to training, all features were standardized using a `StandardScaler`. The best-performing model from the grid search was then retrained on the training set and used to generate predictions on the held-out test set.

Ada Boost
For the final boosting experiment, we trained an AdaBoost classifier using decision trees as weak learners. Each base estimator was a decision tree constrained to a maximum depth of six with a minimum of three samples per leaf to limit overfitting. The ensemble consisted of 300 estimators and used a learning rate of 0.5 to control the contribution of each weak learner. Prior to training, all features were standardized using a `StandardScaler` to ensure comparable feature scales across inputs. The model was trained on the standardized training set and evaluated by generating predictions on the held-out test set.

SVM - RBF Kernel
For the final kernel-based experiment, we trained a Support Vector Machine with a radial basis function (RBF) kernel. The regularization parameter was set to $C = 1.0$, and the kernel coefficient was determined using the default `gamma="scale"` setting. To address class imbalance, balanced class weights were applied during training. Prior to fitting the model, all features were standardized using a `StandardScaler`, as SVMs are sensitive to feature magnitude. The model was trained on the standardized training set and predictions were generated on the standardized test set for evaluation.

Standardized Data Set experiment

For the final experiment, we constructed a new dataset in which character features were standardized *within each individual game* rather than across the dataset as a whole. For each game, character attributes were scaled using z-score normalization, and the resulting standardized values were then combined across all games to form the final dataset. This approach preserves relative differences between characters within the same game while reducing the effect of cross-game variability.

As mentioned earlier in Data Exploration, certain games had features of different scales, where one game might have a stat at a range higher than another game. One method of counteracting this was by also including the game feature, which helps distinguish feature ranges based on the game variable. However, one issue with this is that by including the game itself, the model could 'cheat' and predict gender just based on the game. Since *Dota 2* is mostly male, by passing in a value of 1 for the *Dota* feature it might make the model more likely to predict male. This also meant that were able to remove the game features, as all values were standardized based on distance from the mean within their own game.

Since the Support Vector Machine with an RBF kernel was the best-performing model in earlier experiments, we selected this model for the final evaluation. The dataset was split into training and test sets using an 80–20 split. Prior to training, features were again standardized using a `StandardScaler`. The SVM was trained with balanced class weights and evaluated by generating predictions on the standardized test set.

Evaluation Metrics

In order to evaluate our models, we primarily focused on the classification report package from sklearn, looking at the actual versus predicted classifications. In this report, we also recorded the accuracy, precision, recall, and F1 score. For all models, we used the weighted F1 scores, recall, and precision, in order to weigh the averages across class predictions and be a fair and representative measure. We also recorded the prediction results for each example, looking to see what characters were correctly or incorrectly classified across the different models.

4 Results

In this section, we describe the results and performance of our models, including best hyperparameters and the resulting classification reports and confusion matrices.

Initial Model Results

The initial results across models can be viewed in Table 1.

Random Forest Model:

The Random Forest model achieved an overall accuracy of 0.64, with a weighted F1 score of 0.62. The model performed well on the majority class (class 0), achieving high recall (0.82) and strong precision (0.69), indicating that it was effective at correctly identifying male characters. However, performance on the minority class (class 1) was notably weaker, with a recall of only 0.30 and an F1 score of 0.36. This suggests that while the Random Forest model captured dominant patterns in the data, it struggled to generalize to less frequent outcomes (female characters), reflecting sensitivity to class imbalance.

K-Nearest Neighbors:

The K-Nearest Neighbors (KNN) model achieved an accuracy of 0.66 and a weighted F1 score of 0.63. Similar to Random Forest, KNN performed well on class 0 (male characters), with a recall of 0.83 and precision of 0.70. KNN demonstrated slightly improved performance on the minority class compared to Random Forest, with higher precision (0.50) and F1 score (0.39). This suggests that the KNN's distance-based classification was somewhat more effective at capturing local patterns relevant to the stats of female characters, though recall for male character predictions remained relatively low.

Linear SVM:

The Linear SVM model had an accuracy of 0.62 and a weighted F1 score of 0.62. Although its overall accuracy was lower than that of Random Forest and KNN, the Linear SVM demonstrated noticeably stronger recall for the minority class (0.50). This improvement in minority-class recall indicates that the model was better able to identify female characters, which is a critical objective in our task. However, this gain came at the cost of reduced performance on the majority class, suggesting that a strictly linear decision boundary limited the model's ability to separate the classes effectively.

Comparative Analysis of Initial Models:

While KNN achieved the highest overall accuracy among the initial models, its recall on the minority class remained low, indicating limited effectiveness in identifying underrepresented outcomes. Random Forest similarly favored the majority class, achieving strong recall for male characters but struggling to correctly classify female characters. In contrast, the Linear SVM demonstrated the highest minority-class recall among the initial models, despite lower overall accuracy. Since recall of the minority class (female characters) is important in this project, as we want to be able to accurately predict characters regardless of gender, using SVMs and margin-based estimators would be best for our goals. Because of this, we decided on continuing with the SVM model instead of KNN in our next section, adding the RBF kernel.

Table 1: Resulting performance metrics across our Initial Models. Precision, recall and F1 score represent weighted averages.

Model	Measure	Score
Random Forest	Accuracy	0.64
	Precision	0.61
	Recall	0.64
	F1 Score	0.62
K-Nearest Neighbors	Accuracy	0.66
	Precision	0.63
	Recall	0.66
	F1 Score	0.63
Linear SVM	Accuracy	0.62
	Precision	0.63
	Recall	0.62
	F1 Score	0.62

Tuned/Final Model Results

Our tuned model results can be found in Table 2.

Random Forest (Tuned via Cross-Validation):

After hyperparameter tuning using cross-validation, the Random Forest model achieved an accuracy of 0.66 and a weighted F1 score of 0.62. Our best hyperparameters had the model using bootstrapping, balanced class weights, no max depth, using log2 max features, having a minimum leaf split of 2 and minimum sample split of 5, and using 300 estimators.

Performance on the majority class (male characters) remained strong, with a recall of 0.86, indicating improved identification compared to the initial Random Forest model. However, recall for the minority class remained low at 0.27, showing that while this model improved overall accuracy and majority-class performance, it didn't improve predictions for female characters and classifying underrepresented examples.

SVM with RBF Kernel (Class-Weighted, Non-Standardized Data):

The class-weighted SVM with an RBF kernel trained on the non-standardized dataset achieved an accuracy of 0.62 and a weighted F1 score of 0.63. Notably, this model attained the highest recall for female characters (0.57) among all evaluated models, demonstrating a significant improvement over both the tuned Random Forest and AdaBoost. While this came at the cost of reduced recall for the majority class (0.65), this model had the best overall balance and predictions between the two classes. This further supports that nonlinear decision boundaries combined with class weighting are effective in addressing class imbalance.

SVM with RBF Kernel (Class-Weighted, Standardized Data):

When trained on the standardized dataset, the class-weighted RBF SVM achieved a slightly higher accuracy of 0.63 and the highest weighted F1 score (0.64) among the tuned models. Minority-class recall remained high at 0.57, matching the non-standardized version, while precision for the majority class improved slightly. These results indicate that feature standardization provided marginal overall performance gains without compromising the model's ability to identify minority-class instances, reinforcing the robustness of the RBF SVM approach.

The confusion matrix of results can be found in Figure 3.

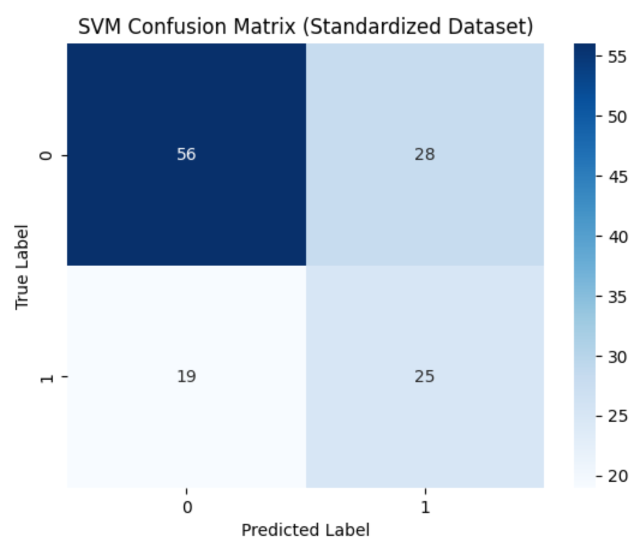

Figure 3: Confusion matrix for RBF SVM, using the standardized dataset.

AdaBoost:

The AdaBoost model achieved an accuracy of 0.66 and a weighted F1 score of 0.64, making it competitive with the best-performing tuned models in terms of overall metrics. While AdaBoost improved minority-class recall to 0.36 compared to Random Forest, it still fell short of the recall achieved by the RBF SVM models. This suggests that although boosting improved class balance relative to tree-based methods, it was less effective than margin-based non-linear classifiers in prioritizing minority-class detection.

The confusion matrix of results can be found in Figure 4.

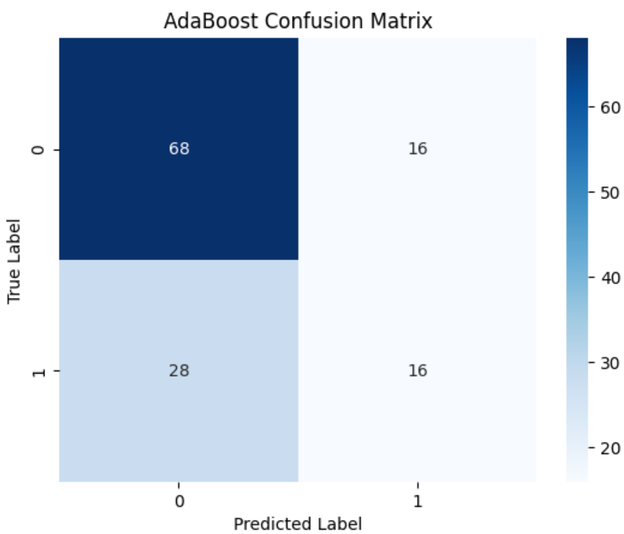

Figure 4: Confusion matrix for AdaBoost model.

Table 2: Resulting performance metrics across our tuned, final models. Precision, recall, and F1 score represent weighted averages.

Model	Measure	Score
Random Forest (Tuned)	Accuracy	0.66
	Precision	0.63
	Recall	0.66
	F1 Score	0.62
RBF SVM (Non-Standardized)	Accuracy	0.62
	Precision	0.65
	Recall	0.62
	F1 Score	0.63
RBF SVM (Standardized)	Accuracy	0.63
	Precision	0.65
	Recall	0.63
	F1 Score	0.64
AdaBoost	Accuracy	0.66
	Precision	0.64
	Recall	0.66
	F1 Score	0.64

5 Conclusions

Overall, our results demonstrated that the gender of MOBA video game characters could be predicted with moderate accuracy, with a higher accuracy for male characters compared to female characters. Our best model, the SVM with the RBF kernel trained on the standardized dataset, was 63% accurate with a weighted F1 score of 0.64. The results across the initial and tuned models emphasize that using non-linear decision boundaries improved classification for the minority class, and had the best balance of results. Some models had higher overall accuracy, but that was due to their prioritization of classifying the majority class, while suffering in accuracy on female characters. Across all models, it appears as if it was a difficult classification task for the minority gender especially (Women), which might also be a result of less representation in MOBA games and our dataset. However, if we had fully accurate results, that would imply that the models were able to clearly discriminate between male and female characters, which would highlight how character stats are possibly dependent on their gender. Because the predictions were moderate, it does not appear that design bias in character design led to significant results that improved classification. However, this paper has a novel attempt for a gender analysis of video game characters, using a Machine Learning approach to predict character gender. It is difficult to necessarily interpret these results, then, as supporting or refuting our hypothesis that accurate results would point towards gender biases by developers in creating characters.

Ethical Considerations and Broader Impacts

This work deals with gender biases and embedded inequalities in video game development. As discussed in the background, the video game industry and community is stereotypically a masculine space, where women and minorities may have experienced discrimination or harassment. This can carry over to developers as well, where character designs and stats may be different depending on the gender of the character. This work aims to highlight these biases, and not further contribute to them, offering an approach to analyze gender differences in video game characters by attempting to distinguish and predict them. The impacts of our models are a bit complicated, as strong results would suggest clear gender differences (high bias/inequality), while weaker results suggest that the classification is more difficult (less bias, characters are designed not according to gender). We believe our results are somewhere in the middle - our overall accuracy was moderate, suggesting that gender was not extremely easy to predict, but our models were much less accurate in correctly predicting female characters. One interpretation of this is that male characters may be designed with a certain archetype or are more similar to one another, having consistent stats such as higher HP or defense, while female characters may be more 'random' with less similarities. However, this would have to be investigated further, and it would be ideal to have a more balanced dataset to test theories further.

Future Work

One limitation of this work was in the size of our dataset. In total, our dataset had 640 characters. Future work could consider increasing the size of the dataset by including other games. However, one difficulty of this project was deciding how to overlap game statistics, since mechanisms work differently depending on the game. For this reason, the dataset had less features than it could have, in order to focus on features that could be somewhat cleanly collapsed on each other. This also meant that all the games included had to stay within-genre. It would be interesting to use a similar approach to analyze character differences in different video game genres, such as Role Playing Games (RPGs) or Hero Shooters. However, MOBAs were also selected due to their large roster sizes, so this analysis on another genre might be even less feasible due to there being less characters in other games.

Additionally, there are several other approaches towards analyzing gender differences based on our dataset. Using statistical testing, such as t-tests, correlation tests, or another analyses can point to how different stats might be higher or lower depending on gender. This can be advantageous towards indicating specific features that are more subject to bias from developers, using test scores and p values to identify where there are significant differences. Because of the scope and goal of this project, we did not employ these measures, but it may be useful for future work.

6 Contributions

Both partners contributed relatively equal work throughout the project. Throughout all steps of the process (dataset building, cleaning, preprocessing, model building and tuning, paper writing), we communicated and shared ideas meeting in person and via Slack, allowing us to collaborate and contribute towards the best possible result and paper.

7 Acknowledgements

AI Acknowledgment

Generative AI (e.g. ChatGPT) was used for debugging, organizing code, and at times graphing data. Portions of code generated or assisted by ChatGPT are documented, and the prompt for this assignment was not fed into any form of LLM to generate our project. The written work in this paper is our own and had no AI involvement.

References

Bailey, E. N.; Miyata, K.; and Yoshida, T. 2021. Gender composition of teams and studios in video game development. *Games and Culture* 16(1):42–64.

Fandom.com. 2025a. Heroes of the storm wiki. `https://heroesofthestorm.fandom.com/wiki/Heroes_of_the_Storm_Wiki`. Retrieved on Aug. 04, 2025.

Fandom.com. 2025b. Mobile legends bang bang heroes. `https://mobile-legends.fandom.com/wiki/List_of_heroes`. Retrieved on Aug. 04, 2025.

Fandom.com. 2025c. Smite wiki. `https://smite.fandom.com/wiki/Smite_Wiki`. Retrieved on Aug. 04, 2025.

Liquipedia, D. . 2025. Hero attributes. `https://liquipedia.net/dota2/Table_of_hero_attributes`. Retrieved on Aug. 04, 2025.

Miller, M. K., and Summers, A. 2007. Gender differences in video game characters' roles, appearances, and attire as portrayed in video game magazines. *Sex roles* 57(9):733–742.

of Legends Wiki, L. 2025. List of champions. `https://wiki.leagueoflegends.com/en-us/List_of_champions`. Retrieved on Aug. 04, 2025.

Sengun, S.; Price, J.; Schlink, L.; and Walker, K. 2021. Azeroth has a workplace gender inequality problem: gendered professions bias in virtual worlds. In *Games and Narrative: Theory and Practice*. Springer. 105–118.