

DSC 412 - Project Proposal and Plan

Student Name: Bruno Perdigao Lobo

Student ID: bperdig

Email: bperdig@ncsu.edu

October 1, 2024

Chosen Project

Title: Dead By Daylight Predictive Killer Tier List Placement Model

Category: Predictive Modeling

Dataset: Community Stats provided by nightlight.gg

Project Proposal

The aim of this project is to analyze the game Dead By Daylight (a.k.a DBD), an asymmetrical horror game in which 2 sides compete for dominance. On one side, the survivors, composed of 4 players, try to complete 5 generators around the map in which they are placed, so they can open exit gates, and escape from the match. On the other side, the killer, summoned by “The Entity” to stop the survivors from escaping, and to sacrifice each one to The Entity.

One somewhat unique feature of DBD is that, just like fighting games, it has multiple characters, each equipped with a different tool to assist in their victory. And, with different ways to play, one thing many people ask themselves as soon as they are introduced to the idea is: which is the best one?

As it turns out, DBD’s cast has different “power levels”. Though survivors are always interchangeable, save some very minor details, some killers have a much easier time applying pressure to the survivors than others, some of them have similar skills, and some of them are better or worse depending on the circumstances. This variety in gameplay has nudged people into “grading” each character based on their expected performance in a match, assuming the

highest possible skill level for the killer, and the survivors going against them. This is done through what is called a “Tier List”, a measurement that puts characters into low, middle and top tier (usually based on letter grades, for example, from E to A).

This project aims to gather unique characteristics about each killer to establish how well would a new killer be graded were it to be announced, solely based on characteristics shared by the developers when the character is announced, before anyone outside of the development team gets to play said killer. From experience, one algorithm with a high cost-benefit ratio to use in this scenario is an entropy/information gain based model, which would be trained using a test/train split from the dataset. This is, however, not decided as of now, since using a different model could provide better results.

Results of this project would benefit many people from many different areas. It would certainly benefit the players that play Dead By Daylight, allowing them to predict killer strengths based on descriptions and raw numbers alone, instead of having to go through the usual lengthy process of figuring out where a killer places on a tier list: Wait for the developers to release the character a couple of weeks after its announcement, wait for killer players to get used to the character’s mechanics, wait for survivors to adapt to said mechanics so they don’t get a false impression of the character’s actual strength, and check how effectively they can go against survivors who know the very basics of playing against that character. However, this is not all. This could benefit many people that play video games, since there is an uncountable amount of games with different playable entities and different power levels. A model like this could be adapted to any game outside of DBD. Outside of gaming, because this model is very generalized, as it focuses on using information gain to gather information rather than relying on specific characteristics of the game, it could help with anything that has multiple entries with different characteristics and must be ranked in one way or another. The possibilities seem endless.

Based on how this semester is going, I believe that the only obstacle to finishing this project will be other classes’ demands. For example, I am enrolled in Senior Design at the moment, which involves way more expected hours than any class that I’m aware of, at least in the Computer Science department, requiring 10 weekly hours at minimum, with logging required to prove

involvement. If anything happens with a project there, I may divert precious working hours to help my team ensure that tasks are completed and that the project carries on without a hitch. I am also in the process of making a Graduate School application, so preparing for the GRE might take some time. Otherwise, I don't see any obstacles getting in the way of finishing this project. In terms of technical issues, there is the issue of how the game is "managed". Dead By Daylight is a competitive game that has been getting constant updates for over 8 years now, and, with each update, killers, perks, add-ons and even game mechanics get introduced, "buffed", "nerfed", reworked, removed and added to the game, in an effort to make things "balanced". This results in an overall shift that can skew the data. Furthermore, the results may change with potential changes to gameplay, and the methodology applied in this project may lose value. However, the point is to establish that it is possible to make this sort of data collection at any point during the game's continued existence, even if the final result may be outdated later.

The most difficult part of the project is the fact that there is no actual scoring system for the characters in the game. All grading is based on the performance that people can see after playing the game for many hundreds, thousands or even tens of thousands of hours. Thus, a new way to calculate the possible score given to each killer in the game had to be figured out by me.

Thankfully, although I haven't played the game in a while, I still understand all the necessary relevant information to grade performances, so it will be just a matter of getting things to fit and hoping that the data provided is enough to give a satisfactory output, in the same way I tried when I did this the last time, hopefully improving on the quality of the model this time around.

The novel element of this project is itself, I believe. That is to say, I have not seen anyone take matters like this and turn them into a predictive model with machine learning algorithms.

Machine learning is a somewhat novel concept, especially considering how long it took for it to get this advanced. Not only that, but a lot of people believe that there is no way to analyze some games like this with a predictive model, as way too many variables need to be taken into account, things that a computer model may overlook, the element of chance intervening in results, or the fact that mistakes are made often. All of these are valid critiques of using a predictive model in this context, which may have deterred interested parties from pursuing this sort of analysis. Thus, if this project shows satisfactory results, I believe it could open the door to

different models in different contexts, that will help people situate themselves in complicated games when they are passionate about their gaming hobby.

Proposed Plan

My plan for this project will be to analyze data collected from nightlight.gg, a website that has received community feedback from over 700,000 match outcomes in the game to date, as well as using general information about each killer in the game, as described by developers and professional players, to set up characteristics that may help predict each character's score. After doing so, the output will be compared to some performance based tier lists made by both professional players and players with lots of knowledge in the game. Because a lot of killers have an expected placement, those results will be almost the same, with some minor variation. If the predicted results match, the model is kept and possibly improved upon. Otherwise, the model is to be reevaluated.

This project is a continuation of something I created for my CSC 422 - Automated Learning and Data Analysis course. Therefore, the vast majority of the data has already been collected, but given it is a one year old project, some additions need to be made. The data was collected in a .csv format, which aggregates multiple qualitative and quantitative features of killers. Some examples being the average movement speed, the skill ceiling, their overall kill rate, whether they can instantly down survivors (more details on this to come on future iterations of the project), whether their individual skill is oriented towards a chase, controlling the map by applying pressure to multiple people or moving very quickly, or to slow down survivors by making them repair generators slower through different ways. All of these categories had their numbers changed into labels: Low, Mid, High and Extreme for quantitative/abstract features, and simple Yes/No for binary features. The data from the website is publicly available, so no permission is needed to share my own interpretation of the information contained therein. Even if information from behind the paywall is used, it will be restricted to the expansive dataset (regular users only have access to 3 months of matches) to ensure the data is accurate.

As mentioned in the project proposal section, the current plan is to use an entropy/information gain model to make the predictions. However, if a better model presents itself, I may change my approach.

The hope with this project is to make a comprehensive algorithm or an overall method capable of predicting where any new killer addition to the game will generally place in a substantial number of tier lists, based on different characteristics and capabilities of a newly released killer. Given different people will have different opinions, the performance metric this project aims to achieve is to correctly predict a killer placement in a tier list, with a possible letter grading from E to S, with a 70% accuracy or better, with a margin of error of 1 letter grade. This would ensure that the model is superior to randomly guessing where they would be placed, which would likely achieve a 50% accuracy rate in the best case.

There are undoubtedly models similar to this online, and the information gain algorithm can be easily encountered in a language, or converted to one of your choice using the mathematical formula. However, I am not aware of a tier list predicting model. If I do come across one, I will make sure to learn from it and use improvements if I deem it necessary.