Features that Impact the Rate of World Records in Video Game Speedrunning

Aidan Ward
Adviser: Xiaoyan Li

Abstract

In this paper, time series prediction models were created with linear and nonlinear regression techniques to predict video game speedrunning world record times as well as the date they occurred for any game with at least seven previous world records. Furthermore, prediction models were trained using additional features about the video games themselves, such as the release year of the game and the total number of people speedrunning it on a speedrunning leaderboard platform called Speedrun.com, and these models were analyzed using their coefficients and permutation importance to determine which features had the greatest impact on what these world record times were and when they occurred, or the rate of world records. It was determined that, of the selected additional features, the number of people following the game on Speedrun.com, the total number of speedruns submitted on Speedrun.com, and the release year of the game had the most impact on the prediction of world records. A greater number of people following the game and total number of speedruns submitted on Speedrun.com caused world records to occur more frequently and with smaller, more iterative improvements, and a larger, more recent release year of the game caused the world records to occur more frequently and with larger improvements.

1. Introduction

1.1. Speedrunning

Speedrun.com defines video game speedrunning simply as "the act of completing a video game as fast as possible"[22]. In other words, speedrunning is an activity that can be compared to track: while the goal of the one hundred meter sprint, for example, is to run a distance of one hundred meters as quickly as possible, the goal of speedrunning a video game is to fulfill the requirements that its speedrunning community defines as completing that game as quickly as possible.

1.2. Motivation

Speedrun.com is a platform in which players can "organize around the activity of speedrunning," allowing users to upload speedrun submissions and displaying leaderboards of the best times for each game, and it is one source which demonstrates the pastime's popularity, stating that twenty million users visit the site every year, 2.2 million users are registered on the site to interact with the game communities, 45.9 thousand games have speedrunning communities on the site, and five million speedruns have been submitted in total[22]. Among the millions of people and speedruns involved in speedrunning, arguably the most interesting aspect of the activity is the world records. The first page which appears when selecting a game on Speedrun.com is the current leaderboard of speedruns, with the fastest time at the top and displayed with some golden icon or medal indicating its placement, the History tab shows the full timeline of world records, who set them, and when they occurred, and the Stats page visualizes this history in graph form showing the decreasing world records over the months and years[21].

In summary, there is a great amount of interest in following the progression of speedrunning world records across a variety of games, and that naturally leads to the question of if new world records can be predicted. What will future world record times be, and when will they occur?

1.3. Goal

Thus, the goal of this project was to collect speedrunning world record data and create a prediction model to answer that exact question of determining what the next world record of a game will be and what day that record will be set based on past world records for a given game. Since this problem has been approached in the past, as will be discussed in Section 2.1 and Section 2.2, the models which were made for this project were intended to improve on these past models as well as investigate another part of this problem: determining how the rate at which speedrunning world records occur vary for different games. Video games can be very different from one another, so there may be other factors about a game which could improve the ability to predict future world records other than simply the past world records while also describing more about the activity of speedrunning in general and what impacts which speedrunning communities progress the world records fastest. Thus, another objective for this project was to analyze additional features about a game, such as its release year and the total number of speedruns submitted for it on Speedrun.com, and determine which ones impacted future world records the most.

2. Background

2.1. Related Works – Sevilla

There are two works which have previously approached the problem of predicting speedrunning world record times. The first work was Jaime Sevilla's "Analysis of World Records in Speedrunning," and this paper also used world record data from Speedrun.com, being 22 speedrunning categories across 15 different games, resulting in a total of 1462 world records[28]. It is important to note that the paper "filtered categories with more than 50 world record (WR) improvements," meaning that each game in this dataset had at least 50 world records[28].

This past work describes some of the observations which will be notable for the results regarding prediction performance in Section 6.1, such as the "diminishing returns curve" of the games' world records which create "a characteristic L-shape" when plotting the world record time on the y-axis of a graph and the date that world record occurred on the x-axis[28]. Sevilla also comments that this curve was "remarkably continuous," possibly due to how, as previously mentioned, all the categories in the dataset had at least 50 world records[28].

A few things about the model creation and evaluation process Sevilla described will be mentioned, as they will be relevant for the models which were created for this project. First, the data for Sevilla's model creation was "train-test split by leaving out the last WR of each category," meaning that the accuracy of the models were only tested on the last world record for each category[28]. The two errors that Sevilla used for assessing the models' performances were mean absolute error, which referred to its usual definition of the absolute value of the difference between predicted value and actual value, and an error denoted as "mean relative error," which referred to "dividing the absolute category errors by the time of the respective current WR"[28]. Since the current world record is the last world record which was the actual world record value used for testing each category, this mean relative error can be generalized to be the absolute value of the difference between predicted value and actual value divided by the actual value, and this error will be referred to as "percent error" for the remainder of this paper. This percent error was the metric which Sevilla stated to be "more informative" due to how "the run times for each category vary widely"[28]. Sevilla also used a baseline model to compare against the models' performances, where "The baseline corresponds to just guessing that the next world record will be equal to the last one"[28].

Next, when analyzing the code included in Sevilla's paper, specifically the cell titled "Least squares model," it can be seen that the prediction method for this past work was to create and train a new model for each individual category[28]. In other words, for Sevilla's 22 speedrunning categories and corresponding sets of world records, 22 prediction models were made. The techniques which were used in this past work were linear regression, log-linear regression, linear grid regression, log-linear grid regression, log-linear grid regression on the twelve most recent weeks of data, and log-linear regression using the ten most recent world records, where the "log" modifier refers to the transformation placed on the inputs and outputs and the "grid" modifier refers only sampling the world record data week by week rather than for each record[28]. Analyzing the code indicates that models were trained using a form of when the world record occurred as the input and a form of the corresponding world record time as the output[28]. The two methods of indicating when the world record occurred were the date and the "WR index," a value which sets the earliest world record to 1 and each subsequent record to 2, 3, 4, and so on[28]. The two methods of indicating the world record time were the world record time and "the relative improvement from record to record" which was included as a part of the Addendum[28].

While the errors will not be reported here, as relevant errors will be included in Section 5.2 when compared with the errors of the models attempting to reproduce these results, Sevilla noted that the models performed better when "modeling the improvement as exponential," meaning the log-transformed models performed better, when "including a grid of datapoints," meaning the grid-sampled models performed better, when "only using recent data," and when using the relative improvements instead of the world record times themselves as training output[28].

2.2. Related Works – Erdil

The second work which has approached the problem of predicting speedrunning world record times was Ege Erdil's "Power-Law Trends in Speedrunning and Machine Learning," edited by Jaime Sevilla[26]. While a significant portion of the paper is concerned with patterns in machine learning in general which will not be discussed due to it not focusing on the problem of predicting speedrunning world records, this paper is a continuation of Sevilla's work in many ways[26]. Once again, this paper utilizes data from Speedrun.com, having slightly more data: "25 categories and 1731 world records," each category having at least 50 world records each[26].[1] The focus of this past work, as indicated in its title, is to investigate how "World record improvements in speedrunning seem to be closely related to power-laws," which a nonlinear time series prediction model called a "random effects model" can "exploit in improving over Sevilla 2021a," referring to the models in Sevilla's work[26].

One of the improvements Erdil makes on the previous models is in regards to "Predicting when the next world record improvement will occur"[26]. In the former paper, Sevilla lists that one of the potential areas of future analysis for speedrunning prediction is determining "information on how frequently world records are achieved," as the structure of the model in this paper used the date of the world record as an input to predict what the world record would be[28]. This is resolved in Section 3.4 of Erdil's paper by creating another regression model for the dates of the world records[26].

Another improvement made on the model is how the models are structured themselves. Unlike the models of the Sevilla's work which were individually created and trained for each category, the code file included in Erdil's paper states that the random effects model is trained on the "first

[1] A footnote comments that 53 categories were used later on, but "This didn't significantly change the results in this section," so it was not commented on further[26].

10 records from each category," and thus this only creates one general model which is capable of predicting world records for any input of world records for a given game rather than requiring a model to be trained for each game[26]. Although this may not seem to change much for the purposes of analyzing a model's performance, I would argue that this makes more sense in terms of the model's use in predicting future world records: the model can be effectively used for any game and category just by inputting the corresponding world records without needing to train it again for each game a user might be curious about. Furthermore, Erdil reported that the model trained on all of the data generally performed better than the model using the same equation which was trained on the category individually, denoted as the "fixed effects model," especially with less data[26].

In conclusion, Erdil reported that the random effects model was the one which performed the best overall, having lower error values than the baseline described in Sevilla's paper which was used once more in this one[26].

2.3. Related Works – Summary and Improvements

While Erdil's paper certainly made many improvements in predicting speedrunning world records, this paper will mainly be focused on utilizing techniques from Sevilla's work due to how I wanted to focus primarily on linear models to emphasize the analysis of the additional features of games and how they impact how often world records occur and what the world records are. However, doing so required that I include many of the improvements made in Erdil's work, such as predicting the date of the world record as well as creating a general model, which will be discussed in Section 3.3, but there will be less of a focus on nonlinear time series prediction models, with the random effects model not being analyzed at all.

3. Approach

3.1. Novel Idea

There are multiple ways in which this project expands on these past works. First, as mentioned previously, I intended to create general models using linear regression techniques for time series prediction, since Sevilla's work created models for each game and Erdil's work focused on nonlinear prediction models. Next, I collected a new set of data which was larger than either of the previous works' datasets, having 99 games with one category each and a total of 3,777 world records, with the emphasis being placed on a larger amount of games in order to collect a large amount of data on additional features.

The main focus and innovation of this project was on analyzing various features of video games to determine if they impacted what the next speedrunning world records for that game would be and when they would occur, since this had not been investigated in previous works: these past works instead mainly focused on the trends of the world records over time themselves as the basis of the predictive models. While I thought that the world records themselves would still be the primary factor in the predictive models, I hypothesized that certain statistics about the games would improve the performance of these models because they would indicate general trends about the rate of the world records. Thus, I chose the following features about the game which were available on Speedrun.com or a reviewing website called Metacritic, including general features about the games as well as ones specific to their speedrunning communities: the release year of the game, the number of years since release, the "Date added" listed on Speedrun.com, the game's "Metascore" according to Metacritic, the game's "User Score" according to Metacritic, the total number of speedruns submitted on Speedrun.com, the "rate" of speedruns submitted, the number of "Followers" a game has on Speedrun.com, the number of "Active" speedrunners for the game on Speedrun.com, the total number of speedrunners for the game on Speedrun.com, and whether or not the game was released on PC, a Nintendo system, a Playstation system, or an Xbox system. These features will be defined in Section 3.2 alongside the potential reasoning as to how they could have impacted the world records.

The process of creating time series prediction models was separated into three parts which can be generalized to many data science problems with a machine learning approach: data collection, processing, and initial analysis, model creation and tuning, and model evaluation.

3.2. Data Collection

First, speedrunning data would need to be collected, with the two types of data for this project being the world record data and the additional features about the game. All of the data was collected from either Speedrun.com or a reviewing website called Metacritic. Speedrun.com, as discussed previously, was used for both past works in the topic of predicting speedrunning world records, so this alongside its popularity and many statistics on various games justified its selection for data collection. Metacritic is a reviewing website which is intended to report "a single score that represents the critical consensus for games, movies, TV shows and albums," and I felt like this would be a reasonable way to numerically represent the reception or enjoyment of various games[9]. The "Metascore" that Metacritic lists, which is on a scale of 0-100[1], "is a weighted average of reviews from top critics and publications for a given movie, TV show, video game, or album"[2], with over 200 publications currently being listed on Metacritic's Support page as sources they gather reviews from for video games[25], and therefore this would be able to encapsulate a wider range of experiences and opinions than one publication or review for each game. Furthermore, Metacritic also has a "User Score," calculated from the user reviews of a given game on the website on a scale from 0-10 including the first decimal place, and I included this as a feature because it might reflect the perspective of the individual players who might be interested in speedrunning rather than publications[8]. These metrics about the games were included because I thought that it might be possible for games with higher ratings to have increased rates of world records, since a more enjoyable game might have a larger speedrunning community interested in speedrunning the game and thus improving more frequently.

The remaining additional features could be found on Speedrun.com. Though most of the statistics given on the pages for each game are not defined formally, their names offer enough information to make reasonable assumptions as to what they are.

The release year of the game is the year the game came out listed alongside the title of the game on Speedrun.com[21]. I thought that newer games might have increased rates of world records due to being more recent and thus more likely to gather interest from speedrunners.

The number of years since release is a metric I created from the data, defined as the year that I collected the data, 2024, subtracted by the release year of the game. This was another way of representing whether a game was newer or older, similarly to the release year.

The "Date added" listed on Speedrun.com is a value given under the Stats tab for a game[21]. The name suggests that it represents the number of years ago that a game was added to Speedrun.com. This was another way of representing whether a game was newer or older, similarly to the release year and the number of years since release, but I thought that this feature might be more accurate since the world record data is collected from Speedrun.com. Therefore, it might be more reasonable for the rate of world records on Speedrun.com to be related to the time that the game was present on Speedrun.com rather than the total time it has been out.

The total number of speedruns submitted on Speedrun.com is a value given both in the side panel called "Game Stats" and the Stats tab for a game as "Total runs"[21]. The name suggests that it represents the number of speedruns that have been submitted for all categories of the game, including speedruns that are not world records. I thought that a game with more speedruns submitted might have increased rates of world records since a greater number of speedruns would indicate that a greater amount of effort has gone into lowering the world records.

The "rate" of speedruns submitted is a metric I created from the data, defined as the total number of speedruns submitted divided by the product of 365.25 and the "Date added," where the multiplication was included to make the rate in number of speedruns per day. I thought that this would be interesting to see if the combination of two different features would potentially impact the rate of world records more or less than its components.

The number of "Followers" a game has on Speedrun.com is a value given both in the Game Stats side panel and the Stats tab for a game[21]. A user on Speedrun.com can "Follow" a given game and it will be added to a list of games that the user is following for their account, and thus the number of "Followers" a game has corresponds to the number of people interested enough in speedruns for a certain game that they want to keep it bookmarked in a certain way[21]. I thought that a game with more followers might have increased rates of world records because a greater number of people being interested in a game might indicate that more effort is being put into lowering the world records.

The total number of speedrunners for the game on Speedrun.com is a value given both in the Game Stats side panel and the Stats tab for a game as "Total players"[21]. The name suggests that it represents the total number of users that have ever submitted a speedrun for that game. I thought that a game with more speedrunners might have increased rates of world records because a greater number of people submitting speedruns for a game might indicate that more effort is being put into lowering the world records.

The number of "Active" speedrunners for the game on Speedrun.com is a value given both in the Game Stats side panel and the Stats tab for a game as "Active players"[21]. The name suggests that, unlike the total number of speedrunners for the game, it represents the number of speedrunners which have submitted a speedrun within a certain time frame.[2] Similarly to the total number of speedrunners, I thought that a game with more speedrunners might have increased rates of world records because a greater number of people submitting speedruns for a game might indicate that more effort is being put into lowering the world records, and the active speedrunners would be an even greater indicator that more effort is being put into speedrunning this game.

Finally, the last features were whether or not the game was released on PC, a Nintendo system, a Playstation system, or an Xbox system. The value of this feature was set to zero if it was not released on a given platform and one if it was released on that platform, and this was determined from the platforms listed for the game on Speedrun.com. PC is one of the platforms which is listed explicitly on Speedrun.com, but the other features are groups of different platforms listed on Speedrun.com[21]. "A Nintendo system" includes FDS, NES, SNES, GB, GBP, N64, GBC, GBA, GCN, DS, Wii, 3DS, New3DS, WiiU, Switch, WiiVC, 3DSVC, New3DSVC, WiiUVC, NESClassic, and SNESClassic[21]. "A Playstation system" includes PS1, PS2, PS3, PS4, PS4Pro, PS5, PS5Pro, PSP, PSVita, PSTV, and PSClassic[21]. "An Xbox system" includes Xbox, X360, XboxOne, XboxOneS, XboxOneX, XboxSeriesS, XboxSeriesX[21].[3] I was not sure whether or not

[2]While it is not stated how long this period of time is, an estimate was made in Appendix A.1.

[3]These sets of platforms are somewhat arbitrary, and details on how they were chosen are described in Appendix A.2.

this would impact the rate of world records, but I was interested in seeing if world records were particularly impacted by the platforms that a game is available on.

3.3. Model Creation

After collecting the needed speedrunning data, the next step was to create time series prediction models which used this data to create some form of prediction of what and when the next world record would be for a given game. Two sets of models were created.

The first set of models would be trained on each game individually and only on the world records, not any of the additional features. In other words, this set of models was created using the structure and some of the methods from Sevilla's work, and this was done to get an overall comparison of the datasets being used. If the same models and techniques are used on different datasets, this could reveal some of the differences between the datasets which might explain characteristics of the other models. Not all of the models were recreated, but the baseline, linear regression, log-linear regression, and log-linear regression on recent world records models were tested, where the input for the measure of when a world record occurred was the date and the input for the measure of the world record time was the world record time, not the relative improvement. I felt like this would allow for a reasonable comparison between the datasets on if linear models were still capable of creating accurate predictions, if log transformations were still more effective, and if utilizing only the most recent world records was still more effective. Furthermore, the baseline model used here would also be used as the baseline model for the entire project, since both past works used the same baseline model and found it to be an adequate measure of how well other models performed.

The second set of models would be trained on every game and accept any world record input to make a prediction, models which I have denoted as general models. There were three different parts of how a model was created for these general models.

The first part of the model was the method of regression used, and I chose to use linear regression for more direct comparison with the linear individual models, lasso regression for its property of more effectively removing features which do not significantly impact the model, and random forest regression for analyzing the performance of a nonlinear model.

The second part of the model was whether or not the additional features of the games themselves would be used or not, and this was done to compare whether or not including the additional features improved the performance of a model which was otherwise the same.

The final part of the model was the method of inputting world record data that was used, where each world record would have its world record time and the date it occurred when used for training and testing the model. The first option for world record input was to train the model by having the training output be the most recent world record and the training input be the seven most recent world records before that. In order to create a general model which accepts any set of world records, it was required that the number of world records input was constant, so the minimum number of world records for each game was chosen such that every game in the dataset could be utilized, with eight world records specifically being chosen for reasons which will be discussed in Section 4.1. The second option for world record input was to train the model by having the training input be seven sequential world records and the training output be the most recent world record after that while stepping through the training data. "Stepping through" the training data will be described using an example. For a given game's set of world records, training by stepping through the data would not just train on the eight most recent world records. Instead, one input-output pair for the model would use the game's first eight world records, using the same split of seven sequential world records being input and the eighth and most recent world record being the output. The next input-output pair for the model would use the game's second world record through the ninth world record, once again where the second through eighth world records would be the input and the ninth and most recent world record would be the output. The next input-output pair for the model would use the game's third world record through the tenth world record, and so on. This was done because only using the eight most recent world records did not utilize all of the data which was collected, but stepping through the data did and could potentially improve the model's performance while still maintaining that the test set was not used for training and the model was always given sequential input-output pairs.

Every combination of these possible regression methods, feature inclusions, and world record inputs of creating a general model was created and evaluated to see which methods were the most effective.

3.4. Model Evaluation

The model's performance was measured using root mean squared error (RMSE), mean absolute error (MAE), and the previously described percent error. While the root mean squared error was included to have an additional metric of evaluation, the errors which were primarily focused on were the mean absolute error and percent error. The mean absolute error provides a direct value indicating the error of the world record prediction in the same unit as the world records, while the percent error, as described by Sevilla previously, provides a way of analyzing the error relative to the length of the speedruns themselves, since they can differ greatly from game to game. Additionally, both of these errors were used in Sevilla's initial past work, thus allowing for direct comparison between models. While the individual models only had errors for what the world record times were from the structure of the model's outputs, the general models had errors for both the dates that the world records occurred and the world record times themselves.

Overall, the error value which was given the most importance for the general models was the mean absolute error for the world record time. The world record time was given emphasis because it is both a smaller unit of measurement than the world record date, being seconds and days respectively, and the focus of speedrunning is primarily on the world record times themselves. The mean absolute error was used because, while it was mentioned in Section 2.1 that Sevilla emphasized the importance of percent error due to the differences in game lengths, the general model is tested for all games, and thus the percent error is still valuable but not needed for scaling the performance of the models. If this is not needed, the focus would be on reducing the difference between the prediction and the actual world record time for any game as much as possible, which is represented through the mean absolute error.

Once a model was reasonably determined to be the best performing from these error values, both the coefficients of the model as well as permutation importance were used to analyze which additional features had the greatest impact on the prediction models. The coefficients of the models would not be standardized, meaning their magnitudes would not necessarily indicate if one feature had a larger impact than another, but whether they were positive or negative indicated how they affected the model. The results of permutation importance were used to determine how greatly the additional features impacted the model, since they were not dependent on the magnitudes of the features' values.

4. Implementation

4.1. Data Collection and Processing

When collecting the data, I found that the APIs offered on Speedrun.com and Metacritic were not able to provide all of the information I was interested in, so I chose to collect the data manually in a spreadsheet alongside the title of the game and the date that I collected the data. All of the world records for any category of any game are available to download on Speedrun.com as a .csv file, containing the date that each world record occurred and the world record time in seconds alongside other information which was not needed for this project, such as the username of the speedrunner[21]. During this process, I created guidelines for how I would collect data which are fully listed in Appendix A.3, but the general principle was to remain as consistent as possible. One guideline which had a significant impact was that the game must have at least ten world records. While this value was chosen somewhat arbitrarily and could have potentially have been different, both of the past works' best performing models used ten world records in training them, so I believed that each game having ten world records to resemble this approach would be reasonable.

Overall, there were a total of 99 games in the collected dataset, each having anywhere between 10 and 188 world records as well as the 14 additional features described previously, and this resulted in a total of 3777 world records.

Once the data for the additional features was collected on a spreadsheet and downloaded as a .csv file and the corresponding world record data .csv files were downloaded, I chose to process the data in Python because it has many libraries that help with the data analysis and model creation process[15]. More specifically, I chose to process the data in a Jupyter Notebook, or .ipynb, file, which allowed individual cells to be created to run code for later parts of the project without having to run earlier data processing parts repeatedly. One of the libraries which was used to organize the data was pandas, where the .csv files containing the dataset were loaded into a pandas dataframe[14].

Then, all of the “Unknown” values for the “Date added” feature were set to be equal to the year the data was processed, 2024, subtracted by the year that the oldest world record for that game took place. Although this may not exactly represent when the game was added to Speedrun.com, the first speedrun which was submitted for a given game would be that game’s oldest world record, and it seemed like a reasonable assumption to state that the first speedrun would occur around the time that the game was added to Speedrun.com.

The data was then split using an 80-20 training-test split. Due to this being time series data, however, the data was split specifically such that the first 80% of world records was used as training data and the last and most recent 20% of world records was used as test data, and therefore the training data did not have any access to future data that it should not have accessed. This is why the models used only the eight most recent world records if they used a set number of recent world records: the minimum number of world records for each game was ten, and the minimum number of world records each game had in its training set after the training-test split was therefore eight. The additional features were not split in this way because, when the general models with additional features are trained and tested, the additional features will be assigned to the input for each world record corresponding to its game.

After, a correlation map of the additional features was made using pandas’ DataFrame.corr() function[13] from the training data and plotted using Matplotlib[7], which can be seen in Figure 3 in Section 5.1. Matplotlib is another useful Python library, and it was used to create all of the graphs and plots for this project.

From this point, I believed that it would be beneficial to convert the pandas dataframe into a NumPy array from the NumPy python library for easier value accessing and modification for data processing[11].

With this NumPy array containing all of the data, the dates of the world records needed to be processed into a form which could be input to the time series prediction models. This was done using Matplotlib's dates.set_epoch() function[6] and dates.date2num() function[5]. The date of the oldest world record in the entire dataset was found, being March 20, 1999, and then the epoch was set to March 19, 1999. Thus, when the dates.date2num() function was called with an input of the date of each world record, it returned the number of days since March 19, 1999, and therefore a consistent scale for all of the dates in the dataset was created. The reason that the day before the date of the oldest world record was used was to make the log transformation of the date simpler, since the log of zero cannot be taken and the oldest world record would therefore cause an error if the log of its number of days since March 20, 1999 was taken.

From here, the last initial data analysis which was performed was plotting the world records in the training data, and the resulting plots can be seen in Figure 1 and Figure 2 in Section 5.1.

4.2. Model Creation and Evaluation

Scikit-learn is an incredibly useful Python library which was used for all of the time series prediction model creation as well as a lot of the evaluation[16]. The linear_model.LinearRegression() class[20] was used to create the linear regression models, the linear_model.Lasso() class[19] was used to create the lasso regression models, and the ensemble.RandomForestRegressor() class[17] was used to create the random forest regression models. Separate cells were used for training and testing each of the models to be able to organize each of the performance analyses for the various models.

For all of the individual models, the root mean squared error, mean absolute error, and percent error for the world record time prediction were calculated and can be seen in Section 5.2.

For all of the general models, the root mean squared error, mean absolute error, and percent error were calculated for both the date of the world record and the world record time and can be seen in Section 5.3. If the model used lasso regression and therefore required hyperparameter tuning, a large number of alpha values were iterated through in a for loop to determine which two values of alpha resulted in the lowest mean absolute error for the date of the world record and the world record time, and these were used in the cell where the model was created and tested. If the model used random forest regression and therefore required hyperparameter tuning, a similar method of determining the two sets of parameters which resulted in the lowest mean absolute error for the date of the world record and the world record time was used, but the for loops iterated over the number of estimators, the maximum depth, and the minimum sample split, since iterating through all of the possible hyperparameters would take a very long time. The results which were reported in this paper were the ones corresponding to the hyperparameters which minimized the world record time error, because, as stated in Section 3.4, the focus was on minimizing the world record time error.

As mentioned previously, once the model with the best performance was determined, then the feature analysis was performed for that model in order to not repeat it for every model with the additional features.[4] First, residual plots for the date of the world record and the world record time were created. Then, if the model used linear regression or lasso regression, the .coef_ attribute of both the linear_model.LinearRegression() class[20] and the linear_model.Lasso() class[19] was used to determine the coefficients of the model corresponding to the features, and these coefficients were plotted on a bar graph. The inspection.permutation_importance() function from scikit-learn was used to determine the magnitude of the features' impacts, since the coefficients were already printed and graphed[18]. This function was called for both the training data and the test data to determine as much information from the additional features as possible, and the resulting importances were graphed and can be seen in Section 5.4.

All of the code used for implementing this project can be found at the following GitHub Repository: `https://github.com/aidanward0804/speedrunning_prediction_research.git`.

[4] More details on the extent to which feature analysis was performed for the models which were not determined to be the best are described in Appendix A.4.

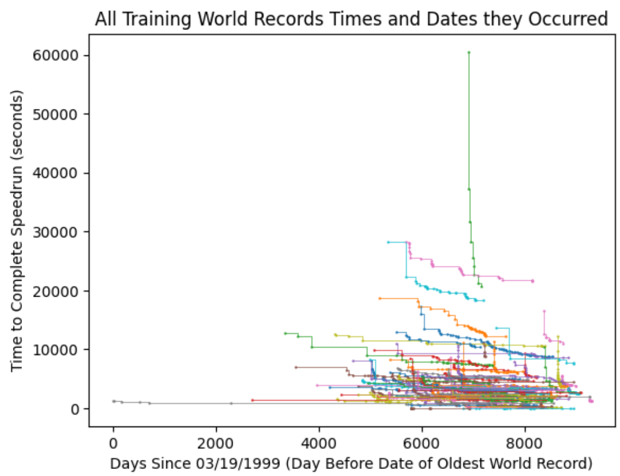

Figure 1: Plot of All World Records in Training Dataset

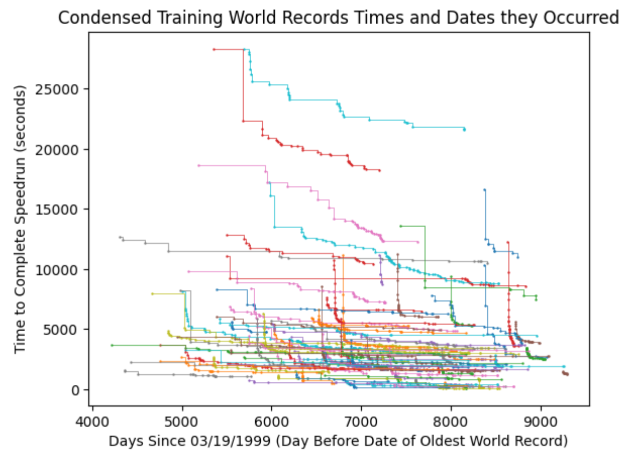

Figure 2: Condensed Plot of World Records in Training Dataset

5. Results

5.1. Data Correlations

Two graphs plotting the training world records have been shown in Figure 1 and Figure 2. Figure 1 displays every world record time in the data set in seconds over the number of days since March 19, 1999. However, a few outliers in the data make it difficult to see many of the details of the graph, so another graph was created, Figure 2, which excludes some of these outliers to have a closer view of some of the trends of the world records.

The correlations between the additional features can be seen in Figure 3. It can be seen that there is a strong correlation between the release year, years since release, and "Date added" to Speedrun.com. This makes sense due to the years since release being a negative offset of the release year, and it is reasonable that most games would be added to Speedrun.com at almost the same time as they are released, though this may not be true for older games which existed before Speedrun.com. There was a strong correlation between games being released on PC, games being released on a Playstation system, and games being released on an Xbox system, and this also makes sense as many of the games in the dataset showed that it was common for games to either only be on Nintendo platforms or to be in a variety of the platform groupings. Finally, there was a strong correlation between the total number of speedruns submitted on Speedrun.com and the rate of speedruns submitted, the number of people following the game on Speedrun.com, and the number of total and active speedrunners, though there is only a weaker correlation between the rate of speedruns submitted, the number of people following the game on Speedrun.com, and the number of total and active speedrunners themselves. This makes sense because the rate of speedruns submitted is derived from the total number of speedruns submitted, and otherwise all of these additional features correspond to the game's popularity for speedrunning in some way or another.

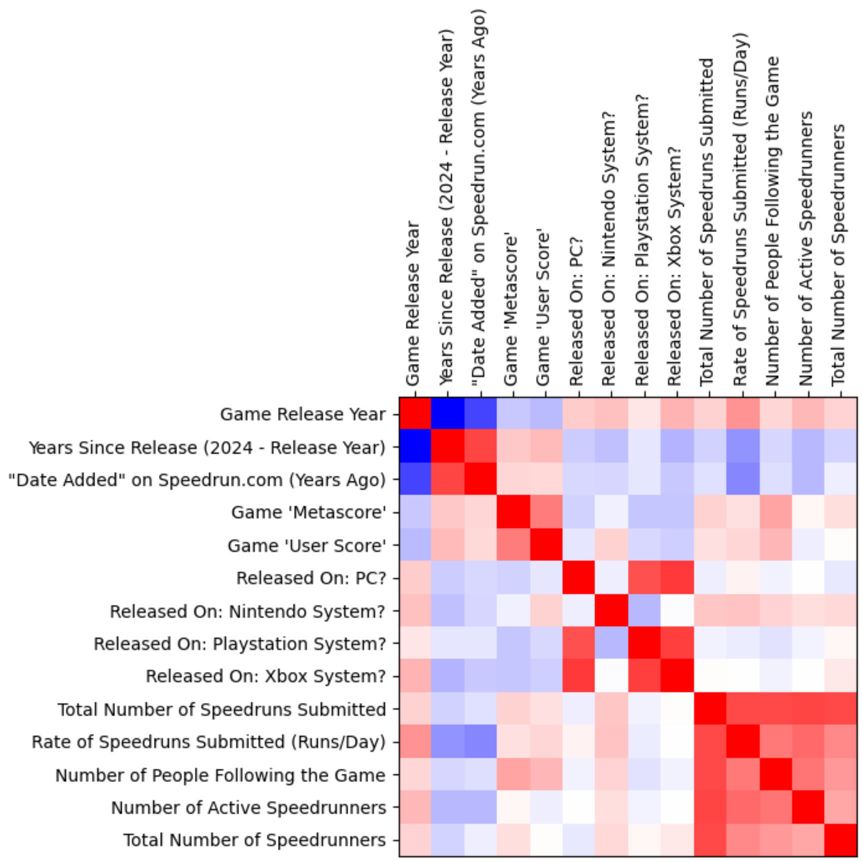

Figure 3: Correlation Map for Additional Features

5.2. Individual Model Performance

Table 1 displays the individual models' performances in this project, and Table 2 displays the performances reported by Sevilla's initial work of the corresponding models,[5] though the log-linear regression models using recent world records are not exactly the same due to the different number of world records used as well as Sevilla's log-linear models with recent data using grid sampling for one model and using the improvements between world records rather than the world records themselves for the other[28]. These tables, along with all of the other tables in this report, were created using LaTeX tables[12]. It can be seen that, while the baseline performance was comparable between the datasets, the other models for this project had much higher error values than the ones reported by Sevilla.

Model	MAE (seconds)	Percent Error
Baseline	30.96	1.13%
Linear	1296.68	69.93%
Log-Linear	477.83	16.35%
Log-Linear from Eight Most Recent Records	227.16	6.66%

Table 1: Individual Model Performances

Model	MAE (seconds)	Percent Error
Baseline	33.22	0.86%
Linear	496.13	15.57%
Log-Linear	308.08	8.22%
Grid Log-Linear from Recent Twelve Weeks of Records	46.77	1.21%
Log-Linear from Ten Most Recent Records by Improvements	23.17	0.71%

Table 2: Individual Model Performances Reported by Sevilla[28]

[5]The baseline errors are the ones reported for the Addendum to Sevilla's paper which is included with it[28]. Although the Addendum does not make any mention of a different dataset or change to how the baseline model was calculated, the errors which were reported before the Addendum were a mean absolute error of 33.23 seconds and a percent error of 0.86%[28]. Given that the percent errors were the same between reported values and the difference is very slight, this difference was not considered to be significant and likely due to rounding.

5.3. General Model Performance

Table 3 displays a table of the performances of the general models which used linear regression, with errors in both the date that the world record occurred and the world record time of the model's prediction. As can be seen by the table, all of the errors reported that the step through method of training performed better for all models, while generally including the additional features did not improve the model in most areas. The only errors which were improved by including the additional features were the root mean squared error of the date the world record occurred and the mean absolute error of the world record time of the step through models.

Linear Models		**Date Errors (Days)**			**Time Errors (Seconds)**		
Includes Additional Features	**Training Input**	*RMSE*	*MAE*	*Percent Error*	*RMSE*	*MAE*	*Percent Error*
No	Most Recent Records Only	211.98	115.17	1.35%	59.99	21.04	1.82%
No	Step Through Records	206.23	98.51	1.15%	56.04	17.70	1.32%
Yes	Most Recent Records Only	222.61	128.73	1.51%	60.13	22.40	4.29%
Yes	Step Through Records	205.07	99.75	1.17%	56.17	17.48	1.51%

Table 3: Performances of General Linear Regression Models

Table 4 displays a table of the performances of the general models which used lasso regression, with errors in both the date that the world record occurred and the world record time of the model's prediction. The first observation was that, other than the root mean squared errors for the step through models, all other errors for all of the lasso regression models were at least as low or lower than the corresponding linear regression model errors. While most of the errors displayed that the models which used the step through method of training performed better like they did for the linear models, the root mean squared error of the date the world record occurred and the percent error of the world record time both increased for the step through models, regardless of whether or not the additional features were included. Additionally, more of the errors for the lasso models displayed that including the additional features improved the performance of the model than the linear models, with the root mean squared error of the date the world record occurred being lower for the step through models and the root mean squared error and mean absolute error of the world record time for all models being lower for the step through models.

Lasso Models		**Date Errors (Days)**			**Time Errors (Seconds)**		
Includes Additional Features	**Training Input**	*RMSE*	*MAE*	*Percent Error*	*RMSE*	*MAE*	*Percent Error*
No	Most Recent Records Only	201.44	105.63	1.24%	58.40	18.77	1.01%
No	Step Through Records	206.86	98.42	1.15%	56.26	17.59	1.21%
Yes	Most Recent Records Only	201.77	105.88	1.24%	58.31	18.62	1.04%
Yes	Step Through Records	206.72	98.82	1.16%	56.23	17.32	1.46%

Table 4: Performances of General Lasso Regression Models

Finally, Table 5 displays a table of the performances of the general models which used random forest regression, with errors in both the date that the world record occurred and the world record time of the model's prediction. It can be seen that this method of regression performed significantly worse than both of the previous methods in all of the error values. The step through training method decreased the errors for both the date of the world record and the world record times for all models, except for all of the date errors for the models which did not include the additional features. Including the additional features lowered all of the errors for the models which used the step through training method, but only the root mean squared error of the world record time was lowered when including the additional features for the models which did not use the step through method. It is important to note that the errors reported here are from one test of this model, and the model's errors and the trends described from them can vary slightly from execution to execution when recreating the model due to the random nature of random forest regression. However, these differences were considered small enough alongside performing significantly worse than the linear regression and lasso regression models that this variance was not investigated further.

Random Forest Models		**Date Errors (Days)**			**Time Errors (Seconds)**		
Includes Additional Features	**Training Input**	*RMSE*	*MAE*	*Percent Error*	*RMSE*	*MAE*	*Percent Error*
No	Most Recent Records Only	372.26	261.97	3.02%	259.11	102.74	23.52%
No	Step Through Records	410.90	275.37	3.16%	191.95	101.78	12.77%
Yes	Most Recent Records Only	408.00	294.95	3.39%	245.80	103.08	23.80%
Yes	Step Through Records	406.05	273.64	3.14%	182.10	99.60	11.58%

Table 5: Performances of General Random Forest Regression Models for One Execution of the Code

From here, it was difficult to exactly determine which model performed the best from the general models, since the linear step through model with additional features, linear step through model without additional features, and lasso step through model with additional features each had certain errors which were lower than the other models' errors while other errors were higher. I decided to choose the lasso step through model with additional features as the best performing model used for further feature analysis, both because it includes the additional features which are needed for the analysis and because it had the lowest mean absolute error for the world record time of all of the

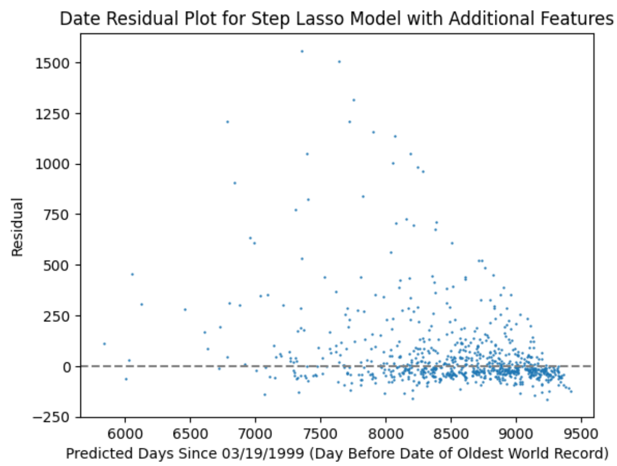

Figure 4: World Record Date Residual Plot for the Step Through Lasso Model with Additional Features

Figure 5: World Record Time Residual Plot for the Step Through Lasso Model with Additional Features

models, which was the error value given the most emphasis, with the other errors still being very comparable.

From the residual plots for the step through lasso model shown in Figure 4 and Figure 5, it can be seen that there are no extreme trends which indicate that the model is particularly biased in its predictions for any date of the world record occurring or world record time.

5.4. Feature Analysis

Graphs for the mean importances which include all of the features of the model make it difficult to discern which additional features are the most important due to the dates of the world records and the world record times being much more important features, as expected. To resolve this, the graphs shown in Figure 6 and Figure 7 only include the additional features, where the black lines denote the standard deviation of the feature. The bar graphs with the additional features and the world record dates and world record times are shown in Figure 9 and Figure 10 in Appendix A.5, and the numerical values of the importances of the additional features are given in Table 7 in Appendix A.5.

From Figure 6 and Figure 7, it can be seen that three additional features were particularly important: the number of people following the game on Speedrun.com, the total number of speedruns submitted on Speedrun.com, and the release year of the game. While it could be argued that the number of total and active speedrunners on Speedrun.com and the Metascore of the game

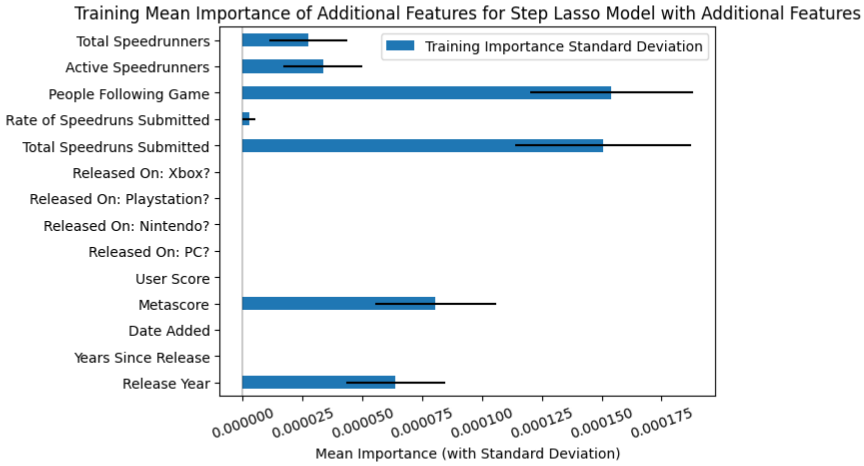

Figure 6: Training Mean Importances of Additional Features for Step Through Lasso Model with Additional Features

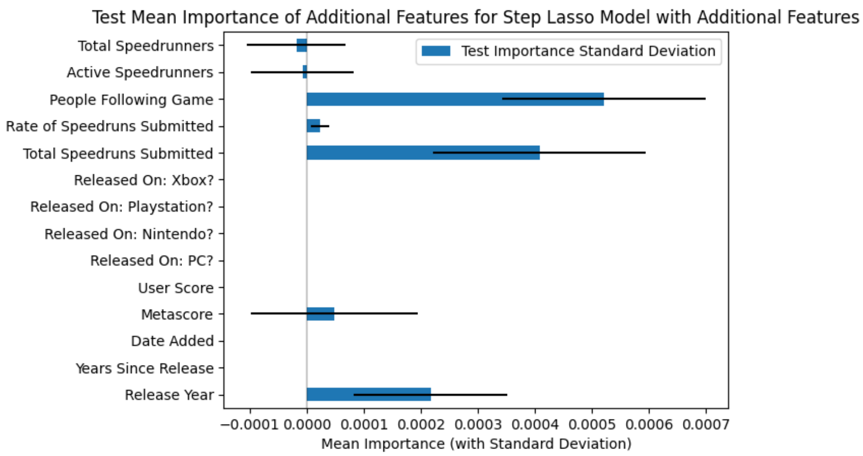

Figure 7: Test Mean Importances of Additional Features for Step Through Lasso Model with Additional Features

could also be particularly impactful based on their importance values for the results of the training permutation importance, the results of the test permutation importance report that their standard deviations are higher than their importances, and thus it is likely that their importances were not significant.

While the bar graph of the coefficients of the step through lasso model with additional features and the numerical values of these coefficients are shown in Figure 8 and Table 6 in Appendix A.5, respectively, only the signs of the three most important features will be stated here because, as mentioned previously in Section 3.4, the coefficient values themselves do not necessarily indicate how important a feature is due to the values not being standardized.

The coefficients for the number of people following the game on Speedrun.com are negative for the date of the world record and positive for the time of the world record, the coefficients for the total number of speedruns submitted on Speedrun.com are negative for the date of the world record and positive for the time of the world record, and the coefficients for the release year of the game are negative for both the date of the world record and the time of the world record.

6. Conclusions

6.1. Comparing Current and Past Individual Models

As stated in Section 5.2, the errors for the individual models were much higher in this project than their corresponding reported values in Sevilla's initial work, excluding the baseline model. One potential reason for this is that all of the games in the dataset of Sevilla's initial work had at least 50 world records, while the games in this project's dataset only had at least 10 world records. This might be particularly impactful for the linear models which were used because, when there are many world records for a diminishing returns trend, many of these world records form a relatively flat line towards the end of the curve, and this may benefit the linear model by causing its fit to the curve to align with that flat line which is likely where the next world record is going to appear. This is especially true when only using recent world records which are only on this line. When there are fewer world records, however, the entire trend is represented with fewer points along the end of this curve, and thus it is possible for the fit of the linear model to have a greater slope downwards which is not beneficial for predicting world records.

Despite the higher errors for the other models, the comparable baseline errors and improvement of the recent record and log models indicate that the dataset of this project likely followed similar trends to the dataset of Sevilla's work and could reasonably be used to analyze the general models.

6.2. Comparing Individual and General Models

From the results in Section 5.3, it can be seen that the mean absolute error of the world record times of all of the general linear and lasso regression models performed better than both the baseline model as well as the best performing model in Sevilla's initial work. Furthermore, while not all of the percent errors of these general models were lower than the percent error of the baseline model and the best performing model in Sevilla's work, it can be noted that the general linear regression models performed significantly better in both mean absolute error and percent error for the world record time than the individual linear regression models. This demonstrates the effectiveness of using recent world records and that general models can be created with linear regression techniques with comparable or even better performance, expanding on the work Erdil performed demonstrating that general models are effective for nonlinear time series prediction techniques.

Additionally, the results report that the step through method of training the model usually performed better than using only the recent world records. While there could be multiple possible reasons why, I believe that part of the reasoning may simply be due to stepping through the data providing significantly more data and therefore more properly training the model.

It is slightly less clear from the results whether or not including the additional features improved the performance of the models overall. One possible interpretation is that many of the additional features, such as the platforms that a game released on, had no significant importance and impact on the model, and therefore including these features when training and testing the model only provided noise which led to worse performance. This is supported by the idea that variations in error were only slight, and furthermore the models using lasso regression did typically benefit from the additional features. Lasso regression has the particular characteristic of being less impacted by features which have low importance values, so it follows that it would be able to take better advantage of the additional features if many of them do not have significant importance.

While it could be argued that either linear or lasso regression were more effective for the general models, it is clear that random forest regression performed the worst out of the methods used.

6.3. Most Important Additional Features

As stated in Section 5.4, the most significant additional features which impact the rate of world records in speedrunning are the number of people following the game on Speedrun.com, the total number of speedruns submitted on Speedrun.com, and the release year of the game. When the coefficient for the date of the world record occurring is negative and the coefficient for the world record time is positive, as they are for both the number of people following the game on Speedrun.com and the total number of speedruns submitted on Speedrun.com, this indicates that higher values for these features caused the world records to occur more frequently but have less improvement in between world records. This is a conclusion which could be potentially explained by the idea that both of these features generally indicate that a game is particularly popular to speedrun. For a more popular game, it is likely that speedruns occur more regularly and therefore world records would likely be broken more frequently in smaller amounts rather than by massive improvements each time. When the coefficients for the date of the world record and the world record time are negative, as they are for the release year of the game, this indicates that higher values for this feature caused the world records to occur more frequently and with larger improvements between records. This is a conclusion which can be potentially explained by the idea that the world records for newer games would be closer to the beginning of the diminishing returns curve which is expected for a world record trend. Therefore, the world records are not at the plateau that might occur later on, and it is more likely that large improvements would be made to the world record in rapid succession.

6.4. Future Work

One potential area of future work based on these results would be performing more feature analysis, especially varying the inputs to training and testing the models more. This project focused on either inputting no additional features or all of the additional features, so it would be interesting to see whether or not inputting only some features would resolve the potential noise from less important features which was discussed previously.[6]

Additionally, it would be interesting to perform further analysis on the role of the additional features in nonlinear time series prediction models. Although random forest regression did not perform as well as the linear regression and lasso regression models, Erdil's work demonstrated that

[6] Supplementary details about work in this area are discussed in Appendix A.6.

nonlinear models can be effective, so it would be interesting to investigate the combination of these additional features and a better performing nonlinear model.

The largest area of future work, however, is likely in regards to the limitations of the data. When inputting the data for training and testing the models, additional features were input with world records according to what game the world record was for. However, some of the features, such as the number of active speedrunners on Speedrun.com, change over time, and therefore the values of these features when they were collected for this project do not necessarily reflect what they were at the time the world record occurred. However, I was unable to find a way on Speedrun.com to determine what these values may have been at any point in the past. One possible method of resolving this problem could be to collect the data at multiple points over time to either be able to assign the values exactly according to the date or to be able to more accurately interpolate the data.

Another limitation with the additional features was with the speedrunning categories. While I limited this project to only using one category for each game, there is a lot of data available if other categories are used, both for world records and for potential additional features.

7. Acknowledgements

I would like to acknowledge the following people for their contributions: my adviser, Dr. Xiaoyan Li, for her guidance in developing this research, leading IW Seminar COS 397 S02 where I learned about the structure of data science projects utilizing machine learning, the properties of lasso regression, and much more, and informing me about the seminar webpage[4] where I could reference past projects for the seminar for formatting help[27]; COS 397 S02 IW Seminar TA Joseph Xu, for his feedback and help with hyperparameter tuning the random forest regression models; COS 397 S03 IW Seminar TA, Aditya Palaparthi, for his help and answering my questions about machine learning concepts; my COS 397 S02 IW Seminar classmates, for their feedback and guiding questions; and the COS Department at Princeton, for advising how to write a research paper[29] and providing the LaTeX template for it[3].

8. Honor Code

This paper represents my own work in accordance with University Regulations. -Aidan Ward

References

[1] "How do I interpret a Metascore?" [Online]. Available: https://metacritichelp.zendesk.com/hc/en-us/articles/15456550496023-How-do-I-interpret-a-Metascore

[2] "How do you compute METASCORES?" [Online]. Available: https://metacritichelp.zendesk.com/hc/en-us/articles/14478499933079-How-do-you-compute-METASCORES

[3] "A Latex Template for Independent Work Reports." [Online]. Available: https://www.cs.princeton.edu/courses/archive/www-coursefiles/iw/IWreport.zip

[4] "Machine Learning and Data Science Seminar Website." [Online]. Available: https://www.cs.princeton.edu/courses/archive/fall24/cosIW03/index.html

[5] "Matplotlib dates.date2num Documentation." [Online]. Available: https://matplotlib.org/stable/api/dates_api.html#matplotlib.dates.date2num

[6] "Matplotlib dates.set_epoch Documentation." [Online]. Available: https://matplotlib.org/stable/api/dates_api.html#matplotlib.dates.set_epoch

[7] "Matplotlib Documentation." [Online]. Available: https://matplotlib.org/

[8] "Metacritic." [Online]. Available: https://www.metacritic.com/

[9] "Metacritic About Page." [Online]. Available: https://www.metacritic.com/about-us/

[10] "Metacritic Nintendo World Championships: NES Edition Page." [Online]. Available: https://www.metacritic.com/game/nintendo-world-championships-nes-edition/

[11] "NumPy Documentation." [Online]. Available: https://numpy.org/doc/stable/

[12] "Overleaf Tables Documentation." [Online]. Available: https://www.overleaf.com/learn/latex/Tables

[13] "pandas DataFrame.corr Documentation." [Online]. Available: https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.corr.html#pandas.DataFrame.corr

[14] "pandas Documentation." [Online]. Available: https://pandas.pydata.org/docs/

[15] "Python Documentation." [Online]. Available: https://docs.python.org/3/

[16] "Scikit-learn Documentation." [Online]. Available: https://scikit-learn.org/stable/

[17] "Scikit-learn ensemble.RandomForestRegressor Documentation." [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html

[18] "Scikit-learn inspection.permutation_importance Documentation." [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.inspection.permutation_importance.html#sklearn.inspection.permutation_importance

[19] "Scikit-learn linear_model.Lasso Documentation." [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.Lasso.html#sklearn.linear_model.Lasso

[20] "Scikit-learn linear_model.LinearRegression Documentation." [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html#sklearn.linear_model.LinearRegression

[21] "Speedrun.com." [Online]. Available: https://www.speedrun.com/

[22] "Speedrun.com About Page." [Online]. Available: https://www.speedrun.com/about

[23] "Speedrun.com Google Snake Page." [Online]. Available: https://www.speedrun.com/snake_game

[24] "Speedrun.com Nintendo World Championships: NES Edition Page." [Online]. Available: https://www.speedrun.com/nwcnes

[25] "Which game critics and publications are included in your calculations?" [Online]. Available: https://metacritichelp.zendesk.com/hc/en-us/articles/14483198627607-Which-game-critics-and-publications-are-included-in-your-calculations

[26] E. Erdil, "Power-Law Trends in Speedrunning and Machine Learning," 2023. Available: https://arxiv.org/pdf/2304.10004

[27] J. Secrest, "From College to the NFL: Draft Prediction through Regression," 2022. Available: https://www.cs.princeton.edu/courses/archive/fall24/cosIW03/jsecrest_written_final_report.pptx.pdf

[28] J. Sevilla, "Analysis of World Records in Speedrunning," 2021. Available: https://docs.google.com/document/d/1hgnVxyjR-EbYT6QDVRAlN87GjXmUDH8M12USwLODrp4/edit?usp=sharing

[29] D. Walker, "How to Write an Independent Work Paper," 2024. Available: https://www.cs.princeton.edu/sites/default/files/2024-12/fall-2024-how-to-write-an-iw-paper.pdf

A. Appendix

A.1. Estimate of Definition of Number of Active Speedrunners

The period of time that a speedrunner is required to submit a speedrun during in order to be considered an "active" speedrunner can be estimated based on the game released most recently in the collected dataset. *Nintendo World Championships: NES Edition* was released on July 18, 2024 according to Metacritic, and its speedrunning data was collected on September 27, 2024[10]. At the time of data collection, the number of active speedrunners was equal to the total number of speedrunners. While it's possible that every person who submitted a speedrun within the first ten days of the game's release submitted another speedrun after that initial ten day period, it's likely that at least one did not and only submitted a few initial speedruns at the game's release. However, this means that two months passed and the number of active speedrunners did not decrease despite one person likely not submitting any new speedruns during this time, and this indicates that the time period for being an active speedrunner is submitting a speedrun within two months or more.

A.2. How Sets of Platforms were Chosen for the Additional Features Regarding What Platform a Game was Released On

The sets of platforms that a game can appear on, although somewhat arbitrary, were created with the intention of grouping together consoles from popular video game console creators. Thus, the platforms which had the Playstation name or seemed to be an abbreviation consisting of the Playstation name were included in the corresponding feature, the platforms which had the Xbox name or seemed to be an abbreviation consisting of the Xbox name were included in the corresponding feature, and the platforms that I was familiar with which either were Nintendo platforms or were abbreviations of Nintendo platforms, such as the DS or Wii, were included in the corresponding feature. While the platforms listed in these sets do not comprehensively cover every platform available on Speedrun.com, every game in the dataset was on at least one of the platforms such that it would be in one of the sets, and this was considered to be a sufficient method of creating a feature to encapsulate the information of what platform the game released on.

A.3. Guidelines for Collecting Data

The guidelines for collecting data were:

- The game must be available on both Speedrun.com and Metacritic under the same title.
- Speedruns for the game must follow a decreasing format of better world records taking less time. One example of a game which does not follow this format is *Google Snake*, where the better speedruns seem to increase in time rather than decrease, likely due to a difference in how scoring works in the game[23].
- The game must have at least ten world records, as described in Section 3.2.
- Metacritic's Metascore and User Score can be different for the same game depending on what platform it is on[8]. Thus, when selecting a value for both of these features, I chose the maximum of each, even if the maximum for Metascore and maximum for User Score were on different platforms. This was primarily just a method of maintaining consistency, and another way of determining the critical reception of the game from these various scores could have been chosen.
- If any data for the additional features is not present, the game was not chosen. The only exception to this was for games with an "Unknown" value for the "Date added" feature because this value is estimated if it is not given, as described in Section 4.1.
- Any game pages listed as "Category Extensions" on Speedrun.com were not included, since these may not represent the speedruns corresponding to the standard game page for that game very well[21].
- Metacritic also lists all of the platforms that a game is available on[8]. If Speedrun.com and Metacritic don't have the same consoles listed, a game was listed as being on a platform if either site has it listed.
- If the year of release for a game on Speedrun.com is different from the one listed on Metacritic for a game of the same name, the game was not included because it may refer to a different game.
- If the "Date added" is less than one year and thus listed in months on Speedrun.com, this

feature was recorded as zero years. The only example of this in the dataset was *Nintendo World Championships: NES Edition*[24]. Since the rate of speedruns submitted feature divides by the "Date added" and thus would not produce a valid rate when dividing by zero, the rate of speedruns submitted was also set to zero for this game.

- If the "Date added" is greater than the years since release year, it was not included because it may indicate that different versions of the game have been released and the data on the page may not be representing the same game.

The following points were not necessarily strict guidelines, but rather indicate how the data was collected for organization purposes.

- Games for data collection were roughly chosen in order from the "Games" page on Speedrun.com when sorted by "Most runs," given that they satisfied the described guidelines[21].
- Only world record data from the category which appears as the default one when clicking on the page for a game was collected. This was because of the focus on having different games to analyze the additional features.

A.4. Feature Analysis on Models which were Not Determined to Perform the Best

The residual plots for the date of the world record and the world record time were actually created for all general models, the coefficients were printed and graphed on a bar graph for all general models which used linear regression or lasso regression and included the additional features, and permutation importance was performed on all general random forest regression models due to not having coefficients corresponding to the features. However, the resulting feature analysis and trends would likely be most reliable for the model which was determined to perform the best. For this reason and because there were many models and therefore many additional graphs, these extra graphs and values were not reported in this paper and can instead be viewed alongside the code used for implementing this project at the following GitHub Repository: `https://github.com/aidanward0804/speedrunning_prediction_research.git`.

A.5. Supplemental Feature Analysis Data of Step Through Lasso Model with Additional Features

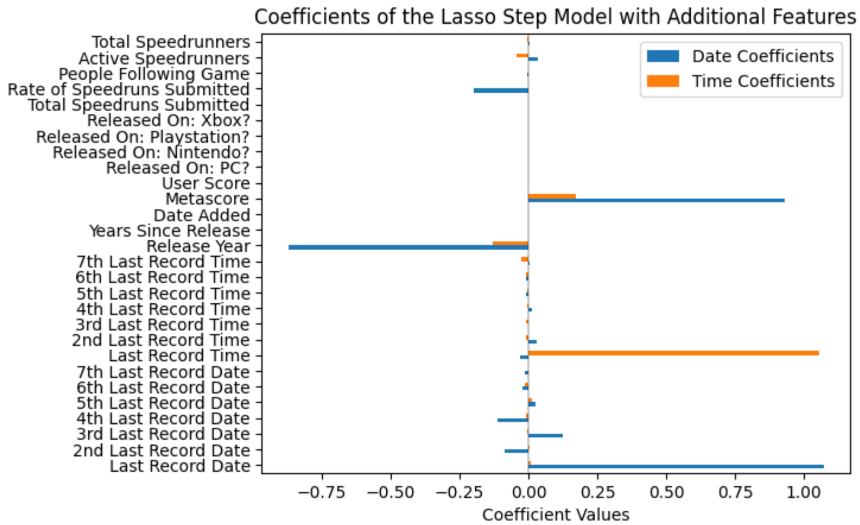

Figure 8: Coefficients of All Features in Step Through Lasso Model with Additional Features

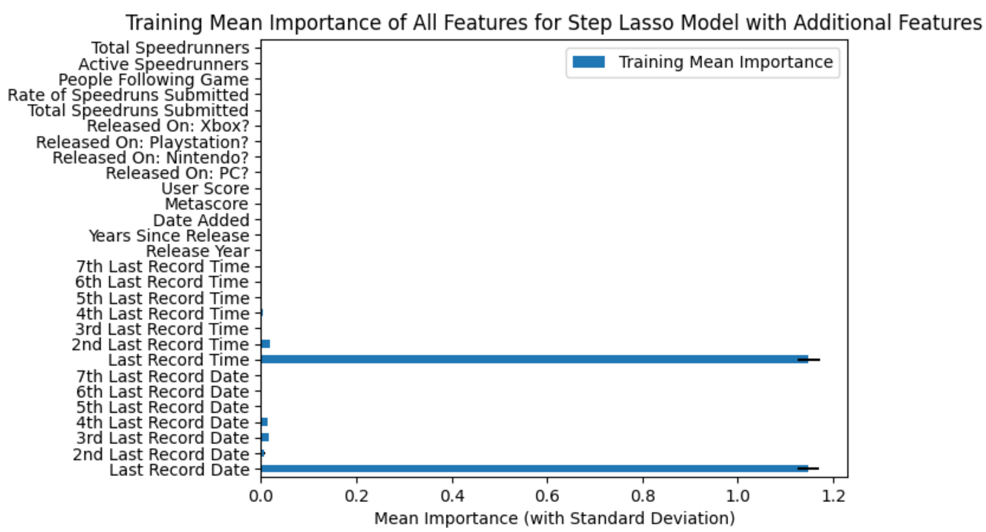

Figure 9: Training Mean Importances of All Features for Step Through Lasso Model with Additional Features

If a numerical value was large enough such that it could be represented by a decimal number with two decimal places, as the values for the rest of the paper have been represented, it was represented in this way. Otherwise, powers of ten were used to represent the coefficient's value. For values of

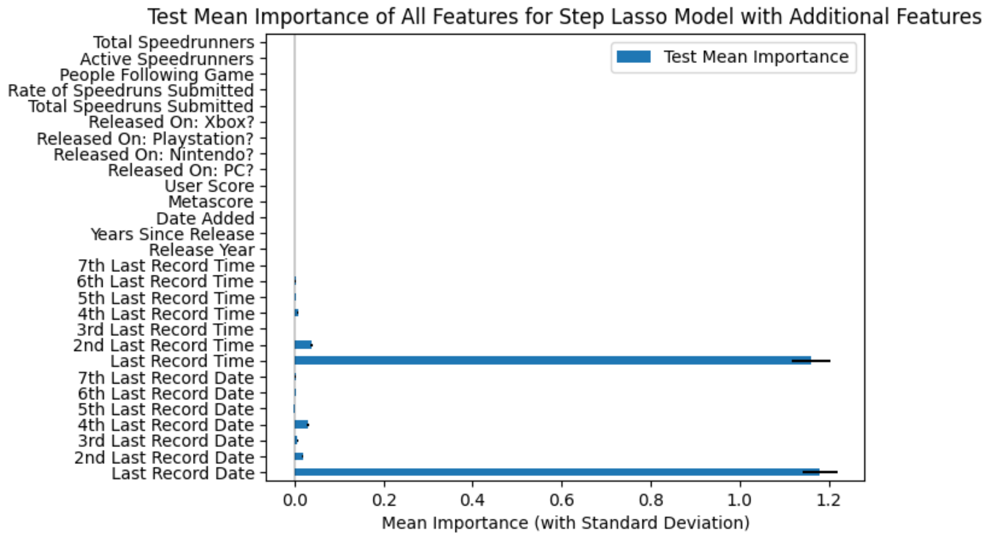

Figure 10: Test Mean Importances of All Features for Step Through Lasso Model with Additional Features

0.00 or -0.00, this meant that the value was listed with a sign when output by the code, but it had no other digits than zero.

Additional Feature	World Record Date Coefficient	World Record Time Coefficient
Release Year	-0.87	-0.13
Years Since Release	3.4×10^{-16}	3.1×10^{-16}
"Date added"	0.00	0.00
Metascore	0.93	0.17
User Score	-0.00	0.00
Released on PC?	0.00	-0.00
Released on Nintendo System?	-0.00	0.00
Released on Playstation System?	-0.00	-0.00
Released on Xbox System?	-0.00	-0.00
Total Number of Speedruns	-1.2×10^{-3}	1.3×10^{-3}
Rate of Speedruns Submitted	-0.20	-0.00
Number of Followers	-2.9×10^{-3}	8.7×10^{-5}
Number of Active Speedrunners	0.03	-0.04
Number of Total Speedrunners	2.3×10^{-3}	-2.3×10^{-3}

Table 6: Numerical Values of Additional Feature Coefficients for Step Through Lasso Model with Additional Features

Additional Feature	**Training Mean Importance**	**Training Importance Standard Deviation**	**Test Mean Importance**	**Test Importance Standard Deviation**
Release Year	6.4×10^{-5}	2.1×10^{-5}	2.2×10^{-4}	1.4×10^{-4}
Years Since Release	0.00	0.00	0.00	0.00
"Date added"	0.00	0.00	0.00	0.00
Metascore	8.1×10^{-5}	2.5×10^{-5}	4.8×10^{-5}	1.5×10^{-4}
User Score	0.00	0.00	0.00	0.00
Released on PC?	0.00	0.00	0.00	0.00
Released on Nintendo System?	0.00	0.00	0.00	0.00
Released on Playstation System?	0.00	0.00	0.00	0.00
Released on Xbox System?	0.00	0.00	0.00	0.00
Total Number of Speedruns	1.5×10^{-4}	3.7×10^{-5}	4.1×10^{-4}	1.9×10^{-4}
Rate of Speedruns Submitted	3×10^{-6}	3×10^{-6}	2.3×10^{-5}	1.6×10^{-5}
Number of Followers	1.5×10^{-4}	3.4×10^{-5}	5.2×10^{-4}	1.8×10^{-4}
Number of Active Speedrunners	3.4×10^{-5}	1.6×10^{-5}	-7×10^{-6}	9.0×10^{-5}
Number of Total Speedrunners	2.7×10^{-5}	1.6×10^{-5}	-1.9×10^{-5}	8.7×10^{-5}

Table 7: Numerical Values of Additional Feature Importances for Step Through Lasso Model with Additional Features

A.6. Comments on Future Feature Selection

A very brief amount of work was done on further analyzing the additional features for different models and can be viewed in the code used for implementing the project at the following GitHub Repository: https://github.com/aidanward0804/speedrunning_prediction_research.

`git`, where different amounts of additional features other than all or none of the additional features were used. However, this work was not discussed because it was not very thorough due to the limited time frame, but the code is included for viewing what is currently there and potential future experimentation.