Can An App's Characteristics Predict Its Success?

Nicolas Feng
Central Bucks South HS
nicolasfeng597@gmail.com
November 30, 2022

Abstract

On the Google Play Store, applications have a variety of characteristics, from category of content and cost to version and rating. In this paper, we analyze the relationship between the success of a Google Play Store app (as determined by an app's rating out of five and its review count) and its characteristics: content category, price, and amount of installs. By using linear regression, random forest, and multi-layer perceptron models, we found that an app's install count, whether or not it costs money, and the type of content all have a significant effect on its success. During modeling, we found that the random forest model was the most successful, with a training and testing RMSE of 0.079 and 0.081, respectively, and a training and testing R2 of 0.185 and 0.180, respectively. Based off these results, we have confirmed that there is a correlation between success and category of content, install count, and price. The results from this paper can inform app developers and investors about optimal statistics for the most successful applications.

1 Introduction

Many developers turn to the Google Play Store when marketing their mobile apps. The store is where Android users can get a variety of applications: newspapers, video games, and streaming services, just to name a few. Users who have installed an app can leave a rating from one to five on the store that tells other users how good the app is. They can also leave a review, a text-based critique, on the app. The Google Play Store also keeps track of other statistics, such as the app's price, how many people have installed it, and its category. To improve their first impression to potential users, developers work to maximize the rating and number of reviews on the store [1]. A better-rated/more popular app is also more likely to be featured on the store's front page, further increasing the app's appeal.

Many have conducted similar experiments, and from the same dataset we used, there have been multiple studies on graphing trends and similarities. Rather than only analyzing the relationships between two aspects of an app, however, we aim to provide a holistic view of the effects of all data points on success by quantifying the relationship between the success of an app and its Google Play Store data.

In this paper, we estimate an app's success, given its installs, whether or not it is free, and its category. This paper also details the relationships between individual variables, such as rating, installs, reviews, and price. When combined to project an app's overall success, these analyses can inform app developers to put an emphasis on key factors, using their estimated success to locate areas of improvement for their applications [3].

2 Related work

In an analysis of the success of Google Play Store applications, Tuckerman (2014) predicts the success of the applications using a combination of install count and rating out of five [4]. However, his success metric used a purely binary grading system, rather than a range from zero to one. Although he addresses the clusters of high installs, this decision somewhat blurs the line between semi-successful apps with hundreds of thousands of installs, and mega-popular apps with installs in the hundreds of millions. Furthermore, his calculation of success would only return a success of one if an app has over 50000 installs, a requirement that, while relevant in 2014, is outdated in current times.

The dataset used in this paper was taken from Kaggle.com, and was scraped from the Google Play Store in 2018. The dataset contains the info of 10841 applications, and consists of the following attributes:

Category
Rating
Reviews
Size
Installs
Type
Price
Content Rating
Genres
Last Updated
Current Ver
and Android Ver [2].

The primary attributes used in this paper are rating (average one to five-star rating), reviews (amount of reviews given to an app), installs (amount of times the app has been installed), type (whether an app is free or paid), and category (the general theme of an app's content – although similar, the "genre" attribute refers to a subsection of a type of category, and is a different section of data in our dataset, and therefore, they cannot be used interchangeably).

3 Exploratory analysis of dataset

We first cleaned the data, converting NaN ratings to 0, formatted installs to integers, and types to ones and zeros. We also decided on taking the log of installs, as an increase from 10 to 100 installs is relatively as important as an increase from 1000000 to 10000000. On inspection of the dataset, we removed two datapoints with the names "Life Made WI-Fi Touchscreen Photo Frame" and "Command & Conquer: Rivals," the first having shifted values due to a missing category, and the second being mostly empty, with values not conforming to the standard formatting, and therefore unable to be put through the cleaning functions.

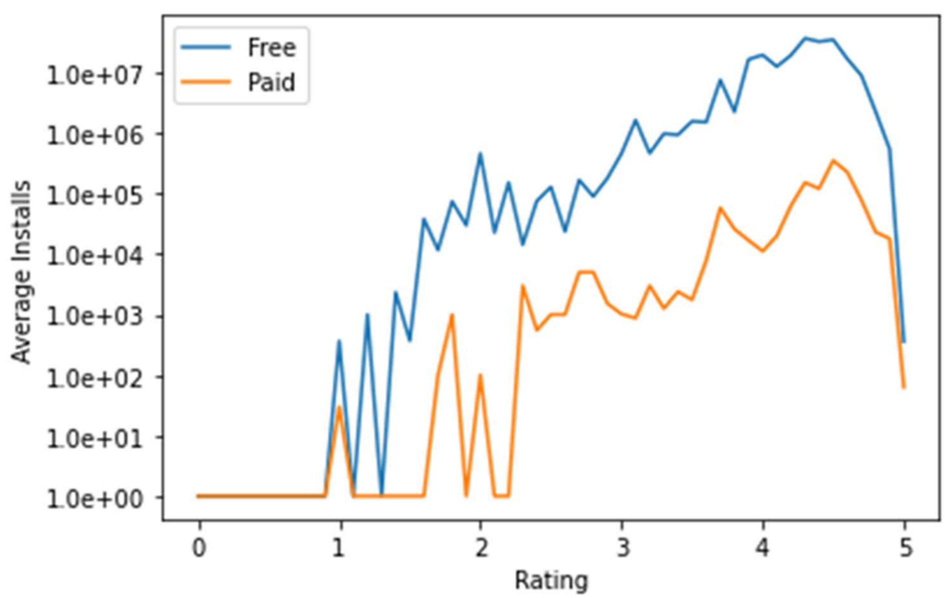
To facilitate the modeling process, we converted rating and reviews into one metric: success. One success metric streamlined the process, as univariate linear regressions cannot have multiple dependent variables. The success metric used in our analysis converts an app's rating and review count into a float from 0 to 1, where an app with zero success would have a rating of zero and zero reviews. An app with one success would have a rating of five and have the highest number of reviews in the dataset, which is 78158306. Because an app with a higher rating is generally more popular and successful than one with just a high review count, we weighed rating three times more heavily than reviews, for a max score of 0.75, with a max value of 0.25 for reviews. We took the log of reviews, with justification similar to the log performed on installs. However, since log(x) returns a negative number from (0, 1), we made it the exponent of two to ensure a positive number. Using this info, the success equation becomes:

$$\text{Success} = \frac{\text{Rating}}{6.\bar{6}} + \frac{2^{\log(\text{Reviews})}}{950.784}.$$

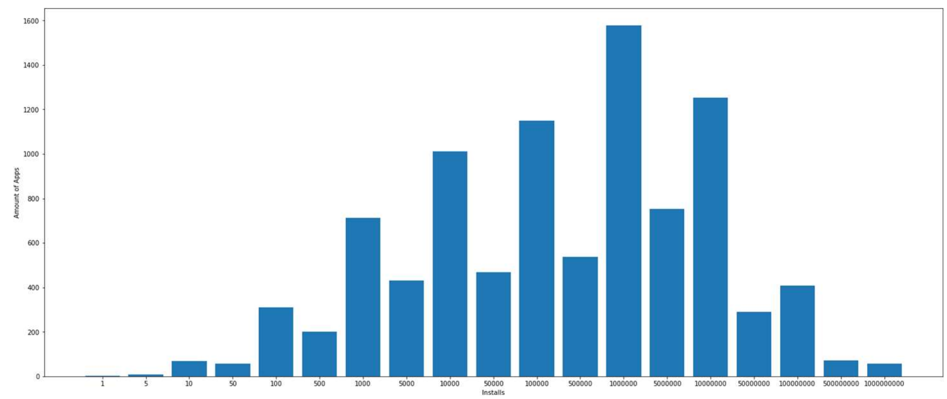
Our initial knowledge of the subject led us to believe the category, install count, and whether an app is free/paid would have the highest impact on an app's success. To confirm this, we used a one-way ANOVA with the success of an app and its installs, type, and category to test the null hypothesis, resulting in a p-value of, respectively:

0.0
0.191
7.834e-84

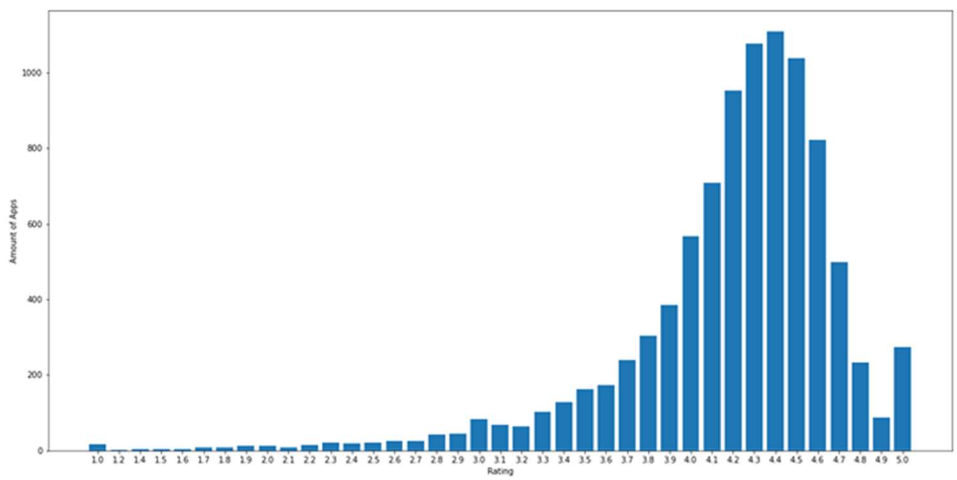
With the low p-values for installs and category, we confirmed our hypothesis about their significance. However, the ANOVA test led us to realize that whether an app was free/paid had less significance to success. Nevertheless, we decided to keep type in future models, because of its effect on installs:



To build the foundations for future observations, we first graphed the number of apps with a certain number of installs.



Similarly, this was also done with rating.

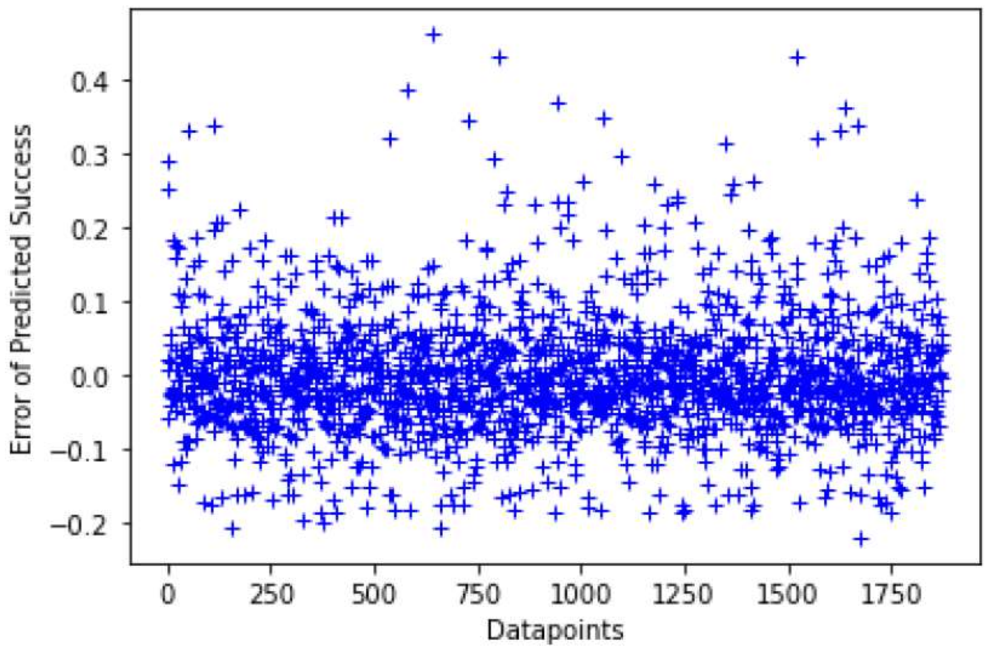


The average rating of an application is around 4.4 stars. However, we noted that while a higher review count implies a higher rating, a higher rating does not necessarily mean it has many reviews. For example, an app with one or two reviews might have a 5-star rating, while an app with hundreds of thousands might have a rating of 4.4. Content rating, storage size, and application name had no significant effect on success.

4 Modeling app success

4.1 Multiple linear regression

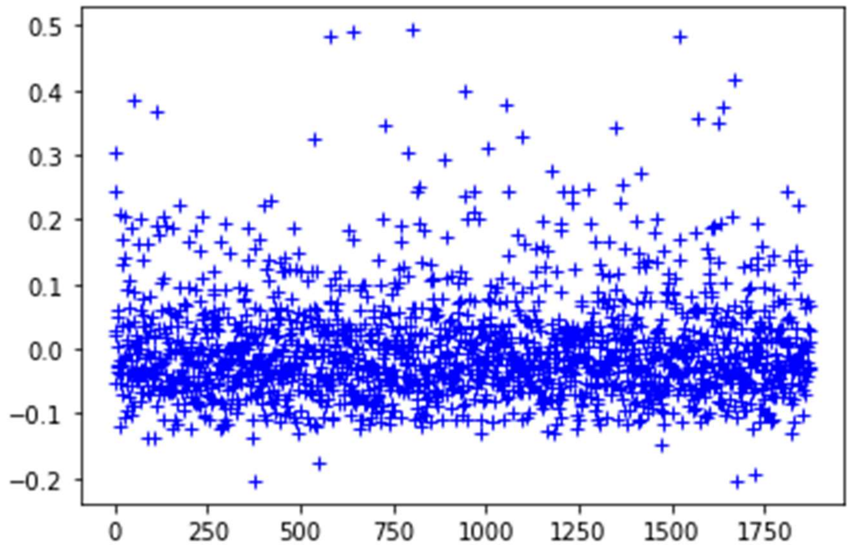
To confirm our hypothesis, we first used a multiple linear regression, with a one-hot encoder for the genres. We separated the data into a training and test set with an 80-20 split. The coefficients of installs and type are, respectively, 2.256e-10 and 1.724e-3, while the coefficients of the genres are around -1.5e-2. The RMSE of the training and testing data, respectively, is 0.08 and 0.082, and the R2 of the training and testing data are, respectively, 0.158 and 0.155. Plotting the error of each data point results in the following scatter plot:



Based on the scatter plot of the error of predicted success, this model is slightly optimistic in its calculation of success, with a maximum error of 0.5 and a minimum error of -0.17. While the R2 value could be higher, a 0.084 value demonstrates a correlation between success and installs, price, and genre. However, we decided to continue the analysis by using more models to get higher predictability.

4.2 Random forest

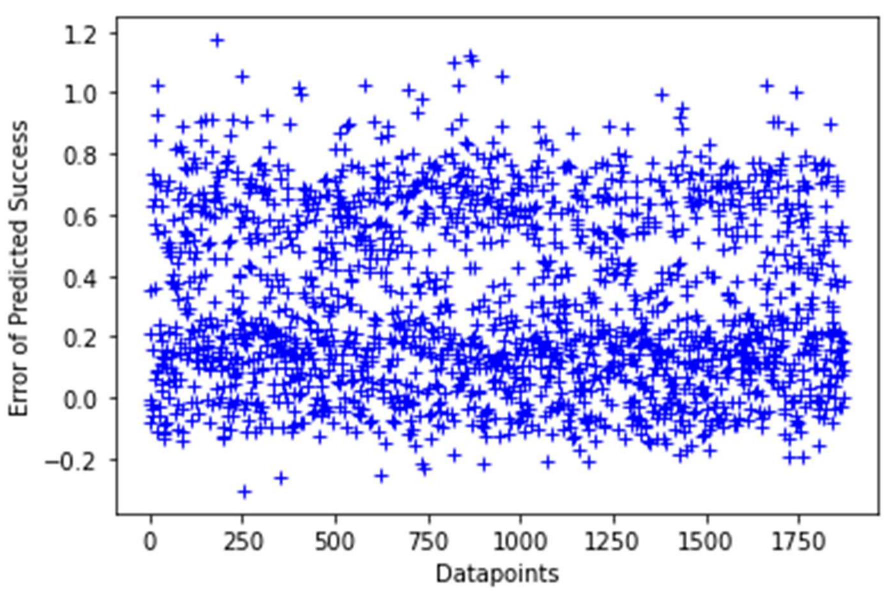
Next, we used a Random Forest, utilizing the training and testing data from the previous models. We also used a grid search for optimal estimator count, depth, and sqrt vs. log2 for determining max features, resulting in 300, 5, and sqrt, respectively. This model has an RMSE of 0.081 and an R^2 of 0.153. The error of each data point is as follows:



Similar to the multiple linear regression, this model is also skewed, having predicted higher successes than reality. It has a higher predictive power, however, with the training and testing RMSE being 0.079 and 0.081, respectively, and a training and testing R2, respectively, of 0.185 and 0.180.

4.3 Multi-layer perceptron

Finally, we used an MLP with a grid search for hidden layers and learning rate, resulting in two hidden layers and an inverse scaling learning rate. Being the most complicated model, however, it was the most sporadic in results, resulting in negative R2 values and up to a 2.5 RMSE. Overall, it was the most inaccurate. One iteration's error is as follows:



5 Conclusions

In conclusion, the random forest had a more accurate estimate of the success of an app than the linear regression, having the lowest RMSE (0.079 training, 0.081 testing) and highest R2 (0.185 training, 0.180 testing). However, both models were slightly skewed, overestimating apps' success. The main inhibiting factor of the study was data – the dataset we used only has three relevant data attributes. A more comprehensive dataset would, ideally, include factors such as whether an app has in-app purchases, consistent updates, and ads. Furthermore, more precise data on points like installs would be beneficial. However, this dataset contains categories relevant to other studies – further analysis of the data could predict several variables, such as the optimal name for an app in a certain genre and a comparison of the size of an application and its install rate. Regardless, we conclude that the characteristics of an app, mainly price, amount of installs, and genre, allow for the prediction of an app's success, a combination of its rating out of five, and its review count. The results of this paper can impact the decisions of app developers looking to optimize their app's success – from the investment into certain categories of apps, to the type of monetization of their app, our paper can provide deeper insights into the creation of a hit Google Play Store application.

References

[1] A. Bhandari and S. Bimo. Why's everyone on tiktok now? the algorithmized self and the future of
self-making on social media. Social Media + Society, 8(1):20563051221086241, 2022.
[2] Lavanya. Google play store apps. https://www.kaggle.com/datasets/lava18/google-play-store-apps, 2018.
[3] M. Pinheiro, M. Serra, and N. Pereira-Azevedo. Predictors of the number of installs in psychiatry
smartphone apps: Systematic search on app stores and content analysis. JMIR Ment Health, 6(11):e15064, Nov 2019.
[4] C. Tuckerman. Predicting mobile application success, 2014.