Analyze the factors affecting the rating of the application on the Google Play Store

Nguyen Thanh Thien Qua[1],Nguyen Hieu Nghia[2], Ngo Thi Phuc[3], Nguyen Thi Nguyet[4]

[1]20521783, [2]20521654, [3]20521765, [4]20521689 - KHDL2020

July 5, 2022

Abstract

App data on Google Play has great potential to drive successful app businesses. Useful insights can be drawn to help developers capture the Android market. Therefore, with this topic, we will conduct some statistical analysis of the data collected from CH Play to find out market trends in order to help developers navigate as well as find the right path to develop android apps. Whenever a user browse or search for apps on Play store, a list of apps are shown to the user in which each app contains the app name along with its rating. Usually the user prefers to download highly rated apps because highly rated apps reflect users' satisfaction. In order to gain high ratings, app developer uses different techniques and tweaks other than the app quality its self. However, there is no scientific approach to finding out the real impact of using catchy headlines or anything else to achieve higher ratings. factors use variable importance. For this purpose, the real-world Google Play store apps data set is used in this paper to determine the significance of these factors. To identify important variables, Random Forest, Linear Regression Model is used. The performance of the model was evaluated using standard performance evaluation techniques. The results show that some factors are more significant and affect app ratings.

1 Introduction

Mobile apps are becoming more and more popular. Mobile app stores like the Google Play store contain millions of apps. Among those apps, some have billions of downloads and active users. These apps are related to different categories including games, communication, books, business, news, sports and many more. Each category has a large number of applications. In the Google Play store, each app is usually identified by its name and rating. App rating is the average rating of all ratings given by users. Before downloading an app, people usually prefer to download apps with high ratings because apps with high ratings are usually of higher quality than the rest. In this regard, A.Montazami et al [4] performed a detailed analysis of the intentions of mobile phone users. This study shows that app ratings have a valuable effect on the user's intention to download and use an app. Therefore, in order to gain better ratings, companies use different techniques to increase the ratings of their apps on the Google Play store. For this purpose, many companies even use fraudulent and deceptive practices to gain more ratings for their apps, so that their apps can show more in the Play store [7]. However, to our knowledge, there is no scientific approach to finding the relationship between ratings and other factors of the app.

It is important to analyze the different elements of the apps on the play store and find out if there is a real connection between these factors and the app. For this purpose, these elements of the Play store app should be tested using Machine Learning and data analytics. This will help to find a real match for these factors and rate the app.

In this project, our sole goal was to determine the influence of various factors of apps on the Google Play store on app ratings. Also in this study, different linear and non-linear regression models were used along with standard evaluation characteristics to evaluate performance. For this purpose, we take real-world data of apps on the Google Play store and use various Machine Learning Models to find the real relationship between these variables and app ratings. We also perform detailed statistical analysis in terms of app category, app name, app size, install count, app type, content rating, and more. with the overall app rating to find the correlation between these other dimensions and their value in terms of app ratings. For this purpose, we use the importance of the Random Forest variable to find out the importance of each variable, we use a Linear Regression model to find the importance of different factors. To evaluate the work, we use Mean Square Error and other performance evaluation techniques to evaluate the performance of our findings.

2 Literature review

As data availability, completeness and accuracy is a big deal in the mobile app market. The Google Play Store ranks each app published in the store. Google's overall app rating system, uses a very complex infrastructure to rate apps. This is why google app rating system ranks higher than other app stores and is so successful. Google Play is a superstar marketplace due to its popular products and apps. Google Play has a good and clear app rating system. apps they like, or look for apps with better ratings. Since the advent of mobile app stores, there are a number of researchers working in the field of mobile app rating prediction and that help find the correlation between app ratings and other factors. Researchers have been working in different power sources to find ways to get better ratings in apps. Some researchers focus on user reviews, some researchers target application properties and features, some researchers work in the area of good software engineering practice than.

However, all of these domains are important in their place, but often application properties are analyzed by researchers to find the relationship between app rating and its attributes. In this section, we would like to introduce some contemporary research works in this area. Tian et al [6] performed a case study using statistical analysis to rank the different factors of apps that effect the app ratings, the size of an app, promotional images of an app are the most influential factors of high-rated apps. Similarly, Ahsan Mahmood [3] investigated the relationship between price, rating and popularity in the blackberry app store and their findings showed that there is a strong correlation between customer ratings and popularity. The researchers compiled a detailed analysis of apps from Android and Apple apps and performed a quantitative analysis of the app's properties and their impact on apps in the app stores different, Ali et al [1]. Furthermore, Liang et al [2] used a feature-driven matrix database to predict mobile app ratings. Searchers discover the factors that influence app ratings for the apple app store and recommend rating prediction models for different apps. They looked at several variations in their model, including package size, app release date, category popularity, and more to find out the importance of these factors, Picoto et al [5].

While different researchers focused on a number of attributes that influence the app ratings, there are still some simple but important factors that are not yet analyzed by the researchers. Moreover, most of the researchers target a few number of attributes and find the importance of those attributes for predicting the apps rating. Usually researchers took the default attributes of the apps and performed their analysis on those attributes. Therefore, there are many attributes that can be computed for each app and its effects can be analyzed for rating of that app. Therefore, to fill the research gap, we conduct a detailed study in which we take a large apps store data set, we use a number of default app features as well as compute a number of attributes for each app and find the importance of these attributes in app ratings .

3 Research methodology

This study determines the importance variable for app ranking according to statistical analysis and regression method. The model is divided into two parts, in the first part, different variables are identified and calculated. These variables are then tested on regression models. The performance of these models was assessed deterministically using different performance evaluation techniques, and the most influential and effective variables were calculated in this section.

3.1 Dataset collection

The dataset was found on Kaggle and the information was gathered through searching over 10,000 apps on Google Play. Since Google Play uses modern techniques like dynamic page loading using JQuery, this makes it harder to find data. Each app has values like category, name, star count, size, download schedule, etc. shown in Table 1. Although Google Play acts as a digital media store, For this analysis, we are only interested in Android apps, so the data includes only mobile apps.

Variable Name	Explanation
App	App Name
Category	Category (Type of application)
Rating	Average rating for the app (From 1 to 5)
Reviews	Number of User Reviews
Size	Dimensions of the application
Installs	Number of Installs
Type	Apps Free or Paid
Price	Price of the application
Content Rating	Age Group Targeted Application
Last Updated	Last Updated App
Current Ver	Current Version
Android Ver	Lowest Android Version required by the app
NameLength	Total number of characters in the title of an app are computed for each app

Table 1: Explanation of variables.

3.2 Preprocessing data

The current data set is quite full of information, but there are still some minor errors such as repeated data points but not many and some missing data. Because we cannot recollect, we will re-filter with python tools and then delete the missing data points to avoid adverse effects on the analysis results.

3.3 Regression models

- **Random forest:** Random forest regression is applied to all the variables The results of random forest determine the importance of all the variable and their influence on the rating. The results of random forest regression are evaluated using Mean Square Error. Random forest model is the first model that is applied to the data set and the results of Random forest classification are computed for a number of variables to find the importance of these variables.

- **Linear Regression:** Linear Regression model is also used to find the variable importance of different variables with ratings. Although linear regression model is a simple regression model, it sometimes produces better results than other complex models. In this model, only numeric values are used. Therefore, when this model is applied to the data set, only the numeric variables are considered.

3.4 Performance evaluation

The performance of different models is evaluated using different performance evaluation techniques. The details are given as under.

- **%IncMSE:** %IncMSE is the most robust and informative measure. The higher MSE shows that the variable is more important while the lower number shows that variable is less important.

- **p_value:** During a hypothesis test, a p value is used to deter-mine the significance of the results. A small p value (less than0.05) indicates strong evidence against the null hypothesis.

3.5 Identification of most important variables

On the basis of results of different models, we identify the most important variables that influence the overall ratings of the app. As app rating is divided between 0 to 5 on the scale of0.01, even a small change in the rating means a lot. Moreover,as a huge number of apps available in the play store, a significant difference in ranking of an app change the app position in play store by a big margin.

4 Question Analytical

- Which app category (Category) has the highest rating?
- Is there a real difference between age-restricted apps and non-age-restricted apps?
- Is there a significant rating difference between the paid and free apps?

4.1 Preliminary analysis of the data set

Are the categories of apps different in ratings? To answer this question, let's take a look at the EDUCATION and ART AND DESIGN categories, to see if the two have a clear difference in ratings. The results of Table 2 show that when randomly taking 10 ratings of applications, the average rating of educational applications (4.3) gives higher results than the remaining applications (4.1), but the difference This difference is not significant. To know exactly whether there is a difference or not we will check in the following section. Likewise, paid apps and free apps have a negligible difference and don't show a clear difference between the two.

EDUCATION	4.4	4.1	4.4	4.3	4.1	4.1	4.4	4.0	4.2	4.3
ART AND DESIGN	4.2	4.3	4.4	4.4	4.5	3.7	4.1	4.0	4.4	4.1

Table 2: Rating 10 random apps.

4.2 Analytical Design

In terms of the characteristics of the study, this study is an observational study because we only observe and record the status of the characteristics, characteristics, and events occurring in the study without any effect. affect the research object. Since we do not have any samples to compare with, the above observational study is simply a descriptive study. As mentioned above, the data is collected by searching the web itself, and because of technical issues, this is quite difficult. So the elements in this data set are completely random and the probability of being selected is the same for all applications. Therefore, the sampling method here is probability sampling and the sampling method is a simple random sampling method. (Sinple random sampling).

4.3 Reflect on the question

Before we begin to analyze the data through visualization and statistical testing, we will predictably answer the questions posed: First, as we all know, now education and entertainment-related applications are very popular, so subjectively we can predict that these applications will often have higher ratings than applications in the remaining fields. Of course, in practice, be limited or forced to authenticate the age when using an application is quite inconvenient for users. So perhaps apps that aren't age-restricted will usually have higher ratings. Finally, an application that has to pay a fee to be able to use it will be more carefully invested in the interface, in terms of user experience as well as always having more tests with its users, and with free applications. Fees will often have a lot of advertising, so customer satisfaction will be worse than paying. Therefore, the rating of paid applications will be higher than that of free applications.

4.4 Data analysis

For the scope of this analysis, we use two types of histograms: boxplots to compare fluctuations and outliers of group data, histograms to describe the distribution of individual samples and to compare probability distributions sample with a standard probability distribution.

- **Which category (Type) of apps has the highest rating**

In the scope of this analysis, we are only interested in some popular and prominent categories in practice such as GAME, BUSINESS, EDUCATION, ENTERTAINMENT, LIFESTYLE, DATING.

1. Define relevant variables:

Explanatory variables: Category, Categorical type. Response variable: Rating, quantitative type.

2. Eploratory Analysis:

To observe the rating difference of different categories, we observe the figure 1a showing: The median rating of EDUCATION applications is the highest (4.4 stars) and the lowest in the list DATING item (4.0 stars), the remaining categories are all worth 4.3 stars. This difference is not really significant, the difference between the highest and the lowest category is only 0.3 stars. Regarding the rating of different categories, there are unequal ranges, the widest is DATING and LIFESTYLE, ranging from 2.3 stars to the highest 5 stars. It shows that these two categories have quite a lot of reviews and the quality of the reviews is quite large and unstable, possibly due to the quality of the applications in this category from many different manufacturers giving customers the best experience. It can be a good experience and it can also be a bad experience. In contrast, the narrowest range is EDUCATION and ENTERTAINMENT from 3.7 stars to 4.7 stars. This range is quite small, the stable rating is a positive reflection of the customer's star count. Regarding outliers, both EDUCATION, DATING, and ENTERTAINMENT are absent, the remaining categories have quite large outliers, especially GAME has an outlier value of 1 star while the highest rating value of this category is 5 Well, we can see that this category is completely unstable with feedback ratings from customers. Similarly, this is further demonstrated in the figure 1b which shows the rating distribution by category in terms of histograms and distribution curves. In which, the red curve is the actual rating distribution and the blue curve is the normal distribution density function, it is easy to see that the above samples have the probability density function close to the normal distribution. This gives us more peace of mind when using the central limit theorem to perform statistical tests.

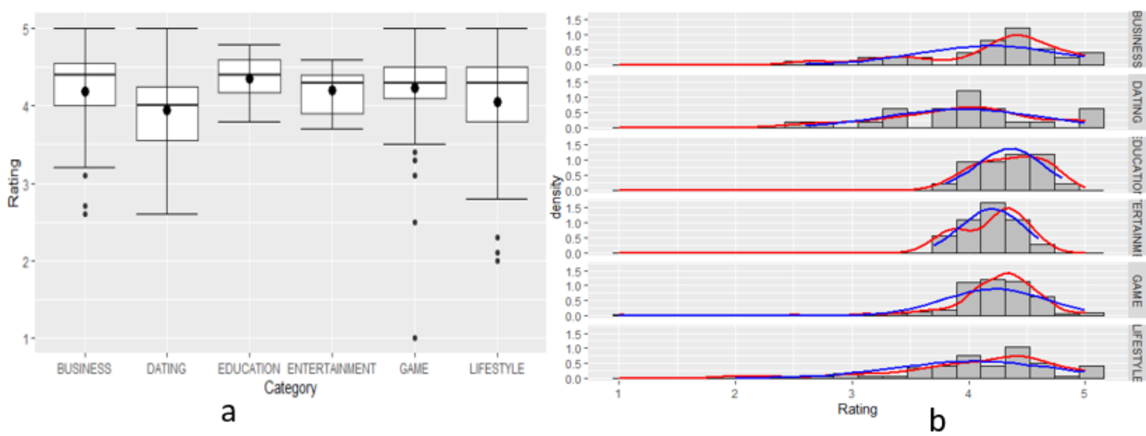

Figure 1: a.Boxplot chart comparing ratings across application categories, b. Histogram and probability distribution density function

3. Interpretation:

Current statistical inference belongs to the category of comparing more than 2 sample means. The appropriate statistical test for this question is ANOVA, which checks whether there is a difference between the ratings of the categories with 95% confidence. The plots are defined as follows:

- Null hypothesis: H_0: There is no difference in rating between different categories of apps
- Alternative hypothesis: H_1: There is a difference in rating between different categories of applications.

4. Test results and conclusions

Based on the results presented in figure 2, the ANOVA F-test indicates that there is substantial statistical evidence to conclude that the true mean rating among all categories is unequal. when p_values = 1.77e-06, this value is too small for significance level 0.05 so we can reject H_0. Based on the HSD table, we can see that the sample pairs of cps p_values less than 0.05 are statistically different in rating.

Exploratory analysis reveals differences in average ratings across different categories of apps. Specifically, we can see that the EDUCATION-BUSINESS, EDUCATION-DATING, GAME-DATING, LIFESTYLE-DATING groups have statistically clear differences. Statistical testing yields a very small value of p (basically close to zero), indicating that the evidence provided by the data is strong enough to disprove H_0 and conclude that there is a difference in ratings between different categories of applications. This is quite true with the fact that online dating applications or online businesses are often less interested at the present time, with only a few exceptions such as Tinder, Dating on Facebook, ... instead, games or educational applications receive quite high ratings such as Lien Quan Mobile, Ghensin Impact, ELSA Speak,...
The data, as well as the results of the above study, do not provide evidence that we Every time an educational or entertainment app is released, user ratings will be high, as this is an observational analysis, the only case where the word has a causal relationship that can be inferred is when the research is a Random experiment, however, to collect the entire list of applications from CHPlay at the present time is still very difficult. This difference can be understood because most Android users use their phones for entertainment or study, so apps in this category are often more satisfying to users. This is completely in line with our expectations and can be motivating to anyone who is looking to develop an entertaining or educational application.

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## Category      5   6.87  1.3734   7.252 1.77e-06 ***
## Residuals   350  66.29  0.1894
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##                                diff          lwr        upr     p adj
## DATING-BUSINESS         -0.24090909 -0.531000866 0.04918268 0.1662106
## EDUCATION-BUSINESS       0.31794872  0.006755921 0.62914151 0.0420021
## ENTERTAINMENT-BUSINESS   0.06333333 -0.251680155 0.37834682 0.9925324
## GAME-BUSINESS            0.18717949 -0.026297010 0.40065598 0.1232011
## LIFESTYLE-BUSINESS       0.08333333 -0.180155606 0.34682227 0.9447610
## EDUCATION-DATING         0.55885781  0.231841991 0.88587363 0.0000219
## ENTERTAINMENT-DATING     0.30424242 -0.026411302 0.63489615 0.0912889
## GAME-DATING              0.42808858  0.192142391 0.66403477 0.0000051
## LIFESTYLE-DATING         0.32424242  0.042240989 0.60624386 0.0137452
## ENTERTAINMENT-EDUCATION -0.25461538 -0.603928432 0.09469766 0.2955536
## GAME-EDUCATION          -0.13076923 -0.392223476 0.13068501 0.7066957
## LIFESTYLE-EDUCATION     -0.23461538 -0.538280543 0.06904977 0.2339009
## GAME-ENTERTAINMENT       0.12384615 -0.142144191 0.38983650 0.7657761
## LIFESTYLE-ENTERTAINMENT  0.02000000 -0.287579372 0.32757937 0.9999688
## LIFESTYLE-GAME          -0.10384615 -0.306191865 0.09849956 0.6832541
```

Figure 2: Inspection results ANOVA

- **Is the rating of an age-restricted app any different from an age-restricted app?**

First of all, we can understand that there is no age limit, which means that the application is aimed at everyone.

1.Define relevant variables

Explanatory variable: Content Rating (categorical type). Response variable: Rating (of Quantitative type).

2.Analyze visits by:

Looking at the figure 3a, we see that the median age-restricted and age-unlimited apps are the same

(4.3 stars) and the rating ranges are quite small, both with can achieve the highest rating (5 stars), tends to lean towards the mean, is relatively stable, and this rating range is relatively equal. In terms of outliers, it seems that the ratings of these two types of applications are quite scattered and have many outliers and exist the lowest rating (1 star). It can be briefly concluded that some applications have a very high rating according to user reviews and the rest have a rather low rating for both types of applications but low rating is only a minority case, the majority still has a fairly high rating and there is no significant difference. This is clearly visualized in the figure 3b, and the red curve shows a rating distribution that closely resembles the normal distribution.

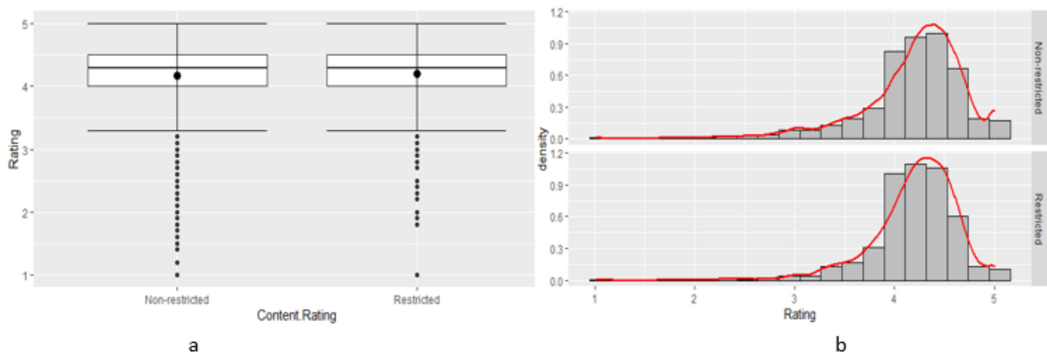

Figure 3: a.Boxplot chart comparing ratings between age-restricted and non-age-restricted categories, b. Histogram and probability distribution density function

3. Interpretation

Statistical inference for the current statistic is of the kind that compares the mean of two samples. The appropriate statistical hypothesis test for this question is a two-sample t-test for the mean (μ_1, μ_2). Hypotheses is defined as follows:

- Null hypothesis: H_0: μ_1 - $\mu_2 = 0$
- Alternative hypothesis: H_1: μ_1 - $\mu_2 \neq 0$

4. Results and conclusions

The results of two-sample t-tests with 95% confidence show that the p_values = 0.00514, which is too small for the 0.05 level of significance, we can reject H_0 and conclude that there is a difference between age-restricted and age-restricted apps, specifically age-restricted apps have a slightly higher rating than the rest.

Exploratory analysis shows that the difference in average ratings between age-restricted and non-age-restricted apps is not large. Statistical testing yields p_values that are a quite small relative to the significance level, so there is an absolute difference between these two types of applications. We see that the average rating of the age-restricted application (4.2023) is higher than the average rating of the age-restricted application (4.1663), but the difference is not significant. This proves that age-restricted applications, although focusing on a smaller group of users, can still satisfy that group of users. The data as well as the results of the above analysis do not provide any evidence that every time we release an age-restricted application, the feedback from the users will be positive, or the rating will be greater than that of an application that does not limited. This difference can be understood because of age-restricted applications such as games or adult entertainment content, these applications, although focusing on a small user group, are aimed at the right audience and serve the needs of the target audience, so they are quite satisfied with the application and highly appreciated. This does not coincide with our point of view, when we think that for every user, the rating will be high if it is popular. This is real proof to give advice to app developers.

- **Is there a significant difference in rating between free and paid apps?**

1.Identify relevant variables

Explanatory variable: Type (categorical type). Response variable: Rating (Quantitative type).

2.Exploratory analysis

Observing the graph of 4a, we find that the median, and the average rating value of these two types of applications are different, the indexes of paid applications (median: 4.4, mean: 4.26) higher than the free app (median: 4.3, mean 4.16) but this difference is not significant. About 75% of ratings are all over 4 stars, the range of data fluctuations is quite small and relatively stable and tends to be distributed upward (from 4 stars or more), all can achieve the number of stars. maximum. Both of these types of applications have many outliers, quite discretely distributed, and at the very least, both reach as low as 1 star. From here we can roughly conclude that there is no significant difference between the paid application and the free application, most of the applications can satisfy customers, so the number of stars is quite high, besides there are still some pretty underrated apps. This is more intuitive when looking at figure 4b, In this plot, the red curve is the probability distribution of the actual data, the blue line is the normal probability distribution. Contrary to the above figures, here we see that the shape of the actual data distribution curve does not match the normal distribution, our approximation of the sample distribution to the normal distribution may be biased and inaccurate, however the sample size we use is quite large so we can still use the central limit theorem.

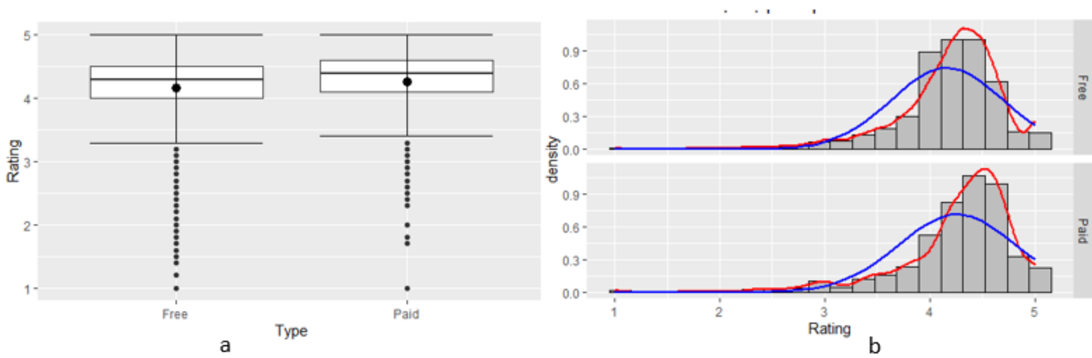

Figure 4: a.Boxplot chart comparing ratings between free and paid apps, b. Histogram and probability distribution density function

3. Interpretation

Statistical inference for the present statistic is of the type of comparing two-sample mean and the appropriate statistic test, in this case, is a two-sample t-test for two samples (μ_1, μ_2). Hypotheses is defined as follows:

- Null hypothesis: H_0: μ_1 - $\mu_2 = 0$
- Alternative hypothesis: H_1: μ_1 - $\mu_2 \neq 0$

4. Results and conclusions

Based on the obtained results, two-sample t-tests indicate that there is significant statistical evidence to conclude that the ratings of free and paid apps are indeed different when p_values are very small. (8.568e-05) relative to the significance level of 0.05, so we can reject H_0 and accept the hypothesis H_1. Exploratory analysis shows that the difference in average rating between free and paid apps is different but the difference is not significant, however, statistical testing results in very small p_values (basically zero), indicating that the evidence is strong enough to conclude that there is indeed a difference and that in particular, the rating of the paid app is higher than that of the free. This is quite in line with our reality and prediction as paid apps usually have better customer service.

The data as well as the results of the above analysis do not provide evidence that we just released a free application and that the rating from users will be lower. Maybe this difference is because the majority of users rate the service as well as the experience of the paid apps. However, this is not necessarily true in Vietnam when paid applications have very little place.

5 Model results and discussion

After computing all the variables and applying the models, the results are computed from different models. Each model has its own significance and importance. Random forest regresssion and variable importance is the most commonly used regression models to find the important variables in a data set. The results of random forest regression are shown in Table 3. The results show that Reviews, Installs, Category are the most influential variables and have high impact on the ratings of the apps. Similarly, Size, NameLength, Android Ver have little importance in terms of predicting the ratings of an app. After the random forest regression, Linear regression model is applied. For this model, only numeric values are used because a linear regression model works best on numeric continuous values. The results of Linear regression models are shown in Table 4. The results show that the correlation between app rating and variable is very low. However, there are still some important variables like Type and Size. Although, in such sorts of analysis, linear regression model doesn't compute well and usually there isn't a direct relation between a variable predictor. However, the results show that p_values are very low in most of the cases. According to the results of the Random Forest regression model, reviews, category, installs are the most important variables among the others. As random forest regression uses ensemble learning methods and aggregates many decision trees, its results have higher importance compared with other regression models. According to the Linear Regression model, type and size are important variables. However, the significance of these two variables is very low while the coefficient values of other variables are also very low. This means that the relationship between apps, rating and other variables doesn't have a linear relationship with each other. Therefore, we can say that none of the variable has a simple relationship with rating and the linear model is notable to find the real importance of any variable.

Variable Name	%IncMSE	IncNodePurity
App	5.61	207.09
Category	25.84	117.94
Reviews	51.50	333.38
Size	23.22	198.71
Installs	40.99	169.21
Type	9.13	10.47
Price	9.96	27.98
Content Rating	10.02	25.77
Last Updated	3.14	209.00
Current Ver	13.86	196.17
Android Ver	19.02	124.15
NameLength	21.67	194.81

Table 3: Random Forest Regression Results

Variable Name	Coefficient	p_value
Reviews	0.00000002944	<0.05
Size	0.001108	<0.05
Installs	0.000000000222	>0.05
Type	0.1101	<0.05
NameLength	0.0001013	>0.05

Table 4: Linear Regression Results

6 Conclusion

Based on the results of our analysis, there is statistical evidence that the "EDUCATION" category has the highest rating. This may be because educational apps are naturally pleasing to users, thus receiving high ratings. There is also statistical evidence showing that Play Store Apps with an age-restricted content rank higher than apps without age-restricted content, and that there is a significant difference in ratings between apps with age-restricted content. free or paid use. For mobile developers or those looking to develop apps, a general idea for an app can help boost Google Play ratings. The results promise that future developers can shape the Android market.

In this project, a detailed analysis is carried out on a number of variable that influence the ratings of google play store apps.Different variables are proposed and computed for this purpose and results of each variable is discussed. The results show that there are some significant variables that are able to influence the rating of apps. As app ratings have a very small scale, even a minor change in the rating can help getting abetter outcome and more downloads and visibility. Therefore, even the lower significant variable important considering the domain. The performance evaluation results show that some of the proposed variables high significance and can be used in a positive way to increase the app ratings. We also performed a detailed keyword analysis which presented a number of important points. The results show that there are some words that promise higher ratings while there are some keywords that usually mean lower ratings. These keywords also reveal that the categories from which these apps belong are also considered as important and unimportant by the users.In the future, we aim to use a much larger database and incorporate app variables as well as app reviews and reviewer variables in order to predict the ratings of an app. This multi-dimensional analysis would help in finding a much better picture of variable importance and helps us finding the factors that contribute in higher app ratings of apps. The above study is an observational study. It is possible to draw conclusions for the "population" because the sample was collected at random, but these results may not be true today because the time of data collection was quite long ago.

References

[1] Mohamed Ali, Mona Erfani Joorabchi, and Ali Mesbah. Same app, different app stores: A comparative study. In *2017 IEEE/ACM 4th International Conference on Mobile Software Engineering and Systems (MOBILESoft)*, pages 79–90. IEEE, 2017.

[2] Tingting Liang, Liang Chen, Xingde Ying, S Yu Philip, Jian Wu, and Zibin Zheng. Mobile application rating prediction via feature-oriented matrix factorization. In *2017 IEEE International Conference on Web Services (ICWS)*, pages 261–268. IEEE, 2018.

[3] Ahsan Mahmood. Identifying the influence of various factor of apps on google play apps ratings. *Journal of Data, Information and Management*, 2(1):15–23, 2020.

[4] Armaghan Montazami, Heather Ann Pearson, Adam Kenneth Dubé, Gulsah Kacmaz, Run Wen, and Sabrina Shajeen Alam. Why this app? how educators choose a good educational app. *Computers & Education*, 184:104513, 2022.

[5] Winnie Ng Picoto, Ricardo Duarte, and Inês Pinto. Uncovering top-ranking factors for mobile apps through a multimethod approach. *Journal of Business Research*, 101:668–674, 2019.

[6] Yuan Tian, Meiyappan Nagappan, David Lo, and Ahmed E Hassan. What are the characteristics of high-rated apps? a case study on free android applications. In *2015 IEEE international conference on software maintenance and evolution (ICSME)*, pages 301–310. IEEE, 2016.

[7] Hengshu Zhu, Hui Xiong, Yong Ge, and Enhong Chen. Discovery of ranking fraud for mobile apps. *IEEE Transactions on knowledge and data engineering*, 27(1):74–87, 2018.