International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 04 | Apr 2021 www.irjet.net p-ISSN: 2395-0072

Play Store App Analysis

Akhlak Ali Sunasara[1], Nancy Jaiswal[2], Suchit Poojari[3], Anil Kumar Chaturvedi[4]

[1,2,3]Student Researcher, Shree L.R. Tiwari College of Engineering, Mumbai University, Mumbai, Maharashtra, India
[4]Assistant Professor, Shree L.R. Tiwari College of Engineering, Mumbai University, Mumbai, Maharashtra, India

---

Abstract - *Software application is vital because specific software is required in almost every industry, in every business ,and for each function. It become more important as time goes on. Mobile app distribution platform such as Google play store gets flooded with millions of new applications uploaded by developers everyday. So in this project, we aim on analyzing Google play store that provides a particular app description and data such as reviews, ratings, price and number of downloads. The objective of this is to analyze the desire of the customer through the reviews provided in the feedback section and apps trend in the market to help the organization & developers. To this end, we provide an idea about app that managed to get maximum and minimum number of downloads and predicting the category of apps that is most likely to be downloaded in the coming years. Moreover, doing sentimental analysis on the apps that generated most positive and negative sentiments, sustainability of app in market on basis of previous data and current market situation. Furthermore, also analyzing the apps that has maximum downloads have they managed to get average rating so that concluding the co-relation between number of downloads and ratings received*

Key Words: Linear Regression Algorithm, Anaconda Navigator, Spyder, Python(3.7),ML Libraries And Packages , Wamp Server for Database.

1. INTRODUCTION

In today's era, the Google Play Store is the largest and most popular android app store. It is flooded with millions of applications and it provides wide collection of data on features like ratings, price and number of downloads and apps description. Many apps are being developed as apps are easy to create and its lucrative. But its important for developers to know which apps are loved by customers and are trending in market so that he develop only those apps and also there is a high competition between app providers producing similar applications. Analysing customer needs is one of the bizarre tasks in the business world today. Hence proposing analyse data to developer that what customer is likely to download, which category got the maximum downloads this all plays a crucial role in app development. Generally, customers download apps depending on number of downloads, positive reviews, negative reviews, ratings and comments. So, in this project we are going to help the users by categorizing positive, negative and neutral reviews and comments of the particular. we are going to help developer by analyzing the desire of the customer through the reviews provided in the feedback section and apps trend in the market to help the organization & developers. Also provide an idea about app that managed to get maximum and minimum number of downloads and predicting the category of apps that is most likely to be downloaded in the coming years. The dataset of google Play Store for analyzing is collected from kaggle.

The project aims at doing this with the help of a sentimental analysis and machine learning that will analyse customer needs and suggest the developers best app for developing. The analysis is acheived using the survey of the user download behaviour on the apps across all the categories on the google play store. Mobile app stores are becoming extremely lucrative. Android is expanding as an operating system and Mobile app industry is increasing in significantly and thus giving rise to more competitions to the one's that are creating applications. Hence, for a developer to know the recent trends, competition is important so that the value of their app in the store do not degrade. Google play store is a digital distribution service and it allows user to browse and download different apps. It serves as official store of apps for android operating system. Play store is additionally a platform which offers music, digital media store, books, movies and tv programs. Mobile app stores are becoming extremely lucrative. Due to the competition in the market and also expansion in order to help our developer understand what kinds of apps are likely to attract more users and what is the motivating factor for the people to download an app we analyze and research relevant data. They will be getting to know the success rate and they will get to decide what features should be added or modified and what should be maintained according to the current state of their app. Hence we found this topic interesting and convincing for our project work.

2. PROBLEM STATEMENT

The Play Store apps data has enormous potential to drive app-making businesses to success.Android is expanding as an operating system and Mobile app industry is increasing in significantly and thus giving rise to more competitions to the one's that are creating applications.Due to the competition in the market and also expansion in order to help our developer understand what kinds of apps are

likely to attract more users and what is the motivating factor for the people to download an app we analyze and research relevant data.For the app development industry where they can analyse the downloads and demand off app download in the industry.

We aim on providing doing sentimental anlysis on the apps that generated most positive and negative sentiments and sustainibility of app in market on basis of previous data and current market.

3. LITERATURE REVIEW

The Literature survey here outlines preceding researches on playstore app analysis, the algorithms and graphs used by them. The writings we present here is the work of many pertinent papers explored by us so by collecting the combination of keywords and snowballing we have improvised our project. The literature review makes us contemplate and understand the earlier innovations related to the project.

Related Work

[1] In this paper, They proposed a completely unique and automatic framework IDEA, which aims to spot Emerging App issues effectively supported online review analysis. They evaluated IDEA on six popular apps from Google Play and Apple's App Store, employing the official app changelogs as their ground truth.Feedback from engineers and merchandise managers shows that 88.9% of them think that the identified issues can facilitate app development in practice.To make the topics comprehensible, IDEA labels each topic with the most relevant phrases and sentences based on an effective ranking scheme considering both semantic relevance and user sentiment.

[2] In this context, traditional recommendation techniques are introduced into APPs recommendation. However, different from traditional context, APPs recommendation is a very unique task since people use APPs for different reasons. In this paper, they analyzed user's usage and download behaviors supported a true Android Market data to hunt useful information which may benefit APPs recommendation task.. APPs usage, Usage of APPs represents user's preferences and demands. Latent Models like matrix factorization can help to factorize user-item preference matrix into user preference vectors and item feature vectors. In their recommendation algorithm, they utilized user's usage history as user's preferences for APPs.

[3]In this paper, They used Sentiment analysis ,Lexicon based sentiment analysis is a method that can be used for determining the sentiment polarization of a review or a comment in the App Store. There are two resources needed in lexicon based sentiment analysis for Indonesian language: machine translation and lexicon resource. In this study, they compared the performance of several different combinations of machine translations and lexicon resources in order to know the best resource combination that can be used in lexicon based sentiment analysis on App Review. The result shows that the combination of Google Translate and SentiWordNet can reach the highest overall accuracy by getting the score 0.72.

[4] They demonstrate an optimization-based aggregation method for ranking extortion and ranking misrepresentation recognition framework for versatile Apps. It is divided into three parts: 1) ranking based evidence, 2) rating based evidence and 3) review based evidence, by demonstrating Apps' ranking, rating and survey practices through measurable theories tests.They used here opinion analysis for finding how much a review is positive or negative. This review score is employed to reinforce the rating score of the user and therefore the emoticons within the reviews or comments. User has provided the rating, review & comments.

[5] In this paper, They represent a large-scale comparative study of cross-platform apps.They mine the characteristics of 80,000 app-pairs (160K apps in total) from a corpus of two .4 million apps collected from the Apple and Google Play app stores. They quantitatively compare their app store attributes, like stars, versions, and costs. They measure the aggregated user-perceived ratings and find many differences across the platforms. Further, they employ machine learning to classify 1.7 million textual user reviews obtained from 2,000 of the mined app-pairs. They also follow up with the developers to know the explanations behind identified differences. they contacted app developers to understand some of the major differences in app-pair attributes such as prices, update frequencies, AUR rates and top rated apps existing only on one platform.

[6] This talk presents results on analysis and testing of mobile apps and app stores, reviewing the work of the UCL App Analysis Group (UCLappA) on App Store Mining and Analysis. The talk also covers the work of the UCL CREST

center on Genetic Improvement, applicable to app improvement and optimization.

[7] In this paper, the google play store is one of the largest and most popular Android app stores. They have used a raw data set of Google Play Store from the Kaggle website. This data set contains 13 different features that can be used for predicting whether an app will be successful or not using different features.. They conduct data modeling by using three models: Gaussian Naive Bayes Model, K-nearest neighbor model, and Decision Tree model. They also discovered how different algorithms work in different cases. They found that the Decision tree is easy to visualize and explain the model implementation and it also saves computational power.

4. PROPOSED SYSTEM

The proposed system is a completely a software based application built in python language and using the data set of Play store. Functionalities This analysis will help individuals /Firms in designing and launching commercially viable mobile apps by using its functionalities like:

1. Percentage download in each category.

2. Category of apps with most least and average downloads.

3. Category of apps with maximum average ratings from users.

4. Apps that have managed to generate the most positive and negative sentiments.

5. The relation between the Sentiment-polarity and sentiment subjectivity of all the apps.

6. Interface where the client can see the reviews categorized as positive, negative and neutral.

About Dataset

Most regularly a dataset relates to the matter of the single database table, or the single factual information framework, where each segment of the table speaks to a specific variable, and each column compares to a given individual from the informational collection being referred to. This information is of Google play store application and is taken from Kaggle, which is the world's largest community for data scientists to explore, analyze and share data. This dataset have 11 columns of varied categories of the appliance. In this project I have analyzed all these various columns of the dataset. The 11 columns of the dataset are as follows:

Parameters	Description
App	Application Name
Category	Category the app belongs to
Rating	Oveall rating of the app
Reviews	Number of user reviews for the app
Size	Size of the app
Installs	No. Of user downloads/installs the app
Type	Paid or free
Price	Price of the app
Content rating	Age group the app is targeted at Children / Mature/ Adult
Genres	An app can belongs to multiple genres For eg-a musical family, game.
Last updated	Date when the app was last updated on google playstore

Table-1*:* Dataset columns and its specifications

We found most popular category of apps on two basis - Number of Installs and Number of reviews. Personalization wins in former criteria whereas Sports wins in later criteria. This data is good to implement machine learning models which was not a part of this project. It are often considered as an improvement for future. A more can be done using Last updated variable where month can be separated and clubbed with a lot of other variable in order to insightful information.

5. IMPLEMENTATION

Mobile app distribution platform such as Google play store gets flooded with millions of new applications uploaded by developers everyday. So in this project, we aim on analyzing Google play store that provides a particular app description and data such as reviews, ratings, price and number of downloads. The objective of this is to analyze the desire of the customer through the reviews provided in the feedback section and apps trend in the market to help the organization & developers. To this end, we provide an idea about app that managed to get maximum and minimum number of downloads and predicting the category of apps that is most likely to be downloaded in

the coming years. Moreover, doing sentimental analysis on the apps that generated most positive and negative sentiments, Sustainability of app in market on basis of previous data and current market situation. Furthermore, also analyzing the apps that has maximum downloads have they managed to get average rating so that concluding the co relation between number of downloads and ratings received. There will be different tabs available for different operations and analysis that user wants to view.

We have used the below given algorithm for implementation.

LINEAR REGRESSION

Logistic Regression may be a Machine Learning algorithm which is employed for the classification problems, it's a predictive analysis algorithm and supported the concept of probability. Linear regression algorithm shows a linear relationship between a dependent (y) and one or more independent (y) variables, hence called as rectilinear regression. Since rectilinear regression shows the linear relationship, which suggests it finds how the worth of the variable is changing consistent with the worth of the experimental variable. When working with linear regression, our main goal is to find the best fit line that means the error between predicted values and actual values should be minimized. The best fit line will have the smallest amount error. The cost function helps us to work out the simplest possible values for a_0 and a_1 which might provide the simplest fit line for the info points. Since we would like the simplest values for a_0 and a_1,

$$minimize \frac{1}{n} \sum_{i=1}^{n} (pred_i - y_i)^2$$

$$J = \frac{1}{n} \sum_{i=1}^{n} (pred_i - y_i)^2$$

Minimization and cost function

We convert this search problem into a minimization problem where we might wish to minimize the error between the predicted value and the actual value. We choose the above function to minimize. The difference between the anticipated values and ground truth measures the error difference. We square the error difference and sum over all data points and divide that value by the entire number of knowledge points. This provides the typical squared error over all the info points. Therefore, this cost function is also known as the Mean Squared Error(MSE) function. Now, using this MSE function we are getting to change the values of a_0 and a_1 such the MSE value settles at the minima.

5.1 USER INTERFACES

The product will exist on a real-life system. The Interface will be a simple user interface.

The Google Play Store Review API lets you prompt users to submit Play Store ratings and reviews without the inconvenience of leaving your app or game.

Fig-1: User Interface

5.2 FLOW CHART

Flowcharts are utilized in designing and documenting complex processes or programs. It shows the flow of our project. How the working of our project is and like other diagrams, they also helpus to see what's happening and thereby help the people to know a process, and maybe also find flaws, bottlenecks, and other less obvious features within it

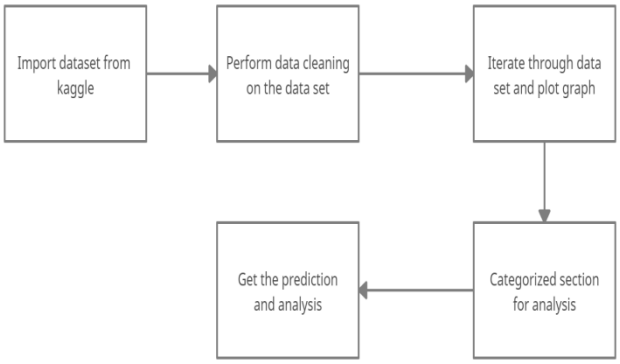

Fig-2: Data Flow

5.3 REQUIREMENTS

Hardware Interfaces

A desktop or laptop with (Minimum i3 5th Gen, 4gb ram or higher).

Software Interfaces

1. Anaconda Navigator

2. Spyder Notebook Installed.

3. PIP Files.

4. Python Ver 3.7 or 2.7.

5. Various additional libraries required for additional functionalities.

6. Various libraries like matplotlib, numpy, pandas, seaborn etc. Are used for invoking special functionalities to the design.

5.4 OUTPUT SCREEN AND RESULT

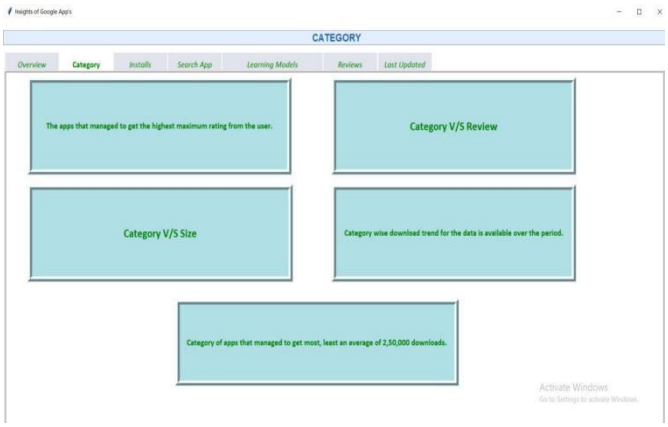

Fig-3: Category(main page)

Figure 3 Shows the user interface where we have different options in the tab section inside those options there are various operations available.

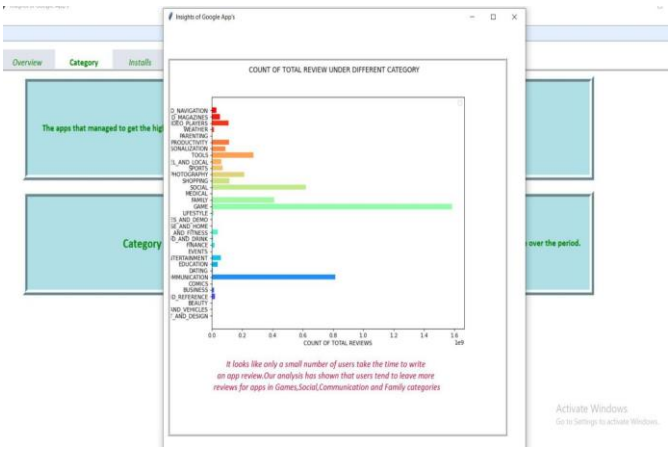

Fig-4 : Category vs Review

The above figure 4, It is the graphical reperesentation of Category v/s Review where each category is shown in different colour. Here Games, Social, Communication and Family categories that have the most reviews received.

Fig-5 *:* Installs(Teen vs. Mature)

In figure 5, It is bar graph reperesentation of installation which is generated in the section of teen and mature which is decided on the age factor of the user.

6. ADVANTAGES

1. Analysis of all the applications on play store app.
2. It will give past trends of all applications.
3. Find out the most rated and most reviewed apps.
4. It will identify the feedbacks of all the application in one interface.
5. Will shows the Category of application in demand.

7. CONCLUSION AND FUTURE WORK

Thus the app development companies could decide what application should be developed and they can also see the prediction of their developed application. In this they also get to see the categorized reviews of all the application in one interface which will help them decide which app is liked by the users and which apps need to be developed more. The dataset contains immense possibilities to improve business values and have a positive impact. It is not limited to the problem taken into consideration for this project.

In future we could display live downloads and top applications of the play store. We could add a system that would create application on its own by using the data set

and creating the best user interface by the highly rated apps.

Future work can also include:

1. Optimization of the pie-charts There are multiple domains in the same slice. The multiple domains could be separated and added to the same field to get a more detailed version of this pie chart.

2. Prediction of the number of reviews and installs by using the regression model.

3. Identifying the categories and stats of the most installed apps.

4. Exploring the correlation between the size of the app, the version of Android, etc. on the number of installs.

Many other interesting possibilities can be explored using this dataset.

8. ACKNOWLEDGEMENT

This project was completed by Akhlak Ali Sunasara, Nancy Jaiswal and Suchit Poojari under the guidance and instruction from Prof. Anil Kumar Chaturvedi. We are extremely grateful to the celebrated authors whose precious works have been consulted and referred in our project work. We also wish to convey our appreciation to our friends who provided encouragement and timely support in the hour of need. The dataset was used from the Kaggle data store.

9. REFERENCES

[1] Cuiyun Gao, Jichuan Zeng, Michael R. Lyu, and Irwin King, "Online App Review Analysis for Identifying Emerging Issues",ACM/IEEE 40th International Conference on Software Engineering, Shenzhen Research Institute of The Chinese University of Hong Kong, China-2018.

[2] Liu Yezheng,Du Fei, Jiang Yuanchun,Liu Xiao, Wang Qiudan, "A Novel APPs Recommendation Algorithm Based on APPs Popularity and User Behaviors", IEEE First International Conference on Data Science in Cyberspace, China, Australia-2016 .

[3] Bayu Trisna Pratama, Ema Utami,Andi Sunyoto ,"A Comparison of the Use of Several Different Resources on Lexicon Based Indonesian Sentiment Analysis on App Review Dataset", Magister of Informatics Engineering Universitas Amikom Yogyakarta, Indonesia-2019.

[4] Varsha A. Patil, Nitin N. Patil, "MOBILE APPS OPINION ANALYSIS USING EMOTICON", International Conference on Global Trends in Signal Processing, Information Computing and Communication, Shirpur, India-2016.

[5] Mohamed Ali,Mona Erfani Joorabchi, Ali Mesbah, "Same App, Different App Stores: A Comparative Study", IEEE/ACM 4th International Conference on Mobile Software Engineering and Systems (MOBILESoft),University of British Columbia Vancouver, BC, Canada-2017.

[6] M. Harman,A. Al-Subaihin, Y. Jia,W. Martin,F. Sarro,Y. Zhang, "Mobile App and App Store Analysis, Testing and Optimisation",IEEE/ACM International Conference on Mobile Software Engineering and Systems,University College London, CREST Centre UCLappA Group, London, WC1E 6BT,UK-2016.

[7] Rimsha Maredia, "Analysis of Google Play Store Data set and predict the popularity of an app on Google Play Store", Texas A&M University College Stattion, Texas ,June-2020.