International Journal of Research in Engineering, Science and Management
Volume-3, Issue-3, March-2020
www.ijresm.com | ISSN (Online): 2581-5792

148

Google Play Store Application Evaluation Using Google Play Scrapper for Analytics

Aishwarya Patil[1], Sakshi Patil[2], Sneha Patil[3], Swapnil Waghmare[4]

[1,2,3]*Student, Department of Computer Engineering, Pillai HOC College of Engineering and Technology,*
Rasayani, India
[4]*Assistant Professor, Department of Computer Engineering, Pillai HOC College of Engineering and*
Technology, Rasayani, India

Abstract: **Now-a-days there are millions of applications present on the Google Play Store of various categories. These apps are downloaded and been used by millions of users. Whenever a user search for apps on Play store, a list of apps is shown to the user in which each app contains the app name along with its rating. The users usually tend to download highly rated apps because that reflects the other users experience. So the different techniques and tricks, like attractive app titles, demanding icons are used by the developers to attract the users ignoring the quality of app itself. However, there is no existing technique to analyze the impact attractive titles or other such things on the ratings of the app. Therefore, to examine a number of factors of Google play store apps and identify the influence of these factors along with visualization using variable importance we developed this application. This application will also enable users to see the quality of apps.**

Keywords: **Google play store, Data scraping, Play store applications, Analysis, Visualisation.**

1. Introduction

Mobile apps are becoming popular every coming day as they are fast, practical and efficient. Their popularity is among users as well as in business owners also as the apps can be used for advertising. App is nothing but a computer program designed to run on devices like phone, tablet to provide users some services. Apps are offered in specific software repositories referred generally as App stores, where the largest shareholders are Google Play1, iPhone App Store 2 and Blackberry App World 3. Google play store is an official app store for Android operating system. These apps belong to various categories like games, communication, books, business, news, sports, and many others.

In Google Play store, each app has its name and its rating. The app rating is the average of all the ratings given by the users. The users tend to download the apps with high ratings as it reflects the good experience of other users who already downloaded and used the apps. The study has showed that the app ratings affect the user's intention to download and use an app [2].

However, there is no application exists that provides relationship between the ratings and other factors of app so that

the developer can ultimately decide how to develop good quality app [3]. Also, many companies even use wrong ways to gain high ratings for their apps such as paid app ratings, so their apps can be more popular and visible in the play store. The proposed system provides an interface to visualize the relationship between the ratings and the other factors of app for developers and the overall quality of app for users.

2. Literature Review

As information accessibility, culmination and exactness could likewise be a huge issue inside the versatile application advertise, the include this territory stays very breaking point. Google Play store positions each application that is distributed in the store. The general application positioning arrangement of Google [1] utilizes complex foundation to rank applications. This is the method of reasoning google application positioning framework outranks different applications store and is effective. As indicated by the researchers, Google Play is even more a genius showcase on account of its well-known items and applications. Google play have a pleasant and clean selection framework that positions applications [6]. Also, these positioning frameworks are unpredictable and rank the applications in a generally excellent way, there are numerous situations where clients look over the pages to the base and discover the applications they like or discover an application with a superior rating. Since the presentation of portable application stores, there are two or three specialists who works inside the universe of portable application rating expectation and variable significance that helps discover a relationship between the applications appraisals and different components. Scientists have worked in various do-mains to discover the approaches to get better appraisals in applications. A few specialists concentrated on client's surveys, a few scientists focus on the applications traits and highlights, a few analysts worked inside the field of higher programming designing practices. Be that as it may, all of these spaces are significant at their places, yet for the most part applications qualities are investigated by the specialists to discover a connection between applications rating furthermore, its qualities. Right now,

International Journal of Research in Engineering, Science and Management
Volume-3, Issue-3, March-2020
www.ijresm.com | ISSN (Online): 2581-5792

149

examine some of the contemporary research works in this territory. Tian et al. [5] played out a contextual analysis utilizing factual investigation to rank the various elements of applications that impact the application appraisals, the size of an application, limited time pictures and target sdk of an application are the first persuasive components of high-appraised applications. Additionally, Finkelstein et al. [4] explores the connection transport between value, rating and acknowledgment inside the blackberry application store and their discoveries show that there's a hearty connection between the client appraisals and fame. Specialists performed an inside and out examination of applications from Android and Apple applications and played out a quantitative concoction investigation of applications traits and their impacts on the applications in various application stores [2]. Scientists reveal factors that impact the application rankings for apple application store and proposed a model that anticipate the appraisals for different applications. They thought about kind of factors in their model, including bundle size, application discharge date, classification fame, and so forth to discover the significance of these factors [4]. Essentially, specialists utilized versatile application appraisals for the application recommender framework, master frameworks and information based framework for various areas. In any case, the universe of application variable significance is amazingly restricted and there's a hole right now. Likewise, at one hand, designers and enterprises attempt their best to make applications to increase better rankings. Albeit a few specialists propose that application rating isn't considered significant or is variable like Liu et al [6] performed a nitty gritty investigation of Google play store. According to their discoveries the survey appraisals have lower sway in the event of free applications. While various specialists concentrated on various qualities that impact the application evaluations, there are still some straightforward yet significant elements that are not yet examined by the re-searchers. In addition, the vast majority of the specialists focus on a couple of number of properties and discover the significance of those properties for foreseeing the applications rating. Normally scientists took the default qualities of the applications and played out their investigation on those qualities. Along these lines, there are numerous traits which will be processed for each application and its belongings are regularly examined for rating of that application. Accordingly, to fill the exploration hole, we con-channel a top to bottom investigation during which we take a enormous applications store informational collection, we use assortment of default application includes moreover as register assortment of traits for each application and discover the significance of these characteristics in application appraisals.

3. Implemented method

A. System Overview

The figure below shows how the system is going to work, in here first the Data is scraped from the Google Play Store. Then the dataset is created from the scraped data and the pre-

processing of Data is done on the dataset. After this the applications are categorized according to the attributes which we have taken into consideration. And this all is visualized to the user through the interface which we have built.
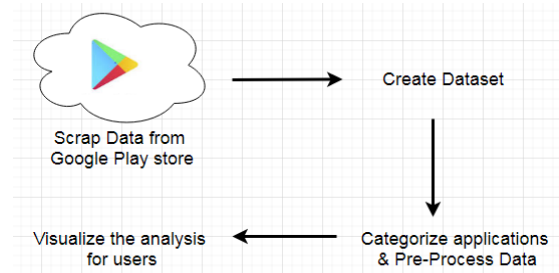

Fig. 1. Diagrammatic representation of the system
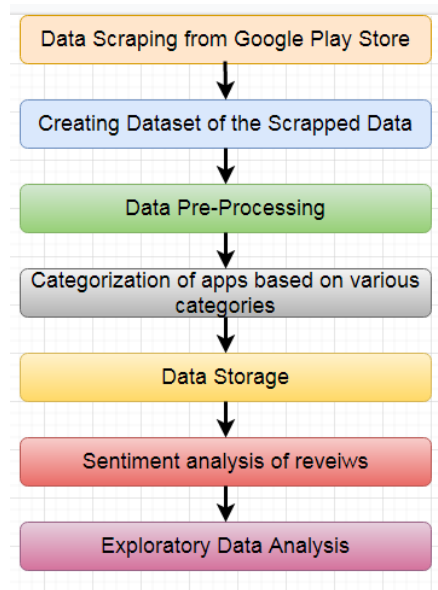
B. System Architecture

Fig. 2. Architecture of the system

The above architecture shows the flow of how the procedure of how the system is going to work and how the interface is built. In the above architecture we can see the different steps that are used for the working of the system and the same are explained below:

Data Scraping from Google Play Store: The data is scrapped from the google play store using the play store scrapper.

Creating Dataset: The dataset that we have created has 5223 unique apps and contained over 170000 reviews of the Google Play Store Apps that are scrapped from Google Play Store.

Data Pre-Processing: Since there were no null/missing values in the dataset. However, the raw data extracted needed to be pre-processed to turn it into some valuable information. To be able to perform EDA on our dataset, installs, total number of ratings, rating distribution were converted to integers. Sizes of app having kilobytes were converted to megabytes to have consistent units.

Sentiment Analysis: Further, to find out the sentiment of the

International Journal of Research in Engineering, Science and Management
Volume-3, Issue-3, March-2020
www.ijresm.com | ISSN (Online): 2581-5792

150

users, the review text was processed using Sentiment Intensity Analyser. Polarity and subjectivity of each of the reviews were found with polarity values between -1 and 1 and subjectivity values between 0 and 1. A polarity value of 1 means a strongly positive feedback by the user and a polarity value of 0 means otherwise. On the other side, a higher value of subjectivity means the review is more subjective and hence is not good.

Data Storage: We used Microsoft Sql Server for storing the sentiments and the pre-processed data so we can use the data for further EDA.

Exploratory Data Analysis: In this part we design different Graphs on the data that we have pre-processed and stored in the database.

From the above architecture we can easily understand the work flow of how the system was built and also all the blocks are explained at a detail. So in it we can see that we have created a dataset by scraping the data from google play store and the dataset is then pre-processed to categorize the applications as per the attributes which we are going to take under consideration and also according to various analysis we have to do. The dataset it is pre-processed and it is stored in the database. The description about these variables are given below:

- App id: It indicates the id of a particular app.
- App name: It specifies the name of the app.
- Category: It includes the category of app.
- Ratings: The no of ratings given by the users for particular app.
- Size: It includes the memory required for the app.
- Reviews: The reviews given by the users for the app.
- Type: It indicates the type of the app. i.e. It is free or paid.
- No of installs: It gives the count of installs associated with app.
- Current version: It specifies the current version of the app.
- Android version: It gives information about android version requires to install the app

After that the applications are reviewed on their genuineness using the sentiment analysis on the reviews of customer we have scrapped and collected in our dataset. Then all this analysed data is stored for further EDA. Also all this data is represented on the interface which we have built and the users can easily access the data about the applications.

4. Results and Analysis

Daily developers upload various apps on the google play store. The users install this app without checking whether the app is authenticating or fake. The developer does not provide proper information about the application which defines authenticity of application. Even though many research is conducted on the same there is not a specific application which is built to solve the problem. So we have built an interface

which provides information about the applications according to the categories and which can be easily accessed by the users.
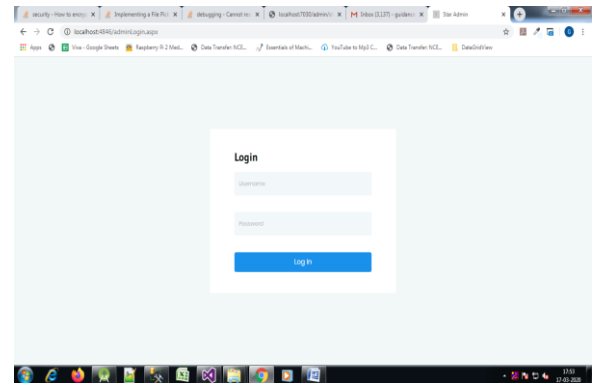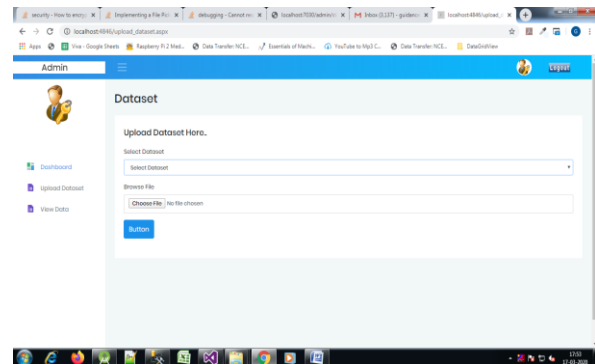

Fig. 3. Login page for users

Fig. 4. Uploading the dataset

This application provides interface which includes categorization of applications using attributes which include app id, app name, category, ratings, size, reviews, type, no of installs, current version, android version. The analysis of this information will help the user to identify the genuine app present on google play store. It will help the user to prevent their confidential information. It will also help to increase the trust of user on Google Play store. This interface can create a huge impact as the user before actually downloading the applications can check whether the claims made by the developer and checking the genuineness of applications
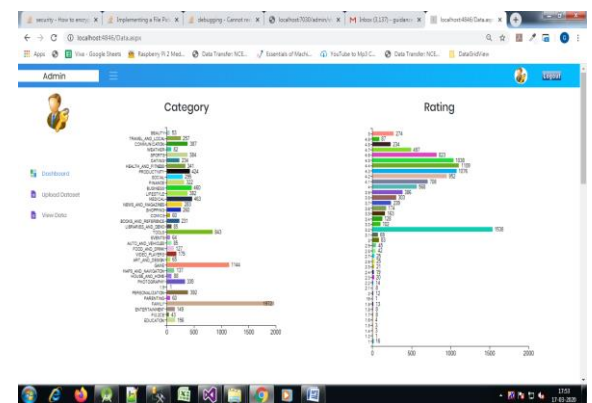

Fig. 5. Categorization of applications

International Journal of Research in Engineering, Science and Management
Volume-3, Issue-3, March-2020
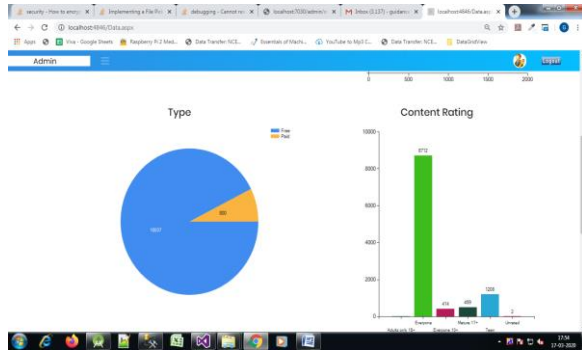**www.ijresm.com | ISSN (Online): 2581-5792**

151

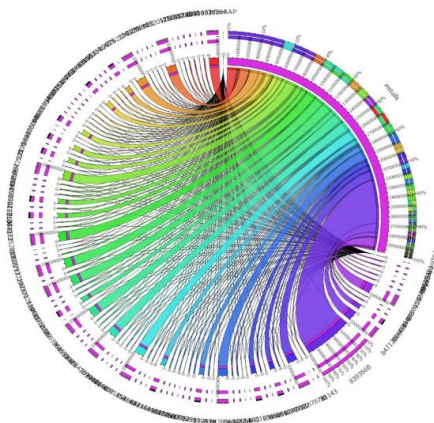Fig. 6. Visualisation of separate Categories

Fig. 7. Visualisation of correlations between selected attributes

Table 1
Comparison between existing technique and implemented technique

Features	Existing Technique	Implemented Technique
Scraping of Data	Yes	Yes
Creating Dataset	Yes	Yes
Built Interface	No	Yes
Analysing the Data	Yes	Yes
Uploading Dataset	No	Yes
Sentiment Analysis	No	Yes

Also data storage is provided so the data which is analyzed can be further used for further analysis if required. The interface which is built shows us the various co-relations of attributes which we have selected and how it effects the overall ratings of the applications.

When we upload a Dataset it takes time loading onto the system and the information is pre-processed and represented onto us by doing all the analysis including sentiment analysis of the reviews.

5. Conclusion

In this paper we have proposed an interface which is built for analyzing and representing the Applications of Google play store. In this, we focused on analyzing Google Play, the largest Android application store. The overall objective of this analysis effort is to provide users an interface where users can check the genuineness of applications which uploaded on the Google play store. Using this approach, we identified a strong negative correlation between price, number of downloads and price, participation and a strong positive correlation between number of download, participation. This application can help the users in understanding the application without downloading the same first.

References

[1] R. M. Amir Latif, M. Talha Abdullah, S. U. Aslam Shah, M. Farhan, F. Ijaz and A. Karim, "Data Scraping from Google Play Store and Visualization of its Content for Analytics," 2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Sukkur, Pakistan, 2019, pp. 1-8.

[2] Chin-Lung Hsu and Judy Chuan-Chuan Lin. 2015. What drives purchase intention for paid mobile apps? - An expectation confirmation model with perceived value. Electron. Commer. Rec. Appl. 14, 1 (January 2015), 46–57.

[3] Aralikatte, R., Sridhara, G., Gantayat, N., and Mani, S. (2018). Fault in your stars: an analysis of android app reviews. In Proceedings of the ACM India Joint International Conference on Data Science and Management of Data, pp. 57–66. ACM.

[4] Roberto Minelli and Michele Lanza. 2013. Software Analytics for Mobile Applications--Insights & Lessons Learned. In Proceedings of the 2013 17th European Conference on Software Maintenance and Reengineering (CSMR '13). IEEE Computer Society, USA, 144–153.

[5] Fu, B., Lin, J., Li, L., Faloutsos, C., Hong, J., and Sadeh, N. (2013). Why people hate your app: Making sense of user feedback in a mobile app store. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 1276–1284. ACM.

[6] Harman, M., Jia, Y., and Zhang, Y. (2012). App store mining and analysis: Msr for app stores. In 2012 9th IEEE Working Conference on Mining Software Repositories (MSR), pp. 108–111.