

Prophecy - Google Apps Analysis and Prediction

Gunjan Khandelwal¹, Gaurav Rajpurohit², Gaurav Goyal³, Loveleen Kumar⁴

¹B.Tech Student, Department of CSE/IT, Global Institute Of Technology, Jaipur, India
gunjan.k8411@gmail.com

²B.Tech Student, Department of CSE/IT, Global Institute Of Technology, Jaipur, India
grv1881997@gmail.com

³B.Tech Student, Department of CSE/IT, Global Institute Of Technology, Jaipur, India
gauravgoyallalsot@gmail.com

⁴Assistant Professor, Department of CSE/IT, Global Institute Of Technology, Jaipur, India
loveleen.kumar@gitjaipur.com

Abstract— “Prophecy – Google Apps Analysis and Prediction” was built with an objective to help the companies to identify the overall rating of their apps based on the reviews and allow the new companies to enter the market of apps with moderate/more/fewer competitors. The project is built with a user friendly interface so as to make it easy for user. The complete project inbuilt various technologies like html, CSS, machine learning, python, NLP etc. The users are also provided with other services in order to help them identify the current market demand. This web-app portal provides a platform for app owner companies to upload the file of their apps review and find out the how many reviews are positive, negative and neutral. Based on this the companies can identify the overall impression their app is making on the users. The system is built using machine learning algorithm which showed the best score among all the algorithm which makes the system highly reliable. An easy to use interface for accessing the services provides an extra advantage to the portal. Other than this it provides the current stats of the apps market, searching between the apps, navigating to the websites of most popular applications etc. all these services will help the user to understand the requirement of the market. In all it can be concluded that this portal will turn to be true friend as the name in providing the solutions.

Keywords – Data mining, Machine learning, web app portal, HTML, CSS, API

I. INTRODUCTION

Google Play Store is an Android Market which is maintained by Google itself. It serves as an official app store for the Android operating system, allowing users to browse and download applications developed with the Android software development kit and published through Google. Any type of application whether paid or unpaid, it can easily be downloaded through Google App Store. The Google Play store had over 82 billion app downloads in 2016 and reached over 3.5 million apps published in 2017. Over 2.7 million apps were available in 2017, the numbers have only increased in the coming years.

The studies have shown that an 18 to 24-year-old spend around 66% of their digital media time using smartphone apps. More than 2 out of 3 millennials say they're always looking for new apps and wish they could do more with the apps they already have and they are also ready to pay for the mobile apps. In 2017 it was recorded that 21% of millennials deleted some or the other mobile apps. Hence every day thousands of apps are being launched in the market at the

same time crores of apps are deleted or uninstalled from the mobile.

In 2018 around 6140 apps were released through the Google Play Store per day. And this number is climbing the mountain day by day. Out of all these only few gets the success and the rest get out of the market in their first month only. Research says that out of 1000, 999 newly launched app fails these days due to user unfriendliness, pivot failure, bad technology etc. which lead to great economic loss. The reasons for such failure can also be lack of understanding of the needs or demands of the market.

Here comes the need for a system which let the app owners not only to identify or understand the app market but also suggest the categories that are highly popular among the users. The system also identifies the overall rating of the apps by doing sentiment analysis on the reviews.

Basically, it is a web portal which will provide a complete solution to all the above mention problems. This web portal provides a platform where the user can upload the file of reviews and get an overall ranking of the app. The system is

built using machine learning algorithm which makes the system highly reliable. The app owner is suggested with the currently popular category of the application so to launch the app that app user actually wants. An easy to use interface for accessing the services provides an extra advantage to the portal. Other than this the it provides other services like searching the required application from the whole dataset, navigating to the websites of the most popular application. In all it can be concluded that this portal will turn to be a true friend to the user and provide solution to their problems.

II. LITERATURE SURVEY

Data mining is an analysis tool which allow the user to extract useful information from a large dataset. Data mining combines several branches of computer science and analytics, relying on intelligent methods to uncover patterns and insights in large sets of information. Machine learning is a field of computer science that uses statistical techniques to give computer systems the ability to ‘learn’ with data, without being explicitly programmed. Data visualization algorithms create images from raw data and display hidden correlations so that humans can process the information more effectively.

III. METHODOLOGY

The project was divided into multiple tasks and each task each further divided into multiple tasks. Firstly, all the tasks were collected to be there in the project for achieving the projected plan. Then the tasks were divided according to the different modules. Different modules identified were:

- **Dataset Collection:** The requirement of the project was analysed first. Accordingly, the data of various apps were collected from various sited for training the model. First of all, we generate the Google Play Store Apps dataset and Reviews dataset through two sequential steps:
 - Data extraction, which extracts the unstructured data from various apps websites viz. Kaggle and other web sources during the period of 2000-2018. After extraction we got two datasets: Google Play store apps and Google Play Store apps reviews dataset.
 - Data pre-processing cleans, integrate and reduces the overall size of the two datasets. Google Play store apps is represented by 4 attributes and Google Play store apps reviews is represented using 2 attributes.

Index	App	Category	Rating	Reviews	Size
0	Photo Editor & Candy Came...	ART_AND_DESI...	4.1	159	19M
1	Coloring book moana	ART_AND_DESI...	3.9	967	14M
2	U Launcher Lite - FREE ...	ART_AND_DESI...	4.7	87510	8.7M
3	Sketch - Draw & Paint	ART_AND_DESI...	4.5	215644	25M
4	Pixel Draw - Number Art C...	ART_AND_DESI...	4.3	967	2.8M
5	Paper flowers instructions	ART_AND_DESI...	4.4	167	5.6M
6	Smoke Effect Photo Maker ...	ART_AND_DESI...	3.8	178	19M
7	Infinite Painter	ART_AND_DESI...	4.1	36815	29M
8	Garden Coloring Book	ART_AND_DESI...	4.4	13791	33M
9	Kids Paint Free - Drawi...	ART_AND_DESI...	4.7	121	3.1M
10	Text on Photo - Fonteee	ART_AND_DESI...	4.4	13880	28M
11	Name Art Photo Editor...	ART_AND_DESI...	4.4	8788	12M
12	Tattoo Name On My Photo ...	ART_AND_DESI...	4.2	44829	20M
13	Mandala Coloring Book	ART_AND_DESI...	4.6	4326	21M

Index	App	Translated_Review	Sentiment	Sentiment_Polarity	Sentiment_Subjectiv
0	10 Best Foods for You	I like eat delicious fo...	Positive	1	0.533333
1	10 Best Foods for You	This help eating healt...	Positive	0.25	0.288462
2	10 Best Foods for You	nan	nan	nan	nan
3	10 Best Foods for You	Works great especially g...	Positive	0.4	0.875
4	10 Best Foods for You	Best idea us	Positive	1	0.3
5	10 Best Foods for You	Best way	Positive	1	0.3
6	10 Best Foods for You	Amazing	Positive	0.6	0.9
7	10 Best Foods for You	nan	nan	nan	nan
8	10 Best Foods for You	Looking forward app,	Neutral	0	0
9	10 Best Foods for You	It helpful site ! It he...	Neutral	0	0
10	10 Best Foods for You	good you.	Positive	0.7	0.6
11	10 Best Foods for You	Useful information ...	Positive	0.2	0.1
12	10 Best Foods for You	Thank you! Great app!! ...	Positive	0.75	0.875
13	10 Best Foods for You	Greatest ever Completely a...	Positive	0.992188	0.866667

Fig 1: Data Set

- Displaying Google Apps Statistics:** We have taken the data of Google App Store which contains the information of all the Application that have been launched into the market. On this dataset, data mining is applied to analyze the app market structure. Upon analyzing, we found: number of apps available in the market based on content rating, top 10 mostly installed application, percentage of paid and unpaid application in the Google play store etc.



Fig 2: Google Play Store Statistics

- Preparing Machine Learning Algorithm:** An algorithm was prepared for predicting the number of installs a app can have. The Apps database was supplied to prediction and the review database was supplied to sentiment. The prediction module uses the Decision Tree Regressor which uses a flow chart like tree which consist of branches and nodes. Here the branches represent the result of the node and the nodes represent either condition or result (end node). With the help of this algorithm the future number of installs are identified. The sentiment module uses the Natural Language Processing (NLP) which programs the computers to process and analyse

large amounts of natural language data. Based on the processing the polarity of the single review is checked or a file containing thousands of reviews is uploaded on the website and the overall rating of the app is predicted. This will help the user to determine the performance of the application and also about the future of the application.



Fig 3: Prediction Page



Fig 4: Sentiment Page

- Web Interface Designing:** This task comprised of designing a friendly user interface for the public to access the functionalities of the project. Different pages were designed as per the services provided like Home

Page, Prediction, Sentiments, Filters, etc. They were all designed using html, CSS etc. Complete designing was done to make the project look more attractive and make it easy to use for the user. Overall an interactive user-friendly interface was prepared.



Fig 5: Home Screen

- Integration of algorithm with web interface:** The web interface is integrated with machine learning algorithm using python library Flask. Flask provides extreme simplicity, flexibility and fine-grained control. We have done routing of YRLs through flask. Through Flask we have integrated machine learning algorithm (Python Based) within the functions with ease which made the application more intelligent and cognitive.

IV. EXPERIMENTATION AND RESULTS

The raw data collected was extracted from various sources and then was pre-processed using various data analysis techniques. Then the data was cleaned from various angles and a fresh dataset was obtained. This dataset was converted into Data frame, a datatype in Python language. This dataset was used for analysis and prediction purposes. Various machine learning algorithm was applied on it and algorithm with the best score was selected and finally a model was created of it, which was used for further prediction. Similarly, various data analysis techniques were used to analyze the data. Different types of visualization tools were used to display the statistics of the dataset.

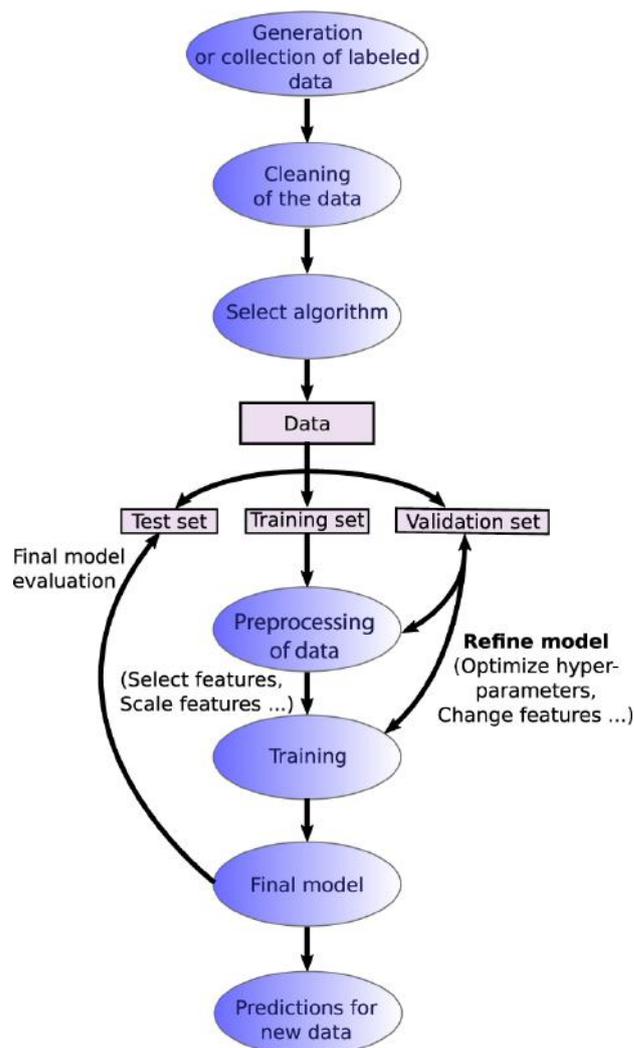


Fig 6: Methodology

V. CONCLUSION

All research work regarding the requirements of the system or project was done in initial phase. Then all the analysis, various facts and figures collected were put together to form a structured plan for the development of the system/project. A complete working web based application integrated with machine learning using python, got developed which provide various information regarding Google Play Store Application.

The user can access the website anytime and form anywhere. This website is available 24*7 to the user. The only requirement is of laptop or desktop and stable internet connection.

Two types user can access this website. First user, who has an application that is launched in the market, and that user want to know the overall performance of his/her application. He needs to just upload the file containing the reviews and overall performance will be generated in terms of number of

positive, negative and neutral reviews. Second user, who is planning to launch an application in the market. He can get an idea of number of installments his application will have, by giving some information regarding his application like, content-rating, type, size etc.

Any general user can also use this website to navigate to some mostly famous websites like: Facebook, Amazon, Instagram etc. The users can see the statistics of the Google Play Store App dataset and can get the idea about the types of application available in the market.

ACKNOWLEDGEMENT

It is our pleasure to be indebted to various people, who directly or indirectly contributed in the development of this work and who influenced our thinking, behavior, and acts during the course of completion. This formal piece of acknowledgement is an attempt to express the feeling of gratitude towards people who help us in successful completion of our project. We are extremely grateful to **Dr. I C Sharma**, Principal, Global Institute of Technology, Jaipur and **Dr. Girraj Khandelwal**, Head of Department (Computer Science & Engineering), Global Institute of Technology, Jaipur for providing all the required resources for the successful completion of our major project. We would like to express our deepest gratitude to **Mr. Loveleen Kumar**, our project guide for his guidance, precious time and necessary advices. He was always there with his competent support and valuable suggestion throughout the development phase of the project. We would also like to thanks the supporting staff of Department of Computer Science & Engineering (Global Institute of Technology, Jaipur) and our friends and family who always have supported us at every step, guided us, inspired us, and provided us all facilities so that we can achieve our goals.

FUTURE SCOPE

In future we wish to expand our project to analyze the performance of multiple application and determine the best one. The statistics of Google Play Store can be displayed according to the input of the users. A platform can be provided to launch the application in much easy way.

REFERENCES

- [1]. Python, R. (2019). *How to Write Beautiful Python Code With PEP 8 – Real Python*. [online] Realpython.com. Available at: <https://realpython.com/python-pep8/> [Accessed 6 Oct. 2019].
- [2]. Nltk.org. (2019). *Natural Language Toolkit — NLTK 3.4.5 documentation*. [online] Available at: <https://www.nltk.org/> [Accessed 10 Sep. 2019].

- [3]. Kaggle.com. (2019). *Google Play Store Apps*. [online] Available at: <https://www.kaggle.com/lava18/google-play-store-apps> [Accessed 9 Aug. 2019].
- [4]. GeeksforGeeks. (2019). *Python | Decision Tree Regression using sklearn - GeeksforGeeks*. [online] Available at: <https://www.geeksforgeeks.org/python-decision-tree-regression-using-sklearn/> [Accessed 1 Sep. 2019].