Google Play Store Rating Prediction

Milestone: Project Proposal

Group 5

Niraj Sai Prasad
Sindhu Swaroop

857-269-4546
857-269-4536

saiprasad.n@northeastern.edu
swaroop.s@northeastern.edu

Percentage of Effort Contributed by Student 1: 100%
Percentage of Effort Contributed by Student 2: 100%

Signature of Student 1: NIRAJ SAI PRASAD

Signature of Student 2: SINDHU SWAROOP

Submission Date: 01/29/2023

Problem Setting:

Google Play Store, formerly known as the Android Market, has come a long way since its birth in 2008. With over 3.5 million apps and various other content such as books, movies and games listed on the platform, it is the most popular digital distribution service amongst android app developers as well as android cell-phone users. Google Play has won consumers' trust because of the security protocols it has in place.

While Google Play provides developers ease of app distribution, apps do require to be marketed for better reach and more downloads. Success of an app does not solely depend on how well optimized it is. It depends on user experience, ratings, and reviews. Google Play has a star-based rating system in place with 5 stars being the highest on the scale. Every decision we take in today's world is in the least, based on a quick glance at the rating and the number of ratings. A user's experience in using a particular app and how they would rate the app helps other consumers determine if they would want to install the app.

Problem Definition:

User ratings can be considered as an accurate representation of consumer impression on the application. Lower-rated applications are often at the bottom of search results, have no reach and may even be flagged and removed from the platform. Consequently, every developer's main aim is to maximize the rating of the app.

The goal of this project is to predict the overall rating of an app as ratings play the most important role in gaining users' confidence. Higher-rated apps and apps with a higher number of ratings are more likely to be recommended by the search algorithm on Google Play and more likely to be trusted by users that find the app while browsing the app store. This project aims to employ supervized machine learning, data mining, and visualization concepts to gain insights into what leads to higher number of installations and high user ratings - which loosely translates to the success of applications.

Few of the business questions we are looking to answer through this analysis are:

1. What category of apps has the greatest number of downloads? Group downloads by category to gain insight into what kind of apps are most popular among consumers.
2. When a user installs an app is it likely he will rate it? Is there any relation between number of ratings and number of downloads?
3. What is the mean and median rating of apps on play store? Plot a frequency distribution of ratings to visualize this.
4. Are paid apps more likely to be rated and reviewed by users on the platform? Does paying a price for apps mean that users feel more obligated to give feedback?
5. Does size of the app influence the number of installs? Do users prefer apps which occupy lesser space on their phones?

Some machine learning models that we plan to use for rating prediction are:

1. KNN Regressor
2. Random Forest Regressor

Data Source:

https://www.kaggle.com/code/taghredsalah199/google-playstore-regression-model/data

Data Description:

The data has been selected from Kaggle. The raw dataset comprises 10,841 records and 13 attributes. The different data fields in the dataset are as follows:

Sr no.	Attribute	Description
1.	App	Name of the application
2.	Category	Category of the application
3.	Rating	Rating provided by the users to the application
4	Reviews	Number of reviews provided by the users to the application
5.	Size	Size of the application in kilobytes

6.	Install	Number of users who have installed the application
7.	Type	Free or Paid
8.	Price	Price of the application to download in USD
9.	Content Rating	Age of the target audience
10.	Genres	Subcategory
11.	Last Updated	Date when the application was last updated on Play Store
12.	Current Version	Current version of the application available on Play Store
13.	Android Version	Android version required to install the application