# Predective Regression Model on Category Developed Applications on Google Play Store to Get High Downloads

Annapoorna Shetty, Mohamed Rizan , Franklin Fernandes

Dept. Of MCA, AIMIT, St Aloysius College (Autonomous), Mangalore

***Abstract* - This paper predicts the category of applications that are developed in large number of installed apps, it is often tedious for users to search for the app they want to download on bases on their liking or needs. In this paper we used K NN classifier for predicting the data. We examine that category of applications are dependent on number of downloads, reviews and ratings**

**Keywords:** ***K NN classifier, Decision Tree, Data Mining.***

## I. INTRODUCTION

The mobile App Store now stores more than two million apps which download billions of times per year in a Google play store as in a mobile platform. Development and updates of apps is one of major technical challenges affecting the mobile development community. As possible solution is in large companies where developers are investing resources and effort. Basically developers consumes much time to choose which category of apps is to be developed [1] [2].

## II. LITERATURE REVIEW

Related to their writing survey, it holds genuine that factors Literature discussed in this section concerns the study of app reviews, downloads and ratings. Here we focus on the Google Play store, with a minority focusing on Apple App Store and there are greater numbers of requirements as well as reviews literature each endeavour to end up more dynamic found that the reviews were mostly positive, and there were significant differences in the distributions between categories, and also between free and paid. Free apps had more reviews but a lower mean. Due to the higher numbers of reviews for free apps, which give an app credibility, the authors argue that in-app purchasing revenue models are a good way to make money for developers, especially if used as a ‘teaser’ for a paid version [3].

Another large sample was used to study apps. In which the authors analysed to reviews for summarisation. They designed a system that enables summarisation of reviews at a per-review, per-app or per-market level. This tool can be useful for large-scale overviews of competitor apps. The weakness of the system is the need for a large complete sample of reviews to be mined first, and the associated mining [4].

It performs static and dynamic analysis on Android apps, in order to help users complete bug reports. The system focuses on the steps to reproduce a bug, using dynamic analysis to walk through Android system events also study the Google Play reviews from 100 open source Android apps, and link the reviews to code changes. Where we find that a mean accuracy of 99% reviews are implemented in new releases, and that the apps with changes more directly implementing the content of user reviews improve their ratings with new releases.

## III. DATA AND EMPIRICAL ANALYSIS

The information we gathered from Google play store apps. In this section, we present the design of our study, the data collection and the processing methods that are used in our study

We select the Google Play Store as our app store of interest. Our criteria for an app store is based on the popularity [5]

Here out of 32 different types of category on the Google app store. Here the top 5 category of apps are taken into consideration as there is a lot of downloads and reviews and those 5 apps are specified as follows 12 represents family,15 represents game,29 represents tools,20 represents medical and 6 represents productivity[7].

### *Decision tree*

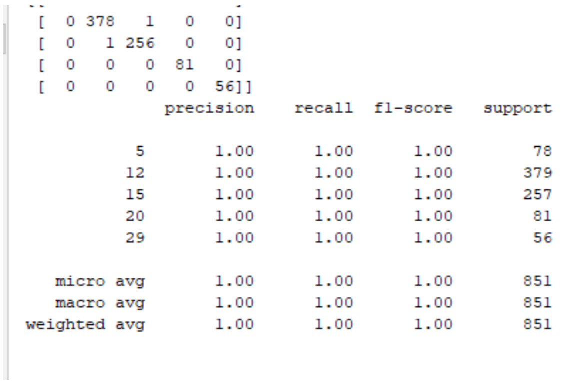
Decision tree constructs regression or classification models in the form of a tree structure. It breaks a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed [6]. The final result is a tree with decision nodes and leaf nodes. A decision node has two or more branches (e.g., Sunny, Overcast and Rainy), each representing values for the attribute tested. Leaf node. Represents a decision on the numerical target.

### *K-nearest neighbours (KNN)*

K-Nearest Neighbours is the most fundamental yet basic grouping calculations in Machine Learning [8]. It has a place with the administered learning area and finds serious application in example acknowledgment, information mining and interruption discovery. It is generally dispensable, in actuality, situations since it is non-parametric, which means, it doesn't make any hidden suppositions about the dispersion of information. We are given some earlier information which arranges facilitates into gatherings distinguished by a trait.

## IV. RESULTS

### KNN



```
 [  0 378   1   0   0]
 [  0   1 256   0   0]
 [  0   0   0  81   0]
 [  0   0   0   0  56]]
              precision    recall  f1-score   support

           5       1.00      1.00      1.00        78
          12       1.00      1.00      1.00       379
          15       1.00      1.00      1.00       257
          20       1.00      1.00      1.00        81
          29       1.00      1.00      1.00        56

   micro avg       1.00      1.00      1.00       851
   macro avg       1.00      1.00      1.00       851
weighted avg       1.00      1.00      1.00       851
```

FIGURE-1(a)

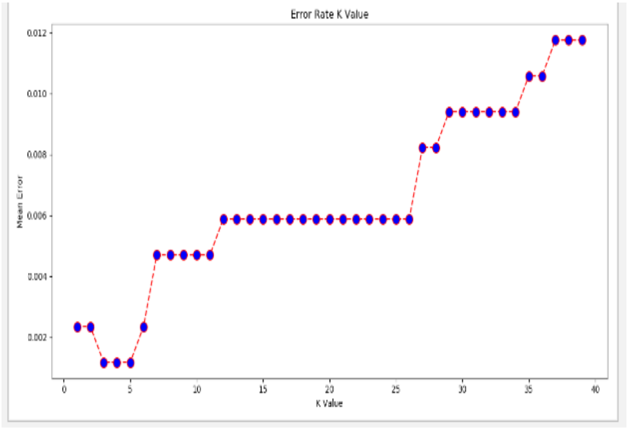
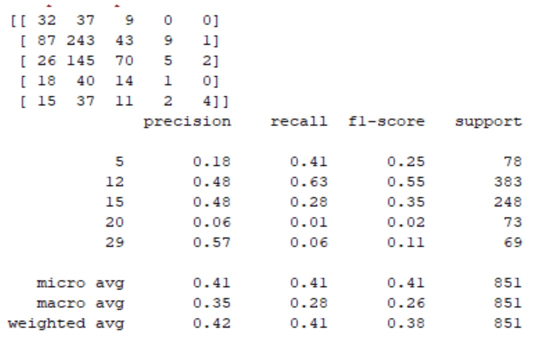


FIGURE-1(b)



```
[[ 32  37   9   0   0]
 [ 87 243  43   9   1]
 [ 26 145  70   5   2]
 [ 18  40  14   1   0]
 [ 15  37  11   2   4]]
              precision    recall  f1-score   support

           5       0.18      0.41      0.25        78
          12       0.48      0.63      0.55       383
          15       0.48      0.28      0.35       248
          20       0.06      0.01      0.02        73
          29       0.57      0.06      0.11        69

   micro avg       0.41      0.41      0.41       851
   macro avg       0.35      0.28      0.26       851
weighted avg       0.42      0.41      0.38       851
```

FIGURE-2(a)

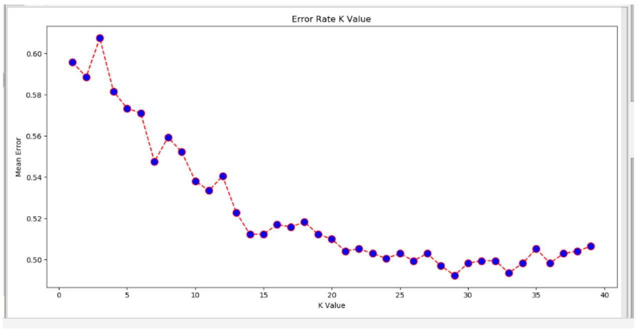
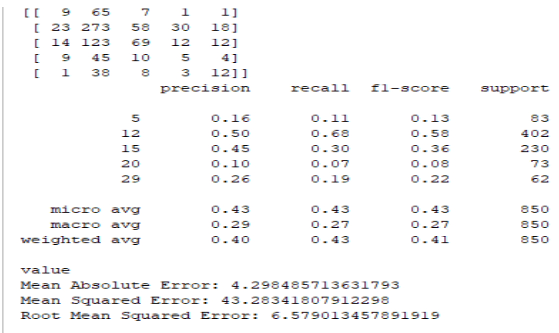


FIGURE-2(b)

### *DECISION TREE*



```
[[  9  65   7   1   1]
 [ 23 273  58  30  18]
 [ 14 123  69  12  12]
 [  9  45  10   5   4]
 [  1  38   8   3  12]]
              precision    recall  f1-score   support

           5       0.16      0.11      0.13        83
          12       0.50      0.68      0.58       402
          15       0.45      0.30      0.36       230
          20       0.10      0.07      0.08        73
          29       0.26      0.19      0.22        62

   micro avg       0.43      0.43      0.43       850
   macro avg       0.29      0.27      0.27       850
weighted avg       0.40      0.43      0.41       850

value
Mean Absolute Error: 4.298485713631793
Mean Squared Error: 43.28341807912298
Root Mean Squared Error: 6.579013457891919
```

*FIGURE-3*

## COMPARISION

Here the KNN algorithm (figure 1(a)) stands much better fit when compared to decision tree regression (figure-3) when rating is not considered. The KNN algorithm has a higher accuracy when compared to decision tree regression for our set of data.

## V. CONCLUSION

The factor effecting category are the number of downloads, review and rating. Here the KNN algorithm does not give a high accuracy if it takes on rating as a factor. But when the rating is removed as a factor then KNN has a higher accuracy. We conclude that KNN regression is a better fit than decision tree regression for predicting category based on the accuracy.

## REFERENCES

[1] https:// mypages.valdosta.edu/ hbelliott/ Haley %20and%20Caitlin%20-BMI %20 Research%20 Paper.pdf
[2] https://scielosp.org/pdf/rpsp/2013.v33n5/349-355/en
[3] Jiawei Han and Micheline Kamber, Decision Tree Induction, Data Mining concepts and techniques, Second Edition
[4] N.P gopalan, B.Sivaselva, Data Mining Techniques and Trends, PHI Learning , 2009.
[5] https://en.wikipedia.org/wiki/An-_empirical_study_on_applying_data_mining_ techniques_ for_the_analysis
[6] https://en.wikipedia.org/wiki/Data_mining
[7] J.han, M.Kambar and J.Pie,"data Mining:Concepts and Techniques", third Edition, Morgan Kaufmann(2012).
[8] Raj Kumar, Dr.Rajesh Velma, "classification Algorithm for Data Mining: A Survey" IJIET, 2012.