International Journal of Gender, Science and Technology
ISSN: 2040-0748
UGC Care Group I Journal.
Vol-11 Issue-02 Aug 2022

Predictive Analysis of Google Play Store App Ratings Using Decision Tree Classifier

BONTALA VENKATA AJITHA1, MENTA VIJAYABHASKAR2

#1Assistant Professor, Department of CSE, PBR Visvodaya Institute of Technology and Science, Kavali

#2Assistant Professor, Department of CSE, PBR Visvodaya Institute of Technology and Science, Kavali

ABSTRACT

As a language model, I understand the theoretical framework you have presented for building two inference engines based on the KNN and Random Forest algorithms to investigate the features that determine the best correlation for the rating of the applications and compute regression metrics using the Google Play Store database. The work is organized into sections, with section II presenting the theoretical framework and relevant themes, while section III goes over the materials and methods, with step-by-step instructions based on the CRISPDM for learning about the database

It appears that the main objective of the work is to explore the factors that contribute to the success of apps on app stores, such as the number of downloads, comments, and star ratings. The KNN and Random Forest algorithms are being used to infer relationships between these features and the rating of applications. Additionally, the work aims to assess the quality of the regression models produced by these algorithms by computing regression metrics

Overall, it seems that the focus of the work is on using machine learning techniques to gain insights into the characteristics of successful apps and develop models that can predict the rating of new apps based on their features

.

Keywords: Machine Learning, Decision Tree, Extra tree , Support vector and gradient boosting regressor ML techniques, evaluation

1.INTRODUCTION

Based on your text, it seems that the focus of the work is on addressing the challenges faced by app developers in the marketplace. The work aims to develop two inference engines using the KNN and Random Forest algorithms to investigate the features that determine the best correlation for application ratings. Additionally, the work

aims to compute and evaluate regression metrics using the Google Play Store database to identify the key factors that influence customer interest in apps.

To achieve these objectives, the work employs statistical hypothesis testing and the Wilcoxon logic to compare the two algorithms' robustness in forecasting application ratings. The results of the study could be valuable to the software development market by providing insights into the features that are most likely to generate customer interest and increase app sales

Overall, it seems that the work is structured to address the challenges of app development in the marketplace by utilizing machine learning techniques to identify the factors that drive customer interest and increase the likelihood of success for new apps

:

2. LITERATURE SURVEY

[1]. N. H. M. Shamsuddin, N. A. Ali, and R. Alwee "An overview on crime prediction methods," 6th International Student Project Conference

Recently, crime analyses have been required to demonstrate the dataset's complexity. Law enforcement personnel will benefit from this procedure in apprehending criminals and directing measures to prevent crime. The ability to predict future crimes based on their location, pattern, and timing may be a significant source of information from a strategic or tactical perspective. However, it is challenging to accurately and more effectively anticipate future crime given the rise in crime in contemporary society. Methods for crime prediction are essential for reducing crime rates and identifying potential crimes. Currently, academics are conducting a study to forecast crime based on specific inputs. The performance of prediction models can be evaluated using a variety of different prediction methods, such as support vector machines, artificial neural networks, and multivariate time series. Their findings are still subject to restrictions that prevent them from accurately predicting where crimes will occur. Numerous research papers have already been published on this topic. Consequently, we examined each one in depth and presented the findings in this paper. Our objective is to identify the system's current applications and investigate ways to enhance it for future requirements..

[2]. S. Kim, P. Joshi, P. S. Kalsi and P. Taheri "Crime Analysis Through Machine Learning,"

The first method assigned a unique number to each neighborhood and crime category when a particular crime occurs in a particular neighborhood. A binary number was given to the neighborhood and the day of the week when the crime occurred in the second method. If the crime occurred on that day in that neighborhood, the number was 1 and it was 0 otherwise. The use of machine learning in crime prediction is the subject of this study. In order to evaluate Vancouver's crime statistics over the past 15 years, this study employs two alternative data processing methods. Machine-learning predictive models like K-nearest neighbor and boosted decision trees are used to predict crime. Speech recognition, web search, self-driving cars, and a better understanding of the human genome are some of the current applications of machine learning. It has also made it possible to predict crime based on referenced data. Nominal class labels are possible with the supervised prediction method known as classification..

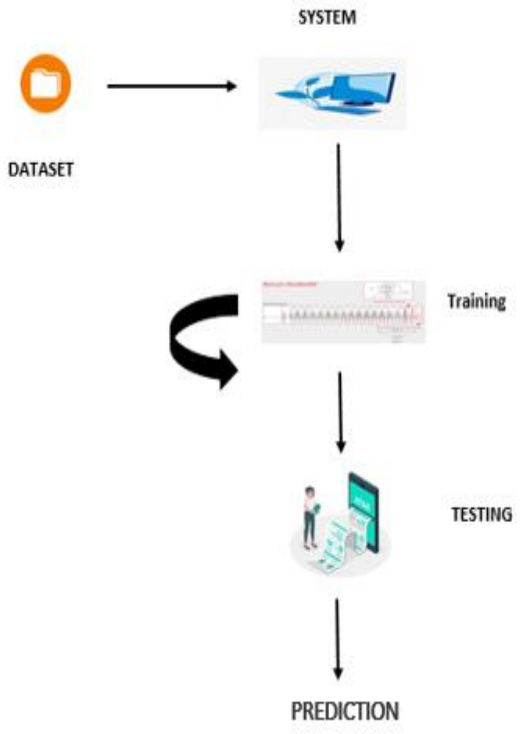
3. PROBLEM STATEMENT

KNN and Random Number Generators have been implemented in the current system. In the current approach, constructing models is done by mathematical computations, which can be very difficult and time-consuming. We employ machine learning tools from the Scikit-Learn library to get around all of this. extreme intricacy. It takes time..

4. PROPOSED SYSTEM

Many machine learning models were put up to categorise app ratings on the Google Play Store. Moreover, comparable studies that have presented models for evaluating this performance classification frequently ignore the heterogeneity and amount of the data. Hence, we suggest the machine learning techniques of Decision Tree, Gradient Boosting, Extra Tree, and Support Vector Regressors. maximum accuracy. eases the intricacy of time. Simple to use

5. ARCHITECTURE



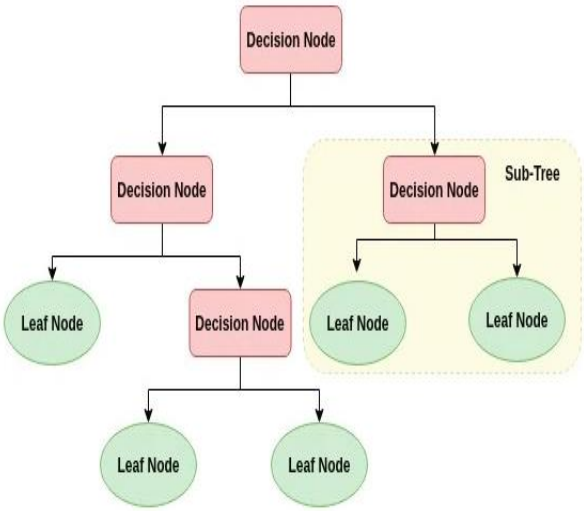
6. METHODOLOGY AND ALGORITHMS

6. 1. DECISION TREE:

Decision trees are a popular machine learning algorithm used for classification and regression tasks. They work by recursively partitioning the data into subsets based on the values of the input features until a stopping criterion is met, such as a maximum tree depth or a minimum number of samples per leaf.

Each internal node in the tree represents a decision based on one of the input features, and each branch represents one of the possible feature values. The leaves of the tree represent the predicted output values for the given input. Decision trees are easy to interpret and visualize, making them a popular choice for exploratory data analysis and predictive modeling in various fields

..



7. IMPLEMENTATION

7.1 User:

1.1 View Home page:

Here user view the home page of the play store application.

1.2 View about page:

In the about page, users can learn more about the rating platform.

1.3 View load page:

In the load_data page , the user will load the dataset for modelling.

1.4 Input Model:

The user must provide input values for the certain fields in order to get results.

1.5 View Results:

User view's the generated results from the model.

1.6 View score:

Here user have ability to view the r2_score in %

7.2 System

1.7 Working on dataset:

System checks for data whether it is available or not and load the data in csv files.

1.8 Pre-processing:

Data need to be pre-processed according the models it helps to increase the accuracy of the model and better information about the data.

1.9 Training the data:

After pre-processing the data will split into two parts as train and test data before training with the given algorithms.

1.10 Model Building

To create a model that predicts the personality with better r2_score, this module will help user.

1.11 Generated Score:

1.12 Here user view the score in %

1.13 Generate Results:

We train the machine learning algorithm and predict the google play store app

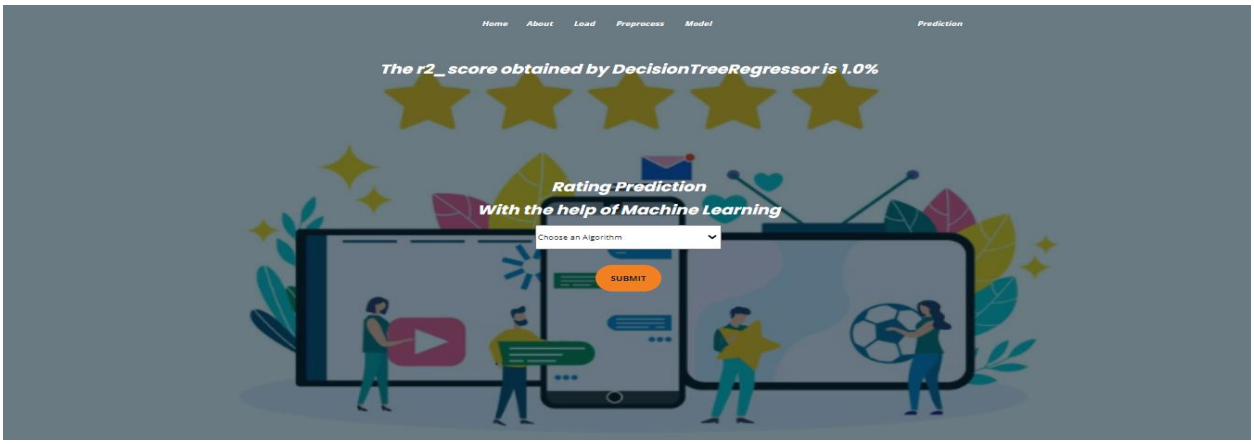
s.

8. RESULTS

Model:

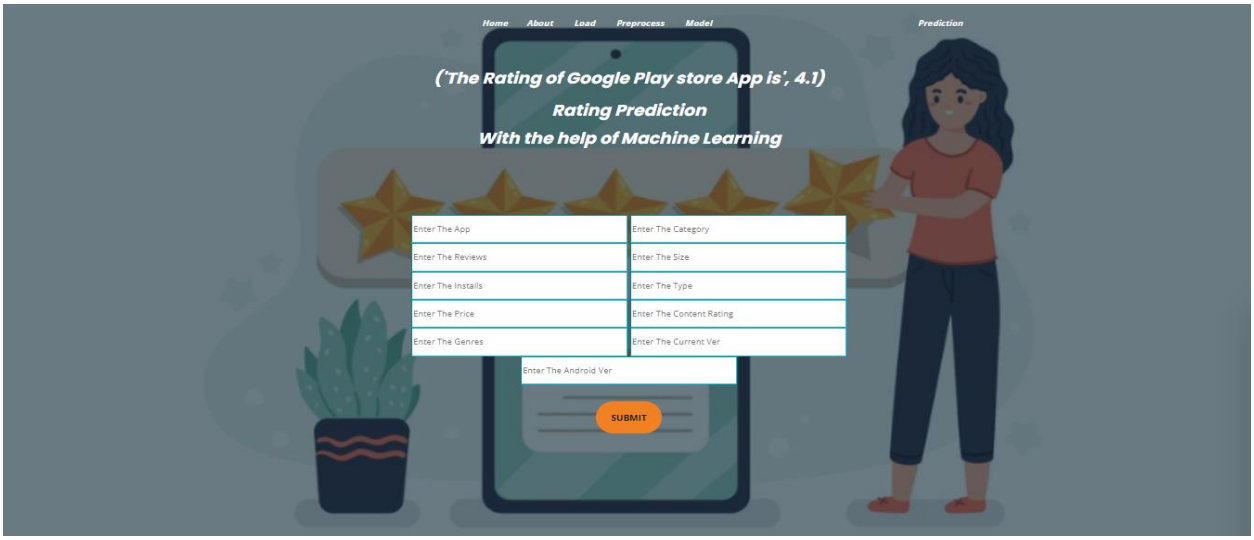
Here we train our data with different ML algorithms.

`



Prediction:

This page show the result of the data.



9. CONCLUSION

such as Linear Regression, Support Vector Regression, and Gradient Boosting Regression. The study found that Random Forest and Gradient Boosting Regression produced the best results in terms of predicting app ratings.Overall, this paper highlights the effectiveness of using machine learning techniques to extract knowledge from data and make predictions. It also emphasizes the importance of selecting appropriate models for specific data sets and problems. Finally, it demonstrates the potential for using these models in various contexts and applications beyond the Google Play store.

REFERENCES:

[1] "Google play store: number of apps 2018 — statista," https://www.statista.com/statistics/266210/number-of-available-applications-in-the-google-play-store/, (Accessed on 12/21/2018).

[2] "Basic concepts in machine learning," https://machinelearningmastery. com/basic-concepts-in-machine-learning/, (Accessed on 12/21/2018).

[3] Z. Zhang, "Introduction to machine learning: k-nearest neighbors," Annals of translational medicine, vol. 4, no. 11, 2016.

[4] L. Breiman, "Random forests," Machine learning, vol. 45, no. 1, pp. 5–32, 2001.

[5] G. Kesavaraj and S. Sukumaran, "A study on classification techniques in data mining," in Computing, Communications and Networking Technologies (ICCCNT), 2013 Fourth International Conference on. IEEE, 2013, pp. 1–7.

[6] E. A. Gehan, "A generalized wilcoxon test for comparing arbitrarily singly-censored samples," Biometrika, vol. 52, no. 1-2, pp. 203–224, 1965.

[7] M. R. Islam, "Numeric rating of apps on google play store by sentiment analysis on user reviews," in 2014 International Conference on Electrical Engineering and Information & Communication Technology. IEEE, apr 2014. [Online]. Available: https://doi.org/10.1109/iceeict.
2014.6919058

[8] Y. Tian, M. Nagappan, D. Lo, and A. E. Hassan, "What are the characteristics of high-rated apps? a case study on free android applications," in Software Maintenance and Evolution (ICSME), 2015 IEEE International Conference on. IEEE, 2015, pp. 301–310.

[9] X. Amatriain, J. M. Pujol, N. Tintarev, and N. Oliver, "Rate it again: increasing recommendation accuracy by user re-rating," in Proceedings of the third ACM conference on Recommender systems. ACM, 2009, pp. 173–180.

[10] S. Bidmon, R. Terlutter, and J. Rottl, "What explains usage of mobile ¨ physician-rating apps? results from a web-based questionnaire," Journal of medical Internet research, vol. 16, no. 6, 2014.