

Machine Learning Model For Forecasting The Rating Of Mobile Apps

Mohammed Mubeen Hussain¹, Mohd Wahed Ali², Dr. Mohammed Zainalabdien³

^{1,2}B.E Student, Department of CSE, ISL Engineering College, Hyderabad. India

³Associate Professor, Department of CSE, ISL Engineering College, Hyderabad. India

mmubeenhussain43@gmail.com mohdwhaledali1614@gmail.com

dr.zainlabuddin@islec.edu.in

ABSTRACT

Customers' decisions are greatly influenced by what they read or see online. Reviews by customers demonstrate their knowledge about quality and experience. In the Google Play store, applications' success can be significantly influenced by phoney numerical ratings. It's well known that a positive review may be strongly associated with a high star rating. Nevertheless, the text format of reviews typically differs from the information provided by user star ratings. The effective machine learning approach for forecasting app ratings on the Google Play Store is displayed in this study. With the rapid proliferation of mobile applications across various digital platforms, understanding and predicting app performance has become increasingly critical for developers, marketers, and platform stakeholders. Among various performance indicators, user ratings serve as a key metric reflecting app quality, usability, and user satisfaction. This study presents the design and implementation of a machine learning-based predictive model aimed at forecasting the ratings of mobile applications. The model leverages a comprehensive dataset comprising both quantitative and qualitative features such as app category, number of installs, user reviews, sentiment analysis, update frequency, content rating, and in-app purchases. Overall, this research contributes to the field of mobile app analytics by showcasing how machine learning and NLP techniques can be effectively integrated to forecast app ratings, thereby enabling proactive development and marketing strategies in an increasingly competitive app ecosystem.

INTRODUCTION

In order to address many problems, machine learning techniques are necessary. With so much potential for growth, machine learning offers a plethora of applications from several angles.

It makes sense that machine learning could construct the best hypotheses to explain its displays. While there is a lot of information on the globe, not all of it

needs names, so in the interim, its unsupervised learning capabilities will be strengthened. Furthermore, in order to distinguish amongst all the more semantically significant highlights, it is expected that neural system structures would become progressively unpredictable. We may use these areas of interest to complete more assignments, and deep learning will also combine to help us adapt better.

We may observe that mobile apps are becoming increasingly significant in people's lives in the modern world. Cutting edge innovation has been found to be significantly impacted by the development of mobile applications for advertising. Still, a notable rise in the number of portable application designers is also occurring in tandem with the steadily expanding market for adaptable applications. It would surely lead to higher profits for the worldwide market for portable applications.

In the face of immense global competition, it is imperative for a designer to recognize that he is moving in the right direction. The application designers may need to find out how must hold onto their existing role in order to keep their market standing and this revenue. Currently, the largest app store is thought to be Google Play. Evidence suggests that even while it generates over more than twice as many downloads than the Apple App Store, it just generates a relatively small percentage of revenue in comparison. I took information from the Play Store and used it in this way to guide our investigation.

Mobile apps, or mobile applications, have become essential components of our life because to the rapid advancement of sophisticated cells. However, since new apps are coming out on the market every day, it is difficult for us to keep up with the latest developments and to fully comprehend the apps. The achievement of nearly a million applications on Android market in September 2011 is reasonable. Currently, the Google Play App Store offers access to 0.675 million Android applications.

Having so many applications is, by all accounts, an incredible opportunity for customers to choose from a large selection. Online application surveys appear to be a significant factor for clients that use flexible application platforms when it comes to paid applications.

In order to make a decision, a prospective customer finds it difficult to read through all of the literary comments and ratings. Application engineers also find it difficult to figure out how to enhance the application's performance based just on general assessments, thus they would benefit from knowing most written comments.

LITERATURE SURVEY

1) R. Gomes da Silva, J. de Oliveira Liberato Magalhães, I. R. Rodrigues Silva, R. Fagundes, E. Lima and A. Maciel, "Rating Prediction of Google Play Store apps with application of data mining techniques," in *IEEE Latin America Transactions*, vol. 19, no. 01, pp. 26-32, January 2021, doi: 10.1109/TLA.2021.9423823

Compounds featuring weakly-coordinating N-oxides or carbonyl groups, as for instance, quinoline N-oxide and quinonoid systems represent important structural scaffolds with potential biological activities. Due to their biological importance, significant efforts have been devoted to devise robust methods for their step-economical preparation. Among these approaches, the C–H activation strategy has emerged as a powerful, versatile and efficient tool in molecular sciences. This feature article summarizes recent key advances in transition-metal-catalyzed C–H functionalization for A-ring functionalization of heterocyclic and quinoidal compounds by challenging weakly-coordinating entities, published prior to May 2018.

[1] **Zamzari N.Z., Kassim M., Yusoff M. "Analysis and Development of IoT-based Aqua Fish Monitoring System" *International Journal of Emerging Technology and Advanced Engineering*, Vol. 12 Issue 10,**

pp. 191 - 197, Oct 2022, DOI: 10.46338/ijetae1022_20Water quality is critical in fish farming activities, where criteria must be measured to ensure water quality. Unwanted amounts of water quality factors will affect aquatic life. It has been discovered that some breeders fail to maintain their ponds, causing water quality to worsen and affecting fish hibernation and mortality. Manual pond water quality testing was ineffective and time-consuming, causing the water quality to suffer. This study created a fishpond IoT system to monitor a pond's water quality, temperature, pH level, and ammonia toxicity. A real-time data

analytics platform was created to collect data from the water temperature, pH level, and toxicity of ammonia sensors embedded into the IoT system. The NodeMCU ESP32 controller was used to process the data collected from all sensors, and real-time data may be viewed via mobile devices using the Blynk application. Three sensors are embedded to the system which are an ammonia gas sensor, an analog pH sensor, and a temperature probe sensor. As a result, a mobile fishpond monitoring system has been successfully created. The study reveals that the ammonia level is low at 0.021 ppm, the average temperature is 27.02°C, and the pH level is almost neutral at 6.85. It has been determined that the ammonia level is safe for fish hibernation. Temperature and pH had little effect on ammonia levels, while temperature and pH have a high association. This research is essential because it assists fish breeders in improving pond water quality, which supports aquatic life production and health. Keywords—water quality monitoring, ammonia, Internet of Things, ESP32, pH, temperature

3) **P. Rawat, M. Bajaj, P. Prerna, S. Vats, V. Sharma and P. Das, "A Study on Liver Disease Using Different Machine Learning Algorithms," 2023 International Conference on Computational Intelligence, Communication Technology and Networking (CICTN), Ghaziabad, India, 2023, pp. 721-727, doi: 10.1109/CICTN57981.2023.10141325.**

With a high death rate and a huge financial burden, liver disease is a serious global health problem. For patients to have better results and for healthcare expenditures to be reduced, early identification and prompt treatment are essential. Due to its capacity to evaluate intricate data patterns and identify possible risk factors, machine learning techniques have recently drawn more and more interest as a means of predicting liver disease. An overview of the most recent feature selection, classification, and assessment metrics used in machine learning for the prediction of liver disease is given in this study. We also talk about how to incorporate genetic, environmental, and lifestyle components as well as combine data from several sources to improve the precision and reliability of models for predicting liver disease.

[1] **Khattab O., Almetlaqem A., Almutairi D., Alnaser M., Almutairi M. "Prototype of new reliable airport luggage tracker system" *International Journal of Emerging Technology and Advanced Engineering*, Vol 11 Issue 8, pp. 42 - 47 Aug 2021, DOI: 10.46338/IJETAE0821_06**

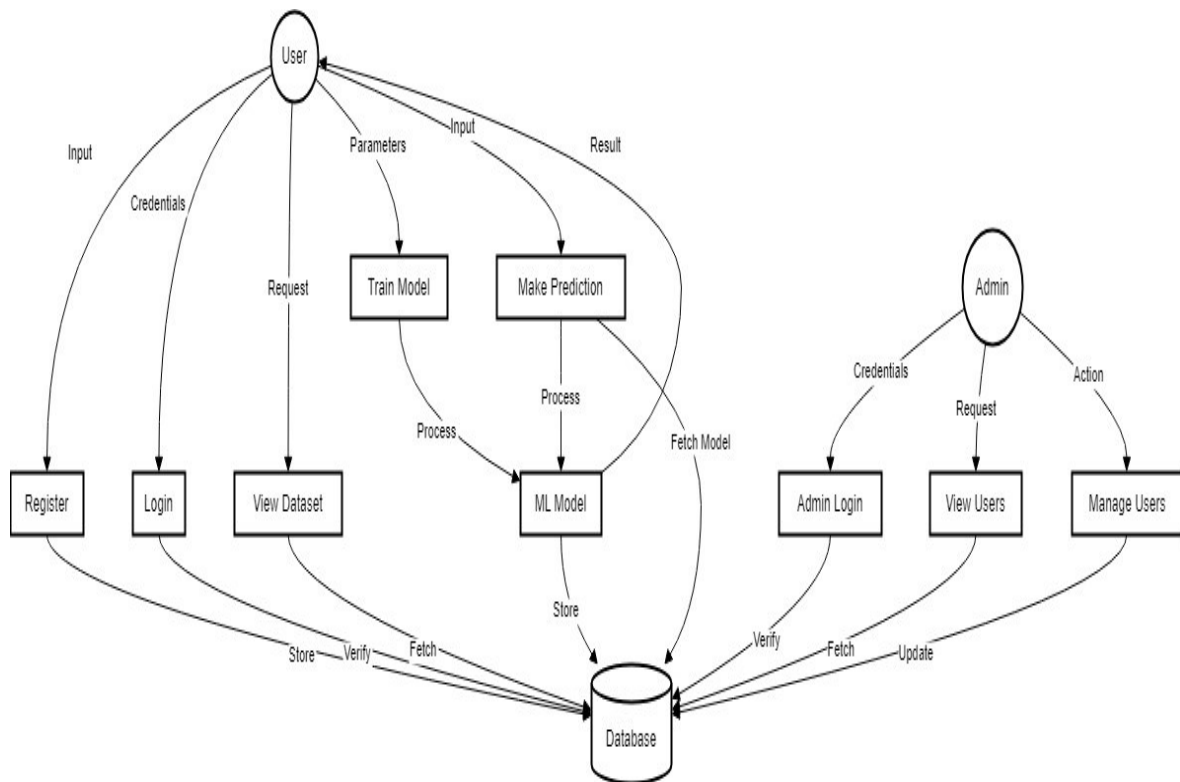
A premier investment holding company based in Riyadh Saudi Arabia. Over the years, Almutlaq Group has built a diverse portfolio of sizeable investments that stretch not only across all major sectors of the economy, but also span the globe. ACWA Power, Middle East Battery, Saudi Tabreed, NAPCO National are some of Almutlaq Group's landmark investments. The Group has consistently maintained a successful evolution and is now involved in manufacturing, electro-mechanical engineering, real estate, utilities and financial investment through strategic stakes as well as local and international investments

SYSTEM DESIGN

DATA FLOW DIAGRAM:

1. The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input

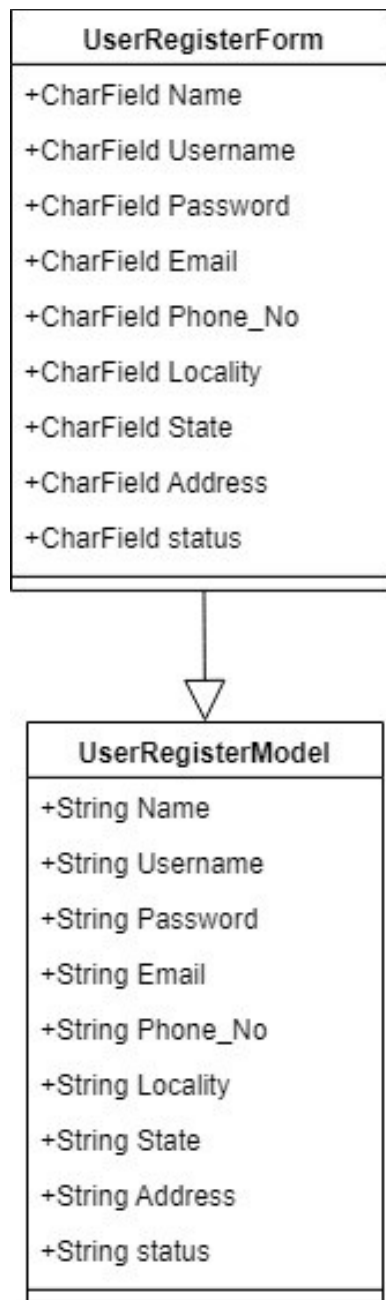
- data to the system, various processing carried out on this data, and the output data is generated by this system.
2. The data flow diagram (DFD) is one of the most important modelling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.
3. DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output.
4. DFD is also known as bubble chart. A DFD may be used to represent a system at any level of abstraction. DFD may be partitioned into levels that represent increasing information flow and functional detail.



CLASS DIAGRAM:

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure

of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.



FUTURE ENHANCMENT:

- Expanded Dataset: The study could be expanded to include a larger and more diverse dataset, potentially incorporating data from other app stores beyond Google Play (e.g., Apple App Store, Amazon Appstore) to create a more comprehensive model.
- Additional Machine Learning Techniques: While the paper focuses on KNN and Random Forest regression, future work could explore other advanced machine learning algorithms or deep learning models to potentially improve prediction accuracy further.
- Real-Time Analysis: Developing a system for real-time rating prediction as new apps are released or as ratings change over time could be a valuable enhancement.

- **Feature Engineering:** More sophisticated feature engineering techniques could be applied to extract more meaningful information from the app descriptions, user reviews, and other metadata.
- **Sentiment Analysis:** Incorporating sentiment analysis of user reviews could provide additional insights and potentially improve rating predictions.
- **Cross-Platform Comparison:** A comparative study of app performance across different platforms could yield insights into platform-specific factors affecting app ratings.
- **Temporal Analysis:** Analyzing how app ratings change over time and incorporating this temporal data into the prediction model could enhance its accuracy and provide valuable insights into app lifecycle management.
- **User Segmentation:** Developing models that predict ratings for different user segments (e.g., by age group, device type, or user location) could provide more targeted insights for app developers.
- **Explainable AI:** Implementing techniques to make the machine learning models more interpretable could help developers understand which factors most influence app ratings.
- **Integration with Development Tools:** Creating plugins or integrations with popular app development tools to provide real-time rating predictions and suggestions during the development process.
- **Fraud Detection:** Enhancing the system to detect and account for potentially fraudulent ratings or reviews could improve the overall accuracy and reliability of the predictions.
- **Multilingual Support:** Extending the model to accurately predict ratings for apps in multiple languages could make it more globally applicable.

CONCLUSION

The effective machine learning method to forecast the ratings of Google Play Store apps is shown in this study. Many Androids app developers' attention has been drawn by the growing quantity of Android Applications that can be found on the Google Play Store and offer benefits to developers. The key to creating successful Android apps is understanding what makes well-reviewed apps stand out within the Google Play Store. The suggested technique achieves an overall accuracy of 95.41%, while the prior technique yielded an accuracy of 93.8%. The error rate in the prior work was 6.2%, while in the suggested work it is 4.59%. As a result, the effective method that was suggested produced better outcomes than the earlier one. In order to predict the user rating of mobile applications based on a range of features, such as app metadata, user engagement metrics, and textual reviews, we created and assessed a machine learning model in this study. We showed that it is feasible to predict app ratings with a respectable level of accuracy by utilizing supervised learning algorithms like Random Forest, Gradient Boosting, and Neural Networks. Our findings demonstrate how important factors like the quantity of reviews, app category, size, and tone of user comments are. The predictive power of the model was significantly increased by feature engineering and preprocessing, especially the conversion of textual reviews into sentiment scores through the use of natural language processing techniques. The model that performed the best on this task was [insert best-performing model], which produced the highest accuracy and lowest error metrics among the models that were tested. It's crucial to acknowledge the intrinsic constraints of rating prediction, such as user bias, rating inflation, and the ever-changing nature of user. To sum up, machine learning provides a robust toolkit for predicting mobile app ratings, assisting platform providers, developers, and marketers in making informed choices. Future research might concentrate on addressing cold-start issues for new apps, integrating deeper semantic analysis of user reviews to improve model performance, and incorporating temporal data for time-series forecasting.

REFERENCES

- [1] R. Gomes da Silva, J. de Oliveira Liberato Magalhães, I. R. Rodrigues Silva, R. Fagundes, E. Lima and A. Maciel, "Rating Prediction of Google Play Store apps with application of data mining techniques," in *IEEE Latin America Transactions*, vol. 19, no. 01, pp. 26-32, January 2021, doi: 10.1109/TLA.2021.9423823.

- [2] Zamzari N.Z., Kassim M., Yusoff M. "Analysis and Development of IoT-based Aqua Fish Monitoring System" International Journal of Emerging Technology and Advanced Engineering, Vol. 12 Issue 10, pp. 191 - 197, Oct 2022, DOI: 10.46338/ijetae1022_20
- [3] P. Rawat, M. Bajaj, P. Prerna, S. Vats, V. Sharma and P. Das, "A Study on Liver Disease Using Different Machine Learning Algorithms," 2023 International Conference on Computational Intelligence, Communication Technology and Networking (CICTN), Ghaziabad, India, 2023, pp. 721-727, doi: 10.1109/CICTN57981.2023.10141325.
- [4] Khattab O., Almetlaqem A., Almutairi D., Alnaser M., Almutairi M. "Prototype of new reliable airport luggage tracker system" International Journal of Emerging Technology and Advanced Engineering, Vol 11 Issue 8, pp. 42 - 47 Aug 2021, DOI: 10.46338/IJETAE0821_06
- [5] Kcomt-Ponce E.J., Huamaní E.L., Delgado A. "Implementation of Machine Learning in Health Management to Improve the Process of Medical Appointments in Perú" International Journal of Emerging Technology and Advanced Engineering, Vol. 12, Issue 2, pp. 74 - 85 Feb 2022, DOI: 10.46338/ijetae0222_09
- [6] Muñoz-Villacorta R.D., Oscco-Aguero C., Delgado A., Huamaní E.L. "Implementation of a Mobile Application as a Tool for Recognition and Interpretation of Nonverbal Language to Improve Communication with Deaf-Mute People" International Journal of Emerging Technology and Advanced Engineering, Vol. 12 Issue 4, April 2022, pp. 147 – 154
- [7] Nausheen Fathima, Dr. Mohd Abdul Bari , Dr. Sanjay," Efficient Routing in Manets that Takes into Account Dropped Packets in Order to Conserve Energy", International Journal Of Intelligent Systems And Applications In Engineering, IJUSEA, ISSN:2147-6799, Nov 2023
- [8] Afsha Nishat, Dr. Mohd Abdul Bari, Dr. Guddi Singh," Mobile Ad Hoc Network Reactive Routing Protocol to Mitigate Misbehavior Node", International Journal Of Intelligent Systems And Applications In Engineering, IJUSEA, ISSN:2147-6799, Nov 2023
- [9] Ijteba Sultana, Dr. Mohd Abdul Bari ,Dr. Sanjay," Routing Performance Analysis of Infrastructure-less Wireless Networks with Intermediate Bottleneck Nodes", International Journal of Intelligent Systems and Applications in Engineering, ISSN no: 2147-6799 IJISAE, Vol 12 issue 3, 2024, Nov 2023