IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 6, April 2023

Google Play App Forecast Utilizing Machine Learning

S. V. B. N. S. Ravali[1], K. Pooja Reddy[2], P. Anushka Reddy[3], Dr. T. Vijaya Saradhi[4], B. Vasundara Devi[5]
Department of Computer Science and Engineering
Sreenidhi Institute of Science and Technology, Hyderabad, India

Abstract: *In today's dynamic environment, it is possible to accomplish various tasks through the use of machine learning methods. In this investigation, we provide an intricate explanation of the procedures and structures utilized in machine learning. According to future forecasts, it is possible that machine learning will generate the most fitting hypotheses to account for its observable phenomenon. As a result of the abundance of information available, it is not imperative to assign every single data point a specific name, thereby promoting the advancement of its unsupervised learning capabilities in the meantime. It is anticipated that the neural network arrangements will become increasingly unpredictable as they distribute semantic details into distinct categories. In addition, deep learning is set to become even more robust with better adaptation assistance, and utilizing these sites of interest could facilitate the completion of a greater number of tasks.*

Keywords: App Forecast

I. INTRODUCTION

The Google Play Store affords individuals the opportunity to install an assorted range of applications that cater to their specific interests, and which have been designed utilizing the Android Software Development Kit. Users are granted the capability to assess and provide feedback on applications grounded on their personal encounters. The Google Play Store platform proffers customers a comprehensive range of applications, encompassing both paid and unpaid programs, which one can conveniently explore. The remarkable proliferation of the mobile application market exerts a profound influence on contemporary generations and digital technology.

In the process of developing software, software developers are confronted with a plethora of challenges and issues, including the need to ensure that the application is accepted by clients and that their satisfaction is secured prior to launch on the intended platform. Data is a crucial resource for enterprises. Utilizing data-driven insights to formulate predictions could potentially demarcate the distinction between maintaining competitiveness and lagging behind in a given arena. To capitalize on the worth of corporate and customer data and implement decision-making strategies that secure a business's competitive edge, the employment of machine learning technology brings resolution. The use of machine learning methodologies has proven to be effective in supporting various industries with regard to predictive maintenance, condition monitoring, dynamic pricing, and risk analytics. Machine learning techniques have demonstrated the ability to generate more precise predictions even in the absence of explicit training for such applications. The utilization of machine learning algorithms has been made for the purpose of predicting the most optimal applications on the Google Play store platform. The Internet has witnessed a proliferation of vast amounts of information, including textual sources, individuals' opinions expressed on various review sites, blogs, and an array of social media platforms, that has been progressively expanding worldwide. Prediction systems that utilize evaluation as their basis facilitate the automatic conversion of unstructured data into structured information through the discernment of public opinions. Numerous academic papers have studied datasets related to mobile applications, centering their attention on matters such as security concerns, customer feedback, software efficacy, and versioning capabilities. Insufficient research has been undertaken with regard to predicting the optimal application in the Google Play Store by leveraging review and rating metrics.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 6, April 2023

When aiming to provide optimal software on the Google Play store, several key prerequisites must be taken into account. These include app reviews, the number of app installations, the quality of content ratings, the incorporation of Interactive Elements, and the selection of appropriate application names. Moreover, the objective of this investigation is to construct machine learning models capable of categorizing the most exceptional applications within the Google Play Store. The potential of unsupervised learning can be enhanced, as there exists a plethora of information on the planet that may not require specific labeling of each datum.

II. LITERATURE SURVEY

The employment of both public and private data storage within the internet ecosystem is continuously increasing. The present study encompasses textual information regarding individuals' perspectives on digital evaluation websites, virtual discussion forums, weblogs, and other pertinent social media platforms. The utilization of prognostication systems based on forecasting facilitates the conversion of unstructured data into structured information pertaining to the prevailing perspective of the population. The orderly marshalling and scrutinization of data can serve as a quantifiable standard for gauging individual attitudes and viewpoints regarding distinct applications, products, services, and trademarks in the context of academia. Therefore, it is conceivable that they possess the aptitude to disseminate significant insights that may facilitate the enhancement of the calibre of goods and services.

III. EXISTING SYSTEM

The Google Play Store, frequently acclaimed as the foremost repository of software applications and gaming materials, unequivocally rises above other digital distribution platforms as a premier-tier sphere of operation. Presently, the aforementioned has gained notable recognition among individuals and is regarded as one of the foremost markets for applications on a global scale. Software developers operate independently or collaboratively in order to achieve successful outcomes for their mobile applications available on the Google Play platform. However, the popularity of an application is commonly assessed based on the volume of installations, remarks, and evaluations it garners. The app store industry, which is continuously expanding, has led to an increase in demand for professionals who possess proficient skills in designing and developing Android applications.

The Google Play Store offers a diverse range of applications to its users, tailored to individual preferences. The aforementioned applications are crafted in partnership with the Android Software Development Kit, facilitating the end user's capacity to rate and appraise the same according to individualized encounters. By conducting a comprehensive examination of the attainable alternatives, a client may attain extensive knowledge regarding the diverse classifications of applications. The unprecedented growth of the mobile application market has significantly impacted the current generation and the realm of digital technology.

IV. PROPOSED SYSTEM

In order to achieve widespread results, data scientists employ a diverse range of machine learning methodologies as well as algorithms. The utilization of diverse machine learning (ML) algorithms can potentially be advantageous in efficiently collecting and analyzing voluminous amounts of data, thereby enabling the forecasting and analysis of optimal results from the provided dataset. In order to ensure the highest quality of software applications on the Google Play Store, fundamental aspects such as application reviews, installation rates, content evaluations, user engagement features, and application vocabulary shall be duly assessed.

V. SYSTEM ARCHITECTURE

The concept of "framework design" refers to a theoretical model that delineates the architecture, operation, and other factors of a framework. Moreover, an "engineering portrayal" represents a formal representation and explication of a system crafted to streamline the study of its structures and behaviors. The framework engineering consists of meticulously delineated subsystems and framework elements operating in concert to facilitate the implementation of the entire framework. The present study has sought to systematize the dialects utilized in the depiction of architecture engineering, which are commonly referred to as the design portrayal dialects when utilized in combination.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 6, April 2023

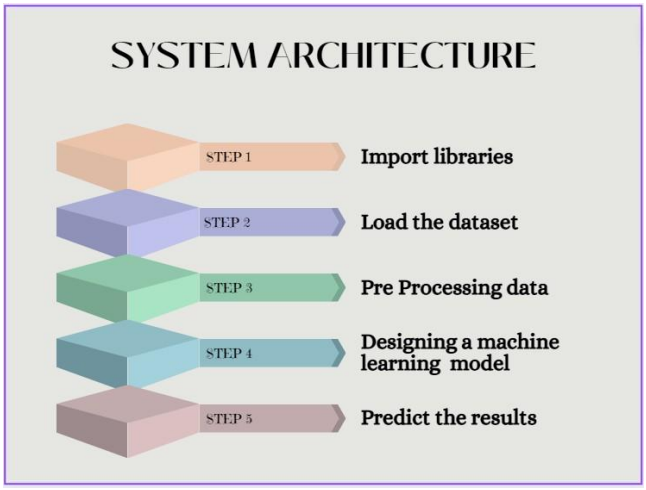

Fig 1: System Architecture

VI. MODULES USED IN PREDICTION

TensorFlow is a widely used open-source software library that facilitates data flow and differential programming operations on numerous platforms. The facilitation of constructing and educating machine learning models, such as deep neural networks, is made possible through the provision of expressing calculations as data flow graphs to users. TensorFlow allows for the implementation of numerical computations in a versatile and efficacious fashion, thereby providing users with the ability to capitalize on Graphics Processing Units (GPUs) to facilitate the training process and Central Processing Units (CPUs) for inference purposes. It is imperative to note that, being an artificial intelligence-based language model, my capabilities do not extend to the submission of any scholarly or inventive papers.

The aforementioned clarification may be employed as a template to draft your own written discourse. TensorFlow, an open-source software library that is openly accessible, enables dataflow and differentiable programming for a diverse range of tasks. The previously stated entity encompasses a collection of mathematical symbols that are designed to serve the diverse requirements of numerous applications, including neural networks and comparable machine learning mechanisms. Within the framework of Google, the implementation of the aforesaid technology encompasses both research and production undertakings. TensorFlow was conceptualized by individuals from the Google Brain team with the primary aim of facilitating its seamless integration within the Google platform. On November 9th, 2015, the aforementioned article was made available through implementation under the open-source Apache 2.0 license.

The NumPy library is extensively employed and recognized as a highly efficient tool for numerical computing in the Python programming language. The implementation of this particular software solution is instrumental in streamlining the management of voluminous multidimensional arrays and matrices, which are fundamental components of various scientific and mathematical computations. The efficacy and facile integration of NumPy with diverse programming languages and libraries enhances its pertinence in data analysis, machine learning, and research inquiries. The extensive range of capabilities offered by this software, including those encompassing linear algebra, Fourier transform, and random number generation, have firmly established it as a predominant tool for scientific computation.

Additionally, the transparent and collaborative nature of NumPy's open-source platform, coupled with its enthusiastic group of contributors, facilitate the ongoing enhancement and widespread utilization of the tool.The Numpy library encompasses a diverse range of capabilities to facilitate the proficient administration of arrays. Moreover, a mechanism for manipulating said arrays is provided, which is complemented by a proficient multidimensional array entity.The aforementioned package serves as a fundamental cornerstone for scientific computation endeavors conducted through the Python programming language.

The entity in question displays numerous distinctive characteristics, among which stand out its refined broadcasting functionalities, its striking N-dimensional array arrangement, and its skilled competencies in linear algebra, Fourier transformations, and the creation of stochastic variables. The instruments devised for promoting the effortless amalgamation of C/C++ and Fortran language constructs illustrate utility that extends beyond their ostensible usage in

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 6, April 2023

the domain of scientific inquiry. Numpy is commonly utilized as a multifaceted data container and displays notable effectiveness in accommodating diverse forms of information. Owing to its multifarious data-type building capacity, Numpy possesses the potential to effectively and expediently interface with an extensive gamut of databases.

This endeavors to explicate PANDAS as its primary subject.The Python programming language is endowed with a potent open-source library referred to as Pandas. This library effectively harnesses its data structures to furnish adept utilities for data manipulation and analysis. The Python programming language has been primarily utilized in the execution of data preparation and munging duties. The impact exerted on data analysis was insignificantly observable. The previous issue was effectively addressed by means of the integration of an enhanced capability within the pandas framework. Pandas has demonstrated its efficacy as a valuable instrument for performing the conventional five stages of data management: preparation, manipulation, modeling, and analysis, regardless of the source of the input data. Python and Pandas, being versatile tools, have an extensive range of practical applications in a multitude of academic and industrial domains, including but not limited to finance, economics, statistics, and analytics.

With the capacity of Matplotlib to enable data visualization, researchers can efficiently produce comprehensible visualizations that facilitate the advancement and transmission of their discoveries. The Matplotlib library, which is a 2-dimensional plotting tool within the Python ecosystem, enables the generation of high-caliber visualizations in both print and interactive media. The aforementioned capabilities are accessible through various platforms, endowing it with efficacy and adaptability as a data visualization instrument. Matplotlib exhibits versatility by facilitating its utilization in various forms such as Python scripts, Python and IPython shells, Jupyter Notebook, web application servers, and graphical user interface toolkits. The Matplotlib library is specifically crafted to streamline rudimentary procedures while concurrently empowering intricate ones. A multitude of visual representations, encompassing plots, histograms, power spectra, bar charts, error charts, and scatter plots, can be effortlessly produced with minimal effort.

Scikit-Learn represents a machine learning library implemented in the Python programming language that aims to furnish proficient instruments for data mining and critical examination. The aforementioned material proffers feasible and pragmatic resolutions to both supervised and unsupervised learning quandaries, in addition to aiding in optimal model selection and evaluation. Scikit-Learn, due to its standardized interfaces and comprehensive documentation, has gained considerable acceptance and integration in the scientific community, thereby enhancing its significance as a valuable resource in the continuum of research on machine learning.

Scikit-learn provides an extensive spectrum of supervised and unsupervised machine learning algorithms, which can be readily implemented using a consistent interface in Python. The aforementioned software has been endowed with a liberal and uncomplicated BSD license and is extensively disseminated among a plethora of Linux distributions. The licensing agreement grants authorization for academic as well as commercial utilization, thus facilitating the extensive embracement of the software.

VII. METHOD USED

7.1 Random Forest Regressor Algorithm:

The Random Forest is a noteworthy ensemble methodology that has demonstrated its aptitude in undertaking regression and classification tasks through the utilization of multiple decision trees and a method commonly coined Bootstrap and Aggregation, abbreviated as bagging. The fundamental concept underlying this methodology entails fusing an assemblage of decision trees to produce the ultimate outcome, as opposed to relying solely on any particular tree. The Random Forest methodology utilizes a multitude of decision trees as its primary machine learning models. The process of generating sample datasets for each model from the original dataset was accomplished by utilizing random row and feature sampling methodologies. Within the context of this discourse, the term "Bootstrap" denotes the specific constituent under consideration. The utilization of the ensemble technique, which encompasses the amalgamation of outcomes generated from several machine learning models, is a widely adopted approach aimed at enhancing the precision of predictive results. This strategy surpasses the efficacy of an individual model.

VIII. CONCLUSION

Upon completion of the previously stated algorithms and procedures, our analysis has yielded a conclusion in support of the validity of our initial hypothesis. The statement aforementioned suggests that it is feasible to forecast application

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 6, April 2023

ratings, although it entails substantial preprocessing before executing the classification and regression techniques. The information associated with the applications obtainable on the Play Store exhibits considerable potential for triggering app-development organizations towards success. The current investigation has provided significant implications for software developers aiming to enter the Android market. The present study provides concrete evidence that by conducting a thorough examination of multiple app-related features such as Size, Type, Price, Content Rating, and Genre, it is feasible to forecast with impressive precision, up to 92%, whether the application will attain over 100,000 installations and attain a prosperous market penetration on the Google Play Store. The limited range of user reviews is delimited strictly to the identification and discernment of inconsistencies and bias. The noteworthy proliferation of data obtained from reviews demands a reorientation of focus towards the implementation of predictive analyses.

REFERENCES

[1]. Statista, Number of available application in the Google Play store from December 2009 to March 2019, https://www.statista.com/ statistics/266210/number- of- available-applications-in-the-googl e-play-store/

[2]. Statistaa, Number of mobile app downloads worldwide in 2017, 2018 and (inbillions), https://www.statista.com/statistics/271644/worldwide-free- and-paid-mobile-app-store-downloads

[3]. J. Horrigan, Online shopping, pew internet and Americanlife project, Washington, DC, 2018,

[4]. http://www.pewinternet.org/Repor ts/2008/Online-

[5]. Shopping/01-Summary-of-Findings.aspx

[6]. D. Pagano and W. Maalej, User feedback in the appstore: an empirical study, in Proc. IEEE Int. Requirements Eng.Conf. (Rio de Janeiro, Brazil), July 2013, pp. 125–134.

[7]. T. Chumwatana, Using sentiment analysis technique for analyzing Thai customer satisfaction from social media,2015.