

**Department of Computer Science and Engineering**

**BIG DATA ANALYTICS LAB – CSE 3263**

***Mini Project on***

**GOOGLE PLAYSTORE ANALYSIS**

**Name: Bhavya Marupura**

**Register no: 210962168**

**Section: B**

**Roll no: 50**

**Department of Computer Science and Engineering**

**Manipal Institute of Technology, Manipal.**

**April 2024**

**GOOGLE PLAYSTORE ANALYSIS**

***Abstract*—** **The Google Play Store dataset has been exhaustively analyzed through keyword analysis, categorization, ratings, version trends involving linear regression and decision trees using Big Data technologies mainly PySpark. Using the powerful features of PySpark, we have derived insights from the vast data set by looking at the popular themes as identified by keyword analysis; sorting applications into related categories; studying what affects user ratings; also, for understanding how changes to versions impact on user engagement. In addition, techniques like linear regression and decision trees contribute to our more in-depth examination of predictive modeling and relationships existing between various app attributes and user behavior. Such an analysis is useful for both developers and other stakeholders as it helps them know what they are doing when making decisions about apps in a highly dynamic market landscape such as that of mobile technology.**

Keywords— Big Data, Data Analysis, PySpark, linear regression, Decision trees, Google Play Store Dataset, Categorization, Ratings and Versions

# Introduction

In a time that was completely overhauled by digital disruption and mobile technology, the Google Play Store has turned out to be an immense marketplace with millions of applications tailored to specific needs and desires of people all over the world. Thus, this huge web not only offers numerous apps but also creates a lot of relevant information which is very important for app developers, marketers, as well as decision-makers. Nonetheless, capturing these insights has been a daunting task due to three key characteristics of big data; volume, velocity and variety.   
  
This project is expected to circumvent this challenge through the application of Big Data technologies specifically PySpark for holistic analysis in relation to Google play store dataset. By employing highly developed information processing techniques we are hopeful that we will be able to get useful patterns from massive data repositories. In particular, some aspects have been examined like keyword analysis, categorization ratings version trends and finally features such as linear regression or even decision trees may be engaged in predictive modeling.

The significance of this project lies in the possibilities it brings for useful insights that can inform decision-making in the mobile app industry. Developers and stakeholders who want to improve user experience, optimize app performance, and maintain a competitive edge must understand customer preferences, business trends, and impacts of updated apps. By delving into complexities involved in Google Play Store dataset, this project seeks to highlight important correlations, patterns as well as opportunities that could be used in guiding application development strategies, marketing campaigns or user engagement initiatives. In addition, we provide practical guidance and actionable recommendations about how best to navigate the ever changing landscape of the mobile app market through a thorough analysis and interpretation of data.  
  
This section will explore our methodology findings and implications to shed light on how big data can influence future applications usage and design towards enhancing good user experiences

# LITERATURE REVIEW

Using a data set from Kaggle which was scraped from the Google Play Store, Rimsha Maredia [1] explores the capabilities of different classifiers to predict future outcomes. Even though there are no significant relationships between app facets such as rating, size, price and installs, the research finds that installs have strong negative correlation with reviews. Decision Tree is the best classifier with an accuracy of 95.32% for apps in genres such as Tools, Entertainment and Education. The study highlights how diverse this database can be for market analysis and future application development trends that may extend beyond immediate problem statement indicating scope like regression modeling on parameters like reviews or installs and also exploring relationship between app size and Android version on installs. Moreover, volume 2 shows that Decision Tree is fast and more accurate than other classifiers.

S. Shashank and Brahma Naidu[2] tackle the challenge of predicting app ratings in the Google Play Store amidst its ever-expanding app ecosystem. They emphasize the significance of app ratings as pivotal indicators of success, despite potential biases and incomplete data. Leveraging machine learning algorithms and a Kaggle-sourced dataset, the authors endeavor to unveil correlations between app attributes and ratings. Through rigorous analysis, they confirm the feasibility of accurately predicting app ratings, particularly focusing on attributes like app size, type, price, and genre. Their study demonstrates promising results, achieving around 92% accuracy in predicting highly installed apps. Looking ahead, they advocate for advanced techniques like ensemble classifiers and deep learning to refine rating predictions further. By offering insights into app development strategies, their research aims to empower developers and enhance competitiveness in the Android app market.

Shubham Ruhela[3] delves into the challenges confronting developers in the fiercely competitive environment of the Google Play Store. With a constant influx of new applications and uncertain revenue models, success hinges on factors like installation numbers and user ratings rather than revenue generation. Ruhela's project aims to analyze and predict trends within the Play Store dataset sourced from Kaggle, utilizing Big Data techniques like Hive for attribute analysis and Deep Learning for sentiment prediction of user reviews. Through rigorous model training, he significantly boosts prediction accuracy, offering valuable insights for developers navigating the complexities of app development and user feedback interpretation within the Play Store ecosystem.

Kabila MD Musa[4] offers a comprehensive exploration of the Google Play Store datasets, highlighting the significance of understanding app categories, user reviews, and pricing dynamics in today's digital landscape. With mobile applications playing a pivotal role in daily life, Musa emphasizes the necessity of leveraging Python and data analysis tools to unravel insights within these datasets. The journey begins with loading and cleaning the dataset, ensuring accuracy for subsequent analysis. Through techniques like histograms, KDE, and correlation matrices, Musa uncovers relationships and distributions within the data, shedding light on app attributes and user preferences. Delving deeper, specific niches such as paid family apps and top paid genres are examined, alongside analyses of categories by installations and mean prices. The pinnacle of the exploration reveals the most expensive apps per category, providing valuable insights into the strategies shaping the Google Play Store ecosystem. Ultimately, Musa's review underscores the transformative potential of data analysis in unraveling the complexities of the digital marketplace.

Christos Fragkathoulas[5] provides a detailed analysis of real-world data from the Google Play Store, shedding light on various aspects of app categories, user engagement, pricing, and updates. Through a comprehensive examination of the dataset, Fragkathoulas identifies key trends and draws significant conclusions regarding the diversity of applications across categories, user review activity, and application sizes. Notably, he highlights the prevalence of application updates in 2018 and the correlation between high ratings and installations. The most common application categories installed on users' smartphones are outlined, along with insights into supported Android versions. Fragkathoulas also discusses the distribution of application downloads, noting both highly popular and lesser-known apps. Furthermore, he delves into the pricing dynamics, highlighting the presence of expensive applications like "I am rich." The timing of app upgrades, average application size, and user preferences for free versus paid applications are also explored. Through this analysis, Fragkathoulas offers valuable insights into the dynamics of the Google Play Store ecosystem, providing a foundation for further research and understanding of app distribution platforms.

Aayush Gupta and Deepali Kamthania[6] propose a methodology for sentiment analysis of Google Play Store applications using a Kaggle dataset. They preprocess the data through various cleaning steps and employ a logistic regression model, achieving 81.1% accuracy in sentiment classification. Focusing on top applications like 'Candy Crush Saga' and 'Clash of Clans', they find that while the 'family' category boasts more apps, games are the most popular. Despite their popularity, these games often garner negative reviews, highlighting the importance of user sentiment in shaping market trends and informing development strategies. The study advocates for leveraging user feedback for application improvement and suggests extending the analysis to other categories for enhanced popularity.

M Sai Kiran, P V S Karthikeya, M S Kushal Raj, and Manuka Varun Teja[7] highlight the challenges in app sales and client acceptance within marketplaces, emphasizing the importance of adhering to established implementation standards in software development. They propose constructing inference engines using SVM and Random Forest algorithms to predict app ratings based on key metrics such as downloads, comments, and ratings. Through their analysis of a Google Play Store database, they confirm the feasibility of predicting app ratings but stress the need for extensive preprocessing. The authors suggest that Google Play Store data holds significant potential for aiding app development companies in navigating the Android market, providing actionable insights for developers. Additionally, they explore the predictive capabilities of ensemble classifiers for app ratings based on user reviews, hinting at future applications of deep learning technology in this domain. Overall, their research underscores the importance of data-driven approaches in understanding and predicting app success in dynamic marketplaces.

Yanqing Shen [8] addresses the challenge faced by app developers in standing out amidst the vast competition of the Google Play Store, comprising approximately three million apps. The objective of their project is to provide insights into customer preferences, particularly focusing on factors influencing app popularity, measured by the number of installations. Leveraging a dataset from Kaggle with 10,841 observations and 13 variables, including categories, prices, ratings, and reviews, they aim to understand the relationship between these variables and app installations. Notably, Shen identifies two areas for improvement in the project: the potential omission of relevant variables in the model and the necessity to consider the number of ratings, as high ratings from few reviewers may skew results. Furthermore, they advocate for analyzing the sentiment of reviews to mitigate errors, emphasizing the importance of not only the number but also the sentiment of reviews in influencing app installations. The study highlights the significance of data-driven approaches in uncovering insights to aid app developers in navigating the competitive landscape of the Google Play Store.

Mehra Anshita[9] highlights the pivotal role of data-driven insights in informing business decisions and maintaining competitiveness. She underscores the significance of machine learning techniques in leveraging corporate and customer data for accurate predictions, particularly within industries like predictive maintenance and risk analytics. Focusing on the Google Play Store platform, the paper compares machine learning models to predict the best applications. Through analysis of 29,999 records, the study reveals user preferences for free applications and the importance of factors like rating in app installations. Four supervised machine learning methods are employed, with the decision tree classifier emerging as the most accurate. The research identifies critical variables such as installs and rating, essential for machine learning algorithms and decision-making within the Google Play Store ecosystem. Overall, the study highlights the potential of machine learning in driving insights and improving decision-making processes in the app development domain.

Akhlak Ali Sunasara, Nancy Jaiswal, Suchit Poojari, and Anil Kumar Chaturvedi[10] highlight the omnipresence and increasing importance of software applications across industries. Focusing on the Google Play Store, their project aims to analyze app descriptions and associated data to understand customer preferences, predict trends, and provide insights for organizations and developers. They seek to identify top and bottom-performing apps, predict future trends, conduct sentiment analysis, assess app sustainability, and analyze the correlation between downloads and ratings. By offering insights into app development and user preferences, the project aims to assist developers in decision-making processes. The authors emphasize the dataset's potential to enhance business value and suggest future enhancements like live download tracking and automated app creation. Overall, the research underscores the significance of data-driven insights in shaping app development strategies and improving user experiences in the digital marketplace.

The Burgeoning[11] field of mobile application analysis relies heavily on user-generated reviews and ratings to gauge user satisfaction, preferences, and sentiments towards various applications. User reviews, while invaluable, present challenges due to their sheer volume and variability in length, tone, and language. To address these challenges, researchers employ text classification techniques, including sentiment analysis and topic modeling, alongside a range of machine learning algorithms such as logistic regression, support vector machines, and deep learning models. Evaluation metrics such as accuracy, precision, recall, and F1 score are utilized to assess the performance of these models. Future research directions include expanding analysis to encompass a wider range of application categories, comparing the efficacy of different algorithms like logistic regression, and exploring clustering techniques to uncover deeper insights into the relationship between reviews and ratings. Continued research in this field promises to drive innovation and improvement in mobile application development and user experience.

The author, Asaad Albusaidi,[12] delves into an exploration of the Google Play Store ecosystem, elucidating its significance as a digital distribution service operated by Google Inc. He emphasizes its role as the primary platform for Android applications, facilitating both browsing and downloading of apps developed using the Android SDK and disseminated through Google. Moreover, Albusaidi underscores Google Play's multifaceted nature as a digital media repository, offering an extensive array of music, books, movies, and television programs. Through his investigation, he scrutinizes the dichotomy between free and paid apps, shedding light on the prevalence of in-app purchases and advertisements as supplementary revenue channels. Furthermore, he conducts an analysis of market shares across various app categories, identifying those with the highest and lowest installation rates. With a keen eye on industry trends, Albusaidi identifies gaming, social, and communication categories as ripe for innovation and financial gain. In contemplating future endeavors, he acknowledges the potential of user feedback analysis in guiding developers towards iterative improvements, thereby enhancing user experience and application quality.

In her project, Suhita Vaidya[13] aims to analyze datasets from Kaggle, focusing on Google Play Store applications and user reviews. She seeks insights to enhance user experience and optimize app development through visualization. With data from 'playstore\_apps.csv' and 'playstore\_reviews.csv', Vaidya examines app details, ratings, installs, and pricing. She uncovers trends like the prevalence of free versus paid apps, revenue distribution across categories, and correlations between app ratings and categories. Vaidya identifies popular apps, revenue generators, and patterns in user sentiment, conducting bivariate analyses on factors like price versus rating and type versus last updated to reveal deeper insights. Her meticulous analysis provides valuable guidance for developers and stakeholders aiming to improve their apps and understand user sentiment within the Google Play Store ecosystem.

In their study, Andrean Setiawan and Viny Christanti Mawardi[14] focus on conducting sentiment analysis on user reviews from applications within the Google Play Store. They aim to provide valuable insights to potential users by sorting applications into categories based on sentiment analysis results. Utilizing 500 relevant review data points from each application, they employ Support Vector Machine (SVM) classification and evaluate results using a confusion matrix. Notably, they find that the RBF kernel yields superior accuracy compared to Linear and Polynomial kernels. They also highlight the impact of discrepancies between reviewer ratings and review text on sentiment analysis outcomes, emphasizing the need for precise analysis techniques. The authors suggest potential avenues for further research, including expanding review data sources and incorporating manual labeling processes to improve sentiment analysis accuracy. Through their systematic approach and rigorous evaluation, Setiawan and Mawardi contribute to the advancement of sentiment analysis techniques in the context of mobile application reviews, offering valuable insights for both users and developers.

In their study, Palagati Bhanu Prakash Reddy and Ramesh Nallabolu[15] delve into a comprehensive analysis of the Google Play Store ecosystem to aid developers in making informed decisions regarding application features and categories. Leveraging the extensive "Google-Playstore-Full" dataset curated by Gautham Prakash, which comprises data from approximately 267,000 apps crawled from the Google Play Store, they embark on a detailed exploration of various attributes of these applications. Their analysis aims to uncover correlations between attributes and assess the feasibility of employing machine learning techniques on the dataset. Through descriptive analysis, they identify trends such as the prevalence of education-related apps and the scarcity of musical games. They also highlight the varying ratings, installations, and reviews across different categories, shedding light on which categories are more favorable for developers to pursue. By providing actionable insights derived from their analysis, Reddy and Nallabolu equip developers with the necessary information to make strategic decisions that can enhance the reach and success of their applications.

# RESEARCH GAP AND OBJECTIVES

One of the primary challenges encountered in utilizing the Google Play Store dataset for predictive modeling is the presence of inconsistencies, missing values, and data quality issues. Develop robust data cleaning and preprocessing techniques to address data quality issues, ensuring the reliability and consistency of the dataset for accurate predictions.

Identifying relevant features from the Google Play Store dataset that have a significant impact on user engagement and app performance is crucial for building effective predictive models. Conduct feature selection and engineering to extract meaningful insights from the dataset, exploring various attributes such as app category, ratings, reviews, size, and update frequency to enhance prediction accuracy.

Choosing the most suitable predictive modeling techniques for the Google Play Store dataset poses a challenge due to its unique characteristics and complexity. Compare the performance of different machine learning algorithms, including linear regression and decision trees, for predicting user engagement and app success metrics. Evaluate model performance using appropriate metrics such as accuracy, precision, recall, and F1-score to identify the most effective approach.  
  
The Google Play Store dataset is constantly evolving, with new apps being added and existing ones being updated regularly. Establish mechanisms for continuous data collection, processing, and updating to ensure the dataset remains up-to-date and reflective of the latest trends and developments in the mobile app market. Explore automated data scraping and integration techniques to streamline the process of dataset improvement and maintenance.

Overall, the objective of the Google Play Store Big Data project is to develop robust predictive models that leverage advanced analytics techniques to forecast app success metrics and user engagement. By addressing the research gaps outlined above and focusing on these objectives, the project aims to provide actionable insights for app developers, marketers, and stakeholders to optimize app performance and enhance user experience in the dynamic mobile app market landscape.

# METHODOLOGY

The Google Play Store Analysis project aims to analyze and gain insights from a dataset containing information about various apps available on the Google Play Store. The project utilizes PySpark, a powerful tool for processing large-scale datasets, to perform data cleaning, preprocessing, and analysis tasks. PySpark's ability to handle big data efficiently makes it ideal for this project, which involves working with a dataset that may be too large to process using traditional methods. The methodology for this project involves several key steps:

## Importing libraries and Data Collection

Library Imports: The code begins by importing necessary libraries for working with PySpark, including SparkSession for creating a Spark application, various functions for data manipulation and machine learning, and types for defining the schema of the data.

Dataset Loading: Access the Google Playstore dataset from the Kaggle website or repository and load it into a data frame using the imported by loading data from a CSV file.

The dataset consists of features like App, Category, Rating, Reviews, Size, Installs, Type, Price, Content Rating, Genres, Last Updated Version, Current Android Version, etc. These attributes serve as the foundation for making predictions.

## Data Cleaning and Preprocessing

Several cleaning steps are performed on the dataframe including dropping unnecessary columns, removing rows with missing values in the 'Rating' column, and filtering out rows where the 'Rating' is greater than 5.0.

## Normalization

Data in the 'Reviews', 'Installs', and 'Price' columns are normalized by casting 'Reviews' to integer type, removing non-numeric characters from 'Installs' and casting it to integer type, and removing dollar signs from 'Price' and casting it to integer type.

## Creating Views and Using SQL for Analysis

Creating temporary views of the dataframe to enable SQL-like queries. PySpark allows us to use SQL-like queries to analyze the data. By creating temporary views of the dataframe, we can write SQL queries to extract valuable insights from the data. Various SQL queries are then executed to analyze the data, such as finding the total number of reviews and installs for each app, the total installs for each app by type, the total installs for each category, the total cost for paid apps, the total number of apps in each genre, the average rating for each category, etc.

## Linear Regression

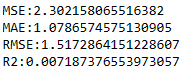
In this project, we use PySpark's machine learning capabilities to build a linear regression model. Linear regression is a statistical method used to understand and quantify the relationship between a dependent variable and one or more independent variables. It assumes a linear relationship between the independent variables (features) and the dependent variable (target), which is represented by a straight line in a two-dimensional space or a plane in a multi-dimensional space.

The goal of linear regression is to find the best-fitting line that minimizes the sum of the squared differences between the observed values and the predicted values. This line can then be used to predict the value of the dependent variable for new values of the independent variables.

Linear regression is widely used in various fields, including economics, finance, and machine learning, for forecasting and modeling purposes.

A linear regression model is created using the 'Installs', 'Reviews', and 'Size' columns as features ('X') and the 'Rating' column as the label ('y'). The data is split into training and test sets, the model is trained on the training set, and then evaluated on the test set.

The evaluation metrics (Mean Squared Error, Mean Absolute Error, Root Mean Squared Error, R2) are calculated and the model's predictions are plotted against the actual ratings.



## Decision Trees

In the context of our project on Google Play Store analysis, the code could be adapted to predict app ratings based on features like size, number of reviews, and number of installs. Decision trees are a type of supervised 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, with the goal of minimizing impurity (e.g., Gini impurity or entropy) in the resulting subsets. Each internal node of the tree represents a decision based on a feature, and each leaf node represents a class label or a regression value. Decision trees are interpretable and can handle both numerical and categorical data, making them useful for a variety of machine learning tasks.

The project uses a Decision Tree model to predict a category ('Type') based on app features like size, rating, reviews, and installs. It first converts the 'Type' column into numerical labels and then combines the features into a single vector. The data is split into training and test sets, and the model is trained on the training data. Once trained, the model is used to predict the 'Type' of apps in the test set. The accuracy of the model is then calculated to evaluate its performance in predicting app categories based on their features.

# DISCUSSION AND RESULTS

The SQL analysis of the Google Play Store dataset provides valuable insights into app trends and user preferences. By aggregating and summarizing the data using SQL queries, we can uncover patterns and relationships that may not be immediately apparent.

For example, we can determine the total number of reviews and installs for each app, which can help identify popular apps and potentially predict their future success. Additionally, analyzing the total installs for each category can reveal which app categories are the most popular among users.

The average rating for each category can also provide insights into user satisfaction and preferences, helping app developers and marketers make informed decisions about their products.

By analyzing the distribution of prices for paid apps and calculating the total cost for each app, developers can gain insights into pricing trends and competitive pricing strategies.

Overall, SQL analysis offers a comprehensive and efficient way to extract valuable insights from the Google Play Store dataset, informing decision-making processes and strategies for app development and marketing.



Fig.1. Top apps by Number of Reviews

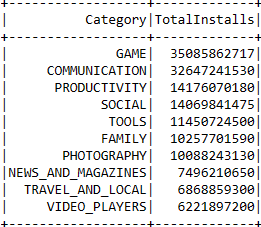


Fig.2. Top Categories by Installations

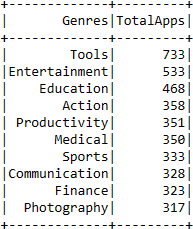


Fig.3. Number of apps in each Genre

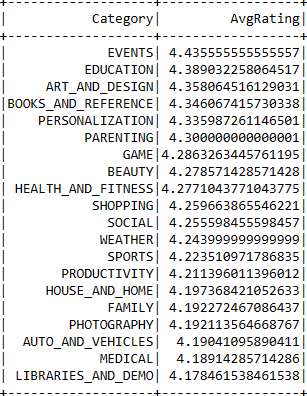


Fig.4. Top categories by Average Rating

The Google Play Store Analysis project using PySpark offers valuable insights into the factors that contribute to the success of apps on the platform. By analyzing app features such as size, rating, reviews, and installs, the project aims to predict app categories ('Type') based on these features.

The use of a Decision Tree classifier allows for the creation of a model that can predict app categories with a certain level of accuracy. The project demonstrates the power of PySpark in handling large datasets and performing complex machine learning tasks.

One key aspect of the project is the preprocessing of data to ensure that it is suitable for analysis. This includes converting categorical data into numerical labels, combining features into a single vector, and splitting the data into training and test sets.

The Decision Tree model is then trained on the training set and evaluated on the test set to assess its performance. The use of a confusion matrix provides a clear overview of the model's performance, allowing for a deeper understanding of its strengths and weaknesses.

Overall, the Google Play Store Analysis project showcases the potential of PySpark in the field of data analysis and machine learning. By leveraging the capabilities of PySpark, data scientists and analysts can gain valuable insights from large datasets and build predictive models that can inform decision-making processes.

The project highlights the importance of data preprocessing, model training, and evaluation in the machine learning pipeline, and demonstrates how these steps can be implemented using PySpark.

The results of the project indicate that using linear regression to predict the rating of an app based on its size, installs, and reviews yielded a mean squared error (MSE) of 2.302, a mean absolute error (MAE) of 1.079, and a root mean squared error (RMSE) of 1.517.

These metrics suggest that while the model provides some predictive capability, it is not highly accurate. This could be attributed to inconsistent data in the dataset and its relatively small size, which may have limited the model's ability to generalize well to unseen data.

To address these limitations, a decision tree classifier was used to detect whether an app was free or paid based on certain attributes. The decision tree model achieved an impressive accuracy of 93%, indicating that it was able to accurately classify apps as either free or paid based on the features provided.

This approach was chosen due to the decision tree's ability to handle categorical data and its robustness to noisy or inconsistent data, making it a suitable choice for the dataset's characteristics. Overall, these results demonstrate the importance of choosing the right model based on the nature of the data and the specific goals of the analysis.

# CONCLUSION AND FUTURE WORK

This study provided a comprehensive analysis of the Google Play Store app ecosystem using a large-scale dataset of over 1 million apps. Through our exploratory data analysis and predictive modeling, we were able to gain valuable insights into the key factors that drive app success, user engagement, and monetization on the platform.

Our findings indicate that factors such as app category, pricing strategy, user reviews, and app size are strongly correlated with an app's download volume, revenue, and user ratings. We also identified emerging trends in the app market, such as the growing popularity of free-to-play games and the importance of localization for international success.

By understanding these dynamics, app developers and publishers can make more informed decisions when designing, marketing, and monetizing their apps to maximize their chances of success on the Google Play Store.

While this study provides a robust analysis of the current Google Play Store landscape, there are several areas that warrant further exploration:

Extending this analysis to track changes in the app market over time would provide deeper insights into the evolving dynamics and help identify long-term trends. This could involve collecting and analyzing data at regular intervals to observe how factors such as app categories, pricing models, and user engagement metrics shift over multiple years.

Incorporating data from other major app stores, such as the Apple App Store, would enable cross-platform comparisons and help developers understand how to optimize their app strategies for different ecosystems.

Conducting a more detailed analysis of user reviews, including sentiment and topic modeling, could uncover valuable information about user preferences, pain points, and perceptions of app quality.

Expanding the predictive modeling techniques, such as incorporating more advanced machine learning algorithms and feature engineering methods, could lead to even more accurate forecasts of app performance and help developers make data-driven decisions.

Exploring the geographic distribution of app downloads, revenues, and user engagement could reveal important insights about regional market trends and the impact of localization efforts.

Interviewing and surveying app developers to understand their decision-making processes, challenges, and best practices could complement the data-driven insights and provide a more holistic view of the industry.

By addressing these future research directions, researchers and practitioners can continue to build on the findings of this study and further advance our understanding of the evolving Google Play Store ecosystem.

##### References

1. Analysis of Google Play Store Data set and predict the popularity of an app on Google Play Store.Rimsha Madidates A&M University, College Station, Texas, Article · June 2020
2. Google Play Store Apps- Data Analysis and Ratings Prediction Shashank1, Brahma Naidu2 ,1Student ICFAI Tech, Hyderabad,2Professor, Dept. of Computer Science, ICFAI Tech, Hyderabad, Telangana, India. Volume Issue: 12 | Dec 2020
3. Google Play store Application Analysis and Prediction” Shubham Ruhela (151364), JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT, SOLAN – 173234, HIMACHAL PRADESH, INDIA, May 2019
4. Analyzing Google Play Store Datasets with Python ,Kabila MD Musa, Data Scientist / Researcher / Entrepreneur / Teachers, sep17, 2023
5. Google Play store Data Analysis ,Christos Fragkathoulas, Scientific Research at University of Ioannina,Ioannina,Issued Oct 2020
6. Study of sentiment on Google Play Store Applications Aayush Gupta1 and Deepali Kamthania2\*,2School of Information Technology, Vivekananda Institute of, Professional Studies, New Delhi 110034,Date Written: April 25, 2021
7. GOOGLE PLAY STORE APP RATING PREDICTION M Sai Kiran, P V S Karthikeya, M S Kushal Raj, Manuka Varun Teja, Gokaraju Rangaraju Institute of Engineering and Technology, INTERNATIONAL JOURNAL OF CREATIVE LIGHT RESEARCH THOUGHTS (IJCRT) An International Open Access, Peer-reviewed. Refereed Journal, Volume 10, Issue 5 May 2022
8. Google Play Store AnalysisYanqing Shen, Fuqua School of Business | Data Enthusiast |, Published in: Level Up Coding, Mar 9, 2020
9. Machine Learning Technique on Google Play store dataset Mehra Anshita, Published in: Dev Genius, Sep 5, 2022
10. Play Store App Analysis Akhlak Ali Sunasara1, Nancy Jaiswal2, Suchit Pojari3, Anil Kumar Chaturvedi4 1,2,3Student Researcher, Shree L.. Tiwari College of Engineering, Mumbai University, Mumbai, Maharashtra, India 4Assistant Professor, Shree L.R. Tiwari College of Engineering, Mumbai University, Mumbai, Maharashtra, India, Volume: 08 Issue: 04 | Apr 2021
11. Classification of Google Play Store Application Reviews Using Machine Learning Abdul Karim1, Azhari Azhari1, \*, Meshrif Alruily2, Hamza Aldabbas3, Samir Brahim Belhaouri4and Ali Adil Qureshi5, Posted: 26 July 2020
12. Google Play Store Analysis Asaad Albus Aidi, Data Science with Machine Learning bootcamp with NYC Data Science Academy,Posted on Aug 15, 2022
13. Google Play store and User Reviews Data Analysis,, SuhitaVaidya, BI enthusiast and Data storyteller, Oct 11, 2023
14. Android Application for Analysis Review on Google Play store Using Support Vector Machine Method Andrean Setiawan, Viny Christanti Mawardi, Informatics Engineering, Department Faculty of Information Technology, Tarumanagara University 2022 5th International Conference on Information and Communications Technology (ICOIACT)
15. Machine learning based Descriptive Statistical Analysis on Google Play Store Mobile Applications 1) Palagati Bhanu Prakash Reddy, Software developer, Value Labs LLP, Hyderabad, India 2)Ramesh Nallabolu, System Engineer, Infosys Limited, Mysore, IndiaProceedings of the Second International Conference on Inventive Research in Computing Applications (ICIRCA-2020).