Exploring Google Play Store Apps Using Predictive User Context: A Comprehensive Analysis

Anandh A[1], Ramya R[2], Vakaimalar E[3]*, Santhipriya B[4]
Dept. of Computer Science and Engineering, Kamaraj College of Engineering and Technology,
Virudhunagar, Tamilnadu, India[1, 2, 4]
Dept. of Information Technology, Kamaraj College of Engineering and Technology, Virudhunagar, Tamilnadu, India[3]

***Abstract*—Google Play Store is a digital platform for mobile applications, where users can download and install apps for their android devices. It is a great source of data for mining and analyzing app performance and user behavior. The increasing volume of mobile applications poses a challenge for users in finding apps that align with their preferences. This work aims to utilize predictive user context to analyze user behavior, thereby enhancing user experience and app development. The work focuses on identifying trends in the app market to recommend suitable applications for users. Play Store app analysis involves gathering data, performing comprehensive evaluations, and making informed decisions to improve app performance and user engagement. By applying Naïve Bayes, Random Forest, and Logistic Regression algorithms, this work evaluates the relationship between application attributes such as categories and the number of downloads, determining the most effective profiling algorithm for app performance evaluation. This analysis is crucial for recognizing user engagement trends, discovering new opportunities, and optimizing existing applications.**

Keywords—Naïve Bayes; random forest; logistic regression; mining; Google play store; android; mobile application

I. Introduction

The growth of smartphones and tablets has contributed to the development of various mobile applications called apps. An application is a standalone piece of software with specific goals, rules, and functions. Apps are provided as proprietary software repositories (often called App stores), with the largest vendors being Google Play Store, iPhone App store and Blackberry App World. App stores typically hold three types of data: app developer data, app user feedback (ratings, reviews, and tags), and statistical and organizational data (app categories and download counts).

The availability of these rich data in a software repository provides a unique opportunity to analyze and understand the relationships between data. Interoperability data analysis provides business development applications with insight into the added value of features that can be considered when developing new products.

Due to its increasing popularity and rapid growth in recent times, Google Play is the largest publisher of Android apps. One of the reasons for this popularity is that 96.96% of the products on Google Play Store are free [1]. Google Play Store, which is pre-installed on certified Android devices, serves as the official app store of Google. By granting access to a wide range of content, such as apps, books, magazines, music, movies, and TV shows, Google Play Store offers a diverse selection to users [2]. It enables users to browse, download, and install apps created using the Android Software Development Kit (SDK) and distributed by Google.

The number of mobile apps has grown exponentially due to the growth of smartphones and the app industry. Users can install many applications that provide useful services for many aspects of daily life such as chat, music, video, web browsing and more. The number of applications installed on a mobile phone range from 10 to 90, and on an average, 50 applications are installed.

Although users can install many applications on their smartphones to make the device to work, searching and choosing the right applications can be difficult. Users have the difficulty in finding the right app and have less exposure to most of the apps. The Google Play store offers a wide selection of data on features and descriptions related to application functionality.

Unfortunately, many of these apps remain unused, resulting in fewer installations and loss of business. Additionally, users often spend a lot of time searching through apps, making it difficult to find the right one. To resolve this issue, Data Mining technique can be used to pre-process the real time Google play store dataset in order to predict the user's mobile application usage behavior. We use Logistic Regression, Naive Bayes and Random Forest to measure the relationship between pairs across all clusters by applying and analyzing the performance of each algorithm.

Examples of such examined pairs include price and rating, price and number of downloads, and rating and number of downloads. This approach of analysis can be used to reveal intrinsic properties of software repositories. When applied to Google Play, it can provide a general picture of the current market situation. This helps the developers to understand customer demand and preferences by using the best analyzer algorithm. Image analysis can be performed by leveraging different features extracted through machine learning algorithms [2] [3] [4] or by utilizing a deep learning framework [5]. In our proposed work, machine learning algorithms are employed to perform the intended task.

*Corresponding Author.

II. Related Work

In the data preprocessing phase, data cleaning and imputation processes were performed by Sivakani et al [6]. Mean imputation was utilized to handle missing values. Subsequently, classification and clustering tasks were conducted on the coronavirus dataset, and their validity was assessed through a 10-fold validation process. The classifiers employed in this study were Naïve Bayes and Random Forest. Nedeva proposed an integrated marketing information system that aimed to enhance the effectiveness and efficiency of marketing activities by integrating various components of the information system [7]. It presented the research findings related to marketing information systems, including data collection and analysis. An integrative analysis of marketing studies concentrating on mobile apps is presented by Lara Stocchi et al. [8]. The review's objective is to increase knowledge of how applications affect customer experiences and add value all along the customer journey. Google playstore with sample apps is shown in Fig. 1.

Fig. 1. Google play store with sample apps.

Sutriawan et al. conducted a performance comparison of multiple classification algorithms, including Naive Bayes, K-Nearest Neighbor (kNN), Support Vector Machine (SVM) and Decision Tree (DT), to categorize the polarity attitudes in Indonesian film reviews as positive and negative [9].

Israel J. Mojica Ruiz et al. says that how ratings impact a client's choice to buy a product [10]. Recent exploration shows that the ratings relate explosively with download counts, a crucial measure of a mobile app's success. App store reviews don't take streamlined performances into account; the majority of app store ratings overlook updated versions and rely on a static rating system to distinguish apps with varying levels of user satisfaction. When app stores exclusively showcase current ratings, app developers may find limited advantages in releasing enhanced versions of low-rated apps. The presence of numerous negative ratings for a low-quality initial version could pose challenges in achieving an improved store rating of 4 or more stars [11].

Hong Cao et al. [12] provides an overview of the existing studies in the field of mining smartphone data for understanding app usage patterns. It contributes to the understanding of user behavior and opens up possibilities for improving app usage prediction and recommendations. Martin et al. highlights the significance of App store analysis in studying applications downloaded from app stores [13]. It emphasizes that app stores give precious information that wasn't available with former software distribution styles. Keng-Pei Lin et al. proposed a substantiated mobile app recommender system grounded on the textual data of user reviews available on the App store [14]. Topic modeling methods are applied to abstract concealed ideas of user reviews, and the probability distributions of the topics are employed to represent the features of the apps. Also, the user profile is constructed grounded on the user's installed apps to record their preferences. They showed that user reviews are effective for inferring the features of apps. Real- world data are employed to perform trials, and the experimental results showed that the reviews are effective for substantiated app recommendations.

Shahab et al. suggested conducting a case study centered on analyzing the Google Play Store. This analysis aims to offer a detailed understanding of the fundamental characteristics of these app repositories [15]. Finkelstein et al. formulates App store analysis as a method of mining software repository [16]. The experimenters used data mining ways to extract features from 32,108 priced apps in the App store of Blackberry. They also considered price and popularity information to dissect specialized, business, and client acquainted aspects of the app store.

William Martin et al. performed app store Analysis studies about the applications that are downloaded from App store [17]. It revolved around gathering non-technical information from App stores and integrating it with technical data to unveil trends and behaviors within these software repositories. The insights derived from this analysis significantly influence software development teams, leading to advancements in requirements engineering, release planning, software design, security, and testing techniques. Embracing App store analysis opens up a thrilling avenue for software engineering research, fostering a profound understanding of the interconnections between social, technical, and business aspects in software development and deployment.

Ahlam et al. proposed a model that addresses the challenge of personalized application recommendations by combining content-based filtering and App profiles [18]. It leverages important features and usage data to suggest relevant apps to users based on their preferences and search queries. Mahmood concentrated on examining Google Play store, the largest Android App store, to gain awareness into the basic assets of app sources [19]. The idea is to give a comprehensive understanding of the current state of the app request and help inventors in understanding client solicitations, stations, and request trends.

Helan et al. used SVD to diminish the corpus dimension and prepare the data for mining [20]. The significance of feature selection in the fields of Data Mining and Human Machine Interaction is suggested by Iryna et al. [21]. It suggests a new approach that combines feature selection and feature extraction methods to evaluate the information quantity. This approach aims to reduce the number of features while maintaining their linguistic interpretation. The increasing number of mobile applications poses a challenge for users in finding apps that align with their preferences. So, we propose a system to address this challenge.

III. Proposed System

Online user reviews can provide insight into the features that users find most useful or least useful, as well as any bugs or issues that users have encountered. User reviews can also provide feedback on the overall user experience, such as how easy it is to use the app or how intuitive the interface is. This feedback can be invaluable for developers, as it can help them identify areas for improvement and make changes to the app that will better meet user needs. Fig. 2 illustrates the proposed system architecture. The system consists of two main components: a feature extraction module and a user preference mining module.

The feature extraction module extracts the features of an app from user reviews and represents them as a topic distribution. The topic distributions of installed apps are used by the user preference mining module to profile user preferences. A data pre-processing module is also part of the system, and it cleans and normalizes user reviews.

A. Data Acquisition and Preprocessing

Raw data gathered from the Google Play Store is quite prone to noise, have missing values and consistency issues. Results of data mining depends on the quality of input data. So, raw data is pre-processed in order to enhance the quality of the data and the mining results. Data pre-processing, one of the most important processes in data mining, deals with the initial dataset preparation and modification. Data cleaning, data integration, data transformation and data reduction are all the phases in the pre-processing of data.

Fig. 3 shows Google play Store Dataset. The Google Play Store dataset is a collection of data about the applications available on the Google Play Store. It contains details like the name of the application, its category, rating, the number of downloads, its size, and its cost. It also includes reviews from users, which can be used to gain insights into the quality of the applications. The dataset can be used to analyze the trends in the mobile application market, as well as to identify popular applications and their features. After completing the processes of data cleaning, integration, transformation, and reduction in Google play store data set, the resulting data becomes ready for utilization in the subsequent steps.

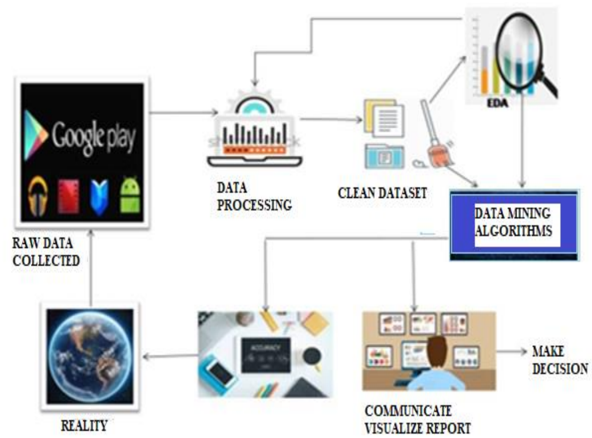

Fig. 2. Proposed system architecture.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content R	Genres	Last Upda	Current Ve	Android Ver	
2	Photo Editor & Candy Ca	ART_AND_	4.1	159	19M	10,000+	Free	0	Everyone	Art & Desi	January 7,	1.0.0	4.0.3 and up	
3	Coloring book moana	ART_AND_	3.9	967	14M	500,000+	Free	0	Everyone	Art & Desi	January 1	2.0.0	4.0.3 and up	
4	U Launcher Lite â€“ FREE	ART_AND_	4.7	87510	8.7M	5,000,000-	Free	0	Everyone	Art & Desi	August 1,	1.2.4	4.0.3 and up	
5	Sketch - Draw & Paint	ART_AND_	4.5	215644	25M	50,000,00	Free	0	Teen	Art & Desi	June 8, 20	Varies wit	4.2 and up	
6	Pixel Draw - Number Art	ART_AND_	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Desi	June 20, 2	1.1	4.4 and up	
7	Paper flowers instructior	ART_AND_	4.4	167	5.6M	50,000+	Free	0	Everyone	Art & Desi	March 26,	1	2.3 and up	
8	Smoke Effect Photo Mak	ART_AND_	3.8	178	19M	50,000+	Free	0	Everyone	Art & Desi	April 26, 2	1.1	4.0.3 and up	
9	Infinite Painter	ART_AND_	4.1	36815	29M	1,000,000-	Free	0	Everyone	Art & Desi	June 14, 2	6.1.61.1	4.2 and up	
10	Garden Coloring Book	ART_AND_	4.4	13791	33M	1,000,000-	Free	0	Everyone	Art & Desi	Septembe	2.9.2	3.0 and up	
11	Kids Paint Free - Drawing	ART_AND_	4.7	121	3.1M	10,000+	Free	0	Everyone	Art & Desi	July 3, 201	2.8	4.0.3 and up	
12	Text on Photo - Fonteee	ART_AND_	4.4	13880	28M	1,000,000-	Free	0	Everyone	Art & Desi	October 2	1.0.4	4.1 and up	
13	Name Art Photo Editor -	ART_AND_	4.4	8788	12M	1,000,000-	Free	0	Everyone	Art & Desi	July 31, 20	1.0.15	4.0 and up	
14	Tattoo Name On My Pho	ART_AND_	4.2	44829	20M	10,000,00	Free	0	Teen	Art & Desi	April 2, 20	3.8	4.1 and up	
15	Mandala Coloring Book	ART_AND_	4.6	4326	21M	100,000+	Free	0	Everyone	Art & Desi	June 26, 2	1.0.4	4.4 and up	

Fig. 3. Google play store dataset.

B. Algorithm Modeling

A model is created from data using data mining algorithms which analyses the data to identify patterns and relationships between variables. The model is then used to predict future data forecasts or judgments. Depending on the type of data and the desired result, data mining methods can be either supervised or unsupervised.

After splitting the final dataset into a training and test set, the feature variable is scaled.

1) Naive bayes: After the pre-processing step, it's time to train the Naive Bayes model using the Training set. A classification method based on the Bayes theorem called Naive Bayes makes the assumption that predictors are independent [22]. The Naive Bayes classifier assumes that the presence of one feature in a class is independent of the presence of any other features. The conditional probability can be obtained using Bayes theorem as illustrated in (1). Finally, the test set is used to predict the performance of the model that is built. It is used to anticipate the Apps that will be most popular.

$$P(A\backslash B) = \frac{P(B\backslash A)\,P(A)}{P(B)} \quad (1)$$

2) Random forest: Random Forest is based on the concept of decision trees. The created ensemble of DT is typically trained using the "bagging" method [23]. Random Forest constructs numerous DT and merges them to get a prediction that is more precise and consistent. With the help of the ensemble of created DT, predictions of the most popular Apps are done.

3) Logistic regression: Logistic Regression operates on the principle of mapping input features to the probability of a binary outcome. The algorithm applies a logistic function to linearly combine feature weights and inputs, constraining the output within [0, 1]. By optimizing the model's coefficients through maximum likelihood estimation, logistic regression effectively learns to classify instances into one of the two classes.

Multi-class logistic regression extends the binary version to handle classification tasks with more than two classes. The algorithm employs the softmax function to calculate the probability of each class for a given instance, ensuring the sum of probabilities equals 1. By iteratively optimizing the model's parameters through gradient descent, multi-class logistic regression effectively learns to distinguish and classify instances across multiple classes.

C. User Reviews Dataset

Fig. 4 shows the Google Play Store user reviews dataset. User reviews for various Apps available on the Google Play Store are included in the dataset. It contains the name of the application, the user's rating, the posting date, and the review's content. To analyze user reviews of Apps on the Google Play Store, this dataset is used. Additionally, it can be utilized to spot patterns in user reviews and locate well-liked applications.

Using Google Play Store reviews dataset, the basic Natural Language Processing (NLP) steps have been applied to pre-process the data. Punctuation, special characters, stop words, are all removed from the data during pre-processing and lemmatization is also done in order to bring together a word's inflected forms for analysis as a single item [24].

And finally, a bag of words is created and used as the model. Data are separated into test and train data. These are used as input for various algorithms to find the accuracy.

1	App	Translated_Review	Sentiment	Sentiment_Polarity	Sentiment_Subjectivity
2	10 Best Foods for You	I like eat delicious food. Tl	Positive	1	0.533333333
3	10 Best Foods for You	This help eating healthy e›	Positive	0.25	0.288461538
4	10 Best Foods for You	nan	nan	nan	nan
5	10 Best Foods for You	Works great especially goi	Positive	0.4	0.875
6	10 Best Foods for You	Best idea us	Positive	1	0.3
7	10 Best Foods for You	Best way	Positive	1	0.3
8	10 Best Foods for You	Amazing	Positive	0.6	0.9
9	10 Best Foods for You	nan	nan	nan	nan
10	10 Best Foods for You	Looking forward app,	Neutral	0	0
11	10 Best Foods for You	It helpful site ! It help fooc	Neutral	0	0
12	10 Best Foods for You	good you.	Positive	0.7	0.6
13	10 Best Foods for You	Useful information The an	Positive	0.2	0.1
14	10 Best Foods for You	Thank you! Great app!! Ac	Positive	0.75	0.875
15	10 Best Foods for You	Greatest ever Completely	Positive	0.9921875	0.866666667

Fig. 4. Google play store user reviews dataset.

IV. Experimental Results

Using the APP review list data set given in Section III, the performance of the suggested technique is assessed here.

A. Basic Exploratory Data Analysis (EDA)

Exploratory data analysis is used to identify trends, patterns and relationships among data. It is used to summarize the data and to gain insights about the data. It is used to explore the data and gain a better understanding of the data. The fundamental goal of exploratory data analysis is to find any patterns or connections between the elements that stand out.

The association between the downloaded App% and the various review characteristics-based App types is shown in Fig. 5. The most popular apps on the market are those for communication. The majority of Apps perform well overall, with an average rating of 4.17. There are a few useless apps as well. The most expensive Apps are those for health and family.

There is a 0.63 moderately favourable association between the number of reviews and downloads. As a result, customers are more likely to download a certain App if more people have rated it. This implies that many people who download an app and are engaged with it typically also write a review or other form of feedback.

To analyze which category of apps has the highest percentage of installs, this work looks at the data from the Google Play Store. Fig. 6 provides a summary of the highest percentage of installed apps in the Google Play Store organized by category.

The category of Games App has the highest rank compared to other categories. This data provides us with the number of installs for each category of Apps, allowing us to compare and determine which category has the highest percentage of installs. This will give us an indication of which type of Apps are most likely chosen by users [25].

B. App Analysis

The analysis determined the average rating change for reviews that received responses, as well as the likelihood and size of the change. After categorizing the reviews, we looked at which subjects were most likely to result in a change for the better usage [26]. The aim for broad, all-encompassing issues that would interest most developers led to this choice of topics. The categories of apps with the highest number of installations by examining the top-ranked apps is analyzed. This is clearly visualized in the following Fig. 7. This clearly indicated that the top most installed Apps are mostly browser related Apps.

Most feedback was provided after releases that positive feedback was often associated with highly downloaded apps, and that negative feedback was often associated with less downloaded Apps and often did not contain user experience or contextual information. Sentiments towards App features show the differences between user sentiments in Google Play Store. Opinions on product quality formed a large portion of reviews, but opinions on service quality had a bigger effect on sales [14].

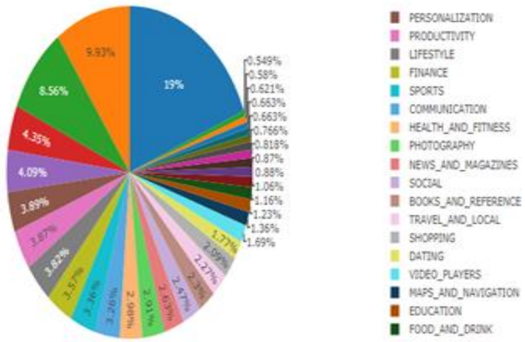

Fig. 5. Downloaded app percentage vs. various types of app in Google play store.

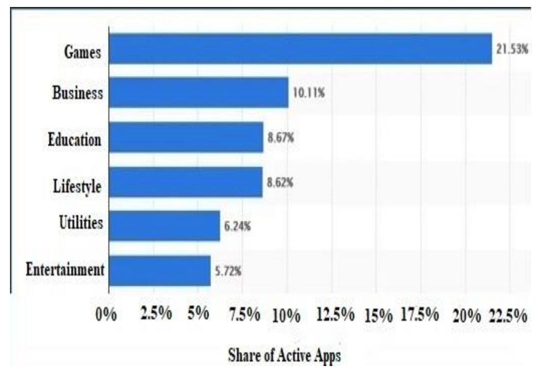

Fig. 6. Number of apps installed based on category.

C. Sentiment Analysis

As mentioned in Section III, the user reviews are pre-processed by performing the lemmatisation and removal of stopwords. The sentiment is derived from reviews using "positive" sentiment words like "good, great, love" or "negative" sentiment words like "bad, hate, terrible". Sentiment represents a user's views or opinions, often as positive or negative in this content. The most common complaints from users are about privacy invasion and unethical behaviour, with hidden costs coming in at number two [27].

	App
347	imo free video calls and chat
378	UC Browser - Fast Download Private & Secure
395	GO SMS Pro - Messenger, Free Themes, Emoji
366	UC Browser Mini -Tiny Fast Private & Secure
346	imo beta free calls and text
345	Yahoo Mail – Stay Organized
413	CM Browser - Ad Blocker , Fast Download , Privacy
9630	Orfox: Tor Browser for Android
4739	WhatsCall Free Global Phone Call App & Cheap C...
4609	Portable Wi-Fi hotspot

Fig. 7. Top 10 apps.

D. Best Analyzer Algorithm

Based on trained data, test data produce accuracy depending on the algorithm characteristic. Naïve Bayes, Random Forest and Logistic Regression algorithms are used for sentiment analysis.

In Fig. 8, the confusion matrix displays the predictions made by logistic regression model. The y-coordinate represents the true labels or ground truth values, y_true and x coordinates are predicted values, y_pred. According to score table, Logistic Regression gives us best accuracy of 90%.

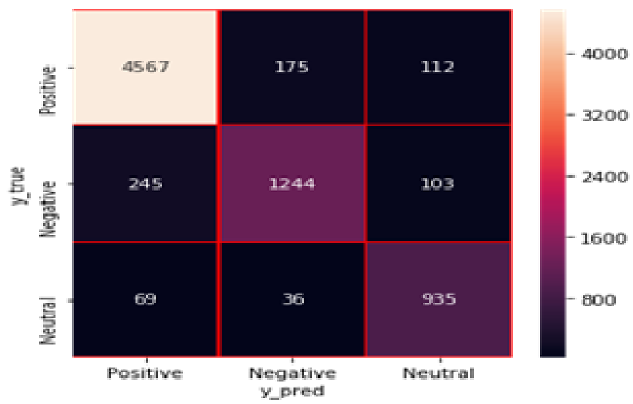

Fig. 8. Logistic regression.

V. Discussion

The analysis of Google Play Store data provided valuable insights into user behavior and App performance. The findings highlighted a moderate correlation between the number of reviews and downloads, suggesting that user reviews significantly influence App popularity. This correlation underscores the importance of positive user feedback in driving downloads and improving App visibility.

The work identified communication Apps as the most popular category, reflecting the high demand for applications that facilitate social interactions and connectivity. Additionally, browser-related Apps emerged as the top most installed category, indicating users' preference for efficient and accessible internet browsing solutions on mobile devices.

Machine learning algorithms such as Naïve Bayes, Random Forest, and Logistic Regression enabled a robust evaluation of the relationship between app attributes and performance metrics. Among these, Logistic Regression demonstrated the highest accuracy (90%) for user sentiment analysis, making it the most effective profiling algorithm in this context. This high accuracy indicates that Logistic Regression can reliably predict user sentiment based on App attributes, providing developers with actionable insights for app optimization.

VI. Conclusion

The research successfully utilized predictive user context to enhance user experience and App development on the Google Play Store. By analyzing user behavior and App performance through comprehensive data evaluations, the work provided critical insights for improving user engagement and app optimization. The identification of trends and patterns in App usage can guide developers in creating more user-centric applications, thereby increasing user satisfaction and App success.

The moderate association between reviews and downloads emphasizes the role of user feedback in App performance, while the dominance of communication and browser-related Apps highlights prevailing user preferences. The superior performance of Logistic Regression in sentiment analysis demonstrates its effectiveness as a profiling tool, offering a reliable method for predicting user sentiment and guiding App development decisions.

Overall, the integration of machine learning algorithms in App analysis provides a powerful framework for understanding and enhancing user engagement, ultimately leading to more successful and user-friendly applications. Analyzing and mining the data from Play Store Apps has a great deal of potential to help app development companies succeed. Developers can get useful knowledge to work on and conquer the Android market.

VII. Future Work

In order to annotate Google Play Store design elements with richer labels, new models could be developed, such as classifiers that explain the semantic function of elements and screens. To train newer varieties of perception-based predictive models, researchers may similarly crowdsource more perceptual annotations (for example, first impressions) over design elements like screenshots and animations. Additionally, a recommendation system that uses the discovered correlation features can be created to suggest applications.

If new apps are not continuously crawled and their database entries are not updated, static research datasets will eventually become out-of-date. Therefore, finding ways to make app mining more sustainable is a crucial area for future research. Making a platform where developers can utilize programs and add their traces to the repository for the benefit of the entire community is one possible route to sustainability.

References

[1] https://www.statista.com/statistics/266211/distribution-of-free-and-paid-android-apps/

[2] A.Anandh, K.Mala, and S.Suganya, “Content based image retrieval system based on semantic information using color, texture and shape features”, Proceedings of IEEE International Conference on Computing Technologies and Intelligent Data Engineering, pp. 1-8, Oct 2016, https://doi.org/ 10.1109/ICCTIDE.2016.7725364.

[3] R.Ramya, K.Mala, and C.Sindhuja, “Student engagement identification based on facial expression analysis using 3D video/image of students”, Taga J., vol. 14, pp. 2446–2454, 2018.

[4] A.Anandh, K.Mala, and R.Suresh Babu, “Combined global and local semantic feature–based image retrieval analysis with interactive feedback”, Measurement and Control, vol. 53(1–2), pp. 3–17, Dec 2019. https://doi.org/10.1177/0020294018824122.

[5] R.Ramya, K.Mala, and S.Selva Nidhyananthan, “3D facial expression recognition using multichannel deep learning framework”, Circuits Syst. Signal Process. vol. 39, pp. 789–804, May 2019. https://doi.org/10.1007/s00034-019-01144-8.

[6] Sivakani, and Syed Masood, “Analysis of COVID-19 and its Impact on Alzheimer’s Patient using Machine Learning Techniques”, International Journal of Computing, vol. 21, issue 4, pp. 468-474, Dec 2022, https://doi.org/10.47839/ijc.21.4.2782.

[7] Nedeva, V. I., ” Analysis of Marketing Information Systems and Conception of an Integrated Marketing Information System”, International Journal of Computing, vol. 3(2), pp. 127-133, Aug 2014. Doi:10.47839/ ijc.3.2.296.

[8] Lara Stocchi, Naser Pourazad, Nina Michaelidou, Arry Tanusondjaja and Paul Harrigan, “Marketing research on Mobile apps: past, present

and future", Journal of the Academy of Marketing Science, vol. 50, pp. 195-225, Nov 2021, https://doi.org/10.1007/s11747-021-00815-w.

[9] S.Sutriawan, P.N. Andono, M. Muljono, and R.A.Pramunendar, "Performance Evaluation of Classification Algorithm for Movie Review Sentiment Analysis", International Journal of Computing, vol. 22, issue 1, pp. 7-14, Mar 2023, https://doi.org/10.47839/ijc.22.1.2873.

[10] Israel J.Mojica Ruiz, Meiyappan Nagappan, Bram Adams, Thorsten Berger, Steffen Dienst, and Ahmed E. Hassan, "Examining the Rating System Used in Mobile-App Stores", IEEE Computer Society, vol. 33, pp. 86-92, Dec 2016, https://doi.org/10.1109/MS.2015.56.

[11] Gabriele Bavota and Mario Linares-Vasquez, "The Impact of API Change- and Fault-Proneness on the User Ratings of Android Apps", IEEE Transactions on Software engineering, vol. 41, pp. 384-407, Apr 2015.

[12] Hong Cao, and Miao Lin, "Mining Smartphone Data For App Usage Prediction and Recommendations: A Survey", Pervasive and Mobile Computing, vol. 37(4), pp. 1-22, Jan 2017, https://doi.org/10.1016/j.pmcj. 2017.01.007.

[13] William Martin, Federica Sarro, Yue Jia, Yuanyuan Zhang, and Mark Harman, "A survey of app store analysis for software engineering", IEEE Trans. Softw. Eng., vol. 43, no. 9, pp. 817–847, Sep 2017, doi: 10.1109/TSE.2016.2630689.

[14] Keng-Pei Lin ,Yi-Wei Chang, Chih-Ya Shen, and Mei-Chu Lin, "Leveraging Online Word of Mouth for Personalized App Recommendation", IEEE Transactions on Computational Social Systems, vol 5, pp. 1061-1070, Dec 2018, https://doi.org/10.1109/TCSS.2018.2878866.

[15] Shahab Mokarizadeh, Mohammad Tafiqur Rahman and Mihhail Matskin, "Mining and Analysis of Apps in Google Play," Proceedings of the 9th International Conference on Web Information Systems and Technologies, Jan 2013, pp. 527–535, 2013, doi: 10.5220/0004502005270535.

[16] Anthony Finkelstein, Mark Harman, Yue Jia, Federica Sarro, and Yuanyuan Zhang, "Mining App Stores: Extracting Technical, Business and Customer Rating Information for Analysis and Prediction", UCL Res. Notes, vol. 13, pp. 1-19, Nov 2013.

[17] William Martin, Federica Sarro, Yue Jia, Yuanyuan Zhang, and Mark Harman, "A Survey of App Store Analysis for Software Engineering" IEEE Transactions on Software Engineering, vol.43, pp. 817-847, Sep 2017.

[18] Ahlam Fuad Abdulghani, Sahar Bayoumi, and Hessah Al-Yahya "A Recommender System for Mobile Applications of Google Play Store", International Journal of Advanced Computer Science and Applications, vol. 11(9), pp. 42-50, Jan 2020, https://doi.org/10.14569/IJACSA. 2020.0110906.

[19] A. Mahmood, "Identifying the influence of various factor of apps on google play apps ratings," J. Data, Inf. Manag., vol. 2(1), pp. 15–23, Mar 2020, doi: 10.1007/s42488-019-00015-w.

[20] V.V. Helen Josephine, and S.Duraisamy, "Novel Pre-Processing Framework to Improve Classification Accuracy in Opinion Mining", International Journal of Computing, vol. 17, issue 4, pp. 234-242, Dec 2018, https://doi.org/10.47839/ijc.17.4.1145.

[21] Iryna Perova, and Yevgeniy Bodyanskiy, "Adaptive Human Machine Interaction Approach for Feature Selection-Extraction Task in Medical Data Mining", International Journal of Computing, vol. 17(2), pp. 113-119, June 2018, https://doi.org/10.47839/ijc.17.2.997.

[22] Jianlin Xu, Yifan Yu, Zhen Chen, Bin Cao, Wenyu Dong, Yu Guo, and Junwei Cao, "MobSafe: Cloud Computing Based Forensic Analysis for Massive Mobile Applications Using Data Mining", Tsinghua Science and Technology, vol. 18, pp. 418-427, August 2013.

[23] Weiwei Lin, Ziming Wu, Longxin Lin, Angzhan Wen, and Jin Li, "An Ensemble Random Forest Algorithm for Insurance Big Data Analysis", Mobile edge computing, vol. 5, pp. 16568-16575, Aug 2017.

[24] Chaoyi Pu, Zhiang Wu, Hui Chen, Kai Xu, and Jie Cao, "A Sequential Recommendation for Mobile Apps: What will User Click Next App", IEEE International Transactions on Web Services, pp. 243-248, Sep 2018.

[25] Ming Liu, Chong Wu, Xiang-Nan Zhao, Chin-Yew Lin, and Xiao-Long Wang, "APP Relationship Calculation: An Iterative Process", IEEE Transactions on knowledge and data engineering, vol. 27, pp. 2049-2063, Aug 2015, https://doi.org/10.1109/TKDE.2015.2405557.

[26] Sha Zhao, Gang Pan, Yifan Zhao, Jianrong Tao, Jinlai Chen, Shijian Li, and Zhaohui Wu, "Mining User Attributes Using Large-Scale APP Lists of Smartphone's", IEEE Systems Journal, vol. 11, pp.315-323, Mar 2017, https://doi.org/10.1109/JSYST.2015.2431323.

[27] Soo Ling Lim, Peter J. Bentley, Natalie Kanakam, Fuyuki Ishikawa, and Shinichi Honiden, "Investigating Country Differences in MobileApp User Behavior and Challenges for Software Engineering" IEEE Transactions on software Engineering, vol. 41, pp. 40-64, Jan 2015.