Acta Scientific COMPUTER SCIENCES

Volume 5 Issue 5 May 2023

Research Article

Application of Data Science in Analysis of Different Usage of Mobile Health Applications

Saheed Yusuf, Tareq Al Jaber* and Majid Bahmanzadeh

School of Computer Science, University of Hull, United Kingdom

***Corresponding Author:** Tareq Al Jaber, School of Computer Science, University of Hull, United Kingdom.

Received: March 14, 2023

Published: April 06, 2023

Abstract

The health sector is one of the most important sectors in any society which also greatly influences economic growth. The adoption of technology in the health sector has contributed a lot of improvements and thus help automate some of the processes, saving time and manpower. Technological advancement in the health sector has led to the evolution of Mobile Health (mHealth) applications to ease some of the medical processes like diagnosis, education, treatment, monitoring etc. This study aims to focus on the usability of the mHealth applications based on the datasets scraped on the popular mobile application stores which are the App store (iOS) and Play store (Android) and what influences the usage of the mHealth applications. The datasets for this study were gathered through web scraping. Data cleaning, feature engineering, data visualization, data analysis and modelling were carried out on these datasets using Python programming language and its libraries. The result of this study shows that mHealth applications that are free to download have higher performance and usability than paid applications. Likewise, applications that provide in-app purchases tend to have higher performance and usability than applications that do not provide in-app purchases. Also, predictive models were trained for predicting the performance of the mHealth application and the XGBoost classifier had the best performance based on accuracy and f1-score. To increase the usability of mHealth applications, it is recommended to promote in-app purchases in mHealth applications rather than asking users to pay to download without having a feel of the service(s) rendered by the applications.

Keywords: HMHealth Applications; Data Science; Performance Evaluation; Feature Engineering; Data Analysis

Introduction

The importance of good health care for individuals in society cannot be over-emphasized. It is a necessity that society is productive and healthy and not a matter of choice. For countries with a large population, the population can be of great advantage to the country as well as liability based on what the country chooses to prioritize [1]. The improvement of healthcare is important not because it is one of the significant government expenses but because it will influence the economic growth of the country as well as improve the standard of living [2]. Technology has contributed immensely to the improvement of healthcare delivery. With no iota of doubt, the introduction of technology to the health sector is unavoidable for the health of individuals. These improvements include and are not limited to health care coordination, patient education and diagnosis. With technological advancement in today's world, there are numerous cheap tiny/small mobile devices which outperform the expensive big computers of years ago in terms of performance and computational power [3]. mHealth is one of the technologies in the health sector that are rapidly growing which involves mobile device usage in discovering and monitoring biological changes in individuals [4]. mHealth technologies are now being adopted into the health sector, laying the groundwork for the effective transformation of medical research and services and how they are being practiced. Aside from the fact that mHealth applications enable individuals to find answers to the problems related to their health, they also help in getting access to health management, fitness and exercise, healthcare as well as other health-related services anywhere and anytime [5]. With the availability of devices with high computing power and advanced miniaturization, people are now increasingly able to personally track, monitor and continuously act on their health metrics in real-time [6]. This research aims to analyze mHealth application usage based on iOS and Android mobile application repositories (i.e. App store and Play store). The research questions for this research are:

- Do paid applications (on download) perform better than the free applications?
- Do applications with in-app purchases perform better than the applications without in-app purchases?
- Which classifier performs best in predicting mHealth application performance?

[7] came up with a new and validated questionnaire for application usability which is known as MAUQ (mHealth application usage Questionnaire) for evaluating mHealth applications. The questionnaire was designed by using a couple of already existing questionnaires that have been used in usability studies of mHealth applications, most importantly the well-validated questionnaires. They evaluated two different mHealth applications using MAUQ, along with the Post-Study System Usability Questionnaire (PSSUQ) and System Usability Scale (SUS) and 128 people responded to the questionnaires. The result shows that the new MAUQ possesses the validity and reliability necessary to evaluate mHealth application usability. [5] carried out a study on the factors that can influence users' acceptability and the usage of mHealth applications to improve the usage of the applications. They reviewed different literatures that are related to the mHealth application and the data source used was the Web of Science core database. The review of the influencing factors was performed from three different perspectives which include application design, society and individual. From the application design perspective, functionality, cost, security and ease of use influence the usage of mHealth applications. The main motivations for individuals to use mHealth applications are e-health literacy and health awareness [5].

Materials and Methods

Data collection

mHealth application usage data was gathered from both App store and Play store websites through web scraping [8-11]. Web scraping is a technique that is commonly used in extracting data from the internet which can be saved in a file system or database for further processing or analysis [12]. Both the App store and Play store have specific categories for mHealth applications on their websites and these categories are medical and fitness categories. The web scraping was achieved using Selenium and eautifulSoup4 python libraries. Selenium is an open-source program which runs on various operating systems such as macOS, Linux and Windows and it is capable of automating activities that can be on major web browsers like Google Chrome through its Web driver components [13]. BeautifulSoup4 is a parser library which is used in parsing XML and HTML contents [14]. Through web scraping, application usage information like average ratings, number of ratings, number of downloads, last updated date, first release date and price of application are extracted. The mHealth applications were alphabetically (A-Z) scraped on the App store based on categories (Medical and Fitness) and are saved as CSV files. Likewise, the mHealth applications scraped on the Play store were based on countries and categories (Medical and Fitness) and are saved as CSV files. The data is scraped and saved into different files because it takes hours to scrape all the data and any internet connection issue while scraping will mean to start scraping over again. In total, 93915 unique mHealth applications data were extracted from the App store and 1571 unique mHealth applications data were extracted from the Play store.

Data pre-processing

The scraped data for each of the mobile application repositories (App store and Play store) are saved into different files. For proper analysis, it is necessary to merge the different App store application data into a single file and also merge all Play store application data into a single file. Without good training data, the performance of the classifier will be poor regardless of the classifier inducer used [15]. For the data scraped on the Play store, the age rating column was cleaned by removing the noisy text in the column and replacing the null cells with None. The price column cells with null values were replaced with 0 and the currency symbol (£) was removed to ease visualization and modelling. The download count column was also cleaned by removing the whitespaces. In addition, the application rating count values that were in thousands using symbol K and millions using symbol M are converted to numbers by multiplying by 1000 and 1000000 respectively. For the data scraped on the App store, the rating count was cleaned by also replacing values that were in thousands using symbol K and millions using symbol M are converted to numbers by multiplying by 1000 and 1000000 respectively. The cells in the In-App purchase column with null values are replaced with False and the cells in the Price column with null values are replaced with 0. Also, the currency symbol ($) in price values was removed to make the price column numeric.

Feature engineering

Identifying and selecting the best features that would aid data modelling is important. With feature engineering, new features can be produced using transformation functions like arithmetic and aggregate functions on features [16]. Feature engineering is separately done on each of the datasets gathered for android and ios applications because some of the features do not exist in both datasets.

A new categorical feature was added to the dataset which is the download cost. This new feature indicates if the download is free or paid. Continuous features like rating counts and prices are grouped based on their ranges as categorical variables. Another categorical variable was created to determine the application performance based on the average rating range and rating count. The application performance is grouped into BAD (rating <2.50), FAIR (3.0 >= average rating < 3.5), GOOD (3.5 >= average rating < 4.0), VERY GOOD (4.0 >= average rating < 4.5) and EXCELLENT (average rating > 4.5).

However, some constraints were added for applications with high average ratings but very low rating counts. For instance, some applications are rated 5.0 with a rating count of less than 5. Applications like this do not necessarily perform well, hence there was a need to add some constraints. The following constraints were added:

- For applications with an average rating of 4.5 and above, if the rating count is less than 50, it is labelled as FAIR. If the count is greater than 49 and less than 100, it is labelled as GOOD. If the count is greater than 99 and less than 500, it is labelled as VERY GOOD. Otherwise, it is labelled as EXCELLENT.
- For applications with an average rating from 4.0 to 4.49, if the rating count is less than 50, it is labelled as FAIR. If the count is greater than 49 and less than 100, it is labelled as GOOD. Otherwise, it is labelled as VERY GOOD.
- For applications with an average rating from 3.0 to 3.99, if the rating count is less than 50, it is labelled as FAIR. Otherwise, it is labelled as GOOD. In addition, some new features including "is app version dependent", "is app updated this year", "is available in multiple countries", "is English supported", "is multiple languages supported", etc. are created for analysis and prediction purposes.

Data analysis and visualization

The mHealth applications on both the App store and play store are categorized under Medical and Health and Fitness. Approximately 57% of the mHealth applications scraped on the Play store are for medical purposes and 43% are health and fitness applications. In contrast, the majority of the mHealth applications on the App store are health and fitness applications. Approximately 65.54% of the mHealth applications are Health and Fitness apps while 34.46% are medical apps on the App store. This can be seen in figure 1. To visualize the data, this study uses the techniques and libraries from [17-19].

Approximately 53% of the mHealth applications scraped on the Play store have zero rating count. This means there is a high probability that no active user is using the applications that have zero rating count. Out of the 1,571 mHealth android applications, only 550 of the applications have up to a thousand rating counts. Some of the android mHealth applications have high number rating counts ranging from 10,000 to 5,000,000 ratings. A high number of rating counts is an indication that the application has a good number of active users. For instance, 11 of the android mHealth applications have between 1 million and 5 million ratings and one of the applications has over 5 million ratings.

In addition, approximately 53% (49886 out of 93915) of the iOS mHealth applications are not rated i.e. zero-rating count. This means there is a high probability that no active user is using the application since it was released. 1752 of the mHealth applications on the App store have rating counts of 1000 and above. Few of the applications have a relatively high number of ratings. For instance, 72 applications have ratings ranging from 100,000 to 500,000, 11 applications have ratings ranging from 500,000 to 1,000,000 ratings and 3 of the applications have ratings over 1 million ratings. Figure 2 shows the percentage rating count distribution.

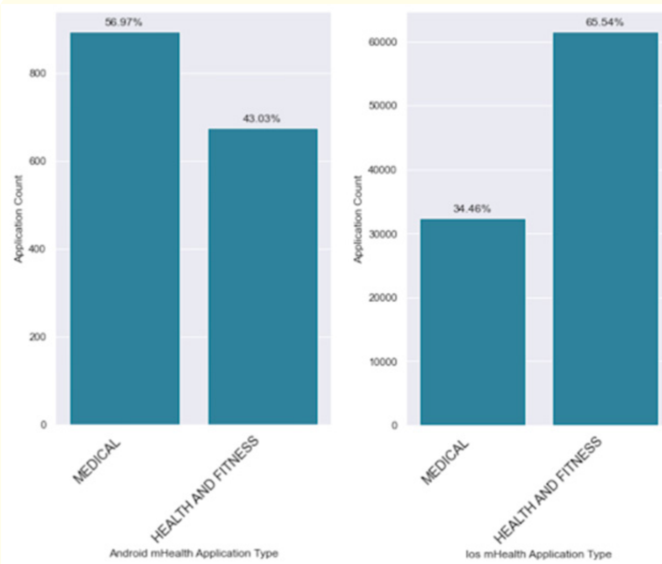

Figure 1: Application counts based on application type (a) Play store: the majority of the mHealth applications are for medical purposes (b) App store: the majority of the mHealth applications in the App store are for fitness purposes.

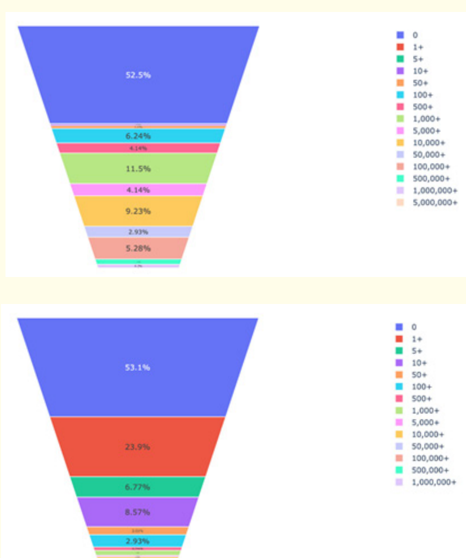

Figure 2: mHealth application count. The count is based on the number of ratings in each store (a) Play store (b) App store.

The majority of the mHealth applications that have high ratings tend to have a high average rating. The android mHealth application with the highest number of ratings has a rating count of 6,710,000 with an average rating of 4.9. Also, the iOS mHealth application with the highest number of ratings has a rating count of 1800000 with an average rating of 4.8. This can be seen in figure 3.

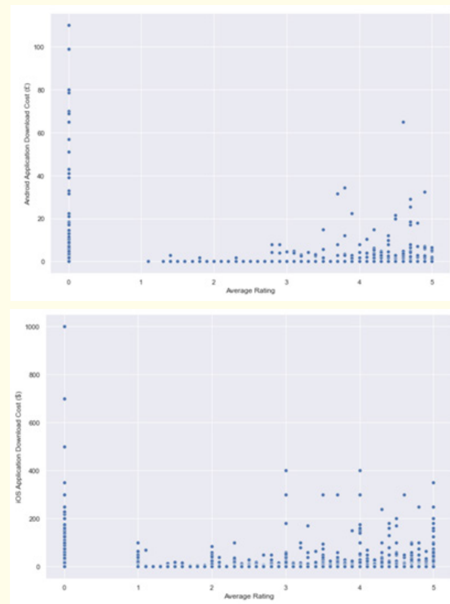

Figure 3: mHealth apps average rating vs rating count (a) Play store (b) App store.

Furthermore, the mHealth applications with higher download costs do not necessarily outperform others with low or no download costs. The most expensive (download cost of £109.99) mHealth application on the Play store is one of the applications with the worst performance i.e. average rating is less than 2.5. Likewise, the most expensive (download cost of $999.99) mHealth application on the App store is one of the applications with the worst performance. The highest download cost of mHealth applications on both the Play store and App store with a good rating (i.e. rating from 3.0 to 5.0) is £64.99 and $299.99 respectively.

There are 10% of the mHealth applications on the Play store that are free to download and are also rated by a minimum of 50,000 users while none of the android mHealth applications that require payment on download has up to 50,000 ratings. Also, there are approximately 0.18% of the mHealth applications on the App store that are free to download and also have a rating count of over 50,000 whereas none of the paid applications on the App store has up to 50,000 ratings. This analysis can be seen in figure 4.

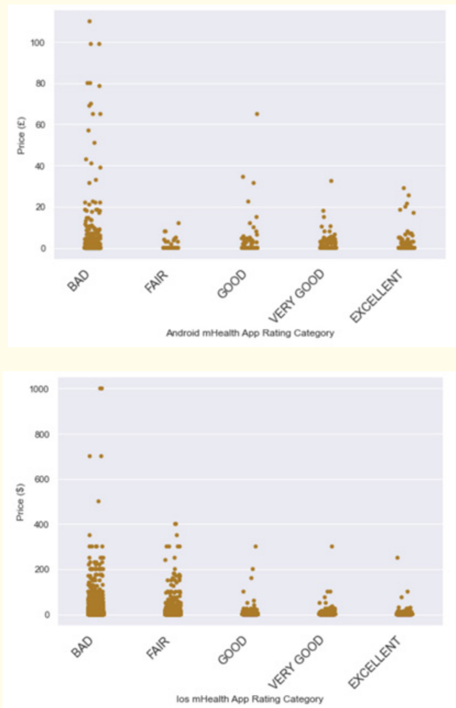

Figure 4: mHealth applications performance vs download price (a) Play store (b) App store.

On both the Play store and App store, the worst performing (based on average ratings) applications have the highest download cost.

Approximately 39% of the mHealth applications on the Play store offer in-app purchase i.e. application users can pay for services directly through the application. The majority of the applications on the App store do not also offer in-app purchases. Approximately 14% of the mHealth applications on the App store offer in-app purchases. Figure 5 shows the statistics of the mHealth applications performance based on support for in-app purchases.

The majority (approximately 93.60%) of the mHealth applications on the App store support English Language and approximately 32% of the applications on the App store support multiple languages. Approximately 98% of the mHealth applications on the

Figure 5: mHealth applications count based on performance and in-app purchase (a) Play store (b) App store.

App store with average ratings ranging from 3.0 to 5.0 are English-supported applications.

Feature selection

Feature selection is important before building a machine learning model. One of the important measures for the evaluation of the relationship between variables before building a machine-learning model is correlation. The features that do not correlate with the target variables should be removed from the dataset prior to model training [20]. Before generating a classification model, it is essential to check the correlation between the variables. Spearman correlation is a measure of a non-linear monotonic relationship between variables [21]. The categorical variables are converted to ordinal variables and spearman correlation was used to get the relationship among the features. Figure 6 below shows the heatmap with the correlation value of the features of both android and ios mHealth datasets.

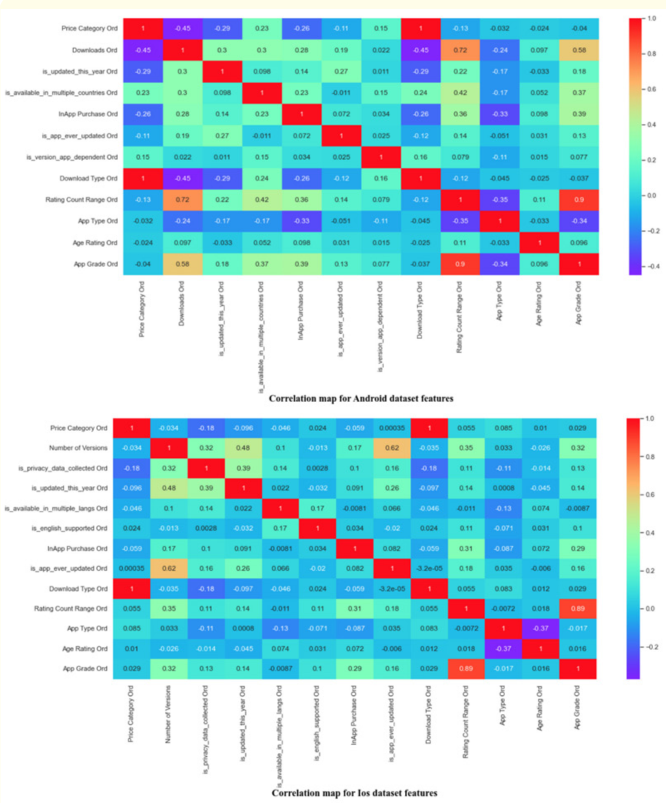

Figure 6: Correlation map shows the relationship between variables in the two different datasets, the Play store and App store.

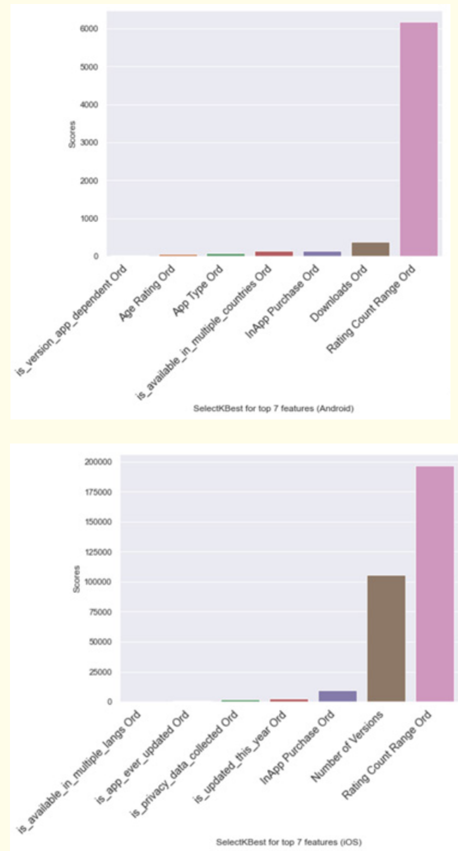

Figure 7: SelectKBest on mHealth application data features a: Play store b: App store.

For the Android dataset, the rating count variable has the highest correlation with the application performance grade with a positive p-value of 0.9. Also, the download counts have a high correlation value of 0.58; in-app purchases and "is available in multiple countries" features have weak positive correlation p-values of 0.39 and 0.37 respectively. In addition, based on the iOS dataset, the rating count variable also has the highest correlation with the application performance grade with a positive p-value of 0.89. This means that the rating count is one of the features that can be considered in predicting the performance of mHealth applications. The number of versions and in-app purchases features have a weak positive correlation value of 0.32 and 0.29 respectively. Furthermore, SelectKBest is a good machine-learning technique to use in selecting features before training the model. With SelectKBest, k number of features can be selected for training the model based on their scores [22]. Figure 7 shows the SelectKBest scores based on both iOS and android datasets.

Classification models

The top 5 features among the features, selected through SelectKBest, were used to generate different predictive models for both iOS and android datasets. Before training the models, each of the datasets was divided into training and test datasets. For the Play store dataset, 30% of the dataset was used as test data and 70% was used as training data while for the App store dataset, 25% was used as test data and 75% was used as training data. Different percentages of the data were used for testing because of the huge difference (in terms of row count) between the two different datasets. The Play store dataset does not contain as many records as the App store dataset. The training datasets are imbalanced and because of this, the training datasets were oversampled for both iOS and android mHealth datasets. SMOTE python oversampling library is used to oversample the training datasets.

Predictive classification models were generated to predict the performance grade of the application. As stated earlier, the application performance was grouped into BAD, FAIR, GOOD, VERY GOOD and EXCELLENT. Predictive models are created and compared using four different supervised learning algorithms.

One of the classifiers used is Naive Bayes which is an inductive algorithm used in data mining and machine learning which works based on the assumption that features are independent [23]. Random Forest was also considered to be one of the ensemble algorithms that use decision trees. The decision tree aims to produce a model capable of predicting dependent variable value through the means of simple decision rules deduced through the features of the dataset [15]. Random Forest makes use of multiple decision trees on different dataset sub-samples and uses averaging to reduce overfitting and increase the classifier accuracy [15]. Among the classifiers considered is logistic regression which predicts the probability of categorical target variables with the assumption that the controlled variables are independent [24]. In addition, XGBoost was used as one of the gradient-boosting frameworks. Using multiple tree models in building a more robust learning model, the XGBoost is capable of performing parallel computing which in turn speeds up computation [25].

In addition, with respect to the iOS dataset, the rating count also has the highest correlation with the application performance grade with a positive p-value of 0.89. The number of versions and in-app purchase features have a weak positive correlation value of 0.32 and 0.29 respectively as shown in figure 6. The top 5 features based on selectKBest scores were used in building predictive models (one for android performance prediction and the other for iOS performance prediction).

Result

The analysis done on the datasets has shown that cost is a vital determining factor in the usability of the mHealth applications. Not only that people prefer to use mHealth applications that are free to download but the majority of the mHealth applications that requires payment on download have bad rating and users do not find them useful. Based on applications that require payment on download, the average download price of the applications on the App store is approximately \$10.25 and £10.04 on the Play store. Figure 8 shows the download price distribution of the mHealth applications on both the Play store and App store based on performance.

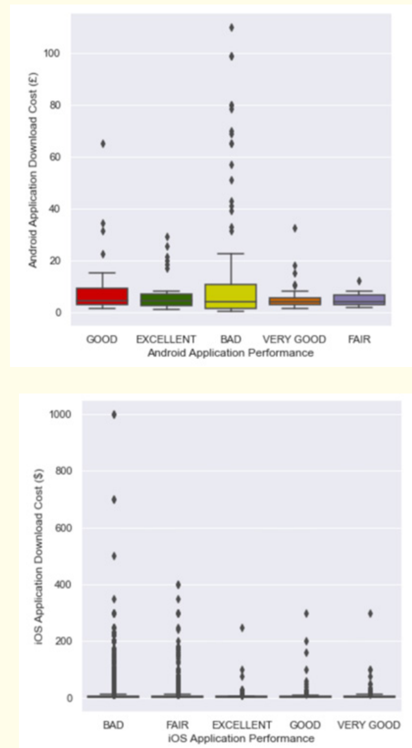

Figure 8: MHealth apps download price distribution. The boxplot shows the distribution of mHealth application prices a: Play store b: App store.

The evaluation metric used for the classifiers is the f1-score because the dataset was oversampled due to the imbalance before training the model. Among the four different classifiers, XGBoost had the best performance on both datasets based on the f1-score. The XGBoost classifier attained an f1-score of 0.90 on the App store dataset and 0.58 on the Play store dataset.

Tables 1 and 2 below represent the classification report of the XGBoost classifier on both android and iOS mHealth datasets. Based on table 1, the android mHealth application performance prediction model (XGBoost) has a f1-score with a weighted average of 0.78. The model has the highest precision and f1-score on class BAD prediction with values of 1.0 and 0.96 respectively. In addition, the prediction on the class label VERY GOOD has the least f1-score with a value of 0.42.

Based on Table 2, the iOS mHealth prediction model (XGBoost) has a f1 score with a weighted average of 0.93. The model has a

	Precision	Precision	F1-Score
BAD	1.00	0.93	0.96
FAIR	0.31	0.75	0.44
GOOD	0.44	0.62	0.52
VERY GOOD	0.58	0.32	0.42
EXCELLENT	0.59	0.64	0.64
Accuracy	-	-	0.77
Macro avg	0.58	0.67	0.6
Weighted avg	0.80	0.77	0.78

Table 1: XGBoost model classification report (Android dataset).

	Precision	Precision	F1-Score
BAD	1.00	0.90	0.95
FAIR	1.00	0.90	0.95
GOOD	1.00	0.90	0.95
VERY GOOD	1.00	0.90	0.95
EXCELLENT	1.00	0.90	0.95
Accuracy	-	-	0.93
Macro avg	0.88	0.92	0.9
Weighted avg	0.94	0.93	0.93

Table 2: XGBoost model classification report (iOS dataset).

perfect precision (i.e. 1.0) on the BAD class prediction and the class label VERY GOOD has the least precision and f1-score of 0.80 and 0.85 respectively.

In addition, the analysis shows that applications that provide in-app purchases perform better than the mHealth applications that do not provide in-app purchases based on the average ratings.

Discussion

This study critically analyzed the datasets and the three research questions were addressed. Below are the findings based on the research questions.

Q1: Do paid applications (on download) perform better than the free applications?

No. The download cost of the mHealth applications on both the Play store and App store does not positively influence the performance of the application. Approximately 85% of the applications that have average ratings between 4.0 and 5.0 are free to download on the Play store and also approximately 93% of the mHealth applications on the App store with average ratings between 4.0 and 5.0 are free to download.

Q2: Do applications with in-app purchases perform better than the applications without in-app purchases?

Yes. Approximately 67% of the mHealth applications on the Play store with ratings between 4.0 and 5.0 provide in-app purchases and approximately 55% of the mHealth applications on the App store with ratings ranging from 4.0 to 5.0 provide in-app purchases.

Q3: Which classifier performs best in predicting mHealth application performance?

Among the five different machine learning classifiers trained to predict mHealth application performance, the XGBoost had the best performance in terms of the f1-score. The XGBoost classifier attained an f1-score with a weighted average of 0.79 on android datasets and an f1-score of 0.93 on ios datasets.

In addition, approximately 93.60% of the mHealth applications on the App store support the English language and approximately 31.55% of mHealth applications on the App store support multiple languages.

MHealth applications are categorized based on countries on the Play store and some of the applications are available in more than one country. Approximately 36.41% of the mHealth applications on the Play store are available in multiple countries.

[5] focused on factors that determine the user's acceptability of the mHealth application and the major factors based on their study are cost, security and ease of use of the application. Similarly, this study shows that none of the mHealth applications that require payment on both the App store and Play store is rated by up to 50,000 users but there are free mHealth applications that are rated by more than 50,000 users to 5,000,000 users. This means that the cost of the application also plays a major role in the usability of the mHealth applications. Hence, the cost of mHealth applications is a very vital factor that influences the adaptation of mHealth applications.

Future Work

This research work can further be extended in the future to analyze the users' comments with respect to the ratings and also by comparing how each application performs generally on both the Play store and App store as opposed to what was done in this research which only analyzed mHealth applications performance on App store and Play store independently.

Conclusions

There are lots of applications that are not updated regularly and also have bad ratings on the App store. The release of mHealth applications should be regulated and reviewed by a central body to avoid misleading apps that can cause more harm to their users. The analysis shows that very few of the applications that require payment before download do not perform well and very few of them have high rating counts. This indicates that the download price of mHealth applications does not positively influence the performance of the applications and this can discourage users from downloading the applications. The applications that provide in-app purchases tend to perform better than the mHealth applications that do not offer in-app purchases, especially for mHealth applications in the Play store. Hence, it is recommended to have an in-app purchase (i.e. users pay for the service they want to use on the application after download) rather than forcing users to pay before downloading or before having a feel of how useful the mHealth application can be to them. Also, the XGBoost classifier had the best performance in predicting the performance of the mHealth applications based on accuracy and f1-score. The XGBoost outperformed other classifiers like Random Forest, Logistic regression and Naive Bayes.

Bibliography

1. Habtemariam MK and Semegn ST. "Setting health sector priorities: a brief overview of Ethiopia's experience". *Cost Effectiveness and Resource Allocation* 16.1 (2016): 1-3.
2. Nolan P. "Measuring productivity in the health sector". Policy Quarterly Special issue: Assessing and Enhancing New Zealand's Productivity 14.3 (2018).
3. Mathews SC., *et al*. "Digital health: a path to validation". *NPJ Digital Medicine* 2.1 (2019): 1-9.
4. Park Y. "Emerging new era of mHealth technologies". *Healthcare Informatics Research* 22.4 (2016): 253-254.
5. Wang C and Qi H. "Influencing Factors of Acceptance and Use Behavior of Mobile Health Application Users: Systematic Review". *Healthcare (Basel)* 9.3 (2021): 357.
6. Steinhubl SR., *et al*. "The emerging field of mHealth". *Science Translational Medicine* 7.283 (2015): rv3.
7. Zhou L., *et al*. "The mHealth App Usability Questionnaire (MAUQ): development and validation study". *JMIR Mhealth Uhealth* 7.4 (2019): e11500.
8. https://apps.apple.com/us/genre/ios-health-fitness/id6013
9. https://apps.apple.com/us/genre/ios-medical/id6020
10. https://play.google.com/store/apps/category/HEALTH_AND_FITNESS
11. https://play.google.com/store/apps/category/MEDICAL
12. Zhao Bo. "Web Scraping". (2017): 1-3.
13. Gheorghe M., *et al*. "Modern techniques of web scraping for data scientists". *International Journal of User-System Interaction* 11.1 (2018): 63-75.
14. Awangga RM., *et al*. "Implementation of web scraping on Github task monitoring system". *TELKOMNIKA (Telecommunication Computing Electronics and Control*) 17.1 (2019): 275-281.
15. Kamiran F and Calders T. "Data preprocessing techniques for classification without discrimination". *Knowledge and Information Systems* 33.1 (2012): 1-33.
16. Nargesian F., *et al*. "Learning Feature Engineering for Classification". In (2017): 2529–2535.
17. Hunter JD. "Matplotlib: A 2D graphics environment". *Computing in Science Engineering* 9.3 (2007): 90-95.
18. Kodinariya TM and Makwana PR. "Review on determining number of Cluster in K-Means Clustering". *International Journal* 1.6 (2013): 90-95.
19. Waskom ML. "Seaborn: statistical data visualization" 6.60 (2021): 3021.

20. Rahaman M., *et al*. "Machine learning to predict the martensite start temperature in steels". *Metallurgical and Materials Transactions A* 50.5 (2019): 2081-2091.

21. M A H Shaikh and K Barbé. "Initial Estimation of Wiener-Hammerstein System with Random Forest". 2019 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Auckland, New Zealand (2019): 1-6.

22. Pedregosa F., *et al*. "Scikit-learn: Machine learning in Python". *The Journal of Machine Learning Research* 12 (2011): 2825-2830.

23. Zhang Harry. "The Optimality of Naive Bayes". Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference, FLAIRS (2004): 2.

24. Susan L. "Building a Logistic Regression in Python, Step by Step" (2017).

25. Li W., *et al*. "Gene expression value prediction based on XGBoost algorithm". *Frontiers in Genetics* 10 (2019): 1077.