Mobile Applications Rating Performance: A Survey

https://doi.org/10.3991/ijim.v16i19.32051

Sabreen Abulhaija(✉), Shayma Hattab, Ahmad Abdeen, Wael Etaiwi
King Talal School of Business Technology, Princess Sumaya University for Technology,
Amman, Jordan
abu20208045@std.psut.edu.jo

Abstract—The use of mobile phones is increasing all the time. These phones have become increasingly vital and beneficial in all parts of our lives, including social and business sides. Mobile applications are expanding with new upgrades and editions every day due to this expansion. This increase makes it more difficult for consumers, particularly those who are not technologically minded, to determine which applications to install and use. It is much more difficult for developers to ensure that their apps will be used and lucrative. Several research papers have been published in the recent five years to investigate mobile applications' rating to aid users and developers in making the best decision possible by employing various classifications and methodologies. This study provides a literature review research analyzed mobile app evaluations from 2018 to 2022 using various datasets. In addition, a new taxonomy is proposed to classify the research papers that looked at the rating of mobile apps into three categories: predictive modeling, sentiment analysis, and priority ranking of the most significant features.

Keywords—machine learning, mobile applications, rating performance, sentiment analysis, predictive modeling

1 Introduction

The Mobile application (AKA: Mobile-App) industry has been developed radically in the last ten years. These applications provide users with unlimited functions that make users' life more entertaining, comfortable, and excited, by delivering services such as online shopping, food ordering, gaming, health management, etc. However, some of these applications are not useful or are not working properly. Hence, users are always looking for an application with a high rating and positive reviews to decide whether to download this application or not. Reference to the recent statistics, more than 2.5 billion people are using a smartphone, and more than twelve million developers have developed applications for these smartphones [1]. Developers were accessible to more than 5 million apps on electronic stores such as Apple, Google, and Amazon, gaining over 200b downloads [1].

This trend is followed by a growing number of mobile software businesses delivering a massive number of mobile applications. Specifically, there are two giant platforms in the market provided by Google Play Store (GPS) and Apple Store (iOS). Mobile applications are offered for free and subscription-based. The app store has a massive number of free applications, making the market very competitive and providing many alternatives for users. And as a part of customer service management, these two platforms allow their users to evaluate the applications and provide reviews and opinions.

Application rating represents all reviews received from users' responses. However, not all applications have excellent ratings, and users would instead download applications with top ratings since they expect them to be more effective and of higher quality. Mobile applications on electronic stores receive on average 22 ratings daily, which depends on the popularity of the application and might reach to few thousands daily. Moreover, only one-third of these reviews are useful to analysts and developers [1].

These reviews and feedback play a vital role in both the developers' business and users' experience in this competitive world [2]. As users provide developers with their feedback which leads to applications update and enhancement, in addition, to increasing security measures, and as a result, attracting more users.

We are motivated to introduce this paper, as it will add value to businesses and developers before launching their mobile applications, by giving them background about mobile stores rating analysis and an overview about the rating analysis categories of mobile applications, available datasets in the market, machine learning performance on these datasets and different modeling's. Moreover, this paper will help developers to understand which attributes are affecting user satisfaction to take them into account during mobile applications development phase. Also, it the first of its kind in the mobile applications industry, by proposing a systematic review, in which, the mobile application's rating analysis are studied and classified. The studied paper researches are classified into three main categories: predictive modeling, sentiment analysis, and priority ranking of most important features.

This research paper is organized as the following: section two includes the background of the survey; section three includes the methodology. Section four includes the analysis of reviewed papers, section five illustrates the discussion and results, and finally the last section shows the conclusion.

2 Background

2.1 Predictive modeling

Fernandez and Gallardo-Gallardo [3] pointed out that predictive analytics term could be detecting what could happen in the future, evaluating historical and past data, and discovering the relationships between these data in an effort to predict future circumstances. The predictive analytics utilizes predictive modeling by using several machine learning techniques. Most of the predictive models give a score to indicate the probability of the event occurrence. A greater score implies the greater possibility of an event happening and a lower score implies the lower possibility of an event happening. Historical data is utilized by the predictive models to reveal solutions for many business problems according to the research of Kumar and Garg, [4]. These models are valuable in recognizing the opportunities and risks for different business domains include; predicting the sales, credit card fraud detection, and distinguishing the stocks that might give a high return on their investment. Hence, it helps the business to be proactive and to be able to take a proper decision at the right time [4]. The Predictive analytics approach has several stages to predict the future starting from collecting the data from different sources, preprocessing, splitting the data into training and testing sets using specific criteria, building the predictive model then assessing the predictive performance [5]. Predictive models utilize different types of machine learning techniques, the most common is supervised and unsupervised learning. Supervised machine learning techniques refer to the labeled training data such as Nearest Neighbor, Gaussian Naive Bayes (GNB), Decision Trees (DT), Support Vector Machine (SVM), and Random Forest (RF). On the other hand, unsupervised machine learning techniques refer to the unlabeled training data such as KNN (k-nearest neighbors) and Neural Networks [6].

2.2 Priority ranking

Priority ranking or feature prioritization is the process of sorting a feature from most to least important [7] using various methodologies such as descriptive statistics or by using the information gain, which can help in the selection of the most relevant features and attributes that have a significant impact on the main objective. The importance of the priority ranking comes from that many of the domains have a large number of attributes some of them are irrelevant and cause noise and will be costly specifically when we have limited time and cost [8].

2.3 Sentiment analysis

According to Yue *et al*. [9] sentiment analysis (or opinion mining) seeks to analyze people's opinions about entities such as individuals, products, services, and companies. Moreover, Bandana [10] mentioned that the term sentiment is an emotion or attitude, and the term sentiment analysis is to identify the opinions and reactions, and subjective feelings toward a certain subject in a sentence or document. Furthermore, he mentioned that sentiment analysis could be applied to several domains like services, products, political elections, movie and book reviews, etc. Dragoni *et al*. [11] pointed out that sentiment analysis intends to categorize the text as positive, negative, or neutral. Sentiment analysis techniques could be classified into symbolic and sub-symbolic approaches, the symbolic approach includes using lexicons, and ontologies, while the sub-symbolic approach includes using machine learning techniques that classified the reactions and feelings based on the frequency of simultaneous words [11]. The growth in the volume and the variety of data and information, make sentiment analysis more vital from different perspectives; from the commercial perspective, sentiment analysis is able to deliver online recommendations for the buyers and sellers, also from the marketing perspective makes it possible to know the customers' preference in specific products and services [9]. Moreover, the importance of sentiment analysis comes as well from that human opinions and reviews are influenced greatly by others' opinions and reviews, when consumers need to know about specific entities such as products, services, or events, they searched for others' feedback. Hence, it is effective for businesses to have an accurate sentiment analysis system that can accurately yield correct sentiment and relevant information [10].

3 Methodology

We searched the literature to identify the research papers that analyze mobile applications performance using a set of predefined keywords including Mobile Apps, Apple Apps, Android Apps, Google Play Store. The datasets included in the search step were: IEEE, ACM, Springer, Wiley, and Elsevier databases. The publication period is restricted to the paper researches that are published during the period 2018-2022. as a result, we identified 27 papers, then we filtered the papers by reviewing the title, the abstract, and the conclusion, filtration phase resulted in excluding five papers for irrelevance. The review included analyzing and summarizing each paper’s objectives, methodology experiment, characteristics of the datasets, results, contributions, limitations, and future work. After conducting the review and analysis a classification for those paper researches is proposed.

3.1 Predictive modeling

The research of Magar *et al*. [12] used the GPS dataset to classify the overall popularity of an app and use the number of installs as the measure. They used six machine learning (ML) algorithms; Logistic Regression (LR), Random Forest (RF), Stochastic Gradient Descent (SGD), KNN, SVM, DT, and the experimental results showed that the SVM classifier produced best results. The models were based only on the top five important features external to the app and they did not include features internal to the app such as the software features and performance of the app.

The study of Sarro *et al.* [2] analyzed 11,537 apps from BlackBerry and Samsung World app stores, they used Natural Language Processing (NLP) technique to obtain from current app types, feature data that capture some of the operations of these apps. They also used Case-Based Reasoning (CBR) to predict the rating of the apps by relying on the claimed features. The results indicated that the ranking of 89% of those apps can be predicted 100% accurately. The findings of the study could help in the requirements engineering of the app stores and provide chances to encourage the needs induction procedure for developers.

The study of Bashir *et al.* [13] proposed a modern structure that offers developers an efficient approach to effectively discover in the competitive Mobile-App industry. By comparing the predicted ranking and downloads numbers with the original dataset. They analyzed the GPS dataset using ML techniques to predict rating and downloads number before going live with the app on the store to help the developers assess their work. The result showed that SVM and KNN can deliver better accuracy than RF.

The research by Suleman *et al.* [14] which aimed to predict rating on GPS using a real-time dataset collected in 2018 contained 10839 records and 8 attributes with the following names; app name, number of reviews, downloads volume, size of the app, categories, content ranking, android version and with a class named as rating. They used several ML techniques including DT, LR, SVM, Naïve Bayesian (NB), K-Means, KNN, and Artificial Neural Networks (ANN). Their methodology contained many processes, including collecting, cleaning, and feature reductions. The authors used MATLAB for the data visualizations. The result showed that DT has the best results in making rating predictions among other techniques.

Daimi *et al.* [15] pointed out that there are many users who are not technically oriented and do not have much deep knowledge about the mobile applications, therefore they depend on the applications rates to choose the most appropriate one, hence the aim of their paper research was to predict user rating of mobile applications using iOS dataset. The authors downloaded the dataset from Kaggle [1]contains 7197 rows and 16 attributes. They used 7 ML techniques including SVM, NN, RF, M5 Rules, LR, and Random Tree, all of them employed by using WEKA (a ML software). The result showed that the RF has yielded the best results for predicting the user rating for the iOS dataset among other techniques.

The proposed paper of Umer *et al.* [16] intended to forecast the numeric ranking of GPS apps using ML classifiers including gradient boosting classifier (XGB), RF, gradient boosting classifier (GBM), extra tree classifier (ET), and the extreme AdaBoost classifier (AB). The dataset was aggregated from the GPS utilizing the Beautiful-Soup (BS) web scraper contained 658 records and included attributes such as: “App_category, App name, App_id, App_review, and App_ rating”. The dataset for this paper was semi-structured which requires several preprocessing techniques to analyze it including features selection. The result showed that GBM and ET have produced the most exact numeric rating predictions. Future work included applying a Deep Learning (DL) algorithm for numeric rating prediction.

The study of Kayalvily *et al.* [17] aimed to predict GPS apps rating using the simplest ML technique which is DT. The dataset was collected from GPS in 2019. They used KDD Methodology to understand and analyze the data and used Tableau for visualization. They concluded that the price and number of downloads have a strong influence on the user ratings.

Sadiq *et al.* [18] aimed in their research to predict numeric reviews and ratings of GPS apps by using DL approaches including Bidirectional Long Short-term Memory (BiLSTM), Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), Long short-term memory (LSTM), and Gated Recurrent Unit (GRU). The dataset collected from GPS contained fourteen different sorts of mobile apps including 658 records that was scrapped by using BeautifulSoup (BS) web scraper, which is a Python package for parsing HTML and XML documents. Different attributes of the dataset are used; “App_id, Appname, App_category, App_review, and App_rating”. The dataset was unstructured and noisy which required preprocessing like cleaning, removing duplicated data, etc. The results showed that the CNN gave the most numeric rating exact predictions than others with the results of 89% recall and 82% precision

In their research paper, Sandag *et al.* [19] created a prediction model using user rating (dataset 2019) for Android apps on GPS utilizing the KKN algorithm. The results

---

[1] www.kaggle.com

showed that education is the most reviewed and book & reference is the highest-rated and dating category is the lowest. On the other hand, KNN performed well in predicting Android applications based on the fivefold cross-validation dataset.

3.2 Priority ranking

Additionally, the research of Mahmud *et al.* [7] proposed a category prioritizing method by studying the rankings and the reviews of the apps, in order to recognize the most important features to consider for a future better release. The authors used the NB and the J48 decision tree classifier, where NB had better results and the results also showed that utilization of resources and application performance had the highest priority rank.

Moreover, the paper of Lengkong and Maringka [20] used RF, KNN, GB, and DT to identify the most influential characteristics of high rated apps in GPS, such as the features of Size, installs, reviews, types (free vs paid), rating, category, content rating, and Price. The research summarized that the GB has the greatest performance with 100% accuracy, 100% Precision, and 100% Recall, it was also noted that the factors install and reviews are the most influential in predicting high-ranked apps.

The study of Mahmood [21] ensured that users seek to download applications that have a high rate as they considered high quality and will be more satisfied. His research question was to know which aspects influence the apps' ranking in GPS. The dataset collected from Kaggle contains 10,840 records, including the following attributes: app name, current version, app id, installs, reviews, size, category, rating, type, content rating, price, last updated, and android version. He used RF, Support Vector Regression (SVR), LR, and Pearson Correlation to know which attributes influence the rating of the app. RF defines the significance of all the factors and their impact on the rating, and it shows that the number of evaluations, genre, app size, and character count in the name are the most important variables than others. The results for SVR showed that content rating and word counting in the name are the most influential factors of rating. Giving the LR model, symbol count in name, and type of app most significant variables. For the Pearson correlation model which is used to calculate the correlation of the binary factors with the rating, the results appeared that symbol count in name has elevated importance. In his future research, he aimed to use high numbers of records and to be able to predict the ratings of the app.

The paper of Dhinakaran *et al*. [1] proposed an active learning technique to decrease the human effort in the review analysis. The proposed app review classification approach utilizes three active learning tactics based on uncertainty sampling. It was found that active learning with comparing to a randomly selected training dataset, generates a much greater prediction accuracy using multiple scenarios.

3.3 Sentiment analysis

The propose research of Martens and Maalej [22] used an iOS dataset for fake reviews detection and used machine learning algorithms including; RF, DT, MLP, SVP, and Gaussian NB, where the RF classifier had the best results. The study resulted in

(35,5%) out of 62,617,037 reviews classed as false. The paper discovered variations between official and false reviews. We noticed that the properties of the corresponding app and evaluator are most valuable to verify if a review is false.

The paper of Jha and Mahmoud [23] analyzed a dataset that contains 6,000 reviews of Apple app categories to identify reviews related to Non-Functional Requirements (NFRs). They used NB and SVM classifiers, where the SVM had better results, the results showed that 40% of the reviews indicate at least one type of NFRs.

The paper of Aralikatte *et al*. [24] discussed the review rating mismatch problem and established the demand for a system that can automatically identify irregularities between evaluations and rankings. The authors applied NB, DT (J48), AdaBoost, KNN, SVM, Holte's 1R & CNN. They examined 8600 reviews from ten apps available for Android through developed multiple models based on machine and deep learning techniques. and found that about 20% of the reviews had ratings and reviews mismatch.

The research of Luiz *et al*. [25] proposed a general approach to allow app developers to review and evaluate user reviews regarding applications on stores, the approach to automatically obtain related features from reviews and study the sentiment related to them. The framework helps in detecting topics that are negatively affecting the overall ranking of a particular application. The topic modeling block is built on the Non-negative matrix factorization (NMF) strategy and they used the SACI strategy for the sentiment analysis.

The research by Ranjan and Mishra [26] aimed to apply sentiment classification of application reviews. They analyzed a dataset downloaded from Kaggle which contains 9659 apps and 13 attributes including the following attributes; app name, rating, and category, etc. They used several ML algorithms; NB, SVM, logistic regression (LR), KNN, and Random. The results showed that the SVM performed best results than others with an accuracy of 93.41%.

The study of Rahman *et al*. [27] analyzed Android-App-Reviews-Dataset downloaded from GitHub. The dataset contained 20,000 records. They used several of machine learning techniques to perform sentiment analysis on the android application which included; KNN, RF, SVM, DT, and NB techniques. The results showed that SVM has the maximum accuracy with a percentage of 88.9% among other techniques.

The study of Pratama *et al*. [28] compared the performance of various mixtures of machine translations and lexicon resources to understand the greatest resource mixture is to be used in lexicon-based sentiment analysis on App Review. The result indicates that the mixture of Google Translate and SentiWordNet can achieve the greatest accuracy.

The study of Soumik *et al*. [29] offered in-depth insight on several current methods to implement sentiment analysis utilizing text classification on a dataset from Bangla extracted from GPS. The results showed that NV, SVM, and LR have presented very encouraging results despite the data limitation. An Ensemble technique is also introduced with Adaptive Boosting revealing a great accuracy score with 5-fold applied. However, SVM has the greatest accuracy score among all the techniques when 5-fold is used, and GBM has the best accuracy score when five-fold is not used.

The research of Suresh *et al*. [30] proposed sentiment analysis on chosen reviews combined with specific features to forecast the star ratings which define the app's success. The testing results showed low MSE rates for SDG and SVR. Hence, people need updated means for the success of business applications prediction.

4 Discussion and results

Based on analyzing the papers, the proposed researches to assess the performance of the mobile application are categorized into three main categories focused on: predictive modeling, sentiment analysis, and priority ranking of most important features. The reviewed researches are summarized in Table 1.

According to the reviewed researches, the contributions are equally divided on predictive modeling and sentiment analysis with 41% each, while 18% of scholars addressed priority ranking. The distribution of the paper according to the proposed classification taxonomy is represented in Figure 1. Furthermore, the scholars mainly used GPS datasets in their experiments as it counted for 75% of the datasets used, this is due to the fact that GPS is the biggest Android app in the market [29].

Additionally, most of the proposed studies focused on using machine learning as the main technique to analyse the mobile application performance. On the other hand, few studies focused on using deep learning and active learning in the proposed models. Scholars who analyzed the mobile apps using predictive modeling used different machine learning algorithms to predict the app performance rating using LR, KNN, SGD, DT, RF, SVM, DT, NB, K-Means, ANN, REP, RF, M5, LR, GBM, XGB, AB, ET and NLP, and deep learning techniques including CNN, RNN, LSTM, BiLSTM, and GRU. According to the proposed study of Ongsulee [31], machine learning is used in predictive modeling since it allows researchers to produce reliable, decisions and support in uncovering hidden insights through learning from historical relationships and trends in the data, while deep learning is a part of a broader family of machine learning techniques based on learning representations of data and its concerned with artificial neural networks and other machine learning algorithms that include more than one hidden layer.

Scholars who analyzed the priority ranking of the features mainly focused on machine learning techniques including SVM, LR, NB, J48, RF, KNN, DT, GB, and active learning. According to the research of Settles [32], active learning, which is also called query learning is concerned with asking queries in the form of unlabeled instances to be labeled to overcome the labeling bottleneck.

Scholars who analyzed the performance of mobile applications using sentiment analysis used machine learning techniques such as; NB, SVM, LR, KNN and RF, DT, SVR, SDG, and deep learning techniques such as CNN, in addition to using lexicon and ensemble methods. According to the paper of Dietterich [33] ensemble techniques are learning algorithms that use a set of classifiers and classify new data points by taking a weighted vote of the predictions and they are used for gaining highly accurate classifiers by combining the less accurate ones.

In addition, Badawood and AlBadri discussed in their paper [34] the purpose of users in the learning environment to accept and adopt mobile learning, their perceptions and considerations that obstruct the implementation of mobile learning in the gulf region. Their paper used a systematic literature review to gather information from post-2017 studies. The model was constructed based on the “Theory of Planned Behavior” and “Unified Theory of Acceptance” and “Use of Technology”. Based on the developed model, main ideas such as performance expectancy, effort expectancy, and social influence are greatly affected by other factors like learner’s creativity and mobility.

Moreover, the app stores unique characteristics could have an impact on the rating importance; as the paper of Strzelecki [35] discussed that some app stores started extending the control over apps description and the developers are only able to edit specific fields, in addition to tightening the pricing policy, those factors may be of interest in reviewing the performance of the applications.

Additionally, the use of mobile applications in education has been strongly depend on client and instructor reviews, developers’ descriptions, and configured with the related collaboration techniques. Moreover, there is a limited experimental proof to give recommendations on the mobile apps value because it has been rated or used by kids [36].

Table 1. Summary table

Type	Target Store	Technique/ Algorithm	Number of Attributes	Dataset Records	Ref.
Predictive Modeling	Google	LR, KNN, SGD, DT, RF & SVM	23	600000	[12]
	BlackBerry & Samsung	CBR & NLP	BlackBerry:1,256 Samsung: 620	BlackBerry: 9,588 Samsung: 1,949	[2]
	Google	RF, KNN & SVM	8	NA	[13]
	Google	DT, LR, SVM, NB, KNN, K-Mean & ANN	8	10839	[14]
	Apple	SVM, ANN, REP, RF, M5, LR and RF	16	7197	[15]
	Google	RF, GBM, XGB, AB & ET	5	658	[16]
	Google	DT	NA	NA	[17]
	Google	CNN, RNN, LSTM, BiLSTM and GRU.	5	658	[18]
	Google	KNN	11	32000	[19]
Priority Ranking	Google	NB and J48	12	7754	[7]
	Google	RF, KNN, DT & GB	13	10842	[20]
	Google	RF, SVM & LR	17	10840	[21]
	Apple & Google	Active Learning	4	4,400	[1]
Sentiment Analysis	Apple	RF, DT, MLP, SVP, NB	15	62 Million	[22]
	Apple	NV & SVM	NA	6000	[23]

	Google	NB, DT (J48), Ada-Boost, KNN, SVM, Holte's 1R & CNN	8	8600	[24]
	Google	NMF & SACI	NA	NA	[25]
	Google	NB, SVM, LR, KNN & RF	13	9659	[26]
	Google	KNN, RF, SVM, DT & NB	40	20000	[27]
	Apple & Google	Lexicon	5	553	[28]
	Google	NB, SVM, LR & Ensemble Methods	3	10000	[29]
	Google	SVR & SDG	NA	NA	[30]

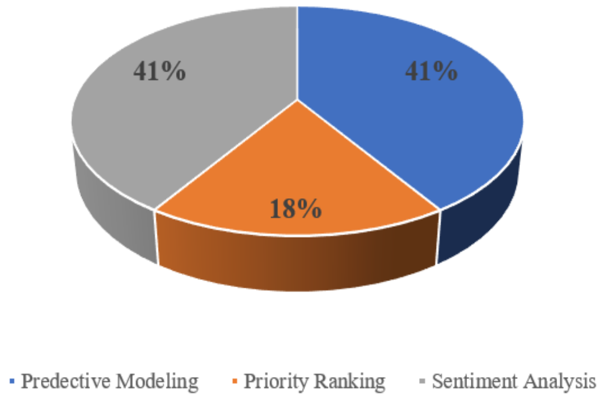

Fig. 1. Distribution of papers

5 Conclusion

The number of mobile applications is enormous and most of the users depend on the mobile applications' rating in their decisions. In this survey, the studies of mobile applications' rating are analyzed and classified into three categories; predictive modeling, sentiment analysis, and priority ranking of most important features.

Given the previous literature review, the studies focused on predictive modeling and sentiment analysis, while few papers focused on priority ranking. The scholars mainly used machine learning techniques and few of them used deep learning, active learning, and ensemble methods. Moreover, the datasets of GPS were the most commonly used, while few models used Apple, Blackberry, and Samsung datasets.

This paper provided insights for scholars on the related research in the domain of mobile applications performance, in addition, to providing app developers information regarding the studies and analysis conducted regarding the application performance, which in turn will help them consider the important factors in their developments.

As future work, considering deep learning in the analyses studies might provide more accurate results since it is useful and powerful in many machine learning applications.

6 References

[1] Dhinakaran, V. T., Pulle, R., Ajmeri, N., & Murukannaiah, P. K. (2018, August). App review analysis via active learning: reducing supervision effort without compromising classification accuracy. In 2018 IEEE 26th international requirements engineering conference (RE) (pp. 170-181). IEEE. https://doi.org/10.1109/RE.2018.00026

[2] Sarro, F., Harman, M., Jia, Y., & Zhang, Y. (2018, August). Customer rating reactions can be predicted purely using app features. In 2018 IEEE 26th International Requirements Engineering Conference (RE) (pp. 76-87). IEEE. https://doi.org/10.1109/RE.2018.00018

[3] Fernandez, V., & Gallardo-Gallardo, E. (2021). Tackling the HR digitalization challenge: key factors and barriers to HR analytics adoption. Competitiveness Review, 31(1), 162–187. https://doi.org/10.1108/CR-12-2019-0163

[4] Kumar, V., & Garg, M. L. (2018). Predictive analytics: a review of trends and techniques. International Journal of Computer Applications, 182(1), 31-37. https://doi.org/10.5120/ijca2018917434

[5] Kuhn, M., & Johnson, K. (2019). Feature engineering and selection: A practical approach for predictive models. CRC Press. https://doi.org/10.1201/9781315108230

[6] Schmitt, J., Bönig, J., Borggräfe, T., Beitinger, G., & Deuse, J. (2020). Predictive model-based quality inspection using Machine Learning and Edge Cloud Computing. Advanced engineering informatics, 45, 101101. https://doi.org/10.1016/j.aei.2020.101101

[7] Mahmud, O., Niloy, N. T., Rahman, M. A., & Siddik, M. S. (2019, June). Predicting an effective android application release based on user reviews and ratings. In 2019 7th International Conference on Smart Computing & Communications (ICSCC) (pp. 1-5). IEEE. https://doi.org/10.1109/ICSCC.2019.8843677

[8] Daeli, N. O. F., & Adiwijaya, A. (2020). Sentiment analysis on movie reviews using Information gain and K-nearest neighbor. Journal of Data Science and Its Applications, 3(1), 1-7.

[9] Yue, L., Chen, W., Li, X., Zuo, W., & Yin, M. (2019). A survey of sentiment analysis in social media. Knowledge and Information Systems, 60(2), 617-663. https://doi.org/10.1007/s10115-018-1236-4

[10] Bandana, R. (2018, May). Sentiment analysis of movie reviews using heterogeneous features. In 2018 2nd International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech) (pp. 1-4). IEEE. https://doi.org/10.1109/IEMENTECH.2018.8465346

[11] Dragoni, M., Poria, S., & Cambria, E. (2018). OntoSenticNet: A commonsense ontology for sentiment analysis. IEEE Intelligent Systems, 33(3), 77-85. https://doi.org/10.1109/MIS.2018.033001419

[12] Magar, B. T., Mali, S., & Abdelfattah, E. (2021, January). App Success Classification Using Machine Learning Models. In 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC) (pp. 0642-0647). IEEE. https://doi.org/10.1109/CCWC51732.2021.9376021

[13] Bashir, G. M. M., Hossen, M. S., Karmoker, D., & Kamal, M. J. (2019, December). Android apps success prediction before uploading on google play store. In 2019 International Conference on Sustainable Technologies for Industry 4.0 (STI) (pp. 1-6). IEEE.

[14] Suleman, M., Malik, A., & Hussain, S. S. (2019). Google play store app ranking prediction using machine learning algorithm. Urdu News Headline, Text Classification by Using Different Machine Learning Algorithms, 57.

[15] Daimi, K., & Hazzazi, N. Using Apple Store Dataset to Predict User Rating of Mobile Applications.

[16] Umer, M., Ashraf, I., Mehmood, A., Ullah, S., & Choi, G. S. (2021). Predicting numeric ratings for google apps using text features and ensemble learning. ETRI Journal, 43(1), 95-108. https://doi.org/10.4218/etrij.2019-0443

[17] Kayalvily, T., Denis, A., Mohd Norshahriel, A. R., & Sarasvathi, N. (2022). Data Analysis and Rating Prediction on Google Play Store Using Data-Mining Techniques. Journal of Data Science, 2022(01).

[18] Sadiq, S., Umer, M., Ullah, S., Mirjalili, S., Rupapara, V., & Nappi, M. (2021). Discrepancy detection between actual user reviews and numeric ratings of Google App store using deep learning. Expert Systems with Applications, 181, 115111. https://doi.org/10.1016/j.eswa.2021.115111

[19] Sandag, G. A., & Gara, F. (2020, October). Android Application Market Prediction Based on User Ratings Using KNN. In 2020 2nd International Conference on Cybernetics and Intelligent System (ICORIS) (pp. 1-5). IEEE.

[20] Lengkong, O., & Maringka, R. (2020, October). Apps Rating Classification on Play Store Using Gradient Boost Algorithm. In 2020 2nd International Conference on Cybernetics and Intelligent System (ICORIS) (pp. 1-5). IEEE. https://doi.org/10.1109/ICORIS50180.2020.9320756

[21] Mahmood, A. (2020). Identifying the influence of various factor of apps on google play apps ratings. Journal of Data, Information and Management, 2(1), 15-23. https://doi.org/10.1007/s42488-019-00015-w

[22] Martens, D., & Maalej, W. (2019). Towards understanding and detecting fake reviews in app stores. Empirical Software Engineering, 24(6), 3316-3355. https://doi.org/10.1007/s10664-019-09706-9

[23] Jha, N., & Mahmoud, A. (2019). Mining non-functional requirements from app store reviews. Empirical Software Engineering, 24(6), 3659-3695. https://doi.org/10.1007/s10664-019-09716-7

[24] Aralikatte, R., Sridhara, G., Gantayat, N., & Mani, S. (2018, January). Fault in your stars: an analysis of android app reviews. In Proceedings of the acm india joint international conference on data science and management of data (pp. 57-66). https://doi.org/10.1145/3152494.3152500

[25] Luiz, W., Viegas, F., Alencar, R., Mourão, F., Salles, T., Carvalho, D., ... & Rocha, L. (2018, April). A feature-oriented sentiment rating for mobile app reviews. In Proceedings of the 2018 World Wide Web Conference (pp. 1909-1918). https://doi.org/10.1145/3178876.3186168

[26] Ranjan, S., & Mishra, S. (2020, July). Comparative sentiment analysis of app reviews. In 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-7). IEEE. https://doi.org/10.1109/ICCCNT49239.2020.9225348

[27] Rahman, M., Rahman, S. S. M. M., Allayear, S. M., Patwary, M., Karim, F., Munna, M., & Ahmed, T. (2020). A sentiment analysis based approach for understanding the user satisfaction on android application. In *Data Engineering and Communication Technology* (pp. 397-407). Springer, Singapore. https://doi.org/10.1007/978-981-15-1097-7_33

[28] Pratama, B. T., Utami, E., & Sunyoto, A. (2019, March). A comparison of the use of several different resources on lexicon based Indonesian sentiment analysis on app review dataset.

In 2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT) (pp. 282-287). IEEE. https://doi.org/10.1109/ICAIIT.2019.8834531
[29] Soumik, M. M. J., Farhavi, S. S. M., Eva, F., Sinha, T., & Alam, M. S. (2019, December). Employing machine learning techniques on sentiment analysis of google play store bangla reviews. In 2019 22nd International Conference on Computer and Information Technology (ICCIT) (pp. 1-5). IEEE.
[30] Suresh, K. P., & Urolagin, S. (2020, January). Android App Success Prediction based on Reviews. In 2020 International Conference on Computation, Automation and Knowledge Management (ICCAKM) (pp. 358-362). IEEE. https://doi.org/10.1109/ICCAKM46823.2020.9051529
[31] Ongsulee, P. (2017, November). Artificial intelligence, machine learning and deep learning. In 2017 15th International Conference on ICT and Knowledge Engineering (ICT&KE) (pp. 1-6). IEEE. https://doi.org/10.1109/ICTKE.2017.8259629
[32] Settles, B. (2009). Active learning literature survey.
[33] Dietterich, T. G. (2000, June). Ensemble methods in machine learning. In International workshop on multiple classifier systems (pp. 1-15). Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-45014-9_1
[34] Badawood, A., & AlBadri, H. (2021). Technology Based Model of a Mobile Knowledge as a Service to Facilitate Education Community. International Journal of Interactive Mobile Technologies, 15(24). https://doi.org/10.3991/ijim.v15i24.27335
[35] Strzelecki, A. (2020). Application of Developers' and Users' Dependent Factors in App Store Optimization. International Journal of Interactive Mobile Technologies (iJIM), 14(13), pp. 91–106. https://doi.org/10.3991/ijim.v14i13.14143
[36] Herodotou, C., Mangafa, C., & Srisontisuk, P. (2022). An Experimental Investigation of 'Drill-and-Practice'Mobile Apps and Young Children. International journal of interactive mobile technologies, 16(7). https://doi.org/10.3991/ijim.v16i07.27893

7 Authors

Sabreen Abulhaija is a business consulting professional with extensive experience in risk management, internal audit, governance and business analytics, she led many advisory projects across different industries in Jordan, Saudi Arabia, UAE and Iraq. she holds Master's Degree in Business Analytics from Princesses Sumaya University for Technology (email: abu20208045@std.psut.edu.jo).

Shayma Hattab is a senior operation officer at an Information Technology (IT) company. She received her BSc degree in Management Information System from the University of Jordan in 2020, and her MSc Degree in Business Analytics in 2022 from Princess Sumaya University for Technology. Shayma has more than 2 years of experience in machine learning and analytics systems. Her research interests include predictive analytics, Data mining, and Machine Learning (email: shy20208104@std.psut.edu.jo).

Ahmad Abdeen is a business consulting professional who has a mixed of practical and advisory experience in governance, risk and compliance functions since 2011 with focus on MENA region. He has a bachelor's degree in business management from the University of Jordan and master's degree in Business Analytics from Princess Sumaya University for Technology (email: ahm20208066@std.psut.edu.jo).

Wael Etaiwi is an assistant professor in the Department of Business Information Technology at Princess Sumaya University for Technology, Jordan. He received his BSc degree in Computer Information Systems from the Hashemite University in 2007, his MSc Degree in Computer Science in 2011 from Al Balqaa Applied University, and his Ph.D. in Computer Science from Princess Sumaya University for Technology in 2020. Dr. Al Etaiwi has 13 years of experience in software development and system analysis. His research interests include, but are not limited to, Artificial intelligence, Data mining, and Natural Language Processing (email: w.etaiwi@psut.edu.jo).

Article submitted 2022-04-28. Resubmitted 2022-08-20. Final acceptance 2022-08-20. Final version published as submitted by the authors.