Peertechz

ENGINEERING GROUP

TRENDS IN Computer Science and Information Technology
OPEN ACCESS JOURNAL

ISSN: 2641-3086 DOI: https://dx.doi.org/10.17352/tcsit

Research Article

Machine Learning Based-prediction of Health Application Effectiveness on Google Play Store

Nathan Andrie Ama*

Department of Agribusiness, Southern Leyte State University, Hinunangan, Philippines

Received: 08 October, 2025
Accepted: 13 October, 2025
Published: 14 October, 2025

***Corresponding author**: Nathan Andrie Ama, Department of Agribusiness, Southern Leyte State University, Hinunangan, Philippines,
E-mail: nathanandrieama@gmail.com

Keywords: Health; Apps; Google play store; Machine learning; AI

https://www.engineegroup.us

Check for updates

Abstract

This study aims to evaluate the effectiveness of health applications on the Google Play Store by analyzing app metadata using machine learning classification models. It investigates whether application features—such as classification, app category, update status, and version—are associated with higher user ratings. A total of 305 health-related applications were selected from the Google Play Store using keyword filters for "Health & Fitness" and "Medical." Key metadata were extracted and pre-processed, including Classification (AI *vs.* Non-AI), Category, Reviews, Developer Type, Version, Release Year, and Recent Update. To address class imbalance, the SMOTE technique was applied, and three machine learning models—Naïve Bayes and K-Nearest Neighbors (KNN)—were used to predict user ratings. The KNN model achieved the most balanced performance with 75.89% accuracy, 82.22% precision, and an AUC of 0.849. Future research should consider larger and more diverse datasets and explore additional features (e.g., user sentiment from reviews, app permissions) to further improve model performance.

Abbreviations

KNN: K-Nearest Neighbors; SMOTE: Synthetic Minority Oversampling Technique

Introduction

In today's digital age, the pervasive influence of technology on nearly every aspect of our lives is undeniable. From the way we communicate and work to how we entertain ourselves, technology has revolutionized human behavior in profound ways [1]. As technology continues to advance at a rapid pace, individuals are becoming more reliant on digital devices and the internet for various aspects of their lives [2]. It has a profound influence of technology in domains such as healthcare, communication, and personal development [3]. Digital technology also revolutionized accessibility to mental health resources, providing avenues for support and intervention [4], the social sphere at the present stage [5], and even integrating digitalization in mental health [6].

Progressing into the Age of Digitalization, there have been unprecedented transformations ongoing in the world and humankind, through the drastic development of algorithms and big data, artificial intelligence, global telecommunication, and cyborgs. There has been progressive and extensive influence of digitalization in every aspect of daily living, including information processing, communication, infrastructure, logistics, finance and commerce, industry, economy, education, healthcare, and entertainment [7]. Nowadays, Digital technologies are dramatically changing healthcare [8]. Due to this significant advancement in the modern world, all people have now drastically switched to these platforms, particularly on the impact of digitalization on physical health and fitness [9]. Mobile health (mHealth) apps have gained significant popularity over the last few years due to their tremendous benefits, such as lowering health care costs and increasing patient awareness [10]. These applications offer the potential for dynamic engagement of patients and providers in health care and a new means of improving health outcomes [11]. The development of health and fitness applications allows users to

Citation: Ama NA. Machine Learning Based-prediction of Health Application Effectiveness on Google Play Store. Trends Comput Sci Inf Technol. 2025;10(3):081-088. Available from: https://dx.doi.org/10.17352/tcsit.000101

conveniently monitor, manage, and improve their overall well-being through digital tools. These applications offer features such as workout tracking, calorie counting, sleep analysis, and personalized fitness plans, which empower individuals to take a more active role in their health. For instance, fitness apps provide various feature sets to assist individuals' physical activity (e.g., running, cycling, working out, health planning, and trackers) for both men and women, allowing for easy access to data and information. Fitness apps typically refer to third-party mobile applications with built-in GPS, social networking capabilities (e.g., users share their exercise records on Facebook or Twitter), and sensor technologies that can help users record physical and physiological data automatically and generate personalized training profiles and schedules [12]. Moreover, Shaw [13] said that many fitness apps have now perfectly marketed themselves to both serve as a resource to use for on-demand fitness content, as well as provide personalized service and include the same type of hands-on dedicated approach one would receive if working directly with a personal trainer or gym class.

This study investigates the effectiveness of Health applications from Google Play Store Metadata with the use of a Machine Learning-Based Prediction model. It also provides a longitudinal study of Google Play app metadata, which will give unique information that is not available through the standard approach of capturing a single app snapshot [14]. Using feature extraction from app analysis, it will be used to find whether an app is effective or not based on user ratings.

Materials and methods

The methods in this study consist of 4 stages. The following are:

Data sourcing and cleaning

In the first stage, data sourcing and cleaning, application data were collected from the Google Play Store using the keywords "Health & Fitness" and "Medical" to filter relevant applications. Metadata, including application name, developer name, number of reviews, user ratings, release year, recent update, application version, and classification (AI or non-AI), were extracted for each app. A total of 234 Health & Fitness applications and 206 Medical applications were initially retrieved. Of these, 11 Health & Fitness apps and 97 Medical apps were excluded due to missing or incomplete data. Further screening identified 9 Health & Fitness apps and 18 Medical apps as irrelevant to the study objectives. As a result, the final dataset comprised 214 Health & Fitness applications and 91 Medical applications, which were included in the subsequent analyses. After data sourcing, basic data cleaning, and data visualization are performed [Figure 1].

Data visualization

The second stage is data visualization, where several categorical variables were transformed into binary-coded formats to enable statistical and machine learning analysis. Developer type, Number of reviews, release year, recent update, application version classification (*AI and Non-AI*), and category (*Health & Fitness, Medical*) are coded with 2-3 binaries, while variable User ratings are also coded (*1=High, 2=Low*).

Synthetic minority over-sampling technique

The third stage is applying SMOTE, since datasets in this study are unbalanced and can lead to biased models that perform well on the majority class but poorly on the minority class. SMOTE (*Synthetic Minority Over-sampling Technique*) is a powerful technique to handle unbalanced datasets of this study, which consists of only 305 collected applications in the Google Play Store. It works by creating synthetic examples for the minority class by interpolating between existing minority instances. This helps in achieving a balanced class distribution without simply duplicating the minority instances. In a 305 collected data from Google Play Store, there are 284 High

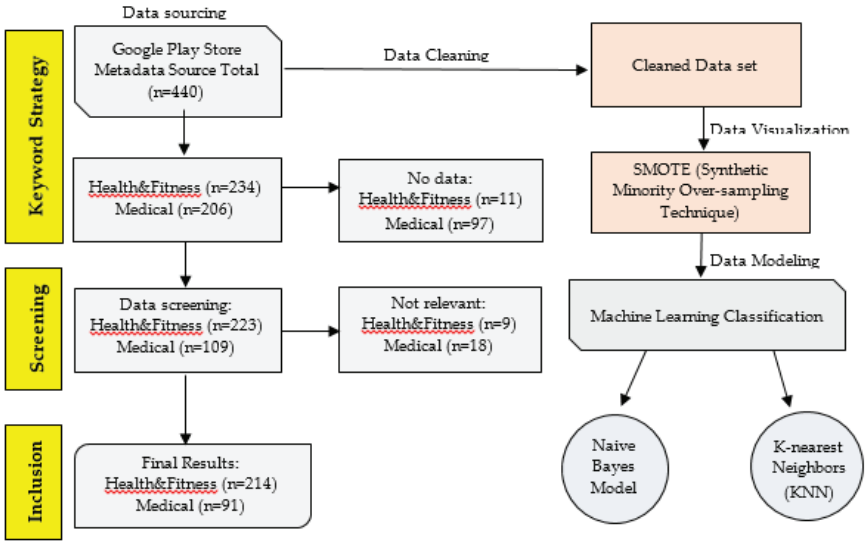

Figure 1: Flowchart of the study's methodology.

Citation: Ama NA. Machine Learning Based-prediction of Health Application Effectiveness on Google Play Store. Trends Comput Sci Inf Technol. 2025;10(3):081-088. Available from: https://dx.doi.org/10.17352/tcsit.000101

Ratings while 21 Low ratings, but since an imbalance can lead to biased models. An application of SMOTE is needed, where it creates 263 synthetic data points to complete low ratings, consisting of an overall 284 low-rated apps. Both High and Low ratings are now both 284, and all data used consists of 568 application data.

Data modeling (machine learning)

Lastly, in the fourth stage, data was analyzed using selected machine learning classification such as Naive Bayes Model and K-nearest neighbors Model Classification, to compare model performance with the highest accuracy percentage and other results. The model demonstrating the highest accuracy and most favorable outcomes will be retained for the experimental analysis. Descriptive statistics for the coded variables are presented first. RStudio v 4.5.1 software was used for data analysis.

Naive bayes classification: We want to predict the user rating R, which derives from a categorical variable with two possible classes: $R \in \{1, 2\}$ (1 = High Ratings, 2 = Low Ratings). We use several observed features to make the prediction.

- x_1 = Classification
- x_2 = Category
- x_3 = Developer and so on.

The general form of Bayes' Theorem is represented below:

$$P(R=r|x_1,x_2,\ldots,x_7)=\frac{P(R=r)\cdot P(x_1,x_2,\ldots,x_7|R=r)}{P(x_1,x_2,\ldots,x_7)}$$

The denominator is the same for all classes, so for classification purposes, ignore it. Then Apply the Naive Assumption. Naive Bayes assumes that all features are conditionally independent given the class. This allows us to break down the joint probability. Formula below:

$$P(x_1,x_2,\ldots,x_7|R=r)=\prod_{i=1}^{7}P(x_i|\ R=r)$$

Substitute this into Bayes' rule:

$$P(R=r)|x_1,x_2,\ldots,x_7)\propto P(R=r)\cdot\prod_{i=1}^{7}P(x_i|R=r)$$

Then define the Classification rule. To make a prediction, we compare the probability score for each class r {1, 2}, and choose the class with the highest value. Formula below.

$$R=arg_{r\in\{1,2\}}^{max}\left[P(R=r)\cdot\prod_{i=1}^{7}P(x_i|\ R=r)\right]$$

where:

R = The predicted class (user rating)

arg *max* = "Choose the class r that gives the maximum result

$r \in \{1, 2\}$= Possible class labels (1 = High, 2 = Low)

$P(R = r)$ = Prior probability of class r, how common this rating is in your dataset

$P(x_i \mid R = r)$= Conditional probability of feature x_i given the class $R = r$

$\prod_{i=1}^{7}$ = Multiply all the conditional probabilities from x_1 to x_7

Compute scores for each class. To classify a new app, plug in observed feature values and compute the score for each rating class:

$$Score(r)=P(R=r)\cdot P(x_1|R=r)\cdot P(x_2|R=r)\cdot\ldots\cdot P(x_7|R=r)$$

Do this for all $r \in \{1, 2\}$, and select the class with the highest score, and that is the predicted user rating. The naive Bayesian model is easy to build, with no complicated iterative parameter estimation, which makes it particularly useful for very large datasets. Despite its simplicity, the Naive Bayesian classifier often does surprisingly well and is widely used because it often outperforms more sophisticated classification methods [15].

K-nearest Neighbors Model (KNN): K-Nearest Neighbors (KNN) are the distances between the test and input data are measured and sorted to find the k nearest neighbors. Then, the majority voting is performed to determine the category of data by selecting the most common vote among the nearest neighbors [16]. The concept of K-nearest neighbors is illustrated in [Figure 2]. KNN classification is used to determine the average accuracy (predicted percentage) of a new data point (estimated rating level of health apps), which serves as an indicator of the effectiveness of using the applications. By using k-fold cross-validation in this study it provides a more reliable estimate of a model's performance by using the entire dataset for both training and validation (test set), reducing bias and variance associated with single train/test splits. The dataset is divided using 10-fold cross-validation, which involves splitting the data into 10 equal-sized subsets, each containing an equal number of samples [Figure 3]. In each iteration, one fold is assigned as the test set while the remaining nine folds serve as the training set, and this process is repeated across all folds (blue segments in Figure 3). For each iteration with one test fold, the distance between each data point is calculated using

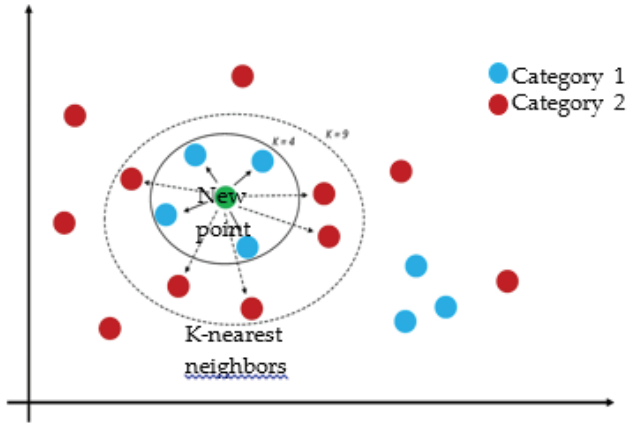

Figure 2: Explanation of k-nearest neighbors.

Citation: Ama NA. Machine Learning Based-prediction of Health Application Effectiveness on Google Play Store. Trends Comput Sci Inf Technol. 2025;10(3):081-088. Available from: https://dx.doi.org/10.17352/tcsit.000101

Euclidean distance, as defined by the formula and illustrated in Figure 4.

Since there are multiple variables in this study, the formula becomes:

$$D = \sqrt{(a_2 + a_1)^2 + (b_2 - b_1)^2 + ...(g_2 - g_1)^2}$$

Distances are computed for all rows within each fold, and the process is repeated across all folds. After calculating the Euclidean distance for each row, the values are sorted in descending order to identify the k nearest neighbors. In this study, k is set to 3, selecting the three highest distances. A majority voting method is then applied by recording the ratings of these three neighbors and predicting the most frequent rating. This process is repeated for each row. The predicted ratings are then compared to the actual ratings; cases where the predicted and actual ratings match are considered correct. The overall accuracy is calculated by counting the number of correct predictions (i.e., matched predicted and actual ratings) and applying the standard accuracy formula below.

$$Accuracy_{Fold\,1} = \frac{\text{Number of correct predictions}}{\text{Total samples in Fold}_i}$$

In this study, the total number of folds is set to 10. The same process is then repeated for all remaining folds, ensuring that the accuracy is recorded for each of the 10 folds. Once all accuracy values are obtained, their average is computed to determine the overall model accuracy.

$$Mean\,accuracy = \frac{1}{10}\sum_{i=1}^{10} Accuracy_i$$

Where:

$\frac{1}{10}$ = Sum of all 10 accuracies divided by 10 folds

$\sum_{i=1}^{10}$ = Summation of accuracies from fold 1 to fold 10

$Accuracy_i$ = Accuracy from the i-th fold

Furthermore, k in KNN is a critical hyperparameter that you adjust based on your dataset's specific characteristics. The optimal value of k is essential for the accuracy of the algorithm's predictions. A smaller k value can make the algorithm sensitive to noise and overly flexible, whereas a larger k value can render it computationally intensive and prone to underfitting. An odd number of k is often chosen to avoid ties in classification.

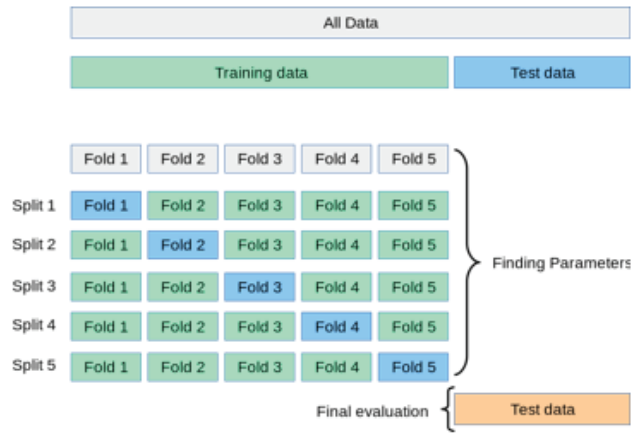

Figure 3: K-fold cross-validation.

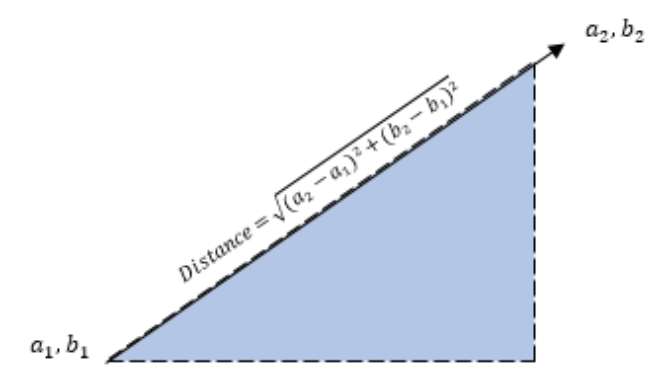

Figure 4: Explanation of Euclidean Distance.

Results

This section presents the results of the statistical analyses conducted to evaluate the effectiveness of health applications. It begins with descriptive statistics summarizing key features of the apps. Subsequently, the performance of each classification model—namely K-Nearest Neighbors and Naive Bayes—is assessed using confusion matrices, ROC curves, and key performance metrics, including accuracy, precision, recall, F1 score, AUC, etc.

Distribution of application features

The distribution of the variables is presented in Table 1, revealing that the majority were AI-based (*n* 328, 57.7%) and primarily categorized under Health & Fitness (*n* = 360, 63.4%). Some effective health & fitness applications filtered in this study are mostly developed by Leap Fitness Group. According to [17], AI algorithms can predictively project individual choices, preferences, geographic behaviors, and patterns by analyzing user data. This enables mobile apps to deliver truly personalized, tailored content, recommendations, and notifications, creating a more engaging and personalized user experience. Furthermore, [14] states that Fitness apps provide various feature sets to assist individuals' physical activity (e.g., running, cycling, working out, and swimming). For example, the data management feature set allows users to collect and manage their exercisers' data, such as recording their steps, running routes, calories burned, and heart rate. A considerable proportion of applications had low user reviews (*n* = 411, 72.4%), indicating limited user engagement or relatively new releases. In terms of effectiveness, applications were evenly distributed between those with high ratings (*n* = 284, 50%) and low ratings (*n* = 284, 50%), justifying the binary outcome modeling in subsequent machine learning analysis.

Furthermore, most apps were developed by small developers (*n* = 530, 93.3%), which may reflect the increasing participation

Citation: Ama NA. Machine Learning Based-prediction of Health Application Effectiveness on Google Play Store. Trends Comput Sci Inf Technol. 2025;10(3):081-088. Available from: https://dx.doi.org/10.17352/tcsit.000101

Table 1: Distribution of Application Features.

Variables		Frequency	Percentage
Classification	AI	328	57.7
	Non-AI	240	42.2
Category	Health&Fitness	360	63.4
	Medical	208	36.6
Reviews	High	42	7.4
	Medium	115	20.2
	Low	411	72.4
Ratings	High Ratings	284	50
	Low Ratings	284	50
Developer	Big Developer	38	6.7
	Small Developer	530	93.3
Recent Update	Old	88	15.5
	Recent	480	84.5
Version	High	43	7.6
	Medium	150	26.4
	Old	375	66.0
Release Year	Old	276	48.6
	Mid	197	34.7
	Recent	95	16.7

of independent developers in the health app market. Recently updated apps comprised the majority (n = 480, 84.5%), showing that developers actively maintain and improve their applications. Regarding versioning, older versions were most common (n = 375, 66.0%), possibly due to compatibility or maintenance constraints. Almost half of the applications were released earlier (n = 276, 48.6%), indicating a longer presence on the market.

Health application effectiveness

A Naive Bayes classifier was applied to evaluate the effectiveness of health applications based on user ratings. [Table 2]. The model achieved an overall accuracy of 57.06%, with a 95% confidence interval of [49.26%, 64.61%]. However, it did not significantly outperform the No Information Rate of 92.94% (p = 1.00), indicating that the model did not perform better than simply predicting the majority class. The agreement between predicted and actual outcomes was weak, with a Cohen's Kappa of (κ = 0.14) suggesting only slight reliability beyond chance. This also means that the Naïve Bayes model performs poorly in predicting highly rated applications.

The confusion matrix revealed that the model successfully identified all apps that were actually rated highly by users, resulting in a recall of 1.000 (100%). However, it also incorrectly classified 73 low-rated apps as highly rated, producing a low precision of 14.12% [Table 3]. In other words, while the model was sensitive to identifying effective apps, most of its predictions of "high rating" were incorrect. The combined effect of high recall and low precision led to an F1 score of 0.25, indicating a weak overall balance between correctly identifying and over-predicting highly rated apps. The model's specificity was 53.80%, reflecting limited ability to correctly identify low-rated apps. The balanced accuracy, averaging performance across both classes, was 76.90%.

Crucially, McNemar's Test was highly significant (p < .001), confirming that the model's misclassifications were not random. Specifically, the model produced many more false positives (73) than false negatives (0), suggesting a strong bias toward predicting high ratings, even when apps were not actually rated highly. In summary, although the Naive Bayes model demonstrated perfect sensitivity in detecting highly rated apps, its very low precision and classification imbalance limit its practical usefulness. The tendency to over-predict effectiveness makes it unsuitable for applications where recommending low-quality health apps must be avoided.

Additionally, the K-Nearest Neighbors (KNN) classification model was also employed to evaluate the effectiveness of health applications based on user ratings [Table 4]. The model achieved an overall accuracy of 75.89%, with a 95% confidence interval of [66.9%, 83.47%], significantly higher than the No Information Rate of 50% (p < .001). The Cohen's Kappa coefficient (κ = 0.52) indicated a moderate agreement between predicted and actual class labels.

The confusion matrix showed that the model correctly classified 37 highly rated apps (true positives) and 48 lower-rated apps (true negatives), while misclassifying 19 highly rated apps (false negatives) and 8 lower-rated apps (false positives) [Table 5]. The model yielded a recall (sensitivity) of 66.07%, meaning it correctly identified two-thirds of truly effective apps. The precision (positive predictive value) was 82.22%, indicating that most apps predicted to be highly rated were indeed so. These values resulted in an F1 score of 0.7326, reflecting a strong balance between recall and precision. The specificity was 85.71%, and the negative predictive value was 71.64%, suggesting reliable identification of both effective and ineffective apps. The balanced accuracy was equal to overall accuracy (75.89%), reinforcing the model's robustness in handling the two classes.

Table 2: Model performance metrics for Naïve Bayes model.

Metric	Value Test	Significance
Accuracy	0.5706	-
95% CI	(0.4926, 0.6461)	-
No Information Rate (NIR)	0.9294	-
p - Value [Acc > NIR]	p = 1.00	Not significant
Kappa	0.1412	-
Sensitivity	1.000	-
Specificity	0.53797	-
F1 Score	0.2474	-
Recall	1.000	-
Precision	0.1412	-
Balanced Accuracy	0.7690	-
Positive Predicted Class	1 (High Rating)	-
Mcnemar's p - Value	p < .001	Highly significant

Note: p < .001 Highly Significant, p <.05 Significant, p > 0.5 Not significant.

Table 3: Confusion Matrix for Naïve Bayes Model.

	Predicted: (1)	Predicted: (2)
Actual: (1)	12	73
Actual: (2)	0	85

Citation: Ama NA. Machine Learning Based-prediction of Health Application Effectiveness on Google Play Store. Trends Comput Sci Inf Technol. 2025;10(3):081-088. Available from: https://dx.doi.org/10.17352/tcsit.000101

Table 4: Model performance metrics for K-nearest Neighbors.

Metric	Value Test	Significance
Accuracy	0.7589	-
95% CI	(0.669, 0.8347)	-
No Information Rate (NIR)	0.5	-
p - Value [Acc > NIR]	*p* < .001	Highly significant
Kappa	0.5179	-
Sensitivity	0.6607	-
Specificity	0.8571	-
F1 Score	0.7326	-
Recall	0.6607	-
Precision	0.8222	-
Balanced Accuracy	0.7589	-
Positive Predicted Class	1 (High Rating)	-
Mcnemar's *p* - Value	*p* = 0.0543	Not significant

Note: *p* < .001 Highly Significant, *p* < .05 Significant, *p* > 0.5 Not significant.

Table 5: Confusion Matrix for KNN Model.

	Predicted: (1)	Predicted: (2)
Actual: (1)	37	8
Actual: (2)	19	48

Although McNemar's Test approached significance (*p* = .0543), it did not reach the conventional alpha threshold (*p* < .05), indicating that the difference in misclassification between false positives and false negatives was not statistically significant. Therefore, the model does not exhibit a strong bias toward one type of misclassification over the other. These findings suggest that the KNN model can effectively classify health applications based on user ratings, offering both sensitivity in detecting highly rated apps and precision in ensuring that positive predictions made by the model are accurate.

Performance metric of the 3 classification models in predicting highly effective applications

Three machine learning models were evaluated in predicting whether health-related mobile applications were perceived by users as highly effective (positive class = 1) or not (class = 2), as shown in Table 6. The K-Nearest Neighbors (KNN) model performed best overall, with an accuracy of 75.89%, high precision (82.22%), and balanced sensitivity (66.07%) and specificity (85.71%). Its F1 Score was 73.26%, and the AUC of 0.849 indicated excellent discriminative ability. The Naïve Bayes model, while achieving perfect recall (100%) for identifying highly effective apps, had very low precision (14.12%), resulting in an F1 Score of 24.74%. This suggests that it overclassifies apps as highly effective, yielding many false positives (Figure 5).

Therefore, the K-nearest Neighbors (KNN) Classification Model demonstrated the most reliable and balanced performance in identifying health apps rated highly (1) by users, making it the most suitable model for this classification task. A study of [3] analyzed largely positive reviews (1=high rated), with 6700 reviews (6700/7929, 84.50%) giving the app a 5-star rating and 2676 reviews (2676/7929, 33.75%) explicitly terming the app "helpful" or that it "helped." Of 7929 reviews, 251 (3.17%) had a less than 3-star rating and were termed as negative reviews for AI health apps. For instance, highly rated health and fitness apps such as MyFitnessPal (Android build 25.26.0) released on July 2, 2025, hit 4.7 (high) ratings with over 2,751,560 downloads on Google Play Store. Tim Holley [18], Chief Product Officer at MyFitnessPal said that "The 2025 Winter Release underscores MyFitnessPal's commitment to supporting our members as they advance the way they approach nutrition and habit development", she added on the post "Integrating tools like Voice Log and Weekly habits, gives members effective solutions to streamline tracking, while reinforcing the importance of progress over perfection in building lasting habits—because true success in nutrition comes from consistency, not perfection.".

Table 6: Performance metrics of three classification models in predicting highly effective health applications (positive class = 1).

Models	Accuracy	Precision	Recall (Sensitivity)	Specificity	FI Score	AUC	Positive Class
Naïve Bayes Model	57.06%	14.12%	100%	53.80%	24.74%	0.669	1 (High)
K-nearest Neighbors (KNN)	75.89%	82.22%	66.07%	85.71%	73.26%	0.849	1 (High)

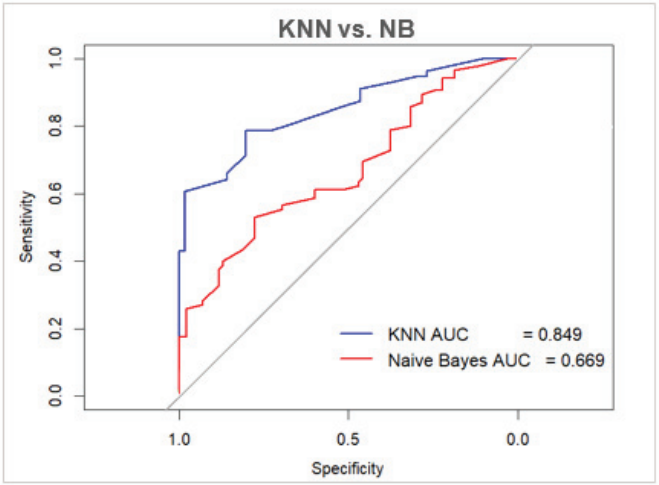

Figure 5: ROC Curves Comparing KNN and Naive Bayes Models in Predicting Health App Effectiveness (User Ratings: 1 = High, 2 = Low).

Furthermore, other survey of [19] for top 20 trending Health & Fitness apps on Google Play as of July 9, 2025, apps like HealthifyMe, Replika, Catzy, and others are currently trending—with user ratings ranging from 4.2 to 4.8 stars, indicating both active use and high satisfaction, This demonstrates that recently updated health apps on Google Play are indeed highly rated, reinforcing the trend that top-performing health apps combine frequent maintenance with strong user approval. Studies of [20] also cited some best and effective health and fitness applications to help you train at home, some are Centr, Nike Training Club, Fiit, Apple Fitness Plus, Sweat, Body Coach, Strava, Home Workout No equipment is among the best fitness applications.

Citation: Ama NA. Machine Learning Based-prediction of Health Application Effectiveness on Google Play Store. Trends Comput Sci Inf Technol. 2025;10(3):081-088. Available from: https://dx.doi.org/10.17352/tcsit.000101

Discussion

This study employed a novel combination of machine learning statistical techniques, including the Naïve Bayes Classifier and K-Nearest Neighbors (KNN). This hybrid methodology enables the analysis and classification of Google Play Store metadata, offering a multidimensional perspective on the effectiveness of healthcare applications based on user ratings.

While this approach provides valuable insights, the researcher acknowledges several limitations that affect the comprehensiveness of the findings. One key limitation is the limited dataset diversity and size—the analysis included only 305 applications, which may not fully represent the wide range of health apps available on the Google Play Store. This constraint potentially limits the generalizability of the findings, especially given the rapid growth and diversity of mobile health applications.

Another challenge is class imbalance and the data cleaning process, which involved the exclusion of entries due to missing or irrelevant data—this may have introduced bias. To address data imbalance for ratings, the study employed the Synthetic Minority Oversampling Technique (SMOTE), which, while effective, can sometimes result in overfitting or generating less realistic representations.

Model performance is also a noted limitation. The models demonstrated moderate predictive performance, averaging around 75%. Notably, the Naïve Bayes classifier, despite achieving high recall, performed poorly overall, suggesting that the current set of features may not adequately capture the determinants of app effectiveness. A promising direction for future research for this study is utilizing a larger and more diverse dataset, coupled with advanced deep learning methodologies, to improve model accuracy and uncover additional predictors of healthcare app effectiveness.

Conclusion

This study successfully demonstrated the predictive capability of machine learning models in evaluating the effectiveness of health applications on the Google Play Store using metadata features such as Classification (AI *vs.* Non-AI), App Category, Developer Type, Version, Reviews, Release Year, and Recent Update. Among the three models tested—Naïve Bayes and K-Nearest Neighbors (KNN)—the KNN model emerged as the most balanced and robust performer with an overall accuracy of 75.89%, strong precision (82.22%), and reliable sensitivity (66.07%). It offered the highest AUC score (0.849), indicating excellent discriminative ability in distinguishing highly rated health apps from low-rated ones. The Naïve Bayes model, while achieving perfect recall (100%), suffered from very low precision (14.12%) and produced many false positives, limiting its utility in real-world applications. Future studies should consider larger and more diverse datasets and explore additional features (e.g., user sentiment from reviews, app permissions) to further improve model performance.

Data availability statement

The data that support the findings of this study are confidential and are not publicly available due to privacy restrictions. Access to the dataset may be granted upon reasonable requests and with permission from the corresponding data owner.

Acknowledgment

The authors thank the reviewers for their comments, which helped to improve the presentation of this manuscript.

Disclaimer/publisher's note

References

1. Baxromovich AD. Impact of technology on physical activity. 2024;2:4. Available from: https://westerneuropeanstudies.com/index.php/4/article/view/707

2. Saedsayad. Naïve Bayesian. 2018. Available from: http://www.saedsayad.com/naive_bayesian.htm

3. McGuire J. We've tested the best workout apps of 2025 to help you train at home. 2025. Available from: https://www.tomsguide.com/best-picks/best-workout-apps

4. Noor NM, Chin YW, Yusoff NH. Unveiling the impact of technological progress on societal advancement: A scholarly analysis of family well-being through the lens of millennial women. Int J Acad Res Bus Soc Sci. 2023;13(7). Available from: https://doi.org/10.6007/ijarbss/v13-i7/17448

5. Raj G, Sharma AK, Arora Y. Analyzing the effect of digital technology on mental health. In: Advances in Web Technologies and Engineering. IGI Global; 2024;54–82. Available from: https://doi.org/10.4018/979-8-3693-6557-1.ch003

6. Balcombe L, De Leo D. Digital mental health challenges and the horizon ahead for solutions. JMIR Ment Health. 2021;8(3):e26811. Available from: https://doi.org/10.2196/26811

7. Chan KT. Emergence of the 'Digitalized Self' in the age of digitalization. Comput Hum Behav Rep. 2022;6:100191. Available from: https://doi.org/10.1016/j.chbr.2022.100191

8. Glauner P, Plugmann P, Lerzynski G. Digitalization in healthcare. Springer; 2021. Available from: https://link.springer.com/book/10.1007/978-3-030-65896-0

9. Shaw B. Why Fitness Apps Have Become So Popular. 2021. Available from: https://sustainhealth.fit/lifestyle/why-fitness-apps-have-become-so-popular/

10. Aljedaani B, Babar MA. Challenges with developing secure mobile health applications: Systematic review. JMIR Mhealth Uhealth. 2021;9(6):e15654. Available from: https://doi.org/10.2196/15654

11. Romanova TF, Klimuk VV, Andreeva OV, Sukhoveeva AA, Otrishko MO. Digitalization is an urgent trend in the development of the social sphere. In: Lecture Notes in Networks and Systems. 2019;931–939. Available from: https://doi.org/10.1007/978-3-030-29586-8_106

Citation: Ama NA. Machine Learning Based-prediction of Health Application Effectiveness on Google Play Store. Trends Comput Sci Inf Technol. 2025;10(3):081-088. Available from: https://dx.doi.org/10.17352/tcsit.000101

12. Hu J, He W, Zhang J, Song J. Examining the impacts of fitness app features on user well-being. Inform Manage. 2023;60(5):103796. Available from: https://doi.org/10.1016/j.im.2023.103796

13. Sandua D. The double sides of technology: Internet addiction and its impact on today's world. 2024. Available from: https://books.google.com.ph/books?hl=en&lr=&id=uS78EAAAQBAJ

14. Maredia R. An analysis of the Google Play Store dataset and predict the popularity of an app. 2020. Available from: https://www.researchgate.net/publication/343769728_Analysis_of_Google_Play_Store_Data_set_and_predict_the_popularity_of_an_app_on_Google_Play_Store

15. Sama PR, Eapen ZJ, Weinfurt KP, Shah BR, Schulman KA. An evaluation of mobile health application tools. JMIR Mhealth Uhealth. 2014;2(2):e19. Available from: https://doi.org/10.2196/mhealth.3088

16. Alshammari AF. Implementation of Classification using K-Nearest Neighbors (KNN) in Python. 2024;186:33. Available from: https://doi.org/10.5120/ijca2024923894

17. Feld J. How AI Is Transforming Fitness Apps. 2024. Available from: https://www.healthandfitness.org/improve-your-club/how-ai-is-transforming-fitness-apps/

18. Holley T. MyFitnessPal Unveils Its 2025 Winter Release. 2025. Available from: https://www.prnewswire.com/news-releases/myfitnesspal-unveils-its-2025-winter-release-302385598.html

19. Appbrains. The top 20 are trending Health & Fitness Apps for Android right now. 2025. Available from: https://www.appbrain.com/apps/trending/health-and-fitness

20. Malik T, Ambrose AJ, Sinha C. Evaluating user feedback for an Artificial Intelligence–Enabled, Cognitive Behavioral Therapy–Based Mental Health app (WYSA): Qualitative Thematic analysis. JMIR Hum Factors. 2022;9(2):e35668. Available from: https://doi.org/10.2196/35668

Citation: Ama NA. Machine Learning Based-prediction of Health Application Effectiveness on Google Play Store. Trends Comput Sci Inf Technol. 2025;10(3):081-088. Available from: https://dx.doi.org/10.17352/tcsit.000101