Received: 14 September 2025 | Revised: 12 November 2025 | Accepted: 11 December 2025 | Published online: 4 January 2026

RESEARCH ARTICLE

FinTech and Sustainable Innovation
2026, Vol. 00(00) 1–14
DOI: 10.47852/bonviewFSI62027666

BON VIEW PUBLISHING

A Crowdfunding Campaign Success Prediction Scheme Using Machine Learning

Suzab-Ud Doula[1] and Mahfuzulhoq Chowdhury[2,*]

[1] *Computer Science and Engineering Department, Chittagong University of Engineering and Technology, Bangladesh*

[2] *Computer Science and Engineering Department, Chittagong University of Engineering and Technology, Bangladesh*

Abstract: Using internet platforms, crowdfunding is a popular method used by entrepreneurs and innovators to raise funds for their initiatives. The success of a crowdfunding campaign depends on many variables, making it challenging to predict how it will play out. Prior research on predicting crowdfunding success had a number of drawbacks, such as the use of incomplete, noisy, and skewed datasets, a deficiency of features, imprecise prediction outcomes, and incompatibility with various campaign kinds or platforms. After examining a sizable dataset, the proposed machine learning (ML)-based crowdfunding campaign success prediction technique presented in this paper is able to predict crowdfunding campaign success results with higher accuracy. Several ML techniques, such as decision trees, logistic regression, support vector machines, K-nearest neighbors, and random forests, were employed in this study to identify the most effective prediction model. Random forest classifiers showed better results than other ML models, according to the resultant data, with 96.42% accuracy, 94.87% precision, 97.36% recall, and 96.98% F1-score. The proposed crowdfunding campaign success prediction model beats existing works by at least 3.1% in accuracy, 2.6% in precision, and 2% in recall score value, according to the performance comparison result. Additionally, this paper offers a mobile application for crowdfunding campaign assistance that has features like police reporting capabilities, donation options, suggestion options, and crowdfunding success prediction. According to the customer feedback results, over 80% of the evaluators were satisfied with the features of our program.

Keywords: machine learning, crowdfunding campaign, finance success prediction, mobile application, random forest

1. Introduction

Modern finance techniques, such as crowdfunding, have recently gained popularity. To support a specific initiative or business, financial contributions are solicited from a large number of contributors or people by using internet platforms [1]. Crowdfunding has significant benefits for both project developers and investors. It provides creators with an alternative source of funding that can help them realize their ideas, confirm there is a market for them, and gain recognition. This allows them to gain access to a large global network of investors and leverage their combined resources. On the other hand, investors have the opportunity to promote innovation, contribute to causes they believe in, and potentially profit financially. While crowdfunding has many benefits, there are some drawbacks, such as the difficulty of predicting campaign success and detecting fraud.

Due to several risk factors, both project developers and investors rely on the success prediction of a crowdfunding campaign [2]. Project developers can improve their plans and raise their funding chances by identifying the factors that influence campaign success. Investors can use success prediction models to assess the potential of a campaign before making an investment decision. To evaluate past campaign data, success prediction models frequently employ a variety of factors and machine learning (ML) methods. Project details such as the funding target, duration, category, and project description are critical components for predicting crowdfunding campaign success [3]. Social factors such as the total number of backers, interaction metrics, and updates may also be considered in the campaign success prediction. Predictive analytics is a tool for forecasting the performance of upcoming campaigns by training models on the outcomes of previous initiatives. On the other hand, detecting fraudulent campaigns is an important part of maintaining the credibility and integrity of crowdfunding platforms. Fraudulent campaigns include invented ventures, improperly handled money, and failure to deliver on promised prizes. Fraud prevention and detection ensure the legitimacy of the crowdfunding ecosystem while protecting investors' interests. Fraud detection methods use a variety of strategies, including rule-based systems, ML algorithms, and data analysis [4–7]. Rule-based systems detect irregular patterns or actions by applying specific criteria and rules. ML techniques use historical data to train models that can distinguish between legitimate and fraudulent campaigns based on project characteristics, user behavior, financial data, and external signals.

The majority of recent research has attempted to determine how the characteristics of the project's founders influence the project's qualities and success [8]. Only a few studies have focused on how the project's description and wording influence predictions

***Corresponding author:** Mahfuzulhoq Chowdhury, Computer Science and Engineering Department, Chittagong University of Engineering and Technology, Bangladesh. Email: mahfuz@cuet.ac.bd

© The Author(s) 2026. Published by BON VIEW PUBLISHING PTE. LTD. This is an open access article under the CC BY License (https://creativecommons.org/licenses/by/4.0/).

of fundraising success and fraud [9]. Previous research works [6, 10–20] did not look into several parameters for predicting crowdfunding campaign success, such as project duration, target funding goals, number of backers, backer feedback, previous campaign performance, project credibility, funding sources, and advertising expenditures. The existing works did not use proper feature selection or hyperparameter tuning techniques to select the best ML model. The majority of existing works suffer from issues such as low accuracy, an imbalanced dataset, a lack of multi-platform or multiple types of campaign support facilities, and a lack of credibility [18, 20]. Existing work did not include the development of a mobile application to assist with crowdfunding campaign success prediction, funding options, suggestions, and police reporting. To address these existing issues, the primary goals of this paper are to provide comprehensive models that not only predict campaign success but also identify potential fraud by combining ML classification techniques. New crowdfunding campaign success is predicted by the suggested model prior to the start of the campaign. Before the new campaign launches, the number of backers and their feedback from previous projects are gathered. The list of significant contributions regarding this article is provided below:

i) This paper introduces an ML-based crowdfunding campaign success prediction model that employs appropriate techniques by examining multiple ML algorithms including the random forest (RF) algorithm, K-nearest neighbor (KNN) algorithm, decision tree (DT) algorithm, SVM (support vector machine) algorithm, and logistic regression (LR) algorithm. This paper also created a dataset for predicting the success of crowdfunding campaigns using a questionnaire-based response collection system. Our dataset includes variables such as project duration, target funding goals, number of backers, backer feedback, previous campaign performance, project credibility, funding sources, and advertising expenses.
ii) To improve accuracy, the proposed prediction model used data preprocessing, handling missing values, level encoding, feature extraction, hyperparameter tuning, normalization, confusion matrix-based examination, and the k-fold cross-validation technique. This paper introduces several new features for forecasting crowdfunding campaign success, such as average backing amounts and supporter growth rates. This paper compared the proposed scheme's results to those of various ML algorithms and existing works using metrics such as accuracy measure, precision measure, recall measure, and F1-score values.
iii) This paper created a crowdfunding campaign assistance mobile application for both investors and contributors that includes features such as crowdfunding campaign success prediction, donation options, suggestions, and police reporting capabilities. This paper also visualizes the app evaluation results by analyzing various features and soliciting user feedback.

Section 2 includes a brief summary of current research on predicting crowdfunding success. Section 3 describes the proposed ML-based crowdfunding success prediction model, including a methodology diagram, in-depth discussion, model selection, and analysis findings. Section 4 describes the crowdfunding campaign assistance mobile application features, including detailed working steps. Section 5 delivers the app evaluation results. Section 6 addresses the proposed work's conclusion, limitations, and future challenges.

2. Literature Review

This section examines related research on crowdfunding campaign success prediction and assistance systems for crowdfunding campaigns. This section also highlights the research gap between proposed and existing work. In the article Yuan et al. [10], the authors demonstrated how textual analysis can boost crowdfunding campaign success and assist project creators in writing better descriptions. They rely heavily on project descriptions for crowdfunding campaign success rather than on other critical factors. They extracted features from texts using domain-constrained latent Dirichlet allocation, also known as the DC-LDA technique. Their work is limited by the fact that it relies on incomplete or inconsistent data from crowdfunding platforms. It would be difficult to analyze text containing slang or typos. Furthermore, due to dataset and factor limitations, their proposed system will not be applicable to other platforms or campaign types. Furthermore, the complexity of semantic models can make interpretation difficult, and changing crowdfunding trends may affect the long-term relevance of the findings. In Kim et al. [11], the authors discussed the impact of founder qualities and project aspects on crowdfunding campaign outcome assessment. They emphasize the importance of managing crowdfunding campaigns while considering both the creator and the project. Their research emphasizes the importance of project developers maintaining a strong social network in order to increase the chances of a successful crowdfunding campaign. The authors emphasized the importance of choosing project elements that appeal to the crowdfunding audience's preferences and interests. However, they did not work on feature analysis or the application of ML to success prediction. In Perez et al. [12], the authors created an ML-based fraud prediction system for crowdfunding campaigns. Their research focused on how text analytics can improve crowdfunding campaigns and assist authors in writing compelling descriptions. They used a multilayer perceptron or MLP classifier to predict fraud. Their proposed model achieves an accuracy rate of 90%, which is significantly lower. They did not provide any mechanism to determine the reason for the successful crowdfunding campaign. In Siering et al. [13], the authors used linguistic and content-based indicators in campaign descriptions, as well as backer comments, to identify fraudulent crowdfunding activity. Their work emphasized the importance of detecting fraud in order to protect contributors and preserve the ecosystem's integrity. They did not address the issues of hyperparameter tuning, dataset credibility, or appropriate feature selection. As a result, their method for detecting fraud is less accurate and reliable than others. They also did not use any ML-based prediction systems. Furthermore, they did not look into the important feature selection for a successful crowdfunding campaign. In Koch et al. [14], the authors looked into the key metrics that influence the outcome of crowdfunding campaigns. They used a large dataset to examine several characteristics for estimating crowdfunding success, including campaign duration, funding goals, project category, description, and the use of video or graphics. However, biases in self-reported data may limit the usefulness of their proposed crowdfunding campaign success prediction system. They did not compare multiple ML classifiers to predict the success of a crowdfunding campaign. They also did not provide any results for the best model selection, such as the accuracy and F1-score values. Sonare et al. [16] used blockchain technology for a crowdfunding platform. Zhou et al. [5] used the fuzzy deep learning method for scientific crowdfunding project identification. Lee et al. [6] investigated the impact of feature selection techniques on crowdfunding fraud prediction. In Zhou et al. [15], the authors focused on developing an automated approach for detecting deception in text-based online communications. They also did not use any ML or Artificial Intelligence (AI) algorithms to predict the outcome of the crowdfunding campaign. Dousios et al. [4] investigated the impact of text and audio features on the success of technology crowdfunding campaigns.

In Zhong et al. [17], the authors investigated the impact of the project founder's online activities on the success of crowdfunding campaigns. They used several factors to predict crowdfunding campaign success, including project goals, duration, rewards, comments, and videos. Their proposed method has an accuracy of only 80%, which is significantly lower. The work in Tang et al. [21] described a deep learning-based attention mechanism for predicting crowdfunding campaign success. They used video and textual data to investigate the outcome of the crowdfunding campaign. Their proposed model has an accuracy of only 72%, which is significantly lower. They also did not create a mobile application-based crowdfunding assistance system for their users. The authors of Oduro et al. [22] used bagging, boosting, and tree-based mechanisms to forecast crowdfunding campaign success. They examined the crowdfunding campaign's success while taking into account various business sectors and states. They investigated the effect of backer count and project duration on crowdfunding campaign success prediction. The article in Yeh et al. [23] looked into the impact of ESG (environmental, social, and governance) data on crowdfunding campaign success prediction. However, they did not create an ML-based prediction model for crowdfunding campaign success and did not present any prediction results. The work in Patil et al. [24] created a Flutter-based mobile app for startup investment facilities. They did not, however, provide features such as ML-based success prediction for crowdfunding campaigns. The article in Wu and Li [25] discussed several factors linked to successful creators and crowdfunding projects, including personality traits, social media engagement, influencer types, and project types. The majority of the previously mentioned works obtained classification results by using unimportant and irrelevant features. They also did not work on multiple features and multiple ML classifiers. Furthermore, they did not create any mobile applications that could help users with crowdfunding campaign success and fraud detection. Unlike previous work, this paper presents a high-accuracy crowdfunding campaign success prediction scheme based on multiple factors and ML classifier models. This paper looks at several factors that influence crowdfunding campaign success, including project duration, target funding goals, number of backers, backer feedback, previous campaign performance, project credibility, funding sources, advertising expenditures, average backing amounts, and supporter growth rates. This paper also uses feature extraction and hyperparameter tuning techniques to improve accuracy, precision, recall, and F1-scores. This article also provides a mobile crowdfunding assistance app with features such as donation, suggestions, profile, login, visualization of crowdfunding success, and police reporting capabilities.

3. Proposed Crowdfunding Campaign Success Prediction Framework

This section will discuss the steps required to predict the outcome of a crowdfunding campaign using ML classifiers.

3.1. System design

Figure 1 depicts the methodology diagram for our proposed crowdfunding campaign success prediction scheme based on ML techniques. The procedure begins with users completing questionnaires about various crowdfunding campaign attributes and characteristics. This data includes details such as project descriptions, financing goals, length, previous campaign activity, supporter feedback, and marketing initiatives. Following that, the information is compiled into a large dataset. Then we prepare our dataset for analysis by cleaning, normalizing, and handling missing values. Then, we performed feature extraction, scaling, and model selection processes. Then we ran model training and testing exercises. The overall dataset is divided into three categories: training, testing, and validation. We examined the RF algorithm, KNN algorithm, SVM algorithm, LR algorithm, and XGBoost algorithm to select a suitable prediction model. The best model is chosen by comparing its accuracy, precision, and F1-score value. Following the best model selection, the best model is used to predict the success of the crowdfunding campaign. The chosen model is implemented in the mobile application to predict the outcome of a campaign.

Figure 2 depicts the data flow diagram for our crowdfunding campaign assistance mobile app. This mobile application used Android Studio as the frontend user interface (UI) and Firebase as

Figure 1
Methodology diagram proposed system

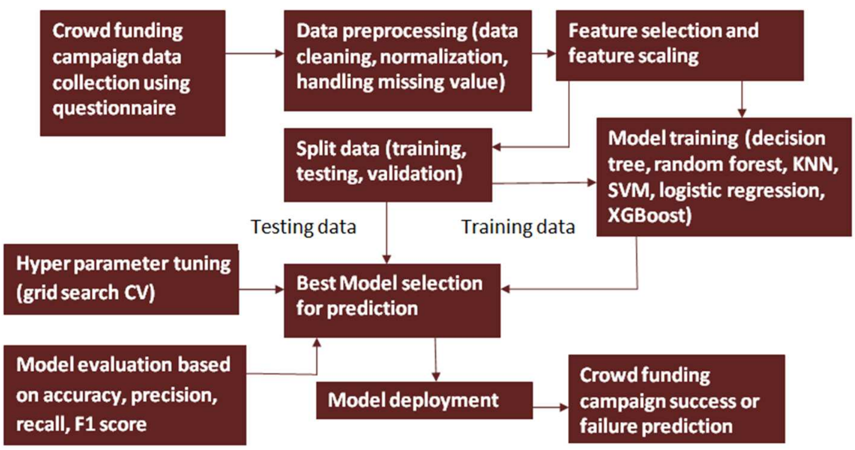
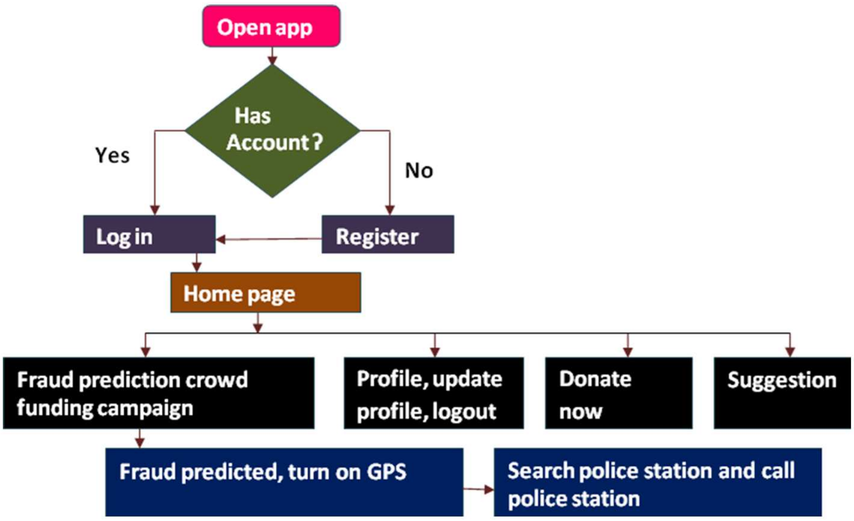


Figure 2
Dataflow diagram for proposed mobile app



the backend system. In this work, the UI data is constantly generated as an XML file. After designing our UI, we wrote backend logic in Java to accept data from the frontend. Firebase, as a backend system, connects mobile applications and the web to backend cloud storage and Application Programming Interface (APIs). To deploy the ML model on Android, we saved it as a pkl file. Then, in PyCharm, we used Python to access the Flask API. When a user submits a post request, the Flask API accepts the data and sends it to the ML model, which predicts the output class. We'll send the predicted class as JSON to the Android app. We used Postman to validate the API for our crowdfunding assistance application. The project's APIs are checked using Postman, an automated and interactive tool. To access the API, a single library called Volley is required. The logic we've devised states that we'll use the Android app's inputs to call the API, and the API's response will be displayed back in the Android app. Figure 2 clearly demonstrates that the Android application includes features such as login, home page, donations, crowdfunding campaign success prediction, profile, fraud detection, police reporting, and suggestions, among others. Section 4 will go over each mobile app feature in detail.

3.2. Data collection

The data collection process for crowdfunding campaign success prediction used a dual approach, with questionnaires administered through Google Forms and real-time crowdfunding campaign data integration from various platforms (e.g., Kickstarter and Indiegogo). These questionnaires addressed a wide range of issues, including project duration, target funding goals, number of backers, backer feedback, previous campaign performance, project credibility, funding sources, and advertising expenses. The questionnaires collected valuable qualitative and quantitative data from stakeholders involved in crowdfunding campaigns. We also gathered data from leading crowdfunding platforms via automated scraping techniques or APIs provided by the platforms themselves. This information included project outcomes (success or failure), funding amounts raised, the number of backers, project descriptions, and feedback from backers. Then we combined the data from both users and crowdfunding platforms. Figure 3 depicts a glimpse of our dataset, while Table 1 highlights the features of our dataset with descriptions. Our dataset consists of 420 rows (incidents) and 13 columns (factors). Eighty percent of the data is training, and 20% is test data. Three renowned economics and business administration faculty members from a university have validated our dataset. By combining direct user feedback and real-world campaign data, this approach ensured that the dataset was robust and suitable for generating valuable insights into crowdfunding campaign success factors.

3.3. Data cleaning and preprocessing

Before being analyzed, raw data must be cleaned, transformed, and organized to ensure that it is suitable for ML algorithms. To avoid duplication and ensure data accuracy, duplicate, missing, and null entries are removed from the dataset. We removed the garbage column from the final dataset because it served no purpose. Based on our expert recommendations, we included all necessary columns with no null values in the final dataset. Figure 4 depicts the status of the dataset following the dataset cleaning operation.

3.4. Feature selection, normalization

After that, we ran the data encoding operation. Data must be properly formatted for analysis. One-hot encoding is a technique for converting categorical data, such as "project duration" or "backer feedback," into numerical representations. This allows ML algorithms to efficiently handle category data. Scaling or normalizing continuous variables ensures that they have a constant range and distribution. This can improve the performance of ML models and accelerate their convergence. Normalization ensures that all features have a similar scale, allowing ML models to converge more quickly and perform better. We used the min-max scaling method to normalize our data. Categorical variables are encoded in numerical formats that are appropriate for ML algorithms. Figure 5 depicts the dataset following completion of normalization and one-hot encoding. We then used the sklearn module to perform the important feature selection operation. The best feature selection is required during the model training process to ensure the best features. Figure 6 depicts the key factors that influence the outcome of a crowdfunding campaign. This graph ranks features according to their impact on the model's predictions. This graph helps campaign creators and backers understand which features are most important regarding campaign success. This information helps to make

Figure 3
Glimpse of dataset

(a)

User id	Project duration	Target amount	No of backers	Backers feedback	Previous record	Address verification	Project credibility
0	long	high	More 1000	good	yes	real	credible
1	small	high	Less 1000	bad	no	fake	Not credible
2	medium	average	More 1000	bad	yes	real	credible
3	medium	low	More 1000	good	no	real	credible
4	long	average	More 1000	bad	no	fake	Not credible

(b)

Campaign realism	Calculation of money	Credible source	Rate of crowd funding platform	Advertisement cost	status
yes	perfectly	yes	high	Average cost	successful
no	imperfectly	no	high	No cost	unsuccessful
yes	imperfectly	yes	medium	Average cost	unsuccessful
yes	perfectly	no	high	No cost	successful
yes	imperfectly	yes	low	No cost	unsuccessful

Table 1
Dataset description with features

Feature	Description
Project Duration	Long, medium, small
Target Amount	High, average, low
Number of Backers	Less than 1000, more than 1000
Backers Feedback	Good, bad
Previous Record	Yes, no
Address Verification	Real, fake
Project Credibility	Credible, non-credible
Campaign Realism	Yes, no
Calculation of Money	Perfectly, imperfectly
Source Credibility	Yes, no
Rate of Crowdfunding Platforms	High, medium, low
Advertisement Cost	High cost, average cost, low cost

better decisions, optimize resource allocation, and improve campaign design. Finally, this leads to higher success rates for crowdfunding campaigns and a better understanding of what factors influence successful outcomes. Now, we will present some analysis of our data visualization, taking into account key factors. Figure 7 shows the frequency of various target amount values across project duration categories. The heat map's cells each display the number of projects with a given target amount falling within a specific project duration range. Each cell's color intensity serves as a visual cue for frequency; darker hues correspond to higher counts, while lighter hues correspond to lower counts. Figure 7 demonstrates that a larger target amount necessitates a longer project duration. Figure 8 depicts the impact of the number of backers on the project's target amount. The figure shows that as the project's target amount increases, so does the number of backers. Figure 8 can be used to identify trends and patterns in the data, such as whether larger target amounts attract more or fewer backers or whether specific target amounts consistently attract a certain number of backers. In this heat map, the most common combinations of target amount and number of backers are highlighted by darker cells, while the least common combinations are indicated by lighter cells. Figure 9 illustrates the relationship between the number of backers and their feedback. Each cell's color intensity indicates the frequency of that specific combination, with darker colors indicating higher frequencies.

3.5. Model selection, hyperparameter tuning

We evaluated the performance of the LR algorithm, KNN algorithm, RF algorithm, SVM algorithm, and DT algorithm to determine the best prediction model. We divided the total dataset into 80% training and 20% testing datasets, respectively. Figure 10(a) depicts the training and testing data splitting processes. Figure 10(b) depicts the model evaluation process for the RF algorithm. The best model is trained using the hyperparameter tuning

Figure 4
Dataset after cleaning operation

(a)

Id	Project duration	Target amount	No of backers	Backers feedback	Previous record	Address verification	Project credibility
0	Long	High	More than 1K	Good	yes	real	credible
1	Small	High	Less than 1K	Bad	no	fake	Non credible
2	medium	Average	More than 1K	Bad	yes	real	credible
3	medium	Low	More than 1K	Good	no	real	credible
4	Long	Average	More than 1K	Bad	no	fake	Non credible

(b)

Is campaign realistic	Money calculation	Source credible	Rate of crowdfund platform	Advertise ment cost	Status
yes	perfectly	yes	high	Avg. cost	successful
no	imperfectly	no	high	No cost	unsuccessful
yes	imperfectly	yes	medium	Avg. cost	unsuccessful
yes	perfectly	no	high	No cost	successful
yes	imperfectly	yes	low	No cost	unsuccessful

Figure 5
Glimpse of dataset after normalization and one-hot encoding

User id	Project duration long	Project duration medium	Project duration small	Target amount average	Target amount high	Target amount low	Backers less than 1000	Backers more than 1000
1	true	false	false	false	true	false	false	true
2	false	false	true	false	true	false	true	false
3	false	true	false	true	false	false	false	true
4	false	true	false	false	false	true	false	true
5	true	false	false	true	false	false	false	true

Figure 6
Important feature selection

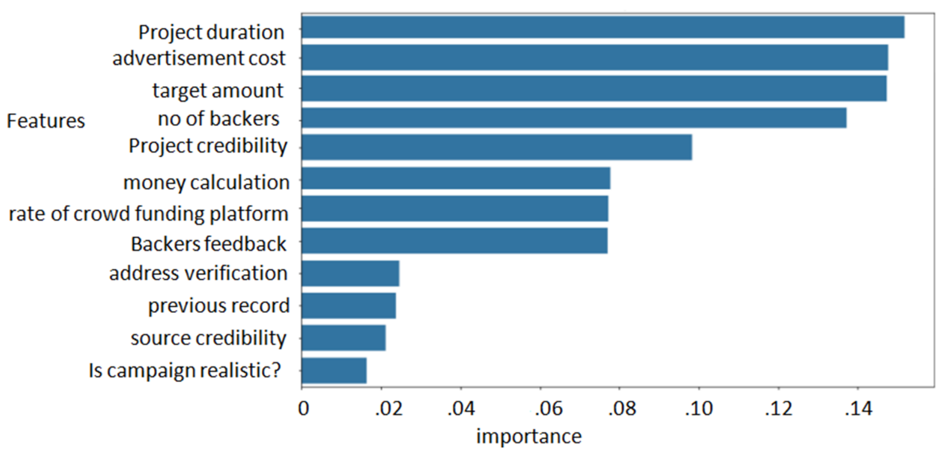


Figure 7
Frequency of various target amount values across different project durations

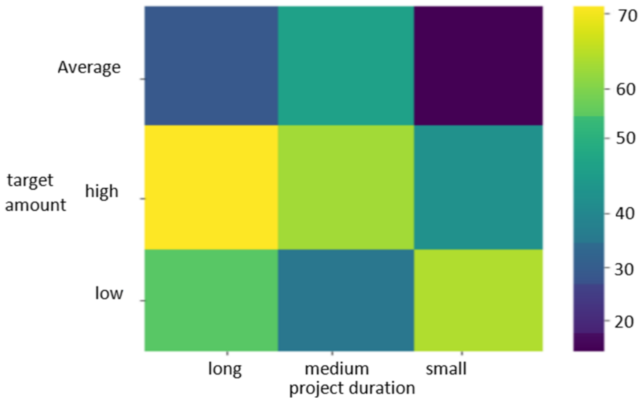


Figure 8
Frequency of number of backers on the target amount

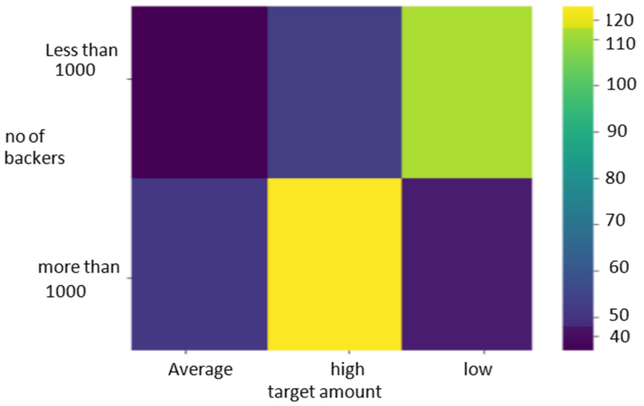
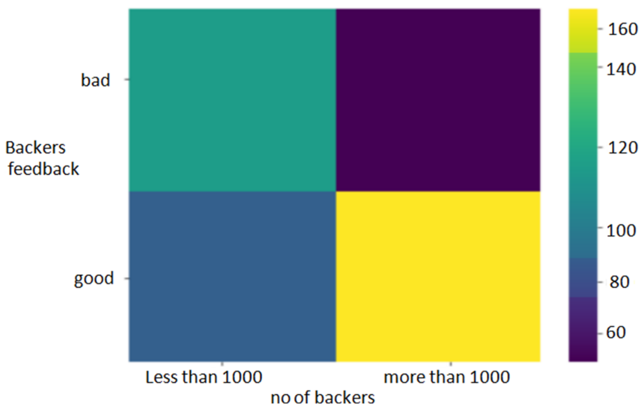


Figure 9
Relationship between the number of backers (no of backers) and the feedback



process shown in Figure 11. We used grid search CV to tune the hyperparameters. For the best crowdfunding campaign prediction model selection, we investigated evaluation results associated with various schemes such as confusion matrix, accuracy value, Receiver Operating Characteristic and Area Under the Curve (ROC-AUC) curve value, recall value, and precision value. We have used k-fold cross-validation where k = 5.

Figure 12(a) and (b) depict the ROC-AUC curve and confusion matrix of the DT algorithm, respectively. Figure 12(b) shows that the DT model has true positives (TP) (correct predictions of campaign status), true negatives (TN) (false positives (FP)), and false negatives (FN) values of 37, 42, 4, and 1, respectively. Figure 12(a) shows that the DT model has an ROC-AUC value of.94. Figure 13(a) and (b) show the ROC-AUC curve and confusion matrix for the RF algorithm, respectively. Figure 13(b) shows that the RF model has TP (correct prediction of campaign status), TN, FP, and FN values of 37, 44, 2, and 1, respectively. Figure 13(a) shows that the RF model has an ROC-AUC value of 0.97. The RF model's ROC-AUC score indicates a high ability to discriminate between successful and unsuccessful campaigns. Figure 14(a) and (b) show the ROC-AUC score and confusion matrix for the KNN algorithm, respectively. Figure 14(b) shows that the KNN model has TP (correct predictions of campaign status), TN (FP), and FN values of 31, 40, 6, and 7, respectively. Figure 14(a) shows that the KNN model has an ROC-AUC value of.898. However, KNN is affected by noisy data and outliers, resulting in errors, as evidenced by FP and FN. It also requires a significant amount of computing power to calculate distances between points. The method struggles with numerous features, rendering distances less useful. Unlike other models, KNN does not generate a general model, making it less efficient and more difficult to understand. Figure 15(a) and (b) show the ROC-AUC score and the confusion matrix for the SVM algorithm, respectively. Figure 15(b) shows that the SVM model has TP (correct prediction of campaign status), TN, FP, and FN values of 27, 37, 9, and 11, respectively. Figure 15(a) shows that the SVM model has an ROC-AUC value of 0.88. SVM may be less effective when dealing with overlapping classes or noisy data, resulting in FP and FN. Furthermore, SVM can be sensitive to kernel and regularization parameter selection, and it may not perform well on large datasets due to its high computational costs. Figure 16(a) and (b) show the ROC-AUC score and confusion matrix for the LR algorithm, respectively. Figure 16(b) shows that the LR model has 27 TP, 38 TN, 8 FP, and 11 FN. The LR model has an ROC-AUC value of.891 (refer to Figure 16(a)). The LR model forecasts campaign success using a weighted sum of input features converted into a binary outcome via a logistic function. However, LR struggles with complex datasets containing nonlinear relationships that are sensitive to outliers. It assumes predictor independence, which may not be valid, resulting in biased predictions. Figure 17(a) and (b) show accuracy and precision value comparisons for various ML models. Figure 18(a) and (b) compare the recall and F1-score values of various ML models. Both figures show that the KNN model has 84.52% accuracy, 83.78% precision, 81.57% recall, and an F1-score of 82.66%. The RF model has an accuracy of 96.42%, precision of 94.87%, recall of 97.36%, and an F1-score of 96.98%. The DT model has 94.04% accuracy, 90.24% precision, 97.3% recall, and an F1-score of 93.67%. The RF algorithm has the best accuracy score and is the most reliable for predicting the positive outcome of crowdfunding projects, according to the accuracy comparison. KNN and DTs both perform well, albeit with slightly lower accuracy scores. The RF algorithm reduces FP and achieves higher precision, recall, and F1-score. Based on the results, we can conclude that the RF is the best predictive model for our crowdfunding campaign outcome prediction.

3.6. Comparison with existing works

In this section, we present a thorough comparison of our proposed method for predicting crowdfunding campaign success with several existing methods reported in the literature. Our evaluation results in Table 2 show that the proposed RF-based crowdfunding

Figure 10
(a) Training and testing data splitting process. (b) Model evaluation and result generation for random forest algorithm

(a)

```
Import matplotlib.pyplot as plt
Import seaborn as sns
#converts the training and testing sets to data frame

X_train_df=pd.DataFrame (X_train, columns=X.columns)
X_test_df=pd.DataFrame(X_test, columns=X.columns)
y_train_df=pd.DataFrame (y_train, columns=['success'])
y_test_df=pd.DataFrame (y_test, columns=['success'])

#combines features and target for pair plots
train_data=pd.concat ([X_train_df, y_train_df], axis=1)
test_data=pd.concat([X_test_df, y_test_df], axis=1)
```

(b)

```
#split the dataset into training and testing sets
X_train, X_test, y_train, y_test=train_test_split(X, y, test_size=.2,
random_state=42)
# random forest classifier model train
model= RandomForestClassifier (n_estimators=100, random_state=42)
model.fit(X_train, y_train)
#evaluate the model
Y_pred=model.predict (X_test)
accuracy=accuracy_score (y_test, y_pred)
precision=precision_score (y_test, y_pred)
recall=recall_score (y_test, y_pred)
f1=f1_score (y_test, y_pred)
```

Figure 11
K-fold cross-validation and hyperparameter tuning

```
Kf=Kfold (n_splits=5, shuffle='true', random_state=42)
grid search=GridSearchCV (estimator=model, param_grid=param_grid)
grid_search.fit(X_train, y_train)
best_params=grid_search.best_param_grid
print ('best parameters', best_params)
#train the best model with entire training data
best_model=grid_search.best_estimator_model
best_model.fit (X_train, y_train)
```

Figure 12
(a) ROC-AUC curve for decision tree algorithm. (b) Confusion matrix for decision tree algorithm

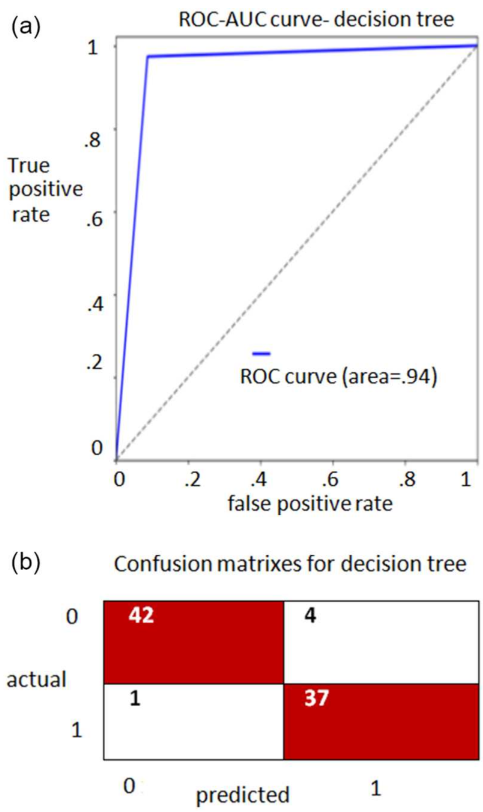
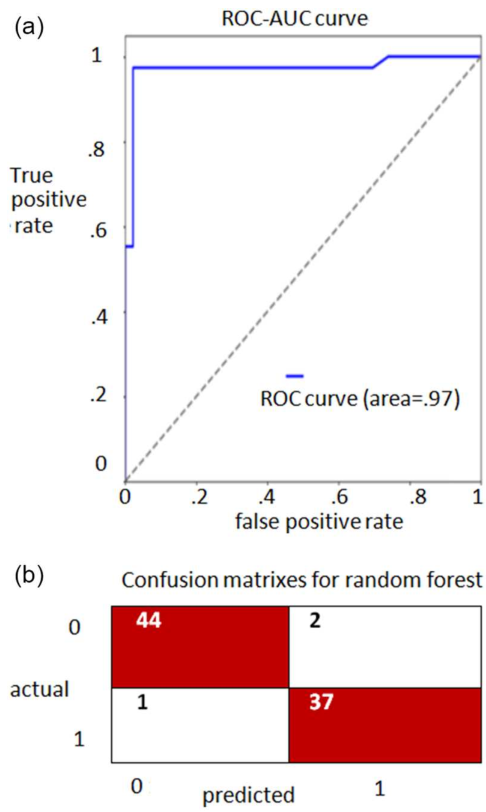


Figure 13
(a) ROC-AUC curve for random forest algorithm. (b) Confusion matrix for random forest algorithm



campaign success prediction model achieves at least 3.1% more accuracy, 2.6% more precision, 2% more recall, and .64% more F1-score value than the existing works (e.g., 12, 17, 20, 21). The article in Perez et al. [12] used an ensemble ML classifier to predict crowdfunding success, with an accuracy of 90.14%. The work in Zhong et al. [17] used a node2vec model to forecast crowdfunding success and achieved an accuracy of 80.5%. The article in Raflesia et al. [20] used the XGBoost algorithm to predict crowdfunding campaign success with an impressive 93.34% accuracy. The work in Tang et al. [21] used attention-based deep learning techniques to predict the outcome of crowdfunding initiatives, with an accuracy of 72.3%.

Figure 14
(a) ROC-AUC curve for KNN algorithm. (b) Confusion matrix for KNN algorithm

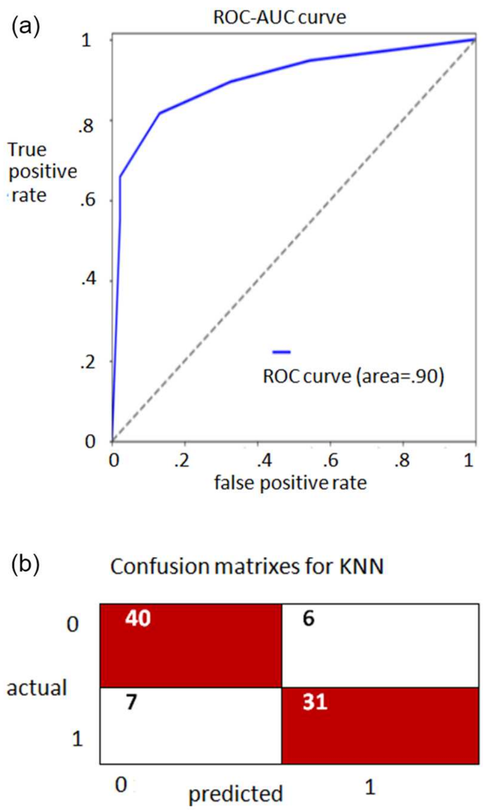


Figure 15
(a) ROC-AUC curve for SVM algorithm. (b) Confusion matrix for SVM algorithm

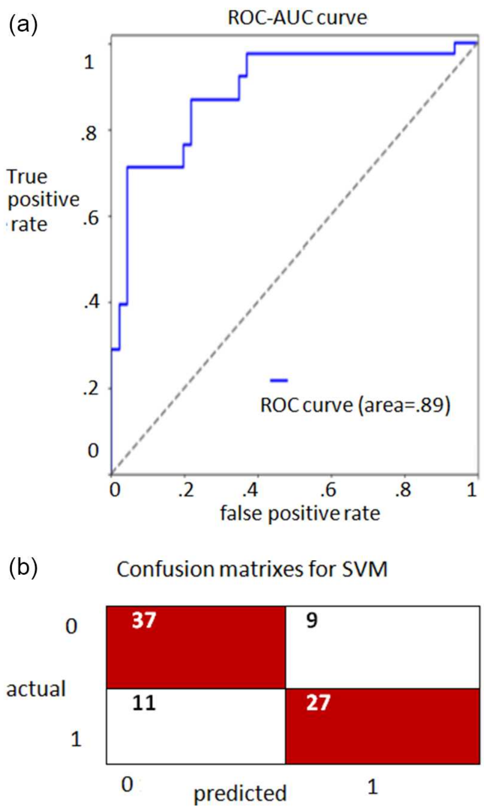


Figure 16
(a) ROC-AUC curve for logistic regression algorithm. (b) Confusion matrix for logistic regression algorithm

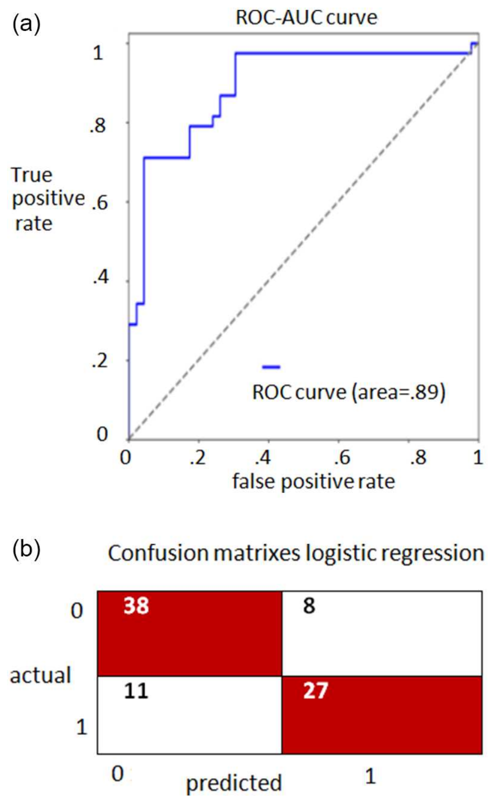


Figure 17
(a) Accuracy comparison. (b) Precision comparison among ML models

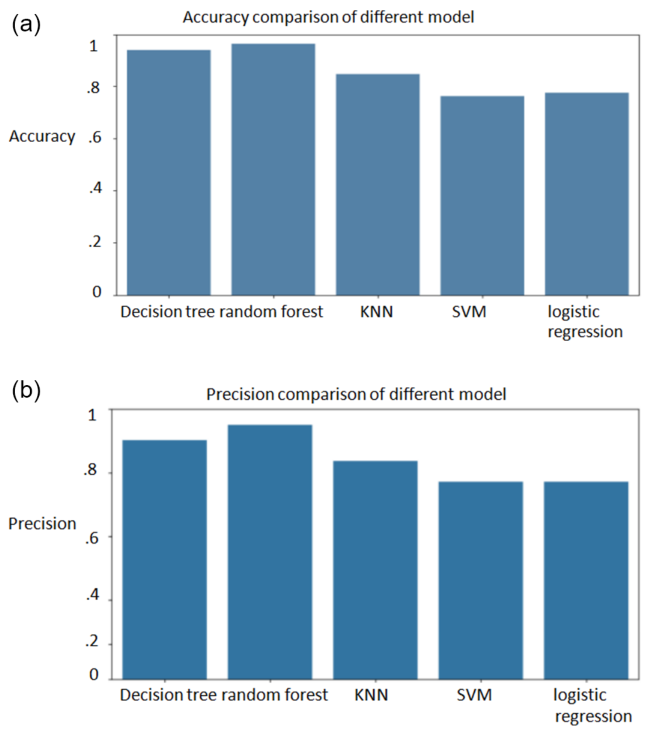
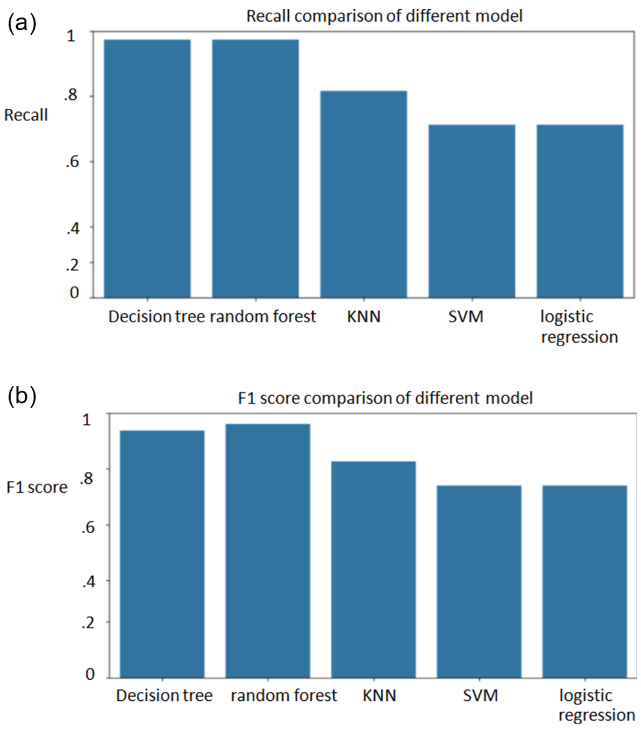


Figure 18
Recall comparison. (b) F1-score comparison among ML models



The proposed RF-based crowdfunding success prediction model outperforms the existing one by incorporating dataset cleaning, validation, best feature selection, normalization, hyperparameter tuning, and cross-validation techniques. Our approach, which includes comprehensive feature engineering, meticulous data preprocessing, and rigorous model tuning, has successfully improved the predictive accuracy of crowdfunding campaign success.

4. Android Application Development

This paper creates a crowdfunding campaign assistance mobile application using Android Studio as the frontend and Google Firebase as the backend platform. The developed Android application includes features such as login, home screen, crowdfunding campaign success prediction, suggestions, donation, and police reporting. Figure 19(a), (b), and (c) show the login, account creation, and home screen of our crowdfunding assistance application, respectively. Users interested in the crowdfunding campaign can access their account by entering their email address and password (see Figure 19(a)). However, the account sign-up requires an email, phone number, address, name, and password (see Figure 19(b)). After logging in, users are taken directly to the homepage (Figure 19(c)). This page provides access to the entire system's functionality. The Location widget allows users to see where they are. Other widgets include donate, check success prediction, report fraud, make suggestions, and report to the police. Figure 20 depicts a key component of the application feature: the check crowdfunding campaign success prediction. To obtain the campaign success prediction result, the user must submit some question answers in Figure 20(a) and (b) on the screen. The user must provide answers to questions such as project duration, target amount, project credibility, number of supporters, backer feedback, and advertising cost, among others. Finally, Figure 20(c) shows the successful campaign prediction results. Similarly, the user can view the unsuccessful campaign results by submitting the question responses and accessing the app screen, as shown in Figure 21(a), (b), and (c). Figure 20 depicts a long-term campaign that aimed to raise a large amount and attracted 1000 backers. The feedback from backers was positive. The project had a verifiable address and a solid track record. The campaign was realistic, with accurate financial calculations, came from reputable sources, received high ratings on crowdfunding platforms, and had an average advertising cost. These responses contributed to the campaign's success, as illustrated in Figure 20(c). Figure 21 depicts a campaign that lasted a few days and had a high target amount, but only 300 backers joined. The feedback was mixed; there was no previous record, and the address was not verified, resulting in medium credibility. The campaign was deemed unrealistic, with incorrect financial calculations and a less credible source. It had a lower rating on crowdfunding platforms and a higher advertising cost, which all contributed to the campaign's failure. Figure 22(a) shows our app's donation feature. The app screen allows the user to select a card number and amount for a donation. Figure 22(b) depicts an app suggestion screen. The Figure 22(b) screen provides the user with information about successful campaigns as well as fraud in crowdfunding campaigns. Figure 23(a) shows the police report feature. Customers can choose a district and police station to report crowdfunding fraud incidents. The customer can also use the Figure 23(b) screen to send messages to police officers and the Figure 23(c) screen to call them. Government police personnel have the authority to keep an eye on the reporting procedure and to take legal action if it is utilized unconstitutionally.

5. Evaluation Results

Figure 24 shows the user evaluation results for our crowdfunding assistance mobile application features. We collected user feedback from 200 users through an offline project showcase process and an online survey. Our work was validated by the users. If users are satisfied, then our goal has been accomplished. We examined the performance of five features, including the home page, prediction, donation, suggestion, and police report (see x-axis). The y-axis represents the number of comments received for each feature. Each feature is evaluated using four comment categories: excellent,

Table 2
Comparison of performance with existing works

Work reference	Method	Accuracy	Precision	Recall	F1-score
Perez et al. [12]	Ensemble ML classifier	90.14	88.90	93.36	95.78
Raflesia et al. [20]	XGBoost	93.34	92.22	95.35	96.34
Zhong et al. [17]	Node2vec model	80.5	77.98	81.78	85.46
Tang et al. [21]	Deep cross attention	72.3	71.4	75.4	77
Our Proposed Method	Random Forest Classifier	96.42	94.87	97.36	96.98

Figure 19
(a) Login. (b) Account creation. (c) Feature search home page

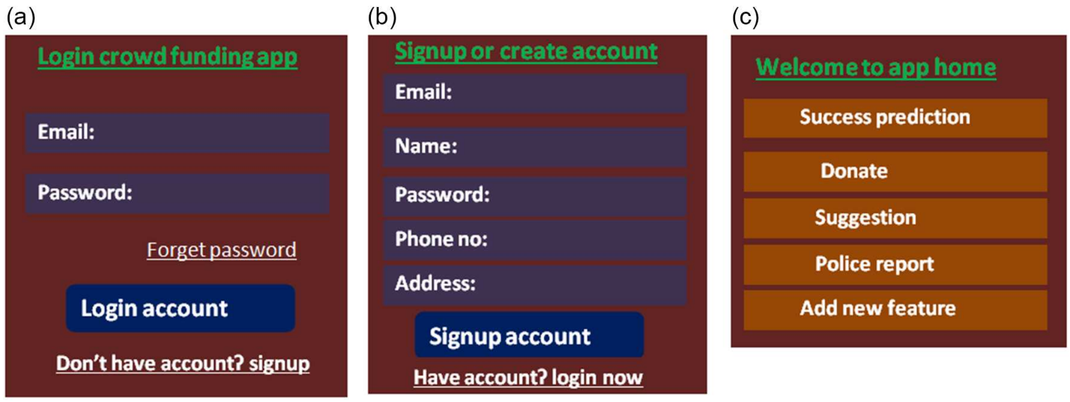


Figure 20
(a) Prediction screen (part 1). (b) Prediction screen (part 2). (c) Success results

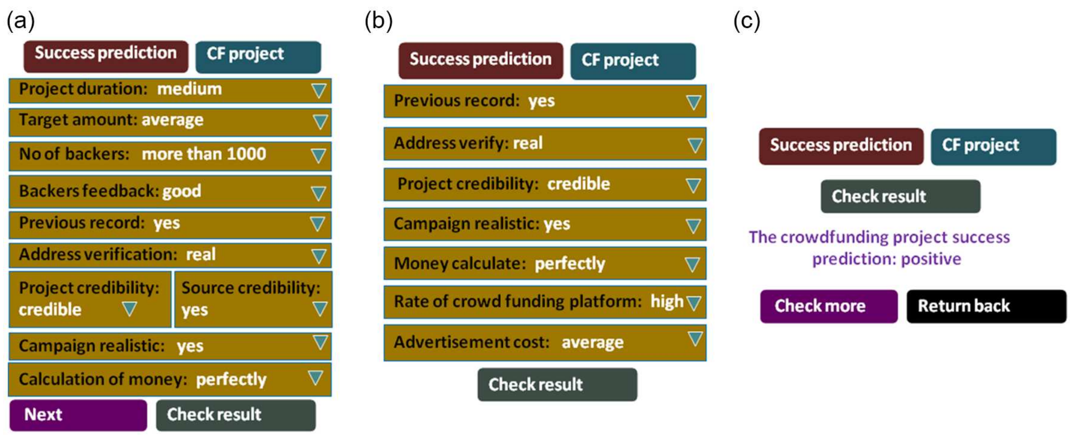


Figure 21
(a) Campaign failure prediction (part 1). (b) Prediction screen (part 2). (c) Failure results

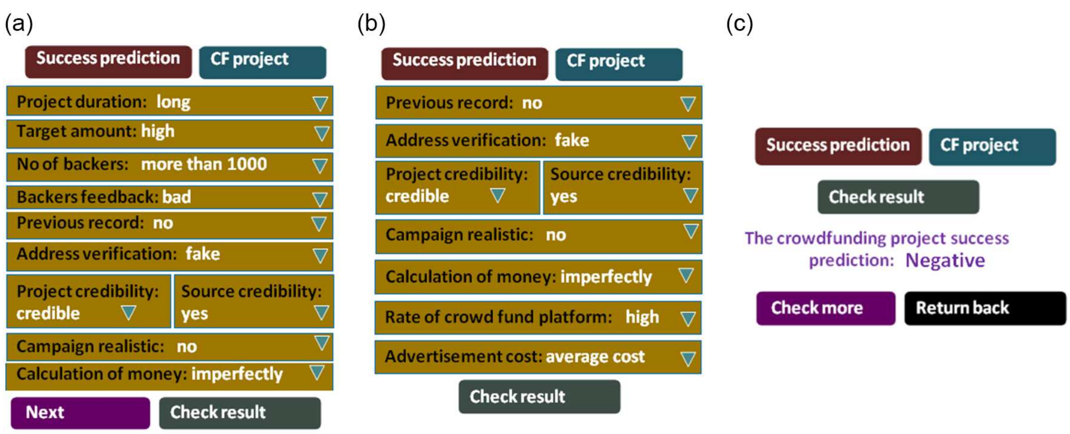
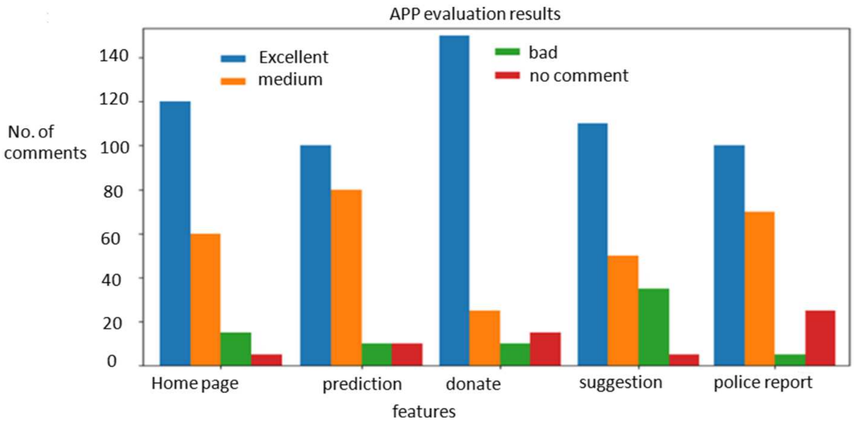


Figure 22
(a) Donation screen. (b) Suggestion screen of our app

Figure 23
(a) Police report screen (part 1). (b) Police report screen (send message). (c) Police report (calling option)

Figure 24
Evaluation of app features



medium, bad, and no comment. The number of users who left Excellent, Medium, Bad, and No Comments for the home page feature is 120, 60, 15, and 5, respectively.

For the campaign success prediction feature, there are 100, 80, 10, and 10 users who left Excellent, Medium, Bad, or No Comments. The number of users who rated the police reporting feature as Excellent, Medium, Bad, or No Comment is 150, 25, 10, and 15. There are 100, 70, 5, and 25 users who have left Excellent, Medium, Bad, or No Comments about the donation feature. There are 110, 50, 35, and 5 users who have left Excellent, Medium, Bad, and No Comments for the suggestion feature, respectively. Thus, the results show that more than 80% of users provided appropriate feedback about the app's features.

6. Conclusion

This paper introduced an ML-based crowdfunding campaign success prediction scheme by investigating various factors and algorithms. This paper developed a crowdfunding campaign success prediction dataset by analyzing a variety of key factors, including project duration, target funding goals, number of backers, backer feedback, previous campaign performance, project credibility, funding sources, advertising expenditures, backing amounts, and supporter growth rates. The prediction model also employs data cleaning, feature selection, normalization, training, testing, cross-validation, and hyperparameter tuning methods. This paper examined several ML classifiers, including the SVM classifier, KNN classifier, DT classifier, LR classifier, and RF classifier, to determine the most accurate prediction model. The results showed that the RF classifier outperformed other classifiers with a 96.42% accuracy value, 94.87% precision value, 97.36% recall value, and 96.98% F1-score value. The comparison results showed that the proposed RF-based prediction model achieves 3.1% more accuracy and 2.6% more precision than previous works. This work also provides users with a crowdfunding assistance Android application that includes features such as donation, crowdfunding campaign success prediction, suggestion, home page, and police reporting. The user survey-based app feature evaluation results confirmed that more than 80 users are pleased with the developed app features. To verify the generalizability and dependability of the suggested model, more tests utilizing bigger and more varied datasets will be required, in addition to validation on data from actual crowdfunding platforms. The suggested model's robustness can be increased by using the confidence interval on testing data. Future extensions of this work could include more data collection, text and video-based crowdfunding campaign success prediction, crowdfunding campaign fraud reason prediction, IoT and ML-based real-time crowdfunding campaign fraud prediction, security enhancement using blockchain and IoT technology, and generative and explainable AI-based crowdfunding campaign fraud classification, among other things.

Recommendation

This paper describes that the RF classifier is most suitable for crowdfunding campaign success prediction.

Ethical Statement

This study did not require formal ethical approval because Chittagong University of Engineering and Technology, Bangladesh, does not have an Institutional Review Board or ethics committee requirement for this type of non-medical social science research. Despite the exemption, the study was conducted in accordance with accepted ethical standards. Participation was voluntary, informed consent was obtained prior to data collection, and no personally identifiable information was collected or disclosed.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support this work are available upon reasonable request to the corresponding author.

Author Contribution Statement

Suzab-Ud Doula: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Visualization. **Mahfuzulhoq Chowdhury:** Conceptualization, Methodology, Validation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration.

References

[1] Kleemann, F., Voß, G. G., & Rieder, K. (2008). Un (der) paid innovators: The commercial utilization of consumer work through crowdsourcing. *Science, Technology & Innovation Studies*, *4*(1), 5–26. https://doi.org/10.17877/DE290R-12790

[2] Levin, F. (2015). Success determinants of crowdfunding projects. *Review of Business and Economics Studies*, *3*(4), 79–87.

[3] Chen, L. -S., & Shen, E. -L. (2019). Finding the keywords affecting the success of crowdfunding projects. In *2019 IEEE 6th International Conference on Industrial Engineering and Applications*, 567–571. https://doi.org/10.1109/IEA.2019.8714815

[4] Dousios, D., Korfiatis, N., & Spanaki, K. (2025). The persuasive impact of text and audio features on technology crowdfunding campaigns: An empirical study on Kickstarter. *IEEE Transactions on Engineering Management*, *72*, 2049–2064. https://doi.org/10.1109/TEM.2025.3549597

[5] Zhou, W., Wang, S., & Chen, J. (2024). Default identification of scientific crowdfunding projects: A hesitant fuzzy deep learning approach. *IEEE Transactions on Engineering Management*, *71*, 15453–15464. https://doi.org/10.1109/TEM.2024.3495496

[6] Lee, S., Park, H., & Kim, H. C. (2025). Fraud detection on crowdfunding platforms using multiple feature selection methods. *IEEE Access*, *13*, 40133–40148. https://doi.org/10.1109/ACCESS.2025.3547396

[7] Prateek, P., Kim, D. J., & Ge, L. (2021). Detection of fraudulent campaigns on donation-based crowdfunding platforms using a combination of machine. In *Proceedings of the 16th Pre-ICIS Workshop on Information Security and Privacy*.

[8] Hou, X., Wu, T., Chen, Z., & Zhou, L. (2022). Success factors of medical crowdfunding campaigns: Systematic review. *Journal of Medical Internet Research*, *24*(3), e30189. https://doi.org/10.2196/30189

[9] Fernandez-Blanco, A., Villanueva-Balsera, J., Rodriguez-Montequin, V., & Moran-Palacios, H. (2020). Key factors for project crowdfunding success: An empirical study. *Sustainability*, *12*(2), 599. https://doi.org/10.3390/su12020599

[10] Yuan, H., Lau, R. Y. K., & Xu, W. (2016). The determinants of crowdfunding success: A semantic text analytics approach. *Decision Support Systems*, *91*, 67–76. https://doi.org/10.1016/j.dss.2016.08.001

[11] Kim, T., Por, M. H., & Yang, S. B. (2017). Winning the crowd in online fundraising platforms: The roles of founder and project features. *Electronic Commerce Research and Applications*, *25*, 86–94. https://doi.org/10.1016/j.elerap.2017.09.002

[12] Perez, B., Machado, S., Andrews, J., & Kourtellis, N. (2022, June). I call BS: Fraud detection in crowdfunding campaigns. In *Proceedings of the 14th ACM Web Science Conference 2022*, 1–11. https://doi.org/10.1145/3501247.3531541

[13] Siering, M., Koch, J. A., & Deokar, A. V. (2016). Detecting fraudulent behavior on crowdfunding platforms: The role of linguistic and content-based cues in static and dynamic contexts. *Journal of Management Information Systems*, *33*(2), 421–455. https://doi.org/10.1080/07421222.2016.1205930

[14] Koch, J. A., & Siering, M. (2015). Crowdfunding success factors: The characteristics of successfully funded projects on crowdfunding platforms. In *23rd European Conference on Information Systems (ECIS 2015)*, 1–15.

[15] Zhou, L., Burgoon, J. K., Nunamaker, J. F., & Twitchell, D. (2004). Automating linguistics-based cues for detecting

deception in text-based asynchronous computer-mediated communications. *Group Decision and Negotiation*, *13*(1), 81–106. https://doi.org/10.1023/B:GRUP.0000011944.62889.6f

[16] Sonare, B., Phursule, R., Dhajekar, J., Gandhi, P., & Kannawar, P. (2024). Improving crowdfunding integrity with blockchain using ethereum-based smart contract. In *2024 8th International Conference on Computing, Communication, Control and Automation*, 1–4. https://doi.org/10.1109/ICCUBEA61740.2024.10775019

[17] Zhong, C., Xu, W., & Du, W. (2022). Success prediction of crowdfunding campaigns with project network: A machine learning approach. *Journal of Electronic Commerce Research*, *23*(2), 99–114.

[18] Shah, A., Shah, P., Savla, U., Rathod, Y., & Baloorkar, N. (2023). Predictive analysis of crowd funding projects. In S. Kautish, P. Chatterjee, D. Pamucar, N. Pradeep, & Singh. D (Eds.), *Computational intelligence for modern business system: Emerging applications and strategies* (pp. 83–95). Springer. https://doi.org/10.1007/978-981-99-5354-7_5

[19] Gangi, F., Daniele, L. M., Scuotto, V., & Tani, M. (2023). Uncovering backers' intention to participate in reward-based crowdfunding: the role of value cocreation. *IEEE Transactions on Engineering Management*, *71*, 12686–12698. https://doi.org/10.1109/TEM.2023.3267088

[20] Raflesia, S. P., Lestarini, D., Kurnia, R. D., & Hardiyanti, D. Y. (2023). Using machine learning approach towards successful crowdfunding prediction. *Bulletin of Electrical Engineering and Informatics*, *12*(4), 2438–2445. https://doi.org/10.11591/eei.v12i4.5238

[21] Tang, Z., Yang, Y., Li, W., Lian, D., & Duan, L. (2022). Deep cross-attention network for crowdfunding success prediction. *IEEE Transactions on Multimedia*, *25*, 1306–1319. https://doi.org/10.1109/TMM.2022.3141256

[22] Oduro, M. S., Yu, H., & Huang, H. (2022). Predicting the entrepreneurial success of crowdfunding campaigns using model-based machine learning methods. *International Journal of Crowd Science*, *6*(1), 7–16. https://doi.org/10.26599/IJCS.2022.9100003

[23] Yeh, J. Y., & Wang, Z. L. (2023, July). Mining ESG semantic features for success prediction in green-oriented crowdfunding campaigns. In *2023 International Conference on Consumer Electronics-Taiwan (ICCE-Taiwan)*, 819–820. https://doi.org/10.1109/ICCE-Taiwan58799.2023.10226784

[24] Patil, A., Arsania, T., Nahar, R., Mohite, A., & Trivedi, A. (2023). Startup funding app using flutter and Machine learning. In *2023 International Conference on Advanced Computing Technologies and Applications (ICACTA)*, 1–7. https://doi.org/10.1109/ICACTA58201.2023.10393813

[25] Wu, T. Y., & Li, Y. M. (2023). Exploring the role of social identity and semantic traits in the success of the creator economy. In *2023 14th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*, 53–56. https://doi.org/10.1109/IIAI-AAI59060.2023.00021

How to Cite: Doula, S. -U., & Chowdhury, M. (2026). A Crowdfunding Campaign Success Prediction Scheme Using Machine Learning. *FinTech and Sustainable Innovation*. https://doi.org/10.47852/bonviewFSI62027666