



# Prediction of fundraising outcomes for crowdfunding projects based on deep learning: a multimodel comparative study

Wei Wang<sup>1</sup> · Hongsheng Zheng<sup>1</sup> · Yenchun Jim Wu<sup>2</sup>

Published online: 7 March 2020

© Springer-Verlag GmbH Germany, part of Springer Nature 2020

## Abstract

As a new financing model, crowdfunding has been developed rapidly in recent years and has attracted the attention of investors and small- and medium-sized enterprises and entrepreneurs. However, many projects fail to be funded; thus, crowdfunding project fundraising outcomes forecasting and multimodel comparisons are meaningful ways to identify project quality and reduce market risk. It is important to reduce participation risk through automated methods, which is of great significance to the sustainable development of Internet finance. First, based on the data from the Kickstarter, preprocessing and exploratory analysis are conducted. Then, we introduce a deep learning algorithm (multilayer perceptron) and apply it to the prediction of crowdfunding financing performance. We compare deep learning with other commonly used machine learning algorithms, including decision tree, random forest, logistic regression, support vector machine, and K-nearest neighbors algorithm. We tune each machine learning algorithm to get the best parameters. The experimental results show that the deep learning model can obtain the best prediction results, with an accuracy of 92.3% when predicting the fundraising outcomes of crowdfunding financing, followed by the decision tree. Deep learning shows significant advantages in many evaluation criteria, which demonstrates the potential for crowdfunding project financing predictions. This study combines machine learning with Internet finance, providing inspiration for future research and resulting in many practical implications.

**Keywords** Crowdfunding · Deep learning · Machine learning · Fundraising prediction · Multilayer perceptron

## 1 Introduction

The rise of the crowdfunding model has broken the limits of traditional financing channels so that every ordinary but creative individual can obtain financial support through the

network. This financing model breaks geographical limitations and has been rapidly developed. In 2014, the global crowdfunding industry market size was \$16.2 billion. In 2015, it doubled to \$34.4 billion, and it reached more than \$50 billion in 2016. This financing model has maintained steady growth in recent years (Testa et al. 2019). It is estimated that the size of the crowdfunding market will grow at an annual growth rate of approximately 30%, and the transaction volume is expected to reach \$26 trillion by 2022. In terms of the global distribution, the top five countries in terms of transaction volume are China, USA, UK, France, and Canada (Simons et al. 2019), which shows the geographical imbalance of online financing. However, challenges come with opportunities, the high failure rate of fundraising results in that the fundraisers are extremely concerned about the fundraising outcomes, and they are eager to know the probability of successful financing before the project goes online. A suitable prediction algorithm would provide a potential solution undoubtedly, which is of great benefit to fundraisers and crowdfunding

---

Communicated by Mu-Yen Chen.

---

✉ Yenchun Jim Wu  
wuyenchun@gmail.com

Wei Wang  
wayswang@gmail.com

Hongsheng Zheng  
hsz\_181811@163.com

<sup>1</sup> College of Business Administration, Huaqiao University, NO. 269, Chenghua North Road, Quanzhou City, Fujian Province, China

<sup>2</sup> Graduate Institute of Global Business and Strategy, National Taiwan Normal University, National Taipei University of Education, 31 Shida Road, Daan District, Taipei, Taiwan

platforms. However, we are not clear about the prediction power of different prediction algorithms for crowdfunding campaigns.

In a broad sense, crowdfunding refers to a business model in which a large number of people (usually Internet users) invest in a project, and each participant invests a small amount of capital. Then, these small amounts are pooled into a certain amount of money to provide financial support for individual founders or enterprises (Long et al. 2019). Generally, in the absence of financial intermediaries, individual entrepreneurs, enterprises, or nonprofit groups can propose project ideas through an Internet platform and complete their financing goals due to the huge number of participants. Therefore, the crowdfunding model is essentially a reflection of the collective wisdom in which the investor group determines the project financing performance collectively (Chaney 2019). Project financing performance is affected by multiple investments from huge group of investors, so the prediction of fundraising performance is different from other models which are only affected by a few variables. In the prediction of crowdfunding fundraising, we must select the appropriate factors carefully as the prediction variables, i.e., feature engineering.

Moreover, the flourishing development of the crowdfunding model has been accompanied by some problems that need to be solved. The asymmetry of information, the uncertainty of financing, the lack of experienced investors, the high opportunity costs, and the low success rate of crowdfunding projects have plagued all involved parties (Miglo and Miglo 2019). Online financing projects reduce information asymmetry through textual descriptions (Miglo and Miglo 2019); therefore, the factors related to project textual signals have become the key factors affecting the financing performance of crowdfunding projects (Gafni et al. 2019). Research on the dynamic attributes of crowdfunding projects shows that personal social networks and geography weaken the information asymmetry (Chinnaswamy et al. 2019; Mollick 2014). However, because investors lack effective means and capabilities to identify project quality, some low-quality projects use abnormal means to obtain short-term financing success, but this has led to a large number of enterprisers being forced to “flee” risks in later implementations. The online environment has greatly increased the risk of the crowdfunding financing mode and produced the phenomenon that “bad money” has driven out “good money,” forming a lemon market. In view of the current situation with high financing failure rates and high market risk, it is of theoretical and practical importance to analyze the influencing factors of crowdfunding financing performance, forecast project financing performance, identify high-quality projects, and guide the

enterprisers to generate attractive campaigns in reasonable ways.

In the existing studies, most of the scholars use the statistical and econometric models to investigate the correlation between variables and the impacts of the variables, and they mainly focus on the factors affecting crowdfunding financing. Important variables are extracted for financing performance predictions by feature engineering. The quality of the feature engineering greatly affects the subsequent prediction outcomes. The commonly used independent variables for crowdfunding predictions include the attributes of the project, the dynamic attributes of the project, and the social connections of the founders. The dependent variables include financing status (1 for success, and 0 for otherwise), the number of backers, the pledged capital amount, financing progress, etc. The methods that have been used mainly include econometrics methods, statistical methods, and classification algorithms. Some researchers have adopted decision trees, logistic regression, support vector machines, but few scholars have tried artificial neural networks and deep learning models. We are still not clear about the effect of the current popular deep learning algorithm on crowdfunding fundraising outcome predictions.

Since feature engineering is required in most of machine learning algorithms, the feature selection will affect the prediction results. Therefore, the prediction results of such algorithms depend largely on feature selection and pre-processing. Deep learning provides an end-to-end solution that simplifies feature engineering and allows us to focus more on the performance of the algorithm itself. Therefore, based on the extant literatures, this study employs a deep learning multilayer perceptron (MLP) to predict the fundraising outcomes of the crowdfunding projects by the attributes of crowdfunding projects, including pledge goals, categories, funding durations, geographical locations, etc. We also make a comprehensive comparison of MLP with other commonly used machine learning algorithms.

The research on the improvement of the crowdfunding project financing success rate and its influencing factors has important implications for founders, investors, and crowdfunding platforms. For founders, if enterprisers know in advance which factors can significantly improve the financing success rate, they would highlight these aspects to improve financing performance and reduce opportunity costs. Investors, through performance predictions, would avoid high-risk, failure-prone projects and use limited funds to support projects with higher success rates, thereby increasing the possibility to be rewarded for their investment. For crowdfunding platforms, improving the project financing success rate affects its profitability; therefore, the financing platform urgently needs to understand the factors affecting the performance of crowdfunding financing, and

these results should be provided to investors and founders to improve the success rate of financing and improve the transparency of online financial markets. To predict the financing performance of crowdfunding projects, this study employs the world's largest reward-based crowdfunding platform—Kickstarter, and introduces machine learning algorithms for the financing performance predictions. The following algorithms are adopted to predict and compare crowdfunding performance: the decision tree, random forest, logistic regression, support vector machine, K-nearest neighbors, and deep learning algorithms. The models are comprehensively evaluated by confusion matrices, accuracy, precision, recall, F1, area under the curve (AUC), average precision-recall score (APRS), and Matthews correlation coefficient (MCC). These results are helpful to improve the financing performance of crowdfunding projects.

## 2 Research progress and literature review

### 2.1 Factors affecting crowdfunding

Online fundraising outcomes are affected by many factors. The participation intentions of backers have a significant impact on project financing performance. The communication between founders and online users will promote investors' willingness to participate in a project. Investors' willingness to participate is extremely important to the success of the project, and online advertisements can also promote successful project financing (Kraus et al. 2016). Research on the dynamic attributes of crowdfunding projects shows that personal social networks, project quality, and geography are all related to fundraising outcomes (Chinnaswamy et al. 2019; Mollick 2014). If we focus on the project itself and the characteristics of the founder, the main influencing factors include the project description, images, videos, and whether the founder has established social relationships. That is, the founder can use appropriate project descriptions (text, images, and videos) and expand their social connections to improve the financing success rate (Kromidha and Robson 2016).

Empirical research has found that the most significant factors affecting the crowdfunding performance are the number of backers, the investment amounts and the pledge targets; the financing purpose, the project category, the reward commitment, the quality signal of the project, and the investment quota have varying degrees of influence (Yao and Zhang 2014). An online crowdfunding project mainly conveys the quality signal of the project through a text description, so the linguistic style of the text description has a significant impact (Parhankangas and Renko 2017).

Online financing projects reduce information asymmetry through textual descriptions (Miglo and Miglo 2019); therefore, the factors related to project textual signals have become the key factors affecting the financing performance of crowdfunding projects (Gafni et al. 2019). Adopting appropriate strategies can greatly improve project financing performance such as strategies related to the social network of financiers (Laurell et al. 2019; Rey-García et al. 2019) and the geography of the campaign (Brent and Lorah 2019). In addition, investment preferences and personalized recommendations for crowdfunding projects are also important factors in fundraising outcomes.

### 2.2 Research on prediction algorithms

From the perspective of data mining, researchers have proposed a variety of prediction algorithms. These numerous algorithms are applicable to a variety of contexts, and the performance of the algorithms is at varying levels (Chen et al. 2020). Algorithms have both advantages and disadvantages, and no algorithm dominates in all fields (Wang et al. 2017a). Models have been established by using Bayesian classification (Pareek et al. 2019), decision trees, support vector machines, K-nearest neighbors, and C4.5, and these algorithms have been compared using evaluation criteria such as the accuracy. Experiments have shown that these predictive algorithms have both advantages and disadvantages in different fields and for different sample sizes (Wahbeh et al. 2011). These findings indicate that in the algorithm selection, it is necessary to select an algorithm according to the characteristics of the sample.

Take the commonly used K-nearest neighbors algorithm and the logistic regression algorithm as examples. These two algorithms are applicable to all kinds of scenarios. The stochastic gradient descent is usually used to solve the parameters of a logistic regression, and the stochastic gradient descent accelerates the convergence of the calculation (Bach 2014). For online financing, the financing results either succeed or fail, so the logistic regression is applicable to the analysis and prediction of crowdfunding performance. The logistic regression and its nonlinear extensions such as multilayer feedforward neural networks can be viewed as converting input or higher-level features into mass functions and aggregating them by combining rules. The probability output of these classifiers is the likelihood of normalization corresponding to the basic combined mass function, so the mass function usually provides more information than the output probability distribution (Denoeux 2019).

The random forest can be used to overcome the disadvantages of overfitting and low stability of a single decision tree (Rokach 2016). In the variable selection, LDA is used to identify text semantics, which can significantly improve

the financing performance forecasting power (Kaminski et al. 2019). Furthermore, the accuracy of the prediction algorithm is also affected by the data itself. For unbalanced data, the performance of prediction algorithms such as the logistic regression, neural network, and decision tree is significantly different. The random forest and gradient enhanced classifiers perform well with unbalanced data. However, the C4.5 decision tree, quadratic discriminant analysis, and the K-nearest neighbors are poor when they process the same data set (Brown and Mues 2012).

In recent years, deep learning algorithm has achieved great success. In many fields, deep learning algorithms based on neural network have achieved preferable prediction results (Li et al. 2020). Compared with the shallow algorithms, the deep learning algorithm provides an end-to-end solution and improves the efficiency. Convolutional neural network (CNN) can learn a continuous representation of the input sequence in a iteration way. A study shows that the two-layer CNN with dropout achieves the best prediction results for text mining (Saumya et al. 2019). The prediction power of deep learning has been verified in many fields. For example, in geographical information system (GIS), power distribution and supply are complex model. When deep learning is used for modeling nonlinear network structure, the prediction accuracy is improved and a better optimization algorithm is obtained (Xue et al. 2019). Similarly, in the prediction of human vision, deep learning shows the advantages of video processing (Wang et al. 2019b). It proves the applicability of deep learning in many fields. In the training of deep learning, distributed and parallel extreme learning machine is often used to improve the prediction performance, and Bayesian model is used to fuse algorithms to achieve real-time prediction (Yao and Ge 2019).

A few studies try to predict the fundraising outcomes of crowdfunding projects with deep learning. In medical crowdfunding, heterogeneity exists in donations across cases, and fundraisers face uncertainty; thus, the motives of investors are different as well. Deep learning could capture the individual differences in a better way (Wang et al. 2019a). The online investment intention is complex, including both the donation recurrence behavior and the donor retention behavior, which poses a technical challenge for predicting the investment behavior. Deep learning provides a solution for predicting the completion recurrence and donor retention on Kiva (Zhao et al. 2019). And the prediction power of fundraising outcomes can be greatly improved by containing text, visual, and linguistic signals (Kaminski and Hopp 2019).

### 2.3 Research on crowdfunding prediction

As investment willingness is affected by many factors (Ulo et al. 2019), adopting appropriate strategies would greatly improve financing prediction such as strategies related to the social network of fundraisers (Laurell et al. 2019; Rey-García et al. 2019) and the geography of the campaign (Brent and Lorah 2019). Based on the statistical characteristics of crowdfunding projects, algorithms such as decision trees and support vector machines are used to construct predictive models, and the financing success rate is predicted before the project is officially launched. The accuracy of the model reaches 68% (Greenberg et al. 2013). Text is one of the most important factors affecting the fundraising outcomes because the project signals are always transmitted by text descriptions. Many researchers analyze the quality signals of online financing projects through linguistic features (Wang et al. 2017b). Based on a text-based framework, potential semantics can be extracted from text descriptions, and numerical features can be used to predict the financing performance of a project by a random forest algorithm (Yuan et al. 2016). The accuracy of the regression model based on a logistic regression is 76.7% by considering social attributes. The characteristics of crowdfunding projects can be divided into static features and dynamic ones. According to the static and dynamic properties of the project, the support vector machine can be used to predict the number of backers with an accuracy of 84% (An et al. 2014). Based on time series, K-nearest neighbors and Markov chain can be used to predict the financing success rate during the fundraising duration, which solves the dynamic prediction problem to some extent (Etter et al. 2013).

Crowdfunding is a relatively easy way for entrepreneurs to obtain capital contribution (Wu et al. 2019), but it is also full of uncertainty (Vismara 2019). In reward-based crowdfunding, potential customers support new unverified products, and entrepreneurs conduct early product testing and market validation, which provides quality signals to potential investors. There is a long-term relationship between crowdfunding and venture capital at the industry level. A comparison of 77,654 Kickstarter projects with 3260 venture capital projects in USA between 2012 and 2017 shows that successful crowdfunding activities led to an increase in subsequent venture capital, especially in hardware-related, electronics-related, and fashion-related projects. These results enhance our understanding on the development of crowdfunding and venture capital, and reward-based crowdfunding helps venture investors assess future trends rather than crowding them out of the market (Kaminski et al. 2019).

### 3 Experimental data and preprocessing

#### 3.1 Data source and preprocessing

The experimental data come from Kickstarter, the world’s largest reward-based crowdfunding platform. Kickstarter was chosen because it is a leading global crowdfunding platform, making the prediction results more reliable and comparable, and it provides many attributes for training and testing the models. The sample includes 85,233 crowdfunding projects (the observation window is from 2015-01-01 to 2018-11-29). Each sample consists of 37 fields, including the number of investors, categories, fundraiser profiles, launch time, fundraising status, etc. Among all the projects, there are 42,927 successful campaigns and 42,306 failed ones, with each group accounting for approximately 50%. Thus, the sample is a balanced data set. The launch time and the end time are recorded by the time stamps in the original data, which are converted into the standard time format in the preprocessing.

For some features (such as category and geographical location), regular expressions are used to extract the necessary information. Research shows that crowdfunding is affected by geographical distance, so the study takes the geographical location of the project as a variable.

There are some missing values in the original data. Through the correlation coefficient thermogram, those features with less influence on the dependent variable are discarded. Finally, all discrete variables are transformed into one-hot format. A matrix of (85,233; 12) dimensions is obtained, which includes 85,233 samples, and each sample is presented by 12 variables, as shown in Table 1. As category variable is transformed into one-hot format, the data trained in the algorithms have more columns than those present in Table 1.

#### 3.2 Exploratory data analysis

Exploratory data analysis (EDA) is an analysis method that explores the distribution and law of a sample under the minimum prior conditions by making charts or other ways. Its main task is to explore the internal characteristics and quantitative relationships of the data and to show the valuable information that is hidden in the data intuitively by using data visualization when the variable changes. With the aid of EDA, we can clearly check the missing values, abnormal values, redundancy, and imbalance of the variables. We use the *pyecharts* for the preprocessing. Table 2 shows the statistical results of the dataset, from which we can see that successful projects tend to have lower pledge targets, higher funded capital amounts, longer funding durations, and more backers. The average pledge target of failed projects is approximately 9 times more than that of successful projects, but the funded capital amount is only approximately 1/15 of that of successful projects. Thus, failed projects positively correlated with high goals, low durations, and low backers, while successful projects have slightly higher funding durations than failed projects.

If the statistics are grouped by category, as shown in Fig. 1, the proportion of Art and Comics projects is the largest, and the categories with the fewest projects are Technology and Theater. The funding success rate for each category is shown in Fig. 2. It can be seen that the success rates of Comics, Dance, and Publishing are the highest at 79%, 76%, and 66%, respectively. Further, Food, Design, and Technology are the lowest with less than 40%. Art, Comics, and Dance are the most popular categories, and the project financing success rates are high. Therefore, it is necessary to consider the project category in predicting crowdfunding fundraising outcomes.

The distributions of successful financing projects and failed projects are compared with logarithmic transformations, as shown in Figs. 3, 4, and 5. Compared with failed

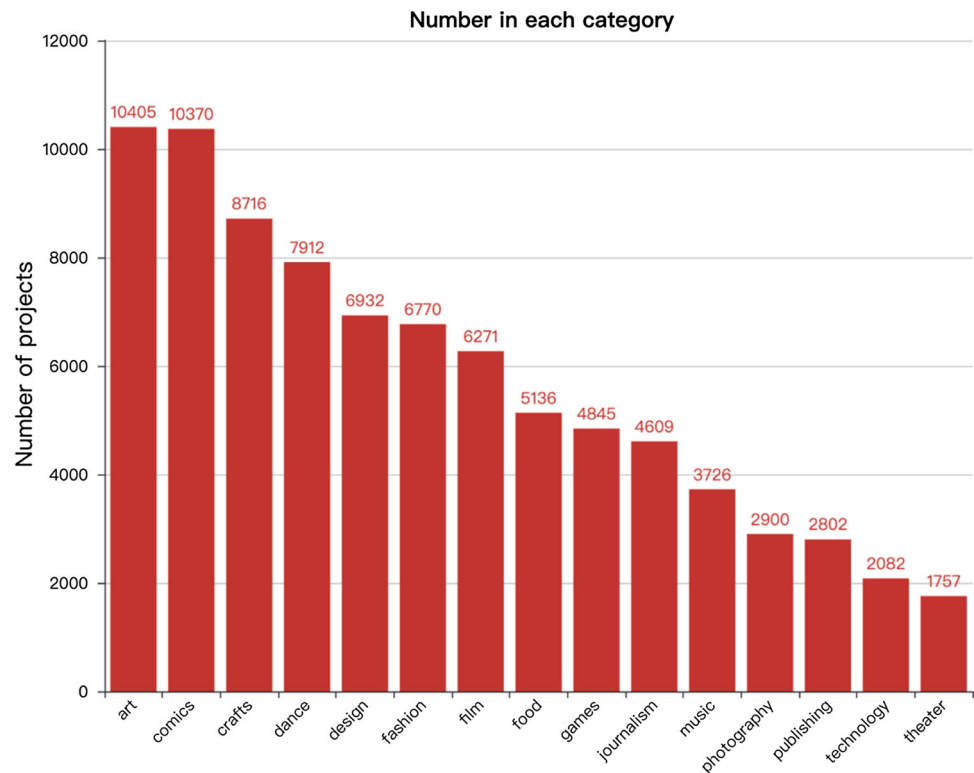
**Table 1** Features and descriptions

Feature	Description
NumBackers	Number of backers
Category	Category of the campaign
Country	Founder’s country
NumCurrency	Campaign’s currency
NumRate	The exchange rate at the time of the project’s initiation, compared to the US dollar
NumGoal	Pledge goal
NumDuration	Funding duration
NumDay	The day of the month that the project is launched
Continent	The continent of the project
City	City of the campaign
State	State of the campaign
Status	1 = success, 0 = failed



**Table 2** Statistical results of the dataset

	Successful	Failed	Total
Count	42,927	42,306	85,233
Proportion (%)	50.36%	49.64%	100%
Total goal	537,985,777	4,596,636,966	5,134,622,743
Goal proportion (%)	10.48%	89.52%	100%
AVG. goal	12,532	108,652	60,242
Amount pledged	1,148,779,075	72,058,145	1,220,837,220
AVG. pledge	26,761	1703	14,332
Pledged versus goal	2.13	0.01	0.24
AVG. Duration (days)	67.96	61.42	64.71
Number of backers	9,604,078	505,887	10,109,965

**Fig. 1** Number of projects per category

projects, successful projects often have advantages in terms of the numbers of investors and the pledged capital amounts. For failed campaigns, the pledge goals are much higher than those of successful financing projects. Figure 6 shows the distribution of project durations, and it shows non-significant difference between successful projects and failed ones.

In order to investigate whether geographical location affects crowdfunding performance, the geographical distribution is depicted, as shown in Fig. 7. It can be seen that most of the projects are launched in North American and European countries. The country where the most projects are launched is the USA, accounting for 67.5%, followed by the UK and Canada.

Since most of the projects are from the USA, the distribution among states is further examined, as shown in Fig. 8. It can be seen that the largest numbers of projects are launched in California, New York, and Texas, accounting for about one-third of the US-initiated campaigns, and in the remaining states, the numbers of launched projects are balanced. From a geographical perspective, the number of crowdfunding projects seems to have a certain relationship with economic status. Therefore, it can be considered that the geographical location of the project is also a factor in determining fundraising outcomes.

Fig. 2 Success ratio per category

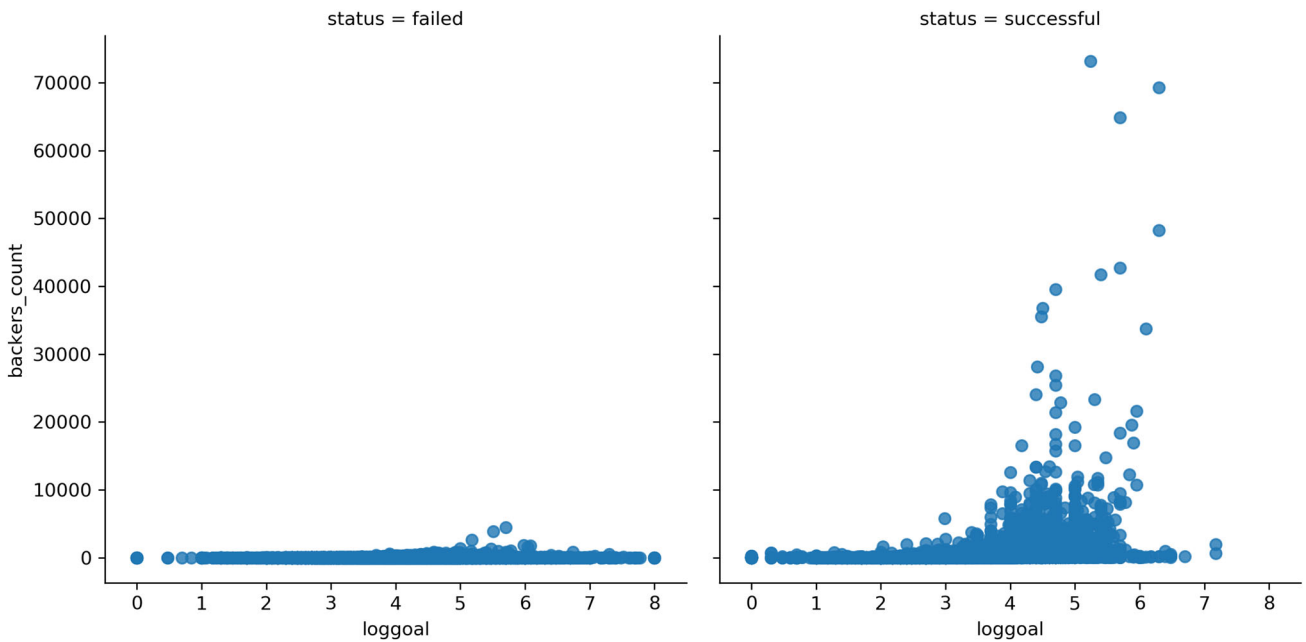
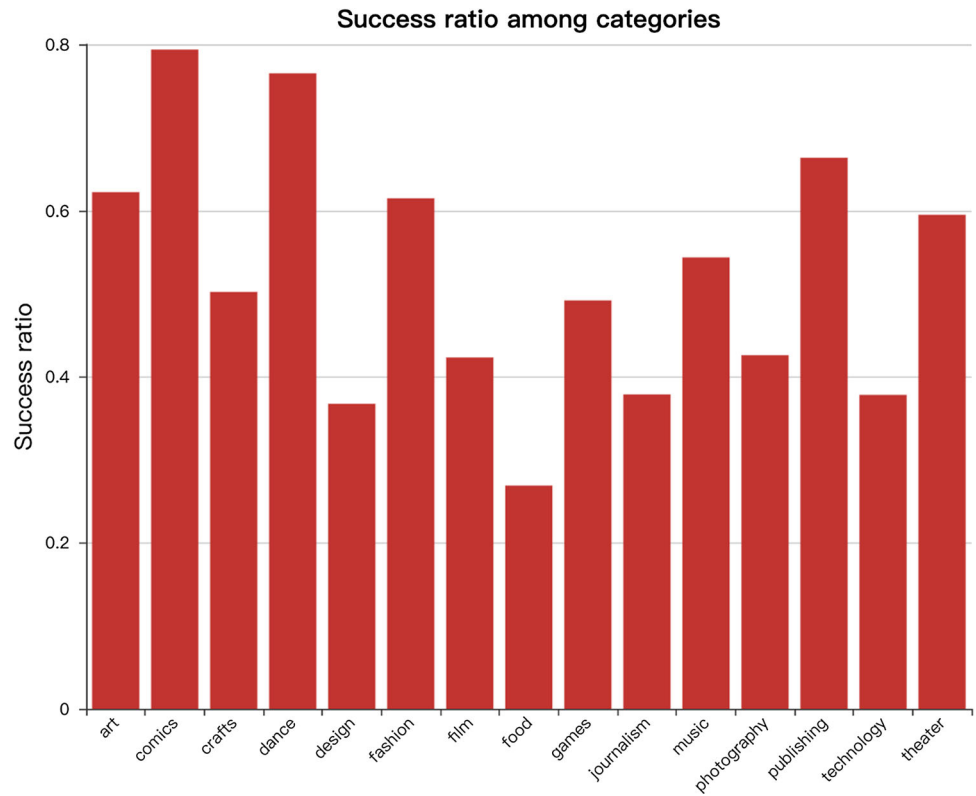
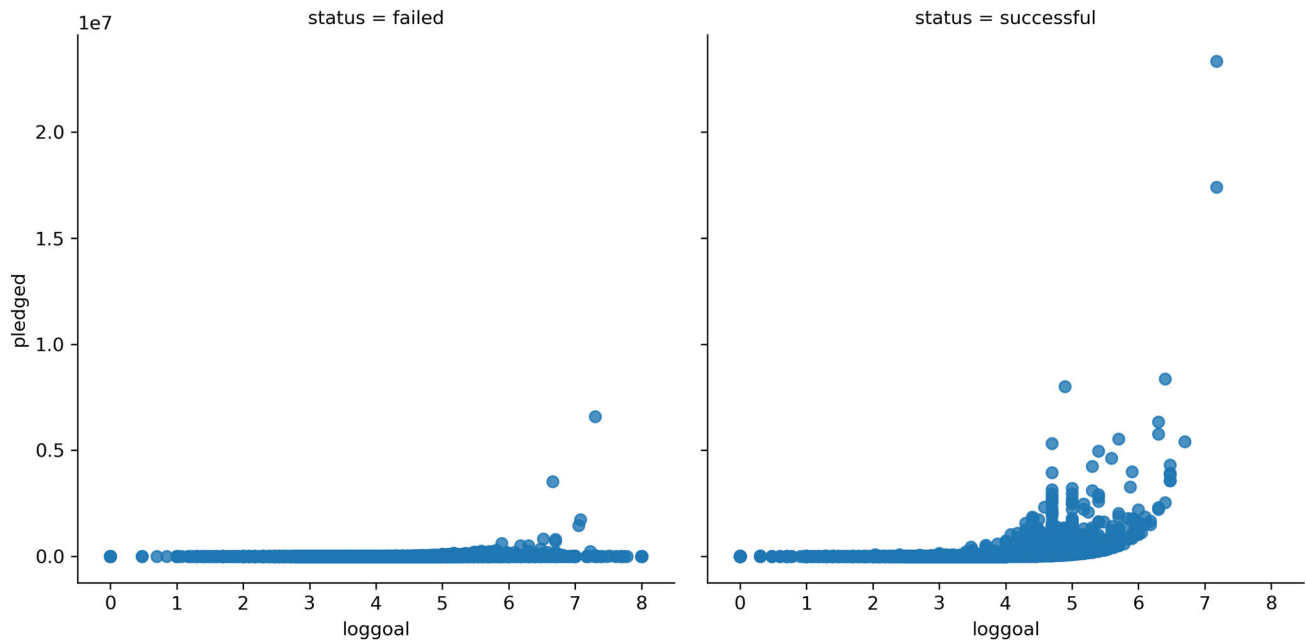
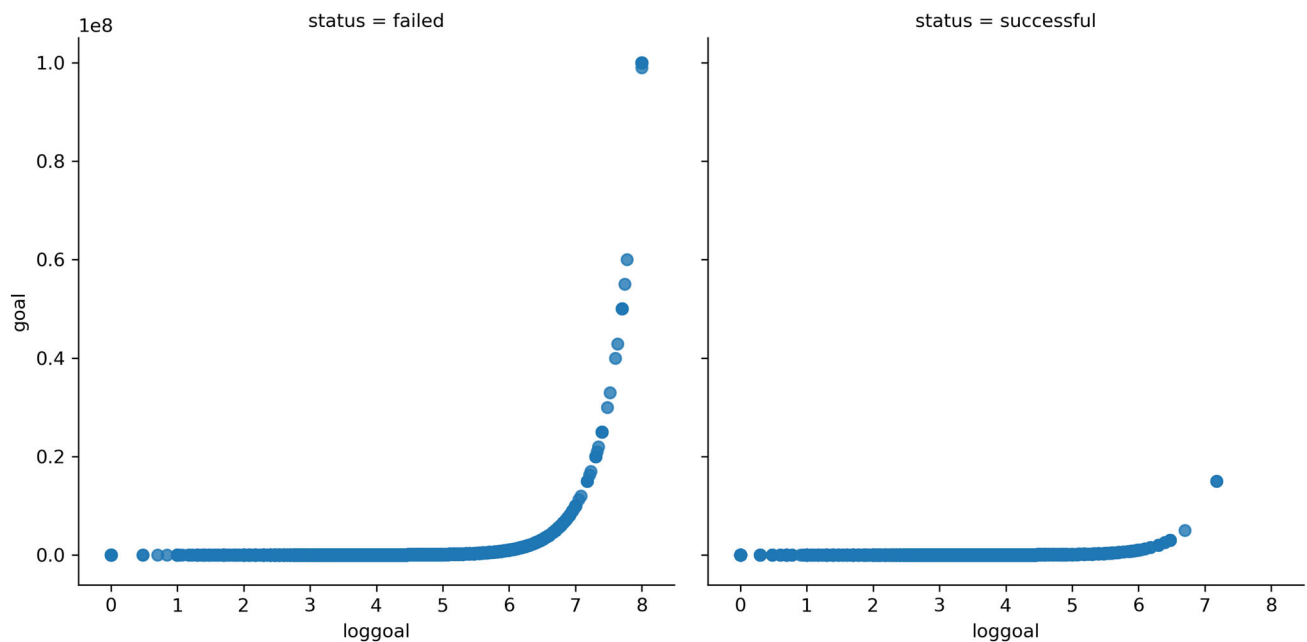


Fig. 3 Distributions of investors



**Fig. 4** Distributions of pledged money



**Fig. 5** Distributions of pledge targets

## 4 Experimental setting

### 4.1 Comparison algorithms

Different classification models have been proposed in extant studies. Since Kickstarter adopts the all-or-nothing financing model, each project is either a financing success or failure. Therefore, the prediction in this scenario is

obviously a classification issue. The machine learning prediction models for binary classification include the decision tree, logistic regression, support vector machine, random forest, KNN, and multilayer perceptron (MLP). Since few studies have focused on the MLP's power on predicting Internet finance projects, we focus on exploring the MLP's predictive power for crowdfunding projects.



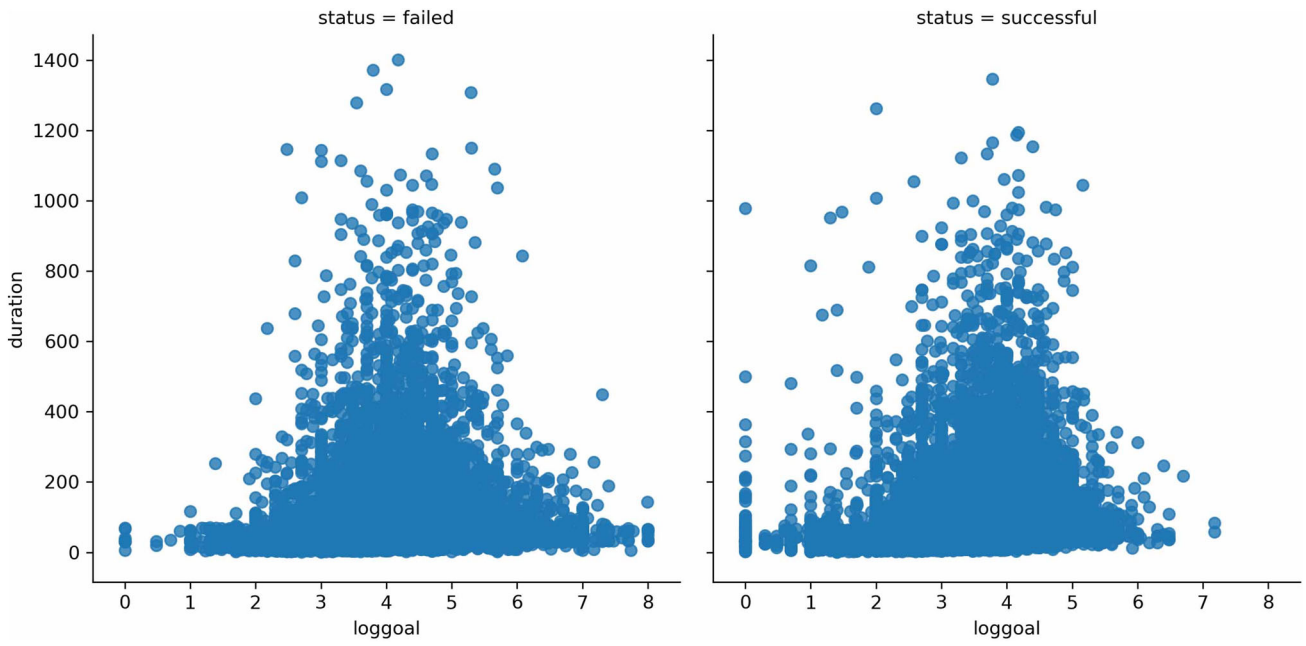


Fig. 6 Distributions of duration

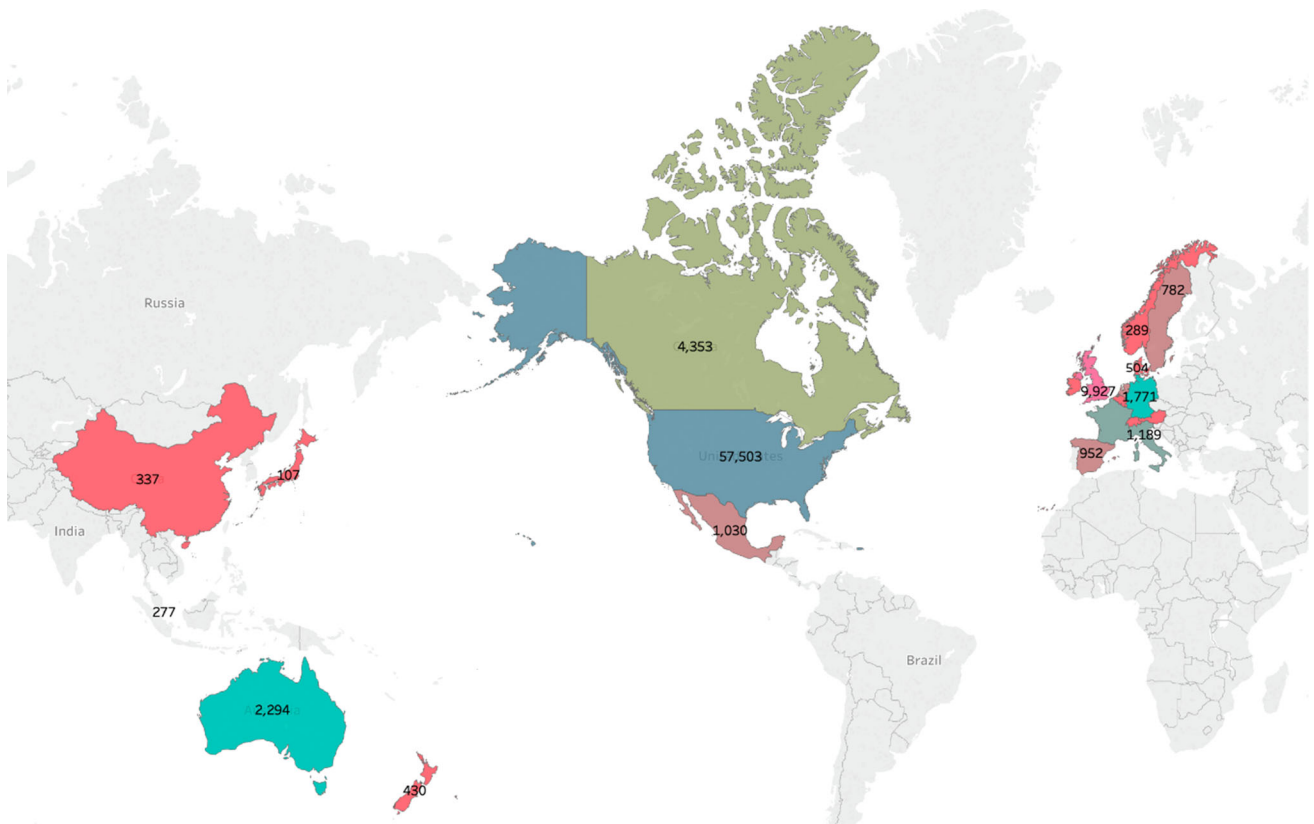


Fig. 7 Number of projects among countries



generated, they are independent of each other and have no causal relationship with each other (Mantas et al. 2019). A random forest is defined as a set of untrimmed classification or regression trees that are trained in the tree generation process using random feature selection to train the bootstrapped samples of the training data. After generating a large number of trees, each tree “votes” to select the most influential classes, which are collectively referred to as the random forest. For random forest classification techniques, two parameters need to be adjusted, namely the number of trees and the number of attributes that are used to generate the trees. The eigenvalues of the nodes are usually split by Gini impurity, and the tree searches through all the features that are used for segmentation to minimize the impurity. Equation (4) shows the calculation of the Gini impurity:

$$IG(p) = 1 - \sum_{i=1}^J p_i^2 \tag{4}$$

where  $p_i$  represents the proportion of the data points in category  $i$ .

#### 4.1.5 KNN

The K-nearest neighbors (KNN) algorithm classifies data points by the plurality of the “votes” of the  $k$  most similar data points. The similarity measurement that is used in this study is the Euclidean distance between two points, as shown in Eq. (5), where  $x_i$  and  $x_j$  are 2 data points:

$$d(x_i, x_j) = \|x_i - x_j\| = [(x_i, x_j)^T(x_i, x_j)]^{1/2} \tag{5}$$

#### 4.1.6 Multilayer perceptron

Deep learning provides an end-to-end approach and has the advantage of not requiring strict feature engineering, which provides a promising means of prediction for multiple domains. As a kind of deep learning, the MLP is also called the feedforward neural network (FNN). It is a forward-structured artificial neural network and is also the most commonly used neural network to solve nonlinear problems. In the MLP, at least a three-layer structure (input layer, hidden layer, and output layer) is used, and each neuron belongs to a different layer. Each layer of neurons can receive the signal from the previous layer and generate a signal to the next layer. The signal passes through the hidden layer from the input layer to the output layer, which is a one-way propagation process, as Fig. 9 shows. The process is essentially a simulated neuron that implements a mapping from  $x$  to  $y$ , namely  $x \xrightarrow{w} f(w \times x + b) \rightarrow y$ .

Since the data in reality are often linear and inseparable, in order to express the complex data distribution, it is

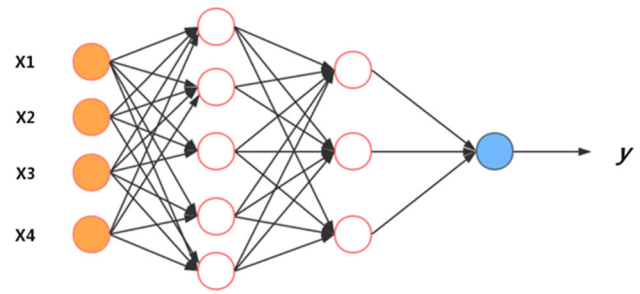


Fig. 9 Multilayer perceptron structure

necessary to introduce a nonlinear model to map the original data. This nonlinear model is called the activation function, which is not available in the built-in neural network. It is also an important determinant of improving the performance of the neural network. In general, the activation function should have the following characteristics: (1) The function should be continuous and nonlinear, which means the derivable activation function can be solved by numerical optimization when performing parameter learning; (2) the function should be as simple as possible, which is beneficial for improving the calculation rate of the model; and (3) the derivative of the activation function should be controlled within a certain range. If the value is too large or too small, the calculation efficiency and stability of the network will be decreased. A commonly used activation function is the sigmoid function (logistic and tanh), as shown in Eqs. (6) and (7), respectively:

$$\sigma(x) = \frac{1}{1 + \exp(-x)} \tag{6}$$

$$\tanh(x) = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)} \tag{7}$$

After the structure of the neural network is defined, the number of neurons in its output, hidden layers, and output layers is determined as well, and then, its weight  $W$  and bias  $b$  need to be determined. The training process determines the weights and biases. By continuously correcting the weights and bias values, the predicted values of the network model are maximized to approximate the true values.

The learning process is as follows: First, the algorithm randomly assigns all weights and biases and random value set to predict the samples in the training set. The results are marked as  $\hat{y}$ , and the true value of the sample is  $y$ . Then,  $W$  and  $b$  are continuously adjusted according to the gap between  $\hat{y}$  and  $y$ . The loss function is introduced to evaluate the error of the training process, and the definition is as shown in Eq. (8). The loss function is a nonnegative real number. For a prediction model, the goal is to make the difference between the predicted value and the true value

smaller. That is, its goal is to minimize the value of the loss function:

$$\text{loss} = (\hat{y} - y)^2 \quad (8)$$

The loss function value is optimized by the gradient descent until the result finally converges. When the loss function value is less than a certain threshold, parameters  $W$  and  $b$  can be determined, and all parameters of the neural network are obtained. Then, the preprocessed data are input into the neural network to obtain the predicted results and the results are compared with the true values.

## 4.2 Algorithm implementation

TensorFlow (1.9.0) is employed as the back-end for deep learning. Since Keras (2.1.6-tf) provides a convenient way to call TensorFlow, it is easy to implement an MLP neural network. The experimental operating system is macOS 10.15 with 8 GB 2133 MHz LPDDR3 memory; GPU is Intel Iris Plus Graphics 655 1536 MB. Using the sequential model in Keras, multiple network layers are linearly stacked to form an MLP model. In the sequential module, the input layer, hidden layer, and output layer and the number of neurons in each layer need to be defined, respectively. The input layer of the MLP model in this study has 11 neurons. After continuous fitting, it was found that the model works best when using two hidden layers and when there are 10 neurons in the first hidden layer and 5 neurons in the second hidden layer. We use the ReLU as the activation function and the  $l_2$  regularization parameter with penalty terms to avoid overfitting.

For the output layer, it accepts values from the last hidden layer and converts them to output values to represent the probability of “success” according to the activation function, which is SoftMax function. In the compilation process, we need to specify the optimizer and loss function of the MLP and the metric that is used to evaluate the performance of the model. At this stage, “Adam” is selected as the optimization function, and “accuracy” is adopted as the evaluation criterion. The optimal parameters of the model are obtained by the grid search.

The following algorithms are implemented by the scikit-learn (0.20.3): logistic regression, decision tree, random forest, SVM, and KNN. In the above algorithms, the grid search is used for the parameter tuning, and cross-validation is employed to avoid overfitting. Table 3 shows the optimal parameters for the algorithms.

**Table 3** Optimal parameters for algorithms

Algorithm	Parameters
Logistic regression	solver = “saga”, penalty = “l1”
Decision tree	criterion = “gini”
Random forest	n_estimators = 2, criterion = “entropy”
SVM	kernel = “rbf”, gamma = 0.1, C = 1
KNN	n_neighbors = 29

## 5 Experimental results and discussion

### 5.1 Evaluation criteria

The confusion matrix is often used to evaluate the predictive power in two-class classification, and the structure is shown in Table 4. The row corresponds to the true value, and the column corresponds to the predicted value.

Based on the confusion matrix, the evaluation criteria that are adopted in this study are the accuracy, precision, recall, and F-score, and their calculations are shown in Eqs. (9)–(12). As the comparison of precisions across models should be conditional on the same level of recalls and the comparison of recalls across models should be conditional on the same level of precisions, the average precision-recall score (APRS) is employed to compare the prediction on the condition of the same level of recalls (Carroll et al. 2010), as shown in Eq. (13), where  $R_n$  and  $P_n$  are the precision and recall at the  $n$ -th threshold. The Matthews correlation coefficient (MCC) is used in machine learning as a measure of the quality of binary and multi-class classifications (Saqlain et al. 2019). It takes into account true and false positives and negatives and is generally regarded as a balanced measure which can be used even if the classes are of very different sizes. The MCC can be calculated directly from the confusion matrix, as shown in Eq. (14):

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (9)$$

**Table 4** Confusion matrix

Confusion matrix		Predict value	
		Positive	Negative
True value	Positive	TP	FN
	Negative	FP	TN



$$\text{Precision} = \frac{TP}{TP + FP} \tag{10}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{11}$$

$$F - \text{score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{12}$$

$$\text{APRS} = \sum_n (R_n - R_{n-1}) P_n \tag{13}$$

$$\text{MCC} = \frac{TP \times TN + FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{14}$$

To comprehensively evaluate the performance of the algorithms, we use both the ROC (receiver operating characteristic) and the AUC (area under the curve). The ROC curve uses the false positive rate (FPR) as the abscissa and the true positive rate (TPR) as the ordinate. A lower FPR or a higher TPR for the ROC indicates better predictive power. Therefore, the ROC curve for a good model always closes to the upper left corner. The AUC is the area under the ROC curve. Normally, the value of the AUC is between [0.5, 1]. AUC = 0.5 indicates that the model “guessed” half of the sample right. It is a perfect model when AUC = 1 where all the samples are “guessed” right.

Cross-validation (CV) is commonly used in machine learning to compare and select a model for a given predictive modeling problem because it generally has a lower bias than other methods. The aim of cross-validation is to ensure that every example from the original dataset has the same chance of appearing in training or testing sets. Given a set of  $m$  examples, the widely used  $K$ -fold cross-validation partitions them equally into  $K$  sets; each of  $K - 1$  is for training the classifier, while the rest are for testing (Wong and Yang 2017). As most extant studies employed tenfold cross-validation (Parisi et al. 2018), we employ a tenfold cross-validation to divide the data into 10 parts randomly, and to train the model on other 9 parts, and finally test the model on the rest 1 part. And we repeat the experiment ten times with different combination; the averaged result for the 10 experiments is adopted as the final results.

## 5.2 Experimental results

Figure 10 shows the learning curve for the MLP. After epoch = 5, the accuracies of the model for the training set and the test set tend to be stable at approximately 0.9 without overfitting. As the epochs increase, the loss value continues to decrease, and it is eventually below 0.5. This

finding demonstrates the usefulness of the MLP and suggests that the algorithm does not overfit.

To intuitively view the performance of the algorithms at predicting fundraising outcomes, Table 5 shows the accuracy, precision, recall, F-score, AUC, APRS, and MCC values for the algorithms. It can be seen that the performance of the MLP is the best among all 6 models, and the model with the second-best performance is the decision tree. Further, logistic regression, SVM, and KNN have almost same predictive powers for crowdfunding campaigns.

Figure 11 shows the comparison of prediction results. From the accuracy, MLP has the highest accuracy of 92.3%, while KNN has the lowest accuracy of 85.2%. But in terms of precision, among all the comparison algorithms, the highest precision is random forest with 93.2%, while the lowest precision is logistic regression with 85.7%. Although the precision of MLP is not the highest, the recall of MLP reaches a very high level of 96.4%. F-score is a comprehensive evaluation criterion, which combines precision and recall. The F-score of MLP algorithm reaches the maximum of 0.921, which indicates the usefulness of deep learning algorithm for crowdfunding project fundraising outcomes prediction. The highest APRS is random forest with 0.898, followed by MLP (0.881) and decision tree (0.869). From the perspective of Matthews correlation coefficient, the MCC of random forest (0.860) and MLP (0.854) is higher than that of KNN (0.713), logistic regression (0.711), and SVM (0.709). From the results, deep learning has potential for the prediction of crowdfunding projects, and random forest has also achieved good prediction results.

Figure 12 shows the confusion matrix for the MLP prediction results. It can be seen that the prediction algorithm has a strong classification power. However, the MLP is more capable of predicting negative cases than positive ones; that is, the MLP is inclined to predict successful projects as failures.

Figure 13 shows the ROC curve for each algorithm. The ROC of the MLP model is the closest to the upward left corner. Based on the above comprehensive indicators, it can be concluded that the MLP can achieve better predictive results than other commonly used machine learning approaches in terms of crowdfunding fundraising outcome predictions. The AUC value is 97.3%, which indicates that the data can be well fit by deep learning.

Many studies have shown that the linguistic features have an impact on the fundraising outcomes of crowdfunding projects (Parhankangas and Renko 2017). Therefore, we try to feed the variables related to the linguistic features of the narratives into the prediction model to observe the improvement of the prediction power. The text description of the crowdfunding project is employed as the

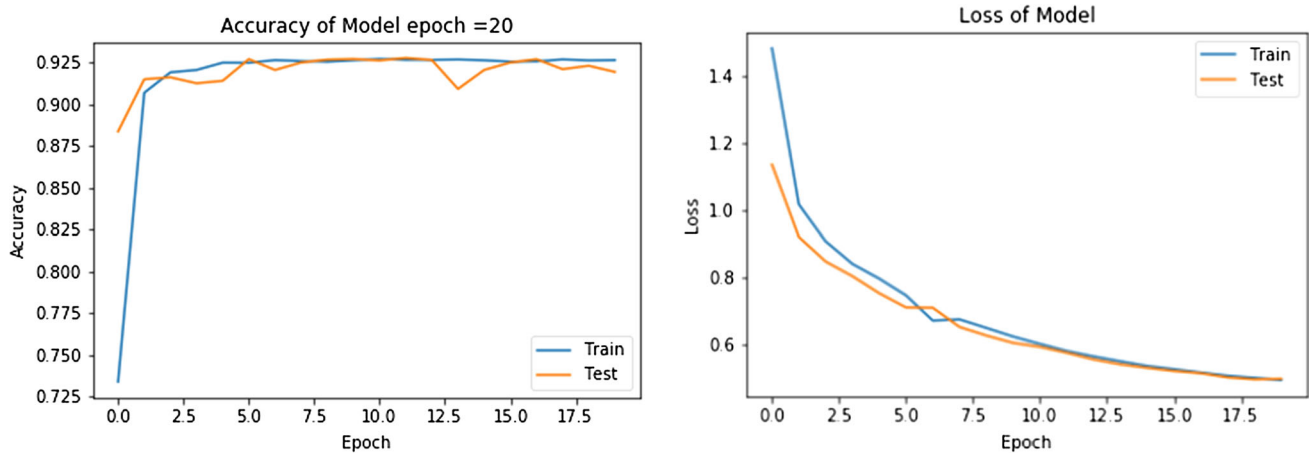


Fig. 10 Learning curve for the MLP model

Table 5 Performance of models

Algorithm	Accuracy	Precision	Recall	F-score	AUC	APRS	MCC
MLP	0.923	0.883	0.964	0.921	0.973	0.881	0.854
Decision tree	0.908	0.898	0.905	0.901	0.900	0.869	0.813
Random forest	0.891	0.932	0.846	0.887	0.939	0.898	0.860
Logistic regression	0.853	0.857	0.852	0.852	0.940	0.815	0.711
SVM	0.853	0.885	0.808	0.845	0.938	0.811	0.709
KNN	0.852	0.858	0.854	0.854	0.930	0.789	0.713

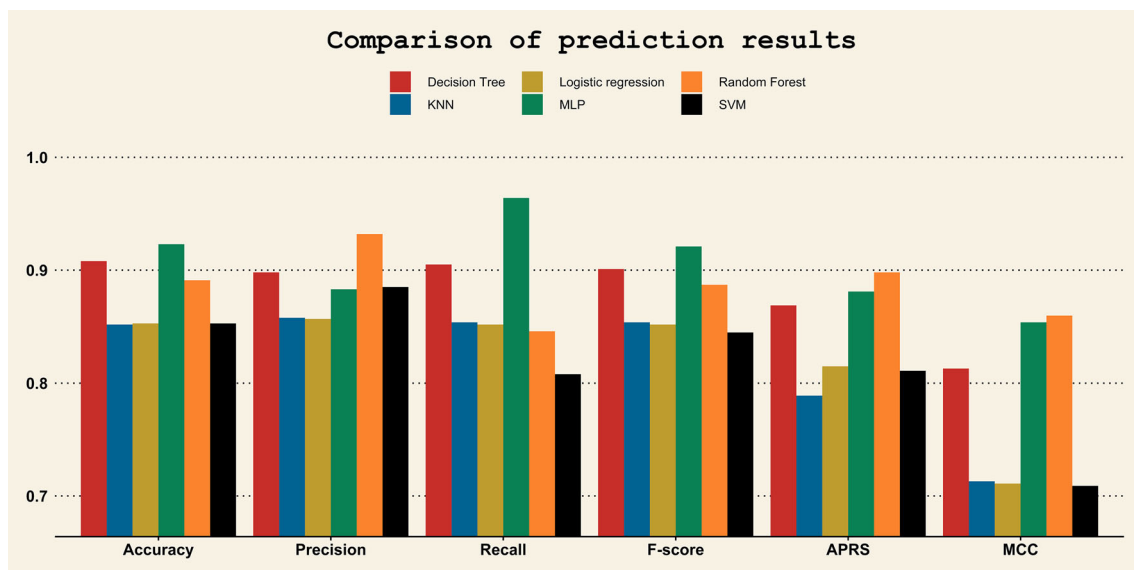


Fig. 11 Comparison of prediction results

corpus. Firstly, we preprocess the corpus, namely convert all the text into lowercase, remove punctuation, and stop words. Then, we transform the text into the document-word matrix. LDA (Blei et al. 2003) is employed to detect text topics. LDA is an unsupervised topic classification; thus, we start from 20 topics to output the results and the top 20 keywords of each topic and observe the topic clustering

results. And we reduce the number of topics if the topics overlap until there is no overlap.

Finally, five topics are obtained: The first topic is the progress-related description; the second topic is the reward-related description; the third topic is the content-related description; the fourth topic is the promotion-related description; and the fifth topic is the description of the



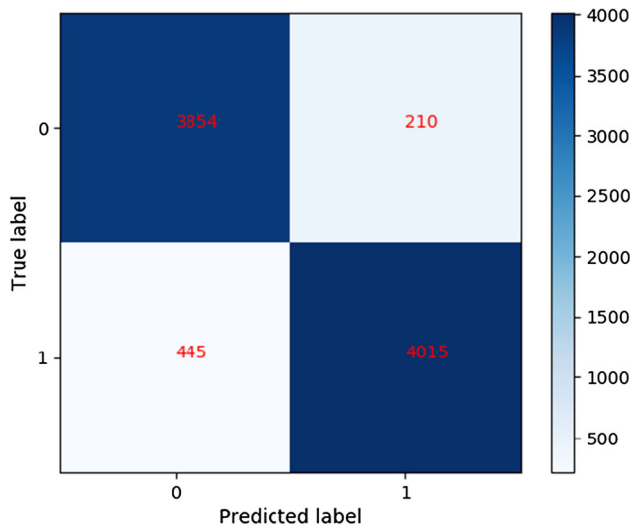


Fig. 12 Confusion matrix for the prediction results

use of fund. Each document belongs to these five topics with a certain probability, and we choose the topic with the highest probability as the topic of the campaign. Figure 14 shows the probability of each topic. It can be seen that many fundraisers introduce the project content in the project description, accounting for 34.42%, followed by the project progress description and promotion description with 21.08% and 16.93%, respectively. However, the description of the use of fund is the least (12.81%).

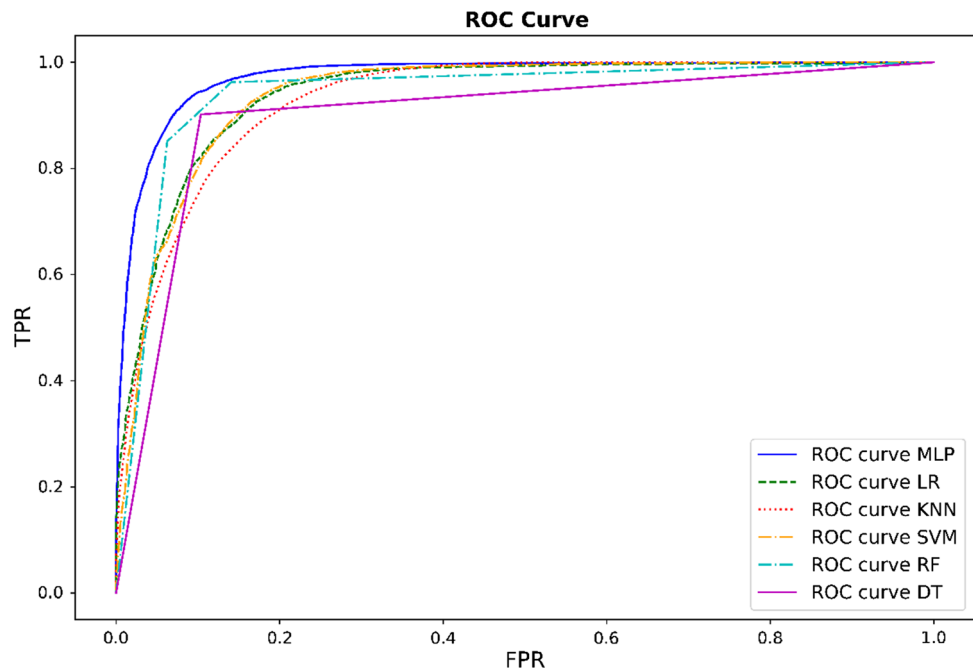
Table 6 shows the performance comparison of models with topic model. From the comparison, LDA can improve the prediction in most algorithms. For example, in deep

learning (MLP), topic model improves the prediction performance in accuracy, precision, F-score, AUC, APRS, and MCC. Take Matthews correlation coefficient (MCC) as an example, it increases from 0.854 to 0.864. However, recall decreases from 0.964 to 0.946 when the topics of the narratives are included in MLP as predictor variables.

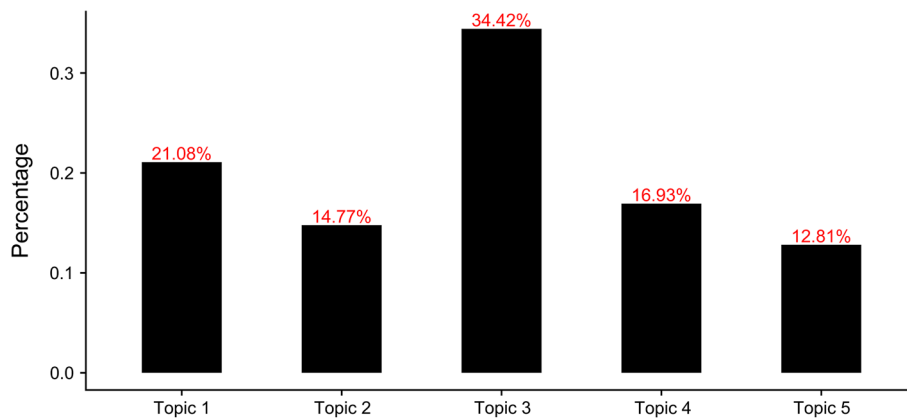
Figure 15 presents the comparison of prediction results with LDA and without LDA. From the results, LDA can improve the prediction for KNN, logistic regression, MLP, but not for SVM and decision tree. For the random forest, the performance of the prediction decreases when LDA is fed in the model. The results indicate that the topic model has different adaptability among prediction algorithms; some of them are improved, while others cannot get the help of LDA.

In the previous experiment, we used a number of backers in the prediction model. However, a prediction model will be not that helpful by using the variable that is performance dependent (i.e., number of backers in this case). Therefore, we try to remove the variable of number of backers in the prediction model and then re-fit the model. The results are shown in Table 7. It can be seen that there are some differences among algorithms. MLP still shows advantages in recall, F-score, and AUC. However, from the perspective of average precision-recall, logical regression has greater advantages. Overall, deep learning has a higher prediction performance even when the number of backers is removed; the result shows the strong adaptability of deep learning to the feature selection.

Fig. 13 ROC curve for each model

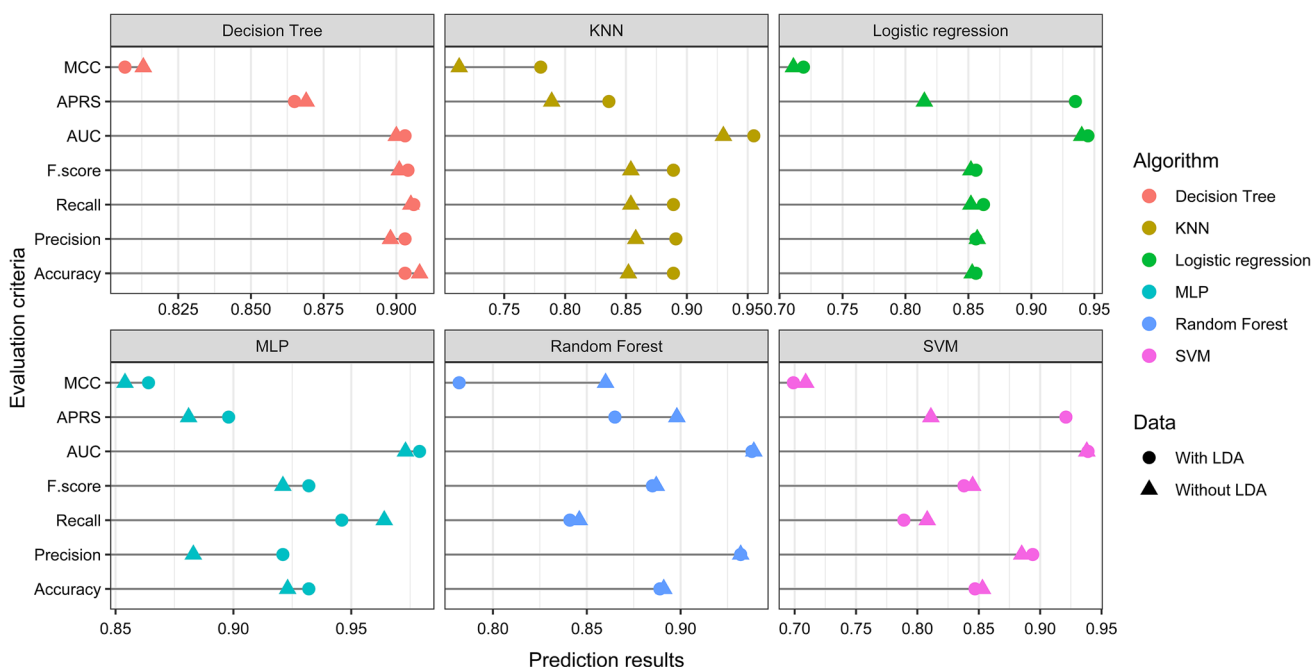


**Fig. 14** The probability of each topic



**Table 6** Performance of models with topic model

Algorithm	Accuracy	Precision	Recall	F-score	AUC	APRS	MCC
MLP	0.932	0.921	0.946	0.932	0.979	0.898	0.864
Decision tree	0.903	0.903	0.906	0.904	0.903	0.865	0.8067
Random forest	0.889	0.932	0.841	0.885	0.938	0.865	0.782
Logistic regression	0.856	0.856	0.862	0.856	0.945	0.935	0.719
SVM	0.847	0.894	0.789	0.838	0.939	0.921	0.699
KNN	0.889	0.891	0.889	0.889	0.955	0.836	0.780



**Fig. 15** Comparison of prediction results with LDA and without LDA

**5.3 Practical implications**

As prediction promotes the development of innovation and entrepreneurship (Biljohn and Lues 2019), the MLP can be used as a user-oriented tool in management. The tool can be used as a real-time predictor of fundraising outcomes to

guide financiers to improve the attributes of their projects before their projects are officially launched and to predict the possibility of successful funding and to reduce the opportunity costs. It can also be used for investor forecasting since it can select the most likely successful projects to invest in, which can avoid projects with high failure

**Table 7** Performance of models (remove the number of backers)

Algorithm	Accuracy	Precision	Recall	F-score	AUC	APRS	MCC
MLP	0.850	0.769	0.999	0.869	0.962	0.769	0.735
Decision tree	0.596	0.606	0.574	0.574	0.6	0.563	0.193
Random forest	0.574	0.618	0.427	0.505	0.619	0.553	0.161
Logistic regression	0.853	0.857	0.852	0.852	0.94	0.931	0.711
SVM	0.618	0.665	0.461	0.544	0.693	0.691	0.245
KNN	0.855	0.858	0.855	0.854	0.93	0.789	0.713

risks. It can also be used for crowdfunding platforms before projects go online, and the projects that may be successful funded can be recommended to investors. Thus, it improves the power to the knowledge management needed for online financing (Briceno and Santos 2019). Moreover, the factors that result in lower success ratios can be summarized to guide the founders to improve their projects.

In addition, we can not only predict the possibility of binary fundraising outcomes, but we can also dynamically predict the other aspects of projects by integrating the temporal information (updates, comments, etc.) of each ongoing project such as the number of project backers, the pledged capital amount, and funding progress. Since the deep learning model results in a better predictive power and generalization ability than other machine learning algorithms, it provides management implications. By using hardware acceleration technology, it is possible to process large amounts of data and obtain the predicted results. Because the deep learning algorithm reduces the dependence on feature engineering and obtains end-to-end prediction results, it helps us to promote real-time management practice. In future managerial applications, powerful deep learning algorithms such as the MLP, convolutional neural networks, and cyclic neural networks can be used to solve practical problems.

## 6 Conclusion and prospects

This study first performs preprocessing and exploratory analysis based on Kickstarter data and then introduces a deep learning algorithm that few researchers have used before. The multilayer perceptron (MLP) is applied to predict crowdfunding project fundraising outcomes, and the results are compared with those generated by other machine learning algorithms, including the decision tree, random forest, logistic regression, support vector machine, and K-nearest neighbors algorithms. The data set is divided by 10-fold cross-validation, and the grid parameters are employed to optimize the parameters and hyperparameters. The same evaluation criteria are chosen for different models. The accuracy, precision, recall, F-value, AUC, APRS, and MCC are calculated from the confusion matrix.

The experimental results show that the MLP model has the best performance in crowdfunding fundraising outcome predictions, and the accuracy reaches 92.3%. Second, the decision tree's accuracy is 90.8%, and the worst performing model is the K-nearest neighbors. There are huge differences between the advantages and disadvantages among algorithms. From the comparison of advantages, MLP model is not always superior to other algorithms. Among evaluation indicators, MLP model has advantages in accuracy, recall, F-score, and AUC, but for precision, MLP model lags behind the random forest (93.2%) and the decision tree (89.8%). Random forest shows advantages on APRS and MCC. The result indicates that random forests and decision trees have a stronger power to predict positive cases (projects that have been successfully financed), but present shortcomings for the negative cases (projects that have failed to fund), and random forests and decision trees have shown significant advantages in precision.

From the point of view of the disadvantages of algorithms, the precision of MLP is smaller than any other evaluation criteria with 88.3%, while other evaluation criteria are all greater than 90%. Similarly, the precision of decision tree is the lowest (89.8%) in its evaluation criteria. The recall of random forest is the lowest (84.6%) as well. The accuracy, precision, recall, and F-score of logistic regression are all lower than 86%. The recall of SVM is the lowest among all algorithms with 80.8%, which means SVM has a poor predictive coverage for the successfully funded projects, that is, SVM is difficult to predict those successfully financed projects. However, it indicates that SVM has a better prediction power for financing failure projects due to the high precision value of SVM (88.5%). The accuracy, precision, recall, and F-score of KNN are approximately 85%. Moreover, the MCC of SVM is lowest with 0.709, and the APRS of KNN is lowest with 0.789. The results demonstrate the potential of deep learning in the prediction of crowdfunding project fundraising outcomes and provide inspiration for guiding follow-up research and management practices.

Although we have evaluated many prediction algorithms and made many comparisons, there are still some shortcomings. We only tested the dummy outcomes (success or failure) of project fundraising. In fact, there are many other

fundraising outcome evaluation criterions such as forecasting the number of investors, forecasting the pledge progress, and forecasting the pledged capital. According to the prediction results, guiding founders to set reasonable attributes is also a potential future direction. For example, we may guide founders to set the optimal funding targets and durations to maximize their respective success ratios. In addition, this study only considers the basic attributes of the crowdfunding project (category, funding target, geographical location, etc.). In future study, it is also necessary to consider the social attributes of the founders (Twitter, Facebook or Flickr), the dynamic attributes of the project (updates, comments), the social promotion attributes (number of followers), etc., to comprehensively examine how various factors affect fundraising outcomes.

**Acknowledgements** This work is partially supported by the NSFC Grant (71601082), Natural Science Foundation of Fujian Province (2017J01132), Huaqiao University's High Level Talent Research Start Project Funding (16SKBS102) and Teaching development reform project for Huaqiao University teachers (17JF-JXGZ17) and Ministry of Science & Technology, Taiwan (MOST 108-2511-H-003 -034 -MY2).

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical approval** This article does not contain any studies with human participants or animals performed by any of the authors.

## References

- An J, Quercia D, Crowcroft J (2014) Recommending investors for crowdfunding projects. In: Proceedings of the 23rd international conference on World wide web, 2014. ACM, pp 261–270
- Bach FR (2014) Adaptivity of averaged stochastic gradient descent to local strong convexity for logistic regression. *J Mach Learn Res* 15:595–627
- Biljohn MIM, Lues L (2019) Social innovation and service delivery in Belgium and South Africa. *Transf Gov People Process Policy* 13:143–158
- Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. *J Mach Learn Res* 3:993–1022
- Brent DA, Lorah K (2019) The economic geography of civic crowdfunding. *Cities* 90:122–130
- Briceno CEB, Santos FCA (2019) Knowledge management, the missing piece in the 2030 agenda and SDGs puzzle. *Int J Sustain High Educ* 20:901–916
- Brown I, Mues C (2012) An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Syst Appl* 39:3446–3453
- Carroll HD, Kann MG, Sheetlin SL, Spouge JL (2010) Threshold average precision (TAP-k): a measure of retrieval designed for bioinformatics. *Bioinformatics* 26:1708–1713
- Chaney D (2019) A principal-agent perspective on consumer co-production: crowdfunding and the redefinition of consumer power. *Technol Forecast Soc Chang* 141:74–84
- Chen X, Wang H, Ma Y, Zheng X, Guo L (2020) Self-adaptive resource allocation for cloud-based software services based on iterative QoS prediction model. *Future Gener Comput Syst* 105:287–296
- Chinnaswamy A, Papa A, Dezi L, Mattiacci A (2019) Big data visualisation, geographic information systems and decision making in healthcare management. *Manag Decis* 57:1937–1959
- Denoeux T (2019) Logistic regression, neural networks and dempster-shafer theory: a new perspective. *Knowl-Based Syst* 176:54–67
- Etter V, Grossglauser M, Thiran P (2013) Launch hard or go home! Predicting the success of Kickstarter campaigns. In: Proceedings of the first ACM conference on Online Social Networks (COSN'13), vol CONF. ACM, pp 177–182
- Gafni H, Marom D, Sade O (2019) Are the life and death of an early-stage venture indeed in the power of the tongue? Lessons from online crowdfunding pitches. *Strateg Entrepreneurship J* 13:3–23
- Gong X-L, Liu X-H, Xiong X, Zhuang X-T (2019) Forecasting stock volatility process using improved least square support vector machine approach. *Soft Comput* 23:11867–11881
- Greenberg MD, Pardo B, Hariharan K, Gerber E (2013) Crowdfunding support tools: predicting success & failure. In: CHI'13 extended abstracts on human factors in computing systems. ACM, pp 1815–1820
- Kaminski JC, Hopp C (2019) Predicting outcomes in crowdfunding campaigns with textual, visual, and linguistic signals. *Small Bus Econ*. <https://doi.org/10.1007/s11187-019-00218-w>
- Kaminski J, Hopp C, Tykvová T (2019) New technology assessment in entrepreneurial financing-does crowdfunding predict venture capital investments? *Technol Forecast Soc Chang* 139:287–302
- Kraus S, Richter C, Brem A, Cheng C-F, Chang M-L (2016) Strategies for reward-based crowdfunding campaigns. *J Innov Knowl* 1:13–23
- Kromidha E, Robson P (2016) Social identity and signalling success factors in online crowdfunding. *Entrepreneurship Reg Dev* 28:605–629
- Laurell C, Sandström C, Suseno Y (2019) Assessing the interplay between crowdfunding and sustainability in social media. *Technol Forecast Soc Chang* 141:117–127
- Li Y, Yang B, Yan L, Gao W (2020) Energy-aware resource management for uplink non-orthogonal multiple access: multi-agent deep reinforcement learning. *Future Gener Comput Syst* 105:684–694
- Long W, Lu Z, Cui L (2019) Deep learning-based feature engineering for stock price movement prediction. *Knowl-Based Syst* 164:163–173
- Mantas CJ, Castellano JG, Moral-García S, Abellán J (2019) A comparison of random forest based algorithms: random credal random forest versus oblique random forest. *Soft Comput* 23:10739–10754
- Miglo A, Miglo V (2019) Market imperfections and crowdfunding. *Small Bus Econ* 53:51–79
- Mollick E (2014) The dynamics of crowdfunding: an exploratory study. *J Bus Ventur* 29:1–16
- Pareek B, Liu Q, Ghosh P (2019) Ask your doctor if this product is right for you: a bayesian joint model for patient drug requests and physician prescriptions. *J R Stat Soc Ser A Forthcom* 182:197–223
- Parhankangas A, Renko M (2017) Linguistic style and crowdfunding success among social and commercial entrepreneurs. *J Bus Ventur* 32:215–236
- Parisi L, RaviChandran N, Manaog ML (2018) Decision support system to improve postoperative discharge: a novel multi-class classification approach. *Knowl-Based Syst* 152:1–10
- Rey-García M, Calvo N, Mato-Santiso V (2019) Collective social enterprises for social innovation. *Manag Decis* 57:1415–1440

- Rokach L (2016) Decision forest: twenty years of research. *Inf Fusion* 27:111–125
- Saqlain SM, Sher M, Shah FA, Khan I, Ashraf MU, Awais M, Ghani A (2019) Fisher score and Matthews correlation coefficient-based feature subset selection for heart disease diagnosis using support vector machines. *Knowl Inf Syst* 58:139–167
- Saumya S, Singh JP, Dwivedi YK (2019) Predicting the helpfulness score of online reviews using convolutional neural network. *Soft Comput*. <https://doi.org/10.1007/s00500-019-03851-5>
- Simons A, Kaiser LF, vom Brocke J (2019) Enterprise crowdfunding: foundations, applications, and research findings. *Bus Inf Syst Eng* 61:113–121
- Testa S, Nielsen KR, Bogers M, Cincotti S (2019) The role of crowdfunding in moving towards a sustainable society. *Technol Forecast Soc Chang* 141:66–73
- Ulo KLM, Hidayanto AN, Sandhyaduhita PI, Fitriani WR, Abidin Z (2019) Factors influencing internet users' intention to sign e-petitions. *Transform Gov People Process Policy* 13:257–275
- Vismara S (2019) Sustainability in equity crowdfunding. *Technol Forecast Soc Chang* 141:98–106
- Wahbeh AH, Al-Radaideh QA, Al-Kabi MN, Al-Shawakfa EM (2011) A comparison study between data mining tools over some classification methods. *Int J Adv Comput Sci Appl* 8:18–26
- Wang W, Tan G, Wang H (2017a) Cross-domain comparison of algorithm performance in extracting aspect-based opinions from Chinese online reviews. *Int J Mach Learn Cybernet* 8:1053–1070
- Wang W, Zhu K, Wang H, Wu Y-CJ (2017b) The impact of sentiment orientations on successful crowdfunding campaigns through text analytics. *IET Softw* 11:229–238
- Wang T, Jin F, Cheng Y (2019a) Early predictions for medical crowdfunding: a deep learning approach using diverse inputs. [arXiv:1911.05702](https://arxiv.org/abs/1911.05702)
- Wang W, Shen J, Xie J, Cheng M-M, Ling H, Borji A (2019b) Revisiting video saliency prediction in the deep learning era. *IEEE Trans Pattern Anal Mach Intell*. <https://doi.org/10.1109/TPAMI.2019.2924417>
- Wong T-T, Yang N-Y (2017) Dependency analysis of accuracy estimates in k-fold cross validation. *IEEE Trans Knowl Data Eng* 29:2417–2427
- Wu YJ, Chen S-C, Pan C-I (2019) Entrepreneurship in the internet age: internet, entrepreneurs, and capital resources. *Int J Semant Web Inf Syst (IJSWIS)* 15:21–30
- Xue J, Wu K, Zhou Y (2019) A risk analysis and prediction model of electric power GIS based on deep learning. *Int J Comput Sci Eng* 18:39–43
- Yao L, Ge Z (2019) Distributed parallel deep learning of hierarchical extreme learning machine for multimode quality prediction with big process data. *Eng Appl Artif Intell* 81:450–465
- Yao H, Zhang Y (2014) Research on influence factors of crowdfunding. *Int Bus Manag* 9:27–31
- Yuan H, Lau RY, Xu W (2016) The determinants of crowdfunding success: a semantic text analytics approach. *Decis Support Syst* 91:67–76
- Zhao H, Jin B, Liu Q, Ge Y, Chen E, Zhang X, Xu T (2019) Voice of charity: prospecting the donation recurrence & donor retention in crowdfunding. *IEEE Trans Knowl Data Eng* 14:1–14

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.