TENCON 2023 - 2023 IEEE Region 10 Conference (TENCON)
31 Oct - 3 Nov 2023. Chiang Mai, Thailand

Leveraging Emotional Features and Machine Learning for Predicting Startup Funding Success

Xiaolu Zhang and Raymond Y.K. Lau (SMIEEE)

Department of Information Systems
City University of Hong Kong
Tat Chee Avenue, Kowloon, Hong Kong SAR
E-mail address: {xzhang2968-c, raylau}@cityu.edu.hk

***Abstract*—Analyzing the crucial factors which help predict startups' funding amounts is important for senior executives of these firms to formulate effective business strategies, which leads to the ultimate startup success. In this research, we crawled real-world startup funding data from the well-known "AngelList" platform which disseminates information about the company profiles of startups, the specific business sectors, and potential investors. Potential investors browse the information posted on AngelList, which may in turn influence their decisions in funding certain startups. Our work aims to evaluate the bundle of factors (e.g., sentiments and emotions embedded in startup profile descriptions, startups' fundamentals, etc.) that may influence startups' funding successes. Moreover, we have examined a variety of state-of-the-art machine learning-based prediction models. In particular, we applied TextCNN, a well-known deep learning method to extract sentimental and emotional features from company profile text to enhance the startup funding prediction task. Our experimental results show that the emotion feature can significantly boost startup funding prediction performance by 12% in terms of F-score, and it is also among the key factors that influence startup funding amounts. To our best knowledge, this work represents the first successful research on examining the relationship emotions captured in company profile text and startup funding success.**

Keywords—Deep Learning; Machine Learning; Emotion Mining; Startup Funding

I. INTRODUCTION

Startups face many problems and challenges in the process of establishment, development and expansion. One of the main challenges that start-up companies first face and must solve is financing, which is the basics and booster for the Startups' survival and development. The success of a startup funding and the level of funding amount to some extent reflect the strength and development potential of a company. The amount of financing also has a crucial impact on the company's future development.

With the development of the economy and society, as well as the continuous emergence of new technologies and business formats, the number of startup companies has surged. According to the calculations in 2017, there are about 137,000 new companies are established every day. However, existing research shows that only 0.25-2% of American startups can obtain venture capital as they wish [1], and almost all startups face difficulties in attracting capital. Predicting whether a startup will succeed in financing and understanding the key factors that affect the amount of financing will help the startup effectively plan and manage the company, enabling it to overcome difficulties further and grow stronger.

Through the literature survey, we found that there are relatively few articles examining the critical factors affecting the startup funding success or predicting success and funding amounts. In fact, previous work mainly focused on the features extracted from structured quantitative data e.g., size of the firm. Most of existing articles identified critical success factors by an ad-hoc manual analysis approach that used linear regressions as the main analysis tool [2][3].

With the development of artificial intelligence, some researchers have gradually introduced machine learning models to analyze unstructured information, such as images and text, and extract new features [4]. However, there are some challenges e.g., the datasets mentioned in papers not publicly available, the varieties of structured versus unstructured data for feature extraction not being explored, and the work being conducted using a small sample only.

So, we aim to build a publicly available dataset and a fusion learning method to improve model feature diversity, and improve the understanding of how unstructured text information and emotion features influence funding results. All of these issues have rarely been studied by previous research. To do so, we crawled the startup data on the AngelList website, attempting to explore the influence of the fusion of basic structured quantitative features and unstructured textual information, providing references for startups seeking to improve funding success rates and funding amounts.

II. RELATED WORK

A. Startup Funding analysis

The team is the basic building block and the essential part of a startup. The founder/team and its development are critical to the survival and growth of a startup [5]. The persuasiveness, integrity, and clear vision of entrepreneurial teams can promote the development of enterprises and teams [6]. Investment based on startup teams is a rational investment strategy for investors. Human assets are of great significance to the early financing and success of startups. In 2017, Bernstein et al. [7] verified this through random experiments on the financing information of startups.

Entrepreneurial experience is one of the main concerns of venture capitalists [8]. Sathaworawong and others [9] found through the investigation of 211 successful financing cases that most successful financing companies have management experience, industry experience, or early entrepreneurial experience. The average working years of company founders in the industry positively impact corporate financing [10]. Previous entrepreneurial experience, appropriate management and professional skills are of great importance for the successful financing of startups. In the company's development process, if the founder's management skills are not enough, hiring experienced managers can promote the company's development [11].

A company's geographic location plays an essential role in startup funding activity. Mollick et al., 2014, [12] using a dataset of more than 48,500 projects, demonstrated that geographic location is correlated with startup financing success. Different regions have different potential talents, which will impact the region's creative productivity. Regions with a higher proportion of creative people have a greater chance of entrepreneurial success. Nahata et al. [13] found that the country where the company is located greatly influences the probability of financing success.

Based on the early financing data of more than 100 startup companies, Miloud and other researchers [14] discussed the differences in the valuation of venture capitalists for early startup companies in different industries through robust statistical analysis in 2012. In 2015, Bocken et al. [15] returned to school through interviews with venture capitalists and relevant experts. Differences in industry development status affect the success of startups, and strong existing industries are their main failure factors.

Investors are the decision makers for the success of financing and also the booster to promote the development of startups. In 2020, Pengnate and Riggins [16] demonstrated the impact of fine-grained emotion of kiva P2P loan description text on loan success. We draw on this research idea to construct the emotion feature.

B. Sentiment and Emotion Detection

Our work is related to a growing literature on Sentiment and emotion detection [17]. We divide emotion detection into 3 categories for summary. Lexicon-based approaches: Lexicon-based approaches involve predefined lists of words and phrases, known as dictionaries, to identify sentiment in text. Each word in the dictionary is scored according to its sentiment polarity (positive or negative), and intensity. Then, the scores for the words in the text are aggregated to determine the overall emotional tone. Widely used dictionaries include SentiWordNet [18][19], OpinionFinder [20], AFINN[21], and LIWC [22]. Dictionaries and rules are often used together, such as the Vader method.

Machine learning-based approaches: Machine learning-based approaches involve training machine learning algorithms, such as support vector machines [23], decision trees [24], and neural networks [25], on large datasets of labeled text data to identify patterns of language use associated with different emotions. Algorithms learn to recognize complex patterns and relationships between words, phrases, and contextual information, classifying text into different sentiment categories.

Pre-trained models-based approaches: Pre-trained models involve machine learning models trained on large text datasets [26]. The models are trained using deep learning algorithms such as neural networks and are designed to identify patterns of language use associated with different emotions. Examples of pre-trained models include BERT[27], GPT-3[28], and ELMo [29].

III. PROPOSED METHOD FOR STARTUP FUNDING PREDICTION

We propose an emotion-based multi-view fusion model for startup funding success prediction. We extract basic sentiment and fine-grained emotions from company descriptions to predict the startup funding success.

A. Data Collection

As mentioned above, the primary challenge in analyzing startup financing is the lack of open and comprehensive datasets. In this section, we introduce the open dataset constructed in this paper, as well as the correspondence between the collected data and qualitative metrics.

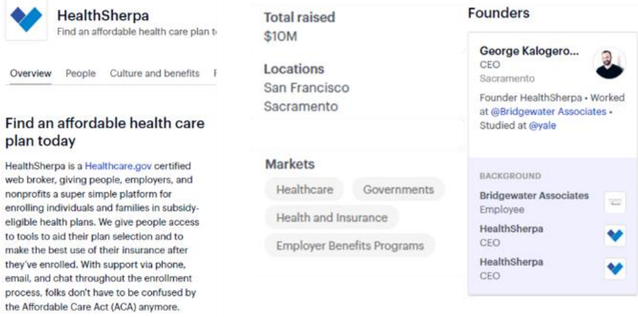

Fig 2. The information items on the website angel.co

AngelList is a leading global information service platform for startups. It provides a market for information sharing for startups to obtain financing, showed in Figure 2. Startup companies can publish their basic information and financing intentions on this platform.In 2022, $1.3B capital raised by Venture Funds. (https://www.angellist.com/2022). We used the Octopus data collector to obtain various information about startups on the AngelList website, including Funding amount, Location of founders, founder Work experiences, Founder team, Company Industries, Location of companies, and Company Profile Text. After completing data cleaning, 139 startups were obtained, of which 2743 companies had founding amounts, accounting for 19.7%。

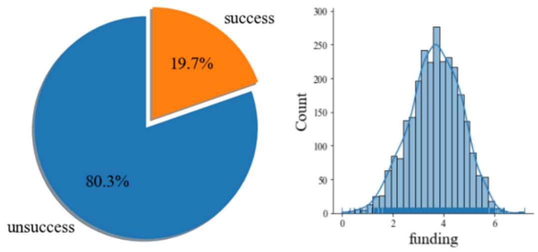

Fig 3. Proportion of funding success & distribution of funding

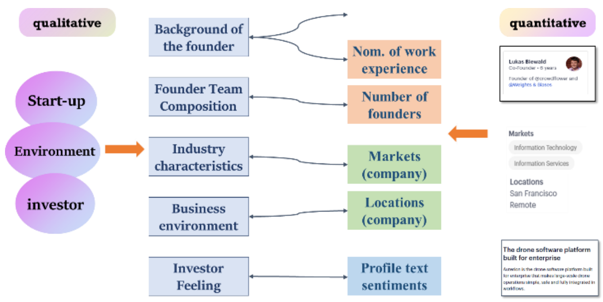

Fig 4. The Mapping relationship between qualitative and quantitative data

B. Feature Extraction

In the related work above, we found that the founder of the startup, the external environment of the startup, and the feeling of investors can affect the funding success. As shown in figure 4, we completed the mapping of qualitative and quantitative indicators based on dataset.

As illustrated in the figure above, we analyze the startup funding from three perspectives. Six types of quantitative data are mapped to five specific qualitative indicators. Among them, founder locations, markets, and company locations are categorical variables. We took the locations and markets that appeared most frequently, and obtained 10 founder locations, 10 company locations, and 20 markets, respectively. Then, we performed one-hot encoding. For startup company C_i, we have three types of features: founder locations $\left(FL_{C_{i_1}}, \ldots, FL_{C_{i_{10}}}\right)$, company locations $\left(CL_{C_{i_1}}, \ldots, CL_{C_{i_{10}}}\right)$, and markets $\left(M_{C_{i_1}}, \ldots, M_{C_{i_{20}}}\right)$. Number of founder work experience and the number of founders are continuous variable FW_{C_i}, FN_{C_i}. The above 5 types of quantitative data collectively constitute Startup basic feature $BF_{C_i} = \left(FL_{C_{i_1}}, \ldots, FL_{C_{i_{10}}}, CL_{C_{i_1}}, \ldots, CL_{C_{i_{10}}}, M_{C_{i_1}}, \ldots, M_{C_{i_{20}}}, FW_{C_i}, FN_{C_i}\right)$. The distribution of number of founders and average work experiences of founders as shown in Figure 5.

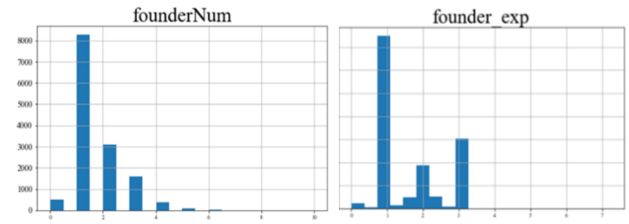

Fig 5. Distribution of founderNum (number of founders, left) and founder_exp (average work experiences of founders, right)

Profile text sentiments are divided into basic sentiments (positive, negative) and fine-grained emotions (7 types: anger, disgust, fear, joy, neutral, sadness,surprise). The sentiment and emotion expressed in the text have transmission characteristics. When humans read a passage, they can sense the emotions conveyed by the author from the text because humans are emotional. For example, investors will feel joy if a company profile outlines the company's bright future and hard work enthusiasm. If the product or service described by the company is uncommon in the current society, or is a special innovation, the text description will often express a kind of surprise emotion, and investors will also feel similar emotions from the text, which will affect investor's decision. Alternatively, suppose the company profile expresses strong dissatisfaction with existing products. In that case, this will also arouse investors' empathy and anger, which has a certain influence on whether investors are willing to invest.

For sentiment features, we use the VADER method to construct. VADER (Valence Aware Dictionary and sEntiment Reasoner) [30] is a sentiment analysis method that uses dictionaries and rules. For startup company C_i, we extracted positive and negative sentiments SP_{C_i}, SN_{C_i} from the company profile text T_{SC_i}.

$$S_{C_i} = \left(SP_{C_i}, SN_{C_i}\right) = VADER\left(T_{C_i}\right)$$

Then, we use the transformer method to build emotion features. Fine-grained emotions require [31][32] a deeper under-standing of the text than essential sentiments. Transformer is a new type of neural network structure based on the attention mechanism, which can learn semantically rich word vectors and contextual relationships, which is conducive to understanding the meaning of the text. We specifically used a pre-trained Roberta model for text emotion extraction. Based on the Transformer structure [31], Roberta has the ability to perceive distant contexts, which is crucial for detecting complex emotions in text. For startup company C_i, we extracted seven emotions (anger EA_{C_i}, disgust ED_{C_i}, fear EF_{C_i}, joy EJ_{C_i}, neutral EN_{C_i}, sadness ES_{C_i}, surprise EU_{C_i}) from the company profile text T_{SC_i}. $E_{C_i} = (EA_{C_i}, ED_{C_i}, EF_{C_i}, EJ_{C_i}, EN_{C_i}, ES_{C_i}, EU_{C_i}) = \mathrm{Roberta}(T_{C_i})$.

TABLE I.
BASIC STATISTICAL DESCRIPTION OF EMOTION SCORES

	anger	*disgust*	*fear*	*joy*	*neutral*	*sadness*	*surprise*
MEAN	0.025	0.011	0.036	0.120	0.740	0.016	0.052
STD	0.046	0.030	0.088	0.168	0.222	0.056	0.062
MIN	0.001	0.000	0.000	0.000	0.002	0.001	0.001
MAX	0.985	0.867	0.979	0.982	0.976	0.984	0.976

We observed that the average score of all emotions were greater than 0.01. In order to obtain strong emotions and retain enough information, we initially set two thresholds of 0.1 and 0.01. The final threshold value will be determined through experimental results.

C. STARTUP FUNDING SUCCESS ANALYSIS

In the previous subsection, for startup company C_i, we first encode its basic features BF_{C_i} and use VADER and Transformer models to obtain S_{C_i} and E_{C_i}. Then, concatenate the above three features. We get vector representation BSE_{C_i}.

$$BSE_{C_i} = \left[BF_{C_i}, S_{C_i}, E_{C_i}\right]$$

For our startup funding dataset, each company is marked with the class label "1" if the it has obtained financing; otherwise, it is marked with the class label "0". In the subsection, we introduced six machine learning methods to predict the startup funding success using vector representation BSE_{C_i}. The methods include:

- Decision Trees [24]: a tree-like structure where each node represents a decision based on the values of a specific feature in the input data.
- Random Forests [33]: an ensemble learning method that combines multiple decision trees to create a more robust and accurate model.
- AdaBoost [34]: Adaptive Boosting, it is an ensemble learning method that combines weak classifiers to create a stronger classifier.
- Gradient Boosting Decision Trees (GBDT) [35]: It uses a boosting approach where subsequent trees are built to correct the errors made by previous trees.
- Support Vector Machines (SVM) [23]: SVM seeks to find a hyperplane that best separates the data points of different classes or predicts the target values of regression tasks.

- Multilayer Perceptron (MLP) [36]: A type of artificial neural network that consists of multiple layers of interconnected nodes (neurons).

To improve the neural network model performance, we construct a sentiment and emotion analysis-based startup funding success prediction fusion model using CNN and MLP. The following Figure 6 shows the overall framework.

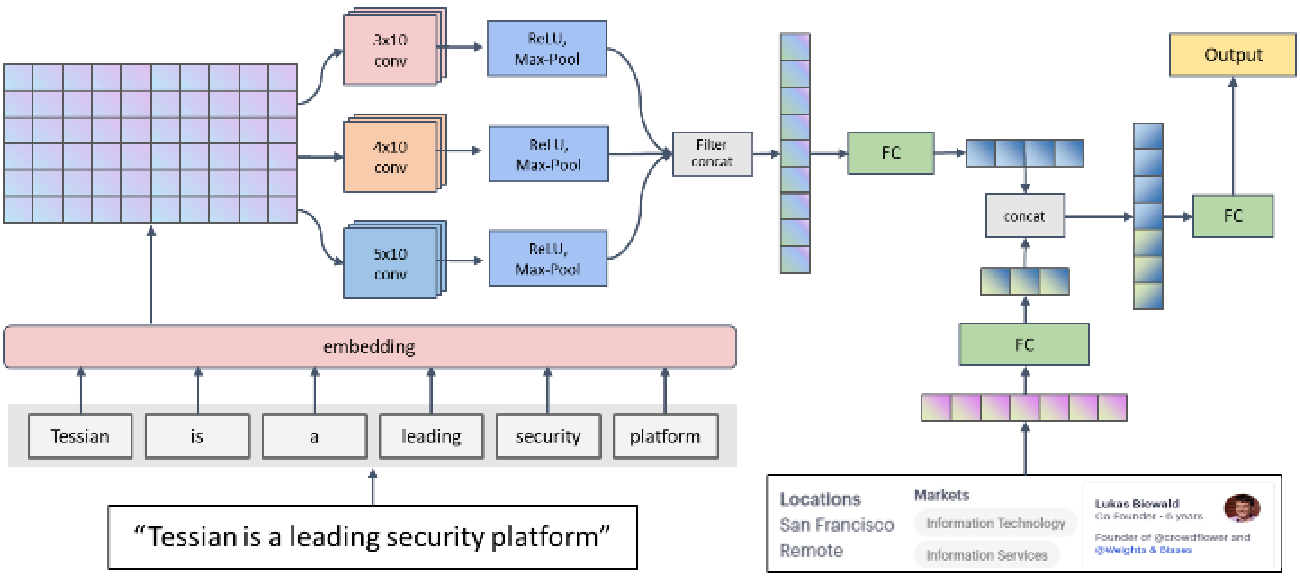

Fig 6. Overview of Startup funding prediction

As shown in the Figure 6, we apply the TextCNN neural network structure to extract the contextual information of the company profile text T_{C_i}, hoping that it implicitly contains the Semantic information of the company's products, fields, development, etc. The company profile text T_{C_i} can be regarded as a sequence of words; each word is represented by a vector x_i (randomly generated) and the dimension of each word embedding is k. So, the text is expressed as a word embedding matrix $X_{C_i} \in \mathbb{R}^{m\times k}$,where m is the length of the text.

$$H_{hj} = \text{maxPooling}\left(\text{Relu}\left(X_{C_i} \otimes V_{hj} + b\right)\right),$$

where convolution kernel $V_{hj} \in \mathbb{R}^{h\times k}, h \in \{2,3,4\}, j = 1, \ldots, 100$, $\otimes$ represents convolution operation. And, Relu(·) is the Rectified Linear Unit. Then, we use the max-over-time pooling operation over the feature map and take the maximum value. So, for each filter, we get the highest value as the most important feature. Since we used three types of filters, each with 100 kernels, we ended up generating a total of 300 features. We concatenate the above features to obtain the semantic vector representation of the text.

$$H_{C_i} = \left(H_{21}, \ldots, H_{2,100}, H_{31}, \ldots, H_{3,100}, H_{41}, H_{4,100}\right)^T \in \mathbb{R}^{300}$$

Next, we then project the BSE_{C_i} and H_{C_i} into the same space, concatenate them, and finally pass through a fully connected network layer for classification prediction.

$$H_{C_i} = \text{Relu}\left(W_1 H_{C_i} + b_1\right),$$

$$BSE_{C_i} = \text{Relu}\left(W_2 BSE_{C_i} + b_2\right),$$

$$H = \left[H_{C_i}, BSE_{C_i}\right], \hat{y} = \text{Softmax}\left(W_3 H + b_3\right)$$

where $W_1 \in \mathbb{R}^{k1\times 300}, W_2 \in \mathbb{R}^{k2\times 300}, W_3 \in \mathbb{R}^{2\times(k1+k2)}, b_1 \in \mathbb{R}^{k1}$, $b_2 \in \mathbb{R}^{k2}, b_3 \in \mathbb{R}^2$ are the trainable weights and biases.

Finally, the classification error is defined by the cross-entropy loss:

$$\text{Loss} = -\frac{1}{n}\sum_{i}^{n}\left(y_1 \log\left(\hat{y}_1\right) + y_2 \log\left(\hat{y}_2\right)\right)$$

IV. EXPERIMENTS AND RESULTS

In this section, we report the evaluation of the proposed model based on our startup funding datasets. We spited the dataset into training set and test set. All models use the five-fold cross-validation method on the training set to determine the optimal parameters and then compare the model performance on the test set. Our code for the implementation is publicly available at https://github.com/AI-lufighting/StartupFunding.

TABLE II.
THE INFORMATION OF TRAINING SET AND TEST SET

Dataset	*#Success*	*#Unsuccess*	*Percentage*
TRAINING	2191	8964	19.64%
TEST	552	2237	19.79%

We adopt five common evaluation metrics to measure the performance: area under the receiver operating characteristic curve (AUC), Accuracy, Precision, Recall and F1-score. AUC describes the probability that the model ranks positive samples more highly than negative samples. In general, a larger AUC implies a better performance. The other four evaluation metrics are calculated as follows:

$$\text{Accuracy} = \frac{TN + TP}{TN + TP + FN + FP}$$

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

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

$$\text{F1} - \text{score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

where TN, TP, FN and FP denote true negatives, true positives, false negatives and false positives, respectively.

A. The experiments on different thresholds

Firstly, we conducted classification experiments using emotion features generated by two different thresholds. The experiment results showed that when the threshold was set to 0.01, AdaBoost and MLP could achieve optimal performance. As shown in Figure 7, the models with threshold 0.01 achieved best AUC. Therefore, in subsequent research, the threshold was fixed at 0.01.

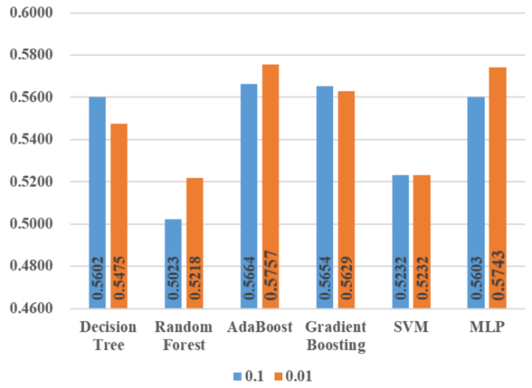

Fig 7. The prediction result on thresholds 0.1 and 0.01

TABLE III.

Model	*Accuracy*	*Recall*	*Precision*	*F1*	*AUC*
Decision Tree	0.8093	0.1141	0.5943	0.1915	0.5475
Random Forest	0.8032	0.0562	0.5254	0.1015	0.5218
AdaBoost	0.8085	0.1902	0.5469	0.2823	0.5757
GBDT	0.811	0.1522	0.5874	0.2417	0.5629
SVM	0.8064	0.0543	0.625	0.1	0.5232
MLP	0.8075	0.1884	0.5389	0.2792	0.5743

According to the experimental results shown in the Table III, all methods have almost the same accuracy, but the other four metrics have significant differences. SVM has the most significant precision, but the recall of this method is only 0.05, which is 1/4 of the Adaboost method and cannot effectively judge positive samples. The AUC of the AdaBoost and MLP models are 0.5757 and 0.5743, respectively. They are better than other models. AdaBoost and MLP are also the best regarding Recall and F1-score metrics and are remarkably higher than other methods.

B. Ablation analysis

To examine the efficacy of sentiment and emotion features of the proposed framework, and we performed an ablation analysis based on the startup funding datasets.

First, we constructed four datasets, namely the all-feature dataset (with the sentiment, emotion, and basic features), the sentiment dataset (sentiment and basic features), the emotion dataset (emotion and basic features), and the basic dataset (excluding sentiment and emotion features). Then, we train the above six machine learning models on these four data sets and comparatively evaluate the model performance.

TABLE IV.
AUC OF ABLATION EXPERIMENT

Model	*All*	*Sentiment*	*Emotion*	*Basic*
Decision Tree	0.5475	0.5632	0.5465	0.5650
Random Forest	0.5218	0.5304	0.5545	0.5300
AdaBoost	0.5757	0.5670	**0.5804**	0.5654
GBDT	0.5629	0.5570	0.5609	0.5606
SVM	0.5232	0.5275	0.5241	0.5275
MLP	0.5743	0.5537	0.5452	0.5623

TABLE V.
F1-SCORE OF ABLATION EXPERIMENT

Model	*All*	*Sentiment*	*Emotion*	*Basic*
Decision Tree	0.1915	0.2476	0.1887	0.2524
Random Forest	0.1015	0.1294	0.2129	0.1312
AdaBoost	0.2823	0.2556	**0.2934**	0.2511
GBDT	0.2417	0.2219	0.2351	0.2336
SVM	0.1000	0.1155	0.1032	0.1155
MLP	0.2792	0.2210	0.1886	0.2493

The results of the ablation experiments are shown in the table above. Tables 4 and 5 give the AUC and F1 scores of the six models on different data sets. The AdaBoost model has the best AUC of 0.5654 based on the basic features. After adding the emotion features, the AUC value is 0.5804 (+1.5%). Considering the F1-score, after adding the emotion feature, the model performance is improved by 4% compared with the basic feature. Therefore, adding emotion features can significantly improve performance.

C. The experiments on fusion model

In order to further improve the performance of the classification model, we introduced the TextCNN and proposed a fusion model suitable for the classification task of startup funding success. Since the TextCNN based model is a deep neural network model, we mainly compared it with the traditional MLP model here.

The experiment results are as follows in Table 6. We conducted three experiments based on imbalanced data, over-sampling data, and under-sampling data. The results indicate that the model based on TextCNN can further improve the performance of the neural network model. Especially when the data is imbalanced, the AUC of the model fused with TextCNN semantic features has increased by 5%, and the F1 score has increased by 12%.

TABLE VI.
ESULT OF NEURAL NETWORK-BASED MODEL EXPERIMENTS

Model	*Accuracy*	*Recall*	*Precision*	*F1*	*AUC*
MLP	0.8075	0.1884	0.5389	0.2792	0.5743
textCNN-based	0.7734	0.3841	0.4206	0.4015	0.6267

V. CONCLUSIONS AND FUTURE WORK

In securing startup funding, sentiments and emotions embedded in company profile text is one of the crucial factors that affect the eventual funding success. The reason is that emotions conveyed in company profile text may largely influence investors who browse startup profile descriptions before making their investment decisions. Based on the

AngelList website, we crawled the data of 14,000 start-ups, including a variety of structured and unstructured data. We then analyzed the key factors of financing success by providing empirical evidence to support the claim that emotions embedded in company profile text may influence startup funding success. In addition, we propose a machine learning-based method for predicting startup funding success and funding amounts. Our experimental results show that emotional features can significantly boost startup funding prediction performance by 12% in terms of F-score. Our research opens the door for startups to polish their emotional profile text to enhance initial funding successes.

ACKNOWLEDGEMENT

This research was partly supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project: CityU 11507219) and CityU SRG (Project: 7005780).

REFERENCES

[1] B. Rose, D. How many start-ups in the US get seed/VC funding per year? http://blog.gust.com/how-many-start-ups-in-the-us-get-seedvc-funding-per-year (accessed January 1, 2023).

[2] Sathaworawong P, Saengchote K, Thawesaengskulthai N. Success factor of start-up fund raising in ASEAN[J]. Asian Administration & Management Review, 2019, 2(2).

[3] Yoo C, Yang D, Kim H, et al. Key value drivers of startup companies in the new media industry—The case of online games in Korea[J]. Journal of Media Economics, 2012, 25(4): 244-260.

[4] Yang K, Lau R Y K. ExeAnalyzer: A deep generative Adversarial Network for Multimodal Online Impression Analysis and startup funding prediction[C]//54th Hawaii International Conference on System Sciences (HICSS 2021). Hawaii International Conference on System Sciences (HICSS), 2021: 2501-2510.

[5] Picken J C. From startup to scalable enterprise: Laying the foundation[J]. Business Horizons, 2017, 60(5): 587-595.

[6] Bussgang, J. Are you suited for a start-up? Harvard Business Review, November–December. https://hbr.org/2017/11/are-you-suited-for-a-start-up (accessed January 1, 2023).

[7] Bernstein S, Korteweg A, Laws K. Attracting early‐stage investors: Evidence from a randomized field experiment[J]. The Journal of Finance, 2017, 72(2): 509-538.

[8] Falik Y, Lahti T, Keinonen H. Does startup experience matter? Venture capital selection criteria among Israeli entrepreneurs[J]. Venture Capital, 2016, 18(2): 149-174.

[9] Sathaworawong P, Thawesaengskulthai N, Saengchote K. Determinant of Startups' Fund-raising Value: Entrepreneur and Firm Characteristic[C]//2018 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM). IEEE, 2018: 1309-1314.

[10] Banerji D, Reimer T. Startup founders and their LinkedIn connections: Are well-connected entrepreneurs more successful?[J]. Computers in Human Behavior, 2019, 90: 46-52.

[11] Prohorovs A, Bistrova J, Ten D. Startup success factors in the capital attraction stage: Founders' perspective[J]. Journal of east-west business, 2019, 25(1): 26-51.

[12] Mollick E. The dynamics of crowdfunding: An exploratory study[J]. Journal of business venturing, 2014, 29(1): 1-16.

[13] Nahata R, Hazarika S, Tandon K. Success in global venture capital investing: do institutional and cultural differences matter?[J]. Journal of Financial and Quantitative Analysis, 2014, 49(4): 1039-1070.

[14] Miloud T, Aspelund A, Cabrol M. Startup valuation by venture capitalists: an empirical study[J]. Venture Capital, 2012, 14(2-3): 151-174.

[15] Bocken N M P. Sustainable venture capital–catalyst for sustainable start-up success?[J]. Journal of cleaner production, 2015, 108: 647-658.

[16] Pengnate S F, Riggins F J. The role of emotion in P2P microfinance funding: A sentiment analysis approach[J]. International Journal of Information Management, 2020, 54: 102138.

[17] Nandwani P, Verma R. A review on sentiment analysis and emotion detection from text[J]. Social Network Analysis and Mining, 2021, 11(1): 81.

[18] Sebastiani F, Esuli A. Sentiwordnet: A publicly available lexical resource for opinion mining[C]//Proceedings of the 5th international conference on language resources and evaluation. European Language Resources Association (ELRA) Genoa, Italy, 2006: 417-422.

[19] Baccianella S, Esuli A, Sebastiani F. Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining[C]//Lrec. 2010, 10(2010): 2200-2204.

[20] Wilson T, Hoffmann P, Somasundaran S, et al. OpinionFinder: A system for subjectivity analysis[C]//Proceedings of HLT/EMNLP 2005 Interactive Demonstrations. 2005: 34-35.

[21] Nielsen F Å. A new ANEW: Evaluation of a word list for sentiment analysis in microblogs[C]//Workshop on'Making Sense of Microposts: Big things come in small packages. 2011: 93-98.

[22] Tausczik Y R, Pennebaker J W. The psychological meaning of words: LIWC and computerized text analysis methods[J]. Journal of language and social psychology, 2010, 29(1): 24-54.

[23] Patil G, Galande V, Kekan V, et al. Sentiment analysis using support vector machine[J]. International Journal of Innovative Research in Computer and Communication Engineering, 2014, 2(1): 2607-2612.

[24] Alshamsi A, Bayari R, Salloum S. Sentiment analysis in English texts[J]. Advances in Science, Technology and Engineering Systems Journal, 2020, 5(6): 1683-1689.

[25] Rajabi Z, Shehu A, Uzuner O. A multi-channel bilstm-cnn model for multilabel emotion classification of informal text[C]//2020 IEEE 14th International Conference on Semantic Computing (ICSC). IEEE, 2020: 303-306.

[26] Adoma A F, Henry N M, Chen W. Comparative analyses of bert, roberta, distilbert, and xlnet for text-based emotion recognition[C]//2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP). IEEE, 2020: 117-121.

[27] Catelli R, Pelosi S, Esposito M. Lexicon-based vs. Bert-based sentiment analysis: A comparative study in Italian[J]. Electronics, 2022, 11(3): 374.

[28] Alexandridis G, Varlamis I, Korovesis K, et al. A survey on sentiment analysis and opinion mining in greek social media[J]. Information, 2021, 12(8): 331.

[29] HUANG C, ZHAO Q. Sensitive information detection method based on attention mechanism-based ELMo[J]. Journal of Computer Applications, 2022, 42(7): 2009.

[30] Hutto C, Gilbert E. Vader: A parsimonious rule-based model for sentiment analysis of social media text[C]//Proceedings of the international AAAI conference on web and social media. 2014, 8(1): 216-225.

[31] Yuan K, Zhang M. Dual-Channel Microblog Emotion Analysis Based on RoBERTa-WWM and Multi-Head Attention[C]//Advances in Natural Computation, Fuzzy Systems and Knowledge Discovery: Proceedings of the ICNC-FSKD 2021 17. Springer International Publishing, 2022: 486-495.

[32] Lai Y, Zhang L, Han D, et al. Fine-grained emotion classification of Chinese microblogs based on graph convolution networks[J]. World Wide Web, 2020, 23: 2771-2787.

[33] Karthika P, Murugeswari R, Manoranjithem R. Sentiment analysis of social media network using random forest algorithm[C]//2019 IEEE international conference on intelligent techniques in control, optimization and signal processing (INCOS). IEEE, 2019: 1-5.

[34] Rathi M, Malik A, Varshney D, et al. Sentiment analysis of tweets using machine learning approach[C]//2018 Eleventh international conference on contemporary computing (IC3). IEEE, 2018: 1-3.

[35] Qiu Y, Song Z, Chen Z. Short-term stock trends prediction based on sentiment analysis and machine learning[J]. Soft Computing, 2022, 26(5): 2209-2224.

[36] Acheampong F A, Wenyu C, Nunoo‐Mensah H. Text‐based emotion detection: Advances, challenges, and opportunities[J]. Engineering Reports, 2020, 2(7): e12189.