



# Prediction Models for Startups Success: A Empirical Analysis

Ramakrishna Allu, Venkata Nageswara Rao Padmanabhuni

**Abstract:** *Small and Medium-Scale Enterprises have been recognized by the government due to their significant role in the country's economy. The risk of capital investment is high in the enterprises and various factors need to be properly analyzed for the prediction of success of an enterprise. Machine learning techniques can be adopted to predict the success of startups that helps the entrepreneur to make a decision accordingly. In this paper, a details analysis has been carried out on the existing methodologies on startup success prediction to analyze the benefits and limitations. Fewer researches has been carried out in the startup success prediction and achieves the considerable prediction performance. Major limitation has been found among startup success prediction model that use irrelevant features. Some researchers have used social media datasets like Twitter data to increase the performance of the developed method. From existing methods, it has been observed that Random Forest classifiers have been out performed than Logistic Regression method.*

**Keywords:** *Logistic regression, Machine learning techniques, Random forest classifiers, Startup success prediction, and, Twitter data.*

## I. INTRODUCTION

Startup success prediction method is developed for small and medium-scale enterprises to increase their profits. The small and medium scale industry has the impact for both developed and developing countries economy [1]. Venture capital investment is a high risk due to presence of lot of uncertainty in startups business models. A simple version of the startup model is usually not sufficient to determine the success of the business [2, 3]. Government has recognized the small and medium-size enterprises due to their impact on economic growth, stability, employment and creation of a new job and developed social cohesion [4]. Recently, crowdfunding is considered as a financial mechanism that has gains the widespread popularity and startups uses the portals like AngelList [5]. Venture Lab Twente has been developed in 2010 to increases the economic growth of business in Northern European for both in private and public sectors. This is developed to help high tech spin offs that involves in

finding their growth of high potential in the program involves in training content, coaching and access to many resources [6].

Companies are applying empirical investigations of methods to find new solutions to assess the value propositions and risks [7]. Small scale companies are facing many challenges due to the standard procedures like require updating 15 various product components in the changes, coordination and communication issues and confusing project overviews [8, 9]. Small technology startups are facing time pressure from market to manage tough competitions, rapidly evolving, uncertain context and operating in chaotic. Despite many successful businesses, a majority of small technology startups self-destructs within two years from their creation [10, 11]. Many research studies also focused on determining the relative performance of different methods and factors involved in startups success [12, 13]. These models that are developed to analyze the startup success factors are interest in knowledge for the entrepreneurs in the pre-startups. This paper provides the detailed analysis about the performance prediction of startups and analyze its limitations and advantages.

The paper is organized as follows: a detailed analysis on various startup prediction model is analyzed in the section 2, comparison analysis of several researches is given in section 3 and conclusion from analysis of existing method is provided in section 4.

## II. LITERATURE WORKS ON STARTUP PREDICTION PERFORMANCE

The prediction model of startup success is an important tool for the business vendors to make decision according to factors. Startup performance prediction is difficult task due to an analysis of various relationships among data. Researches also incorporates the social media dataset to increase the performance of the prediction. Various models were applied in the different data to increase the prediction performance and analysis of important factors. The general block diagram of the startup performance prediction is shown in the Figure 1.

Tomy, et al. [10] analyzed the uncertainties in the surrounding opportunities with evaluation stage of entrepreneurial process. Internal and external data were used to analyze the uncertainty in the startup business performance prediction. The prediction model is implemented based on machine learning techniques and strategic analysis. The data is processed in such a way that the probability of success or failure is calculated in the pre-start-up phase. Naïve Bayes classifier is applied in this method for the prediction of success.

Revised Manuscript Received on March 30, 2020.

\* Correspondence Author

**Ramakrishna Allu\***, Research Scholar, Department of Computer Science and Engineering, GITAM (Deemed to be University), Visakhapatnam, India and Associate Director in Novartis, Hyderabad, India. Email: Ramakrishna.allu@gmail.com

**Venkata Nageswara Rao Padmanabhuni**, Professor, Dept. of CSE, GITAM (Deemed to be university), Visakhapatnam, India. Email: Venkatanageswararao.padmanabhuni@gitam.edu

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license ([http://creativecommons.org/licenses/by-nc-nd/4.0/](https://creativecommons.org/licenses/by-nc-nd/4.0/))

The developed method shows the significant of decision making and technology of entrepreneurship in the startup business success. This method includes uncertainty in the prediction process and achieves considerable performance. Uncertainty can be effectively analyzed by applying feature selection techniques like Particle Swarm Optimization.

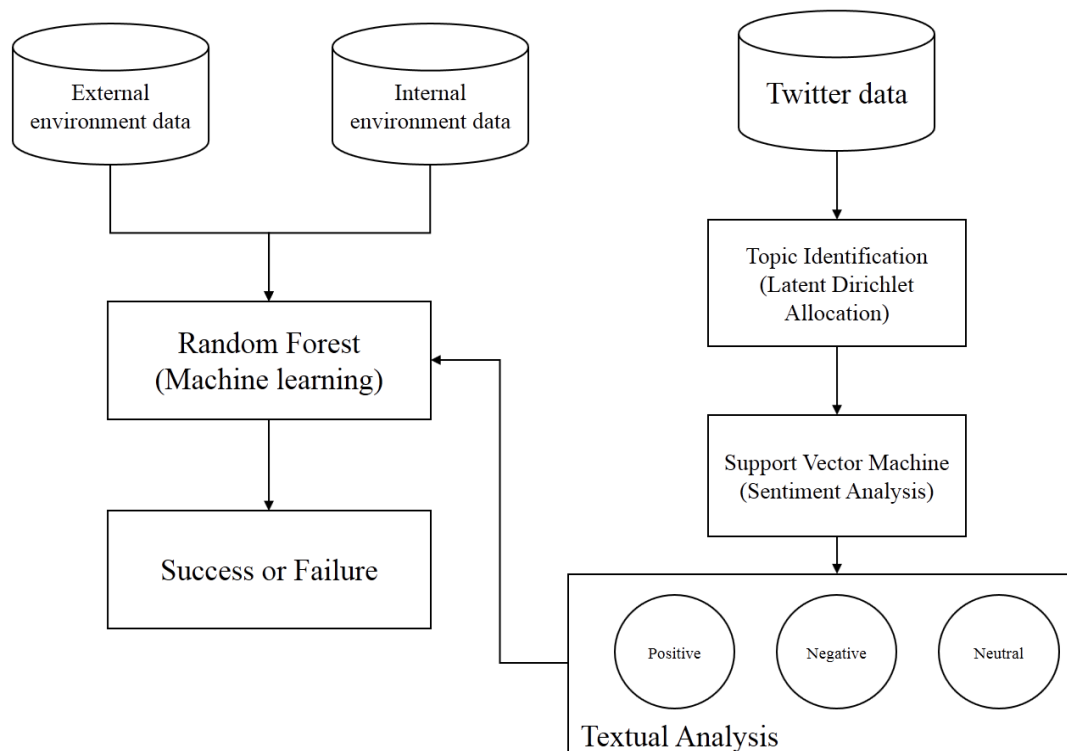
Cacciollattia, et al., [14] developed a framework to analyze the strategic alliance and startups performance prediction. There are 3913 UK high-tech startups engaging in social innovation to analyze the performance of the developed method. The study concludes that the startups scalability is needed between the performance and social mission pursuit. The research found that unlike big companies, startups operates in the particular way and the different form of alliance that are not proper in the business development. Feature selection method can be applied to effectively analysis the factors in startups success.

Balboni, et al. [15] developed a business model to analyze the significant conceptual improvement of study and explanation of development process in technology and science field. The industry structure and marketing dynamics were analyzed in the model for the prediction process. Several factors have been used to analyze the startup performance prediction like entrepreneurial factors, contextual factors and strategic factors. The study states that new venture are need to adopt their business according to the environment and also need to solve the internal consistency

components. Random forest classifier is applied in the factors to effectively analyze the relationship between the factors and a solution to the vendors.

Guzmán, et al. [16] analyzed the success and failure factors of the startups based on the Lussier model on 303 business data records. The data were collected in terms of a personal interview and machine learning techniques were used to analyze the factors. Logistic regression shows that there are three factors that are important in the startup performance. The three factors are professional advice, attract retain employees and partners with success. The experimental analysis shows that the developed method has the accuracy of 66 % in prediction. The Logistic regression has the lower performance in analyzing more number of factors.

Antretter, et al. [17] used social media Twitter data to predict new venture survival. There are more than 187,000 tweets from 253 new venture Twitter accounts are analyzed based on context-specific machine learning method. The developed method achieved 76% of accuracy in prediction. This study analyzes the importance of building legitimacy online and machine learning method in the startup performance prediction. Random forest and gradient boosting were used to predict based on various data including the social media dataset. The Random forest is not effective in handling more data and selecting the relevant features. The deep learning of Long Short Term Memory (LSTM) can be used to increase the performance of the prediction.



**Fig. 1. The general block diagram of startups success prediction**

### III. PERFORMANCE ANALYSIS

This paper analyzes the methods involved in startups performance prediction with various techniques. In recent years, various methodologies have been adopted in the

business including globalization and marketing in Internet. Only few studies have been carried out in the startups performance prediction. The methods involved in the startup prediction performance are compared in Table 1.

**Table- I: Performance Analysis of the startups success prediction**

Author(s)	Proposed Methodology	Advantages	Observations / Limitations	Performance Evaluation
Antretter, et al. [17]	New venture survival has been carried out in this method based on the twitter data.  Large number of data has been collected and prediction model is applied.  Random forest and gradient boosting method was used in this method for the prediction of startups success.	Gradient boosting method increases the performance of the developed method.	Legitimacy data analysis has not used in this method.	This method achieved 76% of accuracy.
Saura, et al. [18]	Latent Dirichlet Allocation (LDA)  Support Vector Machine (SVM)–Sentiment Analysis	Sentiment analysis increases the performance of the developed method.  LDA increases the topic identification and SVM analyzes the sentiment towards the company.	Less number of tweets only analyzed in the method.	Weighted percentage for positive indicator = 2.18.
Xie, et al. [19]	Multilingual sentiment identification is developed based on the Word level sentiment analysis for startup performance prediction	An improved semantic rule algorithm considers the unique semantics in social media texts.  A rule based algorithms are effective in emotional analysis of domain knowledge.	Multilingual learning is required to analyze the startup factors effectively.  Heterogeneous data are need to be analyzed properly.	For English Twitter data, Random Forest achieves 81% of accuracy.  For Korean data, Random forest is achieved 76% of accuracy.
Zhang, et al. [20]	Three algorithms were used for sentiment analysis, namely: Decision tree, Random forest and logistic regression.  Neural Network is also used for the classification.	Heterogeneous datasets were used for the sentiment analysis. Random forest achieves the higher performance compared to other techniques.	The feature selection method is need to be improved to increase the performance of the sentiment analysis.	This method has achieved 89% of accuracy and also achieved 89% of f-measure.
Greenberg, et al. [21]	Adam optimization technique has been applied with various classifier technique.	Adam optimization increase the performance of the developed method.  Random forest achieves the higher performance compared to other method in sentiment analysis.	Feature selection method is required for effective analysis.	Random forest accuracy = 67.53 %.  Logistic regression = 65.09 %.
Dellermann, et al. [22]	Semi-supervised method is used in this research.  Machine learning method has been applied to access, process and structured large amount of information, while individual solve the average systematic errors in the method.	Error value has been much reduced.  Uncertainty in the data has been analyzed effectively.	Time consumption is high in this method.  This method is expensive, when number of data has been increased exponentially.	-
Sharchilev, et al. [23]	Web based startup success prediction has been used in this research.  Business and employment oriented social networking structures has been analyzed in this method.	Success prediction is made in this method based on the collected data.  Heterogeneous data were effectively analyzed by the Gradient Boosting, Logistic Regression and Neural Network.  The developed method achieves the higher performance compared to other existing methods.	Feature selection method is required to improve the performance.	This method achieved 62 % of precision.
Kofanov and Zozul'ov [24].	Bayesian Network with mathematical model has been used for the startups prediction success	Considerable performance has been achieved in this method.  Sentiment analysis has been carried out in this method.	Feature selections like Bag of Words, TF-IDF can be used to increase the performance of the developed method.	Accuracy of the method is low.
Gyimah, et al. [25]	Lussier Model was used in this method for success prediction of small business in Ghana.	Three variables such as capital, economic timing and Marketing skills were used in this method for the analysis.  Logistic regression method was used in this method to increase the performance of the developed method.	Missing data is high in the dataset that affects the performance.  Normalization techniques can be used to solve the missing data problem.	This method achieved 86 % of prediction.

## IV. CONCLUSION

Startup Performance Prediction is a highly useful tool for small and medium scale enterprises to increase their profit. In existing methods, various classification techniques were applied to increase prediction accuracy in startup performance prediction. Social media datasets were incorporated in the existing methods to increase the performance of the prediction. For sentiment analysis in startups prediction performance, topic identification is need to be carried out and Latent Dirichlet Allocation has higher performance in topic identification. Logistic Regression has considerable performance in the analysis of various data related to vendor. From the analysis, Random forest classifiers has achieved higher performance in the social and business datasets. However, Random forest is not efficient in handling number of data points and heterogeneous data. Therefore, efficient classifier with feature selection technique is needed for effective prediction of the performance of the startups.

## REFERENCES

1. L. Krishnan, "Entrepreneurial Competencies and Contextual Determinants of Successful SME Business Startups in Bangalore, India," *International Journal of Engineering and Management Research*, vol. 7, 2017, pp. 1-15.
2. C. Gastaud, T. Carniel, and J. M. Dalle, "The varying importance of extrinsic factors in the success of startup fundraising: competition at early-stage and networks at growth-stage", *arXiv preprint arXiv:1906.03210*, 2019.
3. M. N. Diakomihalis, "Evaluation of high educational institutes economic impact with satellite accounting: an empirical application in Greece", 2014.
4. D. Banerji, and T. Reimer, "Startup founders and their LinkedIn connections: Are well-connected entrepreneurs more successful", *Computers in Human Behavior*, vol. 90, 2019, pp. 46-52.
5. Q. Zhang, T. Ye, M. Essaidi, S. Agarwal, V. Liu, and B. T. Loo, "Predicting startup crowdfunding success through longitudinal social engagement analysis", *In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pp. 1937-1946, ACM.
6. B.G. Englis, P. D. Englis, K. Zalewska-Kurek, and A. Groen, "Business Panel Prediction of Nascent Entrepreneurs Success", *In 24th Annual High Technology Small Firms Conference, HTSF*, 2016.
7. S. Albrecht, B. Lutz, and D. Neumann, "How Sentiment Impacts the Success of Blockchain Startups-An Analysis of Social Media Data and Initial Coin Offerings", *In HICSS*, 2019, pp. 1-10.
8. N. Tripathi, E. Klotins, R. Prikladnicki, M. Oivo, L. B. Pompermaier, L.B., Kudakacheril, A.S., Unterkalmsteiner, M., Liukkunen, K. and Gorschek, T., 2018. An anatomy of requirements engineering in software startups using multi-vocal literature and case survey. *Journal of Systems and Software*, vol. 146, pp. 130-151.
9. D. McKenzie, and D. Sansone, "Predicting entrepreneurial success is hard: Evidence from a business plan competition in Nigeria", *Journal of Development Economics*, vol. 141, 2019, pp. 102369.
10. S. Tomy, and E. Pardede, "From uncertainties to successful startups: A data analytic approach to predict success in technological entrepreneurship", *Sustainability*, vol. 10, 2018, pp. 602.
11. A. A. Abou-Moghli, and M. M. Al-Kasasbeh, "Social network and the success of business start-up", *International Journal of Business and Management*, vol. 7, 2012, pp. 134.
12. M. Islam, A. Fremeth, and A. Marcus, "Signaling by early stage startups: US government research grants and venture capital funding", *Journal of Business Venturing*, vol. 33, 2018, pp. 35-51.
13. H. Janáková, "The success prediction of the technological start-up projects in Slovak conditions", *Procedia Economics and Finance*, vol. 34, 2015, pp. 73-80.
14. L. Cacciolatti, A. Rosli, J. L. Ruiz-Alba, and J. Chang, "Strategic alliances and firm performance in startups with a social mission", *Journal of Business Research*, vol. 106, 2020, pp. 106-117.
15. G. Bortoluzzi, M. Tivan, A. Tracogna, and F. Venier, "The Growth Drivers of Start-up Firms and Business Modelling: A First Step toward a Desirable Convergence", *Management*, 9, 2014.
16. J. Guzmán, and R. Lussier, "Success factors for small businesses in Guanajuato, Mexico", *International Journal of Business and Social Science*, vol. 6, 2015, pp. 1-7.
17. T. Antretter, I. Blohm, D. Grichnik, and J. Wincent, "Predicting new venture survival: A Twitter-based machine learning approach to measuring online legitimacy", *Journal of Business Venturing Insights*, vol. 11, 2019, pp. e00109.
18. J. R. Saura, P. Palos-Sanchez, and A. Grilo, "Detecting indicators for startup business success: Sentiment analysis using text data mining", *Sustainability*, vol. 11, 2019, pp. 917.
19. Y. Xie, Z. Chen, K. Zhang, Y. Cheng, D. K. Honbo, A. Agrawal, and A. N. Choudhary, "MuSES: Multilingual sentiment elicitation system for social media data", *IEEE Intelligent Systems*, vol. 29, 2013, pp. 34-42.
20. K. Zhang, Y. Cheng, Y. Xie, D. Honbo, A. Agrawal, D. Palsetia, K. Lee, W. K. Liao, and A. Choudhary, "SES: Sentiment elicitation system for social media data", *In 2011 IEEE 11th International Conference on Data Mining Workshops*, pp. 129-136.
21. M. D. Greenberg, B. Pardo, K. Hariharan, and E. Gerber, "Crowdfunding support tools: predicting success & failure", *In CHI'13 Extended Abstracts on Human Factors in Computing Systems*, 2013, pp. 1815-1820.
22. D. Dellermann, N. Lipusch, P. Ebel, K. M. Popp, and J. M. Leimeister, "Finding the unicorn: Predicting early stage startup success through a hybrid intelligence method", 2017.
23. B. Sharchilev, M. Roizner, A. Rumyantsev, D. Ozornin, P. Serdyukov, and M. de Rijke, "Web-based startup success prediction", *In Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 2018, pp. 2283-2291.
24. O. Kofanov, and O. Zozulov, "Successful development of startups as a global trend of innovative socio-economic transformations", *International and Multidisciplinary Journal of Social Sciences*, vol. 7, 2018, pp. 191-217.
25. P. Gyimah, K. O. Appiah, and R. N. Lussier, "Success versus Failure Prediction Model for Small Businesses in Ghana", *Journal of African Business*, 2019, pp. 1-20.

## AUTHORS PROFILE



**Ramakrishna Allu** is an Associate Director in Novartis working as a Technology Leader in Web, Social and Mobile CoE. He completed M.S (Software Systems) from BITS Pilani, Currently he is Ph.D Scholar in Computer Science & Engineering at GITAM University, Visakhapatnam. His expertise in Integration and digital programs delivery, extensive experience in managing Global Accounts & Independent business units with P&L responsibility, building high performing teams and Innovative delivery models.



**Venkata Nageswara Rao Padmanabhuni**, received M.Tech (CST) from Andhra University and Ph.D. from AcharyaNagarjuna University and had 25 years of teaching and research experience. Presently working as Professor, Department of CSE, GITAM University, Visakhapatnam. His research interests include Soft Computing, Bioinformatics and Computer Networks