

An efficient and sustainable novel approach for prediction of start-up company success rates through sustainable machine learning paradigms

Bharathi Panduri^{1*}, *P.K. Abhilash*¹, *Chidananda K*¹, *Venkata Naga Tejaswi Bethapud*², *Anjali Naudiyal*³, *Mahitha Kodamunja*²

¹Assistant Professor, Department of Information Technology, GRIET, Bachupally, Hyderabad, Telangana.

²Student, Department of Information Technology, GRIET, Bachupally, Hyderabad, Telangana.

³School of Applied and Life Sciences, Uttaranchal university, Dehradun-248007, India.

Abstract: The primary objective is to construct a sustainable machine-learning model that utilizes multiple variables to forecast the success of a startup enterprise. It incorporates a Flask application for creating a user-friendly interface, where users can input specific parameters related to a startup, such as financial metrics, industry sector, and location. These inputs are then passed through a sustainable machine learning prediction model, which has been trained on a comprehensive dataset of startup information. The model employs sustainable advanced algorithms to evaluate their startup ventures' potential success. Through the development and deployment of the Flask application and the integration of sustainable machine learning prediction model, this model contributes to the field of startup analysis and decision-making. It offers a sustainable and efficient solution for predicting startup success, empowering users to make data-backed decisions and optimize their resource allocation.

1. Introduction¹:

This model aims to develop a sustainable machine learning model that predicts the likelihood of success for startup companies. Startups play a crucial role in driving innovation and economic growth, but their success rate can be uncertain. By leveraging sustainable machine learning algorithms and analyzing various factors such as milestones, funding rounds, relationships, and funding amounts, this model aims to provide valuable insights to entrepreneurs, investors, and stakeholders. The prediction model utilizes a sustainable dataset containing information about startup companies, including their characteristics and outcomes. By training the model on historical data, it learns patterns and relationships that can be used to predict the success or failure of new startups. The model focuses on implementing and evaluating different sustainable machine learning algorithms, such as Random Forest, Logistic Regression, Support Vector Machines (SVM), and Naive Bayes, to

* Corresponding Author: bharathi1284@grietcollege.com

determine the most accurate and effective approach for startup success prediction. The advancement of technology and entrepreneurship has increased the importance of predicting the success of start-up firms. This study aims to explore potential applications for machine learning algorithms in predicting startup company success. Our objective is to develop a prediction model that can assess a start-up's potential for success by considering a variety of elements, including financial information, market trends, and team dynamics. Start-ups are essential for promoting innovation and economic expansion. However, given the high failure rate of startups, it's critical to pinpoint the crucial elements that either contribute to their success or failure. Large-scale data analysis and precise prediction are both possible with the help of machine learning algorithms [21]. In this model, the employment of machine learning techniques is the main goal. Through this model, we aim to empower decision-makers in the startup ecosystem with a reliable tool for assessing the viability and potential of new ventures. By leveraging the power of sustainable machine learning, we can enhance the decision-making process and provide valuable insights that can shape the future of entrepreneurship and innovation.

2. Literature Survey:

Malhar Bangdiwala, Yashvi Mehta et. al. “Predicting Success Rate of Startups using Machine Learning Algorithms” in IEEE Access 2022. Startups must determine whether they are headed for success. Since companies failed at a rate of almost 90% in 2019, it is important to understand how successful a startup is. In-depth forecasting of startup success is the focus of this article. Startups can achieve success in one of two ways: by issuing an IPO (Initial Public Offering) or by merging with or being purchased by another business. [1]

Xuejiao Ren, Xiaozhou, et. al. “Motivation Analysis of Technological Startups Business Models Based on Intelligent Data Mining and Analysis” in Hindawi journal, 2022. To boost the motivation analysis effect of the business model of technological startups, this study combines intelligent data mining technology to analyse pertinent elements and proposes a method of changing the parameters of the machine-learning model based on the Bayesian optimization algorithm. [2]

Nicola Amoroso, Alfonso Monaco, et. al. “Economic Interplay Forecasting Business Success,” in Hindawi Journal, 2021. A startup ecosystem is a dynamic setting where several actors—including investors, venture capitalists, angel investors, and facilitators—play major roles in intricate interactions. The majority of these exchanges involve the movement of money, whose magnitude and direction aid in illustrating the complex web of connections. [3]

Chenchen Pan, Yuan Gao, Yuzi Luo, et. al. “Machine Learning Prediction of Companies Business Success “Each year, thousands of new businesses launch all around the world. Both in the US and China, the number of new businesses has grown quickly during the past few decades. In this study, they used Crunch base data to develop a supervised learning predictive model to categorize which start-ups are successful and which are not. [4]

Prof. Dr. Wolfgang Karl Hardle, Prof. Dr. Weining Wang, et. al. "A Machine Learning Approach towards Startup Success Prediction" This essay discusses how to forecast startup company success. This body of literature on startup success made clear the necessity for more study. The focus of the current literature is on predicting established business success rates. The models in the existing research, however, are challenging to apply to forecast success for startup enterprises since there are differences between corporate and startup success. [5] Other researchers also worked on above related work using various technologies. [6-27]

3. Methodology:

Recognize the issue and collect specifications for the machine learning solution. Gather and pre-process the necessary data, managing outliers and missing values. Select a suitable machine learning algorithm depending on the issue and the data analysis. Establish hyperparameters and design the model architecture. Utilizing the pre-processed data, train the model and assess its effectiveness. Deploy the learned model while keeping scalability and system integration in mind. Keep an eye on the model's performance in real-world scenarios and retrain or update it as needed. Record the whole development process, including the model design, pre-processing steps, and evaluation outcomes [22]. For efficient teamwork and code management, use version control and collaboration technologies. Make sure that the model's capabilities, constraints, and assumptions are communicated clearly.



Fig.1. Data Preparation

3.1 Data Collection and Selection:

This data is collected from crunch base database and consists of columns like first funding, last funding and number of funding rounds etc which helps in calculating the economy and leading to success rate prediction of a company.

Fig.2. Crunch Data Base

3.2 Data Pre-Processing:

A supervised machine learning task's total performance can frequently be significantly impacted by the pre-processing of the input. The pre-processing of data involves these three steps Data Cleaning, Data Selection, and Data Transformation etc.[23], The process stated, Data cleaning, in which the author makes an effort to remove all superfluous and pointless data from the database as well as duplicates, missing data, and outliers. Data selection, limits which data are included in the final dataset by establishing the study's context (i.e., social-demographic criteria). Data transformation involves integrating data from different tables into a single table for the organization or adding new variables.

	relationships	funding_rounds	funding_total_usd	milestones	avg_participants
860	7	3	22487000	2	2.3333
111	1	2	30000000	0	2.5000
524	3	2	675000	1	1.5000
630	1	1	50000	1	1.0000
605	9	4	29288979	1	2.6667

Fig.3. Parameters considered for prediction

3.4 Machine Learning Algorithms:

3.4.1 Random Forest:

The Random Forest method employs an ensemble learning approach by amalgamating multiple decision trees to produce predictions. Its foundations are random feature sub-setting and bagging (bootstrap aggregating). Random Forest uses labeled data, in which each instance has a predetermined class label. The target variable (the projected class label) and the characteristics (the input variables) comprise the dataset's two components.

3.4.2 Logistic Regression:

Logistic regression is a frequently employed classification method to estimate the probability of an instance belonging to a particular class. The logistic function is utilized to establish the relationship between the input variables (features) and the binary outcome. Linear Regression Equation: $z = \theta + 1x1 + 2x2 + \dots + h_x h$, where z is the linear combination of the input features, θ is the intercept term, and $1, 2, \dots, h$ are the coefficients of the input features. Logistic Regression Sigmoid Function: $p = 1 / (1 + e(-z))$. If z is the linear combination of characteristics, e is the base of the natural logarithm (about 2.71828), and p is the anticipated probability.

3.4.3 Support Vector Machine:

Support Vector Machines (SVMs) are robust supervised learning models used for tackling both classification and regression problems.[5] Binary classification and regression use somewhat different SVM working equations. SVM must create a hyperplane that divides the data into two classes with the greatest possible margin of error given a training dataset with input features X and binary class labels y (-1 or 1). The margin represents the distance or separation between the closest data points from each class and the hyperplane. The decision function has the following definition: $f(x) = \text{sign}(w^T x + b)$, where $f(x)$ provides the anticipated class label (-1 or 1) for an input feature vector x .

3.4.4 Naïve Bayes:

A key principle of probability theory and statistics known as Bayes' Theorem explains how to update or adjust probabilities in light of fresh data. According to this rule, the likelihood that event A will occur given that event B has already happened is equal to the likelihood that event B will also occur given that event A has already happened, multiplied by the likelihood that event A will occur, and divided by the likelihood that event B will occur. It can be mathematically stated as $P(A|B) = (P(B|A) * P(A)) / P(B)$.

3.4.5 K Nearest-Neighbour (KNN):

A straightforward yet efficient non-parametric classification and regression approach is the k-nearest Neighbours (k-NN) algorithm. It assigns new instances a classification based on the consensus of their k nearest neighbours.

4. Results:

On executing the code, we will be directed to the initial webpage.

The web page is as follows:

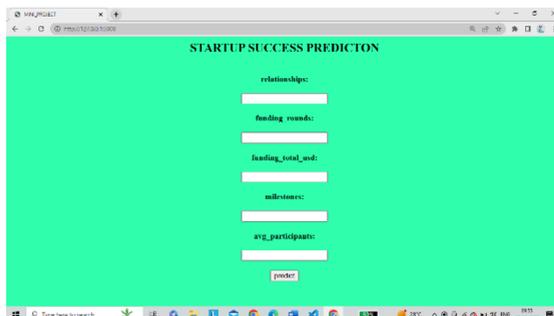


Fig.4. Initial Web Page

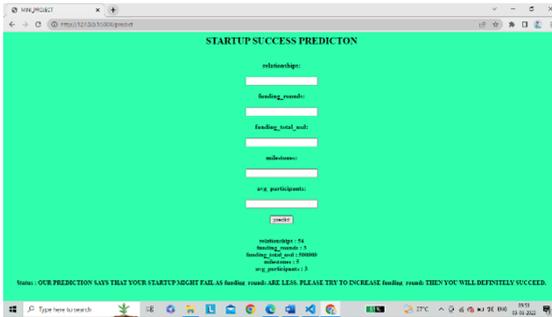


Fig.5. The page showing Startup might fail

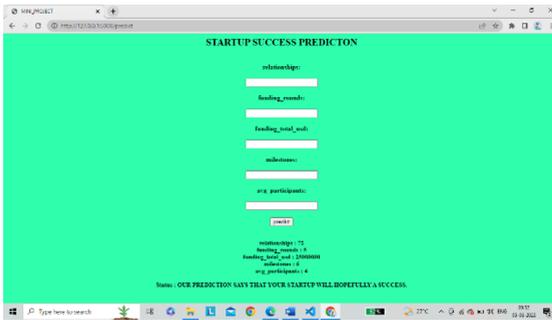


Fig.6. The page showing Startup will become successful

5. Conclusion:

In conclusion, the model successfully addresses the challenge of determining the success potential of startups using machine learning algorithms. This model provides valuable insights into startups' viability and growth prospects by leveraging a comprehensive dataset and advanced prediction models. Throughout the model, we developed a Flask application that serves as a user-friendly interface, allowing users to input specific parameters related to their startup and obtain a prediction regarding success. The integration of the machine learning prediction model enables accurate and data-driven assessments, empowering users to make informed decisions about their startup ventures. The results obtained from the model highlight the significance of utilizing machine learning techniques in startup analysis. By considering various factors such as financial metrics, industry sector, and location, the prediction model exhibits a high level of accuracy in determining the success potential of startups. This can greatly benefit investors, entrepreneurs, and other stakeholders in making investment decisions and formulating effective business strategies.

6. References:

[1] Ries, E. (2022). "Lean Startup." 1-28. Working Paper. <http://doi.org/23>

[2] Holtschke, D., Hermann, B. L., Gauthier, J., Bermann, R. D., & Marmer (2022). The 2015 ranking of the global startup ecosphere. (August), 1-156, The Startup Ecosystem Report Series.

[3] B. Geier (2021). What Did the 2000 Dotcom Stock Bubble Teach Us? from <http://time.com/3741681/2000-dotcom-stock-bust>.

[4] L. Auria and R. A. Moro's article, Support Vector Machines (SVM) as a Technique for Solvency Analysis, was downloaded from www.diw.de in August.

[5] Discriminant analysis, financial ratios, and the likelihood of business bankruptcy, 589–609 in *Journal of Finance*, 23(4), Altman, E. I. 2020. <http://doi.org/10.2307/2978933>

[6] Guyon, I. M., Boser, B. E., and Vapnik, V. N. (2019).144–152). A training method for classifiers with the best margin. The Fifth Annual Workshop on Computational Learning Theory Proceedings, COLT '92. <http://doi.org/10.1145/130385.130401>

[7] Kim, E. (2018). Fastest startups to \$1 billion valuation - Business Insider. Retrieved August 21, 2017, from <http://www.businessinsider.com/fastest-startups-to-1-billionvaluation-2015-8/#1-slack-is-the-fastestgrowing-enterprise-software-ever-11111114>

[8] Zhang, D., & Zhou, L. (2017). Discovering golden nuggets: Data mining in financial application. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*. <http://doi.org/10.1109/TSMCC.2004.82927>

[9] S. Korreck. The Indian Startup Ecosystem: Drivers, Challenges and Pillars of Support. ORF Occasional Paper #210

[10] S. Torkington, "How the Great Resignation is boosting startups from more diverse founders," [weforum.org.https://www.weforum.org/agenda/2022/02/the-great-resignation-boom-in-startups-from-more-diverse-founders/](https://www.weforum.org/agenda/2022/02/the-great-resignation-boom-in-startups-from-more-diverse-founders/) (Viewed on May 7, 2022)

[11] Lead-innovation.com, "S TOP reasons why startups fail," by T. Eschberger.

<https://www.lead-innovation.com/english-blog/reasons-startups-fail> (Completed: 7 May 2022)

[12] N. Bosma, R. Thurik, and M. Van Gelderen. Risk and success aspects during the pre-startup stage. 2005; 24(4):365–380; *Small Business*

[13] W.-L. Tan and T. M. Begley. East Asian and Anglo-Saxon nations' sociocultural environments for entrepreneurship are contrasted. *International Business Studies Journal*, 2001, p. 537–553.

[14] R. Dickinson rate of business failure. *American Journal of Small Business*, Vol. 6, No. 2, 1981, pp. 17–25. W. B. Gartner

[15]. Who is a businessperson? is the incorrect query. 12(4):11–32 of the *American Journal of Small Business*, 1988

[16] Malhotra, N. K. *An Applied Approach to Marketing Research*, 5/E.

The year was 2008, according to Pearson Education India.

[17] Successful entrepreneur traits, D. C. McClelland, *The journal of creative behavior*, 21(3):219-233, 1987.

- [18] "Creating Classification Models from Textual Descriptions of Companies Using Crunchbase," IPMU 2020, CCIS 1237, pp. 695-707; Marco Felgueiras, Fernando Batista, and Joao Paulo Carvalho.
- [19] Chechen Pan, Yuan Gao, and Yuzi Luo "Machine Learning Prediction of Companies' Business Success", CS229: Machine Learning, Fall 2018, Stanford University, CA, 2018.
- [20] Assessment of Machine Learning Performance for Decision Support in Venture Capital Investments*, by Javier Arroyo, Francesco Corea, Guillermo Jiménez-Dáz, and Juan A. Recio-Garca, IEEE Access 7, 2019.
- [21] Ledalla, Sukanya, R. Bhavani, and Avvari Pavitra. "Facial Emotional Recognition Using Legion Kernel Convolutional Neural Networks." Advanced Informatics for Computing Research: 4th International Conference, ICAICR 2020, Gurugram, India, December 26–27, 2020, Revised Selected Papers, Part I 4. Springer Singapore, 2021.
- [22] Indian sign language recognition using convolution neural network. Sukanya L, Tharun E, Anup Raj G, Shreyas Singh T and Srinivas S E3S Web Conf., 391 (2023) 01058. DOI: <https://doi.org/10.1051/e3sconf/202339101058> 5. Racism detection using deep learning techniques. Sukanya L, Aniketh J, Abhiman Sathwik E, Sridhar Reddy M, Hemanth Kumar N E3S Web Conf. 391 01052 (2023). DOI: 10.1051/e3sconf/202339101052
- [23] Ledalla, S. & Mahalakshmi, T.S.. (2019). Sentiment analysis using legion kernel convolutional neural network with LSTM. International Journal of Innovative Technology and Exploring Engineering. 8. 226-229.
- [24] Prasanna Lakshmi, K., Reddy, C.R.K. A survey on different trends in Data Streams (2010) ICNIT 2010 - 2010 International Conference on Networking and Information Technology, art. no. 5508473, pp. 451-455.
- [25] Jeevan Nagendra Kumar, Y., Spandana, V., Vaishnavi, V.S., Neha, K., Devi, V.G.R.R. Supervised machine learning Approach for crop yield prediction in agriculture sector (2020) Proceedings of the 5th International Conference on Communication and Electronics Systems, ICCES 2020, art. no. 09137868, pp. 736-741.
- [26] Sankara Babu, B., Suneetha, A., Charles Babu, G., Jeevan Nagendra Kumar, Y., Karuna, G. Medical disease prediction using grey wolf optimization and auto encoder based recurrent neural network (2018) Periodicals of Engineering and Natural Sciences, 6 (1), pp. 229-240.
- [27] Nagaraja, A., Boregowda, U., Khatatneh, K., Vangipuram, R., Nuvvusetty, R., Sravan Kiran, V. Similarity Based Feature Transformation for Network Anomaly Detection (2020) IEEE Access, 8, art. no. 9006824, pp. 39184-39196.