WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Workflow Paper on Startup Success Prediction using Machine Learning Algorithm

Abhilash Das Mahant[1], Ankit Chandrakar[2], B. Vivekanand Patnaik[3], Priyanshu Awachat[4], Prof. Aparna Pandey[5]

Student 1,2,3,4 Dept. Computer Science Engineering Bhilai Institute of Technology
[5]Assistant Professor 5 Dept. Computer Science Engineering Bhilai Institute of Technology

ABSTRACT –

A start-up is a company that has an original concept. A lot of young people tend to come up with the original concepts for this kind of business. Every year, many start-ups start out with ideas, but only a very tiny percentage of these ideas end up long lasting. There are several variables and quantities that affect its survival. In order to determine if the start-up would be successful in the long run or not, we have developed a concept. We have employed machine learning classification algorithms to identify the best strategy for start-up success. To apply our methods, we used the Kaggle dataset. To determine which method works best for this dataset we have used several algorithm's such as Random Forest, Adaboost, Xgboost, Gradient boosting and Light Gradient Boosting Machine. We determined that "Light Gradient Boosting Machine (LGBM)" algorithm were able to get an accuracy of around 94% in all of them and we found that it will assist newcomers in starting their businesses with a high likelihood of success in the future.

Index Terms or Keywords - Random Forest, Adaboost, Xgboost, Gradient boosting and Light Gradient Boosting Machine. (LGBM).

I. INTRODUCTION

Startups have recently gained a lot of interest on a global scale. The number of new startups throughout the world has significantly surpassed those from the previous year as a consequence of the worldwide pandemic, which has caused a real startup boom. This surge in entrepreneurship is attributed to workers who were laid off and started their own businesses. Startups are widely seen as important drivers of economic expansion and job creation. Startups are now harmful to societal improvement. Through innovation and scalable technology, they may produce significant solutions and hence serve as engines for socio-economic development and transformation.[1] People now have a new source of income thanks to startups. One in five microbusinesses established during the COVID time period were started. by people who were classified as unemployed. 12% was the pre-COVID figure.[2] This demonstrates how a lot of individuals are able to break out from the cycle of poverty thanks to microbusinesses and startups. Since they increase the economy's competitiveness and dynamism, startups are the most dynamic economic entities now operating. As a result, the economy continues to be robust, active, and industrious, and it becomes more difficult for individual businesses to rest on their laurels.[3] However, not every business that was established in recent years was successful. This presents a fairly bleak picture for future business owners. Such failures can be attributed to a wide range of factors. Lack of market demand is one of the causes. Many ventures Centre on a product whose demand has either faded over time or has never been. The absence of proper financial means might also be a factor. Many startups simply run out of the resources they need to continue operating. Unsuitable managerial behaviour might be another factor decisions, acts of God such as Covid, or even strong competition.[4] These machine learning models include Logistic Regression, Decision Trees, Random Forest and Gradient Boost. The information obtained for the same mostly focuses on monetary data, such as valuations and the funds raised in each round of venture capital. These models may be used not only by different startup stakeholders to evaluate their development but also by potential venture capitalists wanting to invest in the business. Since they are the greatest employers of workers, small and medium-sized businesses (SMEs) are regarded as one of the foundational elements of the national economy. By earning money through taxes and levies levied on the goods they created, they help to grow the country's economy. SMEs play a crucial role in economy of nations that contribute to the global economy. These industries are essential for the growth of the nation's physical and people resources in accordance with its objectives and requirements. Because these businesses serve as stepping stones to more established, profitable businesses.

II. LITERATURE REVIEW

Malhar Bangdiwala, Yashvi Mehta, Smrithi Agrawal, Sunil Ghane [5] In This paper they have attempts to determine the success of a startup in terms of getting merged or acquired. With the help of historical data available on startups, five models have been built and compared to predict if a startup would get acquired or not. The models that have been used are Decision Trees, Random Forest, Gradient Boost, Logistic Regression, and MLP Neural networks. The data used to train these models includes key features such as valuations, funding rounds, investments, etc.

Harjo Baskoro, Harjanto Prabowo, Meyliana Meyliana, Ford Lumban Gaol [6] Startup creation and growth have become a worldwide phenomenon in recent years. In many nations, startups have become a crucial component of innovation and economic growth. However, research indicates that the failure probability of a company is roughly 90%. As a result, it is critical for investors, financial advisers, and the government to identify the 10% that will eventually provide higher return rates, create more money, and assure economic development. The goal of this research is to determine what are the essential aspects of startup success that can be utilized to create a predictive model utilizing a machine learning algorithm to forecast startup success.

Fuad Saeed Yousif Saad, Mohamed Abu Elgassim Hassanen, Fath Elrahman Shaa Eldeenm [7] This study aims to recognize a set of variables that have the paramount impact on the performance of small industrial business. It also constructs a statistical model that is used to estimate the probability of faltering for any small industrial enterprise, and to determine its expected survival time. It applies cluster analysis to classify depending on variables, i.e., faltering and non-faltering using Cox & risqué; s regression model.

Ibukun Afolabi, T. Cordelia Ifunaya, Funmilayo G. Ojo and Chinonye Moses [8] The aim of the paper is to presents a design and implementation of a system for the diagnosis, prediction, and provision of recommendation for the success of a Business. The methodology used to develop the prediction model is based on correlation analysis for the data pre-processing and the combination of Naïve Bayes and J48 classification algorithm. Necessary heuristics for the diagnosis were collated from the review of existing business consulting systems, expert systems and human experts in the field of business consulting in Nigeria.

Sunitha Cheriyan, Shaniba Ibrahim, Saju Mohanan, Susan Treesa [9] In this paper they have detailed study and analysis of comprehensible predictive models to improve future sales predictions are carried out in this research. Traditional forecast systems are difficult to deal with the big data and accuracy of sales forecasting. These issues could be overcome by using various data mining techniques. On the basis of a performance evaluation, a best suited predictive model is suggested for the sales trend forecast.

III. PROPOSED METHODOLOGY

Prediction of success of a startup is difficult owing to the numerous variables involved. Machine learning algorithms have been used to predict company success by analysing a variety of characteristics. In this technique, we will look at how to predict startup success using decision trees, random forests, light gradient boosting machines, gradient boosting classifiers, xgboost classifiers, and adaboost classifiers. In addition, precision, recall, F1 score, confusion matrix, AUC, and ROC will be used to determine accuracy.

1.Data Collection and Pre-processing

The first phase in our technique is to gather and pre-process data. We will gather startup data from a variety of sources, including Crunchbase, AngelList, and other public datasets. The data will contain information about the firm, such as the date it was founded, the amount of capital raised, the industry, the number of workers, and the founder's history. We will also provide facts on the market, such as market size and competition. We will pre-process the data after collecting it to eliminate any missing or unnecessary information. In addition, we will apply feature engineering to extract important characteristics from the data. This phase includes converting the data into a more usable format for analysis. To prepare the data for machine learning algorithms, we will employ techniques such as one-hot encoding, scaling, and normalization.

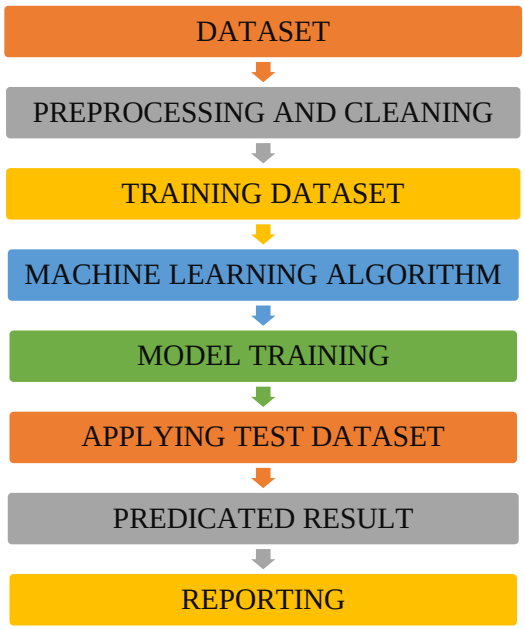

Figure 1 Block or Process Diagram

2. Model Selection and Training the model

The second phase in our technique is to choose the best machine learning algorithm to predict startup success. We will employ methods such as decision tree, random forest, light gradient boosting machine, gradient boosting classifier, xgboost classifier, and adaboost classifier. These methods are often used for classification issues and have been shown to be useful in forecasting startup success. We will divide the data into training and testing datasets in order to train the models. The models will be trained using the training dataset, and their performance will be evaluated using the testing dataset. Cross-validation will be used to verify the models' performance and to modify the hyperparameters to increase accuracy.

3. Evaluation Metrics

We will use a variety of measures to assess the performance of our models, including accuracy, recall, F1 score, confusion matrix, AUC, and ROC. These measures, which are often employed for classification tasks, offer an indication of the model's accuracy and performance. Precision measures the proportion of true positives (TP) in the predicted positive (P) instances. It is defined as TP / (TP + FP). Recall measures the proportion of true positives in the actual positive (A) instances. It is defined as TP / (TP + FN). The F1 score is the harmonic mean of precision and recall and is defined as 2 * (precision * recall) / (precision + recall). The confusion matrix is a table that summarises the performance of a classification method. The number of true positives, false positives, true negatives, and false negatives is displayed. AUC (Area Under the Curve) is a model's ability to differentiate between positive and negative examples. It is frequently used in ROC analysis, which compares the true positive rate (TPR) against the false positive rate (FPR).

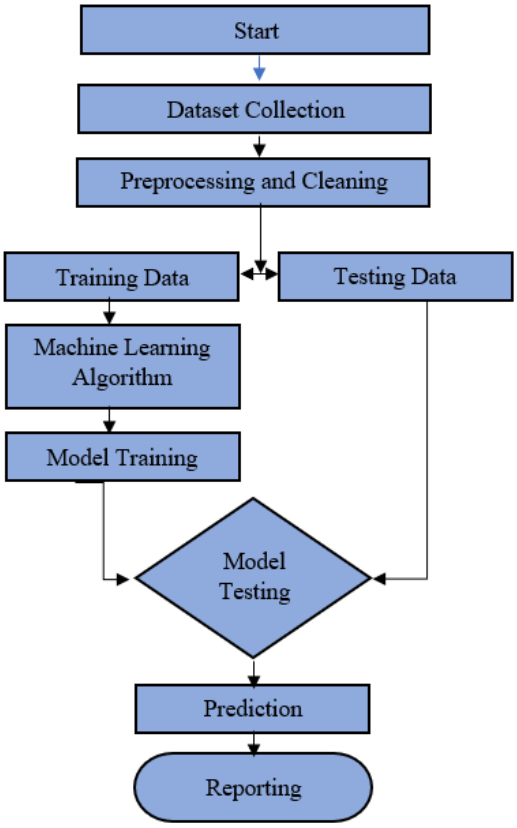

Figure 2 Proposed Flowchart

4. Prediction

The model is trained and optimized using Random Forest, Adaboost, Xgboost, Gradient boosting, Light Gradient Boosting Machine and found that light Gradient boosting machine which help to predict the success of new startups effectively. Monitor the model's performance over time and retrain the model as necessary to improve its accuracy.

5. Reporting

Present the results of the analysis, including the selected model and its performance metrics, to stakeholders in a clear and understandable way. Example: - Heatmap, Bar graph, Pie-chart, etc.

IV. EXPECTED OUTCOME

The proposed technique will increase accuracy for the startup prediction. This technique will extract the large number of features and which will also increase precision, recall values. framework presented in this study can be considered as a decision support system for managers, businessmen and policymakers.

V. CONCLUSION

Our project demonstrated that Light GBM is a powerful gradient boosting framework as well as can be an effective tool for predicting success of a startup. The model achieved high accuracy, predictive performance, and identified important features that influence startup success. The most important achievement of this research in this respect is to code and develop an artificial intelligence system and decision-making help desk to select the appropriate growth strategies for companies in any industry to enable more competition. The introduced model in this project can be offered as a software tool for any industry managers to enable them to become informed about conditions for success of new product development and receive an appropriate strategy for growth only by entering data into the software program. This project can be valuable for investors, entrepreneurs, and stakeholders to make informed decisions in the dynamic and competitive startup landscape.

VI. REFERENCE

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

[2] S. Torkington" How the Great Resignation is driving a boom in startups from more diverse founders" weforum.org. https://www.weforum.org/agenda/2022/02/the-great-resignation-boomin-startups-from-more-diverse-founders/ (accessed: May 7,2022)

[3] T. Eschberger "5 TOP reasons why startups fail" lead-innovation.com. https://www.lead-innovation.com/english-blog/reasons-startups-fail (accessed: May 7,2022)

[4] M. Van Gelderen, R. Thurik, and N. Bosma. Success and risk factors in the pre-startup phase. Small Business Economics, 24(4):365–380, 2005.

[5] M. Bangdiwala, Y. Mehta, S. Agrawal and S. Ghane, "Predicting Success Rate of Startups using Machine Learning Algorithms," 2022 2nd Asian Conference on Innovation in Technology (ASIANCON), 2022, pp. 1-6, doi: 10.1109/ASIANCON55314.2022.9908921.

[6] Baskoro, H., Prabowo, H., Meyliana, M., & Gaol, F. (n.d.). Predicting startup success, a literature review. Retrieved November 15, 2022, fromhttps://prosidingicostec.respati.ac.id/index.php/icostec/article/view/6

[7] Saad, F. S. Y., M. A. E. H. Eldeenm, F. E. S. &, H. M. A. (2022). The Factors of Success and Failure in Small Industrial Business: A Case of Asir Region in Saudi Arabia. Global Journal of Economics and Business, 12 (1), 128-137, 10.31559/GJEB2022.12.1.8

[8] Afolabi, I., Ifunaya, T. C., Ojo, F. G., & Moses, C. (2019). A model for business success prediction using machine learning algorithms. Journal of Physics: Conference Series, 1299, 012050. doi:10.1088/1742-6596/1299/1/012050

[9] S. Cheriyan, S. Ibrahim, S. Mohanan and S. Treesa, "Intelligent Sales Prediction Using Machine Learning Techniques," 2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE), 2018, pp. 53-58, doi: 10.1109/iCCECOME.2018.8659115.