Forecasting Model to Predict Startup Success and Suggested Locations to Optimize Business Expansion

Dr. Akash D. Waghmare
Associate Professor
Department of Computer Engg.,
SSBT's COET, Jalgaon

Dr. Dinesh D. Puri
Associate Professor
Department of Computer Engg,
SSBT's COET, Jalgaon

Dr. G. K. Patnaik
Professor
Department of Computer Engg.,
SSBT's COET, Jalgaon

Ms. Dhanashree S. Tayade
Assistant Professor
Department of Computer Engg.,
SSBT's COET, Jalgaon

Abstract

,

With the use of machine learning, the programme offers an integrated method for creating algorithms that predict startup success and suggest local locations for medium-sized, small, and large startups. The startup success prediction and location recommendation are two of the system's primary features based on market trends, industry, demand, and past data. An informed decision-making process that takes into account market trends, demand, and pandemics is the goal of the holistic approach, which empowers entrepreneurs. Through the use of machine learning, the proposed integrated predicting start-up success and location recommendation. Application gives entrepreneurs precise, data-driven insights that optimise market trend, demand, and industry choices. The project was initiated out of a sincere desire to enable investors and aspiring entrepreneurs to participate in a thriving start-up ecosystem. It will benefit those who aspire to launch their own companies. Establishing a business is an exciting journey, but it can also be rather challenging. People frequently are unsure of the decision they are making. Application is all about helping individuals succeed, facilitating the launch of amazing new ventures, and fostering the development of innovative ideas.

Keywords –Machine Learning, Market trend, Startup

I. Introduction

The startup scene is still thriving with new ideas and initiatives in an era characterized by innovation and entrepreneurship. The fortunes of startups are crucial in determining the direction of sectors, propelling economic expansion, and advancing technology in the context of the changing global economy. With the use of machine learning, this programme attempted to empower investors and prospective business owners by predicting startup success and offering location recommendations. The dynamic nature of the startup environment is attributed to the many impacting elements. a new venture's success or failure. The best company success prediction and location recommendation are frequently hampered by major obstacles that conventional entrepreneur methods frequently encounter. A primary drawback is the absence of accuracy and customized strategies. Because conventional techniques frequently rely on generalized habits, they allocate resources inefficiently. Startup growth and earnings may be negatively impacted by an entrepreneur's inability to quickly adjust to abrupt changes. Furthermore, one of the primary factors contributing to a startup's high success rate is location. The constraints highlight how urgently a successful launch is needed. Anticipating startup success aims to address these issues by utilizing artificial intelligence (AI) to generate customized suggestions grounded in precise data examination. Launching a new company is like setting out on an exciting journey full of opportunities and difficulties. It's thrilling to think that a creative idea could become a profitable business, but there are typically many unknowns along the way.

Entrepreneurs must make critical decisions in the dynamic world of startups, including determining the ideal location for their enterprise, comprehending the needs of their target market, and choosing the right market. These systems are capable of constant learning and adaptation with the use of machine learning algorithms, improving Forecast the location and success of a startup throughout time. The inadequacies of conventional methods—inefficient resource utilization, poor managerial judgment, and imprecise decision-making—give rise to the necessity for startup success prediction and location advice. Prediction is a viable option that gives business owners data-driven insights to increase output.

The remaining portion of the suggested system is divided into the following categories: Section II presents the motivation to work in this area. Section III discusses relevant work information to predict start up success. In section IV, proposed system architecture is discussed. In Section V, the technique for predicting startup success and location recommendation is examined and addressed. The experiment's results and a comparison with other prediction model are covered in detail in Section VI, which is followed by Section VII, which concludes the paper.

II. Motivation

The Predicting Startup Success and Location Recommendation System project was started with the sincere goal of enabling investors and aspiring entrepreneurs to participate in a thriving startup ecosystem. It will benefit those who aspire to launch their own companies. Establishing a business is an exciting journey, but it can also be rather challenging. People frequently are unsure about whether they are making the best choice. The main goals of Project are to help individuals succeed, make it simpler for people to launch fantastic enterprises, and foster the development of innovative new ideas.

Objective of the system are predicting startup success and location recommendation to business.

III. Literature Survey

The literature survey involved an in-depth review of studies focusing on machine learning applications in business Startups particularly and factors regarding startup. Insights were gathered from research highlighting the integration of machine learning algorithms for precise startups success based on industry. Additionally, the survey included examinations of approaches about failure rate of startup and risk analysis tailored to specific losses and shut in early stage.

Harjo Baskoro at el., in [1], emergence and expansion of startups has become a global phenomena in recent times. Startups are now seen as essential to innovation and economic expansion in many countries. Nonetheless, studies show that a company's chances of failing are about 90%. Consequently, it is imperative that investors, financial advisors, and the government pinpoint the 10% of the market that will ultimately yield greater returns, generate additional capital, and ensure economic growth. The purpose of this study is to identify the critical components of startup success that can be used to build a predictive model that forecasts startup success using a machine learning algorithm.

Saeed at el., in [2], purpose of this study is to identify a subset of factors that significantly affect the performance of small industrial businesses. Additionally, it builds a statistical model that can be used to calculate the predicted survival period and assess the likelihood that each small industrial firm will falter. It uses the Cox & Risqué regression model with cluster analysis to categorize based on variables, i.e., non-faltering and altering.

Afolabi, at el., in [3], aims to show a system's design and implementation for the purpose of diagnosing, predicting, and offering recommendations for a business's success. The prediction model was developed using a combination of the Naïve Bayes and J48 classification algorithms, as well as correlation analysis for pre-processing the data. The examination of current business consulting systems, expert systems, and human specialists in the field of business consulting in Nigeria produced the necessary heuristics for the diagnosis.

Sunitha Cheriyan, , at el., in [4], presents a thorough examination and evaluation of easily understood predictive models in order to enhance future sales projections. Traditional forecast methods are tough to deal with the huge data and accuracy of sales forecasting. Several data mining techniques could be used to solve these problems. Based on an assessment of performance, the most appropriate. A predictive model is recommended for the forecast of sales trends.

Malhar at el., in [5] proposed healthy economy factors not a thought about economical crisis. Not focusing on timely success as time always matter in any startup Success and waiting for a successful startup many small enterprises shutdown businesses in early stage of startup.

Denisat el., in [6] Considering a prediction without revenue factor leads to non growing startup and not a space to enter a new startup. Historical data play major role in building new startup idea as it only consider revenue and used SVM with accuracy of 89 %.

Amar Krishna et., in [7] Focusing on less failure more success leads to improve parameters like Funding rounds, IPO, Fundings for finding success and failure but this feature does not focus on long term businesses it may leads to downfall drastically.

M. Bangdiwala et., in [8] The purpose of this study is to assess a startup's likelihood of being acquired or merged. Five models that predict whether or not a firm will be purchased have been developed and compared using historical data on startups. Decision trees, Random Forests, Gradient Boosts, Logistic Regressions, and MLP Neural Networks are the models that have been applied. These models are trained on a variety of important factors, including fundraising rounds, investments, valuations, and so forth.

An overview of the literature on predicting startup success is compiled in Table 3.1.

Table 3.1 Summary of the research on predicting startup success

Author	Methodology	Key Predicators	Findings
Angel et al. [9]	Statistical Analysis	Financial metrics, Founder characteristics	Experience and education of founders have a beneficial effect on startup success.
Chen et al. [10]	Regression Analysis	Prior entrepreneurial experience, Founder traits	Success in entrepreneurship is significantly predicted by prior experience.
Feld [11]	Case Studies	Local economic conditions, Geographical location	Startup results are influenced by local resources and regional conditions.
Klotz et al.[13]	Machine Learning (SVM, Random Forest)	Market conditions, team dynamics, Founder characteristics	By examining intricate relationships, machine learning models can increase the accuracy of predictions.
Gentry & Gable [12]	Hybrid Predictive Models	Financial and non-financial metrics	Accuracy is increased by hybrid models that combine many predictors.

With a focus on different methodologies and their contributions to the understanding of startup success prediction, this table offers a thorough overview of the body of extant literature. By going through literature, we came to know some of the gaps in prediction of startup success.

More extensive and superior quality datasets are required. Numerous models now in use rely on data that might not accurately reflect the dynamic and varied character of startups. The intricate relationships between variables may not be included in the current models or they may be overly simplified. More sophisticated machine learning methods or hybrid models might provide more accurate forecasts.

Generic predictors are used in a lot of models. It would be interesting to investigate the ways in which industry-specific factors affect startup success and how these characteristics differ in other industries. Regional differences may be substantial in startup success predictors. Enhancing prediction accuracy could be achieved by comprehending how regional economic, cultural, and legal aspects affect success.

By filling in these gaps, governments, investors, and entrepreneurs may all benefit from more precise and useful forecasts of startup success.

IV. Methodology

The proposed system collects information about the parameters like Funding, Founding year, revenue for Startup Success Prediction and parameters like season, age group, year, business type for location recommendation based on machine learning algorithm. The dataset for the application uses two different dataset for Startup Success Prediction and location recommendation considering separate parameters for each module. The initial approach is to trained model Random Forest on both the data. Models are trained to classify given input success or failure and recommended location on the input given to the system.

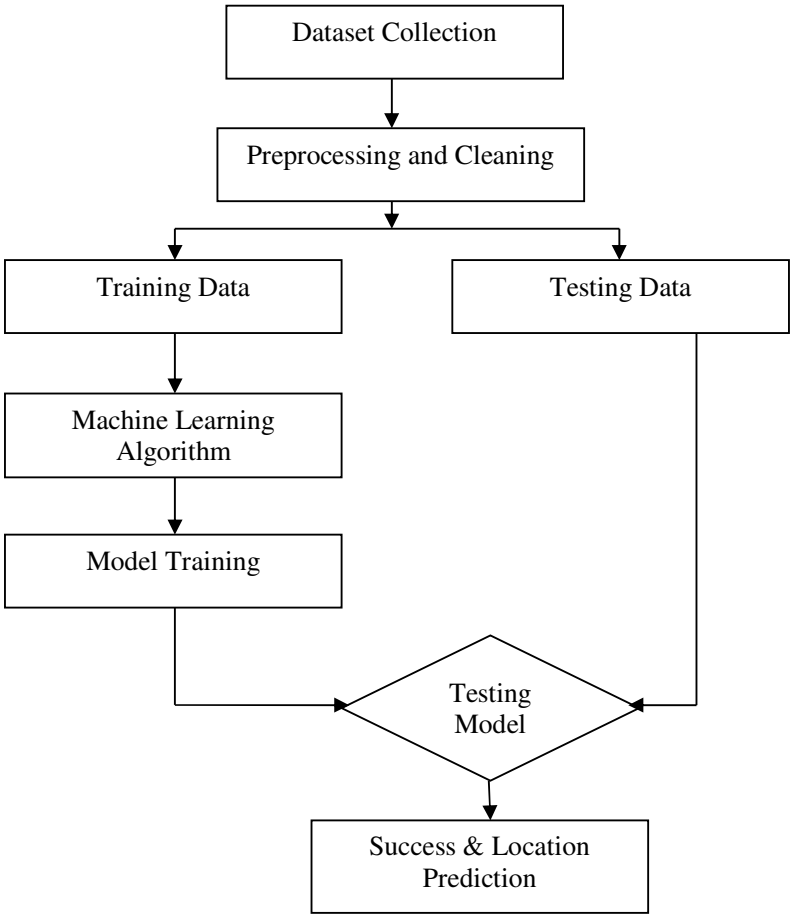

Figure 1: System Architecture or Process Diagram

Figure 1 show the system architecture, which works in the following steps:

1. Location: Starting with choosing a domain for recommend location.
2. Startup: Starting with choosing a domain for predicting success or failure of startup.
3. Data sources: As per the domain, data sources will be different, if success of startup is chosen dataset is considered to have startup features and if location is preferred then features are different about location recommendation.
4. Success factors: Consist of factors responsible for success.
5. Data extraction: Collects gathered data from data sources.
6. Pre-processing: Gather suitable data for models

7. ML model: Choose appropriate machine learning algorithms on the basis of accuracy of model for startup success prediction and location recommendation by separating training and test data for building model.
8. Prediction: By using ML model of startup success prediction predicts startup success or failure. Recommend location using ML model of location recommendation.
9. Local location recommendation: Getting output from prediction and display recommended local location using ML model.
10. Success/ Failure: Getting output from prediction model and display success or failure of startup.

The collection and pre-processing of data is the initial step in our methodology. We plan to collect startup data from multiple sources, such as AngelList, Crunchbase, and other publicly available databases. The data will include details about the company, like the date of founding, the amount of cash raised, the industry, the number of employees, and the background of the founder. We will also offer market data, like market size and competitiveness. After the data is gathered, we will pre-process it to remove any irrelevant or missing information. Furthermore, feature engineering will be utilized to extract significant attributes from the data. During this stage, the data is formatted to make it easier to use for analysis.

Selecting the optimal machine learning algorithm to forecast startup success is the second step in our methodology. We'll use techniques like decision-making tree, random forest, SVM, Naïve bayes, and logistic regression. These techniques have been demonstrated to be helpful in predicting startup success and are frequently applied to classification problems. The data will be split into training and testing datasets. The testing dataset will be used to assess the models' performance once they have been trained using the training dataset. The models' performance will be confirmed by cross-validation, and the hyper parameters will be adjusted to improve accuracy.

Using decision-making tree, random forest, SVM, Naïve bayes, and logistic regression, the model is trained and refined. It is discovered that random forest is a useful tool for predicting the success of new firms. To increase the model's accuracy, track its performance over time and retrain it as needed.

The application presents an integrated approach for Prediction and Recommendation by employing machine learning algorithms. Collect information and decide which aspects of the data are most important for understanding success and failure of the specific startup and suitable location for startup. Divide the labelled data into two sets, one for training the model and one for testing its accuracy. Evaluate the accuracy of the trained model by feeding it with the testing data and comparing its predictions with the actual startup status and location labels. Gather required data of startup from users and make prediction of their startup success status and suitable location.

Algorithm for Predicting Startup Success and Location Recommendation:

1. Random Sampling: Randomly select n samples from the training dataset. This creates a subset of the original dataset
2. Feature Subset Selection: Randomly select a subset of features of startup success prediction and location recommendation from the dataset.
3. Decision Tree Construction: Grow a decision tree using the selected subset of samples and features. Randomly select m features from the feature subset. Choose the best split point for the node among the selected features. Split the node into child nodes based on the selected split point. Repeat this process recursively for each child node until a stopping criterion is met, such as reaching a maximum depth.
4. RepeatSteps1-3: Repeat steps 1-3aspecified number of times to create a collection of decision trees
5. Aggregation: Aggregate the predictions of all decision trees to make a final prediction:
For regression tasks: Average the predictions of all trees.
For classification tasks: Use majority voting to determine the final class label.
6. Final Prediction:
For regression tasks, the final prediction is the average prediction of all trees.
For classification tasks, the final prediction is determined by majority voting among the predictions of all trees
7. Evaluation: Evaluate the performance of the Random Forest model using a separate validation dataset or through cross-validation techniques.

The techniques or libraries used for implementation with the dataset are :

1. **NumPy**: Numerical Python(NumPy), is an open source python library broadly used in scientific calculating. It backs large, numerous-dimensional arrays and matrices, Along a wide array of high-level mathematic functions For manipulating these arrays. NumPy's role in Python Numerical calculations is significant, shaping the core of Scientific calculations libraries.

When handling and modifying data for predictive modeling, NumPy is essential. It helps with several numerical computations, feature engineering, data preparation, and model validation. NumPy may be integrated with other libraries, such as scikit-learn, pandas, and matplotlib, to create strong prediction models and learn more about the elements that lead to company success.

2. **Pandas**: Pandas have beena Python library that is quite Popular, majorly used for data manipulate and analysis. It has been built on NumPy library and provides data Structure and operation to manipulate numerical tables and series time. The name "Pandas" come from "panel Data", which is an econometrics term for datasets that Involve observation over many time periods for identical individuals. To handle missing values, clean up the data, and get rid of duplicates, use libraries like pandas.
3. **Matplotlib**: Matplotlib being a vast library for static, animated, and interactive visualizations in Python. The popularity of it lies in its strong plotting abilities, laying The groundwork for Seaborn and Pandas plotting, making it highly utilized in the Python programming sphere. To visualize the distributions, correlations, and linkages of data, use libraries such as

matplotlib, seaborn, or plotly.

4. **Seaborn**: Seaborn being a Python data visualization library consist on matplotlib. It gives an elevated level connection to draw captivating and informative statistical graphics. Seaborn is especially suited to visualize intricate datasets that will be analyzed with statistical techniques. It is build on top of matplotlib and nearly integrated with pandas data structures.
5. **Sklearn**: Scikit-learn, or known as Sklearn is a famous and flexible machine learning library for Python. Built on NumPy, SciPy, and Matplotlib, it provides basic and efficient tools for prediction data analysis and statistics modelling. Scikit-learn is broadly used in academia and industry because of its easy usability, performance, and wide array of machine learning algorithms. Utilize methods such as Principal Component Analysis (PCA) available at sklearn to minimize the quantity of features while preserving crucial data.

V. Results and Discussions

The application aims to provide a flexible system that can be easily used. The results of a Startup Success Prediction and Location Recommendation typically depend on the accuracy of the model used and the data train and test. A successful system would provide reliable and most accurate predictions. Figure 2 shows some of the algorithms which are considered for the development of startup success prediction model with accuracy rate. On x axis accuracies of algorithms are given and on y axis the name of algorithms is plotted. The highest accuracy is of Random Forest Algorithm would be the top priority.

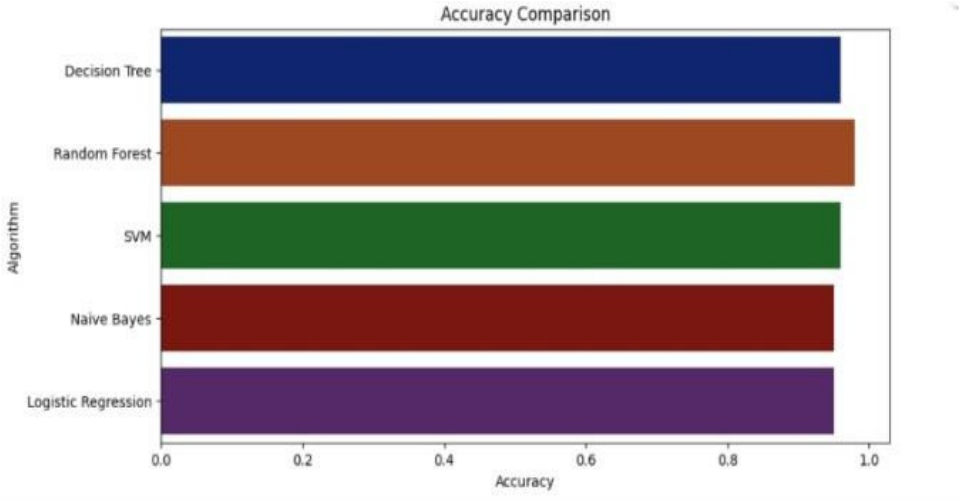

Figure 2: Start Success Model Evaluation

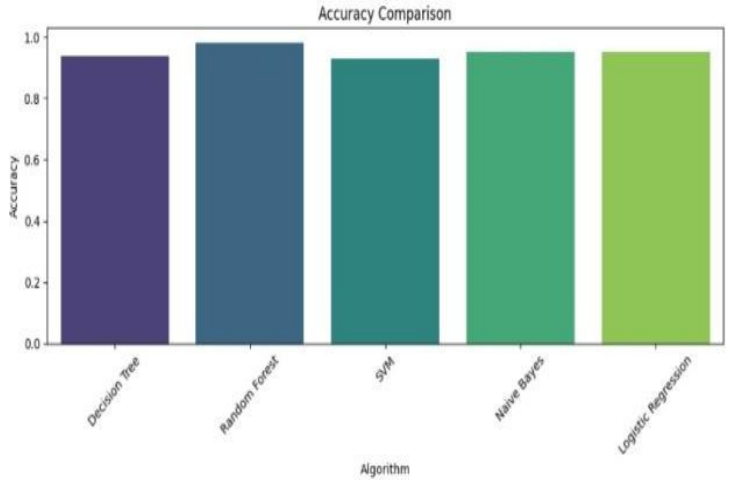

Figure 3: Location Recommendation Model Evaluation

Figure 3 shows some of the algorithms which are considered for the development of Location Recommendation Model with accuracy rate. On x axis name of algorithms are given and on y axis accuracies of the algorithm is plotted.

The implementation has done on open-source python environment, Jupyter notebook for train and test machine learning algorithms The device runs with an INTEL Pentium 4 or Higher processor and 4 GB RAM with a distributed manner on

the Python 3.8, Jupyter analytic platform. Startup dataset is use for startup success prediction and local location dataset used for location recommendation. The above Figure 2 and Figure 3 and demonstrates the accuracy of the proposed system and comparative analysis with various algorithms. The proposed system, Random Forest algorithm for both the model as it having more accuracy than other algorithms

VII. Conclusion and Future work

Predicting Startup Success and Location Recommendation System redefines with precision and efficiency through machine learning. To get more accurate prediction about startup, we consider the factors such as market trend, demand and supply, and risk analysis. Further we added location recommendation for entrepreneurs to start their business. Creating accurate predictive models involves selecting the right algorithms and dealing with imbalanced datasets.

There exist multiple avenues for future research to improve the startup acquisition status forecast through machine learning projects.

Adding New Features: To anticipate the acquisition state, the model currently uses a collection of predefined features. Future research can investigate and incorporate more pertinent features into the model, such as sentiment analysis on social media, industry-specific data, and financial indicators.

Time-series analysis approaches can be incorporated into the model, taking into account the temporal character of the startup data. This would make it possible to record patterns, seasonality, and other time-varying occurrences that could affect the acquisition status.

Increase the Dataset: Although the present model was trained on a particular dataset, the generalizability of the model can be improved by increasing the dataset to encompass a broader and more varied range of startups. By adding startups from other sectors, regions, and phases of development, the model's predictive power can be increased and a wider range of acquisition scenarios can be captured. The future work is to improve accuracy and enhance more feature of the system. The startup acquisition status prediction project can be improved further by pursuing these directions for future research, which will result in more precise forecasts and insightful information for investors and other startup ecosystem stakeholders.

References

1. Baskoro, H., Prabowo, H., Meyliana, M., & Gaol, F. “Predicting startup success, a literature review”. Retrieved November 15, 2022, from https://prosidingicostec.respati.ac.id/index.php/icostec/article/view/6

2. 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

3. 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

4. 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.

5. Malhar Bangadiwala, Yashvi Mehta, “Predicting Success Rate of Startups using Machine Learning Algorithms”, October 2022.

6. Denis Ozornin, Michael Roizner, “Web-based Startup Success Prediction”, November 2021.

7. Amar Krishna, Ankit Choudhary, “Predicting the Outcome of startups: less failure, more success”, June 2022.

8. 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.

9. Angel, J.J., et al., "The Role of the Angel Investor in the Early-Stage Financing of Technology Startups", 2008

10. Chen, J., et al., "The Influence of Founder Characteristics on Startup Success", 2009

11. Feld, B. , "Startup Communities: Building an Entrepreneurial Ecosystem in Your City", 2012

12. Gentry, R., & Gable, M., "Hybrid Predictive Models for Startup Success",2020

13. Klotz, A.C., et al., "Predicting Startup Success: Machine Learning Approaches", 2014