Journal of Namibian Studies, 37 S1 (2023): 427-443 ISSN: 2197-5523 (online)

Examining The Impact Of Entrepreneurial Characteristics On Startup Success Using Machine Learning Techniques

VIVEK PAULOSE[1], Dr. I. JOSEPHRAJ[2] and Dr. JACOB P.M[3]

[1] Research Scholar, Department of Business Administration, Annamalai University.
[2]Assistant Professor, Department of Business Administration, Annamalai University.
[3]HOD & Professor, Naipunnya Business School.

Abstract

The process of foretelling a startup's success or failure is known as startup prediction, involving the forecast of a company's trajectory. Successful prediction requires a blend of analytical expertise, industry knowhow, and a dash of intuition to decipher signals indicating a venture's potential for sustained growth and innovation. The task is formidable due to the multitude of factors influencing a startup's fate, which can swiftly change in the dynamic startup ecosystem. Despite these challenges, the growing availability of data has propelled the popularity of data-driven approaches, including machine learning, for startup prediction. This research delves into the impact of entrepreneurial attributes on predicting startup success. Entrepreneurial attributes encompass the personal traits and skills inherent to entrepreneurs. Specifically, the study examines the influence of certain entrepreneurial attributes—age, gender, education, and network—on startup success prediction. Crunchbase, a platform offering insights into startup companies, venture capital firms, and private equity firms, provides the data for this study. Various machine learning classification techniques are employed for analysis and compared for performance. The outcomes aim to furnish entrepreneurs, investors, and policymakers with valuable insights on identifying and cultivating attributes crucial for startup success. Furthermore, this study contributes to the existing literature on startup prediction by underscoring the significance of entrepreneurs' personal characteristics in the prediction process.

Key Words: Classification Techniques, Entrepreneurial Attributes, Entrepreneurial Success, Feature Importance.

Introduction

In January 2022, India marked the recognition of over 61,000 startups, solidifying its position as the third-largest startup ecosystem globally, trailing only behind the United States and China. This burgeoning startup landscape has evolved into a pivotal catalyst for the country's growth in recent years, aspiring to establish itself as a prominent global tech hub. Many Indian startups are experiencing significant growth, attracting and empowering the younger generation to actively participate and gain expertise in their respective fields (Tomy & Pardede, 2018). The significance of startups extends beyond their sheer numbers, encompassing various crucial roles. Primarily, they serve as crucial drivers of economic expansion and innovation, often spearheading the development of novel products, services, and technologies. This innovation, in turn, has the potential to birth new industries and propel the growth of existing ones. Furthermore, startups contribute to job creation and income generation, thereby enhancing overall economic well-being. Another critical aspect lies in how startups can provide individuals with the opportunity to seize control of their professional trajectories and pursue entrepreneurial dreams. Initiating a business enables individuals to transform ideas and passions into reality, fostering financial independence and success on their own terms. Additionally, startups promote heightened market competition, bringing superior products, quality services, and better prices for consumers (A Nonfinancial Business Success versus Failure - ProQuest, n.d.). This competitive environment compels established companies to maintain honesty, continuously innovate, and enhance their offerings. Moreover, startups play a vital role in local communities, offering support through job provision, backing local suppliers, and contributing to community development via taxes and other economic activities. In essence, startups are deemed crucial for propelling innovation and economic growth, offering opportunities to individuals and communities, and cultivating a more competitive and dynamic business landscape.

Startups play an essential role in the growth of developing economies, exemplified by their significance in countries like India. The term "startup" denotes a fledgling company in its initial operational phases, concentrating on a specific product or service intended for the market. Typically initiated by one or more entrepreneurs, startups aim to fulfil a perceived demand with the founders often providing the initial funding. Entrepreneurs, defined as individuals assuming the majority of risks and reaping great rewards, are viewed as innovators responsible for introducing novel ideas, products, services, and business processes. Despite the inherent high risk, entrepreneurship is recognized for its potential to stimulate economic growth, generate wealth, and foster innovation. Statistics reveal that nearly 90% of startups face failure, underscoring the importance for investors and venture capitalists to discern promising ventures. Predicting the success of a startup, however, remains challenging due to the inherent uncertainty associated with these ventures. This study focuses on evaluating the impact of entrepreneurial attributes on startup success, acknowledging the multifaceted factors influencing their performance.

Limited research has been conducted on the factors predicting the success of startups in the Indian context, resulting in a scarcity of relevant literature in this domain. Previous studies predominantly focused on market conditions, financial performance, and overall company evaluation, yet these investigations encountered limitations with small datasets, typically comprising sample sizes ranging from 200 to 300. Recognizing the multifaceted nature of factors influencing startup success in India, this study aims to fill the gap by specifically examining the impact of entrepreneurial attributes such as age, gender, educational qualification, and prior funding experience. To enhance the robustness of our findings, we employ a larger dataset with a sample size of 11,288. Additionally, this research endeavours to identify the most effective machine learning model for predicting startup success in the Indian context.

The role of startups in driving substantial economic growth and employment opportunities is undeniable. Identifying the key factors contributing to a startup's success is crucial for policymakers and business leaders, as it enables the creation of more conducive environments for startup flourishing. This, in turn, has the potential to stimulate heightened economic activity and job generation. Valuable insights for aspiring entrepreneurs and investors seeking to initiate or invest in new ventures can be gathered from the experiences of startups (Sharchilev et al., 2018). Among the most influential determinants of startup success, the attributes of the entrepreneur stand out prominently. This becomes especially significant in the initial stages of a startup when historical data is unavailable, making the entrepreneur and their attributes the primary measurable factors. Understanding these factors is paramount for investors interested in the early stages of startups, as it provides a crucial guide for making informed investment decisions.

This paper exclusively concentrates on Indian startups, seeking to comprehend the specific characteristics of entrepreneurs that play a role in the success of these ventures.

Literature Review

Forecasting the success of a venture has always presented a formidable challenge, yet it holds immense significance for both public and private entities interested in providing funds, making investment decisions, and establishing new enterprises. Statistics reveal the precarious nature of startups, with a staggering 90% ultimately meeting failure. Breakdowns indicate that 21.5% falter within the inaugural year, 30% succumb in the second, 50% by the fifth, and a substantial 70% by the tenth year. Within the realm of growing startup ecosystems, financial backing emerges as a key determinant. Particularly for technology-centric startups, financial constraints are a common hurdle. Notably, well-funded startups tend to outshine their underfunded counterparts (Żbikowski & Antosiuk, 2021), underscoring the key role of funding in determining success. Paradoxically, even among ventures backed by capital, over 75% struggle with failure or manage to sustain only a marginal existence. While predicting business success remains a formidable task, stakeholders in both public and private spheres grapple with this challenge as they navigate funding decisions, investments, and the establishment of new enterprises. As a startup matures, undergoes rigorous product-market fit testing, and navigates the selection process of angel investors and venture capital funds, the predictive task becomes somewhat easier. Ultimately, a startup can be deemed successful when it launches an IPO (Initial Public Offering) or undergoes a merger or acquisition.

Numerous research studies have probed into the dominion of startups, exploring a diverse array of subjects. These investigations span a broad spectrum, addressing factors influencing startup success, the repercussions of government policies on the initiation and expansion of startups, and the significance of entrepreneurial networks in fostering the evolution of nascent enterprises (A Nonfinancial Business Success versus Failure - ProQuest, n.d.).

In recent years, significant focus has been directed towards researching startup ecosystems. Scholars in this domain explore the multifaceted components that shape the establishment and advancement of startup communities. These elements encompass factors like funding availability, the existence of supportive entrepreneurial organizations, and the cultural and regulatory backdrop (Makridakis, 1996). The objective of these studies is to unravel the intricate interactions among diverse factors, elucidating how they collectively forge an environment conducive to the inception and progression of new ventures (Picken, 2017).

Another extensively explored research domain pertains to the influence of technology on the success of startups. Scholars have scrutinized how technological advancements facilitate novel business models, exemplified by digital platforms and the sharing economy. Additionally, investigations have been conducted on the repercussions of these changes on startups in sectors such as artificial intelligence, biotechnology, and green tech (Tomy & Pardede, 2018). Another area of research is Venture Capital and fundraising, which explores the intricacies of startup funding processes involving angel investors, venture capitalists, and other investors. This involves analysing factors such as the company's stage, investment size, and the valuation's impact on fundraising, subsequent outcomes, and startup performance. The collective findings from studies in the startup realm provide a valuable repository of insights into the elements influencing the establishment, evolution, and triumph of emerging ventures (Tomy & Pardede, 2018).

The process of startup prediction is complex due to numerous influencing factors, including the calibre of the founding team, the robustness of the business model, the size and growth potential of the market, and the competitive landscape (Bangdiwala et al., 2022). Researchers often employ statistical techniques to construct predictive models based on historical data pertaining to startup outcomes. These models enable the assessment of the likelihood of success or failure for new startups based on their distinctive attributes. Machine learning algorithms come into play as researchers analyze extensive datasets encompassing financial metrics, social media activity, and news articles to discern patterns associated with success or failure. Additionally, insights from industry experts, such as venture capitalists or accomplished entrepreneurs, may be incorporated to gauge the likelihood of a startup's success (Ang et al., n.d.). The evaluation of critical aspects of the startup business model, such as scalability, market size, and competitive landscape, is integral to this predictive process.

Forecasting the outcomes of startups poses considerable challenges, given that success and failure are not often binary and are influenced by multifaceted factors that resist easy measurement and prediction. Moreover, the futures of many startups remain uncertain, subject to substantial shifts based on new information or external events (Why Startups Fail, n.d.). Notably, startups backed by venture capital or other forms of investment tend to attract more attention and scrutiny, resulting in comprehensive and readily accessible data. In contrast, startups without funding may lack comparable visibility, potentially impacting the precision of prediction studies.

The success of a startup hinges on a myriad of factors, encompassing the startup's stage, industry, and the founders' objectives (Wasserman, 2017). A pivotal element is the quality of the business model; startups equipped with a lucid and sustainable business model, defining a target market and establishing a competitive edge, are more prone to success compared to those lacking a well-defined plan (Fujita et al., 2022). Another critical determinant is the founding team;

startups boasting robust and diverse teams, featuring complementary skills and experience, stand a higher chance of success than those with weaker teams. Additionally, teams with a history of successfully launching and growing companies are more likely to thrive. The availability of funding represents a third crucial factor, while the size and potential of the market constitute a fourth (Andrew H. Van de Ven et al., 1984). Startups targeting expansive and burgeoning markets are generally more poised for success than those focusing on smaller or stagnant markets. Beyond these factors, external elements such as government policies, cultural attitudes towards entrepreneurship, the quality of local support networks, and access to resources and customers also exert influence on a startup's success.

Decrease in market demand can be one of the reasons for the failure of a startup. Initially, a startup may expect high demand for its product or service, but to realize their expectations to be futile in future. The success of a startup can be influenced by external factors, such as natural disasters or pandemics like Covid, as emphasized by (Ünal, 2019). Inadequate financial resources, in addition to market dynamics, present a significant challenge. Startups that secure sufficient funding for the development and scaling of their products or services are more likely to thrive, in contrast to those facing difficulties in fundraising. The survival of a startup depends largely on essential financial support. Internal factors, including managerial decisions, are equally critical in shaping the success of a startup, as noted by (von Gelderen et al., 2000). A strategic mindset that embraces change is crucial for companies aiming for future success. The capacity to foresee and evaluate the implications of impending changes is intricately linked to performance. Additionally, the success of startups is influenced by their developmental stage; some may secure capital effectively but struggle to generate revenue or achieve profitability. Others may not achieve substantial scalability but maintain a steady revenue stream. Ultimately, the goals and aspirations of founders and the company itself contribute to the varying degrees of startup success.

This study employs a machine learning approach to investigate these dynamics, wherein machine learning (ML) denotes a system's capability to acquire and assimilate knowledge on a large scale. Unlike traditional programming, ML enables systems to learn and adapt by integrating new information. ML operates as a subset of artificial intelligence (AI), providing machines the ability to learn without explicit programming (Woolf, 2009). Computer systems, utilizing ML, execute tasks such as clustering, calculations, and pattern recognition. The learning process involves employing diverse algorithms and mathematical structures to analyze information, which is characterized by features. ML is employed to establish relationships between these features and corresponding output values referred to as labels. This technique is particularly well-suited for addressing problems related to regression, classification, as well as the determination of collection and association rules.

Research Methodology

i. Dependent Variable

Success: Success, in this context, is the outcome variable determined by the startup's achievement or lack thereof. This is assessed based on whether the startup has undergone acquisition, undergone an initial public offering (IPO), or is actively operating post the completion of series B funding round. The mathematical representation is as follows:

y=1 , if xstatus = acquired ∨ x status = ipo ∨ (xstatus=operating ∧ xstatus=seriesb)
y=0 , otherwise

The outcome variable is binary where 1 represents success and 0 represents failure. The startup's current standing, as well as the funding secured, can be extracted from Crunchbase data.

ii. Independent Variables

Age: Age here refers to the entrepreneur's age at the initiation of the startup, which is a continuous positive integer variable.

Gender: Gender of the entrepreneur. It is classified into male and female. It is encoded as male = 0 and female = 1

Degree: The highest educational qualification of the entrepreneur. It is divided into 3 – Masters, Bachelors and Primary education. They are encoded as Primary= 0, Bachelors= 1, Masters= 2.

Previous Funding Experience: The measurement is based on the likelihood of a startup's success when an investor makes an investment. In this assessment, the investor takes into account the entrepreneur's past funding experiences as a network attribute. Consequently, the proxy for network strength is determined by the measured probability of investment success for a startup, with values ranging between 0 and 1.

iii. Process Flow Chart

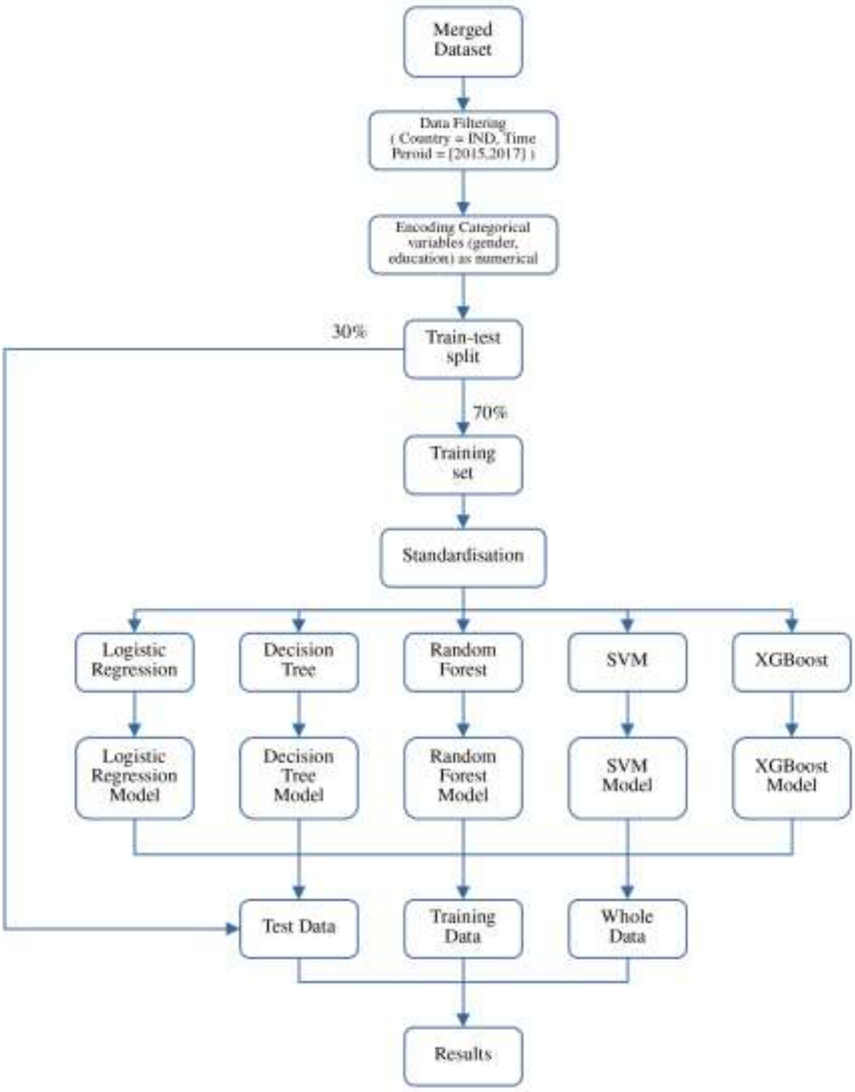

Figure 3.3.1: Conceptual Model

iv. Research Design

Data for this study is secondary data taken from Crunchbase. The files for data analysis are merged to form a single dataset and used for analysis. The dependent variable, which is success, is calculated using the operational status and funding of the company.

Classification technique is used for analysis. This study utilises five algorithms for classification: -

• Logistic Regression
• Decision Tree
• Random Forest
• Support Vector Machine
• XGBoost

The data is split into train set and test set in 70:30 ratio. The machine models will be formed by training the train dataset and will be applied on train data, test data and whole data for analysis. The models will be tested based on 4 metrics:

• Accuracy
• Precision
• Recall
• f1 score

All these scores are classified based on confusion matrix

Confusion Matrix	**Actual**		
Predicted		**True**	**False**
	True	a	b
	False	c	d

Table 3.4.1: confusion matrix

Classification accuracy metric is commonly utilized to evaluate the efficacy of a classification model. This represents the proportion of correctly classified instances relative to the total instances in a dataset. To calculate accuracy, one divides the number of instances correctly classified by the total number of instances and then multiplies the result by hundred to express it in percentage."

$$\text{Accuracy}=\frac{a+d}{a+b+c+d}$$

Precision in classification serves as a metric gauging the percentage of true positive predictions out of all positive predictions made by a classification model. Precisely, it calculates the ratio of true positive predictions to the total positive predictions generated by the model. True Positives (a) refer to instances correctly identified as positive (Shung, 2020), whereas False Positives (b) indicate instances wrongly labelled as positive. A higher precision value implies that the model is effective at minimizing false positives, meaning that when it predicts an instance as positive, it is generally correct. It is important to note that precision is often assessed in conjunction with recall, which evaluates the proportion of true positive predictions compared to all actual positive instances, providing a comprehensive evaluation of a classification model's performance.

$$\text{Precision}=\frac{a}{a+b}$$

In classification, recall measures the proportion of true positive predictions (instances accurately identified as positive) relative to all actual positive instances within a dataset. True Positives (a) are instances correctly identified as positive, while False Negatives (c) represent instances incorrectly labeled as negative. A higher recall value indicates that the model effectively identifies a significant portion of positive instances in the dataset. This metric is particularly crucial in applications like medical diagnosis, where the focus is on identifying all positive cases, even if it means accepting some false positives. It's important to note that recall is often evaluated alongside precision, which assesses the ratio of true positive predictions to all positive predictions generated by a classification model, providing a comprehensive evaluation of the model's performance.

$$\text{Recall}=\frac{a}{a+c}$$

The F1 score provides a comprehensive evaluation of a model's accuracy, considering both precision and recall. Calculated as the harmonic mean of precision and recall, the F1 score offers a balanced assessment on a scale from 0 to 1. A higher F1 score indicates superior model performance. This metric is particularly beneficial in scenarios with imbalanced data, as it equally values both precision and recall. Essentially, the F1 score proves invaluable when seeking a trade-off between precision and recall in the presence of uneven class distribution. It's essential to note that an F1 score of 1 signifies perfect precision and recall, while a score of 0 suggests that the classifier is not making any correct predictions.

$$\text{F1Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

This study utilizes a sample comprising 11,288 startups from India, sourced by merging data from various files: organizations.csv, people.csv, degrees.csv, funding_rounds.csv, and investors.csv, all obtained from Crunchbase and taking uuid as common column. The dataset was filtered to include startups founded between 2015 and 2017, specifically those with the country designation 'India.' This filtering was applied to focus on startups in their 5–7 years of existence, aligning with the conditions for success that define the dependent variable. This approach is essential as recently established companies are in their early operational stages, making it challenging to conclusively determine their success based on the criteria used in this study. Analyzing older companies would introduce complications, as the impact of independent variables tends to change over time. Therefore, restricting the study to a specific timeframe allows for the development of more accurate models."

Findings and Discussions

Inferential statistics

Since the dependent variable is binary, classification technique used for machine learning analysis. The models used will be compared using accuracy, precision, recall and f1 score. The model is created and analysis is done for test data, train data and whole data. Every model created is analysed individually and compared among themselves. Also the model are compared among different datasets.

Train data model comparison				
	Accuracy	**Precision**	**Recall**	**f1**
Logistic Regression	0.83000	0.72009	0.77879	0.74829
Decision Tree	0.86103	0.74863	0.83272	0.78844
Random Forest	0.85850	0.76985	0.81142	0.79009
SVM	0.83167	0.72338	0.77499	0.74830
XGBoost	0.84825	0.76107	0.79208	0.77626
Test data model comparison				
	Accuracy	**Precision**	**Recall**	**f1**
Logistic Regression	0.82964	0.71147	0.75691	0.73349
Decision Tree	0.80278	0.63262	0.73237	0.67885
Random Forest	0.81665	0.70699	0.72853	0.71760
SVM	0.82994	0.73656	0.74457	0.74054
XGBoost	0.82669	0.73656	0.73722	0.73689
Whole data model comparison				
	Accuracy	**Precision**	**Recall**	**f1**
Logistic Regression	0.83000	0.71499	0.77369	0.74318
Decision Tree	0.83815	0.70876	0.79441	0.74914
Random Forest	0.84080	0.74435	0.77896	0.76126
SVM	0.83133	0.72305	0.76857	0.74511
XGBoost	0.83992	0.75344	0.77169	0.76246

Table 4.1.1: Model Comparison

All machine learning algorithms employed in this study demonstrated the ability to generate accurate models for predicting startup success based on entrepreneurial attributes. The evaluation metrics consistently yielded high values, approximately 0.8, indicating strong predictive capabilities across all models. Overall, the models exhibited robust performance, boasting accuracy scores surpassing 0.83 and F1 scores exceeding 0.73 in various tests. Remarkably, the performance metrics exhibited comparable values across all models and datasets.

However, slight variations were observed in model performance across different datasets. In the test dataset, the decision tree model displayed the lowest accuracy and F1 score among all models, while SVM and XGBoost models achieved the highest F1 scores. Contrarily, in the training dataset, the Decision Tree emerged with the highest accuracy and F1 score, whereas Random Forest demonstrated the highest precision. When considering the entire dataset, Random Forest stood out with the highest F1 score, while the decision tree model exhibited the highest recall. In summary, Random Forest emerged as the superior model across all metrics and datasets.

Feature Importance

Feature importance is a crucial phase in constructing a machine learning model, wherein the scores for all input features are computed to determine their significance in the decision-making process. A higher score assigned to a feature indicates a more substantial impact on the model's ability to predict a specific variable.

	Degree	Gender	Age	Prev_fund_prob
Logistic Regression	0.59480	0.18051	1.72824	0.72578
Decision Tree	0.08475	0.02485	0.69807	0.19233
Random Forest	0.10900	0.01673	0.64024	0.23403
SVM	0.40698	0.13310	1.30073	0.47373
XGBoost	0.34711	0.06093	0.49646	0.09551

Table 4.2.1: Feature Importance

When examining the feature importance across various models, it becomes evident that the age feature consistently ranks as the most crucial predictor in all models, indicating its strong influence on the outcome variable. The degree feature holds significant importance as well, particularly in the XGBoost model where it is second only to age. It also exhibits notable importance in Logistic Regression and SVM models but is less emphasized in Random Forest and Decision Tree models. Conversely, the gender feature emerges as the least important in all models, suggesting its limited predictive power on the outcome variable. The prev_fund_prob feature is identified as the second most important in Logistic Regression, Random Forest, Decision Tree, and SVM models, although it lacks significance in the XGBoost model. This comparative analysis implies that age is the foremost influential entrepreneurial feature affecting startup success, followed by the feature related to previous funding experience. Degree also wields a considerable influence, while gender exerts negligible impact on predicting startup success.

Limitations of the study

The study is constrained to Indian startups and focuses exclusively on those founded between 2015 and 2017. The dynamic nature of conditions over time necessitates the periodic updating of the model to maintain relevance. Predicting startup success is contingent on numerous interacting variables, some of which are inherently challenging to analyze, adding complexity to the study. Moreover, the inherent uncertainty of startups exposes them to a diverse array of risks that can significantly impact their success. Unforeseeable, high-impact events such as pandemics, natural disasters, or unexpected technological breakthroughs may occur, exerting a profound influence on startup outcomes but remain challenging to incorporate into the model. Additionally, attributes like vision, resilience, problem-solving capabilities, leadership skills, and risk-taking abilities, while influential, are complex to measure and quantify. Consequently, this study is confined to a subset of influencing factors, recognizing the limitations posed by the complexities of entrepreneurial attributes.

Implications of the Study

The study underscores the significance of certain attributes for investors considering startup investments. Notably, middle-aged entrepreneurs exhibit a higher probability of success, making age a primary factor in evaluating startup viability. This prominence of age in the predictive models suggests that experience gained over time positively influences entrepreneurial success, with industry experience contributing valuable insights into business dynamics.

Educational qualifications emerge as another pivotal factor, ranking as the second most important feature. A higher level of education corresponds to enhanced knowledge and skillsets, increasing the likelihood of entrepreneurial success. Consequently, entrepreneurs with advanced educational backgrounds are positioned for greater success.

The third crucial feature identified in the study is the network attribute, highlighting the importance of an entrepreneur's professional connections in influencing startup outcomes. Conversely, gender exerts a relatively lower influence on the machine learning models, emerging as the least significant among the four attributes.

These findings provide valuable insights for investment decisions, suggesting that age, education, and network attributes are key considerations for predicting startup success. However, it's important to acknowledge the existence of qualitative factors, such as soft skills, which, while challenging to measure, could further enhance the predictive accuracy of these models if incorporated.

REFERENCES

Lussier, R. N. (1995). A nonfinancial business success versus failure prediction mo. Journal of Small Business Management, 33(1), 8.

Ang, Y. Q., Chia, A., & Saghafian, S. (2022). Using machine learning to demystify startups' funding, post-money valuation, and success (pp. 271-296). Springer International Publishing.

Bangdiwala, M., Mehta, Y., Agrawal, S., & Ghane, S. (2022, August). Predicting success rate of startups using machine learning algorithms. In 2022 2nd Asian Conference on Innovation in Technology (ASIANCON) (pp. 1-6). IEEE.

Decision Tree. (n.d.). Geeks for Geeks. https://www.geeksforgeeks.org/decision-tree/

Van de Ven, A. H., Hudson, R., & Schroeder, D. M. (1984). Designing new business startups: Entrepreneurial, organizational, and ecological considerations. Journal of management, 10(1), 87-108.

Fujita, M., Okudo, T., Nishino, N., & Nagane, H. (2022). Analyzing startup ecosystem through corporate networks based on investment relation of venture capitals in unicorns. Procedia CIRP, 112, 11-16.

Makridakis, S. (1996). Factors affecting success in business: management theories/tools versus predicting changes. European Management Journal, 14(1), 1-20.

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

Introduction to Random Forest Algorithm. (n.d.). https://www.analyticsvidhya.com/blog/2021/06/understanding-random- forest/

Sharchilev, B., Roizner, M., Rumyantsev, A., Ozornin, D., Serdyukov, P., & de Rijke, M. (2018, October). Web-based startup success prediction. In Proceedings of the 27th ACM international conference on information and knowledge management (pp. 2283-2291).

Shung, K. P., 2020. Accuracy, Precision, Recall or F1? Medium. https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9

Tomy, S., & Pardede, E. (2018). From uncertainties to successful start ups: A data analytic approach to predict success in technological entrepreneurship. Sustainability, 10(3), 602.

Ünal, C. (2019). Searching for a unicorn: A machine learning approach towards startup success prediction (Master's thesis, Humboldt-Universität zu Berlin).

Von Gelderen, M., Frese, M., & Thurik, R. (2000). Strategies, uncertainty and performance of small business startups. Small Business Economics, 15, 165-181.

Wasserman, N. (2017). The throne vs. the kingdom: Founder control and value creation in startups. Strategic Management Journal, 38(2), 255-277.

What is Logistic Regression? (n.d.). Statistics Solutions. https://www.statisticssolutions.com/ free- resources/directory-of-statistical-analyses/what-is-logistic-regression/

What is support vector machine (SVM)? (n.d.). TechTarget. https://www.techtarget.com/whatis/definition/support-vector-machine- SVM

What is XGBoost? An Introduction to XGBoost Algorithm in Machine Learning. (n.d.). Simplilearn.Com. https://www.simplilearn.com/what-is-xgboost-algorithm-in-machine-learning-article

Feinleib, D. (2012). Why Startups Fail: And How Yours Can Succeed. Berkeley, CA: Apress.

Woolf, B. P. (2010). Building intelligent interactive tutors: Student-centered strategies for revolutionizing e-learning. Morgan Kaufmann.

Żbikowski, K., & Antosiuk, P. (2021). A machine learning, bias-free approach for predicting business success using Crunchbase data. Information Processing & Management, 58(4), 102555.