National Higher School of Technology
Department of Industrial Engineering & Maintenance

Start-up Success Prediction Using Machine Learning Algorithms

Master's thesis in
Industrial Engineering

MAHIEDDINE Maroua,

Advisor:
Dr. GHOMARI Leila

Academic year:
2022-2023

Abstract: Despite the growing interest in entrepreneurship, the success rate of startups is very low. This emphasizes the necessity for additional research to improve the chances of success for startups. By analyzing key features, machine learning (ML) models can provide valuable insights of the factors contributing to startup success. In this article, we delve into the application of ML techniques to predict the outcome of startups. By identifying and analyzing key success factors, we explore the potential of ML models to improve the success rate of startups.

Key-words: Startups, entrepreneurship, Machine Learning, success rate, prediction, IPO, M&A, funding

Contents

1 Introduction 2
2 Data Sources 3
3 Machine Learning Algorithms 4
4 Discussion 6
5 Conclusion 7

Proc G. I. ss of M. E. B. 2022/2023

1. Introduction

Startups have become a popular avenue for individuals and teams to pursue their entrepreneurial dreams. The effect of startups on innovation, job creation and the economy as a whole is a frequently discussed phenomena in the modern economy[1]. Some authors contend that a small portion of the new startup firms are the primary drivers of innovation, while the majority of big companies only have a minimal influence[2]. Nevertheless, launching a successful startup can be very challenging, given the many uncertainties involved. Therefore it is not surprising that a significant portion of startups fail. According to statistics, 90% of startups completely fail, of which 10% fail within the first year, and 70% fail during the second through the fifth year [3].

A startup is a type of business enterprise that is designed to create innovative products or services while operating under conditions of high uncertainty and risk. To gain a better understanding of the unique characteristics of startups, several experts in the field have provided their own definitions of what a startup is. Each definition emphasizes a different aspect of a startup's operations, shedding the light on the many complexities and challenges that entrepreneurs face as they work towards building and scaling their businesses.

Eric Ries defines a startup as: " A human institution designed to create a new product or service under conditions of extreme uncertainty. " [4]. He emphasizes the need for startups to manage high levels of uncertainty in their business operations. Steve Blank, however, defines a startup as a temporary organization that is established to search for a scalable and repeatable business model [5]. Blank's definition highlights the importance of finding a sustainable business model that can grow as the startup develops. Paul Graham, in his definition of a startup, emphasizes the importance of growth. Graham argues that a startup is a company designed to grow rapidly, regardless of whether it operates in the technology sector, receives venture funding, or intends to have a profitable exit strategy [6]. In contrast, Neil Blumenthal defines a startup as a company working to solve a problem where success is not guaranteed [7]. Blumenthal's definition highlights the challenges inherent in the startup process, including the uncertainty surrounding a new product or service's reception in the marketplace.

Although these definitions emphasize different aspects of a startup's operation, they share a common theme: a startup is a business designed to solve problems and create innovative products or services. All definitions acknowledge the high levels of uncertainty and risk involved in a startup's operations, including factors such as financial resources, team dynamics, competencies, market demand, and competition.

Decision makers have traditionally relied on experience, intuition, and trial and error to overcome these challenges. However, with the advent of machine learning and artificial intelligence, there has been an increasing interest in using these technologies to analyze data and predict the success of startups. In 2013, the term "Unicorn" was first introduced to describe a startup that has achieved a valuation of over $1 billion, referring to their rarity[8]. The unprecedented growth of unicorns has made predicting the success of startups a hot topic in the startup ecosystem. Investors, venture capitalists, and entrepreneurs alike are all looking for ways to identify the next unicorn and achieve similar levels of success.

The first step to appropriately formulate the Startup success prediction problem, is defining "success". Generally speaking, it is hard to give a clear and specific definition for "success" and "a successful startup", since it can be viewed differently from different stakeholders' points of view. Let's take the investor's point of view for instance, with the ROI (Return on investment) being mainly his biggest concern, his definition of success is profitability. For the startup founder, on the other hand, success definition is more complicated and subjective. Some founders' views and goals intersect with those of the investors', and seek financial gain. Others perceive success as having a positive social or environmental impact [9].

While constructing the startup success predictive model, researchers had to define startups' success according to their different outcome scenarios recorded in the database they used.
Many researchers define success as profitability and high ROI through going public (IPO - Initial Public Offering, the process by which a startup becomes publicly traded by offering its shares to the public), or through a process of getting merged or acquired (M&A, the process by which the startup gets combined with or acquired by another company) [9–16]. Others criticize this definition, emphasizing that financial metrics are barely relevant to startups especially in their early stage, since it takes time for a startup to become profitable [17, 18]. Instead, funding events

Proceedings of M. I. E. G. I. M. - 2022/2023

were used as a proxy for success. The authors of [9] also expanded their definition of success by including raising follow-on funding amounts, as long as these amounts are greater than the previous funding round (Funding rounds refer to the stages of financing that a startup receives as it grows. The first round is known as seed funding, where a startup raises initial capital to grow and develop its product or service. The next funding round is Series A, it takes place when a startup has achieved a certain level of success. Subsequent funding rounds such as B, C, D, usually occur when a startup is already profitable, they are utilized to boost the startup's valuation before going for an IPO [14]), since such startups have a tendency to grow over time.

The purpose of this article is to explore how machine learning algorithms and data sources can assist us in predicting the success of startups and identifying the key factors that contribute to their success. In addition to reviewing existing research in this field, we will also evaluate the performance of various machine learning algorithms specifically applied to this problem. This approach will help us to identify the algorithms that are most effective in predicting the success of startups and identifying the key factors that contribute to their success. By analyzing the performance of these algorithms, we can determine which ones are the most reliable and accurate tools for making predictions. Through this research, we hope to gain valuable insights into the complex and dynamic world of startups, and to enhance our ability to predict their success.

The rest of this article is structured as follows, in section 2 we will present the data sources leveraged by researchers in this field, this includes an exploration of their strengths and weaknesses. Section 3 will be dedicated to machine learning algorithms used to solve this problem as well as a comparison between them. In section 4, we will present the results of researchers in this field, discussing their findings and potential for future research. Finally, we will summarize the paper with a conclusion.

2. Data Sources

In order to investigate the factors contributing to the startup success, various data sources were used by researchers, ranging from publicly available databases to proprietary sources.

Crunchbase[1] is a plateform that provides data of companies, particularly startups. It contains a large historical dataset of startups accumulated since its launch in 2007. In addition to investors, entrepreneurs, and journalists, who use the platform to track the funding and growth of startups, it has been widely used by researchers working on the startup success prediction problem, to train their models, [9–21] among others. This historical dataset is valuable but it has its shortcomings. some of them were highlighted in the literature. According to [10], the dataset has too many missing values and information, and there was a change in the reporting format. Furthermore, it has a lack of background information about the founders and teams [11]. [18] enriches the dataset with a substantial amount of freely accessible data from both LinkedIn and the web in general.

Another inconvenience is that crunchbase's dataset has a survivorship bias [14], the successful startups are over-presented because in most cases, the ones that failed were not even recorded . Also, some of the variables are not dated, hence, they may not reflect the situation at every single point in time. In order to improve this database and solve this problem (time aspect being neglected), [14] and [15] adopted a time-aware approach. [14] defined 2 time windows:

- Simulation window: represents the period where the company demonstrates its success (3 years).
- Warmup window: the period before an investor decides to invest in the startup (4years).

These windows ensure that an early stage startup will not be older than 4 years at the time of the investment, and that it will show success within the three years that follow (here success is raising new funds). [15] however, only defined one time window within which a startup should evince its success, it starts when an investor makes the decision of investing in a particular startup, and is 18 month long because the time segment between two funding rounds is usually that long [22].

Mattermark[2] is another proprietary database that provides insights on startups markets and investors. Just like Crunchbase, this dataset is also biased towards successful startups, and additional surveys had to be conducted in order to make it consistent enough to be used in research [23].

Another data source; Autopsy, has been used to in-

[1] https://www.crunchbase.com/
[2] https://mattermark.com/

Proceedings of M. F. E. G. I. M. - 2022/2023

vestigate the reasons leading to startup failure [2, 23]. Autopsy is a public database of failed startups' post-mortem reports, built to help entrepreneurs learn from the mistakes of failed startups by providing insights into the reasons that led to the failure.

Data collection is also used, as it provides convenience in ensuring that the sample is representative, with a unified reporting method, and ensures completeness of the dataset. For instance, the Entrepreneurial Research Consortium (ERC) is an international research project that investigates random and representative samples of entrepreneurs during the startup process [24]. Another example of data collection is the Danish Business Authority, an administrative unit under the authority of the Danish Ministry of Business, which generates and collects data on startups [25].

Other sources include Dealroom[3], Deadpool, Indiegogo[4], Kickstarter, LinkedIn and the web in general. Many researchers combined data from multiple sources in order to form a complete database of a representative sample.

The data that these databases contain, represent the different factors and criteria used to evaluate startup success. Some datasets are made specifically to evaluate startups in their early stages [24, 25], therefore factors related to funding were not included. Instead, other factors were taken into consideration, some of them are listed below:

- General information about the founders: such as gender and age
- Human capital: degrees, work experience, and skills of the founder and team
- Motivation: How motivated and what's keeping them motivated
- Process: whether or not they have a business plan, whether or not they receive information and guidance.
- Financial environment: Startup capital
- Network: Industry experience
- Intended organization: Industry type, market, ambition to grow

The studies made on startups in a more advanced stage, used additional factors, such as:

- Average time between funding
- Number of funding rounds
- Number of investors
- Total funding amount

[3]https://dealroom.co/
[4]https://www.indiegogo.com/

3. Machine Learning Algorithms

The startup success prediction problem can be binary or a multi-class depending on how success is defined. The existing solutions to this problem used various algorithms, some of the most used are presented below:

Decision Trees

Decision Trees are a type of tree-based classifiers that classify instances by sorting them based on feature values. A decision tree is built by recursively partitioning the data into subsets based on the values of one of the input features. The goal is to create a tree-like model where each internal node represents a decision based on a feature value, and each leaf node represents the predicted output. Instances are classified starting at the root node, based on their feature values. [26]

Random Forest

Random forest is an ensemble classification learning algorithm that builds multiple decision trees and combines their predictions to make a final one. The algorithm works by selecting a random subsets of features and a subset of the training data, to build each tree. To make a prediction, the algorithm aggregates the predictions of all the trees in the forest and outputs the class that has the most votes. [20]

Logistic Regression

Logistic regression is a binary classification algorithm, which models the probability of an instance belonging to a certain class. During prediction, the algorithm estimates the probability of a new instance belonging to one class and makes a binary decision based on a decision threshold. [19]

K-Nearest Neighbor

K-nearest neighbors (kNN) is one of the simplest machine learning algorithms. Building the model only consists of storing the training dataset. To make a prediction for a new data point, KNN finds the k data points in the training set that are closest to the new data point, and assigns the new data

Proceedings of M. I. E. G. I. M. - 2022/2023

point to the most common class among those k neighbors.

Support Vector Machine

SVMs find a hyperplane in a high-dimensional space that best separates data points of different classes. In binary classification, the hyperplane that maximizes the margin, or distance between the closest data points of each class, is found. SVMs can also handle non-linearly separable data by mapping the data to a higher-dimensional space using a kernel function. SVMs have been widely used in various fields, as they are very powerful for solving complex classification problems. [27]

Neural Network

Neural networks is an algorithm inspired by the structure and function of the human brain. It consists of interconnected nodes called artificial neurons, that process and analyze data to identify patterns and relationships. The nodes are organized in layers, with the input layer receiving the data, one or more hidden layers processing it, and an output layer producing the results. [28]

Table 1: Comparison between ML algorithms

Algorithm	Advantages	Drawbacks	Reported accuracy %
Decision Trees	Ability to handle both categorical and numerical data, ease of use and interpretation [20]	Increased memory consumption [29] [27]	77 [14]
Random Forest	Ability to handle high-dimensional data and provide robust predictions [19]	Computationally expensive	94.18 [11]
Logistic Regression	Easy to implement and train	Sensitive to the outliers in dataset	92.50 [13]
kNN	Ease of implementation	Large storage requirements [27]	73.33 [16]
SVM	Effectiveness in handling high-dimensional data [27, 30]	Poor interpretability [30]	81.70 [14]
Neural Network	Ability to work with incomplete knowledge [31]	Poor interpretability [27, 31]	94.87 [32]

Proceedings of M. F. E. G. I. M. - 2022/2023

The aforementioned techniques have a wide range of applications across various domains and have demonstrated outstanding performance. Nonetheless, it is crucial to thoroughly study the problem at hand, have a good understanding of the strengths and limitations of each algorithm, and then select the most appropriate one.

Table 1 provides an overview of some of the major strengths and drawbacks of each algorithm, along with the accuracy achieved by researchers who have employed the respective technique for the startup success prediction problem. It should be emphasized that the information presented in the table was extracted from the surveyed articles that are cited within the table.

It is important to note that accuracy was the only metric available and common for all the research studies explored, however, other evaluation metrics such as precision, recall, and F1-score should also be considered. Accuracy is the proportion of correctly classified instances out of the total number of instances. Preci-

sion is the proportion of true positive instances among all instances predicted as positive, while recall is the proportion of true positive instances that are correctly identified by the model. The F1-score is the harmonic mean of precision and recall, which provides a balance between the two metrics [33].

We should also highlight that each of these evaluation metrics provides a different perspective on the performance of a model, and the choice of metrics should depend on the specific objectives of the study. For instance, if the goal is to identify as many successful startups as possible, then recall would be a more appropriate metric, as it would measure the proportion of truly successful startups that are correctly identified by the model. On the other hand, if the goal is to minimize false positives and avoid investing in unsuccessful startups, then precision would be a more appropriate metric [12, 14].

In addition to individual algorithms, some researchers have also explored the use of ensemble methods to predict startup success. Ensemble methods involve aggregating the predictions of multiple models, which can be made of the same or different algorithms, to improve the accuracy of the model. For example, [32] used an ensemble neural network approach to predict startup success, which combined multiple neural networks that had different structures (number of hidden layers, and number of neurons in each layer). The authors found that the ensemble neural network approach significantly outperformed individual neural networks and other traditional machine learning algorithms, such as linear regression and logistic regression. The advantage of using ensemble methods is that they can leverage the strengths of multiple models and reduce the risk of individual models making incorrect predictions. However, the main drawback of ensemble methods is that they may require significant resources to implement effectively.

4. Discussion

In the reviewed articles, the authors used various data sources to gather information on startups (Section 2). Some authors designed their own surveys and conducted interviews with with startup stakeholders in order to collect relevant data [24], or to complete an existing database [23], while others were limited by the required time and resources and only used data available on online platforms such as Crunchbase.

Figure 1 illustrates the process of utilizing supervised machine learning to tackle a real-world problem.

The next step in the process is data preparation. This step is essential for building an accurate machine learning model. Researchers used various methods to prepare data. Straightforward data cleaning ways such as removing entire samples with missing values were most commonly used, when the size of the dataset allowed it[9, 11]. However, in the case of a reduced dataset, imputation methods such as mean/mode/median imputation, and kNN imputation should be used [12].

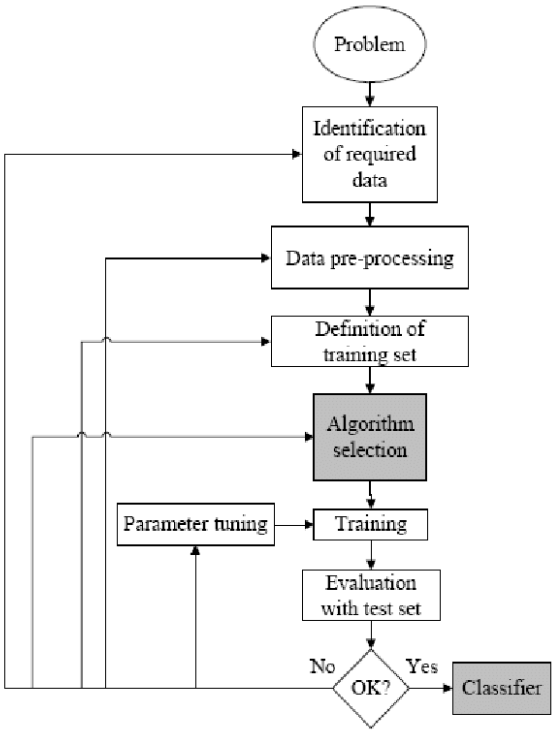


Figure 1: The Process of Supervised ML [27]

Balancing the dataset is also important for improving the performance of machine learning models, as unbalanced classes can cause bias towards the majority class, and therefore a poor model performance. Some studies used oversampling methods such as SMOTE (Synthetic Minority Over-sampling Technique) or ADASYN (Adaptive Synthetic Sampling), to adaptively generate data points for the underpresented class [11, 15]. Adjusting class weights during model training is another technique used to deal with unbalanced classes. This technique involves assigning higher weights to the minority class and lower weights to majority class, which can help improve the model's predictive performance for the minority class. A study

Proceedings of G. I. M. · F · E · 23

compared this technique with oversampling and found that adjusting class weights resulted in superior performance [15], as the model performed better. In addition, oversampling can also increase the training time of the model, especially if the dataset is already large.

Feature engineering is another critical aspect of building predictive models for startup success. Some common techniques used by researchers include selecting essential features [16, 21], and removing unnecessary features such as redundant features, or features that have the same value for all of the instances [19]. Deriving new features is another common technique that can help to capture more information from the data [10]. Additionally, transforming categorical values into binary variables is used to enable the use of categorical data [20].

Researchers employed different algorithms to train their models, as outlined in Section 3, including decision trees, random forests, support vector machines, and logistic regression. Each algorithm performed differently, depending on the data used and the model parameters selected. To optimize the performance of their models, researchers used various techniques for hyperparameter tuning. Grid search was used to tune hyperparameters for decision trees, random forests, and support vector machines [10, 12, 21]. Randomized search was used to tune hyperparameters for random forest and k-nearest neighbor [16]. Bayesian optimization was used to tune hyperparameters for decision trees and XGBoost [9, 15].

Tree-based classifiers are a popular and effective approach in predictive modeling, largely due to their interpretability, which enables researchers to measure the relevance of features used in the prediction models [14, 15]. This is particularly helpful as identifying success drivers is as important as predicting the outcome of startups. Several factors have been identified as having a significant impact on a startup's probability of success. The startup's age, funding amounts, total number of funding rounds, and the location of the startup were found to be among the most influential factors by various studies. Additionally, the reputation of investors and the number of investors associated with a startup were also found to be important factors in predicting startup success [9, 15]. Another study found that incorporating a startup's web presence can improve the quality of prediction. This is because a startup's web presence can indicate its visibility and perception among its target audience [18]. On the other hand, a study on the failure of startups was conducted using a database of failed startups' post-mortem reports [2]. The study shows that the most common reasons of startup failure are: lack or non-existence of a business model, lack of business development, and then comes the financial factor; running out of cash. the first two were not identified as success drivers despite their significant importance (35% and 28% respectively), which highlights the need to study this issue from both perspectives.

However, there are still several limitations in the current research. For instance, most studies only considered direct effects and did not incorporate the "how" variable, which includes the strategies and actions taken by entrepreneurs to develop resources and gain information. Moreover, due to the data being collected at a single point in time, the temporal aspect was overlooked, potentially compromising the reliability of the predictions. Furthermore, the datasets used consisted of startups from diverses industry sectors, which limited the possibility of incorporating customized features that could provide detailed information about startups.

Addressing these limitations can enhance the understanding of startup success prediction, and improve decision-making processes for entrepreneurs, investors, and policy-makers alike.

5. Conclusion

The process of predicting the success of a startup is a complex and challenging task that involves various factors. These factors include general information about the startup, such as its age, its stage of development, and product or service innovation, as well as industry-specific variables like market demand and competition, founder characteristics like experience and leadership skills, and financial metrics such as funding rounds, funding amounts, and number of investors. All of these factors play a crucial role in making accurate predictions.

Although many machine learning algorithms have been found to be effective in predicting startup success, interpretable algorithms were preferred by researchers in this field. Transparent algorithms that can explain the reasoning behind their predictions enable us to understand why the model made a particular prediction. Therefore, in order to learn from the predictions, we choose algorithms that are explainable and interpretable.

Proceedings of M. F. B. G. I. M. - 2022/2023

Future research can focus on developing industry-specific models by incorporating more data sources and features to refine the models. This can help in creating more accurate and customized models that are better suited for each specific industry. In addition, building specific models for different countries can lead to more accurate predictions and customized models that are tailored to the unique characteristics of each country's policy and business environment. For instance, building specific models for Algerian startups can help in better understanding the unique characteristics and challenges of the Algerian startup ecosystem, and in predicting startup success more accurately. This could lead to more accurate predictions and insights, helping entrepreneurs, investors, and policymakers make informed decisions to drive economic growth and innovation in Algeria.

Developing specific classifiers for different countries or business sectors may be challenging, but it can lead to more accurate predictions and customized models that are better suited for each specific industry and region. It is also important to define success and failure in a meaningful way to ensure that the models are measuring what truly matters to stakeholders.

Overall, improving startup success prediction can have a significant impact on driving economic growth and innovation, benefiting not only entrepreneurs and investors but also society as a whole.

References

[1] Aija Vonoga. Start-ups–an element for economic growth and innovativeness. *Latgale National Economy Research*, 1(10):159–167, 2018.

[2] Marco Cantamessa, Valentina Gatteschi, Guido Perboli, and Mariangela Rosano. Startups' roads to failure. *Sustainability*, 10(7):2346, 2018.

[3] Startup Genome. The state of the global startup economy. https://startupgenome.com/article/the-state-of-the-global-startup-economy, 2022. Accessed on 28 February 2023.

[4] Eric Ries. *The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses*. Crown Business, 2011.

[5] Steve Blank. Why companies are not startups. https://steveblank.com/2014/03/04/why-companies-are-not-startups/, March 2014. Accessed on March 18, 2023.

[6] Paul Graham. Startup = growth. http://www.paulgraham.com/growth.html, July 2012. Accessed on March 18, 2023.

[7] Muataz HF AlHazza, Islam Faisal Bourini, MH Zubaidah, and Norsyafira Bt Selamat. Success factor in new product development for startup companies using fuzzy logic approach. In *2019 International Conference on Electrical and Computing Technologies and Applications (ICECTA)*, pages 1–5. IEEE, 2019.

[8] Aileen Lee. Welcome to the unicorn club, 2015: Learning from billion-dollar companies, Jul 19 2015. Copyright - Copyright AOL Inc. Jul 19, 2015; Last updated - 2022-11-08.

[9] Yu Qian Ang, Andrew Chia, and Soroush Saghafian. *Using machine learning to demystify startups' funding, post-money valuation, and success*. Springer, 2022.

[10] Greg Ross, Sanjiv Das, Daniel Sciro, and Hussain Raza. Capitalvx: a machine learning model for startup selection and exit prediction. *The Journal of Finance and Data Science*, 7:94–114, 2021.

[11] Cemre Ünal and Ioana Ceasu. A machine learning approach towards startup success prediction.

Proceedings of M. I. E. G. I. M. - 2022/2023

Technical report, IRTG 1792 Discussion Paper, 2019.

[12] Lele Cao, Vilhelm von Ehrenheim, Sebastian Krakowski, Xiaoxue Li, and Alexandra Lutz. Using deep learning to find the next unicorn: A practical synthesis. *arXiv preprint arXiv:2210.14195*, 2022.

[13] Malhar Bangdiwala, Yashvi Mehta, Smrithi Agrawal, and Sunil Ghane. Predicting success rate of startups using machine learning algorithms. In *2022 2nd Asian Conference on Innovation in Technology (ASIANCON)*, pages 1–6. IEEE, 2022.

[14] Javier Arroyo, Francesco Corea, Guillermo Jimenez-Diaz, and Juan A Recio-Garcia. Assessment of machine learning performance for decision support in venture capital investments. *Ieee Access*, 7:124233–124243, 2019.

[15] Dafei Yin, Jing Li, and Gaosheng Wu. Solving the data sparsity problem in predicting the success of the startups with machine learning methods. *arXiv preprint arXiv:2112.07985*, 2021.

[16] Chenchen Pan, Yuan Gao, and Yuzi Luo. Machine learning prediction of companies' business success. *CS229: Machine Learning, Fall 2018, Stanford University, CA*, 2018.

[17] Dominik Dellermann, Nikolaus Lipusch, Philipp Ebel, Karl Michael Popp, and Jan Marco Leimeister. Finding the unicorn: Predicting early stage startup success through a hybrid intelligence method. *arXiv preprint arXiv:2105.03360*, 2021.

[18] Boris Sharchilev, Michael Roizner, Andrey Rumyantsev, Denis Ozornin, Pavel Serdyukov, and Maarten de Rijke. Web-based startup success prediction. In *Proceedings of the 27th ACM international conference on information and knowledge management*, pages 2283–2291, 2018.

[19] Eliganti Ramalakshmi and Sindhuja Reddy Kamidi. Predictions for startups. *International Journal of Engineering & Technology*, 7(3.12), 2018.

[20] Amar Krishna, Ankit Agrawal, and Alok Choudhary. Predicting the outcome of startups: less failure, more success. In *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, pages 798–805. IEEE, 2016.

[21] Mariia Garkavenko, Hamid Mirisaee, Eric Gaussier, Agnès Guerraz, and Cédric Lagnier. Valuation of startups: A machine learning perspective. In *Advances in Information Retrieval: 43rd European Conference on IR Research, ECIR 2021, Virtual Event, March 28–April 1, 2021, Proceedings, Part I 43*, pages 176–189. Springer, 2021.

[22] M Rungi, E Saks, and K Tuisk. Financial and strategic impact of vcs on start-up development: Silicon valley decacorns vs. northern-european experience. In *2016 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, pages 452–456. IEEE, 2016.

[23] Markus Böhm, Jörg Weking, Frank Fortunat, Simon Müller, Isabell Welpe, and Helmut Krcmar. The business model dna: Towards an approach for predicting business model success. 2017.

[24] Marco van Gelderen, Roy Thurik, and Niels Bosma. Success and risk factors in the pre-startup phase. *Small business economics*, 24:365–380, 2005.

[25] Ulrich Kaiser and Johan M Kuhn. The value of publicly available, textual and non-textual information for startup performance prediction. *Journal of Business Venturing Insights*, 14:e00179, 2020.

[26] Shai Shalev-Shwartz and Shai Ben-David. *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.

[27] Sotiris B Kotsiantis, Ioannis Zaharakis, P Pintelas, et al. Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160(1):3–24, 2007.

[28] FY Osisanwo, JET Akinsola, O Awodele, JO Hinmikaiye, O Olakanmi, J Akinjobi, et al. Supervised machine learning algorithms: classification and comparison. *International Journal of Computer Trends and Technology (IJCTT)*, 48(3):128–138, 2017.

Proceedings of M. F. G. I. M. - 2022 / 2023

[29] Wolfgang Fuhl, Gjergji Kasneci, Wolfgang Rosenstiel, and Enkeljda Kasneci. Training decision trees as replacement for convolution layers. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 3882–3889, 2020.

[30] Sasan Karamizadeh, Shahidan M Abdullah, Mehran Halimi, Jafar Shayan, and Mohammad javad Rajabi. Advantage and drawback of support vector machine functionality. In *2014 international conference on computer, communications, and control technology (I4CT)*, pages 63–65. IEEE, 2014.

[31] Maad M Mijwel. Artificial neural networks advantages and disadvantages. *Retrieved from LinkedIn https//www. linkedin. com/pulse/artificial-neuralnet Work*, 2018.

[32] Jen-Yin Yeh and Chi-Hua Chen. A machine learning approach to predict the success of crowdfunding fintech project. *Journal of Enterprise Information Management*, (ahead-of-print), 2020.

[33] Hercules Dalianis and Hercules Dalianis. Evaluation metrics and evaluation. *Clinical text mining: secondary use of electronic patient records*, pages 45–53, 2018.

Proceedings of M. F. E. G. I. M. - 2022/2023