

Predicting Startup Success, a Literature Review

Harjo Baskoro¹, Harjanto Prabowo², Meyliana³, Ford Lumban Gaol⁴

*Doctor of Computer Science Program,
Bina Nusantara University, Jakarta, Indonesia*

[1harjo@telkom.co.id](mailto:harjo@telkom.co.id)

[2harprabowo@binus.edu](mailto:harprabowo@binus.edu)

[3meyliana@binus.edu](mailto:meyliana@binus.edu)

[4fqaol@binus.edu](mailto:fqaol@binus.edu)

Abstract – The development and growth of startups around the world nowadays have become a global phenomenon. Startups have become an essential element of innovation and economic growth in many countries. But literature shows that the failure rate of a startup is around 90%. Therefore it is crucial for investors, financial advisors, and the government to spot the 10% which eventually will generate higher return rates, bring in greater revenue and ensure economic growth. This research aim is to study what are the critical factors of the startup's success that can be used to make a predictive model using a machine-learning algorithm to predict the success of a startup.

Keywords – *startup, new venture, success, performance, prediction, systematic literature review.*

I. INTRODUCTION

The birth and growth of startups in Indonesia are quite good at least in terms of numbers. In 2004 there are 50 startups, 250 in 2015, 400 in 2016, 700 in 2017 and 900 in 2018 [1]. Also good in terms of valuation, at the end of 2019, there are six startups already become a unicorn (value more than one billion US\$).

Startup plays an important role all around the world, including Indonesia. INDEF report (Institute for Development of Economics and Finance) said that in 2018 digital economy in Indonesia contribute Rp 814 Trillion, equal to 5,5% of GDP (gross domestic product). Minister of communication and informatics, Rudiantara said that he hopes in 2020 digital economy value (including startup) could reach Rp 1.830 Trillion, equal to 11% GDP.

On the other side data shows that the failure rate of startups is very high, Cemre Unal said that 90% of startups failed in the first year, less than 40% from the remaining 10% survive in the next 5 years [2]. Markus Bohm said that the failure rate of startups is between 50% and 83% [3]. Sixty percent of startups can not survive during five years since it was founded and 75% of startups that receive funding finally fail [4].

The same phenomenon happens in Indonesia. Lupiyoadi (2004) said that almost 80% of startups in Indonesia failed in the first year, Wirasmita reported that the failure rate of startup and small enterprises reach 78%, the worst statement

comes from Yudi Candra that only 1% of startups in Indonesia succeed and 99% fail [5].

Here is the challenge we are facing, that is how to enhance the success rate of a startup. This is very important for all of the stakeholders: government, investors, and startups themselves.

There are two research questions in this study. The first is what is the critical factors of early-stage startup success in Indonesia. The second is what is the model to predict early-stage startup success. We focus on an early-stage startup because many studies report that this is the toughest stage in the life cycle of a startup when the highest failure rate occurred.

This paper is a systematic literature review to study what has been done from the previous researches. So we can build our study, we also hope that we can fill the research gap that is still an open issue especially for the Indonesian context.

II. OVERVIEW AND METHODOLOGY

There are several things we need to study from the literature. The first is about startup theory, the definition of a startup, how we measure the performance, and how to define its success. These are very important because many authors have their definitions of these terms, we need to know the thinking background and also what is the implication. The second thing is about how to build the predictive model of startup success, what are the variables, how to select the appropriate variables, and which algorithm to use.

This research is a literature review to answer the research questions by studying and reviewing previous research with a similar topic. These will be done in several steps:

- determining searching keyword
- search literature source
- determining criteria and limitation
- data extraction
- data analysis to answer the research question

The keyword or keyword combinations used to search are:

- startup/new venture
- success/survival
- performance
- prediction

The search period of the literature is between the years 1985 – 2020. All search is done using Google Scholar. Table 1 shows the source of the paper gathered.

The collected papers are then reviewed and analyzed and we select whether it is relevant to the aim of this study.

TABLE 1.
DATA EXTRACTION

No.	Literature Database	Studies Found	Studies Selected
1	Conference	17	7
2	Journal	51	10
3	Open Access Document	4	
4	IEEE Transaction	1	1
5	Master's thesis	6	5
6	Book	4	4
		83	27

III. RESULT AND DISCUSSION

This research has intended to explore several important aspects in predicting the success of a startup. Four aspects need to be considered, that is startup definition, success definition, startup success factor, and algorithm used in predicting model. All these four aspects need to be defined at the beginning, this is important because it will determine the next step to be done, for example, startup definition will determine how the survey or interview will be conducted. The startup stage is also very important because different stage are related to different success factor, for example, most early-stage startup does not generate profit or does not have enough financial data [2] so financial criteria do not apply to startup at this stage.

Next, we need to define a success definition, that also will be used as the dependent variable in predicting model. Success definition will determine the question list in the questionnaire.

The third aspect is startup success factors. Here we need to carefully select the factors. Previous research showed that several variables may contradict the other [6]. Some variables may apply only to specific startup stages, for example, entrepreneurial characteristics and spirit are considered more important in the early stage [7], Some variables may apply with a certain prerequisite.

Last is the machine learning algorithm used for predicting the model. Many researchers use several algorithms with different results in terms of accuracy, sensitivity, etc. In this study, we show some of the previous research that later we can use as a reference to start or to compare.

The fifth reference we get from the literature is the predicting model that depicts the relationship between success factors in a graphic format that easy to interpret. This will be

very useful for the related stakeholder to utilize. This study will show five previous research as a reference.

TABLE 2
VARIABLES AND SOURCE

No.	Variables	Source
1	Startup definition	[8], [9], [10], [11], [12], [13]
2	Success definition	[2], [14], [15], [16], [17], [18], [19], [20], [21], [22]
3	Startup success variables	[7], [14], [20], [23], [24]
4	Machine learning algorithm	[2], [3], [14], [17], [25]
5	Predicting Model (graphically)	[26], [27], [28], [29], [30]

A. Startup Definition

Many researchers develop their definition of a startup, depending on their focus and point of view.

TABLE 3
STARTUP DEFINITION

No.	Author	Startup Definition	Source
1	Andreas Levin	A startup is a novel institution, not older than 15 years, of moderate size, with a headcount lower than 250 employees and annual turnover lower than 50 Million Euros, that aims to deliver new products or services to a market while incurring a high level of uncertainty.	[8]
2	Paul Graham	A startup is a company designed to grow fast. Being newly founded does not in itself make a company a startup. Nor is it necessary for a startup to work on technology, or take venture funding, or have some sort of "exit." The only essential thing is growth.	[9]
3	Peter Thiel	startups are about creating new technological innovation and then startups must aim to create a monopoly in a niche market and only then expand to new markets.	[10]
4	Eric Ries	a startup is a "human institution designed to create a new product or service under condition of extreme uncertainty"	[11]
5	Steve Blank	startup as a "temporary organization used to search for a repeatable and scalable business model"	[12]
6	Agnieszka Skala	A startup life cycle is divided into three basic stages: initial, expansion, and maturity. In the	[13]

	initial phase, a startup is an organization with limited resources which identifies a market problem, recognizes demand, or verifies its solution; at the expansion stage, it is an organization that grows rapidly (even at double-digit rates per month); and at the maturity stage it is a hyper scalable organization	
--	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--

In our context, we will use a combination of Ries's and Skala's definition of a startup. This is based on the consideration that we can not deny the uncertainty a startup must face. The second reason is that in our case we choose to analyze startups that are in the initial stage until the expansion.

B. Success Definition

Startup success is a multidimensional concept and has a subjective definition depending on the angle of view of the founder. Some researchers developed their definition of startup success, some in specific terms and some in more general. Also, startup success has a different meaning according to venture capital or investor. Table 4 show some of the startup success definition.

TABLE 4
STARTUP SUCCESS DEFINITION

No.	Startup Success Definition	Source
1.	Startups get total funding above 1 Million Euros, having more than 10 employees or more, having revenue growth of at least 20% in the past three years.	[14]
2.	A startup that is operating, acquired or issued an IPO (initial public offering)	[2]
3.	Get funding from a reputable financial institution, make a strategic alliance with big corporations	[15]
4.	Startup succeeded to get funding	[16]
5.	Startup already have an IPO (initial public offering) or M&A (merger and acquisition)	[17]
6.	Success is measured by its ROI (return on investment) and profitability	[18]
7.	Startup in the first stage of development with a high level of innovation, inherent risk, and scalable business model which operates for four or more years.	[19]
8.	A startup that achieves series A funding	[20] [21]
9.	A startup that exits and returns at least five times the initial funding invested	[22]

In this case, we will consider a startup as success if it receives a follow on funding, no matter how big it is. This is related to our definition of a startup, that we are focused on startups that are in the initial stage until expansion, that need to get funding. Getting funding implies that a startup is considered as success or potentially success by the investor.

C. Startup Success Variables

The most important thing in this study is the variables used to predict the success of a startup. Many researchers have developed their list. We select studies from 11 researchers and get 132 factors that were considered as the basis to build our success variables. Before using this list, it is necessary to filter out to eliminate the overlapping and multi-collinearity factors if any.

As the first step (but not sufficient) we try to clarify the list with three manager-level key people from Indigo Accelerator Program as the assessor. Indigo Accelerator is an accelerator owned by a state-owned Telco company in Indonesia. We use the Indigo program as the source of preliminary study about startups. The result is as follows 70 factors are considered important by 3 assessors, 28 factors are considered important by 2 assessors, 19 factors are considered unimportant by 2 assessors, and finally 15 factors as unimportant by all the assessors. We will use this result as the starting point to build our list. The list is not shown in the paper, but we select some variables to show in table 5. This shows us that each variable is important to be identified individually to get a complete understanding and know the background behind it.

TABLE 5
STARTUP SUCCESS VARIABLES

No.	Important Variables	Explanation	Source
1	number of founders, time dedication, employees incentives, university ranking, team size, time to market, data orientation, pivoting, societal relevance	the first three variables are very straightforward, the second three variables are relatively quantifiable but are slightly more prone to errors, the last three variables are the most conflicting.	[14]
2	experience in establishing a company, management skills, specialized education & skills	at the earliest stage, the entrepreneurial characteristics and spirit of the founder are more important, but as the company develops the role of administrative management skills and experience are getting more important	[7]
3	supply chain integration, market scope, firm age, size of the founding team, financial resources, marketing experience, industry experience, patent protection	except for supply chain integration, most factors may not be controllable, enlarging team may improve performance, the quality of the resources are more important than its quantity.	[23]
4	utilization of social capital	a strong positive relationship between the utilization of social capital and startup survival, more among	[24]

		high tech firms than low tech firms, how firms utilize the available social capital depends on their personality traits	
5	social capital	founder's professional network connections are critical for the success of the startup because of the information, resource, and status benefit they provide, human capital has a lower impact on startup success than social capital	[20]

D. Predicting Model

Several researchers not only define their list of startup success variables but also illustrate the predicting model graphically, based on the variables defined. The models built based on specific variables depend on the focus and interest of the researcher, also depend on the success definition as the independent variable.

Displaying graphically helps us easily see the relationship between the independent variables and the dependent variables as the result. In this literature review, there are 5 models studied, from these 5 models 3 were built based on the structural equation model. The structural equation model is a good statistical method to analyze the relationship between success factors in an event [25].

Figure 1 shows the Limsong model [26]. According to Limsong, success is defined as receiving financial returns. Six factors positively lead to success, demographic profile, personality traits, the competence of entrepreneur, opportunity, resource, and business characteristics.

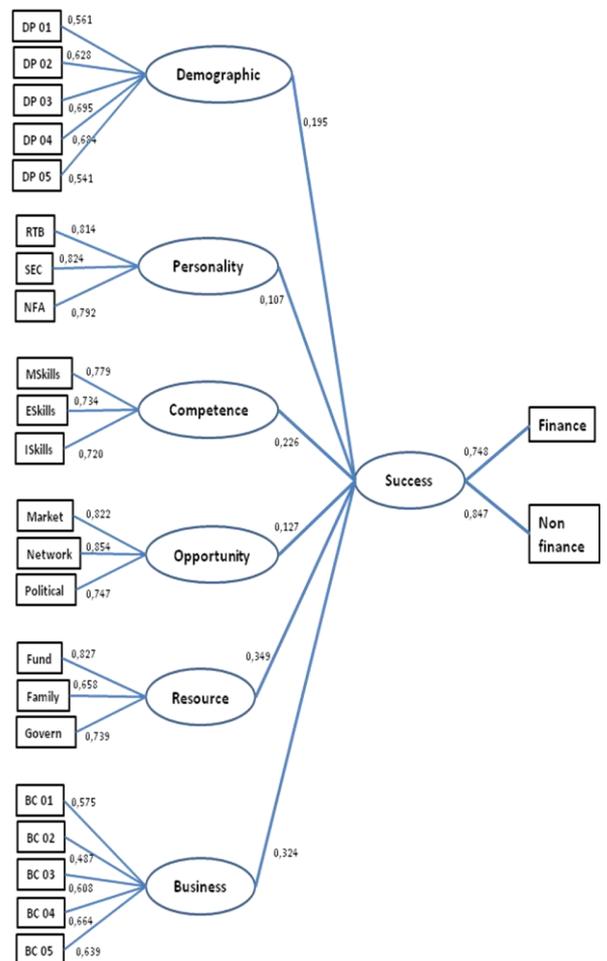


Fig. 1 Limsong model

Figure 2 shows the Nalintippayawong model [25]. Success is defined as a profit of operation and fundraised. Nalintippayawong's model consists of five factors: the potential of a startup, support partner, business model, market opportunity, and customer perspective. The relation between factors are as follows: customer perspectives have a positive effect on market opportunity, market opportunity has a positive effect on a business model and support partner, business model, and support partner have a positive effect on potential startup, while potential startup related with success.

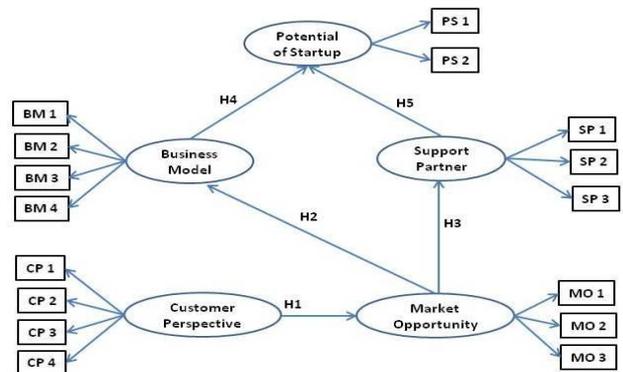


Fig. 2 Nalintippayawong model

Figure 3 shows Tan Le Trinh model [27]. In this model startup performance is considered as the dependent variable, it is affected by two variables including entrepreneurial ecosystem and startup ecosystem. Moreover, these two variables both are affected by five variables that are human capital, financial capital, cultural factors, social factors, and government policy. According to Tan Le Trinh, the startup ecosystem affects startup performance the strongest, while the entrepreneurial ecosystem shows direct relation to performance but also indirect relation to performance via startup ecosystem.

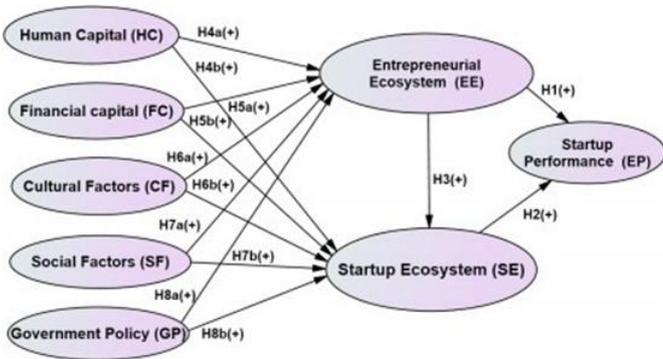


Fig. 3 Tan Le Trinh model

Figure 4 shows the success prediction model proposed by Yankov [28]. Yankov defines a successful startup as one that survived during the last five years and increased its size. In the model, success is affected by four variables that are entrepreneurial team, business strategy, industry structure, and resource. Each of these variables consists of several factors.

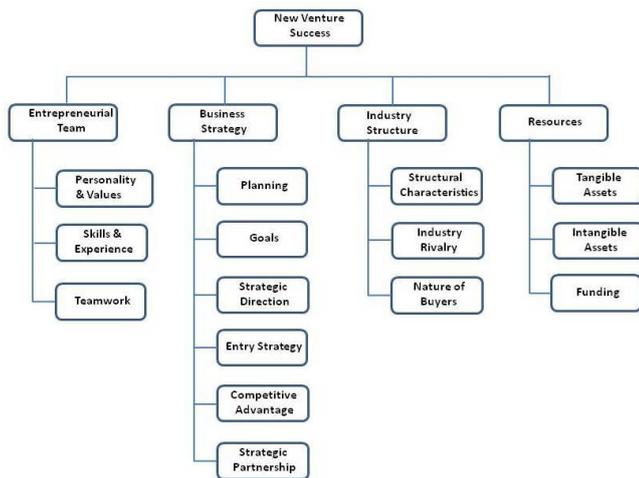


Fig. 4 Yankov model

Figure 5 shows the model proposed by Poon [29]. Poon focuses on three self-concept traits that are an internal locus of control, achievement motive, and generalized self-efficacy, along with entrepreneurial orientation. Firm performance is

considered as the success definition and dependent variable. Poon's study resulted that from three self-concept traits only one factor has direct relation and influence firm performance, that is internal locus of control, while generalized self-efficacy influences firm performance indirectly via entrepreneurial orientation.

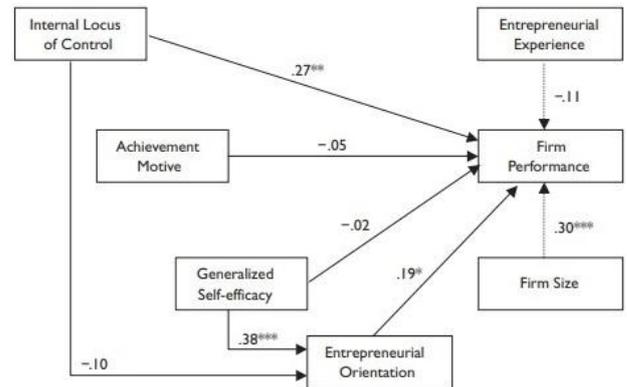


Fig. 5 Poon model

We will define and display our model after we collect the startup dataset and make our model to predict startup success. At the time we hope that we already found the relationship between important factors. We will use many previously defined models displayed here as the reference to make a comparison.

E. Machine Learning Algorithm used in Predicting Model

So far we discussed predicting models that show the relationship between important factors that affect the success of a startup. Further, we also need to evaluate models that were built to predict startup success.

Table 6 shows several predicting models developed by using the machine learning method from previous research. Different algorithms were chosen and tried. Generally, machine learning can be used to predict startup success, each algorithm shows different performance results in terms of accuracy, sensitivity, or area under the curve (not shown in the table).

TABLE 6
MACHINE LEARNING ALGORITHM IN PREDICTING MODEL

No.	Dependent Variable	Algorithm Used	Accu ¹	Sens ²	Src ³
1	Total funding above 1 Million Euro	Logistic regression	71%	17%	[14]
	Having 10 employees or more		71%	75%	
	Annual revenue growth		76%	63%	
2	A startup that	Logistic	77%	80%	[2]

	is operating, acquired or issued an IPO	regression			
		Random forest	94%	98%	
3	The startup already have an IPO or M&A (merger and acquisition)	Logistic regression	92%	92%	[17]
		Support vector machine	92%	92%	
		Random forest	93%	93%	
4	Startup survived	Support vector machine	66%		[3]
	Revenue growth		83%		
5	Not clearly defined	Random forest		83%	[30]
		Logistic regression		84%	

¹. Accu = accuracy

². Sens = sensitivity

³. Src = source

IV. CONCLUSION AND FUTURE WORK

Many factors need to consider in studying the development of a startup and make predictions of its success because different environments imply different consequences and variables. First, we need to define what is the focus and interest (better based on the research gap that still has not resolved yet) and then cautiously and carefully learn what is the issues surrounding the topic.

This paper is a literature review that will be used as a basis to build a prediction model of startup success. After we get enough data from survey, questionnaires, interviews, etc. and we analyze the data by using a statistical method it is very likely (based on literature) that we find some contradictory results, in this case, we need to dive deeper into the subject to know and to make sure what is happened.

Also, we need to realize that understanding the important success factor and making successful predictions is a very different thing. We can build an accurate prediction model based on the machine learning method, but in many cases, we are facing difficulties in interpreting what factors affected the success of such model [31]. So we need to combine it with a statistical model from which we get a model that is easy to interpret and we can directly use the result to enhance the possibility of startup success.

If we can build a robust prediction model (maybe several models for different cases) in the future we can develop a user-friendly platform that can be utilized by all the stakeholders in the startup ecosystem. On the inside, this platform uses the predicting model we build and on the outside this platform provides a user interface for easy access and use.

V. REFERENCES

- [1] Media Indonesia, "Genjot Start Up Mengangkat Ekonomi Masyarakat Bawah," <https://mediaindonesia.com/humaniora/255840/genjot-start-up-mengangkat-ekonomi-masyarakat-bawah>, 2019.
- [2] C. Unal, "Searching for a Unicorn : A Machine Learning Approach Towards Startup Success Prediction," no. Master's Thesis Humboldt Universitat Berlin, 2019.
- [3] M. Böhm, J. Weking, F. Fortunat, S. Müller, and I. Welpé, "The Business Model DNA : Towards an Approach for Predicting Business Model Success," *13th Int. Conf. Wirtschaftsinformatik*, pp. 1006–1020, 2017.
- [4] C. Giardino, S. S. Bajwa, and X. Wang, "Key Challenges in Early-Stage Software Startups," *Int. Conf. Agil. Softw. Dev. 2015*, pp. 52–63, 2015.
- [5] Y. Winosa, "Ini Alasan 90 Persen Startup di Indonesia Gagal.," 2019.
- [6] M. Kakati, "Success criteria in high-tech new ventures," *Technovation J.*, vol. 23, no. 5, pp. 447–457, 2003.
- [7] A. Prohorovs, J. Bistрова, and D. Ten, "Startup Success Factors in the Capital Attraction Stage : Founders' Perspective," *J. East-West Bus.*, pp. 1–26, 2018.
- [8] A. Levin, "Analyzing startup performance: How methods of machine learning and multi-criteria optimization may help to uncover successful startups," p. 81, 2014.
- [9] P. Graham, "Startup = Growth," 2012.
- [10] P. Thiel, *Zero to One: Notes on Start-Ups, Or How to Build the Future*. Crown Publishing, 2014.
- [11] E. Ries, *The Lean Startup*. Crown Business, Crown Publishing Group, 2011.
- [12] S. Blank, *The Four Steps to the Epiphany: Successful Strategies for Startups that Win*. Wiley, 2007.
- [13] A. Skala, *Digital Startups in Transition Economies : Challenges for Management, Entrepreneurship and Education*. Palgrave Macmillan, 2019.
- [14] D. C. Martinez, "Startup Success Prediction in The Dutch Startup Ecosystem," no. Master's thesis Delft University Technology, 2019.
- [15] E. Erdogan and S. Koohborfardhaghghi, *Delivering a Systematic Framework for the Selection and Evaluation of Startups*. Springer International Publishing, 2019.
- [16] B. Sharchilev, M. Roizner, A. Rummyantsev, D. Ozornin, and P. Serdyukov, "Web-based Startup Success Prediction," *CIKM 2018*, pp. 2283–2291, 2018.
- [17] F. R. da S. R. Bento, "Predicting Start-up Success with Machine Learning," *Master's Diss. Univ. Nov. Lisboa*, 2017.
- [18] A. Cirjevskis, "Exploration of qualitative success factors of innovative e-business startups : Blue ocean strategy versus dynamic capabilities Exploration of qualitative success factors of innovative e-business startups : blue ocean strategy versus dynamic capabilities A," no. June, 2018.
- [19] D. Santos da Silva, "Portuguese Startups : a Success Prediction Model," no. Master's Dissertation Universidade Do Porto, 2016.
- [20] O. Spiegel, P. Abbassi, M. P. Zylka, D. Schlagwein, K. Fischbach, and D. Schoder, "Business model development , founders ' social capital and the success of early stage internet start-ups : a mixed-method study," 2015.
- [21] J. A. C. Baum and B. S. Silverman, "Picking winners or building them ? Alliance, intellectual, and human capital as selection criteria in venture financing and performance of biotechnology startups," vol. 19, pp. 411–436, 2004.
- [22] J.-G. Strelitzki and R. Schulte, "Which Venture Capital Selection Criteria Distinguish High-flyer Investments?," *An Int. J. Entrep. Financ.*, vol. 15, no. August, pp. 37–41, 2013.
- [23] M. Song, K. Podoyunitsyna, H. Bij, and J. I. M. Halman, "Success Factors in New Ventures: A Meta Analysis," *J. Prod. Innov. Manag.*, vol. 25, pp. 7–27, 2008.
- [24] C. Bandera and E. Thomas, "The Role of Innovation Ecosystems and Social Capital in Startup Survival," *IEEE Trans. Eng. Manag.*, vol. 66, no. 4, pp. 542–551, 2019.
- [25] S. Nalintippayawong, N. Waiyawatpattarakul, and S. Chotipant,

- “Examining the critical success factors of startup in Thailand using structural equation model,” *Proc. 2018 10th Int. Conf. Inf. Technol. Electr. Eng. Smart Technol. Better Soc. ICITEE 2018*, pp. 388–393, 2018.
- [26] S. H. Srun Limsong, Sok seang, Phou Sambath, “Model of Entrepreneurial Success: Linking Theory and Practice,” no. 1e, pp. 97–103, 2016.
- [27] T. Le Trinh, “Factors affecting startup performance of small and medium-sized enterprises in Danang city,” *Entrep. Bus. Econ. Rev.*, vol. 7, no. 3, pp. 187–203, 2019.
- [28] B. Yankov and P. Ruskov, “Models and Tools for Technology Start-Up Companies Success Analysis,” no. 3, pp. 15–24, 2014.
- [29] J. M. L. Poon, R. A. Ainuddin, and S. H. Junit, “Effects of self-concept traits and entrepreneurial orientation on firm performance,” *Int. Small Bus. J.*, vol. 24, no. 1, pp. 61–82, 2006.
- [30] A. Krishna, A. Agrawal, and A. Choudhary, “Predicting the Outcome of Startups: Less Failure, More Success,” *IEEE Int. Conf. Data Min. Work. ICDMW*, vol. 0, pp. 798–805, 2016.
- [31] M. Stewart, “The Actual Difference Between Statistics and Machine Learning,” 2019.