International Journal of Advanced Culture Technology Vol.12 No.2 167-179 (2024)
DOI https://doi.org/10.17703/IJACT.2024.12.2.167

IJACT 24-6-19

Predictive Model for Evaluating Startup Technology Efficiency: A Data Envelopment Analysis (DEA) Approach Focusing on Companies Selected by TIPS, a Private-led Technology Startup Support Program

[1]Jeongho Kim, [2]Hyunmin Park, [3*]JooHee Oh

[1] Dept. of Industrial Systems Engineering, KAIST, Korea
E-mail: jeonghokim123@gmail.com
[2] Dept. of ICT Convergence, Handong Global University, Korea
E-mail gusalsgmldud@handong.ac.kr
[3] Professor, Dept. of Management and Economics, Handong Global University, Korea*
E-mail: jooheeoh@handong.edu

Abstract

This study addresses the challenge of objectively evaluating the performance of early-stage startups amidst limited information and uncertainty. Focusing on companies selected by TIPS, a leading private sector-driven startup support policy in Korea, the research develops a new indicator to assess technological efficiency. By analyzing various input and output variables collected from Crunchbase and KIND (Korea Investor's Network for Disclosure System) databases, including technology use metrics, patents, and Crunchbase rankings, the study derives technological efficiency for TIPS-selected startups. A prediction model is then developed utilizing machine learning techniques such as Random Forest and boosting (XGBoost) to classify startups into efficiency percentiles (10th, 30th, and 50th). The results indicate that prediction accuracy improves with higher percentiles based on the technical efficiency index, providing valuable insights for evaluating and predicting startup performance in early markets characterized by information scarcity and uncertainty. Future research directions should focus on assessing growth potential and sustainability using the developed classification and prediction models, aiding investors in making data-driven investment decisions and contributing to the development of the early startup ecosystem.

Keywords: *Startup, technical efficiency, machine learning, TIPS, data envelopment analysis, Crunchbase*

1. INTRODUCTION

This study focuses on developing technical efficiency measurement indicators using market data of early-stage startups and proposes a model to classify startups exhibiting high efficiency. Startups, characterized as new and unlisted companies leveraging innovative technology, serve as pivotal contributors to economic

Manuscript received: March 30, 2024 / revised: April 23, 2024 / accepted: May 15, 2024
Corresponding Author: jooheeoh@handong.edu
Tel:+81-54-260-1260
Professor, Dept. of Management and Economics, Handong Global University, South Korea Univ., Korea

new and unlisted companies leveraging innovative technology, serve as pivotal contributors to economic growth, job creation, and technological advancement. Recognizing their significance, governments worldwide are fostering support for startups. In Korea, the TIPS (Tech Incubator Program for Startup Korea) initiative stands out as a private sector-led program offering comprehensive support for startup commercialization and global outreach, thereby providing a stable growth platform for technology startups (Go, Park, Kim, 2022). The Ministry of SMEs and Startups reported in the '2022 Startup Trends' announcement a substantial increase in the number of technology-based startups, peaking at 230,000, attributed primarily to the streamlining of online registration processes. Despite this surge, there remains a deficiency in methodologies for assessing the performance of technology-based startups. Technology-based entrepreneurship (TBE) delineates ventures characterized by heightened levels of innovation and technological acumen (Lee & Joo, 2019). Such startups are founded upon the entrepreneur's competencies, experiences, and expertise (Kim, 2016), embodying a commitment to innovation, growth, and the strategic acquisition of external resources (Spender et al., 2017). This study endeavors to address this lacuna by analyzing data from successful IPOs of Korean firms, aiming to construct technical efficiency indicators tailored to technology-based startups. Through the examination of IPO data, the study aims to gain insights that can enhance the evaluation of TIPS-selected companies. Ultimately, the objective is to furnish stakeholders with robust evaluation metrics conducive to discerning high-performing startups.

Predicting the initial value of a startup is important to investors and government agencies that regulate and support industry growth. According to the '2019 Startup Business Trends' announced by the Ministry of SMEs and Startups, the rate of exit through IPO among startups is about 0.7%, indicating a low indicator of continuous success. As the importance of early indicator development has been emphasized, existing research has analyzed key factors that determine the initial success of startups, such as investment in technology development and financing (Lee, Hwang, Gong, 2017). As an alternative to support this, it is important to develop models that can serve as the basis for data-based investment decisions, as well as efficiency measures and information signals. There is a need to provide opportunities for various investors in the market to discover new business opportunities, perform effective portfolio management, and discover good companies.

This study is focused on developing quantitative evaluation indicators tailored specifically to technology-related startups, with the aim of deriving technology-based metrics. Technology, in this context, encompasses a synthesis of components, parts, or subsystems, constituting all tools that enhance communication and drive innovation to optimize a company's performance (Arthur, 2009). Additionally, technical efficiency (TE) is quantified as the ratio between observed output and maximum output under constant inputs, or alternatively, as the ratio between minimum inputs and observed inputs under constant output conditions (Farrell, 1957). We aim to forecast the initial technology efficiency indicators of startups selected through the TIPS program.

To this end, we developed a model to predict the technical efficiency of TIPS-selected companies with a high probability of successful IPO or merger and acquisition (M&A). Based on the technical efficiency index, it was possible to distinguish companies with relatively high technical efficiency among the TIPS-selected companies. This study is expected to provide the following values to investors, startups, and government agencies. We support efficient investment decision-making for venture capitalists and investors. Entrepreneurs can clearly understand the efficiency of their company's technology and formulate a growth strategy. Government agencies will help establish and evaluate related support policies. The new indicators in this study are expected to provide insight into the development of innovative business models through a multi-faceted approach to startup evaluation.

2. RELATED STUDIES

2.1 Initial startup success factors

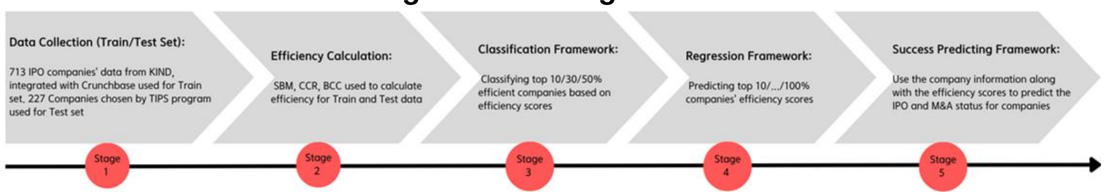
Predicting the initial value of a startup is an important issue for government agencies and investors, so various studies are being conducted. Lee, Hwang, and Gong (2017) derived priorities related to success factors for startups through the Analytic Hierarchy Process (AHP) technique. Among the eight representative success factors for the success of startups' initial market spending, research and development, business management, and marketing were selected as important factors, and the AHP comprehensive analysis showed that initial funding had a preference rate of 51%, which is more than half. You can check it. The most important factor in realizing a start-up is the availability of financing through investment. In the case of technology-based startups, the efficiency and innovativeness of the technology possessed are important factors in determining investment, and technological efficiency is an important indicator that can determine whether or not there will be financing through investment. Startup support policies that help private startups settle in the market can also be a major factor in raising initial funds. Lee (2017) argues that in order to prevent market stagnation due to information asymmetry and risk of failure in the startup market, the government should directly intervene in the startup market and invest budget to foster startups. As such, the importance of start-up support policies is emerging.

A representative example of a private-level startup support policy is TIPS (Tech Incubator Program for Startup Korea). Go, Park, and Kim (2020) conducted an AHP analysis on the determinants of TIPS' investment decision. As a result of the analysis, the importance of the top factors was confirmed in the following order: entrepreneur, market, products and services, finance, and network. Kim, Kim (2023) classified the growth stages of TIPS-supported startups based on quantitative performance criteria. They identified the characteristics of corporate groups at each stage using the start-up's investment period and investment amount. Kim (2022) conducted an analysis of the impact of the TIPS program on entrepreneurship. As a result of in-depth interviews with TIPS start-ups, it was found that the influence was in the order of challenging spirit, innovation, risk taking, and leadership.

2.2 Data Envelopment Analysis (DEA)

Data envelopment analysis (DEA) is based on linear programming and was first presented by Charnes, Copper, and Rhodes (1978) as a statistical method to measure the relative efficiency of decision-making units in similar environments (Go, 2013). There are many papers that measure the productivity and technical efficiency of various companies, industries, and institutions using DEA methodology. Cho and Kwon (2022) analyzed the efficiency and productivity of ICT industry R&D investment, and Lee and Hong (2021) analyzed the efficiency of retail outlets based on the case of fashion companies. Jung and Kim (2006) evaluated how the efficiency and productivity of Korean non-life insurance companies are changing after the IMF. Han (2019) analyzed the efficiency of major domestic Korean food franchise brands. Ha, Han (2019) analyzed the production efficiency of broiler breeder farms by analyzing breeding performance data from 43 domestic broiler breeder farms and suggested directions for improving productivity. Lee and Kim (2011) suggested that a policy to adjust the combination ratio of doctors and nurses to an appropriate level is needed to improve the productivity of emergency medical institutions. Kim (2008) presented various implications by comparing and analyzing the management efficiency of Korea's exchange market before and after merger with overseas exchange markets using a modified DEA model. DEA, which is used in such a variety of fields, differs from existing research in that this study developed a model to generate efficiency indicators, classify them according to the top percentage, and make predictions. The modeling process of this study is presented in Figure 1.

Figure 1. Modeling Process



3. EMPIRICAL ANALYSIS

3.1 Data

For this study, data was collected from Crunchbase and KIND (Korea Investor's Network for Disclosure System), for corporate information acquisition. Crunchbase serves as a repository for comprehensive data on companies, investors, and individuals associated with companies, with a specific focus on company information for the purposes of this study. On the other hand, KIND represents the official conduit through which Korean listed companies furnish regulatory documents and mandatory disclosures mandated by the Financial Services Commission of Korea (FSC). The KIND platform offers investors, analysts, and other stakeholders access to a wide array of corporate information, including financial statements, annual reports, shareholder meeting notices, and various regulatory filings, thus constituting a rich resource for this research endeavor.

(1) Main variables

Table 1. Key Variables Table

Trend.Score.7.Days	Movement in Rank over the last 7 days using a score from -10 to 10	Rank & Scores
Trend.Score.30.Days	Movement in Rank over the last 30 days using a score from -10 to 10	Rank & Scores
Trend.Score.90.Days	Movement in Rank over the last 90 days using a score from -10 to 10	Rank & Scores
Estimated_Revenue_Range_in_mil	Estimated revenue range for organization	Basic Info
Similar.Companies	Total number of organizations similar to the given organization	Similar Companies
SEMrush.Monthly.Visits	Total (non-unique) visits to site for the last month; includes desktop and mobile web.	Web Traffic by SEMrush
SEMrush.Visit.Duration	Average time spent by users on a website, per visit in seconds. Includes both desktop and mobile web.	Web Traffic by SEMrush

SEMrush.Page.Views...Visit	Average number of pages viewed by users in each visit to a site in the last month. Includes both desktop and mobile web.	Web Traffic by SEMrush
SEMrush.Bounce.Rate	Percentage of visitors to site who navigate away after viewing only one page. Includes both desktop and mobile web. Includes both desktop and mobile web	Web Traffic by SEMrush
SEMrush.Global.Traffic.Rank	Traffic rank of site, as compared to all other sites on the web.	Web Traffic by SEMrush
BuiltWith.Active.Tech.Count	Total number of technologies currently in use by this company, as detected by BuiltWith.	Website Tech Stack by BuiltWith
IPqwery.Trademarks.Registered	The number of Trademarks that have been granted to the given company, as detected by IPqwery.	Patents and Trademarks by IPqwery
Number.of.Articles	Number of news articles that reference the Organization	Basic Info
Total.Funding.Amount.Currency..in.USD.	Total amount raised across all funding rounds in USD	Funding
Number.of.Investors	Total number of investment firms and individual investors	Investors
CB.Rank.Company	Algorithmic rank assigned to the top 100,000 most active Companies	Rank & Scores
IPqwery.Patents.Granted	The number of patents that have been granted to the given company, as detected by IPqwery.	Patents and Trademarks by IPqwery
Number.of.Employees	Total number of employees	Team

The main variables selected in this study are as follows. Crunchbase data used data directly related to the company, such as investment round, number of investors, patents, technology, and investment amount. Crunchbase Ranking trend scores for 7, 30, and 90 days can be viewed as dynamically showing a company's performance over time and show ranking changes in "Crunchbase Ranking." “Estimated_Revenue_Range” provides insight into the company’s financial size and performance. “Similar.Companies” is a variable that provides context for understanding the company’s competitive environment. Web traffic statistics provided by SEMrush, including monthly visits, visit duration, page views per visit, bounce rate and global traffic ranking, provide a window into a company's digital footprint and online engagement levels. “IPqwery...Trademarks.Registered” serves as an indicator of how much emphasis a company places on protecting its intellectual property. “Total.Funding.Amount.Currency.in.USD” and “Number.of.Investors” indicate the financial support received by the company and prove its growth potential as reflected from the investors’ perspective. "Number.of.Employees" and "BuiltWith.Active.Tech.Count" are directly related to a company's operational scale and technological capabilities. IPqwery.Patents.Granted represents the company's innovation capabilities, and "CB.Rank.Company" represents the industry's prominence as a whole, making it an important variable in evaluating efficiency.

(2) Input/output variables

Table 2. Survey the literature on efficiency analysis

Researchers	Subject	Input Variables	Output Variables
Lee, Yoon (2021)	83 firms disclosed on the Venture Company Disclosure System (DIVA)	Number of employees, Capital	Startup investment amount, operating profit, net income
Kim, Kang, Park, Yeo (2013)	8 R&D projects from the Ministry of Trade, Industry and Energy	Government contributions, Private sector burden, Number of participating institutions	Patent outcomes, paper outcomes, generated sales
Park, Moon (2010)	342 regional industrial technology development projects	Research costs, Knowledge holdings, Development period	Patents, Papers, Sales, Job creation
Lee, Chung (2014)	National Defense Technology R&D projects	Research costs, Research personnel, Research period	Patents, Papers, Practical applications

Previous studies have commonly employed variable costs such as labor force equivalents and capital as input variables, coupled with technology-related metrics such as patents, academic papers, and sales as output variables. In the context of research and development (R&D) and defense projects, input variables have included research funds and government contributions, while output variables encompass sales and patent publications. In studies pertinent to startups, input variables typically encompass metrics such as the number of employees and capital, directly impacting the production change process. Output variables, on the other hand, often comprise financial indicators such as startup investment amount, operating profit, and net profit.

Building upon this existing research framework, our study selects input and output variables from Table 3 to serve as representative technical efficiency indicators for early-stage startups.

Table 3. Available Input/Output Variables

Input Variable	Output Variable
Number.of.Employees(Crunchbase)	CB.Rank.Company(Crunchbase)
BuiltWith...Active.Tech.Count(Crunchbase)	IPqwery...Patents.Granted(Crunchbase)
Asset(KIND)	ROA(KIND)
Equity(KIND)	EPS(KIND)

The selected input variables encompass the number of employees and technology-related factors, both recognized as influential determinants of technology efficiency. Previous research has underscored the significant impact of human resources on startup performance, with findings indicating that larger team sizes are positively correlated with success and efficiency (Pasayat, Bhowmick, 2020; Aldrich, Auster, 1986). This correlation suggests that larger teams facilitate greater diversity in technologies, ideas, and division of labor, thereby enhancing startup efficiency.

Moreover, a variable representing the number of technologies utilized by a startup is indicative of its technological capabilities. Startups leveraging diverse technologies tend to exhibit higher levels of innovation, adaptability, and efficiency in acquisitions (Greenberg, Guinan, Pablo, Javidan, 2004). Additionally, variables such as Crunchbase (CB) Rank and the number of patents granted to startups are included as calculation variables. CB Rank serves as a metric of a company's prominence, incorporating factors such as community engagement, funding events, and media coverage (Stephan, 2019). A higher CB Rank signifies greater attractiveness to investors, customers, and potential partnerships, thereby augmenting a startup's value. This metric has been utilized in prior research as a gauge of startup success and influence (Kim, San Kim, Sohn, 2020; Schoenberg, 2022).

3.2 Regression analysis

Based on the main variables, the regression equation with efficiency indicators (SBM, CCR, BCC) as dependent variables can be formulated. The independent variables are categorized into popularity-related variables, financial-related variables, and non-financial-related variables. Variables such as 'Trend Score', 'SEMrush', and 'Number of articles' are designated as popularity-related variables, while 'Total Funding Amount' and 'Number of Investors' are classified as financial-related variables. Other variables are categorized as non-financial variables.

The prominence and commerciality of a company are integral to its reputation and merger and acquisition activities (Lee, Geum, 2023). Research has shown that companies with high technological and financial capabilities tend to acquire companies with high popularity. Moreover, companies need the technical expertise to integrate the operations of acquired entities and the financial resources to facilitate the acquisition process. This is particularly pertinent when acquiring highly popular companies, highlighting the significance of technology and finance in such transactions.

Table 4. SBM Linear Regression results

Variable	SBM model	CCR model	BCC model
Trend Score 7 Days	0.0000 (0.9910)	0.0072 (0.2839)	-0.0005 (0.9665)
Trend Score 30 Days	-0.0019 (0.2458)	0.0006 (0.8996)	0.0100 (0.2533)
Trend Score 90 Days	0.0017 (0.1147)	-0.0083 ** (0.0076)	-0.0009 (0.8750)
Similar Companies	-0.0003 (0.4197)	-0.0019 . (0.0936)	-0.0022 (0.2954)
SEMrush Monthly Visit	0.0031 (0.1046)	0.0018 (0.7460)	0.0026 (0.8013)
SEMrush Visit Duration	-0.0030 (0.1933)	-0.0058 (0.3812)	-0.0097 (0.4348)
SEMrush Page Views Visit	0.0002 (0.8239)	0.0021 (0.3886)	0.0043 (0.3441)
SEMrush Bounce Rate	-0.0001 (0.4854)	-0.0001 (0.8433)	-0.0003 (0.7261)
SEMrush Global Traffic Rank	-0.0013 (0.1177)	-0.0048 . (0.0540)	-0.0076 . (0.0993)
IPqwery Trademarks Registered	0.0127 *** (0.0000)	0.0037 (0.6556)	0.0075 (0.6306)
Number of Articles	0.0095 ** (0.0059)	0.0047 (0.6402)	-0.0157 (0.3984)
Total Funding Amount Currency in USD	-0.0016 . (0.0747)	-0.0061 * (0.0172)	-0.0089 . (0.0634
Number of Investors	-0.0012	-0.0037	-0.0077

	(0.6603)	(0.6519)	(0.6129)
Estimated Revenue Range	-0.0037 *** (0.0006)	-0.0172 *** (0.000)	-0.0320 *** (0.000)

P-value Significance: 0 *** , 0.001 ** , 0.01 * , 0.05 . , 0.1 None

In the SBM model, the intercept term demonstrates statistically significant results, exerting a notable influence on the efficiency score. IPqwery Trademarks Registered, Number of Articles, Total Funding Amount, and Estimated Revenue Range were found to significantly impact the efficiency score at given levels of significance. Similarly, the CCR model yields statistically significant results for the intercept term, signifying its substantial impact on efficiency indicators. Furthermore, Trend score 90 days, Similar Companies, SEMrush Global Traffic Rank, Total Funding Amount, and Estimated Revenue Range were identified as influential variables affecting the efficiency score within the CCR model.

In the BCC model, the intercept also emerges as a significant factor, indicating its pivotal role in determining the efficiency score. Notably, SEMrush Global Traffic Rank, Total Funding Amount, and Estimated Revenue Range demonstrated significant results in relation to efficiency scores within the BCC model. Comparing the analysis results across each model, it is evident that Total Funding Amount and Estimated Revenue Range consistently play key roles in determining the efficiency index in all models. It is noteworthy that while variables deemed statistically insignificant may not display a direct impact on the efficiency indicator, this could suggest the presence of non-linear relationships between the independent variables and the efficiency indicator.

4. MACHINE LEARNING MODEL

In this study, data from the KIND database were utilized, comprising information on 713 Korean companies that successfully underwent initial public offerings (IPOs) as the learning dataset. The evaluation dataset comprised 227 TIPS-selected companies deemed to have a high likelihood of successful IPOs or merger and acquisition (M&A) activities. The study aimed to classify highly efficient companies among those selected for TIPS, with technical efficiency categorized as either 0 or 1 based on whether the company fell within the top 50th percentile.

To achieve this objective, a model was developed utilizing a combination of popularity-related, financial-related, and non-financial variables to ascertain their impact on classification outcomes. Initially, non-financial variables were designated as fundamental variables, with popularity-related and financial-related variables subsequently added to assess any resultant performance changes. This comprehensive approach allowed for an exploration of how various factors, including popularity and financial metrics, influenced the classification results.

Table 5. Default variable classification model

Classification Accuracy	Top 50th percentile SBM		Top 50th percentile CCR		Top 50th percentile BCC	
	ACC	F1	ACC	F1	ACC	F1

Random Forest	0.57	0.42	0.46	0.61	0.57	0.71
XGBoosts	0.49	0.45	0.46	0.56	0.57	0.66
Logistic Regression	0.58	0.67	0.46	0.61	0.57	0.71
SVC	0.59	0.68	0.45	0.11	0.56	0.15
AdaBoost	0.51	0.31	0.48	0.11	0.59	0.31
Average	0.54	0.51	0.46	0.38	0.57	0.51

In the evaluation of classification algorithms, five models were assessed: Random Forest, XGBoost, Logistic Regression, Support Vector Classifier, and Adaboost. Initially, the basic variable model achieved an accuracy of 53%. Upon the incorporation of financial and popularity variables, the accuracy of the models improved, reaching 54%, 58%, and ultimately 59.4%. This observation underscores the influence of variables encapsulating a company's popularity and financial information on the classification of efficiency indicators. The incremental accuracy gains suggest that integrating these factors enhances the effectiveness of the classification models in discerning highly efficient companies among TIPS-selected firms.

Table 6. Top 50% Classification model results

Classification Accuracy	Top 50th percentile SBM		Top 50th percentile CCR		Top 50th percentile BCC	
	ACC	F1	ACC	F1	ACC	F1
Random Forest	0.62	0.71	0.62	0.49	0.59	0.56
XGBoosts	0.61	0.67	0.58	0.47	0.55	0.52
Logistic Regression	0.60	0.55	0.66	0.53	0.52	0.37
SVC	0.60	0.47	0.65	0.73	0.61	0.64
AdaBoost	0.60	0.44	0.60	0.67	0.53	0.59
Average	0.61	0.57	0.62	0.58	0.56	0.54

The classification performance notably improved as the classification groups narrowed down to the top 10%, 30%, and 50%. This improvement is attributed to the top 50% encompassing a diverse range of characteristics, thereby accommodating various types of companies. However, when classifying based on the top 10% and top 30%, only the Adaboost and Soft Vector Classifier (SVC) models demonstrated consistent performance across all efficiency indicators. This suggests that these two models exhibit strengths in classification according to more refined criteria, providing valuable insights into the nuances of classification performance based on model selection.

Subsequently, a model was developed to predict efficiency indicators of companies that have achieved high technical efficiency, with companies segmented across the top 10% to 100%. Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) were computed using the Random Forest and XGBoost models. The results obtained from the XGBoost model are presented below.

Table 7. XGBoost Prediction model results

XGBoost MAPE & RMSE	SBM		CCR		BCC	
	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE
Top 10%	0.83	0.54	0.92	0.64	0.81	0.83
Top 20%	0.79	0.38	0.88	0.47	0.82	0.82
Top 30%	1.07	0.32	0.86	0.39	0.73	0.69
Top 40%	1.33	0.27	1.08	0.34	0.75	0.59
Top 50%	2.27	0.25	1.35	0.31	0.89	0.53
Top 60%	4.85	0.22	1.88	0.29	0.95	0.52
Top 70%	9.16	0.20	2.67	0.28	1.03	0.48

Top 80%	12.13	0.19	3.93	0.27	1.26	0.46
Top 90%	17.88	0.18	6.81	0.26	1.61	0.44
Top 100%	25.12	0.17	14.45	0.24	1.99	0.42

The improvement in prediction accuracy for top-performing companies bears significant implications. Firstly, the observed changes in predictive performance across detailed efficiency indicators underscore the substantial impact of the variables utilized in classifying companies with exceptionally high efficiency. This suggests that these variables play a crucial role in accurately predicting startup efficiency, thereby informing investors and stakeholders about the key determinants of startup success.

Secondly, the enhanced accuracy in predicting the efficiency of top-performing companies implies the presence of common characteristics or factors among highly efficient startups. This insight can serve as a valuable guide for investors and startups alike in developing successful strategies. However, the relative difficulty in accurately predicting efficiency values for a broader spectrum of companies highlights the inherent volatility in the performance indicators of various startups. Consequently, there is a pressing need for further research to refine models and accurately predict the efficiency of a wider variety of startups.

Ultimately, this model holds promise in identifying highly efficient startups and forecasting their efficiency levels. Such insights empower investors to make informed investment decisions in startups with high efficiency, while enabling startups to tailor or refine their strategies to enhance efficiency. Through the application of this model, stakeholders can navigate the dynamic landscape of startup investment with greater confidence and precision.

5. CONCLUSION

This study focuses on selecting technological efficiency as a key indicator for evaluating early-stage startups, aiming to develop a model capable of classifying companies with high technical efficiency among those selected for the TIPS program.

To construct the classification model, input and output variables from various companies sourced from Crunchbase and KIND databases were categorized into popularity, financial, and non-financial variables to assess their respective influences. It was found that popularity and financial variables significantly impact the classification model's efficacy. Subsequently, classifications were refined into the top 50%, 30%, and 10%, with the model exhibiting highest accuracy in classifying top companies.

Efficiency indicators of classified companies were then predicted using an XGBoost and RandomForest model, with prediction accuracy improving as companies ascended the efficiency rankings. Specifically, the XGBoost model for the top 10% demonstrated Mean Absolute Percentage Error (MAPE) values of 0.83 for SBM, 0.74 for CCR, and 0.81 for BCC, while the RandomForest model yielded MAPE values of 0.79 for SBM, 0.74 for CCR, and 0.70 for BCC.

This study offers a novel evaluation method that leverages various input and output-based data to objectively and accurately assess startup performance. The proposed indicators, derived from diverse data sources including Crunchbase ranking, funding amount, number of employees, website-based patent information, and initial sales, hold significant value in evaluating the performance of early startups. Given the inherent difficulty in evaluating newly emerging startups due to information gaps, the indicators proposed in this study offer a solution to this challenge, aiding investors and startup founders in measuring and predicting performance accurately. Furthermore, this research can be expanded to encompass models predicting the efficiency of classified top companies, as well as IPO and M&A prediction models.

In conclusion, this study contributes to the understanding and development of the startup ecosystem by mitigating information uncertainty in the early startup market and introducing new indicators for more precise performance evaluation. It is anticipated that this approach will facilitate the evaluation of sustainability and growth potential of early startups, enabling informed investment decisions and fostering a conducive environment for startup development.

REFERENCES

[1] Kim, S., Kim, K. (2020). A Study on the Startup Growth Stage in Korea, *The Korean Society of Business Venturing and Entrepreneurship*, 15(2), 127-135, doi:10.16972/apjbve.15.2.202004.127.

[2] Kim, S., Kang, C., Lee, W. (2016) The Impact of Leading Universities for Start-up Business Program on the Entrepreneurial Self-Efficacy: Focusing on the Mediating Effects of Entrepreneurship and Start-up Capability, *Journal of The Korean Data Analysis Society*, 18(5), 2627-2636, doi:http://dx.doi.org/10.29214/damis.2019.38.3.011.

[3] Go, B., Park, S., Kim, D. (2022). A Study on the Critical Factors Affecting Investment Decision on TIPS, *The Korean Society of Business Venturing and Entrepreneurship*, 17(5), 31-47, doi:10.16972/apjbve.17.5.202210.31.

[4] Spender, J.-C., Corvello, V., Grimaldi, M., and Rippa, P. (2017). Startups and open innovation: a review of the literature, *European Journal of Innovation Management*, 20(1), 4 – 30, doi:10.1108/ejim-12-2015-0131

[5] Bird, B. (1988). Implementing entrepreneurial ideas: In case for intention, *Academy of Management Review*, 13(3), 442-453, doi:10.5465/amr.1988.4306970

[6] Gartner, W. B. (1989). “Who is an entrepreneur?” Is the wrong question, *Entrepreneurship Theory & Practice*, 13(4), 47-68, doi:10.1177/104225878801200401

[7] Lee, M., Joo, Y. (2019). The Effect of Technology-Based Entrepreneurship (TBE) Activities on Firms Growth, *Asia-Pacific Journal of Business Venturing and Entrepreneurship (APJBVE)*, 14(6), 59-76, doi:

[8] Kim, J., Choi, J. Lim, C., Go, H., Lee, Y., (2016). *INNOVATIVE START-UP*, TopBooks.

[9] Lee, H., Hwang, B., Gong, C. (2017). A Study on the Factors that Determine the Initial Success of Start-Up, *The Korean Society of Business Venturing and Entrepreneurship*, 12(1), 1-13, doi: 10.16972/apjbve.17.4.202208.17

[10] Lee, M., Joo, Y. (2019). The Effect of Technology-Based Entrepreneurship (TBE) Activities on Firms Growth, *The Korean Society of Business Venturing*, 14(6), 59-76, doi: 10.16972/apjbve.14.6.201912.59

[11] Lee, H. G. (2017). A study on the determinants of performance in technology-based start-up enterprises: Focused on the effects of the government support policy, Korea University doctoral dissertation, doi:10.23186/korea.000000076804.11009.0000906

[12] Kim, S., Kim, K. (2023). Analysis of the Growth of Startup Ecosystem: Focusing on the TIPS teams, 43(1), 60-79, doi:10.33219/jsmef.2023.43.1.003

[13] Kim, M. (2022). The Effect of Startup Support Policy on Entrepreneurship: Focusing on TIPS Support Projects by the Ministry of SMEs and Startups, *The Korean Society of Business Venturing and Entrepreneurship*, 17(3), 1-17, doi: 10.16972/apjbve.17.3.202206.1

[14] Farrell, M. J. (1957). The measurement of productive efficiency, *Journal of the Royal Statistical Society*, 120, 253-290, doi: https://doi.org/10.2307/2343100

[15] Hu, S., Sun, Y., Zhi, R., Lee, J. (2022). The research on technical efficiency and influence factors of

Chinese cabbage production in China, *Journal of Korea Academia-Industrial cooperation Society*, 23(3), 383-394, doi: 10.5762/KAIS.2022.23.3.383

[16] Charnes, A., Cooper, W.W., Rhodes, E. (1978). Measuring efficiency of decision making units, *European Journal of Operations Research*, 2, 429-444, doi: https://doi.org/10.1016/0377-2217(78)90138-8

[17] Go, C. (2013). The Effect of Managerial Ability on Firm Performance and Tax Avoidance, Hanyang University, doi: http://hanyang.dcollection.net/common/orgView/200000422342

[18] Kim, Y., Ha, H. (2010). Analysis of the productivity changes of the Korean airport industry with DEA-Malmquist productivity index, *Journal of the Aviation Management Society of Korea*, 8(1), 15-28

[19] Kim, J. Y. (2015). Analysis of efficiency and productivity in global arms-producing and military services companies, *International Journal of IT-based Business Strategy Management*, 1(1), doi: 10.5762/KAIS.2015.16.12.8378

[20] Tone, K. (2011). Slacks-Based Measure of Efficiency, In Handbook on Data Envelopment Analysis, 195–209, doi: https://doi.org/10.1016/S0377-2217(99)00407-5

[21] Cho, Y., Kwon, S. (2022). Efficiency and Total Factor Productivity of ICT Industries, Journal of the Korean Data Analysis Society, 24(3), 1083-1097, doi: 10.37727/jkdas.2022.24.3.1083

[22] Lee, E., Hong, J. (2021). The Efficient Analysis Case of Distribution: Focusing on Company A Fashion Retail Stores, *Journal of the Korean Data Analysis Society*, 23(6), 2863-2877, doi: 10.37727/jkdas.2021.23.6.2863

[23] Jung, J., Kim, B. (2006). An Analysis of The Korean Non-Life Insurance Industry's Efficiency by Using DEA Model, *Journal of the Korean Data Analysis Society*, 8(2), 685-696.

[24] Han, Y. (2019). Efficiency Analysis of Korean Restaurant Franchisors Using Data Envelopment Analysis, *Journal of the Korean Data Analysis Society*, 21(1), 335-350, doi: 10.37727/jkdas.2019.21.1.335

[25] Ha, S., Han, Y. (2019). A Study on the Efficiency Factors of Broiler Breeder Farms Using Data Envelopment Analysis, *Journal of the Korean Data Analysis Society*, 6(13), 3141-3153, doi: 10.37727/jkdas.2019.21.6.3141

[26] Lee, K., Kim, J. (2011). Technical Efficiency Measurement and Optimal Combination Ratio of Nurse to Doctor in Emergency Medical Center, *Journal of the Korean Data Analysis Society*, 13(6), 2997-3007.

[27] Kim, B. (2008). A Study on Efficiency Change of the Korea Exchange by using a MDEA Model after Merger, 10(3), 1645-1660.

[28] Pasayat, A. K., Bhowmick, B., & Roy, R. (2020). Factors responsible for the success of a start-up: A meta-analytic approach. *IEEE Transactions on Engineering Management*, 70(1), 342-352, doi: 10.1109/tem.2020.3016613

[29] Aldrich, H., & Auster, E. R. (1986). Even dwarfs started small: Liabilities of age and size and their strategic implications. Research in Organizational Behavior.

[30] Greenberg, D., Guinan, P. J., Pablo, A. L., & Javidan, M. (2004). Mergers and acquisitions in technology-intensive industries: The emergent process of knowledge transfer. *Mergers and Acquisitions, Creating Integrative Knowledge*, 135-155.

[31] Stephan, D. (2019). What is Crunchbase rank and trend score? Crunchbase.

[32] Kim, H.J., San Kim, T., & Sohn, S.Y. (2020). Recommendation of startups as technology cooperation candidates from the perspectives of similarity and potential: A deep learning approach. *Decision Support Systems*, 130, 113229, doi:10.1016/j.dss.2019.113229

[33] Schoenberg, M. (2022). Mapping the EU's landscape of high performing healthcare startups.

[34] Mann, R. J., & Sager, T. W. (2007). Patents, venture capital, and software start-ups. *Research Policy*, 36(2), 193-208, doi:10.1016/j.respol.2006.10.002

[35] Hsu, D. H., & Ziedonis, R. H. (2008). Patents as quality signals for entrepreneurial ventures. *Academy of Management Proceedings*, 2008(1), 1–6, doi:10.5465/ambpp.2008.33653924

[36] Feraud, C., Kim, B., & Kim, U. (2019). Patenting dilemma for startups: number of applied patents, patent imitability, and level of VC funding, 14(3), 169-183, doi:10.16972/apjbve.14.3.201906.169

[37] Arthur, W. B. (2009). The nature of technology: What it is and how it evolves, Simon and Schuster, 10-48, doi:10.5367/ijei.2013.0133

[38] Lee, Y., & Geum, Y. (2023). Identifying patterns of mergers and acquisitions in startups: An empirical

analysis using Crunchbase data, *IEEE Access*, 11, 42463-42472, doi: 10.1109/ACCESS.2023.3270623

[39] Lee, J. H., and Yoon, J. S. (2021). A Study on the Analysis of Management Efficiency of Start-up Investment Companies. *Journal of Digital Convergence*, 19(5), 353-363.

[40] Kim, H. K., Kang, W. J., Park, J. H., Yeo, I. K. (2013). Comparing Efficiencies of R&D Projects Using DEA: Focused on Core Technology Development Project. *Journal of the Society of Korea Industrial and Systems Engineering*, 36(3), 126-132.

[41] Park, J. H., & Moon, J. B. (2010). An efficiency analysis using DEA for national R&D program for regional industrial technology. *Journal of The Korean Industrial Economy*, 23(4), 2047-2068.

[42] Lee, H., & Chung, S. (2014). Study on the Strategies for Promoting Defense International Technology Cooperation: Based on the Korean and Indian Defense Joint Technology Cooperation. *Journal of Korea Technology Innovation Society*, 17(3), 584-603.