science, engineering and health studies

Research Article

Science, Engineering and Health Studies
https://li01.tci-thaijo.org/index.php/sehs
ISSN (Online): 2630-0087

Factor analysis and prediction of startups and ways to exit based on decision tree classification models with adaptive k with SMOTE method for imbalance problem

Wararat Songpan[1] and Ploypailin Kijkasiwat[2]*

[1] Department of Computer Science, College of Computing, Khon Kaen University, Khon Kaen 40002, Thailand
[2] Faculty of Business Administration and Accountancy, Khon Kaen University, Khon Kaen 40002, Thailand

***Corresponding author**:*
Ploypailin Kijkasiwat
ploypailin@kku.ac.th

***Received**: 22 March 2023*
***Revised**: 17 August 2023*
***Accepted**: 5 September 2023*
***Published**: 19 December 2023*

***Citation**:*
Songpan, W., and Kijkasiwat, P. (2023). Factor analysis and prediction of startups and ways to exit based on decision tree classification models with adaptive k with SMOTE method for imbalance problem. Science, Engineering and Health Studies, 17, 23040007.

ABSTRACT

This paper focuses on factor analysis to combine the information of startups with an synthetic minority over-sampling technique (SMOTE) method via an aspect of the decision tree algorithms that assist investors in project screening for describing important features. However, the investment of a startup company has characteristics of imbalanced data. Improvements in the handling of imbalanced data based on the SMOTE method has been developed by sampling from the minority class. The problem is how to set optimized k-nearest neighbors among the most common feature values. This work purposed a method to fit data in the startup's information that is designed to handle the data value by adaptive k with SMOTE, which manages the problem with an imbalance class label for robustness of evaluation metrics for balancing the portion of multi-class. The adaptive k experimental results can solve the k parameter setting and produce a high accuracy rate of startup companies' class as closed, operating, and acquired status of investment at 0.84, 0.87 and 0.97 respectively. The overall accuracy rate is 0.99; that is the best outcome compared with other methods for handling imbalance. In addition, the results and discussion shown that can meet the needs of investment startup are designed and discussed of business views and machine learning views to work co-operation.

Keywords: startup; SMOTE; imbalance class; k-nearest neighbors; decision tree

1. INTRODUCTION

While startup firms may not typically consider exit plans to be a priority, they are a crucial part of planning and obtaining investor funds. To raise funds, startup firms that show the potential for growth may be able to access finance through private investors, such as angel investors, venture capital (VC), and corporate venture capital. Potential investors will consider various factors when making decisions to support startup firms with funds. Considerations include various factors, such as a founder's character, education and experience, domain knowledge, unfair advantage, traction, market response, a supportive ecosystem (Thanapongporn et al., 2021), and a social network as a non-financial resource (Riepe and Uhl, 2020). Potential investors also want to assess whether startups have exit plans which will increase the return on their investment (Gerasymenko and Arthurs, 2014). For this reason, an exit plan is critical for a startup to be successful in the fundraising process. Such exit plans may include acquisition by others or exiting to an initial public offering (IPO), and different factors are associated with each of these.

Acquisition by other companies which have the potential to provide financial support is more popular than IPO. The statistics provide evidence that a high percentage of startups are acquired status by bigger firms (Dibrova, 2015). A study by Bowen et al. (2019) found an approximately 20% decline in IPOs, and a 49% surge in sell-outs. Acquisition by another company may take less time and involve fewer regulations compared with IPO, particularly for non-cross-border acquisitions (Levine et al., 2020).

Many empirical studies indicate that there are important factors that influence the choice to acquire firms. Some of these considerations include the characteristics of the entrepreneur (Lee and Lee, 2015); the role of private investors in the merger and acquisition activities of startups (Dutta and Folta, 2016); and the survival rate and the life cycle of startups (Fisher et al., 2016). Also significant are the supporting ecosystems of startups (Salamzadeh and Kawamorita Kesim, 2017), the degree of innovativeness (Hyytinen et al., 2015), the internal work system (Bendickson et al., 2017), synergic advantage (Lee, 2019), and the duration of investment (Guo et al. 2015). Through the process of acquisition, startup firms can get various benefits that improve their sustainability (Kwon et al., 2018; Rahaman, 2014). While several studies provide information regarding the factors which are important for enabling startup firms to be acquired status by others, there are gaps in the literature that still need to be addressed. The factors that are the most important for operating or acquisition need to be identified. Moreover, there is limited recent research examining the connection between these factors and determining a best practice model for the acquisition process. These gaps in previous research need to be addressed. Also, there is inadequate research on i-positive and negative signals generated by each factor, and how they are associated with the acquisition process in startup firms (Krishna et al., 2016; Ross et al., 2021). In particular, the use of a financial theory, namely signaling theory, is not utilized to describe the relationship of variables. To address this gap, this paper generates decision trees and explores ways to be acquired status as this method is important to the classification process (Malliaris and Malliaris, 2015). Additionally, the study investigates the factors associated with the acquisition, closing, and operating processes of startups. Grounded in signaling theory, the findings of this study can be used to provide insights into general startup behavior and performance.

Signaling is important for both startups and private investors. For fundraising, startup firms can show positive signals to investors regarding their potential for exiting. A high cash burning rate, an increase in customer acquisition costs, investment in research and development, registering for patents, and having network connections can send positive signals to investors to show firms' growth strategies and survival probability (Bernstein et al., 2017; Farre-Mensa et al., 2020; Matricano, 2020; Rompho, 2018; Song et al., 2021). Higher levels of innovation and employment growth indicate the possibility of exiting acquisitions for new and younger organizations. These characteristics indicate that buyers appreciate the growth potential the company signals through its intellectual property rights and employment growth. As such, high-quality, innovative businesses are the most attractive targets for acquisitions (Cotei and Farhat, 2018). Early start-ups rely on signals to demonstrate the changes in their identity that they had to make as they crossed the threshold of the cross-organization lifecycle. An early start-up in an emerging industrial context usually has few good signs to rely on. Correspondingly, there is an advantage for startups in emerging industries to define signaling strategies with public agencies. Therefore, this research aims to make transitions important via the stages of their organization's life cycle. The role of public agencies can prove to be valuable. Startups should win grants from public agencies that are more willing to acquire them and get subsequent VC funding to plan for a successful exit.

From the perspective of private investors, various elements within startups provide information that can signal to them the potential of startup firms, and thereby strongly impact the probability of a funding decision (Ahlers et al., 2015). Financing rounds are a positive signal for potential future investors (Bernstein et al., 2017). Exit plans involve the number of fundraising rounds (Davila et al., 2003), the sources of funding support, types of finance (Bozkaya, and van Pottelsberghe de la Potterie, 2008), investor characteristics, attributes of the founder (Ewens and Townsend, 2020), and the demographic of startups. Young startups with outside investors are more likely to be targeted. This is because angels or VCs have the first chance to liquidate some of their holdings or all of them, when the business becomes an acquisition target from the point of view of the buyer. Experienced startup entrepreneurs are often popular due to their proven ability to realize the growth potential of their venture, as well as their willingness and ability to reap value for themselves. and investors (Cotei and Farhat, 2018).

Machine learning through decision tree models is argued to be a good method for exit prediction. The structure of the decision tree is like a leaf tree for prediction data (Gong et al., 2018). The most informed attributes that can be computed are selected for the best decision tree. Random forest (RF) is a classification model of supervised learning based on a combination of patterns. It is the same as the bagging method. Random forests are like the bagging method (Hong et al., 2018), but the random forests have diversity of the model randomly and the attributes are considered from various features of the sample data. The process of random forests is gathered decision tree prediction with effective aggregation and bootstrap to use binary class and robustness multi-class classification problems (Eghbali and Montazer, 2017). In the aspect of performance improvement, the accuracy of random forests classification results from predicting the ensemble method of the tree. After generating various trees, the majority genre of the tree is selected to the best tree. This process is known as random forest which is also one of the most authoritative machine learning algorithms. It combines a variety of random decision trees and predictions with mean values.

The contribution of startups and ways to exit prediction has challenges. The model depends on data preprocessing and data handling; the data needs to handle imbalanced data in real business to reduce overfitting and underfitting problems before running the machine learning model. Therefore, our proposed model uses the oversampling of minority class using the synthetic minority oversampling technique based on SMOTE (Chawla et al., 2002). The method is generated for the minority class and equal to the majority class. However, SMOTE has some problem in the number of instances uses for an applied model. The purposed method enhances adaptive k with SMOTE; the setting of parameter k is very important parameter to use in SMOTE methodology. Therefore, the method could find out the best k that can be used for the highest accuracy and the number of instances needed to fit the data before applying the machine learning algorithm.

2. MATERIALS AND METHODS

2.1 SMOTE method

The methodology focuses on the synthetic minority oversampling technique (SMOTE) sampling as one approach to deal with unbalanced data sets. There are two main approaches in this method which are under-sampling and oversampling data. Oversampling is generated more than under-sampling from a minority class because under-sampling tends to remove instances that could be carrying some important instances. The SMOTE algorithm to help with the overfitting problem uses random oversampling techniques. The feature space generates new instances from positive distancing of instances. The development of SMOTE called ADSYN (Haibo et al., 2008) aims to oversample the minority class by generating instances to different density distribution. The hybridization which combines under-sampling and oversampling techniques is called the TomekLink. Wang et al. (2019) focused on overlapping in the same feature space. The TomekLink is observed from the links which are removed and increases the class near the decision boundaries. Therefore, the techniques are applied oversampling minority class, and both remove the class observation the criterion of the TomekLinks. The kind of hybrid methods where SMOTE and ENN (Batista et al., 2004). ENN is used to generate under-sampling technique and estimate the nearest neighbors of each of the majority class. The distance of nearest neighbors which is misclassified is removed. This technique with oversampled data combined by SMOTE helps data cleaning. The SMOTE method is developed to combine oversampling and under-sampling to generate the concept of nearest neighbors from the common feature space precisely.

2.2 Our purposed model

The material is used as the computer has processor Core (TM) i5-1135G7@2.40GHz 2.42 GHz and RAM 16 GB to run python language on Jupyter notebook tools. The data is available from Kaggle website via Crunchbase (Andy, 2020) which provides the information about startup companies founded from 1977 to 2014. The original source has 39 features from the database which were used as the literature review including factors for the analysis of 21 factors in our purposed model. Following the study of Li (2020), Arroyo et al. (2019), and Thirupathi et al. (2021), all features used in the current paper are important for analyzing if startup companies will be acquired, closed or operating status. This feature analysis is collected for generating and finding out the best model. There are two points of investment of startups that are a challenge. First, the problem is various significant attributes that are in the history of the investment company which there are difficult for analysis to find out and summarise. Second, the investment of data has characteristics of imbalanced data which means the startup data has 3 classes: operating, closed, and acquired status. Of these, in the overall startup companies operating status is highest, the then acquired and closed status. The data collection is cleaned without null value that gives overall complete information of the startup company to 22,362 company's profile. However, the imbalanced data is very interesting in the field of machine learning to solve the characteristic of these data.

Moreover, the startup's attributes are involved even if there are a lot of features to analyse which are important for the success or failure of startup companies. The key factors go through the decision tree models that are analysed in aspects of business keys and described as: 1) market indicates the sector that a startup firm operates in; market conditions influence the exit strategy, as venture capital values startups higher in sectors with greater product differentiation and faster growth (Sathaworawong, et al., 2019). 2) funding total in US dollars is the total funding raised by the startup firm (Ahluwalia and Kassicieh, 2022). The total funding amount reflects the financial resources used by the startup. Adequate funding is a significant factor in the success of entrepreneurial firms. 3) country indicates the country where a startup firm is currently settling that presents the trends toward start-up exits. 4) city refers to the city where a startup firm is currently settling (Pisoni and Onetti, 2018). 5) funding rounds indicates the number of rounds that a startup firm raises fund. 6) number founded year identifies the number of years that a startup firm has been operating. 7) interval fund indicates the period between each fundraising round. 8) seed represents the initial fundraising from the private sector for a startup (Venugopal and Yerramilli, 2022). 9) venture indicates the amount of funds supported by VC. VC supports startup firms with funding and seek for exit through Initial Public Offering (IPO) or acquisitions (Miyamoto et al., 2022). 10) equity crowdfunding is the amount of fund raising that occurs when a startup firm issues its shares to crowd investors through equity crowdfunding platforms in different countries (Schwienbacher, 2019). 11) Undisclosed indicates the amount of funds raised by anonymous private investors. These involve families and friends. 12) a convertible note is used by startup firms to secure financial support from private investors during early-stage financing. It allows investors converting debt into equity and serves as a temporary solution to potential funding issues that could lead to discontinuation of the firm (Kolev and Schwartz, 2017). 13) debt financing indicates the amount fund startups raise by acting as debtors while private investors act as borrowers (Rahaman, 2014). 14) angle indicates the amount of funds supporting a startup firm by angel investors. Angel investors provide funding to startup firms. They are successful individuals who invest their own funds in promising business projects for potential returns (Rao and Kumar, 2016). 15) grant indicates the amount of financial support as a grant from both private or public sectors. 16) private equity indicates the amount of funds f raised from private equity. Private equity is often referred to as buyout funds, LBO funds (Cumming et al., 2007). 17) post IPO equity indicates the amount of funds raised after the company has already gone public. 18) post IPO debt indicated the amount of fund when corporates loan a company money after the company has already gone public. 19) secondary market indicates the amount of funds raised in a secondary market. 20) product crowdfunding indicates the amount of fund raising. Startups raise funds by issuing shares to crowd investors through reward crowdfunding platforms in different countries. 21) round A, B, C, D, E, F, G and H indicates the round of fund raising starting after the seed round. After the seed round, subsequent rounds of fundraising begin. In Round A funding, investors seek promising ideas with a strong strategy that can be turned into a profitable business. Round B funding is common for development-stage startups, where investors support market expansion. In Round C funding, startups seek additional funds to develop new products, enter new markets, or acquire other companies. If more funding is needed, startups can continue raising funds in subsequent rounds (Rahaman, 2014). The target for a purposed mode for successful startups and ways to exit prediction are 3 class labels: acquired status (class 0), closed status (class 1) and operating (class 2) respectively. Therefore, adaptive k with SMOTE algorithm can set the number of k from k-nearest neighbor to reach the number of minority classes to balance data. The purposed method for SMOTE is to test the set of k using the accuracy rate value between 0 and 1. They also adjust the number k of SMOTE algorithm within the algorithm. There are steps of adaptive k with SMOTE model algorithm as follows:

Adaptive k with SMOTE model algorithm (Purposed model):

The total of dataset as N, the total number of majority M^- and minority M^+ are set respectively.

1. The iteration starts by first selecting a majority class instance randomly.

2. The grid search space of k setting for generating instances is obtained.

3. For every sample x_i is a set of minority M+ are processed by k-nearest neighbor are obtained using Euclidean distance, and calculate the ratio $TpClass_i$ is calculated as $TPClass_i$ of k <= $\Delta(\sum_{i=1}^{Class} TPClass_i)$ / $\sum_{k=1}^{Kn} TPClass_i)$ TPclass , where the TPclass are calculated by the model selection with generating new instance.

4. Thereafter, the total synthetic samples for each x_i will be minimum difference of $TPClass_i$.

5. The number of k instances is adapted to interpolate new synthetic instances M- and M+.

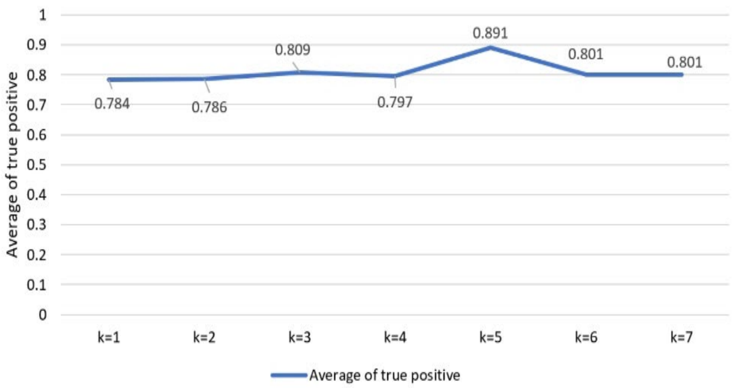
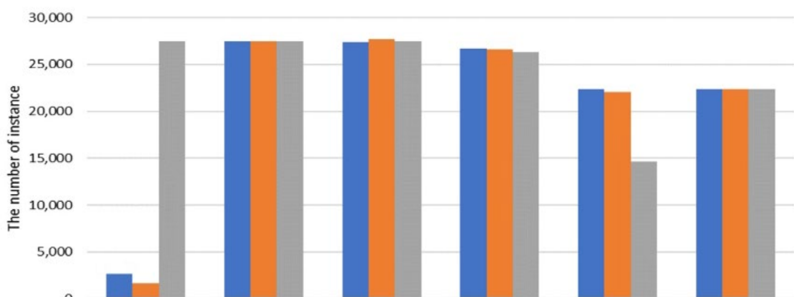


Figure 1. The adaptive k with SMOTE by average of true positive

As Figure 1, the identify k of adaptive k-SMOTE has characteristic of generating to the best evaluating metric focuses true positive class with model training before choose k setting to test data by the true positive (TP) of each class following the algorithm. The adjusted k setting is iteration between 1 and 7 which given k=5 is the best setting to SMOTE algorithm for imbalanced data handling for average of 3 classes: class 0 = acquired status, class 1 = closed status, and class 2 = Operation. Therefore, the adaptive k as 5 is suggested the parameter choosing to train-test model in the next step.



	Actual	SMOTE	ADASYN	SMOTE+Tomek Links	SMOTE+ENN	Adaptive K with SMOTE
Class 0	2,660	27,453	27,392	26,656	22,330	22,362
Class 1	1,692	27,453	27,684	26,595	22,088	22,362
Class 2	27,453	27,453	27,453	26,322	14,650	22,362

Figure 2. The comparison of the number of instances using handling class imbalance methods

Figure 2 shows the comparison of the number of instances using handling class imbalance methods in comparison with actual. The actual dataset as class 2 is majority class and other handling class imbalance methods generated the number of instances is the majority class. The SMOTE, ADASYN, TomekLink are slightly different to generate under sampling of actual class 2. The actual class 2 has 27,453 instances and traditional SMOTE, ADSYN is generated the stable majority class 2 as 27,453 instances and SMOTE and TomekLinks generate majority class 2656 in actual class 0. SMOTE and ENN generate slightly decrease from SMOTE and TomekLink as 22,330 instances. The Adaptive algorithm is generated early iterative by fitting adaptive k=5 with SMOTE method into 22,362 instances for each class, which comparison with actual every class to generate increase instance to 33% from actual dataset.

After the Adaptive k with SMOTE algorithm, the data will be cleaned by the LabelEncoder method, which is a method in data preprocessing that replaces the data value in features using a unique label such as market, city, country, and status need to transform the data because it was exactly calculated to the machine learning model. Therefore, the procedure for preparing the data, this experiment used a startup dataset derived from the investment startup resource. The Machine Learning step handles train-test splits with the number of trainings set as 80% and the number of testing data set as 20% in the decision tree classification models. The experimental results discuss in the results and discussion part.

3. RESULTS AND DISCUSSION

The experimental results show four different models that were included in the test decision tree with the SMOTE method dataset and adaptive k enhancing with SMOTE method. Each model is detailed which describes the different decision tree classifications. The evaluation metrics are calculated from the confusion matrix is the matrix to analysis the insight solving for each class (Tang et al., 2019). As Table 1, the evaluation model described the decision tree algorithm. There are TP for each class between [0, 1]. The high value divergence into 1 given the high accuracy. The highest TP class 0, 1, 2 that was acquired, closed and operating using random forest and SMOTE method was 0.89, 0.90 and 0.91 respectively. And F1-score and Accuracy given highest to 0.90 based on random forest decision tree and SMOTE. Compared with random forest, the decision found that only TP class 2 was operating given high TP to 0.93, however, the other TP class rate, for example, TP class 1 as 0.00 to be used without SMOTE method. The startup's data has the greatest number of operating status and closed and acquired status was important and target class needed analysis. The startup's company was less than decision tree learning. The Decision tree with SMOTE method also had good performance. The TP class 0, 1 and 2 given balance the TP rate classes. Therefore, the processing of imbalanced data will be handled in this experiment.

Table 1. Result of evaluation models

Decision tree classification with imbalance handling methods	**Evaluation metric**					
	TP class 0 (acquired status)	**TP class 1 (closed status)**	**TP class 2 (operating)**	**Average precision**	**Average F1-score**	**Overall accuracy**
DT	0.14	0.02	0.96	0.43	0.38	0.84
RF	0.00	0.00	**1.00**	0.62	0.31	0.86
DT+SMOTE	0.30	0.16	0.97	0.65	0.52	0.87
RF+SMOTE	0.10	0.04	0.99	0.68	0.39	0.87
DT+ADSYN	0.12	0.05	0.95	0.41	0.38	0.83
RF+ADASYN	0.02	0.00	**1.00**	0.49	0.32	0.86
DT+SMOTETomekLink	0.16	0.06	0.94	0.42	0.39	0.83
RF+SMOTETomekLink	0.07	0.01	0.99	0.53	0.36	0.86
DT+SMOTEENN	0.32	0.11	0.87	0.42	0.43	0.78
RF+SMOTEENN	0.26	0.03	0.94	0.48	0.42	0.84
DT+ Adaptive k with SMOTE	0.79	0.84	0.87	0.82	0.83	0.76
RF+ Adaptive k with SMOTE	**0.84**	**0.87**	0.97	**0.89**	**0.90**	**0.90**

Note: **bold** is the best evaluation metric.

As seen in Figure 3, the radar metric comparison between decision tree method TP rate for each class had a calculated value between 0 and 1. That is the number of predicted classes corresponding of the actual class in each sample set divided by the total number in all classes. The first line radar graph shows the visualization of all evaluation metrics. The comparison between decision tree and the secondly line radar graph shows with all evaluation metrics. The expected results should the value of evaluation metric is closed by 1 mean the line of metric looks outer line closed outer boundary. The decision tree and random forest applied adaptive k with SMOTE radar graph are given the best solution and closed to TP as 1 definitely.

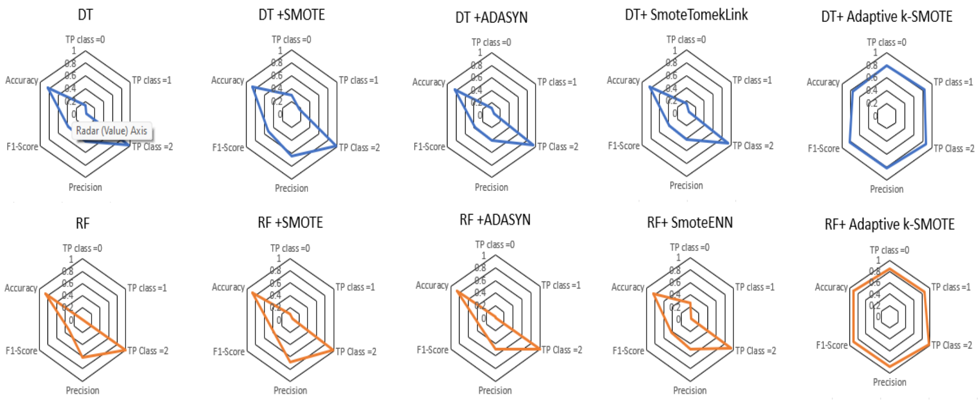


Figure 3. Radar metrics of the comparison between decision tree classification with imbalance handling methods

Table 2. Summary of important features for startup companies results from purposed model

Importance features for each class		
Class 0: Acquired status	**Class 1: Close status**	**Class 2: Operating**
1. Funding total USD	1. Funding total USD	1. Funding total USD
2. Market	2. Number founded year	2. Market
3. Number founded year	3. Country	3. Country
4. Country		4. City
5. Funding round		5. Number founded year
6. Dept financing		6. Funding round
7. Venture		7. Venture

Note: **bold** is the difference importance feature between classes.

For discussion, the advantage of the best decision tree model classification gives the insight feature importance for each class. The class label was divided into 3 classes. Firstly, class 0 was acquired meaning the startup acquisition is a process in a big company by a startup that has gained control over it by purchasing all that company's shares or assets. As Table 2, summary of important features for startup companies results from purposed model. for class 0: acquired were funding total USD, market, country, city, the number founded year, funding round and venture. The difference importance feature based on class 0: acquired and class 1: closed company is debt financing. The dept financing is the highlight of acquired. The class 1: closed company given the importance features only 3 features as funding total USD, number of founded year and country, The startup company which are exiting need to consider three main factors. The first is the startup company was funding the company that will be started. The business model concerns the number of founding especially during the first year, the 7th, 8th and 14th years by experience of company's profile to learn in decision tree. Moreover, the startup company should study the country of the business where location suited for their company.

From a financial perspective, total funding is a crucial factor that influences the exit strategies of startup firms, determining whether they can continue operations, get acquired, or close status. Signal theory suggests that the amount of funding raised by firms can signal the sector in which they operate. A substantial funding amount signals to investors that the startup operates in high-tech sectors, using advanced technology and innovation. Such high-tech firms are often found in countries with supportive regulations and innovative ecosystems (Rungi et al., 2016). Additionally, the location of startup firms plays a role. Entrepreneurial ecosystems, as described by Ahluwalia and Kassicieh (2022), contribute to high growth rates in a region. For instance, Silicon Valley, being the world's largest cluster, attracts abundant venture capital due to its access to capital. The presence of numerous startup firms in such regions signals to investors that these firms have been running for years and have the ability to raise significant funding in each round. All these factors combine to influence the likelihood of startup firms being acquired.

The amount of funding raised by a startup firm can indicate the sector in which it operates. Additionally, the market can provide insights into how long the startup has been running. A startup operating in a particular sector for an extended period signals that the country has a favorable environment for sustained operations. Startups located in countries with supportive ecosystems are more likely to raise funds from the private sector, especially during the early rounds of financing (Zhang, 2011). For startups aiming to achieve the acquired status, they should reconsider their financial support options. Debt finance can be a viable second option, as larger startup firms tend to utilize more business debt, while home-based and growth firms rely more on personal sources of debt (Coleman et al., 2016). High-growth startup firms show smaller increases in debt structure (Flannery, 1994). Startup firms with a less burdensome debt finance structure send a positive signal to venture capital investors, indicating a higher likelihood of continued operations without significant financial burdens.

Regarding closed status, while funding raised is undeniably important for startup firms, insufficient funding for various operational purposes can significantly impact the firm's longevity, leading to financial distress and closure (Megginson et al., 2019). Moreover, incorrect valuation numbers that overvalue a startup can misrepresent its true worth, resulting in inadequate funding to execute business plans (Mustafa, 2021). This can have adverse effects on the country's entrepreneurial environment and ecosystem, sending negative signals to investors and potentially leading to a halt in operations. Proper valuation and securing sufficient funding are crucial to sustain and grow a startup firm successfully.

4. CONCLUSION

This research study proposed handling the imbalance class based on SMOTE to provide the decision tree classification through company profiles to enhance and analyse features of concern for startup and exit status. Startup way should be acquired status and operating and exit way mean startup is closed status. The random forest and classic decision tree with adaptive k-SMOTE as the purposed model gave good results and handled the imbalanced data well compared with other methods. The efficiency of classification and prediction obtained a high TP rate for each classed and average of precision, F1-score and accuracy to reduce the overfitting and underfitting that to appear in unseen of prediction result. The evaluation matrix showed multi-dimension either TP rate of classes as status of acquired, closed and operating status. In order to clearly point out, the classes of acquired startup company have pattern or feature familiar with the operating company. However, the closed company has guidance from our proposed model by feature analysis. In addition, this study encouraged the advantage of decision graph tree to obtain the important feature respectively. The conclusion of random forest tree with adaptive k-SMOTE method given the best TP rate in the aspects of balance of evaluation metrics, and also average of precision, F1-score, and accuracy. Our method shown that the evaluation metrics that analyzed both directly metric by the high performance and feature in decision tree graph. The random forest decision tree shows importance features concerning and guiding the startup and exit effectively. This study presents factors impacting the likelihood of startups' exit strategies in terms of acquired status, operating, or closed status. Additionally, the study adopts signal theory to explain the links between each factor to provide a clearer picture of how the factors are associated. In terms of industrial contribution, startup firms can use the results of this study to plan the exit strategies. Furthermore, entrepreneurs can understand signal transference between each factor, and use that to attract investors for exit plans.

REFERENCES

Ahlers, G. K. C., Cumming, D., Günther, C., and Schweizer, D. (2015). Signaling in equity crowdfunding. *Entrepreneurship Theory and Practice*, 39(4), 955–980.

Ahluwalia, S., and Kassicieh, S. (2022). Effect of financial clusters on startup mergers and acquisitions. *International Journal of Financial Studies*, 10(1), 1–13.

Andy, M. (2020). *StartUp investments (Crunchbase)*. [Online URL: https://www.kaggle.com/datasets/arindam235/startup-investments-crunchbase] accessed on April 14, 2022.

Arroyo, J., Corea, F., Jimenez-Diaz, G., and Recio-Garcia, J. A. (2019). Assessment of machine learning performance for decision support in venture capital investments. *IEEE Access*, 7, 124233–124243.

Batista, G., Prati, R. C., and Monard. M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. *ACM Sigkdd Explorations Newsletter*, 6(1), 20–29.

Bendickson, J. S., Muldoon, J., Liguori, E. W., and Midgett, C. (2017). High performance work systems: A necessity for startups. *Journal of Small Business Strategy*, 27(2), 1–12.

Bernstein, S., Korteweg, A., and Laws, K. (2017). Attracting early-stage investors: Evidence from a randomized field experiment. *The Journal of Finance*, 72(2), 509–538.

Bowen, D. E., Frésard, L., and Hoberg, G. (2019). Technological disruptiveness and the evolution of IPOs and sell-outs. In *Swiss Finance Institute Research Paper Series No. 19–22*. Zürich, Switzerland: Swiss Finance Institute.

Bozkaya, A., and van Pottelsberghe de la Potterie, B. (2008). Who funds technology-based small firms? Evidence from Belgium. *Economics of Innovation and New Technology*, 17(1–2), 97–122.

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16(1), 321–357.

Coleman, S., Cotei, C., and Farhat, J. (2016). The debt-equity financing decisions of U.S. startup firms. *Journal of Economics and Finance*, 40(1), 105–126.

Cotei, C., and Farhat, J. (2018). The M&A exit outcomes of new, young firms. *Small Business Economics*, 50(3), 545–567.

Cumming, D., Siegel, D. S., and Wright, M. (2007). Private equity, leveraged buyouts and governance. *Journal of Corporate Finance*, 13(4), 439–460.

Davila, A., Foster, G., and Gupta, M. (2003). Venture capital financing and the growth of startup firms. *Journal of Business Venturing*, 18(6), 689–708.

Dibrova, A. (2015). Business angel investments: Risks and opportunities. *Procedia - Social and Behavioral Sciences*, 207, 280–289.

Dutta, S., and Folta, T. B. (2016). A comparison of the effect of angels and venture capitalists on innovation and value creation. *Journal of Business Venturing*, 31(1), 39–54.

Eghbali, N., and Montazer, G. A. (2017). Improving multiclass classification using neighborhood search in error correcting output codes. *Pattern Recognition Letters*, 100, 74–82.

Ewens, M., and Townsend, R. R. (2020). Are early stage investors biased against women? *Journal of Financial Economics*, 135(3), 653–677.

Farre-Mensa, J., Hegde, D., and Ljungqvist, A. (2020). What is a patent worth? Evidence from the U.S. patent "lottery." *The Journal of Finance*, 75(2), 639–682.

Fisher, G., Kotha, S., and Lahiri, A. (2016). Changing with the Times: An integrated view of identity, legitimacy, and new venture life cycles. *Academy of Management Review,* 41(3), 383–409.

Flannery, M. J. (1994). Debt maturity and the deadweight cost of leverage: Optimally financing banking firms. *The American Economic Review*, 84(1), 320–331.

Gerasymenko, V., and Arthurs, J. D. (2014). New insights into venture capitalists' activity: IPO and time-to-exit forecast as antecedents of their post-investment involvement. *Journal of Business Venturing*, 29(3), 405–420.

Gong, H., Sun, Y., Shu, X., and Huang, B. (2018). Use of random forests regression for predicting IRI of asphalt pavements. *Construction and Building Materials*, 189, 890–897.

Guo, B., Lou, Y., and Pérez-Castrillo, D. (2015). Investment, duration, and exit strategies for corporate and independent venture capital-backed start-ups. *Journal of Economics & Management Strategy*, 24(2), 415–455.

Haibo, H., Bai, Y., Garcia, E. A., and Li, S. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In *Proceeding of the IEEE International Joint Conference on Neural Networks* (IEEE World Congress on Computational Intelligence), pp. 1322-1328. Hong Kong.

Hong, H., Liu, J., Bui, D. T., Pradhan, B., Acharya, T. D., Pham, B. T., Zhu, A.-X., Chen, W., and Ahmad, B. B. (2018). Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China). *Catena*, 163, 399–413.

Hyytinen, A., Pajarinen, M., and Rouvinen, P. (2015). Does innovativeness reduce startup survival rates? *Journal of Business Venturing*, 30(4), 564–581.

Kolev, J., and Schwartz, E. (2017). To price or not to price? Evaluating convertible-note startup financing. *Academy of Management Proceedings*, 2017(1), 10908.

Krishna, A., Agrawal, A., and Choudhary, A. (2016). Predicting the outcome of startups: Less failure, more. In *Proceeding of the IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, pp. 798–805. Barcelona, Spain.

Kwon, O., Lim, S., and Lee, D. H. (2018). Acquiring startups in the energy sector: A study of firm value and environmental policy. *Business Strategy and the Environment*, 27(8), 1376–1384.

Lee, S. M., and Lee, B. (2015). Entrepreneur characteristics and the success of venture exit: An analysis of single-founder start-ups in the U.S. *International Entrepreneurship and Management Journal*, 11(4), 891–905.

Lee, Y. W. (2019). Synergistic co-operations in the cosmetic industry: Learning and convergence between firms and social media. *Kritika Kultura*, 32, 237–259.

Levine, R., Lin, C., and Shen, B. (2020). Cross-border acquisitions: Do labor regulations affect acquirer returns? *Journal of International Business Studies*, 51(2), 194–217.

Li, J. (2020). Prediction of the success of startup companies based on support vector machine and random forest. In *Proceeding of the 2nd International Workshop on Artificial Intelligence and Education*, pp. 5–11. Montreal, QC, Canada.

Malliaris, A. G., and Malliaris, M. (2015). What drives gold returns? A decision tree analysis. *Finance Research Letters*, 13, 45–53.

Matricano, D. (2020). The effect of R&D investments, highly skilled employees, and patents on the performance of Italian innovative startups. *Technology Analysis & Strategic Management*, 32(10), 1195–1208.

Megginson, W. L., Meles, A., Sampagnaro, G., and Verdoliva, V. (2019). Financial distress risk in initial public offerings: How much do venture capitalists matter? *Journal of Corporate Finance*, 59, 10–30.

Miyamoto, H., Mejia, C., and Kajikawa, Y. (2022). A Study of private equity rounds of entrepreneurial finance in EU: Are buyout funds uninvited guests for startup ecosystems? *Journal of Risk and Financial Management*, 15(6), 236.

Mustafa, M. (2021). Valuation of an early stage business. In *Springer Books*, pp. 137–164. New York City: Springer.

Pisoni, A., and Onetti, A. (2018). When startups exit: Comparing strategies in Europe and the USA. *Journal of Business Strategy,* 39(3), 26–33.

Rahaman, M. M. (2014). Do managerial behaviors trigger firm exit? The case of hyperactive bidders. *The Quarterly Review of Economics and Finance*, 54(1), 92–110.

Rao, S. V. R., and Kumar, L. (2016). Role of angel investor in Indian startup ecosystem. *FIIB Business Review*, 5(1), 3–14.

Riepe, J., and Uhl, K. (2020). Startups' demand for non-financial resources: Descriptive evidence from an international corporate venture capitalist. *Finance Research Letters*, 36, 101321.

Rompho, N. (2018). Operational performance measures for startups. *Measuring Business Excellence,* 22(1), 31-41.

Ross, G., Das, S., Sciro, D., and Raza, H. (2021). CapitalVX: A machine learning model for startup selection and exit prediction. *The Journal of Finance and Data Science,* 7, 94–114.

Rungi, M., Saks, E., and Tuisk, K. (2016). Financial and strategic impact of VCs on start-up development: Silicon Valley decacorns vs. Northern-European experience. In *Proceeding of the 2016 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, pp. 452–456. Bali, Indonesia.

Salamzadeh, A., and Kawamorita Kesim, H. (2017). The enterprising communities and startup ecosystem in Iran. *Journal of Enterprising Communities: People and Places in the Global Economy*, 11(4), 456–479.

Sathaworawong, P., Saengchote, K., and Thawesaengskulthai, N. (2019). Success factor of start-up fund raising in ASEAN. *Asian Administration & Management Review*, 2(2), 1–26.

Schwienbacher, A. (2019). Equity crowdfunding: Anything to celebrate? *Venture Capital*, 21(1), 65–74.

Song, Y., Dana, L. P., and Berger, R. (2021). The entrepreneurial process and online social networks: Forecasting survival rate. *Small Business Economics*, 56(3), 1171–1190.

Tang, L., Tian, Y., and Pardalos, P. M. (2019). A novel perspective on multiclass classification: Regular simplex support vector machine. *Information Sciences*, 480, 324–338.

Thanapongporn, A., Ratananopdonsakul, R., and Chanpord, W. (2021). Key success factors and framework of fundraising for early-stage startups in Thailand. *Academy of Strategic Management Journal*, 20(2S), 1–16.

Thirupathi, A. N., Alhanai, T., and Ghassemi, M. M. (2021). A machine learning approach to detect early signs of startup success. In *Proceedings of the Second ACM International Conference on AI in Finance*, pp. 1–8. New York, USA.

Venugopal, B., and Yerramilli, V. (2022). Seed-stage success and growth of angel co-investment networks. *The Review of Corporate Finance Studies*, 11(1), 169–210.

Wang, Z., Wu, C., Zheng, K., Niu, X., and Wang, X. (2019). SMOTETomek-based resampling for personality recognition. *IEEE Access*, 7, 129678–129689.

Zhang, J. (2011). The advantage of experienced start-up founders in venture capital acquisition: evidence from serial entrepreneurs. *Small Business Economics*, 36(2), 187–208.