Journal of AI and Data Mining
Vol 8, No 2, 2020, 279-287.
DOI: 10.22044/JADM.2020.8248.1963

A Hybrid Business Success Versus Failure Classification Prediction Model: A Case of Iranian Accelerated Start-ups

S. M. Sadatrasoul*, O. M. Ebadati and R. Saedi

Department of Management, Kharazmi University, Tehran, Iran.

Received 13 April 2019; Revised 02 January 2020; Accepted 21 January 2020
**Corresponding author: msadatrasoul@khu.ac.ir (S. M. Sadatrasoul).*

Abstract
The purpose of this work is to reduce the uncertainty in the early stage of start-up success prediction and fill the gap in the previous studies in the field by identifying and evaluating the success variables and developing a novel business success failure (S/F) data mining classification prediction model for the Iranian start-ups. For this purpose, in this paper, we seek to extend the Bill Gross and Robert Lussier S/F prediction model variables and algorithms in a new context of Iranian start-ups, which starts with accelerators in order to build a new S/F prediction model. A sample of 161 Iranian start-ups that are based on accelerators from 2013 to 2018 is applied, and 39 variables are extracted from the literature and organized into five groups. Then the sample is fed into six well-known classification algorithms. The two-staged stacking as a classification model is the best performer among all the other six classification-based S/F prediction models, and it can predict the binary-dependent variable of success or failure with an accuracy of 89%, on average. Also the finding shows that "starting from accelerators", "creativity and problem-solving ability of founders", "fist mover advantage", and "amount of seed investment" are the four most important variables that affect the start-up success, and the other 15 variables are less important.

Keywords: *Start-ups, Accelerator, Business Success Failure, (S/F) Prediction Model, Stacking, Venture Capital.*

1. Introduction

As accelerated start-ups become mature, they change how the people live in the world often by disrupting the old ideas and lifestyles with new solutions. Some of them changed how we live and became unicorn in the last decades, and gradually they organize a significant part of GDP of countries. They also increase the economy productivity [1] and a source of pioneer innovation [2]. Usually start-ups are grow in innovation ecosystems but studying and identification of the factors that describe the survival or failure of start-ups is very important for public policy-makers, professional investors like venture capitalists, angel investors, and even the entrepreneurs because they often want to consume their total time at least for some years and also their money, and therefore, they have the most important opportunity costs.

Many start-ups fail during their lifetime and the failure rate is mostly in their early stage but it continues to mid-stage, and finally, to their growth stage, although the failure rate usually decreases from an early stage to the growth stages [3]. Statistics also confirms the issue; according to the U.S small business administration, over 50% of small businesses fail in their first year of establishment, 33% fail after two years, and finally, 90% of them fail in their first five years [4]; however U.S has been ranked second at the business success rate in 2006. But what are the main failure factors?

Many researchers have targeted predicting the success and failure in the literature. Predicting the entrepreneurial failure and success is an important area of research. It has been revealed that the environmental factors play an important role in business success but the remaining 90% belong to the internal factors [۵]. There are several data mining classification and statistical learning regression models developed to predict the company success/failure in different sizes to predict the business success or failure. Large organization success models are usually used as the financial ratios and are studied under the bankruptcy prediction term. The altman z ratio is one of the most famous models studied [6]. There are also other models developed by others [7] that have studied criticizing the finding and the ability of financial ratios in the predicting success of large businesses (LB) on some datasets [8]. The models based on financial ratios are usually not appropriate for small and mid-sized businesses (SMB). This is mainly due to a less reliable and available data from SMBs compared to LBs. Therefore, models for SMBs have been developed using non-financial data by the researchers including Reynolds and Miller [9], Cooper *et al*. [10], Cooper *et al*. [11], Lussier [12], Lussier and Corman [13], and Lussier and Pfeifer [14]. Start-ups are recognized particularly for their business model uncertainty, repeatability, and scalability. Actually, start-ups are SMB in their early stage and mid-stage of growth, and they are LB in their growth and mature stages. Due to their special issues, researchers including [12, 15-17] have studied the start-up success failure models separately but studies on the start-up success is not mature in terms of variables and a prediction model must feed, and therefore, different researchers have used different factors.

This paper fills the gap of predicting the success failure prediction model firstly for accelerated start-ups, although the paper cannot evaluate and compare the start-ups founded from accelerators and other methods including bootstrapped, studio or angel investment backed start-ups; secondly, it fills the gap of studying the S/F prediction models that have been conducted in Iran. The objective of this work was to combine and test the validated Lussier model (Lussier and Pfeifer 2001) and Gross practical model start-up success evaluation in oral presentation at TED [18] as an orderly as an academic and practical findings for accelerated start-ups in Iran. The selected data mining classification methods were applied for building the papers success/failure prediction model.

2. Review of literature: Success versus failure prediction model

There are many studies investigating the success/failure predictions of start-ups. Lussier and corman have used a stepwise discriminant analysis with 15 independent variables in order to predict the success or failure of 96 companies. These variables were sorted by the ability to discriminate between failure and success, and described as the formula S/F = f (professional advisors, planning, education, minority business ownership, staffing, parents owning a business, record keeping and financial control, capital, industry experience, economic timing). The final model accuracy was 75% [13].

Gelderen *et al*. have estimated the importance of a variety of approaches and variables of a sample of 517 nascent entrepreneurs based on the Gartner's framework of a new venture creation [19]. The Gartner's framework explains that the start-up development endeavours differ in terms of the characteristics of the individual(s) who start the venture, the environment surrounding the new venture, the organization that they create, and finally, the process by which the new venture is started [20]. Finally, they used the logistic regression for their prediction. Krishna *et al*. have collected the data of 7000 start-up companies and 4000 failed ones from the famous crunchbase website [16].

They used 6 classification algorithms using Leave-One-Out Cross Validation (LOOCV) for evaluation, and reached 90% precision and over 0.9 of the area under curve (ROC). Using 150 interactions between entrepreneurs and potential investors, Maxwell *et al*. have studied the early stage angel decisions, showing that angels use an elimination-by-aspects to reduce the available ventures [21]. Dellerman *et al*. have developed a preliminary hybrid intelligence method and introduced a taxonomy of potential predictors that can be generalized for modelling the start-up success predictions [22].

Bohm *et al*. have described the concept of a business model DNA and applied it to 181 start-ups from Germany and USA combined with mattermark, crunchbase, deadpool, and autopsy.io datasets; they showed that there were 12 individual business models each having a distinct growth pattern and a chance to success; they claimed an accuracy rate of 83.6% for predicting the success businesses using dataset and 55 business model patterns [23].

There are also many studies in the area including Alexander *et al*. [24], Antretter *et al*. [25], and Maulana *et al*. [26].It can also found a comprehensive review in [27].

Table 1. Variables introduced and used by different studies that are organized in the Gross's five sectors.

Gross [18]	20 studies Lussier [13]	Krishna *et al*. [16]	Bohm *et al*. [23]	Gelderen *et al*. [19]	Gartner [20]
Appropriate timing (T)	Economic timing, product/service timing	Bad luck or timing, market competition, start date, defunct date, months active	-	-	Accessibility of suppliers, accessibility of customers or new market, governmental influences, proximity of universities, availability of land or facilities, accessibility of transportation, attitude of the area of population, availability of supporting services, high occupational and industrial differentiation, high percent of recent immigrants in the population, large industrial base, large size urban areas, barriers to entry, government rule changes
Team and good execution (TE)	Industry experience, management experience, professional advisors, education, staffing, age, partners, parents, minority, marketing	No focus (lack of traction), no flexibility, no passion or persistence, wrong or incomplete leadership, unmotivated team, no mentor or adviser, no VC experience, social skills-networking with the targeted audience, discipline, determination, ability to adapt to changes, fund raising skills, unwavering belief, low burn rate, good management system, good use of funds and time	Involved people industry/ foundation experts, investors, founding team size, education of founders, location (country & city)	Gender, age, work experience, management experience, experience in firm founding, education, push motivation, ambition become rich, information and guidance, industry experience, ambition to grow large, start out part- or full-time, techno nascent, team	Need for achievement, locus of control, risk taking propensity, job satisfaction, previous work experience, entrepreneurial parents, age, education, presence of experienced entrepreneurs, technically skilled labor force, living conditions, overall cost leadership
Idea truth outlier (I)	Planning	A small similar or non-scalable Idea	Idea closeness to science and patents, idea competition and innovativeness	Business plan	Rivalry among existing competitors, pressure from substitute products/services, bargaining power of buyers, bargaining power of suppliers,
Business model (BM)	Record keeping and financial control	Severity scores (wrong market positioning, no go-to-market strategy, a vision to monetize from the very beginning, weighted average, market value, burn rate, no revenue model, no long term road-map for return of investment, prospects of future earnings)	BM DNA, cluster, scope, focus (B2C or B2B), industry, physical assets, firm age	Risk of the market, industry type	-
Funding (F)	Capital	High burn rate, less capital than needed, composition of capital structure, seed Funding, total rounds of funding, time for seed (in months), series A, B , C..., G funding, valuation, total funds	-	Third party money, start-up capital	Venture capital availability, availability of financial resources

As a practitioner, Gross founded a lot of start-ups and incubated many others. As he got curious about why some start-ups became successful and the others failed, he gathered data from more than 200 companies and found the five key factors that influenced the start-up succession or failure.

These factors were appropriate timing (42%), team and good execution (32%), idea truth outlier (28%), business model (24%), and finally, funding (14%) [18].

Table 1 organizes the variables of the major selected studies in these five sectors.

3. S/F Prediction model method

Building the success/failure prediction model is organized in three steps inducing data gathering and pre-processing, predictive model building, and evaluation; these are described in the following sub-sections.

3.1. Data gathering and pre-processing

In order to identify the relevant S/F variables, a literature analysis is conducted and organized in the Gross five major sections, which are partially shown in table 1. The four steps of this stage are as follow: (a) In this step, interviews are conducted with the industry experts (n = 10; average duration of 45 min) to iteratively combine the findings from appropriate practical variables with variables extracted from the literature review considering that we are seeking the independent variables in order to find the Gross model using formula S/F = f (appropriate timing, team and good execution, idea truth outlier, business model and funding). Finally, 39 appropriate variables are extracted for a S/F prediction model in the context of predicting the success of accelerated start-ups in Iran, shown in table 2. (b) Then the data of a sample of 161 Iranian start-ups (33 failed and 131 success until now; failed/successful ratio = 0.25), which is based on accelerators from 2013 to 2018, is collected carefully using questionnaire and research for 39 variables. Then apparently, numeric variables are converted into nominal e.g. seed funding (yes or no). (c) As there are significant missing values for the variables, substitution strategy using decision tree, naive bayes, and k nearest neighborhood is applied. (d) Three metahuristic methods including particle swarm, genetic algorithm, and greedy search are applied for feature selection in order to gain the most suitable entropy and information gain. (e) Six major classifiers are applied in order to build the final best model.

3.2. S/F Predictive model building

This is where data mining classifiers are employed to construct the S/F prediction models on the pre-processed data. The two steps of this stage are as follow: (a) Splitting the pre-processed data in the previous stage to training and testing sets. The dataset is randomly split ten times into 90% of the data for training and 10% for testing (10-fold cross-validation) since this often leads to better results, and the training data is used to produce the model each time. (b) Constructing a model on the training data using leading classifiers including Naive Bayes (NB), K nearest neighborhood (KNN), Adaboost decision trees (Adaboost DTs), decision trees (DTs), support vector machines (SVMs), and stacking.

3.3. Evaluation

This stage is important mostly for evaluation of the predictive model on the testing data. The performance of the model is measured by an accuracy percentage of the predictions that are correct), specificity (percentage of negatively labelled records that were predicted as negative), sensitivity (percentage of positive labelled records that were predicted as positive), and area under the curve (AUC); the higher values for them are of interest. AUC indicates how the classifier performs in comparison to a random classifier. The random classifier would have an AUC of 0.5, and an AUC of one indicates the best classifier. Also because of a better evaluation of the imbalanced datasets of S/F prediction, AUC is applied in order to better evaluate the discrimination power of classifiers.

4. Results and discussion

Table 2 shows the results of the feature selection. Rapid minder and Weka are the tools that are used to run the results. Using the PSO method, the variable weight is set from zero to one, and the variables above 0.4 are selected to use for the next step, which is classification; it can be seen that 21 variables are selected using PSO for the next step. By applying GA, the weight results are shown to be zero or one, and 18 variables are selected. Using Greedy, there are 18 variables also selected. There are four variable added with one star for which all the three algorithms recognize them suitable including starting from accelerators, creativity and problem solving ability of founders, fist mover, and amount of seed investment. There are also 15 variables that are recognized important by two feature selection methods, and are included with double stars.In the second step, there are different parameter settings done for algorithms in order to work better considering the type of our dataset and problem. For DT, the split gini index is selected, maximum depth of tree is set to 20, minimum size of leafs is set to three, confidence level for prune is set to 0.25, and finally, the minimum number of records to finish is set to four; For SVM settings, the linear kernel function is used, C constant is set to 0.2, and 10000 iterations are considered; for KNN tuning, the number of k is set to 5, and oghlidos distance and weighed class labels are selected; at last, there are no special parameter setting for NB. Two-stage stacking is used to report the results of stacking; in the first round, decision trees (DTs), and Naive Bayes (NBs) are used as the weak classifiers with a lower accuracy, and at the second round, K nearest neighborhood (KNN) is used as a stronger classifier with a higher accuracy. Table 3 shows the results of the classification using four different performance indicators. The best are also marked by stars; the best accuracy 92.68% is achieved by “Stacking+ GA” and marked with one star, although this matters to worsen sensitivity and AUC. The best sensitivity at 90% is achieved by “SVM + Greedy”, “KNN + GA” and “Stacking + GA” simultaniouly, although this matters to worsen accuracy, the authors are also made to consider that “SVM + Greedy” has the best AUC at 0.963 marked with four stars.

Table 2. The weight results for PSO, GA, and greedy; and variable affiliations to Gross categories are also shown in parentheses.

Variable	PSO	GA	Greedy	Variable	PSO	GA	Greedy	Variable	PSO	GA	Greedy
Economic timing (T)**	٠,١٢	١	١	Previous work (TE)	٠,٦١	٠	0.3	Product/service substitution status (I)	٠,٦٥	٠	0
Product/service timing (T)**	٠,٨٧	١	0.2	Creativity and problem-solving abiligy of founders	١	١	0.5	Industry type (I)	٠,٦٣	٠	0
Target market cometition at establishment	٠,٦٢	٠	0.4	Number of founders (TE)**	٠,٥١	١	0.3	Good product/service positioning against competitors (I)	٠,٦٣	٠	0.2
Year of establish (T)	٠,٢٠	٠	0.9	CEO eduation (TE)	٠,٤٨	٠	0.2	Fast follower (I)**	٠,٥٧	1	0.5
Motivated founders (TE)	١	١	0.1	Professional recognition of team members from each	٠,٥٣	١	0.2	Fist mover (I)*	٠,٦٨	0	0.6
Starting from accelerators (TE)*	٠,٩٧	١	١	Average age of founders (TE)	٠	١	0.1	Type of target market players (BM)**	٠,٤٧	1	0.5
CEO ability (TE)**	٠,٩١	١	0.2	Previous entreprenership experience (TE)**	١	١	0.3	Product/service based value propostion/capture	٠,٣٩	1	١
Customer feedback till know (TE)**	٠,٨٩	٠	١	Fisrt round investor (TE)	٠	٠	0.4	Customer base loaylty (BM)	٠,٢١	٠	0.8
Province of activation (TE)**	٠,٨٣	٠	0.9	Type of market penetration (TE)**	١	١	0.3	Business model scalability (BM)	٠,٦٦	٠	0.3
CEO geneder (TE)**	٠,٧٣	٠	0.5	Setteled in the accelerator (TE)**	٠,١٢	١	١	Type of value proposition competition *vs.*	٠,٣١	١	0.1
Complete and balanced team caabilities (TE)	٠,٦٩	٠	0.3	Setteled in the incubator (TE)**	٠,١٩	١	0.5	Product/service stage in its lifecycle (BM)	٠,٣٢	٠	0.3
Team vocational skills (TE)**	٠	١	0.5	Bussiness model clone or new innovation (I)	٠,٨١	٠	0.1	Amount of seed invetment (F)*	٠,٧٠	١	0.7
Team flxibilty for scaling (TE)	٠	٠	0.8	Barriors to entry (I)	٠,١٩	٠	0.3	Start-up is at which round (pre-seed, seed, A, … ? (F)	٠,٣٠	٠	١

Table 3. Results of the analysis.

Missing value handeling	DT mutation				KNN mutation				BN mutation			
Classifier+feature selection	**Accuracy**	**Sensitivity**	**Specificity**	**AUC**	**Accuracy**	**Sensitivity**	**Specificity**	**AUC**	**Accuracy**	**Sensitivity**	**Specificity**	AUC
DT + GA	86.54	62.50	92.11	0.853	80.55	43.33	91.33	0.733	87.87	72.50	93.29	0.837
DT + PSO	82.87	33.33	95.42	0.637	78.12	30	91.33	0.663	82.94	57.50	90.81	0.860
DT + Greedy	85.96	65.83	90.73	0.854	85.48	44.17	97.74* **	0.774	90.18	72.50	95.31	0.916
Adaboost DT + GA	86.54	62.50	92.11	0.853	80.59	36.67	92.93	0.776	89.74	77.50	94.12	0.931
Adaboost DT + PSO	79.82	42.50	89.19	0.759	80.59	37.50	92.29	0.665	83.42	57.50	91.39	0.832
Adaboost DT + Greedy	85.29	64.17	90. 61	0.841	85.48	44.17	97.74* **	0.774	91.43	78.33	95.43	0.940
SVM + GA	87.87	78.33	90.20	0.874	88.16	62.50	90.24	0.831	88.38	85	90.08	0.940
SVM + PSO	74.26	65.83 %	76.14	0.823	76.47	60	81.36	0.868	84.15	70	88.48	0.919
SVM + Greedy	89.67	82.50	91.80	0.930	89.15	81.67	91.45	0.936	90.26	90**	90.62	0.963* ***
KNN + GA	91.51	72.50	96.21	0.903	82.98	50	92.99	0.883	90.29	90**	90.35	0.927
KNN + PSO	88.27	69.17	92.94	0.857	81.88	54.17	89.84	0.865	89.67	85	91.57	0.921
KNN + Greedy	92.10	72.50	96.97	0.919	87.39	64.17	94.73	0.836	90.26	87.50	91.50	0.936
NB + GA	86.62	62.50	93.12	0.842	85.55	59.17	93.30	0.830	92.06	86.67	93.83	0.950
NB + PSO	87.24	70.83	91.76	0.897	83.75	59.17	90.80	0.825	85.99	73.33	89.90	0.928
NB + Greedy	89.67	77.50	93.05	0.927	87.24	67.50	92.87	0.897	89.56	85	91.38	0.947
Stacking + GA	92.68*	75.83	96.86	0.883	86.14	57.50	94.72	0.877	90.29	90**	90.85	0.931
Stacking + PSO	89.08	68.33	94.78	0.886	84.26	50.83	93.70	0.866	88.42	73.33	90.05	0.920
Stacking + Greedy	90.88	72.50	95.31	0.901	85.48	57.50	93.82	0.847	89.60	82.50	92.22	0.925

Finally, the best specificity with three stars belongs to "AdaboostDT + Greedy", and on the other hand, its results are very poor for specificity and AUC.

4.1. Main success/failure model prediction performance analysis

Figure 1 shows the average performance of six different classifiers. DT and AdDT have the worst, and sensitivity (50% near to chance considering the binary situation of class variable successful/fail start-up) and AUC mean that they are not able to recognize the failed start-up well; one the other hand, specificity is superior to others, which shows that they are better to recognize the successful start-ups. SVM has the best sensitivity, and AUC and concurrently compete in accuracy with others.

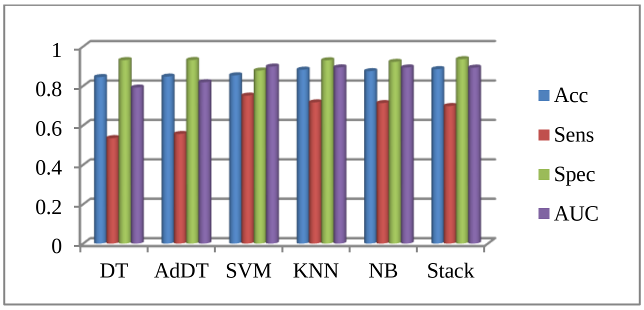

Figure 1. Performance of six classifiers used in the study from four different performance indicator views (average of 12 model performance for each algorithm).

Table 4 shows the rank of algorithms in four different performance indicators. It can be seen that KNN mostly acts better than NB, and stacking is always better than AdDT, and AdDT is better than DT.

Table 4. Overall rank of six technique ranks based on average performance (fist four columns) and repeated patterns extracted from first column (two last columns).

Performance indicator	Overall rank based on average
Accuracy	Stack > KNN > NB > SVM > AdDT > DT
Sensitivity	SVM > KNN > NB > Stack > AdDT > DT
Specificity	Stack > AdDT > DT > KNN > NB > SVM
AUC	SVM > KNN = NB > Stack > AdDT > DT
Total patterns with 100% repeat	KNN >= NB & Stack > AdDT > DT
Total patterns with 75% repeat	SVM > AdDT> DT & NB > AdDT > DT

4.2. Missing value handling performance analysis

Figure 2 shows the average performance of three different missing value handling techniques. It can be seen that BN mutation is the best performer except for the specificity performance indicator, which means that it cannot predict the failed start-ups better than the other techniques, and it is important because the S/F prediction model dataset is often imbalanced. On the other hand, the specificity of KNN is the worst and near 50%, which means that KNN predicts the successful start-ups nearly by chance.

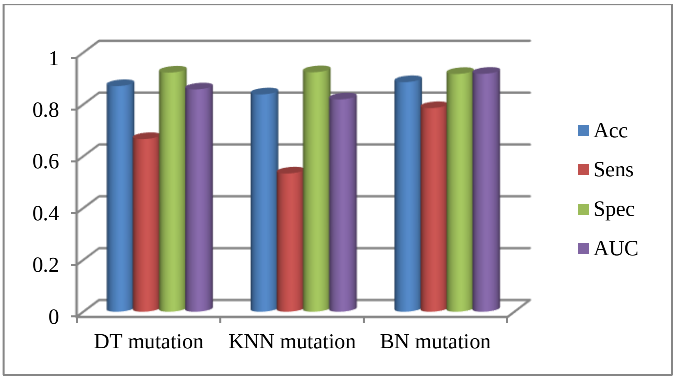

Figure 2. Missing value handling of three applied techniques (average of 18 model performance for each mutation).

Table 5 shows the overall rank of mutation technique rank based on the average performance of three methods. BN is the best performer followed closely by DT, and finally, KNN overall. DT is often a better performer than KNN.

Table 5. Overall rank of mutation technique rank based on average performance.

Performance indicator	Overall rank based on average
Accuracy	BN > DT > KNN
Sensitivity	BN > DT > KNN
Specificity	KNN >= DT > BN
AUC	BN > DT > KNN
Total patterns with 100% repeat	DT >= KNN
Total patterns with 75% repeat	BN > DT >= KNN

4.3. Feature selection method performance analysis

Figure 3 shows the average performance of three different feature selection methods. It can be seen that greedy is always the best performer and GA is mostly better than PSO.

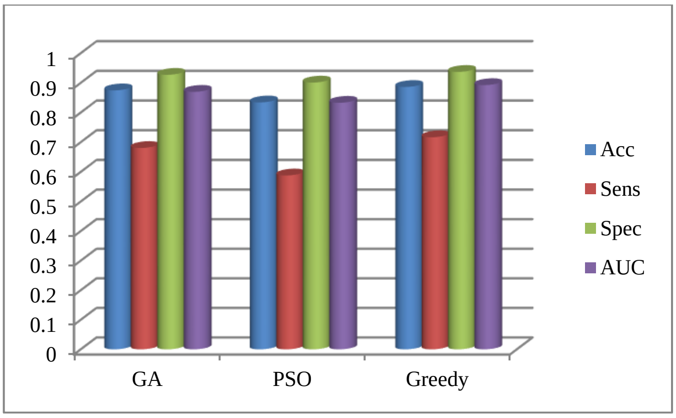

Figure 3. Feature selection method performance (average of 24 model performance for each feature selection method).

Table 6 shows the overall rank of feature selection methods, on average, among all the classification methods that use them. It can be seen that greedy is the best performer undeniably, and PSO is the worst performer.

Table 6. Overall rank of mutation technique rank based on average performance.

Performance indicator	Overall rank based on average
Accuracy	Greedy > GA > PSO
Sensitivity	Greedy > GA > PSO
Specificity	Greedy > GA > PSO
AUC	Greedy > PSO > GA
Total patterns with 100% repeat	Greedy > GA, PSO
Total patterns with 75% repeat	GA > PSO

5. Implications and conclusions

In general, a proposed research model would be applied to evaluate new venture evaluations and build new S/F prediction models in different areas of technology economy. In particular, we have shown and analyzed new variables from the literature review for our own problem for evaluating accelerator success in Iran and build tune classifiers on that. On one hand, our model can predict up to 92.68% accuracy using “Stacking + GA” with DT mutation; on the other hand, someone would use it more practically. For example, “SVM + Greedy” with BN mutation can be used by risk averse venture capitals as an assistant tool because of its ability to recognize failed start-ups by a reported sensitivity of 90% and AUC of 0.963. Also risk taker venture capitals can use “AdaboostDT + Greedy” or “DT + Greedy” with KNN mutation of 97.74%.
Further research work might explore how DNA of business models and their type considering our variables can affect the predictions. Also the literature is very poor to investigate the S/F prediction models for star-tup studios. Finally, a useful S/F prediction model for accelerated start-ups for a practical perdition situations is provided, and that would support “series A venture capital investors” in making better decisions and reduce the frequency of bad investments.

References

[1] Croitoru, A.J.J.o.C.R.i.A. & Sociology, Schumpeter, JA, 1934 (2008 The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest and the Business Cycle, translated from the German by Redvers Opie, New Brunswick (USA) and London (UK): Transaction Publishers, vol. 3, no. 2, pp. 1-13.

[2] Hunt, R.A.J.E.J.o.I.M. (2013). Entrepreneurial tweaking: An empirical study of technology diffusion through secondary inventions and design modifications by start-ups, vol. 16 , no. 2, pp. 148-170.

[3] Kirchhoff, C. J., et al. (2013). Actionable knowledge for environmental decision making: broadening the usability of climate science, vol.38.

[4] Administration, S. B. (2007) [cited 2007 September 21]; Available from: SBA Web site https://www.sba.gov.

[5] Carlson, R.J.J.s.L.R. (2006). The Small Firm Exemption and the Single Employer Doctrine in Employment Discrimination Law, vol. 80, pp. 1197.

[6] Eidleman, G. J. (1995). Z scores-A Guide to failure prediction. The CPA Journal, vol. 65, no. 2, pp. 52.

[7] Sheppard, J. P. (1994). The dilemma of matched pairs and diversified firms in bankruptcy prediction models. The Mid-Atlantic Journal of Business, vol. 30, no. 1, pp. 9.

[8] El-Zayaty, A. I. (1987). Business failure prediction: an analysis of type II prediction errors, City University of New York.

[9] Reynolds, P. (1987). New Firms: Societal Contribution Versus Potential. Journal of Business Venturing.

[10] Cooper, A. C., Gimeno-Gascon, F. J., & Woo, C. Y. (1991). A Resource-Based Prediction of New Venture Survival and Growth. Academy of Management Proceedings, pp. 113–119.

[11] Cooper, A. C., et al. (1990). New Business in America: The Firms & Their Owners. Washington, DC: The NFIB Foundation.

[12] Lussier, R. N. (1996). A startup business success versus failure prediction model for the retail industry. The Mid-Atlantic Journal of Business, vol. 32, no. 2, pp. 79.

[13] Lussier, R. N. & Corman, J. (1996). A business success versus failure prediction model for entrepreneurs with 0-10 employees. Journal of Small Business Strategy, vol. 7, no. 1, pp. 21-36.

[14] Lussier, R. N., & Pfeifer, S. (2000). A Comparison of Business Success Versus Failure Variables between U.S. and Central Eastern Europe Croatian Entrepreneurs. Entrepreneurship Theory and Practice, vol. 4, no. 24, pp. 59–67.

[15] Khan, A. M. (1986). Entrepreneur characteristics and the prediction of new venture success. Omega, vol. 14, no. 5, pp. 365-372.

[16] Krishna, A., A. Agrawal, & Choudhary, A. (2016). Predicting the Outcome of Startups: Less Failure, More Success. in 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW). IEEE.

[17] Van Gelderen, M., R. Thurik, & Bosma, N. (2006). Success and risk factors in the pre-startup phase. Small Business Economics, vol. 26, no. 4, pp. 319-335.

[18] Gross, B. (2015). The single biggest reason why start-ups succeed. [cited 2018 septamber].

[19] Van Gelderen, M., R. Thurik, & Bosma, N. (2005). Success and risk factors in the pre-startup phase. Small business economics, vol. 24, no. 4, pp. 365-380.

[20] Gartner, W. B. (1985). A conceptual framework for describing the phenomenon of new venture creation. Academy of management review, vol. 10, no. 4, pp. 696-706.

[21] Maxwell, A. L., Jeffrey, S. A. & Lévesque, M. (2011). Business angel early stage decision making. Journal of Business Venturing, vol. 26, no. 2, pp. 212-225.

[22] Dellermann, D., et al. (2017). Finding the unicorn: Predicting early stage startup success through a hybrid intelligence method.

[23] Böhm, M., et al. (2017). The business model DNA: Towards an approach for predicting business model success.

[24] Kessler, A., et al. (2012). Predicting founding success and new venture survival: A longitudinal nascent entrepreneurship approach. Journal of Enterprising Culture, vol. 20, no. 01, pp. 25-55.

[25] Antretter, T., I. Blohm, & Grichnik, D. (2018). Predicting Startup Survival from Digital Traces: Towards a Procedure for Early Stage Investors.

[26] Maulana, A., et al. (2008). Requirements for Platform-based Startup Survival: A Qualitative Exploratory Study. in 2018 International Conference on Advanced Computer Science and Information Systems (ICACSIS), IEEE.

[27] sadatrasoul, S. M., et al. (2013). Credit scoring in banks and financial institutions via data mining techniques: A literature review. Journal of AI and Data Mining, vol. 1, no. 2, pp. 119-129.

نشریه هوش مصنوعی و داده کاوی

IAIDM

ارایه یک مدل پیش‌بینی موفقیت و شکست ترکیبی: مورد بررسی شرکت‌های نوپای شتابدهی شده در ایران

سید مهدی سادات رسول*، امید مهدی عبادتی و رقیه ساعدی

دانشکده مدیریت، دانشگاه خوارزمی، تهران، ایران.

ارسال ۲۰۱۹/۰۴/۱۳؛ بازنگری ۲۰۲۰/۰۱/۰۲؛ پذیرش ۲۰۲۰/۰۱/۲۱

چکیده:

هدف از این تحقیق توسعه مدل پیش‌بینی موفقیت یا شکست شرکت‌های نوپای شتابدهی شده به منظور کاهش عدم قطعیت در انتخاب شرکت‌های موفق می‌باشد. بدین منظور مقاله نسبت به توسعه و بسط مدل پیش‌بینی موفقیت و شکست لوزیر از طریق ادغام با مدل تجربی بیل گراس پرداخته و مدل جدیدی را ارایه نموده است. مدل ارایه شده روی یک نمونه ۱۶۱ عددی از شرکت‌های نوپای شتابدهی شده اجرا شده است که از سال ۲۰۱۳ الی ۲۰۱۸ در ایران شتابدهی شده‌اند. اطلاعات این ۱۶۱ شرکت از منظر ۳۹ شاخص جمع‌آوری شده است و در پنج دسته اصلی دسته‌بندی شده است. پایگاه داده ایجاد شده مبنای تنظیم نمودن شش مدل رده‌بند داده‌کاوی قرار گرفته است. مدل استیکینگ از میان شش مدل به کار گرفته شده با ۸۹ درصد نرخ صحت توانسته است بهتر از سایر مدل‌ها شرکت‌های موفق را شناسایی نماید. همچنین شاخص‌های شروع از شتابدهنده ها، توانایی حل مسئله و خلاقیت بنیانگذاران، مزیت شروع کننده اول، میزان سرمایه مرحله بذری به عنوان شاخص‌های مهم در میزان موفقیت شرکت‌های نوپا در این برهه از زمان در ایران شناخته شده‌اند.

کلمات کلیدی: شرکت‌های نوپا، شتابدهنده، مدل پیش‌بینی موفقیت/شکست، استیکینگ، سرمایه‌گذاری جسورانه.