International Journal of Informatics and Communication Technology (IJ-ICT)
Vol. 14, No. 2, August 2025, pp. 446~456
ISSN: 2252-8776, DOI: 10.11591/ijict.v14i2.pp446-456

Machine learning in detecting and interpreting business incubator success data and datasets

Mochammad Haldi Widianto[1], Puji Prabowo[2]
[1]Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia
[2]Entrepreneurship Department, Bina Nusantara University, Jakarta, Indonesia

Article Info

Article history:

Received Jul 26, 2024
Revised Dec 6, 2024
Accepted Dec 15, 2024

Keywords:

AdaBoost
Feature
Interpretation
Machine learning
Model architecture
Select K-best

ABSTRACT

This research contributes to creating a proposed architectural model by utilizing several machine learning (ML) algorithms, heatmap correlation, and ML interpretation. Several algorithms are used, such as K-nearest neighbors (KNN) to the adaptive boosting (AdaBoost) algorithm, and heatmap correlation is used to see the relationship between variables. Finally, select K-best is used in the results, showing that several proposed model ML algorithms such as AdaBoost, CatBoost, and XGBoost have accuracy, precision, and recall of 94% and an F1-score of 93%. However, the computing time the best ML is AdaBoost with 0.081s. Then, finally, the proposed model results of the interpretation of AdaBoost using select K-best are the best features "last revenue" and "first revenue" with k feature values of 0.58 and 0.196, these features influence the success of the business. The results show that the proposed model successfully utilized model classification, correlation, and interpretation. The proposed model still has weaknesses, such as the ML model being outdated and not having too many interpretation features. The future research might maximize with ML models and the latest interpretations. These improvements could be in the form of ML algorithms that are more immune to data uncertainty, and interpretation of results with wider data.

Corresponding Author:

Mochammad Haldi Widianto
Computer Science Department, School of Computer Science, Bina Nusantara University
Bandung Campus, Jakarta, 11480, Indonesia
Email: mochamad.widianto@binus.ac.id

1. INTRODUCTION

Entrepreneurship education continues to grow because of its benefits in developing students' competencies and entrepreneurial skills that the mindset and intention to become entrepreneurs continue to grow from this activity [1]. Activities that support students to learn directly in the field and support programs can improve technical skills as entrepreneurs [2], [3]. In addition, this entrepreneurship education can help develop business performance [4]. Activities that are supported by direct application in the field, an ecosystem that encourages students to be enthusiastic about solving problems, can encourage innovative ideas [5] With universities being able to play a role in business and economic growth, entrepreneurship education can encourage the birth of innovation from new businesses [4].

Universities that focus on developing entrepreneurship, in this case, fully support the birth of new entrepreneurs through the application of entrepreneurship in various courses, creative programs, university-based business incubators, and other supporting facilities that can be a good ecosystem for growing new entrepreneurs [6], [7]. The development of services, teaching, and their influence on the curriculum are required to meet student competencies to obtain the best results [8], [9]. In addition, the existence of a system and periodic evaluation are essential in understanding the needs of business students that the performance of services and facilities can be measured [10].

The existence of sound system and service integration between universities and business incubators can affect the development of businesses being managed [11]. Currently, not many businesses incubators measure and understand their participants' needs in depth. This can help incubators develop supporting facilities such as activities and mentoring [12], [13]. The success of a university-based incubator is greatly influenced by many factors that will affect its performance. Starting from internal university support, facilities in the incubator, entrance selection, human resource capabilities, knowledge of technology and finance, and managerial. Then, external support, such as funding support, business networks, and government policies [14], [15]. Therefore, a manager is needed, as well as good management and measurable evaluation and measurement for success in a business incubator [16].

Predicting, classification, clustering, and interpretation are necessary, especially when building a sustainable business. One way is by utilizing artificial intelligence (AI) [17], [18], However, its use needs to be paid attention to, especially for its business purposes [19], [20] AI also does not stand alone because it has other research branches, such as machine learning (ML) [21], deep learning (DL) [22], and transfer learning (TL). On the other hand, their uses are different. However, in business cases, it is expected to use ML because the data used is usually a lot and can be used as training data [23], [24].

This AI technology also changes how companies do business, allowing them to analyze data and automate decision-making processes. By looking at historical data, ML algorithms can predict future possibilities, which is invaluable for inventory management, demand forecasting, and financial planning. Businesses are using ML to gain deeper insights into customer behavior, preferences, and engagement, resulting in more personalized services and products [25]. However, several questions arise about how to use it or collaborate with it, where some research may only use ML alone without creating a start or architecture for its creation. So, it is felt that the use of ML is not optimal [26], [27].

Previous research has not used ML as much for business classification and interpretation. Martínez *et al.* [28], utilize gradient boosting in its use in business. Kraus *et al.* [29], utilize DL in business conditions. Nakhal *et al.* [30], like others, they also use ML in business. However, there is still not much research that emphasizes architectural creation in classifying and interpreting it. However, some approaches rarely discuss how to interpret the algorithm's results. Therefore, this was a disadvantage in the previous study [31].

This research contributes and investigates by taking advantage of previous research's limitations. Previous research only discussed prediction and classification. This has been clarified by the need for a way to find the interpretation of the ML results. Meanwhile, there is not much clarification regarding the interpretation of the data because this explanation is needed to understand how the ML learning algorithm works. Interpretation of results is usually used so that humans better understand the influence of prediction results on existing features.

Therefore, this research will fill the gaps in previous research by proposing several ML model architectures in this study to provide new solutions for interpreting ML results and can also improve ML performance. This research aims to create an AI model architecture to help see the success and interpretation of the success of the incubator business. The following are his main contributions:

a) This study builds a model architecture based on ML algorithm and looks for a good algorithm based on its classification.
b) The proposed model architecture is used to classify incubator businesses based on existing data.
c) The results of the best ML model need to be interpreted from proposed model using several good algorithms to find dominant features.
d) Present complete datasets in the form of incubator business data.

The article has four sections in the last section. Section 1 explains the problem description and introduction. Section 2 explains the proposed model. Section 3 explains the implementation and trial results. Finally, section 4 is followed by conclusions and suggestions for further research.

2. RESEARCH METHOD

This section will explain the stages of forming an model architecture for an incubator business. Previous researchers have used many general models, but the best ML algorithms must be found when using classification in incubator businesses. Then, the best ML algorithm must be interpreted to obtain the dominant features. Figure 1 shows the research process flow for classifying business success and interpreting the success results, starting from "start" to "Finish." The following are the stages.

a) The first phase is the data introduction phase. Information comes from primary data, namely the Bandung binus business incubator. This data collection was collected by Bandung binus incubator businesses.
   The dataset contains 151 data with 17 features. With the detected variable being "*Target*".
b) Next, we start by handling the data because it will be processed. The author will look for several features that do not affect the classification results. Therefore, the selection is carried out to maximize the results of classification and interpretation. The feature results after elimination are 15 features.
c) The next stage, the fixed data, is carried out by Pearson correlation by looking for a heatmap and examining the correlation results of each variable, especially the primary variable, "*Target*".
d) the data was also treated to conform to a normal distribution because all data was not normally distributed. At this stage, the result range is [0 to 1]. Therefore, it is usually distributed. At this phase, divided into 2 with a comparison between training data and test data, namely 90/10 for training data.
e) At this stage, various suitable ML algorithms are used, which will become a benchmark in the next stage of the interpretation process.
f) At this stage, the test results are the performance of all the selected ML algorithms.
g) After looking for performance, several parameters are evaluated, including accuracy, recall, F1-score, precision, and computing time. This aims to get results from architecture.
h) After that, the results of the best ML model will be interpreted for dominant features using select K-best. This final result is helpful for determining which features are parameters in the incubator business's success.

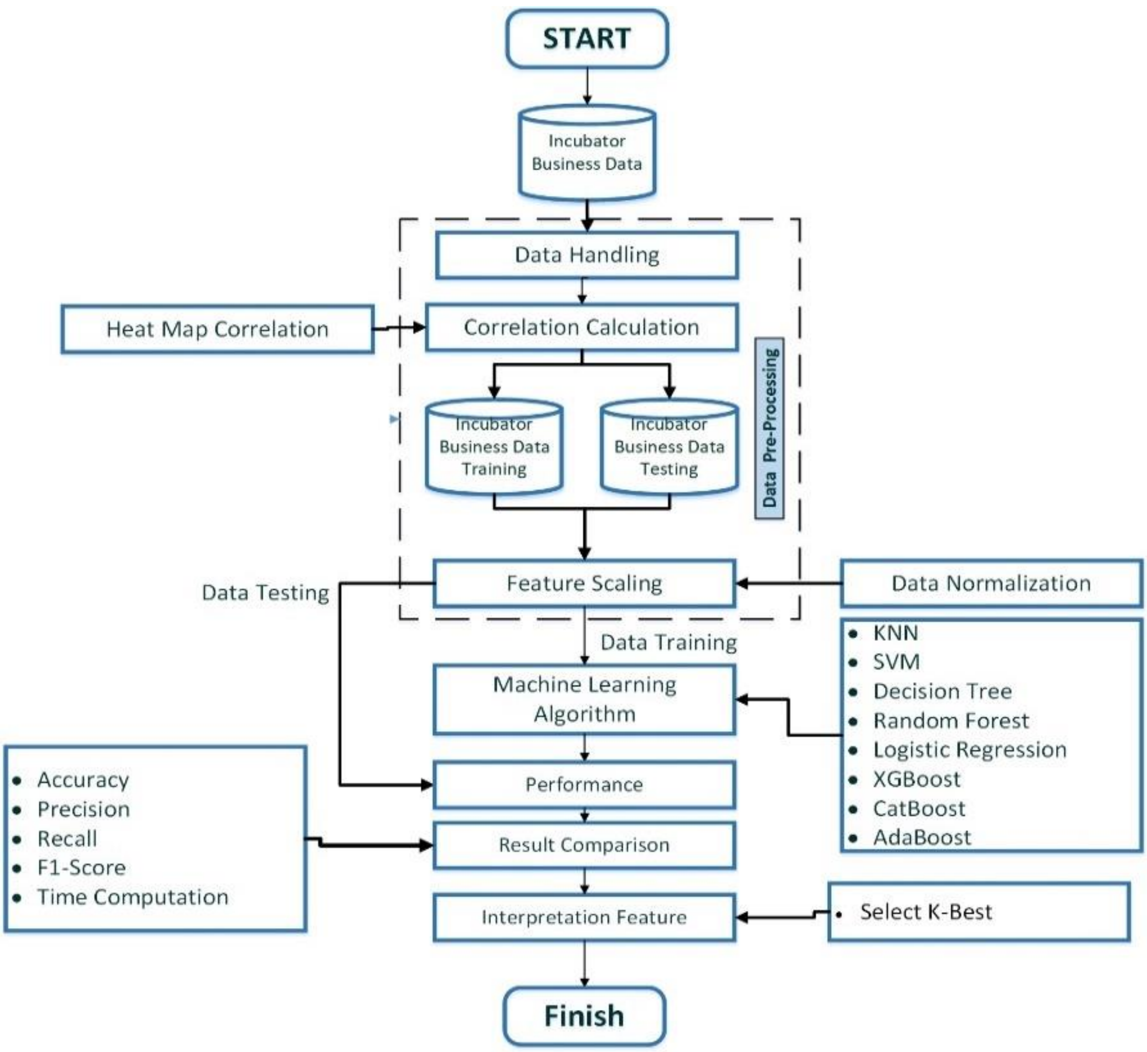

Figure 1. Research proposed model

The proposed model flow is used to incubate a successful process as shown in Figure 1. The data used in this study utilizes primary data from [32]. In Figure 1 the author only uses a few algorithms used in the proposed model, because these ML algorithms are very commonly used in research. Inspired by this widely used model, the ML algorithm in Figure 1 is used.

2.1. Data information

According to the author, the data used is an incubator from Binus Bandung, which was processed to remove useless data and produce processed data as shown in Table 1. Table 1 explains the data used in this study based on a business incubator. Feature no 16's prediction target is "*Target*." The data taken results from an incubator business, which has a target when the final revenue can exceed 20 million rupiahs. Then, the target will be "1," whereas if it does not meet the target, it will be "0". The imbalance issue target has been achieved for each feature so that no more in-depth management is needed to get maximum results.

Table 1. Data information

No	Feature	Number of data sets	Data type
1	Business name	151	String
2	Business category	151	String
3	Business team member	151	Numeric
4	First revenue	151	Numeric
5	Last revenue	151	Numeric
6	Number of employees	151	Numeric
7	External collaboration	151	Numeric
8	Total follows the bazaar	151	Numeric
9	Number of online advertisements	151	Numeric
10	Using TikTok live	151	String
11	Additional investment from external	151	String
12	covered by the media	151	String
13	own trademark rights	151	String
14	have a halal certificate	151	String
15	Number of types of sales methods	151	String
16	Target	151	Numeric

All the features in Table 1 are selected and processed in the data handling section as shown in Figure 1, because there is not too much data, there is no splitting of small data. Because the data used is not too much, the challenge in the proposed model is managing the limited data in order to get good confusion matrix results.

2.2. Correlation

At this phase, a correlation test $r(x_i, y_i)$ is carried out on the dataset. In (1) the correlation formula itself [33]. Correlation is used to see how variables relate to each other, and according to the author this is very important to utilize:

$$r(x_i, y_i) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (1)$$

which:
r (x_i, y_i) : correlation coefficient.
x : x to-i.
$\bar{x}$: x data mean.
y : y to-i.

2.3. Logistic regression

Logistic regression (LR) is used in data analysis techniques, especially in the relationship of several variables used for classification algorithms. This variable is used in grouping classification based on the influence and relationship of several variables [34]. The formula found in (2):

$$log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \quad (2)$$

which:

(p) : the probability of the desired event.
(β) : intercept.
(β_1 ... β_n) : is the coefficient for the predictor variable (x_1 ... x_n).

2.4. SVM classification

Support vector machine (SVM) classification is a tmodel for finding the best hyperplane to divide classes linearly. In SVM, the goal is to looking for a hyperplane in N-dimensional space (N-number of features) that unambiguously classifies data points. It is used to identify data with similar patterns across thereby maximizing the margin between classes. The formula for SVM is as (3) [35]:

$$w.x - b \tag{3}$$

which:
(w) : the weight vector is average to the hyperplane.
(x) : a feature vector
(b) : bias

2.5. Decision tree

ML algorithm that uses a tree structure or decision hierarchy. Where the nodes in the tree represent features, the decision branches and each leaf represent the outcome. The formula for the decision tree (DT) approach is (4) [36]:

$$Entrophy\ (S) = -\sum_{i=1}^{n} p_i log_2 p_i \tag{4}$$

which
p_i : the sample proportion from class-i.

2.6. Random forest

ML method that starts from DT changes and becomes a famous classification (random forest (RF)). The following is the basic formula used in RF for classification presented in (5). This equation is a derivative of the previous use of DT, because RFs are derivatives of DT [37].

$$\hat{y} = modus\{h(x, \theta_k), k = 1, \ldots, K\} \tag{5}$$

Which:
($\hat{y}$) : the predicted class
($h(x, \theta_k)$) : a classification function applied to the input vector (x), with parameters (θ_k) which is unique to the ke tree -(k).
(K) : number of trees in the forest

2.7. XGBoost classification

The XGBoost classification is also a derivative of the DT. This ML model is more focused on the case of gradient boosting algorithms with a DT model as the basis for the classifier. This model can be explained with a (6). Where represents data containing n observations. *x* is the independent variable, and y is the dependent variable [38].

$$D = [x, y] \tag{6}$$

2.8. CatBoost classification

Catboost is a robust ML algorithm for classification. It is a type of boosting algorithm that optimizes gradient-boosting ML. Catboost is a symmetric DT technique that can solve the problems of prediction drift and gradient bias and reduce overfitting with stable and high accuracy and results [39]. In used (7):

$$\hat{y} = argmax_k \sum_{t=1}^{T} f_t\ (x)_k \tag{7}$$

which:

($\hat{y}$) : the predicted class
(f_K $(x)_k$): the score is given by the (t)th DT for the (k)th class.
(T) : is the total number of trees in the model.

2.9. K-nearest neighbors
In the KNN algorithm learning. The model using Manhattan as the distance algorithm is shown in (8) [40]. In (8) are used to find the closest past cases. This equation is beneficial in determining predictive values. The variables X and y from (1) represent only two vectors in the feature space, and xi and yi are their coordinates each [40].

$$\text{Manhattan distance: } d(x,y) = \sum_{i=1}^{n} |x_i - y_i| \tag{8}$$

Which:
i = 1 to n
p = positive integer

2.10. AdaBoost
AdaBoost, is an ML algorithm used as an ensemble or combination technique and is also the basis for deriving DT. It works by combining several weak learners to create a strong learner. Each weak learner usually consists of a simple DT. The following is the formula for AdaBoost [41], [42] (9):

$$w_{i+1} = w_i x \exp\left(\alpha_i x\, y_i\, x\, h_i(x_i)\right) \tag{9}$$

where:
(w_i) : the weight of the (i)-th observation.
(α_i) : the weight of the (i)-th classifier, calculated based on its error rate.
(α_i) : the actual label of the (i)-th observation.
($h_i(x_i)$) : the prediction of the (i)-th classifier for the (i)-th observation.

2.11. Select K-best
This method is used to remove less contributing parts from the attributes or features in the dataset so that it can reduce training time by calculating variables based on the highest correlation between the score function provided and the attribute or feature itself (10) [43].

$$x^2(i) = \sum \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} \tag{10}$$

Which:
(O_{ij}) : Observation frequency of class (j) with feature (i).
(E_{ij}) : is the expected frequency of class (j) with feature (i) if there is no relationship between the features.

3. RESULTS AND DISCUSSION
This section will discuss the results of two experimental proposed models. The proposed model was evaluated by looking at the dataset results based on pearson correlation. Then, the ML results will show which is the best and which will be used. Finally, the results of the best ML will be interpreted using select K-best to see the interpretation of the ML.

3.1. Correlation result
The authors found that the correlation of all data using (1) as shown in Figure 1, where no correlation exceeds 0.5. It can be said that all data is not correlated with each other. However, the best correlation is found in 3 features, namely “first income,” “last income,” and “number of employees.” This third attribute is compared with the target attribute, which correlates around “0.23”, “0.28”, and “0.34”. Where the attribute “number of employees” has the highest correlation value. Weak correlation affects the contribution of features to the algorithm, it is wrong but the author still uses all features because it still allows for a contribution between all features to the performance of the ML algorithm in the proposed model. Therefore, the author suggests that interpretation is necessary to see the influence of classification on the interpretation of the data. The author’s suggestions will also be presented in another section so that you can see the results of ML and interpret them.

3.2. Machine learning result

This section will explain the proposed ML model's results in all aspects, such as F1-score, accuracy, precision, and recall. These results are presented in Table 2. These results are useful in seeing the quality of the ML algorithm given by the proposed model.

Table 2. ML result

No	ML algorithm	Acuraccy	Precision	Recall	F1-score	Time computation
1	KNN	88%	77%	88%	82%	0.1s
2	DT	93%	93%	93%	92%	0.1s
3	LR	88%	77%	88%	82%	0.023
4	SVM	88%	77%	88%	82%	0.021s
5	RF	93%	93%	93%	92%	0.44s
6	XGBoost	94%	94%	94%	93%	0.61s
7	CatBoost	94%	94%	94%	93%	1.49s
8	AdaBoost	94%	94%	94%	93%	0.081s

Table 2 explains the results of the proposed model provided by XGBoost, CatBoost, and AdaBoost. The results were 94% accuracy, 94% precision, 94% recall, and 93% F1-score. However, if the author looks at the computing time between the three algorithms. The author prefers AdaBoost with a computing time of only 0.081s. Meanwhile, Catboost and XGboost have computing times of 1.49s and 0.61s. This computing time was chosen because it is important to see the superiority of the ML algorithm in the proposed model. Computation time also affects how fast processing improves accuracy. Therefore, the algorithm chosen is the best, namely AdaBoost.

The AdaBoost algorithm produces the best results compared to the others. This happens because the three of them are one of the ensemble techniques that are good for classification. As for those who say that the assemblage technique can maximize the algorithm in many ways, especially in its approach to detection [44], [45]. And some reasons why these three techniques are better in terms of:
- Bias and variance reduction
- Focus on mistakes
- Regularization
- Categorical feature processing
- Computational efficiency
- Large scale data handling
- Flexible parameter tuning

In the real world, the advantages of computing time and a suitable confusion matrix affect processing time in the business world because time and accuracy are the ones that influence company operations. Several advantages make this technique better. The three ML techniques will be tried to interpret the results in the next chapter.

3.3. Interpretation result

This section will explain the best results proposed by the model from ML and AdaBoost. This is used to interpret ML results, or what it is usually called, to see how ML works, utilizing (10), namely the select K-best formula. These results are presented in Table 3.

Table 3. Select K-best result

No.	Attribute	K-value
1	Last revenue	0.58
2	First revenue	0.196
3	Number employees	0.077
4	Internal collaboration	0.0242
5	Number of online advertisements	0.024
6	Total follows the bazaar	0.026
7	External collaboration	0.022
8	Business team member	0.016
9	Additional investment from external	0.01
10	own trademark rights	0.009
11	covered by the media	0.008
12	Have halal cerified	0.005
13	Using TikTok live	0.004

The results of the proposed model interpretation of the AdaBoost algorithm show that the most dominant features determining an incubator business's success qualifications are last revenue and first revenue, with values of 0.58 and 0.196. However, this contrasts with the results in Figure 2 which states that the number of employees really influences the target. The K-value is interpreted based on the target features in the ML algorithm and looks at the influence on the business's success.

The results of this interpretation only look at which features are more dominant in the classification. In the context of business success, it can be seen that when the first revenue and last revenue increase, this means that it indicates the success of the business, but not only that, another aspect, namely the number of employees, also influences the success of the business.

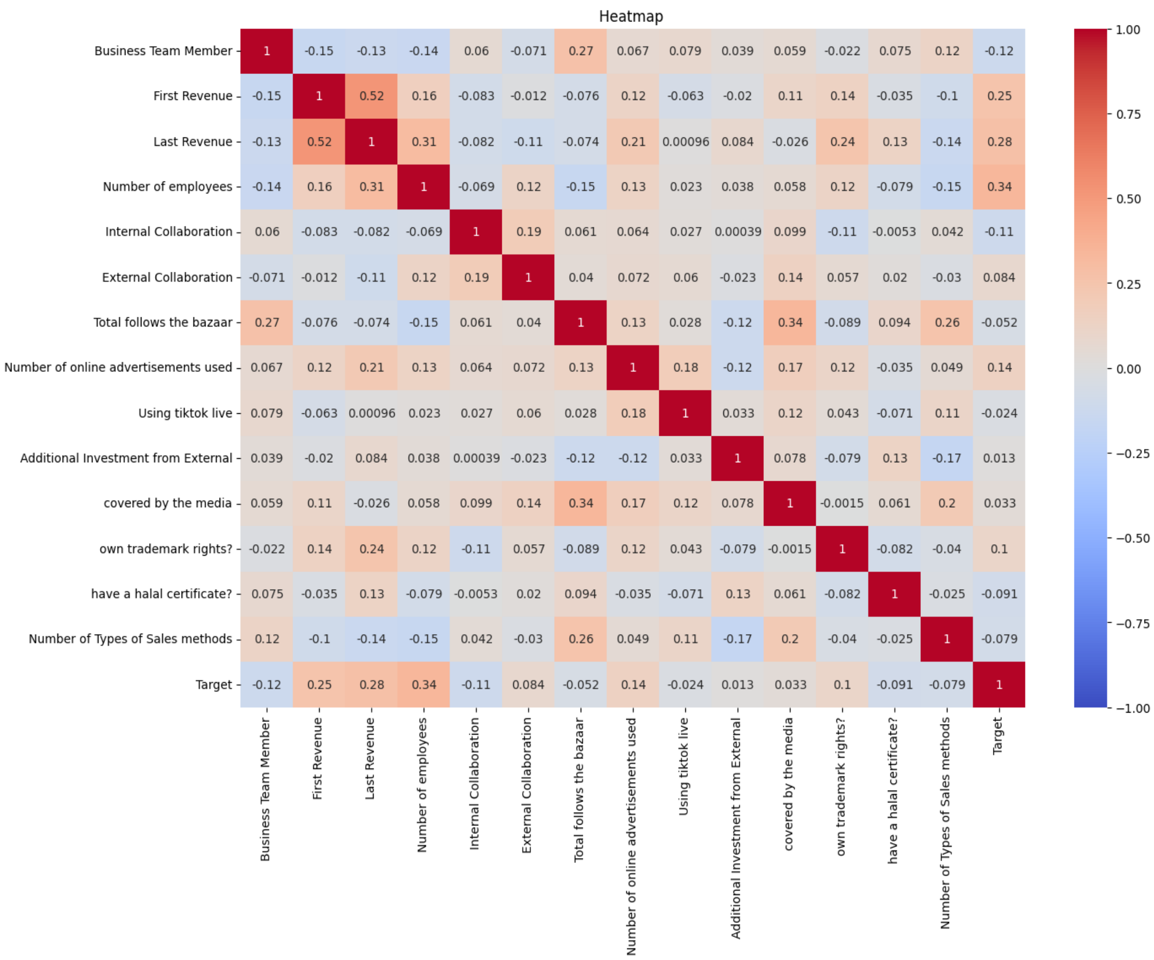

Figure 2. Performance of heatmap correlation incubator datasets

4. CONCLUSION

The results of this research are in search of incubator business success; in this case, the research presents a proposed model architecture for classifying business success; in this case, the proposed model architecture comes from preprocessing, ML models, and model interpretation. The first result is that the best correlation based on the heatmap is the "number of employees," with a correlation value of 0.34. Then, several ML algorithms such as AdaBoost, CatBoost, and XGBoost with accuracy, precision, recall values of 94%, and F1-score of 93%. But if the author looks at the computing time, the best ML is 0.081s, namely AdaBoost. Then, finally, the interpretation results from AdaBoost using select K-best are the best features "last revenue" and "first revenue," with feature K-values of 0.58 and 0.196; these features influence the success of the business. Therefore, in this data, the level of success can be assessed from the correlation results, classification, and interpretation results. Future research, such as using improvements in hybrid ML algorithms and other model interpretations, is important. These improvements could be in the form of ML algorithms that are more immune to data uncertainty, and interpretation of results with wider data.

FUNDING INFORMATION

Bina Nusantara University funds research, especially the RTTO (Research and Technology Transfer Office) unit, which has funded this research for journal registration.

AUTHOR CONTRIBUTIONS STATEMENT

Puji Prabowo provides datasets and data management that can be utilized. Mochammad Haldi Widianto completed the scientific paper by considering revisions from reviewers until completion. Therefore, all authors contributed to this scientific article.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Mochammad Haldi Widianto	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	✓
Puji Prabowo							✓	✓	✓	✓	✓	✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

DATA AVAILABILITY

The data that support the findings of this study are openly available in the mendeley repository at: https://data.mendeley.com/datasets/dd9jv5pw3n/1.

REFERENCES

[1] D. Luketić, "Entrepreneurship education in compulsory education: lessons we learn from recent research practice," *Acta Iadertina*, vol. 18, no. 2, pp. 255–284, Jan. 2022, doi: 10.15291/ai.3607.

[2] G. Elia, A. Margherita, E. Ciavolino, and K. Moustaghfir, "Digital society incubator: combining exponential technology and human potential to build resilient entrepreneurial ecosystems," *Administrative Sciences*, vol. 11, no. 3, p. 96, Sep. 2021, doi: 10.3390/admsci11030096.

[3] M. Guerrero and J. Espinoza-Benavides, "Does entrepreneurship ecosystem influence business re-entries after failure?," *International Entrepreneurship and Management Journal*, vol. 17, no. 1, pp. 211–227, Mar. 2021, doi: 10.1007/s11365-020-00694-7.

[4] R. Mahmood, A. S. M. Zahari, N. Ibrahim, N. F. H. N. Jaafar, and N. M. Yaacob, "The impact of entrepreneur education on business performance," *Asian Journal of University Education*, vol. 16, no. 4, p. 171, Jan. 2021, doi: 10.24191/ajue.v16i4.11947.

[5] Y. Khodkovskaya and O. Malykh, "Project-based learning as an element of the university innovation ecosystem," *SHS Web of Conferences*, vol. 116, p. 00069, Jul. 2021, doi: 10.1051/shsconf/202111600069.

[6] M. Guerrero, D. Urbano, and E. Gajón, "Entrepreneurial university ecosystems and graduates' career patterns: do entrepreneurship education programmes and university business incubators matter?," *Journal of Management Development*, vol. 39, no. 5, pp. 753–775, Apr. 2020, doi: 10.1108/JMD-10-2019-0439.

[7] J. M. Lopes, M. Oliveira, J. Oliveira, M. Sousa, T. Santos, and S. Gomes, "Determinants of the entrepreneurial Influence on academic entrepreneurship—lessons learned from higher education students in Portugal," *Educ Sci (Basel)*, vol. 11, no. 12, p. 771, Nov. 2021, doi: 10.3390/educsci11120771.

[8] T. Ahmed, V. G. R. Chandran, J. E. Klobas, F. Liñán, and P. Kokkalis, "Entrepreneurship education programmes: how learning, inspiration and resources affect intentions for new venture creation in a developing economy," *The International Journal of Management Education*, vol. 18, no. 1, p. 100327, Mar. 2020, doi: 10.1016/j.ijme.2019.100327.

[9] E. M. Swartz, C. B. Scheepers, and T. Toefy, "Women entrepreneurs' opportunity identification of digital platform start-ups: emerging evidence from South Africa," *International Journal of Gender and Entrepreneurship*, vol. 14, no. 3, pp. 352–374, Aug. 2022, doi: 10.1108/IJGE-06-2021-0096.

[10] G. Gabriela, S. E. Cerasela, S. C. Octavian, S. Marcela, and B. C. Larisa, "Creating a sustainable entrepreneurial ecosystem at higher education institution level," *Economic Computation and Economic Cybernetics Studies and Research*, vol. 55, no. 2/2021, pp. 265–280, Jun. 2021, doi: 10.24818/18423264/55.2.21.16.

[11] M. Loganathan and M. H. B. Subrahmanya, “Technological outcome achievements by start-ups at university-based incubators: an empirical analysis in the Indian context,” *Technology Analysis and Strategic Management*, vol. 34, no. 9, pp. 1004–1019, Sep. 2022, doi: 10.1080/09537325.2021.1940921.
[12] K. Lukosiute, S. Jensen, and S. Tanev, “Is joining a business incubator or accelerator always a good thing?,” *Technology Innovation Management Review*, vol. 7, no. 10, pp. 5–15, Jul. 2019, doi: 10.22215/timreview/1251.
[13] A. Rosado-Cubero, T. Freire-Rubio, and A. Hernández, “Entrepreneurship: What matters most,” *Journal of Business Research*, vol. 144, pp. 250–263, May 2022, doi: 10.1016/j.jbusres.2022.01.087.
[14] L. Gozali *et al.*, “Performance factors for successful business incubators in Indonesian public Universities,” *International Journal of Technology*, vol. 11, no. 1, p. 155, Jan. 2020, doi: 10.14716/ijtech.v11i1.2464.
[15] S. Y. Nasution and A. Suzianti, “Development of an incubator performance model based on user perceptions: a PLS-SEM approach,” in *4th Asia Pacific Conference on Research in Industrial and Systems Engineering 2021*, New York, NY, USA: ACM, May 2021, pp. 61–67, doi: 10.1145/3468013.3468310.
[16] R. K. Mavi, H. Gheibdoust, A. A. Khanfar, and N. K. Mavi, “Ranking factors influencing strategic management of university business incubators with ANP,” *Management Decision*, vol. 57, no. 12, pp. 3492–3510, Dec. 2019, doi: 10.1108/MD-06-2018-0688.
[17] M. Madanan, N. A. M. Zulkefli, and N. C. Velayudhan, “Designing a hybrid artificial intelligent clinical decision support system using artificial neural network and artificial bee colony for predicting heart failure rate,” in *2021 International Conference on Computer Communication and Informatics (ICCCI)*, Jan. 2021, pp. 1–7, doi: 10.1109/ICCCI50826.2021.9457007.
[18] R. Sharma, “Artificial intelligence in agriculture: a review,” in *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, May 2021, pp. 937–942, doi: 10.1109/ICICCS51141.2021.9432187.
[19] S. Zel and E. Kongar, “Transforming digital employee experience with artificial intelligence,” in *2020 IEEE / ITU International Conference on Artificial Intelligence for Good (AI4G)*, Sep. 2020, pp. 176–179, doi: 10.1109/AI4G50087.2020.9311088.
[20] S. J. Russell and P. Norvig, *Artificial Intelligence A Modern Approach Third Edition*, Third. 2021.
[21] I. Priyadarshini, S. Sahu, R. Kumar, and D. Taniar, “A machine-learning ensemble model for predicting energy consumption in smart homes,” *Internet of Things*, vol. 20, p. 100636, 2022, doi: 10.1016/j.iot.2022.100636.
[22] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, “Deep reinforcement learning: a brief survey,” *IEEE Signal Process Mag*, vol. 34, no. 6, pp. 26–38, Nov. 2017, doi: 10.1109/MSP.2017.2743240.
[23] M. H. Widianto, V. H. C. Putra, and M. P. Kenardi, “Interpretation and performance machine learning model for detection of brain stroke based on classifier,” in *2023 6th International Conference on Information and Communications Technology (ICOIACT)*, Nov. 2023, pp. 183–188, doi: 10.1109/ICOIACT59844.2023.10455892.
[24] J. Audrey and M. H. Widianto, “Heart attack prediction using machine learning classification models,” in *2023 3rd International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS)*, Aug. 2023, pp. 201–206, doi: 10.1109/ICE3IS59323.2023.10335233.
[25] A. I. Canhoto and F. Clear, “Artificial intelligence and machine learning as business tools: a framework for diagnosing value destruction potential,” *Bus Horiz*, vol. 63, no. 2, pp. 183–193, 2020, doi: 10.1016/j.bushor.2019.11.003.
[26] W. Kratsch, J. Manderscheid, M. Röglinger, and J. Seyfried, “Machine learning in business process monitoring: a comparison of deep learning and classical approaches used for outcome prediction,” *Business and Information Systems Engineering*, vol. 63, no. 3, pp. 261–276, 2021, doi: 10.1007/s12599-020-00645-0.
[27] M. A. Khan *et al.*, “Effective demand forecasting model using business intelligence empowered with machine learning,” *IEEE Access*, vol. 8, pp. 116013–116023, Jun. 2020, doi: 10.1109/ACCESS.2020.3003790.
[28] A. Martínez, C. Schmuck, S. Pereverzyev, C. Pirker, and M. Haltmeier, “A machine learning framework for customer purchase prediction in the non-contractual setting,” *European Journal of Operational Research*, vol. 281, no. 3, pp. 588–596, 2020, doi: 10.1016/j.ejor.2018.04.034.
[29] M. Kraus, S. Feuerriegel, and A. Oztekin, “Deep learning in business analytics and operations research: models, applications and managerial implications,” *European Journal of Operational Research*, vol. 281, no. 3, pp. 628–641, 2020, doi: 10.1016/j.ejor.2019.09.018.
[30] A. J. Nakhal, R. Patriarca, G. Di Gravio, G. Antonioni, and N. Paltrinieri, “Investigating occupational and operational industrial safety data through business intelligence and machine learning,” *Journal of Loss Prevention in the Process Industries*, vol. 73, p. 104608, 2021, doi: 10.1016/j.jlp.2021.104608.
[31] M. H. Widianto, A. A. S. Gunawan, Y. Heryadi, and W. Budiharto, “Interpretation of the dominant features in the prediction results of electrical energy consumption using smart home datasets based on KNN machine learning,” *ICIC Express Letters, Part B: Applications*, vol. 15, no. 4, 2024.
[32] M. H. widianto, “Incubator datasets,” Dataset. Accessed: Jun. 12, 2024. [Online]. Available: https://docs.google.com/spreadsheets/d/1FVoCgDBremBb0RScizy7wF14u1-aASVk/edit?usp=sharing&ouid=100964609793302912523&rtpof=true&sd=true
[33] E. Saccenti, M. H. W. B. Hendriks, and A. K. Smilde, “Corruption of the pearson correlation coefficient by measurement error and its estimation, bias, and correction under different error models,” *Sci Rep*, vol. 10, no. 1, p. 438, 2020, doi: 10.1038/s41598-019-57247-4.
[34] D. Menufandu, R. Fitriani, and E. Sumarminingsih, “estimation of maximum likelihood weighted logistic regression using genetic algorithm (case study: individual work status in malang city),” *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 17, pp. 487–494, Apr. 2023, doi: 10.30598/barekengvol17iss1pp0487-0494.
[35] A. M. Abdulazeez, “Machine learning applications based on SVM classification: a review,” May 2021, doi: 10.48161/qaj.v1n2a50.
[36] L. Cañete-Sifuentes, R. Monroy, and M. A. Medina-Pérez, “A review and experimental comparison of multivariate decision trees,” *IEEE Access*, vol. 9, pp. 110451–110479, 2021, doi: 10.1109/ACCESS.2021.3102239.
[37] M. Jeong, J. Nam, and B. C. Ko, “Lightweight multilayer random forests for monitoring driver emotional status,” *IEEE Access*, vol. 8, pp. 60344–60354, 2020, doi: 10.1109/ACCESS.2020.2983202.
[38] G. Mustafa *et al.*, “Hyperspectral reflectance proxies to diagnose in-field fusarium head blight in wheat with machine learning,” *Remote Sens (Basel)*, vol. 14, no. 12, p. 2784, 2022, doi: 10.3390/rs14122784.
[39] W. Chang, X. Wang, J. Yang, and T. Qin, “An improved catboost-based classification model for ecological suitability of blueberries,” *Sensors*, vol. 23, p. 1811, Feb. 2023, doi: 10.3390/s23041811.
[40] X. Gao and G. Li, “A KNN model based on manhattan distance to identify the SNARE proteins,” *IEEE Access*, vol. 8, pp. 112922–112931, 2020, doi: 10.1109/ACCESS.2020.3003086.
[41] J. Li, G. Li, C. Hai, and M. Guo, “Transformer fault diagnosis based on multi-class adaboost algorithm,” *IEEE Access*, vol. 10, pp. 1522–1532, 2022, doi: 10.1109/ACCESS.2021.3135467.

[42] J. Zheng, D. Lin, Z. Gao, S. Wang, M. He, and J. Fan, "Deep learning assisted efficient adaboost algorithm for breast cancer detection and early diagnosis," *IEEE Access*, vol. 8, pp. 96946–96954, 2020, doi: 10.1109/ACCESS.2020.2993536.
[43] K. Chtouki, M. Rhanoui, M. Mikram, S. Yousfi, and K. Amazian, "Supervised machine learning for breast cancer risk factors analysis and survival prediction," in *International Conference On Big Data and Internet of Things*, Springer, 2022, pp. 59–71, doi: 10.1186/s12889-017-4914-3.
[44] D. Zhang and Y. Gong, "The comparison of LightGBM and XGBoost coupling factor analysis and prediagnosis of acute liver failure," *IEEE Access*, vol. 8, pp. 220990–221003, 2020, doi: 10.1109/ACCESS.2020.3042848.
[45] Y. Niu, "Walmart sales forecasting using XGBoost algorithm and feature engineering," in *2020 International Conference on Big Data, Artificial Intelligence and Software Engineering (ICBASE)*, Oct. 2020, pp. 458–461, doi: 10.1109/ICBASE51474.2020.00103.

BIOGRAPHIES OF AUTHORS

Mochammad Haldi Widianto one of the lecturers at BINUS, and placed in Bandung. He completed his bachelor's in telecommunications and worked at the ministry. Then he continued his master's in the field of electro-communication, now, he is a Ph.D. candidate at the Doctor of Computer Science of Bina Nusantara University. He can be contacted at email: mochamad.widianto@binus.ac.id.

Puji Prabowo is a Creativepreneur lecturer at BINUS, and also a sociopreneur who has been working for 12 years. He is also a coach and business incubator section head at Binus University. He is enthusiastic about the entrepreneurial ecosystem, entrepreneurial skills, innovation, creative ideas, and business development. He graduated with an MBA at ITB. He can be contacted at email: puji.prabowo@binus.ac.id.