Deep Learning to Predict Start-Up Business Success

Lobna Hsairi
Department of Information System and Technology-CCSE, University of Jeddah, Jeddah, Saudi Arabia

***Abstract*—Over the past few decades, there has been rapid growth in the formation of new start-ups around the world. Thus, it is an important and challenging task to understand what makes start-ups successful and to predict their success. Several reasons are responsible for the success and failure of a start-up, including bad management, lack of funds, etc. This work aims to create a predictive model for start-ups based on many key factors involved in the early stages of a start-up's life. Current research on predicting success mainly focuses on financial data such as ROI, revenue, etc. Therefore, in this paper, a different approach is proposed by first investigating other non-financial factors affecting start-up success and failure. Second, the adoption of an algorithm that has not been used much in predicting start-up success, which is Convolutional Neural Network (CNN). The dataset was acquired from Kaggle. The final model was reached through a series of four experiments to determine which model predicts better. The final model was implemented using a CNN with an average accuracy of 82%, an average loss of 0.4, an average 0.9 recall and an average 0.9 precision.**

***Keywords*—Deep learning; Convolutional Neural Network (CNN); prediction; start-up business**

I. Introduction

Start-ups have become an important topic in the economic policies of all developed and emerging economies around the world, not just by being a driver of economic prosperity and wealth but also because of their major impact on innovation and technological development. Start-ups are booming everywhere as more colleges, governments, and private companies invest and stimulate people to pursue their ideas throughout these ventures. Start-ups are raising millions with ease. Examples like Uber and Airbnb are changing societies in such impactful ways that regulation had to be created to keep pace with a new reality [1]. Start-ups are having such an impact that ultimately, it becomes every investor's ambition to be part of a large acquisition, such as Facebook acquiring WhatsApp for nineteen billion dollars, which allowed Sequoia (a Venture Capital fund) to have a 50x Return On Investment (ROI) [2]. But there is a catch: start-ups are companies with an estimated 90% probability of failure [3], which means a lot of investments without proper returns. According to SPA load [3], 90% of start-ups launched in 2023 are failed start-up. Entrepreneurs who experience failure are numerous, and it's important to identify the factors that lead to failure and success too. Those factors, when shared and explored, will assist potential entrepreneurs in the ecosystem in designing their path to success. The consequences of entrepreneurial failure [4] extend beyond the start-up and have an impact on employment and the economy. The ability to predict success is an invaluable competitive advantage for various parties, such as venture capitalists on the hunt for investments since first-rate targets are those who have the potential for growing rapidly soon, which ultimately allows investors to be one step ahead of the competition [5]. The prediction of start-up success will help investors get an idea of whether investing in a start-up will be successful or not. Recently, machine learning algorithms have been considered an effective approach to predicting start-up success. Furthermore, deep learning shows significant promise in the business domain in general and in start-up prediction in particular [6]. Machine learning and deep learning use algorithms to create models that reveal patterns from data, allowing businesses to gain insights and make predictions to enhance operations, better understand customers, and solve other issues. There are numerous algorithms to choose from. These help predict the outcomes of a start-up will be profitable or not.

There are few studies which are performed to understand the reasons for the success of a start-up company [7]. These studies use various criteria of success, varying from predicting funding or follow-up funding, meaning most of the focus is on financial data. These start-ups usually lack enough financial data on their historical performance [8]. Therefore, in this paper, non-financial measures of performance for predicting the probability of start-up success were used. Most existing research is based on machine learning techniques, such as random forest models, Support Vector Machines, and logistic regression [9] (as the most common predictive tools), few researches turn the light to explore deep learning techniques [6]. There is still room for different types of approaches, such as Artificial Neural Network and Convolutional Neural Network, which are used in this research.

In this paper, to predict start-up success, which helps to sustain and grow new businesses, different approach is proposed by first, investigating other non-financial factors affecting start up success and failure. Second, the adoption of deep learning algorithms that have not been used much in predicting start-up success, which are Artificial Neural Network (ANN) and Convolutional Neural Network (CNN). In addition to that, the previous solutions are dependent on memory-based algorithms such as K nearest neighbors [10], the proposed solution will depend on processor-based algorithms which increase the learning, time and velocity of the model.

The reminder of this paper is organized as follows: Section II reviews related work on prediction models of start-up success. Section III describes the dataset, preprocessing techniques, and models used, along with training and hyperparameter tuning approaches. Section IV presents findings, including performance metrics and models' comparison. Section V presents a brief discussion of proposed models' results. Section VI summarizes the main findings and suggests future research directions.

II. Related Work

A number of studies have been conducted to explore the use of machine learning to understand the reasons for the success of a start-up company. These researches may aid in the selection and utilization of machine learning techniques for the prediction of start-up success, primarily based on funding factors at various stages. For instance, C. Pan et al. [11] used three classification algorithms to predict the probability of success of a start-up (Logistic Regression, Random Forest and K-Nearest Neighbors). They conclude that if the investor has a limited investment budget and wants to maximize the proportion of success among its portfolio, it would be better to choose Random Forests model instead of KNN model. However, if the investor has a lot of investment money and wants to maximize the number of successful companies it could invest in, it would be better to choose KNN model. Thus, the model selection to make the best prediction is solely based on the budget. Additionally, another study by B.Yankov et al. [12] presented a quantitative investigation and creation of success prediction models based on the answers to the challenges and questions that start-up companies face. The questionnaire is based on the new venture success prediction model proposed by Yankov [12]. 15 algorithms were used; the most accurate model is J48. Results show that the main challenge Bulgarian high-tech start-ups face is getting adequately funded at the initial stages of the business. Furthermore, D. Fidder [13] identified significant predictors of startup success, namely, technology and B2B/B2C; and he built two models: the first predicts whether a start-up will have a profitable exit for investors, and the second predicts whether the start-up will be able to attract more than 1 million Euros in funding. For the first model, a logistic regression model was built. Where, in the second one, the author built a linear model. On the other hand, T. Żbikowski et al. [14] compared three algorithms: logistic regression, support vector machine, and the gradient boosting classifier. They achieved promising results in terms of precision, recall, and F1 scores for the best model the gradient boosting classifier. The top three important features are the country and region that the company operates in and the company’s industry. In addition, I. Afolabi et al. [15] used both Naïve Bayes algorithm and J48 algorithm for prediction. The result reveals that all the models built for prediction gave a percentage accuracy of above 50%. Other algorithms need to be applied to enhance accuracy. Moreover, S.H. Arshe et al. [16] implemented eight different algorithms and analyzed the percentage of score of them. The deciding factor in the selected data set is the "status" column, which had two values: acquired and closed. The used algorithms are decision tree, Random forest, K-Nearest Neighbor, MLP, Naïve byes, logistic regression and SGD. After using these algorithms, they obtained different success rate scores for each one. The two best algorithms, according to the success rate, are decision tree and Random Forest. In the same way, Ü. Cemre et al. [17] implemented a total of six different models to predict startup success. Using goodness-of-fit measures applicable to each model case, the best models selected were the ensemble methods, random forest and extreme gradient boosting. The top variables in these models are last funding to date, first funding lag and company age. Likewise, V. Shah et al. [18] created a predictive model to predict startup firm success. The key factors used to build the model are seed funding, series funding, rounds of funding, time to get seed funding, valuation after each round of funding, number of milestones, average time taken to achieve each milestone, average time taken to achieve funding, region, degree, university, burn rate, total funding, and category_code. The model implemented using logistic regression reached good accuracy. Nevertheless, T. Kalendová [19] applied four machine learning classification methods (Logistic regression, Random forest, XGB, SVM) to predict startups’ success with a focus on the needs of the venture capital industry. The models’ results have shown the potential of using machine learning algorithms to predict the success of venture capital-backed start-ups in the predefined time period. The Random Forest model proved to be the best predictor from the set, outperforming other methods by having the highest scores of selected performance measures. Also, the rest of the algorithms showed high performance scores, especially extreme gradient boosting.

From the overview above, there is already a lot of knowledge about the most significant predictors of start-up success. Researchers use various criteria for success, varying from predicting funding or follow-up funding, meaning, most of the focus is on financial data. However, these start-ups usually lack enough financial data on their historical performance. Hence, in this paper, non-financial measures of performance for predicting the probability of a start-up success were used. Most of researches are based on machine learning techniques, such as random forest, support vector machines, and logistic regression (as the most common predictive algorithms), few researches turn the light to explore deep learning techniques [6]. There is still room for different types of approaches, such as ANN and CNN which were used in the proposed models.

III. Method

A. DataSet

The dataset was collected from Kaggle website[1]. It contains detailed information about start-up companies. The dataset consists of 116 columns and 473 records. The reason beyond the choice of a dataset with a small number of records; is because unlike other explored datasets, the selected one contains columns that could drive us to make comprehensive and useful insights. It comprises numerical and nominal data, the target factor in the selected dataset is "Dependent-Company Status" column which deliver the current operating status of the start-up and has two values, success and fail. To distinguish the key factors picked out to build models the data exploration is needed. One of the categorical feature is the “industry field of a company”. Startups are categorized into 35 industry fields such as analytics, media, finance etc. An investigation of the company Statues per industry fields allows to conclude that the most common field with the greatest number of successful companies is ‘analytics’. Of the top 10 industries in analysis, ‘Healthcare’ start-ups have a slower average of overall age of success. Meaning, industries such as healthcare would take much time to success unlike ‘Market Research’ start-ups which have a faster age of success. One more fact that should be

[1]https://www.kaggle.com/datasets/ajaygoíkaí/staítup-analysis.

considered to understand what influence the success and failure of a start-up is the prior experience of the founding team (Average years of experience for founder and co-founder, Number of Co-founders, Controversial history of founder or co-founder, etc.). Founding teams with high average years of experience are most likely to form a successful start-up. Another categorical feature is the 'marketplace of a start-up'. A start-up can target two types of marketplace. First, a global market place which is not limited to specific geographic locations but rather involves the exchange of products, services, and employees anywhere in the world. Second, a local marketplace which target and reach potential customers within a certain distance of their business's location. An exploration of the number of successful and failure start-up per marketplace allows for the deduction that targeting a global marketplace could affect the probability of start-up success. It is also important to understand how different business model of a start-up influence its probability of success. B2B (Business to Business) and B2C (business-to-consumer) are distinguished. After inspection, B2B start-ups are more successful. Following examination of different features, the key factors settle on to build models are: age of company in years, industry of company (analytics, e-commerce, advertising, marketing, media etc.), number of investors, number of co-founders, number of advisors, worked in top companies, consulting experience, focus on private or public data, cloud or platform based service/product, local or global player, linear or non-linear business model, disruptiveness of technology, number of direct competitors. In order to more understand the factors affecting start-ups, the investigation of how long a start-up will be considered a failure is needed. Fig. 1 shows that all the failed companies have a normal distribution with 0.4 skew and four years' average age.

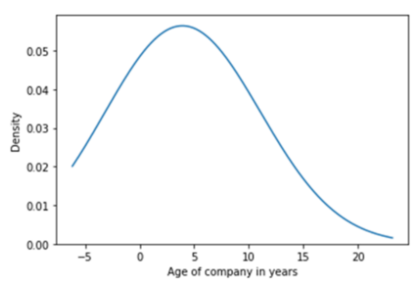

Fig. 1. Average age distribution of failed companies.

B. Data Preparation and Preprocessing

The measurements of success in business are hard to define statistically. There are no correlations between the numerical attributes, even with splitting the dataset classes. It's a subjective and challenging domain, but our mission and role are to help and solve business problems, especially the problems that are associated with finance and decision-making. Since the selected dataset is extracted from Kaggle website, the chances of finding flaws in them are high. There are many problems in the selected dataset, such as noisy data, missing and null values, and duplicated data [20]. Data cleaning is the process of fixing or removing incorrect, duplicated, or incomplete data within a dataset. The records that contain more than 25% null values were removed. Null values are a significant problem in machine learning and deep learning. Several methods were used to deal with them in the adopted dataset. The dataset had 'No info' which was not recognized as a null value. Therefore, 'No info' was replaced with np.nan to let python recognize it as a null value. Also, columns that contain more than 5% null were dropped. Noisy data is detected and removed. The dataset had three types of noisy data, including a column with incorrect entries. To solve it, the column was dropped, and value_counts() was used to reduce the number of repeating words in the column with many items. So, the companies that have "more than one industry" were replaced by "Multi-industry". Data transformation is the process of changing the format. All 'Yes' was changed to 1 and all 'No' to 0; success was changed to 1 and failed to 0. After preprocessing, the dataset consists of 104 columns and 413 records.

C. Model Building

As stated before, ANN algorithm was chosen to build the model. However, after conducting several experiments, we discovered that ANN is not the optimal algorithm in the adopted use case due to its poor results, such as high loss, the model being biased to one class. After evaluating the experiments, we concluded that Convolutional Neural Network (CNN or ConvNet) is a more suitable algorithm to build the model, considering its ability to perform operations that alter the data with the intent of learning features specific to the data [19]. The model is built using Python. For the model experiment setup, the dataset was already prepared by cleaning and preprocessing it, and the necessary columns were selected. Additionally, all the required libraries and extensions for building the model were set up. The dataset was then split using stratified k-fold to overcome the imbalance in the dataset.

A CNN or ConvNet is network architecture for deep learning [21] that learns directly from data. A CNN is composed of an input layer, an output layer, and many hidden layers in between. These layers perform operations that alter the data with the intent of learning features specific to the data. There are common CNN layers were used in building the model, such as Convolution layer (which is the most important component of any CNN architecture). It contains a set of convolutional kernels (also called filters), which gets convolved with the n-dimensional metrics to generate an output feature map [20], activation function (the main task of any activation function in any neural network based model is to map the input to the output), two types of activation function were used in the proposed model: The sigmoid activation function as shown in Eq. (1) and The Rectifier Linear Unit (ReLU) as shown in Eq. (2).

$$f(x)_{sigm}=1/1+e^{-x} \tag{1}$$

$$f(x)_{ReLU}=max(0,x) \tag{2}$$

The explained CNN [22] components are the fundamental component for any model. However, for the proposed model, additional layers are added to handle the requirements, which are a dense layer (containing densely connected neurons) and a dropout layer (overfitting is a serious problem faced by the model. Dropout is a technique for addressing this problem). The key idea is to randomly drop units (along with their connections) from the neural network during training [23]. This prevents units from co-adapting too much. It significantly reduces overfitting and gives major improvements over other regularization methods. The final step in building the model is compilation, during which the "Adam" optimizer is used. To train the proposed model with the given inputs, it is fitted for 30 epochs. Table I illustrates the hyperparameters of different layers.

TABLE I. MODEL'S LAYER AND HYPERPARAMETERS

Layers	Hyperparameters
Conv1D	filters=32, kernel_size=3, activation='relu',input_shape=[66,1]
Conv1D	filters=32, kernel_size=1, activation=sigmoid
Dense	units=50, activation= relu
Dense	units=2, activation='sigmoid'
Dropout	(0.5)

IV. EXPERIMENTS RESULTS

Four experiments were conducted to reach the final model. These experiments will be depicted in the following subsections along with its performance and problems. For the first three experiments, the dataset was split into two main segments: a training set and a test set. The training set was further divided into training and validation sets using an 80:20 approach. For the fourth experiment, the dataset was split into train and test subsets in a stratified fashion. Cross-validation on the training set to ensure that the model does not overfit to the validation set. Cross-validation is recommended in hyperparameter tuning to reduce the problem of selection bias and overfitting. Furthermore, several common metrics are used to obtain valuable information about algorithm performance, such as: Learning curve, Confusion matrix, Accuracy, Loss, Precision and Recall.

A. *Experiment 1: ANN Model*

Experiments are stared by using ANN, as the adopted algorithm. After building the model and evaluate it. The confusion matrix shown in Fig. 2 depicts that the model is highly biased towards success. Moreover, the learning curve shown in Fig. 3 illustrates that the model is overfitting because the validation loss has a lot of vibration. Furthermore, the evaluation metrics in Table II conclude that the results are unacceptable, thus the algorithm needs to be changed for better feature extraction. So, another experiment is conducted using CNN to enhance the model.

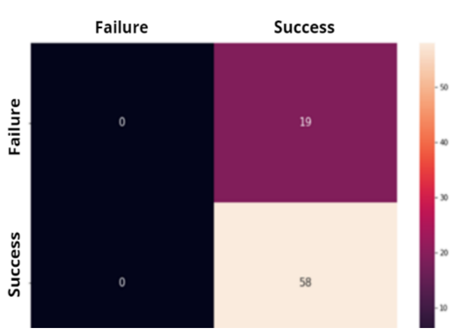

Fig. 2. Confusion matrix for experiment 1.

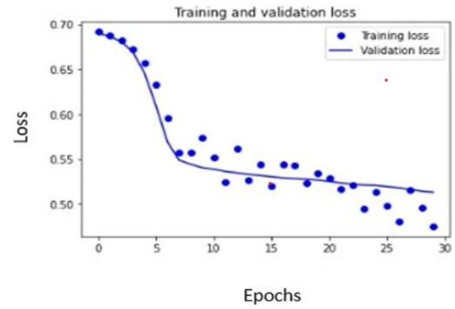

Fig. 3. Learning curve (loss) for experiment 1.

TABLE II. EVALUATION METRICS FOR EXPERIMENT 1

Accuracy	75%
Loss	0.5
Precision	0.7
Recall	0.7

B. *Experiment 2: CNN Model*

In the first experiment, the ANN model is biased towards start-up success because the ANN is fully connected. Thus, the model needs to be changed and build a CNN model and try different feature extraction to make the results more optimal. For the confusion matrix, as it shown in Fig. 4, there is an improvement in the result compared to the ANN model, and in the learning curve in Fig. 5 and Table III the accuracy reach 83% for the CNN model, which is more than the expected accuracy, but in Fig. 6, the loss shows that the model is overfitted. So, additional experiments were made to overcome the problems in this experiment.

Fig. 4. Confusion matrix for experiment 2.

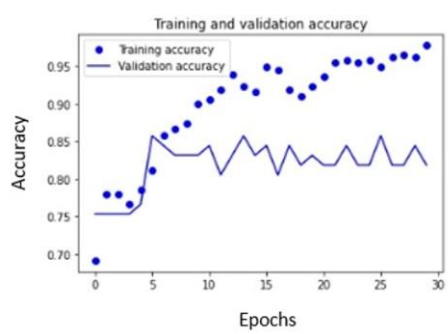

Fig. 5. Learning curve (accuracy) for experiment 2.

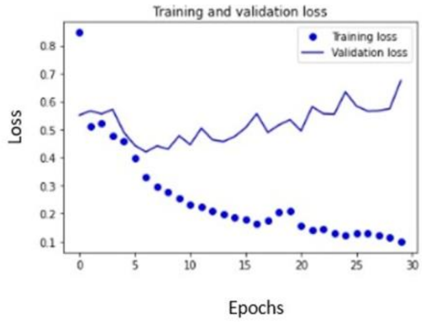

Fig. 6. Learning curve (loss) for experiment 2.

TABLE III. EVALUATION METRICS FOR EXPERIMENT 2

Accuracy	83%
Loss	0.5
Precision	0.83
Recall	0.79

C. Experiment 3: CNN Model with Dropout Layer

As investigated in the previous experiment, the model is suffering from overfitting. In Experiment 3, to overcome this problem, a dropout layer was added. The key idea of dropout is to randomly drop units (along with their connections) from the neural network during training. This prevents units from co adapting too much during training, Table IV shows the evaluation metrics for this experiment. The accuracy and loss values indicate that the model has high accuracy and produces correct outputs. Another way to evaluate the performance of the model is learning curve. In this experiment, the model has a good fit as Fig. 7 shows. A good fit is identified by a training and validation loss that decreases to a point of stability with a minimal gap between the two final loss values. Overall, the model evaluation metrics along with confusion matrix in Fig. 8 indicate that this experiment is good enough, but there is a room for enhancement, so a fourth experiment was elaborated to make the model more accurate and robust.

TABLE IV. EVALUATION METRICS FOR EXPERIMENT 3

Accuracy	83%
Loss	0.7
Precision	0.83
Recall	0.8

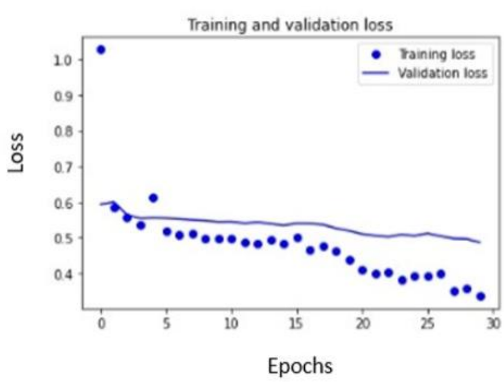

Fig. 7. Learning curve (loss) for experiment.

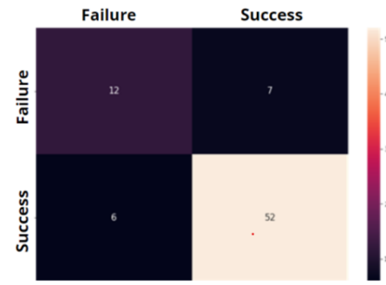

Fig. 8. Confusion matrix for experiment 3.

D. Experiment 4: CNN Model with K-fold

As depicted in Experiment 3, the issue of overfitting was controlled. However, the model can still be enhanced. One way to improve the performance of the model is by using k-fold. Data splitting process can be done more effectively with k-fold cross-validation. The main intention of using k-fold is to develop a more generalized model that can perform well on unseen data. For selecting an appropriate value of k, multiple values are tried until we came to a conclusion that 5 is the optimal value of k for the proposed model. The evaluation metrics in Table V show that after changing the splitting into k-fold, the loss has decreased, which means that the model is doing a good job of predicting the expected outcome: success or failure of start-up business. While the precision and recall increased, which means that the model is performing well.

TABLE V. EVALUATION METRICS FOR EXPERIMENT 4

Avg(Accuracy)	82%
Avg(Loss)	0.4
Avg(Precision)	0.9
Avg(Recall)	0.9

V. DISCUSSION

Trying to reach the optimal model is a series of steps and experiments, which in our case were four experiments (see Table VI). As planned, ANN was used as the starting algorithm. However, it was not optimal for the adopted use case, start-up business success. Which led us to change the algorithm to CNN [20]. Different enhancements were added, such as CNN with a Dropout layer and CNN with k-fold. Among all four experiments, the best one was Experiment 4: CNN with k-fold and dropout layer, due to its reliable performance. It achieved an average 82% of accuracy, an average 0.4 loss, an average 0.9 recall and an average 0.9 precision and the learning curve showed that the model has a good fit. The overall evaluation of the model indicated that the model considered suitable for use and can predict the start-up business success or failure with high accuracy.

TABLE VI. RESULTS OF EXPERIMENTS

	Accuracy	Loss	Precision	Recall
ANN	75%	0.5	0.7	0.7
CNN	83%	0.7	0.8	0.8
CNN with Dropout layer	83%	0.5	0.8	0.8
CNN with K-fold	**Avg(82%)**	**Avg(0.4)**	**Avg(0.9)**	**Avg(0.9)**

VI. Conclusion and Future Work

Predicting start-up success is a challenging task, but it is crucial to many public and private stakeholders who shape economics, make funding and investment decisions, and found companies. Intuitively, the task becomes easier as the company matures and tests its product-market fit. In this article, a deep learning approach is proposed for predicting start-up success at the seed stage, narrowing down the set of features to geographical, demographic, and basic information about the companies. Unlike previous works, financial information is not used. To predict start-up success, deep learning models are built and the performance of two algorithms, ANN and CNN is compared. Four experiments are used: ANN, CNN, CNN with Dropout Layer, and CNN with K-fold. The major problems faced in the models are high loss and overfitting, which are controlled in the last experiment. Experiment 4, CNN with k-fold and Dropout layer, outperforms the others. According to its performance, it achieved an average 0.4 loss, an average 0.9 recall and precision, and the learning curve indicates that the model has a good fit and an accuracy of 82. According to the overall evaluation, the model was deemed suitable for use. Several recommendations for future research can be made. First, gather more data and more completed data. This can be done by accessing different databases and combining them. Second, apply more sophisticated machine learning and deep learning techniques to the data. This allows researchers to make more precise estimations. Researchers should attempt to combine models that include both performance indicators and success metrics, which will lead to more accurate predictions.

References

[1] B. Stone. "The Upstarts: How Uber, Airbnb, and the Killer Companies of the New Silicon Valley Are Changing the World", Large Print. Little, Brown and Company, 2017.

[2] A. L. Deutsch. "WhatsApp: The Best Meta Purchase Ever?". Investopedia. https://www.investopedia.com/articles/investing/032515/whatsapp-best-facebook-purchase-ever.asp

[3] M. Babych. "Startups’ Success and Failure Rate in 2023: In-Depth Overview". SPDLOAD. https://spdload.com/blog/startup-success-rate/

[4] C. Marco, G. Valentina, P. Guido and R. Mariangela. "Startups' Roads to Failure". Sustainability, vol.10, no. 7, pp 2346, 2018.

[5] J. Kim, H. Kim and Y. Geum. "How to succeed in the market? Predicting startup success using a machine learning approach". Technological Forecasting and Social Change, vol. 193, pp. 122614, May, 2023.

[6] A. Mishra, D.S. Jat and D.K. Mishra, (2023). "An Experimental Study of Machine Learning Algorithms for Predicting Start-Up Success". In: Nagar, A.K., Singh Jat, D., Mishra, D.K., Joshi, A. (eds) Intelligent Sustainable Systems. Lecture Notes in Networks and Systems, vol. 578, pp 813--825. Springer, Singapore, 2023.

[7] O. Kofanov and O. Zozul'ov. "Successful Development of Startups as a Global Trend of Innovative Socio-Economic Transformations". International and Multidisciplinary Journal of Social Science, vol. 7, no. 2, pp. 191-217, 2018.

[8] Y. Liu, Q. Zeng, B. Li, L. Ma and J. Ordieres-Meré. "Anticipating financial distress of high-tech startups in the European Union: A machine learning approach for imbalanced samples". Journal of Forcasting, vol. 41, no.6, pp. 1131-1155, 2022.

[9] F. Rodrigues, FA. Rodrigues and TVR. Rodrigues. "Machine learning models for predicting success of startups". Revista de Gestao E Projetos. Vol.12, Issue2, pp. 28-55, 2021.

[10] E. Vasquez, J. Santisteban, and D. Mauricio. "Predicting the Success of a Startup in Information Technology Through Machine Learning". International Journal of Information Technology and Web Engineering, vol. 18. No.1, pp. 1-17, 2023.

[11] C. Pan, Y. Gao and Y. Luo, "Machine Learning Prediction of Companies’ Business Success". Stanford University, CA. 2018.

[12] K. Haralampiev, B. Yankov and P. Ruskov. "Models and Tools for Technology Start-Up Companies Success Analysis". Economic Alternatives, pp. 15-24, 2014.

[13] D. Fidder, "Finding The Most Significant Predictors of Startup Success with Machine Learning". M.S. thesis, Dept. of Mathematics and Computer Science. Eindhoven University of Technology and Tilburg University, Eindhoven, The Netherlands. 2022.

[14] K. Å»bikowski and P. Antosiuk, "A machine learning, bias-free approach for predicting business success using Crunchbase data", Information Processing & Management, vol. 58. No. 4, pp. 102555, 2021.

[15] I. Afolabi, T. Cordelia Ifunaya and F. G. Ojo and Chinonye Moses. "A Model for Business Success Prediction using Machine Learning Algorithms". Journal of Physics: Conference Series. Vol. 1299. No.1, pp. 012050, 2019.

[16] F. R. Akash, M. T. Zoayed and S. H. Arshe, "Startup Success prediction using Classification Algorithms", Dept. of Computer Science and Engineering. East West University, Dhaka, Bangladesh. 2022.

[17] Ü. Cemre and C. Ioana. "A Machine Learning Approach Towards Startup Success". IRTG 1792 Discussion Paper, No. 2019-022.

[18] V. Shah. "Predicting the success of a startup company". Paper 3878-2019.

[19] T. Kalendová. "A Machine Learning Approach to Startup Success Prediction in the Context of Venture Capital Industry". M.S. thesis, University of Economics, Prague, 2020.

[20] K. Goyal. "Data preprocessing in Machine Learning: 7 easy steps to follow". upGrad blog. https://www.upgrad.com/blog/data-preprocessing-in- machine-learning.

[21] Y. Slimani and R. Hedjam, “A Hybrid Metaheuristic and Deep Learning Approach for Change Detection in Remote Sensing Data”, Eng. Technol. Appl. Sci. Res., vol. 12, no. 5, pp. 9351–9356, Oct. 2022.

[22] M. M. H. Milu, M. A. Rahman, M. A. Rashid, A. Kuwana, and H. Kobayashi, “Improvement of Classification Accuracy of Four-Class Voluntary-Imagery fNIRS Signals using Convolutional Neural Networks”, Eng. Technol. Appl. Sci. Res., vol. 13, no. 2, pp. 10425–10431, Apr. 2023.

[23] D. K. Suker, “Deep Learning CNN for the Prediction of Grain Orientations on EBSD Patterns of AA5083 Alloy”, Eng. Technol. Appl. Sci. Res., vol. 12, no. 2, pp. 8393–8401, Apr. 2022.