

## Optimized Deep Learning Model to Predict Business Bankruptcy

Dino Quinteros-Navarro<sup>\*ID</sup>, Ciro Rodríguez<sup>ID</sup>

Faculty of Systems Engineering, Universidad Nacional Mayor de San Marcos, Lima 15081, Peru

Corresponding Author Email: [dino.quinteros@unmsm.edu.pe](mailto:dino.quinteros@unmsm.edu.pe)



Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.290510>

### ABSTRACT

**Received:** 14 March 2024

**Revised:** 5 August 2024

**Accepted:** 14 September 2024

**Available online:** 24 October 2024

#### **Keywords:**

*deep learning, model, predict, business, bankruptcy*

This article studies the ideal and optimized way to predict a company's bankruptcy according to the internal and external environment. Likewise, prognostic factors were identified through machine learning models that are ideal for this type of case. In this context, the choice of the ideal model is the basis for making improvements and updates that allow the optimization of a specific model. For this reason, the optimization of the model is considered a specialized process that allows the precise identification of the factors that lead to the bankruptcy of companies and the identification of the necessary correlations between relevant variables. In this sense, past works on similar contexts were considered for the present study, the analysis method where the methodology is required, and three stages: Data exploration, selection of models, and implementation. Likewise, to determine the results, training was considered to obtain results from the model with optimized characteristics, company bankruptcy factors and correlations, and finally the discussion of results and conclusions was specified.

## 1. INTRODUCTION

Companies find themselves using business tools and strategies with emerging technology in a more recurrent way in a very aggressive business environment. However, under these conditions, the generation of value from products and services is the greatest aspiration of the business. Likewise, one of the main strategies applied is the forecasting of situations and events that could compromise subsistence or growth activities. Therefore, it is important to develop and implement technological strategies in crucial activities that determine the reason for the business. In this context, artificial intelligence is a science that has been making proposals and initiatives to companies that are mainly in subsistence or that require greater efficiency and optimization of their processes. The results required by a company are aligned with institutional goals and strategic plans.

Alsheref et al. [1] considered that companies have accurate and critical targeting activities due to lack of consistency in data and misinformation about financial aspects. As a result, micro, small and medium-sized enterprises often do not use traditional financial channels. In this sense, financial difficulties have a negative impact on the development of business health and sustainability of these companies. Similarly, Brenes et al. [2] note that difficulties in credit evaluations for companies lead to higher interest rates and distort the nature of financing. Additionally, most of these companies are at the subsistence stage and are prone to business bankruptcy. In this sense, they cannot access financial services, and an appropriate credit evaluation. For this reason, at the macroeconomic level, most startups do not constitute the important engine for local economic growth,

especially when facing crises and possible bankruptcies.

Likewise, financial institutions play an important social role in facilitating access to credit and facing the risks associated with lending money. However, there are several difficulties in financing companies and identifying companies that are at risk of bankruptcy. Credit rating and bankruptcy models have been used by financial institutions to assess creditworthiness and identify potential companies that will lose their status as activity [2]. Companies have difficulties in accessing financing and opening up to other markets, motivating them to generate subsistence activities, which over time can be considered a potential bankruptcy [3].

Calabrese [4] defines the credit risk associated with company loan financing as the probability of defaulting on derivative loans, which can lead to a negative credit rating. This rating, in turn, reflects the potential for business bankruptcy. In this sense, data from financial statements are presented to predict the credit and bankruptcy risk of companies. However, behavioral and performance data allow them to qualify for loans through covenants, but once the loan is obtained, they are likely to be unable to meet the payment obligation. According to the study by Cantarero et al. [5], companies that have been in operation for years and those that are new (some with only a few months of start-up) face a highly competitive, complex, ambiguous, volatile and uncertain market, which is not only local or national, but is directly related to the international market through products, services or price.

Tran et al. [6] emphasize that performance measurement is fundamental in a control system and argue that it is essential to measure the financial aspects of companies both before and after they obtain a loan. They also consider that the rating of

financial performance originates the profitability of a company and its continuity in the long term. Muthukumaran and Hariharanath [7] observe that both established companies and new enterprises (some only months old) encounter a highly competitive, complex, ambiguous, volatile, and uncertain market. This challenging environment extends beyond local or national boundaries, directly connecting to the international market through products, services, and pricing.

In this sense, SMEs without financial support cannot invest, running out of cash and ending up in bankruptcy. This result has a significant impact on the entire market chain and financial institutions because it misses the opportunity to obtain a profitable return on investment. For these reasons, the prediction of the risk of company bankruptcy is necessary and essential at all stages. Therefore, it is necessary to develop a predictive learning model to forecast the bankruptcy of companies.

Machine learning models have various algorithms that allow the prediction and forecasting of activities with transcendence for the company. However, treatment with relevant criteria, suitable variables, and consistent data is essential [4]. Several studies require models of credit rating, payment yield and business bankruptcy that provide techniques and determining factors through specific patterns in order to generate predictions according to the selection criteria [8]. Credit rating models for SMEs do not reflect the ability to pay in a timely manner towards an opportune horizon for decision-making. The global financial crisis has shown the importance of having a model to predict default [4].

This article presents an optimized machine learning model to determine the bankruptcy of companies, where various bankruptcy models are required with relevant indicators according to the industry and institutional segment. Likewise, the analysis method according to the optimized model, the results obtained and the conclusions are presented.

## 2. RELATED WORKS

Machine learning is an important part of data science. Likewise, various techniques are used to achieve results with the greatest possible precision and allow decision-making [9]. Guerranti and Dimitri [10] detail that the development of machine learning models involves three key stages: data exploration, model selection, and deployment. Kanzari et al. [11] have determined that the most significant variables should be specified and considered according to the company's context, thus allowing for their integration into the selected model.

Brenes et al. [2], in their research titled "An intelligent bankruptcy prediction model using a multilayer perceptron," developed a neural network approach with multiple layers to forecast the bankruptcy of 95 companies in Taiwan. The companies provided financial ratios to enrich the model and generate reliable results that provided a more complete picture. They considered 4 main metrics: precision (93%), accuracy (92%), sensitivity (92%) and f1-score (93%). This study demonstrates the superiority of deep learning over other machine learning techniques.

In their study 'Explainable Machine Learning in Credit Risk Management,' Bussmann et al. [3] observe that companies with highly inconsistent earnings tend to experience low cash flow. This financial instability increases their likelihood of defaulting on creditors and failing to meet external financial

commitments. In the proposed model, he investigated the accuracy of five artificial neural network models for the prediction of credit risk, bankruptcy risk and payment yield. Likewise, tests were carried out with credit rating data from companies in Australia, where they show that the deep neural network performs more accurately in predicting credit risk than other models. The proposal was a deep learning model to support decision-making in financial risk management, and experiments with trade data. They considered 4 main metrics: precision (94%), accuracy (94%), sensitivity (93%) and f1-score (94%). This study demonstrates the superiority of deep learning over other machine learning and rule-based techniques.

In his study, 'Contagion Effects of UK Small Business Failures: A Spatial Hierarchical Autoregressive Model for Binary Data,' Calabrese [4] specifies that SMEs represent 99.9% of all companies in the United Kingdom, of which 60% are in the private sector and contribute to 52% of all turnover. It also indicates that the number of SMEs has grown by 64%. In this sense, he proposed an expert system for institutions to assess credit and bankruptcy risk, which was designed specifically for SMEs. In the aforementioned study, the logistic regression model and 4 main metrics were considered: precision (85%), accuracy (90%), sensitivity (92%) and f1-score (89%). It was concluded that SMEs face various external factors and that in many cases they cannot be controlled in time, which is why they require timely monitoring so as not to lose the credit rating and the bankruptcy index.

Cantarero et al. [5], in their study titled "Effects of the economic crisis on entrepreneurship in social economy companies in Spain: a spatial analysis", developed a logistic regression model to predict the bankruptcy and credit risk of small and medium-sized enterprises (SMEs). The model utilized financial ratios and corporate governance data from a dataset consisting of 934 small companies in Spain. It is concluded that 70% of companies have obtained a higher rating for future loans and 60% have been able to identify if they are at risk of bankruptcy. They considered 4 main metrics: precision (90%), accuracy (91%), sensitivity (90%) and f1-score (90%).

The authors Muñoz-Cancino et al. [8], in their work with the title "Deep Learning Based Performance Prediction Model for a Corporation", developed a prediction model with a neural network architecture. The model has the objective of forecasting stock prices using a set of economic and financial ratios to obtain variables. In addition, a comparison of results was made between prediction performances with a model that was carried out with a multiple discriminant analysis. Subsequently, he complemented a study for the prediction of company bankruptcy where they used a neural network to obtain results of performance and performance indicators in the financial environment. They also improved the model's prediction accuracy through a preprocessing phase. They considered 4 main metrics: precision (94.5%), accuracy (92.7%), sensitivity (93.5%) and F1-Score (94%). Finally, the model made it possible to predict financial performance using a neural network with a backpropagation algorithm.

The authors Zhang et al. [12], in their study titled "Credit risk prediction of SMEs in supply chain finance by fusing demographic and behavioral data," emphasize the significance of financial indices and transaction data in predicting the credit and bankruptcy risk of SMEs using an improved random forest model. The model incorporates financial ratios and legal judgment data to forecast the credit risk of SMEs. However, the limitation of these models lies in the fact that the data they

utilize only includes static information about the SMEs, and the predictive models they adopt are limited to a single modality. They considered 4 main metrics: precision (88%), accuracy (89%), sensitivity (90%) and f1-score (88%).

### 3. METHODOLOGY

#### 3.1 Method

To obtain data, an instrument consisting of 20 questions classified according to the corresponding dimension was applied: Business Environment (10) and Financial Situation (10), these questions were elaborated according to the analysis

of the business context. The instrument is specified in Table 1, Likewise, the observation technique was used with the instrument of record cards to compile the accuracy of each algorithm to be used.

The exploratory data analysis method was used and was complemented with the model proposed by the authors [1], which determined three stages: data exploration, model selection and implementation. In this sense, 21 variables were determined to be considered for the model, where it was identified that the relevant variables are Bankruptcy, Capital, Letters\_to\_pay, Access\_Credit, Satisfaction\_E and Cost effectiveness (Profitability). The 21 variables and types are specified in Table 2.

**Table 1.** Instrument (survey) for data collection

No.	Dimension	Questions	Interval
1		How many months of activity has the company had?	[0-36]
2		Does the company have regulatory plans?	[0-1]
3		How many employees do you have working in the company?	[0-500]
4		Does the company have information systems?	[0-1]
5	Business Environment	How much do you consider to be business satisfaction/work environment?	[0-5]
6		Does it have access to external markets (national and international)?	[0-1]
7		Does it form a society?	[0-1]
8		How many direct competitors does the company have?	[0-10]
9		Did the company experience any claims?	[0-1]
10		Has the company had any unforeseen events?	[0-1]
11		What is the debt capacity?	[0-5]
12		How much was the initial capital of the company?	[0-999999]
13		Do you have access to credit?	[0-1]
14	Financial situation	Do you have outstanding bills?	[0-1]
15		What is the credit rating?	[0-5]
16		What is the payment yield?	[0-5]
17		What is the profit rate?	[0-100%]
18		What is the profitability index?	[0-5]
19		What is the sales efficiency index?	[0-5]
20		What is the insolvency rate?	[0-5]

**Table 2.** Variables and type

Item	Column	Type	Definition
0	Bankruptcy	int 64	Bankruptcy Identifier
1	Months_A	objects	Months of activity that the company has
2	Capital	objects	Initial capital of the company
3	Plans	int 64	Regulatory plans
4	SI	objects	Information Systems
5	Qualification	int 64	Credit Rating
6	Capacity_E	int 64	Borrowing capacity
7	Satisfaction_E	objects	Business satisfaction / Work environment
8	Credit_Access	int 64	Access to credit
9	Access_Markets	int 64	Access to external markets (national and international)
10	Staff	int 64	Employees you have working in the company
11	Letters_for_pay	objects	Unpaid bills
12	Companies	int 64	Formation of a company
13	Performance_Payment	int 64	Payment performance
14	Competitors_D	int 64	direct competitors the company has
15	Unforeseen	objects	Claims Involvement
16	Claims	objects	Presentation of contingencies
17	Profit_rate	objects	Profit rate
18	Cost effectiveness	int 64	Profitability Index
19	Efficiency	int 64	Sales efficiency index
20	Insolvency	int 64	Insolvency rate

#### 3.2 Data exploration

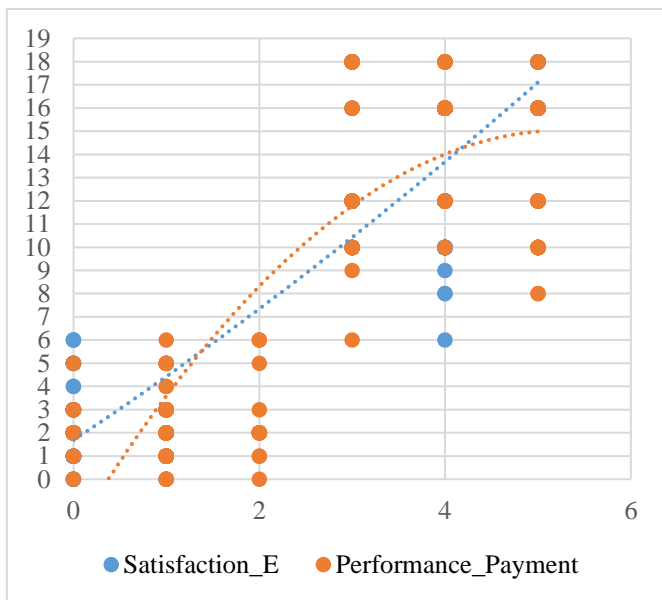
The study population is determined by the companies in the city of Metropolitan Lima. Which is specified according to the National Institute of Statistics and Informatics - INEI in the

technical report on business demographics in Peru IV quarter of 2023, which has as its source the Central Directory of Companies and Establishments (DCEE). In that sense, there are 26,330 companies registered as of the IV quarter of 2023 in Metropolitan Lima. The study sample was determined in

379 companies in the city of Lima, Peru (microenterprises, small companies and medium-sized companies). The sample size has been calculated with a margin of error of 5% and a confidence level of 95% using the finite universe sample calculation formula. Likewise, through an exploration process, 118 records were identified that correspond to microenterprises, 166 to small companies and 95 to medium-sized companies. Table 3 considers the type of company.

**Table 3.** Total de companies

No.	Type	Total of Companies
1	Microenterprise	118
2	Small business	166
3	Medium company	95
	<b>Total</b>	<b>379</b>



**Figure 1.** Time series and correlations (Payment\_Performance with Satisfaction\_E and Months\_A)

An exploratory data analysis was applied to obtain useful factors about business bankruptcy and identify the main causes of it. Likewise, graphs, time series analysis and correlations were used. Nicodemo and Satorra [13] argue that this type of analysis is advantageous because it focuses on essential aspects of a company's context and utilizes important data to generate forecasts. It also allows incorporating new variables for improving or optimizing the model.

Graphs, time series analysis, and correlations were used. This type of analysis allows us to demonstrate that the data patterns are useful to generate a forecast on the effects of the bankruptcy effect of companies, where identifying the factors is very decisive. Figure 1 states that it has been identified that the variables Performance\_Payment and Satisfaction\_E are correlated with the variable Months\_A, so studying these correlations is necessary to understand the characteristics and impact. Correlations were obtained to identify variables that directly impact business bankruptcy.

Data analysis was used in distributions and line graphs of the variables: Bankruptcy, Months\_A, Capital, Plans, If, Qualification, Capacity\_E, Satisfaction\_E, Access\_Credit, Access\_markets, Personnel, Bills\_To\_Pay, Companies, Performance\_Payment, Competitors\_D, Unforeseen, Claims, Profit\_rate, Profitability, Efficiency and Insolvency, where

levels of predominance were identified that require correlation towards each resulting factor.

In addition to this, it was obtained that the study variables have significant attributes for the identification of specific characteristics of bankruptcy. In this way, the bankruptcy forecast can be based not only on the general context of the company but also on the internal context of the processes. It is important that the data set used has variables with a specific level of disaggregation and that it allows relationships to be generated with other variables. To guarantee the effectiveness of the chosen model, it is necessary to consider two essential criteria: 1. The characteristics of the data set, and 2. The correlations between the variables. Figure 2 specifies the characteristics of each variable, where variables that represent greater relevance for the context of business bankruptcy are presented. Likewise, it is important to identify the variables that will be worked on in correlations using the defined criteria.

### 3.3 Model selection

Pratt et al. [14] argue that machine learning algorithms, such as logistic regression, random forests, and neural networks, are well-suited for predicting outcomes using a dataset containing variables that may be correlated. These algorithms are particularly effective in addressing risk identification problems within a specific and detailed context. In the same way, Pandey et al. [15] consider that in order to achieve solid metrics, it is important to implement a machine learning model with characteristics of the context analyzed and that can be aligned with institutional goals. Sulastrri and Janssen [16] also stated that determining the metrics to evaluate the results of an algorithm corresponds to having an exact understanding of the problem in its context and clear objectives. In this context, Yahia et al. [17] consider that in order to obtain better results it is necessary to generate optimization processes that include the incorporation of more suitable variables and correlations between them.

Raza et al. [18] highlight that logistic regression and random forest algorithms are particularly well-suited for incorporating dynamic market information and variables related to the business context. This means that they are operational and suitable for estimating investment probabilities and forecasts of various types, in such a way that data sources are open to any information related to the probability of bankruptcy and potential for loss. For example: stock market information, particularities of the sector in which the activity is carried out, regulation, ease of access to credit, business bankruptcy, etc. In the same way, Muthukumaran and Hariharanath [7] state that deep learning is particularly effective in applications involving Small and Medium-sized Enterprises (SMEs), thanks to its versatility and capability to facilitate numerous training iterations. The main objective of the technique is to determine the financial status of SMEs, which contains the design of the selection of characteristics based on the optimization algorithm and neural network, which is used for data classification. Metrics are necessary and play an important role in defining financial firms' profitability and productivity.

To select a machine learning model, the objectives of the study must be aligned with the activities to be developed. It is also important to have the detail of the general and specific context variables. In this sense, for the selection of the model, application tests were carried out on the dataset with 3 machine learning models that respond to the objectives of the

study. These are: Logistic regression, random forests and neural networks.

Calabrese [4] points out that logistic regression is useful for carrying out analysis activities and prediction processes that allow reducing costs that mean a setback or delay, thereby improving and generating efficiency. Companies to obtain better profitability results and business bankruptcy forecasts require knowing variables and patterns appropriate to their context. They also consider that the essential characteristics of this algorithm are: Identify the question, collect historical data and train the model. Zhang et al. [12] specify that they are important because the structure considers various trees with decision aspects. Likewise, the ease of application and

flexibility has been successfully implemented, as it handles classification and regression problems at a general and specific level. Likewise, decision trees tend to have better power in the result when the variables are correlated. Muñoz-Cancino et al. [8] consider that neural networks are ideal for solutions that generate forecasts because they generate correlations from the identification of variables, which are necessary to establish focus points to achieve effective training and generate the precise learning to obtain ideal results. In this sense, neural networks consider training data to learn and promote improved accuracy over time. In this context, the proposed optimization model considers that training is an indispensable process to improve the results of the metrics.

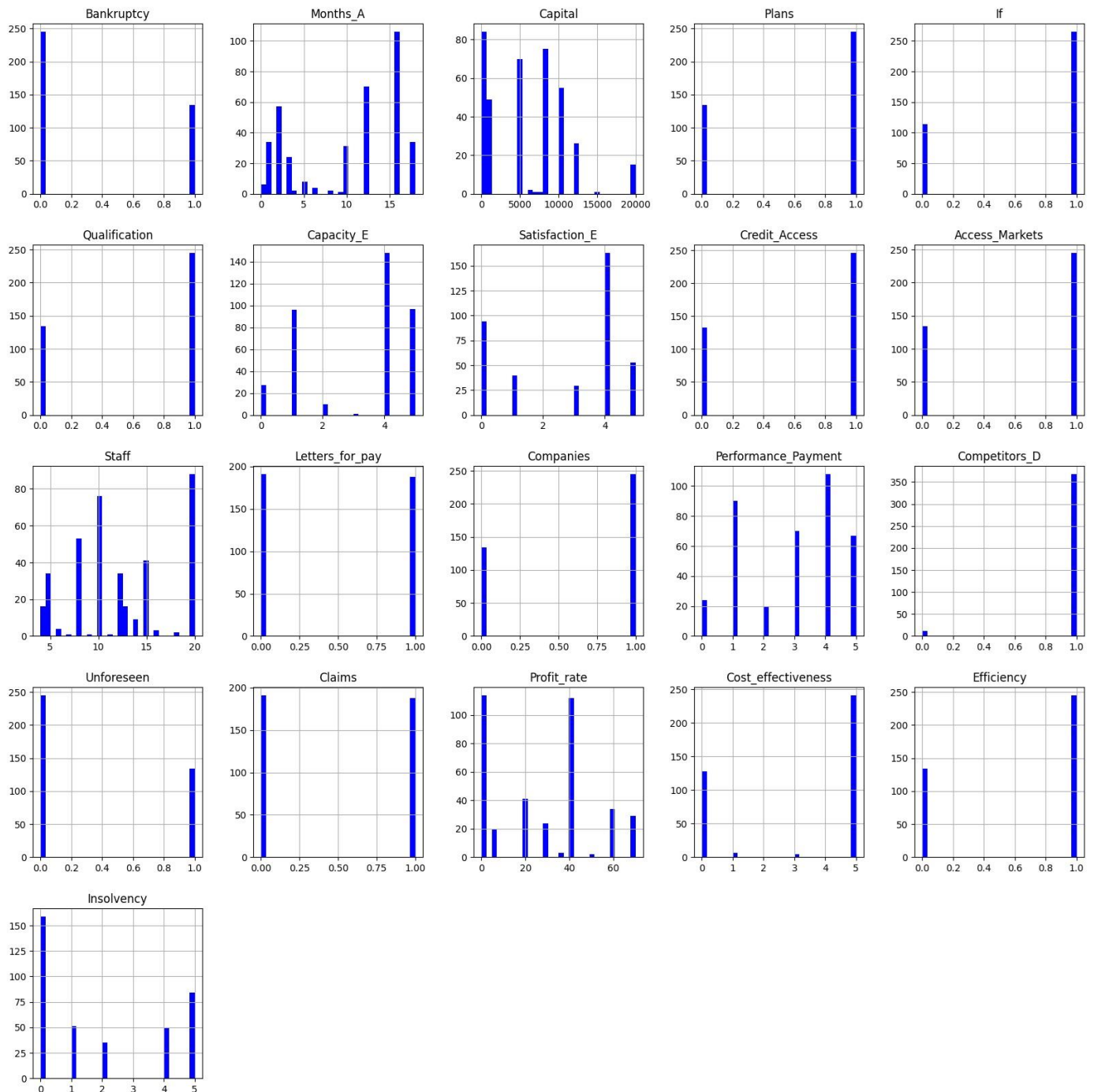
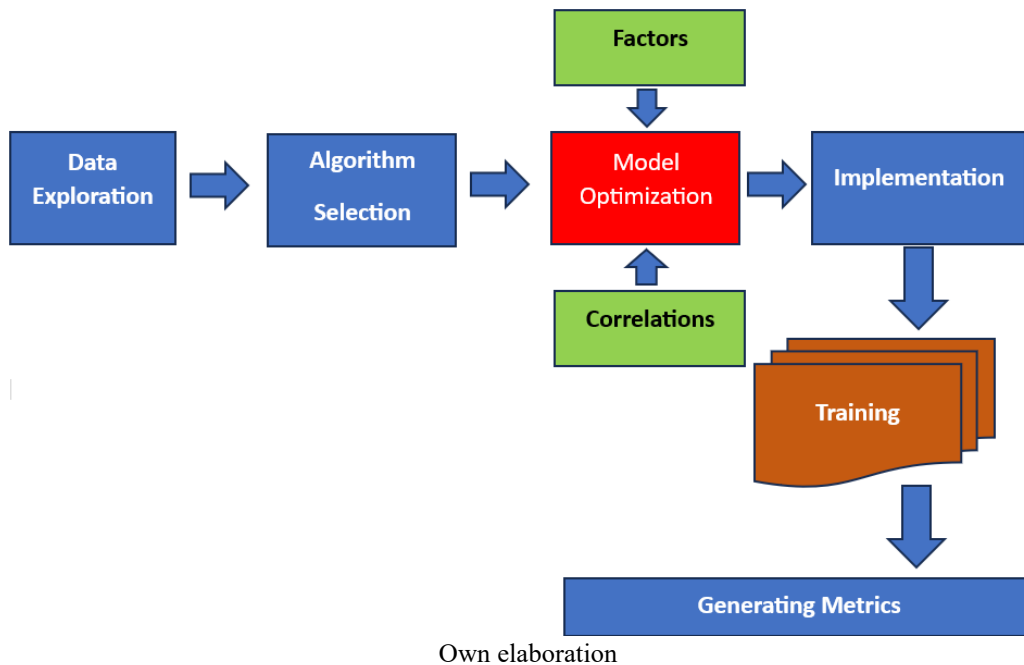


Figure 2. The characteristics of the data sets

**Table 4.** Algorithm comparison

No.	Algorithm	Advantages	Disadvantages	Score
1	Linear Regression	Facility for precise and clear understanding in deployment.	It does not support complex relationships. Tendency to over-adjustments.	2
2	Logistic Regression	Facility for precise and clear understanding in deployment. Support for generating relationships between variables.	Complex relationships do not consider all characteristics. It requires add-ons to improve complex relationships.	4
3	Decision Tree	Facility for precise and clear understanding in deployment. Relationships between variables are generated. Ease of implementation.	Complex relationships are resource-intensive.	3
4	Random forests	Facility for precise and clear understanding in deployment. Support for generating relationships between variables. Combination of individual and group factors and requirements. Efficient and effective training. High performance.	Complex relationships are resource-intensive.	4
5	Gradient Increase	Facility for precise and clear understanding in deployment. Support for generating relationships between variables. High performance.	Small changes in characteristics will produce significant changes.	2
6	Neural network	Facility for precise and clear understanding in deployment. Relationships between variables are generated. Efficient and effective training. High performance.	Delay in the deployment of training results (for some complex cases).	5

Score: 1 al 5 (inadequate to adequate)  
Own elaboration



**Figure 3.** Optimized model architecture

**Table 5.** Metrics

Reference	Algorithm	Metric
[2]	Neural network	Prec-Acc-Sens-F1S
[3]	Neural network	Prec-Acc-Sens-F1S
[4]	Logistic Regression	Prec-Acc-Sens-F1S
[5]	Logistic Regression	Prec-Acc-Sens-F1S
[8]	Neural network	Prec-Acc-Sens-F1S
[12]	Random Forests	Prec-Acc-Sens-F1S

Metric: Precision: Prec. Accuracy: Acc. Sensitivity: Sens. F1-Score: F1S.  
Own elaboration

Various algorithms were compared for the treatment of the specific problem of prediction. The comparison is specified in Table 4.

The metrics were chosen according to the experience generated with research resulting from the similar context on the predictions of a particular problem. This choice of metrics is specified in Table 5. Likewise, it is specified that the chosen algorithms respond to the score generated after the evaluation carried out in Table 4, which is aligned with the background of the topic addressed. It was identified that the neural network is the algorithm that has the highest score and that it is in 3 antecedents on prediction, then the algorithms of random forests and logistic regression have 1 and 2 antecedents of prediction respectively, and 4 of punctuation. The background information chosen is the product of the research carried out as a state of the art, which considers references in contexts similar to the subject treated with similar metrics.



### 3.4 Model optimization

The study variables have significant attributes for the identification of specific characteristics of bankruptcy. In this way, the bankruptcy forecast can be based not only on the general context of the company but also on the internal context of the processes. Sulastri and Janssen [16] consider that generating suitable variables is essential to generate optimal results, thereby ensuring adequate prediction and performance. It is important that the dataset used has variables with a specific level of disaggregation and that it allows generating relationships with other variables. To guarantee the effectiveness of the chosen model, it is necessary to consider two essential criteria: 1. The characteristics of the data set, and 2. The correlations between the variables.

The optimized model was developed based on the model proposed by Alsheref et al. [1] who determined three stages: data exploration, algorithm selection and implementation, in this way it was possible to identify the best machine learning algorithm. In this context, an additional stage was incorporated: Optimization of the model, which included the identification of factors (relevant variables) and the correlations between them. Likewise, it gave greater relevance to the training process, where it was considered as a key step for the optimization of the values of the selected algorithm.

In this sense, 21 variables that were worked on in the data exploration stage were considered. For the algorithm selection stage, 3 algorithms were considered, where the neural network algorithm was the one that obtained results with the highest percentages in the established metrics. For the model optimization stage, it was identified that the relevant variables were: Performance\_Payment, Capital, Letters\_for\_pay, Credit\_Access, Satisfaction\_E and Profitability. The proposed model considered exploratory data analysis to obtain useful factors on business bankruptcy and identify the main causes of bankruptcy. Likewise, graphs, time series analysis and correlations were used. This type of analysis allows us to demonstrate that data patterns are useful to generate a forecast on the effects of the bankruptcy effect of companies, where the identification of factors is very decisive. It has been identified that the Performance\_Payment and Satisfaction\_E variables are correlated with the Meses\_A variable, so the study of these correlations is necessary to know characteristics and impact. Correlations were obtained to identify variables that have a direct impact on business bankruptcy. Figure 3 indicates the architecture of the optimized model.

### 3.5 Implementation

Yao et al. [19] indicate that the implementation stage must specify the choice of the algorithm that has better metrics and that is adapted to the requirements of the business context according to the recorded data. Likewise, Zou et al. [20] consider that the neural network model meets the necessary characteristics to achieve the proposed objectives to generate predictions on a large number of variables and data. In this sense, correlations were established between the relevant variables that have been necessary to achieve better results. To achieve optimal results with optimization characteristics, it is necessary that the selected algorithm has obtained results greater than the results of the other algorithms subject to evaluation. Likewise, it is essential to consider that the training process is a key factor for the learning of the algorithm; all of this will guarantee satisfactory, efficient and truthful results.

## 4. RESULTS

### 4.1 Analysis of algorithm results

#### 4.1.1 Logistic regression model

The logistic regression algorithm has techniques used in companies with a lot of recurrence and where it uses effective classification methods without major complexity, which requires a dataset with a large number of records and related variables for a certain type of context [15].

Table 6 shows the dataset was deployed with a logistic regression algorithm, where an accuracy of 85%, accuracy of 83%, sensitivity of 97% and f1-score of 90% was obtained.

**Table 6.** Logistic regression model results

	Precision	Recall	F1-Score	Support
0	0.85	0.97	0.90	379
1	0.27	0.75	0.58	
	0.58	0.22	0.32	
Accuracy	0.83	379		
Macro avg	0.72	0.59	0.61	379
Weighted avg	0.80	0.83	0.80	379

#### 4.1.2 Random forest model

The random forest algorithm has techniques that combine multiple decision trees that are aligned to obtain a common goal. Likewise, the use of this algorithm is important due to its ease of understanding and alignment with the objectives of the study [21].

Table 7 shows the data set that was implemented, obtaining an accuracy of 83%, accuracy of 83%, sensitivity of 96% and f1-score of 90%.

**Table 7.** Random forest model results

	Precision	Recall	F1-Score	Support
0	0.83	0.96	0.90	379
1	0.23	0.90	0.74	
	0.60	0.06	0.16	
Accuracy	0.83	0.83	379	
Macro avg	0.72	0.54	0.53	379
Weighted avg	0.79	0.83	0.77	379

#### 4.1.3 Neural network model

Neural networks are considered to be an emerging algorithm that has been on the rise in the area of machine learning. In this sense, neural networks manage to generate the processing of a large number of variables and at a significant speed. They also use useful and innovative techniques to solve business problems [18].

**Table 8.** Neural network model results

	Precision	Recall	F1-Score	Support
0	0.90	0.95	0.92	379
1	0.31	0.52	0.43	
	0.59	0.43	0.49	
Accuracy		0.87	379	
Macro avg	0.74	0.69	0.71	379
Weighted avg	0.85	0.87	0.86	379

Table 8 shows the dataset was deployed, where an accuracy

of 90%, accuracy of 87%, sensitivity of 95% and f1-score of 92% was obtained. In this sense, the neural network algorithm is chosen as the basis for the implementation of the optimized model to predict the bankruptcy of companies.

The model was selected under 2 criteria: 1. The variables correlate with each other and 2. Greater precision, accuracy, sensitivity and f1-score.

#### 4.2 Selection model optimized

The selected algorithm considered 21 variables that have characteristics specific to the employee according to the internal context of the company. Initially there is a precision of 90%, accuracy of 87%, sensitivity of 95% and f1-score of 92%. Likewise, the model proposed by the Alsheref et al. [1] was the basis for the complementation towards the optimized model that was proposed in the present research.

The variables identified in the dataset are correlated with each other depending on the context of the company, Figure 4 shows the correlation values between variables.

Data analysis was used in distributions and line graphs of the variables: Bankruptcy, Meses\_A, Capital, Plans, SI, Rating, Capacidad\_E, Satisfaction\_E, Acceso\_Credito, Acceso\_mercados, Personnel, Letters\_for\_pay, Companies, Performance\_Payment, Competidores\_D, Contingencies, Claims, Tasa\_utilidades, Profitability, Efficiency and Insolvency, where levels of dominance were identified that require correlation towards each resulting factor.

Relevant variables related to the company's own context were considered, which were: Meses\_A, Capital, Satisfaction\_E, Credit\_Access, Performance\_Payment, Letters\_for\_pay, Profitability and Insolvency. These variables were required to initiate the training and cleaning of the data.

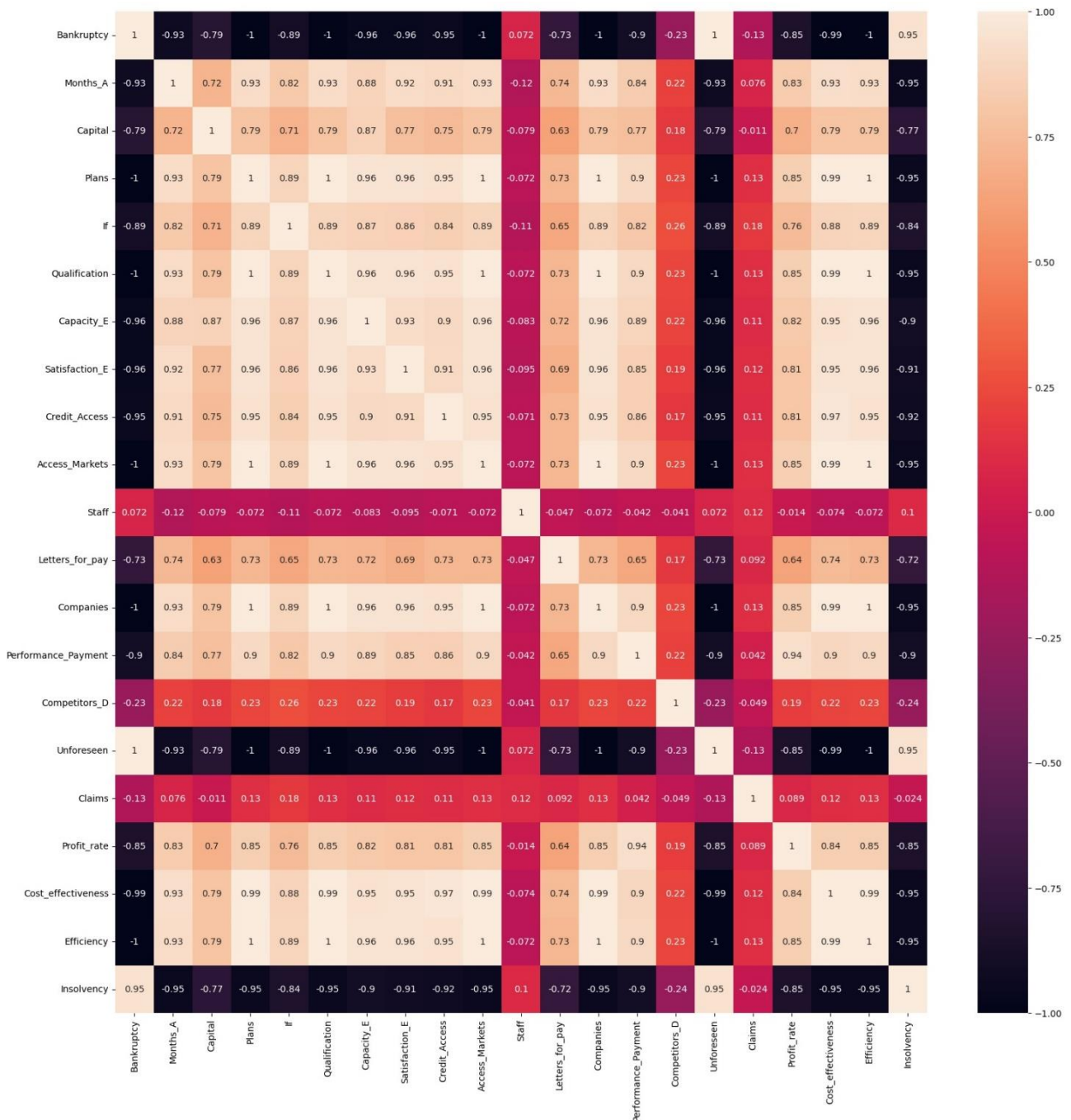


Figure 4. Correlations between variables



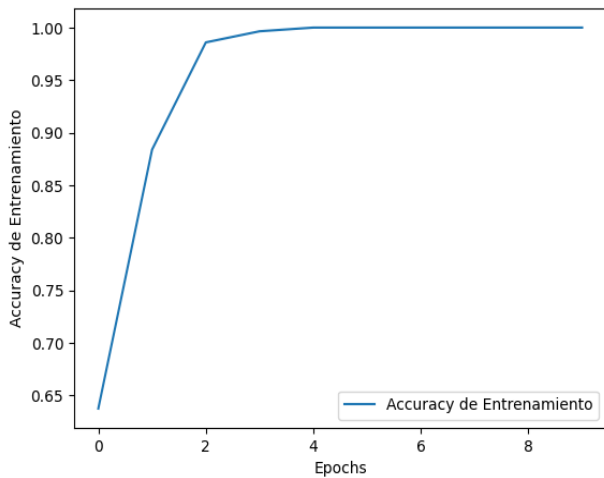
### 4.3 Optimized model training

The process for training the proposed optimized model was carried out, which requires consistent data for the algorithm to generate optimal results according to the learning capacity. Likewise, the data necessary for training requires the ideal response so that the algorithm generates accurate responses through the training patterns that assign the attributes of the input data to the destination. Regarding the data necessary for training the optimized model, it is specified that the relevant

data has been taken as a reference to carry out the correlations. It is necessary to indicate that the data with high and low scores represent ideal prediction moments; However, where more efficient predictions are required are with companies that have sparse data and for which training is necessary to achieve the prediction. The training process is decisive to obtain better metrics in companies with intermediate data. The relevant variables in Table 9 that are necessary for training are considered.

**Table 9.** Relevant variables

No.	Bankruptcy	Performance_Payment	Capital	Months_A	Insolvency	Satisfaction_E	Cost_Effectiveness
0	1	0	1000	2	4	0	0
1	0	3	5000	6	2	4	3
2	1	1	0	4	4	0	0
3	0	5	10000	10	2	3	5
4	0	5	10000	12	0	4	5
...	...	...	...	...	...	...	...
374	1	1	500	4	4	0	1
375	0	5	10000	16	0	4	5
376	0	5	10000	16	0	4	5
377	0	5	10000	16	0	4	5
378	1	0	500	5	4	0	1



**Figure 5.** Model success rate during training

Subsequently, training was carried out with 100 events necessary to improve the metrics. Figure 5 specifies that the events proposed in other models do not reach percentages close to 100%. Therefore, to achieve an optimized model, 100 application events are required.

### 4.4 Analysis of model optimization results

Table 10 shows the data set was deployed, where a precision of 97%, accuracy of 97%, sensitivity of 99% and f1-score of 98% were obtained.

Likewise, after the deployment, the number of companies that would experience bankruptcy was identified with the results for the number of employees who leave the company.

Total enterprises 379.

Number of companies that are at risk of bankruptcy 134

Percentage of companies that are at risk of bankruptcy 35.36%.

Number of companies that are not at risk of bankruptcy 245

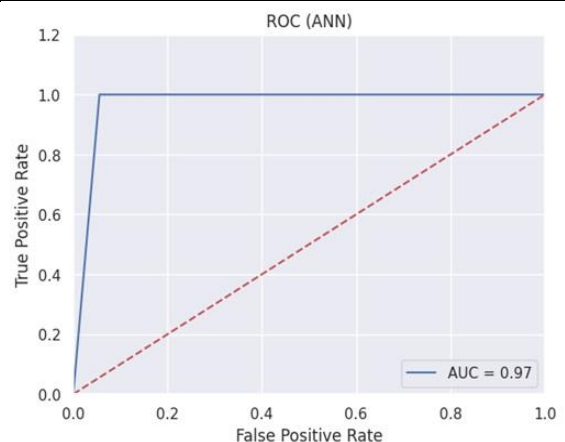
Percentage of companies that are at risk of bankruptcy 64.64%.

Figure 6 specifies that the ROC curve test specifies the binary classification of companies into two groups: one positive and one negative. In this sense, the results obtained demonstrate that the performance of the model has a greater contained area with a predominance to the left and where the AUC classifier value (precision) of 0.97 (close to 1) is required.

The optimized model has required stages of data exploration, model selection and implementation to identify relevant variables. Likewise, for the treatment and simulation with the incorporated variables, 100 training events have been needed to improve the deep learning model that did not have necessary variables.

**Table 10.** Optimized model results

	Precision	Recall	F1-Score	Support
0	0.97	0.99	0.98	379
1	0.32	0.41	0.46	
Accuracy		0.97	0.98	379
Macro avg	0.74	0.69	0.71	379
Weighted avg	0.85	0.87	0.86	379



**Figure 6.** ROC curve test results

### 4.5 Comparison of results

In the Table 11 it can be seen that the algorithm: Model optimization obtained the best results compared to all the other algorithms with precision of 97%, accuracy of 97%, sensitivity of 99% and f1-score of 98%.

**Table 11.** Comparison of results

No.	Algorithm	Precision	Accuracy	F1-Score	Sensitivity
1	Logistic Regression	85%	83%	90%	97%
2	Random Fores	83%	83%	90%	96%
3	Neural Network	90%	87%	92%	95%
4	Neural Network (Model optimization)	97%	97%	98%	99%

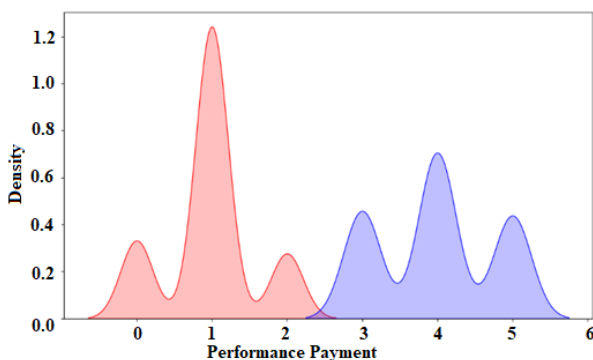
### 4.6 Factors for business bankruptcy

Results were obtained on the relevant variable of Performance\_Payment, where it was identified that a greater number of companies that go bankrupt have a lower payment performance of their obligations and the companies that do not go bankrupt are those that, for the most part, have a higher performance. payment. Likewise, Figure 7 specifies that code to identify payment performance factor. Figure 8 specifies that it was identified that a smaller number of companies that go bankrupt are those that will not be able to recover in the future due to unpayable debts. In that sense, for this segment it is necessary to consider other variables and factors.

```
plt.figure(figsize=(12, 7))

sns.kdeplot(left_df['Performance_Payment'],
            label = "Performance Payment Low",
            color = 'r', fill=True)
sns.kdeplot(stayed_df['Performance_Payment'],
            label = "Performance Payment High",
            color = 'b', fill=True)
plt.xlabel('Performance Payment')
```

**Figure 7.** Code to identify payment performance factor



**Figure 8.** Payment performance factor results

Figure 9 specifies that code to identify the business satisfaction factor. Figure 10 specifies that results were obtained on the relevant variable of Satisfaction\_E where it was identified that a greater number of companies that go

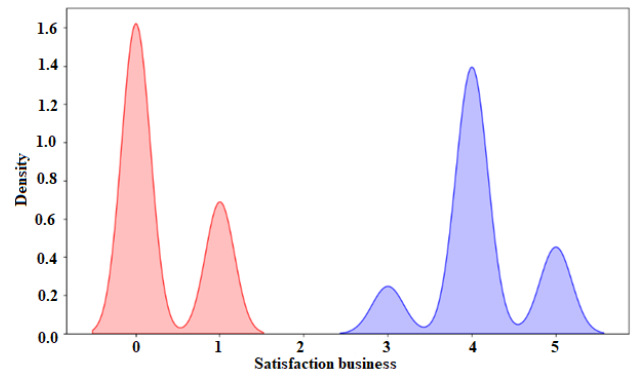
bankrupt have lower business satisfaction (rating of 0) and the companies that do not go bankrupt are mostly those that have a higher business satisfaction (4).

```
plt.figure(figsize=(12, 7))

sns.kdeplot(left_df['Satisfaction_E'],
            label = "Satisfaction business Low",
            fill = True, color = 'r')
sns.kdeplot(stayed_df['Satisfaction_E'],
            label = "Satisfaction business High",
            fill = True, color = 'b')

plt.xlabel('Satisfaction business')
```

**Figure 9.** Code to identify business satisfaction factor



**Figure 10.** Business satisfaction factor results

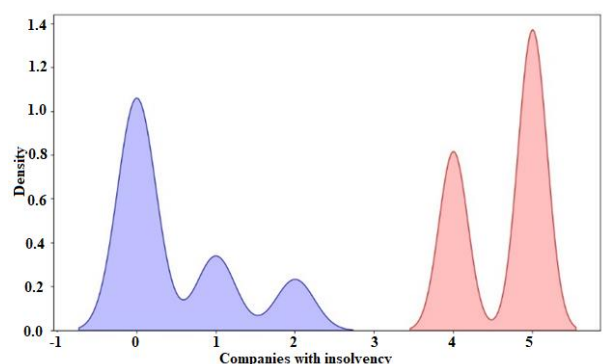
Figure 11 specifies that code for insolvency factor. Figure 12 specifies that results were obtained on the relevant variable of Insolvency, where it was identified that a greater number of companies go bankrupt due to high insolvency. Likewise, companies that have low insolvency were identified, which means they have a greater capacity to sustain themselves over time.

```
plt.figure(figsize=(12, 7))

sns.kdeplot(left_df['Insolvency'],
            label = "Companies with insolvency Low",
            fill = True, color = 'r')
sns.kdeplot(stayed_df['Insolvency'],
            label = "Companies with insolvency High",
            fill = True, color = 'b')

plt.xlabel('Companies with insolvency')
```

**Figure 11.** Code for insolvency factor



**Figure 12.** Insolvency factor results

Figure 13 specifies that code for activity time factor. Figure 14 specifies that results were obtained on the relevant variable of company activity time (Months\_A), where it was identified that a smaller amount of activity time makes the company at greater risk of bankruptcy.

```
plt.figure(figsize=(12, 7))

sns.kdeplot(left_df['Months_A'],
            label = "Company activity time Low",
            fill = True, color = 'r')

sns.kdeplot(stayed_df['Months_A'],
            label = "Company activity time High",
            fill = True, color = 'b')

plt.xlabel('Company activity time')
```

Figure 13. Code for activity time factor

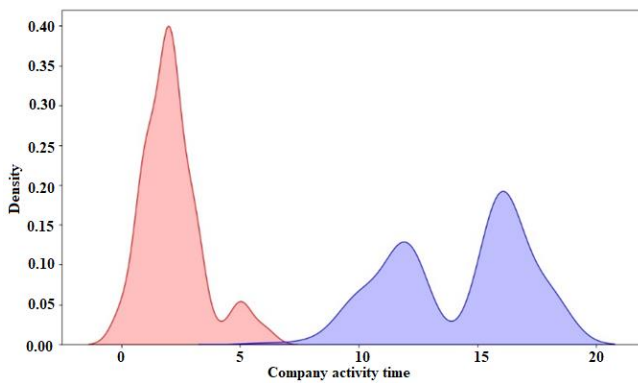


Figure 14. Activity time

#### 4.7 Correlations

Correlations mean the identification of aspects that are generated from the attributes of each variable and that together mean a relevant factor. The following correlations were determined: Payment performance with capital, Profitability, Business Satisfaction and Insolvency with Activity Time (Months\_A). Payment performance is an important variable in any prediction model since it can identify whether its treatment is decisive for business bankruptcy. Likewise, Figure 15 specifies that the higher payment yield has a very high degree of identification within the relevant variable for this optimized model. The opposite of poor performance which is an indicator of possible business bankruptcy.

The Table 12 shows the results of Spearman's rho test and the correlation coefficient, which identifies the degree of relationship between payment yield and capital. It was obtained that  $p = .000 < \alpha = .01$ , therefore, it is specified that there is a highly significant relationship between payment yield and capital; That is, as the company has greater capital, then it will obtain a higher payment yield to meet its medium and long-term obligations. The value of  $\rho = .492$  indicates a moderate positive correlation.

Profitability is a very important variable to consider for every company because it allows you to identify efficiency in income management. Likewise, Figure 16 specifies that it has been identified that when a company has a longer period of activity (months) it has greater cost effectiveness (profitability). For this reason, it is very possible that a

business bankruptcy will originate from the first months of activity.

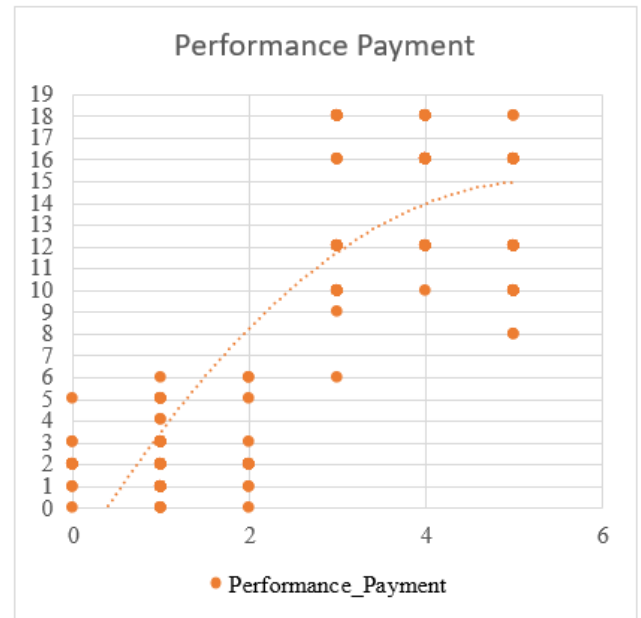


Figure 15. Payment performance results

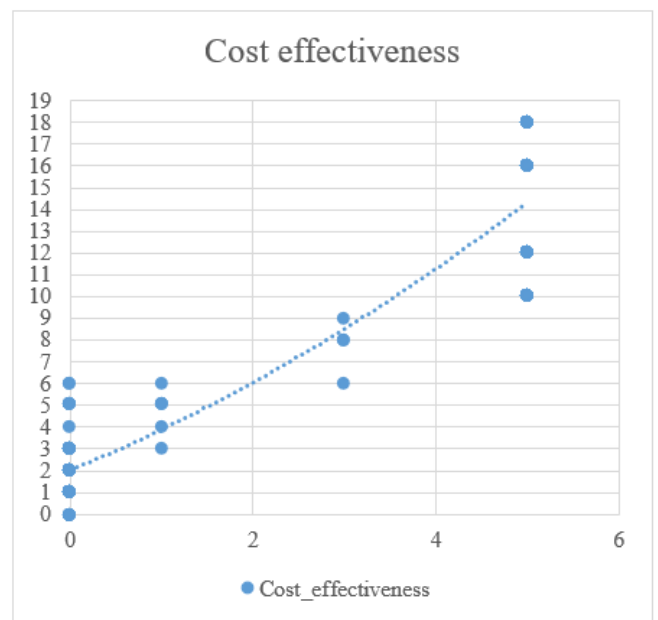


Figure 16. Cost effectiveness (profitability) results with time of activities (months)

Table 12. Correlation between performance and capital

	Correlations	
	Performance	Capital
Corr.Coefficie	1, 000	.492**
nt		
Sig.	.	,000
Nr	709	709
Corr.Coefficie	.492**	1, 000
nt		
Sig.	,000	
Nr	709	709

\*\* . Correlation significant

**Table 13.** Correlation between cost effectiveness (profitability) with time of activities (months)

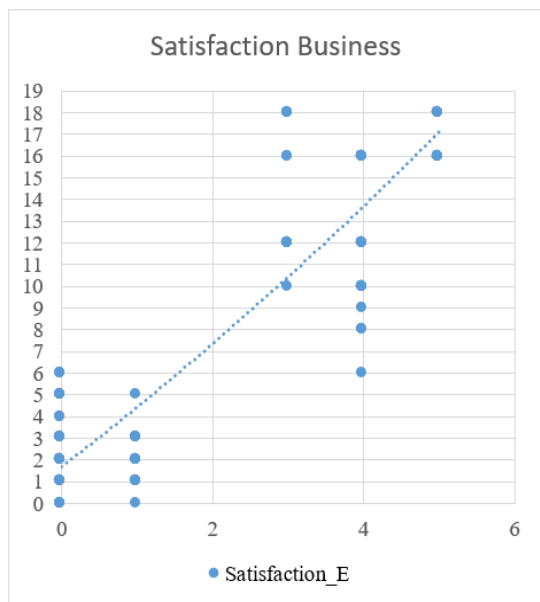
Correlations		
	Cost effectiveness	Activity time
Corr.Coefficient	1, 000	.392**
Sig.		,000
Nr	709	709
Corr.Coefficient	.392**	1, 000
Sig.	,000	.
Nr	709	709

\*\* . Correlation significant

The Table 13 shows the results of Spearman's rho test and the correlation coefficient, which identifies the degree of relationship between distance from home and satisfaction.

It was obtained that  $p = .000 < \alpha = .01$ . Therefore, it is specified that there is a highly significant relationship between profitability and activity time; That is to say, the longer the activity time, the higher the profitability. The value of  $\rho = .392$  indicates a low positive correlation.

Business satisfaction allows us to identify determining aspects not only of the work environment in the company but rather of the environment regarding the context of consolidation and monetary health. Figure 17 specifies that it has been obtained that the degree of satisfaction is greater when the company consolidates over time. The opposite is true for companies that have less time in activities where it is identified that they have lower satisfaction and that could possibly cause business bankruptcy.



**Figure 17.** Business satisfaction results with activity time

**Table 14.** Correlation between satisfaction and activity time

Correlations		
	Business satisfaction	Activity time
Corr.Coefficient	1, 000	.216**
Sig.		.006
Nr	159	159
Corr.Coefficient	.216**	1, 000
Sig.	.006	.
Nr	709	709

\*\* . Correlation significant



**Figure 18.** Business satisfaction results with activity time

**Table 15.** Correlation between insolvency and time of activities

Correlations		
	Insolvency	Activity time
Correlation coefficient	1,000	.210**
Sig. (bilateral)	.	,000
N	709	709
Correlation coefficient	.318**	1,000
Sig. (bilateral)	,000	.
N	709	709

\*\* . The correlation is significant at the 0.01 level (two-sided).

Table 14 shows the results of Spearman's rho test and the correlation coefficient, which identifies the degree of relationship between satisfaction and activity time.  $p = .006 < \alpha = .01$  was obtained. Therefore, it is specified that there is a highly significant relationship between satisfaction and activity time; That is, as the company has more time in activities, it will present greater business satisfaction. The value of  $\rho = .216$  indicates a low positive correlation.

Insolvency and time of activities are important variables within any business bankruptcy model. Figure 18 specifies that it was identified that there is greater insolvency among companies that have less time in activities, and lower insolvency is found in companies that have more time in activities.

Table 15 shows the results of Spearman's rho test and the correlation coefficient, which identifies the degree of relationship between insolvency and activity time. It was obtained that  $p = .000 < \alpha = .01$ . Therefore, it is specified that there is a highly significant relationship between insolvency and time of activities; That is, a relationship is identified between companies that have less time in activities with greater insolvency. The rho value = .210 indicates that there is a low positive correlation.

#### 4.8 Analysis of business importance

It has been identified that the results of the metrics differ with the type of company, where it was obtained that micro companies have lower results than small and medium-sized companies regarding the metrics of precision, accuracy, sensitivity and f1-score. Because the relevant factors (identified in the research) do not have better scores due to their little support or minimal start of income. Likewise, the knowledge and exploration of data and the subsequent processing of the information that companies have is essential to achieve more consistent and solid results. In this regard, it is specified that the identification of these factors is important for micro and small businesses because they generally do not have accurate information and have subsistence activities. All of this allows you to know the current state of your position vis-à-vis the market and financial institutions; In addition, avoid a possible business bankruptcy. Medium-sized companies have more and better tools for identifying factors; However, in this study it has been identified that they are only recently aware of the importance and relevance provided by the results. Table 16 specifies the percentage of metrics by type of company.

**Table 16.** Metrics by type of company

No.	Tipo	Red Neuronal (Model Optimization)			
		Precision	Accuracy	Sensitivity	F1-Score
1	Micro enterprise	94.70%	94.20%	96.60%	97.60%
2	Small business	96.30%	97.40%	98.00%	98.60%
3	Medium company	98.60%	98.60%	98.70%	99.40%

#### 5. DISCUSSIONS

It is specified that the model optimization neural network algorithm exceeded the metrics of the other algorithms considered in the present research where an accuracy of 97%, accuracy of 97%, sensitivity of 99% and f1-score of 98% was obtained. It is also specified that these results are similar to the research carried out by Calabrese [4] that in the study: "Contagion effects of UK small business failures: A spatial hierarchical autoregressive model for binary data", it obtained 4 main metrics: precision (85%), accuracy (90%), sensitivity (92%) and f1-score (89%). These metrics made it possible to assess credit and bankruptcy risk, which was designed specifically for SMEs. Similarly, results similar to the study: "Deep Learning-Based Corporate Performance Prediction Model Considering Technical Capability" by the authors of the authors were obtained [21], who proposed a model based on an artificial neural network that predicts stock prices using economic and financial indices as predictors. They considered 4 main metrics: precision (94.5%), accuracy (92.7%), sensitivity (93.5%) and f1-score (94%). They keep similar results with the aim of predicting the financial performance of companies using an artificial neural network with backpropagation algorithms.

Regarding the general objective of the research, it was necessary to develop a machine learning model to improve the prediction of company bankruptcy, which initially through the 3 stages proposed by Alsheref et al. [1]: data exploration,

algorithm selection and implementation, after which an additional stage was incorporated: model optimization. This allowed the results of the metrics to improve. These results have a similar behavior with what was specified in the study: "Credit risk prediction of SMEs in supply chain finance by fusing demographic and behavioral data", proposed by Zhang et al. [12] who specify the importance of the financial index and transaction data to predict the credit and bankruptcy risk of SMEs with an improved random forest model, where they considered the improvement of the model based on the respective training, for this reason they obtained the metrics: precision (88%), accuracy (89%), sensitivity (90%) and f1-score (88%). Similar results were presented in the study: "Explainable Machine Learning in Credit Risk Management", proposed by Bussmann et al. [3], which requires the treatment of five artificial neural network models for the prediction of credit risk, bankruptcy risk and payment yield. It was shown that the deep neural network performs more accurately in predicting credit risk than other models. The proposal was a deep learning model to support decision-making in financial risk management, and experiments with trade data. They considered 4 main metrics: precision (94%), accuracy (94%), sensitivity (93%) and f1-score (94%). This study demonstrates the superiority of deep learning over other machine learning and rule-based techniques.

The optimized model has considered representative variables according to the context of the company, all of which has allowed us to improve the results. The results obtained have allowed us to know and predict business bankruptcy based on specific and representative variables assumed in the company based on its context. It is important that future work is oriented towards the identification or incorporation of new variables that improve the training process, with respect to the number of events that require learning. All of this will allow obtaining results with greater accuracy, precision and less time.

Companies have a large percentage of money and workforce losses in formalization, continuity, subsistence and credit search processes because not all companies continue in operation after 6 or 12 months. Likewise, business bankruptcy, motivated by various factors, is very significant in small and medium-sized companies. Likewise, it is important to specify that the optimized model has obtained results with great solvency and with this, variables are presented that the company needs to know to strengthen its own critical processes. In this sense, the optimized model has made it possible to identify the variables that generate the possible bankruptcy of the company and the subsequent treatment.

#### 6. CONCLUSIONS

The application of the optimized model allowed us to identify that 134 (35.36%) companies are at risk of bankruptcy out of the total of 379 (100%) employees. Likewise, it was identified that the most determining variables for business bankruptcy in the context of the company under study are: Payment performance, Capital, Profitability, Activity time, Business Satisfaction and Insolvency. In this sense, it is necessary for companies to work on strengthening these variables and thereby help minimize the prognosis of a business bankruptcy. The models studied logistic regression, random forests and neural networks have limitations in terms of identifying specific and relevant variables according to the context of the company. Reason why, they only consider 1 or



2 variables; However, the optimized model considers 8 relevant variables: Months\_A, Capital, Satisfaction\_E, Access\_Credit, Yield\_Payment, Bills\_to\_Pay, Profitability, and Insolvency. Likewise, 4 correlations between variables.

Four relevant correlations were identified: Payment performance with capital, Profitability, Business Satisfaction, and Insolvency with Activity Time (Months\_A). These correlations are determining factors for business bankruptcy.

There is a highly significant relationship between payment performance and capital, a moderate positive correlation was obtained ( $\rho = .492$  and  $p = .000$ ); That is, as the company has greater capital, then it will obtain a higher payment yield to meet its medium and long-term obligations. The value of  $\rho = .492$  indicates that there is a moderate positive correlation. There is a highly significant relationship between distance from home and satisfaction, a low positive correlation was obtained ( $\rho = .392$  and  $p = .000$ ); That is to say, the longer the activity time, the higher the profitability. The value of  $\rho = .392$  indicates a low positive correlation. The value of  $\rho = .392$  indicates a low positive correlation.

There is a highly significant relationship between age and satisfaction, a low positive correlation was obtained ( $\rho = .216$  and  $p = .006$ ); That is, as the company has more time in activities, it will present greater business satisfaction. The value of  $\rho = .216$  indicates a low positive correlation. There is a highly significant relationship between age and performance, a low positive correlation was obtained ( $\rho = .210$  and  $p = .000$ ); That is, a relationship is identified between companies that have less time in activities with greater insolvency. The  $\rho$  value = .210 indicates that there is a low positive correlation.

The model optimization neural network algorithm surpassed the metrics of the other algorithms considered in the present investigation where a precision of 97%, accuracy of 97%, sensitivity of 99% and f1-score of 98% were obtained.

## 7. FUTURE WORK

For future work, the incorporation of variables that represent significant aspects of the company and the market in general should be prioritized. In this way, the model will obtain results closer to 100%. Likewise, consider the study of the identified correlations to determine the factors that determine a greater or lesser relationship between the variables and their causes for the treatment of continuous improvement.

In that context, it is recommended to explore other methodologies besides KDD, such as CRISP-DM and SEMMA to obtain a more complete view and consider different approaches in data mining, which could result in a deeper understanding of the data and generate results. more detailed and specific. It is important to incorporate and consider automation in the process of predicting business bankruptcy through an intelligent system. In this way, for future studies the model would train automatically and generate more specific and complete results. In addition to this, a comparison could be made of other algorithms: research, which would help in the evaluation of a better capacity of the model for prediction, in this way scope and possible limitations could be established with additional information about alternative approaches and their advantages or disadvantages in a specific model. Finally, it is important to establish comparative studies with other institutions that present or share similar characteristics. This approach will not

only help refine the accuracy of the model at the company level, but will also make it possible to obtain common patterns that can be applied in various contexts and situations, even more so in companies of larger size and scope.

## REFERENCES

- [1] Alsheref, F.K., Fattoh, I.E., M. Ead, W. (2022). Automated prediction of employee attrition using ensemble model based on machine learning algorithms. *Computational Intelligence and Neuroscience*, 2022(1): 7728668. <https://doi.org/10.1155/2022/7728668>
- [2] Brenes, R.F., Johannssen, A., Chukhrova, N. (2022). An intelligent bankruptcy prediction model using a multilayer perceptron. *Intelligent Systems with Applications*, 16: 200136. <http://dx.doi.org/10.1016/j.iswa.2022.200136>
- [3] Bussmann, N., Giudici, P., Marinelli, D., Papenbrock, J. (2021). Explainable machine learning in credit risk management. *Computational Economics*, 57(1): 203-216. <https://doi.org/10.1007/s10614-020-10042-0>
- [4] Calabrese, R. (2023). Contagion effects of UK small business failures: A spatial hierarchical autoregressive model for binary data. *European Journal of Operational Research*, 305(2): 989-997. <https://doi.org/10.1016/j.ejor.2022.06.027>
- [5] Cantarero, S., González-Loureiro, M., Puig, F. (2017). Efectos de la crisis económica sobre el emprendimiento en empresas de economía social en España: Un análisis espacial. *REVESCO. Revista de Estudios Cooperativos*, Madrid, España: Universidad Complutense de Madrid, (125): 24-48. <http://dx.doi.org/10.5209/REVE.56133>
- [6] Tran, H., Le, N., Nguyen, V.H. (2023). Customer churn prediction in the banking sector using machine learning-based classification models. *Interdisciplinary Journal of Information, Knowledge & Management*, 18: 087-105. <https://doi.org/10.28945/5086>
- [7] Muthukumaran, K., Hariharanath, K. (2022). Deep learning enabled financial crisis prediction model for small-medium sized industries. *Intelligent Automation & Soft Computing*, 35(1): 521-536. <https://doi.org/10.32604/iasc.2023.025968>
- [8] Muñoz-Cancino, R., Bravo, C., Ríos, S.A., Graña, M. (2023). On the combination of graph data for assessing thin-file borrowers' creditworthiness. *Expert Systems with Applications*, 213: 118809. <https://doi.org/10.1016/j.eswa.2022.118809>
- [9] Dake, D.K., Buabeng-Andoh, C. (2022). Using machine learning techniques to predict learner drop-out rate in higher educational institutions. *Mobile Information Systems*, 2022: 1-9. <https://doi.org/10.1155/2022/2670562>
- [10] Guerranti, F., Dimitri, G.M. (2022). A comparison of machine learning approaches for predicting employee attrition. *Applied Sciences*, 13(1): 267. <https://doi.org/10.3390/app13010267>
- [11] Kanzari, A., Rasmussen, J., Nehler, H., Ingelsson, F. (2022). How financial performance is addressed in light of the transition to circular business models—A systematic literature review. *Journal of Cleaner Production*, 376: 134134. <https://doi.org/10.1016/j.jclepro.2022.134134>
- [12] Zhang, W., Yan, S., Li, J., Tian, X., Yoshida, T. (2022).

- Credit risk prediction of SMEs in supply chain finance by fusing demographic and behavioral data. *Transportation Research Part E: Logistics and Transportation Review*, 158: 102611. <https://doi.org/10.1016/j.tre.2022.102611>
- [13] Nicodemo, C., Satorra, A. (2022). Exploratory data analysis on large data sets: The example of salary variation in Spanish Social Security Data. *Business Research Quarterly*, 25(3): 83-294. <https://doi.org/10.1177/2340944420957335>
- [14] Pratt, M., Boudhane, M., Cakula, S. (2021). Employee attrition estimation using random forest algorithm. *Baltic Journal of Modern Computing*, 9(1): 49-66. <https://doi.org/10.22364/bjmc.2021.9.1.04>
- [15] Pandey, B., Veeramanickam, M., Ahmad, S., Rodriguez C., Esenarro, D. (2023). ExpSSOA-Deep maxout: Exponential Shuffled shepherd optimization based Deep maxout network for intrusion detection using big data in cloud computing framework. *Computers & Security*, 124: 102975. <https://doi.org/10.1016/j.cose.2022.102975>
- [16] Sulastrri, R., Janssen, M. (2022). The elements of the Peer-to-peer (P2P) lending system: A systematic literature review. In *Proceedings of the 15th International Conference on Theory and Practice of Electronic Governance*, Guimarães Portugal, pp. 424-431. <https://doi.org/10.1145/3560107.3560172>
- [17] Yahia, N.B., Hlel, J., Colomo-Palacios, R. (2021). From big data to deep data to support people analytics for employee attrition prediction. *IEEE Access*, 9: 60447-60458. <https://doi.org/10.1109/ACCESS.2021.3074559>
- [18] Raza, A., Munir, K., Almutairi, M., Younas, F., Fareed, M.M.S. (2022). Predicting employee attrition using machine learning approaches. *Applied Sciences*, 12(13): 6424. <https://doi.org/10.3390/app12136424>
- [19] Yao, G., Hu, X., Wang, G. (2022). A novel ensemble feature selection method by integrating multiple ranking information combined with an SVM ensemble model for enterprise credit risk prediction in the supply chain. *Expert Systems with Applications*, 200: 117002. <https://doi.org/10.1016/j.eswa.2022.117002>
- [20] Zou, Y., Gao, C., Gao, H. (2022). Business failure prediction based on a cost-sensitive extreme gradient boosting machine. *IEEE Access*, 10: 42623-42639. <https://doi.org/10.1109/ACCESS.2022.3168857>
- [21] Lee, J., Jang, D., Park, S. (2017). Deep learning-based corporate performance prediction model considering technical capability. *Sustainability*, 9(6): 899. <https://doi.org/10.3390/su9060899>