sciendo Journal of Social and Economic Statistics

Vol. 12, No. 1, Summer 2023

DOI: 10.2478/jses-2023-0005

OPEN ACCESS CC BY NC ND

DOES BOARD GENDER DIVERSITY REALLY IMPROVE FINANCIAL PERFORMANCE AND DEFAULT RISK? EVIDENCE FROM ROMANIAN COMPANIES ENGAGED IN INTERNATIONAL TRADE

Coralia TANASUICA[a*]

[a)] *The Bucharest University of Economic Studies, Romania*

Abstract

Currently, in the business world, important attempts are being made to assess the possibility of verifying the creditworthiness of companies by going beyond traditional financial data and incorporating alternative data as well. Does the fact that the board of a company has a certain structure influence the financial health of that company? The paper's objective is to ascertain if there is a correlation between the board of directors' gender diversity, financial performance, and the probability of default for that respective company. The empirical study examined a sample of Romanian companies engaged in international trade. Using natural language processing techniques, I predicted the gender of the respective director by his/her first name. The second step of the research was the analysis of the correlation between the percentage of women on the board and traditional financial indicators such as profit or turnover, and the correlation between the percentage of women on the board and that company's probability of default. The results show that there is generally not a strong correlation between the percentage of women on the board of the company and the other financial and risk indicators at the entire population level, but there is a strong correlation for some specific industries such as education, meaning that there are industries were the presence of women in board really impacts the performance of the business.

Keywords: Board Behaviour, Diversity, Corporate Default, Gender, Correlation
JEL Classification: C25, G32, J16, M14

[*] Corresponding author, **Coralia Tanasuica** – tanasuicacoralia22@stud.ase.ro

1. Introduction

Many fields have tried to identify the existence of behavioural patterns that differentiate women from men, mainly in terms of their actions but also in terms of the impact of these decisions in the near or further future. Mara Mather (Mather, 2012) and Ruud van den Bos (van den Bos, 2009) have carried out independent research on the decision-making behaviours of women and men under stressful conditions, to identify whether or not there are distinct patterns of behaviour of the two clusters. The bilateral conclusions of the two researchers were that, indeed, decisions made at times of stress display different structure for the two clusters and that they are indeed correlated with the gender of the decision-maker. Male decision-makers tend to accept higher risks and typically focus on very high rewards, but with a very low probability of success compared to female decision-makers, for whom risk aversion is increased as they typically pursue smaller but safer rewards. Therefore, as both papers conclude, women are more risk-averse than men and tend to act more cautiously. The process of decision-making for women will take longer than for men. However, it is true that this result is obtained at the individual level and that a decision taken at the board level, which has a distinct ratio of women and men, will no longer be related to one individual decision, but to a group of individuals deciding together.

This is the direction on which this article focuses, trying to bring the situation in Romania to the forefront, and identifying the behaviour of distinct boards with a distinct concentration of women. The objective of the study is to identify if the board decisions that affect the profitability of the company is corelated with the percentage of the women in board. It also analyses whether there is a correlation between the default risk and the percentage of women in the company board.

The growth of computing power in the last few years proved to be the impulse in experimenting new methods and the usage of new predictors in order to increase the accuracy and performance of credit risk models, which are no longer fully dependent on financial data, but can also incorporate alternative data such as social data, transactional data, climate data and gender data.

This paper adds value especially through the identification of business areas where the increasing of the number of women on boards has no significant influence on probability of default and on company's profitability or where the influence is not a positive one. In these situations, companies should take into account the efficiency of the business, not the trend of ticking the box for gaining a certain number of women on the board. The study also revealed that the size of the company is an important factor in assessing the correlation between the percentage of women on the board, the probability of bankruptcy and the profitability of the company.

The study continues as follows: in the second section, I present a summary of the related literature and in the third section, the types of methods and models used in the present research. In the fourth section, I present the data set, the indicators calculated, the accuracy rates, and the results of the case study. The main results indicate no evident correlation at the level of the entire population. However, there are strong correlations in certain sectors such as P (Education), Q (Health and social care) or K (Financial intermediation and insurance). The last

section of the research includes the conclusions, the limitations of the present work and the future research directions on this topic.

2. Literature review

Over the years, many researchers have tried to identify the impact that board diversity has on a company's profitability. In recent years, researchers have paid more attention to this topic and many studies have been published.

One of the papers that is trying to detect the correlation between gender diversity and company performance focuses on the Malaysian market (Marimuthu, 2009). Board diversity is measured by analysts both in terms of gender mix and ethnic diversity. The sample analysed is small: 100 non-financial companies. The method that the researchers have used to determine the impact of the independent variables on the dependent variables is OLS regressions used on the cross-sectional data. The result of the analysis is partially surprising: gender diversity is not correlated with the financial performance of the company, but ethnic diversity is correlated.

Another paper, based on a sample of 442 companies from India (Akshita, 2021), attempts to identify whether there is a correlation between the percentage of women on the board and the financial performance of the companies. The data sample does not include companies from the financial sector, but instead contains companies from most major business areas. Dynamic panel data analysis techniques were used for the analysis. The results were contrary to those obtained on the Malaysian data sample, with the researchers in this case finding that the inclusion of women on the board of directors has a strong impact on the financial performance of companies. Another hypothesis they tested was that the inclusion of women on the board of directors has a strong impact for higher dividend payouts, but this hypothesis was disproved.

In terms of the results obtained, one of the most interesting articles analyses a larger sample of data, across 26,000 private companies in the UK (Sattar, 2022). The aim of this research is to highlight the effect of company board diversity on the company risk. The researchers quantify risk based on three variables: standard deviation of earnings before interest and taxes (EBIT) deflated by total assets; difference between maximum return on assets (ROA) and minimum ROA; and standard deviation of the difference between the observed firm’s ROA and the yearly industry average ROA of all firms. The results obtained are oriented in several directions: there is a negative correlation between the percentage of women on the board and the company risk, but this result is more evident in small and medium-sized companies. The presence of women on the board reduces volatility, but the company risk is lower only if there are local women on the board. If there are women of other ethnicities on the board, the risk is higher, which is mainly due to a better understanding of the local market held by local women.

Women are more likely to be ethical and more risk-averse than men (Saona, 2018), concludes one study conducted on several highly-developed European countries: Denmark, Finland, France, Germany, Italy, Norway, Portugal, Spain, Sweden, and United Kingdom. Considering these countries, the study tries to highlight if there is any effect of the presence of women in the board of companies on the practices of earnings management. The sample consists of 1269 non-financial, private companies from the observed countries. For the purpose of this analysis, the authors used panel data techniques. The result of the study is that the higher

the percentage of women in the board of directors, the lower the practices of earnings management in the analysed European countries is.

Another study highlighting the importance of women on the board on financial performance is based on a sample of companies in Turkey (Merve, 2016). The sample of companies analysed is small and it includes a number of 149 non-financial companies listed on the Istanbul stock exchange. One of the hypotheses tested by the researchers is that the percentage of women in the company board has a significant and positive impact on the following performance indicators: the return on assets (ROA), the return on equity (ROE) and the return on sales (ROS). The methodology used is the regression analysis and the result confirms the hypothesis put forward. The business conclusion of the final result is that women directors communicate positive signals to the company's shareholders, building a positive image of the company and maintaining a competitive advantage.

Interesting conclusions can be drawn from a research study based on a sample of companies from Spain, which investigates the relationship between the financial performance of 117 corporations listed on the Madrid Stock Exchange (Rodriguez-Dominguez, 2010) and the proportion of women on the companies' boards. The researchers also analysed the relationship in terms of the business context, taking into account: the working environment (hostile or favourable), the proportion of women on the board (high or low), the complexity of the production process (high or low) and the educational level of the women on the board. The conclusion of the analysis was that there is a positive correlation between the proportion of women on the board and the financial performance of the company in those sectors where the working environment is favourable in sense of work conditions, where there is a small number of women on the board, the production process is complex and women have a high level of education. A positive impact on company performance was also found for companies with an equally balanced board.

Another study, focused on the U.S. market of real estate investment trust companies, highlights that gender diversity has a positive impact on market performance (Schrand, 2018), but a negative impact on operational performance. The analysis was done on a sample of 112 companies and the methodology was based on a two-stage Heckman approach applied on the panel dataset. The approach was done in two steps; the first step was to generate the probability for each company to have at least one woman on the board, and the second step focused on analysing the correlation between company performance and the gender mix on the board of directors.

A study in Spain provided important evidence on a specific area of the economy: the agricultural sector. The study observes 672 agricultural cooperatives (Minguez Vera, 2019) and tries to highlight the effect of gender mix on returns, debt level and economic risk. The result shows that those companies with a higher percentage of women on the board have a higher level of returns, a lower level of debt, but also have a higher economic risk. For these analyses they used the ordinary least squares as methodology.

We also have evidence of the positive influence of the presence of women on the board from Kuwait. The study (Issa, 2019), although analysing a small volume of companies of only 23 companies listed on the Kuwait stock exchange, follows the same methods and techniques found in all other scientific research focused on this field, using two-step system generalised method of moments (GMM) estimation in order to determine the impact of board gender

diversity on firm value. The result is one that supports the importance of diversity and inclusion of women on the board of directors: board gender diversity is positively associated with firm value.

It is also worth mentioning the results of a study done on Norwegian and Danish listed companies (Marinova, 2015) which tries to detect whether board gender diversity has a positive effect on firm performance. The study is done on a small sample of companies: 186 listed companies in 2007, of which 102 are Dutch companies listed on Euronext Amsterdam and 84 Danish companies listed on OMX Nordic Exchange Copenhagen. The result of the research is contradictory to other results obtained in the context of developed European countries, namely: there is no relation between board diversity and firm performance. However, the result is not necessarily conclusive, given the small amount of data and the fact that data collected for a study published in 2007 is now relatively old.

There is an interesting theory in the literature, named the Critical Mass Theory (Charles, Redor, & Zopounidis, 2015), which explains that if the number of women on the board increases to reach a certain critical mass, this will not significantly improve corporate governance. Egerová (2019) tests this theory based on a sample of 232 small and medium-sized companies in the Czech Republic. The sample contains only companies active in the manufacturing sector, as this is the most important industry in the Czech Republic. The research attempts to test the existence of the correlation between upper management diversity and the financial performance of the organisation. The methods used in the analysis are: Spearman's correlation coefficients for all indicators and cluster analysis using k-means for better clarifying the relationships between all indicators. The data sample used in this study was ideal to demonstrate the theory mentioned above: one of the conclusions of their research was that increasing the number of women in top management above the critical mass, leads to poorer financial performance. The other result was that, indeed, there is a positive correlation between the mix of board gender and the financial performance of the company, but with the earlier limitation (of critical mass).

It is important to remark that the impact of the presence of women on the board can also be conditioned by the distribution of the organisation's financial performance. A study of the top 100 largest and most well-established companies in America (Charles, 2018) finds that firms that perform poorly and moderately are those that benefit most from women on the board. By contrast, the most successful companies do not necessarily benefit from gender diversity on the board. The authors believe that this is because the firms that perform well require less monitoring, and an increase in the number of women on the board would result in potentially excessive levels of monitoring, which would have a negative impact on the company performance level.

The above study is supported in terms of results by another study, this time conducted on a sample of French companies, a small sample of 105 companies (Dang, 2016), which considers the top companies in terms of market capitalization and the trading volume. The result of the study shows that the gender mix of the board, negatively affects financial performance when the company performs very well, and positively when the company performs poorly.

There is a positive impact of the presence of women on board over the company's financial performance also in Romania. This is the conclusion of a study conducted on a sample of 57 companies (Belenesi, 2022) listed on the Bucharest Stock Exchange. Although the sample is

small, the results converge to the conclusion of the majority of the researchers that the presence of women on the board positively impacts the financial state of the company.

3. Methodology

There are various methodologies and techniques adopted by researchers in studying the impact of women on board members on the financial performance of companies and on the risk of the organisation. Most studies use: panel data analysis (Saona, 2018; Rodriguez-Dominguez, 2010; Schrand, 2018; Belenesi, 2022), regression models (Marimuthu, 2009; Saona, 2018; Merve, 2016; Rodriguez-Dominguez, 2010; Minguez Vera, 2019; Charles, 2018; Dang, 2016; Lenard, 2014) or basic correlation and descriptive statistics (Egerová, 2019; Charles, 2018; Dang, 2016). A smaller number of authors attempt to address the subject using machine learning models, such as segmentation (Egerová, 2019).

The topic was approached using supervised predictive modeling tools, but also unsupervised modeling and deep learning techniques. To identify correlations between variables, I used the correlation matrix, based on the calculation of Pearson correlation coefficients. The construction of the correlation matrix was performed using python, namely the Pandas library. These analyses were performed in order to include in the supervised algorithm only the uncorrelated predictors. To assess the risk of bankruptcy of companies, I used supervised machine learning techniques. Specifically, these were binary classification models. I tested several models in terms of performance and accuracy indicators to select the optimal algorithm in terms of accuracy. The top-performing algorithms in this study were boosting algorithms, which belong to the ensemble modeling technique. Boosting algorithms aim to transform weak learners into strong learners by combining multiple weak classifiers. This technique involves constructing a model using a series of weakly performing models. Initially, a model is built, and subsequent models are developed to correct the errors made by the previous ones. This iterative process continues until the training data set is deemed accurate and comprehensive or until the maximum number of models is reached.

Regarding the databases used for training the algorithms, only the public data available on the Romanian Ministry of Finance was used, such as the Defaulted Companies Database and the databases containing the yearly financial statements for all companies in Romania for the period of time from 2008 until 2021.

Next, I will briefly describe one of the algorithms that have proven to be one of the most efficient in terms of both performance indicators and running time. The algorithm is named LightGBM (LGBM) and it is created by the researchers from Microsoft. LGBM is an effective implementation of gradient-boosted decision trees (GBDT), a popular ensemble method that combines decision trees as weak learners in a sequential manner, known as boosting. LGBM can handle both discrete and continuous data for tasks like classification and regression. It employs a leaf-wise construction strategy, prioritizing the expansion of promising leaves and identifying optimal split points during each iteration. A small number of such cutting-points can lead to massive information loss, but also can reduce overfitting, and a large number of such cutting-points is inefficient in terms of high computational cost. LGBM is extremely useful in this respect, because it divides the data into a fixed number of bins of uniform length, like a histogram, at the level of each independent variable. The algorithm then iterates through

these bins to identify the optimal cutting-points. This translates into much lower computational costs and very good performance.

The indicators used to evaluate the model performance are obtained from the confusion matrix. A confusion matrix can be obtained as follows: the dataset is divided into two parts: training set and test set. Then the model is trained using the data from the train set. Once the model is trained, we use it to predict the test dataset. After obtaining the predicted data, we compare them with the true data from the test set and group the volumes of data in the following groups:

- TP (True-positives): Where the actual label for that column was "Yes" in the test dataset and our model also predicted "Yes".
- TN (True-negatives): Where the actual label for that column was "No" in the test dataset and our model also predicted "No".
- FP (False-positives): Where the actual label for that column was "No" in the test dataset but our model predicted "Yes".
- FN (False-negatives): Where the actual label for that column was "Yes" in the test dataset but our model predicted "No".

Starting from these values, we can calculate the most important performance indicators:

- Accuracy of the model – is defined as the ratio of correctly classified examples among all examples

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{2}$$

- Precision of the model – this is the division of the positive cases correctly identified by all the cases identified as positive (including false-positives).

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

- Recall - (or sensitivity): this is the division of the true-positives by all the real positive cases (including false-negative cases). This matters when you are trying to have more true-positives than true-negatives.

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

- F1 Score – this is a way of combining the precision and recall of the model, and it is defined as the harmonic mean of the model's precision and recall

$$F1 = 2 \times \frac{Precision \times Recall}{Precision+Recall} \tag{5}$$

The accuracy has the advantage that it is very easily interpretable, but the disadvantage that it is not robust when the data is unevenly distributed. In this case, the F1 Score is more robust and it is mostly used when the data are strongly unbalanced.

For text data pre-processing, the Natural Language Toolkit (NLTK) was used for lowercase conversion and for encoding the characters to numbers.

To identify the gender from the first name, were used deep learning technologies. A LSTM (Long-Short Term Memory) neural network was built. This type of neural network is a subtype of recurrent neural networks (RNN). This type of network is used with maximum applicability in the field of language processing. The problem with RNNs is that they only have the ability to have a short memory, retaining only the information prior to the current neuron. This problem is solved by LSTM networks, which have the ability to retain a longer period of information. The special architecture of LSTM allows the algorithm to decide which previous information retained in short-term memory will be retained and which will be discarded. For building and training the LSTM network, the python language was used, namely the Keras library. For network optimization, was used ADAM (adaptive moment estimation) for updating the network weights iteratively, based on the training dataset. It is an optimizer that outperforms classical stochastic gradient descent optimizers, both in terms of low memory and resource requirements, thus being computationally efficient, and in terms of being suitable for large problems in terms of number of terms and/or parameters. For training the algorithm, I build a dictionary of names, including Romanian names. Each name has the correspondent label according to the gender: female or male.

Unsupervised machine learning algorithms were used for segmenting the data. Within this class of algorithms, were used hierarchical clustering algorithms. These algorithms are of two types: agglomerative and divisive. For this study I chose to use agglomerative clustering. For this technique, initially, each data point is considered to be an individual cluster. At each new iteration, similar clusters are grouped into new clusters. The process ends when all clusters have been merged and the final result is a single cluster, which includes all data points. A dendrogram is employed for visualizing the clustering process and assisting in the determination of the optimal number of clusters. To evaluate the performance of the model or the segmentation efficacity, which translates into high inter-cluster variance and a lower intra-cluster variation, we used the Silhouette coefficient. The formula of this coefficient for a data sample is:

$$\frac{b-a}{\max(a,b)} \tag{6}$$

Where:

b = the mean distance between a sample and all other points in the next nearest cluster

a = average intra-cluster distance (The mean distance between a sample and all other points in the same cluster).

The final Silhouette coefficient is the average of all individual sample coefficients. The closer the coefficient is to 1, the better the population is segmented. A coefficient close to 0 indicates that there are clusters that overlap, so the model was not able to segment the population correctly. These algorithms were utilized to identify patterns in the data and discern whether distinct behaviors exist within different groups or levels within the population.

4. Case Study

The first dataset used for detecting the board gender mix contains all the companies in Romania that reported foreign trade activities in the period 2016 to 2022. In total, the data sample contains a volume of 52,000 unique companies. The dataset used in this study is made up of public data obtained from the Romanian Chamber of Commerce and Industry. The dataset contains a lot of variables, most of which are text variables: company name, company registration number, the fiscal code, import sections, other countries with whom the company has commercial transactions, export sections, countries to which the company exports, address, phone number, e-mail address, web site, and board members' names.

The variables used in the first stage of the study to identify the gender composition of the board were the company's tax code and the names of the board members. There are 28,000 companies with board data available for all years, that are active in 2021 and which have all financial information available. The number of distinct board members' first names for this sample of companies (first names for NLP) includes 40,000 distinct names.

The distribution of companies included in the data set on the geographical level can be seen in the following map (Figure no.1):

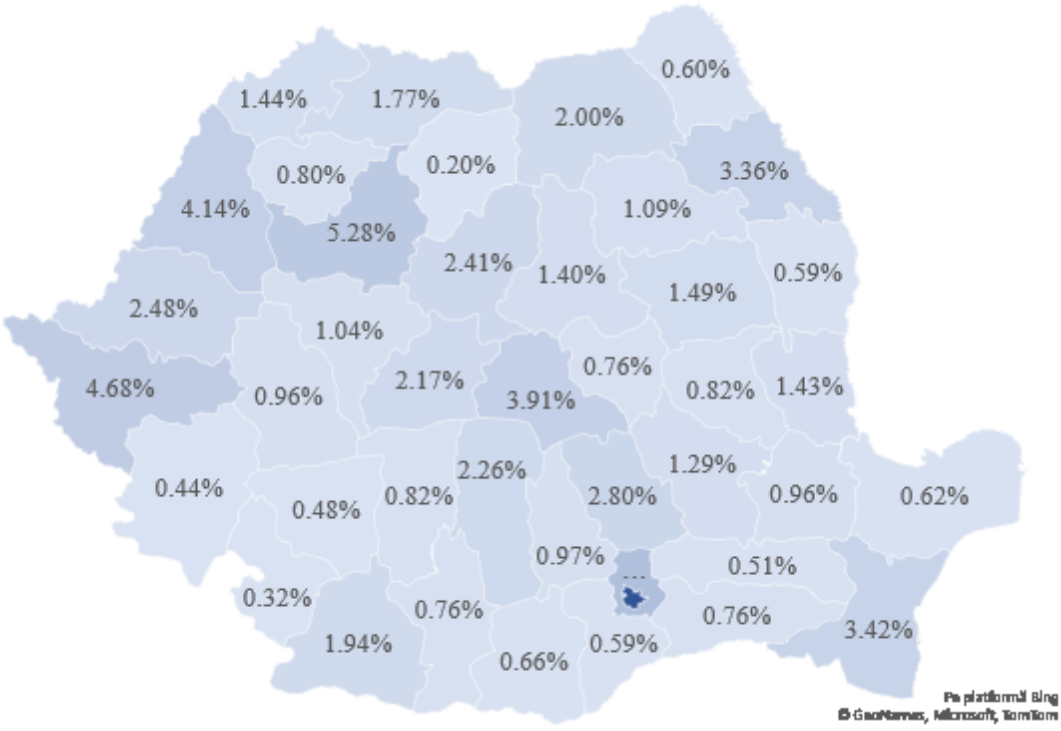


Figure no.1: Geographical distribution of Romanian companies engaged in international trade
Source: Author representation using data from Catalogul Exportatorilor si Importatorilor din Romania, (2022) Eighth Edition, CCIR, using Microsoft Excel

The highest volume of companies is located in Bucharest: 28.36%, followed by Ilfov county with 7.12%, Cluj County with 5.28%, and Timis county with 4.68% from the total volume of companies included in the analysed sample.

To process the text data, the Python text processing library were used: NLTK (natural language toolkit). The library was used for pre-processing, for transforming all characters to lowercase, then for transforming each first name into a character vector. After this process, the

vectors were transformed into vectors of equal length using the padding method. Due to the fact that the neural network works with numeric vectors, the character encoding method was used, so the vectors were transformed into numeric vectors.

In order to train the LSTM model, I created a dictionary of Romanian and foreign names. The dictionary contains female first names and male first names from Romania and abroad. On this dictionary, the neural network has been trained to learn how the characters follow each other in a female name and in a male name. By applying the trained model on the real dataset, the one containing the first names of all board members, the correspondent gender was able to be assigned with an accuracy of 93%. A visual structure of the flow will be added in the following, for a simpler overview of the process. (Figure no.2):

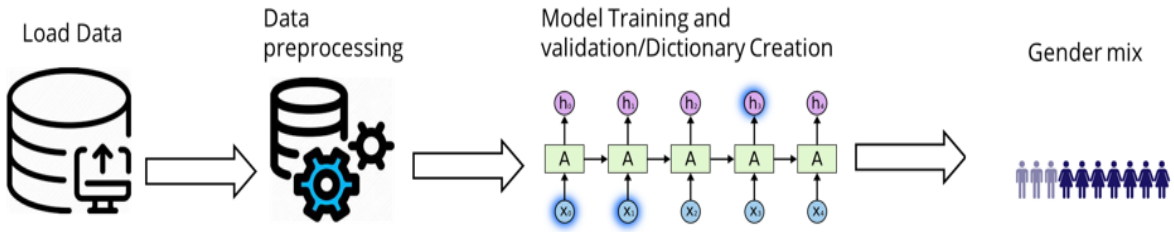


Figure no.2: Steps for getting from data to the output mix of board gender
Source: Author's representation

In Figure no. 3, the data pre-processing process (stage 1 to stage 4) was briefly described for the data to be prepared for training the model. In stage 5, I present the results of the model, which predict the gender for the real data:

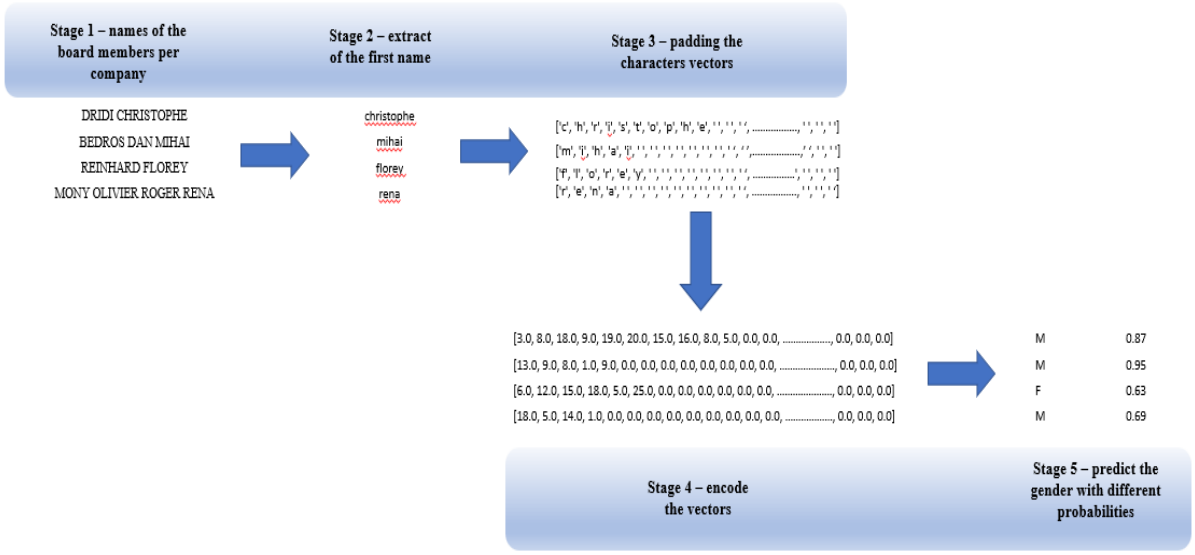


Figure no.3: Data preparation (NLP)
Source: Author's representation

With a brief descriptive analysis of the data, we can get an idea of how the gender mix of the board is distributed nationally by each type of economic sector Table no. A3. Figure no. 4 shows precisely this distribution:

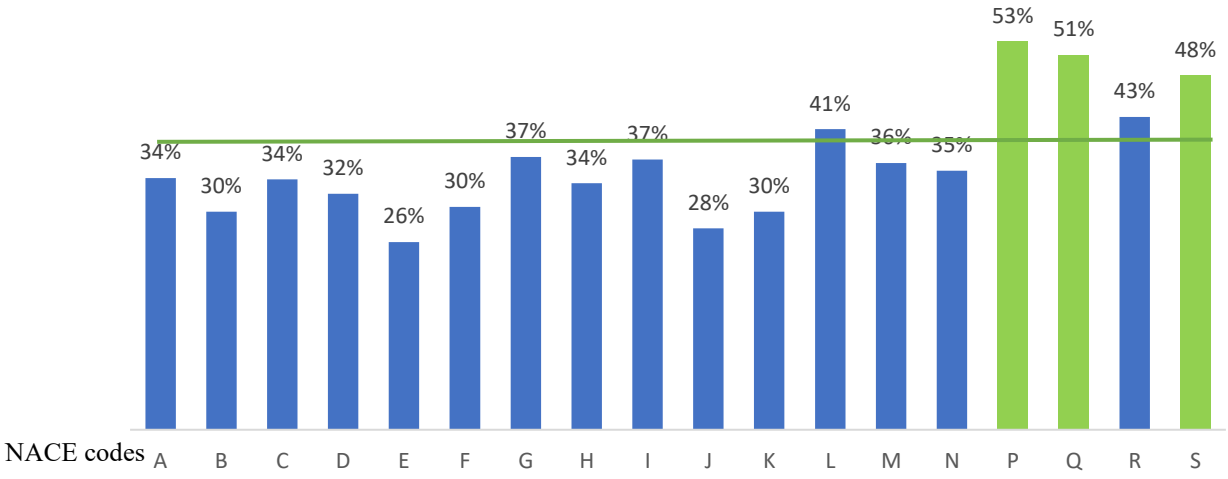


Figure no.4: Average women percentage per industry 2020

Source: Author representation using data from Catalogul Exportatorilor si Importatorilor din Romania, (2020) Sixth Edition, CCIR, using Microsoft Excel

It is obvious that in 2020, there were sectors in the Romanian economy (considering only the sample of companies active in international trade) where women had, on average, a higher representation on the board than man. These are the sectors: P - Education, and Q - Health and social care.

Much more interesting is to see how the mix of women in the board evolved from 2020 to 2021, per each industry Table no. A3 (Figure no. 5).

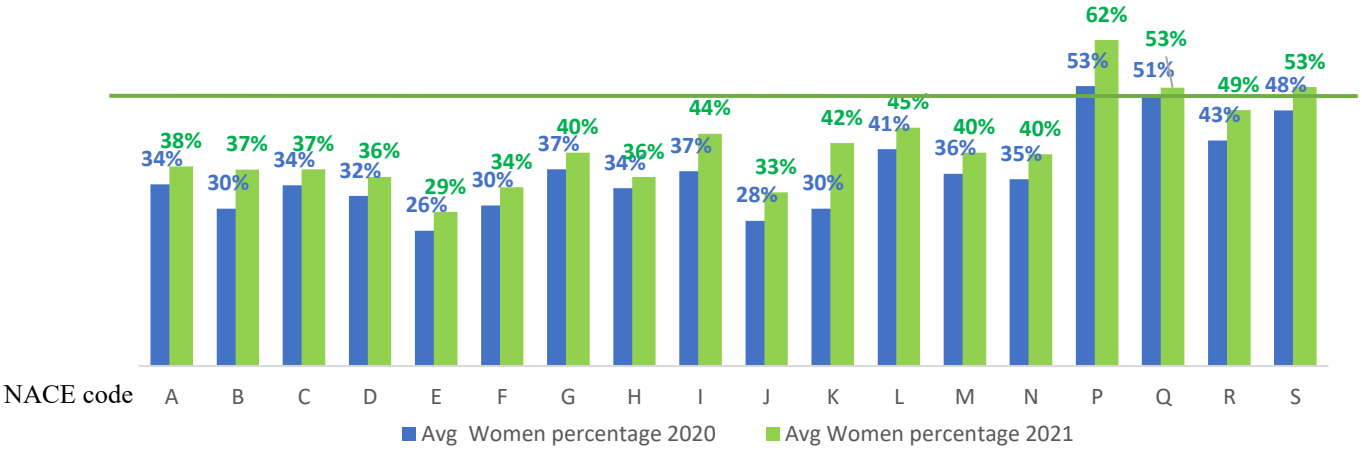


Figure no.5: Average % of Board Women per Industry evolution 2020 to 2021

Source: Author representation using data from Catalogul Exportatorilor si Importatorilor din Romania, (2020) Sixth Edition, CCIR and from Catalogul Exportatorilor si Importatorilor din Romania, (2021) Seventh Edition, CCIR using Microsoft Excel

The number of women on the company's boards increased, on average, for each economic sector, in Romania, in 2021 compared to 2020 (referring to the companies active in international trade).

Regarding the financial indicators corresponding to each company, these data were also extracted from public data sources, including the Romanian trade register, and are represented by the yearly financial statements for each company in Romania from 2008 until 2021. All the

databases used are referenced in the bibliography. The data regarding the companies that defaulted in Romania from 2008 until 2022 was also extracted, using the database of the companies registered with the Trade Registry until July 7th, 2022. These data were used to train supervised machine learning models to generate the probability of bankruptcy for each active company, for the following two years. The variables used are: number of employees, non-current assets, current assets, inventories, receivables, cash and bank, prepaid expenses, payables, advanced payments, provisions, equity, net turnover, total income and shared capital.

The financial databases were extracted for the respective companies for the period 2008 to 2021. Based on this information and the databases containing information on companies that have gone bankrupt, several supervised machine learning algorithms were trained for classifing the company's binary: those that will go bankrupt in the next two years and those that will not. The output of the model is the probability that a company will go bankrupt in the next two years. The model that performed best on this data sample was a boosting model: the LightGBM, which produced 99% accuracy. Considering the skewed nature of data, with significantly fewer companies going bankrupt compared to those remaining active, I opted to use the F1 score as the evaluation metric for the models. The selected model achieved an F1 score of 83%.

The probability of the company's default between 2021 and 2022 will be further named as P2021, and P2022 will refer to the probability that the company will default anytime between 2022 and 2023. To obtain these two distinct outputs, at the industry level Table no. A3, two sets of data spanning one year were used as input for the model. That means that, for the second probability, the financial data for the fiscal year 2021 were added.

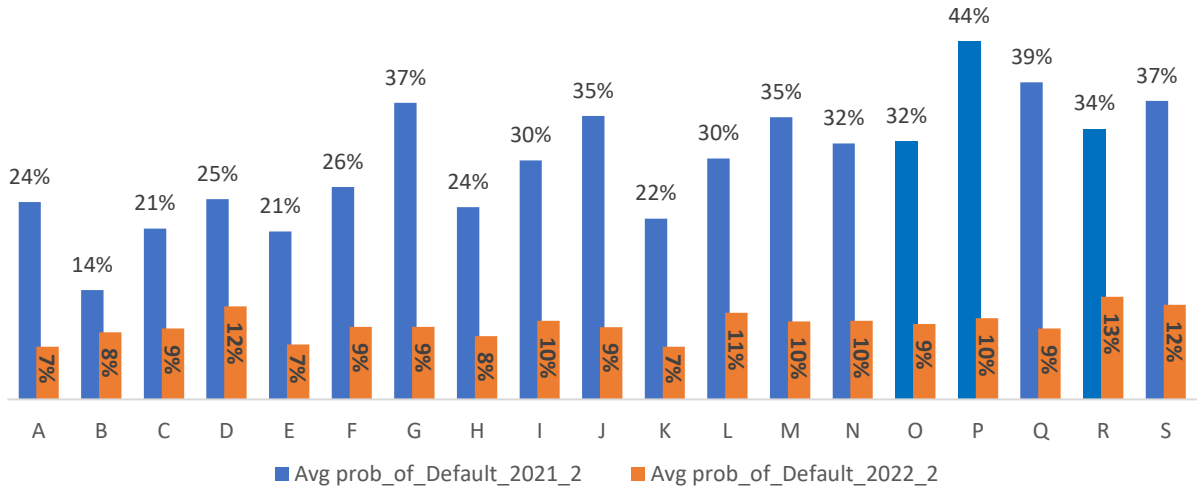


Figure no. 6: Average probability of default in 2020 vs 2021

Source: Author's representation using data resulted from the ML models, trained on data from data.gov.ro/situatii_financiare 2008 to 2021, using Microsoft Excel

Considering these new available probabilities of default for each company, the prediction of the probability of default for each company included in the population analysed for the current study was run, at two distinct periods of time. In Figure no. 6, the average probability of default per industry was observed, for the population analysed. The probability of default is decreasing, on average, for each industry in the Romanian economy. This conclusion is applicable only for this specific population, which is built from companies engaged in international trade between Romania and other countries. At the national level, the outcome

will be different. This behaviour can be translated through an economic revival of international trade. It appears that the war in Ukraine has had a positive impact on Romanian companies that trade internationally, improving their financial situation and economic potential. The lowest average probability of default for 2022 is identified for the K NACE code (Financial intermediation and insurance), meaning that Financial Institutions, included in the research data sample, have the lowest probability of defaulting.

If we take specific samples from the population: the companies that defaulted in 2022 and the companies that are still active and compare the proportion of the women in the board for the analysed companies as an average evolution for the period between 2016-2021, we will have the visual outputs in Figure no. 7.

The increase in the proportion of women in the still active companies is tempered, there is an ascending trend, but with a low amplitude. The companies that defaulted in 2022 started to decrease the proportion of women in the board starting with 2017.

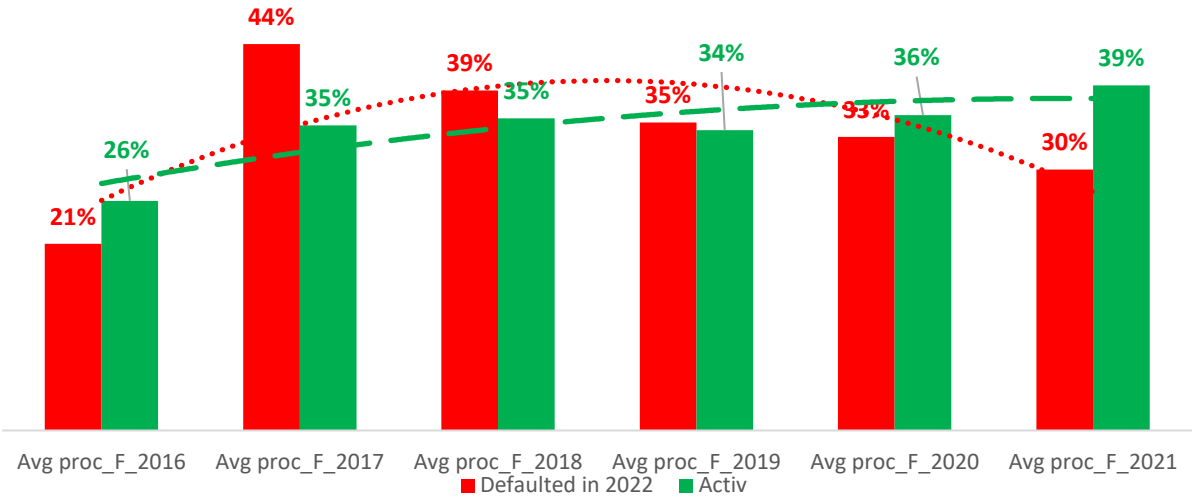


Figure no. 7: Percentage of Women in Board for companies active versus companies that defaulted in 2022

Source: Author's representation using data from Catalogul Exportatorilor si Importatorilor din Romania, (2016-2021), CCIR using Microsoft Excel

In order to identify whether there is a correlation between the percentage of women on the company board and the other variables analysed, the correlation matrix was constructed using the Pearson correlation coefficient. The variables used in the analysis were: number of employees in the company, annually, over the period 2016-2021; total company revenue, annually, over the period 2016-2021; net turnover, annually, over the period 2016-2021; equity, at company level, annually, over the period 2016-20121 and the probability of bankruptcy of each company: P_{2021} and P_{2022}. The corelation coefficients can be reviewed in Table no. A1. In Table no. A1, we can see that there is no correlation between the mix of gender in the board and the other indicators. The highest level of corelation, still very low, can be seen for the first variable P_{2021}.

This was a surprising result, but did not discourage the further development of the study. I split the population by the NACE codes and calculated the correlation matrix for the population included in each individual NACE code, to see if there is a different behaviour at the industry level. This experiment converged to some important outcomes. In Table no.

A2, the correlation matrix for the P NACE code is presented. There is strong correlation between the presence of women in the board and the other indicators. It is obvious the change of behaviours that COVID brought about: the change in the correlation before 2020 and after 2020. The more women there are on the board, the higher the total income is, but until 2020, when the direction of the correlation changed.

Strong correlation was identified in the following economic sectors: Education, Health and social work activities, Entertainment, Cultural and Recreational Activities, Hotels and Restaurants, Real Estate Transactions, Financial Intermediation and Insurance, Administrative and Support Service Activities and Production and Supply of electricity, gas, heat, hot water and air conditioning. On the opposite side are the following sectors: Agriculture, Forestry and Fisheries, Manufacturing, Construction, Wholesale and Retail Trade; Repair of Motor Vehicles and Motorcycles, Transport and Storage and Information and Communication, where no correlation was found.

Another direction of the study was the segmentation of the population for identifying potential patterns that can direct the research on specific clusters of companies from the analysed population. I used an unsupervised machine learning algorithm for generating the clusters of companies: Hierarchical Clustering. For deciding the optimal number of clusters, I used the Silhouette score. The Coefficient for a set of samples is given as the mean of the Silhouette Coefficient for each sample.

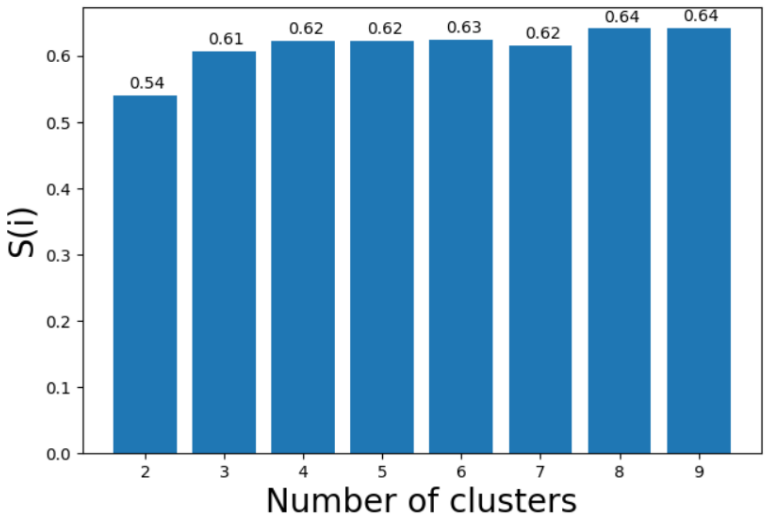


Figure no. 8: Silhouette score for different number of clusters

Source: Author's representation in Python: sklearn & matplotlib using data from Catalogul Exportatorilor si Importatorilor din Romania, (2016-2021), CCIR and data resulted from the ML models, trained on data from data.gov.ro/situatii_financiare (2008 to 2021)

There is an important increase in the score when we pass from 2 clusters to 3. There is also an improvement of the score when deciding to create 4 clusters. If we decide to split the data in more than 4 clusters, there isn't any improvement until reaching 8 clusters. I choose to segment the population in 4 clusters (the silhouette score for 4 clusters = 0.62).

For segmenting the population, we can use one or more dimensions. For the present study, I decided to use 3 dimensions, especially for being able to present the results in a visual format. The dimensions I've chosen for segmenting the population are: P_{2022}, the percentage of women in the board for 2021 and the net profit or loss for 2021. One way of deciding which are the

best dimension to choose for the segmentation is to have a general view of the histograms of distribution of the values for all the potential variables (Figure no. 9)

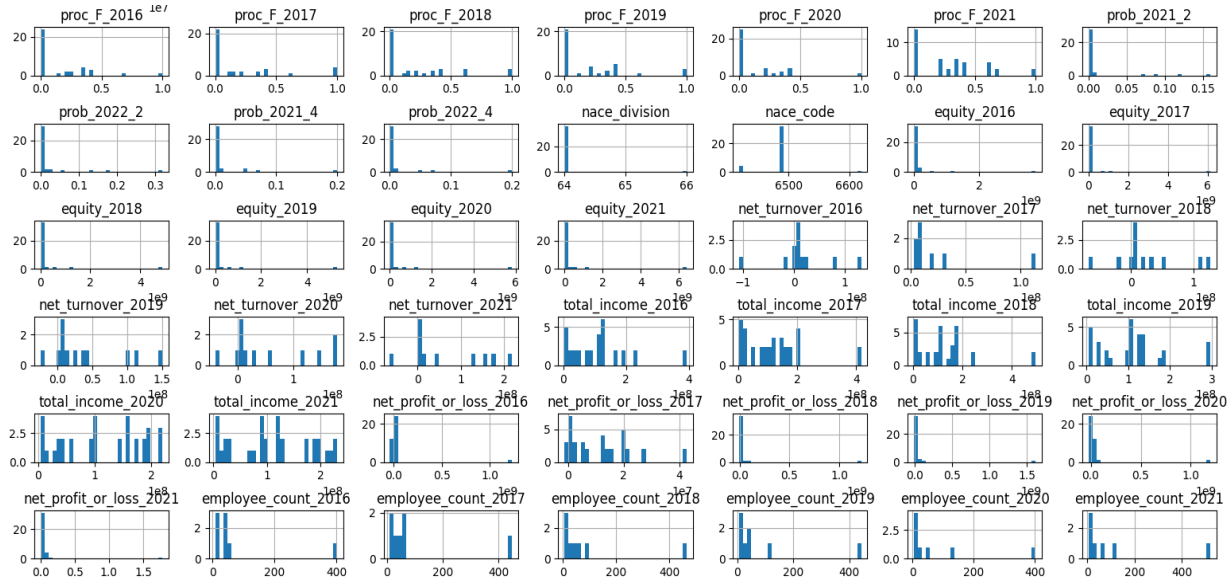


Figure no. 9: Histograms of distribution for distinct variables included in the construction of the correlation matrix

Source: Author's representation in Python: sklearn & matplotlib using data from Catalogul Exportatorilor si Importatorilor din Romania, (2016-2021), CCIR and data resulted from the ML models, trained on data from data.gov.ro/situatii_financiare (2008 to 2021)

From the above set of histograms, we can see that for the variable profit or loss for 2021, the distribution is not concentrated on one or two values, but are spread over an interval of distribution.

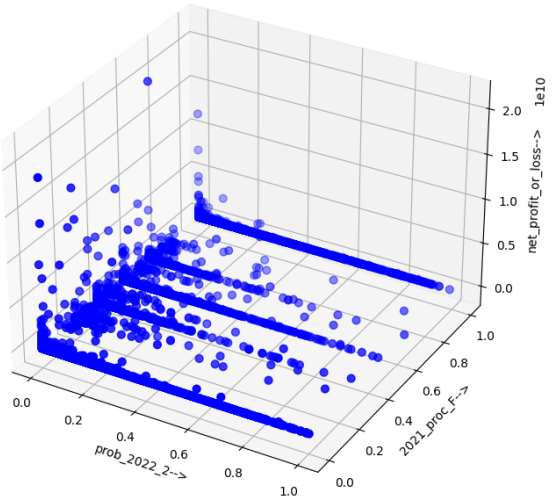


Figure no. 10: Distribution of companies in a 3-dimensional space

Source: Author's representation in Python: sklearn & matplotlib using data from Catalogul Exportatorilor si Importatorilor din Romania, (2016-2021), CCIR and data resulted from the ML models, trained on data from data.gov.ro/situatii_financiare (2008 to 2021)

In a three-dimensional space (built based on the direction chosen for simulating the segmentation): P_{2022}, the percentage of women in the board for 2021, and the net profit or

loss for 2021. The spread of the data can be visualised in Figure no. 10. For making sense of the data, the technique of clustering is very important.

After running the clustering model, the data was visualised in the three-dimensional space of the three indicators that were used for the segmentation. This was done to assess the quality of the segmentation and the efficiency it can bring in subsequent descriptive analyses, primarily by reducing dimensionality and by generating customer segments identifiable by common patterns (Figure no.11, Figure no.12).

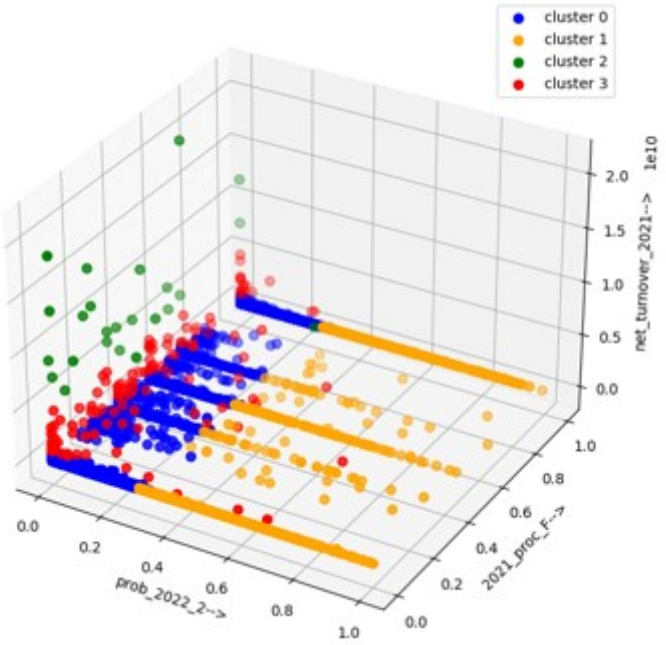


Figure no. 11: Distribution of companies in a 3-dimensional space, after clustering the population

Source: Author's representation in Python: sklearn & matplotlib using data from Catalogul Exportatorilor si Importatorilor din Romania, (2016-2021), CCIR and data resulted from the ML models, trained on data from data.gov.ro/situatii_financiare (2008 to 2021)

Now we have the population clustered in 4 segments. Going deeper inside each cluster for identifying the common patterns inside the clusters and the differentiators between distinct clusters, we will have the descriptive statistics in Table no. 1. Clusters are labelled from 0 to 3.

Table no. 1: Descriptive statistics for each cluster

Values/Cluster No.	0	1	2	3
No. of companies	26,194	1,867	24	138
Avg % of women in board '20	35%	36%	25%	32%
Avg % of women in board '21	39%	40%	34%	39%
Avg P2021	32%	37%	0%	0%
Avg P2022	4%	48%	5%	7%
Avg net_turnover_2020	22,825,113	10,349,505	8,339,000,218	1,573,064,614
Avg net_turnover_2021	25,419,788	9,362,426	9,931,831,970	1,947,658,164
Avg total_income_2020	24,134,439	11,312,968	8,549,895,435	1,674,993,290
Avg total_income_2021	26,824,026	9,965,052	10,150,070,832	2,038,368,526
Avg net_profit_or_loss_2020	1,249,696	(163,011)	237,401,145	54,703,051
Avg net_profit_or_loss_2021	1,839,385	(501,461)	313,727,837	72,212,190

Avg employee_count_2020	48	31	5,316	2,156
Avg employee_count_2021	46	28	5,223	2,040
Avg equity_2020	9,335,717	2,663,091	1,878,875,640	538,457,250
Avg equity_2021	10,375,078	1,818,227	2,173,707,734	578,703,094

Source: Author's calculus using data from Catalogul Exportatorilor si Importatorilor din Romania, (2020-2021), CCIR and data resulted from the ML models, trained on data from data.gov.ro/situatii_financiare (2008 to 2021) using Microsoft Excel

How can each cluster be interpreted? The first cluster (cluster 0) contains small companies, with a high proportion of women on the board, and with a high probability of default. In the second cluster (cluster 1), we have the very small companies, with the highest proportion of women on the board and a high probability of defaulting. In the third cluster (cluster 2), we can find the big corporates with a low proportion of women on the board, and a low probability of default. The final cluster (cluster 3) contains the medium companies, with a high proportion of women in board, and with a low probability of defaulting.

5. Conclusions

The gender composition of company boards was examined in this study, utilizing a comprehensive dataset comprising 52,000 distinct companies in Romania actively involved in foreign trade activities spanning from 2016 to 2022. This dataset encompassed a range of variables, including pertinent company details, trade activities, and the names of board members. Leveraging natural language processing techniques, a Long Short-Term Memory (LSTM) model was trained, achieving an impressive accuracy rate of 93% in predicting the gender of board members. Geographically, the distribution of these companies revealed a concentration in Bucharest, followed by Ilfov, Cluj, and Timis counties. Notably, sectors such as Education and Health and Social Care exhibited a higher representation of women on boards compared to men in the year 2020. Furthermore, there was an observed increase in the average number of women serving on company boards in 2021 as compared to the preceding year, reflecting a positive trend toward greater gender diversity in corporate leadership.

The study examined the relationship between financial indicators and the probability of bankruptcy, utilizing publicly available data sources to extract key financial metrics, including the number of employees, revenue, net turnover, equity, and other relevant factors. Machine learning models were then employed to predict the probability of bankruptcy for each company, with the LightGBM model attaining an impressive accuracy rate of 99%. Notably, financial institutions demonstrated the lowest probability of defaulting, suggesting their inherent stability and resilience. Additionally, an intriguing observation emerged regarding companies that eventually defaulted in 2022, as their boards exhibited a declining representation of women starting from 2017. This finding highlights a potential association between board composition and financial performance, warranting further investigation into the underlying dynamics and implications for corporate governance.

The segmentation of the analysed population provides interesting insights about the impact of women's presence on company boards. The algorithm classified very small and small companies into distinct groups as having the highest number of women on the board, but also the highest probability of bankruptcy. This does not indicate the existence of causality between the number of women and the probability of bankruptcy, because it is natural that a small company is more unstable and has a higher probability of bankruptcy.

This aspect supports the findings of the research conducted on 100 companies in America (Charles, 2018), who identified poorly performing companies as those that attract women on the company board. The segmentation performed in the study classifies medium-sized companies (in terms of turnover) as those that still benefit from women on the board, but at a lower proportion than small companies. These companies are more stable, with a much lower probability of bankruptcy. Large companies, corporations, were grouped in a separate cluster having the lowest proportion of women on the board, but also the lowest risk of bankruptcy. These results support the findings of the study mentioned above (Charles, 2018): that large and stable firms do not benefit from having a large number of women on the board, precisely because the companies are stable and do not need extra monitoring, which could negatively affect the business.

Another output of the research is that there is no correlation between the percentage of women on the board and the probability of the company of defaulting at the entire population level, but there are sectors where there is strong correlation, such as: P (Education), Q (Health and social care), R (Performing, cultural and recreational activities), S (Other service activities), N (Administrative and support service activities), L (Real estate transactions), K (Financial intermediation and insurance).

Some insights of the P NACE code (Education):

- The higher the proportion of women on the board, the lower the probability of default
- The higher the proportion of women on the board, the higher the number of employees
- The higher the proportion of women on the board, the higher the turnover
- The higher the proportion of women on the board, the lower the profit
- Some insights of the K (Financial intermediation and insurance):
- The higher the proportion of women on the board, the higher the probability of default
- The higher the proportion of women on the board, the lower the turnover
- The higher the proportion of women on the board, the lower the number of employees
- The higher the proportion of women on the board, the higher the profit

Some implications of the study are related to the significance of a higher percentage of women in specific sectors, notably the education sector. It is possible that their training and skills may differ from those of women involved in other fields, but it is evident that strategies adopted by a board with a higher percentage of women in Education will be focused on the social and human resources aspects. Within these companies, characterized by a predominantly female board in the education sector, there is a notable emphasis on expanding the workforce, leading to increased turnover due to company growth and investments. However, it is important to note that this expansion may result in a lower level of profit.

The behavior of predominantly female boards, in the financial industry, highlights a completely different approach. In this industry, it seems that women struggle to generate

investments, resulting in lower turnover compared to companies with a predominantly male board. Failing to create an environment conducive to attracting investments, financial companies with a predominantly female board tend to lay off more employees, operating with a smaller workforce and generating higher profits. All these factors increase the likelihood of bankruptcy for these specific companies.

If the question has been asked so far not only about the impact of gender, but also about the cumulative impact of gender and education level on decisions that impact a business, I would go a step further and suggest that variations in behaviour may exist among individuals with the same level of education but different academic backgrounds (schools), assuming that these backgrounds influence behaviours. It is one of the directions that will be addressed in a future paper and the idea started from the difference in decision-making of women in education and finance sectors (even if only at the level of the population currently analysed) Understanding the nuances of decision-making within different sectors can provide valuable insights. The case study highlights the contrasting behaviours of predominantly female boards in the education and finance industries. In educational institutions, a higher proportion of women on the board positively correlates with company growth, reflected in increased employee count, turnover, and investments. Conversely, in the finance sector, a higher proportion of women on the board negatively correlates with the number of employees and turnover, potentially restricting business operations.

This suggests that the impact of gender diversity on business outcomes is not uniform across industries. Different sectors may require distinct approaches and strategies to leverage the benefits of gender diversity effectively. By considering both the presence of women on boards and the sector-specific correlations, organizations can make informed decisions about gender diversity initiatives tailored to their specific industry context.

The limitations of the present research are, first, that it analyses correlation, not causation. Second, it should take into account other potential factors such as nationality, age, level of education or the countries with which transactions are made. As for future developments, it is intended to resume the analysis on at least one industry, which would include as a data sample all companies operating in the industry or industries concerned.

References

[1]. Arora A., 2021. Gender diversity in boardroom and its impact on firm performance, *Journal of Management and Governance,* 26, pp.735-755.

[2]. Bogdan, V., Popa, D.-N. and Beleneşi, M., 2022. The Complexity of Interaction between Executive Board Gender Diversity and Financial Performance: A Panel Analysis Approach Based on Random Effects. *Complexity*, 2022, 1–20, Article ID 9559342, https://doi.org/10.1155/2022/9559342

[3]. Catalogul Exportatorilor si Importatorilor din Romania, 2016 – 2021, Second to Eighth editions, CCIR, Available at: https://ccir.ro/catalogul-exportatorilor-importatorilor [Accessed 20 July 2022].

[4]. Charles, A., Redor, E. and Zopounidis, C., 2015. The Determinants of the Existence of a Critical Mass of Women on Boards: A Discriminant Analysis, *Economics Bulletin*, 35 (3), pp.185-197. https://audencia.hal.science/hal-01188269/

[5]. Charles, A., Dang, R. and Redor, E., 2018. Board Gender Diversity and Firm Financial Performance: A Quantile Regression Analysis. *International Corporate Governance and Regulation (Advances in Financial Economics, 20)*, Emerald Publishing Limited, Bingley,15–55. https://doi.org/10.1108/S1569-373220180000020002
[6]. DATA.GOV.RO ed., 2022. Situații Financiare 2018 – 2021. [online] data.gov.ro. Available at: https://data.gov.ro/dataset/situatii_financiare_<year> [Accessed 20 August 2022].
[7]. Issa, A., Hanayshab, J.R., Elfeky, M.I. and Ullah, I., 2019. The Impact of Board Gender Diversity on Firm Value: Evidence from Kuwait. *International Journal of Applied Science and Research*, 2 (1). http://research.skylineuniversity.ac.ae/id/eprint/99/1/39.pdf
[8]. Kılıç, M. and Kuzey, C., 2016. The Effect of Board Gender Diversity on Firm Performance: Evidence from Turkey. *Gender in Management: An International Journal*, 31 (7), 434–455. https://doi.org/10.1108/GM-10-2015-0088
[9]. Hernández-Nicolás, C.M., Martín-Ugedo, J.F. and Mínguez-Vera, A., 2019. The Effect of Gender Diversity on the Board of Spanish Agricultural Cooperatives on Returns and Debt: An Empirical Analysis. *Agribusiness*, 35 (4), 639–656. https://doi.org/10.1002/agr.21608
[10]. Jane Lenard, M., Yu, B., Anne York, E. and Wu, S., 2014. Impact of Board Gender Diversity on Firm Risk. *Managerial Finance*, 40 (8), 787–803. https://doi.org/10.1108/MF-06-2013-0164
[11]. Marimuthu, M. and Kolandaisamy, I., 2009. Ethnic and Gender Diversity in Boards of Directors and Their Relevance to Financial Performance of Malaysian Companies. *Journal of Sustainable Development*, 2 (3). 139–148. http://scholars.utp.edu.my/id/eprint/12045
[12]. Marinova, J., Plantenga, J. and Remery, C., 2015. Gender Diversity and Firm Performance: Evidence from Dutch and Danish Boardrooms. *The International Journal of Human Resource Management*, 27 (15), 1777–1790. https://doi.org/10.1080/09585192.2015.1079229
[13]. Mather, M. and Lighthall, N.R., 2012. Risk and Reward Are Processed Differently in Decisions Made under Stress. *Current Directions in Psychological Science*, 21 (1), 36–41. DOI: 10.1177/0963721411429452. PMID: 22457564; PMCID: PMC3312579.
[14]. Reddy, S. and Jadhav, A.M., 2019. Gender Diversity in Boardrooms – a Literature Review. *Cogent Economics & Finance*, 7 (1), 1644703. DOI: 10.1080/23322039.2019.1644703 https://doi.org/10.1080/23322039.2019.1644703
[15]. Rodríguez-Domínguez, L., García-Sánchez, I.-M. and Gallego-Álvarez, I., 2010. Explanatory Factors of the Relationship between Gender Diversity and Corporate Performance. *European Journal of Law and Economics*, 33 (3), 603–620. https://doi.org/10.1007/s10657-010-9144-4
[16]. Sattar, M., Biswas, P.K. and Roberts, H., 2022. Board Gender Diversity and Firm Risk in UK Private Firms. *Global Finance Journal*, 54, 100766. https://doi.org/10.1016/j.gfj.2022.100766
[17]. Saona, P., Muro, L., San Martín, P. and Baier-Fuentes, H., 2019. Board of Director's Gender Diversity and Its Impact on Earnings Management: An Empirical Analysis for Select European Firms. *Technological and Economic Development of Economy*, 25 (4), 634–663. https://doi.org/10.3846/tede.2019.9381

[18]. Schrand, L., Ascherl, C. and Schaefers, W., 2018a. Gender Diversity and Financial Performance: Evidence from US Reits. *Journal of Property Research*, 35 (4), 296–320. https://doi.org/10.1080/09599916.2018.1549587
[19]. van den Bos, R., Harteveld, M. and Stoop, H., 2009. Stress and Decision-Making in Humans: Performance Is Related to Cortisol Reactivity, Albeit Differently in Men and Women. *Psychoneuroendocrinology*, 34 (10), 1449–1458. DOI: 10.1016/j.psyneuen.2009.04.016

Annex

Table no. A1: Companies' data for 2016-2021 period

	proc_F 2016	proc_F 2017	proc_F 2018	proc_F 2019	proc_F 2020	proc_F 2021
prob_2021_2	7%	6%	5%	6%	5%	4%
prob_2022_2	2%	0%	0%	1%	1%	1%
equity_2016	-1%	-1%	-1%	-1%	-1%	-1%
equity_2017	-1%	-2%	-1%	-1%	-1%	-1%
equity_2018	-1%	-2%	-2%	-1%	-1%	-1%
equity_2019	-1%	-1%	-2%	-1%	-1%	-1%
equity_2020	-1%	-2%	-2%	-1%	-1%	-1%
equity_2021	-1%	-1%	-2%	-1%	-1%	-1%
net_turnover_2016	-2%	-3%	-2%	-2%	-2%	-1%
net_turnover_2017	-2%	-3%	-2%	-2%	-2%	-1%
net_turnover_2018	-2%	-2%	-2%	-2%	-1%	-1%
net_turnover_2019	-2%	-3%	-2%	-2%	-2%	-1%
net_turnover_2020	-3%	-3%	-2%	-2%	-2%	-1%
net_turnover_2021	-2%	-2%	-2%	-2%	-2%	-1%
total_income_2016	-3%	-3%	-2%	-2%	-2%	-1%
total_income_2017	-3%	-3%	-2%	-2%	-2%	-1%
total_income_2018	-2%	-2%	-2%	-2%	-1%	-1%
total_income_2019	-3%	-3%	-2%	-2%	-2%	-1%
total_income_2020	-3%	-3%	-2%	-2%	-2%	-1%
total_income_2021	-2%	-2%	-2%	-2%	-2%	-1%
net_profit_or_loss_2016	-1%	-1%	0%	-1%	-1%	0%
net_profit_or_loss_2017	-1%	-1%	-1%	-1%	-2%	-1%
net_profit_or_loss_2018	-2%	-2%	-1%	-1%	-2%	-1%
net_profit_or_loss_2019	-1%	-1%	-1%	-1%	-1%	0%
net_profit_or_loss_2020	-1%	-1%	-1%	-1%	-2%	-1%
net_profit_or_loss_2021	-1%	-1%	-2%	-2%	-2%	-1%
employee_count_2016	-2%	-3%	-3%	-2%	-2%	-1%
employee_count_2017	-3%	-3%	-3%	-3%	-2%	-1%
employee_count_2018	-2%	-3%	-2%	-2%	-2%	-1%
employee_count_2019	-3%	-3%	-3%	-2%	-2%	-1%
employee_count_2020	-3%	-3%	-3%	-2%	-2%	-1%
employee_count_2021	-2%	-3%	-3%	-2%	-2%	-1%

Data source: Author's calculus using data from Catalogul Exportatorilor si Importatorilor din Romania, (2016-2021), CCIR and data resulted from the ML models, trained on data from data.gov.ro/situatii_financiare (2008 to 2021) using Python Pandas

Table no. A2: Companies' data for 2016-2021 period

	proc_F 2016	proc_F 2017	proc_F 2018	proc_F 2019	proc_F 2020	proc_F 2021
prob_2021_2	48%	48%	48%	48%	14%	14%
prob_2022_2	-39%	-39%	-39%	-39%	-37%	-37%
prob_2021_4	23%	23%	23%	23%	-19%	-19%
prob_2022_4	23%	23%	23%	23%	-19%	-19%
equity_2016	-16%					
equity_2017	-10%	-10%				
equity_2018	-28%	-28%	-28%			
equity_2019	-26%	-26%	-26%	-26%		
equity_2020	10%	10%	10%	10%	-12%	
equity_2021	24%	24%	24%	24%	8%	8%
net_turnover_2016	26%					
net_turnover_2017	21%	21%				
net_turnover_2018	27%	27%	27%			
net_turnover_2019	35%	35%	35%	35%		
net_turnover_2020	78%	78%	78%	78%	-17%	
net_turnover_2021	20%	20%	20%	20%	17%	17%
total_income_2016	31%					
total_income_2017	19%	19%				
total_income_2018	34%	34%	34%			
total_income_2019	34%	34%	34%	34%		
total_income_2020	92%	92%	92%	92%	-22%	
total_income_2021	26%	26%	26%	26%	14%	14%
net_profit_or_loss_2016	-39%					
net_profit_or_loss_2017	-56%	-56%				
net_profit_or_loss_2018	-57%	-57%	-57%			
net_profit_or_loss_2019	-59%	-59%	-59%	-59%		
net_profit_or_loss_2020	49%	49%	49%	49%	-25%	
net_profit_or_loss_2021	-14%	-14%	-14%	-14%	-4%	-4%
employee_count_2016	4%					
employee_count_2017	52%	52%				
employee_count_2018	60%	60%	60%			
employee_count_2019	48%	48%	48%	48%		
employee_count_2020	33%	33%	33%	33%	-3%	
employee_count_2021	15%	15%	15%	15%	-5%	-5%

Data source: Author's calculus using data from Catalogul Exportatorilor si Importatorilor din Romania, (2016-2021), CCIR and data resulted from the ML models, trained on data from data.gov.ro/situatii_financiare (2008 to 2021) using Python Pandas

Table no. A3: NACE codes for the Romanian industries

NACE	Explanation
A	Agriculture, forestry and fishing
B	Extractive industry
C	Manufacturing industry
D	Production and supply of electricity and heat, gas, hot water and air conditioning
E	Water distribution; sanitation, waste management, decontamination activities
F	Construction
G	Wholesale and retail trade; repair of motor vehicles and motorcycles
H	Transportation and storage
I	Hotels and restaurants
J	Information and communications
K	Financial intermediation and insurance
L	Real estate transactions
M	Professional, scientific and technical activities
N	Administrative and support service activities

Data source: https://caen.ro/sectiuni