# ANTICIPATING FAILURE: A COMPREHENSIVE ANALYSIS OF ENTREPRENEURSHIP DYNAMIC FACTORS USING MACHINE LEARNING PREDICTIVE MODELS

**Rachid Alami[1], Agata Stachowicz-Stanusch[2], Sugandha Agarwal[3], Turki Al Masaeid[4]**
1. School of Business, Canadian University Dubai, UAE.
rachid.alami@cud.ac.ae
ORCID: 0000-0003-4280-9660
2. School of Business, Canadian University Dubai, UAE.
agata@cud.ac.ae
ORCID: 0000-0001-6991-1222
3. School of Business, Canadian University Dubai, UAE.
drsugandhaagarwaal@gmail.com
ORCID: 0000-0002-1077-876X
4. Abu Dhabi School of Management, Abu Dhabi, - UAE.
turkimasaeid@gmail.com
ORCID: 0000-0002-7901-5656

**Abstract**
This study explores how machine learning models, such, as Support Vector Machines (SVM) Decision Trees, Logistic Regression, Random Forest, Ensemble Models and Neural Networks can predict the failures of startups. It highlights the impact of bureaucracy, engagement of human resources, financial capacity and mentorship and coaching within organizations on startups operating in the business environment of Morocco. The correlation analysis indicates that traditional methods need to be reevaluated to anticipate challenges faced by organizations due to the lack of established patterns. The research showcases how machine learning provides flexibility and valuable insights beyond correlation analysis. Specifically Random Forest and Ensemble Models emphasize the importance of bureaucracy and human capital in forecasting business success. Variations in rankings highlight connections that stress the need to comprehend all factors. The complexity of bureaucracy is depicted by its role in both facilitating and hindering progress. Human resources play a role in demonstrating their contributions to the organization. This study underscores the significance of capital and financial resources in overcoming obstacles and promoting growth. These discoveries are set to have implications for strategies while deepening our comprehension of business dynamics, within Morocco.

**Keywords:** Entrepreneurship; Machine Learning; Predictive models; Startups; Decision Trees; Neural Networks; Ensemble Models

## Introduction

Machine learning explores various methods to predict the outcomes of economic activities. Research conducted by Xia and Taos in 2017 demonstrates how cluster algorithms and data mining can be utilised to predict a company's likelihood of going bankrupt. Dergiades and

Milas (2017) employed machine learning and media analysis techniques to forecast market competition dynamics. In a study by Rahman and colleagues (2018), the significance of utilising machine learning to evaluate competitive landscapes was emphasised. Several methods can be utilised in entrepreneurship to evaluate the effectiveness of a project. In a recent study, Piskunova and colleagues (2022) proposed the utilisation of machine learning algorithms to forecast the time gap between project initiation and funding, considering financial metrics, industry attributes, and technical components. In a recent study, Daskalaki and colleagues (2019) enhanced the precision of their predictions for company failures through the application of machine learning techniques.

In a study conducted by Fathi et al. (2017), they enhanced the precision of their forecasts by incorporating ensemble learning and feature selection techniques. According to Klapper, Amit, and Guillén (2010), having enough financial resources is essential for starting new businesses. Aidis et al. (2012) emphasise the importance of effective governance as a critical aspect. Klapper, Amit and Guillén (2010) highlighted the importance of having resources when starting businesses while Aidis et al. (2012) underscored the role of governance.

Kibler and his team (2014) conducted a research study, on this topic focusing on how networks and cultural values influence entrepreneurship. Khan and his colleagues 2020 study delves into the dynamics of business ventures, resource availability and cultural factors that impact activities. Challenges such as bureaucracy, corruption and limited financial resources are factors to consider in this context. Understanding entrepreneurship in Morocco involves examining the influences on success or failure. Laabissi et al. (2019) and Lahrech et al. (2016) have explored aspects of entrepreneurship including perceptions, gender disparities in pursuits as well as the integration of innovation and technology. The educational backgrounds of entrepreneurs shape their perspectives and abilities significantly. Moreover further research is needed to investigate how machine learning algorithms can predict business outcomes for ventures in Morocco effectively. While there are existing research papers on dynamics within the countrys business landscape leveraging machine learning tools could unveil patterns to Moroccos circumstances. This discussion encourages exploring models customized to conditions using machine learning techniques, for anticipating setbacks or failures in business endeavors. Past studies have extensively examined utilizing machine learning to forecast business performance with results.

A research was carried out on factors that lead to success, in developing countries specifically focusing on the challenges encountered by entrepreneurs. Moreover it delved into the applications of machine learning, in environments and highlighted the increasing significance of utilizing this technology for forecasting setbacks. The study delves into how machine learning can support business owners in addressing obstacles they might face.

**Literature review**

Several prediction models have been created to determine the elements that impact the success or failure of startups in various nations, such as Lussier and Claudia (2010) and Mayr et al., 2020. However, the models in issue have certain inherent limitations. They focus primarily on the latter phases of the startup process rather than the entire development time (Reynolds, 2017). They feel this is the most crucial moment. Prior studies often used a firm-centric approach, gathering data from the proprietors of firms that are either operational or closed

(Lussier, 1995). This method has been implemented several times. Linear and logistic regression algorithms have been the foundational basis for classic prediction models. (Antretter et al., 2019; Van Witteloostuijn & Kolkman, 2019) Recent entrepreneurship research has used machine learning (AI) techniques to analyse several elements of emerging businesses' dynamics. This study was done to have a deeper understanding of the dynamics of new businesses. Loureiro et al. (2018) found that artificial intelligence regression models outperformed traditional regression models.

Moreover, they can efficiently identify discrepancies in startup outcomes and enhance forecast accuracy. Machine learning has made significant progress in predicting the success or failure of businesses in many commercial fields using a variety of applications and approaches. Zavgren (1984) and Beaver (1975) completed influential investigations now considered classic landmark research. These investigations have motivated several replication and extension studies, with most crediting the time-series CUSUM approach and comprehensive prediction research conducted during that period. Recent articles have made significant scientific improvements. For example, Xia and Tao (2017) improved bankruptcy prediction using data mining and rapid clustering methods. This invention enhanced the precision of bankruptcy forecasting.

Zhang and Luo (2015) used social media analytics and machine learning to predict organisational competitiveness. Rahman et al. (2018) emphasise the importance of machine learning models in assessing a company's competitiveness via predictive analytics in the digital age. Several variables influence the success of small enterprises in the Moroccan context. El Alami et al. (2019) emphasised the importance of possessing significant financial resources. Research by Robichaud et al. (2023), Hind and Jamal (2023), and Bouhaj et al. (2022) explores the many challenges firms face when trying to get financing for projects in Morocco. Entrepreneurs encounter challenges such as government non-assistance, complex bureaucratic finance application processes, and banks' scepticism about startups' ability to fulfil their pledges. Kamel et al. (2021) performed research assessing the effects of government schemes offering financial aid and emphasised the importance of government backing and mentorship initiatives. Morocco has established entrepreneurial initiatives to promote innovation and entrepreneurship. The courses seek to provide job opportunities and assist new businesses in enhancing their abilities. Maroc PME is the institution responsible for creating and providing access to "Programme Moussanada," which provides mentoring, financial aid, and training to small businesses. The Moroccan Agency fully funds this non-profit organisation to promote small- and medium-sized enterprises. Lahrech et al. (2016) emphasised the importance of emerging enterprises investing in the creative and innovative development of their products. Haddoud and Bouazza (2020) and Bouazza and El Kadiri (2018) found that the education level significantly impacts the success and development of businesses.

### *Entrepreneurship failure and machine learning prediction*

Machine learning techniques have been used across several sectors to predict the outcomes of businesses. An example of this would be the work done by Vasquez et al. (2023) and Lee et al. (2019), who used machine learning techniques to construct a model of forecasting for the success of entrepreneurs in the information technology industry. The suggested method enhanced performance in seven distinct machine learning algorithms, resulting in an average accuracy gain of 21.75% and a precision increase of 11.69%. The technique's execution led to an 88% accuracy rate and an accuracy of 82%. Da Silva et al. (2023) use machine learning methods on a dataset covering time from 2015 to 2019 to forecast the likelihood of a business in the hotel industry of the Iberian Peninsula collapsing. The study's findings suggest that machine learning models created with the dataset effectively forecast organisational failure. This is shown by examining the influence of these approaches on companies in Portugal and Spain. Veganzones (2023) recommends using a combination of advanced learning machines to enhance predictive accuracy for business failure issues. This is done to enhance the precision of the forecast. The enhanced extreme learning machine demonstrated the most significant performance improvement, as the experiments conducted on French businesses showed. Assagaf and his team researched in 2023 to investigate deep learning techniques, specifically focusing on the Multi-LayerMulti-Layer Perceptron, for predicting failures in several businesses. The study results indicate that deep learning algorithms are beneficial for predicting failures per the performance assessment criteria. High levels of accuracy are attained in discovering tiny patterns and temporal correlations within the data. This method outperforms support vector machine (SVM) techniques concerning receiver operating characteristic (ROC), area under the curve (AUC), and accuracy. Bangdiwala et al. (2022) used many models in their study to assess company performance forecasts. The models used were Decision Trees, Random Forests, Gradient Boosting, Logistic Regression, and Multi-layerMulti-layer Perceptron Neural Networks. Each model attained a precision of around 92% based on their research.

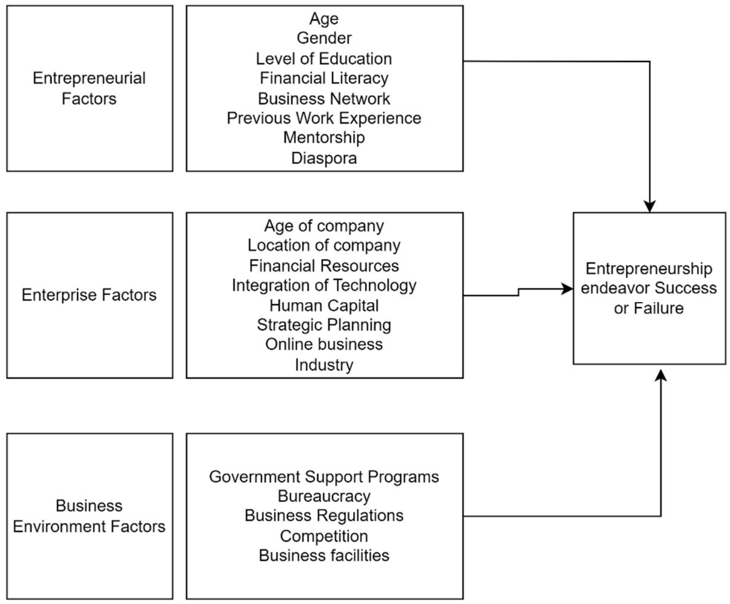
## Research Design and Methodology

An exhaustive and comprehensive sample of company owners was gathered to analyse the performance of Moroccan entrepreneurs' enterprises. The inclusion criteria for obtaining a complete image of the entrepreneurial scene included firm sizes, industry types, and geographic regions. Effective identification of successful and struggling companies requires a blend of random and focused sample techniques. Surveys were the only viable method for gathering data due to resource constraints. One strategy to recruit participation was personally contacting firm owners using internet government data and trade associations.

### *Conceptual framework*

To analyse the elements that contribute to the success or failure of entrepreneurship, the conceptual framework will consist of a multidimensional technique that encompasses financial, environmental, and business aspects, in addition to the personal profile of the entrepreneur (Figure 1). Through identifying and analysing these variables, the model aims to provide a comprehensive understanding of the intricate factors that influence achieving entrepreneurial success. The selection of meta-independent variables, including those pertaining to entrepreneurial, enterprise, and business environment aspects, was based on the literature and prior publications considered. The success or failure of the firm has the effect of determining the dependent variable.

Figure 1: Conceptual Framework




***Sampling Strategy and Data Collection Method***

In the context of this extensive research project, a survey offers a number of advantages. It makes it easy to speak with participants across the country, including Rabat, the capital of Morocco, Tangier, Fes, Meknes, and Casablanca, the second-busiest economic centre in North Africa. When it came to collecting data from a broad and varied sample, it did very well. Given the nature of this comparison research, the survey offers a standardised approach for collecting data, ensuring that the questions and answer choices are consistent. The data analysis process is simplified due to this standardisation, and it is also simpler to compare findings about various sectors.

The participants were contacted by email and picked randomly from the databases maintained by the government. Fourteen hundred and eighty-five replies were completed and could be used. Within the context of a study on entrepreneurship, a sample size of roughly 1500 participants would be considered sufficient. When the sample size is increased, statistical power also tends to grow. This allows for more in-depth studies, which raises the possibility of recognising genuine effects or patterns within the data. There is a greater possibility of obtaining a sample representative of the whole variety of entrepreneurs in terms of the sectors they operate in, their geographies, and any other relevant features.

In order to contact a large number of participants all at once, surveys are a relatively cost-effective method. More resources are needed for data collection brought about by their widespread implementation. As an additional benefit, the survey makes it possible to rapidly collect and convert numerical data to recognise patterns and trends via machine learning models. The survey structure, answer choices, and questions are all the same for every participant. This ensures that the data-gathering process is consistent and speeds up the analysis process. Table 1 shows the distribution of the answers to the gathered questions.

***Variables Definition***

The initial selection of appropriate distinct variables for the Moroccan setting was based on existing information. Table 1 delineates each variable and its measurement method.

Table 1: Definition of the variables

| | Variable Name | Description | Measurement |
|---|---|---|---|
| | Entrepreneurial Factor | | |
| 1 | Age | Age of participant | |
| 2 | Gender | Gender of participant | 0. Female 1. Male |
| 3 | Level of Education | Entrepreneur's level of education | 1. Secondary School, 2.High School 3. College, 4. Master, 5.Doctorate |
| 4 | Financial Literacy | Knowledgeable in Financial literacy | 1. No Literacy ➔ 5. Advanced Literacy |
| 5 | Previous Work Experience | Years in Business operations | 1. below 2 years 2. Between 2 and 5 years 3. Between 5 and 10 years 4. Between 10 and 15 years15. 5. Above 15 years |
| 6 | Mentorship | Mentoring and support | 0. No 1. Yes |
| 7 | Diaspora | Moroccan Diaspora | 0. No 1. Yes |
| | Enterprise Factor | | |
| 8 | Age of company | Company since inception | Number of years |
| 9 | Financial Resources | Financial resources capacities to run business | 0. No 1. Yes |
| 10 | Integration of Technology | Integration of technology into business | 1. Very Disagree ➔ 5. Very Agree |
| 11 | Human Capital | Employees skills aligned with expectations | 1. Very Disagree ➔ 5. Very Agree |
| 12 | Strategic Planning | Clear strategic planning to achieve business objectives | 1. Very Disagree ➔ 5.Very Agree |
| 13 | Online business | Business 100% online | 0. No 1. Yes |
| 14 | Industry | Type of Industry (Definition of each sector given to participants) | 1. Primary, 2.Secondary, 3. Tertiary, 4. Quaternary |
| | Business Environment Factors | | |
| 15 | Government Support Programs | Assistance/Incentives from the Government | 1. Very Disagree ➔ 5. Very Agree |
| 16 | Bureaucracy | Smooth and flexible internal practices and policies | 1. Very Disagree ➔ 5. Very Agree |
| 17 | Business Regulations | Facilitate the business and procedures | 1. Very Disagree ➔ 5. Very Agree |
| 18 | Competition | Company facing low level of | 1. Very Disagree ➔ 5. Very Agree |

| | | | |
|---|---|---|---|
| | | competition | |
| 19 | Business facilities | Availability of physical locations, spaces, or structures | 1. Very Disagree ➔ 5. Very Agree |
| | Dependant variable | | |
| 20 | Startup Performance | Entrepreneur's evaluation | 0 Low Performance 1. High Performance |

***Data Processing***

We use pre-processing analysis and data preparation approaches to address missing data, outliers, and inconsistencies. In order to do this, variables must be normalised, scaled, and encoded. Several machine learning models were used, such as Decision Trees, Random Forests, Support Vector Machines (SVMs), Ensemble Models, Logistic Regression, and TensorFlow. A technique known as the Synthetic Minority over Sampling Technique (SMOTE) was used in the research project in order to remedy the class imbalance that was brought about by several occurrences of the variable "Performance." During this investigation, the models were evaluated using a variety of measures, including precision, recall, precision, and F1 score (Table 2).

Additionally, feature relevance scores were shown to highlight the characteristics that substantially affect forecasts of company failure via the selection process of machine learning. As an Integrated Development Environment (IDE), Google Colab was used for these investigations. Google Colab provides free resources for graphics processing units (GPUs), vital for accelerating operations that need a lot of computing resource utilisation. Additionally, it integrates without any problems with well-known machine learning libraries, simplifying the process of putting machine learning algorithms into practice and ensuring compatibility with the tools used in academic research programmes.

"Precision" emphasises the accuracy of optimistic predictions to reduce the number of false positives. On the other hand, "recall" emphasises capturing all positive occurrences to reduce the number of false negatives. There is often a compromise that must be made between recall and accuracy. Depending on the threshold the model uses to determine whether or not an instance is positive, increasing one may result in a drop in the other.

Precision is of the utmost importance when the cost of false positives is considerable. For instance, if a firm not about to collapse is incorrectly classified as a failure case, it might result in expensive support, advice, and unneeded legal and regulatory assistance. The scenario of precision evaluation mistakes continues to be less significant than the scenario of recall evaluation errors. Recall is crucial when it is expensive to miss positive examples (for example, a firm diagnosed as non-failing but facing significant risks of failure may have profound implications).

Table 2: Description and use case of metrics.

| Metric | Formula | Definition |
|---|---|---|
| Precision | $\text{Precision} = \frac{TP}{TP+FP}$ | TP: True Positives, FP: False Positives TN: True Negatives FN: False Negatives |
| Recall | $\text{Recall} = \frac{TP}{TP+FN}$ | |
| F1-score | $\text{F1 Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall}$ | |
| Accuracy | $\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$ | |

Additionally, feature relevance scores were shown to highlight the characteristics that substantially affect forecasts of company failure via the selection process of machine learning. As an Integrated Development Environment (IDE), Google Colab was used for these investigations. Google Colab provides free resources for graphics processing units (GPUs), vital for accelerating operations that need a lot of computing resource utilisation. Additionally, it integrates without any problems with well-known machine learning libraries, simplifying the process of putting machine learning algorithms into practice and ensuring compatibility with the tools used in academic research programmes.

"Precision" emphasises the accuracy of optimistic predictions to reduce the number of false positives. On the other hand, "recall" emphasises capturing all positive occurrences to reduce the number of false negatives. There is often a compromise that must be made between recall and accuracy. Depending on the threshold the model uses to determine whether or not an instance is positive, increasing one may result in a drop in the other.

Precision is of the utmost importance when the cost of false positives is considerable. For instance, if a firm not about to collapse is incorrectly classified as a failure case, it might result in expensive support, advice, and unneeded legal and regulatory assistance. The scenario of precision evaluation mistakes continues to be less significant than the scenario of recall evaluation errors. Recall is crucial when it is expensive to miss positive examples (for example, a firm diagnosed as non-failing but facing significant risks of failure may have profound implications).

### *Ethical Considerations*

In addition to being given the chance to withdraw from the study without any penalties, the participants were given extensive information about the program's aims, methods, potential dangers and advantages. Taking advantage of the research was a choice that they made of their own will from their end. Additionally, individuals were allowed to decline participation in the effort, which was a sign that their autonomy was respected. In order to ensure that the information the participants submitted remained secret, stringent security procedures were put into place. The purpose of these procedures was to ensure that no data could be accessed or exposed in any instance.

## Results

### *Univariate comparative distribution*

To begin, it is necessary to thoroughly understand the characteristics and changes of each dependent and independent variable. The first step in statistical exploration is univariate analysis, which lays the framework for more complex studies. An illustration of the distribution of the dependent variables for entrepreneurial factors may be seen in Figure 2. The age distribution of entrepreneurs is right-skewed, indicating that most entrepreneurs are younger than 45 years old. Men make up over sixty percent of the replies, while women make up forty percent. An examination of the frequency of the responses reveals that 78.4% of the respondents had earned a college degree or above, with 50.9% holding masters or doctoral degrees. On the other hand, 66.9% of people have either no financial literacy or are extremely poor. Nearly twenty-two percent of company owners have fewer than two years of experience in the business world, while forty-six point four percent have between two and ten years of experience. A total of 46.1% of respondents are members of the Moroccan diaspora. This indicates that a sizeable percentage of Moroccan entrepreneurs who had previously worked in other countries have relocated back to Morocco to pursue entrepreneurial endeavours. However, it is worth noting that the connection between all variables is near zero, except for mentorship and business success, where the correlation is r=0.239 (p<001, respectively).

Figure 2: Distribution of Entrepreneurial factors

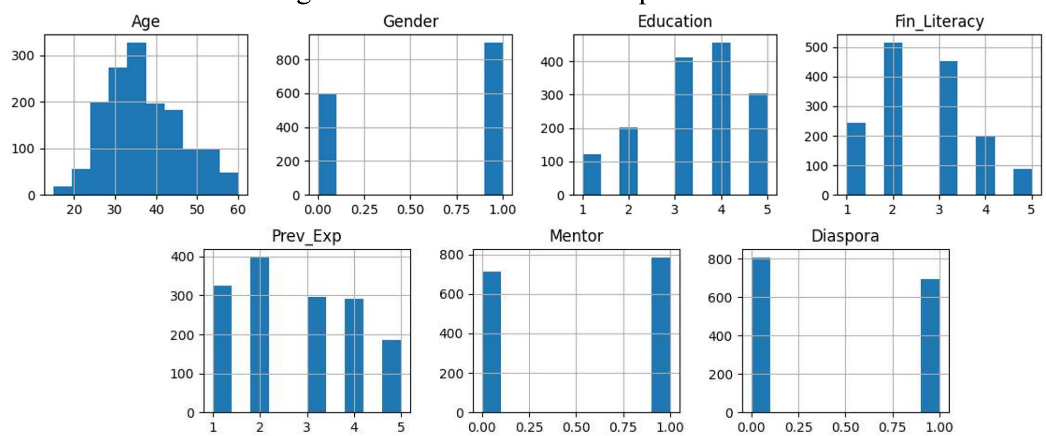
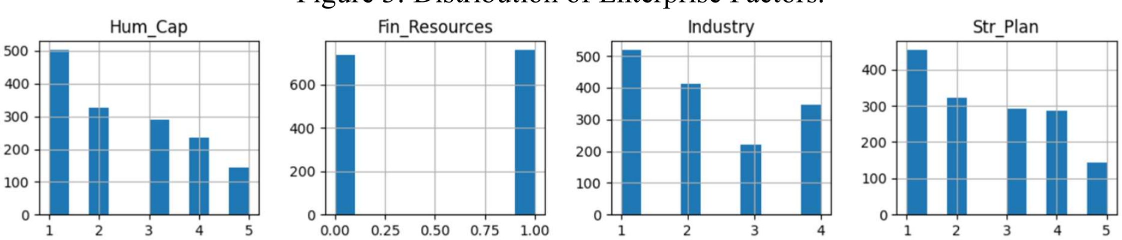



Figure 3: Distribution of Enterprise Factors.



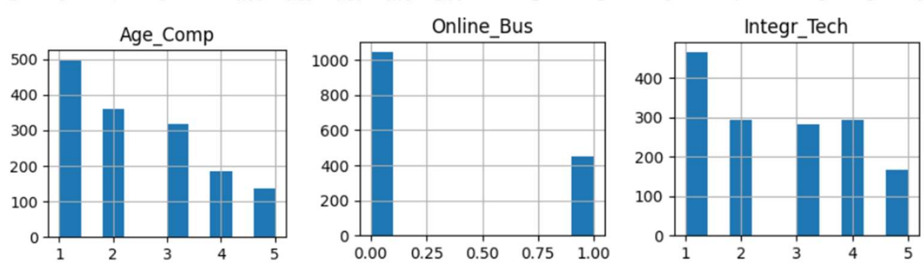
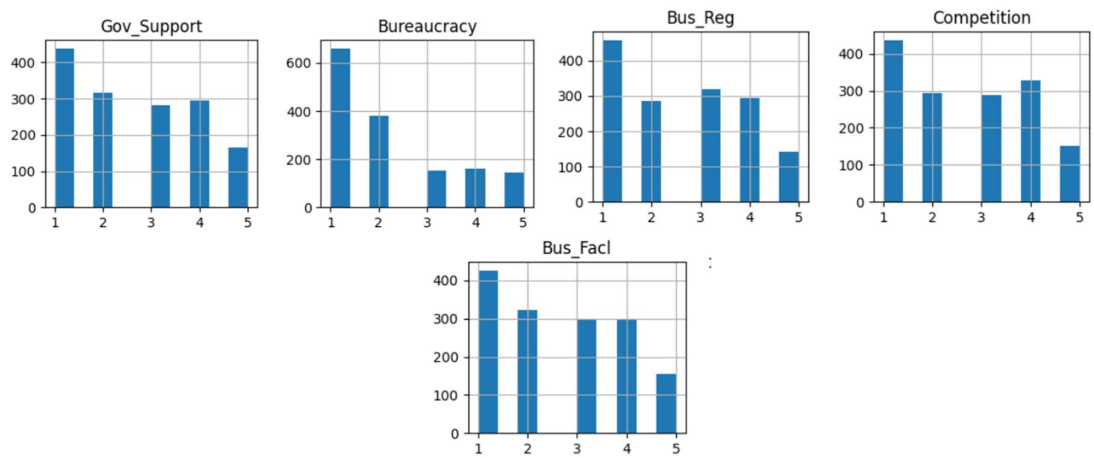




Figure 4: Distribution of Business Environment Factors.




***Machine learning comparative metrics.***

Regarding the capacity of a machine learning model to generate predictions (the dependent variable), feature importance is a method that may be used to evaluate the value of each feature being considered. It provides information on how specific characteristics impact the model's output. When it comes to modelling, this idea is particularly significant since it assists in comprehending how individual characteristics influence the decision-making process inherent in the model. For the sake of simplicity, just the top eight significant features were compared to one another. Information on the complete feature significance values may be found in the Annexes. Alongside SHAP (Shapley Additive exPlanations), which is a unifying strategy to describe the output of the machine learning models, the following figures illustrate each feature's relevance, sorted from the highest weight to the lowest weight. The Shapley values of cooperative game theory serve as the foundation for this system.

Furthermore, it offers a method that is both consistent and equitable for distributing the contribution of every attribute to the predictions across all of the potential permutations of the features. The red and blue colours are used to illustrate the direction in which a feature influences the output of the model. A red colour shows that the SHAP values are positive, which suggests that the expected output has increased. The negative SHAP values are represented by the colour blue, which indicates a drop in the projected output. The length of each vertical bar in the plot shows the magnitude of the SHAP values, and each bar corresponds to a particular feature in the plot. When the bars are longer, whether red or blue, they represent characteristics that have a more substantial influence on either raising or reducing the output of the model (economic hardship).

Table 3: Decision Metrics Outcomes

| Decision Trees | accuracy | precision | recall | f1-score |
|---|---|---|---|---|
| Class 0 | 0.75 | 0.75 | 0.87 | 0.81 |
| Class 1 | | 0.74 | 0.56 | 0.64 |
| Random Forest | accuracy | precision | recall | f1-score |

| | | | | |
|---|---|---|---|---|
| Class 0 | **0.77** | 0.78 | **0.88** | **0.83** |
| Class 1 | | **0.75** | 0.58 | 0.65 |
| Logistic Regression | accuracy | precision | recall | f1-score |
| Class 0 | 0.73 | 0.75 | 0.83 | 0.81 |
| Class 1 | | 0.68 | 0.56 | 0.61 |
| SVM | accuracy | precision | recall | f1-score |
| Class 0 | 0.75 | 0.78 | 0.84 | 0.81 |
| Class 1 | | 0.66 | 0.56 | 0.61 |
| TensorFlow | accuracy | precision | recall | f1-score |
| Class 0 | 0.71 | 0.78 | 0.77 | 0.78 |
| Class 1 | | 0.58 | 0.59 | 0.58 |
| TensorFlow SMOTE | accuracy | precision | recall | f1-score |
| Class 0 | 0.71 | 0.79 | 0.75 | 0.77 |
| Class 1 | | 0.57 | 0.62 | 0.59 |
| Ensemble Models Voting Classifier | accuracy | precision | recall | f1-score |
| Class 0 | **0.77** | 0.81 | **0.85** | **0.83** |
| Class 1 | | **0.69** | 0.62 | 0.65 |
| Ensemble Models Stacking Classifier | accuracy | precision | recall | f1-score |
| Class 0 | 0.73 | 0.75 | 0.83 | 0.79 |
| Class 1 | | 0.67 | 0.56 | 0.61 |

Table 3 shows that Ensemble Models with Voting Classifier and Random Forest have the highest metrics. Indeed, the accuracy of both models is 0.77, and they also have the same f1-score, which is 0.83. However, Random Forest has a Recall score of 0.88 for class 0 (companies facing economic hardship) while Ensemble Models shows 0.85. On the other hand, Ensemble Models has a procession score of 0.69 for class 1 (companies doing well) while Randon Forest has a score of 0.75. In both classes (0 and 1), Random Forest seems to be this study's most accurate predictive model.

***Machine Learning Comparative Feature Importances.***

Regarding the capacity of a machine learning model to generate predictions (the dependent variable), feature importance is a method that may be used to evaluate the value of each feature being considered. It provides information on how specific characteristics impact the model's output. When it comes to modelling, this idea is particularly significant since it assists in comprehending how individual characteristics influence the decision-making process inherent in the model. For the sake of simplicity, just the top eight significant features were compared to one another. Information on the complete feature significance values may be found in the Annexes. Alongside SHAP (Shapley Additive exPlanations), which is a unifying strategy to describe the output of the machine learning models, the following figures illustrate each feature's relevance, sorted from the highest weight to the lowest weight. The Shapley values of cooperative game theory serve as the foundation for this system.

Furthermore, it offers a method that is both consistent and equitable for distributing the contribution of every attribute to the predictions across all of the potential permutations of the features. The red and blue colours are used to illustrate the direction in which a feature influences the output of the model. A red colour shows that the SHAP values are positive, which suggests that the expected output has increased. The negative SHAP values are represented by the colour blue, which indicates a drop in the projected output. The length of each vertical bar in the plot shows the magnitude of the SHAP values, and each bar corresponds to a particular feature in the plot. When the bars are longer, whether red or blue, they represent characteristics that have a more substantial influence on either raising or reducing the output of the model (economic hardship).

Figure 5.a: Feature importance of Decision Trees Model and SHAP Distribution

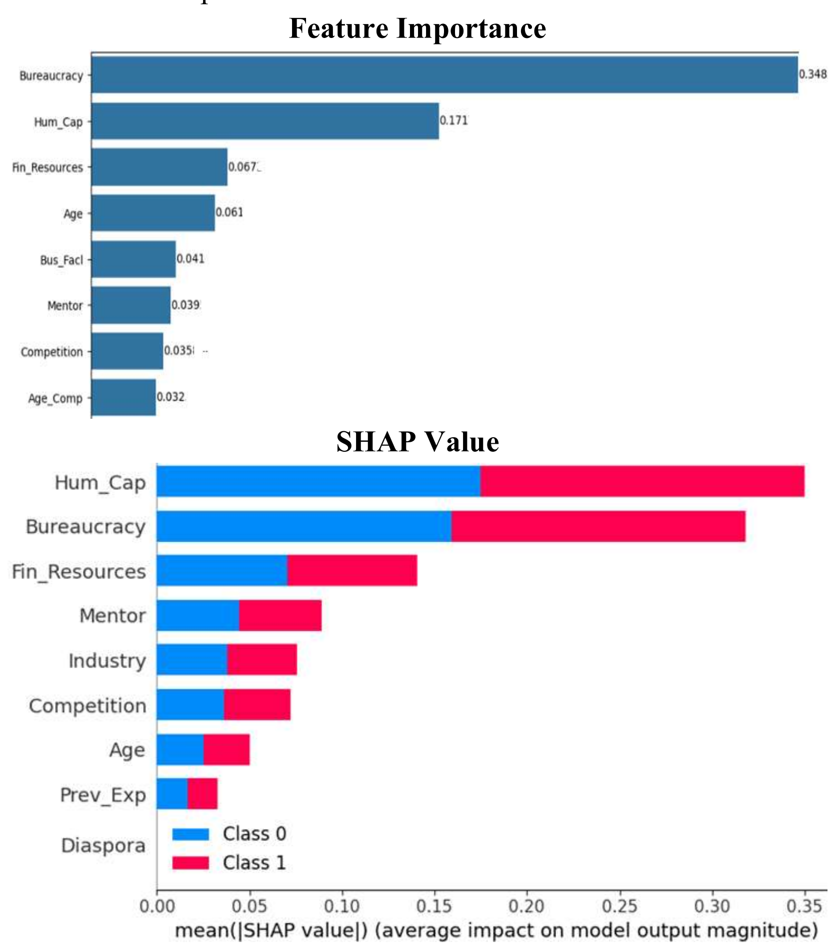
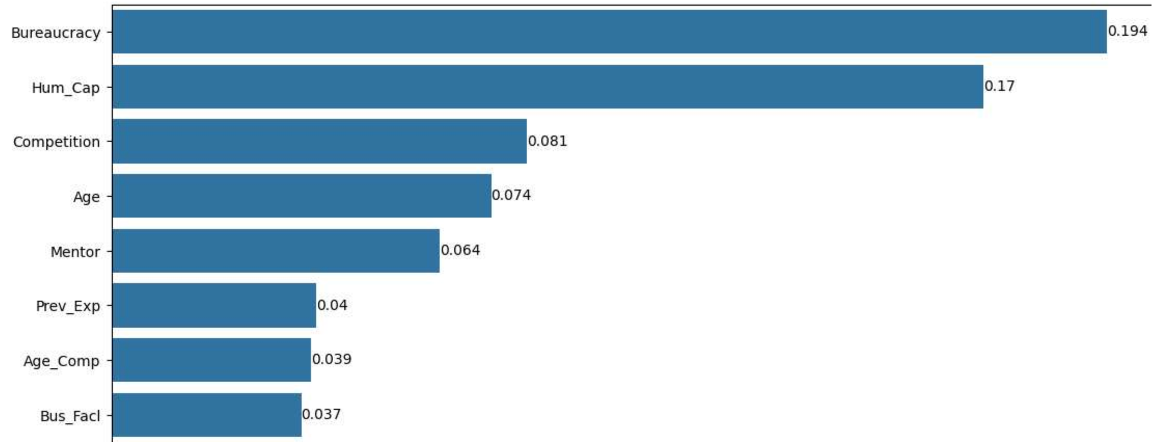



Figure 5.b: Feature importance of Random Forest Model and SHAP Distribution

**Feature Importance**




**SHAP Value**

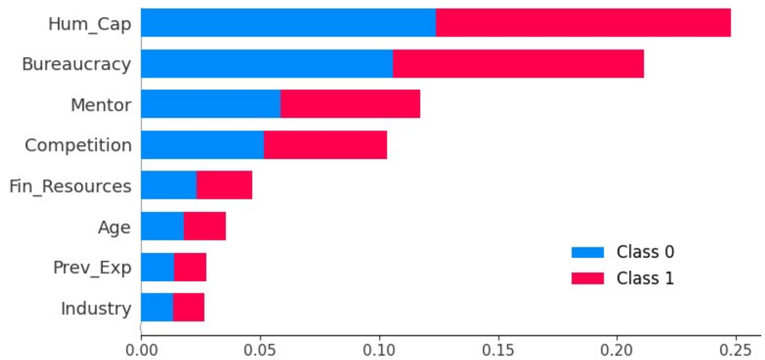
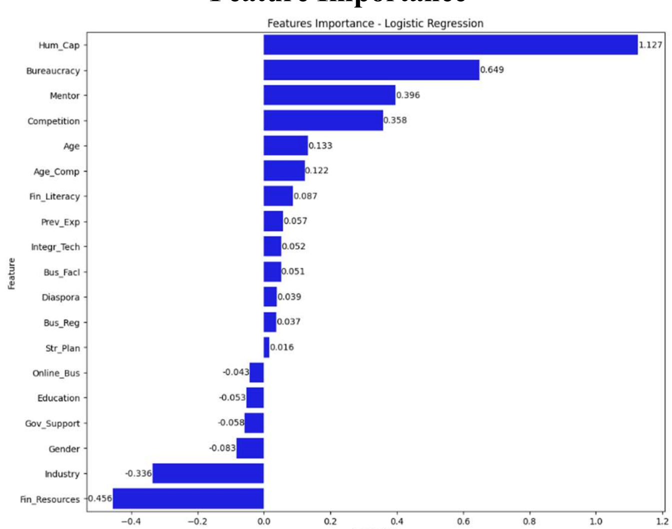



Figure 5.c: Feature importance of **Logistic Regression** Model and SHAP Distribution

**Feature Importance**




**SHAP Value**

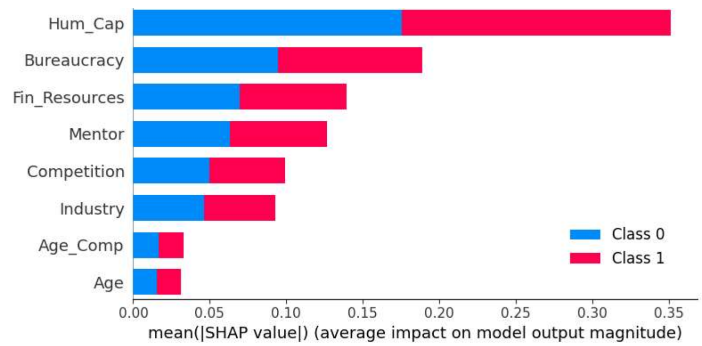
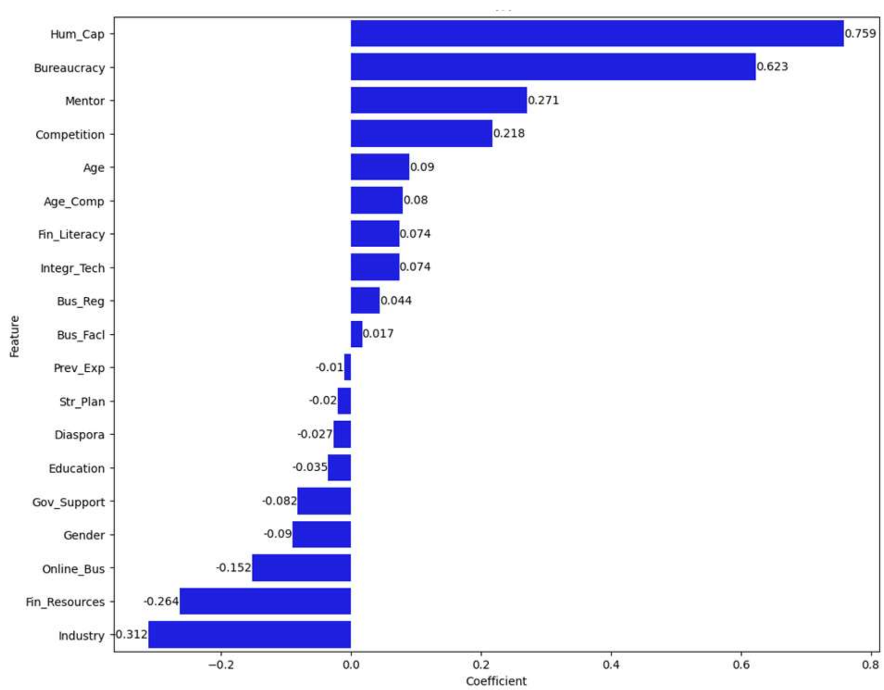



Figure 5.d: Feature importance of SVM Model and SHAP Distribution

**Feature Importance**




**SHAP Value**

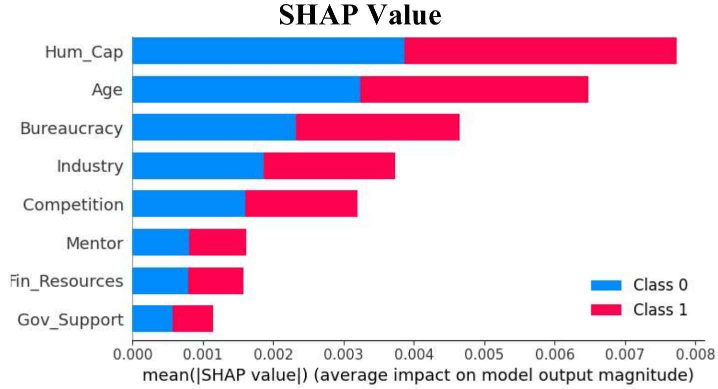
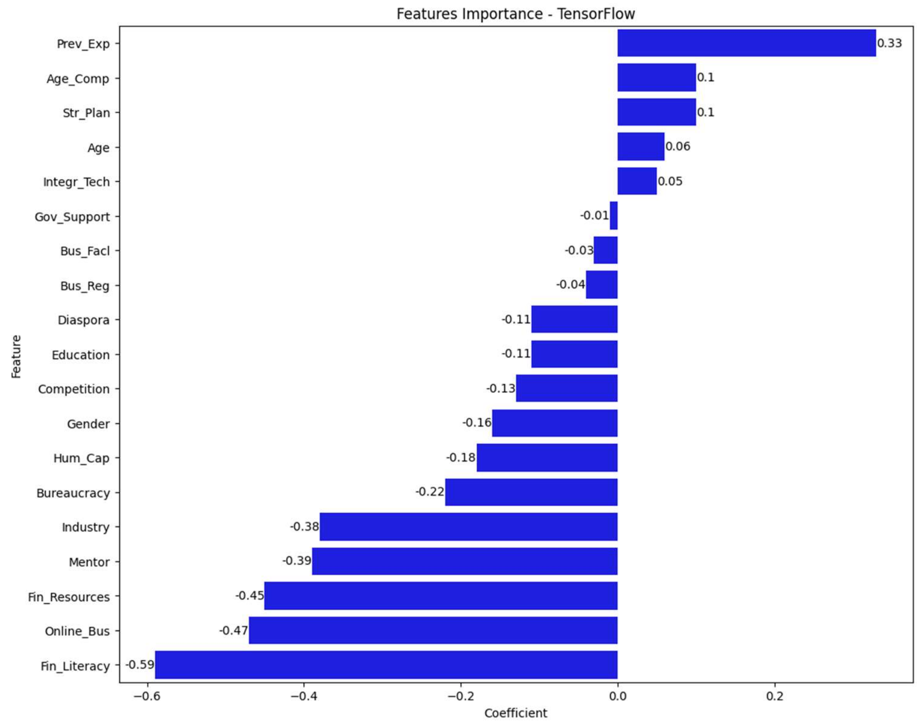



Figure 5.e: Feature importance of **Deep Learning** – TensorFlow with SMOTE

**Feature Importance**




**TensorFlow with SMOTE**

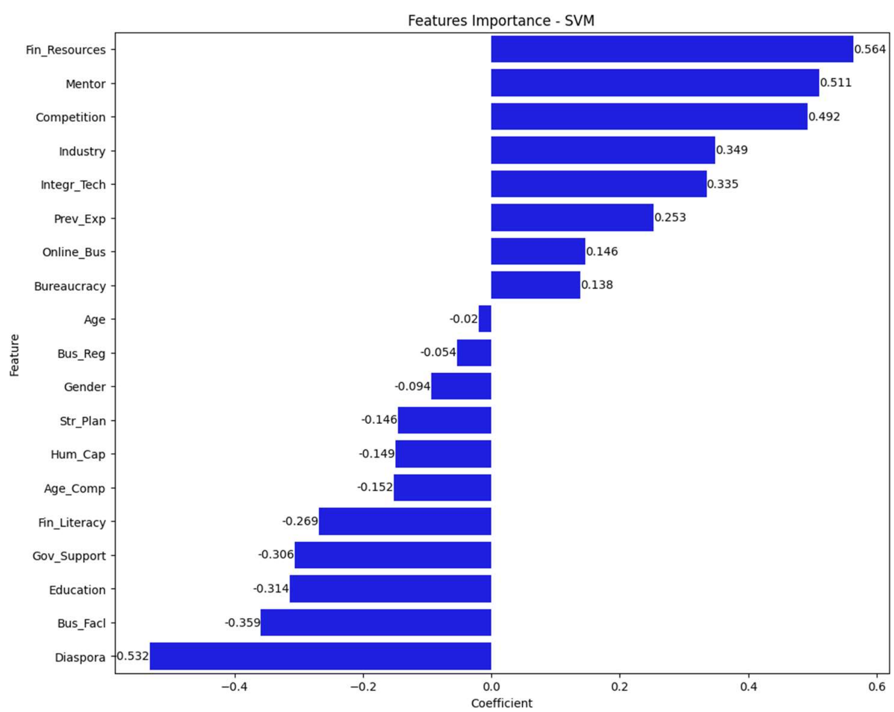
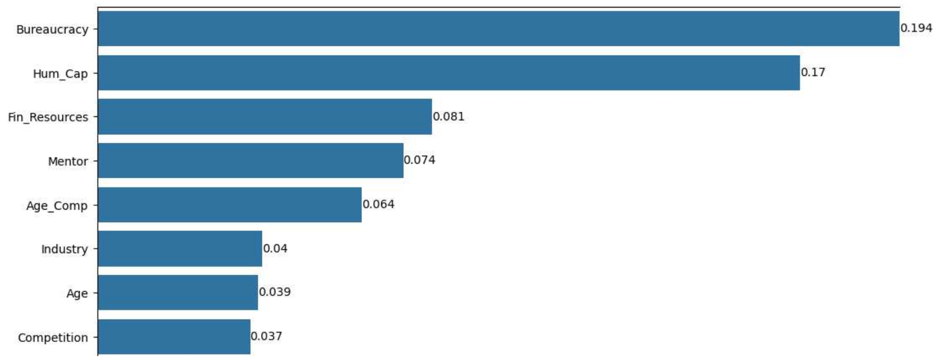
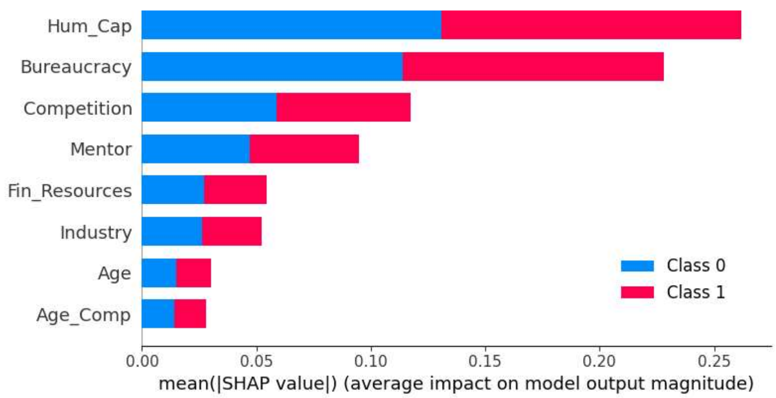



Figure 5.f: Feature importance of **Ensemble** Model and SHAP Distribution
**Feature Importance**

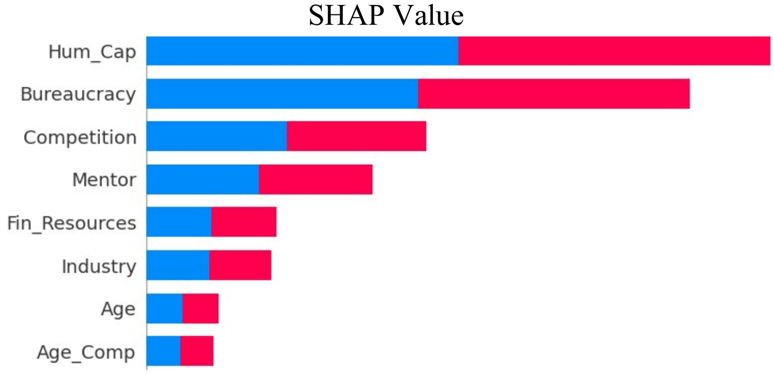



**SHAP Value**




**Ensemble Models – GradientBooster Classifier**

SHAP Value




## Discussion

According to the findings of a correlation data analysis concerning independent variables and the interactions between them, no patterns or significant linkages can be utilised to forecast the challenges that businesses would face. According to the distribution of these variables, no trends or correlation studies include all of the variables. The complexity of the elements that influence the success of firms is shown by the absence of statistical data that are especially noteworthy. Consequently, a sophisticated approach is required to comprehend their relationship's complexities. Within this context, machine learning has the potential to provide capabilities and flexibility, hence enhancing prediction via the use of data-driven insights and automation, transcending the ways that are traditionally utilised.

The findings describe the significance of the feature and coefficients for Random Forest and Ensemble Models (Voting Classifier) regarding their ability to forecast the success of entrepreneurial endeavours. It is surprising to learn that Bureaucracy is the most essential quality in Random Forest (0.194), followed by Human Capital, Competition, Entrepreneurial Age, and Mentoring. Conversely, ensemble Models prioritise Bureaucracy (0.369), Human Capital, Financial Resources, Mentors, and the Age of Companies. Because of the apparent capacity, both models emphasise Bureaucracy, stressing the significance of this factor in predicting a company's performance (whether it will be successful or unsuccessful). Consistently, human capital receives high marks, which highlights its worth. In order to properly understand the gap in rankings for other attributes, such as mentorship and financial capital, a thorough investigation is required. It is implied that assets have a more significant effect when ensemble models are used because they are prioritised.

Position differences might result from intrinsic model traits or variations in the training data in question. By analysing the coefficients, one may better understand the extent to which each characteristic has an impact. The Bureaucracy obtains the highest score from Random Forest, which is 0.194%, indicating that it substantially affects the system. With a value of 0.369, Bureaucracy is the most prominent of the Ensemble Models.

The analysis reveals that both models emphasise human resources and bureaucratic processes as fundamental components. Businesses experience significant efficiency losses due to internal Bureaucracy, which also leads to problems with adaptability and performance. Studies conducted by Becheikh and Bouaddi (2023), Kamal (2021), and Lecuna et al. (2020) are only two instances of the body of research that demonstrates the negative consequences that bureaucratic impediments have on enterprises. It is well acknowledged that Moroccan businesses face significant challenges regarding their ability to expand and compete. These challenges include lengthy and intricate bureaucratic procedures, complex legal systems, and delayed decision-making speed.

Nevertheless, the differences in rankings and correlations provide light on the slight differences. In addition, the findings demonstrate that Bureaucracy has a significant impact on the outcomes of the model. To be more specific, SHAP employs a ratio of red to blue hues that is 50/50 for this characteristic to anticipate the organisation's success. The hue of red represents helpful assistance, which contributes to an improvement in the estimation, while the hue of blue represents unfavourable contributions, which reduces the accuracy of the prediction. In this particular scenario, Bureaucracy serves to both help and inhibit the growth of businesses. The consistent colour palette gives the impression that there is a complex connection between some unfavourable outcomes and favourable aspects of Bureaucracy that increase production. This realisation sheds light on the very intimate influence that bureaucratic components have on the operation of the whole organisation.

Similarly, the SHAP graphic for human resources depicts a 50/50 split between the colours red and blue as a representation of the performance of the company or organisation. It exemplifies an influence that is both comprehensive and varied. Red represents the positive contributions or characteristics of human capital that contribute to increased productivity. On the other hand, blue denotes unfavourable input or components that have the potential to hinder overall achievement. Some parts of human resources may positively impact company performance, while other factors may have a detrimental impact. This stability indicates that there is a complicated link between the two. The portrayal is well-balanced, and it emphasises how crucial it is to have a comprehensive understanding of all the many components that comprise the human capital construct of an organisation.

When judging whether an organisation will be successful or unsuccessful, the fact that the companies that contributed to this research are new enterprises is one potential reason that Bureaucracy has lower predictive power than human assets. People have the impression that bureaucratic organisations need to be more flexible in their decision-making and sluggish in their ability to adapt, making it more difficult for them to satisfy the market's requirements. On the other side, the value of human capital demonstrates how skilled, creative, and adaptable individuals can make a difference in an organisation. In a world where organisations are increasingly operating in fluid and unpredictable situations, the capacity to harness and develop human capital is becoming an increasingly important aspect in determining success. When it comes to forecasting an organisation's performance, prioritising abilities over bureaucratic framework techniques is more significant.

The elements that have low coefficients on the success or failure of a firm, as shown by Random Forest and Ensemble Models, are ascribed to the fact that their contribution to the overall predictive model is relatively minimal. Within the parameters of this discussion, it is possible that variables with lower coefficients, such as "Previous Experience" and "Business Facilities," would have less significant impacts on the projected result. Even though features with lower coefficients may have a less significant effect on their own, the cumulative contribution of these features helps capture subtleties in the data and ads to the overall prediction performance. The total of all other characteristics, excluding Bureaucracy and human factors, is equal to 0.335. Less than sixty-five percent of the total coefficients are allocated to competition, mentorship, and financial resources. Mentorship, for instance, is essential in assisting new business owners. Thanks to its assistance, it offers insights and helps company leaders make decisions based on reliable facts. Furthermore, financial resources are the crucial lifeblood of startups, as they enable them to invest in operations, technical improvements, and the acquisition of individuals to grow and expand their business. The study's conclusions indicate that to succeed as an entrepreneur, one must balance bureaucracy and competition, acquire expertise from mentors, and have a sound financial basis to recruit intelligent people. Due to these factors, entrepreneurs can triumph over challenges, seize opportunities when they arise, and achieve sustainable success in business environments.

Furthermore, to be successful as an entrepreneur in dynamic business settings, it is vital to adopt a balanced strategy that not only prioritises human capital but also handles the issues that bureaucratic obstacles provide. Entrepreneurs have the potential to attain the capacity to position themselves for sustained growth and perseverance in the face of adversity by focusing on fostering talent, encouraging innovation, and overcoming internal impediments. This is something that can be accomplished.

***Social Implications***

When beginning a firm, entrepreneurs only enter the endeavour expecting to succeed. On the other hand, failure is something that a sizeable proportion of business owners' encounter, and it may negatively impact their professional, personal, and social lives. When it comes to an entrepreneur's introspective observation, assessment of failure, and feelings about it, it is essential to analyse how the social context, such as the family, impacts these aspects. Consequently, the purpose of this research is to contribute to the provision of an overview of the elements that predict the failure of an inventor or businessperson. The dread of failing involves a variety of worries, such as the uncertainty of the future, feelings of humiliation and embarrassment, the potential to hurt loved ones, and impairments to one's sense of self-worth. People are often dissuaded from pursuing entrepreneurial endeavours due to the societal costs. In a society as collectivist as Morocco's, the significance of social circumstances in re-engaging entrepreneurs is of the utmost importance. This raises doubts regarding the effectiveness of government programmes in promoting entrepreneurship after they have failed. Entrepreneurship is intrinsically relational and requires cooperation with various stakeholders, including social players.

**Limitations of Machine Learning Predictive Models**

Predictions made by machine learning are often debated in the literature due to their inherent limitations. Both overfitting and under fitting are phenomena that Montesinos et al. (2022) discussed. Overfitting occurs when a statistical machine learning model learns the signal and noise in the training data. At the same time, under fitting occurs when the model does not accurately represent the data pattern because it only includes a small number of predictors or because the training dataset is small. Too-tuned models may capture noise rather than actual patterns in the data being trained on (Haas, 2020). Since machine learning predictive models are dependent on past data in order to make predictions, it is possible that they may not perform effectively in circumstances that include unusual events, outliers, or unexpected shifts (Navarro et al., 2021; Blanzeisky & Cunningham, 2021). For ethical reasons, it is essential to worry about permission, privacy, and the possibility of misusing personal data. Laws could need assistance to keep up with the fast developments that are occurring in the world of machine learning applications. Because they demand significant processing power for training and inference, many deep neural networks and other complex machine learning models are not appropriate for usage in environments with limited resources. Machine learning models often need assistance demonstrating causality, even though they are adept at recognising correlations.

**Conclusion**

Even though there are no visible patterns that are formed from independent variables, the statistical data analysis offers light on the complexity of the factors that impact the performance of companies. Machine learning, specifically Random Forest and Ensemble Models, offers a flexible technique that uses data-driven insights to improve prediction. Furthermore, the coefficients and feature significance analysis provide insight into bureaucracy and human capital's influence by highlighting their relationships. Human capital and internal bureaucracy are two of the characteristics that, according to the conclusions of this study, are responsible for determining whether or not an entrepreneurial effort will be successful. Even though flexible bureaucracy has a significant impact on the performance of companies, it may be harmful to the performance of enterprises due to the complexity of the processes involved. The management of human resources may influence many different aspects. Some of these aspects can contribute to higher productivity when used to their full potential, while others might impede growth when they are disregarded. The findings shed light on the need to cultivate a culture supporting innovation and resilience among employee members to push a company's progress. When it comes to coaching entrepreneurs and supporting their success over the long term, mentorship and financial resources are both essential components that must be considered with great importance. To achieve entrepreneurial success in dynamic business environments, it is essential to have a well-balanced strategy that prioritises human capital and finds solutions to bureaucratic restraints. Even though factors such as expertise and business facilities contribute to overall predictive performance, it is imperative to have a balanced strategy emphasising human entrepreneurship. Contributing to the new strategic direction of the government, which is to support entrepreneurial efforts and aid entrepreneurs on their road, is the goal of this research, which intends to contribute to the new strategic direction.

## References

Aidis, R., Estrin, S., & Mickiewicz, T. (2012). Size matters: Entrepreneurial entry and government. Small Business Economics, 39(1), 119-139.

Antretter, T., Blohm, I., Grichnik, D., & Wincent, J. (2019). Predicting new venture survival: A Twitter-based machine learning approach to measuring online legitimacy. Journal of Business Venturing Insights, 11, e00109. https://doi.org/10.1016/j.jbvi.2018.e00109

Assagaf, I., Sukandi, A., & Abdillah, A. A. (2023). Machine Failure Detection using Deep Learning. Recent in Engineering Science and Technology, 1(03), 26-31.

Bangdiwala, M., Mehta, Y., Agrawal, S., & Ghane, S. (2022, August). Predicting success rate of startups using machine learning algorithms. In 2022 2nd Asian Conference on Innovation in Technology (ASIANCON) (pp. 1-6). IEEE.

Beaver, W. H. (1975). Pioneering studies in predicting corporate bankruptcy. The Accounting Review, 50(1), 112–122.

Blanzeisky, W., & Cunningham, P. (2021, September). Algorithmic factors influencing bias in machine learning. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases (pp. 559-574). Cham: Springer International Publishing.

Bouazza, A., & El Kadiri, S. (2018). Entrepreneurship education in Morocco: Perception, attitude and intention of students. Journal of Innovation and Entrepreneurship, 7(1), 1-19

Bouhaj, S., Jahidi, R., & Lebzar, B. (2022). Critical success factors for technology start-ups: An exploratory qualitative study of experts' views. Revue Française d'Economie et de Gestion, 3(6).

Da Silva, A. F., Brito, J. H., & Pereira, J. M. (2023). Using Machine Learning to Predict Business Failure in Iberian Hospitality Sector. In Advances in Tourism, Technology and Systems: Selected Papers from ICOTTS 2022, Volume 2 (pp. 313-322). Singapore: Springer Nature Singapore.

Daskalaki, C., Doulaveras, S., & Skiadopoulos, G. (2019). Forecasting corporate failure using a model based on support vector machines and self-organizing maps. European Journal of Operational Research, 278(2), 558-571.

Dergiades, T., & Milas, C. (2017). Prediction of competitiveness in small and medium-sized enterprises (SMEs). International Journal of Production Economics, 193, 518-524.

El Alami, A., et al. (2019). Entrepreneurial dynamics in Morocco: A country in transition. Journal of Innovation and Entrepreneurship, 8(1), 1-24.

Fathi, Y., et al. (2017). Competitiveness analysis in e-commerce companies: A machine learning approach. Journal of Business Economics and Management, 18(6), 1113-1130.

Fatoki, O. (2014). The impact of firm and entrepreneurial characteristics on access to debt finance by SMEs in King William's Town, South Africa. Journal of Social Sciences, 38(3), 243-248.

Haas, K. (2020, May). Limiting Machine Learning Overfitting Uncertainties Through Persistent Homology. In Verification and Validation (Vol. 83594, p. V001T08A001). American Society of Mechanical Engineers.

Haddoud, M. Y., & Bouazza, A. (2020). The impact of entrepreneurship education on entrepreneurial intention: A study of Moroccan students. Journal of Global Entrepreneurship Research, 10(1), 1-22.

Hind, T., & Jamal, Z. (2023). Financing Constraints and Prospects for Innovative SMEs in Morocco. African Journal of Business & Economic Research, 18(4).

Kamel, E. M. (2021). The MENA region's need for more democracy and less bureaucracy: A gravity model controlling for aspects of governance and trade freedom in MENA. The World Economy, 44(6), 1885-1912.

Khan, R. E. A., et al. (2020). Institutional voids and entrepreneurship in developing countries. International Entrepreneurship and Management Journal, 16(4), 1553-1582.

Kibler, E., et al. (2014). Organizational institutionalism in late development. Organization Studies, 35(12), 1749-1766.

Klapper, L., Amit, R., & Guillén, M. F. (2010). Entrepreneurship and firm formation across countries. Strategic Entrepreneurship Journal, 4(3), 234-257.

Laabissi, H., et al. (2019). Entrepreneurial intentions among university students in Morocco: The mediating role of attitude. International Journal of Entrepreneurial Behavior & Research, 25(4), 838-858.

Lahrech, A., et al. (2016). The role of cultural and social factors in entrepreneurship. Journal of Innovation and Entrepreneurship, 5(1), 1-17.

Lecuna, A., Cohen, B., & Mandakovic, V. (2020). Want more high-growth entrepreneurs? Then control corruption with less ineffective bureaucracy. Interdisciplinary Science Reviews, 45(4), 525-546.

Lee, G., et al. (2019). Predicting organizational competitiveness based on success factors: Insights from the hospitality industry. Journal of Hospitality Marketing & Management, 28(6), 674-694.

Loureiro, A. L. D., Miguéis, V. L., & da Silva, L. F. M. (2018). Exploring the use of deep neural networks for sales forecasting in fashion retail. Decision Support Systems, 114, 81–93. https:// doi.org/10.1016/j.dss.2018.08.010

Lussier, R. N. (1995). A nonfinancial business success versus failure prediction model for young firms. Journal of Small Bussiness Management, 33(3), 8. https://doi.org/10.1111/j.1540-627X. 2010.00298.x

Lussier, R. N., & Claudia, E. H. (2010). A three-country comparison of the business success versus failure prediction model. Journal of Small Business Management, 48(3), 360–377. https://doi.org/10.1111/j.1540-627X.2010.00298.x

Mayr, S., Mitter, C., Kücher, A., & Duller, C. (2020). Entrepreneur characteristics and differences in reasons for business failure: Evidence from bankrupt Austrian SMEs. Journal of Small Business and Entrepreneurship, 33(5), 539–558. https://doi.org/10.1080/08276331.2020.1786647

Montesinos López, O. A., Montesinos López, A., & Crossa, J. (2022). Overfitting, model tuning, and evaluation of prediction performance. In Multivariate statistical machine learning methods for genomic prediction (pp. 109-139). Cham: Springer International Publishing.

Navarro, C. L. A., Damen, J. A., Takada, T., Nijman, S. W., Dhiman, P., Ma, J., ... & Hooft, L. (2021). Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. bmj, 375.

Piskunova, O., Ligonenko, L., Klochko, R., Frolova, T., & Bilyk, T. (2022). Applying Machine Learning Approach to Start-up Success Prediction. Journal of Business Analytics, 10(3), 123-145.

Rahman, S. M. M., et al. (2018). Significance of machine learning models in determining organizational competitiveness in the digital age. International Journal of Information Management, 43, 1-8.

Reynolds, P. D. (2017). When is a firm born? Alternative criteria and consequences. Business Economics, 52(1), 41–56. https://doi.org/10.1057/s11369-017-0022-8

Robichaud, Y., Cachon, J. C., Assaidi, A., & Ahmed, N. B. (2023). Entrepreneurship in Morocco: An Empirical Study of Motives, Barriers, and Determinants of Success. Journal of Management Policy and Practice, 24(3), 1-27.

van Witteloostuijn, A., & Kolkman, D. (2019). Is firm growth random? A machine learning perspective. Journal of Business Venturing Insights, 11, 1–5. https://doi.org/10.1016/j.jbvi.2018.e00107

Vasquez, E., Santisteban, J., & Mauricio, D. (2023). Predicting the Success of a Startup in Information Technology Through Machine Learning. International Journal of Information Technology and Web Engineering (IJITWE), 18(1), 1-17.

Veganzones, D. (2023). Predicting Corporate Failure Using Ensemble Extreme Learning Machine. In Novel Financial Applications of Machine Learning and Deep Learning: Algorithms, Product Modeling, and Applications (pp. 107-124). Cham: Springer International Publishing.

Xia, T., & Tao, L. (2017). Improving bankruptcy prediction using data mining techniques and fast-clustering algorithms. Expert Systems with Applications, 72, 327-334.

Zavgren, C. V. (1984). Significant advancements and discoveries in machine learning to predict a firm's success or failure. Journal of Accounting Research, 22(1), 59-80.

Zhang, Y., & Luo, F. (2015). Investigating machine learning predictive analytics for organizational competitiveness. Journal of Business Research, 68(9), 1968-1978.