Analyzing Factors Influencing Early-Stage Entrepreneurship with Machine Learning: GEM Database Insights

Abstract

In recent years, understanding the determinants of Entrepreneurial Intentions (EI) among young individuals has gained significant attention worldwide. This study attempts to empirically investigate this phenomenon across 50 economies using the 2019 Global Entrepreneurship Monitor (GEM) dataset of working-age individuals (18–35 years), employing machine learning techniques to uncover influential factors of entrepreneurial intention. We apply machine-learning models such as Decision Trees, Random Forests, and XGBoost algorithms to our predictive model. Among these methods, Random Forest exhibited the highest predictive accuracy. We use 12 variables encompassing cognitive and behavioral factors, economic status, and neighborhood influence as predictors of EI. By running the model separately for low, middle, and high-income economies we draw a contrast between the differences in the factors affecting EI in each. The analysis reveals that networks, skills, and creativity play pivotal roles in shaping entrepreneurial intentions, with education emerging as a crucial determinant, particularly in lower-income countries. Creativity also emerges as a vital driver, especially in middle and high-income countries, emphasizing innovative thinking's role. Furthermore, household situations, such as larger family sizes, exhibit positive correlations with higher entrepreneurial intentions. Neighborhood support is significant in low-income countries, highlighting socio-cultural influences. Continued research is needed to deepen our understanding of entrepreneurial motivations and barriers. Future studies could include longitudinal research to track intentions over time and comparative analyses across cultures. Qualitative methods can complement quantitative analyses by providing insights into the drivers of entrepreneurial aspirations.

Keywords: Entrepreneurial Intentions, Global Entrepreneurship Monitor, Random forest, Network, Perceived skills, perceived creativity, Machine Learning, AI.

JEL Codes: *L26, L31, M13, O30, O57*

1.1 Introduction

Entrepreneurship serves as a fundamental catalyst for economic growth and innovation on a global scale. As such, understanding the factors that shape individuals' intentions to embark on entrepreneurial endeavors is paramount. In our research, we delve into the intricate interplay of cognitive and behavioral factors such as skills, creativity, network role models, vision, and fear of failure. Recognizing the variability of these influences across different countries with diverse economic conditions, our study aims to dissect how these factors interact to impact entrepreneurial intentions. By doing so, we seek to uncover insights that transcend geographical boundaries, shedding light on the universal and context-specific drivers of entrepreneurial aspirations. Building on existing economic literature, an old study by Arenius and Minniti (2005) has emphasized the significance of factors such as alertness to opportunities, fear of failure, and confidence in one's skills in shaping entrepreneurial behavior. Researchers have utilized a large sample across 28 countries to investigate the correlations between these perceptual variables and the decision to start a new business. Findings indicate that these perceptual variables are significantly correlated with entrepreneurial activity across all countries in the sample and among different genders (Graham & Bonner, 2022). While the data did not establish causal relationships, the results suggest that nascent entrepreneurs often rely on subjective perceptions rather than objective expectations of success when making decisions. Therefore, incorporating perceptual variables into economic models of entrepreneurial behavior is crucial for a more comprehensive understanding of entrepreneurial intentions (Beynon and Jones, 2020).

Our comparative analysis extends across multiple countries, each presenting a distinct socio-economic landscape. By traversing these diverse terrains, we endeavor to uncover how cognitive and behavioral factors manifest within different economic environments. Through rigorous examination, we illuminate nuanced insights into the drivers of entrepreneurial intentions, enriching the discourse on entrepreneurship. Moreover, by integrating machine learning methodologies into our study, we not only enhance predictive accuracy but also contribute to the evolution of research methodologies in the field of entrepreneurship. Linan and Urbano (2011) have recognized the significance of environmental factors alongside individual cognitive processes. While cognitive models traditionally focus on internal determinants, recent studies have underscored the importance of considering environmental influences to elucidate regional disparities in entrepreneurial activity. Drawing on theoretical frameworks such as the planned behavior approach, institutional economic theory, and social capital theory, these scholars have explored how factors such as the valuation of entrepreneurship within specific regions shape entrepreneurial intentions (Sanchez & Sahuquillo, 2012). Empirical investigations, including a study on university students from Catalonia and Andalusia, Spain, have employed structural equation techniques to unravel the complex relationship between environmental cognitive elements and entrepreneurial intentions. Findings reveal that the social valuation of entrepreneurship differs between regions, with implications for perceived subjective norms, attitudes towards entrepreneurship, and behavioral control. These insights highlight the importance of policy interventions aimed at promoting positive entrepreneurial values in less developed regions to foster entrepreneurship and spur economic growth.

To realize our research objectives, we employ sophisticated machine-learning techniques to predict entrepreneurial intentions with precision and depth. Leveraging Random Forest, XGBoost, and Decision Trees as predictive models, we harness the power of data analytics to discern patterns within large datasets. This analytical arsenal enables us to unravel the complexities underlying individuals' propensity towards entrepreneurship. Through a comparative analysis of these methodologies, we aim to ascertain the most effective approach for predicting entrepreneurial intentions across economies characterized by varying average household incomes. By bridging the realms of academia and practical application, our research offers actionable insights for policymakers and educators striving to cultivate entrepreneurial ecosystems within their respective regions. In our quest for understanding, we discern that Random Forest emerges as the optimal machine-learning technique for predicting entrepreneurial intentions. This revelation corroborates previous findings, underscoring the robustness of Random Forest in capturing the intricacies of cognitive and behavioral factors associated with entrepreneurship. Armed with this predictive model, we equip policymakers and educators with a potent tool to inform strategic decision-making processes. Our research endeavors to bridge the gap between theory and practice, offering empirical insights that resonate across borders and drive actionable interventions to foster entrepreneurship and spur economic growth globally.

Our study dives into understanding why people in different countries want to start their businesses. We blend insights from psychology and behavioral studies with machine learning to see the full picture of what drives someone to become an entrepreneur. We know that starting a business isn't just about personal traits; it's also influenced by where you live and the conditions around an individual. We try to untangle this complexity by looking at how things like creativity, vision, and spotting business opportunities mix with factors like age and fear of failure to shape people's desire to start businesses. Our findings can help governments, organizations, and programs better support budding entrepreneurs by tailoring help to fit their specific needs. This could mean more jobs, more innovation, and stronger economies. In the end, our research aims to make starting a business easier for everyone, no matter where they are. We believe that by understanding the unique situations people face, we can create environments where entrepreneurship thrives and communities flourish.

2. Literature review

Khefacha and Romdhane (2023) explore the relationship between entrepreneurial aspirations and prosperity, focusing on the impact of entrepreneurship on economic growth, job creation, and income generation. Fostering a conducive entrepreneurial ecosystem and aligning with Sustainable Development Goals can lead to sustainable entrepreneurial activities that benefit society. The results of (Baber and Fanea, 2023) suggest that personal and subjective norms play a mediating role in shaping the intentions of students to choose entrepreneurship in the sustainability field. Acs and Szerb (2007) highlight the positive correlation between entrepreneurial activity and economic growth. Drawing on various theoretical perspectives and empirical evidence, it underscores the critical contribution of entrepreneurship to overall economic prosperity and development. A balanced ratio of opportunity-to-necessity entrepreneurship is indicative of economic development, with policies focusing on strengthening entrepreneurial framework conditions in developed economies (Acs, 2006).

The influence of cognitive, economic, and social factors on entrepreneurial intention among males and females in India delves into the significance of self-efficacy, opportunity perception, and social recognition in shaping entrepreneurial intentions (Saleem, Ali & Arafat, 2022). The study highlights the importance of understanding gender-specific entrepreneurial behaviors and the impact of social valuation on entrepreneurial intentions. A recent study by Lee and Kang (2022) delves into personal characteristics of various factors influencing entrepreneurial intention such as achievement needs, self-efficacy, and innovativeness, environmental characteristics like the presence of an entrepreneurship mentor, and entrepreneurship-related traits such as social awareness, business strategy, and risk sensitivity. Self-efficacy, emotional intelligence, risk-taking propensity, and personal attitude are key factors influencing entrepreneurial intention (Bilgiseven & Kasimoglu, 2019). There exists a positive correlation between attitude and behavior, as well as between planned behavior model factors and entrepreneurial intention. Past studies highlight the importance of individual traits and characteristics in start-up decision-making (Sanchez & Sahuquillo, 2012). Factors such as independence desire, risk propensity, need for achievement, internal locus of control, and preference for innovation are highlighted as key determinants. The role of personal, environmental, and behavioral factors in shaping human behavior, particularly focusing on self-efficacy, outcome expectations, social support, and cultural norms has a lot of emphasis (Munyaradzi, 2023). The study integrates path analysis and fuzzy-set qualitative comparative analysis to examine the complex relationships between entrepreneurial knowledge and skills, fear of failure, self-efficacy, and early-stage entrepreneurial activity. Matic (2023) explores the determinants influencing entrepreneurs in the logistics sector, focusing on demographic factors, growth expectations, internationalization, and motivational aspects. Utilizing the Global Entrepreneurship Monitor (GEM) database, the study analyzes data from 192 logistics entrepreneurs out of 73,806 respondents. They have highlighted the importance of gender, age, and employment status in entrepreneurial decisions within the logistics industry, emphasizing the significance of financial incentives and innovation in driving entrepreneurship in logistics.

The methodological advancements are brought by decision tree models in enhancing predictive accuracy for early-stage entrepreneurship prediction (Graham & Bonner, 2022) and can be further improved using Random forest algorithms. The integration of machine learning techniques is highlighted for its effectiveness in identifying and analyzing factors that impact total early-stage entrepreneurship. Moreover, the review points out the significance of external factors like media coverage and the ease of business establishment in influencing entrepreneurial activity. It stresses the necessity of validating models on unseen data to evaluate their predictive performance accurately. The qualitative aspects of entrepreneurial attitudes have considerable significance in shaping entrepreneurial activity across nations (Beynon and Jones, 2020). By utilizing fuzzy-set qualitative comparative analysis (fsQCA), the study explores the interplay between different entrepreneurial attitudes and their impact on Total Early-Stage Entrepreneurial Activity (TEA). It underscores the importance of understanding the subjective perceptions, motivations, and beliefs of individuals towards entrepreneurship in driving actual entrepreneurial behavior. Predictive models based on ANN can provide insights into entrepreneurial intentions by considering attributes related to family background, social environment, and university factors, despite challenges posed by dataset imbalance (Sandoval & Hernandez, 2021). The use of decision tree algorithms allows for a comprehensive analysis of the factors influencing entrepreneurial intentions without making distribution assumptions, making it a valuable tool for understanding the complex dynamics of entrepreneurial behavior (Xu and Graham, 2022).

Previous studies have underscored the significance of using certified knowledge in GEM data analysis (Sánchez-Escobedo et al., 2012). The application of bibliometric methods has provided valuable insights into the intellectual foundations of gender analysis using GEM data. Research contributions have focused on the qualitative development of the research life cycle, analysis of topics, productive authors, and the involvement of various countries and institutions. The evolution of entrepreneurial activity in Romania and Hungary from 2007 to 2012 was analyzed, utilizing data primarily from the GEM database (Nelu Eugen Popescu, 2013). It delves into the definitions of entrepreneurship and the challenges in establishing clear meanings due to multifaceted interpretations. Dutta & Sobel (2021) have extensively utilized the GEM database to investigate how economic freedom can potentially alleviate the inhibiting effect of fear of failure on entrepreneurial endeavors, offering a novel perspective on the interplay between institutional factors and entrepreneurial behavior. Tunali and Sener (2019) utilize data from the GEM database to investigate entrepreneurship behaviors and attitudes in Turkey. The study draws on GEM's 2013 Adult Population Survey (APS) Global Individual Level Data to analyze demographic, economic, and perceptual variables influencing entrepreneurship aiding policymakers in designing effective strategies to promote entrepreneurship.

A notable void emerges regarding the consideration of a nation's economic status and its impact on individual entrepreneurial intention. This gap underscores the imperative for our study, where we categorize countries into 'High Income,' 'Middle Income,' and 'Low Income' sectors to comprehensively analyze their effects on entrepreneurial aspirations. Moreover, our investigation goes beyond conventional parameters by integrating cognitive dimensions such as 'vision,' 'creativity,' and 'recognizing business opportunities' into our predictive model, alongside established factors like 'age,' 'gender,' and 'fear of failure.' Our research aims not only to address this gap in existing literature but also to offer a nuanced understanding of the complex interplay between macroeconomic contexts, individual characteristics, and entrepreneurial intent.

3 Theoretical Background

We draw the main theoretical understanding from the works of Graham and Bonner, 2022. We use the literature and extend it to three divisions of countries, namely, Low Income, Middle Income, and High-Income countries to study the entrepreneurial abilities in these income regions and thus draw a contrast to the entrepreneurial activities in each region. We delve into early-stage entrepreneurship, leveraging the rich theoretical framework provided by Social Cognitive Theory (SCT). Drawing from the foundational work of Wood and Bandura (1989), we explore the intricate interplay of individual traits, environmental factors, and past behaviors in motivating and shaping entrepreneurial actions. Despite the acknowledged significance of social-cognitive factors in entrepreneurship literature, a comprehensive understanding of the diverse combinations of these factors and their relative importance still needs to be discovered.

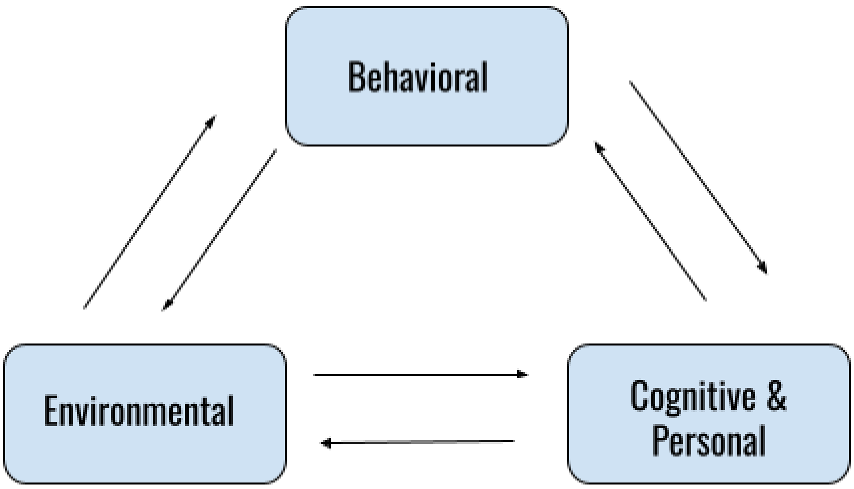


Figure 1: Relationship between SCT factors. Source: based on (Wood & Bandura, 1989)

Behavioral factors: Creativity, Vision, proactiveness, and other behavioral aspects.
Environmental factors: Good neighborhood,
Cognitive and Personal factors: These include factors such as Self-Efficacy, Networks, Opportunity recognition, and Fear of failure.

We also draw upon the Theory of Planned Behavior (TPB) as another foundational model for the theoretical framework. According to TPB, attitudes toward entrepreneurship serve as significant determinants of entrepreneurial intention. This theory posits that an individual's attitudes, subjective norms, and perceived behavioral control collectively influence their intention to engage in entrepreneurial activities. Within this framework, elements such as a good neighborhood and education play a crucial role in shaping attitudes towards entrepreneurship by providing a supportive environment and essential skills. Self-efficacy is central to TPB as it directly impacts perceived behavioral control, influencing both intentions and actions. Networks and opportunity recognition are vital for enhancing subjective norms and perceived behavioral control, as they provide resources and information critical for entrepreneurial activities. Conversely, fear of failure can negatively impact attitudes and perceived behavioral control, potentially deterring entrepreneurial intentions. Creativity, vision, and proactiveness are key behavioral aspects that shape positive attitudes toward entrepreneurship and enhance perceived behavioral control by fostering innovative thinking and a proactive approach to opportunities. These factors interact within the TPB framework to influence entrepreneurial intentions and behaviors, highlighting the importance of both individual traits and external influences in the entrepreneurial process.

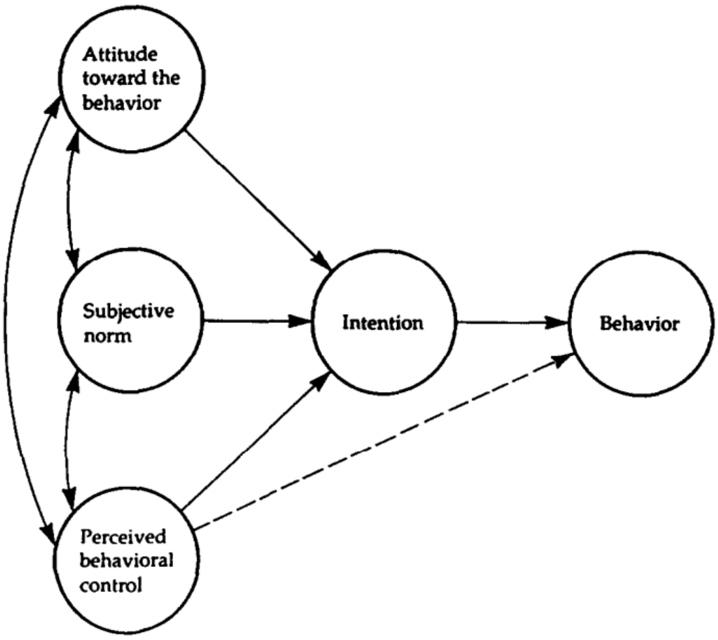


Figure 2: Theory of Planned Behavior. Source: (Ajzen, 1991)

4. Hypothesis Development

Prior entrepreneurial exposure equips individuals with invaluable insights, skills, and networks, fostering a deep understanding of the entrepreneurial journey. This exposure not only imparts practical knowledge but also cultivates confidence in one's ability to navigate the complexities of entrepreneurship. Drawing from the experiences of successful entrepreneurs, individuals glean essential strategies and best practices, strengthening their entrepreneurial acumen. Moreover, these networks established through prior exposure serve as conduits for support, mentorship, and collaboration, further fueling individuals' intentions to embark on entrepreneurial ventures. Therefore, prior entrepreneurial exposure plays a pivotal role in shaping individuals' aspirations and readiness to engage in entrepreneurial endeavors (Graham & Bonner, 2022). Therefore, the hypothesis proposed:

H1: Networks/role models positively influence entrepreneurial intentions.

Research conducted by Anjum, Farrukh, Heidler, and Tautiva (2020) has shown that individuals' perceived disposition for creativity and their attitude toward entrepreneurship significantly affect their entrepreneurial intentions. Individuals with a strong inclination towards creativity are more likely to generate innovative ideas, identify market opportunities, and devise novel solutions to challenges encountered in the entrepreneurial journey. Furthermore, a positive attitude towards entrepreneurship fosters a mindset conducive to risk-taking, resilience, and persistence, essential qualities for navigating the uncertainties inherent in entrepreneurial endeavors. Therefore, the hypothesis proposed:

H2: Creativity positively influences entrepreneurial Intentions

Evidence suggests that individuals who perceive greater opportunities to initiate startups exhibit stronger intentions to engage in entrepreneurial activities, reflecting a positive correlation between perceived opportunity and entrepreneurial intention. This alignment underscores the importance of recognizing and capitalizing on perceived opportunities as catalysts for fostering entrepreneurial aspirations and ventures (Stuetzer, Obschonka, Brixy, Sternberg & Cantner, 2014). Individuals who perceive favorable opportunities are more likely to be motivated to pursue entrepreneurial endeavors, driven by the potential for innovation, growth, and success. Therefore, the hypothesis proposed:

H3: Opportunity perception positively influences entrepreneurial intentions

Fear of failure acts as a formidable barrier to entrepreneurial intention by eroding individuals' confidence, constraining their propensity for risk-taking, and curtailing their eagerness to seize entrepreneurial opportunities. Research by Kong, Zhao, and Tsai (2020) and Ng and Jenkins (2018) supports this notion. This fear undermines individuals' willingness to embark on entrepreneurial ventures, thwarting their aspirations for entrepreneurship. The apprehension of failure can paralyze individuals, preventing them from pursuing innovative ideas or venturing into new markets. Consequently, addressing and mitigating the fear of failure is crucial for fostering entrepreneurial ambitions and cultivating a conducive environment for entrepreneurial success. Therefore, the hypothesis proposed:

H4: Fear of failure negatively influences entrepreneurial intentions

Self-efficacy, a foundational concept in social cognitive theory, plays a vital role in the framework. Elevated levels of self-efficacy, indicative of an individual's confidence in their aptitude to proficiently undertake entrepreneurial endeavors, are anticipated to correlate with intensified entrepreneurial intentions. Those possessing heightened self-efficacy are inclined to view themselves as adept at surmounting obstacles and capitalizing on entrepreneurial prospects. This belief in one's capabilities fosters a proactive mindset, wherein individuals feel empowered to navigate challenges and seize opportunities in the entrepreneurial landscape. Self-efficacy is also directly related to skills and knowledge in a particular area (Luthje & Franke 2003; Liñán et al. 2007; Graham & Bonner, 2022). Consequently, self-efficacy emerges as a critical determinant shaping the strength of entrepreneurial aspirations and the likelihood of engaging in entrepreneurial activities. Therefore, the hypothesis proposed:

H5: Perceived Skills positively influence entrepreneurial intentions.

Perceived feasibility and desirability are pivotal in driving entrepreneurial intentions, as outlined in Ajzen's Theory of Planned Behavior. Feasibility reflects individuals' beliefs in the practicality of starting a venture, while desirability encompasses subjective evaluations of the attractiveness and personal fulfillment associated with entrepreneurship. Past research consistently highlights their significance: individuals are more motivated to pursue entrepreneurial opportunities when perceived as feasible and desirable. These constructs shape attitudes and motivations, aligning with Ajzen's framework, where behaviors are driven by perceived feasibility and desirability. Policymakers and educators can leverage these insights to cultivate environments conducive to fostering entrepreneurial ambitions, emphasizing the importance of addressing subjective perceptions in understanding and promoting entrepreneurial behavior.

Stephan, Hart, Mickiewicz, and Drews (2015) found a positive correlation between the perceived desirability of entrepreneurship at the country level and individuals' entrepreneurial intentions. Countries with a strong entrepreneurial culture and supportive ecosystems tend to foster higher levels of desirability, which in turn influences individuals to pursue entrepreneurial endeavors. Therefore, the hypothesis proposed:

H6: Higher desirability to start up in a country leads to having higher entrepreneurial intentions.

Shane and Venkataraman (2000) and Davidsson (2006) emphasize the significance of opportunity recognition in driving entrepreneurial intentions. Individuals with a keen ability to identify and capitalize on business opportunities are more likely to harbor intentions to engage in entrepreneurial activities, as they perceive entrepreneurship as a means to exploit identified opportunities and create value. Therefore, the hypothesis proposed:

H7: The ability to see business opportunities leads to higher entrepreneurial intentions.

Obschonka et al. (2019) indicate that individuals with a long-term career vision are more inclined toward entrepreneurship. They view entrepreneurship as a pathway to realizing their career aspirations, achieving autonomy, and creating a lasting impact. This long-term perspective fosters stronger entrepreneurial intentions, as individuals perceive entrepreneurship as a viable avenue for fulfilling their career goals and aspirations. Therefore, the hypothesis proposed:

H8: The vision of a long-term career has a positive influence on entrepreneurial intention.

Gender is identified as a crucial factor influencing entrepreneurial intentions. Societal norms, expectations, and cultural perceptions often shape individuals' attitudes and behaviors toward entrepreneurship differently based on their gender identity. Therefore, gender is hypothesized to play a vital role in either supporting or hindering entrepreneurial intentions. It is observed that most entrepreneurs are males. (Zhang, Duysters & Cloodt, 2014). Men exhibit a greater tendency to contemplate the idea of establishing a firm compared to demonstrating a firm determination to actualize it. Nonetheless, the direct influence of this inclination is mitigated by an interaction effect: men who perceive a stronger alignment between traits traditionally associated with masculinity and those conducive to entrepreneurship are more inclined towards firm entrepreneurial intentions. (Díaz-García & Jiménez-Moreno, 2010). Therefore, the hypothesis proposed:

H9: Males have higher entrepreneurial intentions than females

5. Methodology

5.1 Data.

Utilizing the complete dataset of working-age individuals (18–35 years old) sourced from the 2019 Global Entrepreneurship Monitor (GEM) surveys, this analysis examines Entrepreneurial Intentions (EI), defined as the inclination to initiate a business venture in the future. After thorough data cleaning, which involved excluding entries with missing values, the final dataset comprised 36,828 respondents. The variables of interest, including EI, are incorporated in this dataset, representing the most recent release of GEM data available for analysis.

Entrepreneurial Intention is evaluated on a binary scale, denoted by 0 for absence and 1 for presence. This assessment is conducted at the individual level and encompasses factors such as prior behaviors, environmental perceptions, and cognitive and psychological traits. Examples of cognitive elements include fear of failure, networks, opportunities, and self-efficacy, all measured on binary scales (0 = no, 1 = yes), as delineated in Table 1. Additionally, demographic variables such as age, gender, education, and household size represent supplementary personal traits within the analysis.

Table 1. *List of variables.*

GEM Variable	Name	Labels	Label meaning
age	what is your current age (in years)	0-100	
gender	What is your gender?	1,2	"Refused" "Don't Know" "Male" "Female"
hh size	How many members make up your permanent household, including you?	1-100	

UNEDUC	UN harmonized educational attainment	(0,8)	"Pre-primary education" "Primary education or first stage of basic education" "Lower secondary or second stage of basic education" "(Upper) secondary education" ...
knowentR	How many people do you know personally who have started a business or become self-employed	(-2,3)	"Refused" "Don't know" "None" "One" ...
opportL	In the next six months, there will be good opportunities for starting a business in the area where you live.	(-2,5)	"Refused" "Don't know" "Strongly disagree ""Somewhat disagree" ...
suskillL	You personally have the knowledge, skill and experience required to start a new business.	(-2,5)	"Refused" "Don't know" "Strongly disagree ""Somewhat disagree" ...
fearfailL	You would not start a business for fear it might fail.	(-2,5)	"Refused" "Don't know" "Strongly disagree ""Somewhat disagree" ...
nbgoodcL	In my country, most people consider starting a new business a desirable career choice.	(-2,5)	"Refused" "Don't know" "Strongly disagree ""Somewhat disagree" ...

oppismL	You rarely see business opportunities, even if you are very knowledgeable in the area.	(-2,5)	"Refused" "Don't know" "Strongly disagree ""Somewhat disagree" ...
proactL	Even when you spot a profitable opportunity, you rarely act on it.	(-2,5)	"Refused" "Don't know" "Strongly disagree ""Somewhat disagree" ...
creativL	Other people think you are highly innovative."	(-2,5)	"Refused" "Don't know" "Strongly disagree ""Somewhat disagree" ...
visionL	Every decision you make is part of your long-term career plan.	(-2,5)	"Refused" "Don't know" "Strongly disagree ""Somewhat disagree" ...
FUTSUPNO	Entrepreneurial intentions (in 18-25 sample that is not involved in entrepreneurial activity)	(0,1)	"No" "Yes"

Source: GEM database 2019

5.2 Descriptive Statistics

Table 2. *Descriptive Statistics*

GEM Variable	Labels	Frequency
gender	1 = Male	18346
	2 = Female	18482
UNEDUC	0	443
	1	1671
	2	5702
	3	10829
	4	5344
	5	1728
	6	8515
	7	2512
	8	84
knowentR	0	17378
	1	7536
	2	8801
	3	3113
opportL	1	6799
	2	6939
	3	5832
	4	10861
	5	6397
suskillL	1	6799

	2	6939
	3	5832
	4	10861
	5	6397
fearfailL	1	7299
	2	6609
	3	4461
	4	9886
	5	8573
oppismL	1	4157
	2	7278
	3	6918
	4	11807
	5	6668
proactL	1	4729
	2	6623
	3	5330
	4	11705
	5	8441
creativL	1	3406
	2	5542
	3	7912
	4	12643
	5	7325
visionL	1	2968
	2	4666
	3	4791
	4	11521
	5	12882

nbgoodcL	1	3714
	2	6359
	3	5781
	4	12080
	5	8894
FUTSUPNO	0	26381
	1	10447

Source: GEM database 2019

5.3 Machine Learning Methods

This research uses a machine-learning methodology to forecast EI. Algorithms are used in machine learning to identify patterns in data. Machine learning uses algorithms to identify patterns in data. There are two main types of machine learning: supervised and unsupervised. Supervised learning is a type of machine learning that utilizes labeled datasets to train algorithms for predicting outcomes and identifying patterns. In contrast to unsupervised learning, supervised learning algorithms are provided with labeled training data to understand the relationship between inputs and outputs. This paper employs supervised learning methods.

The present work employs a machine learning approach that comprises many stages, namely: selecting training and test sets of data, training the models, assessing accuracy, and analyzing the results. According to Kuhn and Johnson (2013), these actions are characteristic of a supervised machine-learning technique. Below is a more detailed discussion of these phases.

5.3.1 Machine-Learning algorithms & model training

Using random stratified sampling, the data is divided into two halves for training and testing: 70% of the data is utilized for training the 30% of the data is utilized to assess the model's performance using machine learning models i.e. for testing. This division makes room for an adequate amount of data to both train the model and provide a dispassionate evaluation of the forecasting capabilities of the models. By assessing the model's correctness using a holdout test set, we may get a more impartial evaluation of the model's functionality.

5.3.1.1 Decision Tree Algorithm

Decision Trees are widely recognized as one of the most popular classifier creation methods. The decision tree algorithm is a widely used method in supervised learning, employed for tasks including classification and regression. Initially proposed by Hunt et al. in 1966, it operates by recursively partitioning the feature space into smaller regions. At each node of the tree, a decision is made based on the value of a chosen feature to optimize a splitting criterion. Selecting the feature that maximizes the gain or minimizes impurity is crucial and is typically determined through various algorithms, including ID3, CART, or C4.5. This recursive partitioning continues until a termination condition is met, such as reaching a maximum depth or achieving a certain level of purity within the nodes.

When applied to a new data instance, the decision tree moves from the root to a leaf node, determining the path based on the feature values of the instance and the decisions made at each node. The prediction associated with the leaf node reached by the instance provides the output value or class label.

5.3.1.2 Random Forest Algorithm

The Random Forest algorithm, introduced by Breiman in 2001, is a powerful ensemble learning method extensively applied in predictive modeling tasks. In the context of predicting entrepreneurial intention, Random Forest offers several advantages including robustness to noise, scalability to large datasets, and resilience against overfitting. The algorithm operates by constructing multiple decision trees through a process called bootstrap aggregating or bagging. Each decision tree is trained on a random subset of the training data, with replacement. A random subset of features is considered for splitting at each node of the tree, aiming to promote diversity among the individual trees. Once the ensemble of decision trees is constructed, predictions are made by aggregating the outputs of individual trees. For classification tasks such as predicting entrepreneurial intention, the most commonly occurring class among the predictions of individual trees is typically chosen as the final prediction.

In the case of the GEM database, Random Forest effectively leverages the diverse set of features related to entrepreneurial activities, economic conditions, and sociocultural factors to predict entrepreneurial intention. Furthermore, Random Forest's ability to handle categorical and numerical features without extensive preprocessing makes it well-suited for analyzing GEM data.

5.3.1.3 XGBoost Algorithm

XGBoost, short for eXtreme Gradient Boosting, is a highly efficient and scalable gradient-boosting framework introduced by Chen and Guestrin in 2016. The algorithm is based on an ensemble learning technique known as gradient boosting, where weak learners, typically decision trees, are iteratively run and thus trained to correct the errors made by the preceding models, according to J.H Friedman (2001). XGBoost enhances traditional gradient boosting by incorporating several innovative features such as a novel regularization term to control model complexity, a custom loss function tailored to specific tasks, and a tree construction algorithm that optimizes both computational speed and model accuracy.

In the context of predicting EI from the GEM database, XGBoost can effectively handle the diverse set of features present in the GEM data, including socio-economic indicators, cultural factors, and individual characteristics of potential entrepreneurs. Furthermore, XGBoost's ability to automatically handle missing values and robustness to outliers makes it particularly suitable for analyzing real-world datasets with inherent complexities.

5.3.3 Class Imbalance

Class imbalance occurs when one class (in this case, the presence or absence of entrepreneurial intention) is significantly more prevalent than the other in the dataset. This imbalance can lead to biased model predictions, where the model may exhibit higher accuracy in predicting the majority class while performing poorly on the minority class. Addressing class imbalance is crucial to ensure the model's effectiveness in accurately capturing the patterns and relations found in the data. In the analysis of Entrepreneurial Intention (EI), a notable class imbalance was identified within the dependent variable. (Kuhn & Johnson, 2013)

To tackle the class imbalance in the dataset, we used the Synthetic Minority Over-sampling Technique (SMOTE). SMOTE is a popular method in machine learning to address class imbalance. It works by generating synthetic samples for the minority class, thus balancing the class distribution in the dataset. The SMOTE technique was applied after splitting the data into training and testing sets (Burez & Van den Poel, 2009; Chawla et al., 2002). This ensures that synthetic samples are only generated from the training data, preventing information leakage into the testing set and maintaining the integrity of the evaluation process.

5.3.4. Evaluating and interpreting the model

Using the holdout test set, the model's prediction accuracy is assessed. The area under the receiver operating characteristic curve (AUC-ROC) is the primary measure used to assess the model accuracy because of the class imbalance of the dependent variables (Bradley, 1997; Burez & Van den Poel, 2009). Model interpretation makes use of the variable by displaying the tree structure and assigning significance values. The decrease in the loss function resulting from the splits and surrogate splits on the variable is the basis for calculating the variable significance measures (Thereneau & Atkinson, 2015). To strengthen the validity of our findings, we performed additional data analysis using the previously described methodology.

6. Results and Discussion

In this study, we adopted an approach that prioritized the importance of variables in determining entrepreneurial intention, arranging them according to their significance. The primary dependent variable under investigation was entrepreneurial intention. The dataset was partitioned into three categories based on income brackets: low-income countries (10,365 entries), middle-income countries (10,452 entries), and high-income countries (9,588 entries) following data cleaning procedures. The results from three distinct machine learning models—Random Forest, XGBoost, and Decision Trees—have been presented and compared. Notably, Random Forest exhibited the highest accuracy among the models examined.

Table 3. *Summary of 3 Models.*

	Decision Tree	**Random Forest**	**XGBoost**
Accuracy	62.6	71.5	70.9
AUC-ROC	0.56	0.69	0.67

Upon analysis, it was observed that factors such as network size, startup skills, and creativity consistently emerged among the top-ranking variables across all models. Particularly striking was the prominence of network size and startup skills, suggesting a strong correlation between the breadth of an individual's professional network and their entrepreneurial intentions. Additionally, the importance of possessing relevant skills in entrepreneurship underscores their role in shaping an individual's inclination toward starting a venture. Furthermore, creativity emerged as a crucial determinant, highlighting the significance of innovative thinking in fostering entrepreneurial aspirations. These findings collectively suggest that fostering a diverse network, honing startup-related skills, and nurturing creativity may significantly influence an individual's propensity towards entrepreneurship, regardless of their economic context.

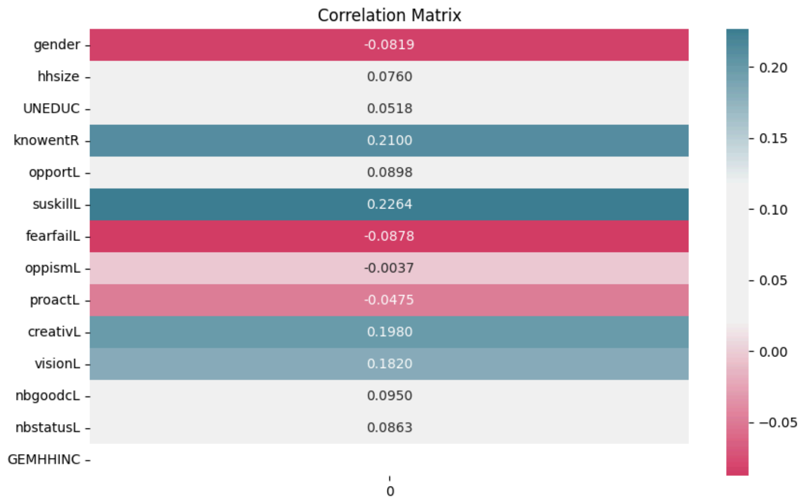


Figure 3: *Correlation matrix of all the variables*

The results of the analysis reveal significant insights into the factors influencing entrepreneurial intentions among individuals aged 18 to 35. The research findings indicate a significant positive correlation between exposure to networks and role models and the development of entrepreneurial intentions, as observed in Figure 1. This suggests that individuals who have access to influential networks and role models are more inclined towards entrepreneurship. The presence of a larger social network and exposure to entrepreneurial cultures appear to stimulate novel ideas and foster motivation toward initiating a startup or pursuing entrepreneurship. These findings align with previous research, underscoring the substantial influence of networks on shaping entrepreneurial intentions. The Network/role models factor is also a significant variable and one of the most important factors in affecting entrepreneurial intentions in countries with middle and high income specifically. In low-income countries, it matters less although it does have a significant role.

Table 4. *Classification report of the Random forest model for LOW, MIDDLE, and HIGH-income Economies.*

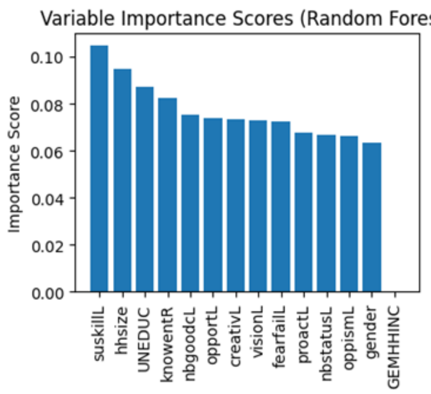
Economy	**Accuracy**	**AUC-ROC**		**Precision**	**Recall**	**f1-score**	**Support**
LOW	71.5	0.69	0	0.75	0.91	0.82	2203
			1	0.53	0.24	0.33	907
MIDDLE	71.9	0.69	0	075	0.89	0.82	2210

			1	0.54	0.30	0.39	926
HIGH	71.2	0.71	0	0.75	0.90	0.82	2049
			1	0.50	0.24	0.32	828

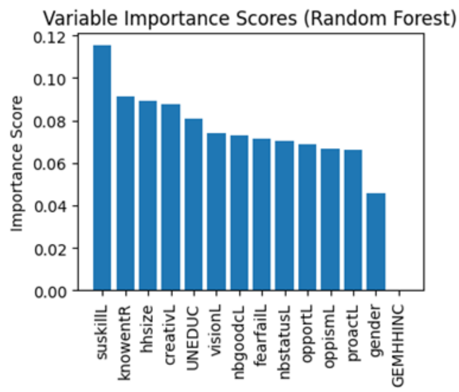
The study identifies creativity as a key driver of entrepreneurial intentions, suggesting that individuals with a greater propensity for innovative thinking are more likely to harbor intentions to pursue entrepreneurial ventures. Also interestingly, creativity has a more significant role to play in middle and high-income countries and not so much in lower-income countries. However, while fear of failure exhibits a negative correlation with entrepreneurial intentions, its significance in determining these intentions is found to be relatively low. While this negative association aligns with prior research, the relatively low significance of fear of failure in predicting entrepreneurial intentions, not ranking among the top five determining features within the studied age group, presents a slight contradiction to previous findings. This is true for all three types of economies. Moreover, the correlation between entrepreneurial competence and problem-solving propensity accentuates the significance of skill development programs in nurturing entrepreneurial ecosystems. These findings align with previous studies, reinforcing the notion that cultivating competencies enhances individuals' propensity towards entrepreneurial endeavors, ultimately contributing to innovation and economic growth. Skills and knowledge are important factors across all types of economies without exception.

Furthermore, the study highlights the significant impact of neighborhood support on entrepreneurial intentions, particularly in low-income countries where environmental factors play a pivotal role in shaping individuals' mindsets and aspirations. Past research consistently highlights the profound impact of sociocultural environments on entrepreneurial behavior, with strong community networks fostering resource mobilization, risk mitigation, and role modeling. In low-income contexts where formal institutional support may be lacking, neighborhood support serves as a vital catalyst for entrepreneurship, providing aspiring entrepreneurs with essential resources and resilience to overcome economic challenges. The analysis unveils notable disparities across income brackets. In low-income countries, education emerges as a critical determinant of entrepreneurial intentions, indicating a higher significance compared to other factors. Conversely, in middle and low-income countries, skills are found to hold greater importance than networks, contrasting with higher-income countries where networks and role models are more influential. Moreover, creativity emerges as a more significant factor in middle and higher-income countries compared to their low-income counterparts.

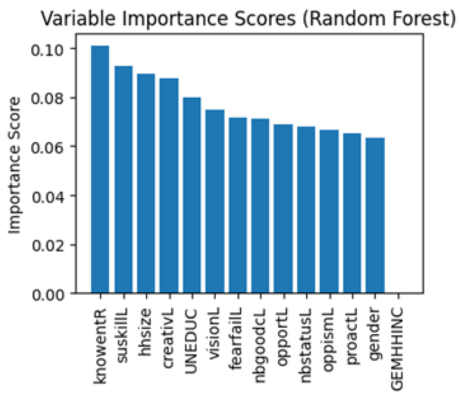
Interestingly, household size is identified as a determining factor in entrepreneurial intentions, with a positive correlation indicating that individuals from larger households are more inclined towards entrepreneurship. Household size seems to be more relevant or important for low-income countries. A possible explanation for this could be that a poor household, with many members in the family, indirectly could influence the family members' tendency to have a business mind and a tendency to build something that has a positive impact on the world.



A.



B.



C.

Figure 4: *Variable Importance for A. Low-income Economies. B. Middle-Income Economies. C. High-Income Economies.*

In our analysis from Figure 2, vision demonstrates a positive correlation with entrepreneurial intentions across all three types of economies. However, the strength of this correlation is not as pronounced as other factors examined in the study. Despite its lesser significance, the relationship between vision and entrepreneurial intentions remains consistent, suggesting that having a clear vision or future outlook plays a role in shaping entrepreneurial aspirations across diverse economic landscapes.

Factors such as the ability to see opportunities for business and the ability to spot profitable ventures do not appear to be significantly influential in affecting entrepreneurial intentions across any type of economy. These factors also exhibit a negative correlation with entrepreneurial intentions, suggesting that they may even deter individuals from pursuing entrepreneurial paths. This observation is further supported by our random forest model, which underscores the limited impact of these particular factors on shaping entrepreneurial aspirations. Thus, while these abilities are often highlighted as essential entrepreneurial traits, our findings indicate that they may not play a substantial role in driving entrepreneurial intentions.

7. Conclusion and Implications

Entrepreneurial Intentions are complex to understand. The multifaceted determinants of entrepreneurial intentions among individuals aged 18 to 35 are observed, drawing upon an analysis of diverse socio-economic contexts. The findings underscore the importance of networks, role models, creativity, and skills in fostering entrepreneurial aspirations. This research highlights the complexity of entrepreneurial intentions, shaped by a combination of personal traits, socio-cultural environments, and economic conditions. Moreover, the study reveals notable disparities across income brackets, highlighting the differential influences of various factors on entrepreneurial intentions. Despite certain limitations, the insights obtained from this study offer valuable implications for policymakers, educators, and practitioners seeking to nurture an environment conducive to entrepreneurship.

Networks and role models take precedence in higher-income countries and they are the most important factor in influencing EI. Entrepreneurial competence and problem-solving propensity highlight the crucial role of skill development programs in nurturing entrepreneurial ecosystems. Having the necessary skills increases the propensity to start a business. Creativity emerges as a key driver, particularly influential in middle and high-income countries, underscoring the importance of innovative thinking in fostering entrepreneurial aspirations. While fear of failure exhibits a negative correlation with entrepreneurial intentions, its overall significance remains relatively low, challenging some previous findings. Neighborhood support stands out as a significant factor, especially in low-income countries, emphasizing the role of socio-cultural environments in shaping entrepreneurial behavior and resilience. Education and skills hold varying degrees of importance across income brackets. In low-income countries, education emerges as a critical determinant, while skills are prioritized in middle and low-income countries. Additionally, household size demonstrates a positive correlation with entrepreneurial intentions, particularly relevant in low-income economies, possibly reflecting the influence of familial entrepreneurship and the drive to create positive impacts within larger family units.

Despite the thorough analysis carried out in this study, it's important to recognize several limitations. First, depending on self-reported data could lead to response biases, potentially impacting the accuracy and reliability of the findings. The cross-sectional nature of the data limits the ability to establish causality between variables. Furthermore, the study's focus on individuals aged 18 to 35 may not capture the full spectrum of entrepreneurial intentions across different age groups. Moreover, the use of machine learning models, while providing valuable insights, may overlook nuanced relationships between variables. The findings of this study hold significant policy implications for fostering an entrepreneurial ecosystem conducive to economic growth and innovation. Policymakers should prioritize initiatives aimed at expanding access to networks and role models, particularly in low-income countries where these factors exhibit a greater influence on entrepreneurial intentions. Efforts to promote creativity and skill development should be integrated into educational curricula to cultivate an entrepreneurial mindset from an early age. Moreover, policies aimed at reducing the fear of failure and providing support structures, such as mentorship programs and access to financial resources, can further stimulate entrepreneurial activity across different socio-economic contexts.

Drawing from the findings of this study, future research could delve into various areas for additional exploration. Future research could delve deeper into the interplay between these factors and explore additional determinants to further enrich our understanding of entrepreneurial intentions globally. Longitudinal studies could be conducted to examine the dynamic evolution of entrepreneurial intentions over time and assess the long-term impact of various factors on entrepreneurial behavior. Qualitative research approaches, like interviews and focus groups, could offer a deeper understanding of the motivations and obstacles affecting EI. Furthermore, comparative studies across diverse cultural and institutional contexts could elucidate the contextual factors shaping entrepreneurial aspirations. Innovative research methodologies, such as experimental designs and neuroeconomic approaches, could offer novel insights into the cognitive processes underlying entrepreneurial decision-making.

8. References

Acs, Z. (2006). How is entrepreneurship good for economic growth? *innovations*, *1*(1), 97-107.

Acs, Z. J., & Szerb, L. (2007). Entrepreneurship, economic growth, and public policy. *Small business economics*, *28*, 109-122.

Acs, Z. J., Arenius, P., Hay, M., & Minniti, M. (2005). 2004 Global entrepreneurship monitor. London Business School and Babson College. London, UK, and Babson Park, MA.

Ajzen, I. (1991). The theory of planned behavior. Organizational behavior and human decision processes, 50(2), 179-211.

Anjum, T., Farrukh, M., Heidler, P., & Díaz Tautiva, J. A. (2020). Entrepreneurial intention: Creativity, entrepreneurship, and university support. Journal of Open Innovation: Technology, Market, and Complexity, 7(1), 11.

Arenius, P., & Minniti, M. (2005). Perceptual variables and nascent entrepreneurship. Small business economics, 24, 233-247.

Baber, H., Fanea-Ivanovici, M., & Sarango-Lalangui, P. (2024). The influence of sustainability education on students' entrepreneurial intentions. International Journal of Sustainability in Higher Education, 25(2), 390-415.

Barba-Sánchez, V., & Atienza-Sahuquillo, C. (2012). Entrepreneurial behavior: Impact of motivation factors on decision to create a new venture. *Investigaciones europeas de dirección y economía de la empresa*, *18*(2), 132-138.

Beynon, M. J., Jones, P., & Pickernell, D. (2020). Country-level entrepreneurial attitudes and activity through the years: A panel data analysis using fsQCA. *Journal of Business Research*, *115*, 443-455.

Bilgiseven, E. B. (2019). Analysis of factors leading to entrepreneurial intention. *Procedia Computer Science*, *158*, 885-890.

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern recognition, 30(7), 1145-1159.

Burez, J., & Van den Poel, D. (2009). Handling class imbalance in customer churn prediction. Expert Systems with Applications, 36(3), 4626-4636.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16, 321-357.

Díaz-García, M. C., & Jiménez-Moreno, J. (2010). Entrepreneurial intention: the role of gender. International entrepreneurship and management journal, 6, 261-283.

Dutta, N., & Sobel, R. S. (2021). Entrepreneurship, fear of failure, and economic policy. European Journal of Political Economy, 66, 101954.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.

Graham, B., & Bonner, K. (2022). One size fits all? Using machine learning to study heterogeneity and dominance in the determinants of early-stage entrepreneurship. Journal of Business Research, 152, 42-59.

Graham, B., & Bonner, K. (2024). The role of institutions in early-stage entrepreneurship: An explainable artificial intelligence approach. Journal of Business Research, 175, 114567.

Khefacha, I., Romdhane, R., & Haj Salem, H. (2023). Unveiling the relationship between entrepreneurial aspirations and prosperity: An international panel study using GEM data. *International Entrepreneurship and Management Journal*, 1-29.

Kuhn, M., & Johnson, K. (2013). Applied predictive modeling (Vol. 26, p. 13). New York: Springer.

Liñán, F., & Santos, F. J. (2007). Does social capital affect entrepreneurial intentions? International Advances in Economic Research, 13, 443-453.

Lüthje, C., & Franke, N. (2003). The ‘making of an entrepreneur: testing a model of entrepreneurial intent among engineering students at MIT. R&d Management, 33(2), 135-147.

Matić, M., Leković, B., & Bobera, D. (2023). DETERMINANTS OF ENTREPRENEURS IN LOGISTIC INDUSTRY: GEM STUDY ANALYSIS. *Facta Universitatis, Series: Economics and Organization*, 261-279.

Munyaradzi, M., Arko-Achemfuor, A., & Quan-Baffour, K. (2023). An Exploration of Comprehensive Student Support Systems in Technical Vocational Education and Training Colleges for Students with Disability. *Community College Journal of Research and Practice*, *47*(2), 106-122.

Obschonka, M., Moeller, J., & Goethner, M. (2019). Entrepreneurial passion and personality: the case of academic entrepreneurship. Frontiers in Psychology, 9, 400253.

Popescu, N. E. (2013). The evolution of entrepreneurship activity indicators in two European countries. *Procedia Economics and Finance*, *6*, 562-572.

Saleem, A., Ali, J., & Arafat, M. Y. (2022). Women entrepreneurs and agricultural start-ups: Cognitive and social capital perspective. In *Driving factors for venture creation and success in agricultural entrepreneurship* (pp. 191-212). IGI Global.

Sánchez-Escobedo, M. C., Fernández-Portillo, A., Díaz-Casero, J. C., & Hernández-Mogollón, R. (2016). Research in entrepreneurship using GEM data. Approach to the state of affairs in gender studies. *European journal of management and business economics*, *25*(3), 150-160.

Sandoval, M., & Hernández, C. (2021). Building Models based on Artificial Neural Networks to Predict Entrepreneurial Intentions among Undergraduate Students. In *Proceedings of the 4th European International Conference on Industrial Engineering and Operations Management*.

Schade, P., & Schuhmacher, M. C. (2023). Predicting entrepreneurial activity using machine learning. Journal of Business Venturing Insights, 19, e00357.

Shane, S., & Venkataraman, S. (2000). The promise of entrepreneurship as a field of research. Academy of Management Review, 25(1), 217-226.

Stephan, U., Hart, M., & Drews, C. C. (2015). Understanding motivations for entrepreneurship: A review of recent research evidence.

Stuetzer, M., Obschonka, M., Brixy, U., Sternberg, R., & Cantner, U. (2014). Regional characteristics, opportunity perception and entrepreneurial activities. Small Business Economics, 42, 221-244.

Therneau, T., Atkinson, B., Ripley, B., & Ripley, M. B. (2015). Package ‘rpart’. Available online: cran. ma. ic. ac. uk/web/packages/rpart/rpart. pdf (accessed on 20 April 2016).

Tunali, C. B., & Sener, S. (2019). The determinants of entrepreneurship in Turkey. Procedia Computer Science, 158, 648-652.

Xu, L., & Graham, B. (2022, August). Using Machine Learning to Study Entrepreneurial Attitudes and Intentions Over Time: Evidence from before, during, and after the 2008 Financial Crisis. In *Irish Academy of Management Conference*.

Zhang, Y., Duysters, G., & Cloodt, M. (2014). The role of entrepreneurship education as a predictor of university students’ entrepreneurial intentions. International entrepreneurship and management journal, 10, 623-641.

9. APPENDIX

High income

	precision	recall	f1-score	support
0.0	0.76	0.86	0.81	2049
1.0	0.49	0.33	0.39	828
accuracy			0.71	2877
macro avg	0.63	0.60	0.60	2877
weighted avg	0.68	0.71	0.69	2877

Middle income

	precision	recall	f1-score	support
0.0	0.76	0.86	0.81	2210
1.0	0.51	0.35	0.41	926
accuracy			0.71	3136
macro avg	0.63	0.60	0.61	3136
weighted avg	0.68	0.71	0.69	3136

LOW income

Accuracy: 70.9967845659164 %

	precision	recall	f1-score	support
0.0	0.75	0.89	0.81	2203
1.0	0.50	0.28	0.36	907
accuracy			0.71	3110
macro avg	0.63	0.58	0.59	3110
weighted avg	0.68	0.71	0.68	3110

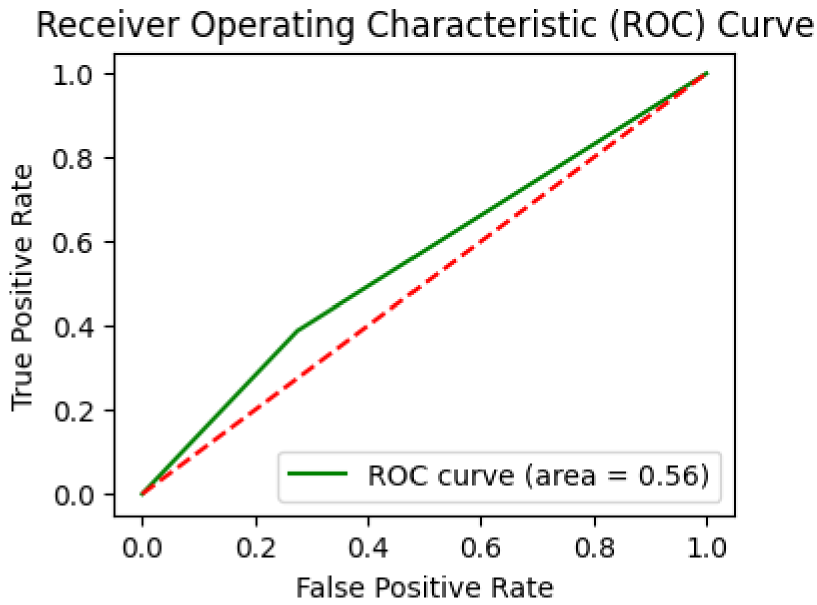


Figure: ROC for Decision tree

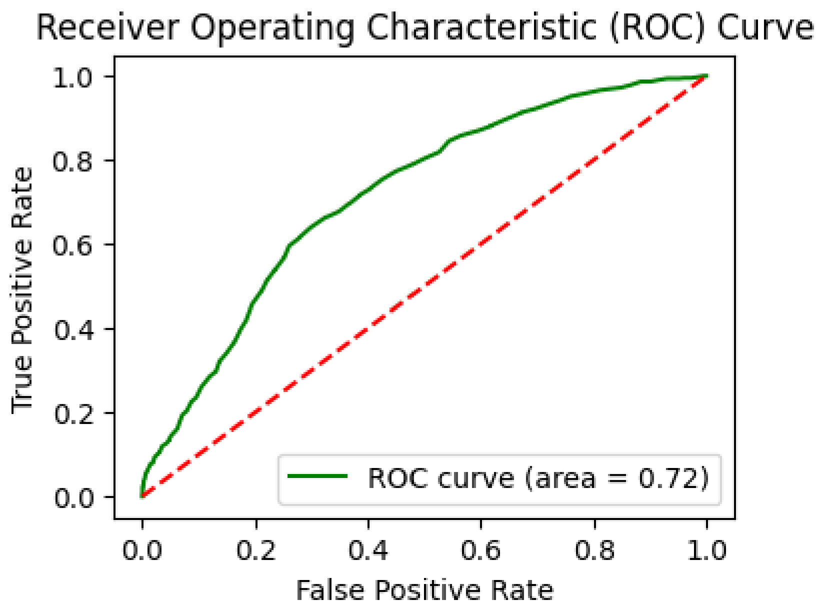


Figure: ROC for Random Forest

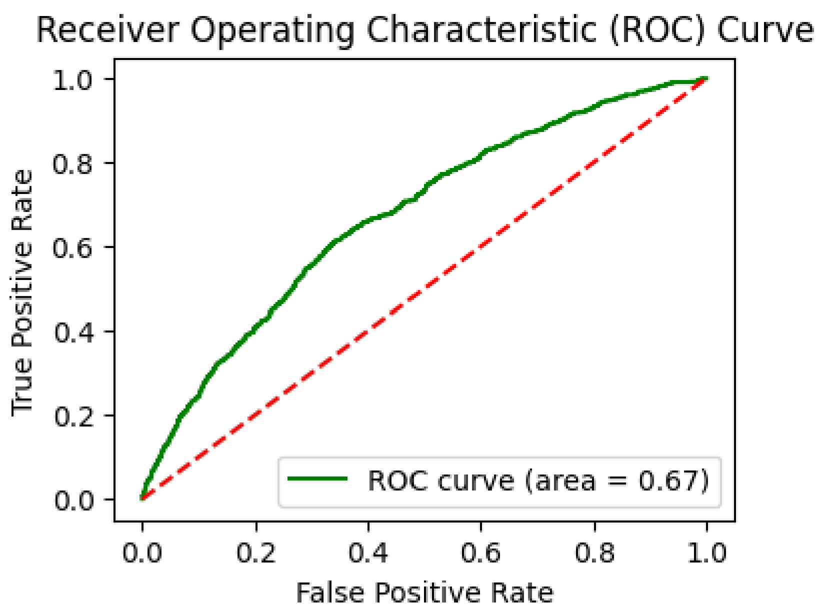


Figure: ROC for XGBoost

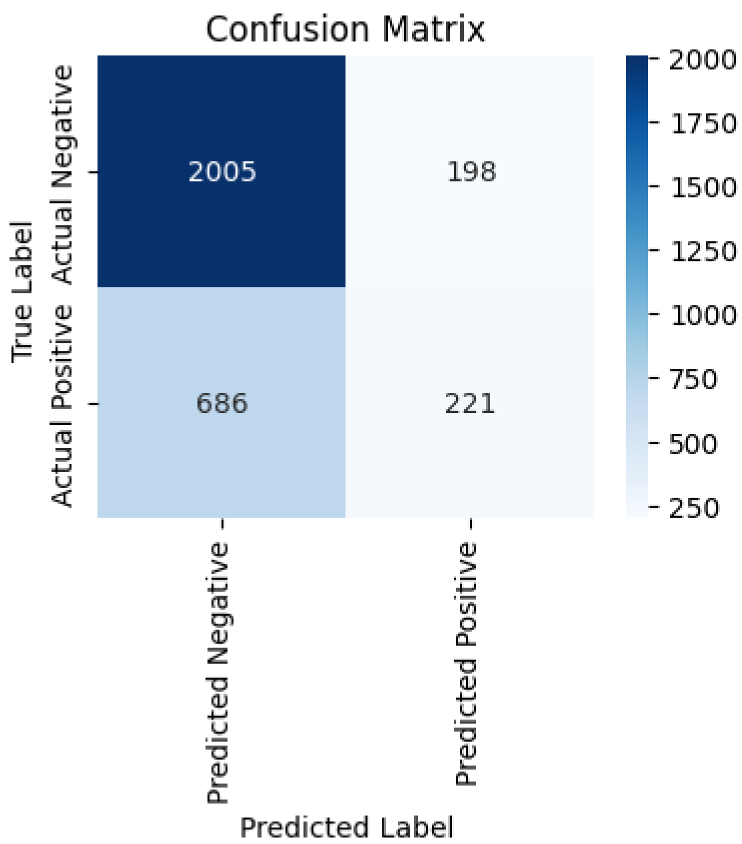


Figure: Confusion matrix of Low-income countries

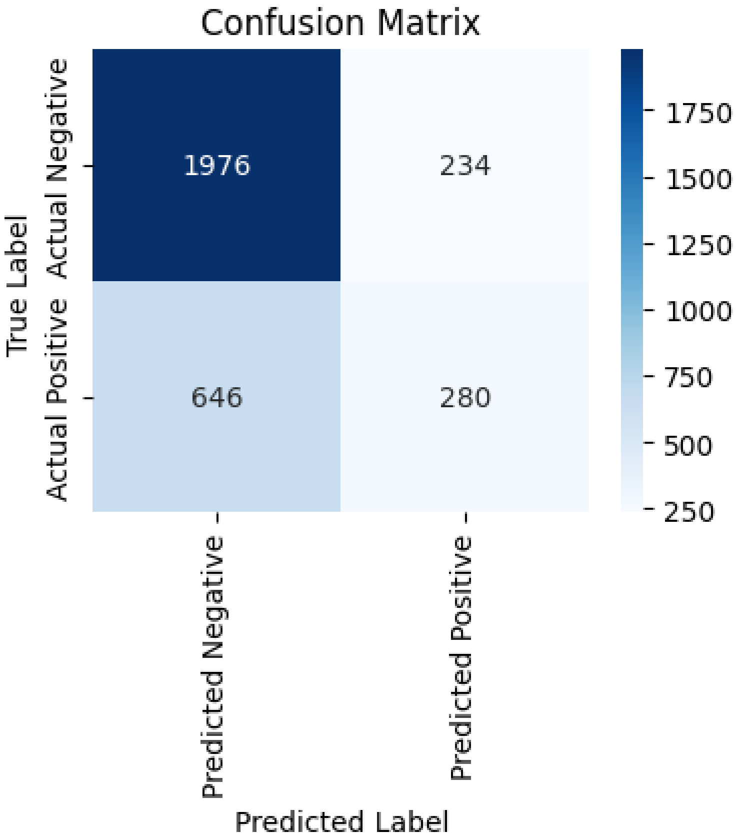


Figure: Confusion matrix of Middle-income countries

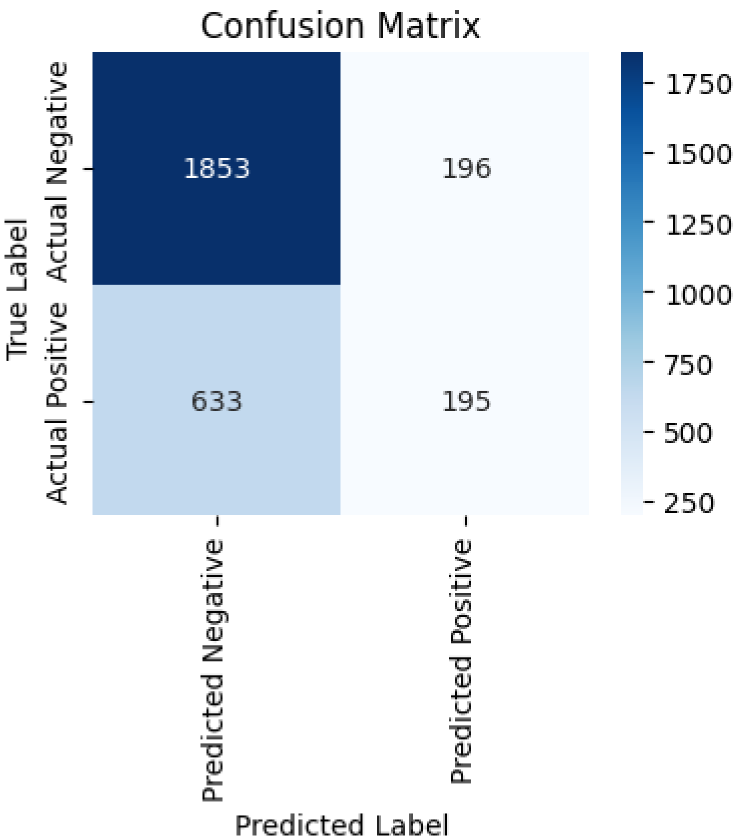


Figure: Confusion matrix of high-income countries