

Article

Startup Sustainability Forecasting with Artificial Intelligence

Nikolaos Takas ¹, Eleftherios Kouloumpris ^{2,*} , Konstantinos Moutsianas ³ , Georgios Liapis ² ,
Ioannis Vlahavas ²  and Dimitrios Kousenidis ⁴ 

¹ Thermi Investment Group, 570 01 Thessaloniki, Greece; president@thermokoitida.gr

² School of Informatics, Aristotle University of Thessaloniki, 541 24 Thessaloniki, Greece; gliapisa@csd.auth.gr (G.L.); vlahavas@csd.auth.gr (I.V.)

³ Division of Business Studies, American College of Thessaloniki, 555 35 Thessaloniki, Greece; moutsian@act.edu

⁴ School of Economics, Aristotle University of Thessaloniki, 541 24 Thessaloniki, Greece; dkous@econ.auth.gr

* Correspondence: elefthenk@csd.auth.gr; Tel.: +30-6949874656

Abstract: In recent years, we have witnessed a massive increase in the number of startups, which are also producing significant amounts of digital data. This poses a new challenge for expert analysts due to their limited attention spans and knowledge, also considering the low success rate of empirical startup evaluation. However, this new era also presents a great opportunity for the application of artificial intelligence (AI) towards intelligent startup investments. There are only a few works that have considered the potential of AI for startup recommendation, and they have not paid attention to the actual requirements of investors, also neglecting to investigate the desirability, feasibility, and value proposition of this venture. In this paper, we answer these questions by conducting a survey in collaboration with three major organizations of the Greek startup ecosystem. Furthermore, this paper also presents the design specifications for an AI-based decision support system for forecasting startup sustainability that is aligned with the requirements of expert analysts. Preliminary experiments with 44 Greek startups demonstrate Random Forest's strong ability to predict sustainability scores.

Keywords: startup; sustainability; forecasting; artificial intelligence; natural language processing



Citation: Takas, N.; Kouloumpris, E.; Moutsianas, K.; Liapis, G.; Vlahavas, I.; Kousenidis, D. Startup Sustainability Forecasting with Artificial Intelligence. *Appl. Sci.* **2024**, *14*, 8925. <https://doi.org/10.3390/app14198925>

Academic Editor: Pedro Couto

Received: 6 September 2024

Revised: 27 September 2024

Accepted: 1 October 2024

Published: 3 October 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent years, startup entrepreneurship has become a critical driver of technological, scientific, and business innovation. Evidently, many well-known companies, like Airbnb and Uber, started as small startups, and today are drastically changing the economy and modes of life through their disruptive business models, services, and products. According to Steve Blank [1], a startup is a temporary organization in search of an innovative business model that is feasible, sustainable, and scalable.

The evolution of startups is highly dependent on the supporting ecosystem, which includes both institutional and individual investors such as venture capital (VC) and angel investors, but also organizations like accelerators and incubators. Furthermore, the conditions of the macroeconomic environment, such as the levels of interest rates and consumer sentiment, also have a great influence on startups.

Due to the high risk associated with startup ventures, only a small percentage of startups succeed in discovering and implementing a successful startup venture. Specifically, only 1 in 10 startups manages to become profitable and sustainable for a period of more than five years [2]. Naturally, researchers and practitioners have raised the question of what are the leading indicators of a successful business. The consensus is that success factors pertain to the quality of the founding team [3], the timing of the venture, the desirability of the developed product or service, and aspects of the market that the latter is targeted towards [4].

Recent surveys suggest that the number of startups is increasing rapidly, which is in accordance with the global expansion of the supporting startup ecosystem. Furthermore,

startups are generating vast amounts of data through landing webpages, digitized presentations, business plans, and their activity on social media. At the same time, startups are becoming more specialized and complex in their business operations.

Furthermore, we must note that personality traits significantly influence entrepreneurship, which are depicted in many models, like the OCEAN (Big Five) Model, which is one of the most well-known models in the field [5]. Openness and low neuroticism enhance risk tolerance and innovativeness. Conscientiousness and extroversion drive proactivity and achievement orientation, while emotional stability aids resilience. Extroversion and agreeableness improve leadership skills, and adaptability is linked to openness. Self-efficacy, associated with extroversion and conscientiousness, empowers challenge tackling. Successful entrepreneurs often exhibit high openness, conscientiousness, and extroversion, and low neuroticism. Understanding these traits helps identify strengths and improve the forecast of business success [6].

The above facts will pose challenges to institutional investors, as the experts do not have the capability to keep track of the increasing number of startups and the vast amounts of data that they produce. Naturally, experts also have limited knowledge in science, technology, and business, which makes the evaluation of multiple specialized companies challenging. Due to these difficulties, the process of empirical startup evaluation is becoming more error-prone, costly, time-consuming, and ultimately, subjective. These challenges of the digital age are not limited to startup entrepreneurship but are universally imposed in multiple fields, including finance, healthcare, and technology.

However, the modern age also presents significant opportunities to modernize startup evaluation by leveraging these challenges. Artificial intelligence (AI) is revolutionizing business processes across every industry, and the timing is just right for its adoption by the startup entrepreneurship community. Unlike expert analysts, AI can analyze vast amounts of data rapidly without human error, in order to support decision-making through rational courses of action.

Startup sustainability and success represent an issue with multiple dimensions that extend far beyond economic growth. Undoubtedly, successful startup ecosystems act as drivers of economic prosperity. Additionally, they play an important role as innovation catalysts by disrupting traditional industries and as leaders in technology advancement. It is noteworthy that the success of startups enhances societal well-being and contributes to personal and community development. Solutions provided by startups on global issues such as climate change and healthcare highlight their contribution to global progress. Considering these dimensions, it is apparent that the prediction of startup sustainability is of utmost importance. The low startup success rate (<https://spdload.com/blog/startup-success-rate/>, accessed on 2 September 2024) by country (i.e., for the US, 20%; Germany, 25%; and the UK, 20%) and by industry (i.e., for the e-commerce industry, 20%, and for Fintech, 25%) for 2023 validates the necessity of startup success forecasting. Many studies have been conducted to shed light on the factors that affect the success of startups [3,4].

Typically, Cox proportional hazard functions are used to measure new-venture survival [7–9] as well as survival-time regressions [10]. In recent years, the rapid development of AI has offered many tools for startup success prediction. Żbikowski and Antosiuk [11] created a predictive model based on supervised learning to identify business success at an early stage. Sadatrasoul et al. [12] evaluate accelerated startups by applying six well-known classification algorithms. Dellermann et al. [13] utilize information regarding team size and entrepreneurial experience to assess startups' survival through a hybrid intelligence model. Prediction of the success probability for new companies in the US and China is conducted by Pan et al. [14] in order to classify the sample into successful and not-successful entities. Ünal and Ceasu [15] employ a set of different models, such as Random Forest, conditional reference tree, and recursive partitioning trees, with aim of predicting the success rates for future new ventures. It should be noted that the great majority of previous works focus on financial information in order to evaluate startup

sustainability, omitting important factors such as founder personality, team dynamics, and information that is available in natural language, such as business plans.

This literature review of the subject also reveals that research on AI-driven startup valuation methods has been limited for several reasons. First, startup data are difficult to access, compared to data of publicly traded companies. Second, contrary to large organizations, startup success depends on the personality traits of the founders and the dynamics of their teams, data on which are also hard to find. Third, close collaboration with multiple industry experts is essential for requirement elicitation and the design of a useful decision support system. To summarize, the lack of data and a multidisciplinary approach makes this a challenging issue with little previous research.

In this paper, we propose the design of a decision support system for investors and startups using AI and a serious game for personality assessment. We also establish the desirability and feasibility of such a system, by surveying four significant institutions of the Greek startup ecosystem. These institutions include the Thermi Investing Group VC (<https://thermi-group.com/en/>, accessed on 2 September 2024), the business accelerator egg—enter grow go of EuroBank (<https://www.theegg.gr/en/>, accessed on 2 September 2024), the Alexander Innovation Zone (<https://www.thessinnozone.gr/en/>, accessed on 2 September 2024) and WALK AUTH (<https://walk.auth.gr/>, accessed on 2 September 2024). A series of experiments with seven machine learning methods was conducted on data collected from 44 Greek startups. The results showcase the effectiveness of Random Forest (RF) in predicting sustainability scores, while the discovered rules offer insights into model explainability.

The collaboration with these institutions removes many barriers that limited previous work in this field. First, they provide a large amount of high-quality startup data. Second, they also provide access to the founders of past and existing startups, the psychometric analysis of whom is essential for the effective forecasting of startup success. The main contributions of this paper are as follows: (1) to provide insights regarding the actual constraints and requirements of startup organizations with regard to the feasibility and desirability of AI-driven systems, (2) the comprehensive design of an AI-driven system for startup sustainability forecasting that is in alignment with the requirements of startup organizations, and (3) a discussion addressing key challenges in startup sustainability forecasting, including optimal feature selection, the prevention of model overfitting, and the incorporation of human factors.

The remainder of the paper is structured as follows. Section 2 defines the survey that was conducted and presents the serious game for personality assessment. Section 3 presents the survey results, the proposed AI-driven system, the initial findings from experiments on Greek startups, and a discussion of the key challenges. Section 4 provides an additional discussion regarding the advantages and limitations of this work. Section 5 concludes the paper.

2. Materials and Methods

2.1. Startup Entrepreneurship Survey Design

The objectives of this survey were, first, to answer fundamental questions regarding the willingness of institutional investors to adopt AI solutions for startup evaluation; second, to understand which problems they are interested in solving using AI; third, to understand how the value added by the AI services is defined and how it can be monitored through key performance indicators (KPIs); and finally, to capture the requirements that will guide the design of the AI system. A set of 12 questions, presented in Table 1, were used to guide the survey discussion.

The surveys were conducted with four expert analysts from the four participating organizations. The results of the survey are presented in the next section, and were also the guiding factors for designing the AI system.

Table 1. Survey conducted with four leading organizations in the Greek startup ecosystem.

Question	Goal
Is your institution planning to integrate AI in decision-making?	Willingness to invest in AI
Are you currently assembling an internal AI development team?	Willingness to invest in AI
Are you currently seeking AI solutions from external software providers?	Willingness to invest in AI
What types of data do you track from previous startups?	Feasibility of implementation
Are the data of previous startups in digitized form?	Feasibility of implementation
Do you maintain an active network of founders?	Feasibility of implementation
What are your reasons for integrating AI technology?	Define value proposition
Which processes of startup evaluation could be automated?	Define value proposition
Which Key Performance Indicators (KPIs) can be used to monitor the benefits of AI?	Define value proposition
Which aspects of startups would you like to forecast with AI?	Requirements that guide implementation
What should be the horizon of the forecasting process?	Requirements that guide implementation
In which way should the results be presented?	Requirements that guide implementation

2.2. Psychometric Evaluation Through a Serious Game

As previously discussed, personality traits can play a crucial role in the assessment process of entrepreneurs. In today's organizational landscape, questionnaires are commonly used to evaluate individuals. However, these methodologies often suffer from being mundane and repetitive, and they are prone to social desirability bias, where respondents answer in a manner they believe is socially acceptable rather than truthful.

Given these limitations and the drawbacks associated with self-assessment questionnaires [16], innovative approaches are being explored. One such promising method is the use of virtual environments, such as games, to objectively assess behaviors and personality traits. These new methods offer a dynamic and engaging alternative to traditional questionnaires, potentially providing more accurate and authentic insights into an individual's personality.

A study found that game-based approaches were instrumental in improving the effectiveness of assessment practices, particularly within corporate training environments [17]. The study highlighted that integrating gaming technology into learning modules can also significantly enhanced the educational experience for employees. This integration led to increased motivation, engagement, and retention rates among participants. For instance, escape room games have gained popularity in corporate settings as tools for assessing both individual and team performance. These games also promote the development of critical teamwork and communication skills, which are essential for achieving collective success.

By leveraging the interactive nature of games, organizations can observe how individuals respond to various scenarios and challenges, thereby gaining a deeper understanding of their personality traits. This approach reduces the likelihood of social desirability bias and provides a more immersive and engaging assessment experience. As technology continues to evolve, the integration of these virtual environments into the assessment process could revolutionize how entrepreneurial potential is evaluated and forecasted.

Despite the potential benefits, existing methods for measuring team effectiveness and performance in real-life escape room (ER) games are limited. Current practices predominantly rely on post-game questionnaires administered to players, which only capture static reflections of their experiences and interactions. To address these limitations, we have developed a new gamified escape room environment designed to enhance the players' assessment process [18].

This innovative gamified ER environment fosters continuous interaction within a dynamic setting, allowing various traits and behaviors to emerge naturally over time. Unlike traditional static questionnaires, the ER game environment offers a comprehensive and nuanced understanding of personality traits. These traits are observed through players' decision-making, problem-solving abilities, and social interactions during the game. This approach provides a more accurate and holistic assessment of individuals, reflecting their true capabilities and behaviors in real-world scenarios.

We utilized MindEscape (v1.0) [19], an innovative digital escape room (ER) game that collects and analyzes each player's behavior, providing a novel, engaging, and effective assessment based on the OCEAN Five-Factor Model. In Figure 1, a screenshot of the game can be seen, from the Openness Escape Room.



Figure 1. MindEscape screenshot, Openness Escape Room.

3. Results

3.1. Survey Findings

The survey revealed that the organizations are unanimously interested in adopting AI solutions, with a preference for outsourcing the development to external software providers. Similarly, organizations tend to track and keep records of startup data. These data mostly include information about the founders (personal websites and CVs), business plans, pitch decks, and other media related to the startups.

Additionally, they record startup deliverables that indicate progress, including frequent revisions of the business model canvas and the value proposition canvas. Whereas a significant amount of data seem to be collected, the organizations do not take full advantage of them. Specifically, expert analysts are well aware of the value of the data that they are collecting, and well educated in the potential of AI.

The requested value proposition involves improved and informed decision support that can lead to faster processing times with fewer errors. The optimal strategy for startup evaluation was described as a hybrid between human and artificial intelligence, each providing a unique view of the investigated startups. Quantitative metrics should be used to evaluate these strategies, including the return on investment (ROI) of the investing decisions recommended by the AI system.

Several forecasting targets were discussed, including profitability, growth, general valuation scores, and future investments. However, the survey revealed that forecasting startup sustainability and longevity should have the most significant impact for decision-makers. Investors heavily question the integrity and longevity of startups before making decisions. Naturally, sustainability depends on multiple factors that must be considered. It is affected mostly by the cohesion of the founding team, the burn rate and runway of the company, and, of course, the potential for profitability and growth. Furthermore, the desirability of the developed product or service by the market and the feasibility of its implementation are critical factors of success, and therefore, also longevity.

3.2. Personality Assessment Results

After collaborating with WALK AUTH, we collected data from 15 new entrepreneurs participating in the second accelerator circle of the program. These entrepreneurs played the MindEscape game, which allowed us to gather detailed insights into their behavior and personality traits. The game sessions took place between 10 and 20 October 2023, providing a comprehensive dataset for analysis.

During this period, each participant's interactions and decisions within the game were meticulously recorded, offering a robust foundation for evaluating their performance through the OCEAN Five-Factor Model. This model assesses key personality dimensions: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism.

The results, showcased in Table 2, highlight individual and collective patterns among the participants. These insights not only reflect the effectiveness of MindEscape as an assessment tool, but also contribute to understanding the entrepreneurial potential and areas for development within this cohort. Such data are invaluable for tailoring future training and support to enhance the skills and success rates of aspiring entrepreneurs.

Table 2. Personality assessment results

Trait	Mean Personality	Characteristics
Openness	high	creativity and adaptability, crucial for innovation and identifying opportunities
Conscientiousness	medium	organization, reliability, and goal-setting, essential for effective business management
Extraversion	medium to high	networking, leadership, and communication
Agreeableness	high	teamwork, negotiation, and maintaining positive interactions with others
Neuroticism	low	emotional stability, good management of stress and ability to make rational decisions

3.3. Artificial Intelligence System for Startup Sustainability Forecasting

The aforementioned survey revealed that stakeholders are primarily interested in an AI system that can forecast the sustainability of pre-seed startups. Furthermore, according to our study, the feasible data sources can be organized into five categories: business plan, web activity, psychometric assessment, financials, and macroeconomics.

The business plan presents well-organized sections that can be extracted through NLP and presented in a structured format. The structured format includes records about the founders and their experience, the qualitative attributes of the product, the size of the potential market, and the competitors. This extraction process can utilize techniques such as Named Entity Recognition (NER), whereas the final output contains both tabular and graph structures.

The data relating to web activity have to be extracted through web crawlers. The web crawlers will scan the startup website, including the available activity on social media. Then, an NLP process that is similar to the one described for business plans can be used to extract more structured data.

The aforementioned sources, along with the personality assessment results through the serious game, form a repository of structured data that mainly describe the founders. However, they also describe other related entities such as their products and services, the market, and their competitors.

A second source of data includes the financials, which may be historical or future projections. These capture the historical or expected profitability and growth of the startup. Another resource regards external macroeconomic indicators, as these can also have an effect on the success rate of startups. Indicators such as interest rates, inflation, and consumer sentiment can impact funding options and the levels of expected revenue. These data are organized as timeseries.

The structured data along with the timeseries data form the final dataset that is used for training the machine learning algorithms. Supervised learning requires that we label a portion of the startups in this dataset. The first option is to use labels assigned by experts, such as empirical sustainability scores. Alternatively, self-supervised learning can be used to automate label extraction by forward-looking the status of previous startups. The exact nature of the label will depend on the task, as described below.

Assuming the existence of labeled datasets, we investigate the potential of four supervised learning paradigms for solving issues related to startup sustainability: regression, classification, ranking, and survival analysis.

Regression can be used to forecast a numerical feature related to a startup, also providing a corresponding confidence interval. Expert-assigned labels in the form of startup sustainability ratings can be used, or alternatively, a self-supervised label depending on future cashflow or runway (the amount of time (usually in months) that a startup has before it needs to secure additional funding). In either case, probabilistic regression can also indicate the worst-case-scenario rating or future cashflow, which should be considered for a potential investment.

Classification is useful for predicting discrete states in which the startup will be or events that will happen. It is more relevant for expert-assigned labels, which may indicate events such as whether the startup will receive funding from a VC or an angel investor. Furthermore, a classification model outputs the probability that a particular event will happen, which can be useful in decision-making.

Ranking allows for the creation of a sortlist of startups according to their expected sustainability. However, this requires that the experts assign a relevance score to each startup through a rating system. Compared to the previous methods, which make absolute predictions, ranking offers the significant benefit of allowing relative comparisons between startups.

Finally, survival analysis can reveal the expected lifetime of the startup. This paradigm works exclusively with self-supervised learning, and we presume that it is the most relevant for sustainability forecasting. Furthermore, it utilizes both censored and uncensored data, which correspond to startups that are still operational and startups that have been shutdown.

Figure 2 illustrates the entire AI-driven system that integrates the aforementioned modules. First, unstructured data regarding business plans, social media activity, and founder personality are processed through NLP and psychometric analysis, to then be stored as structured information. Second, the structured information is combined with financial information to form the final dataset. Third, specialized timeseries and graph models are used to solve various problems that are relevant to startup sustainability forecasting.

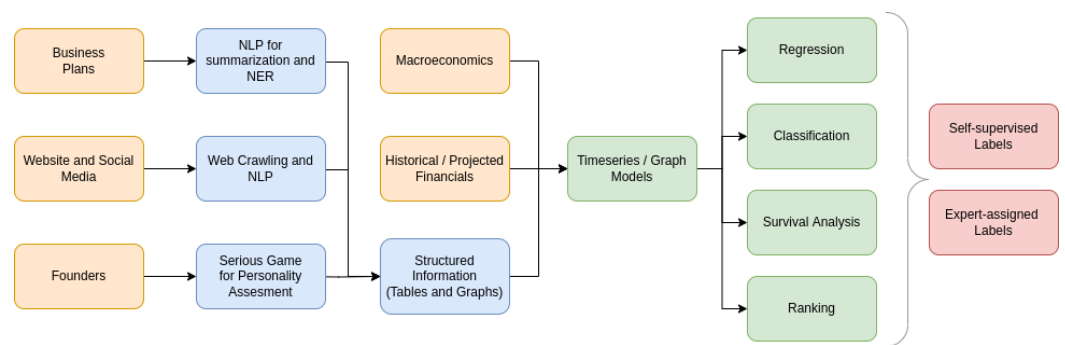


Figure 2. Architecture of the complete AI-driven startup sustainability forecasting system. The architecture is guided by findings regarding the available data sources and actual needs of professional investors.

3.4. Sustainability Forecasting for Greek Startups

Through our collaboration with the aforementioned institutions, we collected an initial dataset of 44 startups. In the following paragraphs, we present several preliminary findings regarding the potential of machine learning methods applied to sustainability forecasting.

The collected dataset contains features related to the progress of the 44 startups during an acceleration program. For each startup, the current income levels and whether they are currently profitable are recorded. The dataset also includes information on their technological readiness level (TRL), the current state of their MVP, and whether they exist in legal form. In addition, it records the number of founders and their participation in exhibitions, as well as evaluations of problem–founder fit and product–market fit. Finally, it details the quality of a comprehensive set of planning and presentation materials, including

the market research report, business model canvas, value proposition canvas, business plan, financial plan, executive summary, website, social media, and pitch deck.

The expert executives of the acceleration program studied these startups in great detail and assigned a discrete sustainability score ranging from 1 to 3; scores of 1, 2, and 3 are interpreted as 'good,' 'average,' and 'bad,' respectively. The input features and the expert-assigned labels are considered in the upcoming supervised learning experiments.

We also conducted experiments with seven ML algorithms for regression, using the RapidMiner (v2024.0.0) implementation. The methods are Linear Regression (LR), artificial neural network (ANN), decision tree (DT), k-Nearest Neighbors (kNN), Support Vector Machine (SVM), Gradient Boosting (GB), and Random Forest (RF) [20]. The models were configured with the recommended hyperparameters, and their performance was evaluated using leave-one-out cross-validation (LOOCV). During the validation process, the Root-Mean-Square Error (RMSE) and the standard deviation of the error (StdDev) were estimated. RMSE indicates the expected error of the model, whereas StdDev reflects the stability of the model. These metrics are provided in Equations (1)–(2), in which y_i and \hat{y}_i denote the actual and predicted scores for the i th startup, respectively.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (1)$$

$$StdDev = \sqrt{\frac{1}{n} \sum_{i=1}^n \left((y_i - \hat{y}_i)^2 - RMSE^2 \right)^2} \quad (2)$$

Table 3 presents the performance results for the aforementioned methods. Using leave-one-out cross-validation (LOOCV), we evaluate their ability to predict the sustainability scores assigned by experts. While every model displays an excellent ability to separate 'bad' from 'good' cases, which differ in score by 2, we observe several performance differences.

The methods LR and ANN are less accurate than the other models, with an RMSE approximately equal to 0.665 and a StdDev of about 0.5. DT improves the RMSE, which reaches 0.578, but increases the variance to 0.518. The methods kNN and SVM outperform DT, achieving RMSEs of 0.534 and 0.505, respectively. They also significantly decrease the StdDev, to 0.392 for kNN and 0.378 for SVM. Finally, GB and RF are the top-performing methods, with very similar RMSE scores (0.489 vs. 0.487). However, GB has a significantly higher StdDev than RF (0.487 vs. 0.361). Therefore, these results indicate that RF is both the best-performing and most stable method for startup sustainability forecasting.

Table 3. Leave-one-out cross-validation results for startup sustainability forecasting using 44 Greek startups. The best-performing method and metric scores are highlighted in bold.

Model	Root-Mean-Square Error	Standard Deviation
Random Forest	0.487	0.361
Gradient Boosting	0.489	0.419
Support Vector Machine	0.505	0.378
k-Nearest Neighbors	0.534	0.392
Decision tree	0.578	0.517
Artificial neural network	0.664	0.493
Linear Regression	0.668	0.504

To provide further insight regarding the workings of the model, a visualization of a good-performing decision tree is shown in Figure 3. By observing the rules, we infer that the tree first considers whether the startup is already profitable. The highest score is achieved for companies that are profitable, have presented a high-quality MVP, and have an annual income that exceeds EUR 57,000. Different branches of the tree also consider whether the company has a legal form, the product–market fit, the number of founders,

and the quality of market research, social network activity, and the pitch deck. The lowest score is given to startups that do not have a legal form and have not presented an MVP.

The aforementioned tree analysis showcases how machine learning models should not necessarily be considered black boxes, but instead, can present transparent reasoning steps. Model explainability is a critical factor for gaining the trust of expert analysts, improving the chances of this technology being adopted by startup organizations. While this section provides evidence for the potential of AI in startup sustainability forecasting, the following section addresses several key challenges that need to be considered in practice.

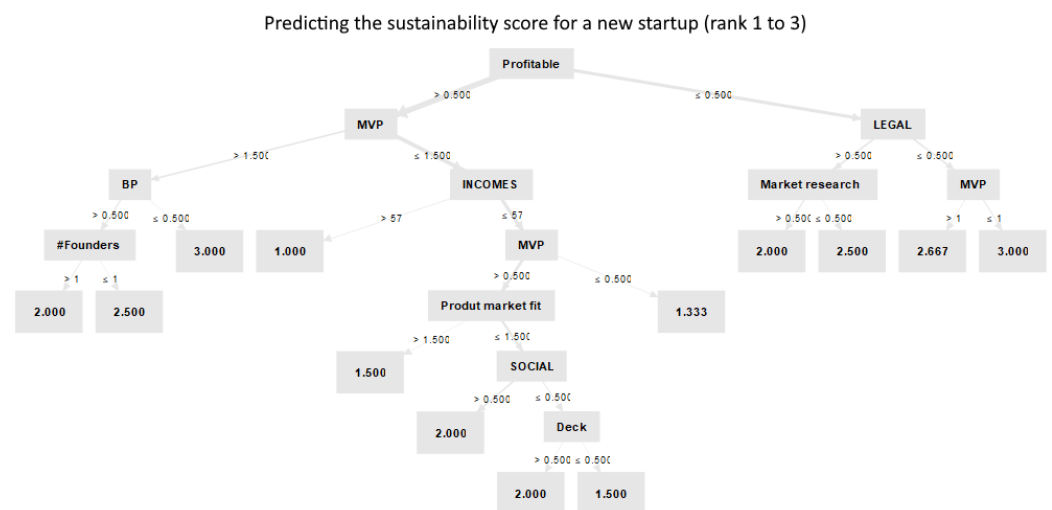


Figure 3. A decision tree visualization that reveals the discovered rules for sustainability forecasting.

3.5. Addressing Key Challenges in Startup Sustainability Forecasting

In the application of AI to forecasting startup sustainability, three critical challenges must be considered: handling correlated features, preventing overfitting, and accounting for the impact of human factors.

Correlated features can lead to the misrepresentation of feature importance and reduce the effectiveness of learning. There are generally two ways to address these issues: feature transformation and feature selection. In feature transformation, the current features are transformed into a new representation where they are not correlated with each other. The primary technique for achieving this is Principal Component Analysis (PCA). This method redefines the feature space into principal components that are orthogonal to each other and capture the largest portions of variance in the data.

While PCA effectively eliminates feature correlations, the principal components often lack a natural interpretation—a significant drawback for model explainability. Given that we are considering models that must support critical, high-level decision-making in investment contexts, we instead opt for feature selection methods. Feature selection techniques choose a subset of the original feature space, preserving the interpretability of the original features.

For this problem, we propose a two-step feature selection process. In the first step, the process ranks the available features according to their predictive power. This can be achieved using the Mutual Information Criterion (MIC) [21], which reveals the degree of statistical dependence between the feature and the target, i.e., the reduction in target uncertainty from knowing the feature. Tree-based feature importance and Boruta are also well-regarded methods for feature selection, but they do not perform well when there are many correlated features.

In the second step, the ranked feature list is scanned sequentially, including features in the subset as long as (a) they are not statistically significantly correlated with the previously selected features and (b) improve the validation performance of the model (step-forward feature selection).

While effective feature selection significantly reduces overfitting, additional methods should be employed to ensure robust model performance. Leave-one-out cross-validation (LOOCV) provides a reliable estimate of model performance; however, the final generalization error should be evaluated using an independent test set. In applications such as startup sustainability forecasting, where data are relatively scarce, fully grown decision trees—being high-variance models—are especially prone to overfitting. Unless the trees are pruned or their depth is restricted, they risk capturing noisy patterns because of their tendency to perfectly fit the training data.

A more effective approach is to use ensemble methods such as Random Forest (RF), which mitigate overfitting by combining fully grown trees without the need for pruning or depth constraints. RF reduces variance through averaging tree predictions and decorrelating individual trees by randomly selecting subsets of features for each split. This method has demonstrated strong performance across various problems. Additionally, hybrid ensembles that combine models with different inductive biases can be trained independently, further reducing overfitting by leveraging their complementary strengths.

Nevertheless, the most effective direct approach to combat overfitting is to collect more data. Increasing the dataset size also enables the use of deep learning approaches, which generally require more examples than tree-based models. For models based on artificial neural networks (ANNs) and gradient descent, techniques such as early stopping, L1/L2 regularization, and dropout should be employed to prevent the network weights from growing excessively.

Additionally, Bayesian neural networks [22] offer increased robustness to overfitting by incorporating uncertainty into parameter estimation. By learning the statistical distribution of neural connection weights instead of relying on point estimates, they prevent the network from converging to an overfitted configuration. For models that incorporate NLP, leveraging large language models pre-trained on extensive corpora significantly reduces the likelihood of overfitting to specific language patterns when fine-tuning on smaller datasets.

Beyond traditional data sources, it is widely recognized that entrepreneurial success strongly depends on qualitative human factors that are challenging to quantify. This is especially true for startups in the seed stage, where recent research highlights that human elements—particularly the personality traits and behavioral patterns of founders—are among the most critical determinants of team cohesion and long-term sustainability.

This motivated us to incorporate a serious game approach, described in Section 2.2, which aims to capture human-related factors such as personality and behavior for the first time in the context of startup sustainability. Additionally, analyzing sentiments expressed on social media can provide valuable insights into customer opinions about products. Early opinions extracted through sentiment analysis can serve as leading indicators of product success, while traditional revenue data often act as a lagging indicator, reflecting a delayed signal of product desirability. Opinion mining should not be limited to only assessing positive or negative sentiment toward a product; through emotion analysis, we can gain deeper insights into the emotional states of customers, including happiness, sadness, anger, fear, and disgust. These approaches allow us to quantify and analyze human-related factors that were previously challenging to measure and understand.

4. Discussion

As we witness the transformation of almost every industry through AI, it is certain that this new technology has a major influence on startup investments. Expert analysts will utilize the predictive power of AI to make better decisions and achieve better returns on investment (ROIs).

The advantage of this work is that it relies on feedback from important organizations of the startup entrepreneurship ecosystem. The proposed AI architecture is well aligned with actual feasibility constraints and limitations of the real environment, and is designed to provide actual value for stakeholders. This section further discusses key challenges that need to be addressed for developing a high-performing system. Experiments conducted

on 44 startups revealed that machine learning, particularly through Random Forest, can achieve high performance in forecasting startup sustainability scores. A limitation of this work is the focus on Greek organizations and startups, which are, however, very representative of the European startup environment.

In future work, we plan to train our AI system with a significant volume of startup data from different countries. This will allow for a thorough evaluation of the system in different conditions, and further prove the added value for startup organizations. The participation of the four organizations was essential for gaining access to an adequate volume of anonymized startup data. They also provided a letter of intent to integrate the final system in their decision-making process for a pilot evaluation.

Another direction for future work is to compare the requirements and restrictions of the European startup ecosystem with the investing environment of other continents. For instance, the American startup ecosystem presents significant differences in terms of startup culture, decision-making processes, regulations, and available data sources. Therefore, it would be necessary to perform more interviews to customize the system for different environments.

5. Conclusions

Startup evaluation is becoming increasingly difficult due to the growing number of companies and the volume of digital data they generate. This difficulty is attributed to multiple factors, including the increasing complexity of new business plans, which require significant expertise for their evaluation. This presents a new challenge for professional investors and analysts, as they have to estimate the feasibility and sustainability of thousands of startup ventures.

AI presents an excellent opportunity to address this issue. Through intelligent decision support tools, expert analysts can improve their efficiency and make fewer errors in their investment decisions. However, until recently, there has been very little research regarding the use of AI for startup sustainability forecasting. Similarly, there are no studies investigating the feasibility and defining the value proposition of AI for startup evaluation. This can be explained by researchers' limited access to startup data, as these are mostly private and collected by accelerators, incubators, and VCs.

In this paper, we conducted a survey that provides insights regarding the desirability, feasibility, value proposition, and specification of AI-driven startup evaluation. The survey was conducted in collaboration with four major startup organizations in Greece. The findings reveal that startup organizations (VCs, accelerators, and incubators) are well aware of the benefits of AI and are currently taking action to integrate this new technology into their decision-making processes. According to the conducted survey, the most desirable feature for an AI-driven system was forecasting startup sustainability. Sustainability is a complicated attribute that depends on multiple factors, including team cohesion, financial projections, burn rate, market conditions, and the developed product/service.

Based on the survey findings, this paper also presents a comprehensive AI-driven architecture for startup sustainability forecasting. Therefore, this architecture is in alignment with the actual feasibility and desirability constraints dictated by the survey. It considers multiple sources of data, including business plans, web activity data, psychometric assessments, company financials, and macroeconomic indicators. Due to the heterogeneous nature of the data, the system integrates multiple machine learning methods, such as NLP for document understanding and timeseries models to process financial projections. Several key challenges have also been addressed, including feature selection, overfitting, and the incorporation of human factors.

This paper also presents the experimental results of seven machine learning algorithms on 44 Greek startups. The purpose of these experiments is to predict startup sustainability scores annotated by expert analysts. These results highlight the high accuracy of Random Forest in predicting sustainability scores, effectively distinguishing good from bad cases.

In addition, the decision tree structure is analyzed to provide insights into the discovered rules, which are highly explainable.

In future work, we plan to train and evaluate our startup sustainability forecasting system with multiple data and in collaboration with the aforementioned organizations. Another direction for future work is to investigate how the requirements and constraints of an AI-driven sustainability forecasting system would differ in the US startup ecosystem.

Author Contributions: Conceptualization, N.T. and E.K.; methodology, E.K. and G.L.; software, G.L.; validation, N.T., K.M., I.V. and D.K.; investigation, E.K. and K.M.; resources, N.T.; data curation, G.L.; writing—original draft preparation, E.K., G.L. and K.M.; writing—review and editing, N.T., K.M., I.V., D.K., E.K. and G.L.; visualization, E.K. and G.L.; supervision, I.V. and D.K.; project administration, E.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Ethical review and approval were waived for this study due to anonymized data collection.

Informed Consent Statement: Informed consent was obtained from all subjects involved in this study.

Data Availability Statement: Restrictions apply to the availability of these data. The data were obtained from Thermi Investment Group and are available from the authors with the permission of Thermi Investment Group.

Acknowledgments: The authors would like to thank B. Chrysovergis from WALK AUTH; R. Bachtalia, E. Georgakopoulos and A. Stamopoulos from the business accelerator egg—enter grow go of EuroBank; and P. Ketikidis from the Alexander Innovation Zone for their participation in the interviews and for providing consent to publish our findings.

Conflicts of Interest: Author Nikolaos Takas is the owner of the company Thermi Group. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial intelligence
ML	Machine learning
NLP	Natural language processing
VC	Venture capital

References

1. Spender, J.C.; Corvello, V.; Grimaldi, M.; Rippa, P. Startups and open innovation: A review of the literature. *Eur. J. Innov. Manag.* **2017**, *20*, 4–30. [[CrossRef](#)]
2. Krishna, A.; Agrawal, A.; Choudhary, A. Predicting the Outcome of Startups: Less Failure, More Success. In Proceedings of the 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW), Barcelona, Spain, 12–15 December 2016; pp. 798–805. [[CrossRef](#)]
3. Diakanastasi, E.; Karagiannaki, A.; Pramadari, K. Entrepreneurial Team Dynamics and New Venture Creation Process: An Exploratory Study Within a Start-Up Incubator. *Sage Open* **2018**, *8*, 2158244018781446. [[CrossRef](#)]
4. Eliakis, S.; Kotsopoulos, D.; Karagiannaki, A.; Pramadari, K. Survival and Growth in Innovative Technology Entrepreneurship: A Mixed-Methods Investigation. *Adm. Sci.* **2020**, *10*, 39. [[CrossRef](#)]
5. Jang, K.L.; Livesley, W.J.; Vernon, P.A. Heritability of the big five personality dimensions and their facets: a twin study. *J. Personal.* **1996**, *64*, 577–591. [[CrossRef](#)] [[PubMed](#)]
6. Rauch, A. Let's Put the Person Back into Entrepreneurship Research: A Meta-Analysis on the Relationship Between Business Owners' Personality Traits, Business Creation, and Success. *Eur. J. Work. Organ. Psychol.* **2007**, *16*, 353–385. [[CrossRef](#)]
7. Cader, H.; Leatherman, J. Small business survival and sample selection bias. *Small Bus. Econ.* **2011**, *37*, 155–165. [[CrossRef](#)]
8. Delmar, F. Does Experience Matter? The Effect of Founding Team Experience on the Survival and Sales of Newly Founded Ventures. *Strateg. Organ.* **2006**, *4*, 215–247. [[CrossRef](#)]
9. Delmar, F.; Scott, S. Legitimizing First: Organizing Activities and the Survival of New Ventures. *J. Bus. Ventur.* **2004**, *19*, 385–410. [[CrossRef](#)]

10. Bosma, N.; Praag, M.; Thurik, R.; Wit, G. The Value of Human and Social Capital Investments for the Business Performance of Start-Ups. *Small Bus. Econ.* **2004**, *23*, 227–236. [CrossRef]
11. Żbikowski, K.; Antosiuk, P. A machine learning, bias-free approach for predicting business success using Crunchbase data. *Inf. Process. Manag.* **2021**, *58*, 102555. [CrossRef]
12. Sadatrasoul, S.M.; Ebadati, O.; Saedi, R. A Hybrid Business Success Versus Failure Classification Prediction Model: A Case of Iranian Accelerated Start-ups. *J. Data Min.* **2020**, *8*, 279–287. [CrossRef]
13. Dellermann, D.; Lipusch, N.; Ebel, P.; Popp, K.M.; Leimeister, J.M. Finding the unicorn: Predicting early stage startup success through a hybrid intelligence method. *arXiv* **2021**, arXiv:2105.03360. [CrossRef]
14. Pan, C.; Gao, Y.; Luo, Y. Machine learning prediction of companies' business success. *CS229: Machine Learning, Fall 2018*; Stanford University: Stanford, CA, USA, 2018.
15. Ūnal, C.; Ceasu, I. A Machine Learning Approach towards Startup Success Prediction. IRTG 1792 Discussion Papers 2019-022, Humboldt University of Berlin, International Research Training Group 1792 "High Dimensional Nonstationary Time Series". 2019. Available online: <https://www.econstor.eu/bitstream/10419/230798/1/irtg1792dp2019-022.pdf> (accessed on 2 September 2024).
16. Byworth, C. Measuring Personality Constructs: The Advantages and Disadvantages of Self-Reports, Informant Reports and Behavioural Assessments. *Enquire* **2008**, *1*, 75–94.
17. Dicheva, D.; Dichev, C.; Agre, G.; Angelova, G. Gamification in Education: A Systematic Mapping Study. *J. Educ. Technol. Soc.* **2015**, *18*, 75–88.
18. Liapis, G.; Zacharia, K.; Rrasa, K.; Liapi, A.; Vlahavas, I. Modelling Core Personality Traits Behaviours in a Gamified Escape Room Environment. In Proceedings of the 16th European Conference on Games Based Learning, Lisbon, Portugal, 6–7 October 2022; Volume 16; p. 731. [CrossRef]
19. Liapis, G.; Zacharia, K.; Rrasa, K.; Vlahavas, I. Serious Escape Room Game for Personality Assessment. In Proceedings of the 12th International Conference on Games and Learning Alliance, Dublin, Ireland, 29 November–1 December 2023; pp. 420–425. [CrossRef]
20. Hastie, T.; Tibshirani, R.; Friedman, J.H.; Friedman, J.H. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer: New York, NY, USA, 2009; Volume 2.
21. Qian, W.; Shu, W. Mutual information criterion for feature selection from incomplete data. *Neurocomputing* **2015**, *168*, 210–220. [CrossRef]
22. Jospin, L.V.; Laga, H.; Boussaid, F.; Buntine, W.; Bennamoun, M. Hands-on Bayesian neural networks—A tutorial for deep learning users. *IEEE Comput. Intell. Mag.* **2022**, *17*, 29–48. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.