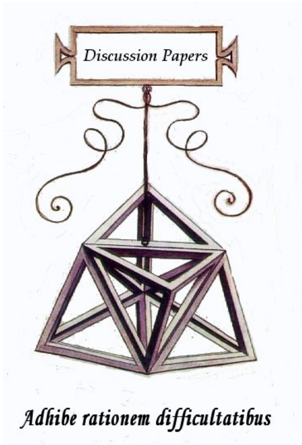
Discussion papers
E-papers of the Department of Economics e Management – University di Pisa



Caterina Giannetti, Maria Saveria Mavillonio

Crowdfunding Success: Human Insights vs Algorithmic Textual Extraction

Discussion paper n. 315

2024

Discussion paper n. 315, presented: **November 2024**

Authors' address/Indirizzo degli autori:

Caterina Giannetti — University of Pisa - Department of Economics and Management, Via Cosimo Ridolfi 10, 56124 Pisa – Italy. E-mail: caterina.giannetti@unipi.it

Maria Saveria Mavillonio — Via Cosimo Ridolfi 10. E-mail: mariasaveria.mavillonio@phd.unipi.it

Please cite as:/Si prega di citare come:

Caterina Giannetti, Maria Saveria Mavillonio (2024), "Crowdfunding Success: Human Insights vs Algorithmic Textual Extraction", Discussion Papers, Department of Economics and Management – University of Pisa, n. 315 (http://www.ec.unipi.it/ricerca/discussion-papers).

Discussion Papers Series contact: pietro.battiston@unipi.it

Discussion paper

n. 315

Caterina Giannetti, Maria Saveria Mavillonio

Crowdfunding Success: Human Insights vs Algorithmic Textual Extraction

Abstract

Using a unique dataset of equity offerings from crowdfunding platforms, we explore the synergy between human insights and algorithmic analysis in evaluating campaign success through business plan assessments. Human evaluators (students) used a predefined grid to assess each proposal in a Business Plan competition. We then developed a classifier with advanced textual representations and compared prediction errors between human evaluators, a machine learning model, and their combination. Our goal is to identify the drivers of discrepancies in their evaluations. While AI models outperform humans in overall accuracy, human evaluations offer valuable insights, especially in areas requiring subtle judgment. Combining human and AI predictions leads to improved performance, highlighting the complementary strengths of human intuition and AI's computational power.

Keywords: Crowdfunding, Natural Language Processing, Human Evaluation

JEL CLassification: C45, C53, G2

Crowdfunding Success: Human Insights vs Algorithmic Textual Extraction

Caterina GIANNETTI Maria Saveria MAVILLONIO*

November 7, 2024

Abstract

Using a unique dataset of equity offerings from crowdfunding platforms, we explore the synergy between human insights and algorithmic analysis in evaluating campaign success through business plan assessments. Human evaluators (students) used a predefined grid to assess each proposal in a Business Plan competition. We then developed a classifier with advanced textual representations and compared prediction errors between human evaluators, a machine learning model, and their combination. Our goal is to identify the drivers of discrepancies in their evaluations. While AI models outperform humans in overall accuracy, human evaluations offer valuable insights, especially in areas requiring subtle judgment. Combining human and AI predictions leads to improved performance, highlighting the complementary strengths of human intuition and AI's computational power.

Keywords: Crowdfunding, Natural Language Processing, Human Evaluation

JEL Classification: C45, C53, G2, G23, L26

*Department of Economics and Management, University of Pisa, Italy.
Corresponding author: Email: mariasaveria.mavillonio@phd.unipi.it
The research acknowledges funding support from the PRIN grant no. 20177FX2A7, provided by the Italian Ministry of University and Research.

1 Introduction

Small and illiquid firms struggle to secure funding for their activities as they face stronger information asymmetry, e.g. Berger and Udell (1998); Hall and Lerner (2010); Cole and Sokolyk (2016). Indeed, crowdfunding platforms have emerged as an alternative source for small and innovative projects. Crowdfunding involves raising small amounts of money from a large number of people, typically via the Internet. These platforms allow entrepreneurs to present their ideas directly to potential backers, bypassing traditional financial intermediaries and mitigating some of the information asymmetry challenges they face (Signori and Vismara, 2018).

One way to convey information to the public on these platforms and attract investors is through the writing of a Business Plan. A business plan is a formal document that outlines the goals, strategies, target market, and financial projections of a business (Guarino and Mariani, 2021). It provides detailed information on the company's mission, products or services, market analysis, marketing strategies, operational plan, and financial forecast. Despite their importance, business plans are not fully exploited to their potential in guiding entrepreneurs on how to write them effectively. In particular, Kaminski and Hopp (2020) emphasize the importance of understanding crowdfunding from an investor's perspective by applying machine learning techniques based on text, speech, and video metadata to predict the outcomes of crowdfunding startup pitches. Moreover, Mason and Stark (2004) point out that different types of funders—bankers, venture capitalists, and business angels—evaluate business plans from distinct perspectives, necessitating customized plans for different audiences. Additionally, the literature reveals mixed evidence on the effectiveness of business plan training programs (McKenzie, 2017; Clingingsmith, Drover, and Shane, 2023).

Our aim in this paper is to understand how the presentation and contents of a business plan can be used to determine the success of a crowdfunding campaign, both by humans and algorithms. We first assess the discrepancy between human evaluation and various algorithms, which incorporate textual information, in predicting campaign success. We then explore differences in judgment between humans and machines by assessing specific features of the business plan, such as readability, clarity, editing quality, feasibility, completeness, and attractiveness.

Recently, Wang et al. (2024) showed how AI analysts outperform human analysts in stock forecasts, especially in transparent and data-rich environments. However, human analysts retain an advantage in interpreting critical information that requires institutional knowledge, such as intangible assets. They also found that text information contributes about 10% to the AI model's price prediction.

Similarly, Hemmer et al. (2022) find that when experts have access to unique human contextual information not available to AI during training—due to technical or economic reasons—the teams achieve a lower mean absolute error (MAE) compared to teams relying solely on AI or without contextual insights. In a related study, McKenzie and Sansone (2019) examined a Nigerian business competition and found that human evaluators were no better than a machine learning algorithm in predicting the success of companies by analyzing their business plans. However, they did not explore when or why one approach might outperform the other.

Kim, Muhn, and Nikolaev (2024) highlight the ability of large language models like GPT-4 to outperform human analysts in structured financial forecasting, particularly in data-rich settings, while Abolghasemi, Ganbold, and Rotaru (2024) demonstrate that LLMs, though adept at processing large datasets, do not always surpass human forecasters, especially in complex scenarios like promotional sales where human judgment still plays a critical role.

In general, empirical evidence shows that humans and machine algorithms have complementary strengths in tasks like face recognition, sports prediction, diagnostic imaging, and classifying medical images (Steyvers et al., 2022). This complementarity arises from different information sources and processing strategies. However, it is influenced by the correlation between human and machine predictions. High correlation limits the accuracy benefits of combined classifiers. Effective AI advice should be both accurate and independent of human judgment (Steyvers et al., 2022).

In this paper, we aim to uncover the synergy between AI computational power and human understanding of soft information to identify which characteristics a business plan needs to lead to campaign success. Are there any differences in human judgments relating to the writing style of the BP? Are certain aspects of the business plan more influential in human evaluations compared to algorithmic assessments? By conducting a Business Plan competition with students evaluating plans and developing a machine learning classifier, we analyze the respective prediction errors and determine how textual representations and business plan characteristics affect the success predictions. This comprehensive compari-

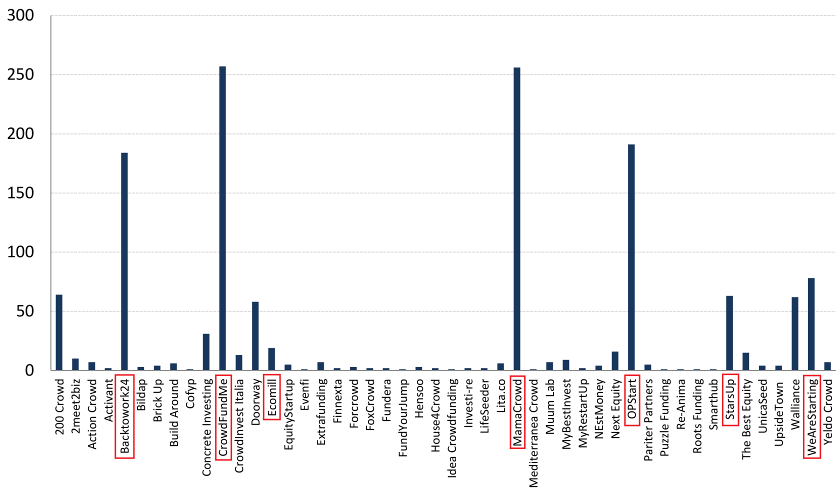


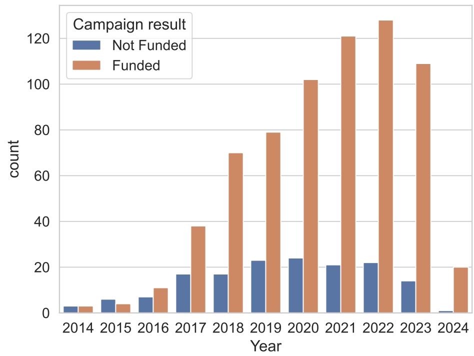
Figure 1: Number of campaigns for each platform taken from Politecnico di Milano – Dipartimento di Ingegneria Gestionale (2024). We highlighted in red the platform we have chosen.

son highlights the complementary strengths of human judgment and AI, offering insights into improving the evaluation process for crowdfunding campaigns.

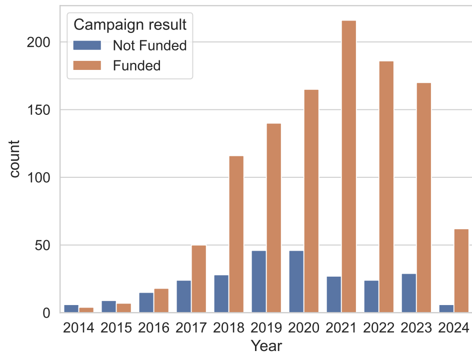
2 Data Description

We collected business plans and campaign data by web scraping publicly available documents from Italian crowdfunding platforms. This included campaign details such as equity offered, pre-money valuation, minimum funding goals, share types (e.g., Type A/B/C/D), and business plans (BPs). To accurately link firm-specific data, we also retrieved the VAT number for each company. Currently, 33 portals are authorized to publish campaigns (Figure 1). From these, we selected 7 platforms: *MamaCrowd* and *Crowdfundme* as large platforms, *BacktoWork24* and *OpStart* as second-tier large platforms, two medium-sized platforms (*StarSup* and *WeAreStarting*), and one small platform (*Ecomill*) from the inception of crowdfunding in 2014 to the end of July 2024. Due to the harmonization of European Crowdfunding Service Providers(ESCP) most platforms had to suspend their offerings at the end of 2023, leading to a decline in 2024. Figure 2 illustrates the trend in the number of online crowdfunding campaigns from 2014 to 2024, comparing the overall landscape in Italy with the specific data from our dataset.

The figure also shows how many of these campaigns were successful during this time frame. In total, there were 1,427 campaigns across Italy, while our dataset includes 874 campaigns. Notably, the success rate has consistently been above 80%, with a steady increase year by year. Furthermore, the trend within our dataset closely aligns with the general trend observed across the Italian crowdfunding market, reflecting similar patterns in the growth and success of campaigns over the years.



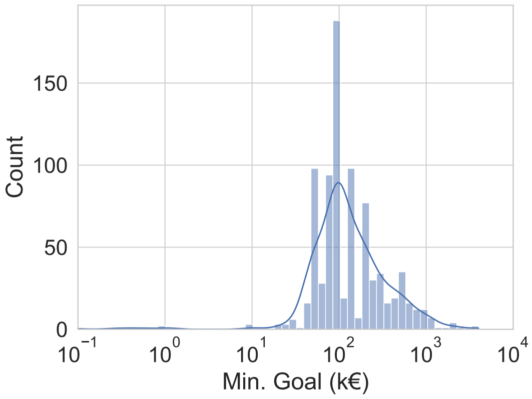
(a) Temporal flow of the equity crowdfunding in our dataset



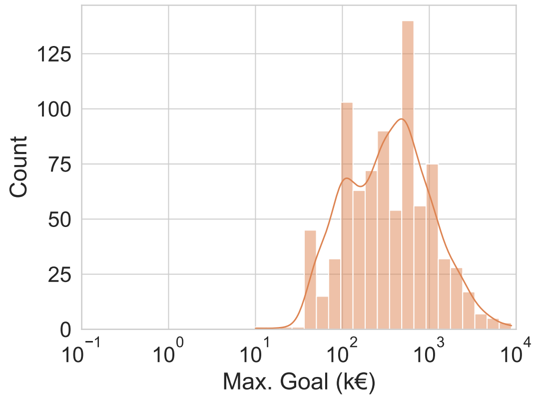
(b) Temporal flow of equity crowdfunding campaigns in Italy Politecnico di Milano – Dipartimento di Ingegneria Gestionale (2024)

Figure 2: Comparison of equity crowdfunding trends.

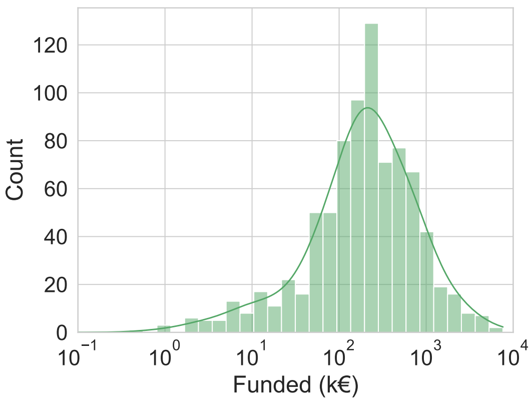
Figure 3 provides a descriptive analysis of key variables in equity crowdfunding campaigns, highlighting significant variability in funding targets, amounts raised, investor participation, and equity offered. The average minimum funding goal is €210k, with campaigns ranging from as low as €0.10k to €4M, while the average maximum goal is €618k, reaching up to €8.82M. The mean amount raised is €409.8k, with substantial variation across campaigns, as indicated by a maximum of €7.6M and a large standard deviation of €687.5k. These results demonstrate the diversity in equity crowdfunding campaigns, both in terms of financial scale and investor engagement. The figure 3d specifically compares the actual amount raised with the minimum and maximum goals to assess campaign performance. When the funded amount is near the minimum goal, the campaign has just met its essential requirements. Conversely, if the funded amount approaches the maximum goal, the campaign can be considered highly successful. On average, the mean of the actual capital raised exceeds the mean of the minimum goal,



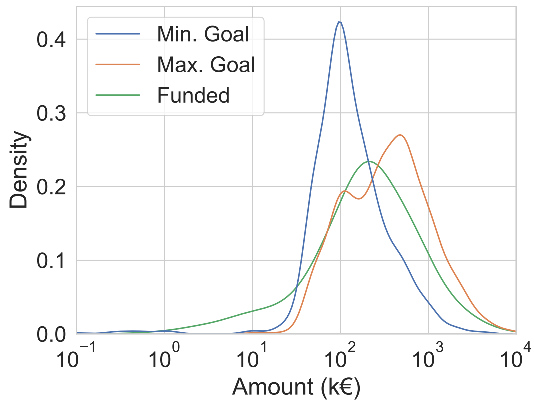
(a) Distribution of Minimum Goal



(b) Distribution of Max Goal



(c) Distribution of Actual Capital Raised



(d) Comparison across all metrics

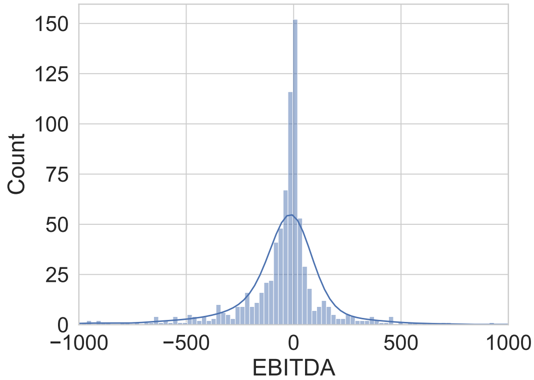
Figure 3: Distribution of Goal, Maximum Goal, and Capital Raised (logarithmic scale on the x-axis).

indicating that equity campaigns generally achieve a level of overfunding close to the maximum goal set by the firm.

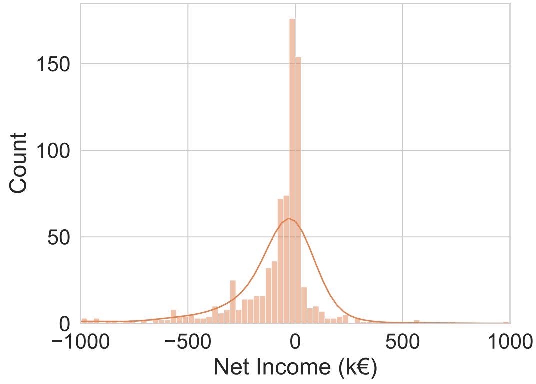
We then augmented our dataset with relevant financial variables sourced from AIDA. Specifically, we have integrated geographic data, financial performance indicators (such as net income and EBITDA[1]), internal variables like shareholders and equity, as well as the classification of each firm.

Figure 4 presents the distribution of Net Income and EBITDA as indicators of firm profitability. Mean values reveal slightly negative profitability, with EBITDA averaging

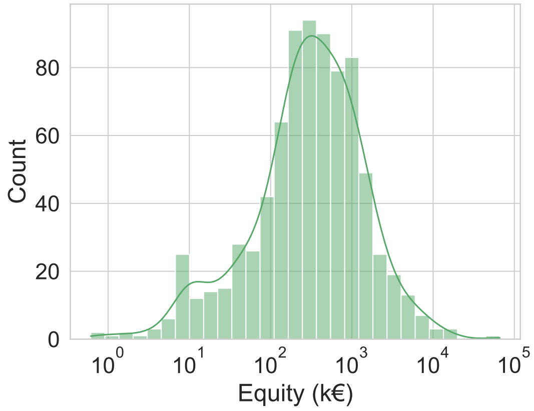
[1]EBITDA, an acronym for Earnings Before Interest, Taxes, Depreciation, and Amortization, serves as an alternative measure of profitability to net income. It is commonly used to evaluate a company's operating profitability and overall financial performance.



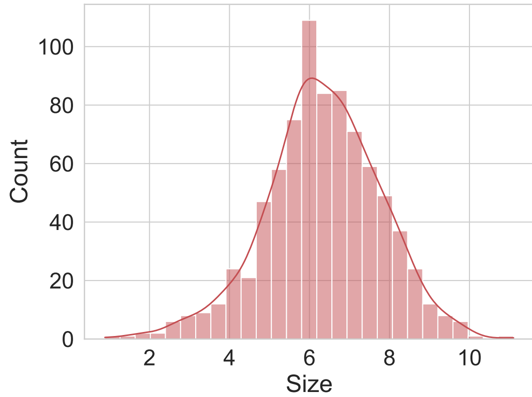
(a) Distribution of EBITDA (in thousands of euros).



(b) Distribution of Net Income (in thousands of euros).



(c) Distribution of Equity (in thousands of euros) on logarithmic scale.



(d) Distribution of Size (ln of Total Assets).

Figure 4: Distribution of EBITDA, Net Income, Equity and Size.

-0.06 k€ and Net Income -0.12 k€, which is typical for startups and SMEs. Shareholders' equity averages at 730 k€, ranging from a minimum of 8 k€ to a maximum of 10,000 k€, reflecting a wide range of capital structures. Additionally, firm size, as indicated by the log of total assets, exhibits high variability, following a normal distribution. These findings suggest a mixed financial profile overall, with assets and equity distributions being generally positive, while profitability remains modestly negative on average.

2.1 Human Evaluators

We invited students from all departments at the University of Pisa to enter a Business Plan competition. Each participant was assigned three Business Plans to evaluate according to the grid described in the Appendix at the end of the paper. Additionally, they were asked to guess several variables for the campaign, such as success and survival at five years. We also asked how much they would invest in the company if they had 1000 Euros. Upon completion of the task, each participant received a voucher to collect a University T-shirt from our city shop. Furthermore, we assigned scores to each participant based on their answers to the campaign questions. For every 50 participants, we created a ranking, and the top-ranked participant would also earn a University hoodie.

Table 1: Summary of Students' Characteristics and Scores

	Non-Expert	**Expert**	**Total Average**
Number of Students	223	25	248
Percentage	89%	11%	100%
Female	48%	32%	47%
Average Score	202.18	168.40	197.53
Average Investment	372.19	334.13	368.49
Correct Prediction (%)	66.27%	53.33%	64.4%

In Table (1), we report a summary of the students' characteristics along with their scores. Out of 248 students, a small percentage (about 11%) can be considered experts as they attended a course on Business Plan evaluation. On average, however, the non-expert evaluators achieved a higher score than the expert evaluators (202.18 vs 168.40) and they would invest slightly more in the company (372.19 Euros vs 334.13 Euros). Considering the prediction of campaign success, students correctly guessed about 65% of the equity offering, with the expert students doing worse than non-expert students (53% vs 66%). The complementary measure represents our first "raw" measure of error, the misclassification rate, defined as:

$$\text{Errors}^{Human} = \frac{\text{Number of Incorrect Predictions}}{\text{Total Number of Predictions}} \tag{1}$$

We will use this variable to gain a rough understanding of human (mis)judgment in comparison to different algorithms. However, it is important to note that this variable is not directly comparable to the errors made by the algorithms, as only the errors made by the model on the test dataset can be directly contrasted with human predictions. As explained below, we will develop an algorithm that incorporates variables based on human judgment and compare its predictive probabilities with those of algorithms that do not rely on human judgment.

3 Methodology

3.1 Different source of information

To build our algorithm and predict the success of equity crowdfunding campaigns, we investigate two distinct sources of information: the first one is the BP representation, through both implicit and explicit features derived from the entire business plan; the second consists of the variables obtained from the evaluation grid of the BP, which is filled in by humans. In greater detail, we employ the following models:

- BERT, which generates a succinct contextualized vector representation of the text by leveraging bidirectional transformers (Devlin, 2018). In contrast to traditional model, BERT is able to captures both the left and right context of a word within a sentence, enables the model to comprehend the meaning of words in relation to their surrounding context. [2]

- PROFILING-UD, which identifies and measures differences and similarities across texts representing various language varieties by examining the distribution of numerous linguistic features (Brunato et al., 2020). It is a web-based tool using the Universal Dependencies (UD) formalism, a widely adopted schema for morpho-syntactic and syntactic annotation in corpus linguistics, grounded in dependency syntax. [3]

- HUMAN_AI, which includes all the answers from the evaluation grid completed by students, according to different sections typically found in a business plan (Guarino

[2] It extracts over 760 features, though each feature is not easily interpretable.

[3] It allows the extraction of more than 130 features, spanning across different levels of linguistic description and it has been specifically devised to be multilingual since it is based on the Universal Dependencies framework.

and Mariani, 2021), such as entrepreneurial feasibility, internal and external feasibility, economic and financial feasibility, as well as business plan writing styles (see the end of the paper for a detailed description of the grid);

- CONTROLS, that is the combination of each of the previous model with control variables, such as the characteristics of the campaign and the firm (see also Signori and Vismara (2018)), as derived from the platform and AIDA.
- ALL in which either BERT or PROFILING-UD is combined with HUMAN_AI and CONTROLS

3.2 Evaluating firm's campaign success

In the following, we describe our model to evaluate the success of the campaign through the evaluation of the business plan, along with other firm and campaign characteristics. Let Y be the success of the financing application, equal either 0 (failure) or 1 (success), and let $\mathbf{X}$ be the feature vector:

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_n \end{pmatrix}$$

where:

$$\begin{aligned} X_1 &= \text{Macroeconomic variables} \\ X_2 &= \text{Textual BP} \\ X_3 &= \text{Company age} \\ X_4 &= \text{Company sector} \\ X_5 &= \text{Company total assets} \vdots \\ X_n &= \text{Other controls} \end{aligned}$$

The Random Forest model can be represented as an ensemble of T decision trees, where each tree t produces an individual prediction $\hat{Y}_t$:

$$\hat{Y}_t = f_t(\mathbf{X}), \quad t = 1, 2, \ldots, T$$

In particular, in X_2 we leverage different models for representing the text. Indeed, Mavillonio (2024) shows that going from a textual analysis to a textual representation can increase the model's ability to predict a firm's campaign success by up to 20%.

The final prediction $\hat{Y}$ is obtained by averaging the predictions of all individual trees:

$$\hat{Y} = \frac{1}{T} \sum_{t=1}^{T} \hat{Y}_t$$

In this case, the Random Forest model predicts the success of the financing application based on the combined information from macroeconomic variables, textual analysis of the BP, and company characteristics. Similar to human evaluators, for classification tasks, errors can be measured in terms of misclassification rate. The misclassification rate is defined as:

$$\text{Errors}^{AI} = \frac{\text{Number of Incorrect Predictions}}{\text{Total Number of Predictions}} \tag{2}$$

In Table 2, we report a summary of the results for different types of models, assessing their performance based on metrics such as training accuracy, test accuracy, and balanced accuracy. Training accuracy refers to the model's performance on the data used to train it, while test accuracy evaluates its generalization capability on unseen data.

Balanced accuracy accounts for class imbalance (i.e., the disproportion between successful and unsuccessful crowdfunding campaigns) by calculating the average recall obtained for each class, and it can be expressed as:

$$\text{Balanced Accuracy} = \frac{1}{C} \sum_{i=1}^{C} \text{Recall}_i \tag{3}$$

where C is the number of classes and Recall_i is the recall for class $i = \{success, failure\}$.[4]

Among the non-control models, BERT achieves moderate test accuracy at 66.19% and

[4]Recall for class i is defined as:

the highest test balanced accuracy at 58.59%. In contrast, PROFILING-UD shows the lowest test accuracy at 60.74%, though it performs comparably in terms of balanced accuracy, reaching 57.49%. The HUMAN_AI model demonstrates significant variability, with a lower test balanced accuracy of 46.96%. Introducing controls, leads to a general performance improvements. BERT_CONTROL and PROFILING-UD_CONTROL exhibit higher test accuracy at 72.77% and 75.35%, respectively, with corresponding balanced accuracy scores of 64.66% and 67.68%. HUMAN_AI_CONTROL stands out by achieving the highest test balanced accuracy at 71.59%. Finally, for the combined models, both ALL_BERT and ALL_UD perform well, with ALL_UD showing strong results in both test accuracy 72.91% and balanced accuracy 66.74%. These combined models highlight the effectiveness of integrating multiple approaches for improved performance. Overall, we can assert that the AI algorithm that relies on Bert textual representation achieves the best performance.

3.3 Understanding and Exploiting differences in judgement

In this section, we begin to model human "raw" error, which refers to the error in human judgment as defined above (see eq. 1). Our goal is to understand which characteristics of the business plan (BP) may lead humans to erroneously evaluate a firm's crowdfunding campaign. Figure 5 presents a confusion matrix comparing the predictions of the "raw" HUMAN and BERT models. It indicates that the two models agree on 391 correct predictions (about 56% of cases). However, in 177 instances (about 25%), the BERT model outperforms the Human model by making correct predictions whereas the Human model fails, while in 84 instances (about 12%), the Human model provides the correct predictions where the BERT model fails.

Indeed, we can observe that errors committed by humans and AI_Bert are not perfectly correlated, and AI_Bert tends to perform better. However, there are still about 7% of cases

---

$$\text{Recall}_i = \frac{\text{True Positives}_i}{\text{True Positives}_i + \text{False Negatives}_i} \tag{4}$$

where True Positives_i represents the number of correctly identified instances of class i, and False Negatives_i represents the number of instances of class i that were incorrectly classified as a different class. Recall measures the model's ability to identify all relevant instances of a given class. In datasets with class imbalance as ours, balanced accuracy provides a more accurate overall evaluation by averaging the recall across all classes.

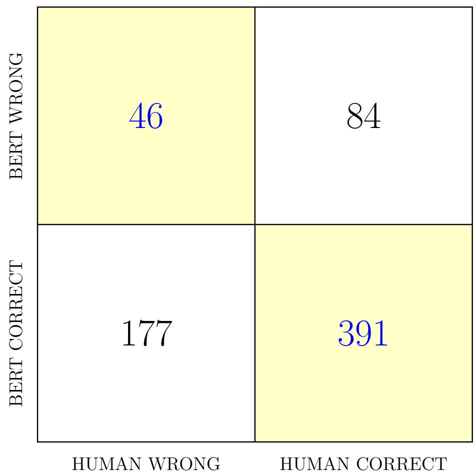


Figure 5: Confusion matrix of the human evaluation and BERT prediction

in which neither of them can guess correctly the campaign success. However, as stated earlier, we cannot directly compare the "raw" human error. Obviously, the algorithm has an advantage in this scenario as some of the guesses are made on the training dataset.

Therefore, we focus on the discrepancy between predictions made by the human-based algorithm, $\hat{y}_{\text{Human\_AI}}$, and those of our best-performing algorithm, $\hat{y}_{\text{BERT}}$. By using continuous prediction values rather than binary 0-1 errors, we gain a more nuanced view of the differences between the algorithms, capturing the magnitude of deviations rather than merely whether predictions are correct or incorrect. This approach allows us to observe how close each prediction is to the target, providing finer insight into performance.

In other words, we rely on the algorithm based on human evaluation of the business plan (BP) to assess the extent of alignment with the predictions of our best-performing AI model. This ensures that the algorithms being compared — one based on human input and the other independent of it — are both trained and tested on the same type of dataset. Specifically, we consider both absolute and squared deviations to capture the discrepancy in predictions:

$$Discrepancy_{abs} = |\hat{y}_{\text{Human\_AI}} - \hat{y}_{\text{BERT}}|$$

$$Discrepancy_{squared} = (\hat{y}_{\text{Human\_AI}} - \hat{y}_{\text{BERT}})^2$$

To further clarify, we can express the difference in predictions as: $\hat{y}_{\text{Human\_AI}} - \hat{y}_{\text{BERT}} =$

$(y - \hat{y}_{\text{BERT}}) - (y - \hat{y}_{\text{Human\_AI}})$. This formulation shows that the discrepancy in predictions is essentially a comparison of the respective errors, where $\text{Errors}^{\text{BERT}} = (y - \hat{y}_{\text{BERT}})$ and $\text{Errors}_{\text{Human\_AI}} = (y - \hat{y}_{\text{Human\_AI}})$ represent the deviations from the true value y for each algorithm. Interpreting the prediction difference in this way allows us to directly measure how each model's error varies relative to the other.

This model enables us to analyze how various characteristics of a business plan impact the discrepancies in prediction values between human evaluators and the AI model. By examining continuous prediction values rather than binary outcomes, we can capture more nuanced variations in judgment. This approach allows us to detect specific aspects of the business plan where human judgment and AI predictions diverge, offering insights into areas where human evaluators may apply subjective criteria that differ from algorithmic processing.

Specifically, to investigate these discrepancies, we perform a regression analysis by regressing the error variable (i.e., the absolute or squared discrepancy between human and AI predictions) on a set of key features that are assessed by human evaluators when reviewing a business plan. These features include:

- **Clarity**: How clearly the business plan communicates its objectives, structure, and value proposition.
- **Completeness**: The degree to which the business plan covers all essential aspects (such as market analysis, financial projections, and competitive positioning).
- **Editing Quality**: The level of attention to detail in grammar, spelling, and formatting, which can affect the perceived professionalism of the business plan.
- **Feasibility**: The practicality and realism of the business plan's goals and strategies, as judged by human evaluators.
- **Attractiveness**: The appeal of the business idea, which may reflect human bias towards certain sectors or innovative concepts.
- **Originality**: The uniqueness and innovativeness of the business concept, which might be valued differently by humans and algorithms.

By regressing the discrepancy variable on these features, we aim to understand which aspects of the business plan contribute most significantly to differences in judgment between humans and the AI model. For example, if the "clarity" or "completeness" of the business plan has a strong influence on the discrepancy, this may indicate that human evaluators place greater emphasis on these elements, whereas the AI model may not prioritize them to the same extent. Alternatively, high discrepancies associated with "attractiveness" or "originality" might suggest that human judgments are influenced by subjective preferences, which are less prominent in algorithmic evaluation.

Ultimately, this analysis provides valuable insights into how subjective and qualitative aspects of a business plan affect its evaluation. It highlights potential biases and priorities that differ between human and AI judgments, allowing us to refine the algorithm or develop hybrid models that better account for the qualitative factors that humans consider essential when assessing business plans.

The results in Table 4, column 1 and 2, show that an increase of 1 point in *Editing* has a positive contribution to increasing the gap between human and algorithmic judgment. Although the coefficient 0.001 in the squared error model (and 0.004 in the absolute deviation model) appears small, the result is statistically significant at the 5% level and economically important (considering that the mean error is equal to 0.070 and 0.007 respectively).

The *Completeness* of the business plan is also statistically significant at the 1% level and economically relevant, suggesting that an increase of 1 point in the completeness rating reduces the discrepancy by 0.002 (squared error) and 0.009 (absolute deviation).

Clarity also has a significant negative effect (*Coef.* = -0.001, significant at the 5% level), although only in the squared error specification. Other variables, such as *Originality* and *Feasibility*, do not significantly affect the discrepancy.

In columns 3 and 4, we control for individual characteristics and performance in the business competition. Although these variables are statistically significant, they do not alter the main coefficients.

In summary, the results show that improvements in *Editing* and *Completeness* significantly reduce the gap between human and algorithmic judgment, highlighting the importance of well-crafted business plans for consistent evaluations. While *Clarity* has a modest impact in the squared error model, factors like *Originality* and *Feasibility* show no significant effect. Controlling for individual characteristics does not alter these findings.

Overall, these results suggest that enhancing certain business plan qualities could improve coherence in assessments across different evaluative methods, with potential implications for optimizing entrepreneurial guidance.

4 Conclusions

This study explores the predictive power of human evaluation versus machine learning models in assessing the success of crowdfunding campaigns, specifically through the lens of business plan (BP) analysis. By comparing human evaluations with algorithmic predictions derived from BERT text representations and linguistic features, we identify both the strengths and limitations of each approach. Our analysis reveals that while AI models, particularly those using BERT representations, consistently outperform humans in overall accuracy, human evaluators retain specific advantages, especially when dealing with nuanced or context-dependent information.

The results indicate that certain business plan characteristics, such as clarity and attractiveness, play pivotal roles in shaping human predictions. However, the machine learning models excel in processing vast textual data, yielding more consistent results in scenarios where large-scale patterns are required for accurate predictions. The differences in errors between human and AI models, particularly their lack of perfect correlation, suggest that combining these two approaches can yield improved predictive performance.

Our findings emphasize the complementarity of human judgment and AI. Humans are adept at evaluating qualitative, intangible aspects of business plans, while AI excels in processing quantitative textual features and large datasets. Future research could explore how to effectively integrate human insights with machine learning techniques to create hybrid models that maximize predictive accuracy. Additionally, as AI tools evolve and gain more contextual understanding, their role in evaluating crowdfunding campaigns and other business decisions will likely expand, offering valuable insights for both entrepreneurs and investors.

References

Abolghasemi, Mahdi, Odkhishig Ganbold, and Kristian Rotaru (2024). "Humans vs. large language models: Judgmental forecasting in an era of advanced AI". In: *International Journal of Forecasting.*

Berger, Allen and Gregory Udell (1998). "The economics of small business finance: The roles of private equity and debt markets in the financial growth cycle". In: *Journal of Banking & Finance* 22(6), pp. 613–673.

Brunato, Dominique et al. (2020). "Profiling-ud: a tool for linguistic profiling of texts". In: *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pp. 7145–7151.

Clingingsmith, David, Will Drover, and Scott Shane (2023). "Examining the outcomes of entrepreneur pitch training: an exploratory field study". In: *Small Business Economics* 60(3), pp. 947–974.

Cole, Rebel and Tatyana Sokolyk (2016). "Who needs credit and who gets credit? Evidence from the surveys of small business finances". In: *Journal of Financial Stability* 24, pp. 40–60.

Devlin, Jacob (2018). "Bert: Pre-training of deep bidirectional transformers for language understanding". In: *arXiv preprint arXiv:1810.04805.*

Guarino, Loredana and Giovanna Mariani (2021). *Business plan smart. Una road map per la redazione di un business plan nella start-up.* Vol. 1. Pisa University Press.

Hall, Bronwyn H and Josh Lerner (2010). "The financing of R&D and innovation". In: *Handbook of the Economics of Innovation.* Vol. 1. Elsevier, pp. 609–639.

Hemmer, Patrick et al. (2022). "On the effect of information asymmetry in human-AI teams". In: *arXiv preprint arXiv:2205.01467.*

Kaminski, Jerzy C and Christian Hopp (2020). "Predicting Outcomes in Crowdfunding Campaigns with Textual, Visual, and Linguistic Signals". In: *Small Business Economics* 55, pp. 627–649.

Kim, Alex, Maximilian Muhn, and Valeri Nikolaev (2024). "Financial statement analysis with large language models". In: *arXiv preprint arXiv:2407.17866.*

Mason, Colin and Matthew Stark (2004). "What do Investors Look for in a Business Plan?" In: *International Small Business Journal* 22, pp. 227–248.

Mavillonio, Maria S (2024). *Textual Representation of Business Plans and Firm Success*. Tech. rep.

McKenzie, David (2017). "Identifying and spurring high-growth entrepreneurship: Experimental evidence from a business plan competition". In: *American Economic Review* 107(8), pp. 2278–2307.

McKenzie, David and Dario Sansone (2019). "Predicting Entrepreneurial Success is Hard: Evidence from a Business Plan Competition in Nigeria". In: *Journal of Development Economics* 141, p. 102369.

Politecnico di Milano – Dipartimento di Ingegneria Gestionale (2024). *9° Report Italiano sul Crowdinvesting*. Tech. rep. Politecnico di Milano – Dipartimento di Ingegneria Gestionale.

Signori, Andrea and Silvio Vismara (2018). "Does success bring success? The post-offering livesof equity-crowdfunded firms". In: *Journal of Corporate Finance* 50, pp. 575–591.

Steyvers, Mark et al. (2022). "Bayesian modeling of human–AI complementarity". In: *Proceedings of the National Academy of Sciences* 119(11), e2111547119.

Wang, Junbo et al. (2024). "From Man vs. Machine to Man+ Machine: The Art and AI of Stock Analyses". In: *Journal of Financial Economics*.

Table 2: Performance comparison of different models

Method	Tr Acc (%)	Test Acc (%)	Tr Bal. Acc (%)	Test Bal. Acc (%)
BERT	83.10 ± 2.72	66.19 ± 3.91	85.28 ± 2.88	**58.59** ± 4.95
PROFILING−UD	67.91 ± 1.25	60.74 ± 1.55	69.34 ± 1.36	57.49 ± 1.26
HUMAN_AI	83.92 ± 14.72	62.45 ± 7.17	85.89 ± 14.33	46.96 ± 1.49
BERT_CONTROL	87.30 ± 2.82	72.77 ± 3.03	89.11 ± 2.45	64.66 ± 5.75
PROFILING−UD_CONTROL	86.44 ± 5.80	75.35 ± 3.61	86.69 ± 6.71	67.68 ± 2.05
HUMAN_AI_CONTROL	82.57 ± 7.11	75.78 ± 3.24	83.63 ± 7.51	**71.59** ± 5.39
ALL_BERT	84.29 ± 2.00	69.06 ± 6.31	86.29 ± 1.73	61.79 ± 8.31
ALL_UD	83.19 ± 2.40	72.91 ± 3.74	84.01 ± 3.34	66.74 ± 4.25

Table 3: Decision Tree Subjective Variables

Variable	**Importance**
Clarity	0.35635716
Completeness	0.10502927
Originality	0.26971111
Feasibility	0.0383106
Attractiveness	0.15327627
Editing	0.07731559

Table 4: Discrepancy errors

	(1)	(2)	(3)	(4)
Discrepancy	Squared	Absolute	Squared	Absolute
Editing	0.001**	0.004*	0.001**	0.004*
	(0.000)	(0.002)	(0.000)	(0.002)
Completeness	-0.002***	-0.009***	-0.002***	-0.009***
	(0.000)	(0.002)	(0.000)	(0.002)
Clarity	-0.001**	-0.004	-0.001**	-0.004
	(0.001)	(0.003)	(0.001)	(0.003)
Originality	0.000	0.002	0.000	0.003
	(0.000)	(0.002)	(0.000)	(0.002)
Attractiveness	-0.000	-0.002	-0.001	-0.002
	(0.001)	(0.003)	(0.001)	(0.003)
Feasibility	0.001**	0.005*	0.001**	0.005*
	(0.001)	(0.003)	(0.001)	(0.003)
Individual Score			0.000	-0.000
			(0.000)	(0.000)
Age 2			-0.001	-0.002
			(0.001)	(0.004)
Age 3			0.005*	0.027**
			(0.002)	(0.013)
Age 4			0.002*	0.009*
			(0.001)	(0.005)
Female			0.001*	0.007*
			(0.001)	(0.004)
Constant	0.009***	0.075***	0.008***	0.072***
	(0.001)	(0.005)	(0.002)	(0.008)
Observations	698	698	698	698
R-squared	0.057	0.041	0.075	0.058

Standard errors in parentheses
**** p<0.01, ** p<0.05, * p<0.1*

A Business Plan Evaluation Grid

Table 5: Business Plan Evaluation Grid

Section	**Criteria**
Index	Presence of the index (1/0)
Executive Summary	Brief description of the project and objectives pursued Characteristics of the product/service Size of the potential market and sales potential Strengths and weaknesses compared to competitors Opportunities to be seized with the business idea Mission, Vision of the company, company history Entrepreneur's profile Summary data on project profitability Financial needs and hypotheses on the mix of sources Type of collaboration requests to the recipient of the BP

Table 5: Business Plan Evaluation Grid (continued)

Section	**Criteria**
Entrepreneurial Feasibility	Analysis of the skills of the entrepreneurial team and key personnel Strategic collaborations Core Competence Economic-financial performance of the past years Existence of a multidisciplinary team suitable for the project (1/0) Existing collaborations Brief description of the team (age, profession, and experience)

Table 5: Business Plan Evaluation Grid (continued)

Section	Criteria
External Feasibility	Analysis of the reference market Needs that the product/service can meet Analysis of the overall economic system Current and prospective reference market Development of overall sector demand Identification and analysis of the chosen market segments Estimate of the target market Analysis of current and potential competition Definition of market penetration percentage Estimate of sales potential (quantity) Profile of potential consumers and purchase/consumption process Presence of a SWOT matrix (1/0) Presence of a positioning map (1/0)

Table 5: Business Plan Evaluation Grid (continued)

Section	Criteria
Internal Feasibility	Operational plans Technical-production plan Marketing and communication plan Organizational plan Feasibility of the product and its production process (1-5) Business model used (e.g., Transactional, SAAS, etc.) Use of the Business Model Canvas (1/0) Intellectual property protection (1/0) Presence of a Go to Market strategy (1/0)

Table 5: Business Plan Evaluation Grid (continued)

Section	Criteria
Economic/Financial Feasibility	Quantification of costs (structural and operational) Production costs Marketing costs Organizational and general costs Full cost hypothesis Definition of sales price and estimated revenue Economic-financial forecasts Financial needs of the initiative Hypothesis of the financing mix and resulting financial structure Forecasted budgets and analysis with indices Overall financial dynamics Break Even Point NPV and IRR Risk level assessment of the project

Table 5: Business Plan Evaluation Grid (continued)

Section	**Criteria**
Section Identifiability and Positioning	Clarity Completeness Originality Feasibility Attractiveness Editing
Importance of Tables	For each section, the tables and/or images were evaluated for their usefulness
Predictions	Clarity Completeness Originality Feasibility Attractiveness Editing Forecast for the company in the next 5 years (growth, acquisition, stability, failure)
Overall Assessment	Assessment of the overall characteristics of the BP (insufficient, sufficient, good, excellent)