Crowdfunding Success: Human Insights vs Algorithmic Textual Extraction

Caterina Giannetti

Maria Saveria Mavillonio*

September 12, 2024

PRELIMINARY AND INCOMPLETE

Abstract

We investigate the synergy between human insights and algorithmic analysis in assessing crowdfunding campaign success by comparing their strengths and weaknesses in evaluating business plans. In a Business Plan competition, students from various departments used a predefined grid to evaluate business plans. We then developed a classifier using the latest textual representations and compared prediction errors among human evaluators, a machine learning model, and a combination of both, considering the characteristics of the plans. Our goal is to identify the drivers of discrepancies in their evaluations.

Keywords: Crowdfunding, Natural Language Processing, Human Evaluation

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

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

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.

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. Mason and Stark (2004) emphasize 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 different algorithms, which also include textual information, in predicting campaign success. We then seek to explore whether human knowledge and assessment of the business can be leveraged by the algorithm to provide an improved synthetic score reflecting the quality of the business plan.

Recently, Wang et al. (2024) showed how AI analysts outperform human analysts in stock forecasts, especially in transparent and data-rich environments. Human analysts instead retain an advantage in interpreting critical information requiring institutional knowledge, such as intangible assets. They also show that text information can contribute 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 the AI during training — due to technical or economic reasons — the teams achieve a lower mean absolute error (MAE) compared to teams without contextual information and the AI alone. Close to our paper, in a Nigerian business competition, McKenzie and Sansone (2019) found that human evaluators are no better than a machine learning algorithm in predicting this success companies' by analyzing their business plans. However, they do not provide insights on whether and when one of the two may have an advantage or disadvantage. 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 comparison highlights the complementary strengths of human judgment and AI, offering insights into improving the

evaluation process for crowdfunding campaigns.

2 Data Description

We gathered business plans and campaign data by web scraping all available documents from seven different Italian crowdfunding platforms (*Startups, Crowdfundme, Ecomill, Opstart, WeAreStarting, MamaCrowd, and BacktoWork*) from the inception of crowdfunding in 2014 to the end of January 2024.

Starting with 848 successfully funded offerings, we excluded 14 mini-bond and 16 real estate offerings, resulting in 818 initial equity offerings. Due to file extension or language issues, 28 business plans were unusable, leaving a final sample of 790 observation- We augmented our database with AIDA *Bureau Van Dijk* to derive additional financial variables and firm characteristics (see also Mavillonio (2024)) and macro-economic variables from Eurostat (to add).

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.

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.

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%

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 measure of error, the misclassification rate, defined as:

$$\text{Errors}^{Humans} = \frac{\text{Number of Incorrect Predictions}}{\text{Total Number of Predictions}}$$

3 Methodology

3.1 Different source of information

We investigate two different sources of information to predict the success of equity crowdfunding campaign. The former is the representation, both implicit and explicit extracted from the entire Business Plan. The latter is all the variables obtained from the evaluation grid. More in deep , we employ the following models:

- BERT, which generates a succinct vector representation of the text;
- PROFILING-UD [1], which computes multiple features about the lexical

[1]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 morpho-syntactic structure.

- HUMAN, which includes all the answer of the evaluation grid made by student;
- the combination with the control variables, such as the characteristics of the campaign and firm.

3.2 Evaluating firm's campaign success

Let Y be the success of the financing application, 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:

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

In Table 3 we report a summary of the results for different type of models. The AI algorithm that relies on Bert textual representation achieves the best performance. The result compares the performance of various models (BERT, PROFILING-UD, HUMAN, and their control versions) based on training and test accuracy, as well as balanced accuracy.

- **BERT** achieves moderate test accuracy (66.19%) and the highest test balanced accuracy (58.59%) among non-control models.
- **PROFILING-UD** has the lowest test accuracy (60.74%) but performs comparably on balanced accuracy (57.49%).
- **HUMAN** shows significant variance, with lower test balanced accuracy (46.96%).

Control models improve performance:

- **BERT_CONTROL** and **PROFILING-UD_CONTROL** show higher test accuracy (72.77% and 75.35%) and balanced accuracy (64.66% and 67.68%).

- **HUMAN_CONTROL** achieves the highest test balanced accuracy (71.59%).

For combined models:

- **ALL_BERT** and **ALL_UD** perform well, with ALL_UD showing strong test accuracy (72.91%) and balanced accuracy (66.74%).

3.3 Exploiting differences in judgement

In this section, we develop a model where the dependent variable is the error discrepancy between human and AI predictions. We first begin by comparing the plain prediction of humans evaluators with that of the Algorithm with rely on Bert representation of the text. This can be expressed as:

$$E_{\text{discrepancy}} = \left|\text{Error}^{\text{human}} - \text{Error}^{\text{AI}}\right|$$

where Errors are defined as above.

Figure 1 presents a confusion matrix comparing the predictions of the Human and BERT models. The matrix consists of four quadrants:

- **Top-left (46)**: Both the Human and BERT models produce incorrect predictions.
- **Top-right (84)**: The Human model predicts correctly, while the BERT model predicts incorrectly.
- **Bottom-left (177)**: The BERT model predicts correctly, whereas the Human model predicts incorrectly.
- **Bottom-right (391)**: Both the Human and BERT models predict correctly.

This confusion matrix indicates that the models agree on 391 correct predictions. However, in 177 instances, the BERT model outperforms the Human model by making correct predictions where the Human model fails, while in 84 instances, 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 6.5% of cases in which neither of them can guess correctly the campaign success. Obviously, the algorithm has an advantage in this scenario as some of the guesses are made on the training dataset.

We therefore also compare the discrepancy between the algorithm based on human evaluations (AI_Human) of the business (as a result of the grid) and the AI_Bert algorithm. In this case, we observe a different situation, as reported in figure **??**.

3.4 Understanding the drivers of error discrepancy

To understand the drivers of the difference we therefore model the error discrepancy is modeled as a function of business plan characteristics, including readability, completeness, and three other characteristics.

$$E_{\text{discrepancy}} = \beta_0 + \beta_1 \text{Clarity} + \beta_2 \text{Completeness} + \beta_3 \text{Novelty} + \\ + \beta_4 \text{Feasibility} + \beta_5 \text{Attractiveness} + \beta_6 \text{Editing} + \epsilon \quad (1)$$

This model allows us to analyze how different aspects of a business plan influence the discrepancy in prediction errors between human evaluators and the AI model. The result in the table 4 shows that the variable *Clarity* has significant negative effect (*Coef.* = -0.1054, *P-value* = 0.001), while *Completeness* and *Attractiveness* are positively significant (*P-value* < 0.05). Other variables like *Originality*, *Feasibility*, and *Editing* are not statistically significant. However, the adjusted R-squared is relatively low (0.043), suggesting limited explanatory power.

Therefore, we leverage the differences between predictions derived solely from human evaluation and those based exclusively on BERT representations.

In contrast, the **BERT model** demonstrates a strong fit with an adjusted R-squared of 0.711. Here, *Clarity*, *Originality*, and *Feasibility* show

statistically significant positive effects (*P-value* 0.05), while *Completeness* and *Attractiveness* are not significant.

The **HUMAN model** achieves the highest adjusted R-squared (0.779), indicating the strongest fit among the models. Similar to the BERT model, *Clarity*, *Originality*, and *Feasibility* are significant predictors (*P-value* $<$ 0.05). Additionally, *Attractiveness* also shows a significant positive effect. However, *Completeness* and *Editing* remain non-significant.

These results highlight the differences in variable significance and model fit across the three models, with the HUMAN model performing the best, followed by the BERT model, and the Error model showing the weakest performance.

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

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.

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.

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*.

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.

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: Decision Tree Subjective Variables

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

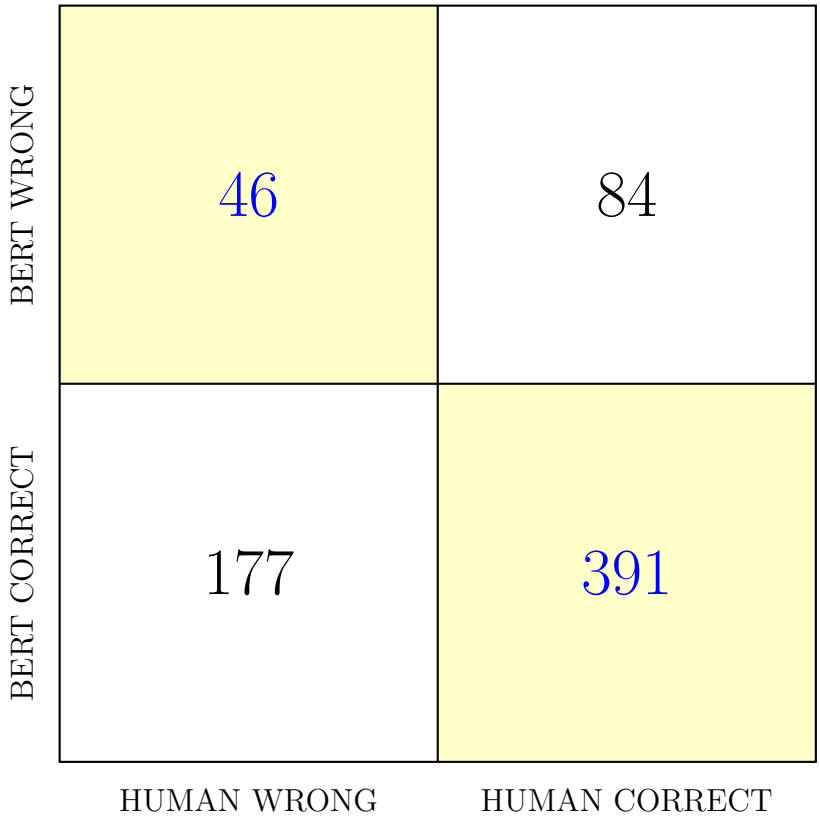


Figure 1: Confusion matrix of the Human and BERT prediction

A Business Plan Evaluation Grid

Table 5: Business Plan Evaluation Grid

Section	**Criteria**
Index	Presence of the index (1/0)

Table 5: Business Plan Evaluation Grid (continued)

Section	**Criteria**
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)

Table 3: 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	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_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 4: Regression Results

Model	**Coef.**	**Std. Err.**	**t**	**P-value**	**[0.025**	**0.975]**
Error: Adj. R-squared = 0.043						
Clarity	-0.1054	0.031	-3.351	0.001	-0.167	-0.044
Completeness	0.0557	0.028	1.993	0.047	0.001	0.111
Originality	-0.0296	0.029	-1.027	0.305	-0.086	0.027
Feasibility	-0.0477	0.030	-1.615	0.107	-0.106	0.010
Attractiveness	0.0817	0.033	2.472	0.014	0.017	0.146
Editing	0.0066	0.026	0.251	0.802	-0.045	0.058
Error BERT: Adj. R-squared = 0.711						
Clarity	0.1108	0.025	4.358	0.000	0.061	0.161
Completeness	-0.0297	0.023	-1.315	0.189	-0.074	0.015
Originality	0.0480	0.023	2.061	0.040	0.002	0.094
Feasibility	0.1902	0.024	7.967	0.000	0.143	0.237
Attractiveness	0.0493	0.027	1.848	0.065	-0.003	0.102
Editing	0.0065	0.021	0.306	0.760	-0.035	0.048
Error HUMAN: Adj. R-squared = 0.779						
Clarity	0.0747	0.023	3.238	0.001	0.029	0.120
Completeness	-0.0318	0.020	-1.554	0.121	-0.072	0.008
Originality	0.0517	0.021	2.444	0.015	0.010	0.093
Feasibility	0.2176	0.022	10.048	0.000	0.175	0.260
Attractiveness	0.1310	0.024	3.054	0.002	0.026	0.121
Editing	0.0215	0.019	1.117	0.264	-0.016	0.059