
Enhancing Crowdfunding Success for Startups Based on Promotional Text Quality: An Advanced Analysis Using Fast Text Models , indiegogo platform Example

Abdallah GUERBI *

Laboratory of Economic Studies and Local.

Development in the Southwest Tahri Mohammed

University - Algeria

guerbi.abdallah@univ-bechar.dz

Received: 28/01/2026

Mohammed Elamine BOUDKHIL

Laboratory of Economic Studies and Local.

Development in the Southwest,Tahri Mohammed

University - Algeria

boudkhil.mohammed@univ-bechar.dz

Published: 31/05/2026

Abstract:

Crowdfunding has revolutionized entrepreneurial finance; however, campaign success remains elusive. While multiple factors contribute to this, this study posits that the linguistic fabric of campaign descriptions is a critical, underexplored predictor. This study aims to answer a key question: to what degree do the linguistic and semantic features of campaign descriptions on Indiegogo forecast their ultimate success? We propose that successful campaigns tend to use language that is more precise and unambiguous than those that do not succeed. Using a quantitative approach, we examined 20,617 Indiegogo campaigns with Natural Language Processing (NLP). Word embeddings from the FastText model were used to capture semantic relationships, supplemented by comparative methods like TF-IDF, lexical diversity, and semantic similarity. The results strongly confirm our hypothesis. Successful campaigns showed notably greater semantic coherence, diversity, and lexical accuracy. Conversely, unsuccessful campaigns displayed linguistic variability, longer average word lengths, and language that was repetitive and less focused. Visualizing the FastText vector space with t-SNE revealed clear groupings for successful campaigns, suggesting linguistic consistency and a well-defined strategic message. This research offers three main contributions: it empirically links linguistic theory with crowdfunding success models, highlighting the importance of persuasive communication; it proves the effectiveness of sophisticated embedding models like FastText for analyzing financial texts; and it lays the groundwork for practical tools, such as Fast and Frugal Trees, to assist creators in refining their campaign language. The findings emphasize that in crowdfunding, the way an idea is presented is just as important as the idea itself.

Keywords: Crowdfunding, Natural Language Processing (NLP), FastText, Semantic Analysis, Persuasive Communication, Indiegogo, Entrepreneurial Finance., startups.

Jel Classification Codes : G32 ; C45 ; G24 ; M13

* Corresponding author.

1. Introduction :

In recent years, crowdfunding has become one of the most innovative and most attention-grabbing funding mechanisms, successfully transforming the idea of crowdfunding into a unique interactive experience for the project community. This mechanism allows entrepreneurs, especially small businesses, to access capital directly without having to resort to traditional channels of funding. It's pretty clear this sector is growing, seeing as its global volume is around \$13 billion. Nine billion dollars in 2023, with reward platforms leading the market.

Indiegogo, which hosts a wide range of campaigns in a wide range of categories, such as technology, culture and media, is a prominent example.

There are many factors that can make a campaign run well, but the words and the text of the campaign ad are extremely important. They show how creators talk about what they need and what they want, and how they connect with people. Just by changing a few words here and there, and telling your story, you can change the way people see you, the way they believe in you, and how they decide to do it. Although academics are increasingly interested in what makes crowdfunding projects successful--and usually look for external things like who the donors are--they are not really in the know. We haven't really digested deeply enough what makes these campaigns tick at the level of language and meaning. Past research has often treated text as a secondary element or skimmed the surface with key words. They really missed how much stories and choice of words could influence people. Subtle differences in the appropriate choice of words and narrative structure can make a significant difference in the perception, trust and timely decision-making of supporters. So the big question we want to answer is this: **how much does the the linguistic and semantic characteristics of the texts in the descriptions of Indiegogo campaigns contribute to predict whether or not these campaigns will succeed?**

This research hypothesizes that successful crowdfunding campaigns are characterised by more specific and clear language patterns than unsuccessful campaigns. Is relevant to advance understanding of persuasive communication in business by integrating language theory into frameworks for crowdfunding success, demonstrating the applicability of the Fast and Slow Tree model to financial decision making and highlighting how optimal wording can improve the chances of success for creators.

2. Theoretical Framework:

2.1. Persuasive Communication in Crowdfunding

2.1.1. Promotional texts in crowdfunding: They are reflected in influencing funders' decisions to contribute to supporting various campaigns by highlighting the level of professionalism and

expertise of the campaign creator. for example, a study conducted on the Indiegogo platform found that the characteristics of the story section in campaigns are directly associated with the amount of funding a campaign attracts. This indicates that funders do not look only at the idea itself but are also influenced by the quality of the narrative and the way it is presented. moreover, analyzing textual characteristics using natural language processing techniques contributes to uncovering hidden patterns in campaign language and helps predict its performance, thereby providing an effective solution to reduce risk and improve the quality of investment decision making(Gunduz 2024:3).

2.1.2. Promotional speech as a persuasive mechanism: This persuasive mechanism is manifested in promotional texts, whether written or visual, and is constructed within a discursive system primarily concerned with producing trust and activating contribution. Analyses of crowdfunding related video content have shown that the discourse relies on carefully designed rhetorical moves that perform four strategic functions: demonstrating competence, diagnosing the problem, presenting the solution, and issuing an explicit call for contribution. Accordingly, the promotional text takes the form of a hybrid discursive model that blends scientific and advertising registers, allowing for an interplay between objectivity and emotional appeal mechanisms, whereby language shifts from a mere tool of expression into an effective persuasive act in shaping the recipient's attitudes and behavior(Vivas-Peraza 2025:4.5).

2.2. Linguistic and psychological characteristics of promotional texts:

Effective promotional texts are based on an integrated combination of linguistic and psychological characteristics that work to enhance the discourse's ability to persuade and influence supporters. Among the most important of these characteristics are:

2.2.1. Emotional Tone and Sentiment Analysis: This is reflected through identifying the emotional tone of campaign texts, whether positive negative or neutral. In the context of crowdfunding campaigns, for example, a positive tone is used to convey enthusiasm and reinforce trust among the audience, while a negative tone is strategically employed in certain cases to highlight the magnitude or severity of the problem addressed by the project. sentiment analysis indicators are among the most prominent predictive factors of campaign success, which reflects the role played by the emotional dimension in shaping funding decision making(Gunduz 2024:2).

2.2.2. Narrative Structure: This is manifested in describing projects through stories characterized by strong expressive capacity and the inclusion of specific characters, which contribute to increasing the level of emotional attachment among funders. The structural features of narrative, such as

coherence, activate cognitive domains associated with emotions, which indicates its high ability to influence funding decisions (Caterina and Maria Saveria 2024:5).

2.2.3. Persuasive Linguistic Techniques: It involves the use of linguistic techniques such as metaphors, questions, and emotional appeals to encourage funders to take swift action. Understanding how these techniques are used can help in delivering more effective messages.

2.2.4. Writing quality: It includes features such as readability, grammar, and spelling. A study using the "BERT Score" as an indicator of writing quality found that it is inversely related to the amount of funding, suggesting that the relationship may be complex and non intuitive.

2.3. linguistic characteristics influencing the success of crowdfunding campaigns:

This is reflected in determining the extent to which the effectiveness of marketing messages impacts crowdfunding campaigns, as they are considered among the key factors contributing to shaping the impressions of potential funders. From this standpoint, we address the most important linguistic characteristics that influence campaign success:

2.3.1. Length and number of sentences and words: Text length is considered a factor with dual effects, as its impact depends on the nature of the target audience and the type of product. The ideal length, ranging between 300 to 600 characters or 15 to 25 sentences, is considered the best for campaign success in crowdfunding, as it provides sufficient information (Mitra and Gilbert 2014), (Kaminski and Hopp 2020:17).

2.3.2. Clarity and syntactic structure: It involves using simple grammatical constructions instead of nominal or passive sentences to increase funders' ability to understand the project idea and implementation mechanisms. Repeating the main ideas helps reinforce the message more quickly in the minds of individuals or entrepreneurs interested in project funding (Peng et al. 2022:9).

2.3.3. Positivity and negativity in expression: The emotional tone of the text is one of the most influential elements in the decision-making process, as using positive language enhances funders' acceptance of the project and increases the likelihood of their participation, with words such as "innovative" or "exciting." In contrast, negative words such as "impossible" or "failure" are factors that reduce its attractiveness (Allison et al. 2017:8). However, excessive positivity may suggest a lack of credibility or indicate that the project owner is unaware of potential risks, so a balance between the two is recommended (Kolbe, Mansouri, and Momtaz 2022:2).

2.3.4. Emotional engagement and use of pronouns: Emotional engagement is an effective means of building a bridge of trust and closeness between the project owner and funders, as using first-person

pronouns such as “we” and “I” enhances the sense of participation and shows the project owner’s commitment to their idea. Additionally, using interactive techniques, such as direct questions or inviting the funder to participate, for example, “We would be happy to have you join us,” increases the target audience’s engagement with the content. This can achieve funding rates up to 20% higher compared to those using a dry, informational style (Parhankangas and Renko 2017:7).

2.3.5. Lexical richness and semantic diversity: Lexical richness reflects the diversity of vocabulary used in the text and is considered an indicator of content quality and the effort invested in its formulation. It is perceived as more professional and reliable, including the use of technical or creative terms related to the project, which demonstrates the project owner’s knowledge of their field. Additionally, it prevents boredom and maintains the reader’s attention (Mollick 2014:14). It increases the success of crowdfunding campaigns by up to 15% in some categories (Peng et al. 2022:6.7).

2.4. Text representation models in machine learning and deep learning:

2.4.1. The concept of text representation in artificial intelligence: It forms the foundation that enables computer systems to process and understand natural texts, involving the transformation of text into numerical or symbolic forms that machines can handle efficiently (Wang and Zhang 2018).

2.4.2. Traditional methods of text representation: It uses representation by words or grammatical units based on statistical measures such as TF-IDF to determine the importance of words in texts, which helps in text classification and information retrieval (Yang 2025:4). Models such as word distribution representations rely on calculating word frequency and importance statistically, and work on measuring text similarity using formulas like the Kullback Leibler distance (Wang and Zhang 2018:396).

2.4.3. Advanced Embedding Models: This involves using models such as Word2Vec and GloVe, which produce dense, vectorized representations of words that capture contextual relationships between them and are used to enhance text understanding in a non abstract way (Yang 2025:3)

2.4.4. Deep language and transformer models: The FastText model demonstrates the ability to process text in a fully contextual manner, relying on learning representations based on the complete context of a sentence or paragraph, which significantly improves meaning comprehension and increases task efficiency. Transformer models, in turn, are based on self-attention technology and produce precise contextual representations, making them ideal for tasks such as entity recognition, classification, and machine translation (Chuan et al. 2021:16.18). Additionally, it improves the process of selecting and generating the most effective promotional texts, as it captures the diverse

semantic and contextual relationships of words accurately, even in the presence of unfamiliar or out of vocabulary words, thanks to its ability to incorporate precise meanings that enhance the strength of the promotional message (Babić et al. 2020:6).

3. Methodology and Data

3.1. Research Design:

This research paper relies on a quantitative analytical methodology by integrating Natural Language Processing (NLP) techniques through examining the linguistic and psychological characteristics of promotional texts that influence the success of crowdfunding campaigns. The FastText model was used to generate high-density numerical representations (word embeddings) that reflect the semantic and contextual structure of the texts, in addition to employing comparative semantic analysis methods (TF-IDF, lexical similarity, and language diversity) to interpret the differences between successful and unsuccessful campaigns. This methodology allowed for combining advanced textual analysis with behavioral economic interpretation, thereby providing a robust framework for understanding the determinants of success in crowdfunding.

3.2. Data Collection

Data related to Indiegogo campaigns were collected via the Kaggle website, with the final sample including 20,617 campaigns randomly selected from various categories (Art, Technology, Design, Games, Publishing, Music). These cover all major projects proportionally according to transparent success metrics, as shown in the following table:

Table N ° 01: Numerical and categorical variables

Variables Name	Variables Types
Categorical	Logo
Categorical	Category
Categorical	Country
Numeric	Funding goal
Categorical	Campaign title
Numeric	Campaign status
Numeric	Duration

Source: Authors' elaborations based on the Indiegogo database.

4. Analytical Methods

4.1. Study tools and methods:

For analytical purposes, the success variable was converted into a binary value (1 = success, 0 = failure)

based on whether the achieved amount exceeded the funding goal. Prior to applying the FastText model, the texts were subjected to a series of preprocessing steps aimed at transforming them into numerical representations, which included:

- Removing non linguistic symbols and marks.
- Passing the texts through the model to obtain numerical outputs.
- Eliminating meaningless words.
- Verifying that the original campaign language was English.

FastText is an advanced model in natural language processing designed to improve the quality of text representation and the accuracy of predictive models. It converts words into high-dimensional numerical embeddings that capture both meaning and linguistic context beyond traditional frequency based methods. By representing words through smaller subword units, it effectively handles rare words and spelling errors and novel terms commonly associated with innovative projects on crowdfunding platforms.

After generating word representations, fastText computes a comprehensive vector representing the entire text, which serves as input for predictive models used in campaign classification, sentiment analysis, and success probability estimation. This enables the model to identify linguistic patterns most strongly associated with campaign success and to extract textual dimensions that have the greatest impact on predictive outcomes.

Thanks to these features, the fastText model is an effective choice for analyzing promotional texts in crowdfunding studies due to its speed, flexibility, and ability to provide rich linguistic representations that support the development of accurate and reliable analytical models. Additionally, it provides indicators of variable importance that assist in interpreting results within the research framework.

5. Results and Analysis:

5.1. Descriptive Analysis of the Variables:

In light of the available descriptive data, it is clear that the database used for the analysis includes 20,617 crowdfunding campaigns distributed across 28 different categories over a period of 11 years and within 10 countries, which reflects a high level of temporal, geographical, and sectoral diversity, and grants the study stronger explanatory power. The results indicate that 2,108 campaigns were successful compared to 18,509 failed campaigns, meaning that the success rate does not exceed approximately 10%, which is an indicator of intense competition and the difficulty of achieving funding goals on digital platforms. As for financial targets, the average requested goal reached USD 26,702, while the median amounted to USD 10,000, representing a large disparity that reflects a positive skewness in the distribution of values; that is, a small proportion of

campaigns targets very high amounts, which raises the overall mean compared to the median. This dispersion confirms that campaigns differ fundamentally in their size and in the types of projects they present, making sectoral analysis and comparative analysis across categories and countries a necessary step to understand the dynamics of success and failure in crowdfunding. the following table illustrates this:

Table N ° 02: Descriptive Analysis of the Variables

Item	Value
Total number of campaigns	20617
Number of successful campaigns	2108
Number of failed campaigns	18509
Number of categories	28
Number of years	11
Number of countries	10
Average target (USD)	26702.507
Median target (USD)	10000

Source: By the researchers based on the Indiegogo database

5.2. Textual Data Analysis:

The results presented in Table No. 03 highlight that linguistic patterns reflect the communication style adopted by startups to mobilize backers on crowdfunding platforms. The average number of words reached 12.9 words, which is due to the fact that most campaigns rely on short and direct texts, with cases reaching up to 35 words as a maximum, corresponding to campaigns that use more detailed content. In addition, the average number of characters was recorded at 78.1 characters, with a relatively high dispersion (Std = 23.27), indicating a noticeable variation among campaigns in terms of the level of detail in the published announcements.

As for the average word length, it reached 6.16 characters, with close median and quartile values, reflecting a general reliance on medium-length words of an explanatory nature. This pattern is common in marketing texts that aim at persuasion without excessive length. The high maximum word length of 72 characters indicates the possible presence of technical terms or long project names in a limited number of campaigns.

Table N ° 03 : Digital Data

Indicator	Mean	Std	min	max	75%	50%	25%
Number of word	12.92104	4.22071	0	35	16	13	11
Number of characters	78.10157	23.278	0	159	94	86	66
Average word length	6.165374	1.303548	0	72	6.692308	6	5.470588

Source: By the researchers based on the Indiegogo database

5.3. Performance Metrics and Classification:

The figures in Table No. 04, related to the performance metric, indicate an accuracy value of 0.7732, showing that the model used for analyzing campaigns and classifying texts achieves an effectiveness level of approximately 77.3%, which is relatively good in applied contexts characterized by data complexity and diversity.

Table N ° 04: Model Performance Metric

Mertic	Value
Accuracy	0.7732

Source: Prepared by the researchers based on the performance test results of the TestFast model

As for the classification report, the results shown in Table 05 highlight the model's ability to distinguish between campaigns through an achieved accuracy of 0.7927, meaning that 79% of the failed campaigns were correctly predicted, while the recall reached 0.7436, reflecting the effectiveness in detecting most failed campaigns. The same applies to successful campaigns, where the model achieved an accuracy of 0.7558, which is good considering that successful campaigns are usually fewer and pose a greater challenge for the model. The high recall (0.8033) demonstrates the model's strong ability to identify most successful campaigns without overlooking them. The F1 score serves as a clear indicator that the model effectively balances avoiding misclassification of campaigns with the ability to accurately detect successful ones.

Table N ° 05: Performance Metrics of the Classification Model for Successful and Failed Campaigns

Class	Score F1	Recall	Precision
Failed Campaigns	0.7674	0.7436	0.7927
Successful Campaigns	0.7788	0.8033	0.7558

Source: Prepared by the researcher based on the model performance results

5.4. Keywords indicative of success in crowdfunding campaigns:

5.4.1 Successful campaigns from the TF-IDE perspective: The results of the analysis in table No 06 highlight that linguistic content plays an important role in enhancing the effectiveness of funding campaigns for startups, and this is evident through the most frequently used words within the discursive and thematic patterns associated with higher probabilities of success. Most of the high-index keywords are linked to technology, innovation, and everyday use products, reflecting the nature of projects with broad market demand. Words such as wireless and smart speaker stand out as clear indicators that successful campaigns often revolve around modern technological products that offer innovative solutions to consumers. The repetition of terms such as charging, audio, powerful, and ultimate indicates that

describing performance features and product quality represents a marketing element that contributes to building funders' trust. Furthermore, words like design, smallest, first, and waterproof reflect a marketing orientation toward highlighting distinction, innovation, and uniqueness, which aligns with modern digital marketing practices that emphasize superiority and novelty to attract audience attention. Meanwhile, the word worlds suggest the use of an expansive or global discourse in describing the idea, thereby reinforcing the perception of the product's importance or wide scope.

Table N° 06: Best Distinctive Keywords for Successful Campaigns

Word	Score Difference
Wireless	0.013478
Worlds	0.012157
Smart	0.011922
Portable	0.011202
Camera	0.007141
Charging	0.006756
Earbuds	0.006627
Ebike	0.006505
Air	0.006391
Design	0.006328
Sound	0.006002
Bluetooth	0.005902
Audio	0.005896
Speaker	0.005543
Headphones	0.005497
Powerful	0.005348
Ultimate	0.005079
Smallest	0.004979
First	0.004974
Waterproof	0.004777

Source: Prepared by the researchers based on TF-IDF results

5.4.2 Successful keywords by year: The results of this table show a tracking of high frequency keywords used in successful campaigns between 2010 and 2020, revealing an evolution in the nature of marketing discourse and in the orientation toward projects with broad demand in addition to improvements in persuasion strategies and mechanisms within digital platforms. This evolution can be interpreted from a linguistic perspective, in terms of the marketing style employed, as well as from a behavioral perspective related to

backers, and it clearly reflects changing market priorities over time, this temporal change in keywords suggests that success in crowdfunding does not depend solely on product quality, but also on how the product is linguistically presented. successful campaigns increasingly rely on an innovative, technology-oriented language that highlights the product's practical features and employs discourse focused on performance and quality. It also reflects the transformation of crowdfunding platforms from spaces supporting arts and creative projects into major arenas for financing technological innovations and smart products, in line with the changing preferences of the global audience.

Table N° 07: Successful keywords by year

2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
repres	photographi	app	speaker	Audio	camera	world	game	design	design
face	miniatur	illeg	camera	Sound	world	game	world	power	power
gender	freeboot	tucker	Creat	Bike	sound	camera	power	world	world
equal	nois	help	photographi	Music	design	first	design	charg	new
today	artist	defens	Hear	Speaker	first	sound	wireless	wireless	portabl
	wayn	power	world	World	smart	power	first	first	comic
	form	frame	wireless	Wireless	time	design	sound	portabl	one
	postcard	amaz	Take	Help	nois	wireless	portabl	smart	smart
	fun	sound	hunter	Refuge	wireless	portabl	charg	usb	charg
	project	architectur	Fit	headphon	power	way	smart	use	high

Source: Prepared by the researchers based on data from the Indiegogo platform

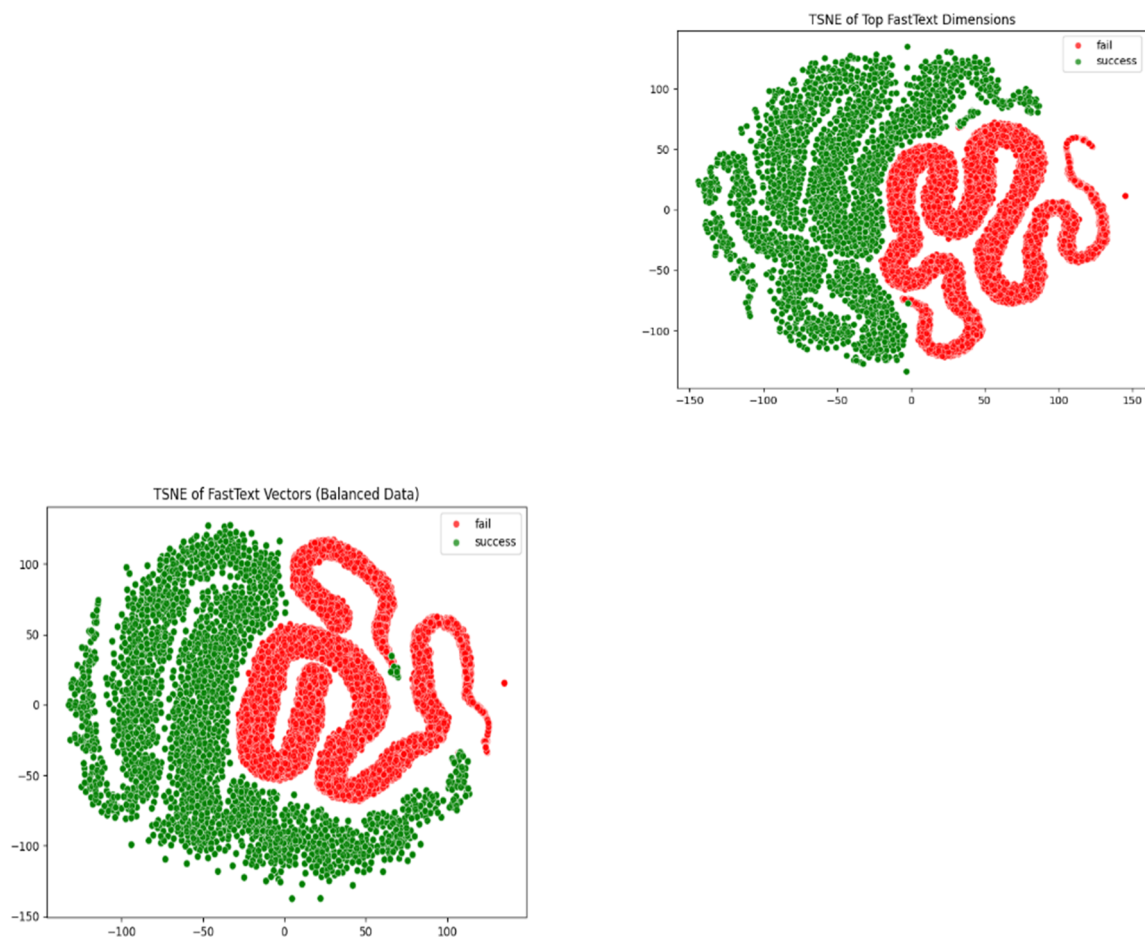
5.4.3 Successful keywords by country: The results in Table No. 08 show that the ability of startups to attract funding through digital platforms does not rely solely on the nature of the product or the size of the financial goal, but also extends to how the campaign message is formulated in a purposeful and appropriate way that aligns with the characteristics of each region, whether within the same country or across different markets. As illustrated in the table above, the terminology used differs between the United States and Sweden, both in terms of volume and diversity, which may be explained by the fact that audiences in Sweden have different preferences compared to those in other countries. Accordingly, it can be argued that certain countries and global markets tend to favor innovative technology driven campaigns, where terms related to design, for example, dominate their promotional content. In contrast, local markets tend to show greater interest in everyday aspects, safety, and personal experience, indicating that campaigns should adapt their language to the targeted audience.

Table N° 08: Shows the most frequently used words in successful campaigns by country

United states	Wester europe	Canada	Great britain	Australia	Hong kong	Sweden	switzerland	singapore	denmark	norway
---------------	---------------	--------	---------------	-----------	-----------	--------	-------------	-----------	---------	--------

The results of dimensionality analysis using the t-SNE technique for FastText vectors support the interpretation that successful campaigns employ more specialized and clear linguistic patterns compared to those that failed. This is evident from the distinct spatial distribution of points in the multidimensional space, where successful campaigns form cohesive clusters, reflecting the high homogeneity of the linguistic features used by their creators. In contrast, failed campaigns appear more scattered and random, indicating significant variation in linguistic styles and a lack of clarity in the message, making it harder for creators to reach digital funders. This distinction in density and clustering suggests that successful campaigns tend to adopt consistent and precise linguistic patterns, characterized by clarity and specialization in content formulation, whereas failed campaigns lack this linguistic coherence. The wide dispersion of failure points indicates the use of multiple, unconnected approaches, reflecting the absence of a clear linguistic strategy.

Figure N° 02: Visualization of the vector space of texts using t-SNE



5.7. Differences in Semantic Coherence and Linguistic Specialization between Successful and Failed Campaigns:

Table 10 shows the existence of significant differences between the texts produced in successful campaigns and those of failed campaigns. The key terms in our research indicate that successful campaigns are characterized by higher semantic coherence and closer connections between concepts, where the main terms are linked to actual content, specific functions, and technical or quantitative components, reflecting a greater degree of specialization and clarity in the discourse.

On the other hand, the semantic associations in failed campaigns reveal linguistic patterns dominated by repetition and morphological derivations of the same words without meaningful expansion, in addition to the appearance of words that are unrelated or weakly connected to the context targeted by the startup founders, indicating a lack of cohesive semantic structure and a weak ability to convey their idea clearly.

It can be concluded that successful campaigns rely on more precise language closely connected to the project's characteristics, whereas failed campaigns depend on general and ambiguous language. This supports the hypothesis that the strength of semantic coherence and the degree of linguistic specialization are influential factors in the success of crowdfunding campaigns.

Table N° 10: Semantic Coherence and Linguistic Specialization

Keyword	Success Similar Words	Fail Similar Words	Similarity
Project	product, crash, Gonzalez	projects, projector, projet	0.366
Campaign	collectors, parisian, drink	campaigns, campaigning, campaña	0.376
App	a7, az, duo	apps, apply, approach	0.377
Fund	funded, breed, tailored	funds, funded, fundraise	0.385
New	haven, oven, born	newly, newark, knew	0.404
Support	700wh, sport, 99	supports, supportive, supported	0.412
Help	helps, where, love	ihelp, helpful, helpline	0.441
Fail	few, golden, Parisian	fácil, faul, foil	0.448
Funding	coding, raising, pending	crowdfunding, founding, fundraising	0.456
Money	connect, online, conditioner	monetary, monkey, kidney	0.456

Source: Prepared by the researchers based on FastText model results

5.8. Semantic Differences and the Role of Function Words in Distinguishing Funding Campaigns:

In light of the semantic comparison of keywords between successful and failed campaigns presented in Table 11, it is evident that campaign success is associated with greater linguistic richness and diversity in semantic connections. Words in successful texts tend to be non-derivative and more contextual, as they are linked to functional concepts or attributes that enhance understanding of the project, reflecting a persuasive discourse based on rich narration and comprehensive description. In contrast, failed campaigns show a clear reliance on direct derivatives of the original word without semantic expansion, indicating linguistic poverty and

repetitive patterns that weaken the text’s ability to influence and convey the added value of the project. additionally, the stopword ratios reveal a more natural language in successful campaigns, reinforcing the informative and narrative style. thus, the analysis highlights that semantic breadth, contextual diversity, and avoidance of morphological repetition are factors that directly contribute to improving the appeal of the marketing message and increasing the likelihood of campaign success on crowdfunding platforms.

Table N° 11: Semantic Comparison of Keywords

Keyword	Successful Similar Words	Failed Similar Words	Successful	Failed
			Stopword Ratio	Stopword Ratio
Product	Stop, evil, pull, well, expert	Produkt, produce, prod, products	0.0000	0.0000
Support	Short, magnet, sport, dirt, shortcut	Supports, supportive, report, supporters	0.0000	0.0000
Campaign	Bamboo, clinic, pencil, paracord	Scampaign, campaigns, campaigner	0.0625	0.0000
New	Crew, away, penny, place	Knew, newly, newish	0.0625	0.0625
Help	Heavy, comics, may, way	Helpushelpthem, helpful, helps	0.2500	0.0000
Create	Ease, sure, sore, else	Recreate, creative, crate	0.0625	0.0000

Source: Prepared by the researchers based on FastText model results.

5.9. Analysis of Textual Cohesion and Linguistic Features in Crowdfunding Campaigns:

The in-depth linguistic analysis, as shown in Table 12, demonstrates that the success of crowdfunding campaigns is closely linked to the degree of textual cohesion and clarity of linguistic structure. Successful campaigns recorded a higher textual coherence rate by 142% compared to failed campaigns 0.6967 versus 0.2876, reflecting a logically consistent structure that makes it easier for the reader to follow the intellectual sequence and connect concepts. In contrast, the data reveal that failed campaigns tend to use longer words 6.96 letters versus 6.41 in successful ones, indicating a tendency toward verbal complexity at the expense of communicative clarity, while linguistic diversity is equal in both groups, reducing its role as a distinguishing factor. The core significance lies in semantic precision, where successful campaigns outperformed by 10.4% in conceptual specificity (0.9987 versus 0.9043), confirming that semantic clarity and thematic focus not formal diversity are the decisive factors in the effectiveness of fundraising discourse.

Table N° 12: Linguistic Characteristics and Semantic Indicators Differentiating Successful and Failed Funding Campaigns.

Metric	Failed Campaigns	Successful Campaigns	Difference	Analysis
Text Cohesion	0.2876	0.6967	0.4091	Successful texts are more cohesive, making it easier for supporters to understand the purpose of the campaign.

Enhancing Crowdfunding Success for Startups Based on Promotional Text Quality:

An Advanced Analysis Using Fast Text Models, indiegogo platform Example

Average Word Length	6.9600	6.4100	-0.5500	Failed texts use longer words, which may increase reading difficulty and distract attention.
Word Variety	1.0000	1.0000	0.0000	Linguistic diversity is equal, and it does not seem to be a decisive factor for success here.
Semantic Specificity	0.9043	0.9987	0.0944	Successful texts are more precise and specific in vocabulary, reflecting clarity of message and strong thematic focus.

Source: Prepared by the researchers based on FastText model results.

6. Discussion:

The results of this research paper show that the use of linguistic features in the text by fund seekers, including founders and entrepreneurs on Indiegogo, has contributed to explain the performance gap between successful and unsuccessful (failed) campaigns. This supports the hypothesis that crowdfunding campaigns with a clear and specialised language are more likely to succeed. Analysis of text representations revealed a homogenous linguistic pattern in successful campaigns compared to a greater dispersion in unsuccessful (failed) campaigns, which is consistent with the behavioural economics literature on the effect of linguistic consistency on increasing trust and reducing perceived risk among funders.

The results also showed that the successful linguistic and psychological characteristics explain the performance gap. Campaigns with higher lexical diversity and higher semantic richness are characterised by higher vocabulary diversity index values than campaigns with a failure rate. This suggests that rich text provides funders with more comprehensive information, has a more cognitive effect and is more likely to convince them that the project presented on the platforms is viable. The high degree of semantic diversity in the successful texts reflects the richness of the language and the specialisation, which is related to the quality of the content and the perceived professionalism. Interestingly, the study found that word diversity is not a distinguishing factor in itself; semantic accuracy is. This suggests that a clear and targeted use of specialised terms (e.g. "charge," "energy") is more effective than a broad but generic vocabulary. This finding aligns with previous studies emphasizing the role of well-crafted narratives and interactive language in increasing funding probabilities.

In addition, the TF-IDF keyword analysis revealed that in successful campaigns technical terms such as wireless and smart, which are normally associated with innovative products, were dominant. The language used in these campaigns also emphasises the practical features of the product and thus increases the credibility of the proposal. The results support the Crowdfunding persuasive communication Theory. The success of the campaign was not linked to the number of words, but rather to their semantic and functional quality. In addition, the development of success indicators over the years reflects the shift in crowdfunding from supporting creative and artistic projects to funding smart innovations, in line with global trends towards digital

consumption and technology products. More generally, the most strongly associated success terms differ from country to country, reflecting cultural and behavioural differences among global audiences. While technical terms are dominant in the US and the UK, the terms relating to security are more common in countries such as Denmark and Norway, underlining the need for campaigns to adapt to the characteristics of target groups rather than using a uniform language. Finally, a very important point should be made: the high degree of textual consistency in successful campaigns reinforces the idea that a coherent and consistent narrative activates the cognitive and emotional areas of supporters, which facilitates the building of trust and decision-making on funding.

Based on the above, it can be concluded that promotional discourse is not a secondary element in crowdfunding campaigns, but an effective element which shapes the perception of the funders and helps them to make their decisions. Investment in linguistic tools and advanced text analysis offer broad opportunities for improving digital marketing practices and business strategies. This is achieved by shifting the focus from the length of the text to the semantic quality of the text. This includes: (a) the use of precise and clear technical terms, (b) the construction of a coherent narrative linking the problem and the solution, (c) the balance of emotion and the avoidance of over-positive language, and (d) the tailor-made language for the target country.

7. Conclusion:

In conclusion, relying on advanced text representation models such as FastText allows for a deeper understanding of the hidden meanings in linguistic content. building accurate predictive models supports project owners in improving their funding opportunities, as successful campaigns rely on clear language, explicit narratives, and distinctive innovative vocabulary, which contributes to enhancing audience trust. furthermore the nature of success-indicating keywords changes over time and across countries, highlighting the importance of adopting flexible linguistic strategies that adapt to the target audience. This also opens avenues for future research to explore the role of cultural and psychological factors in shaping funders linguistic preferences across the world. Integrating promotional text analysis models provides effective tools for project owners and platforms focused on funding, emphasizing the crafting of advertising content. Additionally, their use offers a practical framework for developing decision support algorithms based on linguistic evidence rather than traditional indicators. Therefore, promotional discourse is no longer merely a description of the project but an integral part of the funding strategy, and improving text quality represents an investment in the project's success.

References:

1. Allison, Thomas H., Blakley C. Davis, Justin W. Webb, and Jeremy C. Short. 2017. 'Persuasion in Crowdfunding: An Elaboration Likelihood Model of Crowdfunding Performance'. *Journal of Business Venturing* 32(6):707–25. doi:10.1016/j.jbusvent.2017.09.002.

2. Babić, Karlo, Francesco Guerra, Sanda Martinčić-Ipšić, and Ana Meštrović. 2020. 'A Comparison of Approaches for Measuring the Semantic Similarity of Short Texts Based on Word Embeddings'. *Journal of Information and Organizational Sciences* 44(2):231–46. doi:10.31341/jios.44.2.2.
3. Caterina, Giannetti, and Mavillonio Maria Saveria. 2024. 'Crowdfunding Success: Human Insights vs Algorithmic Textual Extraction'. *E-Papers of the Department of Economics e Management – University Di Pisa*, 531.
4. Chuan, Jiang, Wang Dongbo, hen Si, Ye Wenhao, and Liu Jiangfeng. 2021. 'Building the Extraction Model of the Software Entities from Full Text of Research Articles Based on BERT'. *School of Economics & Management, Nanjing University of Science & Technology, Nanjing 210094, China*, 20.
5. Gunduz, Hakan. 2024. 'Comparative Analysis of BERT and FastText Representations on Crowdfunding Campaign Success Prediction'. *PeerJ Computer Science* 10:e2316. doi:10.7717/peerj-cs.2316.
6. kaggle. n.d. 'Kaggle'. Retrieved 15 November 2025. <https://www.kaggle.com/datasets/quentinmcteer/indiegogo-crowdfunding-data/resource=download>.
7. Kaminski, Jermain C., and Christian Hopp. 2020. 'Predicting Outcomes in Crowdfunding Campaigns with Textual, Visual, and Linguistic Signals'. *Small Business Economics* 55(3):627–49. doi:10.1007/s11187-019-00218-w.
8. Kolbe, Maura, Sasan Mansouri, and Paul P. Momtaz. 2022. 'Why Do Video Pitches Matter in Crowdfunding?' *Journal of Economics and Business* 122:106081. doi: 10.1016/j.jeconbus.2022.106081.
9. Mitra, Tanushree, and Eric Gilbert. 2014. 'The Language That Gets People to Give: Phrases That Predict Success on Kickstarter'. Pp. 49–61 in *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. Baltimore Maryland USA: ACM.
10. Mollick, Ethan. 2014. 'The Dynamics of Crowdfunding: An Exploratory Study'. *Journal of Business Venturing* 29(1):1–16. doi:10.1016/j.jbusvent.2013.06.005.
11. Parhankangas, Annaleena, and Maija Renko. 2017. 'Linguistic Style and Crowdfunding Success among Social and Commercial Entrepreneurs'. *Journal of Business Venturing* 32(2):215–36. doi:10.1016/j.jbusvent.2016.11.001.
12. Peng, Ling, Geng Cui, Ziru Bao, and Shuman Liu. 2022a. 'Speaking the Same Language: The Power of Words in Crowdfunding Success and Failure'. *Marketing Letters* 33(2):311–23. doi:10.1007/s11002-021-09595-3.
13. Peng, Ling, Geng Cui, Ziru Bao, and Shuman Liu. 2022b. 'Speaking the Same Language: The Power of Words in Crowdfunding Success and Failure'. *Marketing Letters* 33(2):311–23. doi:10.1007/s11002-021-09595-3.
14. Vivas-Peraza, Ana Cristina. 2025. 'Persuasive Dispositio and Hybridity in Digital Science Crowdfunding: The Generic Structure of Project Videos'. *ESP Today* 13(2):326–50. doi:10.18485/esptoday.2025.13.2.4.
15. Wang, Yuzhuo, and Chengzhi Zhang. 2018. 'Using Full-Text of Research Articles to Analyze Academic Impact of Algorithms'. Pp. 395–401 in *Transforming Digital Worlds*. Vol. 10766, *Lecture Notes in Computer Science*, edited by G. Chowdhury, J. McLeod, V. Gillet, and P. Willett. Cham: Springer International Publishing.
16. Yang, Sun. 2025. 'Text Classification Based on Machine Learning: A Review'. *Software College, Shenyang Normal University*, 110034 Shenyang, China, March, 7.