



Article

Crowdfunding as an E-Commerce Mechanism: A Deep Learning Approach to Predicting Success Using Reduced Generative AI Embeddings

Hakan Gunduz ^{1,*} , Muge Klein ² and Ela Sibel Bayrak Meydanoglu ²

¹ Faculty of Engineering, Software Engineering Department, Kocaeli University, Kocaeli 41001, Turkey

² Faculty of Economics and Administrative Sciences, Business Administration Department, Turkish-German University, Istanbul 34820, Turkey; klein@tau.edu.tr (M.K.); meydanoglu@tau.edu.tr (E.S.B.M.)

* Correspondence: hakan.gunduz@kocaeli.edu.tr

Abstract

Crowdfunding platforms like Kickstarter have reshaped early-stage financing by allowing entrepreneurs to connect directly with potential supporters. As a fast-expanding part of digital commerce, crowdfunding offers significant opportunities but also substantial risks for both entrepreneurs and platform operators, making predictive analytics an essential capability. Although crowdfunding shares some operational features with traditional e-commerce, its mix of financial uncertainty, emotionally charged storytelling, and fast-evolving social interactions makes it a distinct and more challenging forecasting problem. Accurately predicting campaign outcomes is especially difficult because of the high-dimensionality and diversity of the underlying textual and behavioral data. These factors highlight the need for scalable, intelligent data science methods that can jointly exploit structured and unstructured information. To address these issues, this study proposes a novel AI-based predictive framework that integrates a Convolutional Block Attention Module (CBAM)-enhanced symmetric autoencoder for compressing high-dimensional Generative AI (GenAI) BERT embeddings with meta-heuristic feature selection and advanced classification models. The framework systematically couples attention-driven feature compression with optimization techniques—Genetic Algorithm (GA), Jaya, and Artificial Rabbit Optimization (ARO)—and then applies Long Short-Term Memory (LSTM) and Gradient Boosting Machine (GBM) classifiers. Experiments on a large-scale Kickstarter dataset demonstrate that the proposed approach attains 77.8% accuracy while reducing feature dimensionality by more than 95%, surpassing standard baseline methods. In addition to its technical merits, the study yields practical insights for platform managers and campaign creators, enabling more informed choices in campaign design, promotional tactics, and backer targeting. Overall, this work illustrates how advanced AI methodologies can strengthen predictive analytics in digital commerce, thereby enhancing the strategic impact and long-term sustainability of crowdfunding ecosystems.



Academic Editor: Ercan Oztemel

Received: 30 November 2025

Revised: 24 December 2025

Accepted: 4 January 2026

Published: 8 January 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: e-commerce; crowdfunding success prediction; BERT embeddings; attention-based autoencoder (CBAM); meta-heuristic feature selection; LSTM and GBM classification

1. Introduction

Crowdfunding has emerged as a transformative mechanism for early-stage financing, allowing individuals and startups to raise capital directly from the public through online platforms. It can be regarded as a specialized branch of digital commerce that combines

traditional online retail practices with community-driven financing models. Platforms such as Kickstarter and Indiegogo operate as e-commerce channels when entrepreneurs raise funds by pre-selling products; however, unlike conventional e-commerce where products already exist, crowdfunding involves selling a promise, often for products still under development. In this setting, consumers effectively become co-investors in product development, blurring the traditional distinction between buyers and backers [1]. This expansion of the e-commerce paradigm allows entrepreneurs to secure advance commitments and small-scale investments from a large and distributed pool of supporters [2].

Since the 2008 Global Financial Crisis, crowdfunding has gained substantial traction as an alternative to traditional funding mechanisms such as venture capital and angel investment. Despite its growing popularity, a significant proportion of crowdfunding campaigns fail to reach their funding targets [3,4]. Campaign success is commonly driven by two interrelated factors: the number of backers and the average contribution per backer [5]. Maximizing these outcomes requires project initiators to attract broad attention while offering compelling value propositions and rewards [6]. Consequently, accurately predicting the likelihood of campaign success has become a critical challenge for improving strategic decision-making across the crowdfunding ecosystem.

Unlike traditional e-commerce forecasting, which primarily relies on structured historical data such as product categories, pricing information, and customer demographics [7], crowdfunding success prediction must account for heterogeneous and largely unstructured inputs. These include campaign narratives, visual media, and platform-generated interaction signals that reflect emotional appeal, trust, and perceived credibility [3,4]. The resulting data are high-dimensional and dynamic, making predictive modeling considerably more complex. This complexity highlights the need for artificial intelligence (AI)-driven, scalable decision-support systems capable of extracting actionable insights from rich textual content [8].

In practical deployment scenarios, many influential behavioral and interaction-based signals—such as funding velocity, update frequency, or social engagement—are unavailable or unreliable at very early-stages of a campaign. As a result, there is substantial value in predictive models that operate exclusively on information available at launch time, when creators seek early feedback and platforms conduct initial screening. In this context, the campaign blurb (short description) plays a central role, as it represents the primary medium through which creators communicate project intent, value propositions, and credibility to potential backers.

Recent advances in AI and natural language processing (NLP) have enabled increasingly effective analysis of campaign blurbs. Pre-trained language models such as BERT and FastText have demonstrated strong performance in capturing contextual semantics and sentiment, and have been widely adopted in crowdfunding analytics [9,10]. When combined with classifiers such as Long Short-Term Memory (LSTM) networks or Gradient Boosting Machines (GBM), these embeddings can yield competitive predictive performance. However, most existing approaches rely directly on high-dimensional embeddings, leading to substantial computational overhead, scalability limitations, and increased risk of overfitting.

To address these challenges, this study proposes a novel deep learning framework that integrates a CBAM-enhanced autoencoder to perform attention-guided semantic compression of high-dimensional BERT embeddings prior to feature selection and classification. By reducing redundancy while preserving salient linguistic information, the proposed autoencoder produces compact representations well-suited for early-stage prediction. To further refine these representations, meta-heuristic feature selection algorithms—specifically

Genetic Algorithm (GA), Jaya, and Artificial Rabbit Optimization (ARO)—are applied to identify discriminative and efficient feature subsets.

The proposed model enriches the existing literature through the following contributions:

- A comprehensive end-to-end framework is introduced, integrating contextual generative AI embeddings (BERT), attention-guided dimensionality reduction via a CBAM-enhanced autoencoder, meta-heuristic feature selection techniques (GA, Jaya, and ARO), and sequence-aware classification methods (LSTM and GBM). This architecture achieves a balance between scalability and predictive performance, enabling early-stage campaign evaluation in digital commerce settings.
- This study presents the first application of a CBAM-enhanced autoencoder to compress transformer-based embeddings in the context of crowdfunding, bridging attention-based semantic modeling and scalable feature optimization.
- A hybrid approach is developed that combines deep representation learning, meta-heuristic feature selection, and text-aware classification to improve both predictive accuracy and computational efficiency.
- An extensive evaluation is conducted on a large-scale Kickstarter dataset, demonstrating that the proposed framework achieves state-of-the-art performance with over 95% feature dimensionality reduction and substantially reduced computational cost, while remaining suitable for deployment in real-world platform scenarios.

By explicitly addressing the dimensionality bottleneck of large language models and focusing on early-stage, text-driven prediction, this work contributes both a methodological advance and a practical decision-support tool for enhancing the strategic value and sustainability of crowdfunding ecosystems.

2. Theoretical Background

Crowdfunding has evolved from a niche financial alternative into a globally significant mechanism for mobilizing collective capital. This transformation has been driven by advances in digital platforms, shifts in entrepreneurial culture, and a growing willingness among individuals to participate directly in funding innovation and social initiatives. Beyond its financial role, crowdfunding represents a structural change in how creators, backers, and platforms interact within digitally mediated ecosystems. Establishing a clear conceptual foundation is therefore essential for understanding campaign dynamics and for motivating the development of predictive models of campaign success.

2.1. Crowdfunding Ecosystem and Success Prediction

Crowdfunding, in its broadest definition, refers to a financing model through which entrepreneurs or project initiators seek relatively small contributions from a large pool of individuals via online platforms [3]. It has emerged as a response to structural limitations in traditional capital markets, offering a decentralized and accessible alternative to funding that reduces dependence on banks or institutional investors [11]. Within this ecosystem, funding initiatives, commonly referred to as campaigns, are time-bound efforts through which project creators seek support for a specific idea or goal [12,13]. These campaigns constitute the fundamental unit of participation on crowdfunding platforms and represent varying levels of opportunity and risk for potential backers.

When classified based on the underlying funding mechanism, rather than the type of return offered to contributors, crowdfunding campaigns typically adopt one of two primary models: the “keep-it-all” model or the “all-or-nothing” model. In the keep-it-all approach, creators retain the total amount pledged regardless of whether the funding target is achieved. In contrast, the all-or-nothing model only releases funds to the campaign if the predefined financial goal is met. Among contemporary crowdfunding platforms,

the all-or-nothing model has become the dominant structure due to its perceived capacity to improve trust and reduce risk for both creators and backers [14].

In addition to funding mechanisms, crowdfunding models can also be classified based on the type of return or benefit provided to contributors. This alternative classification distinguishes between four main categories: equity-based, credit-based, donation-based, and reward-based crowdfunding [15–17]:

- Equity-based crowdfunding enables small-scale investors to acquire ownership stakes in companies via online platforms, thereby gaining rights to future financial returns. This model serves as an alternative mechanism for corporate financing, particularly for firms that face barriers to accessing conventional funding sources, such as angel investors, public grants, or personal networks [15,16].
- Credit-based crowdfunding, also known as peer-to-peer lending (P2P), facilitates direct lending from individuals to borrowers without the involvement of financial intermediaries or collateral requirements. Although this model entails substantial risk, it seeks to minimize potential losses through the aggregation of numerous small and diversified investments. Subcategories such as peer-to-business and institution-to-peer lending have gained traction as institutional actors increasingly participate in this space [17].
- Donation-based crowdfunding is primarily philanthropic in nature and supports initiatives in fields such as the arts, scientific research, or personal and social causes. Contributors in this model do not expect any financial or material return; instead, their motivation is often rooted in altruism [6].
- Reward-based crowdfunding, by contrast, offers backers non-financial incentives, typically in the form of products or services, based on predetermined funding tiers. This model is especially prevalent in the creative industries and among early-stage ventures. Campaigns usually operate with fixed funding goals, and higher levels of support are rewarded with more substantial offerings. In some cases, surpassing the funding target allows additional features or “stretch goals” to be introduced [6].

The various crowdfunding models outlined above are operationalized through on-line platforms, which function as the digital infrastructure facilitating interaction between project initiators and prospective backers. These platforms serve as technological intermediaries connecting creators, such as entrepreneurs, innovators, or project managers, with a dispersed audience of potential supporters. Through these platforms, project creators can publicly showcase their funding proposals, while backers can easily navigate, evaluate, and contribute to campaigns that align with their interests. Advancements in digital technologies have significantly reduced barriers to entry for both parties, offering fast, user-friendly, and cost-effective access to crowdfunding ecosystems. In addition, these platforms streamline the campaign development process by providing tools and support for project presentation, communication, and general management of fundraising efforts [6].

In this context, crowdfunding platforms not only enable the technical execution of campaigns, but also shape the broader dynamics of participation and trust within the crowdfunding ecosystem. They also influence how information is structured, how user engagement is cultivated, and how financial transactions are facilitated - factors that collectively affect the likelihood of campaign success [18]. Beyond these operational roles, platforms also contribute to a broader transformation in the financing landscape. Crowdfunding reflects a significant paradigm shift in how innovation, creativity, and social initiatives are funded—one increasingly driven by the collective contributions of distributed online communities, enabled by digital technologies [19]. The availability of diverse crowdfunding models provides notable flexibility for both project creators and backers, while platforms act as essential intermediaries that coordinate and sustain these interactions.

As the global reach of crowdfunding continues to expand, a more nuanced understanding of its mechanisms, stakeholders, and outcomes becomes critical not only for entrepreneurs seeking alternative capital, but also for researchers studying evolving financial ecosystems and for policy makers working to foster inclusive and innovation-led growth [20].

As crowdfunding matures into a globally significant funding mechanism, understanding and predicting the factors driving campaign success becomes increasingly important, particularly for key stakeholders who depend on these outcomes for strategic decision-making. A crowdfunding campaign is generally considered successful when it achieves its funding target within the specified time frame [21]; however, empirical evidence shows that the success rates remain relatively low across most platforms [21,22]. In this context, the development of predictive models to estimate campaign outcomes has gained growing relevance. Accurate predictions can help optimize campaign strategies, allocate resources more efficiently, and enhance the overall functioning of the crowdfunding ecosystem. This ecosystem consists of three primary stakeholder groups: the founders (project creators or fundraisers), who initiate campaigns to secure financial support; the backers (investors or contributors), who provide funding; and the platforms, which serve as intermediaries facilitating communication, trust, and transaction flow between the other two parties [23]. For each of these actors, reliable success prediction models offer distinct benefits, whether by helping founders refine their proposals, guiding backers in choosing viable projects, or assisting platforms in curating high-potential campaigns. In this regard, the ability to accurately estimate the likelihood of success of a project plays a crucial role in enhancing decision-making throughout the ecosystem [24,25]. For founders, predictive insights can inform campaign design and timing strategies; for backers, they offer a way to minimize risk and allocate support more effectively; and for platforms, they enable more efficient selection and promotion of campaigns with higher success potential.

From the point of view of crowdfunding backers, the ability to predict campaign success is a critical tool for managing investment risk, particularly in the context of early-stage projects, where uncertainty is inherently high. Accurate evaluations of the success of a campaign enable sponsors to make more informed decisions about how they allocate both their financial resources and their personal attention [26]. Such predictive insights can help direct support to campaigns with stronger signals of viability, potentially encouraging not only higher individual contributions but also wider promotional efforts within the personal and professional networks of supporters [12]. By identifying projects with greater potential for success, backers, especially those with limited resources, can optimize their funding portfolios and avoid campaigns with low success probabilities, thereby reducing opportunity costs associated with failed investments [21,26]. Beyond individual decision-making, the presence of reliable predictive models can also enhance overall trust in the crowdfunding platform and the projects it hosts, strengthening user confidence and fostering sustained participation in the crowdfunding ecosystem.

From the perspective of project creators, predictive models serve as a valuable decision-support mechanism during the pre-launch phase of a crowdfunding campaign. Early indications that a project is likely to succeed allow creators to plan subsequent steps with greater confidence [27] and to focus their efforts on strategic elements that are empirically linked to better outcomes [28]. These models not only offer quantitative estimates of the probability of success, but also guide creators to allocate limited resources more effectively [29]. Importantly, predictive feedback can help refine key aspects of the campaign, particularly the project pitch, which plays a pivotal role in shaping the initial perceptions of supporters. Given that many potential supporters make quick judgments, crafting a compelling first impression is essential. Visual components such as a well-produced pitch video, along with a clear and engaging project description, can signal professionalism and

boost perceived quality. In addition, some creators may choose to conduct small-scale, time-bound test campaigns to assess real-time reactions. These early interactions can provide information on which elements may deter supporters and allow timely adjustments before full-scale launch [30]. By systematically identifying and addressing weaknesses in areas such as campaign messaging, pricing strategy, or reward structure, creators can improve the likelihood of reaching their funding goals [31]. In cases where predictive models suggest a low probability of success, creators can opt to delay or cancel the campaign, thereby avoiding unnecessary expenditure of time, effort, and financial resources. In contrast, a strong success prediction score can act as a positive signal to prospective backers, helping campaigns gain early traction and momentum [32].

For crowdfunding platforms such as Kickstarter or Indiegogo, success prediction plays a vital role in preserving platform credibility and user trust. A high rate of unsuccessful campaigns can reduce the confidence of the backers and damage the reputation of the platform [33]. Predictive models enable platforms to identify low-potential projects early in the process, allowing them to intervene proactively. These tools not only help detect campaigns with strong success potential, but also inform strategic decisions about which projects to highlight or promote. Using predictive insights, platforms can offer tailored recommendations to creators, such as adjusting funding goals, improving reward structures, or improving media content, to enhance campaign viability [32]. This not only contributes to higher-quality content for users, but also supports creators in optimizing their campaigns before launch. Integrating such predictive functionalities into the platform infrastructure allows for evidence-based guidance and more structured support within campaign guidelines. Ultimately, by embedding these insights into the design and operation of the platform, operators can improve overall success rates and reinforce the value proposition of the platform for all stakeholders [34].

2.2. Related Works

The rapid growth of crowdfunding platforms such as Kickstarter and Indiegogo has generated substantial research interest in predicting campaign success. Early computational studies relied primarily on meta-features, including campaign duration, funding goal, and creator attributes. For example, Greenberg et al. [35] employed Support Vector Machines (SVMs) and decision trees to model the likelihood of success based on simple metadata such as sentence length and project timelines. Although effective to some extent, these hand-crafted features did not capture the rich semantic and emotional content embedded in the campaign narratives.

Subsequent work incorporated textual and sentiment analysis to improve predictive power. Etter et al. [27] combined campaign time-series data with Twitter activity to forecast performance, while Buttice et al. [36] explored psychological factors, such as narcissism, as predictors of success. More recent studies have adopted deep learning and natural language processing (NLP) techniques to extract richer linguistic patterns from campaign descriptions, using word embeddings such as Word2Vec, GloVe, and FastText [37–39].

The advent of transformer-based models, particularly BERT [40], significantly advanced contextual text understanding in crowdfunding analytics. The applications of BERT to prediction of campaign success [41,42] have demonstrated its superiority over traditional embeddings. Gunduz [10] showed that BERT combined with LSTM yields higher predictive accuracy than FastText. However, most of these approaches use raw, high-dimensional embeddings directly for classification, leading to risks of overfitting, computational inefficiency, and reduced interpretability.

To address high dimensionality, feature selection (FS) techniques have been explored. Filter-based methods, such as Chi-Square, Mutual Information, and Fisher's score [43], are

computationally efficient but often overlook complex feature dependencies and contextual interactions. Wrapper-based methods, including Recursive Feature Elimination (RFE) [44], account for feature interactions but become computationally prohibitive as dimensionality grows. Meta-heuristic algorithms such as Genetic Algorithm (GA) [45,46], Jaya [47,48], Particle Swarm Optimization (PSO) [49,50], and Whale Optimization Algorithm (WOA) [51,52] provide a flexible compromise by efficiently exploring large combinatorial search spaces. Nevertheless, their application in crowdfunding analytics remains relatively limited. A notable example is Ryoba et al. [29], who applied WOA-based FS to Kickstarter data and achieved a 90.28% accuracy using a K-Nearest Neighbor classifier.

Despite these advances, a critical gap remains in how high-dimensional transformer-based embeddings are handled prior to feature selection. Most existing studies either apply FS directly to raw embeddings or rely on generic dimensionality-reduction techniques such as Pearson correlation, Mutual Information, or linear projections that are agnostic to contextual saliency [10]. Even when autoencoders are employed, they are typically used as purely reconstruction-driven compressors without explicit attention mechanisms, which limits their ability to preserve semantically important dimensions [53].

In contrast to variance-based projections (e.g., PCA, UMAP) or plain autoencoders, attention-guided compression explicitly models which regions of the embedding space are informative for the downstream task. By integrating Convolutional Block Attention Modules (CBAM) within an autoencoder, the proposed framework performs semantic compression that selectively emphasizes salient channel-wise and spatial patterns before optimization. This ordering, attention-based compression followed by meta-heuristic feature selection, fundamentally differs from prior workflows that treat dimensionality reduction and feature selection as independent or interchangeable steps.

In the following section, we present the proposed methodology in detail, outlining the architectural components, optimization strategies, and classification models employed in the framework.

3. Proposed Methodology

Building upon the gaps identified in the literature, particularly the lack of attention-based embedding compression combined with meta-heuristic feature selection for crowdfunding success prediction, we propose a multi-stage AI framework tailored to the digital commerce crowdfunding domain. The framework is designed to handle the semantic richness, high-dimensionality, and noisy nature of campaign blurbs, while producing computationally efficient predictions suitable for integration into platform-level decision-support systems.

As illustrated in Figure 1, the pipeline integrates contextual text representations from advanced NLP models, a novel autoencoder for deep embedding compression, meta-heuristic algorithms for dimensionality reduction, and two complementary classifiers: LSTM and GBM. The proposed framework consists of the following stages:

1. **Data Preprocessing:** Campaign blurbs are cleaned, normalized, and tokenized to ensure high-quality input for embedding models.
2. **Feature Extraction:** Contextual embeddings are generated using BERT to capture persuasive linguistic and stylistic patterns associated with backer decision-making.
3. **Deep Feature Compression:** A novel CBAM-enhanced autoencoder is introduced to compress high-dimensional BERT embeddings, retaining semantic richness while improving efficiency.
4. **Feature Selection:** Meta-heuristic optimization algorithms (GA, Jaya, and ARO) are employed to identify the most relevant subsets of compressed features, maximizing discrimination between successful and failed campaigns.

5. Classification: Reduced feature sets are used to train LSTM and GBM models, chosen for their ability to capture sequential linguistic dependencies and structured feature interactions, respectively.

Each component of the framework is elaborated in the following subsections, together with implementation details and theoretical motivations.

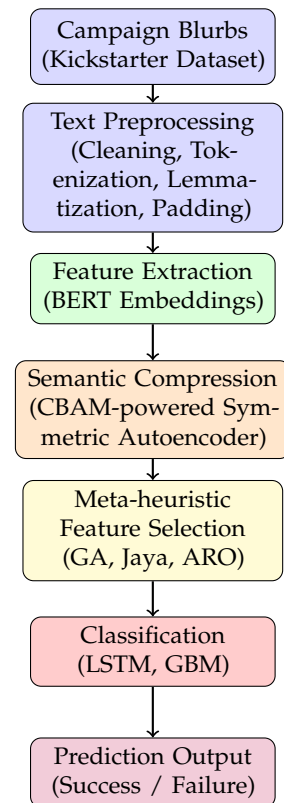


Figure 1. Proposed hybrid AI pipeline for crowdfunding success prediction.

3.1. Dataset Description

The dataset used in this study was sourced from a publicly accessible GitHub repository <https://github.com/rajatj9/Kickstarter-projects> (accessed on 1 July 2025), comprising data from 191,724 Kickstarter campaigns launched during 2018. Each record contains structured and unstructured features including project title, blurb, funding goal, campaign duration, launch and deadline dates, category, and the campaign's final status (successful or failed).

In our study, the primary predictive signal is extracted from the blurb field, which typically contains a short, natural-language description provided by the campaign creator. This aligns with the hypothesis that the wording and sentiment of the campaign blurbs play a vital role in influencing backers' decisions. The target variable is binary: campaigns are labeled as successful or failed.

Although collected in 2018, the data set remains a reliable benchmark in crowdfunding research due to its large size, public availability, completeness, and frequent usage in recent NLP studies [10,41]. From a business standpoint, it reflects the enduring linguistic and structural patterns of crowdfunding blurbs, making it relevant for designing prediction tools that generalize to current campaigns. Class imbalance is inherent, with approximately 40,000 more failed campaigns than successful ones, warranting attention during training and evaluation to ensure balanced learning and avoid bias toward the majority class.

3.2. Text Preprocessing Pipeline

Text preprocessing is a critical step in any NLP pipeline. Poorly pre-processed text can negatively impact embedding quality and introduce noise into learning. In crowdfunding platforms, inconsistent capitalization, creative spelling, and marketing-heavy phrasing are common; preprocessing ensures that these stylistic variations do not degrade model performance. In this work, the following preprocessing steps were implemented using Python's NLTK 3.9.1 and SpaCy 3.8.7 libraries:

1. **Cleaning and Normalization:** All campaign blurbs were stripped of punctuation, special symbols, digits, and URLs. Text was converted to lowercase to maintain uniformity and reduce vocabulary size.
2. **Tokenization:** Sentences were tokenized into word units using the SpaCy tokenizer, which handles contractions, punctuation, and common edge cases effectively.
3. **Stop Word Removal:** Words with minimal semantic value, such as 'and', 'the' and 'to', were removed using a standard stopword list.
4. **Lemmatization:** Each word was reduced to its base or dictionary form. For example, 'running' becomes 'run' and 'better' becomes 'good', allowing equivalent campaign blurbs to be matched despite superficial variation.
5. **Padding and Truncation:** Since neural models expect fixed-length inputs, each campaign description was padded (with special tokens) or truncated to 64 words. This length was chosen based on the distribution of blurb lengths, covering the majority without excessive padding.

These preprocessing steps ensure consistency in the input sequences, reduce the sparsity in the resulting vector representations, and normalize language patterns for fairer model comparisons.

3.3. Feature Extraction with Contextual Embeddings

To extract meaningful features from campaign blurbs, we employed the BERT embedding model for its respective strengths in capturing bidirectional semantic context. BERT is a deep contextual language representation model introduced by Devlin et al. [40], which has achieved state-of-the-art performance in a wide range of natural language understanding (NLU) tasks. Unlike conventional word embedding models such as Word2Vec or GloVe, which generate fixed word vectors independent of context, BERT produces context-sensitive embeddings by capturing bidirectional relationships between words in a sentence. This is achieved through the application of the Transformer architecture [54], which uses self-attention mechanisms to model dependency among all words in the input sequence, regardless of distance.

As shown in Figure 2, BERT is built entirely from transformer encoder layers, each consisting of two subcomponents:

1. **Multi-Head Self-Attention Mechanism:** This mechanism computes attention weights between all token pairs in a sentence, allowing each token to be contextualized with respect to every other token. The multi-head structure enables the model to capture information from different representation subspaces simultaneously.
2. **Position-Wise Feedforward Network:** Each token's attention output is passed through a fully connected feedforward network comprising two linear layers separated by a ReLU activation. This facilitates complex, non-linear transformations of contextual features.

The base BERT model used in this study comprises 12 encoder layers, each with 12 attention heads and 768 hidden units, resulting in rich and high-dimensional token-level outputs.

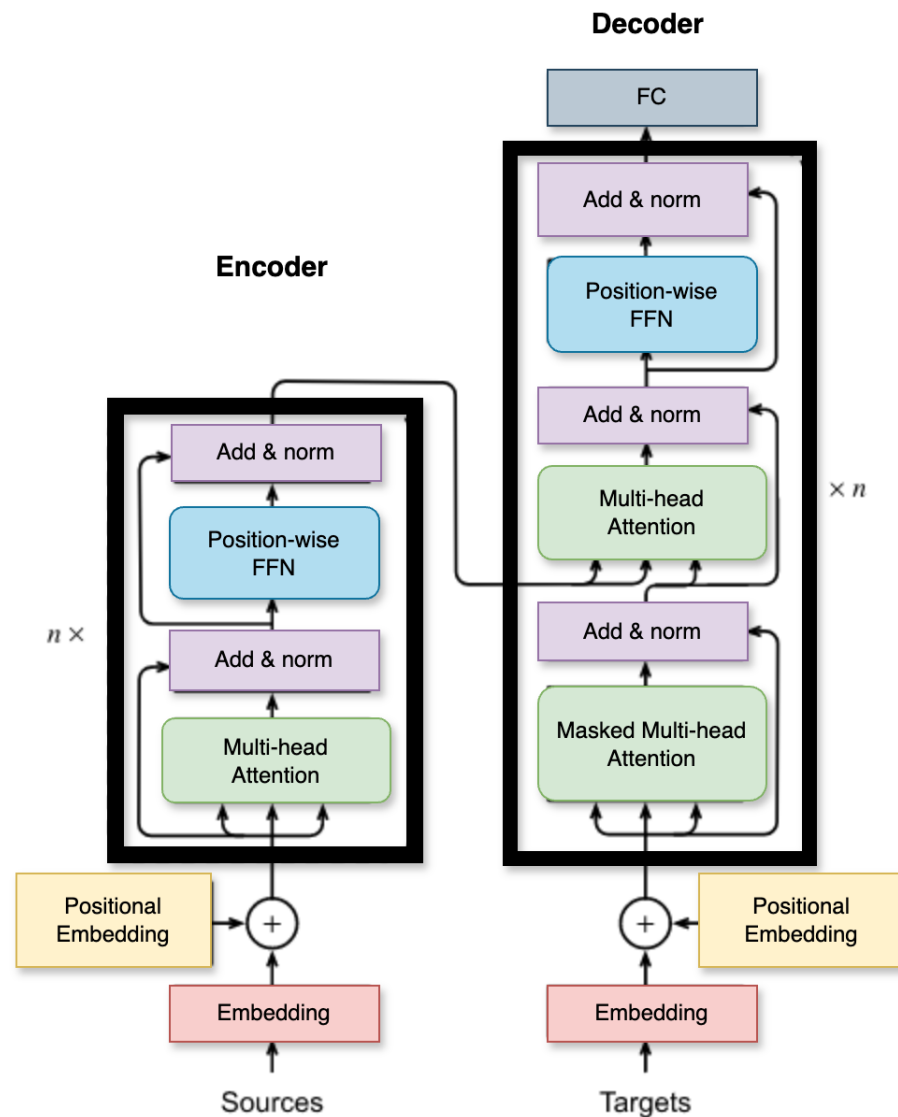


Figure 2. Architecture of the BERT transformer encoder.

In our framework, BERT was selected because of its demonstrated superiority in capturing contextually rich representations of text. Given that crowdfunding campaign success often hinges on subtle linguistic cues, and sentiment, BERT's ability to encode such nuances is particularly advantageous.

The following configuration was employed in our study:

- A pre-trained BERT-Base model from the Huggingface Transformers library was used as the base encoder. This model was initially trained on a large corpus comprising BooksCorpus and English Wikipedia, covering more than 3 billion words.
- Tokenization: Campaign blurbs were first tokenized using the BERT tokenizer, which decomposes words into subword units to effectively handle rare or compound words. The tokenized sequences were padded or truncated to a maximum length of 64 tokens.
- Fine-tuning Strategy: To adapt BERT to the domain of crowdfunding language, the model was fine-tuned using the campaign dataset. Specifically, the lower encoder layers were frozen to retain general linguistic knowledge, while the upper layers were updated to learn domain-specific semantics associated with campaign success narratives. A learning rate of 2×10^{-5} and a batch size of 32 were used during fine-tuning.

- **Embedding Extraction:** For each token in the input sequence, we extracted the hidden states from the last four encoder layers. These were concatenated to form a 3D tensor of shape $64 \times 768 \times 4$, where 64 corresponds to the standardized sequence length (post-padding), 768 is the hidden dimension, and 4 represents the depth of stacked layers. This rich embedding tensor preserves hierarchical contextual information across layers.
- **Compression and Use:** The extracted tensor was subsequently passed through our CBAM-enhanced autoencoder for dimensionality reduction. This step was crucial in minimizing computational costs and suppressing redundant features before applying feature selection and classification.

Using BERT's fine-tuned embeddings, our framework benefits from deep semantic encoding of campaign blurbs, capturing not only surface-level token information, but also underlying intent, sentiment, and characteristics for predicting campaign success.

3.4. Feature Selection with Hybrid Pipeline

The use of deep language models such as BERT for feature extraction introduces high-dimensional semantically rich representations while often include redundant or non-discriminative components. This dimensionality poses significant challenges, including increased computational cost, overfitting risks, and diminished model interpretability. These concerns are particularly acute in resource-constrained environments or real-time inference scenarios.

To address these limitations, we propose a hybrid feature selection pipeline built on top of our CBAM-enhanced autoencoder, which serves as a semantic compression layer for BERT embeddings. This attention-guided module effectively reduces the 3D contextual embeddings into dense latent representations by preserving salient spatial and channel-wise information. However, even after compression, the resulting feature vectors remain highly dimensional relative to traditional structured data, necessitating further refinement [53].

Accordingly, we investigate three complementary meta-heuristic algorithms for feature selection which perform global search to identify optimal feature subsets by maximizing classifier performance metrics such as F1-score or Matthews Correlation Coefficient (MCC).

By applying these selection strategies to the latent representations produced by the CBAM-encoded BERT features, we aim to construct a parsimonious yet highly expressive feature set. This hybrid strategy not only enhances the robustness of the model, but also significantly reduces training and inference costs. The following subsections detail the implementation and performance of each method in the context of crowdfunding campaign success prediction.

3.4.1. CBAM-Enhanced Autoencoder for Embedding Compression

To address the computational and overfitting challenges associated with high-dimensional contextual embeddings, we propose a novel CBAM-enhanced symmetric autoencoder that compresses the output of the last four hidden layers of the fine-tuned BERT model. This component represents a key innovation of the study, serving as a deep semantic feature compressor before applying feature selection and classification.

Each campaign blurb is first encoded using the fine-tuned BERT model, producing a tensor of size $64 \times 768 \times 4$, where 64 is the standardized sequence length, 768 is the embedding dimension, and 4 denotes the number of stacked BERT layers. This tensor is treated as a pseudoimage input to our convolutional autoencoder, which reduces the embedding dimensionality while preserving salient contextual information.

Convolutional Block Attention Module (CBAM)

The Convolutional Block Attention Module (CBAM), integrated into each convolutional layer of the autoencoder, is designed to refine feature representations by sequentially applying attention along two key dimensions: channel and spatial. CBAM enhances feature maps through a two-stage mechanism:

- **Channel Attention Module (CAM):** This sub-module emphasizes informative feature channels while suppressing less relevant ones. It does so by aggregating spatial information using both global average pooling and global max pooling, followed by a shared multi-layer perceptron (MLP). The resulting attention map is applied via element-wise multiplication to the input feature map, selectively enhancing discriminative channels.
- **Spatial Attention Module (SAM):** After refining channel-wise responses, the spatial attention module focuses on localizing salient regions in the spatial domain. It aggregates channel-wise information through average and max pooling operations along the channel axis and applies a convolutional layer to generate a 2D spatial attention map. This map is used to reweight spatial locations in the feature map.

Together, these attention mechanisms allow CBAM to adaptively highlight both “what” and “where” to emphasize in the feature tensor. This improves the network’s ability to learn compact and semantically enriched latent representations. The lightweight nature of CBAM enables its seamless integration into the encoder and decoder without significantly increasing computational overhead.

Encoder Architecture

The encoder consists of five sequential convolutional blocks, each comprising:

- A 2D convolutional layer with increasing filter sizes: 8, 16, 24, 32, and 64;
- Batch normalization and ReLU activation;
- A Convolutional Block Attention Module (CBAM) to apply both channel-wise and spatial attention.

Each block downsamples the input through stride-2 convolutions, progressively reducing spatial dimensions and generating a latent representation of shape $3 \times 24 \times 64$.

Decoder Architecture

The decoder mirrors the encoder with transposed convolutional layers, reversing the filter sizes ($64 \rightarrow 8$). Each upsampling layer restores spatial dimensions. The final output is reconstructed to match the original shape $64 \times 768 \times 4$. A sigmoid activation is applied at the output layer to constrain values to $[0, 1]$.

The autoencoder is trained in an unsupervised manner using Mean Squared Error (MSE) loss. During training, both encoder and decoder weights are optimized. After convergence, the decoder is discarded, and only the encoder is retained to extract low-dimensional latent features for the next stages of the pipeline. CBAM modules enhance the encoder’s ability to emphasize semantically important regions of the input by applying adaptive attention across feature channels and spatial locations. This makes compressed embeddings both compact and information rich. Unlike standard autoencoders or PCA, our design preserves deeper linguistic structures that are vital for downstream classification.

By reducing the input space for meta-heuristic selection and classification, the proposed framework improves training efficiency, reduces model complexity, and enhances generalization, as shown in the experimental results.

3.4.2. Meta-Heuristic Based Methods

- **Genetic Algorithm (GA):** The Genetic Algorithm (GA) [55,56] is a heuristic search approach inspired by the processes of natural selection and genetics, aiming to discover optimal solutions for complex issues. As a form of Evolutionary Algorithm, it uses essential mechanisms to evolve potential solutions over generations: The Selection operator picks individuals for reproduction using methods such as roulette wheel (selecting based on fitness proportion) or tournament selection (randomly selecting and choosing the best). The Crossover operator merges the traits of parents to create offspring, while the Mutation operator ensures diversity by introducing random variations. Elitism retains the best solutions across generations, and a fitness function assesses each solution against the objectives of the problem.
- **Jaya:** The Jaya Algorithm is a powerful optimization strategy introduced in [57] that operates without the need for algorithm-specific parameters such as crossover or mutation probabilities, mutation rates, or inertia weights. Its parameter-free design enhances robustness and simplifies implementation by removing the complications associated with parameter tuning prevalent in other optimization methods. This algorithm optimizes by artfully balancing exploration and exploitation, converging toward the optimal solution while steering clear of the worst option in each iteration. It functions through a cycle of iterative improvement:
 1. Initialization: The procedure starts by creating a diverse set of initial candidate solutions randomly spread across the search space. This diversity is essential for comprehensive exploration and avoiding early convergence to suboptimal solutions.
 2. Calculate Fitness: The fitness of each solution is assessed using the objective function. The algorithm identifies the best S_{best} and worst S_{worst} solutions in the current population, which guide the optimization path.
 3. Update Solutions: A key aspect of Jaya's strength is its update mechanism for solutions. Each candidate solution S_i^k in iteration k is updated according to Equation (1).

$$S_i^{k+1} = S_i^k + r_1 \cdot (S_{\text{best}} - |S_i^k|) - r_2 \cdot (S_{\text{worst}} - |S_i^k|) \quad (1)$$

where r_1 and r_2 are random numbers between 0 and 1. This update method is crafted to pursue two objectives simultaneously: advancing toward the best solution (S_{best}) while retreating from the worst one (S_{worst}). The introduction of randomness with r_1 and r_2 ensures stochasticity, which preserves diversity and helps avoid premature convergence.

4. Iterative Refinement: This process repeats until a stopping criterion, such as reaching a set number of iterations or a satisfactory fitness level, is satisfied. Throughout, the population's quality improves as more effective solutions emerge and inferior ones are discarded.

The algorithm's strength lies in its balanced exploration-exploitation dynamic and its parameter-free structure, making it particularly suited for robust feature selection applications.

- **Artificial Rabbit Optimization (ARO)** method, as presented by [58], imitates the survival strategies of rabbits by integrating exploration (via detour foraging) and exploitation (through random hiding) alongside a dynamic energy reduction mechanism. Its process updates are delineated as follows:
Exploration (Detour Foraging):

$$X_i^{t+1} = X_r^t + R(X_i^t - X_r^t) + C \cdot \mathcal{N}(0, 1) \quad (2)$$

Exploitation (Random Hiding):

$$X_i^{t+1} = X_i^t + R(r_4 \cdot b - X_i^t) \quad (3)$$

Energy Shrink Control:

$$A = 2 \log(1/r) \cdot \theta, \quad \theta = 2(1 - t/t_{\max}) \quad (4)$$

By adaptively balancing exploration and exploitation through energy dynamics, ARO is proficient in escaping local minima in complex search landscapes.

3.5. Classification

For an effective classification of crowdfunding project success, we utilized various machine learning algorithms recognized for their reliability, adaptability, and precision across different applications. Due to the high-dimensionality and complexity of the features derived from the project blurbs, machine learning models were chosen for their ability to handle noisy data, generate accurate predictions, and perform efficiently with high-dimensional feature sets [10]. The models selected for this research, such as Gradient Boosting Machines Forest (GBM) and Long Short-Term Memory (LSTM) networks, are popular in NLP classification tasks, with each providing distinct benefits for our dataset [59]. The next sections detail the advantages and technical details of each model and explain their suitability for predicting the success of crowdfunding initiatives.

3.5.1. Long Short-Term Memory (LSTM)

LSTM networks are a type of recurrent neural network (RNN) designed to capture long-term dependencies in sequential data. They are particularly well-suited for textual data, where contextual relationships between words play a critical role.

The LSTM model consists of input, forget, and output gates that regulate the flow of information. These gates help the network retain relevant information while discarding irrelevant details. LSTM networks were trained on campaign blurbs encoded by BERT embeddings. Dropout regularization was applied to prevent overfitting, and hyperparameter tuning was conducted using grid search to optimize parameters such as learning rate, batch size, and the number of LSTM units. The ability of LSTM to preserve sequential dependencies enables it to effectively model the contextual flow of campaign blurbs, making it ideal for textual classification tasks.

3.5.2. Gradient Boosting Machine (GBM)

GBM is an ensemble learning technique that builds a series of decision trees, where each tree corrects the errors of its predecessors. It is highly effective in handling structured data and is robust against overfitting when properly tuned.

GBM minimizes an objective function, such as classification error, by sequentially adding weak learners (decision trees) to the ensemble. The predictions of all trees are combined to make the final decision. GBM was trained on feature subsets selected by the proposed methods. Hyperparameters, such as the number of estimators, the learning rate, and the maximum depth of the tree, were optimized through grid search. This ensured that the model achieved high accuracy while maintaining generalization capabilities. GBM is robust in handling heterogeneous feature types and missing data. Its iterative nature allows it to effectively capture complex relationships between features and the target variable.

3.6. Performance Metrics

While accuracy is often suggested for evaluating performance, it fails to assess class discrimination. Alternatively, the Matthews Correlation Coefficient and F-Measure serve as a valuable metric for validating class-based model assessments.

Accuracy is defined as the ratio of correct predictions to the total number of cases. However, when the number of false positives (fp) significantly surpasses false negatives (fn), it becomes necessary to employ the F-measure for performance assessment. The F-measure computes the harmonic mean of precision and recall, thus considering both false positive and false negative cases to evaluate class separation.

The Matthews Correlation Coefficient (MCC) provides a balanced evaluation of binary classification performance by incorporating true positives (tp), true negatives (tn), false positives (fp), and false negatives (fn). Unlike accuracy, which may be overly optimistic under class imbalance, and unlike the F-measure, which primarily emphasizes the positive class, MCC captures the overall quality of predictions across both classes. It measures the correlation between predicted and actual labels, returning values between -1 and $+1$, where $+1$ indicates perfect prediction, 0 represents random guessing, and -1 corresponds to complete disagreement [10].

4. Experimental Results

All experiments were conducted in Python 3.11.3 using TensorFlow 2.4.1 and HuggingFace Transformers 4.5.1 on a system equipped with an Intel Ultra7–265K CPU, 64 GB RAM, and an NVIDIA GTX 5070 Ti GPU.

To assess the impact of dimensionality reduction and intelligent feature selection strategies on campaign success prediction, we conducted a series of experiments in three progressive phases. Each phase incrementally enhanced the input representation of the campaign blurbs and evaluated the impact using GBM and LSTM classifiers. Hyperparameters for the LSTM and GBM classifiers were tuned using grid search, with parameter ranges shown in Tables 1 and 2. The optimal configuration was selected based on 5-fold cross-validation performance.

Table 1. Parameters of LSTM network.

Parameters	Value
Number of memory cells	{64, 128, 256}
Dropout rate	{0.2, 0.3, 0.4}
Recurrent Dropout rate	{0.1, 0.2}
Optimizer	Adam
Activation function	{ReLU, ELU}
Number of epochs	50
Batch size	32

Table 2. Parameters of LightGBM.

Parameters	Value
Number of learners	{100, 200, 300, 400, 500, 1000}
Learning rate	{0.1, 0.01}
L2 regularizer	{0.001, 0.0001}
Maximum tree depth	{7, 9, 11}

To ensure reproducibility in the presence of stochastic optimization components, all experiments were executed with fixed random seeds at the levels of NumPy, Scikit, and the meta-heuristic algorithms. In addition, each meta-heuristic configuration (GA, Jaya, and ARO) was evaluated across multiple independent runs within each cross-validation fold. We observed high overlap in the selected feature subsets across runs, indicating stable convergence rather than random fluctuation.

Given the class imbalance in the Kickstarter dataset, we report not only accuracy but also **F1-score** and the **MCC**, the latter being especially informative because it accounts for all entries of the confusion matrix. All results are reported as mean $\pm$ standard deviation across folds. Where appropriate, paired tests and confidence intervals are used to compare configurations.

Throughout our evaluation, we also consider the practical implications for digital commerce: whether a model's prediction quality, dimensionality, and efficiency make it suitable for real-time platform curation, creator feedback systems, or backer personalization. Across the three phases of experimentation, our goal is to understand how different feature representations and selection strategies influence both predictive accuracy and operational feasibility.

4.1. Phase 1: Baseline Using BERT Embeddings (No Feature Selection)

As an initial baseline, each tokenized campaign blurb was processed with a fine-tuned BERT model. Token embeddings (768 dimensions) were averaged to produce a single vector representation per campaign, which was directly fed into the GBM and LSTM classifiers without any feature selection. Table 3 presents the results.

Table 3. Classification performance using BERT features without feature selection.

Metric	GBM (BERT)	LSTM (BERT)
Accuracy ($\pm$ Std)	0.735 $\pm$ 0.010	0.745 $\pm$ 0.026
F1-Score ($\pm$ Std)	0.791 $\pm$ 0.006	0.800 $\pm$ 0.016
MCC ($\pm$ Std)	0.435 $\pm$ 0.006	0.459 $\pm$ 0.015

These baselines confirm that textual blurbs alone capture meaningful predictive signals. However, the resulting 768-dimensional vectors remain computationally expensive for large-scale deployment, and while suitable for offline analytics, this representation is less appropriate for low-latency applications such as real-time campaign evaluation or creator-facing feedback tools.

4.2. Phase 2: Meta-Heuristic Feature Selection on Raw BERT Embeddings

To better exploit BERT's deep contextual structure, we extracted the final hidden states from the last four transformer layers for each token. Layerwise averaging produced four 768-dimensional vectors, which were concatenated into a **3072-dimensional** representation per campaign.

This high-dimensional representation formed the search space for three meta-heuristic optimizers: GA, Jaya, and ARO. Each algorithm was run for **100 epochs** using two objectives—F1-score and MCC—and feature selection was performed separately for GBM and LSTM classifiers. The results in Table 4 show that all optimizers improved MCC and accuracy relative to the Phase 1 baseline, with ARO generally offering the best trade-off between performance and compactness.

Table 4. Meta-heuristic selection results on raw BERT embeddings (4×768 search space, 100 epochs).

Model	Method	Opt.	Acc $\pm$ Std	F1-Score $\pm$ Std	MCC $\pm$ Std	# Features
GBM	GA	F1	0.728 $\pm$ 0.027	0.783 $\pm$ 0.016	0.425 $\pm$ 0.015	120
GBM	GA	MCC	0.741 $\pm$ 0.011	0.794 $\pm$ 0.007	0.450 $\pm$ 0.014	142
GBM	Jaya	F1	0.735 $\pm$ 0.023	0.791 $\pm$ 0.014	0.440 $\pm$ 0.013	202
GBM	Jaya	MCC	0.741 $\pm$ 0.018	0.794 $\pm$ 0.011	0.451 $\pm$ 0.012	251
GBM	ARO	F1	0.740 $\pm$ 0.028	0.799 $\pm$ 0.016	0.460 $\pm$ 0.015	101
GBM	ARO	MCC	0.748 $\pm$ 0.018	0.805 $\pm$ 0.011	0.470 $\pm$ 0.012	124
LSTM	GA	F1	0.743 $\pm$ 0.011	0.797 $\pm$ 0.007	0.465 $\pm$ 0.015	120
LSTM	GA	MCC	0.756 $\pm$ 0.025	0.807 $\pm$ 0.015	0.490 $\pm$ 0.014	142
LSTM	Jaya	F1	0.748 $\pm$ 0.015	0.802 $\pm$ 0.010	0.472 $\pm$ 0.013	202
LSTM	Jaya	MCC	0.758 $\pm$ 0.018	0.809 $\pm$ 0.011	0.495 $\pm$ 0.012	251
LSTM	ARO	F1	0.755 $\pm$ 0.025	0.808 $\pm$ 0.015	0.485 $\pm$ 0.015	101
LSTM	ARO	MCC	0.761 $\pm$ 0.015	0.812 $\pm$ 0.010	0.500 $\pm$ 0.012	124

Features denotes the number of selected embedding dimensions retained by the meta-heuristic feature selection method from the original 4×768 -dimensional BERT embedding space.

Compared to the baseline (Phase 1), the best configuration (LSTM + ARO, MCC objective) improves accuracy by +1.6% and MCC by +0.041, while reducing dimensionality by over 95%. These results demonstrate that meta-heuristic pruning effectively isolates the most discriminative textual cues and reduces computational cost, making the model more suitable for scalable deployment.

4.3. Phase 3: CBAM-Autoencoded BERT with Meta-Heuristic Feature Selection

To ensure a rigorous and reproducible evaluation protocol, all preprocessing and representation-learning steps were performed strictly within each training fold. For every cross-validation split, tokenization, BERT layer extraction, tensor stacking, normalization, autoencoder training, feature selection, and classifier fitting were learned exclusively from the training data. No vocabulary statistics, normalization coefficients, or feature-selection masks were transferred across folds, thereby eliminating any temporal or statistical information leakage.

In this phase, we developed a symmetric convolutional autoencoder augmented with Convolutional Block Attention Modules (CBAMs) to perform attention-guided semantic compression of contextual BERT embeddings. The CBAM-enhanced autoencoder was trained using the Adam optimizer with a learning rate of 1×10^{-4} and a batch size of 64. Training was conducted for a maximum of 100 epochs with early stopping based on validation reconstruction loss (patience = 10) to prevent overfitting. All convolutional and transposed convolutional layers employed 3×3 kernels with L_2 regularization ($\lambda = 1 \times 10^{-4}$) applied to kernel weights to improve generalization. Tables 5 and 6 present the detailed layer-wise architectures of the encoder and decoder components, respectively.

Table 5. Layer-wise structure of the CBAM-enhanced encoder.

Layer	Operation	Output Shape	Description
Input	–	$64 \times 768 \times 4$	Stacked BERT output embeddings
Conv Block 1	Conv2D (8 filters, stride 2) + CBAM	$32 \times 384 \times 8$	Attention-guided downsampling
Conv Block 2	Conv2D (16 filters, stride 2) + CBAM	$16 \times 192 \times 16$	
Conv Block 3	Conv2D (24 filters, stride 2) + CBAM	$8 \times 96 \times 24$	
Conv Block 4	Conv2D (32 filters, stride 2) + CBAM	$4 \times 48 \times 32$	
Conv Block 5	Conv2D (64 filters, stride 2) + CBAM	$2 \times 24 \times 64$	Final latent representation

Table 6. Layer-wise structure of the CBAM-enhanced decoder.

Layer	Operation	Output Shape	Description
Input	–	$2 \times 24 \times 64$	Latent compressed tensor
Deconv Block 1	Transposed Conv2D (32 filters, stride 2)	$4 \times 48 \times 32$	Upsampling
Deconv Block 2	Transposed Conv2D (24 filters, stride 2)	$8 \times 96 \times 24$	
Deconv Block 3	Transposed Conv2D (16 filters, stride 2)	$16 \times 192 \times 16$	Reconstructed embeddings
Deconv Block 4	Transposed Conv2D (8 filters, stride 2)	$32 \times 384 \times 8$	
Deconv Block 5	Transposed Conv2D (4 filters, stride 2) + Sigmoid	$64 \times 768 \times 4$	

Proposed autoencoder reduces the original 196,608-dimensional input ($64 \times 768 \times 4$) to a compact 3072-dimensional latent representation ($2 \times 24 \times 64$). This compact space becomes the new search domain for meta-heuristic feature selection using GA, Jaya, and ARO.

Table 7 summarizes the classification performance obtained from the compressed representations. Across both GBM and LSTM classifiers, ARO optimized for MCC consistently yields the best trade-off between predictive performance and compactness. The best configuration (CBAM + ARO + LSTM) achieves **77.8% accuracy and 0.515 MCC** using only **117 features**, corresponding to a **99.94% reduction** relative to the original BERT embedding space while simultaneously improving predictive performance.

Table 7. Meta-heuristic results on CBAM-compressed BERT embeddings ($2 \times 24 \times 64$ latent space, 100 epochs).

Model	Method	Opt.	Acc $\pm$ Std	F1-Score $\pm$ Std	MCC $\pm$ Std	# Features
GBM	GA	F1	0.745 ± 0.018	0.800 ± 0.012	0.460 ± 0.011	102
GBM	GA	MCC	0.758 ± 0.022	0.809 ± 0.015	0.485 ± 0.014	122
GBM	Jaya	F1	0.736 ± 0.013	0.792 ± 0.010	0.445 ± 0.009	274
GBM	Jaya	MCC	0.752 ± 0.023	0.804 ± 0.015	0.470 ± 0.013	311
GBM	ARO	F1	0.754 ± 0.016	0.808 ± 0.011	0.475 ± 0.010	108
GBM	ARO	MCC	0.762 ± 0.022	0.812 ± 0.015	0.490 ± 0.013	117
LSTM	GA	F1	0.760 ± 0.025	0.808 ± 0.017	0.485 ± 0.015	102
LSTM	GA	MCC	0.771 ± 0.012	0.816 ± 0.009	0.505 ± 0.009	122
LSTM	Jaya	F1	0.766 ± 0.014	0.812 ± 0.010	0.495 ± 0.009	274
LSTM	Jaya	MCC	0.762 ± 0.009	0.809 ± 0.007	0.490 ± 0.007	311
LSTM	ARO	F1	0.771 ± 0.020	0.816 ± 0.014	0.500 ± 0.012	108
LSTM	ARO	MCC	0.778 ± 0.019	0.821 ± 0.013	0.515 ± 0.011	117

Features denotes the number of selected embedding dimensions retained by the meta-heuristic feature selection method from the CBAM-compressed BERT embeddings space.

4.4. Time-Aware Validation

A common source of over-optimism in crowdfunding prediction studies arises from evaluating models using random cross-validation on temporally ordered data. Such protocols may inadvertently allow future linguistic patterns to influence training, thereby inflating performance estimates. To assess temporal robustness under realistic deployment conditions, we conducted an additional time-aware evaluation using a rolling forecasting-origin scheme.

Specifically, the best-performing configuration (CBAM + ARO + LSTM) was trained on campaign descriptions from the first four months of the dataset and evaluated on campaigns launched in the subsequent month. This process was repeated across all contiguous five-month windows within the 2018 Kickstarter dataset. Importantly, only blurbs available at launch time were used; no post-launch signals such as funding velocity, update frequency, or user interactions were included at any stage of training or evaluation.

The results demonstrate that the model generalizes well under realistic temporal drift conditions. Across all rolling folds, the model achieved **an accuracy of 0.756, an F1-**

score of 0.812, and an MCC of 0.465. These scores are only marginally lower than those obtained via random 5-fold cross-validation (0.778 accuracy and 0.515 MCC), indicating that predictive performance does not rely on temporal leakage and remains robust when trained exclusively on past data to forecast future campaigns; while this evaluation does not explicitly include post-2018 data, the stability observed across rolling temporal splits suggests that the linguistic signals captured by the model are not tied to a single snapshot in time, but rather reflect broader narrative patterns common to crowdfunding campaigns.

This time-aware evaluation provides two important insights. First, linguistic and narrative cues embedded in campaign blurbs appear sufficiently stable over short time horizons to support early-stage prediction. Second, the modest performance degradation relative to random cross-validation is consistent with natural shifts in campaign style and platform dynamics, suggesting that lightweight periodic retraining (e.g., weekly or monthly) would be sufficient to maintain accuracy in operational settings. Taken together, these findings confirm that the proposed framework is well-suited for real-world deployment scenarios where predictions must be made before or immediately after campaign launch, using textual information alone. Nonetheless, extending this evaluation to campaigns launched in later years and under evolving platform dynamics remains an important direction for future work, particularly to assess the impact of changing creator behaviors, platform policies, and language trends.

4.5. Cost–Benefit and Deployment Analysis

To evaluate scalability and deployment feasibility, we compared our final pipeline against widely used transformer baselines.

Parameter Efficiency. Fine-tuned BERT contains ~110 M parameters, and DistilBERT contains ~66 M. Our CBAM autoencoder has fewer than 3 M parameters and the downstream 117-dimensional LSTM classifier adds fewer than 0.2 M. Thus, the full pipeline is over 30× smaller than DistilBERT and over 35× smaller than BERT.

Inference Latency. On CPU, fine-tuned BERT requires 20–30 ms per sample; DistilBERT requires 10–12 ms. LSTM using full BERT embeddings requires ~15–20 ms. In contrast, the compressed + ARO-selected 117-dimensional representation enables sub-millisecond inference (<1 ms), providing a 30× speedup over BERT.

Memory Footprint. Fine-tuned BERT occupies ~420 MB; DistilBERT ~265 MB; LSTM baselines 50–100 MB. The 117-dimensional LSTM classifier occupies less than 10 MB.

In summary, the attention-guided compression and meta-heuristic selection pipeline achieves best-in-class accuracy while offering major improvements in latency, memory, and parameter footprint. These characteristics make it highly suitable for real-time deployment in platform-scale settings such as success-aware recommendation engines, automated campaign evaluation dashboards, or creator-support interfaces.

5. Discussion

This study introduced a hybrid framework that integrates fine-tuned BERT embeddings, a CBAM-enhanced convolutional autoencoder, and meta-heuristic feature selection to predict the success of crowdfunding campaigns. By explicitly targeting the challenges posed by high-dimensional textual representations, the framework advances predictive modeling in digital commerce while remaining computationally efficient enough for platform-scale deployment. Empirically, semantic compression combined with biologically inspired optimization yields consistent gains in accuracy, F1-score, and MCC, while simultaneously shrinking the feature space by more than 95% relative to raw BERT embeddings.

Building on these results, it is important to position the contribution of the proposed approach within the broader literature on crowdfunding prediction and e-commerce analytics,

and to understand how each architectural component shapes performance. The following subsections first examine ablation findings, then compare the proposed framework to internal and external baselines, and finally discuss practical, theoretical, and methodological implications alongside key limitations and future research directions.

5.1. Ablation Studies

To isolate the contribution of each architectural component, we conducted a set of controlled ablation experiments using the LSTM classifier for consistent comparison. The ablations examine four factors: (i) the role of CBAM attention within the autoencoder, (ii) the benefit of learned compression versus classical dimensionality reduction, (iii) the impact of meta-heuristic feature selection, and (iv) the effect of different optimization objectives for ARO under class imbalance.

The first ablation (Table 8) shows that replacing CBAM modules with linear dense layers reduces both accuracy and MCC while keeping the number of selected features fixed. This indicates that attention-guided compression produces more discriminative latent representations than a purely convolutional encoder, and that CBAM is not merely redundant with subsequent feature selection.

Table 8. Effect of CBAM attention in the autoencoder.

Model	Accuracy	MCC	# Features
AE (No-CBAM) + ARO (MCC)	0.756 ± 0.018	0.489 ± 0.014	117
CBAM-AE + ARO (MCC)	0.778 ± 0.019	0.515 ± 0.011	117

Features denotes the number of selected embedding dimensions retained by ARO from the AE and CBAM-AE compressed BERT embeddings.

Comparing CBAM-AE to Principal Component Analysis (PCA) and Uniform Manifold Approximation and Projection (UMAP) (Table 9) highlights the advantage of task-adaptive compression over generic projections. Even when the latent dimensionality is matched (3072 features), the autoencoder yields markedly higher accuracy and MCC, suggesting that it preserves contextual and sequential cues that are critical for crowdfunding success prediction.

Table 9. Autoencoder vs. PCA vs. UMAP dimensionality reduction.

Method	Accuracy	MCC	# Features
PCA (3072 dims) + ARO	0.742 ± 0.021	0.461 ± 0.017	118
UMAP (3072 dims) + ARO	0.751 ± 0.016	0.472 ± 0.013	120
CBAM-AE (3072 dims) + ARO	0.778 ± 0.019	0.515 ± 0.011	117

Features denotes the number of retained dimensions after dimensionality reduction and subsequent ARO-based feature selection, starting from the original 3072-dimensional embedding space.

The effect of meta-heuristic feature selection is equally clear (Table 10). Applying ARO on top of the CBAM-compressed embeddings improves accuracy by approximately +2.6 percentage points and MCC by +0.039, while reducing the dimensionality from 3072 to 117 features (around 97% reduction). This confirms that optimization-driven refinement is crucial for both performance and compactness, rather than acting only as a computational convenience.

Table 10. Effect of meta-heuristic feature selection.

Model	Accuracy	MCC	# Features
CBAM-AE (No FS)	0.752 ± 0.017	0.476 ± 0.012	3072
CBAM-AE + ARO (MCC)	0.778 ± 0.019	0.515 ± 0.011	117

Features denotes the number of embedding dimensions used by the model; 3072 corresponds to the full CBAM-AE output without feature selection, while the reduced value indicates the subset selected by the ARO-based meta-heuristic feature selection.

Finally, the comparison of ARO objectives (Table 11) shows that optimizing for MCC yields the strongest overall discrimination under class imbalance, even at the cost of a slightly larger feature subset. This supports the choice of MCC as a primary objective and evaluation metric in imbalanced crowdfunding settings.

Table 11. Comparison of ARO optimization objectives.

Objective	Accuracy	MCC	# Features
ARO (F1 objective)	0.771 ± 0.020	0.500 ± 0.012	108
ARO (MCC objective)	0.778 ± 0.019	0.515 ± 0.011	117

Features denotes the number of embedding dimensions selected by the ARO algorithm under the specified optimization objective (F1 or MCC) from the original embedding space.

Taken together, the ablation results demonstrate that each component of the CBAM + ARO + LSTM pipeline contributes tangible and complementary benefits: CBAM improves semantic selectivity; learned compression outperforms PCA and UMAP; meta-heuristic selection enhances both accuracy and compactness; and MCC-based optimization better aligns the search process with the underlying success/failure imbalance.

5.2. Significance in Relation to Prior Work

To contextualize the contribution of the proposed framework, it is useful to compare its performance against both internal baselines and prior studies in the literature. Classical text-mining pipelines based on TF-IDF and linear or tree-based classifiers provide a strong yet lightweight reference point. As summarized in Table 12, TF-IDF + SVM achieves an accuracy of 0.690, outperforming TF-IDF + Logistic Regression (0.669) and TF-IDF + GBM (0.673). These results are broadly consistent with earlier crowdfunding research that relies on shallow lexical features and highlight the predictive power of simple text statistics.

Table 12. Internal baselines evaluated in this study using 5-fold cross-validation (TF-IDF models) and held-out validation (fine-tuned transformers).

Model	Accuracy	F1-Score	MCC
TF-IDF + SVM	0.690 ± 0.014	0.765 ± 0.008	0.328 ± 0.030
TF-IDF + Logistic Regression	0.669 ± 0.009	0.754 ± 0.017	0.275 ± 0.021
TF-IDF + GBM	0.673 ± 0.014	0.759 ± 0.016	0.284 ± 0.024
Fine-tuned BERT (end-to-end)	0.722 ± 0.015	0.777 ± 0.014	0.408 ± 0.022
Fine-tuned DistilBERT (end-to-end)	0.733 ± 0.011	0.789 ± 0.015	0.428 ± 0.019
This study (CBAM + ARO + LSTM)	0.778 ± 0.019	0.821 ± 0.013	0.515 ± 0.011

Comparison with Transformer Fine-Tuning Baselines

To establish a strong modern baseline, we also evaluate standard transformer fine-tuning configurations commonly used in NLP classification tasks. Specifically, we fine-tune pre-trained BERT and DistilBERT models in an end-to-end manner using a lightweight linear classification head attached to the pooled [CLS] representation. These architectures

represent widely adopted and canonical transformer fine-tuning pipelines in contemporary NLP classification tasks.

Under this setup, the fine-tuned BERT model achieves an accuracy of 0.722, while fine-tuned DistilBERT reaches 0.733, with corresponding improvements in F1-score and MCC. These results confirm the strong representational capacity of transformer models when trained end-to-end on crowdfunding text. DistilBERT, in particular, offers an attractive trade-off between performance and efficiency due to its reduced parameter count and lower inference latency.

Against this backdrop, the proposed CBAM-enhanced autoencoder with ARO-based feature selection and LSTM classification offers a statistically meaningful improvement: 0.778 accuracy, 0.821 F1-score, and 0.515 MCC using only 117 features. The paired *t*-test comparisons in Table 13 confirm that these gains over alternative CBAM-based variants are significant and attributable to the combination of attention-guided compression and imbalance-aware optimization.

Table 13. Paired *t*-test comparison of models against the best performer (LSTM–ARO) using 5-fold cross-validation. Mean differences (Δ) denote LSTM–ARO minus the compared model. 95% confidence intervals are based on the Student’s *t*-distribution ($df = 4$).

Compared Model	Δ Mean	95% CI of Δ	<i>p</i> (<i>t</i> -Test)
GBM–GA	+0.033	[+0.017, +0.049]	0.002
GBM–Jaya	+0.045	[+0.028, +0.061]	0.001
GBM–ARO	+0.022	[+0.007, +0.038]	0.013
LSTM–GA	+0.015	[+0.002, +0.028]	0.037
LSTM–Jaya	+0.012	[−0.001, +0.025]	0.073

Bold values indicate statistically significant differences ($p < 0.05$) in the paired *t*-test when compared against the best-performing model (LSTM–ARO).

To ensure statistical rigor, all reported results are presented as mean $\pm$ standard deviation across cross-validation folds. In addition to descriptive statistics, paired *t*-tests with 95% confidence intervals were conducted to assess whether observed performance differences are statistically significant rather than attributable to random variation. This analysis confirms that the performance gains achieved by the proposed CBAM + ARO + LSTM configuration over competing baselines are statistically meaningful under standard significance thresholds.

Turning to external work, prior studies on crowdfunding prediction have predominantly relied on handcrafted features, traditional embeddings, or multimodal deep architectures. Table 14 summarizes representative results.

Table 14. Comparative accuracy with recent studies in the domain.

Study	Accuracy
Lee et al. (2018) [60]	0.765
Kaminski et al. (2020) [61]	0.710
Tang et al. (2022) [62]	0.722
Gunduz et al. (2024) [10]	0.745
This study (CBAM + ARO + LSTM)	0.778

Lee et al. [60] obtained 0.765 accuracy using GloVe embeddings and Seq2Seq models, underscoring the importance of sequential linguistic cues. Kaminski et al. [61] and Tang et al. [62] incorporated multimodal signals via deep CNNs and cross-attention, but reported lower accuracies (0.710–0.722) with considerably higher computational overhead.

Our previous work [10] reported 0.745 accuracy for BERT + LSTM, highlighting the benefits of contextual embeddings over shallow features.

Compared with these studies, the current framework offers a balanced contribution: it delivers competitive or superior accuracy while drastically reducing the effective feature space. Rather than claiming universal state-of-the-art performance across all platforms and modalities, we position the proposed model as a strong, efficient, and scalable text-based benchmark that advances methodological practice through its integration of attention-guided compression and meta-heuristic feature selection.

5.3. Explainability and Actionable Insights

For predictive models deployed in real-world crowdfunding platforms, performance alone is insufficient; stakeholders require interpretable and actionable insights that explain not only whether a campaign is likely to succeed, but also why such an outcome is predicted and how it may be improved. Although the proposed framework is optimized for predictive accuracy and scalability, it incorporates multiple mechanisms that support explainability at different stages of the pipeline.

First, the Convolutional Block Attention Module (CBAM) embedded within the autoencoder provides intrinsic interpretability by assigning adaptive attention weights across both channel-wise (semantic dimensions) and spatial (token-layer interactions) representations. These attention maps implicitly highlight which semantic regions of the transformer embeddings contribute most strongly to the compressed representation, enabling qualitative inspection of salient linguistic patterns.

Second, the meta-heuristic feature selection stage produces sparse and compact feature subsets, which enhances interpretability by narrowing the decision process to a limited number of highly informative dimensions. Across cross-validation folds, we observe consistent convergence toward similar subsets, suggesting that the selected features capture stable and discriminative textual cues rather than random artifacts.

Third, the downstream classifiers (LSTM and GBM) support post hoc explainability through established techniques such as SHAP or LIME. When applied to the reduced feature space, these tools can be used to trace predictions back to influential embedding dimensions and, by extension, to underlying phrases or stylistic patterns in campaign blurbs. This enables actionable feedback, such as identifying overly vague language, lack of clarity in value propositions, or insufficient emphasis on product uniqueness.

From a practical standpoint, these explainability pathways allow platform operators to provide creator-facing recommendations (e.g., improving narrative clarity or emphasizing concrete deliverables), assist backers in understanding the rationale behind risk assessments, and support transparent decision-making processes within platform governance. As such, the proposed framework balances predictive power with interpretability, aligning technical performance with the practical requirements of digital commerce ecosystems.

5.4. Practical Implications

The findings of this study have several concrete implications for stakeholders in the crowdfunding ecosystem and the broader digital commerce domain.

For **platform operators**, the proposed framework enables scalable, early-stage screening of campaigns using blurbs available at launch. By combining high predictive accuracy with extreme feature compression, the model can be integrated into real-time moderation, recommendation, or curation pipelines without incurring prohibitive computational cost. This capability supports improved campaign visibility allocation, risk-aware promotion strategies, and more balanced platform governance.

For **campaign creators**, predictive scores and language-sensitive feedback can serve as a decision-support tool during the campaign design phase. By identifying linguistic patterns associated with successful outcomes, creators can iteratively refine project descriptions prior to launch, thereby reducing the likelihood of failure and improving funding efficiency.

For **investors and backers**, AI-driven success predictions offer an additional signal for risk assessment in environments characterized by high uncertainty and information asymmetry; while not intended to replace human judgment, such tools can complement existing heuristics and increase transparency in decision-making.

From a **deployment perspective**, the attention-guided compression and meta-heuristic feature selection strategy substantially reduces memory and inference requirements, making the framework suitable for real-world implementation at platform scale. Beyond crowdfunding, the proposed approach generalizes to other digital commerce scenarios involving textual persuasion and early-stage forecasting, such as product launch evaluation, content-driven marketing campaigns, and success-aware recommendation systems.

Overall, these findings demonstrate that carefully designed AI architectures can deliver not only predictive improvements but also operational and strategic value in data-intensive digital commerce environments.

5.5. Theoretical Implications

Beyond its practical and strategic relevance, this study also advances theoretical understanding in the domains of artificial intelligence and digital commerce. First, it demonstrates that integrating attention-based embedding compression with meta-heuristic feature selection offers a viable approach to addressing the persistent challenge of high-dimensionality in text-based predictive modeling. This contributes to the broader literature on representation learning by showing how semantic compression can be systematically aligned with optimization-driven feature refinement.

Second, the findings reinforce the importance of sequence-aware models in capturing narrative and linguistic structures that drive consumer and investor behavior, extending prior work on the role of textual cues in digital decision-making. Third, by framing crowdfunding as a specialized form of e-commerce—blending financial uncertainty with emotionally charged narratives—this study advances theoretical perspectives on crowdfunding as both a commerce and investment ecosystem.

Finally, the work underscores the need to link predictive modeling approaches with theories of trust, persuasion, and platform governance, opening pathways for interdisciplinary research at the intersection of computer science, information systems, and entrepreneurship studies.

5.6. Limitations

Despite the strong empirical and computational results, several limitations should be acknowledged when interpreting the findings.

First, the dataset is restricted to Kickstarter campaigns from a single year (2018), which constrains long-term temporal generalizability. Although a rolling time-aware validation was conducted to mitigate this concern, shifts in creator behavior, platform policies, or backer preferences over extended periods may still affect model robustness. Evaluating the framework on more recent datasets remains an important direction for future work.

Second, the analysis focuses exclusively on English-language campaign blurbs. This introduces linguistic and platform-specific biases, limiting applicability to non-English or

culturally distinct crowdfunding ecosystems. Extending the approach to multilingual datasets and alternative platforms such as Indiegogo or GoFundMe represents a natural extension.

Third, the model relies solely on textual campaign blurbs; while this design choice aligns with early-stage prediction scenarios, it excludes multimodal signals such as images, videos, update frequency, and social interactions, which are known to influence trust and engagement. Consequently, the current framework captures the linguistic dimension of persuasion but not the full multimodal structure of campaign success.

Fourth, although the meta-heuristic algorithms identify compact and highly discriminative feature subsets, their stochastic nature may introduce variability in the selected features across runs. To mitigate this, fixed random seeds, repeated runs, and cross-validation averaging were employed; nevertheless, further work incorporating explicit stability analysis and richer post hoc interpretability tools (e.g., SHAP-based explanations) would enhance transparency.

Finally, category-level heterogeneity across Kickstarter—such as differing baseline success rates between Technology, Arts, and Games—raises potential fairness considerations, and while project category labels were not used as direct predictive features, indirect correlations may still emerge through linguistic patterns. A systematic fairness audit, including category-conditioned performance analysis and bias-aware evaluation metrics, constitutes an important avenue for future research to ensure that predictive accuracy does not disproportionately favor dominant or mainstream campaign types.

These limitations highlight promising directions for future work, including multimodal integration, cross-platform and multilingual validation, fairness-aware modeling, and enhanced interpretability analysis.

5.7. Future Directions

Future work could extend the framework along several dimensions:

- **Multimodal integration:** Incorporating campaign images, videos, and social interactions to capture richer signals of persuasiveness.
- **Cross-platform validation:** Testing on other crowdfunding platforms (e.g., Indiegogo, GoFundMe) to assess generalizability.
- **Real-time and continual learning:** Developing lightweight, on-device models and adaptive retraining strategies to reflect evolving campaign patterns.
- **Explainable and responsible AI:** Applying SHAP, LIME, and attention visualization to enhance transparency, while addressing fairness and bias in predictive outcomes.

These directions not only address current limitations but also pave the way for scalable, responsible deployment of predictive tools in digital commerce platforms.

6. Conclusions

This study presented a hybrid AI framework for predicting the success of crowdfunding campaigns by integrating fine-tuned BERT embeddings, a CBAM-enhanced convolutional autoencoder, and meta-heuristic feature selection. The proposed architecture directly addresses the challenge of high-dimensional textual representations, delivering strong predictive performance on an imbalanced Kickstarter dataset while substantially reducing computational footprint. In its best configuration, the model achieves 0.778 accuracy, 0.821 F1-score, and 0.515 MCC using only 117 features, and maintains robust performance under time-aware validation, indicating that its gains are not solely an artefact of random cross-validation.

The empirical analyses yield several key insights. First, CBAM-powered compression reduces the original $64 \times 768 \times 4$ BERT tensor to a 3072-dimensional latent space and, after feature selection, to 117-dimensional representations, while improving accuracy

and MCC relative to both raw embeddings and classical dimensionality reduction (PCA, UMAP). Second, ARO consistently produces more compact and discriminative feature subsets than GA and Jaya, particularly when optimized with an imbalance-aware objective. Third, LSTM-based classifiers systematically outperform GBM across all phases, underscoring the importance of sequence-aware architectures for capturing narrative structure in campaign blurbs. Ablation studies further confirm that each component—CBAM, learned compression, meta-heuristic selection, and MCC-based optimization—contributes distinct and complementary gains.

Beyond methodological contributions, the framework demonstrates practical relevance for digital commerce. Its compact architecture and sub-millisecond inference times make it suitable for platform-scale deployment in real-time scoring pipelines, enabling creators to receive early feedback on campaign narratives, backers to better assess risk, and platforms to enhance curation, recommendation, and coaching services. At the same time, the model's design is generic enough to be adapted to related tasks such as product launch forecasting, promotion optimization, and success-aware personalization in broader e-commerce contexts.

Overall, this work contributes both technical innovation and applied value: it shows how attention-guided compression and meta-heuristic feature selection can be combined to reconcile expressiveness, interpretive potential, and efficiency in text-based prediction, thereby supporting more informed and scalable decision-support tools in crowdfunding and beyond.

Author Contributions: Conceptualization, H.G., M.K. and E.S.B.M.; methodology, H.G., M.K. and E.S.B.M.; software, H.G.; validation, H.G.; formal analysis, H.G., M.K. and E.S.B.M.; investigation, H.G., M.K. and E.S.B.M.; resources, H.G.; data curation, H.G.; writing—original draft preparation, H.G., M.K. and E.S.B.M.; writing—review and editing, H.G., M.K. and E.S.B.M.; visualization, H.G.; supervision, H.G. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Kocaeli University Scientific Research Projects Coordination Unit under project number FAA-2025-4585.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original data presented in the study are openly available in <https://doi.org/10.5281/zenodo.12735215>. To ensure full reproducibility of our experiments, all scripts of source code have been made publicly available in the following GitHub repository: <https://github.com/hakangunduz86/crowdfunding-success-prediction-with-cbam-autoencoder-extracted-features-and-feature-selection> (accessed on 1 July 2025).

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Belleflamme, P.; Lambert, T.; Schwienbacher, A. Crowdfunding: Tapping the right crowd. *J. Bus. Ventur.* **2014**, *29*, 585–609. [CrossRef]
2. Belleflamme, P.; Omrani, N.; Peitz, M. The economics of crowdfunding platforms. *Inf. Econ. Policy* **2015**, *33*, 11–28. [CrossRef]
3. Mollick, E. The dynamics of crowdfunding: An exploratory study. *J. Bus. Ventur.* **2014**, *29*, 1–16. [CrossRef]
4. Kuppuswamy, V.; Bayus, B.L. Crowdfunding creative ideas: The dynamics of project backers. In *The Economics of Crowdfunding*; Springer International Publishing: Berlin/Heidelberg, Germany, 2018; pp. 151–182.
5. Qiu, C. Issues in Crowdfunding: Theoretical and Empirical Investigation on Kickstarter. 2013. Available online: <https://papers.ssrn.com/abstract=2345872> (accessed on 1 July 2025).
6. Kuti, M.; Madarász, G. Crowdfunding. *Public Financ. Q. Pénzügyi Szle.* **2014**, *59*, 355–366.
7. Chong, A.Y.L.; Ch'ng, E.; Liu, M.J.; Li, B. Predicting online product sales via online reviews, sentiments, and promotion strategies: A big data architecture and neural network approach. *Int. J. Oper. Prod. Manag.* **2017**, *37*, 307–329. [CrossRef]

8. Li, W.; Rakesh, V.; Reddy, C.K. Project success prediction in crowdfunding environments. In Proceedings of the 13th International Conference on Web Search and Data Mining, Houston, TX, USA, 3–7 February 2020; pp. 627–635.
9. Hu, W.; Yang, R. Predicting the success of Kickstarter projects in the US at launch time. In *The Intelligent Systems and Applications, Proceedings of the 2019 Intelligent Systems Conference (IntelliSys), London, UK, 5–6 September 2020*; Springer: Berlin/Heidelberg, Germany, 2020; Volume 1, pp. 497–506.
10. Gunduz, H. Comparative analysis of BERT and FastText representations on crowdfunding campaign success prediction. *PeerJ Comput. Sci.* **2024**, *10*, e2316. [\[CrossRef\]](#)
11. Agrawal, A.; Catalini, C.; Goldfarb, A. Crowdfunding: Geography, social networks, and the timing of investment decisions. *J. Econ. Manag. Strategy* **2015**, *24*, 253–274. [\[CrossRef\]](#)
12. Wash, R. The value of completing crowdfunding projects. In Proceedings of the International AAAI Conference on Web and Social Media, Cambridge, MA, USA, 8–11 July 2013; Volume 7, pp. 631–639.
13. Popescul, D.; Radu, L.D.; Păvăloaia, V.D.; Georgescu, M.R. Psychological determinants of investor motivation in social media-based crowdfunding projects: A systematic review. *Front. Psychol.* **2020**, *11*, 588121. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Cumming, D.J.; Leboeuf, G.; Schwienbacher, A. Crowdfunding models: Keep-it-all vs. all-or-nothing. *Financ. Manag.* **2020**, *49*, 331–360. [\[CrossRef\]](#)
15. Bradford, C.S. Crowdfunding and the Federal Securities Laws. *Columbia Bus. Law Rev.* **2012**, *2012*, 1–150. Available online: <https://journals.library.columbia.edu/index.php/CBLR/article/view/13858> (accessed on 1 July 2025). [\[CrossRef\]](#)
16. Klöhn, L.; Hornuf, L. Crowdinvesting in Deutschland: Markt, Rechtslage und Regulierungsperspektiven. *J. Bank. Law Bank.* **2012**, *24*, 237–266. [\[CrossRef\]](#)
17. Moenninghoff, S.C.; Wieandt, A. The Future of Peer-to-Peer Finance. In *Zeitschrift für Betriebswirtschaftliche Forschung*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 466–487.
18. Ferreira, V.; Papaoikonomou, E.; Terceno, A. Unpeel the layers of trust! A comparative analysis of crowdfunding platforms and what they do to generate trust. *Bus. Horizons* **2022**, *65*, 7–19. [\[CrossRef\]](#)
19. Mora-Cruz, A.; Palos-Sanchez, P.R. Crowdfunding platforms: A systematic literature review and a bibliometric analysis. *Int. Entrep. Manag. J.* **2023**, *19*, 1257–1288. [\[CrossRef\]](#)
20. Escudero, S.B.; Anglin, A.H.; Allison, T.H.; Wolfe, M.T. Crowdfunding: A theory-centered review and roadmap of the multidisciplinary literature. *J. Manag.* **2025**, *52*, 01492063251328267. [\[CrossRef\]](#)
21. Ryoba, M.J.; Qu, S.; Ji, Y.; Qu, D. The right time for crowd communication during campaigns for sustainable success of crowdfunding: Evidence from Kickstarter platform. *Sustainability* **2020**, *12*, 7642. [\[CrossRef\]](#)
22. Liang, T.P.; Wu, S.P.J.; Huang, C.c. Why funders invest in crowdfunding projects: Role of trust from the dual-process perspective. *Inf. Manag.* **2019**, *56*, 70–84. [\[CrossRef\]](#)
23. Kaur, H.; Gera, J. Effect of social media connectivity on success of crowdfunding campaigns. *Procedia Comput. Sci.* **2017**, *122*, 767–774. [\[CrossRef\]](#)
24. Koch, J.A.; Siering, M. The recipe of successful crowdfunding campaigns: An analysis of crowdfunding success factors and their interrelations. *Electron. Mark.* **2019**, *29*, 661–679. [\[CrossRef\]](#)
25. Zhong, C.; Xu, W.; Du, W. Success prediction of crowdfunding campaigns with project network: A machine learning approach. *J. Electron. Commer. Res.* **2022**, *23*, 99–114.
26. Li, Y.; Rakesh, V.; Reddy, C.K. Project success prediction in crowdfunding environments. In Proceedings of the Ninth ACM International Conference on Web Search and Data Mining, San Francisco, CA, USA 22–25 February 2016; pp. 247–256.
27. Etter, V.; Grossglauser, M.; Thiran, P. Launch hard or go home! Predicting the success of Kickstarter campaigns. In Proceedings of the First ACM Conference on Online Social Networks, Boston, MA, USA, 7–8 October 2013; pp. 177–182.
28. Yuan, H.; Lau, R.Y.; Xu, W. The determinants of crowdfunding success: A semantic text analytics approach. *Decis. Support Syst.* **2016**, *91*, 67–76. [\[CrossRef\]](#)
29. Ryoba, M.J.; Qu, S.; Zhou, Y. Feature subset selection for predicting the success of crowdfunding project campaigns. *Electron. Mark.* **2021**, *31*, 671–684. [\[CrossRef\]](#)
30. Schraven, E.; van Burg, E.; van Gelderen, M.; Masurel, E. Predictions of crowdfunding campaign success: The influence of first impressions on accuracy and positivity. *J. Risk Financ. Manag.* **2020**, *13*, 331. [\[CrossRef\]](#)
31. Hohen, S.; Hüning, C.; Schweizer, L. Reward-based crowdfunding—A systematic literature review. In *Management Review Quarterly*; Springer: Berlin/Heidelberg, Germany, 2025; pp. 1–48.
32. Dambanemuya, H.K.; Horvát, E.Á. A multi-platform study of crowd signals associated with successful online fundraising. *Proc. ACM Hum. Comput. Interact.* **2021**, *5*, 1–19. [\[CrossRef\]](#)
33. Elitzur, R.; Katz, N.; Muttath, P.; Soberman, D. The power of machine learning methods to predict crowdfunding success: Accounting for complex relationships efficiently. *J. Bus. Ventur. Des.* **2024**, *3*, 100022. [\[CrossRef\]](#)
34. Yang, J.; Li, Y.; Calic, G.; Shevchenko, A. How multimedia shape crowdfunding outcomes: The overshadowing effect of images and videos on text in campaign information. *J. Bus. Res.* **2020**, *117*, 6–18. [\[CrossRef\]](#)

35. Greenberg, M.D.; Pardo, B.; Hariharan, K.; Gerber, E. Crowdfunding support tools: Predicting success & failure. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*; ACM Digital Library: New York, NY, USA, 2013; pp. 1815–1820.
36. Buttice, V.; Rovelli, P. "Fund me, I am fabulous!" Do narcissistic entrepreneurs succeed or fail in crowdfunding? *Personal. Individ. Differ.* **2020**, *162*, 110037. [\[CrossRef\]](#)
37. Gülşen, E.; Gündüz, H.; Cataltepe, Z.; Serinol, L. Big data feature selection and projection for gender prediction based on user web behaviour. In *Proceedings of the 2015 23rd Signal Processing and Communications Applications Conference (SIU)*, Malatya, Turkey, 16–19 May 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 1545–1548.
38. Yu, P.F.; Huang, F.M.; Yang, C.; Liu, Y.H.; Li, Z.Y.; Tsai, C.H. Prediction of crowdfunding project success with deep learning. In *Proceedings of the 2018 IEEE 15th International Conference on E-Business Engineering (ICEBE)*, Xi'an, China, 12–14 October 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–8.
39. Wang, W.; Zheng, H.; Wu, Y.J. Prediction of fundraising outcomes for crowdfunding projects based on deep learning: A multimodel comparative study. *Soft Comput.* **2020**, *24*, 8323–8341. [\[CrossRef\]](#)
40. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
41. Westerlund, M.; Singh, I.; Rajahonka, M.; Leminen, S. Can short-text project summaries predict funding success on crowdfunding platforms? In *ISPIM Conference Proceedings, Proceedings of the International Society for Professional Innovation Management (ISPIM)*, Nanjing, China, 16–19 September 2019; Researchgate: Berlin, Germany, 2019; pp. 1–15.
42. Faralli, S.; Rittinghaus, S.; Samsami, N.; Distant, D.; Rocha, E. Emotional intensity-based success prediction model for crowdfunded campaigns. *Inf. Process. Manag.* **2021**, *58*, 102394. [\[CrossRef\]](#)
43. Kilinc, M.; Aydin, C. Feature selection for Turkish Crowdfunding projects with using filtering and wrapping methods. *Electron. Commer. Res. Appl.* **2023**, *62*, 101340. [\[CrossRef\]](#)
44. Yan, K.; Zhang, D. Feature selection and analysis on correlated gas sensor data with recursive feature elimination. *Sens. Actuators B Chem.* **2015**, *212*, 353–363. [\[CrossRef\]](#)
45. Fang, Y.; Yao, Y.; Lin, X.; Wang, J.; Zhai, H. A feature selection based on genetic algorithm for intrusion detection of industrial control systems. *Comput. Secur.* **2024**, *139*, 103675. [\[CrossRef\]](#)
46. Sağbaş, E.A.; Korukoglu, S.; Ballı, S. Real-time stress detection from smartphone sensor data using genetic algorithm-based feature subset optimization and k-nearest neighbor algorithm. *Multimed. Tools Appl.* **2024**, *83*, 1–32. [\[CrossRef\]](#)
47. Das, H.; Naik, B.; Behera, H.S. A Jaya algorithm based wrapper method for optimal feature selection in supervised classification. *J. King Saud Univ. Comput. Inf. Sci.* **2022**, *34*, 3851–3863. [\[CrossRef\]](#)
48. Sahu, B.; Dash, S. Optimal feature selection from high-dimensional microarray dataset employing hybrid IG-Jaya model. *Curr. Mater. Sci. Former. Recent Patents Mater. Sci.* **2024**, *17*, 21–43. [\[CrossRef\]](#)
49. Gao, J.; Wang, Z.; Jin, T.; Cheng, J.; Lei, Z.; Gao, S. Information gain ratio-based subfeature grouping empowers particle swarm optimization for feature selection. *Knowl. Based Syst.* **2024**, *286*, 111380. [\[CrossRef\]](#)
50. Tijjani, S.; Ab Wahab, M.N.; Noor, M.H.M. An enhanced particle swarm optimization with position update for optimal feature selection. *Expert Syst. Appl.* **2024**, *247*, 123337. [\[CrossRef\]](#)
51. Zhang, K.; Liu, Y.; Wang, X.; Mei, F.; Sun, G.; Zhang, J. Enhancing IoT (Internet of Things) feature selection: A two-stage approach via an improved whale optimization algorithm. *Expert Syst. Appl.* **2024**, *256*, 124936. [\[CrossRef\]](#)
52. Miao, F.; Wu, Y.; Yan, G.; Si, X. A memory interaction quadratic interpolation whale optimization algorithm based on reverse information correction for high-dimensional feature selection. *Appl. Soft Comput.* **2024**, *164*, 111979. [\[CrossRef\]](#)
53. Dogan, Y. AutoEffFusionNet: A New Approach for Cervical Cancer Diagnosis Using ResNet-based Autoencoder with Attention Mechanism and Genetic Feature Selection. *IEEE Access* **2025**, *13*, 44107–44122. [\[CrossRef\]](#)
54. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008.
55. Holland, J. *Adaptation in Natural and Artificial Systems*; University of Michigan Press: Ann Arbor, MI, USA, 1975; Volume 7, pp. 390–401.
56. Goldberg, D.E. Genetic algorithms in search, optimization, and machine learning. *Addison Wesley* **1989**, *1989*, 36.
57. Rao, R. Jaya: A simple and new optimization algorithm for solving constrained and unconstrained optimization problems. *Int. J. Ind. Eng. Comput.* **2016**, *7*, 19–34. [\[CrossRef\]](#)
58. Wang, L.; Cao, Q.; Zhang, Z.; Mirjalili, S.; Zhao, W. Artificial rabbits optimization: A new bio-inspired meta-heuristic algorithm for solving engineering optimization problems. *Eng. Appl. Artif. Intell.* **2022**, *114*, 105082. [\[CrossRef\]](#)
59. Gunduz, H. Scalable Gender Profiling from Turkish Texts Using Deep Embeddings and Meta-Heuristic Feature Selection. *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 253. [\[CrossRef\]](#)
60. Lee, S.; Lee, K.; Kim, H.c. Content-based success prediction of crowdfunding campaigns: A deep learning approach. In *Proceedings of the Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing*, Jersey City, NJ, USA, 3–7 November 2018; pp. 193–196.

61. Kaminski, J.C.; Hopp, C. Predicting outcomes in crowdfunding campaigns with textual, visual, and linguistic signals. *Small Bus. Econ.* **2020**, *55*, 627–649. [[CrossRef](#)]
62. Tang, Z.; Yang, Y.; Li, W.; Lian, D.; Duan, L. Deep cross-attention network for crowdfunding success prediction. *IEEE Trans. Multimed.* **2022**, *25*, 1306–1319. [[CrossRef](#)]

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