The Edward S. Rogers Sr. Department of Electrical & Computer Engineering
UNIVERSITY OF TORONTO

FundSeer - Final Report

Submitted to:
ECE1786 Teaching Staff

Submitted by:

Mohamed Attia
mohamed.attia@mail.utoronto.ca
Xiaohan Lu
xiaohan.lu@mail.utoronto.ca

Word count: 1941, penalty: 0%

Permissions:

	Mohamed	**Xiaohan**
Post Video	**YES**	**YES**
Post Final Report	**YES**	**YES**
Post Source Code	**YES**	**YES**

1 Introduction

Access to healthcare is universally acknowledged as a fundamental human right. However, financial constraints often obstruct timely medical interventions, creating a critical gap in healthcare accessibility. In response to this challenge, crowdfunding platforms have emerged as innovative avenues for individuals to secure necessary funds for their medical expenses. This paper presents a comprehensive study focusing on the pivotal role of campaign narratives in influencing the success of medical crowdfunding initiatives.

Our primary objective is to predict the probability of success for medical crowdfunding campaigns. In this context, a campaign is considered successful if the amount raised meets or surpasses its stated financial goal. Early identification of unsuccessful campaigns enables strategic interventions, such as narrative enhancement or intensified marketing efforts, to augment their likelihood of success. Natural language Processing (NLP) tools are particularly apt for this task because they can help discern complex patterns in large datasets, which might not be immediately apparent through traditional analysis.

Our architecture employs a scrapper module to extract key metadata from medical crowdfunding campaigns, which are then fed into our main model to calculate the probability of campaign success. An illustration of our proposed system is shown in Fig. 1

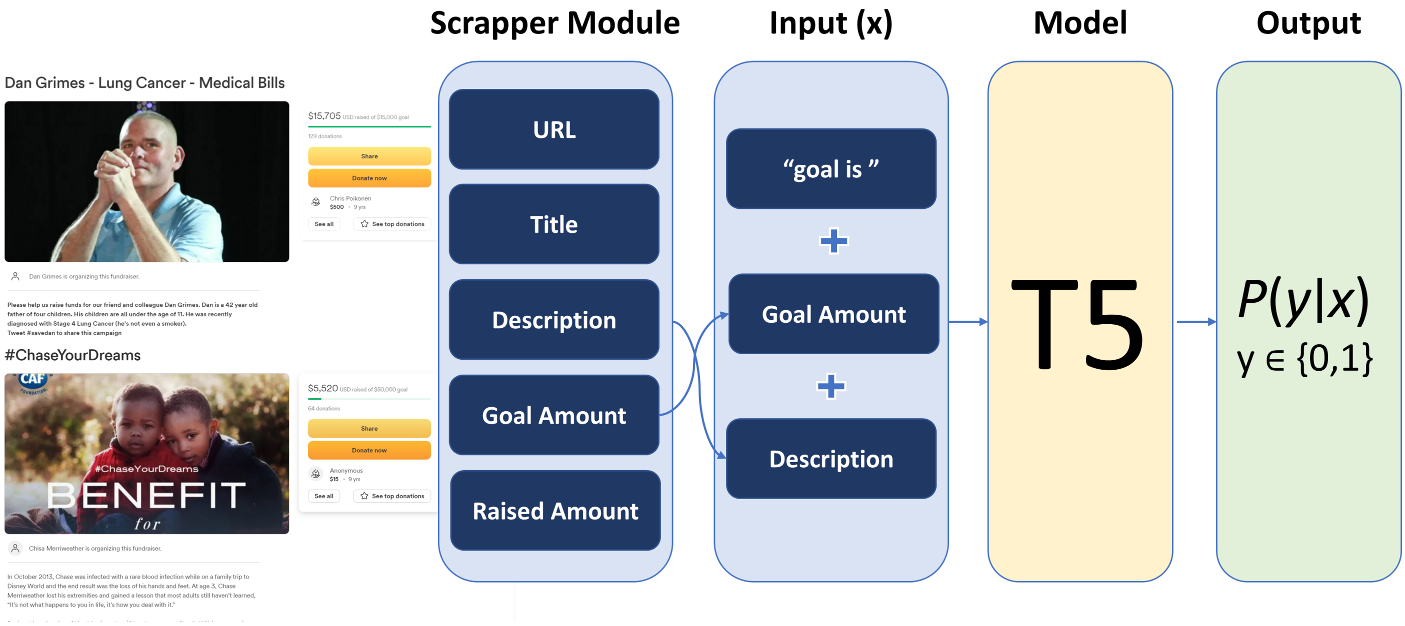


Figure 1: Our proposed system pipeline.

2 Related Work

In exploring the relatively uncharted domain of surgical crowdfunding campaigns, a study [1] gathered a dataset of 66,514 campaigns from 2010 to 2020, which raised $354.8 million out of a total of one billion dollars. This research innovatively applied NLP techniques, specifically using a binary classification Long Short-Term Memory (LSTM) network. It utilized word embeddings derived from campaign texts to predict the success of these campaigns, achieving an accuracy of 0.6042 and an F1 score of 0.3766. This study leverages deep learning to predict the success of surgical crowdfunding campaigns. However, the relatively low F1 score of 0.3766 suggests the dataset was unbalanced. In such scenario, accuracy can be misleading, and the model's predictive performance may not be as robust as it appears, necessitating a more nuanced approach to evaluating model efficacy.

Complementing this, another study [2] focused on the success prediction of broader crowdfunding projects using Multimodal Deep Learning (MDL). It integrated heterogeneous textual and visual information, employing a pre-trained 16-layer VGG model for visual features and Bag of Words (BoW) with Term Frequency-Inverse Document Frequency (TF-IDF) and word embedding for textual representation. While the MDL model showcased superior performance compared to other methods, highlighting the efficacy of multimodal deep learning in forecasting crowdfunding project outcomes, our research focus is distinctly on the narrative aspect of campaigns.

3 Dataset

The dataset encompasses the URL, title, campaign description, as well as financial metrics such as the goal and raised amounts, offering a comprehensive snapshot of each crowdfunding initiative. After processing 1.4 million URLs, we had 80 thousand balanced samples, half of which were successful. A sample of our dataset is shown in Fig. 2.

	url	Title	Description	Goal Amount	Raised Amount	success	language	ids	tokens
0	https://wv	Levi Mayn	As many of you ma	100000	81220	0	en	[1, 2, 7004	['know', 'bill',
1	https://wv	Hayze Jam	Hayze Jameson wa	5000	2630	0	en	[7004, 700	['hayze', 'jam
2	https://wv	Mike Codd	Hi We at JVR Elect	10000	11090	1	en	[326, 7004	['hi', 'jvr', 'ele
3	https://wv	Be a part c	As many of you kn	30000	4765	0	en	[1, 359, 36	['know', 'dau
4	https://wv	Steven Cyr	My son Steven and	300	1295	1	en	[401, 402,	['son', 'steven
5	https://wv	help Mom	Hi my name is Gra	20000	5749	0	en	[326, 412,	['hi', 'grace', '
6	https://wv	Help with I	Hi my daughter in	4100	493	0	en	[326, 359,	['hi', 'daughte

Figure 2: A Sample from our dataset.

3.1 Data Processing

The data collection and preprocessing for our study on medical crowdfunding campaigns involved an initial phase of URL retrieval from GoFundMe along with leveraging the Internet Archive's WayBackMachine to access historical campaigns. Subsequent steps focused on refining this dataset to include only relevant campaigns. Specifically, those within the medical category that were written in English and were raised in USDs. This targeted approach was designed to ensure data uniformity and relevance to the study's objectives. The processed data underwent further analysis, generating descriptive visualizations such as histograms and word clouds to reveal underlying patterns and trends within the campaigns.

We also applied additional preprocessing steps, including removal of emojis and non-standard ASCII characters, elimination of stop words, and tokenization. Then, we applied padding to ensure uniform input length for the models. In the end, we split the data into training, validation and test set in a 80%, 10%, 10% ratio, respectively. Fig. 3 is an illustration of our data processing pipeline.

3.1.1 Insights

We extracted some interesting insights from the data:

- Out of the campaigns that have a goal exceeding \$100,000, approximately 56.20% of them contain the words "surgery" or "cancer" in their descriptions as depicted in Fig. 4a.
- The average goal set for successful campaigns is about \$8,042.33, whereas the average goal set for unsuccessful campaigns is about \$48,150.24 as shown in Fig. 4b.

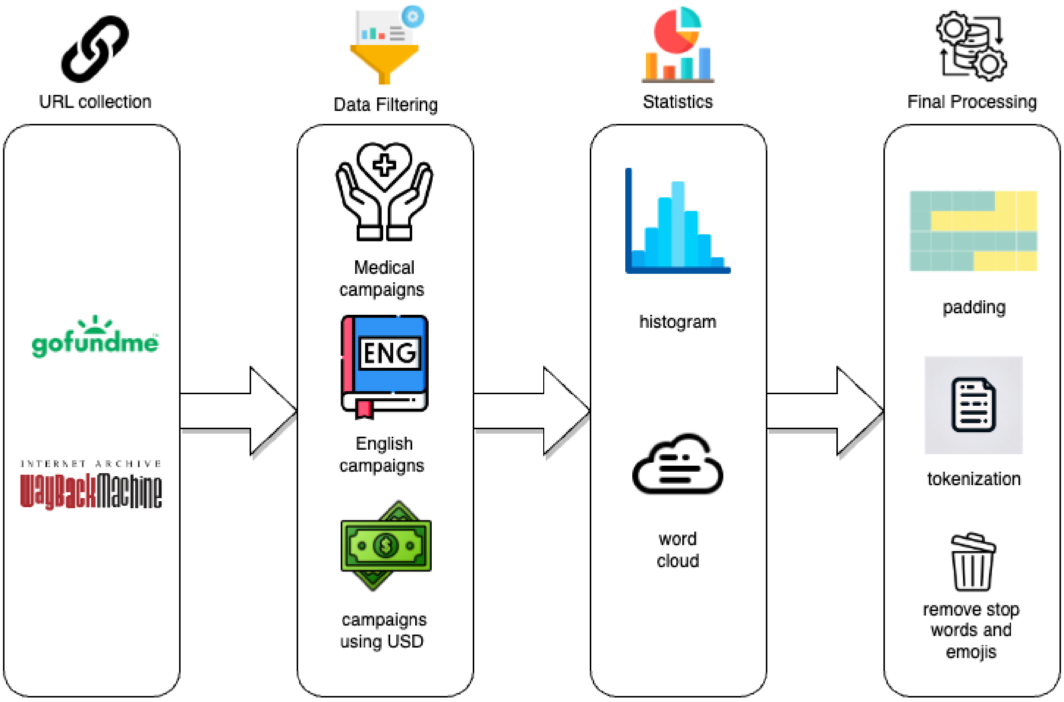
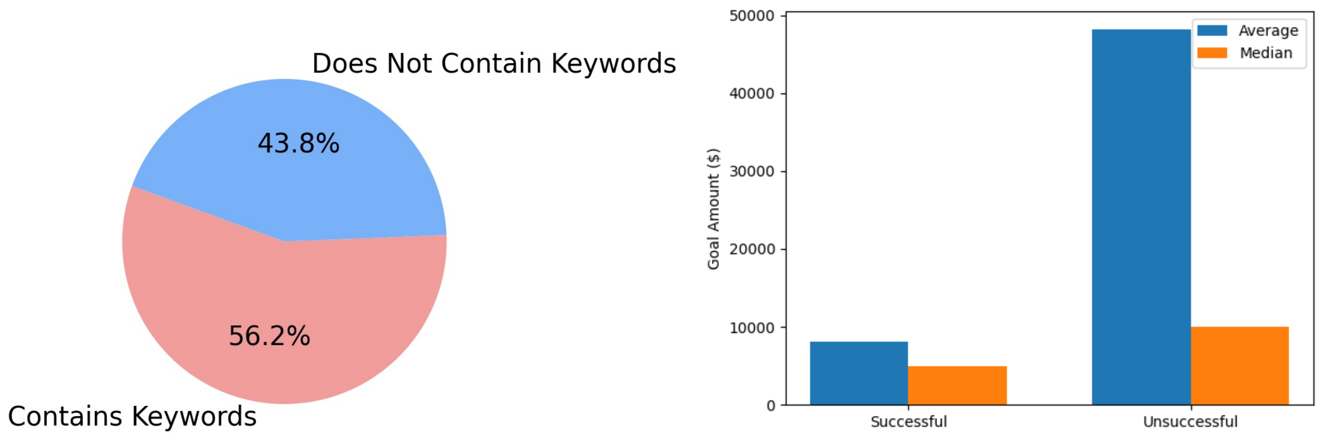


Figure 3: Data processing pipeline.

- The average goal amount set across all campaigns is approximately $28,095.97.



(a) Percentage of campaigns over $100k with keywords "surgery" or "cancer".

(b) Average vs median goal amounts by campaign outcome.

Figure 4: Data statistics

4 Models

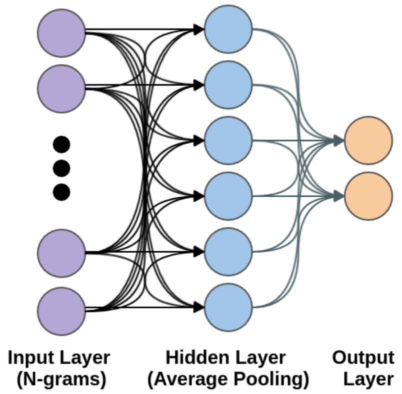
4.1 Baseline Models

4.1.1 Shallow Neural Network (SNN) Model

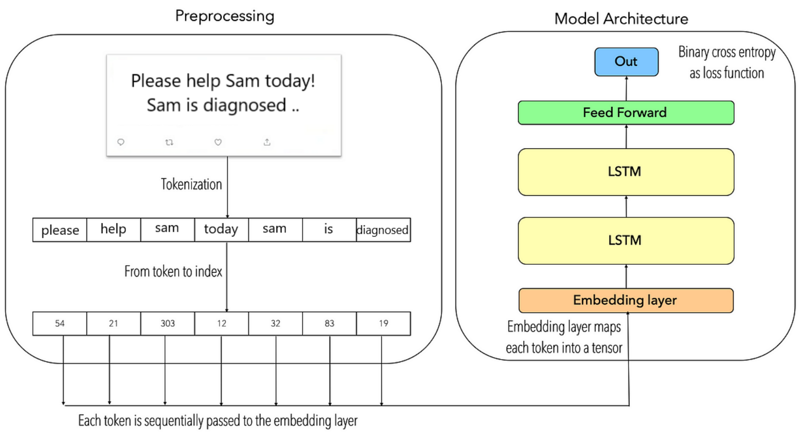
For the initial baseline model, we employed a word embedding model, specifically GloVe [3], to function as the fundamental backbone. Following this, we computed average embeddings associated with the words present in the campaign's narrative, which were then input into a two-layer neural network classifier with a dropout layer of $p = 0.5$ in between, as depicted in Fig. 5a.

4.1.2 Long short-term memory (LSTM) Model

We have chosen our second baseline as an LSTM-based model, aiming to enhance the modeling of temporal dependencies, namely the contextual aspects of the narrative, in our task. The LSTM model starts by embedding each word in the campaign narrative, then an LSTM layer processes each embedded vector individually. Thereafter, the last hidden state is projected to a fully connected layer for predicting the success probability of a campaign, as illustrated in Fig. 5b.



(a) SNN, adopted from [4].



(b) LSTM, modified from [5].

Figure 5: Baseline models architecture.

4.2 Main Models

Our primary model is the Text-To-Text Transfer Transformer (T5) [6], a cutting-edge architecture that has demonstrated state-of-the-art performance across various NLP tasks. Notably, T5-11B has achieved the highest accuracy on the Stanford Sentiment Treebank (sst-2) [7], a widely recognized benchmark for sentiment analysis. This success motivated our decision to leverage the pre-trained smaller versions, specifically the 'small' and 'base' variants of T5, for our task. This choice enables us to effectively capture intricate narrative patterns and details, surpassing the capabilities of baseline models.

In addition, we incorporated A Lite Bidirectional Encoder Representations from Transformers (ALBERT)-base model [8], a parameter-reduced version of the original BERT model designed for improved efficiency with lower GPU memory footprint and faster training. T5 and ALBERT models underwent extensive experimentation, involving training the entire set of model parameters (end2end) or focusing solely on training the classifier layers (clf). The hyperparameters for T5, ALBERT, and baseline models are detailed in Table 1.

For both T5 and ALBERT, we accessed the models and tokenizers through the Hugging Face Transformers library [9]. T5 models were trained on different Colab accounts using a 16-GB T4 GPU, while ALBERT was trained on Kaggle using a P100 16-GB GPU. The input variable, denoted as x, for both models, consisted of the goal amount and campaign narrative combined, aiming to predict the likelihood of a campaign being either successful or not, denoted as $p(y|x)$, where y represents the campaign class (1 for successful and 0 for unsuccessful). This relationship is visually depicted in the block diagram in Fig. 1.

Hyperparameter	SNN	LSTM	T5 & ALBERT
Embedding Dimension	300	256	Default
Learning Rate	1e-3	4e-5	1e-4
Learning Rate Scheduler	Cosine	Cosine	Cosine
Weight Decay	0	0	0.1
Epochs	50	30	5
Batch Size	256	256	16
Sequence Length	512	512	512
GloVe Model	6B	-	-

Table 1: Models Hyperparameters

4.3 Generative Pre-trained Transformer (GPT) Model

A GPT-based classifier was also explored to analyze the characteristics of crowdfunding campaign descriptions to predict their success. Notice that in the main model, the output is the probability of success but in this GPT-based model the output is binary (successful or not). It employs a chain-of-thought prompting approach, where the model is first primed with detailed criteria distinguishing successful and unsuccessful campaigns, considering factors like title clarity, description detail, and goal amount realism. After setting this context, it is provided with the title, description, and goal amount of a specific campaign to determine its likelihood of success. However, a critical observation of this classifier is its tendency to predict a campaign as successful, which could indicate a bias in the model. Our prompt can be found under the `gfm-gpt-classfier.py` script.

5 Results

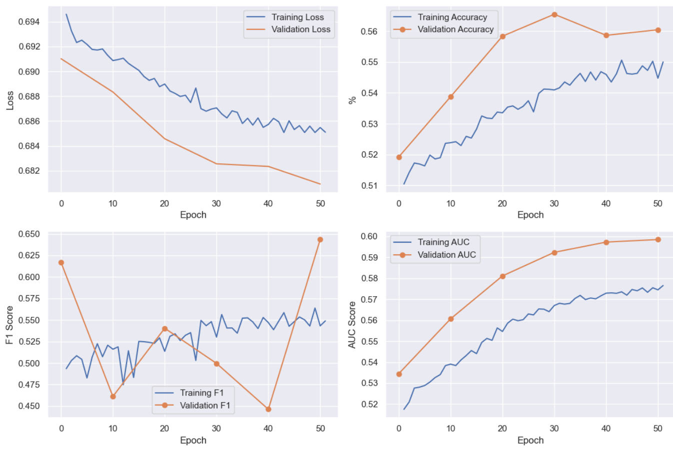
The learning curves for both the baseline models and primary models are presented in Fig.s [6] and [7], respectively. Additionally, classification metrics for all models on the test set are summarized in Table 2. Examination of results reveals that baseline models exhibited poor performance on our balanced test set, achieving accuracy scores of 0.547 and 0.603, for the SNN and LSTM models, respectively.

Notably, training only the classification layers of T5-small yielded the highest accuracy at 0.728 on the test set. Furthermore, all other language models demonstrated accuracies exceeding 0.7 when compared to the best baseline model, LSTM, which achieved a score of 0.603.

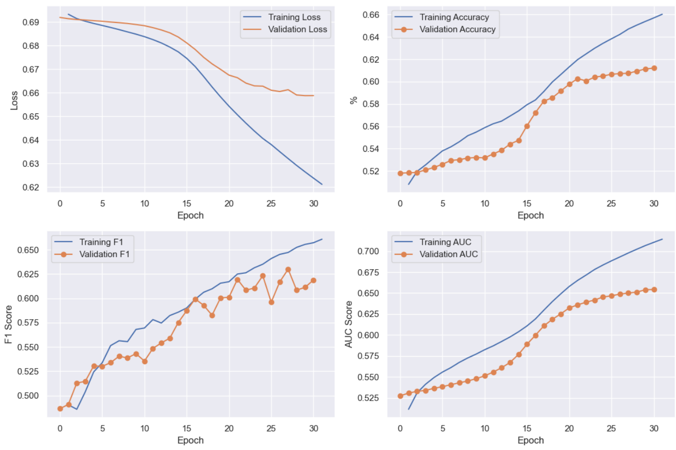
Our data ablation studies confirmed that incorporating the goal amount with the description significantly benefitted language models, resulting in an accuracy increase of 5-6%. Examples of predictions made by the T5-small-clf model are showcased in Fig. 8.

Model	Acc	F1	AUC	#Params
Shallow Neural Network	0.547	0.628	0.582	90.6K
LSTM	0.603	0.605	0.642	2.8M
ALBERT-base-end2end	0.707	0.707	0.787	11.8M
T5-small-clf	**0.728**	**0.716**	**0.806**	60.5M
T5-small-end2end	0.723	0.713	0.806	60.5M
T5-base-clf	0.718	0.712	0.799	223M

Table 2: Models performance on test set.



(a) SNN



(b) LSTM

Figure 6: Learning curves for the baseline models

.

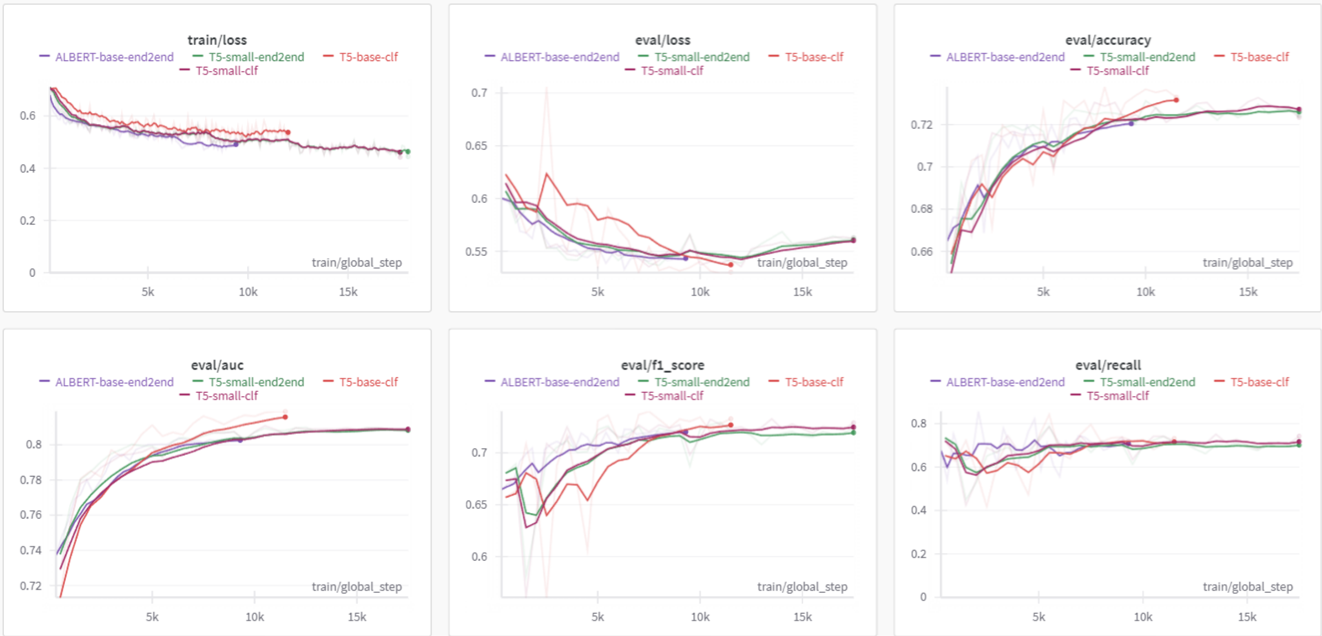
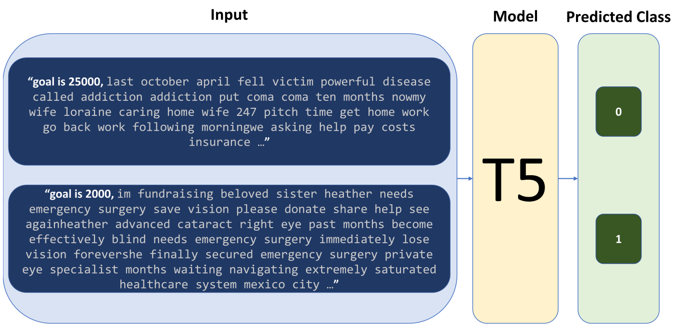
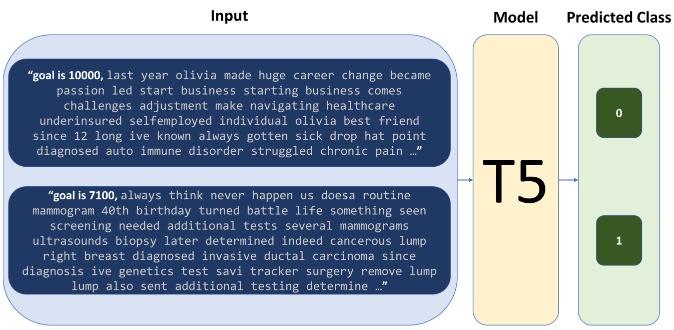


Figure 7: Learning curves for T5 and ALBERT models.



	Description	Goal	Raised Amount	Label	0	1	Prediction
3	last october april fell victim powerful diseas...	25000.0	1870.0	0	1.749076	-1.766922	0
4	im fundraising beloved sister heather needs em...	2000.0	2570.0	1	-0.338364	0.773683	1

(a) Correctly classified.



	Description	Goal	Raised Amount	Label	0	1	Prediction
412	last year olivia made huge career change becam...	10000.0	16564.0	1	1.148222	-1.229415	0
415	always think never happen us doesa routine mam...	7100.0	3120.0	0	0.009986	0.095277	1

(b) Incorrectly classified.

Figure 8: Samples from T5-small-clf predictions.

6 Discussion and Learnings

LSTM exhibited a noteworthy performance advantage, surpassing the SNN model by more than 5.5% in accuracy. This underscores the significance of the narrative context, emphasizing that the arrangement of words provides more informative cues than assigning equal weights to words regardless of their order, as employed in the SNN model.

Remarkably, the language models achieved accuracy and F1 scores exceeding 70%, along with AUC values surpassing 78%. This signifies their capability to discern successful campaigns based solely on their goals and descriptions. Intriguingly, end-to-end training of ALBERT-base, with a modest parameter count of 12 million, successfully captured the patterns distinguishing between the two campaign classes. Additionally, end-to-end training was not imperative for T5-small; training only the classifier head and preserving pretraining knowledge proved superior. While we acknowledge that further exploration of training durations for T5-base could potentially yield enhanced performance, resource limitations constrained us from extensive finetuning. A single trial with 3 epochs alone could take more than 18 hours to complete.

In examining samples in Fig. 8a, it is evident that the model adeptly captures the tendency of individuals to support severe cases, particularly those related to cancer or surgery, as we illustrated previously in Fig. 4a. This may explain its prediction of success for the second example while predicting the first as unsuccessful. After analyzing misclassified narratives in Fig. 8b, we observed that the first narrative involved an individual initially diagnosed with an autoimmune disease and later with cancer. This shift in medical history may have led to model confusion. In the second example, related to a breast cancer campaign, the model predicted success; however, the title wasn't captivating, consisting only of the person's name and "Fight." We assert that a compelling title is a crucial element contributing significantly to the success and better marketing of a campaign.

An area for enhancement involves prompt engineering. Current GPT-based model prompts can be further optimized by examining campaign descriptions in greater depth. Additionally, expanding the outcome categories beyond a binary classification could provide a richer, more dimensional analysis of campaign success. Transforming the classifier into an analyzer, which instead of predicting outcomes, provides insights into the strengths and weaknesses of campaign descriptions, could offer more actionable guidance for campaign organizers.

Our task is inherently complex compared to similar tasks, such as sentiment analysis, where language models can infer sentiment based on certain keywords. In fact, examining word clouds for the two classes of narratives (Fig. 9) reveals a high degree of similarity in the most frequent words. To address the need for capturing more intricate patterns to differentiate between classes, we advocate for the integration of additional modalities—specifically, a campaign's image and metadata. Such integration, as studied in [2], has the potential to enhance performance over depending solely on the description of campaigns.

7 Contributions

Mohamed

- Project conceptualization and management.
- Optimized data scrapper code to utilize multiprocessing.
- Developed the SSN, T5, and ALBERT models.
- Trained all models to yield the best performance for each.

(a) Successful campaigns.

(b) Unsuccessful campaigns.

Figure 9: Campaigns' narrative word cloud.

Xiaohan

- Implemented data scrapper script.
- Optimized data scrapper code to utilize fake user agent.
- Analyzed data to extract insights.
- Preprocessed data for baseline models.
- Developed LSTM and GPT models.

References

[1] A. Patil and H. Lee, "Prediction of surgical crowdfunding campaign success using nlp," *Stanford CS 230*, 2022. [Online]. Available: `https://cs230.stanford.edu/projects_fall_2022/reports/109.pdf`.

[2] C. Cheng, F. Tan, X. Hou, and Z. Wei, "Success prediction on crowdfunding with multimodal deep learning," in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, International Joint Conferences on Artificial Intelligence Organization, Jul. 2019, pp. 2158–2164. DOI: `10.24963/ijcai.2019/299`. [Online]. Available: `https://doi.org/10.24963/ijcai.2019/299`.

[3] J. Pennington, R. Socher, and C. Manning, "GloVe: Global vectors for word representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, A. Moschitti, B. Pang, and W. Daelemans, Eds., Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1532–1543. DOI: `10.3115/v1/D14-1162`. [Online]. Available: `https://aclanthology.org/D14-1162`.

[4] Austin, "FastText for Sentence Classification - Austin G. Walters," *Austin G. Walters*, Jan. 2019. [Online]. Available: `https://austingwalters.com/fasttext-for-sentence-classification`.

[5] F. López, "Text Classification with LSTMs in PyTorch - Towards Data Science," *Medium*, Dec. 2021, ISSN: 7111-1116. [Online]. Available: `https://towardsdatascience.com/text-classification-with-pytorch-7111dae111a6`.

[6] C. Raffel, N. Shazeer, A. Roberts, *et al.*, "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer," *arXiv*, Oct. 2019. DOI: `10.48550/arXiv.1910.10683`. eprint: `1910.10683`.

[7] R. Socher, A. Perelygin, J. Wu, *et al.*, "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank," *ACL Anthology*, pp. 1631–1642, Oct. 2013. [Online]. Available: `https://aclanthology.org/D13-1170`.

[8] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "ALBERT: A Lite BERT for Self-supervised Learning of Language Representations," *arXiv*, Sep. 2019. DOI: `10.48550/arXiv.1909.11942`. eprint: `1909.11942`.

[9] T. Wolf, L. Debut, V. Sanh, *et al.*, "HuggingFace's Transformers: State-of-the-art Natural Language Processing," *arXiv*, Oct. 2019. DOI: `10.48550/arXiv.1910.03771`. eprint: `1910.03771`.