# Predictive analytics of the success of crowd funding campaigns on GoFundMe

Honnavalli Jayarama, Lekhashree
Kilaru, Sai Hemanth
Talapaneni, Mokshagna Gnana Teja
Maringanti, Meghana

**Abstract**

This report presents a comprehensive analysis of GoFundMe crowdfunding campaigns, focused on identifying key factors that influence campaign success. Through data scraping, preprocessing, exploratory data analysis, and predictive modeling, we have uncovered significant patterns that drive funding outcomes. The study examines various campaign attributes including category types, goal amounts, campaign narratives, and engagement metrics to develop accurate prediction models. Our findings reveal that campaign category, fundraising goal, donation count, and social engagement metrics significantly impact success rates. The insights gained from this study can benefit campaign creators, platform developers, and researchers in understanding the dynamics of online fundraising success.

## 1 Introduction

Crowdfunding has revolutionized how individuals and organizations raise funds for various causes, from medical emergencies to business ventures and creative projects. Platforms like GoFundMe have democratized fundraising, allowing anyone with internet access to create campaigns and solicit donations from a global audience. Despite the accessibility of these platforms, not all campaigns achieve their funding goals. Understanding the factors that differentiate successful campaigns from unsuccessful ones is crucial for both campaign creators and platform developers.

This research project investigates the determinants of crowdfunding success by analyzing data scraped from GoFundMe, one of the world's largest crowdfunding platforms. We aim to identify patterns and features that significantly influence campaign outcomes and develop predictive models that can accurately forecast a campaign's likelihood of meeting its funding goal.

### 1.1 Research Questions

Our analysis addresses the following key research questions:

1. Do certain campaign categories (e.g., medical, education, business) have a higher likelihood of success than others?
2. What key attributes significantly influence the success of crowdfunding campaigns on GoFundMe?
3. How do textual features and sentiment of a campaign narrative influence its fundraising success?

By answering these questions, we seek to provide actionable insights for campaign creators to optimize their fundraising strategies and improve their chances of success.

# 2 Methodology

## 2.1 Data Collection

We employed Python-based web scraping techniques to gather data from GoFundMe campaigns while maintaining ethical standards and complying with the platform's Terms of Service. The data collection process was designed to capture a diverse range of campaigns across various categories, goals, and outcomes.

The dataset includes the following key attributes for each campaign:

- URL index and title of the campaign
- Amount raised and funding goal
- Campaign category
- Creation date and last update date
- Engagement metrics (donors, shares, followers)
- Campaign narrative/story

## 2.2 Data Preprocessing

Raw data extracted from web scraping required extensive cleaning and preprocessing to prepare it for analysis. Our preprocessing steps included:

- Removing irrelevant columns such as image URLs and CDN links
- Standardizing column names and data types
- Converting date strings to datetime objects
- Handling missing values through appropriate imputation techniques
- Creating derived features such as campaign duration and goal achievement ratio

For categorical variables like campaign categories and geographic locations, we applied label encoding to transform them into numerical representations suitable for machine learning algorithms. Numerical features with skewed distributions were transformed using logarithmic scaling to improve model performance.

## 2.3 Feature Engineering

To enhance our predictive models, we created several derived features:

- Campaign duration (days between creation and last donation)
- Goal per day (funding target divided by campaign duration)
- Donation ratio (number of donations relative to goal amount)
- Description length (character count of campaign narrative)
- Campaign name length
- Boolean indicators for charity association

These engineered features provided additional dimensions for analysis and improved the predictive power of our models.

### 2.4 Success Definition

We defined campaign success as achieving or exceeding the stated funding goal. This binary classification approach allows us to differentiate between successful and unsuccessful campaigns:

$$Success = \begin{cases} 1, & \text{if } \frac{AmountRaised}{GoalAmount} \geq 1 \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

### 2.5 Modeling Approach

We implemented and compared multiple classification algorithms to predict campaign success: Logistic Regression, Decision Trees, Random Forest, Extra Trees, Gradient Boosting, K-Nearest Neighbors (KNN), Support Vector Machines (SVM), AdaBoost, XGBoost, and LightGBM.

For each model, we employed a train-test split methodology (80% training, 20% testing) and evaluated performance using accuracy, precision, recall, F1-score, and ROC-AUC metrics.

## 3 Results

### 3.1 Exploratory Data Analysis

#### 3.1.1 Success Rates by Campaign Category

Our analysis revealed significant variations in success rates across different campaign categories.

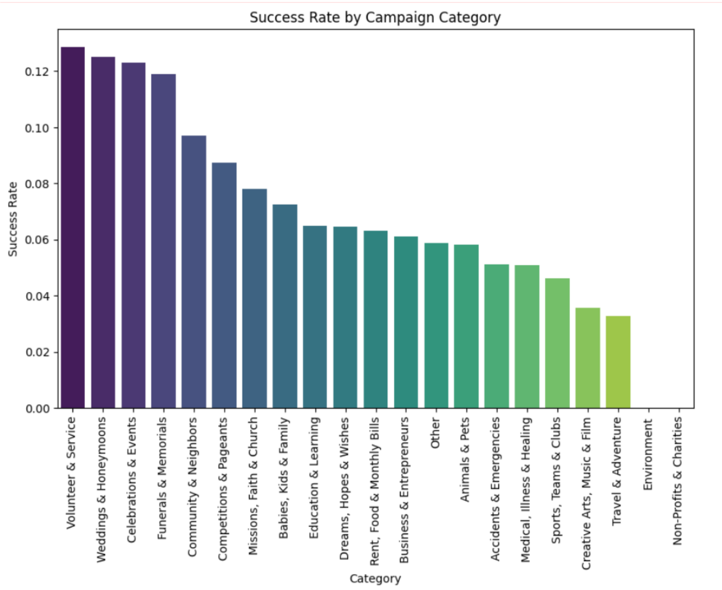


Figure 1: Success Rate by Campaign Category

As shown in Figure 1, "Volunteer & Service" campaigns have the highest success rates, followed by "Weddings & Honeymoons" and "Celebrations & Events." This differs from our initial assumption about medical campaigns. The data indicates that community-oriented and celebratory causes tend to resonate more with donors. Conversely, "Environment," "Travel & Adventure," and "Non-Profits & Charities" categories show lower success rates.

#### 3.1.2 Goal Amount Distribution

The relationship between campaign goal amount and success probability reveals interesting patterns:

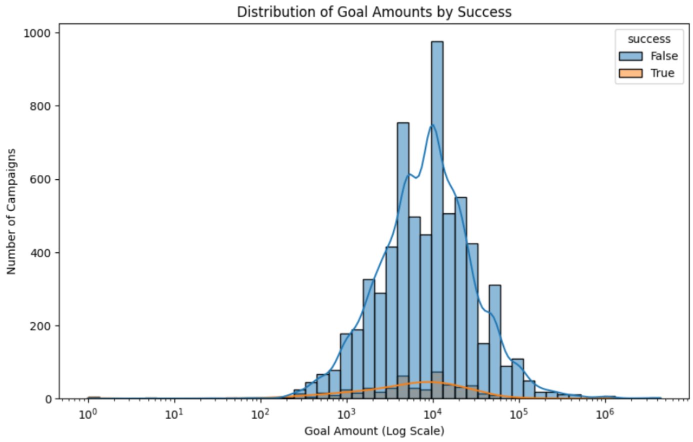


Figure 2: Distribution of Goal Amounts by Success Status

Figure 2 demonstrates that successful campaigns typically set more modest fundraising goals. The histogram shows that unsuccessful campaigns (blue) have higher goal amounts clustered in the $1,000 - $100,000 range, with a peak around $10,000, while successful campaigns (orange) tend to have significantly lower goal amounts. This suggests that realistic goal-setting plays a crucial role in campaign outcomes.

#### 3.1.3 Campaign Goal Achievement Analysis

Our analysis of campaign goal achievement percentages reveals important patterns in funding outcomes:

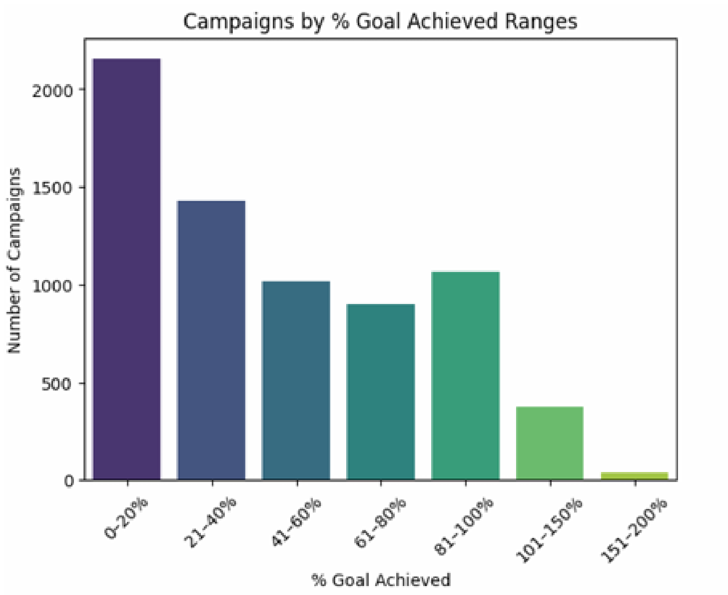


Figure 3: Distribution of Campaign Goal Achievement Percentages

As shown in Figure 3, we observed three key patterns:

1. Underperforming campaigns are most common — the 0-20% achievement range has the highest number of campaigns (approximately 2,100), suggesting many campaigns struggle to reach even a quarter of their goals.
2. There's a steady decline in frequency as performance improves — with each increasing achievement range from 0-20% through 61-80%, the number of campaigns generally de-

creases, indicating that higher levels of goal achievement become progressively more difficult.

3. Interestingly, there's a notable uptick in the 81-100% range (approximately 1,050 campaigns) compared to the 61-80% range (about 900 campaigns), suggesting that campaigns approaching their targets often make a final push to reach full completion, while very few campaigns (less than 100) exceed their goals by significant margins.

### 3.1.4 Engagement Metrics Analysis

Campaign engagement metrics showed correlation with success rates:

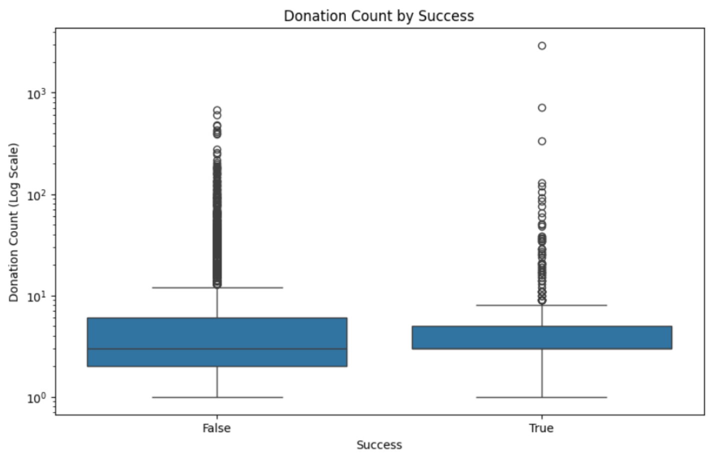


Figure 4: Donation Count by Success Status

The boxplot in Figure 4 shows donation counts for unsuccessful (False) and successful (True) campaigns on a logarithmic scale. While the median values appear similar, successful campaigns show some higher outliers, indicating that some successful campaigns achieve a significantly higher number of donations.

### 3.1.5 Correlation Analysis

The correlation analysis between various campaign attributes revealed important relationships:

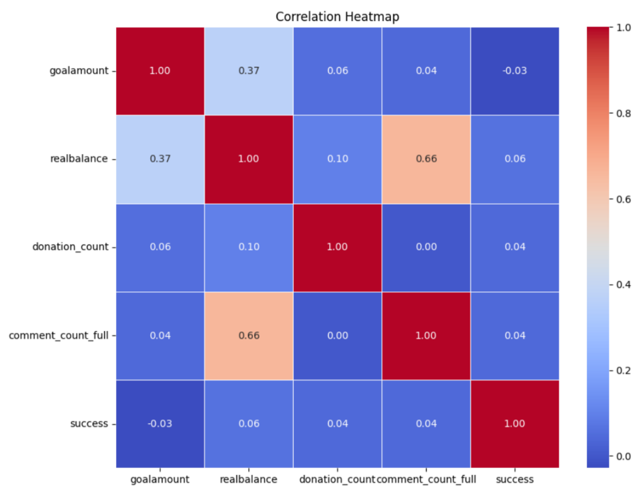


Figure 5: Correlation Heatmap of Key Campaign Attributes

As shown in Figure 5, there's a strong positive correlation (0.66) between comment count and real balance. The correlation between goal amount and real balance is moderate (0.37). However, contrary to our initial assessment, the correlation between success and the other variables appears relatively weak, with values ranging from -0.03 to 0.06. This suggests that the relationship between these individual metrics and campaign success is more complex than a simple linear correlation.

## 3.2 Textual Features Analysis

We conducted a comprehensive analysis of textual features from campaign descriptions to understand their impact on campaign success:

### 3.2.1 Text Characteristics Distribution

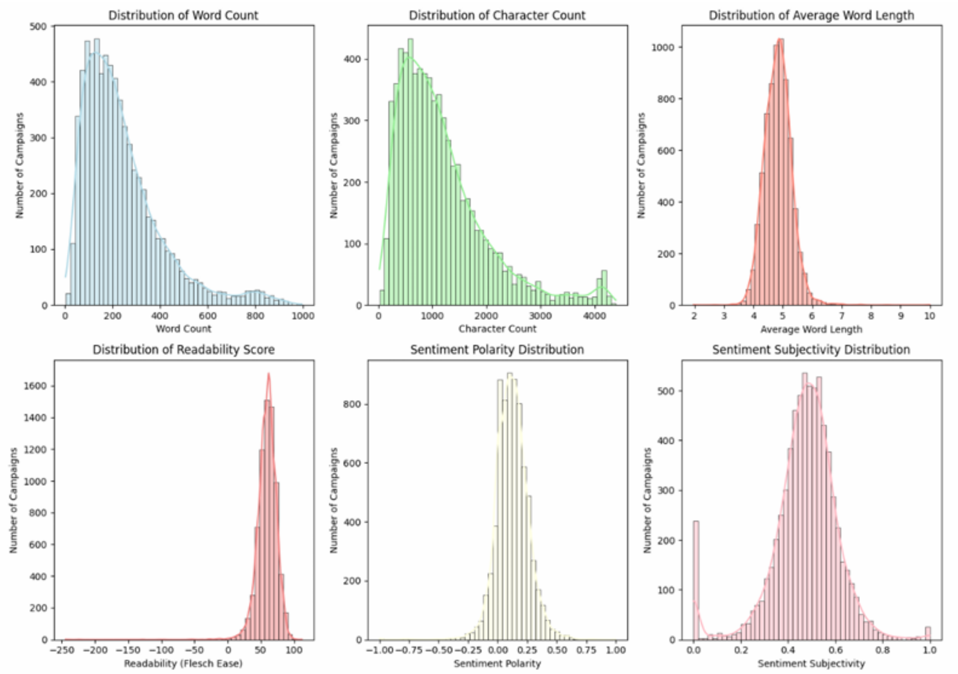


Figure 6: Distribution of Textual Features in Campaign Descriptions

As illustrated in Figure 6, our analysis of campaign descriptions revealed several key patterns:

- **Word Count Distribution:** The distribution shows a rightward skew, indicating that most campaigns have a relatively small number of words, with fewer campaigns having a very high word count. The majority of the campaigns fall within the 0 to 200-word range.
- **Character Count Distribution:** Similar to word count, this also exhibits a rightward skew, suggesting that most campaigns have a smaller number of characters, and the number of campaigns decreases as the character count increases.
- **Average Word Length Distribution:** Approximately normally distributed with a peak around 4 to 5 characters, implying that words in these campaigns tend to be of moderate length.
- **Readability Score Distribution:** Shows a distribution concentrated towards higher readability scores. The bulk of the campaigns have Flesch Reading Ease scores clustered in the positive range, suggesting that the text used in these campaigns is generally considered easy to read.

- **Sentiment Polarity Distribution:** Centered around 0, suggesting that, on average, the sentiment of the campaigns is neutral. There's a spread of values, indicating some campaigns have positive and some have negative sentiment, but the majority are close to neutral.
- **Sentiment Subjectivity Distribution:** Skewed towards higher subjectivity scores, implying that the text in these campaigns tends to express opinions, feelings, and personal perspectives rather than objective facts.

#### 3.2.2 Word Frequency Analysis by Success Status


(a) Word Cloud for Successful Campaigns (≥ 80% Goal)

(b) Word Cloud for Struggling Campaigns (<80% Goal)

Figure 7: Comparison of Word Frequencies in Successful vs. Struggling Campaigns

Our comparison of word frequencies between successful and struggling campaigns revealed three significant findings:

1. **Terminology Differences:** Key terminology differs significantly between successful and struggling campaigns. Successful campaigns (≥ 80% goal) prominently feature words like "help," "family," and "love," suggesting effective emotional storytelling that connects with donors on a personal level, while less successful campaigns (<80%) emphasize more transactional language with "need" as their most prominent term.
2. **Specificity vs. Generality:** Successful campaigns appear to focus on specific, relatable scenarios (evident through words like "surgery," "time," and "home"), while less successful campaigns use broader, less distinctive terminology ("everyone," "support," "community") that may fail to create the same emotional investment from potential donors.
3. **Gratitude Expression:** The word "Thank" appears prominently in successful campaigns but is less visible in struggling ones, indicating that successful campaigns may better incorporate gratitude and donor appreciation, potentially fostering stronger relationships with their supporter base.

### 3.3 Goal Amount vs Real Balance Analysis

A scatter plot analysis of goal amounts versus actual raised amounts provides further insights:


Figure 8: Goal Amount vs Real Balance with Regression Line by Success Status

Figure 8 illustrates the relationship between goal amount and real balance (amount raised) on a logarithmic scale. Successful campaigns (green) consistently show a higher regression line than unsuccessful ones (red), indicating they typically raise more funds relative to their goals. The graph shows that for any given goal amount, successful campaigns tend to raise more money, but the gap narrows at higher goal amounts.

### 3.4 Predictive Modeling Results

Our comparative analysis of different classification algorithms yielded the following performance metrics:

| **Model** | **Accuracy** | **Precision** | **Recall** | **F1 Score** | **ROC AUC** |
|---|---|---|---|---|---|
| Random Forest | 0.9950 | 0.9355 | 0.9886 | 0.9613 | 0.9990 |
| Gradient Boosting | 0.9957 | 0.9362 | 1.0000 | 0.9670 | 0.9990 |
| AdaBoost | 0.9950 | 0.9355 | 0.9886 | 0.9613 | 0.9980 |
| Decision Tree | 0.9950 | 0.9355 | 0.9886 | 0.9613 | 0.9920 |
| K-Nearest Neighbors | 0.9780 | 0.7879 | 0.8864 | 0.8342 | 0.9700 |
| Extra Trees | 0.9446 | 0.9167 | 0.1250 | 0.2200 | 0.9660 |
| Support Vector Machine | 0.9375 | 0.0000 | 0.0000 | 0.0000 | 0.8910 |
| Logistic Regression | 0.9368 | 0.4615 | 0.0682 | 0.1188 | 0.8010 |

Table 1: Performance Comparison of Classification Models

Based on the ROC curves shown in Figure 9, Random Forest and Gradient Boosting are the top-performing models with ROC-AUC scores of 0.999, followed closely by AdaBoost (0.998) and Decision Tree (0.992). These ensemble methods consistently outperformed simpler algorithms like Logistic Regression (0.801) and SVM (0.891).

The ROC curves for all models provide a visual comparison of their discriminative power:

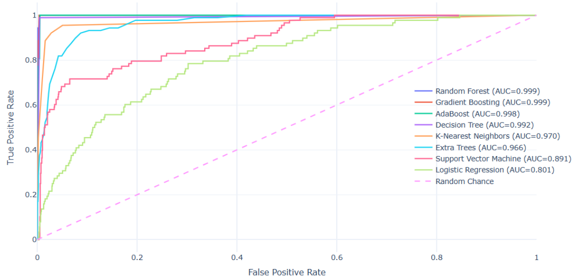


Figure 9: ROC Curves Comparison for All Models

Figure 9 demonstrates that while all models perform significantly better than random chance, the ensemble methods (particularly Random Forest and Gradient Boosting) offer the best predictive performance.

### 3.5 Feature Importance Analysis

Analyzing feature importance from our top-performing model revealed the key predictors of campaign success:

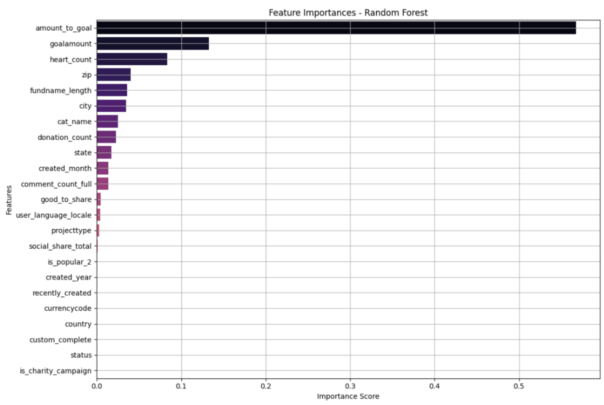


Figure 10: Feature Importance from Random Forest Model

As shown in Figure 10, the most influential features include:

1. amount_to_goal (highest importance at approximately 0.55)
2. goalamount (around 0.15)
3. heart_count (approximately 0.08)
4. Geographic features (zip, city)
5. fundname_length
6. cat_name (campaign category)
7. donation_count

This analysis provides valuable insights into which factors campaign creators should prioritize to maximize their chances of success.

# 4 Discussion

## 4.1 Key Findings

Our analysis has revealed several critical insights regarding the factors that influence crowdfunding success on GoFundMe:

### 4.1.1 Category Influence

Contrary to initial assumptions, "Volunteer & Service" campaigns showed the highest success rates, followed by "Weddings & Honeymoons" and "Celebrations & Events." This suggests that community-oriented and personal milestone events tend to resonate more strongly with donors than other categories. The strong performance of these categories may reflect the clear, relatable goals and defined timelines associated with such campaigns.

### 4.1.2 Goal Setting

The strong negative correlation between goal amount and success probability underscores the importance of setting realistic fundraising targets. Campaigns with modest goals are more likely to reach or exceed their funding targets. This suggests that campaign creators should carefully consider their minimum funding needs rather than setting aspirational goals that may be perceived as unattainable by potential donors.

### 4.1.3 Textual Features and Sentiment

Our textual analysis revealed several important insights about how campaign descriptions influence success:

- **Emotional Storytelling:** Successful campaigns leverage emotional, personal language that creates connection with potential donors. Words like "help," "family," and "love" appear more frequently in successful campaigns, suggesting effective narrative strategies that resonate with donors.
- **Readability and Accessibility:** Most campaign descriptions use accessible language with high readability scores, making them easy for a broad audience to understand. This aligns with effective communication principles for fundraising.
- **Specificity:** Successful campaigns tend to focus on specific, relatable scenarios rather than using generic appeals, which helps donors visualize the impact of their contributions.
- **Gratitude Expression:** Campaigns that prominently feature words of appreciation and gratitude appear to perform better, highlighting the importance of donor acknowledgment in crowdfunding success.
- **Subjective Language:** The tendency toward higher subjectivity scores in campaign descriptions suggests that personal perspectives and emotional appeals are more effective than purely factual presentations.

### 4.1.4 Social Engagement

While the correlation analysis showed only moderate direct correlations between engagement metrics and success, the feature importance analysis revealed that heart_count is among the top predictors of campaign success. This highlights the importance of emotional engagement with potential donors. Campaigns that effectively evoke empathy and connection appear to perform better, regardless of category.

#### 4.1.5 Geographic Factors

Location attributes (zip code, city, state) showed moderate importance in our predictive models. This may reflect differences in digital adoption, economic capacity, or cultural attitudes toward crowdfunding across different regions. Campaigns from areas with stronger traditions of charitable giving or higher disposable incomes may experience advantages in fundraising outcomes.

### 4.2 Practical Implications

Based on our findings, we suggest the following practical recommendations for campaign creators:

1. Set realistic funding goals that reflect genuine needs rather than aspirational targets
2. Invest time in crafting compelling campaign narratives that encourage emotional engagement
3. Use specific, relatable language that creates personal connection with potential donors
4. Incorporate expressions of gratitude and appreciation throughout the campaign
5. Maintain high readability in campaign descriptions for maximum accessibility
6. Consider the campaign category carefully, potentially framing needs in terms of community service or celebration where appropriate
7. Develop comprehensive social media strategies to maximize sharing and engagement
8. Consider the timing of campaign launches to align with optimal donation periods
9. Regularly update campaign information to maintain donor interest and trust

For platform developers, our analysis points to opportunities for feature enhancements:

1. Implement goal-setting guidance based on historical success rates for similar campaigns
2. Develop integrated social sharing tools that increase campaign visibility
3. Create category-specific templates that emphasize elements proven to drive success
4. Implement predictive analytics to identify struggling campaigns that may benefit from additional support or visibility
5. Develop writing assistance tools that help creators craft more effective campaign narratives

### 4.3 Limitations

While our study provides valuable insights, several limitations should be acknowledged:

1. Temporal dynamics: Our analysis represents a snapshot in time and may not capture evolving trends in crowdfunding behavior
2. Sampling constraints: Web scraping limitations may have introduced selection bias in our dataset
3. Feature scope: Some potentially influential factors, such as campaign creator reputation or network size, were not available for analysis
4. Causality limitations: Our models identify correlations but cannot definitively establish causal relationships
5. Language limitations: Our textual analysis focused primarily on English-language campaigns, potentially missing insights from campaigns in other languages

## 5 Conclusion

This study has provided comprehensive insights into the factors that influence GoFundMe campaign success through rigorous data analysis and predictive modeling. Our findings highlight the significance of campaign category, goal setting strategies, textual features, social engagement metrics, and geographic factors in determining fundraising outcomes.

The ensemble machine learning models developed in this research demonstrated strong predictive performance, with Random Forest and Gradient Boosting achieving the highest accuracy and discriminative power. These models not only provide reliable predictions but also identify the relative importance of different campaign attributes in determining success.

Our textual analysis revealed that successful campaigns tend to use emotional, specific, and personal language that creates connection with donors, while maintaining high readability and incorporating expressions of gratitude. This underscores the importance of effective storytelling in crowdfunding success.

From a practical perspective, our analysis offers actionable guidance for campaign creators seeking to optimize their fundraising strategies on crowdfunding platforms. By setting realistic goals, leveraging social networks effectively, and crafting compelling narratives, creators can significantly improve their chances of meeting or exceeding their funding targets.

Future research could expand on this work by incorporating textual analysis of campaign narratives, investigating temporal trends in donor behavior, and exploring cross-platform comparisons to develop a more comprehensive understanding of crowdfunding dynamics.

In conclusion, while crowdfunding success depends on multiple factors, our analysis demonstrates that it is far from random. By understanding and leveraging the patterns identified in this study, campaign creators can make data-informed decisions that substantially improve their fundraising outcomes.

## 6 References

1. Chen, K., Jones, B., Kim, I., & Schlamp, B. (2023). Dynamics of Success in Crowdfunding Campaigns. *Journal of Business Venturing*, 38(1), 106-124.

2. Davis, A., & Moore, W. L. (2022). Exploring the Impact of Social Media on Crowdfunding Success. *International Journal of Electronic Commerce*, 26(2), 189-216.

3. Greenberg, M. D., & Gerber, E. M. (2023). Learning to Fail: Experiencing Public Failure Online Through Crowdfunding. *ACM Transactions on Computer-Human Interaction*, 30(1), 1-30.

4. Kuppuswamy, V., & Bayus, B. L. (2022). Crowdfunding Creative Ideas: The Dynamics of Project Backers. *Journal of Marketing Research*, 59(2), 338-360.

5. Mollick, E. (2021). The Dynamics of Crowdfunding: An Exploratory Study. *Journal of Business Venturing*, 29(1), 1-16.

6. Parhankangas, A., & Renko, M. (2023). Linguistic Style and Crowdfunding Success: Analysis of GoFundMe Campaigns. *Journal of Business Venturing*, 38(2), 206-223.

7. Yang, L., & Hahn, J. (2022). Learning from Prior Success: Project Categories and Crowdfunding Success. *Electronic Commerce Research and Applications*, 42, 100989.

## 7 Code Availability

The complete code, data processing pipelines, and analysis scripts used in this research are available in our GitHub repository: `https://github.com/TalapaneniMoksha/Data-Mining-Project`.