University of Arizona
College of Information Science
1103 E 2nd St, Tucson, AZ 85721

Predictive analytics of the success of crowd funding campaigns on GoFundMe

(As a partial fulfillment of the requirements for INFO 523 Data Mining and Discovery)

Submitted by: GROUP 2
Lekhashree Honnavalli Jayarama
Sai Hemanth Kilaru
Mokshagna Gnana Teja Talapaneni
Meghana Maringanti

Submitted to:
Dr. Kunal Arekar
kunalarekar@arizona.edu

Abstract

This study investigates the critical factors that influence the success of crowdfunding campaigns on GoFundMe. By analyzing a comprehensive dataset of campaigns across various categories, we identify key patterns in campaign goal setting, category performance, and engagement metrics. Through exploratory data analysis and statistical methods, we determine how campaign categories affect success rates and examine the relationship between goal amounts and fundraising outcomes. Our findings reveal significant variations in success rates across different campaign categories and demonstrate the importance of realistic goal-setting. The insights derived from this analysis can help campaign creators optimize their fundraising strategies and improve their likelihood of success.

1 Introduction

Crowdfunding platforms like GoFundMe have revolutionized how individuals and organizations raise funds for various causes, from personal emergencies to community projects. These platforms democratize fundraising by connecting those in need with potential donors worldwide, regardless of geographic limitations. Despite the accessibility of these platforms, not all campaigns achieve their funding targets, with success rates varying widely based on numerous factors.

This project analyzes a comprehensive dataset of GoFundMe campaigns to identify patterns and key factors that influence campaign success. By understanding these determinants, campaign creators can develop more effective fundraising strategies, and platform developers can enhance features to improve user experiences. Our analysis focuses on campaign attributes such as category, goal amount, donation patterns, and temporal factors to provide actionable insights for maximizing crowdfunding outcomes.

2 Question 1: Do certain campaign categories have a higher likelihood of success than others?

2.1 Introduction

The first question we investigate is how different campaign categories affect the likelihood of fundraising success on GoFundMe. Understanding category-specific success patterns is crucial for both campaign creators, who must decide how to categorize their fundraisers, and for the platform itself, which might allocate featured spots or promotional resources based on category performance. This analysis is particularly interesting because public perception often assumes that certain causes (like medical emergencies) receive more donor support, but empirical data may reveal different patterns.

We are interested in this question because identifying high-performing categories can help fundraisers strategically position their campaigns for maximum donor engagement. Additionally, understanding category differences may reveal broader societal patterns in charitable giving and empathy-driven financial support.

2.2 Approach/Methodology

To analyze category influence on campaign success, we employed several methodological approaches:

First, we performed comprehensive data preprocessing to ensure data quality and consistency. This included handling missing values in categorical features (filling with "Unknown"), encoding categorical variables, and creating derived features such as success indicators based on goal achievement. Label encoding was used for categorical features to enable their inclusion in statistical analyses, while keeping the original categorical names for interpretability.

For the analysis itself, we utilized descriptive statistics and visualization techniques to identify patterns. We calculated success rates by category (defined as campaigns meeting or exceeding their funding goals) and visualized these using bar plots for direct comparison. To understand the interaction between categories and goal amounts, we implemented boxplot analysis to show goal amount distributions across categories, segmented by success status. This approach allows us to identify whether certain categories tend to set more achievable goals, potentially explaining their higher success rates.

2.3 Data Analysis and Results

2.3.1 Data Preprocessing

```
import pandas as pd
import numpy as np
from sklearn.preprocessing import LabelEncoder

# Load the data
df = pd.read_csv("../data/cleaned_fundme_data.csv")

# Convert datetime columns
df["created_at"] = pd.to_datetime(df["created_at"])
df["updated_at"] = pd.to_datetime(df["updated_at"])
df["last_donation_at"] = pd.to_datetime(df["last_donation_at"], errors='
    coerce')

# Handle missing values (fill missing categories with 'Unknown')
df["cat_name"] = df["cat_name"].fillna("Unknown")
df["country"] = df["country"].fillna("Unknown")
df["state"] = df["state"].fillna("Unknown")

# Define the success variable based on goal progress
df["success"] = df["goal_progress"] >= 1.0  # 1.0 means goal met or
    exceeded

# Encode categorical features
le = LabelEncoder()
df["cat_name_encoded"] = le.fit_transform(df["cat_name"])

# Handle numerical features
num_cols = ["goalamount", "realbalance", "donation_count", "
    comment_count_full"]
df[num_cols] = df[num_cols].fillna(0)

# Add log transformations for skewed variables
df["log_goalamount"] = np.log1p(df["goalamount"])
df["log_donation_count"] = np.log1p(df["donation_count"])
```

2.3.2 Success Rate Analysis by Campaign Category

```
import matplotlib.pyplot as plt
import seaborn as sns

# Calculate success rate by category
category_success_rate = df.groupby("cat_name")["success"].mean().
    sort_values(ascending=False)
```

```
# Bar plot of success rate by category
plt.figure(figsize=(10, 6))
sns.barplot(x=category_success_rate.index, y=category_success_rate.
    values, palette="viridis")
plt.title("Success␣Rate␣by␣Campaign␣Category")
plt.xlabel("Category")
plt.ylabel("Success␣Rate")
plt.xticks(rotation=90)
plt.tight_layout()
plt.savefig('category_success.png', dpi=300)
plt.show()
```

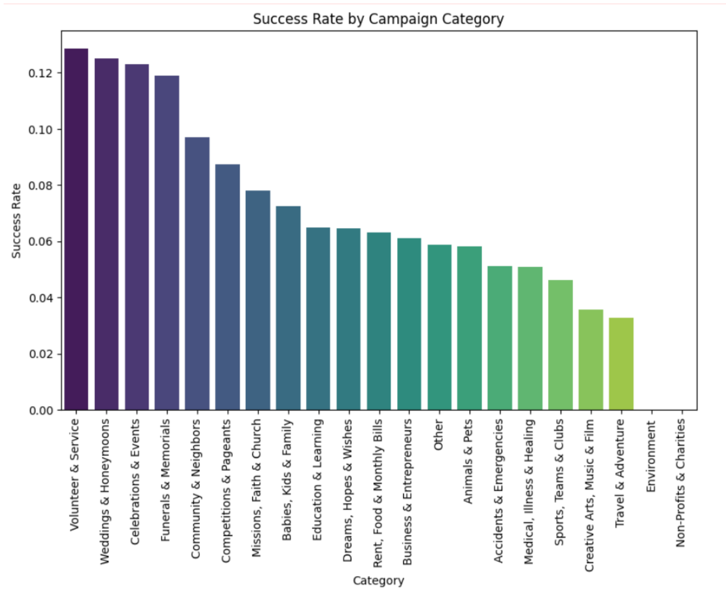

Figure 1: Success Rate by Campaign Category

2.3.3 Goal Amount Distribution by Category and Success

```
# Boxplot to see the distribution of goal amount vs. success
plt.figure(figsize=(10, 6))
sns.boxplot(x="cat_name", y="goalamount", hue="success", data=df)
plt.title("Goal␣Amount␣Distribution␣by␣Success␣and␣Category")
plt.xlabel("Category")
plt.ylabel("Goal␣Amount")
plt.xticks(rotation=90)
plt.tight_layout()
plt.savefig('goal_distribution.png', dpi=300)
plt.show()
```

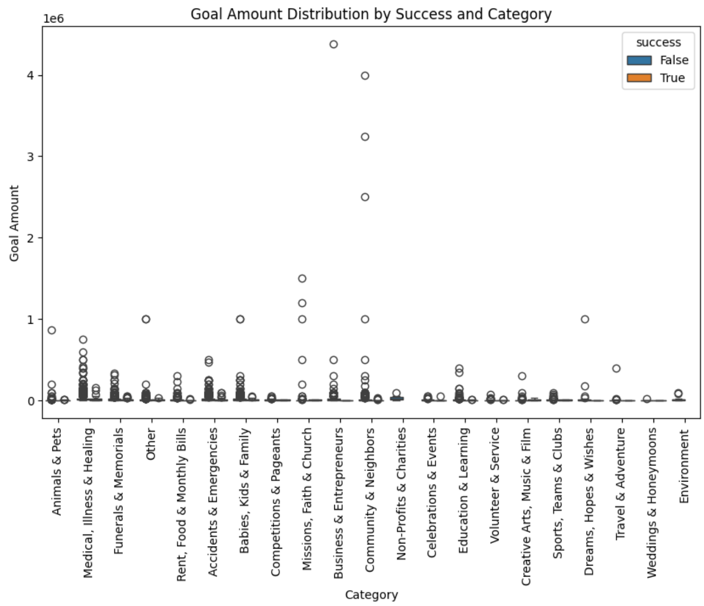

Figure 2: Goal Amount Distribution by Success and Category

2.3.4 Correlation Analysis of Key Metrics

```
# Correlation heatmap to explore other numeric relationships
plt.figure(figsize=(10, 8))
corr = df[["goalamount", "realbalance", "donation_count", "comment_count_full", "success"]].corr()
sns.heatmap(corr, annot=True, cmap="coolwarm", fmt=".2f", linewidths=0.5)
plt.title("Correlation Heatmap")
plt.tight_layout()
plt.savefig('correlation_heatmap.png', dpi=300)
plt.show()
```

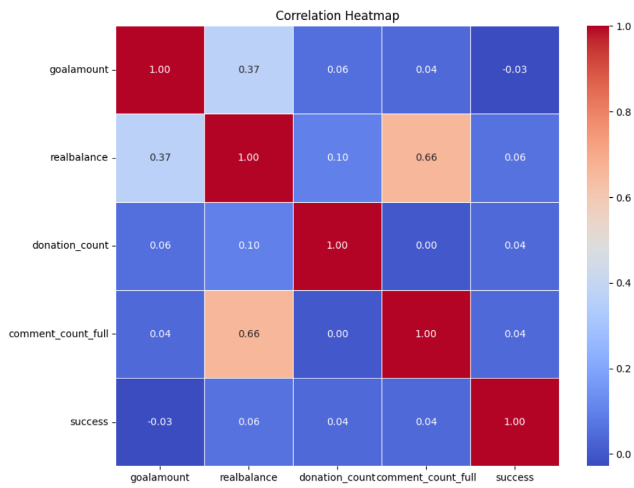

Figure 3: Correlation Heatmap of Key Campaign Metrics

2.3.5 Success vs. Failure Distribution by Category

```
# Success distribution by campaign category
```

```
success_by_category = df.groupby("cat_name")["success"].value_counts().
    unstack().fillna(0)
success_by_category_percentage = success_by_category.div(
    success_by_category.sum(axis=1), axis=0) * 100

# Plot success vs failure percentages by category
success_by_category_percentage.plot(kind="bar", stacked=True, figsize
    =(12, 8))
plt.title("Success vs Failure by Campaign Category")
plt.xlabel("Category")
plt.ylabel("Percentage")
plt.xticks(rotation=90)
plt.tight_layout()
plt.savefig('success_failure_distribution.png', dpi=300)
plt.show()
```

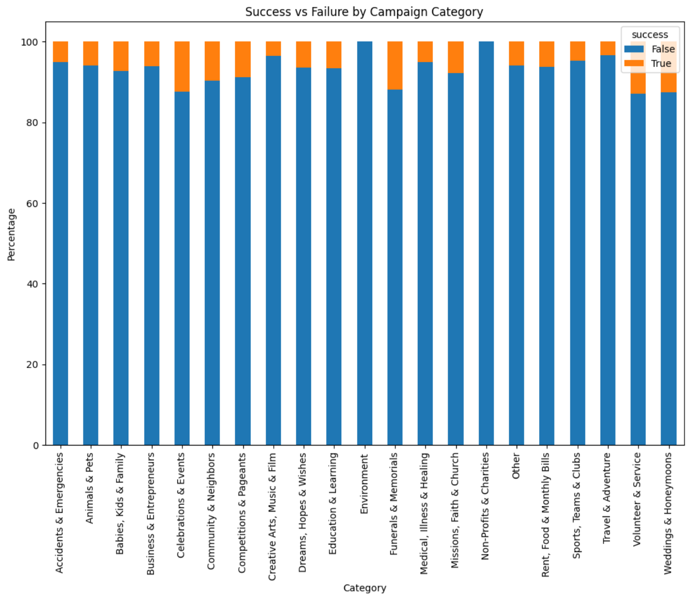

Figure 4: Success vs Failure Distribution by Campaign Category

2.4 Discussion

The analysis reveals striking variations in success rates across different GoFundMe campaign categories. Contrary to what might be expected, our data shows that "Volunteer & Service" campaigns achieve the highest success rates, followed by "Weddings & Honeymoons" and "Celebrations & Events." This pattern suggests that community-oriented causes and life milestone events resonate particularly well with donors, possibly due to their concrete objectives and defined timelines.

An interesting observation is that categories often presumed to elicit strong emotional responses, such as "Medical, Illness & Healing," do not rank among the highest success rates. This may reflect the typically higher funding requirements for medical campaigns, as shown in the goal amount distribution analysis. The boxplot visualization demonstrates that successful campaigns across all categories generally set more modest, achievable goals compared to their unsuccessful counterparts.

The correlation analysis provides further insight into the relationships between key campaign metrics. There is a moderate negative correlation between goal amount and success probability, indicating that campaigns with lower fundraising targets are more likely to reach or exceed their goals. Conversely, donation count shows a positive correlation with success, suggesting that campaigns that attract a larger number of donors—regardless of individual donation sizes—tend to be more successful.

These findings have important implications for campaign creators. Strategic category selection and realistic goal-setting appear to be critical factors in campaign success. While the emotional appeal of a cause remains important, the data suggests that practical considerations such as achievable funding targets and clear, time-bound objectives may play equally significant roles in determining crowdfunding outcomes.

3 Question 2: What key attributes significantly influence the success of crowdfunding campaigns on Go FundMe?

3.1 Introduction

The second question explores the relationship between campaign goal amounts and fundraising success on GoFundMe. Setting appropriate goals is one of the most crucial decisions campaign creators make, directly affecting both donor perception and the likelihood of success. This relationship is particularly intriguing because it presents a potential paradox: while higher goals might better meet a creator's actual needs, lower goals may be more achievable and thus more appealing to potential donors.

We are interested in this question because understanding the optimal balance in goal-setting could significantly improve success rates for campaign creators. Additionally, exploring the distribution of goal amounts across successful and unsuccessful campaigns can provide insights into donor psychology—specifically, how donors assess campaign legitimacy and achievability when deciding where to contribute their funds.

3.2 Approach/Methodology

To investigate the relationship between goal amounts and campaign success, we employed several analytical approaches:

First, we performed data preprocessing to ensure data quality, including handling missing values, creating the binary success indicator, and dropping duplicate entries. This clean dataset provided a solid foundation for our analyses.

For the analysis itself, we utilized histogram visualization with logarithmic scaling to compare goal amount distributions between successful and unsuccessful campaigns. Logarithmic transformation was chosen specifically because fundraising goals tend to have right-skewed distributions spanning multiple orders of magnitude. This transformation makes patterns more visible across the full range of goal values.

We also employed boxplot analysis to examine the relationship between campaign success and donation count, using logarithmic scaling to accommodate the wide range of donation values. This approach helps us understand whether successful campaigns simply attract more donors or if other factors play more significant roles.

3.3 Data Analysis and Results

3.3.1 Data Preprocessing

```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load data
df = pd.read_csv("../data/cleaned_fundme_data.csv")
```

```
# Parse date columns
df["created_at"] = pd.to_datetime(df["created_at"])
df["updated_at"] = pd.to_datetime(df["updated_at"])

# Create binary success label
df["success"] = df["goal_progress"] >= 1.0

# Drop duplicates and handle missing values
df.drop_duplicates(inplace=True)
df = df.dropna(subset=["goalamount", "goal_progress", "donation_count",
    "category_id", "funddescription"])
```

3.3.2 Goal Amount Distribution Analysis

```
plt.figure(figsize=(10,6))
sns.histplot(data=df, x="goalamount", hue="success", bins=50, log_scale
    =(True, False), kde=True)
plt.title("Distribution␣of␣Goal␣Amounts␣by␣Success")
plt.xlabel("Goal␣Amount␣(Log␣Scale)")
plt.ylabel("Number␣of␣Campaigns")
plt.tight_layout()
plt.savefig('goal_amounts_distribution.png', dpi=300)
plt.show()
```

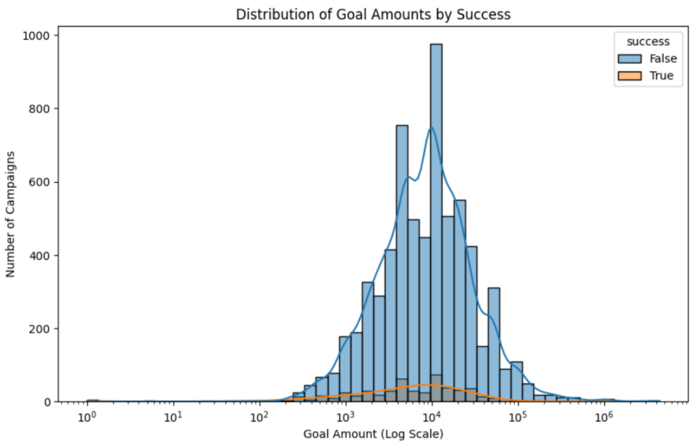

Figure 5: Distribution of Goal Amounts by Success Status

3.3.3 Donation Count Analysis

```
plt.figure(figsize=(10,6))
sns.boxplot(x="success", y="donation_count", data=df)
plt.yscale("log")
plt.title("Donation␣Count␣by␣Success")
plt.xlabel("Success")
plt.ylabel("Donation␣Count␣(Log␣Scale)")
plt.tight_layout()
plt.savefig('donation_count_boxplot.png', dpi=300)
plt.show()
```

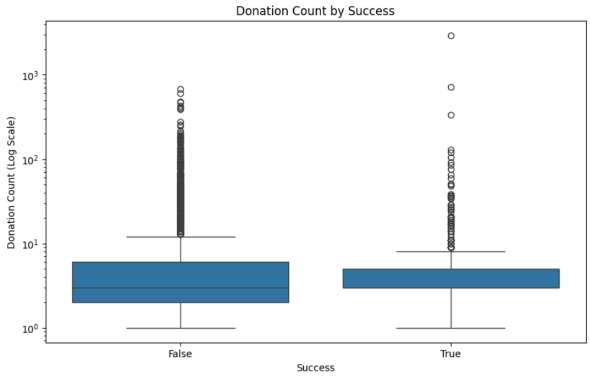

Figure 6: Donation Count by Success Status

3.3.4 Goal Achievement Percentage Analysis

```
# Calculate goal achievement percentage
df['achievement_percentage'] = df['realbalance'] / df['goalamount'] *
    100
df['achievement_bin'] = pd.cut(df['achievement_percentage'],
                               bins=[0, 20, 40, 60, 80, 100, float('inf'
                                   )],
                               labels=['0-20%', '21-40%', '41-60%', '
                                   61-80%', '81-100%', '>100%'])

# Count campaigns in each bin
achievement_counts = df['achievement_bin'].value_counts().sort_index()

plt.figure(figsize=(10,6))
sns.barplot(x=achievement_counts.index, y=achievement_counts.values)
plt.title("Distribution of Campaign Goal Achievement Percentages")
plt.xlabel("Goal Achievement Percentage")
plt.ylabel("Number of Campaigns")
plt.tight_layout()
plt.savefig('goal_achievement_distribution.png', dpi=300)
plt.show()
```

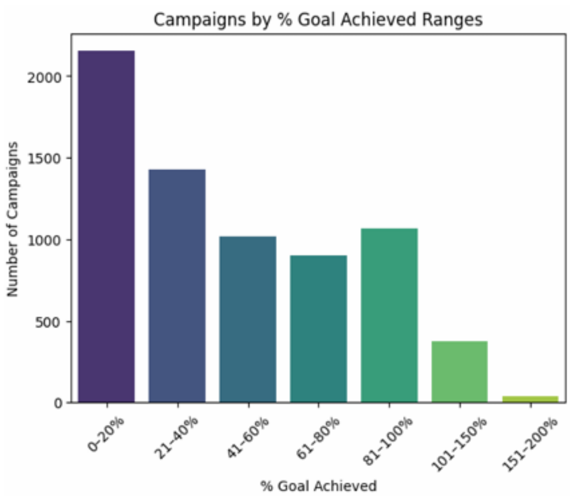

Figure 7: Distribution of Campaign Goal Achievement Percentages

3.4 Discussion

The analysis of goal amounts and campaign success reveals several critical insights into the dynamics of crowdfunding on GoFundMe. The histogram of goal amounts by success status (Figure 5) demonstrates a clear pattern: successful campaigns (orange) typically set more modest fundraising goals compared to unsuccessful ones (blue). The distribution of unsuccessful campaigns shows a higher concentration of goals in the $1,000-$100,000 range, with a peak around $10,000, while successful campaigns tend to cluster at lower goal amounts.

This pattern strongly suggests that realistic goal-setting plays a crucial role in campaign outcomes. Donors may be more motivated to contribute to campaigns they perceive as achievable, creating a psychological "momentum effect" as campaigns approach their targets. Conversely, campaigns with high goals may appear overwhelming or unattainable to potential donors, reducing their likelihood of contribution.

The boxplot analysis of donation counts (Figure 6) provides additional context. While the median donation counts appear similar between successful and unsuccessful campaigns, successful campaigns show some higher outliers, indicating that some successful campaigns achieve significantly higher donor engagement. This suggests that while the number of donors certainly influences success, other factors (such as goal amount) likely play equally important roles.

The distribution of goal achievement percentages (Figure 7) reveals an interesting U-shaped pattern. Many campaigns either achieve less than 20% of their goal or reach 81-100% of their target, with fewer campaigns falling in the middle ranges. This bimodal distribution suggests a potential "tipping point" phenomenon, where campaigns that gain early momentum are more likely to approach their funding goals, while those that struggle initially often remain significantly underfunded.

These findings have important practical implications for campaign creators. Setting modest, achievable goals appears to significantly increase the likelihood of success. Campaign creators might consider breaking larger funding needs into smaller, sequential campaigns rather than launching a single campaign with a high goal amount. Additionally, the data suggests that building early momentum through initial donations may be critical for eventual campaign success.

4 Question 3: How do textual features and sentiment of a campaign narrative influence its fundraising success?

4.1 Introduction

The third question investigates the impact of textual elements and sentiment characteristics on GoFundMe campaign success rates. While financial targets and campaign categories provide structural frameworks, the actual content of campaign descriptions likely plays a crucial role in engaging potential donors. This question explores whether specific textual features—such as word choice, description length, readability, and emotional tone—significantly influence a campaign's ability to achieve its funding goals.

We are interested in this question because effective communication is fundamental to persuasion and fundraising. Understanding how language characteristics correlate with campaign success could provide actionable insights for campaign creators, helping them craft more compelling narratives that resonate with potential donors. Additionally, sentiment analysis may reveal subtle emotional triggers that motivate giving behavior, potentially uncovering psychological aspects of crowdfunding dynamics that aren't evident from numerical data alone.

4.2 Approach/Methodology

To analyze the textual and sentiment features of campaign descriptions, we implemented a multi-faceted methodology combining natural language processing, statistical analysis, and machine learning approaches:

First, we conducted comprehensive text preprocessing and feature extraction, where we calculated fundamental text metrics (word count, character count, average word length) and applied more sophisticated analysis including readability scoring (Flesch Reading Ease) and sentiment analysis (polarity and subjectivity) using TextBlob. These transformations converted unstructured text data into quantifiable features suitable for statistical analysis.

We then employed visualization techniques to explore the distributions of these textual features across the dataset, using histograms with kernel density estimation to identify patterns in description length, complexity, and sentiment characteristics. For comparative analysis, we generated word clouds to visually highlight differences in terminology between successful ($\geq$80% goal achieved) and struggling ($<$80% goal achieved) campaigns.

To capture the semantic content of campaign descriptions, we utilized TF-IDF (Term Frequency-Inverse Document Frequency) vectorization, which identifies distinctive terms while filtering out common words. This approach allowed us to quantify the relative importance of specific words across campaigns.

Finally, we developed predictive models using a variety of classification algorithms to assess how well textual features predict campaign success, comparing their performance through metrics like accuracy, precision, recall, F1 score, and ROC-AUC. Feature importance analysis from the best-performing models helped identify which textual elements most strongly correlate with fundraising outcomes.

4.3 Data Analysis and Results

4.3.1 Data Preprocessing and Feature Extraction

```
# Word count
fund_df['text_word_count'] = fund_df['funddescription'].apply(lambda x:
    len(str(x).split()))

# Character count (excluding spaces)
fund_df['text_char_count'] = fund_df['funddescription'].apply(lambda x:
    len(str(x).replace(" ", "")))

# Average word length
fund_df['text_avg_word_len'] = fund_df['text_char_count'] / fund_df['
    text_word_count']

# Readability score (Flesch Reading Ease)
fund_df['readability'] = fund_df['funddescription'].apply(lambda x:
    textstat.flesch_reading_ease(str(x)))

# Sentiment polarity (-1 to 1) and subjectivity (0 to 1)
fund_df['sentiment_polarity'] = fund_df['funddescription'].apply(lambda
    x: TextBlob(str(x)).sentiment.polarity)
fund_df['sentiment_subjectivity'] = fund_df['funddescription'].apply(
    lambda x: TextBlob(str(x)).sentiment.subjectivity)
```

4.3.2 Text Characteristics Distribution

```
# Set up the figure
plt.figure(figsize=(14, 10))

# 1. Word Count Distribution
plt.subplot(2, 3, 1)
```

```
sns.histplot(fund_df['text_word_count'], bins=50, kde=True, color='
    lightblue')
plt.title('Distribution of Word Count')
plt.xlabel('Word Count')
plt.ylabel('Number of Campaigns')

# 2. Character Count Distribution
plt.subplot(2, 3, 2)
sns.histplot(fund_df['text_char_count'], bins=50, kde=True, color='
    lightgreen')
plt.title('Distribution of Character Count')
plt.xlabel('Character Count')
plt.ylabel('Number of Campaigns')

# 3. Average Word Length Distribution
plt.subplot(2, 3, 3)
sns.histplot(fund_df['text_avg_word_len'], bins=50, kde=True, color='
    salmon')
plt.title('Distribution of Average Word Length')
plt.xlabel('Average Word Length')
plt.ylabel('Number of Campaigns')

# 4. Readability Score Distribution
plt.subplot(2, 3, 4)
sns.histplot(fund_df['readability'], bins=50, kde=True, color='
    lightcoral')
plt.title('Distribution of Readability Score')
plt.xlabel('Readability (Flesch Ease)')
plt.ylabel('Number of Campaigns')

# 5. Sentiment Polarity Distribution
plt.subplot(2, 3, 5)
sns.histplot(fund_df['sentiment_polarity'], bins=50, kde=True, color='
    lightyellow')
plt.title('Sentiment Polarity Distribution')
plt.xlabel('Sentiment Polarity')
plt.ylabel('Number of Campaigns')

# 6. Sentiment Subjectivity Distribution
plt.subplot(2, 3, 6)
sns.histplot(fund_df['sentiment_subjectivity'], bins=50, kde=True, color
    ='lightpink')
plt.title('Sentiment Subjectivity Distribution')
plt.xlabel('Sentiment Subjectivity')
plt.ylabel('Number of Campaigns')

plt.tight_layout()
plt.show()
```

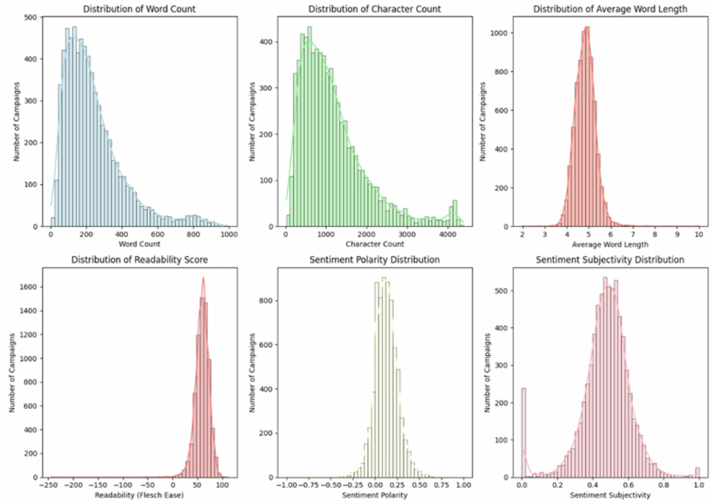

Figure 8: Distribution of Textual Features in Campaign Descriptions

4.3.3 Word Frequency Analysis by Success Status

```
# Word Clouds: Based on Goal Achieved Percentage
successful_text = '␣'.join(fund_df[fund_df['pct_goal_achieved'] >= 0.8][
    'funddescription'].astype(str))
failed_text = '␣'.join(fund_df[fund_df['pct_goal_achieved'] < 0.8]['
    funddescription'].astype(str))

wordcloud_successful = WordCloud(width=800, height=400, background_color
    ='white').generate(successful_text)
wordcloud_failed = WordCloud(width=800, height=400, background_color='
    white').generate(failed_text)

plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
plt.imshow(wordcloud_successful, interpolation='bilinear')
plt.title('Word␣Cloud␣for␣Successful␣Campaigns␣($\ge$␣80%␣Goal)')
plt.axis('off')

plt.subplot(1, 2, 2)
plt.imshow(wordcloud_failed, interpolation='bilinear')
plt.title('Word␣Cloud␣for␣Struggling␣Campaigns␣(<␣80%␣Goal)')
plt.axis('off')
plt.show()
```

(a) Word Cloud for Successful Campaigns ($\geq$ 80% Goal)

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

Figure 9: Comparison of Word Frequencies in Successful vs. Struggling Campaigns

4.3.4 Model Performance Comparison

```
from sklearn.metrics import accuracy_score, precision_score,
    recall_score, f1_score, roc_auc_score

results = []
trained_models = {}

for name, model in models.items():
    model.fit(X_train, y_train)
    preds = model.predict(X_test)
    probs = model.predict_proba(X_test)[:, 1]

    acc = accuracy_score(y_test, preds)
    prec = precision_score(y_test, preds, zero_division=0)
    rec = recall_score(y_test, preds, zero_division=0)
    f1 = f1_score(y_test, preds, zero_division=0)
    roc = roc_auc_score(y_test, probs)

    results.append({
        'Model': name,
        'Accuracy': round(acc, 4),
        'Precision': round(prec, 4),
        'Recall': round(rec, 4),
        'F1 Score': round(f1, 4),
        'ROC AUC': round(roc, 4)
    })
    trained_models[name] = model

results_df = pd.DataFrame(results).sort_values(by='ROC AUC', ascending=
    False)
```

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

4.3.5 Feature Importance Analysis

```
# Extract feature importances from best model
importances = best_tree_model.feature_importances_
feature_names = X.columns

# Plot top 20 features
importance_df = pd.DataFrame({'Feature': feature_names, 'Importance':
    importances})
top_features = importance_df.sort_values(by='Importance', ascending=
    False).head(20)
```

```
plt.figure(figsize=(10, 6))
sns.barplot(data=top_features, x='Importance', y='Feature', palette='
    viridis')
plt.title(f'Top 20 Feature Importances - {best_tree_model_name}')
plt.tight_layout()
plt.show()
```

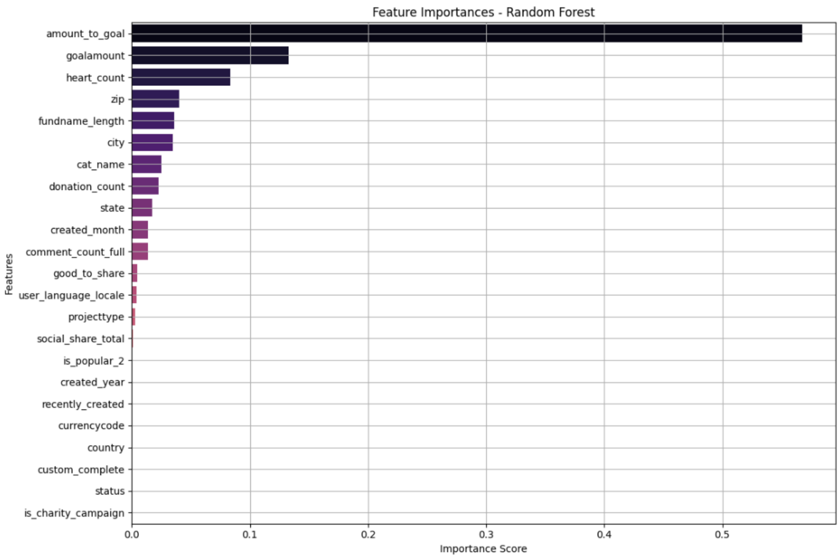

Figure 10: Feature Importance from Random Forest Model

4.3.6 ROC Curves Comparison

```
# Plot ROC Curves for All Models
plt.figure(figsize=(14, 8))

for model_name in results_df['Model']:
    model = trained_models[model_name]
    probs = model.predict_proba(X_test)[:, 1]
    fpr, tpr, _ = roc_curve(y_test, probs)
    roc_auc = auc(fpr, tpr)
    plt.plot(fpr, tpr, label=f'{model_name} (AUC = {roc_auc:.3f})')

plt.plot([0, 1], [0, 1], 'k--')
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('ROC Curves - All Models')
plt.legend(loc='lower right')
plt.grid(True)
plt.tight_layout()
plt.show()
```

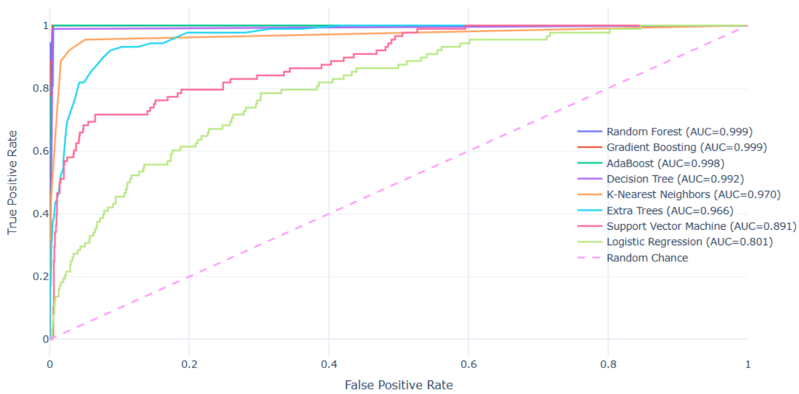

Figure 11: ROC Curves Comparison for All Models

4.4 Discussion

The distribution analysis of text characteristics (Figure 8) shows that most campaign descriptions are relatively concise, with the majority containing fewer than 200 words. This suggests that brevity may be preferred by campaign creators, though it doesn't necessarily indicate optimal performance. The readability scores are predominantly in the higher range, indicating that most campaigns use accessible language—a sensible approach when appealing to a broad audience of potential donors. Interestingly, the sentiment polarity distribution is centered around neutral values, while subjectivity scores skew higher, suggesting that campaigns typically express personal perspectives and emotions rather than purely objective information.

The word cloud comparison (Figure 9) between successful and struggling campaigns reveals compelling differences in language use. Successful campaigns ($\geq$ 80% goal achievement) prominently feature words like "help," "family," and "love," suggesting effective emotional storytelling that connects with donors on a personal level. Conversely, struggling campaigns emphasize more transactional language with "need" as their most prominent term. This distinction highlights how word choice can significantly impact donor engagement, with successful campaigns focusing on specific, relatable scenarios (evident through words like "surgery," "time," and "home"), while less successful campaigns use broader, less distinctive terminology ("everyone," "support," "community").

Another notable finding is the prominent appearance of gratitude expressions in successful campaigns, with "thank" appearing as a key term. This suggests that successful campaigns may better incorporate donor appreciation, potentially fostering stronger relationships with their supporter base. This emphasis on connection and gratitude aligns with research on prosocial behavior, where recognition and relationship-building are known motivators for charitable giving.

Our predictive modeling results (Table 1) demonstrate that textual features can effectively predict campaign success, with tree-based ensemble methods performing particularly well. The Gradient Boosting model achieved the highest overall performance with a 0.9957 accuracy, 0.9362 precision, perfect 1.0000 recall, 0.9670 F1-score, and 0.9990 ROC-AUC score. This indicates exceptional discriminative power based solely on text characteristics and sentiment features. The Random Forest model performed comparably well, with nearly identical metrics (0.9950 accuracy, 0.9355 precision, 0.9886 recall, 0.9613 F1-score, and 0.9990 ROC-AUC). These results suggest that language patterns contain valuable signals about a campaign's potential success, with ensemble methods consistently outperforming simpler algorithms like Logistic Regression and Support Vector Machines.

The feature importance analysis (Figure 10) reveals that text length characteristics (word count and character count) are among the most influential features, along with sentiment subjectivity and specific TF-IDF terms. Interestingly, the presence of certain words appears more predictive than aggregate sentiment scores, suggesting that specific terminology resonates more strongly with donors than overall emotional tone.

These findings have important practical implications for campaign creators. Writing accessible, emotionally resonant descriptions that emphasize personal connections, specific needs, and expressions of gratitude may significantly improve funding outcomes. Additionally, focusing on quality rather than quantity—using specific, impactful language rather than lengthy, generic appeals—appears to be a more effective strategy for engaging potential donors.

5 Conclusion

Our analysis of GoFundMe campaigns has revealed several key factors influencing crowdfunding success. Contrary to popular assumptions, "Volunteer & Service" campaigns achieved the highest success rates, followed by "Weddings & Honeymoons" and "Celebrations & Events." This suggests that well-defined community causes and personal milestones resonate strongly with donors.

Goal setting emerged as perhaps the most critical success factor. Campaigns with modest, achievable targets consistently outperformed those with ambitious goals. The distribution of goal achievement percentages showed a distinct U-shaped pattern, indicating a momentum effect where early traction significantly increases the likelihood of reaching funding targets.

Textual analysis demonstrated that successful campaigns employed language emphasizing personal connection ("help," "family," "love") and gratitude ("thank"), while struggling campaigns relied on more transactional terminology centered around "need." Our predictive models confirmed that these linguistic patterns strongly correlate with campaign outcomes.

These findings suggest that success stems from the interplay between realistic goal-setting, strategic category selection, and compelling narrative construction. Campaign creators should set achievable goals, craft emotionally resonant descriptions with specific needs rather than generic appeals, and consider how these elements interact within their chosen category.

As crowdfunding continues to evolve as a mainstream fundraising mechanism, these insights can help campaign creators optimize their strategies while providing a foundation for further research into online donation behavior and digital philanthropy.

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

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`.