# Disaster Detection: Harnessing NLP and Machine Learning for Crisis Classification in Social Media

Arham Hundia (hundia@purdue.edu)
Sowmya Jayaram Iyer (jayarami@purdue.edu)
Parth Patel (pate1384@purdue.edu)

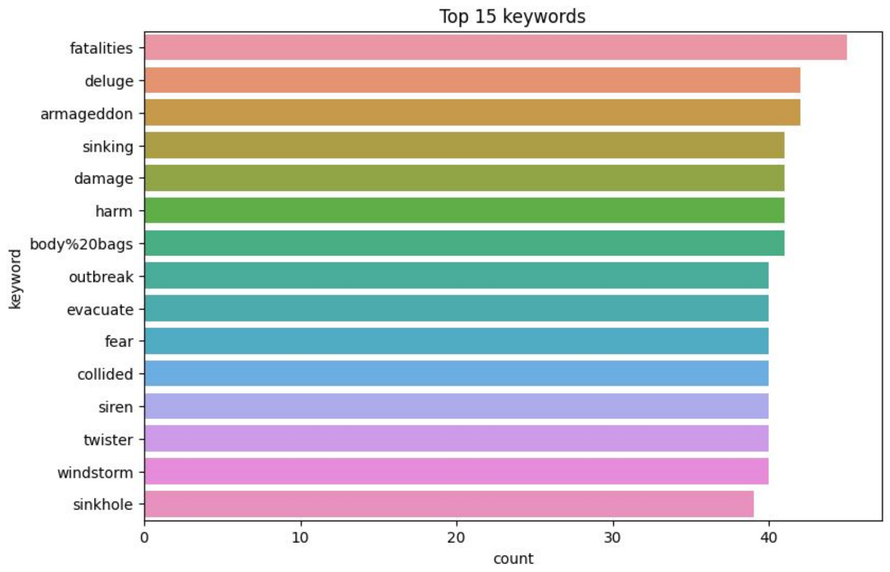
PURDUE UNIVERSITY | Department of Computer Science

CS57700 (Spring 2023)

# INTRODUCTION

- **Objective:** Develop and compare machine learning models for accurate classification of disaster-related tweets
- **Importance:** Timely and accurate information for disaster response and management
- **Models:** Linear SVM, DAN, RNN, LSTM, Bi-LSTM, and BERT
- **Dataset:** A Kaggle competition dataset that consists of 10,000 hand-classified tweet (Target: disaster-related (1) and non-disaster (0))

| | id | keyword | location | text | target |
|---|---|---|---|---|---|
| 0 | 1 | NaN | NaN | Our Deeds are the Reason of this #earthquake M... | 1 |
| 1 | 4 | NaN | NaN | Forest fire near La Ronge Sask. Canada | 1 |
| 2 | 5 | NaN | NaN | All residents asked to 'shelter in place' are ... | 1 |
| 3 | 6 | NaN | NaN | 13,000 people receive #wildfires evacuation or... | 1 |
| 4 | 7 | NaN | NaN | Just got sent this photo from Ruby #Alaska as ... | 1 |






Reference(s): [5]

# INTRODUCTION

The contributions of this paper aim to answer the following questions:

- How can disaster-related tweets be accurately classified using various machine learning and natural language processing techniques?
- How do different techniques perform in the classification task, and which method is the most effective?
- How do simpler models like SVM with TF-IDF perform compared to more complex models like BERT and LSTM?
- What are the limitations of the explored models, and how can they be improved for better performance?

# Methods

- **Preprocessing steps:** Lowercasing, Tokenization, Lemmatization **Feature extraction:** TfidfVectorizer, GloVe word embeddings and BERT Tokenizer
- **Linear SVM:** SVM works by finding an optimal hyperplane that best separates the data points belonging to different classes. SVM minimizes classification errors by maximizing the margin between the classes.
- **DAN:** Deep Averaging Network (DAN) averages the word embeddings of input tokens and feeds the result through a series of fully connected layers.
- **RNN:** Are designed to process sequential data, with connections that allow information to persist from one step of the sequence to the next.
- **LSTM:** Are a type of RNN architecture that use gated cells to regulate the flow of information and overcome the vanishing gradient problem in standard RNNs.
- **Bi-LSTM:** Processes the input sequence in both forward and backward directions using two LSTMs.
- **BERT:** Based on the Transformer architecture that is bidirectionally context-aware, enabling it to understand the context of words using both preceding and following words in a sentence.
- **Evaluation metric:** Test accuracy, precision, recall, and F1-score



Reference(s): [5]

# EXPERIMENTATION

SVM:
Train Accuracy: 0.9775041050903119
Test Accuracy: 0.7977675640183848
Classification Report:

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.80 | 0.87 | 0.83 | 874 |
| 1 | 0.80 | 0.70 | 0.75 | 649 |
| accuracy | | | 0.80 | 1523 |
| macro avg | 0.80 | 0.79 | 0.79 | 1523 |
| weighted avg | 0.80 | 0.80 | 0.80 | 1523 |

**Observation:** reduced performance on the test data could be indicative of overfitting

DAN
Train Accuracy: 0.6218390804597701
Test Accuracy: 0.6014445173998687
Classification Report:

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.62 | 0.77 | 0.69 | 874 |
| 1 | 0.55 | 0.37 | 0.44 | 649 |
| accuracy | | | 0.60 | 1523 |
| macro avg | 0.59 | 0.57 | 0.57 | 1523 |
| weighted avg | 0.59 | 0.60 | 0.58 | 1523 |

**Observation:** ineffectiveness of the DAN model in tasks which require context

RNN
Train Accuracy: 0.7326765188834155
Test Accuracy: 0.7367038739330269
Classification Report:

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.74 | 0.85 | 0.79 | 874 |
| 1 | 0.74 | 0.59 | 0.66 | 649 |
| accuracy | | | 0.74 | 1523 |
| macro avg | 0.74 | 0.72 | 0.72 | 1523 |
| weighted avg | 0.74 | 0.74 | 0.73 | 1523 |

**Observation:** improved language processing capabilities and context understanding of RNN from DAN



Reference(s): [5]

# EXPERIMENTATION

LSTM
Train Accuracy: 0.819047619047619
Test Accuracy: 0.7150361129349967
Classification Report:

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.76 | 0.74 | 0.75 | 874 |
| 1 | 0.66 | 0.67 | 0.67 | 649 |
| accuracy | | | 0.72 | 1523 |
| macro avg | 0.71 | 0.71 | 0.71 | 1523 |
| weighted avg | 0.72 | 0.72 | 0.72 | 1523 |

Bi-LSTM
Train Accuracy: 0.7479474548440066
Test Accuracy: 0.7150361129349967
Classification Report:

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.73 | 0.80 | 0.76 | 874 |
| 1 | 0.69 | 0.60 | 0.64 | 649 |
| accuracy | | | 0.72 | 1523 |
| macro avg | 0.71 | 0.70 | 0.70 | 1523 |
| weighted avg | 0.71 | 0.72 | 0.71 | 1523 |

Train Accuracy: 0.9320197044334976
Test Accuracy: 0.8273145108338804
Classification Report:

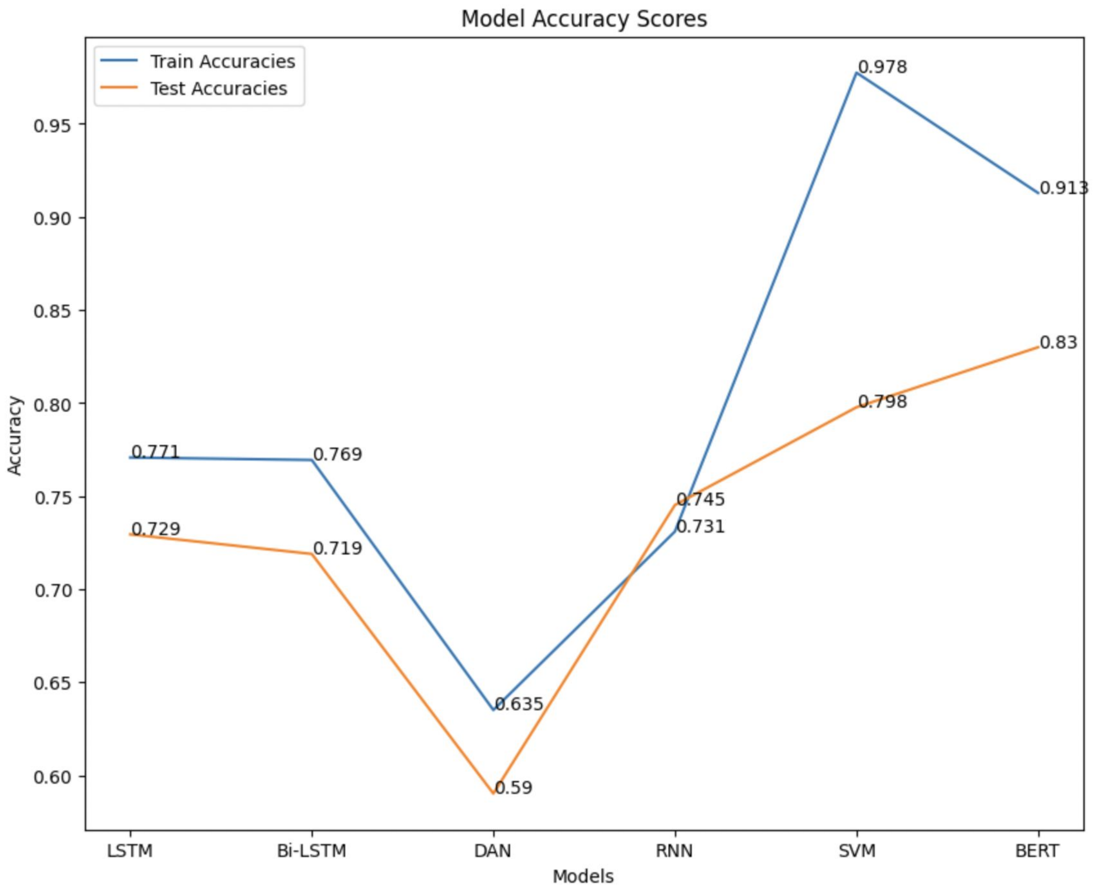
| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.86 | 0.83 | 0.85 | 874 |
| 1 | 0.78 | 0.82 | 0.80 | 649 |
| accuracy | | | 0.83 | 1523 |
| macro avg | 0.82 | 0.83 | 0.82 | 1523 |
| weighted avg | 0.83 | 0.83 | 0.83 | 1523 |

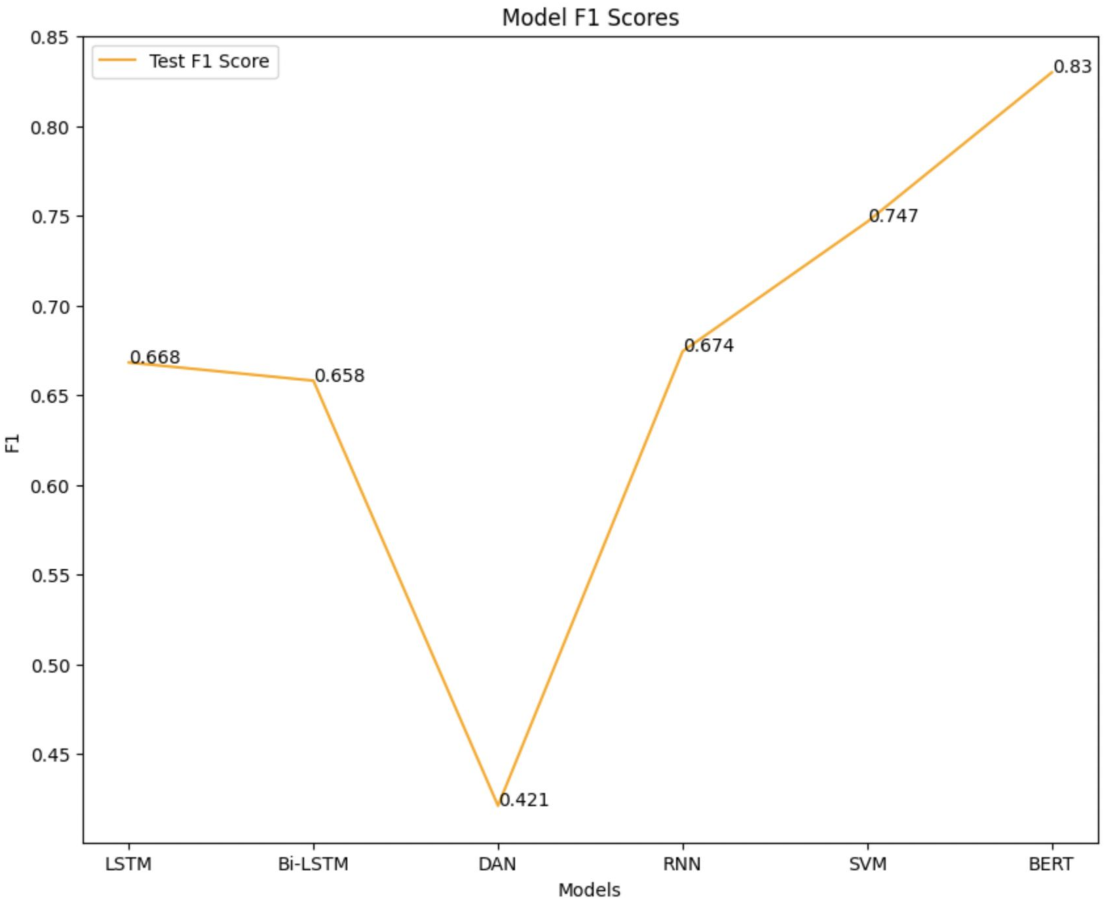
**Observation:** demonstrates the effectiveness of the LSTM model in language processing tasks along with understanding context.

**Observation:** demonstrates the effectiveness of understanding context in both directions for the given input.

**Observation:** the effectiveness of the BERT model with attention mechanism in the binary disaster classification task

# Performance






Reference(s): [5]

# Discussions

- The Linear SVM model shows good performance with test accuracy of 0.79.
- The DAN model falls short with a test accuracy of 0.59.
- The RNN and LSTM models exhibit better language processing capabilities with test accuracies of 0.745 and 0.729, respectively.
- The Bi-LSTM model further enhances the understanding of context by taking into account bidirectional information, achieving a test accuracy of 0.719.
- **Overfitting** observed in some models: Linear SVM, LSTM, and Bi-LSTM
- SVM was able to generalize better than DAN, RNN, LSTM, and Bi-LSTM to the test dataset.
- With transfer learning along with attention mechanism, pre-trained BERT model outperforms all other models, achieving an impressive test accuracy of 83%.
- Overall, our results demonstrate the effectiveness of BERT in NLP tasks that require a deeper understanding of context and reaffirm its position as state-of-the-art.



Reference(s): [5]

# Discussions

- Dataset limitations:
  - Limited size
  - ambiguity
  - subjectivity
- Lack of additional context: External links, non-ASCII characters, emojis
- Language intricacies:
  - Sarcasm,
  - metaphors,
  - misleading keywords or hashtags
- Short length of tweets: Limited context and information

We would do numerous things if we had another month/6 months/full phd time on this subject:

- First, fixing dataset issues like size and imbalance can improve model performance. More data could reduce overfitting and increase generalization.
- We could also leverage attention methods, extra layers,or recent transformer-based models like GPT-3 to better capture context and semantics.
- Addressing ambiguity, subjectivity, sarcasm, metaphors, and misleading hashtags is another topic for improvement.
- Multilingual embeddings or machine translation can assist creating models for more languages and groups



Reference(s): [5]