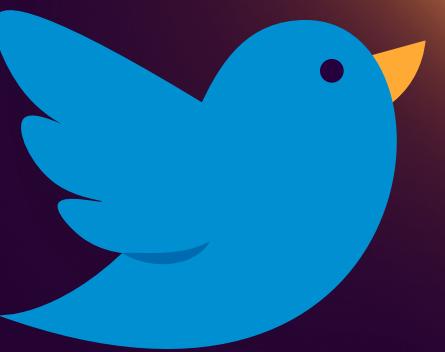


# MACHINE LEARNING

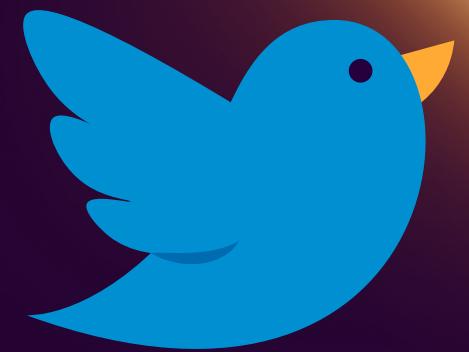
---

NATURAL LANGUAGE PROCESSING  
DISASTER TWEETS

---



“  
The Mavs are on fire after the Dončić trade. Don't expect them to put out the flames anytime soon !!  
”



“

**The fires of hell for Julie #Extant  
writers....she better go down in flames  
when this is all over...#Extant**

”

# THE IMPORTANCE OF DISASTER TWEETS

In many cases, disasters are first reported on Twitter before official news sources. Automating tweet classification can enhance early warning systems and speed up disaster response.

REAL-TIME EMERGENCY  
RESPONSE

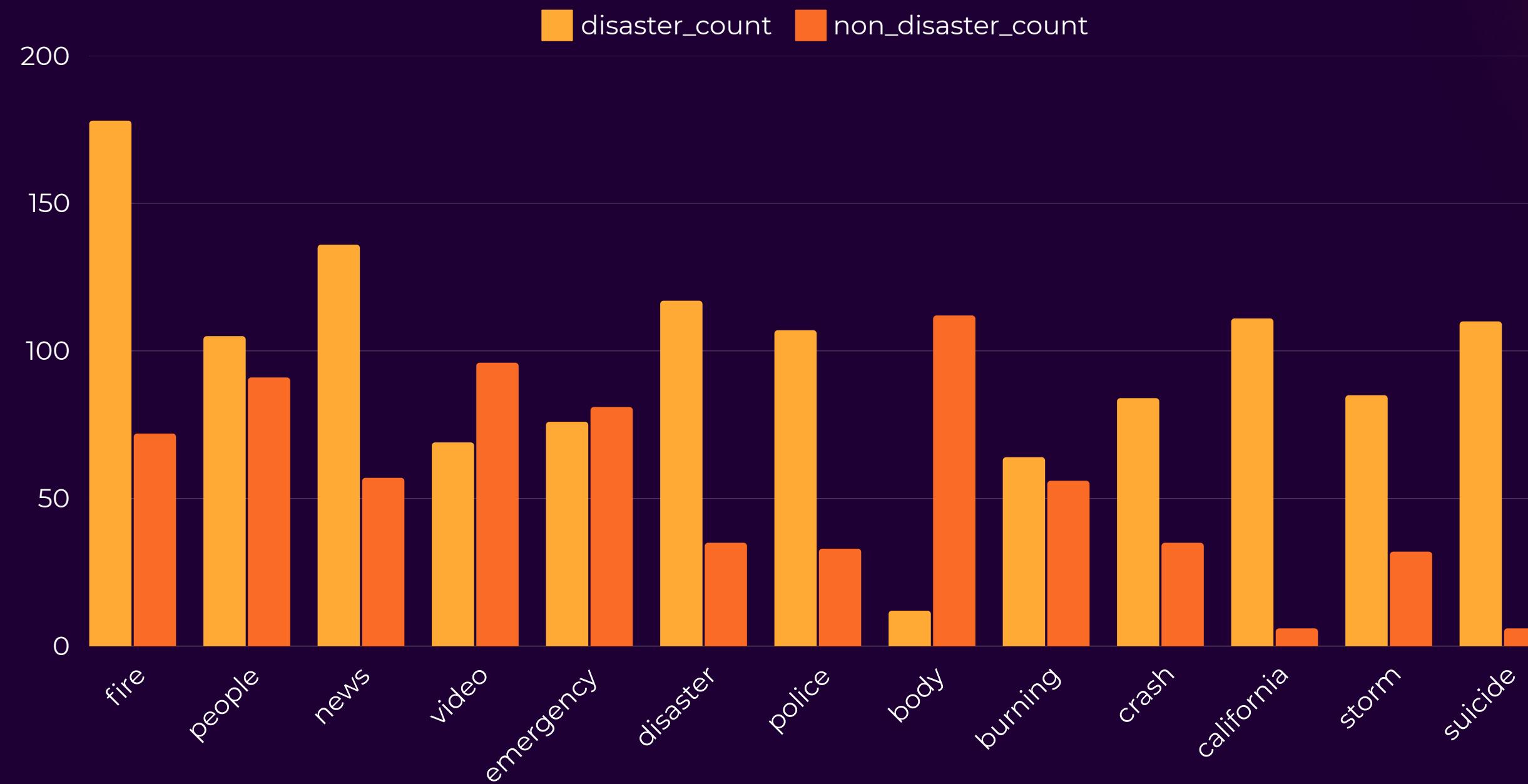
FILTERING OUT NOISE FOR  
RELIEF EFFORTS

COMBATING  
MISINFORMATION & PANIC

# EXPLORATORY DATA ANALYSIS

## Context is crucial

Ambiguous words appear in both categories and hence our model needs context understanding

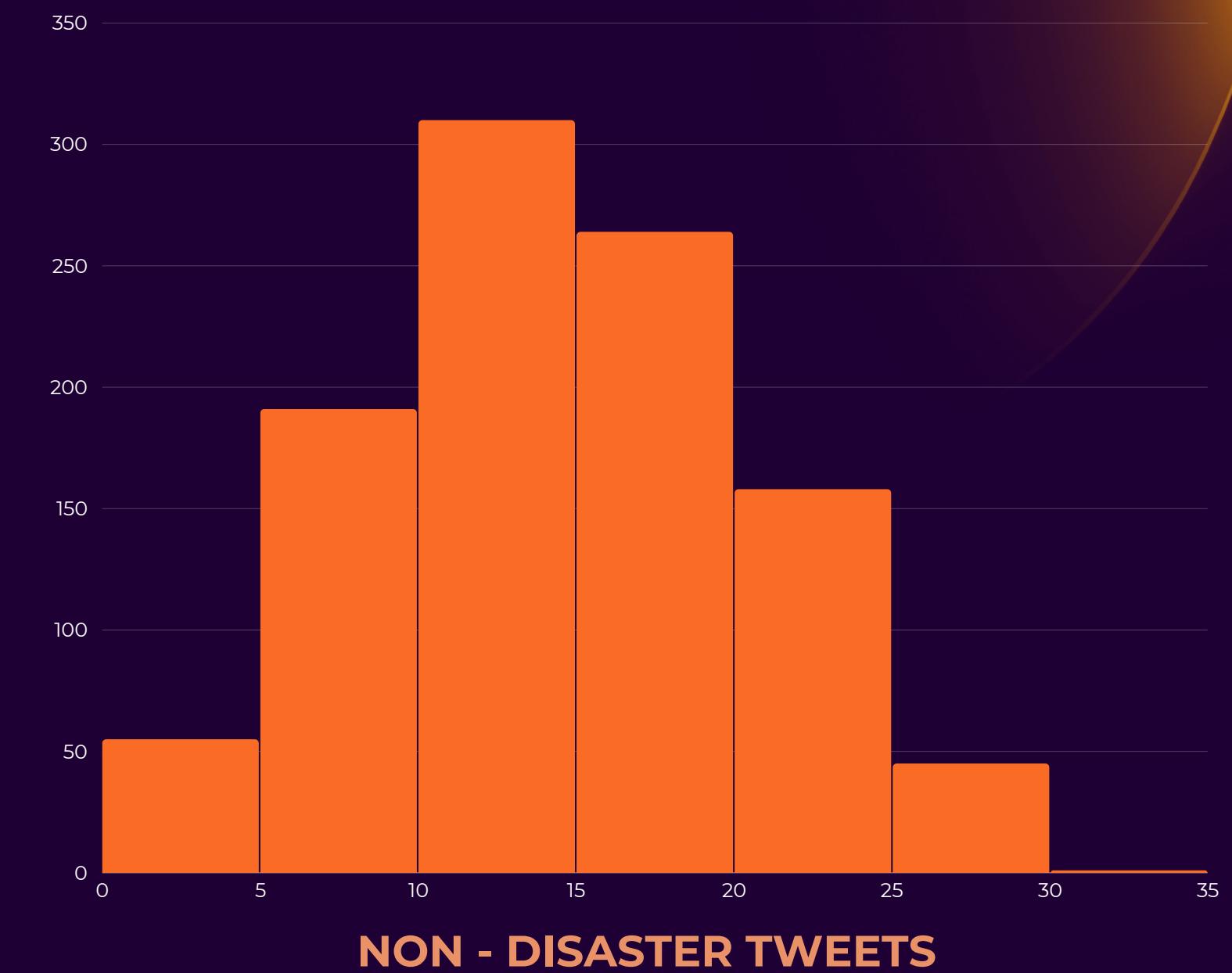
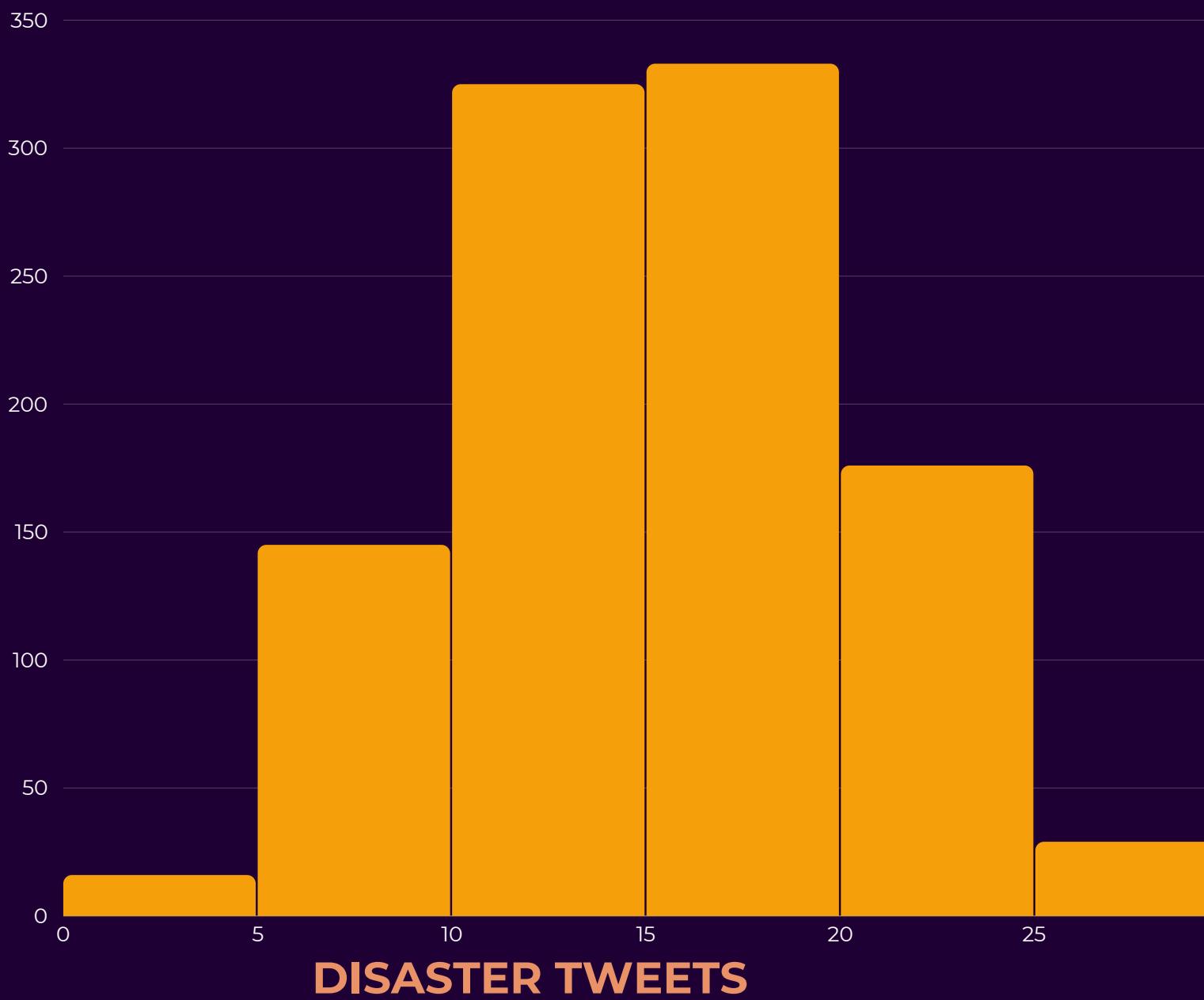


# EXPLORATORY DATA ANALYSIS

## Context Matters More in Longer Tweets

Varying tweet lengths require different processing approaches.

Disaster tweets often provide contextual clues



Words per tweet

# OUR MODELS

---

To effectively classify disaster-related tweets, we leveraged three distinct model architectures



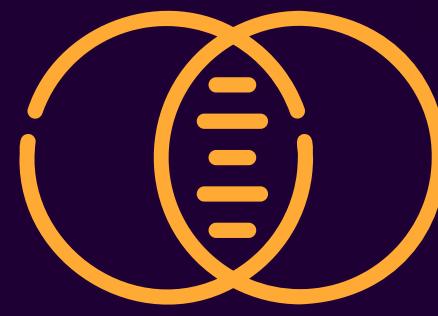
CNN

- Efficient at capturing local text patterns
- Simpler and faster to train.



BERT

- Excels at contextual understanding
- Powerful for nuanced language.



ENSEMBLE

- Blends strengths of multiple architectures for robust
- Higher accuracy.

# CNN

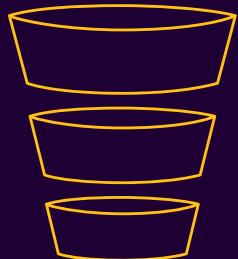
## Fast & Effective Pattern Extraction

A lightweight, convolutional approach focusing on local text patterns for fast, effective classification.



### Convolutional Filters

Learns local text patterns (e.g., phrases or keywords) by sliding filters over embeddings of words.



### Pooling Layers

Reduces dimensionality while retaining the most critical features, speeding up computation.



### Straightforward Architecture

A simpler, layered design with fewer parameters enables quick and efficient training.



### Parallelizable Computations

Convolution operations can run simultaneously, speeding up both training and inference on modern hardware.

# BERT

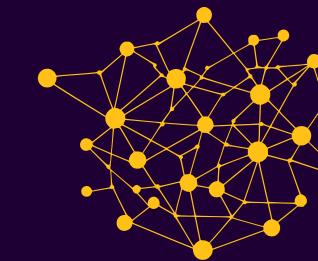
## Deeper Contextual Understanding

A transformer-based language model that excels at understanding contextual nuances in text.



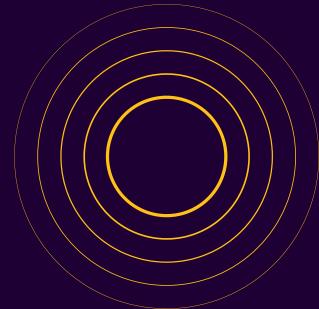
### Large-Scale Pretraining

BERT is pretrained on massive corpora, learning language patterns before fine-tuning on your specific disaster tweet dataset.



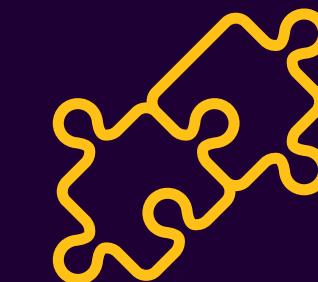
### Transformer Architecture

Leverages multi-head attention to contextualize each word with surrounding words, capturing nuanced meanings and relationships.



### Contextual Embeddings

Generates rich, context-aware embeddings that excel at handling ambiguities like sarcasm or colloquial expressions.



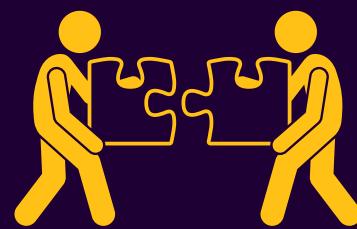
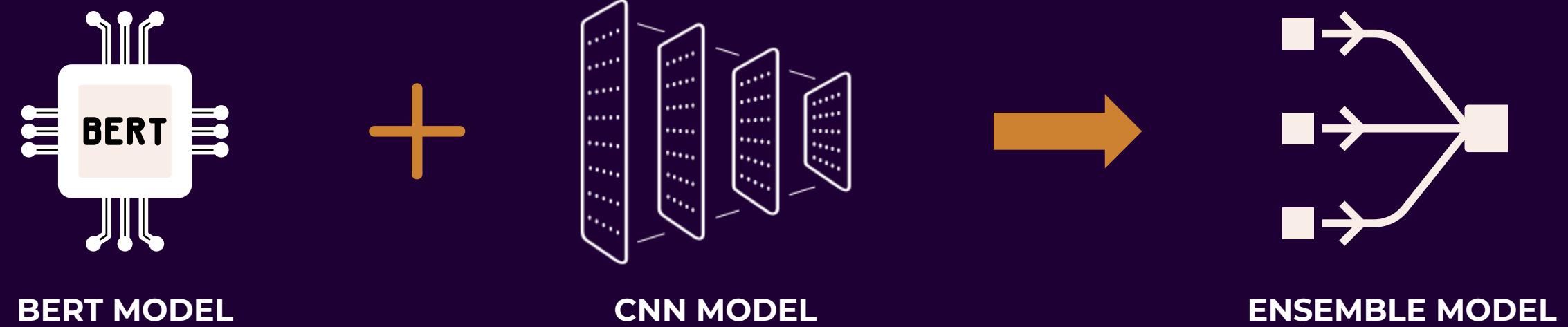
### Fine-Tuned for Disaster Domain

Only minor, task-specific training needed to adapt BERT's deep language understanding for disaster tweet classification.

# ENSEMBLE

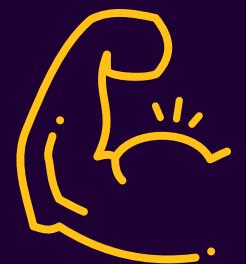
## Strength in Unity

Merges outputs from different models, leveraging their strengths to boost overall accuracy.



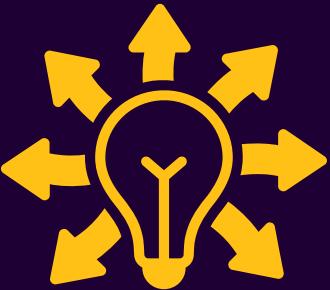
### Combined Predictions

Merges the outputs of BERT and CNN, balancing deep context and quick pattern recognition.



### Robust Performance

Reduces overfitting risk and leverages complementary strengths, typically achieving higher F1.



### Versatility

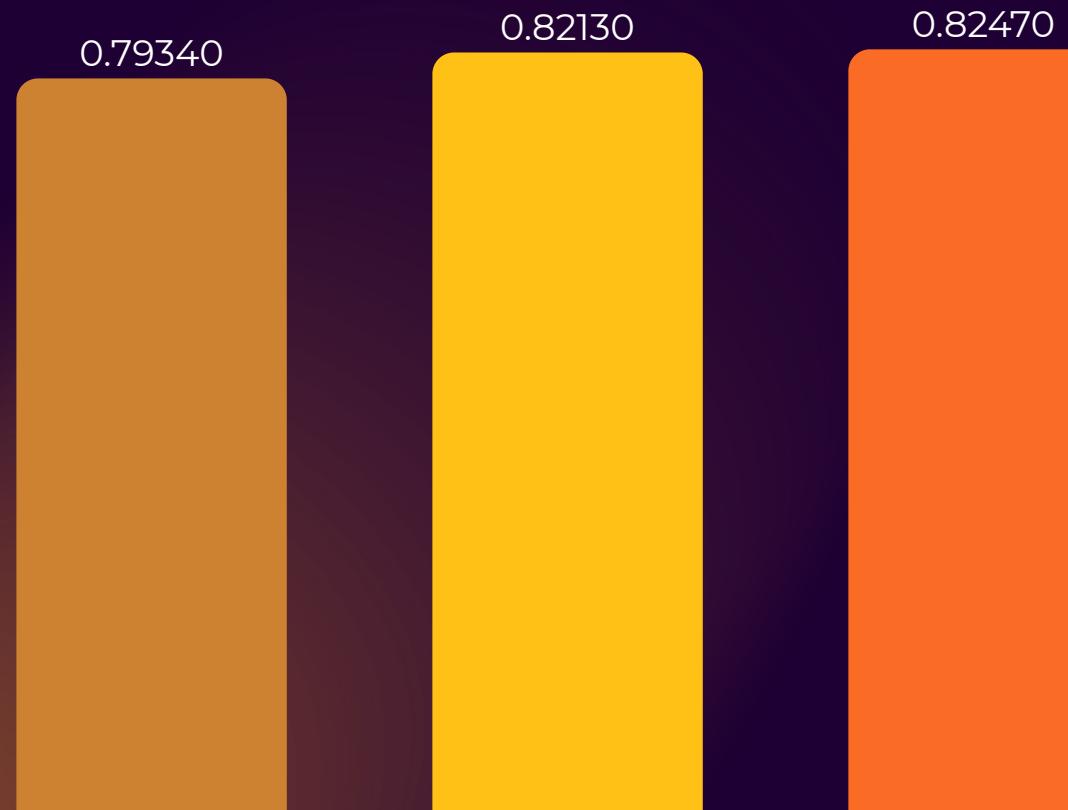
Can integrate additional models or future improvements without major rework to the existing pipeline.

# MODEL SCORES

LEADERBOARD RANK

# 139

CNN BERT ENSEMBLE



CNN



submission (4).csv

0.79344 · Uttish Narayan K · 5d ago

BERT



submission (7).csv

0.82133 · Uttish Narayan K · 4d ago

ENSEMBLE



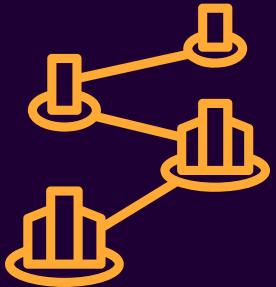
submission.csv

0.82470 · Sanjith Krishna Venkateshkumar · ...

# SCOPE FOR FUTURE WORK



## Enhancing Our Model: Future Steps Toward a More Robust Disaster Classification System



### Scalability

AI systems may collect and analyze large amounts of data, raising concerns about privacy and data protection.



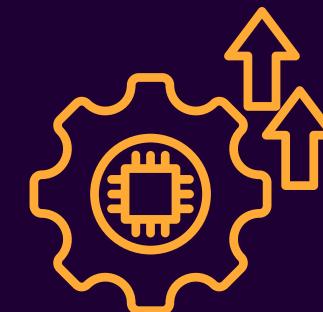
### Real-Time Application

Deploy the ensemble model to live Twitter feeds for early disaster warnings



### Multilingual Support

Fine-tune models for tweets in multiple languages to broaden coverage.



### Advanced Ensemble Methods

Explore stacking or gradient boosting with BERT/CNN outputs.



# THANK YOU !

---