



BT4222 Mining Web Data for Business Insights

# Suicidal Text Detection

in Social Media Posts

**Prepared by: Group 8**

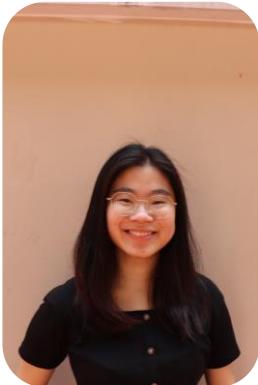
Aeron Sim Shih Win, Goh Jia Yi, Lim Zi Hui, Lin Xiao, Quek Yi Zhen





# Meet the Team

Group 8



Quek Yi Zhen



Lim Zi Hui



Aeron Sim



Goh Jia Yi



Lin Xiao



# Agenda

1. Introduction

2. Data Collection & Wrangling

3. Data Exploration

4. Representation Learning

5. Model Building & Evaluation

6. Model Selection

7. Chatbot Integration

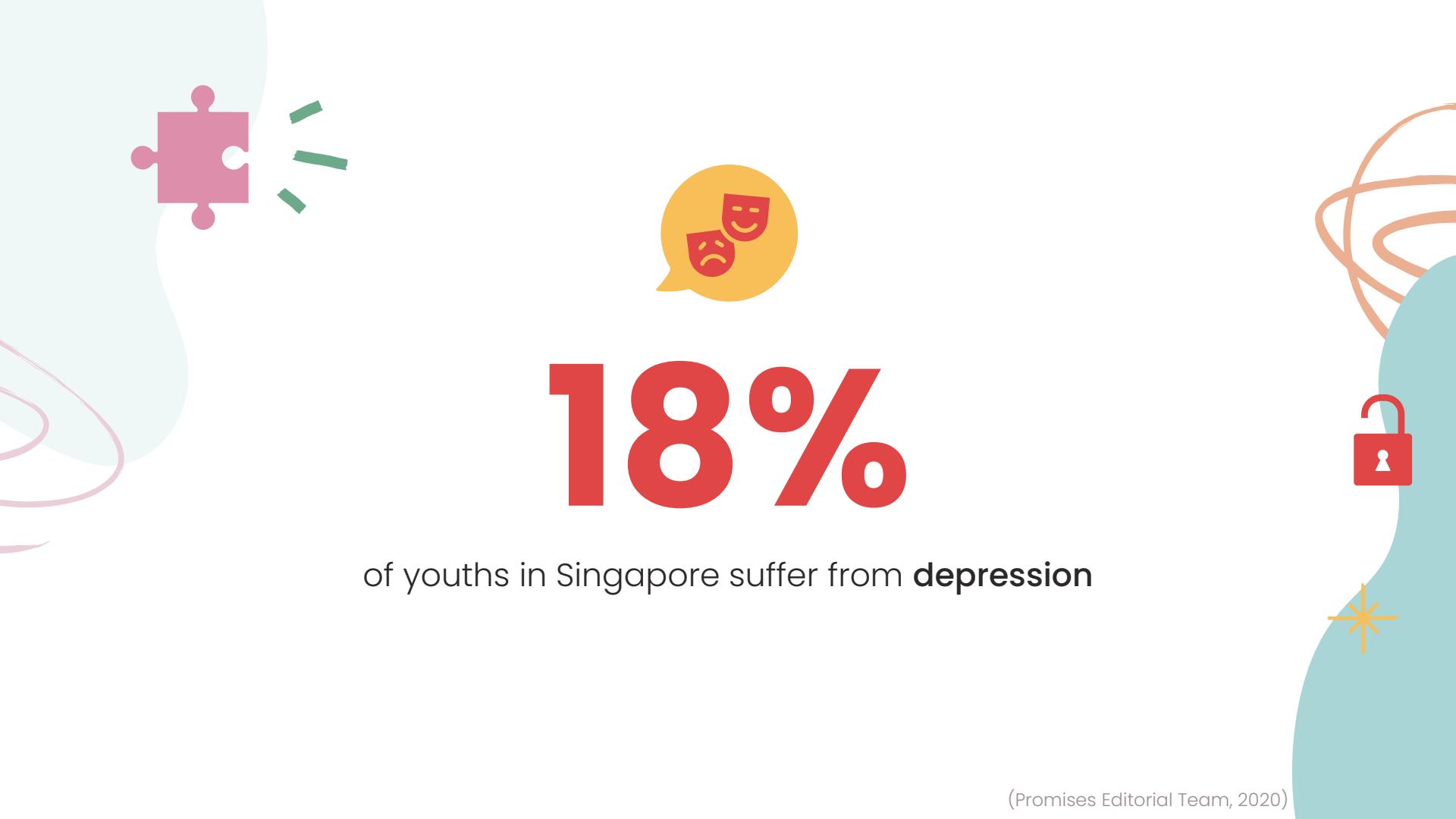
8. Future Improvements

9. Conclusion

1.

# Introduction





**18%**

of youths in Singapore suffer from **depression**

# Problem

Youths are more likely to suffer from **depression** due to



**Chemical changes** in the  
brain during puberty



Stress from **societal and  
academic pressure**

# Problem

 [Sign In](#) [My Feed](#) [Search](#)

[Top Stories](#) [Latest News](#) [Discover](#) [Singapore](#) [Asia](#) [Commentary](#) [Sustainability](#) [CNA Insider](#) [Lifestyle](#) [Watch](#) [Listen](#) [+ All Sections](#)

Singapore

## Number of suicides among those in their 20s highest in Singapore



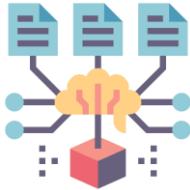
03 Aug 2020 11:20AM  
(Updated: 09 Aug 2020 05:05PM)



# Background



# Objective



## Suicidal Detection Model

- Detect suicide intent in social media posts



## Mental Health Chatbot

- Build intimate conversations with youths
- Direct them to professional resources when required

2.

# Data Collection & Wrangling



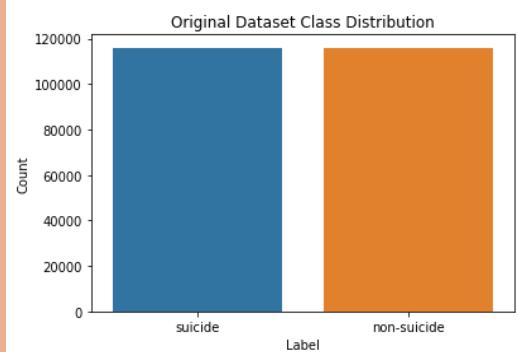
# Data Collection



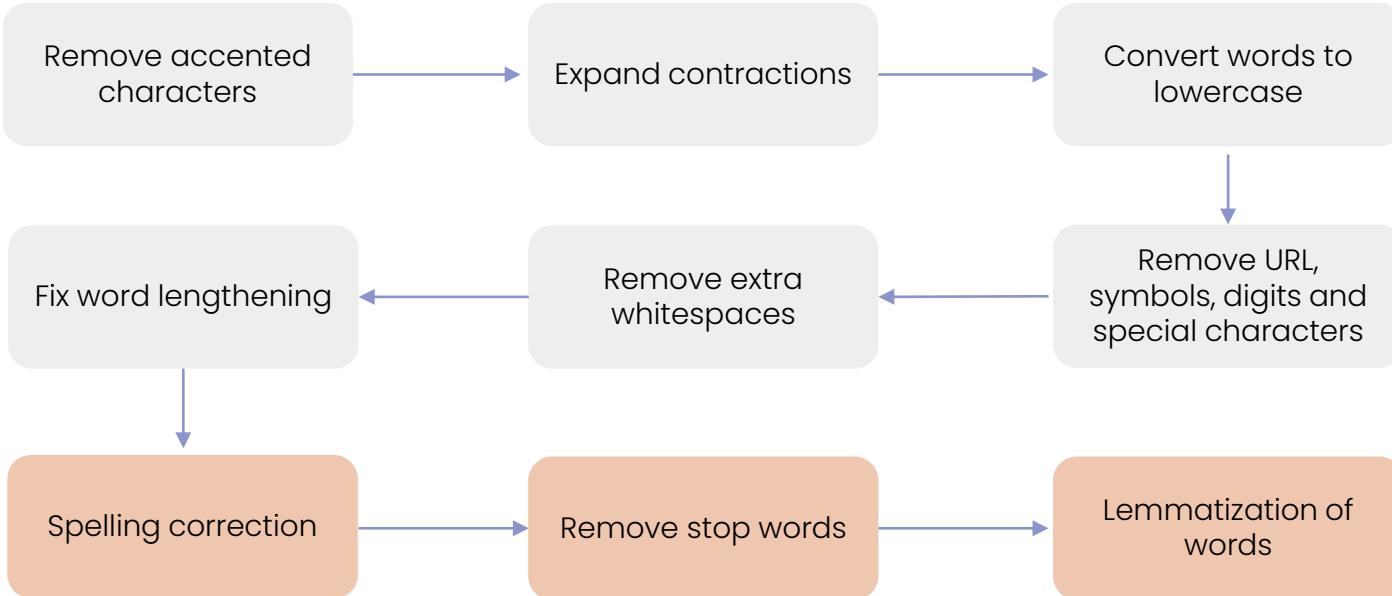
## Suicide and Depression Detection dataset

232,074 rows from 2 classes

	text	class
0	Ex Wife Threatening SuicideRecently I left my ...	suicide
1	Am I weird I don't get affected by compliments... ...	non-suicide
2	Finally 2020 is almost over... So I can never ...	non-suicide
3	i need helpjust help me im crying so hard	suicide
4	I'm so lostHello, my name is Adam (16) and I've...	suicide



# Text Preprocessing



# Data Cleaning

## 1 Remove irrelevant words

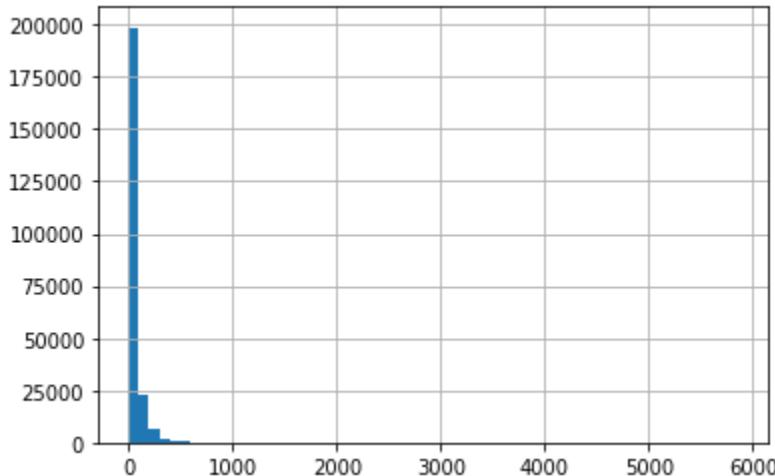
- Anomalous word “*filler*” was found in the top 30 frequently occurring words
- “*filler*” word was removed as it does not contain any meaning

## 2 Remove empty rows

- Rows that do not contain any words were found after preprocessing
- These rows were dropped

# Data Cleaning

## ③ Remove outliers in word count



Metric	Word Count
Min	1
25%	12
50%	25
75%	62
Max	5,850

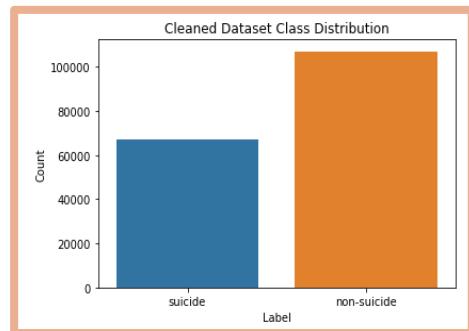
# Final Dataset



## Suicide and Depression Detection dataset

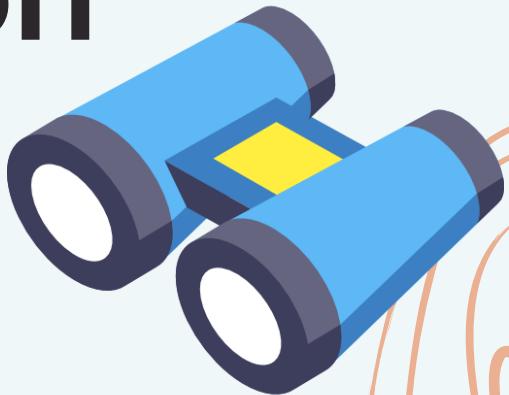
174,436 rows from 2 classes

	class	text	cleaned_text
0	suicide	Ex Wife Threatening SuicideRecently I left my ...	sex wife threaten suicide recently leave wife ...
1	non-suicide	Am I weird I don't get affected by compliments...	weird not affect compliment come know real lif...
2	non-suicide	Finally 2020 is almost over... So I can never ...	finally hear bad year swear fucking god annoying
3	suicide	i need helpjust help me im crying so hard	need help just help cry hard
4	suicide	It ends tonight.I can't do it anymore. \nI quit.	end tonight not anymore quit



3.

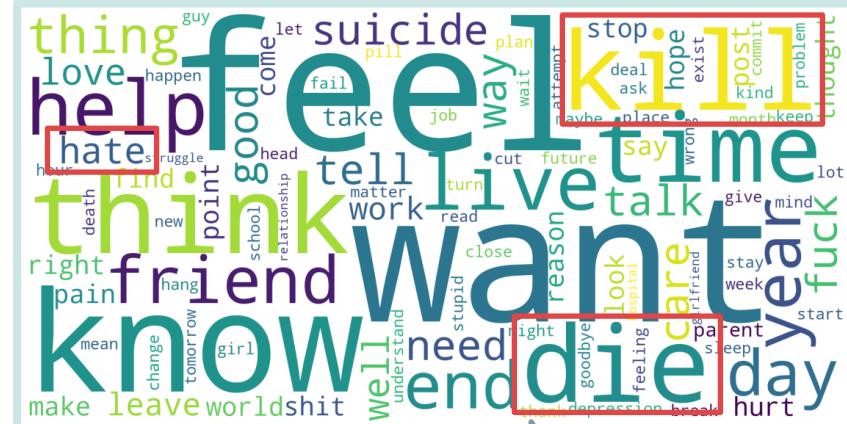
# Data Exploration



# Suicidal

Top Bigrams	Count
('feel', 'like')	10,405
('want', 'die')	5,398
('want', 'kill')	2,365
('want', 'end')	1,872
('commit', 'suicide')	1,653

More negative words

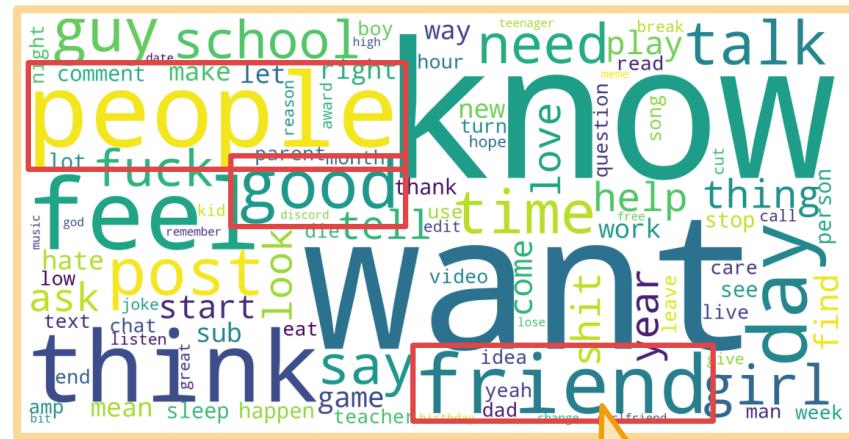


Negative words

# Non-Suicidal

Top Bigrams	Count
('feel', 'like')	3,291
('year', 'old')	1,299
('want', 'talk')	1,248
('sub', 'edit')	1,198
('min', 'craft')	1,144

## More neutral words

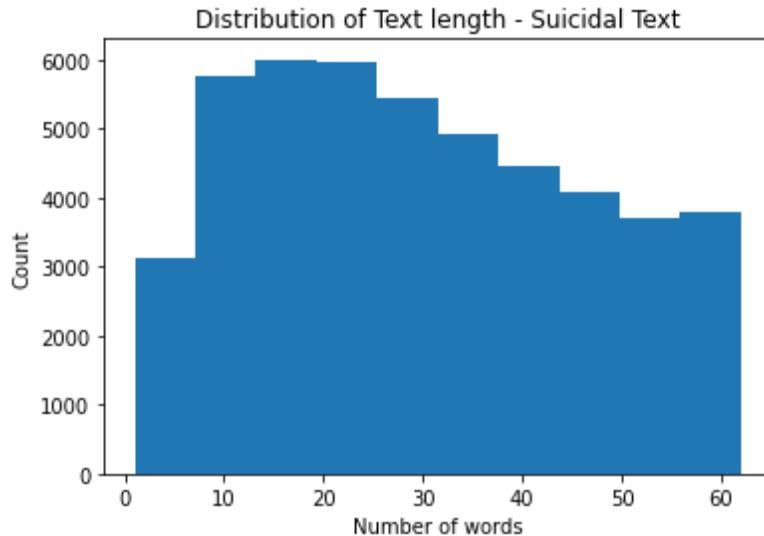


## Neutral and Positive words

# Average Text Length

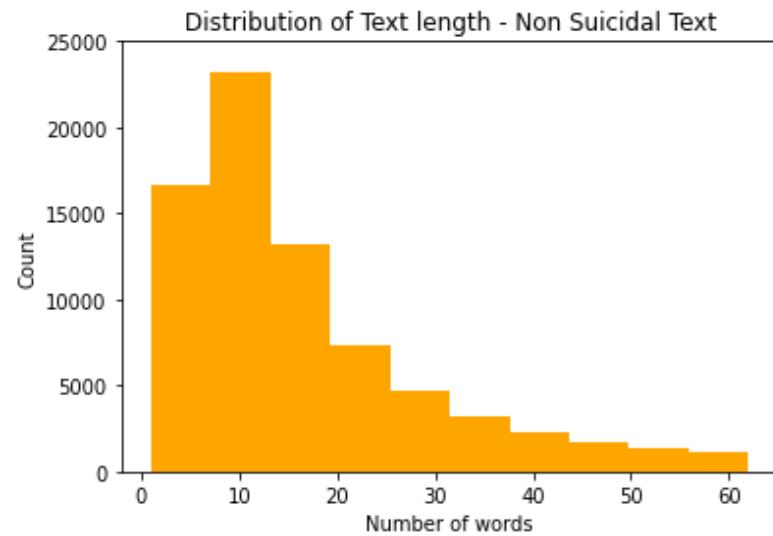
Suicidal

More uniform distribution



Non-suicidal

Right-skewed distribution



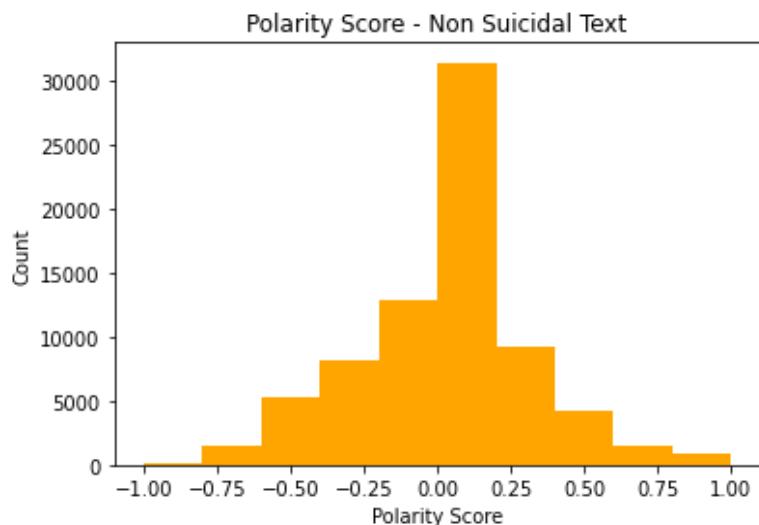
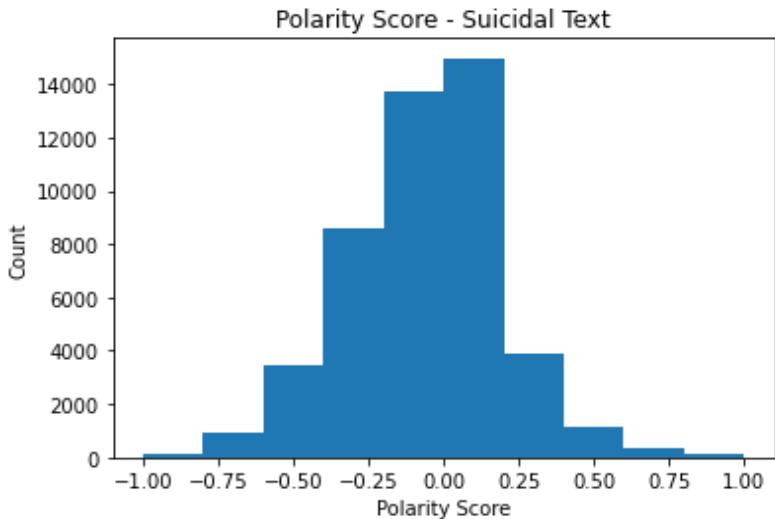
# Polarity Score

Suicidal

More negative datapoints

Non-suicidal

More neutral datapoints



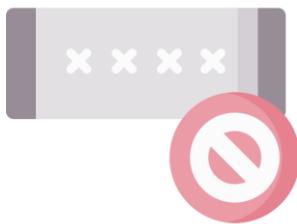
4.

# Representation Learning



# Importance of Word Embeddings

Without pre-trained word embeddings,



Inaccurate representation  
of rare words



Longer training process

# Word Embeddings Used



# Word Embeddings

## Word2Vec

### What is Word2Vec?

- Estimates the meaning of a word through grouping of similar words
- Prediction model built on neural networks

### Implementation

- Vocabulary size: ~20,400
- Training data points: ~139,500
- Dimensions: 300

## GloVe

### What is GloVe?

- Count-based log bilinear model incorporating both local and global context

### Implementation

- Vocabulary size: ~1.2 million
- Training data points: ~2.7 billion
- Dimensions: 200

5.

# Model Building & Evaluation



# Models Built

**Logit**

Logistic  
Regression

**CNN**

Convolutional  
Neural Network

**LSTM**

Long Short-term  
Neural Network

**BERT**

Bidirectional Encoder  
Representations from  
Transformers

**ELECTRA**

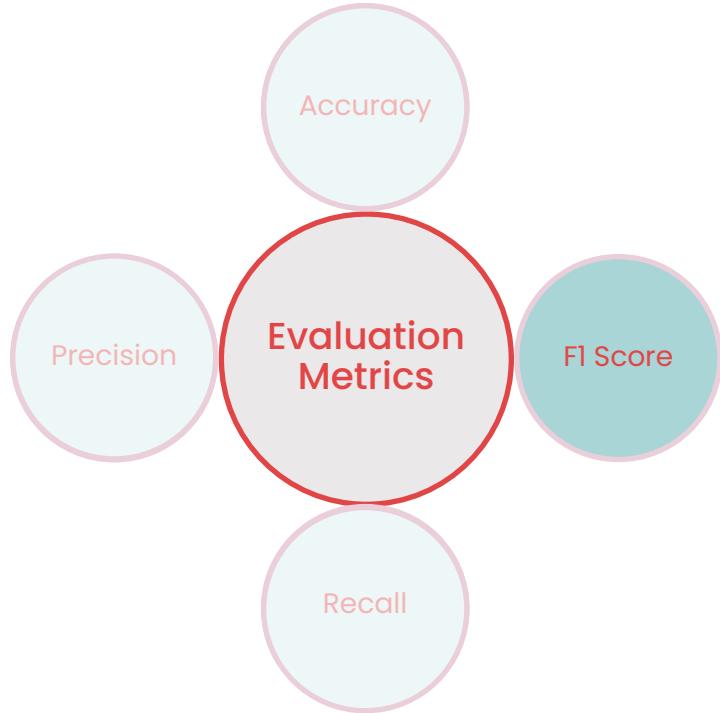
Efficiently Learning an  
Encoder that Classifies  
Token Replacements  
Accurately

Machine Learning

Deep Learning

Transformers

# Evaluation Metrics



# Models Built



## Logit

Logistic  
Regression

## CNN

Convolutional  
Neural Network

## LSTM

Long Short-term  
Neural Network

## BERT

Bidirectional Encoder  
Representations from  
Transformers

## ELECTRA

Efficiently Learning an  
Encoder that Classifies  
Token Replacements  
Accurately



Machine Learning

# Why Logit?

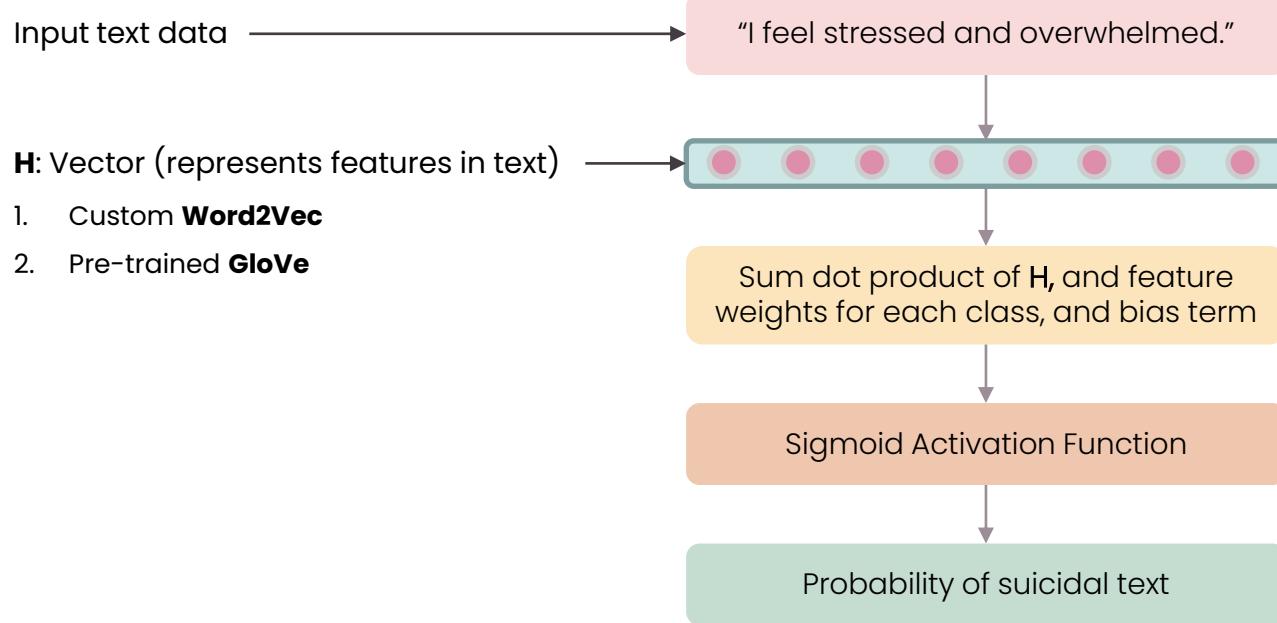


Suitable to classify binary variable



Baseline model

# How does Logit work?



# Logit Evaluation

## Custom Word2Vec Embeddings

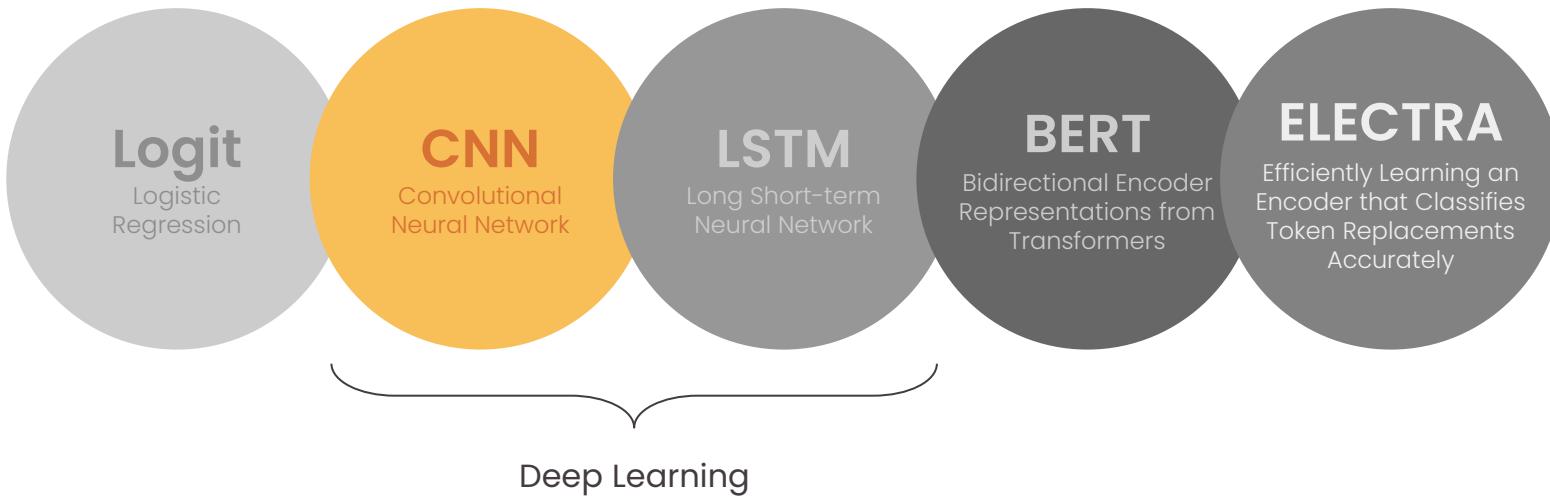
*F1 Score: 0.8851*

Pre-trained GloVe Embeddings

2



# Models Built



# Why CNN?

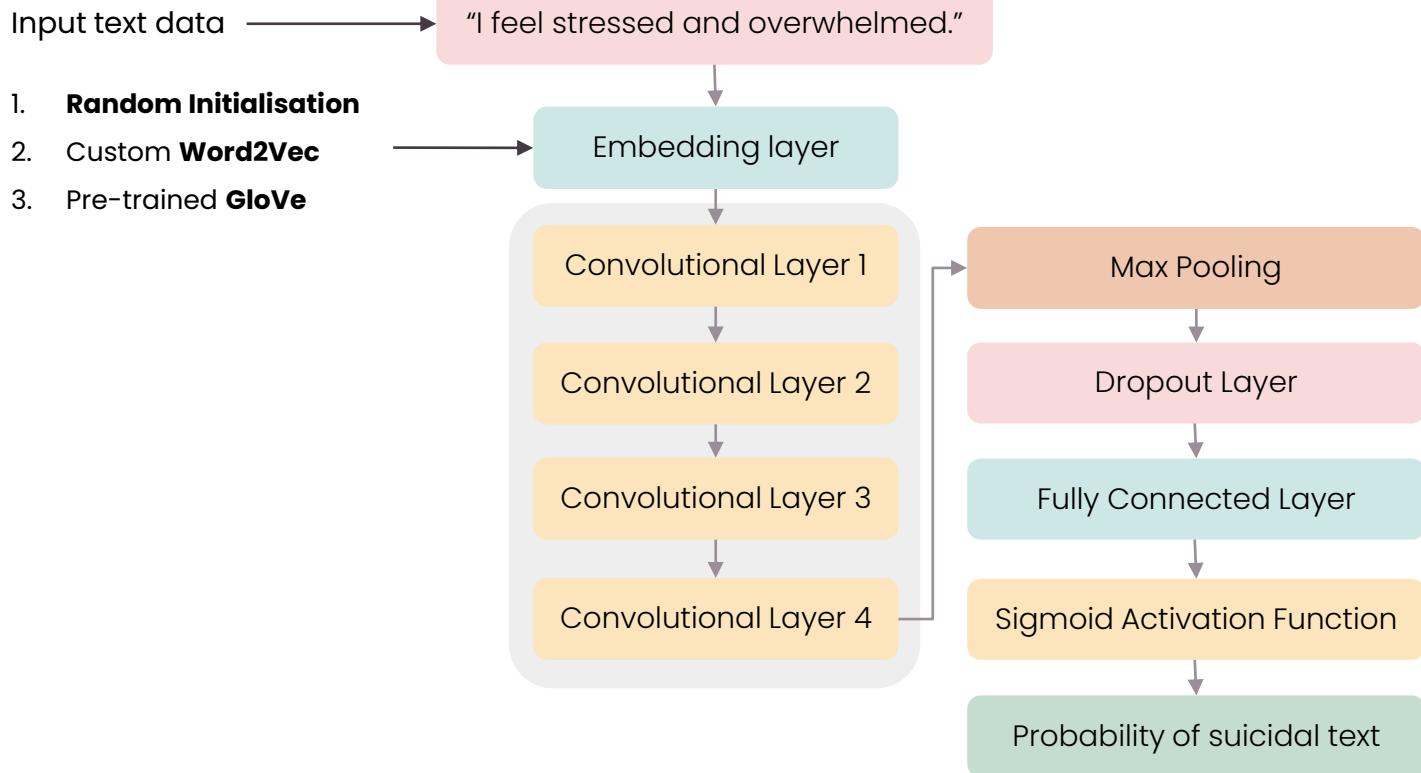


Efficient at classifying text data



Processes sequence of words

# How does CNN work?



# CNN Evaluation

## Custom Word2Vec Embeddings

*F1 Score: 0.9069*

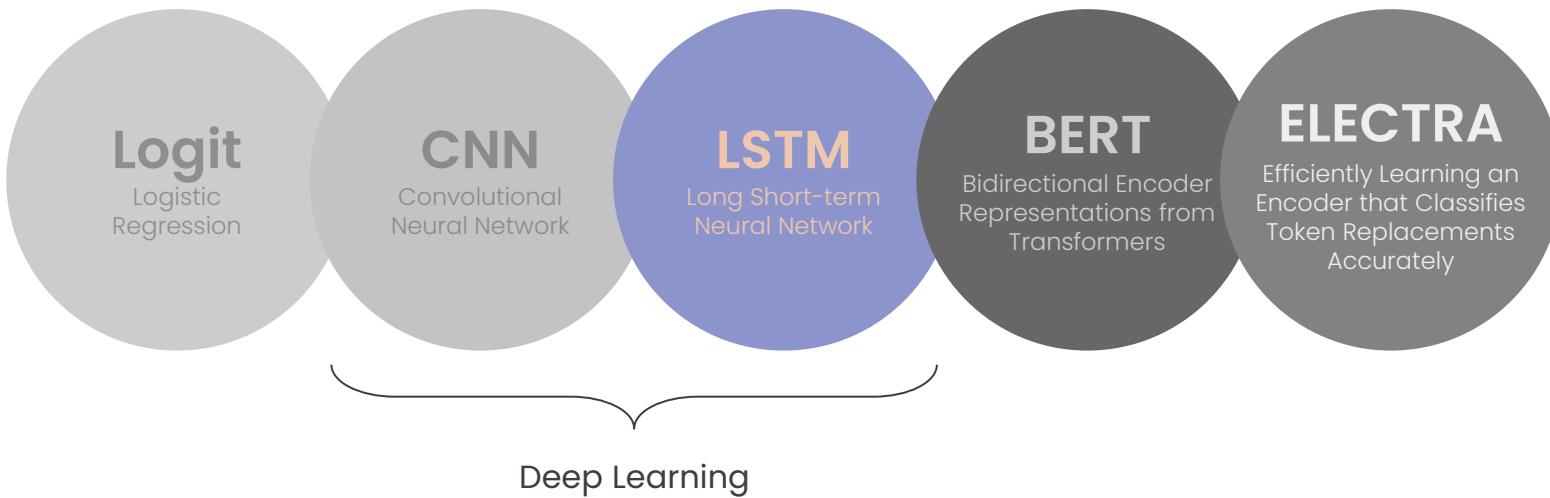
Pre-trained GloVe Embeddings



Random Initialisation



# Models Built



# Why LSTM?

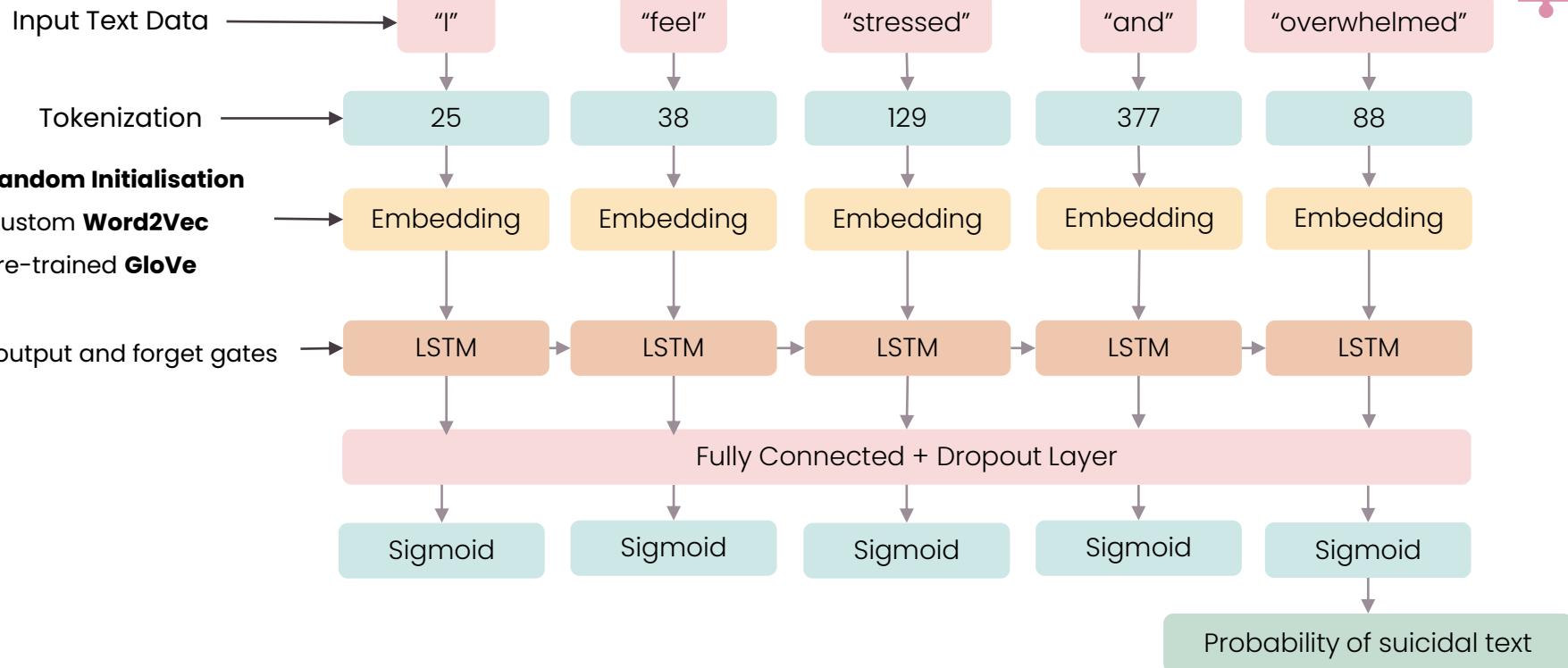


Remembers longer sequence dependencies compared to CNNs



Alleviates vanishing gradient problem of RNNs

# How does LSTM work?



# LSTM Evaluation

## Custom Word2Vec Embeddings

F1 Score: 0.9003

Pre-trained GloVe Embeddings



Random Initialisation



# Models Built

**Logit**

Logistic  
Regression

**CNN**

Convolutional  
Neural Network

**LSTM**

Long Short-term  
Neural Network

**BERT**

Bidirectional Encoder  
Representations from  
Transformers

**ELECTRA**

Efficiently Learning an  
Encoder that Classifies  
Token Replacements  
Accurately

Transformers

# Transformers

## What are Transformers?

- Attention-based Encoder Decoder
- Achieved state-of-the-art performance

## Why Transformers?



Handles long-range interactions



Parallelisation of tasks

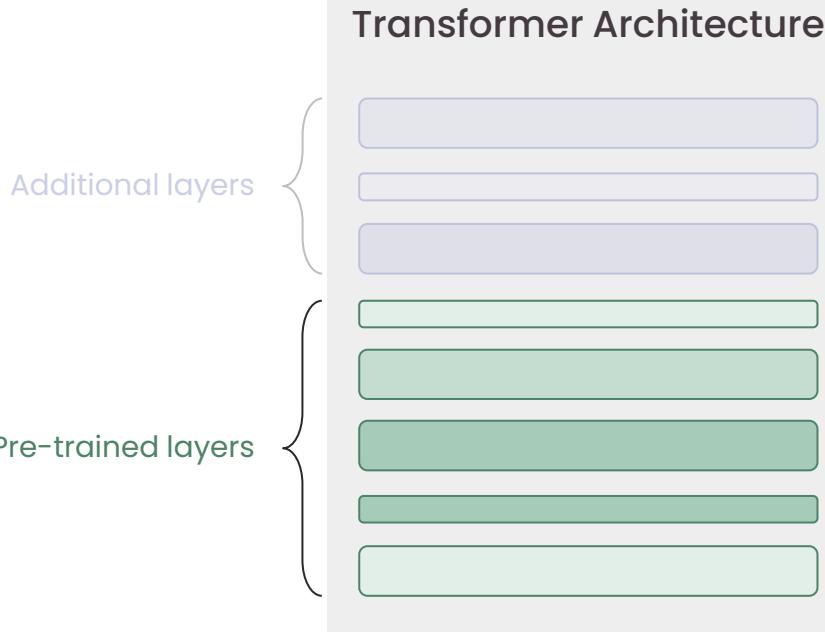


Transfer learning

BERT

ELECTRA

# Transformers Fine-tuning



- Train the entire pre-trained architecture ✓
- Freeze part of the pre-trained architecture and train remaining layers
- Freeze the entire pre-trained architecture and train additional layers



ELECTRA

# Models Built

**Logit**

Logistic  
Regression

**CNN**

Convolutional  
Neural Network

**LSTM**

Long Short-term  
Neural Network

**BERT**

Bidirectional Encoder  
Representations from  
Transformers

**ELECTRA**

Efficiently Learning an  
Encoder that Classifies  
Token Replacements  
Accurately

Transformers

# BERT

## What is BERT?

- Widely used transformer
- Encoder structure for language modelling

## BERT Implementation

- Using the base model
- Customised text preprocessing

## How does BERT work?



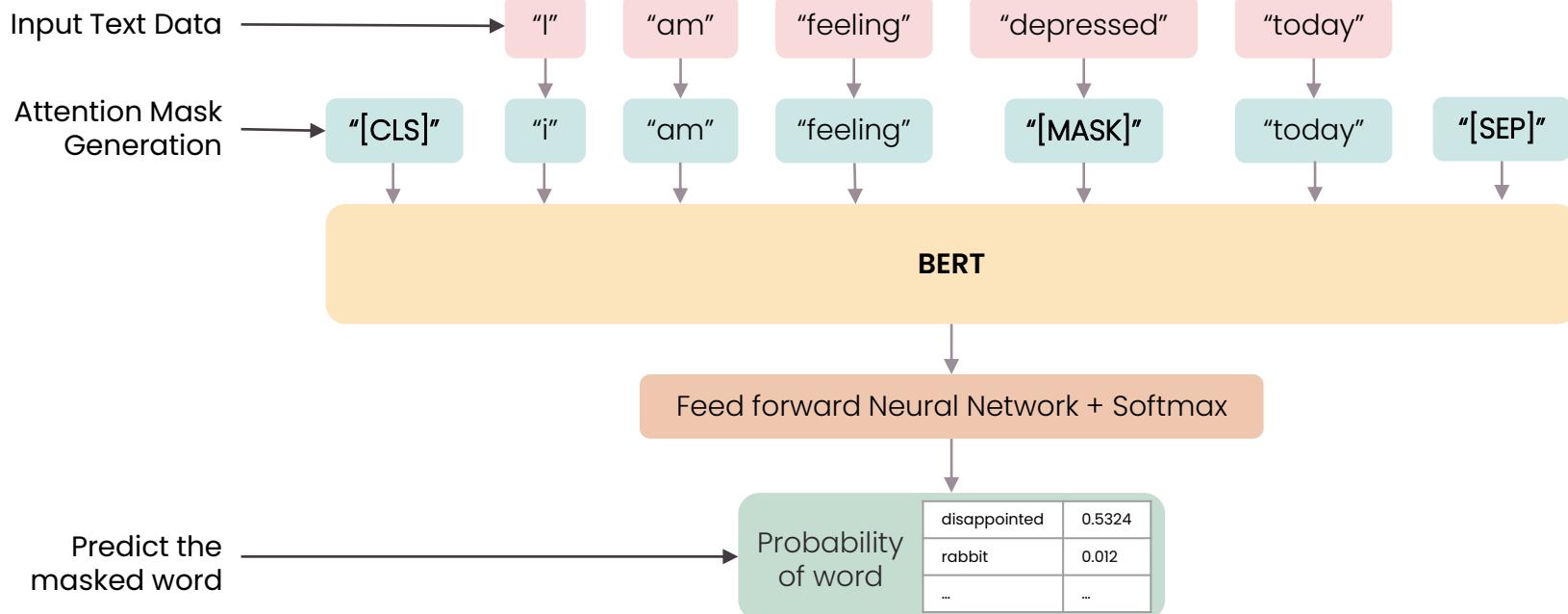
Masked Language Model (MLM)



Next Sentence Prediction (NSP)

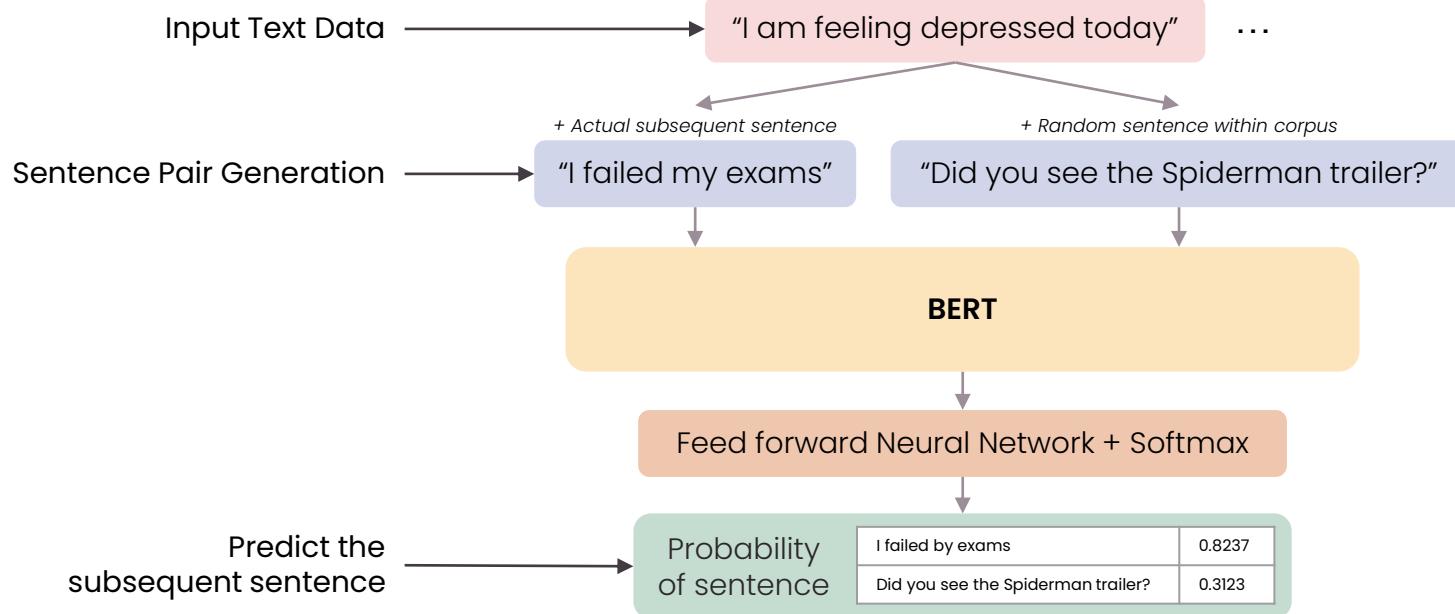
# How does BERT work?

I Masked Language Model (MLM)  
*capturing token-level information*



# How does BERT work?

 Next Sentence Prediction (NSP)  
*capturing sentence-level information*



# BERT Evaluation

## Fine-tuned BERT

*F1 Score: 0.9685*

Pre-trained BERT



# Models Built

**Logit**

Logistic  
Regression

**CNN**

Convolutional  
Neural Network

**LSTM**

Long Short-term  
Neural Network

**BERT**

Bidirectional Encoder  
Representations from  
Transformers

**ELECTRA**

Efficiently Learning an  
Encoder that Classifies  
Token Replacements  
Accurately

Transformers

# ELECTRA

## What is ELECTRA?

- One of the latest transformer model
- Achieved current state-of-the-art performance

## ELECTRA Implementation

- Using the base model
- Customised text preprocessing

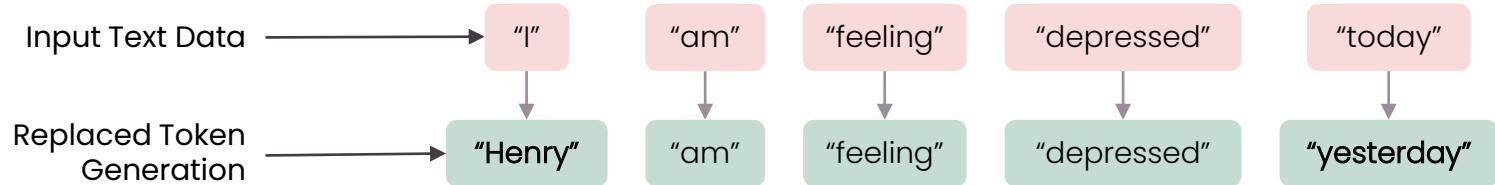
## How does ELECTRA work?



Replaced Token Detection (RTD)

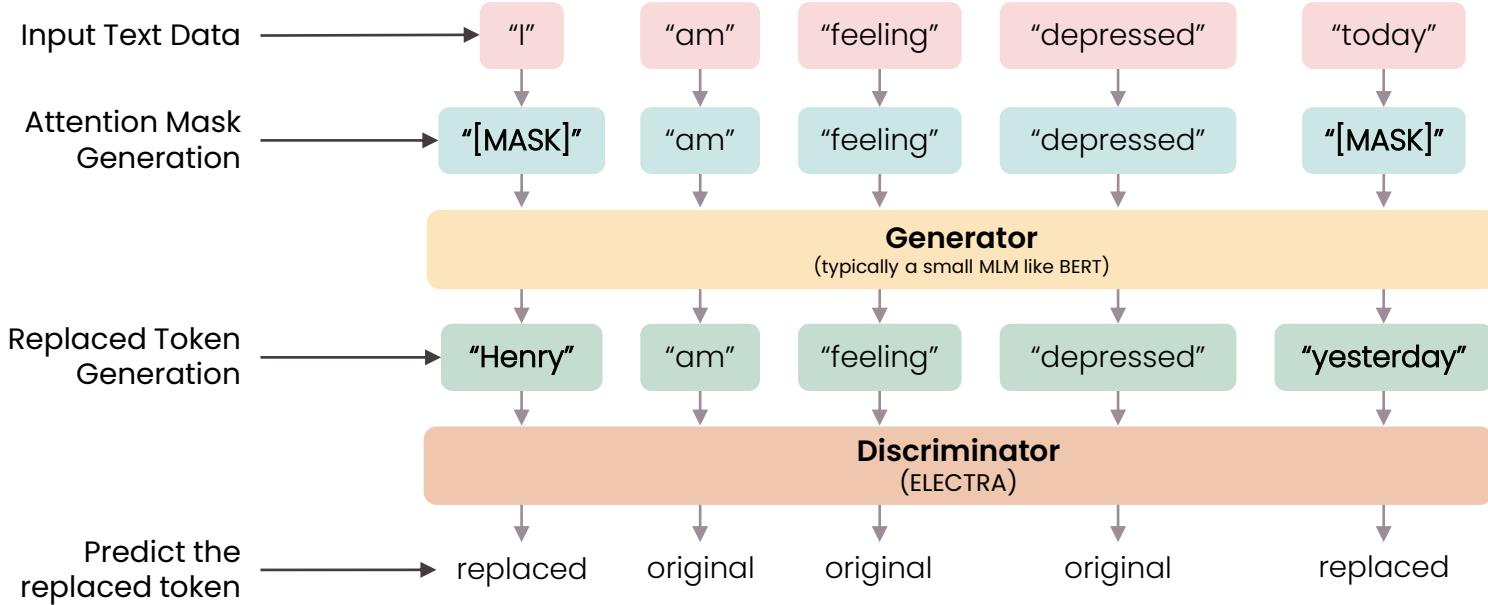
# How does ELECTRA work?

 Replaced Token Detection (RTD)  
*capturing token-level information*



# How does ELECTRA work?

Replaced Token Detection (RTD)  
capturing token-level information



# ELECTRA Evaluation

## Fine-tuned ELECTRA

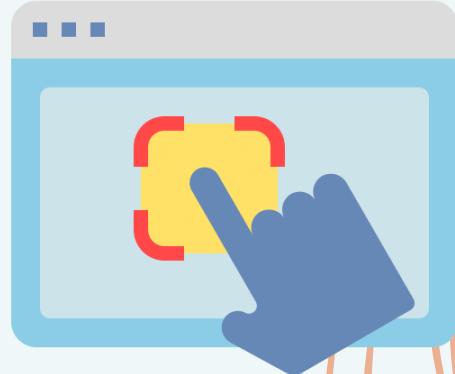
*F1 Score: 0.9732*

Pre-trained ELECTRA



6.

# Model Selection



# Evaluation Metrics



# Model Performance

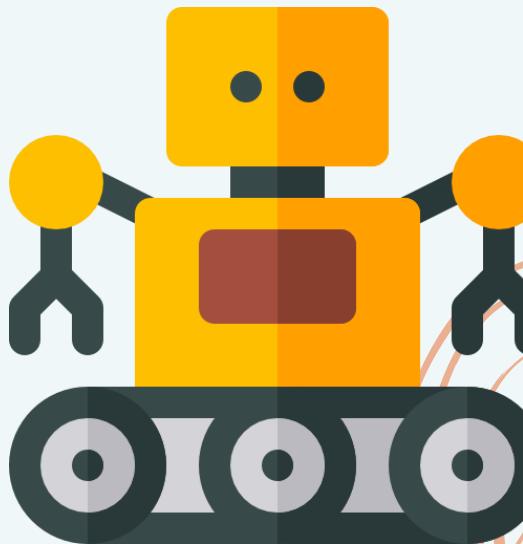
Best Model	Accuracy	Recall	Precision	F1 Score
Logit*	0.9111	0.8870	0.8832	0.8851
CNN*	0.9285	0.9013	0.9125	0.9069
LSTM*	0.9260	0.8649	0.9386	0.9003
BERT <sup>#</sup>	0.9757	0.9669	<b>0.9701</b>	0.9685
ELECTRA <sup>#</sup>	<b>0.9792</b>	<b>0.9788</b>	0.9677	<b>0.9732</b>

\*Using Custom Word2Vec Embeddings

<sup>#</sup>After fine-tuning

7.

# Chatbot Integration





**SCARY TO TALK  
TO COUNSELLOR**

# **SOCIAL STIGMA**

**NO TRUST**      **HOPELESS**      **DESPAIR**  
**STRESSFUL**      **SHOW NO**  
                     **WEAKNESS**  
**AFRAID OF**  
**BEING JUDGED**  
**UNCOMFORTABLE**

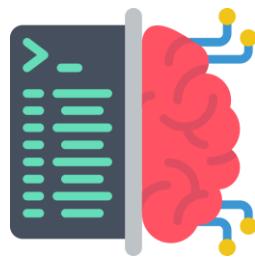
**NO ONE  
TO TALK  
TO**

# Chatbot Integration

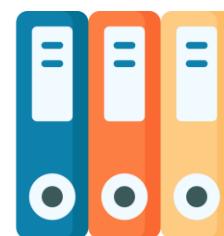


# Chatbot Architecture

Neural Response Generation (NRG) Model



Generative  
Method



Retrieval-based  
Method

# Implementation



## Generative Method

*Conversational purpose*



Implemented DialoGPT transformer



Generates dialogue responses



Responses comparable  
to human response



## Retrieval-based Method

*Suicidal detection purpose*



Customised response



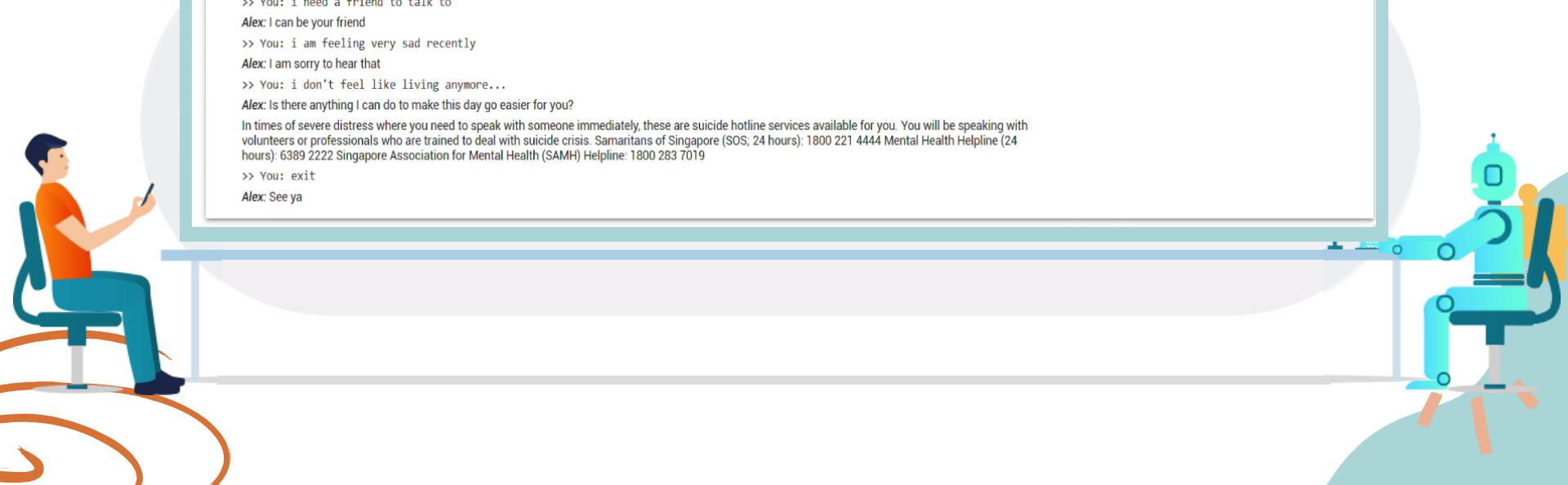
Collection of comforting messages  
and helpline information



Checks for suicidal intent and  
responds appropriately



# Chatbot Demo



```
✓ start_chatbot()
==== Hello! I am Alex and I am your virtual friend. If you need a listening ear, I'm always here. To end the chat, input 'exit' in the chatbox. ====
>> You: hello
Alex: Hiya
>> You: i need a friend to talk to
Alex: I can be your friend
>> You: i am feeling very sad recently
Alex: I am sorry to hear that
>> You: i don't feel like living anymore...
Alex: Is there anything I can do to make this day go easier for you?
In times of severe distress where you need to speak with someone immediately, these are suicide hotline services available for you. You will be speaking with volunteers or professionals who are trained to deal with suicide crisis. Samaritans of Singapore (SOS, 24 hours): 1800 221 4444 Mental Health Helpline (24 hours): 6389 2222 Singapore Association for Mental Health (SAMH) Helpline: 1800 283 7019
>> You: exit
Alex: See ya
```

8.

# Future Improvements



# Business Improvements



## Multilingual Chatbot

Integrating multiple languages  
to expand our outreach



## Social Media Integration

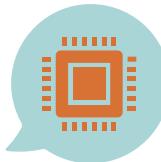
To various platforms to  
boost chatbot usage

# Technical Improvements



## Semi-Supervised Learning

Pseudo labelling of data points to improve data quality



## Larger Transformers Models

Improve model results with more computational resources



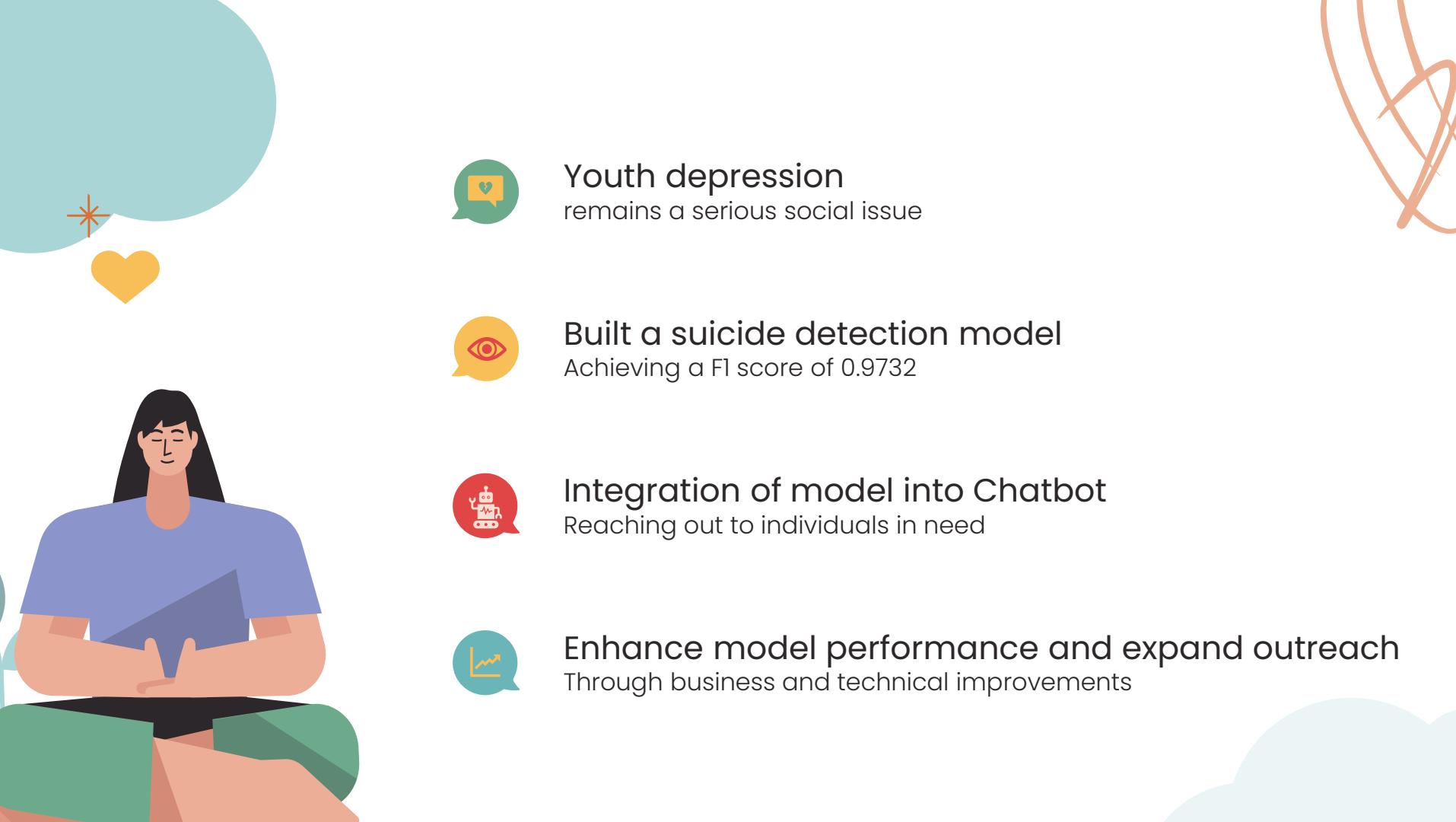
## Reinforcement Learning

Allows chatbot to learn by interacting with end users

9.

# Conclusion





**Youth depression**  
remains a serious social issue



**Built a suicide detection model**  
Achieving a F1 score of 0.9732



**Integration of model into Chatbot**  
Reaching out to individuals in need



**Enhance model performance and expand outreach**  
Through business and technical improvements

# Thank you!

## Get in touch

Aeron Sim	<a href="mailto:e0273924@u.nus.edu">e0273924@u.nus.edu</a>
Goh Jia Yi	<a href="mailto:gohjiayi@u.nus.edu">gohjiayi@u.nus.edu</a>
Lim Zi Hui	<a href="mailto:jlimzihui@u.nus.edu">jlimzihui@u.nus.edu</a>
Lin Xiao	<a href="mailto:e0418026@u.nus.edu">e0418026@u.nus.edu</a>
Quek Yi Zhen	<a href="mailto:e0325941@u.nus.edu">e0325941@u.nus.edu</a>



# Appendix

# Model Variants Performance



Logit Model	Accuracy	Recall	Precision	F1 Score
<b>Custom Word2Vec Embeddings</b>	<b>0.9111</b>	<b>0.8870</b>	<b>0.8832</b>	<b>0.8851</b>
Pre-trained GloVe Embeddings	0.8774	0.8440	0.8394	0.8417



CNN Model	Accuracy	Recall	Precision	F1 Score
Random Initialisation	0.8985	0.8281	0.9010	0.8630
<b>Custom Word2Vec Embeddings</b>	<b>0.9285</b>	<b>0.9013</b>	<b>0.9125</b>	<b>0.9069</b>
Pre-trained GloVe Embeddings	0.9001	0.8511	0.8858	0.8681



LSTM Model	Accuracy	Recall	Precision	F1 Score
Random Initialisation	0.8724	0.7982	0.8611	0.8285
<b>Custom Word2Vec Embeddings</b>	<b>0.9260</b>	<b>0.8649</b>	<b>0.9386</b>	<b>0.9003</b>
Pre-trained GloVe Embeddings	0.8825	0.7613	0.9206	0.8334

# Model Variants Performance

BERT Model	Accuracy	Recall	Precision	F1 Score
Pre-trained BERT	0.4681	0.9295	0.4156	0.5744
<b>Fine-tuned BERT</b>	<b>0.9757</b>	<b>0.9669</b>	<b>0.9701</b>	<b>0.9685</b>

ELECTRA Model	Accuracy	Recall	Precision	F1 Score
Pre-trained ELECTRA	0.4025	<b>0.9908</b>	0.3918	0.5615
<b>Fine-tuned ELECTRA</b>	<b>0.9792</b>	0.9788	<b>0.9677</b>	<b>0.9732</b>