TEXT-BASED EMOTION DETECTION

INTRODUCTION

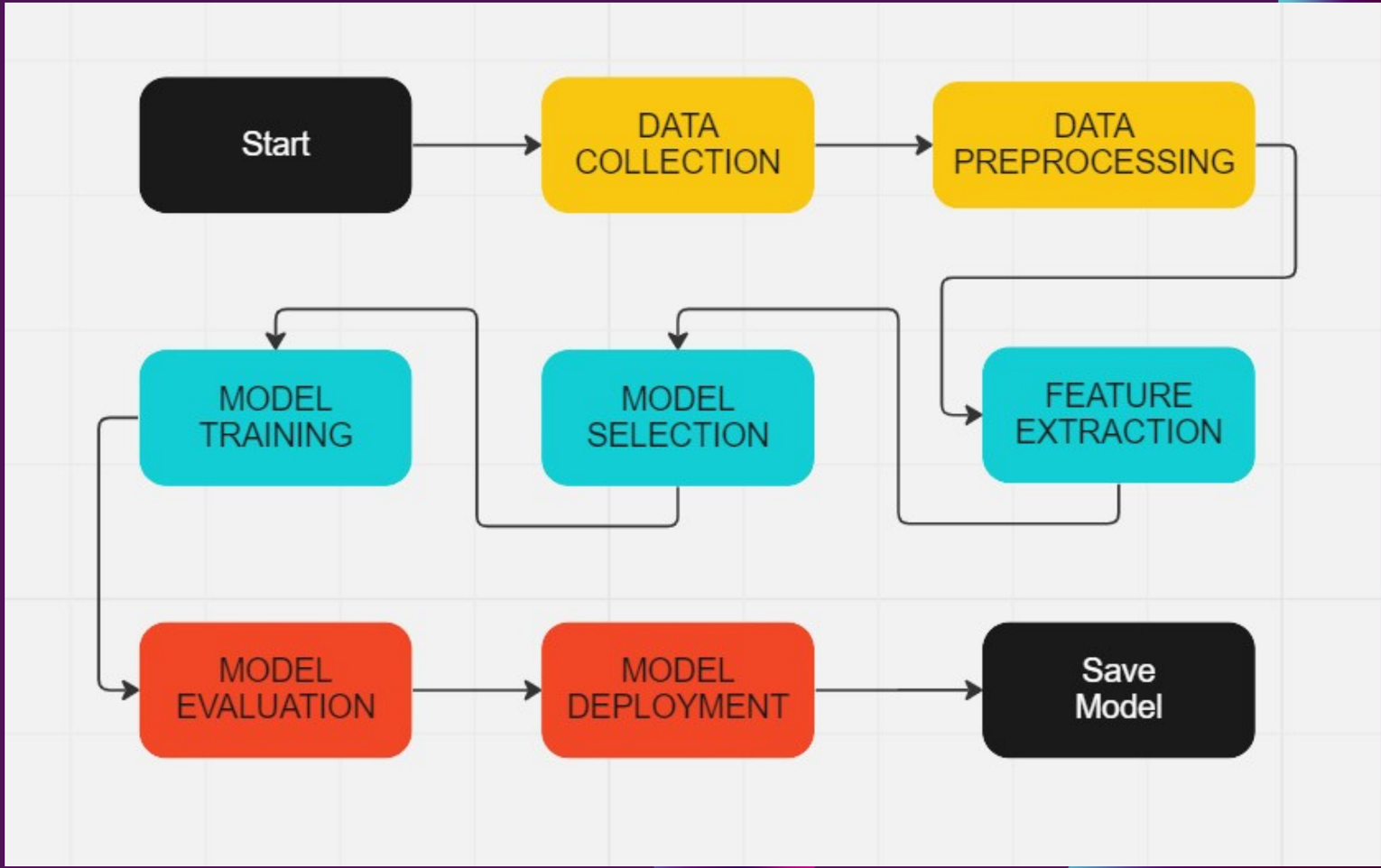
Text-based emotion detection analyzes language patterns to identify the emotional state of a writer. It has many applications and the potential to revolutionize human-computer interaction.

This project aims to develop an emotion detection system using NLP techniques to classify text into different emotions such as happiness, sadness, anger, fear, surprise, neutral, etc. The system will be trained on a large dataset of labeled text documents and evaluated using standard metrics to determine its accuracy and usefulness in real-world applications.

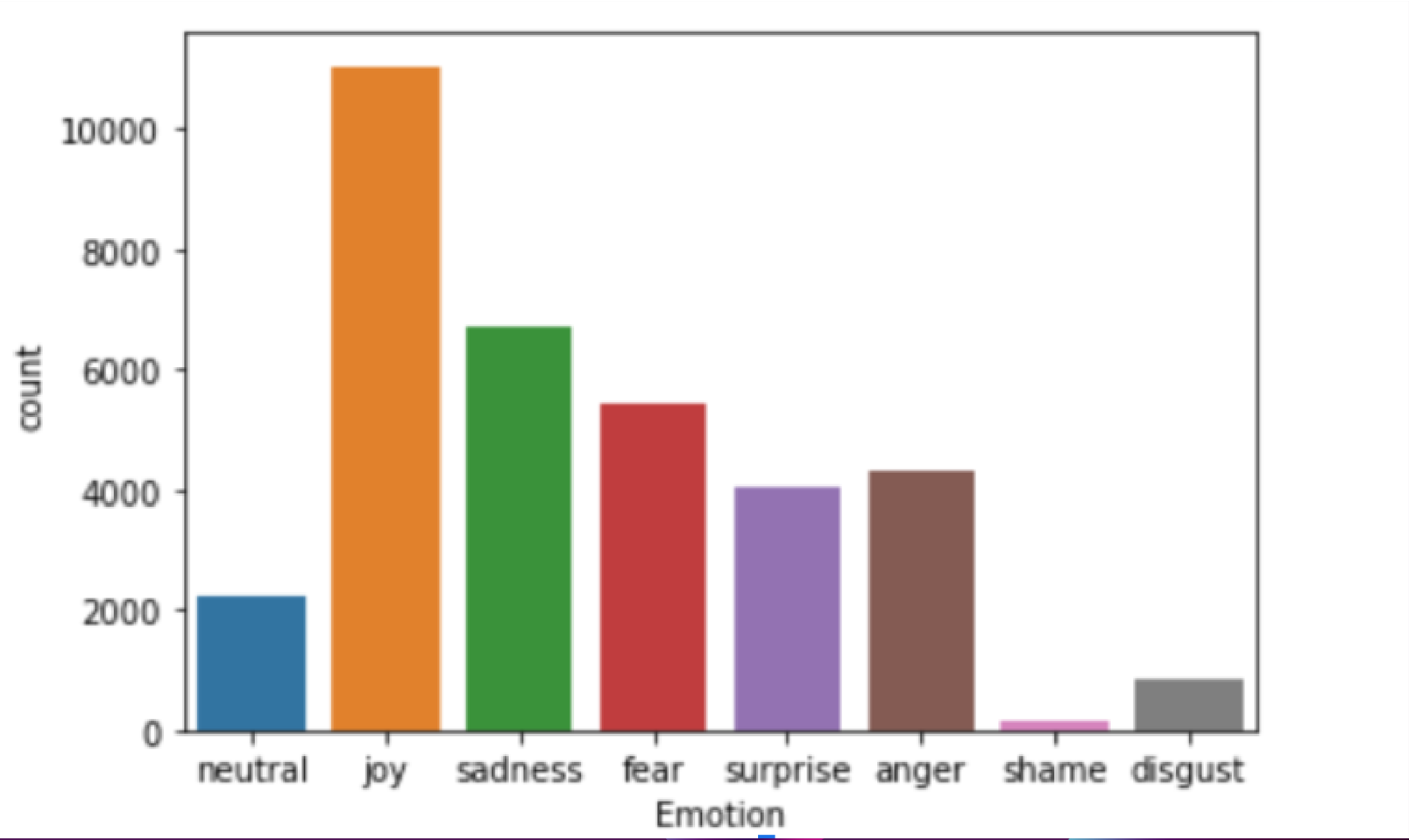
ROLE OF NLP

- NLP is crucial in emotion detection systems that rely on text analysis
- NLP enables computers to understand and process human language
- NLP techniques preprocess text data by removing stop words, stemming or lemmatizing words, and converting text to numerical representation
- These techniques are critical for the model to identify relevant features and patterns associated with different emotions.

FLOW DIAGRAM



Dataset


10000
8000
6000
4000
2000
0
count
neutral
joy
sadness
fear
surprise
anger
shame
disgust
Emotion

DATA PREPROCESSING

REMOVING USER HANDLES, STOP WORDS

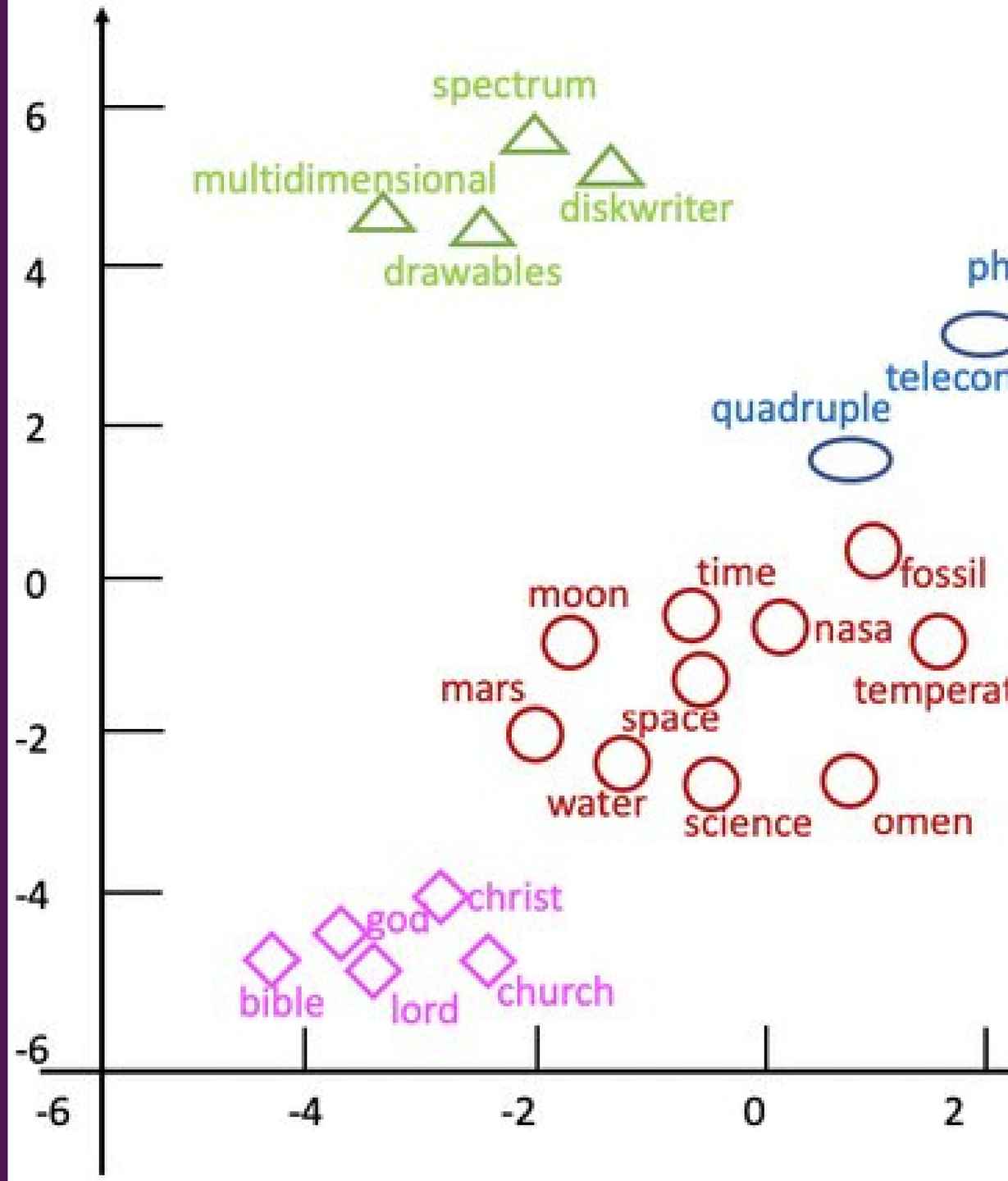
- DATA PRIVACY
- DATA NOISE REDUCTION
- GENERALIZATION

	Emotion	Text	Clean_Text
0	neutral	Why ?	
1	joy	Sage Act upgrade on my to do list for tommorow.	sage act upgrade list tommorow
2	sadness	ON THE WAY TO MY HOMEGIRL BABY FUNERAL!!! MAN ...	way homegirl baby funeral man hate funerals re...
3	joy	Such an eye ! The true hazel eye-and so brill...	eye true hazel eyeand brilliant regular featur...
4	joy	@lluvmiasantos ugh babe.. hugggzzz for u .! b...	ugh babe hugggzzz u babe naamazed nga ako e ba...
...	...	...	...
34787	surprise	@MichelGW have you gift! Hope you like it! It'...	gift hope like hand made wear itll keep warm lol
34788	joy	The world didnt give it to me..so the world MO...	world didnt give meso world definitely cnt tak...
34789	anger	A man robbed me today .	man robbed today
34790	fear	Youu call it JEALOUSY, I call it of #Losing YO...	youu call jealousy call
34791	sadness	I think about you baby, and I dream about you ...	think baby dream time

34792 rows × 3 columns

Feature extraction

- CountVectorizer
- Term frequency-inverse document frequency (TF-IDF)



CountVectorizer

document = ["One Geek helps Two Geeks", "Two Geeks help Four Geeks", "Each Geek helps many other Geeks at GeeksforGeeks."]

	at	each	four	geek	geeks	geeksforgeeks	help	helps	many	one	other	two
document[0]	0	0	0	1	1	0	0	1	0	1	0	1
document[1]	0	0	1	0	2	0	1	0	0	0	0	1
document[2]	1	1	0	1	1	1	0	1	1	0	1	0

TFIDF

Term Frequency: In document d, the frequency represents the number of instances of a given word t

```
tf(t,d) = count of t in d / number of words in d
```

Inverse Document Frequency (IDF) is a term used to measure the importance or rarity of a term within a collection of documents.

```
df(t) = occurrence of t in documents
```

```
idf(t) = log(N/ df(t))
```

```
tf-idf(t, d) = tf(t, d) * idf(t)
```

ML - Based Models

1) Multinomial Naive bayes classifier

The Multinomial Naive Bayes algorithm is a Bayesian learning approach popular in Natural Language Processing (NLP). The program guesses the tag of a text, such as an email or a newspaper story, using the Bayes theorem. It calculates each tag's likelihood for a given sample and outputs the tag with the greatest chance.

2) Linear SVM

Linear Support Vector Machine (SVM) algorithm used for binary and multi-class classification tasks. SVM aims to find the best hyperplane that separates the data points of different classes with the largest margin, which is known as the support vector. In the linear SVM variant, the decision boundary is a hyperplane that is linear in the feature space.

3) Logistics Regression

Logistic Regression is a widely used linear model for binary classification tasks. It's a statistical method that models the relationship between a dependent binary variable (e.g., class label) and one or more independent variables (e.g., features extracted from text data). The goal of logistic regression is to find the optimal coefficients for the independent variables that best predict the binary outcome.

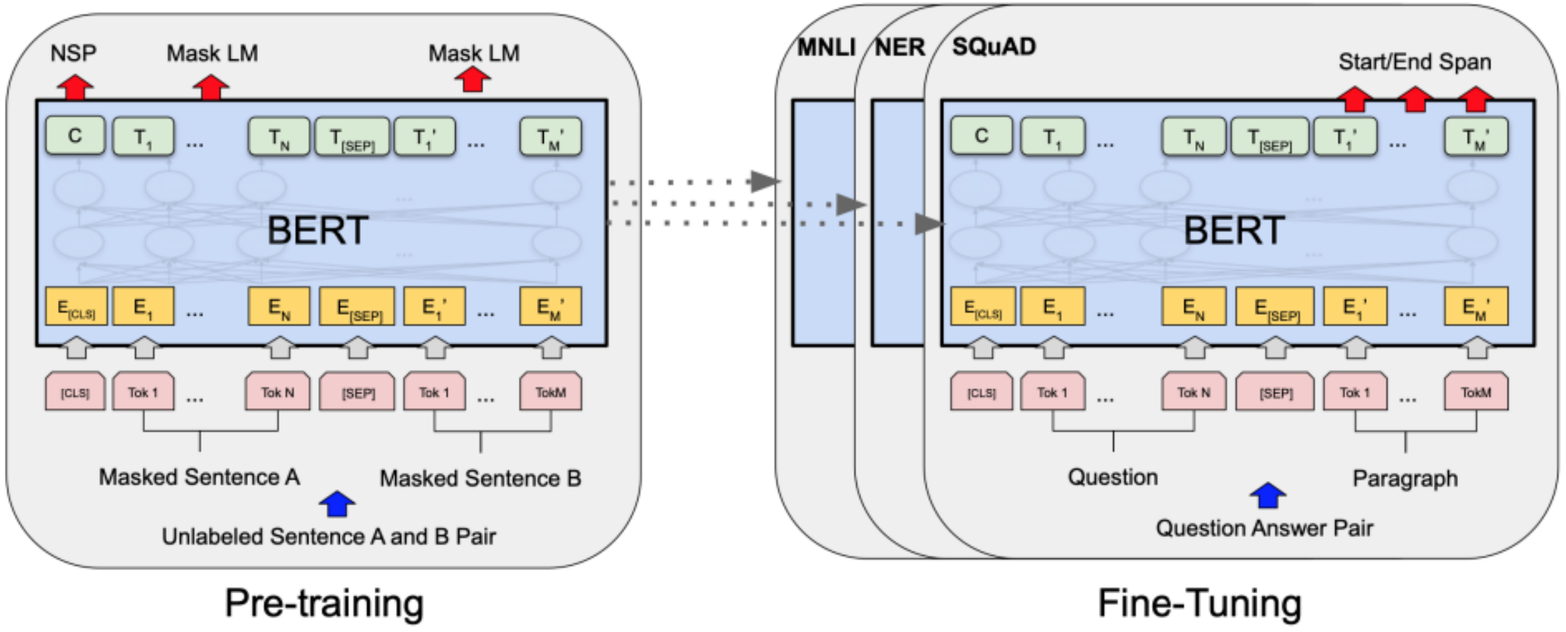
BERT- Bidirectional Encoder Representations from Transformers

BERT, which stands for Bidirectional Encoder Representations from Transformers, is a state-of-the-art language model developed by Google in 2018 for natural language processing (NLP) tasks, such as text classification, named entity recognition, and question answering. BERT is a variant of the Transformer model, which is a type of neural network architecture introduced by Vaswani et al. in the "Attention is All You Need" paper.

Model Evaluation of ML Approaches

	Count Vectorizer	Tf-Idf
Multinomial Naive Bayes Classifier	58.46%	38.37%
Linear SVM	62.00%	38.49%
Logistic Regression	62.47%	40.13%

BERT's execution for these tasks is in two phases viz., pre-training for language understanding, and fine-tuning for a specific task.

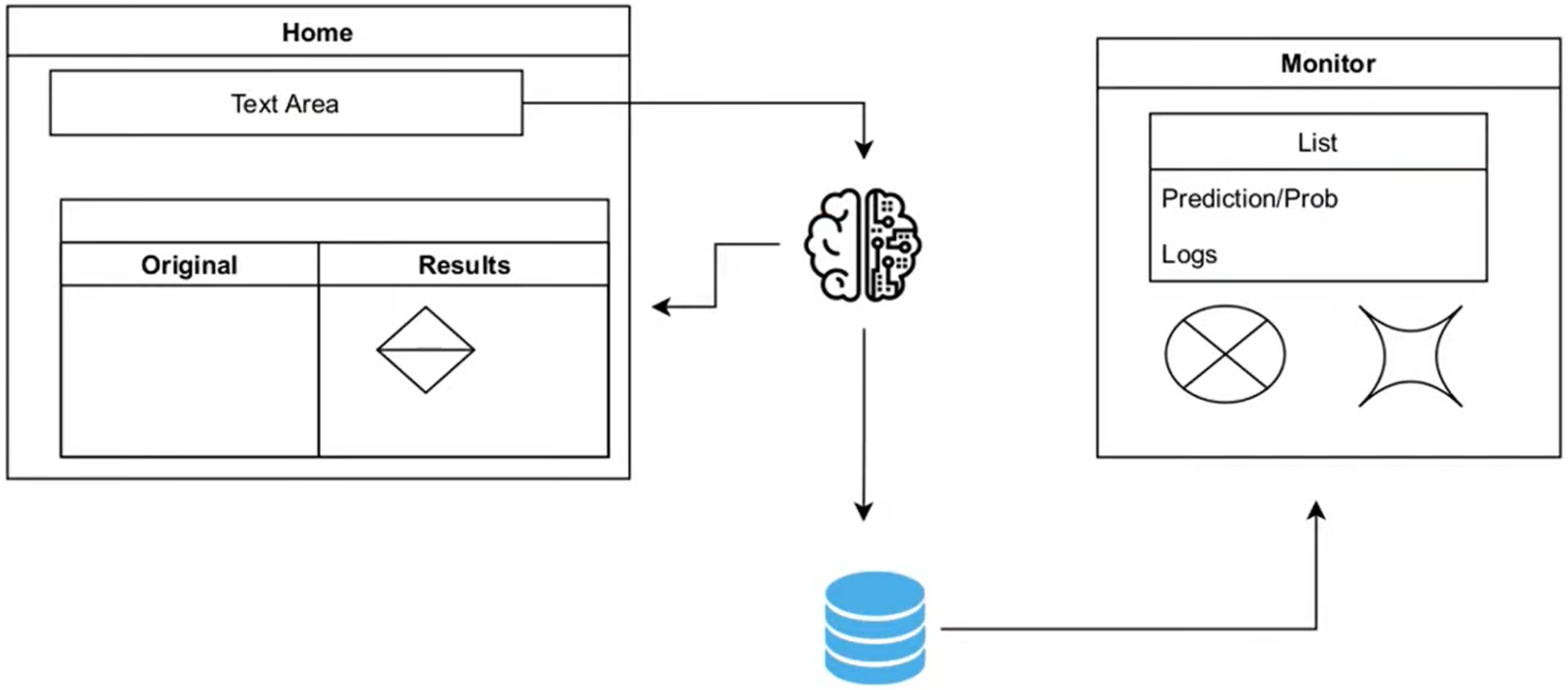


Model Evaluation for BERT

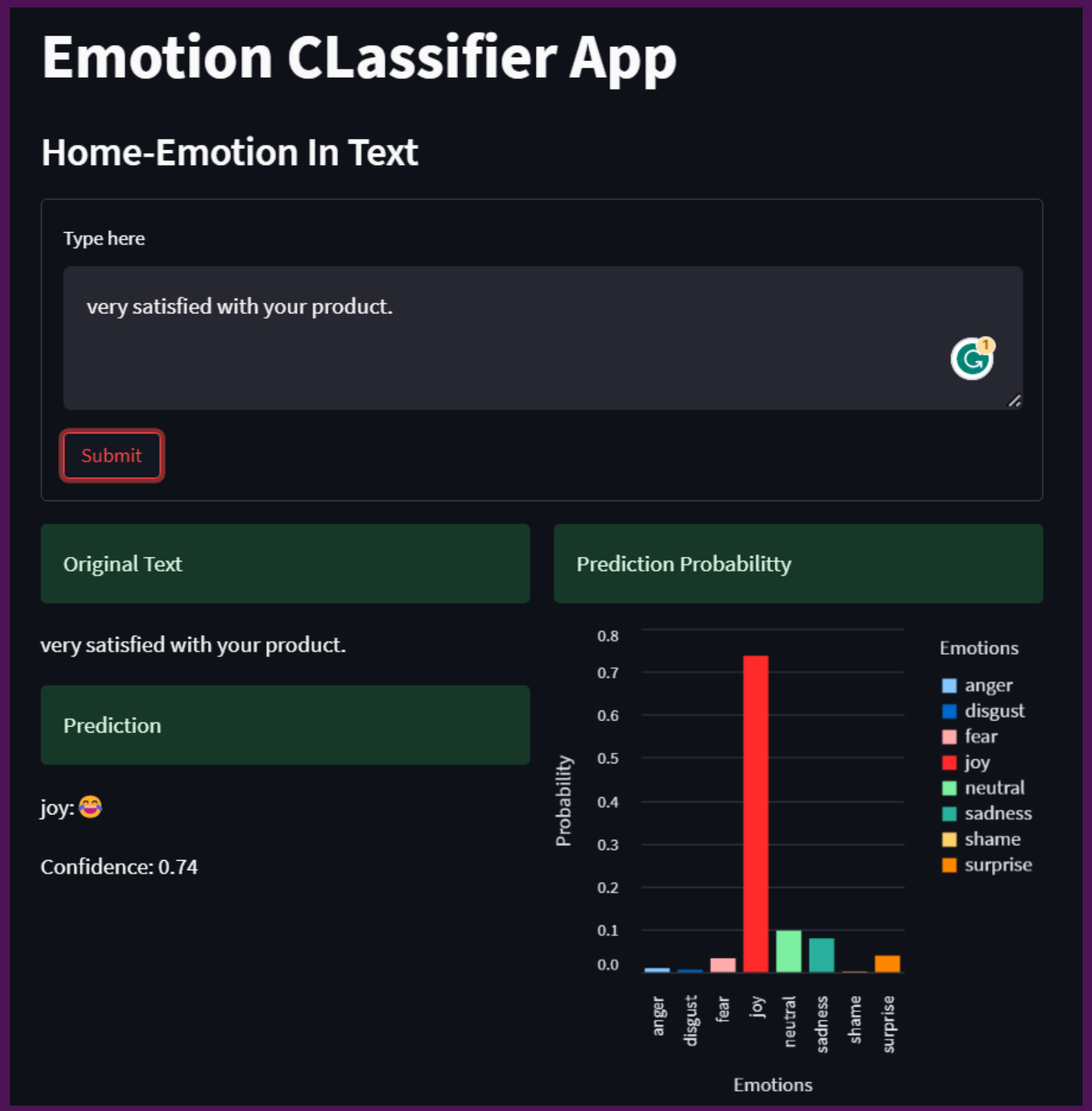
```
Epoch 1
Training loss: 0.9007002512753849
Validation loss: 0.6143069127574563
F1 Score (Weighted): 0.7791319217695921


Epoch 2
Training loss: 0.5381144283001613
Validation loss: 0.6438471145765294
F1 Score (Weighted): 0.8207824902152685


Epoch 3
Training loss: 0.35893184876292417
Validation loss: 0.723008230609435
F1 Score (Weighted): 0.8463474188661483
```

```
Epoch 4
Training loss: 0.2692523200199349
Validation loss: 0.7796335518272365
F1 Score (Weighted): 0.8341132163207956


Epoch 5
Training loss: 0.18156354463565766
Validation loss: 0.8108082735081321
F1 Score (Weighted): 0.8441012614273822
```

App Workflow



OUTPUT


Emotion CLassifier App
Home-Emotion In Text
Type here
very satisfied with your product.
Submit
Original Text
very satisfied with your product.
Prediction
joy: 😂
Confidence: 0.74
Prediction Probabilitty
Probability
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
anger
disgust
fear
joy
neutral
sadness
shame
surprise
Emotions
Emotions
anger
disgust
fear
joy
neutral
sadness
shame
surprise

Emotion CLassifier App
Home-Emotion In Text
Type here
the interface provided for the user is bad
Submit
Original Text
the interface provided for the user is bad
Prediction
sadness: 😔
Confidence: 0.30
Prediction Probabilitty
Probability
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
anger
disgust
fear
joy
neutral
sadness
shame
surprise
Emotions
Emotions
anger
disgust
fear
joy
neutral
sadness
shame
surprise

Scope of Improvement

- System needs improvement to handle code-mixed text such as Hinglish
- Language identification and language-specific features can be incorporated to achieve this
- Language-specific pre-processing techniques and data augmentation can improve system performance on code-mixed text.

Conclusions

- Logistics regression preformed Better than Linear SVM and Multinomial Naive bayes classifier, using CountVectorizers
- Bert based model performed better than classical ML approaches on multi-class emotion Classification

References

- Performance analysis of Ensemble methods on Twitter sentiment analysis using NLP techniques
- https://www.kaggle.com/datasets/pashupatigupta/emotion-detection-from-text
- Text-based emotion detection: Advances, challenges, and opportunities
- https://towardsdatascience.com/bert-explained-state-of-the-art-language-model-for-nlp-f8b21a9b6270
- Transformer Models for Text-based Emotion Detection: A Review of BERT-based Approaches

Thank You

-Presented By:
Rohit Sharma (2020UCO1697)
Vishal Verma (2020UCO1648)
Archit Bansal (2020UCO1709)