



COEP

Technological University

# *Corporate Sentinel*

Harnessing Data to Illuminate Company Reputations

09 May, 2024

# *Group Members*

- |    |                   |           |
|----|-------------------|-----------|
| 01 | Sidhesh Lawangare | 112103138 |
| 02 | Shirshak Tiple    | 112103150 |
| 03 | Vibhav Pande      | 112107043 |

# *Content Overview*

01 Introduction

02 Problem Statement

03 Technologies Used

04 Technical Roadmap

05 Preprocessing and Models

06 Implementation

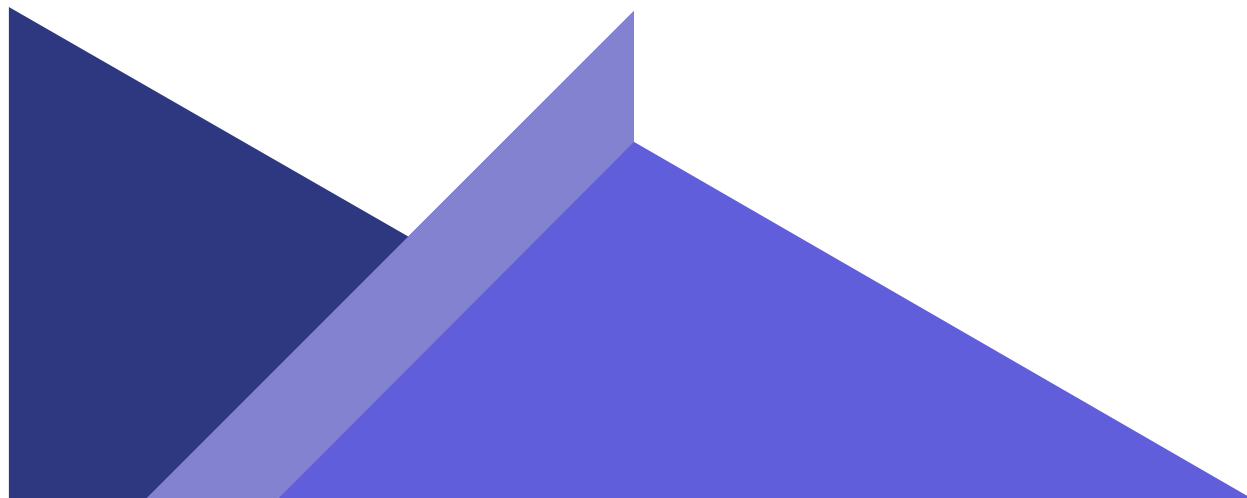
07 Conclusion

08 Future Scope

# *Introduction*

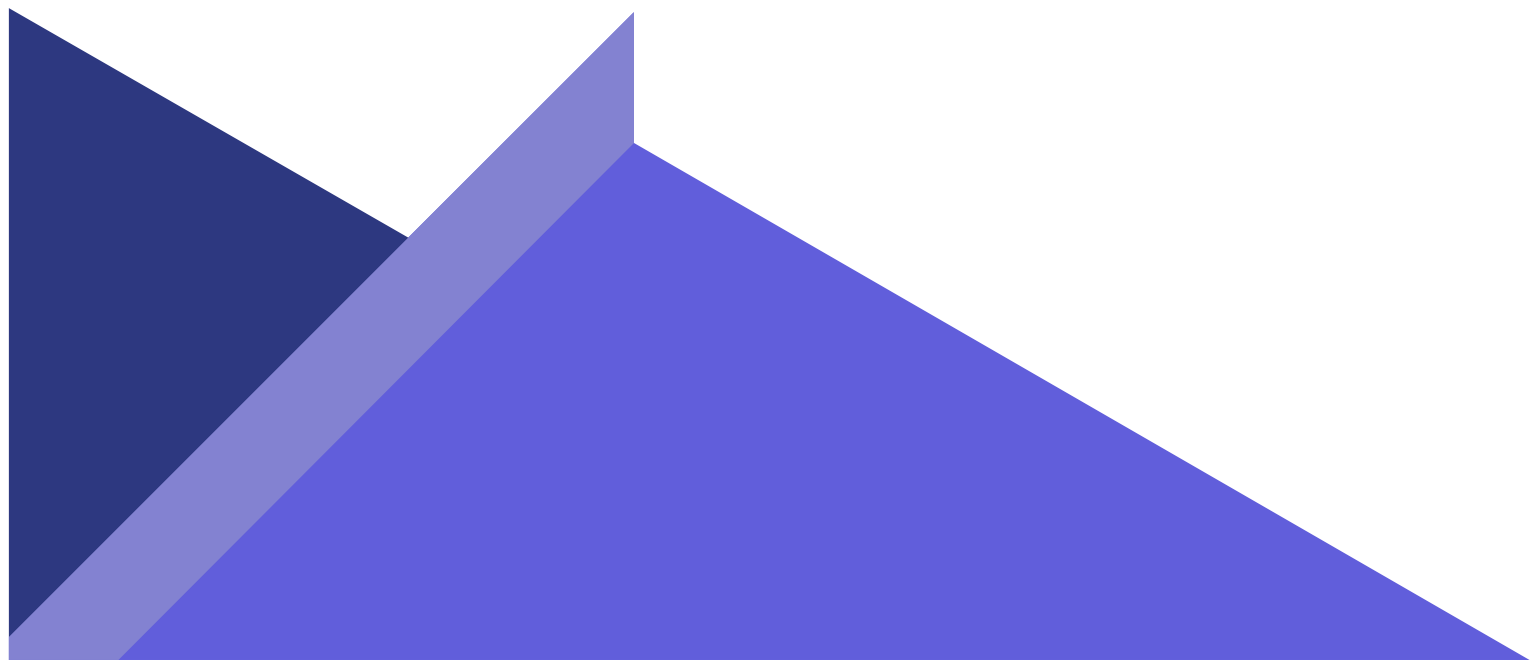
To develop a robust system capable of scraping and analyzing online reviews to identify negative sentiments surrounding companies.

In today's digital age, company reputations are shaped by online discourse. Our project aims to harness the power of web scraping and advanced analytical models to provide companies with actionable insights into their online reputation dynamics.



# *Problem Statement*

Design and implement an automated system to monitor daily news articles from various sources and identify negative sentiment articles related to a specific company.





# *Technologies Used*

**1. Python**

**4. Joblib**

**2. Matplot Lib , Pandas**

**5. Beautiful Soup**

**3. NLTK, sklearn**

**6. TensorFlow/Keras**

# Technical Roadmap

Scraping and Preprocessing	Model Training on Labelled datasets	Model Evaluation	Best Model	Running on scraped data
Scrape relevant data and preprocess it by tokenizing, removing stopwords, and lemmatizing.	Train multiple sentiment analysis models on labeled datasets.	Evaluate models based on metrics like accuracy and loss.	Finding the best model based on the accuracy and losses	Running the model on the scraped data and finding negative articles

# *Preprocessing*

## **1. Data Loading & Integrity**

- Read the CSV file.
- Remove null rows.

## **2. Tokenization**

- Tokenize text into words or tokens.

## **3. Stopword Removal**

- Remove common stopwords.

## **4. Lowercasing**

- Convert tokens to lowercase.

## **5. Apostrophe & Contraction Handling**

- Replace contractions and apostrophes.

## **6. Filtering Non-Alphabetic Tokens**

- Remove non-alphabetic tokens.

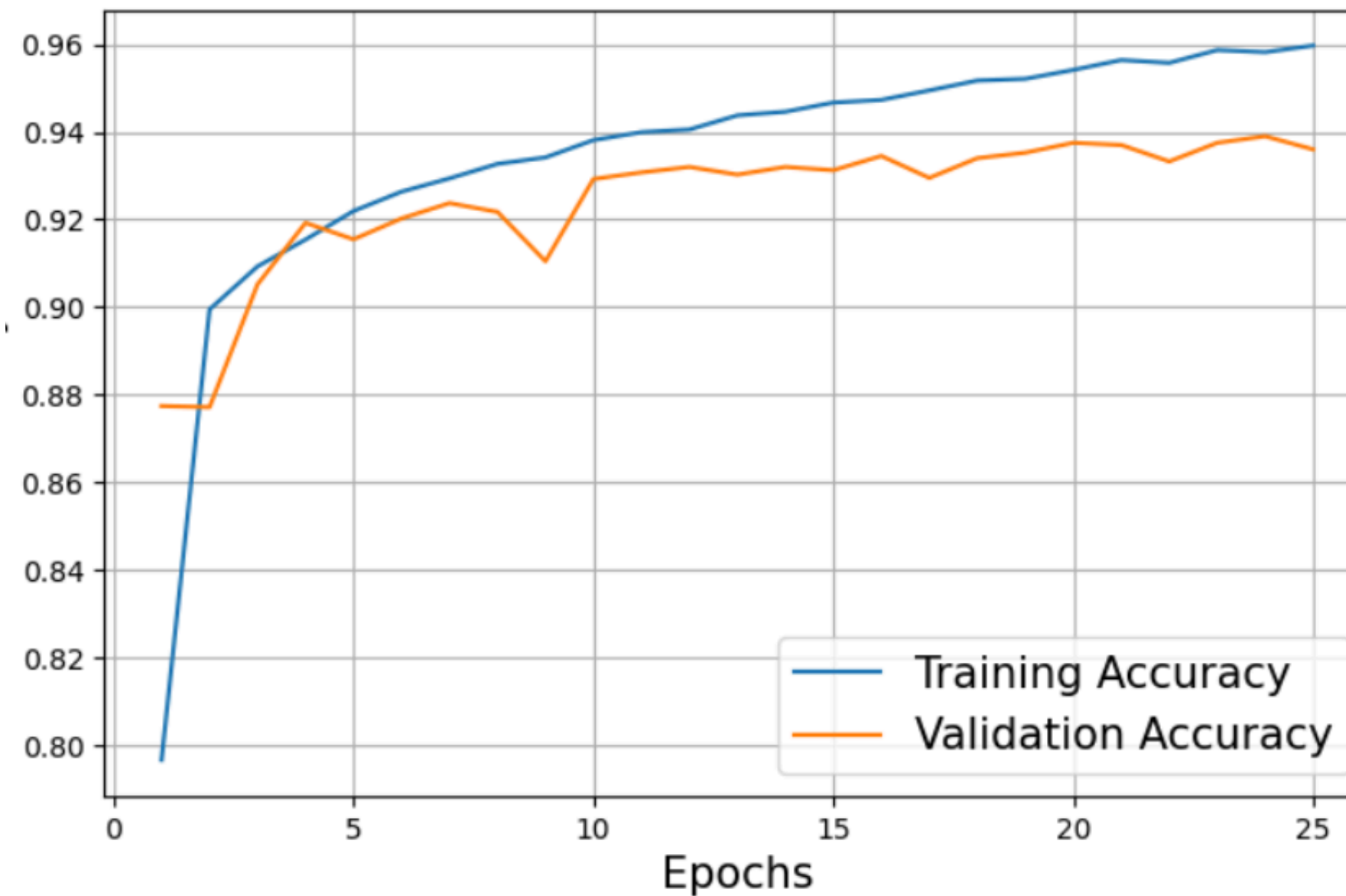
## **7. Lemmatization**

- Reduce tokens to base forms.



# Logistic Regression

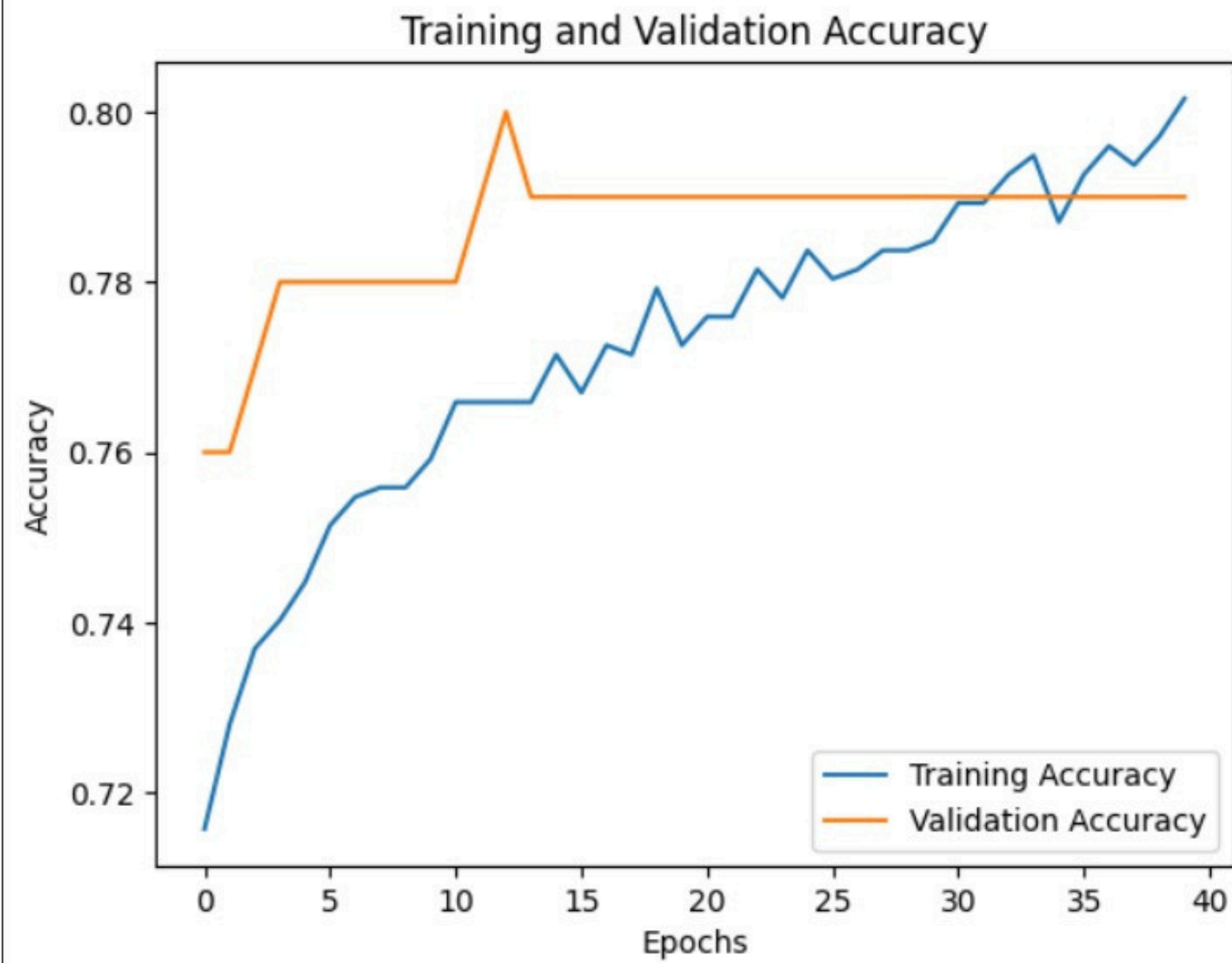
$t(\theta, 0.5, \text{'Accuracy'})$



Estimates the probability of positive or negative sentiment using input features, applying a logistic function to map features to a probability distribution.

```
Epoch 24: val_loss improved from 0.30117 to 0.29846, saving model to BestModel.keras
200/200 - 6s - 28ms/step - accuracy: 0.8821 - loss: 0.2614 - val_accuracy: 0.8523 - val_
Epoch 25/25
```

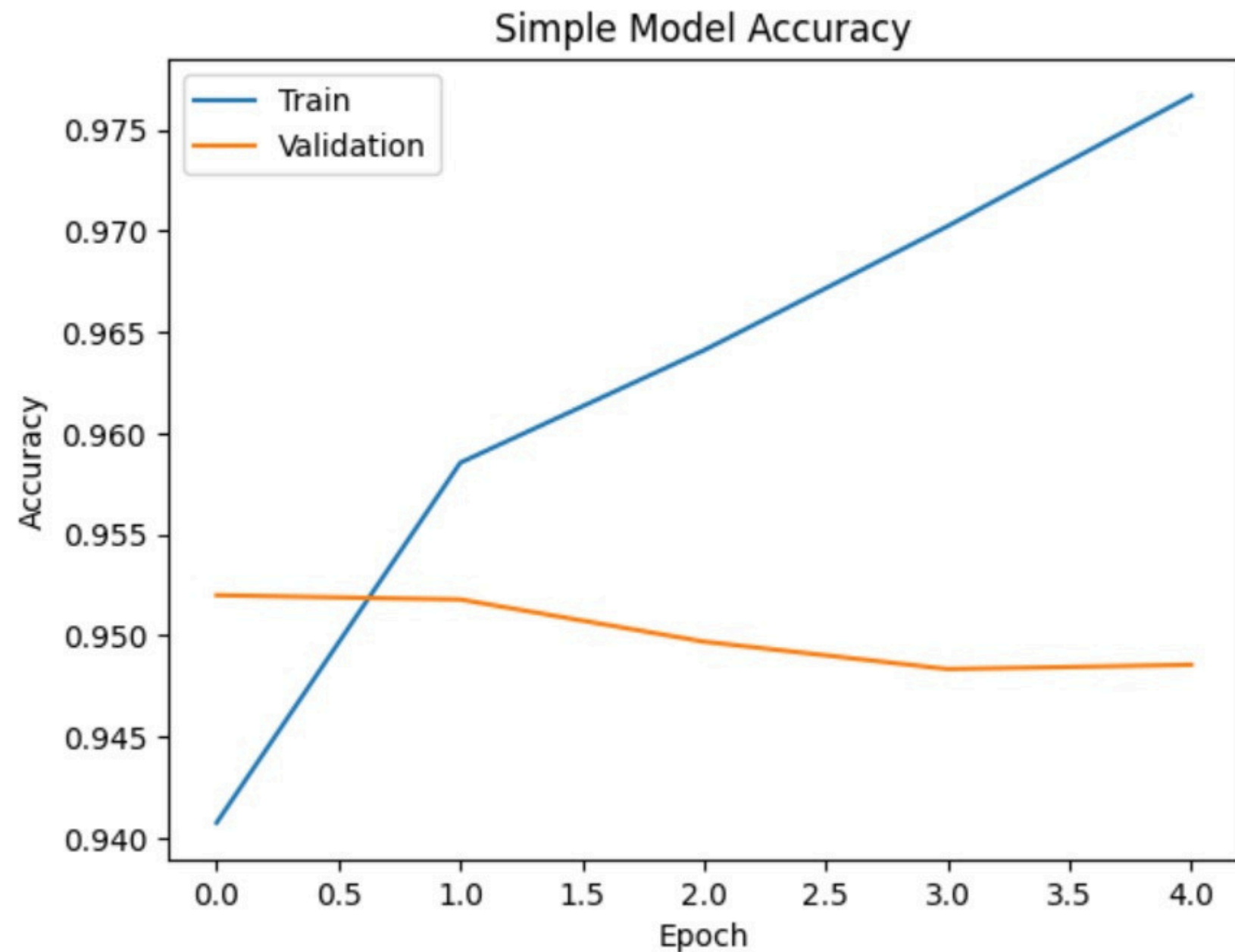
# Convolutional Neural Networks (CNN)



Leverages deep learning architectures to automatically learn intricate patterns and relationships in text data, achieving high accuracy in sentiment classification tasks.

Test Loss: 0.45822933316230774  
Test Accuracy: 0.7720000147819519

# Long-Short Term Memory (LSTM)



LSTM models capture long-range dependencies in text data; linear LSTM processes input sequentially, while bidirectional LSTM considers context from both past and future information, enhancing sentiment analysis performance.

Test Loss: 0.13411585986614227  
Test Accuracy: 0.9524999856948853

# Implementation

The implementation of our project revolves around extracting negative articles about a company from various online sources. Leveraging web scraping techniques and sentiment analysis models, our system identifies and categorizes these negative sentiments.

This output provides companies with valuable insights into public perception and areas requiring attention for reputation management.

```
3/3 [=====] - 1s 12ms/step
Negative Class Instances:

                                     Heading
38  AIMRA accuses Apple India of favouring large r...
53  Apple Researchers Are Building AI Model Called...
65  $14 billion and counting: Apple takes a signif...
80  Apple to Allow Used Parts for Repairs, Startin...

                                     Merged Text
38  As the iPhone 15 series from Apple continues t...
53  Photo Credit: Pexels/Mateusz Taciak\nThe LLM i...
65  My Account\nFollow us on:\nPowered ByhBits - F...
80  Apple todayannouncedthat customers and indepen...
```

# Conclusion

- Bidirectional LSTM outperformed Logistic Regression, CNN, and LSTM models, demonstrating the highest accuracy and the lowest loss in sentiment analysis.
- Logistic Regression had lower accuracy and higher loss compared to LSTM.
- CNN showed higher loss compared to LSTM.
- LSTM, while not the top performer in accuracy, showcased a balance of accuracy and lower loss, making it a competitive choice in sentiment analysis tasks.

# Future Scope

- Integration of text, image, and video data for a more comprehensive understanding of news sentiment across different media types.
- Moving beyond binary sentiment classification to analyze nuanced sentiments such as irony, sarcasm, and mixed emotions.
- Incorporating contextual information such as temporal trends, author credibility, and article topics to improve sentiment interpretation and relevance.



COEP

Technological University

*Thank You*

10 May, 2024