

CSE427

Project Presentation

A Comparative Sentiment Analysis on Tweets using Simple and Complex Models



Presented by:



Md Enayet Ali Labib



Niamul Hasan Chowdhury

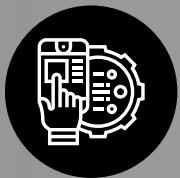


Sifat Sharif

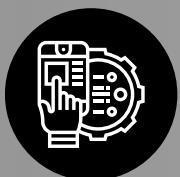
.....

Project Overview

Objectives



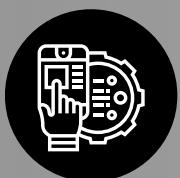
Preprocessing the existing dataset for model training



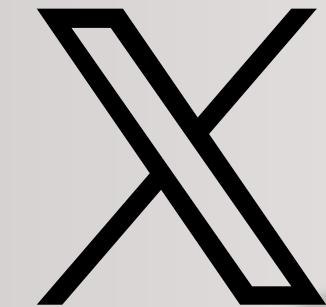
Perform sentiment and emotion classification on the data



Train and Test Multiple Models



Compare traditional ML and deep learning models



Dataset Description



Characteristics

- User-generated noisy social media text
- Contains emotion categories such as Joy, Anger, Fear, Trust



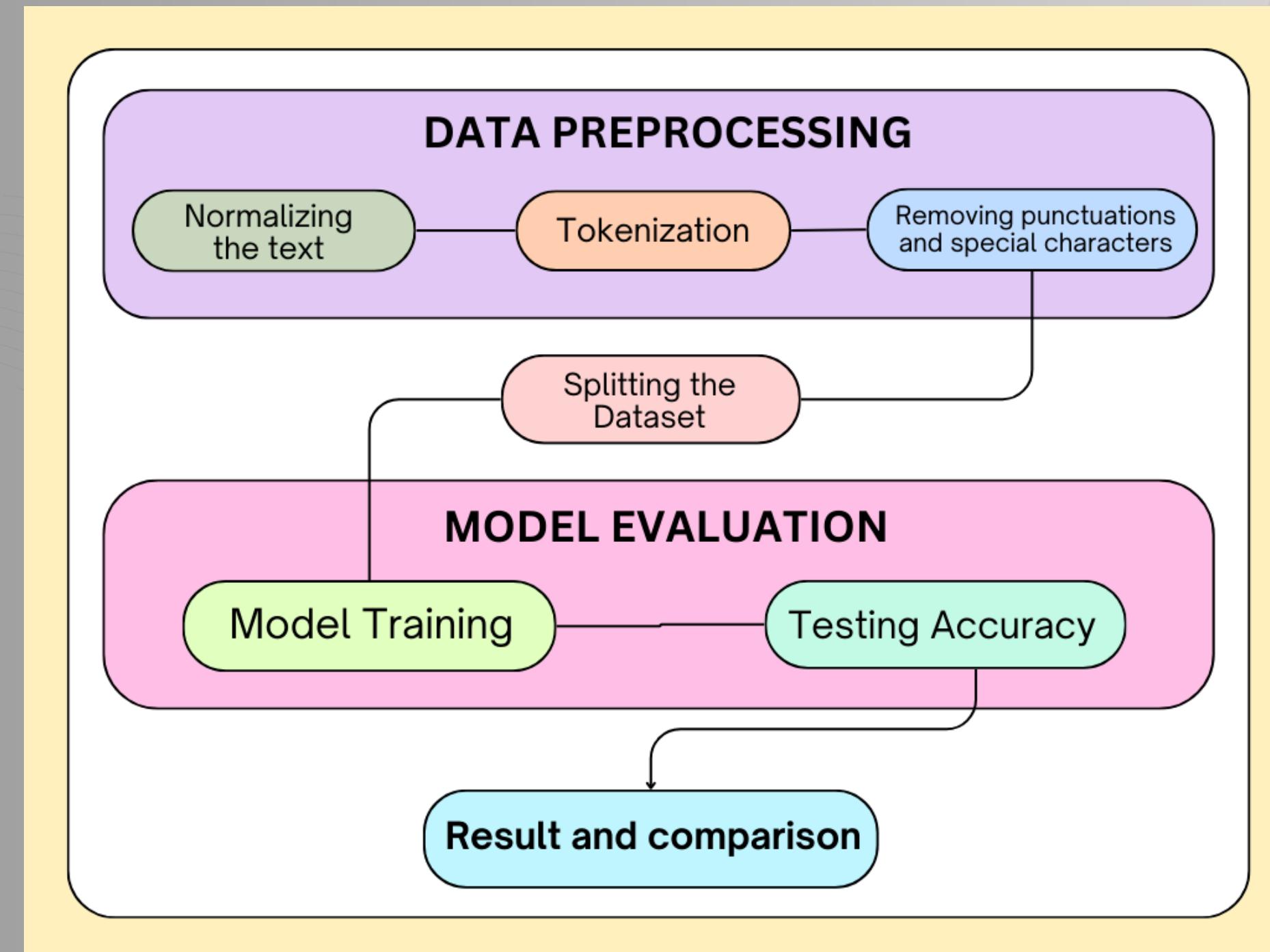
Sentiment Distribution

- Negative: 10,556
- Positive: 7,366
- Neutral: 7,048



Methodology Workflow

○ ○ ○





Data Preprocessing



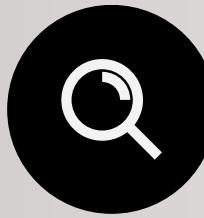
Tokenization and optional stopword removal



Using RegEx for removing punctuations and special characters (#,@)



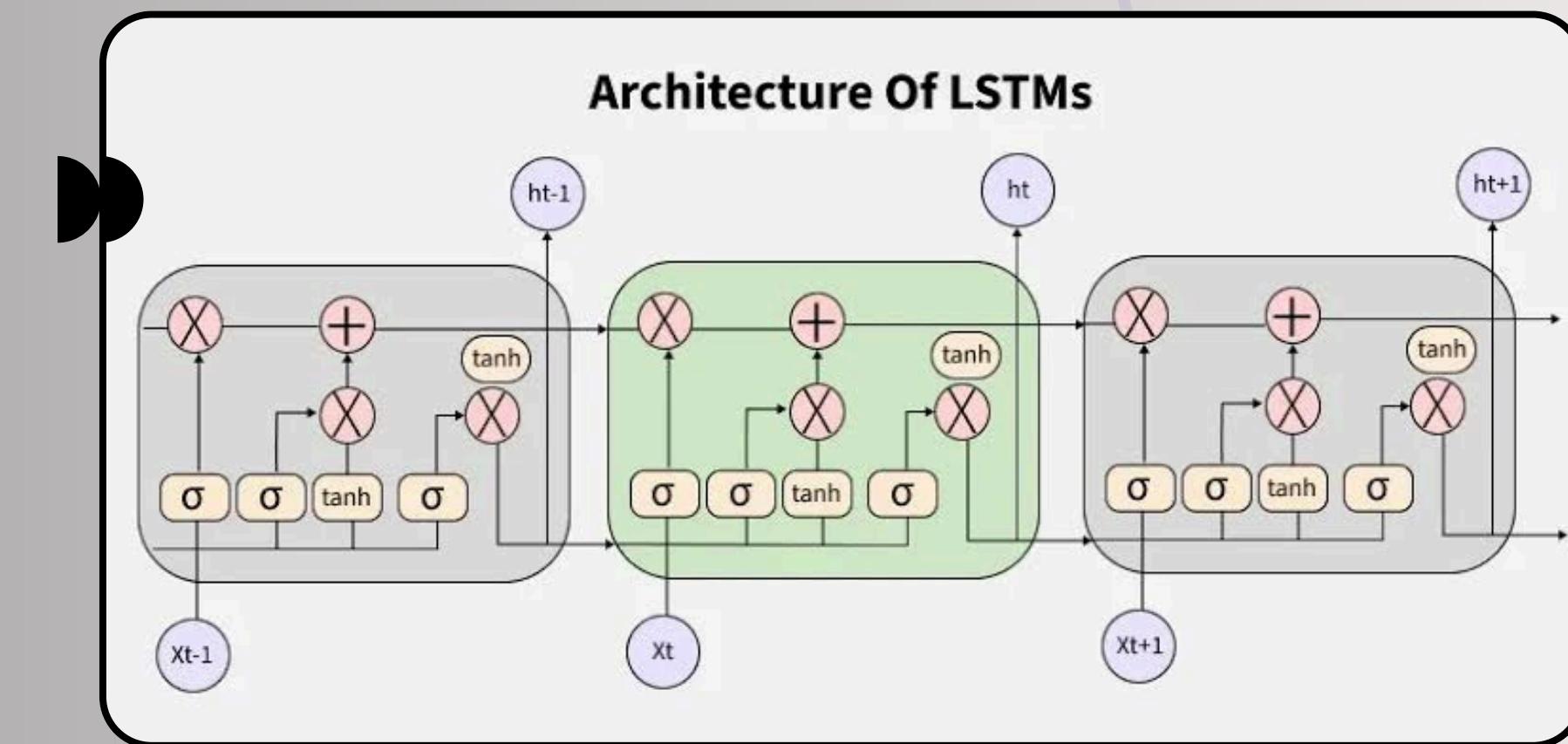
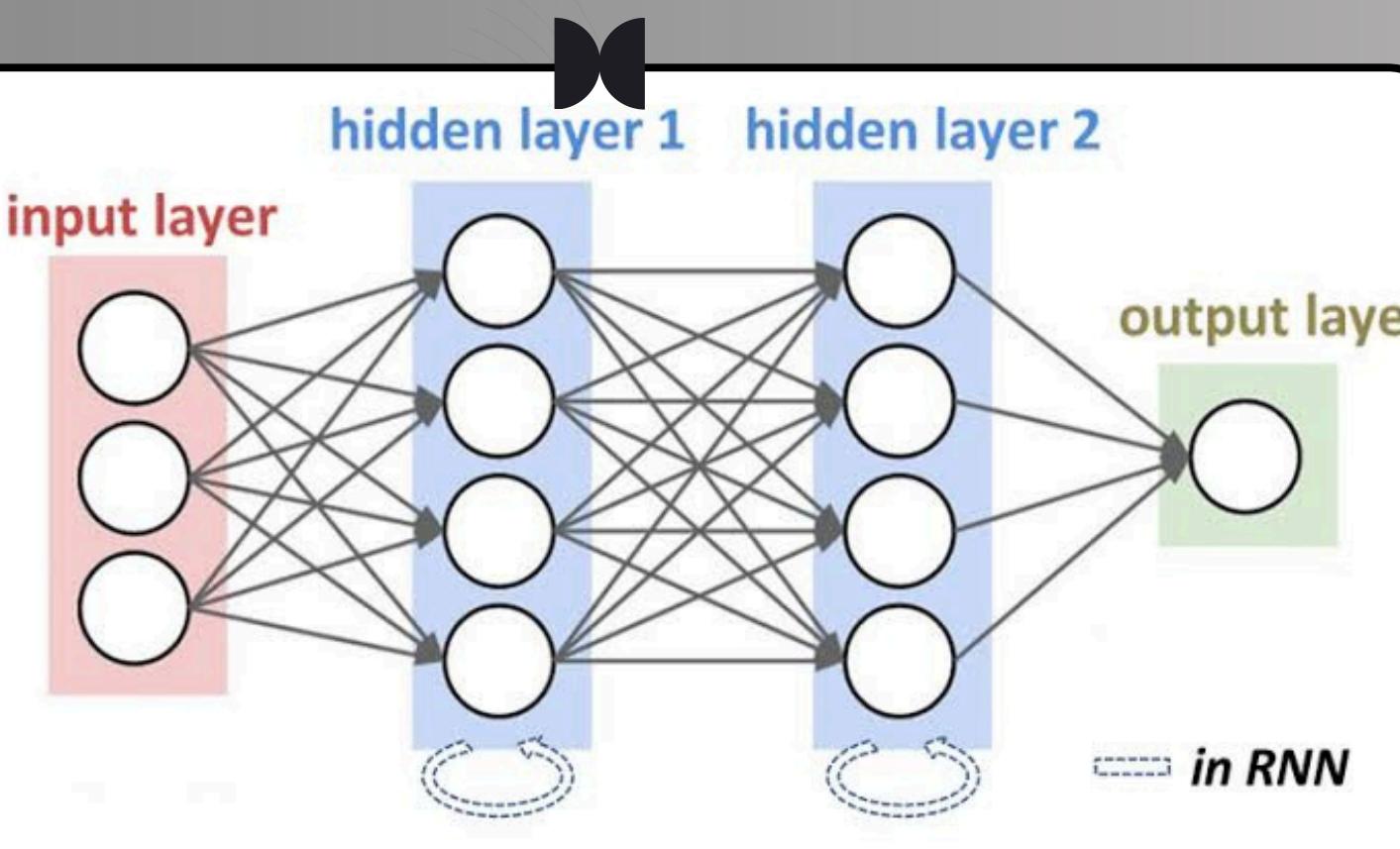
Retained hashtag keywords while removing symbols



Lower-casing and whitespace normalization



Model Implementation



input layer

hidden layer 1 hidden layer 2

output layer

in RNN

Naive Bayes, Logistic Regression, KNN, Random Forest

Tackled model overfitting by Dropout regularization and Early Stopping Strategy

Deep Learning models include RNN, Bidirectional RNN and LSTM

Key Findings and Observations



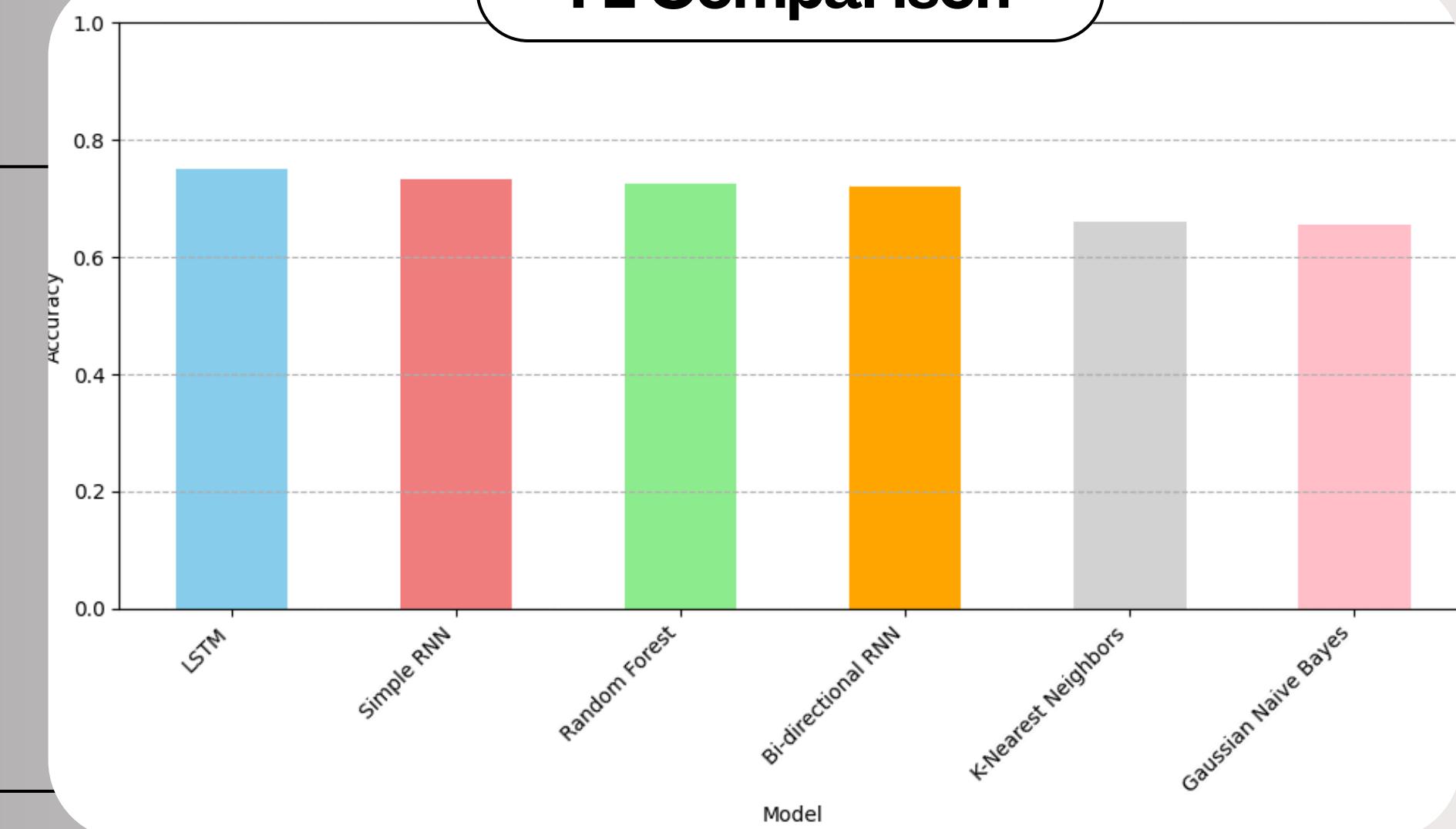
- LSTM achieved the highest accuracy (0.7501)
- Simple RNN and Random Forest also performed strongly
- RNN-based models outperformed traditional ML models
- Naive Bayes and KNN showed the lowest performance

...

...

...

F1 Comparison



Conclusion



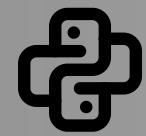
-  Tweet preprocessing significantly improves classification accuracy
-  Models capable of capturing sequential and contextual dependencies in text are better suited for analyzing compared to purely feature-based classifiers.
-  Deep learning models are able to capture contextual and emotional nuances better



Thank you



Sentiment & Emotions Labelled Tweets



Sentiment Analysis from Tweets.ipynb



CSE427 Project Report

• • • •

