SENTIMENT ANALYSIS

TWITTER DATA

Android & Iphone

By GROUP ONE

Business Understanding

✓ Apple and Google are two of the most influential tech companies globally, with millions of users and customers who actively discuss their products on platforms like Twitter.

✓ However, manually analyzing sentiment in these discussions would be both time-consuming and impractical given the sheer volume of tweets.

✓ Therefore, automating sentiment analysis using an NLP model will allow both companies to process and understand consumer opinions quickly and at scale.

✓ This could give them a competitive edge by allowing for a timely response to customer feedback, identifying emerging issues, and improving customer satisfaction and brand loyalty

Project Overview

- Social media platforms like Twitter provide a great amount of realtime insights into customer opinions, especially regarding products and services. For tech gaints like Apple and Google, monitoring sentiments around their products can be a crucial business strategy. Understanding how consumers feel about their products or services, whether positively, negatively or neutrally, helps inform marketing strategies, product development and customer support.

- This project aims to build a Natural Language Processing (NLP) model to analyse the sentiment of tweets related to Apple and Google Products.

OBJECTIVES

Main Objective

To develop an NLP-based machine learning model that can accurately classify the sentiment of tweets related to Apple and Google products as positive, negative, or neutral.
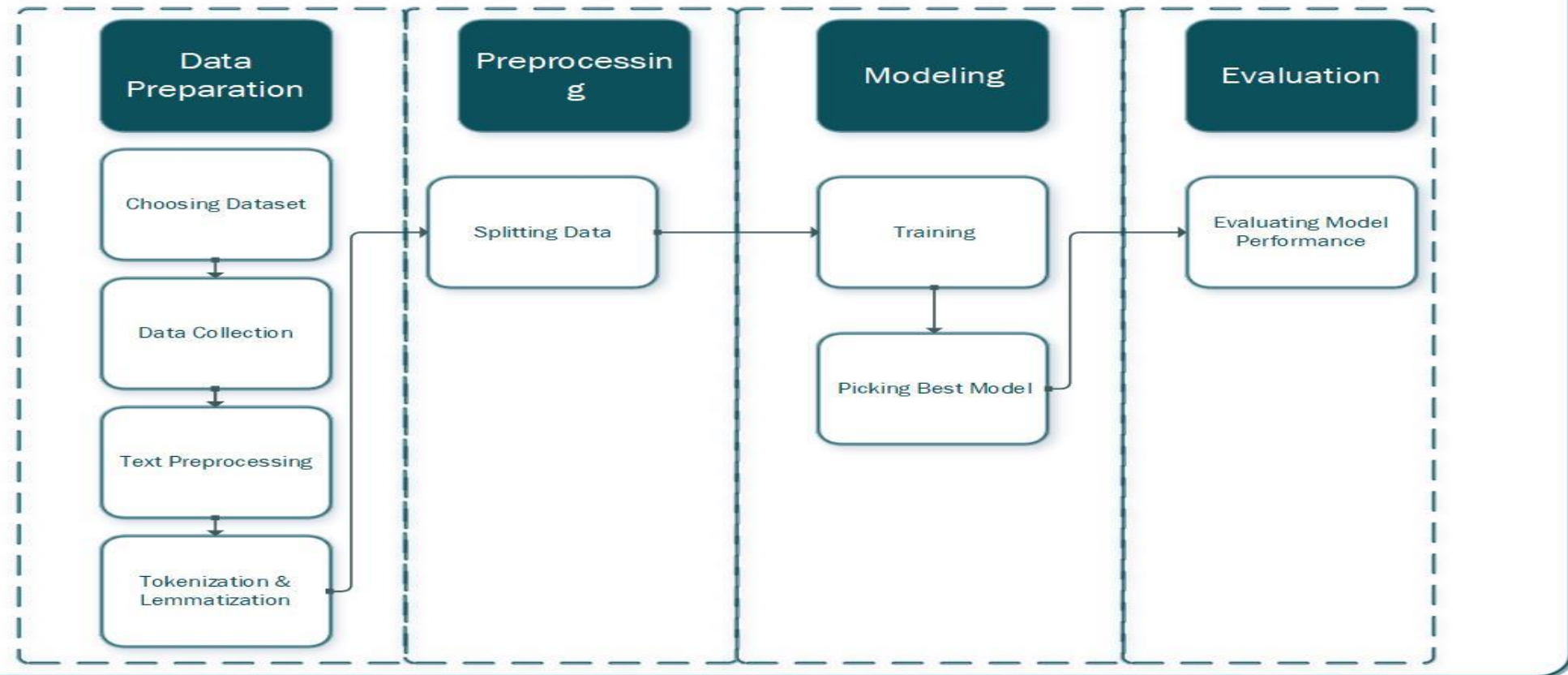
Specific Objectives

Build and evaluate multiple machine learning and deep learning models to classify tweet sentiments.

Preprocess tweet data effectively through removing noise, handling stopwords and tokenization, to improve model accuracy.

Identify patterns in customer sentiments towards Apple and Google products.

METHODOLOGY

CRISP-DM

Data Preparation	Preprocessing	Modeling	Evaluation
Choosing Dataset	Splitting Data	Training	Evaluating Model Performance
Data Collection		Picking Best Model	
Text Preprocessing			
Tokenization & Lemmatization			

Data Collection

The dataset is from CrowdFolder via data.world. The dataset consists of over 9,000 tweets that have been rated by human annotators as positive, negative or neutral sentiment.

Data Description

Column	Data Type	Description
tweet_text	object	It contains text data.
emotion_in_tweet_is_directed_at	object	It contains categorical or string values indicating emotion direction.
is_there_an_emotion_directed_at_a _brand_or_product	object	It holds categorical values ('Negative emotion', 'Positive emotion', 'No emotion toward brand or product' and 'I can't tell').

LIBRARIES

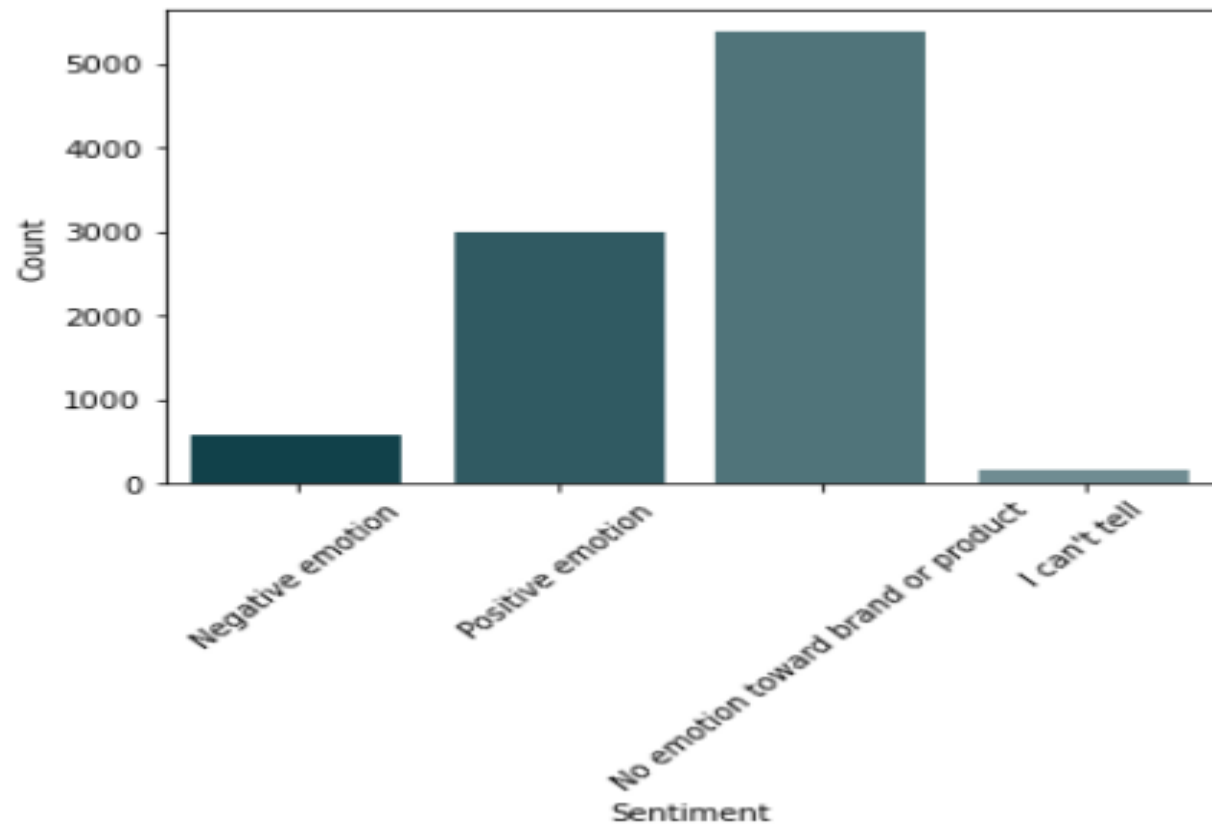| Data Manipulation	Text Processing	Visualization	Preprocessing	Modeling
Numpy	*Natural Language Processing Kit* *Regex*	*Matplotlib* *Seaborn* *Word Cloud*	*Label Encoder* *Tokenizer* *Vader*	Sklearn Tensorflow Keras

ANALYSIS

Tweet Sentiment Distribution

Most people have no emotion towards a most brands and few people can tell

Distribution of Brands/Products Mentioned

Most respondents emotion is directed to unknown barnds and iphone products(iphone and ipad) follow as top 3 brands which emotions are directed at.

Distribution of Brands/Products Mentioned

Most respondents emotion is directed to unknown brands and iphone products(iphone and ipad) follow as top 3 brands which emotions are directed at.

Word Cloud from Tweets

The words in most frequently used words in the tweets is link, retweet
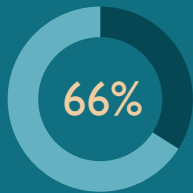
Top 10 Most Frequent Words by Emotion
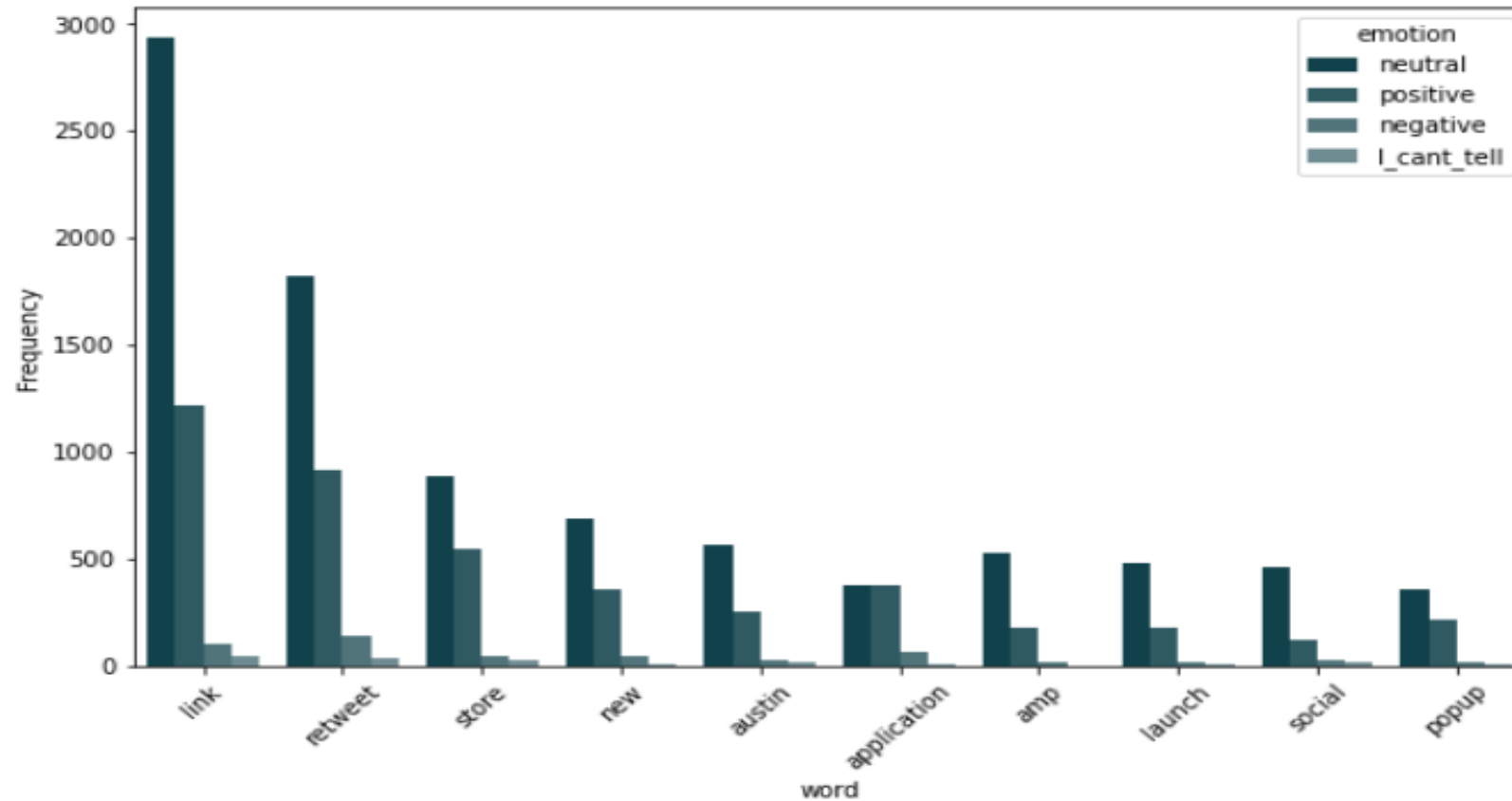
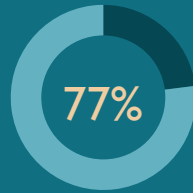Most respondents emotion is associated seems to be distributed evenly across the words.

EVALUATION

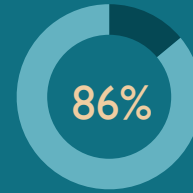
Models

Random Forest — 66%

Logistic Regression — 77%

Multiclass Neural Network — 86%

LSTM Neural Network — 89%

Random Forest: 84% accuracy, 77% macro F1 — better than Logistic Regression, interpretable, captured complex patterns.

Logistic Regression: 66% accuracy — weak across sentiment categories.

Neural Network: 86% accuracy, 82% macro F1 — improved performance but struggled with tone and structure.

LSTM: 89% accuracy, 84% macro F1 — best overall, handled emotion, structure, and complexity well.

CONCLUSION

On Models

LSTM demonstrated the highest effectiveness in understanding language in a better way, making it the most suitable model for sentiment analysis on social media text.

Advancements

Future enhancements could include more training data, use of pre-trained embeddings like GloVe or BERT, or expanding into bidirectional LSTMs for even deeper context learning.

THE TEAM

Bruce Siti | *Group Leader*
Okech Maxmilian | *Member*
Bertha Karuku | *Member*
Roy Gichia | *Member*
Neema Elaly | *Member*
Mich Sego | *Member*

THANK YOU!