NLP

Sentiment Analysis Project

Phase 4 | Project 4| Group 4

Meet the group

Derrick Waititu

Collins Kwata

Knight Mbithe

Levis Gichuhi

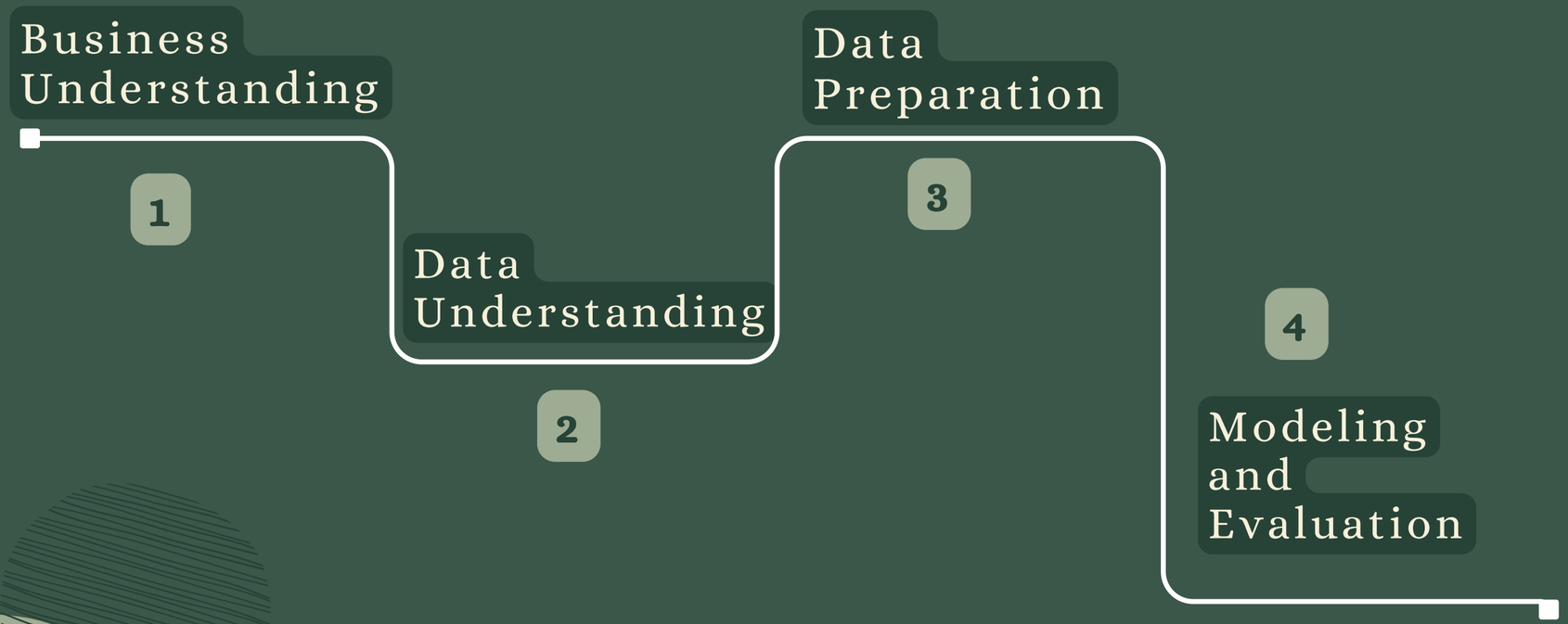
Joy Gitau

Project Overview

Social media platforms are invaluable for understanding customer opinions and experiences with brands and products.

This project analyzes the sentiment of tweets about Apple and Google products to uncover customer perceptions. Using a dataset of over 9,000 tweets labeled, the analysis applies Natural Language Processing(NLP) analysis to uncover insights that can guide development, marketing strategies, and customer engagement.

Project Steps



Project Objectives

1. To develop a Natural Language Processing (NLP) model that can accurately classify the sentiment of social media posts into categories such as positive, negative, and neutral

2. To examine the distribution of sentiment labels and individual variables in the dataset and identify relationships between variables in the dataset.

Data Preparation

We prepared the data for explorative analysis and modelling by carrying out the following steps:

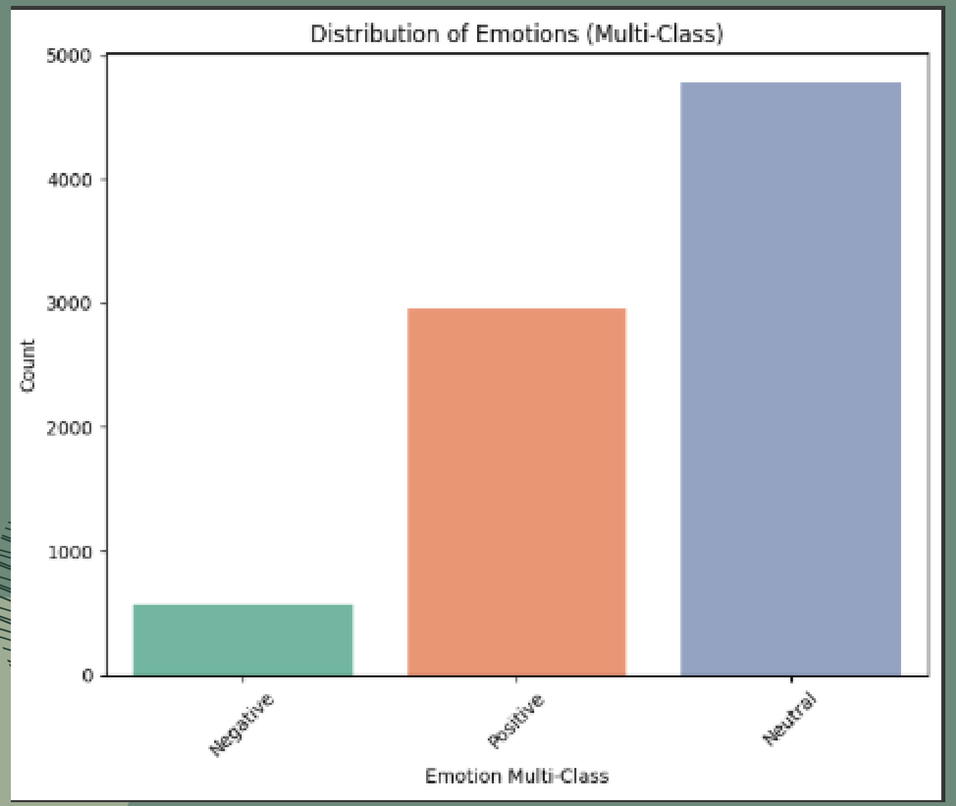
Cleaning- removing irrelevant things, like extra spaces, special characters, or duplicate entries, missing values to make it more useful and clear for analysis.

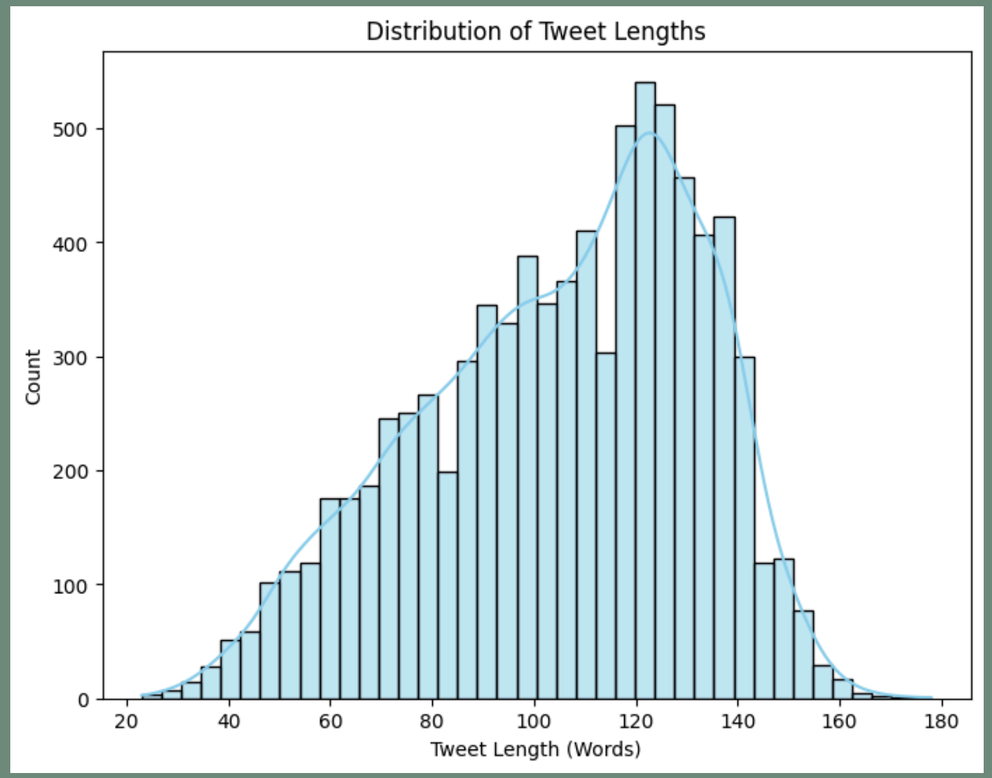
Text preprocessing- cleaning and organizing text into formats a computer can better understand and analyze it.

Encoding- converting categorical columns text or labels into numbers so a computer can process them.

Data Understanding

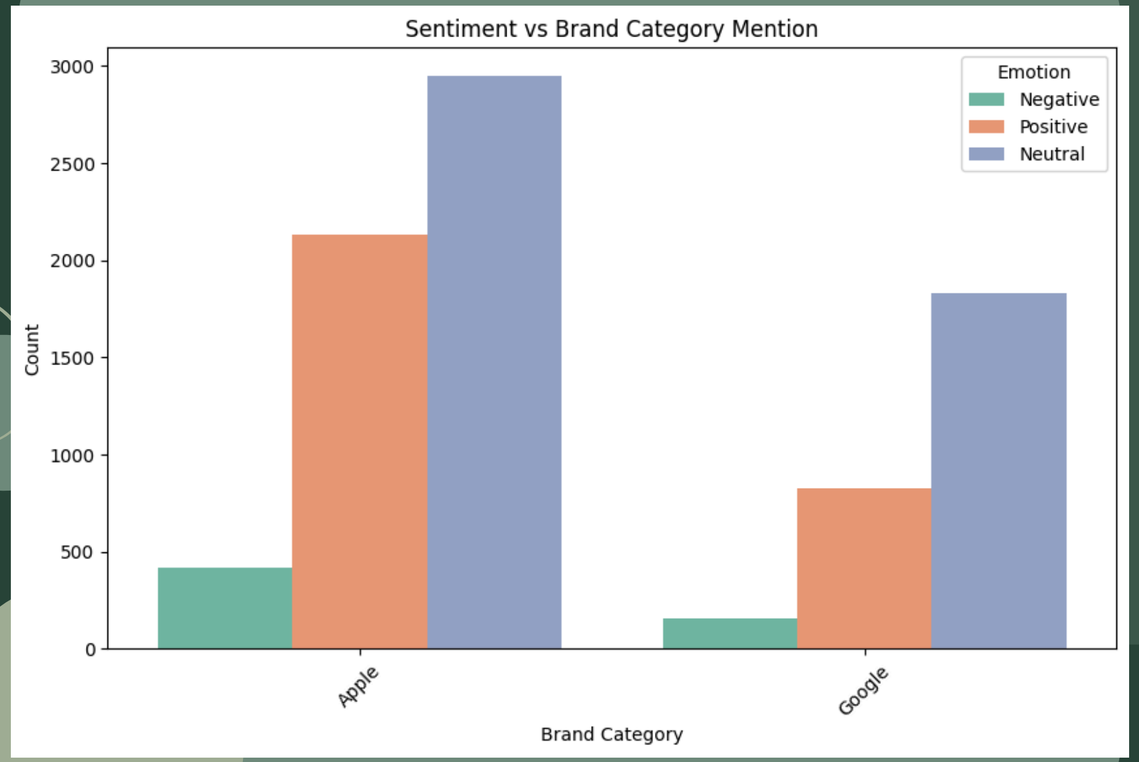
Univariate Analysis:
we visualized the distributions brands, emotions and length of tweets in our data

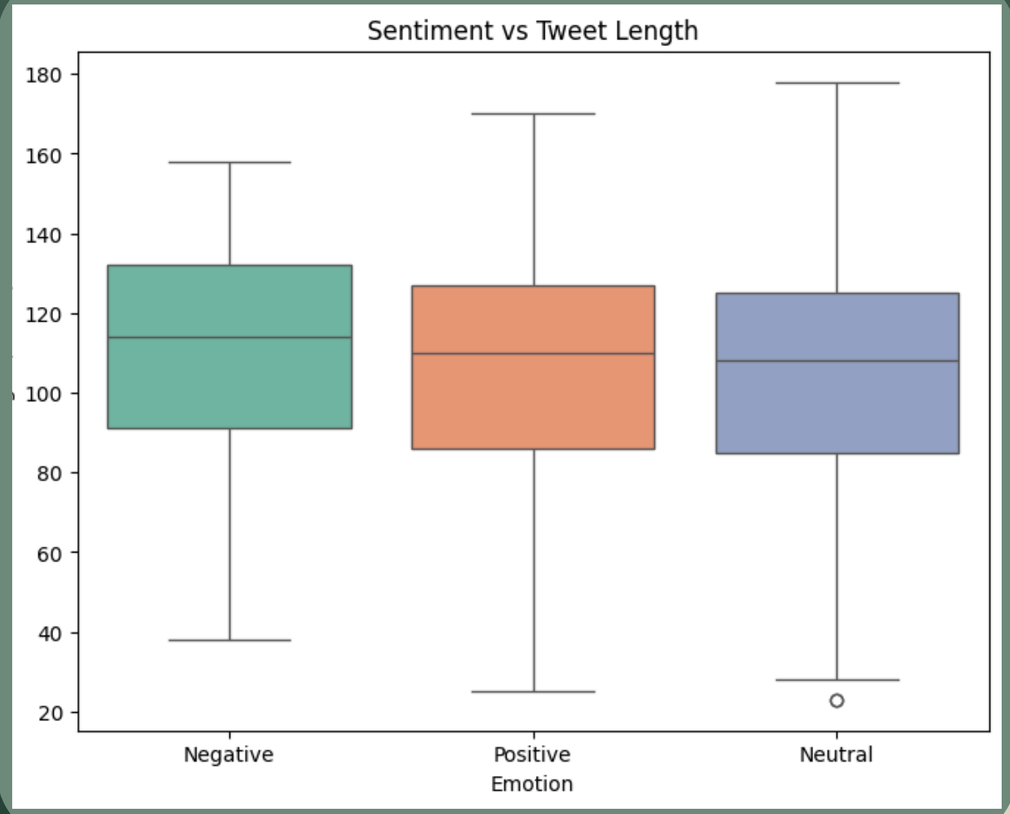




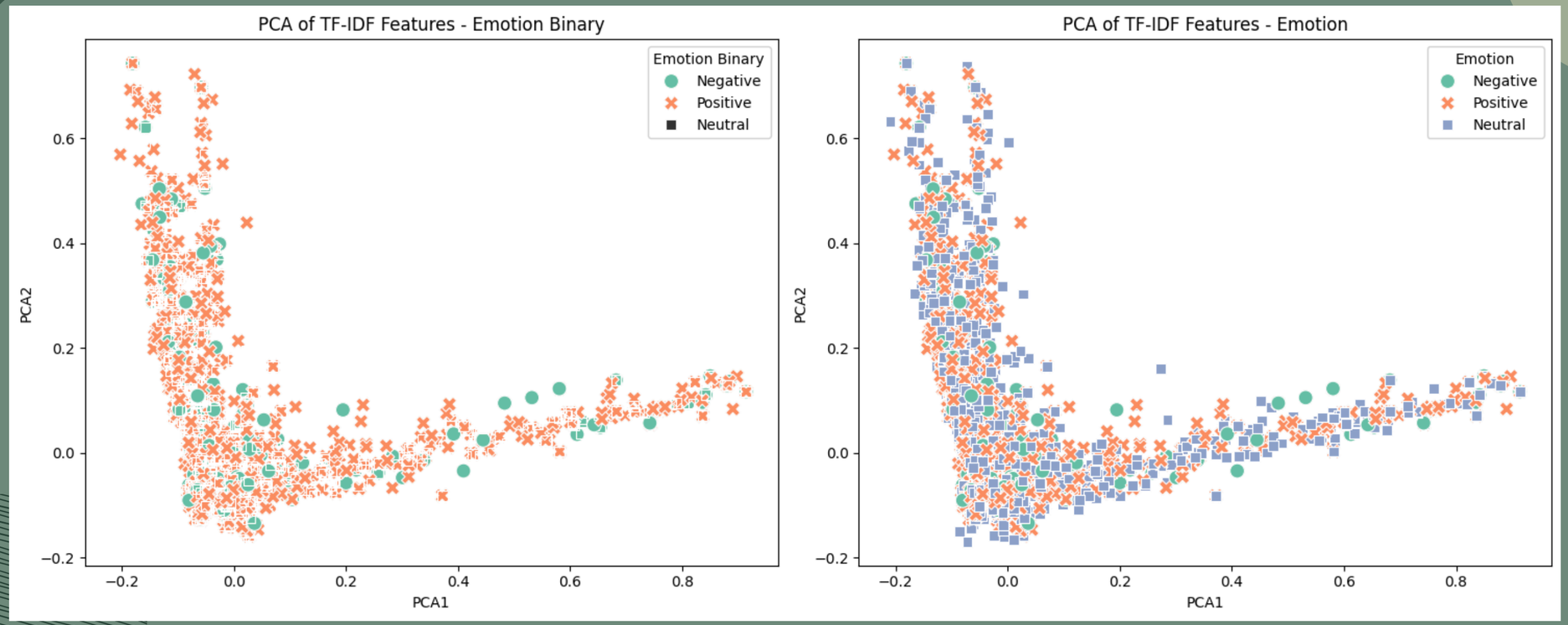
Data Understanding

Bivariate Analysis: we visualized the relationship amongst the features.





Data Understanding



Modeling and Evaluation

We trained and tested the following models; LogisticRegressionModel, RandomForestClassifier, Xgboost Model, SVMModel

SVM and XGBoost consistently deliver superior performance in both binary and multi-class tasks, achieving the highest accuracy and F1-scores suggesting that they are the most reliable options for our sentiment analysis

Conclusion and Recommendation

Focus on using SVM and XGBoost since they work well for identifying emotions in tweets. Improve BERT to better understand complex emotions or combine it with these models for even better results. Try methods that allow tweets to have more than one emotion and use smarter ways to understand the meaning of text, like word embeddings. Also, consider using deep learning models like LSTMs to get more context from tweets and analyze specific product features that people feel strongly about, whether positive or negative.

Thank You