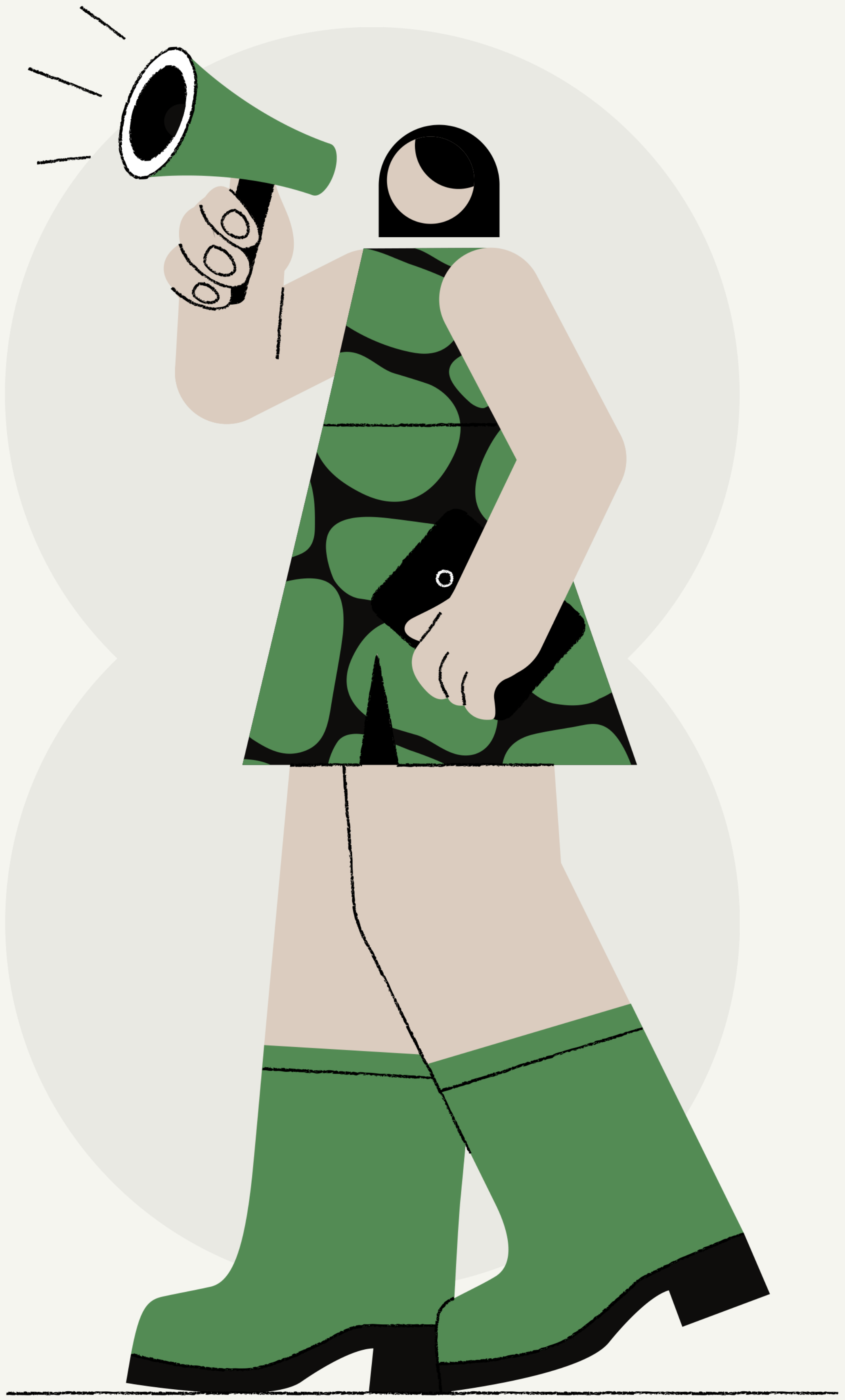
Customer review Sentiments

SC1015 - Data Science and Artificial Intelligence

FCSD_TEAM 4
CHENG YOU JIE ERIC - U2321411D
CHEW WEI HAO KOVAN -U2322227A
JOSHUA LOW SHAO ZHENG - U2320829C

Table of Contents

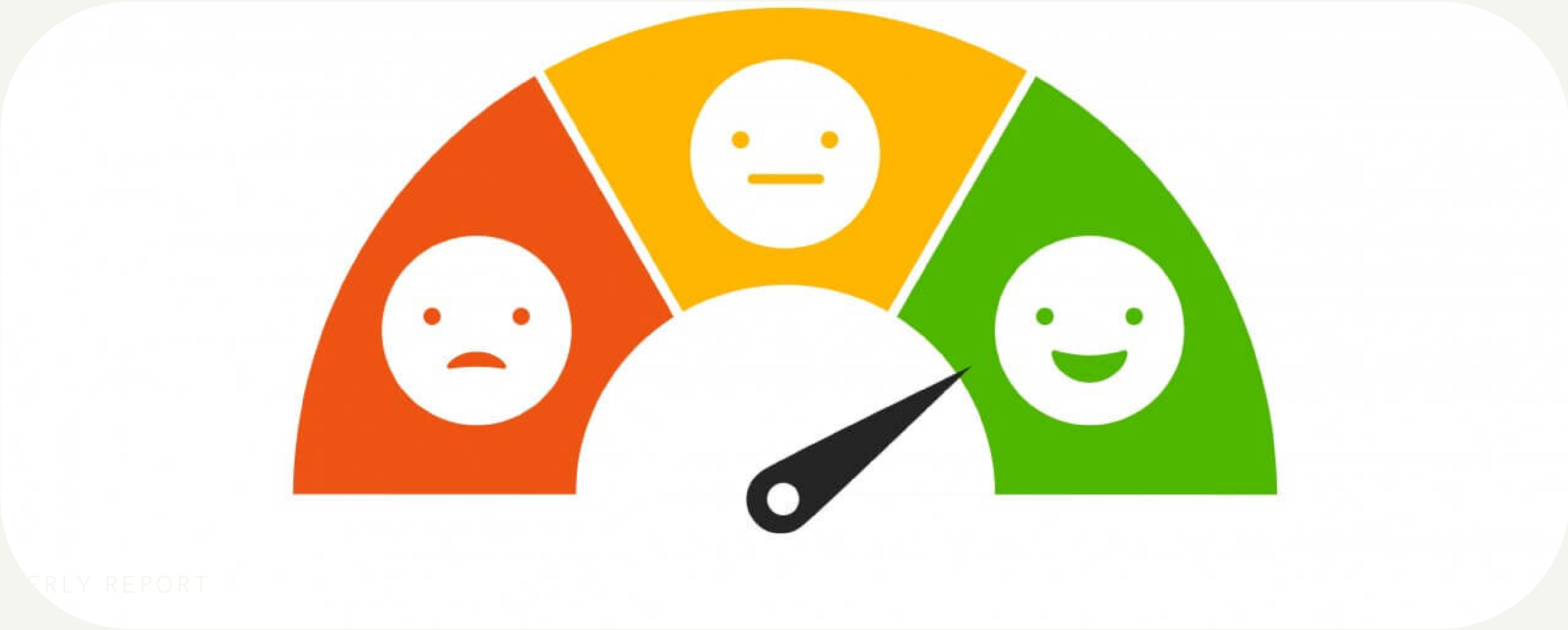


1

Problem Statement and Motivation

Problem Statement

Companies have to spend resources to sort through their customer review to figure out customers' likes and dislikes



Our Motivation

Help companies to streamline the process of determining the experience of their customers by conducting sentiment analysis on customer reviews by focusing on the negative sentiments

Dataset

- Dataset taken from Kaggle regarding customer comments/reviews for tech companies such as Google, Dell, Apple and Microsoft etc
- Dataset consists of 24,970 rows of data

```
                                                Text        Username  \
0  @Logitech @apple @Google @Microsoft @Dell @Len...  ManjuSreedaran
1  @MK_habit_addict @official_stier @MortalKombat...       MiKeMcDnet
2  As @CRN celebrates its 40th anniversary, Bob F...         jfollett
3  @dell your customer service is horrible especi...        daveccarr
4  @zacokalo @Dell @DellCares @Dell give the man ...       heycamella

  sentiment  sentiment_score      emotion  emotion_score
0   neutral         0.853283  anticipation       0.587121
1   neutral         0.519470           joy       0.886913
2  positive         0.763791           joy       0.960347
3  negative         0.954023         anger       0.983203
4   neutral         0.529170         anger       0.776124  )
```

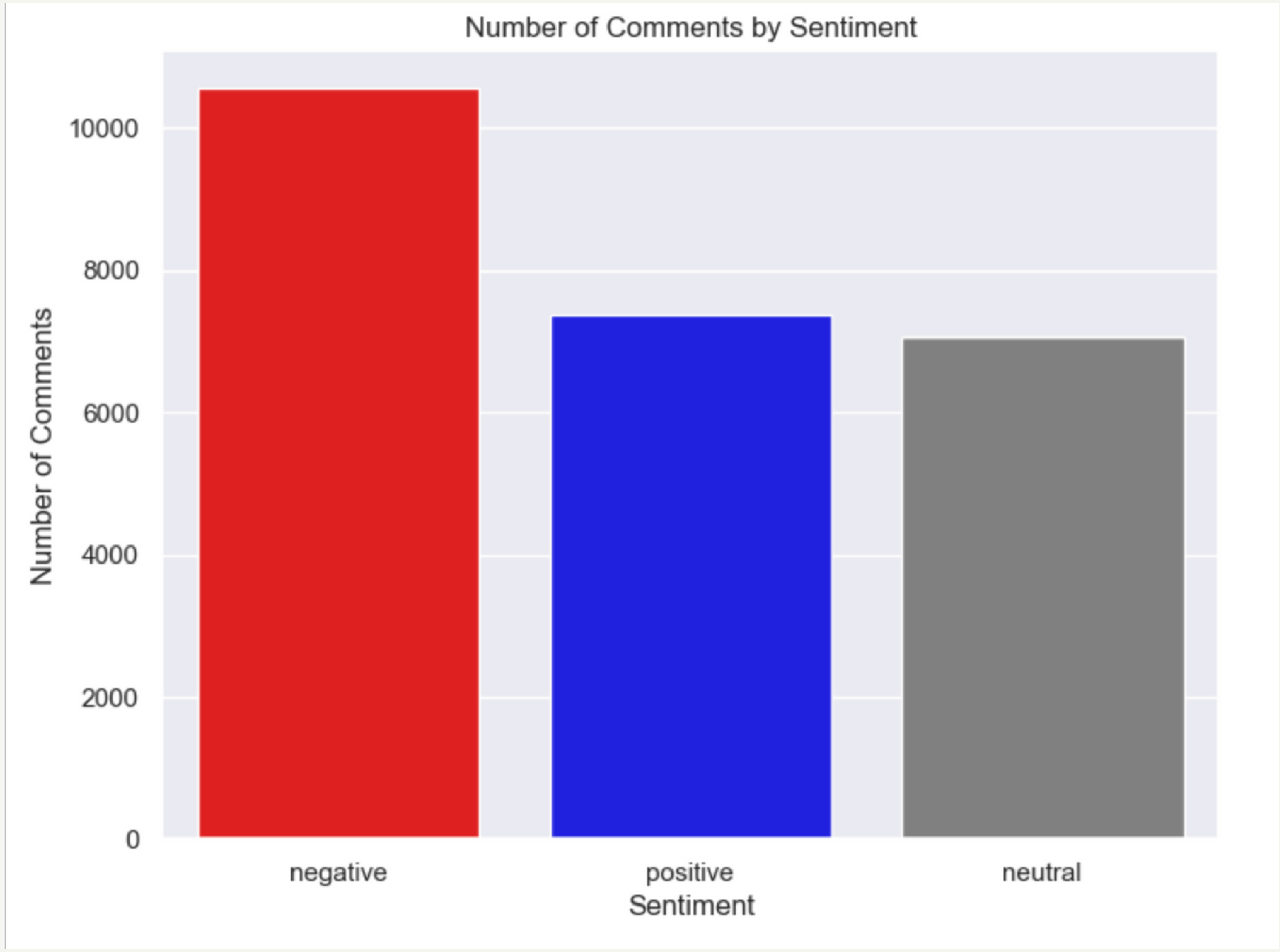
Variables in the data:

1. Datetime
2. Text
3. Tweet id
4. Username
5. Sentiment
6. Sentiment_score
7. Emotion
8. Emotion_score

2

Exploratory Data Analysis

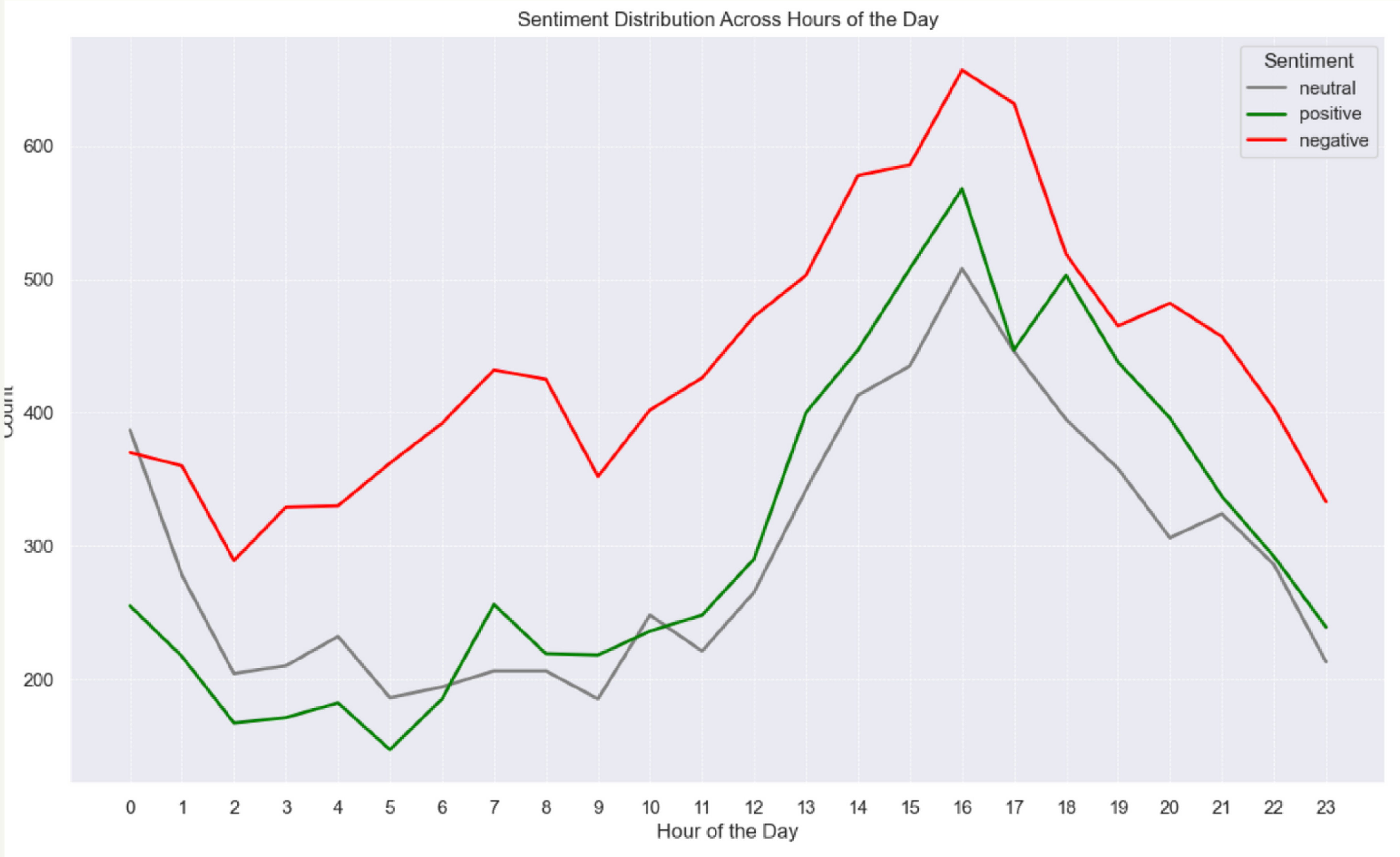
Bar plot: Differentiate the frequency of each sentiment



Counts of sentiment

1. Negative: 10556
2. Positive: 7366
3. Neutral: 7048

Line plot: display the frequency of each sentiment per respective hour

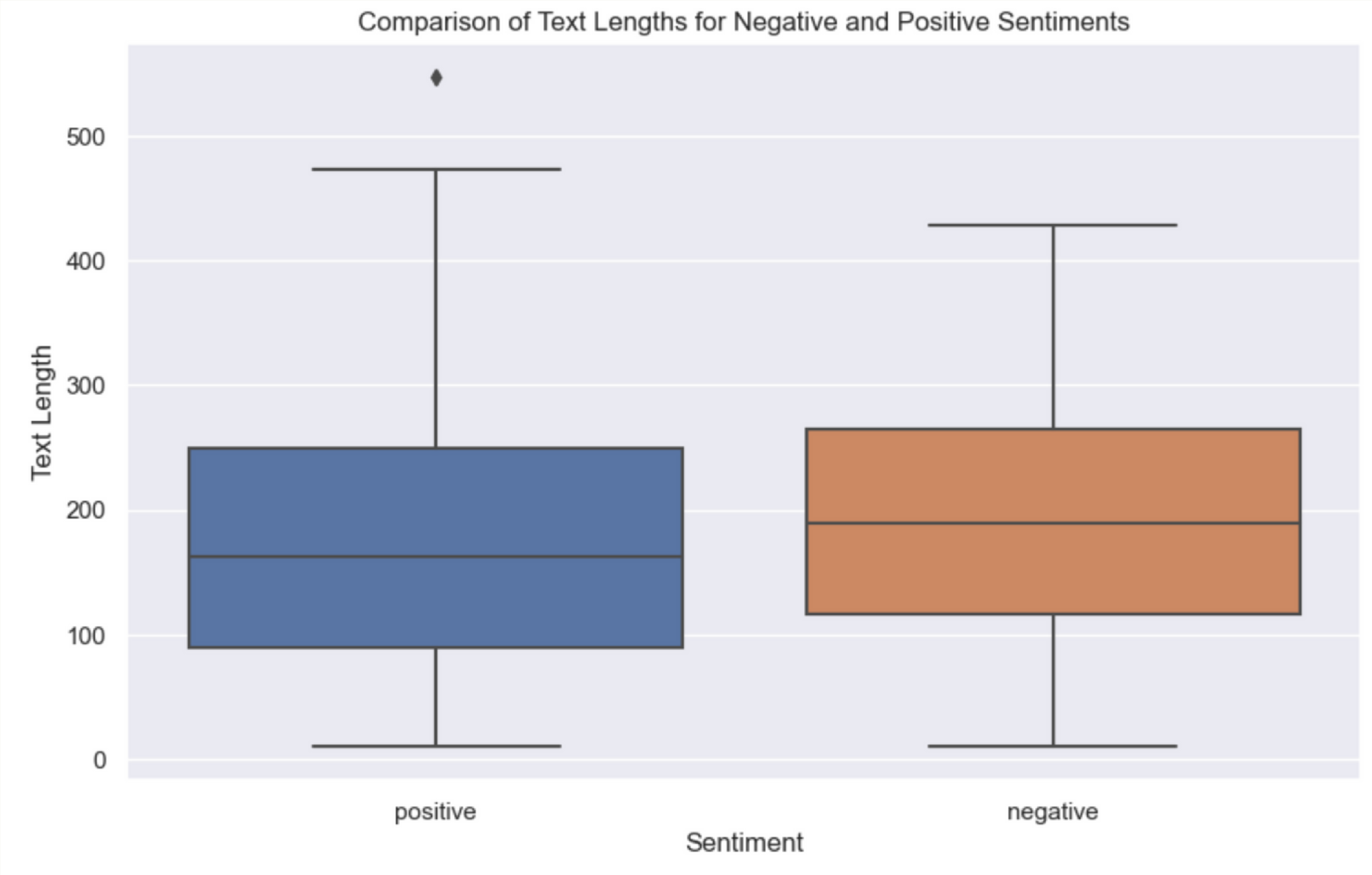


Basic Insight

On 16 Hour of the Day

Most number of negative, Most number of positive and Most number of neutral

Box and whisker: Compare the deviations on the length of the text between Positive and Negative



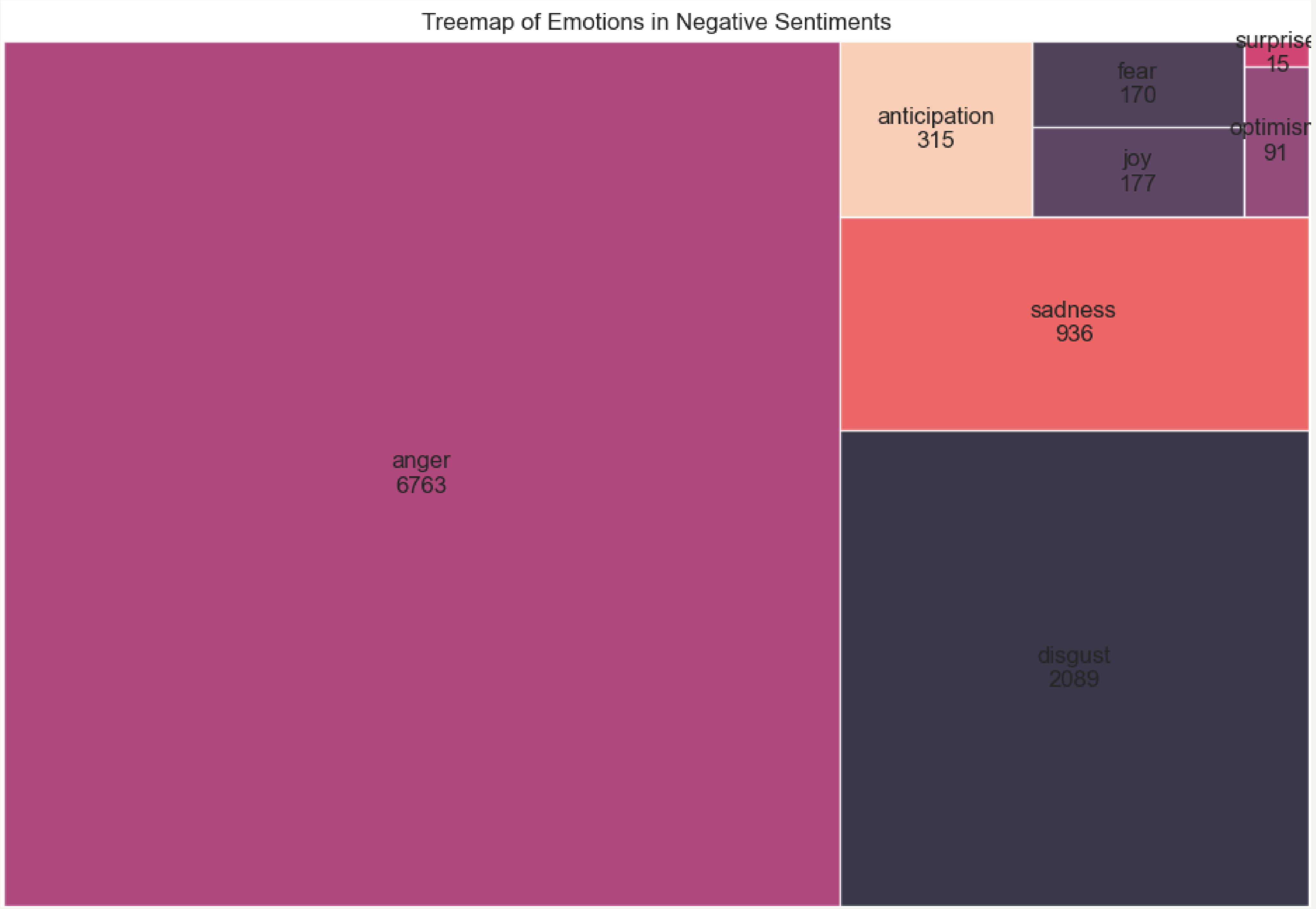
Median length of text

Length of the negative text is longer than postive text

Upper whisker

Positive text has longer length than negative

Treemap: Segmenting types of emotions and it's frequency



Ranking:

1. anger(6763)
2. disgust(2089)
3. sadness(936)

N-GRAM ANALYSIS

0	neutral	et streaming	70
1	neutral	sat amp	65
2	neutral	amp sun	64
3	neutral	every sat	60
4	neutral	tv every	58
5	positive	dell technologies	104
6	positive	join us	71
7	positive	looking forward	61
8	positive	xps plus	60
9	positive	dell xps	57
10	negative	customer service	532
11	negative	dell laptop	295
12	negative	new laptop	201
13	negative	buy dell	198
14	negative	customer care	194

Application:

Identify common phrases (bi grams) associated with each sentiment to understand context and sentiment drivers better.

Fundamental insight (2 words commonly used) :

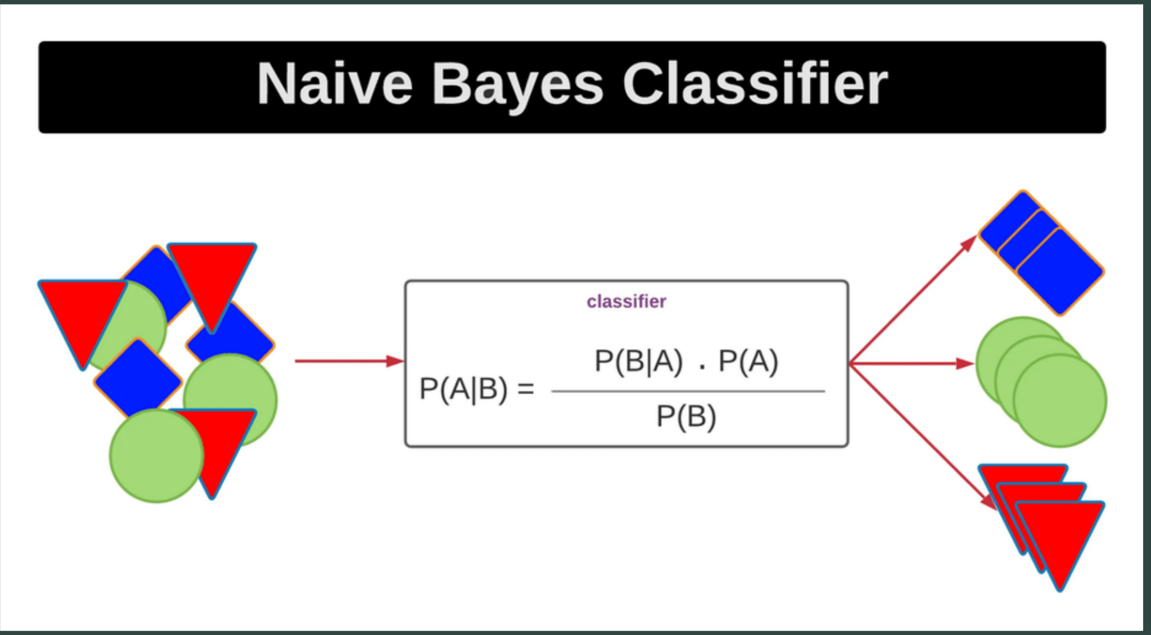
Positive: Dell Technologies

Negative: Customer service

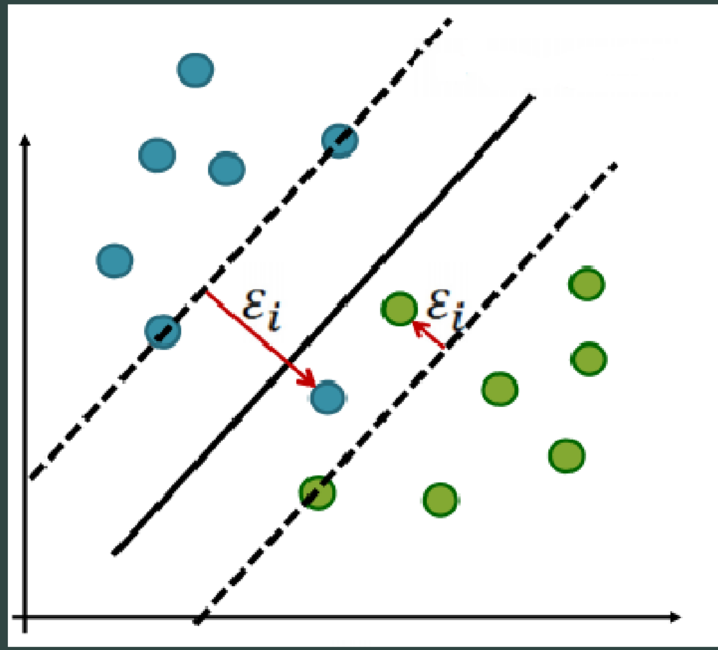
3

Predictive Models

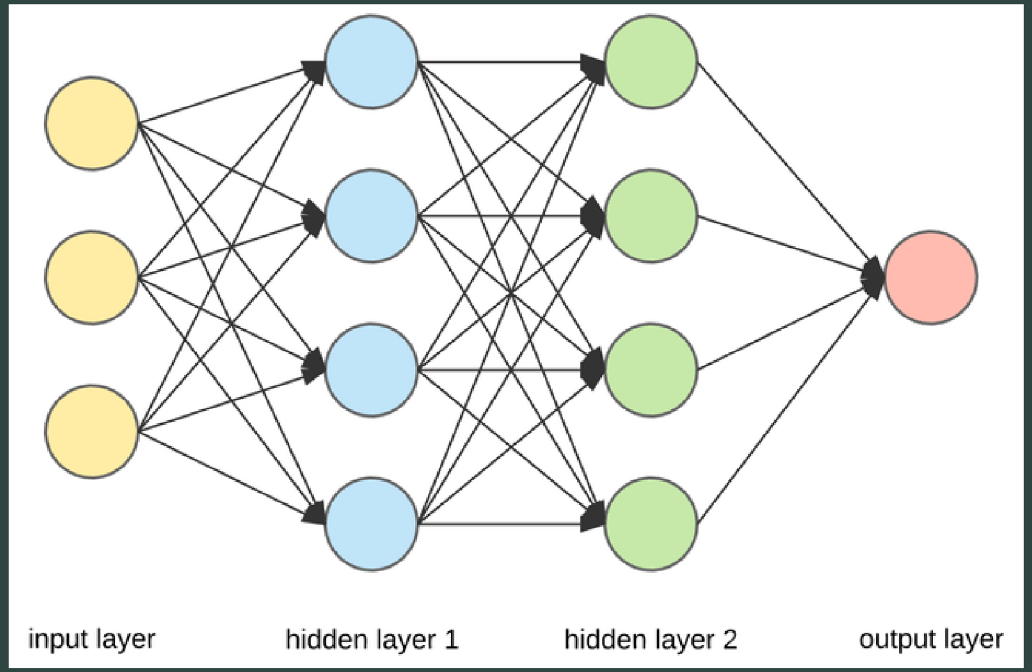
Multinomial Naive Bayes Model Approach



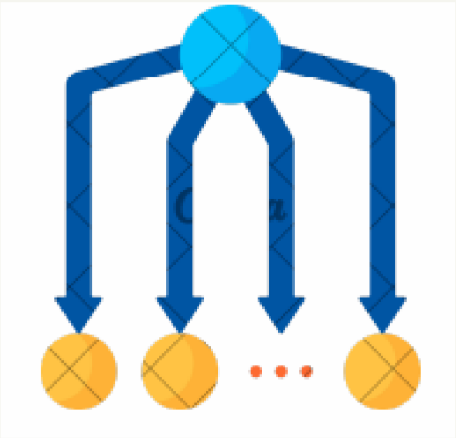
Linear SVC Classifier



Sequential Neural Network Model

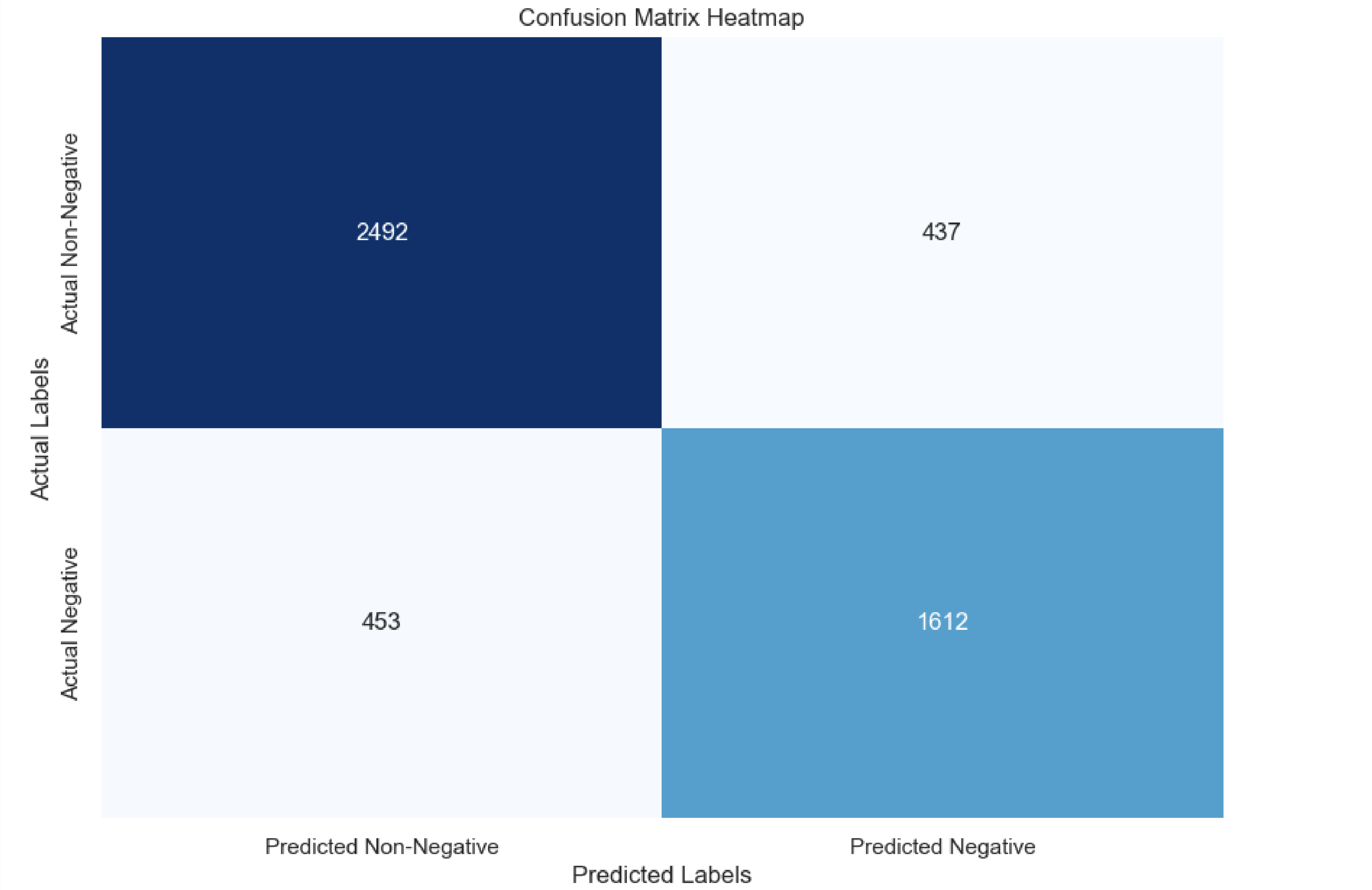


Multinomial Naive Bayes Model Approach



- It is a supervised machine learning algorithm, which is used for classification tasks, like comment classification.
- This model is particularly suited for classification with discrete features (e.g. word counts for comment classification)
- It works by calculating probability of a given text belonging to a particular sentiment class, based on the frequency of words in text , emotions and length of comment
- The algorithm is simple, efficient, and has been shown to perform well in sentiment analysis tasks.

Evaluation of Multinomial Naive Bayes Model Classifier

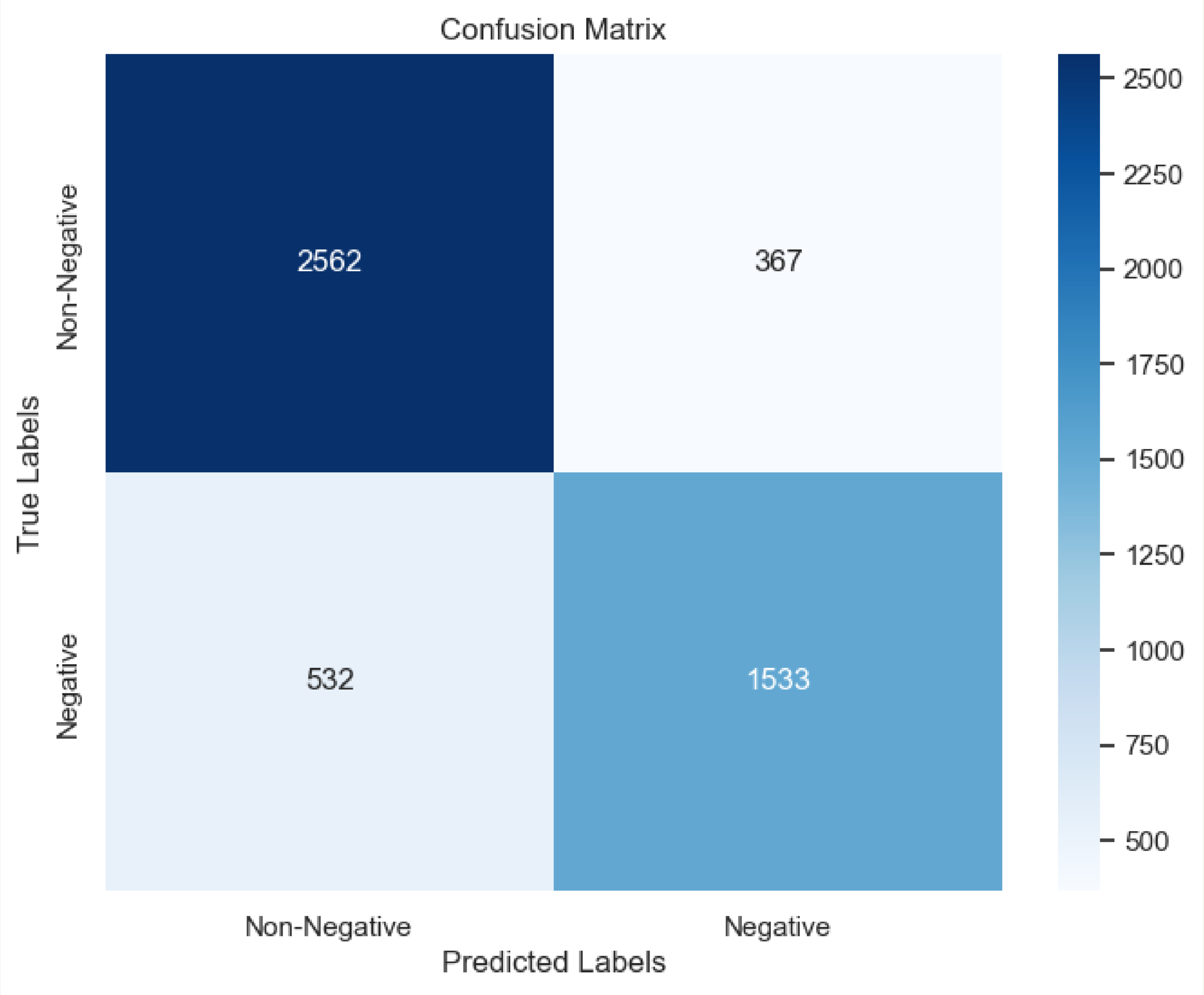


	precision	recall	f1-score	support
non-negative	0.85	0.85	0.85	2929.00
negative	0.79	0.78	0.78	2095.00
accuracy	0.82	0.82	0.82	0.82
macro avg	0.82	0.82	0.82	4994.00
weighted avg	0.82	0.82	0.82	4994.00

Linear SVC Classifier

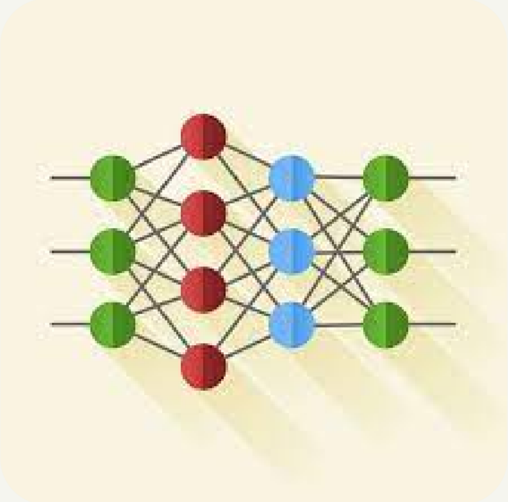
- Similar to Naive Bayes Classifier - used for classification tasks
- trained on **text** and **username** from our dataset
- Linear SVC works by mapping data points to a high-dimensional space and then finding the optimal hyperplane that divides the data into two classes.
- Goal of the model is to maximise the margin width between the 2 support vectors

Evaluation of Linear SVC Classifier



	precision	recall	f1-score	support
non-negative	0.83	0.87	0.85	2929.00
negative	0.81	0.74	0.77	2065.00
accuracy			0.82	4994.00
macro avg	0.82	0.81	0.81	4994.00
weighted avg	0.82	0.82	0.82	4994.00

Sequential Neural Network Model using Keras



```
model = Sequential()
model.add(Embedding(10000, 8))  # Embedding layer for text input, input_length inferred
model.add(Flatten())  # Flatten the 3D tensor to 2D
model.add(Dense(2, activation='softmax'))  # Output layer for binary classification
```

Sequential Neural Network Model using Keras

Embedding Layer

Embedding layer is the first layer, designed to handle text input. It turns positive integers (indexes) into dense vectors of fixed size (here, 8-dimensional). The model expects input sequences of a fixed size (maxlen=100), and the embedding layer is set to the size of the vocabulary (10,000 words).

Flatten Layer

The Flatten layer is used to flatten the 3D output of the embedding layer to 2D, making it possible to add a dense layer afterward.

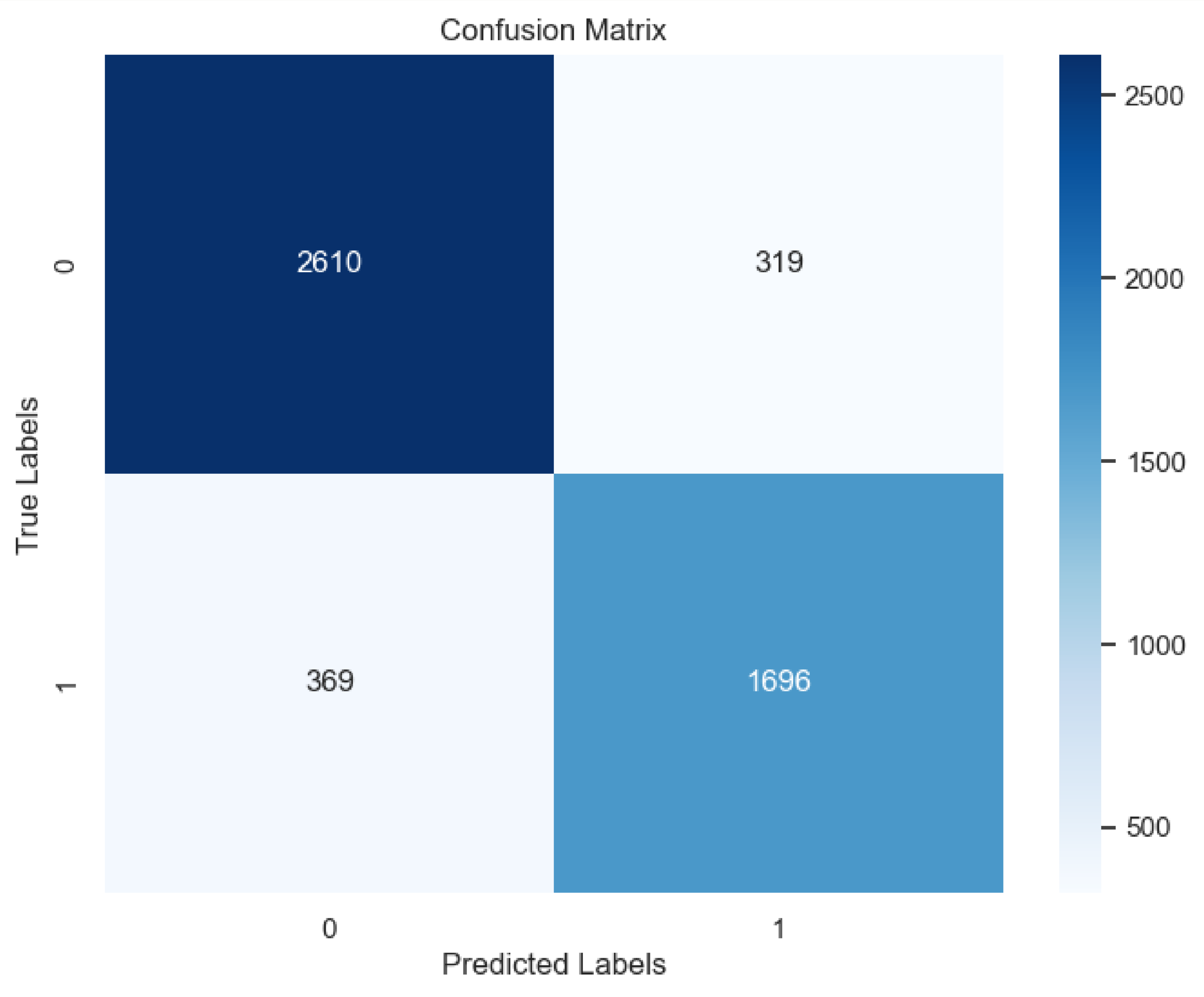
Dense Layer

The Dense layer with a softmax activation function serves as the output layer. It outputs two probabilities, corresponding to the two classes (positive and negative sentiment)

Early Stopping

Implement early stopping to automatically stop training when the validation loss stops improving. This prevents overfitting and reduces the need to manually find the optimal number of epochs. Patient = 3.

Evaluation of Sequential Neural Network



	precision	recall	f1-score	support
non-negative	0.88	0.89	0.88	2929.00
negative	0.84	0.82	0.83	2065.00
accuracy	0.86	0.86	0.86	0.86
macro avg	0.86	0.86	0.86	4994.00
weighted avg	0.86	0.86	0.86	4994.00

4

Conclusion

Conclusion

1.

Our Problem

Finding a dataset coming up with a problem statement

2.

Exploratory Data Analysis

Finding trends in the text data using visualization

3.

Data Preprocessing

Cleaning and preparing the text data using new tools

4.

Machine Learning Model Training

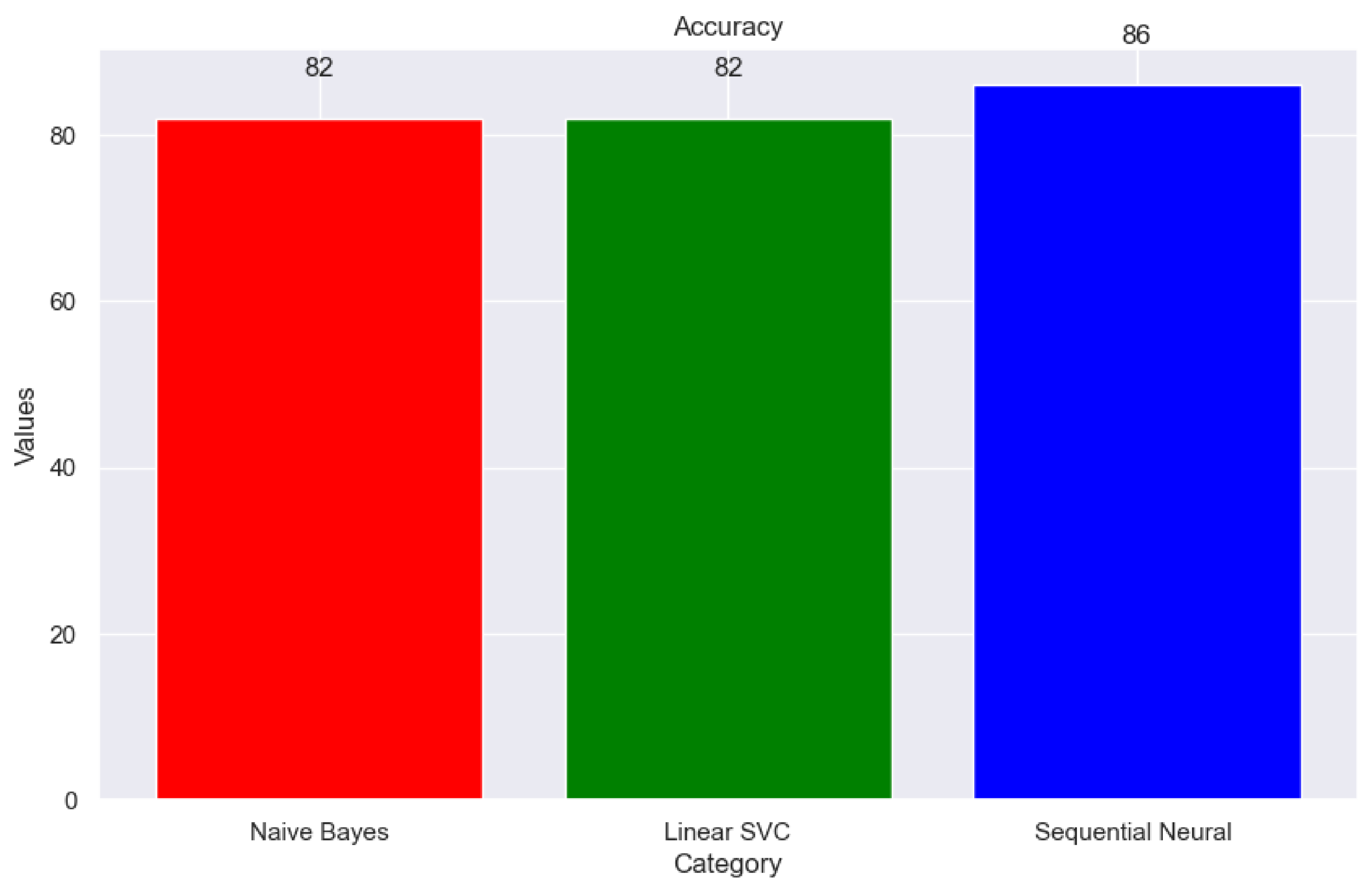
Researching and preparing classification models

5.

Model Evaluation

Interpreting the results of he model predictions and scoring them

Comparison of Models for Accuracy regarding Negative Sentiments



Sequential Neural Network Model predicted negative sentiments with 86% accuracy

Sentiment analysis is a powerful tool used to analyze text data and determine the underlying sentiment or emotion expressed within the text. When it comes to negative comments, sentiment analysis can be particularly useful in understanding the tone, context, and implications of those comments. **The use of AI can be replicated to other social media platforms** such as Instagram, Facebook and Tiktok which are popular among youths, hence creating a stepping stone for a vast range of applications such as cyberbullying detection and social media monitoring

Thank You!