A PROJECT IN
BLOCKCHAIN AND CRYPTOCURRENCIES

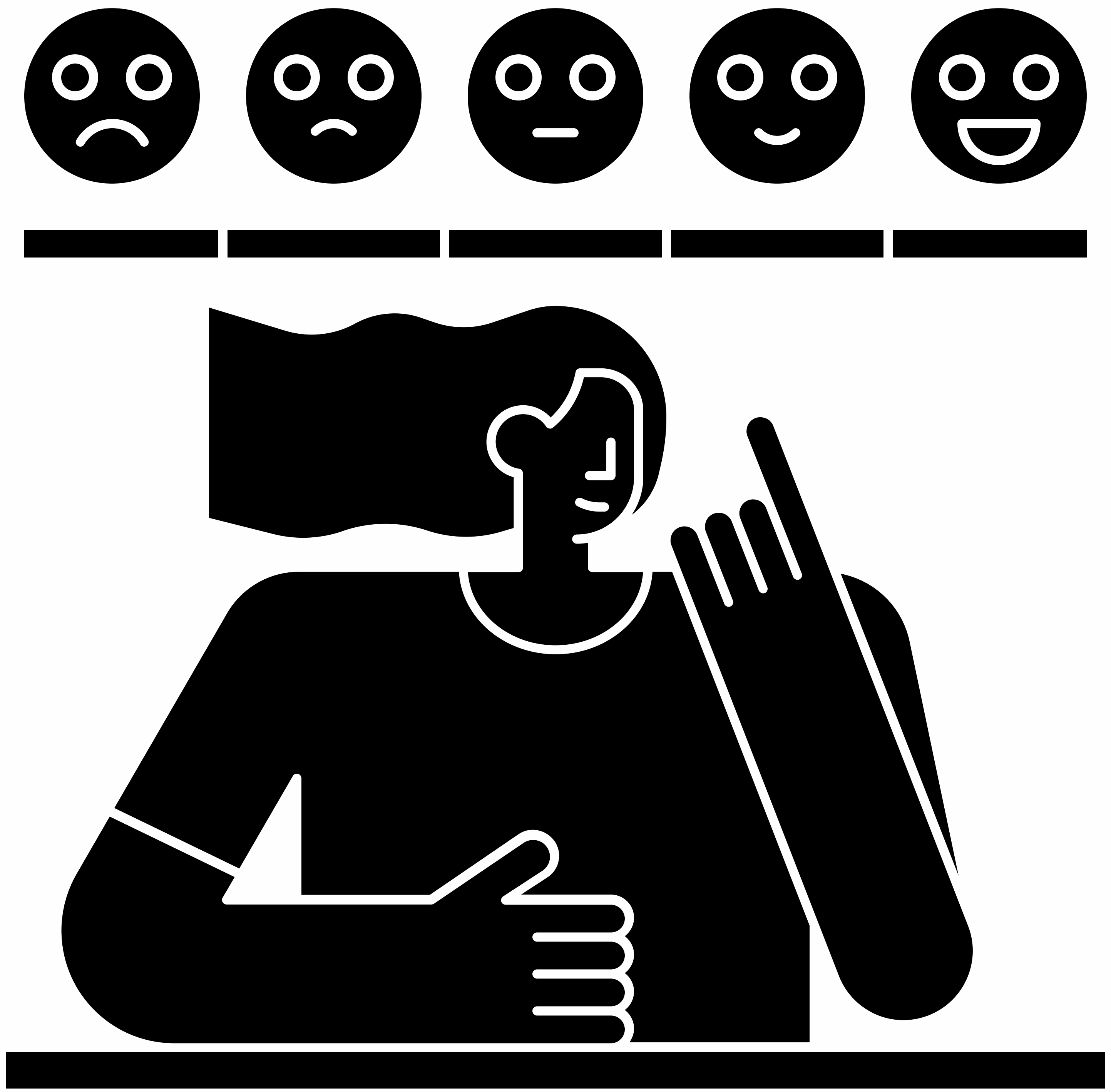
SENTIMENT ANALYSIS USING CRYPTOCURRENCY TWEETS.

Sandeep Kumar Kushwaha
sandeep.kushwaha@studio.unibo.it

INTRODUCTION

KEY TOPICS DISCUSSED IN THIS PRESENTATION

1. Introduction to sentiment analysis and its applications
2. Introduction to cryptocurrency and its role in social media
3. Collecting and preprocessing cryptocurrency tweets
4. Feature extraction and selection for sentiment analysis
5. Model selection and training for sentiment analysis
6. Evaluation and analysis of results
7. Conclusion and future work



SENTIMENT ANALYSIS

OPINION MINING

A field of natural language processing that involves using computational techniques to identify and extract subjective information from text.

Some common applications of sentiment analysis include:

1. Customer service
2. Market research
3. Social media monitoring
4. Political analysis
5. Brand management

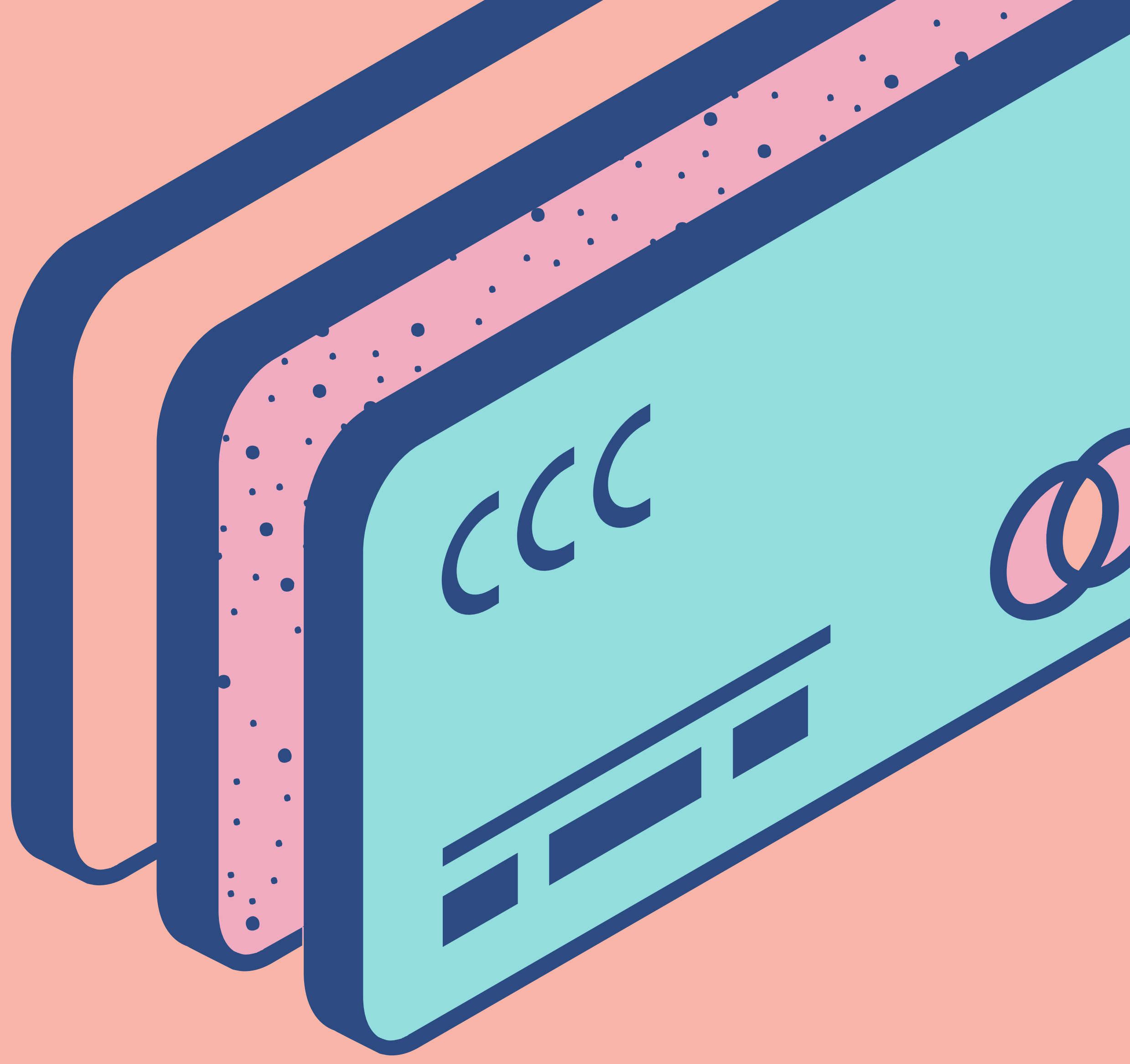
CRYPTOCURRENCIES

A DIGITAL/VIRTUAL CURRENCY THAT USES CRYPTOGRAPHY FOR SECURITY AND IS NOT BACKED BY A CENTRAL AUTHORITY SUCH AS A GOVERNMENT OR BANK.

Cryptocurrencies are frequently discussed on social media, where users trade, share news and opinions about the market and discuss different cryptocurrencies. **This has made social media an important channel for the dissemination of information about cryptocurrencies and for the formation of public opinion about them.**

DATASET PREPARATION

TWEETS FROM API & FEATURE EXTRACTION



1. SETTING UP TWITTER DEV ACCOUNT

Obtain a set of API credentials (a consumer key and consumer secret) to access the tweets.

2. EXTRACTING TWEETS USING API

Access a variety of data and functionality, including the ability to search for and retrieve tweets.

1. FINDING EXISTING DATASET

Since Twitter API has limits on extracting a number of tweets, I referred to an existing dataset.

2. OBSERVATIONS AND FINDINGS

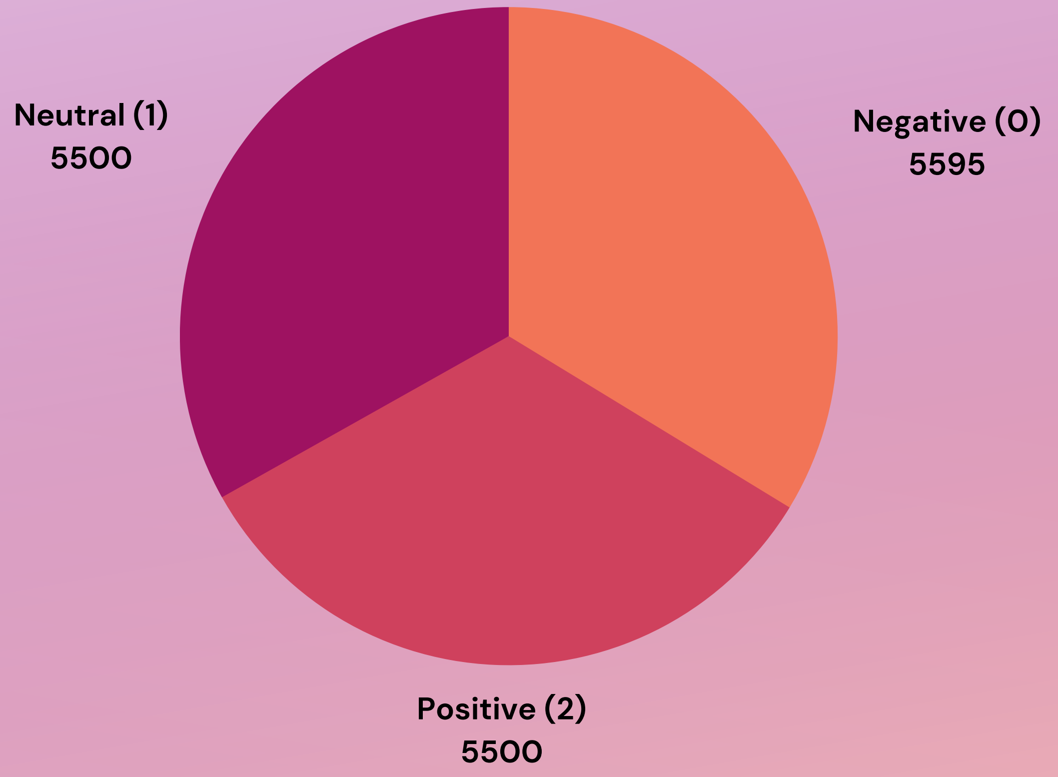
We check the tweets' quality and calculate the portion of positive, negative, and neutral scores.

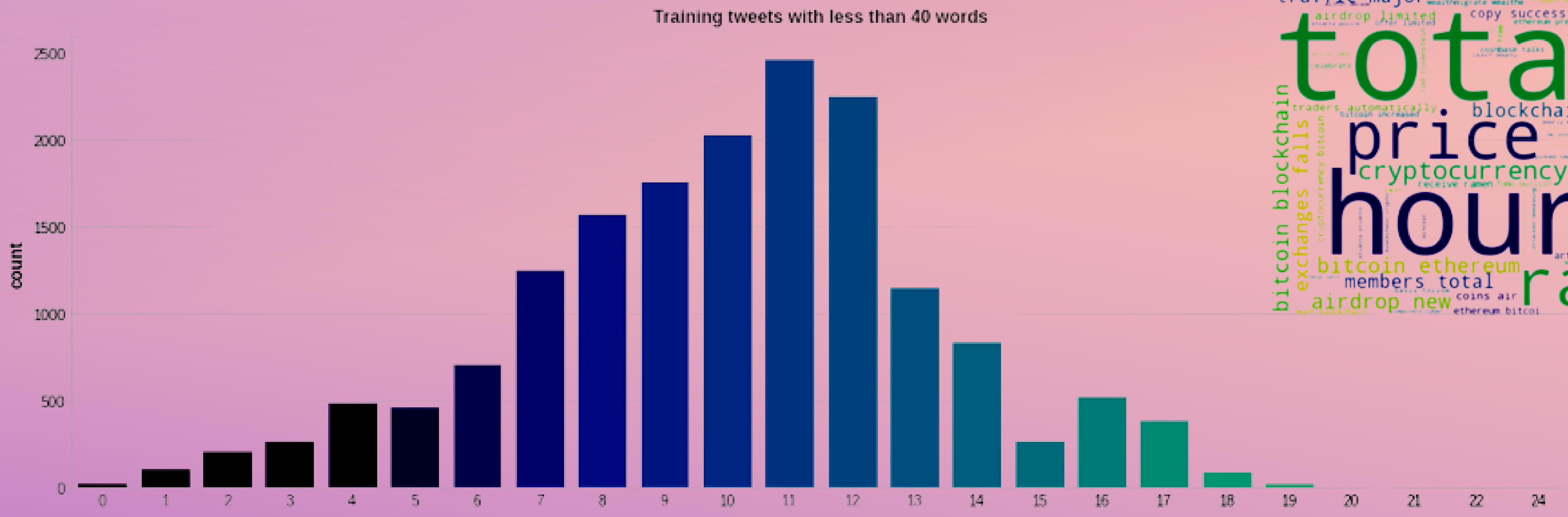
3. CLEANING AND PREPROCESSING

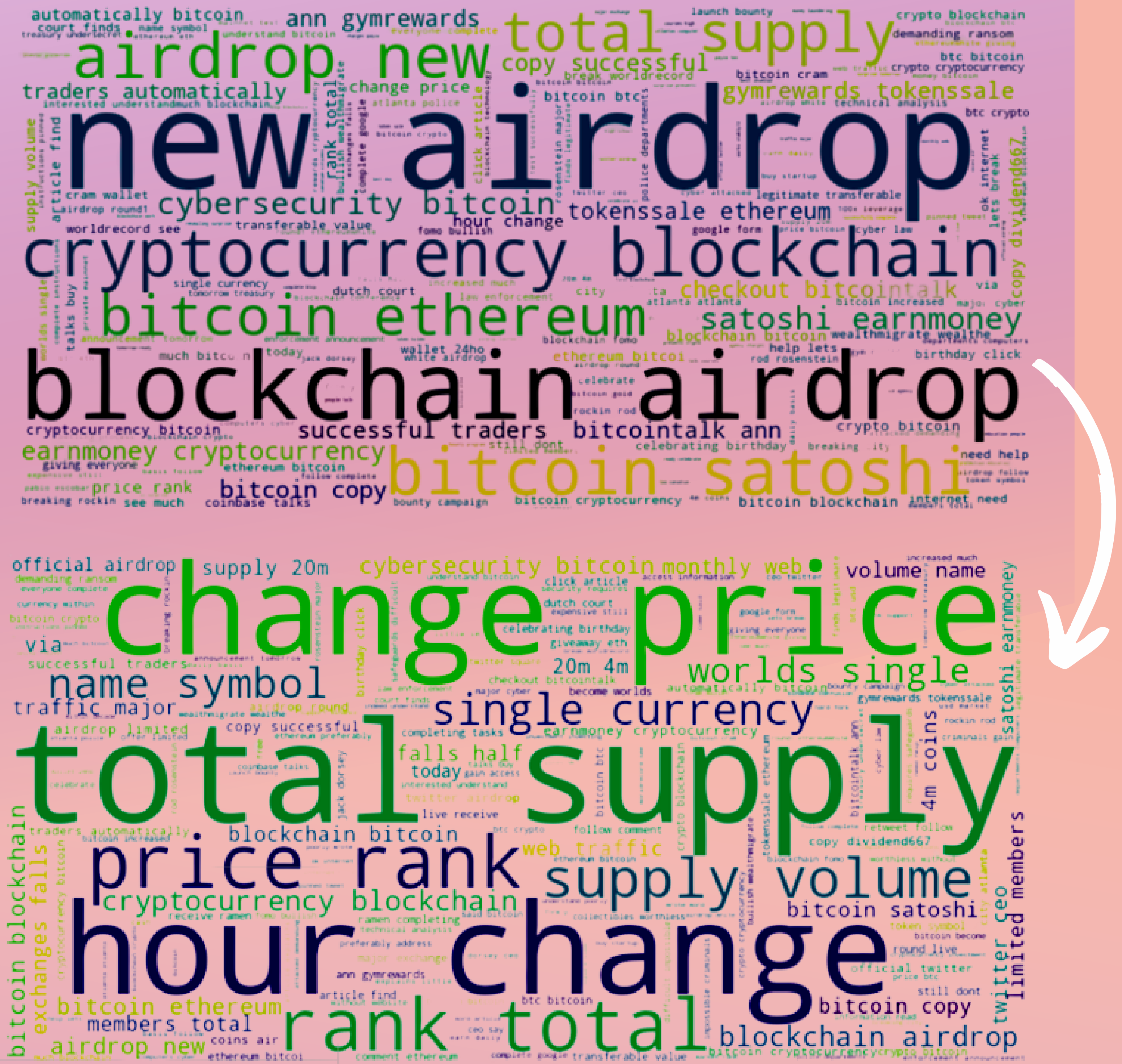
Noise and Null values are removed and only relevant columns are considered to feed into the model.

Dataset Overview

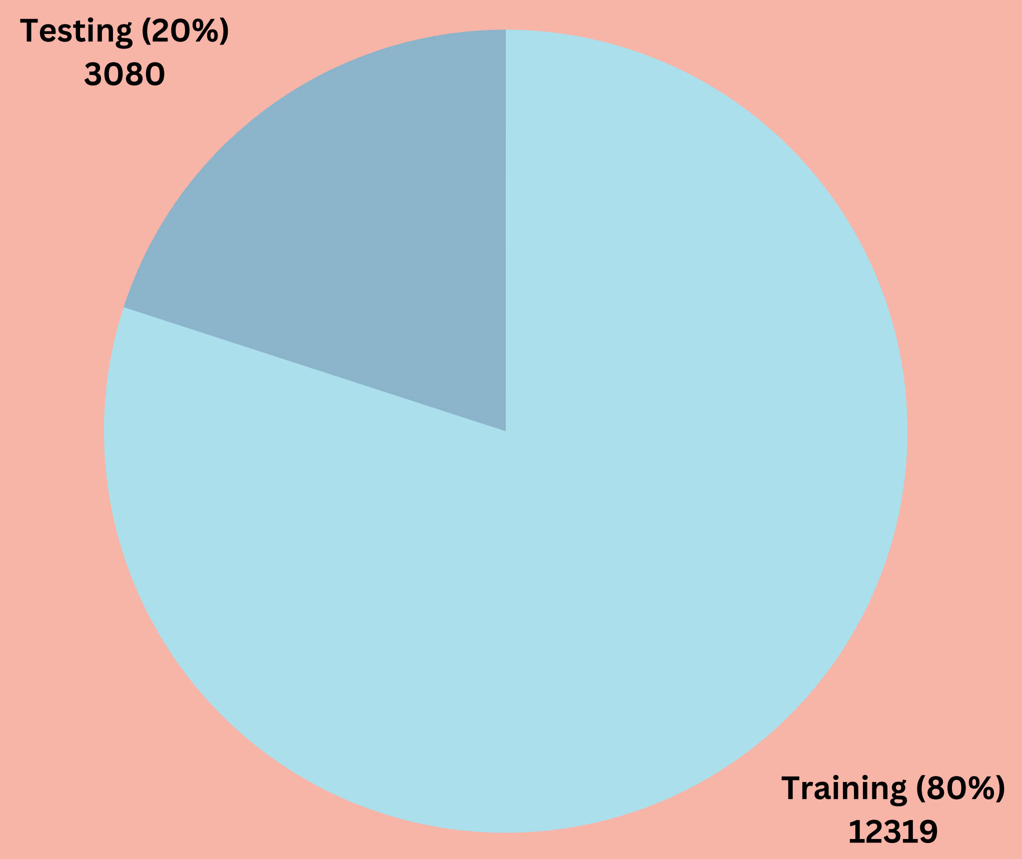
The dataset has well-balanced sentiments for training. Irrelevant words are removed from each tweet. To retain relevant information, I have considered tweets greater than 5 words and less than 40 words.







Model Selection and Training



There are multiple models to choose for Sentiment Analysis:

- Naive Bayes
- BERT (Bidirectional Encoder Representations from Transformers)
- RoBERTa (Robustly Optimized BERT Pre-training Approach)

Training involves

1. Splitting the data into a training set and a test set,
2. then fitting the model to the training set using the fit() method.
3. the model can then be evaluated on the test set using a variety of metrics, to assess its performance:
   - precision
   - Recall
   - F1 score
   - Support

Evaluation and analysis of results

PERFORMANCE OF THE MODEL ON A TEST DATASET

	precision	recall	f1-score	support
Negative	0.90	0.87	0.89	1137
Neutral	0.85	0.81	0.83	896
Positive	0.82	0.89	0.85	1047
accuracy			0.86	3080
macro avg	0.86	0.86	0.86	3080
weighted avg	0.86	0.86	0.86	3080

Classification Report for Naive Bayes

	precision	recall	f1-score	support
Negative	0.93	0.94	0.93	1137
Neutral	0.92	0.92	0.92	896
Positive	0.95	0.93	0.94	1047
micro avg	0.93	0.93	0.93	3080
macro avg	0.93	0.93	0.93	3080
weighted avg	0.93	0.93	0.93	3080
samples avg	0.93	0.93	0.93	3080

Classification Report for BERT

	precision	recall	f1-score	support
Negative	0.94	0.93	0.93	1137
Neutral	0.89	0.93	0.91	896
Positive	0.95	0.92	0.94	1047
micro avg	0.93	0.93	0.93	3080
macro avg	0.93	0.93	0.93	3080
weighted avg	0.93	0.93	0.93	3080
samples avg	0.93	0.93	0.93	3080

Classification Report for RoBERTa

Thank you.