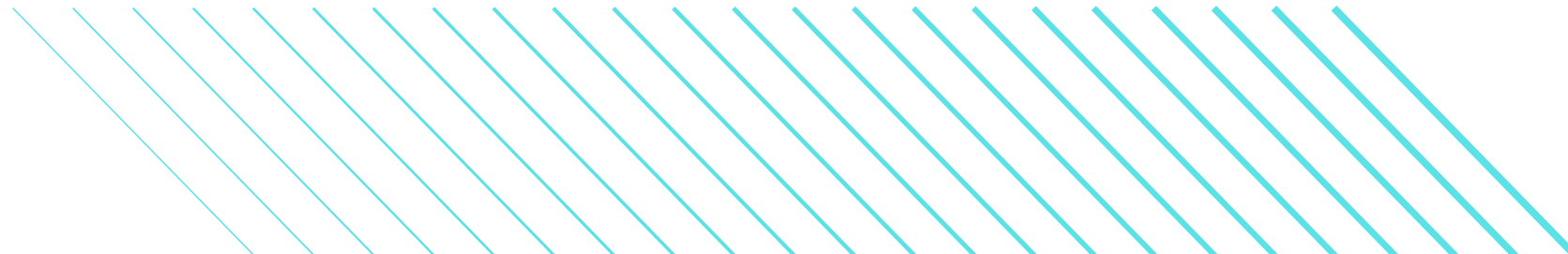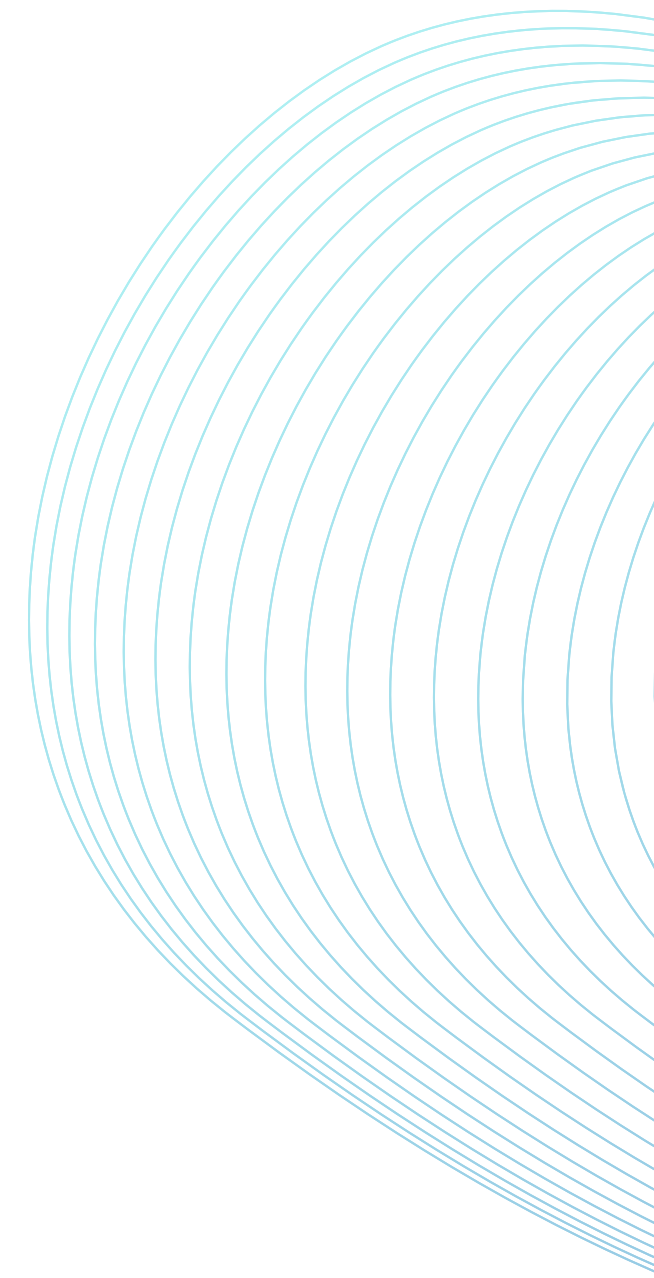# SENTIMENTAL ANALYSIS

THE PROCESS

1 **Data Preprocessing**

Creating a dataframe that contains only the text and the airline sentiment,lowcasing,removing numbers,lemmatization

2 **Tokenization**

Tokenize the tweets into individual words or tokens. This is the process of breaking down text into smaller units, typically words or phrases. Tokenization makes it easier to analyze and process text data

3 **Model Implementaion**

creating a BERT tokenizer using the pre-trained "bert-base-uncased" model

TRANSFORMERS & BERT

Transformers are composed of two main components: the encoder and the decoder. In the context of BERT and most other NLP tasks, only the encoder is used.

Pretrained language models sentiment analysis models are built on top of pre-trained transformer-based language models, such as BERT (Bidirectional Encoder Representations from Transformers)

Transfer learning
Pre-trained models are fine-tuned on specific sentiment analysis datasets, allowing them to adapt to the specific sentiment classification task.
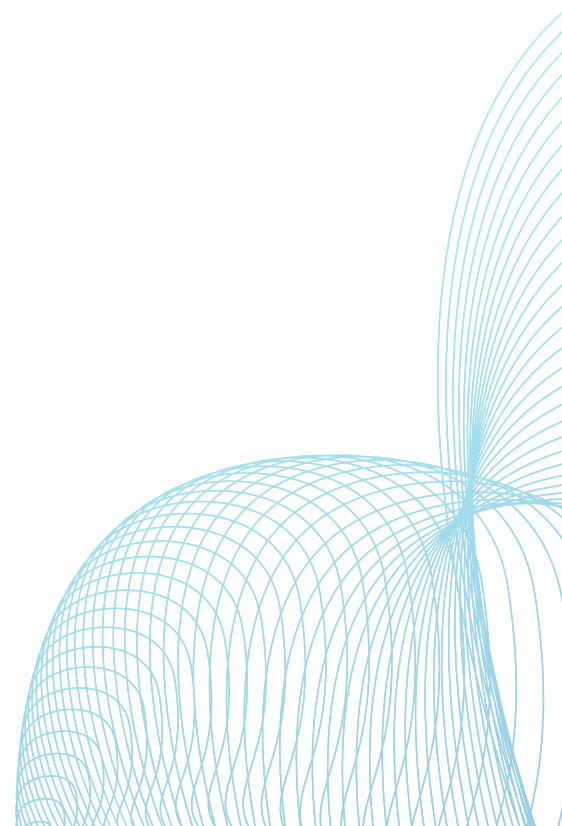
EVALUATION

The model is used to make predictions on the validation data, and the accuracy is calculated using **accuracy_score**. The best model's state is saved if the current validation accuracy is higher than the previous best accuracy.

- The model with the best validation performance is loaded, and evaluation is performed on the test dataset.
- Test predictions are generated and compared to the true labels. The accuracy on the test set is calculated using **accuracy_score**

The **classification_report** function is used, and class names are specified based on the sentiment classes ("negative," "neutral," "positive").

IMPLEMENTATION

```
'Airline has nice hospitality' is: Positive
'The experience was worst while travelling' is: Negative
'I thought it was good but it didn't satisfy me' is: Negative
'The overall experience was great' is: Positive
'This airline increased the price at the last moment' is: Neutral
'This airline increased the price to get profit' is: Negative
```