# Sentiment Analysis

## Introduction

Sentiment analysis is an area of NLP wherein we identify the emotional tone of the given text. Sentiment analysis can be applied to varying scopes such as document, paragraph, sentence and sub-sentence levels.

Sentiment analysis are normally domain dependent and some examples where it can be used are

1. Understand the customer behavior from the comments/reviews they write about a specific product.
2. Predict the connotation of the given text, whether positive/negative/neutral.

## Transformers

Recent advances in neural architectures, such as the Transformer, coupled with the emergence of large-scale pre-trained models such as BERT, have revolutionized the field of Natural Language Processing (NLP), pushing the state of the art for a number of NLP tasks.

## Transfer Learning

To train the transformers from scratch is a computationally expensive task. It is not recommended to train a transform from scratch but use domain-specific data to fine tune the pre-trained transformer models. This is also referred as transfer learning. Transfer learning reduces the time to train the model on domain specific data and can leverage smaller data sets for training. The drawbacks of transfer learning is that it carries the inherent bias present in the pretrained models.

## Data preparation

News headlines about different companies are collected from various sources and hand labelled as Buy, Sell, Neutral. The frequency distribution of the words from different labels represented as word clouds are shown below



Figure 1: Word cloud for text with Buy label

Graphical user interface, text

Description automatically generated

Figure 2: Word cloud for text with Sell label

Text

Description automatically generated

Figure 3: Word cloud for text with Neutral label

From the above word clouds we can see that specific key words can helps us in classifying the text into sell/buy/neutral classes.

## Training

Transformers library in python is used to train a transformer model for predicting the sentiment of the given text, buy/sell/neutral about the stock in the context. **distilbert-base-uncased-finetuned-sst-2-english** is used for transfer learning on the hand labelled data. Transformers library have integrations with MLflow and it is leveraged to track the training experiment runs.

The model is trained for 10 epochs, the below graph shows the loss during training.

Chart, line chart

Description automatically generated

Figure 4: Training loss

The below plot shows the F1-score during training on the training set

Chart, line chart, scatter chart

Description automatically generated

Figure 5: F1 score on the training data

The confusion matrix on the test set is shown below

Background pattern

Description automatically generated

Figure 6: Confusion matrix for the test data

Classification metrics report is below:

Table

Description automatically generated

Figure 7: Classification Report on the test data

## Inference

Results for sample cases were shown below  
Graphical user interface, text, application, email

Description automatically generated

We can see that model is able to label the given text based on the context, for example the model is predicting Neutral for the sentence “Apple is a fruit”.

A picture containing application

Description automatically generated

References:

* <https://doi.org/10.1186/s40537-019-0278-0>
* <http://jalammar.github.io/illustrated-transformer/>
* <https://huggingface.co/docs/transformers/index>
* <https://arxiv.org/abs/2105.00813>