Problem statement

In the era of social media, it is crucial for businesses to understand the sentiments expressed by customers towards their brands or products.
This sentiment analysis project aims to analyze Twitter data and extract valuable insights regarding the sentiments associated with Apple and Google products mentioned in tweets.
By uncovering the public's opinions and emotions, businesses can make data-driven decisions to improve their market positioning and enhance customer satisfaction

Objectives

Our main objective is to create a model that when given a tweet or series of tweets and a product would determine how the user felt about that product. This is trivial for a human to accomplish, our model can do this for thousands or even millions of tweets in a short time.

1. To build a text classifier to accurately distinguish between positive, neutral, and negative sentiments.

2. Competitive Analysis – Compare the sentiment towards Apple and Google products to identify any significant differences in public perception.

3. Give insights as to where the company can increase customer satisfaction.

Success metrics

We will consider this project successful if we can build a model that achieves the following performance metrics on a validation dataset:
1. Accuracy
2. Precision and Recall
3. F1 Score

The outcome of the initial data understanding.

We will be using a dataset from data.world provided by CrowdFlower which has tweets about Apple and Google from the South by Southwest (SXSW) conference. The tweet labels were crowdsourced and reflect which emotion they convey and what product/service/company this emotion is directed at based on the content.

The dataset comprises of Twitter data, including tweet texts, mentions of brands or products, and the associated sentiments or emotions. It had 9093 rows and 3 columns.

Data Quality Check: A thorough assessment of data quality was conducted, focusing on identifying missing values, inconsistencies, and any necessary data preprocessing requirements. This step ensures the subsequent analysis is based on reliable and clean data.