10 Questions and answers

- Why is Twitter sentiment analysis important?
  Sentiment analysis can help businesses understand their audience on social channels, stay on top of what's being said about their brand – and their rivals – and uncover new trends in the market by carefully listening to the voice of the consumer on Twitter.

- How accurate is Sentiment Analysis?
  The accuracy of sentiment analysis depends on the method used, the data quality, and the subjectivity of the sentiment. It is usually measured by precision and recall over the negative and positive categories. Human sentiment analysis has an average accuracy of around 85%, which is higher than most automated tools.

- What are the use cases of Sentimental Analysis models in the real world?
  Sentiment analysis has many use cases in the real world. Here are some examples:
  Product Design and Improvement
  Call Center Sentiment Analysis
  Measuring customer satisfaction.
  Monitoring brand reputation.
  Sentiment analysis can also be applied to social media, customer service, and market research.

- What are the steps were performed for data cleaning process for sentiment analysis of Twitter data?

The below transformations will be done on the dataset to prepare the data for modeling tasks.

- Remove all URLs, hash tags, usernames.
- Remove Emoticons.
- Remove all punctuation, symbols, numbers.
- Remove Stop Words
- make all text lower case.
- Handle NaN values
- Handle duplicate tweets.

- What are some common challenges in Twitter sentiment Analysis?

  There are several challenges of Sentiment Analysis. Here are some examples:

  1. Tone problem: Tone can be difficult to interpret verbally, and even more difficult to figure out in the written word.
  2. Polarity problem: Words such as “love” and “hate” are high on positive (+1) and negative (-1) scores in polarity.
  3. Sarcasm problem: Even the ironic, sarcastic, comparing comments detection is really hard.
  4. Emoji problem: If the data is in the form of emoji, then you need to detect whether it is good or bad.
  5. Idioms problem: Idioms can be difficult to interpret.
  6. Negations problem: Negations can change the meaning of a sentence.
  7. Comparative sentences problem: Comparing a neutral statement is a big task.
  8. Employee bias problem: Employee bias can affect the accuracy of sentiment analysis.

- Can we know what the general theme/mood of people is at a certain period from Sentiment Analysis?
  Yes, Sentiment Analysis can help us know the general theme/mood of people at a certain period. It can be used to analyze social media posts, news articles, and other forms of text data to determine the overall sentiment of people on a particular topic or event. This can be useful for businesses and organizations to understand how their brand is perceived by the public, as well as for researchers and policymakers to gauge public opinion on various issues.

- Which algorithm is used in Twitter sentiment analysis?
  There are several algorithms used in Twitter sentiment analysis. Some of them are:

  Logistic Regression
  Random Forest
  Multinomial NB
  Decision Tree

- What is a Word Cloud?
  A word cloud is a visual representation of text data. It is a collection or cluster of words depicted in different sizes. The bigger and bolder the word appears, the more often it's mentioned within a given text and the more important it is. Word clouds are used to highlight essential textual data points and can make dull data shine and deliver crucial information quickly.

- What is the difference between bag of words and TF-IDF?

Bag of Words (BoW) and TF-IDF are two methods used in natural language processing. BoW simply counts the frequency of words in a document, while TF-IDF contains information on the more important words and the less important ones as well. The key difference between bag of words and TF-IDF is that the former does not incorporate any sort of inverse document frequency (IDF) and is only a frequency count (TF). Bag of Words vectors are easy to interpret, but TF-IDF usually performs better in machine learning models.

- What assumptions were made in the analysis of this data?
  The dataset contained null values and duplicate records. The decision to remove such records from the dataset may have affected the modeling results. But I was unable to find a logical way to fill in the null values without it affecting the final predictions.
  Another assumption is that the target values in the datasets correctly represent the sentiment of the tweet. Usage of some words can mean differently in different contexts.