**Phase 2: INNOVATION**

The problem statement is sentiment analysis for marketing. The goal is analysed the dataset and produced the result as follows; Positive, Negative, Neural.

**PROBLEM DEFINITION:**

The objective of this project is to perform sentiment analysis on customer feedback related to competitor products in order to extract valuable insights that can inform business decisions. By understanding customer sentiments towards competing products, our aim is to identify both strengths and weaknesses, enabling us to improve our own offerings and gain a competitive advantage in the market.

**DESIGN THINKING:**

1.Data Collection

2. Data Preprocessing

3. Sentiment Analysis Techniques

4. Feature Extraction

5. Visualization

6. Insights Generation

**Data collection:** Gather a labeled dataset of Tweets. Where each tweet is labeled as either positive or negative or neural.

**Data Preprocessing:**

**Clean the text data:** Remove any duplicates, special characters, URLs, and mentions.

**Lowercase text:** Converted from uppercase alphabet letters to lowercase letters.

**Tokenization:** Split the text into individual words(tokens)

**Stop words:** Remove the stop words in the dataset like is, was,the etc.

**Join tokens:** Join the tokens back into the words.

**Feature Extraction:** Feature extraction involves deriving meaningful features or representations from text data that can be used for sentiment analysis. Convert the textual data into numerical features that can be used by the Naive Bayes algorithm.

**Split the Data:** Split the dataset into a training set and a testing set. The training set will be used to train the model, while the testing set will be used to evaluate its performance.

Implement Naive Bayes Algorithm: Implement the Naive Bayes algorithm. Naive Bayes classifiers used for classify the result as Positive or Negative or Neural.

**Train the Model:** Train the Naive Bayes classifier using the training data. During this step, the algorithm calculates probabilities based on the occurrence of words in Positive or Negative or Neural.

**Evaluate the Model:** Use the testing set to evaluate the performance of the trained model. using accuracy, precision, recall, and F1-score.

**Hyper-parameter Tuning :** If necessary, perform hyper-parameter tuning to optimise the performance of the model. This might involve adjusting parameters such as the smoothing parameter (Laplace smoothing) in the Naive Bayes algorithm.

**Deployment :** If the model performs well, it can be deployed into a production environment where it can automatically classify incoming tweets as positive or negative or neural.

**Preprocessing:**

In this phase, we'll explore innovative techniques and approaches to building our sentiments classifier.

**Dataset link:** https://www.kaggle.com/datasets/crowdflower/twitter-airline-sentiment

**Tehnology Usage:**

**Data Preprocessing:** Using modules are nltk(Natural language toolkit),stopwords and re.

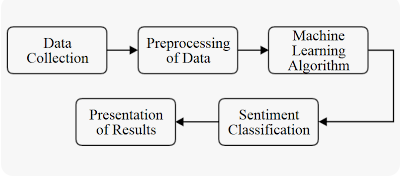
**Sentiment analysis techniques:** Using NLP(Natural language process) technique for sentiment analysis.

**Feature Extraction:** Using technique is TF-IDF(Ter Frequency- Inverse Document Frequency)

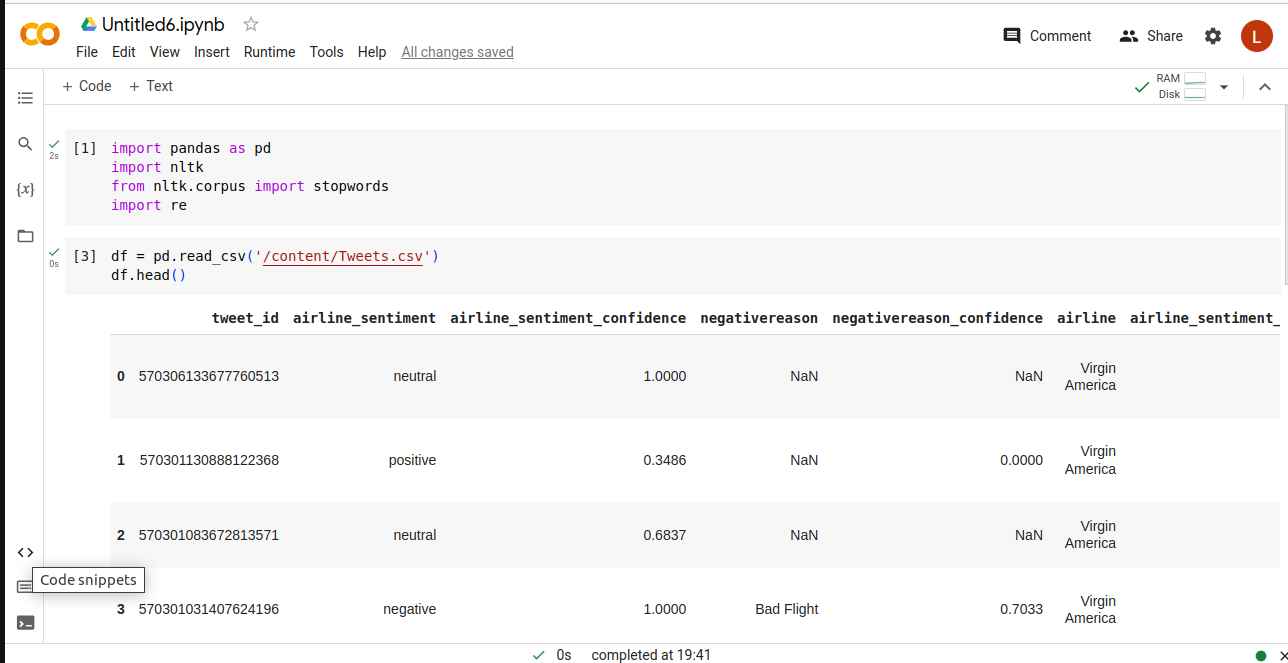
**Visualization:** Using Matplotlib and Seaborn modules.

**Development Environment:** Using google colab platform.

**Architecture diagram:**



**Preprocessing the dataset:**





**After preprocessed the dataset as follows:**

