**IBM NAAN MUDHALVAN AI PHASE 1**

**SENTIMENT ANALYSIS FOR MARKETING**

**PROBLEM DEFINITION:**

From the problem statement that is being given, I could understand that the main task is to with the help of dataset on customer feedback, I’ve to build an AI model that can work efficiently on analysing sentimental feelings of the customer which will help in gaining the insights of the competitor products thereby improving their own feelings. Thereby I’ve to use NLP methods to extract the valuable insights of the customer from the feedback that is being given. As prescribed in the problem statement, the primary objective is to identify the strengths and weakness of the competing products as perceived by customers, and to use this information to improve one’s own offerings.

**DESIGN THINKING:**

**Data Collection**:

* The dataset is given from the Kaggle link.
* https://www.kaggle.com/datasets/crowdflower/twitter-airline-sentiment

**Data Preprocessing:**

* Clean the data by removing noise, such as special characters, HTML tags, and irrelevant information.
* Tokenize the text into words or phrases.
* Perform stemming or lemmatization to reduce words to their base forms.
* Handle imbalanced classes if necessary.

**Exploratory Data Analysis (EDA):**

* Explore the data to gain an understanding of its distribution, common keywords, and patterns.
* Visualize the distribution of sentiments in the data.
* Convert the text data into numerical features that can be used as input for machine learning models. Common methods include TF-IDF (Term Frequency-Inverse Document Frequency) and word embeddings.

**Model Selection:**

* Choose a machine learning or deep learning model for sentiment analysis. Common choices include:
* **Logistic Regression:** A simple yet effective model for text classification tasks.
* **Naive Bayes:** Another simple and fast model often used for sentiment analysis.
* **Convolutional Neural Networks (CNN):** Effective for capturing local patterns in text data.
* **Recurrent Neural Networks (RNN):** Suitable for sequential data but less common for sentiment analysis due to vanishing gradient problems.
* **Transformer-Based Models:** State-of-the-art models like BERT, GPT, and RoBERTa have achieved excellent results in sentiment analysis tasks. You can fine-tune these models on your specific dataset.

**Model Training:**

* Split your dataset into training, validation, and test sets.
* Train the chosen model on the training data, optimizing for accuracy or other relevant metrics.
* Use the validation set to fine-tune hyperparameters and avoid overfitting.

**Evaluation:**

* Evaluate the model's performance on the test set using appropriate metrics (e.g., accuracy, precision, recall, F1-score, ROC-AUC).
* Adjust the model or preprocessing steps if necessary to improve performance.

**Deployment:**

* Once you have a trained model, you can deploy it to analyze customer feedback in real-time or batch processing, depending on your project's requirements.
* Monitoring and Maintenance:
* Continuously monitor the model's performance and retrain it periodically with new data to maintain accuracy and relevance.

**Interpretability:**

* Depending on your project's goals, you may want to analyze the model's predictions to understand which words or phrases contribute to the sentiment classification. Techniques like LIME (Local Interpretable Model-agnostic Explanations) can help with model interpretability.

**Visualization:**

* Create visualizations or dashboards to present the sentiment analysis results and insights to stakeholders.

**Insights Generation:**

Sentiment analysis results can guide business decisions by revealing overall sentiment trends, pinpointing product/service performance, detecting anomalies, and comparing sentiment across channels. Additionally, analyzing sentiment by demographics, keywords, and competitors, along with identifying reasons for negative sentiment, can inform marketing and operational strategies. Furthermore, businesses can leverage sentiment data to understand its impact on sales and revenue, foster product development, engage with advocates and detractors, refine social media campaigns, and manage crises effectively, ensuring data-driven decision-making for improved customer satisfaction and business success.

**Flow Chart:**

**TESTING THE MODEL WITH REALTIME DATA**

**IMPLEMENTING**

**ALGORITHMS**

**SPLITTING TRAIN AND TEST DATA**

**DATA PREPROCESSING**

**DATASET INPUT**