IBM PHASE -1
PROJECT SUBMISSION

Project title:Sentiment analysis

PROJECT DEFINITION:

- The Sentiment Analysis for marketing guidlines aims to develop a system that automatically analyzes and classifies customer feedback to determine the sentiment expressed (positive, negative, or neutral). This project is crucial for businesses to gain insights into customer satisfaction levels and identify areas for improvement

OBJECTIVES:

- Develop a robust sentiment analysis model capable of accurately classifying customer feedback.
- Provide a user-friendly interface for inputting and analyzing customer comments.
- Generate visual reports and dashboards to summarize sentiment trends over time.
- Enable integration with existing customer feedback systems or databases.

DESIGN THINKING:

1. Data Collection :

- Collecting data for a Twitter US Airline Sentiment Analysis project involves extracting tweets related to various US airlines along with their corresponding sentiments (positive, negative, or neutral).
- Collect data from Kaggle by the given link.
- https://www.kaggle.com/datasets/crowdflower/twitter-airline-sentiment

Data Explorer
Version 4 (8.46 MB)

Tweets.csv
database.sqlite

Tweets.csv (3.42 MB)

Detail   Compact   Column   10 of 15 columns

About this file
CSV table of tweets

↔ tweet_id	⚠ airline_sentiment	# airline_sentiment...	⚠ negativereason	# negativereason_c...	⚠ a
	negative 63%		[null] 37%		Unit
	neutral 21%		Customer Service ... 20%		US /
567588279b 570310600b	Other (2363) 16%	0.34 1	Other (6268) 43%	0 1	Oth
570306133677760513	neutral	1.0			Vir
570301130888122368	positive	0.3486		0.0	Vir
570301083672813571	neutral	0.6837			Vir

2. Data Preprocessing :

- Data preprocessing is a crucial step in preparing your dataset for analysis or machine learning models. It involves cleaning, transforming, and organizing raw data into a format that is suitable for your specific task. Here are some common data preprocessing steps:

1. Handling Missing Values:
    - Identify and handle missing data. This can include techniques like imputation (replacing missing values with a sensible estimate) or removal of rows or columns with missing values.

2. Data Cleaning:
    - Remove any irrelevant or redundant information from the dataset. This could include duplicates, outliers, or data that doesn't contribute to the analysis.

3. Data Transformation:
    - Normalize or standardize features to bring them to a similar scale. This is particularly important for models sensitive to the scale of input features.

4.Data Splitting:
    - Divide the dataset into training and testing sets to evaluate model performance. For example, an 80-20 or 70-30 split is common.

5.Scaling and Normalization:
    - Scale features to a similar range. Common techniques include min-max scaling and z-score normalization.

6.Data Integration (if applicable):
    - If you're working with multiple datasets, you may need to integrate them. This could involve combining them, resolving discrepancies, or linking them through common identifiers.

3. Sentiment Analysis Techniques:

- Applying a variety of NLP techniques allows for a comprehensive understanding of the sentiment.
- Choose appropriate NLP techniques for sentiment analysis. Suggested options:
- **Bag of Words (BoW):** A simple but effective technique that represents text as a bag of words, which can be used for sentiment classification.
- **Word Embeddings**: Utilize pre-trained word embeddings (e.g., Word2Vec, GloVe) to capture semantic relationships between words.
- **Transformer Models**: Consider state-of-the-art models like BERT or GPT for more advanced sentiment analysis.

4. Validation and Interpretation:

- Validate the model's performance using metrics like accuracy, precision, recall, and F1score. Additionally, explore misclassified samples to understand the model's limitations.

5. Feature Extraction:

- Convert the preprocessed text data into numerical format. Common techniques include TF-IDF (Term Frequency-Inverse Document Frequency) or word embeddings (Word2Vec, GloVe, etc.).

6. Model Selection:

- Choose a suitable sentiment analysis model. You can start with traditional machine learning models like Support Vector Machines (SVMs), Naive Bayes, or more advanced techniques like recurrent neural networks (RNNs) or transformer-based models like BERT.
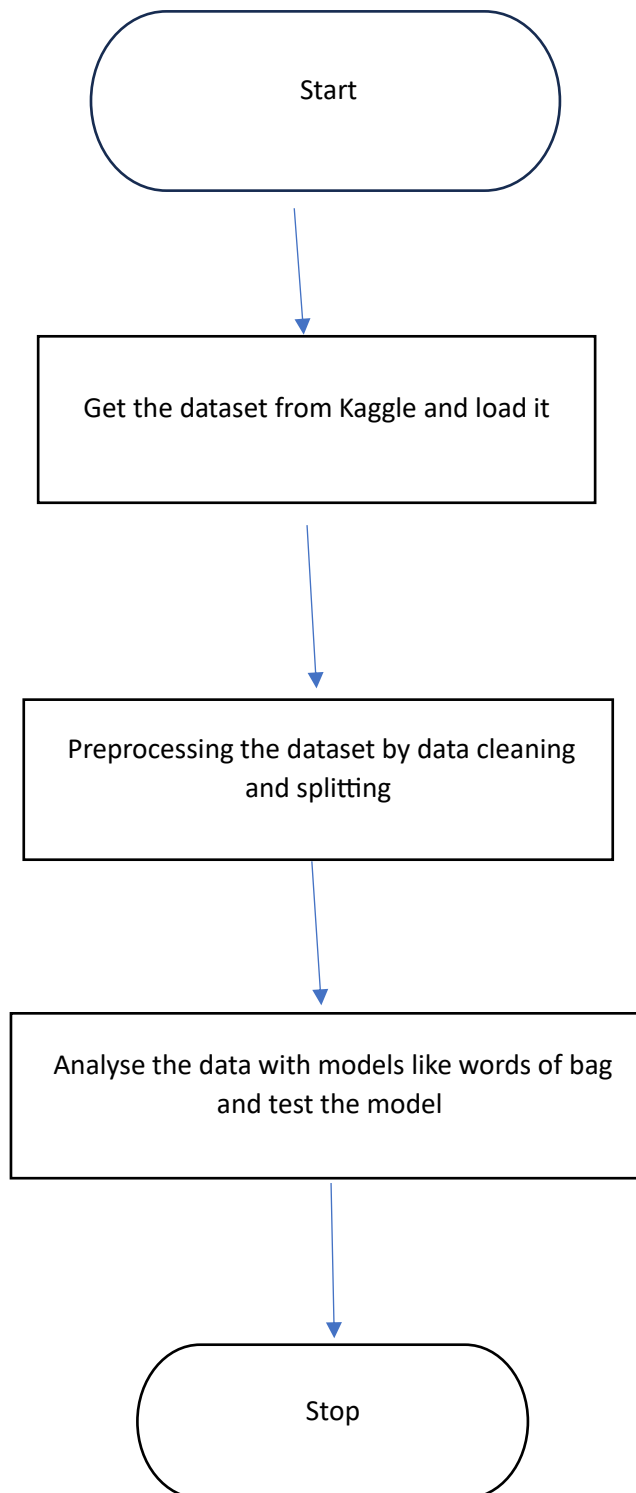
7. Model Training:

- Split your dataset into training and testing sets. Train the chosen model on the training set and evaluate its performance on the testing set.

8. Model Fine-tuning:

- Depending on the performance, fine-tune your model. This might involve hyperparameter tuning or experimenting with different architectures.

FLOWCHART:

```
                    ┌──────────────────────┐
                    │        Start         │
                    └──────────────────────┘
                               │
                               ▼
           ┌──────────────────────────────────────┐
           │   Get the dataset from Kaggle and     │
           │              load it                   │
           └──────────────────────────────────────┘
                               │
                               ▼
           ┌──────────────────────────────────────┐
           │  Preprocessing the dataset by data     │
           │        cleaning and splitting          │
           └──────────────────────────────────────┘
                               │
                               ▼
           ┌──────────────────────────────────────┐
           │  Analyse the data with models like     │
           │   words of bag and test the model      │
           └──────────────────────────────────────┘
                               │
                               ▼
                    ┌──────────────────────┐
                    │         Stop         │
                    └──────────────────────┘
```

CONCLUSION:

The Sentiment Analysis for marketing guidelines aims to provide businesses with a powerful tool to gain actionable insights from customer feedback. By accurately classifying sentiments, businesses can make informed decisions to enhance customer satisfaction and improve their products or services.