SENTIMENT ANALYSIS FOR MARKETING

Team Member
Balamurugan K – 211521243030

Phase 1:

Problem Definition:

This project aims to perform sentiment analysis on customer feedback expressed through tweets on Twitter regarding US airlines, with the goal of gaining actionable insights for marketing strategies. The primary objective is to accurately classify tweets into positive, negative, or neutral sentiments and extract meaningful topics to inform and guide critical business decisions within the US airline industry.

Design Thinking:

1.Data Collection:

The data utilized for the sentiment analysis project has been sourced from Kaggle, a renowned platform for datasets and data science resources. Dataset Link: **https://www.kaggle.com/datasets/crowdflower/twitter-airline-sentiment**

2.Data Preprocessing:

Effective data preprocessing is essential for ensuring that text data is clean, consistent and ready for sentiment analysis, which ultimately helps in extracting meaningful insights for marketing decisions.

1. **Text Cleaning:**

   Convert text to lowercase for uniformity.

   Remove special characters, symbols and punctuation.

   Eliminate URLs and HTML tags.

2. **Tokenization:**

   Break text into individual words or tokens.

3. **Stopword Removal:**

   Remove common stopwords (e.g., "the", "and", "is").

4. **Lemmatization or Stemming:**

   Reduce words to their base or root forms.

5. **Handling Contractions and Negations:**

   Expand contractions (e.g., "can't" to "cannot").

   Recognize and handle negations (e.g., "not good" vs. "good").

6. **Removing Duplicates and Noise:**

   Identify and remove duplicate or near-duplicate texts.

   Filter out noisy or irrelevant data points.

7. **Handling Missing Data:**

   Address missing data through imputation or sample removal.

8. **Text Length Normalization:**

   Normalize text length by padding or truncating.

9. **Exploratory Data Analysis (EDA):**

   Perform EDA to identify patterns and trends in the data.

10. **Text Vectorization:**

    Convert preprocessed text into numerical vectors suitable for machine learning models.

3.Sentiment Analysis Techniques:

The choice of the best NLP technique should align with our project's objectives, available data and resource constraints. It's often beneficial to experiment with multiple techniques and evaluate their performance on a validation dataset to determine which one provide the most meaningful insights for guiding our marketing decisions. Here are some NLP techniques that can be used:

1.Deep Learning Models:

Recurrent Neural Networks (RNNs): Suitable for sequence data, they can capture context in text.

Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) are popular RNN variants.

2.Bag of Words (BoW) and TF-IDF:

BoW represents text as a vector of word frequencies.

TF-IDF (Term Frequency-Inverse Document Frequency) assigns weights to words based on their importance in a document relative to a corpus.

3. Transformer Models:

Transformer models have achieved state-of-the-art performance in various NLP tasks, including sentiment analysis. They can capture complex language patterns and context effectively. They are suitable for large datasets and high computational resources. They are recommended if we aim for top-tier accuracy and have access to pre-trained models. Ex: BERT, GPT-3

4. Word Embeddings:

Word embeddings are useful for capturing semantic relationships between words. They provide meaningful vector representations of words but might not capture context as well as transformer models. Ex: Word2Vec, GloVe

5. Ensemble Methods:

Ensemble methods like combining multiple models (e.g., SVM + Word Embeddings) can enhance accuracy and robustness. Consider using an ensemble approach if you have the resources to train and manage multiple models.

4.Feature Extraction:

Feature extraction is a critical step in sentiment analysis. It involves converting raw text data into numerical features that machine learning models can understand and use for analysis. We can convert text data into numerical features using methods like BoW, TF-IDF, or word embeddings.

5.Visualization:

Visualization plays a crucial role in sentiment analysis as it helps to make the analysis results more interpretable and actionable. It not only helps in understanding the data but also in conveying insights to stakeholders effectively. Visualization techniques such as matplotlib, seaborn can be used.

6.Insights Generation:

1.Aspect-Based Sentiment Insights:

Identification of specific aspects of the airline experience (e.g., customer service, pricing, seat comfort) that are associated with positive and negative sentiments.

Understanding which aspects drive customer satisfaction and which require improvement.

2. Competitor Analysis:

Comparative analysis of sentiment scores among different US airlines.

Insights into which airlines consistently receive positive feedback and which face challenges in satisfying customers.

3. Temporal Trends:

Identification of temporal trends in sentiment, such as seasonal fluctuations or changes over months and years.

Correlation of sentiment trends with specific events or marketing campaigns.

4. Emotion Analysis:

Insights into the prevalent emotions expressed by customers in their feedback (e.g., happiness, frustration, gratitude).How emotions relate to specific aspects of the airline experience.

5. Geographic Patterns:

Regional analysis of sentiment to uncover geographic patterns in customer satisfaction.

Understanding how sentiment varies across different locations and regions.

6. Response Effectiveness:

Evaluation of how airlines' responses to customer feedback on Twitter impact sentiment.

Identification of response strategies that lead to improved customer satisfaction.

7. Complaint Analysis:

Analysis of the most common complaints and pain points mentioned by customers.

Prioritization of issues that require immediate attention and resolution.

8. Sentiment by Customer Segment:

Segmentation of customers based on demographics or behaviors (e.g., frequent flyers, business travelers).

Insights into how different customer segments perceive the airline experience.

9. Visualization Insights:

Insights from visualizations such as charts, graphs, and word clouds that highlight sentiment distribution and key terms.

Visual representations of sentiment trends and patterns.

10. Recommendations:

Actionable recommendations for improving customer satisfaction and addressing pain points based on sentiment analysis findings.

Prioritized strategies to enhance the overall customer experience.

These potential insights can provide a comprehensive understanding of customer sentiments and preferences related to US airlines. They serve as a valuable resource for data-driven decision-making, marketing strategies, customer service improvements, and business decisions aimed at enhancing customer satisfaction and competitiveness in the airline industry.

7.Conclusion:

The project delivers a set of actionable insights and recommendations based on sentiment analysis results. These insights highlight areas of strength and weakness in the airline services, identify aspects that drive positive and negative sentiments, and suggest strategies for improvement.

The project may recommend response strategies for airlines to engage with customers on social media effectively. It assesses the impact of responses on sentiment and suggests ways to improve customer relations.