Phase 2 : Innovation

Data Collection and Expansion

In this phase, we focus on improving our sentiment analysis solution for the "Twitter US Airline Sentiment" dataset. Here's a simplified overview of our approach:

Data Collection and Expansion - To Enhance Dataset Quality

We'll begin by enhancing the dataset through the following steps:

Data Enrichment: We'll add valuable attributes like user demographics, analyze geographical variations, and ensure a balanced sample representation across sentiments.

Data Preprocessing Enhancement - Improving Data Quality

We'll then concentrate on data preprocessing, making it specific to Twitter data:

Customized Preprocessing: Tailoring our preprocessing steps to handle mentions, hashtags, URLs, emojis, and noise reduction for cleaner data.

Sentiment Analysis Model Selection - Opting for Advanced Models

For sentiment analysis, we'll select advanced models and refine them for airline-related sentiments:

Model Evaluation: Assessing the performance of advanced models like BERT and RoBERTa on the dataset.

Fine-tuning: Customizing the selected models to capture nuances in airline-related sentiments.

Emotion Analysis: Exploring the inclusion of emotion analysis alongside sentiment.

Feature Extraction Enhancement

Feature Extraction Enhancement - Improving Data Representation

We'll enhance feature extraction techniques to represent textual data better:

Contextual Embeddings: Experimenting with contextual embeddings like ELMo to capture contextual meaning in tweets.

Subword Tokenization: Utilizing subword tokenization techniques to handle out-of-vocabulary words and slang effectively.

Twitter-specific Features: Extracting features unique to Twitter data, such as tweet length, mentions, and retweet counts.

Visualization Enhancement

Visualization Enhancement - Creating Insightful Visualizations

To make data more accessible, we'll focus on visualization:

Interactive Dashboards: Building dynamic dashboards to visualize sentiment trends, correlation with tweet attributes, and other insights.

Twitter Sentiment Trends: Visualizing sentiment trends over time and relating them to significant events or airline-specific occurrences.

Word Clouds and Hashtag Analysis: Generating word clouds and analyzing popular hashtags associated with different sentiments.

Insights Generation Automation

Insights Generation Automation - Streamlining Insights

We'll streamline insights generation and reporting processes:

Automated Reports: Developing automated reports summarizing key findings, sentiment trends, and notable tweets for each airline.

Root Cause Analysis: Implementing algorithms to identify root causes of negative sentiment spikes and positive sentiment trends.

Real-time Alerts: Setting up real-time alerts for extreme sentiment shifts that may require immediate attention.