Presentation:

The project delves into Aspect-Oriented Sentiment Analysis for Airlines, a study aimed at analyzing customer sentiments towards airline services using data of social media Platform X (Twitter). I embarked on this endeavor to address the inherent limitations of traditional sentiment analysis methods in capturing the nuanced feedback prevalent in complex domains such as airline experiences.

What sets my project apart from previous research is the adoption of aspect-oriented sentiment analysis, a departure from conventional sentiment analysis methodologies.

Sentiment analysis, is the process of computationally identifying and categorizing opinions expressed in text data to determine the sentiment or emotional tone conveyed by the author. It involves analyzing a piece of text to determine whether it expresses positive, negative, or neutral sentiment towards a particular subject, product, service, or topic.

And my project Aspect-oriented sentiment analysis goes beyond traditional sentiment analysis by identifying and analyzing specific aspects or features within text data and determining the sentiment expressed towards each aspect.

Additionally, I expanded the scope by leveraging a diverse array of machine learning classifiers, including Multinomial Naive Bayes, Multi-Layer Perceptron, XGBoost, Random Forest, and Support Vector Machine (SVM). By employing these classifiers, I sought to observe the most effective approach for sentiment analysis in our context.

\*Methodology Overview: Aspect-Oriented Sentiment Analysis for Airlines

Data Collection:

Obtained a dataset of 14,000+ airline tweets from Kaggle.

* Preprocessing:

Filtered tweets with low sentiment confidence (<0.5).

Cleaned text data by removing non-alphabetic characters, converting to lowercase, removing stopwords, and lemmatizing words.

* Aspect Extraction:

Employed aspect-based sentiment analysis to extract key aspects from each tweet.

Defined aspects such as comfort, staff, food, etc., and computed confidence scores.

* Classification and Analysis:

Used ML classifiers like Naive Bayes, MLP, XGBoost, Random Forest, and SVM.

Preprocessed data using CountVectorizer, split into train/test sets, and evaluated model performance.

Generated classification reports and visualized frequent terms using word clouds.

* Top Aspect Identification:

Processed data to identify top aspects driving sentiment for each airline.

Filtered dataset by sentiment, developed a function to identify top aspects, and presented insights.

Through rigorous experimentation and evaluation, we discovered that the Support Vector Machine (SVM) classifier exhibited superior accuracy compared to its counterparts.  
  
Upon obtaining the results, we took a step further by correlating the extracted positive and negative aspects with specific aspects of airline services. By establishing this connection, we aimed to provide actionable insights for airlines to enhance their business operations, improve customer satisfaction, and refine service quality. This approach not only enriches our understanding of customer sentiments but also offers tangible strategies for enhancing the overall airline customer experience.  
  
In summary, Our approach combines advanced natural language processing techniques with machine learning to provide actionable insights for improving customer satisfaction and service quality in the airline industry.

Questions:

1.What is aspect-based sentiment analysis and how is it different from traditional sentiment analysis?

Unlike traditional sentiment analysis, which provides an overall sentiment score for a piece of text, Aspect-based sentiment analysis is a technique used to analyze the sentiment expressed towards specific aspects or features of a product, service, or experience. This approach allows for a more nuanced understanding of customer opinions and preferences, particularly in domains with multiple aspects of interest, such as product reviews or social media posts about experiences.

2.What are the ML classifiers used, and define them? Why did you choose only these and not others?

We employed multiple machine learning classifiers for sentiment analysis, including:

Multinomial Naive Bayes (MNB): A probabilistic classifier based on Bayes' theorem with strong independence assumptions between features.

Multi-Layer Perceptron (MLP): A neural network model with multiple layers of nodes, capable of learning complex patterns in data.

XGBoost: An implementation of gradient boosted decision trees designed for speed and performance.

Random Forest: An ensemble learning method that constructs a multitude of decision trees and outputs the mode of the classes.

Support Vector Machine (SVM): A supervised learning model used for classification and regression analysis, effective in high-dimensional spaces.

We chose these classifiers based on their effectiveness in handling text classification tasks, particularly sentiment analysis. These classifiers are widely used in the literature and have demonstrated good performance in similar tasks. Additionally, the diversity of these classifiers allows us to compare their performance and select the most suitable one for our specific dataset and task.

3.Why did you choose those 4 evaluation metrics and not others?

We chose accuracy, precision, recall, and F1 score as evaluation metrics because they provide comprehensive insights into the performance of the classifiers across different aspects of sentiment analysis.

Accuracy: Measures the overall correctness of the classifier's predictions.

Precision: Indicates the proportion of true positive predictions among all positive predictions made by the classifier.

Recall: Measures the ability of the classifier to correctly identify all positive instances in the dataset.

F1 Score: Represents the harmonic mean of precision and recall, providing a balanced assessment of the classifier's performance.

These metrics are commonly used in classification tasks and provide a balanced evaluation of the classifier's performance in terms of both false positives and false negatives. Additionally, they help us understand how well the classifier performs across different sentiment classes and guide us in selecting the most suitable model for sentiment analysis.