Tech Horizon Internship 2025/TASK-03

Submitted By:Aleem Shoukat
Submitted to: Tech Horizon Internship 2025(Data Analytics)
Date:23/Aug/2025

Task 3: Airline Sentiment Analysis

1. Introduction

The purpose of this task was to analyze customer sentiments regarding different airlines using publicly available Twitter review data. Sentiment analysis is a key Natural Language Processing (NLP) technique that helps companies measure customer satisfaction and identify areas for improvement.

Instead of using live Twitter API access, we relied on a publicly available dataset due to access limitations, which will be discussed later in the report.

---

2. Objective

- Perform sentiment analysis on airline customer reviews.
- Identify which airlines receive the most positive feedback and which face the most negative comments.
- Summarize the findings in a structured way to support data-driven decision-making.

---

3. Methodology

1. **Dataset Selection**
    - Used a publicly available dataset of airline tweets (instead of live API).
    - The dataset contained tweets labeled as **positive, neutral, or negative** along with the airline name.
2. **Data Cleaning**
    - Removed unnecessary symbols, special characters, and emojis.
    - Preprocessed text to make it suitable for sentiment analysis.
3. **Sentiment Analysis**
    - Applied **TextBlob** to calculate polarity scores (ranging from -1.0 to +1.0).
    - Classified tweets as **positive, negative, or neutral**.
4. **Comparison with Original Labels**
    - Built a confusion table to compare original dataset labels with TextBlob results.
    - Achieved an agreement rate of **44.6%** (Exact-label matches).
5. **Result Extraction**
    - Identified top positive and negative tweets for each airline.
    - Summarized the performance of airlines based on customer feedback.

4. Key Findings

- **Best Airlines (Positive Sentiments):**
    - **Delta Airlines** and **Southwest Airlines** received the most positive feedback, mainly praising helpful staff and good service.
- **Worst Airlines (Negative Sentiments):**
    - **United Airlines** and **US Airways** faced the highest number of negative tweets, mostly about delays, poor gate service, and customer dissatisfaction.
- **Mixed Sentiments:**
    - **American Airlines** had both strong positive reviews (appreciation for helpful staff) and strong negative complaints (poor communication and customer handling).

5. Reason for Not Using Twitter Live API

Initially, the task plan was to fetch live tweets directly through the **Twitter API**. However, due to Twitter's updated developer access policy in 2023–24:

- Free API access was discontinued.
- Paid subscription is now required even for basic queries.
- Authentication and approval delays made it infeasible within the task's time constraints.

For these reasons, a **publicly available dataset** was used to ensure the analysis could be completed effectively and on time.

---

6. Conclusion

This project demonstrated how sentiment analysis can be applied to airline customer reviews to evaluate service performance. The results highlight clear distinctions between airlines that maintain customer satisfaction and those facing frequent criticism.

- **Best performing airlines:** Delta, Southwest
- **Worst performing airlines:** United, US Airways
- **Mixed performance:** American Airlines

Although TextBlob achieved only 44.6% exact-label agreement, it provided meaningful insights into customer sentiment trends. With access to more advanced NLP models (e.g., BERT, RoBERTa) or live API data, future analysis could achieve higher accuracy.

---

7. Deliverables

- Cleaned dataset (with preprocessed text).
- Python notebook/code implementing sentiment analysis.
- Confusion table and sentiment classification results.
- Final professional report summarizing methodology, results, and limitations.
