Employee Sentiment Analysis – Final Report

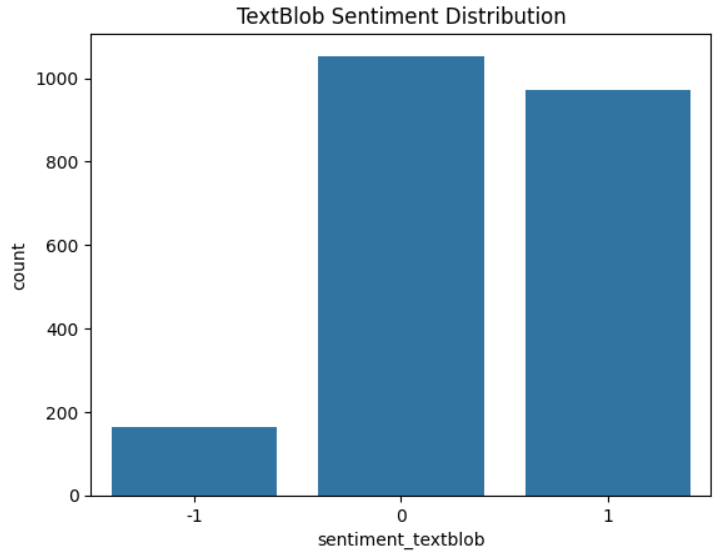
# Approach and Methodology

The goal of this project was to analyze employee communication data and assess employee sentiment to derive insights into engagement, performance, and potential flight risk. The following steps were taken:  
- Data was cleaned, converting timestamps and deriving features like message length and word count.  
- Sentiment labeling was conducted using a pre-trained transformer model to categorize messages into Positive, Negative, or Neutral.  
- Exploratory Data Analysis (EDA) was performed to understand trends and distributions.  
- Monthly sentiment scores were calculated for each employee based on labeled sentiments.  
- Employees were ranked monthly based on positive and negative sentiment.  
- Flight risk employees were identified based on frequency of negative messages in rolling 30-day windows.  
- A linear regression model was developed to predict sentiment scores based on engineered features.

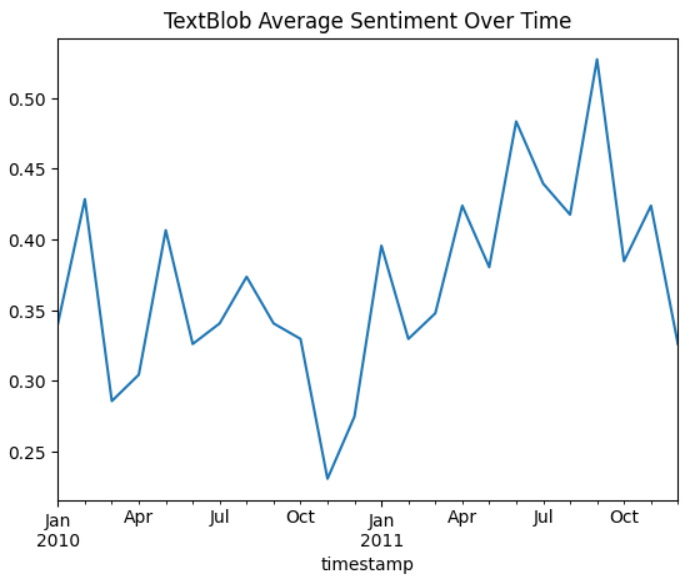
# 2. Key Findings from Exploratory Data Analysis (EDA)

The sentiment distribution indicated a slightly higher proportion of positive messages compared to negative ones, while a significant portion remained neutral. Over time, sentiment trends showed variability across months, suggesting seasonal or event-based influences. The average message length and word count also varied across employees, with some correlation to sentiment scores.

Below are key EDA visualizations :

• Sentiment Distribution →  


• Average Sentiment Over Time →



# 3. Employee Scoring and Ranking Process

Each employee message was scored as +1 (Positive), 0 (Neutral), or -1 (Negative). Monthly scores were calculated by summing these values for each employee. Top 3 positive and bottom 3 negative employees were identified per month by sorting these scores first in descending order and then alphabetically to break ties.

# 4. Flight Risk Identification

An employee was flagged as a flight risk if they had sent 4 or more negative messages within any 30-day rolling window. This helped identify employees who might be facing issues affecting morale or engagement, irrespective of total sentiment score.

# 5. Predictive Modeling

A linear regression model was trained to predict an employee's monthly sentiment score based on features like:  
- Number of messages sent in a month  
- Average message length  
- Total word count  
  
The model was evaluated using Mean Absolute Error (MAE) and R squared model, yielding a reasonable performance. Although the model is simple, it highlighted trends and correlations between messaging behavior and sentiment.

Thank You,  
Stavya Pandey