# Final Report: Employee Sentiment Analysis

## 1. Project Overview

This project analyzed internal employee messages to assess sentiment and engagement. The workflow included sentiment labeling using VADER, exploratory data analysis, monthly sentiment scoring, employee ranking, flight risk detection, and predictive modeling.

## 2. Sentiment Labeling

We used the VADER sentiment analyzer to classify each message as Positive, Negative, or Neutral. The compound score thresholds were defined as follows:  
- Compound > 0.05: Positive  
- Compound < -0.05: Negative  
- Between -0.05 and 0.05: Neutral  
These thresholds were validated using a sample of hand-labeled internal messages.

## 3. Exploratory Data Analysis

We analyzed 2,191 messages across 10 employees. Visualizations included sentiment distribution and monthly message trends. A spike in messaging activity in April may reflect internal announcements, while lower engagement in June could be attributed to vacation periods.

## 4. Monthly Employee Score Calculation

Sentiment scores were calculated by assigning +1 for Positive, -1 for Negative, and 0 for Neutral messages. Scores were summed per employee per month to assess engagement trends.

## 5. Employee Ranking

Employees were ranked each month based on total sentiment scores. The top 3 most positive and most negative employees were highlighted to identify strong contributors and potentially disengaged individuals.

## 6. Flight Risk Identification

An employee was flagged as a flight risk if they sent 4 or more negative messages within any 30-day rolling period.

## 7. Predictive Modeling

We trained a linear regression model using message count, average length, and word count to predict monthly sentiment scores. The model showed:  
- MAE: 3.33  
- RMSE: 4.48  
- R² Score: -19.11  
The poor fit suggests the need for more features and a larger dataset. Still, the model offered insight into how message activity might correlate with sentiment.

## 8. Conclusion & Recommendations

This project demonstrated that employee sentiment analysis can uncover early signs of disengagement. By combining message scoring, flight risk detection, and modeling, organizations can proactively manage retention. Future work should include topic modeling and real-time dashboards for better decision-making.