**EMPLOYEE SENTIMENT ANALYSIS PROJECT**

Internship Submission Report  
By: Dwayne Fernandes

## 1. Project Overview

This project involves analyzing an unlabeled dataset of employee messages to evaluate overall sentiment and engagement. Using Natural Language Processing (NLP) and statistical modeling, the objective was to classify sentiments, derive insights, compute sentiment-based employee scores, identify potential flight risks, and build a predictive model to estimate future sentiment trends.

## 2. Methodology

**Task 1: Sentiment Labeling –** Each message was analyzed using the VADER sentiment analyzer (fast rule-based NLP). Messages were labeled as Positive, Neutral, or Negative based on compound sentiment scores.

**Task 2: Exploratory Data Analysis –** EDA was performed to understand sentiment distribution, monthly patterns, and employee activity. Visualizations included sentiment trends over time, message length, and top active employees.

**Task 3: Employee Score Calculation –** Each message was assigned a score (+1 for Positive, –1 for Negative, 0 for Neutral). Scores were aggregated monthly per employee to obtain monthly sentiment scores.

**Task 4: Employee Ranking –** Employees were ranked monthly based on their sentiment scores. The top three with the highest and lowest scores were identified and visualized.

**Task 5: Flight Risk Identification –** Employees with four or more negative messages in any rolling 30-day period were flagged as potential flight risks.

**Task 6: Predictive Modeling –** A Linear Regression model was developed using monthly features such as message frequency, word count, and length statistics. The model predicted employee sentiment scores to capture engagement trends.

## 3. Key Insights & Findings

• Majority of messages were Neutral, with Positive messages slightly higher than Negative.  
• Monthly sentiment trends revealed periods of low engagement correlating with increased Negative messages.  
• Top-performing employees showed consistently positive communication tone.  
• A few employees exhibited patterns meeting the flight risk criteria (≥4 negative messages in 30 days).

## 4. Predictive Model Evaluation

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset | R² Score | MSE | RMSE | MAE |
| Training | 0.9431 | 0.9782 | 0.9891 | 0.7249 |
| Testing | 0.9103 | 1.0532 | 1.0263 | 0.7431 |

The model achieved strong predictive performance, with high R² values and low errors across both training and testing sets. This indicates that the chosen features explain sentiment score variability effectively.

## 5. Conclusion

The project successfully automated sentiment labeling, analyzed communication trends, and quantified employee engagement. The predictive model demonstrated robust accuracy, confirming its capability to capture sentiment dynamics. These insights can help organizations proactively address employee dissatisfaction and identify potential attrition risks.