Employee Sentiment Analysis Project

Submitted By: Bryan Samuel Das

Email: your.bryansamueldas@gmail.com

Submission Date: 6/22/2025

Internal AI Evaluation — Glynac.ai

# Table of Contents

1. Project Overview

2. Methodology

3. Sentiment Labeling

4. Exploratory Data Analysis (EDA)

5. Employee Sentiment Score Calculation

6. Employee Ranking

7. Flight Risk Identification

8. Predictive Modeling

9. Conclusion & Key Recommendations

10. References

# 1. Project Overview

* Brief intro about project goal
* Dataset description (test(in).csv, columns overview)
* Business context: Understanding employee sentiment & engagement health

# 2. Methodology

Summary of techniques:

* NLP for sentiment labeling (VADER)
* Exploratory data analysis
* Scoring logic (+1, 0, -1)
* Rolling 30-day flight risk detection
* Linear regression predictive modeling

Tools & Libraries: Python, Pandas, NLTK, Seaborn, Scikit-learn

# 3. Sentiment Labeling

Approach: VADER sentiment analyzer

* Classification thresholds:
* Compound ≥ 0.05 → Positive
* Compound ≤ –0.05 → Negative
* Else Neutral

Summary of distribution (include pie chart or bar plot image)

# 4. Exploratory Data Analysis (EDA)

* Data health check: missing values, data types
* Sentiment distribution chart
* Trend of sentiment over time (stacked bar chart)
* Message length and word count distribution graphs
* Top 10 employees by message volume
* Optional: WordCloud of most common words

# 5. Employee Sentiment Score Calculation

* Scoring logic table
* Monthly sentiment score calculation method
* Example table showing employee\_id, month, sentiment\_score

# 6. Employee Ranking

Method to identify:

* Top 3 positive scorers per month
* Top 3 negative scorers per month

Include sample tables or charts showing rankings

# 7. Flight Risk Identification

* Definition of flight risk (≥4 negative messages in rolling 30 days)
* Rolling count method explanation
* List of employees flagged as flight risks

# 8. Predictive Modeling

Features used:

* Avg message length
* Avg word count
* Message count per month

Model chosen: Linear Regression

* Model evaluation:
  + Mean Squared Error (MSE)
  + R² Score
* Feature coefficients interpretation
* Optional: prediction scatter plot

# 9. Conclusion & Key Recommendations

Summary of insights:

* Sentiment trends
* Top performers & at-risk employees
* Model’s predictive reliability

Business recommendations:

* Employee engagement programs
* Proactive outreach for flight risk staff
* Regular sentiment monitoring

# 10. References

* NLTK VADER documentation
* Scikit-learn Linear Regression docs
* Seaborn visualization docs
* Any additional resources you consulted