Final Report:
Employee Sentiment Analysis v2.0

Shaidatullisa Nadia Binti Saipudin

8th June 2025

Table of Contents

Introduction

This report presents a comprehensive analysis of an unlabeled dataset of employee messages. The goal was to assess employee sentiment and engagement using a combination of natural language processing (NLP), exploratory data analysis (EDA), scoring and ranking methods, flight risk identification, and predictive modeling. Each task was designed to systematically transform raw data into meaningful insights that can guide organizational understanding and proactive HR strategies.

1. Approach and Methodology

The project is divided into 6 tasks as follows:

Task 1: Hybrid Sentiment Labeling
I combined a transformer-based sentiment pipeline with tailored keyword rules to classify each message as Positive, Negative, or Neutral. This hybrid approach improved coverage and accuracy by capturing both contextual tone and domain-specific vocabulary.

Task 2: Exploratory Data Analysis (EDA)
I verified data completeness (no missing values), parsed dates into multiple time features, and examined the overall distribution and temporal trends of sentiment. Visualizations: bar charts, line plots, rolling averages, and boxplots

Task 3: Employee Score Calculation
Each message was mapped to +1 (Positive), –1 (Negative), or 0 (Neutral) and summed per employee for every calendar month. This yielded a clear, comparable monthly sentiment score for each individual.

Task 4: Employee Ranking
Using the monthly scores, I identified the top three most positive and top three most negative employees each month. This spotlighted consistently engaged contributors as well as those who might need support.

Task 5: Flight-Risk Identification
I implemented a rolling 30-day window to flag any employee sending ≥ 4 negative messages in that span. Flagged individuals were then prioritized by their total negative count to focus outreach where it's most needed.

Task 6: Predictive Modeling
I trained a linear regression model on features derived from monthly message behavior: volume, word and character statistics, and sentiment ratios. The final model achieved:

- **R^2 Score:** 0.4856
- **Mean Squared Error:** 3.6474

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

Data Completeness & Structure:
All 1,539 messages had valid text and dates, which were expanded into `year_month`, `month`, and `day` fields. Message-level metrics (`message_length`, `char_count`, etc.) revealed wide variability, supporting rich downstream analysis.

Sentiment Distribution:
Positive messages (≈1,068) were most frequent, followed by Negative (≈927) and then Neutral (≈186). Neutral communications are rare, confirming that employees tend to express distinct approval or disapproval more often than ambivalence.

Monthly Sentiment Trends:
Over 24 months, Positive counts typically ranged 40–50 per month, Negative 35–45, and Neutral under 10, with two notable bursts of Neutral in September 2010 and April 2011—likely reflecting all-hands announcements or policy updates.

Message Length by Sentiment:
Positive messages are the longest on average (median ~32 words; outliers up to ~190), Neutral the shortest (~26 words), and Negative in between (~27 words). This suggests that praise or encouragement often comes with more elaboration.

7-Day Rolling Sentiment Trend:
Daily rolling averages show Positive and Negative each averaging ~1–2 messages per day, while Neutral hovers near zero with occasional spikes—indicating that short-lived shifts in tone occur but overall volume is stable.

Per-Employee Sentiment Ratios:
Most individuals maintain a balanced or positive tilt. Notably, **bobette.riner@ipgdirect.com** exhibits a high Negative ratio (~46.1%), while **lydia.delgado@enron.com** has the lowest (~38.4%), signaling where deeper context may be needed.

High-Volume Negative Senders:
lydia.delgado@enron.com and **john.arnold@enron.com** each sent 109 negative messages, topping the list of high-engagement yet negative communicators. They and eight others all exceeded the four-message flight-risk threshold.

See some visual summaries below:

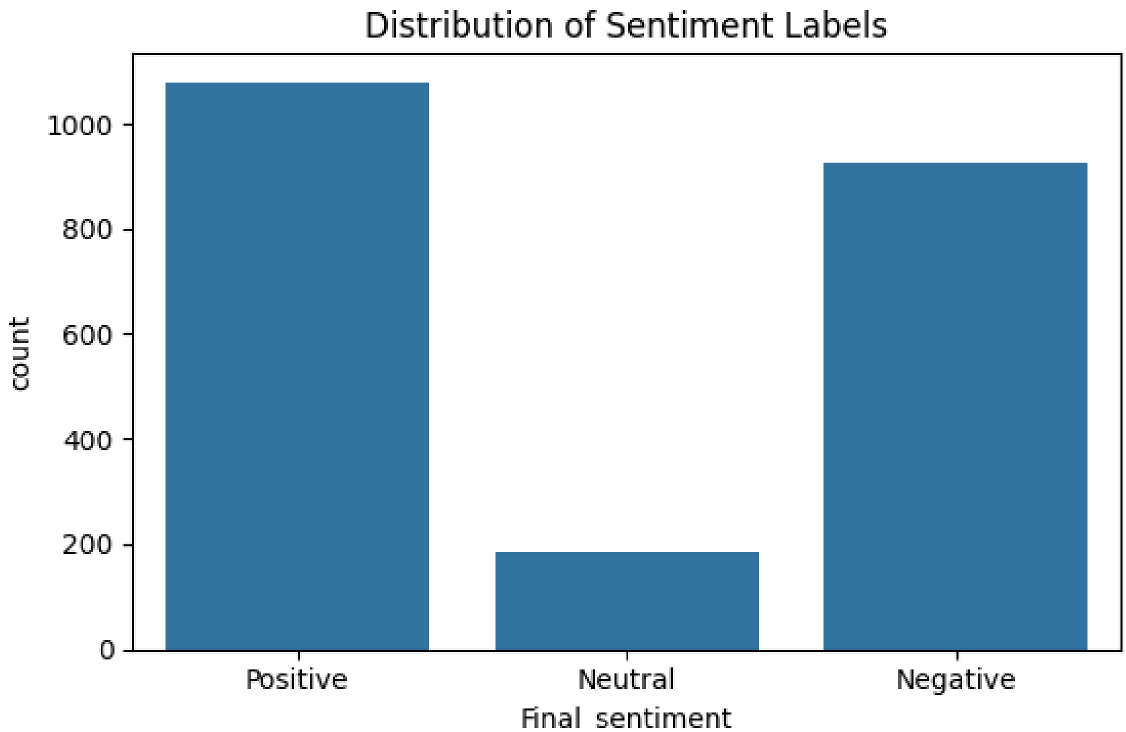

Figure 1: Sentiment Distribution

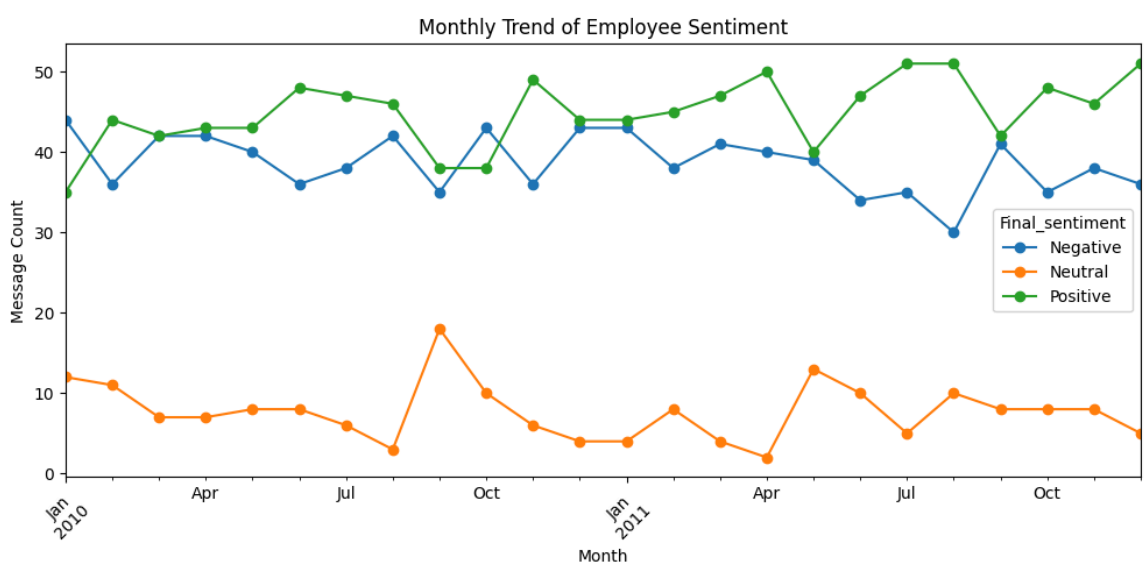

Figure 2: Monthly Trend of Employee Sentiment

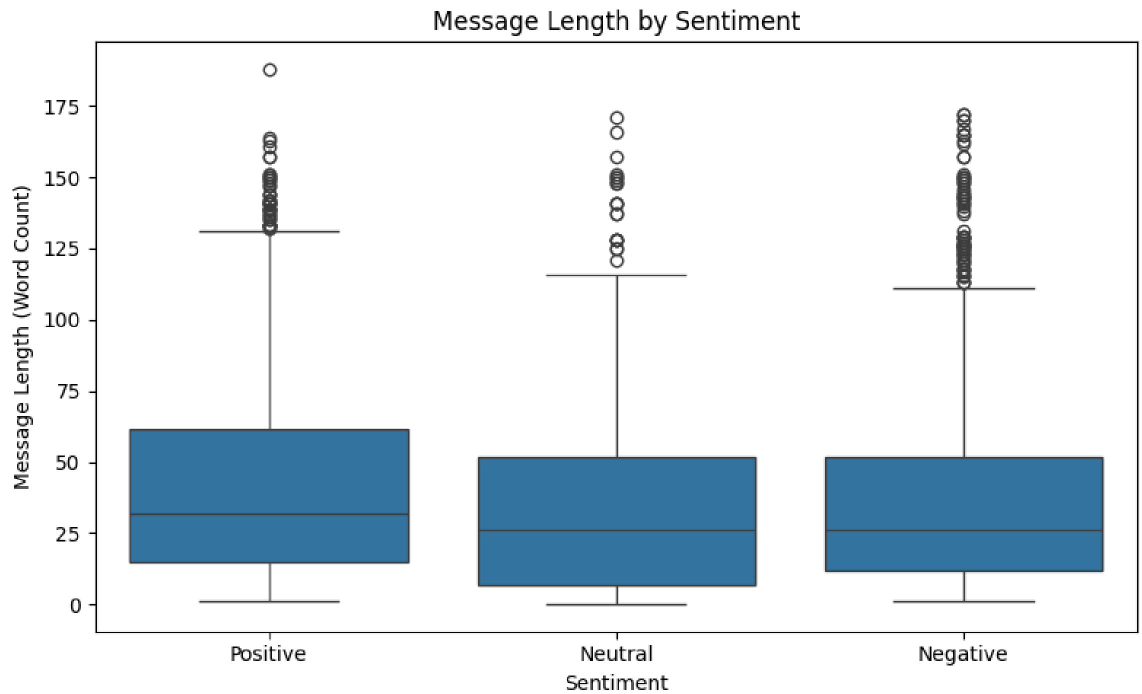

Figure 3: Message Length by Sentiment

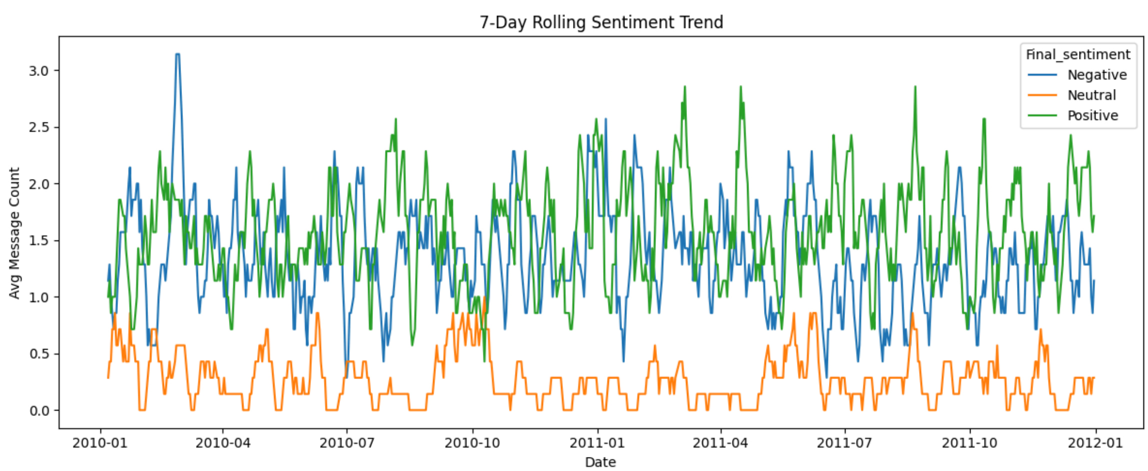

Figure 4: 7-Day Rolling Sentiment Trend

3. Employee Scoring and Ranking

In this task, I calculated a **monthly sentiment score** for each employee based on their labeled messages. This score reflects overall emotional tone and engagement in employee communication.

To compute the score, each message was assigned a value:
+1 for Positive, **–1** for Negative, and **0** for Neutral.
These values were then **summed per employee each calendar month**, creating a rolling sentiment profile that resets monthly for consistency.

This scoring method allowed me to identify which employees consistently expressed positive communication, and which ones tended to be more negative over time. The score provided a quantitative way to compare employee sentiment trends on a monthly basis.

Below are the example visuals:

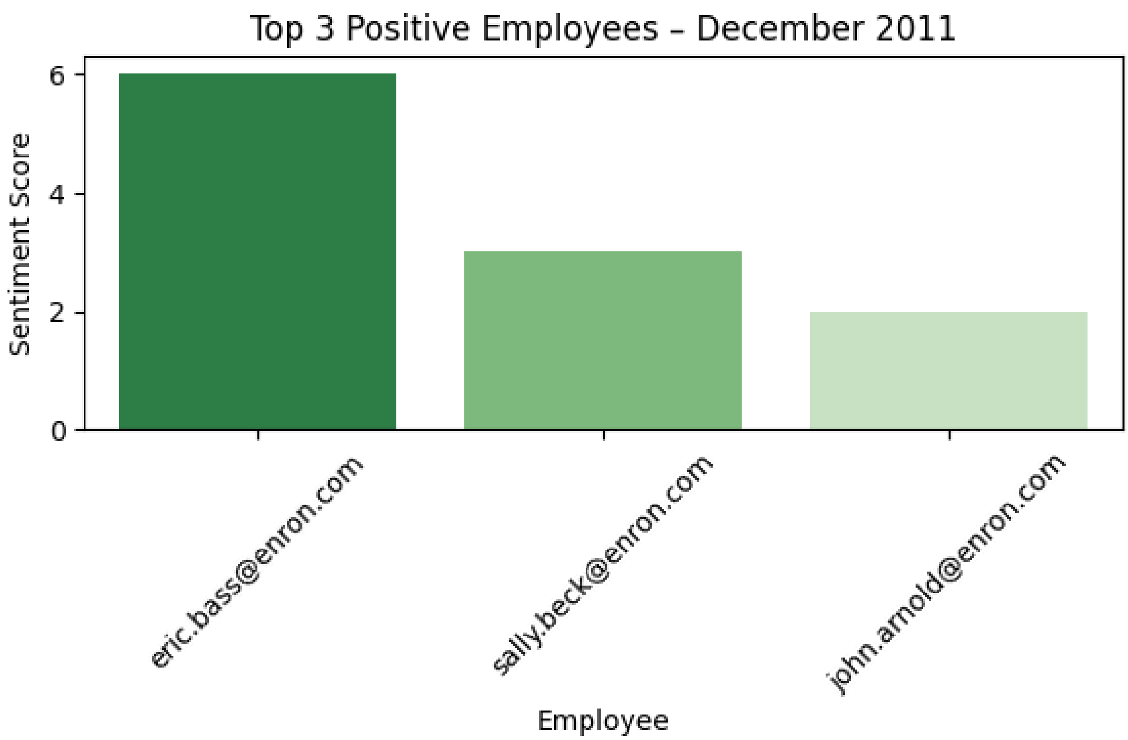

Figure 6: Top 3 Positive Employees of the month

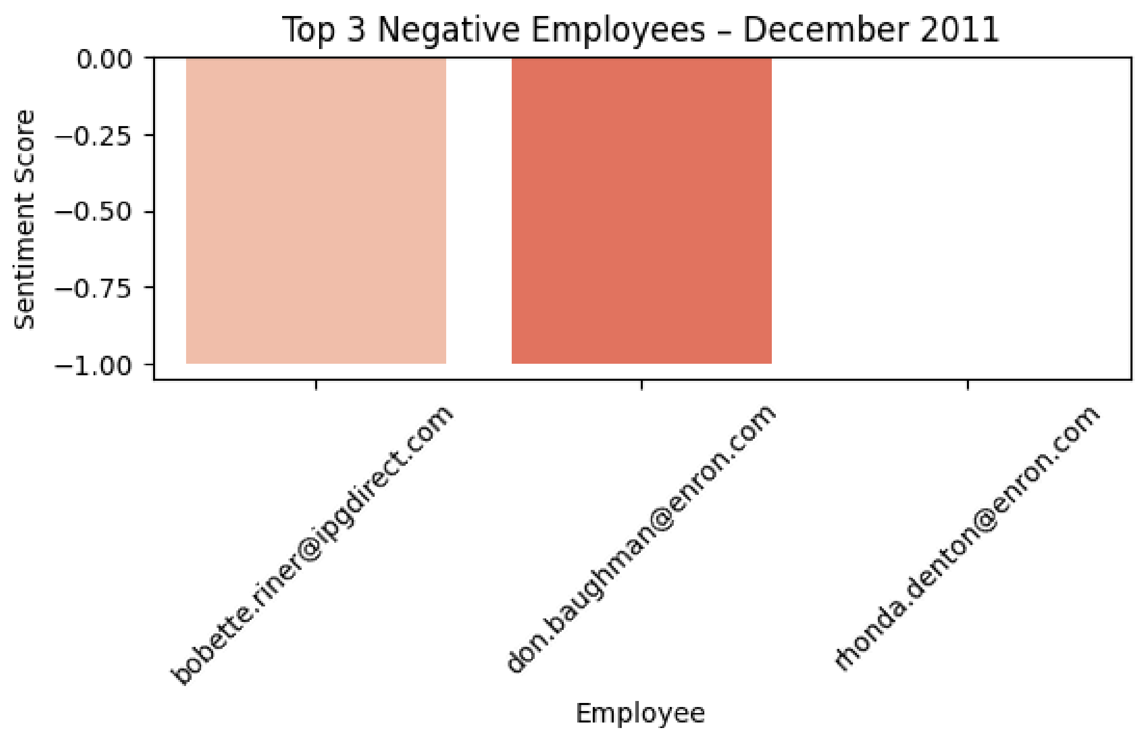

Figure 7: Top 3 Negative Employees of the month

4. Flight Risk Identification

- In this task, I analyzed employee messages to identify individuals at risk of disengagement or resignation. I focused specifically on those who exhibited a pattern of negative communication behavior.
- Using a **rolling 30-day window**, I flagged any employee who sent **4 or more negative messages within any 30-day span**. This method ensures that patterns are captured over time, rather than being restricted to calendar months.
- Employees who met the threshold were labeled as **flight risks**. I then counted the total number of negative messages each flagged employee sent, helping to prioritize who may need closer monitoring or support.

Visualizations and Interpretation:

1. **Bar Chart – Flight-Risk Totals:**
   All ten flagged employees exceeded four negatives in 30 days. At the top were **lydia.delgado@enron.com** (110 negatives) and **john.arnold@enron.com** (109), followed closely by **patti.thompson@enron.com** (100) and **bobette.riner@ipgdirect.com** (100).

2. **Line Chart – Rolling 30-Day Negative Count (Bobette Riner):**
 Peaks reached 10 negatives in early 2011, with multiple intervals above the 4-message threshold—demonstrating persistent frustration rather than isolated incidents.

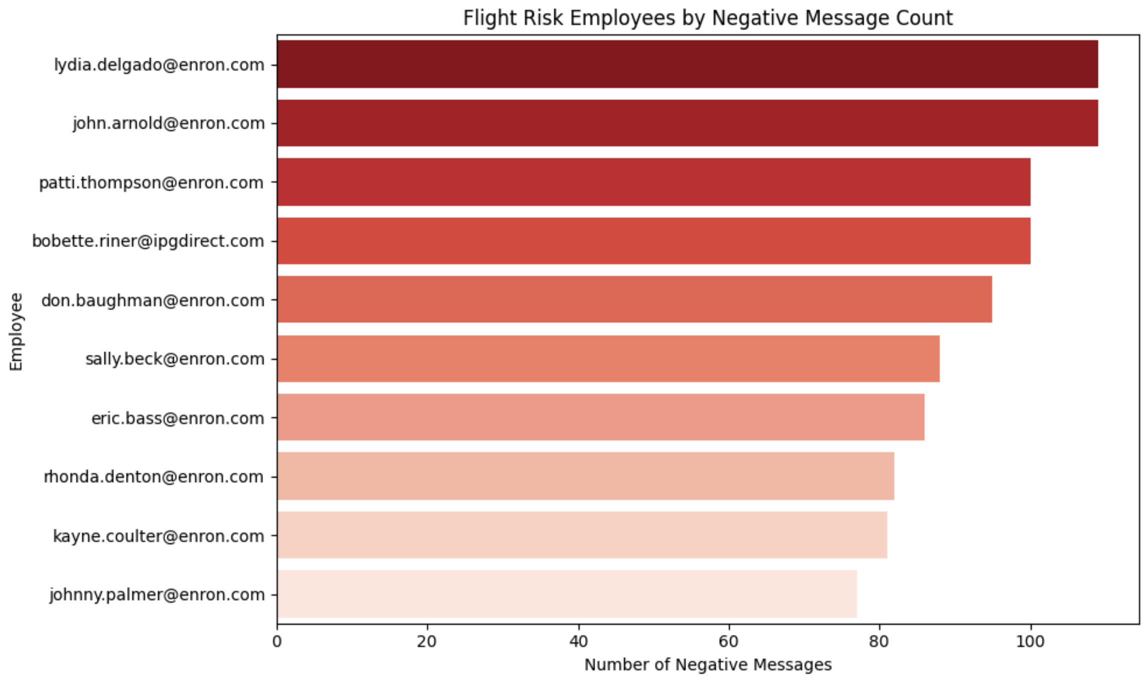

Figure 8: Flight Risk Employees by Negative Message Count

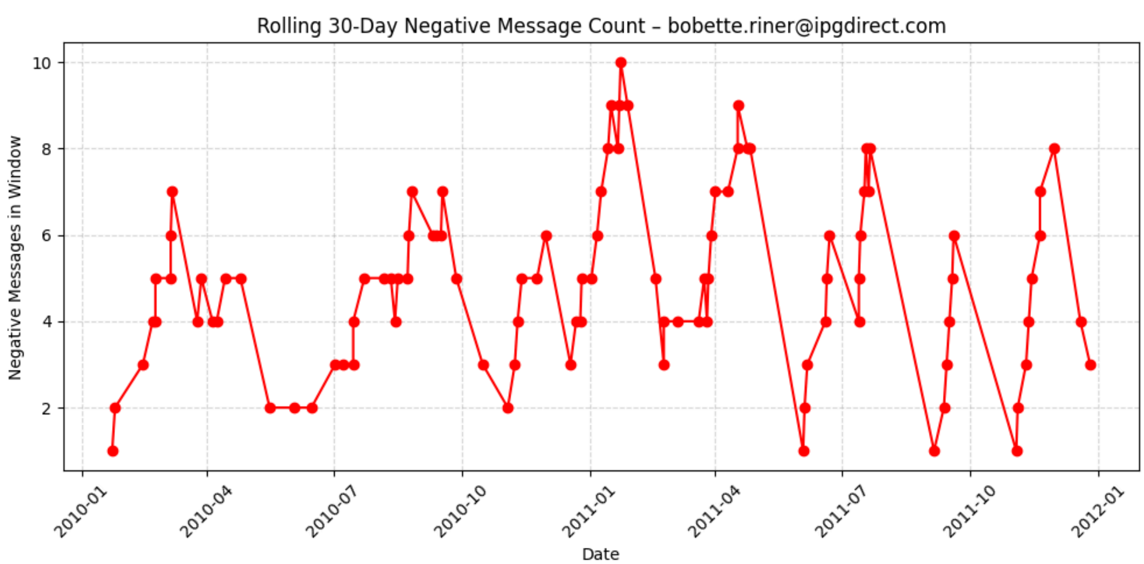

Figure 9: Rolling 30-Day Negative Message Count (one employee example)

5. Predictive Model Overview and Evaluation

A linear regression model used the following feature set:

- **Volume Metrics:** message_count, total_words, total_chars
- **Length/Density Metrics:** avg_message_length, avg_chars_per_message, words_per_message, chars_per_word
- **Sentiment Ratios:** positive_ratio, neutral_ratio, negative_ratio

Key Performance:

- **R^2 Score:** 0.4856 (≈ 49% variance explained)
- **MSE:** 3.6474

An R^2 of 0.4856 means our model explains about 48.6% of the month-to-month variation in sentiment scores, while an MSE of 3.6474 (equivalent to an RMSE of ≈1.91) indicates that, on average, our predictions deviate from the actual scores by about 1.9 points, showing decent overall accuracy but highlighting room to better capture the remaining variance.

Positive and Negative ratios dominate, confirming that sentiment proportions drive most of the predictive power. Other linguistic features play supporting roles.

The **Actual vs. Predicted** scatter confirms good overall alignment along the y = x line, though extreme sentiment scores (±5+) tend to be under-predicted, suggesting benefit from nonlinear models or additional contextual features.

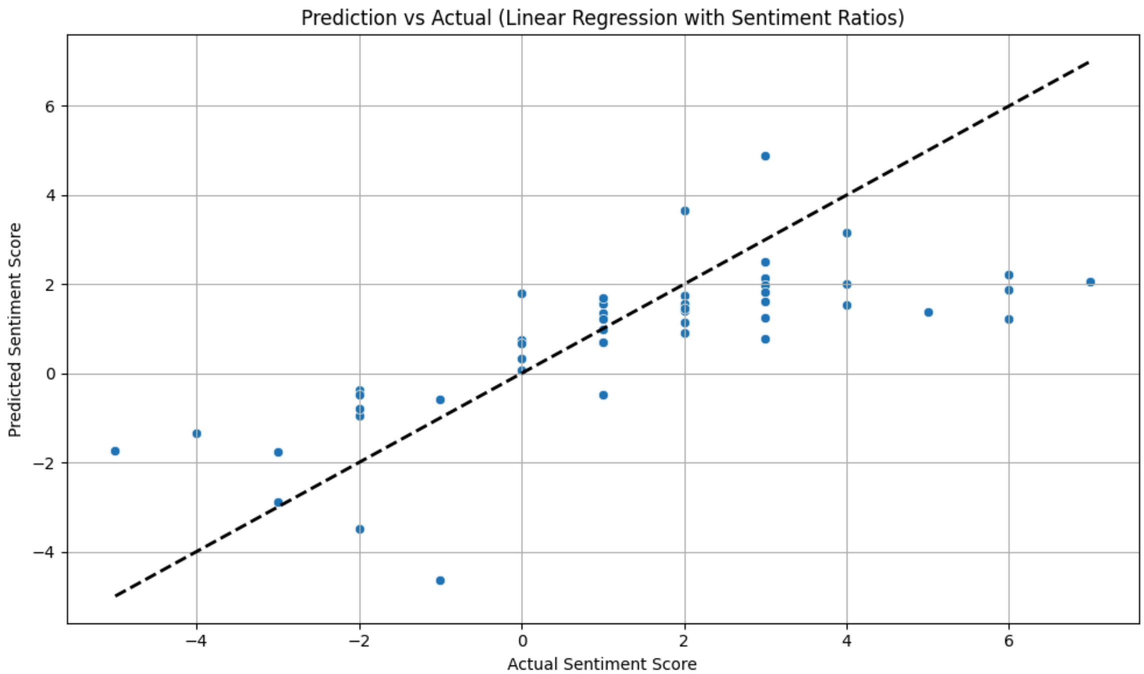

Figure 10: Predicted Vs Actual Sentiment Score Sum

Conclusion

This end-to-end workflow, from hybrid sentiment labeling and EDA to scoring, ranking, risk detection, and modeling demonstrates that internal message data can reliably surface both broad trends and individual cases of engagement or concern. By:

- Automating alerts when rolling 30-day negatives exceed four,
- Recognizing employees who drive positive culture,
- Targeting interventions for high-risk individuals, and
- Refining predictive models with richer features,

organizations can proactively monitor morale, celebrate successes, and address emerging frustrations