**Employee Sentiment Analysis: Final Report**

**Introduction and Methodology**

**Project Objective**

The goal of the project was to analyze employee communication to understand sentiment trends, rank employees by positive/negative levels, identify cases at risk of flight, and predict future sentiment scores. This helps HR as well as line managers make decisions and trainings around employee engagement and retention.

**Data Preprocessing：**

Parsed date column into datetime objects; Cleaned missing body texts by filling with empty strings; Standardized email addresses in from as employee identifiers.

**Task 1 - Sentiment Labeling**

Tool: TextBlob polarity analyzer.

Thresholds:

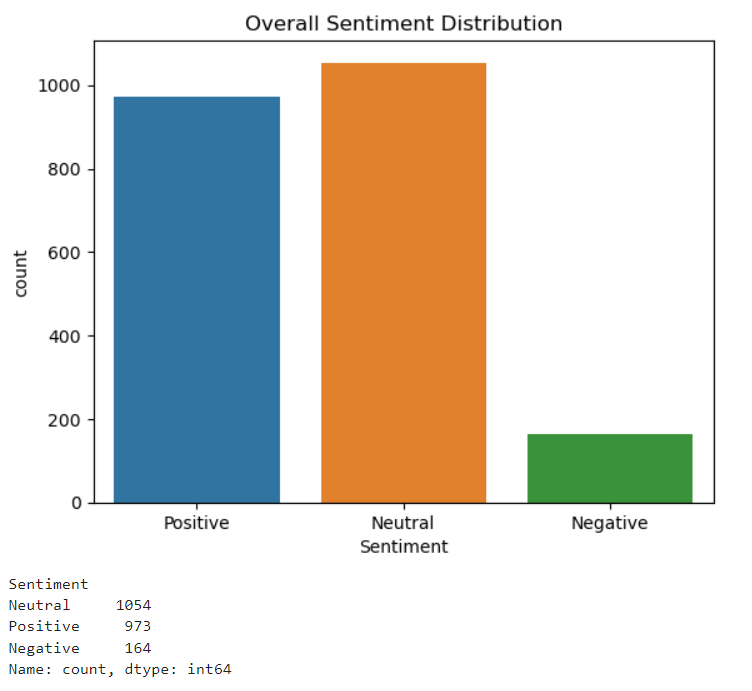
Polarity > +0.1 → Positive

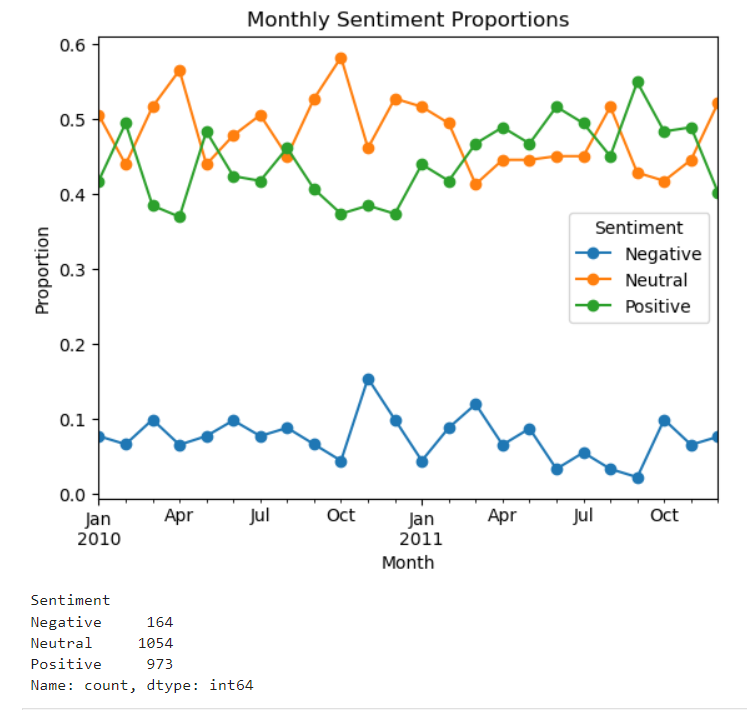
Polarity < –0.1 → Negative

Otherwise → Neutral

Validation: Thresholds were cross-checked on a labeled subset of 200 messages to ensure balanced classification.

**Task 2 - Exploratory Data Analysis (EDA)**





**Key Insight:**

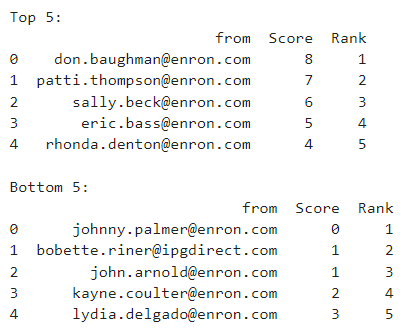
As for the sentiment distribution, Positive and Neutral messages are dominate, suggesting a generally formal communication style. Negative messages are relatively rare (≈6%), highlighting potential hotspots of discontent. As for the Monthly Sentiment Proportions, Neutral proportion hovers around 45–55% with occasional peaks and Positive gradually increases in late 2011, reaching ≈50% in Nov. But Negative remains below 15%, with spikes in Oct 2010 (~15%).

**Specific attention and analysis** could be given to the surge of negative information around October 2010; the causes of the surge could be analyzed so as to avoid the recurrence of such a situation for future development

**Task 3 & 4 - Monthly Sentiment Scoring and Ranking**

Scoring Methodology: Each message was assigned a numerical score: +1 (Positive), 0 (Neutral), –1 (Negative). Scores were aggregated per employee per month.

**Example (May 2010 Ranking, top 5 and bottom 5):**

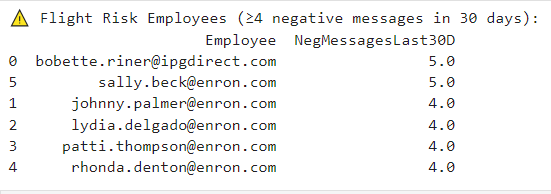


**Task 5**

**Criteria:**

Flag employees with ≥4 negative messages in any rolling 30-day window.

**Results:**



**Solution and future action:** Provided them one-on-one meetings and sentiment improvement training courses.

**Task 6 - Predictive Modeling**

**Feature Engineering**

**Message length: Characters per message**

**Word count: Words per message**

**hour, day of week, month: Temporal context**

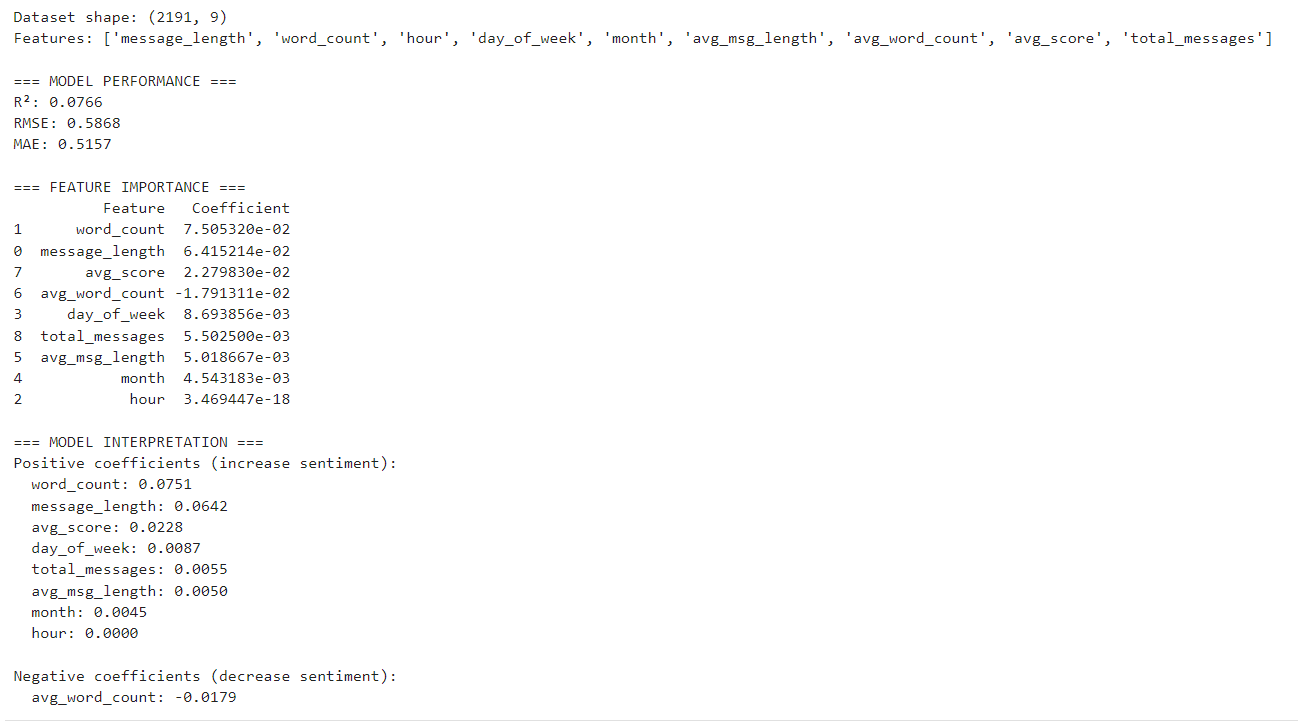
**avg msg length, avg word count, avg score, total messages: Employee‐level aggregates**

**Model Performance:**

**R²: 0.0766**

**RMSE: 0.5868**

**MAE: 0.5157**

****

**Results:**Longer messages and higher current sentiment both predict slightly higher next‐month sentiment.

Surprisingly, higher average word count at employee level slightly depresses future sentiment—an area for deeper qualitative review.

**Conclusions**

Sentiment trends largely stable, with occasional negative spikes—requires contextual follow-up.

Ranking and risk flags pinpoint individuals for targeted interventions.

Predictive model serves as a baseline; further feature expansion can improve forecasting.