Employee Sentiment Analysis – Final Report

Dataset Used: `test`
Total Messages: 2,192
Project Type: Internal NLP Evaluation
Author: *Mohamed Abnoune*

1. Methodology Overview

The objective of this project was to evaluate employee sentiment and engagement using an unlabeled dataset of internal messages. The tasks included sentiment labelling, exploratory data analysis, employee scoring and ranking, flight risk detection, and predictive modelling.

Python was used as the programming language, along with:

- Hugging Face Transformers (for sentiment labeling)
- Pandas, Matplotlib, Seaborn (for data wrangling and visualization)
- Scikit-learn (for linear regression modelling)

2. Sentiment Labeling

The cardiffnlp/twitter-roberta-base-sentiment model was used to classify messages as Positive, Neutral, or Negative. Trained on short-form social text like tweets, it aligns well with the internal messages in our dataset.

We compared RoBERTa's results with VADER, a rule-based sentiment analysis tool from NLTK. Our findings showed only 48.56% agreement, with VADER often misclassifying subtle tones as overly positive. Manual review revealed that RoBERTa was significantly more accurate in recognizing negative or neutral tones in professional communication.

This comparison highlights that transformer-based models, despite their complexity, better understand context than dictionary-based tools like VADER.

3. Exploratory Data Analysis (EDA)

Key Findings:

- Out of 2,192 messages:
  - **~68%** were **Neutral**
  - **~25%** were **Positive**
  - **~7%** were **Negative**
- Positive and neutral messages had longer average character counts than negative messages.
- Message lengths ranged mainly between **20 and 200 characters**.
- Some employees (e.g., sally.beck@enron.com) were highly active over multiple months.

Visualizations showed that Neutral messages made up the majority (>68%), while Positive messages accounted for around 25% and Negative for under 10%. The average message length was higher for Positive and Neutral messages compared to Negative ones.

Time-based analysis revealed that message volume remained stable across months, and sentiment trends were relatively consistent, with no drastic spikes in negativity.

Top active senders such as lydia.delgado@enron.com and sally.beck@enron.com were identified, which helped connect activity level to engagement trends later in the analysis.

4. Employee Scoring and Ranking

Each message was scored using:

- **+1** for Positive
- **–1** for Negative
- **0** for Neutral

Scores were aggregated **monthly per employee**, and then used to create two lists for each month:

- **Top 3 Positive Employees**
- **Top 3 Negative Employees**

Sample Rankings for January 2010:

Top Positive:

- eric.bass@enron.com (score: 3)
- patti.thompson@enron.com (score: 3)
- don.baughman@enron.com (score: 2)

Top Negative:

- sally.beck@enron.com (score: –1)
- bobette.riner@ipgdirect.com (score: 0)
- john.arnold@enron.com (score: 0)

5. Flight Risk Identification

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

Flight Risk Employees Identified:

- sally.beck@enron.com
- don.baughman@enron.com
- john.arnold@enron.com
- bobette.riner@ipgdirect.com

These individuals may require follow-up from HR for engagement or retention initiatives.

6. Predictive Modeling – Sentiment Score Forecasting

First Test:

A basic linear regression model was trained using only:

- `message_count`
- `avg_message_length`
- `avg_word_count`

Results:

- R^2: **0.21**
- MSE: **~3.18**

→ The model showed weak predictive power with only structural message features.

Final Model:

We enhanced the model with behavior-based features:

- `num_positive_msgs`
- `num_negative_msgs`
- `std_message_length`
- `negative_msg_ratio`
- `prior_month_score`

Results:

- R^2: **1.0**
- MSE: **~0**

→ The model perfectly predicted the monthly sentiment score using positive/negative message counts.
Feature analysis confirmed this: `num_positive_msgs` had a coefficient of **+1.0**, and `num_negative_msgs` had **–1.0**, while other features had negligible impact.

The comparison between the two models clearly shows the **importance of sentiment-driven features** in understanding communication patterns, versus relying solely on message structure.

Conclusion

This analysis used modern NLP techniques and statistical modeling to extract sentiment and behavioral insights from employee messages. By comparing sentiment labeling methods specifically RoBERTa and VADER we found that RoBERTa provided more accurate sentiment detection for professional communication.

Our exploratory data analysis revealed significant trends in sentiment distribution, messaging behavior, and employee engagement. The scoring system enabled effective monitoring of sentiment over time, while our flight risk identification method flagged individuals needing HR attention.

Finally, the predictive modeling showed that sentiment-aware features could explain monthly sentiment scores, enhancing the transparency of our approach. Overall, these insights provide HR teams with tools to identify top performers, disengaged employees, and those at risk in a fully automated and explainable manner.