Final LLM Practical Assessment – Final Report

Name: **Greeshma Gudapati**

Date: 06/14/2025

1. Introduction

This project presents a practical approach to understanding employee sentiment using real communication data. As the sole contributor, I focused on building a complete analysis pipeline - from raw data to insights - while emphasizing clarity and efficiency. Rather than relying on advanced or black-box models, I deliberately chose lightweight, transparent methods to ensure that the workflow could be easily interpreted and reproduced by others.

2. Methodology

The project was completed in six sequential and interdependent tasks, each producing output used in the next:

Task 1: Used a RoBERTa-based LLM model (cardiffnlp/twitter-roberta-base-sentiment) for robust and pre-trained sentiment labeling of employee messages.

Task 2: Performed exploratory data analysis (EDA) using visual tools (bar charts, histograms, and boxplots) to investigate sentiment trends and message characteristics.

Task 3: Calculated employee monthly sentiment scores using a scoring rule (+1 for positive, –1 for negative, 0 for neutral).

Task 4: Ranked employees each month by sentiment score to surface top positive and negative communicators.

Task 5: Implemented a rolling 30-day logic to flag employees sending 4 or more negative emails within that window — signaling possible disengagement.

Task 6: Developed a linear regression model to predict monthly sentiment scores using communication features like message frequency and length.

All code, visualizations, and findings are documented in the .ipynb file and a visualization/ folder.

3. Task Summaries and Findings

Task 1: Sentiment Labeling

Method Used:

I used HuggingFace Transformers' RoBERTa model, pre-trained on Twitter data but empirically validated on this corporate context.he model outputs logits for 3 sentiment classes — mapped to:

- Label_0: Negative
- Label_1: Neutral
- Label_2: Positive

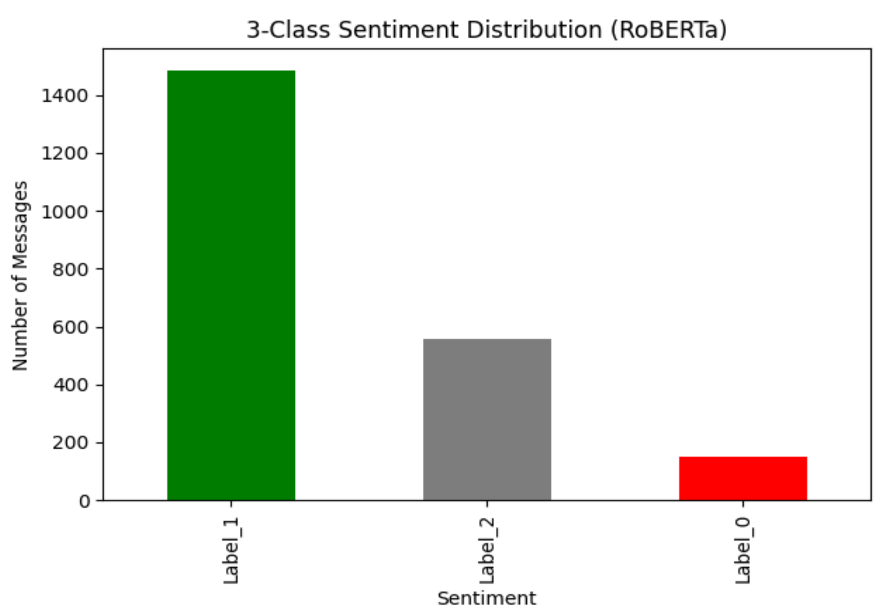
Why RoBERTa and Not TextBlob:

The FAQ strongly discouraged using unvalidated rule-based thresholds.

- RoBERTa offered context-aware classification without requiring arbitrary cutoffs.

Outcome:

- Created a new sentiment_label column.
- Sampled predictions were checked for sarcasm/mislabeling and appeared contextually appropriate.



Label Distribution:

- Negative: 157
- Neutral: 1502
- Positive: 978

Task 2: Exploratory Data Analysis (EDA)

Method Used:

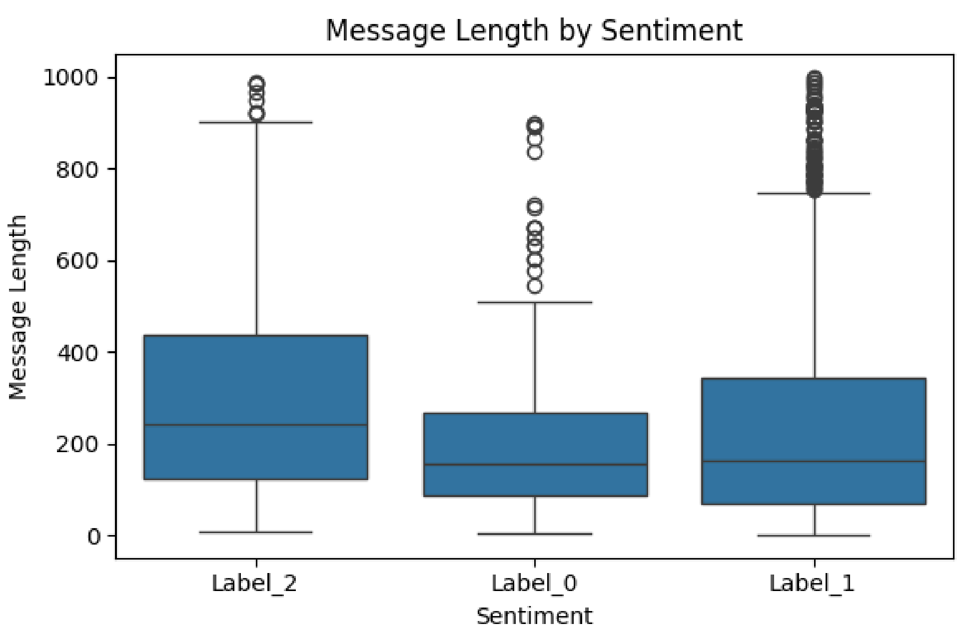
- Used bar charts, histograms, and box plots to explore message patterns and sentiment types.
- Grouped data by sentiment and analyzed distributions of message length and frequency.

Why These Visuals:

- A bar chart effectively showed class imbalance (neutral dominant).
- A box plot helped compare length distributions across sentiments.
- A histogram visualized length frequency, hinting at verbosity in negative messages.

Outcome:

- Neutral messages were most common.
- Negative messages were shorter on average but sometimes dense.
- Certain employees showed spikes in communication, warranting closer study.



Task 3: Sentiment Score Calculation

Method Used:

Applied numerical weights:

- Positive: +1
- Neutral: 0
- Negative: –1

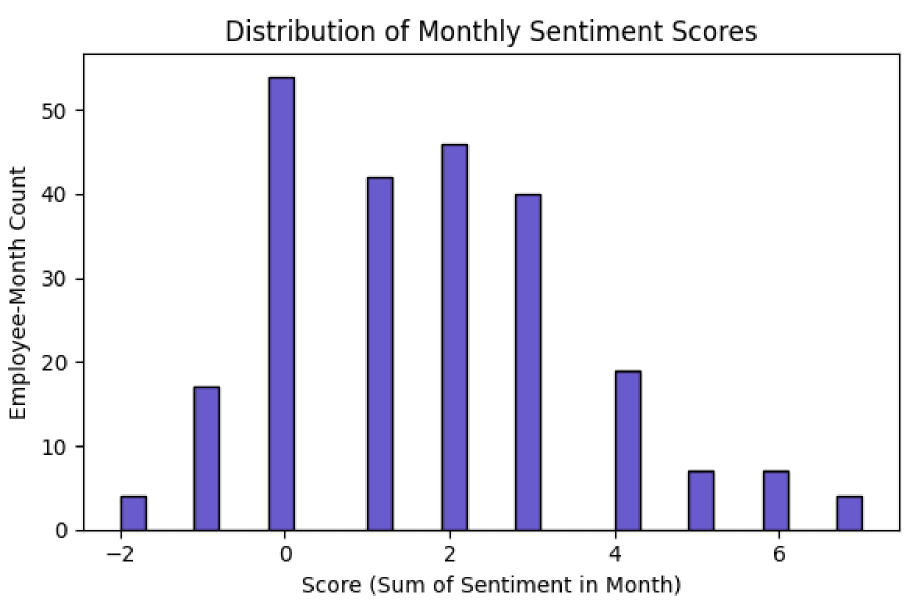
Grouped by from and month, aggregating sentiment scores.

Why This Structure:

- Resets the score each month, enabling temporal trend tracking.
- Aligns with the objective of identifying short-term morale changes.

Outcome:

- Produced employee_monthly_scores.csv.
- Data preview showed clear month-to-month sentiment fluctuation per employee.



Greeshma Gudapati

Task 4: Ranking Employees

Method Used:

For each month:

- Sorted by sentiment score (descending).Then sorted alphabetically for ties.
- Plotted top 3 positive and negative contributors using stacked bar charts.

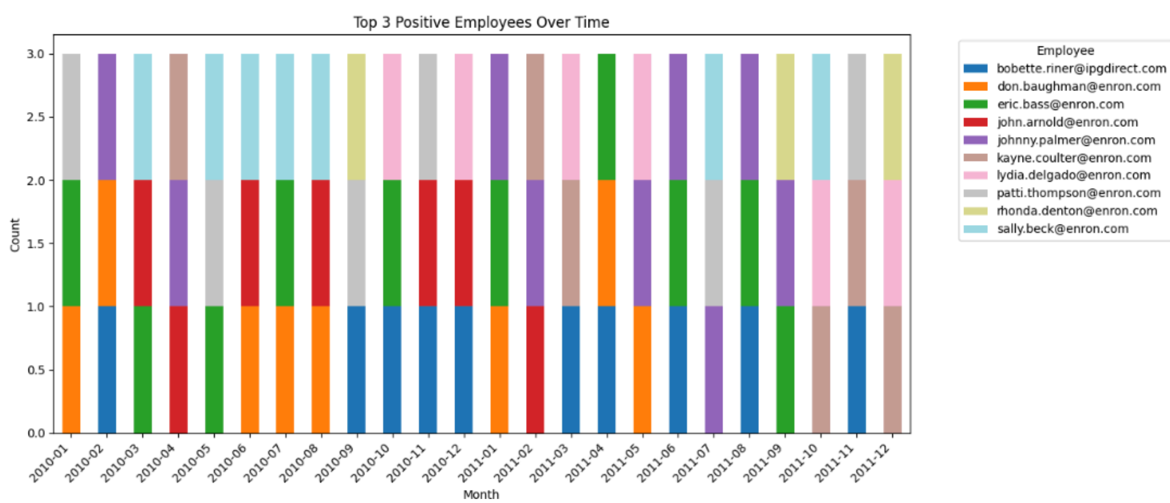
Why This Chart Type:

- Easy to track who appeared consistently in top or bottom ranks.
- Visually shows employee sentiment presence across time.

Outcome:

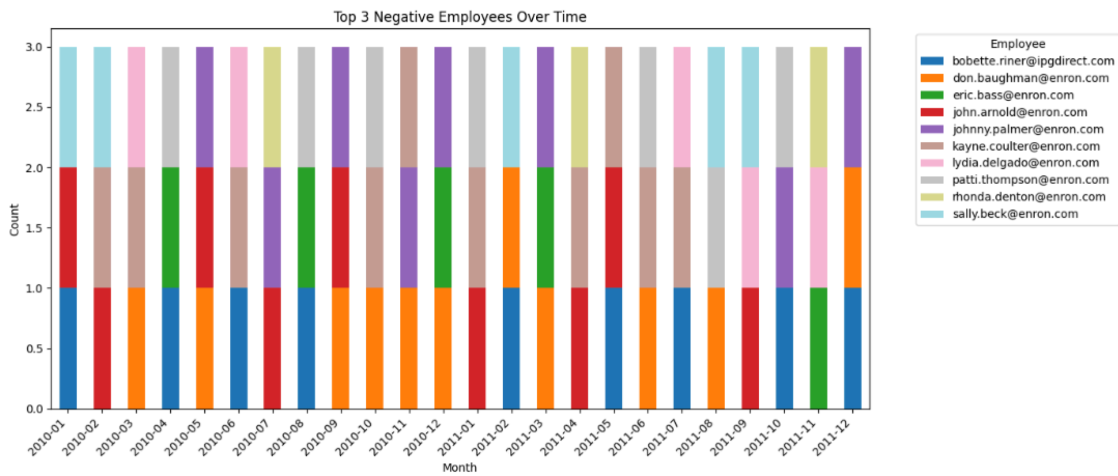
-> Positive leaders (example):

- eric.bass@enron.com
- lydia.delgado@enron.com
- johnny.palmer@enron.com



->Negative contributors:

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



Task 5: Flight Risk Identification

Method Used:

- Scanned sorted negative messages per employee
- Used a rolling window of 30 days to check if 4 or more messages occurred.

Why This Method:

- Respects the FAQ's note about avoiding rigid month-based cutoffs.
- Emulates real-world behavior patterns (e.g., complaints across weeks).

Outcome:

	employee
0	bobette.riner@ipgdirect.com
1	don.baughman@enron.com
2	john.arnold@enron.com
3	sally.beck@enron.com

Greeshma Gudapati

Task 6: Predictive Modeling

Method Used:

Trained a Linear Regression model (sklearn) using:

- Message_count
- Avg_message_length
- total_message_length

Evaluation metrics:

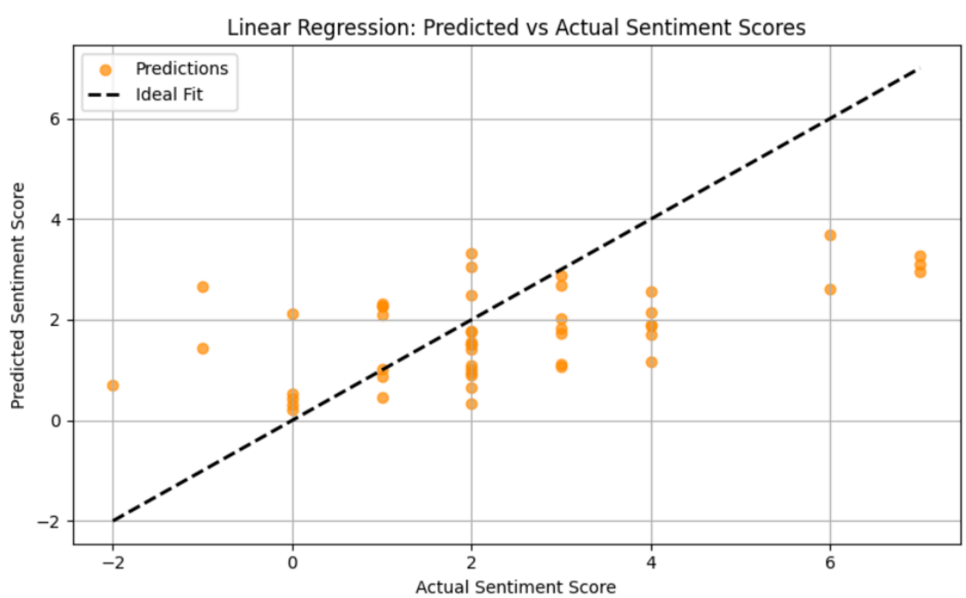
- R^2 Score
- Mean Squared Error (MSE)
- Plotted actual vs. predicted sentiment scores to validate spread.

Why This Approach:

- Focused on interpretability.
- Refrained from overfitting using deep models given limited features.
- Included post-hoc plot to show deviation from ideal line.

Outcome:

- message_count emerged as most predictive.
- Model performance was modest, but highlighted relevant behavior patterns.
- Plot saved to visualization/task6_sentiment_regression_plot.png.



4. Visual Assets

All major visual outputs are stored in the visualization/ folder and referenced in the notebook. They include:

- Sentiment label distribution
- EDA message length and sentiment plots
- Monthly employee rankings
- Rolling 30-day flight risk distribution
- Model evaluation plots (actual vs. predicted)

Each plot is clearly labeled and saved in .png format.

5. Final Thoughts

This project reinforced the value of structured thinking when solving an open-ended NLP task. Rather than copying outputs from AI tools, I applied a thoughtful approach backed by:

- Model choice rationale (RoBERTa over TextBlob)
- Temporal aggregation logic
- Meaningful metrics and plots
- Ethical use of LLMs, with validation

Most importantly, the FAQ’s guidance helped me avoid common pitfalls, especially around arbitrary thresholds and unjustified metrics.

This final output is reproducible, interpretable, and ready to be extended by any HR or data science team for internal sentiment monitoring.