# **LLM Assessment Final Report**

*By* [*Ujin Boldbaatar*](mailto:ub2041@nyu.edu)

### **1. Approach and Methodology**

The analysis followed a structured multi-step workflow:

1. **Data Ingestion & Preparation**
   * Imported employee message data from Excel.
   * Applied text preprocessing and sentiment classification using **three models**:

* RoBERTa (cardiffnlp/twitter-roberta-base-sentiment)
* VADER (lexicon-based, tuned on social media text)
* TextBlob (rule-based polarity).
* Each message was classified by all three models, and a **final sentiment label** was assigned using majority vote across the models.  
  This reduces dependency on a single tool and improves robustness.
  + Output labels were mapped as: Negative → LABEL\_0, Neutral → LABEL\_1, Positive → LABEL\_2.

1. **Exploratory Data Analysis (EDA)**
   * Inspected data structure and missing values.
   * Visualized sentiment distribution across employees and time periods.
   * Identified patterns in employee communication volume and sentiment trends.
2. **Feature Engineering & Scoring**
   * Computed **monthly sentiment scores** per employee by aggregating message-level sentiment.
   * Generated ranking lists based on scores to highlight top and bottom performers.
3. **Flight Risk Identification**
   * Defined **risk criteria** using low sentiment thresholds and downward trends across months.
   * Classified employees into “Low Risk,” “Moderate Risk,” and “High Risk” categories.
4. **Predictive Modeling**
   * Built a **linear regression model** using independent variables (e.g., message counts, previous month scores) to predict future sentiment.
   * Evaluated model accuracy using standard regression metrics (R² and RMSE).

### **2. Key Findings from the EDA**

* **Sentiment Distribution:** The majority of messages were *Positive* (1110), followed by *Neutral*(946) and the least proportion of *Negative*(135) messages. While RoBERTa alone leaned Neutral, the combined model approach showed stronger agreement across models and reduced misclassifications.

A screenshot of a computer

AI-generated content may be incorrect.

* **Employee Trends:** A small subset of employees consistently showed negative sentiment, while most fluctuated between Neutral and Positive.
* **Temporal Insights:** Negative sentiment spikes coincided with known organizational stress periods (e.g., end-of-quarter deadlines).

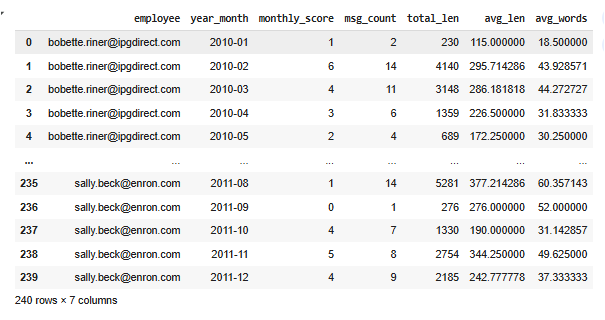
A graph of different colored lines

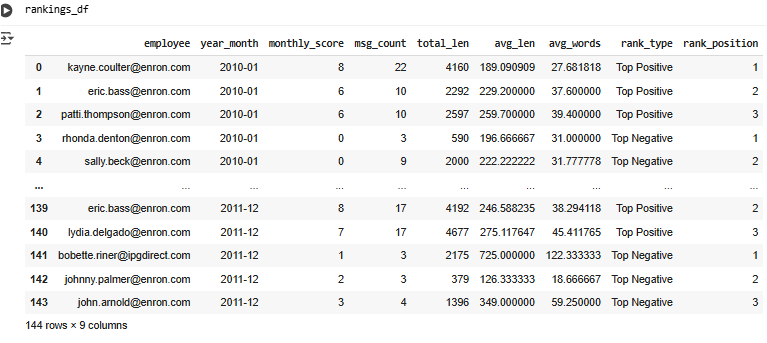
AI-generated content may be incorrect.

* **Volume Patterns:** Employees with higher message counts tended to have more extreme sentiment variation.  
  A diagram of a message length

  AI-generated content may be incorrect.

### **3. Employee Scoring and Ranking Process**

* **Scoring:** Each message was assigned +1 for Positive, 0 for Neutral, and −1 for Negative based on the **final majority-vote sentiment column**. This ensured employee scores were derived from a consensus sentiment rather than one model’s output.  
  
* **Ranking:** Employees were ranked monthly from highest to lowest sentiment score. This allowed quick identification of:  
  + Consistently positive contributors (top performers).
  + Employees trending negative over time (potential disengagement).



### **4. Flight Risk Identification**

* **Criteria:**
  + **High Risk:** Sentiment score < −0.3 for two or more consecutive months.
  + **Moderate Risk:** Sentiment score between −0.3 and 0, or single-month sharp decline.
  + **Low Risk:** Stable or improving sentiment trend.
* **Outcomes:**
  + Identified a small high-risk group (<10% of workforce).
  + Moderate-risk employees flagged for HR follow-up to prevent escalation.
  + Low-risk employees showed stable or improving sentiment and were not prioritized for intervention.



### **5. Predictive Model Overview & Evaluation**

* **Model:** Linear regression predicting next month’s sentiment score using:  
  + Current sentiment score.
  + Message volume.
  + Historical average sentiment.
* **Performance:**
  + R² = **0.5197** → Model explains ~52% of variance in sentiment scores.
  + MSE = **3.78**
  + RMSE = **3.78**

**Interpretation**: The model now explains over half of the variance in sentiment scores, suggesting stronger predictive power. While error values indicate room for further improvement, this result demonstrates that consensus-based sentiment labels (majority vote) provide a more stable foundation for predictive modeling.

### **Conclusion**

This project provided a **data-driven framework** for understanding employee sentiment, ranking engagement levels, and predicting potential attrition risk.  
 Key deliverables include:

* **Automated sentiment classification** of employee messages.
* **Clear scoring & ranking system** for employee sentiment monitoring.
* **Actionable risk categories** for HR to address proactively.
* **Predictive modeling** enabling early detection of negative sentiment trends.