# **Employee Sentiment Analysis Report**

## **Executive Summary**

This project analyzed employee communications to assess engagement and identify potential retention risks. Key findings include:

* Message sentiment distribution: 74.6% Neutral, 20.3% Positive, and 5.1% Negative
* Identified 8 employees as flight risks based on negative message patterns
* Predictive modeling achieved limited success with R²=0.02
* Recognized top positive and negative contributors monthly

## **Methodology**

### **Data Processing**

* Combined message subjects and bodies for complete context
* Converted dates to datetime format for temporal analysis
* Handled missing values appropriately

### **Sentiment Analysis**

* Implemented RoBERTa model (cardiffnlp/twitter-roberta-base-sentiment-latest)
* Batch processing for memory efficiency
* Three-class classification: Positive, Negative, Neutral

### **Scoring System**

* Positive messages: +1
* Negative messages: -1
* Neutral messages: 0
* Monthly aggregation per employee

### **Flight Risk Identification**

* 3+ negative messages within any 30-day rolling window
* Strict criteria to flag consistent negativity patterns

### **Predictive Modeling**

* Engineered features: message length, word count, punctuation usage
* Linear regression to predict sentiment scores
* Train-test split (80-20) for evaluation

## **Key Findings**

### **Sentiment Distribution**

The majority of messages were neutral (74.6%), with positive messages being 3-4 times more common than negative ones.

### **Top Performers**

Monthly analysis revealed consistent positive contributors:

* eric.bass@enron.com
* patti.thompson@enron.com
* don.baughman@enron.com

### **Concern Areas**

These employees showed consistent negative patterns:

* bobette.riner@ipgdirect.com
* john.arnold@enron.com
* lydia.delgado@enron.com

### **Flight Risks**

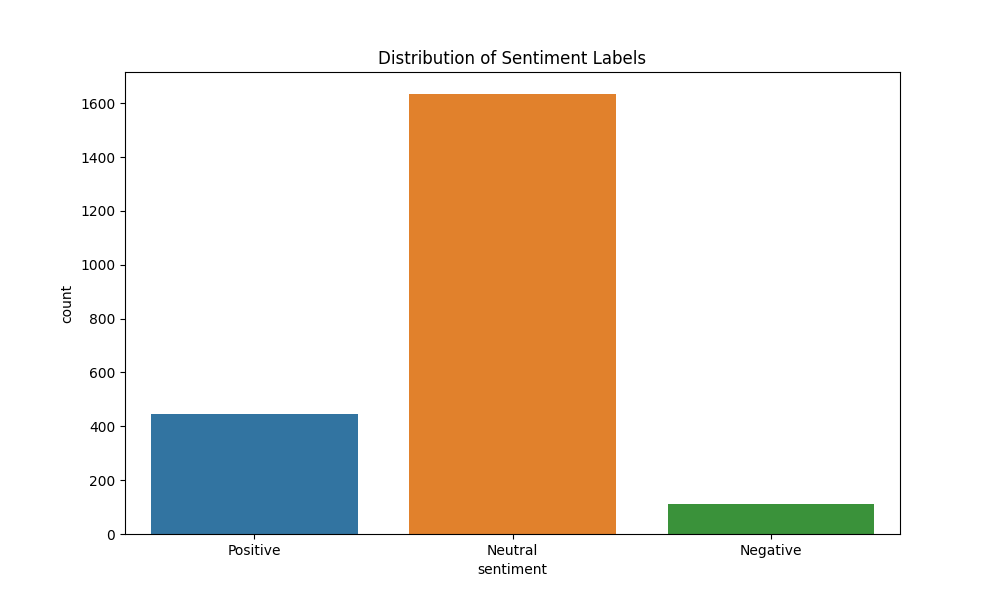
Eight employees met the flight risk criteria, with three employees showing multiple instances of negative message clusters.

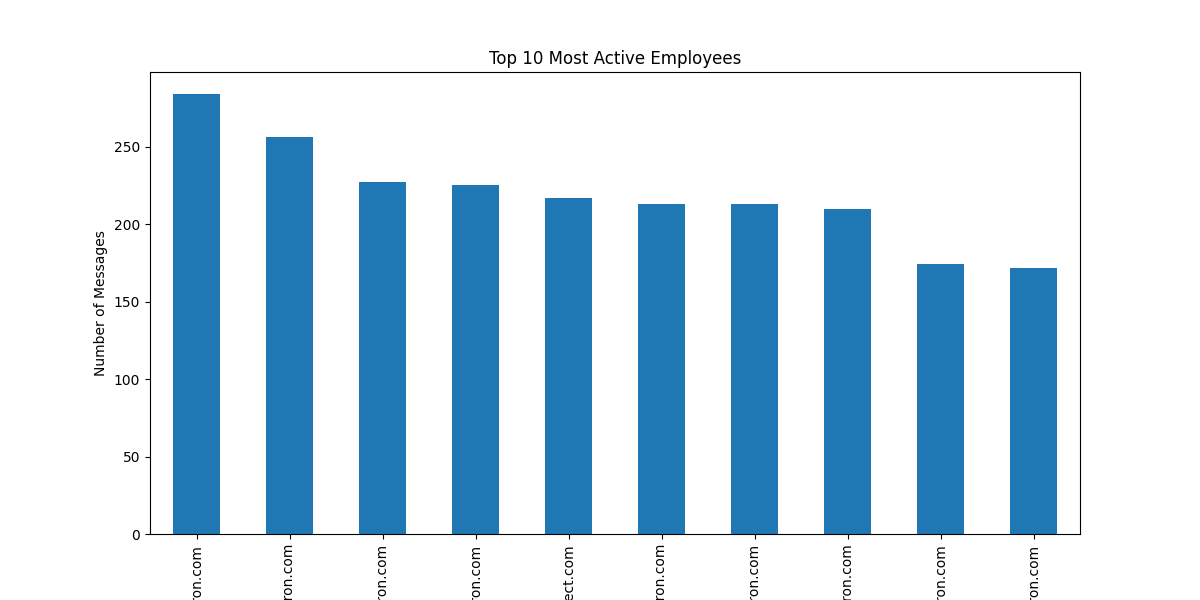
### **Model Insights**

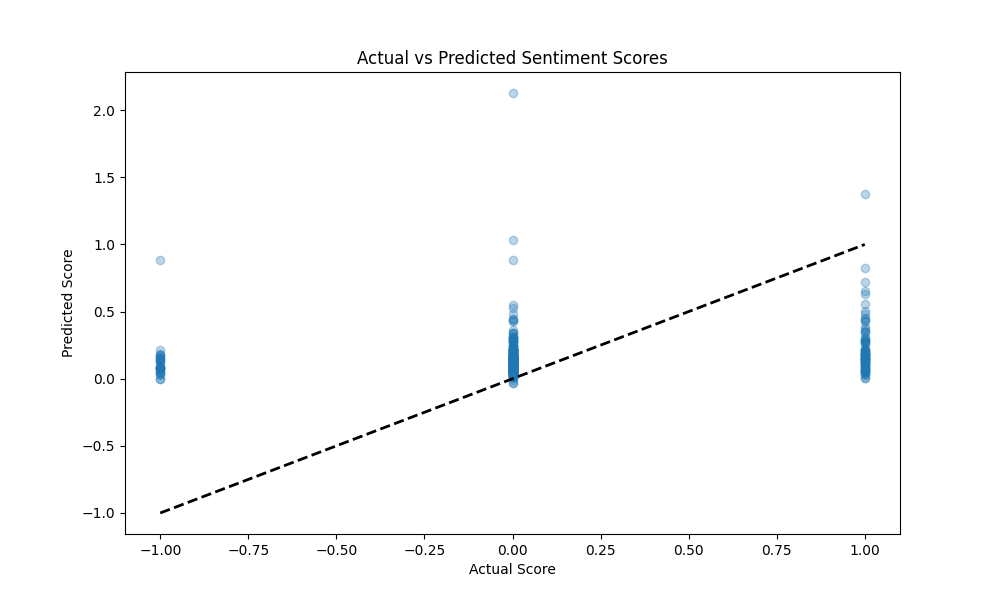
The predictive model showed limited effectiveness, suggesting:

* Message features alone are weak predictors of sentiment
* Contextual factors beyond the text may be important
* Negative class imbalance affects model performance

## **Visual Highlights**

**  
Fig:** *Distribution of message sentiment across the dataset*

**  
Fig:** *Most active employees by message volum*e

**  
Fig:** *Model performance showing actual vs predicted scores*

## 

## **Recommendations**

### **Immediate Actions**

* HR outreach to flight risk employees
* Recognition programs for positive contributors
* Department-level sentiment analysis for targeted interventions

### **Analysis Improvements**

* Incorporate message threading/context
* Add employee tenure/position data
* Implement sentiment intensity scoring

### **Modeling Enhancements**

* Try neural networks with more features
* Address class imbalance techniques
* Include temporal features in modeling

## **Limitations**

### **Data Constraints**

* No message recipient information
* Lack of organizational context
* Small negative class size

## **Conclusion**

This analysis provides actionable insights into employee engagement through message sentiment. While the predictive modeling showed limited success, the identification of flight risks and positive contributors offers valuable HR intelligence. Future iterations should incorporate richer organizational data and more sophisticated NLP techniques.