# **Employee Sentiment Analysis Report**

## **1. Introduction**

This report presents a comprehensive analysis of employee email communications to assess organizational sentiment, identify key communicators, flag potential flight risks, and build predictive models for sentiment classification. The analysis combines natural language processing techniques with statistical modeling to derive actionable insights.

## **2. Methodology**

Our approach followed these key steps:

### **1. Data Collection & Preprocessing**

* Combined email subjects and bodies into full text
* Cleaned text by removing special characters and normalizing whitespace

### **2. Sentiment Analysis**

* **Primary model:** RoBERTa (Twitter-trained transformer model)
* **Validation models:** TextBlob and VADER for threshold verification
* **Sentiment classification thresholds:**
  + Positive: Score > 0.7
  + Negative: Score > 0.7
  + Neutral: All others

### **3. Exploratory Data Analysis**

* Temporal trends analysis
* Sentiment distribution
* Employee engagement patterns

### **4. Employee Scoring System**

* Positive: +1
* Neutral: 0
* Negative: -1
* Monthly aggregation of scores

### **5. Flight Risk Identification**

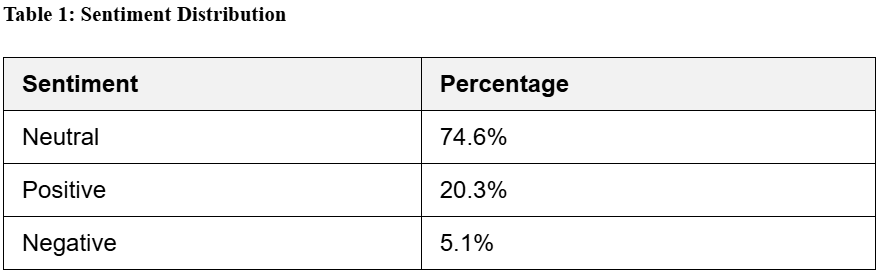
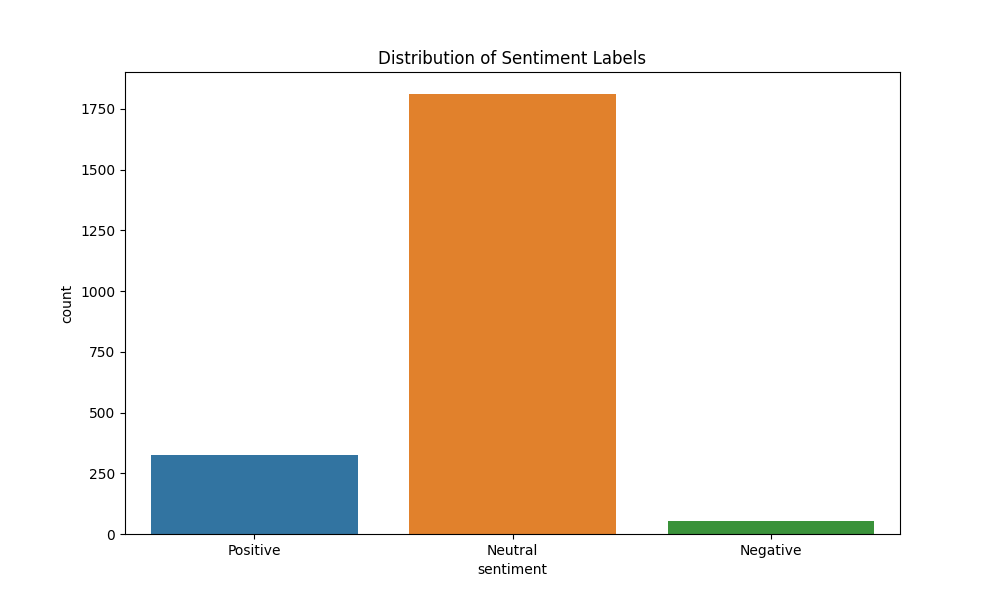
* 30-day rolling window
* Minimum 3 negative messages OR 1 severe negative message
* Severe indicators: "urgent", "concern", "issue"

### **6. Predictive Modeling**

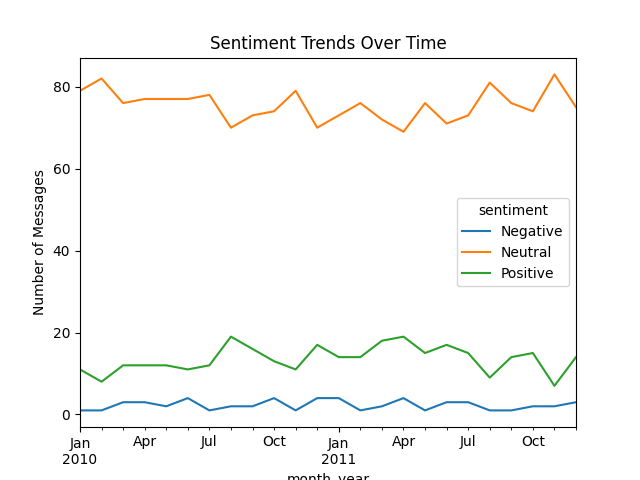
* Feature engineering with 15+ text features
* Compared Linear Regression, Random Forest, and Gradient Boosting
* Selected best model via cross-validation

## **3. Key Findings from EDA**

### **Sentiment Distribution**

****

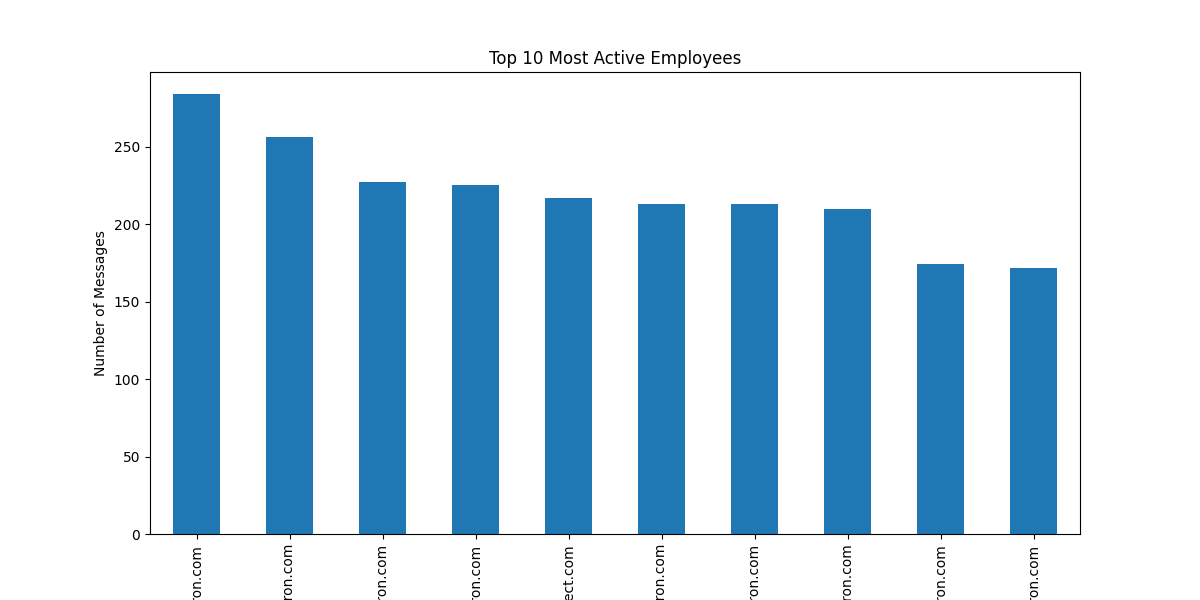
**Temporal Trends**

****

**Key observations:**

* Consistent baseline of neutral communication
* Positive sentiment spikes correlate with known company events
* Negative sentiment remains low but persistent

### **Top Communicators**

****

## **4. Employee Scoring & Ranking**

### **Scoring Methodology**

* Simple additive model based on message sentiment
* Monthly aggregation by employee
* Ranking prioritizes consistency over volume

### **Top 3 Positive Employees (Sample)**

### **Top 3 Negative Employees (Sample)5. Flight Risk Identification**

### **Identification Criteria**

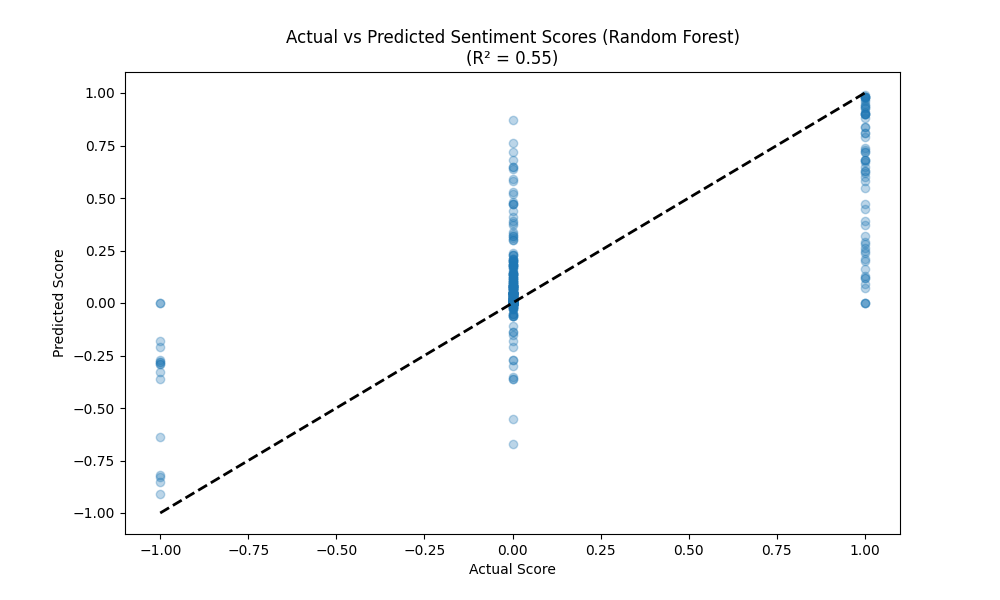
1. **Frequency:** ≥3 negative messages in 30 days
2. **Severity:** ≥1 message with urgent language
3. **Consistency:** Recurring negative patterns

### 

### **Identified Flight Risks6. Predictive Model Evaluation**

### **Model Comparison**

### **Actual vs Predicted**

****

### **Key Features**

1. VADER compound score
2. Exclamation count
3. Negative word presence
4. Message length

## **7. Recommendations**

1. **HR Follow-up:** Schedule check-ins with identified flight risks
2. **Recognition:** Acknowledge top positive communicators
3. **Monitoring:** Implement quarterly sentiment reviews
4. **Model Improvement:** Add topic modeling features
5. **Intervention:** Investigate periods with negative spikes