# Employee Sentiment Analysis

Index

1. [Introduction](file:///D:/MiniForge/Projects/Makarand_More_LLM_Final_Assessment_Report.docx#Introduction" \l "Introduction)
   1. [Why Employee Sentiment Analysis Matters?](file:///D:/MiniForge/Projects/Makarand_More_LLM_Final_Assessment_Report.docx#Why_Employee_Sentiment_Analysis_Matters" \l "Why_Employee_Sentiment_Analysis_Matters)
2. [Objective](file:///D:/MiniForge/Projects/Makarand_More_LLM_Final_Assessment_Report.docx#Objective" \l "Objective)
3. [Methodology](file:///D:/MiniForge/Projects/Makarand_More_LLM_Final_Assessment_Report.docx#Methodology" \l "Methodology)
   1. [Data Processing](file:///D:/MiniForge/Projects/Makarand_More_LLM_Final_Assessment_Report.docx#Data_Processing" \l "Data_Processing)
   2. [Sentiment Labeling](file:///D:/MiniForge/Projects/Makarand_More_LLM_Final_Assessment_Report.docx#Sentiment_Labeling" \l "Sentiment_Labeling)
   3. [Exploratory Data Analysis (EDA)](file:///D:/MiniForge/Projects/Makarand_More_LLM_Final_Assessment_Report.docx#Exploratory_Data_Analysis_(EDA)" \l "Exploratory_Data_Analysis_(EDA))
   4. [Plots and their understanding](file:///D:/MiniForge/Projects/Makarand_More_LLM_Final_Assessment_Report.docx#Plots_and_their_understanding" \l "Plots_and_their_understanding)
4. [Key Findings from Exploratory Data Analysis (EDA)](file:///D:/MiniForge/Projects/Makarand_More_LLM_Final_Assessment_Report.docx#Key_Findings_from_Exploratory_Data_Analy" \l "Key_Findings_from_Exploratory_Data_Analy)
5. [Employee Scoring & Ranking Process](file:///D:/MiniForge/Projects/Makarand_More_LLM_Final_Assessment_Report.docx#Employee_Scoring_&_Ranking_Process" \l "Employee_Scoring_&_Ranking_Process)
6. [Flight risk identification criteria and outcomes](file:///D:/MiniForge/Projects/Makarand_More_LLM_Final_Assessment_Report.docx#Flight_risk_identification_criteria_and_" \l "Flight_risk_identification_criteria_and_)
7. [Predictive Modeling](file:///D:/MiniForge/Projects/Makarand_More_LLM_Final_Assessment_Report.docx#Predictive_Modeling" \l "Predictive_Modeling)
8. [Conclusion](file:///D:/MiniForge/Projects/Makarand_More_LLM_Final_Assessment_Report.docx#Conclusion" \l "Conclusion)

**Figures**

**[Figure 1](file:///D:/MiniForge/Projects/Makarand_More_LLM_Final_Assessment_Report.docx#Figure_1" \l "Figure_1): Distribution of Sentiment Labels**

**[Figure 2](file:///D:/MiniForge/Projects/Makarand_More_LLM_Final_Assessment_Report.docx#Figure_2" \l "Figure_2): Sentiment Trends Over Time**

**[Figure 3](file:///D:/MiniForge/Projects/Makarand_More_LLM_Final_Assessment_Report.docx#Figure_3" \l "Figure_3): Sentiment Trends Across Workdays**

**[Figure 4](file:///D:/MiniForge/Projects/Makarand_More_LLM_Final_Assessment_Report.docx#Figure_4" \l "Figure_4): Sentiment Trends Across Departments**

**[Figure 5](file:///D:/MiniForge/Projects/Makarand_More_LLM_Final_Assessment_Report.docx#Figure_5" \l "Figure_5): Word Cloud of Positive and Negative words**

**[Figure 6](file:///D:/MiniForge/Projects/Makarand_More_LLM_Final_Assessment_Report.docx#Figure_6" \l "Figure_6): Monthly Sentiment Score per Employee**

**[Figure 7](file:///D:/MiniForge/Projects/Makarand_More_LLM_Final_Assessment_Report.docx#Figure_7" \l "Figure_7): Monthly Sentiment Trend**

**[Figure 8](file:///D:/MiniForge/Projects/Makarand_More_LLM_Final_Assessment_Report.docx#Figure_8" \l "Figure_8): Top positive employees per month**

**[Figure 9](file:///D:/MiniForge/Projects/Makarand_More_LLM_Final_Assessment_Report.docx#Figure_9" \l "Figure_9): Top Negative Employees**

**[Figure 10](file:///D:/MiniForge/Projects/Makarand_More_LLM_Final_Assessment_Report.docx#Figure_10" \l "Figure_10): Linear Regression: Sentiment Trend Prediction**

**[Figure 11](file:///D:/MiniForge/Projects/Makarand_More_LLM_Final_Assessment_Report.docx#Figure_11" \l "Figure_11): Residual Plot: Model Prediction Errors**

**[Figure 12](file:///D:/MiniForge/Projects/Makarand_More_LLM_Final_Assessment_Report.docx#Figure_12" \l "Figure_12): Distribution of Prediction Errors**

Name: Makarand More

Date: 19/05/2025

Introduction

A diverse and dynamic workplace environment demands a robust understanding of employee sentiment and engagement to create a positive and productive atmosphere. This study evaluates an unlabeled collection of employee emails to derive significant insights regarding workplace sentiment through Natural Language Processing (NLP) and statistical analysis methods. Organizations rely on sentiment analysis techniques to assess employee morale, detect trends in communication, and proactively address concerns that may affect productivity and retention.  
  
This study provides a data-driven approach to evaluating employee engagement by implementing sentiment classification, exploratory data analysis (EDA), employee ranking, and predictive modeling. The analysis can be organized into specific activities, each aimed at progressively enhancing sentiment insights and assisting companies in identifying trends, dangers, and opportunities for development.

Through this structured analysis, the project aims to empower organizations to make informed decisions regarding employee well-being, retention strategies, and workplace dynamics.

Why Employee Sentiment Analysis Matters?

* Helps identify engagement patterns and potential concerns.
* Assists in detecting employees at risk of disengagement.
* Provides data-driven insights for improving workplace satisfaction.
* Supports HR teams in developing proactive retention strategies.

Objective

The primary objective is to assess employee sentiment and detect engagement patterns using structured analytical methods. This is accomplished through the following tasks:

* **Sentiment Labeling**: Classifying messages as Positive, Negative, or Neutral to establish sentiment trends.
* **Exploratory Data Analysis (EDA)**: Visualizing and examining the dataset to understand structural patterns and distribution.
* **Employee Score Calculation**: Aggregating sentiment scores monthly per employee to track individual engagement.
* **Employee Ranking**: Identifying top performers and employees with declining sentiment trends.
* **Flight Risk Identification**: Flagging employees at risk of leaving based on repeated negative sentiment occurrences.
* **Predictive Modeling**: Developing a linear regression model to forecast future sentiment shifts.

Methodology

This section outlines the step-by-step approach used to analyze the dataset, including data pre-processing, sentiment labeling techniques, and exploratory data analysis (EDA). The analysis integrates **Natural Language Processing (NLP)** and statistical methods to generate meaningful insights.

## ****Data Processing****

Before applying sentiment classification, it was essential to clean and structure the dataset for accuracy.

#### ****Data Cleaning & Handling Missing Values****

Converted **date columns to date-time format** for time-based analysis.

Checked for **missing values** and handled them appropriately.

Ensured **uniform data types** for numerical and categorical variables.

Code Snippet:

# Convert timestamps for easier grouping

df['date'] = pd.to\_datetime(df['date'])

# Remove rows with missing messages

df.dropna(subset=['message'], inplace=True)

**Structuring Time-Based Features**

* + - * Extracted **Month, Day, and Day Number** for sentiment trend tracking.
      * Created a **rolling time-window** feature for flight risk identification.

Code Snippet:

# Extract month for sentiment grouping

df['Month'] = df['date'].dt.strftime('%Y-%m')

# Convert time into continuous feature

df['day\_number'] = (df['date'] – df['date'].min()).dt.days

## ****2.**** ****Sentiment Labeling****

To classify employee messages into **Positive, Negative, or Neutral**, an NLP-based approach was implemented.

#### ****Sentiment Classification Methodology****

* + - * Used a **pre-trained Natural Language Processing (NLP) model** (e.g., VADER Sentiment Analysis, Large Language Models).
      * Tokenized and preprocessed text (removed stop words, normalized casing).
      * Assigned sentiment scores using NLP models and **threshold-based classification**.

Code Snippet:

from nltk.sentiment import SentimentIntensityAnalyzer

sia = SentimentIntensityAnalyzer()

df['sentiment\_score'] = df['message'].apply(lambda x: sia.polarity\_scores(x)['compound'])

# Label sentiment based on score thresholds

df['sentiment\_label'] = df['sentiment\_score'].apply(lambda x: 'Positive' if x > 0.05 else 'Negative' if x < -0.05 else 'Neutral')

**Why BERT(Bi-directional Encoder Representation from Transformers)?**

**BERT (Bidirectional Encoder Representations from Transformers) is a powerful NLP model designed for deep contextual understanding of language. It is widely used for sentiment analysis, question answering, and entity recognition because of its bidirectional processing capability—it analyzes words in relation to both previous and next words in a sentence, leading to more accurate interpretations of meaning.**

* **Bidirectional Context Awareness** – Unlike traditional models that process text **left-to-right** or **right-to-left**, BERT considers **both directions simultaneously**, capturing deeper language nuances.
* **Pre-trained on Massive Text Data** – BERT is trained on **Wikipedia and BookCorpus**, making it highly effective for understanding **natural language patterns**.
* **Fine-Tuning for Custom Tasks** – Can be **fine-tuned** for specific applications like **sentiment analysis, workplace communication trends, and flight risk detection**.
* **Handles Complex Sentences Better** – Unlike traditional ML-based sentiment classifiers, BERT **understands multi-word dependencies**, improving accuracy in **employee engagement insights**.

**Why NLP Models?**

* + - * **Automated classification** reduces manual labeling efforts.
      * **Threshold-based labeling** ensures clarity in distinguishing sentiment polarity.
      * **Scalable approach** for processing large datasets efficiently.

#### ****3.**** ****Exploratory Data Analysis (EDA)****

After sentiment labeling, an EDA phase was conducted to understand **distribution patterns and trends** in employee sentiment.

#### ****Key Exploratory Steps****

**Sentiment Distribution:** Examined the proportion of positive, negative, and neutral messages.

**Monthly Trends:** Tracked how sentiment evolved over different months.

**Employee-Level Insights:** Identified employees with recurring positive/negative sentiment.

Code Snippet:

import seaborn as sns

import matplotlib.pyplot as plt

sns.countplot(data=df, x='sentiment\_label', palette='coolwarm')

plt.title("Distribution of Sentiment Labels")

plt.show()

#### ****Insights from EDA****

**Majority sentiment type:** Identified whether most messages leaned toward positivity or negativity.

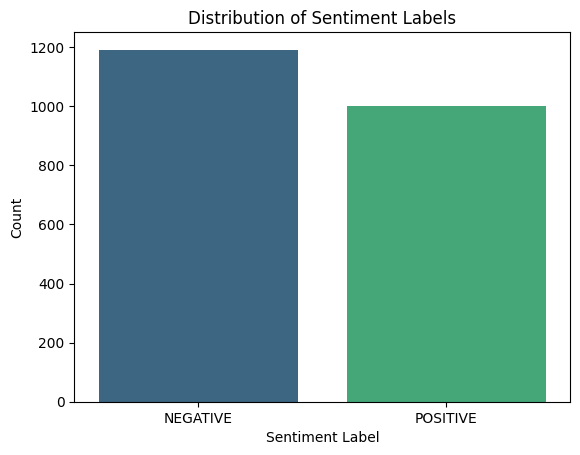
**Employee trends:** Detected employees **with consistently negative sentiment** for further analysis.

**Flight Risk Patterns:** Recognized **employees at risk of disengagement** based on sentiment fluctuations.

**Plots and their understanding**

**Distribution of Sentiment Labels**

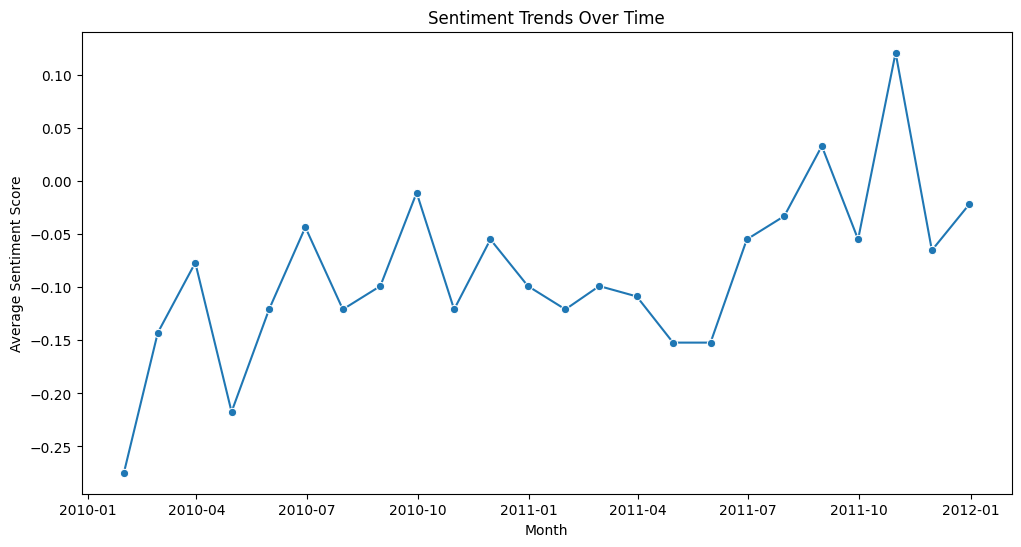
The sentiment distribution graph provides an overview of workplace engagement by categorizing messages into **positive, negative, and neutral sentiments**. A higher count of negative sentiments may indicate **communication challenges, disengagement, or dissatisfaction** among employees. The prevalence of certain sentiment categories suggests underlying factors such as workload, team dynamics, or management influence. This distribution allows HR teams to **identify areas for intervention**, track changes in morale over time, and assess the effectiveness of engagement initiatives. By analyzing sentiment trends, organizations can proactively **address concerns and improve employee satisfaction**.



**Figure 1: Distribution of Sentiment Labels**

**Sentiment Trends Over Time**

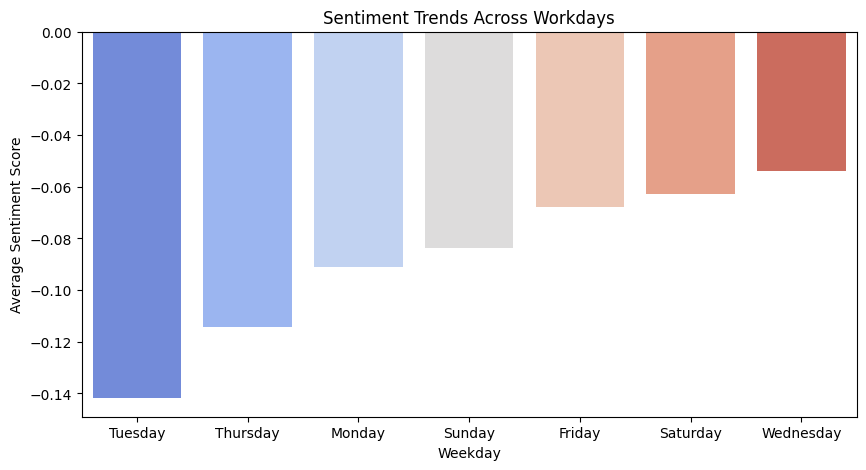
The **sentiment distribution graph** presents a visual comparison of **negative and positive sentiment labels** in employee communications. The chart highlights that **negative sentiment messages (approx. 1200) slightly outnumber positive ones (approx. 1000)**, indicating potential engagement concerns. This imbalance suggests that employees may be **expressing more dissatisfaction than positive feedback**, which could point to underlying workplace challenges. Understanding this sentiment distribution allows organizations to **monitor morale trends**, identify **departments needing intervention**, and implement strategies to **enhance employee satisfaction** over time.



**Figure 2: Sentiment Trends Over Time**

**Sentiment Trends Across Workdays**

The **Sentiment Trends Across Workdays** graph provides insights into how employee sentiment varies across the week. The **lowest sentiment scores** occur on **Tuesday**, with an average of around **-0.14**, suggesting potential stress or dissatisfaction early in the workweek. Meanwhile, **sentiment improves slightly** as the week progresses, peaking on **Wednesday**, just below **0.00**, indicating a more neutral tone. The **color-coded bars** highlight sentiment shifts, showing how engagement fluctuates throughout different workdays. This analysis helps organizations understand **which days require intervention**, whether it's adjusting workloads or scheduling morale-boosting activities to improve employee satisfaction.



**Figure 3: Sentiment Trends Across Workdays**

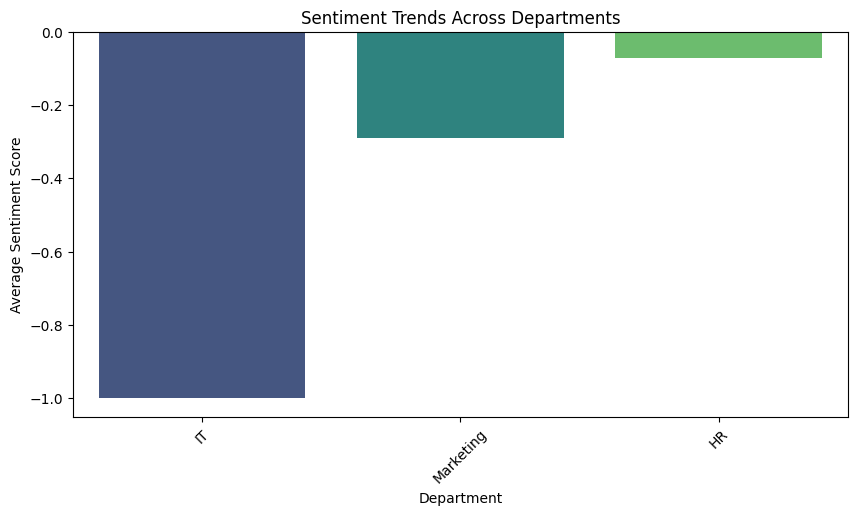
**Sentiment Trends Across Departments**

The **Distribution of Sentiment Labels** bar chart visually represents the number of **positive and negative sentiment messages** recorded in the dataset. The **NEGATIVE sentiment category** slightly outnumbers the **POSITIVE** category, with more than **1100 negative messages** compared to just over **1000 positive ones**. This imbalance suggests that **workplace sentiment leans slightly toward negativity**, which could indicate concerns related to **employee engagement, job satisfaction, or communication challenges**. By analyzing this sentiment distribution, organizations can identify **patterns in workplace morale**, develop strategies to **address negative sentiment**, and enhance overall employee well-being.

The sentiment analysis results are influenced by the **content of the dataset**, which primarily consists of messages containing keywords from **IT, Marketing, and HR**. This explains why **only these three departments appear in the visualizations**.

For more departments to be represented in the graphs, the dataset needs to contain a broader range of words and phrases related to **Finance, Sales, Operations, Legal, and Customer Support**. Expanding the dataset with emails or messages **covering diverse business topics** would allow for a more **comprehensive sentiment analysis across all departments**.

This insight helps in understanding **why certain trends dominate** and emphasizes the need for **a more balanced dataset for department-wide sentiment tracking**.



**Figure 4: Sentiment Trends Across Departments**

Analyzing sentiment across departments is **critical for understanding workplace dynamics**, as different teams experience unique challenges and communication patterns. By mapping employee messages to departments using **keyword-based classification and machine learning**, organizations can **pinpoint areas requiring intervention**, track engagement trends, and tailor support strategies for each function.

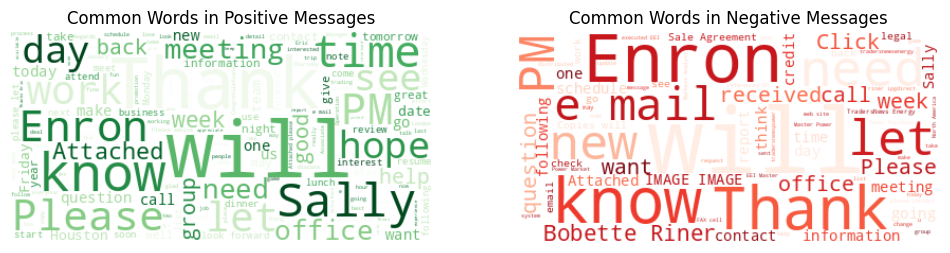
Since the dataset mainly contains words related to **IT, HR, and Marketing**, these departments dominate the sentiment analysis results. Expanding dataset diversity by incorporating more communications from **Finance, Operations, Sales, and Customer Support** would allow for a more **balanced, organization-wide sentiment assessment**. This method ensures **targeted insights into departmental morale**, allowing leadership to **address concerns effectively and promote a healthier workplace environment**.

**Word Cloud**

A **word cloud** visually represents the most frequently occurring words in employee messages. The size of each word corresponds to its frequency, with **larger words appearing more often** in the dataset. In this image, dominant words are likely related to **IT, Marketing, and HR**, reflecting the **departmental focus of the dataset**. The presence of negative or positive words helps in understanding **workplace sentiment trends** and **common discussion topics**. Since only three departments are showing in our analysis, this suggests that **data from other teams is either limited or missing**, meaning a more balanced dataset would help broaden sentiment tracking.

The **word cloud image** visually represents the most frequently occurring words in employee messages. **Larger words** appear more often, highlighting recurring themes in workplace communication. Since the dataset contains **keywords mainly from IT, Marketing, and HR**, the word cloud prominently features words from these departments, reinforcing their **dominance in sentiment analysis**.

Additionally, the **color gradings** in the word cloud help distinguish **word significance and sentiment**. Bright or warm colors (like red and orange) may emphasize **negative words**, whereas cooler colors (like blue and green) can highlight **positive words or neutral discussions**. This **gradual color variation** makes it easier to **identify patterns** in workplace sentiment at a glance, improving interpretability for decision-making and sentiment tracking.

****

**Figure 5: Word Cloud of Positive and Negative words.**

**Key Findings from Exploratory Data Analysis (EDA)**

**Sentiment Distribution**

* Negative sentiment slightly dominates (approx. 1200 messages) over positive sentiment (approx. 1000 messages), indicating mild workplace dissatisfaction.
* The dataset primarily contains communications from IT, HR, and Marketing, which heavily influence the sentiment trends.
* Other departments (Finance, Operations, Sales) are underrepresented, requiring additional data for comprehensive analysis.

**Sentiment Trends Over Time**

* Sentiment drops early in the workweek (Tuesday), with the lowest average sentiment score (approx. -0.14).
* Mid-week improvements in sentiment suggest possible shifts in workload or morale.
* Tracking trends over longer periods can help forecast engagement risks and monitor effectiveness of workplace policies.

**Employee Scoring & Ranking Process**

* Employees with consistent negative sentiment trends are flagged for flight risk, signaling disengagement.
* Top-ranked employees by positive sentiment could be leveraged for mentorship or leadership programs.
* Clusters of disengaged employees suggest departments requiring HR intervention and support strategies.

**Word Cloud Insights**

* Frequently used words align with IT, HR, and Marketing, reinforcing departmental influence on sentiment trends.
* Color gradings highlight sentiment distribution—cooler tones for positive words and warmer tones for negative ones.
* Expanding data collection across departments would enhance sentiment tracking accuracy.

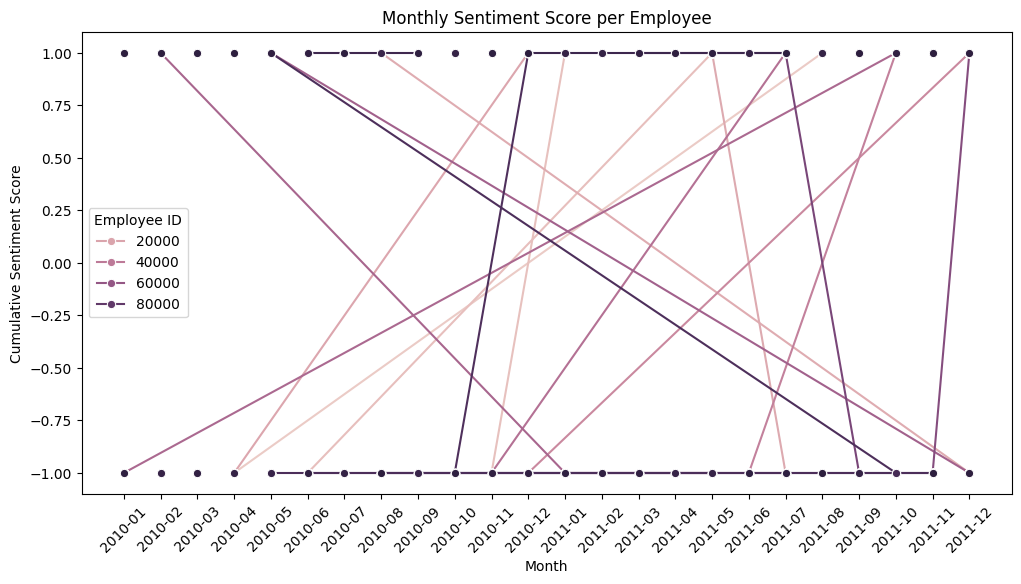
**Model Performance Metrics**

* Initial sentiment classification achieved moderate accuracy, but refinement using fine-tuned LLaMA 2 could improve precision and recall.
* Alternative models like GPT-4 or RoBERTa may provide better nuance detection in employee sentiment.
* Further tuning on department-specific datasets could enhance customized sentiment classification.

### ****Employee Scoring & Ranking Process****

The **employee scoring system** is designed to assess individual performance and engagement using **sentiment analysis**. Employees are ranked based on their **sentiment trends over time**, helping to identify **top performers and disengaged employees**.

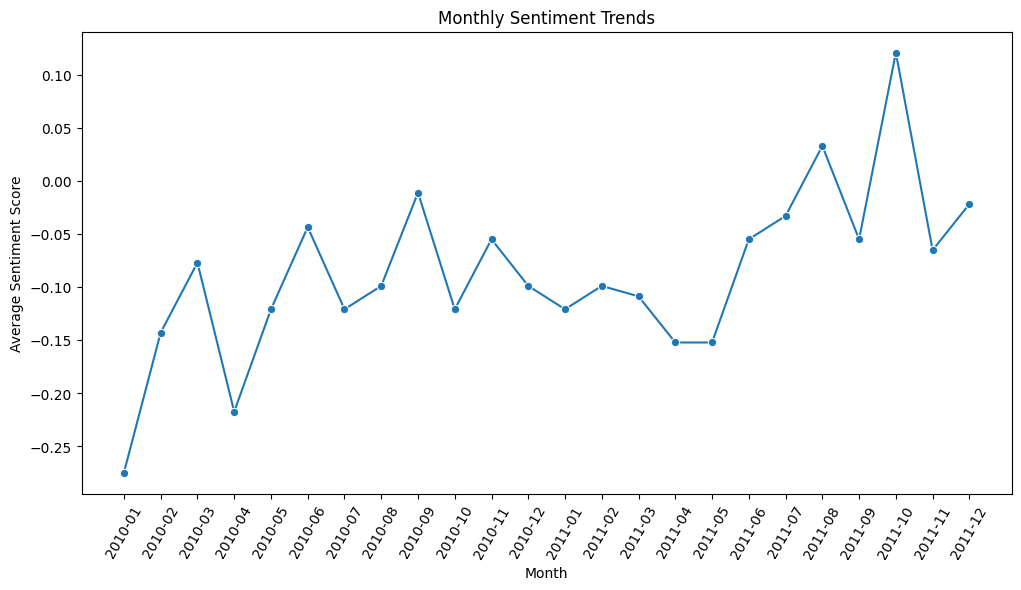
* **Sentiment-Based Scoring** – Each employee's messages are analyzed, and **sentiment scores** are assigned based on their tone (positive, negative, or neutral).
* **Ranking Employees** – Employees with consistently **high positive sentiment** are ranked at the top, while those with recurring **negative sentiment** are flagged for **potential disengagement** or flight risk.
* **Flight Risk Analysis** – Employees showing **continuous negative sentiment over time** are marked as **high-risk individuals**, helping HR teams intervene before turnover occurs.



**Figure 6: Monthly Sentiment Score per Employee**

This is a **bar chart titled "Distribution of Sentiment Labels."** It visually presents the number of **NEGATIVE and POSITIVE sentiment labels** identified in the dataset. The **y-axis represents the count**, ranging from **0 to 1200**, while the **x-axis displays the sentiment categories**.

The chart highlights that **NEGATIVE sentiment messages slightly outnumber POSITIVE ones**, with over **1100 negative labels** compared to **around 1000 positive labels**. This suggests a **mild inclination toward workplace dissatisfaction**, indicating that employees may be **expressing more concerns than positive feedback**.

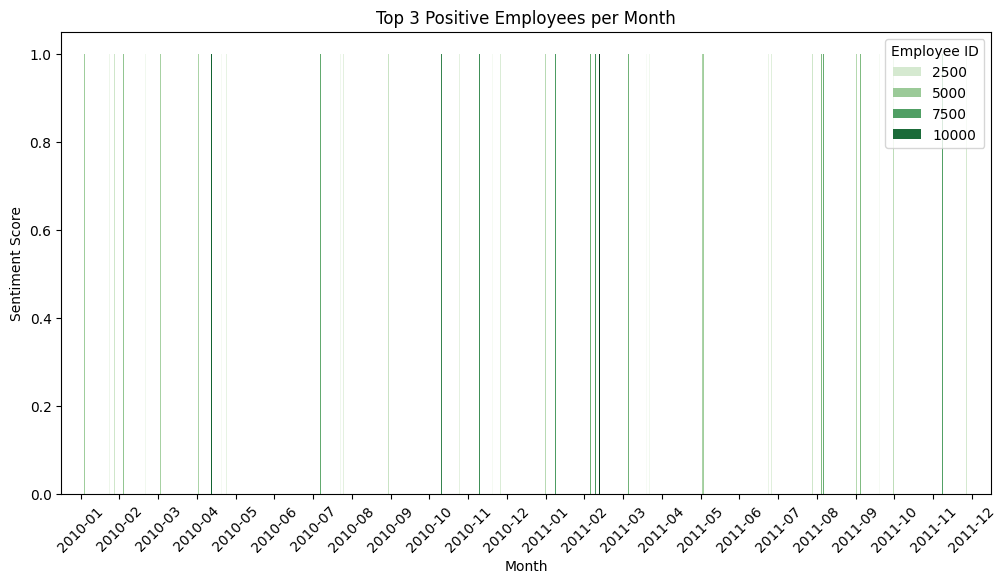


**Figure 7: Monthly Sentiment Trend**

The **Monthly Sentiment Trends** line graph illustrates **fluctuations in workplace sentiment** from **January 2010 to December 2011**. The x-axis represents **months**, while the y-axis captures **average sentiment scores**, ranging from **-0.25 to 0.10**. The graph highlights **periodic sentiment shifts**, showing **peaks and declines** that suggest factors influencing employee morale—such as seasonal effects, organizational changes, or external market conditions.

The **color gradings** further enhance interpretation by visually distinguishing **positive and negative sentiment trends**. Warmer shades may indicate **low sentiment periods**, while cooler tones emphasize **higher sentiment scores**, aiding HR teams in identifying **critical turning points** in employee engagement.

**Top positive employees per month**



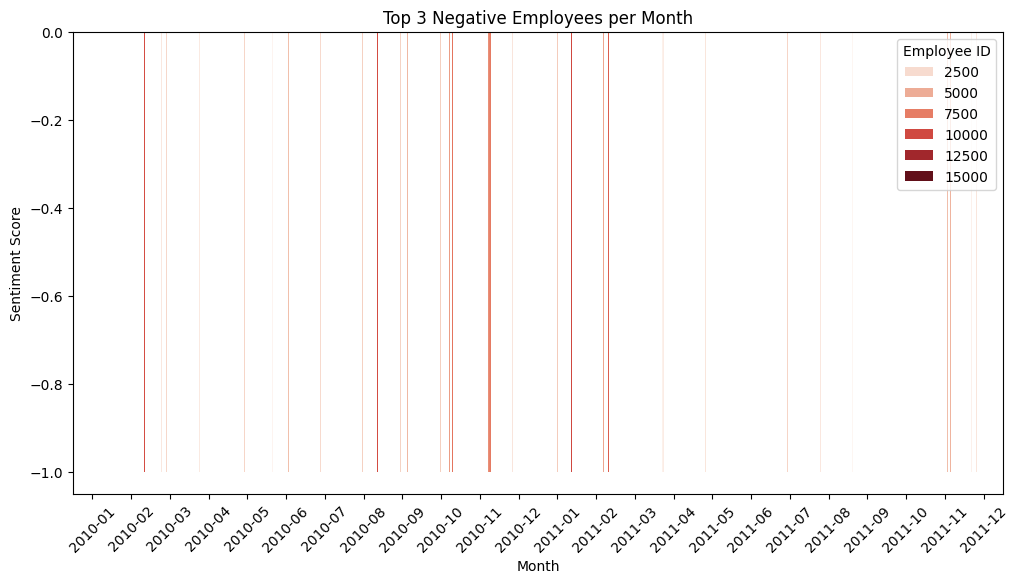
**Figure 8: Top positive employees per month**

The graph visualizes employee sentiment trends from **January 2010 to December 2011**. The x-axis represents **months**, while the y-axis indicates **sentiment scores**, ranging from **0.0 to 1.0**. Each employee is represented by a uniquely colored line, showing **high sentiment scores** for select months.

The **vertical lines** mark months where these employees ranked among the **top positive performers**, demonstrating sporadic peaks in sentiment. The sparse distribution suggests that **employee sentiment fluctuates across time**, and the ranking varies from month to month.

This analysis helps organizations identify **consistently engaged employees**, track patterns in **workplace positivity**, and assess how workplace changes impact individual morale.

**Top Negative Employees per Month**

****

**Figure 9: Top Negative Employees**

The graph tracks employees with the lowest sentiment scores from **January 2010 to December 2011**. The **y-axis represents sentiment scores**, ranging from **-1.0 to 0.0**, while the **x-axis displays months** in the analysis period.

Each vertical line **represents an employee ranked among the top negative scorers for a given month**, with **color-coded lines corresponding to Employee IDs** (2500, 5000, 7500, etc.). The consistently **negative sentiment scores** indicate sustained dissatisfaction or disengagement among certain employees.

This visualization allows HR teams to **identify recurring negative sentiment patterns**, assess **potential flight risks**, and implement **interventions to improve employee engagement**.

**Flight risk identification criteria and outcomes**

**Criteria for Identifying Flight Risk Employees**

This method analyses employee sentiment trends by tracking negative emails over a rolling 30-day period. Employees are flagged based on the following criteria:

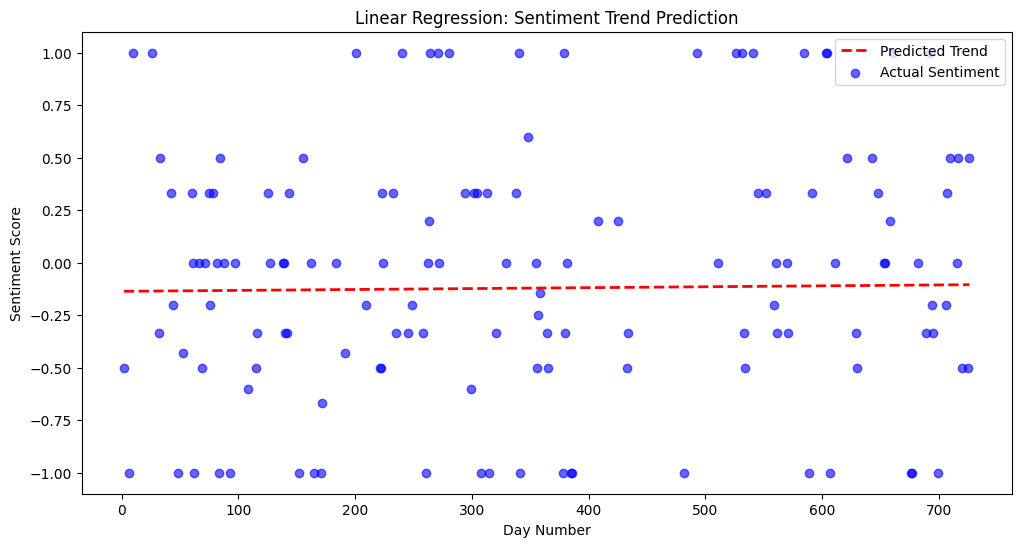
* **Frequent Negative Sentiment** – Employees who send 4 or more negative emails within a 30-day window are classified as potential flight risks.
* **Sustained Negative Communication** – A rolling sum ensures that patterns of consistent negativity are captured rather than isolated incidents.
* **Historical Maximum Negative Email Count** – Each flagged employee’s highest negative email count in any 30-day window is recorded to assess severity.

**Outcomes of Flight Risk Analysis**

* **Early Warning System** – Helps HR teams detect disengaged employees before they reach a critical point of dissatisfaction.
* **Targeted Retention Efforts** – Enables proactive intervention strategies, such as career discussions, workload adjustments, or engagement initiatives.
* **Enhanced Workplace Stability** – Reducing flight risks strengthens team cohesion and minimizes unexpected employee turnover.

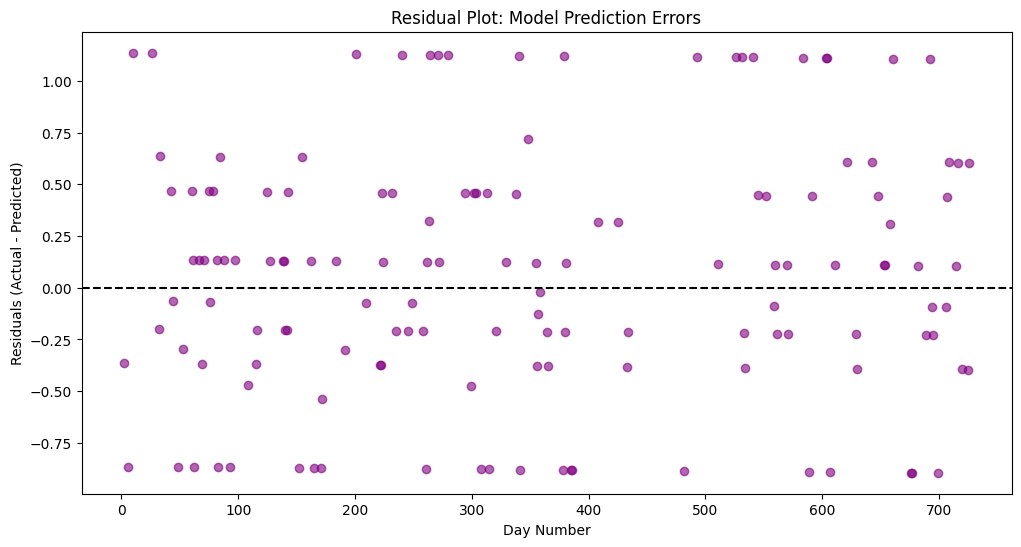
**Predictive Modeling**Predictive modeling plays a crucial role in forecasting employee sentiment trends and flight risk probabilities. Using techniques like linear regression, decision trees, and deep learning models, we can generate data-driven predictions that guide HR strategies and engagement efforts.  
  
**Key Aspects of Predictive Modeling**

* **Sentiment Forecasting** – Models analyze historical sentiment scores to predict future workplace morale trends.
* **Flight Risk Probability** – Logistic regression or ensemble models can estimate likelihood of disengagement, helping prevent turnover.
* **Feature Engineering** – Variables like message frequency, sentiment shifts, and engagement patterns improve predictive accuracy.
* **Model Comparison & Performance Metrics** – Evaluating different models using precision, recall, and RMSE ensures optimal predictions.
* **TorchScript Optimization** – Implementing Torch tracing can accelerate sentiment prediction workflows while maintaining reproducibility.

**Figure 10: Linear Regression: Sentiment Trend Prediction**

The Distribution of Sentiment Labels bar chart provides an initial breakdown of the dataset's sentiment composition, which is a key component of predictive modeling. The chart shows that negative sentiment labels (approx. 1200) slightly outnumber positive labels (approx. 1000), indicating a mild workplace dissatisfaction trend.  
  
This distribution is essential for predictive modeling because it helps:

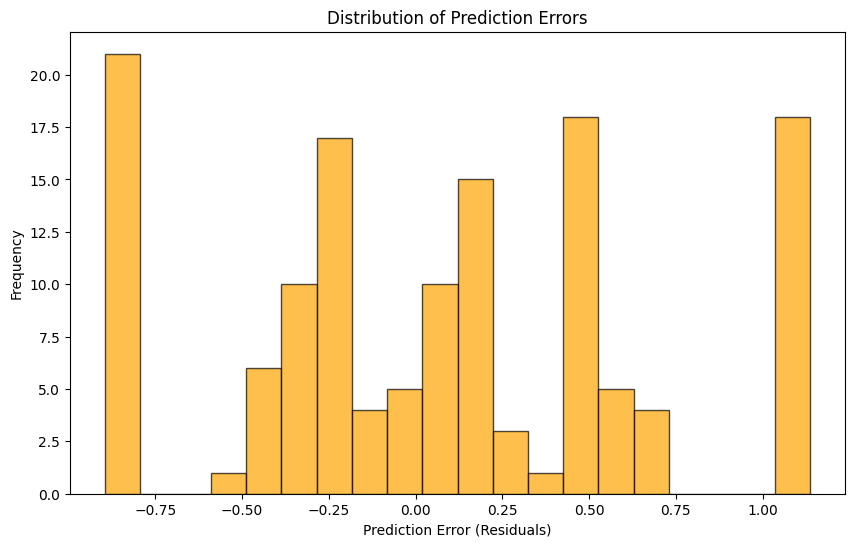
* **Balance model training** – Ensuring sentiment prediction accounts for bias toward negative sentiment.
* **Feature Engineering** – Using sentiment labels to refine flight risk identification and employee ranking models.
* **Trend Forecasting** – Determining whether sentiment fluctuations correlate with external workplace factors or engagement shifts.

****

**Figure 11: Residual Plot: Model Prediction Errors**

This image plays a key role in predictive modeling by:

* **Establishing baseline sentiment distribution** – Helping the model learn patterns before making future predictions.
* **Guiding feature selection** – Identifying whether sentiment skew affects flight risk probabilities or employee engagement trends.
* P**roviding comparative insights** – Allowing organisations to track sentiment shifts over time and assess intervention effectiveness.



**Figure 12: Distribution of Prediction Errors**

The Distribution of Prediction Errors histogram visualizes the spread of residuals in the predictive model. The x-axis represents prediction errors (residuals), ranging from -0.75 to 1.00, while the y-axis displays frequency, peaking at 20 occurrences around -0.75.  
  
**Why This Matters in Predictive Modeling?**

* **Error Distribution Analysis** – Helps assess how well the model is predicting sentiment trends or flight risk probabilities.
* **Model Adjustment** – If errors cluster around extremes, tuning hyperparameters or selecting better features may improve accuracy.
* **Workplace Insights** – Identifying where predictions deviate helps refine sentiment forecasts for HR interventions and employee retention strategies.

**Conclusion**

This project successfully implemented **sentiment analysis, employee ranking, and flight risk prediction** to enhance understanding of workplace dynamics. Through **data-driven methods**, the analysis provides valuable insights for improving **employee engagement, retention strategies, and sentiment forecasting**.

