**Final Report**

# 1. Approach and Methodology

This project applies Natural Language Processing (NLP) and Machine Learning techniques to analyze employee feedback data. The methodology followed a pipeline approach:  
1. **Data Collection** – feedback data was collected and preprocessed (cleaning, handling missing values).  
2. **Sentiment Labeling** – text feedback was sentiment-labeled (positive, neutral, negative).  
3. **Exploratory Data Analysis (EDA)** – to identify trends, patterns, and anomalies.  
4**. Scoring & Ranking** – employees were assigned sentiment-based scores and ranked monthly.  
5. **Flight Risk Analysis** – employees at potential risk of attrition were identified.  
6. **Predictive Modeling** – machine learning models were trained to predict employee risk outcomes.

# 2. Key Findings from the EDA

The exploratory analysis revealed important insights:  
- **Sentiment Distribution**: Majority of feedback was positive, followed by neutral and negative.  
- **Temporal trends**: Monthly analysis showed fluctuations in sentiment aligned with organizational events.  
- **Feedback length analysis**: Negative feedback was generally longer, suggesting detailed concerns.  
- **Contributors**: A small set of employees contributed disproportionately to negative feedback.

# 3. Employee Scoring and Ranking

The employee scoring process mapped feedback sentiment to numeric scores:

* Positive feedback = +1
* Neutral feedback = 0
* Negative feedback = -1

Scores were aggregated at the monthly level per employee. Based on these scores:

* Top Positive Employees: Highest-scoring employees per month.
* Top Negative Employees: Lowest-scoring employees per month.

This ranking highlighted consistently high-performing employees as well as those requiring intervention.

# 4. Flight Risk Identification

Employees were flagged as potential flight risks based on patterns of negative feedback and declining scores.  
Criteria included:

* Consecutive months of negative scores.
* Sudden sharp decline in feedback scores.
* High proportion of negative vs. positive feedback.

This helped identify employees who may be disengaged or dissatisfied, requiring timely HR attention.

# 5. Predictive Model Overview

A supervised machine learning approach was used to build a predictive model for employee risk:

* **Features**: Aggregated sentiment scores, trends over time, feedback length, employee activity.
* **Models tested**: Linear Regression.

# 6. Conclusion

The project successfully transformed qualitative employee feedback into actionable insights. Through sentiment analysis, scoring, ranking, and predictive modeling, HR teams can:

* Recognize top-performing employees.
* Detect early signs of disengagement.
* Prioritize interventions to reduce attrition risk.

Future improvements could include integrating additional behavioral and performance data, as well as fine-tuning predictive models with larger datasets.