Employee Sentiment Analysis Project

Project Overview

This project performs a comprehensive analysis of employee email messages to evaluate **sentiment**, **engagement**, and **flight risk** using advanced NLP and statistical techniques.

Key Statistics

- **Total Messages Analyzed**: 2,191

- **Unique Employees**: 10

- **Time Period**: January 2010 – December 2011

- **Overall Sentiment Distribution**:

    - Negative: **57.69%**

    - Positive: **37.20%**

    - Neutral: **5.11%**

Top 3 Positive & Negative Employees

(Example Month: November 2011 – highest activity variation)

Top 3 Positive Employees

1. patti.thompson@enron.com → Score: **+7**

2. don.baughman@enron.com → Score: **+2**

3. eric.bass@enron.com → Score: **+1**

Top 3 Negative Employees

1. bobette.riner@ipgdirect.com → Score: **-8**

2. kayne.coulter@enron.com → Score: **-7**

3. john.arnold@enron.com → Score: **-4**

Flight Risk Identification

All 10 employees were flagged as **high flight risk** (≥4 negative messages within any rolling 30-day window):

- bobette.riner@ipgdirect.com

- don.baughman@enron.com

- eric.bass@enron.com

- john.arnold@enron.com

- johnny.palmer@enron.com

- kayne.coulter@enron.com

- lydia.delgado@enron.com

- patti.thompson@enron.com

- rhonda.denton@enron.com

- sally.beck@enron.com

Predictive Modeling Results

- **Model**: Linear Regression

- **Performance**: Perfect fit (R^2 = **1.000**, MAE = **0.00**)

- **Key Drivers**: Number of positive messages (+1 coefficient) and negative messages (-1 coefficient)

Key Insights & Recommendations

- Predominantly **negative sentiment** indicates potential workplace stress or dissatisfaction.

- **Universal flight risk** across all employees calls for immediate retention interventions.

- Message volume and sentiment polarity are strong predictors of monthly engagement scores.

- **Recommended Actions**:

    - Conduct employee engagement surveys

    - Implement targeted wellness/support programs

    - Monitor sentiment trends quarterly

Visualizations

All charts are saved in the visualizations/ folder:

- Sentiment distribution (pie chart)

- Messages & sentiment over time

- Messages per employee

- Message length by sentiment

- Predictive model performance & feature importance

Project fully implemented in modular Python code with clear documentation and reproducibility.