Team Name - Intelligent Analyzers
Group Members:

1. Atharv Chandratre - atharvc2@illinois.edu (Captain)
2. Uday Kanth Reddy Kakarla - uk3@illinois.edu
3. Priyanka Awatramani - pma7@illinois.edu
4. Ansh Bilimoria - amb20@illinois.edu

Topic

Sentiment analysis (or opinion mining) is a natural language processing (NLP) technique used to determine whether data is positive, negative, or neutral. Sentiment analysis is often performed on textual data to help businesses monitor brand and product sentiment in customer feedback, and understand customer needs. We plan to use sentiment analysis on subreddits (like r/CSMajors) and posts on a company forum to help estimate employee satisfaction. We would be using 'emotion detection' sentiment analysis, allowing us to go beyond polarity (i.e. very positive, positive, neutral, negative, very negative), to detect emotions, like happiness, frustration, anger, and sadness. Many emotion detection systems use lexicons (i.e. lists of words and the emotions they convey) or complex machine learning algorithms. We would be using the lexicon-based approach. In today's time, employee satisfaction is of utmost importance. Having a satisfied and happy workforce strengthens the company by lowering employee turnover, increasing employee productivity, increasing customer satisfaction, and promoting loyalty, eventually leading to higher profits for the company. Our tool would help companies apply sentiment analysis on multiple forums and find out their employee satisfaction ratio. Based on the results, the tool would recommend some ideas to the managers like planning a team outing/dinner, etc. This would eventually lead to higher employee satisfaction and increased productivity.

Approach Plan

1. Scrape

2. Pre-processing the data
- Tokenizing the data
- Stop-word removal
- Punctuation removal
- POS tagging

3. Lexicon-Based Approach
- Running preprocessed text through a sentiment lexicon

4. Recommendations
- Generating some recommendations for teams to improve communication and their work experience based on sentiment results received by the lexicon model

Tools to be Used

The tools we will be using during the completion of this project will be the following:
Programming Language - Python
Interactive Python Running Environment - Jupyter Notebooks
Python Execution Environment - Google Colab
Natural Language Processing Library - NLTK (Natural Language ToolKit)

Expected Outcome

The outcome of this project would be that companies would be able to find out their employee satisfaction levels and based on it, they could take measures through the tool generated recommendations which would help boost those scores and eventually lead to more satisfied and productive employees.

Evaluation Methodology

Given that Sentiment Analysis is a classification problem we will be comparing the performance of our Sentiment Analysis against labeled data, which is done according to the intuition of the user who labeled the data. We will be using the evaluation metrics of Precision, Recall and F-Score, and Accuracy to determine the performance of our model.

Workload (Estimated)

Name of Task	Expected Hours
Data Scraping	20
Pre-processing the data: Tokenizing the data Stop-word removal Punctuation removal POS tagging	10
Lexicon-Based Approach: Running preprocessed text through a sentiment lexicon	20
Recommendations: Generating some recommendations for teams to improve communication and their work experience based on sentiment results received by the lexicon model	30
Total	80