

# Sentiment Analysis of Student Feedback

User Manual

Joshua M. Holly - ID: 81236763

Angelo State University – CS6399: Project

Professor Roya Choupani

April 2025

## Project Overview

This project focuses on the analysis of anonymized student feedback through the application of Natural Language Processing (NLP). By using sentiment analysis, the system categorizes student comments into Positive, Neutral, or Negative sentiments. These results are visualized through a Power BI dashboard, offering college administrators actionable insights to guide improvements in course structure and instruction.

## System Requirements

- Python 3.8 or higher
- Jupyter Notebook
- Power BI Desktop (free version)
- GitHub (optional, for source code access)

## How to Run the Sentiment Analysis Notebook

### Step 1: Install Required Packages

Use the command below to install all dependencies: `pip install -r requirements.txt`

### Step 2: Execute the Notebook

1. Launch Jupyter Notebook
2. Open notebook/sentiment\_analysis.ipynb

3. Run all cells sequentially

The notebook performs the following tasks:

- Cleans text data (removes punctuation, lowers case)
- Tokenizes using spaCy
- Removes stopwords and applies lemmatization
- Conducts sentiment analysis with VADER
- Outputs a final .csv for use in Power BI

## Using the Power BI Dashboard

### **Step 1: Open the Dashboard**

1. Launch Power BI Desktop
2. Open the file at: powerbi/student\_feedback\_dashboard.pbix

### **Step 2: Explore the Dashboard**

The single-page dashboard contains:

- Pie chart showing sentiment distribution
- Bar chart of sentiment by course
- Bar chart of sentiment by delivery mode
- Table showing the most recent feedback entries

Filtering options include:

- Course name
- Sentiment classification
- Delivery mode

While a second dashboard page for detailed feedback was initially planned, it was excluded to maintain a clean, focused design for this one-semester dataset. This decision is documented in the final report.

### Output Location

The final CSV containing processed feedback and sentiment labels can be found at:

`data/student_feedback_dataset.csv`

### Notes

- All data used in this project is synthetic and anonymized.
- A database was deemed unnecessary; using .csv files with Power BI aligns with common academic/professional reporting workflows.

## **Requirements**

### ***Functional Requirements***

- The system must collect and process text-based student feedback data
- The system must classify sentiment into categories: Positive, Negative, Neutral, and Mixed.
- The system must allow filtering by course, instructor, difficulty level, etc.
- There should be a data visualization component of the project that provides descriptive analytics based on the outcome.

### ***Non-Functional Requirements***

- The system must handle large datasets.
- The system must perform in real-time or batch processing depending on size of dataset.
- The model should be interpretable, ensure transparency in classifications results.

### ***Constraints***

- Data Availability: No dataset currently exists, so it must be collected, processed, cleaned, and synthesized.
- Ethical Considerations: Student and instructor data must remain anonymized.

## UI Design

- A Tableau (or similar) dashboard that provides sentiment trends and insights.
- Wireframe:
  - Landing Page / Dashboard
    - Sentiment Summary: Pie or Bar chart showing sentiment distribution.
    - Word cloud: Highlighting the most frequently used words in student feedback.
    - Recent Feedback: List of the latest analyze comments w/ sentiment labels.
  - Feedback Analysis Page
    - Student Feedback List – Table and column format containing information such as Student ID, Course, Instructor, Feedback Text, Sentiment Label, etc.
    - Detailed View: Clicking on a feedback entry would expand to show all details in user's feedback.

## **System Design**

### ***Architecture***

- Data Collection: Manually collected student feedback.
- Sentiment Analysis Model: NLTK, Vader, spaCy
- Database: Storing Feedback
- Visualization Module: Tableau

### ***Modules***

- Data Collection Module
- Preprocessing Module – Data Cleaning
- ML Model – Perform sentiment classification.
- Storage Module – Store processed feedback and results.
- Visualization Module – Display Sentiment trends.

### ***Tech Stack***

- Programming Language: Python
- NLP Libraries: NLTK, Vader, spaCy
- Database: SQLite
- Visualization: Matplotlib/Seaborn/Tableau

### Project Timeline / Implementation Plan

| Task                        | Duration / Timeline      | Status                                 |
|-----------------------------|--------------------------|--|
| Data Collection             | 1 Week / 2.9.25-2.15.25  | Behind / Data Collection is in process |
| Data Cleaning & Processing  | 2 Weeks / 2.16.25-3.1.25 | On Track                               |
| Model Training & Testing    | 3 weeks / 3.2.25-3.22.25 | On Track                               |
| Analysis & Visualization    | 2 weeks / 3.23.25-4.5.25 | On Track                               |
| Report & Final Presentation | 1 week / 4.6.25-4.11.25  | On Track                               |

### Test Plan

- Unit Testing: Ensure modules function correctly
- Performance Testing: Evaluate model accuracy & processing speed
- Integration Testing: Ensure seamless interaction between modules.

### Risk Plan

| Risk                | Mitigation Strategy    |
|---------------------|------------------------|
| Lack of Dataset     | Use Synthetic Data     |
| Model underperforms | Try alternative models |
| Ethical Concerns    | Ensure anonymization   |



## **Conclusion**

This project outlines a sentiment analysis system that is design to process and ultimately provide insights into university students course feedback. The key modules include data ingestion, preprocessing, machine learning sentiment classification, storage, and finally visualization.

The machine learning model will enable classification of feedback while the visualization model with ensure that insights are available and easy to understand. This system is designed for future iterations and long-term use to the client – the universities.

In conclusion, this project will demonstrate an effective framework for sentiment analysis that turns unstructured text into meaningful insights that can support decision-making.

## References

### Datasets:

- Kaggle.com
- Google Data Search
- Synthetic Data – To be created

### NLP:

- VADER: <https://github.com/cjhutto/vaderSentiment>
- NLTK (Natural Language Toolkit): [NLTK :: Natural Language Toolkit](#)
- spaCy: [spaCy · Industrial-strength Natural Language Processing in Python](#)

### Data Processing:

- Pandas: [pandas.pydata.org](https://pandas.pydata.org)
- NumPy: [numpy.org](https://numpy.org)

### Machine Learning Frameworks:

- Scikit-learn: [scikit-learn.org/stable/](https://scikit-learn.org/stable/)
- PyTorch: [pytorch.org](https://pytorch.org)

### Data Visualization & Analysis:

- Matplotlib: [matplotlib.org](https://matplotlib.org)
- Seaborn: [seaborn.pydata.org](https://seaborn.pydata.org)
- Tableau Public: [Discover | Tableau Public](#)