

Sentiment Analysis of Student Feedback

Project Design and Implementation – Final Report

Joshua M. Holly - ID: 81236763

Angelo State University – CS6399: Project

Professor Roya Choupani

April 2025

Problem Statement

The effectiveness of university course and instructors is measured through student provided feedback at the conclusion of their course. Analyzing this data is time-consuming and can be conditional based on a student's attributes. Using Natural Language Process (NLP) techniques, this project aims to develop a sentiment analysis system that classifies student feedback into Positive, Negative, Neutral, or Mixed sentiments. Insights arriving from this system will be utilized by universities to better understand their faculty, students, and curriculum success.

Requirements

Functional Requirements

- The system must collect and process text-based student feedback data
- Sentiment is classified into three categories: Positive, Neutral, and Negative ("Mixed" was dropped).
- 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

- No real dataset existed, so a fully synthetic dataset was generated.
- All feedback, students, and instructors are anonymized to ensure ethical standards.

UI Design

- A Power BI dashboard was used instead of Tableau due to easier integration with CSV outputs.
- Wireframe:
 - Landing Page / Dashboard
 - Sentiment Summary: Pie chart showing sentiment distribution
 - Recent Feedback: Table of the most recent student feedback entries, with sentiment
 - Sentiment by Course: Horizontal bar chart
 - Sentiment by Delivery Mode: Horizontal bar chart
 - A second page for detailed analysis was considered but not implemented due to scope & lack of value.

System Design

Architecture

- Data Source: Synthetic student feedback generated using Faker and manual input
- Sentiment Analysis Model: spaCy for preprocessing, VADER for classification
- Database: Database was omitted due to ease of use for .csv file implementation into PowerBI.
- Visualization Module: PowerBI

Modules

- Data Generation & Collection Module
- Preprocessing Module – Text Cleaning, Tokenization, Lemmatization
- Sentiment Scoring Module – VADER classifier
- Output Module – Exports cleaned + labeled feedback to CSV
- Visualization Module – Power BI Dashboard

Tech Stack

- Language: Python
- NLP Libraries: spaCy, VADER
- Visualization: Power BI
- Storage Format: CSV (SQLite and Access were considered but excluded)

Project Timeline / Implementation Plan

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

Test Plan

- Functional Testing: Ensured notebook processed feedback correctly and output matched expectations
- Visualization Testing: Verified interactivity of slicers and correctness of Power BI charts
- Manual Inspection: Spot-checked rows of CSV against sentiment scores for accuracy

Risk Plan

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

Conclusion

This project implements a sentiment analysis system to evaluate student feedback using Natural Language Processing (NLP) techniques. By processing raw student comments, the system generates actionable insights. Some of the key details within this project include text preprocessing with spaCy, sentiment classification using VADER, and data visualization through Power BI, these processes provide an accessible and interactive way for academic leaders to analyze course feedback.

The value of this system for academic leaders resides in its ability to quickly convert unstructured student feedback into visualized insights. This allows academic leader to forgo the need to manual review an abundance of student feedback provided at the conclusion of a semester. The PowerBI dashboard highlights key trends which includes sentiment distribution, course-level breakdowns, and comparisons based on delivery mode (online/in person/hybrid). This will enable informed decision-making around curriculum adjustments and instructional improvements which provides a straightforward way for leaders to monitor and enhance the student experience.

In summary, this project offers a flexible and easy to implement framework for universities to leverage sentiment analysis in their decision-making processes. It simplifies the extraction of meaningful insights from student feedback, helping academic leaders drive continuous improvement and better understand the impact of their courses.

References

Datasets:

- Kaggle (<https://www.kaggle.com/>)
- Google Dataset Search (<https://datasetsearch.research.google.com/>)
- Synthetic data manually generated using the Faker library (<https://faker.readthedocs.io/>)

Natural Language Processing:

- VADER Sentiment: <https://github.com/cjhutto/vaderSentiment>
- spaCy NLP Toolkit: <https://spacy.io/>

Data Processing:

- pandas: <https://pandas.pydata.org/>
- NumPy: <https://numpy.org/>
- openpyxl (for Excel file support): <https://openpyxl.readthedocs.io/>

Visualization:

- Power BI: <https://powerbi.microsoft.com/>
- matplotlib: <https://matplotlib.org/>
- seaborn: <https://seaborn.pydata.org/>