**Project Report**

**On**

**Reviewify – College Review Webpage**

**Submitted partial fulfillment of the requirements for the award of the degree of**

**BACHELOR OF TECHNOLOGY**

**in**

**COMPUTER SCIENCE ENGINEERING**

**(ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING)**

**Submitted By**

**Utkarsh Arora**

**Roll No. 2823432**

**Batch (2023-2027)**



**Panipat Institute of Engineering & Technology, (Affiliated to Kurukshetra University Kurukshetra, India)**

**(Session 2022-2026)**

# INTRODUCTION

"Reviewify" is a static web-based project designed to display college reviews in a simplified and visual format. It enables prospective students to gain insights into different universities through user-submitted reviews, represented as word clouds and bar graphs. The website currently features reviews for PIET, Amity University, and Kurukshetra University.

**Project Overview**

"Reviewify" is a college-based review system designed to gather, analyze, and visualize student feedback using Natural Language Processing (NLP). The project collects data through Google Forms and processes it to generate insightful word clouds, highlighting the most frequent and significant feedback themes. This system serves as a transparent and efficient medium for both students and administration to understand common sentiments and areas of improvement.

**Problem Statement**

Currently, student feedback systems in many educational institutions are either manual or lack analytical depth. These systems often fail to present a comprehensive and visual understanding of student opinions. As a result, valuable insights are lost, and decisions based on feedback become ineffective or delayed.

**Scope of the Project**

The project is focused on collecting student reviews from institutions like Kurukshetra University (KUK) and Amity University through a standardized online form. The feedback data is processed using Python libraries to generate word clouds for individual colleges or departments, offering visual representation of the most commonly used words in student reviews. The system is scalable and can be integrated with multiple universities over time.

**Research Objectives**

* To design an automated system for collecting and analyzing student feedback.
* To apply NLP techniques for preprocessing and visualization of text data.
* To generate word clouds that represent key feedback points clearly.
* To provide administrative users with an efficient overview of students' sentiments.

**Motivation for the Project**

This project was inspired by the lack of effective feedback mechanisms in many colleges. Traditional feedback methods are either ignored or do not provide real-time insights. With the rise of data science tools, it is now possible to interpret large volumes of text data meaningfully, thus motivating the development of this project to bridge the communication gap between students and administration.

**Significance of the Project**

Reviewify enables institutions to identify strengths and problem areas based on student inputs. By leveraging NLP for textual analysis, the system brings a modern approach to feedback interpretation. It improves transparency, encourages student participation, and assists in data-driven decision-making for academic and infrastructural improvements.

**Methodology**

**Overview of the Methodology**

The methodology followed in this project is a structured pipeline that involves collecting student reviews, cleaning and processing the text data, analyzing patterns using Natural Language Processing (NLP), and finally, generating word clouds to visually interpret the results. Each step has been carefully designed to ensure accuracy, efficiency, and meaningful insights from the collected data.

**Data Collection**

Data was collected using a Google Form distributed among students from various departments of Kurukshetra University (KUK) and Amity University. The form allowed students to provide open-ended textual feedback on faculty, campus facilities, academic resources, and overall experience. The responses were exported in CSV format for further processing.

**Data Preprocessing**

Preprocessing involved cleaning the raw textual data to remove noise and irrelevant elements. Key steps included:

* Converting text to lowercase.
* Removing punctuation, numbers, and special characters.
* Removing stopwords (e.g., "is", "the", "and") using the NLTK library.
* Tokenization of sentences into words.
* Lemmatization to convert words to their base forms (e.g., "studying" → "study").

**Exploratory Data Analysis (EDA)**

EDA was performed to understand the nature and distribution of the text data. The following analyses were conducted:

* Word frequency count to identify commonly used terms.
* Visualization using bar charts for most frequent words.
* Comparison between different departments or universities based on word usage.

**Feature Engineering**

For this project, word frequencies served as the primary features. Additional feature engineering included:

* Counting the number of unique words per review.
* Identifying sentiment polarity using TextBlob for a basic sentiment analysis.
* Constructing term-frequency vectors for advanced analysis (if needed in future versions).

**Results Interpretation**

The processed data was used to generate word clouds for each university and department. These word clouds helped in:

* Highlighting key themes and areas of concern.
* Identifying frequently mentioned topics like "faculty", "cleanliness", "labs", etc.
* Supporting administrative decision-making based on visual insights from feedback.

Top of Form

Bottom of Form

**SOFTWARE REQUIREMENT ANALYSIS**

**Define the Problem**

The core problem addressed in this project is the lack of a centralized and analyzable feedback mechanism for universities to assess student opinions regarding academic and campus experiences. Existing systems either do not use textual feedback effectively or fail to visualize it in a form that is easy to interpret. The project aims to bridge this gap by using NLP techniques to extract insights from qualitative student reviews and present them through word clouds.

**Define the Modules and Their Functionalities**

1. **Data Collection Module**
   * Collects student feedback via Google Forms.
   * Exports responses into CSV format for processing.
2. **Data Preprocessing Module**
   * Cleans and normalizes text (removal of noise, tokenization, stopword removal, etc.).
   * Prepares text data for further analysis.
3. **Exploratory Analysis Module**
   * Analyzes the word frequency and distribution.
   * Generates basic visualizations and statistics to understand patterns.
4. **Word Cloud Generator Module**
   * Uses cleaned text data to generate word clouds using Python libraries.
   * Outputs department- or university-specific visual representations.
5. **User Interface (Optional/Future Enhancement)**
   * A GUI to upload CSVs and display word clouds interactively.

**Software Requirements**

1. **Hardware Requirements**

* Processor: Intel Core i3 or above
* RAM: Minimum 4 GB
* Storage: At least 500 MB free space
* Internet: Required for Google Form access and downloading dependencies

1. **Software Requirements**

* **Operating System**: Windows 10 / Linux / macOS
* **Programming Language**: Python 3.x
* **Libraries**:
  + pandas – for data manipulation
  + nltk – for text preprocessing
  + matplotlib and seaborn – for data visualization
  + wordcloud – for generating word clouds
  + TextBlob – for basic sentiment analysis
* **IDE/Editor**: Jupyter Notebook / VS Code / PyCharm
* **Browser**: Chrome/Firefox for accessing Google Forms and Goog