Term Project Proposal: Sentiment Analysis on Review Websites Using Supervised and Zero-Shot Methods

Project Overview

In this project, you will scrape review data from popular websites, such as Amazon, TripAdvisor, Yelp, Rotten Tomatoes, and IMDb. Next you will perform **sentiment analysis** using a supervised model on labeled data, and then apply a **zero-shot sentiment analysis** technique to the same data for comparison.

Objectives

1. **Data Collection**: To scrape and preprocess review data from popular websites.
2. **Supervised Sentiment Analysis**: To train a sentiment analysis model on labeled data.
3. **Zero-Shot Sentiment Analysis**: To use a pre-trained zero-shot model to perform sentiment analysis.
4. **Comparison**: To analyze the performance and generalizability of supervised versus zero-shot sentiment models.

Project Steps

1. Literature Review and Theoretical Background

- **Sentiment Analysis**: Overview of sentiment analysis techniques, including the distinction between supervised and zero-shot models.
- **Zero-Shot Learning**: Explanation of zero-shot learning, its advantages, and limitations in natural language processing.
- **Comparison Studies**: A review of studies comparing supervised and zero-shot learning in sentiment analysis.

2. Data Collection and Preprocessing

- **Data Scraping**:
  - Choose a review website such as Amazon, TripAdvisor, Yelp, Rotten Tomatoes, or IMDb.
  - Use web scraping tools such as **Scrapy**, **Selenium**, or **BeautifulSoup** to gather review data.
  - **Data Fields**: Collect relevant fields such as review text, rating, date, and user information.
- **Data Cleaning and Preprocessing**:
  - Preprocess the review text by standardizing (e.g., lowercasing, removing special characters).
  - Apply NLP preprocessing techniques (tokenization, stop-word removal, lemmatization).

3. Supervised Sentiment Analysis

- **Data Labeling**:
  - Manually label a subset of the data for sentiment (e.g., Positive, Neutral, Negative).
- **Model Training**:

        - Train a supervised sentiment analysis model, such as **SVM, Logistic Regression**.
    - **Evaluation**:
        - Evaluate model performance on a test dataset using metrics like **accuracy, precision, recall, and F1-score**.

4. **Zero-Shot Sentiment Analysis**
    - **Zero-Shot Model Selection**:
        - Select a suitable pre-trained zero-shot model (e.g., **GPT-3, BART, or LLaMA**) to perform sentiment classification without task-specific training.
    - **Implementation**:
        - Apply the zero-shot model to classify sentiment as Positive, Neutral, or Negative.
    - **Evaluation**:
        - Evaluate zero-shot performance on the same test dataset.
5. **Result Comparison and Analysis**
    - **Compare Models**:
        - Compare the performance of supervised and zero-shot models.
        - Use metrics such as **accuracy, precision, recall, and F1-score** for direct comparison.
    - **Findings and Insights**:
        - Analyze scenarios where zero-shot models perform comparably to supervised ones or have limitations.
6. **Discussion and Conclusion**
    - **Interpretation**: Summarize the effectiveness of supervised versus zero-shot models in sentiment analysis.

Expected Deliverables

- A **detailed report** with methodology, results, and insights.
- A **code repository** containing all project-related scripts.
- A **presentation** summarizing approach, findings, and comparisons.

Grading Criteria

1. **Data Collection & Preprocessing** - 20%
2. **Supervised Sentiment Analysis Implementation** - 25%
3. **Zero-Shot Sentiment Analysis Implementation** - 20%
4. **Comparison and Analysis** - 20%
5. **Report & Presentation Quality** - 15%