**Project: Artificial Intelligence – Conception phase**

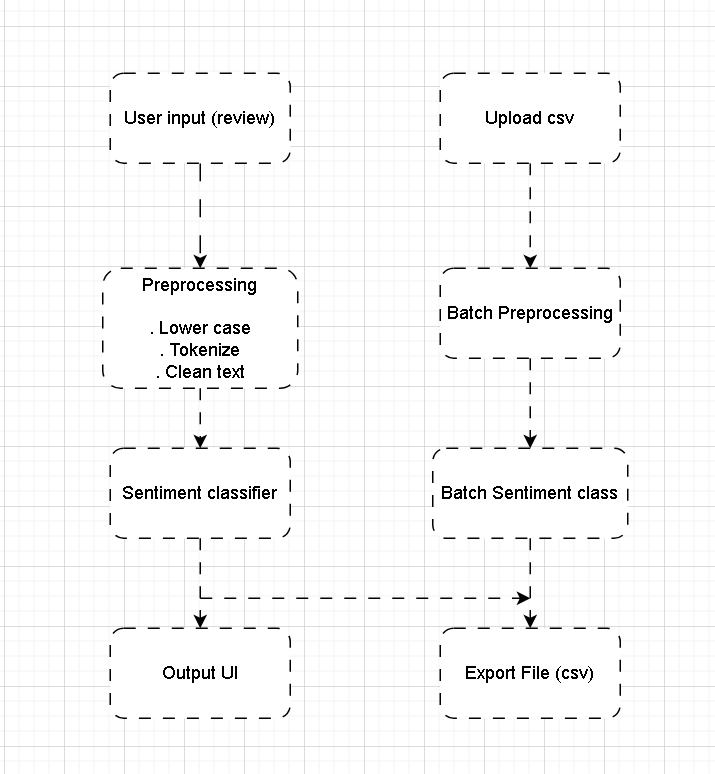
**Overview**

**The goal of this project is to design a sentiment analysis tool that determines how customers feel about a product based on their written reviews. The tool will classify customer reviews into at least three sentiment levels: positive, neutral, and negative. It will also be robust enough to handle challenging textual features such as irony, sarcasm, negation, and multipolarity (mixed sentiments).**

**The system is modular, consisting of components for data input, text preprocessing including batch processing, classification using a fine-tuned BERT-based model, and validation. The solution prioritizes both accuracy and interpretability, ensuring that end-users receive actionable insights.**

**Architecture and component interaction**

The following diagram illustrates the high-level system architecture (see diagram section below).



**Components:**

1. **User Interface (UI)**: Allows marketing users to input or paste a review. Reviews can also be uploaded in bulk as a CSV file.
2. **Preprocessing Module**: Cleans and tokenizes the text, handles punctuation, stopwords, and contractions. This is done both per each review or in bulk.
3. **Sentiment Classifier**: Uses RoBERTa (via Hugging Face Transformers) for context-aware sentiment classification. Hugging face provides the RoBERTa model pre trained on a very large dataset which ensures high accuracy.
4. **Output Module**: Displays predicted sentiment and confidence level. A CSV file is available for download for batch sentiment analysis.
5. **Validation Module**: Evaluates accuracy, precision, recall, and F1-score using labelled test sets. Each review sentiment anlysis is assigned a score and for batch processing, each row with a review is assigned a score.

**Tools & Frameworks**

|  |  |  |
| --- | --- | --- |
| **Component** | **Tool** | **Why** |
| UI | Gradio | User friendly and fast |
| Preprocessing | SpaCY | Efficient processing |
| Model | RoBERTa via Hugging Face | Robust for sarcasm |
| Training | PyTorch/ Transformers | Fine Tuning |
| Evaluation | Scikit-Learn | Confusion Matrix |

This design emphasizes modularity, reusability, and context sensitivity. By using a transformer-based model (RoBERTa), the system handles sarcasm, irony, and negations better than classical models. The separation of concerns (UI, preprocessing, classification, validation) ensures scalability and maintainability.