Project Title	**My Zoom: A Transformer-Based Model for Contextual Feedback Validation**
Skills take away From This Project	**● Text Preprocessing and Data Augmentation** **● Transformer Models (e.g., BERT or RoBERTa)** **● Binary Classification in NLP** **● Model Evaluation and Performance Metrics** **● Deployment using Gradio and Hugging Face Spaces**
Domain	**EdTech (Educational Technology)**

Problem Statement:

Develop a machine learning solution to validate user feedback in an EdTech application. The goal is to determine whether a user's feedback (`text`) aligns with the provided dropdown reason (`reason`). If the feedback aligns, label it as `1`, otherwise label it as `0`.

This validation ensures that only relevant and meaningful feedback is recorded, streamlining the feedback process for enhanced user experience.

Business Use Cases:

Enhanced Feedback Systems:
Validate user feedback before saving to ensure relevance and alignment with predefined categories.

Automated Moderation:
Prevent irrelevant or misleading feedback from being recorded in customer service tools.

Quality Control in Surveys:
Maintain the quality and relevance of user responses in online surveys.

EdTech Analytics:
Use accurate feedback to generate meaningful insights for course improvement.

Approach:

Dataset Preparation:

- Augment the training data to include examples for label `0` (negative class).
- Use techniques like paraphrasing, word swapping, or creating mismatched `text` and `reason` pairs.

Text Preprocessing:

- Clean and tokenize the `text` and `reason`.
- Encode them using a transformer tokenizer (e.g., BERT tokenizer).

Model Development:

- Fine-tune a transformer model (e.g., BERT, RoBERTa) for binary classification.
- Input both `text` and `reason` as pairs to the transformer for contextual understanding.

Evaluation:

- Use the evaluation dataset to calculate metrics like accuracy, precision, recall, and F1-score.
- Ensure the model performs well on both positive and negative classes.

Deployment:

- Create a Gradio-based user interface.
- Host the application on Hugging Face Spaces for public accessibility.

Results

- **Expected Outcomes**:
  - A trained transformer model that can classify whether user feedback aligns with a given reason.
  - Deployment of a working Gradio app for real-time validation.
- **Impact**:
  - Improved feedback validation process with over 85% accuracy.

  - Seamless integration into EdTech platforms.

Project Evaluation Metrics

1. **Accuracy**: Overall correctness of the model.
2. **Precision**: Model's ability to correctly identify relevant feedback (label 1).
3. **Recall**: Model's ability to capture all relevant feedback.
4. **F1-Score**: Balance between precision and recall.
5. **Confusion Matrix**: Visualize performance on both classes.

Technical Tags

- NLP
- Transformers
- Binary Classification
- Text Pair Modeling
- Gradio Deployment
- Hugging Face Spaces

Data Set

- **Link : Dataset link**
- **Format**: Tabular data with columns: `text`, `reason`, and `label`.
- **Variables**:
  - `text`: User feedback
  - `reason`: Dropdown reason
  - `label`: Target variable (1 for aligned, 0 for not aligned)

Data Set Explanation

The dataset contains user feedback (`text`) and a corresponding reason (`reason`) selected by the user from a dropdown. The target `label` indicates whether the feedback aligns with the reason.

- **Preprocessing Steps**:

    - Clean the `text` and `reason` for typos, special characters, and stopwords.
    - Augment negative class (`label 0`) data to ensure balanced training.

Project Deliverables

1. Source Code (Python files or Jupyter Notebook).
2. Documentation (README file explaining the methodology and deployment steps).
3. Gradio-based App hosted on Hugging Face Spaces.
4. Evaluation Report (Metrics and confusion matrix visualization).

Project Guidelines

1. **Version Control**:
    - Use Git for tracking changes.
    - Include detailed commit messages.
2. **Coding Standards**:
    - Follow PEP 8 guidelines for Python.
    - Modularize code into functions or classes.
3. **Best Practices**:
    - Test the model on unseen data before deployment.
    - Monitor class imbalance during training to avoid bias.

Timeline:

1 week

References:

Hugging face course - Link

Streamlit documentation in hugging face - Link

Pre-trained models - Link

Orientation Tamil - Link

Orientation English - Link

PROJECT DOUBT CLARIFICATION SESSION (PROJECT AND CLASS DOUBTS)

About Session: The Project Doubt Clarification Session is a helpful resource for resolving questions and concerns about projects and class topics. It provides support in understanding project requirements, addressing code issues, and clarifying class concepts. The session aims to enhance comprehension and provide guidance to overcome challenges effectively.
Note: Book the slot at least before 12:00 Pm on the same day

Timing: Monday-Saturday (4:00PM to 5:00PM)

Booking link :https://forms.gle/XC553oSbMJ2Gcfug9

For DE/BADM project/class topic doubt slot clarification session:

Booking link : https://forms.gle/NtkQ4UV9cBV7Ac3C8

Session timing:

For DE: 04:00 pm to 5:00 pm every saturday
For BADM 05:00 to 07:00 pm every saturday

LIVE EVALUATION SESSION (CAPSTONE AND FINAL PROJECT)

About Session: The Live Evaluation Session for Capstone and Final Projects allows participants to showcase their projects and receive real-time feedback for improvement. It assesses project quality and provides an opportunity for discussion and evaluation.
Note: This form will Open only on Saturday (after 2 PM) and Sunday on Every Week

Timing:

For BADM and DE
Monday-Saturday (11:30AM to 1:00PM)

For DS and AIML
Monday-Saturday (05:30PM to 07:00PM)

Booking link : https://forms.gle/1m2Gsro41fLtZurRA

Created By:	Verified By:	Approved By:
Aravinth Meganathan		