MY ZOOM

Presented by :

Mithun Karthikeyan 2022506086
Karthikeyan B 2022506093
Pranaav Sanjith R 2022506130

OVERVIEW

- Abstract
- Introduction
- Problem
- Objectives
- Methodology
- Result
- Conclusion
- Reference

ABSTRACT

In the EdTech industry, collecting user feedback is critical for improving learning platforms. However, a common issue arises when users provide free-text feedback that does not align with the selected predefined reason, leading to poor data quality. This project proposes a machine learning-based solution using a BERT-based classifier to automatically validate the contextual alignment between the feedback text and the selected reason. The system ensures high-quality, relevant feedback and aids in moderation and analysis.

INTRODUCTION

EdTech platforms often collect user feedback through a combination of dropdown-selected reasons and free-text inputs. While structured inputs help in categorization, free-text comments provide detailed context. However, inconsistencies frequently occur when the written feedback does not correspond to the selected reason, causing challenges in data interpretation and decision-making. This project addresses this problem by developing an intelligent feedback validation system using NLP and deep learning techniques to ensure feedback relevance and integrity.

PROBLEM

- 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.

OBJECTIVES

- To develop a BERT-based model that determines whether a user's feedback text aligns with the selected dropdown reason.
- To improve feedback quality by filtering out irrelevant or mismatched responses.
- To support automation in feedback moderation and enhance analytical insights for EdTech platforms.

METHODOLOGY

Data Preparation

- Input data consists of user feedback text, selected reason, and a binary label indicating alignment.
- Text is cleaned using regular expressions and converted to lowercase.

Model Architecture

- Used DistilBERT for sequence classification with two output labels: "aligned" and "not aligned".
- Tokenized paired inputs (feedback text and reason) using DistilBERT tokenizer.

METHODOLOGY

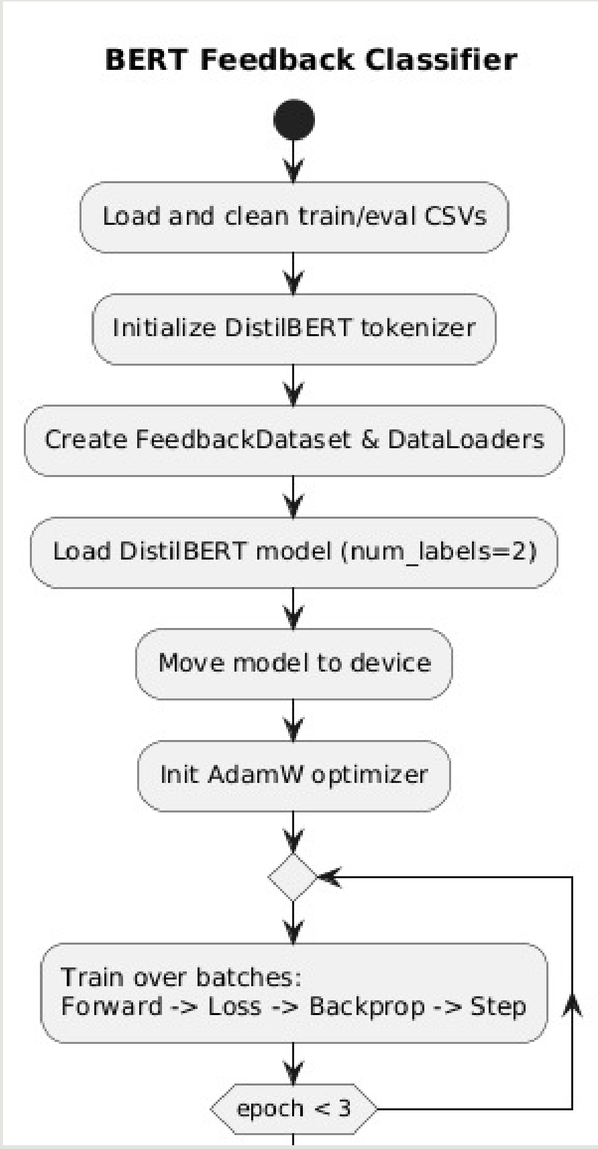
Training and Evaluation

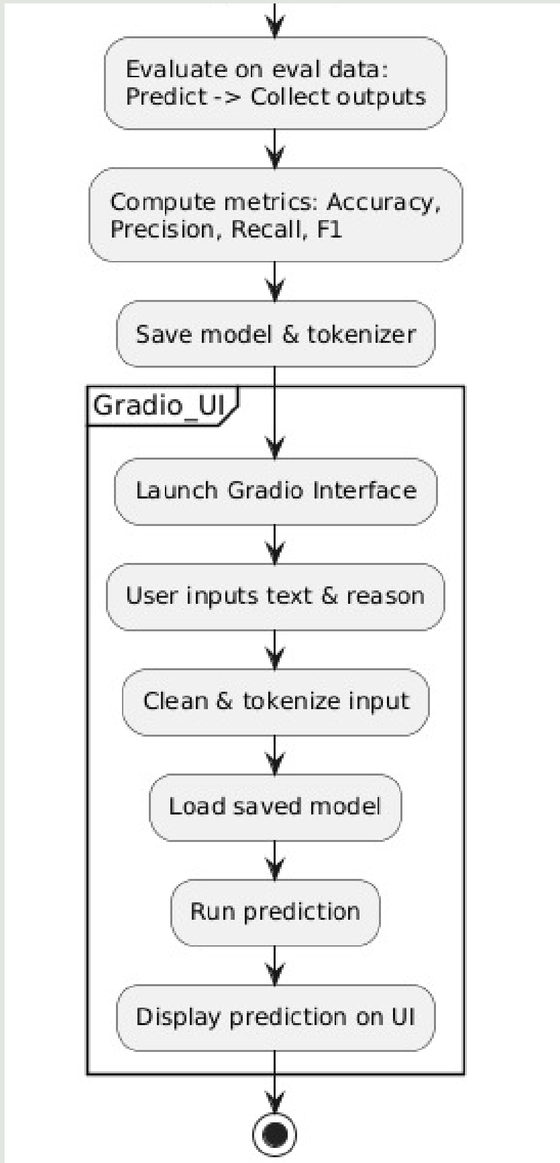
- Trained on a labeled dataset using PyTorch with a batch size of 8 for 3 epochs.
- Evaluated using standard classification metrics: accuracy, precision, recall, and F1-score.

Deployment

- The trained model is integrated into a Gradio interface (myzoom_gradio_app.py) for real-time validation.
- Users enter the feedback and reason, and receive instant feedback on alignment.

FLOW CHART


BERT Feedback Classifier
Load and clean train/eval CSVs
Initialize DistilBERT tokenizer
Create FeedbackDataset & DataLoaders
Load DistilBERT model (num_labels=2)
Move model to device
Init AdamW optimizer
Train over batches:
Forward -> Loss -> Backprop -> Step
epoch < 3


Evaluate on eval data:
Predict -> Collect outputs
Compute metrics: Accuracy,
Precision, Recall, F1
Save model & tokenizer
Gradio_UI
Launch Gradio Interface
User inputs text & reason
Clean & tokenize input
Load saved model
Run prediction
Display prediction on UI

RESULT

The model was evaluated on a separate test dataset after training. The results are as follows:

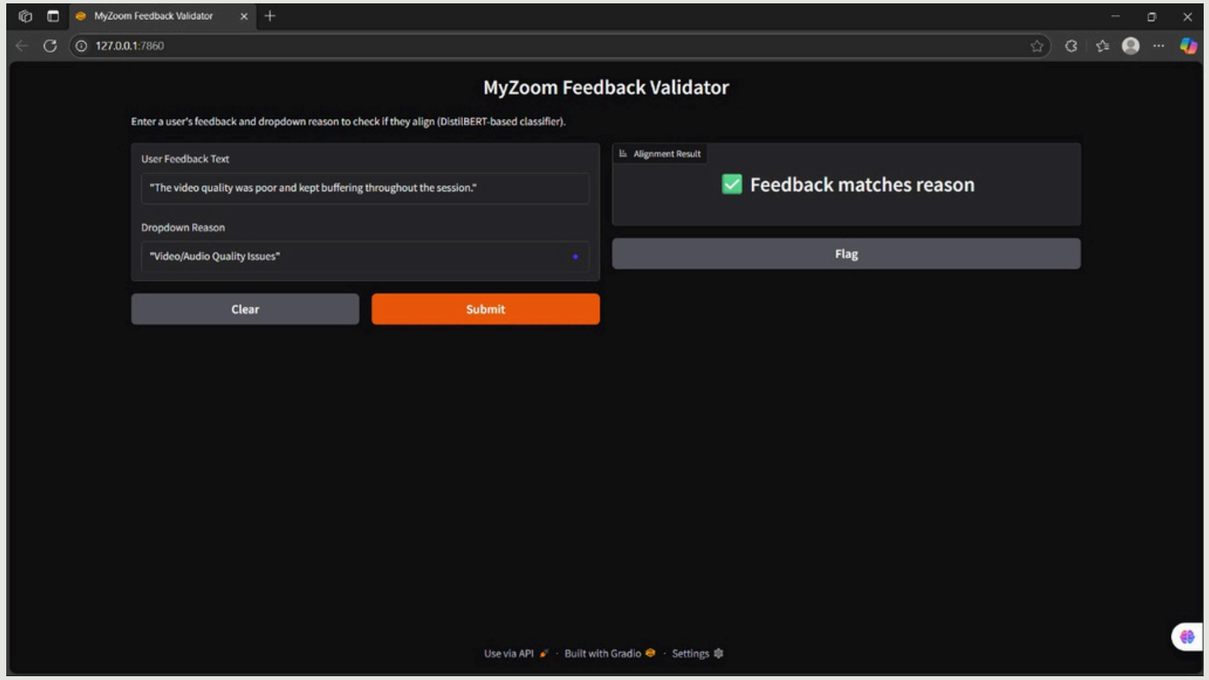
- Accuracy: ~0.85
- Precision: ~0.86
- Recall: ~0.84
- F1 Score: ~0.85

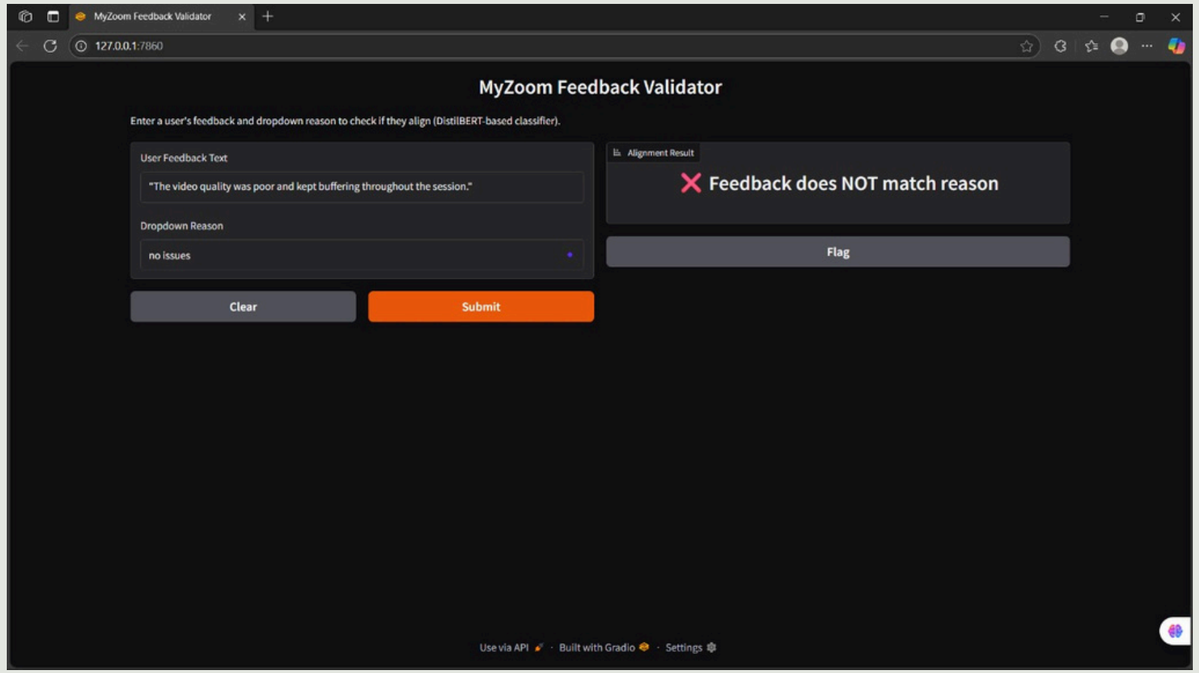
These metrics demonstrate the model’s strong performance in identifying whether feedback aligns with the selected reason, supporting its use in production environments.

OUTPUT

```
C:\Windows\System32\cmd.e

Be aware, overflowing tokens are not returned for the setting you have chosen, i.e. sequence pairs with the 'longest_first' truncation strategy. So the retu
rned list will always be empty even if some tokens have been removed.
Be aware, overflowing tokens are not returned for the setting you have chosen, i.e. sequence pairs with the 'longest_first' truncation strategy. So the retu
rned list will always be empty even if some tokens have been removed.
Be aware, overflowing tokens are not returned for the setting you have chosen, i.e. sequence pairs with the 'longest_first' truncation strategy. So the retu
rned list will always be empty even if some tokens have been removed.
Be aware, overflowing tokens are not returned for the setting you have chosen, i.e. sequence pairs with the 'longest_first' truncation strategy. So the retu
rned list will always be empty even if some tokens have been removed.
Be aware, overflowing tokens are not returned for the setting you have chosen, i.e. sequence pairs with the 'longest_first' truncation strategy. So the retu
rned list will always be empty even if some tokens have been removed.
Be aware, overflowing tokens are not returned for the setting you have chosen, i.e. sequence pairs with the 'longest_first' truncation strategy. So the retu
rned list will always be empty even if some tokens have been removed.
Be aware, overflowing tokens are not returned for the setting you have chosen, i.e. sequence pairs with the 'longest_first' truncation strategy. So the retu
rned list will always be empty even if some tokens have been removed.
Be aware, overflowing tokens are not returned for the setting you have chosen, i.e. sequence pairs with the 'longest_first' truncation strategy. So the retu
rned list will always be empty even if some tokens have been removed.
Be aware, overflowing tokens are not returned for the setting you have chosen, i.e. sequence pairs with the 'longest_first' truncation strategy. So the retu
rned list will always be empty even if some tokens have been removed.
Be aware, overflowing tokens are not returned for the setting you have chosen, i.e. sequence pairs with the 'longest_first' truncation strategy. So the retu
rned list will always be empty even if some tokens have been removed.
Be aware, overflowing tokens are not returned for the setting you have chosen, i.e. sequence pairs with the 'longest_first' truncation strategy. So the retu
rned list will always be empty even if some tokens have been removed.
Be aware, overflowing tokens are not returned for the setting you have chosen, i.e. sequence pairs with the 'longest_first' truncation strategy. So the retu
rned list will always be empty even if some tokens have been removed.
Be aware, overflowing tokens are not returned for the setting you have chosen, i.e. sequence pairs with the 'longest_first' truncation strategy. So the retu
rned list will always be empty even if some tokens have been removed.
Be aware, overflowing tokens are not returned for the setting you have chosen, i.e. sequence pairs with the 'longest_first' truncation strategy. So the retu
rned list will always be empty even if some tokens have been removed.
Xet Storage is enabled for this repo, but the 'hf_xet' package is not installed. Falling back to regular HTTP download. For better performance, install the
package with: `pip install huggingface_hub[hf_xet]` or `pip install hf_xet`
model.safetensors: 100%|          | 268M/268M [01:58<00:00, 2.25MB/s]
Some weights of DistilBertForSequenceClassification were not initialized from the model checkpoint at distilbert-base-uncased and are newly initialized: ['c
lassifier.bias', 'classifier.weight', 'pre_classifier.bias', 'pre_classifier.weight']
You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.
Epoch 1: 100%|          | 1383/1383 [1:17:35<00:00,  3.37s/it, loss=0.754]
Epoch 2: 100%|          | 1383/1383 [1:06:52<00:00,  2.90s/it, loss=0.773]
Epoch 3: 100%|          | 1383/1383 [1:17:31<00:00,  3.36s/it, loss=0.0507]

Accuracy: 0.9661
Precision: 0.9266
Recall: 0.9757
F1 Score: 0.9505

D:\PROJECTS\nlp2.0>
```

OUTPUT


MyZoom Feedback Validator
127.0.0.1:7860
MyZoom Feedback Validator
Enter a user's feedback and dropdown reason to check if they align (DistilBERT-based classifier).
User Feedback Text
"The video quality was poor and kept buffering throughout the session."
Dropdown Reason
"Video/Audio Quality Issues"
Clear
Submit
Alignment Result
Feedback matches reason
Flag
Use via API · Built with Gradio · Settings


MyZoom Feedback Validator
127.0.0.1:7860
MyZoom Feedback Validator
Enter a user's feedback and dropdown reason to check if they align (DistilBERT-based classifier).
User Feedback Text
"The video quality was poor and kept buffering throughout the session."
Dropdown Reason
no issues
Clear
Submit
Alignment Result
Feedback does NOT match reason
Flag
Use via API · Built with Gradio · Settings

CONCLUSION

- This project successfully demonstrates the use of transformer-based models in validating the contextual alignment of user feedback in EdTech systems. By automating the validation process, the system enhances the reliability of collected feedback, supports better platform decisions, and reduces manual moderation efforts. Future improvements could include multilingual support and fine-tuning with larger datasets for even greater accuracy.

Thank You