

Title: Advanced Sentence Transformation Classification

Objective:

Develop an advanced NLP model that classifies **sentence transformation types** (e.g., Active ↔ Passive, Direct ↔ Indirect Speech, Positive ↔ Negative statements), while focusing on **interpretability, generalization, and production-readiness**. The goal is to assess your skills in **text modeling, attention mechanisms, explainable AI (XAI), and model evaluation**.

Task Breakdown:

1. Dataset

- You will be provided (or asked to generate) a dataset of English sentence pairs with labeled transformation types:
 - Active to Passive
 - Passive to Active
 - Direct to Indirect Speech
 - Indirect to Direct Speech
 - Positive to Negative
 - Negative to Positive

Each row will consist of:

Original Sentence | Transformed Sentence | Label

2. Model Development

- Build a model that takes a **transformed sentence** and classifies the type of transformation applied.
- Utilize advanced NLP architectures (e.g., fine-tuned BERT, RoBERTa, LSTM with attention, etc.).
- Incorporate **preprocessing**, **tokenization**, and **data augmentation** where helpful.

3. Data Splitting

- Apply a **stratified 60:20:20** train/validation/test split.

4. Explainability

- Provide clear explanations for:
 - Preprocessing steps
 - Choice of model architecture
 - Any attention layers or interpretability techniques used
- Use tools like **SHAP**, **LIME**, or **attention visualization** to explain predictions.

5. Metrics and Evaluation

- Report:
 - Accuracy
 - Precision/Recall/F1 Score (per class)
 - Confusion Matrix
- Perform **error analysis**: What types of transformations are most/least accurately detected and why?

6. Optional Bonus


- Create a simple **streamlit/gradio** app or API (FastAPI/Flask) that allows inputting a sentence and seeing:
 - Detected transformation type
 - Highlighted words contributing to the classification
 - Confidence score
-

Completion Criteria:

1. A trained classification model for sentence transformation detection.
 2. A modular, readable, and well-documented codebase (preferably in a Jupyter notebook + `.py` scripts).
 3. Explainable modeling decisions and interpretable outputs.
 4. Evaluation using multiple metrics and confusion matrix analysis.
 5. Demonstration of advanced ML/NLP practices (e.g., transfer learning, error analysis).
 6. Optional but desirable: lightweight deployment/demo via an app or API.
-

Skills Evaluated:

- Advanced NLP modeling (transformers, attention mechanisms)
- Preprocessing and data stratification
- Explainability and model transparency (SHAP/LIME/Attention)
- Software engineering practices (clean code, modularity)
- Analytical thinking (error analysis, metrics interpretation)
- Optional: ML deployment (gradio, streamlit, Flask/FastAPI)

Dataset Link:  immverse_ai_eval_dataset

(https://docs.google.com/spreadsheets/d/1_yA6Ut4nHh7BD7OcY6kadG19zCpql1-JRMDCRWePrqE/edit?usp=sharing)

Note: The dataset provided is a sample for reference purposes. You are required to create a similar dataset based on this format. The final dataset must contain a minimum of **200 rows**.