**Evidence Detection with DeBERTa**

A concise end-to-end workflow for detecting relevant evidence in text using [DeBERTa v3](https://huggingface.co/microsoft/deberta-v3-base). It includes data loading, synonym-based augmentation, hyperparameter tuning with Hyperopt, and model inference.

**Requirements Overview:**

* Python 3.10+
* pandas
* scikit-learn
* nltk
* datasets
* torch
* transformers
* hyperopt
* accelerate

Install everything via:

pip install -r requirements.txt

**Data and Attribution**

* Data: Place your train.csv, dev.csv, and test.csv in the data/ folder (or update paths in the notebook).
  + train.csv / dev.csv must have claim, evidence, label columns.
  + test.csv must have claim, evidence.
* Glove Data Source: <https://www.kaggle.com/datasets/thanakomsn/glove6b300dtxt>

**Usage**

1. Install dependencies:
2. pip install -r requirements.txt
3. Open and run the notebook:
   * Make changes to global variables to update paths and hyperparameter search space if required
   * It loads data, performs optional augmentation, hyperparameter tuning, then trains a final model.
   * A file (e.g., best\_deberta\_model.pt) is saved with the best weights.

**Inference**

To load the saved model and run on new data (e.g. test.csv):

1. Skip the training cells in the notebook.
2. Load the model state dict from best\_deberta\_model.pt.
3. Generate predictions.
4. The notebook saves these predictions as a CSV with a single "prediction" column.