Cover Letter

Proposed Idea:

1. Incorporation of Prompts: Rather than directly leveraging raw data, we integrate contextual prompts to offer a clearer frame of reference for the model.
2. Zero-Shot Classification Using BERT: For the task of zero-shot classification, we employ the "facebook/bart-large-mnli" model. This approach assists us in obtaining pseudo labels for our dataset.
3. Prompt Integration Across Datasets: To ensure the model is well-acquainted with the context, we integrate the prompt not only into the training data but also into the development (dev_set) and test datasets. By doing so, we familiarize the model with the input format it will face during actual deployment. This strategic move aims to maintain a seamless alignment between the model's training behavior and its real-world application performance.
4. Model Construction and Training: After updating the training dataset with pseudo labels, we embark on building a robust machine learning model. It's crucial to scrutinize the dataset for any data imbalances. In cases where imbalances are detected, we harness the SMOTE technique to rectify them. Further, to optimize the model's F1 score, rigorous hyperparameter tuning using RandomizedSearchCV is undertaken.

Prediction Comparison:

Dev_Set	Movies		News	
	New_Test Classifier	Step 3 Classifier	New_Test Classifier	Step 3 Classifier
	0.89	0.761	0.746	0.779

Analysis:

1. I've generated pseudo labels for two datasets: 2,999 for the Movies dataset and 6,000 for the News dataset. Given that the dev set was incorporated during the model training, it wouldn't be appropriate to assess performance solely based on the F1 score from the dev set. Instead, I've relied on a validation dataset obtained through a train-test split, which yielded an F1 score of 0.7 for Movies and 0.72 for News after hyperparameter tuning. The results from my new classifier align relatively closely with those from step 3, but there's still room for improvement. A couple of potential reasons for this discrepancy could be:
    a. Not properly designed the machine learning model framework compared with CNN in step 3.
    b. The Gradient Boosting algorithm might achieve enhanced performance if provided with a larger labeled dataset, especially for the Movies category which had only 2,999 labels.
2. On a brighter note, leveraging the "bart-large MNLI" model and incorporating prompts has enabled a standard Gradient Boosting classifier to predict labels for 120,000 documents with an F1 score of 0.72 in under 2 minutes.