DETECTION OF AI-GENERATED TEXTS

300232-ยงสุข

DATA

Sample_Submission.csv

test.jsonl

train.jsonl


```
1 {"id": "test_0001", "abstract": "This paper addresses the problem of evaluating learning systems in safety critical domains such as autonomous driving, where failures can have catastrophic consequences.", "summary1": "The field of few-shot learning has recently seen substantial advancements.The recent advancements in few-shot learning are related to its ability to handle large and diverse datasets. The field has made significant progress in recent years as it has been able to achieve better results than traditional methods on several benchmark datasets.", "summary2": "We show that rare but catastrophic failures may be missed entirely by random testing, which poses issues for safe deployment. Our proposed approach for adversarial testing fixes this."}
```

```
1 {"id": "train_0001", "abstract": "This paper addresses the problem of evaluating learning systems in safety critical domains such as autonomous driving, where failures can have catastrophic consequences.", "summary": "We show that rare but catastrophic failures may be missed entirely by random testing, which poses issues for safe deployment. Our proposed approach for adversarial testing fixes this.", "generated": "This paper addresses the problem of evaluating learning systems in safety critical domains such as autonomous driving, where failures can have catastrophic consequences.The paper \"This paper: Evaluating Learning Systems in Safety Critical Domains\" presents an evaluation study of learning systems in safety-critical domains such as autonomous driving, where failures can have catastrophic consequences. The study evaluates the performance of various learning systems in different safety contexts, and highlights the importance of evaluating the effectiveness of learning systems in these domains. The paper also discusses potential challenges and limitations in the evaluation of learning systems."}
```

PREPROCESSING DATA

TRAIN SET

abstract	**LABEL 0**
summary	**LABEL 0**
generated	**LABEL 1**

TEST SET

abstract

summary1

summary2

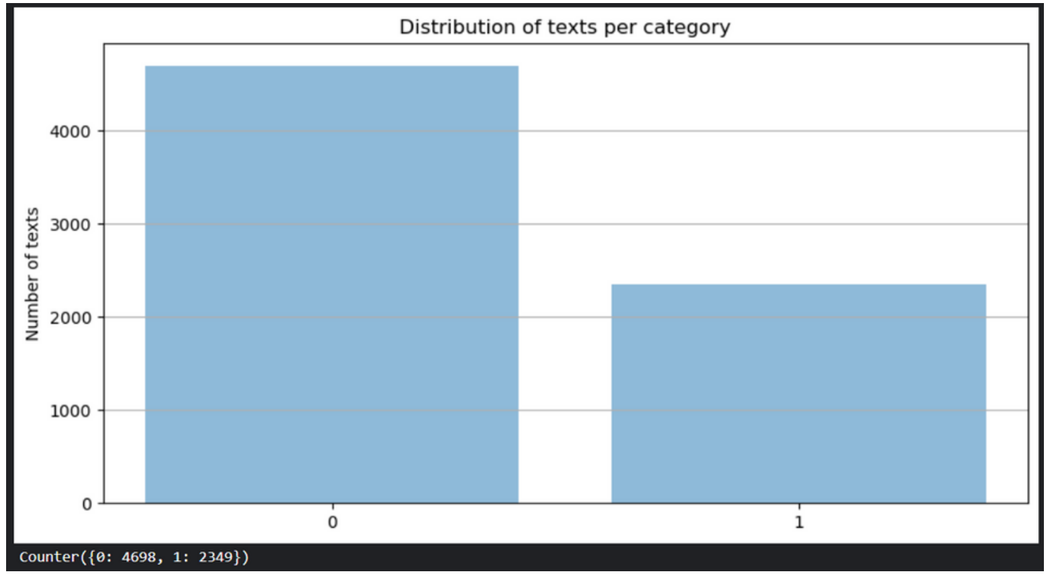
0 = HUMAN 1 = AI

MORE DATA FROM

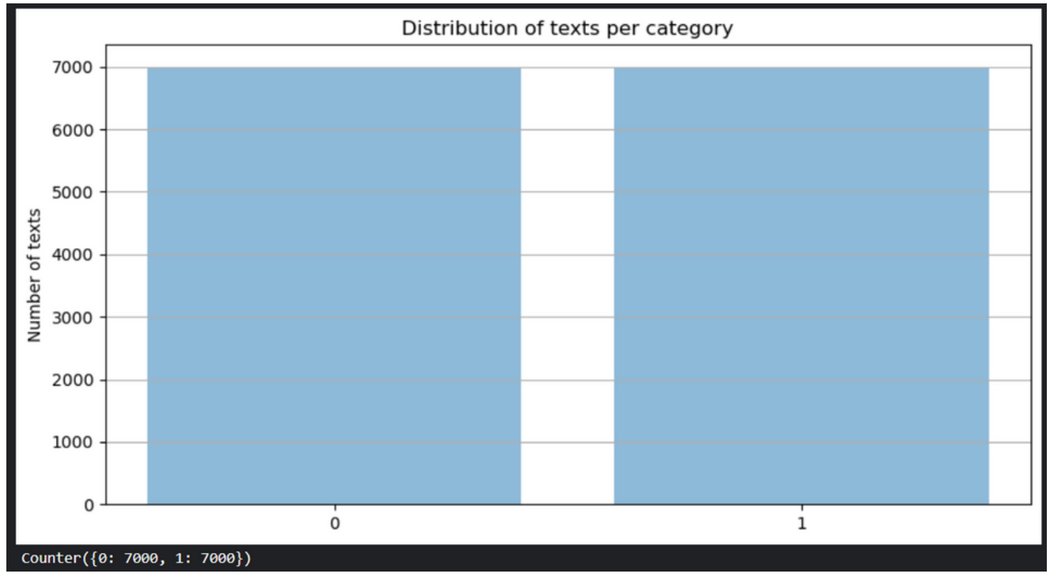
Datasets: aadityaubhat / GPT-wiki-intro like 7

EDA

BEFORE

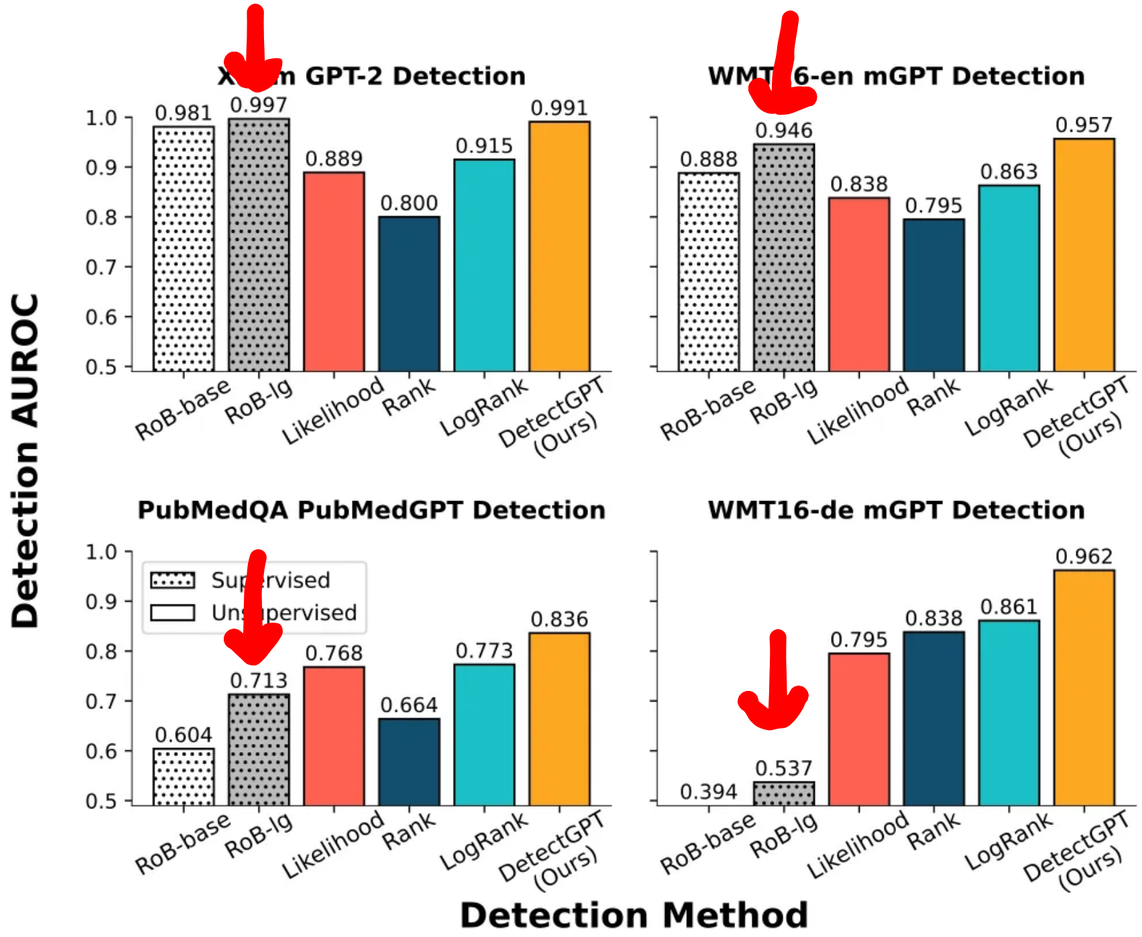


AFTER



0 = HUMAN 1 = AI

MODEL



ROBERTA LARGE

```
# Initialize the tokenizer
tokenizer = RobertaTokenizerFast.from_pretrained("roberta-large")
```

```
model = TFRobertaForSequenceClassification.from_pretrained("roberta-large", num_labels=2)
```

Ref: https://medium.com/@TheHaseebHassan/detectgpt-detecting-ai-generated-text-a0284f1d05de

ACCURACY

```
Epoch 1/3
705/705 [==============================] - 890s 1s/step - loss: 0.0641 - accuracy: 0.9805 - val_loss: 0.0098 - val_accuracy: 0.9965
Epoch 2/3
705/705 [==============================] - 755s 1s/step - loss: 0.0280 - accuracy: 0.9945 - val_loss: 0.0058 - val_accuracy: 0.9993
Epoch 3/3
705/705 [==============================] - 755s 1s/step - loss: 0.0106 - accuracy: 0.9988 - val_loss: 0.0013 - val_accuracy: 1.0000
```

BEFORE

```
Epoch 1/3
1400/1400 [==============================] - 1016s 654ms/step - loss: 0.0508 - accuracy: 0.9806 -
val_loss: 0.1973 - val_accuracy: 0.9357
Epoch 2/3
1400/1400 [==============================] - 856s 611ms/step - loss: 0.0330 - accuracy: 0.9921 -
val_loss: 0.0115 - val_accuracy: 0.9971
Epoch 3/3
1400/1400 [==============================] - 854s 610ms/step - loss: 0.0191 - accuracy: 0.9959 -
val_loss: 0.0123 - val_accuracy: 0.9971
```

AFTER

Submission and Description	Private Score	Public Score	Selected
more-data-submission.csv Complete · 12h ago	1	1	☑
submission02.csv Complete · 1d ago	0.92366	0.95384	☑
Sample_Submission.csv Complete · 1d ago	0.48091	0.53846	☐
submission01.csv Complete · 1d ago	1	1	☑