PROMPT-BASED TEXT MATCHING METHODS FOR FAKE NEWS STANCE DETECTION

CMPT 413 FINAL PROJECT

GROUP AWSL

ZEYONG JIN, YUQING WU, ZHI FENG

TABLE OF CONTENTS

- I. Introduction
  - 1. Problem Statement
  - 2. Input & Output
- II. Related Work
  - 1. Text Matching
  - 2. Pre-trained NLP Models
- III. Data Fake News Challenge (FNC-1)
- IV. Approach
  - 1. Pre-processing FNC corpus
  - 2. Preparing BERT base model
  - 3. Fine-tuning BERT model
- V. Evaluation Accuracy, Precision, F1, AUC, ROC, FNC Score
- VI. Experiment Baselines vs BERT-based Method (Ours)
- VII. Conclusion

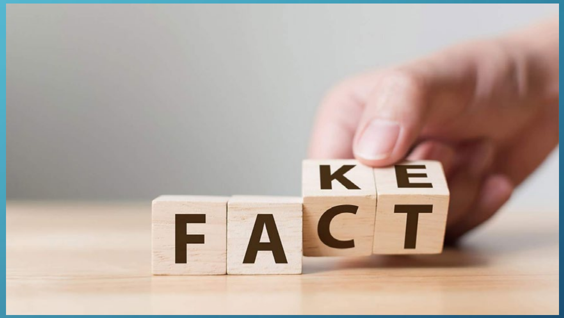


Image Credit: https://www.txstate.edu/commonexperience/pastsitearchives/2019-2020/stories/blog/2020-03-30-truth-about-fake-

PART I: INTRODUCTION

- 1. Problem Statement
  - Exploring how AI could be leveraged to combat fake news.

Image Credit:
http://www.fakenewschallenge.org/

PART I: INTRODUCTION

- 2. Input
  - A headline and a body text - either from the same news article or from two different articles.
- 3. Output
  - Classify the stance of the body text relative to the claim made in the headline into one of four categories:
    - Agrees: The body text agrees with the headline.
    - Disagrees: The body text disagrees with the headline.
    - Discusses: The body text discuss the same topic as the headline, but does not take a position
    - Unrelated: The body text discusses a different topic than the headline

PART I: INTRODUCTION

- 4. Example headline
  - “Robert Plant Ripped up $800M Led Zeppelin Reunion Contract”
- 5. Example snippets from body texts and correct classifications
  - Agree:
    - “… Led Zeppelin’s Robert Plant turned down £500 MILLION to reform supergroup. …”
  - Disagree:
    - “… No, Robert Plant did not rip up an $800 million deal to get Led Zeppelin back together. …”
  - Discusses:
    - “… Robert Plant reportedly tore up an $800 million Led Zeppelin reunion deal. …”
  - Unrelated:
    - “… Richard Branson’s Virgin Galactic is set to launch SpaceShipTwo today. …”

PART II: RELATED WORK

- 1. Traditional Text Matching Models
  - TF-IDF, BM25, Latent Dirichlet Allocation (LDA)
- 2. DNN-based Text Matching Models
  - Deep Structured Semantic Models (DSSM) (Huang et al., 2013)
  - Architecture-I (ACRI) (Hu et al., 2014)
  - Tree-structured LSTM (Tree-LSTM) (Tai et al., 2015)
- 3. Pre-trained-based Text Matching Models
  - ELMo (Peters et al., 2018)
  - GPT (Radford et al., 2018)
  - BERT (Devlin et al., 2018)

PART III: DATA

- Fake News Challenge (FNC-1)
    - 1. Training Set: [headline, body text, label]
        - Pairs of headline and body text with the appropriate class label for each.
    - 2. Testing Set: [headline, body text]
        - Pairs of headline and body text without class labels used to evaluate systems.

```
Headline,Body ID,Stance
Police find mass graves with at least '15 bodies' near Mexico town where 43 students disappeared after police clash,712,unrelated
Hundreds of Palestinians flee floods in Gaza as Israel opens dams,158,agree
"Christian Bale passes on role of Steve Jobs, actor reportedly felt he wasn't right for part",137,unrelated
HBO and Apple in Talks for $15/Month Apple TV Streaming Service Launching in April,1034,unrelated
Spider burrowed through tourist's stomach and up into his chest,1923,disagree
'Nasa Confirms Earth Will Experience 6 Days of Total Darkness in December' Fake News Story Goes Viral,154,agree
"Accused Boston Marathon Bomber Severely Injured In Prison, May Never Walk Or Talk Again",962,unrelated
Identity of ISIS terrorist known as 'Jihadi John' reportedly revealed,2033,unrelated
Banksy 'Arrested & Real Identity Revealed' Is The Same Hoax From Last Year,1739,agree
British Aid Worker Confirmed Murdered By ISIS,882,unrelated
Gateway Pundit,2327,discuss
"Woman detained in Lebanon is not al-Baghdadi's wife, Iraq says",1468,agree
Kidnapped Nigerian schoolgirls: Government claims ceasefire deal with Boko Haram that will bring missing girls home,1003,unrelated
"No, that high school kid didn't make $72 million trading stocks",2132,unrelated
"Soon Marijuana May Lead to Ticket, Not Arrest, in New York",47,discuss
Vandals add rude paint job to $2.5m Bugatti (but luckily for the owner it all turned out to be a hoax),615,unrelated
Boko Haram Denies Nigeria Cease-Fire Claim,2463,discuss
```

PART IV: APPROACH

- 1. Pre-processing FNC corpus
- 2. Preparing BERT base model
- 3. Fine-tuning BERT model

PART IV: APPROACH

- 1. Pre-processing CSV
  - Join stances.csv and bodies.csv

```
       UniqueID  ...      Stance
0             0  ...   unrelated
1             1  ...   unrelated
2             2  ...   unrelated
3             3  ...   unrelated
4             4  ...   unrelated
...         ...  ...         ...
25408     25408  ...       agree
25409     25409  ...     discuss
25410     25410  ...    disagree
25411     25411  ...    disagree
25412     25412  ...       agree

[25413 rows x 4 columns]
```

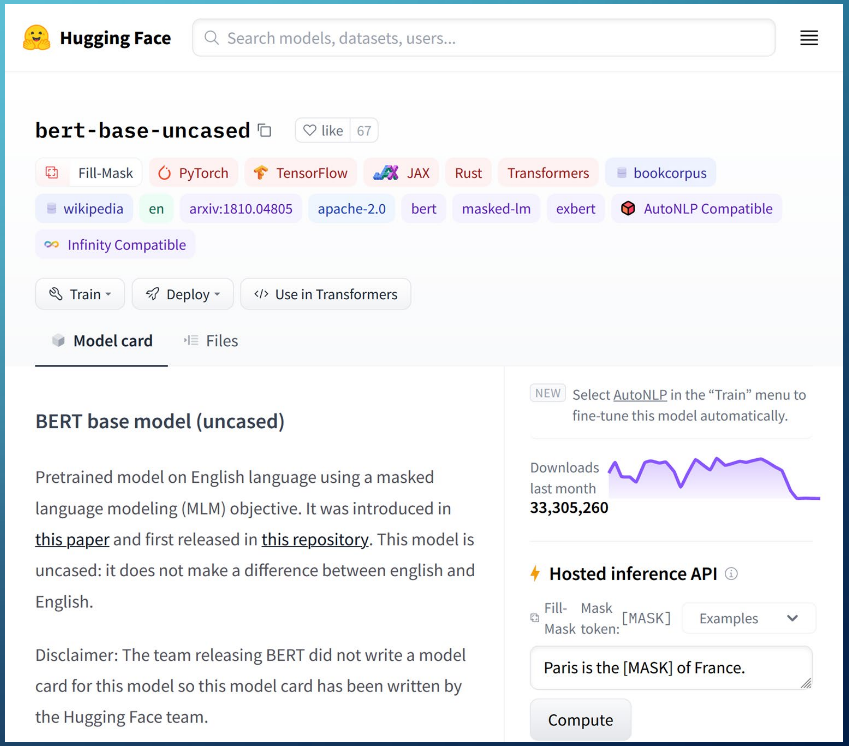
```
                                               articleBody
Body ID
1        Al-Sisi has denied Israeli reports stating tha…
2        A bereaved Afghan mother took revenge on the T…
3        CNBC is reporting Tesla has chosen Nevada as t…
12       A 4-inch version of the iPhone 6 is said to be…
19       GR editor’s Note\n\nThere are no reports in th…
…                                                    …
2582     Congressional Republicans, evidently hoping th…
2583     Did Obamacare work?\n\nIt’s worth reflecting u…
2584     Millions may lose coverage next year if Congre…
2585     Come November, the grim trudge across the incr…
2586     Remember how much Republicans wanted to repeal…

[904 rows x 1 columns]
```

```
       Body ID                                            Mixed  Stance
40261      752  Did Kim Yo-Jong Take Kim Jong Un’s Role? North…       0
33359     1166  ISIS beheads American photo-journalist James W…       0
27620     2272  Rumor: Seth Rogen to Appear as Woz in Sony's S…       0
41445      493  Purported ISIS claim: Jordan airstrike kills f…       0
14838     1605  Catholic priest dies for 48 mins, wakes up cla…       0
…          …                                                 …      …
41291     1427  Mexico hit by student massacre: At least 17 an…       3
14137     2403  New Audio Allegedly Captures Moment Michael Br…       0
5354      1744  Apple Watch to Be Shower-Proof, Have 100,000 A…       0
15195     1489  11 jetliners 'missing' after Islamist takeover…       0
29406      649  NASA Raises Doubts About Reports of Nicaraguan…       0

[49972 rows x 3 columns]
```

PART IV: APPROACH

- 2. Preparing BERT model
  - Using Google Colab Pro + HuggingFace Transformer

```
[ ]: max_length = 512
     tokenizer = BertTokenizerFast.from_pretrained(local_dir, local_files_only=True,
      do_lower_case=True)
     model = BertForSequenceClassification.from_pretrained(local_dir,
      local_files_only=True, num_labels=4).to(device)

[ ]: #
     # model_name = "bert-base-uncased"
     # max_length = 512
     # tokenizer = BertTokenizerFast.from_pretrained(model_name, do_lower_case=True)
     # model = BertForSequenceClassification.from_pretrained(model_name,
      num_labels=4).to(dev)
```



PART IV: APPROACH

- 3. Fine-tuning BERT model
  - Method 1: Tokenize bodies-headline texts

```
[ ]: # tokenize the dataset, truncate when passed `max_length`,
     # and pad with 0's when less than `max_length`
     train_encodings = tokenizer(train_texts, truncation=True, padding=True,␣
      ↪max_length=max_length)
     valid_encodings = tokenizer(valid_texts, truncation=True, padding=True,␣
      ↪max_length=max_length)
     test_encodings = tokenizer(test_texts, truncation=True, padding=True,␣
      ↪max_length=max_length)
```

PART IV: APPROACH

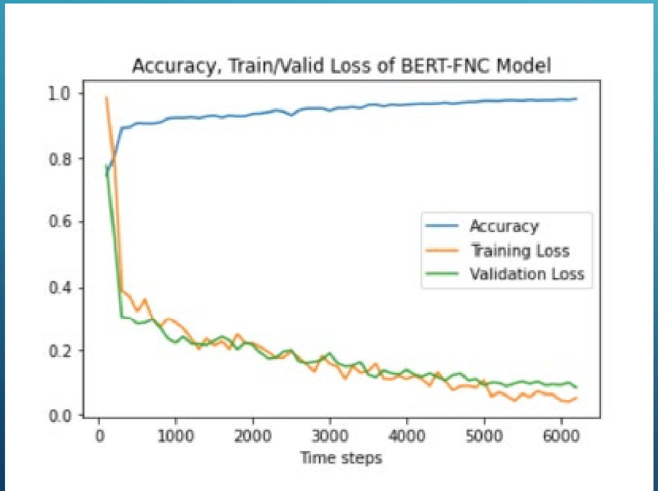
- 3. Fine-tuning BERT model
    - Method 2: Generate PyTorch DataSet and DataLoader

```
[ ]: class FakeNewsDataset(torch.utils.data.Dataset):
         def __init__(self, encodings, labels):
             self.encodings = encodings
             self.labels = labels

         def __getitem__(self, idx):
             item = {k: torch.tensor(v[idx]) for k, v in self.encodings.items()}
             item["labels"] = torch.tensor(self.labels[idx])
             return item

         def __len__(self):
             return len(self.labels)
```

```
[ ]: from torch.utils.data import DataLoader
     test_dataloader = DataLoader(test_dataset, batch_size=128, shuffle=False)
     test_dataloader
```

PART IV: APPROACH

- 3. Fine-tuning BERT model
  - Method 3: Run fine-tuning trainer
    - Set Training Arguments
    - Set Trainer
    - Define Training Metrics



Train BERT model

```
[ ] from sklearn.metrics import accuracy_score, precision_score, f1_score, confusion_matrix

    def compute_metrics(pred):
      labels = pred.label_ids
      preds = pred.predictions.argmax(-1)
      # calculate accuracy using sklearn's function
      acc = accuracy_score(labels, preds)
      return {
          'accuracy': acc,
      }
```

```
[ ] training_args = TrainingArguments(
        output_dir='./results',          # output directory
        num_train_epochs=3,              # total number of training epochs
        per_device_train_batch_size=16,  # batch size per device during training
        per_device_eval_batch_size=64,   # batch size for evaluation
        warmup_steps=500,                # number of warmup steps for learning rate scheduler
        weight_decay=0.01,               # strength of weight decay
        logging_dir='./logs',            # directory for storing logs
        load_best_model_at_end=True,     # load the best model when finished training (default metric is loss)
        # but you can specify `metric_for_best_model` argument to change to accuracy or other metric
        logging_steps=100,               # log & save weights each logging_steps
        save_steps=1000,
        evaluation_strategy="steps",     # evaluate each `logging_steps`
    )
```

```
[ ] trainer = Trainer(
        model=model,                         # the instantiated Transformers model to be trained
        args=training_args,                  # training arguments, defined above
        train_dataset=train_dataset,         # training dataset
        eval_dataset=valid_dataset,          # evaluation dataset
        compute_metrics=compute_metrics,     # the callback that computes metrics of interest
    )
```

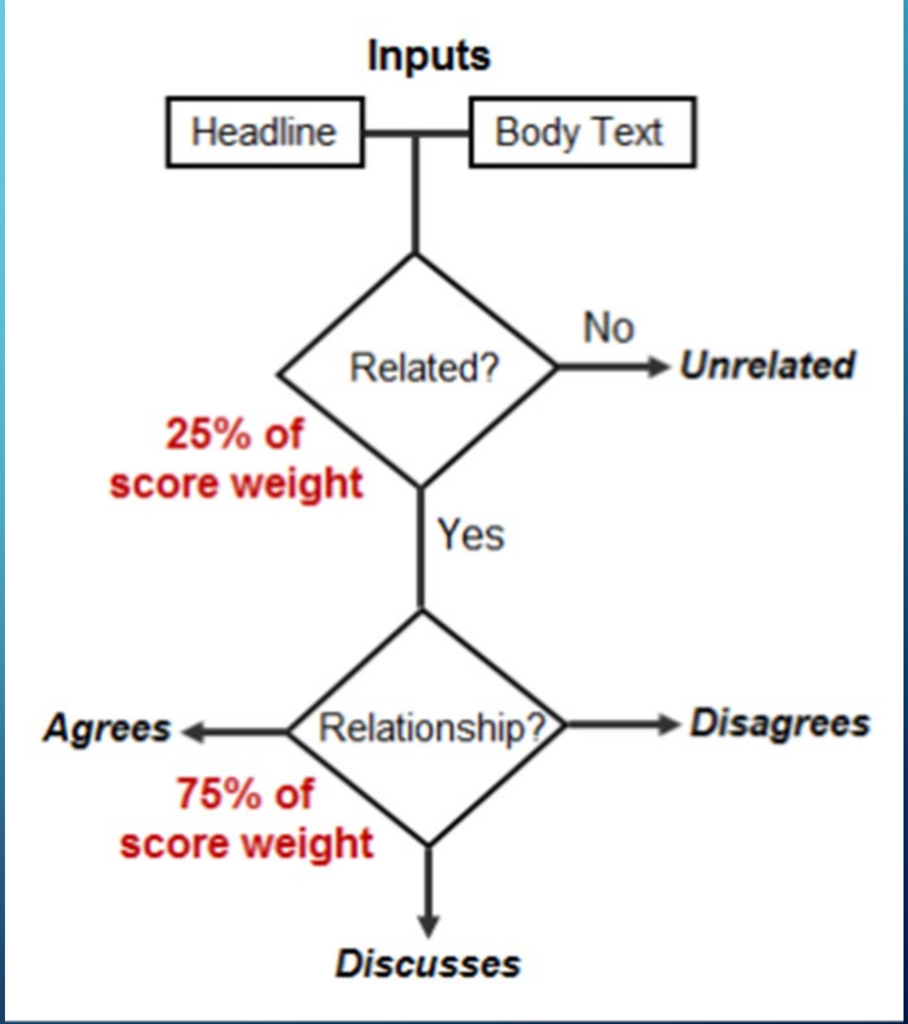
```
[ ] trainer.train()
```

PART V: EVALUATION

- 1. Metrics
  - Accuracy, Precision, F1, AUC, ROC
  - FNC Score (Domain-Specific)

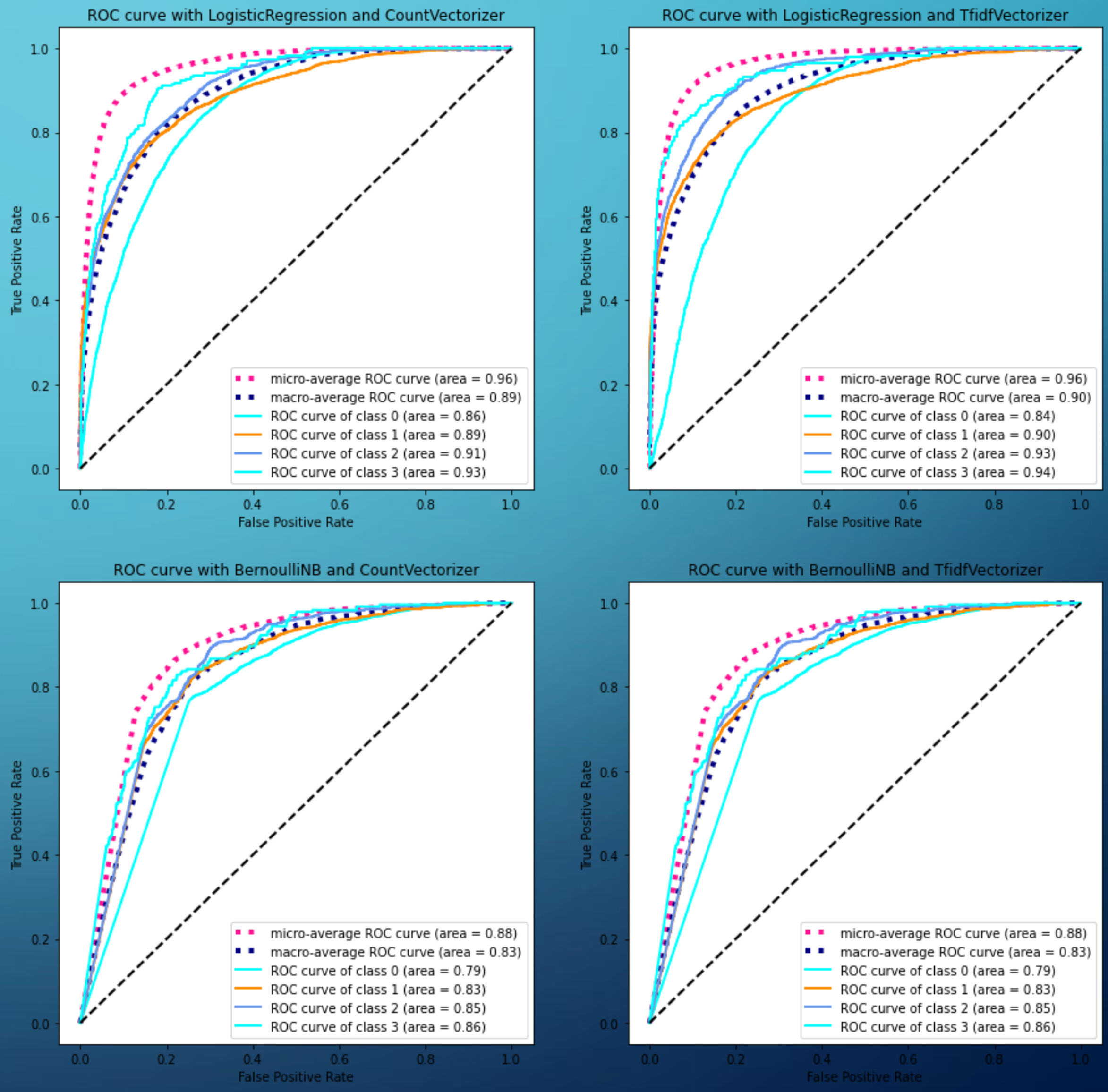
PART V: EVALUATION

- 2. Flowchart
  - Description:
    - Concretely, if a [headline, body text] pair in the test set has the target label unrelated, a team's evaluation score will be incremented by 0.25 if labels the pair as unrelated.
    - If the [headline, body text] test pair is related, a team's score will be incremented by 0.25 if it labels the pair as any of the three classes: agrees, disagrees, or discusses.
    - The team's evaluation score will so be incremented by an additional 0.75 for each related pair if gets the relationship right by labeling the pair with the single correct class: agrees, disagrees, or discusses.



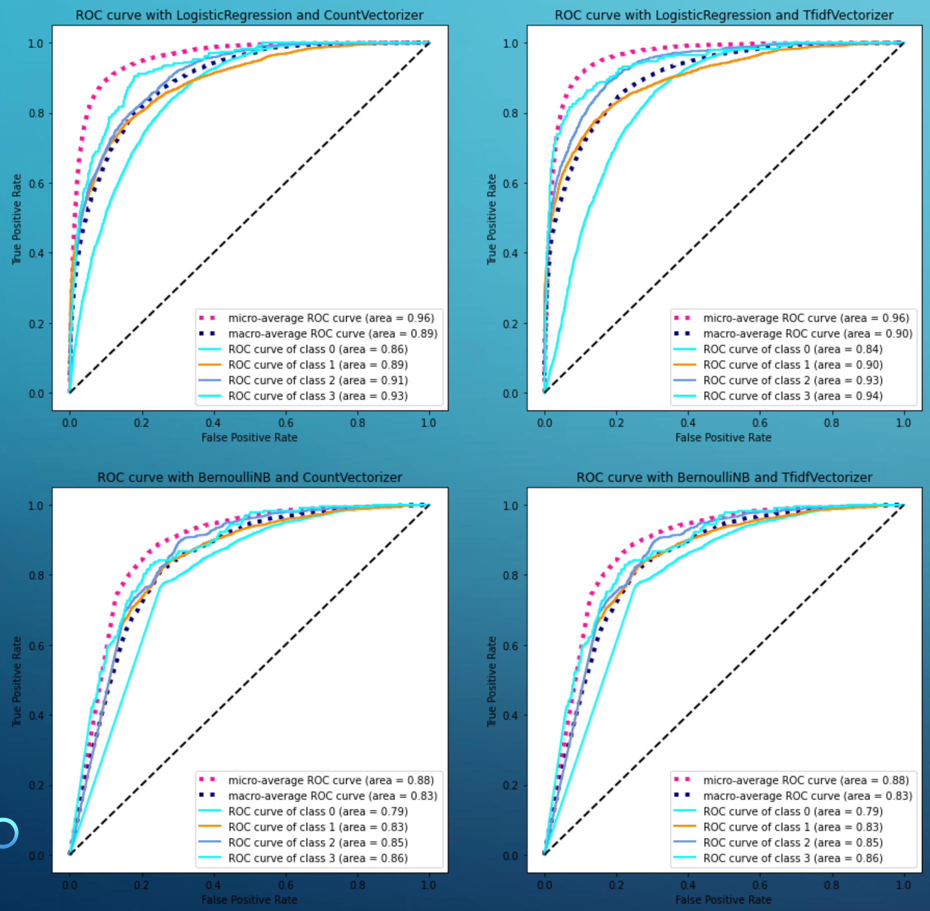
PART VI: EXPERIMENT

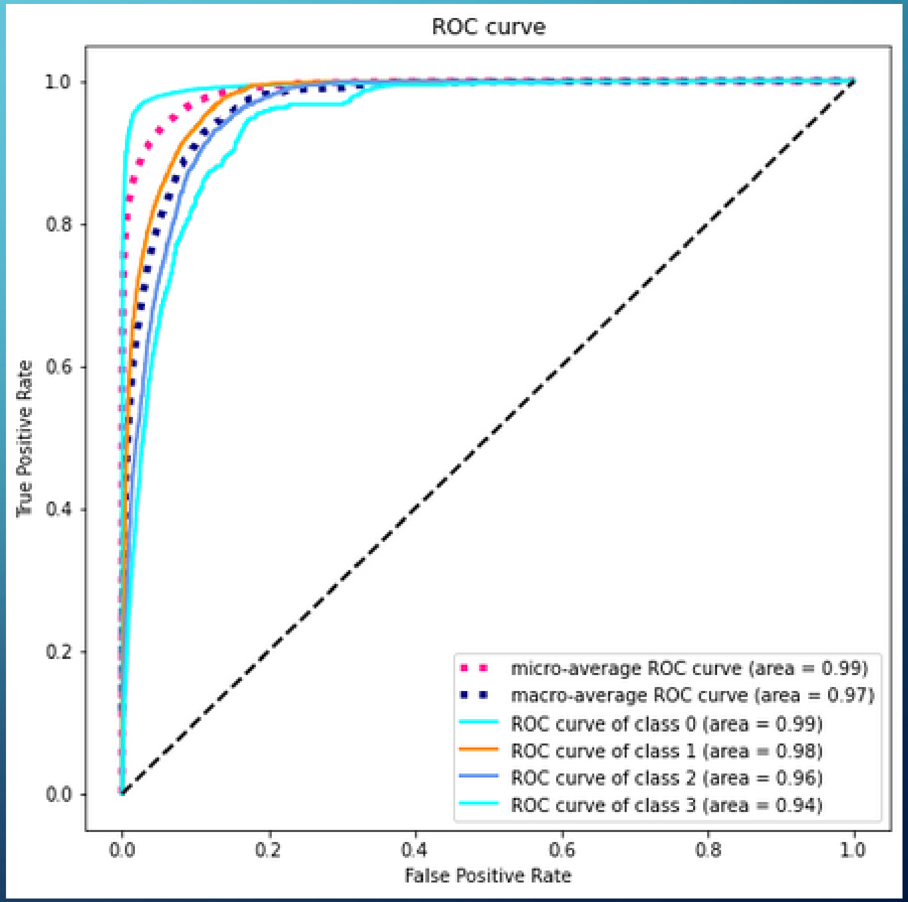
- 1. Baselines (Statistical ML)
  - Bernoulli Naïve Bayes + TF-IDF
  - Bernoulli Naïve Bayes + Count
  - Logistics Regression + TF-IDF
  - Logistics Regression + Count



PART VI: EXPERIMENT

- 2. BERT-based Method (Ours)





PART VI: EXPERIMENT

- 3. Comparison

	Accuracy	Precision	F1	Micro-ROC	Marco-ROC	# FNC	% FNC
LR-Count	71.27%	59.20%	62.58%	96.00%	89.00%	4815.5	41.33%
LR-TFIDF	59.00%	57.94%	58.30%	96.00%	90.00%	5026.75	43.14%
NB-Count	66.51%	56.59%	60.90%	87.00%	84.00%	4834.75	41.49%
NB-TFIDF	66.51%	56.59%	60.90%	87.00%	84.00%	4834.75	41.49%
BERT	**90.37%**	**90.60%**	**90.43%**	**99.00%**	**97.00%**	**10068.25**	**86.41%**

Table 1: Performance Metrices

PART VII: CONCLUSION

- Achievement
  - Good Result: Accuracy > 90%
- Limitations
  - Not very good FNC Score
- Future Work
  - Learn more about PTMs

REFERENCES

1. Blasingame, D. (2020). *Blocks of Fact And Fake.* The Truth about Fake News. Texas State University. Retrieved December 3, 2021, from https://www.txstate.edu/commonexperience/pastsitearchives/2019-2020/stories/blog/2020-03-30-truth-about-fake-news.html
2. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805.*
3. Fake News Challenge. (2021). *Fake News.* Exploring how artificial intelligence technologies could be leveraged to combat fake news. Fake News Challenge. Retrieved December 3, 2021, from http://www.fakenewschallenge.org/
4. Baotian Hu, Zhengdong Lu, Hang Li, and Qingcai Chen. 2014. Convolutional neural network architectures for matching natural language sentences. In *Advances in neural information processing systems*, pages 2042–2050.
5. Po-Sen Huang, Xiaodong He, Jianfeng Gao, Li Deng, Alex Acero, and Larry Heck. 2013. Learning deep structured semantic models for web search using clickthrough data. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, pages 2333–2338.

REFERENCES

6. Matthew E Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. *arXiv preprint arXiv:1802.05365.*

7. Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. Improving language understanding by generative pre-training.

8. Kai Sheng Tai, Richard Socher, and Christopher D Manning. 2015. Improved semantic representations from tree-structured long short-term memory networks. *arXiv preprint arXiv:1503.00075.*

THANK YOU FOR WATCHING!

PRESENTED BY

ZEYONG JIN, YUQING WU, ZHI FENG

CMPT 413 GROUP AWSL

SCHOOL OF COMPUTING SCIENCE, SIMON FRASER UNIVERSITY