



Claim Detection in Tweets

鄒詠霖
Yung-Ling Tsou

指導教授: 許永真博士
Advisor: Jane Yung-Jen Hsu, Ph.D.

2023/7/26



Outline

- Introduction
- Research Questions
- Methodology
- Experiments
- Conclusion

Introduction

Misinformation

- **Misinformation:** false information that is spread, regardless of whether there is intent to mislead
- **Fact-Checking:** the action of confirming the truth of an claim
- **Claim:** an assertion of something as a fact

BREAKING: Wisconsin has more votes than people who are registered to vote.

Total number of registered voters:
3,129,000

Total number of votes cast:
3,239,920

This is direct evidence of fraud.

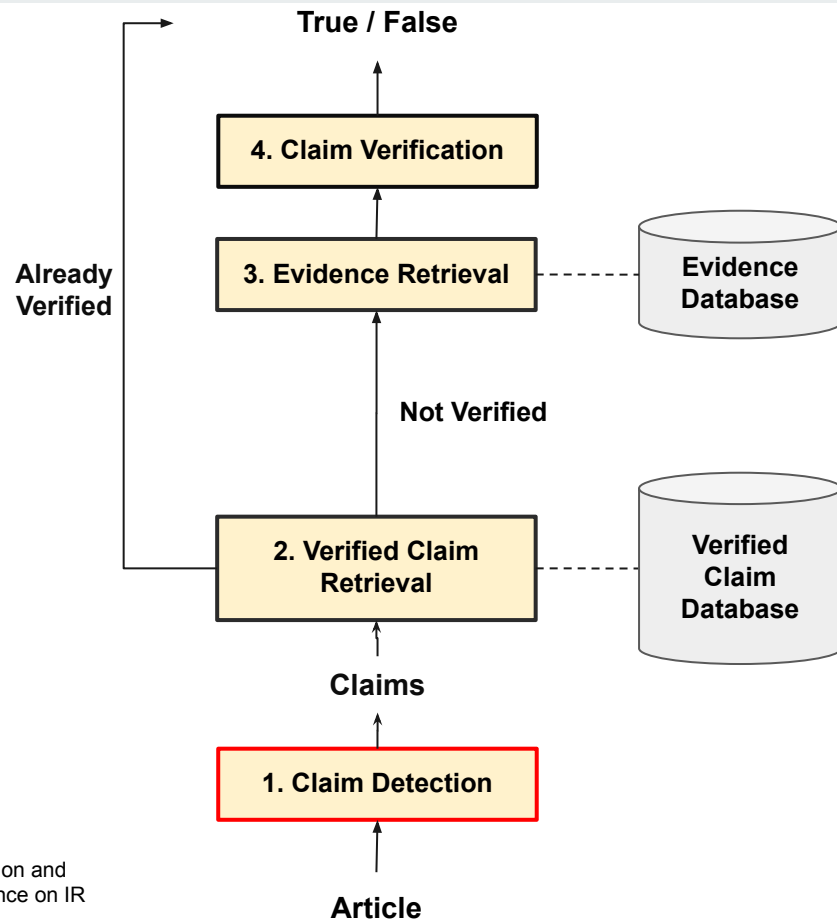


fact check by [PolitiFact](#)

Introduction

Automatic Fact-Checking pipeline

- Machines require a standard process
- Each stage can be a independent subtask



Introduction

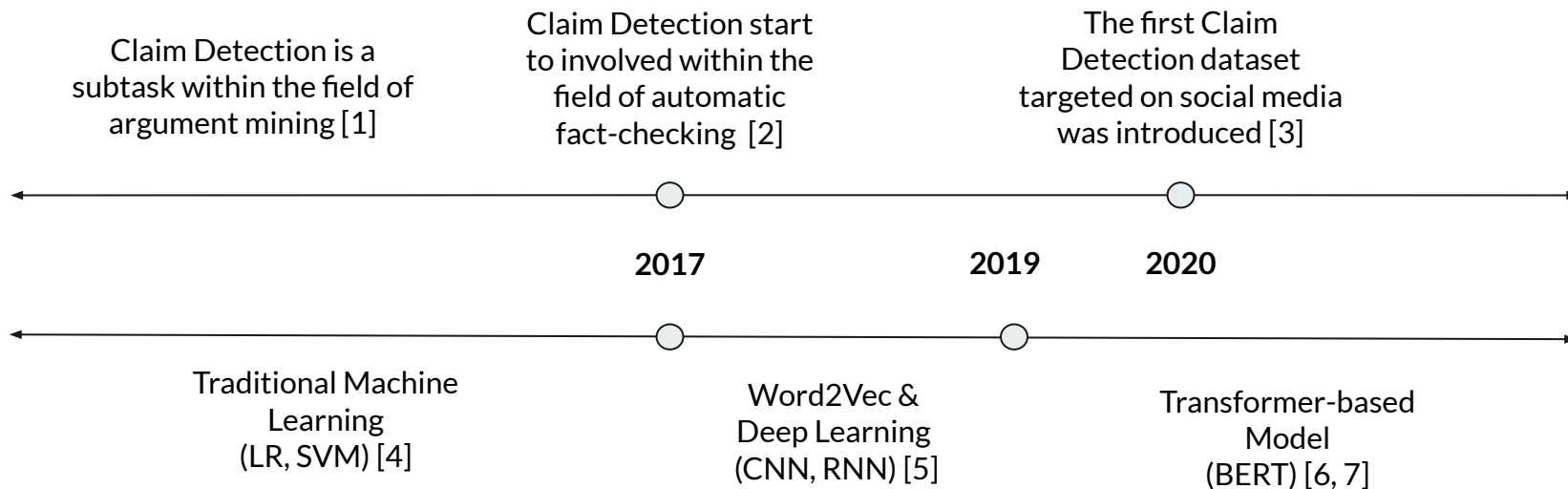
Why Tweet



- No limit when posting on Social Media
- Plenty amount of Social Media post
- Nowadays, people consume information from social media heavily
- Dataset are mostly Tweet

Introduction

Related Work



[1] Ehud Aharoni, et al. 2014. A Benchmark Dataset for Automatic Detection of Claims and Evidence in the Context of Controversial Topics. ACL.

[2] Naeemul Hassan, et al. 2017. Toward Automated Fact-Checking: Detecting Check-worthy Factual Claims by ClaimBuster. ACM.

[3] Shaden Shaar, et al. 2020. Overview of CheckThat! 2020 English: Automatic Identification and Verification of Claims in Social Media. CLEF 2020 Working Notes.

[4] R Levy, et al. 2014. Context Dependent Claim Detection. ACL

[5] Johannes Daxenberger, et al. 2017. What is the Essence of a Claim? Cross-Domain Claim Identification. ACL.

[6] [Check_square at CheckThat! 2020: Claim Detection in Social Media via Fusion of Transformer and Syntactic Features](#)

[7] Shreya Gupta, et al. 2021. LESA: Linguistic Encapsulation and Semantic Amalgamation Based Generalised Claim Detection from Online Content. EACL.

Introduction

Claim Detection in Tweets

- Given a tweet, determines whether it contains any claims.

data:

India's gift of 100,000 COVID-19 vaccines arrived Barbados earlier today. This was a very special moment for all Barbadians and I want to thank Prime Minister Modi for his quick, decisive, and magnanimous action in allowing us to be the beneficiary of these vaccines.

<https://t.co/cSCb40c2mt>

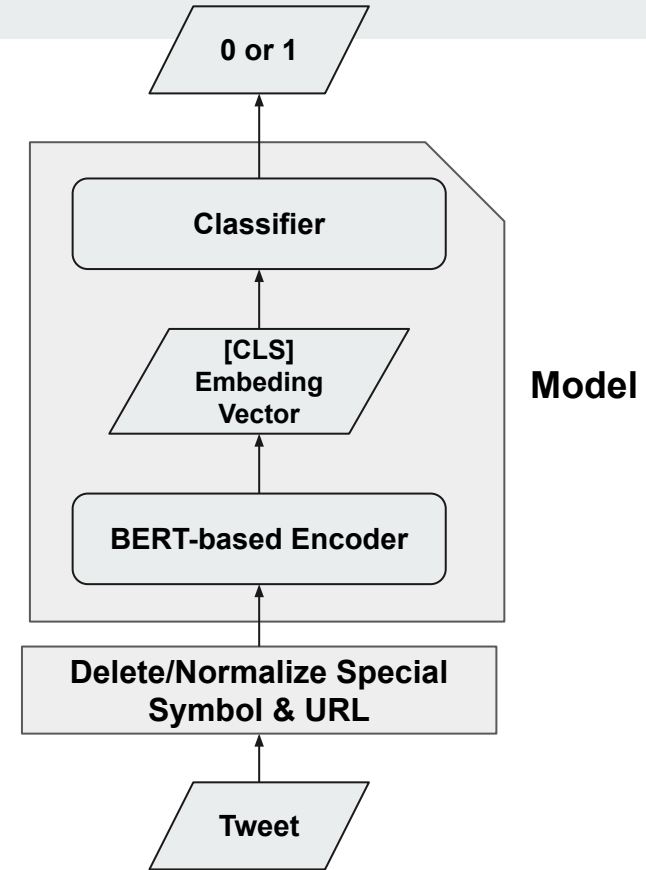


label: 1 (contain claim)

Introduction

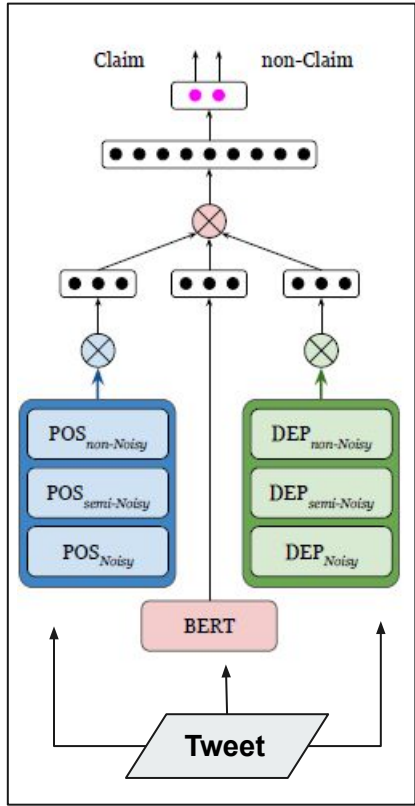
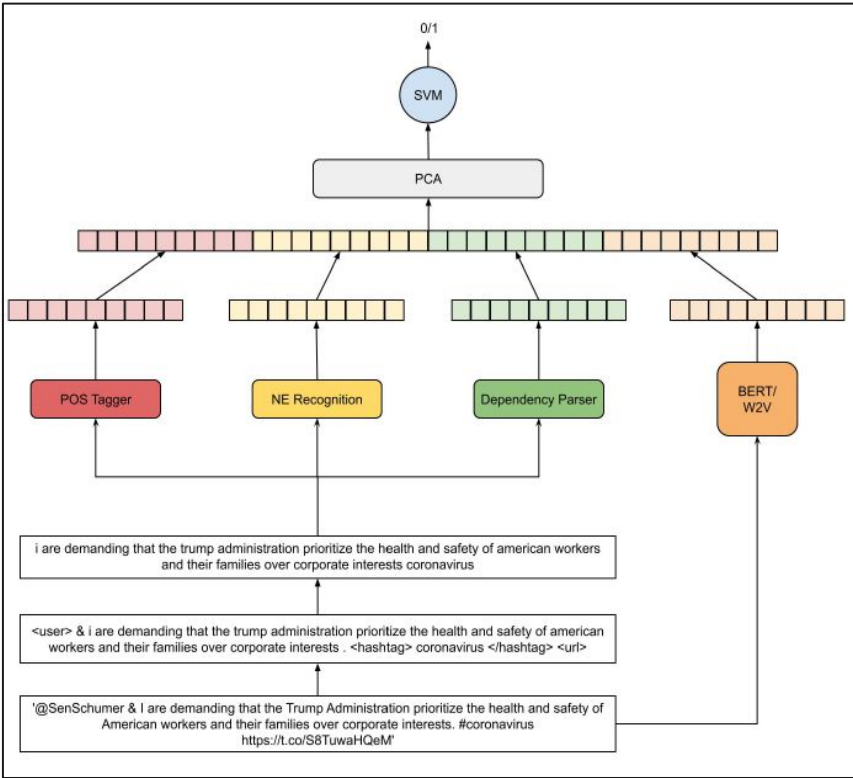
Current Approach

- BERT-based Encoder + Classifier structure
- Some simple data preprocess(Delete/Normalize Special Symbol & URL)
- Bert-based Encoder for capturing semantic features



[1] [Overview of the CLEF-2022 CheckThat! Lab Task 1 on Identifying Relevant Claims in Tweets](#)

[2] Amelie Wühl and Roman Klinger. 2021. Claim Detection in Biomedical Twitter Posts. In Proceedings of the 20th Workshop on Biomedical Language Processing, Association for Computational Linguistics.




[1] [Check square at CheckThat! 2020: Claim Detection in Social Media via Fusion of Transformer and Syntactic Features](#)

[2] Shreya Gupta, et al. 2021. LESA: Linguistic Encapsulation and Semantic Amalgamation Based Generalised Claim Detection from Online Content. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: EACL 2021


Research Questions

- Tweet's noisy formatting
 - url, @, #, emoji
 - informal writing style

#Enugu govt takes delivery of #COVID19vaccines
Enugu State Government alongside Ebonyi and Anambra
states, on Thursday, took delivery of the first batch of
#oxfordastrazeneca COVID-19 vaccines from @NphcdaNG, at
Enugu Airport... <https://t.co/srx1CWZqtK> @Coal_City
#COVID19 <https://t.co/PrO9GSiXuv>



had someone on sc for 2 days before i deleted her ass. she had
to have been on some stupid shit. read one article about 5G
waves and how they cause cancer and stuff and went on to say
the corona virus is fake and it was just our cell phones. is u
stupid



Research Questions

- Tweet's noisy formatting
 - url, @, #, emoji
 - informal writing style



Research Question 1

Does reducing noise in the formatting of tweets help improve the performance of claim detection?

#Enugu govt takes delivery of #COVID19vaccines
Enugu State Government alongside Ebonyi and Anambra
states, on Thursday, took delivery of the first batch of
#oxfordastrazeneca COVID-19 vaccines from @NphcdaNG, at
Enugu Airport... <https://t.co/srx1CWZqtK> @Coal_City
#COVID19 <https://t.co/PrO9GSiXuv>



had someone on sc for 2 days before i deleted her ass. she had
to have been on some stupid shit. read one article about 5G
waves and how they cause cancer and stuff and went on to say
the corona virus is fake and it was just our cell phones. is u
stupid



Research Questions

- Improper definition of Claim Detection
 - Binary classification on tweet-level
 - Claim may only be a part of tweet

- Without exactly claim =>

Can't continue automatic fact-checking

India's gift of 100,000 COVID-19 vaccines arrived Barbados earlier today. This was a very special moment for all Barbadians and I want to thank Prime Minister Modi for his quick, decisive, and magnanimous action in allowing us to be the beneficiary of these vaccines.
<https://t.co/cSCb40c2mt>



Research Questions

- Improper definition of Claim Detection
 - Binary classification on tweet-level
 - Claim may only be a part of tweet

- Without exactly claim =>

Can't continue automatic fact-checking



Research Question 2

Does reducing the inconsistency between the level of input and claim help improve the performance of claim detection?

India's gift of 100,000 COVID-19 vaccines arrived Barbados earlier today. This was a very special moment for all Barbadians and I want to thank Prime Minister Modi for his quick, decisive, and magnanimous action in allowing us to be the beneficiary of these vaccines.

<https://t.co/cSCb40c2mt>



Methodology

Reducing Noise in the Formatting of Tweets

- ~~url, @, #, emoji~~
- informal writing style

- Rule-based approach to eliminate url, emoji, #

#Enugu govt takes delivery of #COVID19vaccines
Enugu State Government alongside Ebonyi and Anambra states, on Thursday, took delivery of the first batch of #oxfordastrazeneca COVID-19 vaccines from @NphcdaNG, at Enugu Airport... <https://t.co/srx1CWZqtK> @Coal_City #COVID19 <https://t.co/PrO9GSiXuv>



Rule-based

Enugu govt takes delivery of COVID19vaccines
Enugu State Government alongside Ebonyi and Anambra states, on Thursday, took delivery of the first batch of oxfordastrazeneca COVID-19 vaccines from NphcdaNG, at Enugu Airport... Coal_City COVID19

Methodology

Reducing Noise in the Formatting of Tweets

- url, @, #, emoji
- informal writing style

- Rewrite by GPT to deal with abbreviation and informal writing style

had someone on sc for 2 days before i deleted her ass. she had to have been on some stupid shit. read one article about 5G waves and how they cause cancer and stuff and went on to say the corona virus is fake and it was just our cell phones. is u stupid
rewrite:

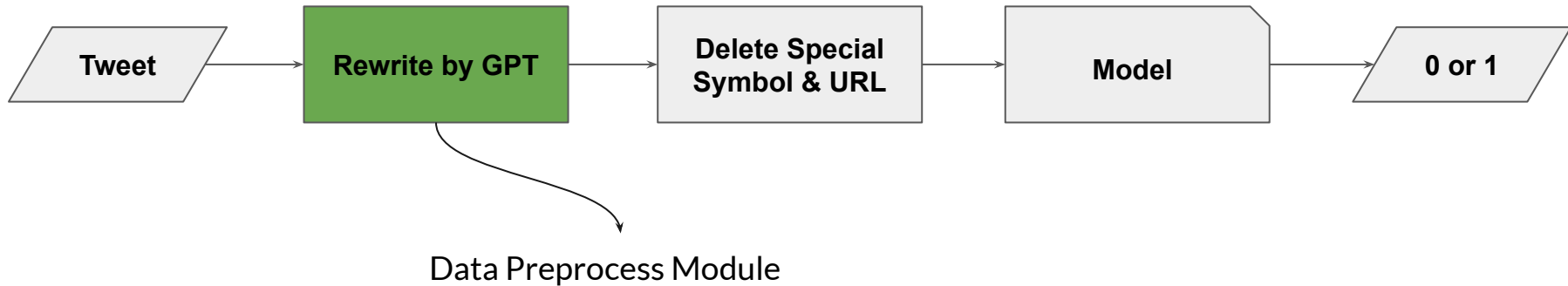


I had someone on Snapchat for two days before I removed them. They must have been on some ridiculous nonsense. They read an article about 5G waves causing cancer and claimed that the coronavirus was a hoax and it was all because of our cell phones. Are they lacking intelligence?

Methodology

Reducing Noise in the Formatting of Tweets

Training & Testing

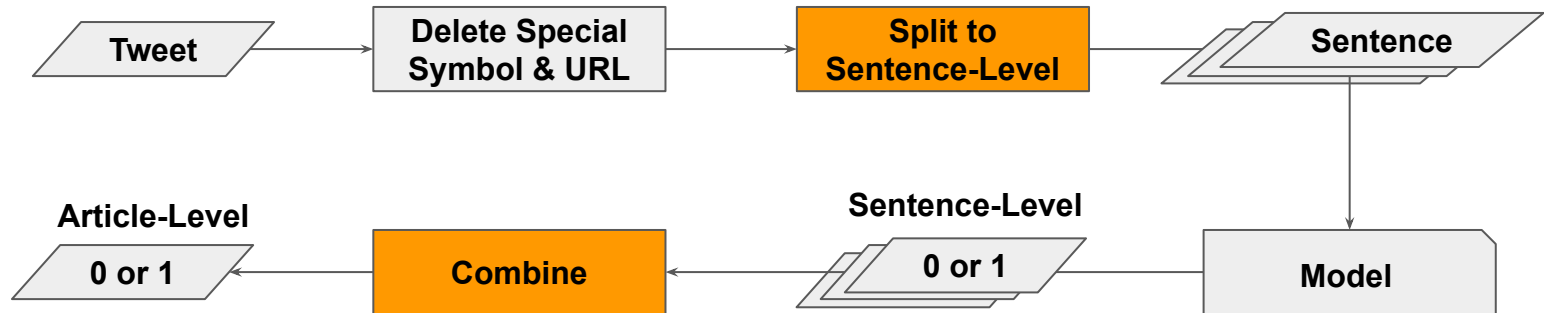


Methodology

Reducing the Inconsistency between the Level of Input and Claim

- Split tweet into sentences when **Testing**

Testing



Methodology

Reducing the Inconsistency between the Level of Input and Claim

- Split tweet into sentences when **Testing**

India's gift of 100,000 COVID-19 vaccines arrived Barbados earlier today. This was a very special moment for all Barbadians and I want to thank Prime Minister Modi for his quick, decisive, and magnanimous action in allowing us to be the beneficiary of these vaccines.



rule-based by [NLTK](#)

India's gift of 100,000 COVID-19 vaccines arrived Barbados earlier today.

This was a very special moment for all Barbadians and I want to thank Prime Minister Modi for his quick, decisive, and magnanimous action in allowing us to be the beneficiary of these vaccines.

Methodology

Reducing the Inconsistency between the Level of Input and Claim

- Leveraging SRL(Semantic Role Labeling), we can extract the semantic more precisely

India's gift of 100,000 COVID-19 vaccines arrived Barbados earlier today. This was a very special moment for all Barbadians and I want to thank Prime Minister Modi for his quick, decisive, and magnanimous action in allowing us to be the beneficiary of these vaccines.

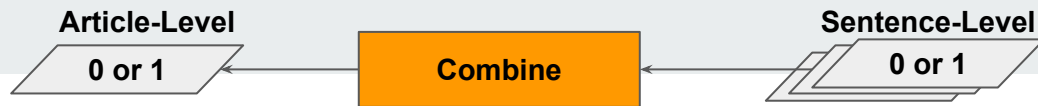


SRL by [AllenNLP](#)

India's gift of 100,000 COVID-19 vaccines arrived Barbados earlier today.

This was a very special moment for all Barbadians.

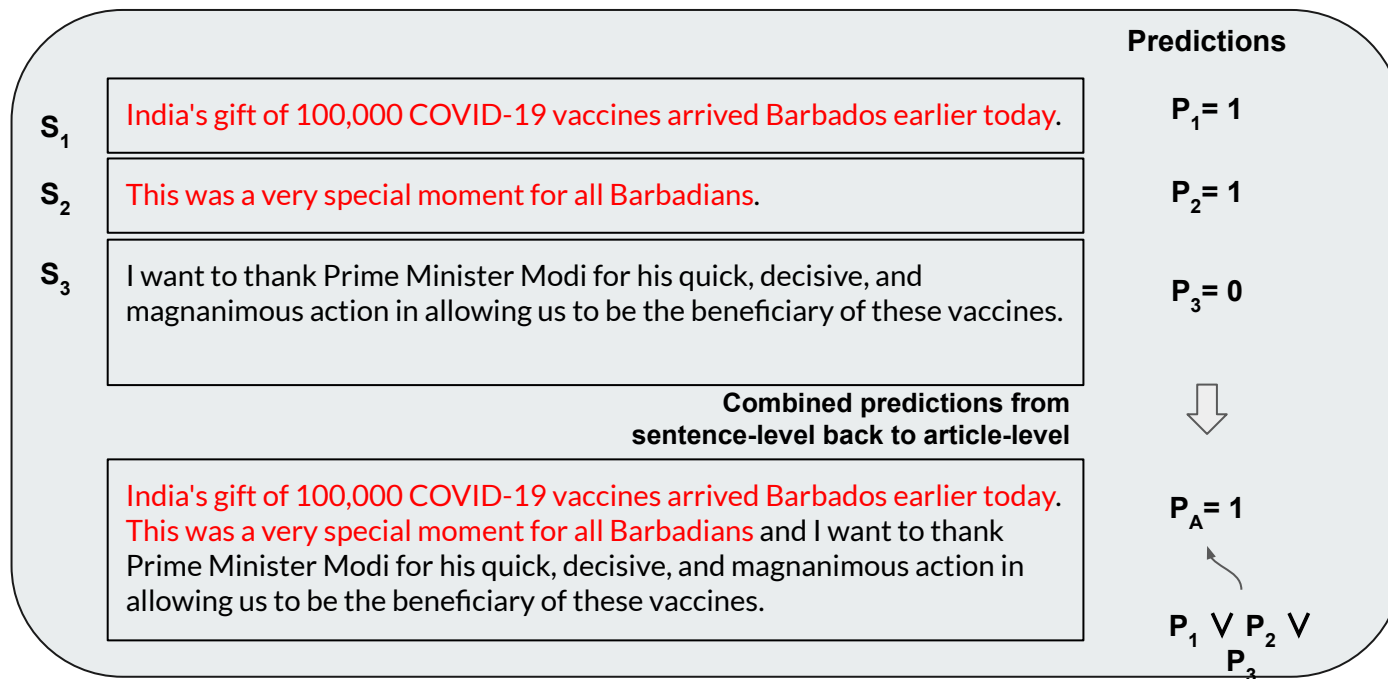
I want to thank Prime Minister Modi for his quick , decisive , and magnanimous action in allowing us to be the beneficiary of these vaccines.



Methodology

Reducing the Inconsistency between the Level of Input and Claim

- How to combine the predictions of N sentences
 - One sentence is predicted as claim => The tweet contain claim



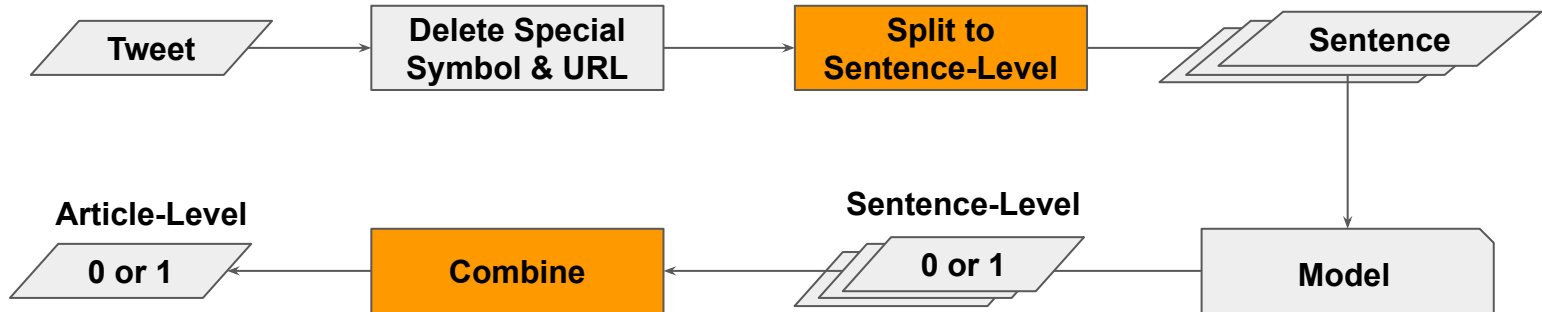
Methodology

Reducing the Inconsistency between the Level of Input and Claim

- Split tweet into sentences when **Testing**

How about Training

Testing



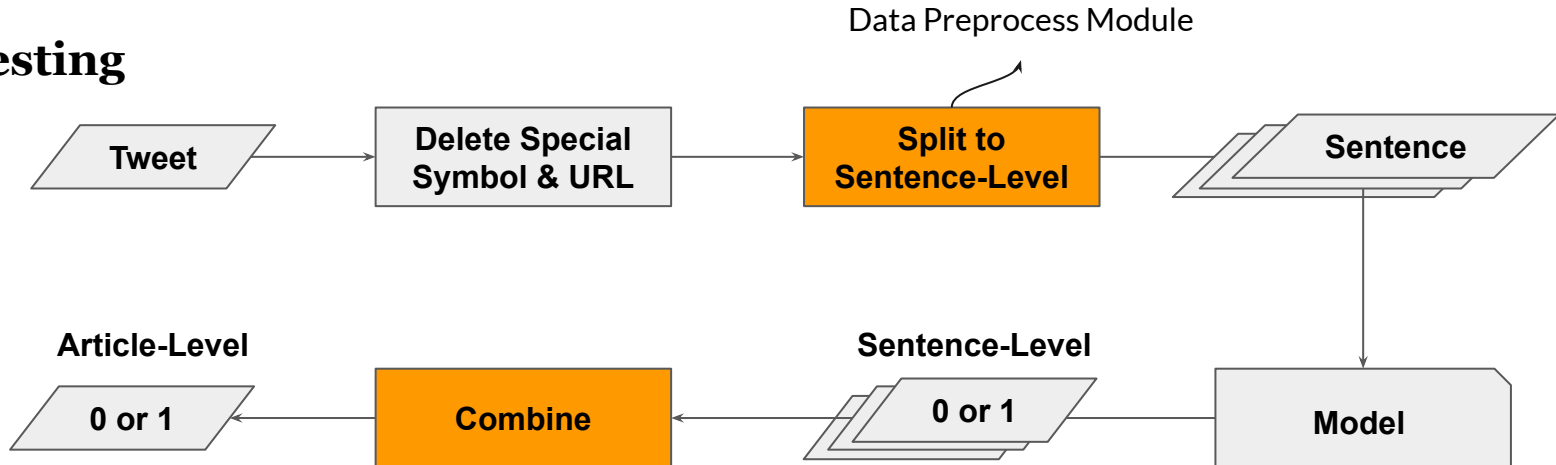
Methodology

Reducing the Inconsistency between the Level of Input and Claim

Training



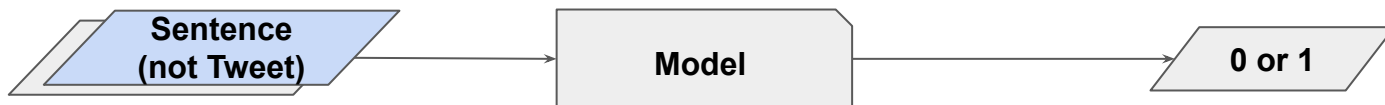
Testing



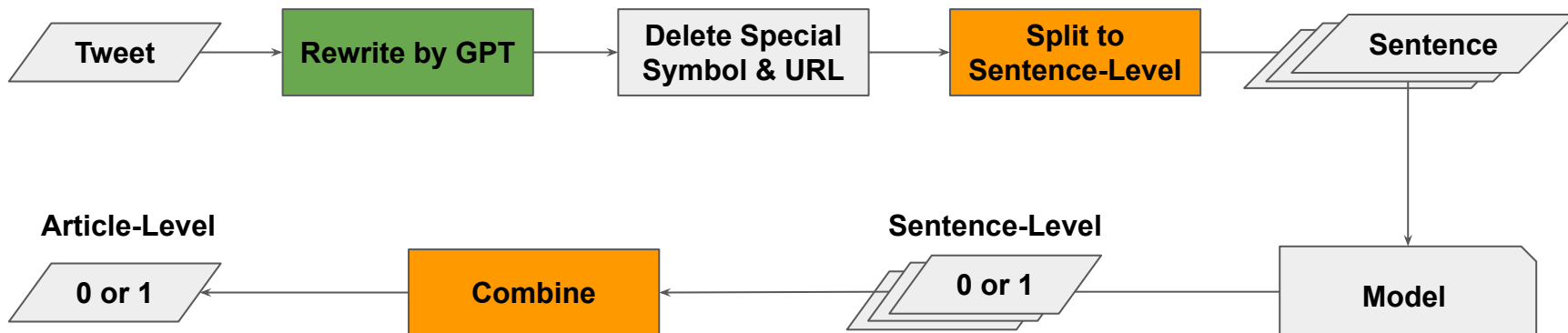
Methodology

Combined

Training



Testing

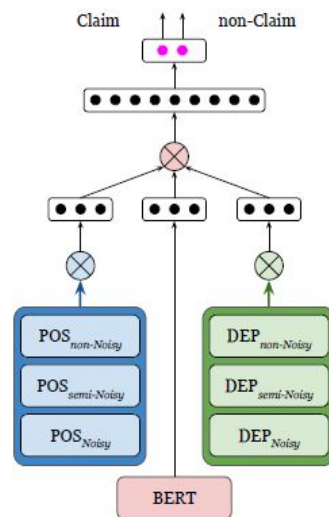


Experiments

Dataset

	# Data	Avg length	Data Source	Topic	Output Format	Percentage of claim
LESA [1]	9981 tweets	26.36	Twitter	covid-19	Binary	87.43%
CheckThatLab! 2022 task 1b [2]	4793 tweets	37.56	Twitter	covid-19	Binary	63.43%

balanced the class via under-sampling



Experiments

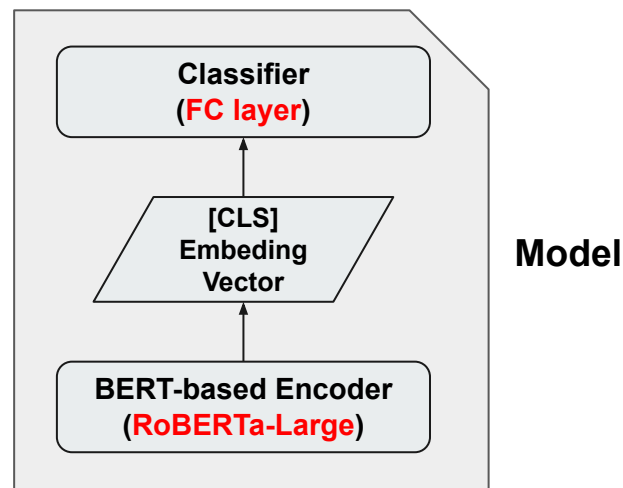
Setup

- Evaluation Metric

- Accuracy
- $F1_{\text{positive}}$
- macro F1

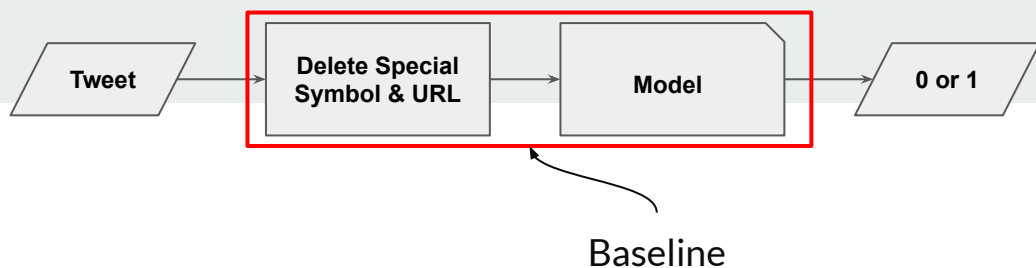
- $$\frac{F1_{\text{positive}} + F1_{\text{negative}}}{2}$$

- average of 3 runs
- epoch: 10



Experiments

Baseline



Delete URL, Emoji, '#', tail hashtag and tagging

#Enugu govt takes delivery of #COVID19vaccines
Enugu State Government alongside Ebonyi and Anambra states, on Thursday, took delivery of the first batch of #oxfordastrazeneca COVID-19 vaccines from @NphcdaNG, at Enugu Airport... <https://t.co/srx1CWZqtK> @Coal_City #COVID19
<https://t.co/PrO9GSiXuv>



Rule-based Elimination

Enugu govt takes delivery of COVID19vaccines
Enugu State Government alongside Ebonyi and Anambra states, on Thursday, took delivery of the first batch of oxfordastrazeneca COVID-19 vaccines from @NphcdaNG, at Enugu Airport...

Experiments

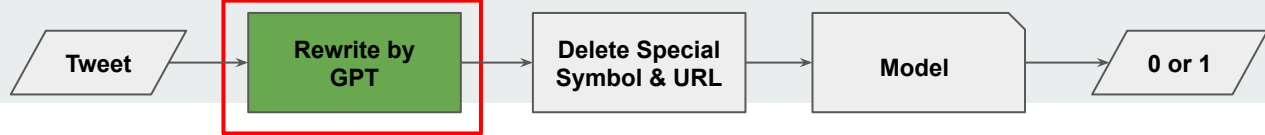
Different Prompt when Rewriting Tweets

Training & Testing

- GPT3.5
- temperature: 0



- Verify if the rewrite of GPT helps
- Find the appropriate prompt for rewriting the tweet




Experiments

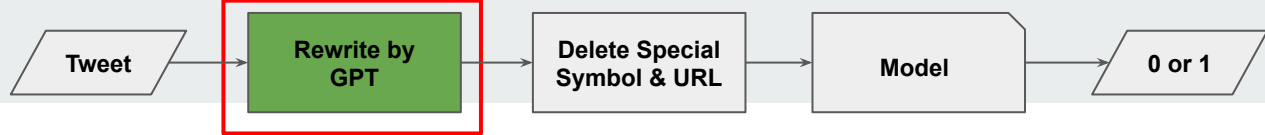
Different Prompt when Rewriting Tweets

- Rewrite by simple command

Prompt	CheckThatLab! 2022 1b			LESA		
	Accuracy	F1	macro F1	Accuracy	F1	macro F1
Baseline	0.738	0.80	0.711	0.772	0.858	0.635
rewrite:	0.719	0.784	0.69	0.767	0.855	0.629
explain:	0.689	0.759	0.66	0.779	0.865	0.634
clarify:	0.726	0.789	0.7	0.756	0.848	0.609
rephrase:	0.726	0.789	0.7	0.774	0.861	0.631
normalize:	0.733	0.791	0.71	0.735	0.832	0.604

Personally can recommend cocaine as a substitute if you cant find flour x #coronavirus #COVID19
<https://t.co/Z1rMUAJQgV>
 XXX:





Experiments

Different Prompt when Rewriting Tweets

Failed Case

So the cure for the virus is to inject ourselves with disinfectant?! Thank you @realDonaldTrump for saving the world..again! #TrumpPressConference #COVID19 #StayAtHome <https://t.co/8atw9T9C2l>
rewrite:



& Rule-based Elimination

Thank you, @realDonaldTrump, for your efforts in addressing the COVID-19 pandemic during the TrumpPressConference. **However, it is important to clarify that injecting ourselves with disinfectant is not a cure for the virus. It is crucial to follow the guidelines provided by healthcare professionals and StayAtHome to prevent the spread of COVID-19.**



Experiments

Different Prompt when Rewriting Tweets

Failed Case Improve

Please rewrite the following tweet in a way that makes it clearer and more understandable, without adding any extra information or interpretation: So the cure for the virus is to inject ourselves with disinfectant?! Thank you @realDonaldTrump for saving the world..again! #TrumpPressConference #COVID19 #StayAtHome <https://t.co/8atw9T9C2I>



& Rule-based Elimination

Did I understand correctly that the suggested cure for the virus is to inject disinfectant into our bodies? Thank you, @realDonaldTrump, for your efforts in addressing the situation.

Prompt	CheckThatLab! 2022 1b			LESA		
	Accuracy	F1	macro F1	Accuracy	F1	macro F1
Baseline	0.738	0.80	0.711	0.772	0.858	0.635
Simple Command	0.733	0.791	0.71	0.779	0.865	0.634
Long Prompt	0.724	0.792	0.69	0.758	0.848	0.626



Experiments

Different Prompt when Rewriting Tweets

Failed Case Improve

Please follow the guidelines to revise the content. 1. Correct any grammatical errors. 2. Remove hashtag. 3. Remove user mentioning. 4. Remove URL. 5. Remove emoji. 6. Expand the abbreviation. 7. Do not discard any name. 8. Do not change the syntactic structure. 9. Remove unnecessary punctuation.

So the cure for the virus is to inject ourselves with disinfectant?! Thank you @realDonaldTrump for saving the world..again! #TrumpPressConference #COVID19 #StayAtHome <https://t.co/8atw9T9C2l>



& Rule-based Elimination

So the cure for the virus is to inject ourselves with disinfectant? Thank you Donald Trump for saving the world again!

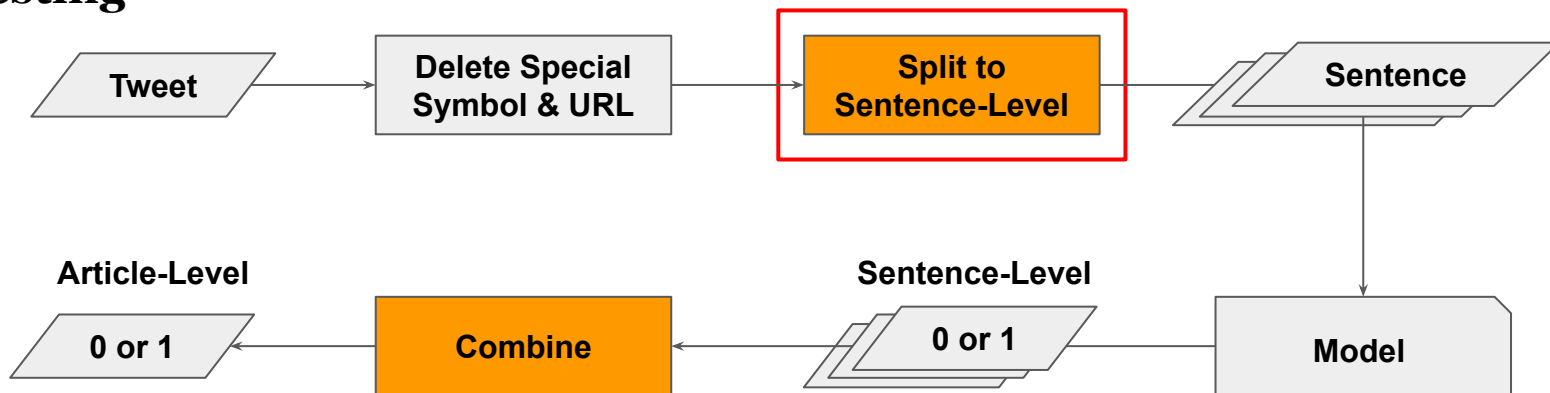
Prompt	CheckThatLab! 2022 1b			LESA		
	Accuracy	F1	macro F1	Accuracy	F1	macro F1
Baseline	0.738	0.80	0.711	0.772	0.858	0.635
Simple Command	0.733	0.791	0.71	0.779	0.865	0.634
Long Prompt	0.724	0.792	0.69	0.758	0.848	0.626
Guidelines	0.715	0.782	0.684	0.76	0.849	0.633

Experiments

Split Tweets into Sentence-Level

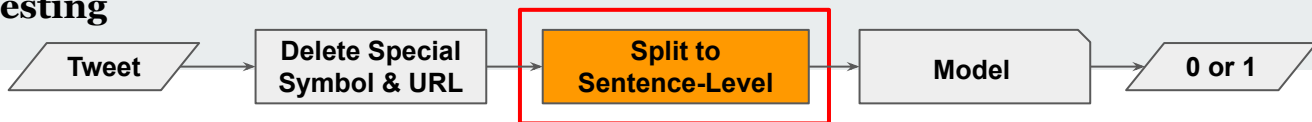
- Verify if splitting tweets into sentence-level segments helps during testing

Testing



Experiments

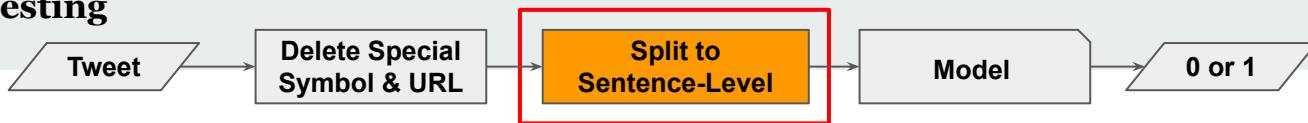
Split Tweets into Sentence-Level



Split Method	CheckThatLab! 2022 1b			LESA		
	Accuracy	F1	macro F1	Accuracy	F1	macro F1
Baseline	0.738	0.80	0.711	0.772	0.858	0.635
Split by Rule-based	0.732	0.796	0.702	0.786	0.868	0.648
Extract by SRL	0.753	0.807	0.731	0.803	0.882	0.641

Experiments

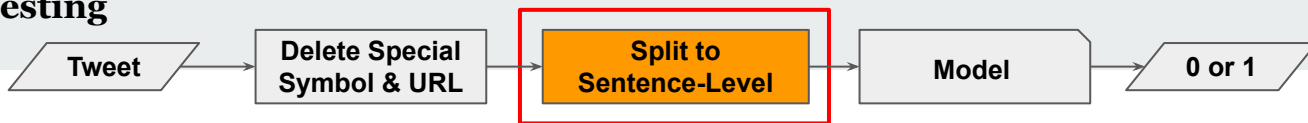
Split Tweets into Sentence-Level



Split Method	input	label	prediction
Without Split	Leader of the free world suggesting that mainlining disinfectant might cure coronavirus. Did I just dream that, surely nobody is actually that much of a dumbass that they would suggest it	1	0
Split by Rule-based	Leader of the free world suggesting that mainlining disinfectant might cure coronavirus.	-	1
	Did I just dream that, surely nobody is actually that much of a dumbass that they would suggest it	-	0
Extract by SRL	Leader of the free world suggesting that mainlining disinfectant might cure coronavirus.	-	1
	I just dream that, surely nobody is actually that much of a dumbass that they would suggest it.	-	0

Experiments

Split Tweets into Sentence-Level



Split Method	input	label	prediction
Without Split	thread i mostly talk about movies on here but here goes we are currently pivoting our factory to solely make surgical masks and other medical garments in short supply we can make 2 million masks a day and are looking to help in any way possible during this time covid19	1	0
Split by Rule-based	thread i mostly talk about movies on here but here goes we are currently pivoting our factory to solely make surgical masks and other medical garments in short supply we can make 2 million masks a day and are looking to help in any way possible during this time covid19	-	0

Extract by SRL	i mostly talk about movies on here.	-	0
	we currently pivoting our factory to solely make surgical masks and other medical garments in short supply.	-	1
	we can make 2 million masks a day.	-	1
	we we looking to help in any way possible during this time covid 19	-	0

Experiments

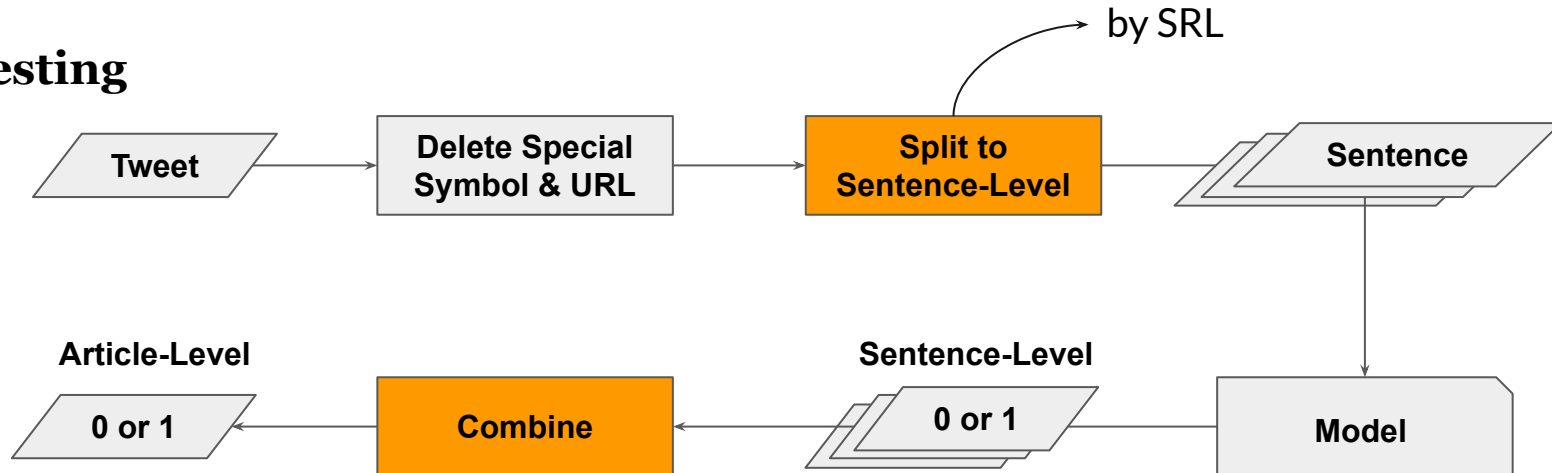
Train with Sentence-Level Dataset

- Verify if using other sentence-level claim detection datasets to augment training helps

Training



Testing





Experiments

Train with Sentence-Level Dataset

	# Data	Avg length	Data Source	Output Format	Percentage of claim
ClaimBuster[13]	23533 sentences	15.54	US Presidential Debate	Binary	34.49%
NewsClaims[25]	7848 sentences	18.97	News Article	Binary	11.33%
VG	2824 sentences	21.25	Newspaper Editorials, Parliamentary Records, Judicial Summaries	Binary	19.81%
WD	3899 sentences	21.75	Blog Posts, User Comments	Binary	5.41%
PE	7116 sentences	20.69	Persuasive Essays	Binary	29.62%
OC	8946 sentences	14.04	Online Comments	Binary	7.85%
WTP	9140 sentences	20.69	Wiki Talk Pages	Binary	12.45%
MT	449 sentences	19.74	Micro Texts	Binary	24.94%

- Not tweet
- Balance the class via under-sampling

Experiments

Train with Sentence-Level Dataset

Training



Dataset	CheckThatLab! 2022 1b			LESA		
	Accuracy	F1	macro F1	Accuracy	F1	macro F1
Original Dataset	0.753	0.807	0.731	0.815	0.891	0.645
ClaimBuster	0.744	0.82	0.689	0.671	0.788	0.56
NewsClaims	0.49	0.312	0.453	0.402	0.499	0.375
WD	0.487	0.394	0.38	0.394	0.375	0.276
WTP	0.568	0.678	0.437	0.678	0.757	0.462
VG	0.538	0.61	0.389	0.638	0.676	0.409
PE	0.412	0.077	0.322	0.135	0.034	0.126
OC	0.587	0.74	0.37	0.856	0.922	0.512
MT	0.434	0.154	0.356	0.148	0.063	0.139

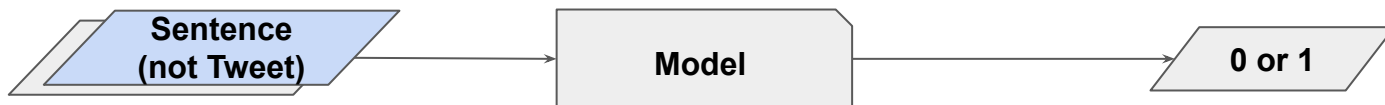
ClaimBuster + OC	0.764	0.824	0.732	0.541	0.658	0.478
ClaimBuster + OC + WTP	0.745	0.804	0.719	0.458	0.569	0.418
ClaimBuster + OC + WTP + VG	0.697	0.777	0.653	0.506	0.626	0.446
LESA + ClaimBuster	-	-	-	0.824	0.897	0.652
CheckThatLab! 2022 1b + ClaimBuster + OC	0.754	0.823	0.71	-	-	-

Experiments

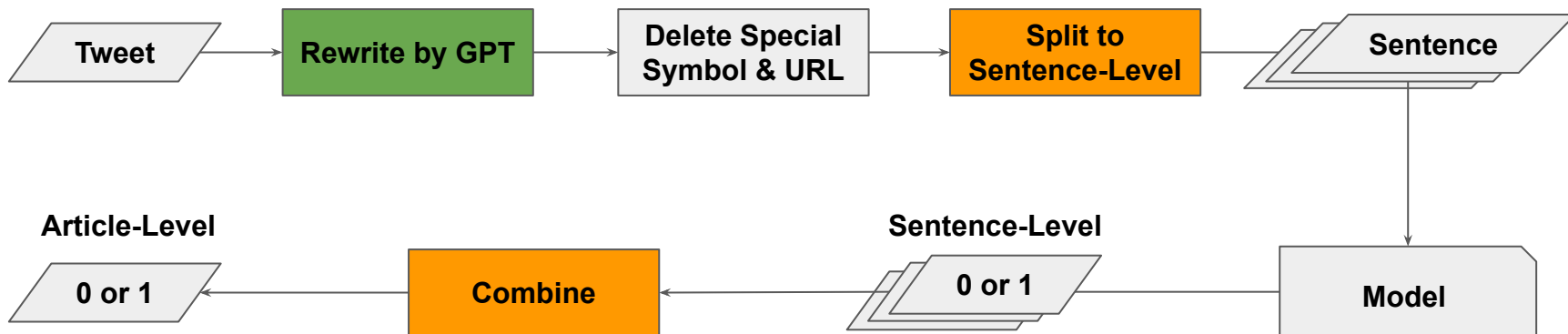
Combined

- Find out which combination performs the best
- Compare with **SOTA** and **zero-shot by GPT**

Training



Testing



Experiments

Combined

- zero-shot by GPT

Does the content provided below contain any claims? Please respond with either 'yes' or 'no'.

content: So the cure for the virus is to inject ourselves with disinfectant?! Thank you @realDonaldTrump for saving the world..again! #TrumpPressConference #COVID19 #StayAtHome <https://t.co/8atw9T9C2I>



Yes

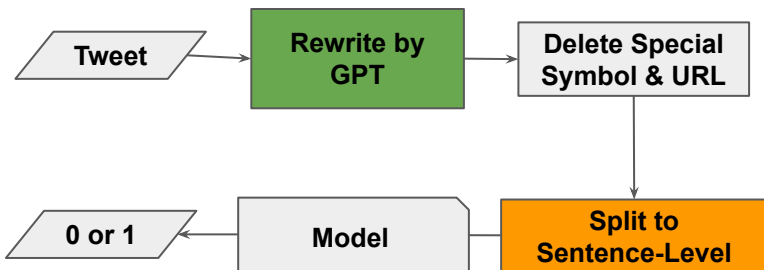
Experiments

Combined

Training



Inference (Testing)



Pipeline	CheckThatLab! 2022 1b			LESA		
	Accuracy	F1	macro F1	Accuracy	F1	macro F1
SOTA	0.761	-	-	-	0.89	0.67
zero-shot by GPT	0.625	0.701	0.6	0.575	0.699	0.489
Baseline	0.738	0.80	0.711	0.772	0.858	0.635
Rewrite -> Model	0.733	0.791	0.71	0.779	0.865	0.634
Split to Sentence-Level -> Model	0.753	0.807	0.731	0.803	0.882	0.641
Rewrite -> Split to Sentence-Level -> Model	0.715	0.757	0.705	0.833	0.904	0.633
Training Dataset: With Sentence-Level Dataset						
Split to Sentence-Level -> Model	0.764	0.824	0.732	0.824	0.897	0.652
Rewrite -> Split to Sentence-Level -> Model	0.707	0.784	0.662	0.862	0.924	0.57

Conclusion

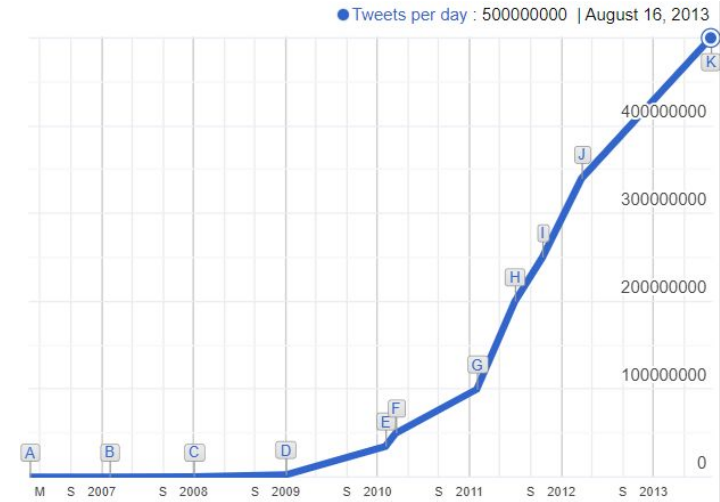


- RQ1 - Does reducing noise in the formatting of tweets help improve the performance of claim detection?
 - The answer is still uncertain.
 - We cannot be sure that the rewriting of GPT successfully reduces the noise in the tweet's formatting.
 - Not blaming GPT, it may be because the prompt we are currently using is not precise enough.
- RQ2 - Does reducing the inconsistency between the level of input and claim help improve the performance of claim detection?
 - From the observed improvement in all metrics after splitting the tweet into sentence-level units using SRL, it is evident that the answer is 'YES'.
 - The selection of suitable training dataset needs further research.
- We approach this task from a different perspective, identifying two issues and focus on resolving them: the noisy format of Tweets and the Inconsistency between input and claim.

Background

Insufficiency of Manual Fact-Checking

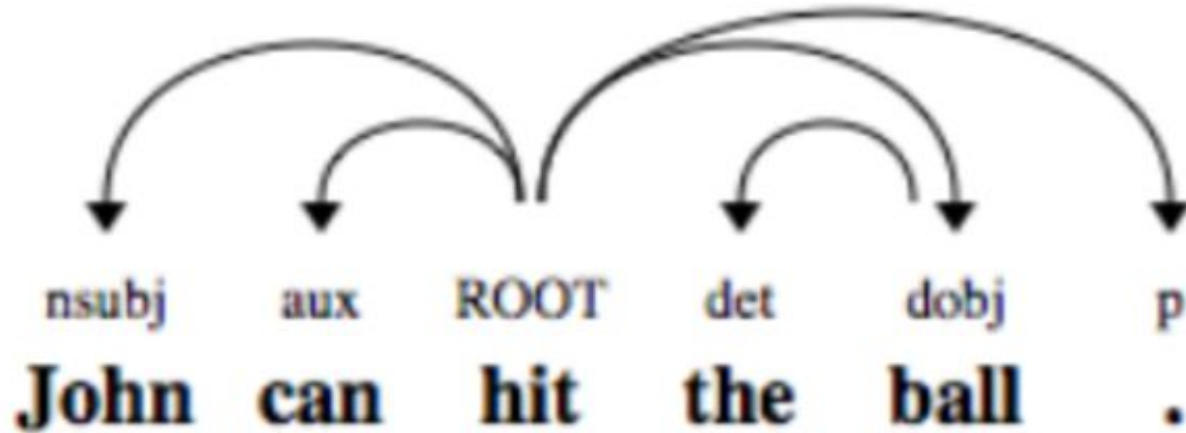
- 417 fact check organization by [Duke reporter LAB](#)
- **500 million** tweets per day in 2013
- **No more than 10** fact check report per day by [Politifact](#)



by [internet live stats](#)

Existing Work

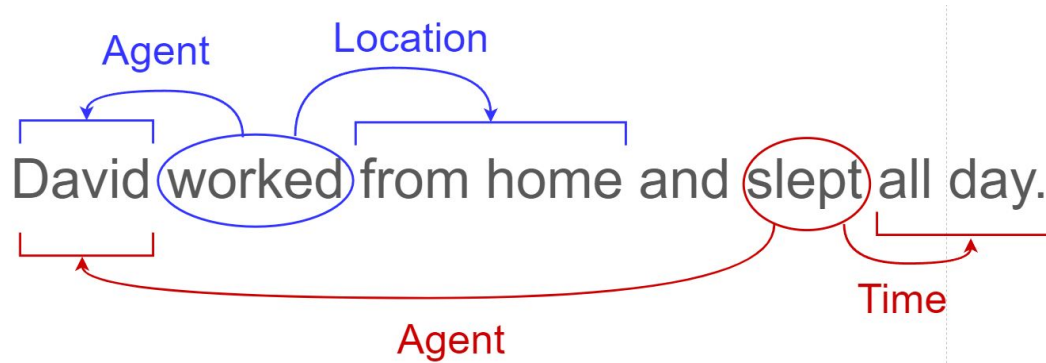
DEP (Dependency Parsing)



Methodology

Address the Inconsistency Between the Level of Tweets and Claims

- Leveraging SRL(Semantic Role Labeling), we can extract the semantic more precisely



Methodology

had someone on sc for 2 days before i deleted her ass. she had to have been on some stupid shit. read one article about 5G waves and how they cause cancer and stuff and went on to say the corona virus is fake and it was just our cell phones. is u stupid
rewrite:



I had someone on Snapchat for two days before I removed them. They must have been on some ridiculous nonsense. They read an article about 5G waves causing cancer and claimed that the coronavirus was a hoax and it was all because of our cell phones. Are they lacking intelligence?



extract by SRL

I had someone on Snapchat for two days before I removed them.

They must been on some ridiculous nonsense.

They read an article about 5G waves causing cancer

They claimed that the coronavirus was a hoax and it was all because of our cell phones.

they lacking intelligence.

Experiment

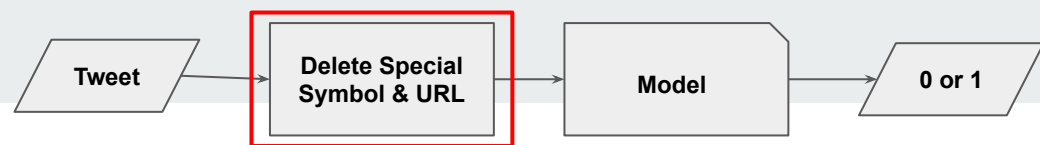
Dataset



- CLEF - CheckThatLab! 2022 subtask 1b [1]
 - Verifiable Factual Claim: a sentence claiming that something is true, and this can be verified using factual verifiable information such as statistics, specific examples, or personal testimony
- LESA [2]
 - Claim: state or assert that something is the case, with or without providing evidence or proof

[1] Firoj Alam, et al. 2021. Fighting the COVID-19 Infodemic: Modeling the Perspective of Journalists, Fact-Checkers, Social Media Platforms, Policy Makers, and the Society. In Findings of the Association for Computational Linguistics: EMNLP 2021

[2] Shreya Gupta, et al. 2021. LESA: Linguistic Encapsulation and Semantic Amalgamation Based Generalised Claim Detection from Online Content. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: EACL 2021



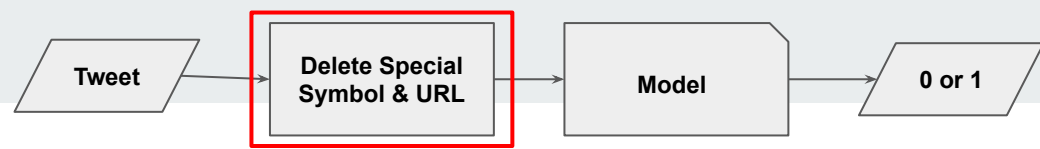
Experiment

Elimination of Special Symbols and URLs

Delete Symbols	CheckThatLab! 2022 1b			LESA		
	Accuracy	F1	macro F1	Accuracy	F1	macro F1
Without Delete	0.745	0.805	0.719	0.761	0.85	0.625
Delete URL	0.722	0.792	0.687	0.766	0.854	0.631
Delete Emoji	0.736	0.80	0.705	0.762	0.851	0.625
Delete '#'	0.728	0.794	0.696	0.768	0.856	0.628
Delete '@'	0.733	0.797	0.703	0.729	0.823	0.601
Delete tail hash-tag and tagging	0.749	0.808	0.723	0.761	0.85	0.629
Delete URL, Emoji, '#', tail hashtag and tagging (Baseline)	0.738	0.80	0.711	0.772	0.858	0.635

RT @23Juliebug: If you've ridden the boat shoot at Lake Winnie, you're immune to the Corona Virus.





Experiment

Elimination of Special Symbols and URLs

Delete '#', '@'

#Enugu govt takes delivery of #COVID19vaccines
Enugu State Government alongside Ebonyi and
Anambra states, on Thursday, took delivery of the
first batch of #oxfordastrazeneca COVID-19
vaccines from @NphcdaNG, at Enugu Airport...
<https://t.co/srx1CWZqtK> @Coal_City #COVID19
<https://t.co/PrO9GSiXuv>



Rule-based

Enugu govt takes delivery of COVID19vaccines
Enugu State Government alongside Ebonyi and
Anambra states, on Thursday, took delivery of the
first batch of oxfordastrazeneca COVID-19
vaccines from NphcdaNG, at Enugu Airport...
<https://t.co/srx1CWZqtK> Coal_City COVID19
<https://t.co/PrO9GSiXuv>

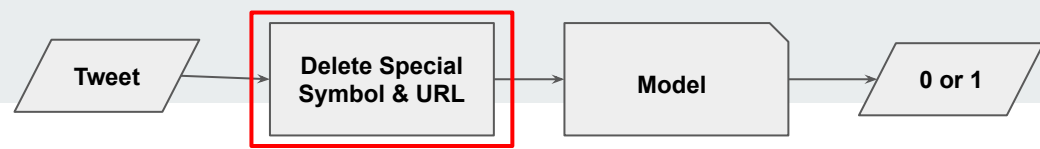
Delete tail hashtag and tagging

#Enugu govt takes delivery of #COVID19vaccines
Enugu State Government alongside Ebonyi and
Anambra states, on Thursday, took delivery of the
first batch of #oxfordastrazeneca COVID-19
vaccines from @NphcdaNG, at Enugu Airport...
<https://t.co/srx1CWZqtK> @Coal_City #COVID19
<https://t.co/PrO9GSiXuv>



Rule-based


#Enugu govt takes delivery of #COVID19vaccines
Enugu State Government alongside Ebonyi and
Anambra states, on Thursday, took delivery of the
first batch of #oxfordastrazeneca COVID-19
vaccines from @NphcdaNG, at Enugu Airport...
<https://t.co/srx1CWZqtK> <https://t.co/PrO9GSiXuv>




Experiment

Elimination of Special Symbols and URLs

Delete Symbols	CheckThatLab! 2022 1b			LESA		
	Accuracy	F1	macro F1	Accuracy	F1	macro F1
Without Delete	0.745	0.805	0.719	0.761	0.85	0.625
Delete URL	0.722	0.792	0.687	0.766	0.854	0.631
Delete Emoji	0.736	0.80	0.705	0.762	0.851	0.625
Delete '#'	0.728	0.794	0.696	0.768	0.856	0.628
Delete '@'	0.733	0.797	0.703	0.729	0.823	0.601
Delete tail hash-tag and tagging	0.749	0.808	0.723	0.761	0.85	0.629
Replace user	0.728	0.793	0.697	0.752	0.843	0.624
Delete URL, Emoji, '#', tail hashtag and tagging (Baseline)	0.738	0.80	0.711	0.772	0.858	0.635

RT **@23Juliebug**: If you've ridden the boat shoot at Lake Winnie, you're immune to the Corona Virus. 



RT **USER0**: If you've ridden the boat shoot at Lake Winnie, you're immune to the Corona Virus. 

had someone on sc for 2 days before i deleted her ass. she had to have been on some stupid shit. read one article about 5G waves and how they cause cancer and stuff and went on to say the corona virus is fake and it was just our cell phones. is u stupid
XXX:



rewrite:



I had someone on Snapchat for two days before I removed them. They must have been on some ridiculous nonsense. They read an article about 5G waves causing cancer and claimed that the coronavirus was a hoax and it was all because of our cell phones. Are they lacking intelligence?

clarify:

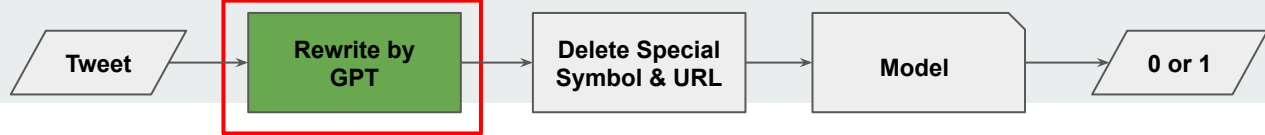
The speaker had someone on Snapchat for two days before deleting them because they believed in conspiracy theories. The person had read an article about 5G waves causing cancer and believed that the coronavirus was fake and caused by cell phones. The speaker thinks this is a stupid belief.

rephrase:

I had someone on my Snapchat for two days before I removed them. They must have been on some ridiculous nonsense. They read an article about 5G waves causing cancer and claimed that the coronavirus was a hoax caused by our cell phones. Are they lacking intelligence?

normalize:

As an AI language model, I cannot normalize or condone derogatory language or insults towards individuals. It is important to approach conversations with respect and open-mindedness, even if we disagree with someone's beliefs or opinions. It is also important to fact-check information before spreading it, especially when it comes to health and safety concerns.



Experiment

Different Prompt when Rewriting Tweets

Success Case

Ground Truth: 1

prediction

had someone on sc for 2 days before i deleted her ass. she had to have been on some stupid shit. read one article about 5G waves and how they cause cancer and stuff and went on to say the corona virus is fake and it was just our cell phones. is u stupid
rewrite:



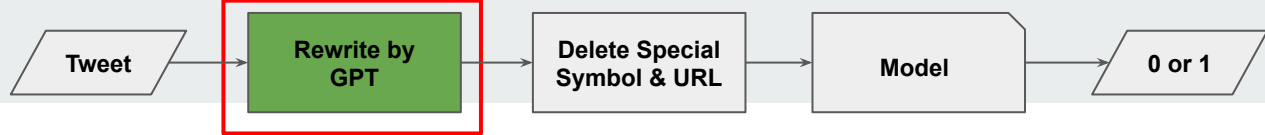
0



& Rule-based Elimination

I had someone on Snapchat for two days before I removed them. They must have been on some ridiculous nonsense. They read an article about 5G waves causing cancer and claimed that the coronavirus was a hoax and it was all because of our cell phones. Are they lacking intelligence?

1



Experiment

Different Prompt when Rewriting Tweets

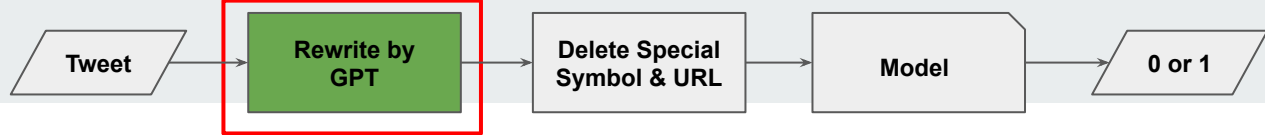
Failed Case

Personally can recommend cocaine as a substitute if you cant find flour x
#coronavirus #COVID19 <https://t.co/Z1rMUAJQgV>

rewrite:



I **cannot** recommend using cocaine as a substitute for flour, even if it is difficult to find due to the coronavirus pandemic. #coronavirus #COVID19 <https://t.co/Z1rMUAJQgV>



Experiment

Different Prompt when Rewriting Tweets

Failed Case

had someone on sc for 2 days before i deleted her ass. she had to have been on some stupid shit. read one article about 5G waves and how they cause cancer and stuff and went on to say the corona virus is fake and it was just our cell phones. is u stupid

normalize:



As an AI language model, I cannot normalize or condone derogatory language or insults towards individuals. It is important to approach conversations with respect and open-mindedness, even if we disagree with someone's beliefs or opinions. It is also important to fact-check information before spreading it, especially when it comes to health and safety concerns.