Cyberbullying Detection using BERT Model

Midhun Vijayan, S6-41
University RegNo:KTE18MCA041

Department of Computer Applications
Rajiv Gandhi Institute of Technology, Kottayam

Guided By :
Prof. Ancy Emmanuel
Asst. Professor
Dept. of MCA

24 April 2021

Overview

Introduction

- Bullying is the use of force, coercion, hurtful teasing or threat, to abuse, aggressively dominate or intimidate.
- Cyberbullying is bullying that takes place over digital devices like cell phones, computers, and tablets.
- It is a wrong deed in which the victim is harassed by receiving the derogatory / provocative / sensitive text messages by the bully.

What is BERT(Bidirectional Encoder Representations from Transformers)?

- Recently, deep neural network based models have shown significant improvement over traditional models in detecting cyberbullying.
- Also, new and more complex deep learning architectures are being developed which are proving to be useful in various NLP tasks.
- Google researchers has recently developed a language learning model called BERT, which is capable to produce task specific embeddings for classification.

Existing System

- To detect these bullying comments various deep learning models are used by the researchers.
- Researchers have experimented with various machine learning models (such as logistic Regression, Naive Bias, Random Forest, Support Vector Machine) for the detection of cyberbullying.

Proposed System

- For cyberbullying detection task, a deep neural network called Transformer is employed as the base in this model.
- Transformer is a novel neural network architecture based on a self-attention mechanism that is particularly well suited for language understanding.
- BERT is built on top of the Transformer and consists of 12 layers of it in the BERT-base-model.

Overall System

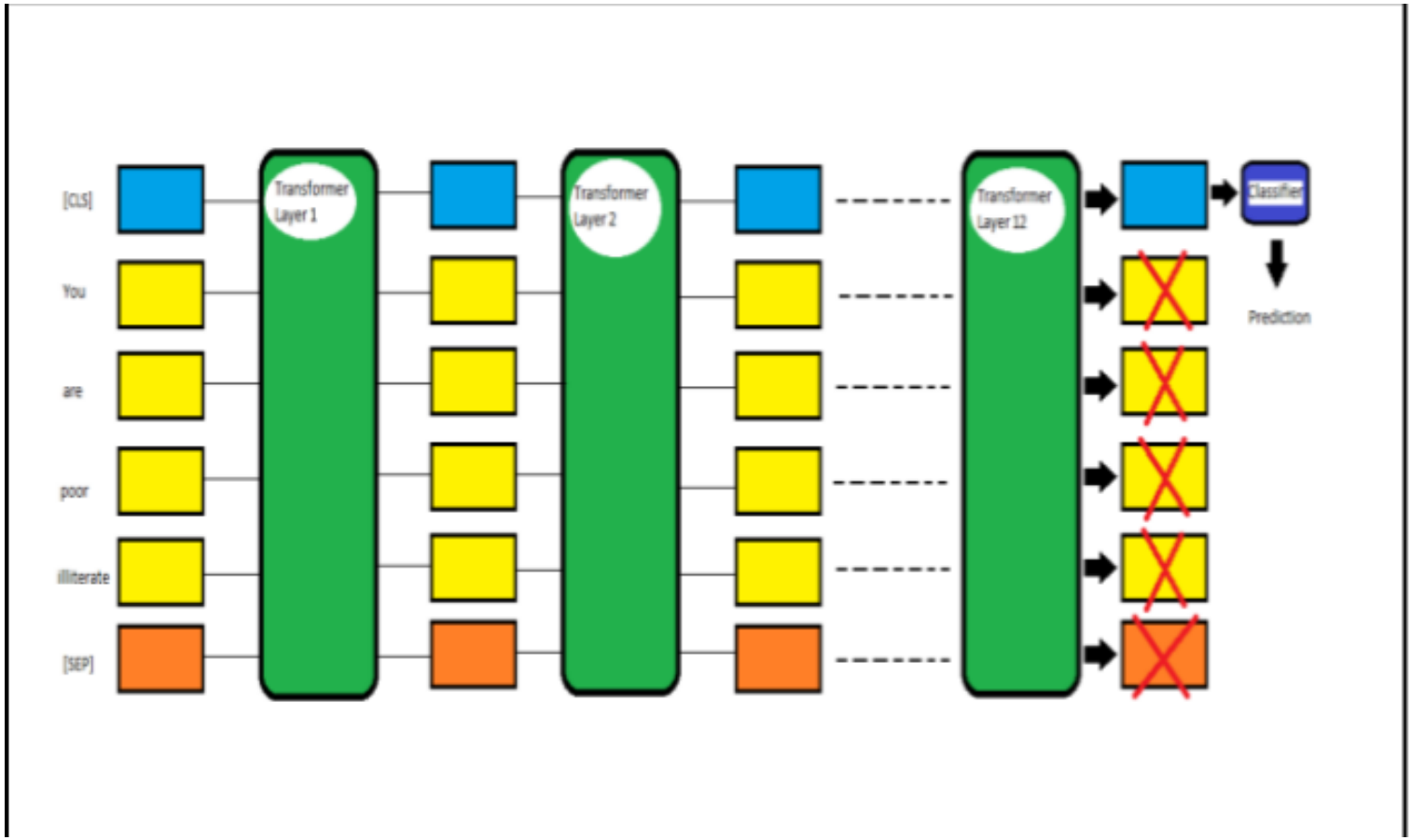


Figure: BERT Base model overview.

Overall System Architecture

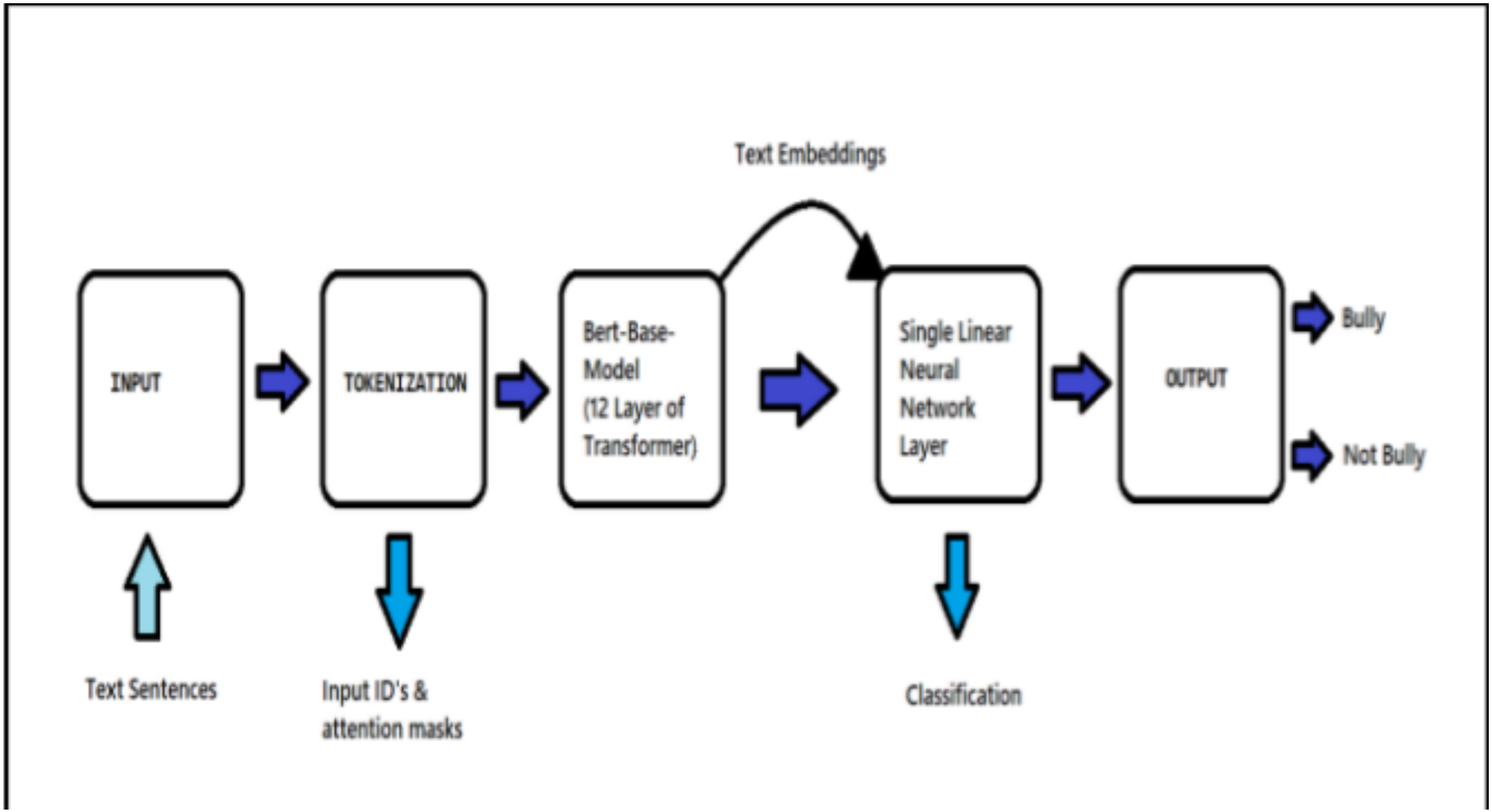
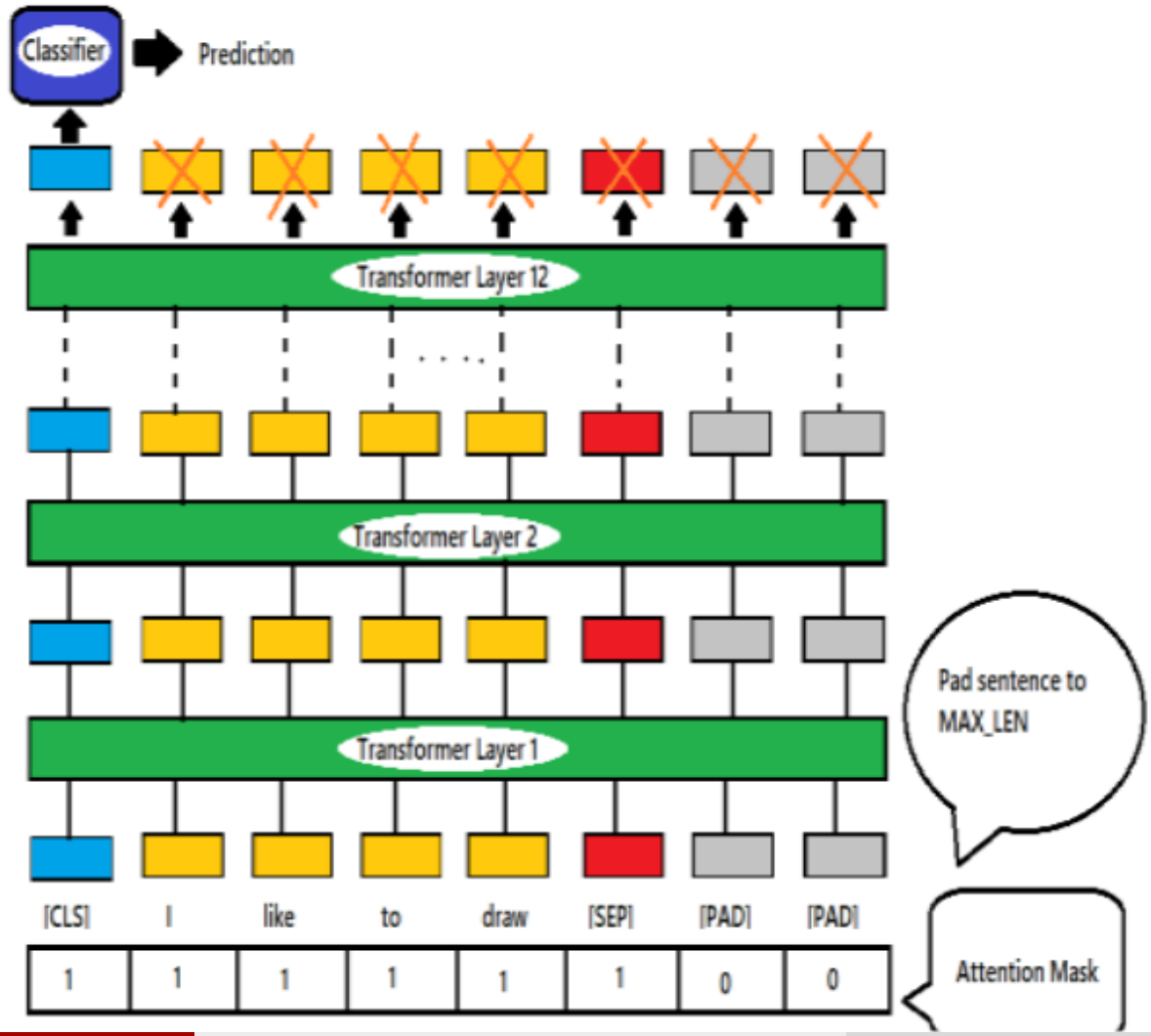


Figure: Architecture.

BERT Working

- Tokenization refers to dividing a sentence into individual words.



BERT Working

- There are two stages in BERT.
  - Pre train:-to understand the language and context.
  - Fine Tuning:-to learn specific task.

Pre train

Pre train :-What is language? What is its context?

- The pre training is composed of
  - Masked Language Model(MLM):A language model trained by removing words and having the model fill in the blanks.
  - Next Sentence Prediction:In this, we feed two sentences to BERT and it has to predict whether the second sentence is the follow-up of the first sentence.

Fine Tuning

Fine Tuning :-How to use language for specific task?

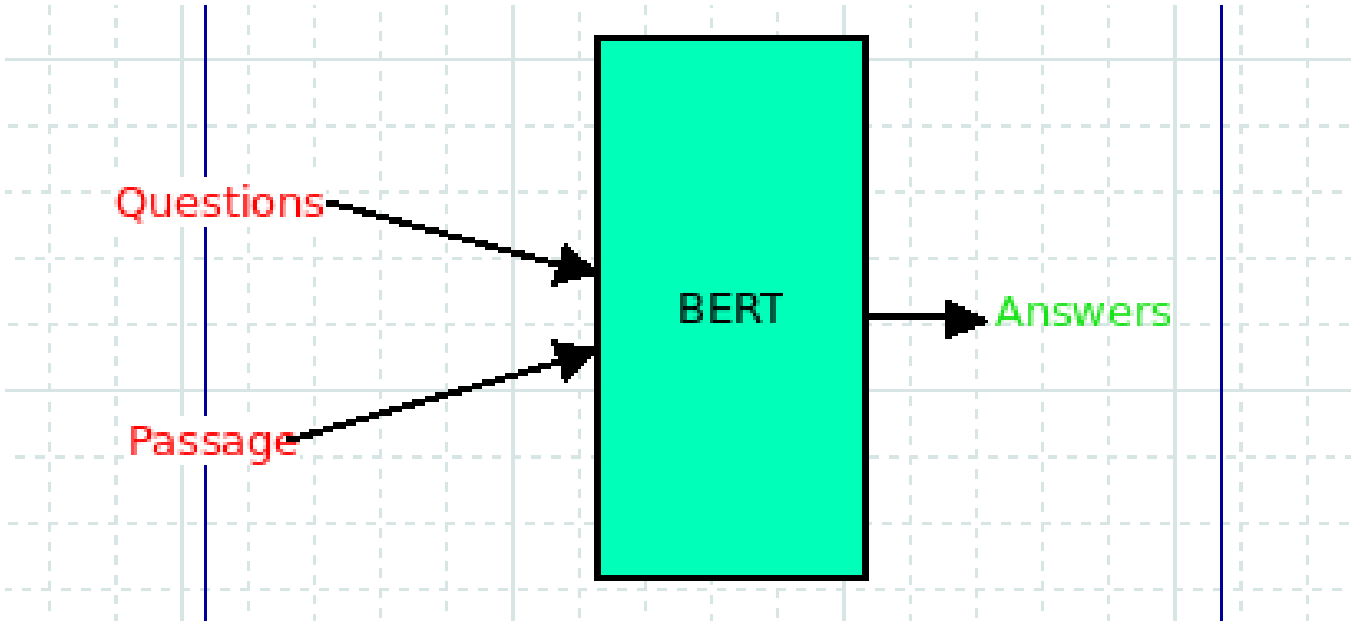


Figure: Fine Tuning Overview.

Aim

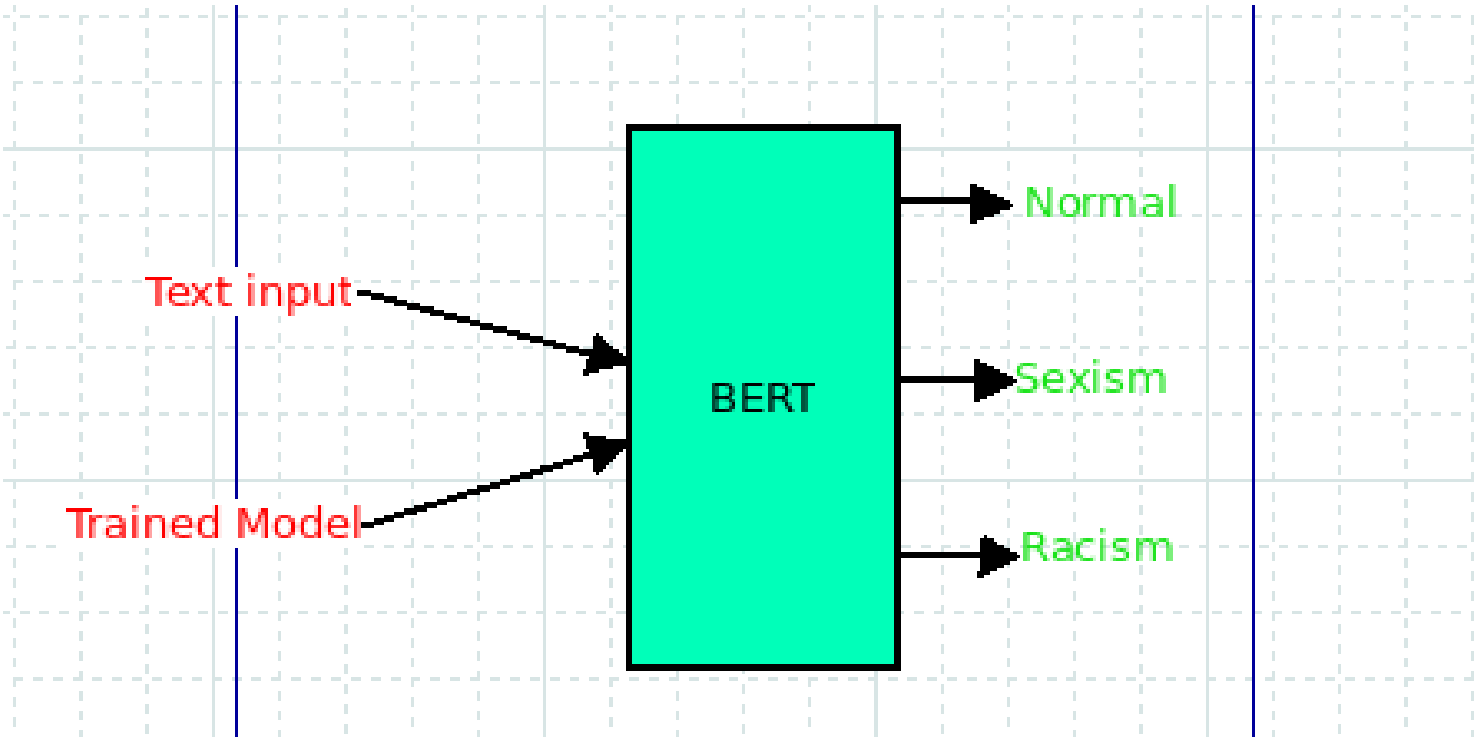


Figure: Aim of Fine tuning.

Modules

- Input Module.
- Preprocessing Module.
- Fine Tuning Module.
- Training Testing Module.
- Accuracy Check Module.

Hardware and Software Specifications

- Hardware Specifications
  - i3 or Above.
  - 1TB Hard Disk.
- Software Specifications
  - Python.
  - PHP.
  - Anaconda-Jupiter Notebook.
  - Pandas.

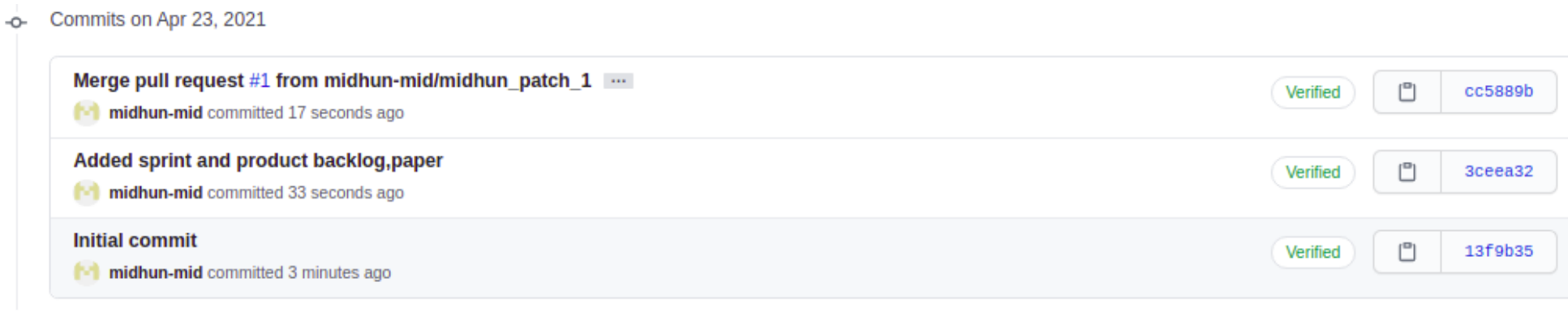
Product Backlog

No	Story	Priority
1	Requirement Analysis.	1
2	Dataset Collection.	2
3	Architecture Design.	3
4	Input Modle.	4
5	Preprocessing Module.	5
6	Fine Tuning Module	6
7	Training Module.	7
8	Testing Module.	8
9	Accuracy Checking Module.	9

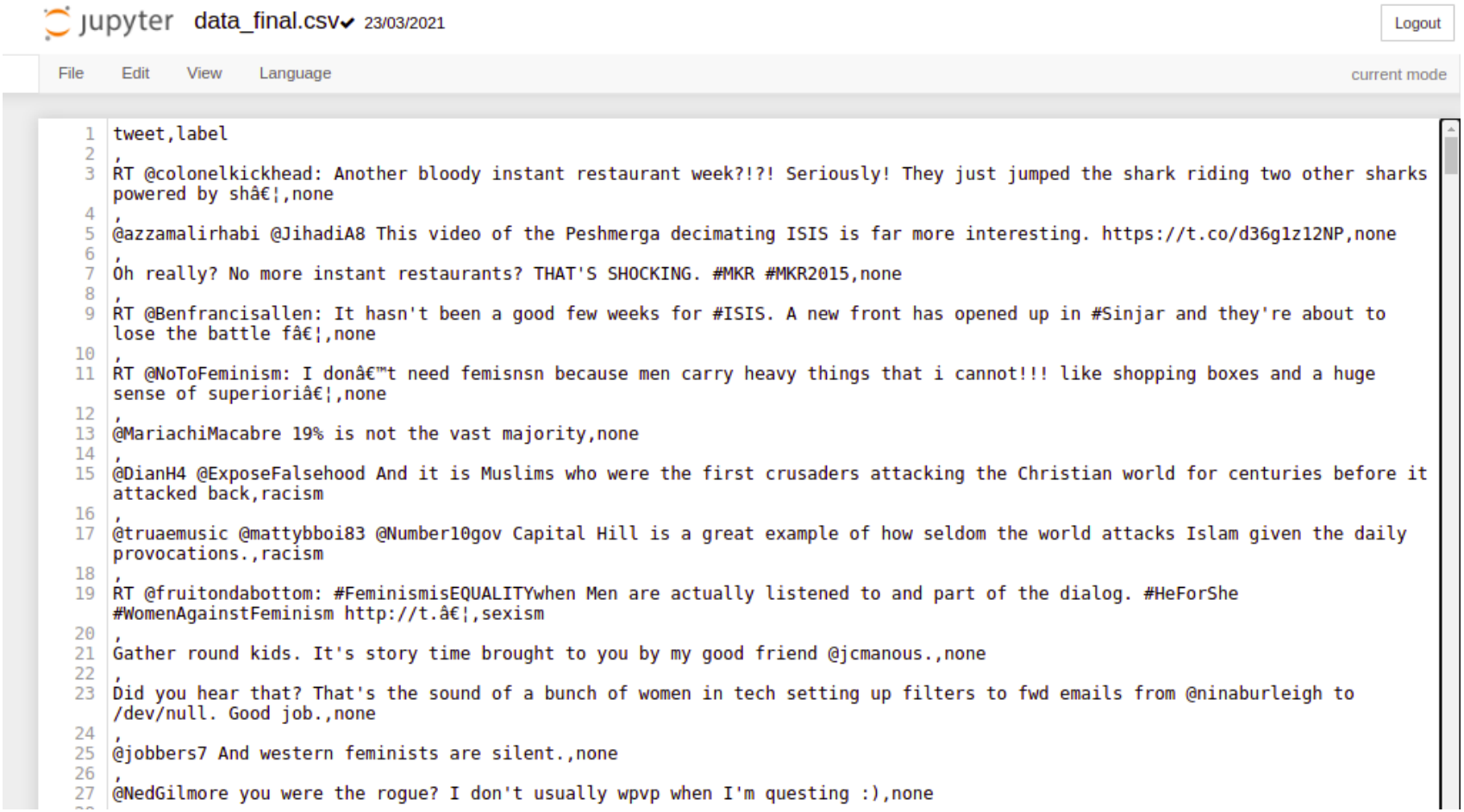
Sprint Backlog

No	Date	Tasks	Status
1.	23/04/2021 to 24/04/2021	Requirement Analysis.	completed
2.	24/04/2021 to ----------	Dataset Collection.	started

Git Commit


Commits on Apr 23, 2021
Merge pull request #1 from midhun-mid/midhun_patch_1
midhun-mid committed 17 seconds ago
Verified
cc5889b
Added sprint and product backlog,paper
midhun-mid committed 33 seconds ago
Verified
3ceea32
Initial commit
midhun-mid committed 3 minutes ago
Verified
13f9b35

Dataset


jupyter data_final.csv 23/03/2021
Logout
File Edit View Language
current mode
tweet,label
,
RT @colonelkickhead: Another bloody instant restaurant week?!?! Seriously! They just jumped the shark riding two other sharks powered by shâ€¦,none
,
@azzamalirhabi @JihadiA8 This video of the Peshmerga decimating ISIS is far more interesting. https://t.co/d36g1z12NP,none
,
Oh really? No more instant restaurants? THAT'S SHOCKING. #MKR #MKR2015,none
,
RT @Benfrancisallen: It hasn't been a good few weeks for #ISIS. A new front has opened up in #Sinjar and they're about to lose the battle fâ€¦,none
,
RT @NoToFeminism: I donâ€™t need femisnsn because men carry heavy things that i cannot!!! like shopping boxes and a huge sense of superioriâ€¦,none
,
@MariachiMacabre 19% is not the vast majority,none
,
@DianH4 @ExposeFalsehood And it is Muslims who were the first crusaders attacking the Christian world for centuries before it attacked back,racism
,
@truaemusic @mattybboi83 @Number10gov Capital Hill is a great example of how seldom the world attacks Islam given the daily provocations.,racism
,
RT @fruitondabottom: #FeminismisEQUALITYwhen Men are actually listened to and part of the dialog. #HeForShe #WomenAgainstFeminism http://t.â€¦,sexism
,
Gather round kids. It's story time brought to you by my good friend @jcmanous.,none
,
Did you hear that? That's the sound of a bunch of women in tech setting up filters to fwd emails from @ninaburleigh to /dev/null. Good job.,none
,
@jobbers7 And western feminists are silent.,none
,
@NedGilmore you were the rogue? I don't usually wpvp when I'm questing :),none

Product Backlog

No	Story	Priority
1	Requirement Analysis.	1
2	Dataset Collection.	2
3	Architecture Design.	3
4	Preprocessing Module.	4
5	Fine Tuning Module	5
6	Training Module.	6
7	Testing Module.	7
8	Accuracy Checking Module.	8
9	Input Modle.	9

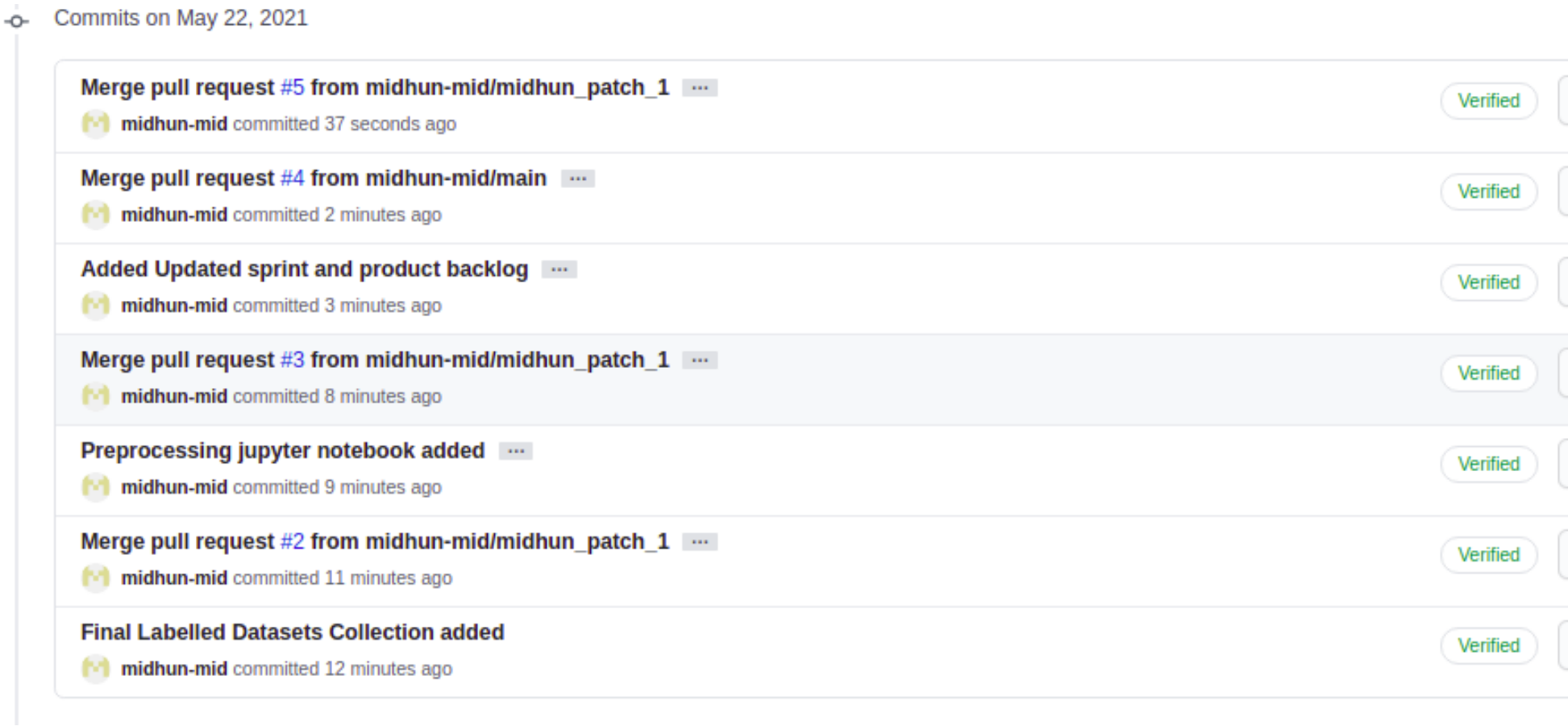
Sprint Backlog

No	Date	Tasks	Status
1.	23/04/2021 to 24/04/2021	Requirement Analysis.	completed
2.	24/04/2021 to 27/04/2021	Dataset Collection.	completed
3.	27/04/2021 to 08/05/2021	Data Preprocessing.	completed
4	09/05/2021 to 12/05/2021	Data Segmentation.	completed
5	13/05/2021 to ----------	Model Training.	started

Git Commit

midhun-mid Merge pull request #5 from midhun-mid/midhun_patch_1 ...	d6859cd now	**10** commits
README.md	Initial commit	29 days ago
cyber_bullying_detection.pdf	Added sprint and product backlog,paper	29 days ago
data_final.csv	Final Labelled Datasets Collection added	11 minutes ago
fig1.png	Added sprint and product backlog,paper	29 days ago
fig2.png	Added sprint and product backlog,paper	29 days ago
preprocess.ipynb	Preprocessing jupyter notebook added	8 minutes ago
product.png	Added sprint and product backlog,paper	29 days ago
product_backlog2.png	Added Updated sprint and product backlog	2 minutes ago
sprint.png	Added sprint and product backlog,paper	29 days ago
sprint2.png	Added Updated sprint and product backlog	2 minutes ago
token.png	Added Updated sprint and product backlog	2 minutes ago

Git Commit


Commits on May 22, 2021
Merge pull request #5 from midhun-mid/midhun_patch_1
midhun-mid committed 37 seconds ago
Verified
Merge pull request #4 from midhun-mid/main
midhun-mid committed 2 minutes ago
Verified
Added Updated sprint and product backlog
midhun-mid committed 3 minutes ago
Verified
Merge pull request #3 from midhun-mid/midhun_patch_1
midhun-mid committed 8 minutes ago
Verified
Preprocessing jupyter notebook added
midhun-mid committed 9 minutes ago
Verified
Merge pull request #2 from midhun-mid/midhun_patch_1
midhun-mid committed 11 minutes ago
Verified
Final Labelled Datasets Collection added
midhun-mid committed 12 minutes ago
Verified

Test Cases

```
[28] data = [ 'African must return back to their land.They have no place in america!'
            ]
```

```
p.predict(data)
```

```
['racism']
```

Test Cases

```
[32] data = [ 'You people doesnot belong to our country!'
            ]
```

```
p.predict(data)
```

```
['racism']
```

Test Cases

```
[30] data = [ 'Womens are lesser than men!'
            ]
```

```
p.predict(data)
```

```
['sexism']
```

Test Cases

```
[39] data = [ 'You nigger people  have to prepay before you eat'
            ]
```

```
p.predict(data)
```

```
['racism']
```

Test Cases

```
[49] data = [ 'Asian peoples are filthy and sucks!'
            ]
```

```
p.predict(data)
```

```
['racism']
```

Test Cases

```
[51] data = [ 'Womens are to be working on the kitchen rather than working on a company!'
            ]
```

```
p.predict(data)
```

```
['sexism']
```

Test Cases

```
[59] data = [ 'Mens are always unhygienic!'
            ]
```

```
p.predict(data)
```

```
['sexism']
```

Test Cases

```
[61] data = [ ' Iam not a sexist but you know mens are always unhygienic!'
            ]
```

```
p.predict(data)
```

```
['sexism']
```

Test Cases

```
[63] data = [ 'Ohh Boy these toys are meant to be used by girls its not for you!'
            ]
```

```
p.predict(data)
```

```
['sexism']
```

Test Cases

```
[65] data = [ 'Ohh you are an Indian  Well that makes sense'
            ]
```

```
p.predict(data)
```

```
['racism']
```

Test Cases

```
[68] data = [ 'Indian peoples are always a trouble maker'
            ]
```

```
p.predict(data)
```

```
['racism']
```

Test Cases

```
Asia covers an area of 44,579,000 square kilometres, about 30 percentage of
Earths total land area and 8 percentage of the Earths total surface area. The
continent, which has long been home to the majority of the human population,
was the site of many of the first civilizations.But asians are crazy,filthy and
such a wimp.The border of Asia with Europe is a historical and cultural
construct, as there is no clear physical and geographical separation between
them. It is somewhat arbitrary and has moved since its first conception in
classical antiquity.



output:


The text is: ['Asia covers an area of 44,579,000 square kilometres, about 30
percentage of Earths total land area and 8 percentage of the Earths total
surface area. The continent, which has long been home to the majority of the
human population, was the site of many of the first civilizations.But asians
are crazy,filthy and such a wimp.The border of Asia with Europe is a historical
and cultural construct, as there is no clear physical and geographical
separation between them. It is somewhat arbitrary and has moved since its first
conception in classical antiquity. '] Result: ['racism']
```

Test Cases

```
[8] data=['Which team do you think will be eliminated tonight? #MKR']
```

```
print("The text is:",data,"Result:",p1.predict(data))
```

```
The text is: ['Which team do you think will be eliminated tonight? #MKR'] Result: ['none']
```

Test Cases

```
[10] data=['No he did not. After killing many he tuned it down to not killing those used for hunting and herding. Pets killed']
```

```
print("The text is:",data,"Result:",p1.predict(data))
```

```
The text is: ['No he did not. After killing many he tuned it down to not killing those used for hunting and herding. Pets killed'] Result: ['none']
```

Test Cases

```
[12] data=['This show is going to air until Christmas isn''t it? ']
```

```
print("The text is:",data,"Result:",p1.predict(data))
```

```
The text is: ['This show is going to air until Christmas isnt it? '] Result: ['none']
```

Test Cases

```
data=['Given that Islam declared war on all humanity 1400 years ago why should we not burn down all Mosques?']
print("The text is:",data,"Result:",p1.predict(data))
```

```
The text is: ['Given that Islam declared war on all humanity 1400 years ago why should we not burn down all Mosques?'] Result: ['racism']
```

Test Cases

```
data=['JUST FELT THE NEED 2 EXPRESS THIS 4 SOME REASON  Dont get me wrong Iam not sexist but seeing female engineers on site is weird']
print("The text is:",data,"Result:",p1.predict(data))
```

```
The text is: ['JUST FELT THE NEED 2 EXPRESS THIS 4 SOME REASON  Dont get me wrong Iam not sexist but seeing female engineers on site is weird'] Result: ['sexism']
```

Conclusion

- The old methods of traditional machine learning were also used by the researchers but they proved to be inefficient and inaccurate by time.
- BERT model which is based on the complex and novel deep neural network, Transformer provides a new approach of detecting bullying.

References

- https://towardsdatascience.com/bert-explained-state-of-the-art-language-model-for-nlp-f8b21a9b6270
- https://www.analyticsvidhya.com/blog/2019/09/demystifying-bert-groundbreaking-nlp-framework/

Thank You