

Github Issue Taxonomy

Team Members

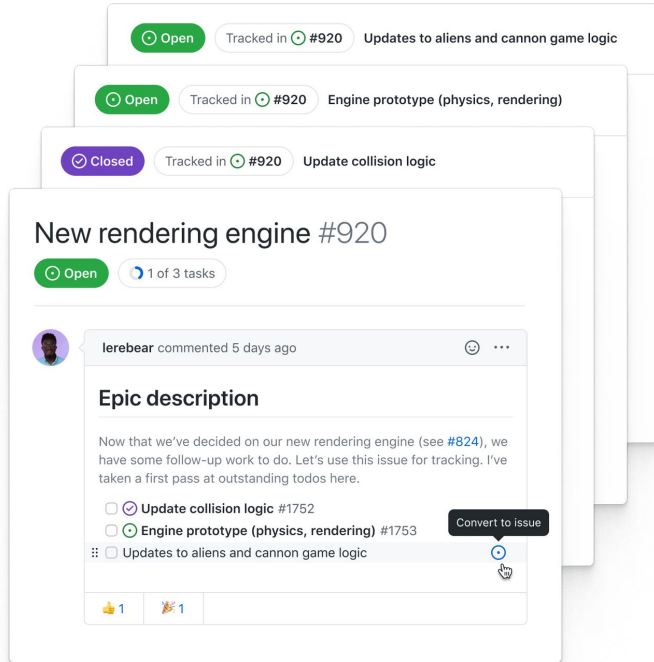
Zaara Zabeen Arpa | ID:200042101

Sadnam Sakib Apurbo | ID:200042135

Nazia Karim Khan Oishee | ID:200042137

[Github Repository Link](#)

What are github issues?



01

Collaborative Issue Tracking

02

Organization and Categorization

03

Integration with Development Workflow

Problem Description



Manual Categorization

- Project managers label the issues manually
- Categories can be bug, feature, question etc



Reduced Efficiency

- Time consuming
- Prone to human error
- Inefficient bug resolving

Solution

01

Automating issue categorization

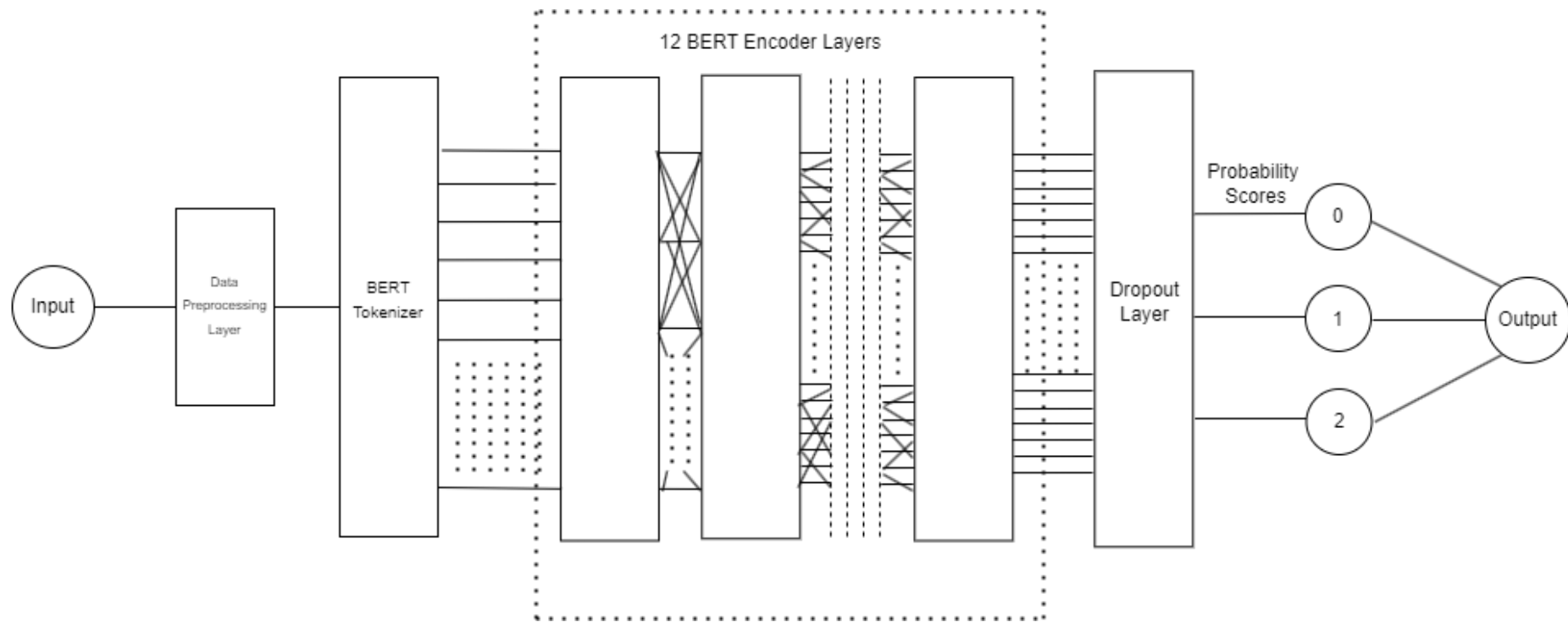
02

Classification using neural network based models

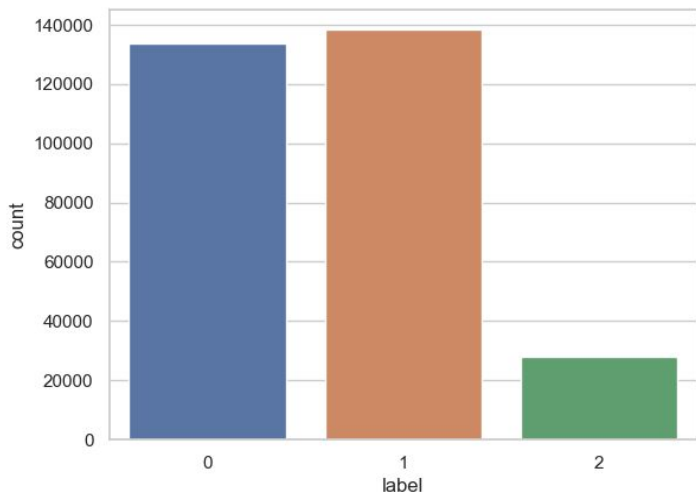
03

Using natural language processing

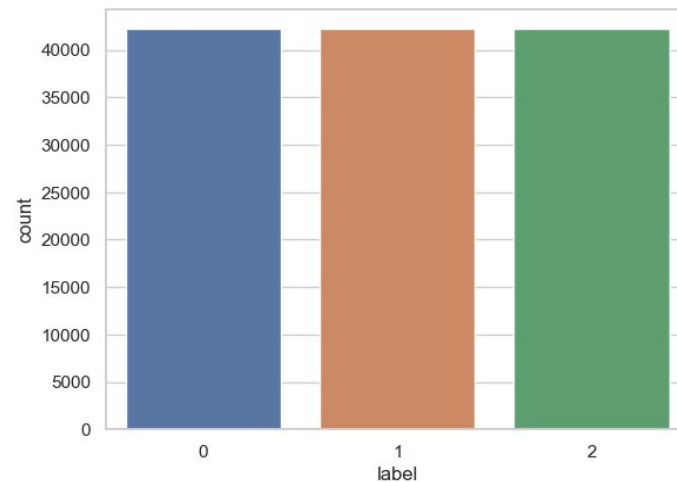
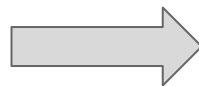
Neural Network Architecture



Implementation



Creating Dataframes & Concatenating Data



Downsampling Data

Data Pre-processing

Input Data

Before preprocessing :

Title :

error creating window on os x 10.12.5

Body :

hi!\r \r when creating window using win, err := window.new \ glfw\ , 800, 600, \



Merge lines, Remove
stopwords, punctuations, URL
etc

After preprocessing :

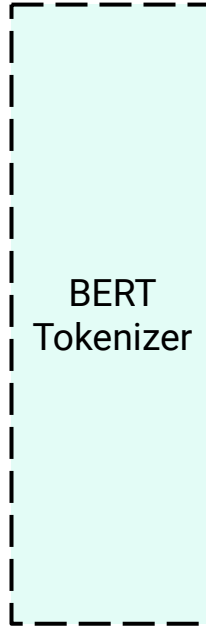
Text :

error creating window os x hi creating window using win err windownew glfw hello

Text before tokenization :

I am facing a bug while installing the package

Text

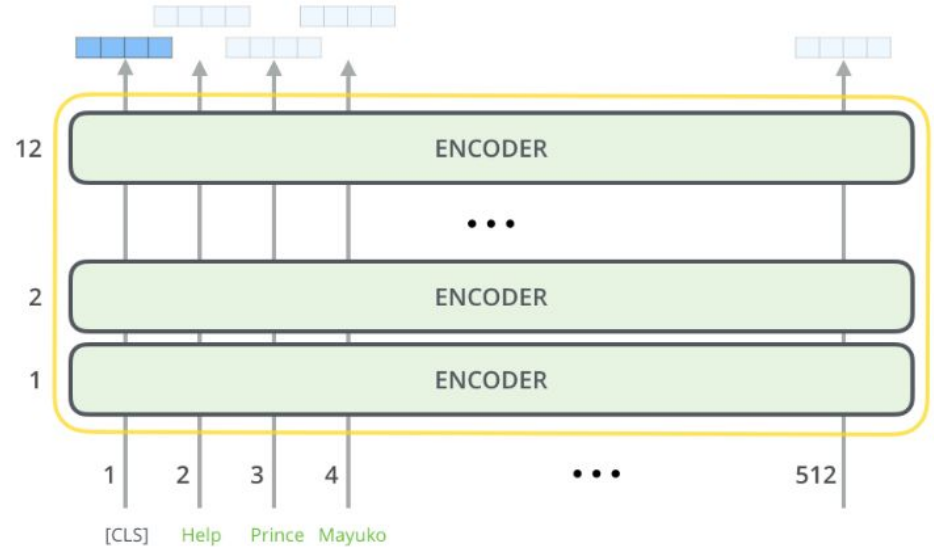


```
After tokenization :  
-----  
['i', 'am', 'facing', 'a', 'bug', 'while', 'installing', 'the', 'package']  
After encoding :  
-----  
tensor([[ 101, 1045, 2572, 5307, 1037, 11829, 2096, 23658, 1996, 7427,  
         102,   0,   0,   0,   0,   0,   0,   0,   0,   0,  
         0,   0,   0,   0,   0,   0,   0,   0,   0,   0,  
         0,   0,   0,   0,   0,   0,   0,   0,   0,   0,  
         0,   0,   0,   0,   0,   0,   0,   0,   0,   0,  
         0,   0,   0,   0,   0,   0,   0,   0,   0,   0,  
         0,   0,   0,   0,   0,   0,   0,   0,   0,   0,  
         0,   0,   0,   0,   0,   0,   0,   0,   0,   0,  
         0,   0,   0,   0,   0,   0,   0,   0,   0,   0,  
         0,   0,   0,   0,   0,   0,   0,   0,   0,   0,  
         0,   0,   0,   0,   0,   0,   0,   0,   0,   0,  
         0,   0,   0,   0,   0,   0,   0,   0,   0,   0])  
tensor([1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0])
```

Input ID
Attention Mask

BERT Base Model

- 12 Encoder layer
- 768 feed-forward networks
- Takes a sequence of words as input
- Each layer passes results through a feed-forward network
- hands it off to the next encoder.



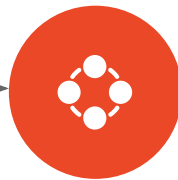
Loss Function



Softmax Activation



**Cross Entropy Loss
Function**



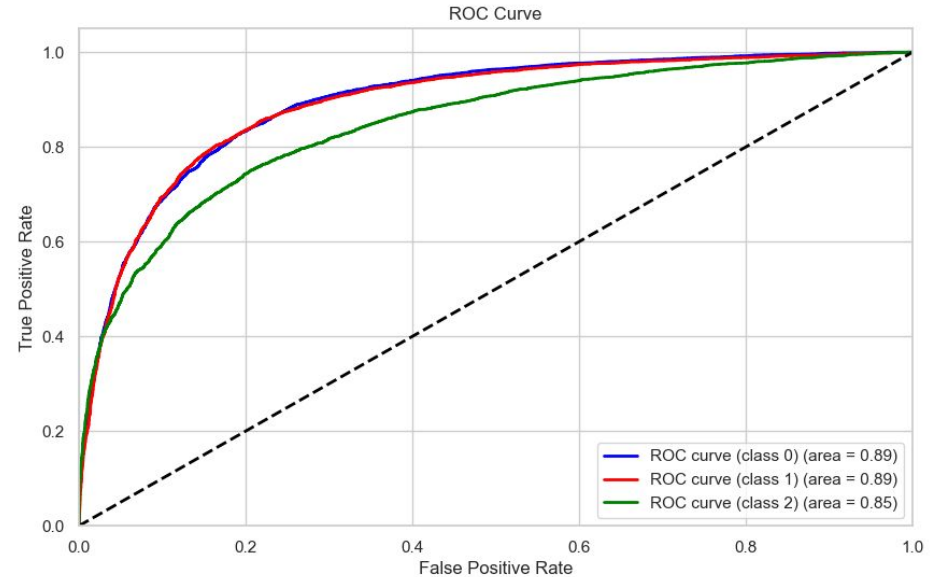
Backpropagation

Model Performance Assessment

F1 Score

- Class 0 : 75%
- Class 1 : 74%
- Class 2 : 70%

Accuracy: 73%



Thank You