

Abstract

1) Algorithm

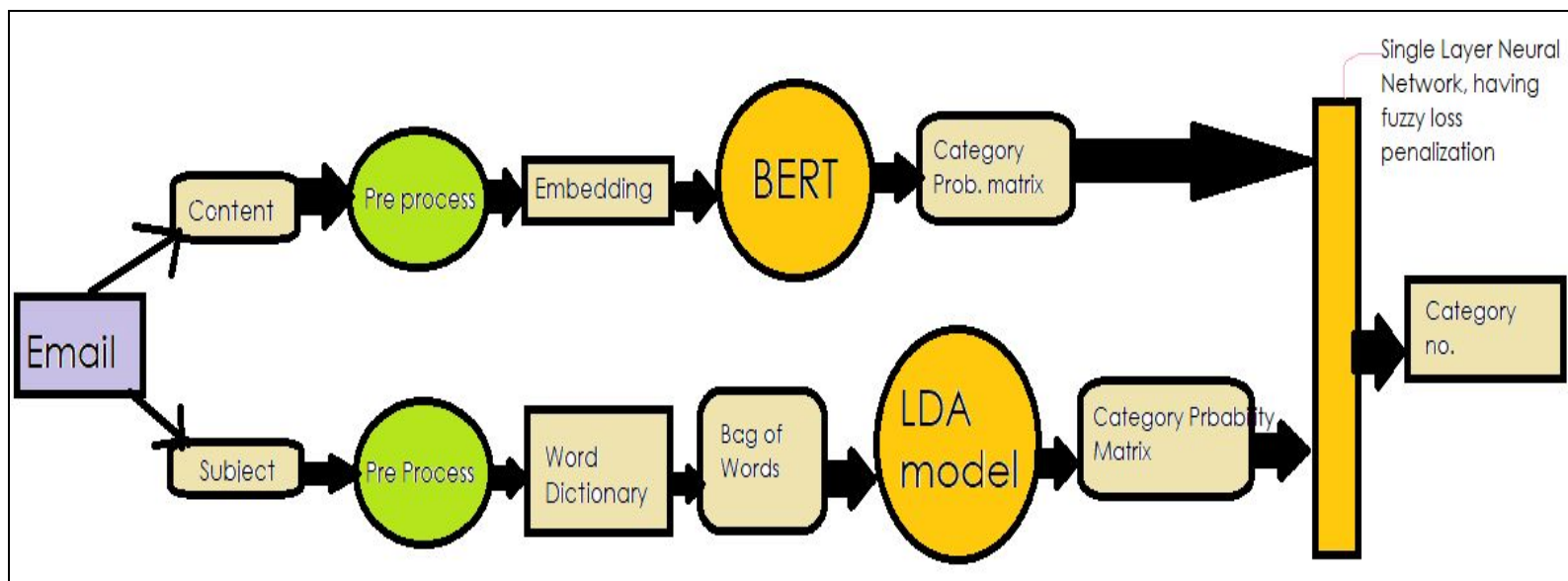
Email is one of the primary means for customers to connect with executives of a company. An email classifier helps a company to bifurcate these emails in order to divide them to different customer support executives.

There are a lot of classification algorithms such as KNN, Ripper, Winnow and MaxEnt[1] to do the task, but the issue with them is that the less data might not give accurate results.

To overcome these limitations, our proposed approach makes use of pretrained BERT and LDA models that require relatively less amount of data to train on. Both these models will be working independently, though at the final layer, the results from both the models would be merged using loss penalization.

BERT will be using the content of the email as an input, while LDA[2] will be using the subject of the email as an input. The raw vector input of BERT and LDA would be sent to the final layer, along with their 1D probability tensor as multi dimensional input to the layer. The output of which will be the final predicted category.

Also, in order to make the models individually capable, we will be using categorical loss between the model's predicted category, and the actual category.



2) Front-end mockups

a) Training page

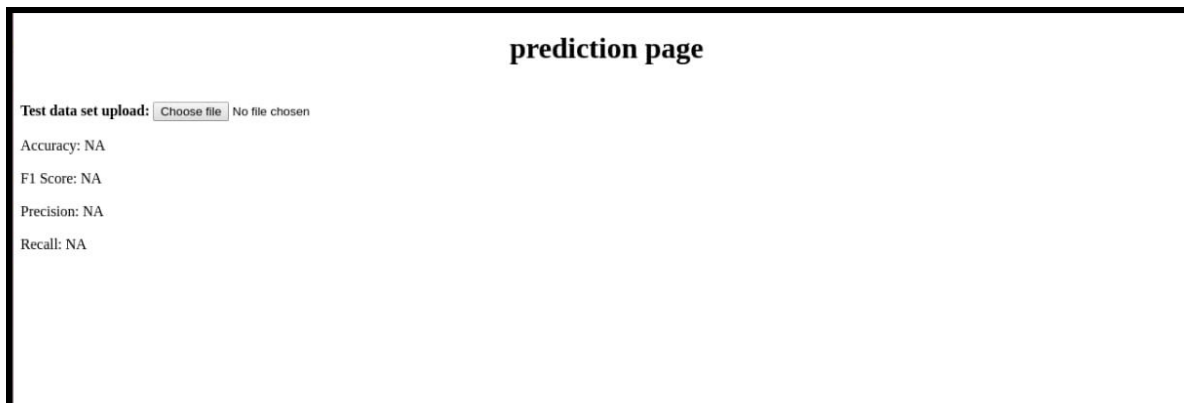
- i) Hyperparameter (text input)
- ii) Dataset upload(file upload)
- iii) F1 score(text field-output)
- iv) Accuracy(text field-output)
- v) Precision(text field-output)
- vi) Recall(text field-output)
- vii) Epoch vs train graph (Image display-plotly)

viii) Epoch vs val graph(Image display-plotly)



b) Prediction page

- i) Test Dataset upload (file upload)
- ii) Accuracy (text field-output)
- iii) F1 score (text field-output)
- iv) Precision (text field-output)
- v) Recall (text field-output)



3) Solution flow

- Get the values/data from frontend

- Preprocess the data
- Train/Test the model using the hyperparameters
- Send the data to frontend

References:

[1] <https://www.sciencedirect.com/science/article/pii/S1319157814000573>

[2] https://github.com/shaival2905/email-classification-using-BERT/blob/master/Future_Technologies_Email_Classification.pdf