COSC 4368

Zitlali Silva

4/9/2022

Data Preprocessing involves the transformation of the raw data into an understandable format. As well, it is a fundamental stage in data mining to improve data efficiency. The steps to do that consist of:

1. Gathering data and import the dataset & libraries
2. Dealing with missing values
3. Dividing the dataset into Dependent (y values on the code) and Independent (x values on the code)
4. Splitting the dataset into training and test set by using from sklearn. model\_selection import train\_test\_split
5. Creating Bert Models: tiny, mini, small, medium
6. Creating Precision, Recall, F1\_score for each of the Bert Models which are the results on the table

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Bert Model | Train Accuracy | Test Accuracy | Precision | Recall | F1\_score |
| Bert Tiny | .802 | 0.805 | 0.83 | 0.76 | 0.80 |
| Bert Mini | .833 | 0.838 | 0.83 | 0.85 | 0.84 |
| Bert Small | .840 | 0.847 | 0.86 | 0.84 | 0.85 |
| Bert Medium | .852 | 0.858 | 0.85 | 0.87 | 0.87 |

The hypermeters and their values were constant throughout the: epoch = 1, learning\_rate = 3e-5, optimizer = tf. keras , batches = np. reshape(sentences,(-1, 50)).

The best Bert model will be Bert Model Medium due the numbers being a higher percentage in both accuracies and closer to 1 value for the F1\_score.

For improvements, you can also change the epoch = 2 to make it more efficiently and better results for accuracy.