**Observations Criteria:**

Note:

1. **preprocessed\_data** was used for training data processing. The training data/tweets are converted into tokens of words after punctuation removal and lowercase conversion.
2. Simple pre-processing also tried to improve the accuracy of the MLP model. But not a significant change in the accuracy.

* **MLP model accuracy with different factors:**
* **MLPClassifier – For better understanding, didn’t change the default features and used the default solver(adam) and activation function (ReLu)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Vector size** | **Number of neurons** | **Accuracy** | **Prediction** |
| **Layer 1** | 300 | 30 | 60% | Off and Non-offensive |
| **Layer 2** | 300 | 30,30 | 28% | Off and Non-offensive |
| **Layer 3** | 300 | 30,30,30 | **71%** | Off and Non-offensive |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Vector size** | **Number of neurons** | **Accuracy** | **Prediction** |
| **Layer 1** | 500 | 30 | **71.5%** | Off and Non-offensive |
| **Layer 2** | 500 | 30,30 | 71% | Off and Non-offensive |
| **Layer 3** | 500 | 30,30,30 | **71.5%** | Off and Non-offensive |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Vector size** | **Number of neurons** | **Accuracy** | **Prediction** |
| **Layer 1** | 500 | 50 | 70.3% | Off and Non-offensive |
| **Layer 2** | 500 | 50,50 | **72.2%** | Off and Non-offensive |
| **Layer 3** | 500 | 50,50,50 | **72.1%** | **Non-offensive** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Vector size** | **Number of neurons** | **Accuracy** | **Prediction** |
| **Layer 1** | 500 | 50 | 70.3% | Off and Non-offensive |
| **Layer 2** | 500 | 30,50 | **71.74%** | Off and Non-offensive |
| **Layer 3** | 500 | 30,50,50 | **71.74%** | Off and Non-offensive |

**SUMMARY OF ANALYSIS OF RESULTS:**

1. When stop-words are removed from the training dataset, the accuracy was around 72%, the same for the MLP model trained on single, two layers, or three layers.
2. Only non-offensive predictions were observed with zero offensive predictions, which shows that the model is over-fitted.
3. Stop-words are removed during pre-processing of the training dataset, and thereafter model was able to capture different accuracy for different layers.
4. MLP model trained on a single layer with a vector size of 300 and a number of neurons of 30 for the single layer, that predicted both offensive and non-offensive labels with 60% accuracy.
5. When the MLP model is trained on two layers with neuron size as (30,30) for both layers the results were not as expected. The accuracy dropped to 28%, which should not be because, with an increased number of layers or networks, the model performs slightly better.
6. But with three layers, vector size 300, and neurons (30,30,30), the model performed well with an accuracy of 71%, which is expected.
7. The second batch of model performance testing is done by increasing the dimensionality of the word vector from 300 to 500 but keeping the number of neurons the same for each layer.
8. It is observed that the change in vector size improved the performance of the MLP model for each combination of layers
9. The accuracy of the MLP model trained on two layers escalated from 28% to 71%
10. However, when the number of neurons unit in each layer has been increased from 30 to 50, the MLP model trained on three layers predicted only non-offensive labels
11. This is overfitting, where the network reached a point where it learned the noise in the dataset
12. To further test the model in order to overcome overfitting, fine-tuned the neuron units to a certain level i.e. (30,50,50) for three layers. The model accuracy is now 71.74% which is the same as the MLP model trained on two layers.
13. Overall, it can be concluded that the feed-forward neural network MLP model, can perform better if the capacity of the network is sufficient enough to learn the given task.
14. Because the capacity of the model is directly related to the bias/variance trade-off
15. If the model network is not strong enough to learn the complex information it might affect the accuracy
16. However, if the model capacity is more than enough then it might lead to overfitting and learning noise in the training dataset
17. Fine-tuning of vector size, number of layers, and number of neuron units with non-removal of stop-words turned out to be the deciding factors for MLP model capacity to perform well.