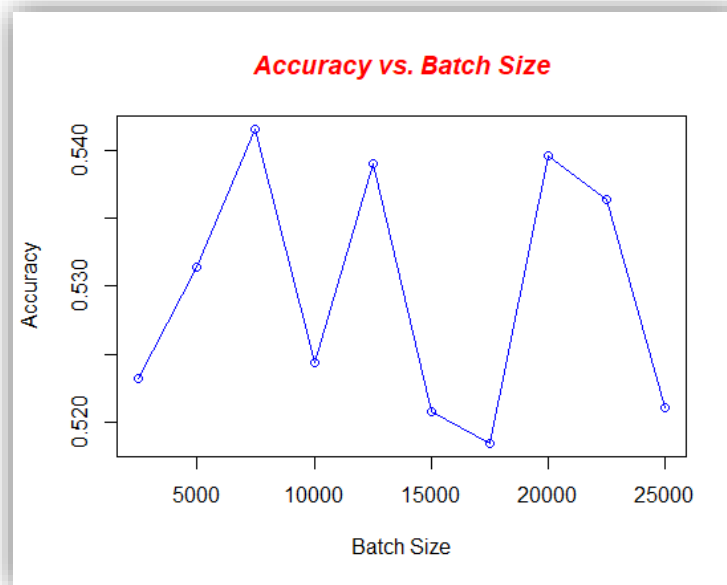


MACHINE LEARNING WITH SPARK MLlib

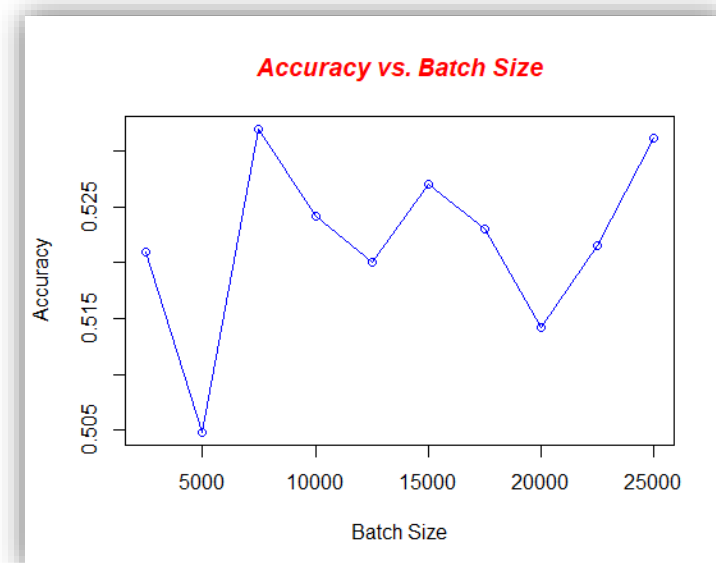
Performance Analysis of Bernoulli Model:

Accuracy vs Batch size: Accuracy is plotted against varying batch sizes(5000, 10000, 15000...) to study the resulting trend

Train Data:



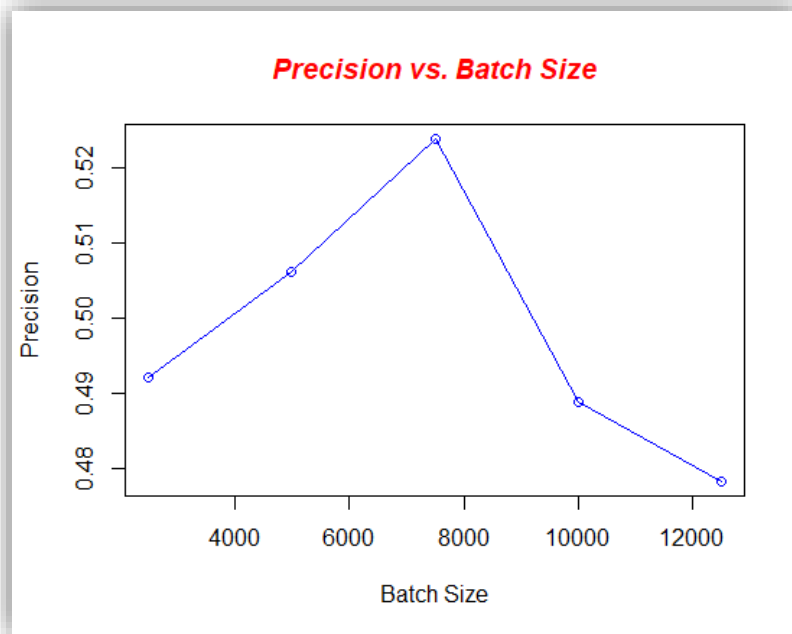
Test Data:



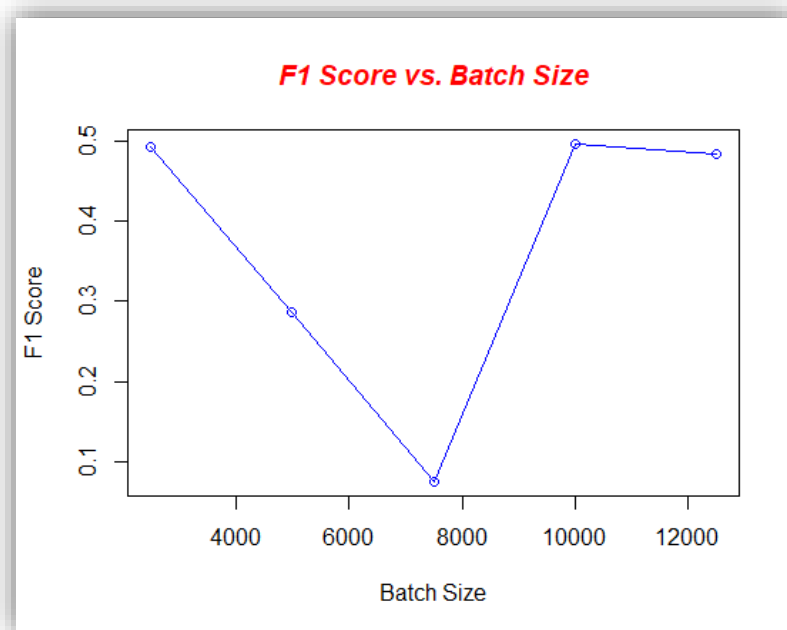
Performance Analysis of Perceptron Model:

Precision vs Batch size: Accuracy is plotted against varying batch sizes(5000, 10000, 15000...) to study the resulting trend

Train Data:



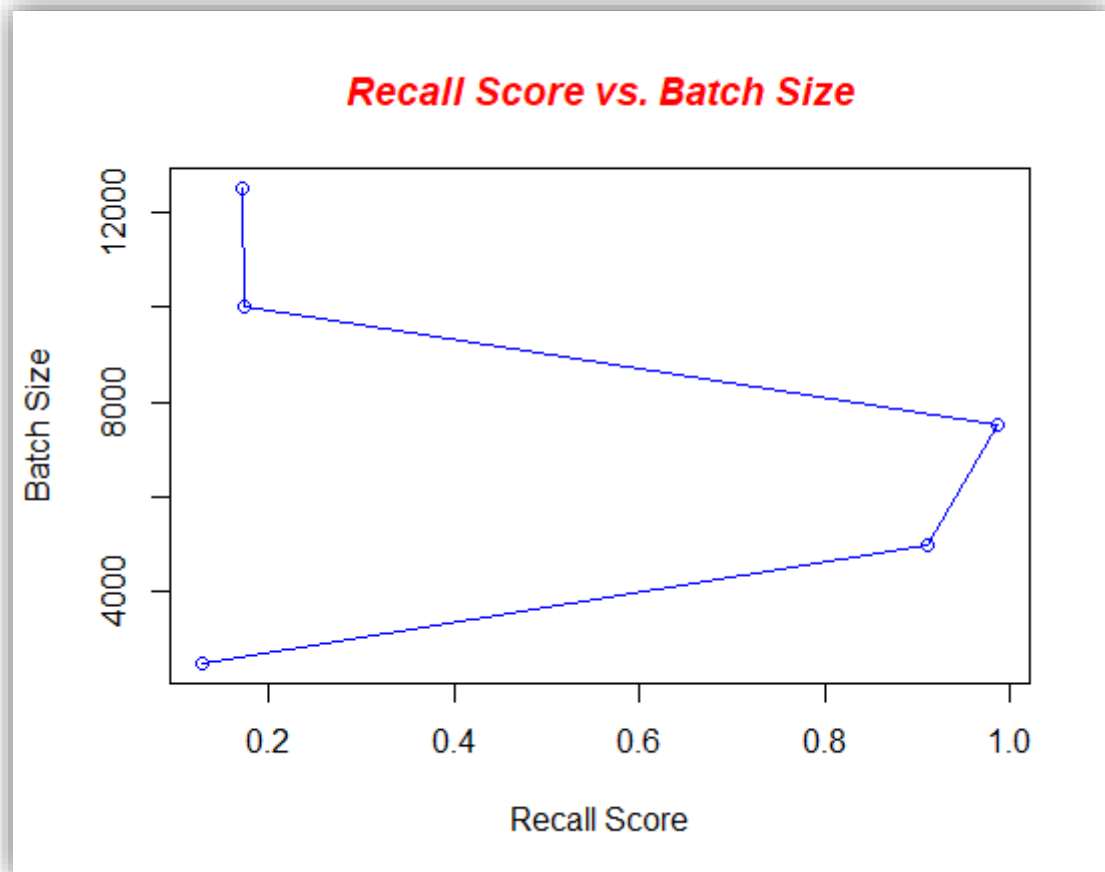
F1 vs Batch Size:



Performance Analysis of SGD Classifier:

Recall vs Batch size: Accuracy is plotted against varying batch sizes(5000, 10000, 15000...) to study the resulting trend

Recall vs. Batch Size:



Confusion Matrix:

Batch Size: 5000

```
Confusion Matrix:  
[[440 383]  
 [412 415]]
```

Batch Size: 10000

```
Confusion Matrix:  
[[760 879]  
 [701 960]]
```

Batch Size: 15000

```
Confusion Matrix:  
[[1437 1076]  
 [1274 1163]]
```