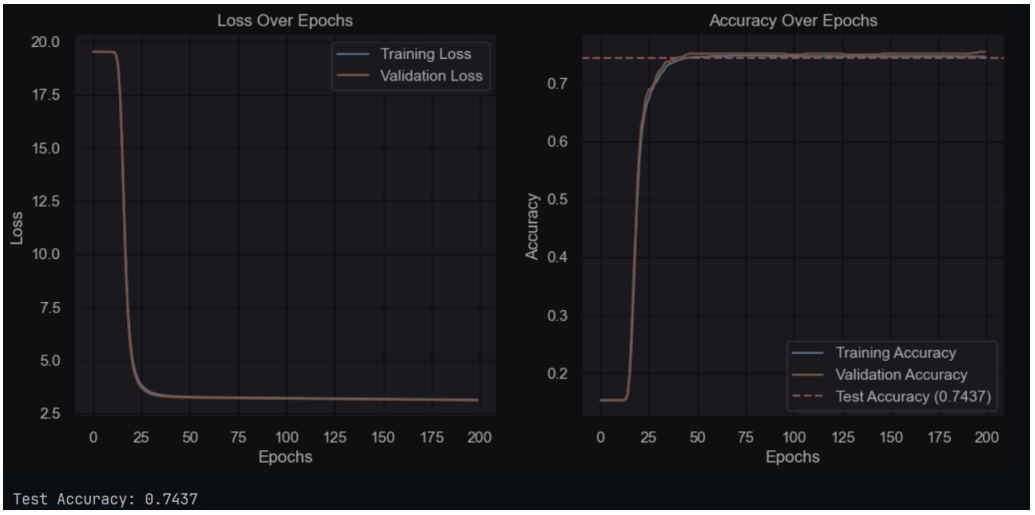
ML-01 Report Logistic Regression

Devyansh Chaudhary [2022156]

I have inserted markdowns in code as answers to my problems asked in the assignment here are plots and some descriptions for them only.

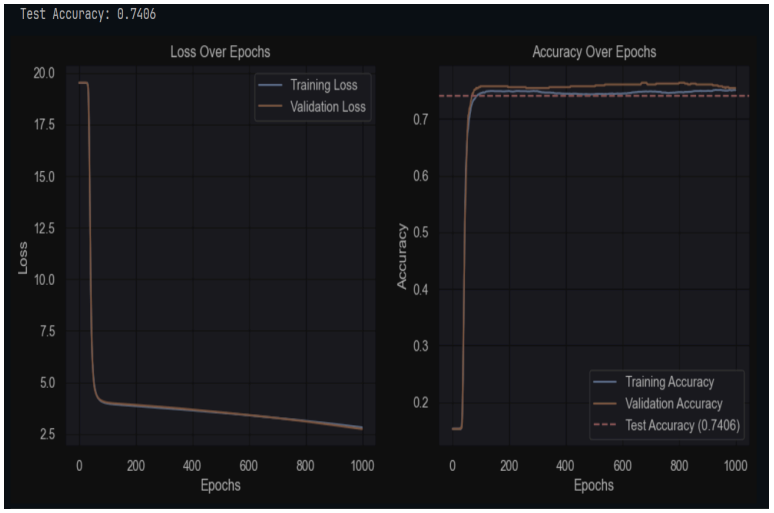
Part-a

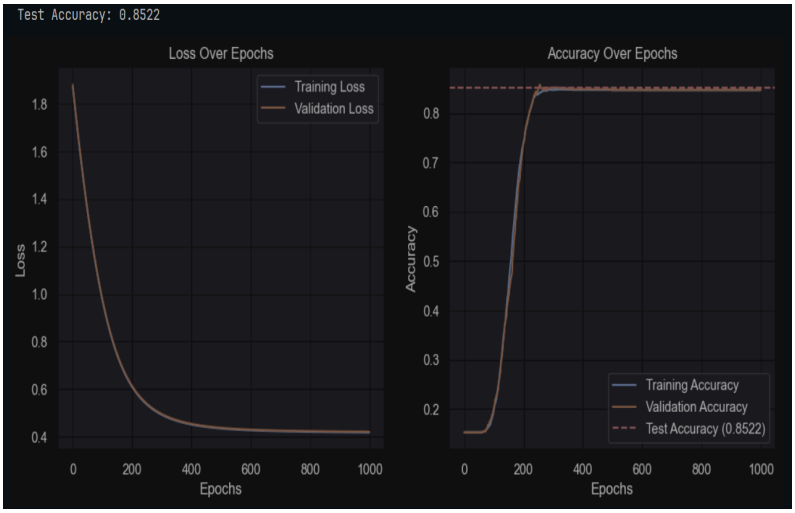
After applying the Batch gradient descent on the given dataset i got the results attached below. The Learning Rate: 0.0001 and epochs: 20. Were used for training this model.
I got an Accuracy of 74.34%. I did not find any evidence of overfitting or underfitting.



Part-b

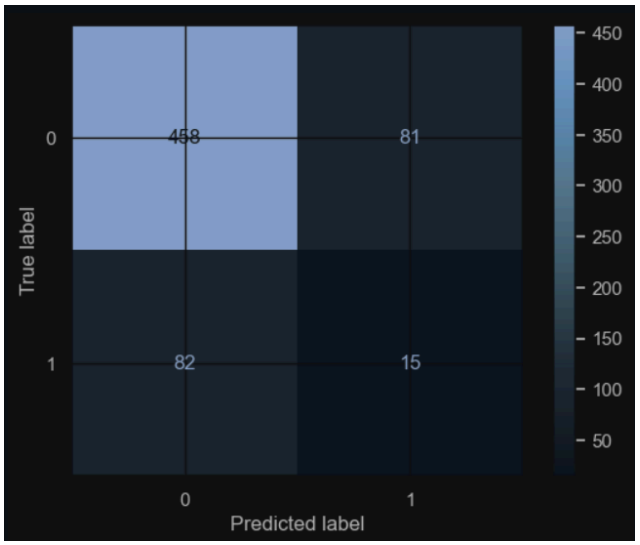
Figure 2 is after applying MinMax to my dataset which shows smoother loss and accuracy function with no early plateau and meaning full gradients also got better accuracy of 85% than NoMinMax. Learning Rate: 0.01 and epochs:1000 used for minmax and Learning Rate: 0.0001 and epochs 1000 in case of no minmax.





Part-C

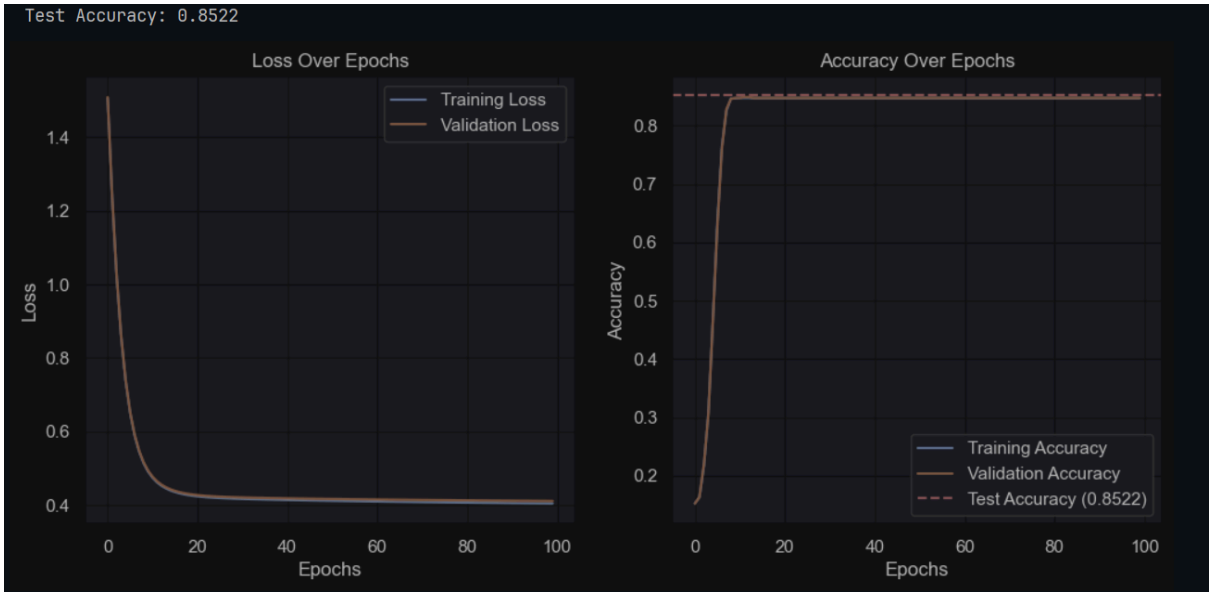
The confusion matrix shows that the model is facing issues in classifying class 1 because of Class Imbalance i fount in the given dataset so the model got more trained over class 0, This result clearly shows and affected my confusion matrix.



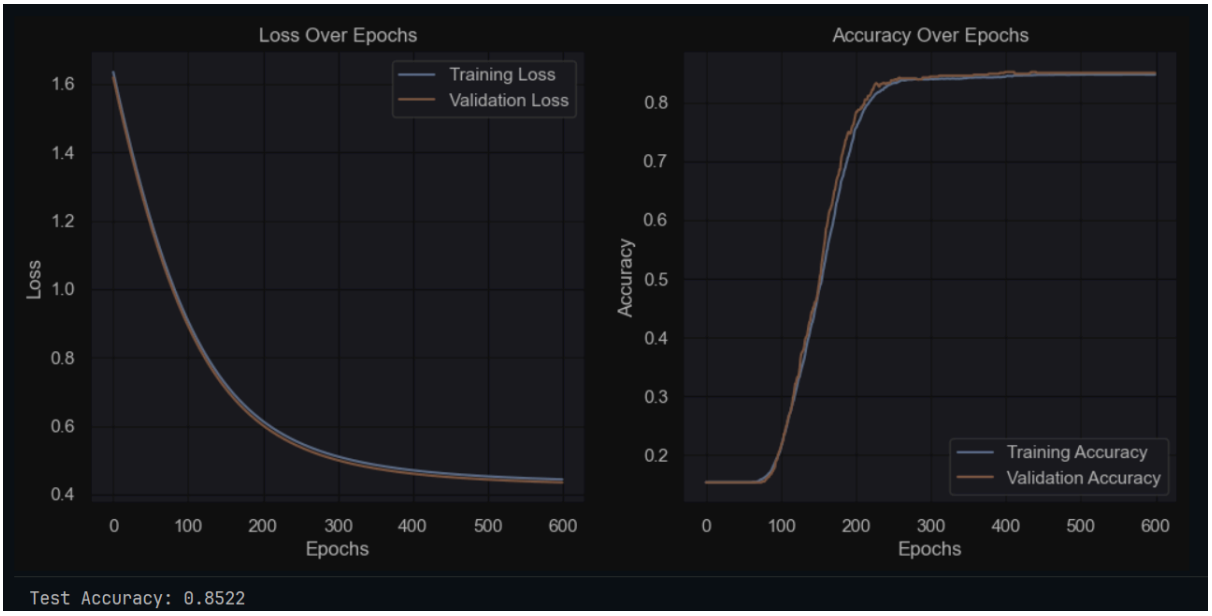

```
Precision: 0.1562
Recall: 0.1546
F1 Score: 0.1554
ROC-AUC Score: 0.4697
```

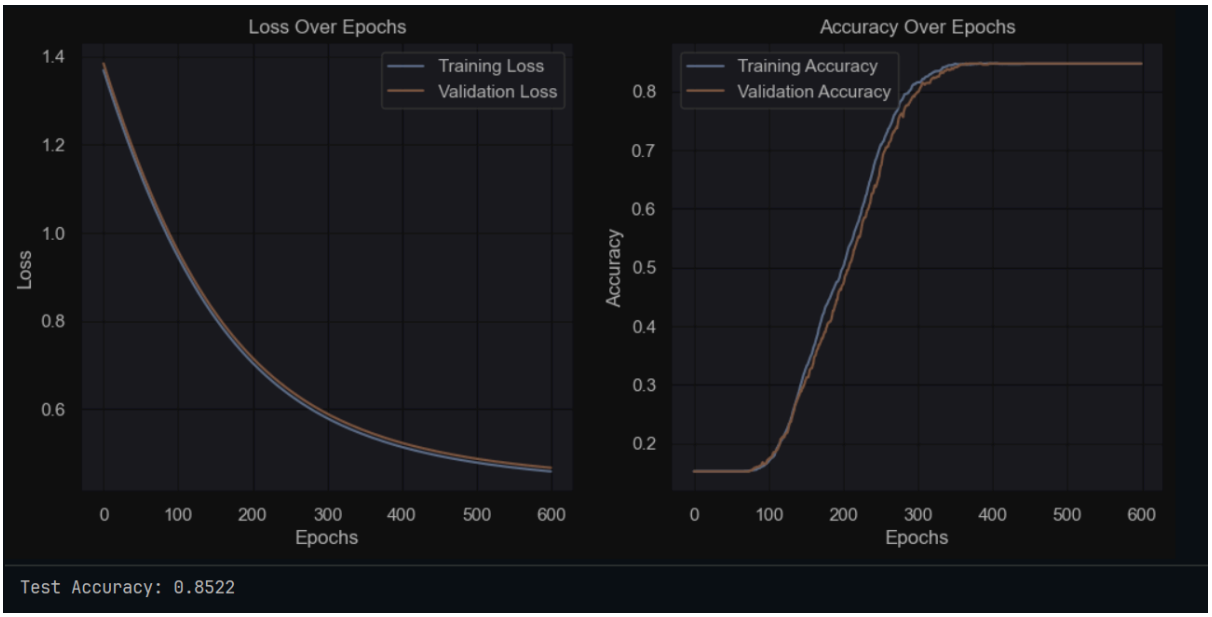
Part-D

Stochastic Gradient Descent:
Got an Accuracy of 85% at just 200 epochs the model saturated its accuracy. The model randomly selects one datapoint and applies gradient descent on that which is shows after applying min-max the dataset which made my all gradient descent models more stable.



Mini Batch:
Batch Size 32 and 52 on increasing the batch size i found that the loss functions become more smoother and stricter which as per theory i read about on increasing batch size the gradient descent becomes more accurate both got similar accuracy of 85.38% and 85.22%.

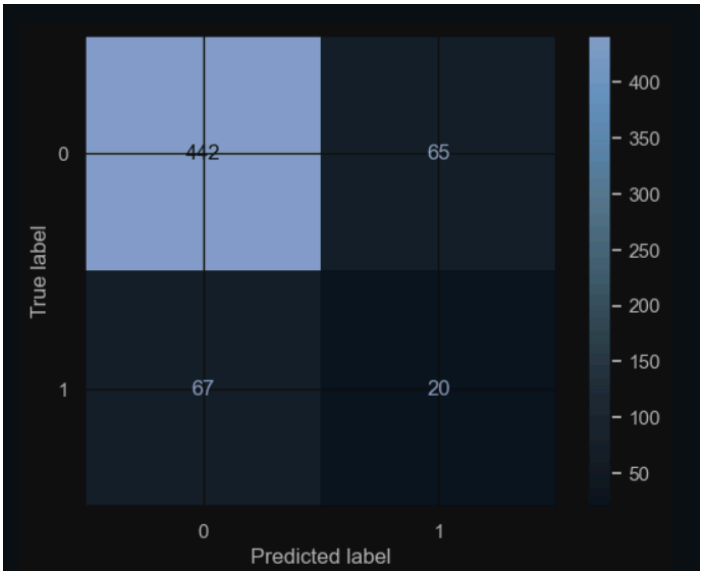




Part - E

K-Fold Cross validation The class imbalance explains the discrepancy between the relatively high accuracy and the low precision/recall/F1 scores.Based on the provided cross-validation results, the model appears to be both stable and have low variance.

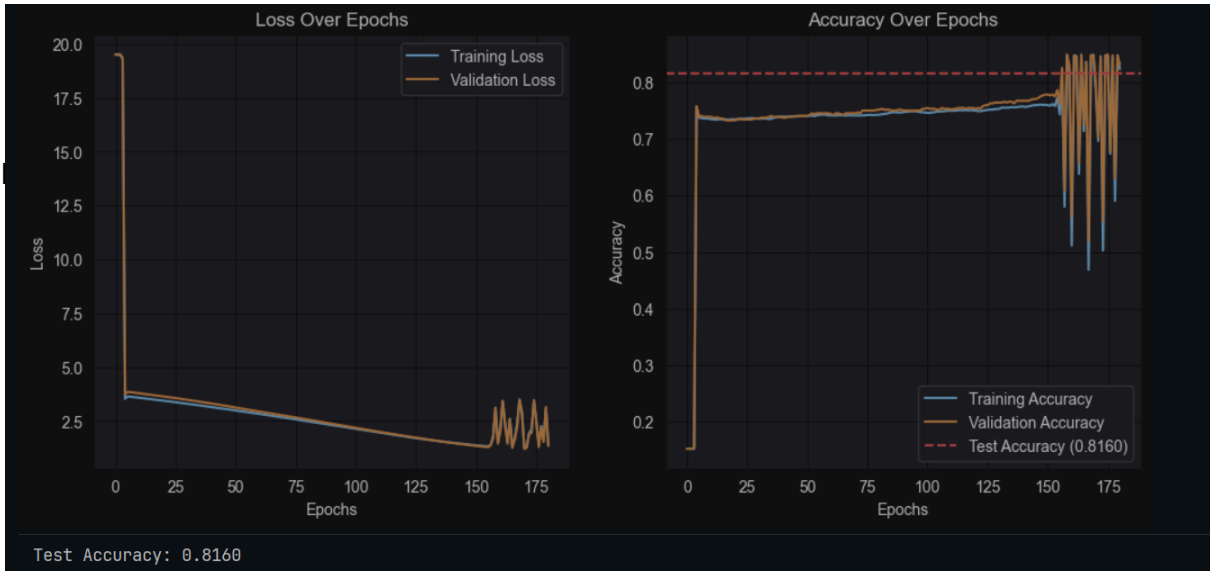
```
Cross-Validation Results:
Average Training Accuracy: 0.7426 ± 0.0113
Average Validation Accuracy: 0.7414 ± 0.0368
Average Training Loss: 2.4390 ± 0.3722
Average Validation Loss: 2.4627 ± 0.5827
Average Precision: 0.2004 ± 0.0639
Average Recall: 0.2174 ± 0.0460
Average F1 Score: 0.2071 ± 0.0523
```



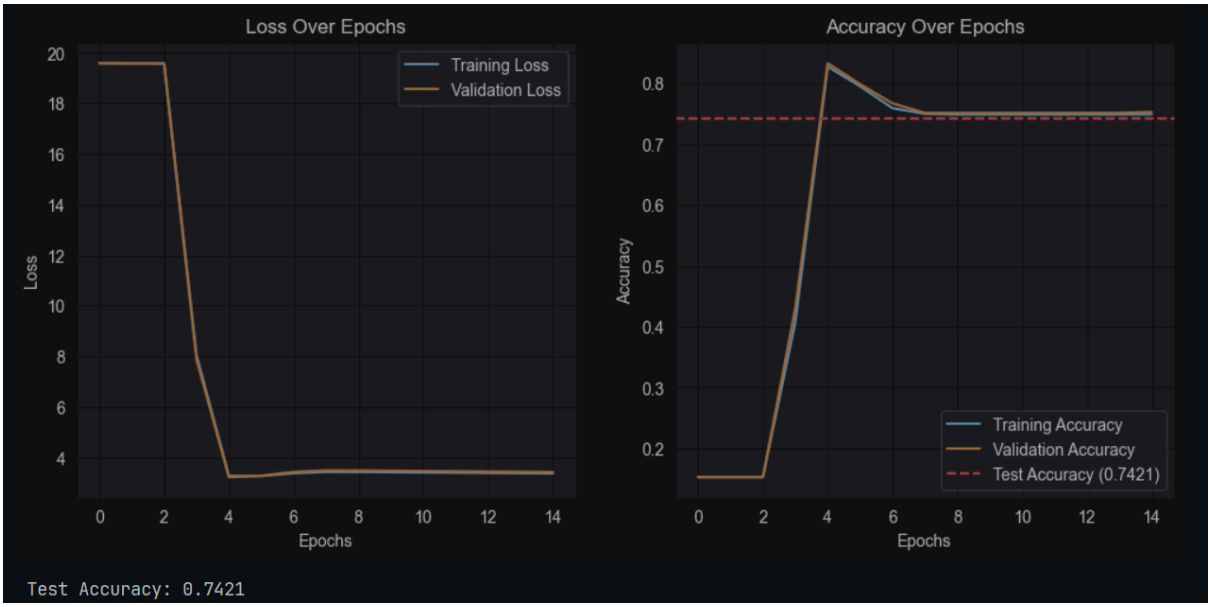
Part-F

In the first two cases (None, L1), early stopping helped prevent large gaps between training and validation performance. L2 regularization has shown the **best result** in test accuracy, indicating that this combination of early stopping and regularization was most effective at generalization.

No Regularization: 81



L1 Regularization: Acc 74.2



L2 Regularization: Acc: 73.74%