CSL7370 : Dependable AI

Assignment1 : Bias - Detection, Mitigation and Evaluation

A report submitted in fulfillment of the requirements for Assignment

by

Nandini Saini (MP19AI003)

Date of submission :27 December 2020

Submitted to
Dr. Richa Singh

Indian Institute of Technology, Jodhpur

1 Question 1 :

1.1 Class-Wise Accuracy

	Classwise Accuracy
Aamir_Khan	26.7%
Aishwarya_Rai	30.3%
Alia_Bhatt	50.0%
Amitabh_Bachchan	21.7%
Ayushmann_Khurrana	13.0%
Hrithik_Roshan	32.1%
Irrfan_Khan	0.0%
Kartik_Aaryan	21.4%
Kriti_Sanon	37.5%
Sonakshi_Sinha	31.6%

Figure 1: Class Wise Accuracy

1.2 Overall Accuracy

```
print ("Overall Accuracy:",'{:.3%}'.format(accuracy))
```

```
Overall Accuracy: 28.938%
```

Figure 2: Overall Accuracy

1.3 Model Bias Estimation

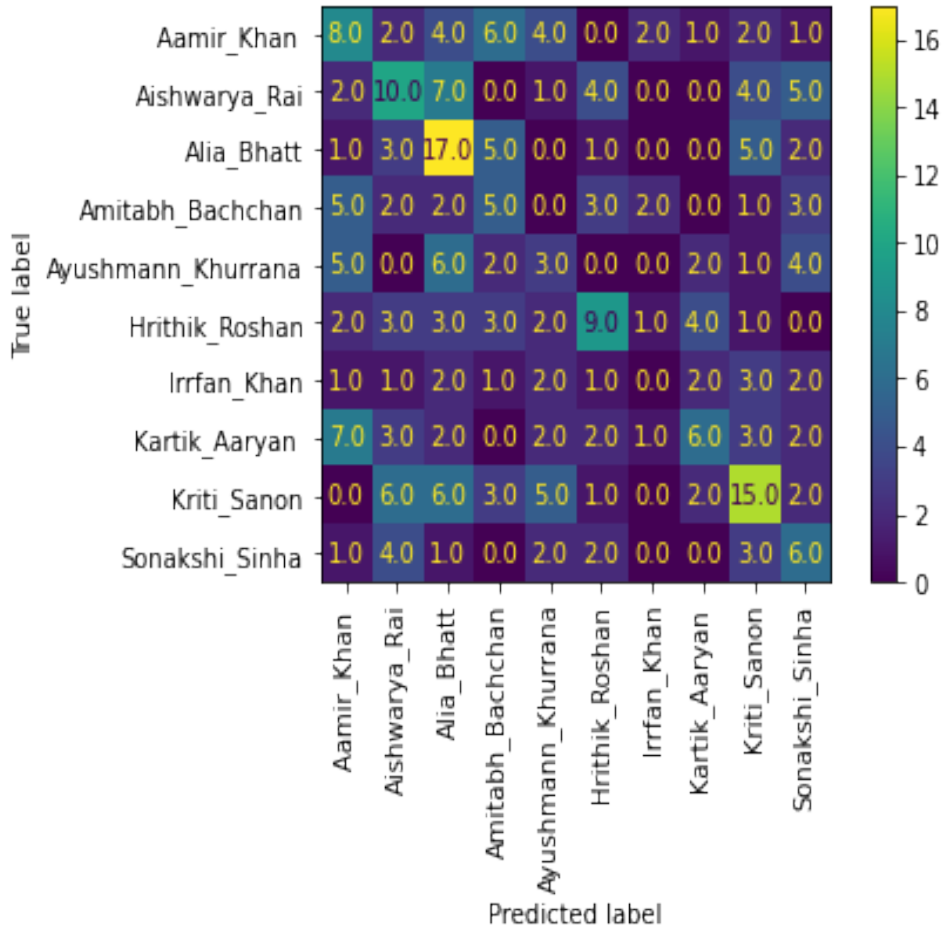

Figure 3: Confusion matrix

```
MSE : 16.382783882783883
 Bias :  12.788461538461538
Variance :  3.594322343223444
```

Figure 4: Degree of Bias (Bias Variance trade off)

1.4 Bias Type Analysis

Based on my analysis, in this question these are biases :

- Sample bias due to non uniform data distribution and size of the dataset
- Algorithm bias due to parameter setting in the svm classification machine learning algorithm
- This suffers from human bias also because we are selecting classes from the data set for the training, so it can call as selection bias which is type by human bias.

2 Question 2 :

2.1 Testing performance (mean ± std)

Testing Performance and Accuracy

Experiment	SVM	Neural Network
1	0.5885	0.7165
2	0.6045	0.6955
3	0.6645	0.729
mean	0.619167	0.713667
standard devation	0.0327143	0.0138223

Figure 5: Testing Performance and accuracy

2.2 Testing Accuracy and Confusion matrix

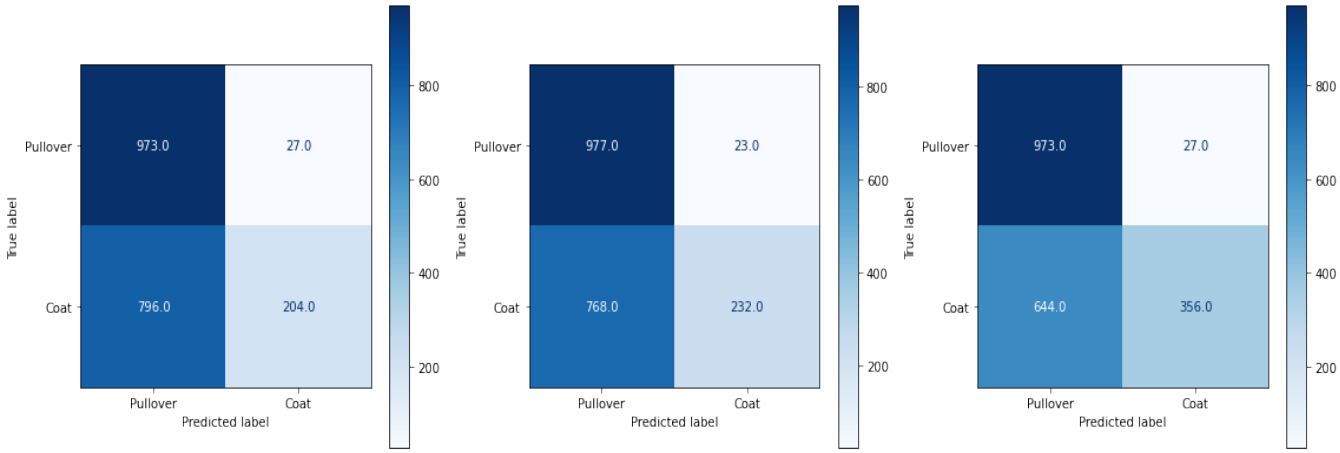

Figure 6: Confusion Matrix for SVM

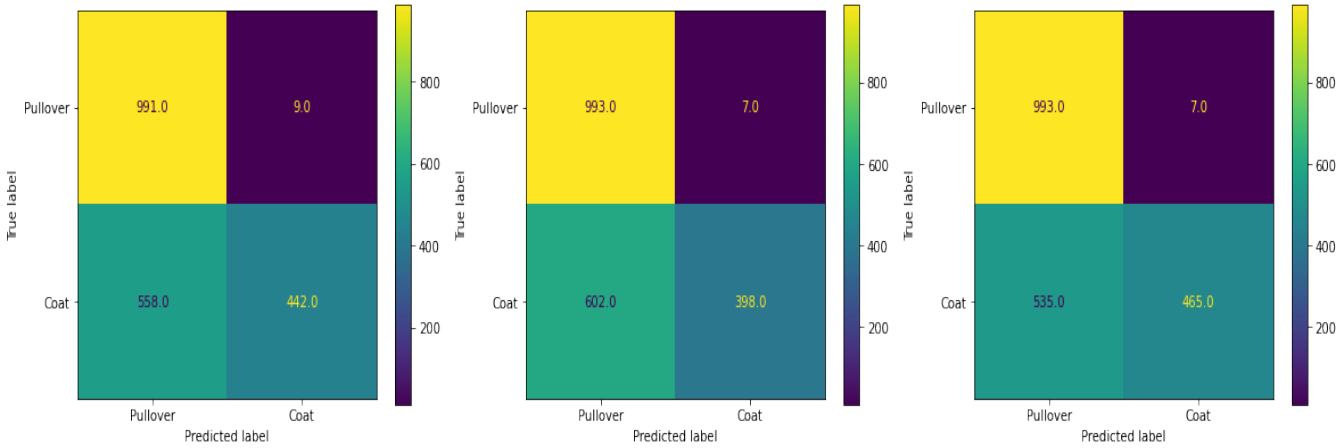

Figure 7: Confusion Matrix for Neural Network

2.3 ROC curve and Equal Error Rate (EER)

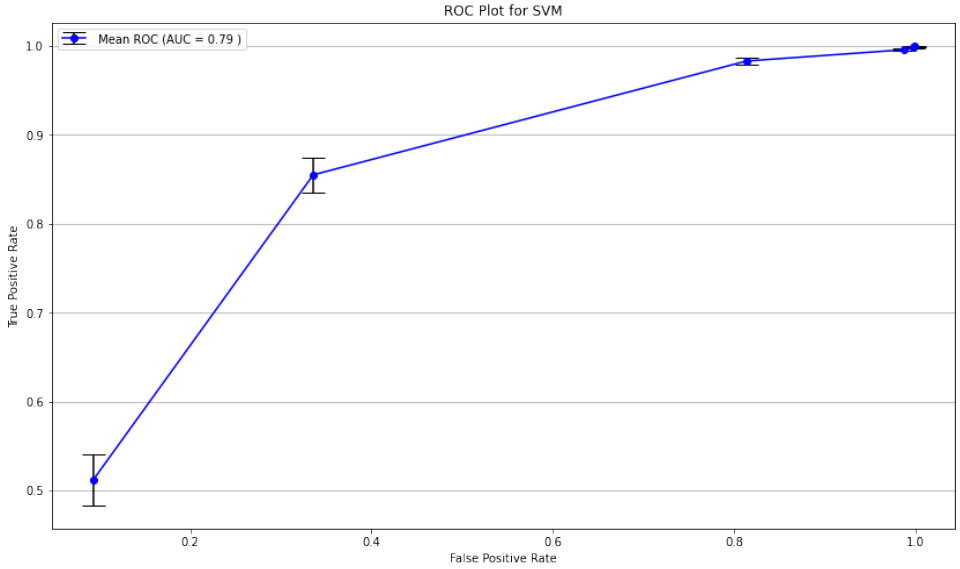

Figure 8: ROC curve for SVM

```
Equal Error Rate in SVM with threshold :
EER 0.2403
Threshold :  0.9
```

Figure 9: EER in SVM

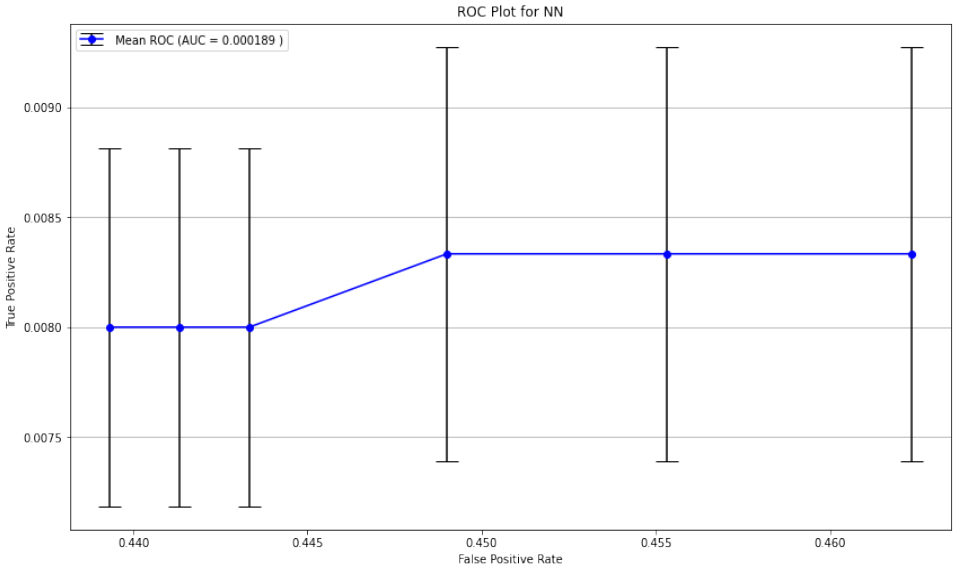

Figure 10: ROC curve for NN

```
Equal Error Rate in 5-layer Neural Network with threshold :
EER 0.7270
Threshold :  0.2
```

Figure 11: EER in NN

2.4 Precision-Recall curve

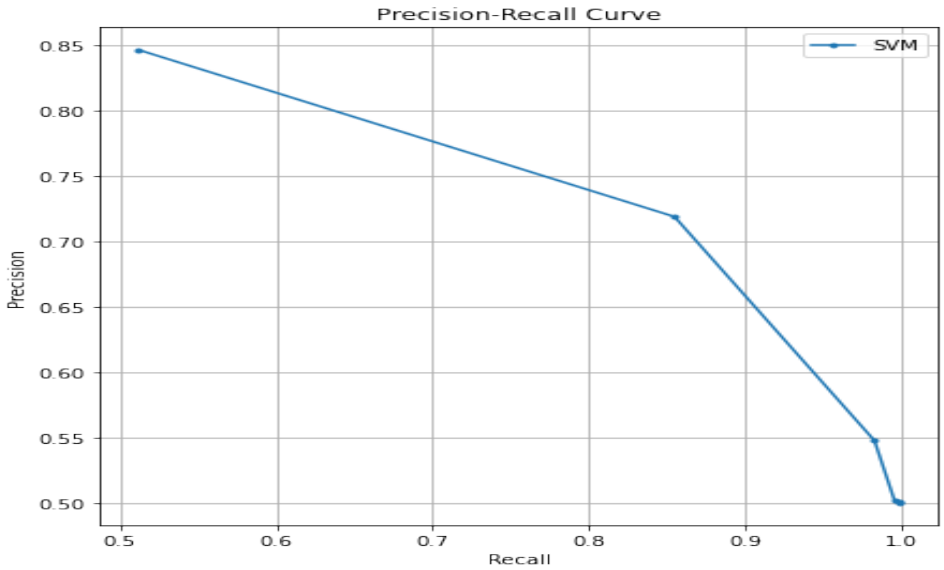

Figure 12: Precision -Recall curve for SVM

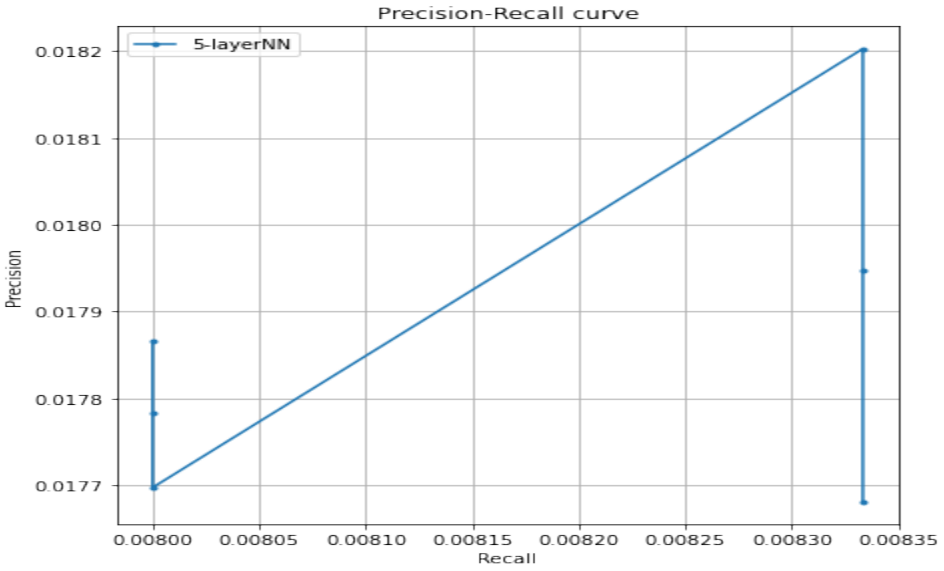

Figure 13: Precision -Recall curve for NN

2.5 Analysis of curve on biased/imbalanced data

ROC Curve : ROC curve is a plot between True Positive Rate (TPR) versus False Positive Rate (FPR) with different thresholds.ROC curve provides a better view of the model on different thresholds. In roc, if the threshold for predicting positive is really high such as 0.99, then we will have a high true positive rate, which is going to be at the top of our chart(Perfect model), but also a high false-positive rate, which is also going to be to the right of our chart(Worse Guessing). If we pick a shallow threshold, say 0.01, then we will have a low false-positive rate to the left of our plot and a low true positive rate, which is our bottom bar plot(Better Plot).ROC curve is appropriate when observations are balanced between each class. The area under the curve (AUC) in ROC measures how well we are separating the

two classes.

Precision- Recall Curve : The precision-recall curve is a plot between TPR and positive predicting label using different thresholds, and it is an unbalanced metric. This will mostly be a decreasing curve. Here area under the curve will depend on how unbalanced our data set is.
The right curve will depend on tying our results, so true positive versus true negative to our outcomes, and the relative costs of false positives versus false negatives.Generally ROC curve use for data with balanced classes. When data with imbalanced classes , then the precision-recall curve will generally be better suited.

2.6 Analysis of algorithm

Accuracy : Based on my analysis and performing classification with both the algorithm, the neural network performs better in every experiment concerning testing performance and accuracy. The neural network's overall accuracy is 71.36%, whereas, in SVM, accuracy was 61.9%. Therefore based on accuracy, NN is performing better than SVM. But accuracy is not only one parameter to decided on a good or bad approach.

Analysis of ROC : The AUC gives the measure of how well classifying two classes. Based on AUC, SVM performed well because the AUC of SVM is 0.79 and the AUC of NN is 0.00189.

Analysis of EER: Equal error rate (EER) is where your false pos rate (fpr) == false neg rate (fnr). Smaller EER is better for the model. For better performance, TPR, TNR should be high, and FNR, FPR should be low. The EER of SVM is 24% and 72% of the neural network. Therefore SVM performance is better.

Overall SVM is better approach compare to NN in this experiment due to smaller dataset and hyper parameter setting.

3 Question 3 :

3.1 Confusion Matrix for 5 classes

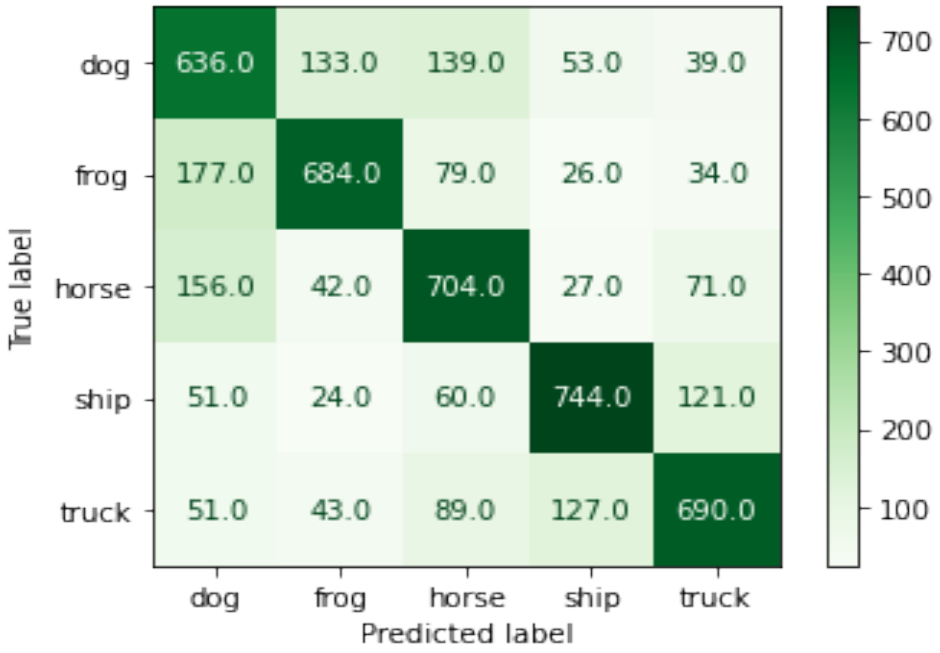

Figure 14: Confusion Matrix

4 Question 4 :

4.1 Bias Estimation Metrics

Fromm the below ,Bias Estimation Metrics values we can say that in the given problem model has bias.

```
Accuracy and Loss on the testing data :  0.9318897724151611 0.06811022013425827
```

Figure 15: Accuracy

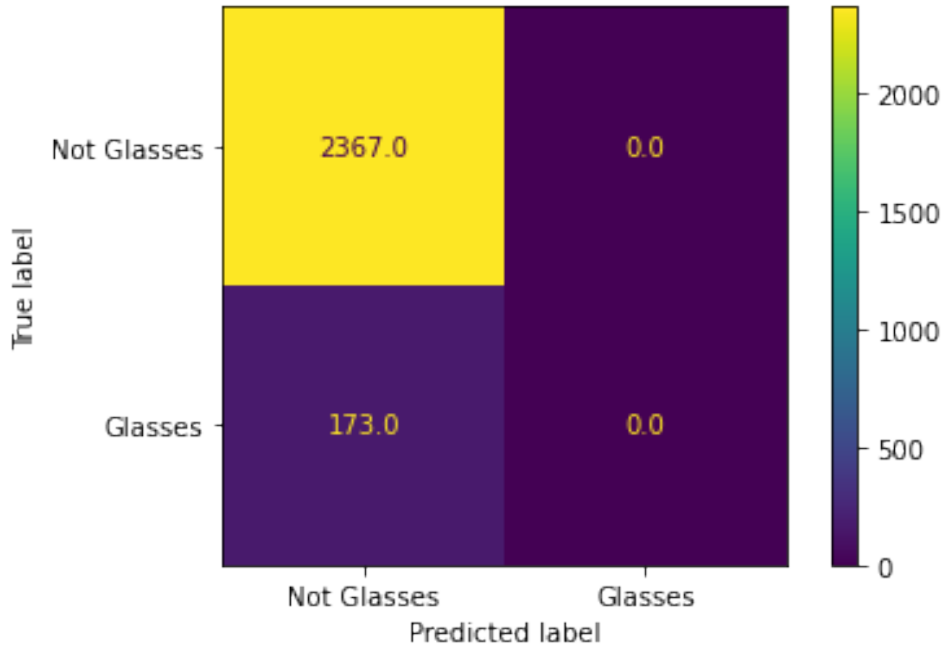

Figure 16: Confusion Matrix

```
[35] disp_impct
     0.0
```

Figure 17: Disparate Impact

```
[37] mse_loss, bias_nn, var_nn
     (0.06811023622047244, 0.06811023622047244, 0.0)
```

Figure 18: Degree of bias

4.2 New Evaluation Metric

Method 1 : The confusion matric tells the performance of the classification model. From the confusion metric, we can calculate TPR, which summarizes how well the positive label predicted. TNR summarizes how well the negative label predicted. If we consider both the values together, then we are concerned about both the classes equally. Both the classes are equally important; then, this score can tell about bias in the model. We can take the arithmetic mean or geometric mean of both values to finding bias.
Method 2 :Like the ROC and PR curve, we can plot TNR versus FNR graph and find AUC to detect bias in the system.

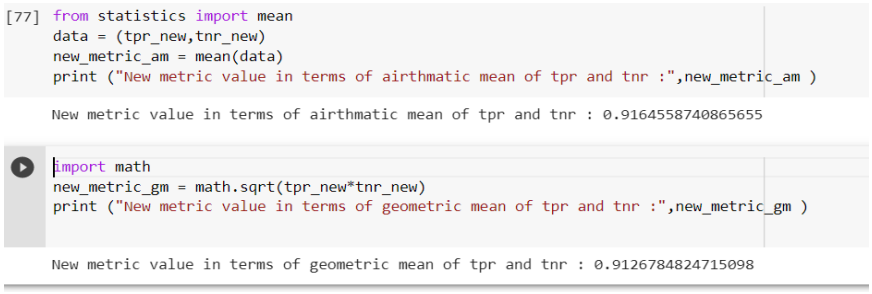

Figure 19: New Evaluation Metric Result

4.3 DATA Method Result

Figure 20: Accuracy and Loss

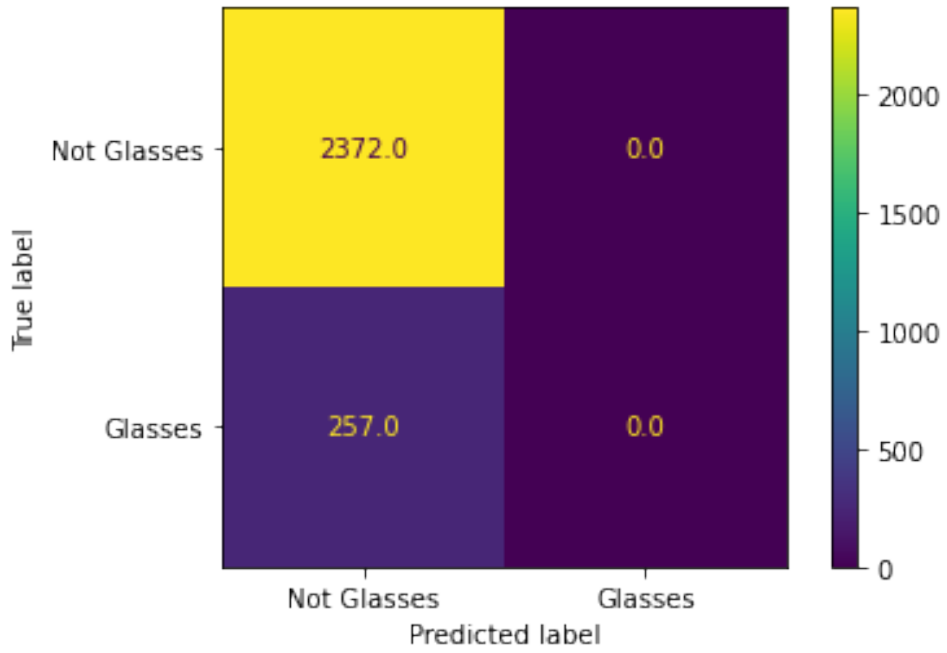

Figure 21: Confusion Matrix

```
[48] disp_impct

     0.0
```

Figure 22: Disparate Impact

```
[50] mse_loss, bias_nn, var_nn

     (0.09775580068467098, 0.09775580068467098, 0.0)
```

Figure 23: Degree of bias

4.4 ALGORITHMIC Method Result

```
print('accuracy and loss on the testing data in algorithmic method', test_acc_data35, test_loss_data35)

accuracy and loss on the testing data in algorithmic method 0.9037656784057617 0.31022512912750244
```

Figure 24: Accuracy and Loss

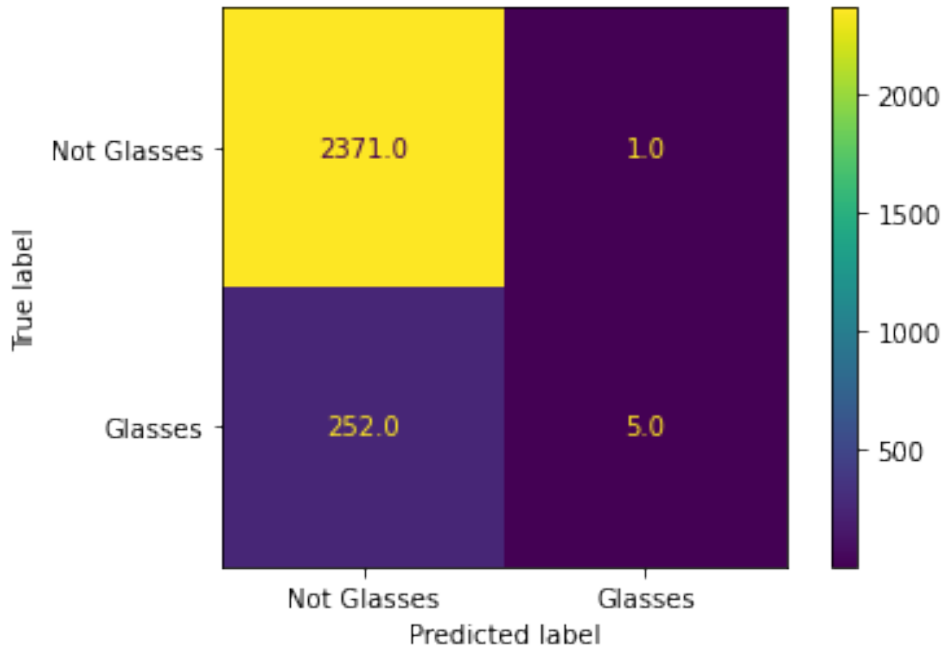

Figure 25: Confusion Matrix

```
[58]  disp_impct
      0.019463458423525513
```

Figure 26: Disparate Impact

```
[60]  mse_loss, bias_nn, var_nn
      (0.08699203324763614, 0.08699203324763614, 0.0)
```

Figure 27: Degree of bias