Troy Krupinski

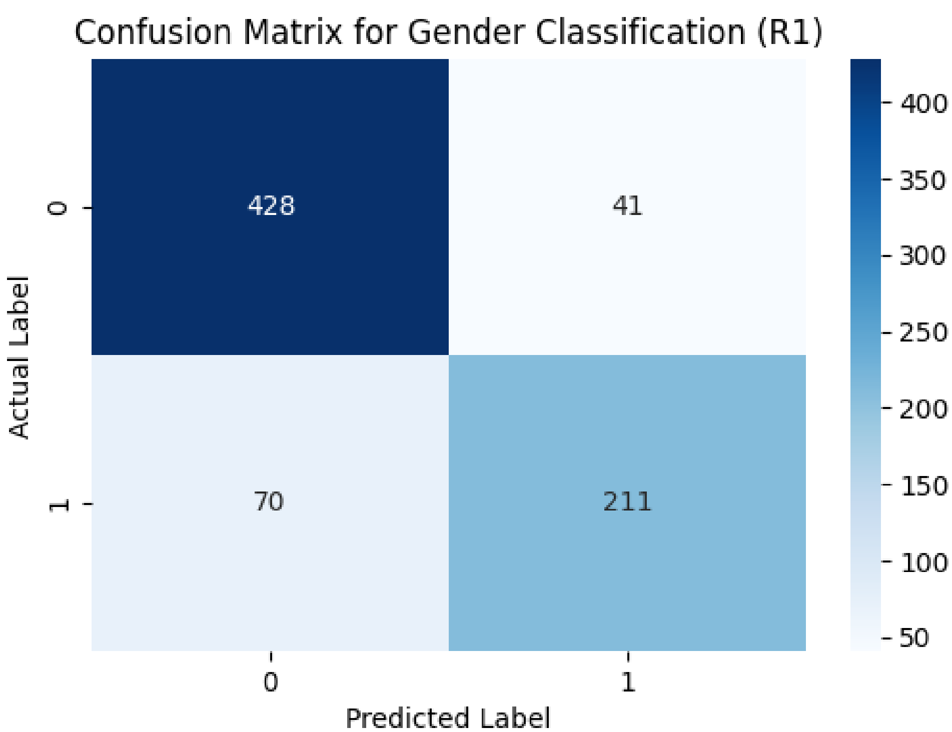
Tsk0064

CSCE 5215 – 10/20/2024

R1

Classify test images using the attribute dataset into classes of individuals by gender into the two classes, Male and Female. This is an example of query answering on the CelebA dataset. Produce the classification accuracy and confusion matrix that is produced by deployed your model on the test partition.

R1 (Gender Classification) Results - Loss: 0.35554325580596924, Accuracy: 0.8519999980926514



R2

The purpose of this requirement is to investigate whether the gender classification model that you produced with R1 generalizes with age. If this is the case, then it should be equally easy to separate males from females, irrespective of their age status, young or old. Classify test images using the attribute dataset into two targets, gender and age. Produce the classification accuracy and confusion matrix that is produced by deployed your model on the test partition. Note that in this case there are 4 combinations of classes.

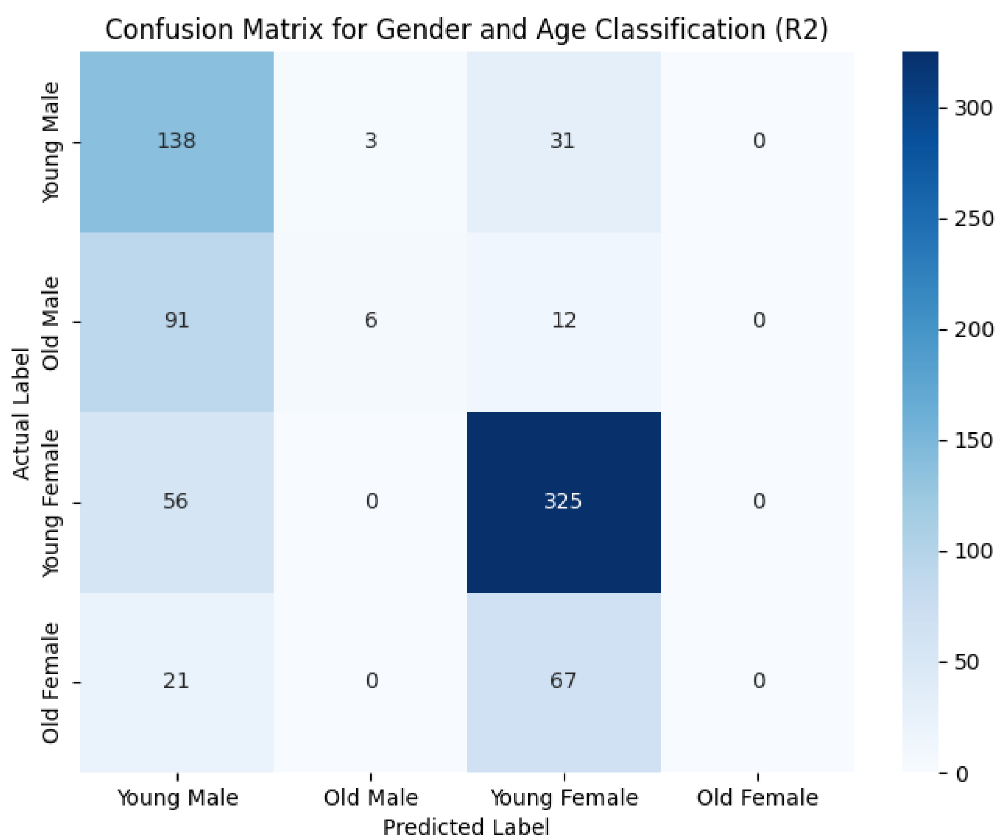
R2 (Gender and Age Classification) Results - Loss: 0.9042612314224243, Overall Accuracy: 0.625333309173584

Accuracy for Young Male: 0.80 (138/172)

Accuracy for Old Male: 0.06 (6/109)

Accuracy for Young Female: 0.85 (325/381)

Accuracy for Old Female: 0.00 (0/88)



R3

In this requirement we will incorporate meta data from the landmarks partition. We will quantify mouth width using 4 quartiles. The mouth width index is obtained from the landmark dataset by computing the horizontal distance between the two mouth positions – this will be given by mouth_width=rightmouth_x - leftmouth_x.

Before balancing Q1 dataset: Smiling

0   2088

1   103

Name: count, dtype: int64

After balancing Q1 dataset: Smiling

0   2088

1   2088

Name: count, dtype: int64

Before balancing Q4 dataset: Smiling

1   1616

0   30

Name: count, dtype: int64

After balancing Q4 dataset: Smiling

0   30

1   30

(a) Compare the classification accuracy of persons who are smiling and whose mouth_width index are in Quartile 1 (Q1) with those in Quartile 4 (Q4). Produce the classification accuracy and confusion matrix that is produced by deployed each of your two models on the test partition.

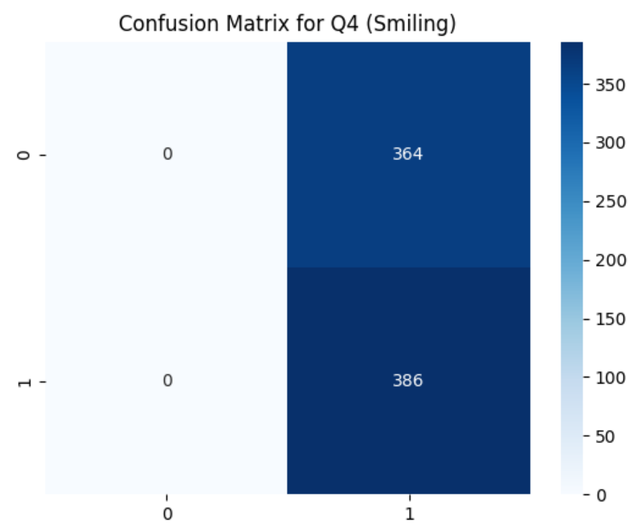
Q1 model (Smiling):

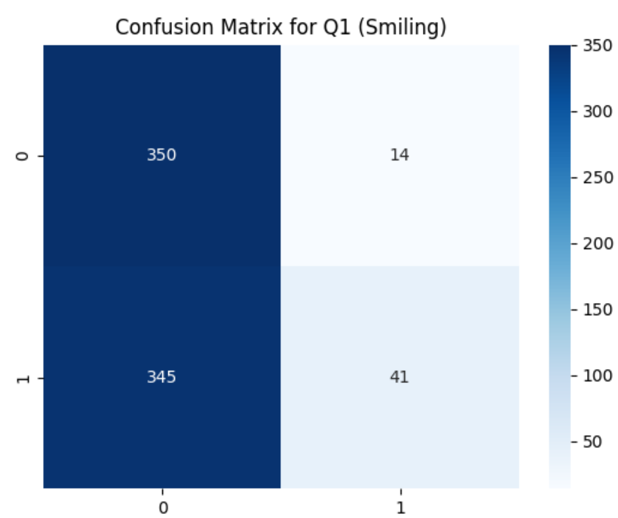
- Precision: 0.7455
- Recall: 0.1062
- F1 Score: 0.1859

Q4 model (Smiling):

- Precision: 0.5147
- Recall: 1.0

- F1 Score: 0.6796





(b) Using the results of (a) above can you conclude whether the mouth_width index is a useful meta data feature for classification of smiling status?

Yes, I believe the mouth width index to be a useful metadata feature for classification of smiling status. The results show that individuals with larger mouth width (in Q4) were more accurately classified in regard to F1 and recall score compared to Q1.

(c) Classify individuals who are female and whose eye_width index are in Quartile 1 (Q1) with those in Quartile 4 (Q4). Produce the classification accuracy and confusion matrix that is produced by deployed each of your two models on the test partition.

Q1 female model (Smiling):

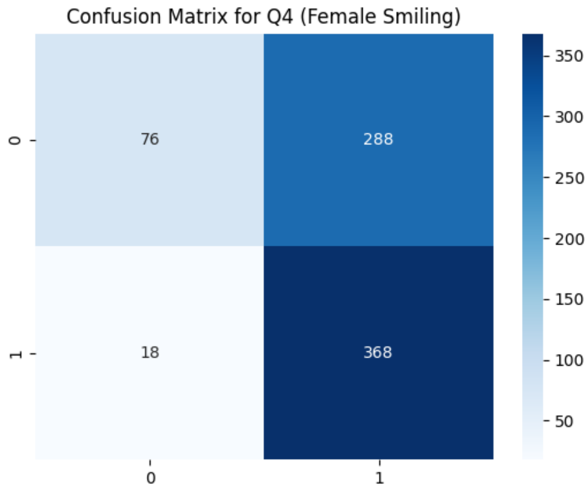
Precision: 0.8155
Recall: 0.4352
F1 Score: 0.5676
Confusion matrix (Q1 Female Smiling):

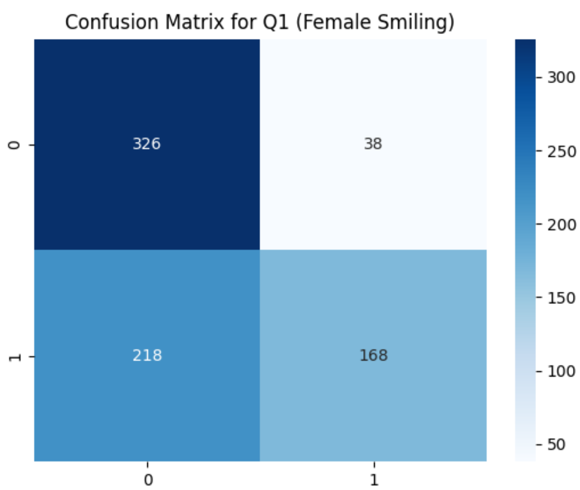
The Q1 female model had moderate precision (81.55%) but lower recall (43.52%), indicating that it missed a substantial portion of the smiling instances.

Q4 female model (Smiling):

Precision: 0.5610
Recall: 0.9534
F1 Score: 0.7063
Confusion matrix (Q4 Female Smiling):

For Q4, the model achieved much higher recall (95.34%) at the expense of lower precision (56.10%), suggesting that it identified most of the smiles but produced more false positives compared to the Q1 model.
Conclusion: Like (a) and (b), Q4 seems to come out with the best results overall, especially regarding recall and F1.





(d) Using the results of (c) above can you conclude whether the eye_width index is a useful metadata feature for classification of female gender status?

Yes, it seems that eye width index is also a somewhat useful metadata feature for the classification of smiling status in females, as the model was more accurate again for individuals in Q4 that have a larger eye width – achieving the higher recall between Q4 & Q1. This suggest that eye width contributes to the model's ability to accurately detect smiles in females.