Introduction to Machine Learning

Assignment 1 report

By:

Omar Atef 7619

Ahmed Moghazy 7397

Nour Mohamed Mahmoud 7591

Model Performance

1- Logistic Regression

For different values of ‘C’ and penalties, the best performance was :

C = 0.01 with the penalty ‘None’. The accuracy was about 77%.

	param_penalty	param_C	mean_test_accuracy	mean_test_precision	mean_test_f1	mean_test_recall
0	None	0.01	0.766704	0.746933	0.782447	0.821753
1	l2	0.01	0.765956	0.741119	0.784055	0.832494
2	None	0.1	0.766704	0.746933	0.782447	0.821753
3	l2	0.1	0.764958	0.744604	0.781175	0.821753
4	None	1	0.766704	0.746933	0.782447	0.821753
5	l2	1	0.766704	0.746907	0.782443	0.821753
6	None	10	0.766704	0.746933	0.782447	0.821753
7	l2	10	0.766455	0.746576	0.782260	0.821753
8	None	50	0.766704	0.746933	0.782447	0.821753
9	l2	50	0.766455	0.746576	0.782260	0.821753
10	None	100	0.766704	0.746933	0.782447	0.821753
11	l2	100	0.766704	0.746933	0.782447	0.821753

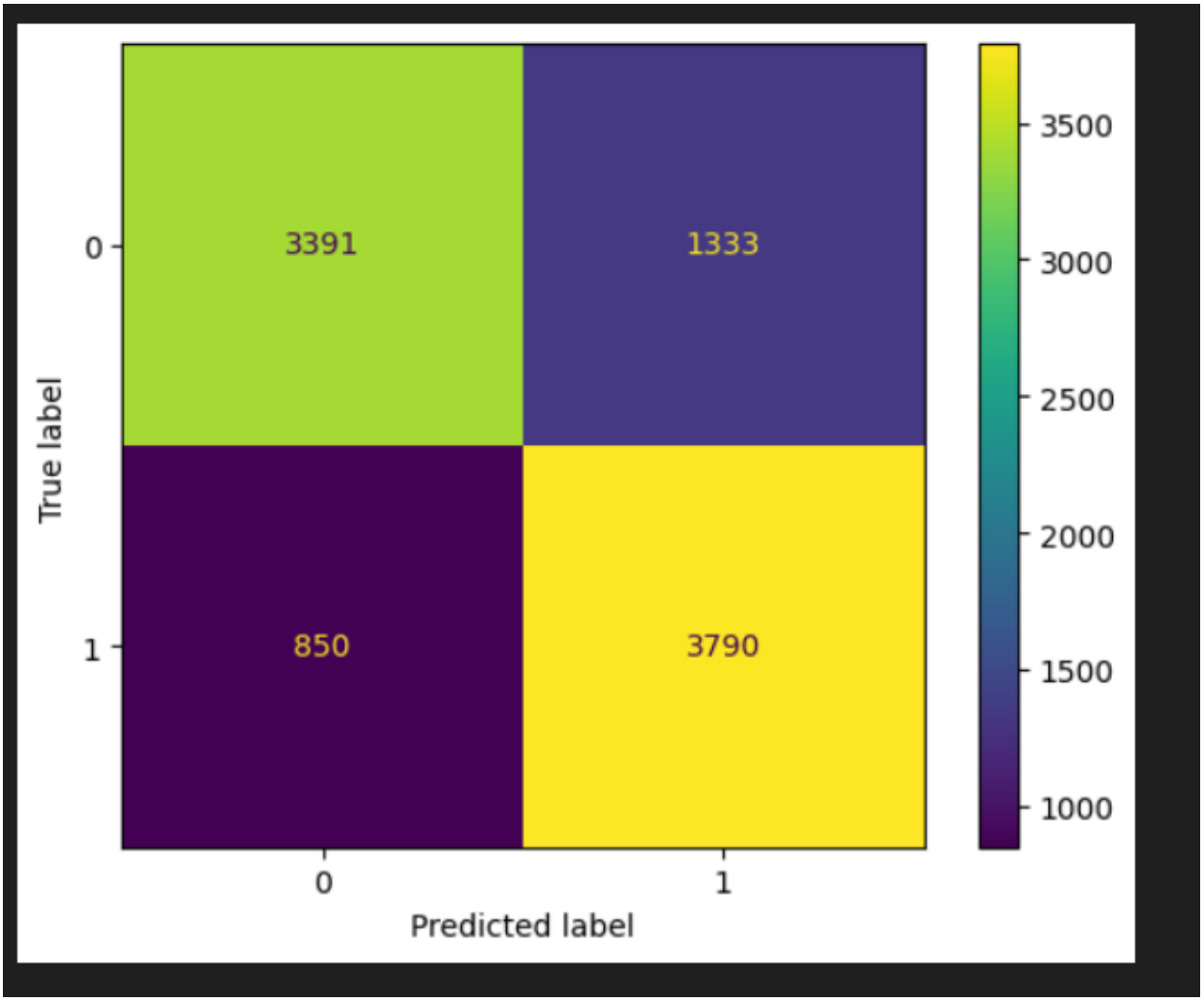
```
lg_cv.best_params_
✓ 0.0s
{'C': 0.01, 'penalty': None}
```

```
print(classification_report(y_eval, lg_prediction))
✓ 0.0s
              precision    recall  f1-score   support

           0       0.80      0.72      0.76      4724
           1       0.74      0.82      0.78      4640

    accuracy                           0.77      9364
   macro avg       0.77      0.77      0.77      9364
weighted avg       0.77      0.77      0.77      9364
```

-Confusion Matrix of Logistic Regression Model

#Note: ‘1’ is gamma signals and ‘0’ is hadrons.

The ‘True Negative’ was about 850 samples.



2- K-NN Model

-For different values of 'K', the best performance was :

K = 7. The accuracy was about 88%.

	param_n_neighbors	mean_test_accuracy	mean_test_precision	mean_test_f1	mean_test_recall
0	3	0.784903	0.756361	0.801831	0.853520
1	5	0.790640	0.757339	0.808949	0.868639
2	7	0.793628	0.757284	0.812618	0.876949
3	9	0.789397	0.751239	0.809685	0.878415
4	11	0.793133	0.752032	0.814149	0.887695
5	13	0.792636	0.750942	0.814118	0.889163
6	15	0.785408	0.743728	0.808024	0.884778
7	17	0.785406	0.742407	0.808611	0.888207
8	19	0.786652	0.746094	0.808588	0.882817

```
knn_cv.best_params_
```

✓ 0.0s

```
{'n_neighbors': 7}
```

```
knn_prediction = knn_cv.best_estimator_.predict(x_eval)

print(classification_report(y_eval, knn_prediction))
```

✓ 1.1s

```
              precision    recall  f1-score   support

           0       0.86      0.72      0.79      4724
           1       0.76      0.88      0.82      4640

    accuracy                           0.80      9364
   macro avg       0.81      0.80      0.80      9364
weighted avg       0.81      0.80      0.80      9364
```

-Confusion Matrix of K-NN Model

#Note: ‘1’ is gamma signals and ‘0’ is hadrons.

The ‘True Negative’ was about 541 samples.