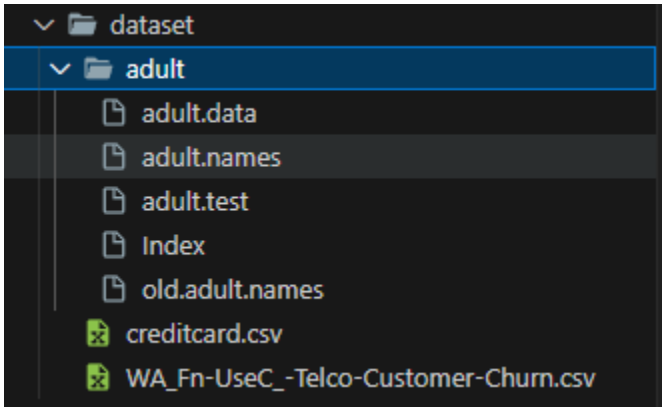
How to Run

- Extract the 1805093.zip
- Create a directory named “dataset” at the root and paste all the dataset there. The directory should look like this -



- Run the 1805093.py file. It will, by default, preprocess, fit, predict and report the Telco Dataset.
- To get other datasets to do the same, please uncomment and comment specific lines found at the very bottom on 1805093.py

```
X_train, X_test, y_train, y_test = preprocess_telco()
# X_train, X_test, y_train, y_test = preprocess_adult()
# X_train, X_test, y_train, y_test = preprocess_creditcard()
```

- Right beneath that, you will find the hyperparameters to tweak

```
# hyperparameters
alpha = 0.01
num_iter = 5000
threshold = 0.0
num_features = 20
num_hyp = 10
```

Performance Measure (Logistic Regression)

Learning Rate, alpha = 0.01
Number of Iteration, num_iter = 5000
Early Stopping Threshold, threshold = 0.0
Number of Features, num_features = 20
Number of Hypotheses, num_hyp = 10

Telco Dataset

Performance Measure	Training	Test
Accuracy	0.7992545260915868	0.794889992902768
Recall	0.5195234943745863	0.46089385474860334
Specificity	0.9017705554208101	0.9086584205518554
Precision	0.6596638655462185	0.632183908045977
False Discovery Rate	0.3403361344537815	0.367816091954023
F1 score	0.5812661977045538	0.5331179321486268

Adult Dataset

Performance Measure	Training	Test
Accuracy	0.7623537360646172	0.766783367115042
Recall	0.01313607958168601	0.012740509620384815
Specificity	1.0	1.0
Precision	1.0	1.0

False Discovery Rate	0.0	0.0
F1 score	0.025931520644511586	0.0251604621309371

Credit Card Dataset

Performance Measure	**Training**	**Test**
Accuracy	0.9760873543585676	0.9756038058062942
Recall	0	0
Specificity	1	1
Precision	0	0
False Discovery Rate	0	0
F1 score	0	0

Performance Measure (Logistic Regression)

Learning Rate, alpha = 0.1
Number of Iteration, num_iter = 1000
Early Stopping Threshold, threshold = 0.1
Number of Features, num_features = 10
Number of Hypotheses, num_hyp = 10

Telco Dataset

Performance Measure	Training	Test
Accuracy	0.6458998935037273	0.6579134137686302
Recall	0.02183984116479153	0.027932960893854747
Specificity	0.8746058695124909	0.8725023786869648
Precision	0.5	0.4166666666666667
False Discovery Rate	0.5	0.5833333333333334
F1 score	0.04185161699429296	0.05235602094240838

Adult Dataset

Performance Measure	Training	Test
Accuracy	0.6389776357827476	0.6486870120652946
Recall	0.01455989410986102	0.0223463687150838
Specificity	0.8678146980354111	0.8620361560418649

Precision	0.4074074074074074	0.38095238095238093
False Discovery Rate	0.5925925925925926	0.6190476190476191
F1 score	0.028115015974440893	0.04221635883905014

Credit Card Dataset

Performance Measure	**Training**	**Test**
Accuracy	0.7036946039740795	0.7062219765370678
Recall	0	0
Specificity	0.9269012944983819	0.9246481704865299
Precision	0	0
False Discovery Rate	0	0
F1 score	0	0