Classification challenge on Alzheimer's Disease

Edwing Yair Ulin
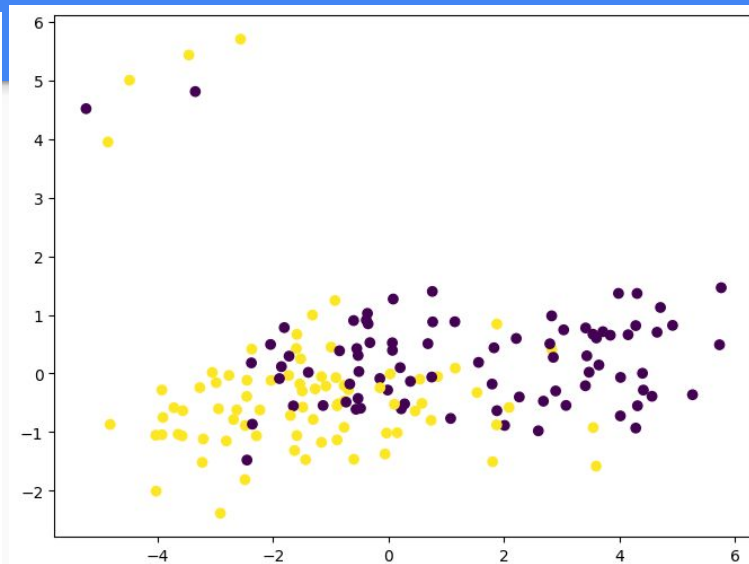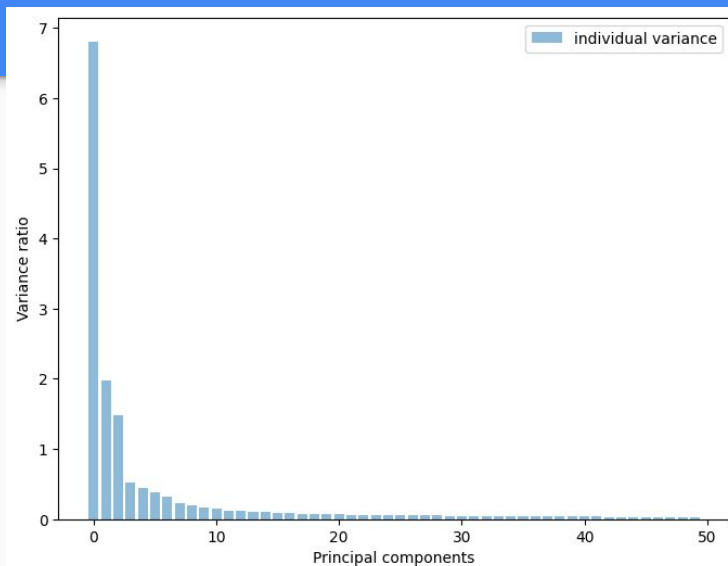
Steps of the code

Data Cleaning
Ingesting the data and training.

1

Scaling and Normalizing

2

Principal Components

3

Reducing the features

4

Cross validation and Metrics

5
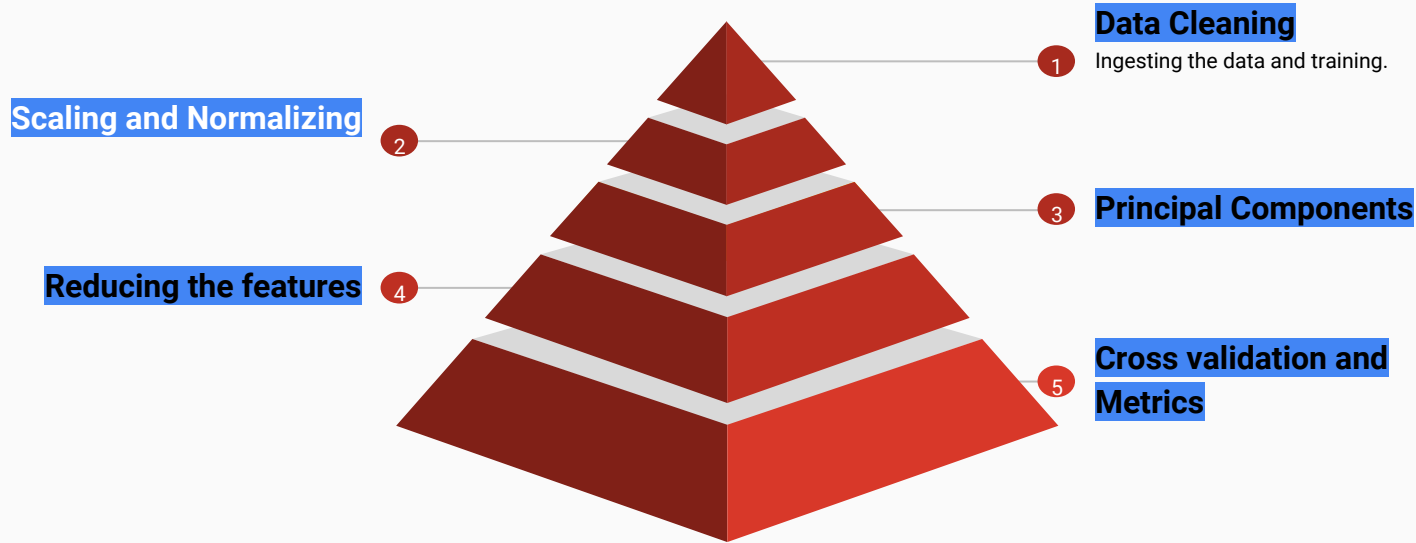
ADCTL Dataset

MaxMinScaler

Plot of 6 features selection

```
['x42' 'x223' 'x276' 'x308' 'x317'
'x358']
```

ADCTL Results

	Accuracy	Precision	F1	AUC	MCC
Logistic Regression	100.000%	100.000%	100.000%	100.000%	100.000%
Support Vector Machines	100.000%	100.000%	100.000%	100.000%	100.000%
Decision Trees	81.818%	85.714%	80.000%	81.618%	63.944%
Random Forest	90.909%	88.235%	90.909%	90.993%	81.985%
Naive Bayes	84.848%	78.947%	85.714%	85.110%	71.008%
K-Nearest Neighbor	93.939%	93.750%	93.750%	93.934%	87.868%

ADMCI Dataset

MaxAbsScaler
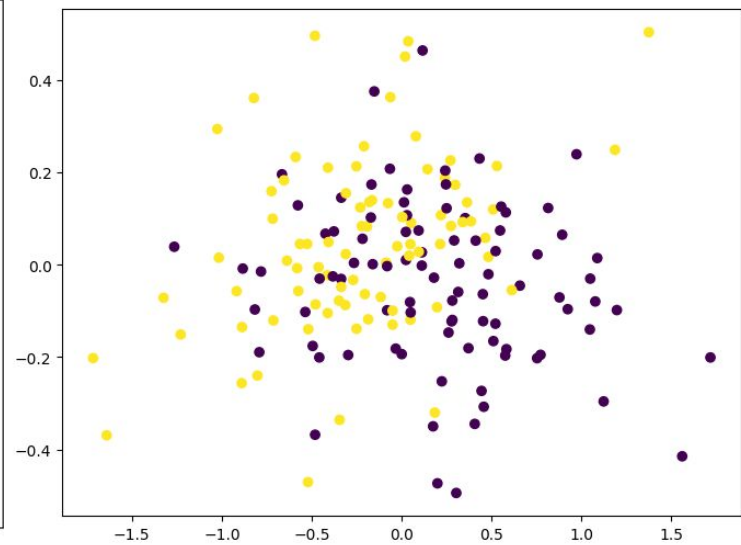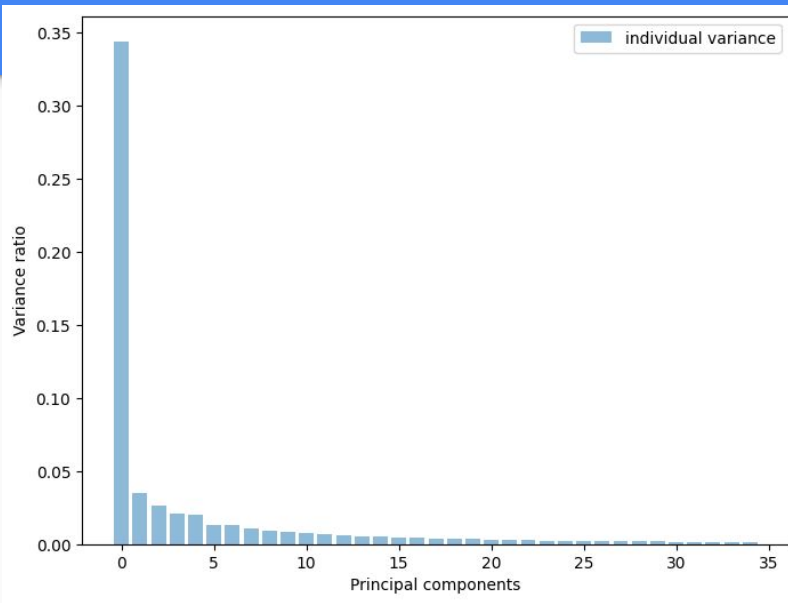

```
['x21' 'x22' 'x34' 'x35' 'x36' 'x37'
 'x38' 'x39' 'x49' 'x50' 'x53' 'x54']
```
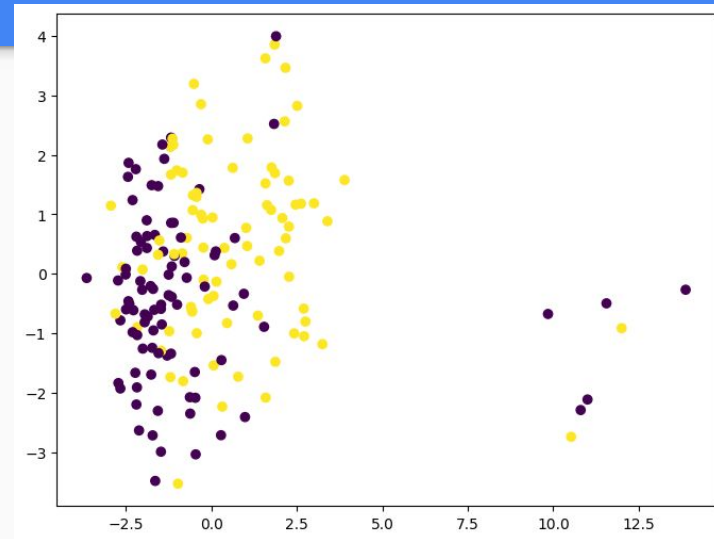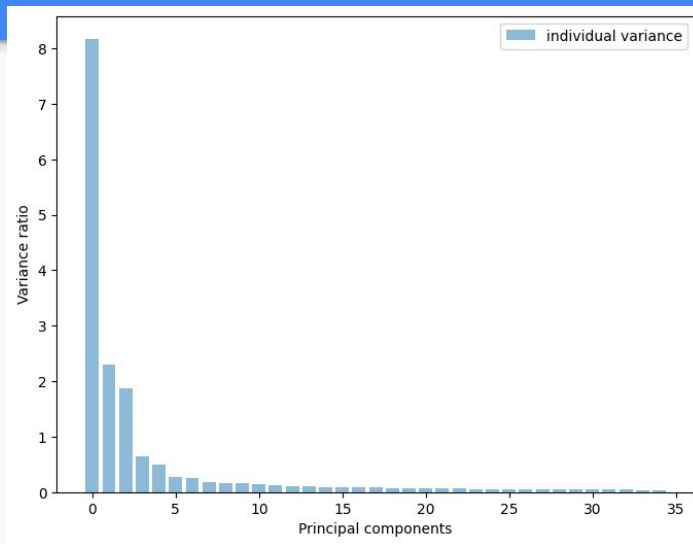
Plot of 12 features selection

ADMCI Results

	Accuracy	Precision	F1	AUC	MCC
Logistic Regression	71.429%	72.222%	72.222%	71.405%	42.810%
Support Vector Machines	71.429%	68.182%	75.000%	71.078%	43.606%
Decision Trees	65.714%	68.750%	64.706%	65.850%	31.803%
Random Forest	80.000%	82.353%	80.000%	80.065%	60.131%
Naive Bayes	74.286%	71.429%	76.923%	74.020%	49.010%
K-Nearest Neighbor	62.857%	69.231%	58.065%	63.235%	27.380%

MCICTL Dataset

MaxMinScaler

`[ 'x335'  'x373'  'x456'  'x458' ]`

Plot of 4 features selection

ADMCI Results

	Accuracy	Precision	F1	AUC	MCC
Logistic Regression	91.429%	84.615%	88.000%	91.486%	81.508%
Support Vector Machines	82.857%	87.500%	70.000%	76.993%	61.025%
Decision Trees	65.714%	50.000%	53.846%	63.949%	27.031%
Random Forest	74.286%	60.000%	66.667%	74.457%	46.916%
Naive Bayes	82.857%	87.500%	70.000%	76.993%	61.025%
K-Nearest Neighbor	77.143%	66.667%	66.667%	74.638%	49.275%