Classification challenge
on
Alzheimer's Disease data

Taiabur RAHMAN

Outline

- Data Preprocessing
- Feature selection
- Data Analysis
- Predict data
- Predicted Scores
- Result

Data Preprocessing

Inspect ***p*** the number of predictors and ***n*** is number of samples each dataset.

- Dataset 1: *n = 164, p = 429* => Very high
- Dataset 2: *n = 172, p = 63* => **Low dimensionality**
- Dataset 3: *n = 172, p = 593* => Very high dimensionality

Data Preprocessing

In dataset to classify patients to 2 classes

- AD: Alzheimer Disease
- CTL: Control

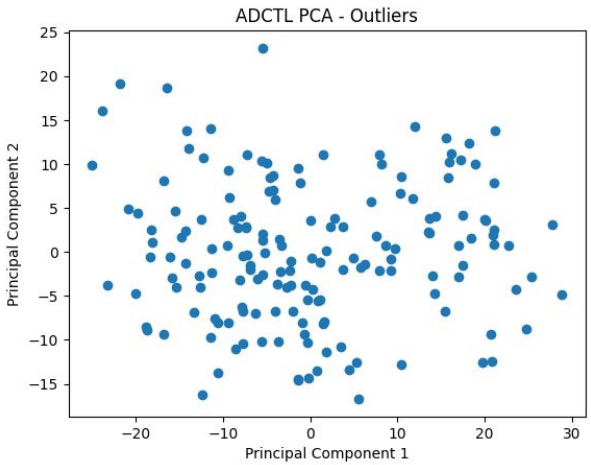
To achieve the result train the following models:

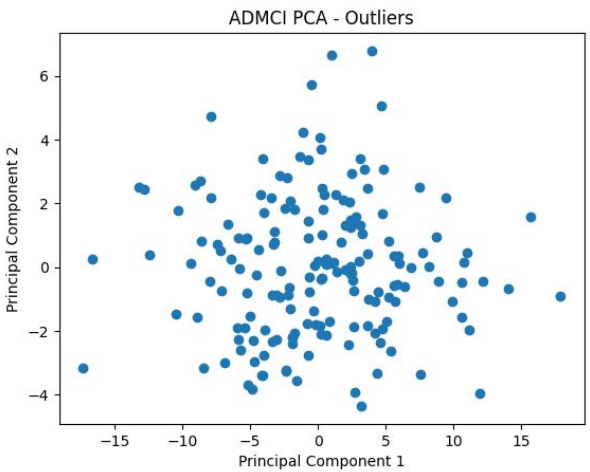
- Logistic Regression
- Support Vector Machine
- Random Forest

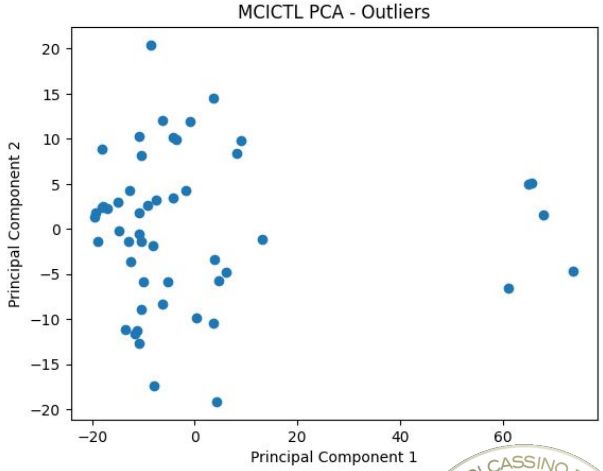
Data Analysis

The first step of the challenge is to analyse the datasets. The result summarised

Principal Component Analysis







Training dataset predicted

Performance on the training datasets

Logistic Regression

	Accuracy	Sensitivity	Precision	F1	AUC	MCC	Balanced Accuracy
ADvsCTL	0.8609	0.8889	0.8276	0.8571	0.8625	0.7236	0.8625
ADvsMCI	0.7438	0.7778	0.6885	0.7304	0.7471	0.4914	0.7471
MCIvsCTL	0.7769	0.7612	0.8226	0.7907	0.7787	0.5544	0.7787

Thank you!