Lesson 8 Assignment

Submit Assignment

Due Oct 14 by 11:59pm **Points** 100 **Submitting** a file upload
Available Oct 1 at 12am - Oct 21 at 11:59pm 21 days

For this assignment, you are going to use the **Breast Cancer Coimbra Data Set (https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Coimbra)** .

Data Set Information:

There are 10 predictors, all quantitative, and a binary dependent variable, indicating the presence or absence of breast cancer. The predictors are anthropometric data and parameters which can be gathered in routine blood analysis. Prediction models based on these predictors, if accurate, can potentially be used as a biomarker of breast cancer.

Attribute Information:

Quantitative	**Attributes**:
Age	(years)
BMI	(kg/m2)
Glucose	(mg/dL)
Insulin	(µU/mL)
HOMA	
Leptin	(ng/mL)
Adiponectin	(µg/mL)
Resistin	(ng/mL)
MCP-1(pg/dL)	(ng/mL)

Labels:

1 = Healthy Controls

2 = Patients

1. Perform Data exploratory analysis on the data (10 points)
2. Use 30% of data as test set and build a Logistic regression model to predict Labels variable (20 points)

3. Build the Naïve Bayes model to predict Labels variable (20 points)
4. Build the Decision tree model to predict Labels variable (20 points)
5. Build Neural network model to predict Labels variable (20 points)
6. Compare their performance, which one is better? (10 points)

When completed, submit your answers to this assignment.