CS 422 Assignment 6 Report

Name(s) of the student(s) completing the assignment:

Ivan Biacan

The dataset you used, its source and characteristics:

Heart Disease

Input Features of the dataset: Age, totChol, sysBP, diaBP, BMI, heartrate, glucose

Output Feature of the dataset: TenYearCHD (10-year risk of coronary heart disease)

A table showing how many iterations it took for training to converge with the three different combinations of model parameters you experimented with.

	Iterations
Model One	88
Model Two	78
Model Three	43

Here are three models with different parameters to compare the performance between the three. Throughout the report, I will refer to them as what is shown in this section

Model One Parameters

- (hidden_layer_sizes = (10, 10, 10, 10, 10, 10, 10), solver = 'adam', max_iter = 100000)

Model Two Parameters

- (hidden_layer_sizes = (20, 20, 20, 20, 20, 20, 20), solver = 'adam', max_iter = 100000)

Model Three Parameters

- (hidden_layer_sizes = (300, 300, 300, 300, 300, 300, 300), solver = 'adam', max_iter = 100000)

A table showing relevant evaluation metrics (accuracy, sensitivity, specificity, f1 score, log loss) for the **training dataset**.

	Model One	Model Two	Model Three
Accuracy	0.8522	0.8546	0.8528
Sensitivity	0.8528	0.8557	0.8529
Specificity	0.7619	0.7568	0.8421
F1 Score	0.9198	0.9029	0.9201
Log Loss	0.3978	0.4003	0.4174

A table showing relevant evaluation metrics (accuracy, sensitivity, specificity, f1 score, log loss) for the **test dataset**.

	Model One	Model Two	Model Three*
Accuracy	0.8443	0.8455	0.8443
Sensitivity	0.8460	0.8478	0.8460
Specificity	0.5000	0.5714	0.5000
F1 Score	0.9154	0.9159	0.9154
Log Loss	0.4141	0.4357	0.4174

** Model three confusion matrix produced the same confusion matrix as model one.*

Comparison of performance metrics for MLP Classifier and Logistic Regression from Assignment 4.

For this comparison, I will use Model Two as the representative for MLP Classifier

- The table below shows the **training data** set used for MLP Regressor and Linear Regression

	Logistic Regression	MLP Classifier (Model Two)
Accuracy	0.8507	0.8546
Sensitivity	0.8525	0.8557
Specificity	0.6429	0.7568
F1 Score	0.9189	0.9029
Log Loss	0.3884	0.4003

- The table below shows the **test data** set used for MLP Regressor and Linear Regression

	Logistic Regression	MLP Classifier (Model Two)
Accuracy	0.8502	0.8455
Sensitivity	0.8522	0.8478
Specificity	0.6667	0.5714
F1 Score	0.9184	0.9159
Log Loss	0.3953	0.4357

Any additional details you would like to include.

I revisited Assignment 4 because the log loss given back then was always around 5 which did not make sense to me at the time. Looking back, I had to predict the probabilities of the training and test input features and apply that in the log_loss function from sklearn. In Assignment 4, I put the predicted binary value array instead of the probability array of it being 0.