Assignment 2: Building Classification Models

Top of Form

|  |  |
| --- | --- |
| Hide Assignment Information | |
| Instructions |  |
| You are provided with a breast cancer dataset ([Breast\_Cancer\_Data.csv](https://nationalu.brightspace.com/d2l/common/dialogs/quickLink/quickLink.d2l?ou=108314&type=coursefile&fileId=CSV+Files%2fbreast-cancer-wisconsin.csv)) taken originally from the UCI data repository.  <https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(original)>.  This dataset has approximately 683 patient data having 10 features and 1 class label describing whether the patient has cancer or not. Each row describes one patient, and the class column describes if the patient tumor is benign (label = 2) or malignant (label = 4). For this dataset, build all the classification models (using Python and Scikit-learn) given below (no need to visualize) and tabulate the accuracy and confusion matrix obtained for each. Split the dataset such that the test data size is 25% of the total dataset.  Make sure to code each classification model in a separate python file. Then, you can tabulate the accuracy and confusion matrix in a Word document table. Finally, submit all the python files and Word documents.              a. Logistic Regression              b. KNN (k = 5)              c. Linear SVM (kernel = linear)              d. Kernel SVM (kernel = rbf)              e. Naïve Bayes              f. Decision Tree              g. Random Forest (estimators = 10)              f. XGBoost | |
|  |  |
| Due on May 12, 2025 1:59 AM | |

Submit Assignment

Bottom of Form