SIG720 Task 2P

Solve the following set of problems using Python and submit the code file with extension .ipynb.

Part A: Logistic Regression (Bank Dataset)

1. Load the "bank-full.csv" dataset and create a logistic regression model to predict term deposit subscription.
2. Evaluate the model using accuracy, precision, recall, and F1-score.
3. Create two regularized logistic regression models for predicting the target variable using the evaluation setting that you have used in 1. and report the performance.
4. Use KNN as a baseline model and compare its performance with logistic regression. Consider the following aspects: number of trainable parameters, training time and model performance. Explain why KNN is worse/better than logistic regression.

Part B: SVM Classification (Grid Stability Dataset)

5. Download and load the Electrical Grid Stability Simulated Data dataset and print the dimension of the dataset.
6. Classify the "Electrical Grid Stability Simulated Data" (target=stabf) available in the dataset using SVM with three different kernels. Select appropriate data splitting approach and performance metrics. Report the performances and the used model hyper-parameters.
7. Tune the "C" parameter for each kernel and report the best performance for each.
8. Compare and discuss the performance of different SVM kernels.