**Assignment 3**

**CS430-01**

**Machine Learning on Cloud**

**Fall 2022**

**Exploratory Data Analysis, train the data using classification techniques (100 points)**

**Goal:** The goal of this assignment is to use Pandas/Matplotlib/Seaborn to explore the dataset, use Sciket-learn libraries do the data preprocessing, split the data and train models with Logistic Regression and SVM classification.

**Instructions:** For this assignment, you work on a Jupyter Notebook. First, create a new notebook titled **Assignement3\_XXX**, where **XXX** are your initials. Also create a GitHub repository titled **Assignement3\_XXX** to which you can push your code. Then complete the following:

1. In this assignment, you will need to work on the **customer\_data** from Credit Risk Classification Dataset (https://www.kaggle.com/datasets/praveengovi/credit-risk-classification-dataset). The dependent variable of the dataset is label-Target Column 0-Low Credit Risk 1-High Credit Risk. Do necessary data preprocessing.
2. Find the best Score of the classifier and print them in a pretty table:
3. **Stratified K Fold Cross validation technique with SVC classifier**
4. **Repeated Random Subsampling technique with SVC classifier**
5. **Randomized search with Logistic Regressor**
6. **Grid search with SVC classifier**

Make sure select the relevant parameters. Some useful documentations are:

1. <https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html>
2. <https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html>
3. <https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.StratifiedKFold.html>
4. <https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.ShuffleSplit.html>
5. <https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html>
6. <https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.RandomizedSearchCV.html>
7. Ensure your notebook is organized and has proper **Markdown comments,** etc. You can assume that after someone see the raw notebook, so it should be clear.

**Assignment Submission:** Upload a link to your GitHub repository for the project in the area provided in Moodle by the deadline specified.