**Assessments for Level 2A AI-ML programs**

**1.** Consider you are given a dataset which consists of details of titanic ship travelers and who survived and who did not on that epic journey. The problem is, given a person who is a traveler on that titanic journey as input, predict if he survived. Create models using all the decision tree techniques such as CART, C4.5 and ID3 and find the best of all your models. Use python to create the models and cross-validate the dataset. Please make sure you also use appropriate performance measures, error values, bias and variance to come to conclusion of the best dataset

Dataset: titanic dataset from Kaggle (create a login to get the dataset)

Model: Classification using decision trees

Algorithm: ID3, C4.5 and KART

Language: python

**2.** You are given a dataset of fake and clean currency notes. The problem is to find if the given currency note is fake or clean. Prepare the data and create models using all the decision tree techniques such as CART, C4.5 and ID3 and find the best of all your models. Use python to create the model and cross-validate the dataset. Please make sure you also use appropriate performance measures, error values, bias and variance to come to conclusion of the best dataset

Dataset: bank note authentication – dataset attached



Model: Classification using Decision Trees

Algorithm: CART, C4.5 and ID3

Language: Python

**3.** You are given a dataset which consists of the medical records of Pima Indian women. The problem is to find if a new person whose medical data is given as input has a chance of getting diabetes in the next 5 years. This is indicated by 1 or 0 in the last column of the dataset. Please use classification models such as KNN and Naïve Bayes and find the best of the two models for the given data set. Use python to prepare the data and create the models and cross-validate the dataset. Please make sure you also use appropriate performance measures, error values, bias and variance to come to conclusion of the best dataset

Dataset: Pima Indians – data set attached



Model: Classification

Algorithm: KNN and Naïve Bayes

Language: Python

**4.** You are given a dataset which consists of the petal and sepal measurements of iris flowers. The problem is to find the category of flower (Iris Setosa, Iris Virginica, Iris Versicolor) given the measurements of a particular flower. Create models using KNN and Naïve Bayes and find the best of the two models used. Use python to prepare the data and create the models and cross-validate the dataset. Please make sure you also use appropriate performance measures, error values, bias and variance to come to conclusion of the best dataset

Dataset: Iris Flowers – dataset attached



Model: Classification

Algorithm: KNN and Naïve Bayes

Language: Python