**CS 6301 - Machine Learning Lab - Week 12**

Date: 20.10.2023

**TITLE**

**IMPLEMENTATION OF K- NEAREST NEIGHBOUR ALGORITHM**

**TASK**

(1) Write Python code that performs supervised classification using the k-NN algorithm which assigns the majority label of the k-closest stored label instances into a new input instance. Use the dataset (data.csv) to classify the new record (#10) using k=9:

I. Use min-max standardization formula and calculate the min-max standardized values for Age and Income attributes.

II. Calculate the distance for the new record (#10) from each of the nine records using standardized values.

III. Using unweighted voting, classify the risk factor for the new record.

**IV.** Record your observations using different standardization methods and distance measures**.**

(2) Data Set Description: Data Filename: data4_19.csv The data set contains 150 data points, there are three classes where each class refers to a type of iris plant. The first four columns represent the attributes listed below. Note that only the first four columns should be used as attributes. The last column is the ground truth class name. 1. sepal length in cm 2. sepal width in cm 3. petal length in cm 4. petal width in cm 5. Ground truth class name: -- Iris Setosa -- Iris Versicolour -- Iris Virginica

Write Python code to find the species of the new flower given Sepal length= 5.2, Sepal width =3.1, Petal length= 1.4, Petal width =0.2 using the k-nearest neighbour algorithm.

I. Consider using different distance measures and tabulate the results.

II. What is the K-value used? How do you find the best k value? Justify your answer.