CSE643: Artificial Intelligence

Assignment-4

Due date: 21-Nov-2020 11:59PM **Marks: 7**

1. The dataset (attached as 'roo-data.csv') provides anonymized information for graduates along with their respective scores in various CS subjects. Specifically, the suggested job role is predicted for every graduate based on the courses they have done.
2. The task is to make a ML-based prediction system to predict the job role for a new graduate.
3. Specifically, you will have to do the following:
    a. Study the given data thoroughly and then modify it/prepare it considering the end goal at hand – suggesting a suitable role.
    b. Divide it into training and testing sets.
    c. Using the sklearn library, train a ML model (ANN) using the training data.
    d. Predict the outcome for the testing data.
    e. Using the predicted labels and actual labels, find out accuracy, confusion matrix and class-wise accuracies.
    f. Also, you need to experiment with data preparation - analyze what modifications produce better results and why. Try out different ratios of train-test splits (60-40, 70-30, 90-10 etc.), shuffling the data etc.
    g. You do not have to connect this data or the prediction system with your electives advisory system that you have already built in assignment #1. This can be independent of that. The aim is to get you to build a ML model for prediction of job role on training data. However, if you want to you can rename some of the courses with the electives that you have considered, but use the given data for those attributes.
    h. You are free to club some job roles like "Business Intelligence Analyst" and "Business Systems Analyst" into "Analyst", etc. and reduce it to 5 to 7 class classification task. Similarly, for the values of various attributes you can bucketize the values into various ranges or categorize them such as "high", "medium", "low".
4. You are required to submit in a zip file consisting of your code (.py file/ .ipynb notebook) along with a report. The report (in pdf format) should contain the steps you did for data preparation/modification, the experiments you performed, results, analysis on the

obtained results, and the listing of the program. Name the ZIP file as: AI-A4-<Name>-<RollNo>. Demonstrate the program to your assigned TAs.