COSC 311 Project 1 (10 points)

Classification Based Wireless Indoor Localization

Please finish the following tasks and submit your project report via MyClasses. Your submission must contain your source code file (one ".py" file for all code) and a PDF document. For each task, please include your explanation and the output/results of your program (may use screenshot).

Dataset:
Please use the "Wireless Indoor Localization Data Set". The description of this dataset can be found at https://archive.ics.uci.edu/ml/datasets/Wireless+Indoor+Localization. You may download the dataset from this website or from the project assignment on MyClasses.

Task 1: Self-test for KNN and DT algorithms (3 points)

- Conduct experiments to show the self-test results (using classification report) for KNN and Decision Tree (DT), respectively.
- Explain how you set parameters for the KNN model. If possible, you are suggested to use data analysis results to support your explanation and conclusion.
- Explain how you set parameters for the DT model. If possible, you are suggested to use data analysis results to support your explanation and conclusion.

Task 2: Independent-test for KNN and DT algorithms (3 points)

- Assume we use 30% data for independent test, conduct experiments to show the independent-test results (using classification report) for KNN and DT, respectively.
- Do you use the same KNN model as the self-test above?

    a) If yes, please exploit whether you could optimize its parameters to get better performance. What are the results (using classification report) of the optimized model?
    b) If no, please explain the difference between the self-test model and the independent-test model and why you make this change.

- Do you use the same DT model as the self-test above?

a) If yes, please exploit whether you could optimize its parameters to get better performance. What are the results (using classification report) of the optimized model?
b) If no, please explain the difference between the self-test model and the independent-test model and why you make this change.

Task 3: Classification model finalization (4 points)

- Based on the above analysis, please further optimize corresponding parameters and select a model (KNN or DT) for this application scenario that can obtain the highest performance.
- Please show the independent-test results (using classification report) of the above model when we use 30% data for test. Draw a figure to show the confusion matrix.
- Conduct independent tests using 10%, 20%, 30%, 40%, and 50% data, respectively, for test. Draw a line or bar figure to show the corresponding accuracies of these independent tests.

Policy

1. Each student MUST finish this project independently. NO TEAM WORK and DISCUSSION are allowed. If you need any help, please feel free to contact the instructor.
2. You need to write your whole program in an editor and save your source code as a ".py" file, which will be submitted to MyClasses together with your PDF report.