Week 9: Independent Project

Agenda

1. **Instructions**
    - Learning Outcomes
    - Deliverables
2. **Assessment**
3. **Submission & Evaluation**

Instructions

During this week's Independent project, you will get to test the skills that you learned this week. More specifically, you will get the test your understanding of the following learning outcomes.

Overall Learning Outcomes

- I can understand and apply supervised learning algorithms such as regression, decision trees, KNN, SVM, naive Bayes, random forests to solving business problems.
- I can understand the benefits, limitations, and requirements of various supervised learning algorithms.

Deliverables

The deliverables for this week's Independent project include:

- A Github repository containing your 2 notebook(s).

Assessment

This week's project requires us to implement a K-nearest neighbor (kNN) classifier and a Naive Bayes classifier. Once we conduct the experiments, we will calculate the resulting metrics:

Experimental Procedure:

1. Download the two datasets from the given links:
    - Dataset 1 Source: [Train Dataset Source: **Link (https://archive.org/download/train5_202002/train%20%285%29.csv)** , Test Dataset Source: **Link (https://archive.org/download/test1_202002/test%20%281%29.csv)**]
    - Dataset 2 Source: [**Link (https://archive.ics.uci.edu/ml/datasets/Spambase)**]
2. Randomly partition each dataset into two parts i.e 80 - 20 sets.
3. For dataset 1, because we don't have the label for the test set, we will use the train set to create train and test data (i.e. splitting further), then perform K-nearest neighbor

classification.

4. For dataset 2, perform classification of the testing set samples using the Naive Bayes Classifier.
5. Compute the accuracy (percentage of correct classification).
6. Report the confusion matrix of each classifier.
7. Repeat step 2 to step 4 twice, each time splitting the datasets differently i.e. 70-30, 60-40, then note the outcomes of your modeling.
8. Suggest and apply at least one of the optimization techniques that you learned earlier this week.
9. Provide further recommendations to improve both classifiers.

Create a notebook for each project.

Submission & Evaluation

The submission to this week's Independent Project should be made here [**Link (https://moringaschool.instructure.com/courses/624/assignments/9924)**]. This submission will be a link to your Github repository.

Do note:

- Late submissions will not be assessed.
- Do not seek to copy someone else's work while working on this Independent project. You deny yourself an opportunity to learn whenever you resolve to plagiarism.

"The most exciting thing is the unknown, crossing all the times the borders that normal people avoid, and sailing permanently in uncharted waters." - Unknown