Strathmore University

School of Computing and Engineering Sciences

Bachelor of Business Information Technology

CAT 2

BBT 3201: Artificial Intelligence

Date: 6th November 2024 **Due Date:** 22nd November 2024

Classification is the machine learning task of predicting a new discrete value given a set of training instances. For classification, eager learners are those that create the model first before predicting, offering a greater chance of generalisation compared to lazy learners.

In your assigned groups for the course, respond to these questions in a Colab Notebook and submit on e-learning before the due date. Each question/response should appear on its own cell.

Instructions:

1. **As assigned in your groups,** download the dataset assigned from these links:
    a. Stars Dataset
    b. Pokemon Dataset
    c. Mobile Price Range Dataset
    d. Employee Attrition Dataset

2. After appropriately cleaning the data, carry out Exploratory Data Analysis processes to retrieve at least 4 insights/observations from the dataset.

3. Create a model based on the k-Nearest Neighbours algorithm for classification based on at least 2 suitable features for one target variable. Verify the optimal k for classification.

4. Create 2 other eager learning classifiers as assigned based on at least 2 suitable features for one target variable. For each classifier, create another model with tuned hyperparameters that enhance the accuracy of classification.

5. Evaluate the performances of the k-NN based model, models based on the 2 eager learners and the two tuned eager learning models. Use the confusion matrix metric to evaluate the performances.

Once done, you will **present your work to the instructor** by scheduling a time through this link. After presentation, submit the work on the appropriate link on e-learning before the due date. **All members are expected to be present during presentation.** A score for teamwork will be assessed during the presentation. Ensure you present before the due date or your group risks losing marks for the entire CAT. Presentation will count for 1 hour attendance for the unit.

Feel free to use any available resources for your work, however the rules against plagiarism apply. **Do not directly use ChatGPT or any other AI-generative text applications to develop your work**, this will be penalised strictly when detected.