FIU | Knight Foundation School of Computing and Information Sciences

# Fall 2025: CAP5610 – HW 4

In HW4, you will learn and apply two supervised machine learning algorithms – KNN and SVM, and one unsupervised machine learning, Kmeans.

**Data for Classifiers:**

- Gene Expression: lncRNA_5_Cancers.csv (available under Module 2)
- Cancer Types: KIRC, LUAD, LUSC, PRAD, THCA

**Task 1**: [30 points] Apply KNN algorithm to classify the given data into 5 classes. Use 5-fold stratified cross-validation. (a) Report on the performance using 4 metrics, including Accuracy, Precision, Recall, and F1 score. (b) Show the overall confusion matrix. (c) Show the overall ROC-AUC for 5 classes. (d) Show the overall PR-AUC for 5 classes.

**Task 2**: [30 points] Apply SVM algorithm to classify the given data into 5 classes using three different kernels (Linear, Polynomial, and RBF). Use 5-fold stratified cross-validation. (a) Compare the performance of three kernels using 4 metrics, including Accuracy, Precision, Recall, and F1 score. (b) Show the overall confusion matrix derived from each kernel. (c) Show the overall ROC-AUC for 5 classes with three kernels. (d) Show the overall PR-AUC for 5 classes with three kernels.

**Task 3**: [40 points] Cluster the given dataset applying K-means algorithm using K = 2, 3, 4, 5, 6, & 7. (a) Visualize the clusters for each K. To visualize, use 5 different colors to represent 5 types of cancer types. (b) Find the number optimal clusters for the dataset using visual inspection (open eye inspection). (c) Find the number optimal clusters for the dataset using evaluation metrics: (i) Elbow method and (ii) Silhouette Score.

You must submit the following items in CANVAS:

- Report (MS word or PDF)
  - Describe the algorithms/approaches/tools used: (a) What it is or What it does, (b) How it does, and (c) Application.
  - Describe results: (a) Put Figure/Table number and Title: On top of the table, and bottom of the figure. (b) Describe the figure and table. (c) Your observation about the figure and table. (d) Conclusion.
- Source code (*.py or Jupyter notebook)
  - Must be well organized (comments, indentation, …)
- File name: HW3_lastName

You must submit the files **SEPERATELY**. DO NOT compress into a ZIP file. If you fail to provide all required information or files, you may be given zero score without grading.

**Deadline:**

The deadline is **11:59pm Wednesday, October 29, 2025**. Late assignments will not be accepted.