

Louden Demers  
CSCD 429-040  
02/21/2023

## **Homework 2 Report**

### **How to Run:**

- 1) `cd` to the **CSCD429HW2\_KNN** directory
- 2) Compile with **`javac *.java`**
- 3) Run with **`java Classification`**
- 4) A file called **`output.txt`** should appear

### **Methods Used (Preprocessing):**

- 1) Handling missing data was not needed due to my usage of KNN, which auto-handles the cases of MANY missing values.
- 2) For interaction data, I stored every interaction as its own unique object and then attempted to assign interaction weights in my *weighted difference* approach to the Gene objects within the training set.

### **Method Used (Classification):**

- 1) I used KNN as my classifier, with a  $k$ -value of 5. This is implemented by first calculating the weighted differences based on attributes *Complex*, *Class*, and *Motif* and the *Interaction* objects (which follow the exact form as the .data file).
  - a) To store the weighted differences, I opted for a hashmap since I really only needed the GeneID  $X$  from the test set and the respective GeneIDs  $Y$  from the training set to store their individual weights.
- 2) After calculating the weights, I then sorted the hash map containing them and pulled the Localization values of the  $k$  (5) highest weighted GeneIDs from the training set and stored them as the test set's localization value.
- 3) Then I just wrote out the GeneID-Localization pairs to a .txt file.

**Accuracy:** I unfortunately ran out of time to check this, but it *appears* to be fairly close.