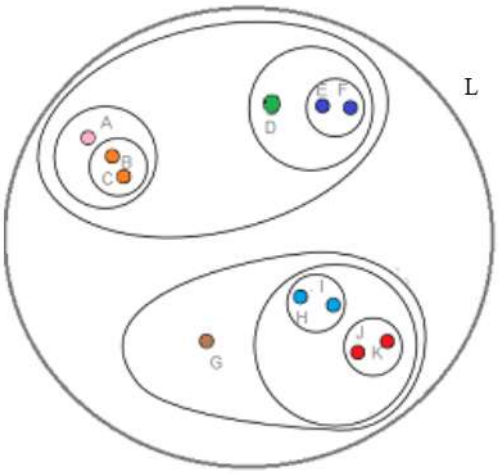
Arkansas Tech University
Graduate College, Information Technology
INFT 6903 Data Science, Summer 2022
Assignment 7 / Due date: July 19th, 2022 before midnight. Total points: 100. Good luck!

Important Note: You can use any kind of Python compiler for this assignment. Also note that, some of online compilers **do not** support/provide some data science libraries like Numpy and Pandas. We will continue using *Anaconda* and *Jupyter notebook* in data analysis and machine learning algorithm development (clustering and classification). Therefore, I recommend using *Anaconda* and *Jupyter notebook* in this assignment. Please also find the *Instructions to Download Anaconda and Jupyter Notebook* pdf file from Blackboard under the Week5 folder.

Anaconda and Jupyter notebook: https://www.anaconda.com/
Google Colab: https://research.google.com/colaboratory/
Onlinegdb: https://www.onlinegdb.com/online_python_compiler
Replit: https://replit.com/languages/python3
Downloading&Installing a Python editor from the Python website: https://www.python.org/

---

1. (*10 points*) Draw the dendrogram of the following hierarchical clustering example manually.



2. (*15 points*) Draw the dendrogram for the following data points in Python.

A = [[4,2], [10,13], [16,12], [20,13], [27,27], [92,78], [70,76], [57,74], [75,61], [80,90]]

3. (*15 points*) Apply DBSCAN clustering algorithm to random data set to develop anomaly (outlier) detection system in Python.

4. (*15 points*) Apply NMF clustering algorithm to the following data set (array) in Python.

Find W and H components.

Data = [[3, 7], [1, 2], [8, 7], [1, 1.9], [2, 6.8], [9, 10]]

5. (*16 points*) Please find the *Several internet searches.pdf* file under the Week7 folder/Miscellaneous section.

Write the answers of the following questions in a few sentences (one or two).

a) Define image pixels in your own words.

b) How can we use bag of words method for text mining?

c) What is the relation between the graph and matrix?

d) In handwritten recognition, what kind of the data set do we need to train the computer with data science/machine learning algorithms?

e) Where can we use human action recognition?

f) Is it possible to use data science in cybersecurity for instance intrusion detection?

g) Is it possible to use anomaly detection for a sensor data? These are some examples of the sensors: Temperature, pressure, distance, humidity, vision, motion, chemical, electromagnetic, acoustic, etc.

h) How can we use data science/machine learning for drug discovery?

6. (*14 points*) Please find the *How to read a paper.pdf* file under the Week7 folder/Miscellaneous section. Write your opinion about this paper in a few sentences (one or two).

7. (*15 points*) Please find the *Sample scientific paper.pdf* file under the Week7 folder/Miscellaneous section. Explain the Figure 4 with one or two sentences.