**DSC540\_0500\_Topic7\_KMeansCluster Analysis**

**Explanation**

For the K-means clustering analysis, the Iris is picked. Using Python, the dataset is then visualized to cluster assignments from K = 2 to K = 5 clusters. This is to test what value of K fits the best.

K=3 is the best K because of the elbow method that I used, and I also visualized it as well to compare it to the other K-mans.

I first started by importing the necessary modules:

Text

Description automatically generated

The data is then printed just to know what kind of data that I’m applying the K-means clustering for. After picking the data, I assigned all the columns except the last one which is the species as x:

Table

Description automatically generated

The iris selection has been assigned to a variable x so that we can use it to fit the data using Kmeans.

I then try to predict with different Ks so that we can determine which K is the best. We also get the predictions from them. The detailed graph can see in the code execution result.

However, we can just use the elbow method to determine the best **K**. Where there is an elbow, it is the best **K**.

Graphical user interface, text, application

Description automatically generated

Chart, line chart

Description automatically generated

As we can see, the elbow occurs when **K=3**

Since we know that the best **K** is 3, we can now go back to fitting the model and predicting them. First thing we need to do is to convert the **species** column into numbers (0,1,2) depending on the species. We make another data frame.

Graphical user interface, text, application, email

Description automatically generated

Now continue to use Kmeans. After that, we will evaluate how well our Kmeans model is doing when **K=3**:

Table

Description automatically generated

From the above result, we can see that the accuracy is at **89%** which is a pretty good accuracy when predicting something, especially a classification problem.

**Ethical aspects of Data analysis and Privacy**

Privacy is a state of being free from others interference and intrusion or one can say it is a right to let alone. It is all about protecting the private or sensitive information from getting exposed. With the digital transformation happening all around and all the processes becoming digitized, now more importance is being given to the data privacy. Since because of digitization more information being is communicated online, it has become very much important to protect the data privacy. All the data being transferred online needs to be managed on its precited significance.

The world that we are living in today has no digital privacy at all. Social Media platforms own our personal chats and other sensitive information which is very dangerous for us. However, they claim that they are encrypted, and they cannot read it but it's not trustworthy at all. It will be fair to think at once that they have access to everything about our account. They use our data to provide us recommendation or the products and services, but this is what they claim. On daily basis we hear news such as xyz company sold abc amount of data to dark web, and it is actually true as our personal data is worth of billion dollar and company do sell them in order to gain profit but obviously they deny that. Now on religious point of view, there are many things that religions tell to follow and to avoid. Now in this digital world there is nothing like confidentiality of the data. Everything can be accessed by these companies who are running digital business. To follow the religious ethics, one must avoid using digital media such as social media platform and other platform completely which is also very difficult to do.