**B.TECH. (2020-24)**

**Artificial Intelligence**

**Open Ended**

**LAB File**

on

**Fundamentals of Machine Learning**

**[CSE313]**

**Logo

Description automatically generated**

Submitted To

**Dr Monika Arora**

Submitted By

|  |  |  |
| --- | --- | --- |
| **Hitesh** | **A023119820027** | **5AI 1** |
| **Gauri Tyagi** | **A023119820028** | **5AI 1** |
| **Kushagra Dubey** | **A023119820029** | **5AI 1** |
| **BHOOMIKA SHARMA** | **A023119820030** | **5AI 1** |
| **SUNIDHI SINGH** | **A023119820032** | **5AI 1** |

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

AMITY SCHOOL OF ENGINEERING AND TECHNOLOGY

AMITY UNIVERSITY UTTAR PRADESH

NOIDA (U.P)

**OPEN ENDED EXPERIMENT**

**Aim**

To implement k means clustering algorithm over a dataset

**Software Used**

Google Colab

**Theory**

K-Means Clustering is an [Unsupervised Learning algorithm](https://www.javatpoint.com/unsupervised-machine-learning), which groups the unlabeled dataset into different clusters. Here K defines the number of pre-defined clusters that need to be created in the process, as if K=2, there will be two clusters, and for K=3, there will be three clusters, and so on. It is a centroid-based algorithm, where each cluster is associated with a centroid. The main aim of this algorithm is to minimize the sum of distances between the data point and their corresponding clusters. The performance of the K-means clustering algorithm depends upon highly efficient clusters that it forms. But choosing the optimal number of clusters is a big task.

### **Elbow Method**

The Elbow method is one of the most popular ways to find the optimal number of clusters. This method uses the concept of WCSS value. **WCSS** stands for **Within Cluster Sum of Squares**, which defines the total variations within a cluster.

To find the optimal value of clusters, the elbow method follows the below steps:

* It executes the K-means clustering on a given dataset for different K values (ranges from 1-10).
* For each value of K, calculates the WCSS value.
* Plots a curve between calculated WCSS values and the number of clusters K.
* The sharp point of bend or a point of the plot looks like an arm, then that point is considered as the best value of K.

**Program Code and Output**





Table

Description automatically generated

Table

Description automatically generated

Graphical user interface, table

Description automatically generated with medium confidence

Table

Description automatically generated

Table

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated



Table

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

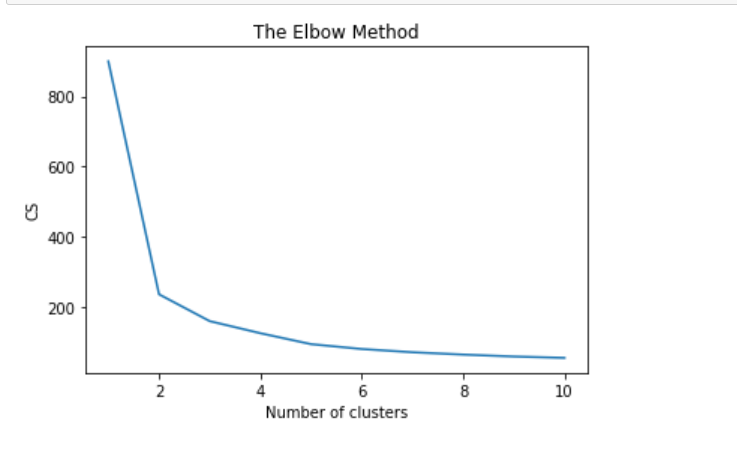
Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

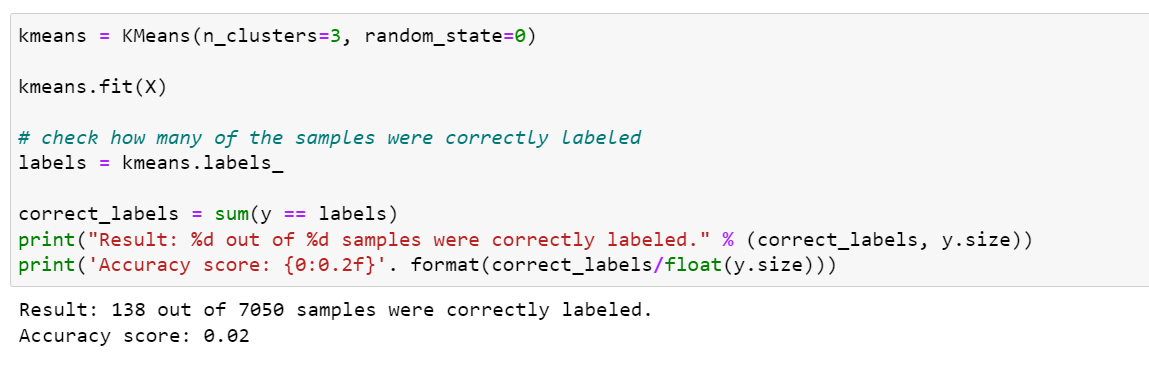
Text

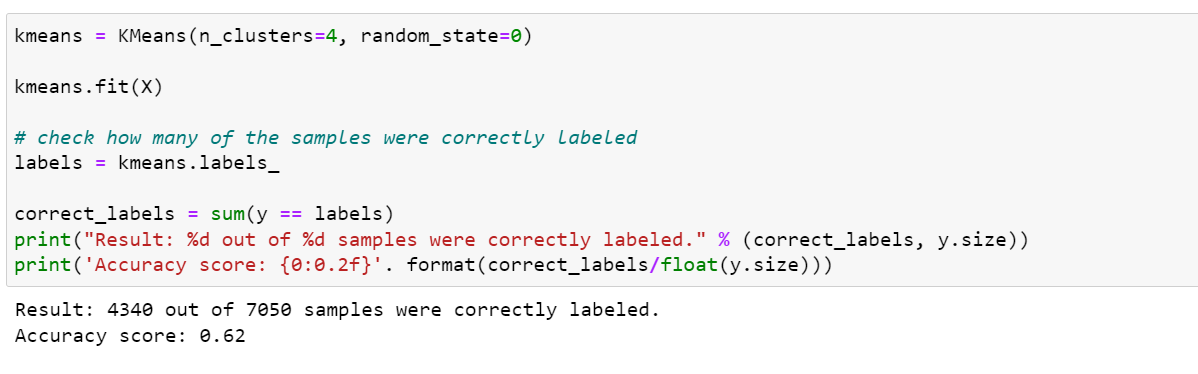
Description automatically generated with medium confidence



Graphical user interface, text, application, email

Description automatically generated





Graphical user interface, text, application, email

Description automatically generated

**Discussion and Conclusion**

The lesser the model inertia, the better the model fit. So, we use the elbow method to find optimal number of clusters. There is a kink at k=2.Hence k=2 can be considered a good number of the cluster to cluster this data. So, we have changed the value of k and found relatively higher classification accuracy of 62% with k=4. Hence, we can conclude that k=4 being the optimal number of clusters.

