



DAVIN BRAVEN - 2K20/SE/09



Amazon User Segmentation

MINI PROJECT - SE391
Prof. Shweta Meena Mam
HoD of SE Ruchika Malhotra

Contents

Amazon User Segmentation

- Objective
- Machine Learning & AI
- Data Preprocessing
- Clustering
- K-Means Clustering
- K-Means Random Initialization Trap
- Future Work
- References
- End - Thank You

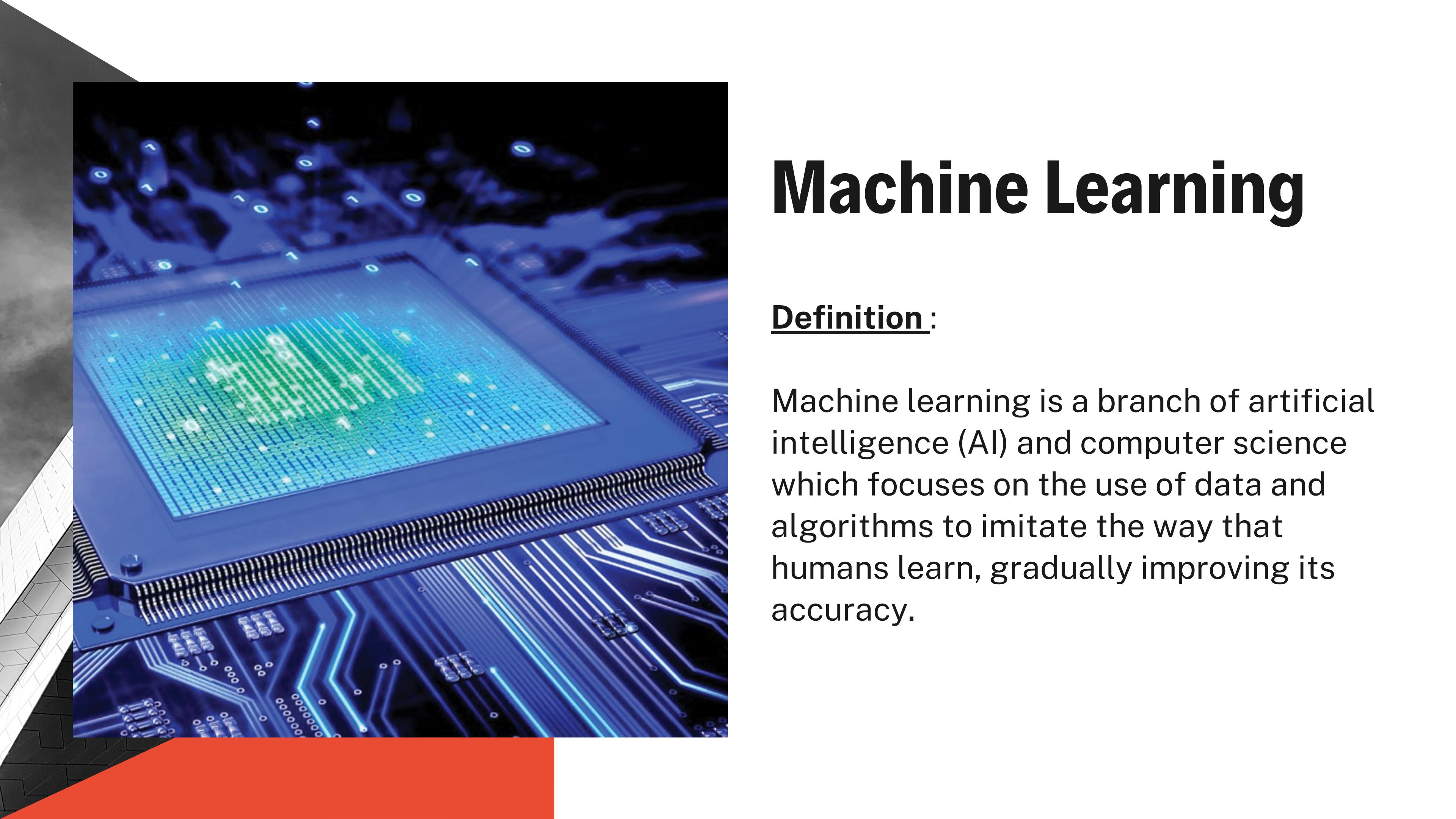
OBJECTIVE

Amazon Personalize now offers intelligent user segmentation which allows you to run more effective prospecting campaigns through your marketing channels. Traditionally, user segmentation has relied on demographic information and manually curated business rules to make assumptions about users' intentions and assign them to pre-defined audience segments. Amazon Personalize uses machine learning techniques to learn about your items, users, and how your users interact with your items. Amazon Personalize segments users based on their preferences for different products, categories, brands, and more. This can help you drive higher engagement with marketing campaigns, increase retention through targeted messaging, and improve the return on investment for your marketing spend.



What is Machine Learning & AI ?

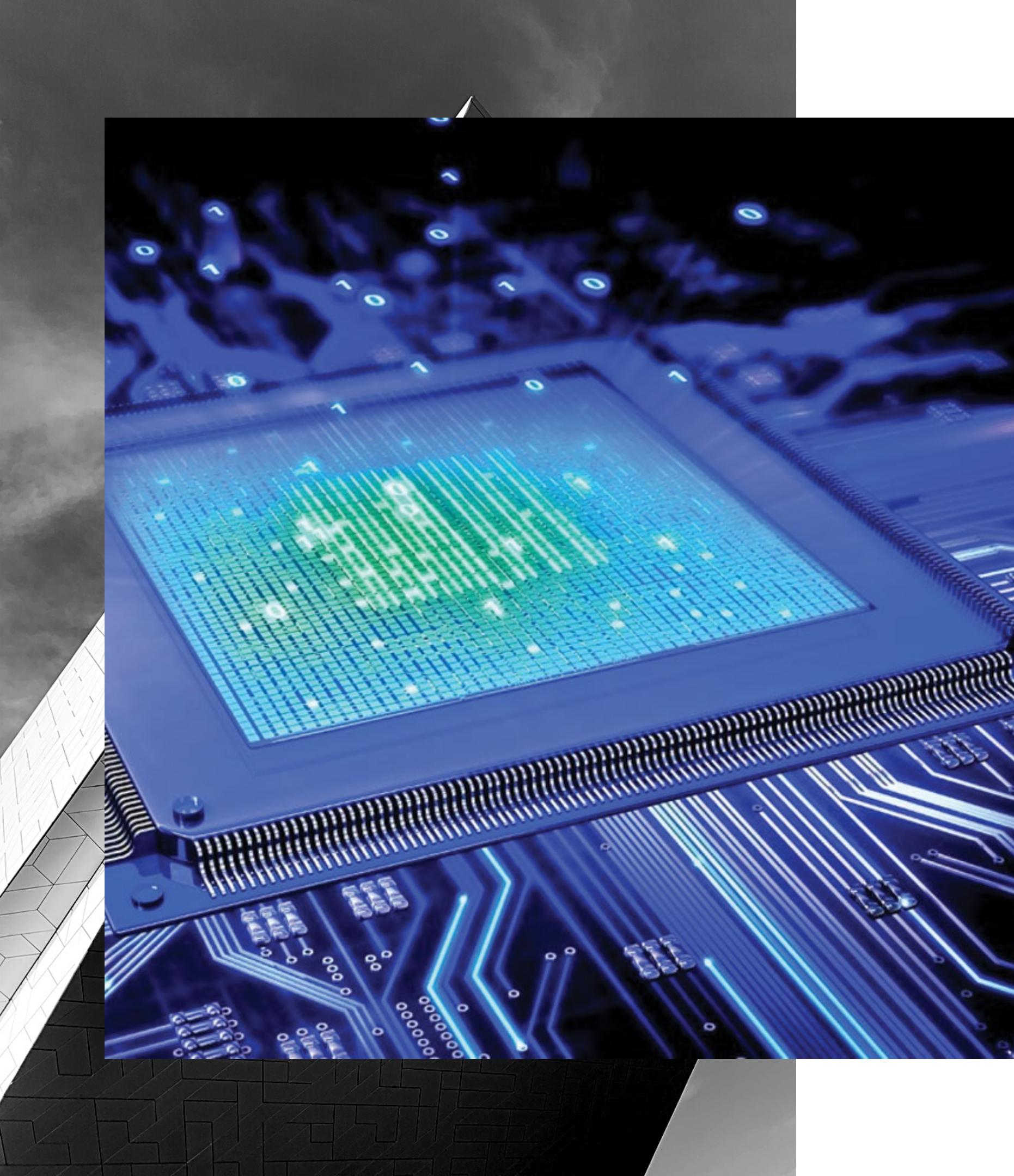


A close-up photograph of a computer circuit board, likely a GPU or AI accelerator, showing its complex network of blue and green printed circuit boards. Superimposed on the board is a grid of glowing blue binary digits (0s and 1s) forming a matrix pattern, symbolizing data processing and machine learning operations.

Machine Learning

Definition :

Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.

A close-up photograph of a computer circuit board. The board is primarily blue and black, with intricate silver-colored metal traces and components. A central component, possibly a microchip or memory module, is highlighted with a vibrant green and blue color overlay, showing a grid-like pattern of glowing dots and lines. The background is dark, making the glowing elements stand out. The overall composition emphasizes the complexity and advanced technology of artificial intelligence.

Artificial Intelligence

Definition:

" It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable."

Data Pre Processing – ML

Data preprocessing in Machine Learning is a crucial step that helps enhance the quality of data to promote the extraction of meaningful insights from the data. Data preprocessing in Machine Learning refers to the technique of preparing (cleaning and organizing) the raw data to make it suitable for a building and training Machine Learning models. In simple words, data preprocessing in Machine Learning is a data mining technique that transforms raw data into an understandable and readable format.

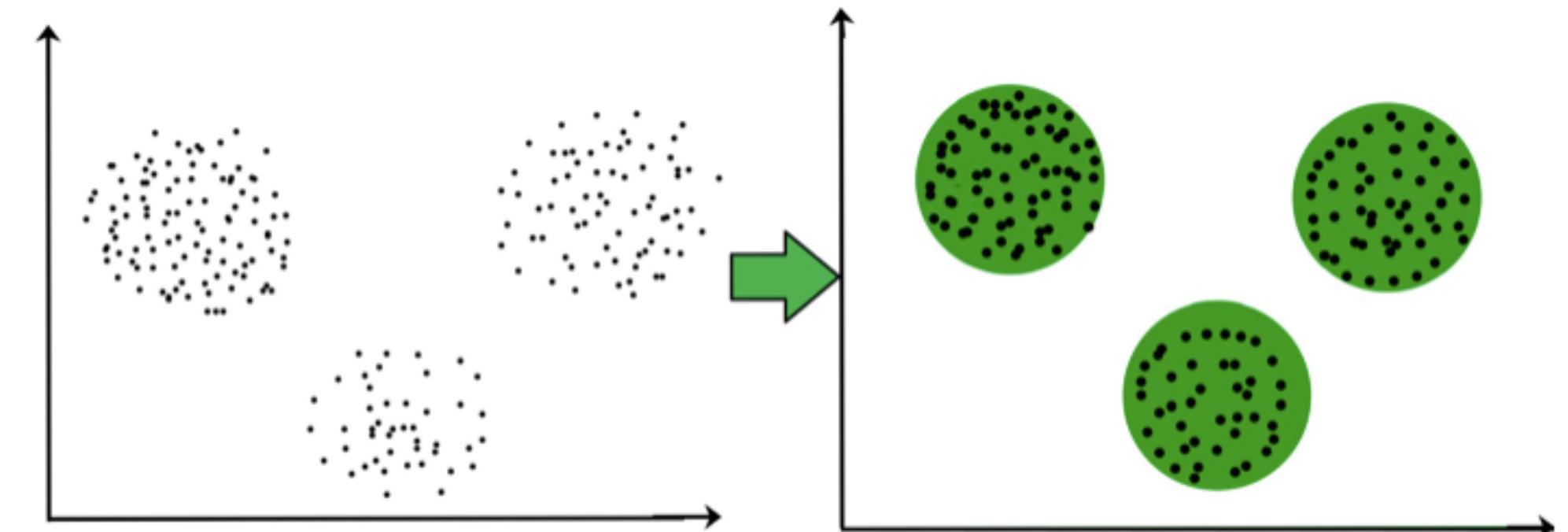
1. Acquire the dataset
2. Import all the crucial libraries
3. Import the dataset
4. Identifying and handling the missing values
5. Encoding the categorical data
6. Splitting the dataset
7. Feature scaling



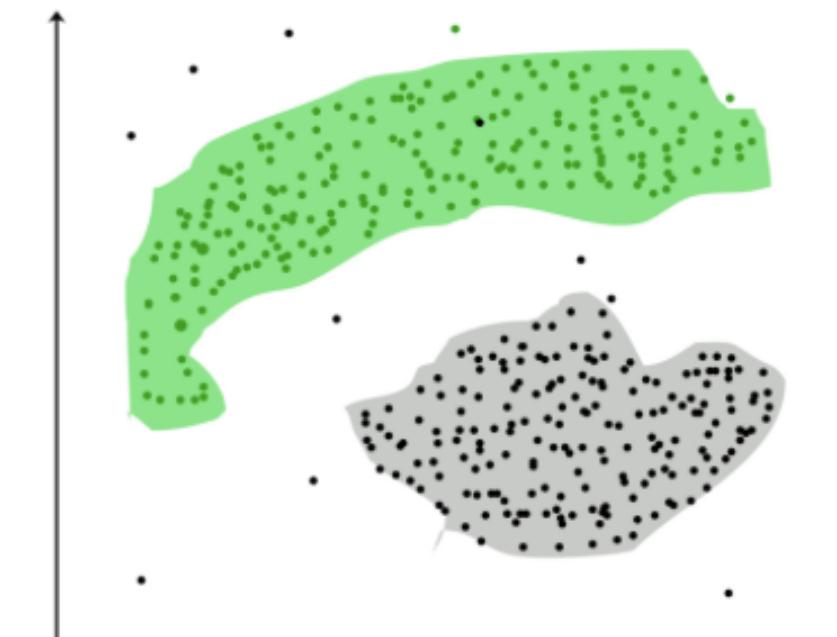
Clustering

Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group and dissimilar to the data points in other groups.

For ex- The data points in the graph below clustered together can be classified into one single group. We can distinguish the clusters, and we can identify that there are 3 clusters in the below picture.



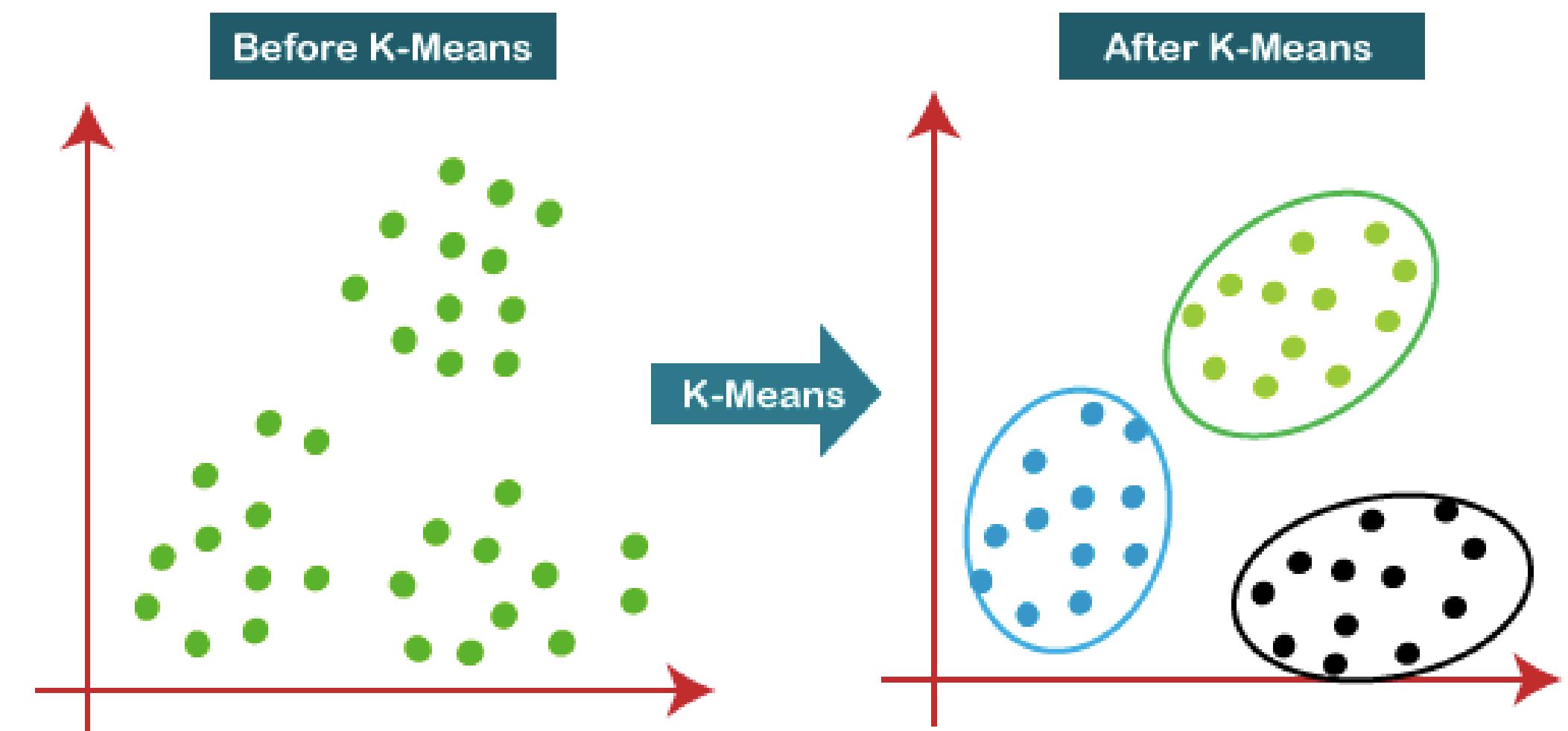
It is not necessary for clusters to be spherical. Such as :



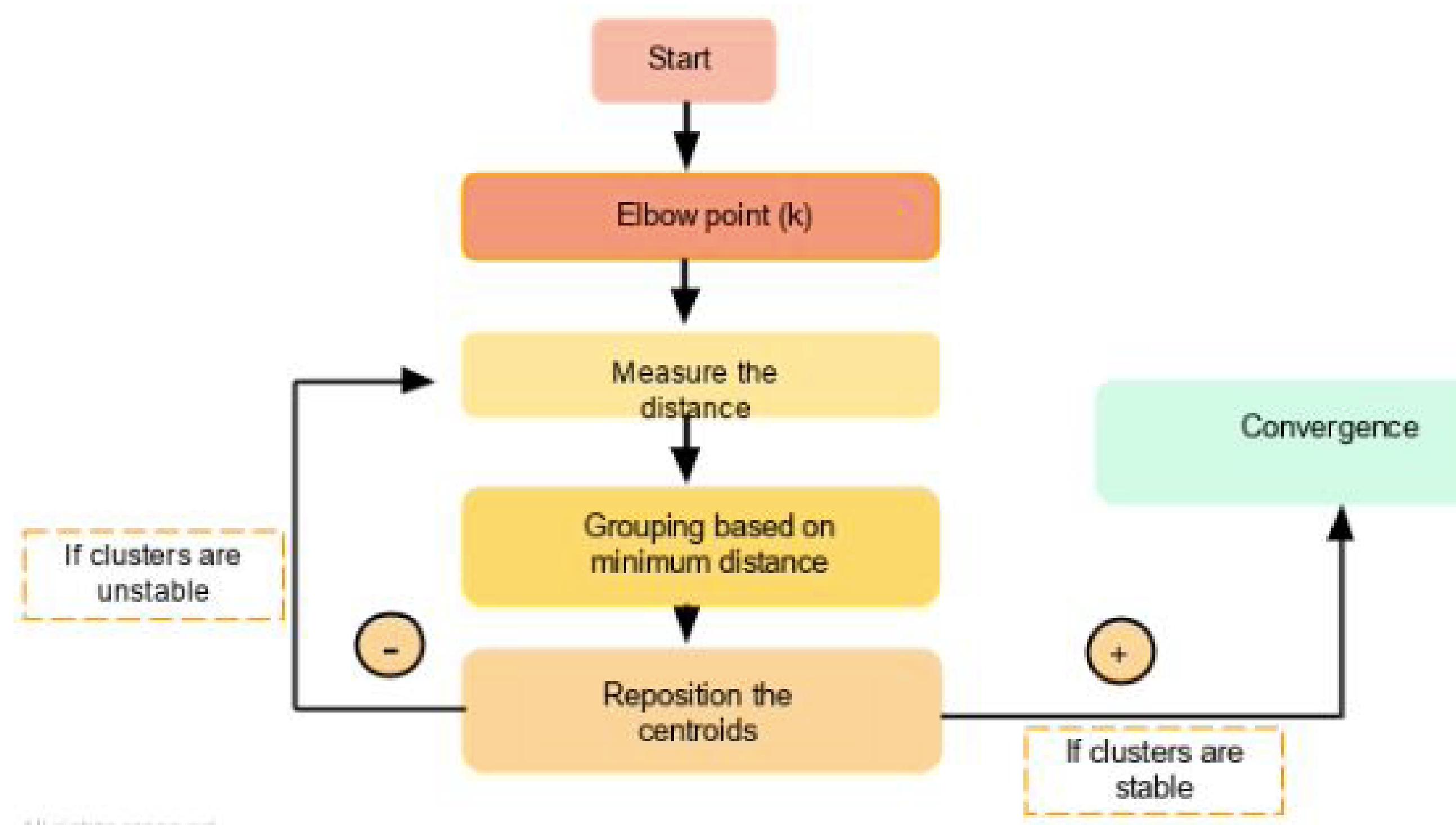
K-Means Clustering

K-Means Clustering is an Unsupervised Learning algorithm, which groups the unlabeled dataset into different clusters.

The algorithm takes the unlabeled dataset as input, divides the dataset into k-number of clusters, and repeats the process until it does not find the best clusters. The value of k should be predetermined in this algorithm.



How Does K-Means Works ?



K-Means Clustering Steps

01

**Data
Pre Processing
Step**

02

**Importing the
Libraries and
Data Set**

03

**Finding the
optimal number
of clusters using
the elbow
method**

04

**Training the
K-means
algorithm on the
training dataset**

05

**Visualizing the
Clusters**

Random Initialization Trap

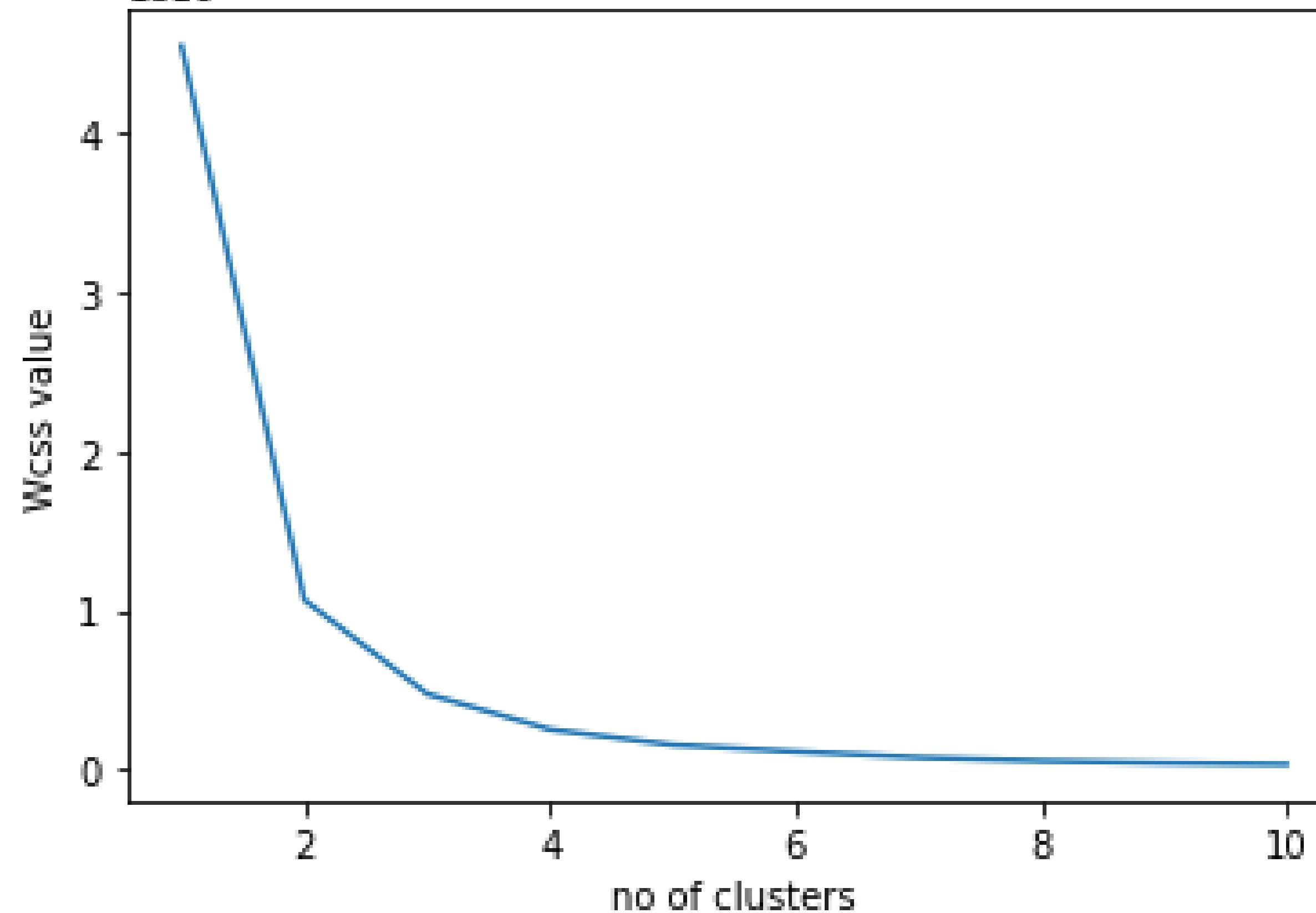
Random initialization trap is a problem that occurs in the K-means algorithm.

Random initialization trap is a problem that occurs in the K-means algorithm. In random initialization trap when the centroids of the clusters to be generated are explicitly defined by the User then inconsistency may be created and this may sometimes lead to generating wrong clusters in the dataset. So random initialization trap may sometimes prevent us from developing the correct clusters.

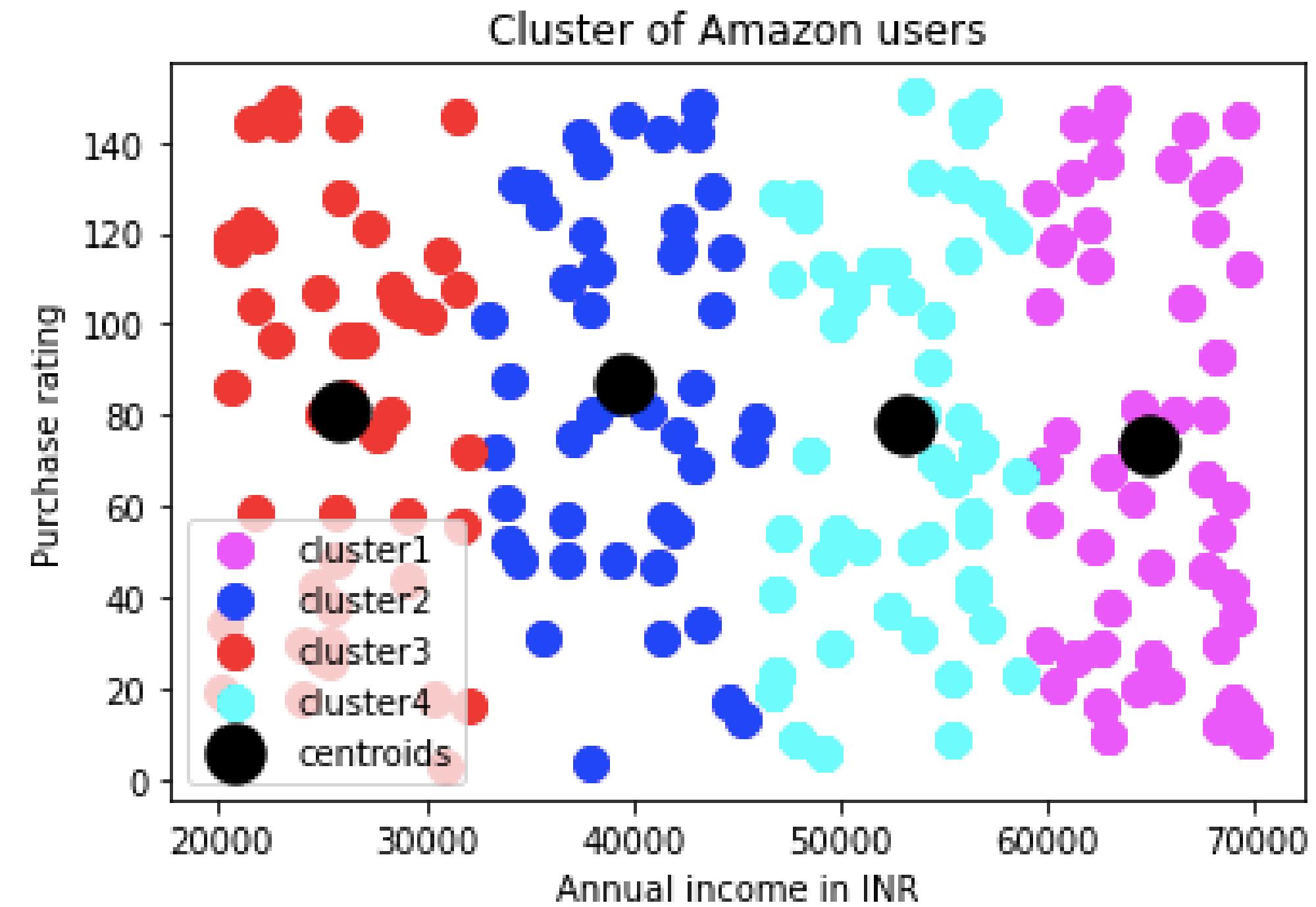
Amazon User Segmentation

wcss via elbow method

le10



Cluster of Amazon Users



Future Work



- 01 Target large enterprises.
- 02 Retrain low-performing coordinators.
- 03 Implement the new system.
- 04 Assign high performers to new territories.
- 05 Present 2025 targets at annual staff meeting.

References



- 01 Maching Learning book (By Dr. Bret Lantz)
- 02 Machine Learning with Python book (By Daniel Geron)
- 03 GeeksforGeeks and Javapoint
- 04 TowardsDataScience - Website
- 05 Google Colab and Jupyter



Thank You!

DAVIN BRAVEN - 2K20/SE/09

