CUSTOMER CHURN ANALYSIS

MENTORS:
RASHI SHARMA
RITIK SHAH
MUDIT AGRAWAL
MRIDUL GUPTA

INTRODUCTION TO DATA SCIENCE

The field of data science involves using scientific methods, processes, algorithms, and systems to extract knowledge and insights from structured and unstructured data.

Libraries: Python libraries useful for Data Analysis - Numpy, Pandas, Matplotlib and Seaborn.

Regression: Regression analysis is a statistical method for predicting a dependent variable based on one or more independent variables.

Any general **linear equation** with **multiple features** can be written as :

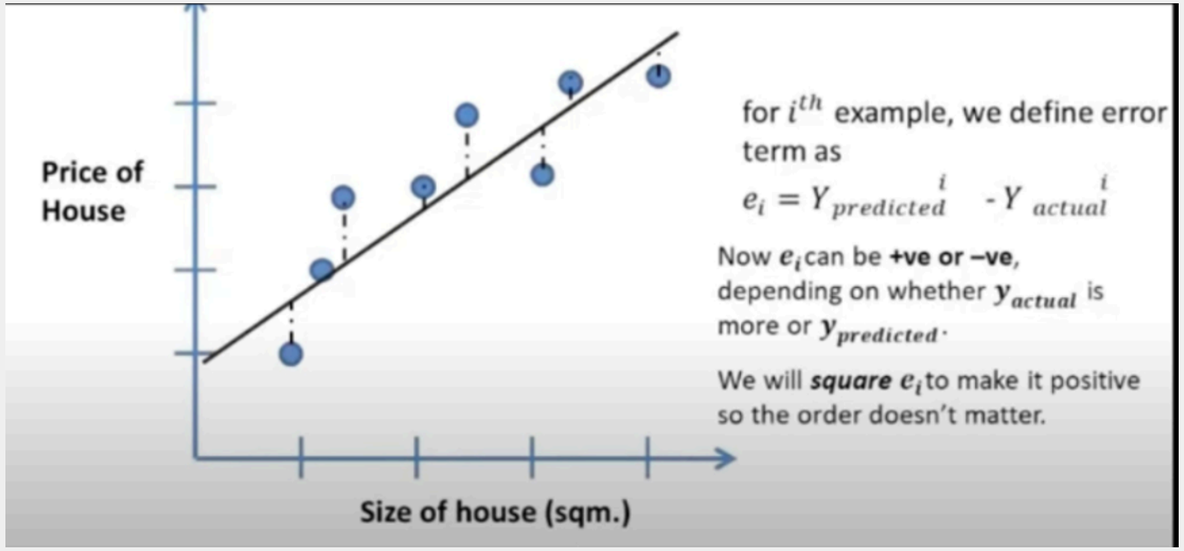
$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + \ldots\ldots\ldots + a_nx_n$$

x_i *are features*

a_i *are model parameters or coefficients*

y *is target variable*

LINEAR REGRESSION



OLS method:

The Ordinary Least Squares (OLS) method estimates unknown parameters in a linear regression model by minimizing the sum of the squared differences between observed and predicted values.

Cost function **(J)** = $\frac{1}{2m}(e_1^{\ 2} + e_2^{\ 2} + e_3^{\ 2} + \ldots\ldots\ldots\ldots + e_m^{\ 2})$

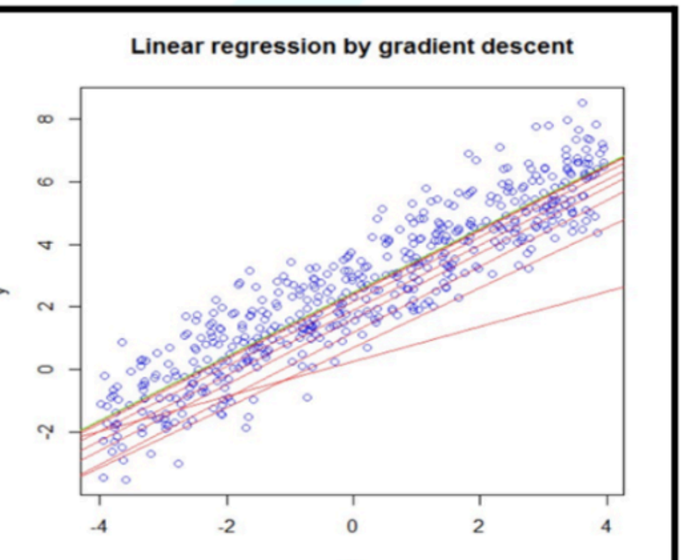
$$J(a) = \frac{1}{2m} \sum_{i=1}^{m} (y_{i\,(pre)} - y_{i\,(act)})^2$$

$$J(a) = \frac{1}{2m} \sum_{i=1}^{m} (a_0 + a_1 x_1^{\ (i)} - y_{i\,(act)})^2$$

Cost function (J) is a function of parameter space $a = (a_0, a_1)$.

Gradient Descent:

Gradient Descent is an optimization algorithm that iteratively adjusts model parameters to minimize the cost function, improving the model's accuracy.



Step 1:

Calculate $\frac{\partial J}{\partial a_0}$ **(slope)** at the current value of parameter a_0.

Calculate $\frac{\partial J}{\partial a_1}$ **(slope)** at the current value of parameter a_1.

Step 2:

$$(new)a_0 = a_0 - \alpha(\frac{\partial J}{\partial a_0})$$

$$(new)a_1 = a_1 - \alpha(\frac{\partial J}{\partial a_1})$$

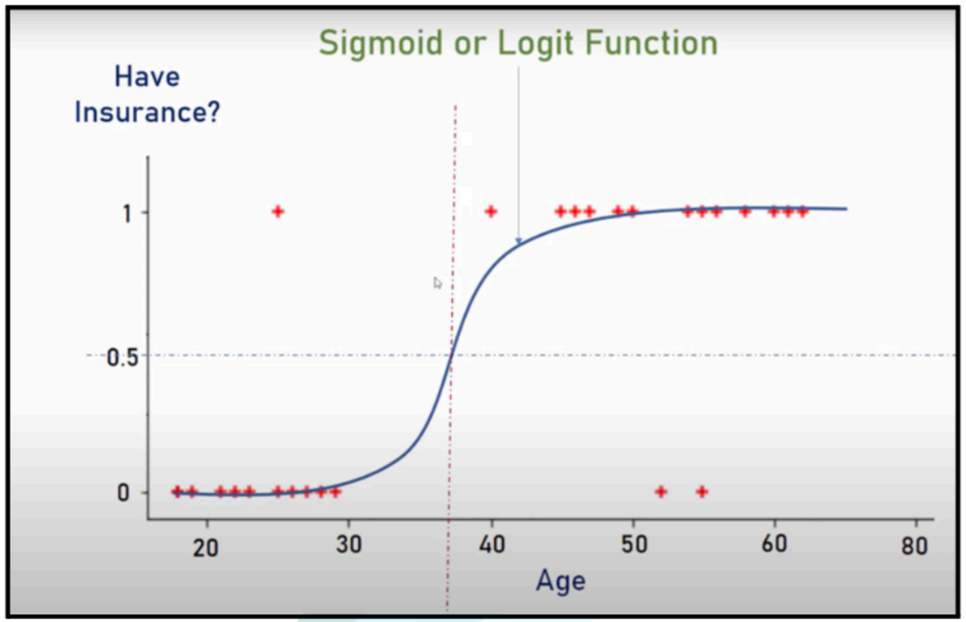
Step 3:

update Cost Function J with **new (a_0 and a_1)**

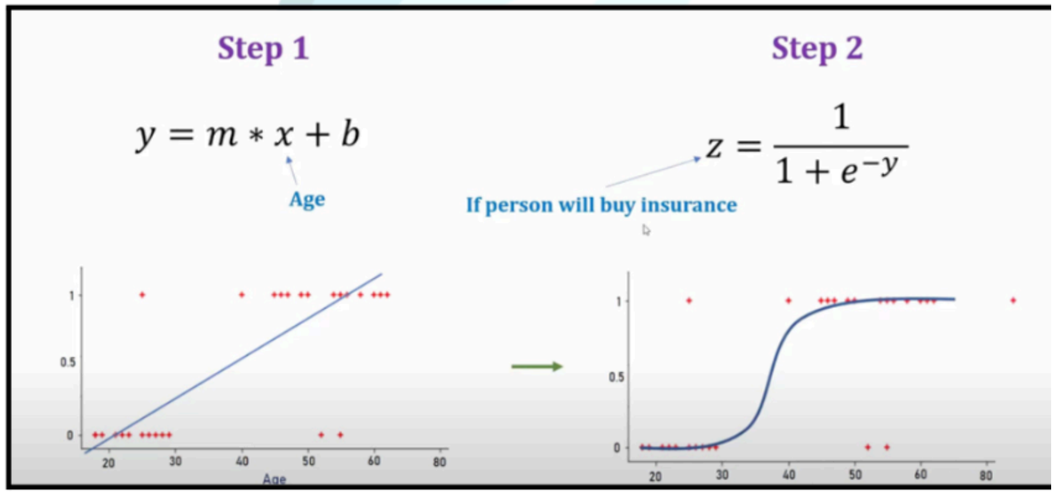
Repeat Step 1

LOGISTIC REGRESSION

Logistic regression is a statistical method for binary classification, predicting the probability that an input belongs to one of two possible categories. (e.g., yes/no, success/failure)



After using the sigmoid function the output values are in the range (0, 1), making it suitable for representing probabilities.




```python
class LogisticRegression():
    def __init__(self):
        self.w = None
        self.b = None

    # return dot product of weight vector with data point
    def dot_pro(self,x):
        return (np.dot(self.w,x) + self.b)

    # return the probability of the point belonging to a class.
    def predict(self, x):
        return 1.0/(1.0 + np.exp(-self.dot_pro(x)))

    # Calculates gradient w.r.t w
    def gradient_w(self,x,y):
        pred = self.predict(x)
        return (pred - y)*x

    # Calculates gradient w.r.t b
    def gradient_b(self,x,y):
        pred = self.predict(x)
        return (pred - y)

    # Fit method
    def fit(self, x_train, y_train, epochs=100, learning_rate=0.01, refit=True):
        # initializing weights with random values.
        if refit:
            self.w = np.random.randn(x_train.shape[1])
            self.b = 0
        for i in range(epochs):
            grad_w = 0
            grad_b = 0
            for x, y in zip(x_train, y_train):
                grad_w += self.gradient_w(x, y)
                grad_b += self.gradient_b(x, y)

            self.w = self.w - learning_rate*grad_w
            self.b = self.b - learning_rate*grad_b
```

EDA

Data Types and Correlation EDA examines the relationships between categorical and numerical data using tools like histograms, box plots, heatmaps, and scatter plots.

1. Analyze the correlation between categorical and numerical data to understand their interactions.
2. Use histograms and box plots for univariate analysis, and heatmaps and scatter plots for multivariate analysis.
3. Automate EDA with Pandas Profiling for concise reports with minimal code.

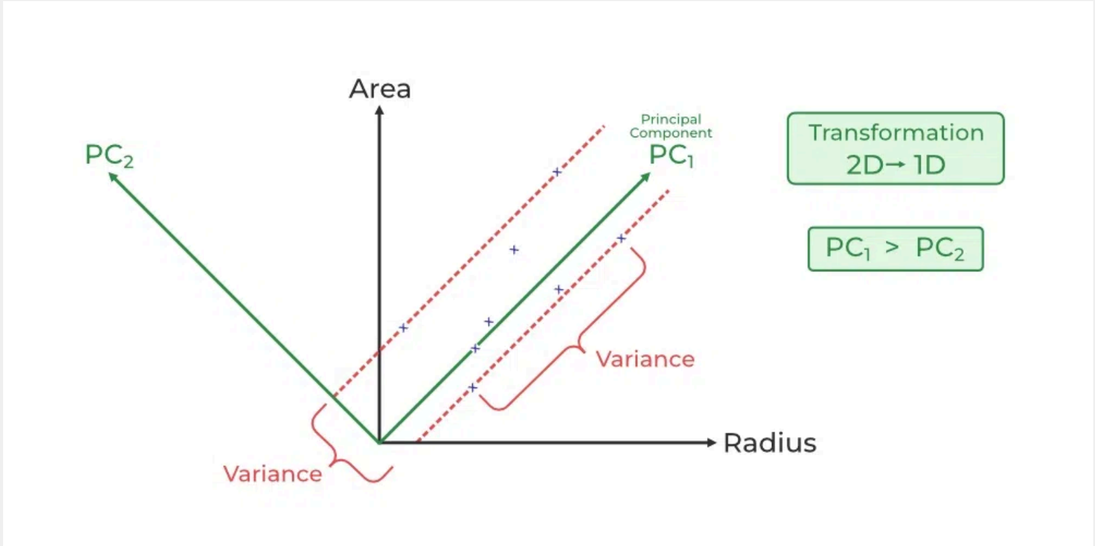
FEATURE ENGINEERING

Feature engineering is the process of transforming raw data into features that are suitable for machine learning models.

Feature engineering consists of the following parts:

- **Feature Transformation**
- **Feature Scaling:**
  - Standardisation: (Xi - Xmean)/σ, where sigma is the standard deviation of the data
  - Normalisation:
    - Min - Max Scaling:
      (Xi - Xmin)/(Xmax - Xmin)
    - Mean Normalisation
    - Robust Scaling
    - Max Absolute Scaling
- **Feature Construction**
- **Feature Extraction**
- **Data Imputation**

PCA



Function and Power Transformation

Handling skewed or unsymmetrical distributions is essential for improving model accuracy.
Techniques include:
1. Power Transformation: **Yeo-Johnson, Box-Cox**
2. Function Transformation: Logarithmic, Exponential, Square Root, Reciprocal
Many machine learning algorithms assume normally distributed data, so handling skewed distributions is crucial to avoid biased or inaccurate models.

Power Transformation:

- Yeo-Johnson
- Box-cox

Function Transformation:

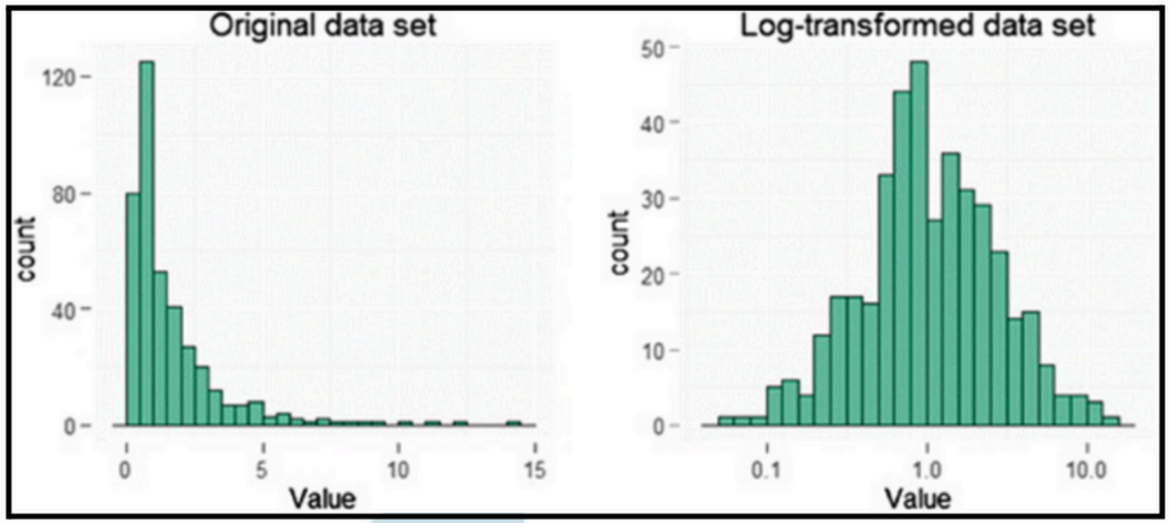
- Logarithmic
- Exponential
- Square Root
- Reciprocal

The Yeo-Johnson transform is given by:

$$x_i^{(\lambda)} = \begin{cases} [(x_i+1)^\lambda - 1]/\lambda & \text{if } \lambda \neq 0, x_i \geq 0, \\ \ln(x_i+1) & \text{if } \lambda = 0, x_i \geq 0 \\ -[(-x_i+1)^{2-\lambda} - 1]/(2-\lambda) & \text{if } \lambda \neq 2, x_i < 0, \\ -\ln(-x_i+1) & \text{if } \lambda = 2, x_i < 0 \end{cases}$$

while the Box-Cox transform is given by:

$$x_i^{(\lambda)} = \begin{cases} \dfrac{x_i^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0, \\ \ln(x_i) & \text{if } \lambda = 0, \end{cases}$$



Logarithmic Transformation

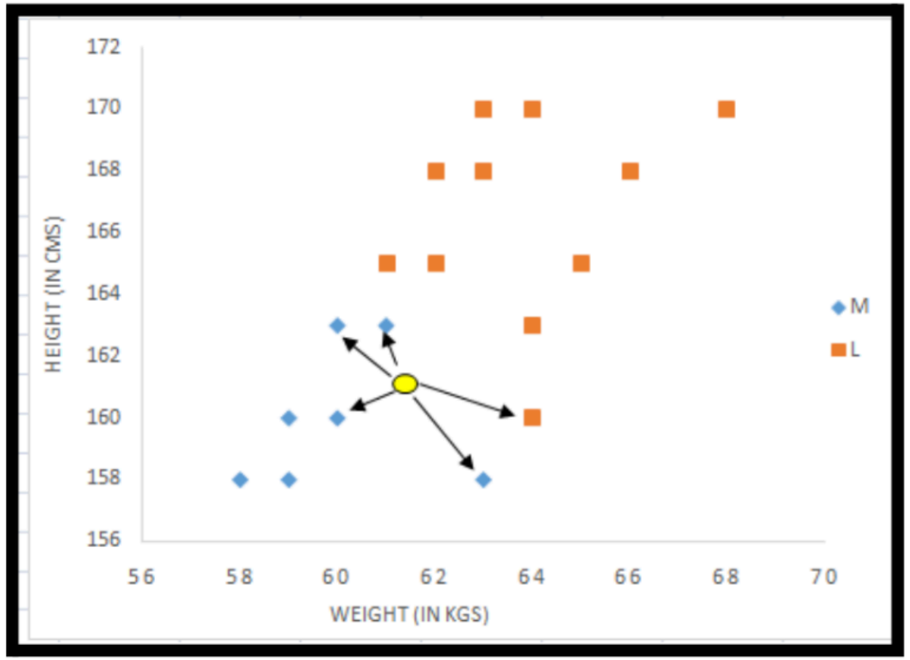
KNN IMPUTER AND ALGORITHM, MICE ALGORITHM

KNN imputation is a more sophisticated approach that considers the proximity of data points. It imputes missing values by averaging or voting based on the values of the k-nearest neighbors in the feature space.

The MICE algorithm is a powerful statistical tool that reconstructs missing data by iteratively estimating values based on relationships between variables

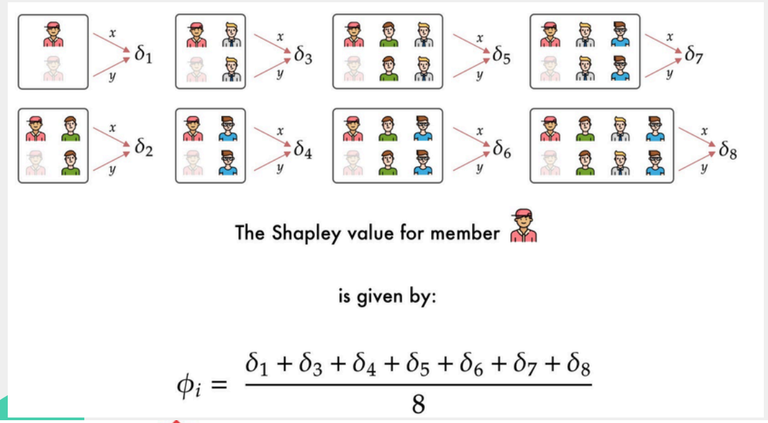
A	B	C	D
45		24	37
62	21	34	
	64	33	57
34	34	56	
	23		48
64			69
56		22	24
86	45	26	
		78	97
21	39	91	

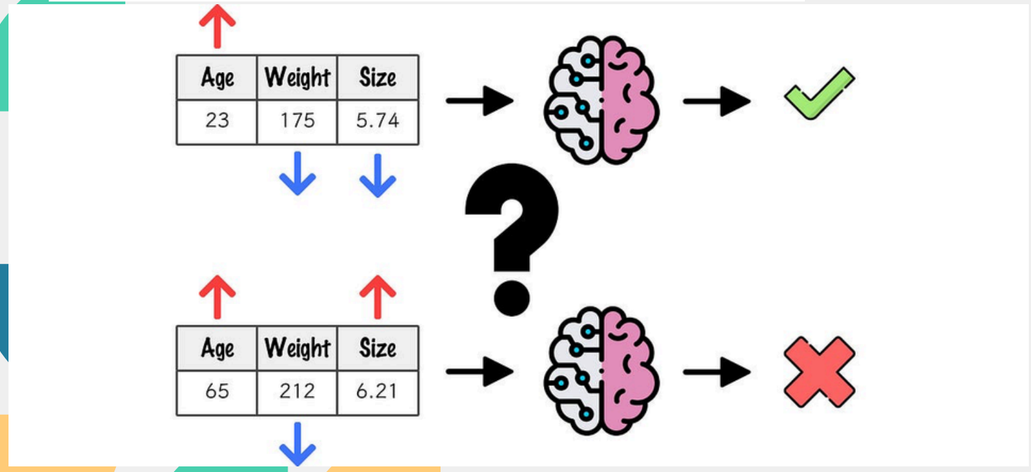
A	B	C	D
45	35	24	37
62	21	34	39
51	64	33	57
34	34	56	41
34	23	36	48
64	65	66	69
56	34	22	24
86	45	26	52
88	87	78	97
21	39	91	50

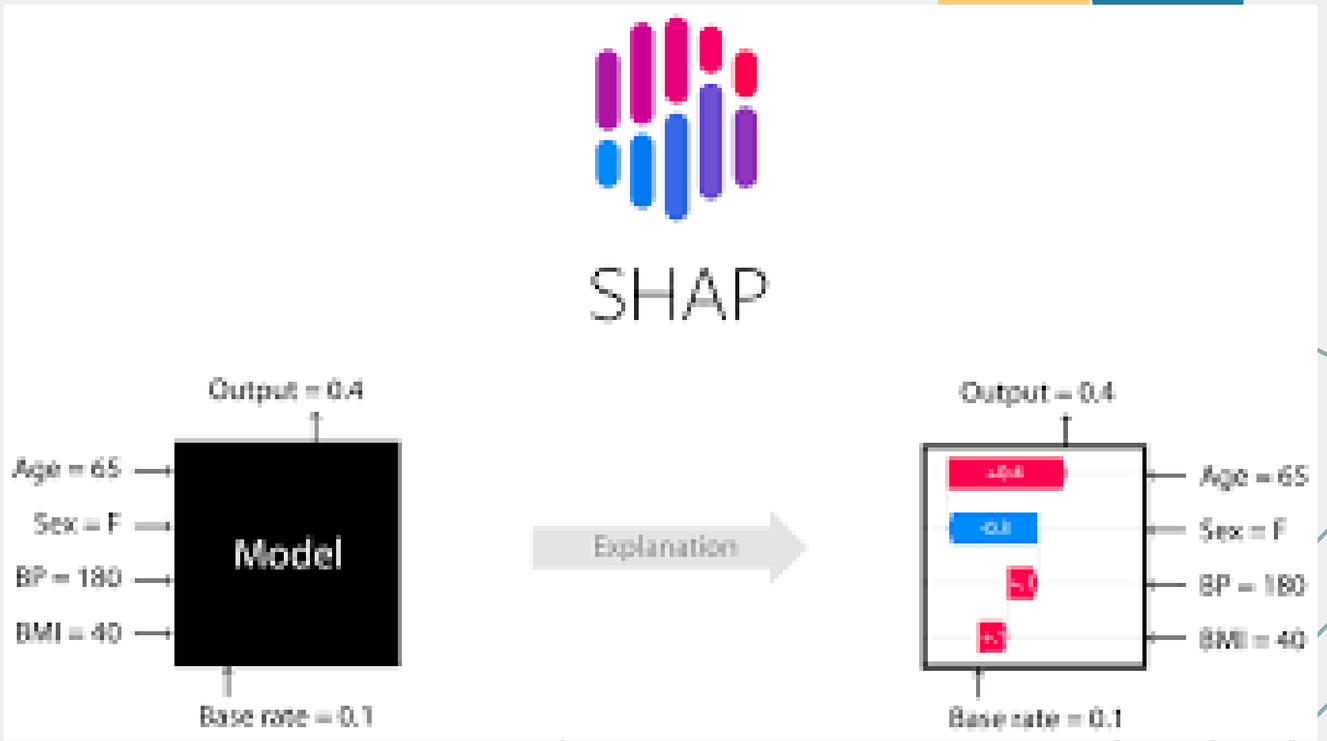


SHAP AND SHAPLEY

SHAP (SHapley Additive exPlanations) values are used to assign an importance value to each feature representing how significantly it contributes to the model's output. SHAP also helps in comparing the significance of each feature compared to others and the model's reliance on the interaction between features.

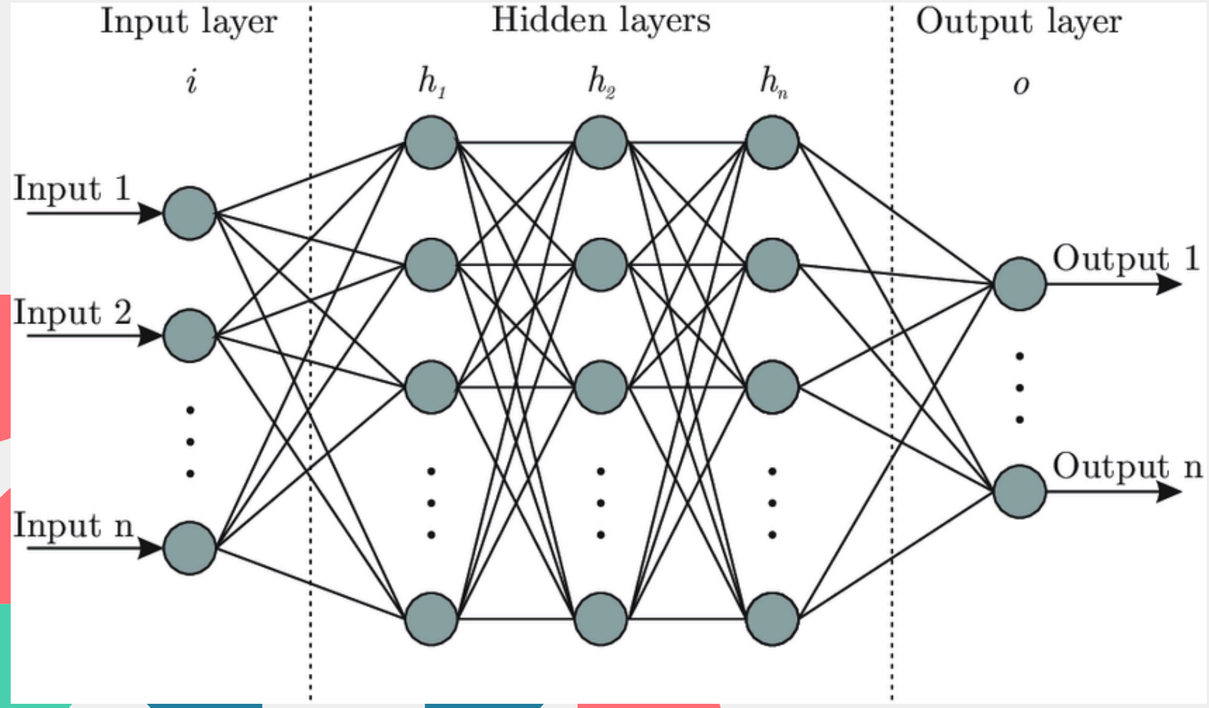




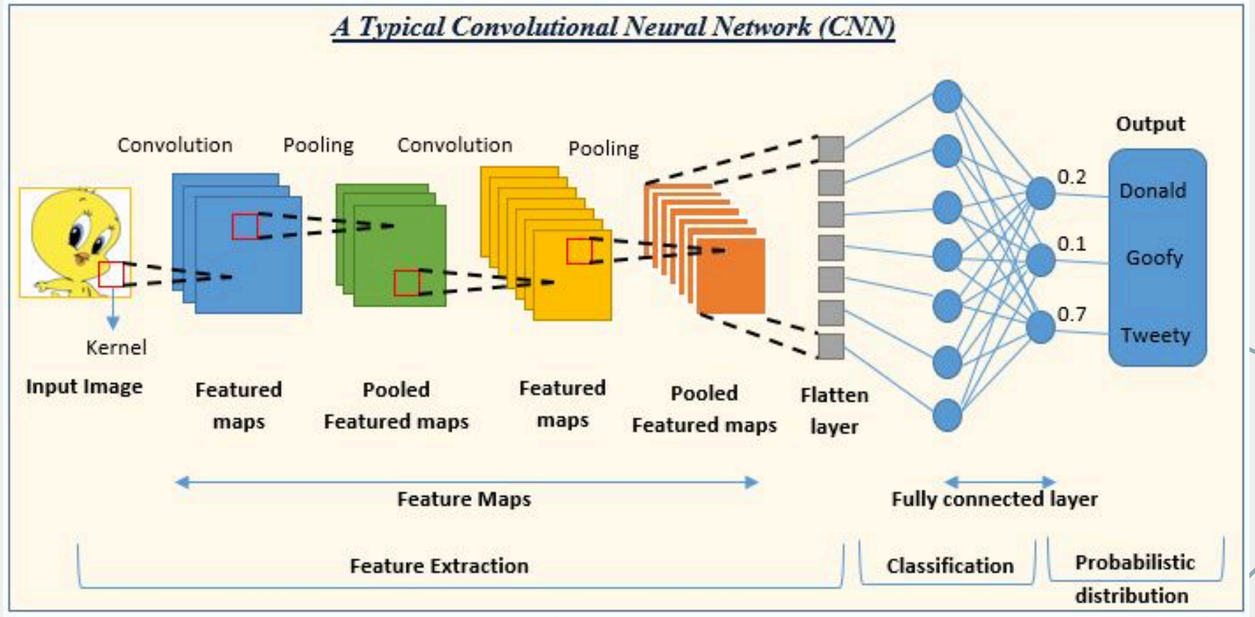


NEURAL NETWORKS

ANN



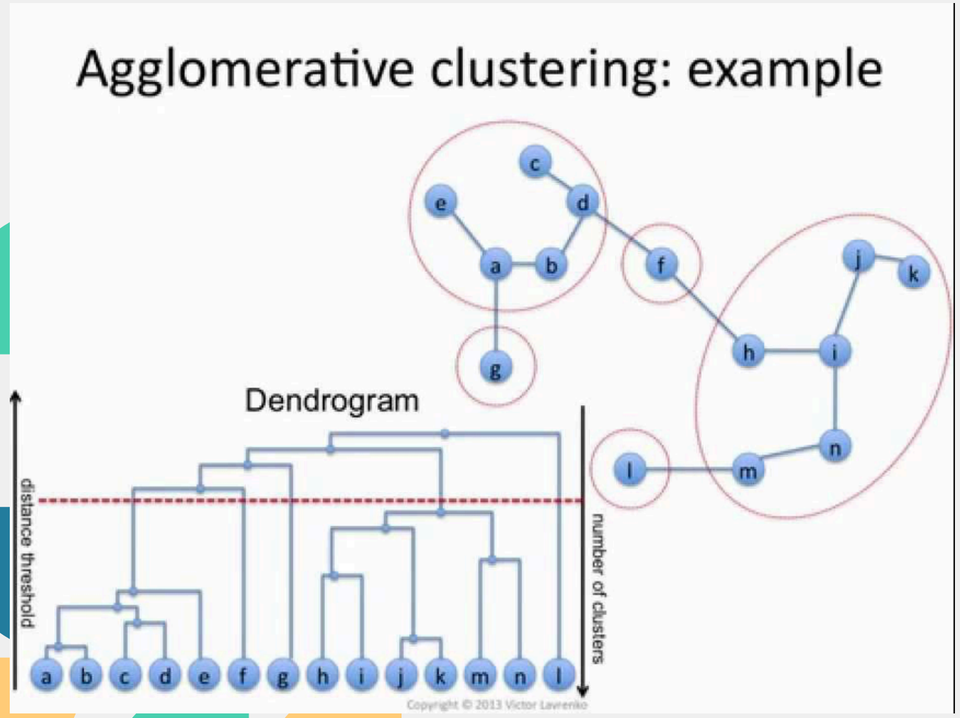
CNN



HIERARCHIAL CLUSTERING(DS)

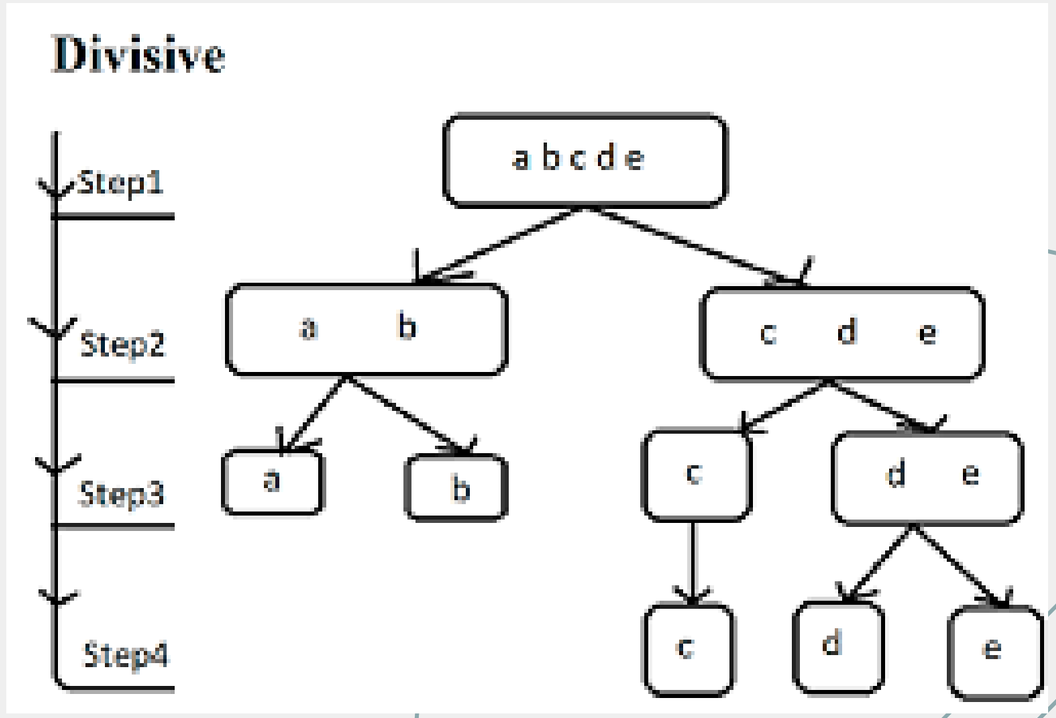
Agglomerative

The algorithm starts by treating each object as a singleton cluster. Next, pairs of clusters are successively merged until all clusters have been merged into one big cluster containing all objects. The result is a tree-based representation of the objects, named dendrogram.



Divisive

Divisive clustering is a hierarchical clustering method that involves dividing every cluster into smaller subsets, starting with each object in a single cluster, until the desired number of clusters is achieved. It uses a top-down approach and splits clusters into two sub-clusters based on the analysis of all possible bipartitions.



CONSULTING

TYPES OF CONSULTING

STRATEGY CONSULTING

MANAGEMENT CONSULTING

OPERATIONAL CONSULTING

FINANCIAL CONSULTING

LEGAL CONSULTING

IT CONSULTING

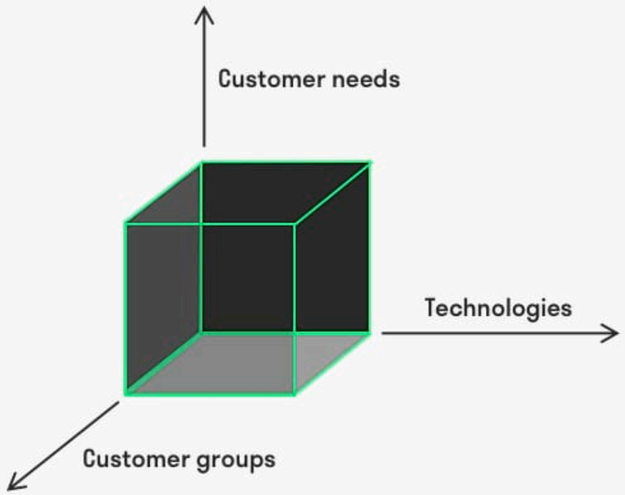
FRAMEWORKS


S
WEAKNESSES
O
THREATS
STRENGTHS
W
OPPORTUNITIES
T

SO WHAT'S SWOT ANALYSIS

A SWOT analysis is a strategic planning tool that helps an organization identify its strengths, weaknesses, opportunities, and threats. This framework aids in understanding the internal and external factors that can impact the organization's objectives, enabling effective strategy formulation.

ABELL MODEL



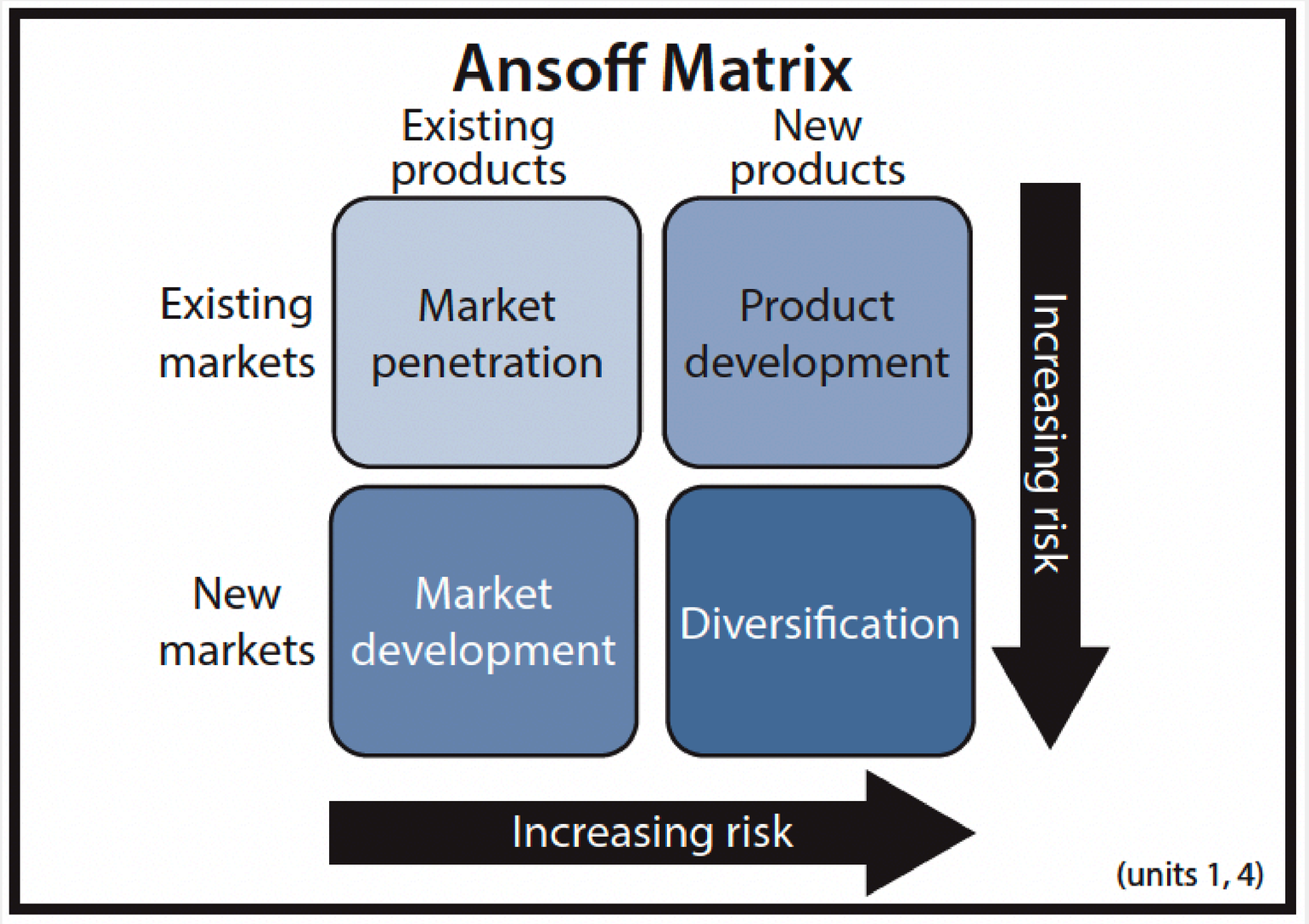
The **Abell Model**, is a strategic tool used to define the business scope and identify growth opportunities. It focuses on three dimensions:

Customer Groups (Who):
- Categorizing groups based on demographics, geography, and behavior.
- Understanding the needs and preferences of these groups to tailor products and services accordingly.

Customer Needs (What):
- Determining the specific needs and wants of the identified customer groups.
- Analyzing how well current products or services meet these needs and identifying gaps for new offerings.

Technologies (How):
- Assessing the technologies used to satisfy customer needs.
- Exploring new technologies that can enhance product offerings or create new market opportunities.

ANSOFF MATRIX



TYPES OF MARKETING STRATEGIES AND CASE STUDIES

CASE STUDIES

udaan

marico

Unilever

Types of Marketing Strategies Used

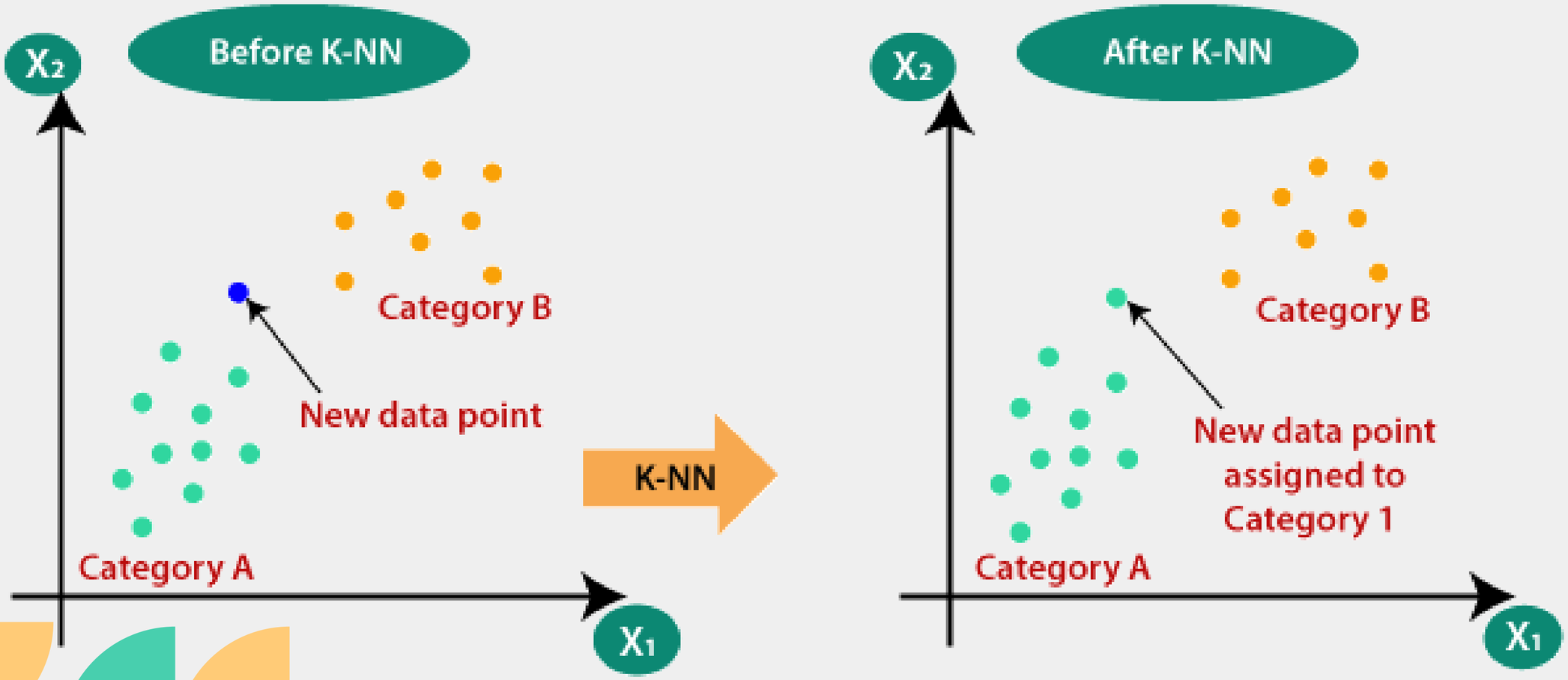
- Market Segmentation
- Target Market
- Factors Considered

- Strategy Making
- QnA
- SWOT

CLUSTERING

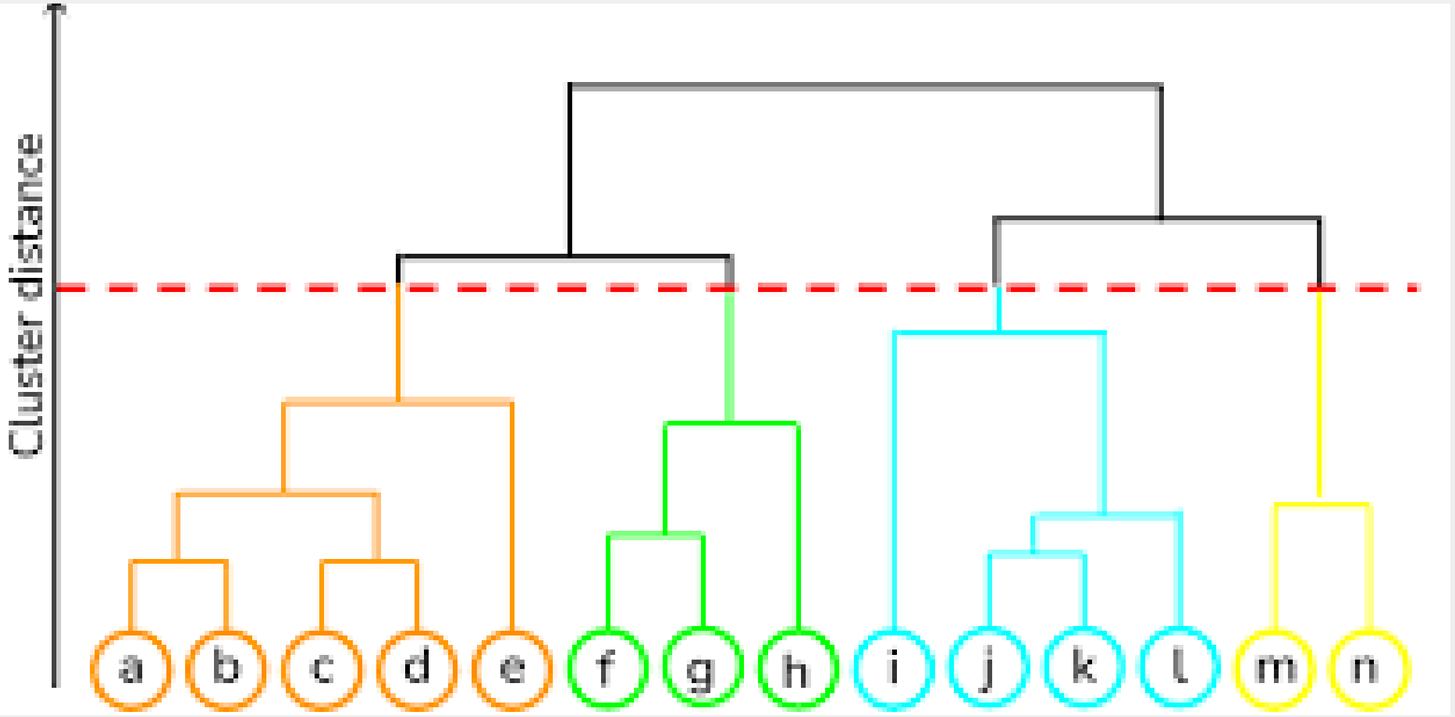
KNN CLUSTERING

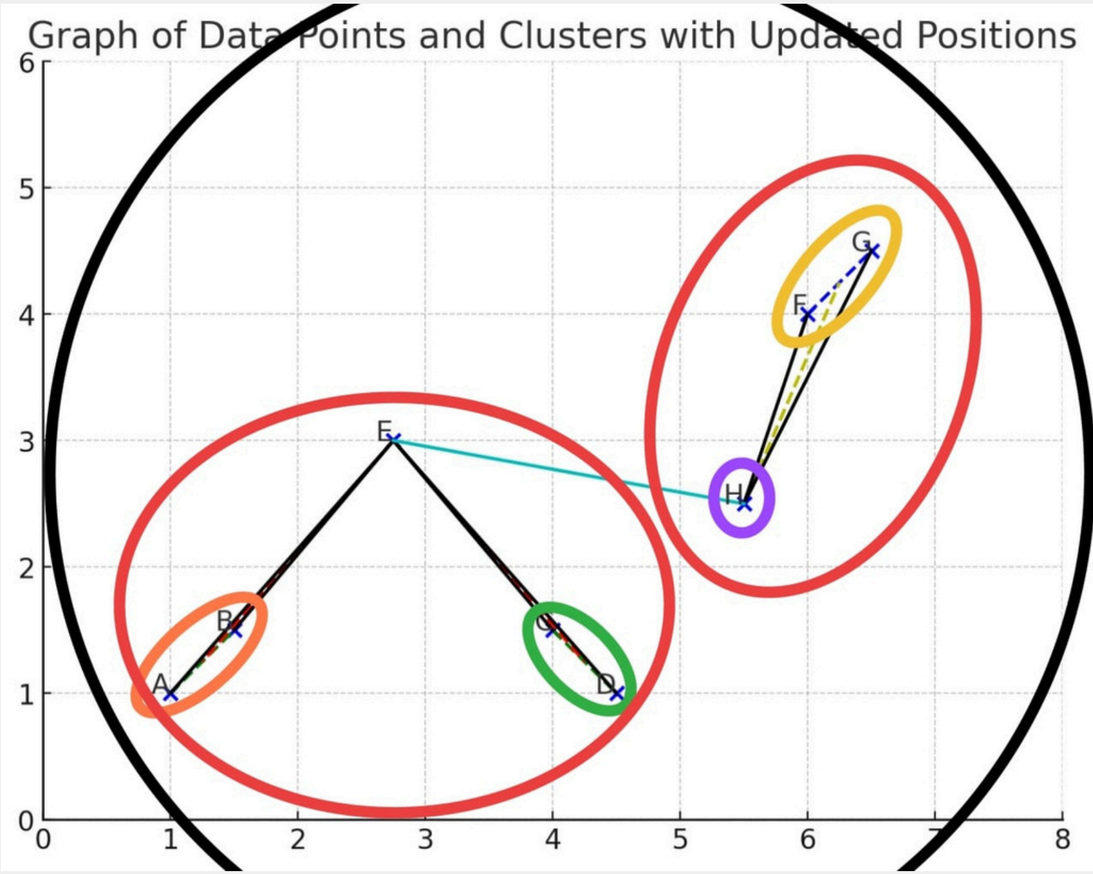
The K-Nearest Neighbors (KNN) algorithm is a popular machine learning technique used for classification and regression tasks. It relies on the idea that similar data points tend to have similar labels or values.



HIERARCHICAL CLUSTERING

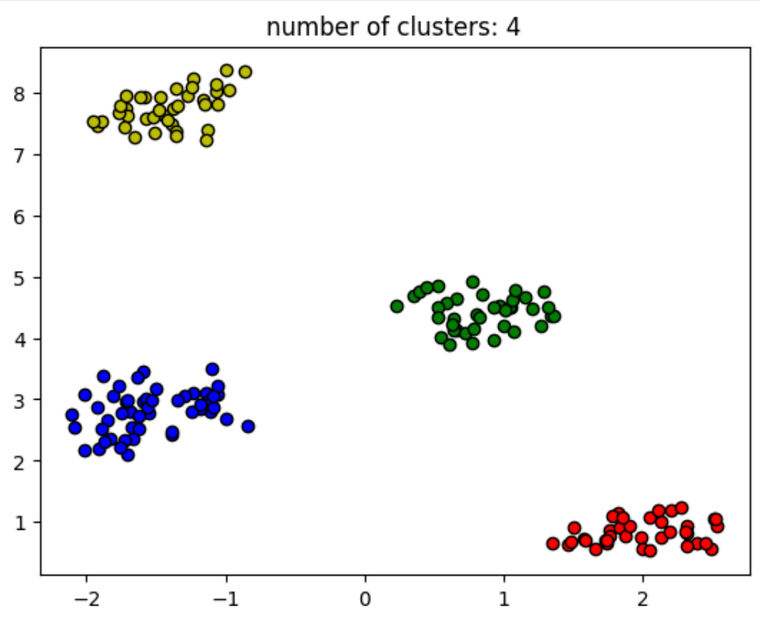
Hierarchical clustering is a clustering analysis method that constructs a hierarchy of clusters, commonly visualized with a dendrogram to illustrate the arrangement of the clusters.

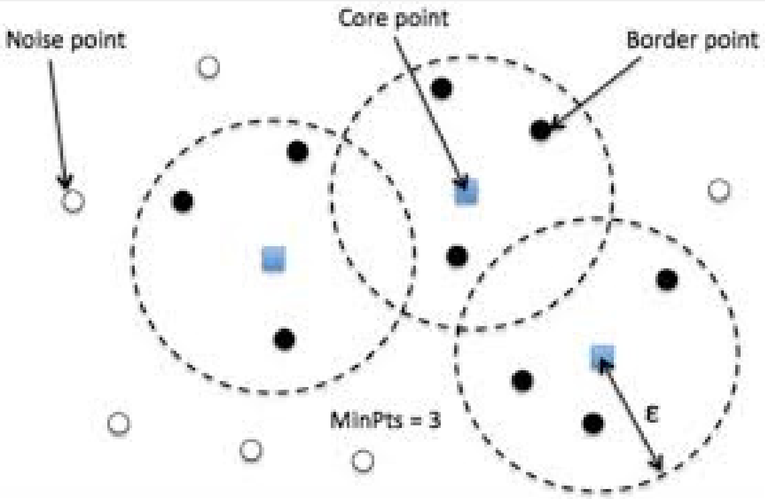

Cluster distance
a
b
c
d
e
f
g
h
i
j
k
l
m
n

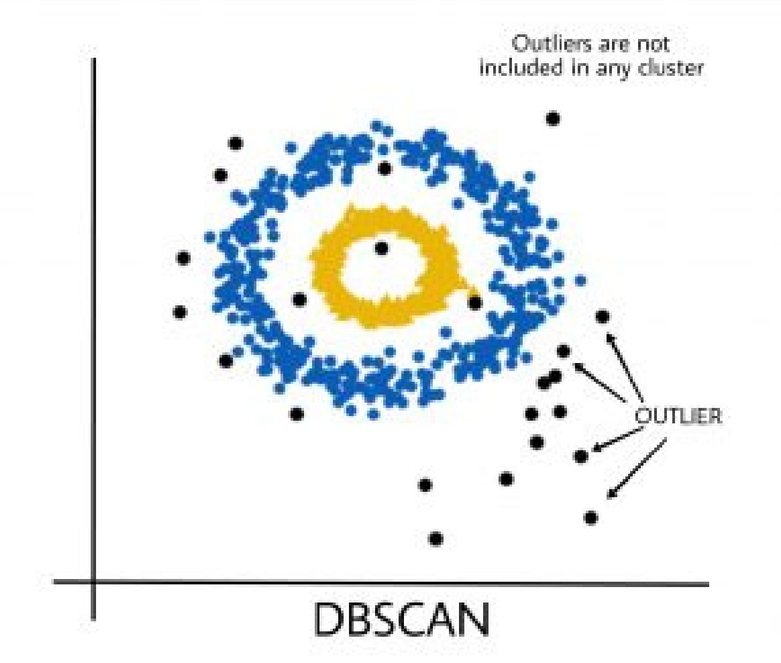

Graph of Data Points and Clusters with Updated Positions
A
B
C
D
E
F
G
H
0
1
2
3
4
5
6
7
8

DBSCAN CLUSTERING

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is a widely used clustering algorithm that detects clusters by analyzing the density of data points within a specific region. It effectively deals with noise and outliers in the data.







THANKYOU