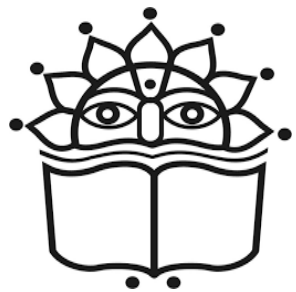
Customer Segmentation Using Machine Learning And Deep Learning

Ms. Vipashyana Jawale

Mr. Anish Date

Mr. Yashraj Devrat

Mr. Shubham Keskar



Guide : Dr. P .M. Paithane

Department of Artificial Intelligence & Data Science

Vidya Pratishthan's Kamalnayan Bajaj Institute of Engineering and Technology

Vidyanagari, Baramati-413133

27/05/2024

Contents

Introduction

- Customer segmentation is crucial for successful advertising campaigns and marketing new products.
- Recency , Frequency and Monetary Values of E-Commerce dataset.
- RFM Analysis : RFM analysis assigns scores to each customer based on these three factors and puts them into segments.
- Models for customer segmentation.

Problem Statement

Current customer segmentation methods result in suboptimal clustering and inefficient processing times, hindering effective marketing strategies and customer retention. To identify an efficient algorithm that accurately segments customers based on multiple behaviors, thereby improving marketing effectiveness and customer satisfaction.

Objectives

- To divide customers into homogeneous clusters based on their RFM values, identify distinct customer segments with specific characteristics and properties.
- Gain insights into customer preferences and needs and understand product development and innovation.
- Develop targeted acquisition strategies to attract new customers who align with existing segments.
- To compare the performance of traditional K-means clustering, SOM network, Gaussian Mixture Model, DBSCAN for customer segmentation.

System Architecture

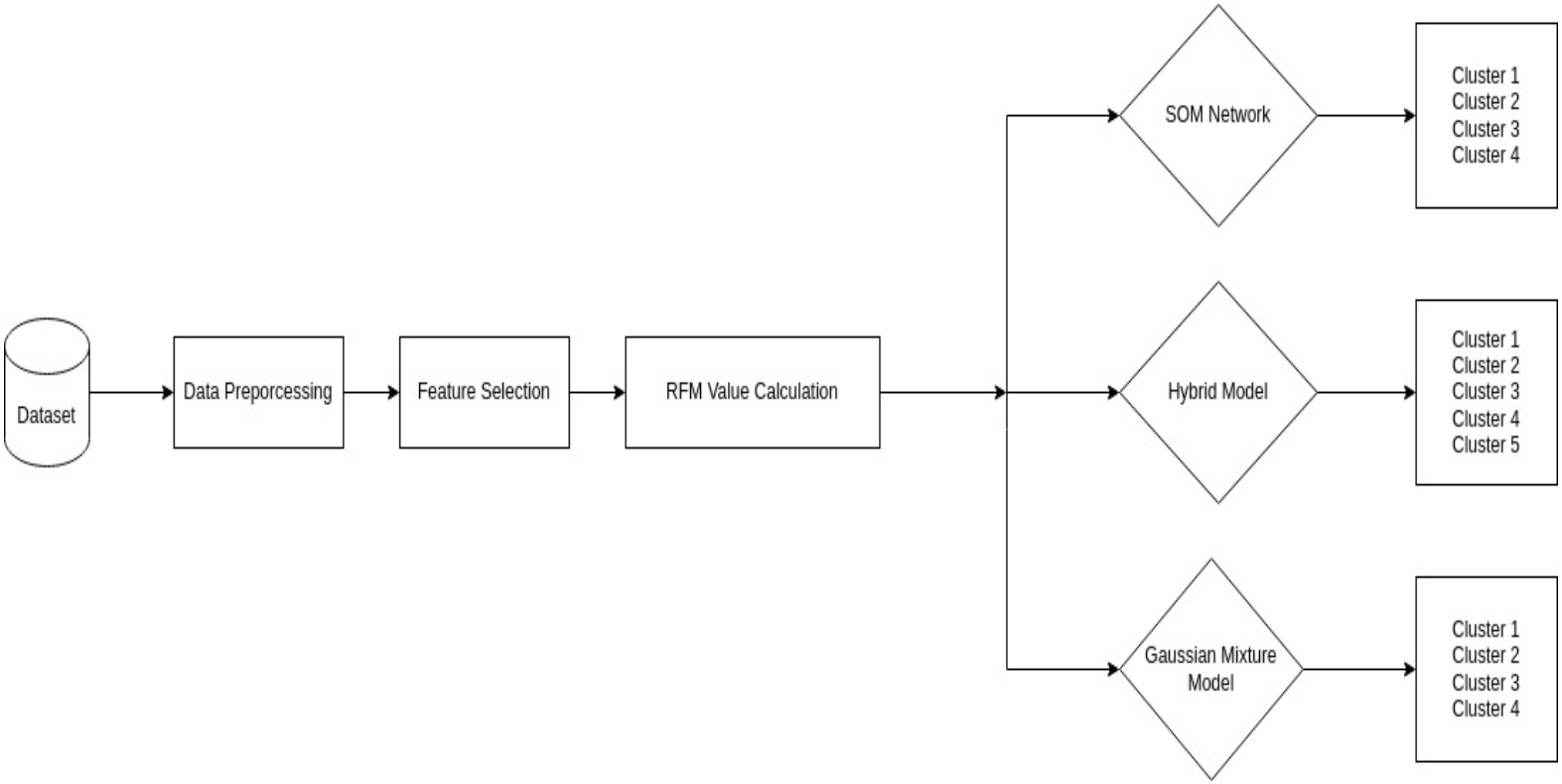


Figure: Architecture

Mathematical Modeling

Self-Organizing Maps (SOM) and K-means:

1. Self-Oragnizing Maps:

$$M(i,j) = \sum_{t=1}^{n} w_{ij}(t)x_t \tag{1}$$

2. K-means:

$$J = \sum_{i=1}^{k}\sum_{j=1}^{n_i} ||x_j^{(i)} - \mu_i||^2 \tag{2}$$

is the objective function to minimize.is the number of clusters.is the number of points in cluster i.is the j-th point in cluster i.is the centroid of cluster i.

Mathematical Modeling

Gaussian Mixture Model (GMM):

1. Probability Density Function

$$P(x) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x|\mu_k, \Sigma_k) \tag{3}$$

$P(x)$ is the probability density function for data point x. K is the number of Gaussian components. π_k is the mixing coefficient for component k. $\mathcal{N}(x \mid \mu_k, \Sigma_k)$ is the Gaussian distribution with mean and covariance Σ_k.

Mathematical Modeling

K-Means and DBSCAN:

1.KMeans Clustering (Objective Function):

$$J = \sum_{i=1}^{k} \sum_{j=1}^{n_i} ||x_j^{(i)} - \mu_i||^2 \tag{4}$$

2. DBSCAN (Core Points and Neighborhood):

$N_\epsilon(p) = \{q \in D \mid \text{dist}(p, q) \leq \epsilon\}$

$|N_\epsilon(p)| \geq MinPts$

Algorithm Development

Self-Organizing Maps (SOM) and K-means

- Normalization: Ensures equal contribution of all features. - Training: Projects high-dimensional data onto a 2D grid of neurons.

- Adjustment: Each data point matched with the closest neuron, adjusting weights to create a topological map.

$$M(i,j) = \sum_{t=1}^{n} w_{ij}(t)x_t$$

Algorithm Development

Gaussian Mixture Model (GMM)

- Initialization
- Model Fitting
- Cluster Assignment
- Expectation-Maximization (EM) Algorithm

KMeans and DBSCAN

- K-Means Clustering
- Cluster Assignment
- Core Points and Neighborhood

Implementation

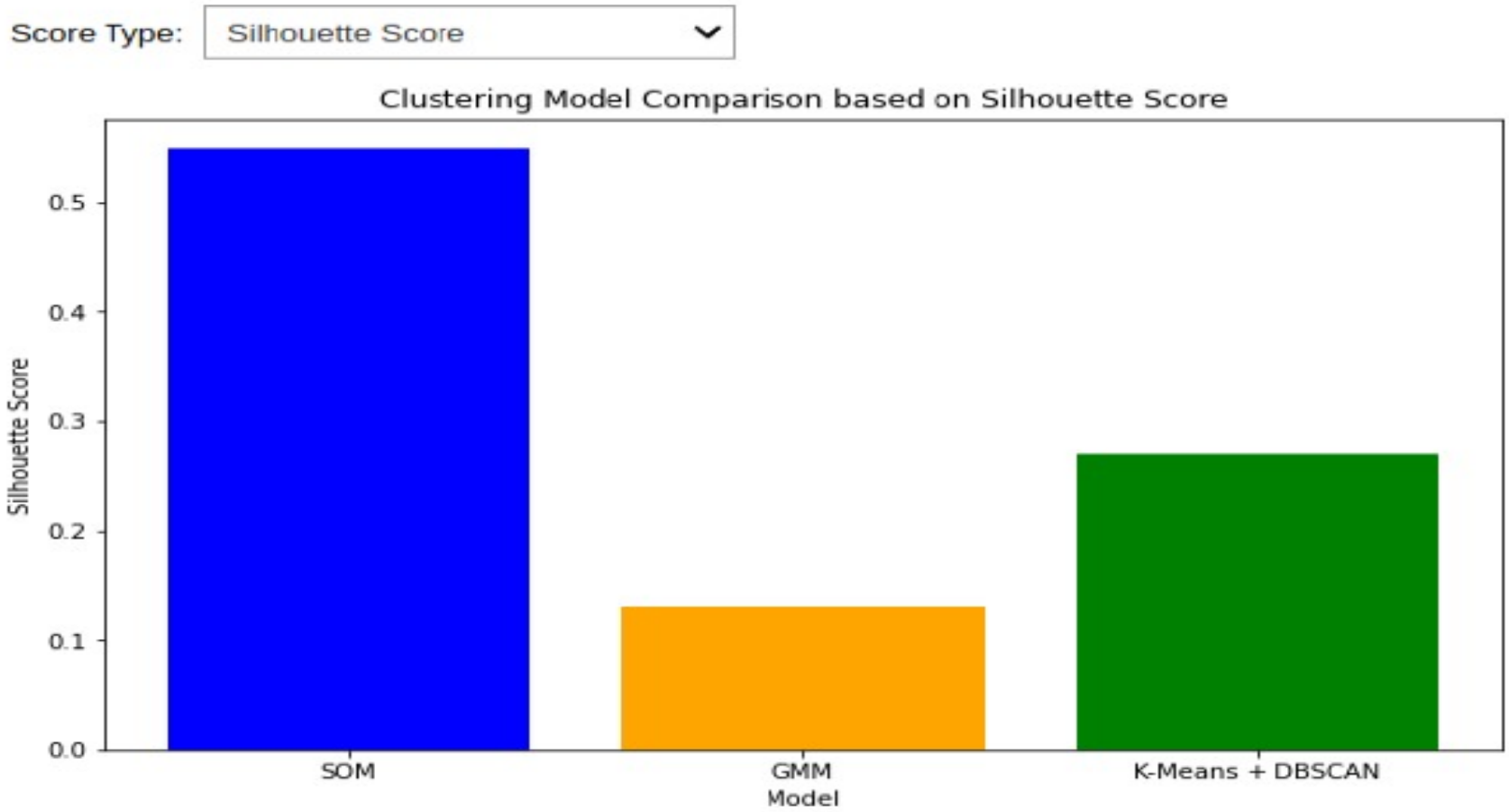
- Github link of code
  `https://github.com/yash9011/Final-Year-Project`

Results

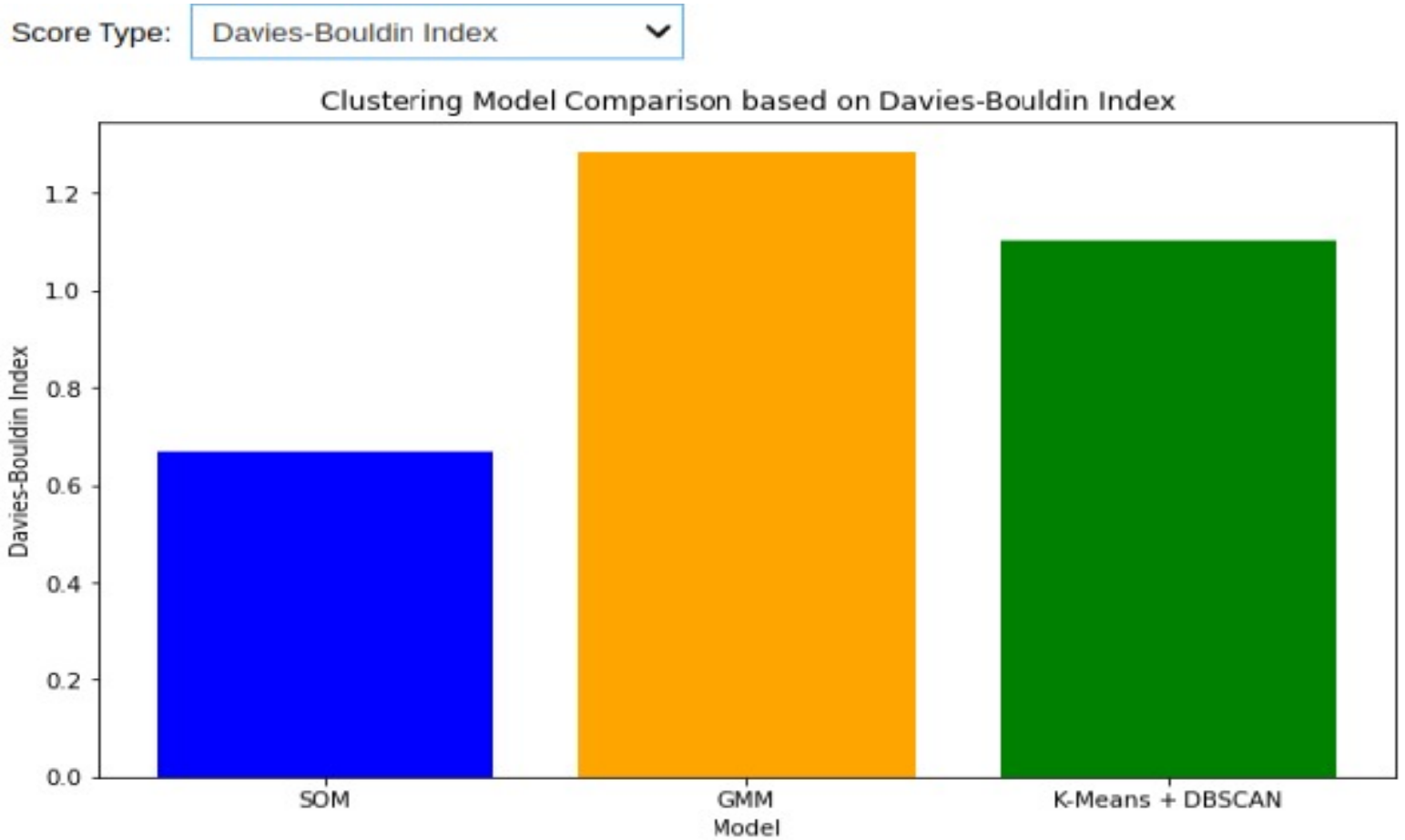
Methods	**SOM**	**GMM**	**K-means**
Silhouette	0.5	0.1300	0.2700
Davies Bouldins	0.6675	1.2830	1.1000
Calliniki Harabasz	12185.00	1148.48	665.28

Performace Evaluation

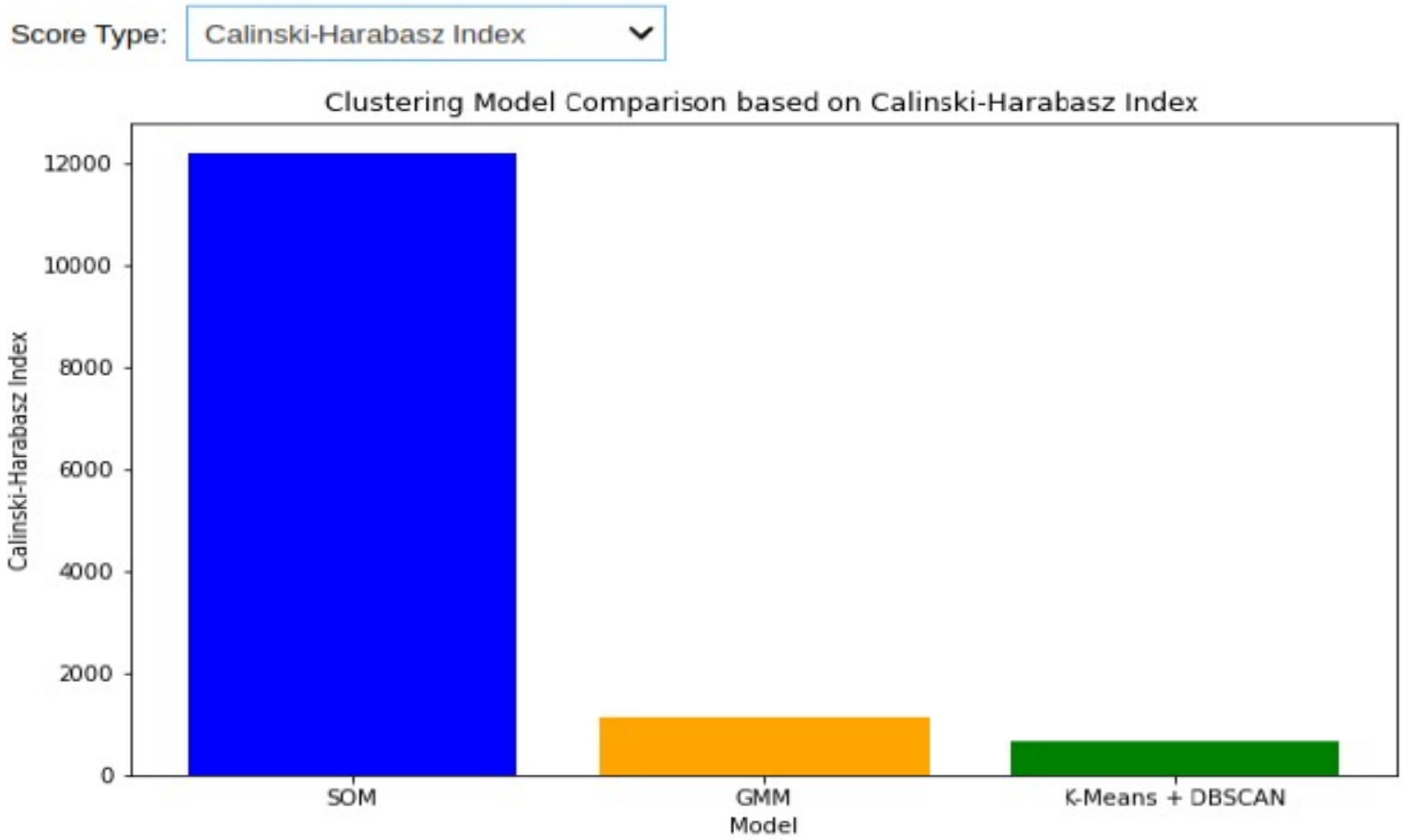
Visualisation:



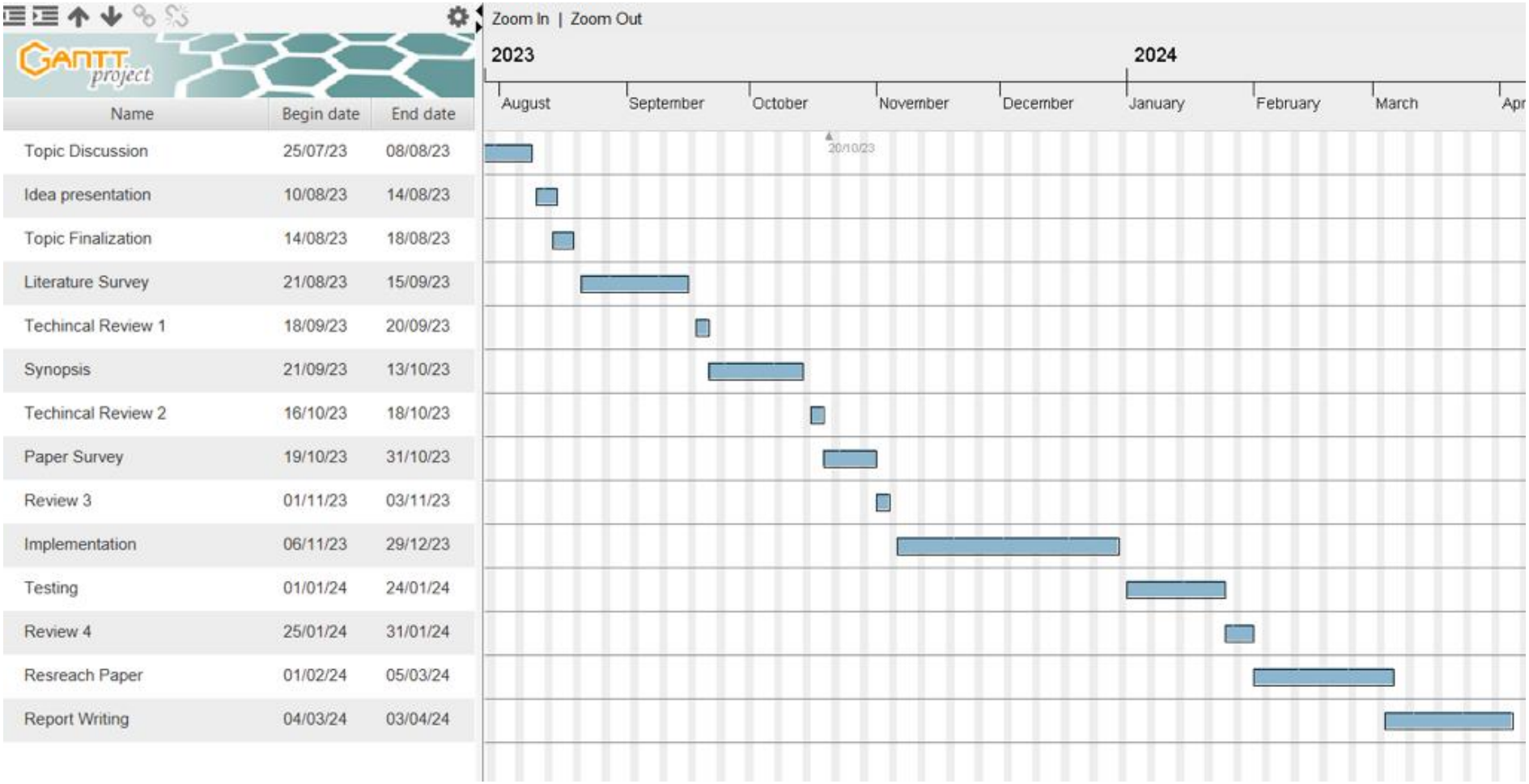
Visualisation:



Visualisation:



Project Planning



Name	Begin date	End date
Topic Discussion	25/07/23	08/08/23
Idea presentation	10/08/23	14/08/23
Topic Finalization	14/08/23	18/08/23
Literature Survey	21/08/23	15/09/23
Techincal Review 1	18/09/23	20/09/23
Synopsis	21/09/23	13/10/23
Techincal Review 2	16/10/23	18/10/23
Paper Survey	19/10/23	31/10/23
Review 3	01/11/23	03/11/23
Implementation	06/11/23	29/12/23
Testing	01/01/24	24/01/24
Review 4	25/01/24	31/01/24
Resreach Paper	01/02/24	05/03/24
Report Writing	04/03/24	03/04/24

Figure: Project Plan

Conclusion

- Customer segmentation is critical for businesses to leverage the power of data analysis for improved profitability. Various clustering techniques divides into data into categories of clusters and patterns gives idea for complementary marketing strategies. Gained valuable insights for effective marketing and contributed to meaningful customer segmentation .

References

- New RFM-D classification model for improving customer analysis and response prediction, Moulay Youssef SMAIL, Hanaa HACHIMI,2023.
- Customer Segmentation Using Fuzzy-AHP and RFM Mode, Anu Gupta Aggarwa, Sweta Yada, 2020.
- RFM ranking – An effective approach to customer segmentation, A. Joy Christy a, A. Umamakeswari a, L. Priyatharsini b, A. Neyaa 2018.
- Customer Segmentation with RFM Model using Fuzzy C-Means and Ge- netic Programming, Anas Syaifudin , Purwanto , Heribertus Himawan , M. Arief Soelema, 2023.
- Customer Segmentation Based on RFM Model Using K-Means, K-Medoids, and DBSCAN Methods, Rahma Wati Br Sembiring Berahmana, Fahd Agodzo Mohammed, Kankamol Chairuang, 2020.

References

- https://towardsdatascience.com/marketing-automation-customer-segmentation-5924b45556b5
- https://towardsdatascience.com/mastering-customer-segmentation-with-llm-3d9008235f41
- https://medium.com/@365datascience/build-your-own-customer-segmentation-model-in-python-af0b3d0d34713

Thank You