International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

CUSTOMER SEGMENTATION

T.Sridivya[1], M.Ramadevi[2], V. Aishwarya[3], K.Srilatha[4]

Department of Computer Science and Engineering, Stanley College of Engineering and Technology for Women, Telangana, India

Abstract

In today's fast moving world of marketing, the management of customer treatment can be seen as a key to achieve revenue growth and profitability. One of the big challenges in customer-based organizations is customer cognition, understanding the difference between them, and scoring them. Business strategies have to be according to the solutions to the above mentioned challenges. In this paper we proposed a Customer Segmentation, which is one of the key aspects of the business decision support system and we are using K-means clustering algorithm which is a partitioning algorithm, to segment the customers according to the similar characteristics to determine the optimal clusters.

We are currently living in a world where a vast amount of data is collected daily. Analyzing such data is an important need in the industry. Generally, companies may create three segments like High (Group who buys often, spends more, and visited the platform recently), Medium (Group which spends less than high group and is not that much frequent to visit the platform) and Low (Group which is on the verge of churning out). This is where Machine Learning provides a crucial solution, several algorithms are applied for revealing the hidden patterns in data for better decision making. Here these algorithms are used to segment the customers according to similar characteristics. To determine the optimal clusters.

Key Words : Revenue growth, Profitability, Segmentation, Hidden Patterns, Optimal Clusters

1. Introduction

1.1 About Project

In today's fast moving world of marketing, the management of customer treatment can be seen as a key to achieve revenue growth and profitability. One of the big challenges in customer-based organizations is customer cognition, understanding the difference between them, and scoring them. Business strategies have to be according to the solutions to the above mentioned challenges. In this paper we proposed a Customer Segmentation, which is one of the key aspects of business decision support system and we are using K-means clustering algorithm which is a partitioning algorithm, to segment the customers according to the similar characteristics to determine the optimal clusters.

1.2 Objectives of the Project

The main objective of market segmentation is to determine the differences among customers. The marketer can frame his marketing policies and strategies on the basis of these differences to provide maximum satisfaction to customers.

- To identify the needs, tastes, priorities, buying- motives of the target consumers.
- To determine marketing strategies, targets and goals of the firm.
- To make the activities of the firm consumer- oriented.
- To identify areas where the customers may be created and market area can be explained.

1.3 Scope

Customer segmentation is the process of separating your **customers** into groups based on the certain traits (e.g.personality, interests, habits) and factors (e.g. demographics, industry, income) they share. **Segmentation** offers a simple way of organizing and managing your company's relationships with your customers.

1.4 Advantages

- Demographic variables are typically easier to collect and measure versus those of other segmentation techniques.
- Targeting is typically more straightforward when using demographics as a metric — e.g., you can target a consumer group such as college-educated millennials, or men between the ages of 35 and 45.
- Consumer Profiles easy to understand across the board, which lends itself to an easier strategy development process in terms of internal collaboration between departments (e.g., sales, customer service, management, etc.).

1.5 Disadvantages

- No-optimal set of clusters: K-means doesn't allow development of an optimal set of clusters and for effective results, you should decide on the clusters before.
- Lacks consistency: K-means clustering gives varying results on different runs of an algorithm. A random choice of cluster patterns yields different clustering results resulting in inconsistency.
- Order of values: The way in which data is ordered in building the algorithm affects the final results of the data set.
- Sensitivity to scale: Changing or rescaling the dataset either through normalization or standardization will completely change the final results.

1.6 Applications

Any Campaign with budget constraints requires customer segmentation. Segmentation is used across the domains and process to increase the ROI and effectiveness of a campaign. Right from the Customer Acquisition to cross sell to long term Retention till the collection and Recovery processes, Segmentation is used throughout the customer relationship life cycle. Some business domains that use customer segmentation extensively are Banking, Insurance, Retail and Telecom. Some of the processes where segmentation is applied are-

- Customer Acquisition: Segmentation is done to identify the profile of customers and design acquisition campaigns accordingly.

- Customer Relationship Management: Banks use customer segmentation techniques combined with predictive model score outputs to identify the (Customer Lifetime Value).

- Collection and Recovery Processes: Customers are segmented based on the propensity/likelihood of default on the next payment, or the clearing of all dues (once in collections). Different segments are exposed to different collection treatment.

1.7 Requirements

1.Software Requirements:

- Python
- MATLAB
- Scikit-Learn
- Matplotlib
- NumPy
- SciPy

2.Hardware Requirements

- Hard disk
- Ram
- Keyboard
- Mouse
-

Literature Survey

2.1 Existing System

K-Means Clustering Algorithm

1. Select initial k centroids
2. Assign data points to nearest centroid
3. Compute new centroids
4. Repeat the previous two steps till convergence

Convergence: No change in the position of centroids from ith iteration to i+1th iteration.

Drawbacks of Existing System

- Similar demographics do not always imply similar needs, values or motivations within a particular demographic group.
- The effectiveness of your marketing message may be hampered by a "one-sizefits-all" approach to the consumers within a given demographic segment.
- Skewed or problematic demographic data within a given region can produce unreliable assumptions, potentially reducing the accuracy of your marketing message and methods.

2.2 Proposed System

- BIRCH Algorithm
- DB-SCAN Algorithm
- MEAN Shift Algorithm

Birch Algorithm

- Multi-level Clustering
- Works for large datasets

DB-Scan Algorithm

- Clustering does not need the total number or amount of clusters to be specified priorly.
- It works great with outliers, anomalies and observations that stand out.

- It is very effective when working with clusters of varying shapes or arbitrary shapes. Mean shift Algorithm
- It is widely used to solve computer vision, where it is used for image segmentation.
- Clustering of data points in real-time without mentioning the number of clusters.
- Performs well on image segmentation and Video tracking

3.Proposed Architecture

Knowledge on consumer habits is essential for companies to keep customers satisfied and to provide them personalised services. We present a data mining architecture based on clustering techniques to help experts to segment customer based on their purchase behaviours. In this architecture, diverse segmentation models are automatically generated and evaluated with multiple quality measures. Some of these models were selected for given quality scores. Finally, the segments are compared.

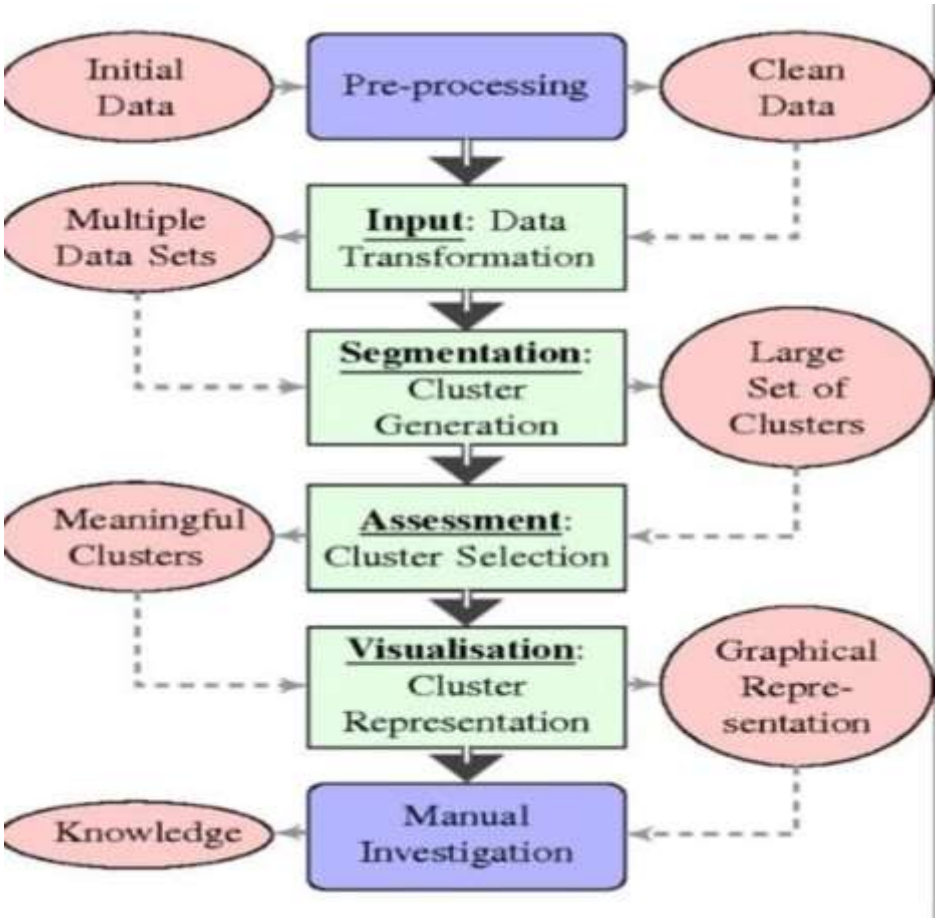

Fig: 3.1 Architecture

4.IMPLEMENTATION

4.1 Algorithm

Step 1: Importing the required libraries

Step 2: Loading the data

Step 3: Preprocessing the data

Step 4: Reducing the dimensionality of the data to make it visualizable

Step 5: Building the clustering model

Step 6: Visualizing the clustering

4.2 Code implementation

```
import numpy as np import pandas as pd import
matplotlib.pyplot as plt import seaborn as sns from
sklearn.cluster import KMeans


# loading the data from .csv file to a pandas dataframe customer_data =
pd.read_csv('/content/Mall_Customers (1).csv')
#printing first 5 rows in the data set customer_data.head()
customer_data.shape
#getting scripted statistics of data frame
customer_data.describe() #getting information about
dataset customer_data.info()
customer_data.isnull().sum() #for checking missing
values customer_data.drop(["CustomerID"],axis=1,
inplace=True) customer_data.head() plt.figure(1,
figsize=(15,6)) n=0 for x in ['Age' , 'Annual Income
(k$)' , 'Spending Score (1-100)']:
 n += 1 plt.subplot(1 , 3 , n) plt.subplots_adjust(hspace =0.5 ,
wspace= 0.5) sns.distplot(customer_data[x] , bins = 20)
plt.title('Distplot of {}' . format(x)) plt.show()

#WCSS
```

```
X1 =customer_data.loc[:, ["Age","Spending Score (1-100)"]].values wcss = [] for k
in range(1,11):              kmeans = KMeans(n_clusters=k, init="k-means++")
kmeans.fit(X1)

   wcss.append(kmeans.inertia_)
plt.figure(figsize=(12,6)) plt.grid()
plt.plot(range(1,11),wcss, linewidth=2, color="red", marker="8") plt.xlabel("k
values") plt.ylabel("WCSS") plt.show()

kmeans = KMeans(n_clusters=4) label =
kmeans.fit_predict(X1) print(label)


#K-MEANS CLUSTERING
# plotting all the clusters and their centroids plt.figure(figsize=(8,8))
plt.scatter(X1[Y==0,0], X1[Y==0,1], s=50, c='green', label='Cluster 1')
plt.scatter(X1[Y==1,0], X1[Y==1,1], s=50, c='red', label='Cluster 2')
plt.scatter(X1[Y==2,0], X1[Y==2,1], s=50, c='yellow', label='Cluster 3')
plt.scatter(X1[Y==3,0], X1[Y==3,1], s=50, c='violet', label='Cluster 4')
plt.scatter(X1[Y==4,0], X1[Y==4,1], s=50, c='blue', label='Cluster 5')

#plot the centroids plt.scatter(kmeans.cluster_centers_[:,0], kmeans.cluster_centers_[:,1], s=100,
c='cyan', label='centroids' ) plt.title("customer groups") plt.xlabel("annual income")
plt.ylabel("spending score") plt.show()

#BIRCH ALGORITHM

import matplotlib.pyplot as plt from
sklearn.cluster import Birch

# Creating the BIRCH clustering model
model = Birch(branching_factor =30, n_clusters = 5, threshold = 2.5)

# Fit the data (Training) model.fit(X)

# Predict the same data pred =
model.predict(X) # Creating a scatter plot
plt.scatter(dataset["A"], dataset["B"], c =
```

```
pred, cmap = 'rainbow', alpha = 1,
edgecolors = 'b')  plt.show()
```

#DBSCAN ALGORITHM

```
y_means = db.fit_predict(x) plt.figure(figsize=(7,5)) plt.scatter(x[y_means == 0, 0], x[y_means == 0, 1], s = 50, c = 'pink') plt.scatter(x[y_means == 1, 0], x[y_means == 1, 1], s = 50, c = 'yellow') plt.scatter(x[y_means == 2, 0], x[y_means == 2, 1], s = 50, c = 'cyan') plt.scatter(x[y_means == 3, 0], x[y_means == 3, 1], s = 50, c = 'magenta') plt.scatter(x[y_means == 4, 0], x[y_means == 4, 1], s = 50, c = 'orange') plt.scatter(x[y_means == 5, 0], x[y_means == 5, 1], s = 50, c = 'blue') plt.scatter(x[y_means == 6, 0], x[y_means == 6, 1], s = 50, c = 'red') plt.scatter(x[y_means == 7, 0], x[y_means == 7, 1], s = 50, c = 'black') plt.scatter(x[y_means == 8, 0], x[y_means == 8, 1], s = 50, c = 'violet') plt.xlabel('Annual Income in (1k)') plt.ylabel('Spending Score from 1-100') plt.title('Clusters of data') plt.show()
```

#MEANSHIFT ALGORITHM

```
features = ['Age','Spending Score (1-100)']
X = df[features]

bandwidth = estimate_bandwidth(X,quantile=0.20)
meanshift = MeanShift(bandwidth=bandwidth,bin_seeding=True) meanshift.fit(X)

labels = meanshift.labels_

X['labels'] = labels
sb.scatterplot(x='Age',y='Spending Score (1-100)',data =
X,hue='labels',palette="Set1")
```

5.Results

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

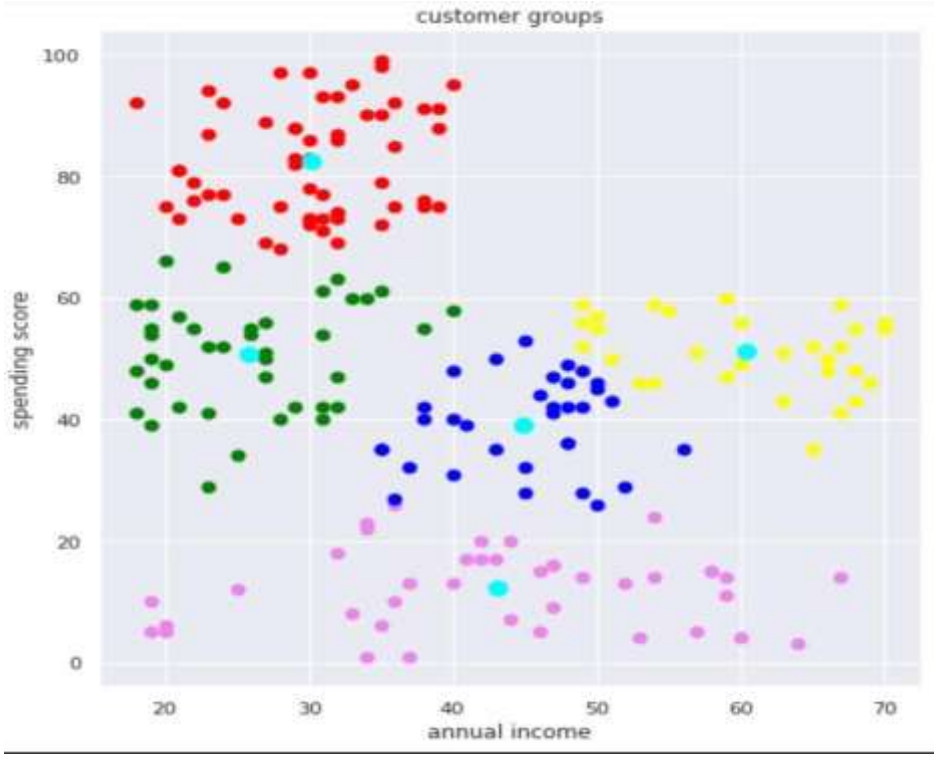

Fig: 5.1 K-Means Clustering Algorithm

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

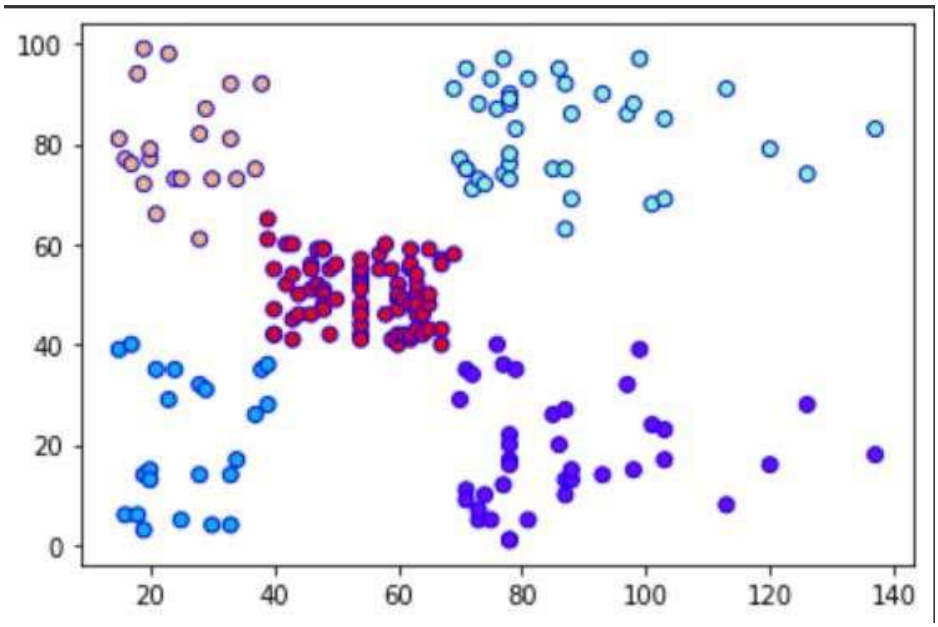

Fig: 5.2 Birch Algorithm

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

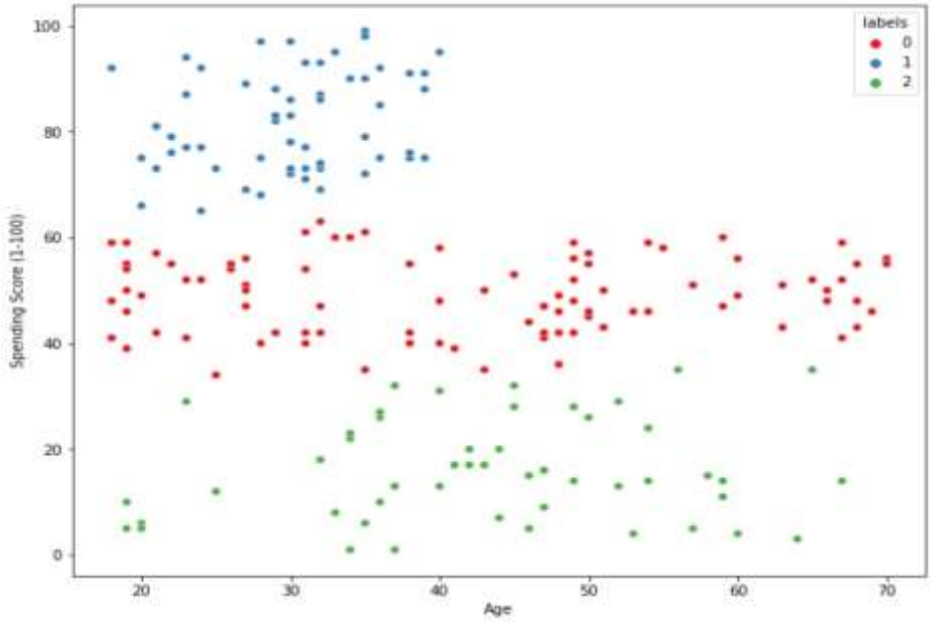

Fig: 5.3 Mean Shift Algorithm

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

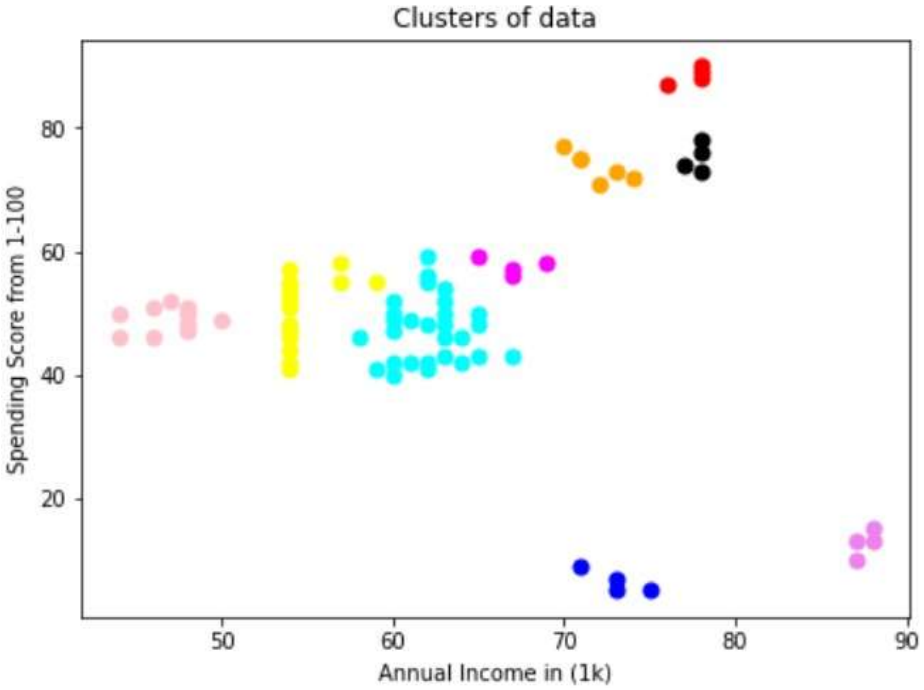

Fig:5.4 DBSCAN Algorithm

6.Conclusion

The process of customer segmentation ensures that your brand is customer-centric and helps you serve them better. It boosts conversions, brings your marketing efforts to fruition, and also helps build everlasting customer relationships. The strategies discussed here will help you organize your segments, but after you have them in place, continue to monitor and make sure your product is still valuable to the groups. The key to successful customer segmentation is the constant research it entails to ensure your brand and product stay relevant and indispensable.

7.Future Scope

The premise of customer segmentation is that to maximize sales to a large population of customers, it is best to divide it into logical subgroups. The assumption is that by dividing one large, amorphous mass into subgroups, you can fine-tune your product, messaging, support, or distribution channels to meet the specific needs of unique customer groups. Thus, the goal is to use a customer segmentation model to improve marketing success and optimize marketing ROI.

Segmentation models vary from basic to complex, and the approaches to developing and applying them is a topic for an entire book itself. But here are some examples:

- Example 1: A telecommunications company selling mobile phone services might segment its market based on complexity of needs. One customer group might only need voice service, and very little volume at that. Another group might primarily use mobile texting services. Yet another might be a heavy user of mobile phones for voice, text, email, and web browsing. By identifying distinct patterns in customer needs, the company can optimize product bundles and target them at the correct audiences.

- Example 2: A hotel chain that caters to families might simply segment its market based on income level and travel frequency. The chain might find that a group of moderate income-frequent travelers exists that is swayed by certain loyalty program rewards. It might find another important group exists—that is more swayed by on-site amenities. With this information, the hotel can optimize its offerings and loyalty programs to appeal to each group's unique needs.

8.References

1. T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y.
2. Wu, “An efficient K-means clustering algorithm,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, pp. 881-892, 2002.
3. MacKay and David, “An Example Inference Task: Clustering,” Information Theory, Inference and Learning Algorithms, Cambridge University Press, pp. 284-292, 2003.
4. N. Made, A. Santika, I. K. Gede, D. Putra, and I. M. Sukarsa, “Implementasi Metode
5. Clustering DBSCAN pada Proses Pengambilan Keputusan,” Lontar Komputer, vol. 6, no. 3, pp. 185–191, 2015.
6. P. Pengelompokan, R. Kost, D. I. Kelurahan, and T. Semarang, “Perbandingan Metode K-Means dan Metode DBSCAN pada Pengelompokan Rumah Kost Mahasiswa di Kelurahan Tembalang Semarang,” Jurnal Gaussian, vol. 5, pp. 757–762, 2016.
7. D. Mining and K. Discovery, “BIRCH : A New Data Clustering Algorithm and Its Applications” vol. 182, pp. 141-182, 1997.
8. H. Ziafat and M. Shakeri, “Using Data Mining Techniques in Customer Segmentation,” vol. 4, no. 9, pp. 70–79, 2014
9. I. V Anikin, "Privacy Preserving DBSCAN Clustering Algorithm for Vertically Partitioned Data in Distributed Systems," International Siberian Conference on Control and Communications, vol. 10, pp.1-4, 2017.

10. Rui Xu and D. Wunsch, "Survey of clustering algorithms", Neural Networks IEEE Transactions on, vol. 16, no. 3, pp. 645-678, 2005.

11. Girish Punj and David W. Stewart, "Cluster analysis in marketing research: Review and suggestions for application", Journal of Marketing Research, vol. 20, no. 2, pp. 134-
12. 148, 1983
13. M. McDonald, "The role of marketing in creating customer value", Marketing from an Engineering Perspective (Digest No. 1996/172), pp. 1/1-111, Nov 1996.