International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8
***Article Received:** 25 June 2023 **Revised:** 12 July 2023 **Accepted:** 30 August 2023*

A Preference Compositional Approach for Client Structured Web Customer Segmentation Using Machine Learning Techniques

Adlin Selva Golda. V[1] , Narayani. V[2],
[1]Research Scholar, Manonmaniam Sundaranar University, Tirunelveli,
Email: adlingolda@gmail.com
[2]Assistant Professor, Department of/Computer Science, St Xavier's College, Tirunelveli

Abstract: The web information system develops in an exponential growth in which the data classification and segmentation are tedious process to handle in an effective way. The process of handling vast amount of web information with the target of segmented grouping entirely depends on the nature of the data along with the approach of segmentation. The existing customer segmentation methods lacks in the areas of scale, modification and verification. The main issues of redundancy, incorrect and irrelevant data plays its substantial role in degrading the performance of segmentation methodology. This research article proposes a machine learning approach for handling client structured web customer segmentation with the preference compositional process based on their requirements of online web requests and responses. In near future this research article leads the path for the incorporation of artificial intelligence based customer segmentation in web information system.

Keywords: *Machine learning, web data, segmentation, information system, customer data*

I. INTRODUCTION

Segmentation:

Segmentation means to divide the marketplace into parts, or segments, which are definable, accessible, actionable, and profitable and have a growth potential. In other words, a company would find it impossible to target the entire market, because of time, cost, and effort restrictions [1]. It needs to have a 'definable' segment - a mass of people who can be identified and targeted with reasonable effort, cost and time [2].

Customer Segmentation:

Customer segmentation is the process of dividing a customer base into distinct groups of individuals that have similar characteristics [3]. This process makes it easier to target specific groups of customers with tailored products, services, and marketing strategies [4]. By segmenting customers into different classes, businesses can better understand their needs, preferences, and buying patterns, allowing them to create more personalized and effective marketing campaigns.

Machine Learning:

Machine Learning is the field of study that gives computers the capability to learn without being explicitly programmed [5]. ML is one of the most exciting technologies that one would have ever come across. As it is evident from the name, it gives the computer that makes it more similar to humans: The ability to learn [6].

Web information System:

Web information system, or web-based information system, is an information system that uses Internet web technologies to deliver information and services, to users or other information systems/applications.

II. METHODOLOGY

The proposed methodology comprises 3 levels of implementation. They are

a. Corrective motion -Issues removal

1. Removal of redundancy

The hash method approach is used to remove the redundant data in the structured web customer segmentation process. Customer information is converted with hash function for matrix representation. The process of comparing the customer data using hash values are time efficient and produces more effective outputs. The deleted redundant data improves the performance in customer segmentation.

2. Removal of incorrect data

The removal of incorrect data includes the process of identifying the source and recognizes the type of error.

The authentication of source data can be done through proper validation on sensitive customer financial information system like PAN, Aadhar etc.

3. Remove irrelevant data

The removal of irrelevant data focuses on structural errors removal and removing the unwanted components in the customer information record system

b. Improvisation-Dealing with existing methods slackness

The process of implementing customer relationship management system improves the scaling of customer segmentation in web information system.

The modification and validation of customer data in a spontaneous updating approach through MySQL, Mongo DB, or Firebase eliminates the frequency changing of data content values.

c. Preference Compositional approach-Optimal approach selection

The proposed preference compositional approach consists of 4 stages.

They are

1. Composition of Customer x Product mapping

2. Composition of Customer Segmentation categorization.

3. Composition of Customer segmentation means.

4. Composition of Customer segmentation preference approach

- ✓ K-Means Clustering
- ✓ Agglomerative Hierarchical Clustering
- ✓ Expectation-Maximization (EM) Clustering
- ✓ Density-Based Spatial Clustering
- ✓ Mean-Shift Clustering

The proposed methodology of preference compositional approach for client structured web customer segmentation using machine learning techniques is as follows in Fig-1.

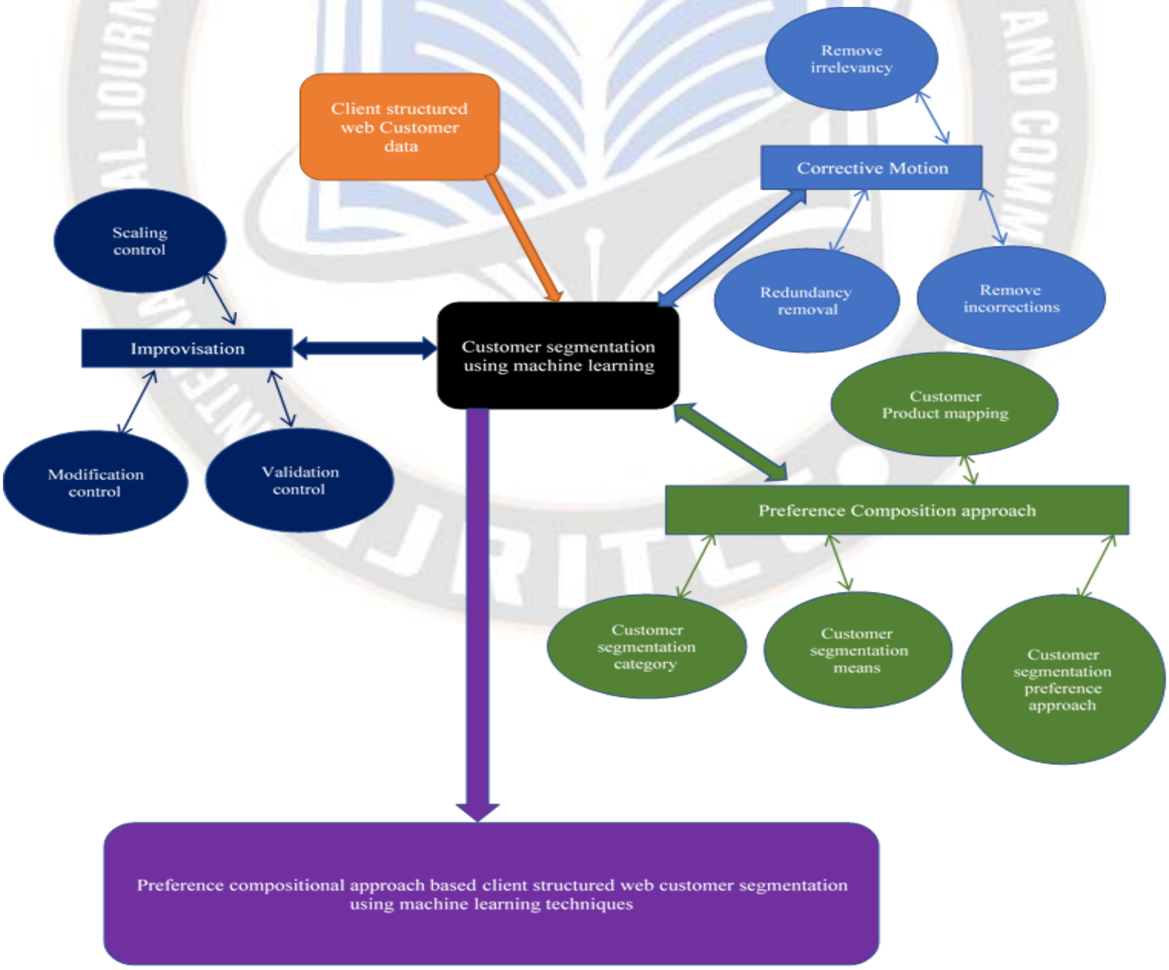


Fig-1: Proposed preference compositional approach for client structured web customer segmentation

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8
***Article Received:** 25 June 2023 **Revised:** 12 July 2023 **Accepted:** 30 August 2023*

The flow chart for the preference compositional approach for client structured web customer segmentation using machine learning techniques is as follows,

Start

Input: Customer data from real time / standard data set

Step-1: Corrective motion -Issues removal using machine learning

a. Removal of redundancy

b. Removal of incorrect data

c. Remove irrelevant data

Step-2: Improvisation using machine learning

a. Scaling control

b. Modification control

c. Validation control

Step-3: Preference Compositional approach using machine learning

1. Composition of Customer x Product mapping

2. Composition of Customer Segmentation categorization.

3. Composition of Customer segmentation means.

4. Composition of Customer segmentation preference approach.

Step-4: Check the general criteria

If Number of valid customer segmentation > 1&

segment is large enough &

forecast future demand &

serve the target audience &

segment's unique characteristics>1 then

Goto step-5;

Else

Goto step-1;

End if

Step-5: Display Customer segmentation

End

III. IMPLEMENTATION

a. Corrective motion -Issues removal

It represents the removal of data error issues in the web information system for customer segmentation.

1. Removal of redundancy

- ❖ *Select non-empty bucket.*
- ❖ *Apply hash for each block if its unique adds it to the bucket else removes.*
- ❖ *Use hash as a key for comparison.*
- ❖ *Add until all the blocks are checked.*

2. Removal of incorrect data:

It deals with the removal of incorrect information issues in the web client structured data for customer segmentation. It includes the following component checking,

- ✓ *Segment entry state,*
- ✓ *Segment collection process,*
- ✓ *Segment integration operation,*
- ✓ *Segment transmission function, and*
- ✓ *Segment storage facility.*

3. Remove irrelevant data

This part focuses on the removal of data which are not needed for processing in customer segmentation. It include the following operations,

- ✓ *Structural errors removal.*
- ✓ *Unwanted outliers identification and removal.*
- ✓ *Missing data handling.*

b. Improvisation-Dealing with existing methods slackness

The tools used for improving the customer segmentation in web information system for effective handling of data scaling are Sales force[8],Hub spot[9] and zoho CRM[10] as in Fig-2,Fig-3 and Fig-4.These tools play the vital role in handling huge amount of client structured customer segmentation from web information system.

1. Sales Force tool

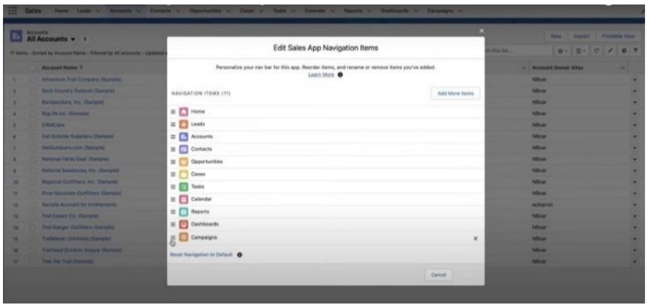


Fig-2: Sales Force CRM tool page

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8
***Article Received:** 25 June 2023 **Revised:** 12 July 2023 **Accepted:** 30 August 2023*

---

2. Hub spot tool:

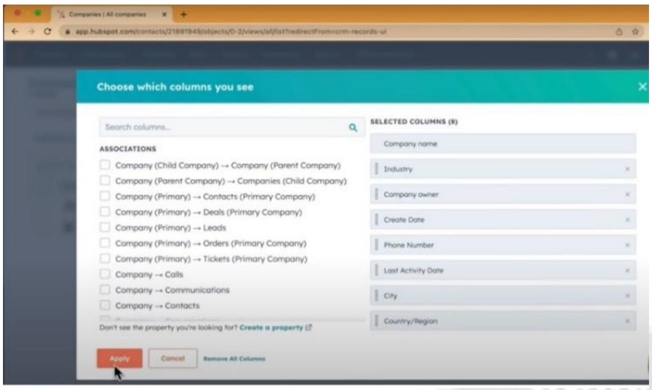


Fig-3: Hub spot CRM tool page

3. Zoho CRM

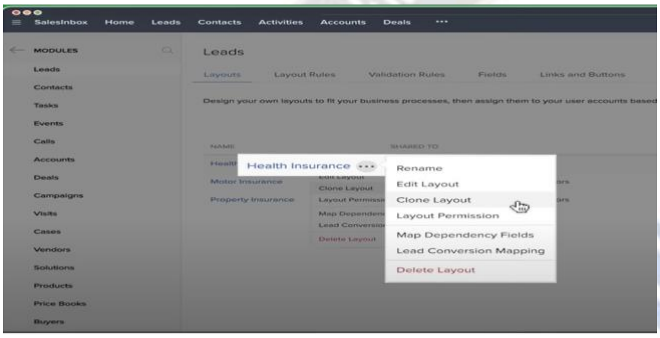


Fig-4: Zoho CRM tool page

c. Preference Compositional approach-Optimal approach selection

The proposed preference compositional approach consists of 4 stages. They are

1. Composition of Customer x Product mapping**-** Customers mapped with specific product based on their frequency in their preferences of the particular product. Generally focuses on the customer matching to the situation scheme.

2. Composition of Customer Segmentation categorization**.**

i. Geographic - customer segmentation based on customer location.

ii. Demographic – customer segmentation based on customer's size structure and density on numbers.

iii. Psychological- Customer segmentation based on customer belief and attitudes.

iv. Customer behavior pattern-Customer segmentation based on past activity trained data used for future predictions.

3. Composition of Customer segmentation means

The implementation is done through the following ways.

i. Survey

Google forms or surveys [11] are used to perform survey as in Fig-5 in the process of customer segmentation. The data retrieved form the survey acts as the machine learning based training for customer segmentation in fast and accuracy.

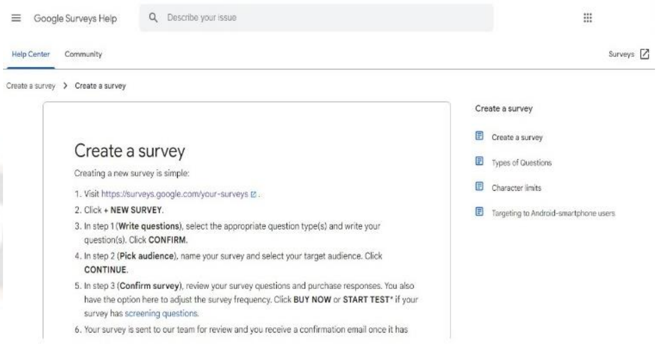


Fig-5: Google survey tool page

ii. Focus groups

Focus group represents the face to face meeting with the customer group in order to improve the customer segmentation process. Zoom [12] and Google meet supports the focus groups customer segmentation process as in Fig-6.

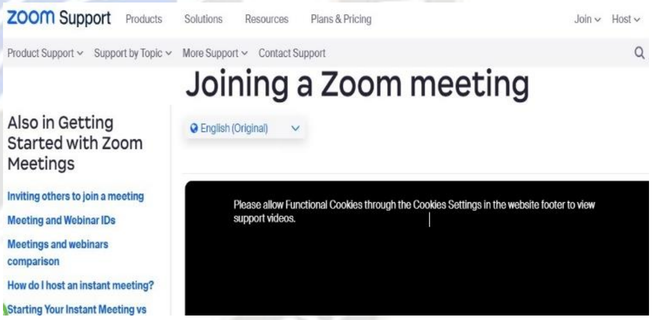


Fig-6: Zoom meeting tool page

iii. Polls

The tools such as Questionpro, Mentimeter [13], slido, poll everywhere, majency, directpoll and vivox are used to perform customer segmentation support polls as in Fig-7.

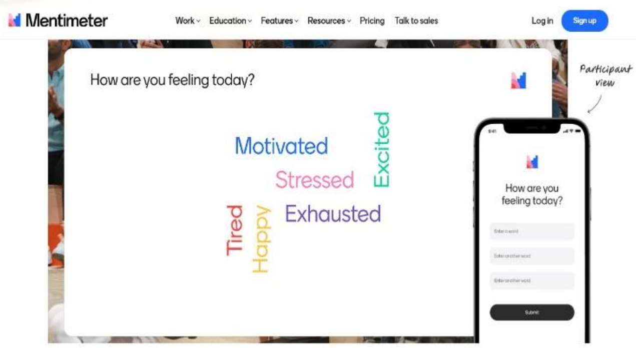


Fig-7: Mentimeter tool page

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8
***Article Received:** 25 June 2023 **Revised:** 12 July 2023 **Accepted:** 30 August 2023*

4. Composition of Customer segmentation preference approach

The preference based compositions describe the choice for machine learning clustering in customer segmentation approach. The structure is represented by the following select case scheme.

Select Case (Customer segmentation preference)

{

Case: - customer data are partially observable

Apply Expectation maximization clustering;

Break;

Case: - customer data are majority based

Apply Mean shift clustering;

Break;

Case: - Number of customer groups is known

Apply K-means clustering;

Break;

Case: - Number of customer groups is unknown

Apply Agglomerative Hierarchical Clustering;

Break;

}

IV. RESULTS AND DISCUSSION

Consider the customer data collection from Kaggle standard data set through Kaggle web domain [7]. The mall customer data sheet is used for the implementation process for this research article is shown in Fig-8 through Fig-15.

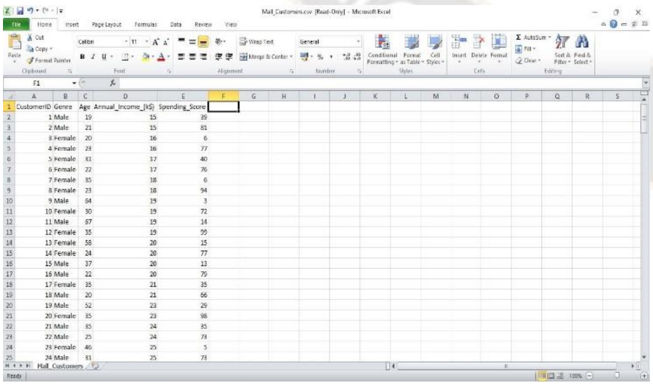


Fig-8: Mall Visitor data collection sheet-1 for customer segmentation

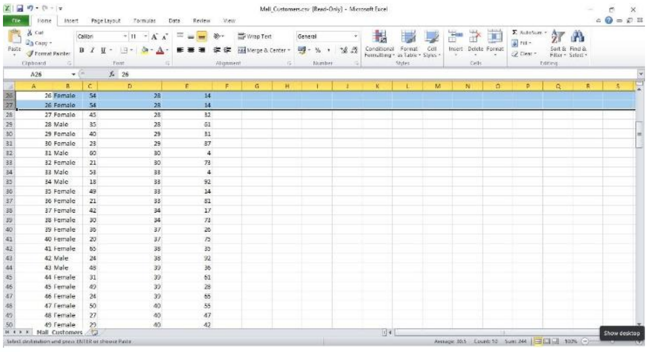


Fig-9: Mall Visitor data collection sheet-2 for customer segmentation

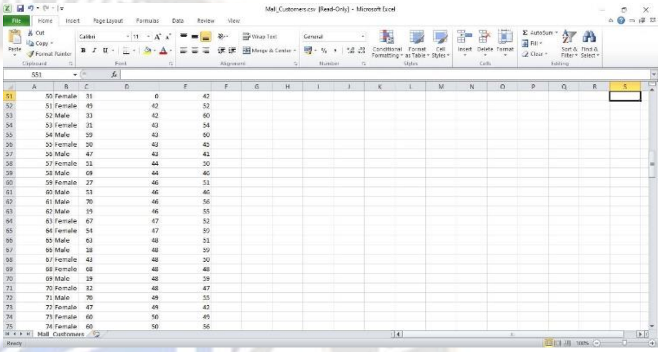


Fig-10: Mall Visitor data collection sheet-3 for customer segmentation

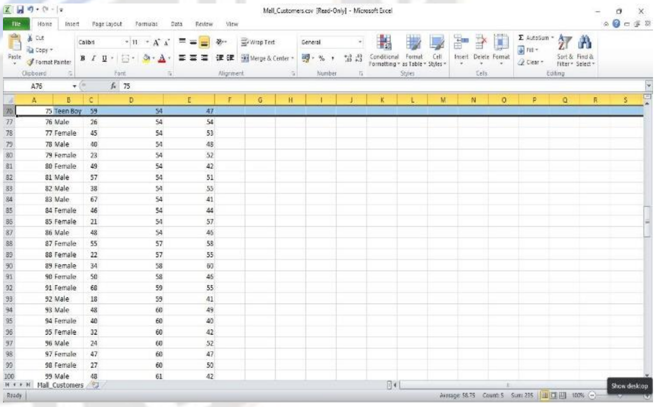


Fig-11: Mall Visitor data collection sheet-4 for customer segmentation

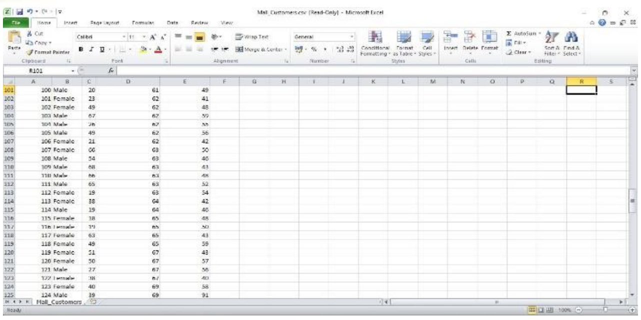


Fig-12: Mall Visitor data collection sheet-5 for customer segmentation

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8
Article Received: 25 June 2023 Revised: 12 July 2023 Accepted: 30 August 2023

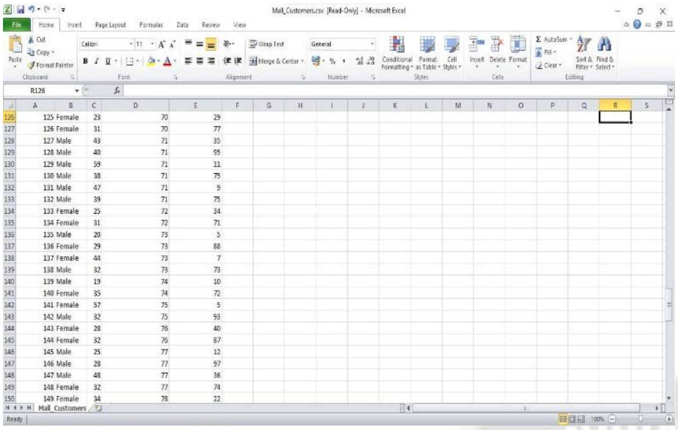


Fig-13: Mall Visitor data collection sheet-6 for customer segmentation

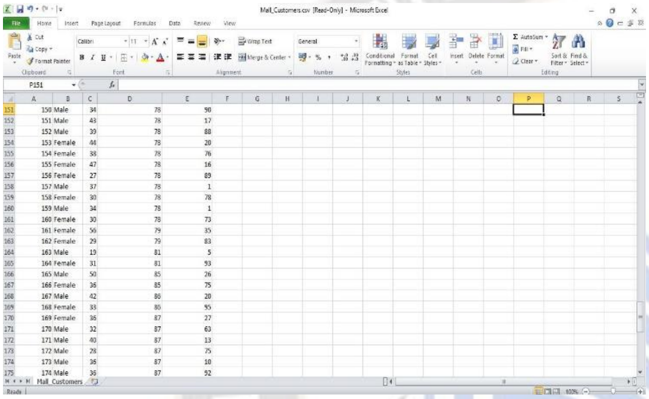


Fig-14: Mall Visitor data collection sheet-7 for customer segmentation

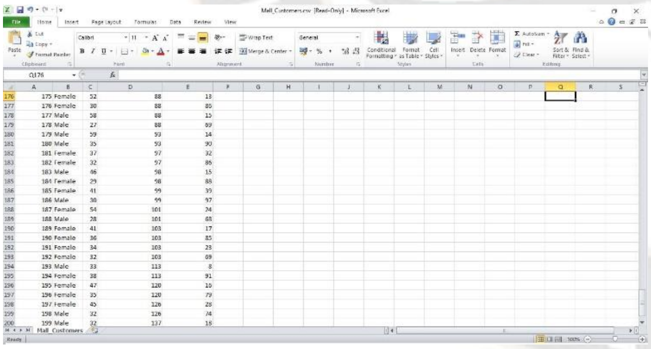


Fig-15: Mall Visitor data collection sheet-8 for customer segmentation

a. Removal of redundancy

The given standard data set for mall visitor's entry record data is taken into consideration for this research article.

Hash (Cust-Id=26) =2

The customerID-26 occurs twice so remove the second entry for redundancy removal such that the total number of records is 200-1=199.

b. Removal of incorrect data

The customer id 50 attains the spending score as 42 but the annual earning is 0 dollars which represents the incorrect data present in the customer records. Remove customer-id-50 such that the total number of records is 199-1=198.

c. Remove irrelevant data

The customer-Id 75 holds the gender value as Teen boy since it's an irrelevant data ,no need for removal, just change the gender to Male (Teen boy). Now the total number of records is still 198.

d. Composition of Customer segmentation preference approach

Now checking the constraints for applying the proper machine leaning based clustering for customer segmentation towards the standard dataset used in this research module.

Select Case (Customer segmentation preference)

{Case: - customer data are partially observable = NO fully observable so skip;

Case: - customer data are majority based= NO Equal rights of men and women so skip;

Case: - Number of customer groups is known=YES, 2 to 10 based on gender, income and spending

Case: - Number of customer groups is unknown=NO so not needed here;}

Now apply the K-means clustering for the standard input data for customer segmentation using XLSTat [14] utility tool as follows in Fig-16 through Fig-16, Fig-17, and Fig-18.

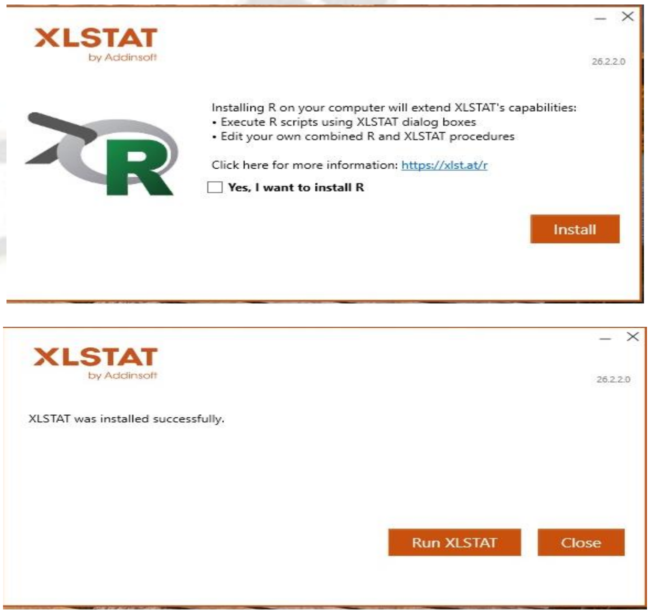


Fig-16: XLStat tool installation

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8
Article Received: *25 June 2023* ***Revised:*** *12 July 2023* ***Accepted:*** *30 August 2023*

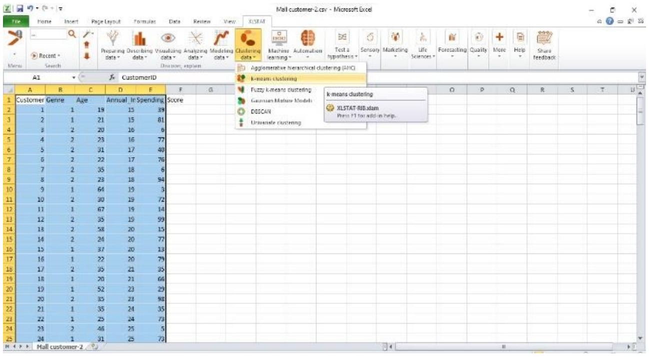


Fig-17: XLStat tool implementation setting for the research dataset

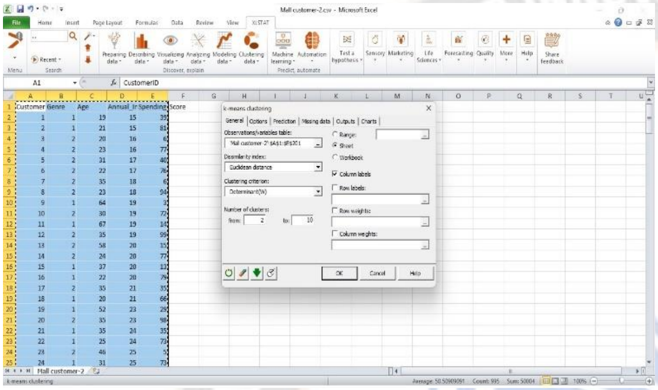
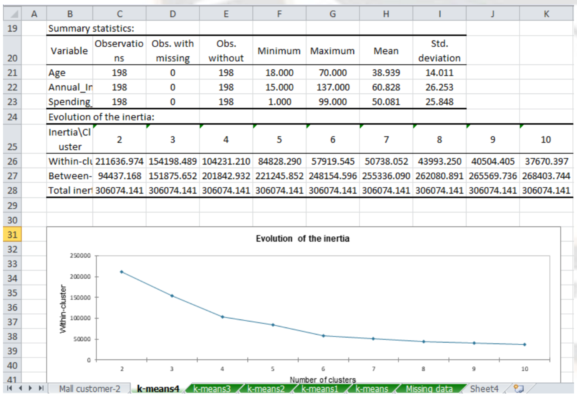


Fig-18: XLStat Execution setting for the research dataset

The outputs for the customer segmentation using k-means clustering with machine learning approach are shown in Fig-19, Fig-20, and Fig-21.



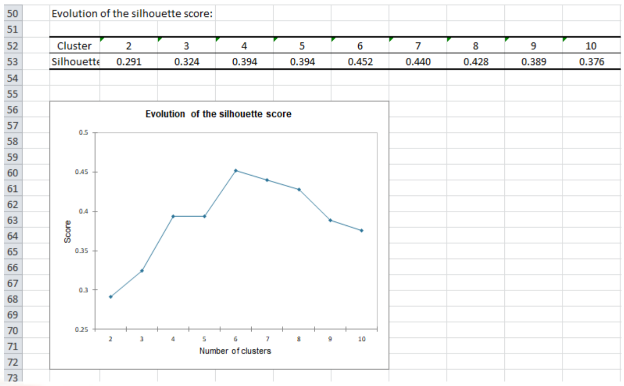
Summary statistics:

Variable	Observations	Obs. with missing	Obs. without	Minimum	Maximum	Mean	Std. deviation
Age	198	0	198	18.000	70.000	38.939	14.011
Annual_In	198	0	198	15.000	137.000	60.828	26.253
Spending	198	0	198	1.000	99.000	50.081	25.848

Evolution of the inertia:

Inertia\Cluster	2	3	4	5	6	7	8	9	10
Within-cl	211636.974	154198.489	104231.210	84828.290	57919.545	50738.052	43993.250	40504.405	37670.397
Between-	94437.168	151875.652	201842.932	221245.852	248154.596	255336.090	262080.891	265569.736	268403.744
Total iner	306074.141	306074.141	306074.141	306074.141	306074.141	306074.141	306074.141	306074.141	306074.141

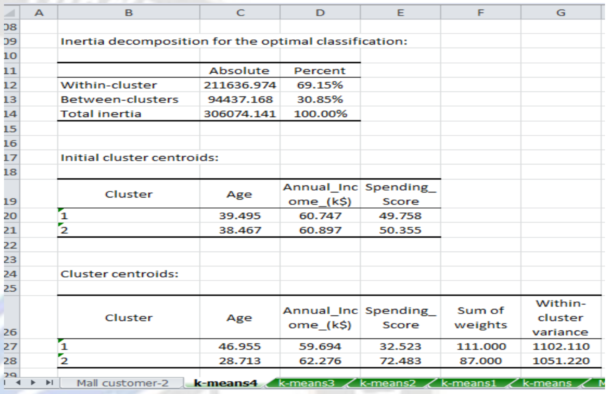
Fig-19: K-means clusters for the research dataset



Evolution of the silhouette score:

Cluster	2	3	4	5	6	7	8	9	10
Silhouette	0.291	0.324	0.394	0.394	0.452	0.440	0.428	0.389	0.376

Fig-20: Different customer clusters for the research dataset



Inertia decomposition for the optimal classification:

	Absolute	Percent
Within-cluster	211636.974	69.15%
Between-clusters	94437.168	30.85%
Total inertia	306074.141	100.00%

Initial cluster centroids:

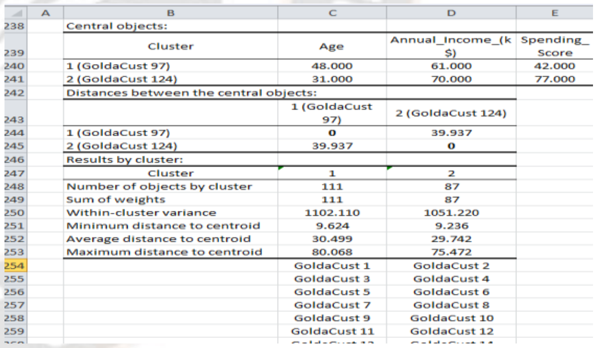
Cluster	Age	Annual_Income_(k$)	Spending_Score
1	39.495	60.747	49.758
2	38.467	60.897	50.355

Cluster centroids:

Cluster	Age	Annual_Income_(k$)	Spending_Score	Sum of weights	Within-cluster variance
1	46.955	59.694	32.523	111.000	1102.110
2	28.713	62.276	72.483	87.000	1051.220

Fig-21: Cluster centroids for the research dataset

The final output shows that the cluster formations for the standard data sets are optimal and up to the mark with 2 major clusters on gender as in Fig-22.



Central objects:

Cluster	Age	Annual_Income_(k$)	Spending_Score
1 (GoldaCust 97)	48.000	61.000	42.000
2 (GoldaCust 124)	31.000	70.000	77.000

Distances between the central objects:

	1 (GoldaCust 97)	2 (GoldaCust 124)
1 (GoldaCust 97)	0	39.937
2 (GoldaCust 124)	39.937	0

Results by cluster:

Cluster	1	2
Number of objects by cluster	111	87
Sum of weights	111	87
Within-cluster variance	1102.110	1051.220
Minimum distance to centroid	9.624	9.236
Average distance to centroid	30.499	29.742
Maximum distance to centroid	80.068	75.472
	GoldaCust 1	GoldaCust 2
	GoldaCust 3	GoldaCust 4
	GoldaCust 5	GoldaCust 6
	GoldaCust 7	GoldaCust 8
	GoldaCust 9	GoldaCust 10
	GoldaCust 11	GoldaCust 12

Fig-22: Major dual clusters for the research dataset

The members for the dual major clusters are shown in the following table -1.

Table-1: Dual major clusters for research data

Customer ID	Genre	Customer ID	Genre
		GoldaCust1	Male
GoldaCust3	Female	GoldaCust2	Male
GoldaCust4	Female	GoldaCust9	Male
GoldaCust5	Female	GoldaCust11	Male
GoldaCust6	Female	GoldaCust15	Male
GoldaCust7	Female	GoldaCust16	Male
GoldaCust8	Female	GoldaCust18	Male
GoldaCust10	Female	GoldaCust19	Male

GoldaCust12	Female	GoldaCust21	Male
GoldaCust13	Female	GoldaCust22	Male
GoldaCust14	Female	GoldaCust24	Male
GoldaCust17	Female	GoldaCust28	Male
GoldaCust20	Female	GoldaCust31	Male
GoldaCust23	Female	GoldaCust33	Male
GoldaCust25	Female	GoldaCust34	Male
GoldaCust27	Female	GoldaCust42	Male
GoldaCust29	Female	GoldaCust43	Male
GoldaCust30	Female	GoldaCust52	Male
GoldaCust32	Female	GoldaCust54	Male
GoldaCust35	Female	GoldaCust56	Male
GoldaCust36	Female	GoldaCust58	Male
GoldaCust37	Female	GoldaCust60	Male
GoldaCust38	Female	GoldaCust61	Male
GoldaCust39	Female	GoldaCust62	Male
GoldaCust40	Female	GoldaCust65	Male
GoldaCust41	Female	GoldaCust66	Male
GoldaCust44	Female	GoldaCust69	Male
GoldaCust45	Female	GoldaCust71	Male
GoldaCust46	Female	GoldaCust75	Male
GoldaCust47	Female	GoldaCust76	Male
GoldaCust48	Female	GoldaCust78	Male
GoldaCust49	Female	GoldaCust81	Male
GoldaCust51	Female	GoldaCust82	Male
GoldaCust53	Female	GoldaCust83	Male
GoldaCust55	Female	GoldaCust86	Male
GoldaCust57	Female	GoldaCust92	Male
GoldaCust59	Female	GoldaCust93	Male
GoldaCust63	Female	GoldaCust96	Male
GoldaCust64	Female	GoldaCust99	Male
GoldaCust67	Female	GoldaCust100	Male
GoldaCust68	Female	GoldaCust103	Male
GoldaCust70	Female	GoldaCust104	Male
GoldaCust72	Female	GoldaCust105	Male
GoldaCust73	Female	GoldaCust108	Male
GoldaCust74	Female	GoldaCust109	Male
GoldaCust77	Female	GoldaCust110	Male
GoldaCust79	Female	GoldaCust111	Male
GoldaCust80	Female	GoldaCust114	Male
GoldaCust84	Female	GoldaCust121	Male
GoldaCust85	Female	GoldaCust124	Male
GoldaCust87	Female	GoldaCust127	Male
GoldaCust88	Female	GoldaCust128	Male
GoldaCust89	Female	GoldaCust129	Male
GoldaCust90	Female	GoldaCust130	Male
GoldaCust91	Female	GoldaCust131	Male
GoldaCust94	Female	GoldaCust132	Male
GoldaCust95	Female	GoldaCust135	Male
GoldaCust97	Female	GoldaCust138	Male
GoldaCust98	Female	GoldaCust139	Male
GoldaCust101	Female	GoldaCust142	Male
GoldaCust102	Female	GoldaCust145	Male
GoldaCust106	Female	GoldaCust146	Male
GoldaCust107	Female	GoldaCust147	Male
GoldaCust112	Female	GoldaCust150	Male
GoldaCust113	Female	GoldaCust151	Male
GoldaCust115	Female	GoldaCust152	Male
GoldaCust116	Female	GoldaCust157	Male

GoldaCust117	Female	GoldaCust159	Male
GoldaCust118	Female	GoldaCust163	Male
GoldaCust119	Female	GoldaCust165	Male
GoldaCust120	Female	GoldaCust167	Male
GoldaCust122	Female	GoldaCust170	Male
GoldaCust123	Female	GoldaCust171	Male
GoldaCust125	Female	GoldaCust172	Male
GoldaCust126	Female	GoldaCust173	Male
GoldaCust133	Female	GoldaCust174	Male
GoldaCust134	Female	GoldaCust177	Male
GoldaCust136	Female	GoldaCust178	Male
GoldaCust137	Female	GoldaCust179	Male
GoldaCust140	Female	GoldaCust180	Male
GoldaCust141	Female	GoldaCust183	Male
GoldaCust143	Female	GoldaCust186	Male
GoldaCust144	Female	GoldaCust188	Male
GoldaCust148	Female	GoldaCust193	Male
GoldaCust149	Female	GoldaCust198	Male
GoldaCust153	Female	GoldaCust199	Male
GoldaCust154	Female	GoldaCust200	Male
GoldaCust155	Female		
GoldaCust156	Female		
GoldaCust158	Female		
GoldaCust160	Female		
GoldaCust161	Female		
GoldaCust162	Female		
GoldaCust164	Female		
GoldaCust166	Female		
GoldaCust168	Female		
GoldaCust169	Female		
GoldaCust175	Female		
GoldaCust176	Female		
GoldaCust181	Female		
GoldaCust182	Female		
GoldaCust184	Female		
GoldaCust185	Female		
GoldaCust187	Female		
GoldaCust189	Female		
GoldaCust190	Female		
GoldaCust191	Female		
GoldaCust192	Female		
GoldaCust194	Female		
GoldaCust195	Female		
GoldaCust196	Female		
GoldaCust197	Female		

The means by the dual major clusters are represented in the following table-2.

Table-2: Dual major cluster means for customer segmentation

Silhouette scores (Means by cluster):	
Component	**Silhouette scores**
Cluster 1	0.281
Cluster 2	0.304
Mean width	0.291

The proposed methodology produces good results without any errors or deviations due to the composition and preference based approach with its initial data cleaning procedures. This research article produces 99% (198 out of

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8
***Article Received:** 25 June 2023 **Revised:** 12 July 2023 **Accepted:** 30 August 2023*

200 record sets) of success rate for the preference compositional approach for client structured web customer segmentation using machine learning techniques. The parametric comparison between existing and proposed methods with precision, accuracy etc. are represented in the below Table-3 format,

Table-3: Proposed methodology parametric comparisons

No	Approach	Accuracy	Precision	Recall	F1 score value
1	Data mining based customer segmentation approach	73%	0.72	0.74	0.71
2	Proposed preference compositional approach for client structured web customer segmentation using machine learning techniques.	99%	0.98	0.99	0.98

The following fig-23 shows the performance comparison between the proposed and existing methodologies.

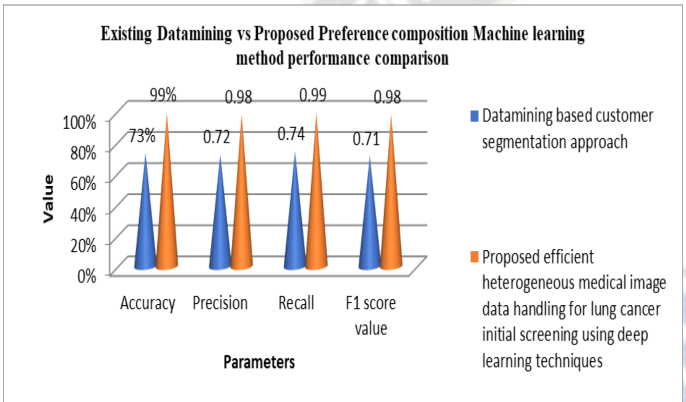


Fig-23: Proposed vs. existing methodology performance comparisons

V. CONCLUSION

Customer segmentation plays the vital role in running the corporate world in a successful manner. The process of handling large amount of data using the manual data base record process is very difficult to handle and very hard to analyses towards future references. The existing methodologies for customer segmentation directly applying the cluster algorithms in which the results are irrelevant in terms of optimized customer segmentation. Web data dealing for customer segmentation requires the proper soft computing tool for dynamic request handling procedures. This research article proposes 3 stages of customer segmentation process, initially the customer data are handled with soft computing based corrective motion, then followed by the data improvisation approach for its effectiveness and finally the preference composition approach using machine learning techniques are used for the proper customer segmentation in an optimized manner. This research article produced 99% success for the Customer segmentation in client structured web information system.

REFERENCES:

[1]. Alden, D.L. and Nariswari, A., 2017. Brand Positioning Strategies during Global Expansion: Managerial Perspectives from Emerging Market Firms. In The Customer is not Always Right? Marketing Orientations in a Dynamic Business World (pp. 527-530). Springer, Cham.

[2]. Boso, N., Hultman, M. and Oghazi, P., 2016, July. The impact of international entrepreneurial-oriented behaviors on regional expansion: Evidence from a developing economy. In 2016 Global Marketing Conference at Hong Kong (pp. 999-1000).

[3]. Boso, N., Oghazi, P. and Hultman, M., 2017. International entrepreneurial orientation and regional expansion. Entrepreneurship & Regional Development, 29(1-2), pp.4-26

[4] Nasrin JOKAR, Reza Ali HONARVAR, Shima AgHAMIRZADEH, and Khadijeh ESFANDIARI, "Web mining and Web usage mining techniques," Bulletin de la Société des Sciences de Liège, vol. 85, pp.321 - 328, 2016.

[5] Anurag Kumar and Kumar Ravi Singh, "A Study on Web Structure Mining," International Research Journal of Engineering and Technology (IRJET), vol. 04, no. 1, pp. 715-720, January 2017

[6] Dutton, T. An Overview of National AI Strategies. Available online: http://www.jaist.ac.jp (accessed on 8 January 2020)

[7]https:www.kaggle.com/datasets/shrutimechlearn/customer-data

[8] www.Salesforce.com

[9] www.hubspot.com

[10] www.zoho.com

[11] www.surveys.google.com

[12] www.zoom.com

[13] www.mentimeter.com

[14] www.xlstat.com