_____

Optimized Collaborative Approach for Client Structured Web Customer Segmentation Using Machine Learning Techniques

Adlin Selva Golda.V[1], Narayani.V[2]

[1]Research Scholar
Department of Computer Science
St.Xavier's College (Autonomous), Palayamkottai – 627 002

[2]Assistant Professor
Department of Computer Science
St.Xavier's College (Autonomous), Palayamkottai – 627 002

Affiliated to Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli – 627 012

Abstract

The process of customer segmentation generally focuses on the customer domain rather than the customer segmentation claimant domain which actually represent the top down design of a business model. If the target or goal is known for the customer segmentation then the design and implementation process adopt the essential components for achieving the goal in an optimized way. Based on requirements the customer segmentation process concentrates on selecting the appropriate data, approaches, methodologies, and tools with prompt measures so as to avoid unnecessary data handling process which produces efficient results. The existing customer segmentation methods are single approach oriented without any proper implementation of data metrics and current trends of machine learning tools. This research article proposes a corporate to customer oriented approach for the implementation of client structured web customer segmentation with the optimized collaborative process based on their requirements for customer segmentation. In near future this research article incorporates the automation based customer segmentation in web information system.

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

I. INTRODUCTION

Collaborative approach:

Collaborative approach includes processes, behaviors, and conversations that relate to the collaboration between two or multiple approaches. This approach specifically aims to increase the success of output as they engage in collaborative problem solving.

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.

1635

_____

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].

II. METHODOLOGY

The proposed methodology comprises 8 phases of implementation. They are

a. Requirement Phase

The corporate or an organization basic requirements for a customer segmentation is as follows,

❖ Data collection for analytics.
❖ Proper Approaches for customer data learning towards effective data access.
❖ Identify all possible segments based on real-time criteria.
❖ Collaboration of recent methodologies.
❖ Identification of interface tools to interact.
❖ Optimize the performance if required.
❖ Effective Presentation.

b. Data collection phase:

The data collection from the web information system includes the data heterogeneous from different sources in different formats. Data collection includes the following basic components.

❖ Data source.
❖ Data type.
❖ Data format.

The data collection procedures entirely depend on the data cleansing process.

c. Design Phase:

The different approaches for customer data learning towards effective data access,

❖ Overall customer examination using supervised learning.
❖ Knowledge of every customer using semi supervised learning.
❖ Explanation of segmentation possibility using unsupervised learning.
❖ Customer segmentation analysis using Reinforcement learning.

d. Segmentation phase:

The segmentation phase contains different possible ways for segmenting the data. The isolation and composite mixtures are based on the corresponding requirements. They are,

i. Location.
ii. Rate.
iii. Product.
iv. Consumer type.
v. Environment.
vi. Offer.
vii. Process.

e. Collaborative approach phase:

The collaborative approach phase includes 2 sub component phases. The computation component and algorithmic component with the individual and composite mixture support. They are,

i. Collaborative computation phase
ii. Collaborative algorithmic phase

f. Tools identification phase:

The tools identified for customer segmentation interfacing are as follows,

1. Google analytics
2. Hub spot
3. User Pilot
4. Mail chimp.

g. Performance tuning phase:

The estimation of needed number of clusters improves the customer segmentation process. The choices for the performance tuning strategy are as follows,

i. Elbow method
ii. Average silhouette method
iii. Gap statistics method.

h. Presentation Phase:

The customer segmentation results must be shown in a proper format for effective visualization in order to reduce the boredom of sequence data evaluation.

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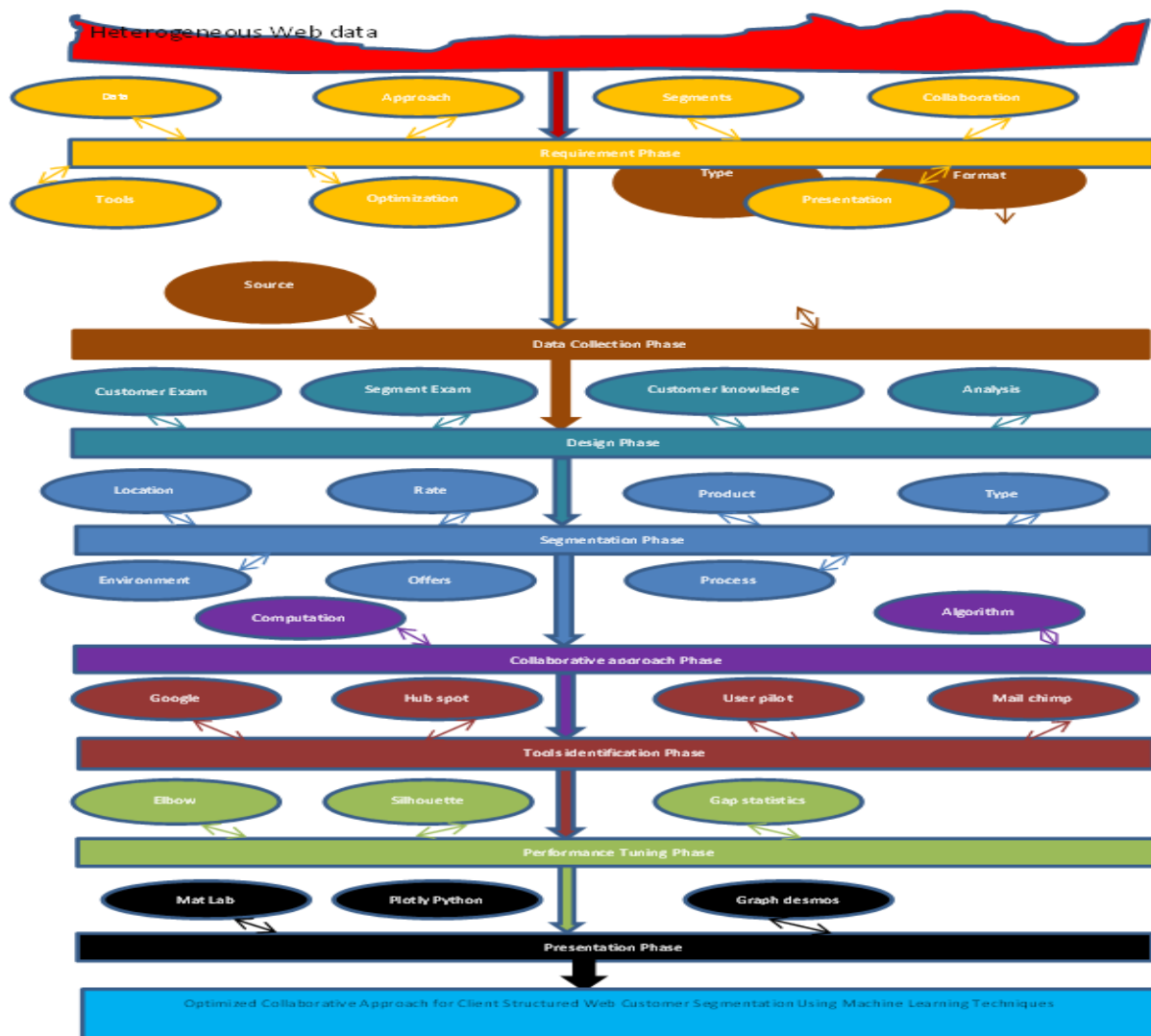
_____

Fig-1: Proposed optimized collaborative approach for client structured web customer segmentation

The flow chart for the optimized collaborative 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: Requirements Phase

Corporate or organization requirements for the customer segmentation are defined with proper terms and conditions. Set the requirements list as the key component for the customer segmentation.

Step-2: Data Collection Phase

This phase collects the data from different data sources with different data types represented in different formats for the unique structured data for customer segmentation.

Step-3: Design Phase

This phase focuses on the examining all the customers with individual knowledge and segmentation examination with its analysis.

Step-4: Segmentation Phase

This phase concentrates on the customer location, product type, rate with promotion and offers along with the environment and process features.

Step-5: Collaboration Phase

This phase collaborates the different soft computing approaches such as fuzzy logic, neural networks, genetic algorithm, and automata theory. It includes different algorithmic approaches. It acts as a multi domain implementation.

Step-6: Tools identification Phase

This phase incorporates the tools for the customer segmentation such as Google analytics, hub spot, user pilot and mail chimp etc.

Step-7: Performance tuning Phase

1637

_____

This phase concentrates on the optimized improvement in performance of customer segmentation through the cluster estimation approaches.

Step-8: Presentation Phase

This phase targets the proper presentation for the flexible and feasible visualization of customer segmentation results based on the proposed methodology implementation.

The completion of all the steps in an effective manner will produce the customer segmentation results in an optimal way otherwise repeat the steps for the corresponding features.

The steps that are needed individually can also be allowed for further tuning in the performance improvement.

END

III. IMPLEMENTATION

The requirement phase outcomes are set as the target goal points for the implementation of the proposed methodology.

1. Data collection phase:

The data collection from the web information system includes the data heterogeneous from different sources in different formats. Data collection includes the following basic components.

a. Data source.

The data sources are categorized into two types based on machine learning categorization schema, one from retrieving the data naturally and other from initiating the consumer to get the data artificially in a smooth way. As shown in the following table-1.

Table-1: Data sources table

Sl.No	Natural source data	Initiated source data
1	Geographical location.	Customer complaint.
2	Browsing history.	Customer review.
3	Customer preferences.	Survey.
4	Purchase History.	Forum.
5	Payment mode.	Community.
6	Profile.	Report.
7	Device info.	Marketing data.
8	Software status.	Product rating.
9	Interests.	Feedbacks.
10	Friends and family.	Suggestions.

b. Data type:

The data types from the corporate or an organization point of view are classified into 3 categories based on the machine learning objectives. They are,

i. Marketing data.

ii. Product development data.

iii. Customer service data.

c. Data format:

The data formats are raw data, structured data, semi structured data and unstructured data. The implementation of converting raw data, semi structured and unstructured towards structured data are completed through data cleansing procedures.

2. Design Phase:

The different approaches for customer data learning towards effective data access,

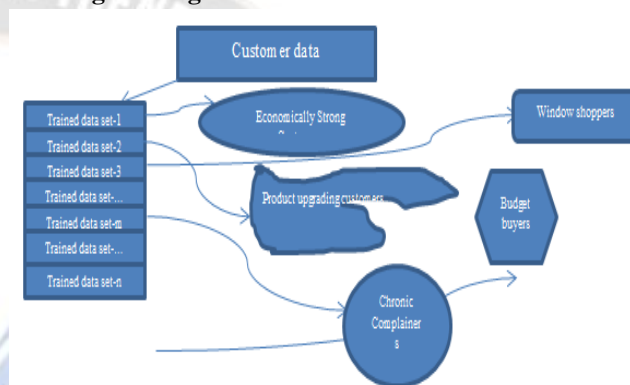### a. Overall customer examination using supervised learning as in fig-2.

Fig-2: Supervised Learning in Segmentation

b. Knowledge of every customer using semi supervised learning.

The knowledge of every customer without a predefined training data set are assessed through the semi supervised learning based on the machine learning based if-then-tree structure.

For each customer
If customer location="Metro" then
    Location ID="L1";
Else if customer location="Urban" then
    Location ID="L2";
Else if customer location="Rural" then
    Location ID="L3";
Else
    Location ID="L4";
End if
If customer education="Professional degree" then
    Education ID="E1"
Else if customer education="Post graduate degree" then
    Education ID="E2"
Else if customer education="Under Graduate degree" then

1638

_____

Education ID="E3"

Else if customer education="HSc/SSLC/ESLC" then

Education ID="E4"

Else

Education ID="E5"

End if
If annual income>50L then

Economic Income status ID="I1";

Else if 25L < annual income <= 50L then

Economic Income status ID="I2";

Else if 12L < annual income <= 25L then

Economic Income status ID="I3";

Else if 6L < annual income <= 12L then

Economic Income status ID="I4";

Else

Economic Income status ID="I5";

End if
End For

The following sample table-2 represents the knowledge about each customer based on the semi supervised learning structure. For the Metro area customer with ID L1, the variation knowledge is,

Table-2: Knowledge of each Customer

Sl.NO	Customer Name	Knowledge Id
1	C1	E1I1
2	C2	E1I2
3	C3	E1I3
4	C4	E1I4
5	C5	E1I5
6	C6	E2I1
7	C7	E2I2
8	C8	E2I3
9	C9	E2I4
10	C10	E2I5
11	C11	E3I1
12	C12	E3I2
13	C13	E3I3
14	C14	E3I4
15	C15	E3I5
16	C16	E4I1
17	C17	E4I2
18	C18	E4I3
19	C19	E4I4
20	C20	E4I5
21	C21	E5I1
22	C22	E5I2
23	C23	E5I3
24	C24	E5I4
25	C25	E5I5

c. Explanation of segmentation possibility using unsupervised learning.

The particular portion of labeled data and the remaining unknown data combination creates the possibility for the customer segmentation. The application of semi supervised learning provides the following segment information's from table-2 in order to develop the segment information's as in table-3 as follows.

Table-3: Segment possibility table

Sl.NO	Customer Name	Knowledge Id	Customer knowledge description
1	C1	E1I1	Metro Area High professional Rich customer
2	C2	E1I2	Metro Area High professional High class customer
3	C3	E1I3	Metro Area High professional Upper middle class customer
4	C4	E1I4	Metro Area High professional Middle class customer
5	C5	E1I5	Metro Area High professional Lower Middle class customer
6	C6	E2I1	Metro Area Post Graduate Rich customer

1639

_____

7	C7	E2I2	Metro Area Post Graduate High class customer
8	C8	E2I3	Metro Area Post Graduate Upper middle class customer
9	C9	E2I4	Metro Area Post Graduate Middle class customer
10	C10	E2I5	Metro Area Post Graduate Lower Middle class customer
11	C11	E3I1	Metro Area Under Graduate Rich customer
12	C12	E3I2	Metro Area Under Graduate High class customer
13	C13	E3I3	Metro Area Under Graduate Upper middle class customer
14	C14	E3I4	Metro Area Under Graduate Middle class customer
15	C15	E3I5	Metro Area Under Graduate Lower Middle class customer
16	C16	E4I1	Metro Area Charm school Rich customer
17	C17	E4I2	Metro Area Charm school High class customer
18	C18	E4I3	Metro Area Charm school Upper middle class customer
19	C19	E4I4	Metro Area Charm school Middle class customer
20	C20	E4I5	Metro Area Charm school Lower Middle class customer
21	C21	E5I1	Metro Area Non-school Rich customer
22	C22	E5I2	Metro Area Non-school High class customer
23	C23	E5I3	Metro Area Non-school Upper middle class customer
24	C24	E5I4	Metro Area Non-school Middle class customer
25	C25	E5I5	Metro Area Non-school Lower Middle class customer

d. Customer segmentation analysis using Reinforcement learning.

The reinforcement learning of reward and punishment way of approach provides the information to create the customer segmentation in a symbolic unique notation. It works as

If symbol=X then customer belongs to group ABC else "NO"

The implementation of reinforcement learning gives the customer category symbols as in table-4,

Table-4: Segment analysis table

Sl.NO	Customer Name	Knowledge Id	Customer knowledge description	Customer Category
1	C1	E1I1	Metro Area High professional Rich customer	O1++++

1640

_____

2	C2	E1I2	Metro Area High professional High class customer	A+++
3	C3	E1I3	Metro Area High professional Upper middle class customer	A++
4	C4	E1I4	Metro Area High professional Middle class customer	A+
5	C5	E1I5	Metro Area High professional Lower Middle class customer	A
6	C6	E2I1	Metro Area High professional Rich customer	O1+++
7	C7	E2I2	Metro Area High professional High class customer	B+++
8	C8	E2I3	Metro Area High professional Upper middle class customer	B++
9	C9	E2I4	Metro Area High professional Middle class customer	B+
10	C10	E2I5	Metro Area High professional Lower Middle class customer	B
11	C11	E3I1	Metro Area High professional Rich customer	O1++
12	C12	E3I2	Metro Area High professional High class customer	C+++
13	C13	E3I3	Metro Area High professional Upper middle class customer	C++
14	C14	E3I4	Metro Area High professional Middle class customer	C+
15	C15	E3I5	Metro Area High professional Lower Middle class customer	C
16	C16	E4I1	Metro Area High professional Rich customer	O1+
17	C17	E4I2	Metro Area High professional High class customer	C+++
18	C18	E4I3	Metro Area High professional Upper middle class customer	C++
19	C19	E4I4	Metro Area High	C+

_____

			professional Middle class customer	
20	C20	E4I5	Metro Area High professional Lower Middle class customer	C
21	C21	E5I1	Metro Area High professional Rich customer	O1
22	C22	E5I2	Metro Area High professional High class customer	D+++
23	C23	E5I3	Metro Area High professional Upper middle class customer	D++
24	C24	E5I4	Metro Area High professional Middle class customer	D+
25	C25	E5I5	Metro Area High professional Lower Middle class customer	D

3. Segmentation phase:

The different customer segment factors are,

       i. Location.
      ii. Rate.
     iii. Product.
     iv. Consumer type.
      v. Environment.
     vi. Offer.
    vii. Process.

The identification of segments based on the association rule mining and mapping of pattern matching provides the following process.

       i. Ordering the data.
         Cost value
         Lifetime value
         Rating
      ii. Identify the patterns.
         Customer satisfaction
         Buying nature
         Discounts oriented

4. Collaborative approach phase:

The computation and algorithmic combinations provides better results than its individual implementation.

a. Collaborative computation phase:

The inclusion of following procedures completes the computation structures.

i. Fuzzy membership valuation.

The customer segmentation based on fuzzy membership values for providing the sponsored content from the corporate or an organization for the chance of successive purchase marketing is as follows in the sample table-5

Table-5: Fuzzy collaboration

Sl.No	Purchased Product	Successive product to buy next	Fuzzy membership value based on existing customer data
1	Glucometer	Strips	1.0
2	Mobile/WM	Mobile/WM cover	0.9
3	TV/Fridge	Stabilizer	0.8
4	Handbag(Women)	Purse(Women)	0.7
5	Laptop	Additional warranty	0.6
6	Samsung mobile	Power adopter	0.5
7	Shirts(Men)	Tie(Men)	0.4
8	Dishwasher	Powder/Liquid	0.3
9	Vermicomposting	Cocopeat	0.2
10	Dining table	Additional chair	0.1
11	TV	Additional Remote	0.0

_____

ii. Genetic cross over.

The corporate or an organization uses the genetic cross over approach by mixing the properties of one product with the other product in order attain several customer segmentations. Sometimes a new segmentation will be formed by using the genetic cross over approach. The entertainment industry or an organization uses the cross over approach as follows in table-6,

Table-6: Entertainment data sector

Sl.No	Gender	Age group	Most Interested Content Genre(With Exceptions)
1	Male	Kid	Games
2	Male	Teen	Action
3	Male	Youth	Romance
4	Male	Adult	Sports
5	Male	Adulthood	Family
6	Female	Kid	Aesthetics
7	Female	Teen	Drama
8	Female	Youth	Fashion
9	Female	Adult	Emotional
10	Female	Adulthood	Spiritual

The process of attracting more possible customer segments by applying the genetic cross over based documents/games/movies/web series from an entertainment industry is as follows in Table-7

Table-7: Genetic algorithm collaboration

Sl.No	Cross-over	New Customer segment
1	Colorful natural games with animals	All kids
2	Action, Romance with modern trend drama	All Teens and youths
3	Modern emotional drama	Female teen, youth and adult
4	Spiritual and family	All Senior citizens.
5	Sportsman biography	Male Teen, youth, and adult.

iii. Neural networks.

Neural networks comprise input layer, hidden/process layer and the output layer. The output is based on the threshold value. If the processed value is greater than the threshold then output is generated otherwise learning takes place until the output is reached.

In an educational organization,

Consider the threshold value 50 marks

Input:

Normal Students (customers)

Hidden layers:

HL-1: Class teaching with text book
HL-2: Written Notes
HL-3: Presentations
HL-4: AV contents
HL-5: Question bank
HL-6: Important Questions

Threshold:

50%

Output:

Customer segmentation category

The neural network based approach for customer segmentation is representation in the following table-8

Table-8: Neural networks collaboration

Student ID	HL-1	HL-2	HL-3	HL-4	HL-5	HL-6	Score/Threshold	Customer segment category
S1	>50						>50	A1
S2	<50	>50	>50				>50	A2
S3	<50	<50	>50				>50	A3
S4	<50	<50	>50	>50			>50	A4
S5	<50	<50	>50	<50	>50		>50	A5

_____

S6	<50	<50	>50	<50	<50	>50	>50	A6

iv. Automata theory.

The corporate to customer side initiation of purchasing the mobile phone through different ways of triggering are represented in the automata theory format.

Consider the scenario in the theory of automata, as in fig-3

With the states q0, q1, q2, q3 and q4 and

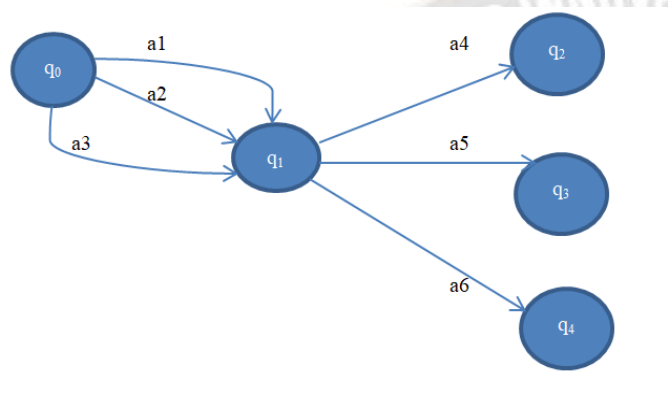With the corresponding actions named as a1, a2, a3, a4, a5, and a6.

Fig-3: Automata theory collaboration

Such that,

The state and action set representations are as follows,

Set-1:

q_0=Starting state =Mobile Purchasing decision from the customer
$a1$=Selecting High premium brand due to pride.
$a2$=Selecting same brand due to upgrade to newer version.
$a3$=Selecting equal level brand Attractive features.

Set-2:

q_1= Mobile Purchasing mode is ready
$a4$=Coupons, Card Discounts and Exchange offer
$a5$=Exchange offer alone
$a6$=No promotion and no exchange

Set-3:

q_2=Purchase the mobile
q_3=Wait for some more time
q_4=NO purchase.

The transition δ function is given by table-9,

Table-9: Transition table for automata collaboration

fun δ: Qx∑-->Q States/Action	a1=Selecting High premium brand due to pride.	a2=Selecting same brand due to upgrade to newer version.	a3=Selecting equal level brand Attractive features.	a4=Coupons, Card Discounts and Exchange offer	a5=Exchange offer alone	a6=No promotion and no exchange
q_0=Starting state =Mobile Purchasing decision from the customer	q_1= Mobile Purchasing mode is ready	q_1= Mobile Purchasing mode is ready	q_1= Mobile Purchasing mode is ready	NULL	NULL	NULL
q_1= Mobile Purchasing mode is ready	NULL	NULL	NULL	q_2=Purchase the mobile	q_3=Wait for some more time	q_4=NO purchase.
q_2=Purchase the mobile	NULL	NULL	NULL	NULL	NULL	NULL
q_3=Wait for some more time	NULL	NULL	NULL	NULL	NULL	NULL
q_4=NO purchase.	NULL	NULL	NULL	NULL	NULL	NULL

_____

b. Collaborative algorithmic phase:

The two main algorithmic approaches used for implementing customer segmentation are

i. K-means clustering:

Start

Step-1: Initialize k for cluster centroids count.

Step-2: Form different groups based on each centroid neighbors.

Step-3: Reduce sum of square point distance.

Step-4: If All data points unite then stop otherwise continue step-1.

Stop

ii. Hidden Markov Model:

Start:

Step-1: Identify state and observation space.

Step-2: Perform initial state distribution

Step-3: Apply state transition probabilities

Step-4: Identify possible observations using forward backward approach.

Step-5: Decode the hidden states.

Step-6: Evaluate the model.

Stop

5. Tools identification phase:

The tools identified for customer segmentation interfacing are as follows,

a. Google analytics

The machine learning tool identification of sample Google analytics [8] tool page is shown in fig-4.
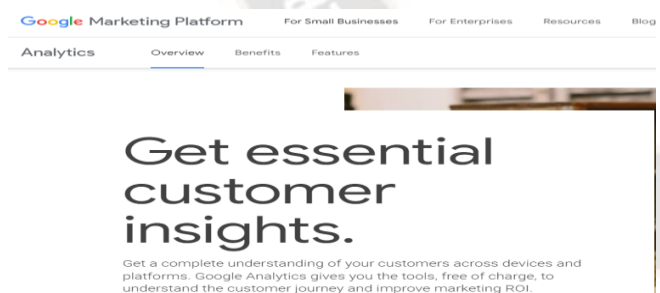

Fig-4: Google analytics tool page

b. Hub spot:

The machine learning tool identification of sample hub spot [9] tool page is shown in fig-5.

Fig-5: hub spot tools page

c. User Pilot

The machine learning tool identification of sample user pilot [10] tool page is shown in fig-6.
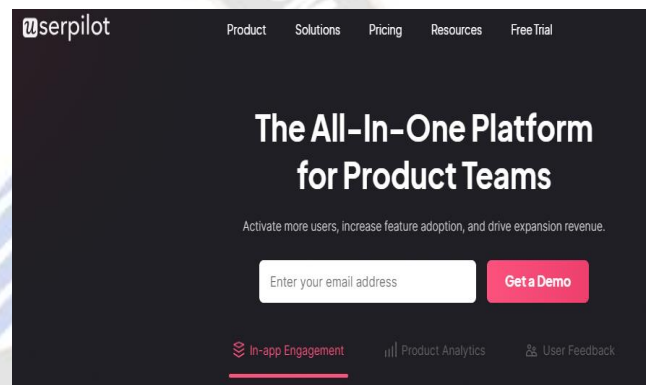

Fig-6: User pilot tools page

d. Mail chimp:

The machine learning tool identification of sample mail champ [11] tool page is shown in fig-7.
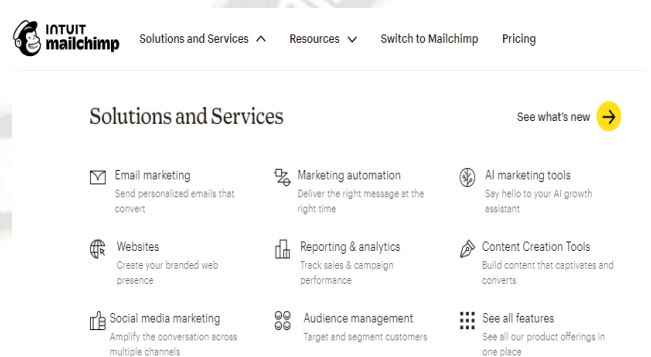

Fig-7: Mail chimp tools page

The selection of tools are based on the following factors,

- ❖ Abstractness
- ❖ Reliability
- ❖ Robustness
- ❖ Easy to use

1645

_____

- ❖ Up gradation and updating possibility
- ❖ Report generation

6. Performance tuning phase:

The performance improvement tuning is based on the estimation of number of clusters in the early stage itself. The performance tuning functions are as follows,

a. Elbow method:

Identifying the number of cluster based on the elbow bend change in ordered data variation.

b. Average silhouette method:

Self-similarity within the cluster counts.

c. Gap statistics method:

Compare the observed deviation and expected deviation for hypothesis count setting.

7. Presentation Phase:

The customer segmentation results must be shown in a proper format for effective visualization in order to reduce the boredom of sequence data evaluation.

The open source tools plays the important role in visualization by the top management in any corporate or an organization.

The Mat Lab, plotly python, and Graph desmos are the sample tools for effective visualization.

IV. RESULTS AND DISCUSSION

The machine learning based implementation of customer segmentation in us structured domain provides the optimized results.

The standard data sets from kaggle [7] and real-time are used for the implementation process. This research module gives 99% (594 out of 600 record sets from standard datasets) with the success rate for the optimized collaborative approach in client structured web customer segmentation using machine learning techniques.

The existing data mining approach gives only 401out of 600 record sets with 66.83% of success rate only due to its traditional standalone dependency.

The parametric comparison between existing and proposed methods metrics variations are represented in the below Table-10 format,

Table-10: Proposed methodology parametric comparisons

No	Approach	Accuracy	Precision	Recall	F1 score value
1	Data mining based customer segmentation approach.	66.83%	0.67	0.68	0.65
2	Proposed optimized collaborative approach for client structured web customer segmentation using machine learning techniques.	99%	0.98	0.99	0.97

The following fig-7 shows the performance comparison between the proposed and existing methodologies.
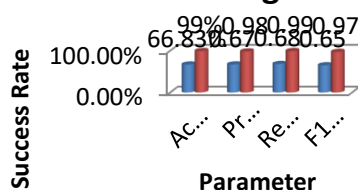

Fig-8: Proposed vs. existing methodology performance comparisons

V. CONCLUSION

The collaboration of machine learning, fuzzy logic genetic algorithm, neural networks and automata theory in data analytics for customer segmentation provides the most optimized results when compare with its each individual standalone implementation.

The process of handling vast data along with the goal of achieving the proper results in customer segmentation with perfect performance tuning of results is the desired one to implement.

The existing methodologies for customer segmentation don't provide good results due to its traditional standalone approaches.

This research article proposes 8 stages of implementation starting with the requirement phase

1646

_____

followed by the data collection phase then with the design phase followed by the segmentation phase along with the tools identification phase followed by the performance tuning phase and finally with the presentation phase.

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. https://marketingplatform.google.com/about/analytics/

9. https://www.hubspot.com/products/crm

10. https://userpilot.com/

11. https://mailchimp.com/