Heuristic Approach for Client Structured Web Customer Segmentation Using Machine Learning Techniques

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

Abstract: The customer segmentation in the web domain needs the data cleaning and improvisation process for the fundamental functionalities. The generic machine learning based customer segmentation produces the better results when compared with the direct clustering process. The immense implementation of machine learning based customer segmentation by considering the component based implementation approach with different types of customer segmentation must strengthen the goal with effective results. The existing customer segmentation methods lacks in the areas of scale, modification and verification. Applying machine learning as a whole is degraded in its efficiency when focusing on the each individual component incorporation of machine learning process. This research article proposes a component wise machine learning approach for the implementation of client structured web customer segmentation with the heuristic process based on their requirements of online web requests and responses. In near future this research article will be extended with the implementation of neural networks based customer segmentation in web information system.

I. INTRODUCTION

Heuristic approach:

Heuristics, or "rules of thumb," are problem-solving methods that are based on practical experience and knowledge. They allow you to use a "quick fix" to solve a minor problem or to narrow down options. They're also a great starting point for brainstorming or exploring new ideas.

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].

II. METHODOLOGY

The proposed methodology comprises 5 levels of implementation. They are

a. Customer Segmentation Category

The customer segmentation category lies on its incorporated attributes which classify the customer into the following segments.

❖   Regional based customer segmentation

❖   Traits based customer segmentation

❖   Attitude based customer segmentation

❖   Values based customer segmentation

b. Heuristic Selection of Machine learning approach

There are many machine learning approaches used for the customer segmentation process but he primary

[1]*Research Scholar, Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, India.*
Email: adlingolda@gmail.com
[2]*Assistant Professor, Department of Computer Science,*
St Xavier's College (Affiliated to Manonmaniam Sundaranar University), Tirunelveli, Tamil Nadu, India.

machine learning approaches related with the customer segmentation are as follows,

❖ Machine learning based Clustering algorithms.

❖ Machine learning based Decision trees.

❖ Machine learning based neural networks.

❖ Machine learning based Association rule mining.

c. Apply preferential mapping based learning style

The different types of learning used for customer segmentation in web information system are as follows,

❖ Supervised learning

❖ Semi supervised learning

❖ Unsupervised learning

❖ Reinforcement learning

## d. Extremity cluster based Machine learning approach for customer segmentation

The extremity cluster identifies the maximum number of clusters in the customer segmentation process in order to obtain the peak classifications for the thorough data analytics procedures in data process improvement. The method used for extremity cluster is known as Elbow method. This method defines the maximum value of K in k-means clustering.

e. Identify the tools for implementing appropriate customer segmentation

The following tools are used for implementing appropriate customer segmentation.

1. XLStat

2. Python

3. Mat Lab

4. Orange

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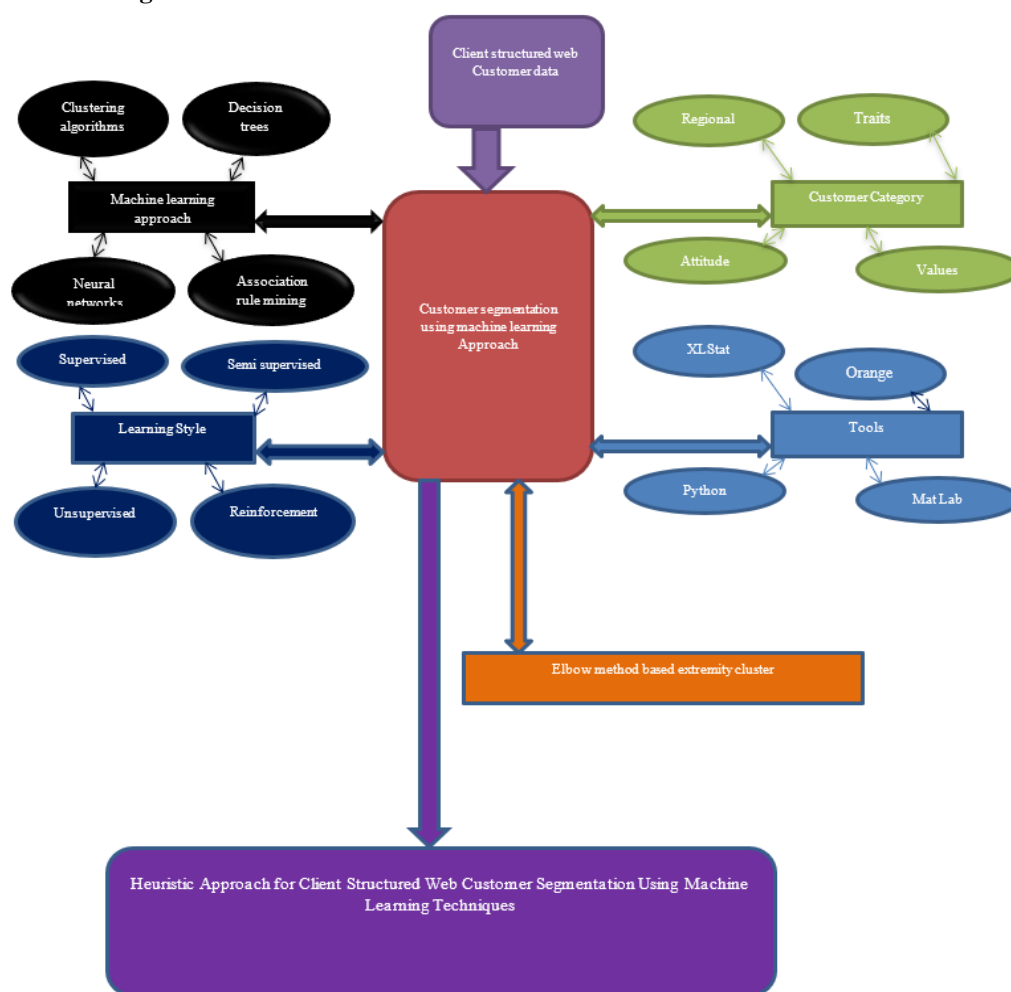
Fig-1: Proposed heuristic approach for client structured web customer segmentation

The flow chart for the heuristic 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: Identify customer segmentation category

a. Check for the region

b. Focus on the traits.

c. Target the attitude.

d. Evaluate the values.

Step-2: Perform criteria based Heuristic selection of machine learning approach

a. Algorithm for clusters.

b. Formation of decision trees.

c. Apply neural networks

d. Implement association rule mining

Step-3: Preference based learning style implementation

1. Supervised learning for known factors.

2. Semi supervised learning for mixed factors.

3. Unsupervised learning for unknown factors.

4. Reinforcement learning with reward and penalty.

Step-4: Compute extreme cluster for customer segmentation

Apply Elbow method with monotonic decrease curve point identification

Step-5: Apply the appropriate tool

1. Excel based XLStat

2. Script based Python

3. Design based Mat Lab

*4. User interface based Orange*

END

III. IMPLEMENTATION

Step-1: Identify customer segmentation category

The customer segmentation category is identified through the attributes associated with the customer details. The customer details include the following fields.

1. Name

2. Location

3. Gender

4. Age

5. Education

6. Income & Occupation

7. Opinion/review about the product

8. Web access time

9. Orders placed

10. Product search history

a. Check for the region

Based on the location the customer segmentation is classified as follows in fig-2,
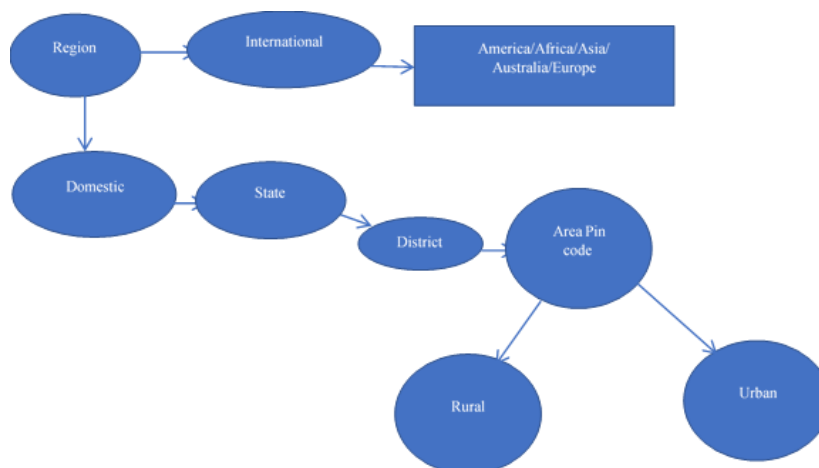
Fig-2: Region based customer segmentation category

b. Focus on the traits.

The customer traits such as age, gender, income, and occupation decide the customer segmentation category as represented in the following table-1.

Table-1: Age and gender based traits category

Sl.No	Age span	Gender	Customer Segmentation category
1	<15	Male/Female/Trans	Kids
2	16 to 19	Male/Female/Trans	Teens
3	20 to 35	Male/Female/Trans	Youth
4	36 to 50	Male/Female/Trans	Adult
5	51 to 60	Male/Female/Trans	Parent/Middle age
6	61 to 80	Male/Female/Trans	Senior citizen
7	>81	Male/Female/Trans	Super Senior citizen

Based on income the following customer segment category are identified as in table-2,

Table-2: Income based traits category

Sl.No	Annual income	Customer Segmentation category
1	< 2.5 Lakhs	Basic class
2	2.5 to 5 Lakhs	Middle class
3	5 to 10 Lakhs	Upper middle class
4	10 to 50 Lakhs	High class
5	>50 Lakhs	Rich class

Based on occupation the following customer segment category are identified as in table-3

Table-3: Occupation based traits category

Sl.No	Occupation	Customer Segmentation category
1	Doctors/Stars/Creators/Business tycoon/Executives etc.	Professional
2	Central/State/Rural/Bank etc.	Government
3	IT/Motor/Electronics etc.	Private A-class
4	Teachers/Employees etc.	Other Private
5	Self-labored	Skilled

c. Target the attitude.

The customer segmentation based on the attitude depends on the following factors as mentioned in table-4.

Table-4: Attitude based category

Sl.No	Attitude	Customer Segmentation category
1	Purchase behavior	Mobile/Electronics/Home/Kitchen
2	Occasion	Festival
3	Timing	Repeat
4	Gain attained	Coupon/card offer/prepaid offer
5	Loyalty	Brand/Seller/Product
6	Journey stage	1 to 3 or 3 to 5 or >=5
7	Yearly purchase	High>1 Lakh/medium 50K/low <50K

8	Satisfaction/Reviews	More positive/Neutral/More negative
9	Refund/cancel	Frequent/sometimes/rare
10	Special membership status	Yes/No

d. Evaluate the values.

The customer segmentation based on the attitude depends on the following factors as mentioned in table-5.

Table-5: Values based category

Sl.No	Value	Customer Segmentation category
1	Openness	Ready to test/ experience new product
2	Life style	Family/party/vacation/tour/office
3	Social status	Sophisticated/expensive/moderate/
4	Opinion	Positive/negative
5	Activities	Outgoing/energetic/sociable/reserved
6	Belief	Religious
7	Aesthetic	Art/natural/replica collectors
8	Neuroticism	Aggressive/calm
9	Interests	Macro/micro times watch/buy
10	Priority	Product/accessory/both

Step-2: Perform criteria based Heuristic selection of machine learning approach

The following select case component defines the heuristic selection of machine learning approach   **.**

The corresponding approaches are evaluated accordingly.

Select case (Customer segmentation category)

Case (Region):

Apply Cluster algorithmic approach as K-means clustering

Break;

Case (Traits):

Apply decision trees approach

Break;

Case (Attitude):

Apply neural networks approach

Break;

Case (values):

Apply association rule mining approach

Break;

End Select

Step-3: Preference based learning style implementation

The following select case component defines the preference based learning style implementation of machine learning approach   **.**

Select case (Customer segmentation category)

Case (Region):   Apply supervised learning style

Break;

Case (Traits):   Apply Semi supervised learning style

Break;

Case (Attitude):   Apply unsupervised learning style

Break;

Case (values):   Apply Reinforcement learning style

Break;

End Select

Step-4: Compute extreme cluster for customer segmentation

Apply Elbow method with monotonic decrease curve point identification by using the following stages.

The stages are:

Stage-1: k=1 compute the within cluster sum of square value.

Stage-2: If the curve is monotonically increasing then continue with the k increment

Stage-3: k=K++ Continue the process until the curve starts decreasing.

Stage-4: Note k, such that k is the extremity value for the cluster count.

Stage-5: Stop

The final value of k determines the maximum number of clusters available in the customer segmentation process using the k-means clustering approach.

For example k=3 represents the maximum of 3 valid clusters.

The valid number of clusters is not optimal clusters always. The optimal number of clusters depends on the quality of data present in the entire architecture.

Step-5: Apply the appropriate tool

1. Excel based XLStat

The Xlstat [8] tool as in fig-3 is used for supervised learning with K-means clustering technique
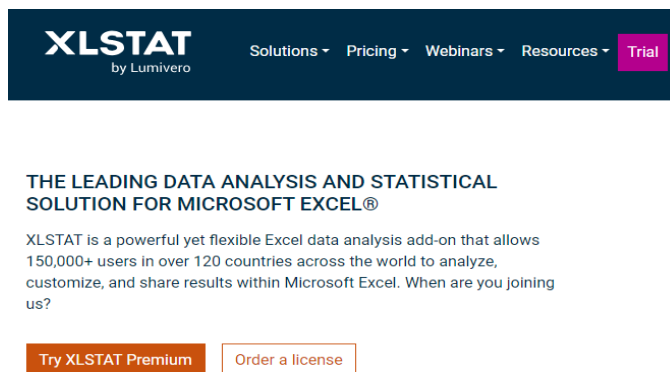

Fig-3: XLStat tool page for implementation

2. Script based Python

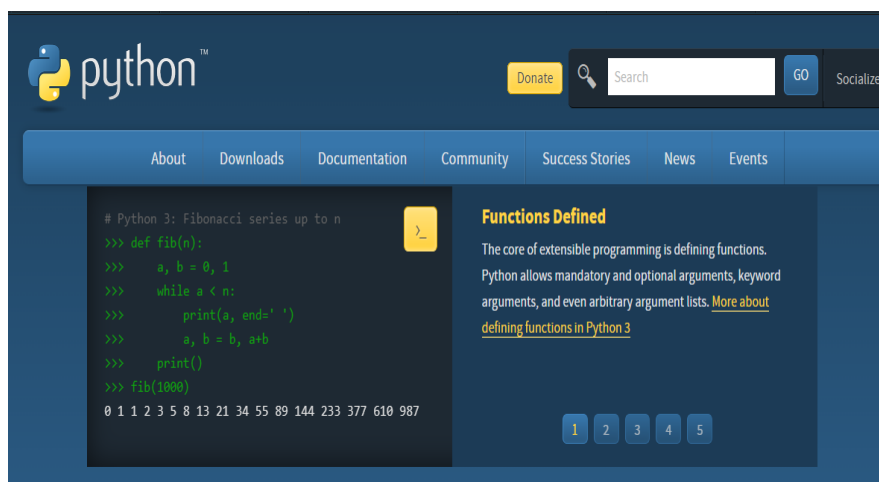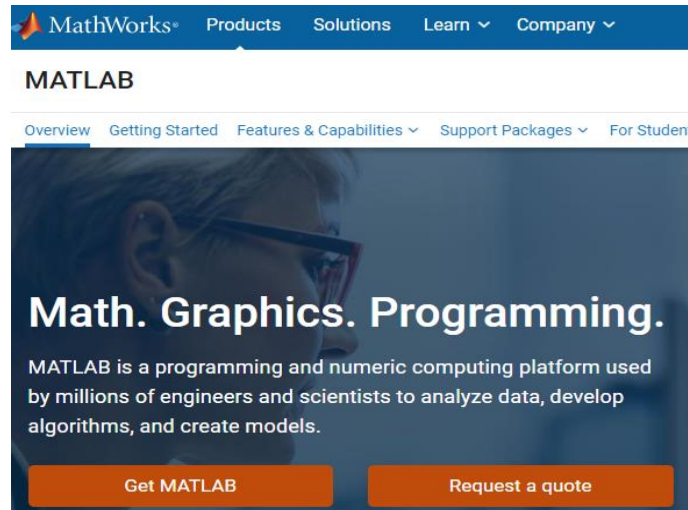The Python [9] tool as in fig-4 is used for semi supervised learning style.

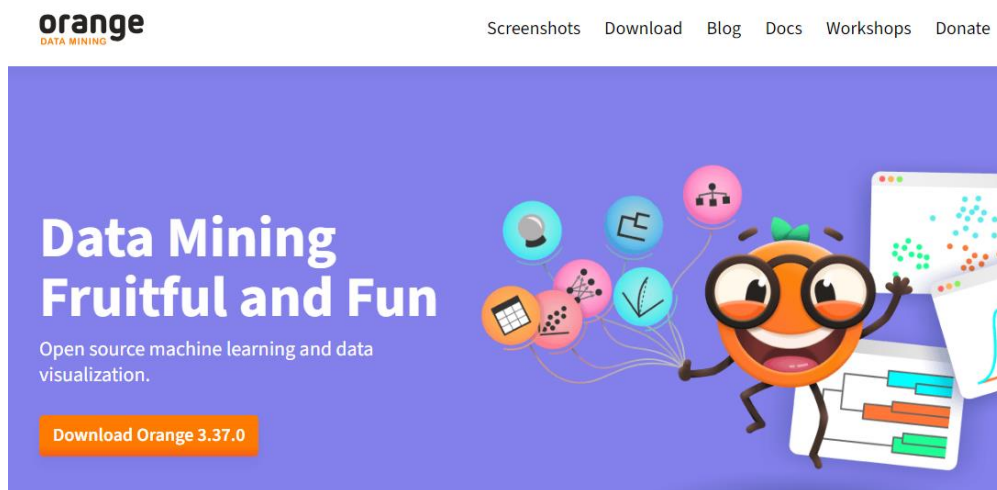**Fig-4:** Python tool page for implementation

3. Design based Mat Lab

The Mat Lab [10] tool as in fig-5 is used for unsupervised learning style.

Fig-5: Mat Lab tool page for implementation

4. User interface based Orange

The Orange [11] tool as in fig-6 is used for Reinforcement learning style.

Fig-6: Orange tool page for implementation

IV. RESULTS AND DISCUSSION

The proposed methodology gives the better results lot of unimaginable customer segmentation counts due to the impact of machine learning approach incorporation in the field of customer data analytics. The standard data sets from kaggle [7] and real-time are used for the implementation process. This research module gives 96% (480 out of 500 record sets from standard datasets) with the success rate for the heuristic approach in client structured web customer segmentation using machine learning techniques. The existing data mining approach gives only 335 out of 500 record sets with 67% of success rate which is low when compared with the proposed approach. The parametric comparison between existing and proposed methods with recall, f1 score etc. are represented in the below Table-6 format,

Table-6: Proposed methodology parametric comparisons

No	Approach	Accuracy	Precision	Recall	F1 score value
1	Data mining based customer segmentation approach	67%	0.68	0.69	0.66
2	Proposed heuristic approach for client structured web customer segmentation using machine learning techniques.	96%	0.95	0.97	0.96

The following fig-7 shows the performance comparison between the proposed and existing methodologies.
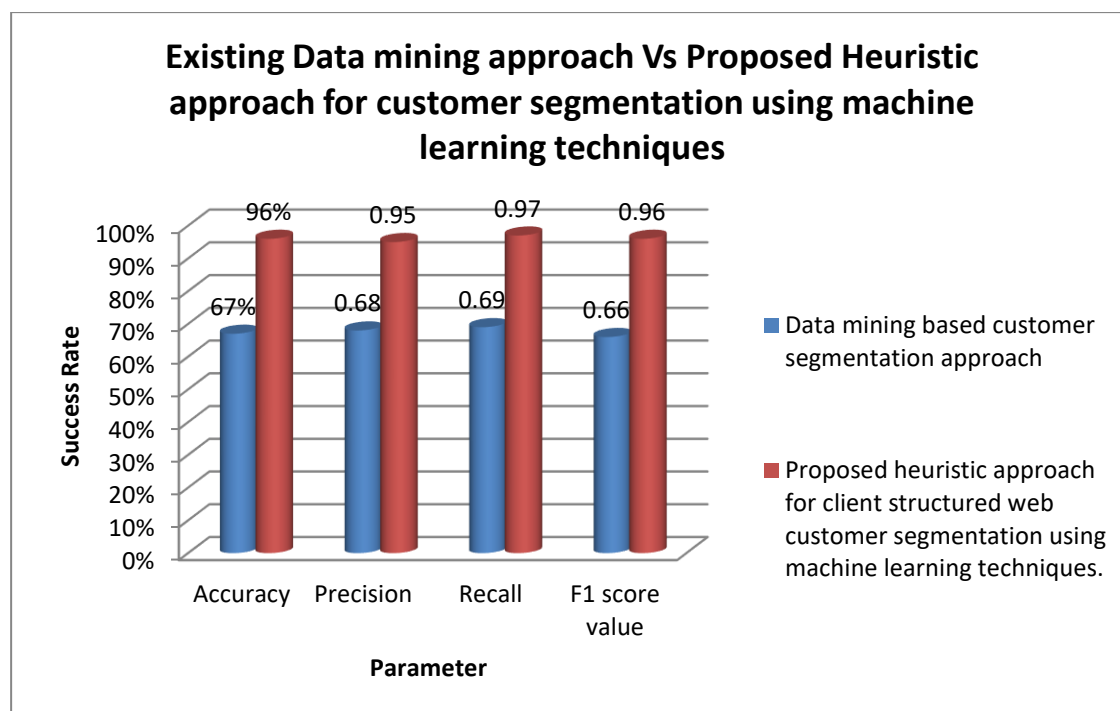
Fig-7: Proposed vs. existing methodology performance comparisons

V. CONCLUSION

The process of implementing customer segmentation is highly effective if it is incorporated with the machine learning domain due to its immense architectural design, learning techniques, decision making skills, and non-retentiveness of procedures.

The accuracy and the art of handling vast amount of information with or without knowledge about the data are the primary objectives I machine learning approaches.

The existing methodologies for customer segmentation directly applying the cluster algorithms do not produces good relevant results.

This research article proposes 5 stages of implementation starting with the customer segmentation category identification followed by the selection of machine learning approaches, then with the appropriate learning style selection, followed by the cluster counter and finally with the preferred tool implementation for the betterment of results.

This research article produced 96% 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.xlstat.com

[9] https://www.python.org/

[10] https://www.mathworks.com/products/matlab.html

[11] https://orangedatamining.com/