

A SURVEY ON CUSTOMER SEGMENTATION IN CLIENT STRUCTURED WEB DOMAIN USING MACHINE LEARNING TECHNIQUES

Adlin Selva Golda. V¹, Narayani. V²,

¹Research Scholar, Manonmaniam Sundaranar University, Tirunelveli,

²Assistant Professor/Computer Science, St Xavier's College, Tirunelveli

Article DOI: <https://doi.org/10.36713/epra9738>

DOI No: 10.36713/epra9738

ABSTRACT

Nowadays the direct communication between the users and resources are diminished which forces the online web domains to take the responsibility in order to maintain the entire business strategies. The entire web domains are maintaining their customer's information for the retaining purpose towards effective business development in the area of Education, Over the Top (OTT), Government services, and E-Commerce domains. The prolonged improvements in the entire business strategies critically depend upon the classification of customer segmentation. Machine learning helps to extract customer analysis information from huge sets of data. It is the procedure of data classification and segmentation of knowledge from customer data. This research can be considered as an efficient data segmentation of different source of customer information web mining domains to extract the useful segments for the enhanced time and space complexities. This paper presents a survey on customer segmentation in client structured web domain using machine learning techniques.

KEYWORDS—Segmentation, Web domain, Client structure, Machine learning, Performance

1. INTRODUCTION

Machine learning (ML) is the study of computer algorithms that can improve automatically through experience and by the use of data. It is seen as a part of artificial intelligence. Machine learning algorithms build a model based on sample data, known as training data, in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as in medicine, email filtering, speech recognition, and computer vision, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks.

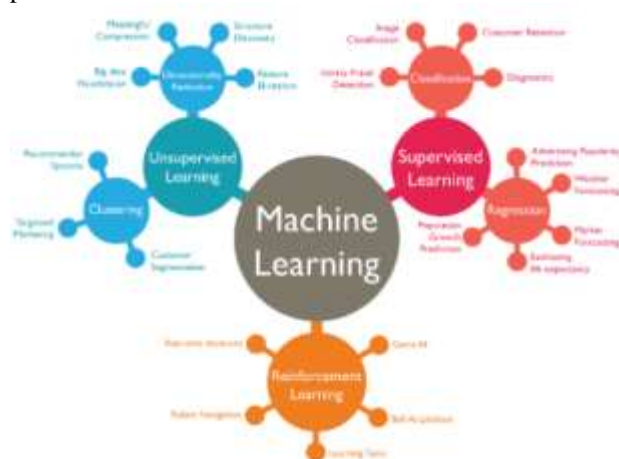


Fig-1: Machine Learning Components [12]

2. CUSTOMER SEGMENTATION

Customer segmentation is the process by which you divide your customers up based on common characteristics – such as demographics or behaviors, so you can market to those customers more effectively. These customer segmentation groups can also be used to begin discussions of building a marketing persona.

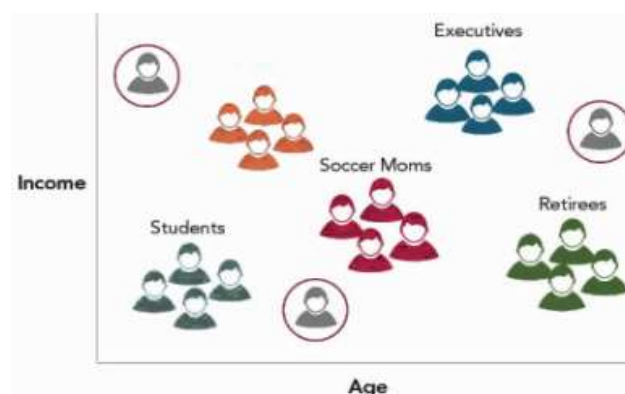


Fig-2: Customer Segmentation [13]

3. COMPONENTS OF MACHINE LEARNING

There are six main components of Machine Learning:

- ❖ Feature Extraction.
- ❖ Feature Selection.
- ❖ Choice of Algorithm.
- ❖ Training.



- ❖ Choice of Metrics/Evaluation Criteria.
- ❖ Testing.

4. BUILDING BLOCKS OF MACHINE LEARNING

- ❖ Machine Learning Building Block #1: Capturing the Input

Every machine learning system needs a lot of data to function. Ultimately, it will take decisions based on the data it captures. And it needs to capture data about the environment it is in, the ambient conditions, user inputs, and so on. Hence, the first building block of any machine learning or AI system is the way it captures and input in the system.

- ❖ Machine Learning Building Block #2: Processing and Storing the Data (Edge and Cloud)

Once capture the data from input units, we will need to either store it or run computations on it. That's basically the choice it boils down to when we're working on a machine learning project. Both of these (processing or storing) can either happen on the system typically called "AI on the Edge" or they can happen on the cloud.

- ❖ Machine Learning Building Block #3: Output or Interaction Unit

Finally, there would be an output or interaction unit in a successful AI or machine learning system. This is the unit where the machine learning system would interact with the outside universe and take action. This could be in the form of a display, voice output, or informal robotic actions. Usually, the output from our machine learning system would have several design considerations as well.

5. LITERATURE SURVEY

In 2018, Ahmad et al. [11] performed an ad hoc literature review, by selecting 166 research papers on the SO that were mainly classified about software development life cycle from the start of the SO website till the year 2016 positively. Similarly, the work of Baltadzhieva and Chrupala [1] thoroughly reviewed and analyzed various questions quality posted on diverse community question answering (CQA) websites like SO. In 2013, Meth et al. [2] conducted an SLR on investigating the works on automated requirements elicitation. Later on, Binkhonain and Zaho [3] conducted an SLR on ML algorithms for identifying and classifying NFRs. Recently, Iqbal et al. [4] presented a survey on ML algorithms and requirements engineering. They provided a bird's-eye view of how ML algorithms are aiding different requirements engineering activities. Besides, there are some surveys, SLR's, and systematic mapping studies done in other areas on sentiment analysis of scientific citations [5], data preprocessing methods for class imbalance problem [6], ML algorithms or techniques based software development effort estimation models [7], usability in agile software development [8, 9], and requirements prioritization [10], among others.

6. ALGORITHMS IN MACHINE LEARNING

Machine Learning is the study of learning algorithms using past experience and making future decisions. Although, Machine Learning has a variety of models, here is a list of the most commonly used machine learning algorithms by all data scientists and professionals in today's world.

- ❖ Linear Regression
- ❖ Logistic Regression
- ❖ Decision Tree
- ❖ Bayes Theorem and Naïve Bayes Classification
- ❖ Support Vector Machine (SVM) Algorithm
- ❖ K-Nearest Neighbor (KNN) Algorithm
- ❖ K-Means
- ❖ Gradient Boosting algorithms
- ❖ Dimensionality Reduction Algorithms
- ❖ Random Forest

7. ISSUES IN MACHINE LEARNING

Although machine learning is being used in every industry and helps organizations make more informed and data-driven choices that are more effective than classical methodologies, it still has so many problems that cannot be ignored. Here are some common issues in Machine Learning that professionals face to inculcate ML skills and create an application from scratch.

- ❖ Inadequate Training Data
- ❖ Poor quality of data
- ❖ Non-representative training data
- ❖ Over fitting and Under fitting
- ❖ Monitoring and maintenance
- ❖ Getting bad recommendations
- ❖ Lack of skilled resources
- ❖ Customer Segmentation
- ❖ Process Complexity of Machine Learning
- ❖ Data Bias
- ❖ Lack of Explain ability
- ❖ Slow implementations and results
- ❖ Irrelevant features

8. IMPLEMENTATION

- ❖ Study and perform background analysis of customer segmentation.
- ❖ Investigate integral functional strategies for web domains.
- ❖ Bridge the gap between computationally feasible and functionally relevant time scales.
- ❖ Improve multi resolution state prediction.
- ❖ Combine classical and modern techniques in machine learning environment.
- ❖ Sample larger sets of dynamical client structured web domains.
- ❖ Realize interactive modeling.
- ❖ Foster the development of machine learning model.



9. TOOLS PERFORMANCE

The following data illustrates machine learning tools with its implementation scope for the betterment of customer segmentation in client structured web domain system.

- 1) Scikit-learn
- 2) PyTorch
- 3) TensorFlow
- 4) Weka
- 5) KNIME
- 6) Colab
- 7) Apache Mahout
- 8) Accord.Net
- 9) Shogun
- 10) Keras.io
- 11) Rapid Miner

10. RECENT TRENDS

❖ Internet of Things

The first and foremost ML trends, for which the majority of computer workers are anxiously anticipating in IoT. A breakthrough in this area will have a big impact on 5G adoptions as it will become the foundation for IoT. Systems will be able to receive and send information at a faster rate because of 5G's incredible network speed. Other machines on the system can be connected to the internet via IoT devices. Every year, we see a large increase in the number of IoT devices linked to the network, resulting in a commensurate increase in the amount of data exchanged.

❖ Automated machine learning

Professionals can design effective tech models that assist them to improve production and efficiency by applying automated machine learning. As a result, we will observe the majority of advancements in the domain of effective task solving. AutoML is mostly used to generate highly sustainable concepts that can aid in the derivation of job efficiency, particularly in the development sector, where experts can develop apps without having much programming skills.

❖ Improved cyber security

With the advancement of technology, most apps and devices have become smart, resulting in significant technological advancement. However, because these smart devices are continually hooked up to the internet, there is a pressing need for them to be more secure. Tech pros may utilize machine learning to create anti-virus models that will block any possible cyber-attacks and reduce dangers.

❖ Ethics in AI

With the advancement of new technologies such as machine learning and artificial intelligence, there is a growing worry about defining some ethical guidelines for these technologies. The more advanced the technology, the more advanced the ethics should be. If these ethics are not followed, machines will be unable to perform efficiently, resulting in poor decisions. This is visible in the self-driving automobiles that are currently on the market. The failure of the self-driving automobile is due to the embedded artificial intelligence, which is the vehicle's brain.

❖ Natural speech understanding automation

A lot of information on home automation is being disseminated, which theoretically works on smart speakers. The procedure is eased because of the use of smart voice assistants like Google, Siri, and Alexa, which establish a connection with smart appliances via non-contact control. In terms of detecting human sounds, these computers already have a high level of accuracy.

❖ General adversarial networks

GANs, or General Adversarial Networks, are new ML trends that produce samples that must be reviewed by networks that are selective in nature and can delete any type of undesired content. GAN, like the government, has numerous branches that provide checks and balances to ensure accuracy and trustworthiness.

Businesses must innovate in order to achieve their objectives, and they must develop new and innovative ways to do it using technology. ML is the wave of the future, and every company is adjusting to this new technology.

11. APPLICATIONS

- Image Recognition.
- Speech Recognition.
- Traffic prediction.
- Product recommendations.
- Self-driving cars.
- Email Spam and Malware Filtering.
- Virtual Personal Assistant.
- Online Fraud Detection.
- Stock Market trading.
- Medical Diagnosis.
- Automatic Language Translation.

12. CONCLUSION

The detailed research survey in the field of web domain towards customer segmentation system with traditional approach when compared to the machine learning approaches and recent techniques with advanced tools shows that the higher level of impact in the field of machine learning in the customer segmentation of web domain approaches with the cope up towards latest trends and systematic pathways for the improvement progress of several advanced strategies. The approaches for machine learning dealt with the various levels of implications towards the selection strategies for the analysis and prediction of customer segmentation focusing along with the factors in the development of customer segmentation system implementations. The tools performance and applications of machine learning provides the several directions for the development of different methodologies to implement in the better way. In future this research will lead the direction of customer segmentation system for the E-commerce in an effective way.

REFERENCES

1. A. Baltadzhieva and G. Chrupala, "Question quality in community question answering forums," *ACM SIGKDD Explorations Newsletter*, vol. 17, no. 1, pp. 8–13, 2015.



2. H. Meth, M. Brhel, and A. Maedche, "The state of the art in automated requirements elicitation," *Information and Software Technology*, vol. 55, no. 10, pp. 1695–1709, 2013.
3. M. Binkhonain and L. Zhao, "A review of machine learning algorithms for identification and classification of non-functional requirements," *Expert Systems with Applications*, vol. 1, 2019.
4. T. Iqbal, P. Elahidoost, and L. Lúcio, "A bird's eye view on requirements engineering and machine learning," in *Proceedings of the 2018 25th Asia-Pacific Software Engineering Engineering Conference (APSEC)*, pp. 11–20, Nara, Japan, December 2018.
5. A. Yousif, Z. Niu, J. K. Tarus, and A. Ahmad, "A survey on sentiment analysis of scientific citations," *Artificial Intelligence Review*, vol. 52, no. 3, pp. 1805–1838, 2019.
6. H. Ali, M. N. M. Salleh, K. Hussain et al., "A review on data preprocessing methods for class imbalance problem," *International Journal of Engineering & Technology*, vol. 8, pp. 390–397, 2019.
7. J. Wen, S. Li, Z. Lin, Y. Hu, and C. Huang, "Systematic literature review of machine learning based software development effort estimation models," *Information and Software Technology*, vol. 54, no. 1, pp. 41–59, 2012.
8. D. A. Magües, J. W. Castro, and S. T. Acuna, "HCI usability techniques in agile development," in *Proceedings of the 2016 IEEE International Conference on Automatica (ICA-ACCA)*, pp. 1–7, Curico, Chile, October 2016.
9. D. A. Magües, J. W. Castro, and S. T. Acuña, "Usability in agile development: a systematic mapping study," in *Proceedings of the 2016 XLII Latin American Computing Conference (CLEI)*, pp. 1–8, Valparaiso, Chile, October 2016.
10. P. Achimugu, A. Selamat, R. Ibrahim, and M. N. r. Mahrin, "A systematic literature review of software requirements prioritization research," *Information and Software Technology*, vol. 56, no. 6, pp. 568–585, 2014.
11. A. Ahmad, C. Feng, S. Ge, and A. Yousif, "A survey on mining stack overflow: question and answering (Q&A) community," *Data Technologies and Applications*, vol. 52, no. 2, pp. 190–247, 2018.
12. <https://towardsdatascience.com/machine-learning-algorithms-in-laymans-terms-part-1-d0368d769a7b>
13. <https://www.youtube.com/watch?v=zPJtDohab-g>