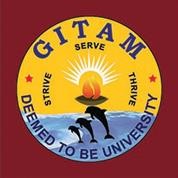
**GANDHI INSTITUTE OF TECHNOLOGY AND MANAGEMENT**

(Deemed to be University)



A Mini Project Report

on

# “Customer Segmentation

# Using

# Machine Learning”

Submitted in partial fulfillment of the requirement for the degree of

## **Bachelor of Technology**

## **in**

**Computer Science and Engineering**

#### Submitted by:

**G. Vishnu Sai (321810301045)**

Under the guidance of **Dr. Chethana Tukkoji**

Assistant professor

**Department of Computer Science & Engineering, School of Technology, GST**

**2021-22**



**GITAM**

### (DEEMED TO BE UNIVERSITY)

**Bengaluru-**561203

Department of Computer Science & Engineering

**Certificate**

This is to certify that the Mini Project titled **“Customer Segmentation Using Machine Learning”** is the bonafide work carried out by **G. Vishnu Sai (321810301045** student of B.Tech (CSE) of GITAM Deemed to be University, Bengaluru campus during the academic year 2021-2022, in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology (Computer Science and Engineering). It is certified that all corrections/suggestions indicated for Internal Assessment have been in corporated in the Report deposited in the departmental library.

Signature of the Guide Signature of HOD

#### **Dr. Chethana Tukkoji** **Dr. Vamsidhar Y**

Assistant professor Head of the Department

**Acknowledgement**

The Satisfaction and Euphoria that accompany the successful completion of any Mini Project would be incomplete without mentioning the people who made it possible, whose constant guidance and encouragements crowned our efforts with success. We take this opportunity to express the deepest gratitude and appreciations to all those who held us directly or indirectly towards the successful completion of the Mini Project.

I would like to thank **Dr. S DINESH**, Ph.D., Director, GITAM School of Technology for all the facilities provided.

I would like to thank **Dr. VAMSIDHAR Y**, Ph.D., Professor-HOD, Department of **Computer Science and Engineering** for his support and encouragement that went a long way in successful completion of this Mini Project.

I consider that as privilege to express our heartfelt gratitude and respect to **Dr . Chethana Tukkoji ,** Assistant professor, department of Computer Science and Engineering for being our internal guide and **Mrs. Benna Pillai** for being our Project coordinator for his integral and incessant support offered to us throughout the course of this project and for constant source of inspiration throughout the Mini Project.

I would like to thank our **parents** and our **friends** for their support and encouragement in the completion of project in time.

Last but not least I would like to thank all the teaching and non-teaching staff members **Departments** of the **Computer Science and Engineering,** for the support in completion of the Mini Project in time.

**Abstract:**

Ecommerce transactions are no longer a new thing. Many people shop with ecommerce and many companies use ecommerce to promote and to sell their products. Because of that, overloading information appears on the customers’ side. Overloading information occurs when customers get too much information about a product then feel confused. Personalization will become a solution to overloading problem. In marketing, personalization technique can be used to get potential customers in a case to boost sales. The potential customer is obtained from customer segmentation or market segmentation. This paper will review customer segmentation using data, methods and process from a customer

segmentation research. The data for customer segmentation were divided into internal data and external data. Customer profile and purchase history were treated as the internal data while server log, cookies, and survey data were as the external data. These data can be processed using one of several methods: Business Rule, Magento, Customer Profiling, Grouping, Supervised Clustering, Customer Likeness Clustering, Purchase Affinity Clustering and Unsupervised Clustering. Those methods were classified into Simple technique, Target technique, and Unsupervised technique and the process was generalized in determining the business objective, collecting data, data preparation, variable analysis, data processing, and performance evaluation. Customer behavior in accessing ecommerce when viewing a product on ecommerce was recorded in server log with time. Duration when seeing the product can be used as customer interest in the product so that it can be used as a variable in customer segmentation.

**TABLE OF CONTENTS**

**S.no Chapter Name Page No.**

1. Introduction 7-8

2. Literature survey 8-10

3. Customer Segmentation 10-11

4. Clustering and K-Means Algorithm 11-12

5. Problem Statement 12

6. Process of customer segmentation 12-13

7. Working methodologies 13-14

8. Advantages of customer segmentation 14-15

9. Disadvantages of customer segmentation15-17

10. Hardware and software requirements 17

11. Tools Required 18

12. Packages Imported 18-19

13. Conclusion 20

14. References 21

1.Introduction:

Ecommerce development began when the internet is growing and growing until today, especially in B2C ecommerce (Business to Customer). When shopping use ecommerce, a user finds it easy and faster. The ease of using ecommerce encourages customers to buy using ecommerce. With these conditions the problem that comes up is the overloading information because of many products offered by ecommerce. Overloaded information can be overcome by an implementation of personalization in ecommerce services such as providing product recommendation, links recommendation, ads or text and graphics that correspond to the users’ characteristics and needs. In addition to solving the problem of overloaded information, personalized services in ecommerce can maintain customer loyalty of existing customer, getting new customers by providing service to customers in accordance with their needs and characteristics. It will generate more profits for the company.

Before the personalization is implemented, customer segmentation should be conducted because the result from customer segmentation process will be used as inputs to personalize ecommerce services, resulting in dynamic personalization ecommerce services based on current customer conditions.

Customer segmentation is currently performed by processing customer database, i.e., demographic data or purchase history. Several researchers discuss the customer segmentation method on their papers, such as Magento, who used several variables to perform customer segmentation, namely transaction variable, product variable, geographic variable, hobbies variable and page viewed variable; Baer and Colica discuss customer segmentation methods of Business Rule, Quantile membership, Supervised Clustering, Unsupervised Clustering, Customer Profiling, Customer Likeness Clustering and Purchase Affinity Clustering. Some of these methods have similarity. Other researchers discuss the implementation of customer segmentation. This paper will classify customer segmentation methods based on data processing.

**2.Literature survey**

Customers can have different types of characteristics and can be of different importance to a company. For companies to know which customers are of significance, a segmentation of customers need to be done (McDonald & Dunbar, 2012).

The theory of segmentation is the process where identifying characteristics of different customers and dividing them into groups. What companies often do when segmenting their customers is to divide them based on how much revenue they contribute to the company based on their purchase volumes (Batt, 2000)

Identifying and classifying customers leads to a better understanding of who the customers are and what type of demand the customers require. Some customer groups can have a high degree of innovation where changes within the customer group over time often occur. For these types of customers, you need to be aware of the requirement changes to meet the customer demand in the best way which also fulfils the customer needs (Bottcher, Spott, Nauck & Kruse, 2009)

When a company has the right knowledge about the customer requirements it will give them the ability to easier divide the customers into segmentation groups. Furthermore, the company can easier find out what satisfies their customers and even surprise them. This kind of information can be used for further improvements into their services or products. These days, customer service is as important for the customers as the actual product or service, and it is important that the companies have this part set. Finally, segmenting customers can simplify the choices of how much and what the company should put emphasis on when it comes to the degree of services that the different groups should get (Buttle, 2009).

A company provides the market with either a service or a product. Because of this it is vital according to (Fang, Palmatier & Steenkamp 2008) for a company to reach the customer service elements to please their customers. Well established service companies have the right skillset and right knowledge to fulfil the demands, expectations and needs of their customers (Mattson, 2004). The concept of customer service can be defined as what a company does to include the purchasers, sellers and other groups that can boost their product or service. A successful customer segmentation within services benefits the company to enhance their relationship with their purchasers and sellers which also contributes to an enhanced competitiveness (Pauline, 2009).

Out in the market there are a huge range of service providing companies with many different types of customers. To be competitive and meet the customer demand in best way, service companies have different strategies of how to target and segment their customers (Anderson, Narus & Narayandas 2009)

**3. Customer Segmentation:**

Over the years, as there is very strong competition in the business world, the

organizations have to enhance their profits and business by satisfying the demands of their customers and attract new customers according to their needs. The identification of customers and satisfying the demands of each customer is a very complex and tedious task. This is because customers may be different according to their demands, tastes, preferences and so on. Instead of “one-size-fits-all” approach, customer segmentation clusters the customers into groups sharing the same properties or behavioral characteristics. According to, customer segmentation is a strategy of dividing the market into homogenous groups. The data used in customer segmentation technique that divides the customers into groups depends on various factors like, data geographical conditions, economic conditions, demographical conditions as well as behavioral patterns. The customer segmentation technique allows the business to make better use of their marketing budgets, gain a competitive edge over their rival companies, demonstrating the better knowledge of the needs of the customer. It also helps an organization in, increasing their marketing efficiency, determining new market opportunities, making better brand strategy, identifying customers retention

**4.Clustering and K-Means Algorithm**

Clustering algorithms generates clusters such that within the clusters are similar based on some characteristics. Similarity is defined in terms of how close the objects are in space. K-means algorithm in one of the most popular centroid based algorithm. Suppose data set, D, contains n objects in space. Partitioning methods distribute the objects in D into k clusters, C1,...,Ck , that is, Ci ⊂ D and Ci ∩Cj = ∅ for (1 ≤ i, j ≤ k). A centroid-based partitioning technique uses the centroid of a cluster, Ci, to represent that cluster. Conceptually, the centroid of a cluster is its center point. The difference between an object p ∈ Ci and ci, the representative of the cluster, is measured by dist(p,ci), where dist(x,y) is the Euclidean distance between two points x and y.

**4.1Algorithm**: The k-means algorithm for partitioning, where each cluster’s center is represented by the mean value of the objects in the cluster. Input: k: the number of clusters, D: a data set containing n objects.

Output: A set of k clusters. Method:

(1) arbitrarily choose k objects from D as the initial cluster centers;

(2) repeat

(3) (re)assign each object to the cluster to which the object is the most similar, based on the mean value of the objects in the cluster;

(4) update the cluster means, that is, calculate the mean value of the objects for each cluster;

(5) until no change.

**5.Problem Statement:**

Customer segmentation is a technique in which we divide the customers

based on their gender, age, annual income, spending score, etc.

It is useful to get this information so that the store can get help in

personalize marketing and provide customers with relevant deals.

With the help of this project, companies can run user-specific campaigns

and provide user-specific offers rather than broadcasting same offer to all

the users.

**6.Process of customer segmentation**

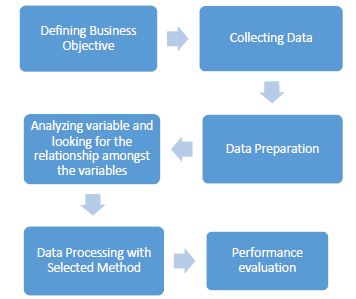
****

Fig.1-process

In the above fig.1, it explains the following the steps. The following steps are

1. Design A Proper Business Case Before You Start.
2. Collect data
3. Prepare the Data.
4. Performing Segmentation Using k-Means Clustering.
5. Visualization of the Results.

**7.Working methodologies:**

* Customer segmentation Dataset is taken.
* The data is preprocessed to clean the data and understand the dataset.
* The preprocessed data is divided as training and testing data.
* The model is built using machine learning algorithms like Logistic Regression
* The model is trained using training dataset and once the model has been trained successfully it has to be tested.
* The trained model is tested using testing dataset and accuracy is calculated.
* The algorithm which gives the best accuracy is taken as our final prediction model.
* The finalized model is converted into pickle model (binary format data) and saved.
* Finally, the predicted output is displayed on the jupyter notebook.

**8.Advantages of customer segmentation:**

Implementing customer segmentation leads to plenty of new business opportunities. You can do a lot of optimizations in:

**1.Budgeting:**

Nobody likes to invest in campaigns that don’t generate any new customers. Most companies don’t have huge marketing budgets, so that money has to be spent right. Segmentation enables you to target customers with the highest potential value first, so you get the most out of your marketing budget.

**2.Product Design:**

Customer segmentation helps you understand what your users need. You can identify the most active users/customers, and optimize your application or offer towards their needs.

**3**.**Promotion:**

Properly implemented customer segmentation helps you plan special offers and deals. Frequent deals have become a staple of e-commerce and commercial software in the past few years. If you reach a customer with just the right offer, at the right time, there’s a huge chance they’re going to buy. Customer segmentation will help you tailor your special offers perfectly.

**4.Marketing:**

The marketing strategy can be directly improved with segmentation because you can plan personalized marketing campaigns for different customer segments, using the channels that they use the most.

**5.Customer Satisfaction:**

By studying different customer groups, you learn what they value the most about your company. This information will help you create personalized products and services that perfectly fit your customers’ preferences.

**9.Disadvantages of customer segmentation:**

#### **1. Limited Production:**

In each specific segment, customers are limited. So, it is not possible to produce products in mass scale for every segment. Therefore, company cannot take advantages of mass scale production; scale of economy is not possible. Product may be costly and affect adversely to the sales.

#### **2. Expensive Production:**

Customer segmentation is expensive in both production and marketing. In order to satisfy different groups/segments of buyers, producers have to produce products of various models, colors, sizes, etc., that result into more production costs. In the same way, the producers are required to maintain large inventory for different styles, colors, and sizes of products.

#### **3. Expensive Marketing:**

Customer segmentation also results into expensive marketing. Due to different groups of buyers, the marketer has to consider all the segments in terms of needs, interests, habits, preferences and attitudes. Marketer has to formulate and implement several marketing strategies for different segments.

#### **4. Difficulty in Distribution:**

Company needs to make the separate arrangement for each of the products demanded by different classes of customers. Salesman’s recruitments, selection, training, payments, and incentives are more difficult and costly. Company has to maintain separate channels and services for satisfying varied customer groups.

#### **5. Heavy Investment:**

Customer segmentation leads to heavy investment. In order to satisfy different needs and wants of various groups, a company has to produce variety of product lines and product items. For the purpose, the company requires to invest more on technology and other inputs that may demand heavy investment.

#### **6. Promotion Problems:**

Customer segmentation also creates promotional problems and multiplies promotional difficulties. It is obvious that different segments are made on the basis of distinguished characteristics of buyers. Each group differs in terms of advertising media, appeal or message. In order to influence various segments of buyers, the company is required to prepare a separate advertising programme or strategy. Similarly, personal selling and sales promotional activities become more complex. Company needs to spend more to take benefits of specialization.

#### **7. Stock and Storage Problems:**

To meet needs and wants of different consumer groups, the company must maintain adequate stock of various products on a continuous basis. This creates problem of stocks, storage, and working capital. Most limitations reflect the impact of situation and inability of manager to segment the market purposively and meaningfully. But, limitations cannot restrict segmentation philosophy and practice. These limitations can be overcome by segmenting market carefully and objectively.

**10.Hardware and software requirements**

**Hardware Requirement:**

* System: Pentium IV 2.4 GHz.
* Hard Disk: 500 GB.
* Ram: 4 GB.
* Any desktop / Laptop system with above configuration or higher level.

**Software Requirements:**

* Operating system: Windows XP / 7
* Coding Language: Python
* Version: Jupyter notebook
* IDE: Jupyter notebook
* ML APIS: NumPy, Pandas, Sklearn, matplotlib, seaborn,
* ML Algorithm: Logistic Regression

**11.TOOLS REQUIRED:**

* **Jupyter Notebook:** Jupyter Notebook provides you with an easy-to-use, interactive data science environment
* **Python:** Python is an interpreted high-level general-purpose programming language. Its design philosophy emphasizes code readability with its use of significant indentation.
* **Excel – customer reviews:** Microsoft Excel is a spreadsheet developed by Microsoft for Windows, macOS, Android and iOS. It features calculation, graphing tools, pivot tables, and a macro programming language called Visual Basic for Applications.

**12.PACKAGES IMPORTED:**

1. **Pandas:**

* Pandas is a Python library used for working with data sets. It has functions for analysing, cleaning, exploring, and manipulating data.

1. **Matplotlib:**

* Matplotlib is a low-level graph plotting library in python that serves as a visualization utility.

**3.Seaborn:**

* Seaborn is a library that uses Matplotlib underneath to plot graphs. It will be used to visualize random distributions.

**4.NumPy:**

* NumPy is a Python library used for working with arrays. It also has functions for working in domain of linear algebra, Fourier transform, and matrices.

**5.Nltk:**

* The Natural Language Toolkit, or more commonly NLTK, is a suite of libraries and programs for symbolic and statistical natural language processing (NLP) for English written in the Python programming language.

**6.Keras:**

* Keras is a high-level neural networks library, written in Python and capable of running on top of either TensorFlow or Theano. It was developed with a focus on enabling fast experimentation. Being able to go from idea to result with the least possible delay is key to doing good research.

**7.TensorFlow:**

* TensorFlow is an end-to-end open-source platform for machine learning. TensorFlow is a rich system for managing all aspects of a machine learning system; however, this class focuses on using a particular TensorFlow API to develop and train machine learning models.

13.Conclusion:

Customer segmentation is a way to improve communication with the customer, to know the wishes of the customer, customer activity so that appropriate communication can be built. Finding an optimal number of unique customer groups will help you understand how your customers differ, and help you give them exactly what they want. Customer segmentation improves customer experience and boosts company revenue. That’s why segmentation is a must if you want to surpass your competitors and get more customers. Doing it with machine learning is definitely the right way to go.

**14. References**

[1] Al-Qaed F, Sutcliffe A. Adaptive Decision Support System (ADSS) for B2C E-Commerce. *2006 ICEC Eighth Int Conf* *Electron Commer Proc NEW E-COMMERCE Innov Conqu* *Curr BARRIERS, Obs LIMITATIONS TO Conduct Success* *Bus INTERNET*. 2006:492-503.

[2] Mobasher B, Cooley R, Srivastava J. Automatic Personalization Based on Web Usage Mining. *Commun* *ACM*. 2000;43(8).

[3] Cherna Y, Tzenga G. Measuring Consumer Loyalty of B2C e-Retailing Service by Fuzzy Integral: a FANP-Based Synthetic Model. In: *International Conference on Fuzzy Theory and Its Applications iFUZZY*.; 2012:48-56.

[4] Magento. An Introduction to Customer Segmentation. 2014. info2.magento.com/.../ An\_Introduction\_to\_Customer\_Segmentation...

[5] Baer D. CSI : Customer Segmentation Intelligence for Increasing Profits. *SAS Glob Forum*. 2012:1-13.

<http://support.sas.com/resources/papers/proceedings12/103-> 2012.pdf.

[6] Colica R. Customer Segmentation And Clustering Using SAS Enterprise Minner Part I The Basics. 2011:1-14.

[7] Schneider G. *Electronic Commerce, 9th Edition*.; 2013:643. doi:10.1002/1521-3773(20010316)40:6<9823::AIDANIE9823> 3.3.CO;2-C.

[8] Venkatesan R. Cluster Analysis For Segmentation. 2007.

[9] Ezenkwu CP, Ozuomba S. Application of K-Means Algorithm for Efficient Customer Segmentation : A Strategy for Targeted Customer Services. 2015;4(10):40-44.

[10] Lieberman M. Target “ golden egg ” consumer to achieve maximum ROI. 2009;(May):50-51.

[11] Dodwell A. Effective Marketing email startegy \_ Segmentation RFM. 2015. <http://www.sailthru.com/marketing-blog/written-effectiveemail-> marketing-strategies segmentation-rfm/