INTERNATIONAL JOURNAL OF

MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.521

Customer Segmentation in E-commerce using Clustering Algorithms

Ms.Zuha Shoaib Khan, Ms.Snehal Nandayya Swami, Mr.Neel Satyajeet Kalshetty,

Mr.Darshan Surykant Lingraj, Mrs. Yogini Prasad Patil

Diploma Student, Department of Computer Engineering, A.G.Patil Polytechnic Institute, Solapur, Maharashtra, India

Diploma Student, Department of Computer Engineering, A.G.Patil Polytechnic Institute, Solapur, Maharashtra, India

Diploma Student, Department of Computer Engineering, A.G.Patil Polytechnic Institute, Solapur, Maharashtra, India

Diploma Student, Department of Computer Engineering, A.G.Patil Polytechnic Institute, Solapur, Maharashtra, India

Lecturer, Department of Computer Engineering, AG Patil Polytechnic Institute, Solapur, Maharashtra, India

ABSTRACT: Customer Relationship Management(CRM) has always played a crucial role as a market strategy for providing organizations with the quintessential business intelligence for building, managing and developing valuable long-term customer relationships. A number of business enterprises have come to realize the significance of CRM and the application of technical expertise to achieve competitive advantage. This study explores the importance of Customer Segmentation as a core function of CRM as well as the variousmodels for segmenting customers using clustering techniques. The available clustering models for customer segmentation, in general, and the major models of K-Means and Hierarchical Clustering.

I.INTRODUCTION

In the contemporary day and age, the importance of treating customers as the principal asset of an organization is increasing in value. Organizations are rapidly investing in developing strategies for better customer acquisition, maintenance and development. The concept of business intelligence has a crucial role to play in making it possible for organizations to use technical expertise for acquiring better customer insight for outreach programs. In this scenario, the concept of CRM garners much attention since it is a comprehensive process of acquiring and retaining customers, using business intelligence, to maximize the customer value for a business enterprise

One of the two most important objectives of CRM is customer development through customer insight. This objective of CRM entails the usage of an analytical approach in order to correctly assess customer information and analysis of the value of customers for better customer insight. Keeping up with the changing times, organizations are modifying their business flow models by employing systems engineering as well as change management and designing information technology(IT) solutions that aid them in acquiring new customers, help retain the present

Clustering techniques reveal internally homogeneous and externally heterogeneous groups. Customers vary in terms of behavior, needs, wants and characteristics and the main goal of clustering techniques is to identify different customer types and segment the customer base into clusters of similar profiles so that the process of target marketing can be executed more efficiently. This study aims to explore the avenues of using customer segmentation, as a business intelligence tool within the CRM framework as well as the use of clustering techniques for helping organizations redeem a clearer picture of the valuable customer base.

II.LITERATURE REVIEW

Li, Zeying have proposed a method in which a retail supermarket was taken as research object, and data mining methods was used to retail enterprise customer segments, and then association rules obtained using Apriori algorithm were used to different groups of customers and get rules about customer characteristics to make customer characteristic analysis efficiently. Finally, the author gave some references to the supermarket's marketing and management work, which helped in understanding it in detail. Data mining was used efficiently to deal with the large number of historical and current data, from the database to find some potential, useful and valuable information for the retail stores which help us target customers.

Wang, Zhenyu, Yi Zuo, Tieshan Li, CL Philip Chen, and Katsutoshi Yada have analyzed customer segmentation based on broad learning system which provides an alternative view of learning in deep structure. Firstly, in addition to customer purchasing behavior, RFID (Radio Frequency Identification) data was also included, which can accurately represent the consumers' in-store behavior. Secondly, this paper used Broad Learning System (BLS) to analyze the consumer segmentation. BLS is one of the finest machine learning techniques, and quite efficient and effective for classification tasks. Thirdly, the customer behavior data used in this paper was collected from a real-world supermarket in Japan. Customer segmentation was considered as a multi-label classification problem based on both of POS data and RFID data.

Kansal, Tushar, Suraj Bahuguna, Vishal Singh, and Tanupriya Choudhury performed customer segmentation using K-means clustering. A python program was developed and the program was trained by applying standard scaler onto a dataset having two features of 200 training sample taken from local retail shop. Both the features are the average of the amount of shopping by customers and average of the customer's visit into the shop annually. By applying clustering, 5 segments of cluster were formed labelled as Careless, Careful, Standard, Target and Sensible customers. However, the authors got two new clusters on applying mean shift clustering labelled as High buyers and frequent visitors and High buyers and occasional visitors.

Bhade Kalyani, Vedanti Gulalkari, Nidhi Harwani and Sudhir N Dhage have proposed a systematic approach for targeting customers and providing maximum profit to the organizations. An important initial step was to analyze the data of sales acquired from the purchase history and determine the parameters that have the maximum correlation. Based on respective clusters, proper resources can be assigned towards profitable customers using machine learning algorithms. K-Means clustering was used for customer segmentation and Singular Value Decomposition was used for providing appropriate recommendations to the customers. This paper also deals with the drawbacks of the recommender system like sparsity, cold start problem etc and how they can be overcome

III. METHODOLOGY OF PROPOSED SURVEY

Customer Relationship Management

Customer Relationship Management is an important business approach for developing and securing steady, long-term customer associations. The modern marketing approach promotes the usage of CRM as part of the organization's business strategy for enhancing customer service satisfaction. CRM enables business enterprises in customer value analysis as well as the targeting of those customers that prove of greater value. It also helps business organizations in developing high-quality and long-term customer-company relationships that increase loyalty and profits. An accurate evaluation of customer profitability and the targeting of highvalue customers are important factors that contribute to the success of CRM

Customer Segmentation

As the market is widening, the rate of competition between all business entities is rapidly growing. Hence, these business enterprises are increasing their expenditure on their marketing strategies to achieve competitive advantage . In this context, the significance of employing Information Technology(IT) solutions to marketing campaigns emerges as a pivotal step in a modern approach to business. Customer Segmentation is a popular technique of partitioning the customer base into externally distinct and internally uniform groups in order to create

varied marketing strategies for targeting each group according to its characteristics. Generally speaking, it is defined as the process whereby the consumers of a business enterprise are divided into groups according to their preferences, characteristics and purchasing behavior.

KEY TAKEWAYS –

•Customer segmentation will help your business increase brand awareness, customer retention, and new sales. It also provides customer data that enables you to offer more relevant ad targeting and improve your product's features and pricing

•It's crucial to set goals and objectives that guide your customer segmentation and marketing efforts. This is especially relevant when deciding between user vs. customer segmentation. Do you want to gather data on users who interact with your product or service or those in charge of purchasing your product or service?

•Regardless of your business type, customer segmentation typically involves grouping customers according to demographics, geographics, psychographics, or behavior.

•Digital companies can gain customer insights through techno graphic, behavioral, needs-based, value-based, or customer status segmentation.

•Customer segmentation mistakes to avoid include creating segments that are too narrow and not evolving or testing your segments over time.

RESEARCH_METHODOLOGY
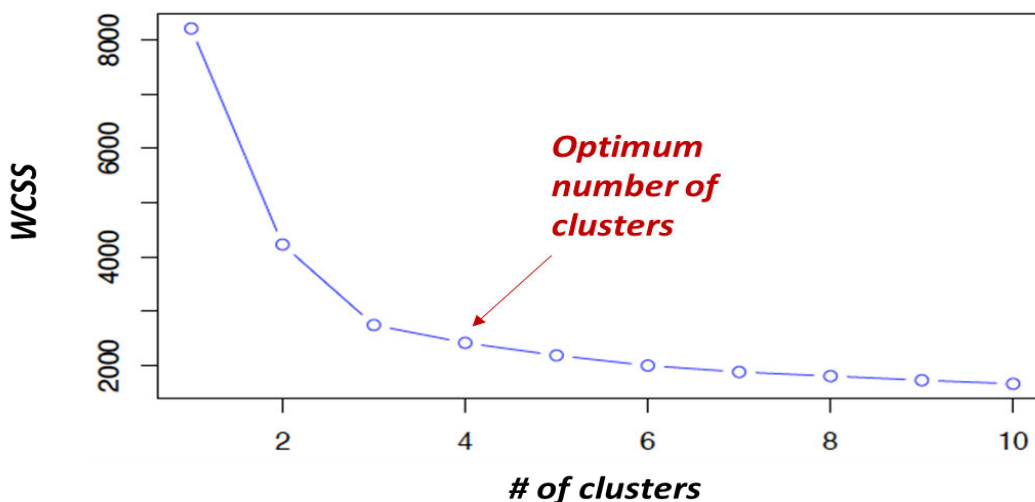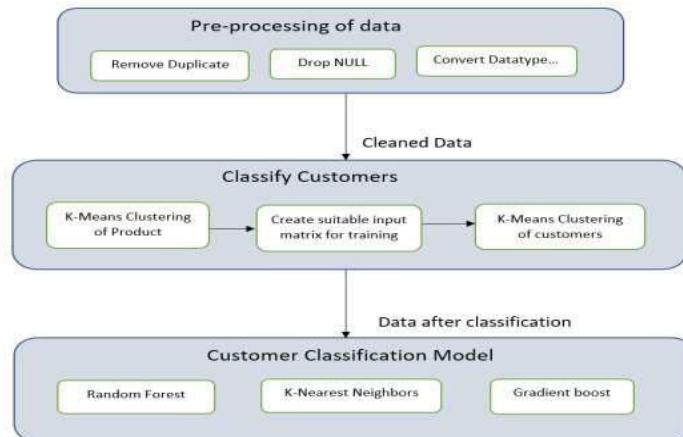
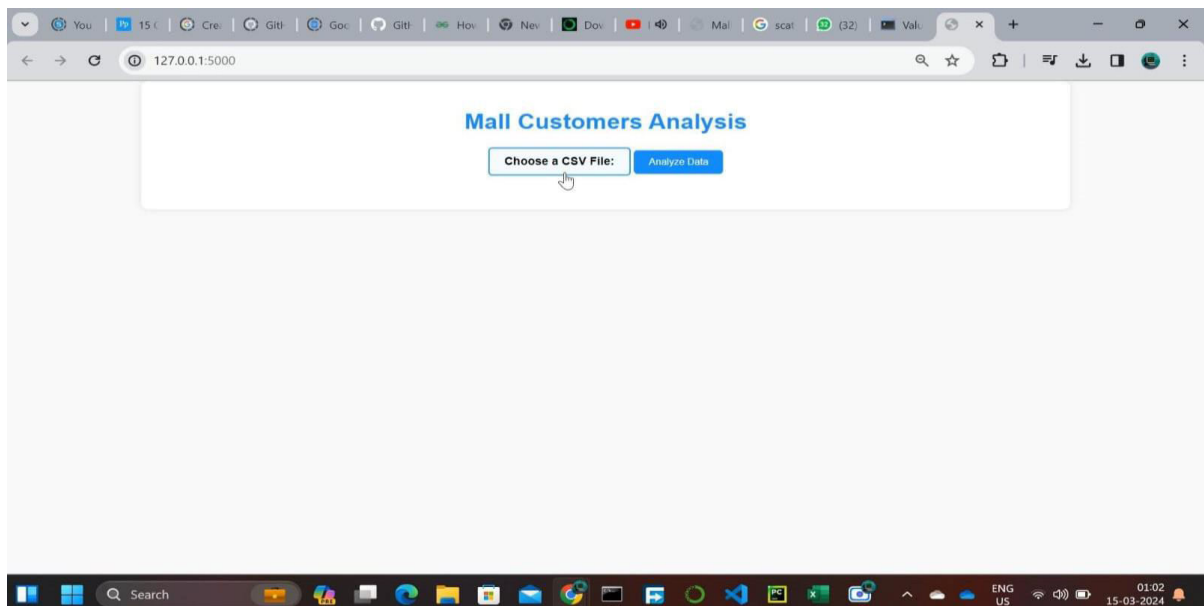### Data Pre-processing:

The dataset was obtained from a machine learning repository in University of California **.** Initially it was found that the dataset has 5,41,909 rows and 8 columns (Invoice No, Stock Code, Description, Quantity, Invoice Date, Unit Price, Customer ID, Country). After removing all the NULL values, the data frame was of 4,06,829 rows and 8 columns. It was observed that almost 25% of customer ID entries were NULL. It was then observed that there were 5,225 duplicate rows in the data frame. It was then observed that there were 37 different countries in the dataset and plotted a choropleth map for the same showing the orders per country. It was observed that 16.47% of the orders were cancelled. Some entries had negative values for quantity so it was assumed that when an entry is cancelled in the dataset, an entry with same Customer ID, description is added but with a negative value of quantity so it was decided to locate the entries. The dataset had discount entries, which got discarded later. Even then the dataset had some negative quantity entries which proved our assumption to be wrong. It was because the dataset has values from 2010 but the remaining values are those which were made before 2010 but cancelled after 2010.

IV. RESULTS

Applications

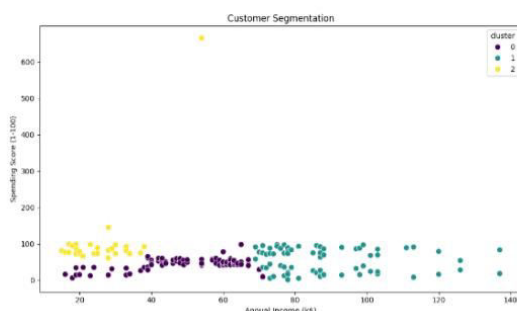**Data Collection**: Gather relevant data such as purchase history, browsing behavior, demographics (age, gender, location), and any other relevant information.

Data Preprocessing: Clean the data, handle missing values, and normalize or scale the features as needed to prepare it for clustering.

Feature Selection: Identify which features will be used for clustering, such as frequency of purchases, average order value, etc.

Clustering Algorithm Selection: Choose appropriate clustering algorithms such as K-means, hierarchical clustering, or DBSCAN based on the nature of the data and the goals of the segmentation.

Clustering: Apply the chosen algorithm to the data to cluster customers into distinct groups based on similarities in their purchasing behavior and demographics.

Interpretation: Analyze the clusters to understand the characteristics of each segment, such as high-spending customers, frequent buyers, etc.

Customer Profiling: Create profiles for each cluster, detailing their demographics, preferences, and behaviors.

Marketing Strategy: Tailor marketing strategies for each segment based on their unique characteristics. For example, sending personalized recommendations, offering discounts on frequently purchased items, or targeting specific demographics with relevant promotions.

Evaluation: Assess the effectiveness of the segmentation by monitoring key metrics such as conversion rates, customer retention, and revenue generated from each segment.

Iterative Improvement: Continuously refine the segmentation and marketing strategies based on feedback and performance metrics to optimize results over time.

By employing customer segmentation through clustering algorithms, e-commerce businesses can enhance customer engagement, improve conversion rates, and ultimately increase revenue by delivering personalized experiences to different customer segments.

V.CONCLUSION AND FUTURE WORK

These clustering models need to possess the capability to process this enormous data effectively. Each of the above discussed clustering algorithms come with their own set of merits and demerits. The computational speed of K-Means clustering algorithm is relatively better as compared to the hierarchical clustering algorithms as the latter require the calculation of the full proximity matrix after each iteration . K-Means clustering gives better performance for a

large number of observations while hierarchical clustering has the ability to handle fewer data points. The major hindrance produces itself in the form of selecting the be provided as an input to this non-hierarchical clustering algorithm. This limitation does not exist in the case of hierarchical clustering since it does not require any cluster centers as input. It

depends on the user to choose the cluster groups as well as their number. Hierarchical clustering also gives better results as compared to K-Means when a random dataset is used. The output or results obtained when using hierarchical clustering are in the form of dendrograms but the output of K-Means consists of flat-structured clusters which may be difficult to analyze. As the value of k increases, the quality(accuracy) of hierarchical clustering improves when compared to K-Means clustering. As such, partitioning algorithms like K-Means are suitable for large

datasets while hierarchical clustering algorithms are more suitable for small datasets .

Both K-Means and Hierarchical clustering have drawbacks that make them unsuitable when used individually. For business use, data visualization forms a major part of efficient data analysis and hierarchical clustering aids in doing so. Furthermore, when the performance aspect is taken into account, K-Means tends to deliver better results. With the vices and virtues of the two techniques pointed out, it comes to light that an amalgam of the best of these algorithms could outperform the individual models. In summary, different clustering algorithms, owing to their

properties towards different kinds of data can be used in succession such that the advantages of these techniques could be harnessed in full. However, the selection process of these suitable techniques as well as their judicious

implementation could require a considerable time investment in studying and processing the data along with an adequate understanding of the goals and requirement

REFERENCES

[1] Li, Zeying. "Research on customer segmentation in retailing based on clustering model." In 2011 International Conference on Computer Science and Service System (CSSS), pp. 3437-3440. IEEE, 2011

[2] Wang, Zhenyu, Yi Zuo, Tieshan Li, CL Philip Chen, and Katsutoshi Yada. "Analysis of Customer Segmentation Based on Broad Learning System." In 2019 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC), pp. 75-80. IEEE, 2019.

[3] Kansal, Tushar, Suraj Bahuguna, Vishal Singh, and Tanupriya Choudhury. "Customer segmentation using K-means clustering." In 2018 international conference on computational techniques, electronics and mechanical systems (CTEMS), pp. 135-139. IEEE, 2018

[4] Bhade Kalyani, Vedanti Gulalkari, Nidhi Harwani and Sudhir N Dhage "A Systematic approach to customer segmentation and buyer targeting for profit maximization" In 2018 9 th International Conference on Communication and Networking Technologies(ICCCNT), pp. 1-6 IEEE,2018.

[5] https://archive.ics.uci.edu/ml/datasets/Online+Retail

[6] https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

[7] https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm

[8] https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingClassifier.html

INTERNATIONAL JOURNAL OF

MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com