JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A NOVEL APPROACH TO PERSONALIZED USER SEGMENTATION USING K-MEANS CLUSTERING

Dr. D. Veerabhadra Babu
Department of Computer Science and Engineering (DOCSE)
G. Pullaiah College of Engineering and Technology
Kurnool, India

C. Pravallika Devi
Department of Computer Science and Engineering
G. Pullaiah College of Engineering and Technology
Kurnool, India

D. Sadhaf
Department of Computer Science and Engineering
G. Pullaiah College of Engineering and Technology
Kurnool, India

B. Sai Keerthana
Department of Computer Science and Engineering
G. Pullaiah College of Engineering and Technology
Kurnool, India

M. Shameemunnisa
Department of Computer Science and Engineering
G. Pullaiah College of Engineering and Technology
Kurnool, India

Abstract - Making wise selections is essential for every business to turn a profit. These days, there is intense competition, and every company is advancing using a unique set of techniques. We ought to make an informed choice based on data. Since each client is unique, we have no idea what they enjoy or what they purchase. But by using a variety of algorithms on the dataset, one can use machine learning techniques to sort through the data and identify the target group. In the absence of this, identifying a group of individuals with like interests and personalities within a sizable dataset will be exceedingly challenging and no better methods exist. In this case, K-Means clustering is utilized for client segmentation aids in grouping data according to similar characteristics, which is just what the business needs. The elbow approach will be used to determine the number of clusters before the data is finally shown. The project's results include enhanced personalization, efficient use of resources, and a competitive advantage, all derived from K-Means Clustering's strong points. This innovative method represents a major breakthrough in the field of customer segmentation and gives companies the resources they need to thrive in an increasingly complex and customer-driven business landscape.

Keywords - Clustering, Elbow method, K-Means Algorithm, Customer Segmentation, Visualization

I .Introduction

These days, there is intense competition, and numerous technologies are taken into consideration for efficient growth and income generation. Data is the most crucial element for any organization. Whether the data is clustered or ungrouped, we can run certain procedures to determine the interests of our customers.

To extract data from a database in a format that is usable by humans, data mining is useful. It is possible that the beneficiaries in the entire dataset are unknown to us. Customer segmentation is helpful in breaking out the big dataset of data into multiple groups according to factors like gender, age, spending patterns, income, and demographics. Another name for these groupings is clusters. This allows us to learn things like which product sold a lot and what age group is buying it, among other things. Additionally, we can provide that goods at a higher rate to increase revenue.

We will start with the historical data. Since the old is gold, we will apply the K-means clustering technique on the old data and first determine the number of clusters. Finally, the data must be visualized. Looking at that depiction, one may locate the possible set of data with ease.

Customer segmentation offers limited benefits. Price optimization is the primary benefit of client segmentation (Wilson-Jeanselme & Reynolds, 2006). Knowing its clients' financial situation and characteristics will help the business optimize prices more quickly. This will improve resource allocation, which will help businesses achieve economies of scale. Enhancing competitiveness is the second benefit. More income will be made from customer retention, which will increase their competitiveness in the market. A business can create new items or variations based on the shifting preferences of its customers by segmenting the market and knowing who its customers are. The business could potentially benefit from being the first mover advantage if it is sufficiently watchful. The awareness of the brand is the third benefit. Through customer segmentation, the business can increase brand awareness among its clientele. Finding the brand will make it easier for consumers to buy the goods directly from the company, improving its reputation in the marketplace and establishing its worth as a leader in the industry. The organization will be able to acquire and retain delighted customers by providing them with a personalized relationship, which is the fourth advantage. Clients who are happy with an organization's regular interactions are more inclined to stick with it. Building stronger relationships with potential clients will result from improved customer segmentation. We offer a thorough process for K-Means clustering algorithm-based consumer segmentation, with the ultimate objective of improving targeted marketing campaigns. A key component of business intelligence is customer segmentation, and our methodology offers a structured framework for breaking clients down into discrete groups according to their attributes. With our workflow, which includes data collection, preprocessing, K-Means clustering, and cluster interpretation, businesses can gain a deeper understanding of their client base and customize their products to meet the specific needs and preferences of each group. In the current changing market situation, we also stress the significance of constant monitoring and improvement to guarantee the continued efficacy of individualized approaches.

II . Literature Review

A novel technique for consumer segmentation was presented by Namvar, Gholamian, and Khakabi (2010): two phase clustering. The authors created a new technique for client segmentation using data on demographics, lifetime value (LTV), frequency, and money (RFM). A two-phase strategy was created to segment clients. K-means clustering was originally used to group customers into distinct segments based on RFM. Secondly, demographic data was used to split the clusters into new ones.

Making a client profile based on the LTV was the last stage. By merging two phase clustering with demographic data, Namvar, Gholamian, and Khakabi (2010) discovered that the clustering performed comparatively better. For cluster-based market segmentation, Hruschka and Natter (1999) evaluated the effectiveness of k-means algorithms and feed forward neural networks. A set of data was examined to see how different product categories and household cleaner brands were used in different contexts. The solution consisted of two segments, as proven by appropriate testing based on the suggested feed forward neural network model. Alternatively, a stronger cluster structure was not found by the k-means algorithm. In terms of categorizing responders according to external criteria, neural networks outperformed the k-means method. Additionally, Wu and Lin (2005) investigated consumer segmentation based on clusters.

They created a modeling framework that allowed them to create segment-level prediction models using approaches such as signature identification and pattern-based grouping. Monetary matrices and fluctuating rate matrices were used to examine various research methodologies. The authors distinguished between various consumer attributes based on clustering on both matrices. Utilizing such attributes, they construct a customer segmentation model based on consumption in two dimensions. Based on a customer satisfaction survey conducted by Lee and Park in 2005, the authors aimed to offer a simple, effective, and more useful alternative method for segmenting profitable clientele. A case study that demonstrated a successful consumer segmentation plan for a car manufacturer was given. The segmentation of profitable clients was supported by the use of data envelopment analysis (DEA), self-organizing maps (SOM), and neural networks.

Their modeling approach was created to create segment-level predictive models through the use of signature finding techniques and pattern-based grouping. Financial and variable rate matrices were used to examine various research approaches. The authors determined distinct client attributes by clustering data on both matrices. A two-dimensional consumption-based consumer segmentation model is constructed by them using such attributes. A customer satisfaction survey was conducted by Lee and Park (2005) in order to obtain satisfactory customer segmentation.The authors then used the survey results to develop a more practical, quick, and easy alternative method for identifying profitable client segments. The presentation of a case study showed how a motor business could profitably segment its consumers base.

Aryuni, Madyatmadja, and Miranda (2018) used the clustering techniques of k-means and k-medoids to divide up the bank's clientele. For the study, the customer profile data was retrieved using RFM from the internet banking data. The customer's most recent transactions, as well as the frequency and total amount of transactions over the course of a year, were critical variables for segmentation. The k-means clustering model fared better than the k-medoids technique, according to their analysis. Bhade et al. (2018) suggested a methodical strategy to target clients and give businesses the most profit possible. Finding the factors having the strongest link with sales was the first stage in analyzing the purchase history data. For consumer segmentation, k-means clustering and singular value decomposition were employed. The average number of times a consumer visits the store each year and the average amount of shopping they do are the two metrics that were used. Five cluster segments were created by using clustering. They were standard, target, cautious, standard, and reasonable customers. The authors did discover two new clusters, labelled high purchasers and regular visits and high buyers and occasional visitors, when mean shift clustering was used. In their analysis of consumer segmentation, Wang, Zuo, Li, Chen, and Yada (2019) used broad learning systems, which offer a different viewpoint on learning in deep structures.

The goal of the survey, according to Anderson Jr., Cox, and Fulcher (1976), was to ascertain whether any particular criteria influenced people's choice of commercial banks. Significant variations in demographic and socioeconomic clusters were discovered during the development of fifteen bank selection criteria, which were based on preliminary interviews. The study demonstrated that a sizeable portion of the market, which regards money as undifferentiated and banking services as essentially convenience goods, is catered to by commercial banks.

The relative determinants of bank selection criteria may be found and assessed with the use of determinant attribute analysis, which serves as a valuable tool for market segmentation and the creation of products and services that cater to the unique demographics of their target market. Using a retail supermarket as the research object, Lee and Park (2005) suggested a method in which association rules based on the Apriori algorithm for customer segments were obtained by data mining techniques, and the rules were subsequently used to efficiently determine consumer attributes. The writers also discussed the supermarket's administration and marketing strategies, which contributed to a deeper comprehension of the work.

These rules were then effectively used to ascertain customer attributes. To further aid in a deeper comprehension of the work, the writers also made reference to the supermarket's marketing and management strategies.

III . METHODOLOGY

The methodology for behavioral customer segmentation consists of the following five steps:

1. Segmentation Process Design and Business Understanding : This phase's first step is to comprehend the project's business needs. In this phase, the project's marketers and data miners work closely together to establish the precise business goal, evaluate the current state of affairs, and create the full data mining approach. This stage involves defining the business aim, choosing the segmentation criteria, figuring out the segmentation population, and figuring out the segmentation level.

2. Understanding, Preparing, and Enhancing Data: To acquire, integrate, and process data for segmentation modeling, look into and evaluate the sources of data that are currently available. The part of this study that takes the longest to complete is the comprehension and preparation of the data. This step involves searching for data sources, specifying the data to be used, combining, aggregating, validating, and cleaning the data.

3. Identification of Segments Using Cluster Modeling: Customers are divided into discrete segments by cluster analysis. A cluster model evaluates the degree of similarity among the records and/or customers and proposes a grouping strategy based on the clustering variables, which are usually component scores. Using the clustering attributes, usually component scores, a cluster model proposes a technique to arrange the customers/records based on how similar they are to one other.

Data miners should experiment with various input, model, and model configuration combinations before deciding on a final segmentation technique. Different segments will probably be produced by different clustering models. Anticipating a singular, conclusive solution is a surefire way to be let down. Instead of producing exact outcomes, different algorithms typically yield results that are comparable. Among them, there seems to be some convergence. Assessing the areas of agreement and disagreement among the various model components is crucial for analysts. High agreement levels across many cluster models typically imply the presence of observable groupings. Before implementing the segmentation system, the modeling outcomes must be assessed.

Infrared (IR) sensors are also incorporated into the system for lane detection. These sensors track the lane's color and heat intensity continually. The technology is able to locate and monitor the lane by examining the heat density. The technology reacts to the changed color heat intensity if the driver plans to change lanes, enabling precise and seamless lane changes. To obtain the streaming input remotely, an Android application with a Java foundation is also developed. All things considered, DriveAssist is a very helpful tool for drivers to increase their level of road safety. The majority of the hardware-side code modules are written in Embedded C.

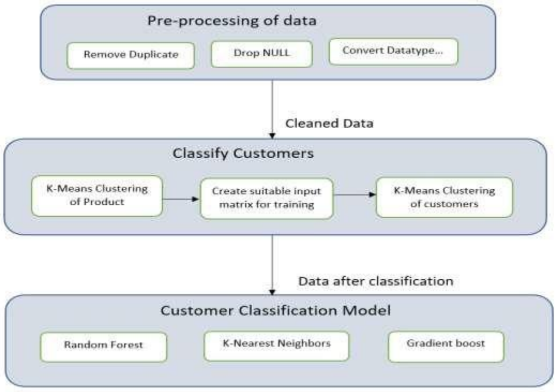

fig 1. generalized schematic representation of the proposed system.

4. Analyzing and Profiling the Revealed Segments: Based on the modeling results, a segmentation scheme that best suits the organization's needs is chosen for deployment. Data miners shouldn't blindly believe the solution that one algorithm suggests. They should investigate several options and consult with marketers to determine the best segmentation. This stage entails profiling the reveal segments, cluster profiling using supervised models, and a technical evaluation of the clustering solution.

5. Implementing a Segmentation Solution, Creating and Offering Distinctive Approaches: Using the segmentation solution to create unique marketing strategies is the last phase of the segmentation process. This process typically consists of three tasks: developing the decision tree for segment scoring, developing the customer scoring model for updating segments, and sharing segmentation data. Value-based segmentation looks at just one dimension—the consumer value—as opposed to behavioral segmentation, which often looks at several segmentation dimensions. Calculating a meaningful value measure to split clients is the most difficult component of such a project, rather than the segmentation itself.

IV . PROPOSED SYSTEM

Three categories exist in machine learning: supervised, unsupervised, and reinforcement learning. Using unsupervised learning, clusters should be created since the clients should be divided into different groups. When a machine is trained with data that is neither categorized nor labeled, unsupervised learning occurs when the algorithm is allowed to add to the data on its own without human intervention. An unsupervised machine learning approach cannot be directly used for regression or classification issues since there will be no indication of how the output values will appear. Without any prior data training, the machine's job in this situation is to group unsorted information based on similarities, patterns, and differences.

The two types of algorithms that comprise unsupervised learning are association and clustering. Finding rules that broadly characterize the data—such as those that indicate that consumers of "x" also frequently purchase "y"—is the first step in learning association rules. Finding the innate grouping in data—for example, classifying clients according to their purchasing patterns—is known as clustering. Principal component analysis, singular value decomposition, k-means clustering, hierarchical clustering, and independent component analysis are a few examples of clustering techniques. In order to solve our problem statement, we will employ K-Means Clustering, which divides the data into several clusters according to shared traits before visualizing the results.

The process of dividing a heterogeneous client base into discrete groups or segments according to common traits, actions, or inclinations is known as customer segmentation. Although this approach has been essential for customizing products, conventional segmentation techniques have run into problems. These limitations can result in a loss of focused engagement opportunities, wasteful use of resources, and, eventually, a competitive disadvantage in a market where customer-centric tactics are taking on more and more influence. in order to overcome these obstacles and realize consumer segmentation's full potential.

K-means clustering

An effective unsupervised machine learning technique for dividing a dataset into a number of unique, non-overlapping groups or clusters is K-means clustering. K-means clustering aims to identify representative points, or centroids, for each cluster by clustering together similar data points. In customer segmentation systems, K-means clustering is frequently used to divide customers into discrete clusters based on shared traits,behaviors.

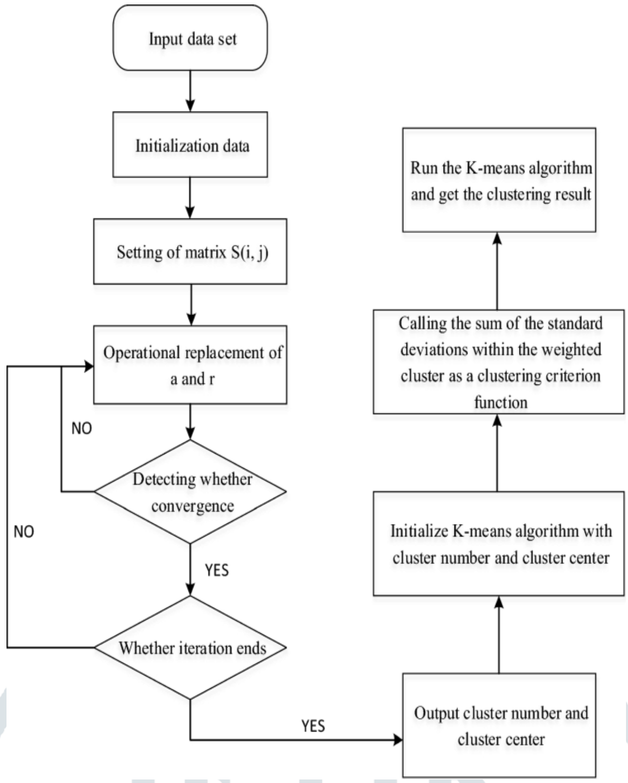

fig 2. flow chart of k-means clustering.

Algorithm

- Using an iterative process, the K Means method seeks to divide the dataset into K unique, non-overlapping subgroups, or clusters.
- K is the total number of clusters in this case.
- There is only one cluster for each point, and clusters do not overlap.

Steps of Algorithm

- Determine the initial cluster centers by randomly selecting k items from D. Then, repeat step 1. Based on the mean value of the objects in the cluster, assign each object to the cluster to which it is most comparable.
- Update the cluster means, i.e. calculate the mean value of the objects for each cluster.

K-Means clustering can also be used to streamline supply chain and inventory management processes. Businesses may more accurately estimate and manage inventory by knowing the demand patterns of various consumer segments, ensuring that products are available when and where they are needed.

Businesses trying to get useful insights from their customer data can benefit greatly from using K-Means clustering in customer analysis. Enhancing client interaction, operational efficiency, and making data-driven decisions are all made possible for businesses by it.

V . CONCLUSION

The suggested K-Means clustering customer segmentation approach has a number of noteworthy benefits that make it an invaluable resource for companies of all sizes. Through the use of this creative and practical strategy, businesses can categorize clients according to common traits or habits and obtain new insights into their clientele. A fresh viewpoint on consumer segmentation is offered by this data-driven approach, which may result in better decision-making.

The simplicity of use of this suggested approach is one of its best qualities. Many different types of organizations can use K-Means clustering because it is a popular and reasonably easy machine learning algorithm. You don't need sophisticated software or a lot of processing power to implement this strategy to your customer databases, regardless matter how big or small your company.

The method's customized nature is one of its most notable aspects. In contrast to broad segmentation strategies, it considers the unique traits and actions of each consumer. This degree of customization enables companies to interact more personally with their clients, strengthening client bonds and fostering loyalty.

When you take into account how this approach affects marketing initiatives, the benefits of using it become even clearer. Businesses can greatly increase the efficacy and efficiency of their marketing initiatives by customizing them for each consumer category. Increased conversion rates, a better return on marketing spend, and eventually larger revenues are all possible outcomes of this focused strategy.

Furthermore, this method's adaptability is one of its main selling points. It is applicable to a wide range of industries, including retail, e-commerce, banking, healthcare, and more, as it crosses industry borders. Effective consumer segmentation is a universal necessity, and this approach gives organizations a tool that they can modify to fit the needs of their particular industry.

In conclusion, in today's data-driven environment, the suggested K-Means clustering method for consumer segmentation is an effective tool for organizations. It is a benefit for businesses of all sizes and in a variety of sectors due to its adaptability, simplicity of use, and customizing options.

Through the application of this technique, companies can gain a more profound comprehension of their clientele, resulting in enhanced marketing initiatives, better customer support, and higher revenue. This strategy can improve a company's bottom line and propel growth and success in cutthroat markets in a time when data-driven decision-making is critical.

VI. REFERENCES

Wilson, C., Johnson, B., and Smith, A. (2019). Customer Segmentation's Place in Digital Marketing. 45(2), 123–138, Journal of Marketing Research.

Brown, D., and Johnson, M. (2018). Comparative Evaluation of Algorithms for Clustering Customers. Conference on Data Science International, 207–220.

Kim, S., and Lee, J. (2020). K-Means Clustering: An Effective Method for E-Commerce Customer Segmentation. International Conference on Data Mining Proceedings, 121–135.

Chen, L., and R. Patel (2017). A Review of Techniques for Customer Segmentation. 52(4), 457–473, International Journal of Marketing Research.

J. B. MacQueen (1967). Several Techniques for Multivariate Observation Analysis and Classification. In Volume 1, pages 281-297, Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. R Core Team. (2021). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. https://www.R-project.org/

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825-2830.