



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 7, July 2022

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.165



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

Customer Segmentation Using Machine Learning

¹Manjushree K R, ²Aishwarya G, ³Kalpak D T, ⁴Ninada Narayan Patgar, ⁵Padmashree T U

¹Assistant Professor, Dept. of ISE, Bapuji Institute of Engineering and Technology, Karnataka, India

²UG Student, Dept. of ISE, Bapuji Institute of Engineering and Technology, Karnataka, India

³UG Student, Dept. of ISE, Bapuji Institute of Engineering and Technology, Karnataka, India

⁴UG Student, Dept. of ISE, Bapuji Institute of Engineering and Technology, Karnataka, India

⁵UG Student, Dept. of ISE, Bapuji Institute of Engineering and Technology, Karnataka, India

ABSTRACT: The emergence of many competitors and entrepreneurs has caused a lot of tension among competing businesses to find new buyers and keep the old ones. As a result of the predecessor, the need for exceptional customer service becomes appropriate regardless of the size of the business. Furthermore, the ability of any business to understand the needs of each of its customers will provide greater customer support in providing targeted customer services and developing customized customer service plans. This understanding is possible through structured customer service. Each segment has customers who share the same market features. Big data ideas and machine learning have promoted greater acceptance of automated customer segmentation approaches in favor of traditional market analytics that often do not work when the customer base is very large. In this paper, the k-means clustering algorithm is used for this purpose.

KEYWORDS: ML, Kmeans Clustering, VS code.

I. INTRODUCTION

Over the years, increased competition among businesses and the availability of largescale historical data has resulted in widespread use of data mining techniques to find critical and strategic information that is hidden in organizations' information. Data mining is the process of extracting logical information from and so on. The importance of customer segmentation includes, inter alia, the ability of a business to customize market plans that would be appropriate for each segment of its customers; Support for business decisions based on risky environments such as credit relationships with its customers; Identify products related to individual components and how to manage demand and supply power; Interdependence and interaction between consumers, between products, or between customers and products are revealed, which the business may not be aware of; The ability to predict customer declines, and which customers are likely to have problems and raise other market research questions and provide clues to find solutions. Buried in a database of integrated data proved to be effective for detecting subtle but subtle patterns or relationships. This mode of learning is classified under supervised learning. Integration algorithms include the K-Means algorithm, K-nearest algorithm, sorting map (SOM), and more. These algorithms, without prior knowledge of the data, are able to identify groups in them by repeatedly comparing input patterns, as long as static aptitude in training examples is achieved based on subject matter or process.

II. EXISTING WORK

Customer Data stored in database is of less use as we cannot sort-it-out or make any meaningful conclusion about customer needs. It does not helps the business to understand the market situation and leads to decrease in revenue generated and consumer dissatisfaction. The existing system is unable to predict the current situation of market which in turn we lose competitive advantage.

III. LITERATURE SURVEY

E-commerce transactions are no longer a new thing. Many people shop with e-commerce and many companies use e-commerce 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. 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 homogeneous groups. The data used in customer segmentation technique that divides the customers into groups depends on various factors like, data geographical conditions, economic conditions, demo-graphical 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. Marketing techniques are entirely based on mutual consumer-retailer relationship. One way to increase profits is to determine customer requirements, through communication with the consumers. Communicating with consumers on personal level is practically not an easy task and without building communication, marketing disasters are inevitable. To tackle this problem retailers can communicate through the data generated by consumers. Retailers can group their consumers according their habits and later on develop business strategies according to it. Customer segmentation is a manner to enhance communication with the consumer, to realize the desires of the consumer hobby in order that suitable communication may be built. Also, strategies of Customer Segmentation may be categorized into Simple technique, RFM technique, Target technique, and Unsupervised technique. In the aggressive marketplace of e-commerce, the hassle of figuring out capacity consumer is gaining increasingly more attention. This paper proposes the solution to identify potential customers using RFM analysis tool and k-means algorithm. One of the important steps for customer segmentation is clustering. A MATLAB implementation of k-Means clustering set of rules for consumer segmentation primarily based totally on records accrued from mega enterprise retail outfit, has a purity of 0.95 indicating 95% accurate segmentation of the customers. Over the years, the commercial world is becoming more competitive, as such organizations have to satisfy the needs and wants of their customers, attract new customers, and hence enhance their businesses. The task of identifying and satisfying the needs and wants of each customer in a business is a very complex task. This is because customers may be different in their needs, wants, demography, geography, tastes and preferences, behaviours and so on. As such, it is a wrong practice to treat all the customers equally in business. This challenge has motivated the adoption of the idea of customer segmentation or market segmentation, in which the customers are subdivided into smaller groups or segments wherein members of each segment show similar market behaviours or characteristics. According to customer segmentation is a strategy of dividing the market into homogenous groups posits that —the purpose of segmentation is the concentration of marketing energy and force on subdivision (or market segment) to gain a competitive advantage within the segment. It's analogous to the military principle of concentration of force to overwhelm energy. Customer or Market segmentation includes geographic segmentation, demographic segmentation, media segmentation, price segmentation, psychographic or lifestyle segmentation, distribution segmentation and time segmentation. The zeitgeist of modern era is innovation, where everyone is embroiled into competition to be better than others. Today's business run on the basis of such innovation having ability to enthrall the customers with the products, but with such a large raft of products leave the customers confounded, what to buy and what to not and also the companies are nonplussed about what section of customers to target to sell their products. This is where machine learning comes into play, various algorithms are applied for unravelling the hidden patterns in the data for better decision making for the future. This elude concept of which segment to target is made unequivocal by applying segmentation. The process of segmenting the customers with similar behaviour into the same segment and with different patterns into different segments is called customer segmentation. In this paper, 3 different clustering algorithms (k-Means, Agglomerative, and Meanshift) are been implemented to segment the customers and finally compare the results of clusters obtained from the algorithms. A python program has been developed and the program is been 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 mean 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 have been formed labelled as Careless, Careful, Standard, Target and Sensible customers. However, two new clusters emerged on applying mean shift clustering labelled as High buyers and frequent visitors and High buyers and occasional visitors.

IV. METHODOLOGY

Data Collection : This is a data arrangement stage. The component generally serves to refine all information things at a standard rate to work on the execution of clustering algorithms and customer classification is done.

Group Analysis : Group analysis is an incorporation or unification, way to deal with purchasers dependent on their likeness. There are 2 primary sorts of unmitigated gathering examination in market strategy: hierarchical group analysis, and characterization. Meanwhile, we will examine how to characterize gatherings, called k-means.

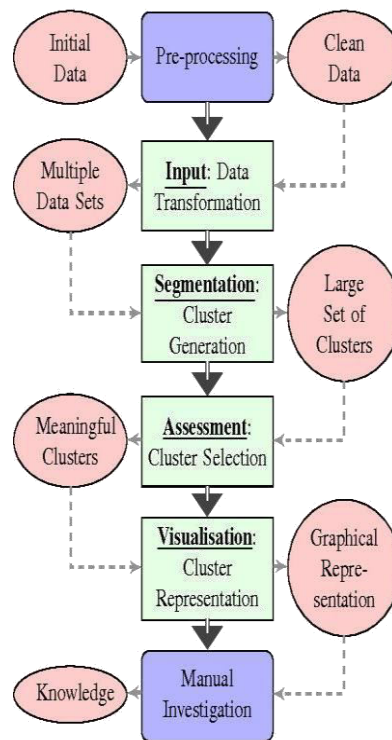


Fig. 3.1 Model Architecture

K-means clustering

K-means clustering algorithm is one of the grouping calculations dependent on division. It's anything but a heuristic iterative cycle to re-partition information articles and re-update group focuses.

The essential thought of the calculation is: assume a set with component objects and the quantity of bunches to be generated.

In the first cycle, an example component is arbitrarily chosen as the underlying bunch centre, and the distance between other example components and the middle point is broke down the groups are individually isolated by the distance.

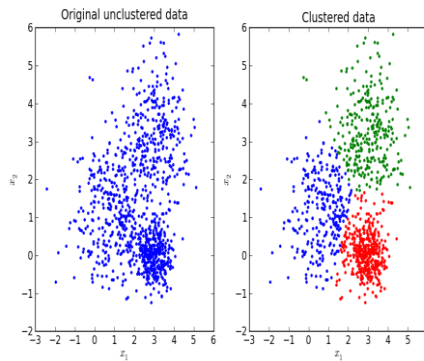


Fig. 3.2 K-means clustering

Data and Sources of Data

1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	
2	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Na	Segment	City	State	Country	Postal Code	Market	Region	Product ID	Category	Sub-Category	Product Name	Sales	Quantity	Discount	Profit	Shipping Cost	Order Priority	
3	32298	CA-2012-124	31-07-2012	31-07-2012	Same Day	RM-15495	Rick Hansen	Consumer	New York Ct	New York	United States	10024	US	East	TEC-AC-1000	Technology	Accessories	Plantronics C	2309.65	7	0	762.1845	933.57	Critical	
4	20341	IN-2013-1781	09-02-2013	09-02-2013	Second Class	RM-16210	Justin Ritter	Corporate	Wellington	New South W	Australia			Oceania	FUR-CH-1000	Furniture	Chairs	Novimec Eve	3709.993	9	0.1	-288.765	923.63	Critical	
5	25330	IN-2013-7121	17-10-2013	18-10-2013	First Class	CR-12730	Craig Reiter	Consumer	Brisbane	Queensland	Australia			Oceania	TEC-PH-1000	Technology	Phones	Nokia Smart	5175.171	9	0.1	919.971	915.49	Medium	
6	13524	ES-2013-1537	28-01-2013	30-01-2013	First Class	MM-16375	Katherine M	Home Office	Berlin	Berlin	Germany			EU	TEC-PH-1000	Technology	Phones	Motorola Sm	2892.51	5	0.1	-96.54	910.16	Medium	
7	47221	SG-2013-4320	05-11-2013	06-11-2013	Same Day	RM-9495	Rick Hansen	Consumer	Dakar	Dakar	Senegal			Africa	TEC-CHA-1000	Technology	Copiers	Sharp Wireless	2832.96	8	0	311.52	903.04	Critical	
8	22732	IN-2013-4238	28-06-2013	03-07-2013	Second Class	MM-15655	Jim Michum	Corporate	Sydney	New South W	Australia			Oceania	TEC-PH-1000	Technology	Phones	Samsung Snc	2862.675	5	0.1	763.275	897.35	Critical	
9	38070	IN-2013-8184	07-11-2011	09-11-2011	First Class	TS-21340	Toby Swindell	Consumer	Poinira	Wellington	New Zealand			Oceania	FUR-CH-1000	Furniture	Chairs	Novimec Eve	1822.08	4	0	564.84	894.77	Critical	
10	31191	IN-2012-8636	14-06-2012	14-06-2012	Standard Clai	MM-18085	Mike Brown	Consumer	Hamilton	Waikato	New Zealand			Oceania	FUR-CH-1000	Furniture	Tables	Chromcraft C	2244.84	6	0	890.48	878.38	High	
11	40553	CA-2014-135	14-10-2014	21-10-2014	Standard Clai	JW-15220	Jaime Wasco	Corporate	Sacramento	California	United States			West	OFF-BH-1000	Office Suppli	Blenders	Fellowes P85	5083.96	5	0.2	1906.485	867.69	Low	
12	40936	CA-2012-116	28-01-2012	31-01-2012	Second Class	HM-15495	Joseph Holt	Consumer	Concord	North Caroli	United States			South	FUR-TA-1000	Furniture	Tables	Chromcraft B	4297.644	13	0.4	-1862.3124	865.74	Critical	
13	30577	CA-2011-202	05-06-2011	06-06-2011	Second Class	GM-16495	Greg Maxwell	Corporate	Alexandria	Virginia	United States			South	OFF-SU-1000	Office Suppli	Supplies	Martin Bull C	4164.05	5	0	81.281	846.54	High	
14	28879	ID-2012-2846	19-04-2012	22-04-2012	First Class	AI-10780	Anthony JACC	Corporate	Kabul	Kabul	Alghanistan			Central Asia	FUR-TA-1000	Furniture	Tables	Bevis Confor	4626.15	5	0	647.55	835.57	High	
15	43794	SA-2011-1838	27-12-2011	29-12-2011	Second Class	MM-17260	Magdalena M	Consumer	Rizau	Rizau	Saudi Arabia			EMEA	TEC-CIS-1000	Technology	Phones	Cisco Smart E	2656.96	4	0	1151.4	832.41	Critical	
16	412	MX-2013-136	13-11-2012	13-11-2012	Same Day	VF-23275	Willy Freyrola	Home Office	Toluca	Parana	Brazil			LATAM	FUR-CH-1000	Furniture	Chairs	Harbour Cove	2221.8	7	0	622.02	810.25	Critical	
17	23704	ES-2013-7395	06-06-2013	08-06-2013	Second Class	RF-19120	Peter Fuller	Consumer	Mudanjiang	Heilongjiang	China			North Asia	OFF-AP-1000	Office Suppli	Appliances	KitchenAid M	3703.52	12	0	1036.08	804.54	Critical	
18	13779	ES-2014-5098	31-07-2014	03-08-2014	Second Class	RP-11185	Ben Petermal	Corporate	Paris	Ile de France	France			EU	Central	OFF-AP-1000	Office Suppli	Appliances	Ireville Berli	1869.588	4	0.1	186.948	801.66	Critical
19	36178	CA-2014-143	03-11-2014	06-11-2014	Second Class	TB-21175	Thomas Bold	Corporate	Henderson	Kentucky	United States			South	TEC-AC-1000	Technology	Accessories	Logitech dHn	2249.91	9	0	517.4793	780.70	Critical	

V. RESULTS & DISCUSSION

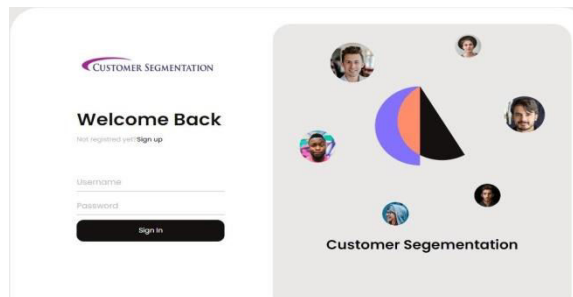


Fig. 4.1 login page

This snapshot shows the welcome page. If a user has not registered earlier has to register before logging in. Once the user is registered, he/she can login to the page with credentials such as username and password.

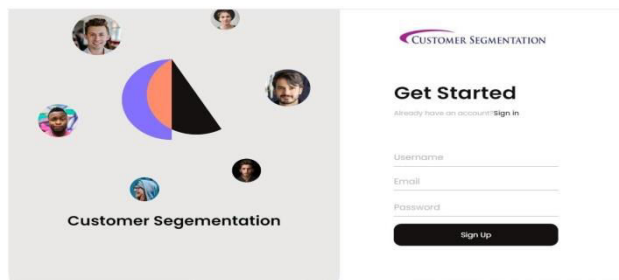


Fig. 4.2 Registration page

This snapshot shows there user registration page for new users. User has to provide details such as username, email id and password for registration.



Fig. 4.3 Home page

This snapshot shows the dashboard of our website. Dashboard consists of options including Home, Analysis and Log Out.



Fig. 4.4 Analysis 1

This snapshot shows customer analysis bar graph where segmentation of category based on county and the profit is selected. Here, categories and sub-categories are displayed. User can view the respective analysis graph by selecting categories.

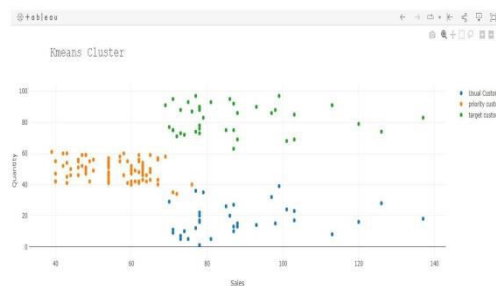


Fig. 4.5 Analysis 2

This figure shows K-means cluster graph. Here, usual customer, priority customer and target customers are shown on graph with respect to sales and quantity of product.

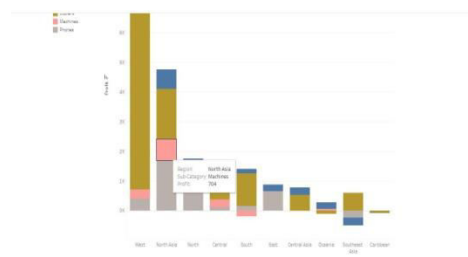


Fig. 4.6 Analysis 3

This snapshot shows bar graph representing the profit in various regions.

Fig. 4.7 Analysis 4

This snapshot shows profit analysis graph. Graph is generated with respect to month of order date, profit and shipping cost.

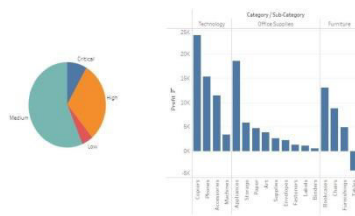


Fig. 4.8 Product categories and shipment

This snapshot shows pie chart to present product categories and their shipment priorities and bar graph to show profit in multiple categories like technology, office supplies and furniture.

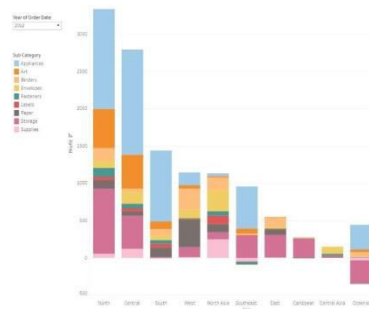


Fig. 4.9 Bar graph of profit of all categories

This snapshot shows bar graph representing the profit of all categories in various regions and the year of order date is 2012.

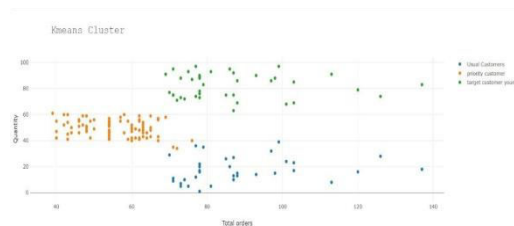


Fig. 5.0 K-means cluster on quality of products



This snapshot shows K-means cluster graph for quality of products with respect to total orders.

VI. CONCLUSION

We used the K-means algorithm to segment our customer in various clusters having similar similarity. Client segmentation in shopping malls is achievable despite the fact that this form of machine learning application is highly useful in the market, a manager can concentrate all of his or her attention on each cluster that has been discovered and meet all of their requirements. Mall managers must be able to understand what customers require and, more importantly, how to meet those needs. analyze their purchasing habits, and establish frequent encounters with customers that make them feel comfortable in order to satisfy their demands.

REFERENCES

- [1] Puwanenthiren Premkanth, - Market Classification and Its Impact on Customer Satisfaction and Special Reference to the Commercial Bank of Ceylon PLC. | Global Journal of Management and Business Publisher Research: Global Magazenals Inc. (USA). 2012. Print ISSN: 0975-5853. Volume 12 Issue 1.
- [2] T . Nelson Gnanaraj , Dr.K.Ramesh Kumar N.Monica. AnuManufactured cluster analysis using a new algorithm from structured and unstructured data. International Journal of Advances in Computer Science and Technology. 2007. Volume 3, No.2.
- [3] McKinsey Global Institute. Big data. The next frontier for innovation, competition, and productivity. 2011. Retrieved from www.mckinsey.com/mgi on 14 July, 2015.
- [4] Jean Yan. —Big Data, Bigger Opportunities- Data.gov’s roles: Promote, lead, contribute, and collaborate in the era of big datal. 2013



INNO  SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

 **doi**[®]
cross **ref**

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  ijircce@gmail.com



www.ijircce.com

Scan to save the contact details