International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT217144

Customer Segmentation Analysis and Visualization

K. Keerthi*[1], G. Lakshmi Thirupathamma [2], N. Vijayalakshmi[3], D. Aparna[4], U. Vineela[5]

*[1]Assistant Professor, Computer Science and Engineering, Qis College of Engineering and Technology, Ongole, Andhra Pradesh, India

[2]Student, Computer Science and Engineering, Qis College of Engineering and Technology, Ongole, Andhra Pradesh, India

Article Info
Volume 7, Issue 1
Page Number: 280-284

Publication Issue :
January-February-2021

Article History
Accepted : 20 Feb 2021
Published : 28 Feb 2021

ABSTRACT

Today's world is all concerning Innovation and new concepts, where everybody desires to contend to measure higher than others. In the business world, it is crucial to know the client's desires and behavior patterns concerning buying merchandise. With the giant number of merchandise the businesses square measure confused to work out the potential customers to sell their merchandise to earn the large profits. To solve this real-time downside we tend to use machine learning techniques and algorithms. We can conclude the hidden patterns of knowledge. So that we can observe choices for earning a lot of profits. For this, we tend to take client information and divides the purchasers into totally different teams conjointly known as segmentation. segmentation permits businesses to create higher use of their selling budgets, gain a competitive edge over rival corporations, and, significantly, demonstrate much better information about your customer's desires and needs. In this project, we tend to square measure implementing k-means agglomeration algorithmic rule to analyze the results of clusters obtained from the algorithmic rule. A code is developed in python and it's trained on an information set having 201 data samples that are taken from the native shopping center. All the offered data within the dataset is placed along to own a concept concerning client age, gender, annual financial gain, and outlay score(Expenditure) of mall customers dataset. Finally, this understanding information is analyzed to the simplest of our knowledge under the abled guidance of our mentor.

Keywords : Pandas, Sklearn, Numpy, Scipy, Matplotlib, Seaborn, Plotly, Mall Customers Dataset.

I. INTRODUCTION

Segmentation of shoppers can even refer as market segmentation or market basket analysis. Customer segmentation could be a procedure of dividing the shoppers into totally different classes supported by their similarities, behavior, and interests. so that company understands their customers and market their merchandise in step with these categories and build their business effective to earn profits. It is all

regarding matching customers with appropriate merchandise and it's a brand new means of an act with customers regarding what you recognize their preferences.Market/customer segmentation is all regarding finding your potential or target customers. Nowadays business individuals suppose one-size-fits-all isn't a good technique to enhance their business profits. If the corporate didn't determine the potential client, it should lead the corporate to fall under losses. We will look into an example of segmentation which is given below.

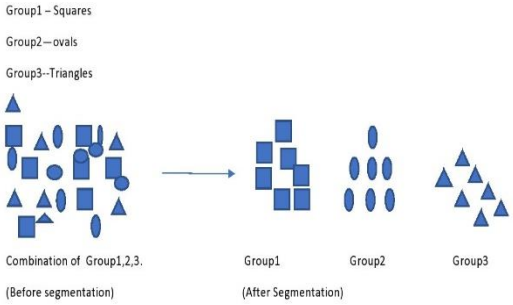

Figure1:Represents the segmentation process.

Customer Segmentation depends on some factors or variables. Those square measure divided into four sorts as given below:

1)Geographic→like location, population, density, climate.

2)Behavioural→like previous purchases, payment ways used.

3)Demographics→like age, gender, marital standing, annual financial gain, educational level, occupation.

4)Lifestyle→like social values, community affiliations.

We have several blessings of client segmentation like ready to determine the foremost valuable and potential customers, we can higher the client service, we can improve revenue, and scale back the wastage of cash.

In this project, we tend to use python language with jupyter notebook from anaconda software. In machine learning, we've got unsupervised learning. We implement the client segmentation exploitation k-means cluster algorithmic program.

II. PROPOSED WORK

In the planned model, the client dataset is organized well to hold out Mathematical and applied math operations(Statistical operations), and therefore the excel file is modified into a comma-separated value file i.e CSV file. Here we tend to use value-based segmentation whenever customers are divided into teams by economic standings like their salary, spending score i.e expenditure, and the alternative factors. Data is planned to the 3D visualized picture, this visualization helps in understanding the results higher. K-means cluster algorithmic rule is employed and plot could be a methodology, accustomed to visualizing the resulted information, sci-kit, pandas and mat-plot lib are some libraries used for on top of mentioned operations.

III. IMPLEMENTATION

To implement the client segmentation first, we must always import the required libraries and modules into the jupyter notebook. Read the information set as 'df' and that we perform some data visualization operations like shape() for knowing the number of rows and columns, info() for obtaining the main points of dataset attributes, and displaying the primary five rows with head().

A count plot() technique is employed on gender attributes that show the counts of observations in every categorical bin using bars. A pair plot permits us to see both distributions of single variables and relationships between two variables. Like that we tend to perform several operations like plots, graphs, etc.Later performing various data visualization operations, Now we focus on the k-means clustering algorithm.

A. K-MEANS CLUSTERING ALGORITHM

K-means clustering algorithmic program computes the center of mass and iterates till we tend to find

optimum centroid. It assumes that the quantity of clusters area unit is already illustrious. It's additionally known as flat clustering algorithmic program. the quantity of clusters known from information by the algorithmic program is portrayed by 'K' in K-means.

In this algorithmic program, the info points area unit is assigned to a cluster in such a fashion that the total of the square distance between the info points and center of mass would be minimum. it's to be understood that less variation at intervals the clusters can cause additional similar information points at intervals in the same cluster.

B. PROCEDURE OF K-MEANS CLUSTERING ALGORITHM

To understand the concept of the k-means clustering algorithm, we follow the following step:

1)first, we'd like to specify the number of clusters, K, that have to be compelled to be generated by this formula.

2) Next, haphazardly choose K purpose | datum \ information}s and assign every knowledge point to a cluster. In straightforward words, classify the info supported the number of knowledge points.

3) Currently it'll figure out the cluster centroids.

4) Next, keep iterating the subsequent till we discover the best center of mass that is the assignment of knowledge points to the clusters that don't seem to be dynamical to any further extent.

(A) Initially, add the square distance between information points and centroids would be computed.

(B) Currently, we've got to assign every information to the cluster that's nearer than a different cluster (centroid).

(C) Finally, eventually cipher the centroids for the clusters by taking the typical of all information points of that cluster.

K-means follows the Expectation-Maximization approach to resolve the matter. The Expectation-step has employed for the distribution of the info points to the nearest cluster and therefore the Maximization-step is employed for computing the center of mass of every cluster.

While operating with K-means algorithmic program we'd like to require care of the subsequent the subsequent.

While operating with clustering algorithms as well as K-Means, it's counseled to standardize knowledge|the info|the information} as a result of such algorithms use distance-based activity to work out the similarity between data points.

Due to the reiterative nature of K-Means and the random format of centroids, K-Means could stick in an exceedingly native optimum and should not converge to an international optimum. that's why it's counseled to use completely different initializations of centroids.

IV. FLOW CHART REPRESENTATION OF CUSTOMER SEGMENTATION:

To understand the customer segmentation we also look into the flow chart of customer segmentation which is given below:

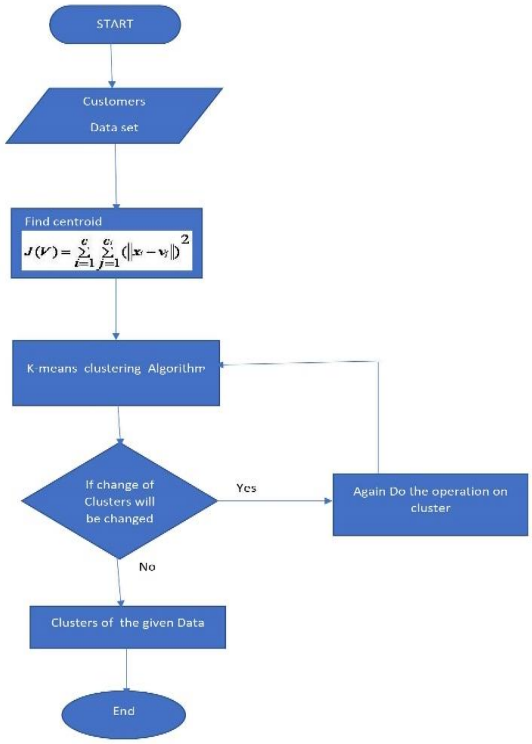

Figure 2. Represents flow chart of k-means algorithm.

V. CONCLUSION

This paper has presented the implementation of K-means clustering algorithmic rule for client segmentation supported information collected from the native shopping center.

Finally, we can conclude, there are 6 types of clusters and customers.

VI. RESULT

A 3D image graph that shows half-dozen kinds of clusters supported attributes like annual financial gain, spending score, and age.

Here :

1) Customers with average regular payment and average outlay.

2) Customers with High financial gain however low pay.

3) Customers with Low financial gain and low pay.

4) Customers with medium financial gain and medium pay of regular payment.

5) customers with medium financial gain regular payment and high annual pay of regular payment.

6)customers with low annual financial gain however it's high expenditure

VII. REFERENCES

[1]. T. Nelson Gnanaraj, Dr.K.Ramesh Kumar N.Monica. Survey on mining clusters using the new k-mean algorithm from structured and unstructured data.

[2]. McKinsey Global Institute. Big data. The next frontier for innovation, competition, and productivity. 2011. Retrieved from www.mckinsey.com/mgi on 14 July, 2015.

[3]. D. D. S. Garla and G. Chakraborty, “Comparison of Probabilistic-D and k-Means Clustering in Segment Profiles for B2B Markets,” SAS Global Forum 2011, Management, SAS Institute Inc., USA

[4]. A. Vattani, “K-means exponential iterations even in the plane,” Discrete and Computational Geometry, vol. 45, no. 4, pp. 596-616, 2011.

[5]. I.S.Dhillon and D. M. Modha, “Concept decompositions for large sparse text data using clustering,” Machine Learning, vol. 42, issue 1, pp. 143-175, 2001.

[6]. Gaurav Gupta and Himanshu Aggarwal International Journal of Machine Learning and Computing, Vol. 2, No. 6, December 2012 10.7763/IJMLC.2012.V2.256 Improving Customer Relationship Management Using Data Mining (20)

[7]. VandanaKorde and C NamrataMahender," Text classification and classifiers:A survey", International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.2, March 2012.

[8]. T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, “An efficient K-means clustering algorithm,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, pp. 881-892, 2002.

[9]. Meeravali Shaik, K. Vivek, dr.Yamarthi Narasimha. (2020). PREDICTION OF TOP-K DOMINANCE QUERIES ON LARGE DATASETS USING SUPPORT VECTOR MACHINE. International Journal of Advanced Science and Technology, 29(05), 10297 - 10302.

[10]. C.L. Blake, D.J. Newman, S. Hettich and C.J. Merz. (2012) UCI machine learning repository databases.
Online].Available:http://mlr.cs.umass.edu/ml/ machine-learningdatabases/0022

[11]. M. A. Nishara Banu, B. Gomathy ,"Disease Forecasting System Using Data Mining Methods", 2014 International Conference on Intelligent Computing Applications

[12]. Fayyad, Ussama; Piattetsky-Shapiro,Gregory; Smyth, Padhraic (1996). "From Data Mining to Knowledge Discovery in Databases" (PDF). Retrieve December 2008

[13]. Tanupriya Choudhury, Vivek Kumar and Darshika Nigam, "Intelligent Classification & Clustering Of Lung & Oral Cancer through Decision Tree & Genetic Algorithm", International Journal of Advanced Research in Computer Science and Software Engineering, 2015.

[14]. Jayant Tikmani, Sudhanshu Tiwari and Sujata Khedkar, "Telecom customer segmentation based on cluster analysis An Approach to Customer Classification using k-means", IJIRCCE, 2015.

[15]. W.Gao,C.Z.He,X.Y.Jiang, Customer segmentation study based on fuzzy clustering ensemble, Journal of Intelligence, 30 (2011), 125- 129.

[16]. H.V. Shashidhar, V. Subramanian, Customer segmentation of bank based on data mining security value based heuristic approach as a replacement to k-means algorithm, International Journal Computer Application, 19 (2011), 13-18.

[17]. Jiawei Han, Michelline Kamber, Data mining: conception and technology, Beijing: Mechanic Industry Publish, 2002

[18]. Yun Chen, Guozheng Zhang, Dengfeng Hu, Shanshan Wang School of Public Economy Administration of Shanghai University of finance & economics, Shanghai, 20043, CUSTOMER SEGMENTATION IN CUSTOMER RELATIONSHIP MANAGEMENT BASED ON DATA MINING

[19]. Sulekha Goyat. The basis of market segmentation: a critical review of literature. European Journal of Business and Management www.iiste.org. 2011. ISSN 2222-1905 (Paper) ISSN 2222-2839 (Online).Vol 3, No.9, 2011

Cite this article as :

K. Keerthi, G. Lakshmi Thirupathamma, N. Vijayalakshmi, D. Aparna, U. Vineela, "Customer Segmentation Analysis and Visualization", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 7 Issue 1, pp. 280-284, January-February 2021. Available at doi : https://doi.org/10.32628/CSEIT217144
Journal URL : https://ijsrcseit.com/CSEIT217144