

# Analysing Customer Buying Habits with Visual Data Mining

Anusha Apparaju

Computer Science and Engineering, JNTUH College of Engineering Jagtial, Jagtial, Telangana, India

## ABSTRACT

### Article Info

Volume 7, Issue 4

Page Number: 584-587

### Publication Issue :

July-August-2021

### Article History

Accepted : 12 Aug 2021

Published : 23 Aug 2021

The ultimate aim of every industry and organization is to make profits by attracting a greater number of customers. To to achieve this motto they need to analyze the priorities of the customer. This is usually done by doing marketing everywhere such as in social media, newspaper, sites, etc. The marketers keep advertising at many sites without even knowing whether the advertisement is useful at that platform or not. Hence, making such huge investments in advertisements at wrong platforms will lead to less profits. There is a need for the companies to provide customer services in such a way that the customer doesn't lose his interest and trust and must maintain a healthy relationship. If such services are provided equally to every customer, then there is a possibility that the company might provide its service to those customers who bring low profits and keep the customers who make high profits in waiting. To avoid such discrepancies, categorization of customers can be done based on their priority. This can be achieved using the Clustering technique, k-means algorithm. Since the customer data is unsupervised, k-means helps us to cluster them. If we use supervised data, then prediction of new customer's priority can also be done using K nearest neighbors' algorithm. The exploration of deep insights of data using exploratory data analysis makes it easy to understand data using visual representations [3][4][6]. These visual representations also lead to less time consumption for exploratory data analysis.

**Keywords :** K-Means, Customers Priority Analysis, Clustering on Customer Data, K-Means For Clustering, Visual Datamining

## I. INTRODUCTION

Now-a-days all the companies collect huge data from the customers, thus data mining comes into existence. This collected data may have many attributes like name, customer id, time of transaction, date of transaction, product details, quantity, price, mobile

numbers, gender, address, date of birth, etc. This enormous amount of data can be subjected to Exploratory Data Analysis (EDA) to extract a lot of vital information. This data includes relevant attributes such as customer age, purchases made recently, frequency of purchases, category and count of products and services availed by the customer.

Such kind of attributes provide a greater scope for clustering, whereas the attributes like address, allow us to identify the locations where purchase rate is more.

Most of the companies claim to identify qualitative and quantitative customers along with their vision of improvising and advancing their products. This objective of identifying customers is the main motive of making this project, which helps the companies in analyzing the customers who provide large profits to the company. In this paper, we consider attributes which provide the information about the amount of transaction, number of transactions made by a customer, the kind of products purchased and then a behavioral attribute age.

Depending upon these main attributes we would separate the customers into different categories and based upon the profits made, the most profitable category is identified. Thus, the company tends to decide the way they need to focus on marketing in order to avoid financial loss on blind advertisements.

It can be a great help when customer services are provided based upon the category the customer belong. Using this we can also tell which product is at the rise and based on the data we get it is likely to identify which product will give high-rate next month. Depending upon the analysis the product manufacturing can be done. Therefore, the data which have the information about customer transactions, the products that are sold by the customer, customer data which includes their age is collected. There are different ways by which a purchase is made by the customer.

## II. LITERATURE SURVEY

### [1] Customer Classification Based on Discrete Data Mining of Customer

Deciding valuable customers based on the customer details which includes a lot of data extraction from big data is involved in the classification of random data concentrated on customer details using data mining. This study is based on the outlier data analysis, customer classification model where the customer details are concerned utmost. This model comprises of four parameters in each dimension, including frequent transactions, variety of services and products, amount of the transaction and age of the client. Clustering is used before classification to divide customer data into four categories of the types they were divided as twenty-five types and certain marketing strategies are also used to classify the outliers of customer data of companies. This can avoid decentralizing the energies to identify the profits of the company and helps in reducing result-less effort. The outliers formed on the basis of clustering algorithm are the least valued customers, but they cannot be segregated out from the valuable customers because they may be the one who has been dealing with huge amount of product purchase. as a template and simply type your text into it.

### [2] Optimization of Decision Support System Based on Data Mining

Clustering algorithm is explored more by using the data more efficiently. Data processing of data mining is studied. In this it is known how effective the data can be utilized and the ways of data processing are examined and explained briefly. To solve the problems of a business decision support system, mathematical statistics such as pattern recognition is done. But in this research, they provided effective decision support for website construction in contemporary e-commerce field. In this paper Ant colony clustering algorithm is effectively used based on information entropy for carrying out path analysis by which five different paths are identified. Ant colony optimization algorithm is known to find good paths from graphical plots using probabilistic techniques for solving computational problems and

provide better paths. This helps in the construction of other e-commerce websites.

### [3] Application of k-means Algorithm for Efficient Customer

Every business tends to keep regular customers and have a healthy relationship between them by gaining new customers they need to make their strategies and marketing methods efficient to attract them and gain their trust. As the business develops the size of customer servicing is also getting larger. In-order to reduce this size and maintain good customers, services should be provided according to their priority and their needs. This helps in providing fruitful services and developing customized strategies in marketing. This can be achieved by systematic customer segmentation based on their purchase and similar patterns which can be done using big data and machine learning. These technologies can be utilized when data available are extremely large. In this paper, the k-means clustering algorithm is applied for this purpose. A MATLAB program of the k-means algorithm was developed and is trained using a z-score normalized two-feature dataset of 100 training patterns acquired from a retail business. The features are the average amount of goods purchased by customer per month and the average number of customer visits per month. From the dataset, four customer clusters or segments were identified with 95% accuracy, and they were labelled: High.

### III. METHODOLOGY

For this paper we have obtained a dataset which consists of more than eighteen thousand transactions and customer data of about five thousand and a limited data about product details. Thus, it involves [1][2]data mining, customer segmentation and predictive analytics. Initially data insights are known using[5]CRISPDM methodology which is followed by many industries in real time. This helps in cleaning the large data and obtaining the key attributes, after

selecting the attributes data is made free from null values and missing rows are eliminated and a finite dataset

is obtained and understood by an analyst. Now the data obtained is divided into categories using k-means [4] clustering algorithm in order to obtain the value of the customers in accordance with their purchases and transactions. Therefore, our unsupervised data is converted into supervised data. This supervised data is now trained by k nearest neighbour[5] classification algorithm to identify the group which a new customer belongs to. Thus, this paper not only helps in customer segmentation and prediction but also in analysing which product is more purchased as per the time period. This analysis can be done using visual representations [10][11][12].

### IV. EVALUATION OF THE CLASSIFICATION EXPERIMENT

The clusters in k-means is identified using elbow method as shown in the fig1, it shows that four clusters can be formed based on the provided data. Then the data is clustered into these clusters and now the converted supervised data is trained using k nearest neighbors algorithm to find the group which a new customer belong to. Later on the accuracy of the model prediction is found, which is about 96%. Giving the data details as follows:

Total accuracy of the model: 0.96032

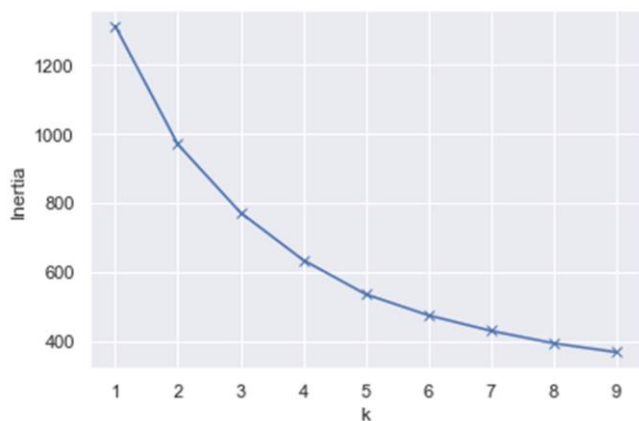
age accuracy: 0.95254

frequency accuracy: 0.96359

amount accuracy: 0.94872

accuracy: 0.97018

Geometric mean accuracy: 0.91914



**Figure 1.** Elbow Method

## V. CONCLUSION

Classification of valuable customers can be done based on their priority which is identified by analyzing their transactions and their relationship with the company. Minimization of cost on advertisements can be made by locating the places where most of the valuable customers are present and spreading their company's awareness at such places can bring most of the customers to them. This also helps in avoiding blind marketing which is done by most of the startup companies to make people know about them which may not work if they do such blind advertisements. This project helps in providing and understanding visual representations in an easy and efficient way. Multiple switching of patterns can be used for fast and clear recognition of prioritized customers based upon the requirement. Depending upon the analysis the product manufacturing can be done. It can be a great help when customer services are provided by attracting them and keeping their trust. Helps in identifying which product will give high rate next month as the accuracy obtained is 96% using KNN algorithm.

## VI. REFERENCES

[1]. P. Adrian and D.Zaninge, "Data Mining," Addison Wesley, 1996.

- [2]. U.M.Fayyad, G.Piatetsky-shapiro, P.Smyth, Uthurusamy, "Advances in Knowledge Discovery and Data Mining", AAAI/MIT Press, 1996, pp1-36.
- [3]. Abello.J, Korn.J, "A System for Visualizing massive multidigraphs", Transaction on Visualization and Computer Graphics, 2001.
- [4]. Research on k-means Clustering Algorithm: An Improved k-means Clustering Algorithm, Shi Na; Liu Xumin; Guan Yong.
- [5]. KNN Model-Based Approach in Classification, Gongde Guo, Hui Wang, David Bell, Yixin Bi, Kieran Greer
- [6]. Crisp-dm: towards a standard process model for data mining, R. Wirth, J. Hipp Published 2000, Computer Science.
- [7]. Kreusler.M, Lopez.N, Schumann.H, "A Scalable frame work for information Visualization", Proc. Inter Vis'2000, Salt Lake city, 2000, pp-27.
- [8]. Augusto, J.C. 2005, Temporal reasoning for decision support in medicine, Artificial Intelligence in Medicine, vol. 33(1), pp. 1-24.

### Cite this article as :

Anusha Apparaju, "Analysing Customer Buying Habits with Visual Data Mining", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 7 Issue 4, pp. 584-587, July-August 2021. Available at doi : <https://doi.org/10.32628/CSEIT2174132>  
Journal URL : <https://ijsrcseit.com/CSEIT2174132>