

Find Out The Top Purchase Product Using k-means Algorithm

Dr.M.Rajeswari

Assistant Professor, Department of B. Com (Business Analytics)
PSGR Krishnammal College for Women, Coimbatore, India.

rajeshwarim@psgrkcw.ac.in

D.Priyanka

UG Scholar, Department of B. Com (Business Analytics)
PSGR Krishnammal College for Women, Coimbatore, India.

Priyankapradeep427@gmail.coms

Abstract:

The aim of this paper is to discover the pinnacle bought product for consumer recommendation using K-Means set of rules. The statistics provided is based on systematic have a look at and medical applications in analysing sales records and purchasing behaviour of the purchasers. The curated and organized facts as an outcome of this scientific look at not best complements commercial enterprise sales and profit, but also equips with clever insights in predicting patron purchasing behaviour and related styles. In order to execute and observe the medical method using K-Means set of rules, the real time transactional and retail dataset are analysed. This objective is based totally at the cluster and deploys dataset segmentation ideas using K-Means Algorithm. A shape of dataset clusters is installation for the calculation of cluster.

Keywords: Identifying top purchase products, Buying pattern of customer, k-means algorithm.

Introduction

The reason of this paper is to find out the top purchased product for purchaser recommendation the use of K-Means set of rules. Customers are divided into set of individuals with wonderful similarities. Some of the attributes relevant to customer segmentation are gender, age, way of life, region, buy and earnings behaviour. To acquire the above targets, customer clustering and segmentation is finished the usage of the K-Means set of rules. Clustering can be described as segmentation of consumer evaluation. The most distinguished manner of segmentations used in K-Means Algorithm are the cluster insights. In the scope of the modern take a look at, cluster perception is used for the cause of segmentation for well-defined client segmentation facilitates in powerful allocation of advertising resources, enables the groups to goal the specific institution of clients and additionally allows in constructing healthful lengthy-term courting with the clients. The effects accordingly received in regards to profits transactions are in assessment with various parameters like unit price and product.

OBJECTIVE

- To analyse the unit price and product of the top purchased product.
- Examine the product recommendation to the customers to make decision.

Related work

P. Anitha Malini, M. Patil 2019, customer segmentation and for data mining can be applied the Retail Industry (Han et al., 2011), because it requires a vast amount of data on sales, transportation, consumption ratio, redelivery service and many others. Also, Retail data mining helps in identifying and effectively mapping customer behavior and related patterns during the entire life-cycle of business transactions.

Deepak Arunachalam, Niraj Kumar 2018, Data collected from sample points are analysed in this paper using four clustering approaches: Hierarchical clustering, K-Medoids, fuzzy clustering, and Self-Organising

Maps (SOM). The findings suggest that Fuzzy and SOM based clustering techniques are comparatively more efficient than traditional approaches in revealing the hidden structure in the data set.

Femina Bahari, Sudheep Elayidom, 2014, Data mining¹⁵ is defined as a process that uses mathematical, statistical, artificial intelligence and machine-learning techniques to extract and identify useful information and subsequently gain knowledge from databases. Information technology tools, advanced internet technologies and explosion in customer data has improved the opportunities for marketing and has changed the way relationships between organizations and their customers.

A. Joy Christy, A. Umamakeswari, L. Priyatharsini, A. Neyaa, 2018, efficient segmentation of customers of an enterprise is categorized into groups of similar behavior based on the RFM (Recency, Frequency and Monetary) values of the customers. The transactional data of a company over is analyzed over a specific period. Segmentation gives a good understanding of the need of the customers and helps in identifying the potential customers of the company.

Xin Fu, Xi Chen, Yu-Tong Shi, Indranil Bose, Shun Caia, 2017, an innovative model for segmenting online players based on data related to their in-game behaviours to support player retention management. This kind of analysis is helpful to explore the potential reasons behind why players leave the game, analyse retention trends, design customised strategies for different player segments, and then boost the overall retention rate.

PeterKolarovszki, 12 September 2016, In the past customers were considered en masse. In the present an individual approach to customer care is the current trend. This individual approach can be ensured by customer segmentation. The introduction of CRM will enable equal and individual approach to customers, however its implementation is time consuming and costly.

Paul Murray, July 2017 separating customers into segments uses descriptive variables to identify similar behavior expectations. In some domains, however, descriptive variables are not available or are not adequate for distinguishing differences and similarities between customers.

WafaQadadeh 2018, Segmentation is an important concept for designing marketing campaigns to improve businesses and increase revenue. Clustering algorithms can help marketing experts to achieve this goal

MEHODOLOGY

The k-means model is used to find the top purchase product using unit price. It is a method of clustering data to partition existing data into one or more clusters or groups. So that the data had the same characteristics was grouped into the same cluster and data that had different characteristics grouped into other groups.

A. PROPOSED SYSTEM

STEP 1: Imported the dataset, modified the dataset and saved in Excel.csv format.

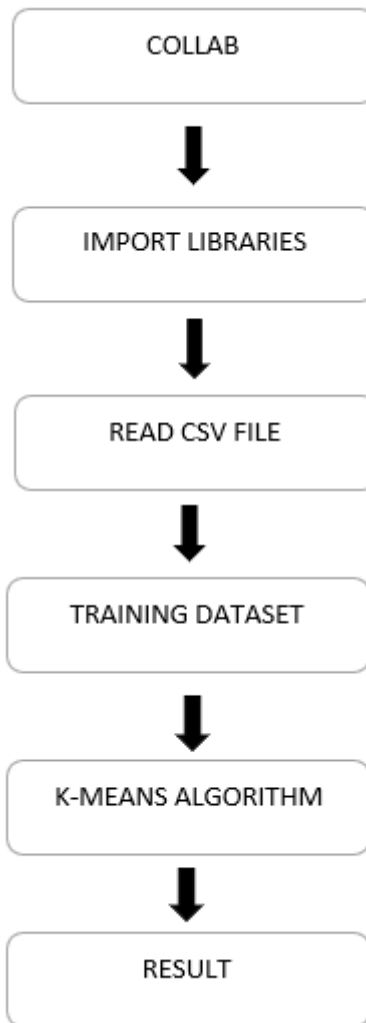
STEP 2: Used Google Colab for executing python coding and removed all unwanted data from dataset.

STEP 3: Then dataset is splitted into training dataset and testing dataset.

STEP 4: Visualization are made in Google Colab for better understanding output are use in k-mean algorithm.

STEP 5: It is used to identifying the most purchased product by analysing buying pattern of the customer.

Work Flow Process



K-means cluster algorithm

- K-Means Clustering is an Unsupervised Learning set of rules, which corporations the unlabelled dataset into one-of-a-kind clusters. Here K defines the variety of pre-described clusters that want to be created inside the process
- It allows facts into different businesses and a handy way to discover the types of businesses within the unlabelled dataset on its own without the need for any training
- It is a centroid-based algorithm, in which every cluster is associated with a centroid. The important goal of this set of rules is to minimize the sum of distances among the records factor and their corresponding clusters.
- The algorithm takes the unlabelled dataset as enter, divides the dataset into k-range of clusters, and repeats the method till it does no longer locate the fine clusters. The value of okay need to be predetermined in this set of rules.

RESULT

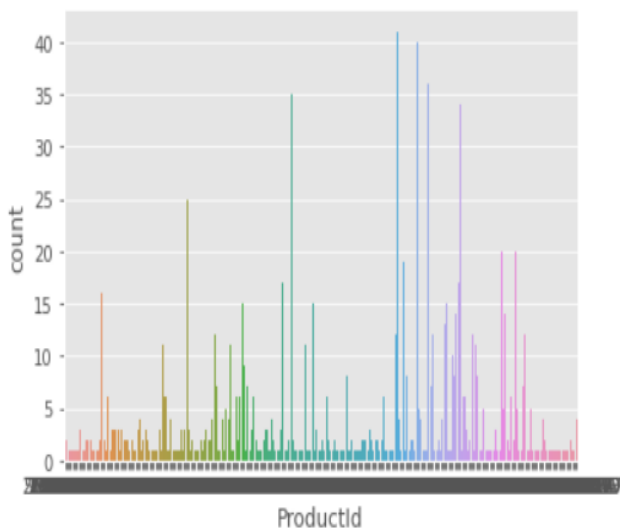


fig 5.1

UnitPrice	ProductId
1.25	106
1.65	102
2.95	93
2.10	64
0.85	60
3.75	54
2.55	53
4.95	44
1.95	39
4.25	35

fig 5.2

- In this paper, in fig 5.1 price of the unit price, product id is compared and this shows the different price ranges of different product, from this it is understood that medium priced products are purchased more when it is compared with luxury products.
- In fig 5.2 it shows the top purchased product by analysing the buying pattern of the customer.

CONCLUSION AND FUTURE WORK

- By importing and modifying the dataset in excel the Colab helps to executing the python coding and it removes the unwanted data in dataset. After that, dataset is splitted into training dataset and testing dataset.
- With the help of Google Colab, the visualization is made to understand the dataset clearly by using k-means algorithm
- By comparing the unit price and product id we got top purchase product which is highly preferred by the customers.
- It shows the top purchased product by analysing the buying pattern.

REFERENCES

- Anitha Malini.P, M. Patil 2019, “RFM model for customer purchase behavior using K-Means algorithm”
- Deepak Arunachalam, Niraj Kumar 2018, “Benefit-based consumer segmentation and performance evaluation of clustering approaches: Evidence of data-driven decision-making”.
- Femina Bahari, Sudheep Elayidom, 2014 “An Efficient CRM-Data Mining Framework for the Prediction of Customer Behavior”
- Joy Christy.A, A. Umamakeswari, L. Priyatharsini, A. Neyaa, 2018, “RFM ranking – An effective approach to customer segmentation”

Paul W Murray, July 2017 Market segmentation through data mining: A method to extract behaviors from a noisy data set

Peter Kolarovszki, 12 September 2016, The New Model of Customer Segmentation in Postal Enterprises

Wafa Qadadeh 2018 Customers Segmentation in the Insurance Company

Xin Fu, Xi Chen, Yu-Tong Shi, Indranil Bose, Shun Caia, 2017, “User segmentation for retention management in online social games”

