IJIRCCE

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

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 2, February 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 10, Issue 2, February 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1002040 |

Data Analysis on Super Market Sales for Betterment of the Sales in the Business using Machine Learning

Aryan Warang, Nikhil Jani, Navneet Giri, Samrat Mishra, Mr.Chintamani Chavan

Diploma Student, Dept. of I.T., Thakur Polytechnic, DTE Region, Mumbai, India

Diploma Student, Dept. of I.T., Thakur Polytechnic, DTE Region, Mumbai, India

Diploma Student, Dept. of I.T., Thakur Polytechnic, DTE Region, Mumbai, India

Diploma Student, Dept. of I.T., Thakur Polytechnic, DTE Region, Mumbai, India

Professor, Dept. of I.T., Thakur Polytechnic, DTE Region, Mumbai, India

ABSTRACT: Data Analysis and Forecasting on Supermarket Sales Transactions is a pro-posed system which focus on the betterment of the sales in the business. The whole proposed system comprises mostly of two sections:
Insights, exploratory data analysis is a method of dealing with breaking down informative collections to compress their main characteristics, frequently using visual tools. Exploratory Data Analysis refers to the process of performing preliminary analyses on data in order to uncover patterns, detect anomalies, test hypotheses, and eliminate doubts using basic measurements and graphical representations. we base our research by dealing with a real-world problem in an enterprise. A RFM (recency, frequency, and monetary) model and K-means clustering algorithm are utilized to conduct customer segmentation and value analysis by using online sales data. Customers are classified into four groups based on their purchase behaviors. On this basis, different CRM (customer relationship management) strategies are brought forward to gain a high level of customer satisfaction. The effectiveness of our method proposed in this paper is supported by improvement results of some key performance indices such as the growth of active customers, total purchase volume, and the total consumption amount.

KEYWORDS: EDA, Customer segmentation, RFM analysis, K-mean clustering.

I. INTRODUCTION

Supermarket sales transactions that are focused on increasing sales in the firm. We are now seeing positive results from businesses that use Machine Learning (ML) and Artificial Intelligence (AI) to outperform their competitors and close more deals. According to the Harvard Business Review, sales teams who use these technologies report an increase in leads of more than 50%, as well as cost reductions of up to 60%. Interpret client data, Improve sales forecasting, anticipate consumer demands, and Efficient transaction sales are just a few of the possibilities.

We handle the information from the beginning when it comes to data analysis. EDA (exploratory data analysis) is used in statistics to: better comprehend the data, develop an intuition about the data, produce hypotheses, discover insights, and visualize the data. We used RFM (recency, frequency, monetary) analysis to segment the client after visualising the data. The RFM model determines when consumers buy (Recency), how frequently they buy (Frequency), and how much they buy (Amount) (Monetary). While a client's past purchases can be used to predict their future purchasing behaviour, an organisation can determine whether clients are worthy. We'll use K-mean clustering to obtain the RFM model score. However, we should inform the K-means algorithm how many clusters we require. We'll use the Elbow Method to figure it out. The Elbow Method simply determines the optimal group size for optimal dormancy. The mean value of Recency, Frequency, and Revenue represents a better comprehension of the result. We may segment the data using the low-value, mid-value, and high-value filters.

By combining RFM and K-means approaches, we achieve customer segmentation and offer management options. On this foundation, we do customer segmentation and value analysis using an RFM model and K-means algorithm.

II. RELATED WORK

EDA (Exploratory Data Analysis)

The first clue that a visualisation is successful is that it reveals a problem in your data, detects outliers or abnormal events, and uncovers interesting relationships between variables [1]. Exploratory data analysis is a statistical way to

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 10, Issue 2, February 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1002040 |

studying data sets in order to summarise their essential properties, frequently using visual tools. With the use of summary statistics and graphical representations, exploratory data analysis refers to the crucial process of doing first investigations on data to uncover patterns, spot anomalies, test hypotheses, and check assumptions.

RFM Model

Hughes of the American Database Institute proposed the RFM concept in 1994 [2]. It has been frequently utilised for measuring customer lifetime value [3] as well as customer segmentation and behaviour analysis [4] as a popular instrument of customer value analysis. We present a brief description of the RFM model in the above literature in the next paragraphs.

Participating in a robust battle is a critical task for advertising if they want to have a successful business. Typically, advertising should use a scientific mode to identify highlight divisions and then execute an effective combat plan to target profitable prospects. The proposed system was meant to segment customers using the RFM (recency, frequency, andmonetary) principle. The RFM approach is used to determine the worth of a client. It's commonly used in database marketing and direct marketing, and it's garnered a lot of attention in the retail and professional services industries. This investigation proposes utilizing the accompanying RFM factors:

Recency(R): when people buy.

Frequency(F): how often they buy.

Monetary(M): how much they buy.

K-MEANS ALGORITHM

Clustering is the technique of grouping comparable objects from a collection of physical or abstract objects. Macqueen originally employed the K-means technique in 1967 [5], and it has since been widely used in a variety of fields, including data mining, statistical data analysis, and other business applications.

The goal of K-means clustering is to divide data into K groups, with each observation belonging to the cluster with the closest mean. Data mining, statistical data analysis, and other business contexts have all used the K-means clustering algorithm. The steps for K-means clustering are as follows: group the items into K initial clusters by associating each observation with the nearest mean; assign an item to the cluster with the closest centroid; and recalculate the cluster's centroid after adding or removing an item. Based on the enlarged RFM model, a K-means algorithm to classify consumer product loyalty in a B2B environment.

CRM AND DATA MINING

CRM stands for "customer relationship management," which is defined as "a business strategy to understanding customer behaviour through meaningful communications in order to increase customer acquisition, retention, loyalty, and profitability" (Swift, 2001). The CRM framework is critical to the firm from an architectural standpoint, and it may be divided into operational CRM and analytical CRM. CRM is at the heart of a number of industries, including telecommunications, insurance, and retail marketing[6]. To improve e-business processes, CRM considers the customer as a focal point. However, real-world applications pose a greater challenge to the CRM categorization model. As a result, data quality is a significant consideration for segmentation models. Other issues, such as different types of data and their differences, make data preparation and segmentation more difficult. Data mining techniques are used to solve data preparation and clustering issues. Missing data from unenthusiastic clients who do not provide all facts, misunderstandings, and customer blunders are common causes of incomplete data sets[7]. It is possible that high-dimensional data contains irrelevant information, resulting in the termination and upset of learning algorithms. As a result, it converts quite well for data mining jobs when it comes to feature selection. Assorted data is together from the specific dataset, mostly indefinite and infinite and more dissimilar formats as well as nominal or numerical. An effective customer-oriented strategy is critical since it aids in the strengthening of customer-business relationships [8]. The majority of strategies are used to identify loyal clients. One factor is demographics (gender, age, etc.).

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 10, Issue 2, February 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1002040 |

III. PROPOSED METHODOLOGY

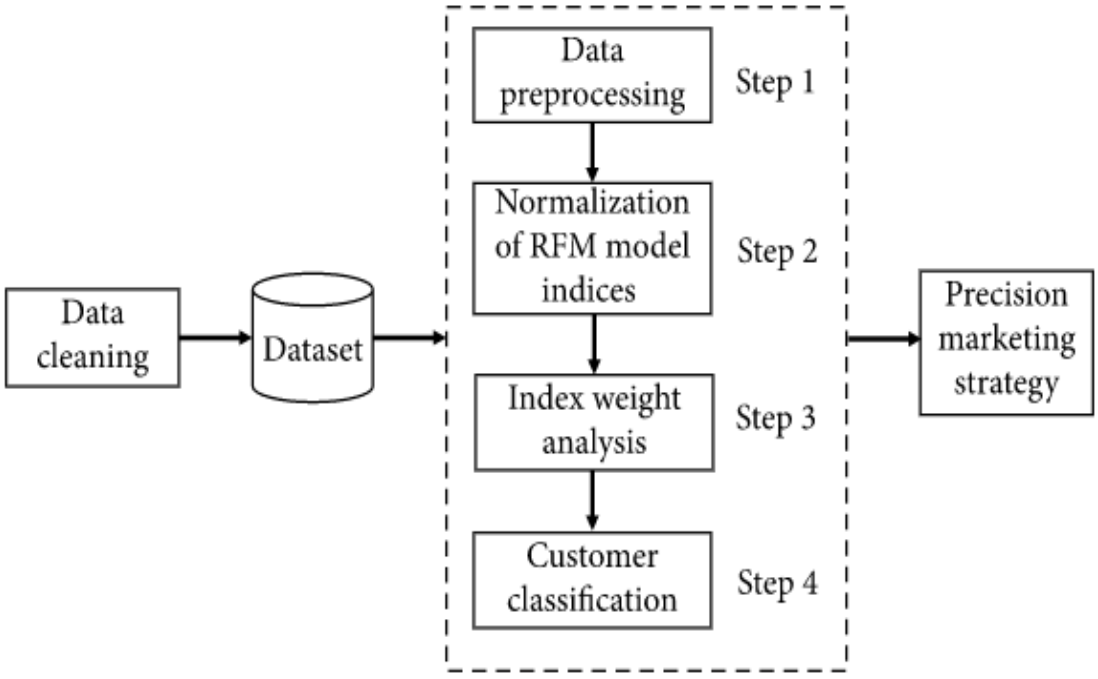

Fig.1.Steps of the proposed process

IV. PROPOSED ALGORITHM

Proposed System work for Data analysis and forecasting of the sales in the business following way:

Step 1: Get the data
Step 2: Understanding dataset
Step 3: Apply Exploratory Data Analysis to visualize data
Step 4: Apply RFM method for analyzing customer value
Step 5: Assign Score to the customer using K-means Algorithm
Step 6: Recalculate the cluster's centroid after adding or removing an item
Step 7: Classify consumer product loyalty in a B2B environment.

V. CONCLUSION AND FUTURE WORK

In my perspective, EDA is a strategy for organising and segregating educational collections in order to describe their rule ascribes using visual approaches. With the use of summaries and graphical representations, it is able to performing preliminary investigations on data to detect patterns, spot anomalies, and test hypotheses. There is a clearer understanding of which clients are priority and which are not, as well as what actions are required to enhance sales for low priority customers.

There are two possible study directions in the future. One is based on theory, while the other is based on experience. With data being updated on a daily basis, more appropriate algorithms are required to fit the new dataset for theoretical analysis.

REFERENCES

1. Karun Thankachan: "Automating Anomaly Detection for Exploratory Data Analytics", 2017.
2. A. M. Hughes, Strategic Database Marketing, Probus Publishing Company, Chicago, USA, 1994.
3. C.-H. Cheng and Y.-S. Chen, "Classifying the segmentation of customer value via RFM model and RS theory," Expert Systems with Applications, vol. 36, no. 3, pp. 4176–4184, 2009.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 10, Issue 2, February 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1002040 |

4. D. Chen, S. L. Sain, and K. Guo, "Data mining for the online retail industry: a case study of RFM model-based customer segmentation using data mining," Journal of Database Marketing & Customer Strategy Management, vol. 19, no. 3, pp. 197–208, 2012.
5. J. MacQueen, "Some methods for classification and analysis of multivariate observations," in Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, pp. 281–297, University of California Press, Berkeley, California, USA, January 1967.
6. Mohammadreza Tavakoli, Mohammadreza MolaviHajiagha., Vahid Masoumi, Majid Mobini: "Customer Segmentation and Strategy Development Based on User Behavior Analysis, RFM Model and Data Mining Techniques A Case Study", 2018.
7. Chen Hsuan-Kai, Hsin-Hung Wu, Jo-Ting Wei, Ming-Chun Lee: "Customer relationship management in the hairdressing industry: An application of data mining techniques, Expert Systems with Applications 40(18):7513-7518", December 2013.
8. Jun Wu,1,2 Li Shi,1 Wen-Pin Lin ,3 Sang-Bing Tsai ,4 Yuanyuan Li,2 Liping Yang,2 and Guangshu Xu5," An Empirical Study on Customer Volume2020|ArticleID8884227|https://doi.org/10.1155/2020/8884227, 19nov2020.

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

Scan to save the contact details