



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 11, November 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

Customer Segmentation using RFM and K-Means

Nityam Joshi¹, Niraj Chordia², Harshit Kadam³, Ishwari Gangadhar⁴

Department of Computer Engineering, Trinity Academy of Engineering, Pune, India

ABSTRACT: In today's competitive business landscape, understanding and catering to customer preferences have become paramount for sustained success. This project aims to leverage advanced data analysis techniques to achieve this objective. By combining the K-means clustering algorithm with the RFM model, the project intends to divide the customer base into meaningful segments that can guide the company's decision-making processes. The project's foundation lies in the collection and preprocessing of customer transaction data, which will include attributes such as purchase dates, transaction amounts, and product categories. Using this data, the RFM model will be applied to calculate three key metrics: Recency, Frequency, and Monetary value. These metrics will provide a comprehensive view of each customer's engagement level and monetary contribution to the business. Subsequently, the K-means clustering algorithm will be employed to group similar customers based on their RFM scores. This clustering process will unveil distinct customer segments that exhibit similar purchasing patterns and behaviours. By thoroughly analysing these segments, the project aims to uncover insights such as the most valuable customer group, the potential high growth segments, and the customers in need of re-engagement strategies. In conclusion, this project's significance lies in its potential to enhance customer understanding and tailor business strategies accordingly. The combination of K-means clustering and the RFM model promises a data-driven approach to customer segmentation, allowing the company to allocate resources more efficiently, design personalized marketing campaigns, and ultimately improve customer satisfaction and loyalty.

KEYWORDS: Customer Segmentation, RFM Model, K-means Clustering, Data Analysis, Customer Engagement, Purchase Patterns, Business Strategies, Customer Satisfaction.

I. INTRODUCTION

In the fast-paced realm of modern business, the ability to understand and cater to diverse customer needs is a critical factor for sustained growth. Customer segmentation emerges as a pivotal strategy, allowing businesses to categorize their clientele based on shared characteristics. This project embarks on the exploration of an advanced segmentation approach, combining the precision of RFM (Recency, Frequency, Monetary Value) analysis with the efficiency of K-means clustering. By scrutinizing transactional data, we seek to uncover nuanced patterns and behaviors, ultimately paving the way for targeted marketing strategies that resonate with specific customer segments.

The overarching goal of this project is to harness the rich insights embedded in transactional data to craft a sophisticated customer segmentation model. RFM analysis serves as the cornerstone, offering a three-dimensional lens through which we gauge the recency, frequency, and monetary value of customer transactions. Building upon this, K-means clustering adds a layer of sophistication by grouping customers with similar RFM profiles, allowing for the creation of highly focused and actionable segments. Through this dual-method approach, the project aims to unlock a deeper understanding of customer dynamics, empowering businesses to tailor their marketing initiatives with precision.

By adopting RFM analysis and K-means clustering, businesses can anticipate several key outcomes. First, the project promises to unveil distinct customer segments, each with its own set of characteristics and preferences. This, in turn, lays the groundwork for targeted marketing strategies, ensuring that promotional efforts resonate with specific customer needs. Furthermore, the application of data-driven techniques like the Elbow Method and Silhouette Analysis guarantees the selection of an optimal number of clusters for robust and meaningful segmentation. Through these endeavors, the project aspires to not only enhance marketing efficacy but also establish a framework for ongoing customer relationship management and personalized engagement.

II. LITERATURE SURVEY

Jiang and Tuzhilin (2009) identified that both customer segmentation and buyer targeting are necessary to improve the marketing performances. These two tasks are integrated into a stepbystep approach, but the problem faced is unified optimization. To solve the problem, the author proposed the K-Classifiers Segmentation algorithm. This approach focuses on distributing more resources to those customers who give more returns to the company. A sizable amount authors had written about different methods for segmenting the customers.

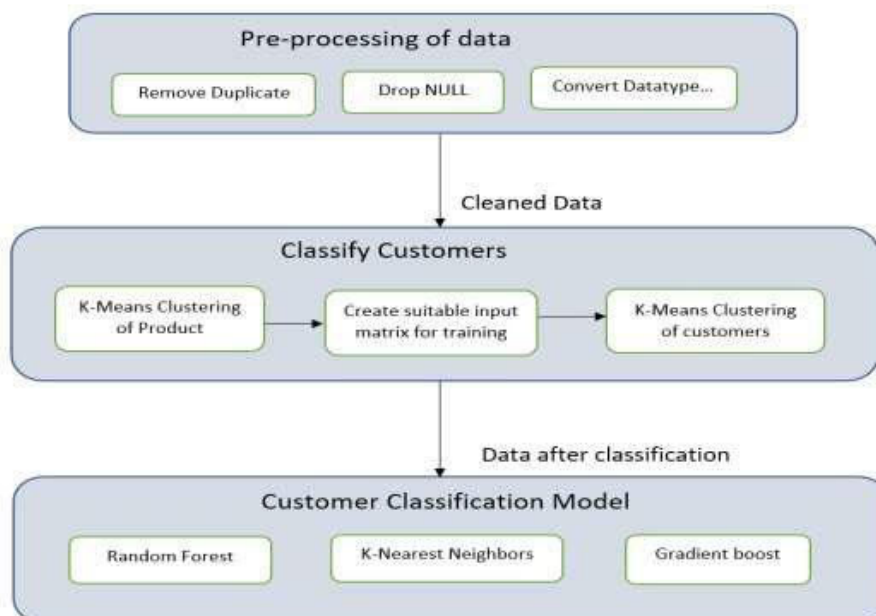
He and Li (2016) suggested a three-dimensional approach to improving the customer lifetime (CLV), the satisfaction of the customer and customer behaviour. The authors have concluded that the consumers are different from one another and so are their needs. Segmentation assists in finding their demand and expectations and proving a good service.

Cho and Moon (2013) proposed a customized recommendation system using weighted frequent pattern mining. Customer profiling is performed to find the potential customers using the RFM model. The author has defined varied weights for each transaction to generate weighted association rules through mining. Using the RFM model will provide a more accurate recommendation to the customer which in turn increases the profit of the firm.

Zahrotun (2017) used the customer data from online to identify the finest customer using Customer Relationship Management (CRM). By applying the CRM concept for online shopping, the author identifies the potential customers by segmenting them which helps us in increasing the profits for the company. To perform customer segmentation and marketing to customers in an accurate way the Fuzzy C-Means Clustering Method is used. Thus, this helps the customers to get special facilities in more than one category in the appropriate marking strategies according to their needs.

RFM is a popular model introduced by Hughes (2011) which has been employed to measure customer lifetime value in various area of applications e.g. retail banking (Hosseini and Shabani 2015; Khajvand and Tarokh 2011), hygienic industry (Parvaneh et al. 2012, 2014), retailing (Abirami and Pattabiraman 2016; Doğan et al. 2018; Hu and Yeh 2014; Serhat et al. 2017; You et al. 2015), telecommunication (Akhondzadeh-Noughabi and Albadvi 2015; Song et al. 2017), and tourism (Dursun and Caber 2016). The RFM model comprises three attributes: recency (R), frequency (F), and monetary (M). Due to the significant importance of the M attribute from the banking viewpoint, it is forecasted in this study.

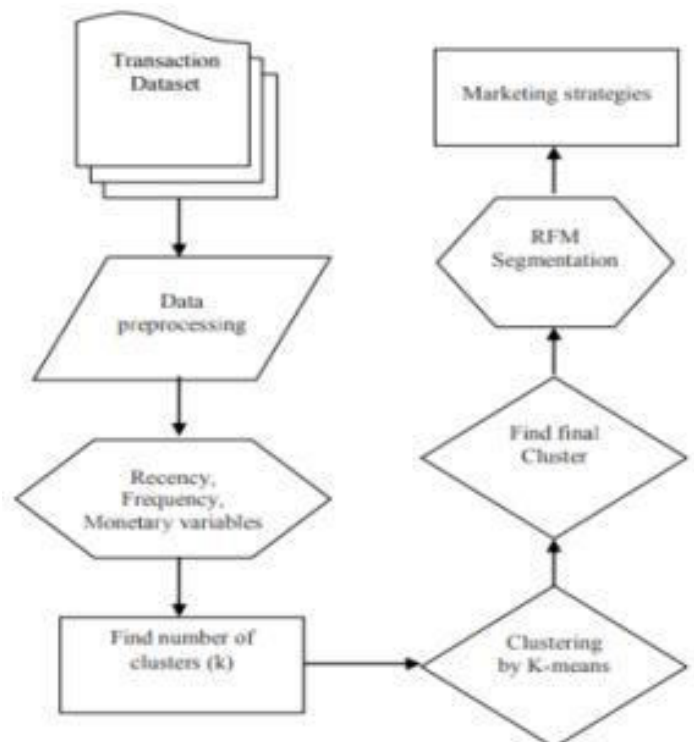
III. SYSTEM ARCHITECHTURE



The architecture of Customer Segment Pro, designed for effective customer segmentation using K-means and RFM analysis, comprises a web-based user interface for user interaction and CSV file uploads. The platform ingests and validates data, conducts Exploratory Data Analysis (EDA), and employs K-means clustering and RFM analysis to

segment customers. Machine learning models ensure accuracy in segmentation. The results are presented through the user interface, offering visualizations and marketing strategies tailored to each segment. Real-time insights and data security measures are integral to the system, and scalability and integration with existing CRM systems are supported. User feedback is encouraged for continuous platform improvement, and machine learning models are managed for ongoing accuracy and relevance. The architecture prioritizes user friendliness, data security, and the generation of insights for data-driven marketing strategies.

IV. DESIGN METHODOLOGY



1. Transaction Dataset: Collect transactional data from various sources such as your company’s sales database, point-of-sale systems, or e-commerce platforms. Ensure that the dataset includes relevant information like customer IDs, transaction dates, and transaction amounts.

2. Data Preprocessing: Handle missing values: Impute or remove missing values, especially in critical fields like customer ID or transaction amount. Remove duplicates: Eliminate duplicate records to ensure the accuracy of analysis. Outlier handling: Identify and address outliers that may skew the RFM analysis.

3. RFM Variables: Calculate Recency (R): Determine the most recent transaction date for each customer and calculate the time difference between that date and the present. Calculate Frequency (F): Count the number of transactions for each customer over a specified period. Calculate Monetary Value (M): Sum the total monetary value of transactions for each customer.

4. RFM Segmentation: Determine quantiles: Divide the customers into segments based on quantiles for each RFM variable (e.g., quintiles - 20). Assign scores: Assign scores (1 to 5, with 5 being the highest) based on the quantile segmentation for each variable. Combine scores: Create an RFM score by combining the individual scores (e.g., R=3, F=4, M=5 results in an RFM score of 345).

5. Number of Clusters Determination: Elbow Method: Run the K-means algorithm for a range of K values and plot the sum of squared distances (inertia) against K. The "elbow" of the plot indicates the optimal number of clusters. Silhouette Analysis: Evaluate the silhouette score for different K values to find the number of clusters that maximizes the score.

6. Clustering by K-means: Apply the K-means algorithm with the determined optimal number of clusters. Assign each customer to a cluster based on the cluster centroids.

7. RFM Segmentation Post-Clustering: Analyze the characteristics of each cluster in terms of the original RFM variables. Refine segments based on the K-means clustering results. This step ensures that clusters are homogeneous

and actionable for marketing strategies.

8. Marketing Strategies: High-value customers: Develop loyalty programs, VIP events, or exclusive offers to retain and encourage repeat business. At-risk customers: Implement reactivation campaigns, personalized communications, or special promotions to bring them back. Regular customers: Provide personalized product recommendations, cross-sell opportunities, or targeted promotions to enhance their engagement.

V. CONCLUSIONS

In conclusion, the utilization of RFM (Recency, Frequency, Monetary) and K-means clustering for customer segmentation has proven to be a valuable strategy. This method has effectively categorized our customer base into distinct segments, each exhibiting unique behaviours and purchase patterns. By tailoring our marketing efforts to these segments, we can anticipate higher customer engagement and conversion rates. Customer segmentation using K-means and RFM analysis is a robust strategy for businesses to categorize their customer base into meaningful groups, aligning marketing efforts with specific customer behaviours and characteristics. This approach not only enables the development of targeted marketing campaigns but also facilitates the creation of a dynamic dashboard to monitor and analyze customer segments in real-time. By incorporating machine learning algorithms, companies can extract deeper insights from their customer data, refine their segmentation models, and provide personalized product recommendations and marketing content. This, in turn, leads to improved customer satisfaction, retention, and business growth, making it an essential tool for modern businesses in today's competitive landscape.

VI. ACKNOWLEDGEMENTS

We extend heartfelt gratitude to Mr. R.B. Lagdive, our project guide, whose expertise and support shaped this research. Thanks to Dr. Nilesh Uke, Dr. M.B. Wagh, and Trinity Academy of Engineering faculty for guidance. Special thanks to friends for valuable insights. Divine blessings kept our morale high. Their collective contributions made this research paper possible.

REFERENCES

1. He X., Li, C., 2016. The research and application of customer segmentation on e-commerce websites. In: 2016 6th International Conference on Digital Home (ICDH), Guangzhou, pp. 203–208. doi: 10.1109/ICDH.2016.050
2. Haiying, M., Yu, G., 2010. Customer Segmentation Study of College Students Based on the RFM. In: 2010 International Conference on E-Business and EGovernment, Guangzhou, pp. 3860-3863. doi: 10.1109/ICEE.2010.968
3. Sheshasaayee, A., Logeshwari, L., 2017. An efficiency analysis on the TPA clustering methods for intelligent customer segmentation. In: 2017 International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Bangalore, pp. 784–788.
4. Memon, K.H., Lee, D.H., 2017. Generalised fuzzy c-means clustering algorithm with local information. In: IET Image Processing, vol. 11, no. 1, pp. 1-12, 1
5. Zahrotun, L., 2017. Implementation of data mining technique for customer relationship management (CRM) on online shop tokodipers.com with fuzzy c-means clustering. In: 2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Yogyakarta, pp. 299–303.
6. Tong, L., Wang, Y., Wen, F., Li, X., Nov. 2017. The research of customer loyalty improvement in telecom industry based on NPS data mining. China Commun. 14 (11), 260–268. <https://doi.org/10.1109/CC.2017.8233665>.
7. Shah, S., Singh, M., 2012. Comparison of a Time Efficient Modified K-mean Algorithm with K-Mean and K-Medoid Algorithm. In: 2012 International Conference on Communication Systems and Network Technologies, Rajkot, pp. 435–437.
8. Srivastava, R., 2016. Identification of customer clusters using RFM model: a case of diverse purchaser classification. Int. J. Bus. Anal. Intell. 4 (2), 45–50.
9. Liu, C.C., Chu, S.W., Chan, Y.K., Yu, S.S., 2014. A Modified K-Means Algorithm – TwoLayer K-Means Algorithm. In: 2014 Tenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Kitakyushu, pp. 447–450. doi: 10.1109/IIH-MSP.2014.118.



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.423



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 **6381 907 438** **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details