

# Customer Segmentation in Banking Dataset using Machine Learning

Prof. Leena Aakone<sup>1</sup>, Aniket Parate<sup>2</sup>, Pranay Dandekar<sup>3</sup>, Dasharath Pache<sup>4</sup>

Department of Computer Science and Engineering, Wainganga College of Engineering & Management, Nagpur.

## ABSTRACT

Machine learning techniques analyze and extract useful information from data sets in order to solve problems in different areas. For the banking sector, knowing the characteristics of customers entails a business advantage since more personalized products and services can be offered. The goal of this study is to identify and characterize data mining and machine learning techniques used for bank customer segmentation, their support tools, together with evaluation metrics and datasets. We performed a systematic literature mapping of 87 primary studies published between 2005 and 2019. We found that decision trees and linear predictors were the most used data mining and machine learning paradigms in bank customer segmentation. From the 41 studies that reported support tools, Weka and Matlab were the two most commonly cited. Regarding the evaluation metrics and datasets, accuracy was the most frequently used metric, whereas the UCI Machine Learning repository from the University of California was the most used dataset. In summary, several data mining and machine learning techniques have been applied to the problem of customer segmentation, with clear tendencies regarding the techniques, tools, metrics and datasets.

**Keywords:** Customer Segmentation, Banking, Data, Clustering Methods.

## I. INTRODUCTION

The customer segmentation has the importance as it includes, the ability to modify the programs of market so that it is suitable to each of the customer segment, support in business decision; identification of products associated with each customer segment and to manage the demand and supply of that product; identifying and targeting the potential customer base, and predicting customer defection, providing directions in finding the solutions. The thrust of this paper is to identify customer segments using the data mining approach, using the partitioning algorithm called as K-means clustering algorithm Company's database. Customer segmentation is one of the applications of data mining which helps to segment the customers with similar patterns into similar clusters hence, making easier for the business to handle the large customerbase.

This segmentation can directly or indirectly influence the marketing strategy as it opens many new paths to discover like for which segment the product will be good, customizing the marketing plans according to the each segment, providing discounts for a specific segment, and decipher the customer and object relationship which has been previously unknown to the company.

Customer segmentation allows companies to visualize what actually the customers are buying which will prompt the companies to better serve their customers resulting in customer satisfaction it, also allows the companies to find who their target customers are and improvise their marketing tactics to generate more revenues from them.

## II. LITERATURE REVIEW

Over the years, the competition amongst businesses is increased and the large historical data that is available has resulted in the widespread use of data mining techniques in extracting the meaningful and strategic information from the database of the organization. According to, Clustering techniques consider data tules as objects. They partition the data objects into groups or clusters.

The key to meaningful segmentation is to define customer variables and attributes that are relevant to your unique business. The customers are becoming more concerned and sophisticated in how they navigate their shopping choices, and the online retailers are discovering the one-size-fits-all marketing approaches that aren't so effective any more.

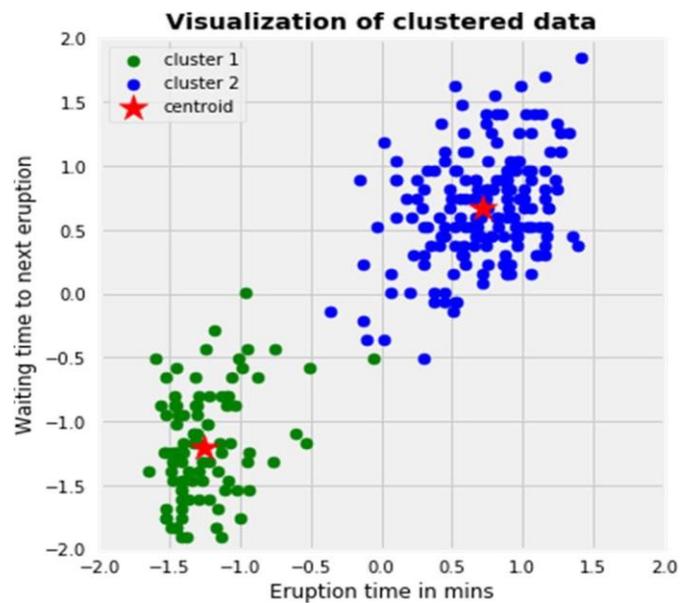
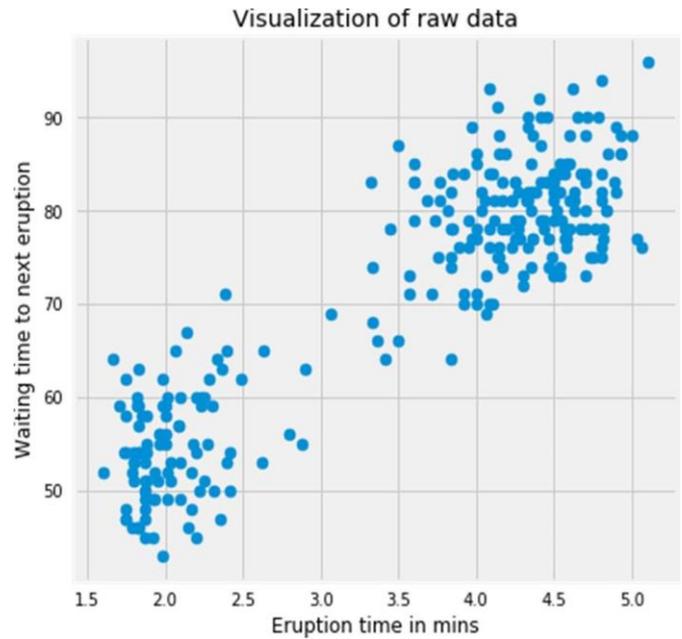
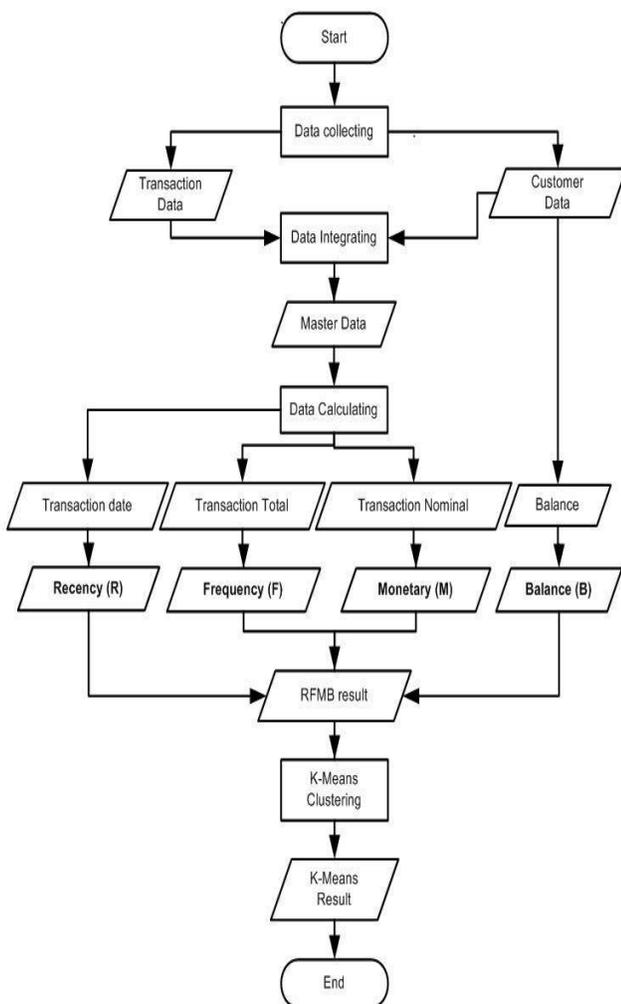
Xiaojun Chen, Yixiang Fang, Min Yang, Feiping Nie, Zhou Zhao and Joshua Zhexue Huang suggested a partitioned clustering algorithm named "PurTreeClust" for the faster clustering of customer's transaction records where one of the major key element in achieving successful modern marketing and customer relationship management is the customer segmentation or the clusters of customers [1].

Segments are not necessarily predictive in nature, although they are generally descriptive and serve as a type of classification that can be used to aid in understanding the future behaviors and the needs of customer.



**3. Mean shift Clustering** This clustering algorithm is a non-parametric iterative algorithm functions by assuming the all the data points in the feature space as empirical probability density function. The algorithm clusters each data point by allowing data point converge to a region of local maxima which is achieved by fixing a window around each data point finding the mean and then shifting the window to the mean and repeat the steps until all the data point converges forming the clusters.

**III.FLOWCHART**

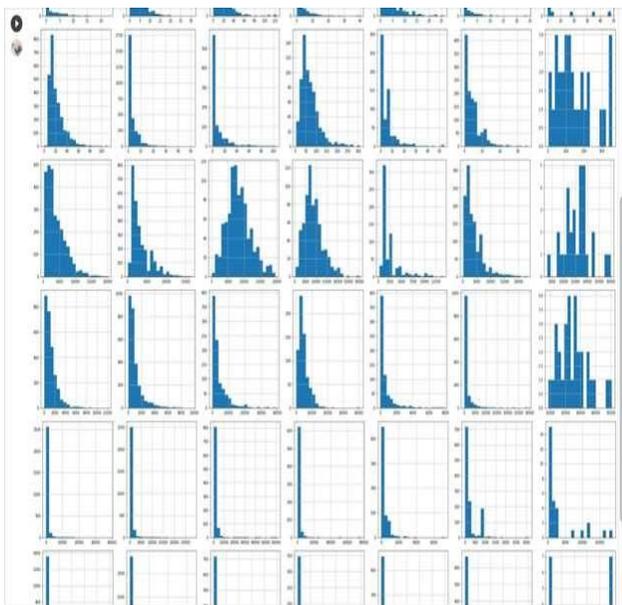


**IV. CONCLUSION**

As our dataset was unlabelled, in this paper we have opted for internal clustering validation rather than external clustering validation, which depend on some external data like labels. Internal cluster validation can be used for choosing clustering algorithm which best suits the dataset and can correctly cluster data into its opposite cluster. From the above visualization it can be observed that

Cluster 1 denotes the customer who has high annual income as well as high yearly spend. Cluster 2 represents the cluster having high annual income and low annual spend. Cluster 3 represents

customer with low annual income and low annual1999. Vol. 31, No. 3. [13] Vishish R. Patel and Rupa G. Mehta. MpImpact for External Removal. spend. Cluster 5 denotes the low annual income but high yearly spend. Cluster 4 and cluster 6 denotes the customer with medium income and medium spending score.



## V. REFERENCES

1. Blanchard, Tommy. Bhatnagar, Pranshu. Behera, Trash. (2019). Marketing Analytics Scientific Data: Achieve your marketing objectives with Python's data analytics capabilities. S.I: Packt printing is limited
2. Griva, A., Bardaki, C., Pramatari, K., Papakiriakopoulos, D. (2018). Sales business analysis: Customer categories use market basket data. Systems Expert Systems, 100, 1-16.
3. Hong, T., Kim, E. (2011). It separates consumers from online stores based on factors that affect the customer's intention to purchase. Expert System Applications, 39 (2), 2127-2131
4. Hwang, Y. H. (2019). Hands-on Advertising Science Data: Develop your machine learning marketing strategies... using python and r. S.I: Packt printing is limited Puwanenthiren Premkanth, - Market Classification and Its Impact on Customer Satisfaction and Special Reference to the Commercial Bank of Ceylon PLC. Global Journal of Management and Business Publisher Research: Global Magazenals Inc. (USA). 2012. Print ISSN: 0975-5853. Volume 12 Issue1.
5. Puwanenthiren Premkanth, - Market Classification and Its Impact on Customer Satisfaction and Special Reference to the Commercial Bank of Ceylon PLC. Global Journal of Management and Business Publisher Research: Global Magazenals Inc. (USA). 2012. Print ISSN: 0975-5853. Volume 12 Issue
6. Sulekha Goyat. "The basis of market segmentation: a critical review of the literature. www European Journal of Business and Management.iiste.org. 2011. ISSN 2222-1905 (Paper) ISSN 2222-2839 (Online). Vol 3, No.9, 2011.
7. By Jerry W Thomas. 2007. Accessed at: www.decisionanalyst.com on July 12, 2015.
8. T.Nelson Gnanaraj, Dr.K.Ramesh Kumar N.Monica. AnuManufactured cluster analysis using a new algorithm from structured and unstructured data.

International Journal of Advances in Computer  
Science and Technology. 2007. Volume 3, No.2

9. McKinsey Global Institute. Big data. The  
next frontier is creativity, competition and  
productivity. 2011. Accessed at:  
[www.mckinsey.com/mgi](http://www.mckinsey.com/mgi) on July 14, 2015. Research.

10.