Journal of Engineering Sciences
Vol 13, Issue 06, June/2022
ISSN NO:0377-9254

CUSTOMERS SEGMENTATION USING K MENAS AND GAUSSIAN MIXTURE

Dr. G Naga Chandrika, Assistant Professor, Ph.D., Department of IT, nchandrika_g@vnrvjiet.in

G. Varshith, BTech, Department of IT, varshithgandhe2k@gmail.com

N.Bhargav Reddy, BTech, Department of IT, nagirthibunny@gmail.com

G.Gurubrahmaiah, BTech, Department of IT, gurunayak4890@gmail.com

G Siva Sai Teja, BTech, Department of IT, sivasrinivas135@gmail.com

ABSTRACT: Customer segmentation is a marketing strategy in improving customer relationships. Customer loyal behavior towards a product or service will greatly benefit the company because customers will continue to look for the product they want. Many companies do not have a segmentation system to know the type of customers and measure customer value, even though the potential of data can be used for company profits. The purpose of this study is to find out the type of customer and measure the value of the customer so that the business owner can determine which customer gives the greatest benefit and which customer does not provide benefits. Identification of customer criteria in cluster formation based on RFM (Recency, Frequency and Monetary) values called clustering. This grouping method uses the K-Means Clustering algorithm and Gaussian mixture algorithms. The results of the Elbow Method are seen from the SSE (Sum Squard Error) of several cluster numbers. The value of the SSE difference is 2.7630 and the average Silhouette Index is 0.7210 in cluster 3. This can provide customer grouping results based on accurate RFM values in customer segmentation.

Keywords: *RFM. K-means clustering algorithm.*

1. INTRODUCTION

Customers are buyers and users of products by the company. Customers have the requirements and the final determinant whether or not the company product is given accordingly. Customers [1] are the company's market share that creates sales and profits for the company. Loyal behavior [2] is customer behavior towards a product or service which of course is profitable because customers will continue to look for the product they want. Just as in obtaining customer satisfaction, every business owner must improve the quality of their existing services. If the number of business customers increases, the sales and profits of the business will increase as well, and vice versa. The use of data mining techniques [3] is one solution to the problem of customer segmentation in the development of marketing strategies. Identification of customer criteria is based on the data obtained and grouped. These groups are called clustering. Clustering [4] [5] [6] is a data processing technique for grouping customers based on interactions that occur between customers and

business. Customer data is customer transaction data that has many attributes [7]. The existence of this attribute results in poor data processing. There needs to be a good selection of attributes to get more optimal results. To overcome the problem of selecting customer attributes in the segmentation process, it is proposed to use the K-Means algorithm [8]. Determining customer segmentation requires a reference alternative as a benchmark in its assessment, the reference in question can be in the form of data and information obtained from several sources such as recency, frequency and monetary values of RFM. Information obtained from related sources is obtained through data observation methods. The data that has been collected can be the basis of assessment for segmentation of customers [9].

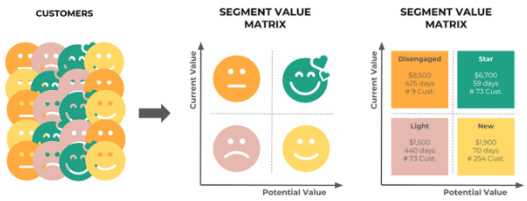

Fig.1: customer segementation

This research aims to find out the type of customer and measure the value of the customer so that the business owner can determine which customer gives a large profit and which customer does not provide benefits.

2. LITERATURE REVIEW

2.1 Customer Segmentation By Using Rfm Model And Clustering Methods: Acase Study In Retailindustry

Companies need to understand the customers' data better in all aspects. Detecting similarities and differences among customers, predicting their behaviors, proposing better options and opportunities to customers became very important for customer-company engagement. Segmenting the customers according to their data became vital in this context. RFM (recency, frequency and monetary) values have been used for many years to identify which customers valuable for the company, which customers need promotional activities, etc. Data-mining tools and techniques widely have been used by organizations and individuals to analysis their stored data. Clustering, which one of the tasks of data mining has been used to group people, objects, etc. In this paper we propose two different clustering models to segment 700032 customers by considering their RFM values. We detected that the current customer segmentation which built by just considering customers' expense is not sufficient. Hence, models that recommended in this research are expected to provide better customer understanding, well-designed strategies, and more efficient decisions. Keywords: Customer segmentation, RFM model, Clustering, K-means clustering.

2.2 A Review ON K-means DATA Clustering APPROACH

In data mining, clustering is a technique in which the set of objects are assigned to a group called clusters. Clustering is the most essential part of data mining. K-means clustering is the basic clustering technique and is most widely used algorithm. It is also known as nearest neighbor searching. It simply clusters the datasets into given number of clusters. Numerous efforts have been made to improve the performance of the K-means clustering algorithm. In this paper we

have been briefed in the form of a review the work carried out by the different researchers using K-means clustering. We have discussed the limitations and applications of the K-means clustering algorithm as well. This paper presents a current review about the K means clustering algorithm.

2.3 Learning orientation, firm innovation capability, and firm performance

Contemporary organizations require a strong learning orientation to gain competitive advantage. Based on in-depth interviews with senior executives and a review of the literature, the present investigation delineates four components of learning orientation: commitment to learning, shared vision, open-mindedness, and intraorganizational knowledge sharing. A framework is tested using data from a broad spectrum of US industries. Learning orientation is conceptualized as a second-order construct. Its effect on firm innovativeness, which in turn affects firm performance, is examined. The results generally support theoretical predictions, and some interesting findings emerge.

3. IMPLEMENTATION

So far, PD KaryaMulya already has 174 customers and records sales transactions every day. So as to facilitate the calculation of data mining that is useful in providing services to customers this research uses recency, frequency and monetary variables using the K-Means Clustering method. To overcome the problem of random initial cluster centers and local optima, this study conducted repeated trials at least 100times then the best optima global values were found. The process of calculating the Silhouette Index by calculating the average distance of the same cluster with different cluster average distances.

Drawbacks:

- The process of calculating the Silhouette Index average will find out in what cluster the results are, so it can define the best number of clusters to determine the characteristics of the RFM value data.

The purpose of this study is to find out the type of customer and measure the value of the customer so that the business owner can determine which customer gives the greatest benefit and which customer does not provide benefits. Identification of customer criteria in cluster formation based on RFM (Recency, Frequency and Monetary) values called clustering. This grouping method uses the K-Means Clustering algorithm. The results of the Elbow Method are seen from the SSE (Sum Squard Error) of several cluster numbers. The value of the SSE difference is 2.7630 and the average Silhouette Index is 0.7210 in cluster 3. This can provide customer grouping results based on accurate RFM values in customer segmentation. For this we are using K-means & Gaussian mixture algorithms in this paper

Advantages:

1. The quality scores generated by the proposed framework closely follow the ground truth quality scores for the movements.

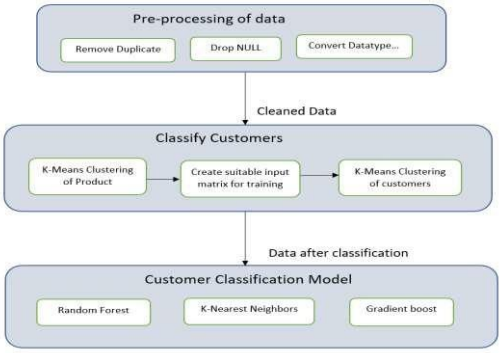

Fig.2: System architecture

4. ALGORITHM

K-MEANS:

k-means is a technique for data clustering that may be used for unsupervised machine learning. It is capable of classifying unlabeled data into a predetermined number of clusters based on similarities (k). In other words, the K-means algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. The 'means' in the K-means refers to averaging of the data; that is, finding the centroid.

The number of clusters identified from data by algorithm is represented by 'K' in K-means. In this algorithm, the data points are assigned to a cluster in such a manner that the sum of the squared distance between the data points and centroid would be minimum. The K-means clustering algorithm is used to find groups which have not been explicitly labeled in the data. This can be used to confirm business assumptions about what types of groups exist or to identify unknown groups in complex data sets.

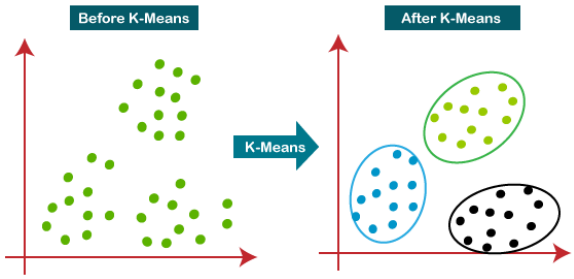

Fig.3: K-means model

GAUSSIAN MIXTURE:

Gaussian mixture models (GMMs) are often used for data clustering. You can use GMMs to perform either hard clustering or soft clustering on query data. To perform hard clustering, the GMM assigns query data points to the multivariate normal components that maximize the component posterior probability, given the data. Gaussian mixture models (GMMs) are a type of machine learning algorithm. They are used to classify data into different categories based on the probability distribution. Gaussian mixture models can be used in many different areas, including finance, marketing and so much more!

Gaussian Mixture Models (GMMs) assume that there are a certain number of Gaussian distributions, and each of these distributions represent a cluster. Hence, a Gaussian Mixture Model tends to group the data points belonging to a single distribution together.

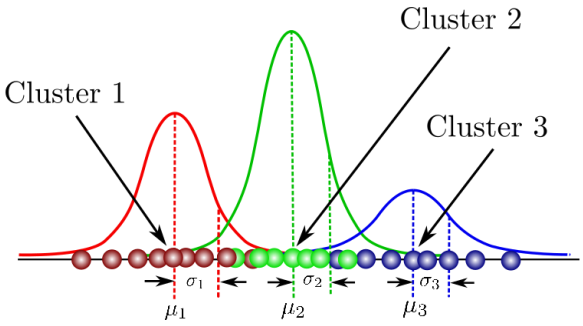

Fig.4: Gaussian mixture model

5. EXPERIMENTAL RESULTS

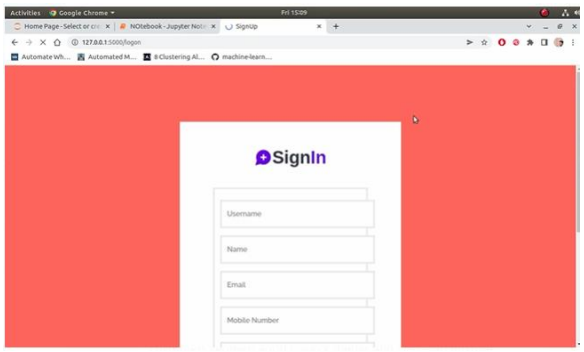

Fig.5: Registration screen

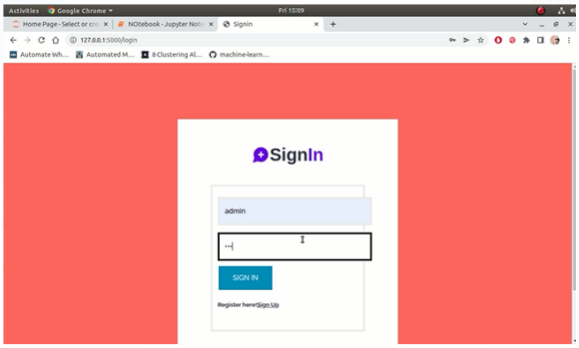

Fig.6: login

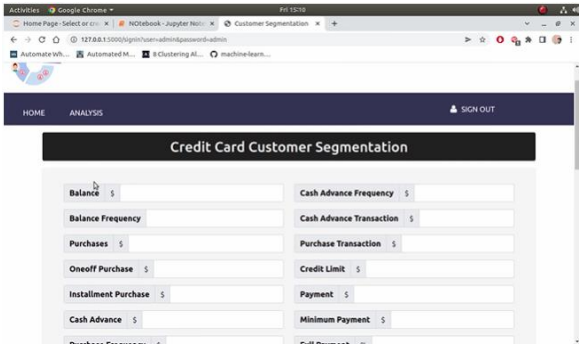

Fig.7: Home screen

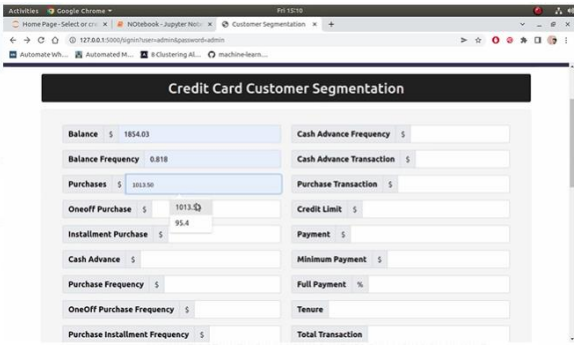

Fig.8: input values

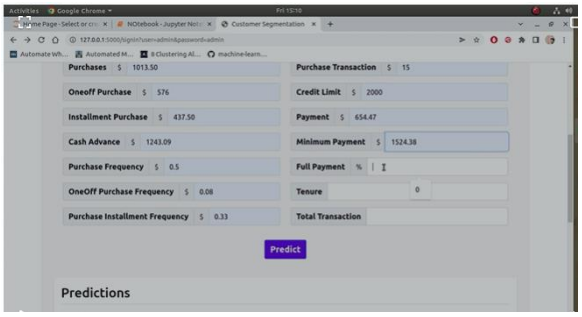

Fig.9: Predict

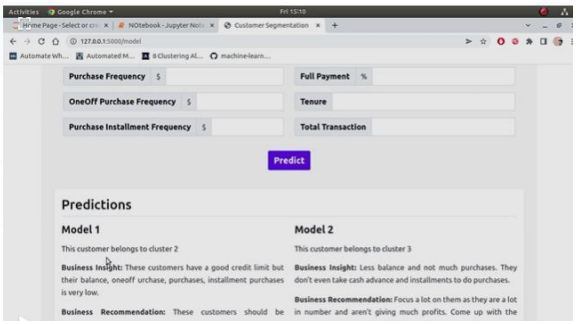

Fig.10: Predicted result

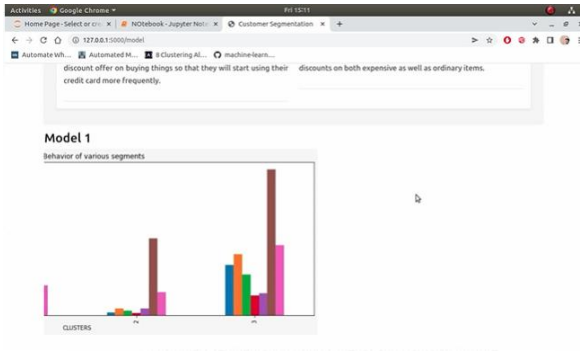

Fig.11: Model-1 graph

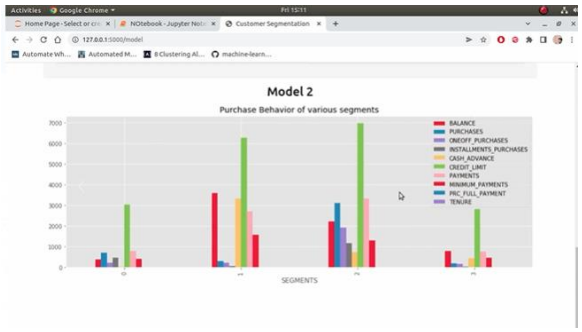

Fig.11: Model-2 graph

6. CONCLUSION

Based on activities during the process of making a customer segmentation system based on RFM values using the K-Means algorithm and Gaussian mixture models in PD KaryaMulya, it can be concluded based on the tested data that there are 3 clusters with their respective characteristics. The first cluster consists of 69 customers who are categorized as everyday shoppers. The second cluster consists of 95 customers and categorized as dormant customers. Then cluster 3 with 11 golden customers. The biggest difference value of Sum of Square Error is 2.7630 and the Silhouette Index is 0.7210. The results of the tests performed show that in the K-Means algorithm and Gaussian mixture models test, cluster 3 is the best cluster that the company can use as a promotional medium to loyal customers.

7. FUTURE SCOPE

Further research can be added with several criteria so that the system can be better and right on target. Development of supporting applications that use other tools and methods can be used as comparisons to the system that have been developed.. Future work can be developing the algorithm better segmented techniques. So there is a scope of improvement in the techniques.

REFERENCES

[1] Y. Li and F. Lin, "Customer segmentation analysis based on SOM clustering," 2008 IEEE International Conference on Service Operations and Logistics, and Informatics, Beijing, 2008, pp. 15-19.

[2] T. Zhang, X. Huang, J. Tang and X. Luo, "Case study on cluster analysis of the telecom customers based on consumers' behavior," 2011 IEEE 18th International Conference on Industrial Engineering and Engineering Management, Changchun, 2011, pp. 1358-1362.

[3] G. S. Linoff. and M. J. A. Berry, “Data Mining Techniques For Marketing, Sales, and Customer Relationship Management Third Edition”, Indiana: Wiley Publishing, Inc, 2011.

[4] X. Chen, Y. Fang, M. Yang, F. Nie, Z. Zhao and J. Z. Huang, "PurTreeClust: A Clustering Algorithm for Customer Segmentation from Massive Customer Transaction Data," in IEEE Transactions on Knowledge and Data Engineering, vol. 30, no. 3, pp. 559-572, 1 March 2018.

[5] U. R. Raval et al.,”International Journal of Computer Science and Mobile Computing”, Vol.5 Issue.5, May- 2016, pg. 191-203 . 7

[6] T. Yuanli and S. Liangshan, "Customer segmentation based on ant clustering algorithm," 2010 Second International Conference on Computational Intelligence and Natural Computing, Wuhan, 2010, pp. 133-136.

[7] J. Han and M. Kamber, “Data Mining Concept and Tehniques. San Fransisco: Morgan Kauffman”, ISBN 13: 978-1-55860-901-3, 2016.

[8] S. Shukla and S. Naganna, “A Review ON K-means DATA Clustering APPROACH”, International Journal of Information & Computation Technology.ISSN 0974-2239 Volume 4, Number 17, pp. 1847-1860, 2014.

[9] M. Namvar, M. R. Gholamian and S. KhakAbi, "A Two Phase Clustering Method for Intelligent Customer Segmentation," 2010 International

Conference on Intelligent Systems, Modelling and Simulation, Liverpool, 2010, pp. 215-219. [10] O. Dogan, E. Aycin and Z. A. Bulut, “Customer Segmentation By Using Rfm Model And Clustering Methods: Acase Study In Retailindustry”, International Journal of Contemporary Economics and Administrative Sciences ISSN: 1925–4423 Volume:8, Issue:1, Year:2018, pp. 1-19

[11] R. J. Calantone, S. T. Cavusgil, and Y. Zhao, "Learning orientation, firm innovation capability, and firm performance," Industrial Marketing Management, vol. 31, pp. 515-524, 2002 [12] D. D. Nimbalkar and P. Shah, “Data mining using RFM Analysis”, ISSN 2229-5518 International Journal of Scientific & Engineering Research, Volume 4, Issue 12, 2013

[13] F. Buttle, “Customer Relation Management, Concept and Technologies, Second Edition, Elsevier Butterworth-Heinemann”. ISBN: 978-1-85617-522-7 [14] A. R. Barakbah and K. Arai, “Centronit: Initial Centroid Designation Algorithm for K-Means Clustering”, EMITTER International Journal of Engineering Technology ISSN: 2443-1168 Vol.2, No.1, June 2014.

[15] D. Kamthania, A. Pahwa and S. S. Madhavan, “Market Segmentation Analysis and Visualization Using K-Mode Clustering Algorithm for ECommerce Business”, Journal of Computing and Information Technology, Vol. 26, No. 1, March 2018, 57–68