IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Vol. 12, Issue 4, April 2023

DOI: 10.17148/IJARCCE.2023.12448

CUSTOMER SEGMENTATION SYSTEM USING MACHINE LEARNING

Kartik Naphade[1], Durgesh Chaudhari[2], Aaditya Salunkhe[3], Suyog Patil[4], Ashish T. Bhole[5]

[1,2,3,4] UG Engineering Student, Department of Information Technology, SSBT's College of Engineering and Technology, Jalgaon, Maharashtra, India

[5]Associate Professor, Department of Information Technology, SSBT's College of Engineering and Technology, Jalgaon, Maharashtra, India

Abstract: Nowadays Customer segmentation became very popular method for dividing company's customers for retaining customers and making profit out of them, in the following study customers of different of organizations are classified on the basis of their behavioural characteristics such as spending and income, by taking behavioural aspects into consideration makes these methods an efficient one as compares to others. For this classification a machine algorithm named as k- means clustering algorithm is used and based on the behavioural characteristic's customers are classified. Formed clusters help the company to target individual customer and advertise the content to them through marketing campaign and social media sites which they are really interested in.

Keywords. Data visualization, Data analysis, Machine learning, Customer segmentation, K-means algorithm

I. INTRODUCTION

Today many of the businesses are going online and, in this case, online marketing is becoming essential to hold customers, but during this, considering all customers as same and targeting all of them with similar marketing strategy is not very efficient way rather it's also annoys the customers by neglecting his or her individuality, so customer segmentation is becoming very popular and also became the efficient solution for this existing problem. This project aims to help businesses better understand their customers and tailor their marketing strategies to meet the unique needs of each customer segment. By dividing customers into groups based on shared characteristics and behaviours, businesses can create more effective marketing campaigns, improve customer retention rates, and ultimately increase their revenue. Through the use of advanced machine learning techniques, this project will develop a system that can automatically segment customers based on various criteria such as demographics, purchase history, and online behaviour. By leveraging the power of artificial intelligence, this system will provide businesses with valuable insights into their customer base and help them make data-driven decisions to optimize their marketing efforts.

II. LITERATURE SURVEY

Over the years, the commercial world has become more competitive, as organizations such as these have to meet the needs and wants of their customers, attract new customers, and thus improve their businesses. The task of identifying and meeting the needs and requirements of each customer in the business is a very difficult task. This is because customers may vary according to their needs, wants, demographics, shapes, taste and taste, features and so on. As it is, it is a bad practice to treat all customers equally in business.

1. As per [1], Sukru Ozan proposes a case study on customer segmentation by using machine learning methods.
2. As per [2], Jayant Tikmani, et.al. proposes telecom customer segmentation based on cluster analysis an approach to customer classification using k-means.
3. As per [3], Chinedu Pascal Ezenkwu, et.al. proposes application of k-means algorithm for efficient customer segmentation: Strategy for targeted customer services.
4. As per [4], Potharaju, et.al. proposes data mining approach for accelerating the classification accuracy of cardiotocography. clinical epidemiology and global health.
5. As per [5], Yogita Rani, et.al. proposes a study of hierarchical clustering algorithm.
6. As per [6], Omar Kettani, et.al. proposes an agglomerative clustering method for large data sets.
7. As per [7], Snekha, et.al. proposes real time object tracking using different mean shift techniques.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Vol. 12, Issue 4, April 2023

DOI: 10.17148/IJARCCE.2023.12448

III. METHODOLOGY

Customer segmentation is an essential process for any business to understand its customers' behaviour and preferences. With the help of machine learning algorithms, you can automate the customer segmentation process and make it more accurate and efficient. Here's a method for developing a customer segmentation system using machine learning:

1. Collect and clean customer data: Collect the data related to customers such as their demographics, transactional history, website behaviour, and customer feedback. Clean and pre process the data to remove any missing values or outliers.

2. Define the customer segments: Determine the segments you want to create based on your business objectives. For example you may want to segment your customers based on their purchasing behaviour, age, or location.

3. Feature engineering: Create new features that can provide better insights into customer behaviour. For example, You can calculate the total spending of a customer, the frequency of their visits, or their purchase history.

4. Select a machine learning algorithm: Choose a suitable machine learning algorithm based on the type of problem you want to solve. For example, if you want to cluster customers based on their behaviour, you can use clustering algorithms like K-Means or DBSCAN.

5. Train the model: Split the data into training and testing sets. Train the model on the training set and tune its hyper parameters to get the best results.

6. Evaluate the model: Evaluate the performance of the model using metrics like accuracy, precision, recall, and F1- score.

7. Deploy the model: Deploy the model to your production environment, and use it to segment your customers in real time.

8. Monitor the model: Monitor the performance of the model regularly and retrain it if necessary. Also the quality of the data inputs to ensure that they are accurate and up-to-date.

K Means Clustering Algorithm:

K-Means is one of the most widely used clustering algorithms and is simple and efficient. The K-Means clustering beams at partitioning the ‘n’ number of observations into a mentioned number of ‘k’ clusters. The K-Means is an unsupervised learning algorithm and one of the simplest algorithms used for clustering tasks. The K-Means divides the data into non-overlapping subsets without any cluster internal structure. The values which are within a cluster are very similar to each other but, the values across different clusters vary extremely. K-Means clustering works really well with medium and large sized data. Despite the algorithm’s simplicity, K-Means is still powerful for clustering cases in data science. K-means technique for customer segmentation due to its following advantages: This technique suits for the data with numeric features and often terminates at local optimum. It is highly scalable and efficient for large data sets. It is fast in modeling and its result is more understandable. The aim of the K-Means algorithm is to divide M points in N dimensions into K clusters (assume k centroids) fixed a priori. These centroids should be placed in a wise fashion so that the results are optimal which otherwise can differ if locations of the centroids change. So, they should be placed as far as possible from each other. Each data point is then taken and associated with the nearest centroid until no data points are pending.

The algorithm works as follows:
Step 1: Specifying the number of clusters – k value.
Step 2: Centroids are initialized by shuffling the dataset and then randomly selecting k data points for the centroids without replacement.
Step 3: Repeat the iteration until there is no change to the centroids. i.e, assignment of data points to the clusters does not change. Recency, Frequency and Monetary are brought to the same scale and the data is normalized before clustering process. It is important to determine the optimum number of clusters i.e, “k value”.

IV. IMPLEMENTATION

The implementation process for a customer segmentation system using machine learning involves collecting and preparing customer data, selecting relevant features, choosing a suitable machine learning algorithm, training and evaluating the model, deploying it to a production environment, and monitoring its performance. The main steps include data

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Vol. 12, Issue 4, April 2023

DOI: 10.17148/IJARCCE.2023.12448

collection and preparation, feature selection, model selection and training, model evaluation, deployment, and monitoring.
Language- Python
Operating System- Windows 10
Python Libraries- Pandas, Matplotlib, Seaborn

V. PROPOSED SYSTEM

Customer segmentation is the process of dividing a customer base into groups of individuals with similar characteristics or behaviours. The goal of segmentation is to identify the most profitable customer segments and tailor marketing efforts to them.

Here is a proposed system for customer segmentation using machine learning:
• Data Collection Module: The module is responsible for collecting customer data from various sources such as sales data, website analytics, social media platforms, surveys, and feedback forms. The data collected can include customer demographics, purchase history, website behaviour, social media interactions, and more.
• Data Cleaning and Preprocessing Module: Module is responsible for cleaning and preparing the collected data for analysis. It involves tasks such as removing duplicates, filling in missing values, and standardizing data formats.
• Segmentation Algorithm Module: The module is the heart of the customer segmentation system. It involves applying segmentation algorithms to the preprocessed data to group customers based on certain criteria. Some commonly used algorithms for customer segmentation include:
• K-Means Clustering: The algorithm groups customers into clusters based on the similarity to each other. It is commonly used for segmentation based on customer demographics and behaviour.
• Decision Trees: The algorithm uses a tree-like model to segment customers based on a set of predefined criteria. It is commonly used for segmentation based on customer preferences and purchase history.
• RFM Analysis: The algorithm segments customers based on three factors - Recency, Frequency, and Monetary value of the purchases. It is commonly used for segmentation based on customer loyalty and value.
• Visualization Module: The module is responsible for presenting the segmented customer data in an easy-to understand visual format such as charts, graphs, and tables. It can help businesses to gain insights into the customer base and make data-driven decisions.

VI. RESULT

K-means algorithm is used with large dataset as its time complexity is almost linear and even it takes less space as it only requires to store only the data points and centroids. K means is fast and one of the simplest algorithms gives best result when data points are distinct but it fails when data points are highly overlapped or are non-linear. Moreover, it requires a pre-defined value of ‘k’ the number of clusters whereas in hierarchical clustering is no need of the value and number of clusters to be formed is easily determined by using a dendrogram. However hierarchical clustering doesn’t work well with large data set as its time complexity is O(n3) and requires O(n2) memory as it requires to store dendrogram.

Table 1: Customer Segmentation Using Machine Learning Approach

Index	Customer	Gender	Age	Annual Income	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40
5	6	Female	22	17	76
6	7	Female	35	18	6
7	8	Female	23	18	94
8	9	Male	64	19	3
9	10	Female	30	19	72
10	11	Male	67	19	14

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Vol. 12, Issue 4, April 2023

DOI: 10.17148/IJARCCE.2023.12448

Table. 1 shows the customer segmentation using machine learning approach shows the elbow method is based on the observation increasing the number of clusters can help to reduce the sum of within-cluster variance of each cluster.

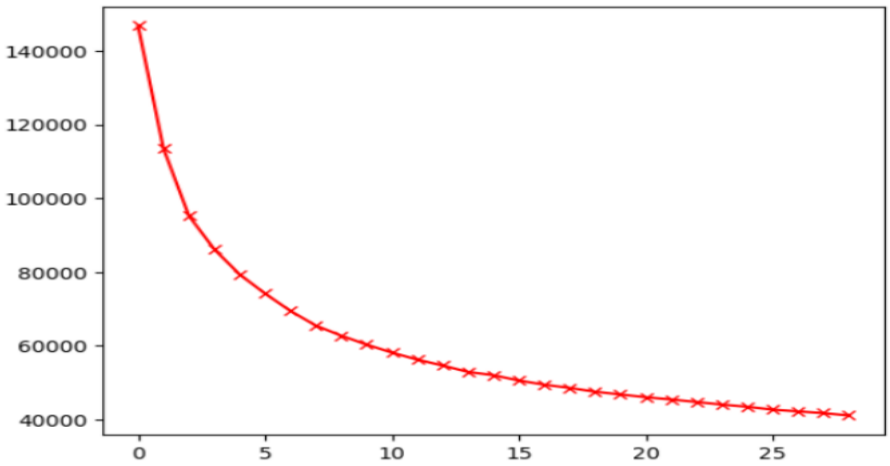

Figure 1: Customer Segmentation Using Machine Learning Approach

Fig. 1 shows the optimal K value is found to be 6 using customer segmentation machine learning approach. It works by evaluating within-cluster sum of squares for a range of cluster numbers and identifying the "elbow" point.

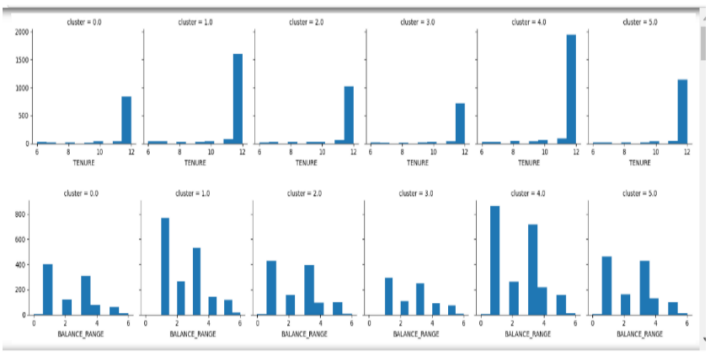

Figure 2: Interpretation of Cluster

Fig. 2 shows the plot indicates that there is a good structure to the interpretation of Cluster. The goal of interpretation is to identify the unique characteristics of each cluster and to use this information to develop effective marketing strategies and improve customer engagement.

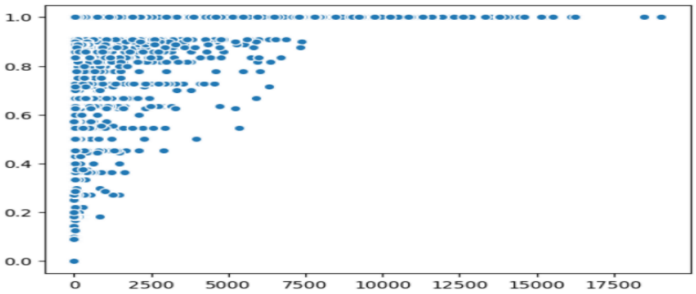

Figure 3: Outlier Detection

Fig. 3 shows the outlier detection code first filters and keeps the data points that belong to cluster label 0 and creates a scatter plot. outlier are data points that are far away from the majority of the observations in the dataset.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Vol. 12, Issue 4, April 2023

DOI: 10.17148/IJARCCE.2023.12448

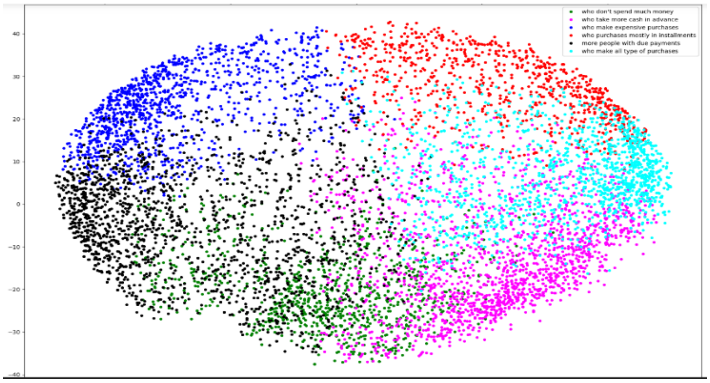

Figure 4: Final Cluster Form

Fig. 4 shows the code iterates filtering the data according to each unique class one iteration at a time. The result is the final visualization of all the clusters.

VII. CONCLUSION

Customer segmentation using machine learning can be a powerful tool for businesses to identify different groups of customers based on their behavior, preferences, and characteristics. This allows businesses to tailor their marketing strategies to each group's unique needs and preferences, improving customer satisfaction and retention, and ultimately increasing revenue. Machine learning algorithms such as clustering, decision trees, and neural networks can be used to automate the segmentation process and provide more accurate and efficient results. The success of a customer segmentation system using machine learning relies on the quality of the data used for training the model and the chosen features to identify customer segments.

REFERENCES

[1] Sukru Ozan, "A Case Study on Customer Segmentation by using Machine Learning Methods", IEEE, Year: 2018.
[2] Jayant Tikmani, Sudhanshu Tiwari, Sujata Khedkar "Telecom customer segmentation based on cluster analysis An Approach to Customer Classification using k-means", IJIRCCE,Year: 2015.
[3] Chinedu Pascal Ezenkwu, Simeon Ozuomba, Constance kalu Electrical/Electronics and Computer Engineering Department, University of Uyo, Akwa Ibom State, Nigeria "Application of K-Means Algorithm for Efficient Customer Segmentation: A Strategy for Targeted Customer Services", IJARAI,Year: 2015.
[4] Potharaju, S. P., Sreedevi, M., Ande, V. K., & Tirandasu, R. K. (2019). Data mining approach for accelerating the classification accuracy of cardiotocography. Clinical Epidemiology and Global Health, 7(2), 160-164.
[5] Yogita Rani and Dr. Harish Rohil "A Study of Hierarchical Clustering Algorithm", IJICT, Year: 2013.
[6] Omar Kettani, Faycal Ramdani, Benaissa Tadili "An Agglomerative Clustering Method for Large Data Sets", IJCA, Year: 2014.
[7] Snekha, Chetna Sachdeva, Rajesh Birok "Real Time Object Tracking Using Different Mean Shift Techniques–a Review", IJSCE, Year: 2013.SulekhaGoyat"The basis of market segmentation: a critical review of literature", EJBM, Year: 2011.