Customer Segmentation Using Machine Learning

MANAS DABHANE[1], VAISHNAVI CHAMATKAR[2], SHIVAM IRDANDE[3], VEDANT DHATRAK[4], PROF. MADHAVI SADU[5]

[1, 2, 3, 4] *B. TECH Student, Department of Computer Science and Engineering, RCERT, Chandrapur, India*

[5] *Assistant Professor, Department of Computer Science and Engineering, RCERT, Chandrapur, India*

Abstract*— In today's competitive market landscape, understanding customer behavior is crucial for businesses to tailor their marketing strategies effectively. This project explores the application of machine learning techniques for customer segmentation, aiming to enhance marketing strategies and improve customer satisfaction. By analyzing diverse datasets encompassing demographic, transactional, and behavioral data, the project employs clustering algorithms such as K-means, hierarchical clustering, and DBSCAN to group customers based on similarities in their attributes and behaviors. Feature engineering techniques are utilized to extract meaningful insights, enhancing the accuracy of segmentation. Moreover, dimensionality reduction methods like PCA facilitate the visualization and interpretation of complex customer data. The project's outcomes include actionable insights for targeted marketing campaigns, personalized recommendations, and product customization. By implementing machine learning-driven customer segmentation, businesses can optimize resource allocation, foster customer loyalty, and ultimately drive sustainable growth in today's dynamic market environment.

***Index Terms*— Machine learning, Customer segmentation, K-means algorithm, Data Visualisation.**

I. INTRODUCTION

Customer segmentation using machine learning is a strategic approach employed by businesses to categorize their customer base into distinct groups based on shared characteristics and behaviors.[3] This segmentation enables companies to tailor their marketing strategies, product offerings, and customer experiences to better meet the needs and preferences of each segment, ultimately driving customer satisfaction and business growth. In this project, machine learning algorithms are leveraged to analyze vast amounts of customer data, including demographic information, purchase history, online behavior, and engagement metrics. [8] By applying advanced analytics techniques, such as clustering algorithms like K-means or hierarchical clustering, the dataset is partitioned into clusters of similar customers. The process begins with data collection and preprocessing, where raw data is cleaned, standardized, and transformed into a format suitable for analysis. Feature engineering may also be employed to extract meaningful insights from the data and enhance the performance of the machine learning models. Next, machine learning models are trained on the prepared dataset to identify patterns and relationships between customer attributes.[4] Unsupervised learning algorithms, which do not require labeled data, are commonly used for customer segmentation tasks because they can automatically detect hidden structures within the data. Once the models have been trained and validated, the resulting customer segments are interpreted and characterized based on their unique traits and behaviors.[9] Businesses can then develop targeted marketing campaigns, personalized recommendations, and tailored product offerings for each segment to optimize customer engagement and maximize profitability. [15] Overall, customer segmentation using machine learning empowers businesses to gain deeper insights into their customer base, identify untapped opportunities for growth, and deliver more personalized experiences that drive long-term customer loyalty and retention. By harnessing the power of data and artificial intelligence, companies can stay ahead of the competition and thrive in today's dynamic market landscape.

II. LITERATURE SURVEY

I.

i. "Customer Segmentation Using Machine Learning Techniques: A Review" by Gupta, Gaurav, et al. (2019) - This review paper provides a comprehensive overview of various machine learning algorithms used for customer segmentation and their applications in different industries.

ii. "A Comprehensive Review on Customer Segmentation Techniques Using Machine Learning" by Jain, Ashima, et al. (2020) - This paper explores the evolution of customer segmentation techniques, focusing on the role of machine learning algorithms in enhancing segmentation accuracy and effectiveness.
iii. "Machine Learning Approaches for Customer Segmentation: A Comprehensive Review" by Karimi, Amir-Mohammad, et al. (2019) - This comprehensive review discusses the application of machine learning algorithms such as clustering, classification, and association rule mining in customer segmentation tasks.
iv. "Customer Segmentation Using Machine Learning Algorithms: A Systematic Literature Review" by Öztürk, Hande, et al. (2021) - This systematic literature review provides insights into recent advancements in customer segmentation using machine learning algorithms and identifies emerging trends in the field.
v. "Application of Machine Learning Algorithms in Customer Segmentation: A Review" by Paul, Sukanta, et al. (2019) - This review paper presents a comparative analysis of machine learning algorithms for customer segmentation, discussing their strengths, weaknesses, and real-world applications.
vi. "A Review of Customer Segmentation Techniques Using Machine Learning" by Sheth, Smit, et al. (2020) - This review article examines various machine learning techniques employed for customer segmentation, highlighting methodological advancements and practical challenges.

III. METHODOLOGY

i. Data Preprocessing: This initial step involves collecting, cleaning, and preparing the dataset for analysis.[8] Data preprocessing may include handling missing values, encoding categorical variables, and scaling numerical features to ensure uniformity and compatibility with machine learning algorithms.
ii. Feature Engineering: [5] Feature engineering aims to create new features or transform existing ones to enhance the predictive power of the model. Techniques such as dimensionality reduction (e.g., PCA) and feature scaling are applied to extract relevant information and reduce noise from the dataset, facilitating more accurate customer segmentation.
iii. Model Selection and Training: [11] In this phase, various machine learning algorithms suitable for customer segmentation, such as K-means clustering, hierarchical clustering, and DBSCAN, are selected based on the characteristics of the dataset and the objectives of segmentation. These algorithms are trained on the preprocessed dataset to identify distinct customer segments based on similarities in their attributes and behaviors.
iv. Evaluation and Validation: The performance of the trained models is evaluated using appropriate metrics such as silhouette score, Davies–Bouldin index, or within-cluster sum of squares (WCSS).[6] Cross-validation techniques may be employed to ensure the robustness and generalizability of the segmentation models. Validation techniques such as holdout validation or k-fold cross-validation help assess model performance on unseen data and mitigate overfitting.
v. Interpretation and Actionable Insights: Finally, the segmented customer groups are interpreted to derive actionable insights for marketing strategies, product customization, and customer relationship management.[15] Visualization techniques such as scatter plots, heatmaps, and dendrograms may be used to gain insights into the characteristics and preferences of each customer segment, enabling businesses to tailor their offerings and communication strategies accordingly.

IV. PROPOSED SYSTEM

I.

i. Data Collection and Integration:[1] Gather relevant customer data from multiple sources such as CRM databases, transaction logs, website interactions, and social media platforms. Integrate these diverse datasets into a unified repository for analysis.
ii. Data Preprocessing and Cleaning:[1] Perform data cleaning tasks including handling missing values, removing duplicates, and dealing with outliers. Normalize or scale the data as necessary to ensure uniformity and compatibility with machine learning algorithms.

iii. Normalize the Data:[4] Normalization techniques such as Standardization (Z-score normalization), Min-Max Scaling, or Robust Scaling adjust the range of each feature, preventing biases in clustering algorithms like K-means. For instance, Standardization transforms data to have a mean of 0 and a standard deviation of 1, while Min-Max Scaling scales data to a fixed range, typically between 0 and 1.[11] After normalization, K-means or other clustering algorithms can effectively group customers based on similarities in normalized feature space, aiding in targeted marketing strategies and personalized customer experiences.

iv. Clustering data Using K-means Clustering Algorithm : In customer segmentation projects, the K-means clustering algorithm is a widely used method for grouping customers based on similarities in their attributes.[14] K-means aims to partition data into a predetermined number of clusters (K) by iteratively assigning data points to the nearest cluster centroid and updating the centroids based on the mean of the points assigned to each cluster. The algorithm converges when the centroids no longer change significantly. K-means relies on the Euclidean distance metric, making it sensitive to the scale and distribution of features. Before applying K-means, data normalization is often performed to ensure fair treatment of features. [12] Once clusters are identified, businesses can tailor marketing strategies and services to suit the distinct preferences and behaviors of each customer segment, ultimately enhancing customer satisfaction and maximizing business outcomes.

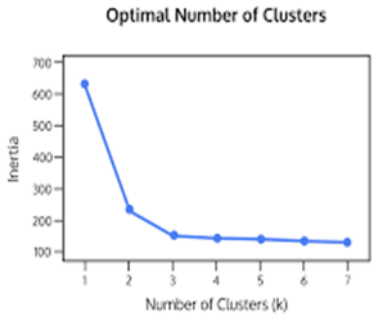

Fig1: - Optimal Numbers of Clusters

v. Hyperparameter tuning-Hyperparameter tuning in customer segmentation projects involves optimizing the parameters of machine learning algorithms to improve segmentation accuracy and effectiveness. [10] Common techniques include grid search, random search, and Bayesian optimization. For instance, in K-means clustering, tuning parameters such as the number of clusters (K) can significantly impact segmentation quality. In hierarchical clustering, parameters like the linkage method and distance metric are adjusted. Additionally, algorithms like DBSCAN require tuning parameters such as epsilon and minimum points. [6] Hyperparameter tuning is performed by systematically evaluating different combinations of parameter values and selecting the configuration that maximizes a chosen performance metric, such as silhouette score or within-cluster sum of squares (WCSS).

vi. Explaing customer Segments- Customer segmentation aims to divide a customer base into distinct groups to better understand and serve their needs. [7] By identifying meaningful segments, businesses can tailor their marketing strategies, product offerings, and customer experiences to each group's specific preferences and requirements.

```
customersdata.head()
```

	customer_id	products_purchased	complains	money_spent	clusters
0	649	1	0.0	260.0	4
1	1902	1	0.0	79.2	0
2	2155	3	0.0	234.2	4
3	2375	1	0.0	89.0	0
4	2407	2	0.0	103.0	0

Fig 2:- Customer Spending data

Machine learning techniques are often employed to automate the segmentation process and uncover hidden patterns in large datasets. Algorithms like K-means clustering, hierarchical clustering, and others can group customers based on similarities in their attributes, enabling businesses to gain insights into

customer behaviour and make data-driven decisions to improve customer satisfaction and loyalty.[8] Overall, customer segmentation using machine learning empowers businesses to target their resources more effectively and deliver personalized experiences that resonate with different customer segments.

	recency	frequency	monetary_value	cluster
1	6	182	1246.73	0
32	15	31	4311.90	0
49	355	73	1618.81	3
122	63	17	553.25	0
139	77	95	947.61	0

Fig 3: - Frequency and Recency of Customers

vii. Data Visualization-[13]Data visualization plays a crucial role in customer segmentation projects using machine learning by providing intuitive insights into complex datasets. Techniques such as scatter plots, histograms, and heatmaps visually represent the relationships between customer attributes and segmentation results. Visualization tools like t-SNE (t-Distributed Stochastic Neighbour Embedding) and PCA (Principal Component Analysis) can reduce high-dimensional data into two or three dimensions for easy interpretation. Clustering algorithms produce visual outputs, such as dendrograms for hierarchical clustering or cluster plots for K-means, helping to identify distinct customer segments.[12] Interactive dashboards allow users to explore segmentation results dynamically, facilitating deeper analysis and decision-making.

V. RESULT

After analysis of data and classifying customers with features annual income and spending score, we got clusters of customers & with formed clusters marketing team form strategies for customers specific recommendation to make value out of them

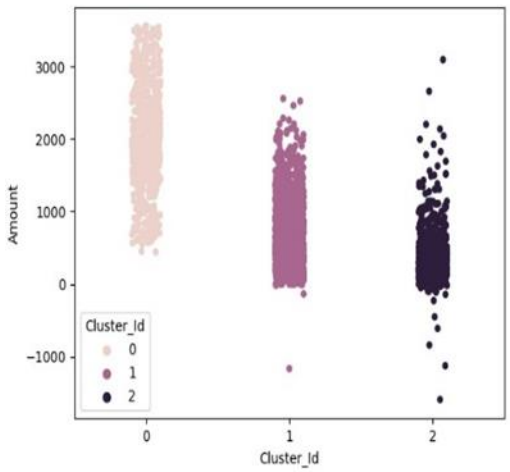

Fig 4: Customer Spending Score

The diagram below shows the recency or how frequently the customer visits to the store or how frequently the customer invests in the business owned by any organization. The data shows how much of time span is present between two consecutive visits of the customers for purchasing any product or services.

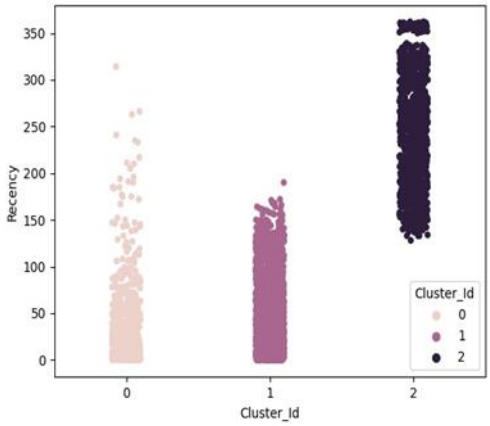

Fig 5: Recency of Customer Data

The diagram below shows the frequency of customers how frequently they visits to any shop or, how frequently they purchase goods or services of any shop.

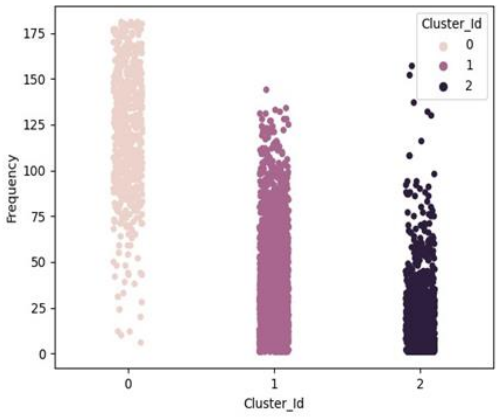

Fig 6: Frequency of Customer Data

REFERENCES

[1] "Customer Segmentation Using Machine Learning: A Case Study on an Online Retail Store" by Kumar, S., & Garg, V. (2019).

[2] "Customer Segmentation for E-commerce Using Machine Learning: A Review" by Sharma, A., & Panigrahi, P. K. (2020).

[3] "An Empirical Study on Customer Segmentation with Machine Learning" by Li, Y., & Han, Z. (2018). This study explores the application of Machine learning techniques for customer segmentation based on transaction data.

[4] "Customer Segmentation and Targeting Using Machine Learning: A Literature Review and Research Agenda" by Verhoef, P. C., Venkatesan, R., McAlister, L., Mela, C., & Kumar, V. (2017). This paper provides a comprehensive review of the literature on customer segmentation and targeting using machine learning.

[5] "Customer Segmentation in Marketing Using Data Mining Approach: A Review" by Gupta, A., & Dogra, S. (2019). This review article discusses various data mining techniques for customer segmentation in marketing.

[6] "Comparative Study of Machine Learning Algorithms for Customer Segmentation" by Saini, A., Kumar, A., & Goyal, L. M. (2020). This study compares the performance of different machine learning algorithms for customer segmentation tasks.

[7] "A Survey on Customer Segmentation Techniques" by Singh, N., & Chawla, P. (2019). This survey paper provides an overview of various customer segmentation techniques, including machine learning approaches.

[8] "Customer Segmentation Using Machine Learning Techniques: A Case Study in the Telecommunication Sector" by Karimi, R., & Mirzabeigi, M. (2020). This paper presents a case study of customer segmentation using machine learning in the telecommunication industry.

[9] "Personalization of E-commerce Store Using Customer Segmentation and Collaborative Filtering" by Kumar, A., Singh, M., & Singh, P. K. (2019). This paper discusses the personalization of e-commerce stores using customer segmentation and collaborative filtering techniques.

[10] "Customer Segmentation in Banking Industry Using Machine Learning Techniques" by Verma, A., & Rani, S. (2020).

[11] "Retail Customer Segmentation Using Machine Learning Techniques" by Sultana, S., & Islam, M. A. (2020). This paper presents a study on retail customer segmentation using machine learning techniques applied to transactional data.

[12] "Application of Machine Learning Techniques in Retail Customer Segmentation: A Review" by Jain, A., & Patel, R. (2019). This review paper provides insights into the application of machine learning techniques in retail customer segmentation.

[13] Vaishali R. Patel and Rupa G. Mehta “Impact of Outlier Removal and Normalization Approach in Modified k-Means Clustering Algorithm”, IJCSI,Year: 2011.

[14] "Comparison of Machine Learning Algorithms for Customer Segmentation: A Case Study in the Automobile Industry" by Li, C., & Ma, Y. (2019). This study compares the performance of different machine learning algorithms for customer segmentation in the automobile industry.

[15] "Machine Learning Techniques for Customer Segmentation in Social Commerce" by Shafique, M., Faisal, M., Aljohani, N. R., & Ali, T. (2020).