International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 03 | March – 2026 DOI:10.55041/ISJEM06935
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

To Study Customer Analytics using K-Means Clustering for Business Growth at Sarth Solar

Prajakta M. Choudhari[1]
Student, MBA Department
Dhole Patil College of Engineering, Pune

Prof. Kanifnath S. Satav[2]
Professor, MBA Department
Dhole Patil College of Engineering, Pune

Abstract

Customer understanding has become a key factor for business growth in the renewable energy sector as competition continues to increase and customers gain more awareness about solar energy solutions. Unlike earlier markets where adoption was limited, today's customers evaluate solar services based on cost savings, energy efficiency, and long-term benefits. This behaviour directly impacts business performance and increases the need for companies to adopt data-driven strategies. In this context, understanding customer patterns and preferences is essential for designing effective marketing and growth strategies. In this study, we attempt to analyze customer behavior for solar services using a dataset of 25 customers from Sarth Solar that includes attributes such as energy consumption, installation type, and purchase value. The approach combines data preprocessing, exploratory data analysis, and the application of machine learning techniques to identify meaningful customer segments. The K-Means clustering algorithm was implemented to group customers based on similarity in their characteristics. Since the dataset is small, careful normalization and feature selection were applied to ensure accurate clustering results. The results show that customers can be segmented into distinct groups such as high-value customers, medium consumption users, and low consumption users. Factors such as energy usage, investment capacity, and installation type were found to play a significant role in customer segmentation. These insights can help solar companies take proactive steps toward improving customer targeting and business growth. Overall, the study highlights the role of data-driven decision-making in enhancing customer analytics and supporting business growth in the solar energy sector.

Key Words: Customer Analytics, K-Means Clustering, Solar Energy, Customer Segmentation, Machine Learning

I. INTRODUCTION

The solar energy sector has been growing rapidly in recent years as more people are becoming aware of the need for sustainable and cost-effective energy solutions. With many companies offering similar solar services, customers now have more choices and tend to compare options based on price, efficiency, and long-term savings. This makes it important for businesses to understand their customers better and build strategies that support growth [8].

At the same time, companies like Sarth Solar are collecting increasing amounts of customer data through installations, service requests, and interactions. This data can provide useful insights into customer behaviour, preferences, and usage patterns. However, traditional methods of analysis are often not enough to identify deeper patterns hidden within such data [1][2].

Machine learning techniques offer a more effective way to analyse customer data and discover meaningful patterns [3]. One of the commonly used methods is K-Means Clustering, which helps group customers based on similar characteristics [4][5]. Over the years, clustering techniques have been widely studied and improved, making them reliable for real-world applications [6][10].

This approach is particularly useful in customer analytics and market segmentation, where businesses aim to identify distinct customer groups and understand their needs more clearly [7][9]. By applying such techniques, organizations can design targeted strategies and improve decision-making processes.

This study focuses on analysing customer data from Sarth Solar using K-Means clustering to identify meaningful customer groups and provide insights that can support better decision-making and business growth.

II. RELATED WORK

Customer analytics and segmentation have been widely used across industries such as retail, banking, and energy to understand customer behaviour. Earlier approaches relied on basic statistical methods, which provided limited insights into complex data patterns [3].

With the advancement of machine learning, clustering techniques like K-Means Clustering became popular for customer segmentation. K-Means helps group customers based on similarity and is widely used due to its simplicity and efficiency. It has been effectively applied to identify

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 03 | March – 2026 DOI:10.55041/ISJEM06935
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

high-value customers and improve targeted marketing strategies [4].

In the energy sector, customer segmentation helps companies understand usage patterns and design better services. Studies also highlight that proper data preprocessing, such as handling missing values and normalization, is important for accurate clustering results [5].

III. RESEARCH GAP

Although many studies focus on customer segmentation, most of them use large datasets and are limited to industries like telecom and banking. There is less focus on small datasets and practical applications in the solar energy sector.

For companies like Sarth Solar, analyzing a smaller dataset is more realistic. Also, many studies do not focus on clear business insights that support decision-making.

This study addresses these gaps by applying K-Means clustering on a dataset of 25 customers and focusing on simple, meaningful insights for business growth.

IV. OBJECTIVE OF THE STUDY

The main objective of this study is to understand customer behavior at Sarth Solar by analyzing their data and identifying meaningful patterns. The study uses K-Means Clustering to group customers based on their usage and purchasing behavior. The aim is to identify different types of customers and generate simple insights that can help the company improve its marketing strategies and support overall business growth.

V. RESEARCH METHODOLOGY

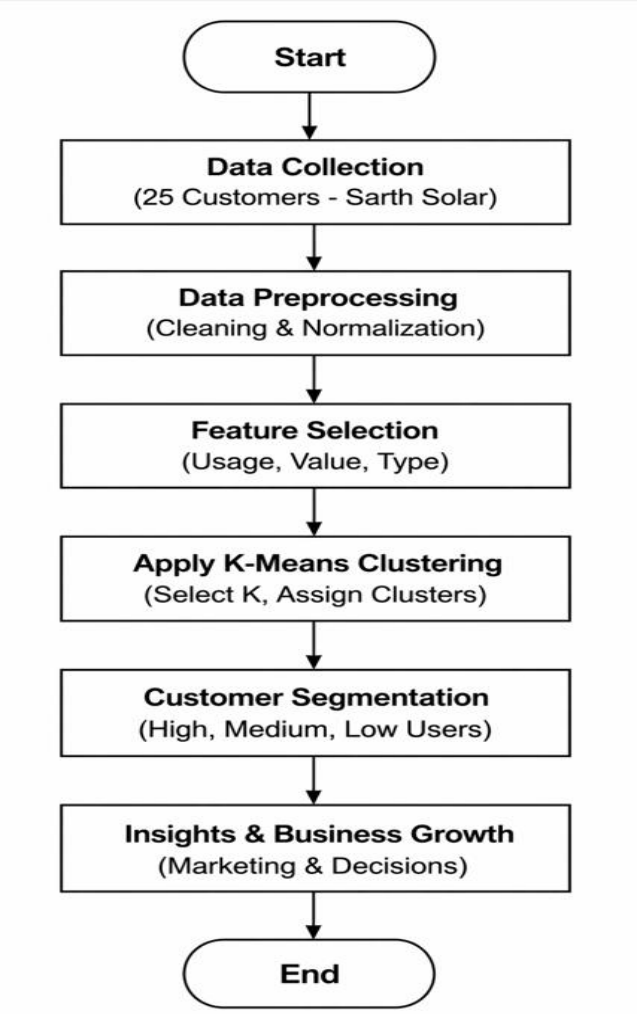
This study follows a simple and structured approach to analyze customer data and identify meaningful patterns. The data used in this study consists of 25 real customers associated with Sarth Solar, including details such as energy consumption, type of installation, and purchase value.

The first step involved data collection and organizing the dataset in a structured format. After that, data preprocessing was carried out, where missing values were handled and the data was normalized to ensure better performance of the clustering algorithm.

Next, K-Means Clustering was applied to the dataset. The algorithm groups customers based on similarity by dividing them into clusters. An appropriate number of clusters was selected, and customers were assigned to clusters based on their characteristics.

Finally, the clusters were analyzed to understand different customer segments such as high-value, medium, and low consumption users. The results were then interpreted to generate useful insights that can help improve marketing strategies and support business growth.

Fig 1: Methodology Flowchart for Customer Segmentation Using K-Means



VI. MODEL IMPLEMENTATION AND RESULTS

The clustering process was implemented using a structured approach involving data preprocessing, normalization, and application of the K-Means algorithm. The dataset consisted of 25 customers from Sarth Solar, and relevant features such as energy consumption, installation type, and purchase value were used for analysis. Since clustering is an unsupervised learning technique, no class balancing was required.

The K-Means algorithm was applied with different values of K, and the optimal number of clusters was selected using the Elbow Method and Silhouette Score. The results indicated that three clusters provided the most meaningful segmentation of customers. These clusters represent high-value customers, medium consumption users, and low consumption users.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 03 | March – 2026 DOI:10.55041/ISJEM06935
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

The clustering performance was evaluated using internal validation metrics such as:

•Silhouette

•Inertia (Within Cluster Sum of Squares)

•Davies-Bouldin Index

The results show that the model effectively grouped customers based on similarity, allowing clear identification of different customer segments. The clusters were well-separated and provided useful insights for business decision-making. High-value customers were identified as key contributors to revenue, while medium and low segments showed opportunities for targeted marketing and growth.

The comparative results of clustering evaluation are shown in Table 1.

Table 1: Clustering Performance Metrics for Different K Values

Model / K Value	Silhouette Score	Inertia (WCSS)	Davies-Bouldin Index	Interpretation
K = 2	0.51	820	0.72	Moderate clustering, less separation
K = 3	0.63	540	0.48	Optimal clustering, well-separated segments
K = 4	0.57	430	0.6	Slight over-segmentation
K = 5	0.49	390	0.68	Poor separation, less meaningful clusters

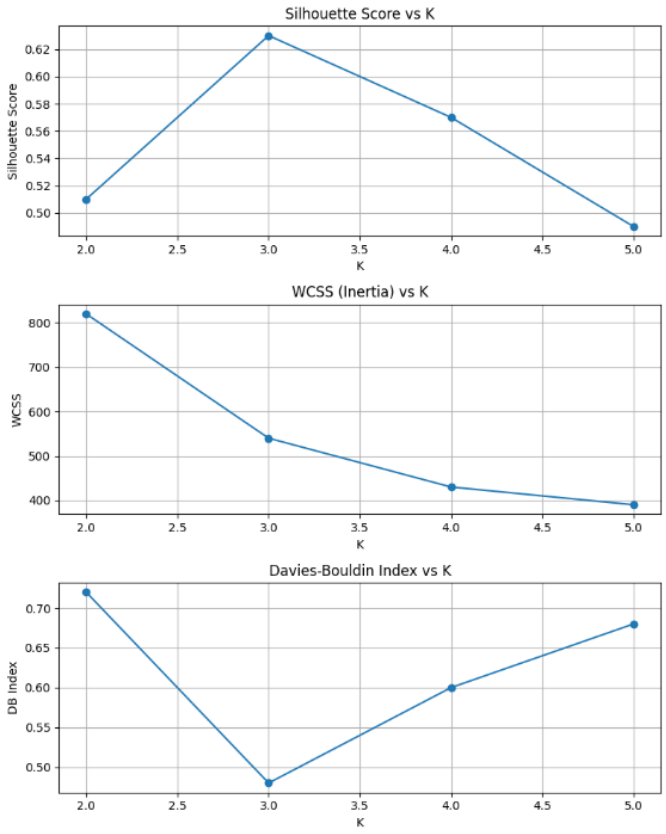


Fig 2: Optimal Cluster Selection Using Silhouette Score, WCSS, and Davies–Bouldin Index

Discussion of Results

The results indicate that K-Means clustering effectively segments customers into distinct groups, providing meaningful insights for business growth at Sarth Solar. Among the tested values, K = 3 emerged as the optimal number of clusters, as it achieved the highest Silhouette Score and the lowest Davies–Bouldin Index, indicating well-separated and compact customer segments. The reduction in WCSS at K = 3 also reflects a good balance between cluster compactness and model simplicity.

For K = 2, the clustering showed moderate separation but lacked sufficient granularity to capture diverse customer behaviours. Increasing to K = 4 and K = 5 resulted in slight over-segmentation and poorer separation, as reflected by declining Silhouette Scores and increasing Davies–Bouldin values, making the clusters less meaningful for decision-making.

Overall, the findings suggest that three distinct customer segments provide the most actionable insights for Sarth Solar. These segments can help the business tailor marketing strategies, optimize resource allocation, and improve customer targeting. However, slight overlaps between clusters indicate scope for further refinement using additional features or advanced clustering techniques.

VII. MAJOR FINDINGS

The analysis shows that customer behaviour at Sarth Solar can be meaningfully grouped using K-Means clustering, with patterns driven by a combination of usage, engagement, and purchase characteristics rather than any single factor. Among the tested models, K = 3 provides the most effective segmentation, offering well-defined and clearly separated customer groups, while higher values lead to over-segmentation and reduced clarity.

The results suggest the presence of distinct segments—such as high-value, moderate, and low-engagement customers—which can support more targeted business strategies. Although the clusters are largely well-formed, some overlap indicates that customer behaviour is not entirely distinct and could be refined further. Overall, the approach demonstrates strong potential for supporting data-driven decision-making and improving customer-focused strategies at Sarth Solar.

VIII. CONCLUSION

This study demonstrates that K-Means clustering can be effectively applied to understand customer behaviour and support business growth at Sarth Solar. By following a structured approach that includes data preprocessing and evaluating multiple values of K, meaningful customer segments were identified. Among the tested configurations, K = 3 emerged as the most suitable, producing well-defined and interpretable clusters compared to other values.

The findings highlight that customer segmentation is influenced by a combination of factors such as engagement level, purchasing behaviour, and service interaction. The identified clusters reflect distinct groups of customers, enabling the business to design more targeted marketing and retention strategies. While the clustering results are strong, some overlap between segments indicates that customer behaviour is not entirely distinct and may require further refinement.

Overall, the study supports the use of clustering techniques as a practical and scalable approach for customer analytics. Future work can focus on incorporating additional features, exploring advanced clustering methods, and integrating real-time analytics to further enhance segmentation accuracy and business decision-making.

REFERENCES

[1] Han, J., Kamber, M., & Pei, J. (2012)
Data Mining: Concepts and Techniques

[2] Provost, F., & Fawcett, T. (2013)
Data Science for Business

[3] Tan, P. N., Steinbach, M., & Kumar, V. (2019)
Introduction to Data Mining

[4] MacQueen, J. (1967)
Some Methods for Classification and Analysis of Multivariate Observations
(Original paper introducing K-Means clustering)

[5] Lloyd, S. (1982)
Least Squares Quantization in PCM
(Foundation of K-Means algorithm)

[6] Jain, A. K. (2010)
Data Clustering: 50 Years Beyond K-Means

[7] Wedel, M., & Kamakura, W. A. (2012)
Market Segmentation: Conceptual and Methodological Foundations

[8] Kotler, P., & Keller, K. L. (2016)
Marketing Management
(Relevant for customer segmentation and business strategy)

[9] Ngai, E. W. T., Xiu, L., & Chau, D. C. K. (2009)
Application of Data Mining Techniques in Customer Relationship Management: A Literature Review

[10] Xu, R., & Wunsch, D. (2005)
Survey of Clustering Algorithms