Customer Segmentation and Engagement Analysis using Swarm Intelligence Approach

Abstract:

In today's fast-paced business environment, **Artificial Intelligence** (AI) based modelling has become a cornerstone for building efficient, automated, and intelligent systems that address contemporary needs. Companies across various sectors leverage these advanced modelling methods to enhance their marketing efforts, striving to increase product development innovatively and drive business growth. The essence of successful marketing lies in offering the right product to the right customer at the right time. Achieving this requires a deep understanding of customer behaviour and purchasing patterns, which can be harnessed to generate significant profits for the company.Customer segmentation, the process of grouping customers based on common characteristics, plays a crucial role in this understanding. However, developing an efficient AI-based customer segmentation model that enhances digital marketing growth presents a considerable challenge. In this paper, we introduce an innovative **unsupervised deep learning model** that combines a **Self-Organizing Map** (SOM) with an modified **Social Spider Optimization** (MSSO) approach for effective customer segmentation.
Our methodology begins with a comprehensive feature engineering process, employing a ***swarm intelligence model*** known as ***Modified Social Spider Optimization (MSSO)*** to meticulously select the behavioural features of customers. These features are then used to cluster customers using a ***Self-Organizing Neural Network (SONN)***, which dynamically groups them based on identified patterns. Following this clustering, we classify the customers using a Deep Neural Network (DNN) model to fine-tune the segments and predict future behaviours.The experimental results demonstrate the robustness and high performance of our proposed model, showcasing its superior clustering and segmentation capabilities. By accurately identifying and grouping customers, our model significantly enhances targeted marketing efforts, leading to improved customer engagement and increased business profitability. This research not only contributes to the field of AI-driven customer segmentation but also provides practical insights for companies seeking to optimise their digital marketing strategies and drive sustainable growth.

1.Introduction:

1.1 Objective

The dynamic and rapidly evolving marketing environment presents businesses with complex and competitive decision-making challenges. To thrive in this landscape, **effective customer segmentation,** targeting, and positioning are paramount. This research aims to develop a sophisticated **AI-based customer segmentation model** leveraging **unsupervised deep learning techniques** to enhance IHCL's digital marketing strategies. The primary objective is to identify and categorise customer groups with similar behavioural patterns to facilitate targeted marketing, thereby increasing engagement and profitability.

1.2 Motivation

Customer segmentation is a critical and competitive endeavour, particularly in the **hospitality industry,** where understanding and catering to diverse customer needs is paramount. This motivates our research to develop an optimal and accurate segmentation strategy for the IHCL brand by delving deep into **customer engagement patterns** derived from **social media** interactions. Effective segmentation helps identify how to correlate and group customers based on their behaviours and expectations, enabling more **personalised and targeted marketing** strategies.
In the context of IHCL, customer segmentation is instrumental for decision-makers to make timely and precise decisions that enhance the marketing of specific services and experiences. Our research focuses on clustering customers based on their engagement with IHCL's Instagram posts, including metrics such as likes, comments, and overall interaction rates. By analysing these data points, we aim to segment customers according to their behavioural patterns and preferences.
The primary goal of this research is to implement a sophisticated AI-based model that leverages advanced techniques such as Modified Social Spider Optimization (MSSO), Self-Organizing Neural Networks (SONN), and Deep Neural Networks (DNN) to achieve accurate and meaningful customer segmentation. This approach not only aims to improve IHCL's marketing effectiveness but also provides a framework that can be adapted to other brands seeking to optimise their digital marketing strategies and enhance customer engagement in a competitive landscape.

1.3 Background

In the current marketing landscape, decision-makers are tasked with processing vast amounts of data to enhance profitability. A critical component of this process is uncovering valuable insights from customer data to inform and develop effective customer segmentation strategies (Glanz et al., 2020). The e-commerce datasets typically consist of raw data related to products and customer shopping behaviours, which makes extracting actionable insights challenging. Without a clear strategy based on this data, businesses struggle to achieve growth and improve decision-making.One common approach to customer segmentation is the use of clustering combined with Recency, Frequency, and Monetary (RFM) analysis. RFM analysis evaluates the most relevant aspects of sales, purchase amounts, and customer behaviours, serving as a foundational preprocessing method for data analysis (Aliyev, Ahmadov, Gadirli, Mammadova & Alasgarov, 2020; Kabasakal, 2020). By clustering customers based on RFM metrics, researchers and marketers can identify segments with similar behaviours, aiding in targeted marketing strategies (Gustriansyah, Suhandi & Antony, 2020; Rahman, 2020).Recent studies, such as He and Cheng (2020), demonstrate that combining RFM analysis with clustering algorithms can yield significant insights for customer segmentation. However, while this combination offers valuable segmentation, it often falls short in terms of interpretation and verification. Additionally, clustering algorithms typically require the initial number of clusters to be specified, which can lead to issues in distinguishing between different customer variants and their usage patterns.This paper is structured as follows: Section 2 reviews related approaches to customer segmentation. Section 3 introduces the proposed feature selection, clustering, and segmentation methodologies. Section 4 presents the experimental results and analysis. Finally, Section 5 concludes the work and discusses future research directions.

2. Related work

The field of customer segmentation has seen significant advancements, particularly with the integration of artificial intelligence (AI) and machine learning (ML) techniques. **Lee et al. (2021)** proposed a hybrid approach combining clustering and classification for customer segmentation. Their method employs **Artificial Bee Colony (ABC)** for **feature selection and RFM analysis** for basic data processing. They utilised clustering algorithms such as **k-means, fuzzy c-means, and the Wald method**, followed by an i**mproved fuzzy decision tree** for classification. Despite satisfactory results, the ABC algorithm's effectiveness in feature selection was limited.
Matias et al. (2021) developed a neural network with transfer learning for time series segmentation to detect similar patterns, using **Mask-based Neural Networks (NN**). Their model, validated on clinical and human activity datasets, achieved high precision and recall values, demonstrating the effectiveness of transfer learning in segmentation tasks.
Ansari (2021) explored deep learning-based customer segmentation for **business-to-business marketing automation** in the United States using **self-organising maps (SOM)**. The study highlighted SOM's capabilities in clustering, dimensionality reduction, and cluster pattern visualisation, providing insights for new marketing channels.

Janardhanan and Muthalagu (2020) investigated market segmentation using machine learning algorithms, focusing on profitable products, customer behaviour, and sales forecasting with the **Auto Regressive Moving Average (ARIMA) model**. They used the k-means algorithm for customer clustering based on weekly sales data, improving forecasting results.
Verma et al. (2021) reviewed the significance of AI in marketing, using conceptual, intellectual, and bibliometric analyses from 1982 to 2020. **Gacanin and Wagner (2019)** discussed autonomous customer experience management and the role of AI and ML in establishing intelligence. **Sha and Rajeshwari (2019)** examined AI advancements in e-commerce, highlighting AI's potential to enhance customer satisfaction and brand loyalty through automated online buying systems.
Chen et al. (2020) combined optimization approaches with ML models for e-commerce assessments, while Dekimpe (2020) applied text mining and ML in various sectors, including banking, finance, retail, art marketing, and tourism, to identify profitable customers. These studies underscore AI's suitability for product design and customer satisfaction.
Despite the growing interest in AI-based customer segmentation, there is a limited application of deep learning and swarm intelligence (SI) models in this domain. This gap motivated the current research to develop a deep learning and SI-based customer segmentation model for digital marketing, leveraging advanced techniques to enhance segmentation accuracy and business profitability.

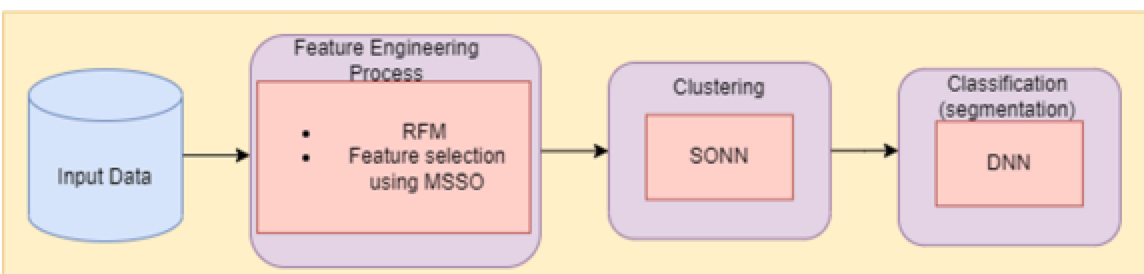


Fig. 1. Overview of the proposed Customer segmentation approach.

3. Proposed Methodology

The proposed customer segmentation approach is summarised in Fig. 1. The process begins with data preprocessing and feature engineering, specifically using Recency, Frequency, and Monetary (RFM) metrics to rank customer categories. Following this, feature selection is conducted using a swarm intelligence approach called MSSO (Modified Social Spider Optimization). Clustering is performed with the Self-Organizing Neural Network (SONN) algorithm to group customers based on their purchasing behaviours. Finally, the clustered data is classified using Deep Neural Networks (DNN) according to purchasing amounts or patterns. This segmentation is then utilised for business analytics to drive business growth.

Feature engineering, a critical component of data processing, enhances the performance of classifiers by building models with the most relevant features. The dataset comprises both numerical and categorical data, with the latter transformed into numerical values using data normalisation (min-max scaling). This transformation is represented by the equation:

$$Y = \frac{X - min(X)}{max(X) - min(X)}$$

RFM analysis ranks customers based on their purchasing behaviours: recency (how recently a customer has come to their page), frequency (how often they comment), and monetary value (how much they hit the like button).

3.1 Swarm Intelligence Model for Feature Selection Using MSSO

The original SSO algorithm assumes a spider web as the search space and the population of the candidate solution represents the spider. According to the fitness value (A. Alzghoul et al., 2021; Liu et al., 2020; Lv, Li, Feng & Lv, 2021, 2022; Mou, Duan, Gao, Liu & Li, 2022; Rostami et al., 2022; Rostami, Berahmand, Nasiri & Forouzandeh, 2021, 2020), each spider gets its weight. The colony cooperative behaviours are simulated with two sets of evolutionary operators. This model is used to solve the online global optimization issues with the constraint denoted in Eq. (2)

minimise F(x) where x = (x1 , x2 , …xd) ∈ **R^d** and x ∈ X

Where, F - non-linear function and X = {x ∈ R^d | Li ≤ x ≤ Ui, i = 1, 2, ..d} - reduced feasible search space with lower (L) and upper (U) limits.

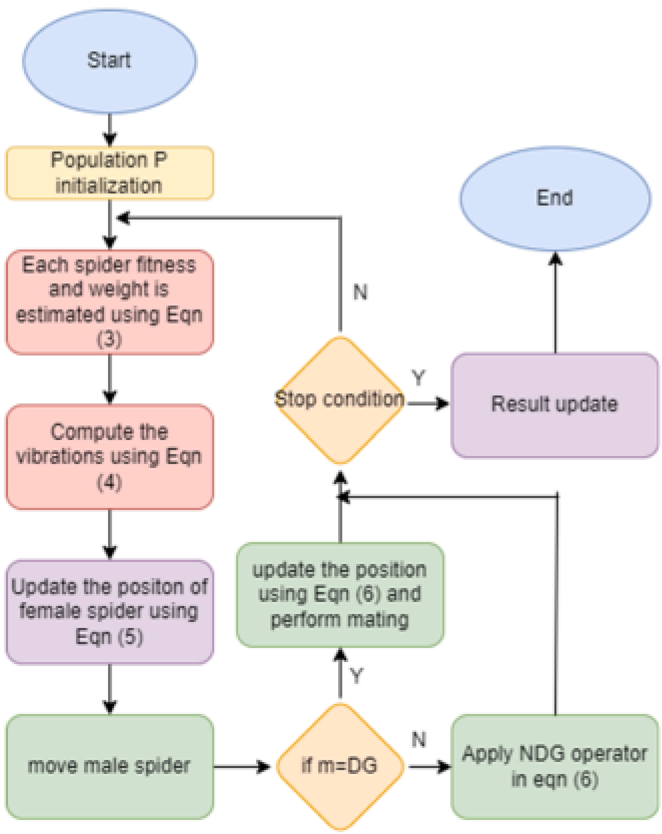


Fig. 3. Flow of standard SSO.

The MSSO algorithm models the search space as a spider web with the candidate solutions represented as spiders. Fitness values assign weights to each spider, and cooperative behaviours are simulated with evolutionary operators to optimise the feature selection process. The objective is to minimise a nonlinear function F(x) over a feasible search space X:

minimise F(x) where x=(x1,x2,…,xd)∈R^d and x∈X

The algorithm divides the population into male and female agents, simulating interactions and updating positions based on computed vibrations, enhancing the feature selection process for clustering high-dimensional data.

Fitness Function

The fitness function F(x)for a candidate solution x is calculated based on the differences in recency, frequency, and monetary values:

$$F(x)=\sum_{i=1}^{N}[(recency_i-x_0)^2+(frequency_i-x_1)^2+(monetary_i-x_2)^2]$$

where N is the number of data points, and x0, x1, and x2 are the dimensions representing recency, frequency, and monetary values, respectively.

Population Initialization

The population is initialised with P candidate solutions, each represented as a position vector in the search space:

population={x1,x2,…,xP}

where $xi \in R^d$

Weight Calculation

The weight for each spider is computed based on its fitness value:

$$Wi = \frac{fitness(i)-worst}{best-worst}$$

where ***fitness(i)*** is the fitness value of the i-th spider, and best and worst are the best and worst fitness values in the population, respectively.

Vibration Calculation

The vibration $v_{i,j}$ from spider j to spider i is calculated as:

$$v_{i,j} = w_j e^{-d_{i,j}^2}$$

where wj is the weight of the jj-th spider, and di,j is the distance between spiders i and j.

Position Update for Female Spiders

The position of each female spider fi updated based on the received vibrations:

$$\mathbf{f}_i^{t+1} = \begin{cases} \mathbf{f}_i^t + \alpha v_{i,n}(\mathbf{s}_n - \mathbf{f}_i^t) + \beta v_{i,b}(\mathbf{s}_b - \mathbf{f}_i^t) + \gamma(r - 0.5) & \text{with probability } P_f \\ \mathbf{f}_i^t + \alpha v_{i,n}(\mathbf{s}_n - \mathbf{f}_i^t) - \beta v_{i,b}(\mathbf{s}_b - \mathbf{f}_i^t) + \gamma(r - 0.5) & \text{with probability } 1 - P_f \end{cases}$$

where α,β,γ are random numbers in the range 0 to 1,Sn is the nearest spider with the highest weight, and sb is the best spider.

Position Update for Male Spiders

Male spiders are divided into dominant groups (DG) and non-dominant groups (NDG). The position of each male spider mi is updated as follows:

For $\mathbf{m}_i \in \mathrm{DG}$:

$$\mathbf{m}_i^{t+1} = \mathbf{m}_i^t + \alpha v_{i,f}(\mathbf{s}_f - \mathbf{m}_i^t) + \gamma(r - 0.5)$$

For $\mathbf{m}_i \in \mathrm{NDG}$:

$$\mathbf{m}_i^{t+1} = \mathbf{m}_i^t + \alpha \left(\frac{\sum_{l \in \mathrm{NDG}} \mathbf{m}_l w_l}{\sum_{l \in \mathrm{NDG}} w_l} - \mathbf{m}_i^t \right)$$

where $\mathbf{s}_f$ is the nearest female spider.

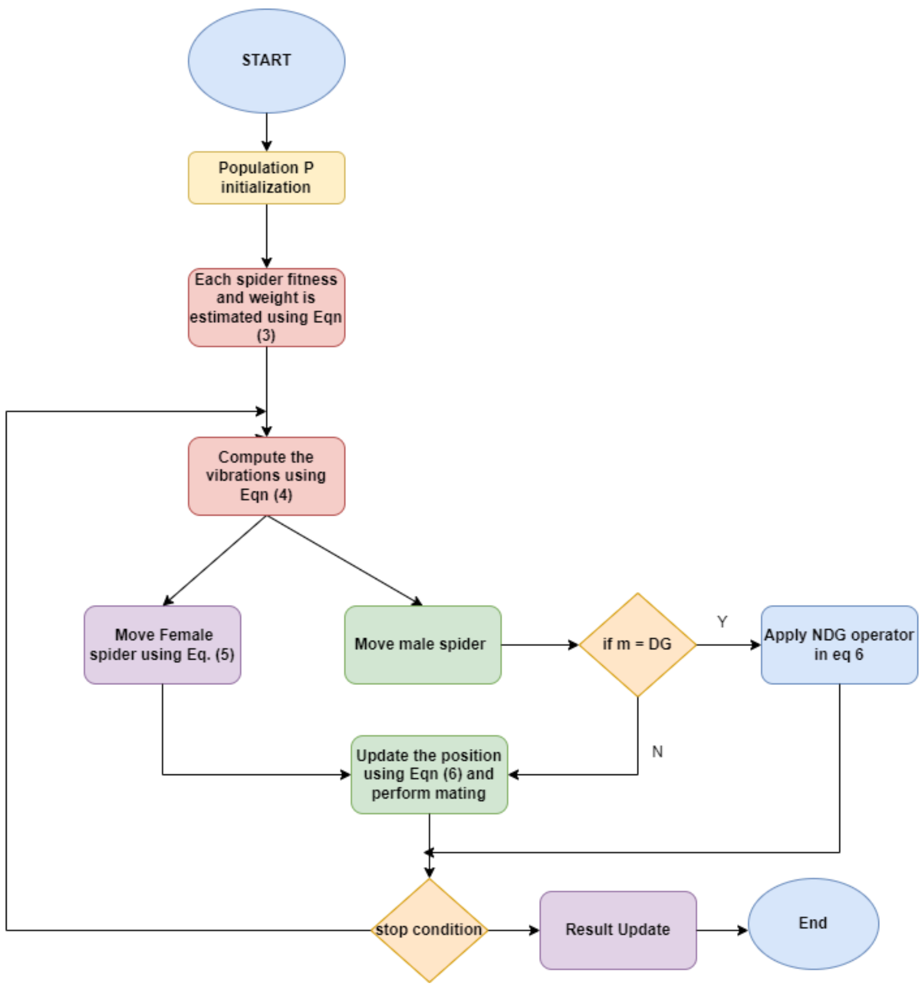


Fig: modified Social Spider Optimization (MSSO)

Mating Process

The mating process generates new spiders Snew. If the fitness of snew is better than that of other spiders, it replaces the worst spider in the population. Otherwise, snew is discarded.

Algorithm Execution

The MSSO algorithm iteratively updates the positions of the spiders and computes their fitness until the stopping condition (number of iterations) is met.

3.2 Clustering Using Self-Organizing Neural Network (SONN)

SONN is an unsupervised neural network utilising competitive learning instead of error correction. It transforms high-dimensional data into lower-dimensional representations while preserving topological properties. The network consists of an input layer connected to a competitive (output) layer. The competitive layer uses neighbourhood and feedback mechanisms to adjust neuron weights based on input data, identifying the best matching unit through Euclidean distance calculations.

The training process involves initialising weights, computing distances, and updating weights for each neuron iteratively until convergence. This unsupervised learning method facilitates effective clustering of customer data.

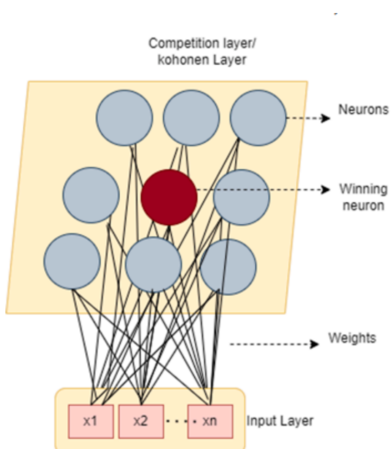


Fig. 5. Structure of SONN.

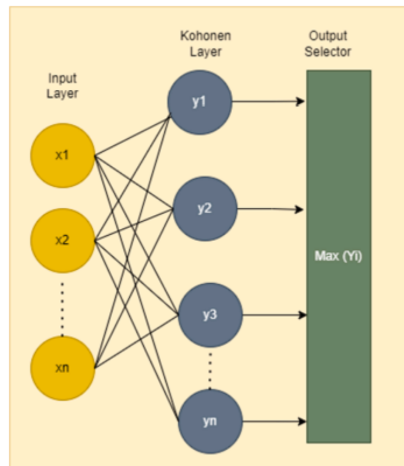


Fig. 6. Neural network structure of SONN.

3.3 Classification Using Deep Neural Networks (DNN)

The clustered data from SONN is classified using a DNN, which includes an input layer, multiple hidden layers, and an output layer. The DNN parameters are optimised to classify customers based on their purchasing patterns. The ReLU activation function is employed in hidden layers, while the softmax function is used in the output layer to segment customers. The classification process involves minimising a binary cross-entropy loss function with fairness constraints to ensure accurate and fair customer segmentation.

```
Input Layer (Size: input_size)
|
v
Hidden Layer 1: Linear (Size: input_size -> 64)
|
v
Activation: ReLU
|
v
Hidden Layer 2: Linear (Size: 64 -> 32)
|
v
Activation: ReLU
|
v
Output Layer: Linear (Size: 32 -> n_clusters)
```

4. Results and Conclusion

The proposed clustering and classification of customer segmentation for digital marketing were evaluated using the IHCL customer segmentation dataset obtained from their Instagram posts. This dataset consists of various engagement metrics detailing the interaction patterns of customers with IHCL's social media content. The dataset includes multiple features such as likes, comments, shares, post type, and time of engagement.The objective of this study was to develop a model to group customers based on their engagement patterns and identify those who are more responsive to digital marketing efforts. The dataset was processed using Python programming, leveraging libraries such as torch, for clustering, and Matplotlib, Numpy, and pandas for visualisation and data handling. The Jupyter notebook environment was used for coding, visualisation, and workflow management.

Feature Selection

The features in the dataset were selected using the SI-based feature selection model. Among the multiple features, the most relevant ones were chosen to improve the performance of the system by reducing complexity. The original and selected features are listed below:

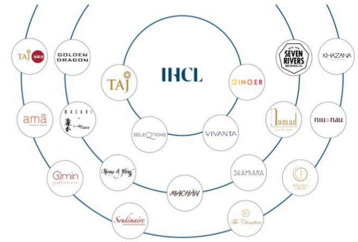
Feature Name	***Description***	***Data type***
`caption`	Description of the post	Nominal
`ownerFullName`	Name of the owner	Nominal
`ownerUsername`	Username in instagram of the owner	Nominal
`commentsCount`	Number of comments on a post	Numeric
`firstComment`	First comment in the post	Nominal
`likesCount`	Number of likes a post received	Numeric
`timestamp`	Time when a user last commented	Numeric
`queryTag`	Queries tag used in the data	Numeric
`hashtags/0-30`	Hashtags used in the posts	Nominal

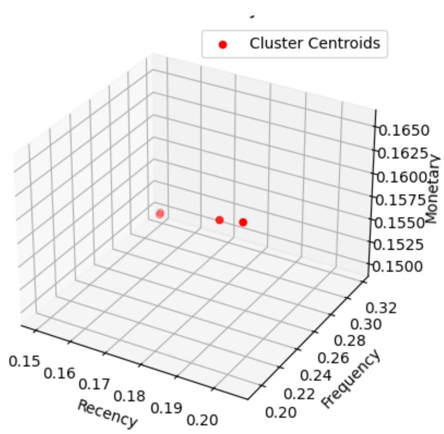
Feature Selected for MSSO:

Feature Name	***Description***	***Data type***
`ownerFullName`	Name of the owner after doing EDA	Nominal
`commentsCount`	Number of comments on a post	Numeric
`likesCount`	Number of likes a post received	Numeric
`timestamp`	Time when a user last commented	Numeric

Business Impacts:

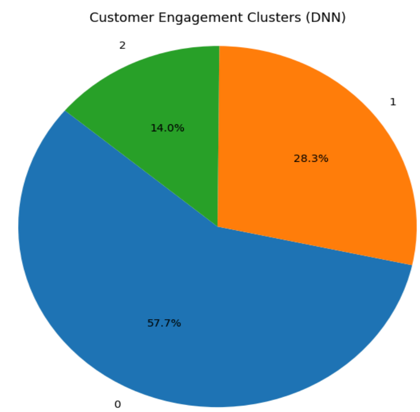
The Indian Hotels Company Limited (IHCL) is one of Asia's largest and most iconic hospitality companies. Established in 1903 by Jamsetji Tata, the founder of the Tata Group, IHCL has a rich heritage and is known for its luxury hotel brand, Taj Hotels. With over a century of experience in the hospitality industry, IHCL operates a diverse portfolio of hotels, resorts, jungle safaris, and palaces, spanning various brands such as Taj, Vivanta, SeleQtions, and Ginger. The company is recognized for its commitment to providing exceptional service, luxurious experiences, and sustainable practices, making it a leader in the global hospitality market. IHCL's extensive network includes properties in India and internationally, offering guests a blend of traditional Indian hospitality and modern luxury.The IHCL brand has many hotels and restaurants who manage their instagram pages.you can see the picture of brands and we have found the engagement level of customer in their insta pages with an efficient combination of algorithms using RFM, SONN and DNN .

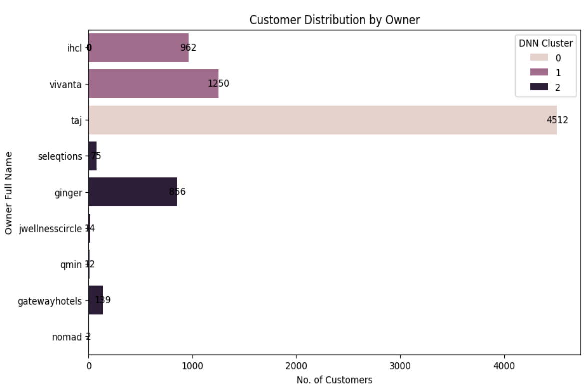




1. Improvement in Targeted Marketing:

The precise customer segmentation achieved through our developed algorithms enables IHCL to tailor its marketing efforts to specific customer groups which can be seen in the diagram that three cluster centroids are formed. A segment of customers (which is referred as 0), who frequently engage with service posts can be targeted with loyalty programme offerings or exclusive promotions, And the customers with medium engagement (which is referred as 1) with different promotional services which will make them engage for longer periods of time, while less engaged segments might receive re-engagement campaigns designed to boost their interaction levels. The use of DNNs allows IHCL to predict future customer behaviours, such as the likelihood of responding to specific types of content or offers. The pie chart shown on the left helps to distribute the customers into clusters on the basis of their interactions (likes, comments). This segmentation helps in the campaigns and loyalty programs that will be given to the customers.





2. Increased Efficiency and Profitability:

The Customer Distribution by Owner bar plot on the left illustrates the number of customers segmented by their engagement levels within different clusters. The plot reveals that Taj has the highest number of customers, along with the highest engagement on their page. In contrast, IHCL and Vivanta, which show moderate engagement, and Ginger, with lower engagement, have a similar number of customers. This insight enables IHCL to tailor its marketing campaigns more effectively, focusing on more responsive customers to enhance customer retention and maximise marketing impact.

The integration of our developed unsupervised algorithms based customer segmentation and engagement analysis into IHCL's marketing strategy has the potential to significantly elevate the company's business impact. By utilising advanced algorithms like MSSO, SONN, and DNN, IHCL can achieve a deeper understanding of its customers, enabling the company to execute highly targeted and personalised marketing campaigns. This precision in targeting not only improves customer engagement but also drives business growth by increasing conversion rates and enhancing customer loyalty. The business impact of this approach is profound. IHCL can expect to see a marked improvement in marketing efficiency, with a higher return on investment as marketing efforts become more focused on the most promising customer segments. Moreover, the predictive capabilities of the model allow IHCL to stay ahead of customer trends, ensuring that the company remains agile and responsive to changing customer needs.In summary, the deployment of this AI-driven customer segmentation model is a strategic move that positions IHCL for sustained success in the competitive hospitality market. By refining how the company understands and interacts with its customers, IHCL is better equipped to deliver personalised experiences that resonate with its audience, ultimately leading to increased profitability and long-term business growth.