2025 International Conference on Computational Robotics, Testing and Engineering Evaluation (ICCRTEE) | 979-8-3315-9709-2/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICCRTEE64519.2025.11052920

Comparative Analysis of Traditional and Deep Learning Based Customer Segmentation Models in Online Retail

Hari Govind A M
Department of Mathematics Amrita School of Physical Sciences
Amrita Vishwa Vidyapeetham Coimbatore, India
cb.ps.p2dls23006@cb.students.amrita.edu

Dr. S. Subburaj
Department of Mathematics Amrita School of Physical Sciences
Amrita Vishwa Vidyapeetham
Coimbatore, India
S_subburaj@cb.amrita.edu

***Abstract*—Customer segmentation is the process by which customers are divided into different groups based on shared features, behaviors, or needs. This assists companies in optimizing resource allocation which enhances customer experience and customizing marketing tactics. This study uses a transactional data set of an online retail company to compare multiple customer segmentation models. It can help us determine how traditional customer segmentation models like RFM can fare against deep learning based models like LSTM. The clustering methods used here are K-Means and Hierarchical Clustering. RFM is a technique based on Recency, Frequency and Monetary features and their scores. This along with clustering is used to make customer segmentation models. LSTM generates feature representations from sequential data after which it is used for segmentation using clustering. Using each of the two clustering methods on the output from both LSTM and RFM, four customer segmentation models can be obtained. Clustering results are compared using Silhouette score, Davies–Bouldin index and Calinski-Harabasz index. It is found that the LSTM based clustering techniques outperform RFM based techniques.**

***Index Terms*—customer segmentation, RFM analysis, hierarchical clustering, LSTM, K-Means Clustering**

I. Introduction

The division of customers into discrete groups according to specific attributes is called customer segmentation. Behavioral tendencies, spending patterns and demographic data are among the traits that all the segments have in common. By optimizing resource utilization, segmentation enables businesses to better engage customers and customize their marketing strategies. Depending on the needs of a particular customer group, businesses should adjust their marketing strategies. In order to achieve this, businesses analyses large datasets and uncover hidden patterns using techniques like machine learning and data mining.

Customers can be separated into groups according to common traits and requirements with the use of clustering. Clustering techniques are essential for finding underlying patterns in big datasets which enables companies to classify clients according to common characteristics and needs. This strategy encourages greater customer loyalty and makes it easier to create individualized experiences and targeted marketing campaigns. These techniques are especially helpful in industries like financial services and internet retail where understanding customer behavior is essential to obtaining a competitive edge [1].

Behavioral customer segmentation using RFM analysis (Recency Frequency and Monetary value) is one of the most popular and conventional methods. RFM analysis takes into account customers spending amounts (monetary value) frequency of purchases (frequency) and recency of their most recent purchase (recentness). When taken as a whole these three metrics provide a quantitative overview of customer activity and total value. Due to its simplicity, interpretability and efficacy in measuring key customer segments, including loyal, high-value and churn-risk customers RFM methodology is the most widely used approach. Using RFM variables in conjunction with clustering algorithms like k-means or hierarchical clustering can help define distinct customer segments that guide strategic decision-making cite [2].

The interpretability of traditional techniques like k-means clustering and RFM analysis makes them popular but they frequently fail to capture intricate nonlinear relationships in sequential behavioral data (e.g., purchase histories or browsing pat-terns). Sequential data processing and long-term dependency learning are made possible by a specific RNN architecture which is known as Long Short-Term Memory (LSTM). The use of LSTMs in customer analytics is growing in popularity as a way to model sequential behavioral data like past purchases or patterns of digital engagement over time. This paper provides a comparative study of conventional segmentation methods (RFM-based segmentation) and deep learning models (LSTM networks) for customer segmentation in online retail.

The main aim of this study is to compare the various traditional customer segmentation models with deep-learning based models. This study adds to academic research and business practice in customer analytics. To academics, it is a comparative framework to assess static and dynamic segmentation models. To business practitioners, it is insightful guidance on how various approaches to segmentation can be utilized based on data availability, business objectives, and computational power.

The results of this paper can inform e-commerce and retail companies how the method of modeling that results in more valuable and actionable customer segments is better. It can inform crucial marketing, customer retention, and customization decisions. As online purchasing continues to expand, these analytical competencies are becoming increasingly essential for remaining competitive.

II. Literature Review

Modern technological advancements especially in the field of big data and its uses are widely accepted across a range of industries in an attempt to manage vast quantities of information. Advances in technology may favor business stakeholders in their decision-making. Research to evaluate the performance of a mix of several big data analytical processes for analyzing interconnected customer information has been carried out. A hybrid model consisting of RFM (Recency Frequency Monetary) model with linked Bloom filters, K-means clustering and the Naive Bayes algorithm to enhance client segmentation and marketing strategies has been suggested. The RFM model is used to score customers according to certain criteria. K- means was used to segment outcome from the RFM model. The inactive customers were stored and can be retrieved using Bloom filters and is classified using Naive Baye's algorithm [3].

The traditional mass marketing principles have given way to more specialized strategies targeted at particular customers as a result of the growing emphasis on customer-focused marketing particularly in e-commerce. A thorough analysis of 105 research studies published between 2000 and 2022 showed that segmentation is still a crucial part of customized targeting. Four stages in the scientific process of segmentation were identified by the research: data collection customer Customer analysis and targeting strategy representation. Regardless of the size of the dataset or the specific application context K- means clustering is the most common segmentation technique when RFM analysis and manual feature engineering are the most commonly employed forms of consumer representation. Its long-lasting popularity over the years is a testament to its reliability interpretability and adaptability [4].

Beyond marketing unsupervised learning algorithms—in particular clustering algorithms—have found widespread application in fields such as supply chain management. Researchers found that unsupervised learning techniques were able to handle the additional complexity brought on by the global logistics disruption and market volatility caused by the COVID-19 pandemic. In optimization problems such as location planning and vehicle routing clustering algorithms are popular due to their durability and simplicity. Notwithstanding more sophisticated generative models have taken longer to catch New studies reveal that widespread use of real-world data shows how advanced and practical unsupervised learning has become. These results validate the efficacy of unsupervised models particularly clustering algorithms such as K-means and hierarchical clustering as powerful knowledge discovery agents for customer and usage. Their capacity to identify patterns in Unstructured data makes them especially suitable for use with RFM models or sophisticated learning methods such as LSTM in customer segmentation issues [5].

A seminal article on big data in retail explains how using large datasets can equalize the playing field in five areas: customer, product, time, location, and channel. The article emphasizes that the value in big data is not merely because of scale, but through the aggregation of multiple sources of data, use of sophisticated statistical methods, and specific theories. It is most important in customer segmentation, where the article talks about the usefulness of Bayesian methods—such as hierarchical modeling, data augmentation, and updating—to derive valuable insights from large retail data. Predictive analytics, with backing from real-world tests, is also essential to making segmentation methods more effective. Moreover, the paper also emphasizes that theory remains essential in organizing analysis and being intelligible even with machine learning and deep learning these days. Last but not least, it highlights important ethical and privacy concerns that must be confronted in any data-driven retail practice. These concerns provide valuable context when contemplating both traditional and deep learning approaches to customer segmentation in online retailing contexts [6].

While clustering methods are useful for revealing hidden patterns in static data they may not be able to capture behavioral or temporal changes in consumer data. Recurrent neural networks (RNNs)—particularly Long ShortTerm Mem- ory (LSTM) networks which have demonstrated remarkable performance in modeling sequential data—help to mitigate this limitation. Long-term dependencies are difficult for traditional RNNs to handle due to vanishing gradients which LSTM networks successfully solve by employing gate mechanisms. They are especially well-suited to modeling time-based customer activity such as engagement trends or purchase behavior because of their architecture which allows them to maintain and use long-range connections in sequences [7].

In finance, the swift growth of consumer credit has made it necessary to create advanced credit scoring models based on large volumes of behavioral data. Researchers have increasingly turned to statistical methods and artificial intelligence (AI) methods to improve decision-making models employed by banks and financial institutions. Neural networks, or deep learning architecture in particular, are now the key constituents due to their capacity to recognize intricate patterns in consumer behavior. A study utilizes a bidirectional Long Short-Term Memory (LSTM) network to predict the likelihood that credit card holders will miss one or more consecutive payments. The model, which is based on real credit card data, generates monthly risk probabilities and is evaluated based on several metrics such as accuracy, AUC Brier score, Kolmogorov-Smirnov test, and H-measure. Calibration outcomes show that the LSTM outputs are good probabilities. Furthermore, the LSTM-based approach showed superior predictive performance compared to traditional machine learning approaches, such as logistic regression, support vector machines, random forests, and multi-layer perceptrons. The results highlight the need to use deep learning to study sequential consumer behavior, thus emphasizing its relevance in customer segmentation in online retailing [8].

III. Materials and Methods

A. Dataset

All transactions for an online retailer that took place between December first of 2010 and December ninth of 2011 are included in the transactional data set used here. Unique gifts for every occasion are the company's primary product. Wholesalers are some of their major customers.

B. RFM

RFM (Recency Frequency and Monetary) is a customer segmentation technique that has been used traditionally in marketing and business analysis. It allows businesses to profit from information about consumer behavior based on past purchasing patterns. It is very helpful for customer retention, targeted marketing campaigns and customer relationship management (CRM). It divides up its customers based on spending patterns, frequency of purchases and recent behavior. Thus, businesses can identify the high-value customers, predict their future demand and thus optimise their marketing strategies.

The elements of RFM analysis are as follows. Recency (R) measures the time period since the consumer's last purchase. Consumers who have recently purchased are more likely to return than consumers whose purchases were previously made. Frequency (F) measures the number of times a customer buys within a given time interval. Active customers are more frequent and more profitable for a company. Monetary Value (M) shows how much a client has spent overall during a certain time period. Consumers with more spending tend to be more valuable to the company.

Collecting transaction information like customer ID, purchase date, total spend and buying frequency is the initial step in RFM. Every Recency, Frequency and Monetary value of customer are scored. Organizations assign scores on a scale on a scale of 1 to 5. They are then utilized to segment the client base into various categories. Some common customer segments are at-risk shoppers (low recency moderate-to-high historical spend), customers of high frequency and moderate value high-value clients (high recency high frequency and high economic value).

C. LSTM

One of the recurrent neural network (RNN) variants specially developed to avoid vanishing gradient problem—a key limitation of basic RNNs that hinders learning long-term relations in sequential data—is the Long-Short-Term Memory (LSTM) network. Due to their capability to learn temporal dependencies and retain information in long sequences LSTMs are applied in numerous applications including speech recognition time series prediction natural language processing (NLP) and anomaly detection.

The three most important gating mechanisms of the LSTM network, apart from memory cells, are the input gate, the forget gate, and the output gate. The input gate determines how much new information is added to the cell state. What parts of the cell state are to be erased is regulated by the forget gate. What information is passed to the next time step is regulated by the output gate.

These gates enable long-range dependencies of sequential data to be learned and preserved by LSTMs in a computationally effective manner. Therefore, they work best with time series data in applications like machine translation, financial market analysis and consumer behavior prediction.

Variants like bidirectional LSTMs and attention-based LSTMs significantly improve performance by processing sequences bidirectionally or by using contextual information from nearby time steps. LSTMs, because of their flexibility and robustness, are still a basic architecture for deep learning tasks dealing with sequential data.

D. K-Means Clustering

K-Means is an unsupervised machine learning algorithm used to segment the data into specified number (k) of segments. The algorithm's simplicity and low computational demands make it especially well-adapted to applications such as consumer segmentation and image compression. The algorithm tries to reduce intra-cluster variance so that the data points within a cluster are similar. At the same time, time, it tries to optimize inter-cluster variance to make the clusters distinctive. The K-Means algorithm operates by using iterative process. The initial step, or the initialization, is the random selection of k initial centroids from the data. The second step is the allocation of each individual data point to the closest centroid with a pre-specified distance metric, typically Euclidean distance, hence creating k clusters. The third stage is the re-calibration of the centroid of every cluster as the mean of all the data points assigned to it. Steps two and three are iterated iteratively until the centroids are converged (i.e., no more updates are constructed) or the highest number of iterations is reached.

E. Hierarchical Clustering

Hierarchical clustering is a machine learning unsupervised algorithm that groups data points that are similar without any prespecified number of clusters. It generates a tree-like result known as a dendrogram, which may be cut off at a chosen level to obtain the final clusters. Often used in social network analysis and market segmentation, hierarchical clustering is divided into two: Agglomerative (bottom-up), where single data points are merged into clusters, and Divisive (top-down), where a single cluster is split again and again. The procedure begins with the calculation of a distance matrix via metrics like Euclidean, Manhattan, or Cosine distance. Linkage measures—like Ward's measure, which tries to reduce variance—set the computation of distances between clusters and govern the process of how clusters merge.

F. Clustering Evaluation Metrics

Evaluating the quality of clustering results is a crucial step in unsupervised learning. Since clustering algorithms do not use labels, internal evaluation metrics are used to determine how well data points are grouped. The three internal metrics used for evaluating clustering performance are discussed be- low. The compactness and separation of clusters are measured differently by each of these metrics.

- Silhouette Score:The Silhouette Score measures the level of similarity between a data point and its assigned cluster [9]. The values obtained are between -1 and 1. The higher the score, the better the data point is grouped in its cluster. Scores close to 1 mean a low similarity with adjacent clusters and a high similarity with its own cluster. If the score is close to -1, the data point has been incorrectly assigned to a wrong cluster, and if it is close to 0, it is on the edge of two clusters.
- Davies-Bouldin Index (DBI): It calculates the average similarity of every cluster to the cluster most similar to it [10]. Similarity is distance between points in the cluster divided by distance between the clusters. Lower is better in clustering. Low DBI means well-separated and dense clusters and high DBI means overlapping or scattered clusters.
- Calinski-Harabasz Index (CHI): The Calinski-Harabasz Index, sometimes referred to as the Variance Ratio Criterion, calculates the ratio of between-cluster to within-cluster dispersion [11].

IV. Implementation

Customer segmentation is a key principle in data-driven marketing strategies that allow companies to customize products and with involvement based on behavioral patterns. This research uses five varied models of segmentation:

A. RFM with K-Means

Three of the most critical customer-oriented mea- sures—recency, frequency, and monetary value—were computed to create RFM scores for each customer. Re- cency refers to the number of days since a customer's last purchase. The frequency measure is the number of times a customer has purchased. Monetary value refers to the rolling sum of a customer's purchases. Each customer's recency measure was compared against a reference date, which was one day after the last transaction date in the data. Then, the grouping of customer-level data in the RFM table was completed. In order to determine the best k value for cluster formation, the Elbow Method was adopted, which consists of plotting the sum of squared errors (SSE) against k values between 1 and 10. The optimal k value, which was found to be 2, was at the elbow point marked as the point at which the SSE curve began its flattening trend. After which, the K-Means algorithm was applied to the normalized RFM data, for which k = 2 was found to be the best parameter. As a result, the customer base was successfully partitioned into two segments by assigning each customer with a corresponding cluster label based on spending behavior.

B. RFM with Hierarchical Clustering

For RFM with Hierarchical Clustering model, RFM data was normalized and then applied to Agglomerative Clustering with Ward Linkage. The hierarchical clustering method was utilized because it is easy to interpret and can uncover esoteric groups.

C. LSTM with K-Means

In the case of models based on LSTM, the following approach was taken to create time-series data suitable for sequence learning. The time gap between every transaction and the customer's last transaction was calculated to calculate days since last purchase. To calculate purchase frequency, invoices of every customer were counted. To provide an equal input scale appropriate for the LSTM model, these features along with quantity and unit price were scaled with MinMaxScaler. The data set was sorted according to the invoice date and grouped by customer ID for further application of the LSTM model. For customers with a reasonable transaction history, ten con- secutive transaction sequences were extracted. This setup facilitated the formation of an LSTM-based autoencoder. After training, the latent representations, or the hidden activations, of every customer sequence were captured using the encoder part of the LSTM. These embeddings were pooled using mean pooling to generate a fixed vector representing every customer. K-Means clustering was then performed to examine these grouped embed- dings. The Elbow Method indicated that two clusters reflected the optimal number. Therefore, the customers were segmented accordingly.

D. LSTM with Hierarchical Clustering

After training the LSTM model like before and extracting the hidden activations, each customer's hidden activation was aggregated using mean pooling. Agglomerative Clustering was used to group customers according to these LSTM-derived features. The ability of this hierarchical clustering technique to uncover nested structures in the data made it the preferred choice particularly in situations where the number of clusters is unclear. Agglomerative clustering with Ward Linkage is the algorithm used. The three metrics are used for evaluation. Clusters are visualized using PCA.

V. Results and Discussion

A. Comparison using clustering evaluation metrics

The experimental findings in Table I and Table II show a major difference in the performance of RFM-based clustering algorithms and those using LSTM. For all the evaluation measures, all the models using the LSTM model show better performance compared to their counterparts. Specifically, the combination of LSTM with K-Means, with k=2, shows well separated and well-distinguishable clusters, with the maximum Silhouette Score of 0.68 and the minimum Davies-Bouldin Index of 0.47. Further, the combination of LSTM with hierarchical clustering at k=7 shows the maximum Calinski-Harabasz score of 13485.24, which indicates maximum separation of clusters. RFM-based methods, however, show poor performance, with greater Davies-Bouldin indices and lesser Silhouette Scores, indicating poor compactness and separation of the clusters. The findings show that LSTM-based features, when used with either K-Means or hierarchical clustering, significantly outperform conventional RFM-based methods in this application.

TABLE I
EVALUATION METRICS FOR RFM AND LSTM-BASED CLUSTERING MODELS

Model	k	Silhouette Score	Davies-Bouldin Score
RFM + K-Means	2	0.42	0.92
	3	0.40	0.93
	4	0.36	0.95
LSTM + K-Means	2	0.68	0.47
	3	0.64	0.55
	4	0.63	0.58
RFM + Hierarchical	3	0.35	0.97
LSTM + Hierarchical	7	0.60	0.55

TABLE II
CALINSKI-HARABASZ SCORES FOR ALL MODELS

Model	k	Calinski-Harabasz Score
RFM + K-Means	2	4351.55
	3	4286.77
	4	4120.81
LSTM + K-Means	2	11354.52
	3	12362.82
	4	12934.78
RFM + Hierarchical	3	3400.09
LSTM + Hierarchical	7	13485.24

B. PCA Visualization of Clusters

As shown in Figure 1, four customer segmentation models were evaluated with PCA-based visualizations. RFM with K-Means produces two well-separated clusters with high visual clarity, making it interpretable and easy to implement. However, the simplicity of the segmentation may limit behavioral in-sights. RFM with Hierarchical Clustering offers a more sophisticated approach towards segmentation, resulting in three well-distinguishable clusters in the PCA visualization, hence a better balance between interpretability and richness of behavioral analysis is obtained. LSTM with K-Means clusters demonstrates some overlap in the PCA visualization; however, the model demonstrates good internal cohesion, indicating good insight into temporal patterns in customer behavior. The LSTM with Hierarchical Clustering demonstrates at least seven strongly separated clusters with good visual clarity, facilitating interpretation and implementation. This is because it captures the temperol patterns learned by LSTM better.

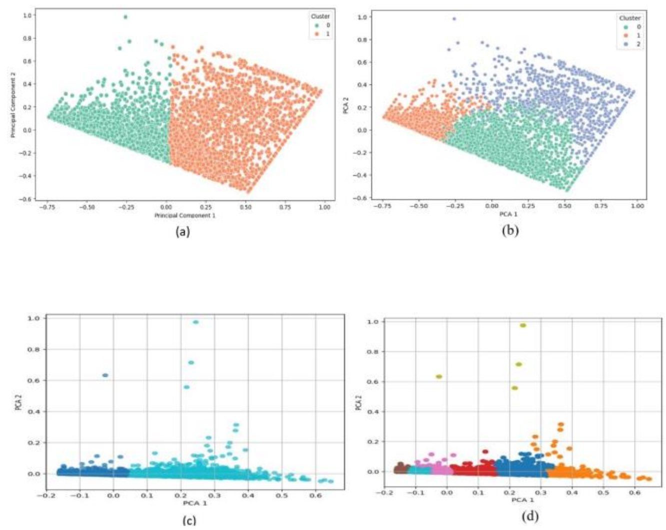

Fig. 1. PCA Visualization of Clusters for: (a) RFM with K-Means (b) RFM with Hierarchical Clustering (c) LSTM with K-Means (d) LSTM with Hierarchical Clustering

C. Time-Based Customer Spend Behavior Analysis

From Figure 2 it is seen that LSTM with K-Means clustering produced two distinct clusters. The first cluster indicates clients with consistently lower and stable mean spend across the year and the second cluster indicates a distinct rising trend peaking in late quarter (in the area of September) and falling marginally towards the end of the year. There is a distinct separation between low-value stable customers and high-value customers whose spend increases over time perhaps as a result of changing buying habits or seasonal sales. Richer segmentation was obtained using LSTM combined with Hierarchical Clustering, which resulted in the formation of seven clusters. Some clusters, such as Cluster 1, have a consistent increase in spending, while others demonstrate either inconsistent or chronically low spending behaviors. For example, Clusters 4 and 6 demonstrate higher month to-month variation, while Cluster 5 demonstrates moderate and consistent spending behavior. Hierarchical methodology seems to capture a variety of temporal behaviors that contain both regular and irregular spending patterns. Overall, although LSTM combined with K-Means produces a simpler and more interpretable segmentation that is useful for focus on macro-marketing strategies, LSTM combined with Hierarchical Clustering offers more in depth insight into the behavior of the customers over time and hence is more appropriate for custom and behavior-specific interventions.

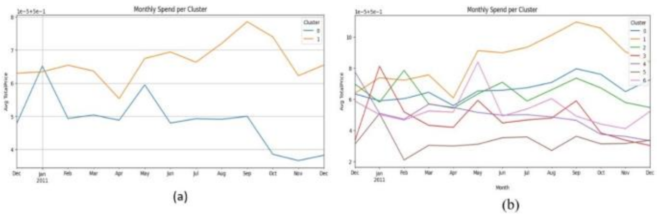

Fig. 2. Time-Based Customer Spend Behavior Analysis for (a) LSTM with K-Means (b) LSTM with Hierarchical Clustering

VI. CONCLUSION

The study indicates that in terms of cluster separation and temporal comprehension, LSTM-based clustering techniques outperform traditional RFM techniques. LSTM with K-Means provides a simpler understandable segmentation suitable for mass marketing cam-paigns while LSTM with hierarchical clustering recognizes nuanced customer behavior suitable for customized actions. Although RFM models are clearly visible in PCA, they are shallower in behavior analysis. The findings favor LSTM based clustering for consumer segmentation and hierarchical techniques offer detailed information for time-sensitive and dynamic approaches.

REFERENCES

[1] P. A. Sarvari, A. Ustundag, and H. Takci, "Performance evaluation of different customer segmentation approaches based on rfm and demographics analysis," *Kybernetes*, vol. 45, no. 7, pp. 1129–1157, 2016.

[2] A. J. Christy, A. Umamakeswari, L. Priyatharsini, and A. Neyaa, "Rfm ranking–an effective approach to customer segmentation," *Journal of King Saud University-Computer and Information Sci- ences*, vol. 33, no. 10, pp. 1251–1257, 2021.

[3] S.-C. Wang, Y.-T. Tsai, and Y.-S. Ciou, "A hybrid big data analytical approach for analyzing customer patterns through an integrated supply chain network," *Journal of Industrial Informa- tion Integration*, vol. 20, p. 100177, 2020.

[4] M. Alves Gomes and T. Meisen, "A review on customer segmen- tation methods for personalized customer targeting in e-commerce use cases," *Information Systems and e-Business Management*, vol. 21, no. 3, pp. 527–570, 2023.

[5] B. Rolf, A. Beier, I. Jackson, M. Mu¨ller, T. Reggelin, H. Stuck- enschmidt, and S. Lang, "A review on unsupervised learning algorithms and applications in supply chain management," *In- ternational Journal of Production Research*, vol. 63, no. 5, pp. 1933–1983, 2025.

[6] E. T. Bradlow, M. Gangwar, P. Kopalle, and S. Voleti, "The role of big data and predictive analytics in retailing," *Journal of retailing*, vol. 93, no. 1, pp. 79–95, 2017.

[7] Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: Lstm cells and network architectures," *Neural computation*, vol. 31, no. 7, pp. 1235–1270, 2019.

[8] M. Ala'raj, M. F. Abbod, and M. Majdalawieh, "Modelling customers credit card behaviour using bidirectional lstm neural networks," *Journal of Big Data*, vol. 8, no. 1, p. 69, 2021.

[9] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics*, vol. 20, pp. 53–65, 1987.

[10] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE transactions on pattern analysis and machine intelligence*, no. 2, pp. 224–227, 2009.

[11] T. Calin´ski and J. Harabasz, "A dendrite method for cluster anal- ysis," *Communications in Statistics-theory and Methods*, vol. 3, no. 1, pp. 1–27, 1974.

[12] C. Rungruang, P. Riyapan, A. Intarasit, K. Chuarkham, and J. Muangprathub, "Rfm model customer segmentation based on hierarchical approach using fca," *Expert Systems with Applica- tions*, vol. 237, p. 121449, 2024.

[13] P. Fra¨nti and S. Sieranoja, "K-means properties on six cluster- ing benchmark datasets," *URL http://cs. uef. fi/sipu/datasets*, pp. 4743–4759, 2018.

[14] Z. Sun, L. Sun, and K. Strang, "Big data analytics services for en- hancing business intelligence," *Journal of Computer Information Systems*, vol. 58, no. 2, pp. 162–169, 2018.

[15] S. F. Ahmed, M. S. B. Alam, M. Hassan, M. R. Rozbu, T. Ishtiak, N. Rafa, M. Mofijur, A. Shawkat Ali, and A. H. Gandomi, "Deep learning modelling techniques: current progress, applications, ad- vantages, and challenges," *Artificial Intelligence Review*, vol. 56, no. 11, pp. 13 521–13 617, 2023.

[16] G. Sakaline and L. Buics, "Advancing towards sustainable retail supply chains: Ai-driven consumer segmentation in superstores," *Engineering Proceedings*, vol. 79, no. 1, p. 73, 2024.

[17] C. Wang, C. Rao, F. Hu, X. Xiao, and M. Goh, "Risk assessment of customer churn in telco using fclcnn-lstm model," *Expert Systems with Applications*, vol. 248, p. 123352, 2024.

[18] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.