An E-Satisfaction Analysis using machine learning evaluates customer behaviors and profiles to measure and improve digital satisfaction

Mr. Chetan Kumar
Computer Science & Engineering, Regional College for Education Research & Technology, Sitapura Jaipur, India kumarchetan@deepshikha.org

Abstract Driven by the convergence of the digital economy and artificial intelligence, this study provides a comprehensive evaluation of consumer behavioral profiling and segmentation architectures, specifically using K-Means, modified K-Means, and K-Prototype clustering algorithms to optimize electronic satisfaction (e-satisfaction). The research offers critical insights into modeling consumer behavior, loyalty, and satisfaction within automated systems, proposing enhancements to elevate market performance. Keywords— effective elements, loyalty, e-satisfaction, machine learning, customer segmentation, e-commerce

INTRODUCTION

In the contemporary digital landscape, evaluating consumer dynamics and establishing data-driven customer segmentation via Machine Learning (ML) paradigms have emerged as pivotal strategies for optimizing electronic satisfaction (E-satisfaction). Marketing literature has long emphasized the significance of ML-driven computational frameworks within electronic marketplaces, particularly in capturing the co-creation of value between enterprises and consumers. This focus has propelled researchers to look beyond traditional assessments and implement advanced Data Mining (DM) methodologies to decode consumer feedback while accounting for the holistic value-driven experiences of the user.Consequently, behavioral profiling and algorithmic segmentation elucidate the distinct personas comprising a company's clientele. By simulating diverse purchasing scenarios, organizations can accurately gauge consumer mindsets and prospective responses, even to entirely novel stimuli. To capitalize on these insights, modern enterprises embed robust behavioral analytics, specialized partitioning models, and optimized ML clustering techniques into their operational architectures to attract and retain consumers.

As illustrated above in Fig. 1, the macro-level architecture categorizes three core user archetypes. At the processing level, an efficient segmentation routine divides these individuals into mutually exclusive clusters to satisfy distinct mathematical requirements.This model relies on an expansive multi-parameter matrix to analyze key Influential Effective Elements

(EEs), which include:Geographic EEs: Regional and spatial consumer distributions.Demographic EEs: Age, gender, and socio-economic attributes.Psychographic EEs: Lifestyle, values, and psychological profiles.Behavioral EEs: Purchasing habits, brand interactions, and engagement levels.Firmographic EEs: Organizational metrics for B2B sub-segments.These EEs systematically group consumers based on quantified characteristics and statistical attributes. Concurrently, behavioral analysis utilizes the foundational Recency, Frequency, and Monetary (RFM) metrics as primary benchmarks for profiling. Within this context, this paper presents a comparative analysis of customer segmentation strategies optimized for E-satisfaction by evaluating K-Means, modified K-Means, and K-Prototype clustering algorithms.

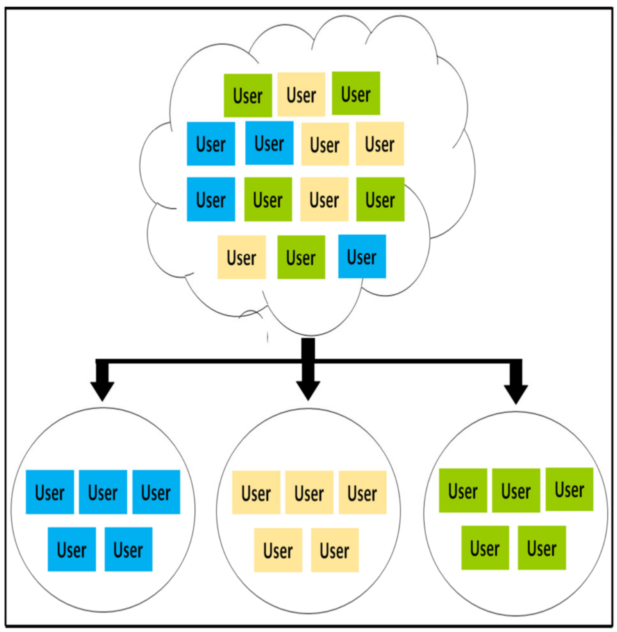

Fig. 1. Basic approach of customer segmentation

CONCEPTUAL PARADIGM: AN ILLUSTRATION OF SEGMENTATION, E-SATISFACTION, AND K-PROTOTYPE

The pervasive daily reliance on digital marketplaces for application deployment, service consumption, and product procurement yields massive, high-velocity datasets. Modern enterprises actively preserve and analyze these data repositories to optimize customer acquisition strategies, maximize sales volumes, and bolster brand equity.A. Foundations of Customer SegmentationAlgorithmic customer segmentation entails partitioning a heterogeneous consumer base into distinct, homogeneous cohorts using multi-dimensional feature vectors. This methodology serves as a cornerstone for data-driven targeted marketing. The feature architectures governing customer-centric profiling and feedback-driven categorization are outlined below.1) Structural Attributes of Customer-Centric SegmentationCustomer profiling operates across two distinct operational dimensions: the consumer persona and the account lifecycle (Fig. 2). Behaviorally, a shopper may manifest as loyal, volatile, newly acquired, historical, infrequent, or highly active. From an administrative standpoint, the enterprise secures critical transactional metadata, including user profile credentials, historical order registries, and customized promotional incentives.

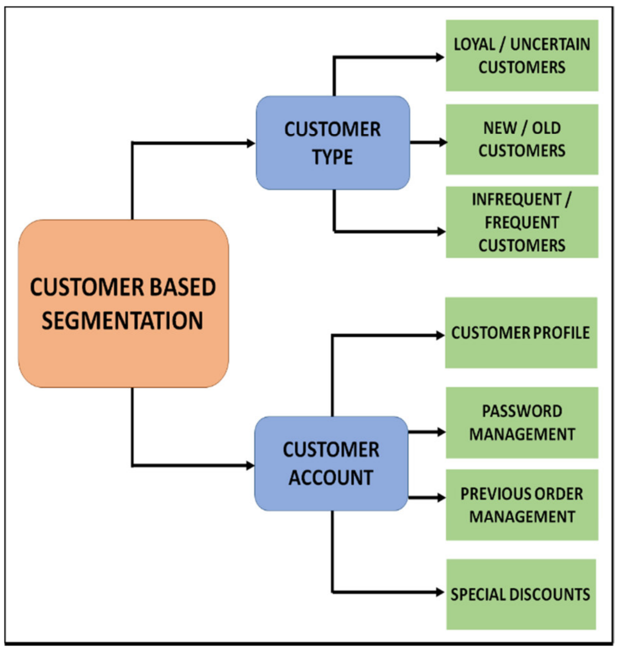

Fig. 2. Feature structuring in customer-based segmentation

1) Features of customer reviews: Fig. 3 illustrates the feature structuring of the seg-mentation perspective based on customer reviews and feedback. It covers the feed-back type, recommendation system, and rating system. The form of the feedback can be simple suggestions or detailed informative ones. These reviews and sugges-tions can be either positive or negative or neutral. The recommendation system helps the customers and users in selecting the products and services. A customer can provide a rating between 1 and 5 for different products and services offered.

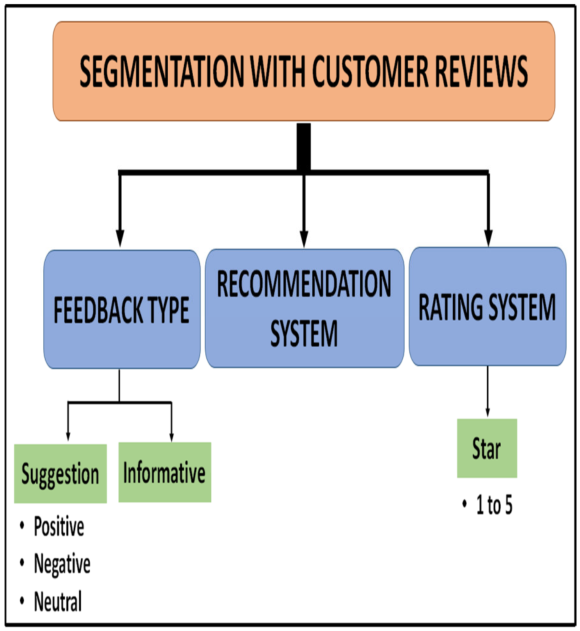

Fig. 3. Feature structuring in review-based segmentation

B. *K-Prototype Clustering Technique*

A K-Prototype clustering technique is an improvement over the K-Means which processes the mixed types in datasets. This partitioning algorithm integrates the means of the numerical part of the K-Means algorithm with the mode of the categorical part. Then it constructs a new hybrid cluster center called 'Prototype' for the K-Prototype algorithm. Fig. 4 shows two numeral feature values with a single categorical variable.

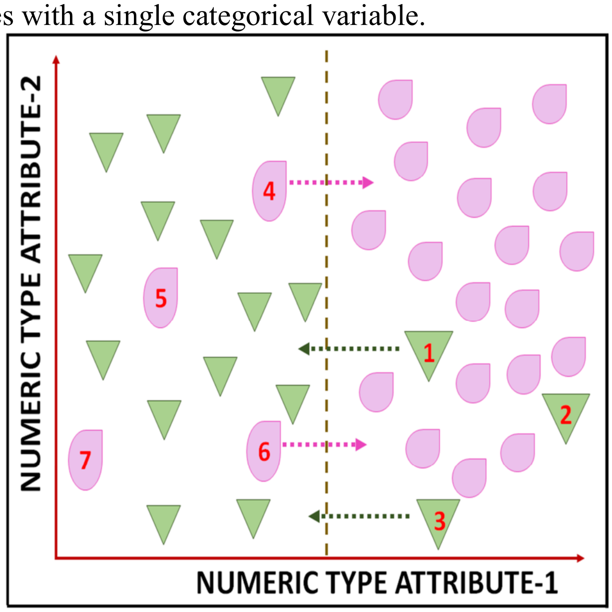

Fig. 4. K-Prototype clustering process

C. *Why Target E-Satisfaction?*

Customer segmentation and analysis define strategic marketing that helps in controlling the expenditure within the budget. In India, its typical market has different ex-pectations and demands of the customers. Such a market depends on various factors to target the customer with greater customization. So, the primary concerns of online enterprises are clients' happiness, satisfaction, and the extraction of behavior patterns.

THEMATIC BACKGROUND

This section explains several existing customer segmentation and E-satisfaction approaches and compares them based on some criteria. Further, it highlights the limitations and gaps of such existing works. The customer segmentation procedure [1] first gathered and analyzed the data and then grouped it using three clustering techniques. It assessed the system using the Sum Square Error method (SSE). Using Naive Bayes (NB), Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), K Nearest Neighbors (KNN), Classification And Regression Tree (CART) in Decision Tree (DT), Logistic Regression (LR), Multi-Layer Perceptron (MLP), and ensemble-based techniques, the next approach [2] compared and investigated many churn prediction algorithms. Ada Boosting Trees (ABT), Random Forest (RF), and Stochastic Gradient Boosting (SGB) were among the ensemble approaches used.

The virtual authenticated network [3] employed the modified E-satisfaction strategy. It used Google forms to collect data from questionnaires and collaborated with Lyceum-Northwestern University's (L-NU) Internet services. Non-Negative Trigonometric Sums (NNTSs), a circular statistical-based method based on the customer's temporal patterns, was provided by the customer segmentation algorithm [4]. It calculated the total Average Silhouettes Weights (ASW) and the ASWs of all groupings for appropriate parameter combinations. By gathering the customer relationship management's SSE knowledge, the approach [5] examined customer behavior for destructive characteristics. The data was gathered and cleaned before the features were chosen. It was then classified and finally assessed using knowledge extraction.

Table 1 illustrates the comparison among several existing research contributions using the criteria of the concerned problem, products and services considered, and performance results.

TABLE I. DEPICTING CONCERNED PROBLEM, PRODUCTS/SERVICES CONSIDERED, AND RESULTS OF EXISTING CONTRIBUTIONS

Ref. No.	Concerned Problem	Year	Products/Services Considered	Performance Results
Behavioral Aspect				
[1]	K-means clustering for customer segmentation	2018	Shopping items from local retail shops.	The ASW of learning approaches ranged from 0.52 to 0.56. Cluster 3 had an SSE difference of 2.763.
[2]	A comparative assessment of ML strategies for customer retention	2018	International consumer service calls in the telecommunication industry.	With ABT, MLP, SVM, DT, NB, LR, and LDA, accuracy was 96%, 94%, 94%, 90%, 88%, 86.7 %, and 86.7%, respectively.
[3]	Virtualized AAA framework to enhance E-satisfaction in LNU	2019	University student database from different courses and years.	Good performance results.
[4]	Segmentation of customers depending on their usage patterns	2020	SMS, voice data, and Foursquare dataset as well as a Global-scale check-in dataset.	SMS data revealed ASW= 0.7246 for cluster-I with 19 people and ASW= 0.7506 for cluster-II with 17 people. With FourSquare, ASW = 0.842 was discovered and ASW = 0.7008, 0.9274, and 0.8332 for clusters 1, 2, and 3.
[5]	Modified K-means to find E-satisfaction	2020	Products from Hamkaran System.	Achieved SSE = 0.3908.
[6]	Customer contribution's past, present, and future	2020	Retail Market database and poll of leading scholars.	-
[7]	Supply chain demand forecasting using predictive techniques and operations	2020	-	-
Demographic Aspect				
[8]	Consumer segmentation focused on RFM	2019	Customer database.	Clusters II to VIII exhibited ASW values ranging from 0.6005 to 0.721. SSE difference ranged from 0.5218 to 10.1316, and SSE ranged from 2.7809 to 10.1316.
Firmographic Aspect				
[9]	Consumer segmentation strategy with buyer	2018	Online grocery shopping bills.	Good performance results.

Ref. No.	Concerned Problem	Year	Products/Services Considered	Performance Results
	targeting to improve profits			
[10]	RFM model to analyze customer's buying behavior	2020	Real & synthetic databases. Customer purchase transactions from the UCI repository.	In Synthetic DB, ASW Score was found to be 0.362159752 for K = 3 and 0.3490755342 for K = 5. In Real datasets, ASW values of 0.362052482 for K = 3 and 0.381312431 for K = 5 were got.
Firmographic and Behavioral Aspects				
[11]	Consumer review methodology with DM	2018	South Indian restaurant database and Facebook.	Good performance results.
[12]	Consumer segmentation analysis for buying behavior using RFM	2020	Customer transaction database from online website.	The overall sales grew by 279%, while the total amount utilization increased by 101.97%.
Firmographic, Demographic and Geographic Aspects				
[13]	Methodology for product bundling and direct marketing	2018	Customer database from online website.	Achieved the confidence range for ARM from 78.41% to 89.45% in scenario 1, and from 90% to 91.13% in scenario 8.
Firmographic, Demographic and Behavioral Aspects				
[14]	RFM analysis based on clustering for customer behavior	2018	Online customer transaction database.	Good performance results.
[15]	Customer behavior analysis with DM	2018	Products from food service & retail stores & industries.	Achieved average expert verification result as 64% for ARA-1 and 88% for ARA-2.

The review [6] highlighted the origin, present, and prospective buying patterns, as well as retail and existing market research. It illustrated technical improvements, customer value variations, and consumption observations. Another study [7] examined the applications, techniques, and research prospects for supply chain demand prediction using big data analytics (BDA). It deliberated how these algorithms were classified and how they were used in time-series forecast, segmentation, KNN, SVM, Artificial Neural Network (ANN), support vector regression, and regression analysis. It concluded that relatively little study had been done into BDA and reverse logistics. The customer segmentation method [8] followed the cross-industry standard process for DM and included the steps of business knowledge, data gathering, data normalization, feature analysis, testing, and Silhouette index determination.

The recommender [9] was designed for the customers while also maximizing profits for industries. It used the dendrogram, Akaike and Bayes criterion, Elbow method, Renyi and Shannon entropies to create the best data clusters. The K-means algorithm was then used to cluster the data using Principal Component Analysis (PCA) to determine the highest variance from recency and frequency. Singular Value Decomposition (SVD) and cosine similarity were also used to recommend to customers. Another method [10] involved assessing the data, pre-processing it, performing RFM analysis, and segmenting the customers. It then used ASW to assess the clusters and re-viewed the customers by measuring RFM sales parameters from one cluster to the next.

Another method [11] gathered data, pre-processed it, used PCA to normalize the RFM scores, assessed the weights, and categorized customers into four categories. The next method [12] employed the WEKA tool to evaluate consumer comments from Facebook to make decisions, improve product quality, and streamline processes. This analysis pre-processed the data, assessed the features, and finally categorized it. Data comprehension, data integration and data cleansing, standardization, and modeling were all steps in the proposed technique [13]. The market was then split and customers were clustered based on loyalty. Following that, it determined product packs, predicted the best product pack frequency, carried out direct marketing, and statistically analyzed the results.

The next segmentation method [14] included data gathering, normalization, data pre-processing, feature selection, and K-means clustering. An improved method [15] used K-Means to cluster purchases of products by customer type before using the Elbow method to identify the suitable clusters. Then, it used the Apriori algorithm to evaluate buying behavior and established the ARM Analysis (ARA-1) and ARA-2. Finally, the accuracy of buying product patterns was assessed.

Various contributors and existing works have implemented their E-satisfaction and customer behavioral systems with the K-Means and with their modified and improved versions. These systems have deployed using K-Means clustering [8] [9] [11] [13] where K was 2 [14], 3 [4] [10], 4 [12], and 5 [1] [10] [15].

A few systems employed the improved K-means clustering technique [5]. Some other systems implemented the learning techniques [3] [6] [7] such as the Agglomerative and Meanshift techniques [1]; the LDA, CART, KNN (K = 7), SVM, LR, RF (500 DTs), ABT, SGB, NB, and MLP techniques [2]; the DBSCAN [5], SVD [9], and J48 [11] techniques; the Kohonen Self-Organizing Maps (SOM), Apriori and SVM techniques [13]; and the ARM & Apriori techniques [15].

All these existing systems had faced many challenges, gaps, limitations, and critical issues as shown in Fig. 5. These gaps must be filled by improving the system's performance and extending the system with many other ML techniques, real-world applications, parameters, and datasets. Some other suggested extensions require time-efficiency and a modified velocity process for mass data.

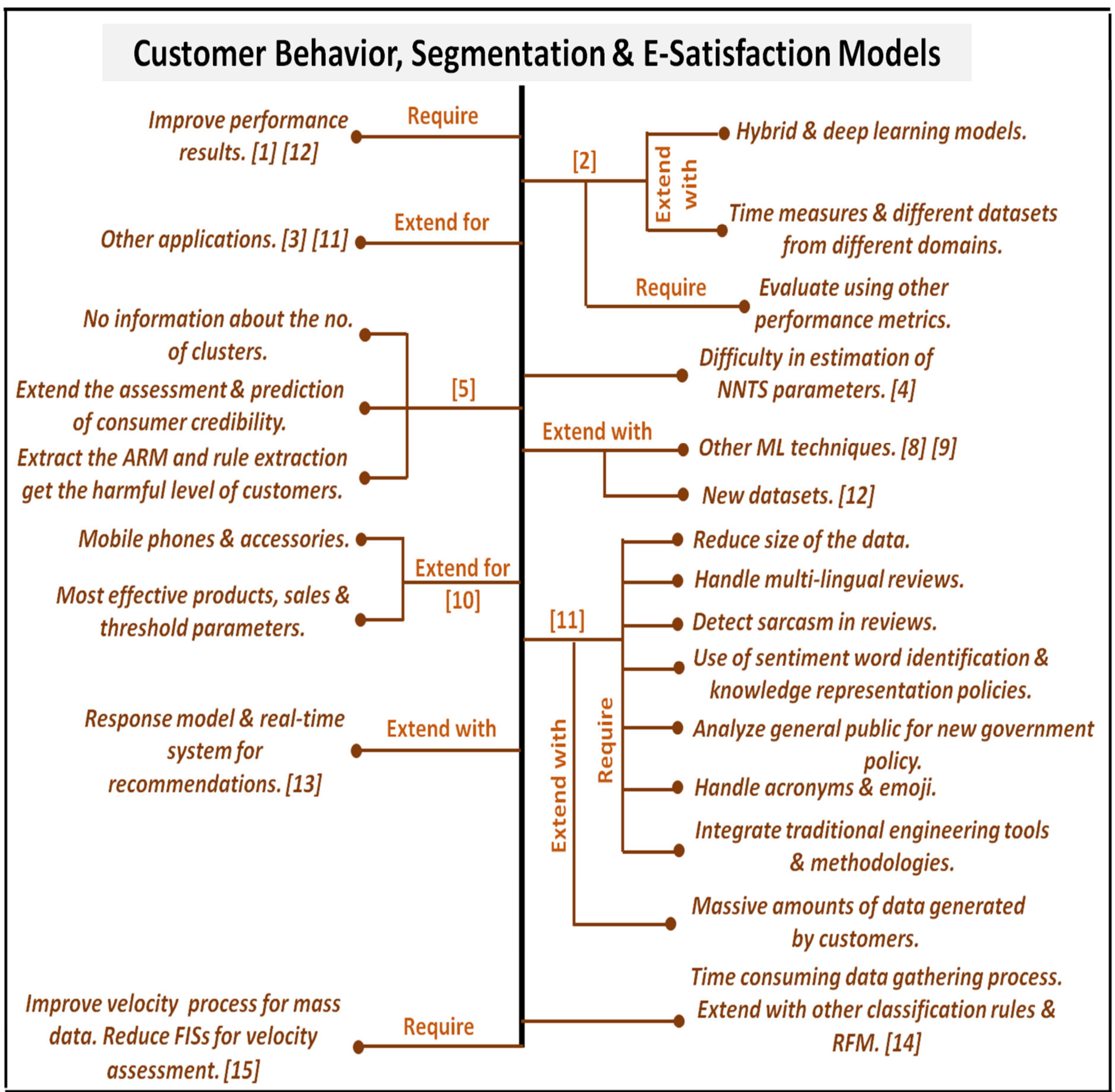

Fig. 5. Depicting gaps, needs, and extensions of various existing systems

Analytical Results in E-Satisfaction

Fig. 6 and Fig. 7 graphically show the analytical and comparative results of various data types and clustering techniques used in the existing contributions, respectively. They both represent the graphs with the % of types and techniques used by the re-search contributors. It is observed that the maximum wok, i. e. 50% was done on the shopping items and grocery dataset. The K-Means clustering methodology, on the other hand, became the favourite among all others, with a contribution of 66.67%.

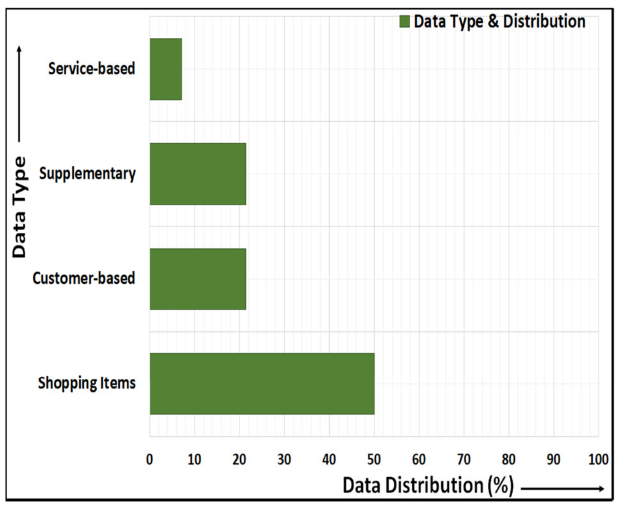

Fig. 6. Data types and their distributions in various existing contributions

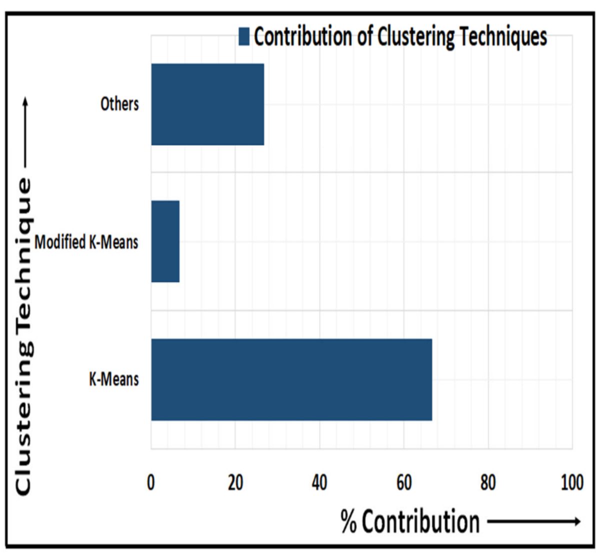

Fig. 7. Clustering techniques and their % distributions in various existing contributions

CONCLUSIONS

This paper provided a study on customer behavior analysis and segmentation to achieve a high level of E-satisfaction. Many research works were compared with each other based on some key criteria and their performance was distinguished from each other. Subsequently, their gaps, limitations, and critical issues were identified. This shows that such research works should be more focused on the usage of multiple ML techniques in different real-world application areas, and should work upon the strong K-Means and K-prototype algorithms to achieve efficient results.

ACKNOWLEDGMENT

We would like to thank all those who have helped us directly or indirectly to complete this work.

REFERENCES

[1] Kansal, T., Bahuguna, S., Singh, V., Choudhury, T.: Customer segmentation using K-means clustering. In: International Conference on Computational Techniques, Electronics and Me-chanical Systems, pp. 135-139. IEEE Press (2018).

[2] Sabbeh, S. F.: Machine-learning techniques for customer retention: a comparative study. In-ternational Journal of Advanced Computer Science and Applications 9(2), 263-281, 2018.

[3] Berba, E. M., Palaoag, T. D.: Improving customer satisfaction on internet services in L-NU using virtualized AAA network architecture. In: Proceedings of the 8th International Confer-ence on Informatics, Environment, Energy and Applications, pp. 178-183. ACM, NY, USA (2019).

[4] Fernández-Durán, J. J., Gregorio-Domínguez, M. M.: Consumer segmentation based on use patterns. Journal of Classification 38, 72-88 (2020).

[5] Zare, H., Emadi, S.: Determination of customer satisfaction using improved K-means algo-rithm. Soft Computing 24, 16947–16965 (2020).

[6] Malter, M. S., Holbrook, M. B., Kahn, B. E., Parker, J. R., Lehmann, D. R.: The past, pre-sent, and future of consumer research. Marketing Letters 31, 137–149 (2020).

[7] Seyedan, M., Mafakheri, F.: Predictive big data analytics for supply chain demand forecast-ing: methods, applications, and research opportunities. Journal of Big Data 7(1), 1-22 (2020).

[8] Dedi, Dzulhaq, M. I., Sari, K. W., Ramdhan, S., Tullah, R., Sutarman: Customer segmenta-tion based on RFM value using K-Means algorithm. In: Fourth International Conference on Informatics and Computing, pp. 1-7. IEEE Press (2019).

[9] Bhade, K., Gulalkari, V., Harwani, N., Dhage, S. N.: A systematic approach to customer segmentation and buyer targeting for profit maximization. In: 9th International Conference on Computing, Communication and Networking Technologies, pp. 1-6. IEEE Press (2018).

[10] Anitha, P., Patil, M. M.: RFM model for customer purchase behavior using K-means algo-rithm. Journal of King Saud University-Computer and Information Sciences, 1-8 (2020).

[11] Varudharajulu, A. K., Ma, Y.: Feature-based restaurant customer reviews process model us-ing data mining. In: Proceedings of the 2018 International Conference on Computing and Big Data, pp. 32–37. ACM, NY, USA (2018).

[12] Wu, J., Shi, L., Lin, W-P., Tsai, S-B., Li, Y., Yang, L., Xu, G.: An empirical study on cus-tomer segmentation by purchase behaviors using a RFM model and K-means algorithm. Mathematical Problems in Engineering, 1-7 (2020).

[13] Beheshtian-Ardakani, A., Fathian, M., Gholamian, M.: A novel model for product bundling and direct marketing in e-commerce based on market segmentation. Decision Science Letters 7(1), 39-54 (2018).

[14] Sheshasaayee, A., Logeshwari, L.: Implementation of clustering technique based RFM analysis for customer behaviour in online transactions. In: 2nd International Conference on Trends in Electronics and Informatics, pp. 1166-1170. IEEE Press (2018).

[15] Ponyiam, P., Arch-int, S.: Customer behavior analysis using data mining techniques. In: In-ternational Seminar on Application for Technology of Information and Communication, pp. 549-554. IEEE Press (2018).

*****