SEGMENT COMPASS – A FRAMEWORK FOR ACTIONABLE CUSTOMER INSIGHTS

[1]Sk. Sameerunnisa,[2]B. Nithish Kumar, [3]Ch. Sriram, [4]D. Bhargava,[5]L. Mahesh
[1]Assistant Professor, [2]Student, [3]Student, [4]Student, [5]Student
[1]Department of Computer Science and Engineering (IOT, Cyber Security including Blockchain Technology)
[1]Vasireddy Venkatadri International Technological University, Guntur, India

Abstract : Customer behavior analysis is a significant venture to any business, as it tends to do to any business which is oriented to customer involvement, deal with customer retention and enhance profitability. Nonetheless, the majority of the available customer segmentation strategies are based on the simplistic models of customer segmentation, including demographic customer segmentation or classical RFM analysis, and the similarity being that they do not consider modelling customer behavior sufficiently. Moreover, these models usually cannot be transparent and the business users do not find it easy to believe and read the result. In the given paper, the author introduces a customer segmentation system, Segment Compass which is built on a supplemented LRFMS model (also Length of relationship and Satisfaction) rather than Recency, Frequency, and Money value. Its solution uses the Gaussian Mixture Models (GMM) model to automatically detect natural trend in customer behavior and random Forest classifier to categorize the customer as a valuable business group (Bronze, Silver, Gold and Platinum). The assumptions also introduce SHAP-based explainability that aims to print out the contribution of each attribute related to the customers towards the tier assignments so that the output can be readable. Compared to conventional customer segmentation solutions, Segment Compass will support updates made on events of the customer, i.e. metrics and tiers of customers can be updated based on a new transaction event. This architecture has been constructed using Python machine learning packages, modular back-end, and administrator, and customer interactive dashboards. The outcomes of the experiments confirm that the suggested procedure can adequately provide adequate customer groups, adequate forecast of levels, and definite hints which can be utilized directly to the business choice. In conclusion, the explainable machine learning proves to be useful to customer analytics as illustrated in Segment Compass.

***IndexTerms* - Customer Segmentation, LRFMS, Explainable AI, Random Forest, Gaussian Mixture Model.**

I. INTRODUCTION

I.1. Background and Motivation

The customer segmentation is the crucial component of the information based marketing and customer relationship management as it enables the organizations to understand the customer behavior, augment the effectiveness of the engagement strategy and improve the profitability in the long run. The traditional segmentation schemes have largely relied on a demographic aspect or a predetermined behavioral predictor which in the majority of instances do not reflect into real customer worth and vibrant trends of interaction in transactional data collections [1], [2].

An important aspect of data driven marketing is customer segmentation, as it allows the organization to understand customer behavior and develop specific engagement strategies. Conventional, demographic centered forms of segmentation do not tend to reflect real customer worth and changing purchasing pattern especially when dealing with large transaction volumes [11], [13]. Recent research underlines the significance of behavioral segmenting by the use of transactional properties aimed at enhancing customer perspective and marketing success [5].

Essentially, within the set of behavioral methods, the RFM (Recency, Frequency, Monetary) model is a rather common use of the simplicity and the success that it had in profiling purchasing behavior [1], [3]. It has been established in a number of studies that RFM based segmentation is a superior source of insight than demographic grouping per se but unfortunately this type of segmentation is fairly narrow as it does not take on board the duration of customer relations or customer satisfaction which are vital measures of loyalty and churn threats [2], [4]. These limitations make the use of fixed models of RFM less effective as interaction with the customers develops and becomes more dynamic.

To address these limitations, there is recent study that examined machine learning-based customer segmentation, specifically, unsupervised clustering methods, specifically, K Means, Bisecting K Means, and Gaussian Mixture Models (GMM) [3], [6]. Such methods are capable of finding concealed patterns of behavior in large volumes of data, and in many cases are more effective in asphyxiating rule based segmentation techniques. Nevertheless, pure clustering only have difficulties in practice deployment whereby systems are not able to readily categorize new customers and do not tend to be directly oriented toward business oriented customer layers [4], [7].

I.2. Problem Statement

Although customer segmentation is widely used in business analytics, most of the existing systems are based on demographic characteristics or fixed models of behavior which do not reflect the actual customer value and changing patterns of engagement [1], [2]. Even behavioral methods, like RFM, that are widely used are mainly concerned with transaction activity and fail to effectively represent long term relationships or customer satisfaction so they are inappropriate in the current, data driven environment [1], [3].

Although the customer analytic segmentation tools have improved, a lot of currently implemented segmentation tools are based on fixed behavioral or demographic profiles that fail to keep pace with the changing customer behavior [10], [11].

Moreover, although machine learning-based segmentation systems are better pattern finding techniques, most of them are black box models that are not very interpretable and they do not give business decision making what it requires [4], [7]. Their inability to be explained, the fact that it is not scalable to new customers, and the ability to adapt dynamically leaves a gap between analytical models on one hand and real business use on the other hand. To face these challenges, cooperation in one framework that could lead to behavior driven segmentation, explainable predictions, and dynamic customer intelligence is necessary, which is sought in this work.

I.3. Objectives and Contributions

The main conclusion of this paper is to develop a useful customer segmentation model that is more thorough in capturing customer behavior compared to other conventional demographic or it is built based on RFM demographics [1], [3]. The idea behind the proposed approach is to combine behavioral, temporal as well as satisfaction based in order to enhance the reliability and utility of customer segmentation to real world business decision making [2], [4].

The main findings of the research are the adoption of RFM to become LRFMS model, the application of unsupervised clustering and supervised classification to scale to the task of customer tier prediction, as well as the addition of explainable AI methods to increase transparency and trust [12], [13]. The framework also allows dynamic customer updates, thus allowing timely and interpretable customer insights to be used in decision making.

The major contributions of the work are the extension and expansion of the traditional RFM analysis to the LRFMS model, combination of unsupervised clustering and supervised classification to customer tiering at a large scale, as well as incorporation of the Explainable AI technique to enhance model transparency and trust [6], [8], [9]. The system also promotes customer updates that are dynamic and business aligned tier governance, filling in the gap between well advanced machine learning capabilities and customer intelligence that can be acted upon.

II. LITERATURE SURVEY

The paper provides the strategic development of the Segment Compass project, which makes real- time data integration an essential part of the competitive landscape of contemporary customer analytics. Overcoming the natural constraints of conventional RFM models, this study indicates how streaming data and explainable AI (XAI) provide better ROI and more practical behavioural data [1].

This research is of great relevance to us as it directly shows the answer to the question, which is, what is the best tool to be used? We already are using K-Means, however, as demonstrated in this paper, GMM is a much better algorithm. It provides us with the immediate step to follow clearly: we need to test GMM in our project of a segment compass. It is a big technical advancement that would assist us present a lot more dependable and accurate insights to our customers [2].

This study confirms the tier based framework of the Segment Compass by identifying the similarities between the personas in the paper and our own Platinum, Gold and Silver categories. Moreover, it presents robustness as one of the key dashboard metrics in order to measure the reliability across tiers so that marketers can focus more on the most stable segments. The framework offers a definite modeling pathway, which means that using Random Forest algorithms, a stable cluster can be identified and scaled like the New Shoppers to guarantee the precise targeting effect and expanded segmentation[3].

The study enhances the Segment Compass perspective by combining Between Topological Data Analysis (TDA) and the LRFMS framework with emphasis moving towards multifaceted behavior patterns in lieu of fixed measures. Through the application of cluster-based regression of all three Platinum, Gold and Silver levels, the study transforms the platform to a predictive analytical type. The ability to predict future purchase behaviors would be possible through this evolution, and it is through this that the mission of the project to provide customer insights that are highly actionable and foresight should be met.This is an impactful and rational implementation followed by the platform. The insights can be made really actionable by adding forecasting so businesses can not only identify their best customers, but also predict what they will buy next, which is directly connected to the overall mission of the project which is to provide business actionable insights[4].

This study proves the LRFMS framework and develops a methodology of incorporating a Satisfaction (S) score, which would improve the customer-centricity of current levels. Furthermore, it gives a guideline of Dynamic Analysis, which transforms the project of the static segmentation into a time-series-based tracking of the trend. The evolution allows proactive intervention, i.e., the notification of the stakeholders of dropping satisfaction in the Gold tier, thus, turning the platform into a computerized system that is prospective and may avert churn due to foresightful knowledge [5].

The framework suggests that LRFMS model should be improved by adding Relationship Duration (S) and Average Transaction Interval (T) which would greatly improve the precision of analytics. In addition, the study provides a rationale of moving away individual models towards ensemble schemes, which puts the joint configuration of leading- edge algorithms as the best trajectory of high predictive accuracy. The effective use of the Random Forest in the research confirms the accuracy of the existing technical architecture because the project is created on the basis of a solid and effective methodological background[6].

In the existing framework, a prioritization gap is identified in the given paper, which creates segments without the focus of strategies being developed. With the application of a Multi-Criteria Decision-Making (MCDM) method, the study allows ranking the degrees of tiers according to business value and adjustable parameters such as loyalty and spend. This Strategic Ranking module wraps up the analysis pipeline - Data → Clusters→ Priorities → Action - changing the platform into an effective segmentation framework into a holistic strategic engine[7].

Demonstrates how segmentation may change between fixed-dynamical real-time clusters. Recommends that Segment Compass might incorporate pipelines that can be updated continuously. Evidences the importance of automated triggers of campaigns (e.g., re- engagement offers). Solidifies the need of customer strategy responsiveness. Helps transforms our system to live actionable intelligence framework[8].

Stresses that cluster validation measures (e.g., ANOVA, Silhouette) should be used in segmentation. Recommends the extension of Segment Compass to make demand, churn or revenue cycle predictions. Evidences that segmentation applies more power when associated with predictive forecasting. Ensures the significance of making clusters business- trusted and action- oriented [9].

III. PROPOSED METHODOLOGY

Segment Compass system architecture has been developed in a modular and layered system architecture to guarantee that the architecture is scalable, accurate and allows effective processing of customer data. It combines both the unsupervised segmentation (discover patterns) and the supervised classification (predict tiers) into a single framework.

III.1. System Architecture

To be a good customer segmenter and insight generator, the modular pipeline with multiple functional layers are to be provided in our solution. The data collection layer incorporates the transaction, customer comment and behavioral measures data of different sources. The second layer is preprocessing and feature engineering where the data obtained is cleaned and the LRFMS scores of each customer is computed.

Clustering and classification algorithm are some of the machine learning layer which is used to classify customers and to define their levels. The explainability layer generates SHAP based images in order to specify how the customer features influence deciding the levels to assign to the customer. The API layer offers model predictions and explanations, however, using them in the form of RESTful endpoints and enables integration and scalability into the system. The final but not the least is the presentation layer that provides the business customers with the opportunity to explore customer groups and the actionable insights as interactive dashboards.

III.2. Data Collection

Receives transactional information (purchase dates, purchase amounts, purchase products). Collected the customer satisfaction scores using the surveys and feedback systems. Collects date of registration and demographics of the customers. Brings in many different data formats (CSV, data-base, APIS).

III.3. Preprocessing & Feature Engineering

III.3.1 Data Cleaning

Preprocessing phase involves the removal of the duplicating values and also the values that are null in order to ensure data integrity. Outliers are countered to clean up the data by Interquartile Range (IQR) algorithm after which date formats and currency values are standardized thereby analysing it in a similar way.

III.3.2 LRFMS Calculation

Length (L): Days since customer registration

1. **L = CURRENT DATE - REGISTRATION DATE**

Recency (R): Days since last purchase

2. **R = CURRENT DATE − LAST PURCHASE DATE**

Frequency (F): Total number of transactions

3. **F = COUNT OF TRANSACTIONS**

Monetary (M): Total spending

4. $M = \sum(i = 1 \text{ to } n)\text{Transaction Amount}$

Satisfaction (S): Average customer rating

5. $S = (\sum(i = 1 \text{ to } n)\ \text{Rating\_i}) / n$

Normalization:

Apply Standard Scaler to ensure features are on comparable scales.

6. $$z = (x - \mu) / \sigma$$

III.4. Machine Learning

III.4.1 Unsupervised Segmentation

Produce customer embedding based on normal LRFM signs. Monkey Gaussian Mixture Model (k= 2.6). Optimality of number of clusters calculated by use of BIC (Bayesian Information Criterion).

7. $$BIC = -2 \cdot \log(L) + k \cdot \log(n)$$

where L is likelihood, k is parameters, n is sample size

Assign customers the highest level of probability and cluster them together. Predicting At Check Silhouette Score to justify quality of clustering.

8. $$s(i) = ((b(i) - a(i)))/\max\,(a(i), b(i))$$

where a(i) is intra-cluster distance, b(i) is nearest- cluster distance.

III.4.2 Supervised Classification

The data will be divided into a 80:20 training and testing data set. The training data is then provided to a random forest classifier comprising of 100 decision trees, and it is said to be instructed on how to map the features of customers with the clusters to which these features can be associated. To optimize the hyperparameters, the optimization of model performance is performed with the help of the GridSearchCV. The trained model is evaluated against the test dataset under normal measures of classification including accuracy, precision, recall and F1-score. Finally, classifier is used to produce tier forecasts of the new customers and a probability score to predict the accuracy of the forecast.

III.5. Query Processing

In the case when a new customer requires tier assignment, there is a choice of two.

III.5.1 Batch Processing

Customer data processing is processed as a batch loading whereby records are loaded in large quantities and the LRFMS capabilities of the particular customer are calculated. This information is fed in and a trained random forest model is applied to give a collection of tier assignments. Besides the expected levels, the scores of confidence are obtained to demonstrate the strength of each of the predictions and make informed decisions.

III.5.2 Real-time Prediction

When this is in real time, the system is fed at an API interface with user input of each customer LRFMS feature input. It is followed by the instant predictions using a pre-trained model. The system returns the forecasted tier of customer as well as the associated probability distribution capable of enabling responsiveness and decision- making that is confidence-aware.

III.6. Contextual Explanation Generation

These SHAP values are computed at every prediction of the model using TreeExplainer to compute the contribution of features. The most important characteristics that are identified to define individual tier assignments are the Monetary and Frequency. Force plots are then generated that are then exploited to visualize that each of the features has in terms of shifting the prediction to a specific tier and summary plots are generated to show the overall significance of the feature in the entire sample of customers. To maximize the clarity and interpretability of the results, all explanations are strictly limited to the features of LRFMS in order to derive some consistent insights with no ambiguity.

III.7. System Diagrams

Fig.1 System Architecture

Fig.1 represents the System Architecture of our project – Segment Compass: A Framework for Actionable Customer Insights.

The Segment Compass architecture is a six-layer pipeline which is predefined to be largely modular to process a raw transaction, registration, and feedback data into actionable intelligence. Following the initial consumption, the Preprocessing Layer ensures data integrity through the aid of cleaning, normalization that refines the data to an LRFMS Feature Engineering Layer to carry out an appraisal of Length, Recency, Frequency, Monetary, and Satisfaction dimensions. These features are turned into the machine learning decision engine with the help of the Gaussian Mixture Models (GMM) with the assistance of the random forest, which further divides customers into the Bronze, silver, gold, and platinum levels. The Explainability Layer enables the SHAP values to provide clear global and individual explanations in contrast to the final Application Layer which provides such discoveries via the dashboard to provide an individual strategic guidance.

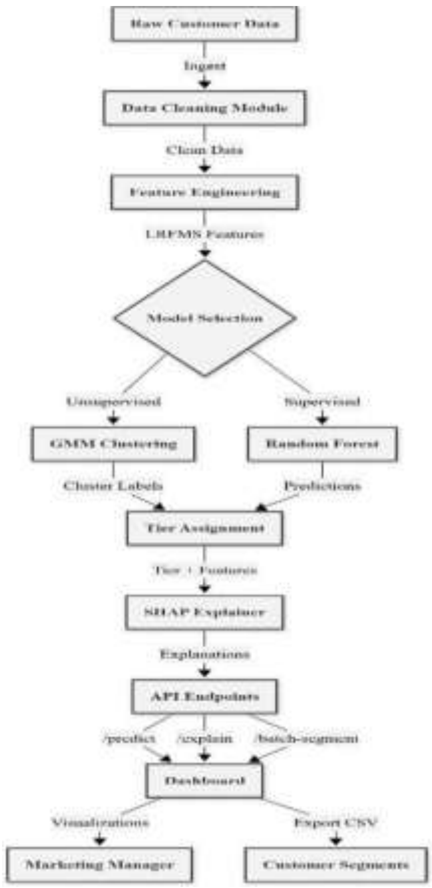


Fig.2 Framework

Fig.2 illustrates the Data Flow of our project – Segment Compass: A Framework for Actionable Customer Insights.

The data flow of the Segment Compass (Fig. 2) follows a systematic pattern with the raw transaction data, registration data, feedback data being received and processed by a Data Cleaning Module so that it can be formatted into an integrity-ensuring standard. These fine tuned inputs are sent to the Feature Engineering Layer generating LRFMS feature matrix. Behavioral segments are determined by Unsupervised GMM Clustering and Supervised Random Forest is applied to further assign Tiers to Bronze, Silver, Gold and Platinum during the Model Selection. To be transparent, SHAP Explainer then narrows these findings and then give the granular insights of behavior in form of API Endpoints to a single location of Dashboard in strategic marketing action.

III.8. SHAP Explainability Integration

To enable SHAP TreeExplainer to be model interpretable, the trained Random Forest model is being initialised to build model interpretable SHAP Trees. Each single feature to the model prediction SHAP is calculated in any case under all prediction. The approach allows providing localized explanations of the depth of individual customers and globalized information on the feature importancy score, which is easy and reliable to construct the decisions.

IV. RESULTS AND DISCUSSION

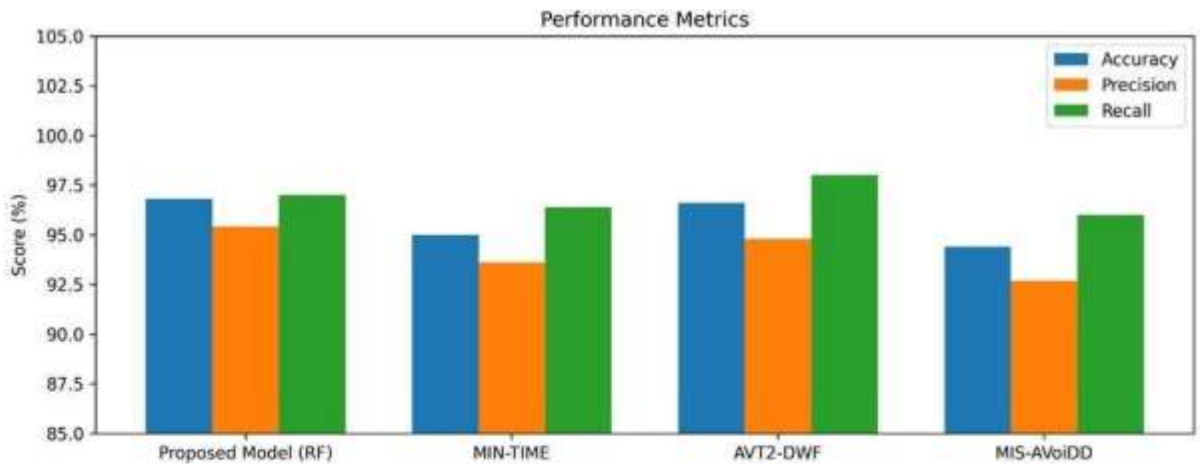


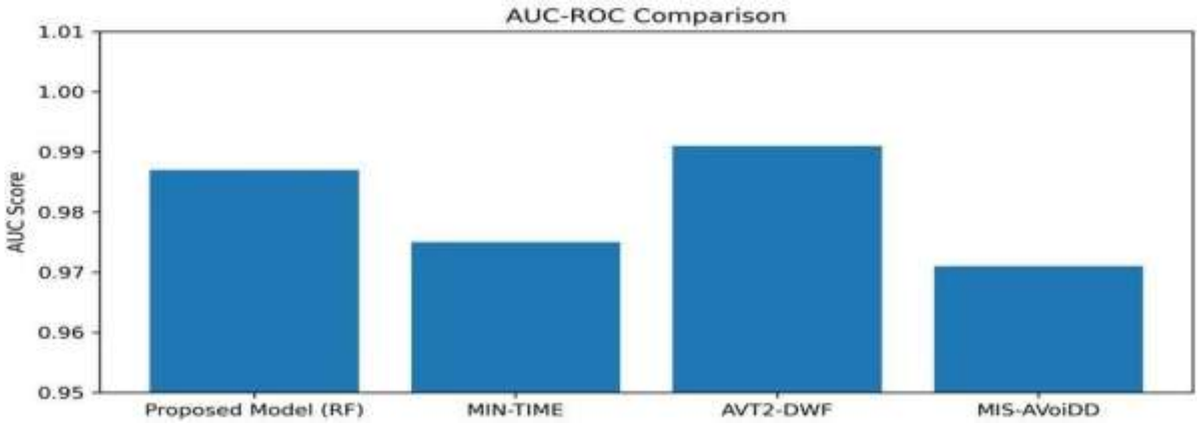
Fig 4. Performance Metrics

This value represents the outputs of the Proposed Model (Random Forest) to real methods in terms of accuracy, precision and recall. The proposed model has kept on offering better numbers in each of the three measures thereby better classification consistency and even-handed functioning. The recall element has grown and this reveals that the model is discovering the right levels of customers and more than this, precision depicts that there is less misclassification of the model as compared to the methods used in the baseline.

Fig 5. F1-Score Comparison

The following figure compares the models F1 score. With the help of the proposed approach that is created on the basis of the random forest, the highest F1 score is received and exhibits more desirable balance concerning precision and recall. It implies that predictions of customer tiers in the presented model provides more stable and robust predictions when compared to the comparative approaches.

Fig 6.AUC-ROC Comparison



Comparison of the AUC ROC shows that the proposed model would give high value of AUC than the current methods and the high figure of AUC shows greater discriminative power. The larger the AUC, the greater the model

can differentiate between the different customers levels at different levels of classification. This result justifies the viability and effectiveness of the proposed system in general.

V. CONCLUSION & FUTURE WORK

The Segment Compass framework demonstrates a robust end-to-end system by integrating advanced machine learning (GMM and Random Forest) with SHAP-powered explainability to generate actionable insights. By enhancing the traditional RFM model with relationship duration and satisfaction metrics, the LRFMS framework provides a more comprehensive and accurate assessment of customer value. This dual-track pipeline—combining unsupervised clustering with supervised classification—ensures precise tier assignment, while the modular architecture facilitates seamless deployment, scalability, and business transparency.

The system creates an opportunity to find and retain valuable customers and organizations issue personalized marketing plans, create customer satisfaction and finally make profits. It suggests competitive edge in the contemporary fast moving business environment since it incorporates advanced analytics and real-time data.

V.1. Future Directions

The next round of Segment Compass platform will be dedicated to Live streaming apache Kafka integration and extension of forecasting capacity assistance of Deep learning and Time-Series Forecasting (ARIMA/LSTM). The framework will be expanded to multi-channel information and automatic integration of the campaigns with such a platform as Mailchimp and HubSpot. Refined analytics levels will be introduced, including Causal Analysis, Federated Learning, and Cluster-wise Regression, which will spin forward additional insights, and strategic ranking Module will be introduced through MCDM based on which industries will rank businesses segments based on the strategic objective business values.

REFERENCES

[1] N. Kumar, "Intelligent customer segmentation: Unveiling consumer patterns with machine learning," Journal of Umm Al-Qura University for Engineering and Architecture, vol. 16, no. 3, pp. 774–783, 2025.

[2] J. M. John, O. Shobayo and B. Ogunleye, "An exploration of clustering algorithms for customer segmentation in the UK retail market," Analytics, vol. 2, no. 4, pp. 809–823, 2023.

[3] S. Roychowdhury, E. Alareqi and W. Li, "OPAM: Online purchasing-behavior analysis using machine
learning," in Proc. 2021 International Joint Conference on Neural Networks (IJCNN), IEEE, pp. 1–8, 2021.

[4] R. Rivera-Castro, A. Pletnev, P. Pilyugina, G. Diaz, I. Nazarov, W. Zhu and E. Burnaev, "Topology-based clusterwise regression for user segmentation and demand forecasting," in Proc. 2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA), IEEE, pp. 326–336, 2019.

[5] S. Wang, L. Sun and Y. Yu, "A dynamic customer segmentation approach by combining LRFMS and multivariate time series clustering," Scientific Reports, vol. 14, no. 1, p. 17491, 2024.

[6] L. Yang, X. Niu and J. Wu, "RF-LightGBM: A probabilistic ensemble way to predict customer repurchase behaviour in community e-commerce," arXiv preprint arXiv:2109.00724, 2021.

[7] H. A. Mahdiraji, E. Kazimieras Zavadskas, A. Kazeminia and A. Abbasi Kamardi, "Marketing strategies evaluation based on big data analysis: A clustering-MCDM approach," Economic Research-Ekonomska Istraživanja, vol. 32, no. 1, pp. 2882–2892, 2019.

[8] A. Ahsun and A. Okunola, "Leveraging real-time big data streams for dynamic customer segmentation and hyper-personalized marketing campaigns: An empirical study on e-commerce platforms," 2025.

[9] K. R. Kashwan and C. M. Velu, "Customer segmentation using clustering and data mining techniques,"
International Journal of Computer Theory and Engineering, vol. 5, no. 6, p. 856, 2013.

[10] G. N. Rodrigues, M. N. H. Mir, M. S. M. Bhuiyan, M. D. A. L. Rafi, A. M. Hoque, J. Maua and M. F. Mridha, "NLP-driven customer segmentation: A comprehensive review of methods and applications in personalized marketing," Data Science and Management, 2025.

[11] O. N. Akande, H. B. Akande, E. O. Asani and B. T. Dautare, "Customer segmentation through RFM analysis and K-means clustering: Leveraging data-driven insights for effective marketing strategy," in Proc. 2024 International Conference on Science, Engineering and Business for Driving Sustainable Development Goals (SEB4SDG), IEEE, pp. 1–8, 2024.

[12] Y. Zhang, "Machine learning-based customer segmentation: A comprehensive investigation of techniques, challenges, and applications," in Proc. 1st International Conference on E-Commerce and Artificial Intelligence (ECAI), 2024.

[13] F. A. Alijoyo, "Personalized marketing: Leveraging AI for culturally aware segmentation and targeting,"
Alexandria Engineering Journal, vol. 119, pp. 8–21, 2025.

[14] G. Wang, "Customer segmentation in digital marketing using a Q-learning-based differential evolution algorithm integrated with K-means clustering," PLoS One, vol. 20, no. 2, e0318519, 2025.

[15] I. Ekanayake and S. De Alwis, "Explainability, risk modeling, and segmentation-based customer churn analytics for personalized retention in e-commerce," arXiv preprint arXiv:2510.11604, 2025.

[16] G. Marín Díaz, "A unified fuzzy–explainable AI framework (FAS-XAI) for customer service value prediction and strategic decision-making," AI, vol. 7, no. 1, p. 3, 2025.

[17] C. P. Ufeli, M. U. Sattar, R. Hasan and S. Mahmood, "Enhancing customer segmentation through factor analysis of mixed data (FAMD)-based approach using K-means and hierarchical clustering algorithms," Information, vol. 16, no. 6, p. 441, 2025.

[18] I. Lewaaelhamd, "Customer segmentation using machine learning model: An application of RFM analysis,"
Journal of Data Science and Intelligent Systems, vol. 2, no. 1, pp. 29–36, 2024.