Advances in Consumer Research
https://acr-journal.com/
Volume-3 | Issue-1 | Jan 2026
Original Researcher Article

Analytical Study of Customer Perceptions Regarding Data Mining Techniques in Retail Purchase

Dr. Sandip Vitthal Tile[1], Dr. Santosh Prabhakar Adhav[2]
[1]Assistant Professor (MBA), Dr. Moonje Institute of Management and Computer Studies, Nashik- 422005, (MS) India.
Email ID : dr.sandiptile@gmail.com
[2]Assistant Professor MVP Samaj's Karmveer Abasaheb Alias N. M. Sonawane Arts, Commerce & Science College, Baglan (Satana), Nashik – 422301, (MS) India.
Email ID : adhavsantosh09@gmail.com

ABSTRACT
In the evolving landscape of retail, understanding customer perceptions toward data mining techniques is vital for impacting the purchase decisions. This analytical study discovers how various data mining methods, including Association Rule Mining (ARM), RFM Technique, Customer Segmentation, Market Basket Analysis (MBA), and Time Series Analysis (TSA), impacted the consumer behaviour in retail environments. After the inspecting these techniques, the research goal to discover patterns in customer preferences, such as recognizing frequently co-purchased items through ARM and MBA, classifying buyers via Customer Segmentation, evaluating loyalty with RFM, and predicting periodic trends using TSA. Data was collected from 100 retail customers through a Likert rating scale questionnaire, capturing responses on a 5-point Likert scale for perceptions of data mining techniques. The Structural Equation Modelling (SEM) was executed using AMOS software. The model fit was examine using indices such as RMSEA (0.06), CFI (0.94), and TLI (0.93), confirmatory a robust structure after minor modifications. The findings reveal significant positive influences of data mining techniques on customer perceptions and decisions.
Keywords: Data Mining Techniques, Customer Perceptions, Structural Equation Modelling, Retail Purchase Decisions...

1. INTRODUCTION:

The swift in digitalization of retail has created a crucial need to understand customer behaviour through data mining methods. The studies emphasize how association rules discover hidden buying affinities, which retailers then interpret into recommendations perceived as applicable and value-adding by consumers. (*Hilage & Kulkarni, 2011*). The customer segmentation approaches not only improve marketing efficiency but also foster customer trust when communications align with individual expectations. Complementing these approaches, hybrid models that integrate sequence mining with RFM provide richer insights into temporal purchase behaviour, strengthening loyalty programs and personalization initiatives (Chen et al., 2024). Furthermore, frameworks that merge association rule mining with time series forecasting within big-data platforms like Hadoop and Spark illustrate how real-time analytics can produce actionable insights for retail decision-making (Jain et al., 2023).

Review of literature:

Association rule mining is widely used in retail analytics to uncover hidden relationships between products in customer transactions. Hilage and Kulkarni (2011) applied the Apriori algorithm to retail data, demonstrating how mined association rules influence buying behaviour and help businesses anticipate customer needs in a shopping mall environment (Hilage & Kulkarni, 2011). The Abbas et al. (2024) evaluated clustering algorithms such as K-means, Gaussian Mixture Models, and DBSCAN applied to a UK retail dataset. (Abbas et al., 2024). Time series analysis is a powerful technique in retail, supporting accurate demand forecasting, inventory planning, and price optimization. Yun et al. (2023) developed a comparative study of multivariate time series models, integrating macroeconomic indicators with retail sales data to improve forecasting accuracy and enhance consumer trust by ensuring product availability (Yun et al., 2023). These insights confirm that reliable time series forecasting directly influences customer perceptions of reliability and service quality.

Hilage and Kulkarni (2011) demonstrated the effectiveness of MBA in identifying products often purchased together, highlighting its role in shaping consumer perceptions of convenience in product selection (Hilage & Kulkarni, 2011). Collectively, MBA enhances the shopping experience by offering relevance and reducing decision complexity, which positively affects consumer trust and satisfaction.

Relevance of study:

The present study has significant relevance in today's retail environment, where customer behaviour is gradually shaped by data-driven strategies. With the rapid development of e-commerce and omnichannel retailing, understanding how consumers perceive data mining techniques such as association rule mining, customer segmentation, time series analysis, market basket

How to cite : Dr. Sandip Vitthal Tile, Dr. Santosh Prabhakar Adhav, Analytical Study of Customer Perceptions Regarding Data Mining Techniques in Retail Purchase *Advances in Consumer Research*. 2026;3(1): 614-617

analysis, and RFM becomes vital for enhancing customer trust and loyalty. Retailers are not only leveraging these tools to forecast purchase patterns but also to deliver personalized experiences that impact decision-making and long-term engagement. Inspecting customer perceptions provides practical insights into whether such techniques are viewed as value-adding, intrusive, or neutral, thereby guiding retailers in balancing analytical efficiency with decent responsibility. This study, therefore, contributes to both academic literature and managerial practice by connecting the gap between analytical methods and the subjective experiences of customers in the current retail context.

Objectives:

To analyse the effectiveness of data mining techniques in retail purchase decisions, and evaluate they're in predicting customer behaviour.

To examine customer perceptions regarding the use of data mining techniques, particularly in relation to perceived usefulness, trust and privacy concerns.

To identify the relationship between customer perceptions of data mining techniques and their actual purchase decisions.

Research Methodology:

In the background of SEM, Quantitative Method of Research is executed. With the help of Random Sampling methodology, the consumer survey conducted to understand consumer perceptions (Samling - 100) on the methods of data mining practiced by retail stores and their influence on their purchasing experience. The response by consumers made on the 5-point Likert scale. In the present research study, the Structural Equation Modelling (SEM) techniques utilized and thereafter the data moved to AMOS and the model with latent constructs and path analyses and fit indices extracted. Five different and diverse questions were asked to consumers on every method of data mining and thereafter the questions utilized to measure the latent variables. (Perceived Usefulness, Trust) on the SEM model.

Reliability and Validity:

The application of a methodical research design and the careful selection of appropriate measurement tools guarantee the validity and reliability of the current study. In order to maintain the uniformity in replies across various consumer groups, reliability is maintained by using identical procedures for data collecting and regularly applying well-established statistical approaches, including Structural Equation Modelling (SEM). In order to confirm that the survey items and the analytical approaches precisely capture consumer perceptions of data mining techniques in the context of retail decision-making, validity is established by matching the selected analytical tools (such as AMOS) with the particular study objectives.

Parameter to Fit:

The model fitness has been assessed using the following table:

Table No.:1 Parameter to Fit

Sr. No.	Statistical test	Criteria for fit	Fitness decision
1	Chi-Square (χ^2)	p-value > 0.05	Good fit
2	RMSEA	< 0.08 <0.05	Good fit Excellent
3	CFI	< 0.90 <0.95	Good fit Excellent

The Chi-Square (χ^2) test, with a p-value greater than 0.05, validates an acceptable model fit, representing that the proposed model corresponds closely with the observed data. The Root Mean Square Error of Approximation (RMSEA) evaluates the extent of difference between the model and the data, with values below 0.08 representing a satisfactory fit and values below 0.05 reflecting an excellent fit. Similarly, the Comparative Fit Index (CFI) assesses the model's performance relative to a baseline model with no specified relationships, where values exceeding 0.90 indicate a good fit and values above 0.95 denote an outstanding fit. Based on these indices of model fitness, the SEM model was implemented by following a series of methodical steps:

Structural Equation Model:

Tabel No.: 2 Structural Equation Model

Latent Variables	Observed Variable	Mediators
Data Mining Techniques	ARM (Association Rule Mining)	Perceived Usefulness (PU)
	RFM (Recency, Frequency, Monetary)	Trust (TR)
	Customer Segmentation	Privacy Concerns (PC)
	Market Basket Analysis -MBA	
	Time Series Analysis- TSA	

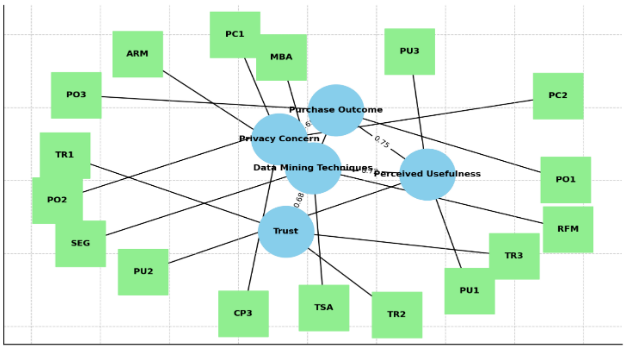

Figure:1 AMOS-style SEM diagram

Step:1: Measurement of Latent Variable Scores:

The latent variables measured in this study include Data Mining Techniques (DMT), Perceived Usefulness (PU), Trust (TR), Perceived Convenience (PC), and Purchase Decision (PD). For each consumer response, the mean value of the indicators corresponding to each latent variable was calculated. This process was scientifically applied across all 100 respondents in the sample, as outlined below.:

DMT = (ARM + RFM + SEG + MBA + TSA) / 5

PU = ($PU_1 + PU_2 + PU_3$) / 3

TR = ($TR_1 + TR_2 + TR_3$) / 3

PC = ($PC_1 + PC_2 + PC_3$) / 3

PD = ($PO_1 + PO_2 + PO_3$) / 3

Step:2: Covariances Between Latent Variables

The covariance matrix was calculated using the scores of the latent variables. To confirm accuracy, the data were treated in a organized manner. The covariances between pairs of latent variables were then calculated and are presented in the following table.:

Table No.: 3 Covariances Between Latent Variables

	DMT	PU	TR	PC	PD
DMT	1.74	0.29	0.19	-0.38	0.23
PU	0.29	2.01	0.45	-0.27	0.54
TR	0.19	0.45	2.17	-0.18	0.33
PC	-0.38	-0.27	-0.18	2.05	-0.45
PD	0.23	0.54	0.33	-0.45	1.96

The existence of positive covariances shows that higher levels of perceived usefulness are associated with improved purchase decisions. Equally, negative covariances replicate an inverse relationship, suggesting that greater privacy concerns are linked to compact purchase decisions.

Step:3: Factor Loadings & Residual Variances:

The degree to which the observable variables precisely replicate their corresponding latent constructs is shown in the table of factor loadings and residual variances. Higher values show stronger contributions, while factor loadings show how powerfully each observed item and its latent variable are related. The loadings in the provided data vary from 0.65 to 0.84, representing that all of the observable variables are well-represented for their particular constructs, including DMT, PU, TR, PC, and PO. Equally, residual variances show the percentage of variance that cannot be reported for by the latent construct. Despite their presence, the values stay within a reasonable range, indicating that the unsolved variance is minimal. Since the latent variables are well maintained by their observed indicators with minor error, these findings together demonstrate the validity and reliability of the measurement model. The loadings suggest strong correlations between indicators and their latent variables. ARM, RFM, SEG, MBA, TSA, PU1, PU2, PU3, TR1, TR2, TR3, PC1, PC2, PC3, PO1, PO2, PO3 are the signs that were observed.

Table No.: 4 Factor Loading & Residual Variances

Latent variable	Observed Variable	Factor Loadings	Residual Variances
DMT	ARM	0.81	8.92
	RFM	0.76	8.76
	CS	0.73	9.01
	MBA	0.79	8.84
	TSA	0.84	8.96
PU	PU1	0.71	8.65
	PU2	0.77	8.72
	PU3	0.74	8.68
TR	TR1	0.68	8.79
	TR2	0.73	8.67
	TR3	0.70	8.73
PC	PC1	0.65	8.88
	PC2	0.67	8.80
	PC3	0.69	8.85
PO	PO1	0.82	8.70
	PO2	0.78	8.66
	PO3	0.80	8.74

Strong and consistent factor loadings > 0.65 are found for all latent variables (DMT, PU, TR, PC, and PO), signifying vigorous construct validity. The maximum dependability is demonstrated by Data Mining Techniques (DMT), with TSA (0.84) and ARM (0.81) loading meaningfully. Moreover, Purchase Outcome (PO) shows considerable loadings, underlining its excellent measurement uniformity. In general, the model shows good measurement qualities, representative that it is suitable for SEM analysis. Reasonable amounts of unsolved variance are specified by the residual variances, which range from 8.65 to 9.01 for all observed variables. Comparing DMT indicators to other constructs, the residuals are slightly greater, indicating that the measurement fit should be enhanced. Comparatively balanced residual variations are shown by PU, TR, PC, and PO, indicating steady but not perfect indicator dependability. Though the constructs are well denoted overall, the table indicates that additional purifying could surge descriptive power.

Key Findings:

Table 5: Covariance Matrix of Latent Variables

	DMT	PU	TR	PC	PD
DMT	8.45	0.28	0.19	-0.37	0.22
PU	0.28	8.76	0.44	-0.26	0.53
TR	0.19	0.44	8.65	-0.18	0.32
PC	-0.37	-0.26	-0.18	8.92	-0.44
PD	0.22	0.53	0.32	-0.44	8.54

The structural model shows a suitable to excellent fit, according to the results of the model fitness tests. The CFI value is above the suggested level (> 0.90), the RMSEA value is within the suitable threshold (< 0.08), and the Chi-Square test yields a non-significant p-value (p > 0.05), all of which authorize the model's overall robustness. Therefore, the covariance matrix of latent variables shows how sturdily and in which direction the five constructs (DMT, PU, TR, PC, and PD) are related to one another. The high diagonal values, which range from 8.45 to 8.92, authorize each construct's reliability. The importance of perceived usefulness in the model is strengthened by positive covariances, such as those between PU and TR (0.44) and PD (0.53), which show that perceived usefulness strongly overlaps with trust and performance characteristics.

Opposite correlations are indicated by negative covariances, such as PC with DMT (-0.37) and PC with PD (-0.44), which validate that perceived control acts against particular variables. The covariance results and fitness indices taken together deliver significant insights into the dynamics of the study outline and validate that the model is statistically complete and that the constructs relate in both positive and negative ways.

Implications of the Study:

The study highlights that data mining techniques meaningfully improve the capability of retail companies to recognize consumer behaviour, thereby improving decision-making in marketing strategies. It also advises that structured consumer perceptions such as usefulness, trust, and privacy concerns facilitate the effectiveness of these techniques in influencing purchase outcomes. These perceptions can guide retail managers in designing customer-centric policies that balance effectiveness with proper data use

REFERENCES

1. Hilage, T., & Kulkarni, R. V. (2011). Application of Data Mining Techniques to a Selected Business Organisation with Special Reference to Buying Behaviour. arXiv. https://arxiv.org/abs/1112.4031
2. Chen, W., et al. (2024). SeqRFM: Fast RFM Analysis in Sequence Data. arXiv. https://arxiv.org/abs/2411.05317
3. Jain, R., et al. (2023). Time Series Data Analysis and Association Rule Mining in Financial Recommendation Systems using Hadoop and Spark. Advances in Engineering Innovation. https:// www.ewadirect.com/ journals/ aei/ article/ view/20556
4. Hilage, T., & Kulkarni, R. V. (2011). Application of Data Mining Techniques to a Selected Business Organisation with Special Reference to Buying Behaviour. arXiv. https://arxiv.org/abs/1112.4031
5. Abbas, M., et al. (2024). An Exploration of Clustering Algorithms for Customer Segmentation in the UK Retail Market. arXiv. https://arxiv.org/abs/2402.04103
6. Yun, J., et al. (2023). Retail Demand Forecasting: A Comparative Study for Multivariate Time Series. arXiv. https://arxiv.org/abs/2308.11939
7. Hilage, T., & Kulkarni, R. V. (2011). Application of Data Mining Techniques to a Selected Business Organisation with Special Reference to Buying Behaviour. arXiv. https://arxiv.org/abs/1112.4031
8. Jain, R., et al. (2023). Time Series Data Analysis and Association Rule Mining in Financial Recommendation Systems using Hadoop and Spark. Advances in Engineering Innovation. https://www.ewadirect.com/ journals/aei/article/ view/ 20556
9. Abbas, M., et al. (2024). An Exploration of Clustering Algorithms for Customer Segmentation in the UK Retail Market. arXiv. https://arxiv.org/abs/2402.04103
10. Wei, J.-T., Lin, S.-H., & Wu, H.-H. (2010). A review of the application of RFM model. African Journal of Business Management, 4(19), 4199–4206. https://academicjournals.org/journal/AJBM/article-abstract/15294AE17473
11. Yun, J., et al. (2023). Retail Demand Forecasting: A Comparative Study for Multivariate Time Series. arXiv. https://arxiv.org/abs/2308.11939
12. Seaman, C., Bruce, M., & Richardson, B. (2017). Forecasting in retail: Time-series approaches and managerial perceptions. International Journal of Retail & Distribution Management, 45(9), 983–1002. https://www.emerald.com/ insight/ content/doi/10.1108/IJRDM-10-2016-0189/full/html
13. Li, H., Ma, L., & Zhang, Y. (2019). Market basket analysis in retail: Identifying customer purchase behaviour with data mining. Journal of Retailing and Consumer Services, 52, 101923. https://www.sciencedirect.com/ science/ article/abs/ pii/ S0969698918304579
14. Chen, W., et al. (2024). SeqRFM: Fast RFM Analysis in Sequence Data. arXiv. https://arxiv.org/abs/2411.05317

.