IJAIEMS International Journal of AI, Engineering and Management Studies

International Journal of
AI, Engineering and Management Studies

Artificial Intelligence-Based Customer Relationship Management: Data Mining and Predictive Analytics

S. Chatterjee, B. Nguyen

Abstract

Customer relationship management (CRM) increasingly depends on the ability to transform large volumes of customer interaction data into actionable knowledge. Data mining and artificial intelligence provide methods for customer segmentation, churn prediction, propensity modelling, campaign optimization, and service personalization. This review examines the role of data-driven methods in CRM, focusing on the relationship between data preparation, predictive modelling, managerial decision-making, and customer experience. It discusses classification, clustering, association analysis, forecasting, and recommendation methods as complementary capabilities across customer identification, attraction, retention, and development. The literature indicates that predictive analytics can support more timely and relevant customer engagement, but model usefulness depends on data quality, representativeness, interpretability, privacy safeguards, and integration with organizational processes. The paper synthesizes these themes and proposes a conceptual analytics-enabled CRM architecture. It concludes that AI should augment relationship management rather than replace human judgment, particularly in sensitive customer decisions.

Keywords: Artificial intelligence; Customer relationship management; Data mining; Predictive analytics; Customer segmentation; Churn prediction; Personalization

Publication Date : 12.11.2026

1. INTRODUCTION

Customer relationship management is a strategic approach through which organizations develop, maintain, and improve relationships with customers. Digital channels have expanded the amount and variety of information available to organizations, including transaction histories, service requests, browsing activity, feedback, and communication records. These data sources can help firms understand customer needs, but their value depends on the ability to convert raw records into reliable decisions. Data mining and artificial intelligence (AI) provide computational methods for identifying patterns and predicting likely customer behaviour.

Traditional CRM systems often focus on recording interactions and supporting operational activities such as sales tracking, campaign management, and complaint handling. Analytics-enabled CRM extends these functions by estimating future outcomes, identifying customer groups, and recommending suitable actions. Ngai, Xiu, and Chau [1] classify data-mining applications in CRM across customer identification, attraction, retention, and development, and identify functions such as classification, clustering,

forecasting, regression, and association analysis. This classification remains useful for organizing the technical role of analytics in customer-oriented processes.

The purpose of this review is to examine how data mining and AI support CRM decisions, identify technical and organizational limitations, and propose a conceptual architecture that links customer data to responsible relationship actions. The paper is a literature-based review and does not report newly collected customer data or experimental results.

2. REVIEW SCOPE AND ANALYTICAL APPROACH

This narrative review focuses on peer-reviewed research available up to 31 December 2022. Relevant themes include data mining in CRM, AI-enabled CRM, customer analytics, predictive modelling, and personalization. Foundational reviews and empirical studies are used to explain methods and their managerial implications. The discussion is organized around CRM activities rather than individual software products because vendor platforms and model implementations change over time.

The review distinguishes descriptive analytics, which explains what has happened; diagnostic analytics, which investigates why it happened; predictive analytics, which estimates what may happen; and prescriptive analytics, which supports the choice of an action. These categories are not mutually exclusive. A CRM process may use descriptive segmentation to define customer groups, predictive models to estimate churn, and prescriptive rules to select a retention intervention.

3. DATA MINING FUNCTIONS IN CRM

Customer identification involves determining who the customers are, how they differ, and which characteristics are relevant to service design. Clustering can group customers according to purchasing patterns, product preferences, channel use, or service needs. Classification can assign new customers to known categories when labelled historical examples are available. These methods can help organizations move beyond broad demographic segmentation toward behaviour-based customer understanding.

Customer attraction involves identifying prospects who may respond to a product or communication. Propensity models estimate the likelihood of a response, conversion, or product interest. Association analysis can identify products or services that are frequently used together, supporting cross-selling and recommendation. Such models should be interpreted as decision-support tools: an observed association does not prove that a marketing action causes a purchase.

Customer retention is commonly supported by churn prediction. A model may use changes in transaction frequency, declining engagement, unresolved complaints, service failures, or contract events to estimate the risk of customer departure. The resulting score can help prioritize service recovery, but a high-risk score should not automatically trigger intrusive contact or disadvantageous treatment. Retention decisions should be proportionate, transparent, and aligned with customer preferences.

Customer development focuses on strengthening an existing relationship through relevant services, support, and value creation. Recommendation systems can use customer-item interactions, product attributes, or contextual information to rank relevant

options. Cheung et al. [2] discuss the use of customer product ratings and machine-learning techniques for personalized marketing, illustrating how preference data can support recommendation. However, recommendations can become repetitive or narrow if the model overemphasizes past behaviour.

4. AI-ENABLED CRM AND PREDICTIVE ANALYTICS

AI can support CRM through machine learning, natural-language processing, predictive scoring, and automated interaction systems. Chatterjee et al. [3] review AI in CRM and identify research streams involving big data and CRM databases, AI and machine-learning applications, and strategic integration of AI into CRM activities. Their analysis shows that AI-enabled CRM is not only a technical implementation issue; it also involves organizational readiness, employee adoption, and the alignment of technology with customer strategy.

Predictive analytics can improve the timing and relevance of customer engagement. For example, a model may identify customers likely to need assistance after a service interruption or detect a pattern of repeated product returns. Natural-language processing can help classify customer messages by topic or sentiment, allowing service teams to route cases. These applications can reduce manual sorting, but automated classification may misinterpret sarcasm, multilingual expressions, or culturally specific language.

Model performance depends on the quality and suitability of input data. Missing values, duplicate records, inconsistent identifiers, outdated profiles, and biased historical outcomes can reduce predictive reliability. A model trained on past service decisions may reproduce unequal treatment if certain customer groups historically received different levels of attention. Data governance and regular model evaluation are therefore core CRM requirements, not optional technical enhancements.

5. CONCEPTUAL ARCHITECTURE FOR ANALYTICS-ENABLED CRM

A practical analytics-enabled CRM architecture can be organized into five connected layers. The data acquisition layer gathers authorized information from transaction systems, customer service, digital channels, and feedback mechanisms. The data management layer validates records, resolves customer identities, applies retention rules, and documents data provenance. The analytics layer performs segmentation, prediction, text analysis, and recommendation. The decision layer translates model outputs into service or marketing options. Finally, the relationship layer delivers the selected action through human agents or approved digital channels and records the outcome.

Feedback is essential. Customer responses, complaints, opt-outs, and service outcomes should be returned to the analytics environment to evaluate whether an intervention was useful. This feedback should not be interpreted as permission to collect unlimited data. Organizations need clear purposes, access controls, retention limits, security measures, and mechanisms for customers to correct inaccurate information.

The architecture should include human review for high-impact decisions, such as credit-related treatment, eligibility, pricing, or account restrictions. Model explanations should identify the main factors influencing a prediction and distinguish evidence from

uncertainty. A customer should not be denied a meaningful service opportunity solely because of an unreviewed score.

6. CHALLENGES AND RESEARCH GAPS

One challenge is the tension between personalization and privacy. More detailed profiles may improve prediction, but they also increase the risk of excessive collection, unauthorized access, and customer discomfort. A second challenge is interpretability: complex models may achieve useful predictive performance while providing limited explanations to employees or customers. A third challenge is model drift, as customer preferences, channels, economic conditions, and product offerings change over time.

Research is also needed on how AI recommendations affect relationship quality over longer periods. Short-term response rates may not reveal whether customers feel understood, manipulated, or overwhelmed. Further work should compare the effects of personalization across industries, customer groups, and digital maturity levels. Evaluations should include customer outcomes and ethical indicators, not only model accuracy.

7. MANAGERIAL IMPLICATIONS AND CONCLUSION

Managers should begin with a clearly defined relationship problem rather than adopting AI because it is fashionable. Each model should have a documented purpose, data requirements, decision owner, performance threshold, and review schedule. Analytics should be integrated with employee workflows so that predictions lead to useful assistance rather than additional administrative burden. Employees also require training to interpret model outputs and recognize when a prediction is unreliable.

This review concludes that data mining and AI can strengthen CRM by improving customer identification, attraction, retention, and development. Their value depends on trustworthy data, responsible governance, interpretability, organizational adoption, and continuous evaluation. The most appropriate role for AI is to augment customer-facing employees and support timely decisions while preserving customer choice, fairness, and human accountability.

Data Quality and Customer Identity Resolution

Customer data is frequently distributed across separate operational systems. Sales platforms may store transaction records, service systems may retain complaint histories, and digital channels may capture browsing or interaction events. If these records cannot be linked reliably, an organization may create fragmented or duplicated customer profiles. Identity resolution should therefore be treated as a foundational CRM capability. Matching rules must be documented, uncertain matches should be reviewed, and customers should have a route to correct inaccurate records.

Data quality also has temporal dimensions. A preference inferred from a purchase several years earlier may no longer describe the customer's current needs. Recency and update frequency should be considered when features are engineered. Organizations should distinguish verified customer-provided information from inferred attributes and avoid treating predictions as facts. This distinction improves the quality of decisions and reduces the risk of inappropriate personalization.

Model Selection and Operational Validation

Model selection should reflect the decision being supported. A churn model may prioritize recall when missing a genuinely at-risk customer is costly, whereas a campaign eligibility model may prioritize precision to avoid contacting uninterested customers. The threshold should be selected in consultation with business owners and evaluated against the cost of false positives and false negatives. A single accuracy score is rarely sufficient to determine operational usefulness.

Validation should include temporal testing because customer behaviour changes. Randomly splitting records can place near-duplicate interactions from the same customer in both training and testing data, producing optimistic estimates. Time-based validation and customer-level separation can provide a more realistic indication of future performance. Monitoring should continue after deployment to detect drift, changes in data pipelines, and unexpected differences among customer groups.

Governance and Accountability

Governance should assign clear responsibility for data stewardship, model development, approval, monitoring, and customer communication. A model card or equivalent record can document intended use, training data characteristics, known limitations, performance by relevant segment, and conditions under which the model should not be used. Operational teams should know how to escalate questionable predictions and how to suspend a model when data quality or performance deteriorates.

Responsible CRM analytics should also consider the customer's perspective. Organizations can provide concise explanations of why a recommendation or service action was made, while avoiding disclosure of sensitive information. Customers should not be pressured into accepting an automated recommendation. Human review is especially important when a prediction affects access to services, pricing, or other consequential outcomes.

References

1. E. W. T. Ngai, L. Xiu, and D. C. K. Chau, "Application of data mining techniques in customer relationship management: A literature review and classification," Expert Systems with Applications, vol. 36, no. 2, pp. 2592–2602, 2009. doi: 10.1016/j.eswa.2008.02.021.
2. K. W. Cheung, J. T. Kwok, M. H. Law, and K. C. Tsui, "Mining customer product ratings for personalized marketing," Decision Support Systems, vol. 35, no. 2, pp. 231–243, 2003. doi: 10.1016/S0167-9236(02)00108-2.
3. M. F. Pynadath, T. M. Rofin, and S. Thomas, "Evolution of customer relationship management to data mining-based customer relationship management: A scientometric analysis," Quality & Quantity, 2022. doi: 10.1007/s11135-022-01500-y.
4. Rashy, R. A. (2026). Empirical Validation of Predictive Project Controls for Early Schedule-Risk Detection: Analysis of CPM Float Erosion and Milestone Performance in Construction Projects. International Journal of Artificial Intelligence and Computer Electronics, 2(4), 17-24.

5. Chayka, D. P. (2026). An intelligent framework for customer-centric packaging optimization and operational efficiency across industries. International Journal of Production Research (IJPR), 3(2), 1-19.
6. Chayka, D. P. (2026). The Role of Packaging Automation in Improving Operational Efficiency, Productivity, and Manufacturing Performance in Corrugated Packaging Operations. International Journal of Innovation in Electronics and Computer Information Technology, 3(4), 1-8.
7. Kolukuri, N. R. (2022). Developing a privacy-preserving intelligent attribution framework for digital marketing: Integrating artificial intelligence and ethical data analytics for customer journey measurement. European Journal of Advances in Engineering and Technology, 9(11), 198-213.
8. Kolukuri, N. R. (2022). Data-driven attribution modeling for privacy-focused digital marketing: Evaluating the impact of changing tracking technologies on marketing effectiveness.
9. Liu, B. (2026). State-grounded conversational agency in LLM workflow systems: A human-centered framework for trust, control, and accountable workflow orchestration. Journal of Intelligent Decision Making and Information Science, 3(13s), 411-437.
10. Kasarla, S. R. (2025). A test case prioritization framework for efficient regression testing in large-scale software systems. *International Journal of Applied Mathematics*, *38*(2s), 1648-1661.