Minnesota Journal of Business Law and Entrepreneurship

Volume 2026, No. 3

ISSN: 1540-3270

Artificial Intelligence in Business Applications: A Systematic Review

Armaan Gugnani
Independence High School, Frisco, Texas

Abstract

The research articles considered in this review paper collectively examine AI and machine-learning applications in business, mainly in three areas: customer segmentation and digital marketing, employee attrition and human-resource analytics, and export-sales forecasting. Most studies report strong predictive or clustering performance, but several have important weaknesses involving class imbalance, small or synthetic and public datasets, possible data leakage, weak validation, and limited business evaluation.

Keywords: customer segmentation, sales forecasting, employee attrition

1. Introduction

Artificial Intelligence (AI) and Machine Learning (ML) have emerged as important technologies for transforming business decision-making by enabling organizations to extract meaningful insights from large and complex datasets. Their applications have expanded across several business functions, including customer segmentation and digital marketing, human-resource analytics and employee-attrition prediction, and export-sales forecasting. In customer analytics, techniques such as K-means, hierarchical clustering, DBSCAN, PCA, and evolutionary optimization methods are used to identify customer groups and support targeted marketing strategies. Similarly, supervised learning approaches, including Random Forest, logistic regression, and decision trees, have been widely applied to predict employee attrition, job-search intentions, and other HR-related outcomes. ML-based forecasting approaches have also been explored for predicting export sales and supporting business planning. Although the reviewed studies generally report promising predictive or clustering performance, their practical reliability is affected by several methodological limitations, including the use of small, synthetic, public, or single-organization datasets, class imbalance, possible data leakage, overfitting, inconsistent validation and evaluation metrics, and limited external validation. Furthermore, many studies focus primarily on technical performance measures such as accuracy rather than demonstrating measurable business outcomes, while complex ensemble and evolutionary models may provide limited interpretability for business decision-makers. The dynamic nature of customer and employee behaviour also highlights the need for adaptive rather than purely static models. Therefore, this review critically examines existing AI and ML approaches used in business applications, compares their strengths and limitations, identifies key methodological and practical research gaps, and highlights future directions toward more reliable, explainable, generalizable, and business-oriented AI/ML solutions.

2. Study Characteristics

Table I presents the key characteristics of the studies reviewed in this paper, including the publication year and the corresponding business application addressed by each study

Table I: Key characteristics of existing studies

Study	Year	Business Application
Wang	2025	Digital-marketing customer segmentation
Turkmen	2022	Online-retail customer segmentation

Minnesota Journal of Business Law and Entrepreneurship

Volume 2026, No. 3

ISSN: 1540-3270

Hung, Lien and Ngoc	2019	Credit-card customer segmentation
Sohrabpour et al.	2021	Export-sales forecasting
Papineni et al.	2021	Prediction of job-search intention
Krishna and Sidharth	2022	Employee-attrition prediction
Alsaadi, Khlebus, and Alabaichi	2022	HR analytics and employee-turnover prediction
Ding	2022	HR administration and HRM prediction
Durga et al.	2022	HR data analysis and employee-attrition prediction

3. Comparative Methodological Assessment

Table II summarizes the major AI and machine-learning methods identified in the reviewed studies, highlighting their strengths, limitations, and appropriate business applications

Table II: Major AI and ML methods in the existing studies

Method	Strengths in Business Applications	Limitations	Appropriate Use
Random Forest	Handles nonlinearities and interactions; robust to many data forms; offers feature importance	Can obscure decision logic; may be slow; inflated performance can occur through leakage or improper validation	Attrition, churn, risk, and business classification
K-means	Simple, scalable, interpretable through centroids	Requires selecting K; sensitive to initialization, scale, and outliers; forces every case into a cluster	Large transactional segmentation tasks
Hierarchical clustering	Produces a dendrogram; does not initially require a fixed K	Computationally expensive; linkage choice affects results; may produce coarse segments	Small and medium datasets requiring structural exploration
DBSCAN	Detects irregular cluster shapes and noise	Sensitive to density parameters; difficult with	Anomaly-rich spatial or

		varying density; noise may include valuable customers	behavioural datasets
PCA	Reduces multicollinearity, noise, and computational burden	Components are less directly interpretable for managers	High-dimensional customer or HR data
Genetic programming	Discovers nonlinear symbolic equations without prespecifying model form	Can generate unstable or overly complicated expressions; overfitting risk with small datasets	Causal forecasting and symbolic regression
SMOTE	Increases minority-class representation and can improve training metrics	Synthetic samples may not improve real holdout performance; must be applied only within training folds	Imbalanced attrition, fraud, and churn classification
Q-learning and differential evolution	Adaptively optimize clustering parameters and initialization	Computationally complex and difficult to explain	Advanced segmentation where performance justifies complexity
Logistic regression	Transparent, fast, and useful as a baseline	Limited for complex nonlinear relationships without transformations and interactions	Explainable binary classification
Decision tree	Easy to visualize and convert into rules	Unstable and prone to overfitting without pruning	Rule-based business decision support

4. Research Gaps

Table III presents the key research gaps identified across the reviewed studies, along with supporting evidence from the existing literature

Table III: Research gaps identified in existing studies

Research Gap	Evidence
Limited external validity	Many studies use one Kaggle dataset or one company
Inadequate imbalance evaluation	Several HR papers emphasize accuracy while attrition is rare
Possible leakage and overfitting	Multiple papers report 99-100% accuracy without rigorous independent testing
Lack of business-impact evidence	Clustering studies rarely test campaign outcomes
Weak explainability	Hybrid evolutionary and ensemble models are difficult for managers to interpret
Static models dominate	Customer and employee behaviour changes over time

Reproducibility is limited	Several papers omit code, seeds, parameter settings, or complete datasets
Comparisons are inconsistent	Different metrics and splits prevent direct comparison

5. Conclusion

It is concluded that the reviewed studies demonstrate that AI and ML techniques can provide valuable insights and support data-driven decision-making across diverse business functions. Clustering and predictive models have generally achieved promising results, indicating their potential for improving customer targeting, employee-retention strategies, and sales forecasting. However, the review also identifies several methodological and practical limitations that affect the reliability and generalizability of existing research. Class imbalance remains a significant challenge, particularly in employee attrition prediction, while the use of small, synthetic, or publicly available datasets may limit the applicability of findings to real-world business environments. In addition, inadequate validation procedures and the possibility of data leakage can result in overly optimistic performance estimates. Another important limitation is the limited consideration of business-oriented evaluation measures, as many studies primarily focus on conventional machine-learning metrics rather than assessing the actual financial, operational, or strategic benefits of the proposed models. Future research should therefore focus on developing robust and generalizable AI/ML solutions using larger, diverse, and real-world datasets. Greater attention should be given to appropriate handling of class imbalance, rigorous cross-validation, prevention of data leakage, external validation, and explainable AI to improve the transparency and trustworthiness of predictive systems. Furthermore, future studies should integrate technical performance with business-oriented metrics such as cost reduction, revenue improvement, customer lifetime value, employee retention impact, and return on investment. The integration of advanced approaches such as explainable AI, ensemble learning, deep learning, and generative or agentic AI may further enhance business analytics. Overall, the reviewed literature confirms the significant potential of AI and ML for business decision-making, while also highlighting the need to move from high predictive performance toward reliable, interpretable, validated, and measurable real-world business impact.

References

[1] Alsaadi, E. M. T. A., Khlebus, S. F., & Alabaichi, A. (2022). Identification of human resource analytics using machine learning algorithms. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*.

[2] Ding, G. (2022). Optimization of human resource administration system based on DM technology and random forest algorithm. *Mobile Information Systems, 2022*, Article 4359535.

[3] Durga, S., Perugu, P., D., N. S., Podile, V., Bhanushali, M. M., & Revathi, R. (2022). Human resource data analysis & prediction using decision tree algorithm and random forest. *NeuroQuantology, 20*(9), 4803-4809.

[4] Hung, P. D., Lien, N. T. T., & Ngoc, N. D. (2019). Customer segmentation using hierarchical agglomerative clustering. In *Proceedings of the 2019 2nd International Conference on Information Science and Systems* (pp. 33-37). Association for Computing Machinery.

[5] Krishna, S., & Sidharth, S. (2022). HR analytics: Employee attrition analysis using random forest. *International Journal of Performability Engineering*.

[6] Papineni, S. L. V., Reddy, A. M., Yarlagadda, S., Yarlagadda, S., & Akkineni, H. (2021). An extensive analytical approach on human resources using random forest algorithm. *International Journal of Engineering Trends and Technology, 69*(5), 119-127.

[7] Sohrabpour, V., Oghazi, P., Toorajipour, R., & Nazarpour, A. (2021). Export sales forecasting using artificial intelligence. *Technological Forecasting and Social Change, 163*, Article 120480.

[8] Türkmen, B. (2022). Customer segmentation with machine learning for online retail industry. *The European Journal of Social and Behavioural Sciences, 31*(2).

[9] Wang, G. (2025). Customer segmentation in digital marketing using a Q-learning-based differential evolution algorithm integrated with K-means clustering. *PLOS ONE, 20*(2), e0318519