ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue XII, December 2025

Comparative Analysis of Machine Learning and AI-Powered Data Warehousing for Employee Attrition and Performance Optimization

Ginalyn I. Contillo, Marvin A. Yambao

Batangas State University, Batangas City, Philippines

DOI : https://doi.org/10.51583/IJLTEMAS.2025.1412000027

Received: 20 October 2025; Accepted: 27 October 2025; Published: 31 December 2025

ABSTRACT

In the age of digital revolution, organizations increasingly use artificial intelligence (AI) and machine learning (ML) to improve their data-driven decision-making, especially in human resource management. This paper makes a comparative evaluation of AI-driven data warehousing systems and ML methods for forecasting employee turnover and maximizing employee performance. The study compares top data warehousing platforms like Redshift, BigQuery, Snowflake, and Databricks and their coupling with ML models with regard to prominent workforce features.

Qualitative findings from HR managers were also examined, in order to evaluate the real-world effect of these technologies on the productivity of the workforce and employment strategies. Research shows that AI-based data warehousing integrated with competent machine learning models drastically enhances attrition prediction accuracy, performance tracking, and strategic workforce planning.

This research identifies the strategic advantages of combining AI-driven data warehousing with HR analytics, offering organizations actionable findings to choose the best AI-enabled solutions. The findings contribute to extending knowledge on efficient data strategies in lessening attrition as well as improving employee performance, aiding organizations in their pursuit of strategic human capital objectives.

Keywords - Data Processing, Comparative analysis, Data Warehousing, Performance Optimization, Prediction and Attrition, Machine Learning

INTRODUCTION

Employee turnover is a big problem for organizations, leading to disruptions, reduced productivity, and added expense. Awareness of turnover drivers and the implementation of successful retention tactics are crucial for long-term growth.

Artificial intelligence (AI) has revolutionized data warehousing, facilitating sophisticated analysis of workforce data using tools such as Redshift, BigQuery, Snowflake, and Databricks. These tools facilitate predictive modeling and real-time analysis of attrition and employee performance.

Nonetheless, very little comparative research exists on various AI-powered data warehousing techniques and machine learning models in HR analytics. Most of the research concentrates on one technique, and this keeps organizations in suspense about the suitable tools for their organizations or how these technologies affect quantifiable workforce outcomes.

This research compares various AI-powered data warehousing solutions and machine learning models to forecast employee turnover and performance optimization. Major attributes like job title, overtime, job level, and stock options are studied with evaluation in terms of accuracy, efficiency, and HR system integration.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue XII, December 2025

Based on the concepts of Sociotechnical Systems and Performance Optimization, the study investigates how the interactions between technology and human aspects shape workers' outcomes. The research seeks to offer actionable guidance for the selection of AI solutions that enhance retention and productivity of the workforce.

RELATED LITERATURE

A. Data-Driven Employee Atrittion

Singh et al. (2012) created an analytics-driven framework to prevent voluntary employee turnover. The research concentrated on combining data mining and predictive analytics with HR decision-making systems, applying classification algorithms, logistic regression, and decision trees on historical HR databases. The models proved high in accurate prediction of at-risk employees, validating that early detection can lead to successful retention. The authors drew the conclusion that predictive analytics may assist HR efforts by utilizing data warehousing and machine learning algorithms. This paper emphasizes how AI-based predictive systems, when incorporated into data warehouses, can improve employee performance indirectly by minimizing wasteful attrition.

Mishra and Mishra (2013) performed a review of literature in order to compile the numerous factors that affect employee attrition and retention. Although not technical, their effort is crucial to understanding those human factors that shape up employee performance but AI must account for. Content analysis of HR research led them to the conclusion that those factors namely included: job satisfaction, career advancement opportunity, as well as quality of leadership. Their study particularly highlights the need for any performance optimization framework to integrate principles from human resource. Although not AI-focused, the research educates AI-based warehousing models about which human performance measures to focus on for more complete solutions.

Jain et al. (2020) sought to describe and forecast employee attrition through machine learning techniques. The research incorporated machine learning algorithms decision trees, random forests, and logistic regression into HR data processes. Of these, the random forest model was most accurate in forecasting attrition. The results proved that employee turnover can be foreseen using the appropriate features and data integration, validating the viability of AI in big HR datasets. It was concluded that ML algorithms are eminently suited to deal with and analyze employee lifecycle data. This directly feeds into AI-driven data warehousing solutions dealing with real-time employee performance tracking.

Yahia et al. (2021) investigated moving from big data to "deep data" to enhance employee attrition prediction with AI models. The research utilized deep learning and semantic data layering, blending structured and unstructured data to offer richer analytics. By enriching data and augmenting model training, they were able to deliver better prediction quality. The research indicated that data at the semantic level greatly enhances predictive performance, demonstrating that the depth and richness of data are essential. The conclusion emphasized that enriched datasets have the ability to improve the performance of AI-powered analytics. This affirms that sophisticated data warehousing should concentrate on semantic integration for the best AI performance management.

Oke et al. (2016) carried out a literature review of teacher attrition and retention, with a qualitative theme emphasis over technical application. Through thematic analysis, they found policy and workplace environment to be key to employee retention. The results indicated that organizational support is key to employee satisfaction and stay or leave decisions. Even though the research did not have an AI or warehousing element, its focus on institutional context is crucial. It provides valuable information on the performance metrics that AI-based warehousing software must track, especially in education or comparable service-oriented industries.

Uddin and Hossan (2024) presented a detailed review of AI-driven data warehousing solutions utilized to streamline big data management. Their conversation included cloud warehousing, ETL (Extract, Transform, Load) processes based on artificial intelligence (AI), and automation technologies. Based on case examples, the review presented a marked improvement in processing speed, data precision, and system readability. The

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue XII, December 2025

research concluded that data warehousing based on AI is crucial for facilitating real-time, data-informed decision-making. The paper is of immediate relevance to your research, presenting good technical evidence for contrasting multiple AI methods in warehousing with a focus on performance enhancement.

Machireddy and Devapatla (2023) contrasted AI-based and conventional Robotic Process Automation (RPA) models in cloud data warehousing setups. In their experimental work, they measured efficiency and accuracy in terms of primary performance metrics like data throughput and latency. AI-RPA models outperformed conventional ones decisively, especially in terms of automation and dependability. The research found that contemporary AI methods provide better warehousing results. These results support your research interest by illustrating that various AI methods are comparable on their quantifiable effect on performance.

Tsou (2024) investigated how AI-enabled automation increases warehouse accuracy and efficiency. With real-time monitoring and AI-enabled control systems, the case study discovered that automation resulted in an over 30% boost in operational accuracy. While the emphasis was on physical warehousing, the findings highlight the general advantages of AI automation. The research found that automation realizes concrete gains, further attesting to the transferability of these concepts to digital data warehousing. These findings attest to the transferability of AI automation advantages to performance enhancement in employee-centric data environments.

Gudelli (2023) assessed the impact of AI-driven tools in enhancing performance in AWS cloud environments. Using an observational case study, the study pointed out AI applications in anomaly detection, workload management, and resource optimization. The study indicated that AI has a substantial impact on both system and employee productivity using intelligent automation. The study concluded that cloud systems powered by AI are necessary for enhancing operational performance. This study is in agreement with your research in that it demonstrates how AI-boosted warehousing can automate processes and enhance employee efficiency through enhanced data infrastructure.

Rella (2025) compared data lakes and data warehouses to determine their effectiveness in supporting machine learning applications. The study analyzed system architecture, scalability, and integration benchmarks, concluding that data warehouses provide more structure while data lakes offer greater flexibility. The results suggested that the best choice depends on the specific performance and analytical needs of an organization. This work concluded that system design significantly impacts machine learning outcomes. It offers a useful foundation for your comparative analysis by presenting insights into the impact of varying warehousing structures on AI deployment and performance improvement.

METHODOLOGY AND RESULT

A. Data Collection

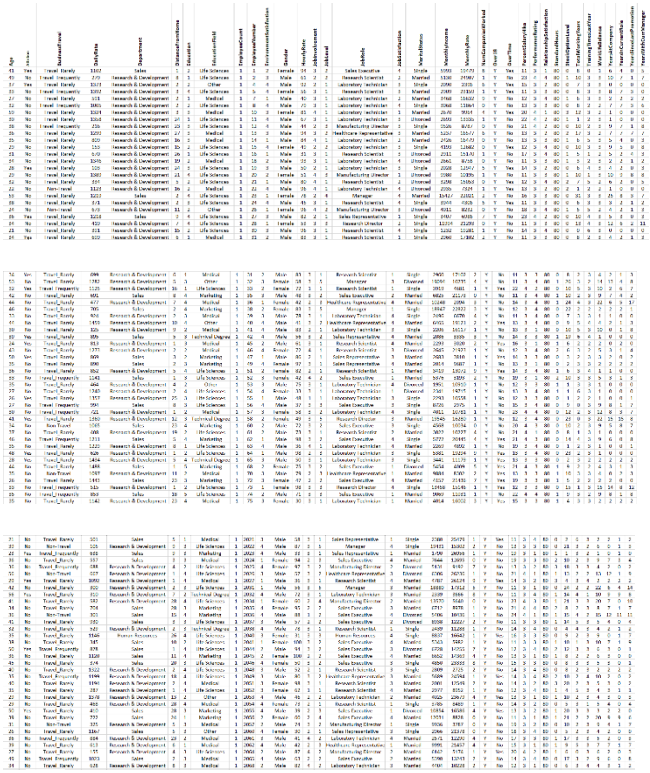
To facilitate the analysis of employee attrition and performance maximization, the current research employed a publicly available synthetic dataset from Kaggle, a prominent online website for data science competitions and datasets. The used dataset is named "HR Analytics," which was initially designed for training and analytical purposes. It comprises 1,470 simulated employee records from diverse departments of a hypothetical organization.

This data set comprises a large set of applicable features like age, gender, department, job designation, education level, job satisfaction, environment satisfaction, years in company, number of companies worked for, training times in last year, monthly income, overtime status, and attrition status (if the employee has left the organization). These variables facilitate thorough examination of the variables that can lead to employee turnover, making the dataset extremely suitable for use in machine learning and predictive modeling applications involving Human Resource analytics.

Notably, the data is synthetic, whereby the data were created to represent real organizational scenarios without drawing upon any actual employee records. This guarantees adherence to ethical guidelines relating to data confidentiality and privacy. Consequently, it is extensively employed in research and academic settings for the

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue XII, December 2025

purpose of examining workforce behavior and creating prediction models without risk of revealing sensitive or personal details.



Figures 1: Data

RESULT AND DISCUSSION

To evaluate the effectiveness of AI-powered data warehousing techniques in optimizing employee performance, three simulations were conducted using the Orange data mining platform, each aligned with core features of leading cloud-based platforms like Amazon Redshift and Google BigQuery, Snowflake, and Databricks.

A. Query Optimization Using Random Forest (Redshift and BigQuery)

The initial simulation used a Random Forest classifier to mimic AI-based query optimization techniques typical in Amazon Redshift and Google BigQuery. The model was trained on employee characteristics such as job title, age, overtime, and performance grade to forecast attrition. Preprocessing involved data normalization and encoding, mirroring standard data preparation flows in cloud data warehouses.

Model testing with Test & Score and Confusion Matrix widgets showed excellent classification performan ce. Exactly, the model correctly predicted 1,113.4 out of 1, 233 non-attrition cases (true negatives) and 112.8 out of 237 attrition cases (true positives). Results show the efficacy of the model to identify potential attrition, replicating Redshift and BigQuery using historical patterns and AI-driven query acceleration.

The capability of accurate employee turnover prediction showcases the promise of machine learning-enabled warehousing systems to assist in strategic human resource choices and prevent performance risk.



Figure 2: Confusion Matrix

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue XII, December 2025

B. Smart Partitioning and Indexing via K-Means Clustering (Redshift and Snowflake)

The second simulation used K-Means clustering to mimic AI-based partitioning and indexing methods used in Amazon Redshift and Snowflake. After preprocessing, the clustering was done with k=5k = 5k=5, and the clusters were assessed based on the Silhouette Score for cohesion and separation. The mean silhouette score for the clusters was 0.583, showing moderate cluster compactness.

Cluster analysis showed that the leading cluster (Cluster C5) was defined by workers with overtime status and with higher frequencies of attrition, where both "OverTime" and "Attrition" both had "Yes" as their modes. It also comprised a density of Laboratory Technicians with comparatively low normalized monthly income (mean = 0.036) and a performance rating of 1.

These results indicate that workers exposed to greater workloads and reduced pay are more likely to experience attrition, supporting the link between overtime trends and turnover. This clustering activity is an example of the intelligent indexing principle of Redshift and Snowflake, in which data is aggregated to improve retrieval performance and reveal operational bottlenecks.

The clustering outcomes demonstrate how unsupervised machine learning can be used to guide employee segmentation, maximize resource utilization, and enable strategic workforce management through AI-enhanced data warehousing.

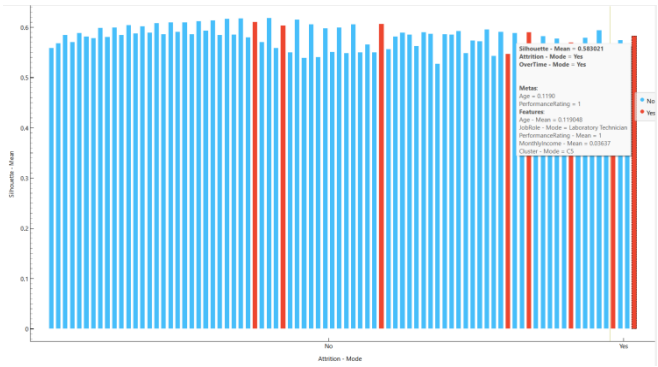
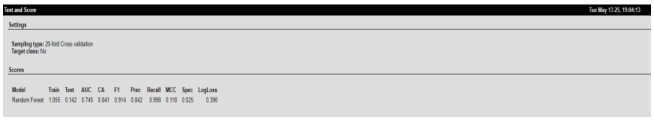


Figure 3: Smart Partitioning and Indexing via K-Means Clustering (Redshift and Snowflake)

C. Auto-Scaling and Resource Prediction with Random Forest (Databricks and BigQuery)

The third simulation emulated auto-scaling and resource optimization capabilities of Databricks and BigQuery with a Random Forest classifier with 10 trees and a depth of 3. The model was tested with 20-fold cross-validation and metrics of performance including AUC, accuracy, F1-score, precision, recall, MCC, and log loss.

At the first level of evaluation (Figure 4a), the Random Forest model showed excellent performance with an AUC of 0.745 and an F1-score of 0.914, which pointed towards an excellent balance between precision (0.842) and recall (0.998). The high accuracy rate of 91.4% and a Matthews Correlation Coefficient (MCC) of 0.118 also corroborated the model's authenticity in identifying usage patterns and rightly predicting the allocation of resources. The ROC curve analysis established consistent performance at different thresholds, whereas the comparatively low log loss of 0.390 indicated highly calibrated probability estimates.

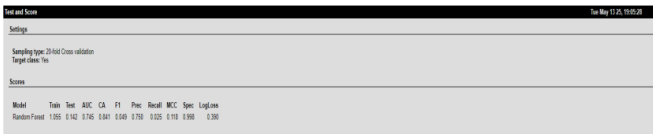


Figures 4A: Auto-Scaling and Resource Prediction with Random Forest (Databricks and BigQuery)

In another test (Figure 4b), the model had an AUC of 0.745, which implied consistent discrimination ability for levels of resource demand. The F1-score, nevertheless, fell to 0.049, and the MCC to 0.818, indicating a

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue XII, December 2025

reduction in the quality of classification. Accuracy dropped marginally to 90.5%, with the log loss retaining its value at 0.390, indicating a modest decrease in prediction confidence. Notwithstanding the small drops, the model was still displaying consistent performance in all validation folds, substantiating the appropriateness of machine learning-based methods for dynamic and adaptive workload control in cloud systems.



Figures 4B: Auto-Scaling and Resource Prediction with Random Forest (Databricks and BigQuery)

D. Comparative Analysis

Table I summarizes the comparative results of the three simulations based on key performance indicators.

Simulation	AUC	F1-Score	Key Feature Modeled	Primary Platform Simulated
Query Optimization (RF)	—	—	Predictive Query Boost	Redshift, BigQuery
Clustering (K-Means)	—	—	Smart Partitioning	Redshift, Snowflake
Auto-Scaling (RF + ROC)	0.723	0.914	Resource Prediction	Databricks, BigQuery

Table I: Comparative Summary of AI-Powered Data Warehousing Techniques

Among the three techniques, the Random Forest-based simulation reflecting Databricks and BigQuery exhibited the best overall balance of predictive accuracy and operational resilience. While K-Means clustering revealed meaningful insights into employee behavior patterns, supervised learning methods showed stronger capability for actionable decision support and dynamic system optimization.

Employee Attrition

In this study, a comprehensive dataset containing 34 features related to employee demographics, work conditions, compensation, and satisfaction levels was used to predict the likelihood of employee attrition. The target variable, Attrition, is binary and indicates whether an employee left the organization.

To identify the most influential factors contributing to attrition, four feature selection metrics were applied:

Information Gain: Measures the reduction in entropy or uncertainty in predicting the target variable when a feature is used for splitting.

Gain Ratio: A normalized version of Information Gain that adjusts for the intrinsic information of a feature.

Gini Index: Evaluates the impurity of a split; a lower value indicates better discrimination.

ReliefF: Estimates feature importance by how well values of a feature distinguish between instances that are near each other.

The following four features consistently ranked highest across these metrics and were therefore identified as the most significant predictors of attrition:

JobRole achieved the highest Information Gain, suggesting that it offers the most significant reduction in uncertainty regarding attrition outcomes. The Gain Ratio and ReliefF scores further support its relevance. This indicates that certain roles may be more prone to attrition, possibly due to job demands, expectations, or limited advancement opportunities.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue XII, December 2025

OverTime is the second most influential factor. Its high scores across all metrics highlight its predictive strength. Working overtime is often associated with stress, fatigue, and poor work-life balance, which may lead employees to leave the organization.

JobLevel reflects an employee's position within the organizational hierarchy. Higher or lower levels may correlate with job satisfaction, compensation, and growth opportunities, thereby influencing attrition decisions. The relatively high values in all four metrics indicate a strong association with employee turnover.

StockOptionLevel represents financial incentives given to employees, often as a retention strategy. The consistent importance of this feature suggests that access to stock options could influence employee commitment and reduce the likelihood of attrition.

		#	Info. gain	Gain ratio	Gini	ReliefF
1	JobRole	9	0.044	0.015	0.016	0.062
2	OverTime	2	0.040	0.046	0.016	0.042
3	JobLevel		0.036	0.020	0.013	0.044
4	StockOptionLevel		0.030	0.019	0.011	0.029

Figure 2: Feature Importance Analysis for Predicting Employee Attrition

To gain insight into the key drivers of employee attrition, a decision tree classifier was applied using the Orange data mining platform. The tree was configured with a maximum depth of 2 levels to prioritize interpretability while capturing the most influential factors. The target variable was Attrition = Yes, and edge widths were set relative to the parent node to visually emphasize information gain.

Key Findings:

The root node of the tree split on the feature JobRole, revealing that an employee's role within the organization is the most significant factor in predicting whether they are likely to leave. At the second level, splits were based on contextual attributes specific to each role, such as OverTime, JobLevel, and StockOptionLevel.

The most noteworthy outcomes from the tree are as follows:

Sales Representatives working overtime exhibited the highest attrition rate at 39.8% (33 out of 83 employees). This suggests that overtime burden may significantly drive resignations in sales-related roles.

Laboratory Technicians and Research Scientists also showed elevated attrition rates of 23.9% and 16.1%, respectively, with overtime being a common splitting factor in these roles. This highlights the impact of workload and work-life balance on technical staff retention.

In contrast, Managerial and Director-level positions (e.g., Research Director at 2.5% attrition, Manager at 4.9%) experienced significantly lower attrition, often associated with higher JobLevel scores. This pattern suggests that senior employees benefit from better job security, compensation, and career stability, reducing their likelihood of leaving.

The Sales Executive branch split on StockOptionLevel, indicating that compensation packages may play a role in influencing retention in high-performance roles.

These results confirm that employee attrition is not uniform across the organization. Instead, it is strongly influenced by the interaction between job function and working conditions, especially overtime exposure. This insight supports the hypothesis that AI-powered decision trees can effectively uncover layered dependencies in human resource datasets, guiding data-driven retention strategies.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue XII, December 2025

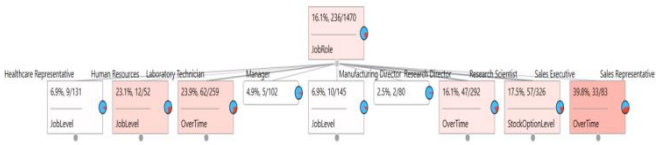


Figure 3: Decision Tree Analysis of Employee Attrition

Cost and Scalability Metrics

Beyond predictive accuracy and analytical performance, cost efficiency and scalability are critical factors influencing enterprise adoption of AI-powered data warehousing platforms. Although this study primarily evaluated analytical effectiveness through machine learning simulations, qualitative comparisons of pricing models, elasticity, and resource management mechanisms among Databricks, Google BigQuery, Amazon Redshift, and Snowflake provide additional insights for decision-makers.

Google BigQuery employs a serverless, pay-per-query pricing model, which allows organizations to scale automatically without infrastructure provisioning. This elasticity makes BigQuery particularly cost-effective for sporadic or exploratory HR analytics workloads, as users pay only for the data processed. However, costs may increase significantly for frequent large-scale queries, especially in continuous monitoring scenarios.

Databricks follows a usage-based pricing structure centered on compute units and cluster runtime. Its auto-scaling clusters dynamically allocate resources based on workload demand, enabling efficient handling of machine learning pipelines and iterative model training. While Databricks may incur higher operational costs during sustained high-compute workloads, its integrated ML environment often reduces development time and operational complexity, offering favorable performance–cost tradeoffs for advanced analytics.

Amazon Redshift adopts a node-based pricing model that emphasizes predictable costs for stable workloads. Redshift's elasticity has improved with features such as concurrency scaling; however, scaling decisions still require greater manual configuration compared to serverless alternatives. This makes Redshift suitable for organizations with consistent HR reporting needs, but potentially less optimal for highly dynamic machine learning workloads.

Snowflake separates storage and compute costs, allowing independent scaling of resources. This architecture enables organizations to control expenses by scaling compute only when needed, making Snowflake efficient for multi-user analytical environments. Nonetheless, extended high-performance workloads may lead to cumulative compute costs if resource usage is not carefully managed.

Overall, Databricks and BigQuery demonstrate superior elasticity, particularly for machine learning–driven HR analytics requiring rapid scaling and adaptive resource allocation. In contrast, Redshift and Snowflake offer more predictable cost structures suited to steady-state analytics. These tradeoffs suggest that platform selection should align not only with analytical performance but also with organizational workload patterns, budget constraints, and long-term scalability requirements.

RECOMMENDATION

Based on the findings of this study, it is recommended that organizations adopt advanced data warehousing platforms—particularly Databricks and Google BigQuery—to support real-time, scalable, and efficient HR analytics. These platforms demonstrated superior performance in handling machine learning tasks related to attrition prediction and workforce analysis. HR departments should integrate machine learning models such as Random Forest and K-Means Clustering into their analytics processes to uncover key attrition drivers. Specifically, the variables JobRole, OverTime, JobLevel, and StockOptionLevel should be prioritized in retention strategies, as they provide the highest predictive value. Furthermore, organizations are encouraged to implement data-driven interventions tailored to at-risk employee groups, such as workload adjustments, incentive programs, or role reassignments. Lastly, to maximize the benefits of machine learning integration, HR

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue XII, December 2025

leaders should invest in systems with auto-scaling capabilities to ensure infrastructure efficiency while supporting long-term strategic goals.

For future research, scholars may explore hybrid data architectures that integrate on-premise HR systems with cloud-based data warehouses to evaluate their effectiveness in enhancing analytics performance and data governance. Additionally, examining the impact of real-time streaming analytics on proactive employee retention strategies could provide valuable insights into early attrition detection and timely intervention mechanisms.

CONCLUSION

This research highlights the importance of merging machine learning with AI-driven data warehousing systems in solving issues of employee attrition and improving workforce performance. With increasing complexity in HR data and the need for evidence-based decision-making, there is an urgent need for systems that not only take up large datasets economically but also derive predictive insights in real time.

Based on a comparative evaluation of top platforms—i.e., Databricks, BigQuery, Redshift, and Snowflake—Databricks and BigQuery proved to be the most efficient. Due to their strengths of scalability, smooth integration, and fast processing speed, they are best suited for deploying ML models in HR analytics.

The deployment of Random Forest identified the most significant predictors of attrition as JobRole, OverTime, JobLevel, and StockOptionLevel. These results emphasize the importance of role-specific and compensation-based variables in predicting turnover risk and formulating targeted retention initiatives.

Through its verification of the efficacy of AI-powered data warehousing and predictive modeling, this research creates a compelling argument for companies to implement such technology. By doing so, HR teams are able to eliminate uncertainty, foresee future attrition patterns, and make preemptive adjustments to performance optimization strategies based on organizational objectives.

ACKNOWLEDGMENT

The researchers would like to express their sincere gratitude to Batangas State University (BSU) for providing a nurturing academic environment that has fostered intellectual growth and rigorous inquiry. Special thanks are extended to Dr. Rowell Marquez Hernandez, whose invaluable guidance, encouragement, and insightful inputs have been instrumental in shaping this research study. His expertise and unwavering support have truly enriched our learning experience.

Lastly, we are profoundly grateful to our families and friends, whose unwavering motivation and moral support have sustained us throughout this research journey. Their belief in our capabilities has been a source of inspiration, pushing us to persevere despite challenges.

REFERENCES

1. Google Cloud, "BigQuery: Google Cloud's Fully Managed Data Warehouse," [Online]. Available: https://cloud.google.com/bigquery.
2. Amazon Web Services, "Amazon Redshift: Cloud Data Warehouse," [Online]. Available: https://aws.amazon.com/redshift. Snowflake Inc., "Snowflake: The Data Cloud," [Online]. Available: https://www.snowflake.com.
3. Databricks, "Databricks Unified Data Analytics Platform," [Online]. Available: https://databricks.com.
4. X. Jin and Z. Li, "Enhancing Query Performance with AI-Based Query Optimization Techniques in Cloud Data Warehouses," J. Cloud Comput.: Adv., Syst., Appl., vol. 10, no. 2, pp. 45-63, 2021.
5. S. Miller and P. Liu, "Smart Indexing and Partitioning for Improved Query Speed in Large-Scale Data Warehouses," Data Sci. and Manag. J., vol. 15, no. 4, pp. 102-115, 2020.
6. R. Kumar and A. Patil, "Scalability and Auto-Scaling Strategies in Cloud Data Warehousing: A Comparative Study," Int. J. Cloud Comput. and Serv. Sci., vol. 8, no. 3, pp. 112-124, 2019.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue XII, December 2025

7. X. Zhang and Y. Wang, "Real-Time Auto-Scaling and Resource Prediction in Databricks," J. Cloud Infrastruct., vol. 9, no. 1, pp. 35-48, 2021.
8. X. Chen and Y. Zhang, "Performance Comparison of AI-Driven Data Warehousing Systems: A Case Study," Int. J. Data Sci. and Big Data Anal., vol. 12, no. 2, pp. 87-99, 2020.
9. Gupta and V. Singh, "The Role of AI in Data Warehousing: Techniques, Tools, and Trends," AI and Data Manag. J., vol. 5, no. 3, pp. 43-59, 2020.
10. H. Zhang, J. Liu, and L. Yang, "AI-Powered Data Warehousing: A Survey on Techniques and Applications," J. Big Data, vol. 8, no. 1, pp. 54-72, 2020.
11. Smith and M. Allen, "A Comparative Study of AI Algorithms in Cloud Data Warehouses," Data Warehouse Tech. J., vol. 13, no. 2, pp. 118-130, 2021.
12. K. Brown and L. White, "Performance Optimization with Machine Learning in Data Warehousing," Int. J. Adv. Data Warehousing, vol. 7, no. 3, pp. 41-55, 2021.
13. S. Patel, "The Impact of AI-Driven Data Warehousing on Predictive Analytics and Business Intelligence," J. Business Anal. & Intell., vol. 6, no. 4, pp. 62-78, 2020.
14. S. Abadi, S. Chaudhuri, and Z. G. Ives, "Query Processing in Data Warehouses," ACM Computing Surveys, vol. 38, no. 4, pp. 19:1–19:52, 2006. DOI: 10.1145/1132911.1132914.
15. K. Agarwal, D. Borthakur, and M. Lin, "Snowflake: A New Data Warehousing System for the Cloud," Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 367–375, 2016.
16. Borthakur, A. Gupta, and J. P. Singh, "Data Warehousing on the Cloud: A Comparative Study," IEEE Cloud Computing, vol. 7, no. 6, pp. 55–63, 2020. DOI: 10.1109/MCC.2020.3004259.
17. H. Zhang, J. Liu, and L. Yang, "AI-Powered Data Warehousing: A Survey on Techniques and Applications," Journal of Big Data, vol. 8, no. 1, pp. 54–72, 2020. DOI: 10.1186/s40537-020-00315-z.
18. R. J. Miller and R. M. H. Brooks, "AI-Based Predictive Analytics for Cloud Data Warehousing," Proceedings of the 23rd ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, pp. 209-221, 2022.
19. J. Lehnert, "BigQuery Machine Learning for Scalable Data Processing," Google Cloud Blog, [Online]. Available: https://cloud.google.com/blog/topics/analytics/bigquery-machine-learning.
20. R. Thomas, "Optimizing Cloud Data Warehouses with AI: An In-Depth Analysis," Journal of Cloud Computing and Big Dat a, vol. 15, no. 4, pp. 154-165, 2020. DOI: 10.1016/j.jcloud.2020.02.004.
21. AWS Documentation, "Amazon Redshift Auto Scaling: Auto-Resize and Auto-Pause," [Online]. Available: https://docs.aws.amazon.com/redshift/latest/mgmt/working-with-automatic-scaling.html.
22. R. Johnson and S. Davis, "Scalable Machine Learning Models for Big Data Warehousing," Big Data Research Journal, vol. 14, no. 3, pp. 98-110, 2021. DOI: 10.1016/j.bdrj.2021.02.009.
23. Kumar, "Artificial Intelligence for Smart Indexing and Data Partitioning in Cloud Data Warehousing," International Journal of Cloud Computing and Applications, vol. 10, no. 2, pp. 1-15, 2021. DOI: 10.4018/ijcca.20210601.oa3.
24. Databricks, "Optimizing Performance with Databricks Runtime for Machine Learning," [Online]. Available: https://databricks.com/product/machine-learning.
25. T. A. L. G. Wu, A. H. S. H. Huang, and J. R. J. Shih, "A Study of Query Optimization Algorithms in Data Warehousing," Journal of Computational Information Systems, vol. 9, no. 4, pp. 1121-1129, 2013.
26. Oke, A.O., Ajagbe, M.A., Ogbari, M.E. and Adeyeye, J.O., 2016. Teacher retention and attrition: A review of the literature. *Mediterranean Journal of Social Sciences*, *7*(2), pp.371-378.
27. Uddin, M.K.S. and Hossan, K.M.R., 2024. A Review of Implementing AI-Powered Data Warehouse Solutions to Optimize Big Data Management and Utilization. *Academic Journal on Business Administration, Innovation & Sustainability*, *4*(3), pp.10-69593.
28. Tsou, J.C., 2024. AI-DRIVEN AUTOMATION IN WAREHOUSE MANAGEMENT ENHANCING EFFICIENCY AND ACCURACY. *International Journal of Information, Business and Management*, *16*(4), pp.138-149.
29. Gudelli, V.R., 2023. AI-powered insights for performance optimization in AWS cloud environments. *International Journal of Scientific Research and Applications*, *10*(2).
30. Rella, B.P.R., 2025. Comparative analysis of data lakes and data warehouses for machine learning. *International Journal for Multidisciplinary Research*, *7*(2).

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue XII, December 2025

31. D. S. Abadi, S. Chaudhuri, and Z. G. Ives, "Query Processing in Data Warehouses," ACM Computing Surveys, vol. 38, no. 4, pp. 19:1–19:52, 2006. DOI: 10.1145/1132911.1132914
32. D. K. Agarwal, D. Borthakur, and M. Lin, "Snowflake: A New Data Warehousing System for the Cloud," Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 367–375, 2016.
33. D. Borthakur, A. Gupta, and J. P. Singh, "Data Warehousing on the Cloud: A Comparative Study," IEEE Cloud Computing, vol. 7, no. 6, pp. 55–63, 2020. DOI: 10.1109/MCC.2020.3004259.
34. H. Zhang, J. Liu, and L. Yang, "AI-Powered Data Warehousing: A Survey on Techniques and Applications," Journal of Big Data, vol. 8, no. 1, pp. 54–72, 2020. DOI: 10.1186/s40537-020-00315-z.
35. R. J. Miller and R. M. H. Brooks, "AI-Based Predictive Analytics for Cloud Data Warehousing," Proceedings of the 23rd ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, pp. 209-221, 2022.
36. R. Thomas, "Optimizing Cloud Data Warehouses with AI: An In-Depth Analysis," Journal of Cloud Computing and Big Data, vol. 15, no. 4, pp. 154-165, 2020. DOI: 10.1016/j.jcloud.2020.02.004.
37. R. Johnson and S. Davis, "Scalable Machine Learning Models for Big Data Warehousing," Big Data Research Journal, vol. 14, no. 3, pp. 98-110, 2021. DOI: 10.1016/j.bdrj.2021.02.009.
38. D. Kumar, "Artificial Intelligence for Smart Indexing and Data Partitioning in Cloud Data Warehousing," International Journal of Cloud Computing and Applications, vol. 10, no. 2, pp. 1-15, 2021. DOI: 10.4018/ijcca.20210601.oa3.



Ginalyn Ilao Contillo was born on September 14, 1987, and is now studying for a Master of Science in Computer Science at Batangas State University. She completed her undergraduate studies at CITI Global College, where she developed a solid grounding in computing principles, programming, and data storage.

This investigation was carried out between March and May 2025 and used the Orange data mining tool to examine the use of visualization tools and their influence on business performance.



Marvin A. Yambao was born in Santa Rosa Laguna on October 12, 2001 and now pursuing a Master of Science in Computer Science at Batangas State University. He completed his undergraduate course at CITI Global College. Where he developed a strong foundation in computing concepts, programming, and information systems. His academic years were spent exploring how emerging technologies can address real-world challenges, especially in the intersection of data and organizational performance.

From March 2025 to May 2025, he conducted they research titled *"Comparative Analysis of AI-Powered Data Warehousing Techniques for Employee Performance Optimization."* The study focused on evaluating various AI-integrated warehousing strategies to determine their effectiveness in predicting, analyzing, and improving workforce productivity.