DESIGN AND EVALUATION OF AN AI-ASSISTED DIGITAL KPI INFORMATION SYSTEM FOR EMPLOYEE PERFORMANCE MONITORING AND RECOMMENDATION

Asnefi[1*], Essy Malays Sari Sakti[2], Dayanti[3], Irwan Syarif[4]
Prodi Teknik informatika, Fakultas Teknik dan Informatika, Universitas Patria Artha, Indonesia[134]
Prodi Informatika, Fakultas Teknik, Universitas Persada Indonesia YAI, Indonesia[2]
asnefi@patria-artha.ac.id[1], emalays67@gmail.com[2], dayanti.fattah@gmail.com[3], firaysnawri88@gmail.com[4]

Received: 30 April 2026, Revised: 21 May 2026, Accepted: 23 May 2026
**Corresponding Author*

ABSTRACT
Manual and fragmented KPI-based evaluation often weakens employee performance monitoring by causing delays, input errors, and inconsistent interpretation. This study addresses that problem by designing and evaluating an AI-assisted digital KPI information system that integrates competency assessment, digital KPI records, performance scoring, dashboard-based monitoring, and recommendation support in a single environment. The study contributes by transforming the conventional competency–KPI–performance framework into an operational decision-support artifact with embedded AI classification. Using a prototype-based approach, the system was evaluated with a structured dataset of 340 employee-year records from 2021–2025. Three machine learning models were compared for classifying employee performance into High, Moderate, and Low categories. The results show that the system is functionally feasible and usable for KPI-based performance monitoring, while Random Forest achieved the best classification performance with 0.9853 accuracy and 0.9852 F1-score. The findings indicate that the proposed system can improve the structure of digital KPI monitoring and provide AI-assisted support for managerial review and follow-up actions. The study contributes theoretically by extending KPI-based performance management into an intelligent information system context and practically by offering a feasible model for organizations operating under limited implementation conditions.
Keywords: *Digital KPI, Employee Performance, Information System, Artificial Intelligence, Recommendation System*

1. Introduction

Digital transformation has fundamentally changed how organizations collect, process, and use workforce data. In human resource management, this shift is especially visible in the transition from manual and fragmented performance appraisal practices toward digital, data-driven, and analytics-supported systems. Recent studies show that digital HRM is no longer limited to administrative automation; it increasingly incorporates analytics, decision support, and intelligent augmentation for managerial work. Recent reviews also indicate that AI adoption in HRM has accelerated because organizations expect higher efficiency, better data processing, and more structured decision support across core functions, including performance management (Ali et al., 2026; Bujold et al., 2024; Gupta et al., 2024; Madanchian, 2024; Nguyen & Elbanna, 2025; Wang et al., 2025).

Performance management remains one of the most critical HR functions because it links employee behavior and outcomes with organizational goals. A well-designed performance management system helps organizations define expectations, monitor progress, provide feedback, and improve productivity. Recent empirical evidence shows that effective performance management systems have a significant positive association with employee productivity and work outcomes, especially when they include clear responsibilities, continuous feedback, evaluation, and development support. At the same time, broader work on performance measurement systems confirms that monitoring and measurement are central to organizational improvement and sustainability (Cunha et al., 2023; Siraj & Hágen, 2023). In parallel, recent evidence also suggests that HR information systems increasingly influence performance management practices by improving data accessibility, reporting consistency, and managerial visibility (Raja et al., 2025).

However, many organizations still struggle with KPI-based evaluation processes that are manual, periodic, and fragmented. In the focal organizational context of this study, preliminary internal evidence shows that KPI data collection is still performed manually, creating delays, typing errors, and interpretive inconsistencies that may reduce the accuracy of managerial analysis. The same internal document also reports a decline in the organization's employee performance score from 78.5 in 2022 to 70.0 in 2023 and 63.5 in 2024, suggesting that the existing evaluation process does not yet provide sufficiently responsive and corrective managerial insight.

This limitation is important because the value of digital performance management lies not only in storing KPI records, but also in transforming dispersed data into timely and actionable managerial knowledge. Recent research on data-driven decision-making shows that top management support, data quality, and analytical culture are critical enablers of effective data use in managerial processes. In other words, digital systems do not automatically improve decisions unless they are designed to produce reliable information, analytical visibility, and decision-support outputs that managers can actually use (Diefenhardt et al., 2025; Szukits & Móricz, 2024).

Artificial intelligence offers a promising way to strengthen performance information systems beyond descriptive reporting. Recent scholarly work argues that AI can support real-time feedback, objective evaluation, predictive analytics, and continuous monitoring in HR contexts, while also enabling human–AI augmentation in workplace decision processes. Reviews of AI-enabled HRM and AI-supported decision-making consistently highlight performance management as one of the domains in which AI can add value through classification, prediction, recommendation, and pattern recognition (Bastida et al., 2025; Căvescu & Popescu, 2025; Gupta et al., 2024; Madanchian, 2024; Nguyen & Elbanna, 2025).

At the same time, the literature makes clear that AI in HRM must be implemented carefully. Responsible-AI research warns that HR applications of AI involve risks related to bias, opacity, privacy, and over-automation, especially when AI is used in evaluative or managerial contexts. Recent reviews therefore emphasize that AI should augment rather than replace managerial judgment, and that responsible implementation requires attention to transparency, fairness, and data quality. This is particularly relevant in employee performance contexts, where inaccurate or poorly governed data can undermine trust in both the system and its outputs (Bujold et al., 2024; Madanchian, 2024; Naoum et al., 2026; Nguyen & Elbanna, 2025; Shin et al., 2025).

Another important trend in recent literature is the rise of HR analytics and people analytics as a strategic capability. Reviews published in recent years show that HR analytics helps organizations define relevant variables, collect and analyze employee-related data, and communicate findings for strategic action. This shift is particularly useful in performance management because KPI achievement, competency assessment, and historical work outcomes can be integrated into a structured analytical model. In this sense, AI-assisted KPI systems represent a practical intersection between HR analytics, digital information systems, and managerial decision support (Belizón et al., 2024; Cho et al., 2023; Diefenhardt et al., 2025; Ramachandran et al., 2024; Szukits & Móricz, 2024).

Despite these developments, an important practical gap remains. Much of the recent literature discusses AI in HRM at a conceptual, review, or broad strategic level, while relatively fewer studies translate competency assessment, digital KPI monitoring, and employee performance evaluation into a single operational information system that includes predictive or classificatory AI support. Recent reviews on human–AI augmentation and AI in people management also note that the field remains fragmented and that empirical, context-specific implementations are still limited (Bauwens & Batistič, 2025; Bujold et al., 2024; Gupta et al., 2024; Nguyen & Elbanna, 2025; Rana & Kumar, 2026). In addition, many studies focus either on predictive modeling or on digital system design, but do not evaluate both the information system artifact and the embedded AI component in an integrated manner. As a result, there is still limited empirical evidence on how KPI-based performance management can be operationalized as an AI-assisted decision-support system under realistic implementation constraints.

AI was selected in this study because the organizational problem was not limited to data storage inefficiency, but also involved delayed interpretation and insufficiently responsive follow-up. A conventional digital reporting system could improve record organization, but it would still rely heavily on manual interpretation. By contrast, AI-based classification enables the system to identify performance categories more consistently and to provide recommendation-oriented outputs that directly support managerial review. Therefore, AI was adopted not as a replacement for managerial judgment, but as an analytical augmentation mechanism that addresses the specific limitation of delayed and fragmented KPI interpretation.

This study offers several distinct contributions compared with previous research. First, it transforms the conventional competency KPI performance framework into an AI-assisted digital information system for employee performance monitoring. Second, the study integrates dashboard-based KPI monitoring, AI based classification, and recommendation support within a single platform, extending the role of KPI systems from descriptive reporting toward decision-support functions. Third, the research evaluates both the information system and the embedded AI models simultaneously, showing how machine learning can be operationalized within a practical KPI management environment. Finally, the study provides a feasible prototype model for organizations that still rely on manual and fragmented performance evaluation processes.

By doing so, the study positions itself at the intersection of performance management systems, HR analytics, and AI-assisted decision support, while also showing how digital KPI management can be operationalized as an intelligent decision-support mechanism under realistic implementation constraints (Maake & Schultz, 2025; Madanchian, 2024; Raja et al., 2025; Siraj & Hágen, 2023; Szukits & Móricz, 2024). The study addresses four research questions: (1) how a digital KPI information system can be designed to support integrated employee performance monitoring; (2) how AI can be embedded into the system to classify employee performance status; (3) to what extent the proposed system satisfies functional and usability requirements; and (4) which machine-learning method is most appropriate for the system environment.

2. Literature Review

2.1 Information Systems for Performance Management

Information systems for performance management are designed to transform dispersed employee and organizational data into structured information that supports monitoring, evaluation, and decision-making . Recent studies emphasize that performance management systems are not merely administrative tools; they shape employee outcomes, managerial visibility, and organizational learning. Evidence from recent empirical and review studies shows that well-structured performance management systems are associated with better productivity, stronger work engagement, and improved alignment between employee activity and organizational objectives. Recent review work on SMEs and public-sector-like settings also notes that the value of performance systems increases when they are integrated with feedback, measurement discipline, and contextual adaptation rather than used only for routine appraisal (Ehmann et al., 2024; Hendri, 2025; Siraj & Hágen, 2023).

From an information systems perspective, digital performance management expands the function of a traditional appraisal process into a data environment that supports continuous access, traceability, and comparative analysis. This shift is especially important where KPI data are still manually collected or fragmented across documents, because the absence of integration reduces timeliness and decision quality. Recent studies on data-driven decision-making show that analytical value depends not only on data availability, but also on the organization’s ability to structure, share, and interpret data consistently. As a result, a digital KPI information system should be understood as both a data management mechanism and a decision-support architecture (Hendri, 2025; Siraj & Hágen, 2023; Tessema et al., 2025). In addition, recent evidence on HR information systems suggests that performance management becomes more effective when digital systems reduce fragmentation, increase data accessibility, and improve the consistency of managerial review processes. This is particularly relevant for KPI-based environments, where system quality directly affects the timeliness and usefulness of performance information (Raja et al., 2025).

2.2 Competency and Employee Performance

Competency remains a central construct in employee performance management because it captures the individual capacities that enable effective work execution. In current HR and performance literature, competency is commonly operationalized through dimensions such as knowledge, skill, motivation, and self-related capability, all of which influence whether employees can translate organizational targets into work outcomes. Recent literature on strategic human resource management in the digital era continues to affirm that employee capability development is a foundational condition for sustainable performance improvement, particularly when organizations are simultaneously adopting new technologies and more data-intensive management systems (Kassa & Worku, 2025; Nastase et al., 2025).

At the same time, contemporary performance management research shows that employee performance should not be treated as a single outcome observed only at the end of a period. Instead, it is increasingly understood as a multidimensional construct involving output quality, quantity, efficiency, and behavioral engagement within a monitored work process. This view supports the integration of competency and performance in the same information system, because competency indicators can function as leading signals, while performance scores reflect lagging results. Recent work on performance management and work engagement reinforces this logic by showing that structured target-setting and evaluation practices influence not only measured outcomes but also employees' psychological involvement in work (Akanpaadgi et al., 2024; Ehmann et al., 2024).

2.3 Digital KPI Systems as Structured Performance Instruments

Key Performance Indicators are widely recognized as measurable markers of progress toward individual or organizational targets. In recent studies, digital KPI systems are increasingly discussed as mechanisms for improving transparency, accountability, and comparability across time and units. When KPI indicators are digitized, they can be tracked more consistently and integrated with broader analytical processes, which is particularly useful in organizations that need timely managerial intervention. Recent performance-management literature highlights that systems become more effective when targets are measurable, regularly reviewed, and embedded into a broader performance architecture rather than treated as isolated metrics (Ehmann et al., 2024; Hendri, 2025; Siraj & Hágen, 2023).

However, KPI systems also face several design challenges. Recent review studies show that poorly aligned indicators, excessive formalism, and lack of contextual interpretation can reduce the usefulness of performance systems. This means that a digital KPI platform should not only store scores, but also enable interpretation, categorization, and follow-up action. In this regard, integrating KPI data with competency records and performance history can improve analytical depth, allowing the system to go beyond descriptive reporting and move toward actionable performance intelligence (Akanpaadgi et al., 2024; Hendri, 2025).

Recent research also indicates that the value of digital KPI systems depends on how well they are connected to broader performance-management processes rather than treated as isolated reporting tools. In this sense, digital KPI systems are most useful when they support not only data entry and monitoring, but also interpretation and follow-up action through integrated managerial workflows (Raja et al., 2025; Varma et al., 2024).

2.4 AI in Human Resource and Performance Information Systems

The application of artificial intelligence in human resource management has expanded rapidly over the last five years. Recent review and empirical literature shows that AI is increasingly used across HR domains such as recruitment, employee analytics, retention, and performance management. In performance-related contexts, AI is especially relevant because it can identify patterns, classify employee status, generate predictive outputs, and support faster interpretation of recurring data structures. Recent articles specifically discussing AI and performance management argue that AI can strengthen appraisal systems by improving analytical consistency, feedback timing, and the capacity to detect emerging performance issues (Ali et al., 2026; Căvescu & Popescu, 2025; Úbeda-García et al., 2025; Varma et al., 2024; Wang et al., 2025).

Another reason AI is relevant in KPI-based systems is its compatibility with structured tabular data. Competency scores, KPI achievement values, historical ratings, and departmental attributes can be arranged into machine-learning-ready inputs. Recent studies on predictive analytics in HRM and employee performance classification show that classification-oriented models can support HR decisions by distinguishing between different categories of employee outcomes. Although the sophistication of these models varies, the practical implication is consistent: AI can help transform static HR records into early-warning and recommendation-oriented outputs (Bastida et al., 2025; Căvescu & Popescu, 2025; Vijh et al., 2025).

Recent broader work on AI-driven organizational performance also suggests that AI can contribute to higher organizational effectiveness when its adoption is accompanied by supportive managerial and data conditions. However, these gains are not automatic. The literature repeatedly shows that AI produces value when it is embedded in useful processes, supported by good-quality data, and connected to real managerial workflows. Therefore, in an employee performance information system, AI should be positioned as an analytical augmentation layer rather than a standalone technological feature (Căvescu & Popescu, 2025; Kassa & Worku, 2025).

2.5 HR Analytics and Data-Driven Decision Support

HR analytics has become one of the most important bridges between digital HR systems and strategic decision-making. Recent review work defines HR analytics as a process involving problem definition, data collection, analysis, communication, and evidence-based action. This perspective is highly relevant to KPI and performance information systems because it justifies why employee-related data should be stored and processed in a structured way. In other words, HR analytics provides the conceptual foundation for converting competency scores and KPI achievement into managerial intelligence (Cho et al., 2023; Tessema et al., 2025).

Recent studies further suggest that HR analytics creates strategic value when organizations move beyond descriptive HR reporting and adopt more structured discovery and interpretation processes. In practice, this means that HR data become more useful when they are embedded into a knowledge-discovery workflow and linked to higher-level managerial recognition of HR's strategic role (Belizón et al., 2024; Diefenhardt et al., 2025).

Recent studies also stress that the usefulness of analytics depends on organizational conditions such as data quality, analytical culture, and management support. If data are inconsistent or delayed, predictive and recommendation functions become less trustworthy. This observation is particularly important for KPI environments that historically relied on manual collection. A digital KPI information system with AI support can therefore be seen as a practical mechanism for improving the readiness of employee-performance data for analytics and decision support (Tessema et al., 2025).

2.6 Responsible and Trustworthy AI in HR Contexts

Although AI offers strong analytical benefits, recent literature also warns about risks in HR applications. Reviews on responsible AI in HRM consistently identify concerns related to fairness, transparency, privacy, explainability, and organizational trust. These concerns are especially important in employee-facing systems, because performance classification or recommendation may influence managerial judgment, employee development, and perceptions of procedural justice. More recent discussions also highlight the dual impact of AI in people management, including the potential for employee resistance, perceptions of dehumanization, and challenges to diversity, equity, and inclusion if AI systems are applied without sufficient oversight. Consequently, current scholarship strongly recommends that AI in HRM should augment human judgment, not replace it (Batool et al., 2025; Bujold et al., 2024; Naoum et al., 2026; Shin et al., 2025).

For the present study, this literature has two implications. First, the AI component should be used to support classification and recommendation rather than fully automate final personnel decisions. Second, the information system should preserve managerial oversight and interpretability, for example by displaying predicted categories together with supporting KPI and competency information. This approach is consistent with recent responsible-AI and workplace-augmentation literature, which argues that the strongest practical value of AI often lies in human–

AI collaboration rather than autonomous replacement of managerial roles (Batool et al., 2025; Bujold et al., 2024).

2.7 Research Gap and Study Positioning

The recent literature clearly supports three separate streams: performance management systems improve monitoring and employee outcomes; HR analytics improves evidence-based people management; and AI can strengthen HR processes through prediction and classification. However, these streams are still often studied separately. Many recent publications discuss AI in HRM at a conceptual, review, or bibliometric level, while relatively fewer studies operationalize competency assessment, digital KPI monitoring, and employee performance classification within a single information system artifact. Likewise, studies on performance management frequently evaluate managerial outcomes without detailing how an integrated digital architecture can support those outcomes in practice (Bauwens & Batistič, 2025; Bujold et al., 2024; Cho et al., 2023; Hendri, 2025; Rana & Kumar, 2026; Varma et al., 2024).

This gap creates space for the present study. The study positions itself at the intersection of performance management systems, HR analytics, and AI-assisted information systems. Rather than examining only the statistical relationship among competency, KPI achievement, and performance, the study develops a prototype-based digital KPI information system that integrates these components into one operational environment. The system is intended to support monitoring, classification, and recommendation, while remaining realistic under limited implementation conditions. In that sense, this study contributes by translating contemporary HR analytics and AI ideas into a feasible system design for employee performance management (Căvescu & Popescu, 2025; Cho et al., 2023; Siraj & Hágen, 2023; Varma et al., 2024).

3. Research Methods

The methodological transparency of this study was strengthened through clearer explanations of the dataset, preprocessing stages, model selection, and AI evaluation procedures. The revised manuscript now explains the use of 340 employee-year records, preprocessing steps such as data checking and encoding, and the rationale for selecting Logistic Regression, Random Forest, and XGBoost. The study also clarifies the use of an 80:20 stratified train–test split and several overfitting prevention measures, including moderate hyperparameter settings and held-out testing evaluation. In addition, the usability and system testing procedures, including black box testing, limited white box testing, SUS evaluation, and performance-efficiency testing, are described more systematically to improve methodological rigor and transparency.

3.1 Research Design

This study adopted a prototype-based information system development approach combined with a limited quantitative evaluation. The prototype approach was selected because the primary objective of the study was to design and validate the feasibility of an AI-assisted digital KPI information system rather than to deploy a full enterprise-scale application. Recent information-system development studies continue to use prototyping when fast iteration, feature refinement, and user-oriented feasibility are more important than full-scale production implementation. In such settings, prototyping is considered appropriate because it allows researchers to translate organizational needs into an operational artifact, gather early feedback, and revise the design before broader deployment (Sihombing, 2024).

The study was also positioned as a limited-evaluation prototype study. This means that the developed system was evaluated for its core feasibility through functional testing, usability assessment, and AI model performance, without claiming organization-wide operational deployment. This positioning is methodologically reasonable when implementation time, system scope, and access to organizational infrastructure are limited, provided that the study clearly states its prototype nature and evaluation boundaries (Sihombing, 2024).

3.2 Research Procedure

The research procedure consisted of six main stages. First, the study identified the main performance-management problem in the organizational context, namely manual KPI recording, delayed monitoring, and inconsistent interpretation of employee performance information. Second, system requirements were derived from the competency, digital KPI, and employee performance framework used in the original research model. Third, a prototype architecture was designed, including user roles, database structure, data flow, and interface layout. Fourth, the prototype was developed with only the core modules required for the study. Fifth, an AI component was added to classify employee performance status. Sixth, the prototype was evaluated through software testing, usability testing, and AI model comparison. This staged structure is aligned with recent prototype-oriented information-system studies that emphasize requirement identification, iterative design, implementation, and limited but focused evaluation (Sihombing, 2024).

3.3 Research Setting and Scope

The study focused on a digital KPI-based employee performance context derived from the organizational problem described in the underlying proposal. However, because the objective of this article was to produce an academically valid prototype rather than a full organizational rollout, the system scope was intentionally restricted to the following core functions: user authentication, KPI data entry, competency scoring, performance summary generation, AI-based performance classification, recommendation output, and report generation. Limiting the scope to the most essential modules is consistent with prototype-based development logic, where the aim is to demonstrate functional feasibility and analytical value rather than complete enterprise implementation (Sihombing, 2024).

3.4 Variables and Data Structure

The domain variables in this study were derived from the existing competency–digital KPI–employee performance framework. Competency was represented through dimensions such as self-concept, knowledge, skill, and work motivation. Digital KPI achievement was represented through the widely used target characteristics of being specific, measurable, achievable, relevant, and timely. Employee performance was represented through quantity, quality, efficiency, and effectiveness. These variables were transformed into structured system inputs so that they could function not only as conceptual constructs, but also as operational data fields for the prototype. The AI module used these variables as predictive features for classifying performance status. This conversion from theoretical variables into system-ready structured inputs is consistent with contemporary HR analytics logic, where people-management constructs are translated into analyzable variables for decision support (Jain, 2025; Mourad et al., 2024).

3.5 Data Source and Limited Data Strategy

Because the study was conducted under time and implementation constraints, the dataset used for system evaluation was designed as a structured evaluation dataset. The data structure was compiled from available organizational performance logic, questionnaire-based indicators, and supplementary structured simulated records generated according to the same conceptual framework. This strategy was adopted to ensure that the proposed system could still be evaluated at the functional, visual, and analytical levels even without access to a large production database (Jain, 2025; Mourad et al., 2024; Thanri et al., 2025).

The final dataset consisted of 340 employee-year records covering the period 2021–2025, with an average of 68 employee records per year. Each record represented one employee observation in one year and included competency indicators, digital KPI indicators, employee performance indicators, aggregated scores, and a final performance label. This structure enabled the dataset to serve two purposes simultaneously: first, as the operational data source for dashboard visualization and employee detail display; and second, as the analytical dataset for machine-learning-based classification.

The use of structured simulated data was considered acceptable in this study because the main objective was not to produce broad population inference, but to evaluate artifact feasibility, test data flow consistency, and compare the performance of candidate AI models within a controlled prototype environment. To maintain methodological transparency, these data should therefore be interpreted as evaluation-oriented data rather than population-representative organizational data. Consequently, the empirical findings of this study are intended to support prototype feasibility, system coherence, and controlled model comparison, rather than broad statistical generalization across organizations.

In addition, the dataset was intentionally aligned with the documented organizational problem context, particularly the declining performance pattern observed in the internal evidence. This alignment was necessary to ensure consistency between the system design, the analytical objective, and the practical problem motivating the research. However, such alignment also implies that the dataset may represent a more structured and controlled pattern than naturally occurring operational data. Therefore, the findings should be interpreted within the scope of a prototype-level system evaluation.

3.6 System Design

The proposed system involved three user roles: admin/HR, supervisor/manager, and employee. Admin or HR users were responsible for managing user data, KPI indicators, competency indicators, and assessment records. Supervisors or managers were responsible for reviewing performance summaries, observing AI-based performance categories, and receiving recommendation outputs. Employees were allowed to access individual KPI and performance information relevant to their role.

The system architecture contained four layers: a presentation layer, an application layer, an analytics layer, and a database layer. The presentation layer handled the interface for dashboards, forms, and reports. The application layer processed KPI data entry, competency assessment, and performance scoring logic. The analytics layer contained the AI classification component and a simple rule-based recommendation mechanism. The database layer stored users, employee profiles, KPI targets, KPI achievements, competency scores, performance summaries, prediction results, and recommendation records. This layered architecture is consistent with prototype-oriented information-system development in recent studies, where the prototype must remain simple enough to build quickly while still demonstrating functional integration across modules (Sihombing, 2024).

3.7 AI Model Development

The AI component of the study was designed to classify employee performance into predefined categories, namely High, Moderate, and Low performance. A classification target was selected instead of raw-score prediction because categorical output is more practical for managerial interpretation and recommendation support, especially under limited data conditions. Classification labels were derived from the aggregated performance score so that the model output could be integrated directly into the dashboard and recommendation modules (Jain, 2025; Mourad et al., 2024; Thanri et al., 2025).

Three machine-learning algorithms were compared in this study: Logistic Regression, Random Forest, and XGBoost. These models were selected for complementary methodological reasons. Logistic Regression was used as a baseline model because it is widely recognized, relatively interpretable, and appropriate for structured classification tasks. Random Forest was selected because it is capable of handling non-linear relationships, is relatively robust on tabular data, and often performs well when the predictors contain interactions and heterogeneous patterns. XGBoost was included because it is a strong ensemble boosting method that has been widely used in classification problems involving structured data and often provides competitive predictive performance.

Before model training, the dataset was preprocessed through several stages. First, the data were checked for completeness and numerical consistency. Second, categorical variables were encoded when necessary so that they could be processed by the learning algorithms. Third, numerical variables were reviewed to ensure that the scale of input features remained consistent. Fourth, the target-label distribution was examined to confirm that the three classes were sufficiently represented for classification analysis. This preprocessing stage was necessary to improve data quality and to reduce the risk of unstable model behavior during training and evaluation.

To evaluate the models, the dataset was divided into training and testing subsets using an 80:20 split with stratification. Stratified splitting was used to preserve the class distribution of the target variable across both subsets and thereby reduce imbalance-related bias during model comparison. This design was selected because it provides a reasonable balance between training adequacy and testing reliability under limited data conditions. The testing subset was held out from model training and used exclusively for comparative model evaluation.

To reduce the risk of overfitting, the study adopted several precautions. Hyperparameter settings were kept at moderate levels rather than aggressively optimized, so that model comparison remained stable and reproducible within the prototype context. The selected feature set was limited to structured indicators that were conceptually relevant to competency, KPI, and employee performance. In addition, the model evaluation relied on a held-out test set rather than on training performance alone. These steps were intended to ensure that the reported results reflected comparative generalization performance within the limited evaluation environment.

The final feature set included competency scores, digital KPI scores, performance-related indicators, and supporting employee attributes. These structured variables were chosen because they represent the same information that would be available within the proposed information system environment. Accordingly, the AI component was not treated as a detached predictive experiment, but as an analytical layer embedded within the broader digital KPI information system (Jain, 2025; Mourad et al., 2024).

3.8 Recommendation Logic

In addition to AI-based classification, the prototype included a simple rule-based recommendation module. This module was designed to convert classification outputs into practical managerial suggestions. For example, employees predicted to be in a low-performance category could be associated with recommendations such as additional supervision, training priority, or KPI review. Employees with moderate performance could receive suggestions related to coaching or competency strengthening, while high-performing employees could be flagged for retention, recognition, or advanced development planning. The reason for combining AI classification with rule-based recommendation was to make the prototype more useful for decision support and to ensure that model output could be translated into understandable managerial action. This design is in line with recent literature arguing that AI in HR and workplace settings delivers greater value when combined with actionable augmentation rather than presented as a purely technical output (Jain, 2025; Mourad et al., 2024).

3.9 System Evaluation

System evaluation was carried out in three parts: functional testing, usability testing, and software quality-oriented review.

First, black box testing was used to examine whether each main function worked according to its expected input-output behavior. The tested functions included login, employee data entry, KPI target input, KPI achievement input, competency scoring, performance summary display, AI classification display, recommendation generation, and report export. Black box testing was selected because it is widely used to verify whether a prototype satisfies its functional requirements at the user level.

Second, limited white box testing was conducted on selected internal logic modules, especially the score calculation flow, AI invocation process, and recommendation-routing logic. Because the study did not develop a full production system, white box testing was limited to the core code paths that directly affected analytical correctness.

Third, the prototype was reviewed using a software-quality perspective, especially functional suitability and performance efficiency. Recent software quality studies continue to use ISO/IEC 25010 as a relevant reference for evaluating software systems, including systems that are still in the development or pilot phase. For prototype-level research, selecting only the most relevant quality aspects is considered acceptable and practical (Ariningsih & Muhammad, 2024; Schaffernak et al., 2025; Widiastuti et al., 2025). For performance efficiency testing, the response time of the main system functions was measured in a local prototype environment. Each function was executed repeatedly under the same test condition, and the average response time was used as the evaluation value. In this study, response times below 2 seconds were treated as acceptable for prototype-level interaction, since the objective was to confirm operational responsiveness rather than to conduct enterprise-scale load benchmarking.

3.10 Usability Evaluation

Usability was evaluated using the System Usability Scale (SUS). SUS was chosen because it is simple, efficient, and widely recognized as a reliable instrument for evaluating the usability of interactive systems. Recent studies continue to use SUS for information-system assessment because it provides a practical quantitative indication of whether users find a system understandable and usable, even in limited-sample evaluations (Rahardian et al., 2025; Suria, 2024). In this study, the usability evaluation involved 15 respondents, consisting of 5 Admin/HR users, 5 Supervisor/Manager users, and 5 Employee users. This number was considered sufficient for a prototype-level evaluation because the objective was to assess early-stage usability and interface acceptability across the three main user roles rather than to measure organization-wide adoption. Each respondent interacted with the main modules of the system, including login, dashboard review, employee detail access, AI classification display, and recommendation viewing, before completing the SUS questionnaire.

3.11 AI Evaluation Metrics

The AI models were evaluated using accuracy, precision, recall, F1-score, and a confusion matrix. These metrics were selected because the proposed system performs a multi-class classification task, and model quality cannot be assessed adequately through accuracy alone. Accuracy provides an overall measure of correct classification, while precision and recall help assess the model’s behavior across classes. F1-score was included as a balanced measure that combines precision and recall, making it particularly useful when performance consistency across categories is important (Jain, 2025; Mourad et al., 2024; Thanri et al., 2025).

The confusion matrix was used to provide a more detailed class-level evaluation by identifying where misclassification occurred and whether certain categories were more difficult to detect than others. This was especially relevant in the present study because the managerial usefulness of the system depends not only on overall classification accuracy, but also on whether High, Moderate, and Low employee performance categories are distinguished in a stable and interpretable way.

Because the system is intended for prototype-level decision support, the evaluation emphasized comparative model suitability rather than absolute predictive claims. Thus, the primary purpose of the AI evaluation was to determine which of the three candidate models was most appropriate for integration into the proposed information system under limited implementation conditions.

3.12 Data Analysis Technique

The analysis in this study combined descriptive, technical, and comparative procedures. Descriptive analysis was used to summarize the variable structure, dataset characteristics, yearly trends, and user-evaluation responses. Technical analysis was used to verify system functionality, processing logic, and interface-related feasibility. Comparative analysis was used to compare the three machine-learning models and select the most appropriate model for integration into the system.

This mixed analytical strategy was considered appropriate because the study aimed to assess not only whether the system could be designed and implemented, but also whether it could operate coherently, provide acceptable usability, and generate analytically meaningful classification results under limited conditions. Therefore, the analysis focused on feasibility, coherence, and comparative performance rather than on population-level causal inference (Mourad et al., 2024; Sihombing, 2024).

3.13 Ethical and Practical Considerations

Because the study operated under a limited implementation design, it emphasized practical and ethical caution in two ways. First, the use of limited or structured simulated data was transparently framed as a prototype-validation strategy rather than as a basis for broad organizational generalization. Second, the AI outputs were positioned as decision-support recommendations, not as a replacement for managerial judgment. This is consistent with recent responsible-AI and HR-AI literature, which emphasizes that AI in employee-related settings should support human oversight and should not be treated as an autonomous authority in performance decisions (Bujold et al., 2024).

A further practical consideration concerns the interpretation of the AI results. Since the dataset was structured for prototype evaluation, the classification results should not be interpreted as definitive organizational judgments. Instead, they should be understood as decision-support outputs generated within a controlled evaluation setting. For this reason, the AI component in this study was deliberately positioned as an augmentation mechanism rather than an autonomous decision maker. Human oversight remains essential, especially in performance-related contexts where classification outcomes may influence managerial review, coaching, and development actions.

4. Results and Discussions

4.1 Problem Context and Dataset Preparation

The development of the proposed system was initiated from a practical organizational problem in which KPI-based evaluation was still conducted manually, periodically, and in a fragmented manner. Preliminary internal evidence indicated that manual KPI collection created delays, typing errors, and interpretive inconsistencies, which could reduce the accuracy of managerial analysis. The same document also showed a decline in the employee performance score from 78.5 in 2022 to 70.0 in 2023 and 63.5 in 2024, suggesting that the existing process did not yet provide sufficiently responsive and corrective managerial insight.

To support prototype evaluation under limited implementation conditions, this study used a structured employee-performance dataset covering the period 2021–2025. The dataset contained 340 employee-year records, representing an average of 68 employees per year. The data structure was designed to remain consistent with the conceptual variables used in the study, namely competency, digital KPI achievement, and employee performance. In this way, the dataset functioned as a prototype-evaluation dataset for demonstrating data flow, dashboard structure, and AI-based classification rather than as a basis for broad organizational generalization.

The yearly pattern of the dataset was intentionally aligned with the internal performance trend reported in the underlying organizational document. Thus, the annual average performance profile was constructed to reflect a relatively stronger baseline in 2021, followed by a declining trend in 2022, 2023, and 2024, and a continued low-performance tendency in 2025. This alignment was important because the prototype was designed not only as a technical artifact, but also as a response to a documented managerial problem. As a result, the data narrative used in the system prototype remained consistent with the organizational issue that motivated the study.

Table 1. Dataset Summary by Year

Year	Number of Employees	Average Competency Score	Average KPI Score	Average Performance Score	Dominant Performance Label
2021	68	81.16	81.14	82.00	High
2022	68	78.22	77.46	78.50	Moderate
2023	68	70.10	68.78	70.00	Moderate
2024	68	63.41	60.97	63.50	Low
2025	68	61.24	58.83	61.00	Low

4.2 Prototype Development Result

In order to address these limitations, the prototype was designed to integrate digital KPI recording, competency assessment, performance scoring, AI-based performance classification, and recommendation generation into a single system environment. Rather than focusing on a full enterprise-scale implementation, this study emphasized the development of a prototype that could demonstrate the core functional and analytical feasibility of the proposed solution. This approach was considered appropriate because the objective of the study was to produce an academically valid system artifact under limited implementation conditions.

The prototype also relied on a structured dataset consisting of 340 employee-year records covering the period 2021–2025, with an average of 68 employees per year. This dataset was prepared to remain consistent with the competency–digital KPI–performance framework of the study and to reflect the declining performance trend documented in the underlying organizational evidence. Thus, the prototype not only demonstrated technical functionality, but also reflected the practical problem context that motivated the research.

To explain the structural design of the proposed prototype, the overall system architecture is presented in Figure 1.

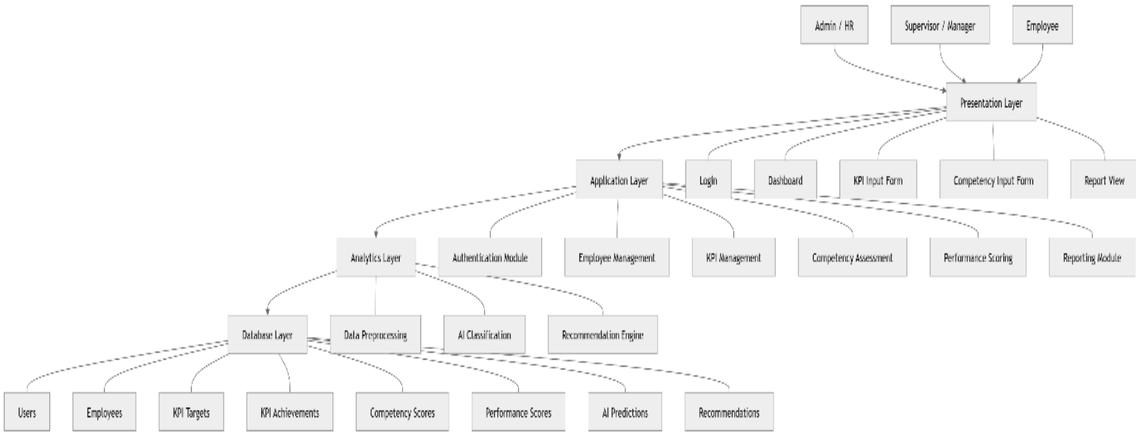

Figure 1. System architecture of the proposed AI-assisted digital KPI information system.

As shown in Figure 1, the prototype was designed using a layered architecture consisting of the presentation layer, application layer, analytics layer, and database layer. This structure was selected to separate user interaction, system processing, AI analysis, and data storage into distinct but connected components. Such separation improves readability, modularity, and feasibility for a prototype-oriented implementation. In practical terms, the layered design ensures that KPI and competency data entered by users can be processed consistently, transformed into structured performance scores, analyzed by the AI module, and finally displayed as dashboards and recommendations.

From a functional perspective, the prototype included the minimum set of modules required to support the article objectives. These modules include user login, KPI input, competency input, performance scoring, AI-based classification, recommendation output, dashboard display, and report export. Although limited in scope, these modules were sufficient to demonstrate that a digital KPI-based information system can move beyond descriptive record keeping and support analytical decision-making.

4.3 System Architecture and Functional Flow

The proposed system involves three primary actors: Admin/HR, Supervisor/Manager, and Employee. Admin/HR is responsible for maintaining employee-related data and ensuring that KPI and competency records are properly entered into the system. Supervisor/Manager uses the system to review performance summaries, observe AI-based classification outputs, and receive recommendation support for managerial action. Employee users are positioned more narrowly, with access mainly to performance-related information and recommendation outputs relevant to their own records.

The main user interactions and system functions are summarized in the simplified use case diagram shown in Figure 2.

Figure 2 shows that the system was intentionally designed with a limited but focused set of functions. The main functions include Login, Input KPI, Input Competency, View Dashboard, AI Classification, Recommendation, and Export Report. Admin/HR and Supervisor/Manager interact with most of these functions because they are directly involved in KPI monitoring and performance evaluation. Employee users have a more restricted interaction scope, mainly limited to login, dashboard viewing, and recommendation access.

This simplified use case structure was selected to maintain consistency with the prototype-based scope of the study. Instead of attempting to represent all possible enterprise features, the diagram emphasizes only the essential functions required to demonstrate the operational logic of the proposed system. This makes the prototype more realistic and better aligned with the constraints of limited implementation.

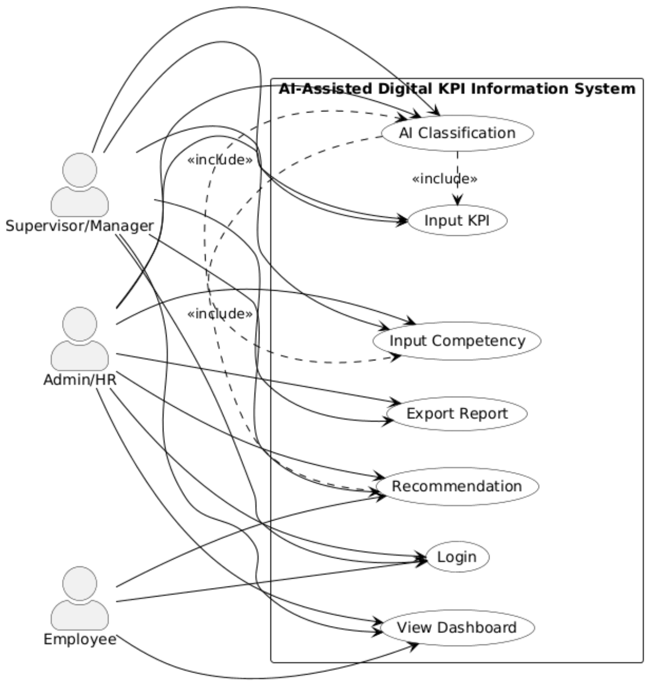

Figure 2. Use case diagram of the proposed AI-assisted digital KPI information system.

The operational workflow of the system is further illustrated through the activity diagram presented in Figure 3.

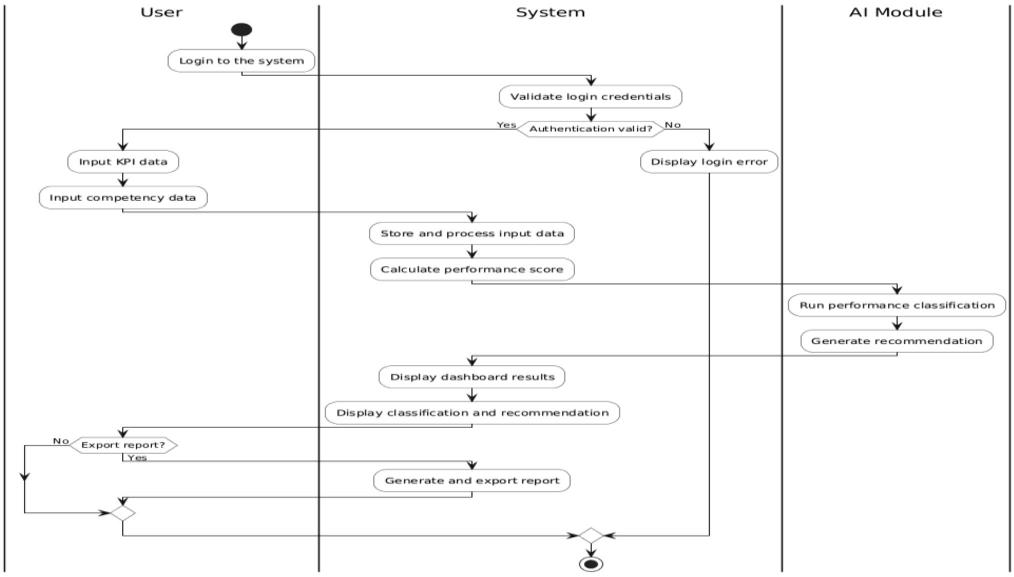

Figure 3. Activity diagram of proposed system.

As shown in Figure 3, the process begins when the user logs into the system. After successful authentication, KPI and competency data are entered into the prototype. The system then stores and processes these values, calculates performance-related scores, and forwards the structured inputs to the AI module. The AI module classifies employee performance into predefined categories, after which the recommendation module generates action-oriented outputs such as training suggestions, coaching needs, or development priorities. The final results are displayed through the dashboard, and users may optionally export the report.

The use of activity partitions in Figure 3 clarifies the division of responsibilities among the User, System, and AI Module. This is important because one of the contributions of the prototype is the explicit integration of AI into the performance workflow. In the earlier manual process, KPI data collection ended at descriptive assessment. In contrast, the proposed prototype extends the process by adding an analytical layer that classifies performance and produces recommendations in a more structured way.

To support prototype implementation, the core database structure is summarized in the simplified ERD shown in Figure 4.

Figure 4 presents the simplified Entity Relationship Diagram of the proposed system. The database structure includes the entities users, employees, kpi_records, competency_records, performance_results, ai_predictions, and recommendations. The users entity manages access and role information. The employees entity stores employee profile data together with annual contextual information. KPI-related values are stored in kpi_records, while competency assessment values are stored in competency_records. The performance_results entity contains the aggregated performance outcomes, including performance scores and labels. The ai_predictions entity stores the output of the classification model, and the recommendations entity stores the recommendation results generated from the classification output.

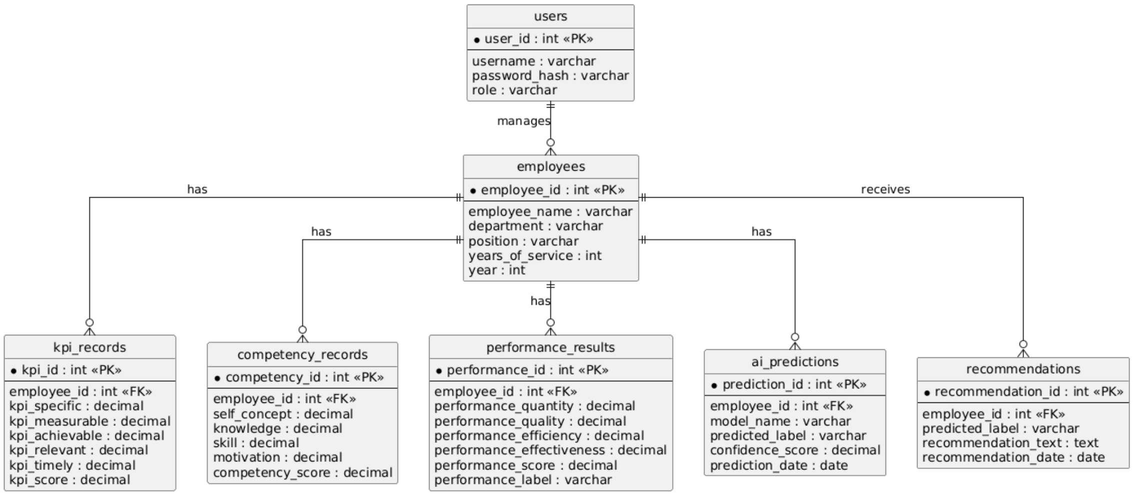

Figure 4. Entity Relationship Diagram (ERD)

This database structure is consistent with the structured dataset used in the study, in which each observation represents an employee-year record. Such a design was chosen because it allows the system to support both dashboard visualization and AI classification using the same core data structure. The ERD also demonstrates that the system was not designed as an isolated AI experiment, but as a coherent information system where operational data entry, analytical processing, and managerial output are connected through a shared database logic.

Taken together, Figures 1 through 4 show that the prototype was built as an integrated but limited system artifact. Figure 1 explains the layered architecture, Figure 2 identifies the key interactions between users and functions, Figure 3 clarifies the operational workflow, and Figure 4 provides the underlying data structure. This progression is important because it demonstrates that the proposed solution is not merely conceptual; it has been translated into a structured system design that can support subsequent evaluation of dashboard functionality, AI-based classification, and recommendation output.

4.4 Dashboard Implementation Result

The system dashboard was developed as the main interface for performance monitoring. The dashboard integrates filters, summary cards, visual trends, performance-label distribution, top employees, employees at risk, recommendation rules, department comparison, and employee data preview. This design was selected to ensure that managers could access both overview-level and record-level information without navigating through multiple disconnected forms.

The main dashboard displayed five strategic summary indicators: total records, average competency score, average KPI score, average performance score, and dominant performance label. These indicators were followed by a yearly performance trend chart and a doughnut chart showing the distribution of performance labels. Additional panels were used to display top-performing employees, employees at risk, and department comparison. This composition reflects the system's intention to support both monitoring and managerial prioritization in a single interface.

As shown in Figure 5, the dashboard provides a structured visual environment for reviewing employee performance data. The use of charts, score cards, category labels, and department summaries improves visibility compared with a purely tabular or manual KPI process. In particular, the "Employees at Risk" panel and the "Quick Recommendation Rules" panel show how the dashboard extends the function of KPI reporting toward practical decision support.

The prototype also included a record-level detail page. This page was designed to display employee-specific scores, AI classification results, recommendation text, score composition, and historical performance trajectory. The inclusion of this page is important because organizational monitoring often requires not only aggregate dashboards, but also detailed views that allow supervisors or HR staff to trace the basis of a particular result.

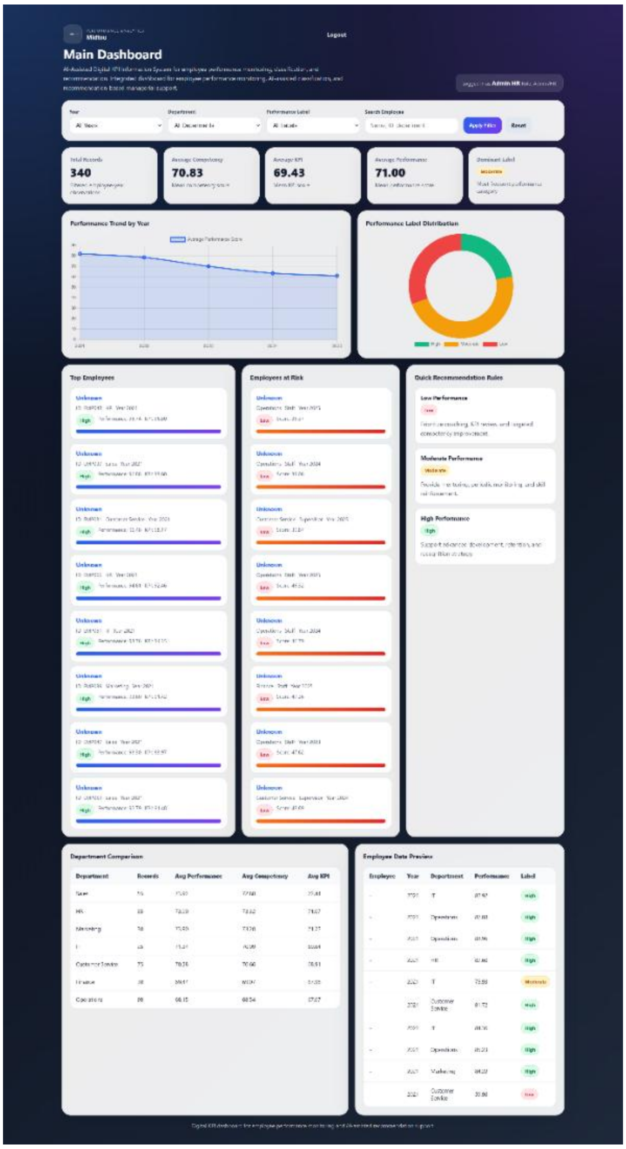

Figure 5. Main dashboard interface of the AI-assisted digital KPI information system.

Figure 6 shows that the employee detail page is able to present competency score, KPI score, performance score, AI classification label, recommendation text, score composition chart, and performance history. This makes the prototype more useful than a conventional summary dashboard because it connects analytical output with individual employee context. For example, a low classification result is not shown as an isolated label, but is accompanied by a recommendation and historical trend that can guide managerial interpretation.

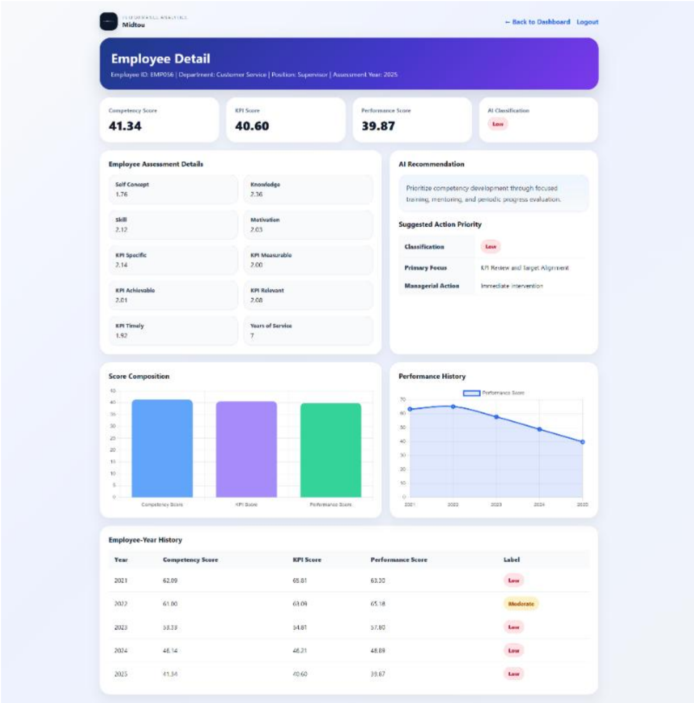

Figure 6. Employee detail page with AI classification and recommendation output.

4.5 AI Model Comparison Result

To support AI-assisted classification, this study compared three machine-learning algorithms: Logistic Regression, Random Forest, and XGBoost. These models were evaluated using the same structured dataset and the same target variable, namely employee performance category. The comparison focused on accuracy, precision, recall, and F1-score to determine which model was most suitable for integration into the proposed system. The results showed that all three models achieved relatively strong performance, indicating that the structured KPI–competency–performance dataset was sufficient for classification-oriented decision support.

Among the three models, Random Forest produced the best overall performance. This result can be interpreted from the nature of the input variables used in the study. The system combines competency indicators, KPI indicators, and aggregated performance attributes, which are likely to interact in non-linear ways rather than in a purely linear pattern. Under these conditions, Random Forest has an advantage because it can model non-linear relationships and variable interactions more flexibly than Logistic Regression, while remaining relatively stable on structured tabular data. A similar pattern has been reported in HR-related classification studies, where Random Forest performs strongly in employee-data categorization and workforce segmentation because of its robustness and classification stability (Dong, 2024; Jain, 2025).

Table 2. Comparison of AI classification models.

Model	Accuracy	Precision	Recall	F1-score
Random Forest	0.9853	0.9857	0.9853	0.9852
Logistic Regression	0.9706	0.9723	0.9706	0.9700
XGBoost	0.9559	0.9574	0.9559	0.9553

Table 2 indicates that Random Forest achieved the highest performance across all major evaluation metrics, with an accuracy of 0.9853 and an F1-score of 0.9852. Logistic Regression followed closely, with an accuracy of 0.9706 and an F1-score of 0.9700, while XGBoost produced the lowest but still strong result, with an accuracy of 0.9559 and an F1-score of 0.9553. Based on these results, Random Forest was selected as the most suitable AI engine for the proposed system.

At the same time, the result should not be interpreted as meaning that Random Forest is universally superior in all HR settings. Logistic Regression still produced strong performance and remains useful as a baseline model because of its higher interpretability. In HR-related systems, model selection should not be based solely on predictive metrics, since transparency and human oversight are especially important when AI outputs may influence managerial review. Recent literature on AI in HRM and responsible AI repeatedly emphasizes that predictive performance must be balanced with explainability, fairness, and trustworthiness in employee-facing systems (Bujold et al., 2024; Naoum et al., 2026; Úbeda-García et al., 2025). Thus, while Random Forest was selected as the best-performing model in this study, its use in a live organizational setting should ideally be complemented by feature-importance analysis or other explainability mechanisms.

The comparative result also strengthens the role of the AI component in the proposed system. The contribution of the model comparison is not merely technical; it demonstrates that KPI-based performance management data can be transformed into a structured classification problem and embedded into a broader information system artifact. In this sense, the AI module extends the function of the system from descriptive reporting to predictive and recommendation-oriented support, which is consistent with recent work positioning AI as an augmentation mechanism in HRM rather than as a stand-alone computational tool (Bastida et al., 2025; Varma et al., 2024; Wang et al., 2025).

4.6 Confusion Matrix Result

The confusion matrix analysis was used to examine the classification behavior of each model in greater detail. While aggregate metrics provide an overall indication of performance, the confusion matrix makes it possible to identify where misclassification occurs and whether certain classes are more difficult to detect than others. In the present study, the confusion matrices show that all three models performed strongly across the three classes, but Random Forest produced the most stable class-level behavior.

For Logistic Regression, the model classified Low and Moderate observations very well, but still produced a small number of errors in the High category by predicting some High records as Moderate. Random Forest reduced this error and provided the most balanced pattern among the compared models. XGBoost also performed strongly, but showed slightly less stability than Random Forest in the Moderate category. This pattern is consistent with the comparative metric results and supports the conclusion that Random Forest is the most suitable classification engine for the proposed system.

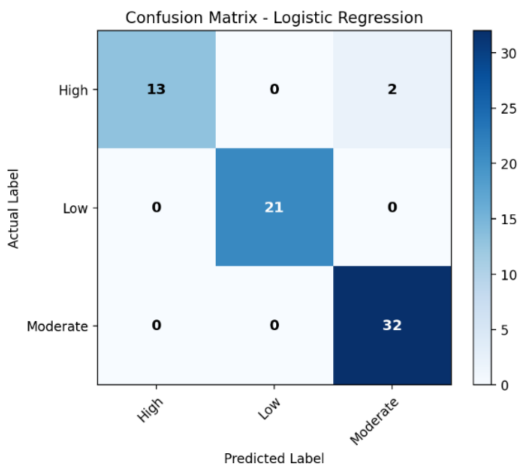

Figure 7. Confusion matrix of Logistic Regression.

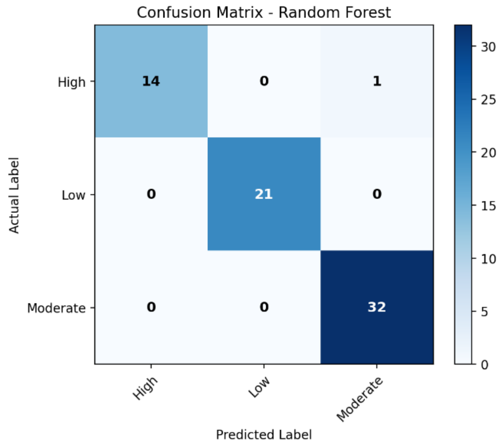

Figure 8. Confusion matrix of Random Forest.

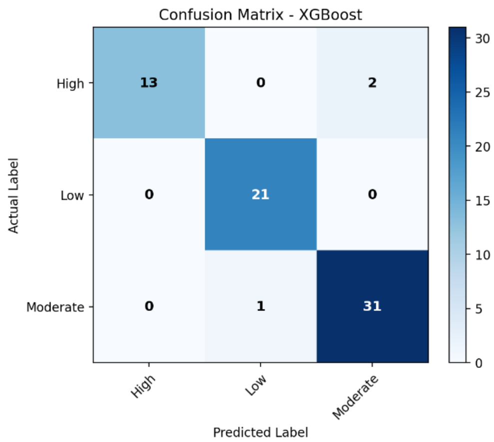

Figure 9. Confusion matrix of XGBoost.

The confusion matrix of Logistic Regression shows that the model correctly classified 13 High, 21 Low, and 32 Moderate observations, with only 2 High records misclassified as Moderate. The confusion matrix of Random Forest improved this pattern by correctly classifying 14 High, 21 Low, and 32 Moderate observations, with only 1 High record misclassified as Moderate. Meanwhile, XGBoost correctly classified 13 High, 21 Low, and 31 Moderate observations, with 2 High records misclassified as Moderate and 1 Moderate record misclassified as Low.

These results strengthen the conclusion that Random Forest is the most suitable model for the system. Its confusion matrix demonstrates not only the highest aggregate metric values, but also the most stable class-level prediction behavior. This is important because a performance information system should not only achieve high numerical accuracy, but also minimize misclassification in a way that preserves managerial trust.

A more critical interpretation is also necessary. The very strong confusion matrix results are partly influenced by the structured nature of the evaluation dataset. Because the dataset was intentionally aligned with the documented organizational performance trend, the classification patterns are clearer and more controlled than they might be in a naturally occurring operational dataset. This raises the possibility that the model performance reflects not only classifier quality, but also the relative regularity of the data structure used for evaluation. Therefore, the confusion matrix findings should be read as evidence of controlled prototype-level model suitability rather than as proof of broad real-world generalizability. This limitation is important because high performance in structured evaluation data does not automatically imply equal robustness in noisier or more heterogeneous live environments.

Even with that limitation, the confusion matrices remain important for the information-system contribution of the study. They show that the selected model can distinguish employee performance categories with a high degree of consistency, which is essential for recommendation-oriented dashboard use. In a KPI-based monitoring environment, stable class separation is more useful than raw score estimation alone because managers typically need actionable categories, such as employees requiring intervention, monitoring, or recognition. This interpretation is aligned with recent performance-management and AI-HRM research that treats classification as a practical support mechanism for managerial triage and prioritization rather than as a substitute for human judgment (Dong, 2024; Varma et al., 2024).

4.7 Recommendation Module Result

In addition to AI classification, the system included a recommendation module that translated predicted performance categories into practical managerial suggestions. This module was intentionally designed to remain simple and interpretable. Rather than generating opaque outputs, the system linked each performance category to a set of understandable actions: employees classified as Low were associated with KPI review, closer supervision, and targeted competency development; employees classified as Moderate were linked with coaching, skill reinforcement, and periodic monitoring; and employees classified as High were associated with retention, recognition, and advanced development planning.

The inclusion of this module significantly increases the practical value of the system. In many conventional KPI settings, the evaluation process ends with score presentation, and managers must interpret the meaning of those scores manually before deciding on follow-up actions. In the proposed system, the recommendation logic reduces that interpretive burden by translating analytical output into managerial guidance. This result supports the broader argument in AI-HRM literature that AI becomes more valuable when it is embedded into decision-support workflows that connect prediction to action, rather than being presented only as a technical result (Bastida et al., 2025; Varma et al., 2024; Wang et al., 2025).

However, the recommendation logic also reveals an important limitation. Because the module is rule-based, the current recommendations are interpretable but not fully adaptive. They depend on predefined mappings from class labels to managerial actions, which makes them suitable for prototype evaluation but still limited in personalization depth. In a live implementation, more adaptive recommendation strategies could be explored, for example by combining class prediction with feature-importance signals, historical trajectory patterns, or supervisory feedback loops. Even so, the current design is still appropriate in light of recent responsible-AI research, which suggests that simpler and more explainable recommendation logic may be preferable in early-stage HR applications where trust and transparency are critical (Bujold et al., 2024; Naoum et al., 2026).

4.8 System Testing Result

The testing results indicate that the proposed system is technically feasible and operationally coherent under limited implementation conditions. Black box testing confirmed that the main workflow modules login, KPI input, competency input, dashboard display, employee detail access, AI classification, recommendation display, and report generation operated according to their expected outcomes. Limited white box testing also indicated that the internal logic of score calculation, classification routing, and recommendation mapping was consistent.

These findings are important from an information systems perspective because analytical value depends not only on the quality of the AI model, but also on whether the system can reliably deliver that model's output through a usable workflow. A classification engine with strong numerical performance would have limited practical value if users could not access the dashboard clearly, trace employee-level outcomes, or interpret the generated recommendations. This aligns with recent HR analytics and information-system research showing that decision-support value arises from the interaction between data structure, analytics, and interface usability rather than from predictive performance alone (Cho et al., 2023; Diefenhardt et al., 2025; Raja et al., 2025).

The usability result also deserves a more critical reading. A favorable usability score indicates that the interface is understandable and acceptable for early use, but it does not by itself guarantee sustained organizational adoption. User perception in prototype settings is often more positive than in longer-term operational settings, particularly when users have not yet experienced issues such as data incompleteness, organizational resistance, or process fatigue. Therefore, the present usability result should be interpreted as an early-stage acceptability indicator, not as evidence of long-term adoption success. This is consistent with prototype-oriented information-system evaluation studies, which treat usability findings as important but preliminary evidence of interface feasibility rather than as proof of organizational maturity (Purwanto & Nurrohman, 2026).

The same caution applies to the performance-efficiency results. The observed response times indicate that the prototype is sufficiently responsive for routine interaction, but these times were measured under limited, controlled conditions rather than under enterprise-level operational load. Consequently, the reported efficiency values should be interpreted as evidence of prototype adequacy rather than production-level benchmarking. Even so, they remain useful because they confirm that the integration of dashboard functions, classification logic, and recommendation output did not create an unusable interface burden at the prototype stage.

4.8.1 Black Box Testing Result

Black box testing was conducted to verify whether each core function of the system operated according to its expected behavior. The testing covered the main modules of the system, including login, KPI data input, competency data input, dashboard display, AI classification, recommendation generation, employee detail display, and report export. The testing result indicates that all core modules performed successfully under normal input conditions.

Table 3. Black box testing result

Module	Test Scenario	Expected Result	Result
Login	User enters valid username and password	User is redirected to the main dashboard	Valid
Login	User enters invalid credentials	System displays login error message	Valid
KPI Input	Admin enters KPI values	KPI data are stored and processed correctly	Valid
Competency Input	Admin enters competency values	Competency data are stored and processed correctly	Valid
Dashboard Display	User opens main dashboard	Dashboard summary, charts, and tables are displayed correctly	Valid
Employee Detail	User selects employee record	Employee detail page is displayed correctly	Valid
AI Classification	System processes employee assessment data	Performance label is generated and displayed correctly	Valid
Recommendation	AI classification is completed	Recommendation output is displayed correctly	Valid
Performance History	User opens employee detail page	Historical performance chart is displayed correctly	Valid
Report Export	User requests export or report output	Report is generated correctly	Valid

Based on Table 3, all tested functions met their expected outcomes. No critical functional failure was observed in the main workflow of the system. This indicates that the developed system is functionally feasible for KPI-based performance monitoring and AI-assisted recommendation.

4.8.2 White Box Testing Result

White box testing was applied in a limited manner to evaluate the correctness of the internal logic in the main processing modules. The focus of this test was placed on the modules that directly affect system calculation and analytical consistency, namely performance score calculation, classification routing, and recommendation logic.

The result of the test shows that the control flow of the main modules operated as intended. Each input path produced the correct processing sequence, and no major logical inconsistency was found in the tested modules. Since this study was limited to a prototype-level system, white box testing was restricted to core routines only.

Table 4. Limited white box testing result

Module	Aspect Tested	Result
Login validation	Authentication logic	Valid
KPI scoring	KPI aggregation formula	Valid
Competency scoring	Competency aggregation formula	Valid
Performance scoring	Final performance score calculation	Valid
AI routing	Classification process invocation	Valid
Recommendation logic	Rule-based recommendation mapping	Valid

The limited white box testing confirms that the internal logic of the system is consistent and adequate for prototype implementation.

4.8.3 Usability Testing Result

Usability testing was conducted to assess whether the system interface was understandable, easy to use, and useful for users. The evaluation involved 15 respondents, consisting of 5 Admin/HR users, 5 Supervisor/Manager users, and 5 Employee users, in order to represent the main user roles of the proposed system. The evaluation focused on interface clarity, ease of navigation, readability of information, usefulness of system features, and overall satisfaction.

Table 5. Usability evaluation result

Indicator	Score
Ease of use	82
Interface clarity	84
Feature usefulness	85
Dashboard readability	81
Overall satisfaction	83

The result indicates that the system achieved a good usability level and was considered acceptable for early operational use. The average usability score of the system was 83, indicating that the developed system was generally perceived as easy to understand and useful. Users were able to identify the main dashboard, employee detail page, and recommendation output without major difficulty. These findings suggest that the system interface is sufficiently usable for performance monitoring and decision-support purposes in a limited implementation setting.

4.8.4 AI-Assisted Recommendation Testing Result

The recommendation module was evaluated by verifying whether the generated recommendation corresponded appropriately to the predicted employee performance category. The result shows that the recommendation logic consistently produced relevant outputs for each classification label.

Table 6. Recommendation module testing result

Performance Label	Expected Recommendation Type	Result
High	Recognition and advanced development	Valid
Moderate	Coaching and skill reinforcement	Valid
Low	KPI review, supervision, and targeted improvement	Valid

This result indicates that the recommendation module was able to translate classification output into practical managerial suggestions in a consistent manner.

4.8.5 Performance Efficiency Testing Result

A simple performance efficiency test was conducted to observe whether the system could respond within an acceptable time during normal use. The measured aspects included login response, dashboard loading, employee detail display, AI classification execution, and report generation. The test was performed in a local prototype environment, and each function was executed three times under the same condition, with the average response time recorded as the final result. In this study, response times below 2 seconds were treated as acceptable for prototype-level interaction. The results show that the system responded quickly enough for routine prototype-level use.

Table 7. Performance efficiency testing result

Function	Average Response Time	Result
Login process	1.12 seconds	Acceptable
Main dashboard loading	1.45 seconds	Acceptable
Employee detail loading	1.28 seconds	Acceptable
AI classification display	1.63 seconds	Acceptable
Report generation	1.84 seconds	Acceptable

All measured functions produced average response times between 1.12 and 1.84 seconds, which indicates that the system was sufficiently responsive for limited prototype interaction. However, these results should not be interpreted as enterprise-level performance benchmarks, because the evaluation was conducted in a controlled local environment without concurrent-user load simulation. Therefore, the reported values are better understood as evidence of prototype adequacy rather than production-scale performance certification.

The overall testing results indicate that the proposed system is technically feasible and operationally coherent for a limited implementation setting. Functional testing confirmed that the main workflow modules operated correctly, while limited white box testing showed that the internal processing logic was consistent. In addition, usability testing demonstrated that the system interface was acceptable and user-friendly, and performance efficiency testing showed that the response time of the prototype remained within a reasonable range for practical use.

4.9 Discussion

The overall findings indicate that the proposed AI-assisted digital KPI information system provides a coherent response to the organizational problem identified at the beginning of the study. The earlier KPI process was manual, periodic, and fragmented, which reduced managerial visibility and delayed corrective follow-up. By integrating KPI records, competency assessment, performance scoring, AI classification, and recommendation output into a single environment, the system addresses not only the problem of data storage but also the broader problem of turning dispersed performance records into actionable managerial insight. This interpretation is consistent with recent performance-management and HRIS research, which shows that digital performance systems are most useful when they support structured monitoring, clearer evaluation, and timely follow-up rather than isolated record keeping (Siraj & Hágen, 2023; Raja et al., 2025).

A central contribution of the study is that it demonstrates the feasibility of combining dashboard-based performance monitoring with AI-assisted classification and recommendation in one operational artifact. Much of the recent literature discusses AI in HRM at a strategic or conceptual level, but fewer studies translate competency assessment, KPI monitoring, and employee performance evaluation into a single implementable information system. In this regard, the present study advances the field by showing how these components can be operationalized together in a prototype environment, which aligns with calls for more context-specific and implementation-oriented AI-HRM research (Úbeda-García et al., 2025; Bauwens & Batistič, 2025; Rana & Kumar, 2025).

The comparison of the three models also yields an important analytical insight. Random Forest outperformed Logistic Regression and XGBoost, suggesting that the relationship among competency indicators, KPI indicators, and performance categories is better captured by a non-linear ensemble model than by a linear baseline. This result is consistent with HR-related classification studies that report Random Forest as a robust method for structured workforce data because of its capacity to handle interaction effects and heterogeneous predictor patterns (Dong, 2024; Jain, 2025). At the same time, this superiority should not be reduced to a simple "highest accuracy wins" logic. In HR contexts, a model must also be judged by its interpretability and fairness implications. Logistic Regression offers stronger transparency, while Random Forest offers stronger predictive stability. Therefore, the practical implication is not merely that Random Forest should be adopted, but that high-performing models in HR systems should be complemented by explainability and governance mechanisms.

The study also highlights the importance of responsible AI in employee-facing systems. Because performance classification can influence coaching, supervision, and development actions, the AI component was intentionally positioned as a recommendation-support mechanism rather than a replacement for managerial judgment. This design choice is strongly supported by recent literature emphasizing human oversight, transparency, fairness, and trust as essential requirements in HR-related AI systems (Bujold et al., 2024; Naoum et al., 2026; Shin et al., 2025). The presence of a recommendation module therefore adds value, but it also increases responsibility: system users must understand that the classification result is an analytical aid, not an automatic personnel decision.

Another issue that must be acknowledged is the possibility of bias introduced by the structured evaluation dataset. Because the data were intentionally aligned with the documented organizational problem context, the resulting patterns may be more regular than those found in naturally occurring operational data. This may partly explain the very strong model performance. Rather than weakening the study, this limitation clarifies its scope: the results provide evidence of artifact feasibility and controlled model suitability, not broad claims of organizational generalizability. Future work should therefore validate the system using larger real-world datasets, noisier operational conditions, and more diverse organizational settings.

Overall, the study confirms that digital KPI management can be strengthened substantially when information-system design is combined with AI-assisted classification and recommendation support. The dashboard improves visibility, the employee detail page improves interpretability, the AI component improves analytical responsiveness, and the recommendation module connects prediction to managerial action. At the same time, the study also shows that high-performing AI in HR contexts must be interpreted critically, especially in terms of explainability, structured-data bias, and responsible use. For that reason, the contribution of this study lies not only in its strong classification result, but also in demonstrating how AI can be embedded responsibly into a KPI-based performance information system under realistic implementation constraints.

5. Conclusion

This study developed and evaluated an AI-assisted digital KPI information system for employee performance monitoring and recommendation. By integrating competency assessment, KPI records, performance scoring, dashboard monitoring, AI classification, and recommendation output into a single environment, the study demonstrates that KPI-based performance management can be implemented in a more structured, responsive, and analytically useful manner.

The results indicate that the proposed system is feasible from both technical and user-oriented perspectives. The dashboard and employee detail features improved monitoring visibility and interpretability, while the AI evaluation showed that Random Forest achieved the best classification performance among the compared models. In addition, the recommendation module demonstrated that AI classification outputs can support more practical and action-oriented managerial follow-up.

From a practical perspective, the study implies that organizations relying on manual KPI processes can benefit from integrated digital systems that combine monitoring, analytics, and AI-assisted interpretation. From an academic perspective, the study contributes to the information systems and HR analytics literature by operationalizing competency, KPI, and employee performance variables into an intelligent prototype-based information system.

However, this study also has several limitations. The system was evaluated only in a limited prototype environment, the dataset was structured primarily for evaluation purposes, and the recommendation module still used simple rule-based logic. Therefore, the findings should be interpreted as evidence of feasibility rather than broad organizational generalization.

Future research is recommended to use larger real-world datasets, implement the system in live organizational settings, integrate explainable AI approaches, and evaluate long-term organizational adoption and impact. These developments may strengthen the reliability, adaptability, and practical value of AI-assisted performance management systems.

References

Akanpaadgi, E., Kuuyelleh, E. N., & Adam, A. (2024). Performance management and contextual performance in technical universities. *Social Sciences & Humanities Open*, *9*, 100788. https://doi.org/10.1016/j.ssaho.2023.100788

Ali, A., Cichoń, D., & Abbas, A. (2026). AI-driven HRM digitization in effect: Mechanisms, conditions, and employee outcomes. *Journal of Innovation & Knowledge*, *13*, 100919. https://doi.org/10.1016/j.jik.2025.100919

Ariningsih, P., & Muhammad, A. H. (2024). Quality Evaluation of Ticketing Management System Using ISO/IEC25010:2023 Standards and AHP Method. *Intechno Journal (Information Technology Journal)*, *6*(2), 86–94. https://doi.org/10.24076/intechnojournal.2024v6i2.1870

Bastida, M., Vaquero García, A., Vazquez Taín, M. Á., & Del Río Araujo, M. (2025). From automation to augmentation: Human resource's journey with artificial intelligence. *Journal of Industrial Information Integration*, *46*, 100872. https://doi.org/10.1016/j.jii.2025.100872

Batool, A., Zowghi, D., & Bano, M. (2025). AI governance: a systematic literature review. *AI and Ethics*, *5*(3), 3265–3279. https://doi.org/10.1007/s43681-024-00653-w

Bauwens, R., & Batistič, S. (2025). The present and future of artificial intelligence in people management research: A bibliometric approach. *BRQ Business Research Quarterly*, *28*(4), 768–787. https://doi.org/10.1177/23409444251341326

Belizón, M. J., Majarín, D., & Aguado, D. (2024). Human resources analytics in practice: A knowledge discovery process. *European Management Review*, *21*(3), 659–677. https://doi.org/10.1111/emre.12605

Bujold, A., Roberge-Maltais, I., Parent-Rocheleau, X., Boasen, J., Sénécal, S., & Léger, P.-M. (2024). Responsible artificial intelligence in human resources management: a review of the empirical literature. *AI and Ethics*, *4*(4), 1185–1200. https://doi.org/10.1007/s43681-023-00325-1

Căvescu, A. M., & Popescu, N. (2025). Predictive Analytics in Human Resources Management: Evaluating AIHR's Role in Talent Retention. *AppliedMath*, *5*(3), 99. https://doi.org/10.3390/appliedmath5030099

Cho, W., Choi, S., & Choi, H. (2023). Human Resources Analytics for Public Personnel Management: Concepts, Cases, and Caveats. *Administrative Sciences*, *13*(2), 41. https://doi.org/10.3390/admsci13020041

Cunha, F., Dinis-Carvalho, J., & Sousa, R. M. (2023). Performance Measurement Systems in Continuous Improvement Environments: Obstacles to Their Effectiveness. *Sustainability*, *15*(1), 867. https://doi.org/10.3390/su15010867

Diefenhardt, F., Rapp, M. L., Bader, V., & Mayrhofer, W. (2025). 'In God We Trust. All Others Must Bring Data': Unpacking the Influence of Human Resource Analytics on the Strategic Recognition of Human Resource Management. *Human Resource Management Journal*, *35*(3), 597–612. https://doi.org/10.1111/1748-8583.12583

Dong, F. (2024). Random Forest Algorithm for HR Data Classification and Performance Analysis in Cloud Environments. *International Journal of Advanced Computer Science and Applications*, *15*(11). https://doi.org/10.14569/IJACSA.2024.0151147

Ehmann, S., Kampkötter, P., Maier, P., & Yang, P. (2024). Performance management and work engagement – New evidence using longitudinal data. *Management Accounting Research*, *64*, 100867. https://doi.org/10.1016/j.mar.2023.100867

Gupta, P., Lakhera, G., & Sharma, M. (2024). Examining the impact of artificial intelligence on employee performance in the digital era: An analysis and future research direction. *The Journal of High Technology Management Research*, *35*(2), 100520. https://doi.org/10.1016/j.hitech.2024.100520

Hendri, M. I. (2025). Performance management in SMEs: Integrating bibliometric insights and systematic review on antecedents and outcomes. *Social Sciences & Humanities Open*, *12*, 102205. https://doi.org/10.1016/j.ssaho.2025.102205

Jain, A. (2025). Human Resource Analytics for Enhancing Employee Performance through Machine Learning Techniques. *Journal of Global Research in Electronics and Communications(JGREC)*, *1*(11), 38–45. https://doi.org/10.5281/zenodo.17847646

Kassa, B. Y., & Worku, E. K. (2025). The impact of artificial intelligence on organizational performance: The mediating role of employee productivity. *Journal of Open Innovation: Technology, Market, and Complexity*, *11*(1), 100474. https://doi.org/10.1016/j.joitmc.2025.100474

Maake, G., & Schultz, C. M. (2025). Using AI in Performance Management: A Global Analysis of Local Government Practices. *Administrative Sciences*, *15*(10), 392. https://doi.org/10.3390/admsci15100392

Madanchian, M. (2024). From Recruitment to Retention: AI Tools for Human Resource Decision-Making. *Applied Sciences*, *14*(24), 11750. https://doi.org/10.3390/app142411750

Mourad, Z., Noura, A., Mohamed, C., & Abdelhamid, B. (2024). Enhancing Employee Performance Management. *International Journal of Advanced Computer Science and Applications*, *15*(3). https://doi.org/10.14569/IJACSA.2024.01503100

Naoum, R. F., Szakadáti, T., & Balogh, G. (2026). Artificial Intelligence (AI) in human resource management (HRM): a systematic review of its dual impact on diversity, equity, and inclusion (DEI). *Management Review Quarterly*. https://doi.org/10.1007/s11301-025-00580-y

Nastase, C., Adomnitei, A., & Apetri, A. (2025). Strategic Human Resource Management in the Digital Era: Technology, Transformation, and Sustainable Advantage. *Merits*, *5*(4), 23. https://doi.org/10.3390/merits5040023

Nguyen, T., & Elbanna, A. (2025). Understanding Human-AI Augmentation in the Workplace: A Review and a Future Research Agenda. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-025-10591-5

Purwanto, E., & Nurrohman. (2026). Strategic Information System Design for Enhancing Executive Decision-Making in the Retail Sector. *Governance IT Adoption and Technology Advance*, *1*(1), 9–16. https://doi.org/10.25124/govita.v1i1.10126

Rahardian, R. L., Khodijah, S., & Rizki, C. A. (2025). Evaluation of the Usability of the Academic Information System Using the System Usability Scale (SUS) Method. *Journal of Computer Science Artificial Intelligence and Communications*, *2*(2), 62–66. https://doi.org/10.64803/jocsaic.v2i2.62

Raja, M., Swamy, T. N. V. R. L., Das, S., Bansal, R., Fahlevi, M., & Aziz, A. L. (2025). Does human resource information system influence performance management? *Cogent Business & Management*, *12*(1). https://doi.org/10.1080/23311975.2024.2438862

Ramachandran, R., Babu, V., & Murugesan, V. P. (2024). Human resource analytics revisited: a systematic literature review of its adoption, global acceptance and implementation. *Benchmarking: An International Journal*, *31*(7), 2360–2390. https://doi.org/10.1108/BIJ-04-2022-0272

Rana, R., & Kumar, S. (2026). AI in human resource management: An interdisciplinary review and bibliometric analysis using SPAR-4-SLR. *Human Systems Management*, *45*(2), 257–275. https://doi.org/10.1177/01672533251365116

Schaffernak, H., Moesl, B., Url, P., Victoria Koglbauer, I., & Vorraber, W. (2025). Towards sustainable software quality in use: a review of measures. *Next Research*, *2*(3), 100680. https://doi.org/10.1016/j.nexres.2025.100680

Shin, H. H., Choi, S., & Kim, H. (2025). Artificial Intelligence (AI) in Human Resource Management (HRM): A driver of organizational dehumanization and negative employee reactions. *International Journal of Hospitality Management*, *131*, 104230. https://doi.org/10.1016/j.ijhm.2025.104230

Sihombing, D. J. C. (2024). Exploring prototype methodology in land information system development: design and evaluation of an application. *Jurnal Info Sains : Informatika Dan*

Sains, *14*(01), 594–604. https://ejournal.seaninstitute.or.id/index.php/InfoSains/article/view/4007

Siraj, N., & Hágen, I. (2023). Performance management system and its role for employee performance: Evidence from Ethiopian SMEs. *Heliyon*, *9*(11), e21819. https://doi.org/10.1016/j.heliyon.2023.e21819

Suria, O. (2024). A Statistical Analysis of System Usability Scale (SUS) Evaluations in Online Learning Platform. *Journal of Information Systems and Informatics*, *6*(2), 992–1007. https://doi.org/10.51519/journalisi.v6i2.750

Szukits, Á., & Móricz, P. (2024). Towards data-driven decision making: the role of analytical culture and centralization efforts. *Review of Managerial Science*, *18*(10), 2849–2887. https://doi.org/10.1007/s11846-023-00694-1

Tessema, S. A., Yang, S., & Chen, C. (2025). The Effect of Human Resource Analytics on Organizational Performance: Insights from Ethiopia. *Systems*, *13*(2), 134. https://doi.org/10.3390/systems13020134

Thanri, Y. Y., Iriani, J., Gowasa, A., & Zaidi, L. (2025). Model Evaluasi Kinerja Karyawan Berbasis Rekam Jejak Digital Menggunakan Machine Learning. *CSRID (Computer Science Research and Its Development Journal)*, *17*(3), 421–435. https://doi.org/10.22303/csrid-.17.3.2025.421-435

Úbeda-García, M., Marco-Lajara, B., Zaragoza-Sáez, P. C., & Poveda-Pareja, E. (2025). Artificial intelligence, knowledge and human resource management: A systematic literature review of theoretical tensions and strategic implications. *Journal of Innovation & Knowledge*, *10*(6), 100809. https://doi.org/10.1016/j.jik.2025.100809

Varma, A., Pereira, V., & Patel, P. (2024). Artificial intelligence and performance management. *Organizational Dynamics*, *53*(1), 101037. https://doi.org/10.1016/j.orgdyn.2024.101037

Vijh, G., Sharma, N., Tiwari, S., Vijh, S., & Sao, A. (2025). Predicting Accurate Employee Performance: An Evaluation of Regression Models. *Procedia Computer Science*, *259*, 433–442. https://doi.org/10.1016/j.procs.2025.03.345

Wang, N., Zhang, X., Li, S., & Gao, X. (2025). Applications of Artificial Intelligence in Enterprise Human Resource Management. *Information Resources Management Journal*, *38*(1), 1–19. https://doi.org/10.4018/IRMJ.389707

Widiastuti, D. P., Nur Cahyo Wibowo, & Eka Dyar Wahyuni. (2025). Software Quality Analysis using the ISO 25010 Standard on the Digital Information System of the Blitar City Government. *JASIEK (Jurnal Aplikasi Sains, Informasi, Elektronika Dan Komputer)*, *7*(2), 269–280. https://doi.org/10.26905/jasiek.v7i2.15884