IJMR | International Journal of Multidisciplinary Research
Международный институт управления и предпринимательства

Research Article Print ISSN 3105-8884 Online ISSN 3105-8892 December 2025, Vol. 1, No. 3

https://doi.org/10.65231/ijmr.v1i3.84

Explainable AI Prevention Pathways for Employee Turnover in Digital Transformation Enterprises: Model Construction and Strategy Optimisation Based on Ensemble Learning + SHAP

Liu Ziqi[1], Jingyi Huang[2],Chen Yu[3],QI Hongwei[4*]

[1]School of Business of Belarusian State University,Belarus,Minsk, Oboynaya Street 7,220004, Belarus

[2]Hello China Inc.Chicago,IL60657,USA

[3]Kyrgyz National University named after Jusup Balasagyn, Frunze Street 547, Bishkek 720033,Kyrgyzstan

[4*]Department of English Language and Translation, Institute of Modern Languages and International Studies, M.K. Ammosov North-Eastern Federal University; Yakutsk, Russian Federation;

KEYWORDS

Digital transformation;

Human resource risk management;

Employee attrition prediction;

Explainable AI;

SHAP method

ABSTRACT

In the context of digital transformation, employee turnover has become a critical risk for corporate development. Traditional human resource management and purely machine learning models have limitations in prediction or face ‘black box’ issues. This study constructs an Ensemble Learning + SHAP dual-track framework, using digital enterprises as a case study. It’s trying to achieve a high-precision turnover prediction through Random Forest, while leveraging SHAP to reveal key influencing factors,most notably a U-shaped relationship between tenure and attrition risk,and proposes targeted retention strategies such as a hierarchical career development framework and a closed loop workload optimisation system. This approach effectively addresses the accuracy-interpretability trade off, providing a new pathway for intelligent, proactive human resource risk management in digitally transforming enterprises.

INTRODUCTION

Against the backdrop of deepening digital transformation, employee attrition has emerged as a critical strategic risk, directly constraining corporate sustainability and innovation capacity. Traditional human resource management (HRM) practices, which often rely on managerial intuition and retrospective exit interviews, are theoretically unsound as a mechanism for proactive risk mitigation. They fail to capture the complex, non-linear interactions between myriad factors—such as compensation equity, career trajectory, workload stress, and organisational culture—that collectively drive turnover decisions. Moreover, they suffer from an inherent predictive lag, typically triggering intervention only after an employee has mentally disengaged.

The advent of data analytics and machine learning (ML) promised a change. Algorithms like Random Forests have demonstrably improved predictive accuracy. However, their widespread adoption in HRM has been hampered by a critical barrier which is the black box problem. When a model flags an employee as a high turnover risk, HR practitioners are left in the dark regarding the specific reasons behind this prediction. This opacity hinders the design of personalised retention interventions, undermines managerial trust, and precludes the audit of the model for potential biases related to gender, age, or other protected attributes. Consequently, high precision models often fail to translate into tangible managerial effectiveness.

To bridge this gap, this study proposes and implements a dual track analytical framework that synergistically combines the predictive power of ensemble learning with the explanatory clarity of SHAP . We here by using digital enterprises in Guangzhou, China, as an empirical case study,

* Corresponding author. E-mail address:qihongwei@yandex.ru

Received date: December 01, 2025; Revised manuscript received date: December 10, 2025; Accepted date: December 20, 2025; Online publication date: December 31, 2025.

this research aims to: (1) construct a robust, high performance attrition prediction model; (2) employ SHAP to transform model outputs into transparent, actionable intelligence about key risk drivers; and (3) derive targeted, evidence-based retention strategies. By doing so, it offers a novel, integrated pathway for moving HR risk management from a reactive, intuitive function towards a proactive, intelligent, and explainable strategic pillar.

1.Literature Review： From Prediction to Explainable Intelligence

The quest to understand and predict employee turnover has evolved through distinct methodological stages. The first stage relied heavily on traditional statistical methods for example logistic regression and theory-driven models from organizational behavior, such as the Mobley turnover chain. While providing interpretability, these approaches often struggled with the high-dimensionality and complex interactions present in real-world HR data.

The second stage embraced classical machine learning algorithms, including Decision Trees, Support Vector Machines (SVM), and K-Nearest Neighbors (KNN). These models improved handling of non-linear patterns but introduced a trade-off between performance and transparency. Ensemble methods, notably Random Forest (RF) and Gradient Boosting Machines like XGBoost, marked a significant advance in this stage, consistently achieving state of the art predictive accuracy by aggregating multiple weak learners. However, their nature further deepened the black box dilemma[1].

This has ushered in the current, third stage, focused on Explainable AI (XAI). The core challenge is no longer solely predictive accuracy but achieving interpretability without sacrificing performance. Techniques like LIME (Local Interpretable Model-agnostic Explanations) and SHAP have been adapted to HR analytics. SHAP, grounded in cooperative game theory, is particularly powerful as it provides a unified, theoretically sound measure of feature importance that is both consistent and locally accurate.

Despite these advances, a critical gap persists in the literature. Most studies conclude with model explanations or instance-level insights but stop short of systematically translating these data-driven discoveries into comprehensive, actionable management strategies that are integrated with established HR theories. Furthermore, there is limited research focusing specifically on the unique attrition dynamics within digital transformation enterprises, which operate at a faster pace and under different pressures. This study aims to fill this gap by not only applying the RF+SHAP framework but also using its outputs to construct a theoretically grounded, strategic retention toolkit for this specific context.

2.Research Methodology： A Dual-Track Framework

2.1.The Dual-Track Framework: Ensemble Learning + SHAP

To simultaneously achieve high predictive accuracy and essential interpretability, this study constructs a dual-track analytical framework. The Predictive Track employs Random Forest (RF) as its core engine, chosen for its robustness against overfitting and superior ability to model the intricate, non-linear interactions characteristic of HR data. The Explanatory Track utilizes the SHAP (SHapley Additive exPlanations) framework post-hoc to attribute model predictions to individual input features, transforming opaque outputs into intelligible insights. This is not a sequential application but an integrated design where the predictive model's output becomes the direct input for explanatory analysis.

2.2 Data and Preprocessing

The analysis is grounded in a multi-dimensional human resources dataset from digital enterprises. A rigorous preprocessing pipeline was executed, involving the encoding of categorical variables, treatment of outliers in key numerical features like monthly income and overtime hours using robust statistical methods, and the critical removal of variables that could lead to target leakage, ensuring the model learned genuine predictors of attrition risk rather than procedural artifacts[2].

2.3 Model Optimisation and Integrated Evaluation

The RF model was optimized via a 5-fold cross-validated grid search to tune hyperparameters such as the number of trees (n_estimators) and maximum depth (max_depth), seeking to maximize the area under the ROC Curve (AUC). Model efficacy was assessed through a tripartite framework: (1) Predictive performance using standard metrics; (2) Interpretability fidelity by checking SHAP explanations against domain knowledge; and (3) Operational stability by testing the consistency of outputs across different data

SHAP-Based Insight	Theoretical Correlate	Implication for HR Practice
U-shaped tenure-risk relationship	Career Stage Theory; Social Exchange Theory	Retention strategies must be phased and tailored to employee lifecycle stages (new hire, core, veteran).
Overtime as a primary risk driver	Job Demands-Resources (JD-R) Model; Conservation of Resources Theory	Workload and well-being are not just some soft issues but quantifiable risk factors requiring systemic management.
Job level and income as stabilizers	Equity Theory; Human Capital Theory	Competitive compensation and clear advancement paths are validated as foundational retention tools.

samples.

3.Result and Interpretation： From Data to Insight

3.1.Predictive Performance

The comparative model evaluation yielded decisive results. The Random Forest model demonstrated superior and robust performance, achieving an AUC of 0.9998 and maintaining a precision-recall balance above 0.97. This indicates an exceptional ability to discriminate between employees at risk of departure and those likely to remain. The failure of other models like SVM underscores the importance of aligning algorithmic choice with data characteristics in HR analytics[3,4].

3.2.Explainable Insights via SHAP

The application of SHAP provided the crucial link between prediction and understanding. Global analysis identified the three most influential predictors of turnover: Resignation status (administrative indicator), Total working years (Tenure), and overtime patterns. The most significant finding was the non-linear, U-shaped relationship between tenure and attrition risk. SHAP dependence plots clearly illustrated elevated risk for new employees (<3 years), decreased risk for core employees (3-10 years), and a resurgence of risk for veteran employees (>10 years). Furthermore, SHAP confirmed a strong positive relationship between overtime frequency and attrition probability, while higher job level and income acted as stabilizing factors.

Table.1. Strategic Insights Derived from SHAP Explainability Analysis

4.Theoretical Integration： Explain the WhyBehind the Data

The empirical findings necessitate integration with organizational theory to move from correlation to causal understanding. The U-shaped tenure-risk curve empirically validates the progression of psychological contracts and career stages. The high initial risk mirrors the fragility of early social exchange; the mid-career stability reflects fulfilled exchanges and growth per career stage theory; the late-career risk resurgence aligns with career plateau and conservation of resources theory, where challenge resources deplete. Similarly, the strong link between overtime and attrition provides quantitative validation for the health impairment pathway in the JD-R Model, where excessive demands deplete energy reserves. This synthesis allows us to posit that attrition risk in digital enterprises is a function of the dynamic equilibrium between structural resources and psycho-physiological demands across the career lifecycle which is a proposition made measurable by our framework.

5.Core Strategy Recommendations

Integrating the SHAP-derived insights with the context of digital enterprises, here we proposed a targeted, dual-core strategic framework.

5.1.Phased Career Development Framework

Based on the U-shaped tenure-risk insight, a differentiated approach is mandated:

For new hires (<3 years): Implement structured onboarding and early career navigator programs to accelerate role clarity and social integration, fortifying the initial psychological contract.

For Core Employees (3-10 years): Sustain engagement through Dual-career pathway initiatives, combining transparent vertical promotion with formal lateral rotation programs to provide continuous growth and prevent stagnation.

For veteran employees (>10 years): Mitigate plateau risks by creating formal Senior Expert roles, leveraging their institutional knowledge through mentorship and strategic advisory duties to facilitate value reconstruction.

5.2.Closed-Loop Workload Optimisation System

Addressing the overtime-driven risk requires a systemic JD-R-informed approach:

Assess: Institute an Overtime Necessity Audit to eliminate non-essential overtime.

Compensate and Recover: Enforce a Time-off in Lieu + Premium Subsidy policy to fairly compensate and mandate recovery.

Optimise at Source: Deploy process automation and AI tools to streamline workflows, reducing the root cause of excessive demands.

SHAP Insight	Proposed Intervention	Responsible Function	Success Metric (KPI)
U-Shaped Tenure Risk	1. Launch Early Career Navigator program. 2. Establish Internal Talent Marketplace for lateral moves. 3. Create Senior Fellow roles.	Talent Development HR Business Partners Strategy andOps	Attrition rate by tenure cohort.
Overtime as Key Driver	1. Implement Overtime Transparency Dashboard. 2. Policy: Mandatory Time-off in Lieu. 3. Audit projects for Sustainable Workload Design.	Team Leads / HR HR / Finance Project Management Office	Avg. monthly overtime hours; Well-being survey scores.
Job Level andIncome as Stabilizers	1. Review compensation equity for high-risk cohorts. 2. Accelerate promotion cycles for high-potentials in risk groups.	Compensation andBenefits Leadership Committee	Attrition rate of high-potentials; Pay equity ratio.

Table.2. Actionable Retention Strategy Matrix

6.Ethical Considerations and Implementation Roadmap

Deploying predictive analytics in HR demands ethical vigilance and structured change management.

6.1.Ethical Guardrails

To ensure responsible use, deployment must be governed by principles of transparency and consent (informing employees about analytic purposes), human-in-the-loop decision making (using outputs for supportive dialogue, not automated punishment), and continuous bias auditing (using SHAP to check for unfair disparities across demographic groups).

6.2.A Phased Implementation Roadmap

Transitioning from prototype to practice requires a structured approach:

Phase 1: Pilot and Validation (Months 1-3): Test in a single business unit to refine predictions and explanations.

Phase 2: Process Integration and Training (Months 4-6): Train managers on interpreting risk reports and SHAP insights for coaching conversations.

Phase 3: Full Deployment with Ethics Board (Months 7-12): Enterprise-wide rollout overseen by a cross-functional ethics board.

Phase 4: Continuous Learning (Ongoing): Integrate intervention outcomes to learn which actions are most effective, evolving from predictive to prescriptive analytics.

Conclusion

This study successfully demonstrates that the Ensemble Learning + SHAP framework effectively bridges the accuracy-interpretability divide in employee attrition prediction. By integrating data-driven insights with organizational theory, it provides not just a predictive tool but a diagnostic system for understanding the underlying mechanisms of turnover. The proposed phased career framework and workload optimization system offer a concrete pathway for proactive retention.

Limitations include the cross-sectional nature of the data and the specific cultural context. Future research should pursue longitudinal studies to establish causality, integrate external pull factor data, and replicate the framework across diverse cultural and industrial settings. Ultimately, the goal is to deeply integrate this explainable intelligence into corporate HR systems, creating a closed-loop ecosystem for strategic human capital management that is both empirically grounded and ethically sound[5].

Funding

This research did not receive any specific financial support from public, commercial, or non-profit funding agencies.

REFERENCES

1. Reddy, A. M., Yarlagadda, S., & Akkinen, H. (2021). An extensive analytical approach on human resources using random forest algorithm. arXiv preprint arXiv:2105.07855.
2. Vijayan, N. E. (2025). Mitigating Attrition: Data-Driven Approach Using Machine Learning and Data Engineering. arXiv preprint arXiv:2502.17865.
3. Căvescu, A. M., & Popescu, N. (2025). Predictive Analytics in Human Resources Management: Evaluating AIHR's Role in Talent Retention. AppliedMath, 5(3), 99.
4. Dutraj, R., & Sengupta, P. R. (2025). Aligning Human Resources to Businesses through Human Resource Analytics. Asian Journal of Management, 16(2), 73-81.
5. Chaudhary M., Gaur L., Chakrabarti A., Singh G., Jones P., Kraus S. (2025)An integrated model to evaluate the transparency in predicting employee churn using explainable artificial intelligence.Journal of Innovation and Knowledge.