




Predicting Employee Turnover Using Explainable Machine Learning Models: A Comparative Study of XGBoost, LightGBM, CatBoost, and Deep Neural Networks

Gabriela. Torres¹, Priya. Nair^{2*}, Mikko Lahtinen³


¹ Department of Business Administration and Innovation, Tecnológico de Monterrey, Monterrey, Mexico

² Department of Human Resource Management, Ted Rogers School of Management, Toronto Metropolitan University, Toronto, Canada



³ Department of Industrial Engineering and Management, Aalto University, Espoo, Finland

* Corresponding author email address: priya.nair@torontomu.ca

Editor

Mohammed Abdul Imran Khan
Department of Financial
Management and Economics,
Dhofar University, Oman
mimran@du.edu.om

Reviewers

Reviewer 1: Mehrdad Bayat
Assistant Professor, Department of Management, Payam Noor University, Tehran,
Iran.
Email: bayatmehrdad60@pnu.ac.ir
Reviewer 2: Marco Yamba-Yugsi
Unidad Académica de Posgrado, Universidad Católica de Cuenca, Azuay 010101,
Ecuador
Email: marco.yamba@ucacue.edu.ec

1. Round 1

1.1. Reviewer 1

Reviewer:

In the paragraph describing data collection through an “integrated Human Resource Analytics Framework,” the manuscript would benefit from a detailed conceptual diagram or table illustrating how data sources were integrated. Currently, it remains unclear how survey measures, HR records, and performance indicators were synchronized and linked at the employee level while preserving data quality and confidentiality.

The paragraph beginning “Employee engagement was assessed using the Utrecht Work Engagement Scale” describes the instrument but does not report psychometric evidence for the current sample. Please provide Cronbach’s alpha, composite reliability, and, where appropriate, confirmatory factor analysis results for all survey-based measures used in the study.

Similarly, for the paragraph “Job satisfaction was measured using the Job Satisfaction Survey developed by Spector,” the authors should report the number of items retained, any modifications made to the original instrument, scale scoring procedures,

and psychometric properties within the present dataset. Without such information, measurement validity cannot be adequately assessed.

In the paragraph “Comparative analyses were conducted to determine statistically significant differences among model performances,” no statistical comparison results are presented. Please report the statistical procedures used (e.g., DeLong test, McNemar test, paired bootstrap resampling) and provide significance values demonstrating whether performance differences between CatBoost, XGBoost, LightGBM, and DNN were statistically meaningful.

Table 2 reports impressive performance metrics, particularly the CatBoost model’s AUC-ROC of 0.958. However, confidence intervals are absent for all evaluation metrics. Reporting 95% confidence intervals would allow readers to assess the precision and stability of model performance estimates and strengthen the robustness of the findings.

In the Results section, the authors state that “The superior performance of CatBoost may be attributed to its advanced handling of categorical variables and its ability to minimize prediction bias through ordered boosting procedures.” This explanation is speculative because no empirical evidence is presented to verify that categorical processing specifically drove the observed advantage. Consider conducting ablation analyses or additional experiments to support this interpretation.

Authors revised the manuscript and uploaded the new document.

1.2. Reviewer 2

Reviewer:

The paragraph describing “Organizational commitment was assessed using the Three-Component Organizational Commitment Scale developed by Meyer and Allen” should clarify whether affective, continuance, and normative commitment were modeled separately or aggregated into a composite score. Combining theoretically distinct dimensions may obscure meaningful relationships with turnover behavior.

In the preprocessing section, the authors state that “Feature engineering procedures were employed to derive additional predictors, including tenure-to-promotion ratios, engagement-performance interaction terms, absenteeism trends, compensation growth trajectories, and workload balance indicators.” Greater transparency is required regarding the mathematical formulation of these engineered features. A supplementary appendix containing feature definitions would substantially improve reproducibility.

The paragraph indicating that “Feature importance screening techniques and recursive feature elimination procedures were subsequently applied” does not specify which variables were eliminated, the criteria used for feature retention, or the final number of predictors entering each model. Please provide a feature-selection workflow and associated thresholds.

The statement “To address class imbalance, Synthetic Minority Oversampling Technique (SMOTE) and class-weight adjustment methods were implemented where necessary” requires clarification. It is important to specify whether SMOTE was applied before or after data splitting, how synthetic observations were generated, and whether separate experiments were conducted to compare the impact of oversampling versus class weighting.

In the paragraph describing hyperparameter optimization, the authors note that “Hyperparameter optimization was conducted using grid search and Bayesian optimization approaches.” However, the manuscript does not provide the search spaces, optimization criteria, number of iterations, or computational resources used. These details are essential for methodological reproducibility and should be included in a dedicated table.

The description of the Deep Neural Network architecture is overly general. The sentence “The Deep Neural Network architecture consisted of multiple fully connected hidden layers” should be expanded to report the exact number of layers, neurons per layer, dropout rates, activation functions, optimizer settings, batch size, learning rate, and training epochs.

Authors revised the manuscript and uploaded the new document.

2. Revised

Editor's decision after revisions: Accepted.

Editor in Chief's decision: Accepted.