International Journal of Engineering Inventions
e-ISSN: 2278-7461, p-ISSN: 2319-6491
Volume 15, Issue 11 [November 2026] PP: 142-155

Predicting Employee Turnover Using Machine Learning Algorithms: A Comparative Study in Human Resource Analytics

Vakkas Kuşdoğan
Department of Management Information Systems, Selcuk University, Konya, Türkiye Corresponding Author:

Ilknur Cevik Tekin
Department of Management Information Systems, Selcuk University, Konya, Türkiye,

Yavuz Selim Taspinar
Department of Mechatronic Engineering, Selcuk University, Konya, Türkiye, ytaspinar@selcuk.edu.tr

ABSTRACT: *Employee turnover constitutes a significant challenge in terms of the sustainability of organizations' human resources, preservation of organizational knowledge, and human resources planning. Predicting employees' risk of leaving their jobs in advance enables organizations to develop proactive employee retention strategies against potential personnel losses. The aim of this study is to compare the performance of different machine learning algorithms used to predict employee turnover and to identify the models that are more successful in predicting employees' departure from the organization. The study used an open-access employee turnover dataset obtained from the Kaggle platform containing 9,541 records; 9,540 observations were included in the Orange analyses. Eight machine learning algorithms were evaluated: Artificial Neural Networks, K-Nearest Neighbors, Decision Tree, Random Forest, Support Vector Machines, Logistic Regression, Naive Bayes, and AdaBoost. Model performance was assessed using 5-fold stratified cross-validation. In the analysis using all features, the Artificial Neural Network produced the highest reported numerical performance with an AUC of 0.913, classification accuracy of 0.856, and F1 score of 0.854. Random Forest produced closely comparable results (AUC=0.904, CA=0.853, F1=0.849), whereas SVM showed substantially weaker overall discrimination (AUC=0.584, CA=0.687, F1=0.630). In the second analysis, the four highest-ranked features according to Information Gain Ratio were review, avg_hrs_month, tenure, and promoted. The results indicate that model selection in employee turnover prediction should not rely on accuracy alone; class-specific recall, precision, F1 score, specificity, false-negative results, and overall discrimination should be considered jointly. The findings provide a data-driven framework for evaluating employee turnover risk while also demonstrating the importance of interpreting aggregate and class-specific performance measures separately.*

Keywords: Employee Turnover, Human Resource Analytics, Machine Learning, Predictive Analytics, Employee Retention

---

Date of Submission: 05-11-2026 Date of acceptance: 16-11-2026

I. INTRODUCTION

In today's competitive business environment, retaining qualified employees has become a strategic priority for organizations seeking sustainable competitive advantage. Employee turnover increases recruitment and training costs and may lead to organizational knowledge loss, lower motivation, reduced service quality, and weaker performance. Predicting turnover and identifying employees at risk of leaving are therefore central objectives of contemporary human resource management. As digital transformation accelerates, data-driven decision-making is becoming increasingly widespread in human resource management. Human Resource Analytics supports managerial decisions by applying statistical and computational methods to employee data. Machine learning extends this approach from explaining employee behavior to predicting potential organizational outcomes, including turnover, workforce requirements, and retention risks. Previous turnover research has used conventional methods such as Logistic Regression together with Decision Tree, Random Forest, Support Vector Machine, Artificial Neural Networks, XGBoost, and other ensemble algorithms. Although many studies report high classification performance, comparisons often rely on limited metrics, and their managerial implications are not always discussed. Evidence comparing multiple algorithms on the same data structure also remains limited. A considerable proportion of existing research treats accuracy as the primary performance indicator, while giving less attention to Recall, Specificity, F1-score, and ROC-AUC. In turnover

prediction, false negatives may leave employees at high risk of departure unidentified. Evaluations based only on overall accuracy may therefore be inadequate for managerial decision-making. This study compares Artificial Neural Network (ANN), K-Nearest Neighbors (KNN), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), Logistic

Regression (LR), Naive Bayes (NB), and AdaBoost using an open-access employee analytics dataset from Kaggle. The models were implemented in Orange Data Mining and evaluated using five-fold stratified cross-validation. Aggregate metrics and class-specific confusion-matrix measures were considered together. The contribution of the study lies in examining three complementary evaluation levels within the same experimental design: aggregate Test and Score measures, confusion-matrix measures for the positive "yes" class, and the effect of reducing the feature set through Information Gain Ratio. Recall, F1-score, Specificity, and False Negative Rate were interpreted from a decision-support perspective, particularly in relation to the managerial implications of false negatives. Employee turnover is a major concern in human resource management because of its implications for recruitment and training costs, organizational knowledge loss, and business continuity [1]. Artificial intelligence and machine learning methods have increasingly been used to predict turnover, including Decision Tree, SVM, Logistic Regression, Random Forest, XGBoost, and Artificial Neural Networks [3, 4, 10].

Research in organizational behavior indicates that turnover is associated with leadership and employee attitudes and should not be treated exclusively as a classification problem. Mutlu et al. reported that paternalistic leadership positively predicted job performance and negatively predicted turnover intention, while gender did not significantly moderate these relationships [6]. Sector-specific studies indicate that retention problems vary across work environments. Managerial practices have been examined qualitatively in private security [2], while systematic reviews in nursing report that machine learning can predict turnover and turnover intention, although external validation and generalizability remain concerns [9].

Taner, Areta Hızıroğlu, and Hızıroğlu compared Decision Tree, SVM, Logistic Regression, Random Forest, XGBoost, and Artificial Neural Network using IBM's open dataset. The models were implemented in R and evaluated with accuracy, sensitivity, specificity, precision, F1-score, and ROC-AUC. Logistic Regression produced the highest overall performance; class imbalance and reliable evaluation were identified as limitations [3]. Al Akasheh, Al Hujran, Faisal Malik, and Zaki enriched tabular employee data with knowledge graph features and explainable artificial intelligence. Among five models evaluated on 1,470 IBM employee records, a linear SVM supported by knowledge graph features achieved 0.925 accuracy. Work environment, job satisfaction, and job involvement were identified as relevant factors [4]. Mutlu et al. surveyed 426 white-collar employees in Konya. Correlation analyses and PROCESS Macro applications indicated that paternalistic leadership positively predicted job performance and negatively predicted turnover intention; gender did not significantly moderate these relationships [6]. Using the 2005–2019 Turkish Household Labour Force Surveys, Polat and Ulus examined worker flows and wage differences. Increased worker turnover was accompanied by greater labor-market segmentation, and employees who changed jobs within the preceding 12 months earned lower wages than those remaining in their positions [8]. Aykan et al. examined 393 manufacturing employees in the Kayseri Organized Industrial Zone. PROCESS Macro Models 1, 4, and 7 indicated that job satisfaction partially mediated the relationships between fatigue and organizational outcomes, while life satisfaction mitigated the adverse effect of fatigue on job satisfaction [7]. Talebi, Khatibi Bardsiri, and Khatibi Bardsiri systematically reviewed 58 peer-reviewed studies published between 2015 and 2025. Random Forest was among the most frequently used methods and often achieved high predictive performance; job satisfaction was among the most frequently identified predictors [5]. Shafique, Fatima, and Shafique reviewed conventional and contemporary turnover models. They concluded that turnover is influenced by numerous work-related factors and noted the increasing relevance of contemporary approaches such as job embeddedness [1]. Luamba, Chagadama, and Blye examined strategies for reducing turnover and improving customer satisfaction in the private security sector in Northern Virginia. Semi-structured interviews with six senior security executives generated practice-oriented recommendations concerning retention and organizational performance [2]. Bilenler, Gül, and Uçar compared Gradient Boosting Tree, Random Forest Trees, and XGBoost Regression on an open-source personnel dataset using MAE, MSE, R^2, and adjusted R^2. Error values were similar, while Random Forest produced the highest R^2; its feature-importance analysis identified employee satisfaction as the leading variable [10]. Avrahami, Pessach, Singer, and Chalutz Ben-Gal analyzed approximately 700,000 employee records covering ten years. Relationships between competence, commitment, trust, cultural values, and turnover varied by employee role, individual characteristics, and cultural background [11]. Alpsoy and Tükel developed a hybrid CNN–XGBoost turnover model using a synthetic Kaggle dataset. Because the Turkish and English abstracts reported different accuracy values, the present study does not cite a specific percentage [12]. Alıcı, Ertek, and Aktemur evaluated 28 international airlines using employee-oriented social and financial indicators. LOPCOW was used for criterion weighting and COCOSO for performance ranking. Low workforce turnover and favorable financial indicators occurred together in the highest-performing group [13]. Sönmez and Üngüren analyzed 400

Generation X, Y, and Z employees in four- and five-star hotels in Alanya using PLS-SEM. Job satisfaction positively affected work engagement, while the Welch–Satterthwaite test found no significant moderating effect of generation [14]. Bakal et al. examined the role of artificial intelligence in human resource management and recruitment, addressing potential gains in speed and efficiency together with adoption risks and limitations [15]. Yıldız examined quiet quitting through interviews with nine managers. Participants associated quiet quitting with job dissatisfaction, unmet expectations, long working hours, authoritarian management, and perceived inadequate rewards [16]. Ekiz Kavukoğlu used content analysis to assess the visibility of human resource policies on the websites of private hospitals in Türkiye, including recruitment, training, careers, performance, compensation, and discrimination prevention [17]. Serbes reviewed the use of artificial intelligence in recruitment, data analysis, performance management, workload reduction, and workplace efficiency, while also addressing ethical, technical, and managerial challenges [18]. Kılıç Özkaynar examined employee performance, workplace happiness, and social media use in textile companies in the TR72 Region. Analyses using SPSS 25, AMOS 20, and PROCESS Macro indicated a positive association between performance-related factors and workplace happiness and an adverse effect of social media use on the examined relationship [19]. Hudays et al. screened 596 records and included eight eligible studies in a PRISMA-based systematic review. Several models achieved high internal-validation performance, but none of the studies conducted external validation. Salary and age were among the most frequently reported predictors [9].

The literature centers on two areas: organizational and behavioral factors associated with turnover, and data-analytics approaches for predicting turnover. Together, these perspectives permit examination of both turnover antecedents and predictable patterns. Organizational-behavior research has examined leadership, job satisfaction, employee fatigue, life satisfaction, and work engagement [6, 7, 14]. These findings indicate that turnover cannot be explained solely by demographic indicators; work experience and organizational conditions should also be considered. Predictive research has applied Decision Tree, Random Forest, SVM, Logistic Regression, XGBoost, ANN, CNN, and hybrid models [3, 4, 5, 10, 12]. The studies are heterogeneous in datasets, algorithms, metrics, and validation designs. Although Random Forest is frequently used and often performs well [5], the highest-performing model varies with the data and evaluation framework [3, 4, 10, 12]. Recent research also addresses explainability, external validation, class imbalance, and managerial applicability. However, systematic separation of aggregate performance from the error structure of the positive turnover class within the same experiment, followed by model-level reassessment after feature reduction, remains limited. This study provides a detailed complementary evaluation without claiming to be the first.

The research gap is defined as the need to examine multiple levels of model performance under the same experimental design. The study jointly evaluates aggregate Orange metrics, measures specific to the positive "yes" class, and two scenarios using either all features or four Gain Ratio-selected features under the same dataset and five-fold stratified cross-validation framework. An open-access employee analytics dataset from Kaggle was used to compare eight algorithms. Analyses were conducted in Orange Data Mining using five-fold stratified cross-validation. Aggregate AUC, CA, Precision, Recall, F1, and MCC were evaluated together with positive-class measures calculated from confusion matrices.

- Data: An open-access employee analytics dataset obtained from Kaggle was used.
- Method: ANN, KNN, Decision Tree, Random Forest, SVM, Logistic Regression, Naive Bayes, and AdaBoost were applied.
- Analysis environment: Data analysis, modeling, and performance evaluation were conducted using Orange Data Mining.
- Model validation: Five-fold stratified cross-validation was used.
- Performance evaluation: Models were assessed using Accuracy, Precision, Recall, F1-score, Specificity, and ROC-AUC.
- Comparative evaluation: Results were compared to identify higher-performing algorithms and interpret their implications for human resource management.

The contributions to literature and practice can be summarized as follows:

- Separating aggregate Test and Score metrics from positive-class confusion-matrix metrics provides a more detailed comparison than a single accuracy value.
- Evaluating the same eight algorithms with all features and four Gain Ratio-selected features permits direct comparison of feature-reduction effects across algorithms.
- Interpreting Recall, F1-score, Specificity, and FNR in a managerial context accounts for employees at risk of leaving who may be missed through false-negative predictions.
- This integrated approach supports model selection based on overall discrimination, positive-class detection, and robustness to feature reduction.

The remainder of the study is organized as follows. Section 2 describes the dataset, Orange environment, analytical workflow, algorithms, performance metrics, and validation procedure. Section 3 presents the complete- and reduced-feature results. Section 4 discusses the findings, and Section 5 presents the conclusions and recommendations.

II. MATERIAL AND METHODS

This section describes the dataset, the analytical workflow developed in Orange Data Mining, the machine learning algorithms, the model evaluation metrics, and the cross-validation approach. It was structured to compare

the baseline analysis using all features with the second analysis using four features selected according to Information Gain Ratio within the same evaluation framework.

2.1 Dataset: Employee Turnover Dataset

An open-access Employee Turnover dataset obtained from Kaggle was used to predict whether employees would remain with or leave the company [20]. The source dataset contains 9,541 employee records. Orange included 9,540 observations in its analytical outputs, and model performance was calculated using these observations. Nine independent variables and one target variable were included. The dataset was reported to contain no missing values, and all usable observations were included. The dependent variable was "left," which indicates whether an employee had left the organization. This binary variable was displayed as "no" and "yes" in Orange. For class-specific evaluations based on the confusion matrices, "yes" was defined as the positive class, allowing correct and incorrect classifications of turnover observations to be examined separately. Independent variables: The predictors included department, promotion status (promoted), performance review score (review), number of projects (projects), salary level (salary), tenure, job satisfaction (satisfaction), bonus status (bonus), and average monthly working hours (avg_hrs_month). The numerical variables were number of projects, performance review score, job satisfaction score, tenure, and average monthly working hours. The categorical variables were department, promotion status, salary level, and bonus status. For confidentiality, salary was masked as low, medium, or high. The bonus variable indicates whether the employee received a bonus during the preceding 24 months. Table 1 presents the descriptions, data types, and possible values of all variables used in the analysis.

Table 1. Variables Used in the Employee Turnover Prediction Model

Variable Name	Variable Type	Data Type	Description and Possible Values
left	Target	Categorical Binary	Indicates whether the employee left the company. 0: Left; 1: Did not leave
department	Independent Feature	Categorical	Department or organizational unit in which the employee works
promoted	Independent Feature	Categorical Binary	Indicates whether the employee received a promotion during the preceding 24 months. 1: Promoted; 0: Not promoted
review	Independent Feature	Numerical Continuous	Most recent performance review score, ranging from 0 to 1
projects	Independent Feature	Numerical Discrete	Total number of projects in which the employee is actively involved
salary	Independent Feature	Categorical Ordinal	Salary level masked for confidentiality: low, medium, or high
tenure	Independent Feature	Numerical Continuous	Total period worked for the company in years
satisfaction	Independent Feature	Numerical Continuous	Employee satisfaction score, ranging from 0 to 1
bonus	Independent Feature	Categorical Binary	Indicates whether the employee received a bonus during the preceding 24 months. 1: Received; 0: Not received
avg_hrs_month	Independent Feature	Numerical Continuous	Average monthly working hours

Note. Target refers to the dependent variable, whereas Feature refers to an independent predictor. Categorical variables were converted into the numerical representations required by the machine learning models. The dataset was obtained from Kaggle [20].

2.2 Analytical Workflow and Orange Data Mining Platform

Orange Data Mining version 3.36 was used to develop and evaluate the machine learning models. Orange is an open-source platform that supports data analysis and machine learning through visual workflows [21]. Two parallel analyses were conducted. The first used all nine independent variables, while the second used

four features selected in the Rank widget according to Information Gain Ratio. The dataset was imported through the File widget, after which the target and independent variables were defined. In the baseline branch, data containing all features were directly connected to the eight classifiers and Test and Score, providing baseline performance without feature selection. In the second branch, the features were ranked according to Information Gain Ratio. The "Best ranked: 4" option was selected, and the four highest-ranked features were transferred through Reduced Data. The selected variables were review, avg_hrs_month, tenure, and promoted. The same eight algorithms were evaluated in both analyses: Artificial Neural Network (ANN), k-Nearest Neighbors (kNN), Decision Tree, Random Forest, Support Vector Machine (SVM), Logistic Regression, Naive Bayes, and AdaBoost. This design enabled the effects of feature reduction to be compared across algorithms.

Model evaluations were conducted in Test and Score using five-fold stratified cross-validation. Because "None, show average over classes" was selected, AUC, CA, F1, Precision, Recall, and MCC were interpreted as aggregate or average results across classes. Confusion matrix was also used. TN, FP, FN, and TP were determined from the actual and predicted "no" and "yes" classes. With "yes" defined as positive, class-specific Precision, Recall, F1, Specificity, and False Negative Rate were calculated separately to avoid conflating aggregate performance with turnover-class performance. Discriminatory performance was also evaluated using ROC analysis and AUC. Model comparisons therefore considered AUC, Classification Accuracy, F1, Precision, Recall, MCC, and class-specific confusion-matrix measures together. The first analysis measured performance using all features, whereas the second used the four features selected through Gain Ratio. Both used the same dataset, algorithms, and five-fold stratified cross-validation approach. In the reduced-feature workflow, Rank was placed before Test and Score through Reduced Data. Feature ranking was therefore conducted outside the cross-validation folds. Because supervised feature selection using target information before cross-validation may introduce data leakage, the four-feature results were interpreted with this limitation in mind. A stricter design would repeat feature selection within each training fold.

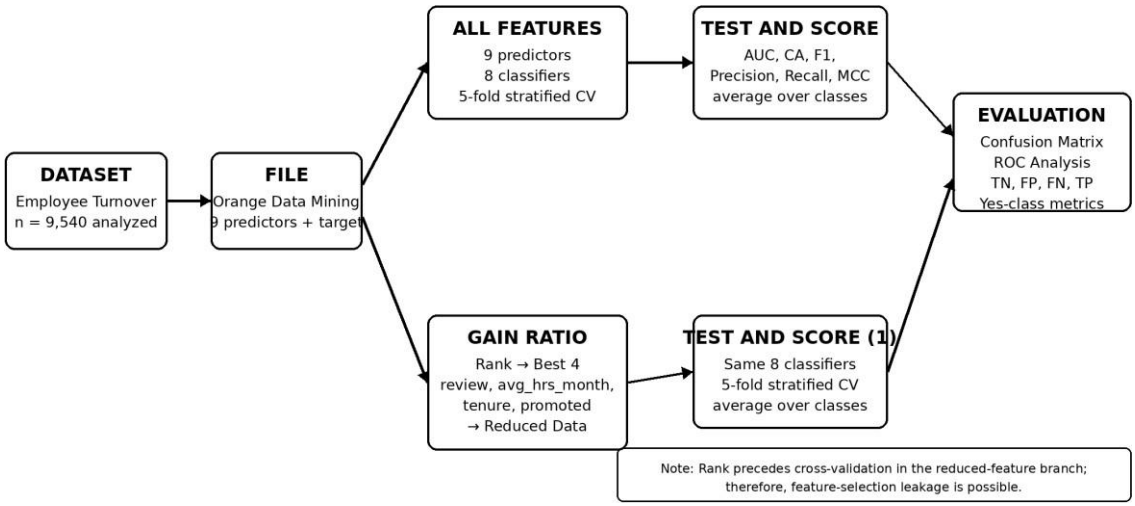


Fig. 1. Machine Learning Analysis Workflow for Employee Turnover Prediction

Fig. 1 presents two analysis branches. The first evaluated all nine predictors directly using eight algorithms and five-fold stratified cross-validation. The second ranked the features using Information Gain Ratio and evaluated review, avg_hrs_month, tenure, and promoted using the same algorithms and validation procedure. Performance was examined through Test and Score, confusion matrices, and ROC analysis. Because Rank precedes cross-validation in the reduced-feature branch, its results were interpreted by considering the potential risk of feature-selection leakage. Data preparation: After import into Orange, the target and nine independent variables were defined. No missing-data imputation was applied because the dataset contained no missing values. The final workflow did not include a separate Preprocess widget; therefore, unverified normalization or manual encoding procedures were not reported. Required data representations were handled within Orange and the learner widgets. Feature importance analysis: Feature selection was performed in Rank using Information Gain Ratio. The four highest scores were review (0.034), avg_hrs_month (0.008), tenure (0.006), and promoted (0.005). These variables were transferred through Reduced Data. Because the final output did not verify Information Gain, Gain Ratio, and Gini as three separate procedures, the method description was limited to the verified Gain Ratio implementation.

2.2 Machine Learning Algorithms

Eight algorithms representing neural-network, instance-based, tree-based, ensemble, margin-based, linear

probabilistic, and Bayesian learning approaches were used to predict employee turnover. The same algorithms were applied in the complete-feature and four-feature analyses.

2.2.1 Artificial Neural Network

Artificial Neural Networks are machine learning models inspired by the information-processing mechanisms of neurons in the human brain. They consist of interconnected artificial neurons and learn complex nonlinear relationships for prediction and classification [22]. A neural network generally consists of an input layer, one or more hidden layers, and an output layer. The input layer receives the predictors, the hidden layers learn relationships among them, and the output layer generates the final prediction. Each neuron weights and sums its inputs and produces an output through an activation function [23]. During training, the error between predicted and actual values is calculated and used to update connection weights, generally through backpropagation [24].

ANN parameters used in this study:

- Hidden-layer neurons: 100
- Activation function: ReLU (Rectified Linear Unit)
- Optimization algorithm: Adam
- Regularization parameter (α): 0.0001
- Maximum iterations: 200

2.2.2 K-Nearest Neighbors

K-Nearest Neighbors is a non-parametric supervised learning method based on similarity among observations [25]. To classify a new observation, it considers the classes of the k nearest observations in the training data. KNN stores the training data rather than estimating a conventional model, allowing local patterns to be considered without distributional assumptions [26].

KNN parameters used in this study:

- Number of neighbors (k): 5
- Distance metric: Euclidean
- Weighting: Uniform

2.2.3 Decision Tree

Decision Trees divide a dataset into branches based on its features and model decisions hierarchically. Beginning at the root node, the model branches through internal nodes and produces classifications at leaf nodes. Criteria such as Gini impurity or information gain are used to determine splits. Their principal advantage is interpretability and visualizability [27].

DT parameters used in this study:

- Splitting criterion: Gini impurity
- Maximum depth: Unlimited
- Minimum number of samples per leaf: 2

2.2.4 Random Forest

Random Forest is an ensemble method comprising multiple decision trees. Each tree is constructed from bootstrap samples, while random feature subsets are considered at node splits. This reduces variance and the risk of overfitting. The final classification is determined through majority voting [28].

RF parameters used in this study:

- Number of trees: 100
- Maximum depth: Unlimited
- Minimum number of samples required to split a node: 5

2.2.5 Support Vector Machine

Support Vector Machine seeks to construct an optimal separating hyperplane between classes. For nonlinearly separable problems, kernel functions map the data into a higher-dimensional space in which an optimal separation can be identified [29].

SVM parameters used in this study:

- Kernel function: RBF (Radial Basis Function)
- Regularization parameter (C): 1.0
- Gamma: Auto

2.2.6 Logistic Regression

Logistic Regression is a parametric statistical method used particularly for binary classification problems. The model uses a linear combination of the independent variables to estimate the probability that an observation belongs to a particular class. The sigmoid function transforms this linear output into a probability between 0 and 1. One advantage of Logistic Regression is the interpretability of its coefficients [30]. LR parameters used in this study:

- Regularization: L2 (Ridge)
- Regularization coefficient (C): 1.0
- Maximum iterations: 100

2.2.7 Naive Bayes

Naive Bayes is a probabilistic classification algorithm based on Bayes' theorem. It calculates the probability that an observation belongs to a particular class from the conditional probabilities of its features. The model assumes that the features are mutually independent. Although this assumption is rarely satisfied completely in real-world datasets, Naive Bayes can produce effective results, particularly for high-dimensional data and limited sample sizes [31].

2.2.8 AdaBoost

AdaBoost is an ensemble learning method that combines multiple weak learners to construct a stronger classifier. At each iteration, the algorithm increases the weights assigned to misclassified observations, causing subsequent learners to focus more heavily on these cases. Through this adaptive procedure, the model learns difficult observations more effectively. The final classification is obtained from the weighted combination of all weak learners [32].

AdaBoost parameters used in this study:

- Base learner: Decision Tree (max_depth = 1)
- Number of estimators: 50
- Learning rate: 1.0

2.3 Model Evaluation Metrics

Classification performance was not assessed using a single metric. Aggregate and class-specific indicators were considered together. AUC, Classification Accuracy (CA), F1, Precision, Recall, and MCC were obtained from Orange's Test and Score widget. TN, FP, FN, and TP values were obtained from the Confusion Matrix outputs and used to calculate Precision, Recall, F1, Specificity, and False Negative Rate (FNR) for the positive “yes” class.

2.3.1 Confusion Matrix

A confusion matrix is one of the principal methods used to evaluate classification performance in detail. It compares the classes predicted by a model with the actual class labels, enabling correct and incorrect classifications to be examined separately for each class. In employee turnover prediction, the confusion matrix identifies which employees were classified correctly or incorrectly and therefore provides information that cannot be obtained from overall accuracy alone [33]. For binary classification, a confusion matrix consists of four components: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). TP indicates that an employee who actually left the organization was correctly predicted as belonging to the turnover class. TN indicates that an employee who remained was correctly predicted as belonging to the non-turnover class. FP occurs when an employee who remained is incorrectly predicted to leave. FN occurs when an employee who left is incorrectly classified as remaining [33].The structure of the confusion matrix used in the analysis is shown in Table 2.

Table 2. Example Confusion Matrix

Actual / Predicted	Turnover (Positive)	No Turnover (Negative)
Turnover (Positive)	TP	FN
No Turnover (Negative)	FP	TN

The TP, TN, FP, and FN values obtained from the confusion matrices were used to calculate the performance measures. The FN value is particularly relevant to employee turnover prediction because a high number of false negatives indicates that employees who actually have a risk of leaving are classified as not being at risk.

2.3.2 Performance Metrics

Accuracy, Precision, Recall, Specificity, F1-score, MCC, and ROC-AUC were used to evaluate classification performance. Because "None, show average over classes" was selected in Test and Score, the reported Precision, Recall, and F1 values were not interpreted as representing the employee turnover ("yes") class alone. Positive-class measures were calculated separately from the TP, TN, FP, and FN counts in the confusion matrices. The definitions and formulas for the performance metrics are presented in Table 3.

Table 3. Classification Performance Metrics and Formulas

Metric	Description	Formula
Accuracy	Proportion of all observations classified correctly.	Accuracy = (TP + TN) / (TP + TN + FP + FN)
Precision	Proportion of positive predictions that are actually positive.	Precision = TP / (TP + FP)
Recall	Proportion of actual positive observations identified correctly.	Recall = TP / (TP + FN)
Specificity	Proportion of actual negative observations classified correctly.	Specificity = TN / (TN + FP)
F1-score	Harmonic mean of Precision and Recall.	F1 = 2 × (Precision × Recall) / (Precision + Recall)
ROC-AUC	Ability to discriminate between the positive and negative classes.	AUC = $\int_0^1$ TPR(FPR) d(FPR)

Joint evaluation of the confusion matrices and aggregate performance measures makes it possible to assess both overall classification performance and the models' ability to identify the employee turnover class [34]. FN and the corresponding FNR received particular attention because they represent employees who actually left but were classified by the model as "no."

2.3.3 ROC Analysis and AUC

Receiver Operating Characteristic (ROC) analysis is a graphical method used to evaluate classifier performance across different decision thresholds. It provides a visual representation of a model's ability to discriminate between classes. The ROC curve plots the True Positive Rate against the False Positive Rate and shows how model performance changes across classification thresholds [34]. The Area Under the Curve (AUC) summarizes a model's overall discriminatory performance as a single numerical value. Values approaching 1 indicate greater discriminatory ability, whereas values close to 0.5 indicate performance near random classification. AUC is therefore widely used when comparing classification models [34].

2.4 Cross-Validation

Five-fold stratified cross-validation was applied in Orange's Test and Score widget to evaluate model generalizability. The dataset was divided into five subsets. In each iteration, four subsets were used for training and the remaining subset was used for testing. The procedure was repeated five times so that each subset served as the test set once. Stratification was used to preserve the class distribution across folds as closely as possible. The performance values reported by Test and Score were based on this cross-validation procedure [35].

3.1. Feature Importance Analysis

Information Gain Ratio was applied in Orange Data Mining's Rank widget to identify the features used in the second analysis. The "Best ranked" option was set to four, and the ranking was performed using all 9,540 observations. The highest Gain Ratio values were obtained for review (0.034), avg_hrs_month (0.008), tenure (0.006), and promoted (0.005). The values displayed for satisfaction, bonus, department, projects, and salary were 0.000.The feature ranking obtained using Information Gain Ratio is presented in Table 4.

Table 4. Feature Ranking Based on Information Gain Ratio

Rank	Feature	Gain Ratio
1	review	0.034
2	avg_hrs_month	0.008
3	tenure	0.006
4	promoted	0.005
5	satisfaction	0.000

6	bonus	0.000
7	department	0.000
8	projects	0.000
9	salary	0.000

This ranking defines the criterion used to select the four variables for the second analysis. Gain Ratio scores indicate the amount of discriminative information that a variable provides in relation to the target class; they were not interpreted as causal effects.

3.2. Model Results Using All Features

All independent variables were used in the first analysis. Eight classification models were evaluated in Orange Data Mining's Test and Score widget using five-fold stratified cross-validation. Because "None, show average over classes" was selected for the target class, the F1, Precision, and Recall values represent average results across classes rather than performance for the "yes" class alone.The aggregate performance results obtained using all features are presented in Table 5.

Table 5. Test and Score Evaluation Results Using All Features

Model	Train	Test	AUC	CA	F1	Prec.	Recall	MCC	Spec.	LogLoss
Neural Network	72.806	0.158	0.913	0.856	0.854	0.853	0.856	0.643	0.772	0.333
kNN	0.261	7.137	0.732	0.742	0.730	0.727	0.742	0.334	0.561	1.826
Tree	6.777	0.004	0.743	0.831	0.829	0.828	0.831	0.583	0.738	4.216
Random Forest	1.249	0.128	0.904	0.853	0.849	0.849	0.853	0.632	0.751	0.482
SVM	6.261	0.947	0.584	0.687	0.630	0.625	0.687	0.076	0.364	0.618
Logistic Regression	4.591	0.067	0.717	0.733	0.678	0.712	0.733	0.238	0.416	0.545
Naive Bayes	0.118	0.016	0.733	0.705	0.661	0.663	0.705	0.162	0.412	0.542
AdaBoost	0.819	0.113	0.786	0.822	0.822	0.822	0.822	0.570	0.750	6.423

The Neural Network produced the highest AUC (0.913), CA (0.856), F1 (0.854), and MCC (0.643) values in the complete-feature analysis. Random Forest achieved 0.904, 0.853, 0.849, and 0.632 for the same measures, respectively. Although the results were similar, the Neural Network produced numerically higher average values. SVM generated the lowest overall discriminatory performance, with an AUC of 0.584 and an MCC of 0.076.

Because the "yes" class required separate evaluation, Precision, Recall, F1, Specificity, and False Negative Rate were calculated for the positive class from the confusion matrices. "Yes" was defined as the positive class and "no" as the negative class.The class-specific performance results for the employee turnover "yes" class are presented in Table 6.

Table 6. Class-Specific Performance for the Employee Turnover "Yes" Class Using All Features

Model	TN	FP	FN	TP	CA	Yes Prec.	Yes Recall	Yes F1	Spec.	FNR
ANN	6,177	579	799	1,985	0.856	0.774	0.713	0.742	0.914	0.287
kNN	5,874	882	1,575	1,209	0.742	0.578	0.434	0.496	0.869	0.566
Tree	6,055	701	911	1,873	0.831	0.728	0.673	0.699	0.896	0.327
Random Forest	6,241	515	891	1,893	0.853	0.786	0.680	0.729	0.924	0.320
SVM	6,171	585	2,402	382	0.687	0.395	0.137	0.204	0.913	0.863
Logistic Regression	6,451	305	2,244	540	0.733	0.639	0.194	0.298	0.955	0.806
Naive Bayes	6,152	604	2,209	575	0.705	0.488	0.207	0.290	0.911	0.793
AdaBoost	5,894	862	838	1,946	0.822	0.693	0.699	0.696	0.872	0.301

The class-specific results demonstrate that average Test and Score metrics and performance for the "yes" class should not be interpreted as equivalent. The Neural Network generated 1,985 true-positive and 799 false-negative

predictions. Its Recall, Precision, and F1 for the "yes" class were 0.713, 0.774, and 0.742, respectively. Random Forest achieved the highest Precision for the "yes" class at 0.786, while its Recall was 0.680. AdaBoost achieved a Recall of 0.699 and was therefore close to the Neural Network in detecting actual turnover cases. The difference between overall accuracy and positive-class performance was more pronounced for SVM, Logistic Regression, and Naive Bayes. SVM produced a Recall of 0.137 and an FNR of 0.863 for the "yes" class. Logistic Regression achieved high Specificity at 0.955, but its Recall for the "yes" class was only 0.194. Naive Bayes produced a Recall of 0.207 and an FNR of 0.793. These findings indicate that using CA alone as the model-selection criterion may conceal the error structure for the positive turnover class.

3.3. Model Results Using Four Features Selected Through Information Gain Ratio

The second analysis used the four highest-ranked variables according to Information Gain Ratio: review, avg_hrs_month, tenure, and promoted. The same eight algorithms were re-evaluated using five-fold stratified cross-validation. Average results across classes were again displayed in Test and Score.The aggregate performance results obtained using the four selected features are presented in Table 7.

Table 7. Test and Score Evaluation Results Using Four Selected Features

Model	Train	Test	AUC	CA	F1	Prec.	Recall	MCC
Neural Network	33.729	0.068	0.866	0.799	0.797	0.796	0.799	0.506
kNN	0.209	1.510	0.808	0.769	0.766	0.764	0.769	0.428
Tree	5.325	0.005	0.654	0.754	0.748	0.745	0.754	0.383
Random Forest	1.331	0.079	0.819	0.768	0.765	0.763	0.768	0.427
SVM	4.554	0.934	0.513	0.517	0.538	0.609	0.517	0.048
Logistic Regression	0.379	0.021	0.697	0.733	0.677	0.715	0.733	0.239
Naive Bayes	0.097	0.016	0.741	0.695	0.643	0.643	0.695	0.113
AdaBoost	0.556	0.030	0.686	0.739	0.740	0.740	0.739	0.371

In the four-feature analysis, the Neural Network produced the highest numerical aggregate results, with an AUC of 0.866, CA of 0.799, F1 of 0.797, and MCC of 0.506. Random Forest ranked second according to AUC at 0.819, while kNN achieved an AUC of 0.808. The differences between kNN (CA = 0.769, F1 = 0.766, MCC = 0.428) and Random Forest (CA = 0.768, F1 = 0.765, MCC = 0.427) were minimal. SVM again produced the lowest overall discriminatory results, with an AUC of 0.513 and MCC of 0.048.The class-specific performance results for the employee turnover "yes" class using four features are presented in Table 8.

Table 8. Class-Specific Performance for the Employee Turnover "Yes" Class Using Four Selected Features

Model	TN	FP	FN	TP	CA	Yes Prec.	Yes Recall	Yes F1	Spec.	FNR
ANN	5,858	898	1,024	1,760	0.799	0.662	0.632	0.647	0.867	0.368
kNN	5,762	994	1,212	1,572	0.769	0.613	0.565	0.588	0.853	0.435
Tree	5,761	995	1,354	1,430	0.754	0.590	0.514	0.549	0.853	0.486
Random Forest	5,740	1,016	1,201	1,583	0.768	0.609	0.569	0.588	0.850	0.431
SVM	3,397	3,359	1,252	1,532	0.517	0.313	0.550	0.399	0.503	0.450
Logistic Regression	6,475	281	2,262	522	0.733	0.650	0.188	0.291	0.958	0.812
Naive Bayes	6,185	571	2,334	450	0.695	0.441	0.162	0.237	0.915	0.838
AdaBoost	5,502	1,254	1,232	1,552	0.739	0.553	0.557	0.555	0.814	0.443

The Neural Network generated 1,760 true positives and 1,024 false negatives. Its Recall, Precision, and F1 for the "yes" class were 0.632, 0.662, and 0.647, respectively. The F1 value for the "yes" class was approximately 0.588 for both kNN and Random Forest. Although SVM's Recall increased to 0.550, the model generated 3,359 false positives, while its Precision and Specificity remained at 0.313 and 0.503. The increase in SVM Recall should therefore not be interpreted in isolation as an overall improvement. Logistic Regression and Naive Bayes classified the negative class with higher Specificity but demonstrated low sensitivity for the positive class. Recall for the "yes" class was 0.188 for Logistic Regression and 0.162 for Naive Bayes. Their FNR values were 0.812 and 0.838, respectively. AdaBoost achieved a Recall of 0.557 and an F1 of 0.555 for the

"yes" class.

3.4. Overall Comparative Evaluation

The changes in model performance between the complete-feature and four-feature analyses are summarized in Table 9.

Table 9. Comparison of the Complete-Feature and Four-Feature Analyses

Model	AUC All Features	AUC Four Features	Δ AUC	CA All Features	CA Four Features	Δ CA
ANN	0.913	0.866	-0.047	0.856	0.799	-0.057
kNN	0.732	0.808	+0.076	0.742	0.769	+0.027
Tree	0.743	0.654	-0.089	0.831	0.754	-0.077
Random Forest	0.904	0.819	-0.085	0.853	0.768	-0.085
SVM	0.584	0.513	-0.071	0.687	0.517	-0.170
Logistic Regression	0.717	0.697	-0.020	0.733	0.733	+0.000
Naive Bayes	0.733	0.741	+0.008	0.705	0.695	-0.010
AdaBoost	0.786	0.686	-0.100	0.822	0.739	-0.083

Feature reduction did not affect all models in the same direction. For the Neural Network, AUC decreased from 0.913 to 0.866 and CA from 0.856 to 0.799. For Random Forest, AUC declined from 0.904 to 0.819 and CA from 0.853 to 0.768. Both principal measures also decreased for Decision Tree and AdaBoost. kNN was the model that benefited numerically from feature reduction. Its AUC increased from 0.732 to 0.808, CA from 0.742 to 0.769, and Recall for the "yes" class from 0.434 to 0.565. Although the AUC of Naive Bayes increased from 0.733 to 0.741, its CA decreased from 0.705 to 0.695 and Recall declined from 0.207 to 0.162. Logistic Regression's CA remained at 0.733 in both analyses. When SVM was reduced to four features, Recall for the "yes" class increased from 0.137 to 0.550, but false positives rose from 585 to 3,359. AUC decreased from 0.584 to 0.513, CA from 0.687 to 0.517, and MCC from 0.076 to 0.048. The increase in positive-class sensitivity was therefore not interpreted as an improvement in overall discriminatory performance.

The Neural Network produced the highest AUC and CA values in both analyses. Model selection, however, should not be based solely on AUC or accuracy. When the managerial cost of false negatives is considered, Recall, Precision, F1, FNR, and Specificity for the positive class should be evaluated jointly. In the complete-feature analysis, the Neural Network, Random Forest, and AdaBoost produced more balanced positive-class results than the remaining models. In the second analysis, the four features were selected using the complete dataset in the Rank widget and subsequently transferred to Test and Score. Because feature selection was conducted outside the cross-validation folds, the resulting performance estimates may be optimistic. The four-feature results were therefore treated as an exploratory comparison of how feature reduction affected model behavior. The cross-validation results obtained with all features were used as the primary basis for evaluating generalization performance. Future validation should use a nested or pipeline-based approach in which feature selection is repeated within each training fold.

IV. DISCUSSION

This study compared eight machine learning algorithms for employee turnover prediction using the same dataset. The models were evaluated through five-fold stratified cross-validation in Orange Data Mining 3.36. Performance varied according to the algorithm, feature set, and evaluation metric. Because Test and Score reported Precision, Recall, and F1 as averages across classes, results specific to the positive "yes" class were calculated separately from the confusion matrices. In the complete-feature analysis, the Neural Network produced the highest numerical aggregate results: AUC = 0.913, CA = 0.856, average F1 = 0.854, average Precision = 0.853, average Recall = 0.856, and MCC = 0.643. Random Forest achieved AUC = 0.904, CA = 0.853, F1 = 0.849, Precision = 0.849, Recall = 0.853, and MCC = 0.632. Although the Neural Network produced numerically higher aggregate performance, the statistical significance of the differences was not tested; statistical superiority should therefore not be claimed. The confusion-matrix results show why aggregate measures alone are insufficient. The Neural Network generated 1,985 true positives and 799 false negatives, with Recall = 0.713, Precision = 0.774, F1 = 0.742, and Specificity = 0.914 for the "yes" class. Random Forest achieved Recall = 0.680, Precision = 0.786, F1 = 0.729, and Specificity = 0.924. The Neural Network detected more actual turnover

cases, whereas Random Forest produced slightly greater positive precision and negative-class specificity. Model choice should therefore reflect the organization's relative costs of false negatives and false positives. AdaBoost and Decision Tree also produced relevant positive-class results. With all features, AdaBoost achieved Recall = 0.699 and F1 = 0.696 for the "yes" class, while Decision Tree achieved 0.673 and 0.699, respectively. AdaBoost generated 1,946 true positives; despite lower aggregate CA and AUC than the Neural Network and Random Forest, its ability to identify turnover cases should be considered. The divergence between overall accuracy and positive-class performance was more pronounced for SVM, Logistic Regression, and Naive Bayes. SVM achieved AUC = 0.584 and CA = 0.687, but its Recall for the "yes" class was 0.137 and its FNR was 0.863. Logistic Regression achieved Specificity = 0.955 but Recall = 0.194, while Naive Bayes produced Specificity = 0.911 and Recall = 0.207. Their relatively high negative-class performance may therefore conceal limited ability to identify employee turnover. Information Gain Ratio ranked review (0.034), avg_hrs_month (0.008), tenure (0.006), and promoted (0.005) as the four leading features. These scores indicate comparatively greater information for distinguishing the classes in this dataset, but they do not establish causal effects. The Neural Network again produced the highest **numerical** aggregate results in the four-feature analysis, with AUC = 0.866, CA = 0.799,

F1 = 0.797, and MCC = 0.506. Random Forest ranked second by AUC at 0.819, while kNN achieved 0.808. Their CA and F1 values were almost identical: kNN achieved CA = 0.769 and F1 = 0.766, whereas Random Forest achieved CA = 0.768 and F1 = 0.765. Their ranking therefore depends on the selected performance metric.

Feature reduction affected the algorithms differently. Neural Network AUC decreased from 0.913 to 0.866 and CA from 0.856 to 0.799; Random Forest AUC decreased from 0.904 to 0.819 and CA from 0.853 to 0.768. Decision Tree and AdaBoost also declined. By contrast, kNN AUC increased from 0.732 to 0.808, CA from 0.742 to 0.769, and F1 from 0.730 to 0.766. This increase does not demonstrate that the excluded variables were noise; additional ablation or repeated-validation analyses would be required. Class-specific results support the same interpretation. Neural Network Recall and F1 for the "yes" class declined to 0.632 and 0.647, whereas kNN Recall increased from 0.434 to 0.565. Recall was 0.569 for Random Forest, 0.557 for AdaBoost, and 0.514 for Decision Tree. Although SVM Recall increased to 0.550, it generated 3,359 false positives; its Precision fell to 0.313, Specificity to 0.503, AUC to 0.513, and MCC to 0.048. SVM therefore detected more positive cases at the cost of many false-positive predictions. The findings support a multi-metric approach. If the objective is to identify employees at risk of leaving, Recall and FNR should receive greater weight. If reducing unnecessary warnings is the priority, Precision and Specificity become more relevant. AUC should be considered for overall discrimination, while MCC and F1 provide information about classification balance. The Neural Network offered a favorable balance between aggregate performance and positive-class sensitivity, whereas Random Forest provided higher positive-class Precision and Specificity. The results are generally consistent with studies reporting that machine learning and ensemble methods can perform effectively in employee turnover prediction [3, 4, 5]. Performance values across studies should not be treated as directly equivalent because sample size, class distribution, features, preprocessing, hyperparameters, and validation design can substantially affect results. The present findings should be interpreted within the dataset and experimental design used. Prediction models may serve as early-warning tools that support human resource professionals. A high-risk prediction should not be converted into an automatic personnel decision. When less interpretable models such as neural networks are used, organizations should consider explainability, employee privacy, data security, fairness, and discrimination risks. Model outputs should be assessed together with human judgment. A methodological limitation concerns the feature-selection workflow. Features were selected from the complete dataset in Rank before Reduced Data was transferred to cross-validation in Test and Score. Because supervised feature selection was conducted outside the validation folds, this workflow introduces a risk of data leakage. The four-feature results should therefore be treated as exploratory. A stricter estimate would repeat feature selection within every training fold. Other limitations include the use of a single open-access dataset, the absence of external validation, and the lack of systematic hyperparameter optimization. The dataset also omits variables such as organizational commitment, burnout, leadership, work–life balance, and manager–employee relationships. Future studies may use data from different organizations and sectors, repeated or nested cross-validation, external or temporal validation, systematic hyperparameter optimization, class weighting, and cost-sensitive learning. Reporting means with standard deviations or confidence intervals would permit more reliable comparison. The Neural Network produced the highest numerical aggregate cross-validation performance for this dataset, while Random Forest provided comparable aggregate performance with higher positive-class Precision and Specificity. Model selection should jointly consider false-negative and false-positive costs, explainability, computational requirements, and intended organizational use rather than assuming a universally best model.

V. CONCLUSION

This study compared Neural Network, kNN, Decision Tree, Random Forest, SVM, Logistic Regression, Naive Bayes, and AdaBoost using an employee turnover dataset containing 9,540 observations. The models were evaluated through five-fold stratified cross-validation. The first analysis used all features, whereas the second used review, avg_hrs_month, tenure, and promoted, selected according to Information Gain Ratio. Performance for the positive "yes" class was calculated separately from the confusion matrices in addition to the aggregate Test and Score metrics. In the complete-feature analysis, the Neural Network produced the highest numerical aggregate results, with AUC = 0.913, CA = 0.856, F1 = 0.854, and MCC = 0.643. Random Forest achieved AUC = 0.904, CA = 0.853, F1 = 0.849, and MCC = 0.632. For the positive class, Neural Network Recall was 0.713 and FNR was 0.287; Random Forest Recall was 0.680 and FNR was 0.320. Random Forest had higher positive-class Precision (0.786 versus 0.774) and Specificity (0.924 versus 0.914). Their relative merits should therefore be assessed according to institutional error costs. AdaBoost and Decision Tree produced positive-class Recall values of 0.699 and 0.673, respectively. The corresponding values for SVM, Logistic Regression, and Naive Bayes were 0.137, 0.194, and 0.207. Overall accuracy is therefore insufficient as a standalone performance measure when the positive class has direct managerial relevance. In the four-feature analysis, the Neural Network again produced the highest numerical aggregate performance, with AUC = 0.866 and CA = 0.799. Random Forest ranked second by AUC at 0.819. kNN achieved CA = 0.769 and F1 = 0.766, close to Random Forest's CA = 0.768 and F1 = 0.765. kNN was the principal model to benefit from feature reduction, with AUC increasing from 0.732 to 0.808; performance declined for Neural Network, Random Forest, Decision Tree, and AdaBoost. Although SVM Recall for the "yes" class increased to 0.550, it generated 3,359 false positives. Its positive-class Precision was 0.313, Specificity was 0.503, AUC was 0.513, and MCC was 0.048. An increase in a single class-specific metric should not be interpreted as an improvement in overall performance. Likewise, the high Specificity of Logistic Regression and Naive Bayes should be considered together with their low positive-class Recall. The principal methodological conclusion is that AUC, CA, F1, MCC, Precision, Recall, Specificity, and False Negative Rate should be considered jointly. Recall and FNR are more relevant when failing to identify employees at risk is costly, whereas Precision and Specificity should receive greater weight when false alarms are costly. Because feature selection was performed before cross-validation, the four-feature workflow involves a risk of data leakage. Future research should use a nested validation design in which feature selection is repeated within each training fold. In practice, the Neural Network and Random Forest may be considered candidate models for employee turnover early-warning systems. Their outputs should not be converted directly into automated personnel decisions. Predictions should support assessments by human resource professionals, with due consideration for explainability, employee privacy, data security, fairness, and discrimination risks. Future studies should conduct external and temporal validation using datasets from different organizations and sectors. They should implement systematic hyperparameter optimization, class weighting, and cost-sensitive learning; report performance with standard deviations or confidence intervals; and employ explainable artificial intelligence methods such as SHAP and LIME. Machine learning models can support decisions concerning employee turnover risk, but their performance cannot be summarized adequately by a single accuracy value. For the dataset and experimental design used in this study, the Neural Network produced the highest numerical aggregate performance, while Random Forest provided a strong alternative. The final model should be selected by jointly considering positive-class detection, false-alarm and missed-case costs, explainability, and organizational implementation conditions.

REFERENCES

[1]. Shafique, M., Fatima, M., & Shafique, A. (2024). Employee turnover models: A review of conventional and contemporary approaches and future research direction. Leadership and Organizational Behavior Journal, 4(1), 42–60.

[2]. Luamba, D. S., Chagadama, J., & Blye, M. L. J. (2026). Mitigating employee turnover and increasing customer satisfaction: A case study of the private security industry. American International Journal of Business Management, 9(1), 13–24.

[3]. Taner, Z., Areta Hızıroğlu, O., & Hızıroğlu, K. (2024). Leveraging machine learning methods for predicting employee turnover within the framework of human resources analytics. Journal of Intelligent Systems: Theory and Applications, 7(2), 145 –158. https://doi.org/10.38016/jista.1440879

[4]. Al Akasheh, M., Hujran, O., Faisal Malik, E., & Zaki, N. (2024). Enhancing the prediction of employee turnover with knowledge graphs and explainable AI. IEEE Access, 12, 77041–77053. https://doi.org/10.1109/ACCESS.2024.3404829

[5]. Talebi, H., Khatibi Bardsiri, A., & Khatibi Bardsiri, V. (2025). Machine learning approaches for predicting employee turnover: A systematic review. Engineering Reports. https://doi.org/10.1002/eng2.70298

[6]. Mutlu, M., Kurt, B. S., Öven, Ç., & Şen, N. (2026). Paternalist liderlik, iş performansı ve işten ayrılma niyeti arasındaki ilişkilerin incelenmesi: Cinsiyetin düzenleyici rolü. Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 28(1), 485 –503. https://doi.org/10.16953/deusosbil.1670573

[7]. Aykan, E., Karakuş, G., Akçadağ, M., & Çataltepe, O. (2026). Yaşam tatmini düzenleyici rolde iken çalışan yorgunluğunun çalışanların işten ayrılma niyeti ve performansları üzerindeki etkisinde iş tatmininin aracı rolü. Uluslararası İktisadi ve İdari Akademik Araştırmalar Dergisi, 6(1), 1–29.

[8]. Polat, S., & Ulus, M. (2026). Türkiye'de işçi devri ve yeni işlerin getirisi: 2005–2019 Hanehalkı İşgücü Anketlerinden çıkarımlar. Ekonomi-tek, 15(1), 159–191. https://doi.org/10.65653/ekonomitek.1851833

[9]. Hudays, A., Schiltz, N. K., Alrashidi, M., Arishi, A., Khormi, J., Darraj, A., Alkhadrah, A., Flimban, A., Aljohani, R., Sailah, M. A.,

Bahari, G., & Alqahtani, N. (2025). Machine learning models for predicting nurse turnover and turnover intention: A systematic review. Saudi Journal of Nursing and Health Care, 8(6), 148–162. https://doi.org/10.36348/sjnhc.2025.v08i06.003
[10]. Bilenler, B., Gül, S., & Uçar, T. (2024). Makine öğrenmesi teknikleri ile işten ayrılacak personelin tahminlenmesi ve tekniklerin performanslarının karşılaştırılması. Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi, 15(4), 807–816. https://doi.org/10.24012/dumf.1517176
[11]. Avrahami, D., Pessach, D., Singer, G., & Chalutz Ben-Gal, H. (2022). A human resources analytics and machine-learning examination of turnover: implications for theory and practice. International Journal of Manpower, 43(6), 1405–1424. https://doi.org/10.1108/IJM-12-2020-0548
[12]. Alpsoy, A., & Tükel, D. (2026). Çalışan devir hızını tahmin etmede CNN ve XGBoost algoritmaları ile geliştirilen hibrit model önerisi. Doğuş Üniversitesi Dergisi, 27(1), 1–18. https://doi.org/10.31671/doujournal.1547333
[13]. Alıcı, A., Ertek, A., & Aktemur, Ş. (2026). Havayolu sektöründe çalışan odaklı sosyal ve finansal performans analizi: Bütünleşik LOPCOW-COCOSO yaklaşımı. Black Sea Journal of Engineering and Science, 9(1), 334–350. https://doi.org/10.34248/bsengineering.1831229
[14]. Sönmez, A. R., & Üngüren, E. (2026). Kuşak farklılıkları bağlamında iş tatmini ve işe adanmışlık ilişkisi: Otel çalışanları üzerine ampirik bir araştırma. Alanya Akademik Bakış, 10(1), 196–216. https://doi.org/10.29023/alanyaakademik.1684747
[15]. Bakal, C., Ayden, C., Kişman, Z. A., & Eşidir, O. V. (2026). İşe alım süreçlerinde yapay zeka kullanımının avantajları ve dezavantajları. Iğdır Üniversitesi Sosyal Bilimler Dergisi, 41, 211–231. https://doi.org/10.54600/igdirsosbilder.1807282
[16]. Yıldız, E. (2026). Yöneticilerin gözünden işletmelerde sessiz vazgeçiş. Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 35, 1–17. https://doi.org/10.35379/cusosbil.1609772
[17]. Ekiz Kavukoğlu, T. (2026). Özel hastanelerin insan kaynakları politikalarının incelenmesi: Kurumsal web siteleri üzerinden bir analiz. Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Dergisi, 27(1), 350–372. https://doi.org/10.37880/cumuiibf.1811366
[18]. Serbes, K. (2026). Yapay zekâ teknolojilerinin insan kaynakları yönetimine etkisi: Fırsatlar, zorluklar ve tehditler. İstanbul Aydın Üniversitesi Sosyal Bilimler Dergisi, 18(1), 1–35. https://doi.org/10.17932/IAU.IAUSBD.2021.021/iausbd_v18i1001
[19]. Kılıç Özkaynar, G. (2026). İşgören performansını etkileyen faktörler ve iş mutluluğu ilişkisinde sosyal medya kullanımının etkisi: TR72 Bölgesi tekstil sektörü örneği. Gümüşhane Üniversitesi Sosyal Bilimler Dergisi, 17(1), 195–212.
[20]. Canlı, E. (2026). Employee turnover [Data set]. Kaggle. https://www.kaggle.com/datasets/elifcanli/employee-turnover/data
[21]. Demšar, J., Curk, T., Erjavec, A., Gorup, Č., Hočevar, T., Milutinović, M., Možina, M., Polajnar, M., Toplak, M., Starič, A., Štajdohar, M., Umek, L., Žagar, L., Žbontar, J., Žitnik, M., & Zupan, B. (2013). Orange: Data mining toolbox in Python. Journal of Machine Learning Research, 14, 2349–2353.
[22]. Haykin, S. (2009). Neural networks and learning machines (3rd ed.). Pearson Education.
[23]. Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. Neural Networks, 2(5), 359–366. https://doi.org/10.1016/0893-6080(89)90020-8
[24]. Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323(6088), 533–536. https://doi.org/10.1038/323533a0
[25]. Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 13(1), 21–27. https://doi.org/10.1109/TIT.1967.1053964
[26]. Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. The American Statistician, 46(3), 175–185. https://doi.org/10.1080/00031305.1992.10475879
[27]. Quinlan, J. R. (1986). Induction of decision trees. Machine Learning, 1(1), 81–106. https://doi.org/10.1007/BF00116251 [28]. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324
[29]. Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297. https://doi.org/10.1007/BF00994018
[30]. Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression (3rd ed.). John Wiley & Sons.
[31]. McCallum, A., & Nigam, K. (1998). A comparison of event models for naive Bayes text classification. In AAAI-98 Workshop on Learning for Text Categorization (Vol. 752, pp. 41–48). AAAI Press.
[32]. Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, 55(1), 119–139. https://doi.org/10.1006/jcss.1997.1504
[33]. Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. Information Processing & Management, 45(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002
[34]. Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010
[35]. Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of the 14th International Joint Conference on Artificial Intelligence (Vol. 2, pp. 1137–1143). Morgan Kaufmann.