August 2026| Volume 05 | Issue 03 | Pages 319-330

Future Technology
Open Access Journal
ISSN 2832-0379

Journal homepage: https://fupubco.com/futech

https://doi.org/10.55670/fpll.futech.5.3.28

Future Publishing LLC

Article

An explainable AI-based workforce intelligence framework for integrating future skill demand and employee attrition prediction with risk-aware decision analytics

Prathap D L[1]*, Thimmaraju S N[2]

Department of CS & E, VTU-RRC, PG Centre, Karanatks, India

ARTICLE INFO

Article history:
Received 24 January 2026
Received in revised form
26 May 2026
Accepted 30 June 2026

Keywords:
Explainable AI, Workforce intelligence, Employee attrition, Skill demand analysis, Machine learning, Risk-aware decision analytics

*Corresponding author
Email address:
prathapadl28@gmail.com

DOI: 10.55670/fpll.futech.5.3.28

ABSTRACT

Artificial intelligence supports workforce analytics by improving skill assessment, attrition prediction, and talent planning. However, external labor-market skill demand and internal employee attrition risk are often analyzed separately. This study presents a workforce intelligence framework based on explainable AI that combines skill-demand cluster analysis, attrition prediction, explainability (SHAP), and aggregate risk-aware decision support. Two public datasets were used: the Jobs and Skills Mapping for Career Analysis dataset and the IBM HR Analytics Employee Attrition Dataset. TF-IDF was applied to job-related text to generate clusters of job-required skills and to predict auxiliary pay grades, and Logistic Regression, Random Forest, and XGBoost were evaluated for attrition prediction. Logistic Regression was the best-performing model for identifying the risk of attrition, with a recall of 0.6170, an F1-score of 0.4328, and an ROC-AUC of 0.7954. The best recall and F1-score were obtained at a threshold of 0.40, with values of 0.7872 and 0.4901, respectively, as determined by threshold analysis. Over time, SHAP identified frequent business travel, job level, lab technician position, and total years of work as important factors in attrition. The Workforce Risk Score was a combination of normalized skill demand and normalized aggregate attrition risk, with the highest-ranked skill-demand cluster being moderate at 0.396. The framework provides actionable summary-level decision support for workforce planning.

1. Introduction

The rapid advancement of artificial intelligence (AI) and machine learning (ML) has shifted workforce analytics from descriptive reporting toward predictive and decision-support applications. Organizations increasingly rely on data-driven methods to understand employee behavior, manage talent more effectively, and anticipate future workforce requirements. In this context, interpretability has become essential because algorithmic outputs used in employment, retention, and workforce planning must be transparent and reliable. Early research in explainable machine learning emphasized the need to make complex models understandable in high-stakes contexts [1], while SHAP provided a consistent method for explaining model predictions via feature-level contributions [2]. These developments have strengthened the argument that interpretable systems are preferable to opaque black-box models in high-impact applications [3], and broader interpretability frameworks have also been formalized for practical use [4]. Machine learning adoption in organizations has increased due to scalable algorithms and widely available computational tools. Libraries such as Scikit-learn have simplified model development [5], while ensemble and boosting methods, such as XGBoost and LightGBM, have improved predictive modeling of structured HR data [6,7]. These developments are relevant in labor markets shaped by automation, digitalization, and changing skill requirements. As job roles evolve, organizations need to monitor external skill demand while maintaining internal workforce capability. Autor [8] argued that technological change does not necessarily make human skills obsolete, but reshapes their composition, making continuous skill monitoring and workforce adjustment necessary. Nevertheless, as much as these have been achieved, a key difficulty is that external labor market trends and internal employee retention are usually studied independently. Job market data show demand for skills, while attrition models focus on internal organizational variables. This division restricts workforce planning, preventing a holistic view of talent requirements, skills shortages, and retention risks. Consequently, organizations might have difficulty integrating hiring, upskilling, and retention plans with broader labor market trends. Three gaps are identified in the literature. First, external skill-demand analysis and internal attrition prediction are rarely integrated within a single framework. Second, although machine learning models can achieve strong predictive performance, limited interpretability remains a barrier to trust and adoption in HR settings. Third, existing approaches often emphasize prediction accuracy rather than decision-oriented analytics that can support managerial action.

To address these gaps, this study proposes an explainable AI-based workforce intelligence framework that integrates external job market analytics with internal HR data. The framework uses natural language processing to identify skill-demand patterns in job-related text and applies machine learning models to predict employee attrition. Explainable AI, including SHAP, is used to improve transparency, while the Workforce Risk Score converts predictive outputs into aggregate decision-support information. This study's scientific contribution lies not in the creation of a novel machine learning algorithm, but in the formulation of an integrated and explicable workforce intelligence framework that interlinks three analytical components: external skill-demand cluster analysis, internal employee attrition prediction, and comprehensive risk-aware decision support. The proposed paradigm integrates both attrition prediction and labor-market skill analysis inside a unified decision-support system, in contrast to research that addresses only one of these aspects. The framework provides four primary contributions. Initially, it uses job market data to pinpoint skill-demand clusters that reflect external workforce needs. Second, it applies supervised machine learning models to predict the risk of internal employee attrition. Third, it uses SHAP-based explainability to identify the main drivers of attrition predictions. Fourth, it introduces a Workforce Risk Score that combines normalized skill demand and normalized aggregate attrition risk to support strategic prioritization. The score is intended as an aggregate decision-support index rather than a causal employee-level prediction model.

This contribution is therefore positioned as a framework-level integration of predictive modeling, explainability, and workforce decision analytics. It supports practical workforce planning by linking skill-demand information, attrition-risk modeling, and interpretable decision-support outputs. The research objectives are summarized as:

- To develop an integrated AI-based framework combining job market analysis and employee attrition prediction.
- To predict employee attrition using machine learning models with optimized performance.
- To provide explainable insights and a Workforce Risk Score for decision-oriented workforce analytics.

2. Literature review

The use of machine learning (ML) to predict employee attrition has been increasing rapidly, driven by the growing volume of organizational data and the need to proactively manage the workforce. Initial research assessed the predictive power of classification models, and Sisodia et al. [9] demonstrated that ensemble approaches tend to outperform single classifiers and that preprocessing and feature selection are essential for achieving good results. The predictive analytics underpin the predictive HR analytics of Alaskar et al. [10], who further demonstrated the managerial relevance of attrition modeling by linking predictive analytics to minimized turnover-related costs. The field was further developed by subsequent studies, which introduced more effective algorithms, optimization methods, and strategies. Raza et al. [11] have indicated that nonlinear workforce patterns are better captured by optimized ML models, whereas Iyiade et al. [12] have revealed important trends in the literature, such as the emergence of ensemble learning, the expanded use of real-world HR datasets, and growing interest in explainability. Their review also highlighted the continued methodological concerns of the imbalance between classes, heterogeneous features, and inconsistent evaluation practices, which show that performance improvement will not help with general analytical limitations.

Interpretability has become as important as predictive performance in HR analytics. Explainable AI can reveal the key factors driving employee turnover and improve trust in model outputs. Konar et al. [13] also showed that Bayesian-optimized stacked ensemble models can support both predictive performance and interpretability. Marín Díaz et al. [14] highlighted XAI's ability to translate model behavior into retention-relevant insights, while AL-Ali et al. [15] demonstrated the practical use of ML and XAI in HR analytics. More broadly, Sakib and Islam [16] showed that AI-based HR research is expanding across recruitment, performance management, and workforce planning, indicating that attrition prediction is now part of a broader intelligent decision-making system. Workforce intelligence also requires greater attention to external labor-market dynamics. Although internal HR analytics have been developed significantly, external skill-demand analysis remains loosely related to attrition modeling. NLP has facilitated the identification of emerging skills and labor market patterns from job advertisements and associated textual data, but these findings are seldom combined with internal workforce analytics. This lack of connection limits the strategic use of current systems, as organizations are forced to handle internal retention and external changes in required competencies simultaneously.

Model evaluation is also central to workforce analytics. Fawcett [17] stated that ROC analysis is an effective method for imbalanced classification, but Sokolova et al. [18] argued that multi-metric assessment is more effective than one-dimensional accuracy. The relevance of the principles is particularly high in attrition studies, where precision-recall trade-offs strongly affect managerial usefulness. Simultaneously, the developer of the powerful algorithm for high-dimensional, interaction-rich data, Breiman [19], introduced Random Forests as an effective ensemble learning method for such data.

Although evident improvements have been made, three constraints remain. To begin with, most studies examine either internal HR data or external labor market data in isolation. Second, explainable AI is often implemented at the model level but is not integrated into broader workforce intelligence systems. Third, the current literature is focused on prediction rather than decision support. Therefore, the literature justifies the need to integrate AI-based workforce intelligence technologies that combine ML to predict attrition, NLP to analyze skill demands, and explainable AI to interpret these predictions transparently. This type of framework may enable a more successful alignment between internal retention risk and external labor market change, as well as inform a more workforce strategy that considers risk.

3. Methodology

This study proposes an explainable artificial intelligence-based workforce intelligence framework that integrates external job market analytics with internal workforce behavior modeling to support risk-aware decision-making. The methodology combines natural language processing techniques for skill demand extraction with supervised machine learning models for employee attrition prediction, followed by a strategic integration layer that enables workforce risk assessment.

3.1 Proposed framework

The proposed framework is in three layers: a data input layer, an analytics layer, and an explainability and decision-support layer. The skill-demand analysis module identifies common clusters of skill-demand in the job data set using job-related text. The attrition prediction module helps analyze employee turnover behavior using structured HR data. The decision-support layer combines standardized skill demand and standardized aggregate attrition risk to derive the Workforce Risk Score, enabling risk-informed workforce planning. The overall design of the proposed system is shown in Figure 1. It has three main layers: Data input layer, Analytics layer, and Decision-support layer. Data Input layer – This consists of external Job and internal HR attrition data. The analytics layer analyzes job posting text to identify skill-demand clusters and uses supervised machine learning models to predict attrition. The decision-support layer combines the standardized skill demand and standardized aggregate attrition risk through the Workforce Risk Score. Explainability is used with SHAP to provide insights into the attrition model and understand the key drivers of the expected employee attrition. The system integrates external labor market data with internal HR attrition data to provide comprehensive decision support. The skill-demand module identifies skill-demand clusters from job-related text by preprocessing and applying TF-IDF, whereas the attrition module forecasts employee attrition risk using supervised machine learning models. SHAP is used to elucidate attrition forecasts, while the Workforce Risk Score combines normalized skill demand and normalized aggregate attrition risk to support informed workforce decision-making.

3.2 Data description

Two publicly accessible datasets on Kaggle were used to capture both external labor market dynamics and internal organizational workforce characteristics.

Jobs dataset: The first dataset, Jobs and Skills Mapping for Career Analysis, contains information on job markets, such as job titles, job descriptions, required skills, industry types, and pay-grade categories [20]. This data is particularly oriented for use in job analytics, skills, and career intelligence systems, to discover patterns in skills demand and industry trends.

HR dataset: The second dataset, IBM HR Analytics Employee Attrition Dataset, comprises employee-level organizational data, including demographic attributes, work-related data, and an attrition indicator [21]. The dataset is widely used in predictive analytics for workforce loss and personnel analytics.

Dataset summary and selection rationale: This study used two publicly available datasets to capture two dimensions of workforce intelligence: external skill demand and internal employee attrition. The Jobs and Skills Mapping for Career Analysis dataset was selected because it contains job titles, descriptions, skill-related text, industry information, and pay-grade categories, making it suitable for skill-demand cluster analysis. The IBM HR Analytics Employee Attrition Dataset was selected because it is a widely used benchmark dataset for employee attrition prediction and includes demographic, job-related, satisfaction-related, and attrition-label variables. Table 1 summarises the datasets used in the study.

Although the datasets are useful for reproducible workforce analytics, they have limitations. The IBM HR dataset represents a single organizational context, whereas the jobs dataset does not provide real-time labor market information. In addition, the two datasets do not share employee-level identifiers, job IDs, organizational IDs, or directly matched skill inventories. Therefore, the integration was performed at an aggregate decision-support level rather than through direct record-level merging. Accordingly, the Workforce Risk Score is interpreted as a strategic prioritization index, not as a causal estimate of employee-level risk.

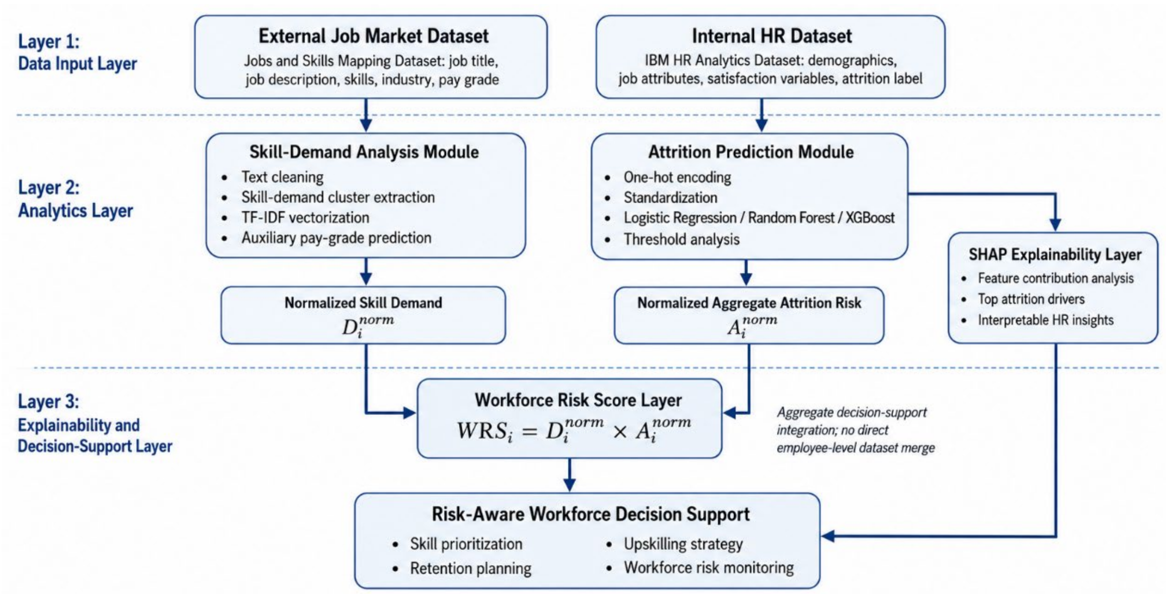

Figure 1. Proposed explainable AI-based workforce intelligence framework

Table 1. Summary of datasets used in the study

Dataset	Source	Records	Features/ Columns	Target Variable	Class Distribution	Missing Values	Duplicate Rows	Purpose
Jobs and Skills Mapping for Career Analysis	Kaggle	970	6	Pay_grade	Average paying: 545; High paying: 285; Low paying: 140	0	0	Skill-demand cluster analysis and auxiliary pay-grade prediction
IBM HR Analytics Employee Attrition Dataset	Kaggle	1470	35	Attrition	No attrition: 1233; Attrition: 237	0	0	Employee attrition prediction and explainability analysis

3.3 Data preprocessing and feature engineering

Both datasets were pre-processed separately because they served different analytical purposes. The jobs dataset was used for skill-demand cluster analysis and auxiliary pay-grade prediction, while the HR dataset was used for employee attrition prediction. For the jobs dataset, text fields were converted to lowercase, unnecessary punctuation was removed, and whitespace was standardized. Job title, short description, and skill-related text were combined into a single textual feature. The skills_required field was used to extract skill-demand clusters. Since the dataset contains grouped skill phrases rather than consistently separated individual skills, the extracted outputs were treated as skill-demand clusters rather than atomic skills. No external skill ontology, synonym normalization, stemming, lemmatization, or named entity recognition model was applied. The combined job-text feature was converted into numerical form using TF-IDF vectorization with English stop-word removal, unigram representation, and a maximum of 500 features. Since the jobs dataset does not contain time-stamped postings, the study does not detect temporal trends. Therefore, the results are interpreted as skill-demand patterns observed in the dataset. For the HR dataset, the attrition label was converted into a binary target variable, where attrition was coded as 1 and non-attrition as 0. Categorical variables were one-hot encoded, and numerical variables were standardized. To avoid data leakage, the train-test split was performed before scaling; the scaler was fitted only on the training data and then applied to the test data. A stratified 80:20 train-test split with a random seed of 42 was used. Missing-value and duplicate-record checks were performed, and no missing values or duplicate rows were found in the cleaned datasets.

3.4 Data preprocessing and leakage control

Both datasets were used to train supervised machine learning models for classification tasks. For the job's dataset, classification models were trained to predict pay-grade categories using textual features from job descriptions and skill-demand clusters. For the HR dataset, employee attrition was predicted using organizational and demographic variables.

Pay-grade prediction as an auxiliary task: Pay-grade prediction was included as an auxiliary classification task for the job's dataset. Its purpose was to examine whether job descriptions and skill-demand cluster features contain meaningful workforce intelligence signals that can distinguish compensation-related categories. This task supports the evaluation of the textual feature representation derived from the job's dataset. However, pay-grade prediction was not used directly in the Workforce Risk Score. The primary role of the jobs dataset in the proposed framework was to identify and rank skill-demand clusters, while pay-grade prediction served as a supporting analytical task.

Model configuration and validation strategy: Logistic Regression, Random Forest, and XGBoost were employed for both classification tasks. A stratified 80:20 train-test split with a random seed of 42 was employed to maintain class distribution between training and testing groups. In addition, stratified 5-fold cross-validation was performed for the HR attrition models to assess robustness. Table 2 summarises the main hyperparameters used in the models. No exhaustive grid search, random search, or Bayesian hyperparameter optimization was performed. The models were trained using fixed hyperparameter settings, while class imbalance was handled using class_weight='balanced' for Logistic Regression and Random Forest, and scale_pos_weight for XGBoost.

Since the attrition dataset is imbalanced, evaluation was not based solely on accuracy. Precision, recall, F1-score, and ROC-AUC were used to assess model performance. Synthetic oversampling methods such as SMOTE and ADASYN were not applied because the HR dataset is relatively small, and synthetic examples may alter the original distribution of employee-related attributes. Instead, class weighting and threshold analysis were used to improve the identification of attrition-risk cases.

3.5 Model evaluation and optimization

Various classification metrics, such as accuracy, precision, recall, F1-score, and ROC-AUC, were used to assess model performance. Since there is a class imbalance in the attrition dataset, particular attention was paid to recall and F1-score to effectively identify at-risk employees. The attrition model was optimized using a threshold to improve its predictive power. The default probability threshold was not used, but a set of threshold values was tested to find the best decision boundary.

Table 2. Model hyperparameter settings

Dataset	Model	Key Hyperparameters
Jobs	Logistic Regression	max_iter=2000, random_state=42
Jobs	Random Forest	n_estimators=200, class_weight='balanced', random_state=42, n_jobs=-1
Jobs	XGBoost	n_estimators=200, max_depth=6, learning_rate=0.1, subsample=0.9, colsample_bytree=0.9, objective='multi:softprob', eval_metric='mlogloss', random_state=42
HR Attrition	Logistic Regression	max_iter=2000, class_weight='balanced', random_state=42
HR Attrition	Random Forest	n_estimators=200, class_weight='balanced', random_state=42, n_jobs=-1
HR Attrition	XGBoost	n_estimators=200, max_depth=6, learning_rate=0.1, subsample=0.9, colsample_bytree=0.9, scale_pos_weight=negative_count/positive_count, objective='binary:logistic', eval_metric='logloss', random_state=42

The maximum F1-score threshold was chosen to guarantee a balanced trade-off between precision and recall and to enhance the model's practical relevance in identifying workforce risks.

3.6 Explainable AI and workforce risk scoring

To improve model transparency, SHAP was applied to the selected Logistic Regression attrition model. Logistic Regression was selected for explanation because it provided the best balance among recall, F1-score, and ROC-AUC for the attrition prediction task. Since the selected model is linear, SHAP LinearExplainer was used to estimate the contribution of each feature to the model prediction. Global feature importance was calculated using mean absolute SHAP values, which helped identify the most influential factors affecting attrition-risk predictions. The Workforce Risk Score was formulated as an aggregate decision-support index that combines normalized skill-demand intensity with normalized attrition-risk indication. Because the jobs dataset and HR dataset do not share employee-level identifiers, job IDs, organizational IDs, or directly matched skill inventories, the score was calculated at an aggregate level. Therefore, the Workforce Risk Score should be interpreted as a strategic prioritization measure rather than as a causal employee-level prediction. The Workforce Risk Score for the skill-demand cluster i is defined as:

$$WRS_i = D_i^{norm} \times A_i^{norm} \quad (1)$$

where WRS_i represents the Workforce Risk Score for the skill-demand cluster i, D_i^{norm} represents the normalized skill-demand value for the cluster i, and A_i^{norm} Represents the normalized aggregate attrition-risk indicator. Skill demand was normalized by dividing each skill-demand cluster frequency by the maximum observed cluster frequency:

$$D_i^{norm} = \frac{D_i}{D_{max}} \quad (2)$$

The attrition-risk indicator was derived from aggregate job-role-level attrition patterns in the HR dataset and normalized relative to the maximum observed attrition rate:

$$A_i^{norm} = \frac{A_i}{A_{max}} \quad (3)$$

The final Workforce Risk Score was obtained by multiplying the normalized skill-demand value and the normalized attrition-risk indicator. Risk categories were assigned using the following thresholds: High risk for scores of 0.66 and above, Moderate risk for scores from 0.33 to 0.65, and Low risk for scores below 0.33. In the revised aggregate scoring output, no High-risk cluster was identified because the maximum observed Workforce Risk Score was 0.396. This reflects the conservative aggregate-level design of the proposed score.

3.7 Integration strategy

A strategic-level integration was performed rather than a record-level merge due to the lack of a direct mapping between job market data and organizational workforce data. The skill demand analysis and attrition prediction modules aggregated their insights to generate decision-oriented outputs. This methodology will ensure methodological consistency and avoid unrealistic assumptions about dataset alignment, thereby enabling the development of a scalable workforce intelligence framework.

4. Results

4.1 Skill demand analysis

The Jobs and Skills Mapping dataset was used to identify prevailing skill-demand groups in employment-related text. Following preprocessing and TF-IDF-based feature representation, the Skills_required field was examined to identify recurring patterns in skill-demand clusters across various employment roles. The findings indicate that multiple clusters integrate concepts associated with communication, analytics, technology, leadership, and reporting. These clusters signify observed skill-demand patterns in the dataset rather than temporal industry trends. Figure 2 illustrates the predominant skill-demand groups derived from the employment information.

4.2 Attrition prediction results

The IBM HR Analytics Attrition Dataset was used to train and evaluate Logistic Regression, Random Forest, and XGBoost models for employee attrition prediction. In addition, the Jobs and Skills Mapping dataset was evaluated through the auxiliary pay-grade prediction task. Model performance was assessed using accuracy, precision, recall, F1-score, and ROC-AUC. The results are presented in Table 3. For the HR attrition task, Logistic Regression achieved the highest recall, F1-score, and ROC-AUC among the evaluated models, while XGBoost achieved the highest accuracy and precision. Since the objective of attrition modeling is to identify employees at risk of leaving, recall and F1-score are especially important. Therefore, Logistic Regression was retained for threshold analysis and SHAP-based interpretation. For the auxiliary pay-grade prediction task, Logistic Regression also achieved the strongest overall performance, with an accuracy of 0.8093 and an ROC-AUC of 0.8993.

Figure 2. Distribution of top skill-demand clusters extracted from the jobs dataset

Table 3. Model performance comparison

Dataset	Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
HR Dataset — Attrition Prediction	Logistic Regression	0.7415	0.3333	0.6170	0.4328	0.7954
HR Dataset — Attrition Prediction	XGBoost	0.8503	0.5600	0.2979	0.3889	0.7578
HR Dataset — Attrition Prediction	Random Forest	0.8231	0.4138	0.2553	0.3158	0.7780
Jobs Dataset — Pay-Grade Prediction	Logistic Regression	0.8093	0.8320	0.8093	0.8009	0.8993
Jobs Dataset — Pay-Grade Prediction	XGBoost	0.7629	0.7656	0.7629	0.7588	0.8642
Jobs Dataset — Pay-Grade Prediction	Random Forest	0.7526	0.7525	0.7526	0.7523	0.8928

To further examine robustness, stratified 5-fold cross-validation was performed for the HR attrition models. The results are reported in Table 4 as mean ± standard deviation. The cross-validation results support the model-selection decision. Logistic Regression achieved the highest recall and ROC-AUC, while XGBoost and Random Forest achieved higher accuracy and precision. Since workforce-risk applications require effective identification of potential attrition cases, Logistic Regression was used for threshold tuning and explainability analysis. Figure 3 presents the confusion matrix for the selected Logistic Regression attrition model after threshold adjustment.

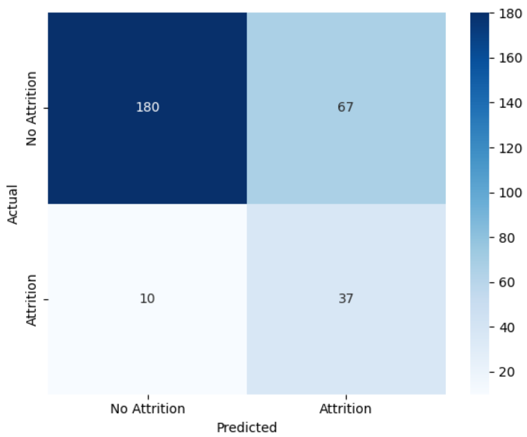

Figure 3. Confusion matrix for attrition prediction model

The confusion matrix shows that the model is good at distinguishing between non-attrition (true negatives = 180) and has a decent predictive power of attrition cases (true positives = 37). But the false positives (67) also indicate that some employees are being mistaken for at-risk, and the false negatives (10) indicate that some cases of attrition are missing. This is a trade-off in imbalanced classification problems between sensitivity and precision.

4.3 Threshold optimization results

Since the HR attrition dataset is class-imbalanced, the default classification threshold of 0.50 may not strike the best balance between precision and recall. Therefore, threshold analysis was performed for the selected Logistic Regression attrition model to examine the precision-recall trade-off. Threshold values from 0.10 to 0.90 were evaluated, and the F1-score was used to identify the most balanced threshold. From the research results, the threshold value of 0.40 was identified as the best F1 score for the assessed split. With a default threshold of 0.50, the model attained an accuracy of 0.7415, a precision of 0.3333, a recall of 0.6170, and an F1-score of 0.4328. When the threshold was increased to 0.40, the recall increased to 0.7872 and the F1-score to 0.4901, while accuracy slightly decreased to 0.7381. This means that the adjusted threshold increased the model's ability to identify potential attrition cases, an important aspect of workforce-risk applications (Table 5). Figure 4 illustrates the threshold tuning curve for the Logistic Regression attrition model.

Table 4. Stratified 5-fold cross-validation results for HR attrition models

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Logistic Regression	0.7537 ± 0.0142	0.3649 ± 0.0145	0.7090 ± 0.0460	0.4813 ± 0.0185	0.8228 ± 0.0231
Random Forest	0.8639 ± 0.0194	0.6617 ± 0.1294	0.3503 ± 0.0293	0.4557 ± 0.0516	0.8057 ± 0.0281
XGBoost	0.8639 ± 0.0173	0.6461 ± 0.1046	0.3880 ± 0.0292	0.4805 ± 0.0313	0.8077 ± 0.0214

Table 5. Threshold comparison for the logistic regression attrition model

Threshold	Accuracy	Precision	Recall	F1-score	Predicted Positive Count
0.50	0.7415	0.3333	0.6170	0.4328	87
0.40	0.7381	0.3558	0.7872	0.4901	104

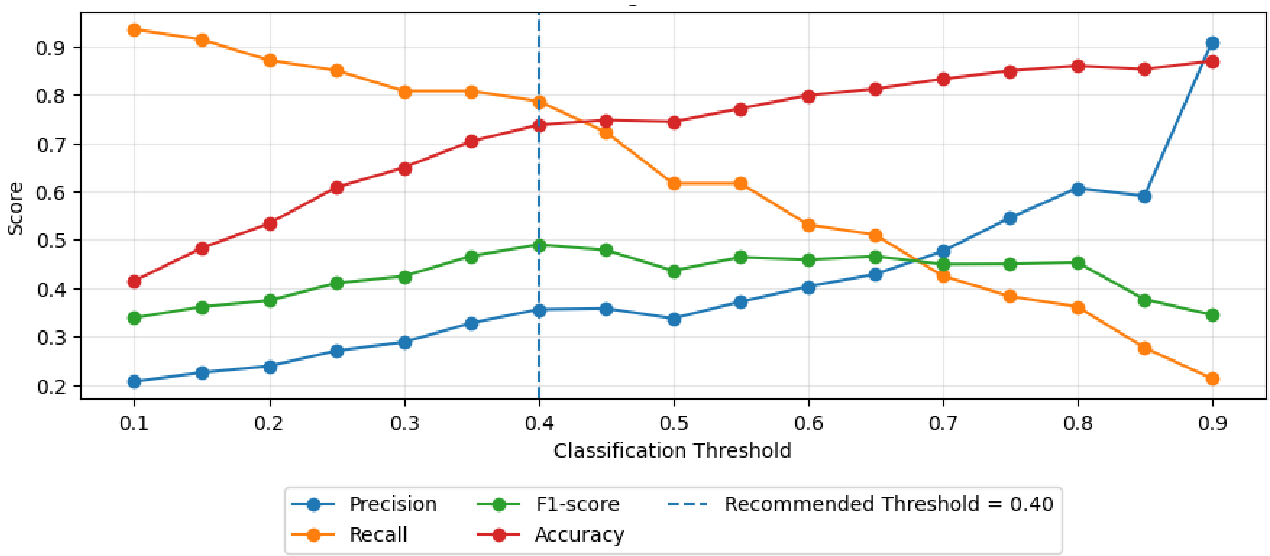

Figure 4. Threshold tuning curve

Lower thresholds increased recall but reduced precision, whereas higher thresholds improved precision but reduced recall. A threshold of 0.40 produced the best F1-score balance in the evaluated split and was therefore used to report the threshold-adjusted attrition prediction results.

4.4 Explainability results

SHAP analysis was applied to the selected Logistic Regression attrition model to improve interpretability and identify the features that contributed most strongly to attrition-risk predictions. Logistic Regression was selected for SHAP-based explanation because it provided the most suitable balance of recall, F1-score, and ROC-AUC in the attrition prediction task. SHAP LinearExplainer was used, and global feature importance was measured using mean absolute SHAP values. Figure 5 presents the SHAP summary plot for the attrition prediction model. The plot shows the magnitude and direction of feature contributions across the evaluated samples. Figure 5 shows the relative contribution of employee and job-related features to attrition-risk predictions across the evaluated dataset. The most influential features were overtime, frequent business travel, job level, laboratory technician role, and total years of work. Table 6 reports the top-ranked features according to mean absolute SHAP values.

Figure 5. SHAP summary plot

Table 6. Top SHAP features for attrition prediction

Rank	Feature	Mean Absolute SHAP Value	Direction of Higher Risk	Interpretation
1	OverTime_yes	0.6879	Positive	Employees working overtime showed a higher predicted attrition risk, suggesting that workload pressure may contribute to a tendency toward turnover.
2	BusinessTravel_travel_frequently	0.5857	Positive	Frequent business travel increased the predicted attrition risk, indicating possible work-life balance pressures and travel-related burdens.
3	JobLevel	0.5647	Lower levels increase risk	Lower job levels were associated with higher attrition-risk predictions, possibly reflecting lower stability, compensation, or career progression.
4	JobRole_laboratory technician	0.5504	Positive	The laboratory technician role contributed positively to the model's attrition prediction.
5	TotalWorkingYears	0.5301	Lower values increase risk	Employees with fewer total years of work showed a higher predicted attrition risk, indicating greater mobility among less-experienced employees.
6	EducationField_life sciences	0.4864	Mixed	This feature contributed to attrition prediction, although its direction varied depending on other employee characteristics.
7	BusinessTravel_travel_rarely	0.4526	Negative to neutral	Rare business travel generally showed a lower attrition-risk contribution than frequent travel.
8	EducationField_medical	0.4329	Mixed	This feature had a moderate contribution to model predictions depending on its combination with other variables.
9	Department_sales	0.4327	Positive	Membership in the sales department was associated with an elevated predicted attrition risk.
10	MaritalStatus_single	0.4192	Positive	Single employees showed higher predicted attrition risk, possibly reflecting greater career mobility.

Table 7. Workforce risk score ranking of skill-demand clusters

Rank	Skill-Demand Cluster	Normalized Skill Demand	Normalized Attrition Risk	Workforce Risk Score	Risk Category
1	analytics branding communication technology reporting	1.000	0.396	0.396	Moderate
2	retention strategy analytics communication technology reporting	0.800	0.396	0.316	Low
3	ux leadership communication technology analytics	0.800	0.396	0.316	Low
4	growth strategy leadership communication technology analytics	0.800	0.396	0.316	Low
5	analytics data interpretation communication technology reporting	0.600	0.396	0.237	Low
6	content analysis communication technology analytics reporting	0.600	0.396	0.237	Low
7	growth strategy analytics communication technology reporting	0.600	0.396	0.237	Low
8	partnerships leadership communication technology negotiation	0.600	0.396	0.237	Low
9	creative thinking, strategic planning, communication leadership, project management	0.400	0.396	0.158	Low
10	biology programming data analysis mathematical modelling research skills	0.400	0.396	0.158	Low

The SHAP results provide statistical and managerial insights into the attrition model. High overtime and business travel can indicate pressures on workload and work-life balance. A lower job level and fewer years in the job overall may indicate moving up the job ladder early or limited career progression. Satisfaction-related factors and departmental factors indicate that the probability of attrition depends on both individual work conditions and the organization's environment. These clarifications help to move beyond relying solely on categorization scores and instead use model predictions to guide HR action.

4.5 Workforce risk score results

The Workforce Risk Score was obtained by merging a normalized skill-demand intensity with a composite normalized attrition-risk indicator. The score was treated as an aggregate decision-support index because there were no direct connections between the occupations and HR datasets at the employee or job-record level, so it was not considered a causal risk estimate for employees. The Workforce Risk Score for the top skill-demand groups is shown in Table 7. The highest-ranked cluster was "analytics branding communication technology reporting", having a normalized skill demand of 1.000, a normalized attrition risk of 0.396, and a Workforce Risk Score of 0.396. This cluster was deemed to be a moderate risk. The remaining clusters were categorized as Low risk using the aggregate score framework. The results show that workforce risk is not uniformly distributed across skill clusters but rather concentrated in a small number of strategically important ones. Figure 6 shows the ranking of Workforce Risk Scores for the top skill-demand groups. The illustration shows that the highest-ranked cluster had a moderate aggregate risk score, while the other clusters were classified as low risk. This figure shows the total WRS values based on normalized skill-demand frequency and normalized aggregate attrition-risk indicators. The top-ranked cluster had a moderate risk score of 0.396, while the other clusters were low risk based on the aggregate scoring matrix. Only the top-ranked clusters are indicated by the red bars and do not indicate high-risk classification. Four aggregation methods were used to calculate the Workforce Risk Score ranking, and a sensitivity analysis was performed to assess the stability of the ranking across these methods: multiplicative scoring, equal-weight additive scoring, demand-heavy weighting, and attrition-heavy weighting. The results are given in Table 8. The sensitivity study demonstrated consistent ranking patterns across several aggregating approaches. This stability is partially affected by the aggregate attrition-risk design, which uses a uniform, normalized attrition-risk indicator across skill-demand clusters due to the lack of a direct correlation between employees and skills. Consequently, future research requires external validation against tangible organizational outcomes, such as actual turnover, vacancy duration, hiring costs, or signs of skill shortages.

4.6 Integrated workforce insights

The combined analysis provides an aggregate view of workforce intelligence by linking observed skill-demand clusters with attrition-risk modeling outputs. The results show that the highest-ranked skill-demand cluster reached a moderate Workforce Risk Score, while the remaining clusters were categorized as low risk under the aggregate scoring approach.

This suggests that the framework can support strategic prioritization by identifying skill clusters that may require closer managerial attention. Because the jobs and HR datasets are not directly linked at the employee level, these insights should not be interpreted as causal evidence that specific skills directly cause attrition risk. Instead, the framework provides a decision-support structure for combining external skill-demand signals with internal attrition-risk information. In practical terms, the framework can support workforce planning discussions related to retention, upskilling, and skill prioritization.

5. Discussion

This research demonstrates the benefits of integrating machine learning, natural language processing, and explainable AI within a workforce intelligence context. The results indicate that supervised learning models can detect important patterns in employee attrition data. These findings suggest that predictive analytics can be used to identify potential employee attrition early and to support evidence-based workforce planning. The SHAP interpretation layer addresses the lack of transparency in traditional predictive models by indicating which features are most influential in predicting the risk of attrition. This enhances the model's usefulness, as HR decision-makers can examine its predicted outcome and the factors that affect it. For high-impact organizations like human resource management, explainability plays a crucial role in fostering trust, accountability, and responsible AI usage [1]. The Workforce Risk Score takes the framework one step further to provide an overall measure of decision support based on the predictive and skill-demand outputs. The study considers not only the model's accuracy in prioritizing the workforce but also several other evaluation metrics, threshold optimization, explainability, and risk stratification. The framework allows for a structured approach to relating internal retention risk to external labor-market signals by combining an attrition prediction model with skill-demand analysis.

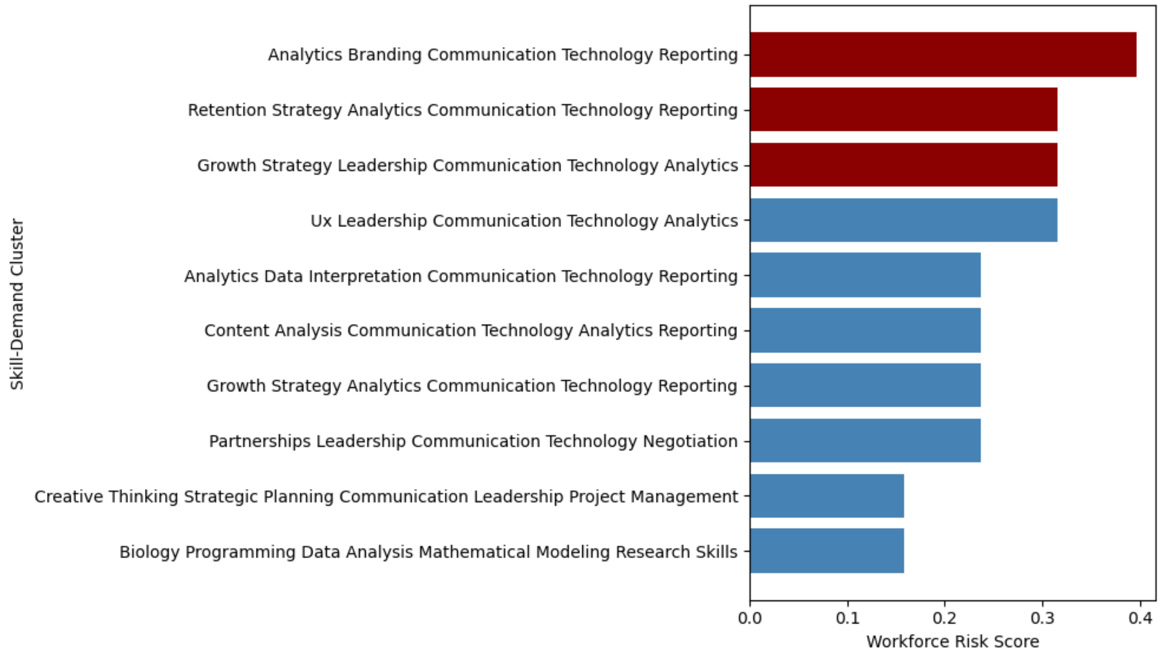

Figure 6. Workforce risk score ranking of skill clusters

Table 8. Workforce risk score sensitivity analysis

Skill-Demand Cluster	Multiplicative Rank	Equal-Weight Rank	Demand-Heavy Rank	Attrition-Heavy Rank
analytics branding communication technology reporting	1	1	1	1
retention strategy analytics communication technology reporting	2	2	2	2
ux leadership communication technology analytics	2	2	2	2
growth strategy leadership communication technology analytics	2	2	2	2
analytics data interpretation communication technology reporting	5	5	5	5
content analysis communication technology analytics reporting	5	5	5	5
growth strategy analytics communication technology reporting	5	5	5	5
partnerships leadership communication technology negotiation	5	5	5	5

It is all the more important in a rapidly changing labor market, where demand for new skills is still reshaping labor market priorities [8]. The evaluation of models also plays a crucial role in the case of class imbalance. Recall, F1-score, ROC-AUC, and threshold analysis are used to minimize reliance on accuracy alone and provide a more balanced evaluation of attrition prediction performance [17]. This enhances the validity of the results and ensures the chosen threshold is applied appropriately in practice. The results are not necessarily viewed as evidence of a fully validated workforce management system, but rather as evidence of a feasible approach to decision support. The proposed framework integrates the organization of external skill-demand clusters, internal attrition prediction, and explainable AI into a single analytical structure. However, because the two data sets are not record-level linked, the methodology does not make a causal claim about the relationship between job-market skill demand and employee attrition. The Workforce Risk Score should be considered as an all-encompassing strategic prioritizing index. It can help identify clusters of skill demand that may require managerial action, but should not be the sole basis for employee-level decisions. The framework's usefulness is in driving conversations about the workforce plan around retention, upskilling, and skill prioritizing. Further improvements in the framework's scalability could include incorporating longitudinal HR data, individual skills inventories, and data from numerous organizations. To improve the study of external skill demand, real-time job posting data from labor market websites and/or APIs can be used.

These improvements would enable the methodology to measure changes over time in skill demand within the organization and to validate the Workforce Risk Score results against actual measures, such as vacancy duration, hiring costs, turnover rates, and skill-shortage indicators. Despite these assets, there are certain drawbacks. Structured HR data can reduce cross-organizational generalisability due to differences in workforce composition, and the integration of external job market data can be improved by using real-time, more diverse data. Additionally, the models are static and may not capture changes in employees' behavior and the skills they need over time. Future studies ought thus to concentrate on scalable and evolutionary structures that involve real-time data, advanced NLP, deep learning, and extended explainability procedures.

This study validates AI-powered workforce analytics, offering a clear and repeatable decision-support model that can link skill-demand trends, attrition forecasts, and overall workforce risk prioritization.

6. Ethical and responsible AI considerations

When using AI for employee attrition predictions, careful ethical considerations must be made to prevent the AI's predictions from influencing decisions about employees. The proposed method is intended to support managerial decision-making and is not intended to be punitive or to fully automate employment decisions. Appropriate implementation is vital and requires employee privacy and data minimization, secure management of HR data, and transparency. There can also be biases in historical labor data that can be reflected in attrition models. If not used carefully, variables that reflect age, gender, marital status, employment status, or department can raise fairness concerns. Organizations that adopt this framework should therefore conduct fairness audits, monitor subgroup performance, and ensure that predictions are reviewed by human decision-makers. Explainability measures, such as SHAP, can provide transparency but are not necessarily fair. Future deployments will be based on responsible AI principles, including human oversight, accountability, bias monitoring, and adherence to applicable data protection and AI governance standards.

7. Conclusion

In this study, we presented a workforce intelligence framework based on explainable AI that comprises four stages: skill-demand cluster analysis, employee attrition prediction, SHAP explainability, and aggregate risk-aware decision support. To identify skill-demand clusters, job-market text was used, and structured HR data were used to predict employee attrition using supervised machine learning models. Logistic Regression yielded the best balance of recall, F1-score, and ROC-AUC, and SHAP found that, over time, frequent business travel, job level, laboratory technician position, and total working years were important attribute drivers of attrition. The Workforce Risk Score proposed is a combination of the normalized skill demand and the normalized aggregate attrition risk. One skill-demand cluster received a moderate overall risk rating, and the remaining clusters were rated low risk. For this reason, the score should be used as an index to support decision-making, not as a causal risk estimate at the employee level. The overall framework facilitates an interpretable workforce planning process by connecting labor-market skill signals, attrition prediction, and workforce risk prioritization. The framework should be validated in future studies using actual organizational data, employee skill inventories (ESIs), longitudinal HR data, and real-time labor market data.

Ethical issue

The authors are aware of and comply with best practices in publication ethics, specifically regarding authorship (avoidance of guest authorship), dual submission, manipulation of figures, competing interests, and compliance with research ethics policies. The authors adhere to publication requirements that the submitted work is original and has not been published elsewhere.

Data availability statement

The manuscript contains all the data. However, additional data will be provided by the corresponding author upon reasonable request.

Conflict of interest

The authors declare no potential conflict of interest.

References

[1] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why Should I Trust You?': Explaining the Predictions of Any Classifier," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, in KDD '16. New York, NY, USA: Association for Computing Machinery, Aug. 2016, pp. 1135–1144. doi: 10.1145/2939672.2939778.

[2] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," Advances in neural information processing systems, vol. 30, 2017, Accessed: Apr. 07, 2026. [Online]. Available: https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

[3] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," Nat Mach Intell, vol. 1, no. 5, pp. 206–215, May 2019, doi: 10.1038/s42256-019-0048-x.

[4] C. Molnar, Interpretable Machine Learning. Lulu.com, 2020. Publisher: Leanpub, ISBN: 9780244768522, 0244768528

[5] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," The Journal of machine Learning research, vol. 12, pp. 2825–2830, 2011.

[6] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, in KDD '16. New York, NY, USA: Association for Computing Machinery, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.

[7] G. Ke et al., "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," in Advances in Neural Information Processing Systems, Curran Associates, Inc., 2017. Accessed: Apr. 07, 2026. [Online]. Available: https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html

[8] D. H. Autor, "Why Are There Still So Many Jobs? The History and Future of Workplace Automation," Journal of Economic Perspectives, vol. 29, no. 3, pp. 3–30, Sep. 2015, doi: 10.1257/jep.29.3.3.

[9] D. S. Sisodia, S. Vishwakarma, and A. Pujahari, "Evaluation of machine learning models for employee churn prediction," in 2017 International Conference on Inventive Computing and Informatics (ICICI), Nov. 2017, pp. 1016–1020. doi: 10.1109/ICICI.2017.8365293.

[10] L. Alaskar, M. Crane, and M. Alduailij, "Employee Turnover Prediction Using Machine Learning," in Advances in Data Science, Cyber Security and IT Applications, vol. 1097, A. Alfaries, H. Mengash, A. Yasar, and E. Shakshuki, Eds., in Communications in Computer and Information Science, vol. 1097. , Cham: Springer International Publishing, 2019, pp. 301–316. doi: 10.1007/978-3-030-36365-9_25.

[11] A. Raza, K. Munir, M. Almutairi, F. Younas, and M. M. S. Fareed, "Predicting Employee Attrition Using Machine Learning Approaches," Applied Sciences, vol. 12, no. 13, p. 6424, Jan. 2022, doi: 10.3390/app12136424.

[12] A. V. Iyiade, F. Ayankoya, and A. Kuyoro, "Machine Learning Applications in Predicting Employee Turnover: A Systematic Review," Asian Journal of Engineering and Applied Technology, vol. 14, no. 2, pp. 44–50, Dec. 2025, doi: 10.70112/ajeat-2025.14.2.4329.

[13] K. Konar, S. Das, S. Das, and S. Misra, "Employee Attrition Prediction Using Bayesian Optimized Stacked Ensemble Learning and Explainable AI," SN COMPUT. SCI., vol. 6, no. 6, p. 672, Jul. 2025, doi: 10.1007/s42979-025-04204-w.

[14] G. Marín Díaz, J. J. Galán Hernández, and J. L. Galdón Salvador, "Analyzing Employee Attrition Using Explainable AI for Strategic HR Decision-Making," Mathematics, vol. 11, no. 22, p. 4677, Jan. 2023, doi: 10.3390/math11224677.

[15] M. AL-Ali, M. Alwateer, S. A. Alsaedi, H. M. Balaha, M. Badawy, and M. A. Elhosseini, "Integrating machine learning and explainable AI for employee attrition prediction in HR analytics," Sci Rep, vol. 16, no. 1, p. 6344, Feb. 2026, doi: 10.1038/s41598-026-36424-2.

[16] Md. N. Sakib and S. Islam, "The impacts of machine learning on human resource management: a systematic literature review and bibliometric analysis," Futur Bus J, vol. 12, no. 1, p. 7, Jan. 2026, doi: 10.1186/s43093-025-00704-6.

[17] T. Fawcett, "An introduction to ROC analysis," Pattern Recognition Letters, vol. 27, no. 8, pp. 861–874, Jun. 2006, doi: 10.1016/j.patrec.2005.10.010.

[18] M. Sokolova, N. Japkowicz, and S. Szpakowicz, "Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation," in AI 2006: Advances in Artificial Intelligence, A. Sattar and B. Kang, Eds., Berlin, Heidelberg: Springer, 2006, pp. 1015–1021. doi: 10.1007/11941439_114.

[19] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[20] Akhter, E. (2025). Jobs and skills mapping for career analysis [Data set]. Kaggle.

https://www.kaggle.com/datasets/emaadakhter/jobs-and-skills-mapping-for-career-analysis

[21] Pavan, S. (2017). IBM HR analytics employee attrition & performance [Data set]. Kaggle. https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset