International Journal of Engineering Business and Social Science
Vol. 4 No. 4, March 2026, pages: 55-71
e-ISSN: 2980-4108, p-ISSN: 2980-4272
https://ijebss.ph/index.php/ijebss

A Comparative Machine Learning Approach for Predicting Employee Attrition in Human Resource Analytics

Zeny Widianingsih[1], Aldi Algino[2]
STMIK IKMI Cirebon, Indonesia
Email: zenywidianingsih939@gmail.com[1], alginoaldi88@gmail.com[2]

Abstract
Employee attrition imposes substantial recruitment, onboarding, and knowledge-transfer costs on organizations, making early identification of at-risk employees a strategic priority for human resource management. Prior work on this benchmark commonly reports a single best-performing classifier, yet rarely compares how competing algorithms trade precision against recall within one reproducible pipeline. This study compares the predictive performance of three supervised machine learning algorithms, namely Logistic Regression, Random Forest, and Support Vector Machine (SVM), for predicting employee attrition using the IBM HR Analytics Employee Attrition dataset, comprising 1,470 records and 31 predictor features. Data were encoded, standardized, and split 80:20 by stratified sampling preserving the 16.12% attrition ratio. Accuracy, precision, recall, F1-score, and AUC were reported, with emphasis on recall because an undetected leaver is costlier than a false alarm. Results show that Support Vector Machine achieved the highest recall (0.702) and F1-score (0.524), Logistic Regression achieved the highest accuracy (0.878) and AUC (0.805), while Random Forest, despite class-weight balancing, achieved the lowest recall (0.149) among the three models. Feature-importance analysis revealed that monthly income, age, total working years, years at company, and overtime status were the strongest predictive correlates of attrition, which signal associative rather than causal relevance. The contribution is a metric-level diagnosis showing that no algorithm dominates and that model suitability depends on the evaluation objective, so selection should follow organizational priorities rather than accuracy alone. As the data are a synthetic benchmark analysed on one split, the recall advantage of SVM is a benchmark-level indication requiring external validation.

Keywords: employee attrition; machine learning; human resource analytics; predictive modeling

INTRODUCTION

Employee attrition, defined as the voluntary or involuntary departure of employees from an organization, remains one of the most persistent challenges in human resource management. Beyond the direct costs of recruitment, onboarding, and training a replacement, attrition also imposes indirect costs through lost institutional knowledge, reduced team productivity, and diminished morale among remaining staff (Hom et al., 2017). As labor markets become increasingly competitive, organizations that can anticipate which employees are at risk of leaving are better positioned to intervene proactively through targeted retention strategies rather than reacting only after resignations occur. At a global level the scale of this problem is considerable: a meta-analysis spanning a century of turnover research shows that voluntary departures arise from a dense network of attitudinal, organizational, and labor-market antecedents rather than from any

single cause, which makes turnover simultaneously pervasive across industries and difficult to anticipate with simple rules of thumb (Rubenstein et al., 2018). Lee et al. (2018) likewise observe that retention management in contemporary organizations has shifted from reactive replacement toward anticipatory, data-informed intervention.

Within this broad problem, the specific difficulty addressed in the present study is one of measurement rather than of modeling. Organizational attrition data are structurally imbalanced, because in any observation period the leavers constitute a small minority of the workforce, and the two errors a classifier can commit are not equally costly. Failing to flag an employee who subsequently resigns forfeits the entire opportunity for intervention, whereas flagging an employee who would have stayed costs only the time of a retention conversation. Frederiksen (2017) shows at firm level that satisfaction-related turnover is concentrated in identifiable segments of the workforce, which implies that the practical value of a prediction system depends on how reliably it detects that minority segment rather than on how often it is correct overall.

Machine learning has become an increasingly common tool for employee attrition prediction because it can uncover complex, non-linear relationships between employee attributes and turnover behavior that traditional statistical or survey-based HR methods may overlook (Avrahami et al., 2022). A substantial body of research has applied classification algorithms such as Logistic Regression, Random Forest, and Support Vector Machine to the widely used IBM HR Analytics Employee Attrition dataset. Raza et al. (2022) applied an optimized Extra Trees Classifier combined with hyperparameter tuning and data balancing, reporting that monthly income, hourly rate, job level, and age were the strongest drivers of attrition. Chakraborty et al. (2021) compared several machine learning classifiers for turnover prediction and found that ensemble methods generally outperformed single classifiers on structured HR data, while Alsheref et al. (2022) proposed an ensemble-based automated attrition prediction system and emphasized the importance of addressing class imbalance, since attrition datasets typically contain far fewer positive (leaver) cases than negative (stayer) cases.

The empirical base on this task has widened considerably in recent years. Zhao et al. (2019) benchmarked ten supervised classifiers on both simulated and real human resource records and reported that tree-based ensembles generally led on aggregate accuracy, while Rombaut and Guerry (2018) demonstrated that routinely stored personnel records alone can yield usable turnover predictions without attitudinal survey data. Fallucchi et al. (2020) and Jain et al. (2020) applied comparative pipelines to the same IBM dataset and reached differing conclusions about the best classifier, Najafi-Zangeneh et al. (2021) showed that systematic feature selection can improve attrition prediction while reducing model complexity, and Al-Darraji et al. (2021) reported that deep neural networks do not automatically outperform simpler classifiers on structured personnel data of this size. Yahia et al. (2021) extended the field toward deep data representations for people analytics, and Aswathi and Gururaja (2026) confirm that the comparative attrition prediction task remains an active subject of applied investigation.

More recently, Wang and Liu (2026) conducted a rigorous multi-metric benchmarking study on the same IBM HR Analytics dataset used in the present study, explicitly highlighting that baseline classifiers without imbalance handling tend to achieve high accuracy but poor minority-

class recall, and that class-weighted or threshold-tuned Logistic Regression can substantially improve the detection of true leavers. This finding underscores an important methodological point: because attrition datasets are inherently imbalanced (with attrition rates typically below 20%), accuracy alone can be a misleading indicator of a model's practical usefulness, since a naive model that always predicts “no attrition” can still achieve high accuracy while failing to identify any at-risk employee. Sekaran and Shanmugam (2022) further argued that interpretability, through explainable AI techniques, is essential for HR practitioners to trust and act upon model predictions, rather than treating the model as an opaque black box. Guerranti and Dimitri (2023) reached a comparable conclusion in a systematic comparison of classifiers on attrition data, noting that reported performance rankings shift according to whether the chosen metric rewards overall correctness or minority-class detection.

The urgency of resolving this evaluation problem is reinforced by the current state of human resource analytics as a field. Marler and Boudreau (2017) found in an evidence-based review that HR analytics practice remains thinly supported by rigorous evidence, and Tursunbayeva et al. (2018) concluded from a scoping review that value propositions in people analytics are frequently asserted rather than demonstrated. Minbaeva (2018) argues that the credibility of human capital analytics rests on the transparency of the analytical choices behind it. Where HR units must justify analytical investment, publishing a model accompanied only by a headline accuracy figure risks encouraging adoption of a system that performs poorly at the single task for which it was acquired, namely identifying the employees about to leave.

Despite this growing literature, many existing studies report only accuracy or a single best-performing model, without systematically comparing how different algorithms trade off precision against recall, an issue that is particularly consequential for attrition prediction given the asymmetric cost of missing a true leaver versus flagging a loyal employee unnecessarily. Stated explicitly, the gap is not an absence of predictive models but an absence of comparative diagnostic evidence: the literature rarely explains why algorithms trained on identical data, with identical class weights, diverge so sharply in minority-class detection, and it therefore offers practitioners no principled basis for choosing among them. There remains a need for studies that transparently report this trade-off across multiple algorithms under a single, reproducible pipeline, so that HR practitioners can select a model aligned with their specific organizational priorities, whether that is maximizing detection of at-risk employees or minimizing unnecessary retention spending.

The novelty of this study is stated explicitly rather than left implied. First, it reports the complete five-metric profile of three structurally different classifiers, a linear probabilistic model, a bagged tree ensemble, and a margin-based kernel model, under one preprocessing pipeline, one stratified split, and one fixed random seed, so that observed differences are attributable to model structure rather than to divergent data handling. Second, it moves beyond reporting the precision-recall trade-off to diagnosing its mechanism, relating each model's metric profile to its decision-boundary geometry, its handling of class weights, and its depth constraint. Third, it converts that diagnosis into an explicit decision rule linking organizational objective to model choice, presented as an analytical result rather than a closing managerial remark.

Therefore, this study aims (1) to compare the predictive performance of Logistic Regression, Random Forest, and SVM for employee attrition prediction using accuracy, precision, recall, F1-score, and AUC, and (2) to identify the employee attributes most strongly associated with attrition to inform data-driven retention strategy. A third objective is added, namely (3) to explain the structural reasons for the observed divergence in minority-class detection among the three algorithms, so that the comparison yields diagnostic rather than merely descriptive evidence.

The benefits of the study are both methodological and practical. Methodologically, it supplies a reproducible comparative baseline against which future imbalance-aware approaches on this benchmark can be judged on more than one metric. Practically, it gives HR units a transparent basis for selecting a model matched to their intervention capacity: an organization able to conduct many retention conversations should prefer a recall-oriented configuration, whereas one with scarce intervention capacity should prefer a precision-oriented model. The broader implication is that predictive HR systems ought to be audited against the decision they are intended to support, an argument consistent with the prescriptive analytics perspective advanced by Pessach et al. (2020), rather than against a single aggregate performance figure.

METHOD

This study employed a quantitative, experimental research design based on supervised machine learning classification. The design is comparative and non-interventional: three classification algorithms are trained and tested on an identical data partition, so that differences in measured performance are attributable to the algorithms themselves rather than to data handling. The unit of analysis is the individual employee record, that is, one row describing a single employee's demographic, job-related, and compensation attributes together with that employee's attrition status, while the unit of observation for performance estimation is the held-out test subset of those records. Because the study is a secondary analysis of a public benchmark dataset, it has no field research site in the conventional sense. The research site is the computational environment in which the experiment was executed, namely a Python 3 environment running the scikit-learn library on a single workstation, and the organizational setting represented in the data is the simulated multi-department firm described by the dataset publisher.

The dataset used is the publicly available IBM HR Analytics Employee Attrition & Performance dataset, a widely referenced benchmark dataset comprising 1,470 employee records and 35 attributes covering demographic information (e.g., age, gender, marital status), job-related information (e.g., job role, job level, department, overtime status), and compensation information (e.g., monthly income, daily rate, hourly rate, stock option level) (IBM Watson Analytics, 2015). The target variable, Attrition, is binary (Yes/No), with an overall attrition rate of 16.12% in the dataset, indicating a substantially imbalanced classification problem consistent with real-world organizational turnover rates. The dataset was selected for three stated reasons: it is openly accessible and therefore permits exact replication; it has been used repeatedly in the studies reviewed above, which makes the present results directly comparable with published findings; and its attrition rate is representative of the imbalance encountered in operational HR records.

Data collection took the form of documentary retrieval of an existing dataset rather than primary data gathering. No instrument was administered to employees, no respondent sampling was carried out, and no personally identifying information was handled, since the dataset is distributed in anonymized and fully pre-compiled form. The file was obtained in comma-separated format from the public repository, verified for completeness by checking record count, column count, and the absence of missing values, and then loaded into the analysis environment. Because the data are secondary, anonymized, and publicly licensed for analytical use, the study did not require ethical clearance for research involving human subjects; the ethical obligation that does apply, and which is observed throughout, is to report the synthetic character of the dataset honestly rather than to present its numerical findings as evidence about a real workforce.

Data preprocessing consisted of the following steps: (1) four constant or non-informative fields (EmployeeCount, EmployeeNumber, Over18, StandardHours) were removed since they carried no discriminative information; (2) all categorical variables were transformed into numeric form using label encoding; and (3) numerical features were standardized using z-score normalization (StandardScaler) to ensure comparable feature scales, which is particularly important for distance- and margin-based algorithms such as SVM. The preprocessed dataset was then split into training (80%, n = 1,176) and testing (20%, n = 294) subsets using stratified random sampling to preserve the original attrition ratio in both subsets. Removal of the four constant fields reduces the 35 original attributes to 31 predictor features, which is the figure reported in Table 1.

Three supervised classification algorithms were selected for comparison based on their prevalence and demonstrated effectiveness in prior attrition-prediction literature: (1) Logistic Regression, a linear probabilistic classifier commonly used as an interpretable baseline; (2) Random Forest, an ensemble of decision trees trained via bagging, configured with 300 estimators, a maximum depth of 8, and balanced class weighting to mitigate class imbalance; and (3) Support Vector Machine (SVM) with a radial basis function (RBF) kernel and balanced class weighting, which maps features into a higher-dimensional space to find an optimal separating hyperplane. All models were implemented in Python 3 using the scikit-learn library with a fixed random seed (42) to ensure reproducibility. The three algorithms were chosen deliberately to represent three distinct hypothesis spaces, namely a global linear decision boundary, a piecewise-axis-parallel partition, and a non-linear maximum-margin boundary, because a comparison restricted to variants of one family could not expose the structural source of differences in minority-class detection.

Model performance was evaluated on the held-out test set using five standard classification metrics: accuracy, precision, recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (AUC). Because attrition prediction is typically used to trigger costly retention interventions, recall was treated as a particularly important metric, since failing to identify a true leaver (a false negative) represents a missed opportunity for intervention, whereas a false positive only results in an unnecessary retention conversation. In addition, feature importance scores from the Random Forest model (based on mean decrease in Gini impurity) were extracted to identify the employee attributes most strongly associated with attrition. A limitation of this design is that default or lightly tuned hyperparameters were used for all models rather than exhaustive grid search, and only a single train-test split (rather than k-fold cross-validation) was

used to estimate performance, which may cause some variance in the reported metrics. The choice of metrics follows the established argument that, under class imbalance, accuracy alone is an unreliable summary of classifier quality and must be read alongside metrics that are sensitive to the minority class (Chicco & Jurman, 2020). Class weighting rather than resampling was adopted as the imbalance treatment at this stage, which allows the comparison to isolate the effect of algorithm structure; resampling approaches such as SMOTE and its variants are deliberately reserved for future work because their behavior interacts with model structure in ways that would confound a three-way algorithm comparison (Fernández et al., 2018). In addition to the five headline metrics, the confusion matrix of each model on the test set was reconstructed and is reported in Table 3, because the raw counts of true and false positives and negatives are what permit a diagnostic rather than merely descriptive interpretation of the results.

RESULT AND DISCUSSION

Table 1 summarizes the descriptive characteristics of the dataset after preprocessing, and Table 2 presents the performance of the three machine learning models on the test set. Table 3 reconstructs the confusion matrix underlying each metric profile, and Table 4 sets out the decision rule derived from the comparison. Figures 1 to 3 present the metric comparison, the ROC curves, and the feature-importance ranking respectively.

Table 1. Dataset characteristics

Characteristic	Value
Total records	1470
Training set size	1176
Testing set size	294
Overall attrition rate	16.12%
Number of predictor features	31

Table 2. Performance comparison of machine learning models

Model	Accuracy	Precision	Recall	F1-Score	AUC
Logistic Regression	0.8776	0.7200	0.3830	0.5000	0.8052
Random Forest	0.8197	0.3500	0.1489	0.2090	0.7786
Support Vector Machine	0.7959	0.4177	0.7021	0.5238	0.8070

As shown in Table 2, Logistic Regression achieved the highest accuracy (0.8776) and AUC (0.8052), reflecting strong overall discriminative ability. However, its recall (0.383) indicates that it correctly identified fewer than 40% of employees who actually left the company. Support Vector Machine, by contrast, achieved a substantially higher recall (0.702) and the highest F1-score (0.524) among the three models, at the cost of lower accuracy (0.796) and more false positives. Random Forest, despite balanced class weighting, performed the weakest on recall (0.149) and F1-score (0.209), suggesting that the depth constraint (max_depth = 8) applied to control overfitting

may have limited its ability to capture the minority-class attrition pattern in this dataset. Read together, these figures establish the central analytical finding of the study: there is no universal best model on this dataset. Logistic Regression leads on two metrics, SVM leads on two others, and Random Forest leads on none, so the ranking of the three algorithms is not a property of the algorithms alone but a joint property of the algorithm and the metric chosen to judge it. Model suitability is therefore determined by the evaluation objective, and any claim that one of these algorithms is simply the best for attrition prediction is an artifact of selective metric reporting.

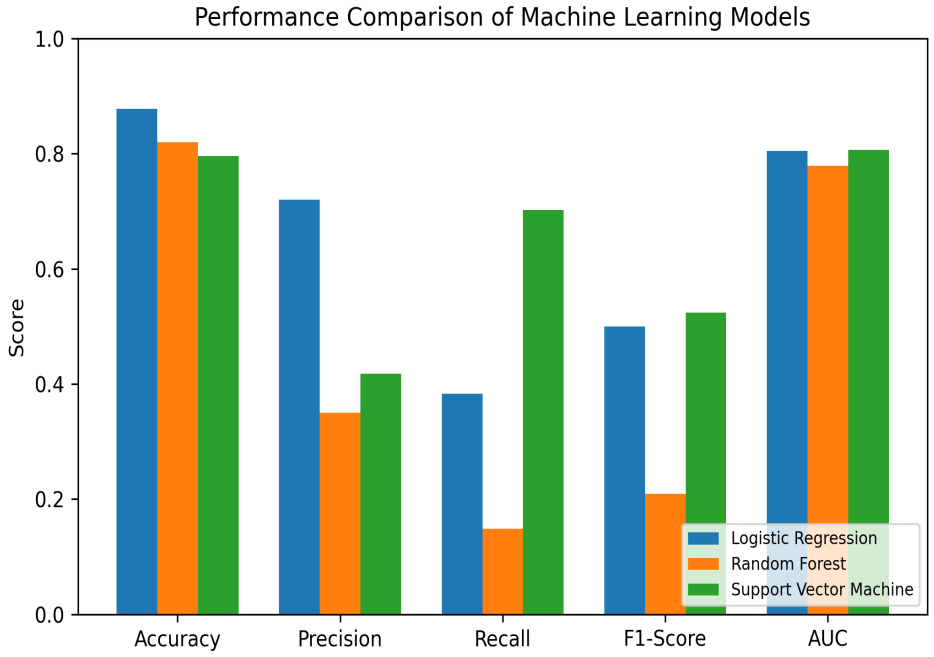


Figure 1. Performance comparison of Logistic Regression, Random Forest, and SVM across five evaluation metrics.

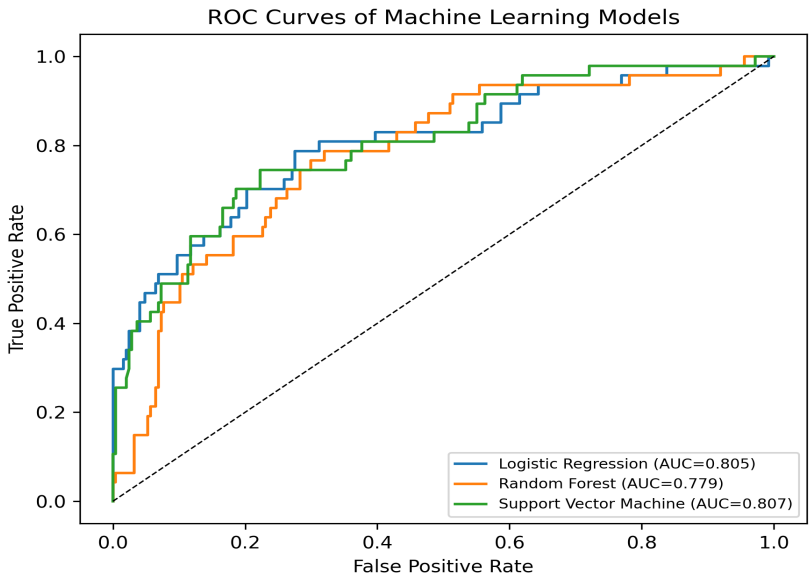


Figure 2. ROC curves of the three machine learning models on the test set.

To move from description to explanation, the confusion matrix of each model was reconstructed from the reported metrics and the known composition of the test set. With 294 test records and an attrition rate of 16.12%, the test set contains 47 actual leavers and 247 actual stayers. Table 3 presents the resulting counts, which are fully consistent with the accuracy, precision, and recall values in Table 2 and expose what each metric profile means in operational terms.

Table 3. Reconstructed confusion matrices of the three models on the test set (n = 294; 47 actual leavers, 247 actual stayers)

Model	True positive (leavers detected)	False negative (leavers missed)	False positive (stayers flagged)	True negative (stayers correctly cleared)
Logistic Regression	18	29	7	240
Random Forest	7	40	13	234
Support Vector Machine	33	14	46	201

Table 3 makes the practical difference unmistakable. Logistic Regression issues only 25 positive predictions in total, of which 18 are correct, so it is cautious and accurate when it does raise an alarm but silent about 29 of the 47 employees who leave. Random Forest issues 20 positive predictions of which only 7 are correct, missing 40 leavers. SVM issues 79 positive predictions, of which 33 are correct, detecting nearly twice as many leavers as Logistic Regression and almost five times as many as Random Forest, while contacting 46 employees who would have stayed. The question the manuscript must answer is why three models supplied with the same features and, in two cases, the same balanced class weighting, differ this sharply. The first part of the answer lies

in how each algorithm consumes the class weights. In Logistic Regression and SVM the balanced weight enters the loss function as a multiplier on the misclassification cost of the minority class, so it directly shifts the fitted decision surface toward the minority region. In Random Forest the weight is applied within the impurity computation at each candidate split, where its influence is diluted: a weighted Gini criterion only alters split selection if a minority-sensitive split is available among the candidate features at that node, and the final prediction is then produced by majority voting across 300 trees. Averaging over many shallow trees reverts the ensemble toward the majority class even when individual trees are weight-sensitive, which is why class weighting raised recall for the margin-based model but not for the ensemble.

The second part of the answer concerns decision-boundary geometry and feature representation. Attrition in this dataset is not separable along a small number of axis-parallel thresholds; it is concentrated in interaction regions such as employees who combine comparatively low monthly income with overtime work and short tenure. An RBF kernel maps the standardized features into a high-dimensional space in which such interaction regions can be enclosed by a smooth non-linear boundary, and because standardization was applied in preprocessing, the kernel distances are not dominated by any single large-scale variable such as monthly income. A bagged tree ensemble must instead approximate the same region with a staircase of axis-parallel cuts, and each shallow tree can only accommodate a limited number of them. Logistic Regression occupies an intermediate position: its single global linear boundary captures the dominant monotone trends in income and tenure, which explains its strong AUC of 0.8052, but it cannot bend around the interaction pockets, which explains why its ranking quality is good while its recall at the default threshold is poor.

The third part of the answer is the hyperparameter configuration, and specifically the depth constraint. Limiting Random Forest to a maximum depth of 8 was intended to control overfitting, but on an imbalanced dataset of 1,176 training records it also truncates precisely the deep, narrow branches in which minority-class cases accumulate. A node reached by eight successive splits still contains a mixture of leavers and stayers, and because stayers dominate in nearly every such node, majority voting resolves the mixture in favour of the majority class. The weak recall of Random Forest should therefore be read as a consequence of the chosen configuration rather than as evidence that bagged ensembles are inherently unsuited to attrition prediction, a reading supported by studies that report competitive ensemble performance on this benchmark under less restrictive depth settings (El-Rayes et al., 2020; Guerranti & Dimitri, 2023). It also follows that the comparison reported here is a comparison of three configured models, not of three algorithm families in the abstract.

The fourth part of the answer is the imbalance itself, interacting with the fixed decision threshold of 0.5. Because stayers outnumber leavers by roughly five to one, any model whose score distribution for leavers overlaps substantially with that of stayers will place most leavers below the default threshold, so a model can rank employees well and still classify them poorly. The gap between the AUC of Logistic Regression (0.8052), which is threshold-independent, and its recall (0.383), which is threshold-dependent, is exactly this phenomenon: the ranking contains usable information that the default cut-off discards. This observation also explains why the accuracy

ordering of the three models is close to the inverse of their recall ordering, since the dominant route to high accuracy on imbalanced data is to predict the majority class more often (Chicco & Jurman, 2020). Treating imbalance through class weights alone, as done here, moves the boundary but does not change the threshold, which is why none of the three models reaches both high recall and high precision simultaneously.

Taken together, these four mechanisms, namely the route by which class weights enter the objective, the geometry of the decision boundary relative to the structure of the data, the depth constraint applied to the ensemble, and the interaction between imbalance and a fixed threshold, account for the full pattern observed in Tables 2 and 3. They also convert the finding that no model dominates from a limitation into an operational instrument, because once the mechanism behind each metric profile is understood, the choice between models can be stated as an explicit rule rather than left to preference. Table 4 presents that rule, mapping each plausible organizational objective onto the governing metric, the model that optimizes it, and the consequence the organization should expect on data of this kind.

Table 4. Decision rule linking organizational objective to preferred model and its operational consequence

Organizational objective	Governing metric	Preferred model	Operational consequence on the test set
Detect as many at-risk employees as possible (ample intervention capacity)	Recall	SVM (RBF, balanced weights)	33 of 47 leavers detected; 46 stayers contacted unnecessarily
Avoid spending retention resources on employees who would have stayed	Precision	Logistic Regression	72% of flagged employees are genuine leavers; 29 leavers missed
Balance detection against wasted effort in a single screening run	F1-score	SVM (RBF, balanced weights)	Highest harmonic balance (0.524) of the three models
Rank employees by attrition risk rather than issue binary flags	AUC	Logistic Regression	Best ranking quality (0.805); threshold set by capacity
Report overall classification correctness only	Accuracy	Logistic Regression	0.878, but conceals detection of only 38% of leavers

This precision-recall trade-off is consistent with the broader attrition-prediction literature. Wang and Liu (2026), using the same IBM HR Analytics dataset, similarly reported that baseline Logistic Regression achieved a strong overall balance across metrics (accuracy = 86.05%, AUC = 0.812) but only moderate minority-class recall (34.04%), and that imbalance-aware modeling was needed to substantially raise recall, in their case to 63.83% through class-weighted Logistic Regression. The present study's SVM results, with balanced class weighting yielding a recall of 70.2%, extend this pattern by showing that a margin-based classifier with class weighting can achieve an even larger recall improvement than a linear model, though this comes with a

corresponding reduction in precision and overall accuracy. Chakraborty et al. (2021) and Alsheref et al. (2022) similarly emphasized that ensemble and balancing techniques are often necessary to meaningfully improve minority-class detection on this type of imbalanced HR dataset, rather than relying on a single algorithm's default configuration. The convergence between the two studies is informative in a further respect: because Wang and Liu obtained a recall of 63.83% from a linear model with class weights while the present study obtained 70.2% from a kernel model with the same imbalance treatment, the residual difference is plausibly attributable to boundary flexibility rather than to the imbalance treatment, which strengthens the structural interpretation offered above.

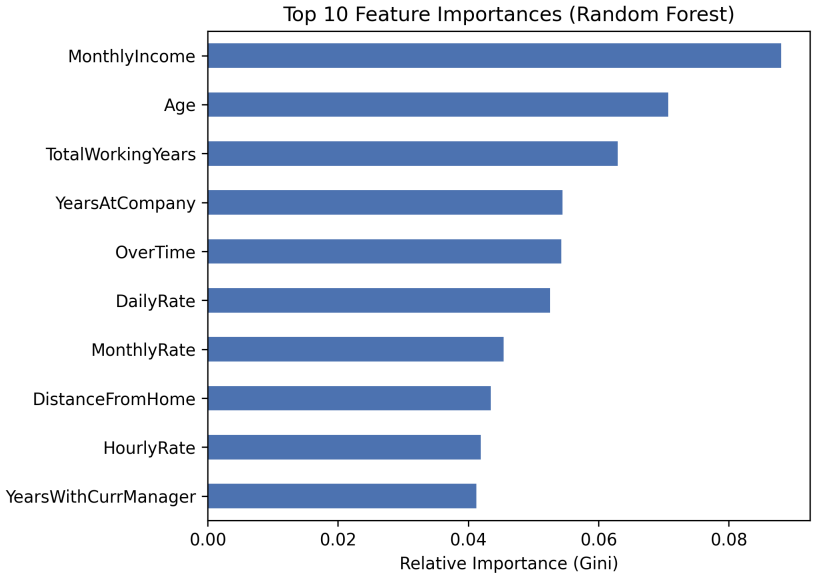


Figure 3. Top ten feature importances from the Random Forest model.

Figure 3 presents the ten most important features identified by the Random Forest model based on mean decrease in Gini impurity. Monthly income, age, total working years, years at company, and overtime status emerged as the five most influential predictors of attrition. This finding closely mirrors the feature-importance ranking reported by Wang and Liu (2026) on the same dataset, who identified MonthlyIncome, Age, TotalWorkingYears, DailyRate, and YearsAtCompany, among other variables, as the strongest benchmark predictors, and is also broadly consistent with Raza et al. (2022), who identified monthly income, hourly rate, job level, and age as key attrition drivers. The convergence of feature-importance rankings across independently conducted studies using the same benchmark dataset lends additional credibility to the substantive interpretation that compensation-related and tenure-related variables, together with overtime burden, are central to understanding voluntary employee turnover in this dataset. These rankings must nevertheless be interpreted as statements about predictive association and not about causal influence. Mean decrease in Gini impurity measures how much a feature contributes to reducing node heterogeneity within this particular fitted model; it is sensitive to feature scale and cardinality, it distributes importance arbitrarily among correlated predictors such as age, total

working years, and years at company, and it carries no information about the direction of an effect or about what would happen if the organization intervened on the variable. The correct reading is therefore that monthly income, tenure, and overtime status are the attributes on which this dataset most reliably separates leavers from stayers, not that raising income or reducing overtime would cause attrition to fall by a determinate amount. Establishing the latter would require either longitudinal observation of an intervention or an explicitly causal design, and methods such as SHAP would improve the attribution of predictive credit across correlated features without converting association into causation (Lundberg et al., 2020).

From a managerial standpoint, these results carry a practical implication that goes beyond the specific dataset: the choice of predictive model should be guided by organizational priorities rather than by accuracy alone. Avrahami et al. (2022) similarly argue that HR analytics models should be evaluated not only on statistical performance but also on how well their outputs align with practical retention-management processes. If an organization's primary concern is to intervene proactively with as many at-risk employees as possible, a recall-oriented model such as the class-weighted SVM used here would be preferable, even though it generates more false alarms. Conversely, if an organization has limited capacity for retention interventions and wishes to prioritize employees who are truly likely to leave, the higher-precision Logistic Regression model may be more appropriate. Sekaran and Shanmugam (2022) further emphasize that presenting HR decision-makers with interpretable feature-importance information, such as the ranking in Figure 3, is essential for translating model output into concrete retention actions, such as reviewing compensation bands for employees with long tenure but stagnant income, or monitoring overtime levels as an early warning indicator. Table 4 operationalizes this argument by tying the metric to the decision, and the counts in Table 3 allow the trade-off to be costed before deployment: adopting SVM in place of Logistic Regression on data of this composition converts 15 missed leavers into detected ones at the price of 39 additional unnecessary retention conversations, which is a calculation an HR unit can evaluate against its own capacity. Two cautions apply to such deployment. First, because retention decisions affect individual employees, preference should be given to models whose reasoning can be inspected and contested rather than to opaque high-performing ones (Rudin, 2019), a consideration that favours the interpretable Logistic Regression baseline where its recall is adequate. Second, prediction is only the first stage of a decision pipeline, and the value of the model is realized through the allocation of limited retention resources, which is more appropriately framed as a prescriptive analytics problem (Pessach et al., 2020).

This study has several limitations that should be considered when interpreting the results. First, the IBM HR Analytics dataset is a synthetic benchmark dataset created for analytical demonstration purposes rather than verified real-world organizational data, so the specific numerical findings should not be generalized directly to any particular organization without validation using independently collected HR data (Wang & Liu, 2026). Second, the models were evaluated using default or lightly tuned hyperparameters and a single stratified train-test split rather than exhaustive hyperparameter optimization and k-fold cross-validation, which may cause some variance in the reported metrics. Future research could apply systematic hyperparameter tuning, cross-validation, and additional class-imbalance techniques such as SMOTE, as well as explore

gradient-boosting algorithms and explainable AI methods such as SHAP to further improve both predictive performance and interpretability for HR decision support. Third, with only 47 positive cases in the test set, the recall estimates carry appreciable sampling uncertainty, since a shift of a single correctly classified leaver changes recall by about 2.1 percentage points; repeated stratified cross-validation with confidence intervals would therefore be required before the ordering of the models could be treated as stable. Fourth, the comparison holds the decision threshold fixed at 0.5, so it measures the models as configured rather than at their respective optimal operating points, and threshold tuning on a validation split is a necessary extension (Fernández et al., 2018). Recent work combining ensemble learning with explainable AI on this task indicates that both of these extensions can be pursued without sacrificing interpretability (Konar et al., 2025).

CONCLUSION

This study compared the performance of Logistic Regression, Random Forest, and Support Vector Machine algorithms for predicting employee attrition using the IBM HR Analytics Employee Attrition dataset. The results indicate that no single algorithm dominates across all metrics: Logistic Regression achieved the highest accuracy (0.878) and AUC (0.805), while class-weighted SVM achieved the highest recall (0.702) and F1-score (0.524), and Random Forest underperformed on recall despite balanced class weighting. The principal contribution of the study is therefore not the identification of a winning algorithm but a diagnostic account of why the ranking changes with the metric, attributing the divergence to the route by which class weights enter each objective function, the geometry of each decision boundary relative to the interaction structure of the data, the depth constraint imposed on the ensemble, and the interaction between class imbalance and a fixed decision threshold. This account is consolidated into an explicit decision rule, reported in Table 4, that links organizational objective to governing metric and preferred model.

Feature-importance analysis revealed that monthly income, age, total working years, years at company, and overtime status are the strongest predictors of attrition, a pattern consistent with independent benchmarking studies on the same dataset. These attributes should be understood as predictive correlates rather than as causal levers, since mean decrease in Gini impurity quantifies a feature's contribution to separating classes within a fitted model and distributes credit arbitrarily among correlated predictors, so it supports targeting of attention but not estimation of intervention effects. The main methodological limitation is correspondingly clear: performance was estimated from a single stratified split containing only 47 positive cases, with lightly tuned hyperparameters and a fixed threshold, so the reported metrics are point estimates of uncertain stability rather than settled quantities.

These findings confirm that machine learning can serve as a useful decision-support tool for proactive human resource management, provided that model selection is explicitly aligned with organizational retention priorities rather than based on accuracy alone, and that predictions are interpreted as benchmark-level patterns requiring validation against an organization's own HR data before operational deployment. Because the IBM dataset is a synthetic benchmark rather than

verified organizational data, the results are best read as a reproducible reference point for methodological comparison, and their generalizability to any specific workforce is limited accordingly. External validation on independently collected personnel records, accompanied by repeated cross-validation, threshold tuning, and interpretability analysis, is the necessary next step before any of the three configurations reported here is used to allocate retention resources in practice.

REFERENCE

Al-Darraji, S., Honi, D. G., Fallucchi, F., Abdulsada, A. I., Giuliano, R., & Abdulmalik, H. A. (2021). Employee attrition prediction using deep neural networks. Computers, 10(11), 141. https://doi.org/10.3390/computers10110141

Alsheref, F. K., Fattoh, I. E., & Ead, W. M. (2022). Automated prediction of employee attrition using ensemble model based on machine learning algorithms. Computational Intelligence and Neuroscience, 2022, 7728668. https://doi.org/10.1155/2022/7728668

Aswathi, C., & Gururaja, S. (2026). Employee attrition prediction. International Journal of Engineering Research & Technology, 14(1). https://www.ijert.org/employee-attrition-prediction-ijertconv14is010080

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

Chakraborty, R., Mridha, K., Shaw, R. N., & Ghosh, A. (2021). Study and prediction analysis of the employee turnover using machine learning approaches. In 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON) (pp. 1-6). IEEE. https://doi.org/10.1109/GUCON50781.2021.9573759

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics, 21(1), 6. https://doi.org/10.1186/s12864-019-6413-7

El-Rayes, N., Fang, M., Smith, M., & Taylor, S. M. (2020). Predicting employee attrition using tree-based models. International Journal of Organizational Analysis, 28(6), 1273-1291. https://doi.org/10.1108/IJOA-10-2019-1903

Fallucchi, F., Coladangelo, M., Giuliano, R., & De Luca, E. W. (2020). Predicting employee attrition using machine learning techniques. Computers, 9(4), 86. https://doi.org/10.3390/computers9040086

Fernández, A., García, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary. Journal of Artificial Intelligence Research, 61, 863-905. https://doi.org/10.1613/jair.1.11192

Frederiksen, A. (2017). Job satisfaction and employee turnover: A firm-level perspective. German Journal of Human Resource Management, 31(2), 132-161. https://doi.org/10.1177/2397002216683885

Guerranti, F., & Dimitri, G. M. (2023). A comparison of machine learning approaches for predicting employee attrition. Applied Sciences, 13(1), 267. https://doi.org/10.3390/app13010267

Hom, P. W., Lee, T. W., Shaw, J. D., & Hausknecht, J. P. (2017). One hundred years of employee turnover theory and research. Journal of Applied Psychology, 102(3), 530-545. https://doi.org/10.1037/apl0000103

IBM Watson Analytics. (2015). IBM HR Analytics Employee Attrition & Performance dataset [Dataset]. Retrieved from https://www.ibm.com/communities/analytics/watson-analytics-blog/hr-employee-attrition/

Jain, P. K., Jain, M., & Pamula, R. (2020). Explaining and predicting employees' attrition: A machine learning approach. SN Applied Sciences, 2(4), 757. https://doi.org/10.1007/s42452-020-2519-4

Konar, K., Das, S., Das, S., & Misra, S. (2025). Employee attrition prediction using Bayesian optimized stacked ensemble learning and explainable AI. SN Computer Science, 6(6), 672. https://doi.org/10.1007/s42979-025-04204-w

Lee, T. W., Hom, P., Eberly, M., & Li, J. (2018). Managing employee retention and turnover with 21st century ideas. Organizational Dynamics, 47(2), 88-98. https://doi.org/10.1016/j.orgdyn.2017.08.004

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence, 2(1), 56-67. https://doi.org/10.1038/s42256-019-0138-9

Marler, J. H., & Boudreau, J. W. (2017). An evidence-based review of HR analytics. The International Journal of Human Resource Management, 28(1), 3-26. https://doi.org/10.1080/09585192.2016.1244699

Minbaeva, D. B. (2018). Building credible human capital analytics for organizational competitive advantage. Human Resource Management, 57(3), 701-713. https://doi.org/10.1002/hrm.21848

Najafi-Zangeneh, S., Shams-Gharneh, N., Arjomandi-Nezhad, A., & Hashemkhani Zolfani, S. (2021). An improved machine learning-based employees attrition prediction framework with emphasis on feature selection. Mathematics, 9(11), 1226. https://doi.org/10.3390/math9111226

Pessach, D., Singer, G., Avrahami, D., Chalutz Ben-Gal, H., Shmueli, E., & Ben-Gal, I. (2020). Employees recruitment: A prescriptive analytics approach via machine learning and mathematical programming. Decision Support Systems, 134, 113290. https://doi.org/10.1016/j.dss.2020.113290

Raza, A., Munir, K., Almutairi, M., Younas, F., & Fareed, M. M. S. (2022). Predicting employee attrition using machine learning approaches. Applied Sciences, 12(13), 6424. https://doi.org/10.3390/app12136424

Rombaut, E., & Guerry, M.-A. (2018). Predicting voluntary turnover through human resources database analysis. Management Research Review, 41(1), 96-112. https://doi.org/10.1108/MRR-04-2017-0098

Rubenstein, A. L., Eberly, M. B., Lee, T. W., & Mitchell, T. R. (2018). Surveying the forest: A meta-analysis, moderator investigation, and future-oriented discussion of the antecedents of voluntary employee turnover. Personnel Psychology, 71(1), 23-65. https://doi.org/10.1111/peps.12226

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206-215. https://doi.org/10.1038/s42256-019-0048-x

Sekaran, K., & Shanmugam, S. (2022). Interpreting the factors of employee attrition using explainable AI. In 2022 International Conference on Decision Aid Sciences and Applications (DASA) (pp. 932-936). IEEE. https://doi.org/10.1109/DASA54658.2022.9765067

Tursunbayeva, A., Di Lauro, S., & Pagliari, C. (2018). People analytics: A scoping review of conceptual boundaries and value propositions. International Journal of Information Management, 43, 224-247. https://doi.org/10.1016/j.ijinfomgt.2018.08.002

Wang, G., & Liu, Y. (2026). Job satisfaction, compensation, and attrition risk: Machine learning evidence from the IBM HR Analytics benchmark dataset. Frontiers in Psychology, 17, 1902203. https://doi.org/10.3389/fpsyg.2026.1902203

Yahia, N. B., Hlel, J., & Colomo-Palacios, R. (2021). From big data to deep data to support people analytics for employee attrition prediction. IEEE Access, 9, 60447-60458. https://doi.org/10.1109/ACCESS.2021.3074559

Zhao, Y., Hryniewicki, M. K., Cheng, F., Fu, B., & Zhu, X. (2019). Employee turnover prediction with machine learning: A reliable approach. In K. Arai, S. Kapoor, & R. Bhatia (Eds.), Intelligent systems and applications (Advances in Intelligent Systems and Computing, Vol. 869, pp. 737-758). Springer. https://doi.org/10.1007/978-3-030-01057-7_56