

Improving minority-class detection in employee attrition with ensemble learning and SHAP-Based explanations

Topluluk öğrenme yöntemleri ve SHAP tabanlı açıklamalar kullanılarak çalışan ayrılmalarında azınlık sınıfının tespit performansının artırılması

Cevher Özden¹ 

¹Asst. Prof. Dr., University of Çukurova, Faculty of Arts and Sciences, Department of Computer Sciences, Adana/Türkiye



Abstract

This study examines how traditional ensemble techniques and recent deep learning approaches perform in predicting employee attrition, particularly when working with small and imbalanced tabular datasets. In addition, it proposes an interpretable decision-support framework designed not only for technical evaluation but also for practical HR use. The analysis was conducted on the IBM HR Analytics dataset using a specially structured validation design to prevent data leakage. To address class imbalance, SMOTE was applied strictly within the training folds, without touching the validation data. The models compared include Random Forest, XGBoost, CatBoost, Artificial Neural Networks (ANN), and TabNet. Rather than relying on default probability thresholds, dynamic threshold adjustment was introduced to improve sensitivity to the minority class. Model performance was evaluated using Recall, Precision, F1-score, and confusion matrices, while SHAP analysis was employed to enhance interpretability and support transparent decision-making in HR contexts. To strengthen the reliability of the evaluation, a Stratified 5-Fold Cross-Validation scheme was adopted. The findings show that CatBoost produced the most balanced and consistent results, achieving a mean F1-score of 0.468 ± 0.053 together with a Brier score of 0.097. After dynamic threshold adjustment, TabNet demonstrated the highest sensitivity (Recall: 0.573 ± 0.064), making it particularly effective for early risk detection. According to the SHAP-based interpretation, OverTime and Stock Option Level emerged as the most influential predictors.

Keywords: Employee churn prediction, tabnet, xgboost, explainable ai (xai), shap, imbalanced classification.

Citation:

Özden, C. (2026). Improving minority-class detection in employee attrition with ensemble learning and SHAP-Based explanations. *OPUS- Journal of Society Research*, 23, e1837461. <https://doi.org/10.26466/opusjrsr.1837461>

Open Access Statement:

This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). This license permits unrestricted use, distribution, and reproduction in any medium, provided that the original author(s) and the source are properly credited. The original publication in *OPUS Journal of Society Research* must be cited in accordance with accepted academic practice.

Review Note:

Evaluated by Double-Blind Peer Review

Ethics Reporting:

To report potential ethical concerns, contact: editorialoffice@opusjournal.net

Similarity screening was conducted via intihal.net.



Öz

Bu çalışma, özellikle küçük ve dengesiz tablosal veri setlerinde, geleneksel topluluk (ensemble) yöntemleri ile güncel derin öğrenme yaklaşımlarının çalışan yıpranmasını (işten ayrılma) tahmin etmedeki performansını incelemekte; ayrıca yalnızca teknik değerlendirmeye değil, pratik İK kullanımına da uygun, yorumlanabilir bir karar destek yapısı önermektedir. Analiz, veri sızıntısını önlemek amacıyla özel olarak yapılandırılmış bir doğrulama tasarımıyla IBM İK Analitiği veri seti üzerinde yürütülmüştür. Sınıf dengesizliğini gidermek için SMOTE, doğrulama verilerine müdahale edilmeden yalnızca eğitim katmanlarında uygulanmıştır. Karşılaştırılan modeller Rastgele Orman, XGBoost, CatBoost, Yapay Sinir Ağları (ANN) ve TabNet'tir. Varsayılan olasılık eşikleri yerine, azınlık sınıfına duyarlılığı artırmak amacıyla dinamik eşik ayarı kullanılmıştır. Model performansı Recall, Precision, F1-skoru ve hata matrisleriyle değerlendirilmiş; yorumlanabilirliği güçlendirmek için SHAP analizinden yararlanılmıştır. Güvenilirliği artırmak üzere Tabakalı 5-Katlı Çapraz Doğrulama uygulanmıştır. Bulgular, CatBoost'un 0,097 Brier skoru ve $0,468 \pm 0,053$ ortalama F1-skoruyla en dengeli sonuçları verdiğini göstermektedir. Dinamik eşik ayarı sonrasında TabNet en yüksek duyarlılığı (Recall: $0,573 \pm 0,064$) sağlamış, bu da onu erken risk tespitinde öne çıkarmıştır. SHAP sonuçlarına göre en etkili değişkenler Fazla Mesai ve Hisse Senedi Opsiyon Seviyesidir.

Anahtar Kelimeler: Çalışan ayrılma tahmini, tabnet, xgboost, açıklanabilir yapay zekâ, shap, sınıf dengesizliği problem.

Introduction

Employee turnover, frequently described as churn, poses a significant threat to both organizational continuity and financial sustainability. Its impact extends well beyond recruitment and training expenses. When experienced employees leave, organizations often face less visible consequences, including weakened team cohesion, erosion of accumulated knowledge, and interruptions in operational routines (Saradhi & Palshikar, 2011). With the expansion of data-driven practices, HR management has gradually moved away from intuition-led decisions toward analytical approaches, commonly referred to as HR Analytics or People Analytics. Within this shift, identifying employees who may be at risk of departure has become a central concern. Anticipating attrition enables organizations to design targeted retention strategies rather than reacting after turnover occurs.

In earlier research, employee attrition was commonly examined using statistical techniques such as Logistic Regression. While these models offer interpretability and simplicity, their linear structure makes it difficult to fully capture the more complex and non-linear patterns that often characterize human behavior.

Over time, Machine Learning (ML) approaches have gained prominence, largely due to their ability to model intricate relationships within HR datasets. Among these, tree-based ensemble techniques, most notably Random Forest and gradient boosting frameworks such as XGBoost and CatBoost, have emerged as leading methods for structured tabular data (Grinsztajn et al., 2022). Empirical findings reported by Fallucchi et al. (2020) indicate that these models tend to outperform traditional statistical approaches, particularly in predictive accuracy and resilience to noisy data.

Although tree-based approaches continue to dominate tabular prediction tasks, the broader progress in Deep Learning (DL) has encouraged researchers to experiment with Artificial Neural Networks (ANN) in structured domains as well. Neural models have proven highly effective for unstructured data such as images and text; however, their

advantage is less clear when applied to structured HR datasets. This has led to ongoing discussion regarding whether DL genuinely offers added value in such contexts.

More recently, Google Cloud AI proposed TabNet, an architecture designed to integrate aspects of tree-based interpretability with the feature representation strength of deep neural networks (Arik & Pfister, 2021). In theory, TabNet attempts to narrow the gap between these two modeling traditions. Nevertheless, direct empirical comparisons between contemporary DL-based models and well-established Gradient Boosting techniques, particularly on small and imbalanced HR datasets, are still relatively scarce in the existing literature.

Beyond model performance, there remains a practical challenge in translating predictive results into managerial action. High accuracy alone does not guarantee adoption within organizations. Decision-makers typically expect clarity about the reasoning behind a risk prediction, especially when it concerns individual employees. When models operate as opaque systems, they tend to generate skepticism rather than trust (Molnar, 2020). Another recurring issue in attrition research relates to class imbalance. In most HR datasets, employees who leave represent a relatively small minority compared to those who stay (Yurtsever, 2024). Studies frequently report overall accuracy without sufficiently addressing how well the minority class is detected. As a result, performance metrics can appear strong while failing to meaningfully identify at-risk employees. The interpretability problem further complicates implementation. Organizations may receive a list of employees predicted to resign, yet remain uncertain about the underlying drivers of these predictions. Without insight into contributing factors, the practical usefulness of such outputs becomes limited. To mitigate this concern, Explainable Artificial Intelligence (XAI) approaches, particularly SHAP (SHapley Additive exPlanations), have gained prominence as tools capable of clarifying model behavior at both local and global levels (Lundberg & Lee, 2017).

This study seeks to respond to these limitations by introducing a hybrid methodological framework

that integrates systematic performance evaluation with structured interpretability analysis. Rather than focusing solely on predictive strength, the approach emphasizes both reliability and transparency within an applied HR context. The main contributions of this paper to the existing body of literature can be summarized as follows:

1. Comparative Evaluation on Structured HR

Data: This study offers a systematic empirical comparison between widely used tree-based models (XGBoost, CatBoost, and Random Forest) and neural approaches (ANN and TabNet). The analysis is conducted using strictly separated outer test folds within a nested cross-validation framework, allowing for a controlled assessment of performance. In doing so, the study contributes to the ongoing discussion regarding the relative effectiveness of tree ensembles and neural architectures in HR analytics settings.

2. Methodologically Sound Handling of Class

Imbalance: In contrast to approaches that apply oversampling before data splitting, thereby introducing potential data leakage, this study confines the use of SMOTE (Synthetic Minority Over-sampling Technique) strictly to the training data. By preserving the integrity of validation and test sets, the evaluation reflects more realistic and reliable performance estimates.

3. Adaptive Threshold Calibration:

The study shows that relying on the conventional 0.5 classification threshold can be inadequate in the context of imbalanced attrition datasets. To address this limitation, a dynamic threshold tuning strategy is introduced, leading to a marked improvement in Recall and more effective identification of at-risk employees.

4. Explainable Decision-Oriented Framework:

Through the integration of SHAP-based interpretation, model outputs are translated into insights that can be meaningfully evaluated by management. Rather than presenting isolated

risk scores, the analysis clarifies how factors such as OverTime, Stock Option Level, and Monthly Income interact and contribute to attrition risk, thereby making predictive results more actionable in organizational settings.

Theoretical background

Research on employee attrition has gradually shifted in scope over time. What was once primarily examined within organizational psychology has increasingly become a focal issue in computational intelligence and data mining. As analytical capabilities expanded, predictive modeling moved to the forefront of HR research.

Existing studies in this area can generally be grouped into three main strands: applications based on traditional machine learning techniques, investigations exploring deep learning models for structured tabular data, and more recent work centered on Explainable AI (XAI) within human resource management contexts.

Initial work on attrition prediction largely employed conventional statistical techniques, including Logistic Regression and Naive Bayes classifiers. While these approaches offered interpretability and ease of implementation, their linear assumptions limited their ability to represent more intricate relationships within employee-related variables (Alao & Adeyemo, 2013). As HR datasets expanded in both scale and structural complexity, these limitations became more apparent. In response, ensemble-based learning methods began to attract greater attention as more flexible alternatives.

Jain and Nayyar (2019) conducted a comparative analysis of Support Vector Machines (SVM) and Random Forest using the IBM HR dataset. Their findings indicated that Random Forest achieved higher predictive performance, which they attributed to its capacity to manage high-dimensional feature spaces and mitigate overfitting through the bagging mechanism.

In a related study, Fallucchi et al. (2020) examined large-scale employee records with multiple classification algorithms. Their results showed that

Gradient Boosting models, particularly XGBoost, consistently surpassed individual learners such as Decision Trees and k-Nearest Neighbors (k-NN) in terms of predictive reliability.

The strong performance of tree-based methods in attrition modeling is often linked to their resilience to outliers and their ability to process heterogeneous data structures, including both categorical and numerical variables, without requiring extensive preprocessing steps (Grinsztajn et al., 2022).

Although Deep Learning (DL) has transformed fields such as computer vision and natural language processing, its effectiveness on structured tabular HR data is still debated. Conventional Artificial Neural Networks (ANN) do not always perform optimally in this setting, partly because tabular datasets lack the spatial or sequential structures that architectures like CNNs and RNNs are designed to leverage (Shwartz-Ziv & Armon, 2022). In addition, neural networks typically benefit from large training samples, whereas HR datasets often consist of only a few thousand observations, limiting their capacity to generalize.

To address these limitations, Arik and Pfister (2021) proposed TabNet, a deep learning model specifically tailored for tabular inputs. TabNet introduces a sequential attention mechanism that selectively focuses on relevant features at each decision stage. In this respect, it resembles the stepwise reasoning of decision trees, while still maintaining the end-to-end optimization characteristic of neural networks.

Despite these conceptual strengths, direct empirical comparisons between TabNet and established gradient boosting techniques such as XGBoost remain limited, particularly in scenarios involving small and imbalanced datasets. This lack of evidence highlights a clear area for further investigation.

A frequently cited concern in HR analytics is the so-called “black box” issue. Even when models such as XGBoost or ANN achieve strong predictive performance, their internal decision processes are not easily interpretable. This opacity can reduce managerial confidence, particularly in contexts

where personnel decisions carry ethical and organizational implications.

More recently, researchers have incorporated SHAP (SHapley Additive exPlanations) to make model behavior more transparent. However, a number of studies remain limited to ranking individual feature importance. Such an approach may overlook more nuanced, non-linear interactions, for instance how OverTime and Stock Option Level jointly influence turnover risk, despite the fact that XAI methods are capable of uncovering these layered relationships.

At the same time, methodological rigor in model validation has received increasing attention. Fang and Zhang (2025) stressed the importance of strictly leakage-free resampling procedures when dealing with imbalanced HR datasets, warning that improper preprocessing can lead to overly optimistic performance results. In a related benchmark study, Thapliyal et al. (2024) observed that although deep learning architectures such as TabNet may offer improved interpretability and enhanced recall under certain conditions, tree-based ensemble models often retain greater stability when working with relatively small tabular datasets. These findings suggest that combining explainability tools with careful threshold calibration may provide a more balanced and practically relevant evaluation strategy.

Although prior research has produced substantial findings, several important limitations persist:

1. **Data Handling and Leakage Concerns:** Many studies utilizing the IBM HR dataset implement oversampling methods such as SMOTE before dividing the data into training and test subsets. Applying resampling at this stage can introduce data leakage, as information from the full dataset may inadvertently influence model evaluation. This practice often inflates reported performance and yields results that may not generalize reliably beyond the experimental setting (Santos et al., 2018).
2. **Limited Comparative Evidence Between TabNet and Gradient Boosting:** Systematic performance evaluations contrasting TabNet with

leading tree-based ensemble models remain scarce, particularly in scenarios involving small and imbalanced HR datasets. As a result, it is still unclear whether the theoretical advantages attributed to TabNet translate into consistent empirical gains under these specific data conditions.

- 3. Limited Focus on Practical Actionability:** A considerable portion of the literature concludes at the stage of predictive modeling, without extending the findings into implementable decision frameworks. Only a small number of studies integrate adaptive threshold calibration with XAI techniques to move beyond prediction and offer a structured, actionable support mechanism for managerial decision-making.

This research is designed to address these identified limitations through a carefully structured methodological approach. By combining controlled evaluation procedures with interpretability-focused analysis, the study seeks to provide both empirical clarity and practical relevance. The key contributions can be outlined as follows:

- **Leakage-Free Evaluation Procedure:** SMOTE is applied strictly within the training data, while the test set is left entirely untouched. This separation preserves the integrity of the evaluation process and allows performance metrics to better reflect real-world deployment conditions.
- **Broad Comparative Benchmarking:** The study contrasts three groups of models within a unified evaluation framework: established ensemble methods (Random Forest, XGBoost, and CatBoost) and more recent neural architectures (ANN and TabNet). This structure enables a balanced assessment across traditional, state-of-the-art, and contemporary approaches.
- **Strategic Interpretive Insight:** Through SHAP-based summary analyses, the study extends beyond predictive output and identifies decision-relevant patterns for HR practition-

ers. In particular, it clarifies how workload indicators such as OverTime interact with compensation-related factors like Stock Option Level, revealing practical trade-offs that can inform targeted retention strategies.

Methodology

The empirical analysis is based on the publicly accessible IBM HR Analytics Employee Attrition dataset, which is frequently used as a benchmark in attrition-prediction studies (Subhash, 2017). It includes a range of demographic, occupational, and organizational variables, such as age, tenure, job role, overtime status, performance evaluation scores, compensation-related indicators, and measures of work–life balance.

The outcome variable, *Attrition*, is defined as a binary label indicating whether an employee left the organization (Yes/1) or remained (No/0). Following the removal of duplicate records and basic consistency checks, the final dataset comprised 1,470 employees.

An initial exploratory analysis highlighted a pronounced class imbalance. Of the total sample, 237 employees (16.1%) were labeled as having left, whereas 1,233 (83.9%) remained with the company. Such imbalance can distort model behavior, as many standard classifiers tend to favor the majority class when minimizing overall error (He & Garcia, 2009).

All predictors were systematically reviewed for missing entries, irregular category definitions, and potential sources of leakage prior to modeling. The dataset does not contain any personally identifiable information. In total, 35 input features were included in the analysis.

For categorical variables, Label Encoding was applied in line with the structural characteristics of tree-based ensemble methods. One-Hot Encoding was deliberately not used, as it substantially increases feature dimensionality and often produces sparse representations. In relatively small samples ($N = 1,470$), this expansion can reduce computational efficiency and complicate split optimization in decision tree algorithms.

Variables with an inherent ranking structure such as performance rating and education level were preserved in their ordinal form. Continuous features were standardized only for scale-sensitive models, while tree-based algorithms were trained using the original, unscaled values.

The methodological framework is structured around four sequential stages: (i) separation of training and test data, (ii) treatment of class imbalance, (iii) model development and hyperparameter tuning, and (iv) performance assessment combined with interpretability analysis.

At each stage, particular attention was given to preventing information leakage and maintaining a realistic evaluation setup. The overall design was intended to approximate conditions that would be encountered in practical deployment rather than an artificially optimized experimental environment.

To assess model stability and strengthen statistical reliability, evaluation was conducted using a Stratified 5-Fold Cross-Validation framework rather than relying on a single train-test partition. The dataset was divided into five folds while preserving the original class distribution in each subset. During each outer cycle, four folds were allocated for model development and one fold was reserved as the outer test fold.

Importantly, class rebalancing through SMOTE was performed only on the training folds within each iteration. The corresponding outer test fold retained its natural class proportions to prevent any form of leakage. Reported performance values represent the mean and standard deviation calculated across the five outer test folds.

To further avoid bias, threshold optimization was examined on an inner validation split derived from the training folds, without using the outer test fold for decision-threshold selection. This design reduces the risk of optimistic bias and allows the effect of threshold adjustment, particularly for TabNet, to be interpreted separately from the main benchmark comparison.

To mitigate the imbalance in the target distribution, the Synthetic Minority Over-sampling Technique (SMOTE) introduced by Chawla et al. (2002)

was utilized. In contrast to simple random over-sampling, which duplicates existing minority observations, SMOTE creates new synthetic instances by interpolating between neighboring samples from the minority class.

More specifically, for a given minority observation x , a synthetic instance x_{new} is generated according to the following formulation:

$$x_{new} = x_i + \lambda (x_{z_i} - x_i)$$

Here, x_{z_i} denotes one of the k -nearest neighbors of x_i , and λ represents a random value drawn from the interval between 0 and 1. To avoid data leakage, an issue frequently observed when synthetic samples inadvertently influence validation data, SMOTE was implemented strictly within the training set. The test data retained its original class distribution so that model performance could be evaluated under conditions closer to practical deployment.

The standard SMOTE algorithm was deliberately preferred over more advanced variants such as Borderline-SMOTE or SMOTE-ENN. This choice was made to maintain methodological clarity and reproducibility. Considering the moderate sample size ($n = 1,470$), the conventional SMOTE approach provides a reasonable balance: it improves minority class representation without substantially increasing the likelihood of introducing artificial noise, which can become more pronounced with boundary-focused resampling techniques.

A heterogeneous group of predictive models was included in order to reflect both established machine learning techniques and more recent deep learning architectures tailored to tabular data. This selection enables a balanced comparison across methodological traditions rather than privileging a single modeling paradigm.

Random Forest (RF): As a reference ensemble model, Random Forest was included in the analysis. This algorithm builds multiple decision trees during training and aggregates their outputs to produce the final prediction. By relying on bootstrap aggregation (bagging), it reduces the variance typically associated with single decision trees and helps control overfitting (Breiman, 2001).

XGBoost (eXtreme Gradient Boosting): XGBoost represents an optimized and scalable realization of the gradient boosting framework. It incorporates a regularized objective function that penalizes model complexity, thereby reducing the likelihood of overfitting. Owing to this balance between flexibility and regularization, XGBoost has consistently demonstrated strong performance in structured tabular prediction tasks (Chen & Guestrin, 2016).

Artificial Neural Networks (ANN): A custom Multi-layer Perceptron (MLP) was implemented using the PyTorch framework. The network architecture comprised three fully connected layers. Given the relatively limited sample size, Batch Normalization and Dropout ($p = 0.3$) were incorporated between hidden layers to enhance generalization and reduce overfitting risk. Model training was conducted using the Adam optimizer, with binary cross-entropy as the loss function.

TabNet: TabNet, introduced by Google Cloud AI as a neural architecture tailored for tabular datasets (Arik & Pfister, 2021), was also incorporated into the experimental design. The model relies on a sequential attention mechanism that selectively focuses on relevant features at each decision stage. Through this process, it generates soft feature selection masks, allowing it to approximate the step-wise reasoning behavior typical of tree-based methods while preserving the end-to-end learning structure of neural networks.

In classification tasks involving imbalanced classes, the conventional probability threshold of 0.5 may not be appropriate. Under this default rule, an employee assigned a predicted risk of 0.45 would be labeled as “Stay,” even if that probability reflects substantial underlying risk. Such cases contribute to False Negatives, where individuals who are likely to leave are not correctly identified.

To enhance the identification of at-risk employees while maintaining a reasonable balance with precision, a Dynamic Threshold Tuning procedure was applied. Instead of fixing the cutoff at 0.5, probability thresholds t were systematically varied

within the interval $[0, 1]$. For each candidate value, classification performance was recalculated, and the threshold that yielded the highest F1-score was selected as optimal:

$$F1 = 2x \frac{Precision \times Recall}{Precision + Recall}$$

Model performance was evaluated on the untouched outer test folds using multiple complementary metrics, including F1-score, Precision, Recall, ROC-AUC, and the Confusion Matrix. Assessing several indicators allowed for a more balanced interpretation of classification behavior rather than relying on a single summary statistic. Beyond statistical performance, results were also considered in light of organizational priorities. For instance, in an early-warning HR framework, a model with higher recall even, if accompanied by moderate precision, may be more suitable, as identifying potentially at-risk employees can be more critical than minimizing false alarms.

To better understand how the top-performing models arrived at their predictions, SHAP (SHapley Additive exPlanations) values were calculated (Lundberg & Lee, 2017). This method provides interpretive insights at two levels: global explanations, which summarize overall feature influence across the full dataset, and local explanations, which clarify the reasoning behind predictions for individual employees.

To maintain consistency in interpretability across structurally different models, SHAP implementations were selected according to the underlying architecture. For tree-based methods (Random Forest, XGBoost, and CatBoost), the exact TreeSHAP algorithm was applied, allowing Shapley values to be computed efficiently in polynomial time.

For neural architectures (ANN and TabNet), a model-agnostic approach was required; therefore, KernelSHAP was employed. Given its computational intensity, the background reference set was reduced using K-means clustering ($k = 100$) to obtain a representative summary of the data. The number of coalition samples was kept at the default configuration to preserve methodological transparency.

Additionally, a consistency check was conducted for TabNet. A qualitative comparison showed that the model’s internal attention masks reflecting feature selection during inference were broadly aligned with the SHAP-based explanations. Both approaches consistently highlighted *OverTime* and *MonthlyIncome* as the most influential predictors.

Results and Discussion

This section summarizes the findings from the comparative evaluation of multiple modeling strategies, including a baseline ensemble approach (Random Forest), gradient boosting techniques (XGBoost and CatBoost), and neural architectures (ANN and TabNet).

was applied, TabNet exhibited the strongest sensitivity, achieving the highest recall in identifying employees at risk of attrition.

As shown in Table 1, both ensemble and neural models produced relatively strong overall accuracy levels, with CatBoost achieving the highest mean Accuracy (0.872 ± 0.015). CatBoost also delivered the most balanced overall profile in terms of Precision and F1-score. In contrast, the threshold-adjusted analyses reported in Tables 2 and 3 show that TabNet achieved the strongest sensitivity to attrition cases, reaching Recall values up to 0.573 ± 0.064 under the optimized decision setting. This distinction is important because threshold adjustment was introduced specifically to improve minority-class detection without contaminating the outer test folds.

Table 1. Comparative Performance Metrics of the Models

Model Architecture	Accuracy	Precision	Recall	F1-Score	ROC-AUC	PR-AUC	Brier Score
XGBoost	0.856±0.011	0.599±0.069	0.354± 0.045	0.442± 0.039	0.811± 0.013	0.548± 0.030	0.113± 0.007
CatBoost	0.872±0.015	0.731±0.126	0.350± 0.052	0.468± 0.053	0.825± 0.017	0.609± 0.030	0.097± 0.006
ANN (PyTorch)	0.832±0.009	0.476±0.027	0.422± 0.038	0.447± 0.033	0.761± 0.015	0.485± 0.039	0.152± 0.007
TabNet	0.761±0.028	0.346±0.043	0.531± 0.080	0.417± 0.050	0.706± 0.047	0.371± 0.080	0.194± 0.022

Note: Reported values are mean ± standard deviation across the five outer test folds. Table 1 summarizes the main benchmark results used for model comparison, while the specific effect of threshold adjustment is illustrated separately in Tables 2 and 3 for TabNet.

Model performance was assessed using Accuracy, Precision, Recall, and F1-score. However, due to the pronounced class imbalance in the dataset (16% attrition rate), particular emphasis was placed on Recall and F1-score. From an HR management perspective, correctly identifying employees who are likely to leave, thereby maximizing True Positives, holds greater practical importance than optimizing overall accuracy alone.

The classification outcomes summarized in Table 1 represent the mean and standard deviation of performance metrics obtained across the outer test folds of the stratified 5-fold nested evaluation procedure. Table 1 reports the main benchmark results used for model comparison, while the specific effect of threshold adjustment is illustrated separately in Tables 2 and 3 for TabNet.

The ensemble-based tree models, particularly XGBoost and CatBoost, demonstrated consistent and stable performance in terms of overall accuracy. However, once dynamic threshold calibration

Within the Gradient Boosting group, CatBoost delivered the most balanced profile. It achieved the highest F1-score (0.468 ± 0.053) while also maintaining the strongest Precision (0.731 ± 0.126), indicating a comparatively stable trade-off between correctly identifying churners and limiting false alarms.

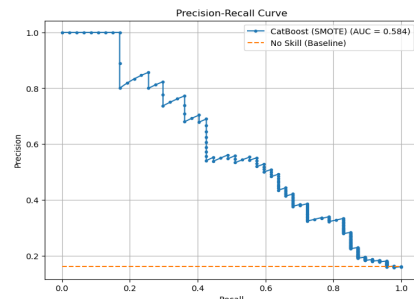


Figure 1. Precision-Recall Curve

To examine model behavior beyond a single cutoff value, Precision–Recall (PR) curves were constructed (see Figure 1). In contrast to ROC

curves, PR curves are generally more suitable for heavily imbalanced datasets because they emphasize performance on the minority class rather than overall discrimination.

The plotted curves indicate that CatBoost maintains relatively stable performance across threshold variations. In comparison, TabNet exhibits a wider operational region in which higher recall can be achieved, making it particularly relevant when the primary objective is to detect employees at risk of leaving. This graphical analysis supports the decision to further examine threshold adjustment through the supplementary TabNet analyses reported in Tables 2 and 3.

Table 2. Illustration of Threshold Adjustment for TabNet

Model	Threshold Setting	Threshold Value	Precision	Recall	F1-Score
Tab-Net	Default	0.50	0.346 ± 0.043	0.531 ± 0.080	0.417 ± 0.050
Tab-Net	Optimized	0.38	0.316 ± 0.051	0.573 ± 0.064	0.407 ± 0.048

Note: The table illustrates the practical effect of threshold adjustment for TabNet. The default setting corresponds to the conventional cutoff of 0.50, whereas the optimized setting reflects the fold-specific threshold selected on the inner validation split, with an average value of 0.38 across folds.

Table 2 provides a compact illustration of how threshold adjustment affected TabNet within the same modeling setup. Lowering the average cutoff from the conventional 0.50 to 0.38 increased Recall from 0.531 to 0.573, indicating improved sensitivity to minority-class attrition cases. This gain, however, was accompanied by a reduction in Precision and a slight decrease in F1-score, reflecting the expected trade-off between sensitivity and decision conservatism in imbalanced prediction settings.

From a risk-oriented standpoint, the threshold-adjusted TabNet model delivered the strongest Recall performance (0.573 ± 0.064). In practical terms, this means that the model correctly identified nearly 57% of employees who ultimately left the organization, substantially higher than the tree-based alternatives (for example, XGBoost Recall: 0.354).

This increase in sensitivity was accompanied by a decline in Precision (0.316 ± 0.051), reflecting a

higher number of false positives. Nevertheless, within an HR early-warning framework, prioritizing the detection of potential churners is often more valuable than minimizing false alerts, as it enables timely and preventive retention measures.

For the TabNet architecture in particular, threshold adjustment played a decisive role. Across the outer folds, the average optimized decision threshold was 0.38, compared with the conventional default of 0.50. This lower cutoff increased the model's sensitivity to minority-class cases and helped raise Recall to 0.573 ± 0.064 in the threshold-adjusted evaluation, illustrating that the default threshold would have been overly conservative for an imbalanced attrition setting. This shift highlights the importance of threshold calibration when working with imbalanced HR datasets, where default settings may under-identify at-risk cases.

Table 3. Compact Ablation Analysis for TabNet

Configuration	Resampling	Threshold Strategy	Precision	Recall	F1-Score
TabNet + SMOTE	SMOTE	Default (0.50)	0.346 ± 0.043	0.531 ± 0.080	0.417 ± 0.050
Full TabNet Pipeline	SMOTE	Optimized (avg. 0.38)	0.316 ± 0.051	0.573 ± 0.064	0.407 ± 0.048

To isolate the practical contribution of threshold selection within the same architecture, Table 3 presents a compact ablation analysis for TabNet under a fixed SMOTE-based resampling setup. The comparison shows that lowering the decision threshold from the default 0.50 to the optimized average of 0.38 increased Recall from 0.531 to 0.573, confirming that part of TabNet's advantage in minority-class detection stemmed from threshold calibration rather than model architecture alone. This gain, however, was accompanied by a reduction in Precision and a slight decrease in F1-score, indicating the expected trade-off between sensitivity and decision conservatism in imbalanced attrition prediction.

To gain a clearer understanding of prediction patterns, the Confusion Matrices of the top-performing models were examined in detail. Figure 2

presents the distribution of predicted outcomes for the TabNet model, illustrating how true positives, false positives, true negatives, and false negatives are distributed under the optimized threshold.

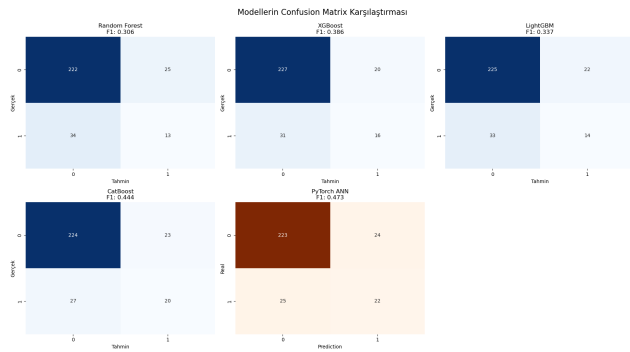


Figure 2. Confusion Matrix of the Optimized TabNet Model.

Within a representative outer test fold, 47 employees were labeled as having left the organization. Under the optimized configuration, TabNet correctly identified 26 of these cases, whereas the default XGBoost setup detected 16. This contrast indicates that, in relatively small tabular datasets, calibrating decision thresholds in neural architectures may enhance minority-class detection beyond what standard tree-based configurations achieve.

However, predictive performance alone is insufficient for managerial application. Interpreting the underlying drivers of attrition remains equally important for strategic planning. To examine feature influence and directional effects, SHAP (SHapley Additive exPlanations) summary visualizations were generated based on the Gradient Boosting model (XGBoost). The resulting SHAP summary plot is shown in Figure 3.

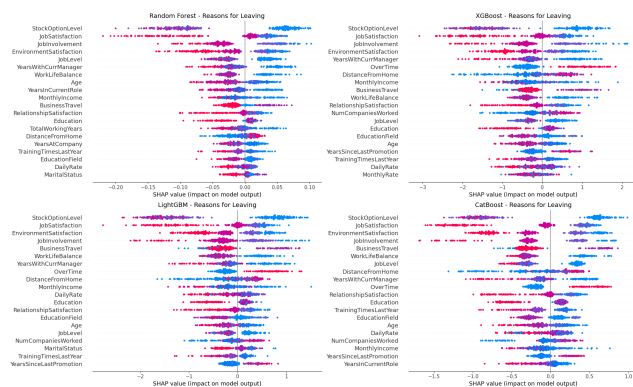


Figure 3. SHAP Summary Plot illustrating the impact of top features on attrition risk.

The SHAP-based interpretation highlights three primary determinants that exert the strongest influence on employee attrition outcomes:

1. **OverTime:** This variable exerts the strongest influence on the model's predictions. In the SHAP summary visualization, observations marked in red (corresponding to "Yes" for OverTime) are predominantly positioned on the positive side of the SHAP value axis. This pattern indicates that over-time work contributes positively to the predicted probability of attrition, suggesting a clear association between extended working hours and an increased likelihood of employee departure.
2. **StockOptionLevel:** The analysis indicates a pronounced inverse association for this variable. Observations representing lower stock option levels (blue markers) tend to correspond with higher SHAP values, implying an elevated probability of attrition. In contrast, higher stock option levels (red markers) are generally linked to negative SHAP values, shifting predictions toward retention. This pattern suggests that equity-based compensation may function as a meaningful stabilizing factor within the organization.
3. **MonthlyIncome:** A comparable trend is evident for monthly income. Lower salary levels are generally linked with positive SHAP contributions, indicating a higher probability of leaving the organization. Nevertheless, the broader dispersion of SHAP values suggests that income alone does not fully determine retention outcomes. Compared with the workload variable (OverTime), compensation appears to function

as an important but not independently decisive factor in attrition dynamics.

Additional influential variables include JobLevel and Age. Lower hierarchical levels and younger employees tend to display a greater likelihood of turnover, as reflected by their positive SHAP contributions. DistanceFromHome also emerges as a relevant predictor; employees with longer commuting distances show increased attrition probability, suggesting that logistical strain may contribute to departure decisions.

To assess the impact of the resampling procedure, an ablation analysis was performed. When SMOTE was not applied, the baseline configuration produced an F1-score of 0.390 ± 0.040 . After introducing SMOTE within the training folds, the F1-score increased to 0.468 ± 0.053 .

A paired t-test was conducted to examine whether this improvement was statistically meaningful at the fold level. The resulting p-value of 0.022 suggests that the performance gain associated with SMOTE was consistently observed across folds. Nevertheless, this result should be interpreted as supportive evidence, given the limited number of cross-validation folds and the exploratory role of this comparison within the broader benchmark design.

In addition to resampling, a cost-sensitive learning strategy based on class weighting was also examined. This approach also produced a competitive performance profile. While class weighting proved particularly effective for tree-based ensembles, SMOTE was maintained as the principal imbalance-handling technique. The rationale for this choice lies in the training dynamics of TabNet: balanced mini-batch distributions contribute to more stable optimization and improved convergence behavior in neural network architectures.

To further examine architectural differences, a paired fold-level comparison was conducted between the strongest tree-based model (CatBoost) and the threshold-optimized TabNet model with respect to Recall. The observed difference was statistically notable at the fold level ($p < 0.01$), suggesting that TabNet consistently achieved higher

sensitivity to minority-class instances than the gradient boosting baseline under the evaluated settings. However, this comparison should be interpreted cautiously, as the fold-based test is intended to provide supportive rather than definitive inferential evidence.

Discussion

This study seeks to construct a structured benchmark that contrasts established tree-based ensemble techniques with more recent deep learning architectures in the context of employee attrition modeling. At the same time, it explicitly addresses concerns related to model opacity by incorporating Explainable AI (XAI) methods into the evaluation framework.

The empirical findings contribute on two levels: they shed light on how different algorithms perform under small-scale tabular conditions, and they clarify the underlying factors that shape workforce turnover from a strategic management perspective.

The results point to a differentiated performance profile rather than a clear-cut winner. After applying threshold calibration, TabNet reached the highest Recall (0.573 ± 0.064), indicating stronger sensitivity to at-risk employees. However, when broader stability indicators are considered, the Gradient Boosting models remained highly competitive, with CatBoost achieving the strongest overall F1-score (0.468 ± 0.053) and XGBoost following with 0.442 ± 0.039 . In other words, TabNet's gain in recall did not translate into a decisive overall advantage.

This pattern is consistent with the large-scale benchmark analysis reported by Grinsztajn et al. (2022), which concluded that tree-based ensembles continue to represent the most reliable choice for medium-sized tabular datasets (fewer than 10,000 observations).

The comparatively lower Precision observed in the Deep Learning models (ANN and TabNet) relative to CatBoost may be linked to the inductive characteristics of neural architectures. Neural networks generally benefit from larger sample sizes to

learn stable and well-separated feature representations (Shwartz-Ziv & Armon, 2022). In smaller datasets, this structural tendency can limit their ability to achieve high precision without careful regularization and calibration.

Accordingly, the findings suggest a practical implication for HR analytics contexts, where datasets are often modest in scale. Gradient Boosting methods appear to provide a stronger balance between predictive performance and computational efficiency under such conditions. In contrast, TabNet may be more appropriate in settings where the primary objective is to maximize sensitivity, particularly when early detection of potential attrition cases outweighs concerns about false positives.

With respect to probabilistic reliability, the Brier Score results (Table 1) indicate that the tree-based ensembles especially CatBoost (0.097) produced comparatively well-calibrated probability estimates without additional adjustment. By contrast, TabNet yielded a higher Brier value (0.194), suggesting less stable probability calibration and greater uncertainty in borderline predictions.

Although post-hoc calibration techniques such as Platt scaling or Isotonic Regression were evaluated as potential remedies, the study prioritized dynamic threshold adjustment as a more direct intervention. This choice was made to improve decision sensitivity while preserving the original probability structure of the models.

Empirical testing indicated that probability calibration methods can compress predicted scores in highly imbalanced settings. In practice, this often lowers the estimated probabilities assigned to minority-class cases, which may inadvertently decrease Recall.

Given that the central aim of the study was to enhance the identification of at-risk employees, preserving sensitivity was prioritized over probability smoothing. For this reason, adjusting the classification threshold directly proved more aligned with the study's objective than applying post-hoc calibration techniques.

In this respect, the reported Brier scores should be interpreted as complementary calibration indi-

cators rather than as the primary optimization target of the study. While lower Brier values for the tree-based models indicate more stable probability estimates, the benchmark was intentionally designed to prioritize operational sensitivity in identifying potential attrition cases. Accordingly, the calibration findings are informative for model interpretation, but they do not override the study's practical emphasis on Recall-oriented decision support.

A closer examination of prior research underscores the relevance of the leakage-free validation design adopted in this study. Earlier investigations using the IBM dataset, such as Jain and Nayyar (2019) and Alao and Adeyemo (2013), reported accuracy values above 90%. However, in several cases, oversampling procedures were performed before the data were partitioned into training and testing subsets. This sequencing introduces the risk of information leakage, potentially inflating performance metrics.

By contrast, the present study preserved strict separation between training and test data, applying SMOTE exclusively within the training folds. As a result, the reported F1-scores likely provide a more conservative and realistic estimate of model performance under deployment-like conditions.

Although the reported performance values may appear more modest at first glance, they reflect a more reproducible and deployment-oriented benchmark. In this sense, the results align with the concerns articulated by Santos et al. (2018), who cautioned against inflated performance estimates arising from improper validation practices in imbalanced classification settings.

Beyond the comparative evaluation of algorithms, the interpretability analysis offers practical insights for managerial decision-making. The SHAP summary visualization clearly identifies OverTime as the most influential determinant of attrition risk.

This empirical pattern lends support to theoretical perspectives in organizational psychology, particularly arguments associated with burnout dynamics. In this context, sustained workload pressure appears to function as a stronger withdrawal

driver than compensation-related factors alone (Saradhi & Palshikar, 2011).

An additional insight emerges from the negative association between StockOptionLevel and attrition probability. While monthly income remains relevant, the model suggests that long-term, equity-based incentives exert a stronger stabilizing influence than short-term salary adjustments.

From a strategic standpoint, this pattern implies that organizations may achieve more sustainable retention outcomes by emphasizing ownership-oriented compensation structures. Rather than focusing solely on increasing base pay, designing incentive systems that promote long-term organizational attachment could prove more effective in reducing voluntary turnover.

Fairness, Bias, and Ethical Implications

While predictive effectiveness was a central focus of this study, responsible implementation in HR settings requires careful attention to fairness considerations. Beyond the qualitative evidence drawn from SHAP feature rankings—which indicated that protected variables such as Gender appeared relatively low in importance (rank > 20) a formal quantitative fairness assessment was also performed using established evaluation metrics.

The fairness evaluation produced a Demographic Parity Difference (DPD) of 0.0422 and an Equalized Odds Difference (EOD) of 0.0511. Both figures fall comfortably below the commonly referenced 0.10 threshold used to flag potential disparate impact.

These results suggest that the model's decisions are largely influenced by work-related and behavioral variables such as OverTime and JobSatisfaction rather than demographic attributes. In practical terms, the predictive patterns appear to be aligned more with organizational conditions than with protected characteristics.

Although fairness diagnostics were included to examine whether predictive gains were associated with disproportionate group-level disparities, fairness was not positioned as the primary research objective of the study. Instead, it was used as a

complementary diagnostic layer alongside recall-oriented attrition prediction. Extending the analysis to a larger number of demographic groups and intersectional subgroups could offer additional nuance; however, such extensions may reduce estimate stability in smaller partitions of the dataset. Accordingly, the present study reports compact fairness indicators while leaving broader subgroup auditing to future work.

Nevertheless, predictive accuracy alone does not eliminate potential risks associated with deployment. Different types of errors carry distinct organizational consequences. A False Positive, classifying an employee as likely to leave when they have no such intention, may unintentionally influence managerial behavior. For example, reduced developmental investment or exclusion from long-term projects could occur if decision-makers interpret the prediction as a signal of low commitment.

On the other hand, a False Negative reflects a missed retention opportunity. Failing to identify an employee who is genuinely at risk of departure limits the organization's ability to intervene proactively. For this reason, the operational use of attrition models must be accompanied by careful governance and clear guidelines to mitigate unintended side effects.

To reduce the potential harms associated with predictive misclassification, the model should operate strictly within a Human-in-the-Loop framework. In this setup, predictions serve as early indicators rather than automated decisions. For example, a high-risk signal should initiate a supportive dialogue such as a discussion about workload or job satisfaction rather than trigger disciplinary or exclusionary actions.

In addition, any deployment must adhere to applicable data protection standards, including regulations such as GDPR. Automated outputs should not produce binding employment consequences without meaningful human oversight. Maintaining this supervisory layer helps ensure that predictive analytics remains a guidance tool rather than a determinant of employment outcomes.

Conclusion

This research investigated employee attrition prediction through a structured comparative design encompassing traditional machine-learning techniques, gradient-boosted ensembles, a multilayer perceptron, and a recent deep-learning model tailored for tabular data (TabNet). Using the IBM HR Analytics dataset as the empirical basis, the study placed particular emphasis on minority-class detection, restricted SMOTE application to the training data, and incorporated threshold adjustment to achieve a more balanced trade-off between precision and recall.

The findings indicate a differentiated performance pattern. Although deep learning models such as TabNet show notable strength in identifying high-risk employees, particularly in terms of Recall, gradient boosting methods (CatBoost and XGBoost) demonstrate greater overall stability and practical reliability for conventional HR analytics applications involving structured tabular data.

By combining imbalance-handling through SMOTE with interpretability via SHAP, the predictive framework extends beyond classification accuracy and evolves into a structured decision-support mechanism. The SHAP-based explanation layer adds analytical depth, enabling stakeholders to understand not only which employees are at risk, but also the underlying factors influencing those predictions.

The analysis showed that a limited subset of variables, most notably overtime status, compensation-related indicators, tenure, and satisfaction measures, carried the majority of the predictive weight. This concentration of signal is consistent with prior empirical findings in the attrition literature.

Taken together, these results suggest that carefully optimized ensemble models, when combined with suitable XAI techniques, can achieve a meaningful balance between predictive strength and interpretive clarity.

The findings suggest that dependable and transparent attrition-prediction frameworks do

not necessarily require highly intricate model architectures. Instead, a methodologically disciplined pipeline, comprising structured preprocessing, principled imbalance treatment, ensemble-based modeling, and SHAP-driven interpretability, can provide a solid and reproducible foundation.

For organizations aiming to implement data-informed early-warning mechanisms, such an approach offers a balanced baseline: technically rigorous, operationally practical, and sufficiently transparent to support responsible decision-making.

For subsequent investigations, incorporating unstructured information such as narrative performance evaluations or exit interview transcripts may provide additional predictive depth. Applying Natural Language Processing (NLP) methods to these textual sources would enable a multi-modal modeling framework that combines structured HR variables with contextual language signals.

Such an expansion could create a more suitable setting for deep learning architectures, which tend to benefit from richer and higher-dimensional inputs. Under these conditions, neural models may overcome the performance limitations observed in purely tabular settings and potentially exceed the capabilities of tree-based ensembles.

One limitation of the present study lies in its dependence on a single benchmark dataset (IBM HR Analytics). Although this dataset is widely used, results derived from it may not fully reflect patterns observed in other organizational environments. A further limitation is that probability calibration was assessed primarily through aggregate Brier scores; future studies may extend this analysis with reliability diagrams or subgroup-level calibration diagnostics. Another limitation is that fairness was evaluated through a compact set of group-based indicators; future research may extend this analysis to broader demographic and intersectional subgroup assessments.

Future research should therefore replicate and validate the proposed methodological framework using external datasets drawn from different in-

dustries and cultural contexts. Such cross-organizational validation would strengthen the generalizability of the conclusions and clarify whether the observed performance dynamics hold across varying workforce structures.

Declarations

Funding: The author received no financial support for the research, authorship, and/or publication of this article.

Conflicts of Interest: The author declares no conflicts of interest related to this study.

Ethical Approval: This study is based exclusively on a publicly available, anonymized secondary dataset (IBM HR Analytics Employee Attrition dataset) and does not involve direct contact with human participants, intervention, or the collection of personally identifiable data. Therefore, ethics committee approval was not required for this study.

Data Availability: The data used in this study are derived from the publicly available IBM HR Analytics Employee Attrition dataset, which is widely used for benchmarking in employee attrition prediction research. The processed data and analytical framework are available from the author upon reasonable request, subject to the terms and conditions of the original data source.

AI Disclosure: Artificial intelligence and machine learning methods (including Random Forest, XGBoost, CatBoost, ANN, TabNet, and SHAP-based explanations) were used as the subject matter and analytical tools of this research. No generative AI tool was used to produce the scientific conclusions of the study unless otherwise stated by the author.

References

- Alao, D., & Adeyemo, A. B. (2013). Analyzing employee attrition using decision tree algorithms. *Computing, Information Systems & Development Informatics*, 4(1), 17-28.
- Arik, S. O., & Pfister, T. (2021). TabNet: Attentive interpretable tabular learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(8), 6679-6687. <https://doi.org/10.1609/aaai.v35i8.16826>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321-357. <https://doi.org/10.1613/jair.953>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794. <https://doi.org/10.1145/2939672.2939785>
- Fallucchi, F., Coladangelo, M., Giuliano, R., & William De Luca, E. (2020). Predicting employee attrition using machine learning techniques. *Computers*, 9(4), 86. <https://doi.org/10.3390/computers9040086>
- Fang, Y., & Zhang, Z. (2025). Employee turnover prediction model based on feature selection and imbalanced data handling. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2025.3589491>
- Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on typical tabular data? *Advances in Neural Information Processing Systems*, 35, 507-520. <https://doi.org/10.48550/arXiv.2207.08815>
- He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263-1284. <https://doi.org/10.1109/TKDE.2008.239>
- Jain, R., & Nayyar, A. (2019). Predicting employee attrition using XGBoost machine learning approach. *Proceedings of the 2019 International*

- Conference on System Modeling & Advancement in Research Trends (SMART), 113-120. <https://doi.org/10.1109/SYSMART.2018.8746940>
- Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30. <https://doi.org/10.48550/arXiv.1705.07874>
- Molnar, C. (2020). *Interpretable machine learning: A guide for making black box models explainable*. Retrieved from: <https://christophm.github.io/interpretable-ml-book/>
- Santos, M. S., Soares, J. P., Abreu, P. H., Araujo, H., & Fujii, J. (2018). Cross-validation for imbalanced datasets: Avoiding overoptimistic performance estimation. *IEEE Computational Intelligence Magazine*, 13(4), 59-79. <https://doi.org/10.1109/MCI.2018.2866730>
- Saradhi, V. V., & Palshikar, G. K. (2011). Employee churn prediction. *Expert Systems with Applications*, 38(3), 1999-2006. <https://doi.org/10.1016/j.eswa.2010.07.134>
- Shwartz-Ziv, R., & Armon, A. (2022). Tabular data: Deep learning is not all you need. *Information Fusion*, 81, 84-90. <https://doi.org/10.1016/j.inffus.2021.11.011>
- Subhash, P. (2017). *IBM HR analytics employee attrition & performance* [Data set]. Kaggle. <https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset>
- Thapliyal, N., Solanki, S., Pandey, N. K., & Papola, S. (2024). Employee attrition analysis using XGBoost. In *2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE)* (pp. 1-6). IEEE. <https://doi.org/10.1109/IC3SE62002.2024.10593326>
- Wang, X., & Zhi, J. (2021). A machine learning-based analytical framework for employee turnover prediction. *Journal of Management Analytics*, 8(3), 351-370. <https://doi.org/10.1080/23270012.2021.1961318>
- Yurtsever, M. (2024). Çalışan yıpranmasını tahmin etmede analitik bir yaklaşım: Topluluk öğrenme yöntemi. *Afyon Kocatepe Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 26(Özel Sayı), 150-160. <https://doi.org/10.33707/akui-ibfd.1462567>