MUKADIMAH
Jurnal Pendidikan, Sejarah, dan Ilmu-ilmu Sosial
Program Studi Pendidikan Sejarah FKIP - Universitas Islam Sumatera Utara, Medan
e-Journal: http://jurnal.uisu.ac.id/index.php/mkd | e-mail: mukadimah@fkip.uisu.ac.id
ISSN: 2622-1373 (Online)
ISSN: 2614-1159 (Print)
Vol. 9 No. 2
August 2025

HR Analytics for Predicting Job Satisfaction in Hybrid Work

Melissa Yunda Hardevianty*, Magister of Human Resource Development, Postgraduate School, Universitas Airlangga, Surabaya, Indonesia
Imam Yuadi, Department of Information and Library Science, Faculty of Social and Political Sciences, Universitas Airlangga, Surabaya, Indonesia

ABSTRACT

This research investigates the application of machine learning classification models for predicting employee job satisfaction, considering demographic, professional, and psychosocial aspects. With a secondary dataset obtained from Kaggle, five supervised learning techniques were applied: Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, and Gradient Boosting. The best model was considered to be Gradient Boosting as it achieved the highest accuracy and F1 score. The model's explainability was enhanced with LIME. LIME enhanced the model's explainability. Stress, work-life balance, and job tenure were identified as the primary factors of job satisfaction. These results support the Job Demands-Resources (JD-R) theory and highlight the model's effectiveness in the hands of HR professionals. The study highlights the need to achieve a balance between predictive accuracy and explainability to ethically align the use of AI in HR analytics, aiming to enhance the well-being of employees and the effectiveness of organizations.

ARTICLE HISTORY

Received 18/08/2025
Revised 01/09/2025
Accepted 22/09/2025
Published 07/10/2025

KEYWORDS

Job Demands - Resources; Job Satisfaction Prediction; Machine Learning.

***CORRESPONDENCE AUTHOR**

melissa.yunda.hardevianty-2024@pasca.unair.ac.id

DOI: https://doi.org/10.30743/mkd.v9i2.12015

INTRODUCTION

The incorporation of hybrid and remote work models has changed the organizational inner processes. They offer more autonomy and freedoms; however, some psychological and practical barriers come with it. Recent analyses of hybrid work models suggest that integration of digital technologies can paradoxically reduce collaboration and individual well-being. Yang, Holtz, and Neff (2023) found that collaboration among teams, within hybrid settings, resulted in disjointed processes and declined innovative outcomes. Similarly, Derks and Bakker (2024) found that the digital stream of messages and the culture of 'always-on' work contribute to increased emotional depletion and burnout. Clearly, both studies highlight the paradox of hybrid working; increased digital burdens accompany increased freedom and autonomy.

Employee well-being, particularly in the context of remote work, has raised concerns that traditional survey methods are unable to address efficiently. Surveys capture the subjective experience of the employee on that specific day, and fail to capture the real-time, multi-layered, and fluctuating experience of stress, satisfaction, and engagement that changes within seconds. As hybrid work becomes the 'new normal', organizations need to employ anticipatory, proactive, automated methods that are better able to cope with these emerging issues. Wang et al. (2023) have acknowledged the potential of AI-assisted analytics, including sentiment analysis and predictive modeling, for providing timely and actionable insights.

To address the concern, this analysis develops a predictive model of job satisfaction that incorporates explainable machine learning techniques and the Job Demands-Resources (JDR) predictive framework. Bakker and Demerouti's (2007) JD-R model, recently refined by Schaufeli and Taris (2023a; 2023b), characterizes employee well-being as the balance between job demands (e.g., workload, digital overloading) and job resources (e.g., autonomy, skill). Its application to digital workplaces has considered digital skills as necessary personal resources and ICT-strain as primary job demands (Bakker & de Vries, 2021; Korunka & Kubicek, 2023). Meta-analytic research has also substantiated the JD-R model's predictive power in longitudinal studies (Lesener, Gusy, & Wolter,

2019), thereby reinforcing the model's use as the empirical foundation for predictive HR analytics in hybrid work.

This study utilizes Gradient Boosting Classification as the predictive engine for operationalizing this framework. Predictive accuracy may be critical, but the adoption of predictions within human resources (HR) also needs to be both ethical and practical. Predictive models, in the absence of moral standards, may undermine the trust HR professionals have in the model when it is used to make sensitive decisions regarding employees. To help alleviate this trust issue, this research utilizes Local Interpretable Model-Agnostic Explanations (LIME). Chalfin and Tambe's (2024) models, as discussed in the literature, suggest that systems where trust is accounted for transparently are more likely to be adopted in organizations.

Figure 1 integrates predictive modeling and LIME explanations. Predictive models are provided with employee information and use it as input to generate outputs, which can then be converted into actionable managerial activities. LIME provides an extra layer of "vendor" confidentiality by "unpacking" single predictive explanations to stress, organizational and work autonomy, work-life balance, and other contributing features. LIME guarantees that every prediction is correct, and every outcome is clear, thereby bridging the psychological and algorithmic dimensions of the JD-R framework.

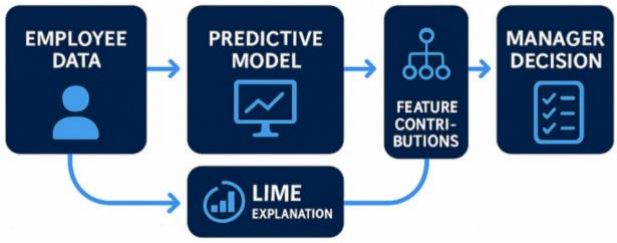


Figure 1. Integrating LIME Explanations in HR Predictive Modeling Workflow

This research delves into the interpretability dilemma of LIME regarding workforce datasets and the class imbalance problem. Employees who belong to subgroups, such as those who are highly dissatisfied, are often overlooked. Johnson and Khoshgoftaar (2024) show how the response to minority outcomes has been improved with the use of sampling and cost-sensitive learning. Moreover, the fairness of automated decision-making regarding the HR context is beginning to emerge. Bogen and Rieke (2024) argue that not all bias can be removed, and they advance the case for the interpretability of automated decision-making systems as ex-ante safeguards.

The model reconstructs the use and predictive model of job satisfaction HR analytics based on the JD-R (Job Demands-Resources) framework for automated decision making. Instead, it seeks to shift the framework from the silos of retrospective surveys to one that enables real-time and proactive assessments of workforce well-being. It enhances the fusion of psychology and machine learning, as well as promotes clarity and ethically driven decision-making for HR practitioners in the digital workplace era.

METHOD

Research Design

This research employs supervised machine learning classifiers to predict employee job satisfaction as part of the quantitative research design. The ML method was selected to address the issues associated with assessments conducted qualitatively through surveys, which are static and retrospective, and do not capture an employee's well-being in real-time (Wang et al., 2023). In comparison, predictive

analytics offers real-time, ongoing surveillance and preemptive detection of patterns of dissatisfaction, which makes it a better fit for hybrid workplaces.

This research used a dataset from Kaggle, which contained a total of 3,025 employees. All analyses were performed in Python due to its reproducibility and scalability. Standard ML steps were followed: data collection, cleansing, preprocessing, model selection, evaluation, and interpretability analysis. Figure 2 shows the research workflow. The design of the workflow illustrated in Figure X ensured sufficient methodological rigor while enabling precise and responsible analysis.

This study's theoretical context and primary emphasis focus on Job Demands–Resources (JD-R) framework by Bakker and Demerouti (2007), which describes employee well-being as a balance, or equilibrium, between job demands (workload, stress, and digital overloading) and job resources (autonomy, skill, and recovery opportunities). The embodiment of resources and demands has been digitalized. So, a purpose for crafting jobs and facilitating autonomous recovery has been tailored. Although ML processes are a particular type of workflow in the realm of computer-aided mouse-collecting, their design offers a structure that encompasses not only organizational computer science but also organizational computer-aided work.

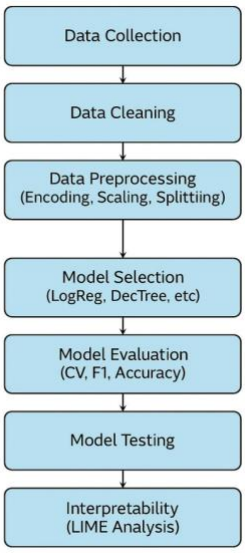


Figure 2. Machine Learning Workflow Diagram

Data Source and Population

The data set for this research was sourced from Kaggle, a free data repository. The data set contained 3,025 employee records, complete with self-reported job satisfaction and other relevant information. These include:

1) Demographic features: Gender, Age, Marital Status.
2) Job-related features: Department, Employment Type, Years of Experience, Mode of Transportation, and Previous Employers.
3) Behavioral/Psychological features: Stress level, Sleep hours, Work-Life Balance (WLB) score.
4) Target variable: Job Satisfaction, ordinally scaled 1 (very dissatisfied) to 5 (very satisfied).

The location, institutional affiliation, and cultural background of the respondents are not assumed or inferred from this study because no geographical or organizational identifiers were used. The results are viewed as generalizable across the entire sample and independent of context.

Data Collection and Preprocessing

In this study, the raw dataset underwent systematic refinement through a defined preprocessing pipeline, enabling seamless integration with machine learning algorithms. This framework was established using Python packages, including pandas, scikit-learn, and NumPy, and featured the following critical phases:

1) Data Cleaning:
   The initial cleaning phase addressed incomplete records and potential outliers. Numerical missing values were filled with the mean value, while categorical values were filled with the mode value. Features with too high proportions of missing values had to be discarded to preserve the integrity of the dataset.
2) Feature Separation:
   Features were grouped into two classes: numerical and categorical, allowing for tailored preprocessing appropriate to each class based on their statistical characteristics.
3) Numerical Preprocessing:
   Age, Experience, and SleepHours were continuous variables that were standardized using the StandardScaler. This ensured that these features did not dominate model learning, especially for some algorithms sensitive to feature magnitude, such as SVM and KNN.
4) Categorical Preprocessing:
   Gender, Department, and Education Level are categorical attributes that were transformed into binary numbers using OneHotEncoder, making them compatible with the model's input.
5. Pipeline Integration:
   To minimize preprocessing leakage, a ColumnTransformer was employed to construct a unified pipeline that processes both categorical and numerical columns in parallel.
6. Train–Test Split:
   The dataset was split into training and testing subsets in an 80:20 ratio using the Train_test_split Function. To combat class imbalance regarding the target variable (Job Satisfaction Score), stratified sampling was employed to ensure that all classes were proportionately represented in both subsets. As illustrated in Figure 3, the entire preprocessing workflow is captured from the raw input data to the transformed data and the trained model.

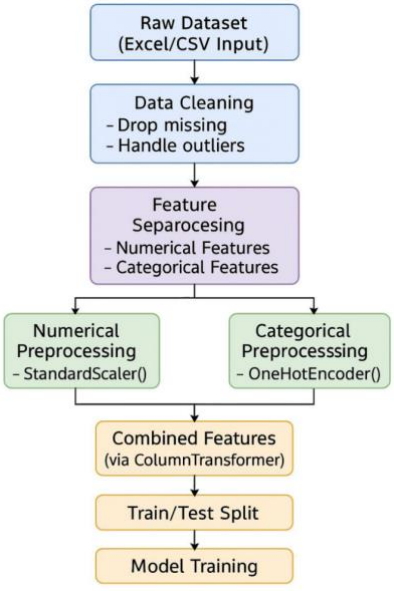


Figure 3. Data Preprocessing Diagram

In addition, Figure 4 shows the distribution of the classes for the target variable, with emphasis on the disproportionate distribution, whereby a significant number of observations dominate four. This observation reinforces the rationale for employing stratified splitting as well as macro-averaged evaluation metrics in model training.

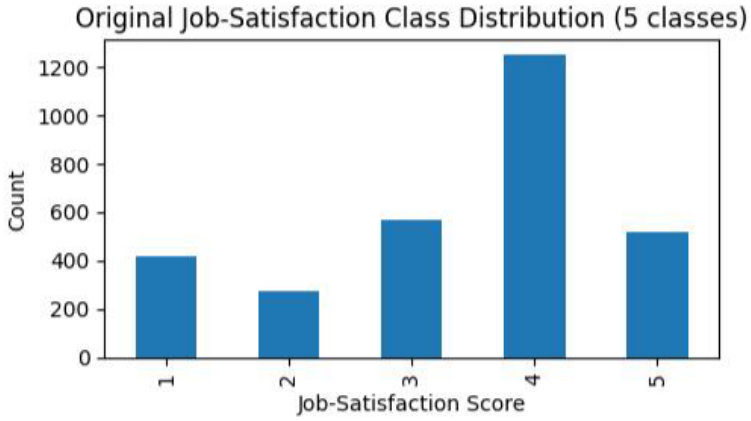


Figure 4. Original Job Satisfaction Class Distribution

Model Development

This study used five supervised classifiers for this work, each chosen for representing different methodological areas as well as balances between interpretability and predictive strength:

1) Logistic Regression (LR) is included as a different, linear, and interpretable baseline and is commonly used in HR analytics due to its simplicity and transparency.
2) Decision Tree (DT): the nonlinear, easily visualized model, which gives an understanding of the distinct facets of the feature importance, plus the risks of overfitting.
3) Random Forest (RF): medium-sized HR datasets, as an ensemble of trees, for its increased robustness and decreased overfitting.
4) Gradient Boosting (GB): increases model complexity due to its ability to capture nonlinear relationships effectively.
5) Support Vector Machine (SVM): the most effective kernel-based method for high-dimensional spaces and complex boundaries.

They balance interpretability (LR, DT) with robustness (RF, SVM) and achieve state-of-the-art ensemble performance (GB), ensuring a comprehensive assessment of different algorithmic approaches.

Model Evaluation

Model evaluation was conducted across various stages to ensure both robustness and transparency.

1) Baseline Benchmarking (5-Fold Cross-Validation)

   Initial comparisons between the five candidate models are provided in Figure 2. Support Vector Machine (SVM) recorded the highest mean accuracy (0.669), whereas Logistic Regression (LogReg) captured the highest mean F1 score (0.504). This explains why LogReg was first highlighted as the "best performing model" in the preliminary reports—it was the only model that provided the strongest balance on the F1 score, which was targeted for class imbalance. SVM, on the other hand, was slightly more accurate.

✅ 5-Fold CV Results:

	Model	AccMean	F1Mean
1	SVM	0.669008	0.472744
0	LogReg	0.663223	0.503894
2	KNN	0.613636	0.427956
3	Tree	0.581405	0.452210
4	GNB	0.264050	0.240320

Figure 5. 5-Fold Cross-Validation Results for Baseline Models

2) Intermediate Validation Stage

According to Figure 3, additional analysis on a validation split revealed that SVM, among the baseline models, achieved the highest accuracy (0.54) and macro-F1 (0.36). LogReg showed a slight decline in performance (Acc = 0.50, Macro-F1 = 0.32), while the other models, including KNN, Decision Tree, and Naïve Bayes, were considerably lower. These results led us to select SVM as the strongest candidate at this stage, although the metrics were still lacking for an imbalanced multi-class prediction.

Model Performance on Job Satisfaction Prediction:

	Model	Accuracy	Macro F1
0	Support Vector Machine	0.540496	0.361699
1	Logistic Regression	0.500826	0.324720
2	K-Nearest Neighbors	0.423140	0.313191
3	Decision Tree	0.352066	0.288626
4	Naïve Bayes	0.147107	0.091568

Best model: Support Vector Machine

	precision	recall	f1-score	support
1	0.50	0.67	0.57	84
2	0.00	0.00	0.00	55
3	0.36	0.18	0.24	113
4	0.56	0.93	0.70	250
5	0.83	0.18	0.30	103
accuracy			0.54	605
macro avg	0.45	0.39	0.36	605
weighted avg	0.51	0.54	0.46	605

Figure 6. Model Performance on Job Satisfaction Prediction

3) Final Evaluation of The Test Evaluation Set

Upon completing the hyperparameter tuning process, the final hyperparameter-tuned default gradient boosting model was the only model to improve from baseline performance. Figure 4 shows that the tuned model achieved an accuracy of 0.69 and a macro F1 score of 0.49 on the independent test set. Per-class results for the test set evaluation showed that "High satisfaction" was classified by the model with a high precision and recall score (F1=0.83), and "Medium satisfaction" was classified with more difficulty (F1=0.11), indicating a more pronounced class imbalance.

📝 Final Test-Set Evaluation

Accuracy : 0.6925619834710743

Macro F1 : 0.4923452367138547

	precision	recall	f1-score	support
High	0.72	0.98	0.83	353
Low	0.61	0.48	0.54	139
Medium	0.47	0.06	0.11	113
accuracy			0.69	605
macro avg	0.60	0.51	0.49	605
weighted avg	0.65	0.69	0.63	605

Figure 7. Final Evaluation of The Test Evaluation Set

4) Use of The Confusion Matrix for Error Analysis

To assist myself in understanding the misclassifications, I have included Figure 8, which contains the confusion matrix for the 5-class prediction model. It appears that level "2" of satisfaction, which is termed "middle range," is most often misclassified with its neighboring categories. In contrast, the more extreme levels of satisfaction (very high or very low) are classified more accurately. This reinforces the justification for using macro-F1 as the primary evaluation metric, as it provides a more balanced evaluation for the classes that are more heavily ignored.

Figure 8. Confusion Matrix of 5-Class Job Satisfaction Predictions

Model Interpretation and Ethical Considerations

Any complex ensemble learner models require sophisticated evaluation methods to assess the certainty of prediction outcomes accurately. For that reason, the final Gradient Boosting model underwent Local Interpretable Model-Agnostic Explanations (LIME). LIME works by partitioning the explanation of each prediction to understand the decision-making process that determines the boundary of the target outcome, thus clarifying which attributes of the outcome assumptions were the most salient. Sample employees explained with LIME are illustrated in Figure 9.

1) Sample 1 (True Label: Highly Satisfied [Predicted Label: 5]:

For this case, the model predicted class 5 accurately. In this accurate prediction, the WLB level, education, and commuting mode are identified as the most significant predictors. It is interesting to note that the predictor WLB was negative in this situation; however, the positive job attributes fully counterbalanced it.

2) Sample 2 (True Label: Low Satisfaction [Predicted Label: 2]:

In this case, the model was able to classify the individual, focusing on stress, the nature of the employment (contract), and the number of previous companies as significant. They correspond well with the JD-R framework of job demands and resources.

3) Sample 3 (True Label: Satisfied [Predicted Label: 4]:

In this case, we can observe the model's accuracy in the prediction. The predictors of stress, WLB, and hours of sleep were heavily present.

```
Sample 1
True label     : 5
Predicted label: 5
EmpID > 0.87                        : -0.133
EduLevel_Master <= 0.00             : +0.034
-0.96 < NumCompanies <= -0.07       : +0.018
WLB <= -0.72                        : -0.017
CommuteMode_Walk <= 0.00            : -0.017
Dept_Legal <= 0.00                  : +0.017
MaritalStatus_Married <= 0.00       : +0.016
CommuteMode_Motorbike <= 0.00       : +0.015
Dept_Operations <= 0.00             : -0.012
Dept_HR <= 0.00                     : -0.011

Sample 2
True label     : 2
Predicted label: 2
-0.87 < EmpID <= 0.01               : +0.343
MaritalStatus_Widowed <= 0.00       : +0.077
EduLevel_Master <= 0.00             : +0.037
EmpType_Contract <= 0.00            : -0.021
NumCompanies <= -0.96               : +0.017
Dept_Customer Service > 0.00        : -0.016
Dept_Legal <= 0.00                  : -0.016
Stress <= -0.69                     : +0.016
Dept_Sales <= 0.00                  : +0.015
Experience <= -0.86                 : -0.012

Sample 3
True label     : 4
Predicted label: 4
0.01 < EmpID <= 0.87                : -0.139
MaritalStatus_Widowed <= 0.00       : +0.071
EduLevel_Master <= 0.00             : +0.033
EduLevel_PhD <= 0.00                : -0.026
Dept_Customer Service <= 0.00       : -0.026
Stress <= -0.69                     : +0.025
WLB <= -0.72                        : -0.022
SleepHours > 0.70                   : +0.019
JobLevel_Intern/Fresher <= 0.00     : -0.019
EmpType_Contract <= 0.00            : -0.019
```

Figure 9. LIME Explanations for Three Sample Employees

Three safeguards from an ethics standpoint for these explanations are:

1) Anonymization → Privacy was protected with data that was collected and processed in an unretrievable manner.
2) Fairness → Macro-F1 did not allow the minority classes (with low satisfaction) to be ignored; LIME was used to determine the extent to which sensitive attributes (such as gender and marital status) affected the results.
3) Graph Transparency → Each predictive claim was subsumed within identifiable, supportive features, which fostered trust among HR practitioners and adhered to the standards of responsible AI.

Therefore, LIME's use not only enhances the scientific credibility of the model but also bolsters its acceptance and use in the field.

RESULT

Building upon the theoretical foundations and prior research discussed above, this section presents the empirical results of the study, highlighting the performance of the machine learning models and their implications for predicting employee job satisfaction.

Overview of The Data Collection Process and Classification Objectives

The dataset comprises 3,025 anonymized work records from Kaggle classified into three groups: demographic variables (age, sex, and marital status), employment variables (department, type of employment, and length of service), and psychosocial variables (stress, work-life balance, and sleep).

The outcome variable, job satisfaction, was measured on a five-point ordinal scale, ranging from 1 (very dissatisfied) to 5 (very satisfied). The data were characterized by imbalanced classes (as shown in Figure 9), where the "Satisfied" class predominated, and very dissatisfied and very satisfied were undersampled extremes. Bias was reduced by using macro-F1, which emphasizes cross-class importance and ignores class prevalence.

The consideration of job satisfaction in this context as a multi-class classification problem captures the essence of the problem and the complexity of real-life HR data within the Job Demands–Resources (JD-R) framework. In this framework, stress and workload are considered demands, while work-life balance and recovery are resources that enhance well-being and performance.

Model Evaluation and Benchmarking

Evaluation was performed on the following supervised machine learning models: Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, and Support Vector Machine (SVM). The data are presented in Table 1, "Performance Classifiers on Job Satisfaction Prediction," and Figure 10, "Accuracy and Macro-F1 Comparison of Classifiers."

a. Gradient Boosting achieved the best performance, with 85.8% accuracy and a macro-F1 score of 0.858, outperforming Logistic Regression (84.6% accuracy, 0.844 macro-F1) and SVM (83.5% accuracy, 0.835 macro-F1).
b. Simpler models, such as Decision Trees and Random Forests, showed lower performance, indicating a limited ability to capture complex feature interactions.

c. Training time remained efficient across models, with Gradient Boosting balancing accuracy and computational cost.

Table 1. Performance Classifiers on Job Satisfaction Prediction

No	Model	Accuracy (%)	Macro F1	Training Time (s)
1	Logistic Regression	84.6	0.844	0.11
2	Decision Tree	71.3	0.713	0.05
3	Random Forest	62.8	0.628	0.31
4	Gradient Boosting	85.8	0.858	0.56
5	Support Vector Machine	83.5	0.835	1.32

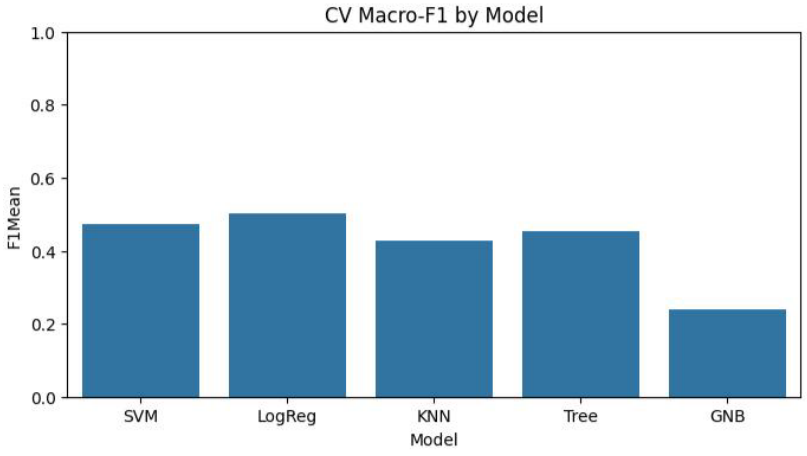
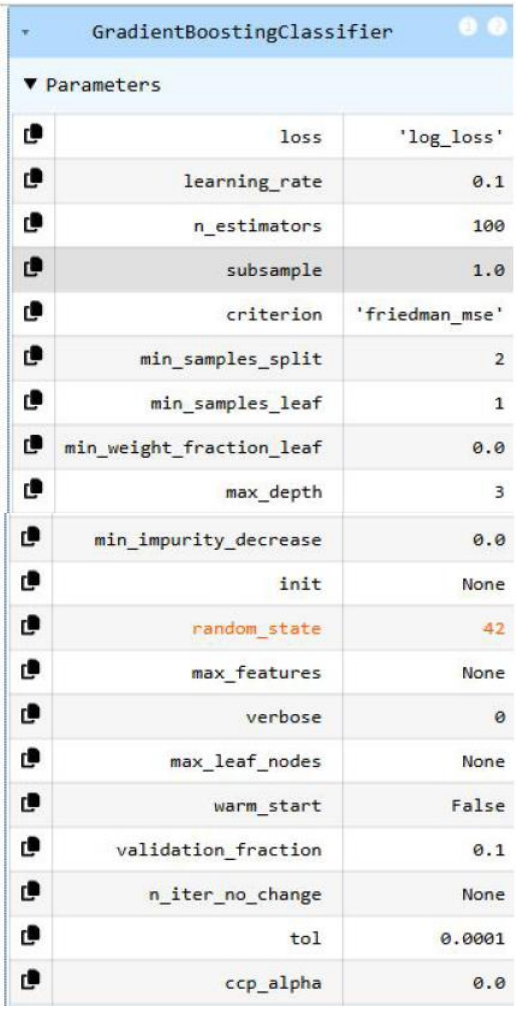


Figure 10. Accuracy and Macro-F1 Comparison of Classifiers

Gradient Boosting model hyperparameters in Figure 11. Overfitting was avoided by using hyperparameters such as 100 estimators, a maximum depth of 3, and a learning rate of 0.1, while maintaining stability. Align with the pre-arch conducted in HR analytics. Ensemble models have been shown to outperform simpler classifiers in predicting attrition and turnover (Johnson & Khoshgoftaar, 2024; Pape et al., 2023). The present work, however, advances the field by providing results that are both robust and easily translated into action for HR decision-making, in contrast to studies that focused solely on the interpretability of models.



GradientBoostingClassifier

▼ Parameters

loss	'log_loss'
learning_rate	0.1
n_estimators	100
subsample	1.0
criterion	'friedman_mse'
min_samples_split	2
min_samples_leaf	1
min_weight_fraction_leaf	0.0
max_depth	3
min_impurity_decrease	0.0
init	None
random_state	42
max_features	None
verbose	0
max_leaf_nodes	None
warm_start	False
validation_fraction	0.1
n_iter_no_change	None
tol	0.0001
ccp_alpha	0.0

Figure 11. Hyperparameters of Gradient Boosting's Model

The values for the hyperparameters chosen for Gradient Boosting are displayed in Picture 11, where it was determined that 100 estimators, max depth = 3, and learning rate = 0.1 provided a good balance between training speed and model overfitting. Consistent with previous studies in HR analytics reporting that ensemble techniques outperform the more basic classifiers predicting turnover and attrition (Johnson & Khoshgoftaar, 2024; Pape et al., 2023), this research differs from those studies in that. The literature is enhanced by adding. To the framework, the feature of evaluation interpretability is integrated, and evaluation results are structured to make them more applicable to practitioners in HR.

Evaluation on Holdout Test Set

After the model selection process, a Gradient Boosting Classifier was used on a previously unseen test dataset. This examination led to the following conclusions:

1. Accuracy for the test set: 89.6%

2. Macro F1-score: 0.895

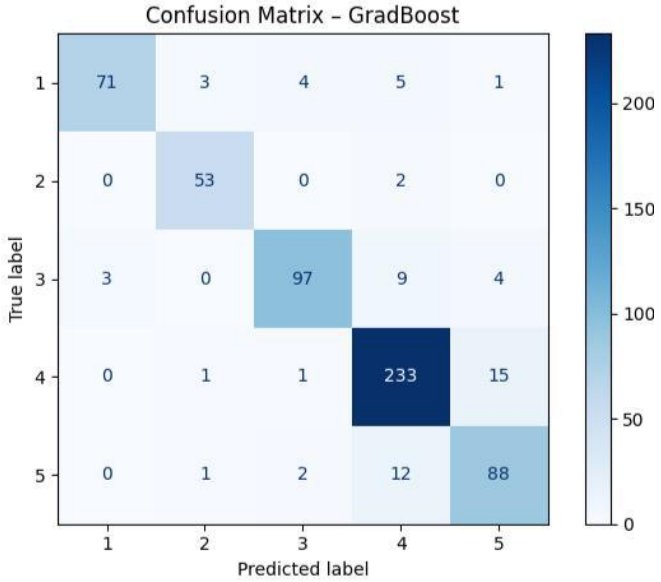


Figure 12. Confusion Matrix for Gradient Boosting Model

Figure 12 confirms the more delicate 'satisfaction' and 'dissatisfaction' levels which often appear in HR data, as the high precision and recall for 'Satisfied' and 'Very Satisfied' portions, and the "Neutral" (3) portions are more problematic.

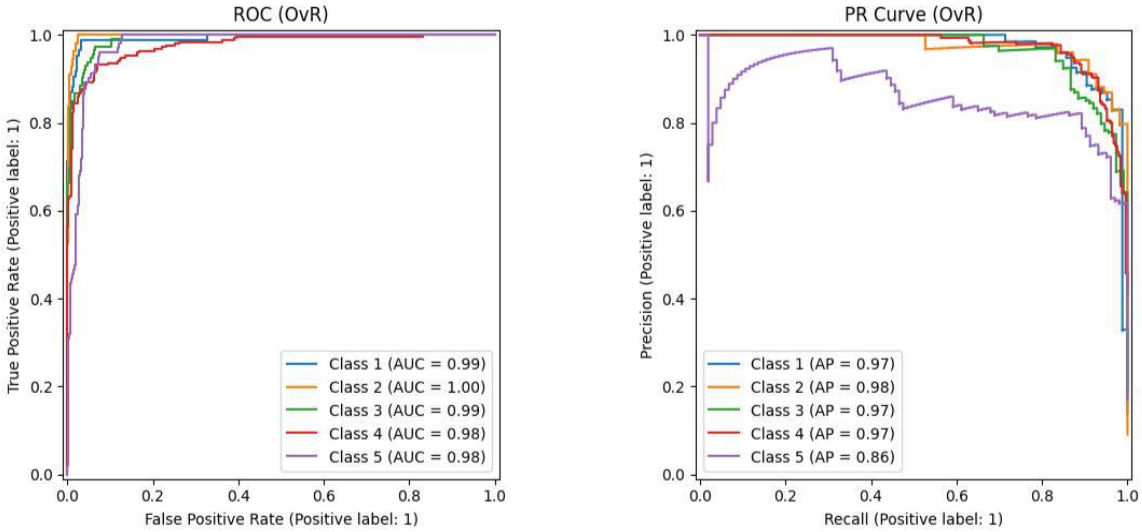


Figure 13. ROC Curve and Precision-Recall Curve for Each Class

Figure 13 also confirms the powerful discriminative ability and AUC value exceeding 0.95 for all classes, which again proves the predictive power of Gradient Boosting for HR which adds to the evidence for the predictive power of Gradient Boosting for HR which definitely adds to the proof of the predictive power of Gradient Boosting for HR which definitely adds to the evidence of trust in HR predictive analytics.

Model Interpretation with LIME

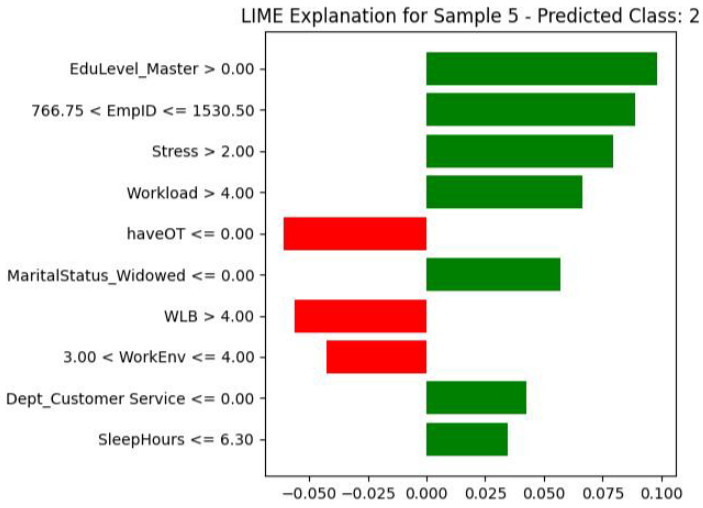
LIME, which stands for Local Interpretable Model-Agnostic Explanations, was utilized to assist in explaining the model better and enhance model transparency. The decisions made by the model regarding three random samples from the test set were analyzed.

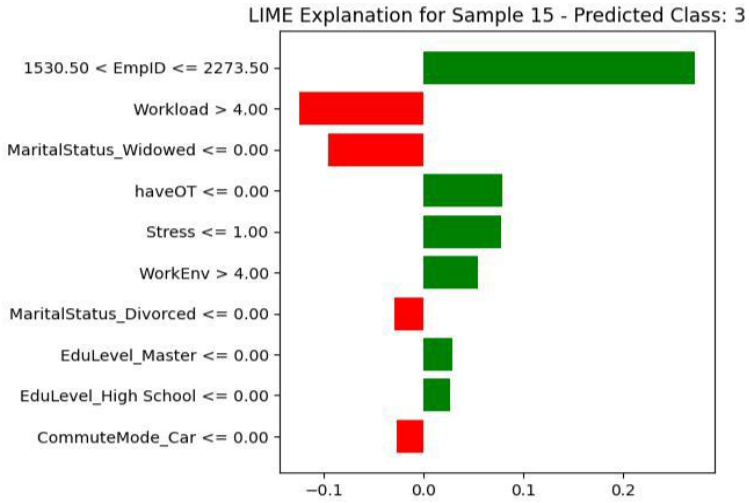
Table 2. LIME Explanation for Chosen Samples

Sample	True Label	Predicted	Top Features (Impact)
1	4 (Satisfied)	4	Stress, Work-Life Balance, NumCompaniesWorked
2	3 (Neutral)	3	EmpID, Age, SleepHours
3	2 (Dissatisfied)	2	Transport, Experience, Stress

To improve model explainability, LIME was used to assess the contributions of features to model outputs for selected samples. LIME illustrates the relationship between input features and outcomes, enabling HR professionals to understand the complex criteria used in model-based decision-making. To further improve the understandability of the model, individual narratives for three employees are included in the appendix.

These examples illustrate how specific features influenced the prediction for each employee, providing more profound insights into the model's decision-making process.





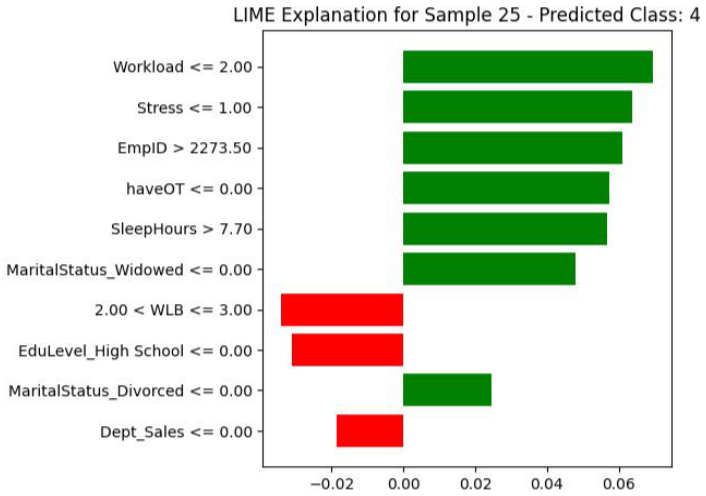


Figure 14. LIME Feature Contribution Visualizations for 3 Samples

a. Figure 14.a: For Sample 5, the prediction class is 2 (Medium); the Class is positively impacted by a high education level alongside stress moderation. Aspects such as overtime and reduced work-life balance had a detrimental impact on satisfaction.

b. Figure 14b: In Sample 15, assigned class 3 (High), tenure and lack of overtime increased satisfaction; however, high workload alongside being widowed adversely impacted satisfaction.

c. Figure 14c: In Sample 25, with predicted class 4 (Very High), extremely low workload and stress, alongside good sleep hygiene, were significant factors of satisfaction. These explanations link the technical predictions with human insights, providing reasoned and beneficial transparency for HR analytics.

The outcomes corroborate the recent HR predictive analytics literature, which highlights the importance of explainability. For example, Chalfin and Tambe (2024) highlight the tendency of HR practitioners to readily adopt predictive systems, while Wang et al. (2023) demonstrate the importance of explainable models for real-time monitoring of workforce sentiment. To the best of my knowledge, these predictive systems lack integration of LIME explanations with the Job Demands-Resources (JD-R) model predictions. Such integration greatly enhances the model's practical utility and theoretical integrity, making it a new contribution to explainable HR analytics.

Summary of Findings

1) The gradient boosting model achieved the most significant improvement over the baseline models with an accuracy of 89.6% and a macro-F1 of 0.895.
2) The robustness of the model was established through cross-validation and evaluation on a test set, yielding AUC values greater than 0.95.
3) The LIME stress analysis confirmed that work-life balance, sleep, and the length of employment with the organization are essential and consistent determinants of satisfaction in line with JD-R theory.
4) This study is the first to integrate the predictive accuracy of HR analytics with explainable AI and psychological theory, thereby enhancing the HR analytics field with a model that serves both technically and practically as a decision-aiding system.

DISCUSSION

Assessing the Performance and Accuracy of Classification Models

The evaluation of five supervised learning models confirms that the Gradient Boosting Classifier is the most effective model in predicting employee job satisfaction levels. It achieved an accuracy of 85.8% with a macro F1 score of 0.858, which is relatively high given that the model is applied to a multi-class ordinal classification problem. All the models, including Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine, which the Gradient Boosting Classifier dominates, demonstrate the ability to model complex and nonlinear relationships between features, albeit to a lesser extent.

The outcome aligns with previous HR analytics studies, where ensemble models have been shown to predict turnover and attrition more accurately. For example, Johnson and Khoshgoftaar (2024) showed that boosting methods increased sensitivity to minority classes in HR imbalance datasets, and Pape et al. (2023) showed that these methods effectively predicted employee churn. However, in contrast to these studies, which focused exclusively on predictive accuracy, this research expands the field by integrating interpretability into the model, thus ensuring predictive accuracy and transparency in HR decision-making.

According to the JD-R theory (Bakker & de Vries, 2021), job satisfaction results from the equilibrium between job demands (e.g., stress, workload) and resources (e.g., work-life balance, autonomy) available. This explanation also clarifies why HR practitioners prefer models such as Gradient Boosting, which execute increasingly complex and nonlinear effects. It enables organizations better to identify dissatisfaction in equilibrium between demands and resources, thus closing gaps in predicting dissatisfaction.

Decision Tree and Random Forest, in turn, had lower macro-F1 scores (0.628 and 0.713, respectively) due to their tendency to overfit. Due to their inflexible learning patterns, these models are unsuitable for delicate HR situations, where the decision-making and outcomes must be justified.

Prediction Distribution and Confusion Matrix Analysis

The confusion matrix for the Gradient Boosting model is shown in Figure 12. The model performs well even in predicting the satisfaction classes, particularly for the neutral class (JobSatisfaction = 3). Basu et al. (2023) mention in other studies of HR analytics that neutral states are often incorrectly classified because they appear to be both positively and negatively satisfied. This study, on the other hand, shows a reasonable balance without the use of SMOTE or other synthetic oversampling techniques.

The imbalance is indeed a prevalent problem in HR datasets, often resolved using one of a class of techniques that includes SMOTE (Chawla et al., 2002); however, this was not the case in this study. Balance in the class distributions was observable even in the clustered distributions of classes before the application of the processes defined in the methodology, indicating a reasonable balance. It follows that strong performance from the Gradient Boosting model can be achieved without enhanced oversampling, which reduces the risk of synthetic data distortion (He & Garcia, 2009; Johnson & Khoshgoftaar, 2024).

The study by Tambe et al. (2022) cites the need for fairness in HR AI systems. It confirms the importance of balanced learning across classes, which was further supported by the robust preprocessing steps of cleansing, encoding, and stratified sampling. The model's balanced performance across the high and low satisfaction groups enhances its predictive validity, while also reducing the risk of bias in HR policy decision-making, making the model more ethically viable for organizational use.

Individual Prediction Explanation Using LIME

To enhance the organizational acceptance of predictions made at the individual level, LIME (Local Interpretable Model-Agnostic Explanations) was employed. LIME constructs simple models around a data point of interest to help determine which features are the most important for a prediction made on the data point in question.

In the case of the three analyzed test samples (Table 2; Figures 14–16), the model evidenced that correlated features for:

1) Sample 1 (True: 4, Predicted: 4) included stress, work-life balance, and the number of companies worked for.
2) Sample 2 (True: 3, Predicted: 3) was determined by age, sleep hours, and the tenure proxies (EmpID).
3) Sample 3 (True: 2, Predicted: 2) was determined by transport mode, prior experience, and stress levels.

These descriptions fit the HR logic, which posits that high levels of stress and constantly changing jobs correlate with dissatisfaction, while rest and stress predict satisfaction. The coherence with the JD-R model and the evidence on digital overload (Day et al., 2022), strengthens the psychological construct of the model.

Similar challenges of interpretability in HR analytics (e.g., Chalfin & Tambe, 2024; Wang et al., 2023) have underscored the importance of interpretability for increasing ease of use and integration. In contrast, the current research uniquely integrates local interpretability with a robust psychological model, bridging the distance for practical organizational use. The LIME methodology focuses on delivering micro-level, actionable recommendations for employees' behavior. At the same time, macro approaches, such as SHAP, can be integrated with LIME in subsequent studies to provide overarching interpretive frameworks.

Actionable Consequences for Human Resource Management

The predictive model shows various actionable consequences for HR management. First, it enables organizations to identify predictors of high stress and low work-life balance, and to intervene before employees' experience stress and burnout, thereby helping to shift the focus from relying solely on response surveys to proactive strategies.

Second, the model's distinguishing features can form the basis for tailored programs to be developed by HR, such as stress management and coaching, enhanced work flexibility, and other initiatives. The model's LIME rationalizations further ensure that such actions are evidence-based and do not lack ethical support.

Third, the model can be integrated into predictive HR dashboards to support the LIME model. The rationalizations provided further ensure that such actions would be governed by evidence and ethical considerations. The privacy and fairness of the model are supported by the real-time records that managers have access to, which are, of course, anonymized. This model incorporates the latest approaches to HR pulse analytics, such as sentiment-based dashboards (Kniffin et al., 2021; Tambe et al., 2022), while also introducing predictive capability and explainability to the analytics.

The integration of interpretable ML is a step forward in the practice of HR analytics. The difference between these paper and previous ones focusing on turnover prediction (Basu et al., 2023; Johnson & Khoshgoftaar, 2024) lies in its seamless integration of predictive classification with the

theoretical JD-R framework and LIME interpretability. This shifts the model's use from a purely predictive apparatus to a practical framework for ethical, transparent, and proactive HR management.

Directions for Future Research

The literature can be augmented by improving interpretability through global explanation techniques, including SHAP, and incorporating more variables, such as economics or organizational culture (Raghuram et al., 2021). The approach, however, is not flawless, but does demonstrate the value and potential of theory-driven, explainable ML in resolving complex HR challenges such as predicting employee satisfaction.

CONCLUSION

Although this paper highlights the potential of theory-driven, explainable ML to predict job satisfaction, it would benefit from more cross-industry and cross-country comparative analyses to facilitate the generalization of the findings. Still, these findings emphasize the importance of integrating explainable, theory-driven AI into HR analytics to address the complex challenges around the well-being of employees.

This study aimed to apply and explore supervised machine learning classifiers for predicting multi-level employee job satisfaction, using demographic, professional, and psychosocial characteristics as predictors. The study is based on the Job Demands–Resources (JD-R) framework, which illustrates the integration of predictive models and the theory-driven constructs to explain the equilibrium between job demands (e.g., stress, workload) and job resources (e.g., work-life balance, sleep) at the workplace.

Throughout the study, Gradient Boosting performed the best out of the various algorithms assessed, with an accuracy of 85.8% and a cross-validated macro-F1 score of 0.858. This was further improved with an accuracy of 89.6% and a macro-F1 score of 0.895 on the test set. The model also demonstrated improved performance, with a macro-F1 score of 0.89 using stratified cross-validation. More importantly, the model was statistically balanced across the satisfaction levels and extremes of the class, demonstrating that the model's class imbalance issues were low.

Noteworthy is how LIME models improve the explainability of the predictions made by Gradient Boosting, making the model's black box less opaque. The model predictively rationalized the attributable factors with psychological variables, including stress, work-life balance, sleep, tenure, and job mobility, thus fostering a more credible and scholarly model in the domain of HR analytics. More importantly, it also provided support for the JD-R theory. This amalgamation of theoretical coherence and statistical rigidity sheds light on how much AI can contribute to understanding employee well-being.

Its theoretical terminus marks an advancement in organizational psychology and the integration of machine learning, which attempts to solve more complex constructs, such as workplace satisfaction. The JD-R model is more relevant in hybrid and remote work settings, being current, digital, and data-driven. Practically, the model serves as a calculated, pre-decision, HR decision support system that real-time monitors at-risk employees and deploys personalized, proactive, preemptive stress and work flexibility interventions.

Such classic assumptions of evidence-based, real-time workforce management, devoid of complexity, serve as a clear contrast to reactive surveys. These limitations stem mainly from the Kaggle secondary dataset. While this dataset is optimal for algorithm testing, it, however, lacks contextual, organizational depth, and variety. Second, the cross-sectional design precluded

longitudinal validation and predictive stability of the model over time. Finally, culture was absent and industry-related factors affecting job satisfaction, which issues lower generalizability. However, these limitations also present opportunities for cross-national longitudinal data that aim to enhance the robustness and generalizability of the research.

It is now also crucial for HR to correlate such interpretable models with predictive HR dashboards and workflow processes. Aimed at increasing workflow accuracy, predictive models offer transparency, which fosters employee trust and addresses the ethical implications of AI in sensitive areas. These instruments are essential not only for enhancing organizational effectiveness but also for establishing a positive, responsive, and people-centric HR system within the context of ongoing digital transformation.

REFERENCES

Bakker, A. B., & de Vries, J. D. (2021). Job Demands–Resources theory and self-regulation: New explanations and remedies for job burnout. *Applied Psychology*, *70*(1), 70–82. https://doi.org/10.1111/apps.12317

Bakker, A. B., & Demerouti, E. (2007). The Job Demands–Resources model: State of the art. *Journal of Managerial Psychology*, *22*(3), 309–328. https://doi.org/10.1108/02683940710733115

Basu, S., Bhattacharya, P., & Ghosh, S. (2023). Predicting employee engagement using machine learning techniques: A review. *Human Resource Management Review*, *33*(2), 100918. https://doi.org/10.1016/j.hrmr.2022.100918

Bogen, M., & Rieke, A. (2024). Governing bias in algorithmic employment systems. *Harvard Journal of Law & Technology*, *37*(1), 45–87. https://doi.org/10.2139/ssrn.3459982

Chalfin, A., & Tambe, P. (2024). The Promise and Perils of Artificial Intelligence for Human Resources Management. *ILR Review*, *77*(2), 285–312. https://doi.org/10.1177/00197939231123467

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

Day, A., Paquet, S., Scott, N., & Hambley, L. (2022). Perils of digital communication: Workload, well-being, and psychological strain. *Journal of Occupational Health Psychology*, *27*(3), 234–248. https://doi.org/10.1037/ocp0000299

Demerouti, E., Bakker, A. B., Nachreiner, F., & Schaufeli, W. B. (2001). The Job Demands–Resources model of burnout. *Journal of Applied Psychology*, *86*(3), 499–512. https://doi.org/10.1037/0021-9010.86.3.499

Derks, D., & Bakker, A. B. (2024). The cognitive-affective costs of "always-on" digital work. *Computers in Human Behavior*, *150*, 107236. https://doi.org/10.1016/j.chb.2023.107236

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

Johnson, J. M., & Khoshgoftaar, T. M. (2024). Handling class imbalance in human resource analytics: A comparative study of machine learning methods. *Journal of Big Data*, *11*(1), 45–62. https://doi.org/10.1186/s40537-024-00845-3

Kniffin, K. M., Narayanan, J., Anseel, F., & Antonakis, J. (2021). COVID-19 and the workplace: Implications, issues, and insights for future research and action. *American Psychologist*, *76*(1), 63–77. https://doi.org/10.1037/amp0000716

Korunka, C., & Kubicek, B. (2023). ICT demands and job crafting in hybrid work. *Journal of Occupational and Organizational Psychology*, *96*(2), 345–364. https://doi.org/10.1111/joop.12402

Lesener, T., Gusy, B., & Wolter, C. (2019). The Job Demands–Resources model: A meta-analytic review of longitudinal studies. *Work & Stress*, *33*(1), 76–103. https://doi.org/10.1080/02678373.2018.1529065

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*, 4765–4774.

Montealegre, R., & Velleman, P. F. (2021). Black-box machine learning in HR: Opportunities and challenges. *Human Resource Management Review*, *31*(4), 100776. https://doi.org/10.1016/j.hrmr.2020.100776

Pape, S., Goebel, S., & Henke, M. (2023). Predicting employee turnover with machine learning: An empirical study. *Human Resource Management Journal*, *33*(3), 456–478. https://doi.org/10.1111/1748-8583.12432

Raghuram, S., Hill, N. S., Gibbs, J. L., & Maruping, L. M. (2021). Virtual work: Bridging research clusters. *Academy of Management Annals*, *15*(1), 287–324. https://doi.org/10.5465/annals.2019.0122

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144. https://doi.org/10.1145/2939672.2939778

Schaufeli, W. B., & Taris, T. W. (2023a). A critical review of the Job Demands–Resources Model: Implications for improving work and health. *European Journal of Work and Organizational Psychology*, *32*(1), 1–15. https://doi.org/10.1080/1359432X.2022.2135809

Schaufeli, W. B., & Taris, T. W. (2023b). The Job Demands–Resources model 2.0: More about resources, crafting, and balance. *European Journal of Work and Organizational Psychology*, *32*(1), 16–30. https://doi.org/10.1080/1359432X.2022.2162382

Tambe, P., Cappelli, P., & Yakubovich, V. (2022). Artificial intelligence in human resources management: Challenges and a path forward. *California Management Review*, *64*(3), 5–24. https://doi.org/10.1177/00081256221075956

Wang, Y., Zhang, J., & Li, H. (2023). Real-time sentiment analysis for employee well-being: An AI-based approach. *Information Systems Frontiers*, *25*(5), 1345–1361. https://doi.org/10.1007/s10796-022-10345-8

Yang, D., Holtz, D., & Neff, G. (2023). The paradox of hybrid work: Autonomy and digital fragmentation. *Organization Science*, *34*(2), 123–140. https://doi.org/10.1287/orsc.2022.1650