Enhancing Organizational Information Systems Through Explainable Artificial Intelligence

Kian Jazayeri
Department of Management Information Systems, Cyprus International University, Nicosia, Türkiye

***Abstract*—This study examines Workplace Perceptions among Finnish employees through the application of Artificial Intelligence within the domain of Human Resource Analytics. An integrated analytical framework combining Clustering Analysis, supervised classification, and Explainable Artificial Intelligence is proposed to uncover and interpret latent employee perception profiles. Using 23 perception-related indicators from the Finnish Working Life Barometer 2022, K-means clustering identified two distinct employee groups: one characterized by consistently positive evaluations of fairness, leadership, well-being, and motivation, and another reflecting systematically negative workplace perceptions. A LightGBM model was subsequently employed to predict cluster membership based on demographic and occupational variables, and SHapley Additive exPlanations (SHAP) were used to provide transparent global and local interpretations of the predictive outcomes. The results show that employment duration, age, industry affiliation, gender, and socioeconomic status are the most influential determinants of cluster membership. By embedding Explainable Artificial Intelligence into Human Resource Analytics, the study demonstrates how employee perception data can be transformed into interpretable knowledge that supports organizational Decision-Support Systems. The proposed framework advances data-driven and transparent HR decision-making and contributes to the United Nations Sustainable Development Goal 8, Decent Work and Economic Growth, by identifying structural disparities in employee experience and enabling more equitable and inclusive workplace interventions.**

Keywords*—*Artificial intelligence; Human Resource Analytics; Explainable Artificial Intelligence; Decision-Support Systems; workplace perceptions; Clustering Analysis; Decent Work and Economic Growth

I. INTRODUCTION

Understanding employee perceptions has long been recognized as a cornerstone of effective organizational development [1] [2]. Employees' views on fairness, leadership, inclusivity, and well-being are closely tied to job satisfaction, productivity, and retention, factors that directly influence organizational resilience and long-term growth [3]. From an information systems perspective, workplace perception data represents a valuable informational asset that, if effectively processed and interpreted, can guide both organizational policies and national development strategies. As workplaces evolve under pressures from globalization, digital transformation, and post-pandemic shifts in work arrangements, the need for robust, data-driven approaches to capture and interpret these perceptions has become increasingly urgent [4][5].

Traditional HR analytics methods, often based on descriptive statistics or simple segmentation, risk overlooking latent patterns within employee populations. Subtle but meaningful differences in workplace perceptions, shaped by combinations of demographic and occupational factors, may remain hidden when analyses are restricted to surface-level groupings. Without uncovering these underlying profiles, organizational information systems risk generating one-size-fits-all policies that fail to meet the diverse needs of employees.

The existing literature demonstrates two complementary but disconnected streams. Research on clustering in HR analytics shows the potential to segment employees into coherent groups based on survey responses or performance indicators. Meanwhile, Explainable Artificial Intelligence (XAI), particularly SHapley Additive exPlanations (SHAP), has proven effective in interpreting predictive models for HR outcomes such as attrition or performance. However, clustering studies typically stop at descriptive profiling without quantitatively linking segment membership to demographic or occupational drivers, while SHAP applications remain confined to supervised prediction tasks and seldom explain why certain latent clusters arise from perception data. To date, no study has integrated unsupervised clustering with supervised classification and SHAP-based explainability to simultaneously uncover and interpret hidden employee perception profiles.

This study addresses that methodological gap by proposing and applying a three-phase framework:

- Phase 1: Clustering: K-means clustering is used to identify latent employee segments based solely on 23 perception-related survey items.
- Phase 2: Classification: A supervised LightGBM model predicts cluster membership from employees' demographic and occupational attributes.
- Phase 3: Explainability: SHAP is employed to quantify and visualize the global and local drivers of cluster membership, offering interpretable insights into the structural determinants of workplace perception.

Guided by this objective, the study seeks to answer the following research questions:

- RQ1: What distinct latent clusters of employee workplace perceptions can be identified using unsupervised machine learning?

- RQ2: To what extent can demographic and occupational attributes predict membership in these clusters?
- RQ3: Which factors emerge as the most influential drivers of cluster membership when examined through global SHAP analysis?
- RQ4: How do these drivers vary for individual employees, and what can local SHAP explanations reveal about unique cases?
- RQ5: How can the integration of clustering and explainable AI support the design of targeted, equitable HR interventions?

To the best of the author's knowledge, this study is the first to combine unsupervised clustering with supervised SHAP-based explainability in the context of employee perception analysis. By moving beyond descriptive segmentation and quantitatively linking latent profiles to their underlying demographic and occupational drivers, it contributes both methodologically and practically to the fields of HR analytics and information systems. The framework is replicable and adaptable to other organizational contexts and datasets, ensuring relevance beyond the Finnish case study presented here.

Importantly, the findings contribute to the United Nations Sustainable Development Goal 8, Decent Work and Economic Growth. By uncovering structural and demographic disparities in workplace perceptions and providing a transparent, data-driven framework for addressing them, the study supports the development of fairer, more inclusive, and more productive workplaces.

At a structural level, the remainder of this study is organized as follows: The next section reviews relevant literature on clustering techniques in HR analytics and Explainable Artificial Intelligence, highlighting the specific research gap addressed by this study. The methodology section then details the dataset, preprocessing steps, and the three-phase analytical framework. This is followed by the presentation of results, including cluster profiles, predictive performance, and SHAP-based explanations. The discussion section interprets the findings in relation to prior research and practical HR implications, and the study concludes by summarizing key contributions, outlining limitations, and identifying directions for future research.

II. Literature Review

A. Clustering Techniques in HR Analytics

Unsupervised clustering is widely used in HR analytics to group employees into meaningful profiles based on multidimensional data. Prior studies have applied K-means and hierarchical clustering to segment employees by well-being, engagement, competence, and risk profiles, often validating clusters using methods such as the elbow criterion, silhouette scores, or within-cluster sum of squares [6]–[10]. Other work has highlighted the availability of rich psychological, relational, and justice-related variables that can serve as clustering inputs, including information anxiety, fairness perceptions, and online learning experiences [11][12].

Machine learning has further expanded HR analytics, with clustering and classification methods supporting recruitment, employee segmentation, and retention strategies [13]. Hybrid frameworks combining clustering with supervised learning, such as K-means followed by ANN models, have shown improved predictive performance, particularly when combined with techniques like GAN-based data augmentation [14]. Additional applications include ML-enhanced Human Resource Information Systems [15], clustering of digital competencies in academia [16], health and well-being profiling using self-organizing maps or Gaussian mixture models [17][18], and pandemic-related segmentation of employee concerns followed by regression-based analysis of cluster predictors [19]. Despite these advances, most clustering-based HR studies remain descriptive and provide limited insight into the factors driving cluster membership.

B. Explainable AI and SHAP in Employment and Social Science

As ML adoption in HR grows, transparency and ethical concerns have increased, leading to the adoption of explainable AI techniques such as SHAP. SHAP has been extensively used to interpret supervised HR models, particularly for employee attrition and turnover prediction, revealing key drivers such as job satisfaction, tenure, overtime, and commute distance [20][21]. Some studies have extended SHAP by clustering explanation values to identify common exit patterns [22]. Complementary methods like Anchors have been used to generate rule-based explanations suitable for non-expert stakeholders [23], while recent research has integrated SHAP with fairness-aware deep learning to address bias in HR decision-making [24].

The incorporation of XAI into HR systems has been advocated to enhance transparency, ethical governance, and organizational responsibility [25] [26]. Related work in knowledge management and group decision support systems further connects explainable analytics with collective organizational decision-making [27]. Empirical evaluations show that interpretable ML models can balance predictive performance with clarity [28] [29]. Recent frameworks combining gradient boosting models with SHAP have successfully identified both established and novel predictors of HR outcomes [30], while hybrid optimization and decision-making approaches have improved the identification of influential turnover factors [31].

C. Gaps and Contributions of this Study

The literature reveals a clear separation between clustering-based employee segmentation and explainable supervised learning in HR analytics. Clustering studies effectively uncover latent employee groups based on perception or well-being data but largely remain descriptive and do not explain why individuals belong to specific clusters [6]–[10], [17]–[19]. In contrast, SHAP-based XAI research focuses on explaining supervised outcomes such as attrition or performance, without addressing the interpretability of latent segments derived from unsupervised analysis [20]–[22], [30],

[31]. Consequently, the formation of employee perception clusters remains largely unexplained.

Table I reveals a methodological gap in HR analytics, as prior studies focus either on clustering or on SHAP-based supervised models, but not both, within a unified pipeline. This study addresses this gap by proposing a three-phase framework that integrates unsupervised clustering, supervised classification, and SHAP-based explainability. K-means is used to identify latent perception profiles, LightGBM predicts cluster membership, and SHAP explains the drivers of these predictions. By treating cluster membership as an explainable outcome, the approach enables transparent interpretation of employee segments and represents a novel, replicable framework for data-driven HR decision-making.

TABLE I. COMPARISON OF THE CURRENT STUDY WITH PREVIOUS WORKS

Study (Ref)	Unsupervised Clustering	Supervised Prediction	Explainability (SHAP/XAI)	Purpose
[6]	Yes	No	No	Descriptive segmentation
[7]	Yes	No	No	Cluster profiling
[8]	Yes	No	No	Well-being profiling
[9]	Yes	No	No	Group comparison
[10]	Yes	No	No	Risk stratification
[17]	Yes	No	No	Pattern discovery
[19]	Yes	Yes	No	Predicting cluster membership
[20]	No	Yes	Yes	Explaining attrition prediction
[21]	No	Yes	Yes	Identifying turnover drivers
[22]	No	Yes	Yes	Explaining model outcomes
[30]	No	Yes	Yes	Interpreting supervised outcomes
This study	Yes	Yes	Yes	Explaining latent cluster formation

III. METHODOLOGY

The methodological approach adopted in this study follows a structured, three-phase framework designed to uncover latent workplace perception profiles and interpret their underlying drivers. In the first phase, K-means clustering is applied to survey responses capturing multiple aspects of employees' workplace experience, producing distinct perceptual groups. In the second phase, a supervised classification model predicts cluster membership based on demographic and occupational attributes, enabling the identification of structural determinants linked to each group. The final phase employs SHAP-based explainability to quantify and visualize both global and local feature contributions, ensuring that the predictive insights remain transparent and actionable. The overall process is illustrated in Fig. 1, which provides a step-by-step overview of the analytical pipeline implemented in this study.

A. Dataset and Data Pre-Processing

This study is based on data from the FSD3784 Finnish Working Life Barometer 2022, an annual survey produced by Statistics Finland and the Ministry of Economic Affairs and Employment and accessed via the Finnish Social Science Data Archive's Aila service. The dataset is licensed for non-commercial academic use under Aila's terms, requiring proper citation and the protection of individual confidentiality [32]. It provides nationally representative information on Finnish employees' perceptions of working life, including leadership, well-being, workplace culture, and employment conditions.

For this study, 23 perception-related survey items were selected as inputs for the Clustering Analysis. Data preprocessing involved removing responses coded with placeholder values (9, 999, 888) and excluding rows with missing values to ensure data quality. Variable naming conventions and categorical coding schemes are documented in Appendices A and B to support transparency, reproducibility, and accurate interpretation, with all definitions preserved in accordance with the original dataset and licensing conditions.

B. Phase 1: Clustering Perceptions

In the first phase, an unsupervised learning approach was used to identify latent employee segments based on workplace perceptions. Twenty-three survey items capturing fairness, inclusion, well-being, leadership trust, psychological safety, and organizational culture were selected from the Finnish HR dataset and are listed in Appendix A with their cluster-wise means.

After data cleaning, ordinal responses were one-hot encoded to preserve interpretability and support effective clustering, resulting in a higher-dimensional but suitable feature space. PCA was then applied for dimensionality reduction and two-dimensional visualization. The optimal number of clusters was identified using silhouette analysis for K = 2 to 10, with the highest score at K = 2, indicating two distinct employee perception clusters (see Fig. 2).

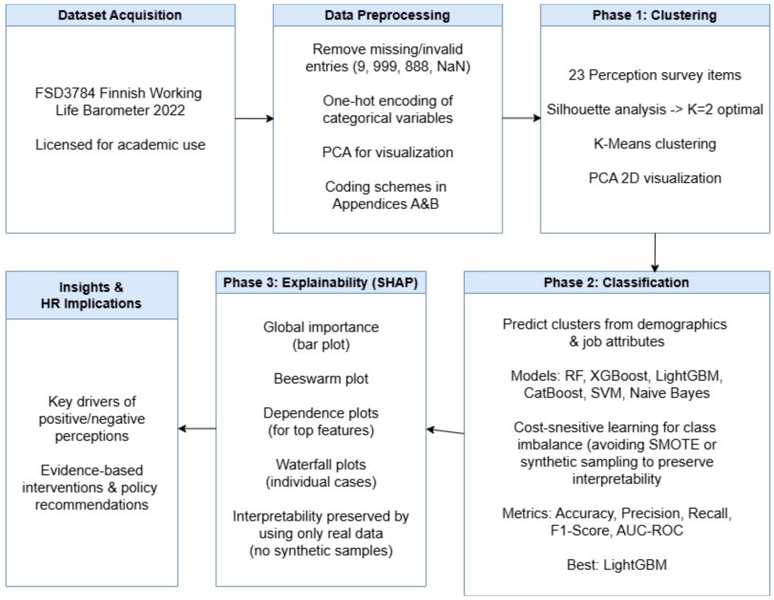

Fig. 1. Three-phase methodological framework for uncovering and interpreting latent workplace perception profiles.

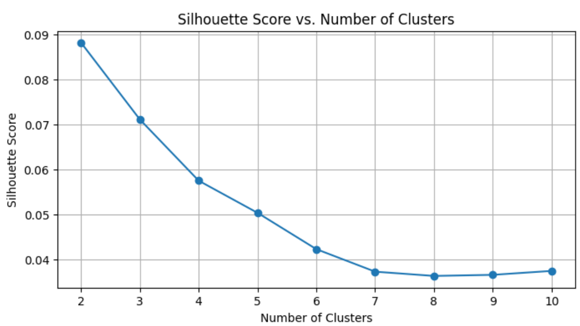

Fig. 2. Silhouette scores for K-means clustering across different numbers of clusters (K = 2–10).

- Once the optimal number of clusters was established, the K-Means algorithm was applied with K = 2 to segment employees based on their perception profiles. The clustering process grouped individuals with similar response patterns into the same cluster, without prior knowledge of their identities or job roles. The resulting clusters were then visualized using the previously computed PCA components, as depicted in Fig. 3.

Fig. 3. Two-dimensional PCA projection of employee survey responses, showing the distribution of employees across two identified clusters.

- The final clustering output included 1,793 employees, distributed across two groups: Cluster 0, comprising 766 individuals, and Cluster 1, with 1,027 individuals. These clusters served as the foundation for the next stage of the analysis, where their defining characteristics and interpretation would be examined in detail.

C. Phase 2: Performing Machine Learning on Cluster Membership

Following the clustering of employees based on their survey responses, the second phase of the analysis aimed to determine whether cluster membership could be predicted using employees' demographic and occupational attributes. Unlike the first phase, which employed an unsupervised learning approach driven solely by perceptual survey data, this phase adopted a supervised machine learning framework. The goal was to examine how well structured background variables could classify employees into their respective clusters, thereby revealing how deeply rooted personal and job-related characteristics contribute to patterns in workplace perception.

To this end, a range of machine learning models was trained and evaluated using a consistent set of explanatory features. These included variables such as gender, age, occupation, employer type, broad and detailed industry categories, full-time versus part-time status, employment permanence, side job status, side work hours, overall employment duration, and indicators of socioeconomic status and class. Each of these features was selected for its potential relevance in shaping employees' lived experiences within their professional environment.

The following classification models were employed:

- Random Forest (RF): An ensemble-based method that applies the bagging strategy by constructing a collection of T decision trees, each trained on a randomly sampled subset of the training set. For a given input vector x, the final class prediction is obtained by combining the outputs of all trees through majority voting.

$$\hat{y} = mode\,\{h_t(x)\}_{t=1}^{T} \quad (1)$$

where, $h_t(x)$ represents the prediction from tree t.

- XGBoost: A highly efficient implementation of gradient boosting that constructs an ensemble of weak learners in a sequential manner. Each successive tree is trained to reduce the residual errors made by the previous ones, aiming to minimize a specified loss function.

$$F_m(x) = F_{m-1}(x) + \gamma h_m(x) \quad (2)$$

where, γ is the learning rate and $h_m(x)$ is the newly added tree.

- LightGBM: A fast and scalable gradient boosting framework that builds decision trees using a leaf-wise growth strategy rather than the traditional level-wise approach. This strategy focuses on splitting the leaf with the highest loss reduction, which typically leads to better accuracy and faster convergence. Like other boosting methods, LightGBM minimizes a regularized objective function:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \sum_{t=1}^{T} \Omega(f_t) \quad (3)$$

where, $\mathcal{L}$ is the loss function (e.g., logistic loss), $f_t(x)$ is the newly added tree at iteration t, and $\Omega(f_t)$ is the regularization term controlling model complexity.

- CatBoost: A gradient boosting framework designed to handle categorical features effectively and perform well on imbalanced datasets. It offers strong predictive performance with minimal need for extensive feature engineering. Like other gradient boosting approaches, it works by minimizing a regularized loss function:

$$\mathcal{L} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{t=1}^{T} \Omega(f_t) \quad (4)$$

where, $\hat{y}_i = \sum_{t=1}^{T} f_t(x_i)$, and f_t are the base decision trees learned sequentially.

- Support Vector Machine (SVM): A margin-oriented classifier that seeks to maximize the separation between classes. By using a radial basis function (RBF) kernel, it maps the input features into a higher-dimensional space, enabling the algorithm to capture complex, nonlinear decision boundaries.

$$f_x = sign\left(\sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b\right) \quad (5)$$

where, $K(x_i, x) = \exp(-\gamma \|x_i - x\|^2)$ is the RBF kernel, and α_i, y_i, and b are model parameters.

- Naive Bayes: A probabilistic classifier grounded in Bayes' theorem, built on the assumption that features are conditionally independent given the class label. This simplifying assumption allows efficient computation of posterior probabilities. For a class c, the posterior probability given features $x_1, x_2, \ldots, x_n$ is expressed as:

$$P(c \mid x_1, \ldots, x_n) \propto P(c) \prod_{i=1}^{n} P(x_i \mid c) \quad (6)$$

The predicted class is:

$$\hat{y} = \arg\max_{c \in C} P(c) \prod_{i=1}^{n} P(x_i \mid c) \quad (7)$$

All models were trained using a cost-sensitive learning approach and assessed on a separate test set using a consistent set of standard classification metrics:

- Accuracy: Measures the proportion of all correctly classified instances relative to the total number of predictions across every class.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

- Precision: Represents the fraction of instances labeled as positive that are truly positive, indicating the reliability of the model in predicting positives.

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

- Recall (Sensitivity): Denotes the share of actual positive cases that are correctly identified by the model, emphasizing its effectiveness in capturing positives.

$$Recall = \frac{TP}{TP + FN} \quad (10)$$

- F1-Score: Defined as the harmonic mean of precision and recall, this metric provides a balanced assessment, particularly useful in scenarios with imbalanced class distributions.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (11)$$

- AUC-ROC (Area Under the Receiver Operating Characteristic Curve): Evaluates the model's ability to distinguish between classes over all possible decision thresholds, offering a threshold-independent measure of classification performance.

These metrics together offer a well-rounded assessment of each model's capability to predict cluster membership, particularly in light of the dataset's inherent class imbalance. A detailed comparison of model performance is provided in Section IV(B).

D. Phase 3: Explainability of Cluster Membership

The third analytical phase focused on identifying which background features most strongly influenced employees' cluster assignments, thereby linking predictive classification with interpretable insights. While prior results showed that cluster membership could be predicted accurately, this phase examined why employees were more likely to belong to one cluster rather than the other.

To achieve this, SHAP was applied to explain the predictions of the best-performing classifier, LightGBM. SHAP decomposes model outputs into feature-level contributions, providing both global importance across the dataset and local explanations for individual predictions. The analysis used the same feature set as the classification stage, including continuous and categorical variables. Categorical attributes such as occupation, employer type, and industry were given particular attention, with detailed encodings documented in Appendix B, while variables like age and employment duration required no additional interpretation.

The SHAP analysis revealed the key demographic and occupational factors driving membership in Cluster 0 or Cluster 1. By converting accurate but opaque predictions into transparent explanations, this phase clarified the structural determinants of employee sentiment and supported the development of targeted, evidence-based interventions to improve employee well-being and organizational climate.

IV. Results

A. *Cluster Profiles*

Following the identification of two employee clusters using K-means, a profile analysis was conducted by calculating mean responses for each of the 23 perception-based survey items within both clusters. These means served as summary indicators of how each group perceived the work environment and organizational culture. As reported in Appendix A, clear and consistent differences emerged across all items, with Cluster 0 showing more favorable perceptions than Cluster 1.

For items measured on a 4-point agreement scale, lower mean values in Cluster 0 indicated stronger agreement with positive statements, while higher means in Cluster 1 reflected more negative perceptions. For items measured on a 5-point frequency scale, higher mean values corresponded to more positive outcomes, and Cluster 0 consistently scored higher than Cluster 1, indicating greater motivation, well-being, and perceived inclusivity.

To illustrate the magnitude of these differences, a difference plot was produced by subtracting Cluster 0 means from Cluster 1 means for each item and sorting the results. As shown in Fig. 4, the direction of the differences reflects the underlying measurement scales, with negative or positive values indicating less or more favorable perceptions in Cluster 1 relative to Cluster 0.

This analysis supports the interpretation that Cluster 0 represents a more engaged, supported, and satisfied group of employees, whereas Cluster 1 captures employees with more negative perceptions of their work environment. This finding directly addresses RQ 1 by demonstrating that unsupervised machine learning can reveal two clear and meaningful latent clusters of employee workplace perceptions. The distinct patterns of engagement, well-being, and organizational support between Cluster 0 and Cluster 1 confirm that perception-based segmentation is both feasible and informative in this context. Such distinctions provide a valuable foundation for targeted human resource interventions aimed at improving workplace conditions and employee experience for the more vulnerable group.

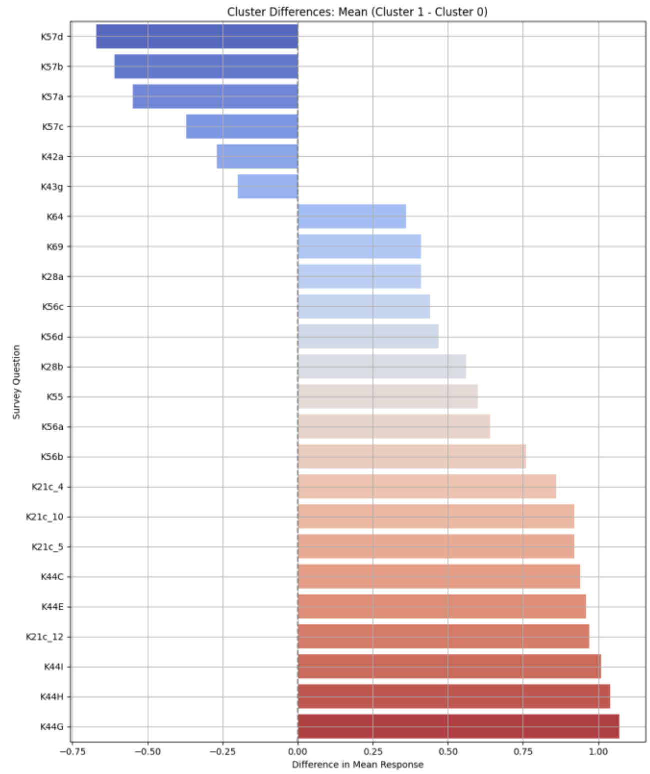

Fig. 4. Mean response differences between cluster 1 and cluster 0 for the 23 perception-related survey items, sorted from largest to smallest disparity.

B. *Machine Learning Results*

To assess the predictive power of structural employee attributes for determining cluster membership, a comparative evaluation was conducted across seven machine learning classifiers. These included tree-based ensemble models and baseline algorithms trained on identical data splits and feature sets. The classifiers evaluated were Random Forest, LightGBM, XGBoost, CatBoost, Naive Bayes, and SVM. The use of ensemble-based models alongside baseline classifiers follows established practices in HR analytics research, where comparative modeling is commonly employed to balance predictive performance and interpretability [13], [28], [29].

All models were trained using a stratified 70/30 train-test split and evaluated using standard classification metrics, including accuracy, precision, recall, F1-score, and the area under the ROC curve (AUC). This evaluation strategy is consistent with prior supervised HR analytics studies that emphasize robust performance assessment when predicting employee-related outcomes [20], [21], [30].

Among the models tested, LightGBM demonstrated the highest overall performance, achieving an accuracy of 90.65%, a precision of 91.80%, a recall of 89.02%, and an F1-score of 90.38%. It also recorded the highest AUC score of 0.9826, indicating excellent discrimination capability. The strong performance of gradient boosting models aligns with previous HR analytics research showing that tree-based boosting frameworks outperform traditional classifiers when modeling complex, non-linear relationships in employee data [14], [30], [31]. These consistent results across multiple metrics motivated the selection of LightGBM as the foundation for the explainable AI analysis in the subsequent phase.

Other ensemble models, including XGBoost, Random Forest, and CatBoost, also showed strong performance, though slightly below LightGBM, consistent with prior HR analytics studies where boosting and bagging methods outperform simpler classifiers [13], [20], [21]. Naive Bayes and SVM exhibited lower performance but served as informative baselines, reflecting the interpretability–accuracy trade-off commonly reported in HR machine learning research [28], [29]. Table II summarizes the evaluation metrics, and the ROC curves in Fig. 5 further confirm the superiority of gradient boosting models and support the use of LightGBM for SHAP-based explainability, in line with recent XAI studies in HR analytics [30], [31].

TABLE II. PERFORMANCE COMPARISON OF MACHINE LEARNING MODELS IN PREDICTING CLUSTER MEMBERSHIP FROM DEMOGRAPHIC AND OCCUPATIONAL ATTRIBUTES

Model	Accuracy	Precision	Recall	F1-score	AUC
LightGBM	0.9065	0.918	0.8902	0.9038	0.9826
Random Forest	0.8897	0.8882	0.9153	0.9015	0.9679
XGBoost	0.8748	0.8735	0.8633	0.8684	0.9713
CatBoost	0.8486	0.8598	0.8376	0.8486	0.9604
SVM	0.7383	0.748	0.7143	0.7308	0.8389
Naive Bayes	0.7103	0.7411	0.7182	0.7295	0.8251

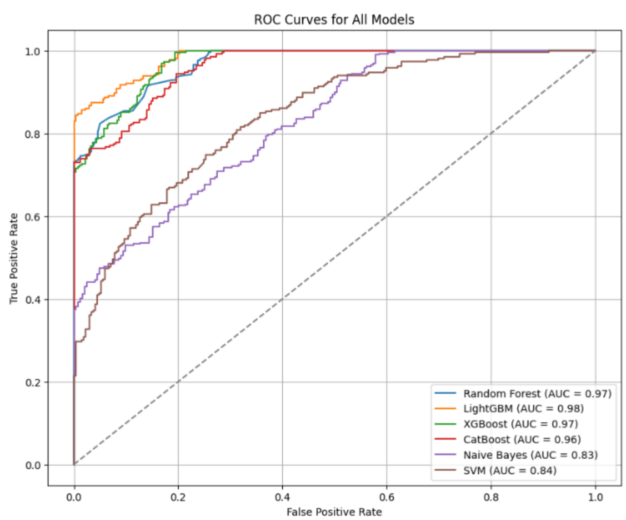

Fig. 5. Receiver operating characteristic (ROC) curves comparing all classifiers on the test set.

This outcome directly addresses RQ 2 by showing that demographic and occupational attributes can predict cluster membership with high accuracy. In particular, the top-performing LightGBM model achieved over 90% accuracy and an AUC of 0.9826, indicating that structural employee characteristics carry substantial predictive power for distinguishing between the two latent perception clusters. Given its strong and consistent performance across all evaluation metrics, LightGBM was selected as the final model for further explainability analysis. Its ability to handle complex interactions and provide insight into feature importance made it especially well-suited for understanding the key drivers behind cluster assignments. The next section explores how this model was interpreted to uncover the most influential variables shaping employee perceptions.

C. SHAP Insights

To interpret the decisions made by the LightGBM classifier and gain a deeper understanding of the structural variables influencing cluster membership, SHapley Additive exPlanations (SHAP) was employed as a model-agnostic explainability method. SHAP values provide a theoretically grounded framework for attributing the contribution of each feature to a specific prediction, enabling both global and local interpretability.

At the global level, the SHAP bar plot (Fig. 6) summarizes the mean absolute SHAP values of all input features, thus ranking them by their average impact on the model's output. According to this plot, the two most influential features in determining cluster membership were Employment Duration and Age, suggesting that the length of time an individual has spent in the workforce and their age are the primary differentiators between those in the more positive (Cluster 0) versus the more negative (Cluster 1) group. These were followed by IndustryDetailed, Gender, and SocioeconomicStatus, which also showed substantial contributions to model output. This directly answers RQ 3 by demonstrating how global SHAP analysis can pinpoint the most influential predictors of cluster membership, offering clear and interpretable evidence of the structural and demographic factors that shape workplace perception profiles.

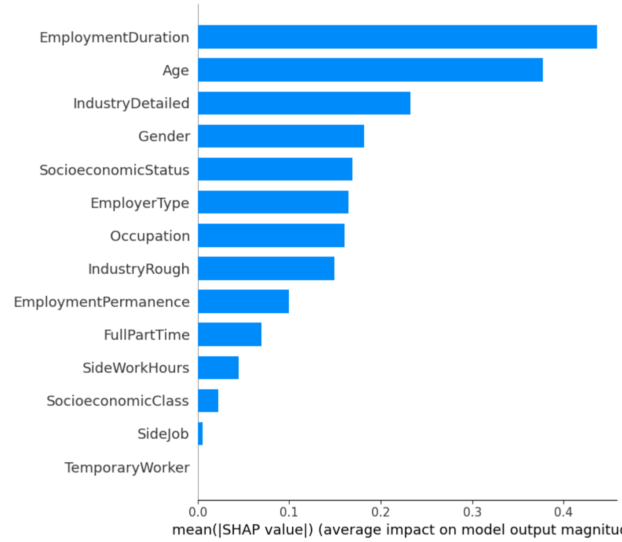

Fig. 6. Global SHAP feature importance plot ranking demographic and occupational variables by their average impact on LightGBM model predictions.

The SHAP beeswarm plot (Fig. 7) further extends this insight by showing how different values of each feature affect the model's predictions. Each point in the beeswarm plot corresponds to a single employee in the dataset, with the horizontal position indicating the SHAP value (i.e., the effect on the model's prediction) and the color encoding the original feature value (from low to high). This view reveals not only which features are important but also how their specific values push the model toward predicting Cluster 0 or Cluster 1. To support the interpretation of these color gradients and categorical values, the complete coding schemes and enumerated representations of all categorical features used in the model are presented in Appendix B.

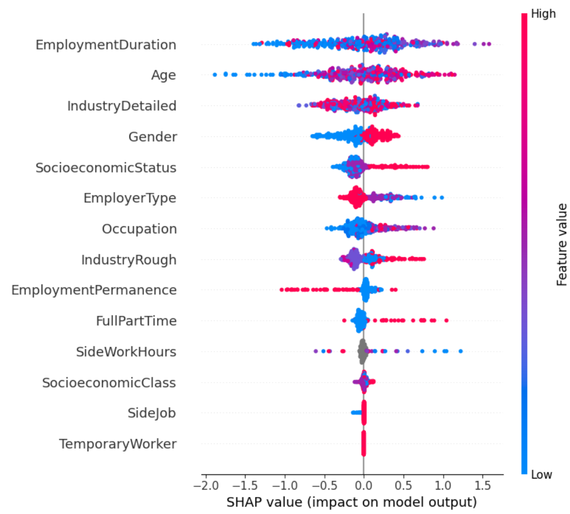

Fig. 7. SHAP beeswarm plot illustrating the direction and magnitude of feature effects on individual predictions, with color representing feature values.

The SHAP analysis highlights how demographic and occupational features influence cluster assignment, providing interpretable insight into the structural drivers of employee workplace perceptions. For Employment Duration, the SHAP beeswarm plot shows a mix of effects, indicating a non-linear relationship between tenure and cluster membership. The corresponding dependence plot reveals that moderate tenure, approximately 0 to 15 years, increases the likelihood of belonging to Cluster 1, while very long tenure shifts predictions toward Cluster 0. This pattern suggests higher dissatisfaction during mid-career stages, followed by more favorable perceptions among long-tenured employees, a finding that aligns with prior HR analytics research linking tenure-related dynamics to changes in employee well-being and engagement [20], [21], [30].

Age exhibits a clearer and more monotonic pattern. Higher age values consistently push predictions toward Cluster 1, while younger employees are more likely to fall into Cluster 0, indicating a largely linear association between increasing age and less favorable workplace perceptions. Similar age-related gradients have been reported in earlier explainable AI studies of employee outcomes, where age emerged as a significant contributor to supervised HR predictions and risk stratification [20], [21]. The ability of SHAP to reveal both the direction and magnitude of this effect reinforces its suitability for interpreting demographic influences in HR contexts [25], [26].

Gender also shows a distinct directional effect, with female employees more likely to be classified into Cluster 1 and male employees into Cluster 0. This result is consistent with prior research emphasizing gender-based disparities in workplace experiences and the value of explainable AI in uncovering such structural inequalities in HR decision-making systems [24], [25]. Similarly, higher Socioeconomic Status values corresponding to workforce-level positions increase the likelihood of Cluster 1 membership, whereas senior or lower-level official roles appear more protective. These findings resonate with earlier clustering and explainable AI studies that highlight the role of hierarchical position and job security in shaping employee perceptions and outcomes [17], [18], [30].

Employer Type further differentiates clusters, as public sector employment at the state or municipal level increases the probability of belonging to Cluster 1, while private sector employment reduces it. Comparable sectoral effects have been observed in previous clustering-based analyses of employee well-being and organizational risk profiles, particularly in studies examining institutional constraints and public sector stressors [6], [9], [19]. Categorical variables such as IndustryDetailed, Occupation, and IndustryRough require interpretation through SHAP dependence plots due to the lack of ordinal meaning in their encodings, an approach that mirrors best practices in recent SHAP-based HR analytics research [22], [30].

Other variables, including EmploymentPermanence, FullPartTime, SideWorkHours, SocioeconomicClass, SideJob, and TemporaryWorker, show minimal influence on predictions. This pattern is consistent with prior explainable machine learning studies in HR contexts, where a subset of structural features typically dominates model behavior while others contribute marginally once core demographic and occupational drivers are accounted for [28], [29].

Overall, the SHAP visualizations provide transparent and theoretically grounded insight into how background characteristics shape cluster membership and validate the LightGBM model's predictions. By combining global feature importance with feature-level dependence analysis, the results extend earlier SHAP-based HR studies that focused exclusively on supervised outcomes, demonstrating how Explainable Artificial Intelligence can also be applied to interpret latent clusters derived from unsupervised analysis [20]–[22], [30], [31]. The detailed examination of dependence plots clarifies how individual attributes contribute to classification into more or less favorable employee perception clusters, with categorical coding schemes documented in Appendix B.

The SHAP dependence plot for employment duration [Fig. 8(a)] shows a clear nonlinear pattern. The likelihood of belonging to Cluster 1 increases during the early to mid-career phase, peaking around 15–20 years of tenure, and then declines for employees with very long service. This suggests that mid-career employees are more prone to unfavorable workplace perceptions, whereas long-tenured employees tend to report more positive or stabilized experiences.

Similarly, the SHAP dependence plot for age [Fig. 8(b)] indicates a steady increase in the probability of being classified into Cluster 1 as age rises. Younger employees are less likely to exhibit negative workplace perceptions, while the risk of dissatisfaction becomes more pronounced after the age of 30. Overall, both tenure and age emerge as important predictors, highlighting mid-career and older employees as groups more vulnerable to negative workplace sentiment.

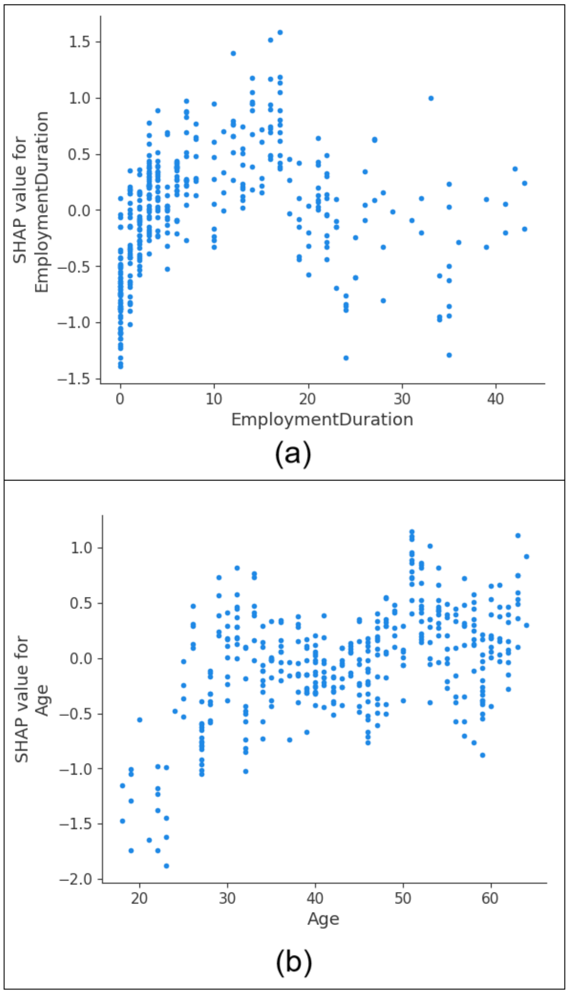

Fig. 8. SHAP dependence plots for: (a) employment duration, and (b) age.

The SHAP dependence plot for IndustryDetailed [Fig. 9(a)] shows that several sectors are strongly associated with classification into the less favorable Cluster 1. Industries such as electricity, gas and heating supply, administrative and support services, public administration and defense, health and social services, and arts and recreation exhibit higher SHAP values, suggesting elevated risk of negative workplace perceptions. These effects likely stem from sector-specific conditions, including high stress, burnout, emotional demands, irregular employment, or job instability.

In contrast, transportation and storage display the lowest SHAP values, indicating a strong association with the more favorable Cluster 0, potentially due to stable demand and expanding logistics-related employment. Education also tends toward lower SHAP values, which may reflect formal employment structures, long-term contracts, and stronger job protections. Finance and real estate show SHAP values near zero, suggesting a balanced risk profile that reflects internal heterogeneity between stable and volatile roles. Retail trade and accommodation and catering services exhibit a wide dispersion of SHAP values, indicating mixed employment conditions and diverse job types within these sectors.

The SHAP dependence plot for Gender [Fig. 9(b)] reveals a clear disparity, with female gender associated with higher SHAP values and a greater likelihood of belonging to Cluster 1, while male gender shows lower values and a protective effect. This finding indicates that gender meaningfully influences cluster assignment and may reflect underlying structural or socio-economic inequalities present in the data.

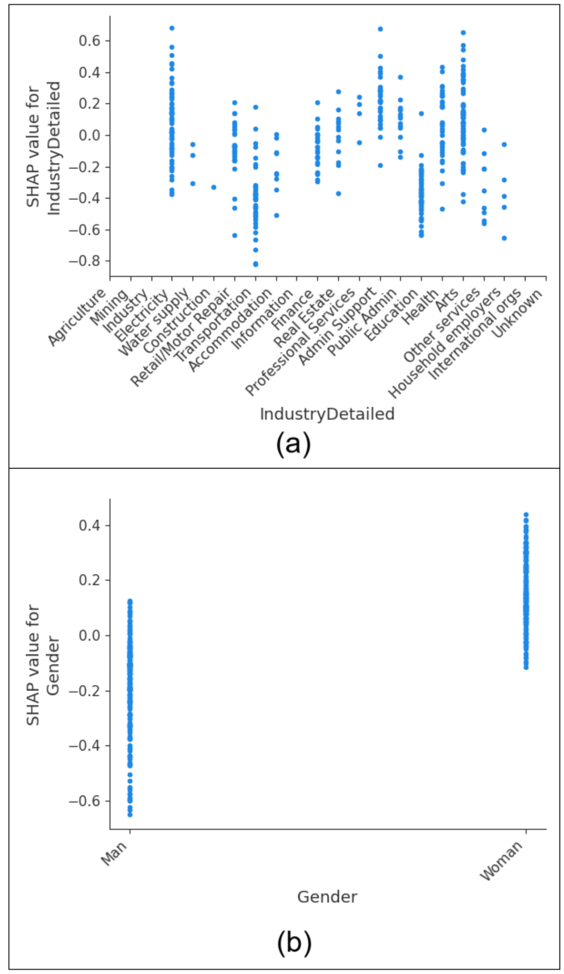

Fig. 9. SHAP dependence plots for: (a) IndustryDetailed, and (b) Gender.

The SHAP dependence plot for Socioeconomic Status [Fig. 10(a)] reveals a clear upward trend in SHAP values from senior officials to workforce-level employees, indicating that lower socioeconomic standing is associated with a higher likelihood of being placed in the unfavorable cluster (Cluster 1). Specifically, individuals categorized as "workforce" exhibit the highest SHAP values, suggesting that their status strongly contributes to predicting membership in the bad cluster. In contrast, senior officials tend to have negative SHAP values, meaning their status lowers the model's probability of assigning them to the adverse group. This pattern suggests that individuals occupying more precarious or lower-ranking positions within the employment hierarchy may face increased vulnerabilities or risks that are being captured by the model. It also reflects how structural inequalities manifest in employment contexts and are detectable in predictive clustering models. The SHAP dependence plot for Employer Type [Fig. 10(b)] highlights significant differences in risk association based on the nature of the employer. Individuals employed by the state exhibit the highest SHAP values, meaning that this group is more strongly associated with Cluster 1, the undesirable or vulnerable group. This could reflect structural stressors or bureaucratic instability within state institutions. Employees working under municipalities or joint municipal authorities also show positive SHAP values, though slightly lower than those in state employment. In contrast, those employed in the private sector predominantly have lower SHAP values, suggesting a reduced likelihood of belonging to the high-risk cluster. This pattern may seem counterintuitive, as private sector jobs are often perceived as less secure, but it may be reflective of better job-matching, more competitive compensation, or other contextual factors that favor private employment outcomes.

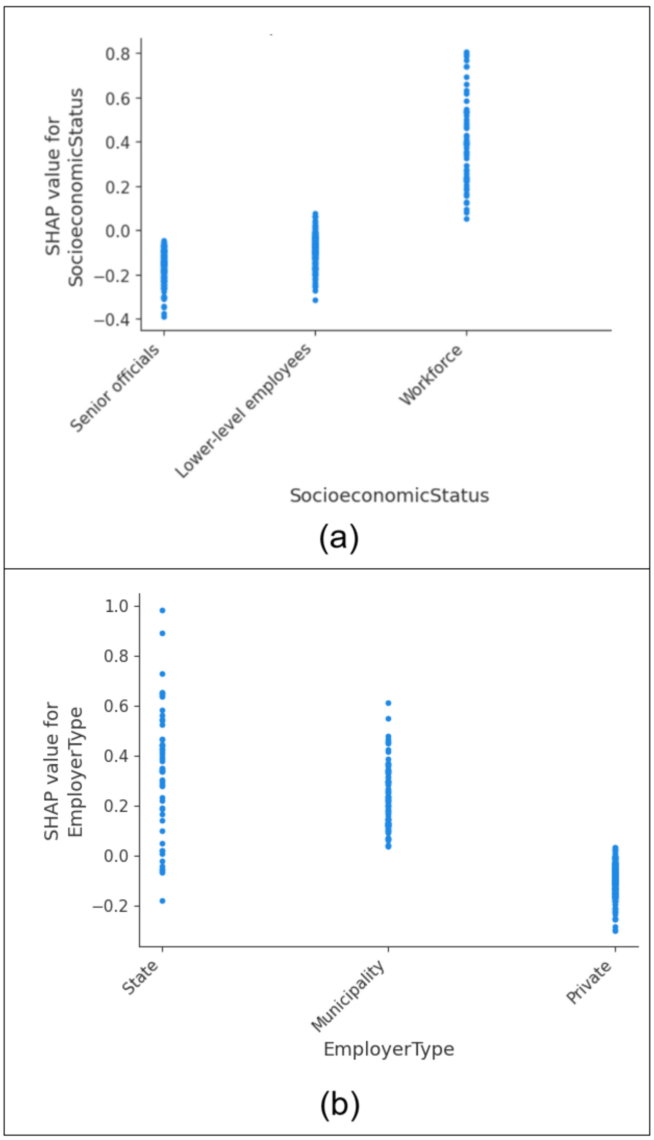

Fig. 10. SHAP dependence plots for: (a) Socioeconomic status, and (b) Employer type.

The SHAP dependence plot for Occupation [Fig. 11(a)] demonstrates clear differences in how job categories influence cluster assignments. Service and sales employees, office and customer service workers, and process and transportation workers exhibit higher SHAP values, indicating a greater likelihood of belonging to the higher-risk Cluster 1. These roles are often characterized by high turnover, customer-facing stress, or physically demanding conditions. Construction, repair, and manufacturing workers also show relatively elevated SHAP values, consistent with the contract-based and physically intensive nature of this work.

In contrast, management, special experts, and cognoscenti display lower SHAP values, suggesting stronger alignment with the more favorable cluster. These occupations typically offer greater stability, clearer career progression, and stronger institutional support. Overall, the results indicate occupational stratification, with higher risk associated with customer-facing and physically intensive roles, and lower risk linked to white-collar or institutional positions.

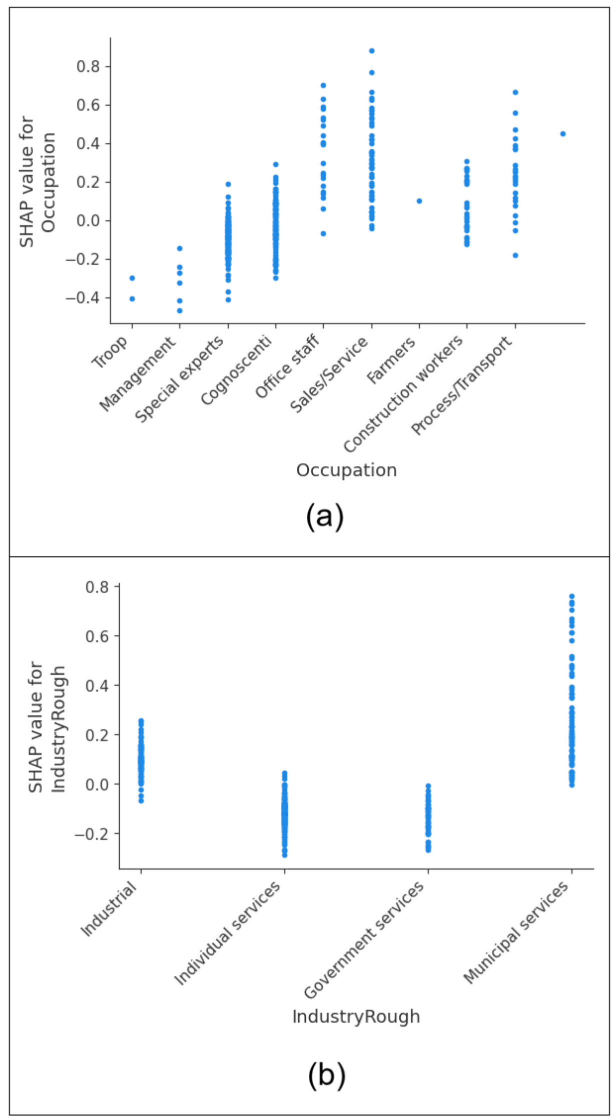

Fig. 11. SHAP dependence plots for: (a) Occupation, and (b) IndustryRough.

The SHAP dependence plot for the rough industry classification [Fig. 11(b)] further emphasizes institutional differences. Municipal services show the highest SHAP values, indicating increased vulnerability, potentially due to contract conditions or budgetary constraints. Government services and individual services exhibit lower SHAP values, while the industrial, applied, and excavated sector shows moderately elevated values, possibly reflecting physical demands or economic volatility. These findings underscore the importance of broad industry context in shaping predicted employment risk.

After examining the SHAP dependence plots, the analysis is extended to individual-level explanations to illustrate how the LightGBM model assigns cluster membership. SHAP waterfall plots are used for three randomly selected employees representing low, medium, and high predicted probabilities of belonging to Cluster 1, which reflects less favorable workplace experiences. While dependence plots summarize average feature effects, waterfall plots show how combinations of features interact differently for each individual, highlighting that similar attributes may have varying impacts depending on context.

For the first employee (Fig. 12), the SHAP waterfall plot shows a strongly negative logit value of −0.852, corresponding to a 29.9% probability of belonging to Cluster 1 and a clear alignment with Cluster 0. The dominant factor driving this outcome is the employee's short employment duration of one year, which strongly reduces the predicted risk, consistent with earlier findings that early-career employees tend to report more positive perceptions. The employee's industry, Financial and Insurance Activities, further lowers the SHAP value, reflecting the generally favorable association of this sector with workplace perceptions.

Additional protective factors include male gender, which slightly reduces risk, and senior occupational status, which is associated with lower SHAP values due to greater autonomy and organizational support. Although the employee is 42 years old, the age effect remains moderate, as mid-career employees lie below the steeper risk increase observed at older ages. Private sector employment also contributes to lowering the predicted risk, in line with earlier results showing favorable SHAP values for this employer type. Other variables, such as rough industry classification, occupation, and side work hours, have minimal impact on the final prediction.

Overall, the SHAP explanation confirms the model's classification and demonstrates how short tenure, private sector employment, seniority, and sector characteristics jointly reduce the likelihood of negative workplace perceptions, placing this employee firmly within the more favorable cluster.

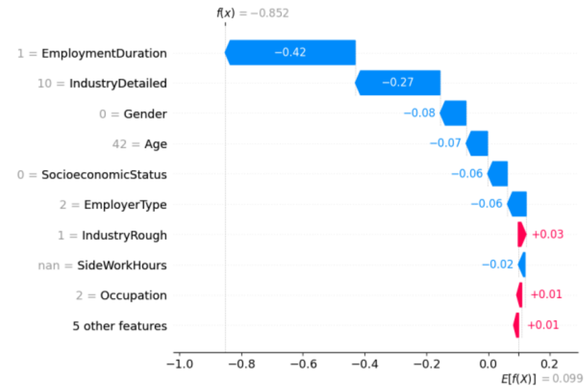

Fig. 12. Local SHAP waterfall plot for an employee with a low probability of belonging to Cluster 1 (negative cluster).

The SHAP waterfall plot for the second employee (Fig. 13) shows a logit value of approximately 0.044, corresponding to a 51.1% probability of belonging to Cluster 1 and a 48.9% probability of Cluster 0. This indicates a borderline case, with the prediction lying close to the model's baseline and no strong inclination toward either cluster.

The strongest factor reducing the likelihood of Cluster 1 membership is the employee's short employment duration of two years, which produces a negative SHAP value. As observed in the dependence analysis, early tenure is generally associated with more favorable workplace perceptions. In contrast, the employee's age of 39 contributes positively to the SHAP value, reflecting the gradual increase in risk observed with increasing age.

The rough industry classification, "Individual services", slightly increases the predicted risk, consistent with the variability and demands often associated with this sector. This upward effect is partially offset by the employee's senior socioeconomic status, which has a protective influence, and by male gender, which also contributes to a lower predicted risk.

The occupation category "Special Experts" adds a small positive contribution, suggesting modest risk, while private sector employment and full-time status both slightly reduce the likelihood of belonging to Cluster 1. Employment permanence has only a negligible positive effect. Overall, the combination of protective and risk-enhancing factors results in an equivocal prediction, illustrating the nuanced interplay of demographic, occupational, and institutional characteristics in shaping workplace perceptions.

The SHAP waterfall plot for the third employee (Fig. 14) shows a strongly positive logit value of +2.197, corresponding to a 90% probability of belonging to Cluster 1 and indicating a high likelihood of negative workplace perceptions. This classification is driven by the cumulative effect of several risk-enhancing factors.

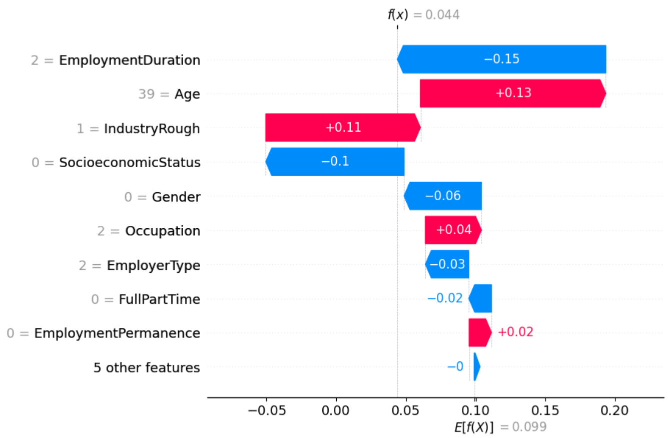

Fig. 13. Local SHAP waterfall plot for an employee with a borderline probability of belonging to Cluster 1 (negative cluster).

The employee's long employment duration of 21 years is a major contributor, as extended tenure is associated with higher SHAP values and increased risk of dissatisfaction. The female gender also substantially increases the SHAP value, consistent with earlier results showing a higher likelihood of women being classified into Cluster 1. Socioeconomic status, categorized as workforce, further elevates risk, reflecting the greater vulnerability of lower-ranking positions.

Industry and occupation also play significant roles. Employment in accommodation and catering, as well as service and sales occupations, both contribute positively to the SHAP value, aligning with prior findings that these contexts are associated with higher stress and instability. Age adds additional upward pressure, as SHAP values increase with advancing age. Smaller positive contributions from rough industry classification, employer type, and employment status further reinforce the prediction.

Together, these demographic and occupational characteristics generate a high SHAP score, leading to confident assignment to Cluster 1 and illustrating how multiple structural factors can interact to produce elevated negative workplace sentiment.

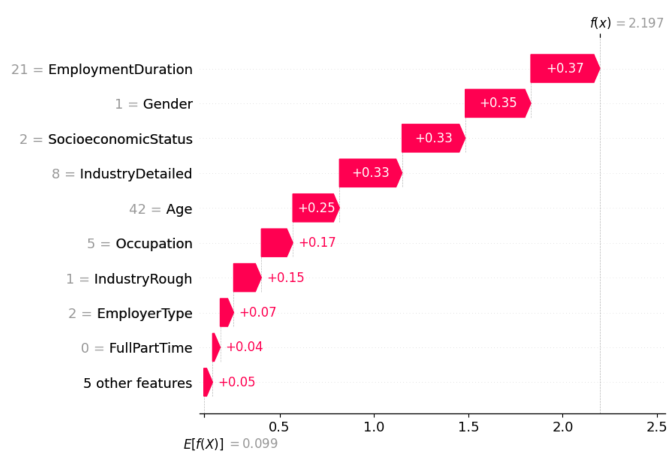

Fig. 14. Local SHAP waterfall plot for an employee with a high probability of belonging to cluster 1 (negative cluster).

This directly addresses RQ 4 by showing how local SHAP explanations uncover the unique combination of factors influencing individual employees' cluster assignments. These individualized insights reveal that even when employees share certain attributes, differences in their broader profiles can lead to distinct prediction outcomes, underscoring the value of local interpretability in HR decision-making.

V. Discussion and Future Work

The identification of two distinct employee perception clusters reveals pronounced differences in engagement, well-being, and organizational support. Employees in Cluster 0 consistently report more favorable workplace experiences, including stronger perceptions of fairness, leadership quality, developmental opportunities, motivation, and overall well-being. In contrast, Cluster 1 reflects systematically less favorable perceptions, characterized by lower trust, weaker inclusion, and reduced job satisfaction. These findings are consistent with prior clustering-based HR analytics research that has demonstrated the existence of heterogeneous employee profiles shaped by differences in well-being, engagement, and organizational support rather than random variation [6]–[10], [17]–[19]. Together, these results reinforce the view that workplace perceptions are structured outcomes influenced by identifiable demographic and occupational conditions embedded within organizational contexts.

The application of SHAP-based explainability provides deeper insight into the mechanisms underlying these differences. Employment duration and age emerge as the most influential predictors of cluster membership, indicating that workplace perceptions evolve across career stages. Shorter tenure and younger age are generally associated with more positive perceptions, whereas mid- to late-career stages exhibit an increased likelihood of negative sentiment. Similar tenure- and age-related effects have been reported in supervised HR analytics and explainable AI studies examining employee satisfaction, attrition, and well-being, where cumulative exposure to organizational stressors and unmet career expectations were identified as key drivers [20], [21], [30]. In addition, industry affiliation, gender, socioeconomic status, and occupation play important roles, revealing how institutional environments and labor market positions shape employees' lived experiences. Elevated risk in sectors such as health and social services, electricity, gas and heating supply, arts and recreation, and administrative support aligns with earlier clustering studies that identified sector-specific vulnerability profiles and occupational risk patterns [6], [9], [17], [19]. The consistency between global SHAP importance rankings and individual-level explanations further strengthens the credibility of these findings and reflects best practices in explainable HR analytics [22], [30].

From a human resource management perspective, the results offer clear guidance for targeted and evidence-based interventions, addressing RQ 5. Mid-career employees may benefit from structured career development pathways, mentoring programs, and role redesign initiatives aimed at restoring motivation and engagement, echoing recommendations from prior HR analytics and well-being studies [13], [18]. Sector-specific interventions are particularly relevant for high-risk industries, where workload intensity, emotional labor, or institutional constraints may drive negative perceptions, as documented in earlier clustering-based analyses of employee well-being and stress [6], [17], [19]. Additionally, the observed disparities linked to gender and socioeconomic status highlight the importance of equity-oriented HR policies that address structural inequalities, promote inclusive leadership, and enhance access to resources and advancement opportunities, reinforcing calls from recent explainable AI and ethical HR systems research [24]–[26]. Well-being initiatives that focus on psychosocial support and work-life balance may further mitigate burnout and improve retention among vulnerable employee groups [20], [21].

Beyond practical implications, the study contributes to methodological advancement by demonstrating how Explainable Artificial Intelligence can bridge the gap between complex analytics and managerial decision-making. By transforming opaque model predictions into transparent explanations, SHAP-based insights enhance trust, accountability, and interpretability in HR analytics, consistent with prior research emphasizing the role of XAI in ethical and transparent HR Decision-Support Systems [25]–[27]. This integration enables organizations to move beyond descriptive segmentation toward an actionable understanding of *why* certain employee groups experience systematically different workplace conditions. As such, the proposed framework extends earlier SHAP-based HR studies that focused exclusively on supervised outcomes by applying explainability to latent cluster formation [20]–[22], [30], [31].

Despite these contributions, several limitations should be acknowledged. First, the analysis is based on cross-sectional survey data, which limits the ability to infer causal relationships or track changes in employee perceptions over time, a constraint also noted in prior clustering and HR analytics studies [6], [17], [19]. Second, the reliance on self-reported survey responses may introduce response bias or reflect temporary emotional states, as highlighted in earlier well-being and perception research [11], [12]. Third, while the Finnish Working Life Barometer provides high-quality and nationally representative data, the findings may reflect context-specific institutional and cultural characteristics, potentially limiting generalizability to other labor markets, as observed in previous country-specific HR analytics research [18], [19]. Finally, the use of predefined survey items constrains the analysis to structured perceptions and does not capture emergent themes that may arise in unstructured employee feedback [27].

Future research can build on this work in several directions. Longitudinal studies could examine transitions between perception clusters across career stages, enabling analysis of how workplace experiences evolve over time and how interventions influence these trajectories, addressing limitations identified in prior cross-sectional HR analytics research [6], [17]. Cross-national applications of the proposed framework would allow comparative analysis of institutional effects on employee perceptions and test the generalizability of the findings [18], [19]. In addition, integrating unstructured data sources such as employee comments, exit interviews, or organizational communications using natural language processing could enrich perception profiling and align with recent advances in explainable and knowledge-driven HR systems [27], [30]. Methodologically, future studies may explore alternative clustering techniques, causal inference approaches, or fairness-aware explainable models to further enhance robustness, transparency, and ethical accountability in Human Resource Analytics [24]–[26], [31]. Together, these extensions would strengthen the framework's role as a comprehensive decision-support tool for Human Resource Analytics.

VI. Conclusion

The study demonstrates that integrating unsupervised clustering with supervised classification and SHAP-based explainability provides actionable insights into employee workplace perceptions. Two distinct perceptual profiles emerged: one reflecting positive evaluations of fairness, leadership, well-being, and motivation, and another characterized by more negative sentiment. By interpreting these latent profiles through supervised modeling and SHAP, the analysis transparently identified the demographic and occupational factors driving cluster membership.

Key predictors included employment duration, age, industry affiliation, gender, and socioeconomic status, underscoring the combined influence of structural and individual factors on workplace experience. Methodologically, the framework bridges descriptive segmentation and explanatory modeling, producing results that are both robust and interpretable for decision-makers. This transparency supports targeted, equitable HR interventions and directly contributes to the United Nations Sustainable Development Goal 8 by promoting fair and inclusive employment conditions.

Declaration on Generative AI

Generative AI tools (specifically large language models) were used solely to assist with drafting, clarifying wording, and refining the language of this manuscript. All conceptual development, methodological design, data analysis, interpretation of results, and final decisions regarding the content were fully conducted by the author. The author takes full responsibility for the integrity, accuracy, and originality of all scientific and analytical components of the work.

References

[1] H. Jo and D. Shin, "The impact of recognition, fairness, and leadership on employee outcomes: A large-scale multi-group analysis," PLOS ONE, vol. 20, no. 1, 2025.

[2] G. Binyamin, "Perceptions matter: Do divergent views of management and colleagues on social climate drive employee engagement through organizational identification and growth satisfaction?," Sage Open, vol. 14, no. 1, 2024.

[3] Y. Liu, Y. Fang, L. Hu, N. Chen, X. Li, and Y. Cai, "Inclusive leadership and employee workplace well-being: The role of vigor and supervisor developmental feedback," BMC Psychology, vol. 12, no. 1, 2024.

[4] S. Akter, K. Biswas, D. Vrontis, S. C. L. Cooper, and S. Y. Tarba, "Mastering digital transformation in workforce management," Production Planning & Control, vol. 35, no. 13, pp. 1525–1532, 2023.

[5] J. B. Carnevale and I. Hatak, "Employee adjustment and well-being in the era of COVID-19: Implications for human resource management," Journal of Business Research, vol. 116, pp. 183–187, 2020.

[6] A. Suvarna, "AI-enabled predictive analytics for employee well-being: A cluster-based approach to proactive HR strategies in the digital era," ICTACT Journal on Management Studies, vol. 11, no. 2, pp. 2124–2129, 2025.

[7] Ș.-A. Catană, S.-G. Toma, C. Imbrișcă, and M. Burcea, "Teleworking impact on wellbeing and productivity: A cluster analysis of the Romanian graduate employees," Frontiers in Psychology, vol. 13, 2022.

[8] C. Bosle, J. E. Fischer, and R. M. Herr, "Creating a measure to operationalize engaged well-being at work," Journal of Occupational Medicine and Toxicology, vol. 16, no. 1, 2021.

[9] F. Niebuhr, G. M. Steckhan, and S. Voelter-Mahlknecht, "New work poses new challenges—The importance of work design competencies revealed in cluster analysis," International Journal of Environmental Research and Public Health, vol. 19, no. 21, 2022.

[10] A. Anttila, M. Nuutinen, R.-L. Leskelä, M. van Gils, A. Pekki, and R. Sauni, "Cluster analysis reveals subgroups with different risk profiles and sickness absence patterns in an occupational health cohort," Journal of Occupational Rehabilitation, 2025.

[11] M. A. Naveed and M. A. Anwar, "Information anxiety in the workplace: Scale development and validation," Information Development, vol. 38, no. 3, pp. 406–423, 2021, 2021.

[12] M. Kim and M. J. Park, "Effect of good meeting practices and interactional justice on online workplace informal learning in Korea," Information Development, vol. 38, no. 4, pp. 609–622, 2021.

[13] R. M. Cabrera et al., "Impact of machine learning in human resource management: Towards the modernization of leadership," Journal of Positive School Psychology, vol. 6, no. 2S, pp. 290–299, 2022.

[14] M. R. Shafie, H. Khosravi, S. Farhadpour, S. Das, and I. Ahmed, "A cluster-based human resources analytics for predicting employee turnover using optimized artificial neural networks and data augmentation," Decision Analytics Journal, vol. 11, 2024.

[15] S. R. K. Indarapu et al., "Exploring human resource management intelligence practices using machine learning models," The Journal of High Technology Management Research, vol. 34, no. 2, 2023.

[16] K. Chahuán-Jiménez et al., "Cluster analysis of digital competencies among professors in higher education," Frontiers in Education, vol. 10, 2025.

[17] K. A. Brown et al., "An unsupervised cluster analysis of post-COVID-19 mental health outcomes and associated comorbidities," AMIA Annual Symposium Proceedings, 2022, pp. 289–298, 2023.

[18] K. Kauppi et al., "Utility of an online well-being assessment in targeting employee well-being programmes: A cross-sectional survey study in Finland," BMJ Open, vol. 14, no. 6, 2024.

[19] J. Bughin et al., "The great divide between employees: Clustering employee 'well-being' during a pandemic," PLOS ONE, vol. 20, no. 3, 2025.

[20] D. K. Shaik, D. S. G. S. Rao, A. M. Babu, D. D. V. L. Reddy, and A. A. Ahmed, "Predictive modeling for HR decision-making: A study of employee turnover," Journal of Marketing and Social Research, vol. 2, no. 4, pp. 92–98, 2025.

[21] G. Marín Díaz, J. J. Galán Hernández, and J. L. Galdón Salvador, "Analyzing employee attrition using explainable AI for strategic HR decision-making," Mathematics, vol. 11, no. 22, 2023.

[22] L. Gadár and J. Abonyi, "Explainable prediction of node labels in multilayer networks: A case study of turnover prediction in organizations," Scientific Reports, vol. 14, no. 1, 2024.

[23] A. Abonamah, D. La Torre, M. Poulin, and M. Repetto, "Explainable artificial intelligence in human resources: A computational study," in 2022 International Conference on Data Analytics for Business and Industry (ICDABI), pp. 1–6, IEEE, 2022.

[24] K. K. K. K et al., "Integrating machine learning in HR: Ethical dilemmas and balancing efficiency with fairness," in 2024 International Conference on Advances in Computing, Communication and Materials (ICACCM), pp. 1–6, IEEE, 2024.

[25] M. T. Nuseir et al., "Role of explainable artificial intelligence (EAI) in human resource management system (HRMS)," in Studies in Big Data, Springer, 2024, pp. 249–263, 2024.

[26] W. Rodgers, J. M. Murray, A. Stefanidis, W. Y. Degbey, and S. Y. Tarba, "An artificial intelligence algorithmic approach to ethical decision-making in human resource management processes," Human Resource Management Review, vol. 33, no. 1, 2023.

[27] W. Shu, S. Pang, and M. Chen, "Achieving structured knowledge management with a novel online group decision support system," Information Development, vol. 38, no. 1, pp. 23–39, 2020.

[28] P. Yang and J. Li, "Construction and application of intelligent human resource management system based on machine learning algorithm," Journal of Computational Methods in Sciences and Engineering, vol. 25, no. 4, pp. 3684–3696, 2025.

[29] A. Heidemann, S. M. Hülter, and M. Tekieli, "Machine learning with real-world HR data: Mitigating the trade-off between predictive performance and transparency," The International Journal of Human Resource Management, vol. 35, no. 14, pp. 2343–2366, 2024.

[30] M. Seo, H. Lee, and S. Kim, "Using machine learning in HRD research: Applications to advance AI-driven organizational analytics," Human Resource Development International, pp. 1–24, 2025.

[31] N. Pourkhodabakhsh, M. M. Mamoudan, and A. Bozorgi-Amiri, "Effective machine learning, meta-heuristic algorithms and multi-criteria decision making to minimizing human resource turnover," Applied Intelligence, vol. 53, no. 12, pp. 16309–16331, 2022.

[32] Statistics Finland and Ministry of Economic Affairs and Employment, Finnish Working Life Barometer 2022, dataset, version 1.0, Finnish Social Science Data Archive, 2023.

APPENDIX A

Variable names and coding schemes were derived from the FSD3784 Finnish Working Life Barometer 2022 dataset. The dataset is distributed by the Finnish Social Science Data Archive (FSD) and is used here under the permitted license for academic research.

			Mean Survey Responses Per Cluster	
ID	Survey Question	Response	Cluster 0	Cluster 1
K21c_4	Information is shared openly at my workplace. Are you:	1- Totally agree 2- Somewhat agree 3- Somewhat disagree 4- Completely disagree 9- Can't say	1.41	2.27
K21c_5	At my workplace, employees are treated equally. Are you:		1.29	2.21
K21c_10	My workplace is able to handle and resolve conflicts. Are you:		1.33	2.25
K21c_12	My workplace has a confidential atmosphere. Are you:		1.24	2.21
K44C	My supervisor gives me feedback on how well I have done at work.		1.48	2.42
K44E	My superior asks for my opinion on decisions that affect me.		1.21	2.17
K44H	My superior is interested in my well-being at work.		1.3	2.34
K44G	My supervisor treats employees fairly and equally.		1.16	2.23

K44I	My supervisor encourages me to develop my work, for example, to adapt the content and working methods of my work to suit me.		1.4	2.41
K56a	How often do you feel mentally exhausted at work:	1- You never know 2- Rarely 3- Sometimes 4- Often 5- You always feel 8- Don't want to answer 9- Can't say	2.3	2.94
K56b	How often do you feel that you are not interested or enthusiastic about your work:		2.08	2.84
K56c	How often do you feel that you cannot concentrate well when working:		2.35	2.79
K56d	How often do you feel that you cannot control your emotions at work?		1.59	2.06
K57a	How often do you feel full of energy when doing your work:		3.71	3.16
K57b	How often are you enthusiastic about your work:		3.9	3.29
K57c	How often are you completely immersed in your work:		3.65	3.28
K57d	How often do you have a sense of community and working together:	1- Never 2- Rarely 3- Sometimes 4- Often 5- Always 8- Don't want to answer 9- Can't say	3.84	3.17
K55	Stress refers to a situation where a person feels tense, restless, nervous or anxious, or has difficulty sleeping due to constant worries. Do you currently feel this kind of stress:	1- Not at all 2- Just a little 3- Some 4- Quite a lot 5- Very much 8- Don't want to answer 9- Can't say	2.21	2.81
K69	How meaningful do you find your work? Is it:	1- Very meaningful 2- Quite significant 3- Not particularly relevant 4- Not relevant at all 8- Don't want to answer 9- Can't say	1.55	1.96
K28a	If you think about the physical demands of your current job, is your work capacity:	1- Very good 2- Fair 3- Moderate 4- Pretty bad 5- Very bad 8- Don't want to answer 9- Can't say	1.52	1.93
K28b	What about mental requirements? Is your work capacity:		1.61	2.17
K64	When you think about balancing work and other life, are your working hours flexible:	1- Sufficiently 2- Somewhat, but not enough 3- Not enough at all 9- Can't say	1.17	1.53
K42a	How often do you work to tight schedules or at a very fast pace:	1- Daily 2- Weekly 3- Monthly 4- Randomly 5- Not at all 9- Can't say	2.38	2.11
K43g	Have you worked outside of working hours without compensation in the past 12 months to perform your job duties:		4.37	4.17

APPENDIX B

IndustryDetailed	0	Agriculture, forestry, fisheries
	1	Mining and quarrying
	2	Industry
	3	Electricity, gas and heating supply, refrigeration business
	4	Water supply, sewerage and wastewater management, waste management and other environmental sanitation
	5	Construction
	6	Wholesale and retail trade; repair of motor vehicles and motorcycles
	7	Transportation and storage
	8	Accommodation and catering activities
	9	Information and communication
	10	Financial and insurance activities
	11	Real estate activities
	12	Professional, scientific and technical activities

	13	Administrative and support services
	14	Public administration and national defense; compulsory social insurance
	15	Education
	16	Health and social services
	17	Arts, Entertainment and Recreation
	18	Other service activities
	19	Activities of households as employers; undifferentiated activities of households for the production of goods and services
	20	Activities of international organizations and institutions
	21	Industry unknown
Gender	0	Man
	1	Woman
SocioeconomicStatus	0	Senior officials
	1	Lower-level employees
	2	Workforce
EmployerType	0	State
	1	Municipality/joint municipal authority
	2	Private
Occupation	0	Troop
	1	Management
	2	Special experts
	3	Cognoscenti
	4	Office and customer service workers
	5	Service and sales employees
	6	Farmers, forest workers, etc.
	7	Construction, repair and manufacturing workers
	8	Process and transportation workers
IndustryRough	0	Industrial/applied/excavated
	1	Individual services
	2	Government services
	3	Municipal services
EmploymentPermanence	0	Continuous (indefinitely valid)
	1	Fixed-term or temporary
FullPartTime	0	Full-time job
	1	Part-time work
SocioeconomicClass	0	Employee or unknown position
	1	Lower-level employee
	2	Senior employee
SideJob	0	Yes
	1	No/was away all week/unknown
TemporaryWorker	0	Yes
	1	No