Employee Productivity Analytics and Prediction: A Dynamic Productivity and Burnout Modelling Engine (DPBME) for Privacy-Aware Workforce Analytics

Lakmali Karunarathne[1], Swathi Ganessan[2], Nalinda Somasiri[3], Kavindu Karunarathne[4]

[1]Lecturer, [2]Deputy Associate Dean, [3]Associate Dean, [4]Research Assistant, Department of Computer Science and Data Science, York St John University, London Campus, London, United Kingdom.

E-mail: [1]l.karunarathne@yorksj.ac.uk, [2]s.ganesan@yorksj.ac.uk, [3]n.somasiri@yorksj.ac.uk,

[4]k.karunarathne@yorksj.ac.uk

Abstract

Digital collaboration tools and project management systems provide continuous data about workforce behavior and activities. Conventional workforce analytics are mainly concerned with historical performance and turnover of employees, with no predictive abilities with respect to future productivity or signs of burnout. In this paper, we propose the Dynamic Productivity and Burnout Modelling Engine (DPBME) that combines the productivity model, designation cohort decline index, and burnout risk score while maintaining explainability and privacy principles. Tested on the dataset of 35,000 employee records, Ridge Regression, Random Forest, and Gradient Boosting Machine (GBM) algorithms are evaluated, with GBM being the most effective one (R^2 = 0.9072). The obtained burnout risk score corresponds well to the real Burn Rate (Pearson r = 0.6562). The sensitivity analysis demonstrates stable parameters' values for the risk of burnout, while the SHAP analysis revealed the importance of mental fatigue and workload as predictors of productivity. The results are presented associatively and cover limitations, governance, and privacy aspects in accordance with the GDPR Article 22 standards.

Keywords: Workforce Analytics, Burnout Prediction, Productivity Modelling, Explainable AI, SHAP, GDPR Compliance, Predictive HR Analytics.

DOI: https://doi.org/10.36548/rrrj.2026.2.007
Received: 12.07.2026, received in revised form: 14.08.2026, accepted: 03.09.2026, published: 15.09.2026

1. Introduction

There are ongoing transformations of organizations in terms of measuring and evaluating employee performance using technology. Digital collaboration solutions, task management systems, communication tools, time-tracking software and cloud computing project management environments are essential parts of organizational structure. They provide continuous flows of structured and unstructured data regarding how employees spend their time, how they communicate with their colleagues and how far they move on with their work. In this regard, the HRM and workforce planning are able to use advanced analytical methods, including machine learning.

Organizations that adopt real-time workforce analytics show higher employee productivity, increased transparency in operations, and quick response from management [1]. Human Resource departments often use predictive modeling for making workforce decisions, especially predicting employee turnover. The stacking of logistic regression with random forest has been successful in predicting turnover risks [2], machine learning has been applied to predict employee attrition rate from HR data [3], and deep learning has been suggested to improve retention using predictive HR analytics [4]. However, the majority of existing systems for work analytics operate retrospectively by producing reports on attrition based on historical data. While knowledge of attrition is important, the productivity decline and other signs of burnout become evident long before the individual leaves the organisation [5, 6]. This means that companies which discover the productivity decline at an early stage can reassign duties and take care of their employees.

Current models of workforce analytics are mostly limited to single output, either predicting employee turnover or short-term performance measurement but not multiple correlated outputs [2], [3]. They usually work with static views of the data but not dynamic, although productivity depends on workload, communication, engagement and psychological stress dynamically. Another issue is architectural one. Workforce data comes from different types of sources such as semi-structured communication log, structured task completion rate or workload data. Integrating this heterogenous data streams into one pipeline for predictions is both a methodological and engineering problem. Big data approaches have been used in the area of people analytics [4], but the approach linking heterogeneous behavioral data into a predictive model has not been developed yet.

The use of predictive analytics in workforce management creates issues of transparency, fairness, and ethics. Decisions made by managers now depend on machine learning results that should be explainable to ensure AI's responsible use, but modern algorithms lack the balance between accuracy and explainability [5], [6]. In the absence of privacy-preserving explainability, it is impossible to build trust in such technologies among employees. This research study addresses the research gap with the introduction of the Dynamic Productivity and Burnout Modelling Engine (DPBME). It represents a conceptual and computational model for the prediction of productivity as a continuous value with an explainable algorithm and calculation of a burnout risk score from it, with embedding of the pipeline in a governance layer compliant with the GDPR principles.

2. Literature Review

Workforce analytics has come a long way since its roots as a reporting function in human resources. Early HR measurements used administrative measures like head count, absenteeism, turnover ratio, and performance appraisals, which had a backward focus and did not contribute much towards strategic planning. Marler Boudreau have described these first-generation HR measurement systems as compliance-driven and process-oriented, lacking business relevance [11]. HR Information System brought in the concept of computerized records of employees, but even then, HR analytics tools that were developed were primarily reporting and diagnostic in nature. Rasmussen Ulrich describe these systems as reporting add-ons that would help managers monitor processes and did little for strategy [12]. The emergence of people analytics through systematic use of statistical modelling, data mining and machine learning techniques brought a new definition to the use of HR data in organizations [13].

2.1 Predictive HR Analytics and Attrition Modelling

The use of machine learning in HRM has advanced greatly, with employee attrition prediction being the most frequent use case. Traditional predictive HR models applied Logistic Regression to predict attrition based on demographics, tenure, performance metrics, and compensation [11]. As the workforce datasets became more complicated and numerous, tree-based algorithms, such as Random Forest and XGBoost Gradient Boosting, started to perform better than the previous statistical algorithms in predicting turnover [3], [15]. The Random Forest algorithm not only predicts well on the unseen data but also captures complex

interactions between variables [3], while XGBoost is very popular due to its scalability in dealing with structured tabular HR data [14].More recently, deep learning approaches have been adopted to predict HR outcomes. Artificial Neural Networks and multilayer perceptrons identify non-linear associations between workforce attributes [17], performing especially well when integrating disparate forms of information like behavior monitoring and engagement metrics [7]. In exchange for increased predictive accuracy, there is inevitably decreased transparency, leading to challenges in communicating model predictions for HR practitioners.

2.2 Productivity Analytics Research

Productivity analytics studies have evolved in parallel with availability of digital data about businesses, focusing on evaluation of performance and operational efficiency instead of turnover rates. Some examples of particular modelling techniques used in this area of research include a combined CNN-GRU model for forecasting productivity through analysis of patterns of workstation usage including typing and mouse movements [8], data analytics workflow that comprises data pre-processing, feature selection, and algorithm tuning for optimising the performance appraisal of employees [9], and business analytics that combines machine learning with operational data to provide guidance on allocation of workforce resources [10]. Workforce monitoring tools utilise operational data and behavioural analytics to increase transparency of an organisation and help manage its activities [1]. Performance-optimisation tools using predictive analytics have also been applied to field work and remote working contexts, combining machine learning with operational data [5].

KPI-based dashboards remain the dominant form of productivity analytics in practice, typically reporting task completion rates, deadline adherence, attendance, overtime hours, and communication volume. Marler and Boudreau note that most HR analytics programmes function as operational reporting tools rather than strategic modelling systems [11], and Tursunbayeva et al. find that many people-analytics projects produce visualisation tools with limited ability to anticipate performance decline [16]. Task-completion and performance-metrics research measures output through resolved tickets, delivery rates, and feature-completion counts in software engineering and other knowledge-intensive settings. Studies of data-driven performance-enhancement systems show that machine learning can detect performance anomalies and optimise workload distribution [6], [7], but these approaches typically treat productivity as

an aggregated performance metric rather than a composite behavioural function, and generally omit qualitative factors such as collaboration quality and mental effort.

The analysis of time tracking and work pattern is another strand. Digital time tracking records the way and time during which workers perform, and the combination of too much working time with non-standard work patterns is connected to decreased performance sustainability and increased risks of burnout [4]. At the same time, the data on time tracking is used by organizations mainly for monitoring and not for modeling.

2.3 Burnout and Well-Being Prediction Research

Employee burnout is a well-established predictor of depression, absenteeism, lack of motivation and turnover, and it has been empirically researched for quite some time [19]. Burnout measurement has traditionally been done by using validated self-report questionnaires, with the Maslach Burnout Inventory (MBI) being one of the most commonly used and recognized measures of symptom intensity on standardized subscales [19]. Survey methods are widely accepted for their validity but suffer from biases, are conducted in irregular intervals and are not appropriate for early warning applications since burnout occurs gradually due to increasing demands. The Job Demands-Resources (JD-R) theory is the prevailing theoretical explanation for burnout in the occupational health literature, which holds that burnout escalates with the presence of an unfavorable balance between job demands (workload, time pressure, and emotional demands) and job resources (control, social support, and opportunities for recovery) [18].

The machine learning techniques have been applied to make predictions about burnout and associated stress-related outcomes based on clinical, system, and behavior information. The supervised classification techniques such as Random Forest, Gradient Boosting, and Support Vector Machines have been used to identify the presence of burnout risks using workload and psychosocial metrics [6]. This is supported by studies on machine learning for predicting depression, stress, and overall well-being [20].

2.4 Governance, Privacy, and GDPR in Workforce Analytics

Algorithmic management has raised governance needs in workforce analytics. Workforce analytics has an impact on the individual assessments, workloads and career of employees, hence leading to higher requirements for justification, transparency, and fairness. Research has therefore shifted from focusing on the possibility of conducting workforce analytics to

its organisational and social consequences acknowledging that system design and not just its accuracy is important [21]. Under regulation, workforce analytics mostly deals with personal data collected through various monitoring applications such as email metadata, collaboration logs, time recording and task management logs. All the principles of GDPR including legality, fairness and transparency, purpose limitation, data minimization, accuracy, storage limitation, integrity and confidentiality, and accountability apply to both descriptive and predictive analytics. Therefore, there is need to clearly state purposes of processing data, justify their necessity and proportionality, minimize data processing, store the data for limited periods, and ensure data protection.

These ethical considerations go beyond legal compliance, as AI-assisted systems can affect performance management, recruitment, compensation, and disciplinary actions. Ethical AI models in HRM value such principles as understandability, explainability, and fairness, encouraging their inclusion into the system design process instead of treating predictive accuracy as the only goal [22]. Critical research also points to such issues as surveillance, loss of trust, and behavioral changes that might happen because of the lack of control over algorithmic management [21]. As a result, Explainable Artificial Intelligence (XAI) became a key form of governance, allowing to make HR decisions transparent and contestable. The SHAP (SHapley Additive exPlanations) framework allows receiving feature-attribution values that ensure transparency and oversight in such critical situations [23]. Nonetheless, in HR, XAI is mostly used for attrition prediction, while its role in workforce well-being and productivity governance remains understudied [23].

2.5 Research Gap

This review identifies four gaps that motivate the DPBME framework. First, predictive HR analytics has relied almost exclusively on employee turnover as the outcome of interest. Logistic Regression, Random Forest, XGBoost, and deep learning models have all been applied extensively to predicting the probability that an employee leaves an organisation, but turnover is a comparatively late and severe outcome; most predictive HR systems therefore function reactively, flagging risk only after substantial performance or engagement decline has already occurred, rather than providing an early warning. Second, productivity analytics research has concentrated on real-time monitoring, KPI dashboards, task-completion metrics, and time tracking. These

tools support operational transparency but treat productivity as a static output rather than a quantity that evolves over time. Formal mathematical treatments of productivity as a composite, time-varying construct, incorporating decline or forecasting components, remain uncommon. Third, burnout and wellbeing prediction research has advanced along two largely separate tracks: self-report survey instruments, and algorithmic classification of burnout status. Comparatively little work connects continuous productivity signals to burnout modelling, in part because most available datasets are survey-based rather than derived from continuously collected behavioural or operational data.

Fourth, ethical, privacy, and governance considerations are frequently discussed in the abstract but rarely embedded within the architecture of predictive HR systems themselves. Most predictive HR models prioritise accuracy and treat data protection, transparency, access control, and auditability as implementation details to be addressed separately, rather than as first-class design constraints.

Table 1. Systematic comparison of closely related studies against DPBME framework components

Study	Continuous productivity modelling	Decline / early-warning signal	Burnout risk scoring	Explainability (XAI)	Privacy / governance by design
Marler & Boudreau [11]	No	No	No	No	No
Talebi et al. [15]	No	Partial	No	Varies	No
Tursunbayeva et al. [16]	No	No	No	No	Partial

Demerouti et al. [18]; Bakker & Demerouti [19]	No	Theoretical only	Theoretical	No	No
Giermindl et al. [21]	No	No	No	No	Critique only
Marín Díaz et al. [23]	No	No	No	Yes	No
DPBME (this study)	Yes	Yes (cohort proxy)	Yes	Yes	Yes

As Table 1 shows, no reviewed study combines more than two of the five dimensions, and none combines continuous productivity modelling with an early-warning decline signal, a probabilistic burnout score, explainability, and embedded governance simultaneously; the closest individual matches are Marín Díaz et al. [23] on explainability alone and Demerouti et al. [18] and Bakker and Demerouti [19] on a purely theoretical account of burnout risk, neither of which is operationalised as a predictive, explainable pipeline.

3. Methodology

3.1 Research Paradigm

This study uses Design Science Research (DSR) as its methodology to develop and evaluate a new prediction-based workforce analytics framework, the Dynamic Productivity and Burnout Modelling Engine (DPBME). DSR is well suited to research that produces and assesses novel artefacts, including conceptual frameworks, mathematical models, and system architectures, that address real-world organisational problems while contributing to theoretical understanding. The DPBME artefact combines mathematical modelling, machine learning, and governance structures into a single architecture for privacy-aware productivity and burnout prediction.

3.2 Dataset

This study uses the publicly available Employee Burnout dataset (Blurredmachine, 2020; DOI: 10.34740/KAGGLE/DS/949779), collected from a technology and services-sector organisation. The combined dataset comprises 35,000 employee records: a labelled training set (n = 22,750) and an unlabelled test set (n = 12,250) constructed by the dataset's originators for competition scoring, which structurally never contains a Burn Rate column and is therefore excluded entirely from this supervised analysis. Within the training partition, 1,124 records (4.94%) additionally carry a missing Burn Rate value; combined across all 35,000 records this yields the often-quoted figure of 13,374 (38.21%) missing values, but that figure is dominated by the test partition's structural absence of labels rather than genuine non-response, and the rate relevant to this study is the 4.94% within the training partition. Because Burn Rate is the supervision target rather than a predictor, these 1,124 records were excluded rather than imputed: imputing a regression target would substitute a statistically derived value for the exact quantity the pipeline is trained to predict and later validated against, a circularity that does not arise when imputing predictor variables. After exclusion, 21,626 labelled records remain and form the analytic sample used throughout this paper.

Exclusion is safe only if missingness is unrelated to the excluded records' characteristics; this was checked directly within the training partition, where mean Designation, Resource Allocation, Mental Fatigue Score, and tenure did not differ significantly between records with and without an observed Burn Rate value (t-tests, all $p > 0.35$), and missingness rate did not differ significantly by Gender, Company Type, or WFH availability (chi-square tests, all $p > 0.07$), supporting the retained sample as reasonably representative rather than systematically biased. Each record contains nine attributes: an anonymised Employee ID, Date of Joining, Gender, Company Type (Service/Product), WFH Setup Available (Yes/No), Designation (0–5, mean = 2.18), Resource Allocation (1–10, mean = 4.47), Mental Fatigue Score (0–10, mean = 5.73), and Burn Rate (0–1, mean = 0.452, std = 0.198). Resource Allocation and Mental Fatigue Score, both predictors rather than the target, contain 3.95% and 6.05% missing values respectively and were median-imputed, since filling a missing input feature carries none of the circularity involved in filling the outcome itself.

3.3 Feature Engineering

The raw dataset attributes were transformed into five model variables grounded in the DPBME architecture. The Workload Index, W_t, normalises Resource Allocation onto [0,1] as W_t = Resource Allocation / 10, with a population mean of 0.448. The Communication Density proxy, C_t, encodes WFH Setup Available as a binary indicator (1 = Yes, 0 = No), with a mean of 0.540, indicating that approximately 54% of employees had remote-work access. The Engagement Consistency term, E_t, is derived from Designation, normalised to [0,1] as E_t = Designation / 5, with a mean of 0.436. The Stress Proxy, S_t, normalises Mental Fatigue Score as S_t = Mental Fatigue Score / 10, with a mean of 0.573, the highest population mean of the five variables, indicating that the average employee in this sample reports above-moderate fatigue. The Overtime Ratio, O_t, is derived as the positive difference between workload and engagement, O_t = clip(W_t E_t, 0, 1), capturing the extent to which an employee's workload exceeds their designation-implied engagement capacity; its population mean is low (0.049), reflecting that W_t and E_t are closely matched across this population.

The Engagement Consistency measure, E_t, is obtained from the variable Designation because positions of higher designation imply higher levels of continuous engagement, longer-term responsibilities, and decreased task discretion. The relationship is assumed not for the reason that higher designation implies higher productivity but, according to the Job Demands-Resources framework [18], for the reason that higher designation implies higher total job demands that may lead to higher risk of burnout and decreased productivity if resources do not increase simultaneously. The given dataset demonstrates a clear empirical relationship in precisely the same direction, i.e., higher designation implies lower productivity.

Three interaction terms were included to account for synergistic joint effects on the basis of occupation stress model. The term $W_t \times S_t$ considers the compounding effect of high workload and high stress which is in accordance with the demand-resource imbalance in COR theory (Hobfoll, 1989). The $E_t \times C_t$ interaction term is based on the assumption that communication may moderate the association between engagement and other variables. The term $O_t \times S_t$ corresponds to the compounding effect of overtime and stress. Tenure was computed using Date of Joining and the resulting normalized variable is denoted as Tenure_norm. Binary variables were considered for Gender (Male = 1), Company Type (Product = 1), and WFH availability

(Yes = 1), leading to a design matrix of 12 variables. Missing data was observed for W_t (n = 1,278, 5.91%), S_t (n = 1,945, 8.99%), and their interaction-based variables (n = 3,036, 14.04%). Column-wise median imputation with the help of scikit-learn's SimpleImputer was used to impute missing values, keeping all 21,626 labelled records. Productivity was considered as P_t = 1 Burn Rate (mean = 0.548, [0, 1]).

3.4 Productivity Model

The productivity model estimates the correlation between the engineered features and the productivity index P_t. Regression models of linear and non-linear types have been examined to see if productivity can be predicted better with non-linear interactions among the features than with linear models only. The model was built on the standardized matrix of the 12 engineered features. The dataset was divided into a training subset of 17,300 observations (80%) and a separate test subset of 4,326 observations (20%) using the same random seed (random_state = 42). In addition, the models were checked for their ability to generalize using 5-fold cross-validation on the training subset based on R^2 measure; since the target feature is numeric, regular K-fold cross-validation was used instead of class-stratified cross-validation.

Linear Ridge Regression acted as the linear baseline model, which was the extension of linear regression using L2 regularization (alpha = 1.0) to prevent overfitting because of the moderate multicollinearity that existed in W_t, S_t, and the interaction effect. Random Forest Regression acted as the non-parametric baseline model, in which the average value from 200 bootstrap-aggregated decision trees (random_state = 42) was computed for reducing the variance of the predictions compared to a single decision tree while still maintaining low bias. Gradient Boosting Machine (GBM) was considered the main candidate model, in which an ensemble model of 200 trees was constructed additively, and each new tree aimed at reducing the residuals of the previous trees through function space gradient descent; the sequential structure of the GBM would enable it to model non-linear interaction terms that were theoretically relevant in the interactions between W_t and S_t.

All 12 features were scaled to have a mean of zero and variance of one by the StandardScaler method. Parameters for scaling were trained on the training set and then applied on the testing set only to prevent information leakage. Productivity feature was calculated as P_t = 1 – Burn Rate and remained the same for all three models. Model selection procedure

was based on test set R^2 score and cross-validated R^2 score to detect potential overfitting. GBM model was chosen as the best one (test R^2 = 0.9072, cross-validated R^2 = 0.9085) and was used in all subsequent steps of the pipeline, such as computing the decline index and assessing burnout risks.

3.5 Decline Model

The decline module measures how far an employee's predicted productivity falls below the expectation set by their designation-level peer group. Productivity decline is conceptually a period-over-period rate of change, D* = (P_t P_{t-1}) / P_{t-1}, where P_t and P_{t-1} denote an employee's productivity in the current and immediately preceding period. This conceptual definition cannot be estimated from the present dataset, which records a single cross-sectional observation per employee: no P_{t-1} value exists for any employee, so D* is not computable as specified.

Estimating a quantity analogous to D* therefore requires substituting an observable reference for the unobserved P_{t-1} term. This study uses the mean predicted productivity of an employee's designation-level peer group, mean_g, as that reference: employees at the same designation level occupy structurally similar roles, so, absent individual decline, their productivity would be expected to track the cohort's central tendency rather than diverge from it. A cohort's mean productivity thus functions as a cross-sectional stand-in for the counterfactual prior-period baseline a genuine panel dataset would otherwise supply. This is an approximation, not an equivalence, and only genuinely longitudinal data could test its accuracy directly; this is stated as an explicit limitation below.

Formally, the implemented decline index D_t for employee i at designation level g replaces the unobserved P term with mean_g: D_t = (P_t mean_g) / (mean_g +), where P denotes the GBM model's predicted productivity for employee i, mean_g the mean predicted productivity across employees at designation level g, and = 1 × 10 a numerical-stability constant. Resulting D_t values are clipped to [1, 1]. A D_t value of zero indicates an employee's predicted productivity exactly matches their cohort's mean; negative values indicate underperformance and positive values outperformance. Throughout the remainder of this paper, D_t refers exclusively to this operational, cohort-referenced proxy rather than the conceptual D* it approximates.

The group mean centering approach ensures a zero mean in the population by design, as evidenced empirically (D_t mean = 0.0000, n = 21,626); what becomes interesting here is the

standard deviation value, which is 0.2385, and implies that roughly 16% of workers lag behind the mean of their designation group by more than 24%, forming the major at-risk category the decline factor is concerned about. D_t gets plugged into the burnout risk equation with a negative sign; hence, decreasing values will result in higher burnout risk scores and vice versa, enabling the module to act as an early warning amplification tool where workers lagging behind their peer group expectations get flagged as high risk regardless of their stress levels.

3.6 Burnout Risk Score

The burnout risk score combines three complementary signals, overtime ratio, productivity decline, and stress proxy, into a single probability-calibrated score quantifying burnout likelihood for each employee. The burnout risk score B_t is defined as a sigmoid-weighted linear combination:

$$B_t = \sigma\left(\theta_1 \cdot O_t \;\theta_2 \cdot -D_t \;\theta_3 \cdot S_t\right)$$

where (x) = 1 / (1 + e) is the logistic sigmoid, O_t is the overtime ratio, D_t is the productivity decline index, S_t is the stress proxy, and , , are component weights set to 0.5, 0.3, and 0.2 respectively. The negation applied to D_t ensures that declining productivity relative to a designation peer group contributes positively to burnout risk, consistent with the theoretical expectation that productivity deterioration is a precursor signal of burnout onset. The sigmoid maps the unbounded weighted sum to [0,1]: a linear combination of zero corresponds to B_t = 0.50 (moderate, undifferentiated risk), with increasingly positive and negative combinations approaching B_t = 1.0 and B_t = 0.0 respectively, and the steepest sensitivity of B_t to its inputs occurring near the midpoint $\sigma'0 = 0.25$.

The three weights have been assigned theoretically based on the body of knowledge in the field of occupational health rather than using a data-driven approach to find an optimum value for each. The first parameter, O_t – the overtime ratio, got the highest weight (= 0.5), due to the established direct causation between long working hours and resource depletion in Conservation of Resources theory. The second parameter, D – the productivity decline, got the second weight (= 0.3), due to its nature as a lagging behavioural factor that indicates disengagement. The stress proxy, S, got the third weight (= 0.2): despite the fact that mental fatigue is the strongest correlation (r = 0.9445) and the most important SHAP feature in this dataset and that it gets into the productivity formula directly, it is still substantially represented.

To address the subjective origin of these weights directly, a grid-search sensitivity analysis was conducted over all non-negative weight combinations satisfying + + = 1 at a 0.05 resolution (231 combinations), scoring each combination by its Pearson correlation and RMSE against the dataset's actual Burn Rate. Table 2 summarises a representative subset of this grid. The theory-derived weights (0.50, 0.30, 0.20) achieve r = 0.6562 and RMSE = 0.1985. The single combination maximising correlation with actual Burn Rate places almost all weight on the stress proxy alone (= 0.10, = 0.00, = 0.90, r = 0.8960); this is expected given that Mental Fatigue Score is itself the strongest raw correlate of Burn Rate, but such a formula would collapse the burnout score into a near-restatement of the stress input alone, defeating the theoretical purpose of a multi-component score deliberately designed to surface overtime and decline signals not already visible in raw stress. Across the full grid, Pearson r ranged from 0.053 to 0.896 and RMSE from 0.178 to 0.247; local perturbations of ±0.1–0.2 around the chosen weights shift r within a comparatively bounded 0.51–0.78 range, indicating the theory-derived weighting is not a fragile, knife-edge choice, even though it is not the empirical optimum for correlation with Burn Rate alone.

Table 2. Sensitivity analysis of burnout weight parameters (, ,) against actual burn rate

(O)	(D)	(S)	Pearson r	RMSE	Note
0.50	0.30	0.20	0.6562	0.1985	Chosen (theory-derived)
0.60	0.30	0.10	0.5545	0.1972	
0.40	0.30	0.30	0.7305	0.2008	
0.50	0.20	0.30	0.7339	0.2040	
0.50	0.40	0.10	0.5792	0.1941	
0.40	0.40	0.20	0.6619	0.1955	
0.33	0.33	0.34	0.7510	0.2011	Equal weights
0.10	0.00	0.90	0.8960	0.2399	Max r (stress-dominated)
0.00	1.00	0.00	0.5786	0.1777	Min RMSE (decline-only)

The continuous value of B_t was divided into four ordered categories to assist the HR department. Workers with B_t values in the range of [0.00, 0.35) were considered to be in the

Low-risk category and therefore not in need of immediate action. Those with B_t in the interval [0.35, 0.60) were assigned the Moderate-risk category and needed periodic check-up on their wellbeing. The High-risk category consisted of people whose B_t values were in the range [0.60, 0.80) and entailed certain HR actions such as decreased workload, consultation for psychological issues, or flexible work options. The Critical risk group had B_t in the range of [0.80, 1.00]. These thresholds have been defined based on the empirical distribution of B_t scores and severity ranges found in the occupational psychology literature; they operate as a triage mechanism for translating a continuous score to categories, and not as a validation of any multi-class classifier, since there is no independent ground-truth burnout risk category available for classification accuracy evaluation.

The empirical evaluation of B_t relative to the actual Burn Rate from the dataset yielded an r-value of 0.6562 and RMSE of 0.1985, suggesting a moderate level of consistency between the computed measure and the concept it represents. The values of B_t ranged narrowly between 0.4509 and 0.6557, with 98.1% of all staff identified as being Moderate risk and 1.9% as High risk. None of the individuals were designated Low or Critical. This level of narrowing is a consequence of the small value of the Overtime Ratio at the population level (0.049), and increasing the range is an important issue to consider in the future development of the framework.

3.7 Experimental Setup

All modelling was implemented in Python 3.11.13 within a Jupyter Notebook environment, using pandas and NumPy for data handling; scikit-learn for regression modelling (Ridge, Random Forest, Gradient Boosting), cross-validation, feature scaling, and median imputation; SciPy for the sigmoid function and correlation statistics; the shap package for explainability; and Matplotlib and Seaborn for visualisation. All stochastic components used a fixed random seed (random_state = 42) throughout to ensure reproducibility. The pipeline runs on standard CPU hardware with no GPU requirement; a fully reproducible execution environment accompanies the DPBME implementation.

4. Experimental Results

4.1 Exploratory Data Analysis

Correlation analysis on the data reveals a very strong positive correlation between Mental Fatigue Score and Burn Rate (r = 0.9445), a very strong positive correlation between Resource Allocation and Burn Rate (r = 0.8563), and a very strong positive correlation between Designation and Burn Rate (r = 0.7376), where Designation is not a minor predictor. The median value of Burn Rate is 0.54 for employees without WFH facility as against 0.39 for those having WFH facility, which is an approximate 28% relative difference and agrees with the results of Taser et al. (2022). The type of company does not make any significant difference in terms of Burn Rate (identical median value of 0.45). The gender makes a small difference (male employees have a median value of Burn Rate of 0.50 while that of females is 0.42).

4.2 Feature Engineering Results

From the engineered set of features, we can observe the following distribution of values over the entire labeled data (n=21,626): W_t = 0.448, C_t = 0.540, E_t = 0.436, S_t = 0.573, O_t= 0.049. The small mean value of O_t (0.049) is a consequence of its definition, clip(W_t E_t, 0, 1), where W_t and E_t are similar to each other in this population, thus limiting the ability of O_t to affect the burnout formula.

4.3 Productivity Model

In Table 3, we have reported comparative performance results for the three regression models used for productivity predictions. The RMSE for Ridge Regression, which represents a linear baseline, is 0.0689, while MAE = 0.0531, R^2 = 0.8753 and CV R^2 = 0.8781, thus demonstrating that a linear model accounts for a considerable amount of variance, even though it does not cover some non-linear patterns. The improvement for Random Forest model (RMSE = 0.0632, MAE = 0.0486, R^2 = 0.8951, CV R^2 = 0.8948) is rather moderate, as expected from an ensemble method. The Gradient Boosting Machine (GBM) model has shown the best performance in all four metrics: RMSE = 0.0594, MAE = 0.0468, R^2 = 0.9072, CV R^2 = 0.9085, which means that around 90.7% of the variance in employee productivity in this particular dataset can be explained by this model. A close match of R^2 and 5-fold CV R^2 values for all three models allows us to assume that the results generalize well and there is no overfitting

to training data. That is why GBM has been selected as the primary predictive model for the following index computations.

Table 3. Productivity model comparison

Model	RMSE ↓	MAE ↓	R^2 ↑	CV R^2 (5-fold) ↑
Ridge Regression	0.0689	0.0531	0.8753	0.8781
Random Forest	0.0632	0.0486	0.8951	0.8948
GBM	0.0594	0.0468	0.9072	0.9085

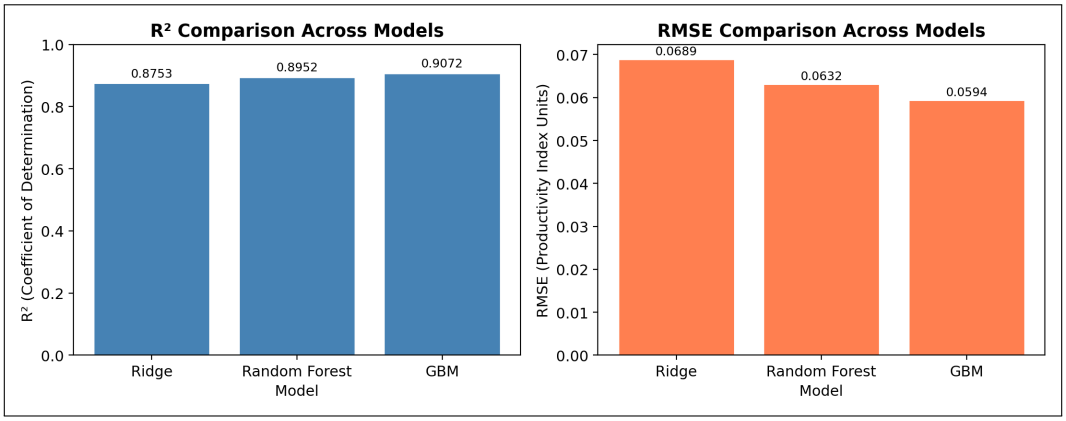

Figure 1. Productivity model comparison (R^2 and RMSE across ridge, random forest, and GBM)

4.4 Decline Model Results

Use of the decline module results in the calculation of an average of about 0.0000 for D_t with a standard deviation of 0.2385 for the entire data set. This is not really an empirical result but just follows from the definition of D_t, since it must have an average value of zero for group-mean centering; the standard deviation, on the other hand, tells us that about 16% of employees are more than 24% lower than their predicted productivity, based on designation cohort mean.

Figure 2 (left panel) plots the mean of the predicted productivity values, $\hat{P}_t$, for each of the six levels of designation. This plot exhibits an obvious monotonic downward trend where those workers belonging to Designation Level 0 are the most productive (approximately 0.85), with the productivity level declining down to the least productive level (about 0.14) for Designation Level 5, in line with the raw correlation above (Designation and Burn Rate, r = 0.7376). This

means that in this data set, the designation variable is associated with productivity in a negative sense, i.e., high designation equals low productivity and high burnout.

The histogram for the Decline Index, D_t, in Figure 2 (right panel) is normally distributed and centered around zero, as seen from the reference line provided, with most of the observations being distributed within the range [0.50, +0.50]. The near-zero centering suggests that there is no systematic and unidirectional decline in productivity of the employees as compared to their peer group; instead, the dispersion on either side of zero is indicative of real differences in relative productivity between individuals, which is what the decline module aims to capture.

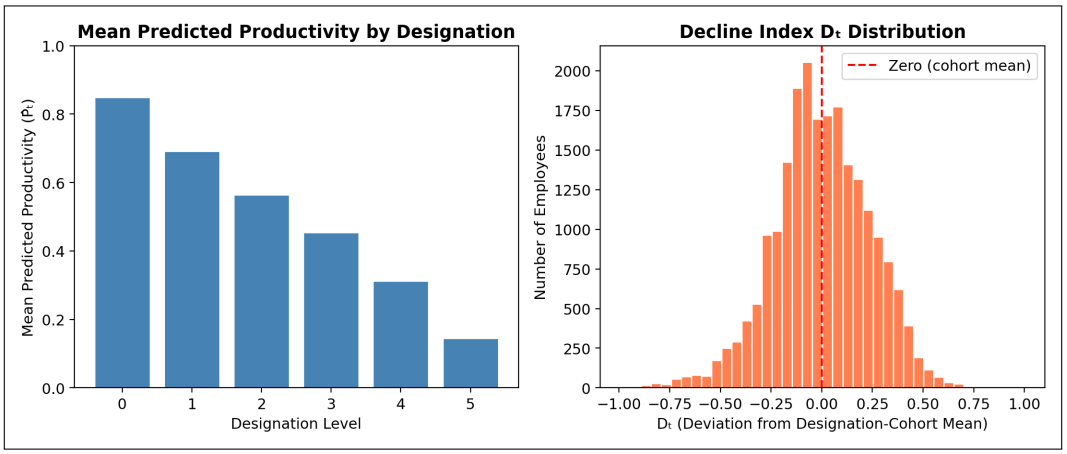

Figure 2. Mean predicted productivity by designation and decline Index (D_t) distribution

4.5 Burnout Risk Results

Figure 3 provides three different views of the burnout risk score, B_t. The left view represents the relationship between the Actual Burn Rate and B_t, where the optimal one-to-one correlation is depicted by a dashed line. There is a positive correlation between the two factors; however, the model fails to identify higher levels of burnout when the Actual Burn Rate is high, because the values of the score cluster together in a narrow range of 0.45 to 0.65. Thus, although B_t can reflect the general trend, it becomes less sensitive on the extreme ends, so the score needs to be calibrated to help identify employees at higher risk of burnout.

The centre panel shows the distribution of employees across the four burnout risk tiers. The large majority of employees (n = 21,221) fall into the Moderate tier, with a smaller group (n = 405, 1.9%) in the High tier; no employees fall into the Low or Critical tiers in this dataset. The concentration of employees in a single tier, together with the B_t compression noted above, is best read as a limitation of the current formula's dynamic range rather than as a claim that the

workforce is uniformly at moderate risk in a clinical sense.

The right panel shows the distribution of B_t scores by WFH status. Both groups show overlapping, right-skewed distributions concentrated between approximately 0.50 and 0.60, with WFH-enabled employees showing a modestly higher share of higher B_t scores in this particular projection; because this reflects the compressed B_t range rather than the underlying Burn Rate difference above, the WFH association is more reliably read from the raw Burn Rate comparison (median 0.54 vs 0.39) than from this panel alone.

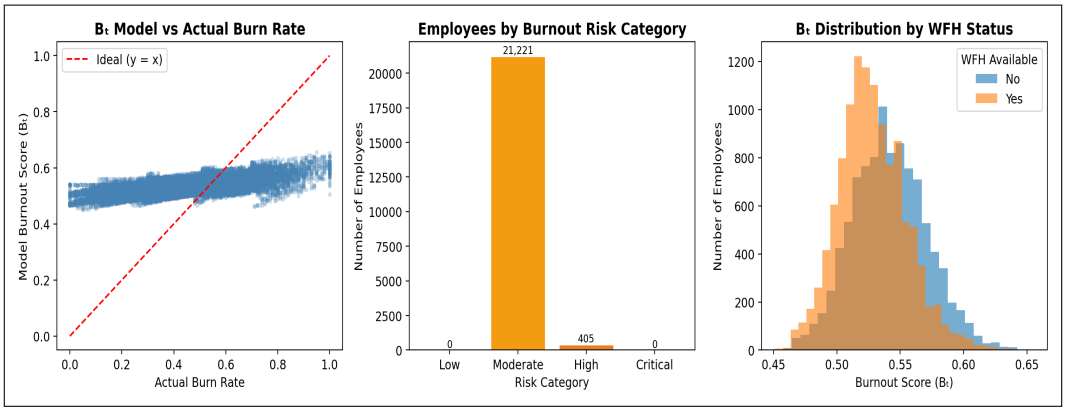

Figure 3. Burnout Score model validation, employee risk level classification, and B_t distribution by WFH status

4.6 SHAP Explainability

Figure 4 shows the SHAP beeswarm summary for the GBM productivity model, with the horizontal axis indicating each feature's SHAP value (its signed contribution to the predicted productivity P_t for a given employee) and colour indicating the underlying feature value, from low (blue) to high (red).

Mean absolute SHAP values establish a clear importance ranking. The stress proxy, S_t, dominates (mean |SHAP| 0.116), more than three times the influence of the next feature, workload W_t (0.034). The $W_t \times S_t$ interaction and the engagement index E_t each contribute a mean |SHAP| of approximately 0.010, indicating a modest non-additive compounding effect between workload and stress, which supports retaining the interaction terms in the feature set. Overtime ratio O_t, WFH encoding, communication density C_t, the $O_t \times S_t$ interaction, Gender, Tenure, the $E_t \times C_t$ interaction, and Company Type each contribute mean |SHAP| values below 0.005, indicating limited independent predictive signal beyond what S_t, W_t, and their interaction already capture.

The beeswarm's directional pattern is consistent with this ranking. High S_t (high mental fatigue) is associated with strongly negative SHAP contributions, pulling predicted productivity down; low S_t is associated with strongly positive contributions. High W_t similarly produces negative contributions, consistent with heavier workload suppressing predicted productivity. Low values of the WFH and communication-density features are associated with negative SHAP contributions, consistent with reduced remote-work access being associated with lower predicted productivity, matching the WFH association above. Company Type, $E_t \times C_t$, Tenure, Gender, and $O_t \times S_t$show SHAP patterns close to zero, suggesting these variables could be dropped from a future, more parsimonious version of the model without a material loss of predictive quality.

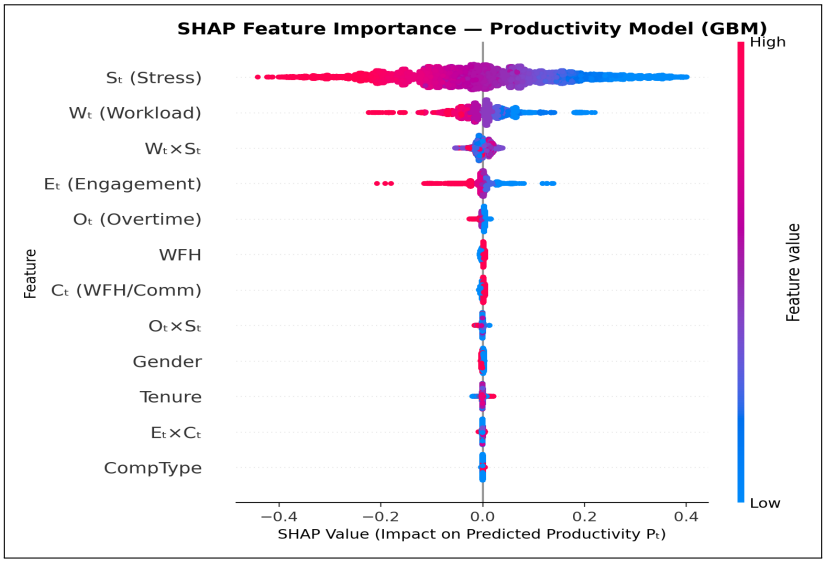

Figure 4. SHAP feature importance — productivity model (GBM)

At the individual level, SHAP's force-plot representation decomposes a single prediction into the specific feature contributions that push it above or below the model's baseline expected output. This is directly usable within the DPBME decision-support layer: it allows an HR practitioner to see which specific mix of workload, stress, and engagement factors is associated with a given employee's elevated burnout risk, rather than only a population-level risk statistic, supporting the individualised explanation envisaged under GDPR Article 22.

5. Discussion

5.1 Key Findings and Implications

The strong held-out fit of the GBM productivity model confirms the key premise of the framework that productivity is an explainable, continuous phenomenon, but this is the model fit on this particular population only and does not provide any insight about workplace productivity prediction in general. The stress proxy S_t being the strongest raw correlation predictor of burnout as well as the most prominent SHAP feature confirms that mental fatigue-related variables alone contain much explanatory power, but using them without considering the independent information provided by decline and overtime through the DPBME framework framework will be ignoring the more novel aspect of its design. The WFH connection, being based on a cross-sectional data set, should be considered purely observational – employees working from home may have other systematic differences from their colleagues, such as position or level of seniority. Finally, GBM outperforming the linear Ridge model is perfectly consistent with the theoretical expectations regarding the non-linear nature of productivity depending on workload and stress interaction.

5.2 The Accuracy-Explainability Trade-off

The DPBME framework pairs a high-performing but comparatively opaque ensemble model (GBM) with a post-hoc explainability layer (SHAP) rather than a simpler, inherently interpretable model. This design choice reflects a trade-off rather than a free improvement: the R^2 gain over Ridge Regression is achieved at the cost of relying on approximate, model-specific explanations rather than directly inspectable coefficients. SHAP's local accuracy and consistency properties mitigate this cost by guaranteeing that reported feature contributions sum correctly to each prediction and do not understate a feature's true marginal effect, but they do not make the underlying GBM model itself simpler. For HR contexts where a decision may be contested under GDPR Article 22, this trade-off is defensible provided the explanation layer, and the governance safeguards described above, are treated as integral to the deployed system, rather than as an optional add-on computed only when a dispute arises.

5.3 Limitations

Three limitations bear directly on how these results should be interpreted. First, the dataset is cross-sectional, one observation per employee, so the decline index D substitutes a designation-cohort mean for the unobserved prior-period productivity term; this proxy is reasonable but not equivalent to observing an individual's productivity trajectory over time, and only genuinely longitudinal data could confirm its accuracy. Second, the burnout risk score's four-tier discretisation functions as an interpretability aid rather than a validated classifier, since the dataset provides no independent categorical ground truth for burnout risk category, so no classification accuracy, precision, recall, or F1 metrics are reported for the tiers. Third, the burnout score's compression into a narrow numeric band limits its practical discriminative power at the individual level; the sensitivity analysis shows the weighting parameters are reasonably stable rather than fragile, but stability is not the same as optimal calibration, and the compressed range remains an open problem for the formula as specified.

5.4 Reproducibility and Generalisability

All reported results are reproducible from the accompanying implementation under the fixed random seed and environment specified during model development. Because the dataset originates from a single organisation in the technology and services sector, the specific coefficients, correlations, and SHAP rankings reported here describe this population and should not be assumed to transfer unchanged to organisations with different role structures, industries, or work arrangements; validating the DPBME framework's structure, rather than its fitted parameters, on additional datasets is a natural next step.

6. Conclusion

This paper introduced the Dynamic Productivity and Burnout Modelling Engine (DPBME), which unites productivity modelling, decline detection through designated cohorts, burnout risk score based on a theoretical model, and SHAP explainability in a privacy-aware pipeline. Being evaluated on 21,626 labelled instances drawn from a publicly available employee burnout dataset, the framework proves that digital HR data regularly collected can be used to monitor productivity and burnout with explainable models and can be used alongside traditional turnover-oriented analysis. The major drawback is the limited range of the burnout risk score

because of low population-level variance of the Overtime Ratio component and sensitivity to parameters. The key innovation of this framework lies in its architectural design, which brings together predictive modeling, early decline identification, interpretability, and privacy, fairness, and governance systems into a single pipeline. Future work needs to include validation of the decline module with longitudinal data, richer behavioral cues at the task and communication level, and tests for generalizability across firms through privacy-preserving methods such as federated learning or differential privacy.

References

[1] U. Ragavee, K. S. S. Jasmin, and A. Sundaram, "Smart Workforce: Enhancing Employee Productivity with Real-Time Data Analytics," in 2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), IEEE, Oct. 2024, 748–755.

[2] S. L. P. T, M. Gunda, E. Rishikesh, and G. Kandagatla, "Predicting Employee Layoffs with Machine Learning: A Social Network and Data Mining Approach," Global Journal of Engineering Innovations and Interdisciplinary Research, Jun. 2025, vol. 5, no. 4, 1–5.

[3] H. Alqahtani, H. Almagrabi, and A. Alharbi, "Employee Attrition Prediction using Machine Learning Models: A Review Paper," International Journal of Artificial Intelligence & Applications, Mar. 2024, vol. 15, no. 2, 23–49.

[4] R. Sharma and A. Singla, "Deep Learning in HRM: Transforming Employee Retention through Predictive Analytics," in 2024 4th Asian Conference on Innovation in Technology (ASIANCON), IEEE, Aug. 2024, 1–6.

[5] V. Chilamkurthi, B. Deka, J. Nikitha, A. Marapatla, and D. K. Babu, "Enhancing Field Employee Productivity and Performance with Android Software Solution and Machine Learning-based Predictive Analytic Model," in 2024 5th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI), IEEE, Jan. 2024, 336–343.

[6] R. Mishra, N. Tyagi, and S. Tyagi, "Evaluating Data-Driven Models to Enhance Employee Retention and Performance," in 2024 International Conference on Progressive Innovations in Intelligent Systems and Data Science (ICPIDS), IEEE, Dec. 2024, 212–218.

[7] N. Ben Yahia, J. Hlel, and R. Colomo-Palacios, "From Big Data to Deep Data to Support People Analytics for Employee Attrition Prediction," IEEE Access, 2021, vol. 9, 60447–60458.

[8] L. Singla, N. Jyani, A. Bhalwal, P. Gupta, and V. Ahuja, "Enhancing Employee Productivity Prediction: A CNN-GRU Approach to Workstation Activity Analysis," in 5th International Conference on Intelligent Technologies (CONIT), IEEE, Jun. 2025, 1–5.

[9] L. G. Tanasescu, A. Vines, A. R. Bologa, and O. Vîrgolici, "Data Analytics for Optimizing and Predicting Employee Performance," Applied Sciences, Apr. 2024, vol. 14, no. 8, 3254.

[10] M. R. Hasan, R. K. Ray, and F. R. Chowdhury, "Employee Performance Prediction: An Integrated Approach of Business Analytics and Machine Learning," Journal of Business and

Management Studies, Feb. 2024, vol. 6, no. 1, 215–219.

[11] J. H. Marler and J. W. Boudreau, "An Evidence-Based Review of HR Analytics," The International Journal of Human Resource Management, Jan. 2017, vol. 28, no. 1, 3–26.

[12] T. Rasmussen and D. Ulrich, "Learning from Practice: How HR Analytics Avoids Being a Management Fad," Organizational Dynamics, Jul. 2015, vol. 44, no. 3, 236–242.

[13] S. Bottesch, C. Schwenke, M. Förster, and M. Klier, "Driving Business Value Through People Analytics: Literature Review and Research Agenda from an Information Systems Perspective," Electronic Markets, Dec. 2025, vol. 35, no. 1, 106.

[14] S. Polisetti, M. Bhargavi, S. Chitneni, S. Eluri, N. Kattamuri, and R. R, "Stacking Models for Employee Attrition Prediction: Leveraging Logistic Regression and Random Forest," 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), IEEE, Oct. 2024, 863–867.

[15] H. Talebi, A. Khatibi Bardsiri, and V. K. Bardsiri, "Machine Learning Approaches for Predicting Employee Turnover: A Systematic Review," Engineering Reports, Aug. 2025, vol. 7, no. 8.

[16] A. Tursunbayeva, S. Di Lauro, and C. Pagliari, "People Analytics—A Scoping Review of Conceptual Boundaries and Value Propositions," International Journal of Information Management, Dec. 2018, vol. 43, 224–247.

[17] C. Maslach and S. E. Jackson, "The Measurement of Experienced Burnout," Journal of Organizational Behavior, Apr. 1981, vol. 2, no. 2, 99–113.

[18] E. Demerouti, A. B. Bakker, F. Nachreiner, and W. B. Schaufeli, "The Job Demands-Resources Model of Burnout," Journal of Applied Psychology, Jun. 2001, vol. 86, no. 3, 499–512.

[19] A. B. Bakker and E. Demerouti, "The Job Demands-Resources Model: State of the Art," Journal of Managerial Psychology, Apr. 2007, vol. 22, no. 3, 309–328.

[20] A. B. R. Shatte, D. M. Hutchinson, and S. J. Teague, "Machine Learning in Mental Health: A Scoping Review of Methods and Applications," Psychological Medicine, Jul. 2019, vol. 49, no. 9, 1426–1448.

[21] L. M. Giermindl, F. Strich, O. Christ, U. Leicht-Deobald, and A. Redzepi, "The Dark Sides of People Analytics: Reviewing the Perils for Organisations and Employees," European Journal of Information Systems, May 2022, vol. 31, no. 3, 410–435.

[22] 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, Mar. 2023, vol. 33, no. 1, 100925.

[23] G. Marín Díaz, J. J. Galán Hernández, and J. L. Galdón Salvador, "Analyzing Employee Attrition Using Explainable AI for Strategic HR Decision-Making," Mathematics, Nov. 2023, vol. 11, no. 22, 4677.