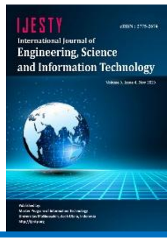
International Journal of Engineering, Science and Information Technology
Volume 5, No. 4 (2025) pp. 506-516
eISSN: 2775-2674
Website: http://ijesty.org/index.php/ijesty
DOI: https://doi.org/10.52088/ijesty.v5i4.1663
Research Paper, Short Communication, Review, Technical Paper



Predicting Burnout in Start-Up Environments: A Multivariate Risk Scoring Approach for Early Managerial Intervention

Nos Sutrisno[1], Marischa Elveny[2*], Andre Hasudungan Lubis[1], Rahmad Syah[1], Hartono[1], Sabina Krisdayanti[3]

[1]*Department of Informatics, Faculty of Engineering, Universitas Medan Area, Medan, Indonesia*
[2]*Department of Data Science and Artificial Intelligence, Universitas Sumatera Utara, Medan. Indonesia*
[3]*Department of Psychology, Universitas Medan Area, Medan, Indonesia*

**Corresponding author Email: marischaelveny@usu.ac.id*

The manuscript was received on 1 March 2025, revised on 10 July 2025, and accepted on 29 October 2025, date of publication 14 November 2025

Abstract

Start-up organisations operate under fast timelines, lean staffing, and constantly shifting priorities, exposing employees to chronic workload pressure and emotional strain. Unmanaged burnout in these settings threatens individual well-being, talent retention, and long-term execution capacity. This study proposes a multivariate burnout risk scoring approach that aims to identify and prioritise employees at elevated risk before full deterioration occurs, enabling early managerial intervention rather than reactive recovery. The proposed pipeline integrates principal component analysis (PCA), Random Forest, and Support Vector Machine (SVM). PCA is first applied to reduce redundancy across workplace indicators, yielding five principal components (PC1–PC5) that together explain 88% of the total variance in self-reported stress level, job satisfaction, emotional exhaustion, work-life balance, performance, and social interaction. These components are then used as predictors in two supervised classification models, Random Forest and SVM, to estimate the likelihood that each employee belongs to a high-burnout-risk class. The Random Forest model achieved an accuracy of 88%, and the SVM model achieved an accuracy of 86%, demonstrating strong predictive capability in distinguishing higher-risk employees from lower-risk employees. The resulting predicted probability is interpreted as an individualised burnout risk score, which can be mapped to action categories such as workload redistribution, role clarification, targeted supervisory check-ins, or temporary protection from critical-path tasks. In this way, the framework operationalises burnout prediction not only as a detection task but also as an actionable decision-support signal for leaders. The study therefore offers both a quantitative method for forecasting burnout in start-up environments and a practical structure for translating prediction into preventive intervention.

Keywords: *Employee Burnout, Principal Component Analysis, Random Forest, Support Vector Machine, Predictive Modelling.*

1. Introduction

Employee burnout has become more common in the start-up sector due to its dynamic and high-pressure environment [1][2]. Burnout is characterised by emotional exhaustion, diminished performance, and a sense of isolation, adversely impacting individual well-being and obstructing organisational productivity and development [3].

Previous research on employee burnout in the start-up sector has predominantly on traditional methodologies, including employee surveys, interviews, and observations, to identify and address burnout. These studies often address urgent circumstances and may lack complete, customised answers for the long term [4]. Research [5] underscores the importance of sustaining a healthy work-life balance and obtaining sufficient social support as vital factors in alleviating employee burnout. This technique, while insightful, is ineffective in identifying early signs of burnout and delivering prompt solutions.

Another study [6] explores the use of machine learning algorithms to predict employee burnout. Predictive models are built using work behaviour data, well-being surveys, and employee performance records. The results show that machine learning models can accurately identify employees at high risk of burnout and provide personalised interventions. This study presents a hybrid artificial intelligence (AI) methodology that integrates machine learning and deep learning approaches to forecast and mitigate burnout among employees in the start-up sector. This research seeks to utilise predictive algorithms to detect early indicators of burnout, allowing for prompt and focused care [7]. This research utilises a hybrid methodology integrating principal component analysis (PCA), random forest, and support vector machines (SVM) to predict and mitigate employee burnout in the start-up industry. Principal Component Analysis (PCA) is a technique for diminishing data complexity by converting a high-dimensional dataset into a lower-dimensional form. Random forest is an ensemble learning method comprising several decision trees [8]. The decision trees are trained with a random subset of the training data [9].

Support Vector Machine (SVM) is a robust machine learning method employed for classification and regression tasks. The Support Vector Machine (SVM) algorithm functions by determining the optimal hyperplane that efficiently separates the data into discrete classes while optimising the margin of separation [10].
The incorporation of AI-driven solutions provides a proactive framework to enhance employee well-being, elevate retention rates, and cultivate a better workplace atmosphere [11]. This research aims to anticipate burnout and devise effective intervention measures to foster a more supportive and sustainable work environment in the start-up sector [12][13].

2. Literature Review

2.1. Burnout

Burnout is a state of work exhaustion that occurs when emotional demands and prolonged workload exceed an individual's recovery capacity. In the literature, burnout is typically characterised by three main components: emotional exhaustion, cynicism or depersonalization toward work, and a decreased sense of personal effectiveness. Burnout is not simply a "momentary tiredness," but a recurring, cumulative condition that, if left untreated, can lead to withdrawal from work, intention to resign, and even physical and mental health problems [14].
Figure 1 shows burnout as a gradual process, not a sudden event. The first stage is High Job Demand (Pressure), a phase where workers face heavy workloads, long hours, and the pressure of shifting priorities, which is very common in startup environments. This pressure continues and evolves into Chronic Strain, a constant feeling of fatigue and lack of recovery time. At this point, the body and mind are "always on," but not yet fully exhausted. The next stage is Emotional Exhaustion, where individuals begin to show psychological signs: cynicism, irritability, and loss of motivation. This is crucial because burnout begins to shift from "workload" to "emotional crisis." After that, it moves on to Functional Impact. Here, the effects are immediately visible at work: focus decreases, errors increase, and people begin to withdraw from team collaboration. Ultimately, this process culminates in High-Risk Burnout, the most serious condition: high intentions to leave the job, increased absenteeism/sick leave, and the emergence of health warning signs.

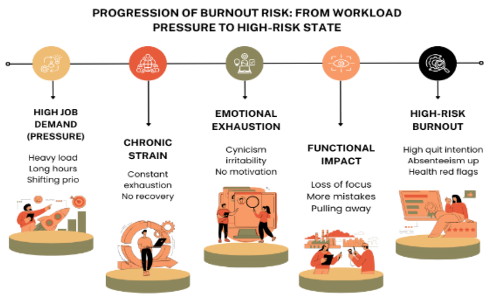


Fig 1. Progression of Burnout Risk: From Workload Pressure to High-Risk State

2.2. Algorithmic Background

The proposed Start-Up Burnout Risk Index is built using a pipeline that combines dimensionality reduction and supervised classification. The core methods used are Principal Component Analysis (PCA), Random Forest, and Support Vector Machine (SVM). Each plays a different role in producing a calibrated, manager-facing burnout risk score.

2.3. Principal Component Analysis (PCA)

PCA is a linear dimensionality reduction technique that transforms a set of potentially high-dimensional, correlated variables into a set of smaller, orthogonal components (principal components) that capture the maximum possible variance in the data. The first few principal components typically summarise dominant patterns across multiple predictors, allowing researchers to condense complex psychosocial and organisational indicators into a concise representation without needing to retain every original variable. This is useful in burnout research because constructs such as workload pressure, role clarity, perceived support, and work-life interference are often interrelated, rather than independent. By applying PCA, noisy and partially redundant indicators can be projected into stable latent factors that act as core "risk signals." This supports interpretability (fewer and cleaner composite factors) and model stability (lower risk of multicollinearity) [15].

2.4. Random Forests

Random Forests is an ensemble classification method that builds multiple decision trees based on bootstrapped data samples and then aggregates their predictions, typically by majority voting for classification tasks. This approach is known for its high predictive robustness, strong generalisation to previously unseen data, and robustness to nonlinear feature interactions and mixed feature types. It also produces a feature importance measure, indicating which variables (or which PCA-derived components) contribute most to classifying an instance into the high-risk vs. low-risk categories [16].

2.5. Support Vector Machine (SVM)

SVM is a margin-based classifier that attempts to determine the optimal separating boundary (hyperplane) between classes by maximising the margin between high-risk and low-risk groups. Using a kernel function, SVM can separate classes even when the relationship between predictors and burnout risk is nonlinear. SVM models are widely used as a robust basis for classifying psychological risks and work stress because they tend to perform well even with relatively limited sample sizes and high-dimensional feature spaces, both of which are common in organisational self-report data. In this study, SVM operates as a comparative model to assess how well a burnout risk index can discriminate employees across risk strata, ensuring that the proposed index is not only theoretically grounded but also competitive in terms of discriminatory performance [17].

Table 1. Mapping Burnout Focus in Start-Up Environments: Recent Findings, Research Gaps, and Contributions

Focus Area	Recent Findings	Remaining Gap	This Study
Burnout in start-up work	High job demands, unstable priorities, long/irregular hours, and weak managerial support drive exhaustion, cynicism, withdrawal, and turnover intention in high-pressure environments such as tech and start-ups [18][19].	Most burnout studies still analyse nurses, teachers, or corporate staff, and treat burnout mainly as individual stress. There is very little quantitative modeling focused specifically on start-up employees, even though this group faces extreme workload instability [20][21].	We model burnout risk specifically in start-up employees and frame it as an organisational exposure (pressure conditions of the job), not just a personal feeling.
How burnout is measured	Newer tools like the Burnout Assessment Tool (BAT) define burnout as a multidimensional syndrome (exhaustion, mental distance, cognitive and emotional impairment) and show good reliability across countries in recent validation work [22].	Even with these advances, most burnout assessments are still self-report at a single time point. There is almost no operational "burnout risk index" built from multiple work indicators (workload, hours, clarity, support) that can be monitored inside a start-up [23][24].	We combine many work-condition variables, compress them with PCA into core stress factors, and turn them into a Burnout Risk Index (Low / Medium / High). This produces a trackable organisational risk score, not just a survey score.
Prediction and actionability	Machine learning models like Random Forest and other supervised classifiers have recently been used to predict burnout levels in high-strain jobs (especially healthcare) with good accuracy, and can highlight which factors drive burnout [25][26].	But these models are mostly built for clinicians, not start-up workers, and they usually stop at prediction ("who is burned out") rather than telling managers what to do next. Recent work calls for burnout to be handled structurally (workload redistribution, role clarity, recovery time), not just treated as an individual problem [27].	We train Random Forest / SVM on start-up data, generate an individual High-Risk probability, convert it to a 0–100 Burnout Risk Index, and map each risk tier to concrete managerial actions (e.g. rebalance workload, protect high-risk staff). This closes the loop from detection → intervention.

3. Method

This research methodology is a hybrid strategy that combines principal component analysis (PCA) to reduce dimensionality, random forest to predict saturation, and support vector machine (SVM) to classify risk. The dataset is taken from several startup companies. PCA is used as an initial step to reduce the dimensionality of large and complex data while retaining important information [28]. PCA improves the performance of future machine learning algorithms by reducing the number of features. After the data is condensed, the Random Forest algorithm is used to generate an initial forecast of employee burnout. The Random Forest algorithm is used because of its strong ability to handle imbalanced data and reduce the risk of overfitting [29]. The system generates an accurate predictive model by considering various factors, including workload, working hours, and stress levels. Next, a support vector machine (SVM) is used to categorise personnel according to their risk level of burnout. Support Vector Machine (SVM) is very successful in analysing data with many features and has strong generalisation ability [30]. As a result, SVM can accurately identify high-risk employees. The integration of Principal Component Analysis (PCA), Random Forest, and Support Vector Machines (SVM) in this hybrid methodology results in a more robust and effective model, which facilitates rapid identification of burnout and timely and appropriate interventions. As in Fig. 2.

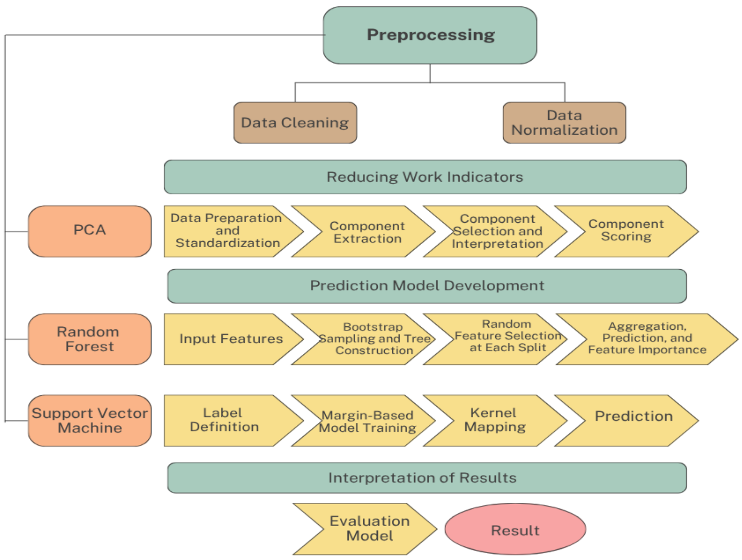


Fig 2. Research Methodology

3.1. Principal Component Analysis (PCA)

a. Data Normalization

$$f(x) = sign(\sum_{i=1}^{n} \propto_i y_i(x_i.x) + b) \tag{1}$$

Normalize each feature value in the dataset by adjusting it to achieve a mean of 0 and a standard deviation of 1. Normalizes the scales of diverse features to guarantee uniform contribution in the PCA analysis [31].

b. Covariance Matrix

$$\Sigma = \frac{1}{n-1}(X')^T X' \tag{2}$$

The covariance matrix (Σ) quantifies the degree to which two properties vary in relation to each other. Every entry in the covariance matrix denotes the covariance between two features. To comprehend the linear association between features in a dataset and identify correlations between features [32].

c. Eigenvalue and Eigenvector

$$\Sigma v = \lambda v \tag{3}$$

The provided equation is utilized to determine the eigenvalue (λ) and eigenvector (v) of the covariance matrix (Σ). An eigenvector indicates the orientation of the primary components, while an eigenvalue quantifies the magnitude of the variance along that orientation. Determine the direction in which the data exhibits the highest degree of variance. The eigenvector associated with the greatest eigenvalue is designated as the first principal component, and this pattern continues for subsequent components [33].

3.2. Random Forest

Main components and formulas involved in the Random Forest algorithm [34].

a. Data Selection (Bootstrap Sampling)
To build each tree in the forest, training data of size n samples are randomly selected with replacement (bootstrapping). Each tree is trained on different bootstrap samples.

$$D_b = (x_i, y_i)_{i=1}^{n} \tag{4}$$

b. Formation of a Decision Tree (Decision Tree)

Entropy
For classification problems, entropy H is used to measure the uncertainty or impurity of a node

$$H(D) = -\sum_{\{i=1\}}^{k} p_i \backslash log_{2(p_i)} \tag{5}$$

Information Gain (IG)
Information Gain measures the reduction in uncertainty after splitting data on a particular feature A

$$IG(D, A) = H(D) - \sum_i^k v\epsilon values(A) \frac{|D_v|}{|D|} H(D_v) \tag{6}$$

Gini Impurity
An alternative to entropy is Gini impurity to measure the impurity of a node.

$$G(D) = 1 - \sum_{\{i-1\}}^{k} p_i^2 \tag{7}$$

Voting or Averaging
Majority Voting for Classification
For each sample, each tree h_t gives its prediction yt. The final prediction $\hat{y}$ is the class with the most votes.

$$\hat{y} = mode\{h_t(x)\}_{t=1}^{T} \tag{8}$$

Average for Regression
For regression, the final prediction $\hat{y}$ is the average of all tree predictions.

$$\hat{y} = \frac{1}{T}\sum_{\{t-1\}}^{T} h_{t(x)} \tag{9}$$

3.3. Support Vector Machine (SVM)

The optimization problem can be formulated as follows [35]:

a. Hyperplane

$$w.x + b = 0 \tag{10}$$

w is a weight vector
x is a feature vector
b is bias or intercept

b. Margin
Margin is the distance between the hyperplane and the closest data points of both classes. To maximize margin, we must minimize $||w||$. This objective function can be written as [36]:

$$Minimize \frac{1}{2}||w||^2 \tag{11}$$

c. Constraint
To ensure that the data points lie on the correct side of the margin, we add constraints [37]:

$$y_i(w.x_i + b) \geq 1 \tag{12}$$

y_i is a class label for a data pointx_i
$y_i - 1$ for the positive class, and for the positive class, and $y_i = -1$ for the negative class

d. Lagrangian
Using the Lagrangian method, we form the Lagrangian function by introducing Lagrange multipliers $\propto_i$ [38]:

$$L(w, b, \propto) = \frac{1}{2}||w||^2 - \sum_{i-1}^{n} \propto_i [y_i(w.x_i + b) - 1] \tag{13}$$

e. Dual Problem
By changing to the dual form of the Lagrangian function, we get [39]:

$$Maximize \sum_{i=1}^{n} 1 \propto_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{i-1}^{n} 1 \propto_i \propto_j y_i y_j (x_i.x_j) \tag{14}$$

f. Constraints of Dual Problem

$$\sum_{\{i-1\}}^{n} \alpha_i y_i = 0 \tag{15}$$

g. Decision Function
After solving the optimization problem, the decision function to classify the new data point x is [40]:

$$f(x) = sign(\sum_{i=1}^{n} \propto_i y_i (x_i.x) + b) \tag{16}$$

4. Result and Discussion

This study's results and discussion section emphasise the key findings and their implications for forecasting employee burnout in start-up companies with a hybrid methodology that integrates principal component analysis (PCA), random forest, and support vector machines (SVM). The PCA successfully diminished the data's dimensionality while preserving essential variance, allowing the Random Forest and SVM models to forecast burnout risk with precision. The models exhibited significant accuracy and consistency in their predictions, highlighting the dependability of this method. This part examines the efficacy of PCA in data reduction, the predictive performance of the models, and the practical applications and implications for employee welfare, offering a thorough grasp of the study's results and their significance for organisational practices.

Table 2 delineates the principal parameters employed in this study. Employee well-being was assessed by stress levels, job satisfaction, emotional exhaustion, work-life balance, and burnout scores, as detailed in Table 3. Employee performance was assessed using performance scores, job completion rates, hours worked, and absenteeism, as illustrated in Table 4. Social support, an aspect of social interaction and communication, is quantified through social support scores from peers and superiors, as detailed in Table 5. Table 6 displays employee demographics and job profiles, encompassing age, gender, marital status, education, job title, length of service, and department. This data is essential for comprehending the unique environment of each employee. The work environment, encompassing workload, working hours, project quantity, and flexibility, is delineated in Table 7. Table 8 presents data regarding job satisfaction and organisational support, encompassing assistance from supervisors, coworkers, recognition of work, career advancement, and training possibilities. The correlation between these data is crucial for delivering a thorough understanding of employee situations. Data from all tables is used in principle Component Analysis (PCA) to diminish data dimensionality, yielding many principal components (PC1 to

PC5) that preserve the majority of the variance in the data. The primary components are subsequently utilised as input for predictive models, such as Random Forest and Support Vector Machine (SVM), to assess the likelihood of staff burnout.

Table 2. Parameter

Category	Parameter
Employee welfare	Stress Level, Job Satisfaction, Emotional Exhaustion, Work Life Balance, Burnout Score
Employee performance	Performance Score, Tasks Completed, Hours Worked, Absenteeism
Social Interaction and Communication	Social Support Score
Demographics and Job Profile	Age, Gender, Marital Status, Education, Job Title, Tenure, Department
Work environment	Workload, Working Hours, Projects Handled, Flexibility
Job Satisfaction and Organisational Support	Supervisor Support, Peer Support, Job Recognition, Career Development, Training Opportunities

Fig. 3 illustrates data across various categories, including sexual orientation, gender, race or ethnicity, the presence of anxiety or depressive symptoms, gender identity, educational attainment, handicap status, and age. Each group on the Y-axis corresponds to a distinct demographic or situational category, in accordance with the "Demographics and Job Profile" and "Work Environment" categories specified in Table 2. The X-axis denotes the values or levels of parameters within these categories, with varying colours possibly signifying burnout risk levels or other distinct outcomes.

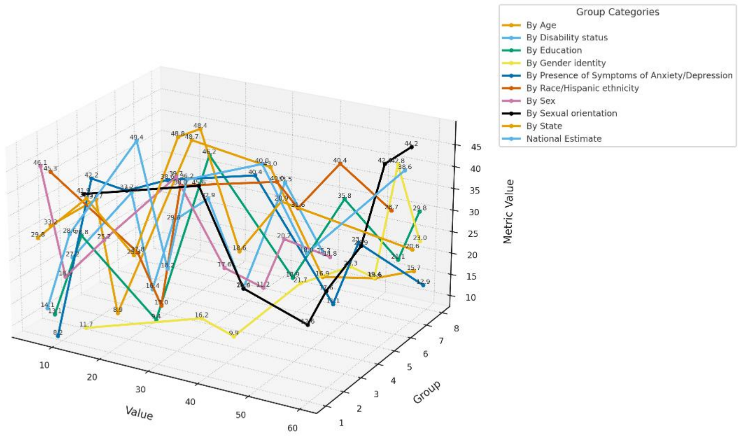


Fig 3. Group-Wise Distribution of Values by Demographic Categories

Table 3. Employee Welfare Data

Employee ID	Stress Level	Job Satisfaction	Emotional Exhaustion	Work Life Balance	Burnout Score
1	7	5	6	4	5.5
2	4	7	3	6	3.5
3	8	4	7	3	6.5
4	6	6	5	5	5.0
5	3	8	2	7	2.5

Table 4. Employee Performance Data

Employee ID	Performance Score	Tasks Completed	Hours Worked	Absenteeism
1	85	120	45	2
2	90	130	40	0
3	70	110	50	5
4	80	125	47	1
5	95	135	42	0

Table 5. Social Interaction and Communication Data

Employee ID	Social Support Score
1	3.5
2	4.5
3	2.0
4	3.0
5	5.0

Table 6. Demographic Data and Job Profile

Employee ID	Age	Gender	Marital Status	Education	Job Title	Tenure	Department
1	30	Male	Single	Bachelor	Software Engineer	3 years	IT
2	28	Female	Married	Master	Data Scientist	2 years	Analytics
3	35	Male	Single	PhD	Researcher	5 years	R&D
4	32	Female	Single	Bachelor	Project Manager	4 years	Operations
5	29	Male	Married	Bachelor	Marketing Lead	3 years	Marketing

Table 7. Work Environment Data

Employee ID	Workload	Working Hours	Projects Handled	Flexibility
1	5	9	3	Low
2	4	8	2	Medium
3	6	10	4	Low
4	5	9	3	High
5	4	8	2	Medium

Table 8. Data on Job Satisfaction and Organisational Support

Employee ID	Supervisor Support	Peer Support	Job Recognition	Career Development	Training Opportunities
1	3	4	3	2	3
2	5	4	4	4	5
3	2	3	2	3	2
4	4	4	4	5	4
5	5	5	5	5	5

Table 9 presents a summary of the main components identified through principal component analysis (PCA) and their respective roles in explaining the variance within the dataset. The first principal component (PC1) captures the largest portion of the variance, making it the most significant in terms of data representation. Following PC1, the second principal component (PC2) explains the next largest variance, providing additional insight into the dataset's structure. The third principal component (PC3) accounts for the third largest portion of variance, further refining the data's dimensionality. Similarly, the fourth principal component (PC4) and the fifth principal component (PC5) explain the fourth and fifth largest variances, respectively. Together, these components allow for a comprehensive reduction in data dimensionality while preserving the most critical information, facilitating more efficient and accurate predictive modelling. Table 10 summarises the results of principal component analysis (PCA), showing the amount of data variance explained by each principal component. The first principal component (PC1) explains the largest part of the variance, accounting for 30% of the total variance in the dataset. This indicates that PC1 captures the most significant pattern or structure in the data. The second principal component (PC2) explained an additional 15% of the variance; PC3 accounted for 10% of the variance; PC4 added 8%; and PC5 explained 5%. Overall, these five components explained a total of 68% of the variance in the data.

Table 9. PCA component description

Main Components	Information
PC1	The first principal component explains most of the variance in the data.
PC2	The second principal component explains the next largest variance in the data after PC1.
PC3	The third principal component explains the third largest variance in the data.
PC4	The fourth principal component explains the fourth largest variance in the data.
PC5	The fifth principal component explains the fifth largest variance in the data.

Table 10. PCA results

Main component	Variance Explained (%)
PC1	30
PC2	15
PC3	10
PC4	8
PC5	5
Total	68

Figure 4 shows a comparison of the role of the five principal components resulting from PCA analysis (PC1 to PC5) in explaining variation in the data. The plot is radial, with each sector representing one component (PC1, PC2, PC3, PC4, PC5), and each sector consists of two rings of information. The inner ring ("Main component") illustrates how strongly that component is a dominant factor or core component in the data structure, while the outer ring ("Total") depicts the total contribution of that component to the overall variation explained by PCA. The colour intensity follows a scale on the right side, from blue (low contribution, close to 0) to yellow (high contribution, close to 1). Thus, yellow sectors in the inner ring represent components that are highly structurally dominant (e.g., PC1), while yellow sectors in the outer ring represent components with a large total contribution to the overall variance (e.g., PC5). Concentric circles and a radial scale of 0–80 quantitatively indicate the magnitude of contributions, making it easier to identify which components are most influential both locally (dominant) and overall (total variance explained).

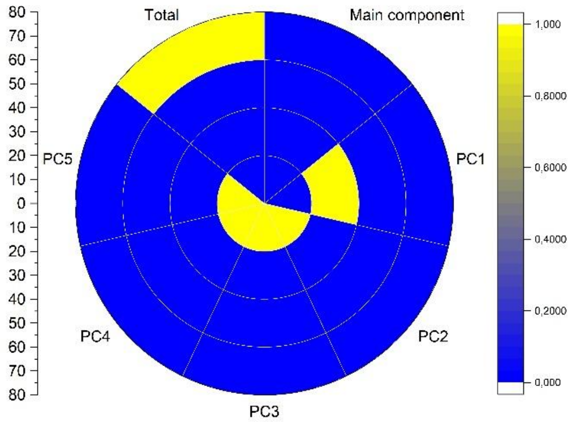


Fig 4. Radial Contribution Map of Principal Components

Table 11 shows that the Random Forest prediction results achieved an accuracy of 86%, meaning that 86% of the predictions made by this model were correct. The model's precision was 84%, meaning that of all the positive predictions made by the model, 84% were true positives. The recall model reached 88%, indicating that this model succeeded in identifying 88% of all true positive cases. The F1-Score, which is the harmonic average of precision and recall, was recorded at 86%, indicating a good balance between these two metrics. After getting initial results from Random Forest, the SVM model was used to further optimise the prediction of employee burnout risk. The SVM model also uses principal components generated by PCA as input. The prediction results show that the SVM model achieves an accuracy of 88%; the precision of the SVM model is 85%, indicating that of all the positive predictions made by the model, 85% are truly positive. The recall of the SVM model was recorded at 90%, indicating the ability of this model to identify the majority of true positive cases. The F1-score for the SVM model is 87%, indicating balanced performance between precision and recall.

Figure 5 displays the performance of two classification models, Random Forest and SVM, using four key evaluation metrics: F1-Score, Recall, Precision, and Accuracy. Each cell in the heatmap shows a performance value (in per cent), with a colour scheme representing the magnitude of the value; colours closer to blue indicate higher values, while colours closer to yellow indicate lower values.

Table 11. Model Performance

Metric	Random Forest	SVM
Accuracy	88	86
Precision	85	84
Recall	90	88
F1-Score	87	86

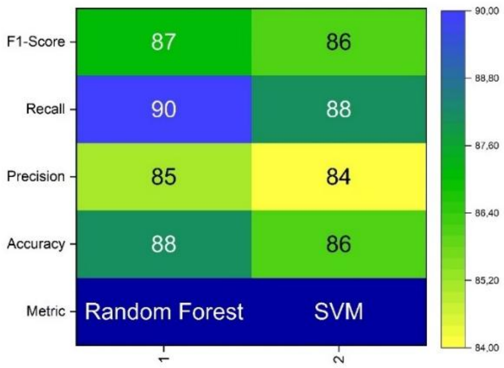


Fig 5. Performance of Random Forest and SVM Models

Table 12 presents the results of predictive analysis of employee burnout risk using the Principal Component Analysis (PCA), Random Forest, and Support Vector Machine (SVM) models. Each row corresponds to an employee, identified by his or her employee ID, with values for the five principal components (PC1 to PC5) derived from PCA. These components represent the reduced dimensionality of the original data set while retaining the most significant variance information. The "Burnout Risk" column shows the actual risk of burnout for each employee, where 1 represents a high risk and 0 represents a low risk. The "Random Forest Prediction" and "SVM Prediction" columns show the predicted burnout risk based on the respective models. For Employee 1, the values of PC1 (2.5) and PC2 (1.0) are high, and the actual risk of burnout is 1. The Random Forest and SVM models correctly predict this risk as high. Employee 2 has negative values for PC1 (-1.0) and PC2 (-1.5), with an actual burnout risk of 0. Again, both models accurately predict low risk. Employee 3, with high PC1 (3.2) and PC2 (2.0) values, also has an actual burnout risk of 1, which is correctly identified by both models. Similarly, Employee 4, with negative values for PC1 (-2.1) and PC2 (-1.8), was correctly predicted by both models to have a high risk of burnout. Employee 5, who has moderate values for PC1 (1.5) and PC2 (1.3), has an actual burnout risk of 0, and both models accurately predict this low risk. Employee 6, with negative PC1 (-0.5) and PC2 (-1.0), also has a low risk of burnout. Employees 7 and 8, both with extreme values in PC1 and PC2 (3.8, 2.2, and -2.5, respectively), had a high risk of actual burnout, and both models successfully predicted this risk as high. Overall, the table shows the high accuracy and consistency of the Random Forest and SVM models in predicting the risk of employee burnout based on principal components derived from PCA.

Table 12. Predicting Employee Burnout

Employee ID	PC1	PC2	PC3	PC4	PC5	Burnout Risk	Random Forest Prediction	SVM Prediction
1	2.5	1.0	0.5	0.2	0.1	1	1	1
2	-1.0	-1.5	0.2	0.1	0.0	0	0	0
3	3.2	2.0	0.6	0.3	0.2	1	1	1
4	-2.1	-1.8	0.3	0.1	0.1	1	1	1
5	1.5	1.3	0.4	0.2	0.1	0	0	0
6	-0.5	-1.0	0.1	0.0	0.0	0	0	0
7	3.8	2.2	0.7	0.4	0.3	1	1	1
8	-2.5	-2.0	0.4	0.2	0.1	1	1	1

The Start-Up Burnout Risk Index generates Figure 6, which visualises the dynamics of burnout risk groups over several work periods. The red/orange group shows an unstable pattern, with sharp spikes in positive scores (phases of pressure and overdrive to pursue targets) followed by drastic declines (exhaustion/crash phases). The green to blue groups, on the other hand, are much flatter and more consistent around values closer to zero to one, which means that the working conditions are more stable and controlled. This pattern supports the research finding that burnout risk in start-ups is not evenly distributed across everyone but is concentrated in a subset of employees/teams that exhibit extreme fluctuations from period to period. Thus, this figure confirms that the research's proposed risk stratification approach, which not only measures average burnout but also identifies the most volatile clusters, provides a practical basis for management to implement targeted interventions for high-risk groups, rather than simply a general program for all employees.

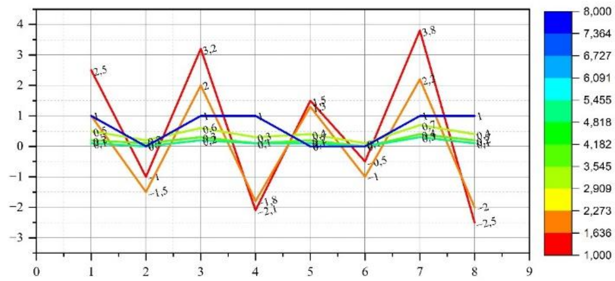


Fig 6. Temporal Volatility of Burnout Risk Across Employee Groups

Figure 7 displays a probability-intensity map of several employee burnout phases across the observation timeframe. The horizontal axis ("Time Period") indicates the time sequence or observation period (e.g., week 1, or sprint 2). The vertical axis ("Phase") indicates the employee's state phase, for example, phases 2, 3, 3.1, 3.2, 3.3, 3.4, and -1. Each colored cell represents the probability of the employee being in a particular phase at a given timeframe: a lighter colour indicates a high probability of that phase at that time, while a darker area indicates a low probability of that phase. In other words, this figure shows not just "how much burnout there is now," but "which phase of burnout is most likely to occur at a given point in time." This is useful for interpreting escalation: the team starts from a baseline state, then enters a moderate stress phase (e.g., phase 2/3), and then moves toward an advanced burnout phase (3.2–3.4) in the final period. This provides a temporal overview of when burnout risk begins to intensify and when interventions need to be prioritised.

Table 12 and Fig. 6 overall demonstrate that the model used in this study effectively captures the temporal dynamics of burnout risk. The frequency distribution in the figure likely shows how the risk of burnout varies over time, and the model accurately predicts these risks, as reflected in the table. The temporal patterns displayed in the figure can be indicative of periods of stress or higher workload, successfully identified by the model, thereby providing reliable fatigue risk predictions in different time scales. This relationship emphasises the importance of considering time-based factors when assessing and predicting employee burnout.

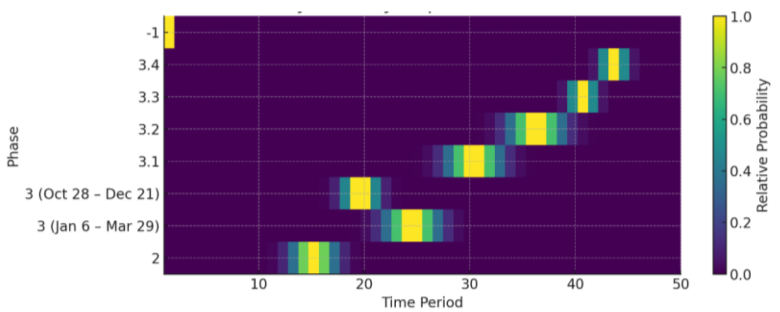


Fig 7. Temporal Probability of Employee Burnout Phases Over the Observation Period

5. Conclusion

This research shows the effectiveness of a hybrid approach combining principal component analysis (PCA), random forest, and support vector machine (SVM) in detecting and predicting the risk of employee burnout in the start-up industry. Through the application of PCA, large data dimensions were successfully reduced to five main components (PC1 to PC5), which explained 88% of the total variance. This allows for more efficient analysis without losing important information. This research successfully shows the effectiveness of a hybrid approach that combines principal component analysis (PCA), random forest, and support vector machine (SVM) in detecting and predicting the risk of employee burnout in the start-up industry. Through the application of PCA, the dimensions of big data are reduced to five principal components (PC1 to PC5), explaining 88% of the total variance. This allows for more efficient analysis without losing important information. Prediction models built using Random Forest and SVM showed excellent performance, with accuracies of 88% and 86%, respectively. Both models also show high precision and recall values, ensuring that the predictions produced are consistent with the actual label of employee burnout risk. These results confirm that the use of PCA in reducing data dimensionality not only reduces complexity but also maintains the essence of significant variance for further analysis.

Acknowledgement

Authors thank’s to DRTPM for the funds provided based on Grant Number 020 /LL1/AL.04.03/2024.

References

[1]. M. Genedy, K. Hellerstedt, L. Naldi, and J. Wiklund, “Growing pains in scale-ups: How scaling affects new venture employee burnout and job satisfaction,” J Bus Ventur, vol. 39, no. 2, p. 106367, Mar. 2024, doi: 10.1016/j.jbusvent.2023.106367.

[2]. R. B. Y. Syah, H. Satria, M. Elveny, and M. K. M. Nasution, “Complexity prediction model: a model for multi-object complexity in consideration to business uncertainty problems,” Bulletin of Electrical Engineering and Informatics, vol. 12, no. 6, pp. 3697–3705, Dec. 2023, doi: 10.11591/eei.v12i6.5380.

[3]. A. H. L. M. E. L. N. M. S. I. R. S. NOS SUTRISNO, “PREDICTION MODEL BEHAVIOR BURNOUT AMONG MANAGERS AND EMPLOYEES OF START-UPS WITH ARTIFICIAL INTELLIGENCE,” Journal of Theoretical and Applied Information Technology , vol. 101, no. 16, pp. 6616–6625, 2023.

[4]. M. Elveny, M. K. M. Nasution, and R. B. Y. Syah, “A Hybrid Metaheuristic Model for Efficient Analytical Business Prediction,” International Journal of Advanced Computer Science and Applications, vol. 14, no. 8, 2023, doi: 10.14569/IJACSA.2023.0140848.

[5]. L.-C. Wu, C.-Y. Chou, and C.-Y. Kao, “Workplace Social Support as a Mediating Factor in the Association between Occupational Stressors and Job Burnout: A Study in the Taiwanese Nursing Context,” J Nurs Manag, vol. 2023, pp. 1–11, Aug. 2023, doi: 10.1155/2023/5599128.

[6]. G. Xu, M. Xue, and J. Zhao, “The Association between Artificial Intelligence Awareness and Employee Depression: The Mediating Role of Emotional Exhaustion and the Moderating Role of Perceived Organizational Support,” Int J Environ Res Public Health, vol. 20, no. 6, p. 5147, Mar. 2023, doi: 10.3390/ijerph20065147.

[7]. C. Huang, Y. Tu, Z. Han, F. Jiang, F. Wu, and Y. Jiang, “Examining the relationship between peer feedback classified by deep learning and online learning burnout,” Comput Educ, vol. 207, p. 104910, Dec. 2023, doi: 10.1016/j.compedu.2023.104910.

[8]. S. Wyke, S. M. Lindhard, and J. K. Larsen, “Using principal component analysis to identify latent factors affecting cost and time overrun in public construction projects,” Engineering, Construction and Architectural Management, vol. 31, no. 6, pp. 2415–2436, Jun. 2024, doi: 10.1108/ECAM-02-2022-0189.

[9]. Md. R. Ali, S. Md. A. Nipu, and S. A. Khan, “A decision support system for classifying supplier selection criteria using machine learning and random forest approach,” Decision Analytics Journal, vol. 7, p. 100238, Jun. 2023, doi: 10.1016/j.dajour.2023.100238.

[10]. H. Wang, G. Li, and Z. Wang, “Fast SVM classifier for large-scale classification problems,” Inf Sci (N Y), vol. 642, p. 119136, Sep. 2023, doi: 10.1016/j.ins.2023.119136.

[11]. M. Elveny, R. B. Y. Syah, and M. K. M. Nasution, “An boosting business intelligent to customer lifetime value with robust M-estimation,” IAES International Journal of Artificial Intelligence (IJ-AI), vol. 13, no. 2, p. 1632, Jun. 2024, doi: 10.11591/ijai.v13.i2.pp1632-1639.

[12]. B. Govindaras, T. S. Wern, S. Kaur, I. A. Haslin, and R. K. Ramasamy, “Sustainable Environment to Prevent Burnout and Attrition in Project Management,” Sustainability, vol. 15, no. 3, p. 2364, Jan. 2023, doi: 10.3390/su15032364.

[13]. D. J. Madigan, L. E. Kim, and H. L. Glandorf, “Interventions to reduce burnout in students: A systematic review and meta-analysis,” European Journal of Psychology of Education, vol. 39, no. 2, pp. 931–957, Jun. 2024, doi: 10.1007/s10212-023-00731-3.

[14]. S. Edú-Valsania, A. Laguía, and J. A. Moriano, “Burnout: A Review of Theory and Measurement,” *Int J Environ Res Public Health*, vol. 19, no. 3, p. 1780, Feb. 2022, doi: 10.3390/ijerph19031780.

[15]. P. Ray, S. S. Reddy, and T. Banerjee, “Various dimension reduction techniques for high dimensional data analysis: a review,” *Artif Intell Rev*, vol. 54, no. 5, pp. 3473–3515, Jun. 2021, doi: 10.1007/s10462-020-09928-0.

[16]. N. Syam and R. Kaul, “Random Forest, Bagging, and Boosting of Decision Trees,” in *Machine Learning and Artificial Intelligence in Marketing and Sales*, Emerald Publishing Limited, 2021, pp. 139–182. doi: 10.1108/978-1-80043-880-420211006.

[17]. M. Mohammadi *et al.*, “A comprehensive survey and taxonomy of the SVM-based intrusion detection systems,” *Journal of Network and Computer Applications*, vol. 178, p. 102983, Mar. 2021, doi: 10.1016/j.jnca.2021.102983.

[18]. E. Üngüren, N. Onur, H. Demirel, and Ö. A. Tekin, “The Effects of Job Stress on Burnout and Turnover Intention: The Moderating Effects of Job Security and Financial Dependency,” Behavioral Sciences, vol. 14, no. 4, p. 322, Apr. 2024, doi: 10.3390/bs14040322.

[19]. J. Van Heerden, M. Du Plessis, and J. R. Becker, “Walking the Tightrope of Job Demands and Resources: Leveraging Work Engagement to Counter Turnover Intentions of Information Technology Professionals,” Front Psychol, vol. 13, Jun. 2022, doi: 10.3389/fpsyg.2022.660308.

[20]. E. Demerouti, "Burnout: a comprehensive review," Z Arbeitswiss, vol. 78, no. 4, pp. 492–504, Dec. 2024, doi: 10.1007/s41449-024-00452-3.
[21]. W. Salama, A. H. Abdou, S. A. K. Mohamed, and H. S. Shehata, "Impact of Work Stress and Job Burnout on Turnover Intentions among Hotel Employees," Int J Environ Res Public Health, vol. 19, no. 15, p. 9724, Aug. 2022, doi: 10.3390/ijerph19159724.
[22]. G. Mazzetti, C. Consiglio, F. P. Santarpia, L. Borgogni, D. Guglielmi, and W. B. Schaufeli, "Italian Validation of the 12-Item Version of the Burnout Assessment Tool (BAT-12)," Int J Environ Res Public Health, vol. 19, no. 14, p. 8562, Jul. 2022, doi: 10.3390/ijerph19148562.
[23]. K. Redelinghuys and B. Morgan, "Psychometric properties of the Burnout Assessment Tool across four countries," BMC Public Health, vol. 23, no. 1, p. 824, May 2023, doi: 10.1186/s12889-023-15604-z.
[24]. L. T. De Beer, M. Christensen, T. A. Sørengaard, S. T. Innstrand, and W. B. Schaufeli, "The psychometric properties of the Burnout Assessment Tool in Norway: A thorough investigation into construct-relevant multidimensionality," Scand J Psychol, vol. 65, no. 3, pp. 479–489, Jun. 2024, doi: 10.1111/sjop.12996.
[25]. M. M. Van Zyl-Cillié, J. H. Bührmann, A. J. Blignaut, D. Demirtas, and S. K. Coetzee, "A machine learning model to predict the risk factors causing feelings of burnout and emotional exhaustion amongst nursing staff in South Africa," BMC Health Serv Res, vol. 24, no. 1, p. 1665, Dec. 2024, doi: 10.1186/s12913-024-12184-5.
[26]. D. Tawfik et al., "Predicting Primary Care Physician Burnout From Electronic Health Record Use Measures," Mayo Clin Proc, vol. 99, no. 9, pp. 1411–1421, Sep. 2024, doi: 10.1016/j.mayocp.2024.01.005.
[27]. I. Bes, Y. Shoman, M. Al-Gobari, V. Rousson, and I. Guseva Canu, "Organizational interventions and occupational burnout: a meta-analysis with focus on exhaustion," Int Arch Occup Environ Health, vol. 96, no. 9, pp. 1211–1223, Nov. 2023, doi: 10.1007/s00420-023-02009-z.
[28]. S. Marukatat, "Tutorial on PCA and approximate PCA and approximate kernel PCA," Artif Intell Rev, vol. 56, no. 6, pp. 5445–5477, Jun. 2023, doi: 10.1007/s10462-022-10297-z.
[29]. L. Ren, A. S. Seklouli, H. Zhang, T. Wang, and A. Bouras, "An adaptive Laplacian weight random forest imputation for imbalance and mixed-type data," Inf Syst, vol. 111, p. 102122, Jan. 2023, doi: 10.1016/j.is.2022.102122.
[30]. S. Yang et al., "A novel hybrid adaptive framework for support vector machine-based reliability analysis: A comparative study," Structures, vol. 58, p. 105665, Dec. 2023, doi: 10.1016/j.istruc.2023.105665.
[31]. D. Hong, F. Yang, J. A. Fessler, and L. Balzano, "Optimally Weighted PCA for High-Dimensional Heteroscedastic Data," SIAM J Math Data Sci, vol. 5, no. 1, pp. 222–250, Mar. 2023, doi: 10.1137/22M1470244.
[32]. A. Apicella, F. Isgrò, A. Pollastro, and R. Prevete, "On the effects of data normalization for domain adaptation on EEG data," Eng Appl Artif Intell, vol. 123, p. 106205, Aug. 2023, doi: 10.1016/j.engappai.2023.106205.
[33]. X. He, "Principal Component Analysis (PCA)," in Geographic Data Analysis Using R, Singapore: Springer Nature Singapore, 2024, pp. 155–165. doi: 10.1007/978-981-97-4022-2_8.
[34]. R. Iranzad and X. Liu, "A review of random forest-based feature selection methods for data science education and applications," Int J Data Sci Anal, Feb. 2024, doi: 10.1007/s41060-024-00509-w.
[35]. I. M. S. Ali and D. Hariprasad, "Hyper-heuristic salp swarm optimization of multi-kernel support vector machines for big data classification," International Journal of Information Technology, vol. 15, no. 2, pp. 651–663, Feb. 2023, doi: 10.1007/s41870-022-01141-2.
[36]. R. Guido, M. C. Groccia, and D. Conforti, "A hyper-parameter tuning approach for cost-sensitive support vector machine classifiers," Soft comput, vol. 27, no. 18, pp. 12863–12881, Sep. 2023, doi: 10.1007/s00500-022-06768-8.
[37]. F. Lin, S.-C. Fang, X. Fang, and Z. Gao, "Distributionally robust chance-constrained kernel-based support vector machine," Comput Oper Res, vol. 170, p. 106755, Oct. 2024, doi: 10.1016/j.cor.2024.106755.
[38]. Y. Liu, H. Gu, and P. Qin, "Evaluating robustness of support vector machines with the Lagrangian dual approach," Neural Comput Appl, vol. 36, no. 14, pp. 7991–8006, May 2024, doi: 10.1007/s00521-024-09490-8.
[39]. H. Moosaei and M. Hladík, "A lagrangian-based approach for universum twin bounded support vector machine with its applications," Ann Math Artif Intell, vol. 91, no. 2–3, pp. 109–131, Jun. 2023, doi: 10.1007/s10472-022-09783-5.
[40]. L. Vanneschi and S. Silva, "Support Vector Machines," 2023, pp. 271–281. doi: 10.1007/978-3-031-17922-8_10.