Computer Science (CO-SCIENCE)
Volume 6 July 2 July 2026
Accreditation **Sinta 3** No. SK : 295/C/C3/KPT/2026

Comparative Analysis of Machine Learning Models for Burnout Prediction in Generation-Z

Anis Fitri Nur Masruriyah[1*], Sanggi Bayu Ardika[2], Ade Hikma Tiana[3], Budi Arif Dermawan[4]

[1,2] Department Informatics, Faculty Computer Science, Universitas Pembangunan Nasional "Veteran" Jakarta, Jakarta, Indonesia

[3,4] Department Information System, Faculty Computer Science, Universitas Pembangunan Nasional "Veteran" Jakarta, Jakarta, Indonesia

e-mail: [1] masruriyah@upnvj.ac.id, [2]sanggi.ardika@upnvj.ac.id, [3]tiana@upnvj.ac.id, [4]budiarifd@upnvj.ac.id

(*) Corresponding Author

Article Info: Received: 10-06-2026 | Revised : 24-07-2026 | Accepted : 27-07-2026

Abstracts - Burnout has become an increasingly prevalent mental health issue among Generation Z due to the interaction of psychological and behavioral factors in a highly digitalized environment. This study aims to predict burnout risk levels using a multi-class machine learning classification approach. The research follows the Cross-Industry Standard Process for Data Mining (CRISP-DM), encompassing data understanding, preprocessing, modeling, and evaluation. A synthetic dataset containing 10,000 records and 22 psychological, behavioral, and lifestyle attributes was used to classify burnout risk into three categories: low, medium, and high. To address class imbalance and ensure reliable performance estimation, Stratified K-Fold cross-validation was employed. Logistic Regression was implemented as a baseline linear model, while Random Forest represented a non-linear approach. Experimental results demonstrate that Random Forest achieved the best performance, obtaining a macro F1-score of 0.988 and outperforming Logistic Regression in multi-class burnout prediction. Feature importance analysis further revealed that psychological variables, particularly the wellbeing index and anxiety score, contributed more substantially to burnout prediction than behavioral variables such as screen time. These findings indicate that internal psychological conditions are stronger predictors of burnout risk than external digital behaviors. This study provides a comparative evaluation of linear and non-linear machine learning models in an imbalanced multi-class setting and offers an interpretable framework to support data-driven strategies for early burnout detection and mental health intervention.

Keywords : Class Imbalance; CRISP-DM; Generation-Z; Machine Learning; Mental Health

INTRODUCTION

In recent years, mental health has emerged as a critical global concern, particularly among Generation Z, who are highly exposed to digital environments and increasing academic and social pressures. The rapid growth of social media and digital technology has created a hyperconnected environment that simultaneously offers opportunities and risks for psychological well-being, including anxiety, depression, and social comparison (Fadillah & Long, 2025; Saura et al., 2025). Previous studies have shown that excessive digital engagement and social media usage are strongly associated with mental health issues such as stress, emotional fatigue, and declining well-being among young individuals (Mojtahe, 2022). These psychological challenges are further exacerbated by lifestyle factors such as reduced sleep quality, increased screen time, and diminished physical activity, which collectively increase vulnerability to burnout. Burnout itself has become a widespread phenomenon that negatively affects not only mental health but also productivity, academic performance, and overall quality of life. Recent global evidence further highlights the substantial mental health burden among younger populations. The World Health Organization reports that approximately one in seven adolescents aged 10–19 years experiences a mental health condition, while depression and anxiety remain among the leading causes of illness and disability in this age group (World Health Organization, 2025). Similarly, the American Psychological Association reported that mental health was a significant source of stress for 72% of adults aged 18–34 in its 2023 Stress in America survey (American Psychological Association, 2023). These findings reinforce the importance of early identification and preventive mental health strategies among younger populations. In addition, recent studies indicate that mental health problems among adolescents are increasing globally, emphasizing the urgency of early detection and intervention strategies (Rothenberg et al., 2023). As a result, early identification of burnout risk has become

increasingly important for preventive intervention and mental health management. In practice, early risk identification may support preliminary screening by helping educational institutions, workplaces, or mental health practitioners identify individuals who may benefit from further assessment or preventive support. However, predictive models should complement rather than replace professional psychological assessment, particularly when predictions are derived from synthetic data. However, the complexity of mental health conditions makes accurate detection challenging, as multiple psychological and behavioral factors interact dynamically.

Despite increasing awareness of mental health issues, accurately identifying individuals at risk remains a significant challenge due to the multifactorial nature of mental health conditions. Mental health problems are influenced by a combination of psychological, behavioral, social, and environmental factors, which often interact in complex and non-linear ways. Traditional assessment approaches, including self-reported questionnaires and clinical evaluations, remain essential for mental health assessment but may be constrained by response subjectivity, assessment time, and scalability when applied to large populations. In data-rich settings, conventional analyses may also have difficulty representing complex and non-linear interactions among multiple psychological and behavioral factors. Machine learning complements these approaches by enabling the simultaneous analysis of numerous predictors and by capturing complex, non-linear relationships that are difficult to identify using conventional analytical techniques (Chung & Teo, 2022, 2023). Nevertheless, machine learning prediction should be regarded as a complementary analytical tool rather than a replacement for professional assessment.

Despite the growing adoption of machine learning in mental health prediction, several methodological challenges remain insufficiently addressed. First, many previous studies formulate burnout prediction as a binary classification problem, whereas burnout severity is more naturally represented as multiple risk levels. Consequently, existing models may not adequately distinguish gradual transitions between low, medium, and high burnout risk. Second, relatively few studies systematically compare linear and non-linear machine learning algorithms under identical experimental settings, making it difficult to determine whether observed performance differences arise from algorithmic characteristics or experimental design. Third, although class imbalance frequently occurs in mental health datasets, its influence on multi-class burnout prediction has received comparatively limited attention. Finally, while synthetic datasets offer important advantages for privacy-preserving research, relatively few studies discuss their implications for model interpretability and external validity. These unresolved methodological issues motivate the comparative framework proposed in this study.

Although machine learning techniques have been widely applied in mental health prediction, several critical limitations remain unresolved. Previous studies demonstrate that machine learning models are capable of identifying patterns and predicting mental health conditions; however, many of these studies rely on limited datasets and fail to capture the full complexity of mental health dynamics (Madububambachu et al., 2024). Additionally, most existing research focuses on binary classification tasks, which do not adequately represent real-world mental health conditions that often exist across multiple levels of severity. Another important limitation is the lack of consideration for class imbalance, where certain categories are significantly underrepresented, resulting in biased models that favor majority classes. Furthermore, existing studies often lack consistency in evaluation metrics, making it difficult to compare model performance across different research contexts (Chung & Teo, 2022, 2023). These issues highlight the need for more robust and comprehensive modeling approaches that can address both data complexity and imbalance. Without addressing these challenges, machine learning models may fail to generalize effectively in real-world applications. Therefore, further investigation is required to improve both the reliability and fairness of predictive models in mental health.

In addition to data-related challenges, there is a lack of comprehensive comparative studies that evaluate different machine learning models under consistent experimental settings. Existing research has explored various algorithms, including Logistic Regression, Support Vector Machines, and Random Forest; however, no single model has been consistently identified as the best across different datasets and problem domains (Chung & Teo, 2023). This indicates that model performance is highly dependent on data characteristics, such as feature distribution, class imbalance, and the presence of non-linear relationships. Moreover, limited studies explicitly compare linear and non-linear models in multi-class classification scenarios, particularly in the context of mental health prediction. Another significant challenge lies in the interpretability of machine learning models, where complex models often act as “black-box” systems, reducing trust and usability in real-world applications (Higgins et al., 2023). Additionally, there is a lack of studies that integrate feature importance analysis to identify key contributing factors to mental health outcomes. This limitation reduces the practical value of predictive models for decision-making and intervention design. Therefore, a systematic and interpretable comparative approach is needed to address these gaps.

In order to address these challenges, this study adopts the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework to develop a systematic burnout prediction model. The proposed methodology comprises exploratory data analysis, data preprocessing, model development, and performance evaluation. Logistic Regression and Random Forest are employed to represent linear and non-linear learning approaches, while Stratified K-Fold cross-validation and class weighting are used to mitigate class imbalance. Model performance is

evaluated using precision, recall, and macro F1-score, followed by feature importance analysis to identify the key predictors of burnout risk.

This study makes four main contributions. First, it develops a reproducible multi-class burnout prediction framework based on the CRISP-DM methodology. Second, it provides a systematic comparison between linear (Logistic Regression) and non-linear (Random Forest) machine learning models under identical preprocessing and evaluation settings. Third, it investigates the effect of class imbalance using class weighting and Stratified K-Fold cross-validation. Finally, it analyzes feature importance to improve model interpretability while discussing the methodological implications and limitations associated with synthetic mental health datasets.

RESEARCH METHOD

This study adopts CRISP-DM as a structured framework connecting the research problem with the analytical workflow. In the business understanding phase, the problem is formulated as identifying different levels of burnout risk among Generation Z, which is translated analytically into a three-class classification problem. Data understanding examines feature distributions, class imbalance, correlations, and potential data-quality issues. Data preparation transforms the available numerical and categorical variables into model-compatible representations. The modeling phase compares linear and non-linear learning approaches, while the evaluation phase assesses both overall and class-specific predictive performance. These interconnected stages ensure that model development is guided by the characteristics of the data and the analytical objective rather than by predictive performance alone (Rahmawati et al., 2024; Wibowo et al., 2025).

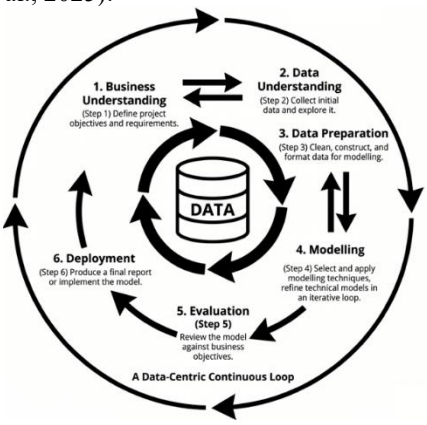

Figure 1. Stages of Crisp-DM
Source: Rahmawati et al. (2024)

1. Data Description

The dataset used in this study consists of 10,000 records with a total of 22 variables representing psychological, behavioral, and lifestyle factors related to burnout risk (Zahid, 2026). These variables include measures such as anxiety score, emotional fatigue, sleep duration, screen time, motivation level, and overall well-being index. The target variable is burnout risk, categorized into three classes: low, medium, and high. Initial analysis reveals that the dataset is highly imbalanced, with the majority of instances belonging to the medium and high categories, while the low-risk class is significantly underrepresented. This imbalance reflects real-world conditions where extreme cases are less frequently observed. The dataset contains both numerical and categorical features, requiring appropriate preprocessing techniques before modeling. Additionally, the presence of correlated variables suggests the need for careful feature analysis to avoid redundancy. It is important to note that the dataset used in this study is fully synthetic and does not represent real individuals. The use of synthetic data provides several methodological advantages, including the elimination of privacy concerns, reproducible experimentation, and controlled class distributions. However, synthetic datasets cannot fully capture the complexity, variability, and hidden biases present in real-world mental health populations. Therefore, the objective of this study is to evaluate the behavior of machine learning algorithms under controlled experimental conditions rather than to develop a clinically deployable prediction model. The psychological variables are simulated numerical indicators designed for research and modeling purposes. Therefore, the results of this study should not be interpreted as clinical diagnoses but rather as a methodological exploration of burnout prediction using machine learning techniques.

Table 1. Attributes of Data

Attributes	Detail
Age	Age of the respondent (in years), representing the demographic profile.
Gender	Gender identity of the respondent.
Country	Country of residence of the respondent.
Student_Working_Status	Indicates whether the respondent is a student, working individual, or both.
Daily_Social_Media_Hours	Average number of hours spent on social media per day.
Screen_Time_Hours	Total daily screen time across all digital devices.
Night_Scrolling_Frequency	Frequency of using social media or devices late at night.
Online_Gaming_Hours	Average number of hours spent on online gaming per day.

Attributes	Detail
Content_Type_Preference	Preferred type of digital content (e.g., educational, entertainment, social).
Exercise_Frequency_per_Week	Number of times the respondent exercises per week.
Daily_Sleep_Hours	Average number of hours of sleep per day.
Caffeine_Intake_Cups	Number of caffeinated drinks consumed per day.
Study_Work_Hours_per_Day	Average number of hours spent on study or work activities per day.
Overthinking_Score	Score representing the tendency to overthink situations.
Anxiety_Score	Score indicating the level of anxiety experienced by the respondent.
Mood_Stability_Score	Measure of emotional stability over time.
Social_Comparison_Index	Level of tendency to compare oneself with others on social platforms.
Sleep_Quality_Score	Self-reported quality of sleep.
Motivation_Level	Level of motivation in daily activities.
Emotional_Fatigue_Score	Degree of emotional exhaustion experienced.
Wellbeing_Index	Overall psychological well-being index combining multiple mental health indicators.
Burnout_Risk	Target variable indicating burnout risk categorized as Low, Medium, or High.

Source: Zahid (2026)

The dataset primarily represents psychological, behavioral, demographic, and lifestyle-related attributes. However, broader contextual factors such as family relationships, socioeconomic status, academic environment, workplace conditions, and access to mental health support are not explicitly represented. Consequently, the proposed model captures only the dimensions available in the synthetic dataset and should not be interpreted as a comprehensive representation of all determinants of burnout.

2. Data Preprocessing

Data preprocessing is a crucial step to ensure the quality and reliability of the modeling process. In this study, preprocessing includes data cleaning, transformation, and normalization (Figure 2). The dataset was first examined for missing values and duplicate records, and no missing values or duplicate records were identified. Categorical predictor variables were transformed using one-hot encoding, where each category was represented as a binary indicator variable. One reference category from each categorical attribute was omitted to prevent redundant feature representation. The target variable (Burnout_Risk) was retained as multiclass class labels throughout the modeling process. Numerical variables were standardized only for Logistic Regression to ensure comparable feature scales, whereas the tree-based Random Forest model was trained using the original feature values. Outliers were identified during data inspection but retained because no domain-specific evidence indicated that these observations represented erroneous or invalid data. Since the dataset was synthetically generated to preserve realistic behavioral variation, retaining these observations was considered more appropriate than removing potentially informative samples. However, the effect of outlier retention on predictive performance was not independently evaluated and is therefore acknowledged as a limitation of this study. In addition, class imbalance was identified as a major issue and therefore influenced both the model selection process and the evaluation strategy.

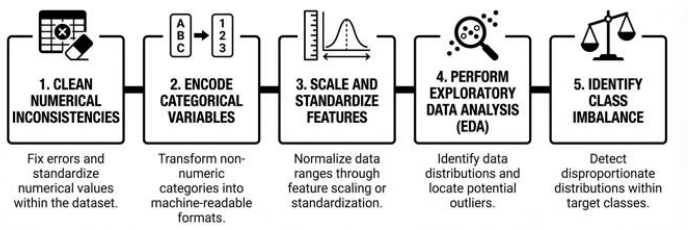

Figure 2. The Overall Preprocessing Workflow
Source: Raja et al. (2022)

3. Modeling Techniques

This study employs multiple machine learning algorithms to analyze burnout risk, including Logistic Regression and Random Forest. Logistic Regression is used as a baseline model due to its simplicity, interpretability, and strong theoretical foundation in classification tasks (Hikmayanti et al., 2023; Rahmawati et al., 2024; Wibowo et al., 2025; A. M. Widodo et al., 2021). Logistic Regression estimates the probability of each burnout category by first computing a linear combination of the predictor variables as expressed in Equation (1), where z is the linear predictor x_i denotes the i-th predictor variable, β_i represents the corresponding regression coefficient, and $\beta_0$is the intercept term. Since this study addresses a multiclass classification problem, the linear scores are subsequently transformed into class probabilities using the Softmax function (Equation 2). where K is the total number of burnout categories and $P(y = k \mid x)$ denotes the probability that an observation belongs to class k.

The predicted burnout category is then determined by selecting the class with the highest posterior probability (Equation 3).

$$z = \beta_0 + \sum_{i=1}^{n} \beta_i x_i \qquad (1)$$

$$P(y = k \mid x) = \frac{e^{z_k}}{\sum_{j=1}^{K} e^{z_j}} \quad (2)$$

$$\hat{y} = arg\ max\ P(y = k \mid x) \quad (3)$$

In contrast, Random Forest is a non-linear ensemble learning method that can capture complex interactions between features through multiple decision trees (Imani et al., 2025; Lechner & Okasa, 2025). Random Forest constructs an ensemble of decision trees, where each tree is trained using a bootstrap sample of the original dataset. The final prediction is obtained through majority voting across all trees. For an ensemble consisting of T decision trees, the final prediction is expressed as Equation 4. Where $h_t(x)$ represents the prediction generated by the t-th decision tree and mode(·) denotes the majority voting operation. Each decision tree recursively partitions the feature space by selecting the split that minimizes node impurity. The impurity of a node is measured using the Gini Indeks (Equation 5). where C denotes the number of classes and p_i is the proportion of observations belonging to class i within the node. By aggregating predictions from multiple decorrelated decision trees, Random Forest generally achieves lower variance and improved generalization compared with a single decision tree.

$$\hat{y} = mode\{h_1(x), h_2(x), \dots, h_t(x)\} \quad (4)$$

$$G = 1 - \sum_{i=1}^{C} {P_i}^2 \quad (5)$$

This study employs Logistic Regression as a baseline model and Random Forest as a non-linear model to capture complex relationships. In the implementation, Logistic Regression was configured using the 'lbfgs' solver with a maximum of 1000 iterations to ensure convergence in the multi-class setting. Random Forest was configured with n_estimators = 100, max_depth = None, and random_state = 42. These parameter settings were selected to provide a balance between model performance and computational efficiency while maintaining reproducibility. The selected hyperparameters were determined based on recommendations reported in previous literature rather than exhaustive hyperparameter optimization. This strategy was adopted to ensure fair model comparison while minimizing the influence of extensive parameter tuning. Consequently, performance differences primarily reflect the intrinsic learning capability of each algorithm instead of optimization bias.

4. Model Evaluation

In order to obtain a reliable estimate of model performance while preserving the original class distribution, this study employed five-fold Stratified K-Fold Cross-Validation (Bhagat & Bakariya, 2022; Rofik et al., 2025; S. Widodo et al., 2022). Unlike conventional random partitioning, stratified cross-validation maintains approximately the same class proportions in every fold, thereby reducing evaluation bias caused by class imbalance. Let the complete dataset D be partitioned into K mutually exclusive subsets as expressed in Equation (6). During the i-th iteration, one fold is used as the testing set while the remaining $K - 1$ folds are combined to form the training set, as defined in Equation (7). The evaluation procedure is repeated for all K folds, and the overall model performance is computed as the arithmetic mean of the evaluation scores obtained from each iteration (Equation 8). Where $Score_i$ denotes the evaluation metric obtained from the i-th validation fold. In this study, $K = 5$ was selected because it provides an effective balance between computational efficiency and the stability of performance estimation while ensuring that each observation is used for both training and validation exactly once. Consequently, the reported performance reflects the average predictive capability of the model across multiple stratified data partitions rather than the outcome of a single train-test split.

$$D = \{D_1, D_2, \dots, D_k\} \quad (6)$$

$$Train_i = D - D_i, \quad Test_i = D_i \quad (7)$$

$$Score = \frac{1}{K} \sum_{i=1}^{k} Score_i \quad (8)$$

5. Evaluation Matrix

Model performance is evaluated using multiple metrics, with a primary focus on macro-averaged F1-score (1) due to the imbalanced multi-class nature of the dataset. Unlike accuracy, which can be misleading in imbalanced scenarios, the macro F1-score provides an equal contribution from each class, ensuring fair evaluation (Bharathi, 2024; Masruriyah et al., 2023). Precision (2) and recall (3) are also calculated for each class to analyze model performance in detail, particularly for the minority class (Masruriyah et al., 2023; Nugraha et al., 2021; Sonjaya et al., 2022). Confusion matrices (Table 2) are used to visualize classification results and identify patterns of misclassification. In order to provide a structured and interpretable evaluation of classification performance in the multi-class setting, a confusion matrix is employed to capture the distribution of predictions across all classes.

Let $C_{i,j}$ denote the number of instances belonging to actual class i that are predicted as class j. In this representation, diagonal elements $C_{i,j}$ correspond to correctly classified instances, while off-diagonal elements reflect misclassification patterns between classes. This formulation enables a detailed analysis of model behavior beyond aggregate metrics, particularly in identifying class-specific errors. The confusion matrix for the burnout risk classification problem (Low, Medium, High) is presented in Table 2.

Table 2. Confusion Matrix

Actual \ *Predicted*	***Low***	***Medium***	***High***
Low	$C_{1,1}$	$C_{1,2}$	$C_{1,3}$
Medium	$C_{2,1}$	$C_{2,2}$	$C_{2,3}$
High	$C_{3,1}$	$C_{3,2}$	$C_{3,3}$

Source: Bharathi (2024)

In multi-class classification, the computation of evaluation metrics such as precision, recall, and F1-score is performed using a one-vs-all (one-vs-rest) approach. For each class k, the problem is transformed into a binary classification setting where class k is treated as the positive class and all other classes are treated as negative. Based on the confusion matrix $C_{i,j}$ the true positive (TP), false positive (FP), and false negative (FN) for class k are defined in equations (9)–(11). Furthermore, the precision, recall, and F1-score for each class can then be calculated as follows equations (12), (13) and (14).

$$TP_k = C_{k,k} \tag{9}$$

$$FP_k = \sum_{i \neq k} C_{i,k} \tag{10}$$

$$FN_k = \sum_{j \neq k} C_{k,j} \tag{11}$$

$$Precision_k = \frac{TP_k}{TP_k + FP_k} \tag{12}$$

$$Recall_k = \frac{TP_k}{TP_k + FN_k} \tag{13}$$

$$F1 - Score_k = \frac{2 \times (Precision_k \times Recall_k)}{Precision_k + Recall_k} \tag{14}$$

RESULTS AND DISCUSSION

Exploratory Data Analysis (EDA) was conducted to understand the underlying characteristics of the dataset, including feature distributions, relationships, and potential data issues. The analysis reveals that the dataset exhibits a significant class imbalance, where the majority of observations belong to the medium and high burnout categories, while the low category is substantially underrepresented (Figure 3). This imbalance simulates patterns commonly observed in real-world scenarios where extreme conditions are less frequently observed. In addition, distribution analysis shows that psychological variables such as anxiety score, emotional fatigue, and wellbeing index demonstrate clearer separation across burnout risks compared to behavioral variables.

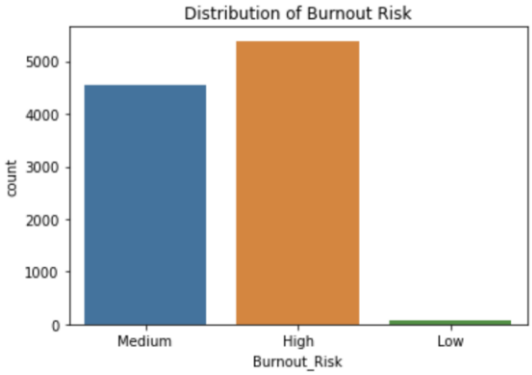

Figure 3. The Distribution of Burnout Risk
Source: Research Result (2025)

Furthermore, correlation analysis provides deeper insights into relationships between variables. As illustrated in Figure 4, several strong correlations are observed among psychological features. For instance, the wellbeing index shows a strong positive correlation with sleep quality and motivation level, while exhibiting a strong negative correlation with anxiety score and emotional fatigue. Similarly, anxiety score is positively correlated with overthinking and emotional fatigue, indicating consistent psychological patterns. In contrast, behavioral features such as screen time, social media usage, and gaming hours show relatively weak correlations with psychological indicators, suggesting that their influence on burnout may be indirect.

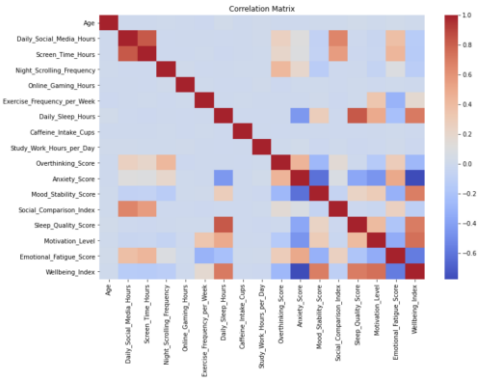

Figure 4. The Correlation Matrix
Source: Research Result (2025)

These findings indicate that burnout is more strongly associated with internal psychological conditions than external behavioral factors. Additionally, the presence of moderate correlations among several predictors suggests potential multicollinearity, which may influence model behavior, particularly for linear models. Overall, the EDA results provide important justification for the use of non-linear models capable of capturing complex interactions between variables.

This study evaluates the performance of multiple machine learning models, including Logistic Regression, Logistic Regression with class weighting, and Random Forest, using Stratified K-Fold cross-validation. The results show that Random Forest achieves the highest performance with a macro F1-score of 0.988, significantly outperforming Logistic Regression (0.919) and Logistic Regression with class weighting (0.908). The superior performance of Random Forest can be attributed to its ability to model non-linear interactions among psychological variables, which were previously observed during exploratory data analysis. Logistic Regression assumes a linear relationship between predictors and the target variable, limiting its ability to capture complex interactions among emotional, behavioral, and lifestyle indicators. This indicates that non-linear models are more effective in capturing complex relationships within the dataset. This substantial performance gap indicates that capturing non-linear feature interactions is critical in burnout prediction. The comparison between linear and non-linear models highlights the importance of selecting appropriate algorithms based on data characteristics, particularly in imbalanced multi-class scenarios.

Logistic Regression, as a linear model, still demonstrates relatively strong performance, suggesting that the dataset contains meaningful and partially separable patterns. However, its limitation lies in its inability to model non-linear interactions among features, which are evident from the correlation and distribution analysis. Interestingly, the application of class weighting does not improve performance, indicating that simple imbalance handling techniques may introduce trade-offs that reduce overall model effectiveness. These findings confirm that model selection is critical, particularly in datasets characterized by complexity and imbalance. The comparison of model performance is presented in Table 3.

The dataset used in this study is highly imbalanced, with the low burnout class representing only a small fraction of the total observations. This imbalance significantly affects the model's ability to learn minority class patterns. Although the overall macro F1-score is high, detailed analysis reveals that the model struggles to accurately classify the low-risk category.

Table 3. Comparison of Model Performance

Attributes	Detail
Logistic Regression	0.919480
Logistic Regression + Balance	0.908076
Random Forest	0.987681

Source: Research Result (2025)

This issue is evident in the confusion matrix, where several low-risk instances are misclassified as medium or high. The limited effectiveness of class weighting indicates that the minority class is not only underrepresented but also partially overlaps with neighboring classes. Consequently, merely adjusting class weights cannot sufficiently improve class separability. More advanced techniques such as synthetic resampling, ensemble balancing, or cost-sensitive learning may provide greater improvements. Such behavior is typical in imbalanced classification problems, where models tend to prioritize majority classes to minimize global error. Furthermore, the results indicate that class weighting does not sufficiently address this issue, suggesting that more advanced techniques may be required. These findings emphasize that relying solely on overall performance metrics can be misleading, and that class-sensitive evaluation is essential. The confusion matrix is shown in Figure 5.

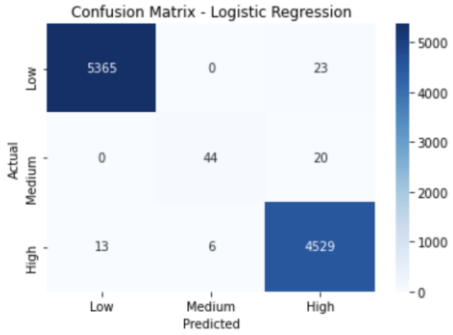

Figure 5. The Confusion Matrix
Source: Research Result (2025)

A deeper class-level analysis reveals significant performance disparities among classes. The model achieves near-perfect classification for medium and high burnout categories, with accuracy values close to 0.996. In contrast, the low burnout class achieves a significantly lower accuracy of approximately 0.688.

This discrepancy is primarily due to the limited number of low-class samples, which restricts the model's ability to learn representative patterns. As a result, the model tends to misclassify low-risk instances into adjacent categories, particularly medium. Despite this limitation, the model still demonstrates some ability to distinguish low-risk cases, indicating that relevant patterns exist within the features.

These findings highlight the importance of evaluating model fairness and robustness across all classes, rather than relying solely on aggregate performance metrics. Improving minority class representation or applying more advanced imbalance handling techniques could further enhance classification performance. The detailed classification report is presented in Table 4.

Table 4. Classification Report

	Precision	Recall	F1 Score	Support
High	1.00	1.00	1.00	5388
Low	0.88	0.69	0.77	64
Medium	0.99	1.00	0.99	4548
Accuracy			0.99	10000
Macro Avg	0.96	0.89	0.92	10000
Weighted Avg	0.96	0.99	0.99	10000

Source: Research Result (2025)

Feature importance analysis using the Random Forest model reveals that the wellbeing index is the most dominant predictor, with a significantly higher importance score compared to other features. This suggests that overall psychological well-being plays a central role in determining burnout risk.

Other important features include anxiety score, motivation level, sleep quality, and mood stability, all of which are psychological indicators. These findings are consistent with the correlation analysis observed in the EDA phase, where strong relationships were identified among these variables. In contrast, behavioral features such as screen time and social media usage show relatively low importance, indicating that they are less influential in predicting burnout directly.

This result challenges the common assumption that digital behavior is the primary driver of burnout. Nevertheless, these findings should be interpreted carefully because the feature importance values are derived from a synthetic dataset whose statistical properties are artificially generated. Although the ranking provides useful methodological insights, validation using real-world clinical or psychological datasets remains necessary before practical implementation. Instead, it suggests that internal psychological conditions are more critical determinants. The dominance of the wellbeing index also indicates that it may act as a composite variable capturing multiple aspects of mental health. The feature importance ranking is presented in Figure 6.

The findings of this study provide several important insights into burnout prediction using machine learning. First, the superior performance of Random Forest confirms the importance of modeling non-linear relationships in complex mental health data. This is consistent with EDA findings, which reveal intricate relationships among psychological variables.

Second, the limited effectiveness of class weighting suggests that handling class imbalance requires more sophisticated approaches beyond simple weighting techniques. Third, the strong influence of psychological variables over behavioral variables indicates that burnout is primarily driven by internal mental states rather than external digital behaviors.

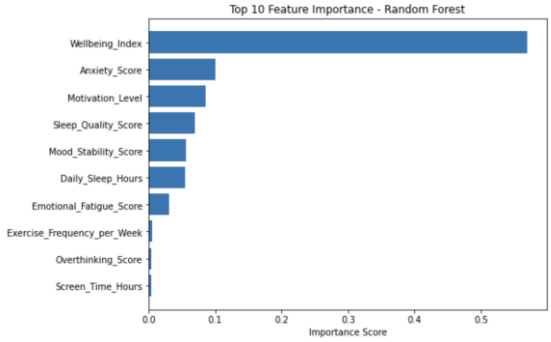

Figure 6. The Feature Importance Ranking
Source: Research Result (2025)

These findings align with existing literature emphasizing the multifactorial nature of mental health conditions. Additionally, the consistency between EDA results and feature importance analysis strengthens the validity of the findings. The integration of interpretable models and feature analysis also enhances the practical relevance of the study, particularly for early detection and intervention strategies. Overall, this study demonstrates the importance of combining data exploration, modeling, and interpretation to gain meaningful insights.

Despite the promising results, this study has several limitations. First, the dataset used is synthetic, which may not fully represent the complexity and variability of real-world mental health conditions. Therefore, the results should be interpreted as a methodological exploration rather than a clinical conclusion.

Second, the significant class imbalance affects the model's ability to accurately classify minority classes, particularly the low burnout category. Future research could explore advanced techniques such as hybrid resampling methods or ensemble strategies to address this issue.

Additionally, the presence of correlated features suggests potential redundancy, which could be further investigated using feature selection or dimensionality reduction techniques. Another important direction for future work is the validation of the proposed approach using real-world datasets to improve generalizability. Finally, incorporating explainable AI techniques could further enhance model transparency and trust in practical applications. Another limitation concerns the use of synthetic data generated under predefined statistical assumptions. Although this approach enables controlled benchmarking and reproducibility, it may not adequately represent demographic diversity, behavioral variability, and latent psychological factors encountered in real-world populations. Therefore, the reported performance should be interpreted as evidence of methodological effectiveness rather than clinical applicability.

CONCLUSION

This study proposes a data-driven approach for predicting burnout risk among Generation Z using a multi-class classification framework based on the CRISP-DM methodology. The results show that Random Forest outperforms Logistic Regression, highlighting the importance of modeling non-linear relationships in complex mental health data. The comparative evaluation confirms that non-linear models consistently outperform linear models in this context. In addition, the findings reveal that psychological factors, such as wellbeing index and anxiety score, have a stronger influence on burnout prediction than behavioral factors.

Compared with previous studies, this research demonstrates that combining multi-class classification, class imbalance handling, and feature importance analysis within a unified CRISP-DM framework provides a transparent methodology for evaluating burnout prediction models. However, because the experiments rely on synthetic data, the reported findings primarily demonstrate methodological feasibility. Future work should validate the proposed framework using real-world mental health datasets and investigate more advanced imbalance handling techniques and explainable AI approaches.

REFERENCES

American Psychological Association. (2023, November). *Stress in America 2023: A Nation Recovering from Collective Trauma*. American Psychological Association. https://www.apa.org/news/press/releases/stress/2023/collective-trauma-recovery

Bhagat, M., & Bakariya, B. (2022). Implementation of Logistic Regression on Diabetic Dataset using Train-Test-Split, K-Fold and Stratified K-Fold Approach. *National Academy Science Letters*, *45*(5). https://doi.org/10.1007/s40009-022-01131-9

Bharathi. (2024). Confusion Matrix for Multi-Class Classification. *Analytics Vidhya*.

Chung, J., & Teo, J. (2022). Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges. In *Applied Computational Intelligence and Soft Computing* (Vol. 2022). Hindawi Limited. https://doi.org/10.1155/2022/9970363

Chung, J., & Teo, J. (2023). Single classifier vs. ensemble machine learning approaches for mental health prediction. *Brain Informatics*, *10*(1). https://doi.org/10.1186/s40708-022-00180-6

Fadillah, D., & Long, B. (2025). Empowering Gen Z's Mental Health in a Hyperconnected World. *International Journal of Social Psychiatry*. https://doi.org/10.1177/00207640251384510

Higgins, O., Short, B. L., Chalup, S. K., & Wilson, R. L. (2023). Artificial intelligence (AI) and machine learning (ML) based decision support systems in mental health: An integrative review. In *International Journal of Mental Health Nursing* (Vol. 32, Number 4, pp. 966–978). John Wiley and Sons Inc. https://doi.org/10.1111/inm.13114

Hikmayanti, H., Nurmasruriyah, A. F., Fauzi, A., Nurjanah, N., & Nur Rani, A. (2023). Performance Comparison of Support Vector Machine Algorithm and Logistic Regression Algorithm. *International Journal of Artificial Intelegence Research*, *7*(1), 1. https://doi.org/10.29099/ijair.v7i1.1.1114

Imani, M., Beikmohammadi, A., & Arabnia, H. R. (2025). Comprehensive Analysis of Random Forest and XGBoost Performance with SMOTE, ADASYN, and GNUS Under Varying Imbalance Levels. *Technologies*, *13*(3). https://doi.org/10.3390/technologies13030088

Lechner, M., & Okasa, G. (2025). Random Forest estimation of the ordered choice model. *Empirical Economics*, *68*(1). https://doi.org/10.1007/s00181-024-02646-4

Madububambachu, U., Ukpebor, A., & Ihezue, U. (2024). Machine Learning Techniques to Predict Mental Health Diagnoses: A Systematic Literature Review. *Clinical Practice & Epidemiology in Mental Health*, *20*(1). https://doi.org/10.2174/0117450179315688240607052117

Masruriyah, A. F. N., Novita, H. Y., & Sukmawati, C. E. (2023). *Performance Evaluation of Popular Supervised Learning Algorithms Towards Cardiovascular Disease*. *8*(3), 420–426. https://doi.org/10.32493/informatika.v8i3.34103

Mojtahe, K. (2022). How social media affecting the mental health of Gen-Z: A mixed method study. *Revista de Psiquiatria Clinica*, *49*(1). https://doi.org/10.15761/0101-60830000000345

Nugraha, R. G., Yoga Wibowo, M., Ajie, P., Handayani, H. H., Fauzi, A., & Nur Masruriyah, A. F. (2021). Implementation of Deep Learning in Order to Detect Inapposite Mask User. *2021 6th International Conference on Informatics and Computing, ICIC 2021*, 4–9. https://doi.org/10.1109/ICIC54025.2021.9632994

Rahmawati, S., Wibowo, A., & Masruriyah, A. F. N. (2024). Improving Diabetes Prediction Accuracy in Indonesia: A Comparative Analysis of SVM, Logistic Regression, and Naive Bayes with SMOTE and ADASYN. *Jurnal RESTI*, *8*(5). https://doi.org/10.29207/resti.v8i5.5980

Raja, R., Nagwanshi, K. K., Kumar, S., & Laxmi, K. R. (2022). *Data Mining and Machine Learning Applications*.

Rofik, Unjung, J., & Prasetiyo, B. (2025). Enhancing costumer churn prediction with stacking ensemble and stratified k-fold. *Bulletin of Electrical Engineering and Informatics*, *14*(1). https://doi.org/10.11591/eei.v14i1.8112

Rothenberg, W. A., Bizzego, A., Esposito, G., Lansford, J. E., Al-Hassan, S. M., Bacchini, D., Bornstein, M. H., Chang, L., Deater-Deckard, K., Di Giunta, L., Dodge, K. A., Gurdal, S., Liu, Q., Long, Q., Oburu, P., Pastorelli, C., Skinner, A. T., Sorbring, E., Tapanya, S., … Alampay, L. P. (2023). Predicting Adolescent Mental Health Outcomes Across Cultures: A Machine Learning Approach. *Journal of Youth and Adolescence*, *52*(8), 1595–1619. https://doi.org/10.1007/s10964-023-01767-w

Saura, J. R., Gelashvili, V., & Martínez-Navalón, J. G. (2025). The Impact of Social Media on Gen Z's Mental Health and Privacy. *Journal of Competitiveness*, *17*(1), 233–252. https://doi.org/10.7441/joc.2025.01.11

Sonjaya, C. B., Masruriyah, A. F. N., Kusumaningrum, D. S., & Pratama, A. R. (2022). The Performance Comparison of Classification Algorithm in Order to Detecting Heart Disease. *INTERNAL (Information System Journal*, *5*(2), 166–175. https://doi.org/10.32627

Wibowo, A., Masruriyah, A. F. N., & Rahmawati, S. (2025). Refining Diabetes Diagnosis Models: The Impact of SMOTE on SVM, Logistic Regression, and Naïve Bayes for Imbalanced Datasets. *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, *7*(1), 197–207. https://doi.org/10.35882/jeeemi.v7i1.596

Widodo, A. M., Anggraeni, Y. S., Anwar, N., Ichwani, A., & Sekti, B. A. (2021). Performance of K-NN, J48, Naive Bayes, and Logistic Regression as Diabetes Classification Algorithms. *SISFOTEK*, *5*(1). https://www.seminar.iaii.or.id/index.php/SISFOTEK/article/view/253

Widodo, S., Brawijaya, H., & Samudi, S. (2022). Stratified K-fold cross validation optimization on machine learning for prediction. *Sinkron*, *7*(4). https://doi.org/10.33395/sinkron.v7i4.11792

World Health Organization. (2025, September 1). *Mental Health of Adolescents*. United Nations. https://www.who.int/vietnam/news/fact-sheets/detail/adolescent-mental-health

Zahid, H. (2026). *Gen Z Mental Wellness & Digital Lifestyle Patterns*. Massachusetts Institute of Technology. https://doi.org/https://doi.org/10.34740/kaggle/dsv/14927563