IJCCS (Indonesian Journal of Computing and Cybernetics Systems)
Vol.19, No.2, April 2025, pp. 117~128
ISSN (print): 1978-1520, ISSN (online): 2460-7258
DOI: 10.22146/ijccs.103551

Optimization of Gradient Boosting Method for Predicting Narcissistic Personality Disorder (NPD) in Employees

Achmad Solichin*[1], Bagas Pramudita[2], Painem Painem[3], Anindya Putri Pradiptha[4]
[1,3,4]Universitas Budi Luhur, Jakarta, Indonesia
[2]Magister Ilmu Komputer, Universitas Budi Luhur, Jakarta, Indonesia
e-mail: *[1]**achmad.solichin@budiluhur.ac.id**, [2]bagaspramudita@gmail.com, [3]painem@budiluhur.ac.id, [4]anindya.putri@budiluhur.ac.id

Abstrak

Gangguan Kepribadian Narsistik (NPD) merupakan tantangan serius di lingkungan tempat kerja modern; namun, deteksi dini dan intervensi yang tepat masih belum terpenuhi. Penelitian ini bertujuan untuk mengatasi masalah tersebut dengan mengusulkan model sistem cerdas berbasis pembelajaran mesin, dengan memanfaatkan metode Gradient Boosting untuk memprediksi NPD pada karyawan. Metode Gradient Boosting dipilih karena kemampuannya menangani data yang kompleks dan secara bertahap meningkatkan kinerja prediksi. Model ini diintegrasikan dengan data karyawan, termasuk berbagai variabel psikologis, perilaku, dan demografi yang relevan dengan NPD. Kontribusi utama penelitian ini adalah pengembangan model prediktif yang dapat membantu organisasi dalam mengidentifikasi dan memberikan intervensi dini kepada karyawan yang berisiko mengalami NPD. Dengan demikian, diharapkan dapat mengurangi dampak negatif NPD di tempat kerja, seperti konflik interpersonal dan penurunan produktivitas. Studi ini menunjukkan hasil yang signifikan dalam kinerja klasifikasi model setelah menerapkan Recursive Feature Elimination (RFE) untuk mengoptimalkan metode Gradient Boosting. Tingkat akurasi mencapai 82%, peningkatan dari 79% sebelumnya yang dicapai dengan menggunakan Gradient Boosting Classifier. Hal ini menunjukkan bahwa model RFE-Gradient Boosting memiliki potensi lebih besar dalam mengklasifikasikan karyawan yang benar-benar memiliki gangguan kepribadian narsistik dibandingkan dengan mereka yang tidak.

Kata kunci*—narcissistic personality disorder, mental health, employee, machine learning, RFE-Gradient Boosting*

Abstract

Narcissistic Personality Disorder (NPD) is a serious challenge in modern workplace environments; however, early detection and appropriate intervention remain unmet needs. This research aims to address the issue by proposing an intelligent system model based on machine learning, utilizing the Gradient Boosting method to predict NPD in employees. The Gradient Boosting method was chosen for its ability to handle complex data and gradually improve prediction performance. This model is integrated with employee data, including a range of psychological, behavioral, and demographic variables relevant to NPD. The primary contribution of this research is the development of a predictive model that can assist organizations in identifying and providing early intervention to employees at risk of developing NPD. In doing so, it is expected to reduce the negative impact of NPD on the workplace, such as interpersonal conflicts and decreased productivity. The study shows significant results in the model's classification performance after applying Recursive Feature Elimination (RFE) to optimize the Gradient Boosting method. The accuracy rate reached 82%, an improvement from the previous 79% achieved using the Gradient Boosting Classifier. This indicates that the RFE-Gradient Boosting model has greater potential in classifying employees who genuinely have narcissistic personality disorder versus those who do not.

Keywords— *narcissistic personality disorder, mental health, employee, machine learning, Recursive Feature Elimination, Gradient Boosting*

Received January 7th,2025; Revised March 20th, 2025; Accepted April 28th, 2025

1. INTRODUCTION

Narcissistic Personality Disorder (NPD) is a mental disorder characterized by excessive thoughts and behaviors regarding self-importance, a need for admiration, and a lack of empathy toward others. Key traits of narcissism include arrogant behavior, a drive for validation, and an excessive obsession with physical appearance [1]. The belief that they possess high attractiveness is part of this behavioral pattern [2]. The impact of NPD affects not only the individual experiencing it but also their social and professional environments, including the workplace. Moreover, this mental disorder can trigger violence in intimate relationships [3].

In the modern workplace, early detection of mental disorders such as NPD is crucial due to its significant negative impact, including interpersonal conflicts, decreased productivity, and disharmony within team dynamics. Concerns about NPD in the workplace, particularly in the information technology sector, are growing [4]. According to research [5], biological, psychoanalytic, and social factors play a role in the development of NPD. Additionally, as stated by [6], narcissistic disorders develop from a failure to learn empathy during childhood, primarily due to a lack of empathetic role models from parents. Other factors that contribute to NPD include social changes such as the emphasis on success, individualism, competition, and short-term hedonism. However, detecting NPD in employees is often a challenge for organizations, with obstacles such as stigma, limited understanding of mental health, and a lack of proper evaluation tools making identification and intervention processes more difficult.

Various previous studies have emphasized the importance approaches in detecting mental disorders such as NPD in the workplace. One promising approach is machine learning techniques [7], [8], which enable in-depth data analysis and the development of predictive models based on complex data patterns. The types of mental health issues discussed include depression, anxiety, and schizophrenia. Furthermore, research by [9], [10] compared different machine learning methods to address imbalanced class problems in predicting mental health. The findings indicated that the Random Forest method performed exceptionally well in handling imbalanced data. Therefore, this study employs the Gradient Boosting method, which is similar to Random Forest but offers greater accuracy. Additionally, to address the issue of imbalanced data, the SMOTE method is also utilized, as applied in [10].

Table 1 presents the state-of-the-art research on predicting various mental health conditions. Although numerous studies have discussed mental health, few have focused specifically on NPD, as this study does. Research by [11] focused on predicting NPD using a machine learning approach, but it relied on neurological and psychological features. In contrast, this study utilizes features derived from respondents' personal profiles, particularly their interactions with digital devices. This research aims to address the issue of early detection of NPD in employees by proposing the development of a predictive model based on machine learning using the Gradient Boosting method [12]. This method was chosen for its ability to handle complex data and gradually improve prediction accuracy, making it expected to produce more precise predictions for NPD.

Additionally, the application of machine learning in this study allows for the integration of broader employee data, including psychological, behavioral, and demographic information relevant to NPD. With comprehensive data analysis, the predictive model developed is expected to provide a deeper understanding of the factors influencing the emergence of NPD in the workplace. This research produce a predictive model that helps organizations identify employees at risk of NPD more quickly and accurately, thus supporting more effective intervention and prevention efforts. Ultimately, this can improve the psychological well-being and performance of employees in the workplace. In addition to its practical contributions, this research also expand the understanding of NPD dynamics in the workplace.

Table 1. State-of-the-art related to mental health prediction

Paper	Year	Type of Mental Health / Data source	Method	Notes
[13]	2020	Depression, anxiety, stress (quessionaire)	DT, RF, SVM, NB, KNN	Best method is Random Forest
[14]	2020	Depression, bipolar, anxiety, etc (Reddit)	DL, XGBoost, CNN	DL is better, but need the combination with NLP
[15]	2021	Social anxiety disorder (EEG signal)	CNN + LSTM	Focus on anxiety
[16]	2022	Bipolar disorder, depression	RNN, LSTM	Best method is LSTM with 3-layer
[17]	2023	Depression, traumatic, stress, etc (cancer patient)	RF	RF have high accuracy
[11]	2023	NPD	Kernel Ridge, SVR	Using the brain and psychological features
[18]	2023	Mental health	Single vs ensemble classifier	Gradient boosting have best performance
[19]	2023	BPD, bipolar, depression, and Anxiety (social media data)	Machine learning	LightGBM better than others
[10]	2024	Electronic Health Records from Ferrera, Itali	Machine learning, SMOTE-NC	Focus on imbalanced dataset

Overall, the development of a machine learning-based predictive model for early detection of NPD in employees is an important step toward creating a healthier and more productive work environment. By leveraging advanced technology, it is hoped that we can be more effective in detecting and managing mental disorders in the workplace, thereby creating a more inclusive, competitive, and sustainable environment.

2. METHODS

2.1 Research Framework

The research framework for detecting Narcissistic Personality Disorder (NPD) in the workplace begins by addressing a key problem: the difficulty of early detection and intervention for NPD, which remains suboptimal despite its serious impact on the work environment. To tackle this issue, the framework, as seen at Figure 1, outlines initial steps of conducting a literature review and consulting with psychologists to establish a foundational understanding of NPD and identify key indicators for prediction. This dual approach ensures a robust theoretical and expert-driven basis for the research, enabling the identification of relevant data points for further modeling.

The next phase involves data acquisition and modeling. Data is collected through expert assessments of employees' NPD levels and employee profile information based on psychologists' recommendations. This multimodal dataset becomes the input for developing a smart NPD prediction model using Gradient Boosting. The model is rigorously evaluated using metrics such as accuracy, precision, recall, and F1-score to ensure its reliability and effectiveness. The expected outcomes include a prototype of the prediction system, comprehensive analytical results, and academic contributions through publications and intellectual property rights (IPR). These deliverables aim to provide a practical and scientifically validated solution for early detection and management of NPD in workplace settings.

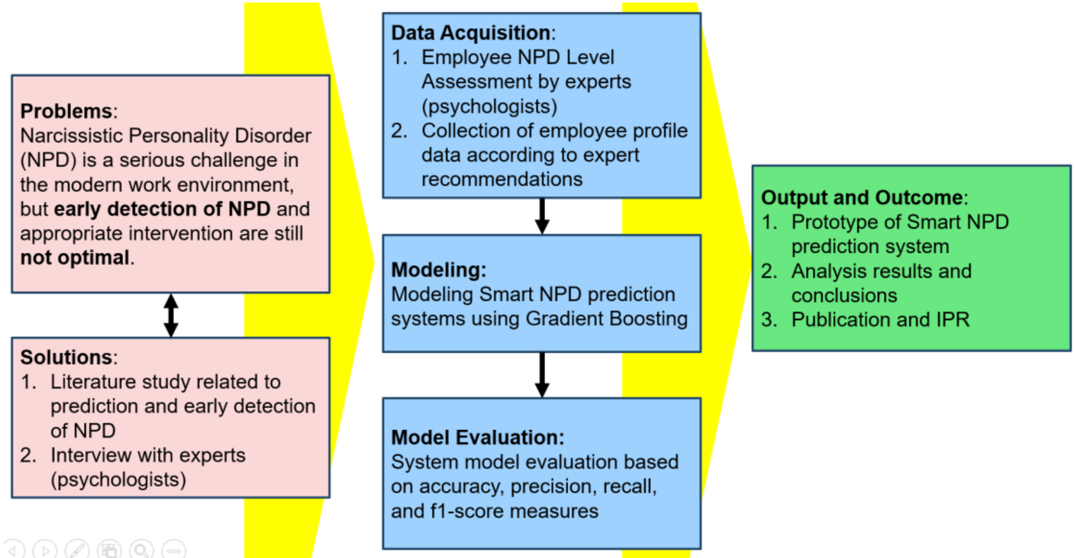


Figure 1. Research Framework

2.2 Gradient Boosting

Gradient Boosting is a machine learning technique used for both regression and classification tasks. It works by iteratively building an ensemble of decision trees, where each subsequent tree attempts to correct the errors made by the previous ones. The process begins with a dataset (X, y), where X represents the features and y the target labels. The first decision tree is trained to predict the target labels, and its errors (residuals) are calculated. These residuals are then used as the target for the next tree. This process is repeated, with each tree focusing on reducing the residuals of the previous model, resulting in a sequence of trees that collectively improve prediction accuracy.

In Gradient Boosting, the contribution of each tree is weighted by a learning rate, which controls the impact of each tree on the final model. The method optimizes a loss function by using gradient descent, where the gradient of the loss function guides the correction of residuals at each step. The final prediction is made by combining the outputs of all the trees, effectively forming a strong predictive model. The iterative nature of Gradient Boosting allows it to handle complex data patterns, making it a powerful and widely used approach in machine learning. However, it requires careful tuning of parameters like the number of trees, learning rate, and tree depth to avoid overfitting.

The tree-building process in this study follows the principles of ensemble learning, specifically using Gradient Boosting, XGBoost, and LightGBM. These models construct decision trees sequentially, where each new tree corrects the errors made by the previous ones. Initially, a weak learner (a shallow decision tree) is created using a portion of the dataset. The model then calculates the residual errors and assigns higher weights to misclassified instances, allowing the next tree to focus more on difficult cases. This iterative process continues until the model reaches the specified number of estimators, minimizing the loss function (log_loss in this study) to improve classification accuracy.

This research adopts a Gradient Boosting approach, as displayed in Figure 2, to detect potential narcissistic personality disorder among employees in the IT sector. The process includes stages of data collection, cleaning, and splitting the dataset into training and testing data. The Gradient Boosting models, including GBC, XGBoost, and LightGBM, are trained using the training data to identify complex patterns associated with narcissistic personality disorder. Evaluation is conducted by measuring accuracy, precision, recall, and F1-score to assess the model's performance on the previously unseen testing data. It is expected that the results of this model analysis can aid in formulating more targeted prevention and intervention strategies to address narcissistic personality disorder in the workplace of the IT industry.

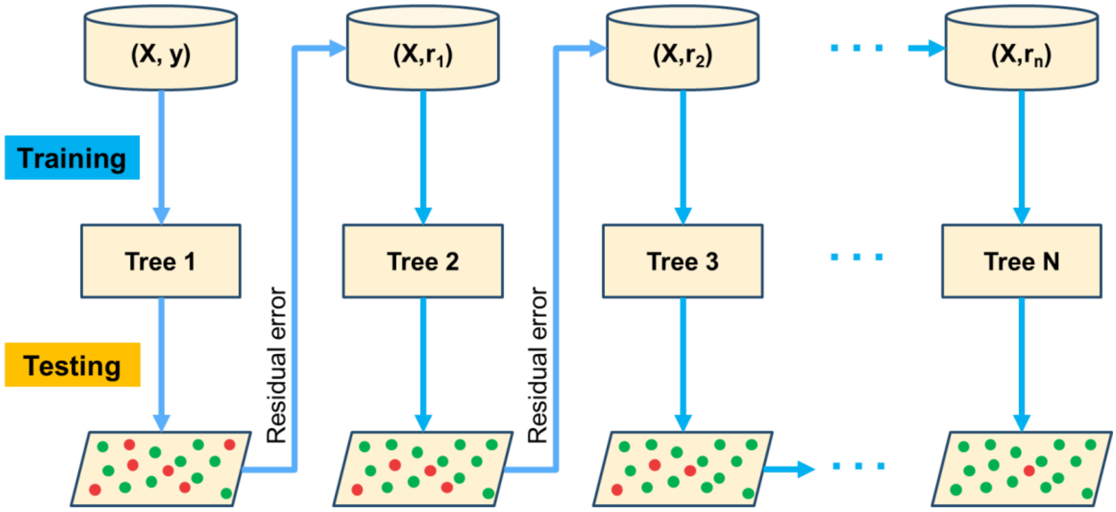


Figure 2. Gradient Boosting Method

2.3 Dataset

This research involves 100 employees from the information technology (IT) sector as the sample. The purposive random sampling method is applied to ensure acceptable representation from various backgrounds and behaviors of the respondents. The purpose of this approach is to provide a comprehensive overview of the characteristics of narcissistic personality disorder among IT employees in general. The questionnaire is designed to explore the habits and preferences of social media usage and their potential relationship with narcissistic personality disorder. The primary instrument in this research is a questionnaire developed by experienced psychologists to identify the characteristics of narcissistic personality disorder. The questionnaire validation process is conducted to ensure the accuracy and relevance of the questions posed.

Table 2 summarizes the respondents' demographic and social media behavior data, with categorical variables converted into numerical codes for analysis. Most respondents are men (65%), aged between 26–30 years (the largest group), and earn between 5,000,000 and 10,000,000 IDR. Regarding smartphone usage, 65% use Android devices. In terms of social media habits, most have 3–7 accounts, spend over 2 hours daily on social media, and rarely update their social media content.

For social media interactions, "Likes" are the most common form of engagement, followed by "Comments" and "Shares/Reposts." Respondents are also classified into four NPD levels: Non-NPD, Low-NPD, Middle-NPD, and High-NPD. The numerical encoding of demographic and behavioral characteristics enables the dataset to be utilized in quantitative analyses, facilitating pattern recognition and modeling through machine learning techniques.

The data labeling process in this study involved collaboration with a psychologist (expert) who used a psychological instrument to determine the respondents' levels of Narcissistic Personality Disorder (NPD). This instrument, designed to measure narcissistic tendencies based on established psychological criteria, provided a structured and objective method for categorizing respondents. The psychologist assessed each respondent and assigned them to one of four categories: Non-NPD, Low-NPD, Middle-NPD, or High-NPD. These labels were critical for ensuring the accuracy and reliability of the classification process.

Table 2. Respondents

Field	Original Value	Total	Value after Transformation
Gender	Men	65	1
	Women	35	2
Age	<20	2	1
	21-25	12	2
	26-30	60	3
	31-35	12	4
	36<	14	5
Monthly Income (IDR)	< 5,000,000	26	1
	5,000,000 – 10,000,000	34	2
	10,000,000 – 15,000,000	17	3
	15,000,000 – 25,000,000	12	4
	> 25,000,000	11	5
Smartphone Type	Android	65	1
	Apple/Iphone	35	2
Number of social media account	< 3 accounts	25	1
	3-7 accounts	69	2
	> 7 accounts	6	3
Social media accounts that are always used	1 account	15	1
	2 accounts	16	2
	3 accounts	24	3
	4 accounts	26	4
	>= 5 accounts	19	5
Daily access time using social media	< 1 hours	11	1
	1-2 hours	35	2
	> 2 hours	54	3
How often do you update social media?	Very Rare	53	1
	Rarely	5	2
	Quite Often	22	3
	Often	4	4
	Very Often	15	5
Post achievements on social media	Never	40	1
	Sometimes	53	2
	Always	7	3
Other people's favorable responses to social media posts	Share / Repost	30	1
	Comment	33	2
	Like	37	3
Label	Non-NPD	4	0
	Low-NPD	56	1
	Middle-NPD	35	2
	High-NPD	5	3

By using a validated psychological instrument, the labeling process was grounded in scientifically recognized methods, which enhanced the dataset's credibility. The involvement of a psychologist ensured that the labels were not only based on observable social media behaviors but also aligned with psychological constructs measured through the instrument. This expert-driven approach provided a robust foundation for analyzing patterns and training machine learning models to predict NPD levels based on demographic and behavioral data. The structured and scientifically rigorous labeling process added significant value to the study.

After preprocessing, the categorical values in the 'NPD' column were replaced with numerical representations: 'No NPD' became 0, 'Low NPD' became 1, 'Moderate NPD' became 2, and 'High NPD' became 3. This step is crucial to ensure that the categorical data is ready for use in machine learning models such as Gradient Boosting Classifier.

2.4 Dataset Balancing with SMOTE

To address the issue of class imbalance in the training data (See Table 2), the researcher applied Synthetic Minority Over-sampling Technique (SMOTE) [20], [21]. Class imbalance occurs when the distribution of samples among different classes is uneven, which can lead to biased model performance. In this study, the first step in implementing SMOTE was to analyze the class distribution and identify the smallest minority class. This allowed the researcher to determine the appropriate number of synthetic samples needed to balance the dataset. The k_neighbors parameter in SMOTE was carefully configured to ensure that it did not exceed the number of samples in the smallest minority class. Specifically, the value of k_neighbors was set to one less than the number of samples in the smallest class to prevent the generation of unrealistic synthetic samples and maintain the integrity of the data.

The application of SMOTE transformed the imbalanced dataset into a more balanced one by generating synthetic samples for the minority class. This method works by creating new data points in the feature space based on the nearest neighbors of the existing samples in the minority class. These synthetic samples are not duplicates but are interpolations between actual data points, which helps retain the diversity within the dataset. By applying this technique, the researcher ensured that the model would be exposed to sufficient examples from both the majority and minority classes, improving its ability to learn from all classes effectively. Table 3 provides the dataset before and after applying SMOTE.

Table 3. Number of Record Before and After SMOTE

Label	**Before SMOTE**	**After SMOTE**
Non-NPD (0)	4	56
Low-NPD (1)	56	56
Middle-NPD (2)	35	56
High-NPD (3)	5	56

Balancing the dataset using SMOTE was crucial in this study to address the imbalance among NPD levels, which could otherwise bias the model toward the majority class. By generating synthetic samples for the minority classes, SMOTE helped the Gradient Boosting Classifier learn more representative patterns, improving its ability to generalize and make accurate, unbiased predictions across all categories. This balancing step significantly enhanced the model’s fairness, reduced the risk of overfitting to dominant classes, and ensured better performance in real-world applications.

2.5 Recursive Feature Engineering (RFE)

Recursive Feature Elimination (RFE) is a widely used method for selecting the most important features in a dataset by recursively removing the least significant ones and retraining the model to find the best-performing feature subset. It helps reduce model complexity, eliminates redundancy, and improves generalizability, particularly in datasets with many features. In this study, RFE was applied using the GBC, where features were evaluated based on their contribution to model predictions, and the least important ones were systematically removed.

While RFE significantly enhances model interpretability and performance by selecting features with strong predictive power, it can be computationally expensive due to the repeated model training required, especially with large datasets. Despite this, its benefits in improving model accuracy and reducing overfitting make it a highly effective technique, particularly when combined with cross-validation to ensure the selected features generalize well to unseen data.

3. RESULTS AND DISCUSSION

3.1 Experimental Setup

In this study, the dataset was divided into 60% training data and 40% testing data to ensure a balanced evaluation of the model's performance. The dataset consisted of 224 records, which were obtained after applying the SMOTE to address class imbalance. The training set was used to train the machine learning models, while the testing set was reserved for evaluating their predictive accuracy. This split ratio was chosen to provide sufficient data for model learning while maintaining a reliable portion for performance validation.

Table 4 provides an overview of the experimental setup for evaluating three machine learning models—Gradient Boosting Classifier (GBC), XGBoost, and LightGBM (LGBM)—both before and after applying Recursive Feature Elimination (RFE). The parameters in the table include the core hyperparameters used during model training, such as `n_estimators`, `learning_rate`, `max_depth`, and `random_state`, along with the number of attributes (features) used in the training process.

Before applying RFE, the models were trained using the full set of 10 features. All three models—GBC, XGBoost, and LGBM—shared the same hyperparameter values for consistency in evaluation. The `n_estimators` parameter was set to 100, indicating that each model trained 100 individual boosting iterations. The `learning_rate` of 0.1 ensured a steady learning process, balancing between convergence speed and overfitting prevention. The `max_depth` parameter was fixed at 3, which restricted the depth of each decision tree to avoid overly complex models. Lastly, the `random_state` value of 30 ensured reproducibility of results by controlling the randomization process.

Table 4. Experimental Setup

Parameter	**Before RFE**			**After RFE**		
	GBC	**XGBoost**	**LGBM**	**GBC**	**XGBoost**	**LGBM**
n_estimators	100	100	100	100	100	100
learning_rate	0.1	0.1	0.1	0.1	0.1	0.1
max_depth	3	3	3	3	3	3
random_state	30	30	30	30	30	30
number of attributes	10	10	10	5	5	5
Loss function	log loss	log loss	log loss	log loss	log loss	log loss

At this stage, all features in the dataset were included, which means the models might have been trained on some irrelevant or redundant features. While the models could potentially achieve good accuracy, the inclusion of unnecessary features can lead to increased computational overhead and reduced interpretability. After applying RFE, the number of features was reduced to the five most important attributes identified during the feature selection process. By eliminating less relevant features, the models were expected to become more efficient and focus on the features that contributed the most to the prediction task. The same hyperparameter values were retained for the models, maintaining `n_estimators` at 100, `learning_rate` at 0.1, `max_depth` at 3, and `random_state` at 30. This ensured a fair comparison between the models trained with the full feature set and the reduced feature set.

Reducing the number of attributes from 10 to 5 provided several advantages, including simplifying the models to reduce overfitting, improving computational efficiency, and enhancing interpretability by focusing on the most significant predictors. By integrating Recursive Feature Elimination (RFE) into the experimental pipeline, the study optimized model performance while maintaining simplicity, enabling the models to achieve strong results without compromising accuracy. The comparison between results before and after RFE emphasizes the critical role of feature selection in boosting models like GBC, XGBoost, and LGBM, ultimately producing models that are both efficient and robust.

The loss function used in this study is log_loss, which is commonly applied in classification tasks to measure the difference between predicted probabilities and actual class labels. Log loss penalizes incorrect predictions more heavily when the predicted probability is far from the true label, ensuring that the model produces well-calibrated probability estimates. This function was chosen because it effectively handles probabilistic outputs from boosting algorithms such as GBC, XGBoost, and LGBM, optimizing their performance in distinguishing between classes while minimizing classification errors.

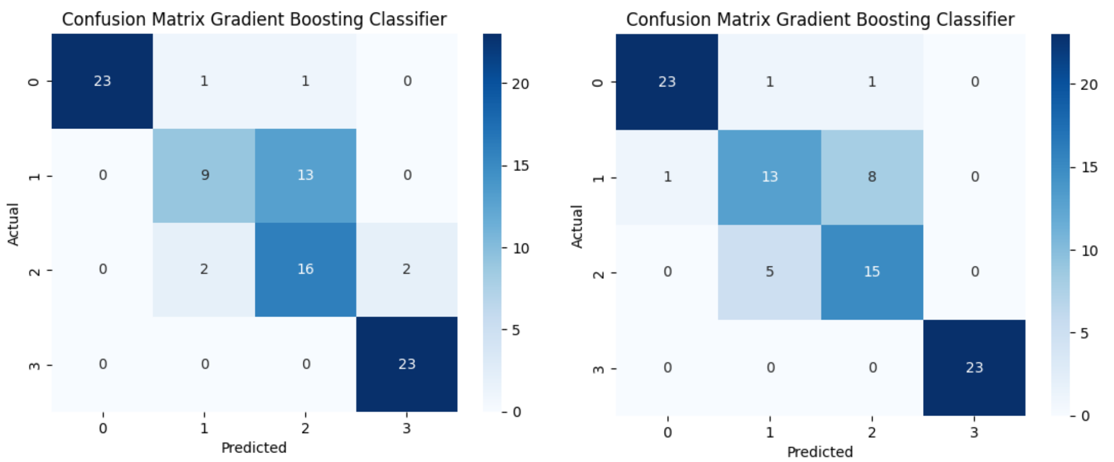
3.2 Experimental Result

In this study, k-fold cross-validation was employed to obtain the best-performing model by systematically evaluating different subsets of the training data. This approach helps ensure the model's generalizability and prevents overfitting. After selecting the optimal model, testing was conducted using 40% of the dataset to assess its final performance. This methodology ensures a robust evaluation of the model's predictive capabilities.

Table 5 provides a comparison of the performance metrics for Gradient Boosting Classifier (GBC), XGBoost, and LightGBM (LGBM) models before and after applying Recursive Feature Elimination (RFE). The metrics evaluated include **accuracy**, **precision**, **recall**, and **F1-score**, which offer insights into the predictive capabilities of each model.

Table 5. Experimental Result

Metrics	**Before RFE**			**After RFE**		
	GBC	**XGBoost**	**LGBM**	**GBC**	**XGBoost**	**LGBM**
Accuracy	0.79	0.79	0.77	**0.82**	0.76	0.79
Precision	0.80	0.81	0.81	**0.82**	0.75	0.78
Recall	0.78	0.78	0.76	**0.82**	0.74	0.78
F1-Score	0.77	0.77	0.74	**0.81**	0.74	0.77
Time (ms)	302.55	45.63	21.17	**582.58**	336.79	324.21



(a) (b)

Figure 3. Confusion Matrix of the GBC method (a) before applying RFE and (b) after applying RFE

Before applying RFE, the GBC and XGBoost models both achieved an accuracy of 0.79, while LGBM was slightly lower at 0.77. Precision was highest for XGBoost at 0.81, followed by GBC and LGBM. Recall scores for GBC and XGBoost were identical at 0.78, while LGBM scored 0.76. In terms of F1-score, GBC and XGBoost both scored 0.77, outperforming LGBM at

0.74. Overall, GBC and XGBoost demonstrated stronger performance than LGBM before feature selection, but there was still room for optimization.

After applying RFE, GBC showed notable improvements across all metrics, with accuracy rising to 0.82 and precision, recall, and F1-score each reaching 0.82. This indicates that feature selection helped GBC focus on the most critical attributes, enhancing its performance. Conversely, XGBoost experienced a decline in performance, with accuracy dropping to 0.76 and decreases observed in precision, recall, and F1-score, suggesting that RFE might have removed important features for this model.

Meanwhile, LGBM showed modest but consistent improvements after RFE. Its accuracy increased to 0.79, and precision, recall, and F1-score also improved to 0.78, 0.78, and 0.77, respectively. These results suggest that RFE helped LGBM achieve a better balance between precision and recall, enhancing its overall classification capability without causing significant overfitting or underfitting.

Table 5 shows that before applying Recursive Feature Elimination (RFE), LGBM had the fastest processing time at 21.17 ms, followed by XGBoost at 45.63 ms, while GBC was much slower at 302.55 ms, highlighting LGBM's suitability for time-sensitive applications. After RFE, all models experienced increased processing times, with GBC rising significantly to 582.58 ms, and XGBoost and LGBM increasing to 336.79 ms and 324.21 ms, respectively. These results suggest that although RFE enhances model accuracy and performance, it also raises computational complexity, emphasizing the need to balance execution speed and model quality when selecting methods for practical use.

3.2 Discussion

The experimental results demonstrate that feature selection significantly improves model performance, particularly for Gradient Boosting Classifier (GBC), which achieved the highest scores after applying Recursive Feature Elimination (RFE). While GBC adapted well to feature reduction, XGBoost's performance declined, indicating its greater reliance on a full set of features. LGBM showed moderate improvement, suggesting some robustness to feature reduction but still falling short of GBC's performance.

These findings highlight that feature selection techniques like RFE must be tailored to the specific characteristics of each model. Although RFE improved model interpretability and efficiency, its effectiveness varied across algorithms. Therefore, carefully balancing feature reduction with each model's requirements is crucial for achieving optimal results in practical applications.

The study also provides important insights for HRM practitioners, particularly in the IT industry, by demonstrating that the RFE-GB approach can improve the accuracy of predicting narcissistic personality disorder (NPD) behaviors. Integrating such machine learning techniques into HR practices enables more targeted, data-driven decisions, enhancing employee well-being and promoting more sustainable and responsive human resource management policies.

From a scientific perspective, this study emphasizes the critical role of machine learning, particularly feature selection, in addressing mental health issues. Selecting the right features significantly enhances model accuracy and reliability in predicting employees at risk of narcissistic personality disorder (NPD). By focusing on the most relevant attributes, predictive models can deliver more effective and consistent outcomes. Additionally, the research highlights the strong performance of Gradient Boosting methods, especially after applying the RFE-GB technique, which proved to be highly effective in predicting NPD among employees in the IT industry. These findings suggest that adopting advanced machine learning approaches like RFE-GB can greatly benefit human resource management and industrial psychology, enabling organizations to develop targeted intervention strategies and improve workforce well-being. Furthermore, this research provides a foundation for future studies to explore innovative solutions that combine technology and psychology for better workplace mental health management.

4. CONCLUSIONS

This research demonstrates the successful application of the Recursive Feature Elimination with Gradient Boosting (RFE-GB) technique to predict narcissistic personality disorder (NPD) among employees in the information technology industry. The study emphasizes the importance of precise feature selection, showing that focusing on the most relevant attributes significantly improves model performance across accuracy, precision, recall, and F1-score metrics. Gradient Boosting methods proved robust for handling imbalanced data and capturing complex behavioral patterns, making them a valuable approach for addressing workplace mental health challenges.

The findings highlight the potential of integrating machine learning into industrial psychology and human resource management, enabling organizations to better detect and manage mental health issues through data-driven strategies. Future research is encouraged to incorporate additional data modalities, such as audio, video, and biometric signals, to enhance prediction capabilities. Longitudinal studies could provide deeper insights into the development of NPD over time, while explainable AI techniques may offer better interpretability of model decisions. Testing the RFE-GB technique across different industries and cultural contexts is also recommended to validate its broader applicability, ensuring its effectiveness in diverse workplace environments.

ACKNOWLEDGEMENTS

We sincerely thank the Ministry of Education, Culture, Research, and Technology for funding this research and Universitas Budi Luhur for their support and resources. Your contributions have been essential to the success of this study.

REFERENCES

[1] S. Grapsas, E. Brummelman, M. D. Back, and J. J. A. Denissen, “The ‘Why’ and ‘How’ of Narcissism: A Process Model of Narcissistic Status Pursuit,” *Perspectives on Psychological Science*, vol. 15, no. 1, pp. 150–172, Jan. 2020, doi: 10.1177/1745691619873350.

[2] E. L. Kacel, N. Ennis, and D. B. Pereira, “Narcissistic Personality Disorder in Clinical Health Psychology Practice: Case Studies of Comorbid Psychological Distress and Life-Limiting Illness,” *Behavioral Medicine*, vol. 43, no. 3, pp. 156–164, Jul. 2017, doi: 10.1080/08964289.2017.1301875.

[3] E. Y. L. Ong *et al.*, “Narcissism, extraversion and adolescents’ self-presentation on Facebook,” *Pers Individ Dif*, vol. 50, no. 2, pp. 180–185, Jan. 2011, doi: 10.1016/j.paid.2010.09.022.

[4] E. Grijalva *et al.*, “Gender differences in narcissism: A meta-analytic review.,” *Psychol Bull*, vol. 141, no. 2, pp. 261–310, Mar. 2015, doi: 10.1037/a0038231.

[5] W. D. Ningrum and S. Nuzulia, “Behavioral Manifestations of Dark Personality Narcissism: an Indigenous Study on Indonesian Workers,” *Journal of Social and Industrial Psychology*, vol. 12, no. 1, pp. 9–24, Jun. 2023, doi: 10.15294/sip.v12i1.68844.

[6] S. Kristanto, “Tingkat Kecenderungan Narsistik Pengguna Facebook,” *Journal of Social and Industrial Psychology* , vol. 1, no. 1, pp. 1–6, 2012, [Online]. Available: http://journal.unnes.ac.id/sju/index.php/sip

[7] P. Kaushik, K. Bansal, and Y. Kumar, “Deep Learning in Mental Health: An In-Depth Analysis of Prediction Systems,” in *2023 International Conference on Communication, Security and Artificial Intelligence (ICCSAI)*, IEEE, Nov. 2023, pp. 364–369. doi: 10.1109/ICCSAI59793.2023.10421590.

[8] L. S. Khoo, M. K. Lim, C. Y. Chong, and R. McNaney, "Machine Learning for Multimodal Mental Health Detection: A Systematic Review of Passive Sensing Approaches," *Sensors*, vol. 24, no. 2, p. 348, Jan. 2024, doi: 10.3390/s24020348.

[9] T. Mokheleli, T. Bokaba, and T. Museba, "An In-Depth Comparative Analysis of Machine Learning Techniques for Addressing Class Imbalance in Mental Health Prediction.," in *ACIS 2023 Proceedings*, 2023. [Online]. Available: https://aisel.aisnet.org/acis2023

[10] E. Gentili *et al.*, "Machine learning from real data: A mental health registry case study," *Computer Methods and Programs in Biomedicine Update*, vol. 5, Jan. 2024, doi: 10.1016/j.cmpbup.2023.100132.

[11] K. Jornkokgoud *et al.*, "Predicting narcissistic personality traits from brain and psychological features: A supervised machine learning approach," *Soc Neurosci*, vol. 18, no. 5, pp. 257–270, Sep. 2023, doi: 10.1080/17470919.2023.2242094.

[12] K. Katiyar, H. Fatma, and S. Singh, "Predicting Anxiety, Depression and Stress in Women Using Machine Learning Algorithms," in *Combating Women's Health Issues with Machine Learning*, Boca Raton: CRC Press, 2023, pp. 22–40. doi: 10.1201/9781003378556-2.

[13] A. Priya, S. Garg, and N. P. Tigga, "Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms," in *Procedia Computer Science*, Elsevier B.V., 2020, pp. 1258–1267. doi: 10.1016/j.procs.2020.03.442.

[14] J. Kim, J. Lee, E. Park, and J. Han, "A deep learning model for detecting mental illness from user content on social media," *Sci Rep*, vol. 10, no. 1, Dec. 2020, doi: 10.1038/s41598-020-68764-y.

[15] A. Al-Ezzi, N. Yahya, N. Kamel, I. Faye, K. Alsaih, and E. Gunaseli, "Severity Assessment of Social Anxiety Disorder Using Deep Learning Models on Brain Effective Connectivity," *IEEE Access*, vol. 9, pp. 86899–86913, 2021, doi: 10.1109/ACCESS.2021.3089358.

[16] F. Ceccarelli and M. Mahmoud, "Multimodal temporal machine learning for Bipolar Disorder and Depression Recognition," *Pattern Analysis and Applications*, vol. 25, no. 3, pp. 493–504, Aug. 2022, doi: 10.1007/s10044-021-01001-y.

[17] K. Kourou *et al.*, "Personalized prediction of one-year mental health deterioration using adaptive learning algorithms: a multicenter breast cancer prospective study," *Sci Rep*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-33281-1.

[18] J. Chung and J. Teo, "Single classifier vs. ensemble machine learning approaches for mental health prediction," *Brain Inform*, vol. 10, no. 1, Dec. 2023, doi: 10.1186/s40708-022-00180-6.

[19] K. Nova, "Machine Learning Approaches for Automated Mental Disorder Classification based on Social Media Textual Data," *Contemporary Issues in Behavioral and Social Sciences*, vol. 7, no. 1, pp. 70–83, 2023.

[20] P. Setiyadi, N. Prayogi, and A. Solichin, "Optimalisasi Prediksi Kehilangan Karyawan Menggunakan Teknik RFE, SMOTE, dan AdaBoost," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 9, no. 4, pp. 2131–2145, 2024, doi: 10.29100/jipi.v9i4.5642.

[21] F. H. Alfebi and M. D. Anasanti, "Improving Cardiovascular Disease Prediction by Integrating Imputation, Imbalance Resampling, and Feature Selection Techniques into Machine Learning Model," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 17, no. 1, p. 55, Feb. 2023, doi: 10.22146/ijccs.80214.