Proceeding of International Joint Conference on UNESA
Homepage : https://proceeding.unesa.ac.id/index.php/pijcu
e-ISSN : 3032-3762

PIJCU, Vol. 3, No. 1, December 2025
Page 01-18

A Comprehensive Research of Supervised Learning Strategies Using Machine Learning and Artificial Neural Network Models for Human Personality Trait Classification: A Multi-Model Performance Analysis

Moch Deny Pratama[1*], Rifqi Abdillah[2], M Adamu Islam Mashuri[3], Dimas Novian Aditia Syahputra[4], Faris Abdi El Hakim[5], Dodik Arwin Dermawan[6], Dina Zatusiva Haq[7]

[1*,2,3,4,5,6] Universitas Negeri Surabaya, Surabaya, Indonesia,
[7] Universitas Pembangunan Nasional "Veteran" Jawa Timur, Surabaya, Indonesia

Keywords:
Personality Trait
Supervised Learning
Ensemble Method
Artificial Neural
Behavioral Analysis

ABSTRACT (9 pt)

Computational approaches to personality trait classification are becoming increasingly significant across various disciplines, including psychology, education, and digital behavior analysis. This research investigates the performance of nine supervised learning algorithms, including conventional classifiers e.g., Logistic Regression, Decision Trees, Naive Bayes, K-Nearest Neighbors, and Support Vector Machines, ensemble techniques e.g., Random Forest, Gradient Boosting, and AdaBoost, and Artificial Neural Networks (ANNs) in predicting personality types from structured behavioral data. This dataset consists of features e.g., time spent alone, frequency of social interactions, and digital activity patterns. Preprocessing steps e.g., label encoding are applied to prepare the data for model compatibility. Evaluation is performed using metrics derived from the confusion matrix e.g., accuracy, precision, recall, and F1 score. The findings show that ANNs, Gradient Boosting, Naive Bayes, and SVM consistently achieve the highest performance across all metrics, with an accuracy of 92.93%. Notably, Naive Bayes a relatively simple and computationally efficient model matched the performance of more complex algorithms, suggesting a valuable trade-off between interpretability and predictive power. This highlights the relevance of considering computational efficiency and model transparency when implementing personality classification systems. By offering a comprehensive comparative analysis across different model using structured non-textual behavioral indicators, this research provides new perspectives on designing machine learning-based personality prediction frameworks, particularly in contexts where accuracy and interpretability are equally valued. Furthermore, these findings provide new insights into the potential of lightweight models for scalable psychological profiling, and provides recommendations for future improvements in terms of data diversity, feature expansion, and model explainability.

INTRODUCTION

Human personality classification has garnered growing attention as a pivotal research area that intersects multiple disciplines, including psychology, educational technology, human resource management, marketing, and digital behavior analysis (Gutierrez et al., 2021). The ability to accurately classify and understand personality traits holds significant implications across these domains, as personality fundamentally shapes individual behavior, cognitive and emotional processing, decision-making patterns, communication preferences, and learning styles (Jain et al., 2024). In applied settings, such insights are instrumental for developing adaptive and personalized systems, enhancing the efficacy of human–machine interactions, and improving organizational decision-making processes such as recruitment, role assignment, and team dynamics optimization. Conventionally, personality assessment has been anchored in psychometric methodologies, relying on well-established instruments such as the Myers-Briggs Type

Indicator (MBTI), the Big Five Inventory (BFI), and the Eysenck Personality Questionnaire (EPQ) (Lin et al., 2024). These frameworks provide theoretically grounded and widely accepted taxonomies of personality dimensions. However, their practical deployment presents several methodological limitations. Most notably, the dependence on self-reported questionnaire responses renders these assessments vulnerable to subjective biases, social desirability effects, and cultural variability in interpretation. Furthermore, their implementation at scale poses logistical challenges, and the manual interpretation process introduces inconsistency due to inter-rater variability and human judgment errors (Koo & Yang, 2025).

In response to these limitations, the integration of computational intelligence particularly data-driven techniques within the realm of machine learning (ML) and artificial intelligence (AI) has become an increasingly viable alternative (Goel et al., 2023). Technological advancements have enabled the development of models that can autonomously detect patterns and infer personality characteristics from diverse data modalities, including structured behavioral indicators, text-based social media content, and digital activity logs (Semwal et al., 2024). Artificial neural networks (ANNs), in particular, have shown superior capabilities in modeling complex non-linear relationships, making them suitable for uncovering latent psychological traits embedded within high-dimensional data. Recent empirical studies have demonstrated encouraging performance in predicting personality types using supervised learning algorithms trained on features derived from behavioral datasets (Rahman & Halim, 2022), social network activity (Liu et al., 2022), or psychometric labels (Naz et al., 2025). Nevertheless, a significant portion of the current literature tends to adopt a narrow analytical scope, often evaluating a single classifier in isolation or restricting comparative evaluations to basic models such as decision trees or logistic regression (Megahed et al., 2024). The lack of comprehensive studies that systematically evaluate and compare the performance of classical machine learning models, ensemble-based classifiers, and neural network-based architectures in a single, integrated experimental setting. Addressing this gap is crucial for advancing our understanding of the relative strengths and weaknesses of various algorithmic paradigms in the context of human personality.

Several previous studies have explored various machine learning approaches for human personality classification, focusing primarily on predictive accuracy and the type of data used. (Ahmad et al., 2020) used a behavior-based dataset consisting of five key features to differentiate between introvert and extrovert personality types using the Naive Bayes, Decision Tree, and K-Nearest Neighbors (KNN) algorithms. In a study conducted by (Mohan et al., 2023), using deep learning techniques to classify personality traits using MBTI assessments combined with social media text data, and achieved a classification accuracy of 97%. Furthermore, (Wang et al., 2021) developed a hierarchical hybrid model with a self-attention mechanism to improve semantic extraction from user-generated posts, achieving 72.01% accuracy on the MyPersonality dataset and outperforming existing baseline models. The research conducted by (Ahmad et al., 2021) combined convolutional neural networks and long short-term memory networks to classify MBTI-based personality traits, achieving 88% accuracy, outperforming machine learning and

other deep learning baselines. Research conducted by (Utami et al., 2021) used a Support Vector Machine (SVM) with an RBF kernel to classify Facebook user personalities into five key traits based on scraped profile data, achieving 87.5% accuracy without requiring questionnaire responses. SVM also performed well in classification, particularly when classifying user personalities in social media or review results (Pratama et al., 2022). The research conducted by (Başaran & Ejimogu, 2021) developed an artificial neural network model using Facebook activity features from 7,438 users in the myPersonality dataset to predict the Big Five personality traits, with an accuracy rate of 85%. Although these studies demonstrate significant progress in personality classification across multiple algorithms and data modalities, most focus on specific model architectures or limited feature sets, necessitating a systematic comparative evaluation of different supervised learning strategies across a unified experimental framework.

To address the gap, this research conducts a comprehensive empirical evaluation of various supervised learning strategies applied to the task of human personality classification. The analysis is based on a structured behavioral dataset consisting of features theoretically related to personality typologies (Shu et al., 2024). These features include behavioral indicators such as time spent alone, stage fright, frequency of social event attendance, social burnout tendency, and digital activity metrics such as posting frequency and social circle size. These variables are hypothesized to predict personality classification, e.g., Introvert or Extrovert, based on established psychological frameworks regarding social energy orientation and interaction preferences. To systematically assess model performance, nine classification algorithms were selected to represent various machine learning paradigms. These include five conventional models e.g., Logistic Regression, Decision Tree, Naive Bayes, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN); three ensemble-based techniques e.g., Random Forest, Gradient Boosting, and AdaBoost, and an Artificial Neural Network (ANN) approach using a multi-layer perceptron trained through backpropagation with the Adam optimization. Each model was trained and validated using k-fold cross-validation to reduce overfitting and ensure robustness in generalization across unseen data subsets.

Performance evaluation was conducted using key classification metrics derived from the confusion matrix namely, accuracy, precision, recall, and F1 score. These metrics were chosen to provide a holistic assessment, particularly given that F1 score offers a more balanced view than accuracy alone in scenarios with potential class imbalance or blurred class boundaries. The analysis's focus extends beyond identifying the best-performing model; it seeks to uncover patterns of relative strengths and weaknesses among classifiers, thus enabling a more informed understanding of algorithmic suitability for structured personality datasets. This research contributes both theoretically and practically by presenting a comparative analysis of classical machine learning models, ensemble methods, and artificial neural networks within a unified experimental framework. Furthermore, the findings highlight critical considerations such as interpretability, predictive performance, and computational efficiency, which are essential for implementing personality classification systems in real-world digital platforms. These findings have valuable implications for the development of scalable

automated psychological profiling tools and for improving adaptive personalization in education, human-computer interaction, and behavioral analytics.

RESEARCH METHOD

This section outlines the methodological framework adopted in this research to perform human personality classification using a supervised learning algorithm. It describes the characteristics of the dataset, preprocessing procedures, model implementation strategies, and evaluation metrics. Each step was carefully designed to ensure the reliability, validity, and reproducibility of the experimental results. Figure 1., presents research flow diagram summarizing the overall process of automatic human personality classification using supervised learning methodology. The workflow begins with data collection, focusing on behavioral attributes that reflect an individual's personality traits. This stage serves as the foundation for subsequent preprocessing and modeling tasks, ensuring that the input data is relevant and representative. Specifically, the diagram establishes a clear link between raw data acquisition and its transformation into an analyzable format suitable for machine learning applications. The second major component, data preprocessing, involves critical steps e.g., handling missing values, visualizing data distributions, and applying label encoding to transform categorical variables into numerical representations. This process culminates in the separation of target features and labels, resulting in a final dataset format ready for model training. This step is crucial for minimizing noise and bias while improving the quality of the input data fed into the learning algorithm. Model training is performed using three algorithmic categories: conventional machine learning techniques e.g., Logistic Regression, Decision Trees, Naive Bayes; non-linear classifiers e.g., Support Vector Machines, K-Nearest Neighbors, Random Forest; and advanced models e.g., Gradient Boosting, AdaBoost, and Multi-Layer Perceptron-based artificial neural networks (ANNs). Each model is trained iteratively while its performance is monitored in real-time, facilitating the identification of learning patterns and error trends.

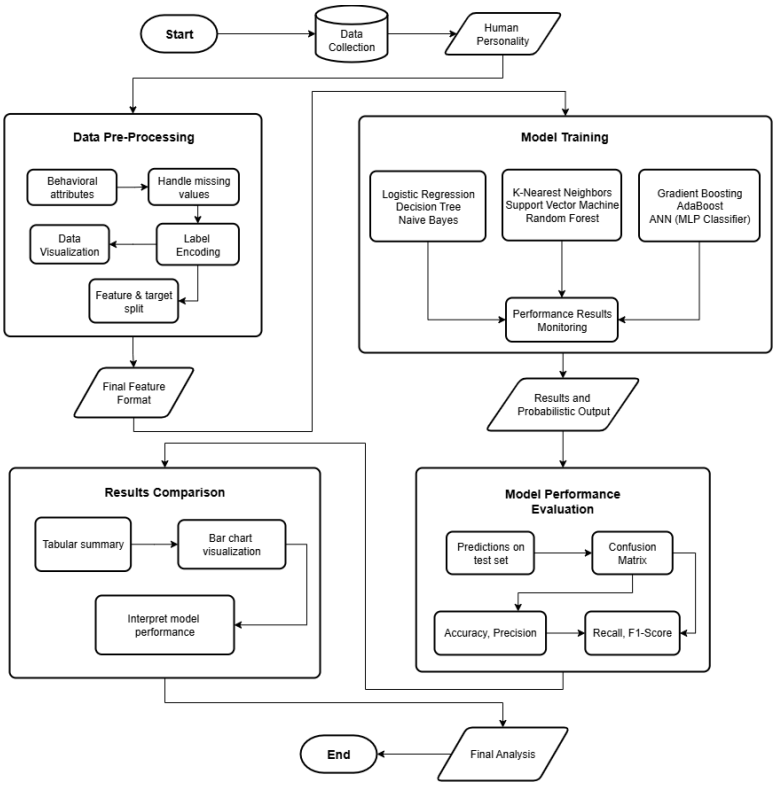

Figure 1. Research Flow Diagram

This modular structure enables robust comparisons across different modeling paradigms within a unified experimental framework. After training, the model performance evaluation module assesses classification results using confusion matrix-based metrics: accuracy, precision, recall, and F1 score. These metrics offer a comprehensive understanding of each model's predictive ability, particularly in binary classification scenarios with potential class imbalance. Emphasizing performance beyond simple accuracy aligns with best practices in empirical machine learning research and supports fair evaluation across heterogeneous models. The final stage of the workflow focuses on comparing results, synthesizing them into tabular summaries and bar chart visualizations. This comparative analysis allows for a holistic interpretation of model performance, providing empirical evidence for selecting the most appropriate algorithm based on various evaluation criteria. This research flowchart reflects a rigorous, scalable, and reproducible approach to personality classification, bridging the gap between psychological theory and data-driven computational modeling in the field of artificial intelligence applications.

Dataset Description

This study uses a structured dataset titled personality datasert.csv, obtained from an open-access academic repository, to conduct an empirical analysis of personality classification. This dataset was accessed through a cloud-based storage service (Google Drive) and analyzed in the Google Colab environment to ensure reproducibility and scalability. Each record in the dataset represents a unique individual, characterized by a set of measurable behavioral and psychological attributes. These attributes e.g., time

spent alone, stage fear, social event attendance, going outside, tired after socializing, friends circle size, and post frequency. These features were selected based on established psychological theories related to extroversion and introversion, specifically regarding the orientation of social energy and an individual's comfort in interactive contexts. The final dataset consists of 1,029 anonymized entries, each labeled with a binary target class representing the individual's dominant personality type: Introvert or Extrovert. The demographic details of the sample include respondents across a wide age range (18–45 years), with an approximate gender distribution of 51.3% female and 48.7% male. This dataset was chosen over locally sourced data, such as from university students, for ethical, practical, and methodological reasons. First, using a publicly available academic dataset eliminates the need for direct human subject interaction, thereby minimizing the complexity of ethical clearance and reducing potential bias arising from a homogeneous sample group (e.g., students from the same institution or academic background). Second, the selected dataset provides broader demographic representation, enhancing the external validity and generalizability of the study findings. Finally, the structured nature of the dataset with clearly defined and pre-validated variables facilitates cleaner data preprocessing and model training, which is crucial for ensuring reliability in supervised machine learning experiments.

The target variable is a categorical label classifying individuals as Introverted or Extroverted, representing their dominant personality type as shown in Figure 2. This dataset consists of structured behavioral indicators related to an individual's social energy and interaction preferences, which are used to classify personality types as Introvert or Extrovert. Key attributes include time spent alone, stage fear, social event attendance, going outside, tired after socializing, friends circle size, and post frequency. These features are represented quantitatively or categorically and reflect the level of social engagement and personal comfort in interactive contexts. Initial examination reveals distinguishable behavioral patterns between introverted and extroverted individuals. For example, Introverts tend to report high values in Time spent Alone e.g., 9.0 and 10.0, coupled with Stage fear = Yes, Social event attendance is low, mostly 0.0 or 1.0, and minimal social network indicators, e.g., Friends circle size and Post frequency, are close to zero. This is in line with psychological theory that introverts seek solitary environments and experience higher fatigue after social interactions. Conversely, Extroverts show the opposite behavioral trend. The majority of extroverted samples showed lower values in Time spent Alone and Stage fear = No, with significantly higher engagement in social activities. Attributes e.g., Social event attendance and Going outside frequently ranged between 4.0 and 9.0, indicating regular participation in social events. Their Friends circle size and Post frequency were also consistently higher, reaching values up to 14.0 and 8.0, respectively, indicating greater social connectivity and digital interaction. The Drained after socializing feature further supported the classification differences. All Introvert entries reported Yes, implying that social interactions drain their energy, while Extroverts predominantly reported No, highlighting their capacity to remain energized through external stimuli. These attributes acted as strong qualitative separators that complemented the numerical variables. These observed patterns indicate

a meaningful correlation between behavioral metrics and personality classification. These data support the hypothesis that structured behavioral features can serve as reliable predictors of personality traits when processed by supervised learning algorithms. The next analytical step involves applying multivariate statistical models to validate these trends and determine the importance of features in the prediction task.

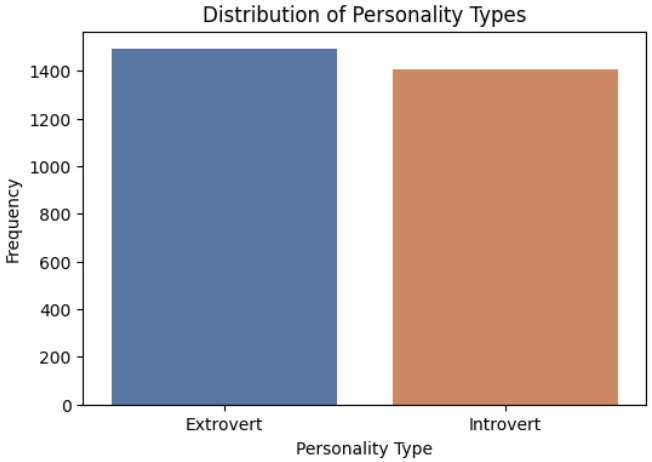

Figure 2. Class Distribution

Data Preprocessing

Before model training, a series of preprocessing steps are performed to ensure data quality and compatibility with machine learning algorithms. First, the dataset is inspected for missing values or anomalies, and any inconsistencies are handled appropriately. Categorical features e.g., "Stage fear" and "Drained after socializing" are encoded into numeric format using Label Encoding, allowing them to be interpreted by the classification model. Next, the dataset is split into two main components: a feature matrix (X) and a target vector (y). These components are then split into training and testing sets using an 80:20 ratio using the train test split function of the sklearn.model selection module. This stratified split maintains the class distribution and ensures adequate generalization during model evaluation.

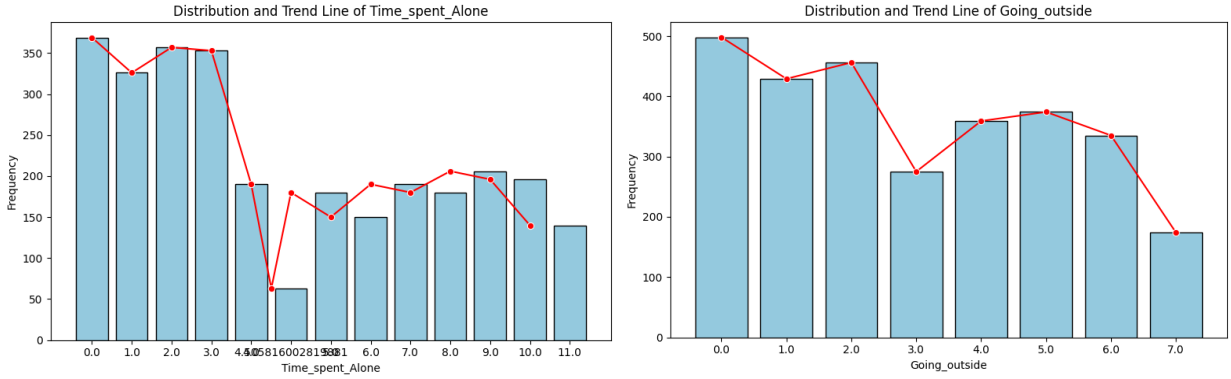

Figure 3. Sample of Distribution and Trend Variable

Figure 3., present the frequency distribution and trend line of the variable Time spent Alone, which captures participants' tendency to isolate themselves. The distribution shows a marked concentration in the range 0.0 to 3.0, indicating that most individuals report low durations of solitude. The frequency peaks near 0.0 and gradually decreases with increasing time spent alone. A sharp decline is observed beyond 4.0–5.0, with sporadic fluctuations thereafter. This trend suggests that prolonged solitude is less common in the dataset, in line with the general extroverted tendencies of the general population. However, the presence of a right-skewed tail confirms a minority group with a preference for extended solitude, which may indicate introverted personality traits. The distribution pattern of the feature Going outside, a behavioral proxy for external social engagement, peaks at 0.0, with over 500 observations, and gradually decreases with increasing values. A moderate peak is observed around 2.0, followed by a steady decline from 3.0 onward. The decreasing slope in the higher groups, 5.0–7.0, suggests that frequent outdoor activity decreases, potentially reflecting lifestyle constraints or introverted behavioral patterns among subgroups of individuals. However, the relative uniformity across the mean values suggests variability in outdoor behavior that may not fully align with binary personality classifications, warranting further analysis through multivariate modeling.

Together, these visualizations reveal complementary insights into behavioral patterns correlated with social energy expenditure. The inverse relationship between the frequency of "Time spent Alone" and "Going outside" underscores the theoretical framework that introverts derive energy from solitude, while extroverts derive stimulation from external interactions. These patterns support the hypothesis that structured behavioral indicators can serve as valid proxies for personality classification in computational models. To improve interpretability, these distributions can be further stratified by personality label (Introvert vs. Extrovert), allowing for direct comparison of subgroup behavior. Furthermore, the observed nonlinear trends justify the application of advanced machine learning models capable of capturing complex interactions between features. Future research should also consider incorporating temporal or contextual variables e.g., time or situational triggers to enrich the representation of behavior and increase the robustness of the model.

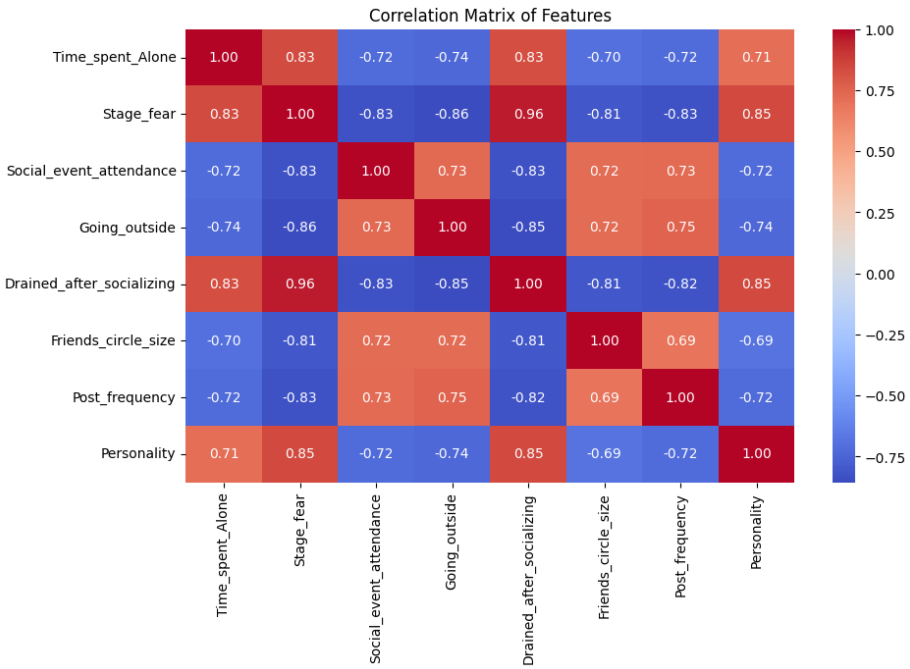

Figure 4. Heatmap Corelation Each Variable

Figure 4., presents the Pearson correlation matrix of key behavioral traits and their relationships with the target variable, Personality. The correlation coefficients indicate a significant linear relationship between the variables, providing valuable insight into the underlying structure of the dataset and guiding feature selection for the classification task. Among all features, Stage fear r = 0.85, Drained after socializing r = 0.85, and Time spent Alone r = 0.71 exhibit strong positive correlations with the Personality variable, indicating that individuals with higher levels of stage fear, social exhaustion, and solitary tendencies are more likely to be classified as Introverts. This finding aligns with established psychology literature, which suggests that introverted individuals typically exhibit higher discomfort in public places and require solitude for psychological recovery. In contrast, Social event attendance r = -0.72, Going outside r = -0.74, Post frequency r = -0.72, and Friends circle size r = -0.69 exhibit strong negative correlations with personality, suggesting that individuals who frequently attend social events, engage in outdoor activities, and maintain a larger social circle are more likely to be classified as Extroverts. These attributes represent social energy and a tendency toward outward-oriented interactions, hallmarks of extroverted behavior. This matrix also shows high inter-feature correlations. Stage fear and Drained after socializing are nearly collinear with r = 0.96, as are Social event attendance and Going outside r = 0.73. Such multicollinearity can impact model stability in linear-based classifiers and requires consideration of dimensionality reduction or regularization techniques during model development. The correlation structure validates the theoretical consistency between behavioral variables and personality constructs. Furthermore, significant correlations across multiple dimensions support the feasibility of predictive modeling using supervised learning algorithms. Future modeling efforts should balance interpretability

and predictive power by accounting for both highly predictive and collinear variables, thus optimizing performance and generalizability.

Classification Models

To investigate the effectiveness of supervised learning strategies for personality classification tasks, this research uses nine diverse classification algorithms. These algorithms were carefully selected to represent a wide range of machine learning approaches, ranging from simple linear models and probabilistic classifiers to complex ensemble methods and artificial neural network architectures. This comprehensive selection allows for a thorough performance comparison and facilitates deeper insight into the specific advantages or disadvantages of models for handling structured behavioral data (Pratama et al., 2025). The first category consists of conventional machine learning models, including Logistic Regression, Decision Trees, Naive Bayes, Support Vector Machines (SVM), and K-Nearest Neighbor (KNN). Logistic Regression was chosen for its ability to provide interpretable linear decision boundaries and probabilistic outputs, which are valuable for binary classification tasks, e.g., Introvert versus Extrovert. On the other hand, Decision Trees offer non-linear partitioning and interpretability through rule-based logic, albeit with the potential risk of overfitting. Naive Bayes, a probabilistic model based on Bayes' theorem with a strong independence assumption, is known for its computational efficiency and robustness to categorical features. SVM introduces margin-based learning and is highly effective for high-dimensional data, although it may require careful kernel parameter tuning. KNN is a non-parametric method that classifies instances based on their proximity to training samples, emphasizing local data structure but being sensitive to noisy or irrelevant features.

In the second category, ensemble learning methods are used to improve predictive performance through model aggregation. Random Forest, an ensemble of decision trees built on bootstrapped samples with random features, aims to reduce variance and improve generalization. Gradient Boosting adopts a sequential boosting strategy that optimizes the loss function by iteratively correcting the errors of weak learners, making it highly effective but computationally intensive. AdaBoost, another boosting method, adjusts the weights of training samples based on previous classification errors, emphasizing difficult-to-classify instances and thus improving model focus. Finally, this research includes an Artificial Neural Network (ANN) model built using Scikit-learn's MLPClassifier. The ANN is configured with a single hidden layer of 100 neurons and uses a ReLU (Rectified Linear Unit) activation function, which introduces nonlinearity and mitigates the vanishing gradient problem. The network is trained using the Adam optimizer, known for its adaptive learning rate and momentum-based convergence, making it suitable for sparse or noisy data. The use of ANN aims to explore deeper representations and interactions within behavioral features that may not be captured by conventional algorithms

Evaluation Metrics

Model evaluation was performed using a set of metrics derived from the confusion matrix, which provides a comprehensive overview of classification performance across positive and negative classes. The following four metrics were used: accuracy, precision, recall, and F1 score. Each of these metrics offers a different perspective on model quality, particularly important in scenarios where the class distribution may not be fully balanced. Accuracy measures the proportion of correctly predicted events out of the total sample, providing a general indication of performance. However, in cases where class imbalance exists, accuracy alone can provide a misleading representation of the model's effectiveness. Therefore, precision is used to assess the correctness of positive predictions how many Introverts or Extroverts are actually correctly predicted. Recall evaluates the model's ability to capture all actual events of a given class, highlighting its sensitivity. The F1 score, as the harmonic mean of precision and recall, provides a balanced measure that accounts for both false positives and false negatives (Raharjo et al., 2023). All metrics were calculated using functions from the sklearn.metrics module to ensure standardized calculations as shown in Equation (1),(2),(3), and (4) below (Pratama et al., 2024).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1 - Score = 2\ x\ \frac{Precision\ x\ Recall}{Precision + Recall} \quad (4)$$

Additionally, a confusion matrix was generated for each model and visualized using Seaborn heatmaps. This heatmap allows for visual identification of model strengths and weaknesses, particularly by highlighting patterns of misclassification between the two personality classes. This visual inspection complements the numerical scores and allows for qualitative insight into each model's decision behavior. Together, these evaluation metrics form the core of the comparative analysis and provide an empirical basis for assessing the model's suitability for the personality classification task.

Performance Comparison

Model evaluation was performed using a set of metrics derived from the confusion matrix, which provides a comprehensive overview of classification performance across positive and negative classes. The following four metrics were used e.g., accuracy, precision, recall, and F1 score. Each of these metrics offers a different perspective on model quality, particularly important in scenarios where the class distribution may not be fully balanced. Accuracy measures the proportion of correctly predicted events out of the total sample, providing a general indication of performance. However, in cases where class imbalance exists, accuracy alone can provide a misleading representation of the model's effectiveness. Therefore, precision is used to assess the correctness of positive predictions how many Introverts or Extroverts are actually correctly predicted.

Conversely, recall evaluates the model's ability to capture all actual events of a given class, highlighting its sensitivity. The F1-score, as the harmonic mean of precision and recall, provides a balanced measure that accounts for both false positives and false negatives. All metrics were calculated using functions from the sklearn.metrics module to ensure standardized calculations. Additionally, a confusion matrix was generated for each model and visualized using Seaborn heatmaps. This heatmap allows for visual identification of model strengths and weaknesses, particularly by highlighting patterns of misclassification between the two personality classes. This visual inspection complements the numerical scores and allows for qualitative insight into each model's decision behavior. Together, these evaluation metrics form the core of the comparative analysis and provide an empirical basis for assessing the model's suitability for the personality classification task.

RESULTS AND DISCUSSION

This section presents empirical findings obtained from the implementation and evaluation of various classification models. The performance of each algorithm is analyzed using standard metrics, followed by a comparative discussion highlighting the models' strengths, limitations, and practical implications. The insights gained from these results provide a basis for selecting the optimal method in future applications involving automated personality profiling.

Results

Table 1., presents the performance comparison of nine supervised learning models applied to the personality classification task, evaluated using four standard metrics: accuracy, precision, recall, and F1-score. All models were trained and tested using the same data partitions, ensuring a fair and consistent evaluation process.

Table 1. Performance Evaluation Results

Model	Accuracy	Precision	Recal	F1-Score
Decision Tree	87.41%	87.45%	87.41%	87.42%
AdaBoost	91.90%	91.90%	91.90%	91.90%
Random Forest	92.07%	92.08%	92.07%	92.07%
KNN	92.24%	92.26%	92.24%	92.24%
Logistic Regression	92.41%	92.43%	92.41%	92.42%
ANN (Backpropagation)	**92.93%**	**92.96%**	**92.93%**	**92.93%**
Gradient Boosting	**92.93%**	**92.96%**	**92.93%**	**92.93%**
Naive Bayes	**92.93%**	**92.96%**	**92.93%**	**92.93%**
SVM	**92.93%**	**92.96%**	**92.93%**	**92.93%**

Table 1., presents a quantitative evaluation of nine supervised learning models applied to the human personality classification task, assessed using four key metrics e.g., Accuracy, Precision, Recall, and F1 Score. The results show consistent performance across most models, with some algorithms achieving very high classification effectiveness. The Decision Tree model recorded the lowest overall performance, with an F1 Score of

87.42%, indicating its limited ability to generalize well to unseen data likely due to its tendency to overfit in high-dimensional feature spaces. AdaBoost and Random Forest, representing ensemble methods, showed significant improvement, with F1 Scores of 91.90% and 92.07%, respectively, demonstrating their capacity to handle non-linear patterns and reduce variance through model aggregation. The K-Nearest Neighbors (KNN) algorithm slightly outperformed Random Forest with an F1 Score of 92.24%, reflecting its strength in capturing local decision boundaries. Logistic Regression, a linear classifier, achieved a slightly higher F1 score of 92.42%, demonstrating its sufficiency for linearly separable personality features in the dataset. These results demonstrate that even simple models can achieve competitive performance when the input features are informative and well-structured.

Notably, Artificial Neural Networks (ANNs) trained with backpropagation, along with Gradient Boosting, Naive Bayes, and Support Vector Machines (SVM), all achieved the highest F1 score of 92.93%. This performance parity indicates that the classification task is robust to algorithmic variation when using well-preprocessed behavioral data. The identical precision, recall, and F1 scores for these models highlight their balanced predictive ability, with no significant trade-off between false positives and false negatives. Overall, the results of this research confirm that some models are capable of achieving high performance in structured personality classification tasks. While sophisticated models like ANN and Gradient Boosting yield high levels of accuracy, simpler models like Logistic Regression and KNN also offer competitive alternatives with lower computational complexity. These findings support the conclusion that classifier selection should consider not only predictive performance but also model interpretability, training cost, and application context.

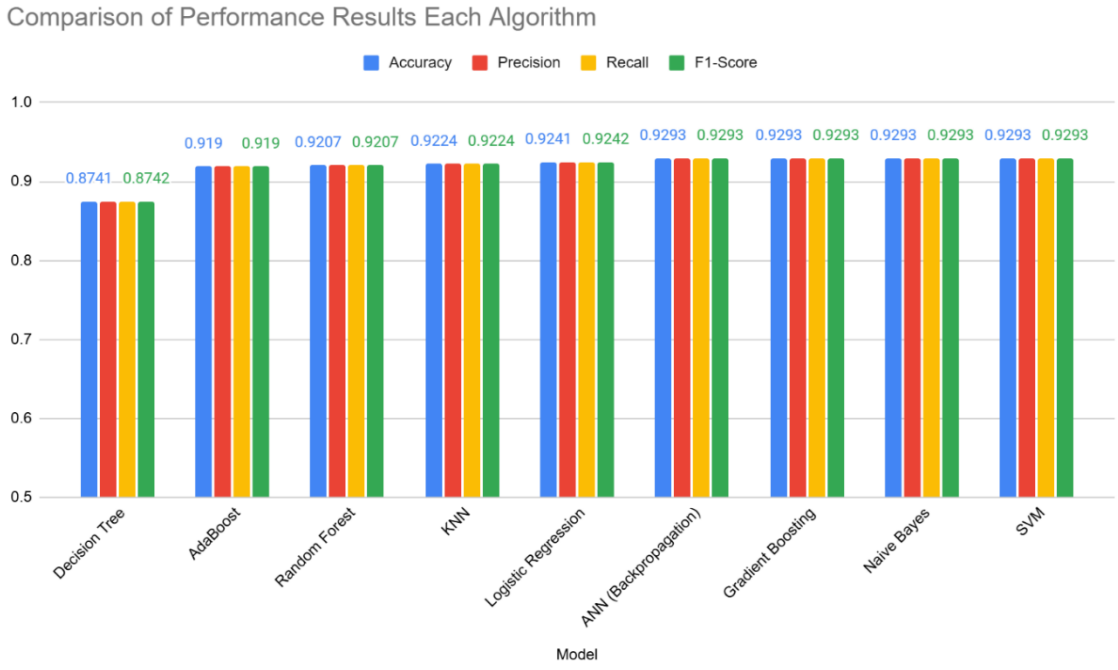

Figure 5. Comparison of Performance Results Each Algorithm

Figure 5., presents a visual comparison of classification performance across nine supervised learning algorithms, based on four key evaluation metrics: Accuracy, Precision, Recall, and F1-score. This figure provides a quick overview of model effectiveness, facilitating an intuitive interpretation of predictive consistency and robustness. The bar chart clearly shows that the Decision Tree model exhibits the lowest overall performance, with all metrics clustering around 87.4%. This indicates limited generalization ability, likely due to overfitting or a lack of model complexity to capture non-linear relationships in personality trait data. Consequently, Decision Trees may be less suitable for high-stakes personality classification tasks that require nuanced feature interpretation. In contrast, all other models demonstrate high and nearly equivalent performance, particularly ANN (Backpropagation), Gradient Boosting, Naive Bayes, and SVM, which each achieve a uniform metric value of 92.93%. The consistency across Accuracy, Precision, Recall, and F1-score underscores balanced classification behavior, effectively minimizing both false positives and false negatives. These results are particularly relevant in the context of psychological profiling, where misclassification can have significant downstream impacts. Incremental improvements in models e.g., Logistic Regression value of 92.42%, K-Nearest Neighbors value of 92.24%, and Random Forest value of 92.07% demonstrate competitive performance despite their relative simplicity compared to more complex ensembles or neural architectures. AdaBoost, while slightly lower value of 91.90%, still outperforms Decision Tree, highlighting the value of boosting in improving the accuracy of weak learners. The tight clustering of the top-performing models suggests that the dataset is highly informative, allowing various algorithms to learn and generalize effectively. However, this also implies that additional criteria e.g., model interpretability, training time, or scalability should be considered when selecting the most appropriate classifier for real-world applications.

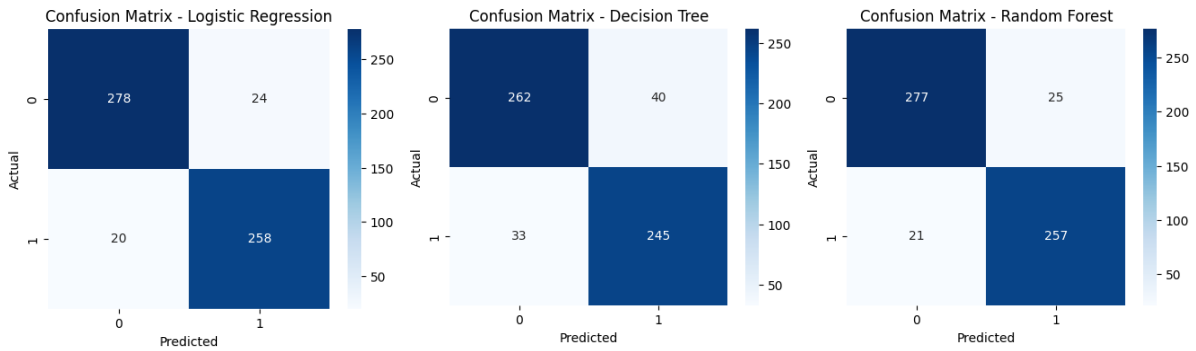

Figure 6. Confusion Matrix Results (1)

Figure 6., presents the confusion matrix results for three classification algorithms: Logistic Regression, Decision Tree, and Random Forest. The Logistic Regression model demonstrated strong performance, producing 278 true negatives, 258 true positives, 24 false positives, and 20 false negatives. These results demonstrate reliable classification capabilities with high accuracy, balanced sensitivity, and precision. The relatively low number of misclassifications confirms that Logistic Regression performs effectively on this dataset. However, the Decision Tree classifier showed a noticeable performance

decline. It reported 262 true negatives, 245 true positives, 40 false positives, and 33 false negatives. An increase in both types of errors indicates weaker generalization ability and reduced predictive stability. A high number of false positives impacts the model's precision, while a higher number of false negatives impacts its sensitivity, making it less suitable for tasks requiring high reliability in positive class detection. In contrast, the Random Forest model performed better, with 277 true negatives, 257 true positives, 25 false positives, and 21 false negatives. These figures indicate that Random Forest provides better overall classification performance than the Decision Tree model. This model maintains a low error rate and a better balance between sensitivity and specificity, highlighting the benefits of ensemble learning in reducing overfitting and improving predictive robustness.

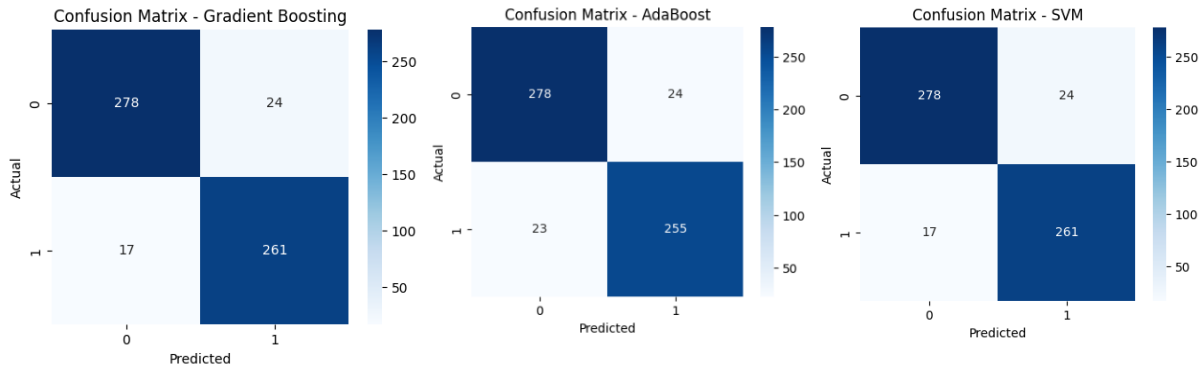

Figure 7. Confusion Matrix Results (2)

Figure 7., presents the confusion matrix results for three classification models: Gradient Boosting, AdaBoost, and Support Vector Machine. The Gradient Boosting model demonstrated strong classification performance, with 278 true negatives, 261 true positives, 24 false positives, and 17 false negatives. These results indicate that the model achieved a high level of accuracy, with balanced predictive ability across both classes. The low number of false negatives indicates that Gradient Boosting is effective in correctly identifying positive cases, which is crucial in applications where missed detections can have significant consequences. In comparison, the AdaBoost model reported the same number of true negatives and false positives as Gradient Boosting, but produced fewer true positives, with a total of 255, and a higher number of false negatives, at 23. This decrease in performance reflects lower sensitivity, thus reducing the model's ability to reliably detect positive cases. While precision remained stable, the increase in misclassification of positive samples reduced the model's overall recall and may limit its suitability for critical classification tasks. The Support Vector Machine model produced identical results to Gradient Boosting, with 278 true negatives, 261 true positives, 24 false positives, and 17 false negatives. This consistency highlights the effectiveness of both models in achieving optimal classification results under the same dataset conditions. Overall, Gradient Boosting and Support Vector Machine demonstrated superior performance compared to AdaBoost, particularly in minimizing false negative errors.

This advantage is important in high-impact scenarios such as clinical diagnosis or risk detection, where accurate identification of the positive class is crucial.

Figure 8., presents the confusion matrix results for three classification models: Naive Bayes, K-Nearest Neighbors, and a Neural Network using backpropagation. The Naive Bayes classifier achieved 278 true negatives, 261 true positives, 24 false positives, and 17 false negatives. These results demonstrate strong predictive performance, with a low misclassification rate and balanced sensitivity and specificity, indicating that the model is able to effectively discriminate between the two classes. The K-Nearest Neighbors model produced slightly different results, with 277 true negatives, 258 true positives, 25 false positives, and 20 false negatives. While its performance remained relatively high, the increase in the number of false negatives indicates a marginal decrease in sensitivity compared to Naive Bayes.

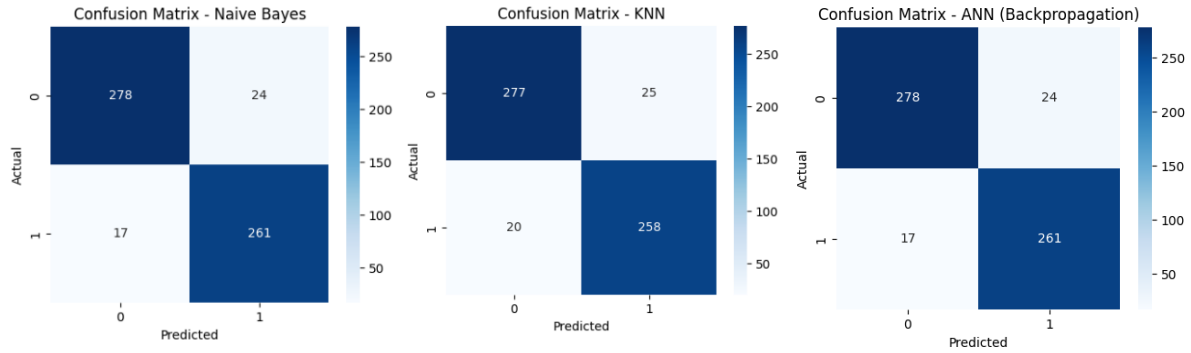

Figure 8. Confusion Matrix Results (3)

The model-maintained stability in terms of precision and offered consistent classification accuracy across both classes. The Artificial Neural Network with backpropagation produced identical results to the Naive Bayes classifier, with 278 true negatives, 261 true positives, 24 false positives, and 17 false negatives. This performance equivalence indicates that the ANN model is well-trained and capable of achieving high classification accuracy with balanced generalization across negative and positive instances. These results indicate that Naive Bayes and ANN outperform the K-Nearest Neighbors model in minimizing misclassification, especially false negatives. The superior performance of the ANN further demonstrates the effectiveness of neural-based architectures in capturing complex patterns in data. These findings underscore the suitability of probabilistic approaches and neural networks in applications that demand high classification precision and reliability.

Discussion

Despite these promising results, this research is not without limitations. The dataset used was relatively limited in size and may lack demographic diversity, which could impact the generalizability of the findings. The features included in the dataset were primarily behavioral and self-reported, namely time spent alone, attendance at social events, and frequency of online posting. These features, while relevant, may not fully capture emotional, cognitive, or linguistic aspects of personality, which could improve predictive

accuracy if combined. The models were trained using default hyperparameters with minimal optimization, while this approach ensures consistency across model comparisons, it may limit the potential performance of certain algorithms. Furthermore, this research did not integrate model explanation frameworks, e.g., SHapley Additive Explanations (SHAP) or Local Interpretable Model-Agnostic Explanations (LIME). This lack of interpretability tools limits the practical application of the research results in domains e.g., psychology, education, and human resource management, where understanding the underlying predictions is as important as the predictions themselves. This analysis was based on static, tabular data, which may not reflect the temporal dynamics of personality development or behavioral changes over time.

To address these limitations, future research should consider e.g., First, future studies should consider utilizing larger and more demographically diverse datasets to improve the generalizability and robustness of classification models. Combining data from different age groups, cultures, and socioeconomic backgrounds can provide a more holistic view of the distribution of personality traits and model behaviors across the population (Singh et al., 2024). Second, expanding datasets to include multimodal features such as emotional signals, cognitive test scores, linguistic characteristics of social media texts, or physiological data can significantly improve the depth and accuracy of personality prediction models. This would also allow for better integration of affective and cognitive components of personality, which are often not captured by behavioral metrics alone (Zhu et al., 2024). Third, applying advanced hyperparameter tuning techniques e.g., grid search, Bayesian optimization, or random search can further optimize model performance. While default settings improve consistency, customized configurations can unlock even better predictive capabilities, especially for models like artificial neural networks or ensemble methods (Alibrahim & Ludwig, 2021). Fourth, integrating model interpretability frameworks e.g., SHapley Additive Explanations (SHAP) or Local Interpretable Model-Agnostic Explanations (LIME) will make models more transparent and actionable. This is important in applied fields like psychology and human resource management, where understanding the rationale behind predictions is crucial for ethical and informed decision-making (Parisineni & Pal, 2024).

CONCLUSION

This research presents a comprehensive comparative analysis of nine supervised learning algorithms for the task of human personality classification using structured behavioral data. By evaluating models including conventional classifiers (Logistic Regression, Decision Tree, Naive Bayes, K-Nearest Neighbors, and Support Vector Machine), ensemble methods (Random Forest, Gradient Boosting, AdaBoost), and an Artificial Neural Network (ANN) with backpropagation, the research highlights the capability of machine learning approaches in accurately predicting personality types. The experimental results reveal that ANN, Gradient Boosting, Naive Bayes, and SVM achieved the highest classification performance across all metrics, each attaining an accuracy of 92.93%. These findings demonstrate that both complex architectures and

simpler statistical models can perform equally well under the right data conditions. Moreover, the consistently high precision, recall, and F1-scores across top-performing models affirm the robustness of supervised learning strategies in personality trait identification. Despite the promising outcomes, the research also acknowledges several limitations, including the relatively small dataset, the limited diversity of feature types, and the lack of interpretability frameworks. These factors point to the need for further exploration in future research. As a recommendation, future studies should focus on expanding the dataset, enriching feature dimensions e.g., including emotional, linguistic, and physiological indicators, applying explainable AI methods, and experimenting with time-dependent or hybrid architectures. Such developments will not only enhance classification accuracy but also promote model transparency and real-world applicability in fields e.g., psychology, education, recruitment, and behavioral analytics.

ACKNOWLEDGEMENTS

The author would like to express his deepest gratitude to the Faculty of Vocational, Universitas Negeri Surabaya, which has supported this research through funding and access to laboratory facilities.

REFERENCES

Ahmad, H., Asghar, M. U., Asghar, M. Z., Khan, A., & Mosavi, A. H. (2021). A hybrid deep learning technique for personality trait classification from text. *IEEE Access*, *9*, 146214–146232.

Ahmad, H., Asghar, M. Z., Khan, A. S., & Habib, A. (2020). A systematic literature review of personality trait classification from textual content. *Open Computer Science*, *10*(1), 175–193.

Alibrahim, H., & Ludwig, S. A. (2021). Hyperparameter Optimization: Comparing Genetic Algorithm against Grid Search and Bayesian Optimization. *2021 IEEE Congress on Evolutionary Computation (CEC)*, 1551–1559. https://doi.org/10.1109/CEC45853.2021.9504761

Başaran, S., & Ejimogu, O. H. (2021). A neural network approach for predicting personality from Facebook data. *Sage Open*, *11*(3), 21582440211032156.

Goel, A., Goel, A. K., & Kumar, A. (2023). The role of artificial neural network and machine learning in utilizing spatial information. *Spatial Information Research*, *31*(3), 275–285.

Gutierrez, E., Karwowski, W., Fiok, K., Davahli, M. R., Liciaga, T., & Ahram, T. (2021). Analysis of human behavior by mining textual data: current research topics and analytical techniques. *Symmetry*, *13*(7), 1276.

Jain, V., Wadhwani, K., & Eastman, J. K. (2024). Artificial intelligence consumer behavior: A hybrid review and research agenda. *Journal of Consumer Behaviour*, *23*(2), 676–697.

Koo, M., & Yang, S.-W. (2025). Questionnaire Use and Development in Health Research. *Encyclopedia*, *5*(2), 65.

Lin, Q., Hu, Z., & Ma, J. (2024). The personality of the intelligent cockpit? exploring the personality traits of in-vehicle llms with psychometrics. *Information*, *15*(11), 679.

Liu, D., Feng, X. L., Ahmed, F., Shahid, M., & Guo, J. (2022). Detecting and measuring depression on social media using a machine learning approach: systematic review. *JMIR Mental Health*, *9*(3), e27244.

Megahed, F. M., Chen, Y.-J., Jones-Farmer, L. A., Rigdon, S. E., Krzywinski, M., & Altman, N. (2024). Comparing classifier performance with baselines. *Nature Methods*, *21*(4), 546–548.

Mohan, G. B., Kumar, R. P., & Gorantla, S. (2023). Enhancing personality classification through textual analysis: a deep learning approach utilizing MBTI and social media data. *2023 International Conference on Network, Multimedia and Information Technology (NMITCON)*, 1–6.

Naz, A., Khan, H. U., Bukhari, A., Alshemaimri, B., Daud, A., & Ramzan, M. (2025). Machine and deep learning for personality traits detection: a comprehensive survey and open research challenges. *Artificial Intelligence Review*, *58*(8), 239.

Parisineni, S. R. A., & Pal, M. (2024). Enhancing trust and interpretability of complex machine learning models using local interpretable model agnostic shap explanations. *International Journal of Data Science and Analytics*, *18*(4), 457–466. https://doi.org/10.1007/s41060-023-00458-w

Pratama, M. D., Abdillah, R., & Haq, D. Z. (2024). Water Quality Identification Using Ensemble Machine Learning and Hybrid Resampling SMOTE-ENN Algorithm. *Fountain of Informatics Journal*, *9*(2), 83–91. https://doi.org/http://dx.doi.org/10.21111/fij.v9i2.12489

Pratama, M. D., Azizah, A. N., Asy'ari, M. F., Syahputra, D. N. A., Mashuri, M. A. I., Kholifah, B., Abdillah, R., Pratiwi, A. P., & Haq, D. Z. (2025). Enhancing Clickbait Headline Identification Performance Without Preprocessing Through Feature Reduction and Sentiment Analysis. *Journal of Applied Informatics Research*, *1*(01), 30–44.

Pratama, M. D., Sarno, R., & Abdullah, R. (2022). Sentiment Analysis User Regarding Hotel Reviews by Aspect Based Using Latent Dirichlet Allocation, Semantic Similarity, and Support Vector Machine Method. *International Journal of Intelligent Engineering & Systems*, *15*(3).

Raharjo, A. B., Pratama, M. D., & Purwitasari, D. (2023). Ensemble Oversampling For Financial Fraud Classification Of Imbalanced Data. *IPTEK The Journal for Technology and Science*, *34*(3), 175–185.

Rahman, A. U., & Halim, Z. (2022). Predicting the big five personality traits from hand-written text features through semi-supervised learning. *Multimedia Tools and Applications*, *81*(23), 33671–33687.

Semwal, R., Tripathi, N., Rana, A., Pandey, U. K., Parihar, S., & Bairwa, M. K. (2024). AI-Driven Insights: Enhancing Personality Type Prediction with Advanced Machine Learning Algorithms. *2024 7th International Conference on Contemporary Computing and Informatics (IC3I)*, *7*, 815–822.

Shu, Z., Sun, X., & Cheng, H. (2024). When llm meets hypergraph: A sociological analysis on personality via online social networks. *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, 2087–2096.

Singh, L., Barokova, M. D., Baumgartner, H. A., Lopera-Perez, D. C., Omane, P. O., Sheskin, M., Yuen, F. L., Wu, Y., Alcock, K. J., & Altmann, E. C. (2024). A unified approach to demographic data collection for research with young children across diverse cultures. *Developmental Psychology*, *60*(2), 211.

Utami, N. A., Maharani, W., & Atastina, I. (2021). Personality classification of facebook users according to big five personality using SVM (support vector machine) method. *Procedia Computer Science*, *179*, 177–184.

Wang, X., Sui, Y., Zheng, K., Shi, Y., & Cao, S. (2021). Personality classification of social users based on feature fusion. *Sensors*, *21*(20), 6758.

Zhu, X., Guo, C., Feng, H., Huang, Y., Feng, Y., Wang, X., & Wang, R. (2024). A review of key technologies for emotion analysis using multimodal information. *Cognitive Computation*, *16*(4), 1504–1530.