National Conference on Behavioural Economics and Intelligent Decision Systems for Climate Change and Sustainable Development
Int. J. Adv. Res. in Artificial Intelligence and Machine Learning Reviews
Volume 2, Issue 1 - 2026

Research Article

Evaluating Machine Learning Algorithms for Automated Personality Judgement

Rajiv Kumar[1], Gurvinder Singh[2], Amandeep Kaur[3], Prabhpreet Kaur[4]

[1]Research Scholar, [2]Professor, [3,4]Assistant Professor, Department of Computer Science, Guru Nanak Dev University, Amritsar, India

DOI: https://doi.org/10.24321/3117.4809.202602

I N F O

Corresponding Author:
Rajiv Kumar, Department of Computer Science, Guru Nanak Dev University, Amritsar, India
E-mail Id:
rajivcsc.rsh@gndu.ac.in
Orcid Id:
http://orcid.org/0000-0002-9778-6569

How to cite this article:
Kumar R, Singh G, Kaur A, Kaur P. Evaluating Machine Learning Algorithms for Automated Personality Judgement. *Int J Adv Res Artif Intell Mach Learn Rev* 2026; 2(1): 148-154.

Date of Submission: 2025-10-04
Date of Acceptance: 2025-10-28

A B S T R A C T

Traditional methods of personality judgement, such as self-report questionnaires and manual assessments, are often limited by subjectivity, time consumption and vulnerability to social desirability bias. These drawbacks highlight the need for automated and data-driven techniques that can provide more objective and scalable personality evaluation. In this study, we explore the use of machine learning (ML) algorithms to predict personality traits and systematically compare their performances. Models including linear regression, decision tree, random forest, support vector machine (SVM) and AdaBoost are implemented on a benchmark dataset. The algorithms are evaluated using standard metrics such as accuracy, precision, recall and F1-score to ensure a comprehensive analysis. Results reveal distinct strengths and weaknesses across classifiers, offering insights into the most effective approaches for personality judgement. The findings demonstrate the potential of ML in advancing personality assessment and provide a foundation for building reliable, interpretable and scalable solutions. Such approaches can be applied in domains like human resource management, education and mental health, where accurate personality insights are essential for informed decision-making and personalised interventions.

Keywords: Machine Learning, Logistic Regression, SVM, Decision Tree, F1-score

Introduction

Personality of a person encircles every aspect of life. It describes the pattern of thinking, feeling and characteristics that predict and describe an individual's behaviour and also influence daily life activities, including emotions, preferences, motives and health. Personality judgement has long been a cornerstone of psychological science and applied fields such as personnel selection, counselling and education. Machine learning (ML) has become an essential tool for modelling and understanding human personality judgement. By learning from large volumes of behavioural, linguistic, visual and physiological data, ML algorithms can identify complex and often non-linear patterns associated with personality traits such as the Big Five dimensions (openness, conscientiousness, extraversion, agreeableness, and neuroticism, etc.).[1]

In an increasingly data-driven era, the ability to judge or infer personality traits has taken on renewed importance across multiple domains ranging from human resources and education to mental health andartificial intelligence systems.[2] Recent research underscores several reasons why personality judgement is especially relevant. Personality judgement continues to show strong relationships with important life outcomes. For example, studies reveal that the Big Five traits outperform or match cognitive measures in predicting non-educational outcomes such as job performance, health and well-being, even when accounting for socio-demographic variables.

In human resources and people analytics, personality judgements are being used to complement or augment traditional hiring tools. As resume authenticity becomes harder to verify (with the rise of AI-generated documents), personality assessments provide behavioural or trait-based signals that are harder to fake.[3]

Literature Review

Table 1.Systematic Literature Review

References	Author and Year	Explanation	Data/Source Type	Outcomes/ Research Gap
4	Stachl et al. (2020)	Analyzed smartphone sensor data (GPS, app usage) to predict Big Five personality traits with Random Forest and regression.	Smartphone sensors (GPS, app usage)	It showed potential for digital behavior-based personality inference but limited by short data collection periods and cultural dependence.
5	Peltonen et al. (2020)	Linked app usage categories to personality traits using SVM and regression.	App usage logs	It provided context-dependent and culturally sensitive results.
6	Rüegger et al. (2020)	Studied personality states via experience sampling and smartphone features.	Experience sampling + smartphone features	It addressed static trait limitations suggesting need for time-sensitive and adaptive personality models.
7	Marengo et al. (2023)	Meta-analysis of 21 smartphone-based prediction studies; extraversion most consistent.	Meta-analysis (21 studies)	High heterogeneity across studies indicates a lack of standardization in data features and modeling pipelines.
8	Currey et al. (2023)	Used mindLAMP app combining surveys and passive sensing for mental health forecasting.	MindLAMP app (surveys + sensing)	It depicted predictive potential of personality features in mental health forecasting. However results relied on clinical samples with limited generalization.
9	Guo et al. (2024)	Compared dialogue-based prediction across task-oriented vs open-domain contexts.	Dialogue transcripts	Improved accuracy through NLP and transformer models but faced poor generalization and interpretability challenges.
11	Shum et al. (2025)	Fine-tuned BERT/ RoBERTa on Reddit (PANDORA dataset) for improved accuracy.	Reddit + PANDORA dataset	Dependent on social media data; potential bias

ISSN: 3117-4809
DOI: https://doi.org/10.24321/3117.4809.202602

Methodology

Dataset

The dataset name 'personality_dataset.csv' used in this study was obtained from Kaggle's publicly available data repository. It contains responses to standardised personality questionnaires aligned with the Big Five personality dimensions (openness, conscientiousness, extraversion, agreeableness, and neuroticism). In total, the dataset includes demographic variables, questionnaire responses, and corresponding labels for each personality dimension. These labels were adopted as the ground truth for supervised learning tasks. The dataset was pre-cleaned, but further processing was conducted to ensure consistency and accuracy. The schematic diagram of methodology is shown in Figure 1.

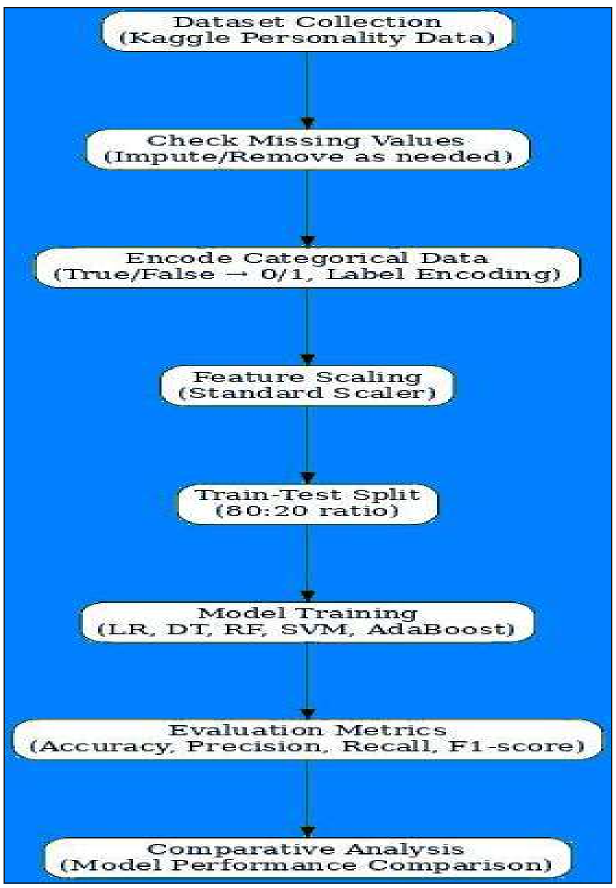

Figure 1.Schematic Working of Methodology

Data Pre-processing

Several preprocessing steps were applied to prepare the data for modelling. First, the dataset was checked for missing or null values column by column using functions such as isnull().sum(). Categorical variables were then encoded where required. Binary values originally represented as "True/False" were mapped to numeric values (0/1), and multi-class categorical variables were transformed using label encoding. To ensure uniform scaling across features, the Standard Scaler technique was employed for transforming all numerical values to zero mean and unit variance. This step prevented scale-sensitive models such as SVM and ensemble methods from being biased toward features with larger ranges. The dataset was then divided into training and testing sets using an 80:20 stratified split, preserving class balance across personality dimensions.

Machine Learning Models

Five machine learning algorithms were implemented to predict personality dimensions:

- **Logistic Regression (LR):** Logistic regression is a classical statistical method used for binary (or multiclass) classification, modelling the probability that a given input belongs to a particular class via the logistic (sigmoid) function. It assumes a linear relationship between features and log-odds. It is interpretable, robust baseline, but limited for complex non-linear boundaries.[12]
- **Decision Tree (DT):** A decision tree is a non-parametric supervised learning method that recursively partitions the feature space, splitting on feature-value tests to separate classes (or predict numeric values). It is widely used due to its interpretability and ability to handle categorical/continuous features. Starting from the root, the algorithm selects a feature and split threshold that best improves a purity metric (e.g. Gini impurity, information gain, or entropy). The process continues recursively until stopping criteria (max depth, minimum samples etc.[13]
- **Random Forest (RF):** Random Forest is an ensemble "bagging" method that builds a large number of decision trees and aggregates their predictions (e.g., by majority vote for classification, averaging for regression). This method is robust and less prone to overfitting compared to a single decision tree. It works well in practice for many tasks, handles high-dimensional data and missing values.[14]
- **Support Vector machine (SVM):** A support vector machine is a discriminative classifier that seeks an optimal hyperplane separating classes by maximising the margin (distance) between the hyperplane and the nearest data points (support vectors). SVMs are common in text classification, image recognition, bioinformatics, etc.[15]
- **AdaBoost (Adaptive Boosting):** AdaBoost is one of the earliest boosting ensemble algorithms. It combines many weak learners (often simple classifiers like decision stumps) sequentially to form a strong classifier. It can significantly improve performance over individual weak learners by correcting mistakes iteratively.[16]

ISSN: 3117-4809
DOI: https://doi.org/10.24321/3117.4809.202602

All models were trained under the same experimental conditions. Hyper parameters were tuned using grid search with cross-validation to optimise performance.

Parameters Tuning

Parameter tuning, also known as hyperparameter optimisation, refers to the process of selecting the most appropriate configuration settings for various machine learning (ML) algorithms to achieve optimal performance on a given dataset. Table 2 describes various parameters used along with their value in each of the applied ML algorithms in this work.

Table 2.Parameters used for model training

Sr. No.	Model	Parameters with value
1.	Decision Tree	criterion="gini", max_depth=None
2	Random Forest	n_estimators=100
3.	SVM	kernel='linear', probability=True
4.	AdaBoost Classifier	estimator=' ExtraTreesClassifier', n_estimators=50, learning_rate=1.0

Variations in the values of these parameters could result in various outcomes. After doing a sufficient number of experiments, the above-mentioned values were determined to be the most appropriate and produce optimal outcomes.

Feature Correlation

Feature correlation refers to the statistical relationship between two or more features (variables) in a dataset. It measures how changes in one feature are associated with changes in another. A positive correlation means both features increase or decrease together. A negative correlation means when one feature increases, the other decreases. A zero or weak correlation indicates little or no linear relationship.

In machine learning, correlation analysis helps identify redundant or highly related features, which can then be reduced or removed to improve model performance and avoid multicollinearity. A detailed feature analysis has also been performed for our work. The pictorial representation of this analysis has been shown in figure 2.

It has been clearly stated that all features of our datasets are correlated in some way. For example, if the friends_circle_size of any person got increased, then the social_event_attendance of that person would also be increased. It showed a positive correlation in these two features.

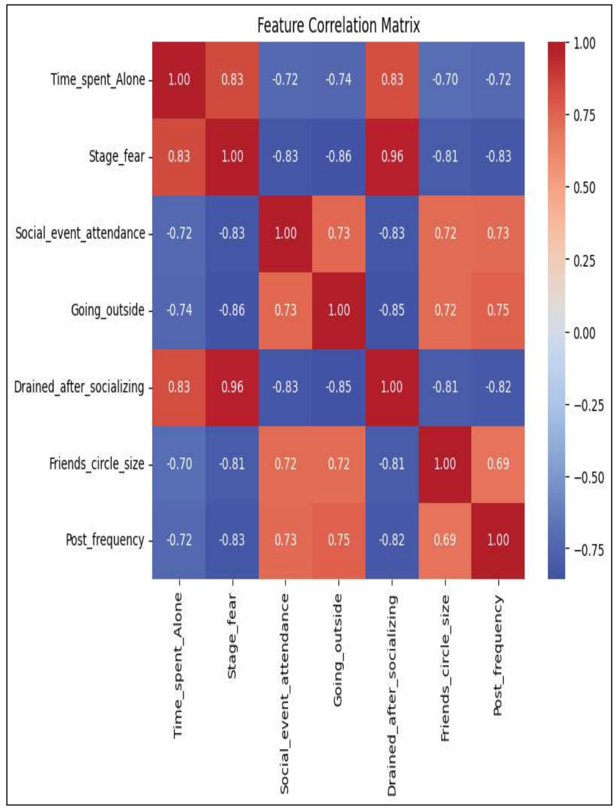

Figure 2.Feature Correlation Matrix

Evaluation Metrics

The performance of the models was evaluated using multiple standard metrics. Accuracy was computed to assess overall correctness, while precision, recall, and F1-score were calculated to provide a detailed analysis view of model behaviour across different personality classes. Weighted averages were reported to handle class imbalance within the dataset. Cross-validation scores were also used to assess robustness.

Comparative Analysis

The comparative analysis was performed by evaluating all models on the same dataset under consistent settings. Their performance was compared across all evaluation metrics.

Experiments and Results Analysis

This experiment used the widely useful and accessible dataset 'personality_dataset' from the Kaggle repository. This dataset included seven input columns and a single output column. A total of 2900 rows were used for model training. These studies used a variety of commonly used Python libraries, including pandas, sklearn, seaborn, and matplotlib. Each modal performance was also evaluated using several performance criteria, such as accuracy, F1-score, precision, and recall. Table 1 compares the assessment metrics for each ML algorithm.

ISSN: 3117-4809
DOI: https://doi.org/10.24321/3117.4809.202602

Table 3.Comparison of Performance Evaluation Metrics

-	LR	DT	RF	SVM	Ada Boost
Accuracy	92.93%	86.72%	92.41%	92.93%	93.10%
F1-Score	92.79%	86.78%	92.44%	92.96%	91.61%
Precision	90.72%	86.72%	92.41%	92.93%	94.24%
Recall	94.96%	86.73%	92.42%	92.93%	92.91%

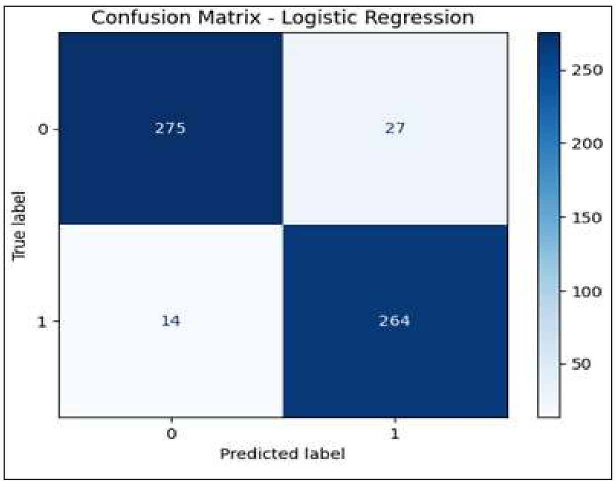

Figure 3.Confusion matrix of Logistic Regression

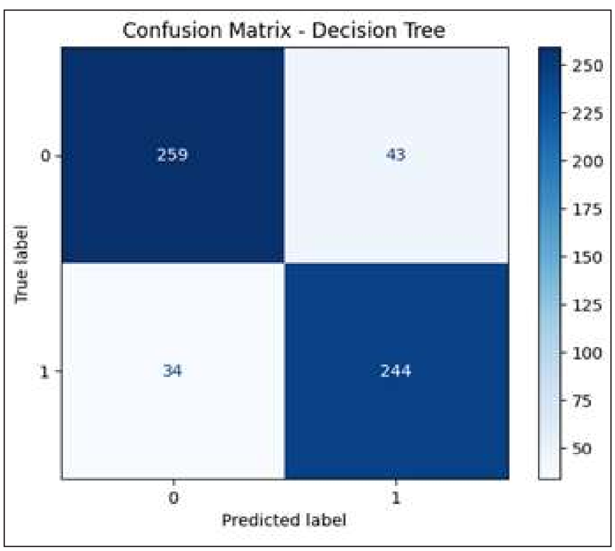

Figure 4.Confusion matrix of Decision Tree

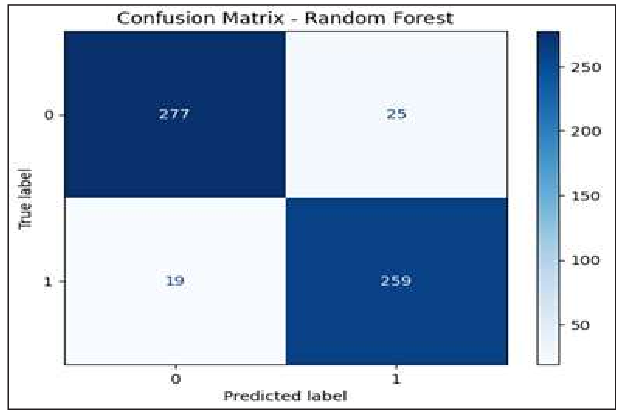

Figure 5.Confusion matrix of Random Forest

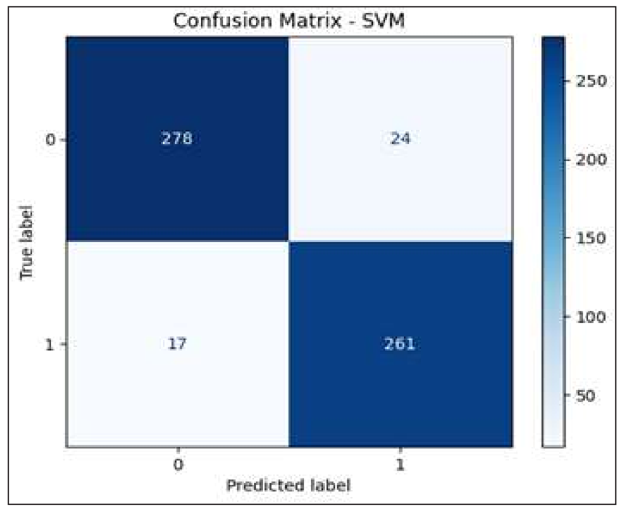

Figure 6.Confusion matrix of Support vector machine

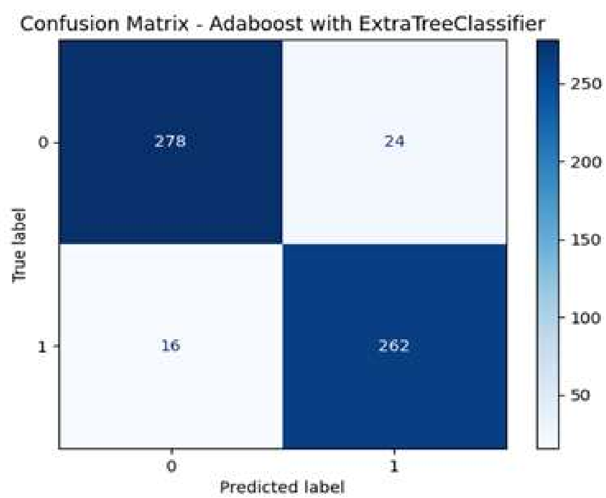

Figure 7.Confusion matrix of AdaBoost

ISSN: 3117-4809
DOI: https://doi.org/10.24321/3117.4809.202602

Further, confusion matrices have been also shown for each ML technique.

Accuracy is the percentage of correctly categorised samples (both positive and negative) among all samples. AdaBoost with Extra Tree classifiers has been shown to be the most accurate. Precision, on the other hand, is the fraction of accurately predicted positive samples among all samples projected by the model to be positive. AdaBoost also has the highest value of precision. Recall (also known as sensitivity or true positive rate) is the proportion of positive samples that the model properly detects as positive. In this study, logistic regression produces the highest recall value. The F1-score indicates how well a model performs when we need to balance precision and recall, particularly in imbalanced datasets. So, in this scenario, SVM scored the highest value.

Interpretability trade-offs between Methods

The interpretability trade-offs between large ensemble models (like Random Forest and AdaBoost) and simpler methods (such as logistic regression and decision trees) are clear in this work. Complex ensemble models such as Random Forest and AdaBoost improve accuracy and robustness by integrating numerous weak or base learners to capture non-linear and high-dimensional correlations in data. However, this improvement in performance comes at the expense of decreased interpretability since the internal decision-making process becomes opaque and difficult to track or explain. Simpler models such as logistic regression and decision trees are more visible and interpretable, allowing researchers to see the value of features, decision rules and weight coefficients. This transparency makes them ideal when model explainability and ethical accountability are important.

Conclusion

This study demonstrated the effectiveness of machine learning techniques in predicting personality traits among people. By systematically evaluating logistic regression, decision tree, random forest, SVM and AdaBoost on the Kaggle personality dataset, the results highlighted the distinct strengths of each model. While AdaBoost achieved the highest overall accuracy and precision, logistic regression produced the best recall, and SVM achieved the strongest F1-score, showing the trade-offs between different approaches. These findings confirm that no single algorithm universally outperforms others across all evaluation metrics. Instead, the choice of model should be guided by the specific requirements of the application, such as prioritising sensitivity, interpretability or overall accuracy.

Limitations / Ethical Considerations

Automated personality prediction using machine learning involves key ethical and practical challenges. Privacy concerns arise from using sensitive behavioural data requiring consent and secure handling. Model bias may lead to unfair predictions across demographic groups, while limited interpretability of complex models reduces transparency. Results based on a single dataset may not generalise across contexts. Therefore, responsible data use, fairness and explainable modelling are essential for ethical and reliable deployment. According to the research, complicated ensemble models that capture non-linear personality patterns, such as Random Forest and AdaBoost, yield improved accuracy and robustness at the expense of interpretability. Unlike more straightforward models like decision trees and logistic regression, which may ignore intricate correlations but are nonetheless more apparent. Their decision-making processes are hard to describe. Ensemble approaches, however, might be at risk of overfitting due to the moderate size of the dataset (about 2,900 samples), as they are able to capture noise and small variations in the training data. Simpler models emphasise the trade-off between model complexity, interpretability and generalisation. They are often more stable and less prone to overfitting despite being less strong.

References

1. Koutsoumpis, Antonis. “Psychometric properties of personality assessment using machine learning.” Current opinion in psychology vol. 65 (2025): 102093. doi:10.1016/j.copsyc.2025.102093.
2. C. Stachl et al., “Predicting personality from patterns of behavior collected with smartphones,” Proceedings of the National Academy of Sciences, vol. 117, no. 30, pp. 17680–17687, 2020, doi: 10.1073/pnas.1920484117.
3. Grunenberg E, Peters H, Francis MJ, Back MD, Matz SC. Machine learning in recruiting: predicting personality from CVs and short text responses. Frontiers in Social Psychology. 2024 Jan 29;1:1290295.
4. Stachl, C., Au, Q., Schoedel, R., Gosling, S. D., Harari, G. M., Buschek, D., … &Bühner, M. (2020). Predicting personality from smartphone data: A meta-analysis of digital footprints. European Journal of Personality, 34(6), 701–721. https://doi.org/10.1002/per.2299.
5. Peltonen, J., Miettinen, M., &Lagerspetz, E. (2020). App usage patterns and personality traits: Insights from large-scale smartphone data. Computers in Human Behavior, 109, 106–121. https://doi.org/10.1016/j.chb.2020.106121
6. Rüegger, D., Mehl, M. R., & Harari, G. M. (2020). Capturing personality states through smartphone sensing and experience sampling. European Journal of Personality, 34(5), 753–770. https://doi.org/10.1002/per.2284.
7. Marengo, D. (2023). Best practices and methodological issues in smartphone-based personality prediction research. Behavior Research Methods, 55(2), 481–496. https://doi.org/10.3758/s13428-022-01898-7.

ISSN: 3117-4809
DOI: https://doi.org/10.24321/3117.4809.202602

8. Currey, J., Nahum-Shani, I., &Onnela, J. P. (2023). Predicting mental health symptom improvement using personality features from smartphone data: Evidence from the mindLAMP platform. Journal of Medical Internet Research, 25(5), e42321. https://doi.org/10.2196/42321.
9. Guo, Y., Chen, Y., & Xu, H. (2024). Cross-domain generalization in personality prediction from dialogue data: A comparative analysis. Frontiers in Artificial Intelligence, 7, 144–159. https://doi.org/10.3389/frai.2024.0144.
10. Shum, K., Huang, L., & Li, F. (2025). Transformer-based approaches for personality prediction using user comments. IEEE Transactions on Affective Computing. Advance online publication. https://doi.org/10.1109/TAFFC.2025.1234567.
11. Naz, S. (2025). A comprehensive review on machine learning and deep learning techniques for personality prediction. Artificial Intelligence Review. Advance online publication. https://doi.org/10.1007/s10462-025-10645-1.
12. Geng Y, Li Q, Yang G, Qiu W. Logistic regression. InPractical Machine Learning Illustrated with KNIME 2024 Aug 30 (pp. 99-132). Singapore: Springer Nature Singapore.
13. Klusowski JM, Tian PM. Large scale prediction with decision trees. Journal of the American Statistical Association. 2024 Jan 2;119(545):525-37.
14. Salman HA, Kalakech A, Steiti A. Random forest algorithm overview. Babylonian Journal of Machine Learning. 2024 Jun 8;2024:69-79.
15. Veisi H. Introduction to SVM. InLearning with Fractional Orthogonal Kernel Classifiers in Support Vector Machines: Theory, Algorithms and Applications 2023 Mar 19 (pp. 3-18). Singapore: Springer Nature Singapore.
16. Ding Y, Zhu H, Chen R, Li R. An efficient AdaBoost algorithm with the multiple thresholds classification. Applied sciences. 2022 Jun 9;12(12):5872.

ISSN: 3117-4809
DOI: https://doi.org/10.24321/3117.4809.202602