JESICA: Journal of Enhanced Studies in Informatics and Computer Applications
Journal homepage: jesica.itsk-soepraoen.ac.id
e-ISSN: 3046-6938 p-ISSN: 3046-6997
Vol. 2, No. 1, February 2025, Pp. 1-7

Comparing Discriminant Analysis Function for Early Prediction of Smartphone Addiction

Mufid Musthofa[1], Mochammad Anshori*[2]
1 Brawijaya University, Indonesia
2 Institut Teknologi, Sains, dan Kesehatan RS.DR. Soepraoen Kesdam V/BRW, Indonesia

***Corresponding author**
E-mail address:
moanshori@itsk-soepraoen.ac.id

Keywords:
Smartphone addiction, health informatics, LDA, QDA, machine learning, discriminant analysis

Abstract
The pervasive use of smartphones in daily life has led to significant benefits, but excessive use has caused alarming behavioral and health issues, particularly among adolescents. Addressing smartphone addiction requires early detection to enable timely interventions. This study investigates the application of machine learning, specifically Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA), for the early prediction of smartphone addiction. The research used a dataset containing 394 instances categorized into "addicted" and "non-addicted" classes. Dataset is derived from questionnaire responses. After preprocessing steps, including feature selection and ordinal encoding, the data was split using 10-fold cross-validation to ensure robust evaluation. The models were assessed using metrics such as accuracy, precision, recall, and F-measure. Results indicate that LDA significantly outperforms QDA across all metrics, achieving an accuracy of 94.16%, a precision of 94.2%, a recall of 94.2%, and an F-measure of 94.2%. Additionally, the Receiver Operating Characteristic (ROC) curve analysis showed an Area Under the Curve (AUC) of 0.9875 for LDA, indicating its high reliability and stability in classifying smartphone addiction. QDA, while effective, has a slightly lower performance due to the linear separability of the dataset. This study concludes that LDA is a robust and effective method for early prediction of smartphone addiction, offering valuable insights for health monitoring systems. The findings provide a foundation for future applications of discriminant analysis in addressing behavioral health issues.

1. Introduction

Smartphones have become an essential necessity in the digital era due to their multifunctionality, akin to computers but in a more compact form. These devices offer numerous benefits for communication, information access, education, and entertainment [1]–[4]. However, the excessive use of smartphones can have negative consequences, such as behavioral changes, including gambling disorders and prolonged internet use [5]. The extensive features provided by smartphones often lead to overuse and dependency.

The global adoption of smartphones has surged over the past decade. In 2012, there were approximately 1.08 billion users worldwide, and this number increased to 1.3 billion devices by 2015. Indonesia ranked fifth as the largest smartphone user globally during this period [3], illustrating the widespread acceptance of these devices. As smartphones simplify daily tasks, the number of users continues to grow, especially among young adults and students. However, this increasing dependence on smartphones has raised concerns about misuse of technology, particularly addiction among young users [6].

Excessive smartphone use has significant implications on both physical and mental well-being. In Indonesia, internet penetration has reached 27.7 billion. 80% of internet users access it via smartphones. Research shows that excessive smartphone use has a negative impacts on students' academic performance and physical health, contributing to decreased muscle mass, lack of physical activity, stress, timidity, and sleep disturbances [7], [8]. Alarmingly, these issues are also experienced by children who are greatly exposed to smartphones due to their parents' busy schedules and relied on technology as a distraction [7]. Such parenting practices inadvertently causes smartphone addiction among children [9].

Addiction to smartphones among teenagers has been connected to mental health problems such anxiety, sleeplessness, and changes in social behavior [9], [10]. Medical issues such as computer vision syndrome, wrist pain, and "smartphone thumb" might result from excessive smartphone use [11]. Furthermore, frequent use of smartphone by young people have been linked to decreased levels of self-confidence [12]. Reducing and overcoming smartphone addiction is essential due to the serious consequences.

In order to reduce the negative effects of smartphone addiction, it is crucial to recognize the signs of addiction early on, especially in kids and adolescents whose actions might be difficult for parents to understand [2]. It is essential to predict smartphone addiction in children since it enables prompt intervention to solve this problem. In order to anticipate smartphone addiction in early teens, this study suggests a machine learning (ML)-based method. Previous studies have demonstrated the efficacy of machine learning algorithms, achieving remarkable precision in detecting smartphone addiction. For instance, Decision Trees achieved an accuracy of 78.74% [11], Random Forest reached an accuracy of 82.59% [6], SVM with a polynomial kernel achieved an accuracy of 73% [4], LVQ3 achieved an accuracy of 96.79% [9], and Decision Tree with an accuracy of 84% [13].

This illustrates how smartphone addiction can be reliably predicted by machine learning. In order to develop a prediction model, the author of this study conducted and experiment using discriminant analysis. Discriminant analysis, particularly linear and quadratic discriminant analysis (LDA and QDA), is the machine learning technique used in this work. Discriminant analysis is a preferred machine method for dimensional reduction that is also capable of doing classification tasks and is able to handle binary and even multiclass tasks. LDA is a technique that is increasingly being used to decrease high dimensional data. Nonetheless, LDA can be applied to two different data classifications. The ability of LDA to produce exceptionally accurate results on linearly separated data is one of its intriguing features [14]. Although QDA and LDA are similar, QDA is utilized for classification on nonlinearly separated data [15]. There are numerous instances in which the healthcare industry can benefit from the effectiveness of discriminant analysis. We will compare the capabilities of both methods to see which classifier performs the best. The main objective of this study is to use discriminant analysis, specifically LDA and QDA, to do an early identification of smartphone addiction.

2. Research Method

This section will be explaining the research methodology of this research. The flow is shown in the figure below. As illustrated in Figure 1, this study will generally be carried out in four stages. Obtaining a dataset is the first phase, which is followed by preprocessing it to make it usable, creating a discriminant analysis model, testing it, and evaluating it to identify the best model. The specifics will be discussed more profoundly below.

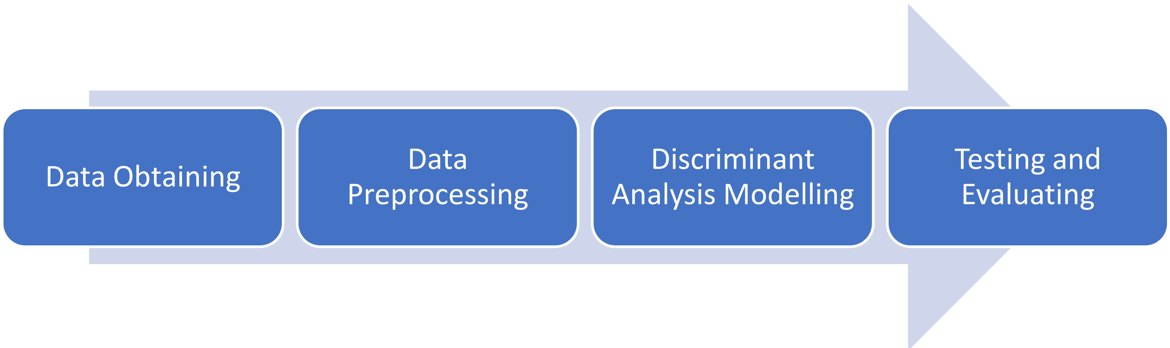

Figure 1. Research methodology

The secondary data used comes from the Mendeley dataset [16]. The dataset is divided into two categories: yes and no addiction in smartphones. It consists of 14 features that are derived from questionnaire questions.

The dataset is then used as input for the preprocessing step. Some features will be removed because they are irrelevant. Eliminating irrelevant features also reduces the dimensional dataset, which may improve classifier performance and minimize computation time. Since we are aware that the data contains categorical values, encoding is also used. We can convert data into numerical values with the use of data encoding. Ordinal encoding is used in this stage [17]. Each categorical is given an integer value as part of the process. The data is now prepared for use in discriminant analysis to build the classification model.

Data will be divided into test and training sets before the classifier model is built. K-fold cross validation with k = 10 is the technique used. The method of dividing data evenly in this study uses k=fold cross validation. The mechanism is folding the data according to a calculated k value. One-fold is used for training, and the other fold is used for testing. The fold of data testing will be shifted iteratively up to k times. The benefit of cross validation is that the model yields more generic results [18] and can improve classifier performance by detecting overfitting [19], [20]. The discriminant analysis function must then be constructed to produce the classifier. Reducing homogenous data and increasing diverse data is the fundamental idea behind discriminant analysis [21]. The two variables that will be compared are the quadratic and linear discriminant analyses.

$$\delta(x) \coloneqq 2(\Sigma^{-1}(\mu_2 - \mu_1))^T x + ((\mu_1 - \mu_2)^T(\Sigma^{-1}(\mu_1 - \mu_2)) + 2\,ln(\frac{\pi_2}{\pi_2}) \quad (1)$$

$$\delta(x) \coloneqq x^T(\Sigma_1 - \Sigma_2)^{-1}x + 2(\Sigma_2^{-1}\mu_2 - \Sigma_1^{-1}\mu_1)^T x + (\mu_1^T\Sigma_1^{-1}\mu_1 - \mu_2^T\Sigma_2^{-1}\mu_2) + \ln\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right) + 2\ln\left(\frac{\pi_2}{\pi_2}\right) \quad (2)$$

$$\hat{C}(x) = \begin{cases} 1, & if\,\delta(x) < 0 \\ 2, & if\,\delta(x) > 0 \end{cases} \quad (3)$$

Equation (1) displays the LDA formula for the binary class, whereas Equation (2) displays the QDA formula and Equation (3) displays the label decision [22]. In LDA, decision-making boundaries are shown by straight lines since it is expected that the data classes have equal covariance matrix. In QDA, on the other hand, pseudo lines in the form of a quadratic divide the borders between differentiating data classes. Covariance matrix, which are the foundation of LDA and QDA, are formed in the preceding formulation by calculating the mean vector for each class. Then, the data point will be categorized into the designated class using Equations (1) and (2). LDA and QDA differ in that LDA works well with linearly separated variables, while QDA works well with nonlinear combinations of independent variables [23].

Following the development of the discriminant analysis model, the last step involves conducting research based on experimental scenarios. The results of the experiment will be evaluated, and a thorough analysis will be conducted. Accuracy, precision, recall, and F-measure are the metrics used in evaluation. The best model will be identified by comparing these evaluation measures.

$$acc = \frac{true\ positive + true\ negative}{n\ data} \quad (4)$$

$$prec = \frac{true\ possitive}{true\ positive + false\ positive} \quad (5)$$

$$rec = \frac{true\ positive}{true\ positive + false\ negative} \quad (6)$$

$$F\ measure = 2\,\frac{prec \cdot rec}{prec + rec} \quad (7)$$

Equation (4) provides an accurate estimate of how well the model fits the data. Equation (5) provides a precise description of the ratio of correctly anticipated observations to all positive observations. In order to assess the number of true positives that the model predicts, equation (6) displays recall and sensitivity. Equation (7) displays the F measure, which displays a harmonic value between recall and precision [24]. The F-measure can be applied when there are uneven classes in the data distribution.

3. Results and Discussion

The results of the experiment will be presented in this section. First, the data that has been obtained will be examined further in a descriptive analysis. As can be seen, there are 394 data instances. The class distribution is shown below.

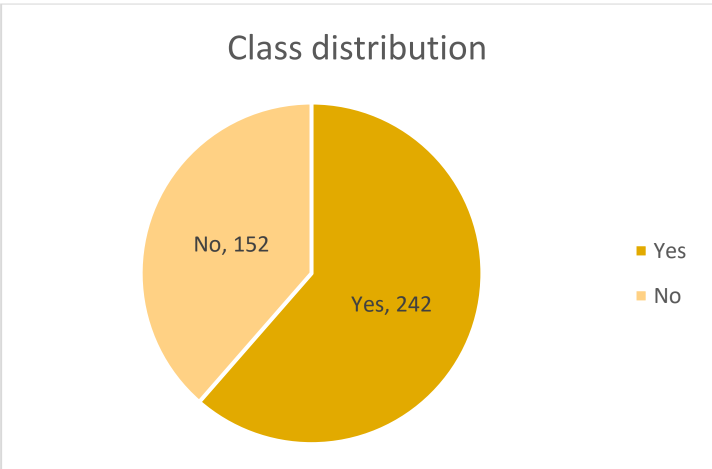

Figure 2. Dataset class distribution

The class distribution is shown in Figure 2. Addiction to smartphone can be classified as either yes or no. There are 152 data points in the No class, representing roughly 39% of the total data. 242 data points or around 61% of the total data indeed make up the class. The F-measure is preferred in this situation because it indicates whether the class distribution is unbalanced. After that, some features will be eliminated since they are irrelevant to the dataset.

Table 1. Dataset details

Feature	Description	Data type
No	Number of respondent	Numeric
Gender	Gender	Categorical
1	Question 1	Numeric
2	Question 2	Numeric
3	Question 3	Numeric
4	Question 4	Numeric
5	Question 5	Numeric
6	Question 6	Numeric
7	Question 7	Numeric
8	Question 8	Numeric
9	Question 9	Numeric
10	Question 10	Numeric
Total	Sum of each answer	Numeric
Result	Class target	Binary

Details of the dataset are displayed in Table 1. Qualities No and Total will be eliminated since they are irrelevant to the dataset. On the other hand, the total is the sum of all the questions. Each response to the ten questions uses a likert scale. Numerical values are the data type as a result. The dataset contains categorical data types; encoding will be used to convert these to numerical data types. After that, discriminant analysis is carried out utilizing k=10-fold cross validation. Weka is being used in this project as a tool for data mining [25].

Table 2. Discriminant evaluation comparison

Model	Acc	Prec	Rec	F-Measure
LDA	0,941624	0,942	0,942	0,942
QDA	0,908629	0,908	0,909	0,908

The evaluation of LDA and QDA performance is shown in Table 2. Table 1 demonstrates that LDA outperforms QDA. The accuracy, precision, recall, and F-measure of LDA are 0.941624, 0.942, and 0.942, respectively. Compared to LDA, the QDA assessment value is lower. The accuracy increases by 0.908629, with a gap of around 0.033. The precision, recall, and F-measure of the QDA reached 0.908, 0.909, and 0.908, respectively. It has been determined whether LDA is the best model based on the assessment table. Due to the linear separation of the smartphone addiction dataset, QDA produces a lower result in this instance.

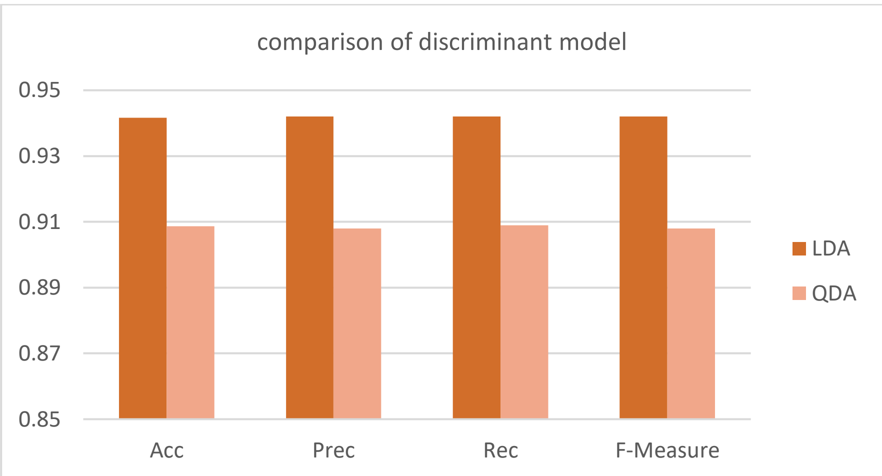

Figure 3. Graphic comparison of LDA and QDA

Figure 3 shows that LDA dominated the high outcome. LDA performs better than QDA in terms of accuracy, precision, recall, and F-measure. It also produces more stable results than QDA. All evaluation measures have a difference of roughly 0.034. Since there isn't much of a difference between the evaluation results, LDA is more dependable than QDA. It is not possible to fit the dataset separately using the quadratic function. According to the performance result above, LDA is typically used in this research for additional discussion.

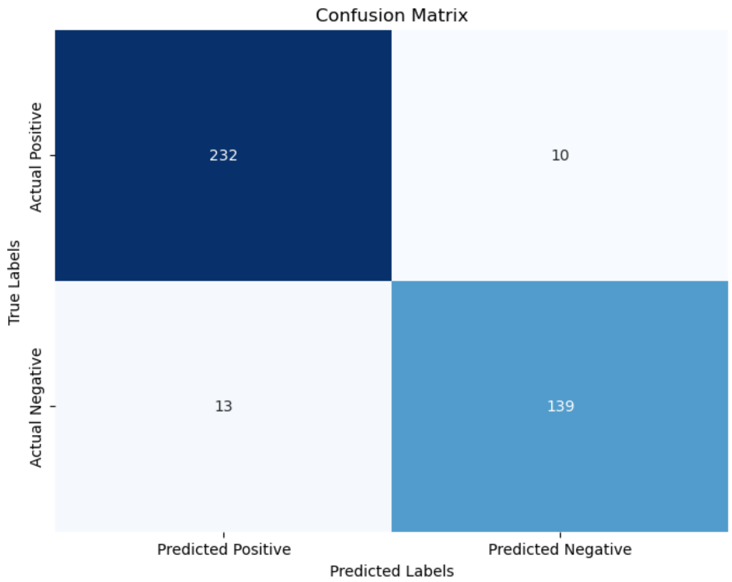

Figure 4. Confusion matrix of LDA

The LDA confusion matrix is displayed in Figure 4. A matrix that summarizes the LDA classifier's performance is called the confusion matrix. A clear visual representation of the LDA model is provided by the confusion matrix. False positive (FP) = 13, true negative (TN) = 139, false negative (FN) = 10, and true positive (TP) = 232 are noted. The numbers that LDA accurately anticipated were TP and TN. False predictions, sometimes referred to as missclassifications, are FN and FP. The matrix shows that there are 371 correct predictions overall, and there are roughly 23 worng predictions. The little amount of misclassifications in the confusion matrix above shows that the LDA model has been successful in producing accurate predictions.

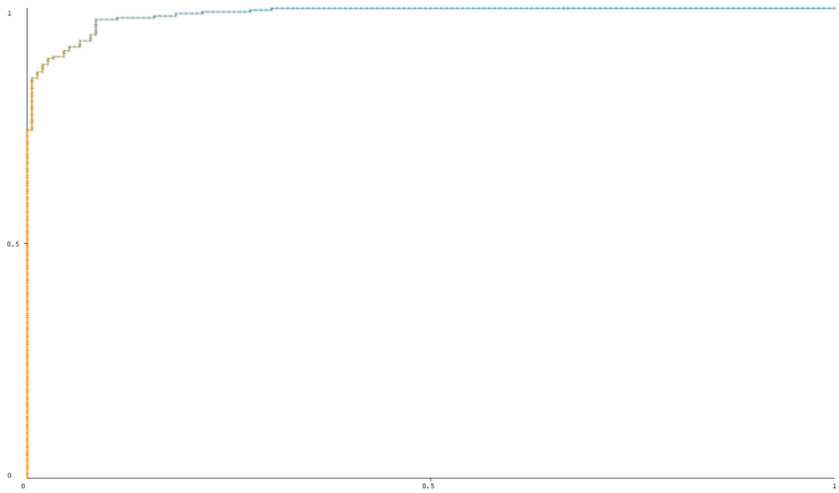

Figure 5. ROC curve of LDA

The receiver operating characteristic (ROC) curve graph is shown in Figure 5. The ROC graph above was produced using the results of the LDA model. False positive rate (FPR) is displayed on the horizontal X axis, while true positive rate (TPR) is displayed on the vertical Y axis. The area under the curve or AUC value is 0.9875. A better performing model is indicated by a higher AUC. AUC LDA is closest to 1, indicating that LDA performance is consistent with a pure model. High sensitivity and a low false positive rate are indicated by a curve that hugs the top-left corner of a perfect classifier.

Accurately predicting smartphone addiction in its early stages is crucial. Supporting decisions in the provision of suitable therapy for patients with smartphone addiction is the aim. This study demonstrates the reliability of discriminant analysis, particularly LDA, for early prediction. Its high evaluation value of around 0.942 and AUC of 0.9875 serve as indicators. This model advances the subject of health systems and has low classification errors, making it reasonably good.

4. Conclusion

Smartphone addiction, particularly among adolescents, has become a pressing issue due to its detrimental effects on mental health, academic performance, and physical well-being. Early detection of smartphone addiction is critical to mitigating these negative consequences and enabling timely interventions. This study proposes the use of discriminant analysis—specifically Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA)—to predict smartphone addiction based on behavioral data. The research utilized a dataset comprising 394 instances, categorized into "addicted" and "non-addicted" groups. After preprocessing steps including feature selection and ordinal encoding, the models were trained and evaluated using 10-fold cross-validation. The evaluation metrics, including accuracy, precision, recall, and F-measure, demonstrated that LDA outperformed QDA. LDA achieved an accuracy of 94.16%, a precision of 94.2%, a recall of 94.2%, and an F-measure of 94.2%. Additionally, the Receiver Operating Characteristic (ROC) curve analysis yielded an Area Under the Curve (AUC) of 0.9875 for LDA, highlighting its robustness and reliability in classifying smartphone addiction. QDA, while effective, exhibited slightly lower performance due to the dataset's linear separability. These findings establish LDA as an effective and reliable method for predicting smartphone addiction. Its stable performance, low misclassification rate, and high sensitivity make it a valuable tool for early intervention strategies in behavioral health monitoring systems. Further research is recommended to optimize the discriminant analysis models, extend the dataset to include more diverse populations, and explore the integration of additional machine learning techniques to enhance prediction accuracy. Investigating the application of these methods in real-time systems and other behavioral health contexts could also yield significant contributions to the field.

References

[1] J. Park, J. E. Jeong, S. yeon Park, and M. J. Rho, "Development of the Smartphone Addiction Risk Rating Score for a Smartphone Addiction Management Application," *Front. Public Heal.*, vol. 8, no. September, pp. 1–11, 2020, doi: 10.3389/fpubh.2020.00485.

[2] J. E. Chung *et al.*, "Smartphone addiction risk and daytime sleepiness in Korean adolescents," *J. Paediatr. Child Health*, vol. 54, no. 7, pp. 800–806, 2018, doi: 10.1111/jpc.13901.

[3] M. Dhamayanti, R. G. Dwiwina, and R. Adawiyah, "Influence of Adolescents ' Smartphone Addiction on Mental and Emotional Development in West Java , Indonesia Pengaruh Kecanduan Gawai pada Perkembangan Mental dan Emosional Remaja di Jawa Barat , Indonesia," *J. Med.*, vol. 51, no. 38, pp. 46–52, 2019, doi: 10.15395/mkb.v51n1.1577.

[4] C. F. Giraldo-jiménez, J. Gaviria-chavarro, A. Urrutia-valdés, J. F. Bedoya-pérez, and M. O. Sarria-paja, "Machine-Learning Predictive Models For Dependency On Smartphones Based On Risk Factors," pp. 1–29, 2021, doi: 10.21203/rs.3.rs-886633/v1.

[5] J. E. Ding *et al.*, "Development Of A Smartphone Overuse Classification Scale," *Addict. Res. Theory*, vol. 27, no. 2, pp. 150–155, 2019, doi: 10.1080/16066359.2018.1474204.

[6] J. Lee and W. Kim, "Prediction of problematic smartphone use: A machine learning approach," *Int. J. Environ. Res. Public Health*, vol. 18, no. 12, 2021, doi: 10.3390/ijerph18126458.

[7] A. Sediyono and A. Ariwibowo, "Software requirement specification of intelligent system for monitoring and preventing smartphone addiction," *Proceeding 2017 Int. Conf. Smart Cities, Autom. Intell. Comput. Syst. ICON-SONICS 2017*, vol. 2018-Janua, pp. 54–58, 2017, doi: 10.1109/ICON-SONICS.2017.8267821.

[8] J. A. Olson *et al.*, "Smartphone addiction is increasing across the world: A meta-analysis of 24 countries," *Comput. Human Behav.*, vol. 129, pp. 1–35, 2022, doi: 10.1016/j.chb.2021.107138.

[9] O. Okfalisa, E. Budianita, M. Irfan, H. Rusnedy, and S. Saktioto, "The Classification of Children Gadget Addiction: The Employment of Learning Vector Quantization 3," *IT J. Res. Dev.*, vol. 5, no. 2, pp. 158–170, 2020, doi: 10.25299/itjrd.2021.vol5(2).5681.

[10] I. D. Amalya, A. Rizqi, and F. A. Fadjri, "The Relationship Between Smartphone Addiction and Loneliness Among Students," *Proc. ICECRS*, vol. 8, pp. 4–9, 2020, doi: 10.21070/icecrs2020421.

[11] W. Lawanont and M. Inoue, "A development of classification model for smartphone addiction recognition system based on smartphone usage data," *Smart Innov. Syst. Technol.*, vol. 73, pp. 3–12, 2018, doi: 10.1007/978-3-319-59424-8_1.

[12] L. N. Filaili, H. Rosyita, and P. Rahmawati Handayani, "Kecanduan Smartphone Dengan Kepercayaan Diri Pada Remaja Usia 12-15 Tahun Tingkat Sekolah Menengah Pertama Di Kecamatan Kramat Jati, Jakarta Timur," *J. Teknol. Inf.*, vol. 7, no. 2, pp. 116–123, 2021, doi: 10.52643/jti.v7i2.1905.

[13] R. Katarya and S. Maan, "Predicting mental health disorders using machine learning for employees in technical and non-technical companies," *Proc. 2020 IEEE Int. Conf. Adv. Dev. Electr. Electron. Eng. ICADEE 2020*, no. Icadee, 2020, doi: 10.1109/ICADEE51157.2020.9368923.

[14] R. Wang, "Comparison of Decision Tree, Random Forest and Linear Discriminant Analysis Models in Breast Cancer Prediction," *J. Phys. Conf. Ser.*, vol. 2386, no. 1, 2022, doi: 10.1088/1742-6596/2386/1/012043.

[15] B. Ghojogh and M. Crowley, "Linear and Quadratic Discriminant Analysis: Tutorial," no. 4, pp. 1–16, 2019, doi: 10.48550/arXiv.1906.02590.

[16] W. Suprianto and D. Wanda, "Datasets for Smartphone Addiction and Cyberbullying Prevalence in Early Adolescents," vol. 3, 2021, doi: 10.17632/M6S3D8KVSM.3.

[17] Kedar Potdar, Taher S. Pardawala, and Chinmay D. Pai, "A Comparative Study of Categorical Variable Encoding Techniques for Neural Network Classifiers," *Int. J. Comput. Appl.*, vol. 175, no. 4, p. 375, 2017, doi: 10.5120/ijca2017915495.

[18] M. Anshori and M. S. Haris, "Predicting Heart Disease using Logistic Regression," *Knowl. Eng. Data Sci.*, vol. 5, no. 2, p. 188, 2022, doi: 10.17977/um018v5i22022p188-196.

[19] M. Anshori, F. Mar'i, and F. A. Bachtiar, "Comparison of Machine Learning Methods for Android Malicious Software Classification based on System Call," *Proc. 2019 4th Int. Conf. Sustain. Inf. Eng. Technol. SIET 2019*, pp. 343–348, 2019, doi: 10.1109/SIET48054.2019.8985998.

[20] M. Anshori, F. Mahmudy, and A. A. Supianto, "Preprocessing Approach for Tuberculosis DNA Classification using Support Vector Machines (SVM)," *J. Inf. Technol. Comput. Sci.*, vol. 4, no. 3, pp. 233–240, 2019, doi: https://doi.org/10.25126/jitecs.201943113.

[21] G. B. G. Pereira, L. P. Fernandes, J. M. R. D. S. Neto, H. D. D. M. Braz, and L. D. S. Sauer, "A comparative study of linear discriminant analysis and an artificial neural network performances in breast cancer diagnosis," *2020 Ieee Andescon, Andescon 2020*, 2020, doi: 10.1109/ANDESCON50619.2020.9272057.

[22] H. W. Wibowo, M. Hasbi, and M. Anshori, "Use Discriminant Analysis to Identify Eroticism-Related Terms in The Lyrics of Dangdut Songs," *J. Enhanc. Stud. Informatics Comput. Appl.*, vol. 1, no. 1, pp. 17–22, 2024, doi: 10.47794/jesica.v1i1.5.

[23] S. Guo and H. Tracey, "Discriminant Analysis for Radar Signal Classification," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, no. 4, pp. 3134–3148, 2020, doi: 10.1109/TAES.2020.2965787.

[24] M. M. Islam, A. K. M. Masum, S. Abujar, and S. A. Hossain, "Prediction of chronic Insomnia using Machine Learning Techniques," *2020 11th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT 2020*, 2020, doi: 10.1109/ICCCNT49239.2020.9225570.

[25] E. Frank, M. A. Hall, and I. H. Witten, "The WEKA workbench," *Data Min.*, pp. 553–571, 2017, doi: 10.1016/b978-0-12-804291-5.00024-6.