Department of Digital Business
Journal of Artificial Intelligence and Digital Business (RIGGS)
Homepage: https://journal.ilmudata.co.id/index.php/RIGGS
Vol. 4 No. 2 (2025) pp: 3526-3534
P-ISSN: 2963-9298, e-ISSN: 2963-914X

Improving Random Forest Evaluation in Mental Health Disorder Identification with Cross Validation

Rosyida Choirunnisa, Mochammad Anshori, Wahyu Teja Kusuma
[1,2,3] Department of Informatics, Faculty of Science and Technology, Institut Teknologi, Sains, dan Kesehatan RS DR Soepraoen Kesdam V/ BRW, Malang, Indonesia
[1]rosyidachoirunnisa81@gmail.com, [2]moanshori@itsk-soepraoen.ac.id*, [3]wtkusuma@itsk-soepraoen.ac.id

Abstract

Mental health disorders are often difficult to detect and diagnose, causing misdiagnoses which lead to inappropriate treatment and have a negative impact on the sufferer's quality of life. This research aims to develop an accurate and efficient model for identifying mental health disorders by utilizing the Random Forest method and Cross Validation techniques. Random Forest was chosen because of its ability to improve prediction accuracy and training speed. Cross Validation is used to train and test models with various combinations of data, and reduces the risk of Overfitting. The dataset consists of 120 data with 18 behavioral attributes and diagnoses, with four target classes: Bipolar Type-1, Bipolar Type-2, Depression, and Normal. Four Cross Validation experimental scenarios were tested: k=5 and k=10, and k=5 and k=10 with Stratified to reduce data bias. Experimental results show that k=10 stratified cross-validation produces the highest accuracy (87.5%), with precision, recall, and F1-score also reaching 87.5%. The Stratified technique is proven to improve the balance of class distribution and reduce the risk of Overfitting. These findings confirm that Random Forest with k=10 Stratified Cross-Validation is the optimal approach for diagnosing mental health disorders. The implications of this research include the potential for applying models in AI-based systems to assist medical personnel in more accurate and efficient early diagnosis.

Keywords: *Random Forest, Cross Validation, Identification, Mental Health Disorders*

1. Introduction

Mental health disorders are conditions where there are differences in behavioral patterns, thoughts, and emotions that can affect daily life [1]. This condition often hinders sufferers from carrying out normal daily activities, including work or social interactions. Sufferers will find it more difficult to develop, are susceptible to stress and can even have an impact on criminal behavior [2]. If sufferers are not treated immediately, they can be at risk of becoming more severe, such as having suicidal tendencies [3].

There are several common mental health disorders, including Bipolar Disorder and Depression. According to WHO 2013, at least one in four people in the world experience mental problems, where there are around 35 million people suffering from Depression and 60 million people suffering from Bipolar [4]. Depression is a mental disorder associated with feelings of deep sadness such as hopelessness, helplessness, and suicidal tendencies [5]. Meanwhile, Bipolar disorder is a disorder of extreme emotional and mood changes that cannot be controlled. There are two types of Bipolar disorder, namely Bipolar Type-1 and Bipolar Type-2 [6]. The similarity of symptoms of Bipolar Type-1, Bipolar Type-2 and Depression tends to be difficult to recognize, so misdiagnosis often occurs [7]. Misdiagnosis can result in delays or mishandling which can worsen the patient's condition [8].

The above problems indicate the need for a technology that is able to identify mental health disorders accurately, especially in Bipolar Type-1, Bipolar Type-2 and Depression, in order to minimize misdiagnosis. One solution that can be applied is the use of Machine Learning. Where Machine Learning is part of the field of artificial intelligence with the ability to identify, recognize patterns, predict, and classify [9]. The advantages of using Machine Learning are that it can increase the speed and accuracy of the diagnosis process [10]. In addition, Machine Learning can learn independently and improve performance as the data used increases [11]

To support the application of Machine Learning in diagnosis, a method is needed that is able to produce accurate predictions. The Random Forest method is one method that can be used. In its development, the Random Forest method consists of a set of decision trees from the dataset [12]. Then voting is carried out from the results of the decision tree, the process aims to determine the target class that appears most often from the results of each tree and provides predictions. The word "Random" in this method refers to the random process on the data used before creating each tree [13]. This method is one of the classification methods that can be said to be very accurate in making predictions [14].

One study by Dony Benaya used the Random Forest method to create a lung cancer classification model with an accuracy of 78% [13]. Another study by Raja Darmawan applied Random Forest to diagnose Tiro's disease with an accuracy of 99% [15]. There is also another study that uses Random Forest to predict heart failure by Edric with an accuracy of 82.6087% [16]. From the success of previous studies, the Random Forest method is able to identify a disease and produce an accurate model in its prediction. Therefore, the Random Forest method is suitable for identifying health data.

The success of the Random Forest method is greatly influenced by the division of the dataset during training [17]. One of the dataset division techniques that can be applied is the Cross Validation technique [18]. This technique trains and tests the model with various data combinations so that it can reduce the risk of Overfitting [19]. In addition, Cross Validation helps estimate model prediction errors and provides information on how well the method used in identifying new data [20]. Therefore, researchers propose to apply Cross Validation as a method in creating a Random Forest model.

This study aims to create an identification model using the Random Forest method in identifying mental health disorders, especially Bipolar Type-1, Bipolar Type-2 and Depression. In determining the best model, this study also conducted a comparison of Cross Validation results to obtain the most optimal and accurate model. The results of this study are expected to help minimize early diagnostic errors, so that sufferers get the right treatment.

2. Research Methods

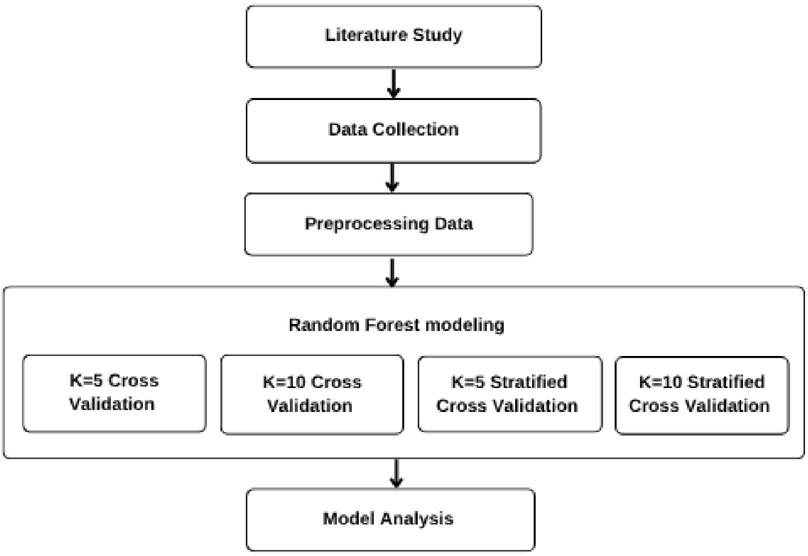

Figure 1. Research Flow

Referring to Figure 1 the flow of the stages of the research is shown. Starting from the researcher conducting a literature study, data collection, data preprocessing, Random Forest modeling with Cross Validation experiments k-5, k = 10, k = 5 stratified, k = 10 stratified and finally conducting data analysis. A more detailed explanation of each stage is explained in the next sub-chapter.

2.1. Literature Study

At this stage, the researcher conducted a literature study with the aim of deepening the theoretical knowledge used as the basis for the research. The literature study conducted by the researcher focused on the Random Forest theory in identifying health data. Where the author looks for theories about Machine Learning, Random Forest methods, Cross Validation and mental health interests through journals obtained online.

DOI: https://doi.org/10.31004/riggs.v4i2.1053

2.2. Data Collection

After conducting a literature study, the researcher collected the dataset. The dataset that will be used in this study was obtained from the Harvard Dataverse platform [21]. The dataset has 120 rows of data and consists of 18 attributes. These attributes include information such as patient numbers, several behavioral patterns and diagnosis results. The dataset also there is a target attribute, namely disease diagnosis. The target attribute is divided into 4 classes, namely Bipolar Type-1, Bipolar Type-2, Depression and Normal.

2.3. Preprocessing Data

The next stage is to perform data preprocessing. This stage is carried out to eliminate data errors and ensure data quality before modeling. In this study, Encoding Treat as Ordinal is applied, the mechanism is by changing categorical data into numeric data. After the data value is changed, the data range value of each attribute will be visible. In the dataset, there are several attributes that have very different value ranges, so researchers normalize using the min max method. This normalization changes the data size from its original range, so that all values have a uniform range in the range of 0 and 1 [14]. The min max normalization formula can be seen in equation (1).

$$V_{norm} = \left(\frac{V_i - V_{min}}{V_{max} - V_{min}}\right) \quad (1)$$

The Information:

Vnorm = Result value applied Min-Max Normalization

Vmin = The minimum value of an attribute in the dataset.

Vner = The maximum value of an attribute in the dataset.

V_i = The i-th value of the attribute to be normalized

In determining the min max normalization value, the result of the difference between the value to be normalized and the minimum value of an attribute is needed. Then divide it by the difference between the maximum and minimum. So that it will produce a new normalization value in the range of 0 to 1.

2.4. Modeling

In the modeling process, researchers apply the Cross Validation technique to divide the dataset into two parts, namely training and testing data. Cross Validation is often known as k-fold Cross Validation, which is a process that is repeated k times by dividing the data into k parts of the dataset [22].

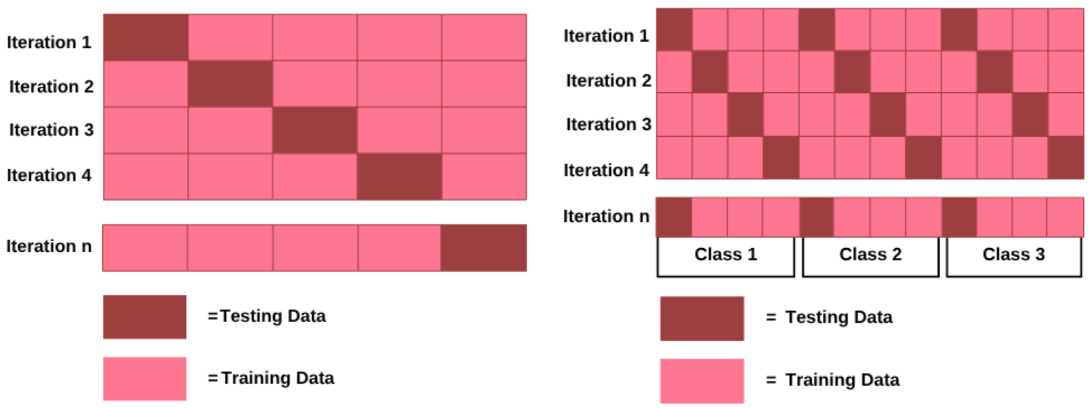

Figure 2. Illustration of Cross Validation (left) and Stratified Cross Validation (right)

In Figure 2 (left) is an illustration of Cross Validation where in its division using k-n then the data will be divided into n parts of the set. In each iteration it consists of (n-1) parts as training data and one part as test data. With

DOI: https://doi.org/10.31004/riggs.v4i2.1053

Applying Cross Validation can obtain a more stable level of performance, can measure the quality of the classification model built, and can increase the accuracy of predictions [14]. In addition, by applying Cross Validation, the resulting prediction model will be more general and the model can avoid overfitting [19]. Figure 2 (right) is an illustration of stratified Cross Validation where the data division technique ensures that in the training data and testing data there must be representatives from all existing classes [23]. Stratified is done to ensure that each part is a good representation of the data.

In this study, an experiment was conducted using the values k-5 and k-10. In this experiment, the dataset was divided randomly without considering the class distribution. Meanwhile, in the next experiment, stratified Cross Validation was applied with the values k = 5, and k = 10. The division of the dataset in stratified considers the class distribution in each fold so that the class distribution is even. This experiment was conducted to evaluate the performance of the model on various cross-validation schemes and determine the best scheme.

One method that is suitable for classification that is often used is Random Forest [24]. The Random Forest method is an algorithm that consists of a large number of decision trees that work independently and produce predictions [25]. The reliability of Random Forest is that it produces predictions and a fast training process. In addition, this method can increase accuracy results and can be used for classification on data that has a large amount [26].

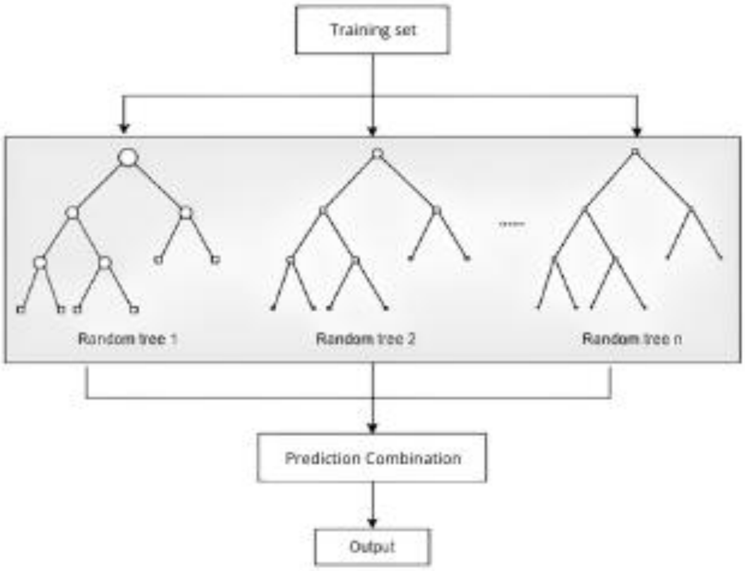

Figure 3. Random Forest Workflow [14]

The workflow of Random Forest is shown in Figure 3 above. Based on the figure, the mechanism is that the incoming training data is used as input after which a random decision tree is built. At each node in the tree, the algorithm randomly selects an attribute that will be used to make a decision. After all decision trees are complete, Random Forest will combine the predictions from each tree to provide the final prediction. Usually, the final result is obtained through majority voting [14]. The following Random Forest formula can be seen in formula (2).

$$F(x) = \frac{1}{j}\sum_{j=1}^{j} h_j(x) \quad (2)$$

where F(x) is the output of the Random Forest; J is the number of trees in the ensemble, and h, is the output of the - (j)th tree. This formula calculates the output of the Random Forest model, which works by combining the prediction results of many decision trees.

2.5. Model Analysis

The result analysis stage is the stage for evaluating the results of the Random Forest model performance. This stage compares the evaluation results from the experiment. The evaluations include accuracy, precision, recall and F1 score.

DOI: https://doi.org/10.31004/riggs.v4i2.1053

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

$$F1 = \frac{2 * (Recall * Presisi)}{Recall + Presisi} \quad (6)$$

Where TP = True Positive; TN = True Negative, FP = False Positive; and FN = False Negative. Accuracy is an evaluation method that measures how close the predicted value is to the actual value. The level of accuracy can be obtained through the results of the analysis of data that is classified precisely [27]. The calculation of accuracy can be seen in equation (3). Precision is a method that measures the comparison of the amount of relevant information taken by the system with the total amount of information taken by the system, both relevant and irrelevant [27]. The precision equation is shown in equation (4). Recall is a testing method that compares the amount of relevant information obtained by the system with the total amount of relevant information in the information collection [27]. The following formula for Recall is shown in equation (5). Then there is F1-Score, which is the measurement result to determine the average comparison between the Recall value and the Precision value. Equation (6) displays the formula used to obtain the F1-Score value. In its analysis, a comparison of evaluation values was carried out with Cross Validation experiments and Stratified Cross Validation at values k-5 and k-10. The results will be recorded and then compared with each other to determine the most optimal Cross Validation model value in cases of identifying mental health disorders.

3. Results and Discussions

This section discusses the results and discussion based on the research methodology that has been presented previously. The data used in the study uses secondary data obtained from the Harvard Dataverse platform and is shown in Table 1.

Table 1. Attributes With Data Type and Value

No	Attribute	Data Type	Data
1	Sadness	Categorycal	Seldom, Some time, Usually, Most-Often
2	Euphoric	Categorycal	Seldom, Some time, Usually, Most-Often
3	Exhausted	Categorycal	Seldom, Some time, Usually, Most-Often
4	Sleep dissorder	Categorycal	Seldom, Some time, Usually, Most-Often
5	Mood Swing	Categorycal	Yes, No
6	Suicidal thoughts	Categorycal	Yes, No
7	Anorxia	Categorycal	Yes, No
8	Authority Respect	Categorycal	Yes, No
9	Try-Explanation	Categorycal	Yes, No
10	Aggressive Response	Categorycal	Yes, No
11	Ignore & Move-On	Categorycal	Yes, No
12	Nervous Break-down	Categorycal	Yes, No
13	Admit Mistakes	Categorycal	Yes, No
14	Overthinking	Categorycal	Yes, No
15	Sexual Activity	Numerical	1, 2, 3, 4, 5, 6, 7, 8, 9
16	Concentration	Numerical	1, 2, 3, 4, 5, 6, 7, 8, 9
17	Optimisim	Numerical	1, 2, 3, 4, 5, 6, 7, 8, 9

Table 1 shows that the dataset has 17 attributes where 14 attributes with categorical data type and 3 other attributes with numeric data type. Attributes with categorical data type need to be converted into numeric using Encoding Treat as Ordinal. This process gives order to the data into numeric values so that it is easier for the model to process the data [28]. After the Encoding process is complete, the range of data values in each attribute is visible as in table 2 below.

Table 2. Attribute and range data

No	Attribute	Range Data
1	Sadness	0- 3

DOI: https://doi.org/10.31004/riggs.v4i2.1053

2	Euphoric	0- 3
3	Exhausted	0- 3
4	Sleep dissorder	0-1
5	Mood Swing	0-1
6	Suicidal thoughts	0-1
7	Anorxia	0-1
8	Authority Respect	0-1
9	Try-Explanation	0-1
10	Aggressive Response	0-1
11	Ignore & Move-On	0-1
12	Nervous Break-down	0-1
13	Admit Mistakes	0-1
14	Overthinking	0-1
15	Sexual Activity	0-8
16	Concentration	0-8
17	Optimisim	0-8

From Table 2 it can be seen that each attribute has a different range of data values. Differences in the range of data values can cause attributes with larger values to dominate the calculation while attributes with smaller values are ignored. This is the reason for applying min max normalization to change the range of values from 0 to 1. Thus, all attributes have the same scale and can work in balance in data analysis.

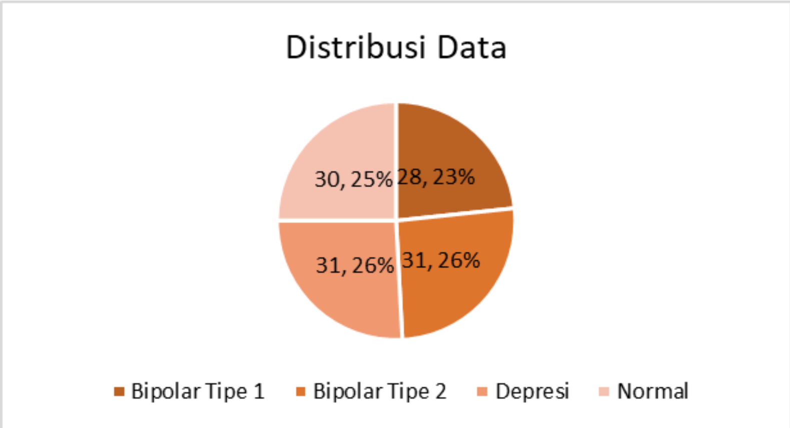

Figure 4. Class Distribution Graph

Based on Figure 4 It can be seen that the class distribution in the dataset is unbalanced, where the Bipolar Type-2 and Depression classes have the largest number of data, each 31 data with a percentage of 26%, followed by the normal class, which is 30 data with a percentage of 25% and Bipolar Type-1 as many as 28 data with a percentage of 23%. This imbalance has the potential to cause problems in the model training process, because the model will tend to predict the majority class (Bipolar Type-2 and Depression) with high accuracy but has difficulty recognizing the minority class (Normal and Bipolar Type-2). At the modeling stage, four Cross Validation scenario experiments were carried out, namely the first scenario using k-5, the second scenario using k-10, the third scenario using k-5 stratified and the fourth scenario k = 10 stratified.

Table 3. Comparison of Random Forest Model Evaluation against Cross Validation

Cross Validation	Presisi	Recall	F1 score	Akurasi
k= 5	0.791	0.792	0.790	0.792
k=10	0.808	0.808	0.807	0.808
k= 5 stratified	0.816	0.817	0.816	0.817
k=10 stratified	**0.875**	**0.875**	**0.874**	**0.875**

Table 3 presents the evaluation results of the Random Forest model with the Cross Validation test scenario that has been explained previously. First with k-5, the model produces a precision of 0.791, recall = 0.792, F1-score = 0.790, and accuracy of 0.792. The second test, with k-10, there was an increase in evaluation, namely Precision - 0.808, recall = 10.808, Fl-score -8.807, and accuracy increased to 0.808. Furthermore, the third test with k-5 stratified produced an evaluation of precision-i 0.816, recall 0.817, Fl-score 0.816, and accuracy 0.817. The highest evaluation increase was seen in the fourth scenario, with k-10 stratified, namely with the results of precision = 0.875, recall = 0.875, F1-score -0.874, and accuracy reaching the highest value with a value of 0.875. This can

DOI: https://doi.org/10.31004/riggs.v4i2.1053

indicate that the use of stratified Cross Validation can improve the performance of the Random Forest classification model evaluation. It is known from the table above that k-5 stratified produces a better model than k-5 without stratification. This is evidenced by an increase in evaluation of approximately 2.5%. In addition, k-10 stratification also experienced an increase in the model when compared to k-10 without stratification, as evidenced by the evaluation results which increased by approximately 6.7%.

Referring to the test results based on four Cross Validation scenarios, it is proven that with different k values and the use of stratified has an impact on the Random Forest classification model. In this study, using k = 10 stratified gave the best results than the test scenarios with k-5, k = 10, and k-5 stratified. This is evidenced by the evaluation of precision, recall, F1-score and accuracy which are the highest than other Cross Validation tests.

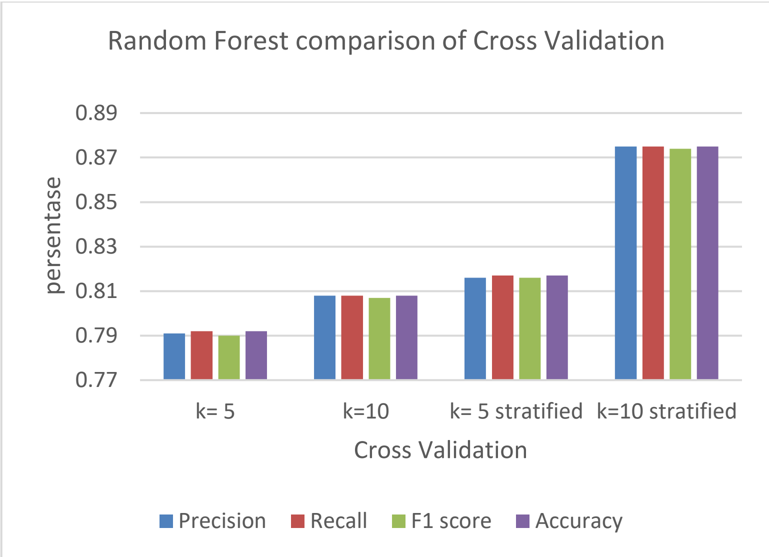

Figure 5. Visualization of Model Comparison

From Figure 5, the graph above shows a comparison of the performance of the Random Forest model based on various Cross Validation techniques, namely k-5, k = 10, k = 5 stratified, and k = 10 stratified, with evaluation using precision, recall, F1-score, and accuracy metrics. It can be seen that the evaluation values at k-5 and k-10 are relatively lower compared to the stratified method. The application of stratified Cross Validation, especially at k-10 stratified, shows a significant increase in all evaluation metrics, with a value of almost 0.88. This shows that stratified Cross Validation provides a more balanced data distribution in each fold, resulting in a more stable and accurate model. Thus, in the implementation of the Random Forest model, the use of stratified Cross validation, especially with k = 10, can be a more optimal choice to improve model performance.

Actual \ Predicted	Bipolar Type-1	Bipolar Type-2	Depression	Normal	Σ
Bipolar Type-1	20	4	0	4	28
Bipolar Type-2	0	31	0	0	31
Depression	0	0	29	2	31
Normal	0	1	4	25	30
Σ	20	36	33	31	120

Figure 6. Confusion Matrix k=10 Stratified Cross Validation

Figure 6 above is a confusion matrix of the results of the Random Forest model evaluation using Cross-Validation with k-10 stratified on the classification of mental disorders, namely Bipolar Type-1, Bipolar Type-2, Depression, and Normal. The rows show the actual labels, while the columns show the model predictions. From the evaluation results, the model has a fairly good performance in classifying Bipolar Type-2 and Depression, with a level of correct predictions were 31/31 (100%) and 29/31 (93.5%) respectively. However, there were some

DOI: https://doi.org/10.31004/riggs.v4i2.1053

misclassifications, especially in Bipolar Type-1 which was often classified as Bipolar Type-2 (4 cases) and Normal (4 cases), and Normal individuals who were sometimes predicted as Depression (4 cases) or Bipolar Type-2 (1 case). These misclassifications could occur due to overlapping features between categories which made it difficult for the model to distinguish several classes well. However, overall, the model showed good performance, especially with the stratified Cross Validation method which ensures a more balanced data distribution in the training and testing process.

4. Conclusion

This study shows that the use of the Random Forest method with the k-10 stratified Cross Validation technique provides the best results in the classification of mental health disorders, with an accuracy of 87.5%. These results prove that a balanced division of the dataset through stratified Cross Validation can improve model performance compared to the method without stratification. This model successfully identified Bipolar Type-1, Bipolar Type-2, and Depression more accurately, although some misclassifications were still found, especially between Bipolar Type-1 and Bipolar Type-2. The implication of this study is that Machine Learning methods, especially Random Forest, can be an effective tool in assisting the early diagnosis of mental health disorders. Integration of this model into an application-based system or digital platform can improve the efficiency and accuracy of diagnosis in the medical world. This study contributes to the development of an Al-based classification system in the field of mental health and strengthens the evidence that the stratified Cross Validation technique can reduce bias in Machine Learning models. Further studies can explore larger datasets, compare Random Forest with other algorithms such as XGBoost or Deep Learning, and develop Al-based applications for direct implementation in the medical world

Reference

[1] F. Siska and S. Heni, “Analisis Data Hasil Diagnosa Untuk Klasifikasi Gangguan Kepribadian Menggunakan Algoritma C4.5,” *J. Teknol. dan Sist. Inf.*, vol. 2, no. 4, pp. 89–95, 2021, doi: 10.33365/jtsi.v2i4.1373.

[2] A. W. Widiarni, S. Sarah, R. A. A. Astri, and M. Mustakim, “Diagnosis Penyakit Mental Pada Remaja Menggunakan Metode Simple Multi Attribute Rating Technique Exploiting Rank (SMARTER),” *Indones. J. Inform. Res. Softw. Eng.*, vol. 2, no. 2, pp. 100–108, 2022, doi: 10.57152/ijirse.v2i2.444.

[3] M. I. Maulana, Z. Lessy, U. Islam, N. Sunan, U. Islam, and N. Sunan, “Upaya Penanganan Dan Peningkatan Kesehatan Mental,” *Koloni*, vol. 2, no. 4, pp. 90–98, 2023, doi: 10.31004/koloni.v2i4.549.

[4] F. I. Kesehatan, U. Kusuma, H. Surakarta, and S. Mahmudah, “Asuhan keperawatan jiwa pada pasien dengan gangguan halusinasi,” 2020.

[5] S. Y. Prasetyo and G. Z. Nabiilah, “Perbandingan Model Machine Learning pada Klasifikasi Tumor Otak Menggunakan Fitur Discrete Cosine Transform,” *J. Teknol. Terpadu*, vol. 9, no. 1, pp. 29–34, 2023, doi: 10.54914/jtt.v9i1.605.

[6] M. Astriliana and E. R. Kustanti, “Pengalaman Sebagai Pasien Dengan Gangguan Bipolar Tipe I (Sebuah Interpretative Phenomenological Analysis),” *J. EMPATI*, vol. 13, no. 1, pp. 78–89, 2023, doi: 10.14710/empati.2024.27722.

[7] U. Fisabilih, “Analisis Gejala Gangguan Afektif Bipolar Tipe II Berdasarkan DSM V,” *J. Ris. Kesehat. Mod.*, vol. 6, no. 3, pp. 205–212, 2024.

[8] A. A. C. A. Sukmana, P. D. A. Putra, and R. D. S. Dinata, “Perancangan Komik Strip Sebagai Sarana Kampanye Dampak Orang Tua Abusive Terhadap Perkembangan Kesehatan Mental Remaja Di Denpasar,” *J. Selaras Rupa*, vol. 2, no. 1, pp. 60–68, 2021, [Online]. Available: https://jurnal.idbbali.ac.id/index.php/selarasrupa

[9] N. R. Muntiari and K. H. Hanif, “Klasifikasi Penyakit Kanker Payudara Menggunakan Perbandingan Algoritma Machine Learning,” *J. Ilmu Komput. dan Teknol.*, vol. 3, no. 1, pp. 1–6, May 2022, doi: 10.35960/ikomti.v3i1.766.

[10] I. M. Faiza, G. Gunawan, and W. Andriani, “Tinjauan Pustaka Sistematis: Penerapan Metode Machine Learning untuk Deteksi Bencana Banjir,” *J. Minfo Polgan*, vol. 11, no. 2, pp. 59–63, 2022, doi: 10.33395/jmp.v11i2.11657.

[11] I. Sari, Fivrenodi, E. Altiarika, and Sarwindah, “Sistem Pengembangan Bahasa Isyarat Untuk Berkomunikasi dengan Penyandang Disabilitas (Tunarungu),” *J. Inf. Technol. Soc.*, vol. 1, no. 1, pp. 20–25, 2023, doi: 10.35438/jits.v1i1.21.

[12] S. P. R. Yulianto, A. Z. Fanani, A. Affandy, and M. I. Aziz, “Analisis Metode Smoote pada Klasifikasi Penyakit Jantung Berbasis Random Forest Tree,” *J. Media Inform. Budidarma*, vol. 8, no. 3, p. 1460, 2024, doi: 10.30865/mib.v8i3.7712.

[13] D. Benaya, “Implementasi Random Forest dalam Klasifikasi Kanker Paru-Paru,” *JOINTER J. Informatics Eng.*, vol. 5, no. 01, pp. 27–31, 2024, doi: 10.53682/jointer.v5i01.331.

[14] Gde Agung Brahmana Suryanegara, Adiwijaya, and Mahendra Dwifebri Purbolaksono, “Peningkatan Hasil Klasifikasi pada Algoritma Random Forest untuk Deteksi Pasien Penderita Diabetes Menggunakan Metode Normalisasi,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 114–122, Feb. 2021, doi: 10.29207/resti.v5i1.2880.

[15] R. Darmawan, Y. H. Chrisnanto, and G. Abdillah, “KLASIFIKASI DIAGNOSA PENYAKIT TIROID MENGGUNAKAN METODE,” vol. 7, no. 2, pp. 203–209, 2024, doi: 10.36595/jire.v7i2.1214.

[16] S. P. Tamba and E. -, “Prediksi Penyakit Gagal Jantung Dengan Menggunakan Random Forest,” *J. Sist. Inf. dan Ilmu Komput. Prima(JUSIKOM PRIMA)*, vol. 5, no. 2, pp. 176–181, 2022, doi: 10.34012/jurnalsisteminformasidanilmukomputer.v5i2.2445.

[17] F. Hamami and I. A. Dahlan, “Klasifikasi Cuaca Provinsi Dki Jakarta Menggunakan Algoritma Random Forest Dengan Teknik Oversampling,” *J. Teknoinfo*, vol. 16, no. 1, p. 87, 2022, doi: 10.33365/jti.v16i1.1533.

[18] W. A. Firmansyach, U. Hayati, and Y. Arie Wijaya, “Analisa Terjadinya Overfitting Dan Underfitting Pada Algoritma Naive Bayes Dan Decision Tree Dengan Teknik Cross Validation,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 262–269, 2023, doi: 10.36040/jati.v7i1.6329.

[19] M. Anshori, N. Rikatsih, and M. S. Haris, “Prediksi Pasien Dengan Penyakit Kardiovaskular Menggunakan Random Forest,” *TEKTRIKA - J. Penelit. dan Pengemb. Telekomun. Kendali, Komputer, Elektr. dan Elektron.*, vol. 7, no. 2, p. 58, 2023, doi: 10.25124/tektrika.v7i2.5279.

DOI: https://doi.org/10.31004/riggs.v4i2.1053

[20] W. Wijiyanto, A. I. Pradana, S. Sopingi, and V. Atina, “Teknik K-Fold Cross Validation untuk Mengevaluasi Kinerja Mahasiswa,” *J. Algoritm.*, vol. 21, no. 1, pp. 239–248, 2024, doi: 10.33364/algoritma/v.21-1.1618.

[21] V. Karbalaeipour, Hengameh; Damari, Siavash; Zolfagharnasab, Mohammad Hossein; Haghdadi, Amin, DVN/0FNET5 , Harvard Dataverse, “Dataset-Mental-Disorders,” 2023. doi: doi.org/10.7910.

[22] H. Azis, P. Purnawansyah, F. Fattah, and I. P. Putri, “Performa Klasifikasi K-NN dan Cross Validation Pada Data Pasien Pengidap Penyakit Jantung,” *Ilk. J. Ilm.*, vol. 12, no. 2, pp. 81–86, 2020, doi: 10.33096/ilkom.v12i2.507.81-86.

[23] A. Farmadi and M. Muliadi, “Deteksi Penyakit Tanaman Padi Menggunakan Ekstraksi Firur Lbp Dan Klasifikasi Modified Knn,” *J. Komputasi*, vol. 11, no. 2, pp. 129–137, 2023, doi: 10.23960/komputasi.v11i2.13238.

[24] M. B. Hanif, H. A. D. Rani, A. Pratama, and R. I. Sudomo, “Klasifikasi Persalinan Prematur Menggunakan Perbandingan Algoritma C4.5 dan Random Forest,” *Joined J. (Journal Informatics Educ.*, vol. 5, no. 2, pp. 61–72, Dec. 2022, doi: 10.31331/joined.v5i2.2503.

[25] M. M. Mutoffar and A. Fadillah, “Klasifikasi Kualitas Air Sumur Menggunakan Algoritma Random Forest,” *Naratif J. Nas. Riset, Apl. dan Tek. Inform.*, vol. 4, no. 2, pp. 138–146, 2022, doi: 10.53580/naratif.v4i2.160.

[26] N. Ambika Hapsari and A. Dwi Indriyanti, “Analisis Sentimen pada Aplikasi Dompet Digital Menggunakan Algoritma Random Forest,” *J. Emerg. Inf. Syst. Bus. Intell.*, vol. 04, no. 03, pp. 186–192, 2023.

[27] R. R. R. Arisandi, B. Warsito, and A. R. Hakim, “Aplikasi Naïve Bayes Classifier (Nbc) Pada Klasifikasi Status Gizi Balita Stunting Dengan Pengujian K-Fold Cross Validation,” *J. Gaussian*, vol. 11, no. 1, pp. 130–139, 2022, doi: 10.14710/j.gauss.v11i1.33991.

[28] W. Andriyani *et al.*, *Matematika pada Kecerdasaan Buatan*. Makassar: CV. Tohar Media, 2024.