G-Tech: Jurnal Teknologi Terapan
Vol. 9, No.3, July 2025, p. 1614-1623
E-ISSN: 2623-064X | P-ISSN: 2580-8737

Classification of Remission Data for Prisoners in Tangerang Class IIA Correctional Institution using K-Nearest Neighbor Algorithm

Refa Maulana Abdillah[1]✉, Halim Agung[2], Syaipul Ramdhan[3]

[1, 2, 3] Informatics Engineering Study Program, Institut Teknologi dan Bisnis Bina Sarana Global, Indonesia

✉ **Corresponding Author :** Refa Maulana Abdillah (e-mail: 1119110042@global.ac.id)

Article Information

Article History

Received : June 26, 2025
Revised : July 3, 2025
Published : July 26, 2025

***Keywords*:**
Data mining;
K-Nearest Neighbor;
Remission;
Correctional Institutions;

ABSTRACT

The classification of remission eligibility for prisoners is a critical issue in correctional institutions, as it directly impacts prison management and the rehabilitation process. Special remission is a reduction of sentence granted to prisoners based on specific criteria, including religious status and the type of remission granted. This research aims to address the challenge of classifying special remission data for prisoners at the Class IIA Tangerang Correctional Facility using the K-Nearest Neighbor (KNN) algorithm. The dataset used in this study includes four indicators: Length of Sentence, Remaining Sentence, Crime Type, and Risk Dimension, which are analyzed to predict the remission status to be granted. The KNN model, with a parameter of k=1, achieved an accuracy of 93.94%. However, the model struggled to accurately classify the "No Remission" class, resulting in failures to detect prisoners who are not eligible for remission. The data processing steps included converting categorical data into numerical format, data normalization, and splitting the data into training and testing sets. Model evaluation was conducted using Confusion Matrix, Precision, Recall, and F1-Score. The findings suggest that while the KNN algorithm can be effectively used to classify remission status, further improvements are needed to address class imbalance and optimize results.

INTRODUCTION

Based on the Government of the Republic of Indonesia Regulation No. 32 Article 1 Paragraph 6 of 1999 concerning the conditions and procedures for the implementation of the rights of Correctional inmates states that remission is a reduction in the period of serving a criminal sentence given to prisoners and criminal children who meet the requirements specified in the legislation (regulation of the government of the Republic of Indonesia, 1999). Meanwhile, special remission is the remission given on the religious holiday adopted by the prisoner and the child of the criminal in question, provided that if a religion has more than one religious holiday and a year, the one chosen is the one that is most honored by the adherents of the religion in question (Khotiah et al., 2022).

Based on Presidential Decree No. 174 of 1999, article 3 Paragraph (2) stated, that the provision of special remission is carried out on: 1) every Eid al-Fitr for prisoners and Criminal children who are Muslims; 2) every Christmas Day for prisoners and Criminal children who are Christians; 3) every Nyepi Day for prisoners and Criminal children who are Hindus; 4)every Vesak Day for prisoners. The Tangerang class IIA Correctional Institution is one of the institutions that has the authority to provide estimates of remission and the proposal of the names of prisoners who will subsequently receive remission, then the Correctional Institution reports to the Minister of Law and legislation of the Republic of Indonesia and is then determined by the decree of the Minister of Law and legislation stipulated in Law Number 12 of 1995 on corrections(law of the Republic of Indonesia, 1995)

This is an open access article under the CC BY license

To address this challenge, this research explores the application of data mining techniques, specifically the K-Nearest Neighbor (KNN) algorithm, for classifying prisoners' remission eligibility. By applying classification methods, the goal is to develop an efficient system that enhances decision-making and supports the accurate prediction of remission status based on factors such as sentence length, remaining sentence, crime type, and risk dimension.

However, in the process of development, Tangerang Penitentiary class IIA faced various problems, especially for the leadership and officers. These problems are related to the way data is processed in correctional institutions, making it difficult to produce accurate and fast decisions, especially when there are a sufficient number of decisions required. For example, there are difficulties in finding alternative verdicts to determine the names of prisoners entitled to remission.

With the advancement of technology, especially in terms of databases, various technologies for new applications have emerged, one of which is data mining (Danny et al., 2024). According to Jiawei Han in his famous book, namely Data Mining: Concepts and Techniques, data mining can be grouped by function, which includes estimation, description, classification, prediction, Association, and clustering (Han & Kamber, 2007). Classification of samples by KNN helps to optimize the sample carrier and to minimize Qe. KNN is a supervised learning approach in which it classifies data based on the training data provided. The results obtained from the KNN algorithm can also be applied in the multi-dimensional data classification described in the results section of this paper (Wan et al., 2024)

This research on the classification of remission data has been carried out by Khotiah et al. (2022) with the title prisoner remission Data classification Model using the K-Nearest Neighbors algorithm with genetic algorithm optimization. The calculation Model used in this research is classification and optimization. The algorithm used in the study is K-Nearest Neighbor (KNN), with the reason that the advantages of this algorithm lies in its resistance to training data that contains a lot of noise, as well as good performance when faced with large test data (Tiwari, 2023). The implementation process is also relatively easy when compared to other algorithms, and this model can achieve high accuracy in various areas of the dataset.

The achievement of high accuracy in this algorithm has also been studied by Nuraeni et al., (2024) which discusses the classification of company employee performance that can produce accuracy of 94% with k = 2. Khotiah and colleagues used a research method that involves data analysis with Rapidminer, where the data is selected and tested using the k-NN algorithm, with the results of k-NN calculations are validated through the Confusion Matrix, resulting in an accuracy value of 87.78% for k=12 which has been validated with a 10-Fold Cross.

The difference with this study is that it does not explain the type of remission studied, as well as different prison units that make the data recorded in the study different. In the research conducted by Khotiah and colleagues only use the term remission while remission itself has several types, namely in the form of general remission, special remission, humanitarian remission and additional remission, as stated in Article 3 Permenkumham No.3 of 2018 (Permenkumham, 2018) and Article 4 of Permenkumham No.7 Of 2022 (Permenkumham, 2022). In the study, data collection was carried out at the penitentiary Class I Semarang. Then, a similar study was conducted by Susilo et al (2024) which aims to predict the best-selling digital products in PT. Global Indo Multimedia uses KNearest Neighbor (K-NN) method with Rapidminer application. With this high level of accuracy, the K-NN model can be one of the solutions to predict digital product transactions based on pre-existing sales data (Susilo et al., 2024).

Gap this study lies in the absence of a deep research focus on the classification of remission data specifically in Tangerang Penitentiary class IIA, which has different characteristics and challenges with other institutions. *remission* data using the K-Nearest Neighbor (KNN) algorithm, applied to inmate records from Tangerang Class IIA Correctional Institution. While previous research, such as Khotiah et al. (2022), applied KNN to general remission data without distinguishing between remission types or institutional contexts, this study narrows its scope to a more specific and less-explored area. Special remission involves additional eligibility factors tied to religious holidays, which introduces a different data structure and classification challenge. This

makes the current study unique in both its dataset and its practical implications, offering a more targeted decision-support system for correctional institutions dealing with remission allocation By using quantitative-based data mining classification methods and the application of tools such as Google Colab, this study aims to introduce more efficient and accurate solutions in processing remission data in correctional institutions, which can speed up the decision-making process and increase transparency and objectivity in remission provision. The urgency of this study is very great, given the need for a better system to support Correctional policies that are fair, transparent, and in accordance with applicable regulations.

Based on the background described, this study focuses on the classification of remission data for prisoners at the Tangerang Class IIA Correctional Institution, specifically for a special type of remission, using the K-Nearest Neighbor (KNN) algorithm. The primary objective is to develop an effective data classification system to predict remission eligibility based on factors such as sentence length, remaining sentence, crime type, and other relevant indicators (Nazar, 2024).

RESEARCH METHODS

The method of this study was conducted with data mining approach to analyze data on remission provision specifically for inmates in class IIA Tangerang Correctional Institution. This study uses a classification method to predict remission status, aims to accelerate the decision-making process in the provision of special remission. The experimental design used refers to the Machine Learning Workflow or pipeline, which includes several stages, ranging from data collection, preprocessing, experimentation, model testing, evaluation, and validation of results Lau et al., (2019)

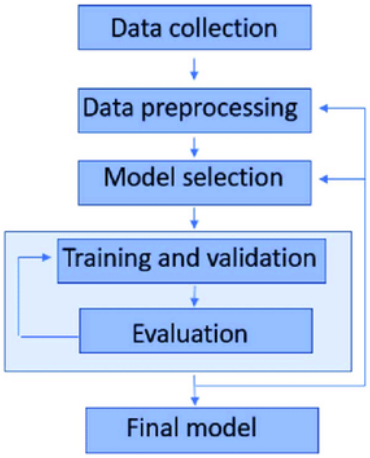

Figure 1. Flowchart Pipeline Process Machine Learning

The Data used in this study were collected from the Tangerang Penitentiary class IIA with quantitative methods. The Data obtained included 164 records of prisoners with four indicators analyzed, namely the length of the crime, the rest of the crime, crime, and risk dimensions. Each data is provided with a remission status label, which indicates whether the prisoner is entitled to receive remission or not. The Data will be used to build a classification model that can predict remission status based on these variables.

The preprocessing process begins with cleaning and converting the data obtained into a format that is easier to analyse Cholil et al., (2021) The values of categories present in the data such as the type of criminal offense or religious status are converted into a numerical format using the one-hot encoding technique. In addition, normalization is carried out to ensure that all variables have a uniform scale. This process is important for preparing data for use in classification algorithms, especially to reduce potential bias and improve model accuracy (Sakti & Daulay, 2024)

After the data is processed, the next stage is experimentation and testing of the classification model. In this study, k-Nearest Neighbors (K-NN) algorithm is used, which is a

popular method in data mining for classification (Ujianto et al., 2025). The selection of parameter k is very important in K-NN, because the value of k that is too small or too large can affect the classification results. To determine the optimal value of k, the basic formula used is :

$$d_i = \sqrt{\sum_{i=1}^{p}(x_1 - x_2)^2} \quad (1)$$

Description:
d_i = distance
i = data variable
p = Dimension data
x_1 = sample data
x_2 = test data

The process of calculating the accuracy value can be done using the equation

$$accuration = \frac{\text{Number of correct classifications}}{\text{number of Test Data}} \; x \; 100\% \quad (2)$$

In this experiment, the value of k is tested and adjusted to obtain the best results. The evaluation process is carried out using a Confusion Matrix, which will give an idea of the accuracy of the classification results. The Confusion Matrix measures the model's accuracy in predicting remission status by showing how much data is correctly and incorrectly classified (Creswell, 2014).

After the experiment is completed, the evaluation phase of the model is carried out to assess the performance of the classification model. Evaluation with the Confusion Matrix generates metrics such as accuracy, precision, recall, and F1-score, which provide complete information about the model's ability to correctly predict remission status. To ensure that the model does not get caught in overfitting, validation of the results is carried out using cross-validation, which results in the average value of the various evaluations carried out in each fold.

RESULT AND DISCUSSION

a. Inmate Data

In this study, the authors used special remission data of prisoners in Tangerang class IIA Correctional Institution in 2024. In this study the authors took the data of prisoners as many as 22 recordings with 4 indicators and 1 label. Of these data, there is 1 label that contains categorical and must be converted into numerical format to facilitate data processing.

Table 1. Inmate Dataset

No	Name Of Inmate and Child	Criminal Old Value	Criminal Residual	Value Criminal Offense	Risk Dimension Value	Remission
1	Cut Marsidah	2,89	1.22	20	20	No
2	Tri Wulandari	3,32	1,42	20	20	Yes
3	Ana Pariha Alias Aray Binti Edi Sutrisni (Alm)	3,91	1,58	0,1	0,1	Yes
4	Andririni Yaktiningsasi Binti Rm Oemar Said (Alm)	3,91	1,58	0,1	0,1	Yes
5	Anggi Rusidi Binti Rusdiman	3,91	1,58	0,1	0,1	Yes
6	Anisya Gelora Binti Tan Priasa	3,07	1,25	0,1	0,1	Yes
7	Anita Maha Pratiwi Alias Tiwi Binti Abdi Pria Dinata	3,07	1,25	0,1	0,1	Yes
8	Annis Pitria Binti Mawardi	3,07	1,25	0,1	0,1	Yes
9	Astri Delima Binti Alm Suwandi	3,35	1,36	0,1	0,1	Yes
10	Elisa Julia Ningsih	3,92	1,46	9,9	9,9	No

No	Name Of Inmate and Child	Criminal Old Value	Criminal Residual	Value Criminal Offense	Risk Dimension Value	Remission
11	Aulia Andayani Alias Aulia Bin Asep Rustandi	3,35	1,36	0,1	0,1	Yes
12	Ayu Novita Sari Binti Tulam	3,91	1,58	0,1	0,1	Yes
13	Bainah Bt. Usman Bin Usman	4,19	1,7	0,1	0,1	Yes
14	Budiyah Bt Punas	2,79	1,13	0,1	0,1	Yes
15	Chofifah Aulia Binti Gunawan	2,51	1,02	0,1	0,1	Yes
16	Dayu Rizki Limayuli Alias Yuris Binti Dadang M. S.	2,51	1,02	0,1	0,1	Yes
17	Debby Novelia Binti Febrial	3,91	1,58	0,1	0,1	Yes
18	Delvia Agustin Binti Agus Matatula	2,65	1,08	0,1	0,1	Yes
19	Eliza Mahdelia	3,8	1,88	9,9	9,9	No
20	Dena Rahayu Binti Dede Tarlan	2,79	1,13	0,1	0,1	Yes
21	Dewi Nurlela	3,91	1,58	0,1	0,1	Yes
22	Diah Ayu Kusumaningrum Binti I Made Suela	2,79	1,13	0,1	0,1	Yes

Source: The Tangerang City Correctional Institution

Table 1 above presents a dataset containing information regarding inmates in 2024, focusing on four indicators measured: criminal legacy value, Criminal residual value, criminal offense value, and risk dimension value, as well as remission status (whether or not remission is granted). This Dataset consists of 22 inmates, all of whom are status with complete and detailed data. Although the total number of inmates in the system is larger, this table illustrates the relevant and updated inmate data according to the categories analyzed.

In the table, each row represents one inmate, with columns that record the inmate's name, criminal old value, Criminal residual value, criminal offense value, risk dimension value, and their remission status. For example, Cut Marsidah inmates have a criminal length of 2.89 years, a criminal remainder of 1.22 years, a criminal offense of 20, and a risk dimension of 20. However, the status of remission given to Cut Marsidah is not. This shows that although the inmate has a lower criminal and residual criminal old score, other factors, such as the criminal offense and the risk dimension, influence the decision of remission given.

On the other hand, inmates such as Ana Pariha (aka Aray Bint Edi Sutrisni, who had died) had a much lower score on criminal length (3.91) and a very small risk dimension value (0.1). Despite the low value of the criminal offense and the risk dimension, these prisoners were given remission with Yes status, reflecting the possibility that other indicators (for example, behavior during the criminal period or other criteria) may have influenced the decision.

Similar things can also be seen in other prisoners, such as Tri Wulandari with a criminal history of 3.32, which has a criminal residual value of 1.42, criminal acts and risk dimensions of 20 each. The remission Status given is no. This signifies that although some inmates have similar figures for criminal remaining and criminal offenses, remission decisions still vary based on other dimensions of risk and behavior.

This table shows the complexity of remission decisions, which are not only determined by the length of the criminal or residual criminal, but also influenced by criminal factors and risk dimensions that need to be analyzed further to understand more in-depth patterns. By converting the yes and no States into numerical values, further statistical analysis can be performed to look for patterns or relationships between these variables and remission decisions.

b. Preprocessing

The Dataset in this study was imported using pandas library. The target column named remission contains the categorical values 'yes' and 'no', which are then converted to numerical values via the map function. The value of ' YES 'is changed to 1, while' no ' is changed to 0. This conversion is done so that the target data can be used properly by machine learning algorithms.

Independent variables that include The Old value of the crime, the residual value of the crime, the value of the crime, and the value of the risk dimension is selected as a feature (X), while the remission column is used as a target (y). Next, the dataset is divided into training data (80%) and test data (20%) using the train_test_split function.

Tabel 1. Conversion Results Dataset

No	Name Of Inmate and Child	Criminal Old Value	Criminal Residual	Value Criminal Offense	Risk Dimension Value	Remission
1	Cut Marsidah	2,89	1.22	20	20	0
2	Tri Wulandari	3,32	1,42	20	20	0
3	Ana Pariha Alias Aray Binti Edi Sutrisni (Alm)	3,91	1,58	0,1	0,1	1
4	Andririni Yaktiningsasi Binti Rm Oemar Said (Alm)	3,91	1,58	0,1	0,1	1
5	Anggi Rusidi Binti Rusdiman	3,91	1,58	0,1	0,1	1
6	Anisya Gelora Binti Tan Priasa	3,07	1,25	0,1	0,1	1
7	Anita Maha Pratiwi Alias Tiwi Binti Abdi Pria Dinata	3,07	1,25	0,1	0,1	1
8	Annis Pitria Binti Mawardi	3,07	1,25	0,1	0,1	1
9	Astri Delima Binti Alm Suwandi	3,35	1,36	0,1	0,1	1
10	Elisa Julia Ningsih	3,92	1,46	9,9	9,9	0
11	Aulia Andayani Alias Aulia Bin Asep Rustandi	3,35	1,36	0,1	0,1	1
12	Ayu Novita Sari Binti Tulam	3,91	1,58	0,1	0,1	1
13	Bainah Bt. Usman Bin Usman	4,19	1,7	0,1	0,1	1
14	Budiyah Bt Punas	2,79	1,13	0,1	0,1	1
15	Chofifah Aulia Binti Gunawan	2,51	1,02	0,1	0,1	1
16	Dayu Rizki Limayuli Alias Yuris Binti Dadang M. S.	2,51	1,02	0,1	0,1	1
17	Debby Novelia Binti Febrial	3,91	1,58	0,1	0,1	1
18	Delvia Agustin Binti Agus Matatula	2,65	1,08	0,1	0,1	1
19	Eliza Mahdelia	3,8	1,88	9,9	9,9	0
20	Dena Rahayu Binti Dede Tarlan	2,79	1,13	0,1	0,1	1
21	Dewi Nurlela	3,91	1,58	0,1	0,1	1
22	Diah Ayu Kusumaningrum Binti I Made Suela	2,79	1,13	0,1	0,1	1

Table 2 above shows the conversion dataset that contains information about 22 inmates along with indicators of Criminal duration, criminal residual, criminal offense, risk dimensions, and their remission status. Remission values are converted into a numerical format, with a value of 1 for prisoners who get remission and 0 for those who do not get it. With this change, the dataset becomes better prepared for further analysis or machine learning model training. In general, the criminal length value in this dataset varies between 2.51 to 4.19 years, reflecting how long inmates have served their sentences. The criminal residual value also shows that most prisoners have only relatively little remaining sentence, which is between 1.02 to 1.88 years. This shows that they are close to completing their sentences.

However, even though inmates have served most of their penal terms, their remission status does not necessarily depend solely on the old value of the Penal and residual Penal. Most of

the inmates in this dataset have low crime scores and risk dimensions, indicating that the crimes committed and the level of risk posed by those inmates are relatively low. For most inmates, the criminal offense and risk dimensions are worth 0.1, which means they are considered to have an insignificant risk. However, there are some exceptions that attract attention. Inmates such as Cut Marsidah and Tri Wulandari have very high criminal values and risk dimensions (20 each), which is most likely the reason they do not get remission even though they have already served most of their sentences. In addition, Elisa Julia Ningsih and Eliza Mahdelia had very high values of crime and risk dimensions (9,9), which showed that the severity of their crime and the level of risk posed were important considerations in the decision to grant remission.

From these data, it can be seen that although most prisoners have a relatively low value of Criminal length and criminal Residual, the granting of remission is not only based on the duration of the sentence that has been served. The decision to grant remission is strongly influenced by the value of the criminal act and the risk dimension, which reflects the severity of the criminal acts committed by the convict as well as the potential risks that may arise from them. Thus, although the prisoner has served most of his sentence, factors such as the severity of the criminal act and the risk to society are the main considerations in the decision on remission given.

c. Normalisasi Data

Normalization process is done by using StandardScaler method to equalize the scale of all features so that it has an average of 0 and a standard deviation of 1. This step is very important because the KNN algorithm is sensitive to differences in data scale (Ravat, 2023). The training Data is normalized using the fit_transform function, while the test data is normalized using transform to prevent information leakage from the test data to the training process.

d. Using the K-Nearest Neighbor algorithm

Application of K - Nearest Neighbor calculation as follows:

1) Determine the value of k=1

KNN Model is built using KNeighborsClassifier library with initial parameter k=1. The Model is trained on training data that has been normalized and then used to predict the class on the test data (Iwan Sudipa et al., 2024).

2) Selecting feature and target columns

The feature column Data are old criminal value, residual criminal Value, criminal offense value and risk dimension value. Target column data in the form of remission Status

3) Determine the Training Data and Testing Data

Training Data used is 80% which is 131 records and Testing Data is 20% which is 22 records. The 80:20 ratio between training and testing was chosen because it provides an optimal balance between the amount of data used to train the model and the data used to test its performance. This ratio allows the model to be trained with enough data so that it can learn effectively, while still providing sufficient data for a representative evaluation of the generalizability of the model on data that has not yet been seen. An 80:20 ratio is a common and efficient practice in this study, given the size of the dataset used.

Alternatives such as 70:30 cross-validation or k-fold cross-validation are also considered. Although k-fold can provide more accurate estimates of model performance, this method requires more computational time because it involves training and testing the model multiple times. Taking into account computational efficiency and the need to ensure representative results, the 80:20 ratio was chosen as the appropriate method to provide effective results without worsening computational time.

4) Model experiment and testing and model evaluation

Testing is done through several metrics, such as:

a) Model accuracy: the evaluation results showed an accuracy of 93.9% (obtained from the accuracy_score function).

b) Classification report: displays precision, recall, f1-score, and number of samples for each class.
c) Confusion Matrix: displayed both in visual form using heatmap to provide insight into the distribution of model errors.

In Figure 2, the evaluation results of the model shown in the figure show that this model has a fairly good overall performance, with an accuracy rate of 93.94%. Nonetheless, a more in-depth evaluation using the precision, recall, and f1-score metrics provides a more complete picture of the model's performance in classifying two classes: no remission (Class 0) and yes remission (Class 1). In the classification report, the precision for the non-remitting class was recorded at 0.00, which indicates that the model did not manage to correctly predict the cases that should be in the non-remitting class. In contrast, all predictions made by the model for the no-remission class are precisely false, with the model predicting them all as the Yes-remission class. This indicates a significant misclassification of the non-remitting class, which is a major concern for further improvement.

Meanwhile, for the Yes remission class, the model's precision was recorded at 0.94, which means that 94% of all positive predictions made by the model for this class were correct. This shows that the model is quite good at recognizing prisoners who are entitled to remission, although there are still some errors in predicting non-remission classes. Recall metrics provide information about the model's ability to detect all relevant cases in the dataset. For non-remitting classes, the recall model is 0.00, which means that the model does not manage to detect at all inmates who should not get remission. In contrast, the recall for the Yes remission class reached 1.00, which indicates that the model managed to predict all the cases that should belong to the Yes remission class perfectly, without missing anything.

The F1-score value, which combines precision and recall, gives a more balanced picture of the model's performance. The F1-score for the non-remitting class was 0.00, corroborating that the model could not predict this class well. For Class Ya remission, f1-score was recorded at 0.97, which indicates that the model has excellent performance in predicting class Ya remission, although there is still potential for improvement regarding classification no remission.

The Confusion matrix shown in the form of a heatmap further illustrates the distribution of predictions made by the model. It can be seen that the model correctly predicted 31 samples as remission Yes class, which shows that the model has excellent performance in recognizing prisoners entitled to remission. However, for the class of no remissions, the model predicts only 2 samples, and both are misclassified as Yes remissions. This indicates that although the model has high accuracy overall, the main problem lies in its inability to predict the class of non-remission correctly.

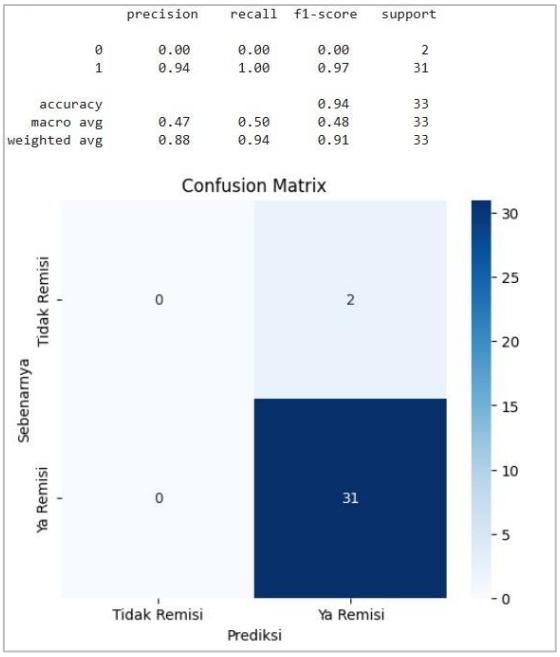

Figure 2. Result Model Evaluation

Overall, although this model shows a high degree of accuracy, performance on the class is not very bad remission. The Model managed to recognize almost all prisoners who were entitled to remission, but failed significantly in classifying those who should not have remission. Therefore, improvements in the model's ability to predict non-remission classes are important steps to improve overall performance, so that the model can provide more balanced and accurate predictions in both classes.

CONCLUTION

The conclusion of this study summarizes the analysis of the dataset, the evaluation model's results, and the model's performance in predicting the remission status of prisoners. The dataset, which includes 22 inmates with indicators such as criminal duration, residual sentence, crime type, and risk dimensions, shows that the applied model performs well in many areas but also highlights areas requiring improvement.

The model achieved a high accuracy of 93.94%, indicating that, overall, it can predict remission status effectively. However, further evaluation using precision, recall, and F1-score metrics reveals that while the model performed excellently for the "Yes" remission class (with a precision of 0.94, recall of 1.00, and F1-score of 0.97), it faced significant challenges with the "No Remission" class. For this class, all metrics (precision, recall, F1-score) scored 0.00, showing the model's failure to predict prisoners who are not entitled to remission. Analysis of the confusion matrix further confirms that the model misclassified all "No Remission" cases, predicting them as "Yes Remission." This imbalance indicates the need for improvements to address the classification issue. While the overall accuracy is high, the inability to predict "No Remission" correctly is a significant limitation of the model.

To enhance the model's performance, future research should focus on improving the prediction of the "No Remission" class. This can be achieved by addressing data imbalance through techniques like oversampling or undersampling, adjusting model parameters, or exploring different classification algorithms better suited for imbalanced datasets. Additionally, future studies could examine the incorporation of more features or data from different correctional institutions to improve the model's generalizability. By implementing these improvements, future

research can develop more balanced and accurate models, leading to more reliable remission status predictions.

REFERENCES

Cholil, M., & others. (2021). Optimizing k-Nearest Neighbor Algorithm for Classification Problems. *Jurnal Teknologi Dan Sistem Komputer*, *7*(1), 42–58.

Creswell, J. W. (2014). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE Publications.

Danny, M., Muhidin, A., & Jamal, A. (2024). Application of the K-Nearest Neighbor Machine Learning Algorithm to Preduct Sales of Best-Selling Products. *Brilliance: Research of Artificial Intelligence*, *4*(1), 255–264. https://doi.org/10.47709/brilliance.v4i1.4063

Han, J., & Kamber, M. (2007). Data Mining: Concepts and Techniques The Explosive Growth of Data: from terabytes to petabytes. *Data Mining: Concepts and Techniques*.

Iwan Sudipa, I. G., Azdy, R. A., Arfiani, I., Setiohardjo, N. M., & Sumiyatun. (2024). Leveraging K-Nearest Neighbors for Enhanced Fruit Classification and Quality Assessment. *Indonesian Journal of Data and Science*, *5*(1), 30–36. https://doi.org/10.56705/ijodas.v5i1.125

Khotiah, T., Sugianela, Y., Setiowati, D., & Arianto, F. (2022). Classification Model of Prisoner Remission Data Using K-Nearest Neighbors Algorithm with Genetic Algorithm Optimization. *Jurnal Ilmiah Intech : Information Technology Journal of UMUS*, *4*(01), 87–96.

Lau, B. P. L., & others. (2019). A Survey of Data Fusion inSsmart City Applications. *Information Fusion*, *52*(January), 357–374. https://doi.org/10.1016/j.inffus.2019.05.004

Nazar, R. (2024). Implementasi Pemrograman Python Menggunakan Google Colab. *Jurnal Informatika Dan Komputer (JIK)* , *15*(1), 50–56.

Nuraeni, F., Kurniadi, D., & Diazki, M. H. (2024). Algoritma K-Nearest Neighbor pada Kasus Dataset Imbalanced untuk Klasifikasi Kinerja Karyawan Perusahaan. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *11*(3), 557–568. https://doi.org/10.25126/jtiik.938144

Peraturan Pemerintah RI. (1999). *Peraturan Pemerintah Republik Indonesia tentang Syarat dan Tata Cara Pelaksanaan Hak Warga Binaan Pemasyarakatan (PP Nomor 32 Pasal 1 ayat 6 Tahun 1999)*.

Permenkumham. (2018). *Peraturan Menteri Hukum dan Hak Asasi Manusia Nomor 3 Tahun 2018 Pasal 3 Tentang Syarat dan Tata Cara Pemberian Remisi, Asimilasi, Cuti Mengunjungi Keluarga, Pembebasan Bersyarat, Cuti Menjelang Bebas, dan Cuti Bersyarat.*

Permenkumham. (2022). *Peraturan Menteri Hukum dan Hak Asasi Manusia Nomor 7 Pasal 4 Tahun 2022 Tentang Perubahan Kedua Atas Peraturan Menteri Hukum dan Hak Asasi Manusia.*

Ravat, A. (2023). *K-Nearest Neighbor (KNN) model for carrier data classification in security domain*. *1*(1), 35–47.

Sakti, R., & Daulay, A. (2024). *Analisis Kritis dan Pengembangan Algoritma K-Nearest Neighbor (KNN): Sebuah Tinjauan Literatur*. *4*(2), 131–141.

Susilo, B., Ramdhan, N. A., & Bachri, O. S. (2024). *Application of the K-Nearest Neighbor Algorithm for Predicting Digital Product Sales Penerapan Algoritma K-Nearest Neighbor untuk Prediksi Penjualan Produk Digital*. *4*(October), 1466–1476.

Tiwari, V. R. (2023). Developments in KD tree and KNN searches. *International Journal of Computer Applications*, *975*, 8887.

Ujianto, N. T., Fadilah, H., Fanti, A. P., & Saputra, A. D. (2025). *IT-EXPLORE Penerapan algoritma K-Nearest Neighbors (KNN) untuk klasifikasi citra medis*. *02*, 33–43.

Undang-Undang RI. (1995). Undang-Undang Republik Indonesia Nomor 12 Tahun 1995 Tentang Pemasyarakatan. In *tentang Permasyarakatan* (Vol. 66, Issue September, pp. 37–39).

Wan, B., Sheng, Z., Zhu, W., & Hu, Z. (2024). Compactness-Weighted KNN Classification Algorithm. *International Journal of Advanced Computer Science and Applications*, *15*(9), 229–238. https://doi.org/10.14569/IJACSA.2024.0150922