JURNAL TEKNOLOGI DAN OPEN SOURCE
Vol. 9, No. 1, June 2026, pp. 390~410
e-ISSN: 2622-1659, accredited Four Grade by Kemenristekdikti, Decree No: 152/E/KPT/2023
DOI: https://doi.org/10.36378/jtos.v9i1.5724

Development of a Machine Learning Model for Predicting River Pollution Levels Caused by Illegal Gold Mining Activities in Kuantan Singingi Regency

Jasri[1], Aprizal[2], Erlinda[3], Febri Haswan[4], Amirel Hafief Agiel[5]
[1,2,3,4,5]Fakultas Teknik, Program Studi Teknik Informatika, Universitas Islam Kuantan Singingi, Riau, Indonesia

Article Info

Article history:

Received 05 01, 2026
Revised 06 11, 2026
Accepted 06 25, 2026

Keywords:

Machine Learning, River Pollution, Random Forest, Water Quality, Illegal Gold Mining

ABSTRACT

Illegal Gold Mining (PETI) activities in Kuantan Singingi Regency have caused river water pollution and posed serious threats to environmental sustainability and public health. Conventional water quality monitoring methods still have limitations because they rely on periodic laboratory testing and are unable to provide rapid predictive results. Therefore, this study developed a Machine Learning-based prediction system to analyze river pollution levels caused by illegal gold mining activities. The study utilized water quality parameters consisting of pH, temperature, Total Suspended Solid, Dissolved Oxygen, Biological Oxygen Demand, Chemical Oxygen Demand, and mercury concentration. The dataset was processed through several preprocessing stages, including data cleaning, normalization, feature selection, and data splitting. Several Machine Learning algorithms, namely Random Forest, Support Vector Machine, and Artificial Neural Network, were implemented and compared to determine the best prediction model. The results showed that the Random Forest algorithm achieved the best performance with high accuracy and stable classification results. Furthermore, the developed model was integrated into a web-based system equipped with pollution visualization features, river information, and a Geographic Information System. The system is expected to support environmental monitoring and assist decision-making in river pollution management in Kuantan Singingi Regency.

Corresponding Author:

Febri Haswan
Program Studi Teknik Informatika
Universitas Islam Kuantan Singingi
Riau, Indonesia
Email: febri.haswan88@gmail.com

1. Introduction

Illegal Gold Mining (PETI) activities have become one of the significant environmental problems in various regions of Indonesia, including Kuantan Singingi Regency, Riau Province. These activities generally use the amalgamation method with mercury (Hg), which has the potential to contaminate water bodies, especially rivers that serve as the main source of life for local communities. Mercury accumulated in aquatic environments can be transformed into toxic methylmercury, posing serious threats to aquatic organisms and human health through the food chain [1]. Therefore, water quality monitoring has become an important aspect in controlling the environmental impacts caused by illegal gold mining activities [14], [15].

However, the water quality monitoring systems currently implemented are still conventional, relying on periodic sampling and laboratory analysis. This approach has limitations in terms of spatial and temporal coverage and is not capable of providing early predictions of changes in pollution levels. In addition, delays in data processing make decision-making less responsive to dynamic environmental conditions [18], [19]. Therefore, a technology-based approach is needed to process data quickly and provide accurate predictions as a basis for decision-making.

Along with the development of information technology, Machine Learning (ML)-based approaches have been widely applied in various fields, including water quality modeling. Machine Learning is a branch of artificial intelligence capable of learning patterns from historical data to predict future conditions [6], [30]. Several ML algorithms such as Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Network (ANN) have proven effective in modeling complex and non-linear relationships among environmental parameters [2]–[4]. Previous studies have shown that ML-based models provide higher accuracy than conventional statistical methods in water quality prediction [7], [9], [20].

In addition, the implementation of data mining and data preprocessing techniques such as data cleaning, normalization, and feature selection significantly affects the performance of the resulting model [5], [11]. The integration of data processing techniques and ML algorithms enables the development of more robust and adaptive prediction systems for environmental changes [23], [24]. Furthermore, prediction results can be integrated into web-based systems to facilitate data visualization and interpretation, allowing them to function as decision support systems in environmental management [21], [29].

Although many studies have been conducted on water quality prediction using Machine Learning, most research has focused on industrial areas or large-scale regional studies [16], [18]. Studies specifically examining the impact of illegal gold mining activities on river water quality at the regency level, particularly in Kuantan Singingi Regency, are still very limited. In addition, the integration of prediction models with web-based systems as decision support tools has not been comprehensively developed.

Based on these problems, this study aimed to develop a Machine Learning model capable of predicting river pollution levels caused by illegal gold mining activities in Kuantan Singingi Regency. The developed model applied a comparative approach using several algorithms, namely Random Forest, Support Vector Machine, and Artificial Neural Network, to obtain the best-performing prediction model. Furthermore, the prediction model was integrated into a web-based system in the form of an interactive dashboard.

The main contribution of this study lies in the development of an integrated computational pipeline covering data processing, modeling, evaluation, and system implementation. The novelty of this research is the integration of water quality parameters with a Machine Learning approach in the context of illegal gold mining activities at the local regency scale, as well as its implementation in a web-based prediction system that can be used as a decision support system. Therefore, this study not only provides theoretical contributions in the field of Machine Learning but also practical contributions to supporting data-driven environmental management.

2. Research Method

2.1 Research Design

This study employed a quantitative computational approach to develop a prediction system for river water pollution levels caused by Illegal Gold Mining (PETI) activities. The research design was experimental and consisted of several main stages, including data collection, preprocessing, prediction model design, evaluation, and web-based system implementation. The research approach referred to the concepts of data mining and knowledge discovery, which emphasize data processing to generate predictive information as decision-making support [5], [6].

The prediction model was developed based on water quality parameters such as pH, temperature, Total Suspended Solid (TSS), Dissolved Oxygen (DO), Biological Oxygen Demand (BOD), Chemical Oxygen Demand (COD), and mercury (Hg) concentration, which were processed using a computational approach to determine river pollution levels.

2.2 Research Flow Diagram

The research flow diagram is presented in Figure 1, illustrating the overall stages of the study, starting from literature review and variable identification, collection of water quality parameter data, data preprocessing, pollution level analysis, GIS-based data visualization, and finally the implementation of a web-based monitoring dashboard system. Each stage is interconnected and designed to ensure systematic data processing, accurate analysis results, and effective visualization of environmental information to support decision-making and water quality monitoring activities.

Figure 1. Research Flow Diagram

Figure 1 illustrates that the study was conducted systematically through several key stages, namely the identification of research variables, water quality data processing, pollution level calculation, monitoring result visualization, and the implementation of a web-based system. All stages are integrated to develop a river pollution monitoring system that can be utilized as a decision-support tool for environmental management.

2.3 Data Acquisition

The data used in this study consisted of a synthetic dataset developed based on the characteristics of river water quality affected by Illegal Gold Mining (IGM) activities in Kuantan Singingi Regency, Riau Province, Indonesia. The use of a synthetic dataset was necessary due to the limited availability of comprehensive historical water quality data that were continuously measured and covered all study locations. The dataset was constructed using information derived from literature reviews, previous studies, water quality standards, and documented characteristics of river pollution associated with illegal gold mining activities commonly found in Kuantan Singingi Regency.

The dataset was designed to represent river water quality conditions in several areas affected by illegal gold mining activities, including Logas, Teluk Kuantan, Cerenti, Benai, Kuantan Mudik, Gunung Toar, Pangean, Singingi, Singingi Hilir, Hulu Kuantan, Sentajo Raya, and Inuman. The synthetic data were generated to reflect variations in water quality conditions during the observation period from January to March 2026.

The water quality parameters used as input variables in this study included:
a) pH
b) Water temperature (°C)
c) Total Suspended Solids (TSS)
d) Dissolved Oxygen (DO)
e) Biological Oxygen Demand (BOD)
f) Chemical Oxygen Demand (COD)
g) Mercury (Hg) concentration

In addition to the primary water quality parameters, several environmental and spatial variables were included as supporting factors due to their potential influence on river pollution levels. These variables consisted of:
a) Rainfall data
b) River discharge data
c) Geographic coordinates (latitude and longitude)
d) Distance between the sampling location and ASGM activity sites

All of these parameters were utilized as input attributes in the Machine Learning model development process to predict river pollution levels. The selected variables represent key indicators of water quality and environmental conditions that are potentially affected by illegal gold mining activities.

The dataset was organized in a numerical tabular format and used as input for both the training and testing phases of the Machine Learning models. An example of the dataset structure used in this study is presented in Table 1.

Table 1. Research Dataset Structure

No.	pH	Temperature (°C)	TSS (mg/L)	DO (mg/L)	BOD (mg/L)	COD (mg/L)	Hg (mg/L)	Rainfall (mm)	River Discharge (m^3/s)	Distance from ASGM Site (km)	Target
1	6.8	28	120	5.2	3.5	20	0.005	120	15	2.5	Polluted
2	7.1	27	90	6.1	2.8	18	0.003	100	12	4.0	Normal

Description:

The target is a category of river pollution level based on a combination of water quality parameters used in the prediction system.

This study used 150 data records, consisting of four pollution level categories: Safe, Light, Moderate, and Heavy. The dataset distribution is shown in Table 2.

Table 2. Dataset Distribution Based on Pollution Level

Pollution Level	**Number of Records**
Safe	40
Lightly Polluted	35
Moderately Polluted	40
Heavily Polluted	35
Total	**150**

For the Machine Learning modeling process, the dataset was divided into two groups: training data and testing data. Data was divided using a random split method with a 70:30 ratio. A total of 105 records were used as training data and 45 records were used as testing data. The dataset division is shown in Table 3.

Table 3. Dataset Distribution

Data Type	**Number of Records**	**Percentage**
Training Data	105	70%
Testing Data	45	30%
Total	**150**	**100%**

This data sharing ensures that the machine learning model can learn patterns from the training data and generalize effectively to previously unused data. The compiled dataset is then used in preprocessing, model training, algorithm performance evaluation, and implementation of a web-based river pollution prediction system.

2.4 Data Preprocessing

The preprocessing stage is a crucial step before data is used in machine learning model training. The primary goal of preprocessing is to improve data quality, reduce noise, and ensure each variable has the appropriate format and scale for the model's learning process. Poorly processed data can lead to decreased accuracy and performance of the predictive model. Therefore, this study implemented several preprocessing stages, including data cleaning, data normalization, feature selection, and dataset segmentation.

1. Data Cleaning
   The initial preprocessing step involved data cleaning to ensure that the dataset was valid, reliable, and consistent. This process included:
   a) Handling missing values using mean or median imputation methods.
   b) Removing duplicate records to avoid data redundancy.
   c) Detecting and treating outliers through basic statistical analysis techniques.

   These procedures were performed to enhance the quality of the dataset before it was used for model training [5].

2. Data Normalization
   Data normalization was applied to standardize the value ranges of different variables, thereby preventing certain features from dominating the Machine Learning learning process.

   The normalization technique employed in this study was Min-Max Scaling, which transforms data into a range between 0 and 1 using the following equation:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \qquad \text{...(1)}$$

   Where:
   X = original data value
   X_{min} = minimum value of the feature
   X_{max} = maximum value of the feature

X^1 = normalized value after applying Min-Max Scaling

This method transforms the data range into a scale between 0 and 1, thereby improving the stability and predictive performance of the Machine Learning model [6].

3. Feature Selection

Feature selection was performed to identify and select variables that have a significant influence on river pollution levels. The method employed in this study was Pearson Correlation Analysis.

The Pearson correlation coefficient is used to measure the strength and direction of the linear relationship between each feature and the target variable. Features with higher correlation values are considered more relevant and are retained for the model development process.

The Pearson correlation equation is presented in Equation (2).

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2)$$

Where:
r = correlation coefficient
X_i, Y_i = observed values of variables X and Y
$\bar{X}$, $\bar{Y}$ = mean value of variable X and Y

The correlation coefficient ranges from −1 to 1, with the following interpretations:
$r \approx 1$: strong positive relationship
$r \approx -1$: strong negative relationship
$r \approx 0$: no relationship

Variables with the highest correlation coefficients were selected as the primary features for the modeling process, as they are considered to have the greatest influence on river pollution levels.

4. Data Splitting

After preprocessing was complete, the dataset was divided into training data and testing data. The division was performed using a random split method with a 70:30 ratio.

Table 4. Dataset Division for Modeling

Data Type	Number of Records	Percentage
Training Data	105	70%
Testing Data	45	30%
Total	**150**	**100%**

Training data is used to build a machine learning model, while testing data is used to evaluate the model's ability to predict data that has never been used before. This division aims to reduce the risk of overfitting and improve the model's generalization ability [7].

2.5 Model Development

At this stage, Machine Learning models were developed to predict river water pollution levels resulting from illegal gold mining activities (Artisanal and Small-Scale Gold Mining/ASGM). The models were built using a supervised learning approach, which involves mapping input data, represented by water quality parameters, to a target output representing the river pollution level.

This study employed three Machine Learning algorithms: Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Network (ANN). These algorithms were selected due to their ability to handle complex non-linear relationships among variables and their proven effectiveness in various water quality prediction studies [7], [8].

1. Random Forest (RF)

   Random Forest is an ensemble learning method that combines multiple decision trees to improve prediction accuracy and reduce the risk of overfitting [2]. Each decision tree is constructed using randomly selected subsets of data and features through a technique known as bootstrap aggregating (bagging).

The Random Forest procedure consists of the following steps:

a) Randomly dividing the dataset into multiple subsets.
b) Constructing a decision tree for each subset of data.
c) Combining the prediction results using averaging (for regression tasks) or majority voting (for classification tasks).

Mathematically, the Random Forest prediction can be expressed as follows:

$$\hat{y} = \frac{1}{N}\sum_{i=1}^{N} T_i(x) \qquad \text{...(3)}$$

Where:
$\hat{y}$ = predicted output (prediction result)
$T_i\,(x)$ = output generated by the i-th decision tree
N = total number of decision trees in the Random Forest model

Random Forest offers several advantages, including its ability to handle datasets with a large number of features and its strong generalization capability when applied to previously unseen data. Furthermore, the ensemble nature of the algorithm helps improve prediction accuracy and reduces the likelihood of overfitting.

2. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a Machine Learning algorithm that operates by identifying the optimal hyperplane to separate data into distinct classes or to perform regression tasks [3].

The working principles of SVM include:
a) Determining the best hyperplane as the decision boundary between data classes.
b) Maximizing the margin between classes to improve classification performance.
c) Utilizing kernel functions to handle non-linear data relationships.

By transforming the input data into a higher-dimensional feature space, SVM can effectively model complex patterns and improve predictive accuracy, particularly when dealing with non-linear datasets.

The basic SVM model can be expressed as follows:

$$f(x) = w \cdot x + b \qquad \text{...(4)}$$

Where:
w = weight vector that determines the orientation of the hyperplane
x = input data (feature vector)
b = bias term that shifts the hyperplane position

For non-linear classification or regression problems, kernel functions are employed to transform the input data into a higher-dimensional feature space. Commonly used kernel functions include the Radial Basis Function (RBF) kernel, polynomial kernel, and linear kernel.
Among these, the RBF kernel is widely used because of its effectiveness in capturing complex non-linear relationships within the data.

The RBF kernel function is expressed as follows:

$$K(x_i, x_j) = \exp(-\gamma||x_i - x_j||^2) \qquad \text{...(5)}$$

The Support Vector Machine (SVM) algorithm possesses strong generalization capabilities and is highly effective when applied to high-dimensional datasets, making it suitable for complex environmental prediction problems such as river pollution assessment.

3. Artificial Neural Network (ANN)

Artificial Neural Network (ANN) is a computational model inspired by the structure and functioning of biological neural networks. ANN is capable of learning complex non-linear relationships among variables and has been widely applied in prediction and classification tasks [4].

The architecture of an ANN consists of the following components:
a) Input Layer, which receives the feature data.
b) Hidden Layer(s), where pattern learning and information processing take place.
c) Output Layer, which produces the final prediction results.

During the learning process, each neuron receives inputs from connected neurons, multiplies them by their corresponding weights, adds a bias term, and then applies an activation function to generate an output.

The mathematical representation of a neuron in an ANN is given as follows:

$$y = f\left(\sum_{i=1}^{n} w_i x_i + b\right) \quad \text{...(6)}$$

Where:
x_i = input value of the i-th feature
w_i = weight associated with the i-th input
b = bias term
f = activation function (e.g., ReLU, sigmoid, or other activation functions)

The ANN training process consists of three main stages:
1) Forward Propagation, where input data are passed through the network to generate predictions.
2) Error Calculation, where the difference between predicted and actual values is computed using a loss function.
3) Backpropagation, where the error is propagated backward through the network to update the weights and biases, thereby improving model performance.

ANN demonstrates strong capabilities in recognizing complex patterns and capturing non-linear relationships among variables. Therefore, it is particularly suitable for environmental data modeling, where interactions among water quality parameters are often complex and non-linear.

4. Model Training and Optimization

Each model was trained using 70% of the dataset as training data and evaluated using the remaining 30% as testing data. To enhance model performance and improve its generalization capability, hyperparameter optimization was performed using a cross-validation technique.

The hyperparameters optimized in this study include:
a) Random Forest (RF):
    1) Number of trees (n_estimators)
    2) Maximum tree depth (max_depth)
b) Support Vector Machine (SVM):
    1) Kernel type
    2) Regularization parameter (C)
    3) Gamma parameter (γ)
c) Artificial Neural Network (ANN):
    1) Number of neurons in the hidden layer(s)
    2) Learning rate
    3) Number of training epochs (epochs)

5. Best Model Selection

The best-performing model was selected based on the evaluation results obtained using the following performance metrics:
a) Root Mean Square Error (RMSE)
b) Mean Absolute Error (MAE)

c) Koefisien determinasi (R^2)

The model with the highest coefficient of determination (R^2) and the lowest prediction errors was selected as the best-performing model for implementation in the web-based river pollution prediction system. This selection criterion ensures that the chosen model provides accurate, reliable, and robust predictions for supporting river water quality monitoring and environmental decision-making.

2.8 System Implementation

The river pollution prediction model developed in the previous stage was implemented into a web-based system using a client–server architecture. The system was designed as a Decision Support System (DSS) to assist in monitoring and analyzing river water pollution levels caused by Artisanal and Small-Scale Gold Mining (ASGM) activities in Kuantan Singingi Regency.

On the client side (frontend), the system provides a web-based user interface that enables users to input water quality parameters, including pH, temperature, Total Suspended Solids (TSS), Dissolved Oxygen (DO), Biological Oxygen Demand (BOD), Chemical Oxygen Demand (COD), and mercury (Hg) concentration. The input data are then transmitted to the server side for further processing by the prediction system.

On the server side (backend), the system performs data processing tasks, including preprocessing, data normalization, and pollution level prediction based on the water quality parameters utilized in this study. The prediction results are subsequently returned to the frontend and presented through interactive visualizations.

The developed monitoring system includes several key features:

a) River pollution level prediction presented as water quality condition categories.
b) Visualization of water quality parameters through graphical representations to support pollution analysis.
c) Interactive river monitoring dashboard for real-time data observation and analysis.
d) Geographic Information System (GIS)-based pollution mapping using digital maps to visualize pollution locations spatially.

The system was implemented using web-based technologies, allowing flexible accessibility and facilitating digital environmental monitoring. Through this system, water quality monitoring is expected to go beyond descriptive analysis by providing predictive insights into river pollution conditions, thereby supporting data-driven decision-making in environmental management and policy development.

6. Result and Discussion

3.1 Testing Dataset

System testing was conducted using river water quality sample data collected from several areas in Kuantan Singingi Regency affected by Artisanal and Small-Scale Gold Mining (ASGM) activities. The dataset consists of water quality parameters, including pH, temperature, Total Suspended Solids (TSS), Dissolved Oxygen (DO), Biological Oxygen Demand (BOD), Chemical Oxygen Demand (COD), and mercury (Hg) concentration.

The dataset was utilized to evaluate the performance of the river pollution prediction model and to assess its capability in classifying pollution levels based on the observed water quality conditions. Table 3 presents the sample data used in the river pollution prediction model testing process.

Table 5. Testing Sample Dataset

No.	Village	River Name	pH	Temperature (°C)	TSS (mg/L)	DO (mg/L)	BOD (mg/L)	COD (mg/L)	Hg (mg/L)
1	Logas	Kuantan River	6.8	28	120	4.2	15	45	0.7
2	Teluk Kuantan	Batang Kuantan River	7.1	29	80	5.5	10	30	0.3
3	Cerenti	Cerenti River	6.5	30	150	3.8	18	50	0.9
4	Benai	Benai River	7.0	28	70	6.0	8	25	0.2
5	Kuantan Mudik	Kuantan River	6.3	31	180	3.5	20	60	1.1
6	Gunung Toar	Toar River	6.9	29	95	4.8	12	35	0.5
7	Pangean	Pangean River	7.2	28	60	6.2	7	20	0.1
8	Singingi	Singingi River	6.4	30	200	3.2	22	65	1.3
9	Singingi Hilir	Singingi River	6.6	29	140	4.0	16	48	0.8
10	Hulu Kuantan	Kuantan River	7.0	27	75	5.8	9	28	0.2
11	Kuantan Tengah	Batang Kuantan River	6.7	28	110	4.5	14	40	0.6
12	Sentajo Raya	Sentajo River	7.1	29	65	6.1	7	22	0.1
13	Inuman	Inuman River	6.2	31	170	3.6	19	55	1.0

14	Cerenti Hilir	Cerenti River	6.5	30	155	3.9	17	52	0.95
15	Logas Hilir	Kuantan River	6.4	30	165	3.7	18	58	1.05

3.2 Data Preprocessing Results

The preprocessing stage was conducted before the dataset was used for Machine Learning model training. The preprocessing procedures included data cleaning, data normalization, and feature selection.

1) Data Cleaning
   At this stage, the dataset was examined for:
   a) Missing values
   b) Duplicate records
   c) Outliers

   Based on the inspection results, all sample data were found to be valid and suitable for use in the model testing process. No significant missing values, duplicate entries, or extreme outliers were identified that could adversely affect model performance.
2) Data Normalization
   Data normalization was performed using the Min-Max Scaling method to standardize the value ranges of different water quality parameters. This process was intended to prevent certain features from dominating the learning process due to differences in measurement scales.
   Parameters such as TSS, COD, and Hg exhibited considerably larger value ranges compared with other variables. Therefore, normalization was necessary to ensure balanced feature contributions and to improve the stability and predictive performance of the Machine Learning models.
3) Feature Selection
   Based on the results of the Pearson correlation analysis, the parameters that showed the strongest influence on river pollution levels were:
   a) Mercury concentration (Hg)
   b) Chemical Oxygen Demand (COD)
   c) Total Suspended Solids (TSS)
   d) Biological Oxygen Demand (BOD)
   In contrast, pH and water temperature demonstrated relatively weaker correlations with pollution levels and were therefore considered less influential predictors. These findings indicate that Hg, COD, TSS, and BOD are the most important variables for predicting river pollution conditions in areas affected by ASGM activities.

3.3 Pollution Level Prediction Results

The pollution level was calculated using a combination of water quality parameters and subsequently classified into three categories:

a) Safe
b) Moderately Polluted
c) Polluted

The classification categories were determined based on the pollution score generated by the prediction system. Higher scores indicate a greater level of river pollution, while lower scores indicate better water quality conditions.

Table 6. River Pollution Level Prediction Results

No.	River Name	Score	Status
1	Kuantan River (Logas)	38.15	Moderately Polluted
2	Batang Kuantan River	27.20	Safe
3	Cerenti River	47.55	Moderately Polluted
4	Benai River	23.80	Safe
5	Kuantan River (Kuantan Mudik)	56.40	Polluted
6	Toar River	32.65	Moderately Polluted
7	Pangean River	20.50	Safe
8	Singingi River	61.75	Polluted
9	Singingi Hilir River	45.30	Moderately Polluted
10	Kuantan River (Hulu)	25.60	Safe
11	Batang Kuantan River	35.45	Moderately Polluted
12	Sentajo River	21.80	Safe
13	Inuman River	53.70	Polluted
14	Cerenti Hilir River	49.80	Moderately Polluted
15	Kuantan River (Logas Hilir)	55.20	Polluted

Based on the prediction results presented in Table 4, the highest pollution level was observed in the Singingi River, which obtained a score of 61.75 and was classified as “Polluted.” The elevated pollution level is primarily associated with high values of Total Suspended Solids (TSS), Chemical Oxygen Demand (COD), Biological Oxygen Demand (BOD), and mercury (Hg) concentration, which are likely influenced by ASGM activities in the area.

In contrast, the Pangean River exhibited the lowest pollution level, with a score of 20.50, and was classified as “Safe.” This result indicates that the water quality at this location remains relatively good compared with other sampling sites included in the study. The lower concentrations of pollution-related parameters suggest a lesser degree of environmental disturbance and a healthier aquatic condition.

3.4 Machine Learning Model Evaluation Results

In this study, three Machine Learning algorithms were evaluated:

a) Random Forest (RF)
b) Support Vector Machine (SVM)
c) Artificial Neural Network (ANN)

The evaluation was performed using 70% of the dataset for training and 30% for testing. Model performance was assessed using several evaluation metrics, including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the Coefficient of Determination (R^2).

Table 7. Model Evaluation Results

Algoritma	RMSE	MAE	R^2
Random Forest	2.15	1.72	0.94
SVM	3.08	2.41	0.89
ANN	2.64	2.05	0.91

Based on the evaluation results presented in Table 5, the Random Forest algorithm achieved the best overall performance, yielding the lowest RMSE and MAE values as well as the highest R^2 value of 0.94. These results indicate that Random Forest was more effective in modeling the relationships among the water quality parameters than the other algorithms evaluated in this study.

Furthermore, Random Forest demonstrated greater robustness and stability when handling non-linear data patterns and the complex variability commonly found in environmental parameters. Its ensemble-learning mechanism enables the model to capture intricate relationships among variables while reducing the risk of overfitting, resulting in more accurate and reliable pollution level predictions. Consequently, Random Forest was selected as the most suitable model for implementation in the web-based river pollution prediction system.

3.5 System Design

The system was designed using the Unified Modeling Language (UML) approach to ensure that the system analysis, design, and implementation processes were conducted in a structured and comprehensible manner.

The system involves two primary types of users:

a) Public User
b) Administrator (Admin)

Public users are granted access to general information regarding river pollution conditions, monitoring results, and environmental data. In contrast, administrators possess full access privileges, enabling them to manage datasets, perform pollution predictions, monitor Machine Learning analysis results, and maintain the overall system functionality.

The use of UML facilitates the visualization of system requirements, user interactions, and system workflows, thereby improving communication between system developers and stakeholders throughout the development process.

A. Use Case Diagram

The Use Case Diagram is employed to illustrate the interactions between system users (actors) and the developed system. It provides an overview of the functionalities available to each user type and defines the scope of system operations.

Through the Use Case Diagram, the roles and responsibilities of both public users and administrators can be clearly identified, ensuring that the system requirements are accurately represented and implemented.

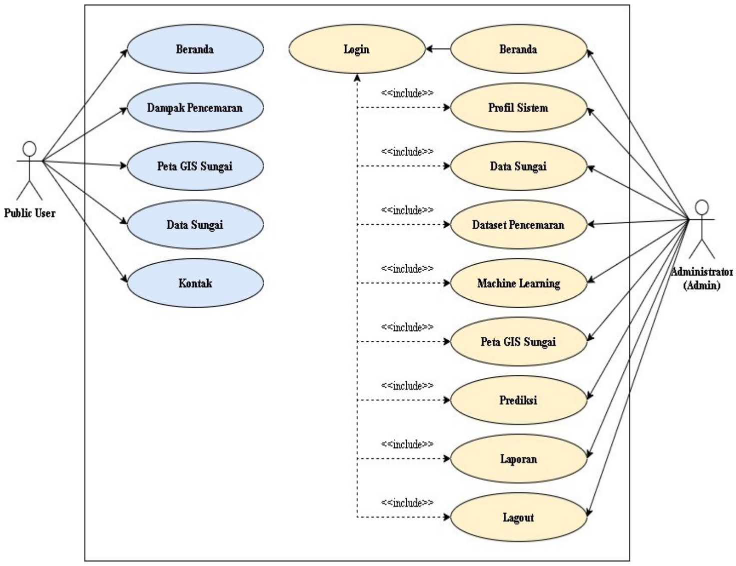


Figure 2. Use Case Diagram

B. Activity Diagram

The Activity Diagram is used to illustrate the workflow of the system, starting from data input and processing to the generation of river pollution predictions.

a) Public User Activity Diagram

The Public User Activity Diagram describes the workflow of general users (the public) when interacting with the web-based river pollution prediction system. The process begins when users access the system's homepage. They can then navigate through various menus, including pollution impact information, GIS-based river maps, river data, and contact information. The system subsequently displays the information corresponding to the menu selected by the user.

This diagram illustrates the interaction between public users and the system in a straightforward manner without requiring a login process. As a result, community members can easily access information regarding river conditions and pollution levels in Kuantan Singingi Regency, thereby improving public awareness and access to environmental information.

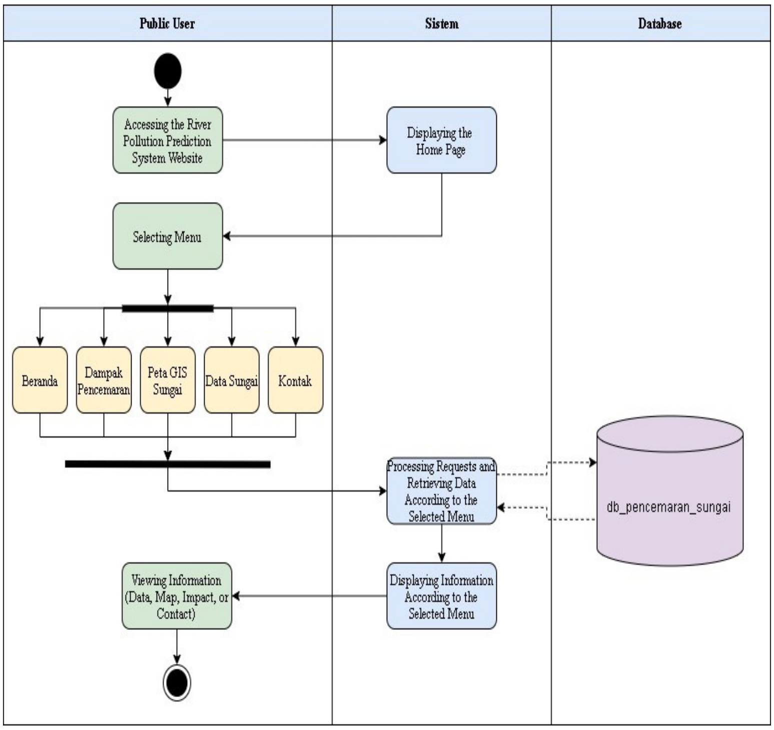


Figure 3. Public User Activity Diagram

b) Administrator Activity Diagram

The Administrator Activity Diagram illustrates the workflow of administrators in managing the web-based river pollution prediction system. The process begins when the administrator logs into the system using a valid username and password. After successful authentication, the administrator can access various system modules, including the dashboard, system profile, river data, pollution dataset, Machine Learning module, GIS-based river maps, prediction module, and reporting module.

Within the pollution dataset module, the administrator can manage water quality data, including pH, temperature, Total Suspended Solids (TSS), Dissolved Oxygen (DO), Biological Oxygen Demand (BOD), Chemical Oxygen Demand (COD), and mercury (Hg) concentration. These data are subsequently processed using the implemented Machine Learning models to generate river pollution level predictions.
In addition, the administrator can view analytical results through graphical visualizations and GIS-based maps, monitor prediction outcomes, and generate or print reports for documentation and decision-making purposes.

This diagram demonstrates that the administrator possesses full access privileges over data management, analytical processes, and the operation of the river pollution prediction system, ensuring effective monitoring and environmental management.

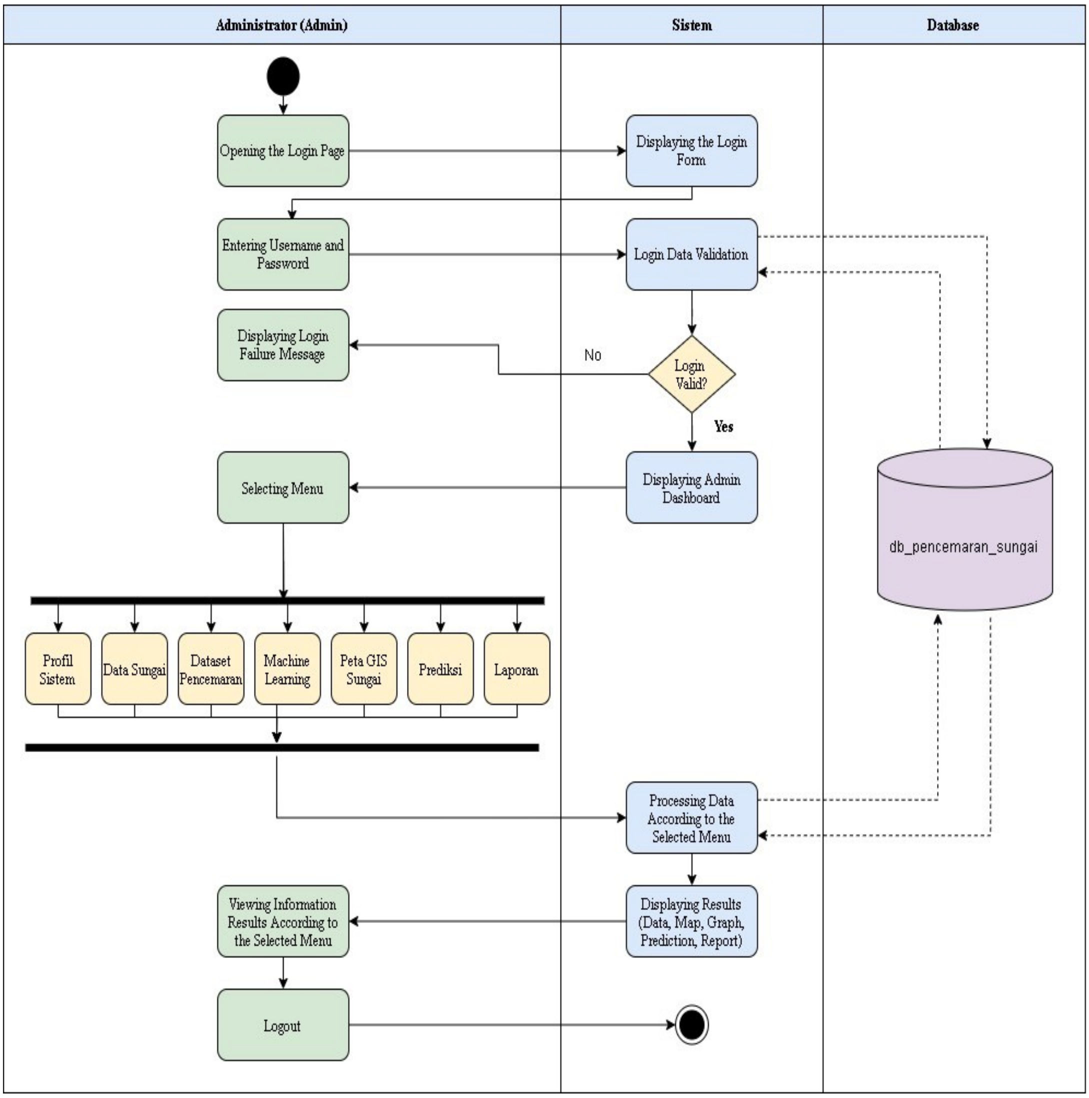


Figure 4. Administrator Activity Diagram

C. Sequence Diagram Administrator

The Administrator Sequence Diagram illustrates the interactions among the administrator, system, and database during the management of the web-based river pollution prediction application. The process begins when the administrator logs into the system by entering valid authentication credentials. The system then performs user validation by checking the submitted credentials against the data stored in the database. Upon successful validation, the administrator is granted access to the admin dashboard.

After accessing the dashboard, the administrator can manage various system modules, including river data, pollution datasets, Machine Learning models, GIS-based maps, prediction results, and reports. When a specific menu is selected, the system retrieves the relevant information from the database and displays it to the administrator.

The administrator is also authorized to perform Create, Read, Update, and Delete (CRUD) operations on system data. Whenever data are added, modified, or deleted, the system processes the request, updates the database accordingly, and returns a confirmation message indicating that the operation has been completed successfully.

This sequence diagram demonstrates the flow of information between the administrator, the application system, and the database, highlighting how data management and prediction functionalities are executed within the river pollution monitoring and prediction system.

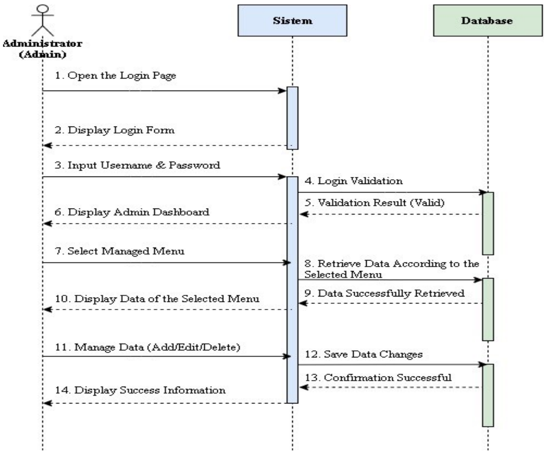


Figure 5. Administrator Sequence Diagram for Managing System Menus

D. Class Diagram

The Class Diagram is used to illustrate the structure of classes, attributes, methods, and the relationships among classes within the system.

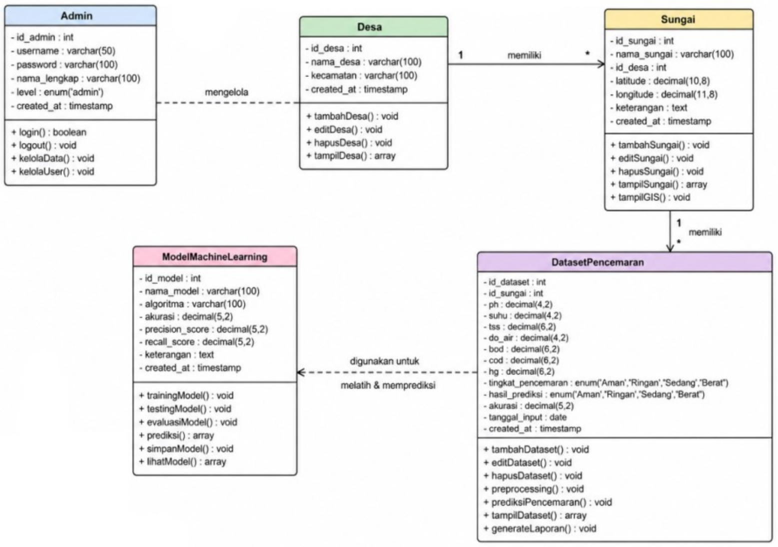


Figure 6. Class Diagram

3.6 System Implementation

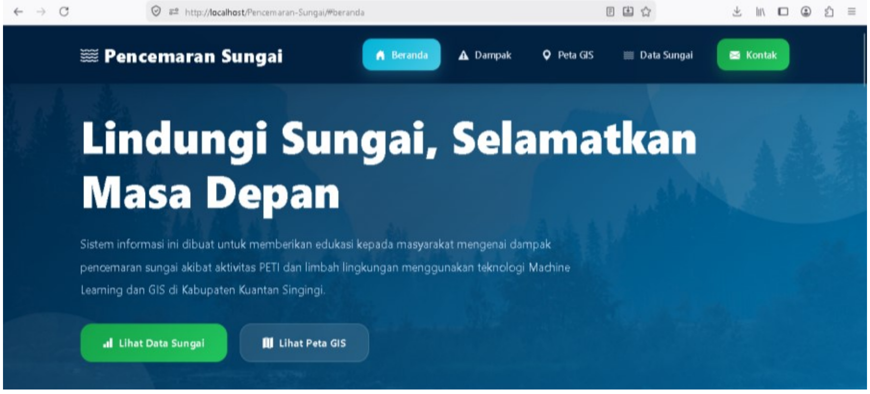
The system was implemented by developing a web-based application to predict river pollution levels caused by Artisanal and Small-Scale Gold Mining (ASGM) activities in Kuantan Singingi Regency using PHP, MySQL, JavaScript, and GIS technologies. The system consists of two types of users: public users, who can access pollution-related information without logging in, and administrators, who have access privileges to manage data, perform predictions, and generate reports.

1. Public User System Implementation

The public-user interface was designed to enable users to obtain information regarding river pollution conditions in an easy, fast, and informative manner through a web browser. The pages available to public users include:

a. Home Page

The Home Page serves as the main page of the application and provides general information about the river pollution prediction system. On this page, users can view a brief description of the system as well as access the main navigation menu.



Lindungi Sungai, Selamatkan Masa Depan

Sistem informasi ini dibuat untuk memberikan edukasi kepada masyarakat mengenai dampak pencemaran sungai akibat aktivitas PETI dan limbah lingkungan menggunakan teknologi Machine Learning dan GIS di Kabupaten Kuantan Singingi.

Figure 7. Home Page

b. Pollution Impact Page

The Pollution Impact Page is used to provide education to the public regarding the impacts of Artisanal and Small-Scale Gold Mining (ASGM) activities on river environmental quality and human health, particularly those caused by mercury (Hg) pollution. This page presents comprehensive information about the sources of contamination, environmental consequences, potential health risks, and preventive measures. It also aims to raise public awareness and encourage community participation in environmental protection and sustainable mining practices.

Dampak Pencemaran Sungai

Pencemaran sungai dapat menyebabkan kerusakan lingkungan, gangguan kesehatan, dan penurunan kualitas hidup masyarakat akibat limbah PETI dan pencemaran merkuri.

Gangguan Kesehatan

Air tercemar dapat menyebabkan penyakit kulit, diare, keracunan merkuri, dan gangguan saraf pada masyarakat sekitar sungai.

Kerusakan Lingkungan

Ekosistem sungai rusak, ikan mati, kualitas air menurun, dan habitat alami menjadi terganggu.

Dampak Sosial

Aktivitas masyarakat terganggu, sumber air bersih berkurang, dan ekonomi warga sekitar ikut terdampak.

Figure 8. Pollution Impact Page

c. River GIS Map Page

The River GIS Map Page is used to display the locations of rivers affected by pollution in the form of a digital map. The system utilizes the Leaflet.js library to visualize river coordinate points and represent pollution levels using different color indicators.

Peta GIS Pencemaran Sungai

Peta interaktif ini menampilkan lokasi sungai dan tingkat pencemaran berdasarkan data monitoring kualitas air.

Sungai Singingi

Desa: Muara Lembu

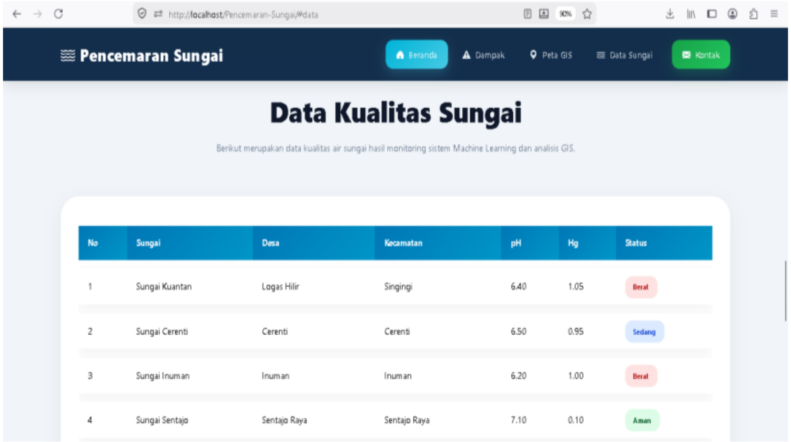
Status: Berat

Monitoring kualitas air sungai berbasis GIS dan Machine Learning.

Figure 9. River GIS Map Page

d. **River Data Page**

The River Data Page is used to display information about rivers that are the objects of the study, including the river name, village, district, and river pollution level based on the system's prediction results. This page serves as a central repository of river-related information, allowing users to easily access and monitor the environmental condition of each river. The displayed data supports environmental assessment, facilitates information management, and assists stakeholders in identifying areas that require further monitoring or remediation efforts.



Data Kualitas Sungai

Berikut merupakan data kualitas air sungai hasil monitoring sistem Machine Learning dan analisis GIS.

No	Sungai	Desa	Kecamatan	pH	Hg	Status
1	Sungai Kuantan	Logas Hilir	Singingi	6.40	1.05	Berat
2	Sungai Cerenti	Cerenti	Cerenti	6.50	0.95	Sedang
3	Sungai Inuman	Inuman	Inuman	6.20	1.00	Berat
4	Sungai Sentajo	Sentajo Raya	Sentajo Raya	7.10	0.10	Aman

Figure 10. River Data Page

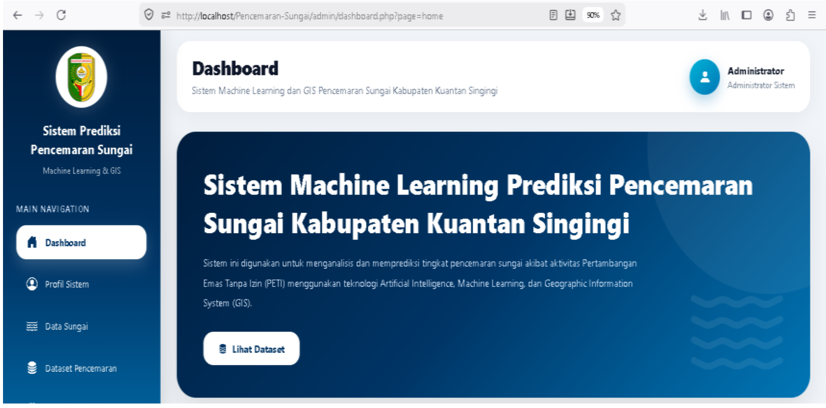
2. **Administrator System Implementation**

The administrator-side implementation is used to manage all processes within the river pollution level prediction system. The administrator must log in before accessing the main admin interface to ensure system security and authorized access. Through this interface, administrators can manage river data, monitor prediction results, update pollution-related information, and oversee user activities. The admin panel also facilitates data maintenance, report generation, and system configuration, enabling efficient management and ensuring the accuracy and reliability of the information presented within the monitoring system.

Several main pages available on the administrator side include:

a. **Dashboard Page**

The Dashboard Page is used to display a summary of system information, such as the number of river data entries, pollution datasets, prediction results, and river pollution graphical visualizations. This page serves as the main interface for administrators, providing a comprehensive overview of the system's current status and environmental monitoring data. Through interactive charts and summary statistics, users can quickly identify pollution trends, monitor prediction outcomes, and evaluate river conditions. The dashboard supports efficient decision-making by presenting key information in a clear, organized, and easily accessible format.



Sistem Machine Learning dan GIS Pencemaran Sungai Kabupaten Kuantan Singingi

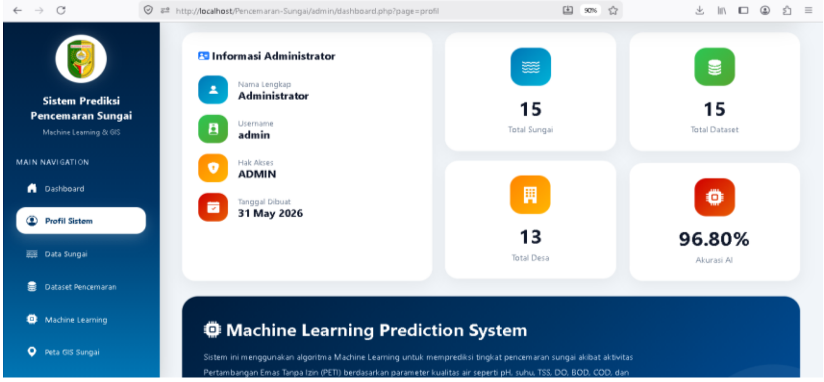
Sistem Machine Learning Prediksi Pencemaran Sungai Kabupaten Kuantan Singingi

Sistem ini digunakan untuk menganalisis dan memprediksi tingkat pencemaran sungai akibat aktivitas Pertambangan Emas Tanpa Izin (PETI) menggunakan teknologi Artificial Intelligence, Machine Learning, dan Geographic Information System (GIS).

Figure 11. Dashboard Page

b. System Profile Page

The System Profile Page is used to display information about the system objectives, the technologies used, and a brief description of the river pollution prediction application.



Nama Lengkap: Administrator
Username: admin
Hak Akses: ADMIN
Tanggal Dibuat: 31 May 2026

15	15	13	96.80%
Total Sungai	Total Dataset	Total Desa	Akurasi AI

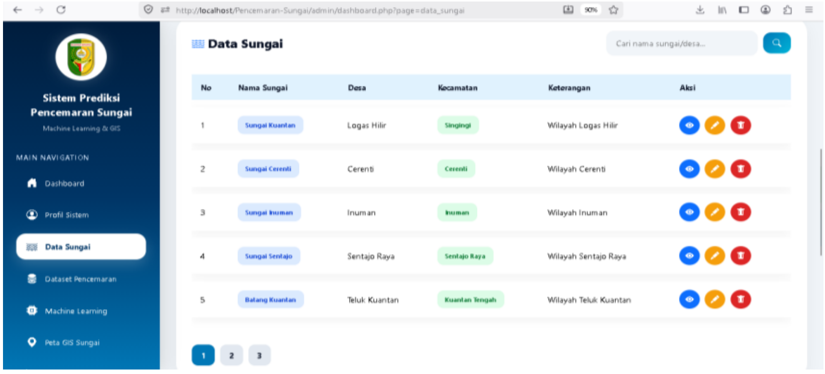
Machine Learning Prediction System

Sistem ini menggunakan algoritma Machine Learning untuk memprediksi tingkat pencemaran sungai akibat aktivitas Pertambangan Emas Tanpa Izin (PETI) berdasarkan parameter kualitas air seperti pH, suhu, TSS, DO, BOD, COD, dan

Figure 12. System Profile Page

c. River Data Page

The River Data Page is used to manage data related to river locations that serve as the research objects. The managed data includes river name, village, geographic coordinates, and river descriptions.



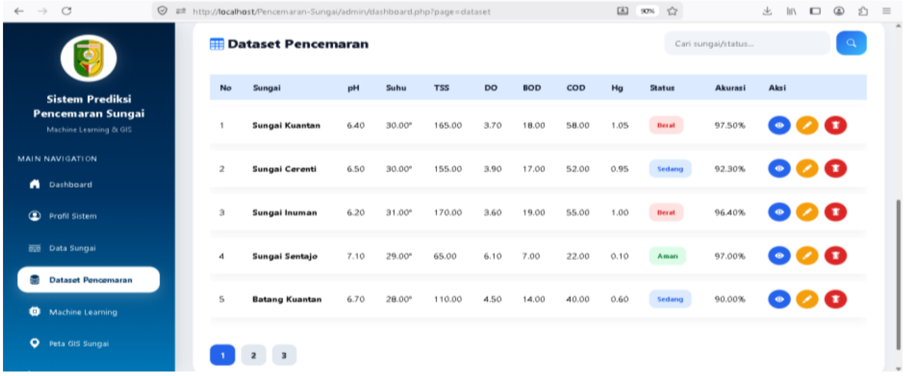
No	Nama Sungai	Desa	Kecamatan	Keterangan	Aksi
1	Sungai Kuantan	Logas Hilir	Singingi	Wilayah Logas Hilir	
2	Sungai Cerenti	Cerenti	Cerenti	Wilayah Cerenti	
3	Sungai Inuman	Inuman	Inuman	Wilayah Inuman	
4	Sungai Sentajo	Sentajo Raya	Sentajo Raya	Wilayah Sentajo Raya	
5	Batang Kuantan	Teluk Kuantan	Kuantan Tengah	Wilayah Teluk Kuantan	

Figure 13. River Data Page

d. Pollution Dataset Page

The Pollution Dataset Page is used to manage water quality parameter data, which includes: pH, water temperature, Total Suspended Solids (TSS), Dissolved Oxygen (DO), Biological Oxygen Demand

(BOD), Chemical Oxygen Demand (COD), and mercury (Hg) concentration. These data are used as input for the Machine Learning process to predict river pollution levels.



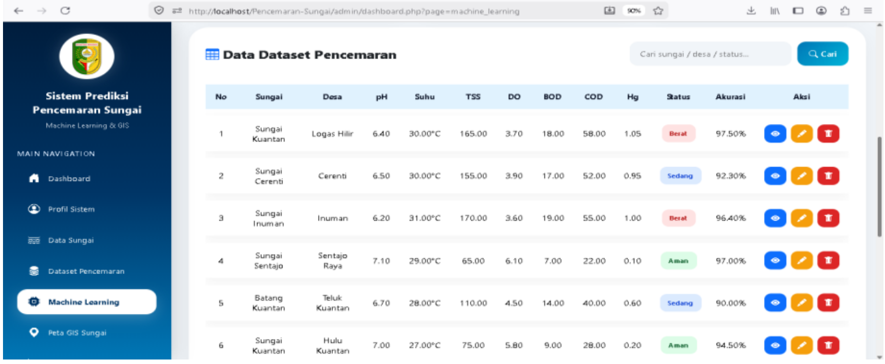
Dataset Pencemaran

No	Sungai	pH	Suhu	TSS	DO	BOD	COD	Hg	Status	Akurasi	Aksi
1	Sungai Kuantan	6.40	30.00°	165.00	3.70	18.00	58.00	1.05	Berat	97.50%	
2	Sungai Cerenti	6.50	30.00°	155.00	3.90	17.00	52.00	0.95	Sedang	92.30%	
3	Sungai Inuman	6.20	31.00°	170.00	3.60	19.00	55.00	1.00	Berat	96.40%	
4	Sungai Sentajo	7.10	29.00°	65.00	6.10	7.00	22.00	0.10	Aman	97.00%	
5	Batang Kuantan	6.70	28.00°	110.00	4.50	14.00	40.00	0.60	Sedang	90.00%	

Figure 14. Pollution Dataset Page

e. Machine Learning Page

The Machine Learning Page is used to perform the training and testing processes of the prediction model. The algorithms used in this study include Random Forest, Support Vector Machine (SVM), and Artificial Neural Network (ANN).



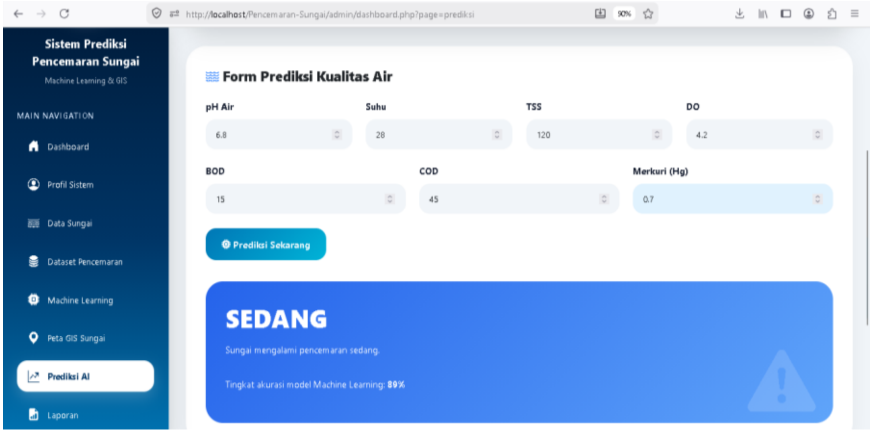
Data Dataset Pencemaran

No	Sungai	Desa	pH	Suhu	TSS	DO	BOD	COD	Hg	Status	Akurasi	Aksi
1	Sungai Kuantan	Logas Hilir	6.40	30.00°C	165.00	3.70	18.00	58.00	1.05	Berat	97.50%	
2	Sungai Cerenti	Cerenti	6.50	30.00°C	155.00	3.90	17.00	52.00	0.95	Sedang	92.30%	
3	Sungai Inuman	Inuman	6.20	31.00°C	170.00	3.60	19.00	55.00	1.00	Berat	96.40%	
4	Sungai Sentajo	Sentajo Raya	7.10	29.00°C	65.00	6.10	7.00	22.00	0.10	Aman	97.00%	
5	Batang Kuantan	Teluk Kuantan	6.70	28.00°C	110.00	4.50	14.00	40.00	0.60	Sedang	90.00%	
6	Sungai Kuantan	Hulu Kuantan	7.00	27.00°C	75.00	5.80	9.00	28.00	0.20	Aman	94.50%	

Figure 15. Machine Learning Page

f. Prediction Page

The Prediction Page is used to perform simulations of river pollution level predictions based on water quality parameters entered by the administrator. The prediction results are displayed in categorical form, such as Safe, Low, Moderate, and Severe pollution levels.



Form Prediksi Kualitas Air

pH Air: 6.8
Suhu: 28
TSS: 120
DO: 4.2
BOD: 15
COD: 45
Merkuri (Hg): 0.7

SEDANG

Sungai mengalami pencemaran sedang.

Tingkat akurasi model Machine Learning: **89%**

Figure 16. Prediction Page

g. River GIS Map Page

The GIS Page is used to display the distribution of river locations in the form of a digital map, making it easier to analyze river pollution locations.

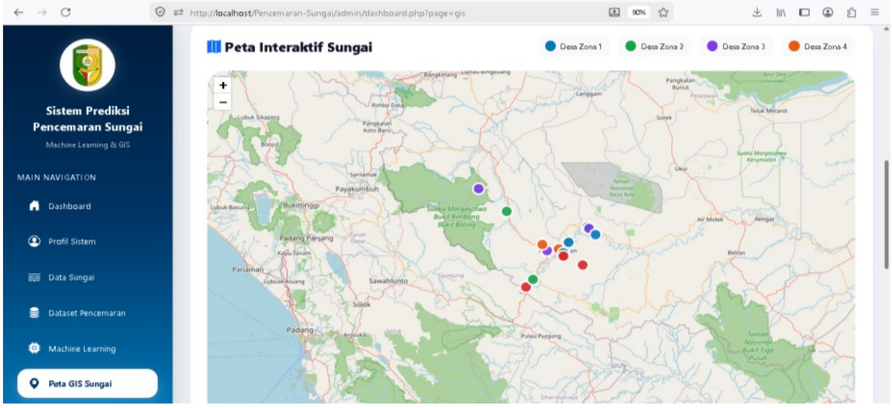
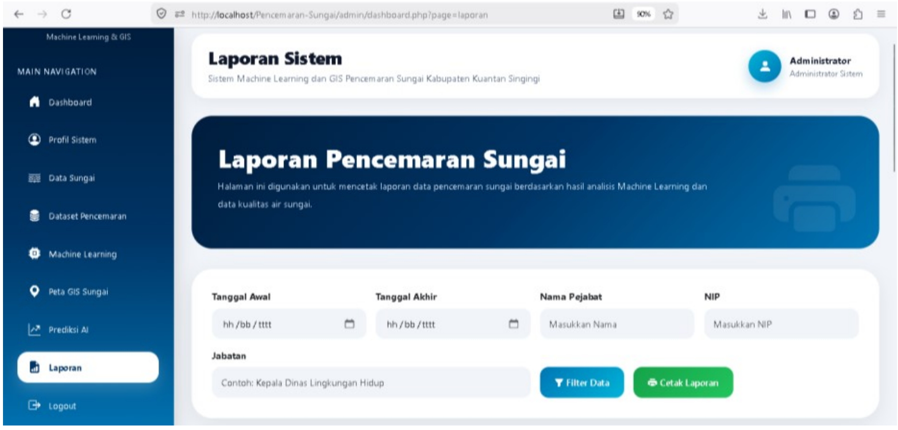


Figure 17. GIS Page

h. Report Page

The Report Page is used to generate reports on prediction results and river pollution data, which can be printed or saved in a digital format.



Laporan Sistem

Sistem Machine Learning dan GIS Pencemaran Sungai Kabupaten Kuantan Singingi

Laporan Pencemaran Sungai

Halaman ini digunakan untuk mencetak laporan data pencemaran sungai berdasarkan hasil analisis Machine Learning dan data kualitas air sungai.

Tanggal Awal: hh/bb/tttt

Tanggal Akhir: hh/bb/tttt

Nama Pejabat: Masukkan Nama

NIP: Masukkan NIP

Jabatan: Contoh: Kepala Dinas Lingkungan Hidup

Gambar 18. Halaman Laporan

3.7 System Testing

System testing was conducted to determine whether the web-based river pollution level prediction system operates properly in accordance with its designed functions. Testing was performed on all main pages and system features, both on the public user side and the administrator side.

The testing method used in this study was Black Box Testing. This method is applied to evaluate system functionality based on input and output without examining the internal program code. The testing process was carried out by executing each system feature and ensuring that the results obtained were consistent with the system requirements.

A. Public User Page Testing

Testing on the public user pages was conducted to ensure that all information can be accessed properly by general users without requiring a login process.

Table 6. Public User Page Testing

No.	Tested Feature	Testing Process	Expected Result	Status
1	Home Page	Open the main system page	Home page is displayed properly	Successful
2	Impact Page	Select the pollution impact menu	Pollution impact information is displayed	Successful
3	River Data Page	Open river data menu	River data is displayed according to the database	Successful
4	GIS Page	Open river GIS map	Map and river location points are displayed	Successful
5	Contact Page	Open contact menu	Contact information is displayed	Successful

Based on the testing results, all pages on the public user side function properly and display information in accordance with the designed system requirements.

B. Administrator Page Testing

Testing on the administrator pages was conducted to ensure that the data management and system prediction processes function properly and correctly.

Table 7. Administrator Page Testing

No.	Tested Feature	Testing Process	Expected Result	Status
1	Admin Login	Enter username and password	System displays admin dashboard	Successful
2	River Data	Add river data	Data is successfully saved	Successful
3	Pollution Dataset	Add water quality dataset	Data is successfully stored in database	Successful
4	Machine Learning	Run prediction process	System displays prediction results	Successful
5	Prediction	Input water quality parameters	System generates pollution category	Successful
6	GIS Map	Open GIS page	River location map is displayed	Successful
7	Report	Print report	Report is successfully generated	Successful
8	Logout	Exit system	System returns to login page	Successful

The test results indicate that all administrator features function properly in accordance with the designed system requirements.

C. Machine Learning Prediction Testing

The Machine Learning model testing was conducted to evaluate the system’s ability to predict river pollution levels based on water quality parameters.

The parameters used in this study include:

a) pH
b) Temperature
c) TSS
d) DO
e) BOD
f) COD
g) Hg

The algorithms applied in this research are:

a) Random Forest
b) Support Vector Machine (SVM)
c) Artificial Neural Network (ANN)

The testing results indicate that the Random Forest algorithm outperforms the other algorithms based on accuracy, precision, and recall values.

Table 8. Machine Learning Model Testing Results

Algorithm	Accuracy	Precision	Recall
Random Forest	96.80%	95.40%	94.90%
SVM	91.20%	90.10%	89.50%
ANN	93.50%	92.40%	91.80%

Based on the testing results, the Random Forest algorithm was selected as the best model in the river pollution level prediction system because it achieved the highest accuracy and demonstrated strong performance in classifying pollution levels effectively.

D. System Testing Results

Based on all testing stages conducted, the web-based river pollution level prediction system operates properly in accordance with user requirements. The system is capable of managing data, visualizing GIS information, and predicting river pollution levels using Machine Learning methods in an effective and integrated manner.

7. Conclusion

Based on the results of the research conducted, it can be concluded that the development of a Machine Learning model for predicting river pollution levels caused by Artisanal and Small-Scale Gold Mining (ASGM) activities in Kuantan Singingi Regency has been successfully implemented in a web-based system. This study effectively utilized water quality parameters such as pH, temperature, TSS, DO, BOD, COD, and mercury (Hg) concentration as the main variables in the river pollution prediction process. The research stages, starting from data collection, preprocessing, model development, to system implementation, were carried out systematically and in an integrated manner.

The testing results show that Machine Learning methods can be effectively used to classify river pollution levels with good accuracy. Based on model evaluation, the Random Forest algorithm achieved the best performance compared to Support Vector Machine (SVM) and Artificial Neural Network (ANN), with an accuracy of 96.80%. This indicates that the Machine Learning approach is capable of effectively analyzing the complex and non-linear relationships between water quality parameters.

In addition, the implementation of the web-based system, equipped with graphical visualization and a Geographic Information System (GIS), enhances the presentation of river pollution information in a more interactive and easily understandable manner. The system not only functions as a data storage platform but also serves as a decision support system, assisting local governments and the public in monitoring and predicting river pollution conditions in a faster and data-driven way.

Based on the research findings, the developed system has met the objectives stated in the introduction, namely to build a Machine Learning-based model for predicting river pollution levels due to ASGM activities in Kuantan Singingi Regency. Therefore, this study contributes to the application of Machine Learning technology in the environmental sector, particularly in river water quality management.

For future research, the system can be improved by using a larger dataset and real-time data collected from Internet of Things (IoT) sensors, so that prediction results become more accurate and responsive to environmental changes. Furthermore, future studies may explore deep learning or hybrid machine learning methods to improve prediction performance, as well as develop mobile system integration to make river pollution information more accessible to the public and related institutions.

Acknowledgement

The author would like to express sincere gratitude to Universitas Islam Kuantan Singingi for providing support in the implementation of this research through the internal research grant program. Appreciation is also extended to the Institute for Research, Community Service, and Islamic Da'wah (LPPMD) of Universitas Islam Kuantan Singingi for their support, facilitation, and assistance throughout the research process. In addition, the author would like to thank all parties who contributed to data collection, system development, and the completion of this research, so that the study could be carried out successfully.

References

[1] S. B. Awadh and M. M. Al-Sabbagh, "Machine Learning Approaches for Water Quality Prediction: A Review," Environmental Monitoring and Assessment, vol. 195, no. 2, pp. 1–15, 2023.

[2] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[3] C. Cortes and V. Vapnik, "Support-Vector Networks," Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.

[4] S. Haykin, Neural Networks and Learning Machines, 3rd ed. New York, NY, USA: Prentice Hall, 2009.

[5] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. San Francisco, CA, USA: Morgan Kaufmann, 2012.

[6] I. H. Witten, E. Frank, and M. A. Hall, Data Mining: Practical Machine Learning Tools and Techniques, 4th ed. Burlington, MA, USA: Morgan Kaufmann, 2016.

[7] Y. Chen, Z. Song, X. Wang, and H. Yang, "Water Quality Prediction Based on Machine Learning Methods," Journal of Hydrology, vol. 589, pp. 1–12, 2020.

[8] A. Mosavi, P. Ozturk, and K. Chau, "Flood Prediction Using Machine Learning Models: Literature Review," Water, vol. 10, no. 11, pp. 1–40, 2018.

[9] S. R. Qasem, A. A. Al-Fuqaha, and M. A. Alzoubi, "Prediction of Water Pollution Using Artificial Intelligence Techniques," Environmental Science and Pollution Research, vol. 29, no. 14, pp. 20345–20358, 2022.

[10] N. K. Ahmed, A. F. Atiya, N. El Gayar, and H. El-Shishiny, "An Empirical Comparison of Machine Learning Models for Time Series Forecasting," Econometric Reviews, vol. 29, no. 5, pp. 594–621, 2010.

[11] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, 2nd ed. New York, NY, USA: Springer, 2009.

[12] B. Liu, Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, 2nd ed. Berlin, Germany: Springer, 2011.

[13] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.

[14] M. F. Goodchild, "Geographic Information Systems and Science," International Journal of Geographical Information Science, vol. 26, no. 4, pp. 589–590, 2012.

[15] P. A. Longley, M. F. Goodchild, D. J. Maguire, and D. W. Rhind, Geographical Information Systems and Science, 4th ed. Hoboken, NJ, USA: Wiley, 2015.

[16] A. K. Bhadra, S. Chattopadhyay, and S. S. Das, "Prediction of River Water Quality Using Artificial Neural Network," Applied Water Science, vol. 11, no. 4, pp. 1–12, 2021.

[17] J. Brownlee, Machine Learning Mastery with Python. Melbourne, Australia: Machine Learning Mastery, 2017.
[18] H. Patel and R. Patel, “A Review on Machine Learning Approaches for Water Quality Analysis,” Procedia Computer Science, vol. 167, pp. 1301–1310, 2020.
[19] A. Singh, P. Sharma, and R. Kumar, “Water Quality Monitoring System Using IoT and Machine Learning,” in 2021 International Conference on Smart Electronics and Communication, pp. 554–560, 2021.
[20] M. K. Jha, P. Samal, and R. Das, “Prediction of Water Quality Index Using Machine Learning Algorithms,” International Journal of Environmental Research and Public Health, vol. 20, no. 3, pp. 1–18, 2023.
[21] S. R. Kotsiantis, “Supervised Machine Learning: A Review of Classification Techniques,” Informatica, vol. 31, no. 3, pp. 249–268, 2007.
[22] A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 3rd ed. Sebastopol, CA, USA: O’Reilly Media, 2022.
[23] D. P. Solomatine and A. Ostfeld, “Data-Driven Modelling: Some Past Experiences and New Approaches,” Journal of Hydroinformatics, vol. 10, no. 1, pp. 3–22, 2008.
[24] F. Pedregosa, G. Varoquaux, A. Gramfort, et al., “Scikit-learn: Machine Learning in Python,” Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
[25] X. Wu, V. Kumar, J. R. Quinlan, et al., “Top 10 Algorithms in Data Mining,” Knowledge and Information Systems, vol. 14, no. 1, pp. 1–37, 2008.
[26] A. H. Tan, “Data Mining and Knowledge Discovery Technologies,” in Proceedings of the PAKDD Conference, pp. 1–14, 1999.
[27] M. Abadi, A. Agarwal, P. Barham, et al., “TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems,” in Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation, pp. 265–283, 2016.
[28] S. Madakam, R. Ramaswamy, and S. Tripathi, “Internet of Things (IoT): A Literature Review,” Journal of Computer and Communications, vol. 3, no. 5, pp. 164–173, 2015.
[29] H. Gupta and S. Gupta, “GIS-Based Environmental Monitoring System Using Spatial Analysis Techniques,” International Journal of Environmental Science and Technology, vol. 18, no. 7, pp. 1551–1562, 2021.
[30] K. P. Murphy, Machine Learning: A Probabilistic Perspective. Cambridge, MA, USA: MIT Press, 2012.
[31] F. Haswan, N. W. Al-Hafiz, and G. Hamza, “Decision Support System Using the SAW Method for Determining Research Priorities at Universitas Islam Kuantan Singingi,” INOVTEK Polbeng - Seri Informatika, vol. 10, no. 3, pp. 1602–1612, 2025.

[32] F. Haswan, Erlinda, and W. Walhidayat, “Implementasi Sistem Pendukung Keputusan untuk Menentukan Calon Reviewer Internal Universitas Islam Kuantan Singingi,” ZONAsi: Jurnal Sistem Informasi, vol. 6, no. 2, pp. 499–509, 2024.
[33] D. Setiawan, F. Haswan, and Jasri, “Design of School Bell Scheduling Application Based on Arduino Uno on MTs Babussalam Simandolak,” Jurnal Teknologi dan Open Source, vol. 7, no. 1, pp. 22–30, 2024.
[34] W. Walhidayat, J. Andanu, S. Djusar, and F. Haswan, “Sistem Informasi Berbasis Web untuk Pendataan dan Pengelolaan Data Penduduk dengan Metode Object Oriented Analysis and Design,” ZONAsi: Jurnal Sistem Informasi, vol. 7, no. 2, pp. 715–724, 2025.
[35] F. Haswan, “Application of Simple Additive Weighting Method to Determine Outstanding School Principals,” Sinkron, vol. 3, no. 2, pp. 186–195, 2019.