Title of ML project: *eg: River Water Quality Prediction Model*

—---

Name: [Your Name]

Organization:

—---
----

1. **Overview of Problem Statement**:

Eg: The quality of river water is crucial for various purposes like drinking, agriculture, and maintaining ecosystem balance. However, predicting river water quality accurately is challenging due to the complex interplay of various environmental factors. Therefore, developing an effective prediction model can aid in proactive management and conservation efforts.

2. **Objective**:

Eg: To develop the best river water quality prediction model using machine learning techniques.

3. **Data Description**:
   - Source: [Specify the source of data]
   - Features: [List the features available in the dataset]

4. **Data Collection**:

Import the dataset from the specified source.

- Gain insights into the data distribution, relationships, and potential patterns.

5. Data Preprocessing - Data Cleaning:

- Handle missing values using appropriate imputation techniques.

- Check for and remove outliers using statistical methods.

- Address skewed data in numerical features through transformations.

6. Exploratory Data Analysis (EDA):

- Gain insights into the data distribution, relationships, and potential patterns.

- Visualizations: Histogram, Boxplot, Pair Plot, Heatmap Correlation, Pie Diagram, Bar Plot, Count Plot, Line Plot, Kernel Density Estimation (KDE).

7. Feature Engineering:

- Identify and encode categorical features using techniques like one-hot encoding or label encoding.

8. Feature Selection:

- Use algorithms like Random Forest and Select K Best to identify relevant features.

- Remove redundant or irrelevant features.

9. Split Data into Training and Testing Sets:

- Divide the dataset into training and testing subsets.

10. Feature Scaling:

- Scale numerical features if necessary to ensure uniform magnitude using techniques like Min-Max scaling or Standardization.

11. Build the ML Model: (at least 5)

   - Implement various classification and regression algorithms:

   - Classification: Logistic Regression, SVM, Decision Tree, MLP Classifier, Naive Bayes, Random Forest, K-Nearest Neighbors, Gradient Boost, Adaboost.

   - Regression: SVR, MLP Regressor, Random Forest Regressor, Linear Regression, Gradient Boost, Adaboost.

12. Model Evaluation:

   - Regression Metrics: MAE, MSE, RMSE, R2 Score.

   - Classification Metrics: Confusion Matrix, Accuracy, Precision, Recall, F1-Score, ROC Curve.

13. Hyperparameter Tuning:

   - Optimize model performance by tuning hyperparameters.

14. Save the Model:

   - Save the trained model for future use.

15. Test with Unseen Data:

   - Assess the model's performance on unseen data.

16. Interpretation of Results (Conclusion):

   - Analyze the model's performance and draw conclusions. Discuss any limitations of the dataset.

17. Future Work:

- Explore deep learning algorithms for potentially higher accuracy.

- Update the model periodically with new data.

- Address imbalanced data through resampling techniques.

- Consider adding more features to enhance predictive power.