WATER QUALITY PREDICTION FOR CONCREATE MIXING

PRESENTATION

PROBLEM STATEMENT

Water quality is a crucial factor in the concrete mixing process, as it directly affects the strength, durability, and overall quality of the final concrete product. The presence of contaminants such as high chloride, organic carbon, solids, sulfate, and turbidity can weaken the concrete structure, leading to costly repairs and replacements. Traditional water quality testing methods may be time-consuming and prone to human error, making them unreliable for real-time decision-making.
The objective of this project is to develop a machine learning-based water quality prediction model that can classify water as suitable or unsuitable for concrete mixing based on key water characteristics.

Good Water (Label = 1)

Chloride: 500 - 6984 mg/L (Mean: 1745)

Organic Carbon: 50 - 598 mg/L (Mean: 164)

Solids: 502 - 11991 mg/L (Mean: 2088)

Sulphate: 20 - 799 mg/L (Mean: 268)

Turbidity: 11 - 5971 NTU (Mean: 1438)

pH: 2.00 - 11.99 (Mean: 4.75)

Bad Water (Label = 0)

Chloride: 500 - 6999 mg/L (Mean: 3994)

Organic Carbon: 50 - 599 mg/L (Mean: 356)

Solids: 502 - 11999 mg/L (Mean: 6110)

Sulphate: 21 - 799 mg/L (Mean: 541)

Turbidity: 16 - 5999 NTU (Mean: 3538)

pH: 2.00 - 11.99 (Mean: 8.19)

OBJECTIVE

- Develop a machine learning model to classify water quality.
- Use a historical dataset with key water parameters (Chloride, Label, Organic_Carbon, Solids, Sulphate, Turbidity, ph).
- Train and evaluate multiple ML models to identify the best-performing one.
- Optimize accuracy using EDA, Normalization & Hyperparameter Tuning.
- Ensure high Precision, Recall, Accuracy, Sensitivity, Specificity, and F1-score.
- Provide a real-time prediction system for construction sites.

Dataset & Features

FEATURES :

pH, Chloride, Organic Carbon, Solids, Sulphate, Turbidity.

TARGET :

1 (Good) /0 (Bad).

EDA

Exploratory Data Analysis

- Visualizations: Histograms, Boxplots, Correlation Heatmap.
- Outlier Detection: Remove/handle extreme values.
- Feature Relationships: Find correlations between parameters.

CORRELATION MATRIX

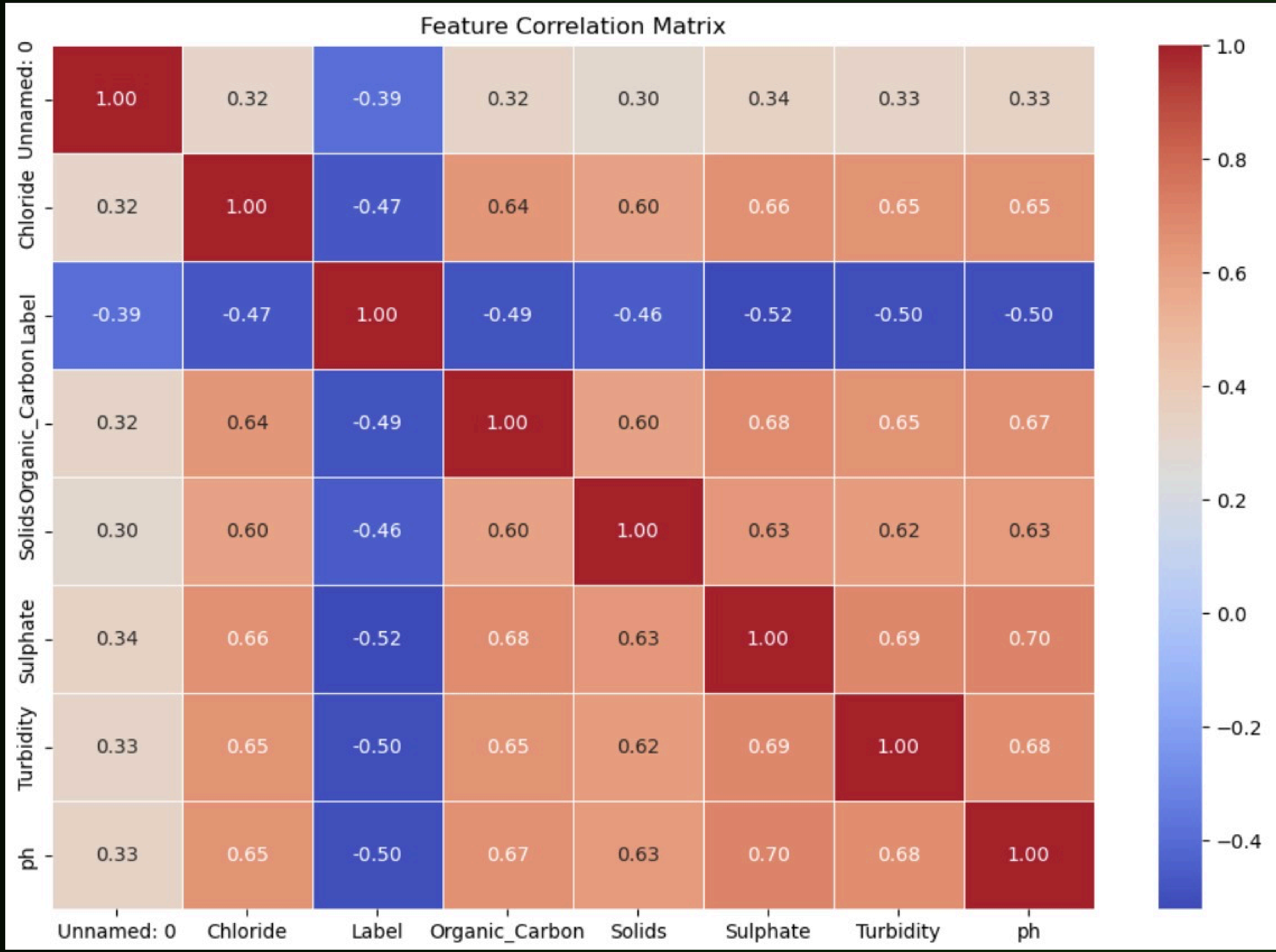
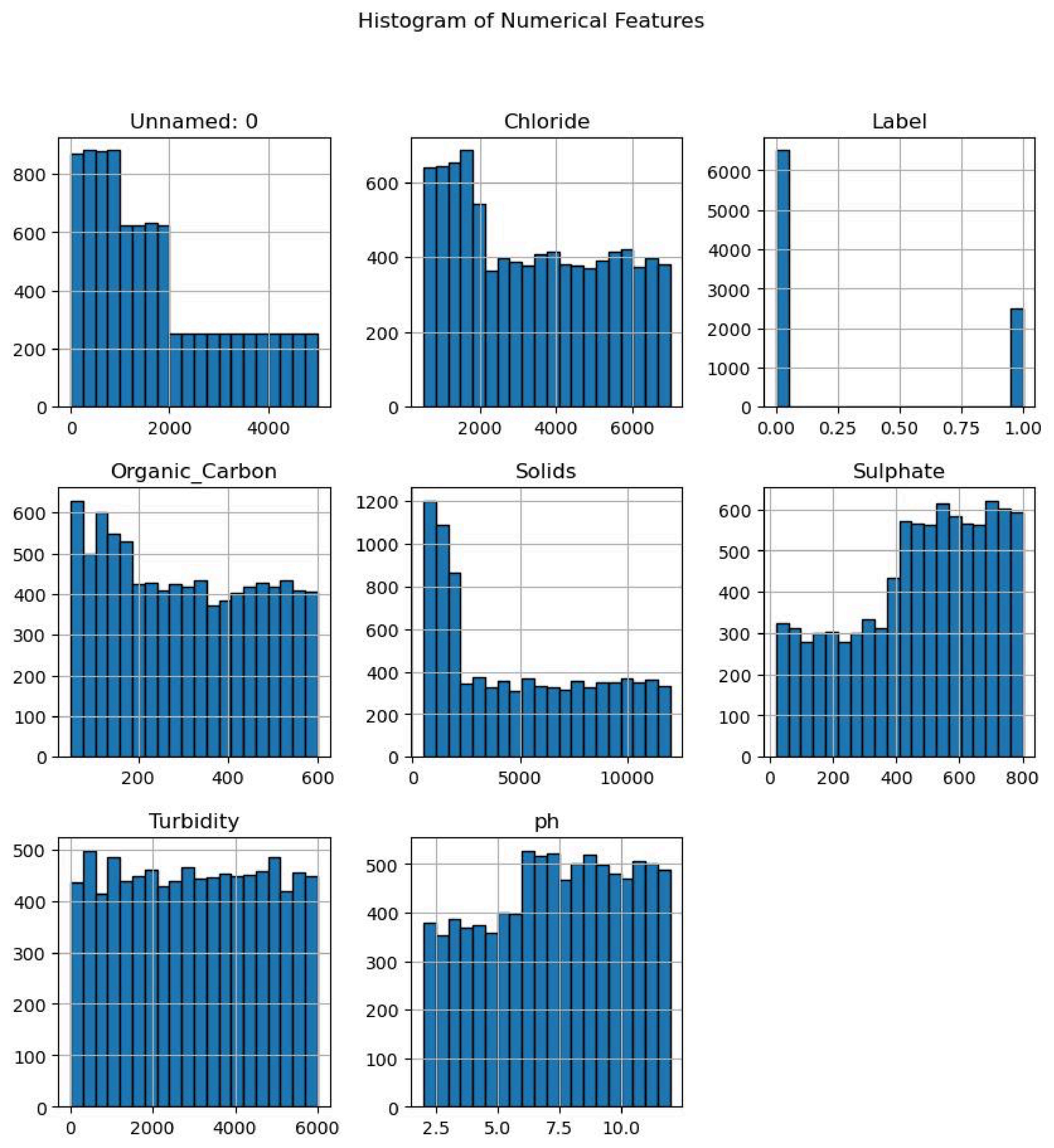


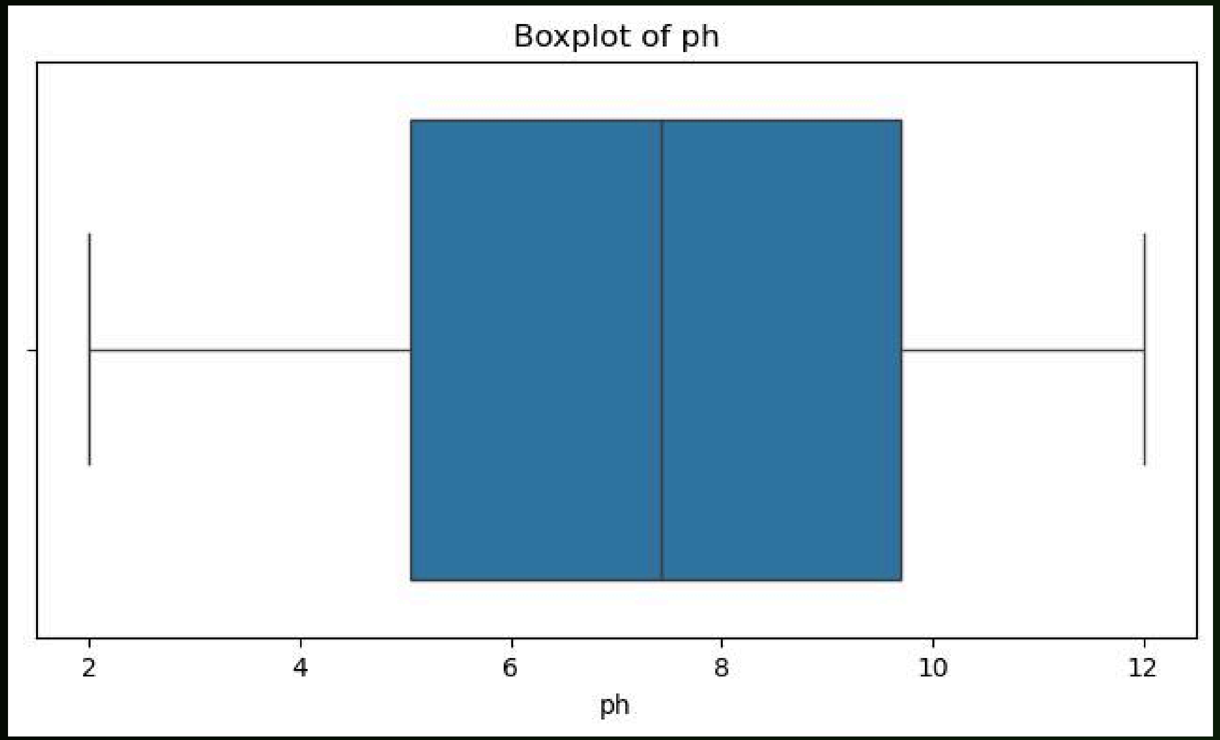
Fig 1-Feature Correlation Matrix

HISTOGRAM PLOT



OUTLIERS DETECTION

- IQR (Interquartile Range) Method is used
- IQR (Q3 - Q1).



Handle outliers using the IQR (Interquartile Range) :

- Calculate Q1 (25th percentile) and Q3 (75th percentile).
- Compute the IQR (Q3 - Q1).
- Define lower and upper bounds using the formula:
- Lower Bound=Q1-1.5×IQR
- Upper Bound=Q3+1.5×IQR Filter out values outside these bounds.]

Data Preprocessing & Normalization

- Handle missing values using imputation.
- Apply standard scaler Standardization for uniform feature scaling

mean -0

variance-1

Model Selection & Training

Algorithms used

- Random Forest
- SVM
- Logistic Regression
- K - Nearesh Neighbours
- Neural Networks.(MLP)

Multi Layer Perceptron

A Multi-Layer Perceptron (MLP) is a feedforward artificial neural network (ANN) that consists of:

1. An input layer (takes in features).
2. One or more hidden layers (processes data using neurons & activation functions).
3. An output layer (makes predictions).

MLP uses backpropagation for learning, which adjusts weights using gradient descent to minimize error.

Hyperparameter Tuning

To improve model performance, GridSearchCV was used to optimize hyperparameters:

Multi-Layer Perceptron (MLP): Tuned Parameters:

Interpretation of Best Parameters

- Activation: 'relu' → Helps in faster training and avoids vanishing gradient problems.
- Alpha: 0.0001 → Small L2 regularization, preventing overfitting while maintaining flexibility.
- Hidden Layers: → Two layers with neurons, balancing complexity and performance.
- Learning Rate: 'constant' → A fixed learning rate helped in stable convergence.
- Solver: 'sgd' → Stochastic Gradient Descent performed better than 'adam'.

CONFUSION MATRIX

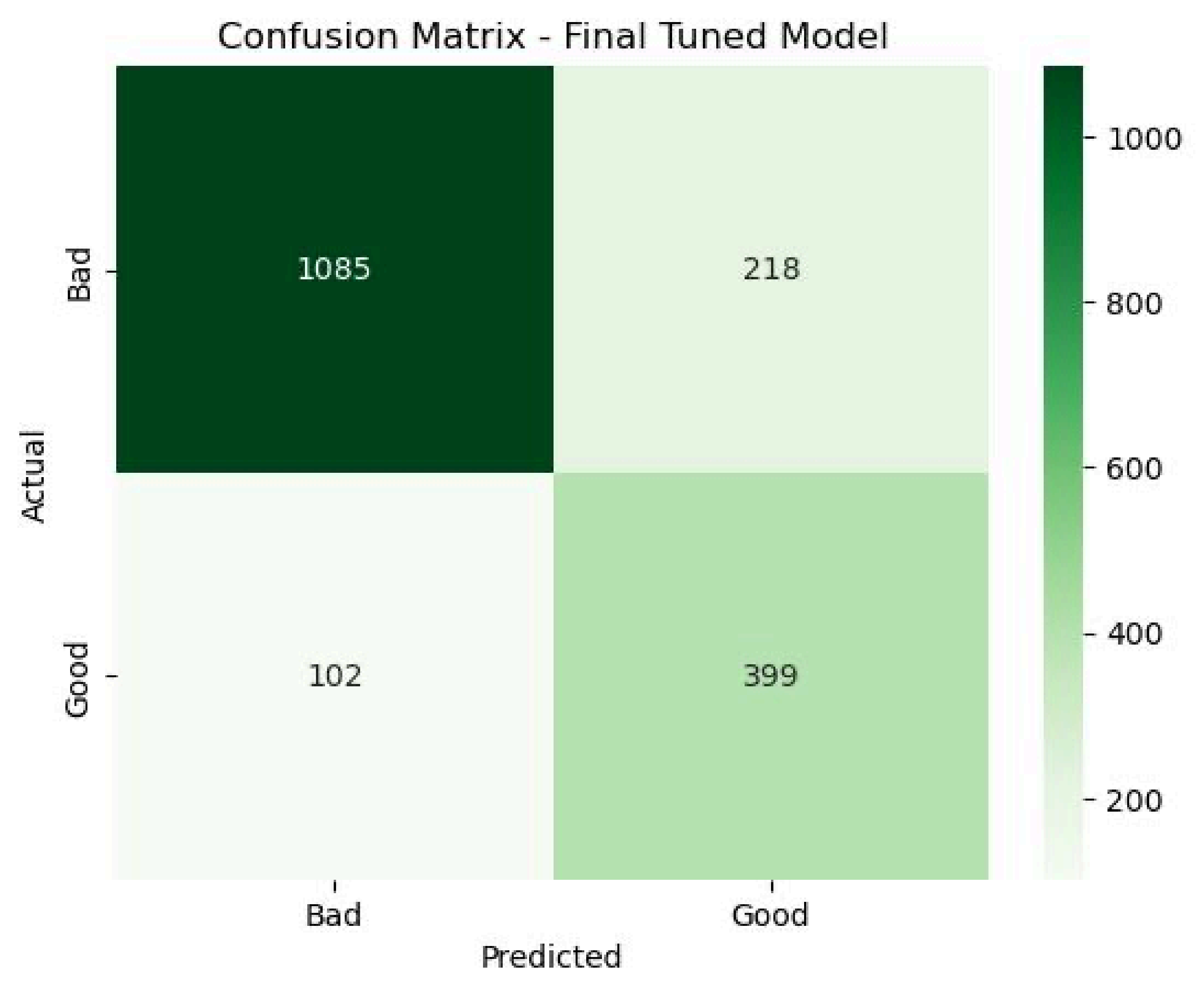


Fig 2-Confusion Matrix - Final Tuned Model

Model Evaluation & Metrics

Accuracy, Precision, Recall, F1-score

Confusion Matrix → Analyze false positives & negatives.

Judging Metrics

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

EVALUATION METRICS

- Accuracy: 0.8226
- Precision: 0.6467
- Recall (Sensitivity): 0.7964
- Specificity: 0.8327
- F1-Score: 0.7138

- Classification Report:

	precision	recall	f1-score	support
Bad	0.91	0.83	0.87	1303
Good	0.65	0.80	0.71	501
accuracy			0.82	1804
macro avg	0.78	0.81	0.79	1804
weighted avg	0.84	0.82	0.83	1804

- Confusion Matrix:

```
[[1085  218]
 [ 102  399]]
```

Fig 3-Best model Evaluation metrics

MODEL COMPARISION

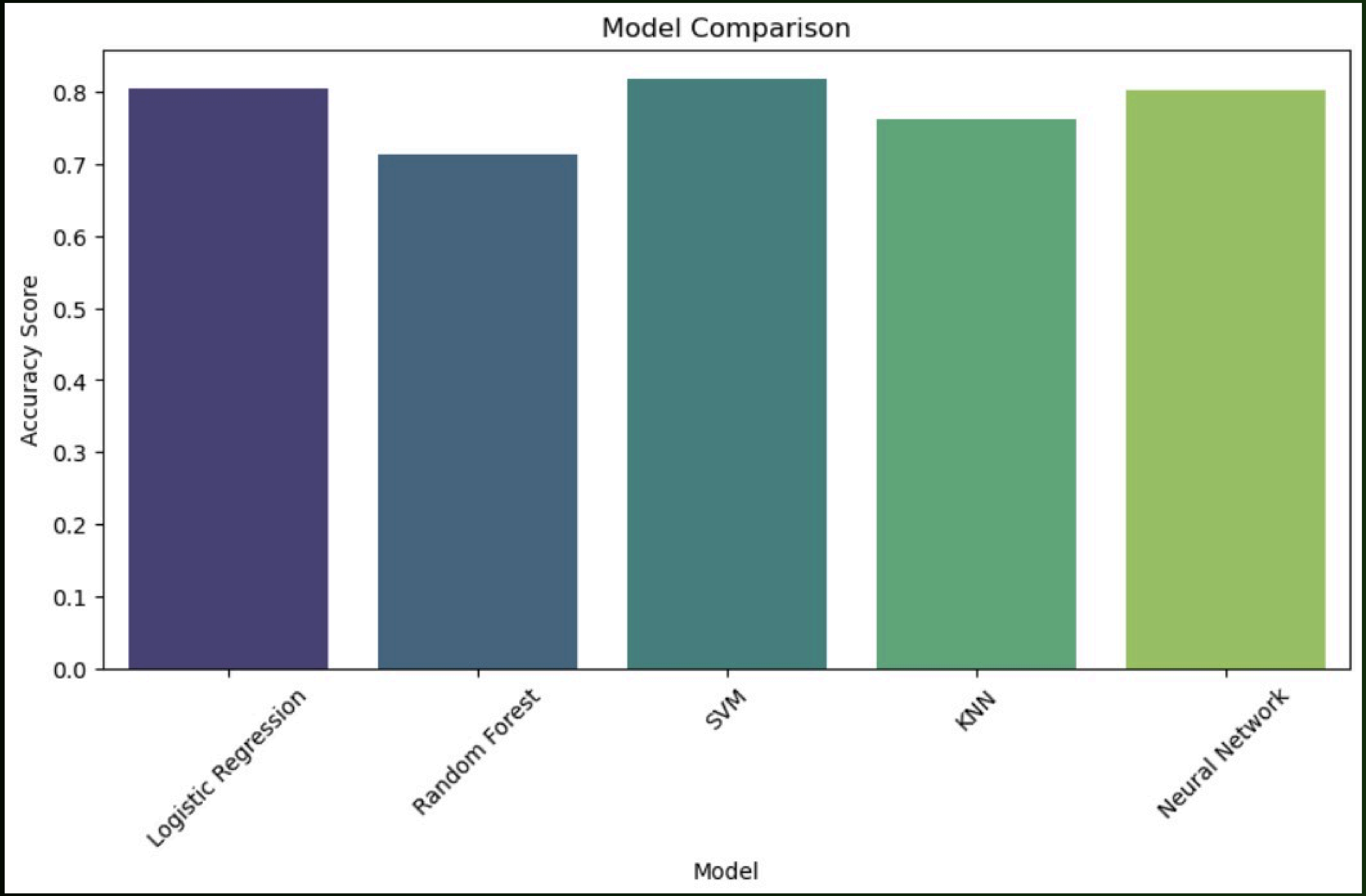


Fig 4-Model Comparison

THRESHOLD VALUES

Good Water (Label = 1) Chloride: 500 - 6984 mg/L (Mean: 1745)

Organic Carbon: 50 - 598 mg/L (Mean: 164)

Solids: 502 - 11991 mg/L (Mean: 2088)

Sulphate: 20 - 799 mg/L (Mean: 268)

Turbidity: 11 - 5971 NTU (Mean: 1438)

pH: 2.00 - 11.99 (Mean: 4.75)

GOOD WATER THRESHOLDS

Bad Water (Label = 0) Chloride: 500 - 6999 mg/L (Mean: 3994)

Organic Carbon: 50 - 599 mg/L (Mean: 356)

Solids: 502 - 11999 mg/L (Mean: 6110)

Sulphate: 21 - 799 mg/L (Mean: 541)

Turbidity: 16 - 5999 NTU (Mean: 3538)

pH: 2.00 - 11.99 (Mean: 8.19)

BAD WATER THRESHOLDS

TEST DATA PREDICTIONS

	Actual	Predicted	Predicted_Label
3353	0	0	Bad
9093	0	1	Good
6415	0	0	Bad
8087	0	1	Good
534	1	1	Good
7964	1	0	Bad
5719	0	0	Bad
9700	0	1	Good
5255	0	0	Bad
4779	0	0	Bad

For first 10 text data values

Results & Inference

- Best Model: MLP
- Achieved Accuracy: 80.2
- Conclusion: Reliable for real-time water quality testing.

OUTPUT

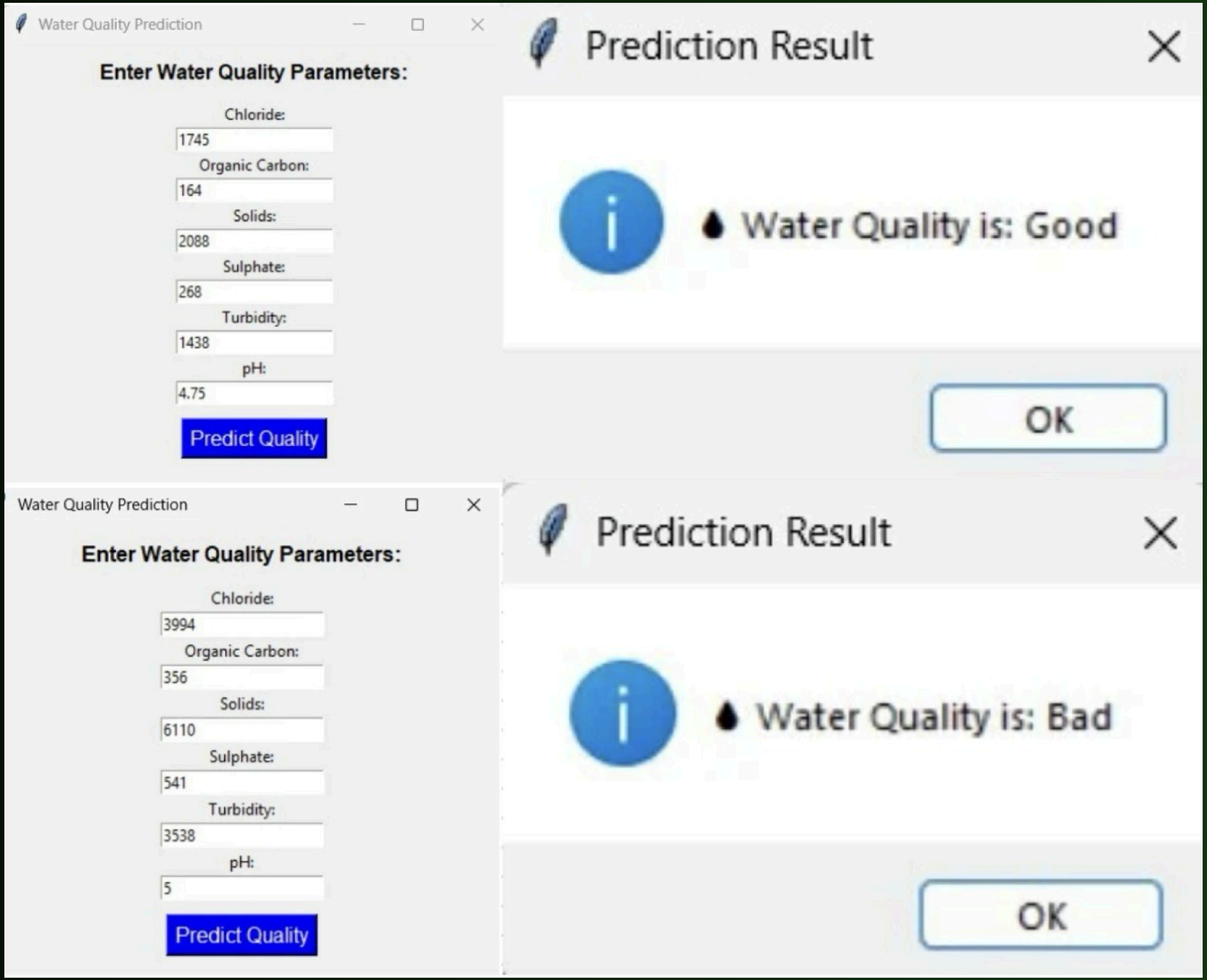


Fig 5- Water Quality Prediction Results

CONCLUSION

Our machine learning-based approach offers faster, automated, and more reliable decision-making compared to traditional water testing methods, which can be time-consuming and prone to human error. This solution enhances construction efficiency and ensures structural integrity by preventing the use of poor-quality water.

Group Members

CB.EN.U4ELC22019	KANISHKA S
CB.EN.U4ELC22037	NIVETHA G K
CB.EN.U4ELC22046	RAHUL B
CB.EN.U4ELC22059	VIKAS K

Thank You!