SMART IRRIGATION REQUIREMENT PREDICTION USING MACHINE LEARNING

A Comprehensive Study on Data-Driven Irrigation Decision Support Systems

This project implements multiple Machine Learning models to predict the level of irrigation required for agricultural fields based on a variety of environmental and soil parameters. The goal is to help farmers and agricultural planners make informed irrigation decisions, optimise water usage, and enhance crop productivity.

Prepared by:

O NITHIN SAI BALAJI

LinkedIn:
https://www.linkedin.com/in/o-ns23/

GitHub:
https://github.com/O-NS23

Project Repo:
https://github.com/O-NS23/Smart-Irrigation-Requirement-Prediction-using-Machine-Learning

Table of Contents

1. Introduction

2. Problem Statement

3. Project Objectives

4. Dataset Description

5. Importance of Each Feature

6. Exploratory Data Analysis

7. Correlation Insights

8. Data Preprocessing Steps

9. Machine Learning Models Used

10. Model Training Strategy

11. Model Evaluation Metrics

12. Model Performance Comparison

13. Confusion Matrix Interpretation

14. Best Performing Model Discussion

15. Real-World Applications

16. Future Enhancements

17. Conclusion

1. Introduction

Water is one of the most essential resources in agriculture. Inefficient irrigation leads to significant water wastage, reduced crop yield, and economic loss. Modern agriculture demands smart, data-driven irrigation systems. This project utilises Machine Learning techniques to predict irrigation requirements categorised into three classes: Low, Medium, and High.

2. Problem Statement

Farmers often rely on manual judgment or traditional methods to decide irrigation quantity. These methods may not always accurately reflect changing environmental and soil conditions. This can lead to either over-irrigation, causing water wastage and root damage or under-irrigation, resulting in plant stress and poor yield.

Therefore, an intelligent system capable of predicting irrigation requirements based on scientific parameters is essential.

3. Project Objectives

• Analyse the irrigation dataset and derive meaningful insights
• Perform Exploratory Data Analysis to understand data behaviour
• Apply preprocessing, including encoding and scaling
• Handle class imbalance using SMOTE
• Train multiple Machine Learning algorithms
• Compare their performance using evaluation metrics
• Identify the best model for real-world usage

4. Dataset Description

The dataset consists of multiple environmental and soil-related parameters that directly influence irrigation requirements. The target variable is Irrigation_Need, which has three categories: Low, Medium, and High.

Target Variable

• Low
• Medium
• High

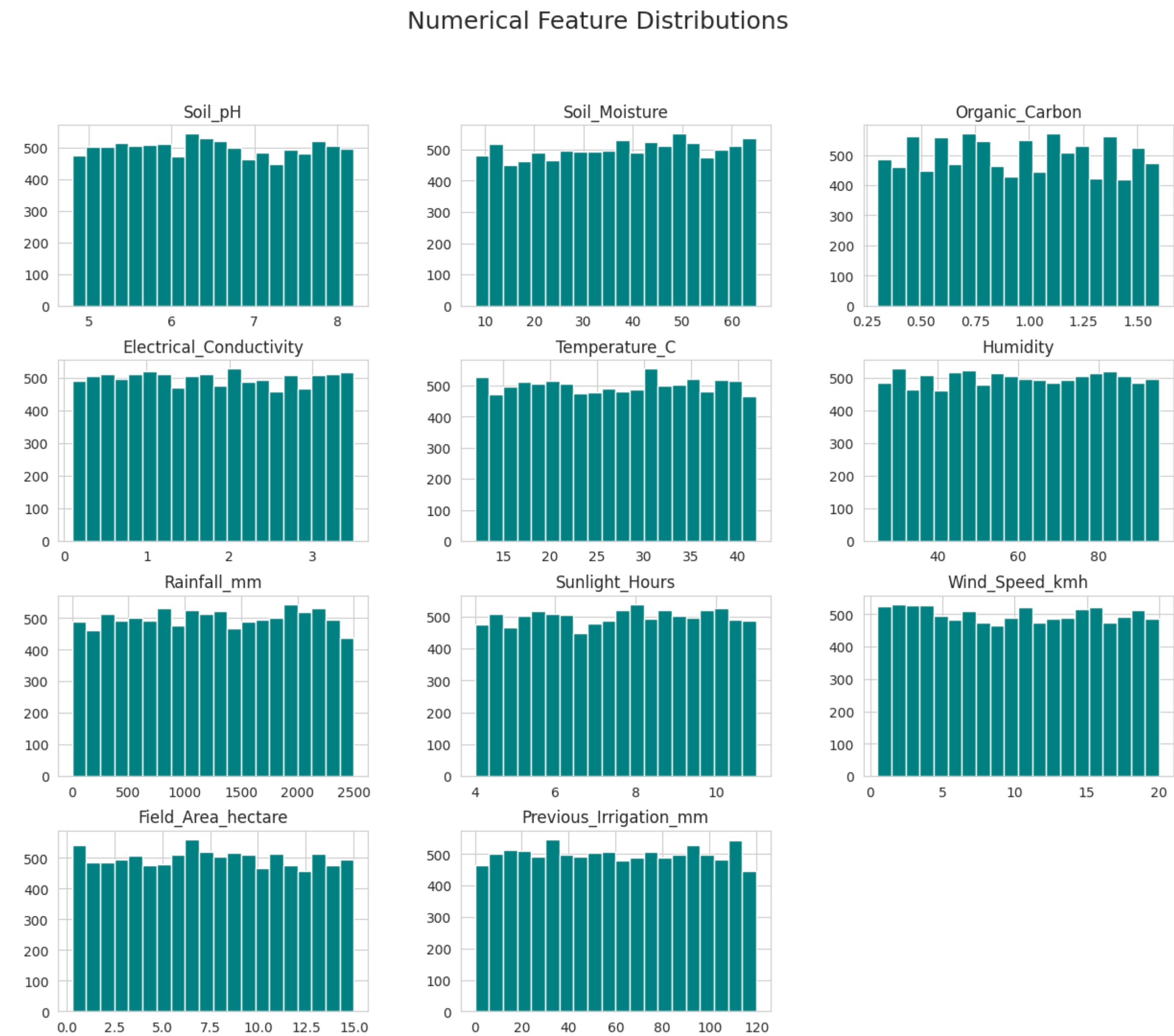
5. Importance of Each Feature

Each feature plays a critical role in determining irrigation levels. Below is the significance of major features:
• Soil Moisture – Primary deciding factor of irrigation requirement
• Temperature – Higher temperature increases evaporation and water demand
• Humidity – High humidity reduces water loss through evaporation
• Rainfall – Natural water input reduces irrigation need
• Organic Carbon and pH – Indicate soil fertility and water retention capability

6. Exploratory Data Analysis

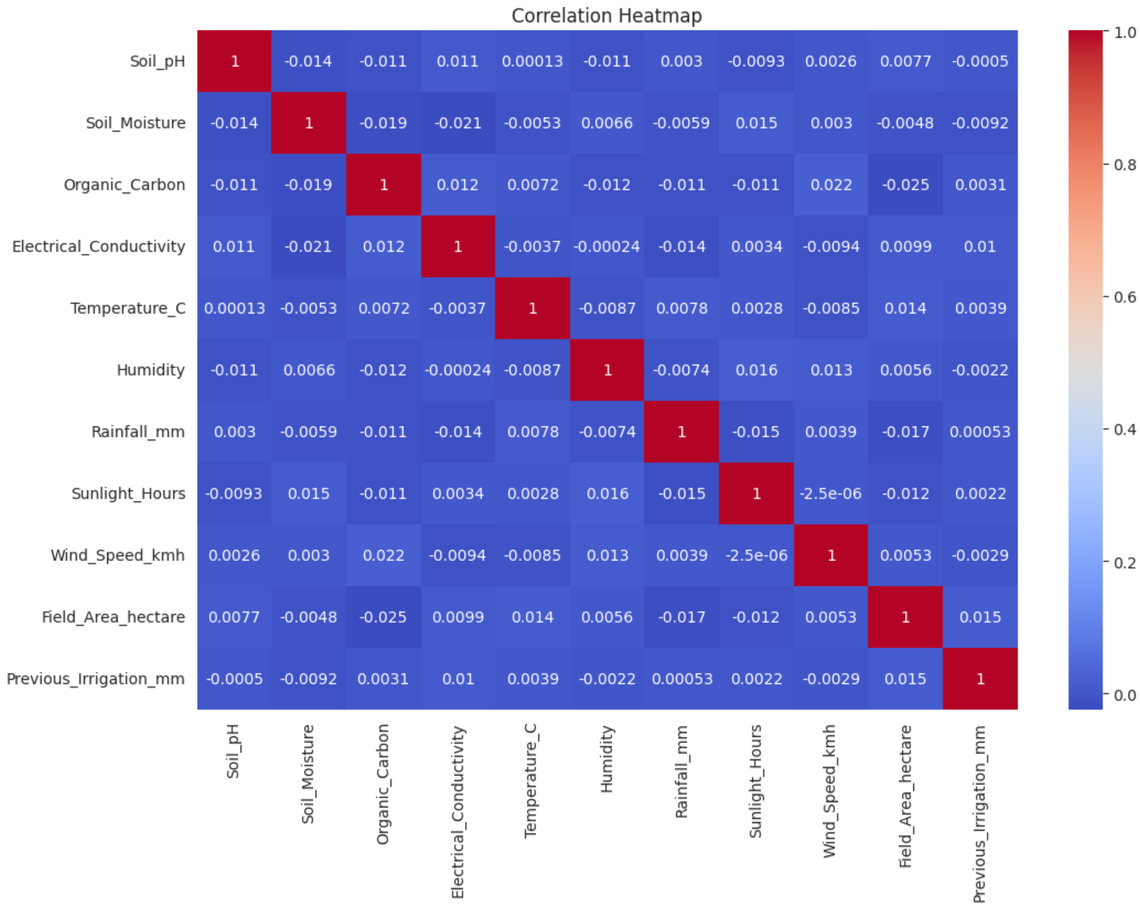
Numerical Feature Distribution

Distribution plots help us observe the spread and skewness of numerical features, ensuring values are meaningful and consistent for modeling.



7. Correlation Insights

The heatmap reveals the relationship between different features. Most features are moderately correlated, indicating each feature contributes meaningful, independent information to the model.



8. Data Preprocessing Steps

1. Label Encoding applied to convert Irrigation_Need into 0,1,2
2. Min-Max Scaling is used to normalize features within the range 0–1
3. SMOTE applied to resolve class imbalance

9. Machine Learning Models Used

The following models were trained:

• Logistic Regression
• Decision Tree
• Random Forest
• Gradient Boosting
• AdaBoost
• KNN
• SVM
• XGBoost
• Naive Bayes

10. Model Training Strategy

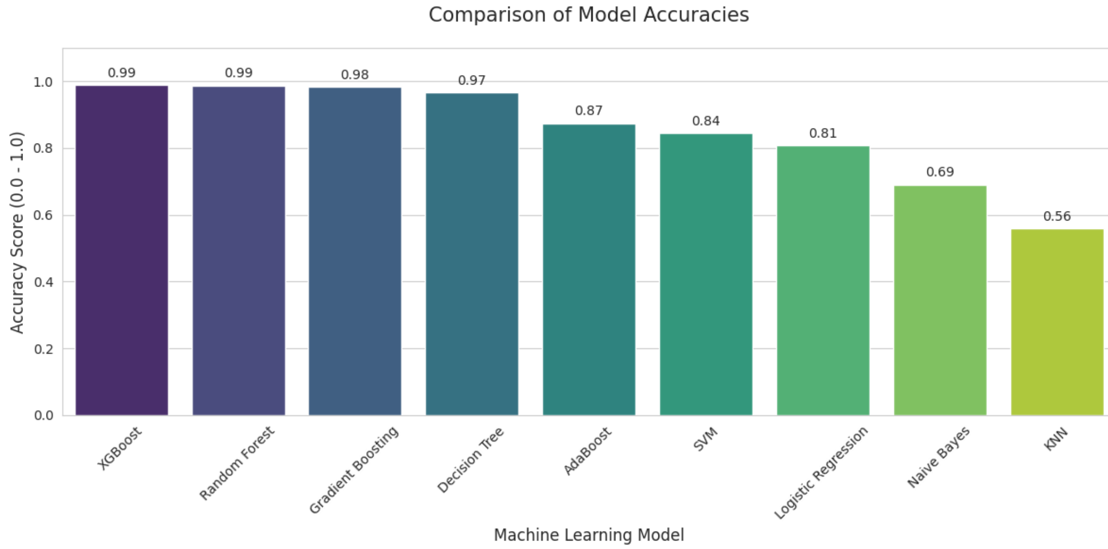
The dataset was split into training and testing sets. Models were trained on scaled and SMOTE-balanced data. Evaluation metrics were computed to judge performance.

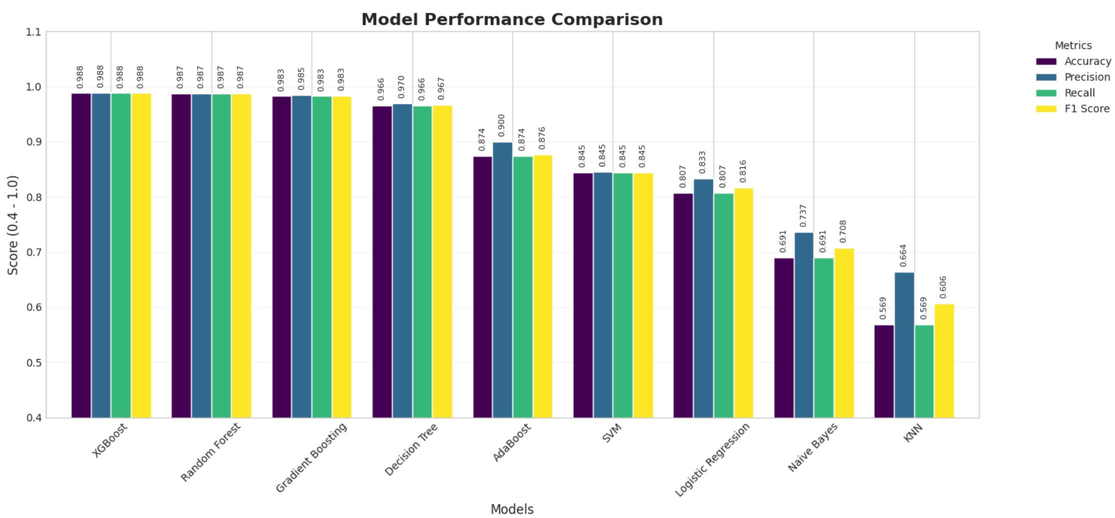
11. Model Evaluation Metrics

Models were evaluated using:

• Accuracy – Overall correctness of prediction
• Precision – Correct positive predictions
• Recall – Ability to detect correct class
• F1 Score – Balance between Precision and Recall

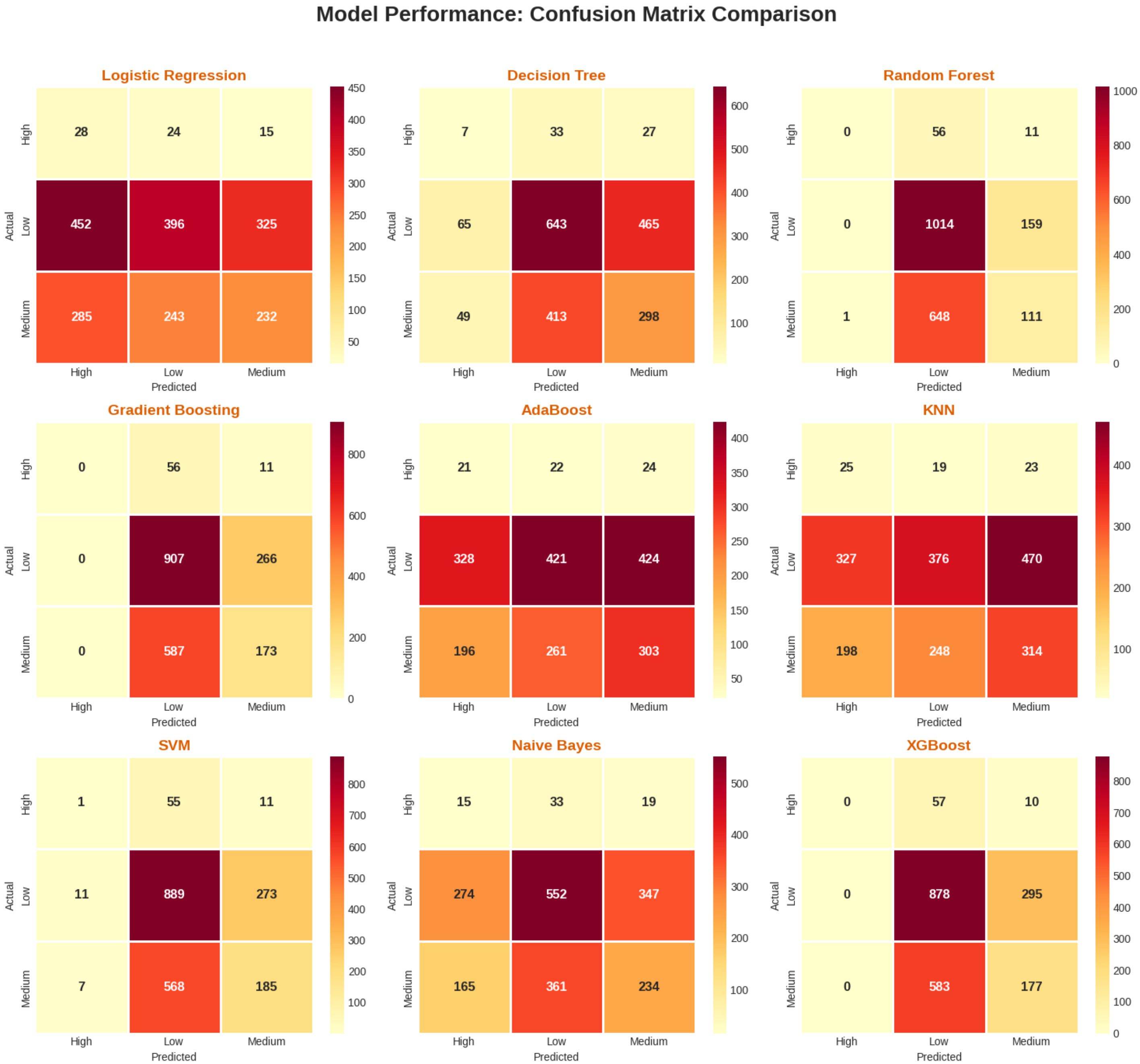
12. Model Performance Comparison



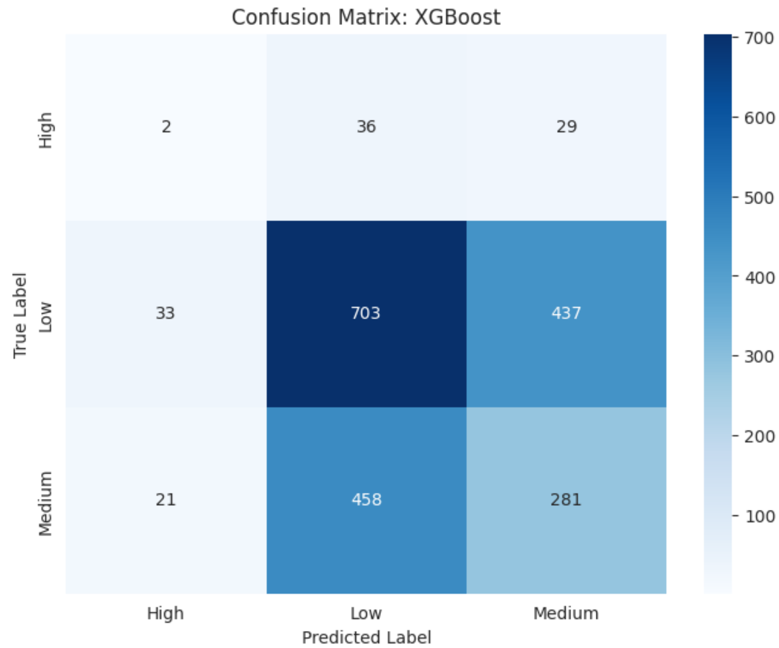


From the visualizations, it is clear that ensemble-based models outperform traditional algorithms.

13. Confusion Matrix Interpretation



Confusion matrices help visualize class-wise performance. The best-performing model showed strong class prediction capability with minimal misclassification.



XGBoost Confusion Matrix

14. Best Performing Model

XGBoost and Random Forest achieved the highest accuracy and most balanced evaluation scores. They handled nonlinear relationships effectively, dealt well with feature interactions, and provided robust predictions.

	Model	Accuracy	Precision	Recall	F1 Score
1	XGBoost	0.9880	0.9881	0.9880	0.9880
2	Random Forest	0.9870	0.9869	0.9870	0.9868
3	Gradient Boosting	0.9825	0.9845	0.9825	0.9831
4	Decision Tree	0.9655	0.9700	0.9655	0.9669
5	AdaBoost	0.8740	0.9005	0.8740	0.8763
6	SVM	0.8445	0.8449	0.8445	0.8445
7	Logistic Regression	0.8070	0.8332	0.8070	0.8163
8	Naive Bayes	0.6905	0.7369	0.6905	0.7077
9	KNN	0.5690	0.6637	0.5690	0.6064

15. Real-World Applications

• Smart irrigation systems
• Precision agriculture
• Automated water management
• IoT integration with soil sensors

16. Future Enhancements

• Deploy as a web or mobile app
• Integrate real-time sensor data
• Apply time-series forecasting
• Use SHAP explainability to justify predictions

17. Conclusion

This project demonstrates the successful application of Machine Learning in agriculture to support irrigation decision-making. With powerful models like XGBoost and Random Forest, irrigation needs can be predicted accurately, supporting sustainable farming, water conservation, and enhanced crop productivity.