Groundwater Level Prediction Using Multiple Linear Regression

# 1. Scenario

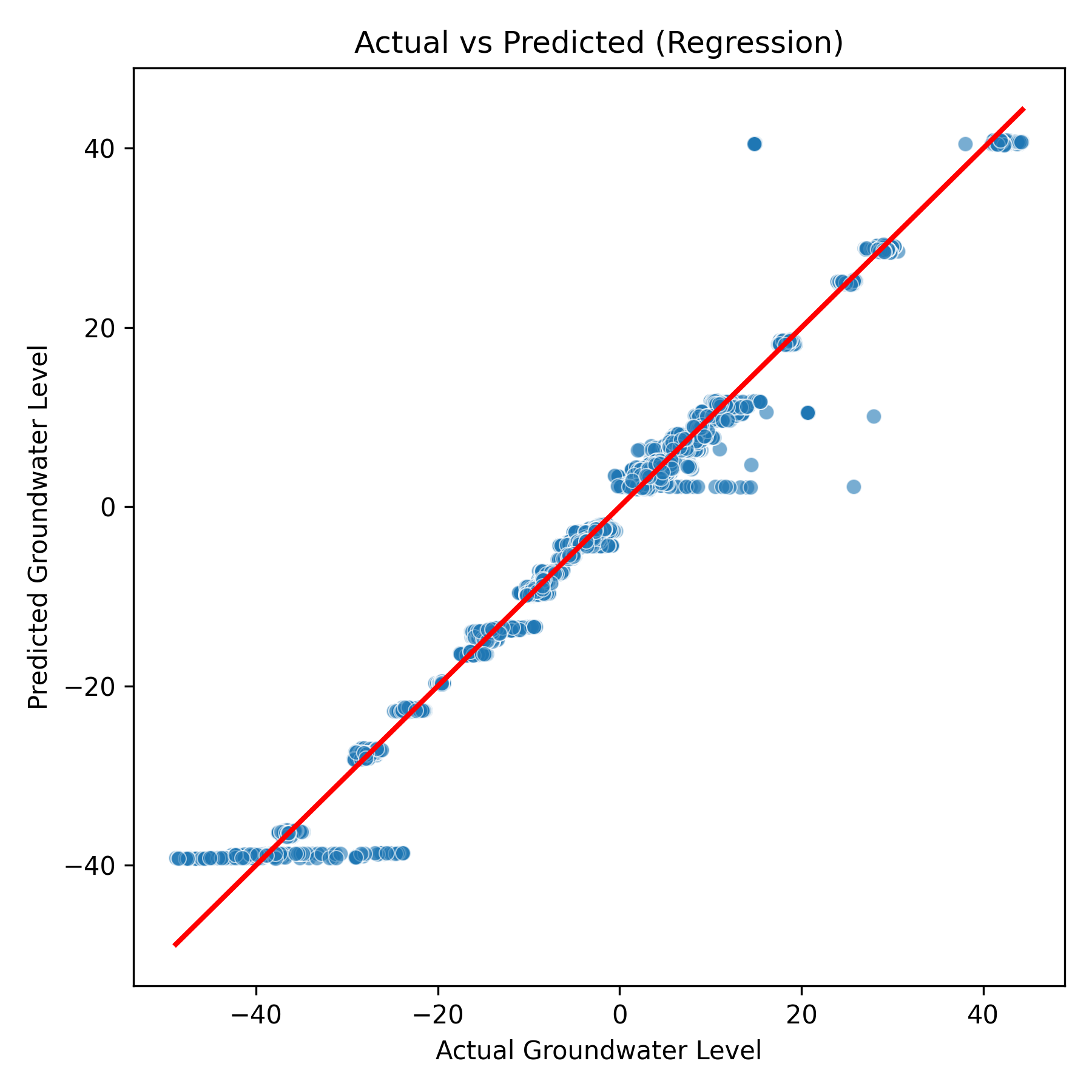
* Delhi NCR
* Objective: Identify key drivers of groundwater depletion and predict groundwater levels at unsampled locations.

# 2. Methodology

* **Data:** Preprocessed groundwater dataset with 175 columns including the target 'data Value'.
* **Dependent Variable**: data value
* **Independent Variables**: BIC criterion used (58 features).
* **Model**: Multiple Linear Regression (OLS).
* **Temporal unit**: Daily/Monthly/Yearly
* **Spatial unit**: District, Latitude, Longitude etc.
* **Data Acquisition**: Data acquired from India WRIS, Bhuvan ISRO, Copernicus Climate Data Store, NICES Portal, SHRUG Atlas.
* **Data Preprocessing**: Missing values handled, outliers removed, data merged appropriately.

# 3. Exploratory Data Analysis (EDA)

* Scatter plots and pair plots were used to explore relationships between features and groundwater levels.
* Spatial and temporal trends were observed.
* Data was cleaned and outliers removed.



# 4. Model Assumptions

* Linearity: Relationships between predictors and target are linear.
* No Perfect Multicollinearity: Checked correlations among predictors.
* Exogeneity: Residuals uncorrelated with predictors.
* Homoscedasticity: Residuals have constant variance.

# 5. Model Selection

* Compared models using AIC and BIC.
* BIC-selected model (58 predictors) was chosen for analysis.

# 6. Model Estimation & Diagnostics

* Top 10 Coefficients:

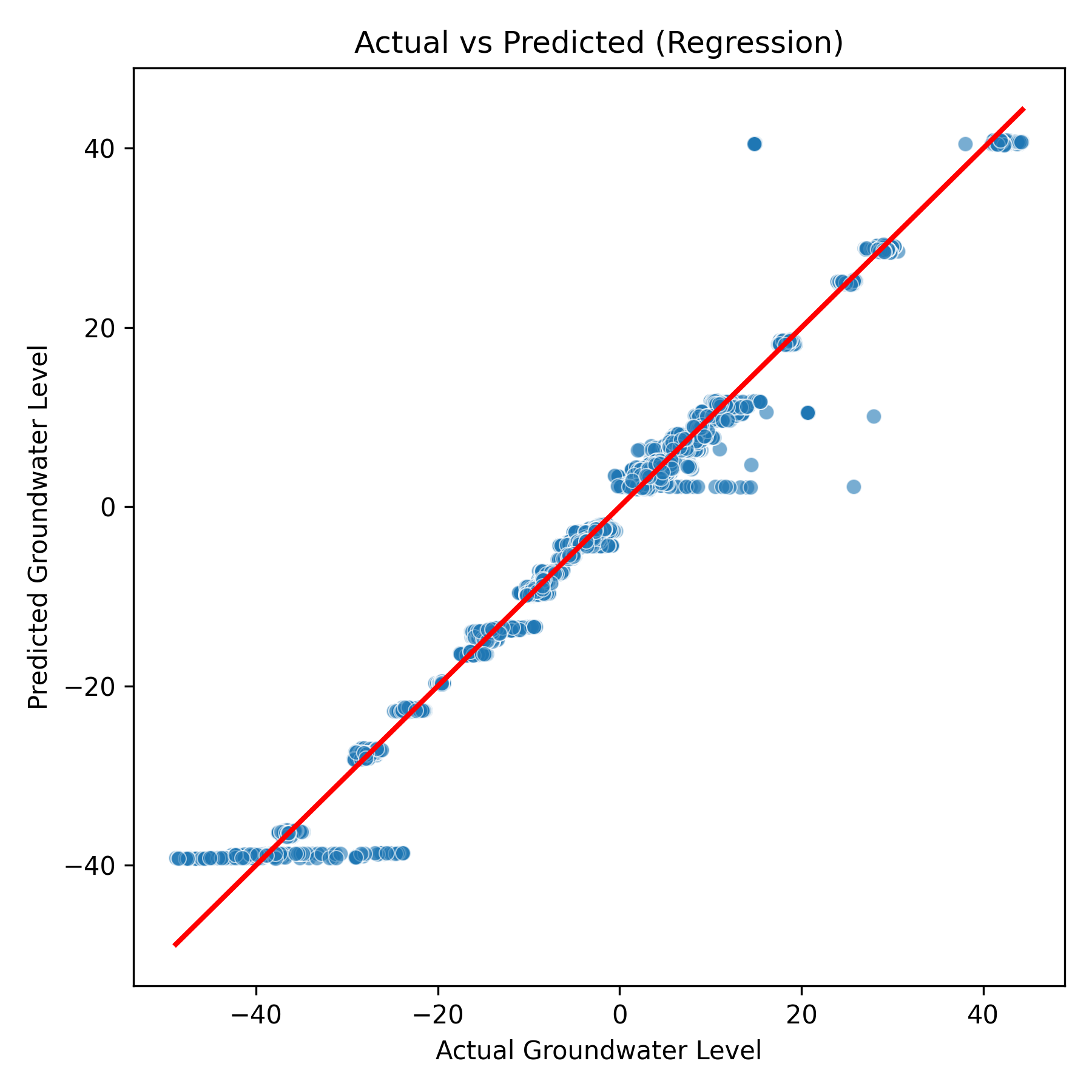
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Feature | Coefficient | Std\_Error | t\_value | p\_value |
| stationCode.1 | -0.22627 | 0.000746 | -303.44491 | 0.0 |
| HISTO\_20 | -81.591015 | 1.554332 | -52.49266 | 0.0 |
| Rainfall (mm)\_2024 | -0.539686 | 0.003682 | -146.55966 | 0.0 |
| \_count | -5.4e-05 | 0.0 | -377.401993 | 0.0 |
| HISTO\_80 | 574.507576 | 9.251323 | 62.100043 | 0.0 |
| shape\_leng | -0.002849 | 1e-05 | -273.703361 | 0.0 |
| Pre Monsoon of GW Trend\_2024 | 47.434549 | 0.648234 | 73.175021 | 0.0 |
| stationName | 0.592188 | 0.002755 | 214.934334 | 0.0 |
| Categorization of Assessment Unit\_2023 | -33.680615 | 0.38376 | -87.764782 | 0.0 |
| sand\_5-15cm\_mean | 1.608464 | 0.01368 | 117.575387 | 0.0 |

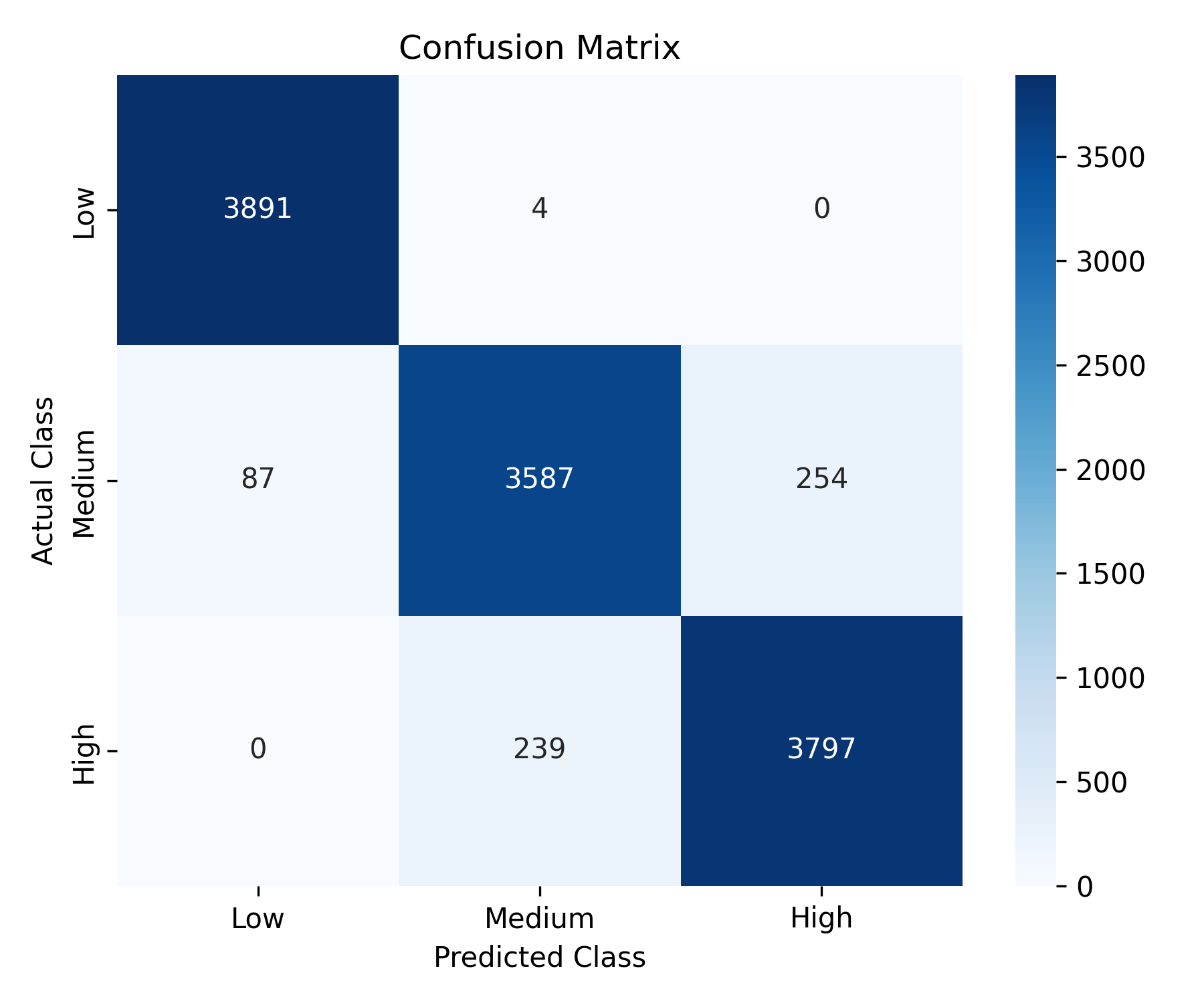
**Model Fit Metrics:**

* R\_squared(training): 0.9904
* Adj\_R\_squared: 0.9904
* F\_stat: 49925.6491
* F\_pvalue: 0.0
* Num\_predictor: 58.0

# 7. Predictions & Evaluation

* Accuracy: 0.9508
* Precision: 0.9505
* Recall: 0.9508
* F1-score: 0.9505
* R-squared (test): 0.9868
* RMSE (regression): 1.7404
* MSE (regression): 3.0291
* Plots:



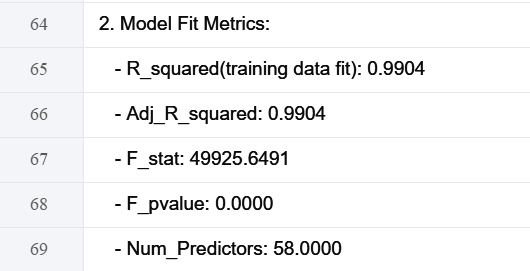


# 8. Significant Features & Interpretation

Top 12 Rows of Significant Features:

A screenshot of a computer

AI-generated content may be incorrect.



A screenshot of a phone

AI-generated content may be incorrect.

**Confidence in Interpretation:**

* The model explains most of the variability in groundwater levels (high R-squared).
* Significant features with p < 0.05 are likely reliable predictors.
* Prediction metrics (RMSE for regression and Accuracy/F1 for categorized classes) indicate reasonable predictive power.

# 9. Conclusion & Policy Implications

* The model identifies key factors affecting groundwater levels.
* Predictions can guide water resource planning.
* Recommendations: Monitor significant drivers and use the model for short-term planning and risk assessment.

# 10. References

* India WRIS: https://indiawris.gov.in/wris/
* Bhuvan ISRO: https://bhuvan-app1.nrsc.gov.in/2dresources/bhuvanstore2.php
* Copernicus Climate Data Store: https://cds.climate.copernicus.eu/#!/home
* NICES Portal: https://nices.nrsc.gov.in/
* SHRUG Atlas: https://www.devdatalab.org/atlas