Groundwater Level Prediction Using Multiple Linear Regression

# 1. Scenario

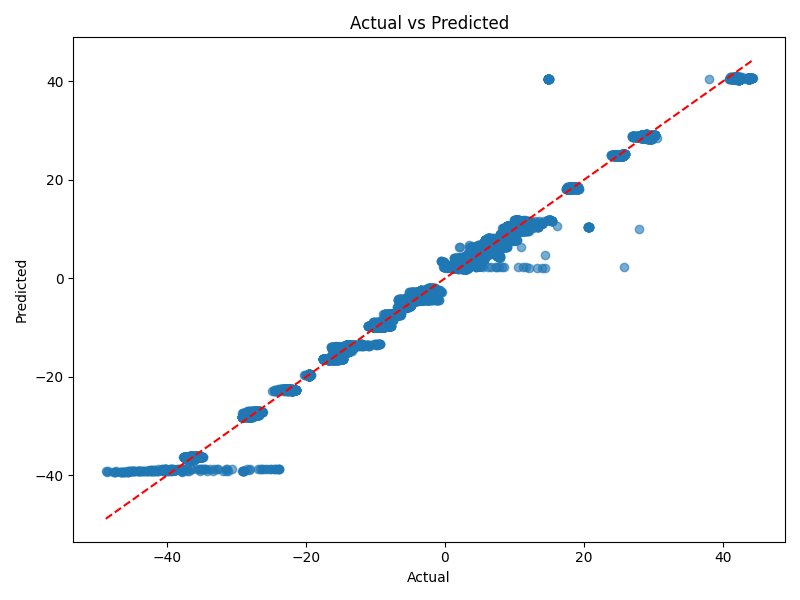
* Delhi NCR
* Objective: Identify key drivers of groundwater depletion and predict groundwater levels at unsampled locations.

# 2. Methodology

* **Data:** Preprocessed groundwater dataset with 175 columns including the target 'data Value'.
* **Dependent variable**: dataValue
* **Independent variables**: 58 selected features (BIC criterion)
* **Spatial unit**: District
* **Temporal unit**: Daily
* **Model equation**: Multiple Linear Regression (OLS)
* **Data Acquisition**: Data acquired from India WRIS, Bhuvan ISRO, Copernicus Climate Data Store, NICES Portal, SHRUG Atlas.
* **Data Merging**: Merged datasets on district and date.
* **Data Preprocessing**: Missing values handled, outliers removed, data merged appropriately.
* **Model Specification**: Defined model structure and selected features.
* **Model Training**: Trained the model using the training dataset.
* **Model Evaluation**: Evaluated model performance using test dataset and metrics like R², RMSE.
* **Model Diagnostics**: Residuals analyzed for patterns.
* **EDA**: Scatter plots, pair plots, and spatial-temporal plots analyzed to observe trends.

# 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

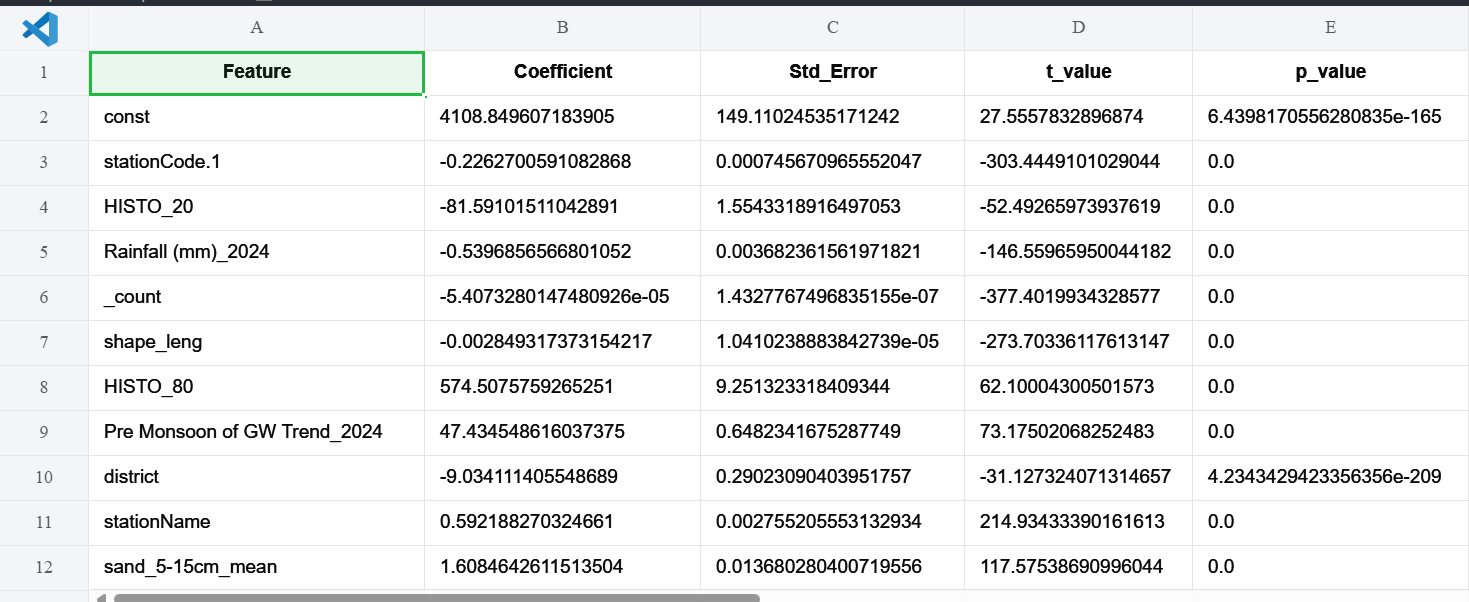
* Linearity: Relationships between predictors and target are linear.
* No Perfect Multicollinearity: Checked correlations among predictors.
* Exogeneity: Residuals uncorrelated with predictors.
* Homoscedasticity: Residuals have constant variance.

# 5. Model Selection

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

# 6. Model Estimation & Diagnostics

* Top 12 Coefficients:

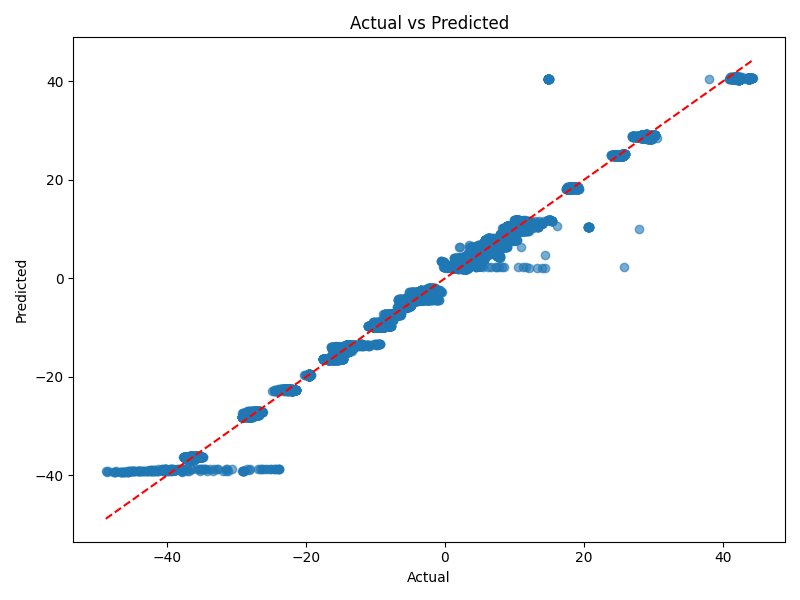


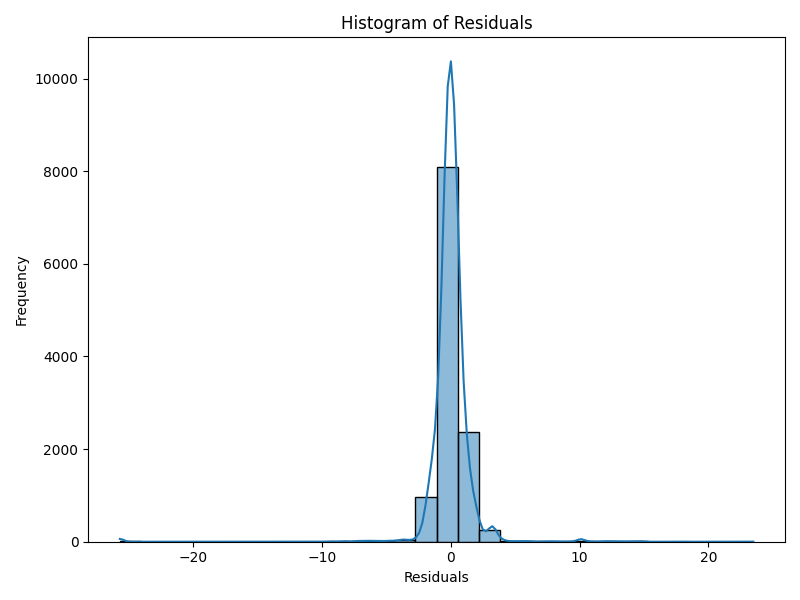
**Model Fit Metrics:**

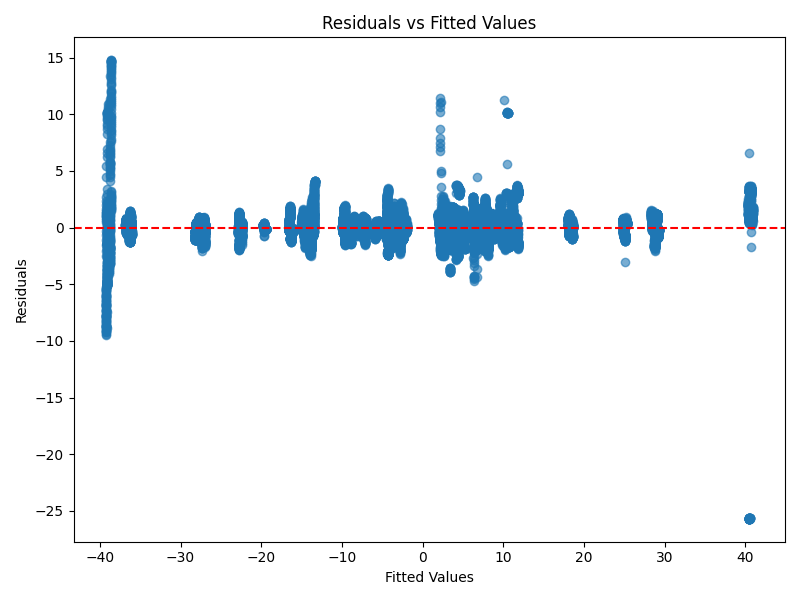
* R\_squared(training): 0.9904
* Adj\_R\_squared: 0.9904
* F\_stat: 49925.6491
* F\_pvalue: 0.0000
* Num\_Predictors: 58.0000

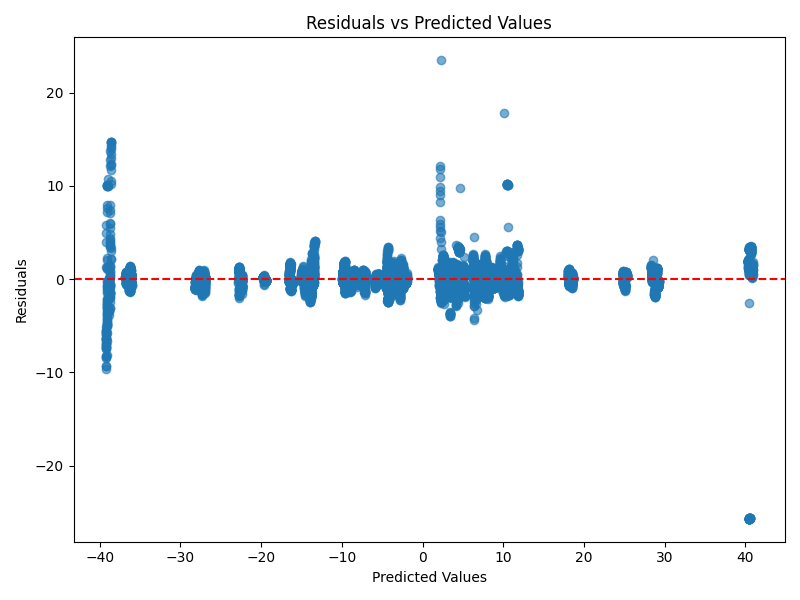
# 7. Predictions & Evaluation

* R-squared(testing data): 0.9868
* RMSE: 1.7404
* MSE: 3.0291
* Plots:



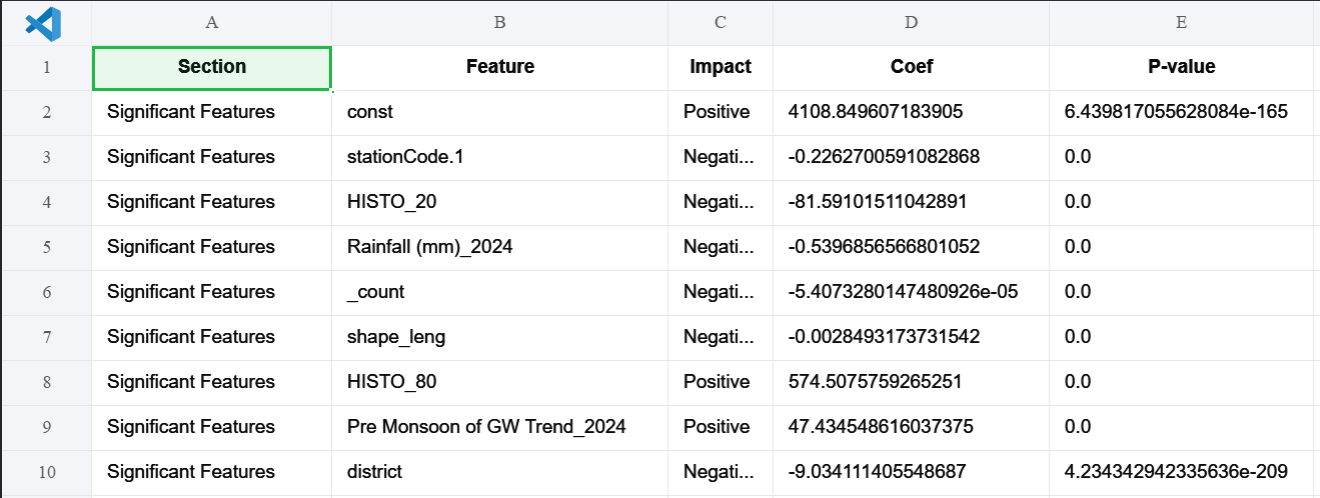




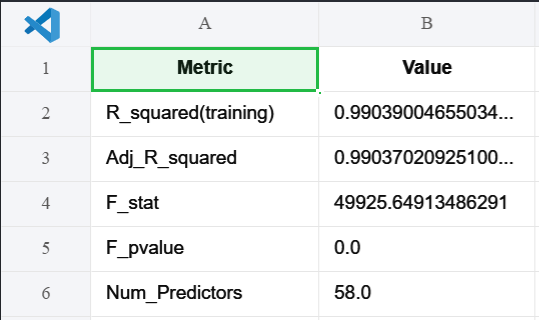


* Residual Plot
* The residual plot shows the difference between observed and predicted values. It helps in diagnosing the model fit.
* Interpretation: No clear pattern suggests a good fit.
* Action: Consider refining the model if patterns are detected.

# 8. Significant Features & Interpretation



**Model Fit Metrics: Model Prediction Metrics:**

 A screenshot of a calculator

AI-generated content may be incorrect.

**Confidence in Interpretation:**

* Description: 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