

# Machine Learning Model Tracking Document

## 1. Dataset Information

Dataset Name:	Well 1225
Number of Samples:	1225
Number of Features:	4 (Tf, Rs, Gg, Api)
Target Variable:	Pb
Outlier Handling:	None
Feature Engineering Applied:	None
Scaling/Normalization Applied:	None
Encoding Applied:	None

## 2. Preprocessing Steps

Step	Description
Train-Test Split	70% - 30%
Shuffling	Yes, using random_state=42
Handling Missing Data	None
Feature Scaling	SVM Only
Feature Selection	None

## 3. Models Used & Hyperparameters

Model	Hyperparameters	Training Time
Linear Regression	None	0.0008
Ridge Regression	alpha = 0.001	0.0011
Lasso Regression	alpha = 0.001	0.0023
Decision Tree	max_depth = 12	0.0032
KNN	n_neighbors = 7	0.0009
SVR	C = 100, kernel = linear	32.7210

## 4. Evaluation Metrics

Model	MSE	RMSE	MAE	R <sup>2</sup> Score	Adjusted R <sup>2</sup>
Linear Regression	258256.3	508.189	348.429	0.8239	0.8219
Lasso Regression	258256.6	508.189	348.429	0.8239	0.8219
Ridge Regression	258256.7	508.190	348.430	0.8239	0.8219
Decision Tree	197204.6	444.077	255.040	0.8655	0.8640
KNN	276717.1	526.039	366.417	0.8113	0.8092
SVR	268059.6	517.745	329.692	0.8172	0.8152

## 5. Cross-Validation Summary (5-Fold)

Model	RMSE Mean	RMSE Std	MAE Mean	MAE Std	R <sup>2</sup> Mean	R <sup>2</sup> Std
Decision Tree	409.140	42.540	256.751	17.748	0.8824	0.0260
KNN	513.670	25.454	367.940	9.634	0.8172	0.0194
Ridge Regression	526.882	63.120	358.851	21.284	0.8068	0.0425
Lasso Regression	526.883	63.120	358.851	21.284	0.8068	0.0425
Linear Regression	526.883	63.120	358.851	21.284	0.8068	0.0425
SVR	607.730	36.829	418.477	19.101	0.7448	0.0230

## 6. Visualizations

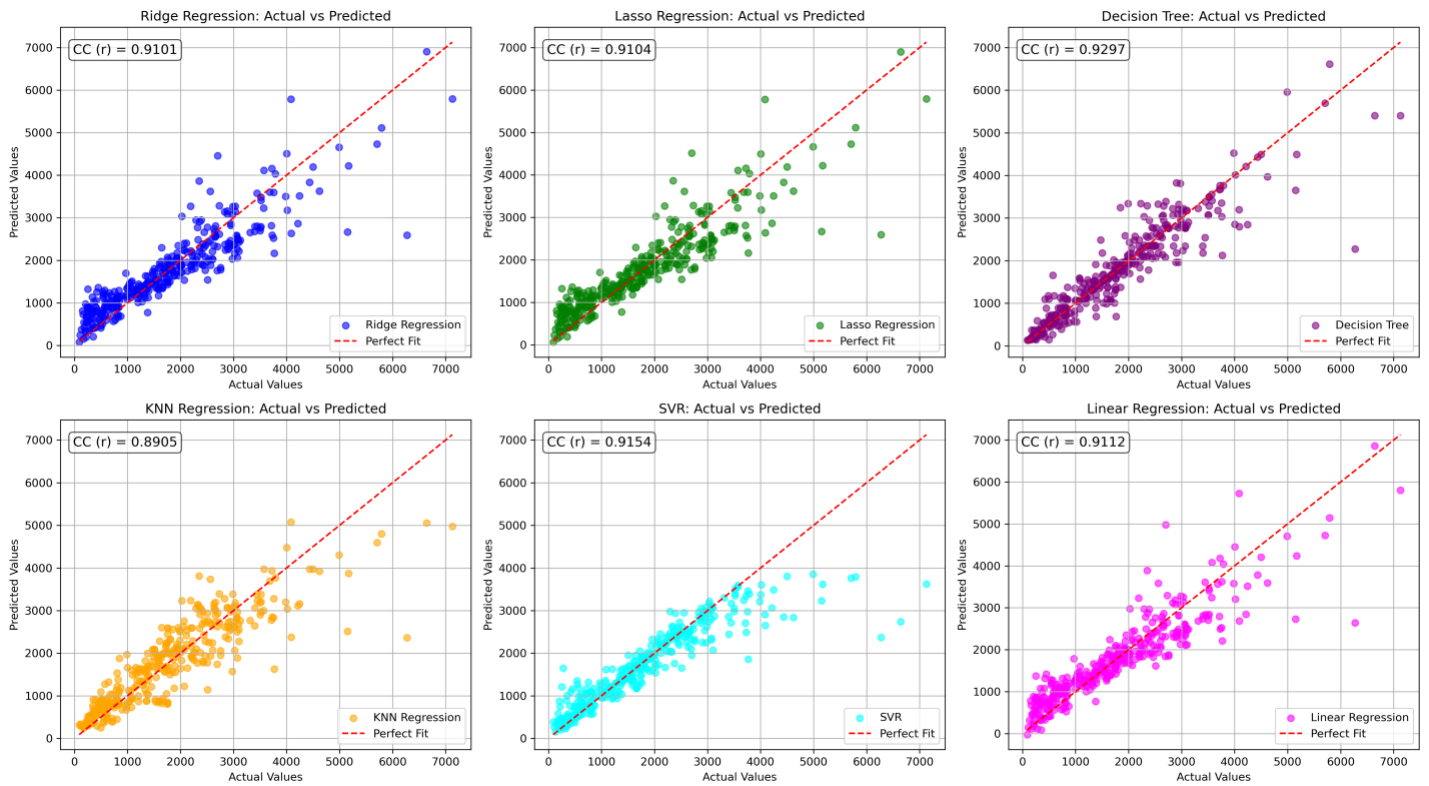


Figure 1: Actual vs. Predicted Values for Pb (Well 1225)

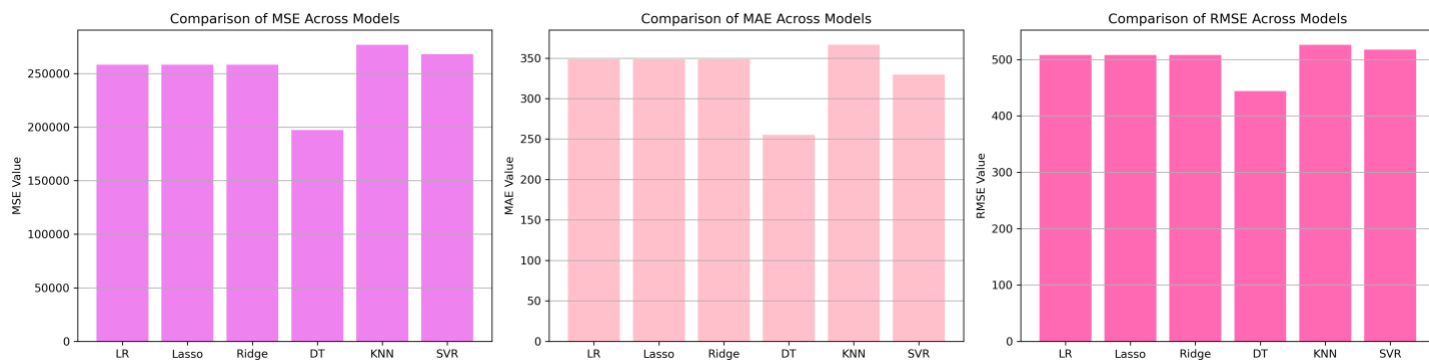


Figure 2: Bar Charts of MSE, RMSE, MAE for Pb (Well 1225)

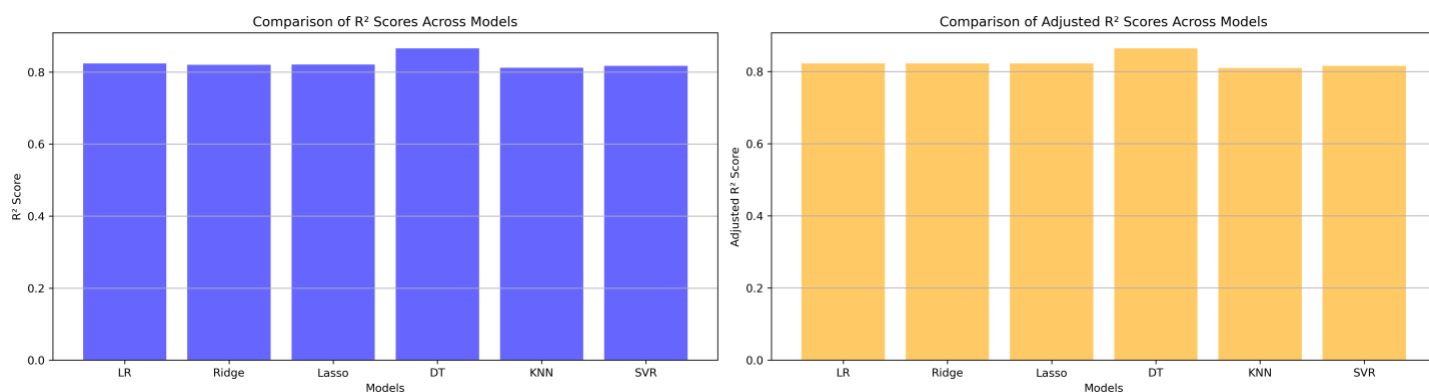


Figure 3: Bar Charts of  $R^2$  and Adjusted  $R^2$  for Pb (Well 1225)

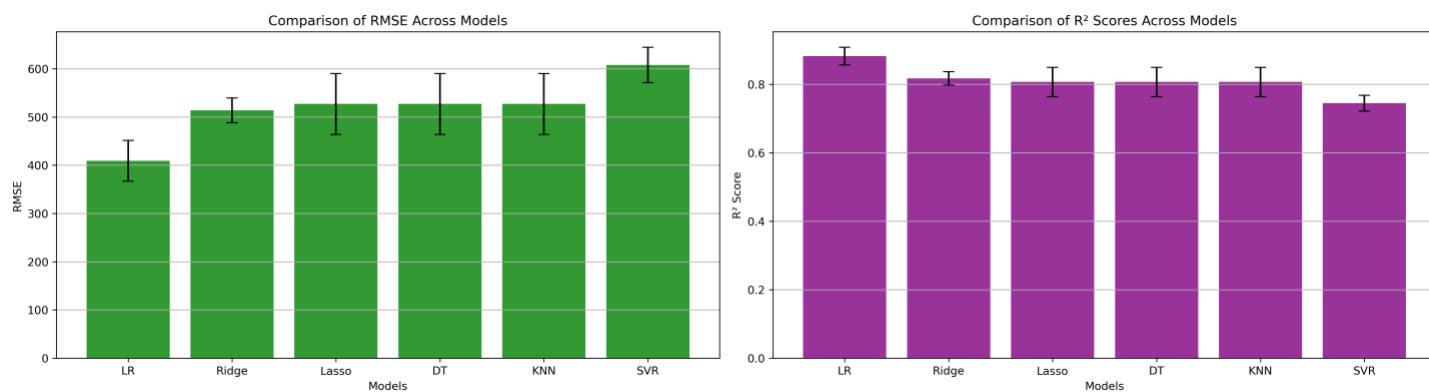


Figure 4: Error Bars for RMSE and  $R^2$  from CV for Pb (Well 1225)

## 7. Observations & Next Steps

- **Best Performing Model:** Decision Tree (RMSE: ~444.08,  $R^2$ : ~0.8655)
  - Linear, Lasso, and Ridge performed similarly
  - Decision Tree outperformed others on both test and CV
  - SVR showed poor results despite long training

## 8. Code Access

The complete source code for data preprocessing, model training, evaluation, and visualization is [available here](#). The repository includes organized Jupyter notebooks for each phase, dataset, and target, as well as requirements for reproducibility.