# Machine Learning Model Tracking Document

## 1. Dataset Information

|  |  |
| --- | --- |
| Dataset Name: | Well 160 |
| Number of Samples: | 160 |
| 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 | Only SVR |
| Feature Selection | None |

## 3. Models Used & Hyperparameters

|  |  |  |
| --- | --- | --- |
| Model | Hyperparameters | Training Time |
| Linear Regression | None | 0.0008 |
| Ridge Regression | alpha = 0.001 | 0.0015 |
| Lasso Regression | alpha = 0.001 | 0.0032 |
| Decision Tree | max\_depth = 6 | 0.0007 |
| KNN | n\_neighbors = 13 | 0.0006 |
| SVR | C = 100, epsilon= 0.01 ,kernel = rbf | 0.0006 |

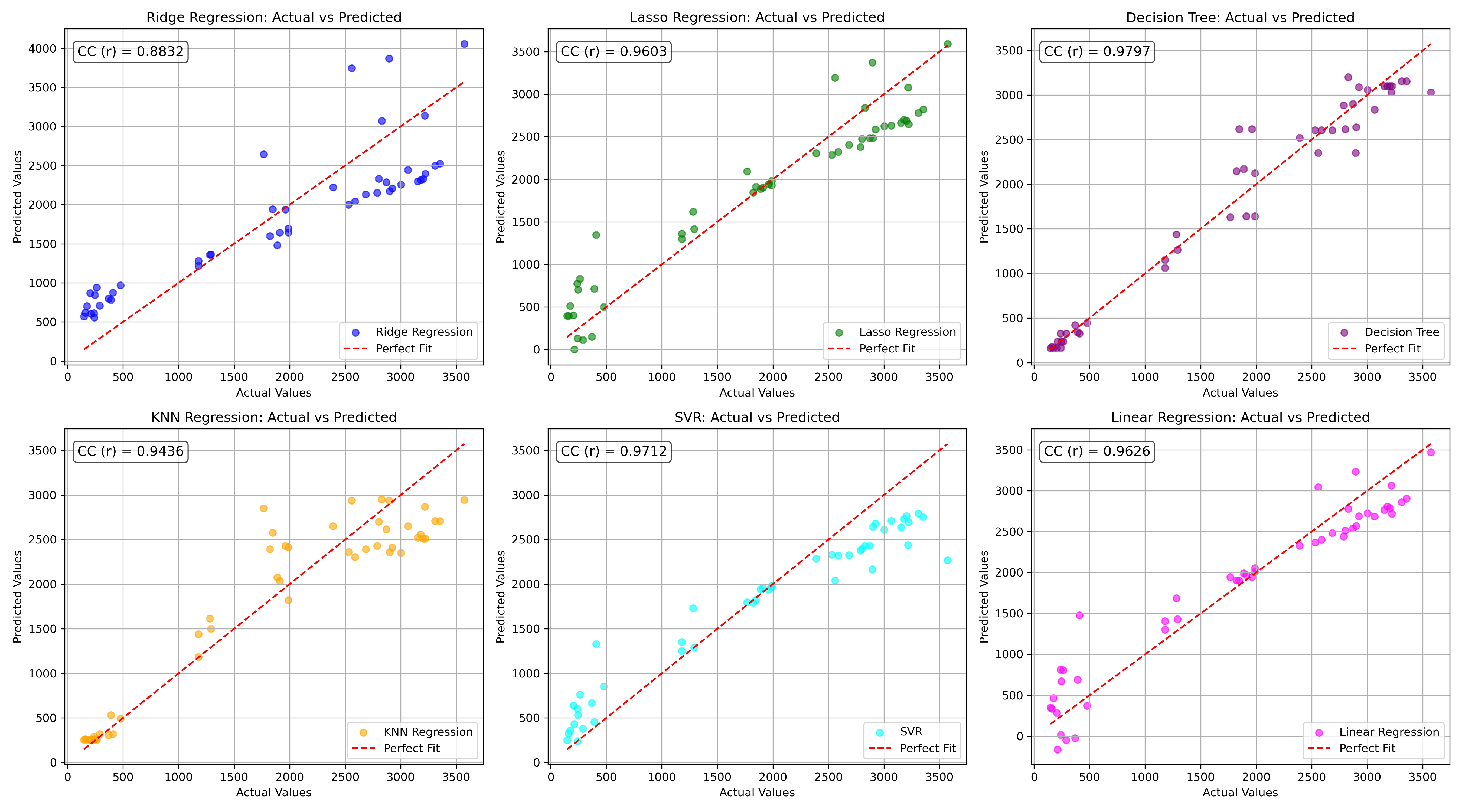
## 4. Evaluation Metrics

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | MSE | RMSE | MAE | R² Score | Adjusted R² |
| Linear Regression | 110361.5 | 332.21 | 272.51 | 0.9190 | 0.9115 |
| Lasso Regression | 110362.2 | 332.21 | 272.52 | 0.9190 | 0.9115 |
| Ridge Regression | 110383.5 | 332.24 | 272.54 | 0.9190 | 0.9115 |
| Decision Tree | 44140.4 | 210.10 | 133.62 | 0.9676 | 0.9646 |
| KNN | 141715.6 | 376.45 | 270.66 | 0.8960 | 0.8864 |
| SVR | 169585.1 | 411.80 | 315.96 | 0.8756 | 0.8640 |

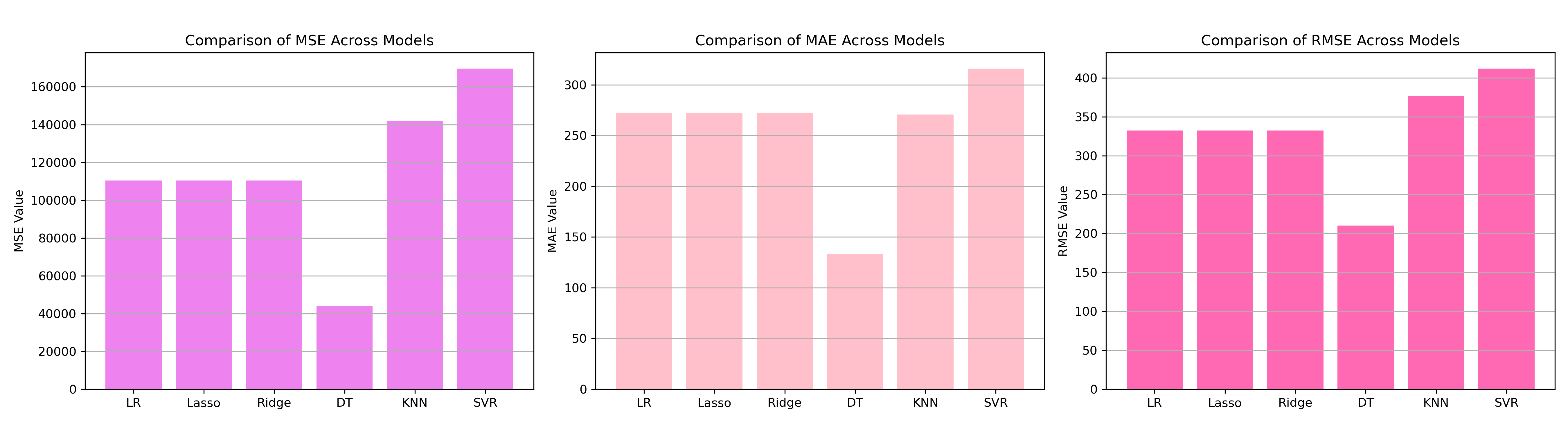
## 5. Cross-Validation Summary (5-Fold)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | RMSE Mean | RMSE Std | MAE Mean | MAE Std | R² Mean | R² Std |
| Decision Tree | 249.94 | 26.11 | 164.76 | 17.23 | 0.9443 | 0.0131 |
| Ridge Regression | 326.36 | 41.78 | 250.87 | 21.68 | 0.9041 | 0.0304 |
| Lasso Regression | 326.37 | 41.78 | 250.88 | 21.66 | 0.9041 | 0.0304 |
| Linear Regression | 326.37 | 41.78 | 250.88 | 21.66 | 0.9041 | 0.0304 |
| KNN | 382.91 | 63.47 | 291.19 | 47.00 | 0.8603 | 0.0665 |
| SVR | 417.55 | 37.10 | 315.86 | 24.32 | 0.8388 | 0.0597 |

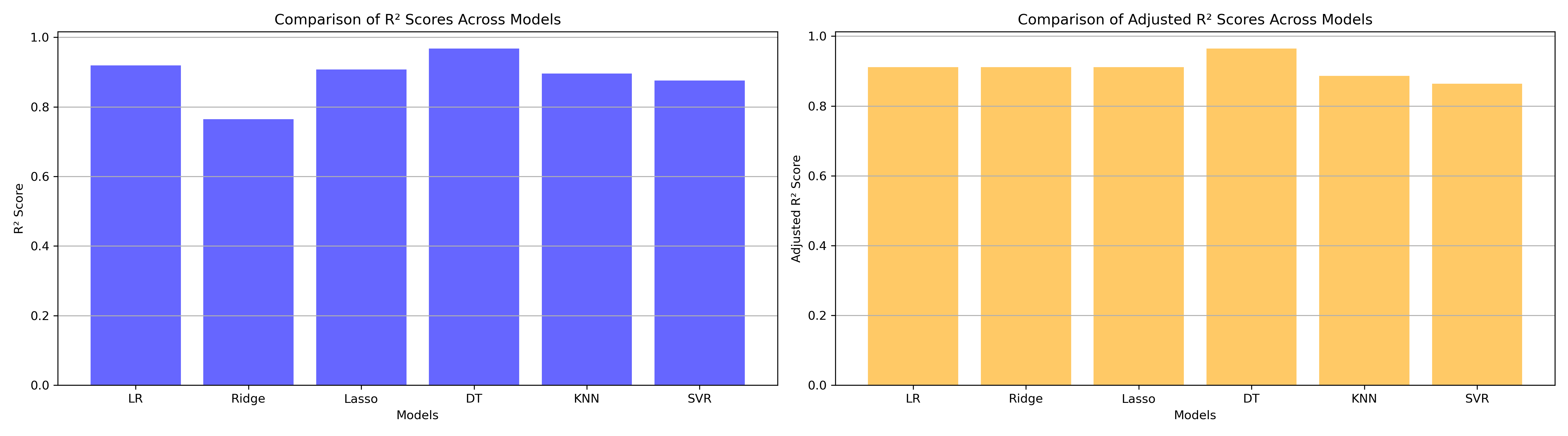
## 6. Visualizations



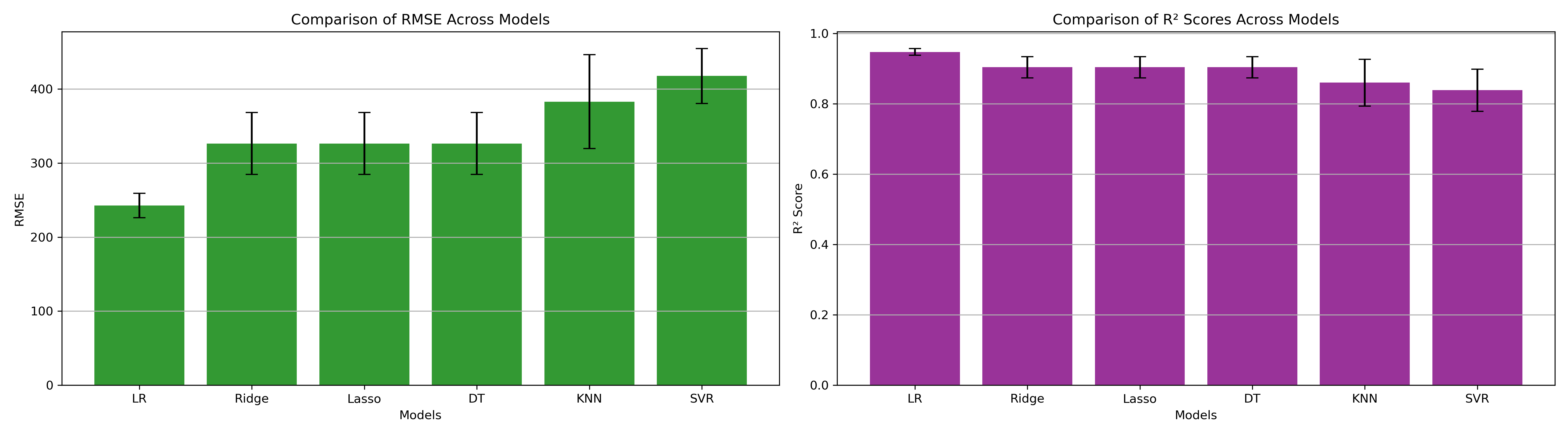
*Figure 1: Actual vs. Predicted Values for Pb (Well 160)*



*Figure 2: Bar Charts of MSE, RMSE, MAE for Pb (Well 160)*



*Figure 3: Bar Charts of R² and Adjusted R² for Pb (Well 160)*



*Figure 4: Error Bars for RMSE and R² from CV for Pb (Well 160)*

### 7. Observations & Next Steps

* **Best Performing Model**: Decision Tree (RMSE: ~210.10, R²: ~0.9676)
  + Decision Tree outperformed regularized linear models
  + SVR shows significant underperformance
  + High variance in KNN and SVR RMSE across folds

### 8. Code Access

The complete source code for data preprocessing, model training, evaluation, and visualization is [available here.](http://github.com/BoushBoo/pvt-prediction-ml-/tree/main) The repository includes organized Jupyter notebooks for each phase, dataset, and target, as well as requirements for reproducibility.