# Machine Learning Model Tracking Document

## 1. Dataset Information

|  |  |
| --- | --- |
| Dataset Name: | Well 283 |
| Number of Samples: | 283 |
| 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 = 1.29 | 0.0006 |
| Lasso Regression | alpha = 10.0 | 0.0006 |
| Decision Tree | max\_depth = 13 | 0.0008 |
| KNN | n\_neighbors = 11 | 0.0006 |
| SVR | C = 100.0, kernel = linear | 2.0789 |

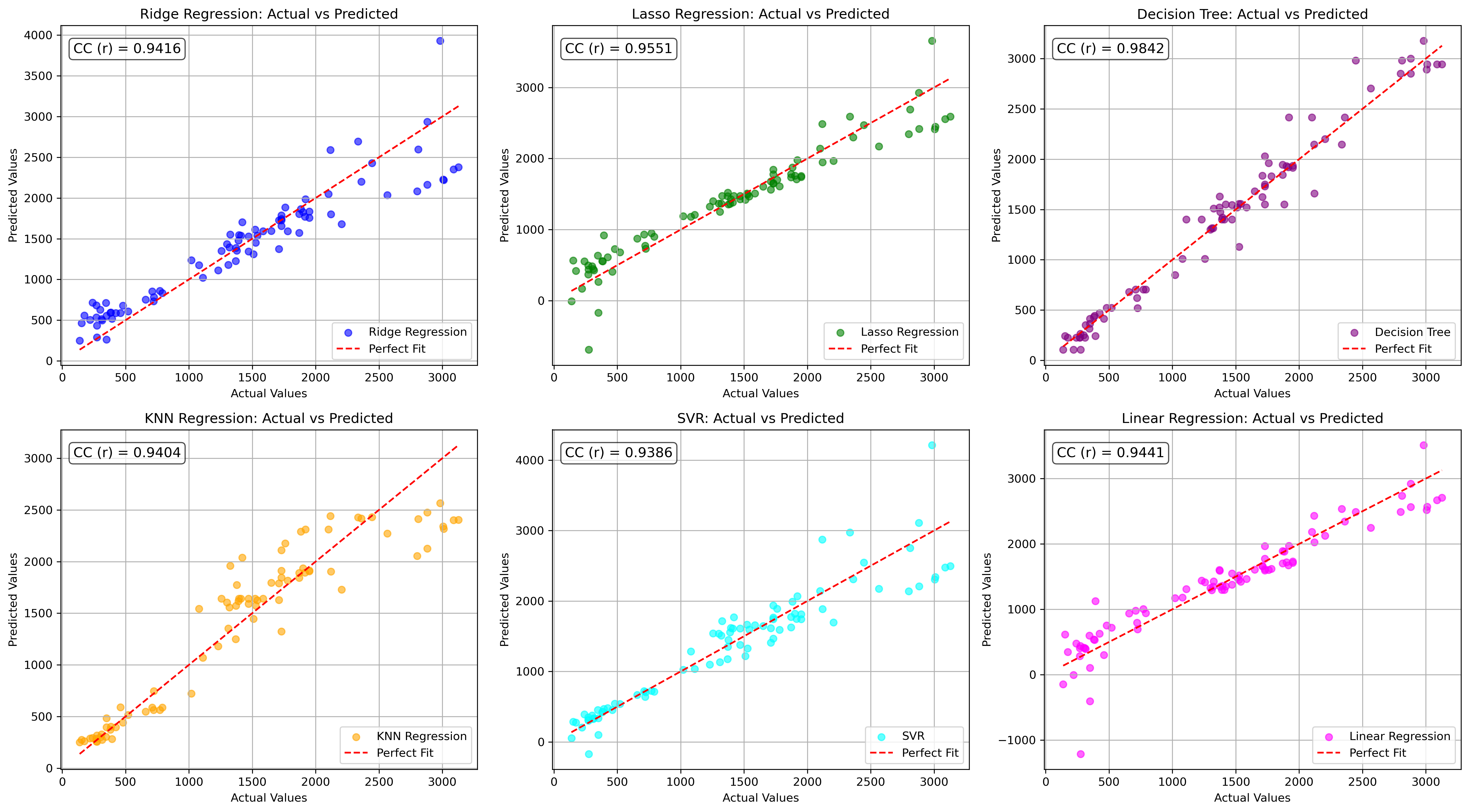
## 4. Evaluation Metrics

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | MSE | RMSE | MAE | R² Score | Adjusted R² |
| Linear Regression | 78159.1 | 279.57 | 190.75 | 0.8900 | 0.8845 |
| Lasso Regression | 63632.3 | 252.25 | 179.87 | 0.9105 | 0.9060 |
| Ridge Regression | 62918.3 | 250.84 | 178.56 | 0.9115 | 0.9071 |
| Decision Tree | 18932.5 | 137.60 | 98.39 | 0.9734 | 0.9720 |
| KNN | 74393.8 | 272.75 | 196.76 | 0.8953 | 0.8901 |
| SVR | 86506.3 | 294.12 | 195.92 | 0.8783 | 0.8722 |

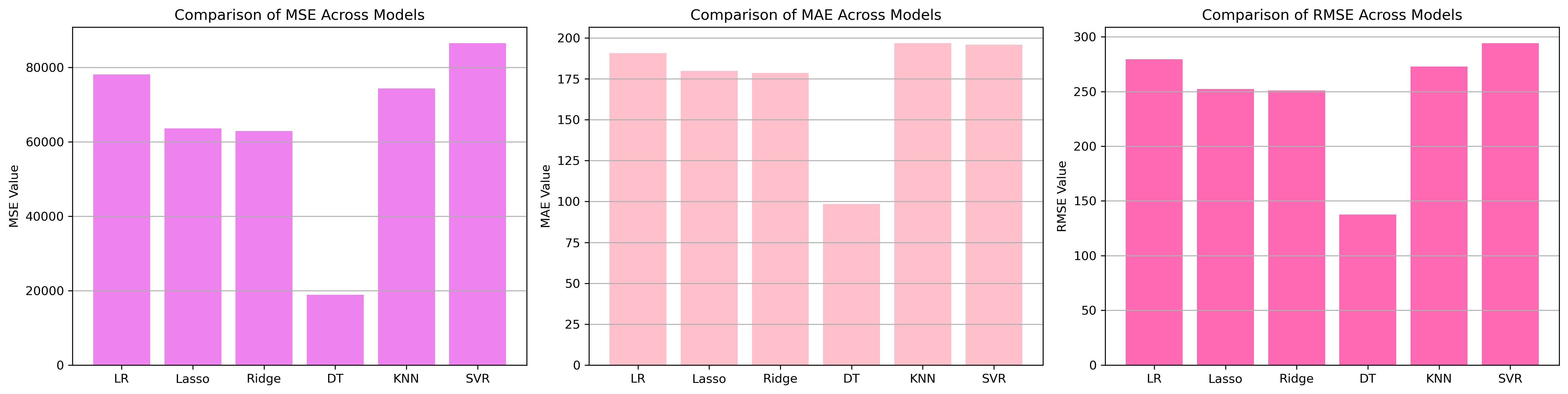
## 5. Cross-Validation Summary (5-Fold)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | RMSE Mean | RMSE Std | MAE Mean | MAE Std | R² Mean | R² Std |
| Decision Tree | 181.06 | 38.18 | 113.59 | 24.43 | 0.9536 | 0.0174 |
| Linear Regression | 248.61 | 45.50 | 180.10 | 21.00 | 0.9116 | 0.0333 |
| Ridge Regression | 259.79 | 45.44 | 187.98 | 25.86 | 0.9049 | 0.0303 |
| Lasso Regression | 262.16 | 47.69 | 189.36 | 27.28 | 0.9029 | 0.0322 |
| KNN | 302.92 | 42.48 | 211.00 | 27.34 | 0.8723 | 0.0312 |
| SVR | 358.53 | 41.30 | 247.24 | 25.61 | 0.8212 | 0.0396 |

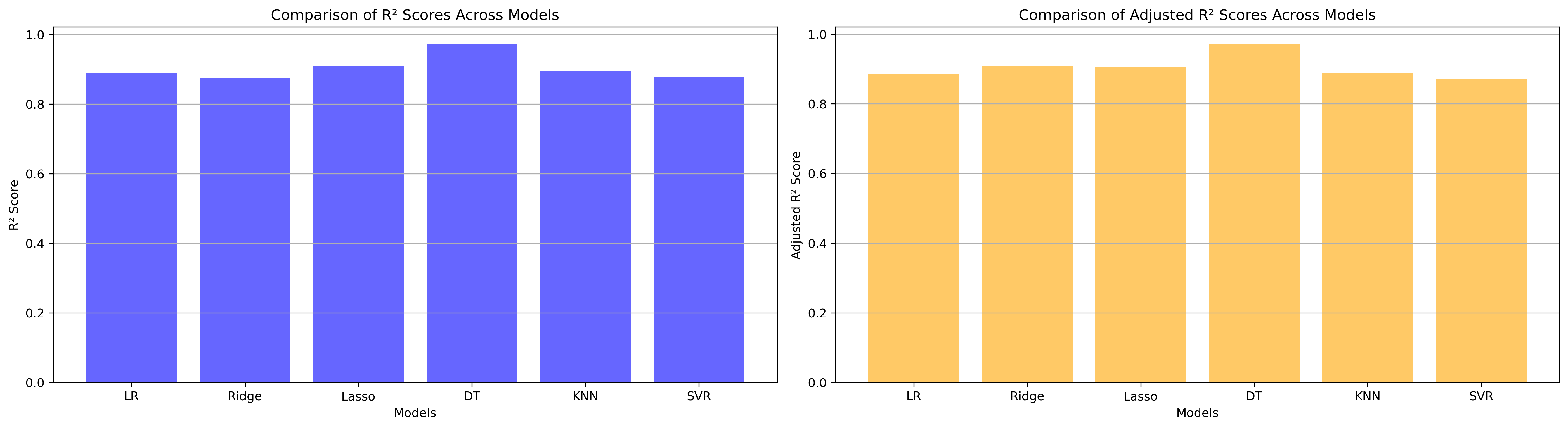
## 6. Visualizations



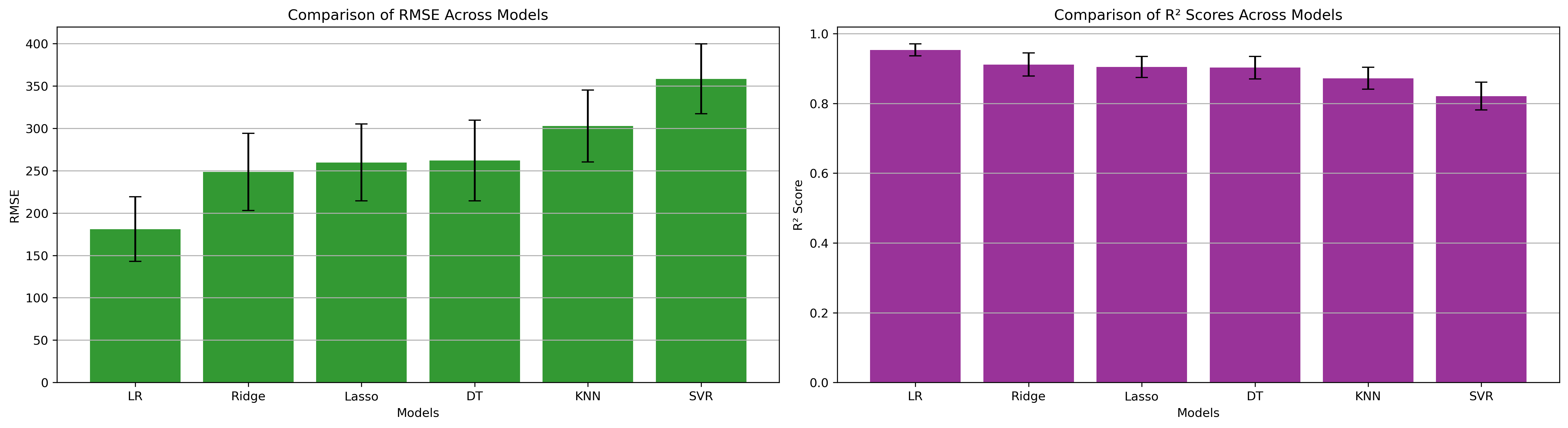
*Figure 1: Actual vs. Predicted Values for Pb (Well 283)*



*Figure 2: Bar Charts of MSE, RMSE, MAE for Pb (Well 283)*



*Figure 3: Bar Charts of R² and Adjusted R² for Pb (Well 283)*



*Figure 4: Error Bars for RMSE and R² from CV for Pb (Well 283)*

### 7. Observations & Next Steps

* **Best Performing Model**: Decision Tree (RMSE: ~137.60, R²: ~0.9734)
  + Decision Tree performed best across all test metrics
  + Regularized Linear Models (Ridge, Lasso) slightly better than Linear
  + SVR was the weakest model

### 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.