# 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: | Bob |
| 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 | SVR Only |
| Feature Selection | None |

## 3. Models Used & Hyperparameters

|  |  |  |
| --- | --- | --- |
| Model | Hyperparameters | Training Time |
| Linear Regression | None | 0.0011 |
| Ridge Regression | alpha = 0.001 | 0.0008 |
| Lasso Regression | alpha = 0.0077 | 0.0006 |
| Decision Tree | max\_depth = 8 | 0.0018 |
| KNN | n\_neighbors = 4 | 0.0009 |
| SVR | C= 0. 7.7426, epsilon= 0.01, kernel = rbf | 0.0805 |

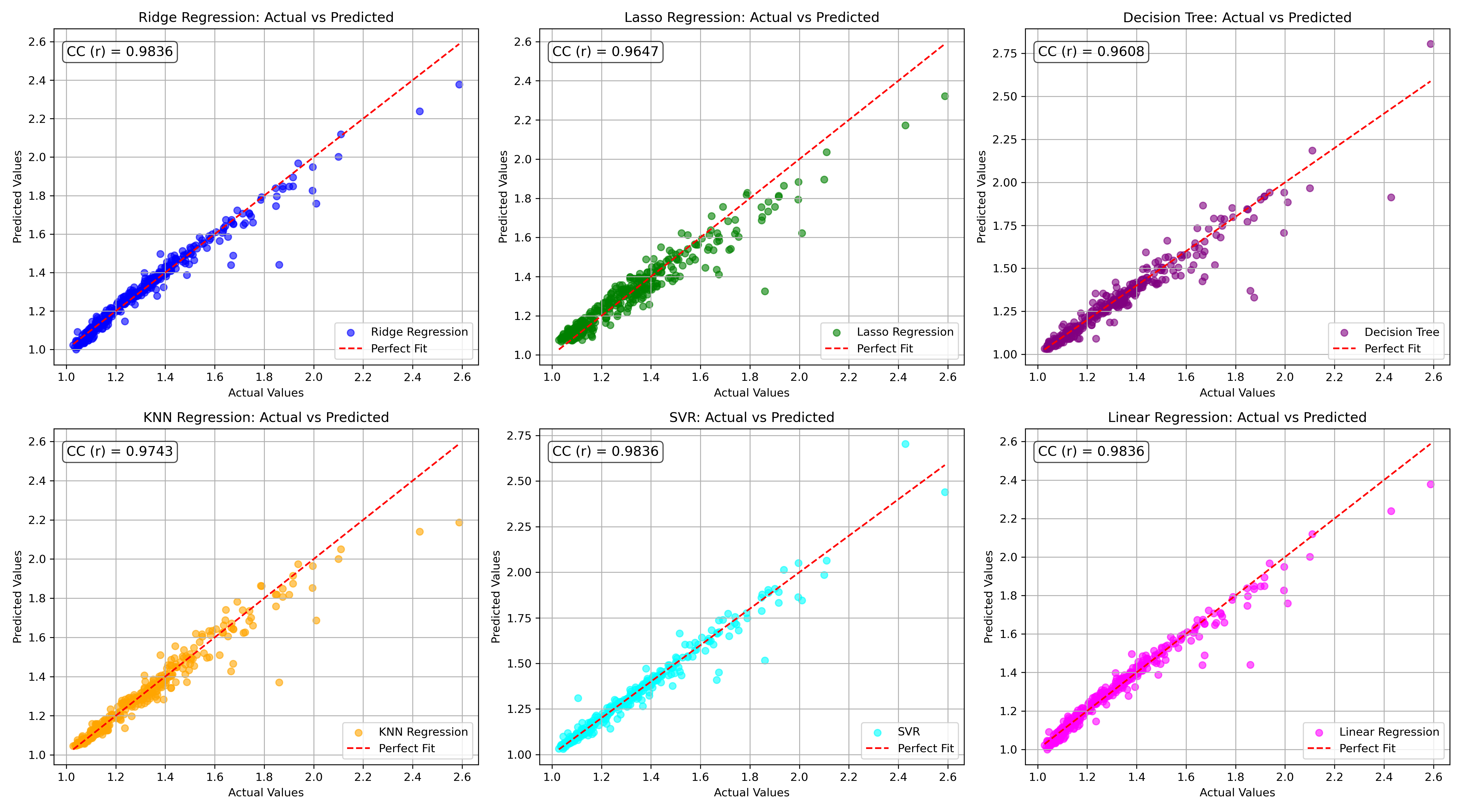
## 4. Evaluation Metrics

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | MSE | RMSE | MAE | R² Score | Adjusted R² |
| Linear Regression | 0.0019 | 0.0440 | 0.0246 | 0.9655 | 0.9651 |
| Lasso Regression | 0.0019 | 0.0440 | 0.0244 | 0.9655 | 0.9651 |
| Ridge Regression | 0.0019 | 0.0440 | 0.0246 | 0.9655 | 0.9651 |
| Decision Tree | 0.0037 | 0.0604 | 0.0317 | 0.9349 | 0.9342 |
| KNN | 0.0023 | 0.0481 | 0.0267 | 0.9588 | 0.9584 |
| SVR | 0.0018 | 0.0492 | 0.0218 | 0.9672 | 0.9669 |

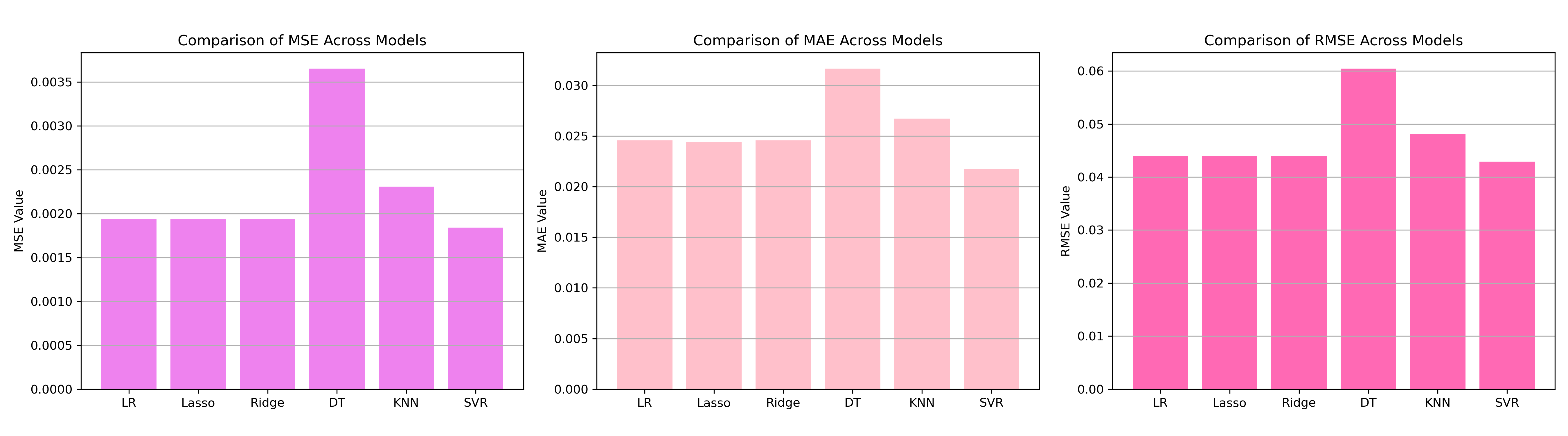
## 5. Cross-Validation Summary (5-Fold)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | RMSE Mean | RMSE Std | MAE Mean | MAE Std | R² Mean | R² Std |
| SVR | 0.044324 | 0.008102 | 0.021951 | 0.002342 | 0.966228 | 0.011101 |
| Lasso Regression | 0.05214 | 0.01212 | 0.02728 | 0.00144 | 0.9539 | 0.01545 |
| Ridge Regression | 0.05216 | 0.01207 | 0.02742 | 0.00141 | 0.9539 | 0.01534 |
| Linear Regression | 0.05216 | 0.01207 | 0.02742 | 0.00141 | 0.9539 | 0.01534 |
| KNN | 0.05452 | 0.00935 | 0.02843 | 0.00231 | 0.9496 | 0.01306 |
| Decision Tree | 0.05980 | 0.00793 | 0.03019 | 0.00271 | 0.9387 | 0.01560 |

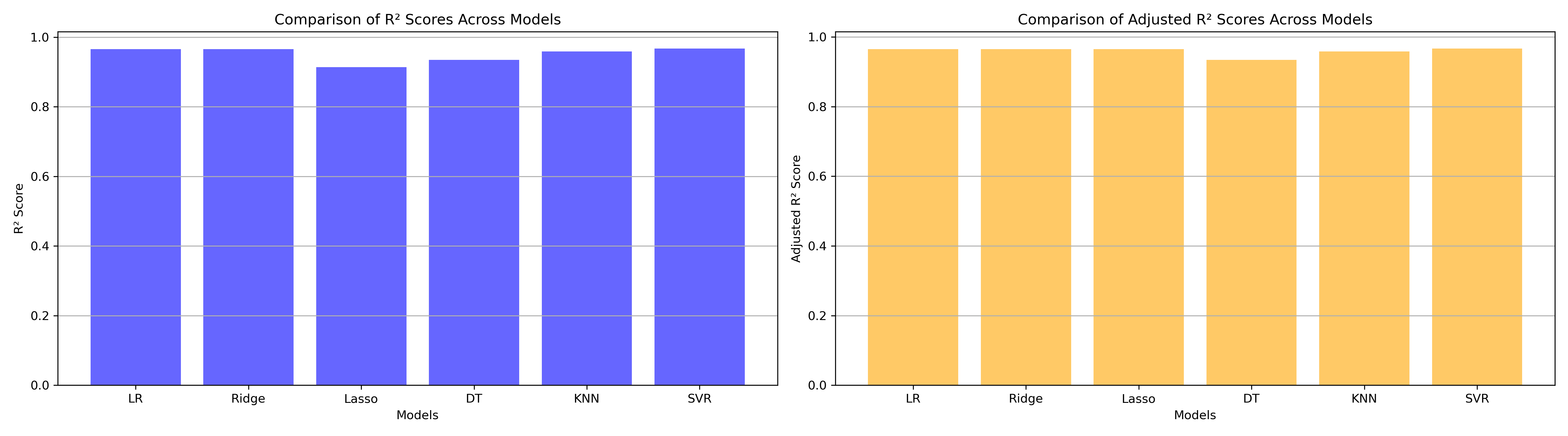
## 6. Visualizations



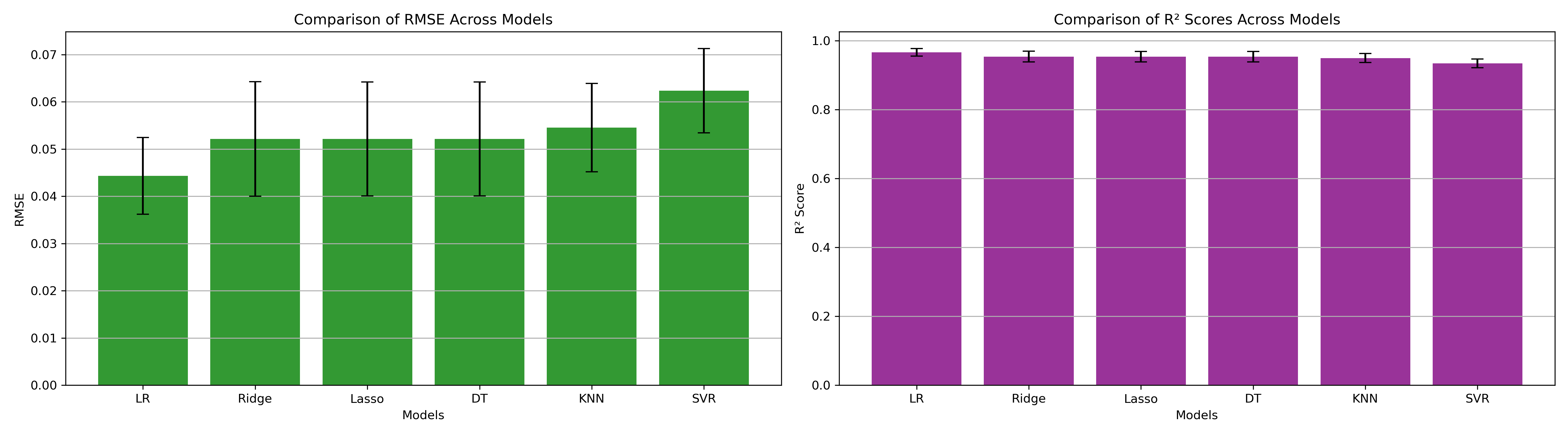
*Figure 1: Actual vs. Predicted Values for Bob (Well 1225)*



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



*Figure 3: Bar Charts of R² and Adjusted R² for Bob (Well 1225)*



*Figure 4: Error Bars for RMSE and R² from CV for Bob (Well 1225)*

### 7. Observations & Next Steps

* **Best Performing Model**: **SVR** (RMSE: ~0.0443, R²: ~0.9662)
  + SVR outperformed all other models across all metrics
  + Lasso, Ridge, and Linear Regression performed similarly and closely behind
  + KNN and Decision Tree performed moderately well

### 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.