# Machine Learning Model Tracking Document

## 1. Dataset Information

|  |  |
| --- | --- |
| Dataset Name: | Well 782 |
| Number of Samples: | 782 |
| 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 | 0.0009 |
| Ridge Regression | alpha = 10.0 | 0.0006 |
| Lasso Regression | alpha = 0.0599 | 0.0006 |
| Decision Tree | max\_depth = 16 | 0.0016 |
| KNN | n\_neighbors = 7 | 0.0010 |
| SVR | C = 0.1668, epslion: 0.01, kernel = rbf | 0.0061 |

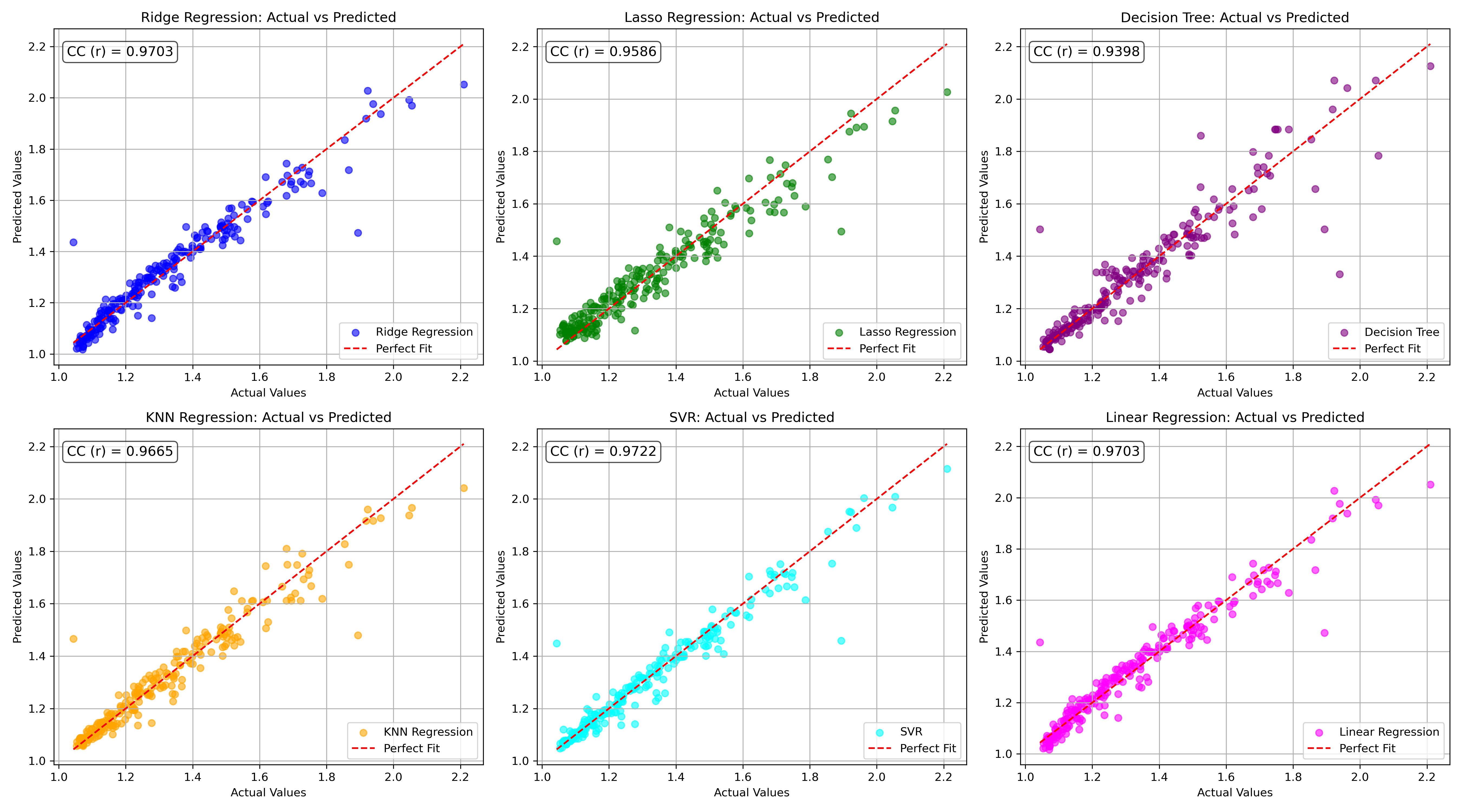
## 4. Evaluation Metrics

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | MSE | RMSE | MAE | R² Score | Adjusted R² |
| Linear Regression | 0.0031 | 0.0556 | 0.0346 | 0.9414 | 0.9404 |
| Lasso Regression | 0.0030 | 0.0549 | 0.0327 | 0.9428 | 0.9419 |
| Ridge Regression | 0.0031 | 0.0555 | 0.0345 | 0.9415 | 0.9404 |
| Decision Tree | 0.0063 | 0.0793 | 0.0410 | 0.8807 | 0.8786 |
| KNN | 0.0032 | 0.0567 | 0.0322 | 0.9390 | 0.9379 |
| SVR | 0.0030 | 0.0545 | 0.0292 | 0.9437 | 0.9428 |

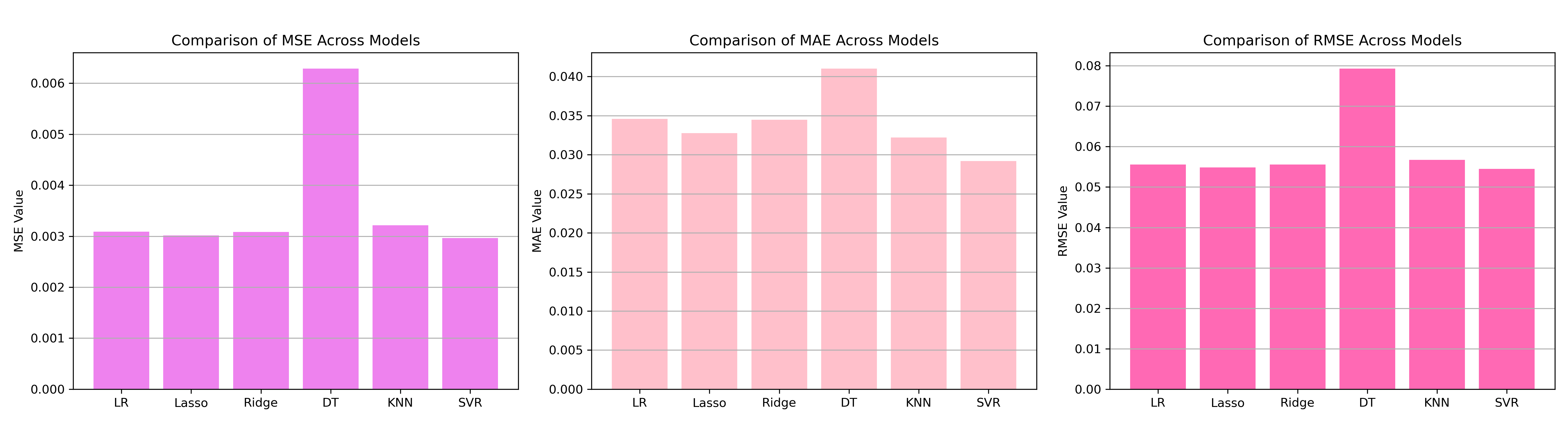
## 5. Cross-Validation Summary (5-Fold)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | RMSE Mean | RMSE Std | MAE Mean | MAE Std | R² Mean | R² Std |
| Ridge Regression | 0.0635 | 0.0182 | 0.0342 | 0.0043 | 0.9403 | 0.0276 |
| Linear Regression | 0.0636 | 0.0182 | 0.0343 | 0.0043 | 0.9402 | 0.0276 |
| Lasso Regression | 0.0638 | 0.0192 | 0.0335 | 0.0051 | 0.9398 | 0.0292 |
| KNN | 0.0708 | 0.0231 | 0.0374 | 0.0077 | 0.9306 | 0.0215 |
| SVR | 0.0735 | 0.0400 | 0.0354 | 0.0099 | 0.9223 | 0.0586 |
| Decision Tree | 0.0918 | 0.0387 | 0.0449 | 0.0095 | 0.8821 | 0.0590 |

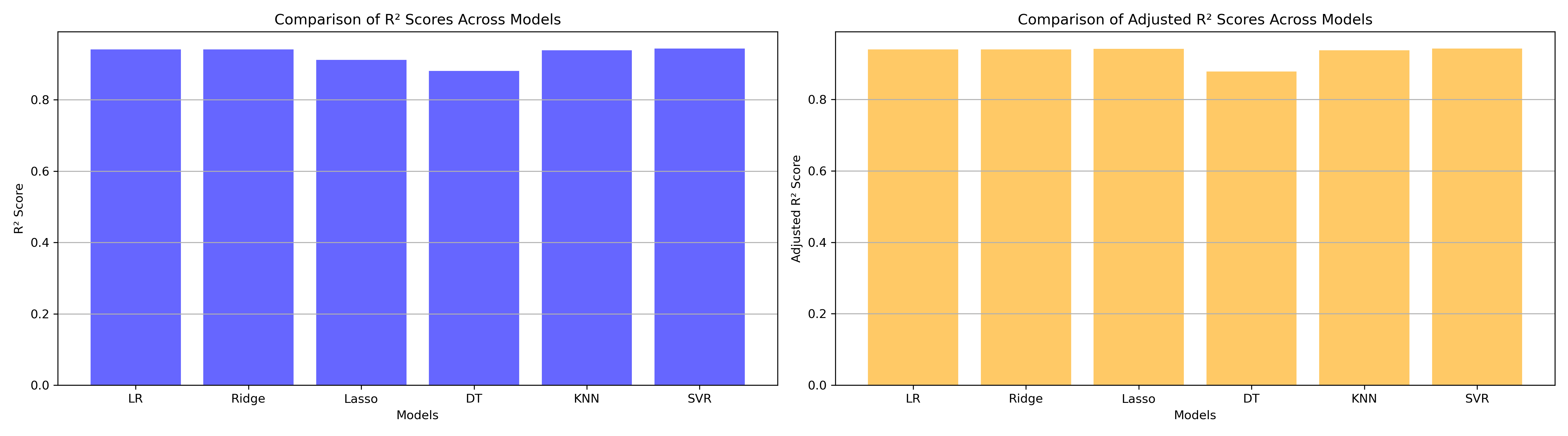
## 6. Visualizations



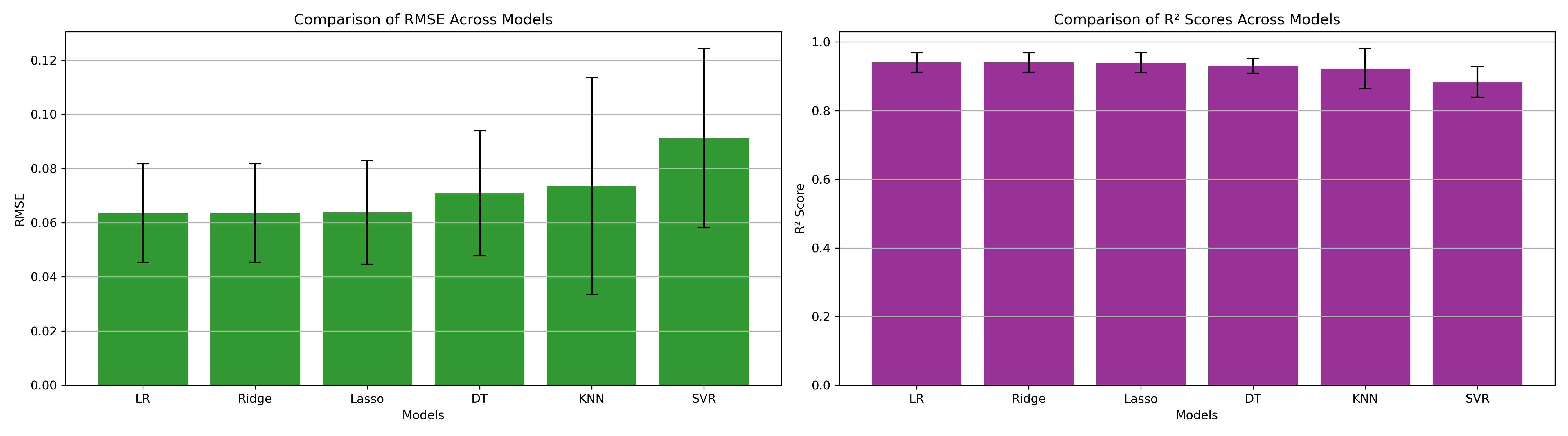
*Figure 1: Actual vs. Predicted Values for Bob (Well 782)*



*Figure 2: Bar Charts of MSE, RMSE, MAE for Bob (Well 782)*



*Figure 3: Bar Charts of R² and Adjusted R² for Bob (Well 782)*



*Figure 4: Error Bars for RMSE and R² from CV for Bob (Well 782)*

### 7. Observations & Next Steps

* **Best Performing Model: Ridge** (RMSE: ~0.0635, R²: ~0.9403)
  + Ridge, Linear, and Lasso models performed nearly identically with R² ≈ 0.94
  + SVR showed the highest variance and the lowest R² (~0.9224), underperforming the linear models
  + KNN and Decision Tree had moderate to lower performance, with Decision Tree being the weakest overall

### 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.