# 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: | 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 = 0.167 | 0.0005 |
| Lasso Regression | alpha = 0.001 | 0.0018 |
| Decision Tree | max\_depth = 12 | 0.0006 |
| KNN | n\_neighbors = 4 | 0.0006 |
| SVR | C= 27.825, epslion= 0.01, kernel = rbf | 0.0029 |

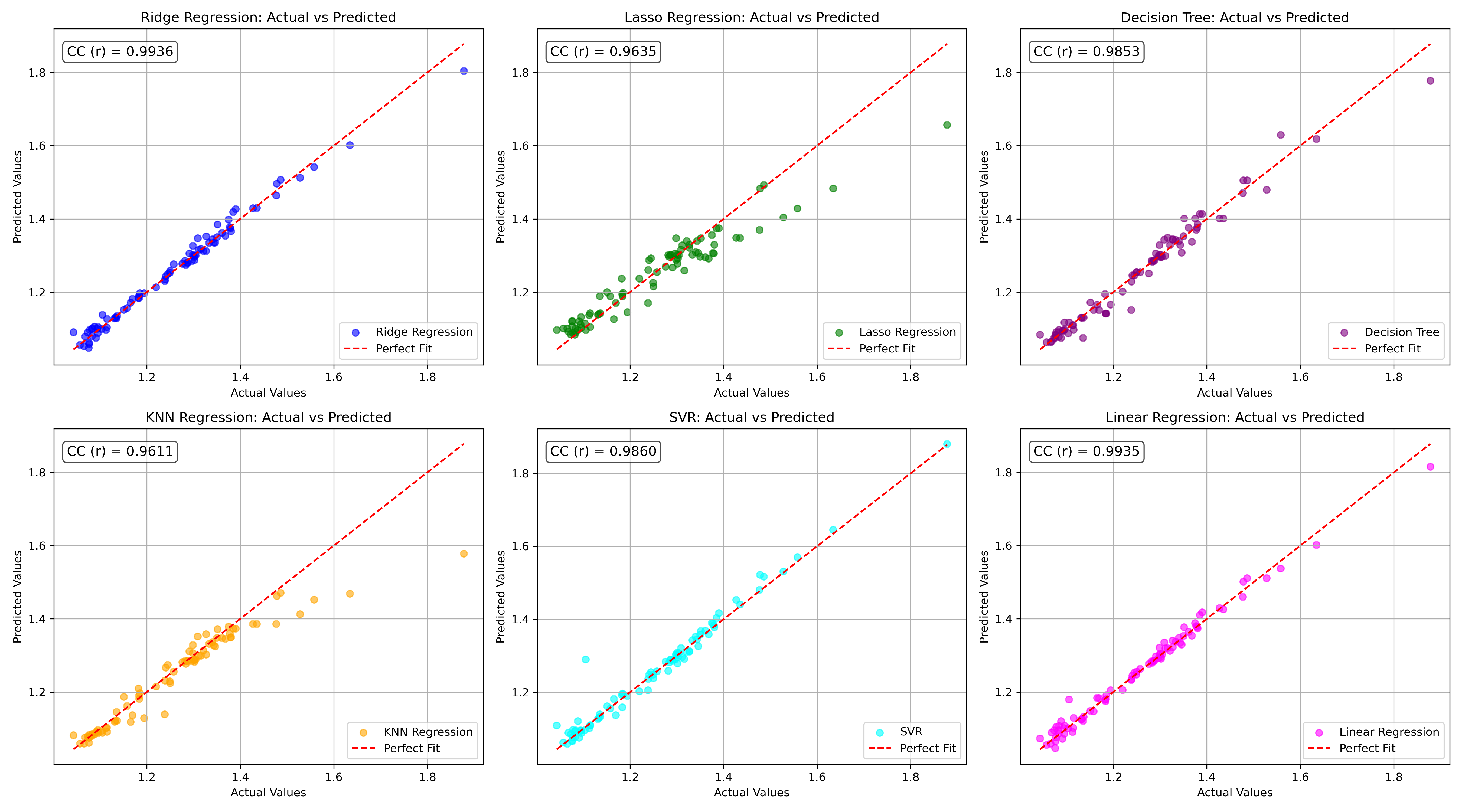
## 4. Evaluation Metrics

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | MSE | RMSE | MAE | R² Score | Adjusted R² |
| Linear Regression | 0.0003 | 0.0168 | 0.0140 | 0.9934 | 0.9927 |
| Lasso Regression | 0.0003 | 0.0184 | 0.0151 | 0.9920 | 0.9913 |
| Ridge Regression | 0.0003 | 0.0167 | 0.0139 | 0.9934 | 0.9928 |
| Decision Tree | 0.0013 | 0.0360 | 0.0252 | 0.9694 | 0.9666 |
| KNN | 0.0005 | 0.0218 | 0.0164 | 0.9888 | 0.9878 |
| SVR | 0.0006§ | 0.0255 | 0.0137 | 0.9707 | 0.9693 |

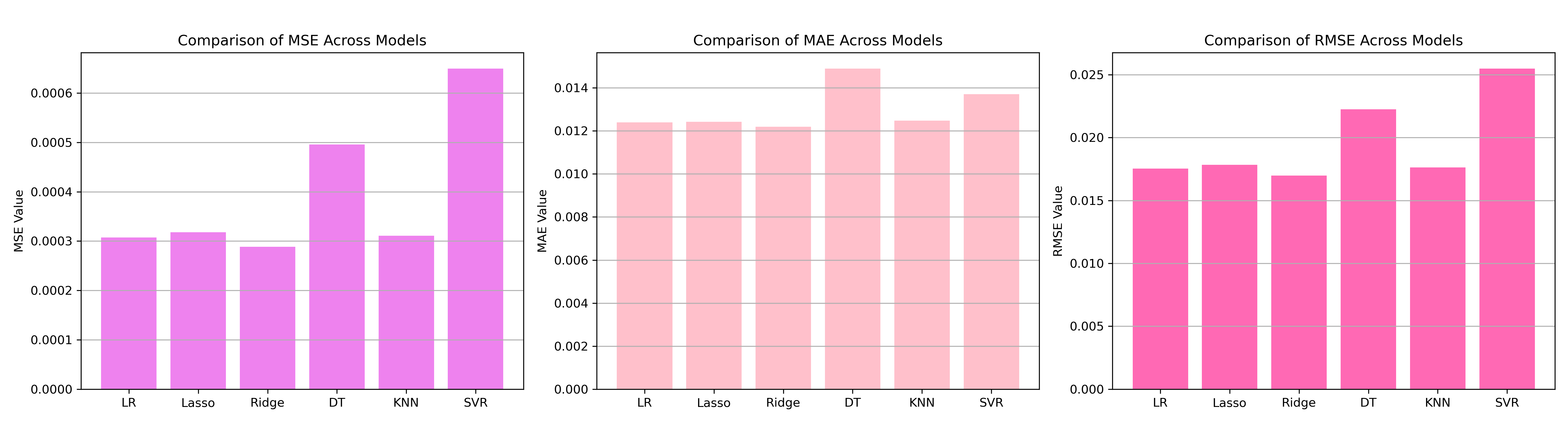
## 5. Cross-Validation Summary (5-Fold)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | RMSE Mean | RMSE Std | MAE Mean | MAE Std | R² Mean | R² Std |
| SVR | 0.011691 | 0.001527 | 0.009008 | 0.001096 | 0.994291 | 0.000790 |
| Ridge Regression | 0.01948 | 0.00407 | 0.01474 | 0.00184 | 0.9905 | 0.00273 |
| Linear Regression | 0.01950 | 0.00414 | 0.01472 | 0.00190 | 0.9906 | 0.00270 |
| Lasso Regression | 0.02084 | 0.00367 | 0.01600 | 0.00171 | 0.9889 | 0.00371 |
| KNN | 0.02509 | 0.00496 | 0.01824 | 0.00418 | 0.9840 | 0.00547 |
| Decision Tree | 0.03050 | 0.00918 | 0.02210 | 0.00413 | 0.9754 | 0.01463 |

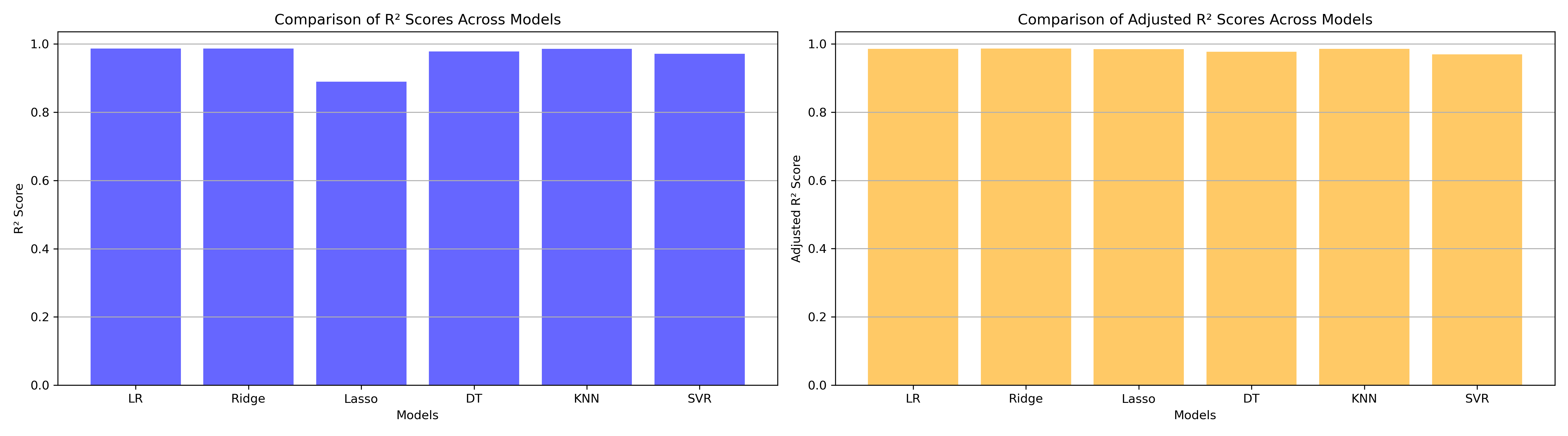
## 6. Visualizations



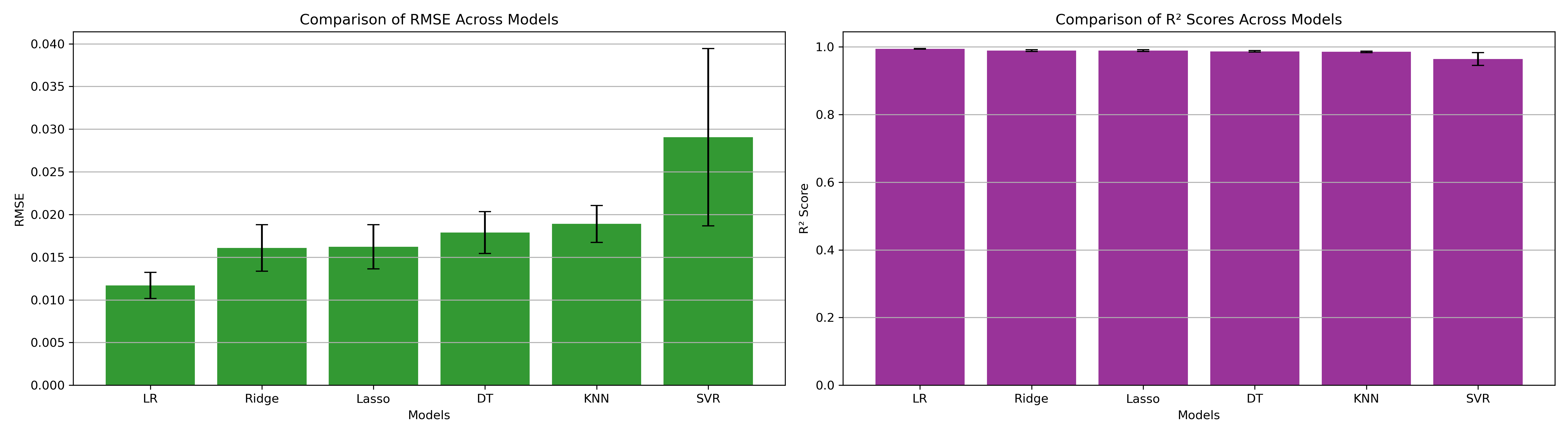
*Figure 1: Actual vs. Predicted Values for Bob (Well 160)*



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



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



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

### 7. Observations & Next Steps

* **Best Performing Model**: SVR (RMSE: ~0.0152, R²: ~0.9944)
  + SVR significantly outperformed all other models with the highest accuracy
  + Ridge, Linear Regression, and Lasso followed closely with strong but slightly lower performance
  + KNN and Decision Tree models performed moderately well compared to the others

### 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.