# 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: | Bob |
| Outlier Handling: | None |
| Feature Engineering Applied: | None |
| Scaling/Normalization Applied: | Only for Neural Networks |
| 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 for Neural Networks |
| Feature Selection | None |

## 3. Models Used & Hyperparameters

|  |  |  |
| --- | --- | --- |
| Model | Hyperparameters | Training Time |
| **XGBoost** | n\_estimators=100, learning\_rate=0.05, max\_depth=6 | 0.0225 |
| **CatBoost** | iterations=1000, learning\_rate=0.05, depth=4, l2\_leaf\_reg=5 | 0.1251 |
| **Neural Network** | [64, 32], epochs=100, batch\_size=16 | 2.6758 |
| **Stacking Ensemble** | Default base models + CatBoost final estimator | 0.5619 |
| **Extra Trees** | n\_estimators=200 | 0.0708 |
| **Deep Neural Network** | [256, 128, 64], epochs=150, batch\_size=32 | 2.8945 |

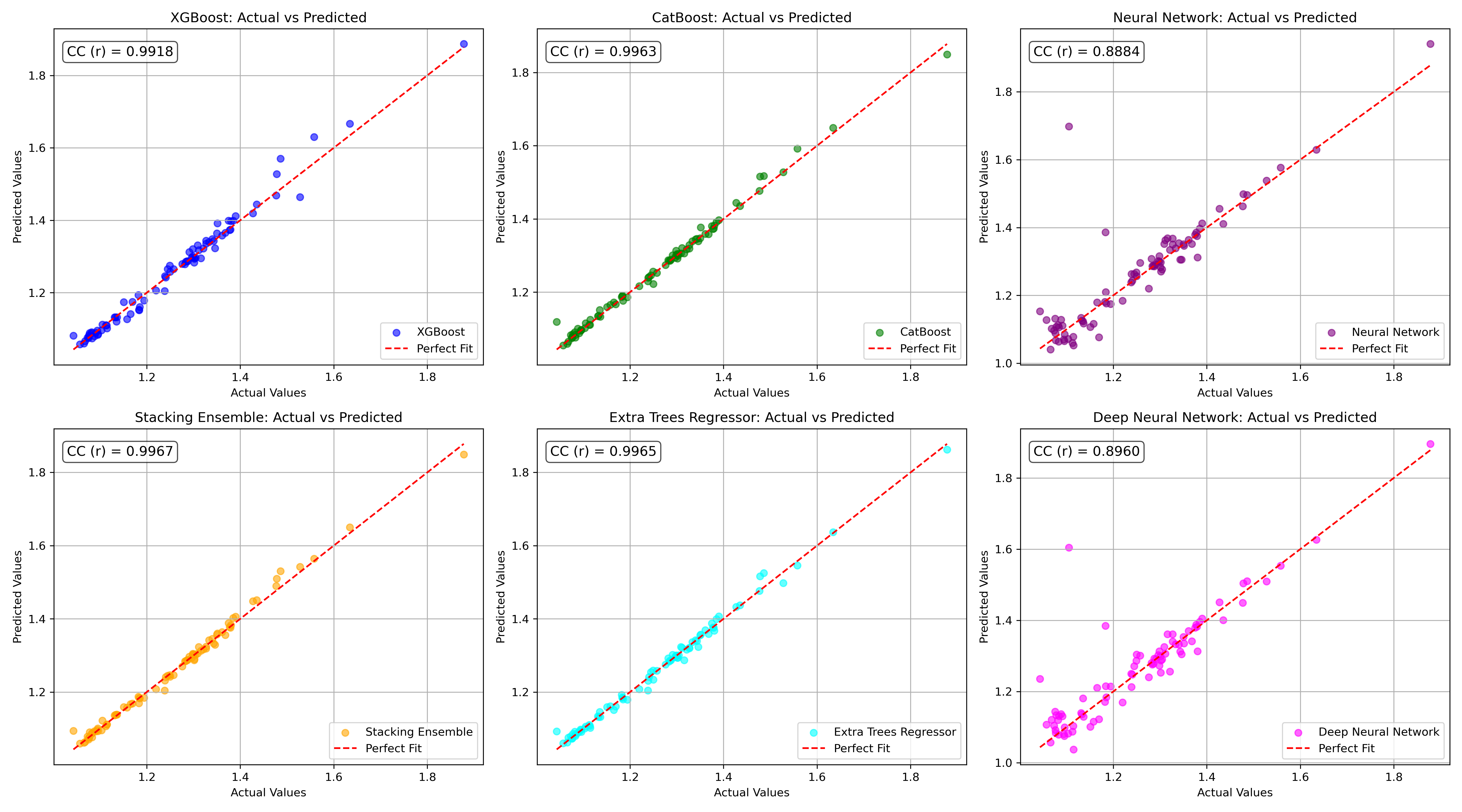
## 4. Evaluation Metrics

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | MSE | RMSE | MAE | R² Score | Adjusted R² |
| **XGBoost** | 0.0003 | 0.0186 | 0.0135 | 0.9845 | 0.9837 |
| **CatBoost** | 0.0002 | 0.0124 | 0.0076 | 0.9931 | 0.9928 |
| **Neural Network** | 0.0043 | 0.0654 | 0.0392 | 0.8071 | 0.7975 |
| **Stacking Ensemble** | 0.0001 | 0.0118 | 0.0078 | 0.9937 | 0.9934 |
| **Extra Trees** | 0.0002 | 0.0125 | 0.0083 | 0.9929 | 0.9926 |
| **Deep Neural Network** | 0.0045 | 0.0671 | 0.0323 | 0.7972 | 0.7870 |

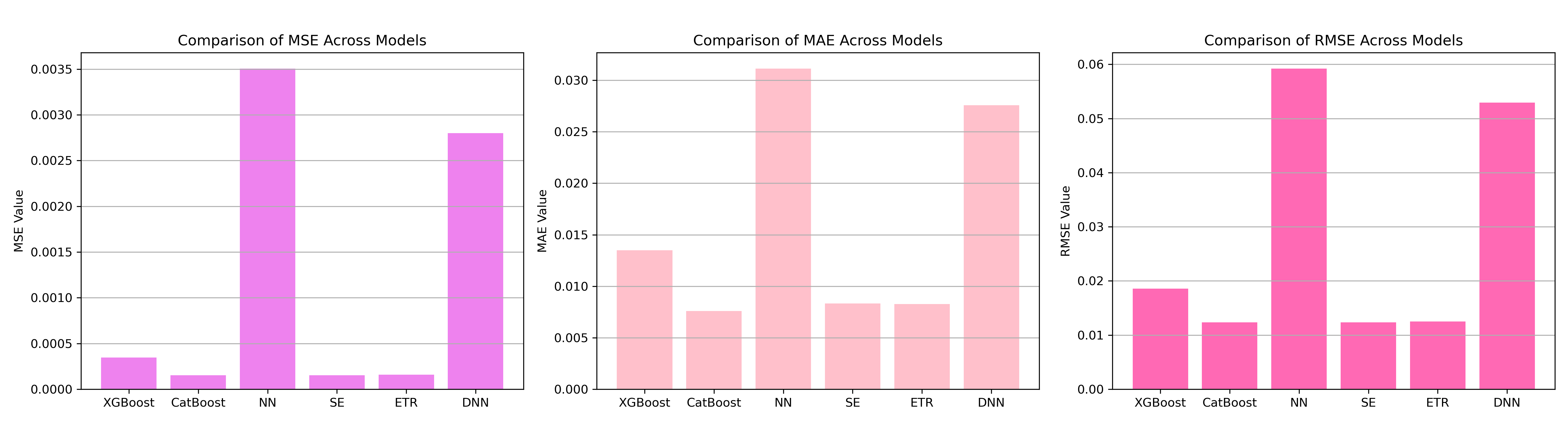
## 5. Cross-Validation Summary (5-Fold)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | RMSE Mean | RMSE Std | MAE Mean | MAE Std | R² Mean | R² Std |
| **Stacking Ensemble** | 0.0107 | 0.0016 | 0.0069 | 0.0009 | 0.9953 | 0.0011 |
| **Extra Trees** | 0.0118 | 0.0015 | 0.0078 | 0.0008 | 0.9942 | 0.0012 |
| **CatBoost** | 0.0138 | 0.0061 | 0.0075 | 0.0026 | 0.9915 | 0.0063 |
| **XGBoost** | 0.0170 | 0.0029 | 0.0116 | 0.0014 | 0.9882 | 0.0020 |
| **Deep Neural Network** | 0.1366 | 0.0279 | 0.1027 | 0.0253 | 0.2138 | 0.2457 |
| **Neural Network** | 0.1633 | 0.0397 | 0.1195 | 0.0250 | -0.1857 | 0.5130 |

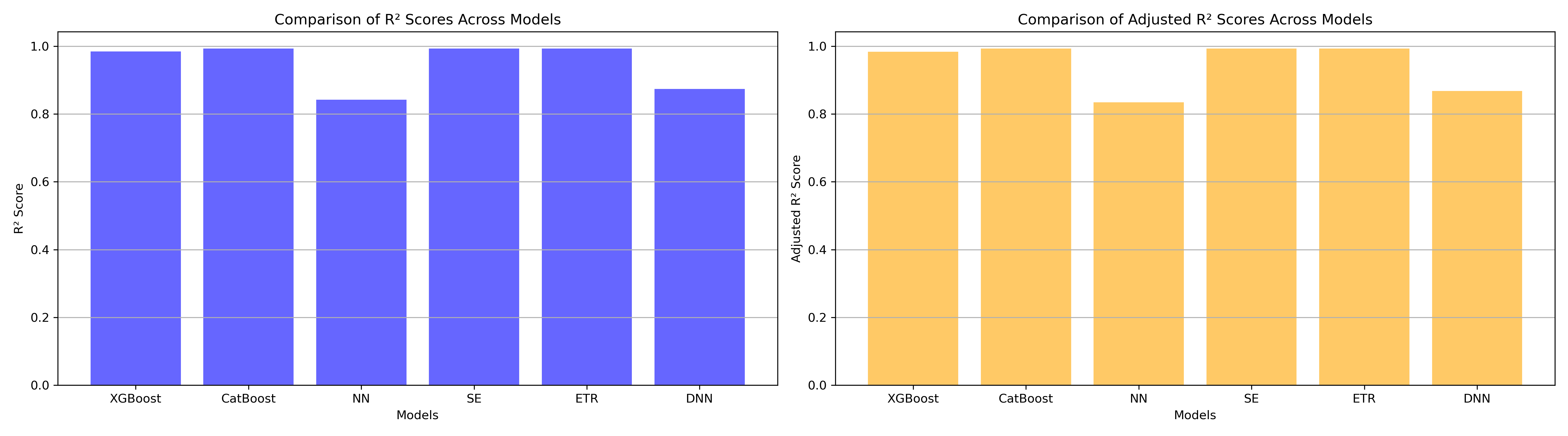
## 6. Visualizations



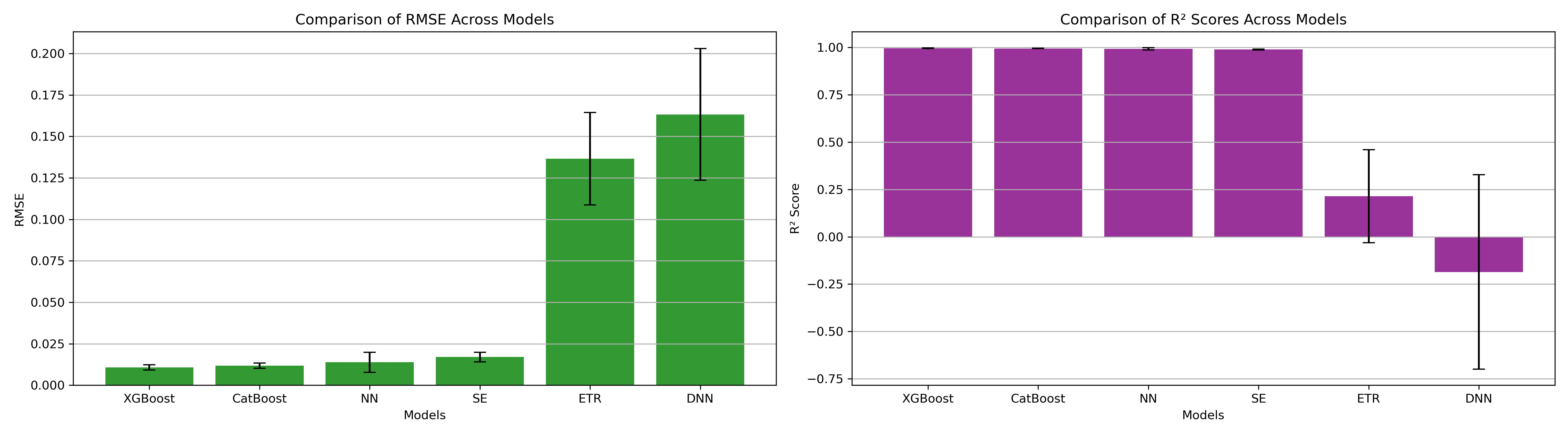
*Figure 1: Actual vs. Predicted Values for Bob (Well 283)*



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



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



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

### 7. Observations & Next Steps

**Best Performing Model:** S**tacking Ensemble** (RMSE: 0.0118, R²: 0.9937)

* CatBoost and Extra Trees also performed competitively with RMSEs below 0.013.
* Neural-based models underperformed, with R² below 0.81.
* All models except the neural-based ones achieved **R² > 0.98**.

### 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.