Oil Production Forecasting Using Machine Learning

IMTIYAZ ALI SYED(2220056)

AKRAM MOHAMMAD(2162967)

FAISAL MALIK MOHAMMED(2214828)

MAHAMMAD SALMAN SHAIK(2215886)

Abstract :

- Using the Volve field dataset and various machine learning methods, forecast oil production.
- Selecting the best machine learning model by comparing them according to their performance measures.

Introduction

- "Can we use machine learning algorithms to accurately predict oil production in the Volve field, and if so, which algorithm performs the best?"
- Accurate prediction of oil production will enable companies to optimize production, reduce downtime, and increase profitability.

DataSet:

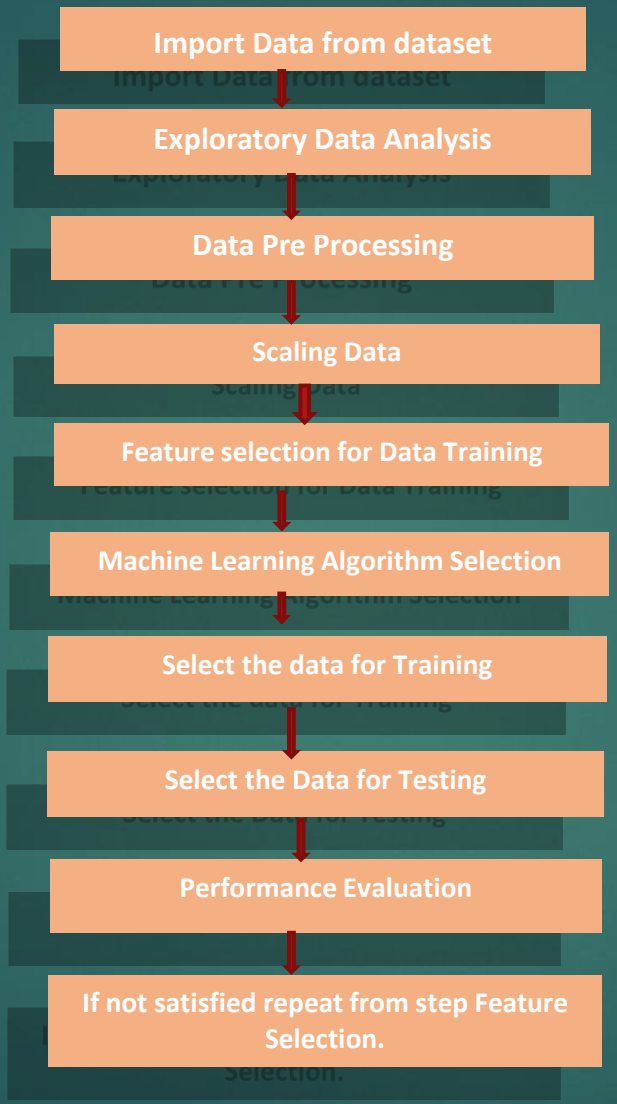
- The Volve Field Production dataset is a high-quality and comprehensive dataset containing production data from the Volve oil field operated by Equinor from 2008 to 2016, and is widely used for studying production optimization and forecasting in the petroleum industry.

- The Volve Field Production dataset is of good quality but covers only a few years of production, with possible missing values and outliers that need careful handling during analysis.

Methodology:

- **Linear Regression**: Predicts continuous numerical values based on input features by fitting a straight line to the data points.
- **Polynomial Regression**: Allows for non-linear relationships between input and output variables by fitting a polynomial function to the data points.
- **XGBoost Regression**: Uses gradient boosting to create an ensemble of decision trees for regression tasks, with each tree trained to correct the errors of the previous tree.
- **Random Forest**: Uses decision trees to create a predictive model by building a large number of trees on randomly sampled subsets of input data and aggregating their predictions.

Work Flow:


Import Data from dataset
Exploratory Data Analysis
Data Pre Processing
Scaling Data
Feature selection for Data Training
Machine Learning Algorithm Selection
Select the data for Training
Select the Data for Testing
Performance Evaluation
If not satisfied repeat from step Feature Selection.

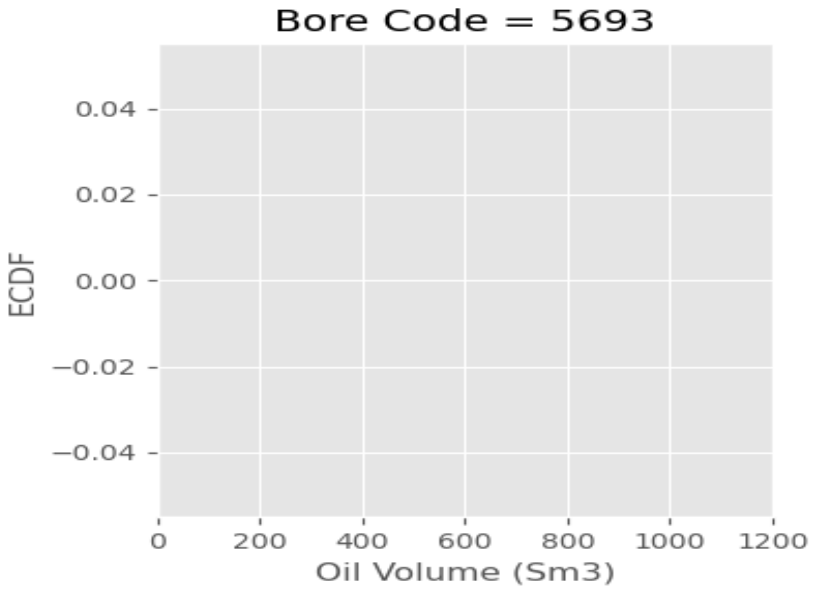
Exploratory Data Analytics:

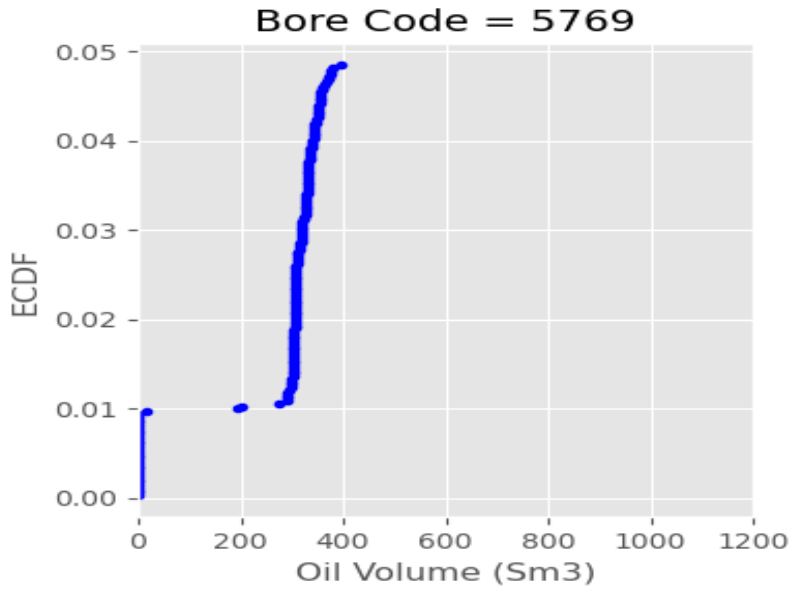
- Check the shape and columns of the data and the number of null values in each column.
- Generate summary statistics of the data, including count, mean, standard deviation, minimum, and maximum values.
- Calculate the correlation between the different columns of the data

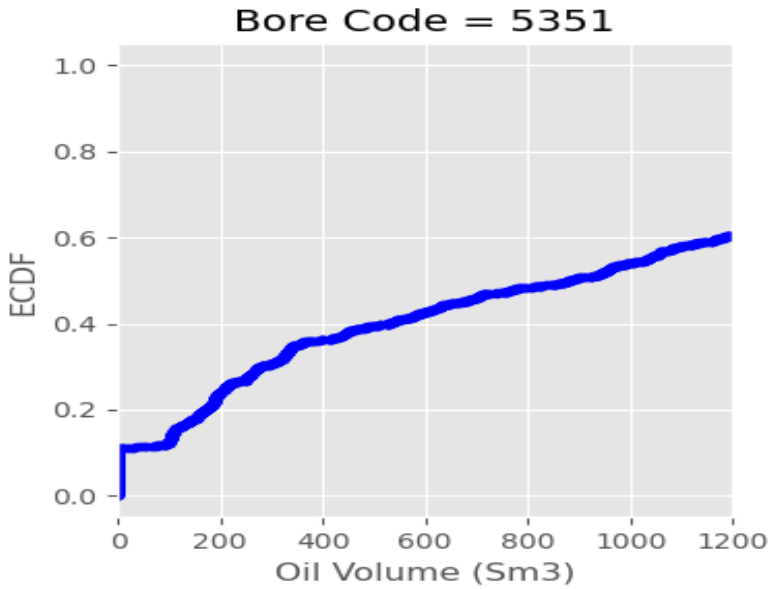
Exploratory Data Analytics(2):

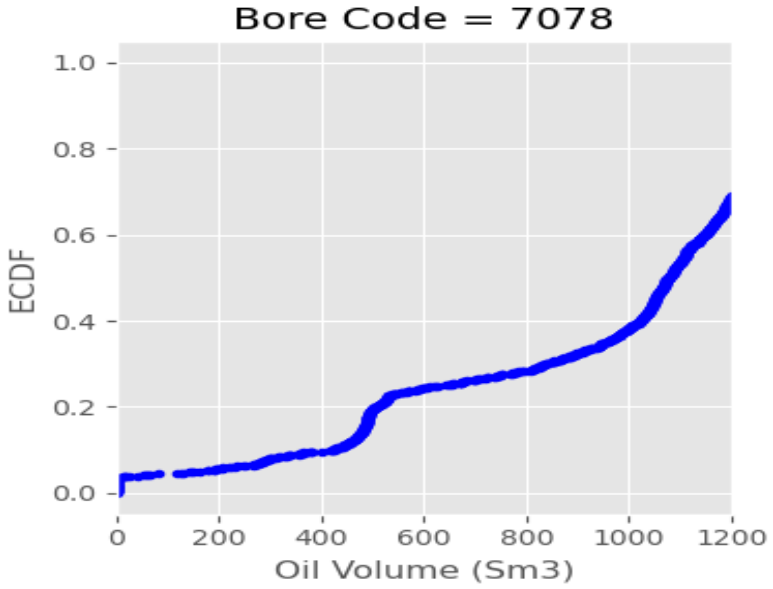
- Using the missingno library to visualize the missing data in the dataset.
- Filtering the data by selecting rows using the 'NPD_WELL_BORE_CODE' to get individual well data.
- Using **Empirical Cumulative Distribution Function** to plot the Oil Production from all wells in order from least to greatest and see the distribution of Oil Production per well.
- Draw the **boxplots** of features for each wellbore for properly understanding the statistics of the dataset, knowing the distribution of features.

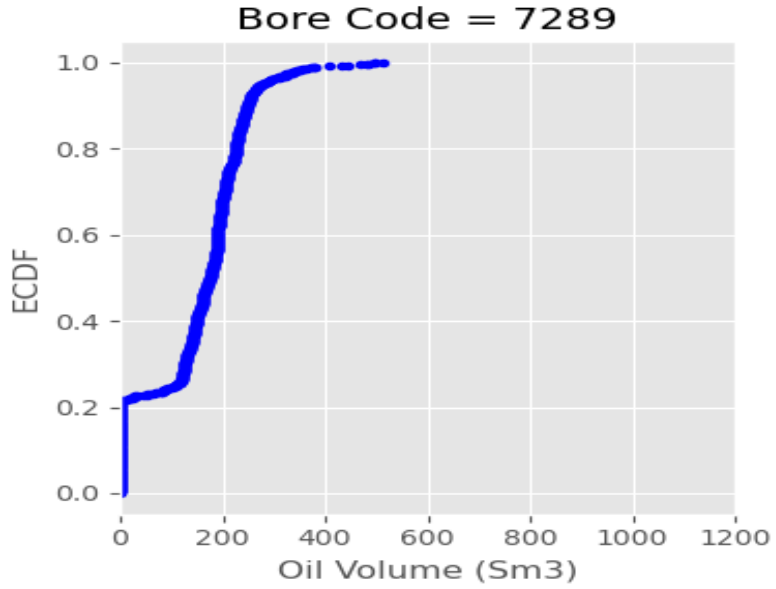
Comparison of Wells using ECDF

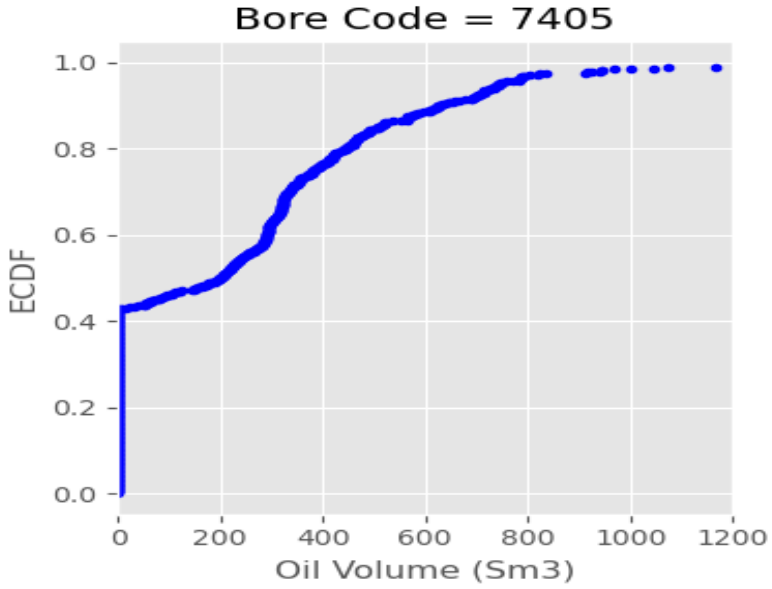












Data Preprocessing and Scaling.

- Using the ECDF function choosing the wells that are appropriate for Machine Learning.
- Handling the skewed values using **Interpolation-linear** method to fill in the missing/NAN values.
- Using **MINMAX** scaler to scale numeric columns to a common range without distorting differences in values or losing information.

Feature Selection:

- Feature selection involves automatically or manually choosing the relevant features that contribute to accurate predictions.
- Generate the Heat map and understand the correlation between different columns .
- Select the appropriate columns that can be used for Training.

Model Building:

- Model building involves selecting and implementing a machine learning algorithm to predict the target variable, oil production.

- It includes selecting the algorithm, identifying relevant features, initializing and training the model, evaluating its performance, and refining it if necessary.

- The goal is to find the best model that accurately predicts oil production based on available data.

Training, Testing and Prediction :

- Model training involves adjusting the model's parameters to minimize the difference between its predicted output and the actual output using input data and output data.

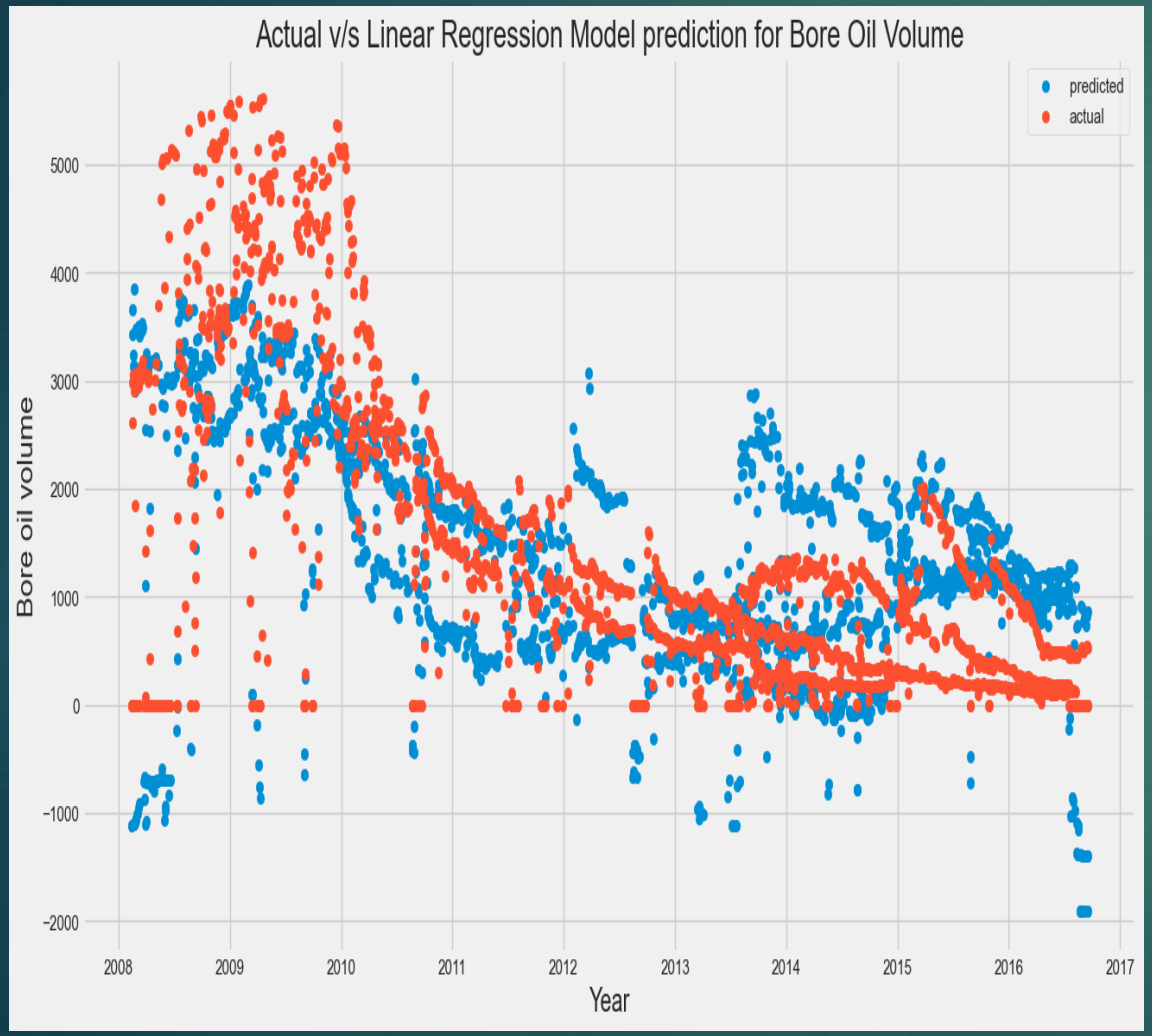
- The model is evaluated using a separate set of data that was not used during training, called the testing set, to assess its performance.

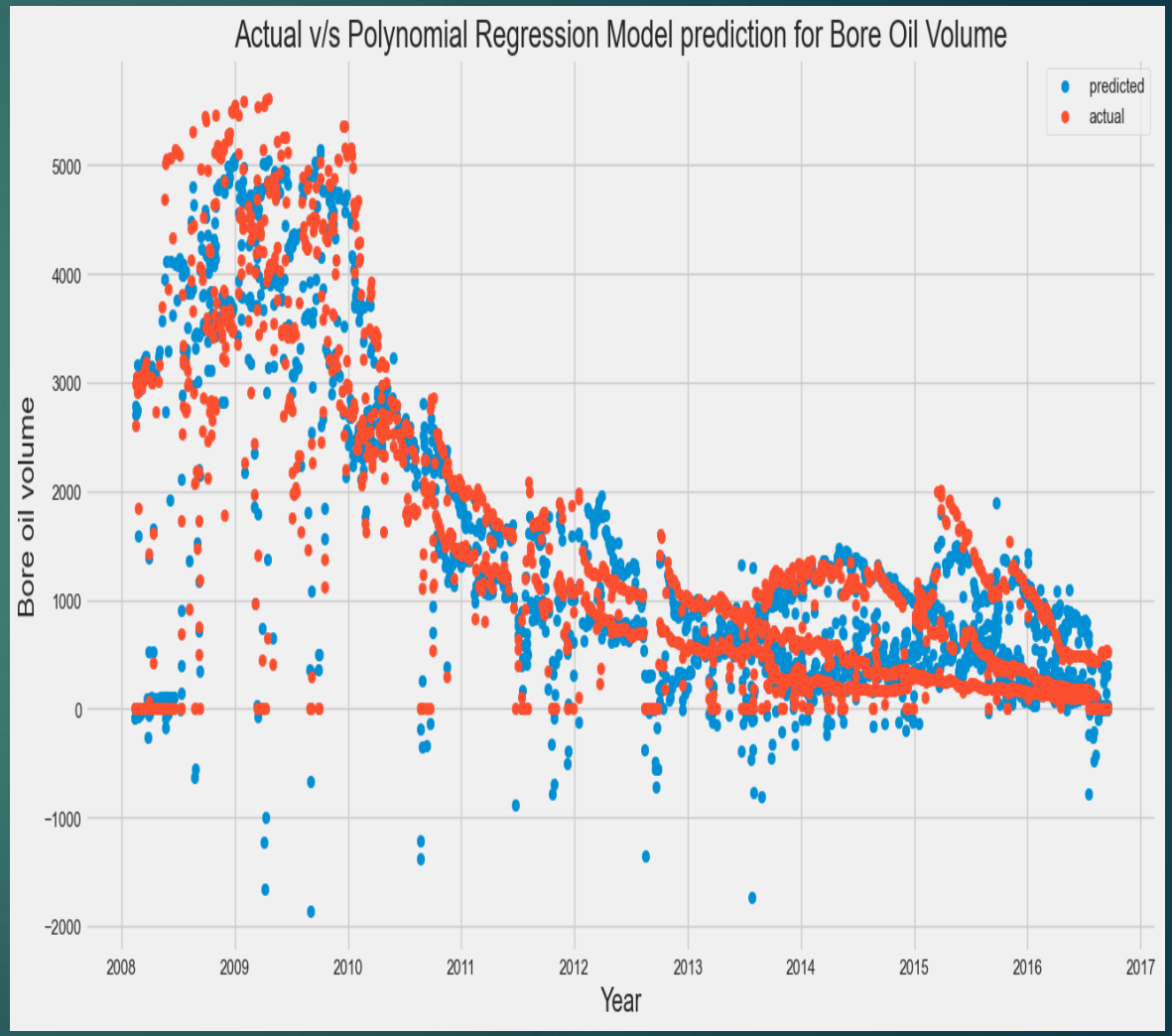
- Once the model is trained and evaluated, it can be used for prediction on new, unseen data to forecast future oil production.

Applying Machine Learning and calculating the performance metrics

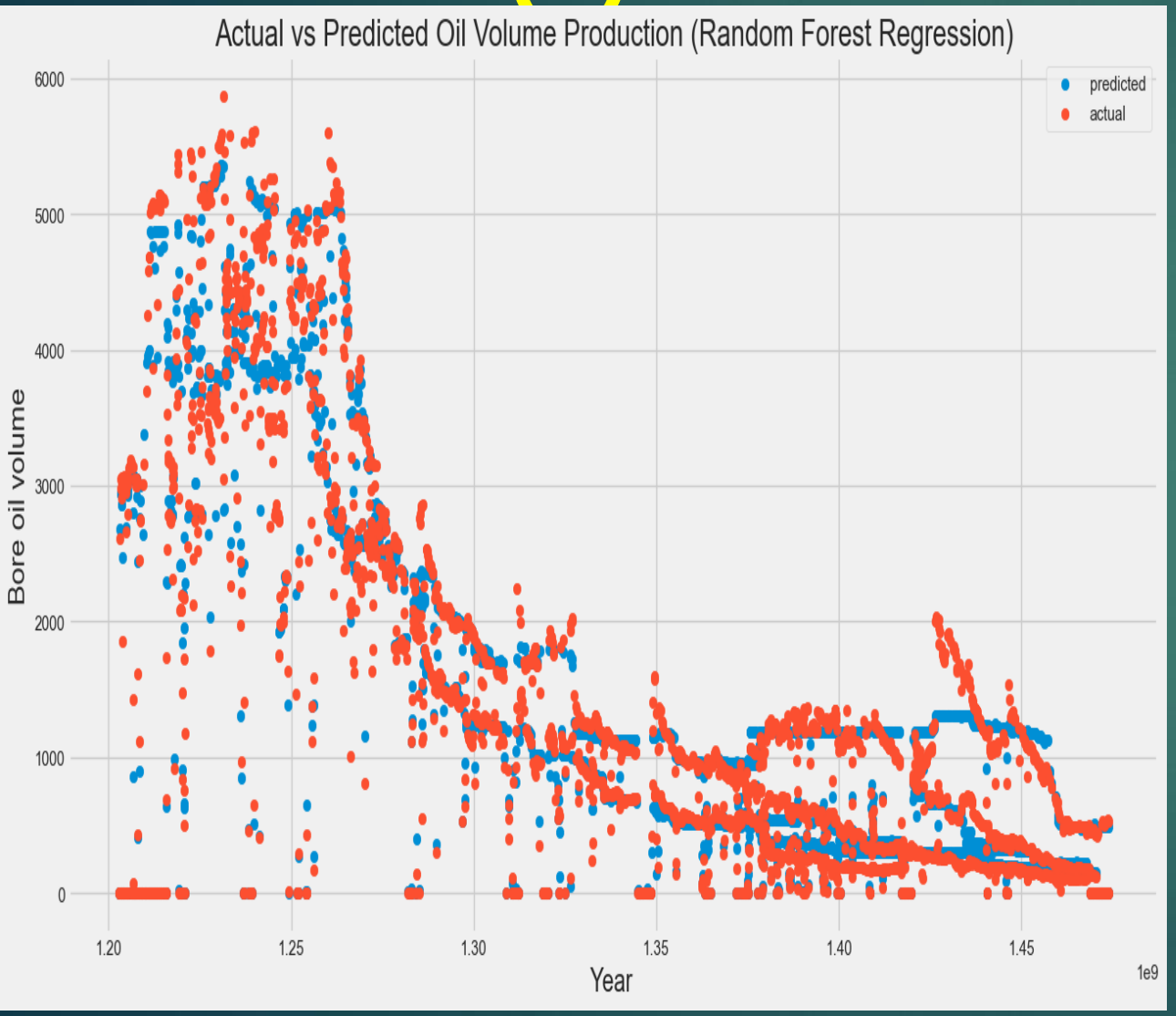
- The Machine Learning Models are applied on the wells using the columns chosen during the feature scaling to predict the Oil Productions.
- Plots are drawn for Actual vs Predicted Values to understand the performance.
- Performance Metrics like R2,MAE,MSE,RMSE are evaluated to find out the best model.

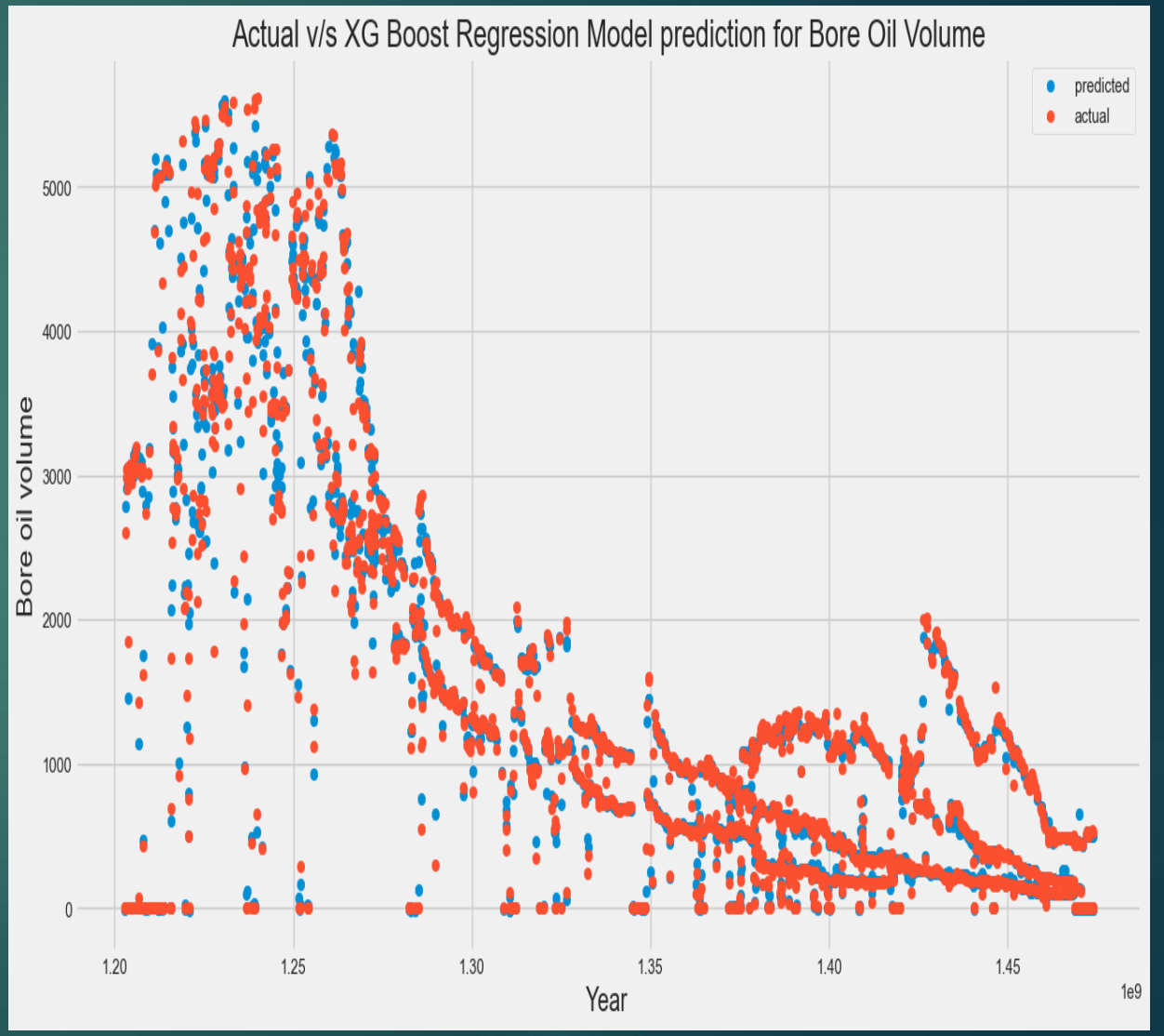
Actual vs Predicted values using different models



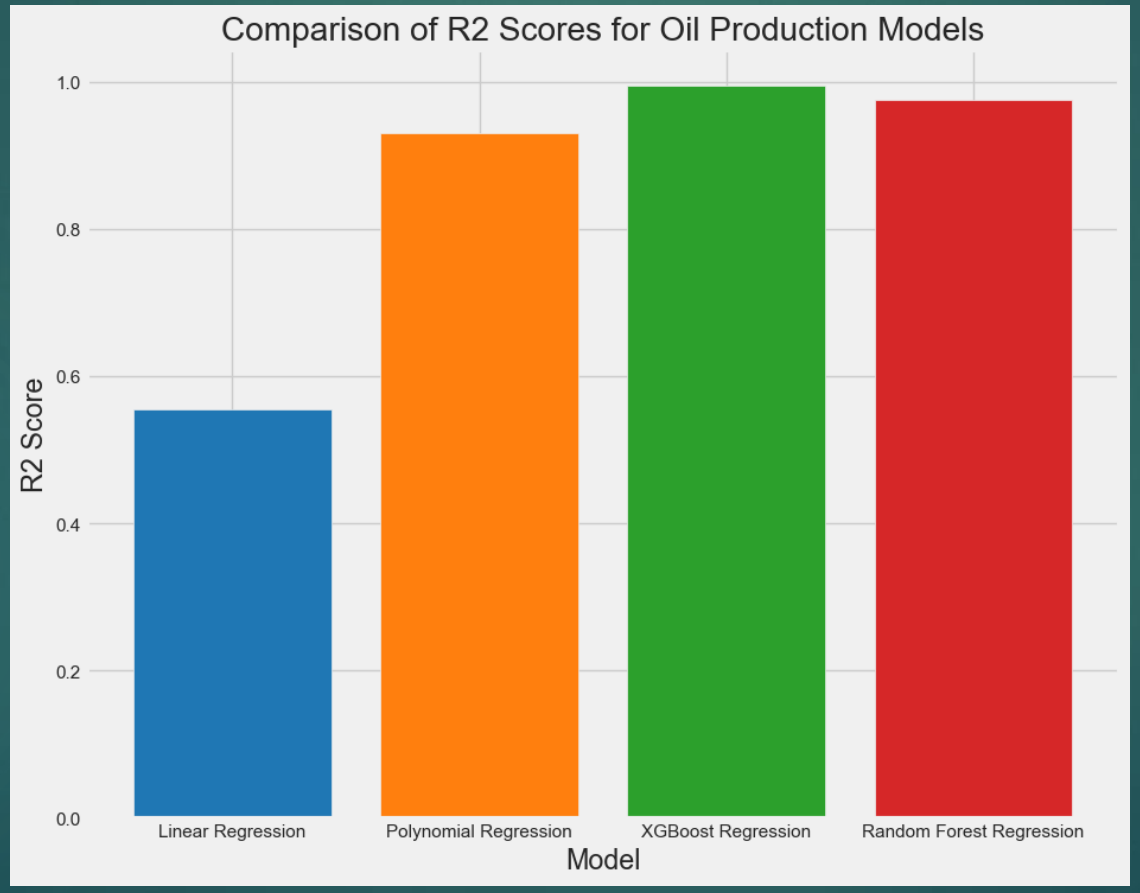


Actual vs Predicted values using different models(2)

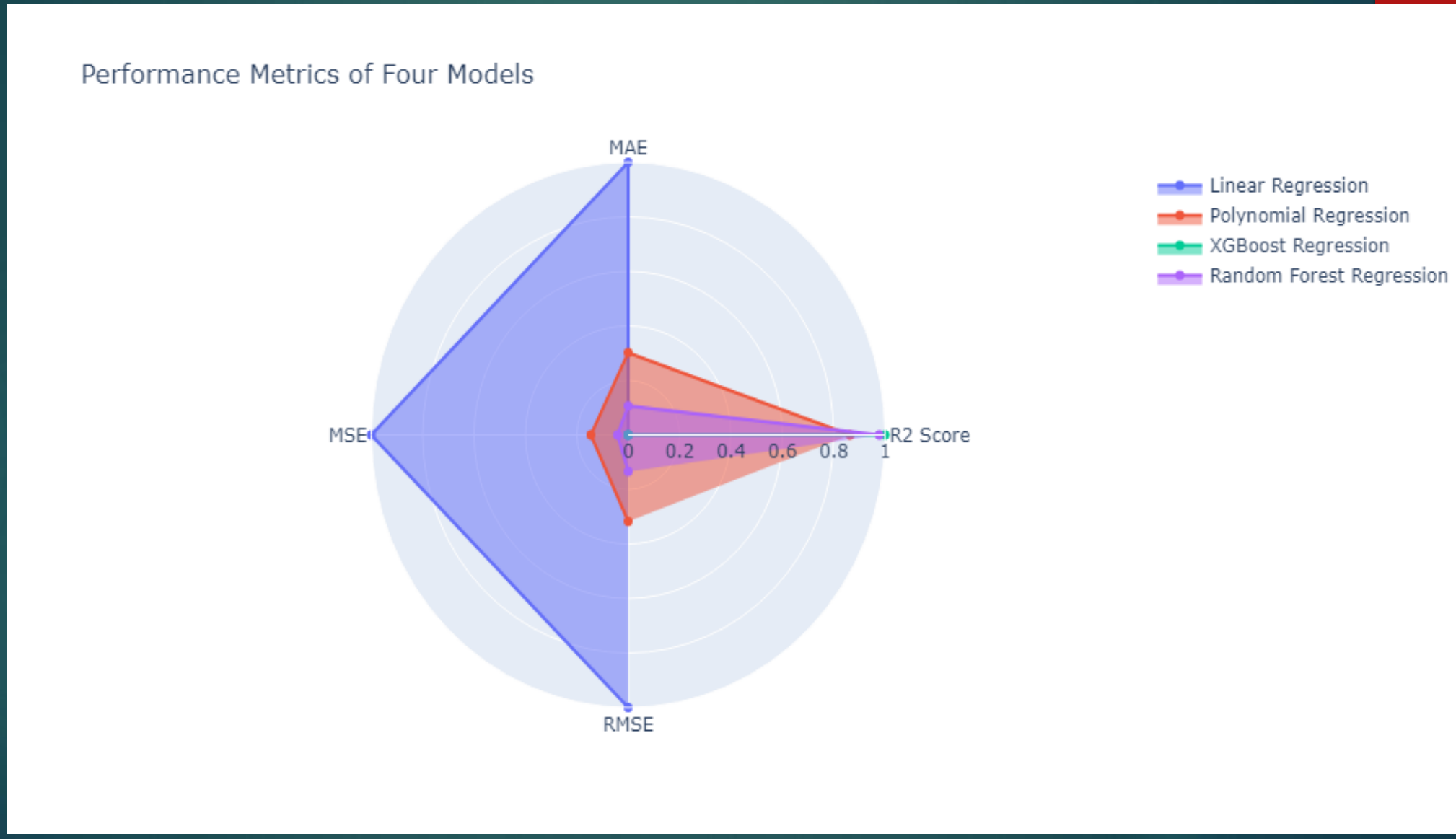




Comparison of R2 squared Values



Performance Metrics of Four Models



Results:

- Using R-squared (R2) as a performance metric to evaluate the models.“
- Also Compared other performance metrics like Mean Absolute Error, Mean Square Error and Root Mean Square Error.
- Results indicate that XGBoost Regression and Random Forest models gave better performance metrics than Linear and Polynomial Regression models.

Conclusion and Recommendations:

- The XGBoost and Random Forest models, in particular, show promise for future use in predicting oil production volume.
- However, it should be noted that the accuracy of the models is dependent on the quality of the data used. Incomplete or inaccurate data can lead to inaccurate predictions.
- The project could improve by adding more data features like well completion design, production history, and reservoir properties

References:

- Chen, J. C., et al. "Prediction of oil production using machine learning algorithms." Journal of Petroleum Science and Engineering 184 (2019): 106448.
- Lu, Z., & Yuan, J. (2020). Machine learning models for predicting oil production: A comprehensive review. Renewable and Sustainable Energy Reviews, 117, 109527.
- Singh, S. K., Kumar, M., & Kumar, D. (2019). A review on machine learning techniques for predicting oil production. Journal of Petroleum Exploration and Production Technology, 9(1), 1-14

Thank You!