Predictive Analytics for Oil Well Failures: A Machine Learning Approach

Presented by:
Titilola Oduwole & Jahir Gutierrez

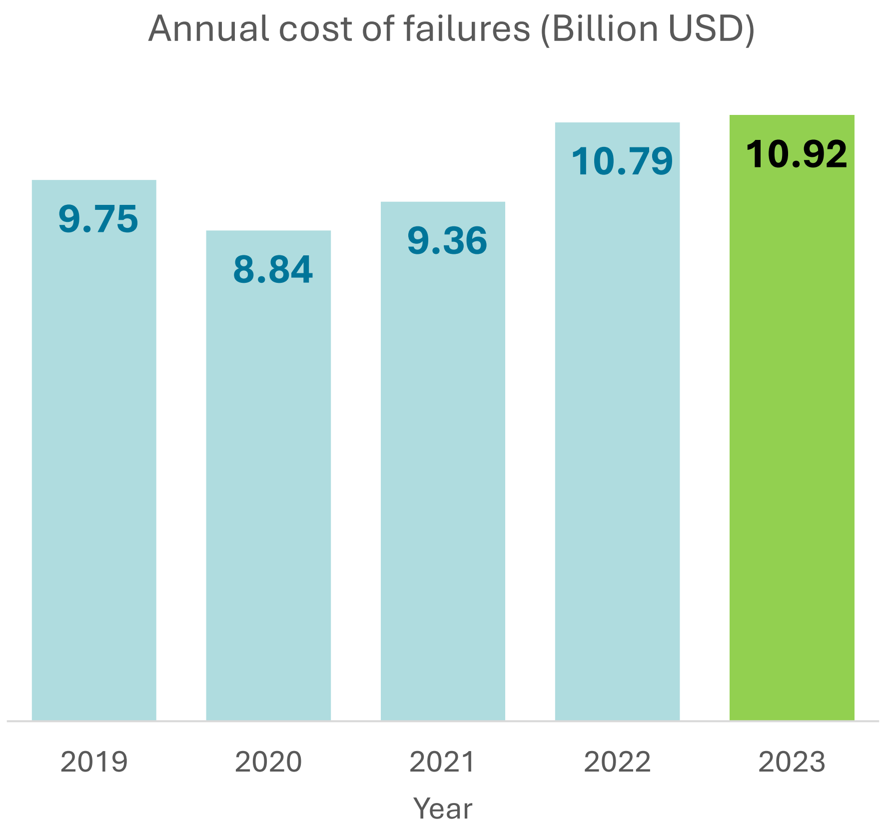
Date: June 2024

NBCC
IT: Business and Advanced Analytics

Table of Contents

NBCC

Introduction The Rationale Behind This Project.



KUSD 130

is the average cost per well to repair a downhole failure (excluding lost production).

The true cost of well failures impacts the ***environment, safety, and your bottom line***.

Problem Statement

To develop a Machine Learning Model to predict failures in oil wells

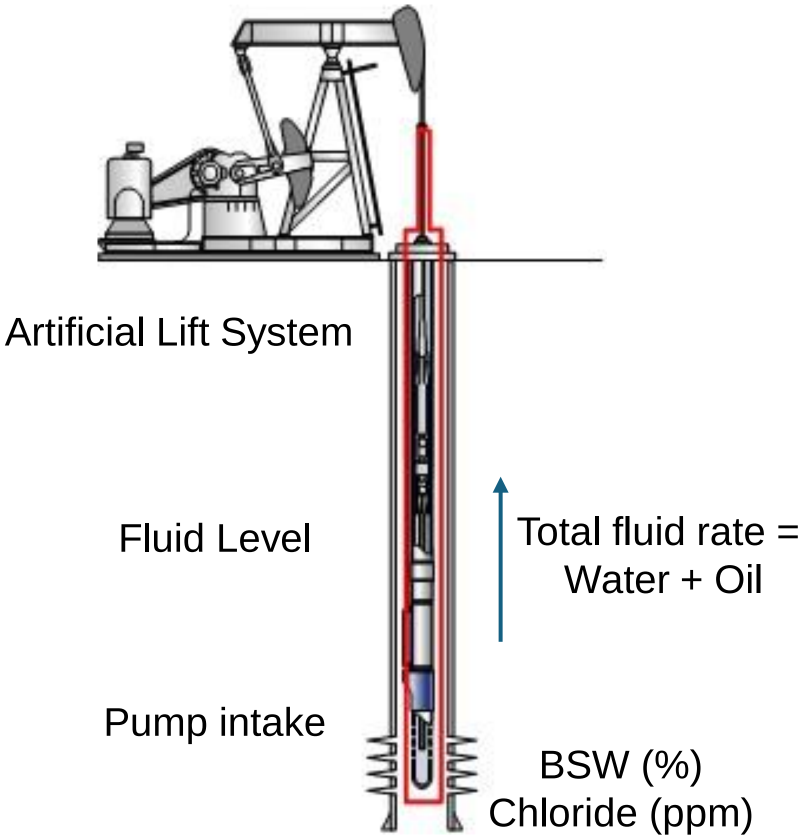
Output:
An actionable prediction ("Intervention Needed" or "No Intervention")

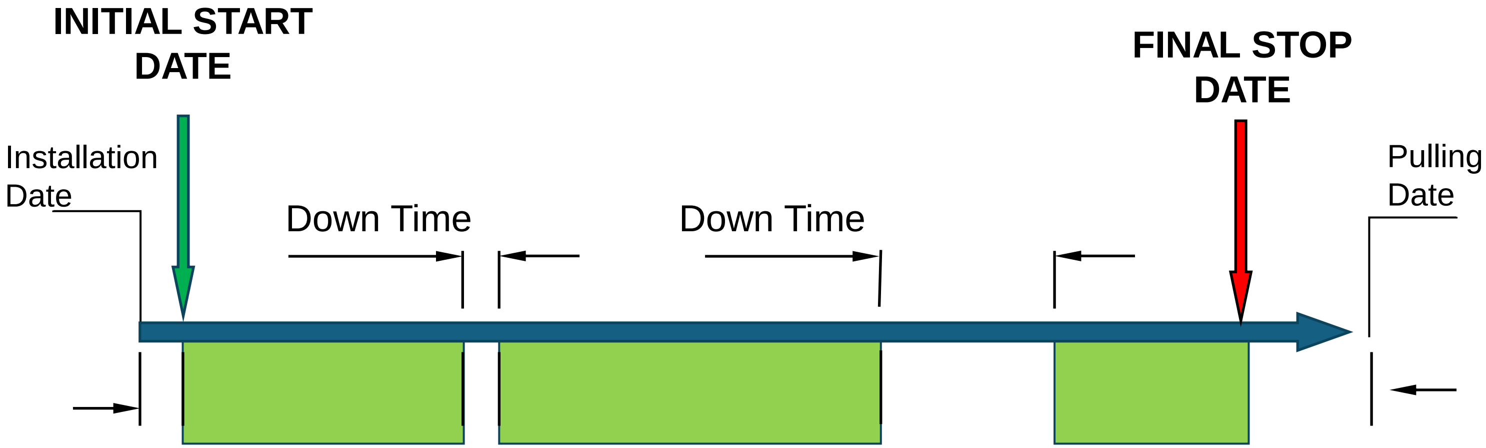
Input:
Operational data (10 years).

Approach:
Evaluate multiple algorithms, feature eng., and rigorous validation

Challenges:
Data: Imbalanced classes (few failures)
Modeling: Complex relationships

Business Understanding





Run Life = Final Stop Date – Initial Start Date

Well Intervention Type:
Failures and No Failures

Project Objectives

Develop a robust Machine Learning model that predicts failures in oil wells before they occur.

Reduce oil well downtime related to producer oil well failures

Improve failure prediction accuracy to reduce economic impacts.

Reduce the economic impact of equipment failures and associated downtimes.

Data Preparation

Data Integration:

Information was record of 10 years:

- Production
- Laboratories
- Fluid level
- Failures

Data Cleaning and Transformation:

- Anonymization techniques employ *hashing*
- Missing values
- Outliers handling using IQR.
- Inconsistencies
- Language differences

Data Preprocessing:

Creation of new features like "BSW (%)“, and “Total fluid rate”.
Encoding Categorical Variables such as “Artificial Lift Systems”

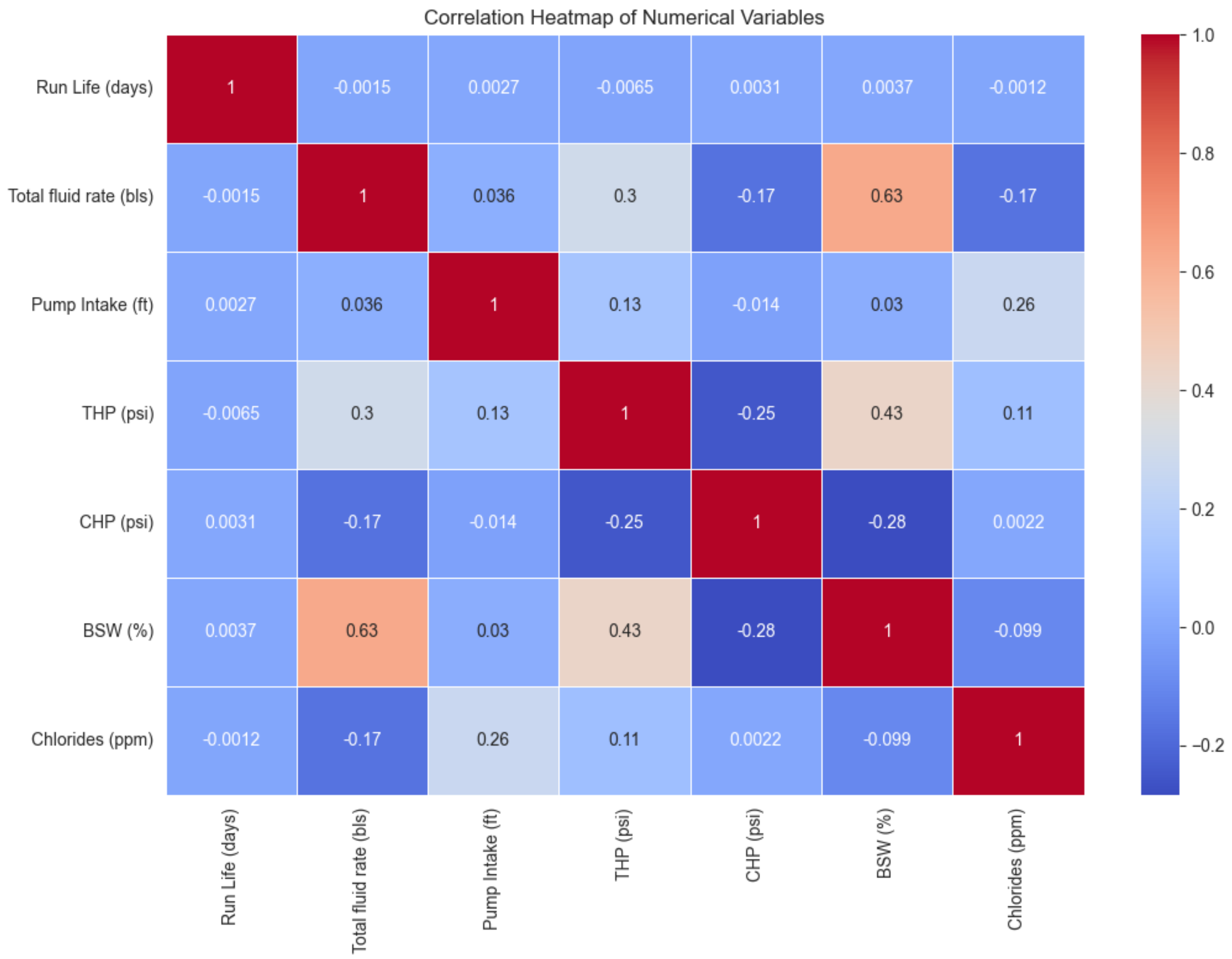
Well Intervention Type selected as the target variable

ML Well Failures Prediction

Database for Modeling

Matrix Correlation

Weak correlation between run life and other variables.



	Run Life (days)	Total fluid rate (bls)	Pump Intake (ft)	THP (psi)	CHP (psi)	BSW (%)	Chlorides (ppm)
Run Life (days)	1	-0.0015	0.0027	-0.0065	0.0031	0.0037	-0.0012
Total fluid rate (bls)	-0.0015	1	0.036	0.3	-0.17	0.63	-0.17
Pump Intake (ft)	0.0027	0.036	1	0.13	-0.014	0.03	0.26
THP (psi)	-0.0065	0.3	0.13	1	-0.25	0.43	0.11
CHP (psi)	0.0031	-0.17	-0.014	-0.25	1	-0.28	0.0022
BSW (%)	0.0037	0.63	0.03	0.43	-0.28	1	-0.099
Chlorides (ppm)	-0.0012	-0.17	0.26	0.11	0.0022	-0.099	1

Strong positive correlation between total fluid rate and BSW.

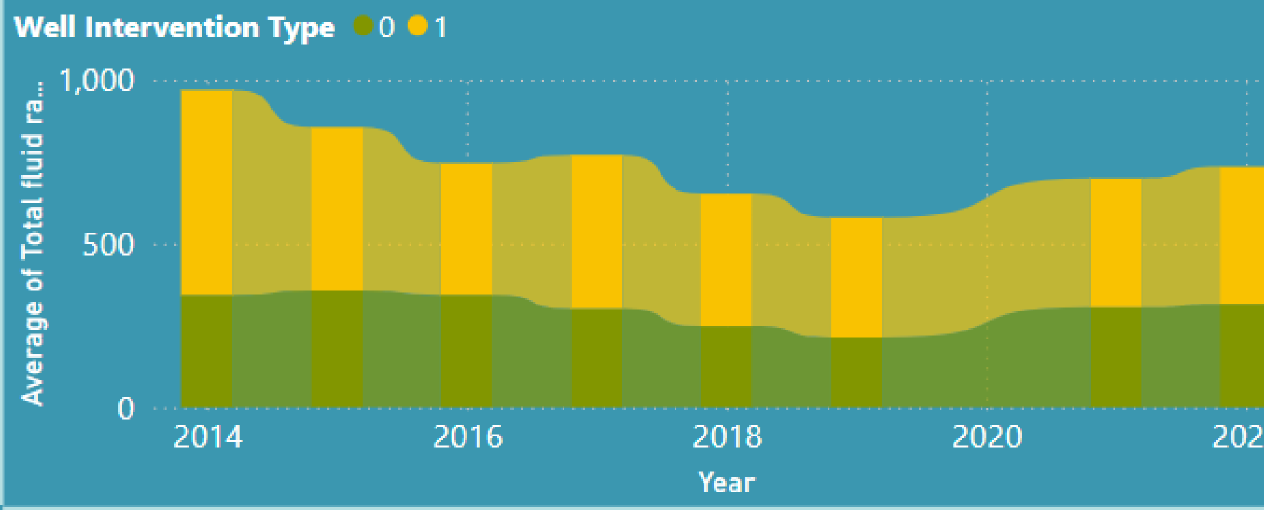
Inverse relationship: THP and CHP.

ANALYSIS

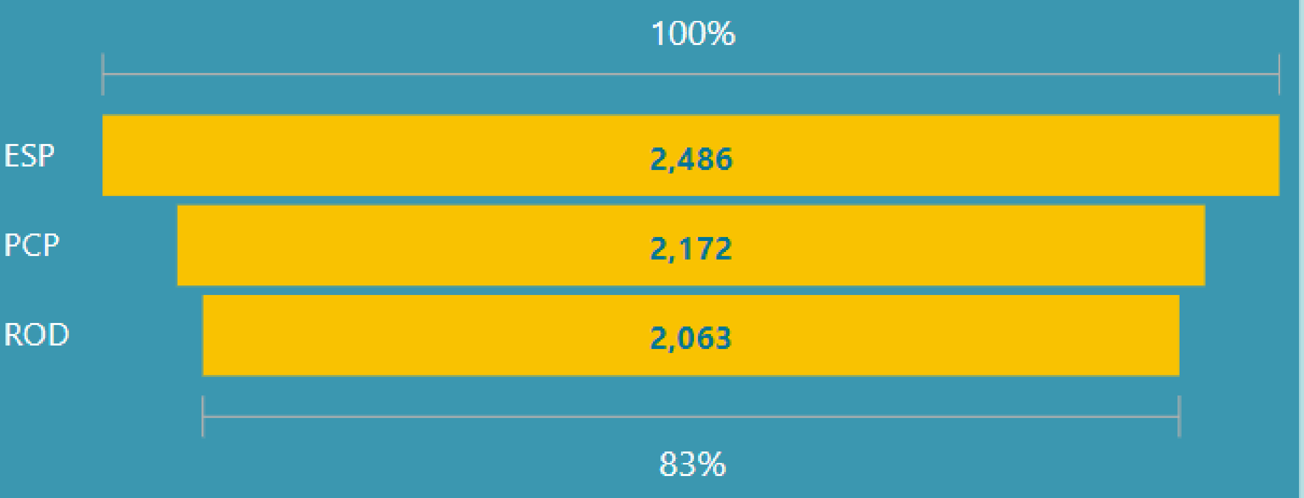
What kind of trends or patterns can prevent a failure?

Avg Run Life by ALS

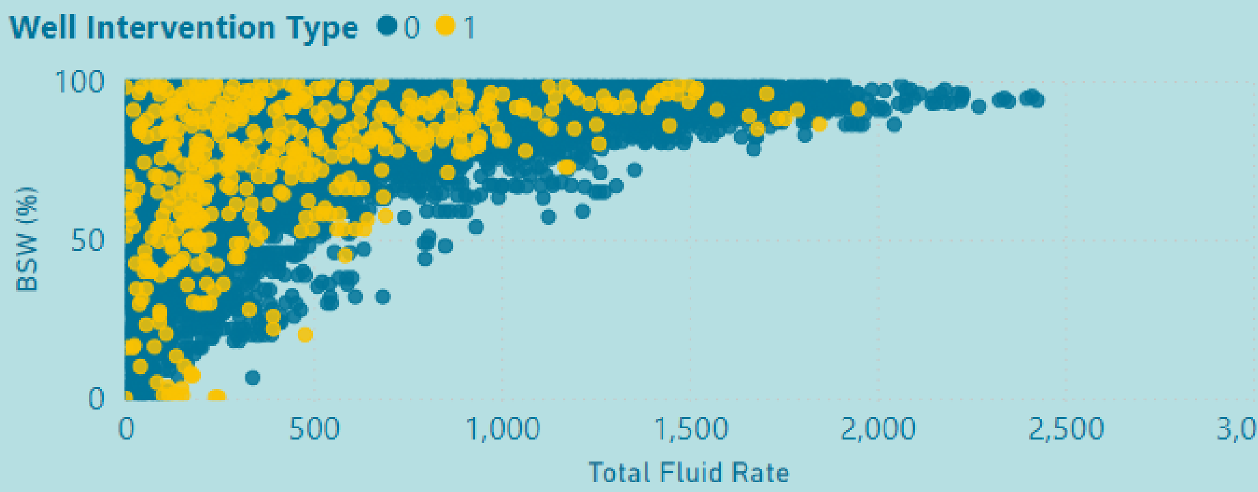
Avg of Total fluid rate by Well Intervention Type



Avg Pump Intake by ALS



BSW (%) and Total Fluid Rate relationship



Modeling

5

Model Algorithms

ML Model algorithms selection

Data Balancing

SMOTE for “oversampling” the class distribution:
Before:
0: 714,142 vs 1: 87,653
After:
1,428,284

Feature Selection

Using *SelectKBest*
RFE:

- Total fluid rate
- Fluid Level
- BSW
- Chlorides
- Run Life

PCA with 95 of variance.

Data Splitting

Split into train and test sets using *StratifiedShuffleSplit*

- Train and test sets (70 / 30)
- Train and validation sets (80 / 20)

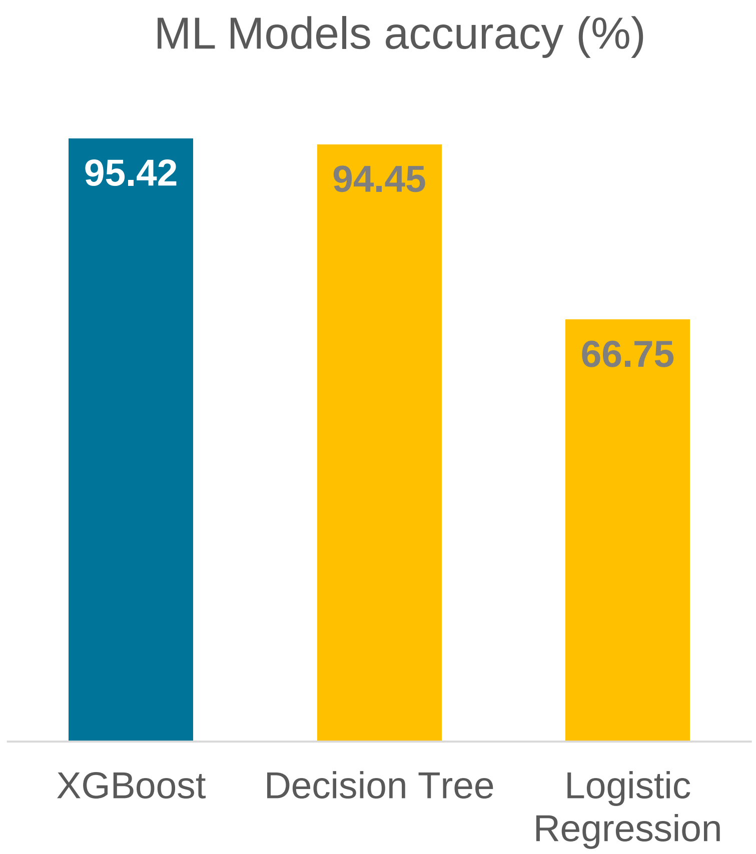
Models Evaluation

XGBoost
Decision Tree
Logistic Regression

Performance metrics:

- Accuracy
- Precision
- Recall:
- F1 Score

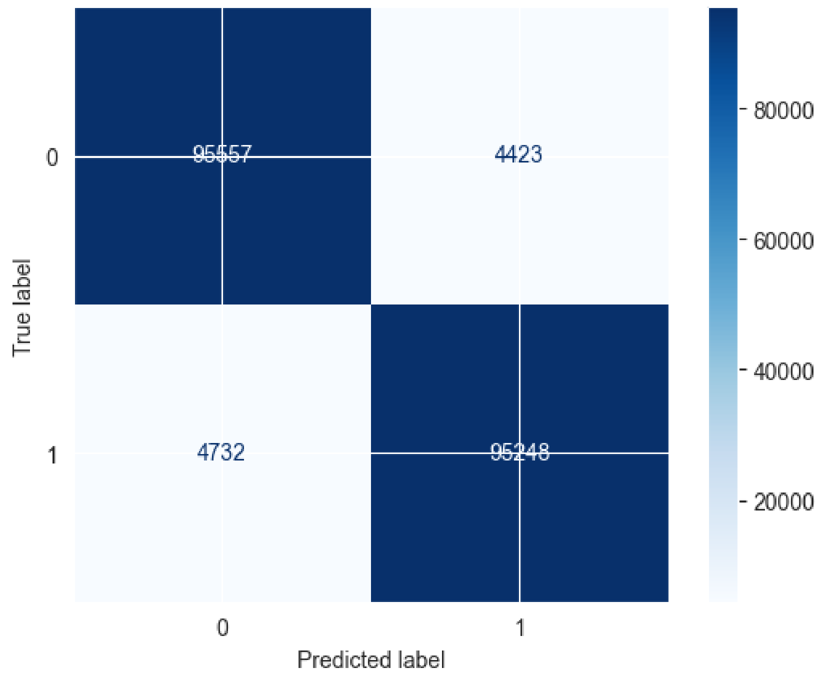
Modeling



XGBoost Model achieved the **highest** accuracy. It has been validated as recommended for primary use due to its reliability in predicting oil well failures.

Precision: 95.56 %	F1 Score: 95.41 %	Recall: 95.27 %

Modeling Results - XGBoost



XGBoos Model demonstrates a **good fit** with increasing accuracy as the training data grows.

CM indicates that the model **performs well** overall, with a balanced number of true positives and true negatives

Conclusions and Recommendations

The XGBoost model achieved the **highest accuracy** (95.42%).

The **Decision Tree model** (accuracy of 94.45%) shows a *balance between accuracy and interpretability*.

Perform a different Analysis **Artificial Lift System** (with and no rod strings) and Run Life.

Perform a future analysis including the impact of water injection on well production performance.

References

- Predictive Analytics for Oil Well Failures: A Machine Learning Approach

- Pennel, M., Hsiung, J., & Putcha, V. B. (2018). Detecting failures and optimizing performance in artificial lift using machine learning models. SPE Western Regional Meeting, Garden Grove, California, USA, 22-27 April 2018. Society of Petroleum Engineers. SPE-190090-MS.

- Reddicharla, N., Ali Sultan Ali, M., Alshehhi, S. S., Elmansour, A., & ReddyVanam, P. (2023). ESP failure prediction in water supply wells using unsupervised learning. Gas & Oil Technology Showcase and Conference, Dubai, UAE, 13-15 March 2023. Society of Petroleum Engineers. SPE-214010-MS.

References

- Pandey, Y. N., Rastogi, A., Kainkaryam, S., Bhattacharya, S., & Saputelli, L. (2020). Machine learning in the oil and gas industry: Including geosciences, reservoir engineering, and production engineering with Python (Chapter 7). Apress.

- Belyadi, H., & Haghighat, A. (2021). Machine Learning Guide for Oil and Gas Using Python (Chapters 4 and 5). Gulf Professional Publishing.

- Holdaway, K. (2014). Optimize Exploration and Production with Data Driven Models (Chapter 8). John Wiley & Sons, Inc.

- Bangert, P. (2021). In Machine Learning and Data Science in the Oil and Gas Industry (Chapters 11 and 12). Gulf Professional Publishing.

Questions?