*International Journal of Scientific Research and Engineering Development-– Volume 8 Issue 2, 2025*

*Available at www.ijsred.com*

RESEARCH ARTICLE OPEN ACCESS

# Survey on Machine Learning Models for Oil Production Forecasting

Aravindh T[1], Giriraaj L[2], Muthuannammalai S V[3], Sivaramakrishnan A[4]

[123]*UG scholar,* [4]*Assistant Professor, Department of Computer Science and Engineering, Chettinad College of Engineering and Technology, Karur, Tamil Nadu, India.*

E-mail : [1]aravindhvel2003@gmail.com, [2]giriraajloganathan2004@gmail.com, [3]muthusivaraman04@gmail.com, [4]sivainccet@gmail.com

----------**********----------

***Abstract:***

**This survey provides a detailed review of artificial intelligence and machine learning methods used for oil production forecasting. It examines different models, their strengths, limitations, and performance comparisons. Traditional models like Multiple Linear Regression (MLR) and ARIMA struggle with capturing complex patterns, while advanced techniques such as Random Forest (RFR), Decision Tree (DTR), XGBoost, and Artificial Neural Networks (ANN) show better accuracy. ANN performs the best overall but requires high-quality data and significant computing power. Hybrid models that combine different techniques offer promising improvements but face challenges like high computational costs and complex tuning. Future research should focus on making these models more efficient, easier to use, and better suited for real-world applications. AI-driven forecasting can greatly improve decision-making and resource management in the oil and gas industry.**

***Keywords*** **—— Oil Production Forecasting, Machine Learning, Artificial Neural Network.**

----------**********----------

## I. INTRODUCTION

The global demand for energy is increasing daily, with the oil and gas industry playing a crucial role in the global energy supply chain. Oil production involves extracting crude oil from wells and transforming these raw resources into petroleum products that can be utilized by consumers. As demand rises, there is a growing need for optimized production strategies to meet human energy requirements efficiently. Traditionally, methods such as reservoir simulation have been used to estimate future production rates.

In recent years, Artificial Intelligence (AI) and Machine Learning (ML) have influenced almost every industry, including oil and gas. The application of AI and ML models has significantly improved oil production forecasting, enhancing resource allocation and decision-making processes.

Effective forecasting depends on selecting the appropriate models, a critical step in any machine learning solution. Hybrid models, which combine two or more algorithms, can also be utilized to improve accuracy. However, selecting inappropriate models can negatively impact outcomes, leading to inadequate predictions.

Thus, it is essential to understand the working principles, advantages, and limitations of each model before implementation. A one-size-fits-all approach does not yield optimal results in oil production forecasting.

The objective of this survey is to provide a comparative study of existing prediction models used for oil production forecasting, analysing their methodologies, advantages, and limitations.

The structure of this paper is as follows: section II provides details on methodologies, section III explains existing research, section IV discusses various forecasting techniques, section V summarizes key findings, section VI references.

## II. MATERIALS AND METHODS

This section outlines the datasets, preprocessing techniques, machine learning models, implementation strategies, and evaluation metrics used in previous studies for oil production forecasting.

### *A. Data Collection and Preprocessing*

The studies utilized real-world datasets from various sources:

[1] *"Application of Artificial Intelligence in Predicting Oil Production Based on Water Injection Rate"* – 1,180 records of water injection and oil production history.

[2] *"Oil and Gas Production Forecasting Using Decision Trees, Random Forest, and XGBoost"* – Well production data from New York State, USA.

[3] *"Classical and Machine Learning Modelling of Crude Oil Production in Nigeria"* – Monthly crude oil production data (2006–2020) from NNPC and Reuters.

  

[4] *"Prediction of Oil Production through Linear Regression Model and Big Data Tools"* – Pressure, downhole temperature, and pressure tubing data.

[5] *"Knowledge-Based Machine Learning Approaches to Predict Oil Production Rate in the Oil Reservoir"* – Feature-engineered production data.

### *B. Machine Learning Models Applied*

Different models were implemented based on dataset complexity and forecasting objectives:

***1) Linear Models:*** Multiple Linear Regression (MLR) and Polynomial Regression (PR) (used in AI in Predicting Oil Production and Prediction of Oil Production through Linear Regression).

***2) Tree-Based Models:*** Decision Tree (DTR), Random Forest (RFR), and XGBoost (applied in Oil and Gas Production Forecasting and Knowledge-Based ML Approaches for handling nonlinear dependencies).

***3) Neural Networks:*** Artificial Neural Networks (ANN) (shown to outperform traditional models in AI in Predicting Oil Production and Classical and ML Modelling).

***4) Other ML Techniques:*** Support Vector Regression (SVR) and K-Nearest Neighbors (KNN) (applied in Knowledge-Based ML Approaches for capturing complex trends).

### *C. Model Implementation and Optimization*

The studies utilized Python-based libraries such as Scikit-learn, TensorFlow, and XGBoost for implementation. Common optimization techniques included: Hyperparameter tuning (e.g., grid search, random search), Feature selection to enhance predictive accuracy, Cross-validation to ensure model generalization.

### *D. Performance Evaluation*

Models were assessed using standard error metrics:

***1) Root Mean Square Error (RMSE):*** Used in all studies for error measurement.

***2) Mean Absolute Percentage Error (MAPE):*** Applied in Classical and ML Modelling to evaluate forecast accuracy.

***3) $R^2$ Score:*** Assessed model fit, with tree-based and ANN models achieving the highest values.

***4) Nash-Sutcliffe Efficiency (NSE):*** Used in Classical and ML Modelling for additional performance validation.

### *E. Comparative Analysis and Hybrid Approaches*

Some studies explored hybrid techniques to improve forecasting:

***1) Ensemble Learning (RFR & XGBoost):*** Enhanced prediction stability (Oil and Gas Production Forecasting).

***2) Neural Networks with Feature Selection:*** Reduced error rates (AI in Predicting Oil Production).

**3) Classical vs. ML Models:** ANN outperformed ARIMA and RF in Classical and ML Modelling.

## III. RELATED WORKS

Oil production forecasting has been a crucial research area, with various machine learning (ML) and deep learning (DL) techniques explored to improve predictive accuracy. Traditional methods, such as Multiple Linear Regression (MLR) and Polynomial Regression (PR), have been widely used but often fail to capture nonlinear dependencies in complex reservoir systems. With advancements in ML, models such as Artificial Neural Networks (ANN), Decision Tree Regressor (DTR), Random Forest Regressor (RFR), XGBoost, Support Vector Regression (SVR), and Long Short-Term Memory (LSTM) have gained prominence due to their ability to handle complex relationships in oil production data.

### *A. Application of Artificial Intelligence in Predicting Oil Production Based on Water Injection Rate (Rosiani et al.)*

Rosiani et al. (2024) explored the use of MLR, PR, and ANN for forecasting oil production based on water injection rates. The study utilized 1180 actual field data points, splitting them into 80% training and 20% testing. The ANN model achieved the best performance with the lowest Root Mean Square Error (RMSE) of 0.142 and the highest test accuracy of 16.2%, outperforming PR and MLR. The study demonstrated that ANN can effectively model oil production variations influenced by water injection rates, providing a more efficient and rapid alternative to conventional reservoir simulation techniques.

### *B. Oil and Gas Production Forecasting Using Decision Trees, Random Forest, and XGBoost (Al Shabaan & Nemer)*

Al Shabaan and Nemer (2024) investigated the effectiveness of tree-based ML models, including Decision Tree Regressor (DTR), Random Forest Regressor (RFR), and XGBoost, for oil and gas production forecasting. Their study aimed to overcome limitations associated with Numerical Reservoir Simulation (NRS) and Decline Curve Analysis (DCA), which are time-consuming and dependent on accurate static models. Using a dataset from New York State, USA, the study found that RFR achieved the highest accuracy (99%), followed by XGBoost and DTR. While RFR improved prediction accuracy and reduced overfitting, it was observed that it became computationally expensive with large datasets. XGBoost, being a gradient boosting technique, optimized feature selection and reduced bias, but required careful hyperparameter tuning to prevent overfitting.

### *C. Classical and Machine Learning Modelling of Crude Oil Production in Nigeria: Identification of an Eminent Model for Application (Obite et al.)*

Obite et al. (2021) conducted a comparative analysis between Autoregressive Integrated Moving Average (ARIMA), Artificial Neural Network (ANN), and Random Forest (RF) for forecasting crude oil production in Nigeria. They utilized monthly crude oil production data from 2006 to 2020 obtained from the Nigerian National Petroleum Corporation (NNPC) and Reuters. Their findings indicated that the ANN model performed significantly better than

  

ARIMA and RF in capturing the temporal dependencies of crude oil production. The ANN model had the best trade-off between RMSE, Mean Absolute Percentage Error (MAPE), and Nash-Sutcliffe Efficiency (NSE), making it the most reliable forecasting method. Their study concluded that ML models outperform classical statistical approaches for crude oil forecasting, reinforcing the need for AI-driven methods in reservoir management.

### *D. Prediction of Oil Production through Linear Regression Model and Big Data Tools (Alharbi et al.)*

Alharbi et al. (2024) proposed a Linear Regression Model combined with Big Data tools for oil production forecasting. Unlike previous studies that focused on complex ML architectures, their work explored a simple yet effective regression model that considered independent variables such as pressure, downhole temperature, and pressure tubing. Their results indicated that even with a basic linear regression approach, accurate predictions could be obtained when big data analytics were integrated. This highlights the importance of feature selection and data preprocessing in enhancing model performance, even in traditional regression methods.

### *E. Knowledge-Based Machine Learning Approaches to Predict Oil Production Rate in the Oil Reservoir (AlRassas et al.)*

AlRassas et al. (2024) introduced a knowledge-based ML framework incorporating Multiple Linear Regression (MLR), Random Forest (RF), Decision Tree (DT), and K-Nearest Neighbor (KNN) for oil production forecasting. Their study emphasized the importance of knowledge-based feature engineering, which helped improve prediction accuracy and reduced computational complexity. The results demonstrated that RF and KNN models performed significantly well, handling nonlinearity effectively while reducing training time compared to deep learning models. The study concluded that a hybrid approach integrating feature selection and multiple ML techniques could enhance prediction accuracy in complex reservoir conditions.

## IV. MACHINE LEARNING TECHNIQUES

Recent advancements in machine learning (ML) have significantly enhanced the accuracy and efficiency of oil production forecasting. Various ML models have been employed to capture the complex and nonlinear relationships inherent in oil reservoir systems. This section presents an overview of the ML techniques explored in prior research and their respective advantages and limitations.

### *A. Multiple Linear Regression (MLR) and Polynomial Regression (PR)*

*[1]"Application of Artificial Intelligence in Predicting Oil Production Based on Water Injection Rate"*

Traditional statistical approaches, such as Multiple Linear Regression (MLR) and Polynomial Regression (PR), have been used to predict oil production rates based on key operational parameters like water injection rates. These models offer a simple and interpretable framework but struggle with capturing nonlinear dependencies in reservoir systems. As demonstrated by Diyah Rosiani et al., the ANN model outperformed both MLR and PR, indicating the necessity of more advanced ML techniques for improved forecasting accuracy.

### *B. Decision Tree (DT) and Random Forest (RF)*

*[2]"Oil and Gas Production Forecasting Using Decision Trees, Random Forest, and XGBoost" & [5]"Knowledge-Based Machine Learning Approaches to Predict Oil Production Rate in the Oil Reservoir"*

Decision Tree Regressor (DTR) has been widely used for its interpretability and ability to model complex relationships. However, it often suffers from overfitting, especially when applied to large datasets. To mitigate this issue, ensemble learning techniques like Random Forest Regressor (RFR) have been introduced. Studies by Al Shabaan et al. and AlRassas et al. demonstrated that RFR significantly improves predictive accuracy by averaging multiple decision trees, making it a robust choice for oil production forecasting. However, as dataset sizes increase, RFR can become computationally expensive.

### *C. XGBoost for Enhanced Feature Selection*

*[2]"Oil and Gas Production Forecasting Using Decision Trees, Random Forest, and XGBoost"*

XGBoost, an advanced gradient-boosting algorithm, has been applied to optimize feature selection and minimize bias. Research by Al Shabaan et al. indicates that XGBoost provides better generalization and reduces overfitting when compared to traditional DT and RF models. Despite its improved predictive performance, XGBoost requires careful hyperparameter tuning, and improper selection can lead to model instability.

### *D. Artificial Neural Networks (ANN) for Complex Nonlinear Relationships*

*[1]"Application of Artificial Intelligence in Predicting Oil Production Based on Water Injection Rate" & [3]"Classical and Machine Learning Modelling of Crude Oil Production in Nigeria"*

Artificial Neural Networks (ANN) have proven effective in capturing highly nonlinear dependencies in oil production systems. Rosiani et al. and Obite et al. found that ANN models outperform classical statistical approaches such as ARIMA and MLR, achieving lower Root Mean Square Error (RMSE) values. ANN's ability to self-learn complex patterns makes it highly suitable for oil production forecasting, though it requires extensive training data and computational resources.

### *E. K-Nearest Neighbors (KNN) for Localized Predictions*

*[5]"Knowledge-Based Machine Learning Approaches to Predict Oil Production Rate in the Oil Reservoir"*

  

K-Nearest Neighbors (KNN) has been utilized for localized oil production predictions. AlRassas et al. reported that KNN effectively models reservoir behaviour when historical data is well-structured and densely populated. However, its performance deteriorates in high-dimensional spaces, making it less suitable for large-scale oil production forecasting.

### *F. Linear Regression with Big Data Tools*

*[4]"Prediction of Oil Production through Linear Regression Model and Big Data Tools"*

Rehab Alharbi et al. explored the application of Linear Regression models integrated with big data tools to improve oil production predictions. By leveraging large datasets with variables such as pressure, downhole temperature, and tubing pressure, their approach demonstrated the potential for scalable ML-driven forecasting. However, linear models often fail to capture intricate nonlinear dependencies, making them less effective for complex reservoir systems.

## V. CONCLUSION

This survey analysed various machine learning and deep learning techniques applied to oil production forecasting, highlighting their strengths, limitations, and computational trade-offs. Traditional statistical models such as Multiple Linear Regression (MLR), Polynomial Regression (PR), and ARIMA have shown limitations in capturing the complex, nonlinear dependencies inherent in oil reservoir systems. In contrast, advanced machine learning models like Random Forest (RFR), Decision Tree Regressor (DTR), XGBoost, and Artificial Neural Networks (ANN) have demonstrated superior predictive accuracy. Among them, ANN consistently outperformed other models, achieving lower RMSE values and better generalization. However, the effectiveness of these models depends heavily on data availability, feature selection, and computational resources. While RFR and XGBoost provided highly accurate results, they required significant computational power and careful hyperparameter tuning. Support Vector Regression (SVR) showed robustness for high-dimensional datasets but struggled with scalability. The study also highlighted that hybrid approaches combining ML and DL techniques offer promising solutions to overcome individual model weaknesses, improving overall prediction accuracy and generalizability. However, real-world implementation challenges, such as computational costs and data quality constraints, remain critical considerations. Future research should focus on optimizing hybrid models, improving real-time deployment, and leveraging advanced feature selection techniques. By addressing these challenges, AI-driven predictive models can significantly enhance decision-making, operational efficiency, and resource management in the oil and gas industry.

## VI. REFERENCES

[1] Diyah Rosiani, Muhamad Gibral Walay, Pradini Rahalintar, Arya Dwi Candra, Akhmad Sofyan, Yesaya Arison Haratua, *"International Journal on Advanced Science Engineering and Information Technology"*, Vol.13 (2023) No. 6 (2023), DOI: 10.18517/ijaseit.13.6.19399.

[2] Mays A. Al shabaan, Zainab N. Nemer, *"Al-Qadisiyah for Computer Science and Mathematics"*, Vol. 16(1) 2024, pp Comp. 9–20 (2024).

[3] Chukwudi Paul Obite, Angela Chukwu, Desmond Chekwube Bartholomew, Ugochinyere Ihuoma Nwosu, Gladys Ezenwanyi Esiaba, *" Energy Reports"*, Volume 7, November 2021.

[4] Rehab Alharbi, Nojood Alageel, Maryam Alsayil, Rahaf Alharbi, and A'aeshah Alhakamy, *"International Journal of Advanced Computer Science and Applications"*, Vol. 13, No. 12, 2022.

[5] Ayman Mutahar AlRassas, Chinedu Ejike, Salman Deumah. Wahib Ali Yahya, Anas A. Ahmed, Sultan Abdulkareem Darwish, Asare Kingsley, and Sun Renyuan,*" IFEDC"*, SSGG, pp. 282–304, 2024.

[6] Elliot, Kizzy Nkem, Damingo, Levi Adawari, *"International Research Journal of Modernization in Engineering Technology and Science"*, Volume:06/Issue:05/May-2024.

  