Optimization of Oil Recovery Factor Using Machine Learning Methods Integrated with Eclipse Simulator

By

Muhammad Rehan Ali

(Registration No: 00000362987)

Department of Chemical Engineering

School of Chemical and Materials Engineering

National University of Sciences & Technology (NUST)

Islamabad, Pakistan

(2024)

Optimization of Oil Recovery Factor Using Machine Learning Methods Integrated with Eclipse Simulator

By

Muhammad Rehan Ali

(Registration No: 00000362987)

A thesis submitted to the National University of Sciences and Technology, Islamabad,

in partial fulfillment of the requirements for the degree of

Master of Science in

Process System Engineering

Supervisor: Dr. Muhammad Nouman Aslam Khan

School of Chemical and Materials Engineering

National University of Sciences & Technology (NUST)

Islamabad, Pakistan

(2024)

THESIS ACCEPTANCE CERTIFICATE

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS thesis written by Mr **Muhammad Rehan Ali** (Registration No 00000362987), of School of Chemical & Materials Engineering (SCME) has been vetted by undersigned, found complete in all respects as per NUST Statues/Regulations, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for award of MS degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said thesis.

Signature: ______________________

Name of Supervisor: **Dr M. Nouman Aslam Khan**

Date: 27/08/2024

Signature (HOD): ______________________

Date: 28/08/2024

Signature (Dean/Principal): ______________________

Date: 28/8/24

TH - 1

Form TH-1

National University of Sciences & Technology (NUST)

MASTER'S THESIS WORK

Formulation of Guidance and Examination Committee (GEC)

Name: Muhammad Rehan Ali NUST Reg No: 00000362987

Department: Department of Chemical Engineering Specialization: Master of Science in Process System Engineering

Credit Hour Completed: 18.0 24 CGPA: 3.58 3.56

Course Work Completed

S/No:	Code:	Title:	Core/Elective:	CH:	Grade:
1.	PSE-801	Process Systems Theory	Compulsory	3.0	B+
2.	PSE-852	Process Modelling and Simulation	Compulsory	3.0	B
3.	TEE-820	Process Intensification	Elective	3.0	A
4.	RM-898	Research Methodology	Additional	2.0	Q
5.	PSE-802	Optimization and Decision Analysis	Compulsory	3.0	A
6.	PSE-823	Advanced Process Dynamics and Control	Compulsory	3.0	B+
7.	CSE-801	Computational Fluid Dynamics	Elective	3.0	B+

*

Date 05 - Dec - 2022 Student's Signature

Thesis Committee

Name: Muhammad Nouman Aslam Khan (Supervisor) Signature
Department: Department of Chemical Engineering

Name: Syed Rafay Hussain Jafri (Cosupervisor) Signature
Department: Department of Chemical Engineering

Name: Umair Sikander (Internal) Signature
Department: Department of Chemical Engineering

Name: Dr. Nouman Ahmad (Internal) Signature
Department: Department of Chemical Engineering

Date: 05 - Dec - 2022 Signature of Head of Department: 20/1/23

APPROVAL

Date: 05 - Dec - 2022 Signature of Dean/Principal: 27-1-2023

School of Chemical & Materials Engineering (SCME) (SCME) H-12 Campus.

* 8.	ENE-902	Numerical methods in Chemical Engineering	Elective	3	B+
9.	ENE-809	Waste water Treatment & Design	Elective	3	B+

TH – 4

Form: TH-04

National University of Sciences &Technology (NUST)

MASTER'S THESIS WORK

We hereby recommend that the dissertation prepared under our supervision by

Regn No & Name: **00000362987 Muhammad Rehan Ali**

Title: **Optimization of Oil Recovery Factor Using Machine Learning Methods Integrated with Eclipse Simulator.**

Presented on: **22 Aug 2024** at: **1430 hrs in SCME (Seminar Hall)**

Be accepted in partial fulfillment of the requirements for the award of Master of Science degree in **Process Systems Engineering**.

Guidance & Examination Committee Members

Name: **Dr Nouman Ahmad** Signature:

Name: **Dr Umair Sikander** Signature:

Name: **Dr Syed Rafay Hussain Jafri** (Co-Supervisor) Signature:

Supervisor's Name: **Dr Muhammad Nouman Aslam Khan** Signature:

Dated: 23/08/2024

Head of Department
Date 23-8-24

Dean/Principal
Date 23/8/24

School of Chemical &Materials Engineering (SCME)

AUTHOR'S DECLARATION

I Muhammad Rehan Ali hereby state that my MS thesis titled "Optimization of Oil Recovery Factor Using Machine Learning Methods Integrated with Eclipse Simulator" is my own work and has not been submitted previously by me for taking any degree from National University of Sciences and Technology, Islamabad or anywhere else in the country/ world.

At any time if my statement is found to be incorrect even after I graduate, the university has the right to withdraw my MS degree.

Name of Student: Muhammad Rehan Ali

Date: 19 February 2024

PLAGIARISM UNDERTAKING

I solemnly declare that research work presented in the thesis titled "Optimization of Oil Recovery Factor Using Machine Learning Methods Integrated with Eclipse Simulator" is solely my research work with no significant contribution from any other person. Small contribution/ help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero-tolerance policy of the HEC and National University of Sciences and Technology (NUST), Islamabad towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS degree, the University reserves the rights to withdraw/revoke my MS degree and that HEC and NUST, Islamabad has the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized thesis.

Student Signature: ____________

Name: Muhammad Rehan Ali

DEDICATION

" To my very supportive, loving and caring family."

ACKNOWLEDGEMENTS

All praise and eminence are due to "ALLAH," the undisputed architect of this world, who gave us the capacity for comprehension and sparked our curiosity about the planet as a whole. Warmest welcomes to the supreme ruler of this world and the hereafter, "Prophet Mohammed (PBUH)," a source of knowledge and benefits for all of humanity as well as for Umah.

I would like to convey my profound gratitude to **Dr. Muhammad Nouman Aslam Khan and Dr. Rafay Jafery** for their leadership, motivation, helpful suggestions, dedication, and thorough oversight during the project. For their aid during my program, I am appreciative to them. It was a fantastic privilege and an honour to work under his direction. I shall cherish this memory for my entire life.

Additionally, I want to express my gratitude to **Dr Nouman Ahmad** and **Dr. Umair Sikandar** from the project's advisory group for their informative and helpful suggestions.

Muhammad Rehan Ali

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS, ABBREVIATIONS AND ACRONYMS

Ultrasonication	US
Liquid Phase Exfoliation	LPE
Energy Density	ED
Inertial Cavitation Dosage	ICD
Cavitation Field	CF
Cavitation Intensity	CI
Cavitation Zone	CZ
Cavitation Bubbles	CB
Acoustic Power	AP
Bulk Titanium Diboride	BTB
Two Dimensional Nanomaterials	2D
Two Dimension TiB_2 Nanosheets	2D TiB_2
Few Layered 2D TiB_2	FLTB
N-methyl-2-pyrrolidone	NMP
De-Ionized Water	DIW

Transition Metal Diborides	TMB
Bulk Form	BF
Nanomaterials	NM

ABSTRACT

As the world grapples with the challenges of sustainable energy, oil and gas remain the main sources of the world's energy supply. While, global oil consumption has increased significantly, the accessible reserves of oil and gas resources have decreased as a result of the usage of common light oils. To maximize extraction from known sources, secondary recovery procedures, often known as EOR, - enhanced oil recovery- are used when relying exclusively on reservoir pressure is no longer adequate. EOR projects depend on variety of factors including economic factors, crude oil market price, and the investment from individuals and firms willing to take on the risk. To optimize extraction beyond reservoir pressure and maximize reservoir recovery and profits, several methods for EOR have been proposed including water injection and polymer and surfactant injection. This paper examines the use of three distinct machine learning simulation models- the random forest regressor, the gradient boost regression, and the KNN Regression- to investigate these EOR methods. The three models were applied to over 1500 data points from field data and extrapolated and expanded using Eclipse Simulator. Analysis of the techniques by these methods show that surfactant injection achieved the greatest return in oil production with an average oil rate of 27.27Mstb per day. Polymer Injection produced the second greatest oil generation per day with an average rate of 25.19 Mstb per day, with water injection generating the least amount of oil at 25 Mstb per day. These results suggest that surfactant injection could be a promising method for enhancing oil recovery and extending the lifespan of existing oil and gas reserves.

Keywords: EOR, water injection, polymer injection, surfactant injection, Eclipse Simulator, KNN Regression, Random Forest regressor, Gradient boost regression, Machine learning.

CHAPTER 1: INTRODUCTION

1.1 Background and Context

Despite the growth of renewable energy sources, oil and gas remain the main sources of the world's energy supply. There is an increasing need to extract more oil from undiscovered reservoirs to meet the expanding energy needs as energy demand rises and conventional supplies become increasingly limited [1]. By 2040, the Organisation of the Petroleum Exporting Countries (OPEC) forecasts that global oil consumption will have increased significantly, rising by 23.1% from present levels to around 111.1 million barrels per day. The accessible reserves of these resources have, however, decreased as a result of the usage of common light oils. The need to access alternate fossil resources is critical given the continuous reliance on fossil fuels for the foreseeable future [2]. Techniques for Enhanced Oil Recovery (EOR) include primary, secondary, and tertiary approaches. Following oil discovery, primary recovery starts, using natural reservoir pressure for early oil production. Secondary recovery procedures, often known as EOR, are used when relying exclusively on reservoir pressure is no longer adequate. Advanced EOR techniques are used in tertiary recovery to extract oil beyond primary and secondary phases [3].

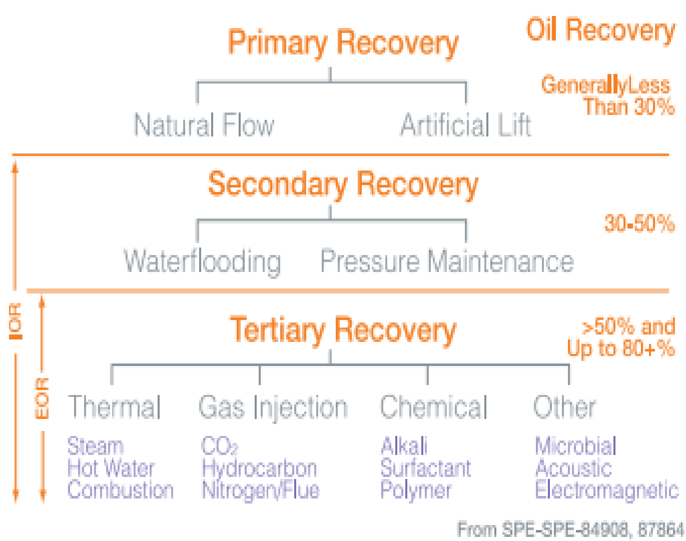


Figure 1.1: IOR/EOR [4]

Eric Delamaide et al. compared polymer flooding in primary, secondary and tertiary stage of a heavy oil field [5]. Alex F. Teixeira et al. proposed machine learning models to support reservoir production optimization [6]. Zhi Zhong et al. predicted reservoir production rates during waterflooding uses a proxy model that makes use of machine learning [7].

Rui ZHANG et al. suggested to use a multivariate time series (MTS) and vector autoregressive (VAR) machine learning model for waterflooding reservoir for oil well output forecasting [8]. Carlos Calad et al. Combined Traditional Reservoir Physics with Machine Learning for Predictive Modelling and Optimisation of a Large Mature Waterflood Project in the Argentinan Gulf of San Jorge Basin [9].

Ehsan Amirian et al. employed ANN to predict how well the water-flooding process will recover [10]. An innovative strategy for predicting the entire oil output from the White Tiger subterranean reservoir was proposed using a supervised learning methodology. This approach was then used to maximise Net Present Value (NPV) by optimising waterflooding tactics. Hien DH et al [11]. To be employed as screening and optimised design tools for polymer injection projects, a forward-looking and an inverse design expert ANN system is being developed.

Qian Sun et al [12]. Imankulov et al. used algorithms of machine learning to address the issues of polymer injection into an oil reservoir [13]. It is shown that the support vector machine-based quantitative prediction models of polymer flooding performance consider both universality and extendibility, and they meet the requirements of engineering computation applications. Jian Hou et al [14].

To show complex polymer gel kinetics and flow dynamics in deep conformance situations with accuracy, the work offers a unique surrogate modelling technique that makes use of machine learning.

A reliable method of predicting oil recovery results for polymer gel treatments in fractured reservoirs is provided by the suggested neural network model. Notably, the model outperforms typical commercial simulators in terms of processing speed and computational efficiency. Mohammad Algazal et al [15]. Random Forest regression was used to build a regional surrogate model for the oilfield. It advances the planning and economic elements inside digital oilfield operations, improving accuracy and efficiency by effectively implementing machine learning approaches for modelling. Fedor Krasnov et al [16].

Yanbin Wang et al. used the Backpropagation (BP) artificial neural network technique, predictive models for enhanced oil recovery (EOR) and internal rate of return (IRR) in polymer flooding situations have been created [17].

A recurrent neural network model has been designed to assist in making practical decisions and forecasting especially for the case that includes a lot of production factors and geological factors in high-dimensional data space. This model should forecast data in heavy oil reservoirs production via Steam Assisted Gravity Drainage (SAGD). It is a useful option to the laborious and tedious traditional means of numerical static and dynamic simulation. Yanwei Wang et al [18]. The incremental recovery of oil has become high enough for our study, which brings it from low oil recovery to high oil recovery. With this development, it will be received how surfactant-induced wettability change is achieved.

We have also developed a quick framework which would make it easier to forecast, analyze and improve the surfactant performance. This approach has an ability to cut down in the amount of time required for experimenting dramatically. Ya Yao et al [19]. This major goal was to develop reliable and accurate models for estimation of interfacial tension (IFT) between ionic surfactants and regular alkanes. Three clever computer-aided methods were used in the study to achieve this goal: decision trees (DT), extra trees (ET), and gradient boosting regression trees (GBRT) models. These methods might be used to generate precise models that would make computing the IFT simple.

Artificial intelligence has been used by Ali Rashidi-Khaniabadi et al. [20] and Mahdi Shayan Nasr et al. to allay concerns about initial evaluations of enhanced oil recovery (EOR) techniques. The research focused on many machines learning techniques, including the radial basis function-artificial neural network (RBF-ANN), multilayer perceptron-artificial neural network (MLP-ANN), and adaptive neuro-fuzzy inference system (ANFIS). Using these techniques, it was predicted that several research employing silica nanofluid flooding in carbonate and sandstone reservoir core samples would be effective [22].

1.2 Statement of Problem

The accurate prediction of oil recovery factor remains a significant challenge, particularly when dealing with different oil recovery methods and varying operating conditions. Traditional EOR techniques may struggle to account for these complexities, leading to suboptimal predictions. Additionally, the computational cost associated with running detailed simulations for large-scale industrial processes can be prohibitive.

To address these challenges, there is a growing interest in integrating Machine Learning (ML) techniques with simulation software for creating prediction models. ML methods, such as artificial random forest regressor, KNN regression, gradient boost regression have shown promise in improving the predictive accuracy of EOR methods. However, developing a seamless integration of EOR methods and ML, along with optimizing model parameters, is an area that requires further exploration.

1.3 Research Objectives

The primary objective of this thesis is to develop a machine learning model and compare the three enhanced oil recovery techniques to find the best machine learning model and enhanced oil recovery method which can produce higher oil rate per day hence increased oil recovery. To achieve the objectives machine learning models are developed on the data produced by the live oil field and eclipse simulator. Some data analytics techniques are employed to understand the data distribution.

1.4 Scope and Limitations

This research will focus on comparing different machine learning models and EOR methods and the prediction of oil rate by each EOR method. The limitations of this research include:

- The study's findings may not be directly transferable to all EOR techniques due to varying process parameters.
- The finding may not be transferable to all oil fields due to different geological parameters.

The performance of model may depend on the quality and quantity of available data.

CHAPTER 2: LITERATURE REVIEW

2.1 Literature Review

Most of the oil output in the world today is from fields that are mature, and hence enhanced oil recovery has come out to be a necessity from these aging resources. This is a matter of major concern to oil companies and governments, as replacement of produced reserves with new finds has slid steadily in recent decades. Therefore, to meet future energy demand, recovery factors have to be optimized from the older sources through both primary and secondary production methods. Improved Oil Recovery (IOR) mechanisms could be considered as Enhanced Oil Recovery (EOR) techniques, the use of new drilling technology, more intelligent reservoir management, better monitoring techniques, and fundamental changes in primary and secondary recovery processes. Economic factors, the price of crude oil, and the willingness of the investor to manage risks and economic exposure—all have a bearing on EOR projects [23].

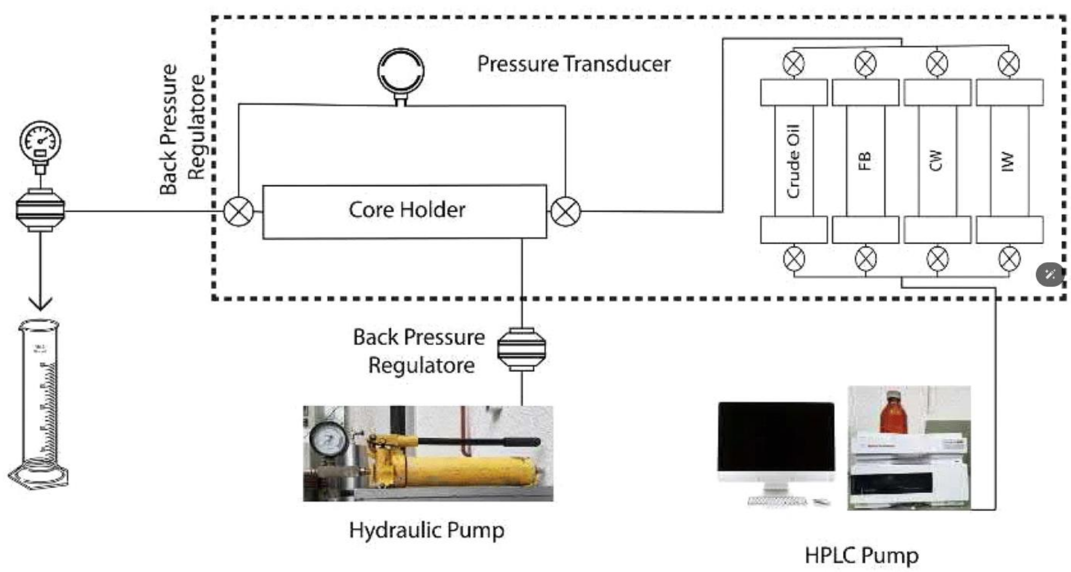


Figure 2.1: Schematic diagram of flooding process [24]

There are three main phases to the development of an oil field: in the primary stage, the initial reservoir pressure produces oil, allowing it to ascend freely to the surface. When the reservoir pressure starts to decline, water injection is typically employed to keep the pressure stable and serve as the secondary stage's driving power. Ultimately, the remaining oil is recovered using a range of techniques in the tertiary (EOR) stage, such as steam, CO2 injection, and natural gas miscible injection.

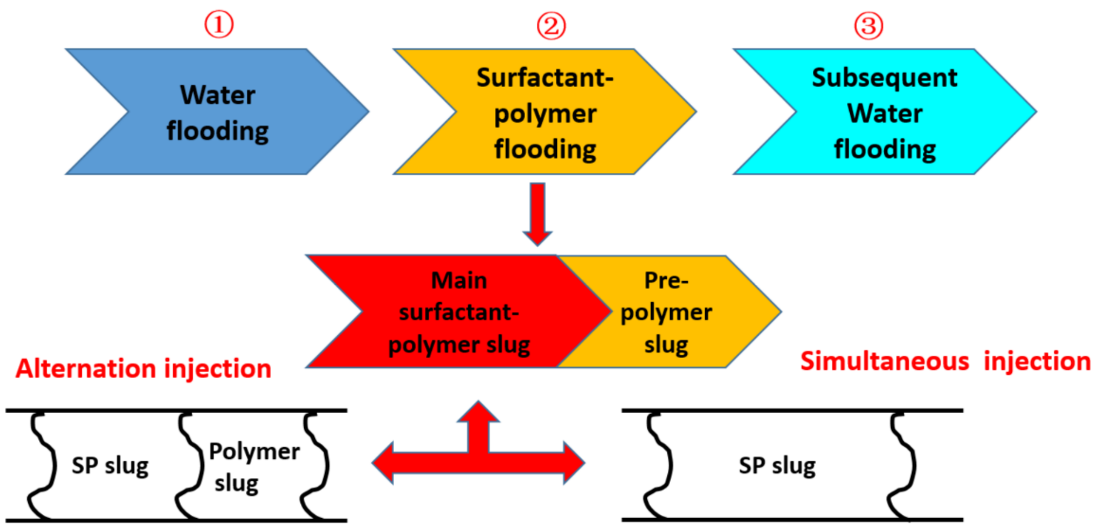


Figure 2.2: Typical injection procedure and slug injection Strategy [23]

These stages are crucial for increasing the oil recovery factor and prolonging the useful life of the oil field [24]. The amount of oil recovered during primary recovery can range from zero to fifty percent or more of the initial oil in situ, contingent upon the hydrocarbon type and the reservoir's driving mechanism. When the gravity drive is effective, a water drive can yield 50% or more in the case of the light oil reservoirs. It is zero in oil sands. Secondary recovery, which is often accomplished through water flooding, ranges from nil for oil sands to a few percent for heavy oils and up to 20-50% for light oils. EOR extends beyond secondary recovery and incorporates a variety of costly approaches, some of which are successful under favourable conditions. EOR is particularly important for highly viscous oils and oil sands with low primary and secondary productivity [25].

2.2 Water Injection

Water injection is critical in maximizing the economic consequences of many oil fields. Successful water injection operations contribute to both expanding oil production and enhancing overall recovery by maintaining reservoir pressure and shifting oil towards producing wells. Water injection can greatly boost the recovery factor in offshore environments, increasing it from roughly 15% with primary recovery to more than 40% [26]. The success of reservoir management is dependent on current and future reservoir performance. The accuracy is affected by the quality of the simulation model utilised. It enables for the evaluation of development approaches and the optimisation of reservoir operation. It is critical to develop a dependable model. These models are used in production optimisation to maximise asset value. Water injection is a technique for optimising recovery that increases oil output and economic productivity. It is popular due to the ease of injection and the efficient oil displacement by water, which extends the productive life of the oil field [27].

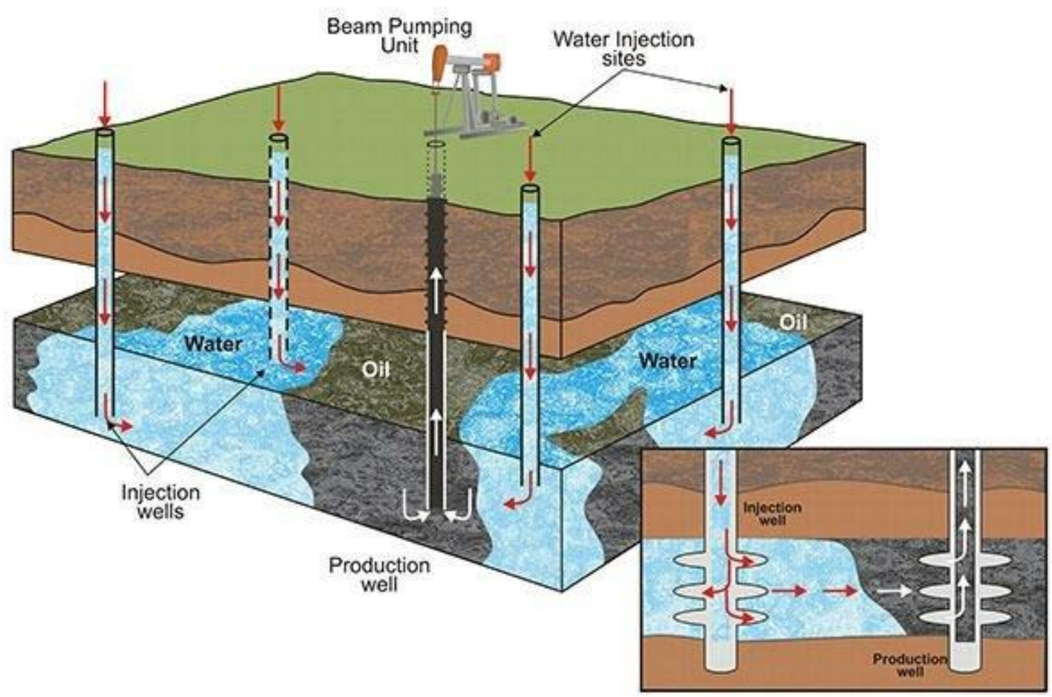


Figure 2.3: Schematic of water injection process

Water is pumped into a reservoir to displace and push existing hydrocarbons into producing wells, hence improving oil recovery following the natural depletion stage. However, the reservoir's geological heterogeneity can have an impact on the efficiency of water flooding. Fortunately, recent developments in downhole equipment and instrumentation have resulted in better sweep efficiency by changing production and injection rates in the reservoir's geological layers. These advancements aid in the optimisation of the water flooding process and the recovery of oil from the reservoir [28]. Numerical simulation models used to analyse subsurface flow injection and production mechanisms can be exceedingly complicated and computationally expensive, making the procedure time-consuming.

In the last few decades, academics have resorted to machine learning approaches such as simple to advanced regression and artificial neural networks to overcome these difficulties. These machine learning techniques could help lower processing costs and improve the efficiency of analysing subsurface flows [29]. Using the most recent industry database, the researchers created an Artificial Neural Network (ANN). The major goal was to develop an ANN model capable of calculating the best Enhanced Oil Recovery (EOR) method based on reservoir rock and fluid parameters. The goal is to obtain accurate estimates in a timely and cost-effective manner. The study will also assess the applicability of the constructed ANN model in real-world circumstances [30].

2.3 Polymer Injection

Polymer injection is a popular Enhanced Oil Recovery (EOR) technology that is used when conventional water flooding techniques fail to produce sufficient oil recovery. Water tends to travel through higher permeability zones and fractures during flooding, bypassing lower permeable zones containing considerable amounts of residual oil. Polymer flooding, first used for secondary and tertiary oil recovery in the 1960s, has found widespread use in both operational and research settings. When water-soluble polymers are added to water, the viscosity of the water increases, resulting in a decrease in the water-to-oil mobility ratio.

As a result, the sweep efficiency and Recovery Factor (RF) are improved. Furthermore, when compared to typical water flooding methods, polymer flooding requires less water [31]. Polymer flooding may have an economic impact because less water is injected and produced when compared to water flooding [32].

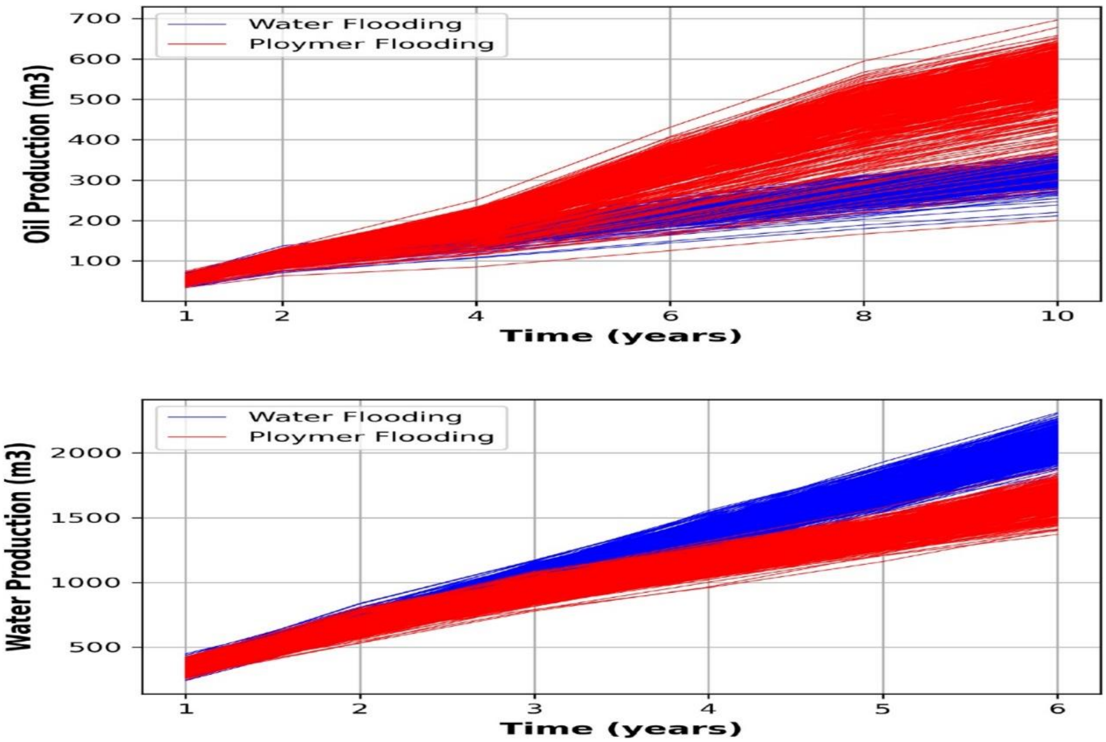


Figure 2.4: Oil Production and Water Production over time [33]

Polymer flooding is a successful enhanced oil recovery process that involves injecting into the reservoir water combined with a tiny proportion of soluble polymer. By raising the viscosity of the injected fluid, this procedure enhances sweep efficiency by reducing viscous fingering and generating a smoother flood front. Maximum sweep efficiency is accomplished by reducing the mobility ratio of the injected fluid in comparison to the oil phase.

Polymer flooding is especially beneficial in heavy oil reservoirs with horizontal injection wells because it allows for bigger polymer slug sizes. This methodology beats other heavy oil recovery systems, such as SAGD and ES-SAGD, in terms of cost and environmental efficiency [34].

2.1 Surfactant Injection

Surfactants, also known as surface active agents, are chemical molecules that have a hydrophilic head and a hydrophobic tail. Surfactants are divided into four types based on the charge on the hydrophilic head: non-ionic (no charge), anionic (negative charge), cationic (positive charge), and zwitterionic (both negative and positive charges). Surfactants often have a chain structure that is mostly affected by the content of the hydrophobic tail. A surfactant's tail group can be made up of a short polymer chain, a long hydrocarbon chain, a siloxane chain, or a fluorocarbon chain.

The head group, on the other hand, is made up of moieties like sulphates, sulfonates, polyoxyethylene chains, carboxylates, alcohols, or quaternary ammonium salts. The presence of these groups confers the amphiphilic property [35].

To boost oil recovery in reservoirs, surfactant Enhanced Oil Recovery (EOR) exploits two essential mechanisms. Surfactants, for starters, reduce the interfacial tension between oil and water, which lowers capillary pressure and allows water to displace trapped oil. Water bypass is a method that mobilises previously immobile oil.

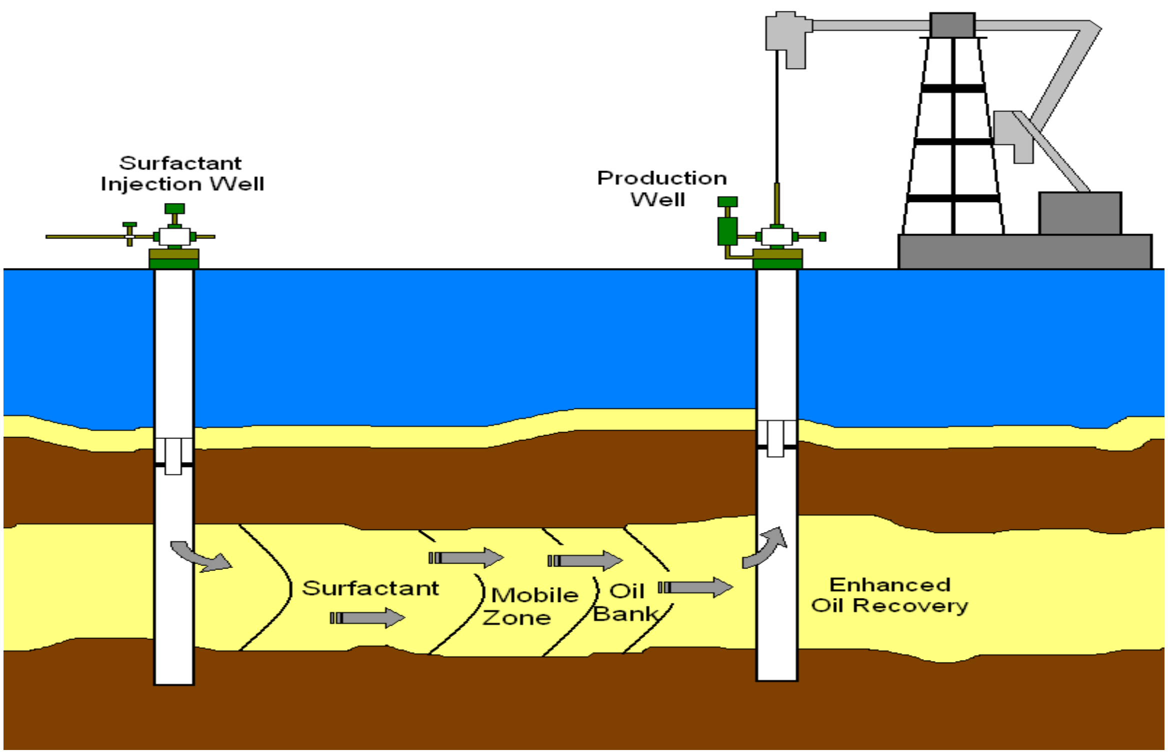


Figure 2.5: Schematic of Surfactant based injection process [36]

Surfactant is injected through the injection well which push the oil toward the production well. Second, surfactants change the wettability of the reservoir to make it more water-wet, assisting in the detachment of oil films from pore walls and minimising residual oil saturation. Surfactant flooding improves oil recovery during water flooding operations by synergistically reducing interfacial tension and altering reservoir wettability, ultimately increasing the amount of oil taken from the reservoir [36].

CHAPTER 3: ML THEORETICAL FRAMEWORK

3.1 Random Forest Regressor

Leo Breiman developed the Bagging method in 1996, which was one of the early-stage algorithms [37]. Amit and Geman generate a large variety of geometric attributes and use a random selection of them to get the ideal split at each note [38]. Dietterich proposed the random split selection hypothesis in 1998 [39]. The division is randomly selected from the top N probable divisions at each internal node. Ho pioneered a large amount of research into the "random subspace" approach, in which each tree is enlarged using a randomly chosen subset of characteristics. Breiman created additional training sets by adding unpredictability into the outputs of the original training set in a separate paper. Notably, the ideas presented in a paper written by Amit and Geman were influential in forming Breiman's thinking about random forests [40,41].

Random forests are a hybrid of machine learning techniques. They combine an ensemble of tree classifiers, with each tree contributing a single vote to the prevailing class, and the ultimate choice is made by averaging their outputs. Random forests have several benefits, including as excellent classification precision, robust handling of outliers and noise, and a lack of overfitting issues. In the realms of data mining and biological research, random forests have earned substantial attention and recognition, emerging as a favoured avenue for investigation [42].

The generalisation error of any decision tree h(x) equals for the input vector A and the output vector B.

$$E_{X,Y}\left(Y - h(X)\right)^2 \dots (1)$$

RFR is projected to be equal to the average of k decision trees h (x, k).

The RFR approach does not induce overfitting as the number of decision trees increases, but it may create generalisation mistake within a specified limit. The following can explain this phenomenon:

Almost definitely, when the number of decision trees h(x) in the forest approaches infinite,

$$E_{X,Y}\left(Y - av_k h(X, \Theta_k)\right)^2 \Longrightarrow E_{X,Y}\left(Y - E_\Theta h(X, \Theta)\right)^2 \ldots (2)$$

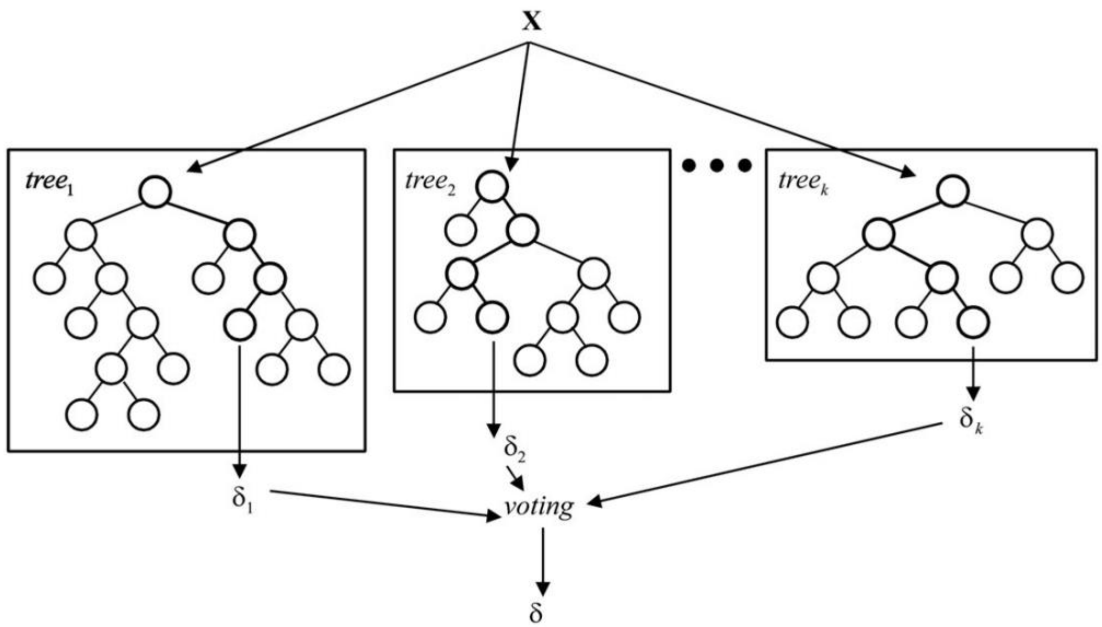


Figure 3.1: Schematic of Random Forest [43]

3.2 Gradient Boost Regression

Boosting tactics combine weak learners by emphasising on the errors made at each level, culminating in the acquisition of a strong learner who is the sum of the consecutive weak learners [44]. Gradient Boosting Regression (GBR) techniques are commonly employed in regression and classification issues and can produce good results [45]. The function estimation or approximation technique is reframed by focusing on numerical optimisation inside the function space rather than the parameter space. This approach integrates stagewise additive expansions with steep descent minimization. For additive expansions, a comprehensive gradient descent "boosting" architecture that is scalable to multiple fitting criteria is provided. The paradigm includes regression techniques that use least-squares, least absolute deviation, Huber-M loss functions, and multiclass logistic likelihood for classification. This methodology is improved by using regression trees as additive components, a technique known as "TreeBoost," which provides competitive, resilient, and interpretable models for both regression and classification, making it especially suitable for dealing with noisy data [46]. Two tuning parameters related to the gradient boosting regression are the shrinkage rate and ntrees. The number of trees that have been grown is thought to be ntrees. The shrinkage parameter is widely understood to indicate the learning rate of every tree that has expanded due to the expansion. [47].

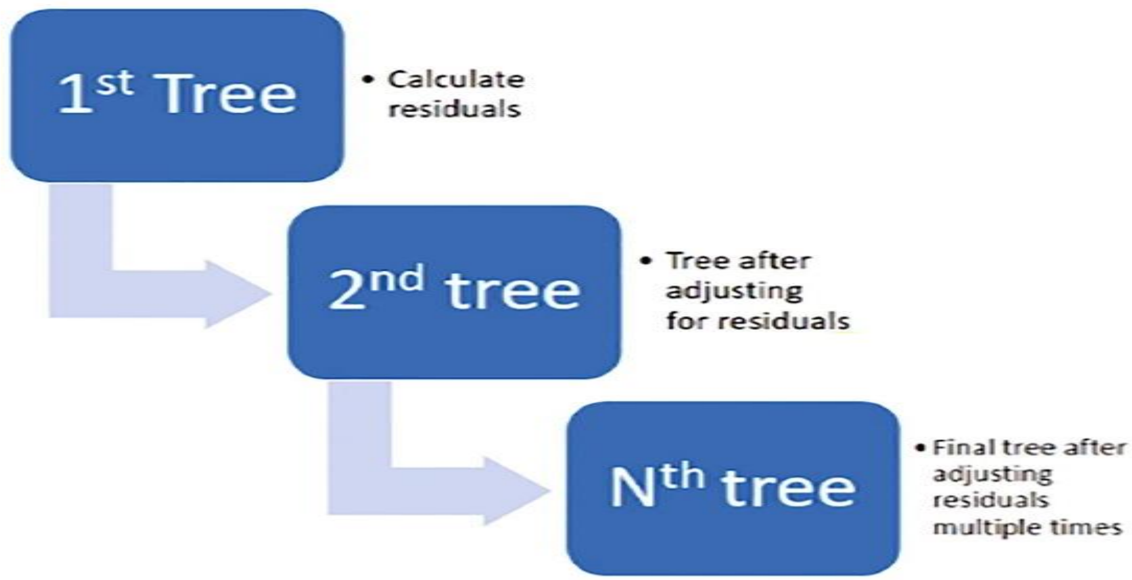


Figure 3.2: Flow Chart Gradient Boosting Regression [48]

3.3 KNN Regression

Cover and Hart introduced the K-Nearest Neighbour method in 1968 and refined it subsequently [49]. Supervisory learning includes algorithms such as the K-Nearest Neighbors (KNN) regression method. It predicts a target value by comparing the supplied instance's similarity to other existing examples. A distance metric, most typically the Euclidean distance, is used to determine similarity. Identifying the K most comparable occurrences, or neighbours, to the testing data point from the complete dataset is the method. KNN uses the Euclidean distance formula to calculate the distances between the numerical values of these locations. This computation calculates the neighbours' proximity and leads the forecast for the testing point's goal value [50]. The kNN algorithm prioritises the closeness of comparable items. When building a kNN model, many metrics such as Euclidean, Manhattan, and Chebyshev distances can be used to find closely related elements. These metrics serve in identifying the similarity of items inside the algorithm [51].

Euclidean

$$\sqrt{(\sum_{i=1}^{n}(y_i} - x_i)\text{^}2 \ldots (3)$$

Manhattan

$$\sum_{i=1}^{n} |\, y_i - x_i| \ldots (4)$$

Minkowski

$$(\sum_{i=1}^{n}(y_i - x_i)^p)\text{^}1/p \ldots (5)$$

Chebyshev

$$\max\ |y_i - y_i| \ldots (6)$$

Some of the factors considered in optimising the model are the distance, point weights, neighbourhood number, and p parameters related with the Minkowski function [52].

CHAPTER 4: METHODOLOGY

4.1 Data Acquisition

In this investigation, three alternative enhanced oil recovery (EOR) methods—water injection, polymer injection, and surfactant injection—were used. Over 1500 data points were included in the characteristic and target variables for each approach. Additional data points were extrapolated using the Eclipse Simulator after the dataset was initially received from an oil extraction field. Python was used to apply machine learning models, data cleaning and preparation and visualizations. The libraries used are pandas, sklearn and seaborn. A few missing values were found during the data preprocessing stage and then eliminated to make sure they wouldn't negatively impact the machine learning results. Further evidence that the removal of these missing values does not create any biassed tendencies comes from the fact that the number of missing values was comparatively low in comparison to the non-missing values.

Feature variables include average pressure psi, Gas Rate MMscf/day, Ratio of Produced Water to Injected Water, Voidage Replacement Coefficient, Water Voidage Production Rate MRB/day, Watercut, fraction, Water Injection Rate Mstb/day, Water Rate Mstb/day, Water-Oil Ratio stb/stb, Number of Injectors Currently Flowing, Number of Producers Currently Flowing, Oil Total MMstb and Oil Rate Mstb/day is the target variable.

The mean pressure applied within the oil reservoir during the extraction process, on the other hand, is known as the average pressure (psi). This becomes significant because it regulates the production and flow rates of fluid. To guarantee maximum oil recovery and to appropriately manage the reservoir, the pressure must be optimal. To reduce the possibility of reservoir damage, this will allow the operators to optimize production rates and set a maximum pressure.

Generated Water to Injected Water Ratio: This ratio considers the full potential of the EOR process. This increases the water injection's efficacy and the reservoir's sweeping efficiency. Thus, a higher ratio corresponds to an improved water flood displacement of oil, which leads to improved recovery and sound reservoir management.

The voidage replacement coefficient describes the amount by which a reservoir replaces fluids that have been lost. It might be seen as a ratio showing how much fluid, typically water, has been put into the reservoir with respect to how much fluid, typically oil, has been produced. A good VRC indicates high efficiency of sweeping and oil displacement by the injected fluid.

A good fluid replacement therefore helps in the pressure maintenance of the reservoir, besides maintaining the performance that is added to the successful recovery.

Meanwhile, the water voidage production rate is a value that indicates the water amount produced from the reservoir, and the unit of the input value may be in the form of millions of stock tank barrels per day (Mstb/day). Moreover, the data support reservoir techniques management and the governance of water handling features through water production behavior provision.

Watercut (fraction): The proportion of water in total generated fluids (oil, gas, and water) is known as the watercut. It is a crucial performance indicator for surveillance and reservoir management. A high watercut might also be a sign of coning or possibly water break-through, which would indicate the need for suitable mitigation and decrease oil recovery.

The water injection rate is expressed in terms of millions of stock tank barrels per day (Mstb/day) to increase recovery. Sustaining the oil displacement, sweep efficiency, and reservoir pressure requires proper control of the water injection rate. It affects oil recovery, total production optimization, and reservoir performance.

Water rate is defined as the quantity of water produced daily in Mstb/day from the reservoir. Measurement of water rates is, therefore, key for purposes of control of water handling facilities, monitoring the behavior of the reservoir and to ensure that practices of the reservoir management give results as desired.

Ratio of water to oil (stb/stb): Such a ratio is given in stb/stb and it measures the proportion in volume between water and oil as produced. It is an anticipated amount of water encroachment into the reservoir, and it helps access the effectiveness of water flooding or injection. Water to oil ratio must be kept in such a manner that high oil recovery is achieved while controlling the water output at the same time.

The number of the injection wells now onstream and thus contributing to the reservoir in accordance with the schemes for improved oil recovery denotes the flowing of the injectors now. A monitored number of injection wells of active injectors would reflect a way of efficiency of reservoir management, technique of injection, as well as displacement of the fluid into the reservoir.

Number of wells flowing in the reservoir are those well activities which are actively producing water, gas, and oil from the reservoir. This, in turn, helps to track the number of active producers, which further aids in assessing how well the reservoirs' function, how efficient the production methods are, and the effectiveness of the whole process of extracting hydrocarbons.

Oil Total (MMstb): Million stock tank barrels (MMstb) is the total amount of oil produced from the reservoir from the beginning of extraction. It helps assess the productivity potential of the reservoir, the efficiency of extraction operations, and the overall oil recovery of the reservoir.

It shows the rate of oil production, which is often expressed in millibars per day (Mstb/day) of stock tank oil. Monitoring shows the amount of oil flowing out of the reservoir and is used to assess how well the production is performing. Rates must be closely monitored to gauge the reservoir's production and enhance extraction processes.

It supports the operators' ability to track changes in oil production over time and makes well-informed choices about reservoir management.

These characteristics are essential to oil extraction because of knowledge about reservoir behavior, fluid dynamics, efficient oil production techniques, and reservoir management information. Therefore, by monitoring and analyzing the data, the operators would be able to manage the reservoir pressure, optimize the pace of production, displace fluid to an ideal level, and assure the highest possible recovery of oil. These characteristics also support the management of water handling facilities, assessment of improved oil recovery techniques, and justifiable decision-making for long-term, reasonably priced oil extraction activities.

4.2 Eclipse Simulator Instruments

Thus, the main goal of this research project is to maximize the oil recovery factor by the collaborative use of reservoir simulation software Eclipse and machine learning techniques. Water Injection Rate (Mstb/day), Water Rate (Mstb/day), Water-Oil Ratio (stb/stb), Average Pressure (psi), Oil Total (MMstb), Ratio of Produced Water to Injected Water, Voidage Replacement Coefficient, Water Voidage Production Rate (MRB/day), Watercut fraction, and Water Injection Rate (Mstb/day) are all significant parameters that can be found in the foundational dataset directly from an active oil field.

To improve the dataset's richness for machine learning applications, generate a few more data points using the Eclipse simulator. With the addition of this dataset, an integrated view of reservoir dynamics—based on characteristics like Watercut and Oil Rate, among others—is provided, which is crucial to the predictive analytics approach.

An inventive approach by machine learning to maximize oil recovery that might further advance reservoir management is blending the simulated inputs with the field's real-world data. Stressing the correctness and dependability of the recently generated datapoints which were produced after the variance was made—while contrasting the outcomes with those obtained from the Eclipse reservoir simulation software is also essential.

Thus, the arduous validation of the simulated data has confirmed the accuracy and consistency of these data with complex dynamics of the real oil field. The study has initiated a strong foundation for the support of machine learning applications in the future through an increasing legitimacy of the dataset through cross-referencing and validation to the machine-generated data and Eclipse outcomes. By means of a compelling comparison between the validation of data through simulation and field data collection, it will serve as the foundation for a more thorough and accurate examination of the optimization techniques meant to maximize the oil recovery factor of the reservoir in question.

The Eclipse 100 simulation research used a very strict set of protocols to fully model and analyze the behavior of the reservoir. The effort began with the creation of a reservoir simulation model called Eclipse (GRID), which symbolized the reservoir using cuboid structures.

The following step was giving each of these three layers the necessary characteristics, such permeability and porosity. In addition, Eclipse (OFFICE) has incorporated other qualities, such as reservoir fluid and a rock characteristic. This helped confirm the model's accuracy because the estimated values were closely compared to the Fluid Initially in Place.

After that, a Base Case simulation was performed without injecting wells to determine the fundamental recovery dynamics. To demonstrate the validity of the model, the resulting production profile was then verified against manually produced numbers. Subsequent simulations were conducted to incorporate injection wells into the network's evolution.

The recovery efficiency of the oil was determined by carefully compiling and calculating the production profiles of the variables in each case.

Notably, simulations of water injection, polymer injection, and surfactant injection were conducted in this comparative investigation. They are designed to make it possible to assess and compare the efficacy of different state-of-the-art rehabilitation techniques. The Eclipse software framework is the foundation of this customized reservoir simulation technique, which illustrates how the reservoir reacts to injections of surfactant, water, and polymers.

4.3 Python

Python's ease of use, adaptability, and rich library support have made it a potent tool in the fields of machine learning and data analysis. In my thesis, I developed and assessed machine learning models for enhanced oil recovery (EOR) using Python and several of its essential libraries. I now present Python and a few of the key libraries—scipit-learn, pandas, numpy, and seaborn—that were crucial in my research.

CHAPTER 5: RESULTS AND DISCUSSION

5.1 Water Injection

Table 5.1: Correlation score among variables for water injection

Variable	Correlation
Oil Rate, Mstb/day	1.000000
Avg. Pressure, psi	-0.710908
Gas Rate, MMscf/day	0.828644
Ratio of Produced Water to Injected Water	-0.682225
Voidage Replacement Coeff	-0.213405
Water Voidage Production Rate, MRB/day	-0.564025
Watercut, fraction	-0.687774
Water Rate, Mstb/day	-0.564118
Water-Oil Ratio, stb/stb	-0.683195
Oil Total, MMstb	-0.609514

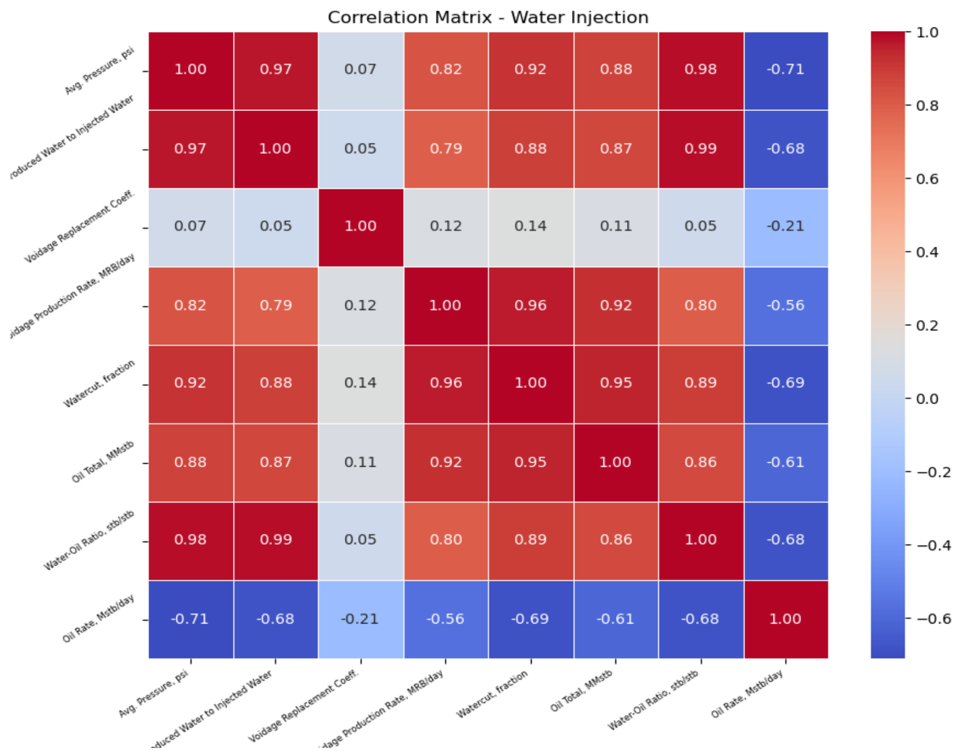


Figure 5.1: Correlation Matrix – Water Injection

The relationship of the aim variable, "Oil Rate, Mstb/day," with the rest of the variables is plotted on the correlation matrix. All those variables, except "Gas Rate, MMscf/day," indicate amongst themselves a quite strong positive relationship with rising gas prices—a link to increasing oil prices. While on the other hand, "Watercut, fraction" and "Water-Oil Ratio, stb/stb" show a significantly negative relationship, meaning that the amount of oil detected in the produced fluid is lower due to the high water cut that has accumulated in it. Similarly, "Oil Total, MMstb" and "Ratio of Produced Water to Injected Water" tend to neutralize relations with the oil rate. It can be understood that the daily rate of oil will reduce to 0 as the ratios of produced water to injected water and cumulative oil output raise. The small negative correlation was associated with "Avg. Pressure, psi," "Water Voidage Production Rate, MRB/day," and "Water Rate, Mstb/day. This highlights that correlations do not connote causation, and that future research will have to be done in order to further comprehend the intricacy of the interrelations between these variables.

5.2 Polymer Injection

Table 5.2: Correlation score among variables for polymer injection

Variable	Correlation
Oil Rate, Mstb/day	1.000000
Ratio of Produced Water to Injected Water	-0.635428
Voidage Replacement Coeff.	-0.180369
Water Voidage Production Rate, MRB/day	-0.528755
Watercut, fraction	-0.661683
Water Injection Rate, Mstb/day	0.290715
Water Rate, Mstb/day	-0.528448
Water-Oil Ratio, stb/stb	-0.675748
Avg. Pressure, psi	-0.548023
Oil Total, MMstb	-0.583521

The relationships between the dependent variable "Oil Rate, Mstb/day" and the other variables in the dataset will be shown in the correlation matrix. It is evident from the strongly negative correlations of other variables like "Watercut, fraction," "Ratio of Produced Water to Injected Water," and "Water-Oil Ratio, stb/stb" that one has high volumes of production water, high ratios of injected water to production water, and high water-oil ratios in general at low oil rates. The following factors show a slightly negative association with the oil rate: average pressure, psi,"Water Rate, Mstb/day" and "Water Voidage Production Rate, MRB/day." Lower oil rates will result from greater average pressure, water voidage production rate, and water rate, respectively. It's crucial to keep in mind that correlation does not imply causality.

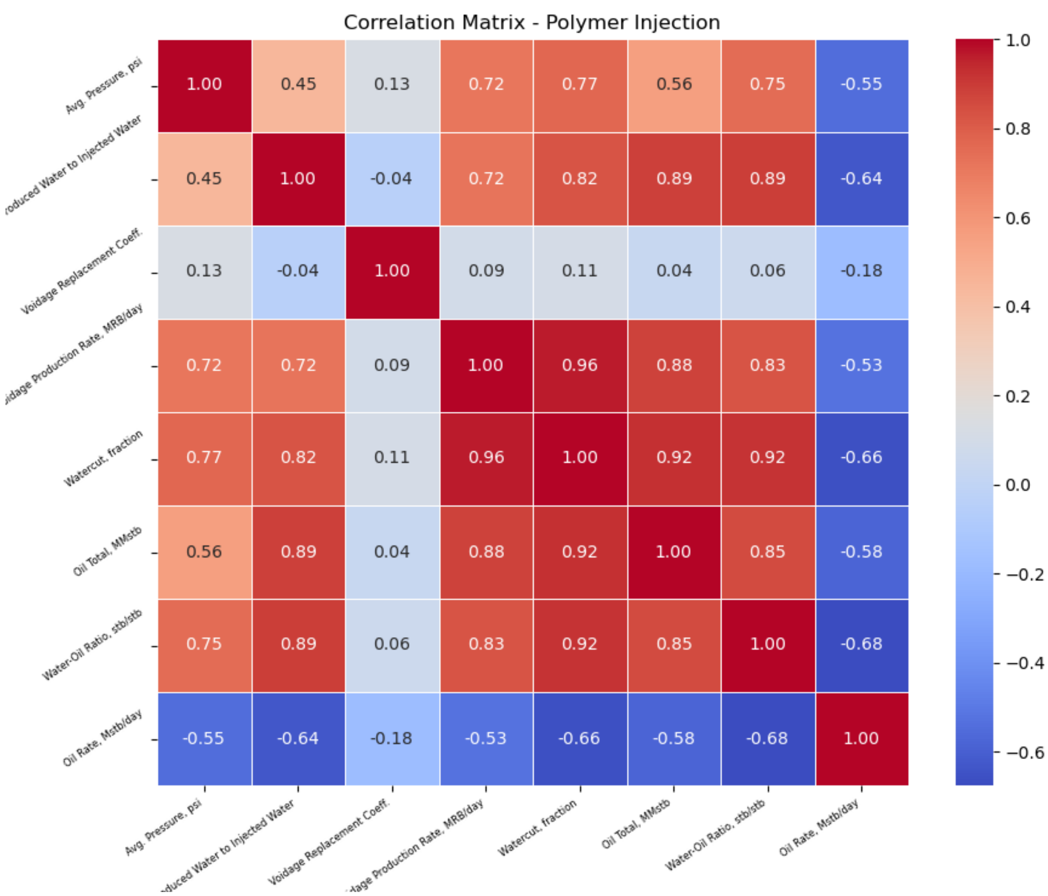


Figure 5.2: Correlation Matrix – Polymer Injection

5.3 Surfactant Injection

The correlation matrix helps in understanding relationships that the goal variable, "Oil Rate, Mstb/day," has with other variables in the dataset. Variables such as "Watercut, fraction," "Ratio of Produced Water to Injected Water," and "Water-Oil Ratio, stb/stb" show severe adverse correlations, signifying that higher contents of water in the produced fluid, higher ratios of the produced water to the injected water, and a higher water-oil ratio infer lower oil rates. The relationship "Water Injection Rate, Mstb/day" and "Oil Rate" is very slightly positively related, i.e., an increase in water injection rates will see an increase in oil rates. Moreover, "Avg. Pressure, psi," "Water Voidage Production Rate, MRB/day," "Water Rate, Mstb/day," and "Oil Total, MMstb," have negative correlation with the oil rate, meaning that as the value of the variables increases, it may affect negatively the amount of oil produced.

Table 5.3: Correlation score among variables for surfactant injection

Variable	Correlation
Oil Rate, Mstb/day	1.000000
Avg. Pressure, psi	-0.292970
Ratio of Produced Water to Injected Water	-0.608177
Voidage Replacement Coeff.	-0.145768
Water Voidage Production Rate, MRB/day	-0.550968
Watercut, fraction	-0.641185
Water Injection Rate, Mstb/day	0.397057
Water Rate, Mstb/day	-0.550580
Water-Oil Ratio, stb/stb	-0.627518

Oil Total, MMstb	-0.503486

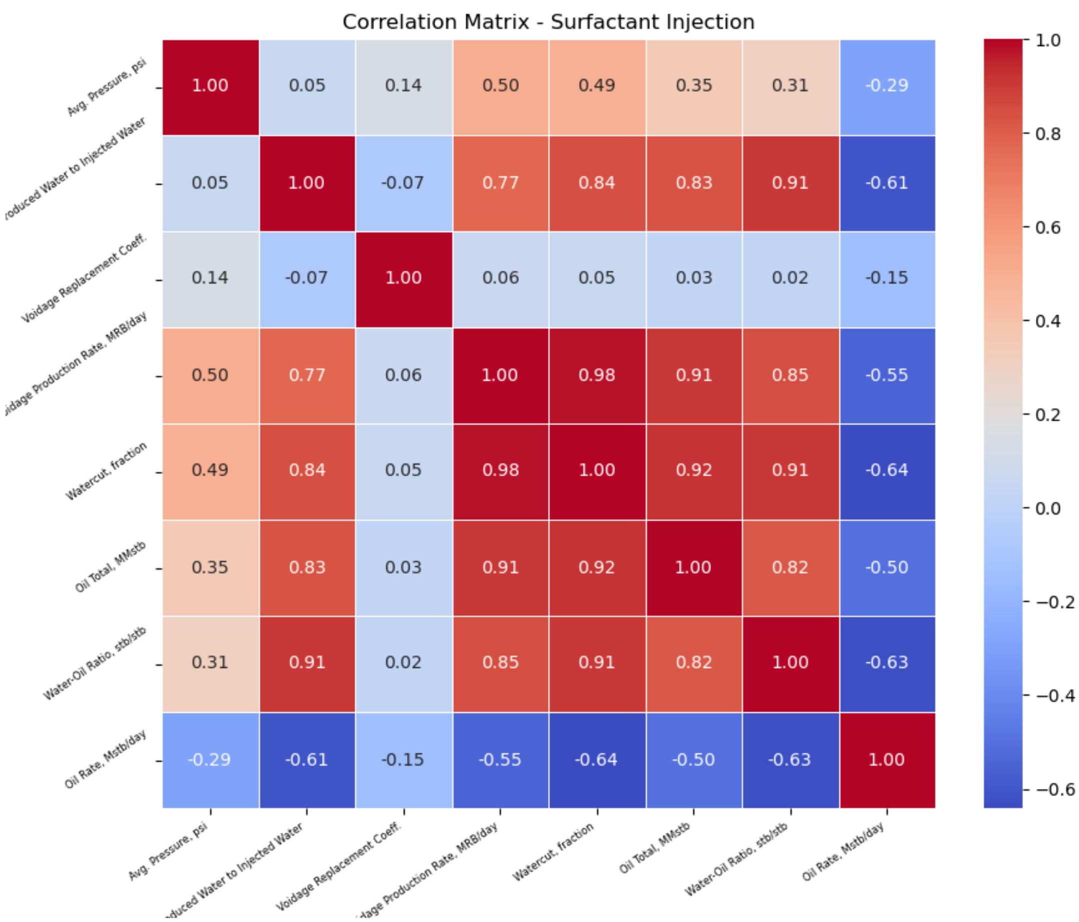


Figure 5.3: Correlation Matrix – Surfactant Injection

5.4 Performance evaluation criteria

MSE and R2 were used to assess preprocessing strategies. The datasets were split into training (70%) and testing (30%) datasets at random to pre-process and deeply model the data. The developed models were validated using 5-fold cross-validation in order to reduce overfitting and data waste. The Regression model toolbox of each model provided ranges for the hyperparameter adjustment, which were subsequently optimised with GA's assistance. The models were then constructed and tested using these hyperparameters. The performance of the validation phase over the modelling process was assessed using the average values of the statistical indices.

Each final machine learning model's prediction performance was evaluated using the following two criteria: 1) Root-mean-squared error (RMSE) and 2) Coefficient of determination (R2). Here are the R2 and RMSE equations:

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i^{exp} - Y_i)^2}{\sum_{i}^{n}(Y_i^{exp} - Y_i^{exp})^2} \ldots (16)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i^{exp} - Y_i)^2} \ \ldots (17)$$

$$Y_i^{exp}\ is\ experimental\ value,\ \ Y_i\ is\ predicted\ data\ and\ n$$

$$represent\ number\ of\ test\ sample$$

5.5 Prediction Performance

Oil rate per day was predicted using random forest regressor, gradient boosting regressor and KNN regressor. This study includes three different enhanced oil recovery methods which are water injection, polymer injection and surfactant injection. All three models given above applied to each dataset. Therefore, nine models have been applied which successfully predicted the oil rate per day.

The KNN Regressor showed respectable prediction accuracy with an R-squared ranging from 0.80 to 0.89 on testing sets, along with greater RMSE and MSE, when the regression models for surfactant, water, and polymer injections were evaluated. In every situation, the Random Forest and Gradient Boosting Regressors performed better than the KNN; on both the training and testing sets, they demonstrated decreased RMSE and MSE and achieved R-squared values that were almost equal to 1. In particular, Random Forest demonstrated exceptionally high accuracy, with an R-squared of 1.00 on the training set for injections of polymers and surfactants. These ensemble approaches demonstrated their applicability for forecasting oil rates in diverse injection scenarios by successfully capturing intricate linkages in the data. Based on the results, Random Forest and Gradient Boosting are better options for real-world applications as they indicate less overfitting. Nonetheless, it is important to take into account the interpretability of ensemble models. Additional research into feature significance and other regression strategies may also help to increase performance and comprehension. The training RMSE, MSE and R-squared values along with their testing values are compared below in the table.

Table 5.4: Injection type, Model and their evaluation parameters

Data set	Model	Training RMSE	Testing RMSE	Training MSE	Testing MSE	Training R Squared	Testing R-Squared
Surfactant Injection	RF Regressor	0.61	1.38	0.37	1.91	0.99	0.98
Surfactant Injection	GB Regressor	1.28	1.55	1.65	2.41	0.99	0.98
Surfactant Injection	KNN Regressor	3.81	4.71	14.48	22.21	0.87	0.80
Polymer Injection	RF Regressor	0.56	1.45	0.31	2.11	0.99	0.98
Polymer Injection	GB Regressor	1.06	1.53	1.12	2.34	0.99	0.98
Polymer Injection	KNN Regressor	3.44	4.02	11.80	16.18	0.89	0.83

Water Injection	RF Regressor	0.56	1.45	0.31	2.11	0.99	0.98
Water Injection	GB Regressor	1.03	1.51	1.07	2.29	0.99	0.98
Water Injection	KNN Regressor	3.42	4.01	11.70	16.07	0.89	0.85

The table uses performance metrics like Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and R-Squared (R^2) for both training and testing data to compare the performance of three regression models: Random Forest (RF), Gradient Boosting (GB), and K-Nearest Neighbours (KNN)—across three datasets: Surfactant Injection, Polymer Injection, and Water Injection. RF exhibits the lowest RMSE and highest R^2 for surfactant injection (0.61 training RMSE, 1.38 testing RMSE, 1 training R^2, 0.98 testing R^2), demonstrating good fit and generalisation. GB (1.28 training RMSE, 1.55 testing RMSE, 0.99 training R^2, 0.98 testing R^2) follows closely behind with minor deviations. In contrast, KNN performs worse as evidenced by its larger RMSE and lower R2 (3.81 training RMSE, 4.71 testing RMSE, 0.87 training R2 and 0.80 testing R2). While GB (1.06 training RMSE, 1.53 testing RMSE, 0.99 training R^2, 0.98 testing R^2) exhibits similar robustness, RF once again performs exceptionally well in Polymer Injection (3.56 training RMSE, 1.45 testing RMSE, 1 training R^2, 0.98 testing R^2), and KNN's higher errors (3.44 training RMSE, 4.02 testing RMSE, 0.89 training R^2, 0.83 testing R^2) imply lower accuracy.

The RF for Water Injection is still high at 0.56 training RMSE, 1.45 testing RMSE, 1 training R^2, and 0.98 testing R^2. GB is next in line at 1.03 training RMSE, 1.51 testing RMSE, 0.99 training R^2, and 0.98 testing R^2, while KNN is once again trailing behind at 3.42 training RMSE, 4.01 testing RMSE, 0.89 training R^2, and 0.85 testing R^2. In general, across all datasets, RF and GB regularly beat KNN with RF typically displaying the strongest performance metrics, demonstrating higher levels of generalisation and predictive power.

5.6 Prediction Curves

Nuanced insights into the performance of the various injection types—water, polymer, and surfactant—are revealed by a thorough examination of regression models. Closely grouped data points around the line of perfect prediction demonstrate the Random Forest Regressor's strong linear correlation with actual values for Surfactant Injection. This suggests that the model successfully captures underlying patterns with little variance and shows high accuracy and dependability. Like this, the Gradient Boosting Regressor exhibits robustness in processing complicated data, with just small variances in its high prediction accuracy for Surfactant Injection. The K-Nearest Neighbours (KNN) Regressor, on the other hand, shows more dispersion, which suggests a higher prediction error and less dependability for this kind of injection.

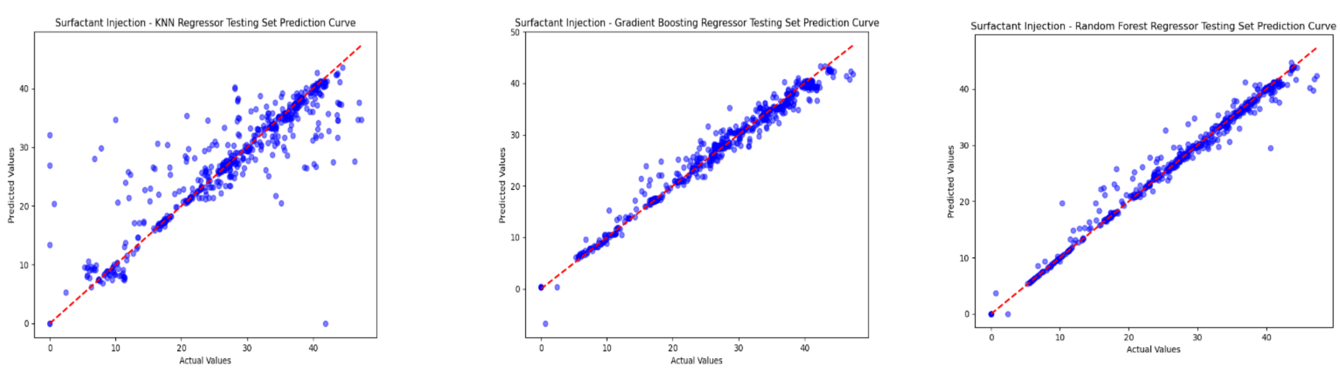
The Random Forest Regressor exhibits remarkable performance in Polymer Injection, as evidenced by data points that are in near alignment with the line of perfect prediction, indicating reliable and precise predictions. Moreover, the Gradient Boosting Regressor performs admirably, exhibiting low variation and strong prediction dependability. The KNN Regressor, however, exhibits substantial variance, especially near the lower end of the real values, suggesting that it has difficulty capturing the intricacy of the data.

The Random Forest Regressor continues to perform well for Water Injection, demonstrating a high degree of correlation between expected and actual values, an indication of trustworthy and precise predictions. High accuracy is also displayed by the Gradient Boosting Regressor, which manages the data well and nearly resembles the line of perfect prediction. But the KNN Regressor exhibits greater dispersion, which suggests that it is less appropriate for this dataset and has a higher prediction error.

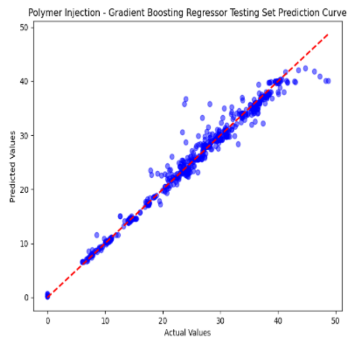
In contrast, the Random Forest and Gradient Boosting Regressors consistently perform well with low prediction variance and good accuracy across all injection types. The robustness of these models is attributed to their capacity to manage complicated datasets and minimise overfitting. The KNN Regressor, on the other hand, exhibits the greatest variability in predictions and has data points that are more widely scattered from the line of perfect prediction, indicating that it is less successful in identifying patterns in the data. Given the larger variance, KNN might not be the optimal option for these datasets, especially when dealing with complicated or high-dimensional data.

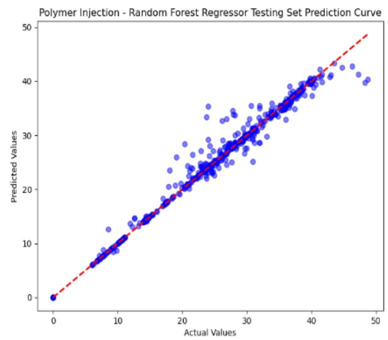
The best models for forecasting injection values are the Random Forest and Gradient Boosting Regressors, both of which exhibit high accuracy and low variation for all injection kinds. Although helpful, the KNN Regressor shows more unpredictability and less precision, suggesting that it might not be the ideal option for these datasets. This research emphasises how crucial it is to choose the right model for a regression task because different models can produce very varied outcomes depending on how well they can identify the underlying patterns in the data. Given that the Random Forest and Gradient Boosting Regressors consistently produce accurate and dependable predictions, it appears that ensemble methods are a good fit for these kinds of injection data. When the relationships between variables are non-linear or involve higher dimensions, the KNN Regressor may have trouble reflecting the complexity inherent in the data, as evidenced by the larger variance seen in its predictions. In the end, the particulars of the dataset and the underlying patterns it exhibits should be considered when selecting a regression model. While more straightforward techniques like KNN could need more tweaking or different strategies to obtain comparable results, ensemble methods like Random Forest and Gradient Boosting provide reliable and accurate answers for datasets like the ones examined here.

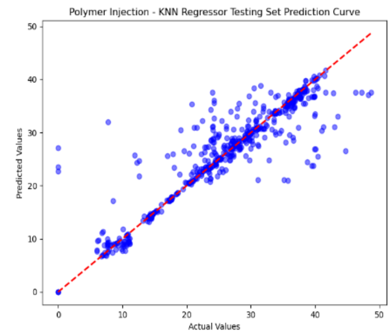
(a) (b) (c)



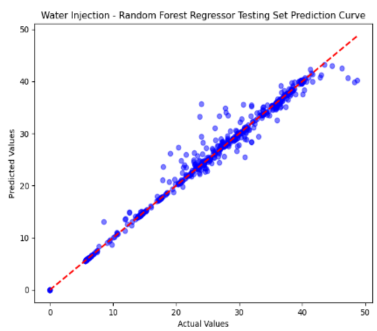
(d) (e) (f)

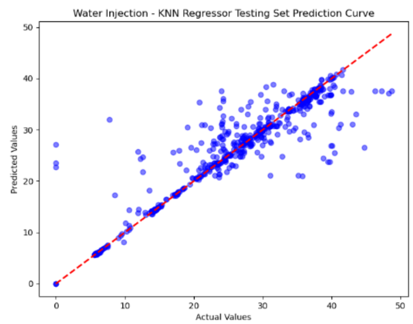






(g) (h) (i)





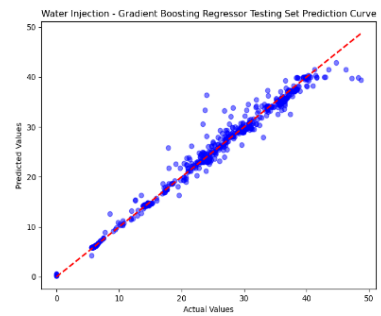


Figure 5.4: Prediction Curves for Surfactant, Polymer and Water Injection

5.7 Feature Importance

The Random Forest Regressor models' feature importances for surfactant, water, and polymer injections make it easier to understand how important each feature is in relation to the forecasts of oil rate. "Water Voidage Production Rate, MRB/day" has the most impact on surfactant injection, scoring 0.2779, closely followed by "Ratio of Produced Water to Injected Water" at 0.2387. Additionally, "Oil Total, MMstb" and "Watercut, fraction" also provide a substantial contribution, scoring 0.1418 and 0.1232, respectively, for relevance. "Water Voidage Production Rate, MRB/day" is the most important factor in water injection, with a value of 0.2481. "Oil Total, MMstb" and "Avg. Pressure, psi" come in second and third, with values of 0.2046 and 0.2043, respectively. "Water Voidage Production Rate, MRB/day" (0.2683), "Oil Total, MMstb" (0.2410), and "Ratio of Produced Water to Injected Water" (0.2208) are the three factors with the largest weight in polymer injection. These results imply that critical factors for all injection methods are the rates of water voidage production and the ratio of generated water to injected water. The bar charts are given below.

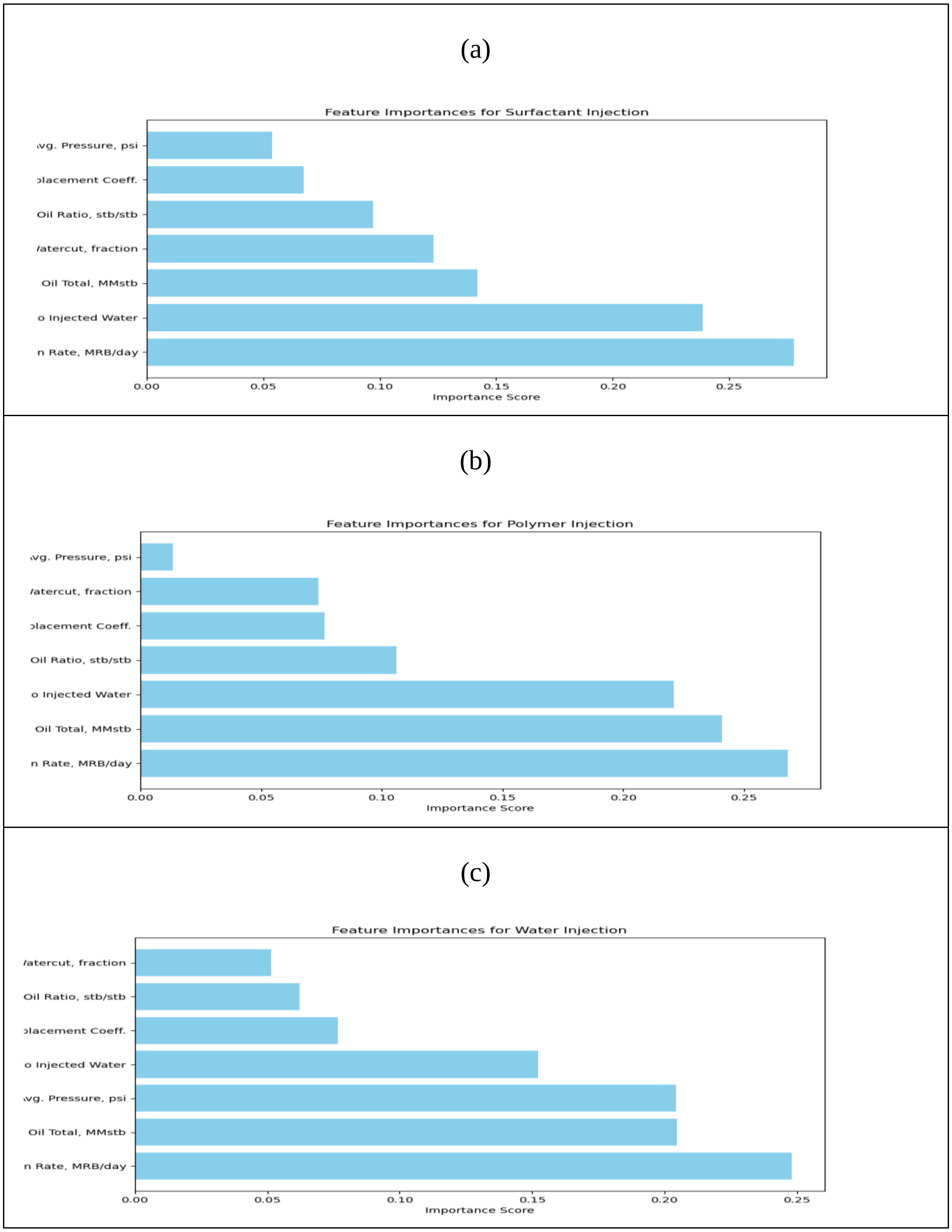


Figure 5.5: Feature Importance for Surfactant, Polymer and Water Injection

5.8 Data Analytics

The three oil recovery techniques have achieved great results, with polymer injection coming in second to surfactant injection in terms of total oil produced, with an average oil rate per day of 27.27 Mstb. At an average oil rate of 25.19 Mstb per day, polymer injection generated more oil than water injection. With an average daily production of 25 Mstb, water injection generated the least oil.

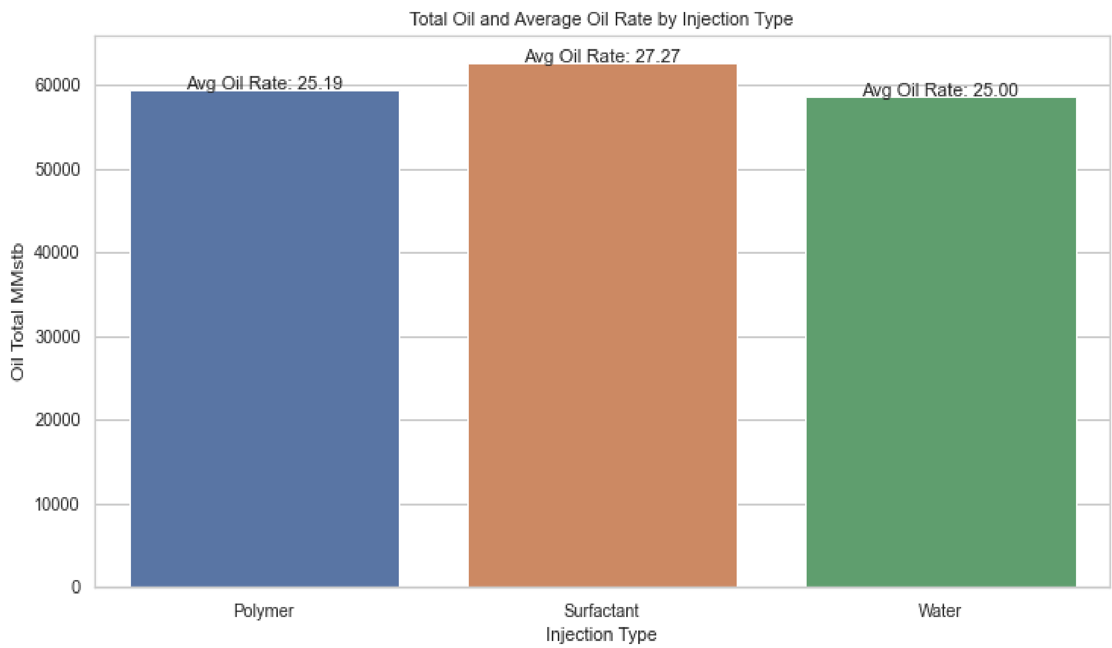


Figure 5.6: Total Oil and Average Oil by Injection Type

The average oil production rates for water injection, surfactant injection, and polymer injection are shown in the bar graph, along with the overall oil output. The total oil output is shown on the y-axis as MMstb (million stock tank barrels), and each bar represents a different type of injection.

A total of around 58,000 MMstb are produced via polymer injection, with an average oil rate of 25.19 MMstb/day. With an average oil rate of 27.27 MMstb/day, surfactant injection produces the most oil overall—more than 60,000 MMstb. With an average oil rate of 25.00 MMstb/day, water injection additionally generates around 58,000 MMstb.

The data unambiguously demonstrates that surfactant injection performs much better in terms of both total oil output and average production rate, while polymer and water injections have comparable total and average production rates. This indicates that among the three techniques for improving oil recovery, surfactant injection is the most successful.

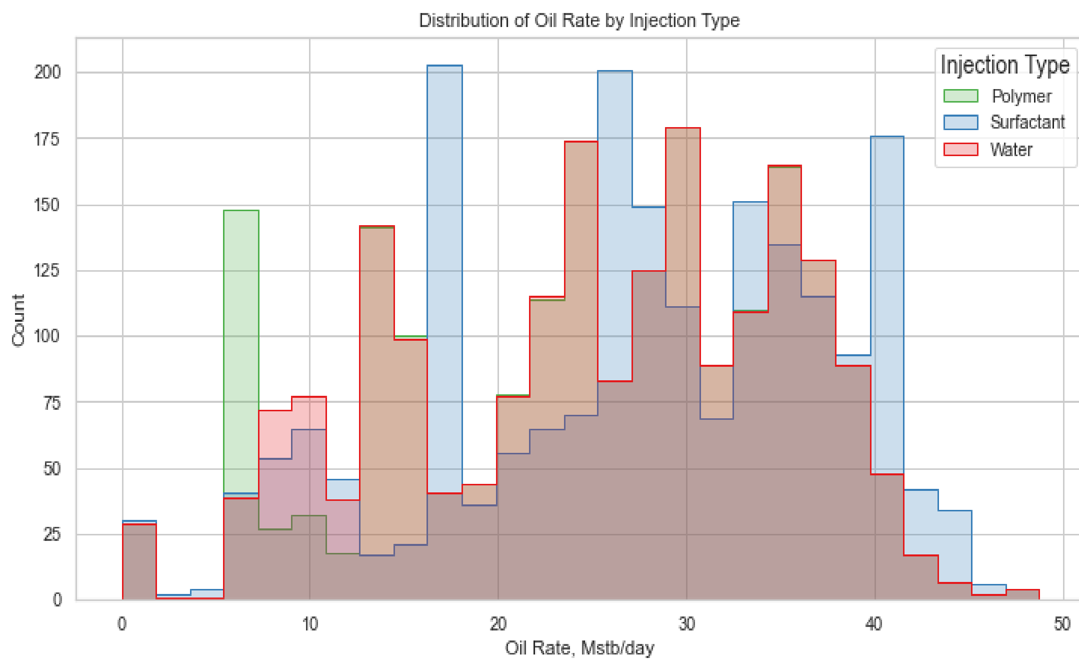


Figure 5.7: Distribution of Oil rate by Injection Type

The distribution of oil production rates, expressed in Mstb/day, for three distinct injection types—polymer, surfactant, and water—is shown in the accompanying histogram. Different colours are used to symbolise each type: water is red, surfactant is blue, and polymer is green. The y-axis shows the number or frequency of wells producing at certain rates, while the x-axis shows the oil production rate in Mstb/day.

Interestingly, compared to polymer and water injections, surfactant injection (blue) has a greater frequency of production rates between 20 and 35 Mstb/day, indicating that it typically results in higher oil production rates.

Water injection (red) has a wider distribution with discernible peaks at 15 and 30 Mstb/day, but polymer injection (green) shows a peak about 10 Mstb/day, indicating a modest production rate. The efficiency of each injection technique is shown in this graphic comparison, with surfactant injection showing to be the most successful overall in terms of obtaining increased oil production rates.

CHAPTER 6: CONCLUSIONS AND FUTURE RECOMMENDATION

6.1 Conclusion

Based on the comprehensive analysis conducted in this thesis, it is evident that the integration of machine learning algorithms with the Eclipse simulator holds great promise for optimizing oil recovery factors. The utilization of surfactant injection, water injection, and polymer injection strategies has been thoroughly investigated with the following key observations derived from the results. Surfactant injection via Random Forest (RF) and Gradient Boosting (GB) regressors achieved significantly lower RMSE and MSE values compared to KNN regressor. RF and GB regressors demonstrated superior predictive performance with high R-squared values, indicating their effectiveness in modeling and optimizing surfactant injection processes. While polymer injection like surfactant injection. RF and GB regressors outperformed KNN regressor, exhibiting lower RMSE and MSE values. High R-squared values obtained from RF and GB regressors signify their robustness in predicting polymer injection outcomes accurately. Water injection on the other hand, employed RF and GB. The regressors showcased remarkable performance in minimizing RMSE and MSE values, surpassing the predictive capabilities of the KNN regressor. The high R-squared values obtained from RF and GB regressors underscore their reliability in optimizing water injection processes effectively.

6.2 Future Recommendations

Overall, the findings suggest that RF and GB regression models are well-suited for predicting the effectiveness of various injection strategies in enhancing oil recovery factors. These results provide valuable insights for reservoir engineers and stakeholders in making informed decisions to maximize oil recovery while minimizing operational costs. Moving forward, further research could explore the integration of additional machine learning algorithms and optimization techniques to enhance the accuracy and efficiency of oil recovery optimization processes.

REFERENCES

[1] Gbadamosi, A.O., Junin, R., Manan, M.A. et al. An overview of chemical enhanced oil recovery: recent advances and prospects. Int Nano Lett 9, 171–202 (2019). https://doi.org/10.1007/s40089-019-0272-8

[2] Kun Guo, Hailong Li, and Zhixin Yu (2016) authored a review titled "In-situ heavy and extra-heavy oil recovery: A review" published in the journal Fuel, Volume 185, pages 886-902. The ISSN is 0016-2361. The DOI for the article is https://doi.org/10.1016/j.fuel.2016.08.047.

[3] Rehman, M.M., Meribout, M. Conventional versus electrical enhanced oil recovery: a review. J Petrol Explor Prod Technol 2, 157–167 (2012). https://doi.org/10.1007/s13202-012-0034-x

[4] Kokal, S., & Al-Kaabi, A. (2010). Enhanced oil recovery: challenges & opportunities. World Petroleum Council: Official Publication, 64, 64-69.

[5] Delamaide, E. (2016) presented a paper titled "Comparison of Primary, Secondary and Tertiary Polymer Flood in Heavy Oil - Field Results" at the SPE Trinidad and Tobago Section Energy Resources Conference in Port of Spain, Trinidad and Tobago. The paper's DOI is https://doi.org/10.2118/180852-MS.

[6] Teixeira, A. F., & Secchi, A. R. (2019). Machine learning models to support reservoir production optimization. IFAC-PapersOnLine, 52(1), 498-501. ISSN 2405-8963. https://doi.org/10.1016/j.ifacol.2019.06.111.

[7] Zhong, Z., Sun, A. Y., Wang, Y., & Ren, B. (2020). Predicting field production rates for waterflooding using a machine learning-based proxy model. Journal of Petroleum Science and Engineering, 194, 107574. https://doi.org/10.1016/j.petrol.2020.107574.

[8] Zhang, R., & Jia, H. (2021). Production performance forecasting method based on multivariate time series and vector autoregressive machine learning model for waterflooding reservoirs. Petroleum Exploration and Development, 48(1), 201-211. ISSN 1876-3804. https://doi.org/10.1016/S1876-3804(21)60016-2.

[9] Calad, C., Gutierrez, F., Pastor, P., & Sarma, P. (2020). Combining Machine Learning with Traditional Reservoir Physics for Predictive Modeling and Optimization of a Large Mature Waterflood Project in the Gulf of San Jorge Basin in Argentina. Paper presented at the SPE Latin American and Caribbean Petroleum Engineering Conference, Virtual, July 2020. Paper Number: SPE-199048-MS. doi: https://doi.org/10.2118/199048-MS. Published: July 20, 2020.

[10] Amirian, E., & Chen, Z. J. (2017). Cognitive data-driven proxy modeling for performance forecasting of waterflooding process. Glob. J. Technol. Optim, 8, 1-8.

[11] Hien, D. H., Hung, L. T., Sang, N. V., Quy, T. X., Sang, N. T., Vy, V. T., ... & Trung, P. N. (2022). MACHINE LEARNING APPROACH TO OPTIMIZE WATERFLOODING WHITE TIGER BASEMENT OILFIELD OFFSHORE VIETNAM.

[12] Qian Sun, Turgay Ertekin (2020). Screening and optimization of polymer flooding projects using artificial-neural-network (ANN) based proxies. Journal of Petroleum Science and Engineering, Volume 185, 106617. ISSN 0920-4105. https://doi.org/10.1016/j.petrol.2019.106617.

[13] Imankulov, T., Kenzhebek, Y., Makhmut, E., & Akhmed-Zaki, D. (2022, September). Using Machine Learning Algorithms to Solve Polymer Flooding Problem. In ECMOR 2022 (Vol. 2022, No. 1, pp. 1-9). European Association of Geoscientists & Engineers.

[14] Jian Hou, Zhen-quan Li, Xu-long Cao, Xin-wang Song (2009). Integrating genetic algorithm and support vector machine for polymer flooding production performance prediction. Journal of Petroleum Science and Engineering, Volume

68, Issues 1–2, Pages 29-39. ISSN 0920-4105. https://doi.org/10.1016/j.petrol.2009.05.017.

[15] Alghazal, M., & Ertekin, T. (2020). Modeling of Deep Polymer Gel Conformance Treatments Using Machine Learning. Paper presented at the Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, UAE, November 2020. doi: https://doi.org/10.2118/203450-MS

[16] Krasnov, F., Glavnov, N., & Sitnikov, A. (2017). Application of Multidimensional Interpolation and Random Forest Regression to Enhanced Oil Recovery Modeling. In Proceedings of the 13th Central & Eastern European Software Engineering Conference in Russia (pp. 2). Association for Computing Machinery. doi: 10.1145/3166094.3166096.

[17] Wang, Y., Hou, J., Cao, X., Wu, G., & Wang, C. (2014). Evaluation of Polymer Flooding Potential Based on Orthogonal Design and BP Artificial Neural Network. Paper presented at the International Petroleum Technology Conference, Kuala Lumpur, Malaysia, December 2014. doi: https://doi.org/10.2523/IPTC-17724-MS

[18] Wang, Y., Liu, H., & Zhou, Y. (2021). Development of a deep learning-based model for the entire production process of steam-assisted gravity drainage (SAGD). Fuel, Volume 287, 119565. ISSN 0016-2361. https://doi.org/10.1016/j.fuel.2020.119565.

[19] Ya Yao, Yue Qiu, Yue Cui, Mingzhen Wei, Baojun Bai (2023). Insights to surfactant huff-puff design in carbonate reservoirs based on machine learning modeling. Chemical Engineering Journal, Volume 451, Part 1, 138022. ISSN 1385-8947. https://doi.org/10.1016/j.cej.2022.138022.

[20] Rashidi-Khaniabadi, A., Rashidi-Khaniabadi, E., Amiri-Ramsheh, B. et al. Modeling interfacial tension of surfactant–hydrocarbon systems using robust tree-based machine learning algorithms. Sci Rep 13, 10836 (2023). https://doi.org/10.1038/s41598-023-37933-0

[21] Shayan Nasr, M., Shayan Nasr, H., Karimian, M. et al. Application of Artificial Intelligence to Predict Enhanced Oil Recovery Using Silica Nanofluids. Nat Resour Res 30, 2529–2542 (2021). https://doi.org/10.1007/s11053-021-09829-1

[22] Daribayev, B., Akhmed-Zaki, D., Imankulov, T., Nurakhov, Y., & Kenzhebek, Y. (2020). Using Machine Learning Methods for Oil Recovery Prediction. In ECMOR XVII, Sep 2020, Volume 2020, p. 1 - 13. European Association of Geoscientists & Engineers. DOI: https://doi.org/10.3997/2214-4609.202035233.

[23] Alvarado, V., & Manrique, E. (2010). Enhanced Oil Recovery: An Update Review. Energies, 3, 1529-1575. doi:10.3390/en3091529.

[24] Alagorni, A., Yaacob, Z., & Nour, A. (2015). An Overview of Oil Production Stages: Enhanced Oil Recovery Techniques and Nitrogen Injection. International Journal of Environmental Science and Development, 6, 693-701. doi:10.7763/IJESD.2015.V6.682.

[25] Ali, S.M. Farouq, & Thomas, S. (1996). The Promise And Problems of Enhanced Oil Recovery Methods. Journal of Canadian Petroleum Technology, 35(07), PETSOC-96-07-07. https://doi.org/10.2118/96-07-07

[26] Palsson, B., Davies, D. R., Todd, A. C., and J. M. Somerville. "Water Injection Optimized with Statistical Methods." Paper presented at the SPE Annual Technical Conference and Exhibition, Denver, Colorado, October 2003. doi: https://doi.org/10.2118/84048-MS

[27] Elmabrouk, S. K., & Mahmud, W. M. (2016). Reservoir Management Strategies for Development of Water Injection Planning Project. In Proceedings of the 2016 International Conference on Industrial Engineering and Operations Management . Detroit, Michigan, USA: IEOM Society International.

[28] Hourfar, F., Bidgoly, H. J., Moshiri, B., Salahshoor, K., & Elkamel, A. (2019). A reinforcement learning approach for waterflooding optimization in petroleum

reservoirs. Engineering Applications of Artificial Intelligence, 77, 98–116. doi:10.1016/j.engappai.2018.09.019

[29] Lee, J.-Y., Shin, H.-J., & Lim, J.-S. (2012). Selection and Evaluation of Enhanced Oil Recovery Method Using Artificial Neural Network. Petroleum Science and Technology, 30(2), 157-164. https://doi.org/10.1080/12269328.2011.10541345

[30] Bahrami, P., Kazemi, P., Mahdavi, S., & Ghobadi, H. (2016). A novel approach for modeling and optimization of surfactant/polymer flooding based on Genetic Programming evolutionary algorithm. Fuel, 179, 289-298. ISSN 0016-2361. https://doi.org/10.1016/j.fuel.2016.03.095.

[31] Sheng, J. J., Leonhardt, B., & Azri, N. (2015). Status of Polymer-Flooding Technology. Journal of Canadian Petroleum Technology, 54(02), 116-126. doi:10.2118/174541-pa\

[32] Tadjer, A., Bratvold, R. B., Hong, A., & Hanea, R. (2021). Application of machine learning to assess the value of information in polymer flooding. Petroleum Research, 6(4), 309-320. ISSN 2096-2495. https://doi.org/10.1016/j.ptlrs.2021.05.006.

[33] Amirian, E., Dejam, M., & Chen, Z. (2018). Performance forecasting for polymer flooding in heavy oil reservoirs. Fuel, 216, 83-100. ISSN 0016-2361. https://doi.org/10.1016/j.fuel.2017.11.110.

[34] Massarweh, O., & Abushaikha, A. S. (2020). The use of surfactants in enhanced oil recovery: A review of recent advances. Energy Reports, 6, 3150-3178. ISSN 2352-4847. https://doi.org/10.1016/j.egyr.2020.11.009.

[35] Chegenizadeh Negin, S., Saeedi Ali, & Quan Xie. (2017). Most common surfactants employed in chemical enhanced oil recovery. Petroleum, 3(2), 197-211. ISSN 2405-6561. https://doi.org/10.1016/j.petlm.2016.11.007.

[36] Breiman, L.: Bagging Predictors. Machine Learning 24, 123–140 (1996)

[37] Amit, Y., Geman, D.: Shape quantization and recognition with randomized trees. Neural Computation 9, 1545–1588 (1997)

[38] Dietterich, T.: An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting and Randomization. Machine Learning, 1–22 (1998)

[39] Ho, T.K.: The random subspace method for constructing decision forests. IEEE Trans. on Pattern Analysis and Machine Intelligence 20(8), 832–844 (1998)

[40] Breiman, L.: Using adaptive bagging to debias regressions, Technical Report 547, Statistics Dept. UCB (1999)

[41] Liu, Y., Wang, Y., & Zhang, J. (2012). New Machine Learning Algorithm: Random Forest. Lecture Notes in Computer Science, 246–252. https://doi.org/10.1007/978-3-642-34062-8_32

[42] Dai, B., Gu, C., Zhao, E., & Qin, X. (2018). Statistical model optimized random forest regression model for concrete dam deformation monitoring. Structural Control and Health Monitoring, 25(6), e2170. https://doi.org/10.1002/stc.2170

[43] Peng Nie, Michele Roccotelli, Maria Pia Fanti, Zhengfeng Ming, Zhiwu Li. (2021). Prediction of home energy consumption based on gradient boosting regression tree. *Energy Reports, 7*, 1246-1255. ISSN 2352-4847. https://doi.org/10.1016/j.egyr.2021.02.006.

[44] Abraham Israeli, Lior Rokach, Asaf Shabtai. (2019). Constraint learning based gradient boosting trees. *Expert Systems with Applications, 128*, 287-300. ISSN 0957-4174. https://doi.org/10.1016/j.eswa.2019.03.011.

[45] Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics, 29(5), 1189-1232. Institute of Mathematical Statistics. Stable URL: http://www.jstor.org/stable/2699986. Accessed: 13/09/2014 03:42.

[46] Nitin Dahiya, Babita Saini, H.D. Chalak. (2021). Gradient boosting-based regression modelling for estimating the time period of the irregular precast concrete structural system with cross bracing. Journal of King Saud University - Engineering Sciences. ISSN 1018-3639. https://doi.org/10.1016/j.jksues.2021.08.004.

[47] Kataria, A., & Singh, M. D. (2013). A Review of Data Classification Using K-Nearest Neighbour Algorithm. International Journal of Emerging Technology and Advanced Engineering, Volume(3), Issue(6), 354. Retrieved from www.ijetae.com, ISSN 2250-2459, ISO 9001:2008 Certified Journal.

[48] Bajaj, P., Ray, R., Shedge, S., Vidhate, S., & Shardoor, N. (Year). SALES PREDICTION USING MACHINE LEARNING ALGORITHMS. MIT School of Engineering, MIT ADT University, Pune, Maharashtra, India.

[49] Buyrukoğlu, S., & Kesriklioğlu, S. (Year). A Comparison of Ensemble and Base Learner Algorithms for the Prediction of Machining Induced Residual Stresses in the Turning of Aerospace Materials. Bitlis Eren Üniversitesi Fen Bilimleri Dergisi / BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE, Volume(11), Number(3), Page Range(861-879). ISSN: 2147-3129, e-ISSN: 2147-3188. DOI: 10.17798/bitlisfen.1130044.

[50] Malakouti, S. M. Ensemble: Mud Loss Prediction at the Marun Oil Field.