

PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
Ministry of Higher Education and Scientific Research

N° Série :/2026

Kasdi Merbah University, Ouargla



Faculty of Hydrocarbons and Renewable Energies and Earth and Universe Sciences
Hydrocarbon Production Department

MASTER THESIS

To obtain the Master's degree Major : Production

Presented By :

**SAIDI DHIKRA MANEL
AMROUN RIHAM**

-THEME-

Machine learning-based Asphaltene Precipitation prediction: Case study- HMD field -

Defended on 11 / 06 / 2026 In front of the examination committee

Jury :

President:	FROUHAT RACHID	Univ. Ouargla
Examiner :	TOUAHRI ABDEL DJEBAR	Univ. Ouargla
Supervisor :	ARBAOUI MOUHAMED ALI	Univ. Ouargla
Co-supervisor:	MERABTI HOCIN	

Academic year 2025/2026

Acknowledgment

I would like to express my sincere appreciation and gratitude to the following individuals and organizations who have played a significant role in the completion of this thesis:

First and foremost, I am deeply grateful to my thesis advisor, Dr. Arbaoui.M Dr. Hocin.M ,D A and r Medjani.F for their guidance, expertise, and unwavering support throughout the entire research process. Their valuable insights, constructive feedback, and continuous encouragement have been instrumental in shaping the direction and quality of this work.

A very special thank you goes to my best friend and binomen. Your friendship, support, encouragement, and countless moments of laughter made this journey much easier. Through stressful days, challenges, and long hours of work, you were always there. I am truly grateful to have shared this experience with you.

Lastly, I want to express my gratitude to any other individuals, not mentioned by name, who have provided support and guidance during the course of this research. Your contributions have made a meaningful impact on this thesis. In conclusion, I would like to emphasize that any errors or shortcomings in this thesis are solely my responsibility. The support and assistance I have received from the aforementioned individuals and organizations have been invaluable, but any mistakes or omissions are mine alone.

Dedication

To the one who gave me life, who stayed up late so I could sleep in peace, who prayed for me in every moment, who sacrificed everything without ever asking for anything in return — my beloved mother, Nadia. You are the reason I stand here today. Every page of this work carries a piece of your love and devotion. No words could ever be enough to express my gratitude for all that you have given me. This achievement is yours before it is mine.

To my dear friend Rahim, who walked beside me through the long nights of hard work and doubt, who offered a word of encouragement when I was ready to give up, who celebrated my small victories as if they were their own. True friendship is rare, and I am grateful every day to have found it in you. You have left a mark on this journey that will never be forgotten. This humble work is the fruit of your love, your support, and your faith in me. I dedicate it to you both, from the bottom of my heart, with infinite gratitude and deepest respect.

Dhikra

Dedication

This thesis is dedicated to my beloved parents, who has been my unwavering source of support, encouragement, and inspiration throughout my academic journey.

To the one who gave me life, who stayed up late so I could sleep in peace, who prayed for me in every moment, who sacrificed everything without ever asking for anything in return — my parents Keltoum and Abdallah .You are the reason I stand here today. Every page of this work carries a piece of your love and devotion. No words could ever be enough to express my gratitude for all that you have given me. This achievement is yours before it is mine it is the fruit of your love.

To my dear friend and my cousin, who walked beside me through the long nights of hard work and doubt, who offered a word of encouragement when I was ready to give up, who celebrated my small victories as if they were their own. True friendship is rare, and I am grateful every day to have found it in you. You have left a mark on this journey that will never be forgotten.

This humble work I dedicate it to you both, from the bottom of my heart, with infinite gratitude and deepest respect.

Riham

Abstract

Asphaltene precipitation is a major flow assurance challenge in the Hassi Messaoud (HMD) field, Algeria, where crude oils containing small amount of asphaltenes can cause severe wellbore plugging. This study develops a machine learning-based predictive framework for asphaltene content using SARA fractionation and PVT data collected from several wells. six supervised algorithms were trained and evaluated via Leave-One-Out Cross-Validation (LOOCV). Gradient Boosting achieved the best performance ($R^2 = 0.969$, $MSE = 0.000584$, $MAE = 0.01347$), outperforming Decision Tree, Random Forest, XGBoost, Linear Regression, and SVR. Feature importance analysis consistently identified the Asphaltene-to-Resin (A/R) ratio as the dominant predictor across all tree-based models, confirming the primacy of colloidal stability in governing asphaltene behavior. The proposed framework offers a reliable, data-driven complement to conventional thermodynamic models for asphaltene management in Algerian oilfields.

Keywords: Asphaltenes, SARA analysis, Machine Learning, Gradient Boosting, Hassi Messaoud, LOOCV, A/R ratio, flow assurance.

Résumé

La précipitation des asphaltènes constitue un défi majeur pour l'assurance de l'écoulement dans le champ de Hassi Messaoud (HMD), en Algérie, où des huiles brutes ne contenant peut d'asphaltènes peuvent provoquer de graves obstructions au niveau du puits. Cette étude développe un cadre prédictif basé sur l'apprentissage automatique, à partir de données d'analyse SARA et de paramètres PVT collectés sur 17 puits. Sept algorithmes supervisés ont été entraînés et évalués par validation croisée LOOCV. Le modèle Gradient Boosting a obtenu les meilleures performances ($R^2 = 0,969$, $MSE = 0,000584$, $MAE = 0,01347$), surpassant l'Arbre de Décision, la Forêt Aléatoire, XGBoost, la Régression Linéaire et le SVR. L'analyse de l'importance des variables a désigné de manière unanime le rapport Asphaltènes/Résines (A/R) comme prédicteur dominant, confirmant le rôle central de la stabilité colloïdale. Le cadre proposé constitue un complément fiable aux modèles thermodynamiques conventionnels pour la gestion des asphaltènes dans les champs pétroliers algériens.

Mots-clés : Asphaltènes, analyse SARA, apprentissage automatique, Gradient Boosting, Hassi Messaoud, LOOCV, rapport A/R, assurance de l'écoulement.

ملخص

يُعدّ ترسّب الأسفلتين من أبرز تحديات ضمان تدفق النفط في حقل حاسي مسعود بالجزائر، إذ يمكن لنفوط خام لا تحتوي إلا على وزناً قليلاً من الأسفلتين تتسبّب في انسداد حادّ على مستوى البئر. تهدف هذه الدراسة إلى بناء إطار تنبؤي قائم على تعلّم الآلة باستخدام بيانات تحليل SARA ومعاملات PVT المستخرجة من 17 بئراً. تمّ تدريب سبعة خوارزميات للتعلّم المُشرف وتقييمها بأسلوب التحقق المتقاطع LOOCV. حقّق نموذج Gradient Boosting أفضل أداء بمعامل تحديد $R^2 = 0.969$ ، وخطأ متوسط مربّع $MSE = 0.000584$ ، وخطأ مطلق متوسط $MAE = 0.01347$ ، متفوّقاً على جميع النماذج الأخرى. وكشف تحليل أهمية المتغيّرات أن نسبة الأسفلتين إلى الراتنجات (A/R) تُمثّل المتنبّي الأكثر تأثيراً، مما يُؤكّد الدور المحوري للتوازن الغرواني في التحكم بسلوك الأسفلتين. يُقدّم الإطار المقترح بديلاً موثقاً وقائماً على البيانات للنماذج الثرموديناميكية التقليدية في إدارة ترسّب الأسفلتين بالحقول النفطية الجزائرية.

الكلمات المفتاحية: الأسفلتين، تحليل SARA، تعلّم الآلة، Gradient Boosting، حاسي مسعود، LOOCV، نسبة A/R، ضمان التدفق.

Table of contents

<i>Acknowledgment</i>	<i>I</i>
<i>Dedication</i>	<i>II</i>
<i>Abstract</i>	<i>IV</i>
<i>Table of contents</i>	<i>VII</i>
<i>List of tables</i>	<i>IX</i>
<i>List of Figure</i>	<i>X</i>
<i>List of abbreviations</i>	<i>11</i>
<i>General Introduction</i>	<i>12</i>
<i>Chapter I:</i>	<i>3</i>
<i>General Characteristics and Properties of Asphaltenes</i>	<i>3</i>
<i>Introduction</i>	<i>4</i>
<i>I.1 Asphaltenes</i>	<i>4</i>
<i>I.2 Resins</i>	<i>5</i>
<i>I.3 Geochemical origins of Asphaltenes</i>	<i>6</i>
<i>I.4 Chemical composition of asphaltenes</i>	<i>6</i>
<i>I.5 Atomic ratio H/C</i>	<i>7</i>
<i>I.6 Molecular Structure of Asphaltenes</i>	<i>8</i>
<i>I.7 Characterization of Asphaltenes</i>	<i>9</i>
<i>I.8 Organic Deposits (asphaltenes)</i>	<i>10</i>
<i>I.9 Asphaltene deposition and flocculation conditions</i>	<i>11</i>
<i>I.10 Influence of temperature and pressure on flocculation</i>	<i>11</i>
<i>I.11 Asphaltene Deposition Mechanism</i>	<i>13</i>
<i>I.11.1 Flocculation</i>	<i>13</i>
<i>I.11.2 The colloidal suspension</i>	<i>13</i>
<i>I.11.3 Aggregation</i>	<i>14</i>
<i>I.12 Deposition of asphaltenes in the reservoir rock</i>	<i>14</i>
<i>I.13 Effect of the deposition on rock wettability</i>	<i>14</i>
<i>I.14 Experimental Determination of Asphaltene Precipitation</i>	<i>16</i>
<i>I.14.1 Sampling Considerations</i>	<i>16</i>
<i>I.14.2 Asphaltene Precipitation Laboratory Testing</i>	<i>16</i>
<i>I.15 Asphaltene Precipitation Modeling</i>	<i>17</i>

I.15.1	<i>Asphaltene Phase Behavior.</i>	17
I.15.2	<i>Thermodynamic Models.</i>	18
I.16	<i>Treatment and control of asphaltenes.</i>	19
I.16.1	<i>Perventive treatment.</i>	19
I.16.1	<i>Curative treateatment.</i>	20
	<i>Conclusion.</i>	22
	<i>Chapter II:</i>	23
	<i>Artificial Intelligence and Machine Learning Fundamentals.</i>	23
	<i>Introduction.</i>	24
II.1	<i>Artificial intelligence.</i>	24
II.2	<i>Machine Learning.</i>	25
II.3	<i>Types of Machine Learning.</i>	25
II.3.1	<i>Supervised Learning</i>	25
II.3.2	<i>Unsupervised Learning</i>	26
II.3.3	<i>Semi supervised</i>	26
II.3.4	<i>Reinforcement</i>	26
II.4	<i>Machine Learning Algorithms</i>	27
II.4.1	<i>Linear Regression</i>	27
II.4.2	<i>Decision Tree.</i>	28
II.4.3	<i>Random Forest.</i>	28
II.4.4	<i>Gradient Boosting</i>	29
II.4.5	<i>Artificial Neutral Networks ANN</i>	29
II.4.6	<i>K-Nearest Neighbors KNN.</i>	30
II.4.7	<i>Support Vector Machine SVM.</i>	30
II.5	<i>The general steps to train ML models.</i>	31
II.5.1	<i>Data collection.</i>	31
II.5.2	<i>Data preprocessing.</i>	31
II.5.3	<i>Data Transformation.</i>	31
II.5.4	<i>Splitting Data.</i>	32
II.5.5	<i>Model Selection</i>	32
II.5.6	<i>Model Training and Validation</i>	33
II.5.7	<i>Model Evaluation</i>	33
II.6	<i>Applications.</i>	34
II.7	<i>Applications in oil and gas.</i>	35

Conclusion.....	37
Chapter III: Machine learning-based Asphaltene Precipitation prediction: Case study- HMD field-	Erreur ! Signet non défini.
Introduction.....	Erreur ! Signet non défini.
III.7 Study area.....	Erreur ! Signet non défini.
III.2 Problematic and objective.....	Erreur ! Signet non défini.
III.3 Methodology.....	Erreur ! Signet non défini.
III.3.1 Data collection.....	Erreur ! Signet non défini.
III.3.2 Data pre-processing.....	Erreur ! Signet non défini.
III.3.4 Data Splitting and validation	Erreur ! Signet non défini.
III.4 Results and discussion	Erreur ! Signet non défini.
III.4.1 Quantitative results.....	Erreur ! Signet non défini.
III.4.2 Qualitative results.....	Erreur ! Signet non défini.
Conclusion.....	Erreur ! Signet non défini.
General conclusion	38
References	41
ANNEXE	Erreur ! Signet non défini.

List of tables

Chapter I. General Characteristics and Properties of Asphaltenes

table I. 1 Elemental Asphaltene Analysis.	7
--	---

Chapter II: Artificial Intelligence And Machine Learning Fundamentals

Table II. 1 Evaluation Metrics for Regression Models.....	34
---	----

Chapter III: : Machine learning-based Asphaltene Precipitation prediction: Case study- HMD field-

Tabel III. 1 Models performance.	Erreur ! Signet non défini.
---------------------------------------	-----------------------------

List of Figure

Chapter I. General Characteristics and Properties of Asphaltenes

Figure I. 1 Different fractions of petroleum (SARA).	4
Figure I. 2 Asphaltene deposit	5
Figure I. 3 Different types of kérogène	6
Figure I. 4 Structural formula of an asphaltene molecule	7
Figure I. 5 Asphaltene solubility in different solvents	8
Figure I. 6 Structure of asphaltenes according to the Yen Model.	9
Figure I. 7 Asphaltene deposit in the production tube	10
Figure I. 8 HMD's asphaltene zones	10
Figure I. 9 Effect of pressure on the flocculation rate at 118°C under reservoir conditions. ...	12
Figure I. 10 Effect of pressure and temperature on deposition (Iranian crude oil)	12
Figure I. 11 Phenomenon of asphaltene flocculation.	13
Figure I. 12 Stages of the structural aggregation of asphaltenes according to the YEN model. (12).....	14
Figure I. 13 Zone affected by asphaltene deposition	14
Figure I. 14 Adsorption of asphaltenes on the surface of rock grains causing a change in wettability.	15
Figure I. 15 Change in the relative permeability of the reservoir rock with the adsorption of asphaltenes on the grain surface.	15
Figure I. 16 Asphaltene phase envelope, pressure-temperature diagram.	18

Chapter II: Artificial Intelligence And Machine Learning Fundamentals

FigureII. 1 Subsets of Artificial Intelligence.	24
FigureII. 2 Classification of ML.	27
FigureII. 3 illustration of linear regression	28
FigureII. 4 Decision Tree structure.....	28
FigureII. 5 Random Forest structure	29
FigureII. 6 Gradient boosting.	29
FigureII. 7 Basic structure of ANN.	30
FigureII. 8 General Workflow for Training Machine Learning Models.	31

Chapter III: Machine learning-based Asphaltene Precipitation prediction: Case study-

HMD field-

Figure III. 1 Hassi Messaoud Well structure.	Erreur ! Signet non défini.
Figure III. 2 Outliers detection.	Erreur ! Signet non défini.
Figure III. 3 Models' performance MSE & MAE comparison.	Erreur ! Signet non défini.
Figure III. 4 R ² score comparison.	Erreur ! Signet non défini.
Figure III. 5 Convergence plot for GBM & XGBoost.....	Erreur ! Signet non défini.
Figure III. 6 Parity plots showing observed versus predicted values for the applied models (without hyperparameter tuning).	Erreur ! Signet non défini.
Figure III. 7 Feature Importance Results from Tree-Based Models.....	Erreur ! Signet non défini.
Figure III. 8 Sensitivity Analysis — Gradient Boosting Model	Erreur ! Signet non défini.

List of abbreviations

Symbol / Term	Description
SARA:	Saturated- Aromatic –Resins – Asphaltenes
H/C ratio	Hydrogen-to-carbon atomic ratio-indicates Aromaticity
wt%	Weight present.
S:	Sulfur content.
O:	Oxygen content.
N:	Nitrogen content.
C:	Carbone content.
H:	Hydrogène content.
HMD:	Hassi Messaoud Field.
EOR:	Enhanced oil recovery.
AI:	Artificial Intelligence.
ML:	Machine Learning.
(x) ;(y):	Variables.
ANNs:	Artificial Neural Networks.
KNN:	K-Nearest Neighbours.
MAE:	Mean Absolute Error.
SVM:	Support Vector machine.
RMSE:	Root Mean Square Error.
R^2	Coefficient of determination.
GR:	Gamma Ray.
T:	temperature.
P:	pressure.
RF:	Random Forest Regressor.
GBM:	Gradient Boosting Machine.
CII:	Colloidal Instability Index.
LOOCV:	Leave-One-Out Cross Validation.
B:	is the number of decision trees.
SVR:	Support Vector Regression.
GBM:	Gradient Boosting Machine.
RBF:	A Radial Basis Function.

General Introduction

In today's world, there is a significant rise in the demand for energy, particularly in industrial sectors, which indicates that we must accept a greater dependency on energy than before. While renewable energy sources can meet some of these requirements, humans still heavily rely on hydrocarbon resources as their primary source of energy. As a result, it is important to acknowledge that the significance of fossil fuels, especially oil and gas, has increased in recent years.

Within the framework of the development techniques for the exploitation and refining of oil, it is important to know with precision the structure and behavior of its various components. Indeed, crude oil is composed of four families (SARA): saturated hydrocarbons, aromatics, resins, and asphaltenes. This distribution is very broad; each part of crude oil is also composed of a range of molecules with varying structures and properties. The exploitation of this source of fossil energy and hydrocarbons is one of the pillars of the contemporary industrial economy because oil provides almost all liquid fuels, and this exploitation requires drilling equipment and surface installations. Very expensive, so the rock that retains the hydrocarbons is the subject of particular attention in order to ensure the longest possible life.

The Hassi Messaoud field in Algeria exemplifies this issue well. Although this crude oil it induces significant clogging at the wellbore, indicating that asphaltene concentration alone is not a dependable predictor of deposition risk. This paradox highlights the intricate, multifaceted character of precipitation, influenced by interrelated elements such as reservoir pressure, temperature, crude oil composition represented by SARA fractions, and the presence of injected fluids or metal ions. **(1)**

Historically, classical thermodynamic methods, including Flory-Huggins's theory and equation-of-state models, were the principal instruments for characterizing asphaltene phase behaviour. Although these methodologies provide significant theoretical insights, they exhibit considerable limits when implemented under actual reservoir circumstances, especially around and below the bubble point pressure, and in the context of chemically heterogeneous crude systems.

In the last twenty years, the rapid advancement of artificial intelligence and machine learning has provided new avenues for addressing nonlinear, data-intensive challenges. In contrast to thermodynamic models, machine learning algorithms do not need predetermined physical assumptions; rather, they derive predicted patterns directly from empirical data. An expanding

General Introduction

corpus of research has validated the pre-eminence of these data-driven methodologies for predicting asphaltene precipitation. Research using techniques like Support Vector Machines, Gradient Boosting, and more recently XGboost has continually surpassed traditional models, attaining elevated coefficients of determination and minimal prediction errors on extensive, varied datasets.

This study introduces a machine learning-based prediction system for asphaltene, developed in Python, building on previous advancements. The model is trained on an extensive dataset that includes essential physicochemical parameters—such as SARA fractions, pressure, temperature, bubble point pressure, and API gravity—aimed at providing dependable and generalizable predictions across diverse reservoir and production conditions also used to assess the risk of asphaltene precipitation during oil flow.

The main purpose of studying asphaltene precipitation is to:

Develop accurate models of ML to better predict and manage asphaltene behaviour under different operating conditions.

Predict and prevent operational issues caused by asphaltene precipitation, such as blockages, equipment fouling, and reduced oil quality.

For this thesis, there are both experimental and modelling approaches.

Chapter I:
General Characteristics and Properties of Asphaltenes

Introduction

A crude oil is a complex mixture of compounds that are distributed according to the traditional liquid chromatography techniques in (Figure I.1).

- ✓ Saturated hydrocarbons: These are aliphatic compounds (n-alkanes, iso-alkanes, cycloalkanes).
- ✓ Aromatic compounds: These are compounds that are physically and chemically distinct from paraffins and aldehydes. They contain one or more atoms that are similar to those of benzene.
- ✓ Resins and asphaltenes: composed of polycyclic molecules with high molecular weight, containing oxygen, sulfur, and azote.

Asphaltic rocks are heavy, viscous, and occasionally solid rocks that are composed of over 50% asphaltenes, resins, and aromatics. They contain less than 10% paraffin and less than 20% naphthene, and they are distinguished by a higher sulfur content, roughly 1%, which can reach 9% in certain raw materials⁽¹⁾:

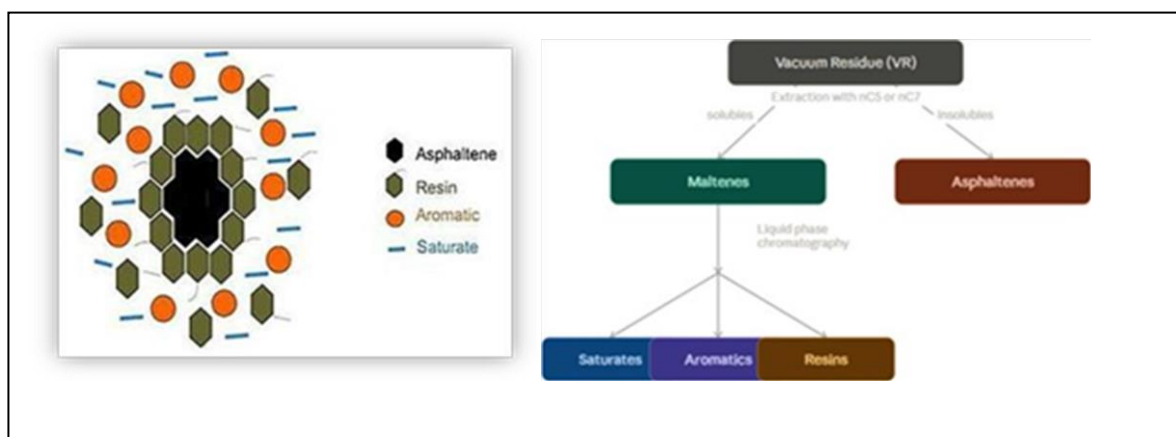


Figure I. 1 Different fractions of petroleum (SARA). (1)

The primary ingredient of petroleum is asphaltene, although it may also come from a variety of carbon-based sources, such as bituminous sand, petroleum-based steel, carbon, or even houille. For this reason, they are named by their source and have a vast variety of chemical compositions and physico-chemical properties.

I.1 Asphaltenes

The solid bodies of asphaltene are dark, brittle, and easily broken. The term "asphalté" was first used in 1837 when the Frenchman Boussingault used it to describe bitumes from the French countryside. When the components of asphalté are separated by distillation, the name

"asphaltene" refers to the solids that are soluble in ethanol but insoluble in alcohol, while "petrol" refers to the volatile components that are soluble in ether. (2)



Figure I. 2 Asphaltene deposit

According to the current theory, petroleum asphaltenes are fractions with high molecular weight that are insoluble in paraffins with low molecular weight (n- heptane, n-pentane), but soluble in certain aromatic solvents like toluene, xylene, or benzene. They are defined as a class of hydrocarbons with an amorphous structure and an aromatic base. They are present in crude petroleum as dispersed colloidal particles. The central part of asphaltene micelles consists of a high molecular weight component. The latter is surrounded and peptized by neutral resins of smaller mass and paraffinic hydrocarbons of low molecular weight, which are Resins. (3)

I.2 Resins

Resins are substances that are soluble in solvents such as pentane or heptane but insoluble in propane. The molecule is composed of a condensed polycyclic ring, composed of aromatic and cyclanic rings and sulfur- or nitrogen-containing heterocyclic. Resins are considered homologues of asphaltenes, with a lower molecular weight and a less condensed structure. Resins are soluble in the alkanes that cause asphaltenes to precipitate but are insoluble in ethyl acetate. Many authors believe that resins can solvate asphaltenes and stabilize their suspensions in solution. Murzakov et al. studied the micellar stability of (asphaltene/benzene) mixtures and found that the addition of resins (2–8 wt%) reduces the amount of asphaltenes precipitated by n-heptane. Lian et al. studied n-pentane flocculation from the (asphaltene/toluene) mixture and their results confirmed Murzakov's conclusion (2).

I.3 Geochemical origins of Asphaltenes

The term "asphaltene" refers to chemically elevated hydrogen pieces that are soluble in petroleum that were expelled from the rock and migrated during the catagenese phase. The presence of nickel, vanadium and chromium reflects the pH and Eh (reduction potential) of the depositional environment of the source rock (Lewan, 1980; 1984). The relationship between these two parameters is used in petroleum exploration to define the oil correlations and determine the rock's petroleum potential.

The structure of kerogen components (insoluble organic matter in rocks) is poorly understood. These are structures of very high molecular weight comprising aromatic cyclane nuclei (S, O, and N). (4)

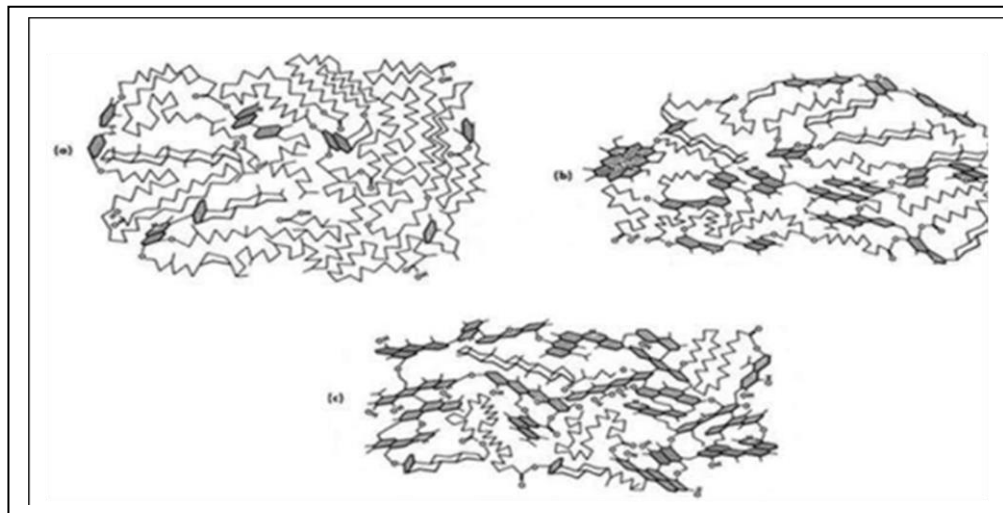


Figure I. 3 Different types of kérogène (4)

I.4 Chemical composition of asphaltenes.

The chemical composition of asphaltic raw materials has been ascertained through a variety of analytical methods, including distillation fractionation, SARA analysis, and elemental analysis. Asphaltenes are essentially composed of Carbon and Hydrogen molecules. Furthermore, they also contain heteroatoms such as Sulfur, Nitrogen (N), and Oxygen in significant amounts, while metals such as Nickel and Vanadium are also present in small quantities and in variable proportions. (Figure I.4). (5)

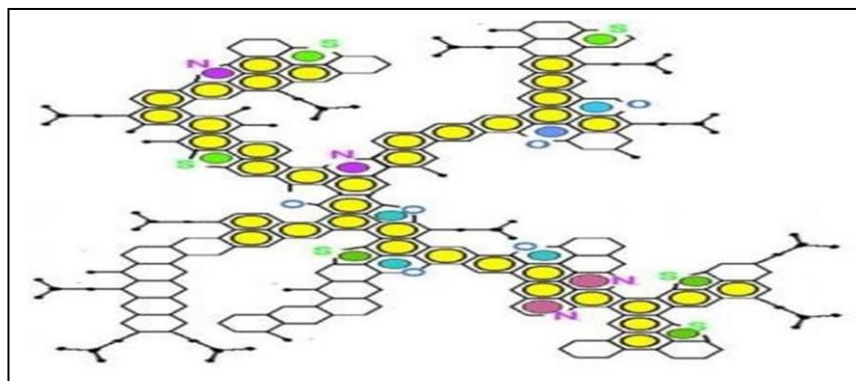


Figure I. 4 Structural formula of an asphaltene molecule

The following table provides the fundamental compositions of asphaltene:

table I. 1 Elemental Asphaltene Analysis.

Elemental composition	Weight (%)
C	82 ± 3%
H	8.1 ± 0.7%
N	0.6 – 3.3%
S	0.3 - 10.3%
O	0.3 – 4.9%
V	1 – 1200 (PPM)
NI	1 – 150 (PPM)

I.5 Atomic ratio H/C

This is the ratio between hydrogen and carbon atoms. It is a simple and quick way to determine the degree of aromaticity of an oil or its fractions, such as asphaltenes and resins. For example, the aromatic compound "benzene" C_6H_6 has 6 hydrogen atoms and 6 carbon atoms with a ratio of $H/C = 1.0$,

while the paraffin "n-hexane" C_6H_{14} has an H/C ratio of 2.3. Most asphaltenes have H/C ratios ranging from 1.0 to 1.3. Elemental analysis shows that asphaltenes are composed mainly of carbon and hydrogen; the atomic ratio H/C is relatively constant, with a value close to 1, indicating that asphaltenes are highly aromatic compounds, as shown in the previous table. (6)

Solubility of the asphaltene mixture in various solvents:

The following histogram shows the results of the solubility test conducted on a sample of asphaltene concrete from the HMD field at ambient temperature (21°C):

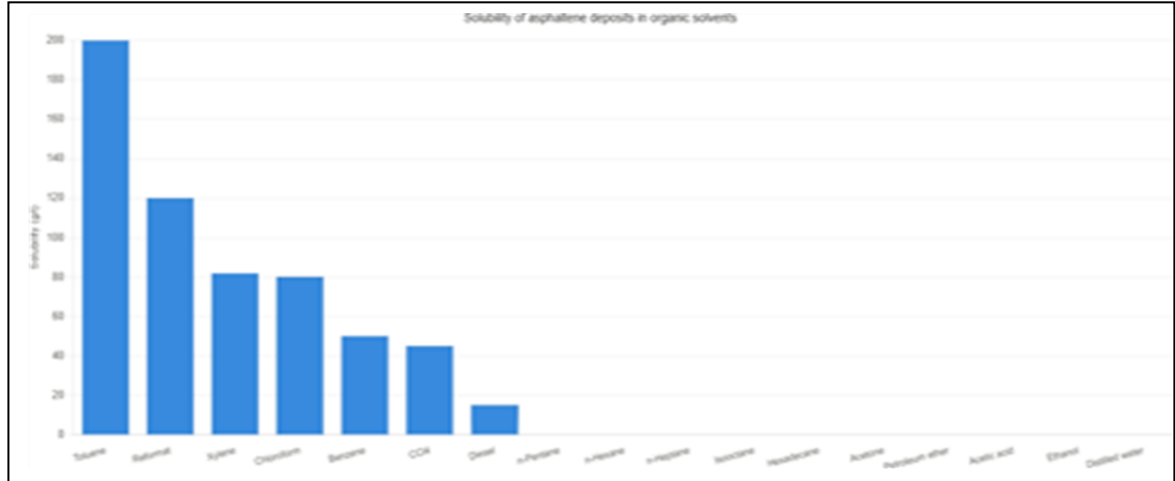


Figure I. 5 Asphaltene solubility in different solvents (7).

The asphaltenes are polar fractions that are soluble in aromatic solvents including benzene, toluene, xylene, and ethylbenzene, but insoluble in alkanes with low molecular weights, such as nC₃, nC₄, till nC₁₂. Studies have classified solvents according to their solubility power on asphaltenes, as follows: Benzene < Toluene < Ortho-xylene < Tetrahydrofuran < Pyridine.

I.6 Molecular Structure of Asphaltenes.

The asphaltene belongs to a class of complex macromolecules. The structure of asphaltenes and resins can be classified on two scales:

- ✓ The macrostructure
- ✓ The microstructure

Particles (more than one molecule) may generate sufficient amounts of micelles in the macromolecular framework. keeping in mind that there are fewer micelles in the micromolecular structure.

The means of analyzing asphaltenes in their natural medium (crude oil or refining products) are scarce; for this reason, most analyses have been carried out on previously precipitated asphaltenes. For this reason, several models have been proposed to describe the structure of asphaltenes. The most widely used is the Yen model (1972). Yen described asphaltenes as molecules and developed a three-dimensional model. The structure of asphaltenes is shown

in Figure 1.6. According to Yen, asphaltene molecules in solution are associated into "particles" (several molecules), and the particles can form larger aggregates (agglomerates) called "micelles."

- **Sheet:** It constitutes the elementary asphaltene molecule with a molecular weight of the order of 500 to 1000 AMU. The sheet is the poly-condensed aromatic core substituted with aromatic chains of naphthenic nuclei composed of 4 to 7 molecules.
- **Particle :** 4 to 6 elementary sheets.
- **Micelle:** grouping of 2 to 3 particles.
- **Aggregate:** the molar mass is of the order of 100,000 to 1,000,000 AMU and is made up of several particles and micelles.(3)

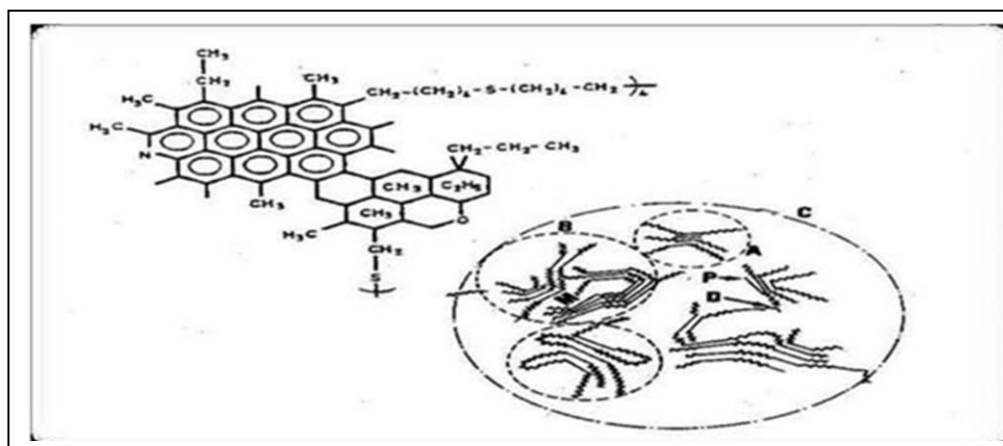


Figure I. 6 Structure of asphaltenes according to the Yen Model. (3).

(A) Particle, (B) Micelle, (C) Aggregate, and (M) Metal atom, (D) Weak bond.

I.7 Characterization of Asphaltenes

- Complexes of aromatics.
- High molecular weight.
- Insoluble in n-alkane.
- Soluble in aromatic solvents like toluene and benzene.
- Their chemical formula is $C_{74}H_{87}NS_2O$ + metals.

The main properties of the oil from the Hassi Messaoud (HMD) field, based on analyses of samples from 52 wells, are as follows: (8)

- Density from 0.8 to 0.835 ($^{\circ}\text{API}$: 37.9 to 45.3).
- Asphaltene fraction in oil: 0.08 to 0.2%

- Average molecular weight of asphaltenes: 840 to 4900

I.8 Organic Deposits (asphaltenes)

Asphaltic deposition is one of the most prevalent issues associated with crude oil production. Some of its consequences include reservoirs damage, the reduction in productivity of petroleum pipelines, and the clogging of pipes and production facilities.

These deposition issues occur not only during the extraction of heavy oils with a high asphaltene content (South American oil contains approximately 10% asphaltene), but also in the case of low asphaltene oils, such as those from the Hassi Messaoud field. (2)



Figure I. 7 Asphaltene deposit in the production tube (2)

In fact, serious problems were found in zones 1 (1A, 1B, and 1C) and 23, which are located in the western part of the field (figure I.7), where some wells were completely closed due to colmatage problem caused by solid asphaltene. Therefore, it is necessary to study the factors, conditions, and causes of this problem.

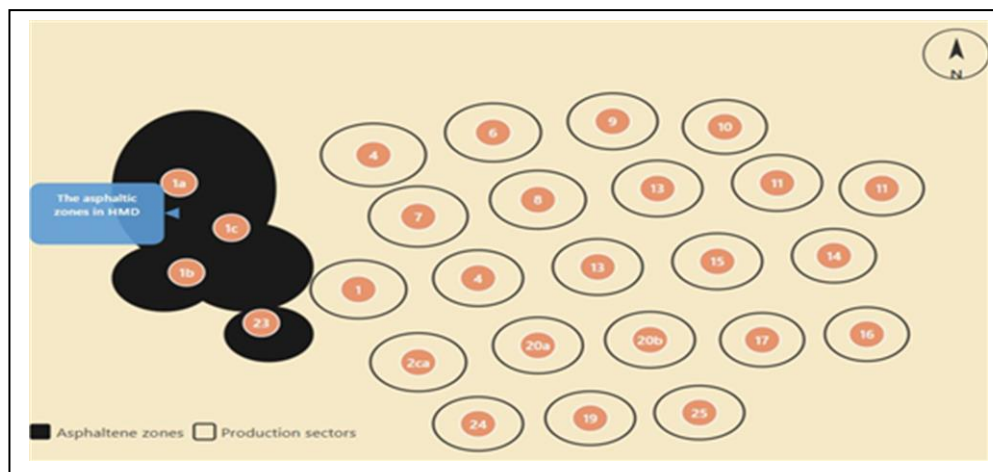


Figure I. 8 HMD's asphaltene zones (9)

I.9 Asphaltene deposition and flocculation conditions.

- Asphaltene precipitation is caused by changes in temperature, pressure, and chemical composition of crude oil.
- A sufficiently low temperature and/or pressure, especially when the pressure is below the bubble point.
- The incompatibility between the formation and injection water.
- The precipitation of asphaltene has also been caused by the injection of CO₂ to increase productivity (EOR).
- Activities like drilling, completion, acid stimulation, and hydraulic fracturing may also cause precipitation in the area around the wells.
- The asphaltene content of a crude oil does not imply the existence of the problem; some reservoirs in Central America and the Middle East, notably the Venezuelan Boscan crude, contain more than 17% asphaltenes but present no problems, in contrast to the Algerian Hassi-Messaoud crude, which contains only 0.1% asphaltenes yet causes a major plugging problem at the well level..
- The impact of ions and particulates of metal-based chlorides on asphaltene has been known for a long time. In particular, ferric chloride, which has relatively high solubility in hydrocarbon environments, causes rapid flocculation of asphaltenes through:
 - Changes in the electrophoretic properties of the oil-asphaltene mixture (displacement under the influence of an electrical field).
 - The precipitation of asphaltene is caused by the development of complexes with metal ions that are suitable for forming a metal-heteroatom liaison.

I.10 Influence of temperature and pressure on flocculation.

The results of experiments carried out to define the behavior of asphaltenes with pressure changes on different crude oils are presented in the following figures. It was observed that above the bubble point, the decrease in pressure causes an increase in the deposition rate. The decrease in pressure can destabilize the oil and cause flocculation of asphaltenes. When the pressure drops, the density of the oil decreases due to the expansion of the oil volume, which leads to a decrease in the concentration of resins in the medium. Subsequently, there will be a migration of resin molecules from the surface of the asphaltene particles toward the solution, creating strong interaction forces between the charged particles, which causes flocculation and aggregation of the particles to form the deposit. (3)

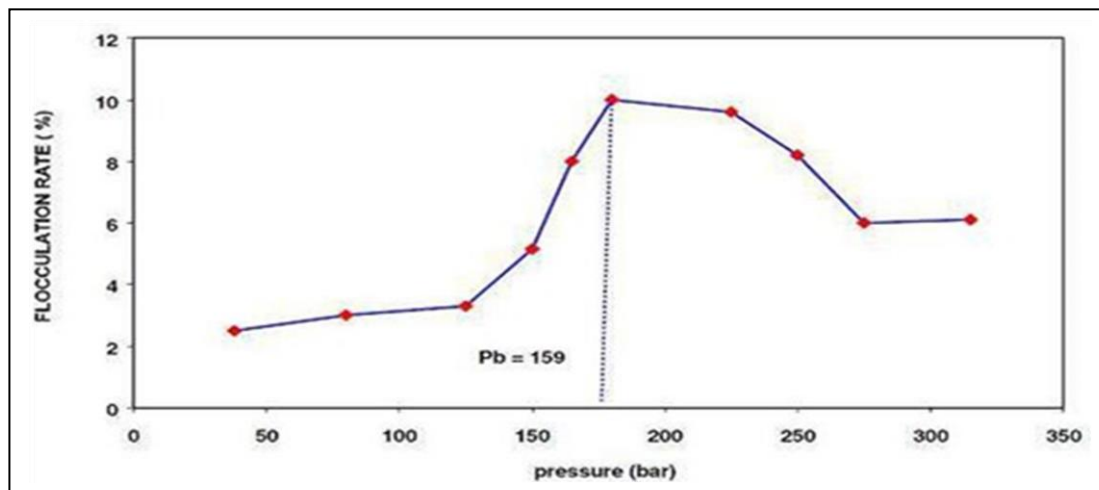


Figure I. 9 Effect of pressure on the flocculation rate at 118°C under reservoir conditions.

When the pressure decreases below the bubble point, the light components of the crude oil separate from the oil to form the gas phase. During this phase, a decrease in the deposition rate was observed, which explains a re-dissolution of the asphaltene deposit in the oil. The figures show that the maximum amount of deposits is observed at a pressure near the bubble point.

Some crude oils may not exhibit this behavior. If the particles are very well stabilized in the oil, either by a sufficient quantity of resins or aromatics, the decrease in pressure would have no effect on deposition.

Another effect that was observed in these experiments is the effect of temperature on precipitation. In our case, the stability of the crude oil increases as a function of temperature, as the latter causes the disorganization of the arrangement of the basic particles, hence a decrease in the size of the aggregates. This can be explained by the change in the molar volume of the crude oil components, which leads to a change in the solubility of the oil. (3)

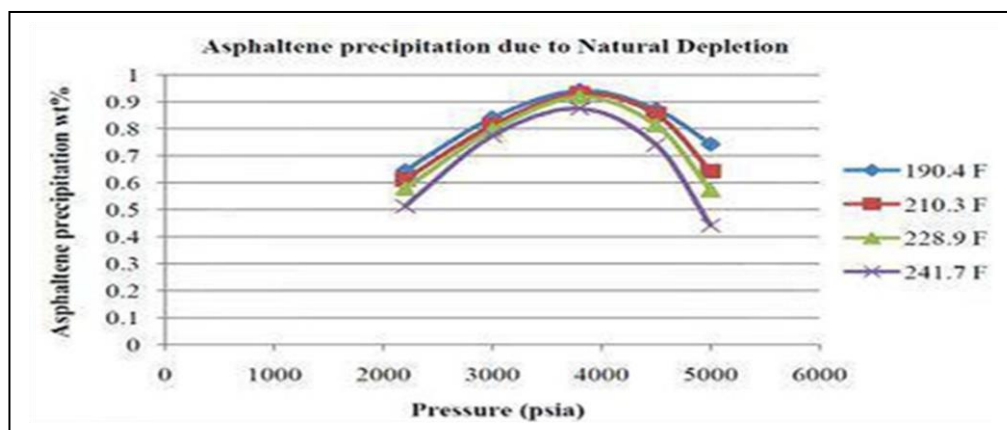


Figure I. 10 Effect of pressure and temperature on deposition (Iranian crude oil) (3).

I.11 Asphaltene Deposition Mechanism

Asphaltenes in solution exist in the form of aggregates. The elementary aggregates, composed according to the Yen model of 5–7 molecules, are formed even in very dilute solutions. These elementary aggregates further combine based on the quality of both the asphaltenes and the solvent used. Asphaltene particles aggregate under the effect of a medium-range attractive interaction potential. The colloidal particles diffuse and meet, but certain contacts between particles lead to aggregation. As the flocculation threshold is approached, aggregation accelerates and large structures are formed, which settle after a time ranging from a few minutes to a few hours depending on the quality and quantity of the flocculant.

I.11.1 Flocculation

In general, flocculation occurs when asphaltene particles or micelles associate to form larger flocs, which are the primary entities that initiate the ultimate precipitation process and are responsible for a significant increase in viscosity. (10)

I.11.2 The colloidal suspension

The presence of a specified amount of resin in the solution prevents asphaltene deposition. The stability of these colloids depends on the concentration of resins in the solution, the surface area occupied by resin particles, and the balance between the amount of resin in the solution and on the particle surface. The change in their concentration in oil causes a change in the number of heavy particles adsorbing on the surface. (11)

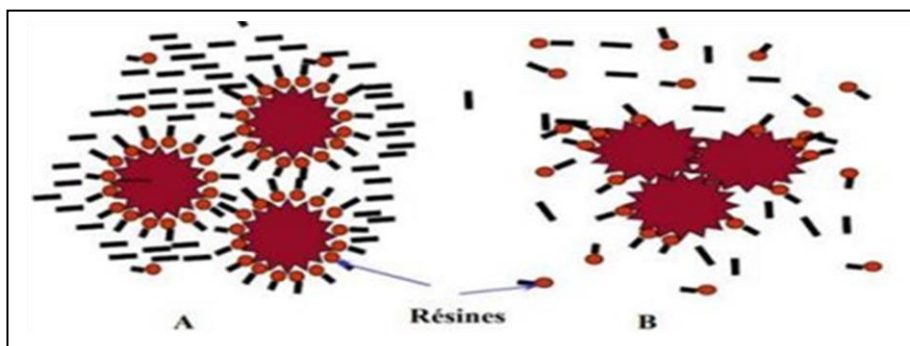


Figure I. 11 Phenomenon of asphaltene flocculation.

A: Steric stabilization of asphaltenes by resins.

B: Destabilization and flocculation of asphaltenes.

I.11.3 Aggregation

The concentration of the peptizing agent in the oil can decrease to a point where the adsorbed quantity is not sufficient to cover the entire surface of the asphaltenes. This leads to the release of active sites and allows asphaltene particles to come together, grow larger and flocculate. (10)

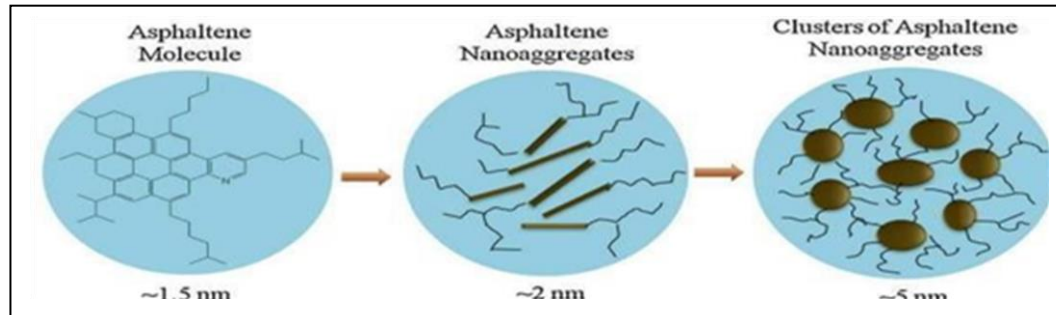


Figure I. 12 Stages of the structural aggregation of asphaltenes according to the YEN model. (12).

I.12 Deposition of asphaltenes in the reservoir rock.

The greater the velocity of fluid flow through the rock pores, the greater the risk of pore blockage by asphaltenes particles. In a producing formation, the area where a high rate of flow is observed is the area of the wellbore where the fluid converges at the perforations. Consequently, the asphaltene deposition issue caused by the decrease in pressure and the electrochemical effect is expected to be concentrated at the wellbore. (13)

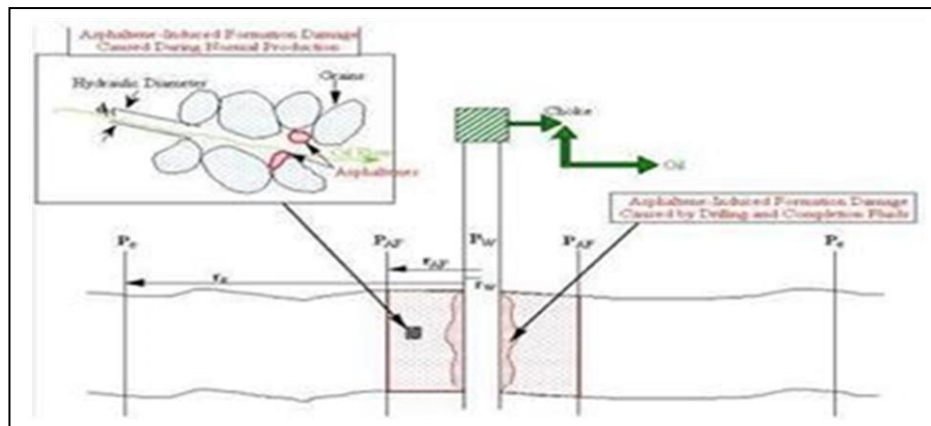


Figure I. 13 Zone affected by asphaltene deposition (14).

I.13 Effect of the deposition on rock wettability.

The deposition of asphaltenes on the pore surface causes changes in the properties of the reservoir rock. These changes are mainly due to the adsorption of particles onto the surface of

the grains constituting the matrix, which causes modification of the porosity, permeability and wettability of the rock.

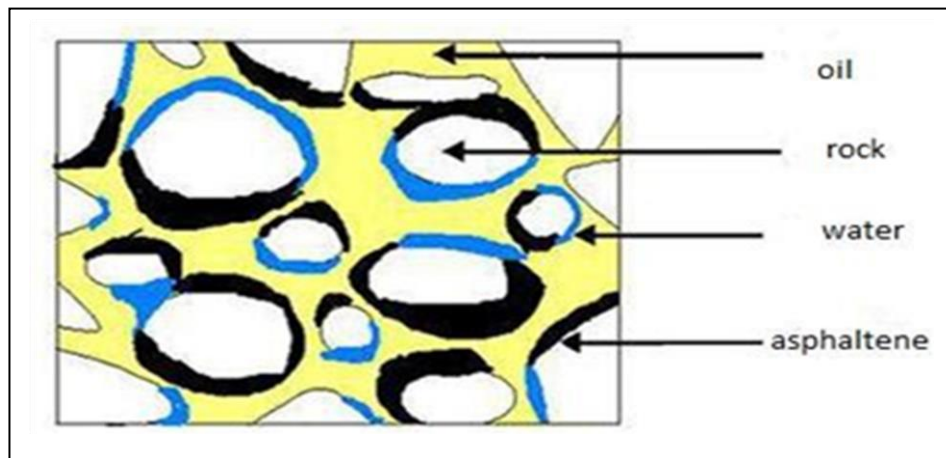


Figure I. 14 Adsorption of asphaltenes on the surface of rock grains causing a change in wettability.

The alteration of reservoir rock wettability by asphaltene deposition has been the subject of investigation for many authors. The adsorption of particles onto the grain surface is a primary factor in changing the wettability of a rock; for example, a water-wet rock becomes oil-wet after adsorption.

Morrow (1990) showed that the change in wettability affects relative permeability, connate water saturation, capillary pressure, dispersion and electrical properties. The alteration of relative permeability and maximum saturation has an important effect on the drainage process. Morrow analyzed the results of water drainage tests carried out on core samples and found that the low wettability of the rock constituted a constraint to efficient fluid drainage and good recovery. The modeling of reservoir rock wettability is a complex subject and requires a great deal of research. Data and models concerning wettability alteration by asphaltene deposition and its effect on the efficiency of fluid displacement in the reservoir are not available. (14)

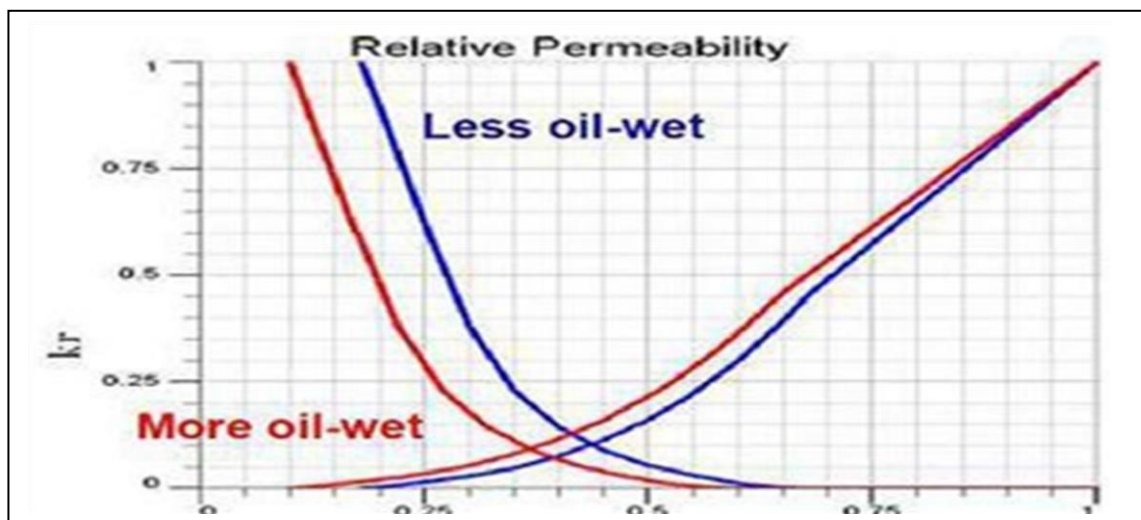


Figure I. 15 Change in the relative permeability of the reservoir rock with the adsorption of asphaltenes on the grain surface. (14)

I.14 Experimental Determination of Asphaltene Precipitation

I.14.1 Sampling Considerations

The most suitable samples for asphaltene testing are single-phase, uncompressed samples collected directly from the wellbore using a pressure-compensated single-phase sampler. Prior to analysis, samples are conditioned in the laboratory at reservoir pressure and temperature for approximately five days, a process that promotes the dissolution of any solids that may have settled during transportation. It is worth noting that reconstituted separator samples are not appropriate for solid-related laboratory analyses, as the significant pressure and temperature drops experienced by the reservoir fluid between the formation and the test separator are likely to cause partial solid deposition and loss within the production line or the separator itself.

I.14.2 Asphaltene Precipitation Laboratory Testing

A laboratory testing program for asphaltene precipitation is typically designed to provide one or more of the following types of information:

- Determination of **the asphaltene onset pressure** at reservoir temperature and at additional temperatures representative of the wellhead and surface separation facilities. The data obtained from these tests are used to construct the Asphaltene Precipitation Envelope (APE) on a pressure-temperature (P-T) phase diagram. This category of test is commonly referred to as the isothermal depressurization test.
- **Fluid compatibility refers** to the maximum concentration of a fluid or solvent that can be injected into the reservoir fluid without triggering asphaltene precipitation. This assessment is carried out through a titration test, typically performed under reservoir conditions to evaluate the compatibility between the reservoir fluid and an injected solvent — such as natural gas, CO₂, or N₂ — or between two reservoir fluids when commingled during co-production operations.
- **Chemical and inhibitor screening tests** are conducted when asphaltene precipitation cannot be avoided under normal production scenarios. These tests follow a procedure similar to the titration and depressurization tests described above, with the key difference that the reservoir fluid is pre-treated with a chemical inhibitor prior to testing, allowing evaluation of the inhibitor's effectiveness in suppressing precipitation.

- **Filtration tests** are designed to quantify the amount of precipitated solids by flowing the oil sample through a filter — typically with a pore size of 0.2 μm — to capture and measure any solids exceeding the filter threshold.

The laboratory equipment used to conduct the above tests, designed and manufactured by Oil-phase-DBR, includes the following instruments:

- **Solids Detection System (SDS):** A laser-based technology for detecting the onset of organic solid precipitation — including wax and asphaltene — from reservoir fluids.
- **High Pressure Microscope (HPM):** Enables direct visual observation of solid particle formation under high-pressure conditions.
- **Particle Size Analysis (PSA):** A software-based system that scans digital HPM photomicrograph images in real time at a high sampling rate, simultaneously discretizing and quantifying particle size distributions.

I.15 Asphaltene Precipitation Modeling.

I.15.1 Asphaltene Phase Behavior.

The phase behavior of asphaltenes is conventionally represented by the Asphaltene Precipitation Envelope (APE) on a pressure-temperature (P-T) diagram, as illustrated in Figure I. This envelope defines the thermodynamic boundaries within which asphaltene precipitation occurs during reservoir depletion.

At pressures above the upper asphaltene boundary, the system exists as a single liquid phase in which both asphaltenes and light components remain fully dissolved in the crude oil. As reservoir pressure declines during production, asphaltene molecules begin to aggregate, forming stacked structures that grow progressively larger until they detach from the bulk fluid — marking the onset of a two-phase system comprising solid asphaltenes and liquid crude oil. This transition point defines the upper asphaltene boundary.

As pressure continues to decrease toward the bubble point, the concentration of precipitated asphaltene particles reaches its maximum. At the bubble point, all light components depart the liquid phase to form a vapor phase, giving rise to a three-phase system consisting of solid asphaltenes, liquid oil, and light vapor components. The departure of these light components effectively increases the solvating power of the remaining heavy oil phase, which becomes a more favorable solvent environment for asphaltenes. Consequently, asphaltene particles begin to redissolve back into the oil phase as pressure continues to fall below the bubble point, until the last asphaltene particle re-enters solution at the lower asphaltene boundary.

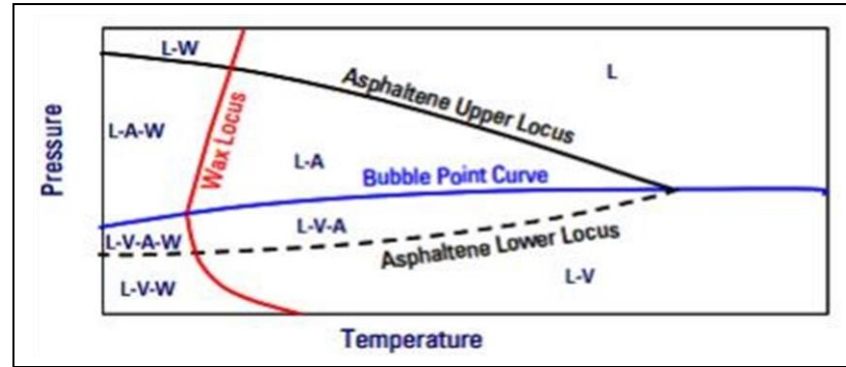


Figure I. 16 Asphaltene phase envelope, pressure-temperature diagram.

I.15.2 Thermodynamic Models.

Given the inherently ambiguous definition of asphaltenes and the considerable complexity of their molecular structures, experimental approaches for accurately and cost-effectively characterizing asphaltene precipitation remain limited in their scope and applicability. This has motivated the development of thermodynamic relationships capable of predicting asphaltene precipitation behavior as a function of independent variables such as pressure, temperature, and crude oil composition. Although these pressure-volume-temperature (PVT) relationships have been the subject of extensive research over several decades, no universally applicable model has yet been established that can reliably predict asphaltene behavior across all crude oil systems and solvent injection scenarios.

The theoretical frameworks proposed for modeling asphaltene precipitation fall primarily into two categories: **colloidal models** and **solubility models**.

The **colloidal model** treats asphaltenes as micelles stabilized in the crude oil medium through polar-polar interactions with resin molecules. Within this framework, asphaltene precipitation is initiated when the resin-to-asphaltene concentration ratio falls below a critical threshold, disrupting the protective colloidal coating and causing asphaltenes to aggregate and separate from the bulk phase.

The **solubility model**, by contrast, assumes that asphaltenes exist in the oil phase in a partially or fully dissolved state. This approach defines a solubility parameter that governs the thermodynamic equilibrium of asphaltenes in solution, relying on weak van der Waals

intermolecular interaction forces rather than the stronger polar-polar interactions central to the colloidal model. Precipitation

occurs when the solubility parameter of the oil phase falls below that of the asphaltene fraction, driving phase separation.

I.16 Treatment and control of asphaltenes

Asphalt flooding is a major issue for the petroleum industry. Additionally, because of the large molecular weight of flocculated asphaltene and their tendency to adsorb on solid surfaces, this asphaltene may develop irreversible deposits that are difficult to remove with simple washing. Several approaches are used to address this issue that leads to significant economic losses in the petroleum industry:

These deposits must be treated either to eliminate existing deposits (curative treatment) or to prevent them in the future (preventive treatment). (14)

I.16.1 Preventive treatment

This treatment involves injecting a solution with different dispersants into wells that are known to have deposit problems over and over again. This is achieved by circulating the solution downward through a concentric tube and returning it to the surface through the annular space. This will prevent or minimize the agglomeration and formation of asphaltene deposits.

The remedy for asphaltene deposits consists of preventing their flocculation in the crude oil. This can be achieved by considering the following three solutions:

- **Modification of thermodynamic parameters:**

Maintaining a high pressure in the reservoir to stay out of the dangerous area is one potential preventive strategy. However, the use of this method has had a significant influence on recovery rate, productivity, and manufacturing cost. Furthermore, this method does not prevent the problem of deposits in wells and surface facilities. Bibliographies show that the only thermodynamic change that can affect the precipitation of asphaltene is pressure. Therefore, it would be necessary to produce the brut at a pressure higher than the onset point, which is the domain of the impossible.

- **Modification of petroleum composition :**

This result can only be obtained by increasing aromaticity. Laboratory experiments have shown that the amount of aromatics required to prevent asphaltene flocculation is significant, which implies costly expenditures.

- **Modification of the ionic environment surrounding the asphaltene:**

This modification involves the use of specific inhibitors (surfactant- or polymer-based). It appears to be an approach with good chances of being both effective and economical.

I.16.1 Curative treatment

- **Mechanical scraping:** This method involves routinely scraping the internal walls of production tubes using wireline operations with a scraper. However, this operation is sometimes difficult and has drawbacks such as repeated interventions and cable breakage.
- **Tube coating:** Another considered solution, which has demonstrated its effectiveness in the laboratory, is the coating of tubing with plastic material on which deposits have low adhesion. The chosen plastic material (generally polymers) must meet two conditions:
Resistance to the temperature of the medium Affinity for asphaltenes
- **Removal with hot fluids:** This means injecting hot oil, water or steam into the formation to open plugged zones (reverse circulation). This method works by melting organic deposits.
- **Cleaning with solvents:** This is the most popular method for hydrocarbon removal. Among the excellent solvents are the following: reformat, carbon disulfide, benzene, toluene, xylene and chlorinated solvents. Circulation of an aromatic solvent solution in the tubing with the aim of dissolving deposits formed on the tubing walls. Squeezing of various solvent solutions such as reformat, with the purpose of dissolving organic deposits in the formation. Carbon disulphide .

While it is one of the best solvents, it is also one of the most dangerous to use. It is toxic and explosive, with a flash point of -22°F (-30°C) and an auto-ignition temperature of 212°F (100°C). Chlorinated solvents:

They are excellent solvents, but they damage the catalyst used in the refining process and represent health hazards.

Benzene:

It's an excellent solvent, but extremely flammable and carcinogenic.

Xylene and le toluene:

Also, they are exceptional solvents. Their saturation point is rapidly reached, and they are unable to dissolve any further.

Reformat:

It is the most widely used solvent; it is economical for the treatment of asphaltene deposits in the Hassi Messaoud field and has proven its effectiveness on several wells with good results after treatment. However, its use is very dangerous. (15)

Chapter I: General Characteristics and Properties of Asphaltenes.

- Features of the Reformat:

Density=0.778 Flash point = -5 °c.

- What the Reformat Is Made Of:

The Reformat is a petroleum distillate produced locally in Algeria. It is obtained from the first distillation charge, consisting essentially of paraffins (42%), naphthenes (5%) and aromatics (53%).

Conclusion

One of the most complicated parts of crude oil is asphaltenes. They are the heaviest composition of petroleum and are distinguished by their high molecular weight, polycyclic aromatic structure, and presence of trace metals like nickel and vanadium as well as heteroatoms like sulfur, nitrogen, and oxygen. They are intrinsically unstable under shifting reservoir conditions due to their peculiar solubility characteristic, which renders them soluble in aromatic solvents but insoluble in light alkanes.

Variations in temperature and pressure, variations in the content of crude oil, injection of fluids like CO₂, and the presence of metal ions like ferric chloride are some of the conditions that cause asphaltenes to deposit and flocculate. Asphaltene concentration alone does not define the severity of deposition, as shown by the Hassi Messaoud field, which shows that even crude oils with extremely low asphaltene content (as little as 0.1%) may produce serious clogging issues.

Asphaltene deposition has serious repercussions, including lower oil recovery, blocked production tubing, changed rock wettability, and decreased reservoir permeability. The petroleum sector suffers significant financial losses as a result of these problems.

Strategies for both preventative and curative therapy have been developed to address this issue. Using chemical inhibitors, modifying the composition of the oil, or preserving ideal

thermodynamic conditions are the basic preventive strategies. Mechanical scraping, solvent cleaning (especially with reformat in the Algerian setting), and hot fluid injection are examples of curative techniques. Every technique has its own set of safety and technological limitations.

Chapter II:
Artificial Intelligence and Machine Learning Fundamentals

Introduction

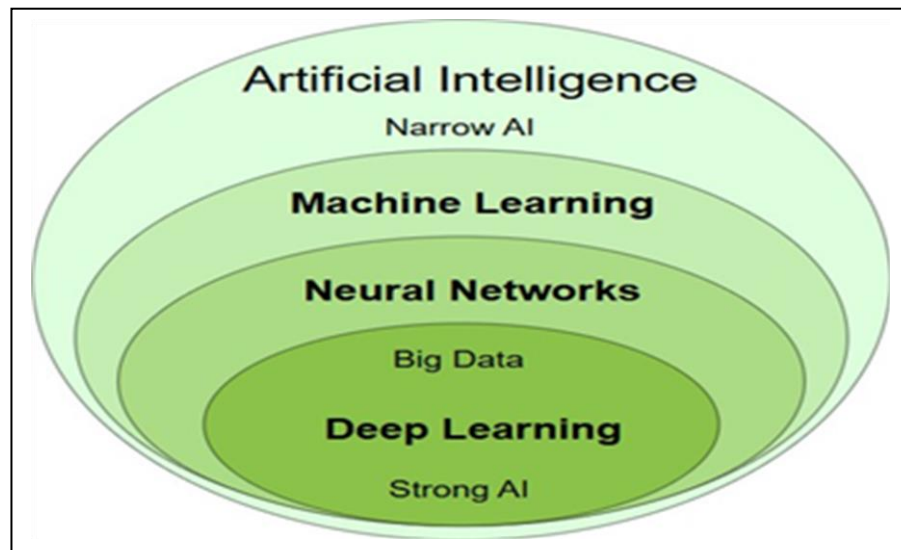
The fast improvement of computational technologies over the past few decades has led to the emergence of artificial intelligence and machine learning as powerful tools capable of transforming complex problem-solving across numerous scientific and industrial domains. In the petroleum industry, where the behavior of reservoir fluids is governed by highly nonlinear and multivariable interactions, these technologies offer a promising alternative to conventional analytical and thermodynamic approaches.

II.1 Artificial intelligence

Artificial intelligence (AI) is a domain of computer science focused on creating systems and software that can execute activities often linked to human cognition. These systems are designed to learn from experience, adapt to new inputs, and make autonomous choices with little human involvement via the integration of data, algorithms, and computer power.

AI exhibits capabilities in natural language processing, pattern recognition, complicated problem-solving, and creative behavior, often exceeding human performance in speed and scale. These capabilities are profoundly altering human-technology interaction and perpetually broadening the operating limits of robots.

The area is divided into many specialized subfields, each focusing on certain facets of AI study and application, such as Machine Learning, Deep Learning, and Natural Language Processing. (17)



FigureII. 1 Subsets of Artificial Intelligence.

II.2 Machine Learning

Machine Learning (ML) is a distinct domain of artificial intelligence that allows computer systems to learn from data without the need for explicit programming directives. This is accomplished by optimizing model parameters using mathematical calculations, so shaping and refining the model's behavior based on the underlying facts or collected experience. The learning method incrementally adjusts parameter values throughout the training phase, enabling the ML model to provide predictions or judgments based on data-driven insights.

(18)

Machine learning fundamentally enables computers to enhance their performance on certain tasks by using experience and statistical learning instead of relying on predetermined rules. Its ability to automate intricate analytical procedures and derive important patterns from extensive data establishes it as a disruptive technology with considerable ramifications across many sectors and application areas. **(19)**

II.3 Types of Machine Learning

Learning can be broadly categorized into four distinct classes based on the nature of the training data and the mode of interaction between the learning system and its environment.

(20)

II.3.1 Supervised Learning

Supervised machine learning is based on labeled training datasets. As data is fed to the ML model, it goes through a cross-validation process which adjusts its weight until it is fitted appropriately. A majority of practical machine learning uses supervised learning. In supervised learning, the system tries to learn from the previous examples that are given. Speaking mathematically, supervised learning is where you have both input variables (x) and output variables (Y) and can use an algorithm to derive the mapping function from the input to the output. The mapping function is expressed as $Y = f(X)$ **(21)**

- **Categories**

Regression: In statistics, a Linear Regression is an approach to modeling a linear relationship between y and x. **(22)**

Classification: A classification problem is when the output variable is a category or a group, such as “black” or “white” or “spam” and “no spam“ **.(23)**

II.3.2 Unsupervised Learning

It is called unsupervised learning because unlike supervised learning above, there are no given correct answers and the machine itself finds the answers. Unsupervised is based on unlabeled datasets. The algorithms are left to themselves to discover interesting structures in the data. Mathematically, unsupervised learning is when you only have input data (X) and no corresponding output variables the objective of unsupervised learning is to teach ML models to detect hidden patterns or structures without human supervision. (24)

- **Categories**

Association: An association rule learning problem is where you want to discover rules that describe large portions of your data, such as “people that buy X also tend to buy Y”.

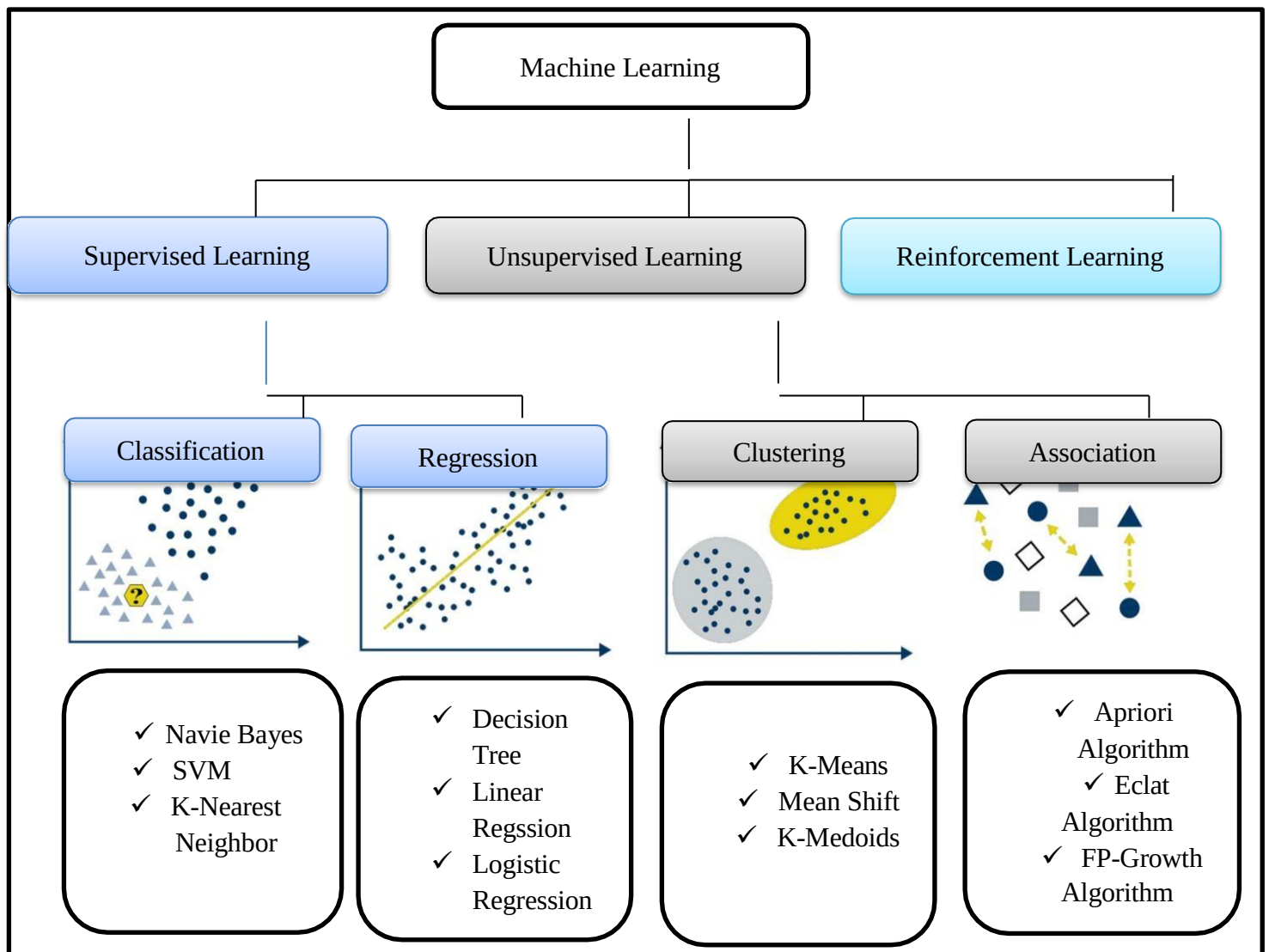
Clustering: Clusters are collections of data based on similarity. (25)

II.3.3 Semi supervised

If some learning samples are labeled, but some others are not labeled, then it is semi-supervised learning. It makes use of a large amount of unlabeled data for training and a small amount of labeled data for testing. Semi-supervised learning is applied in cases where it is expensive to acquire a fully labeled dataset while more practical to label a small subset

II.3.4 Reinforcement

While similar to supervised learning, reinforcement learning relies on trial and error. Without labeled training datasets, reinforcement learning trains ML models to develop best recommendations based on a series of successful outcomes. (21)

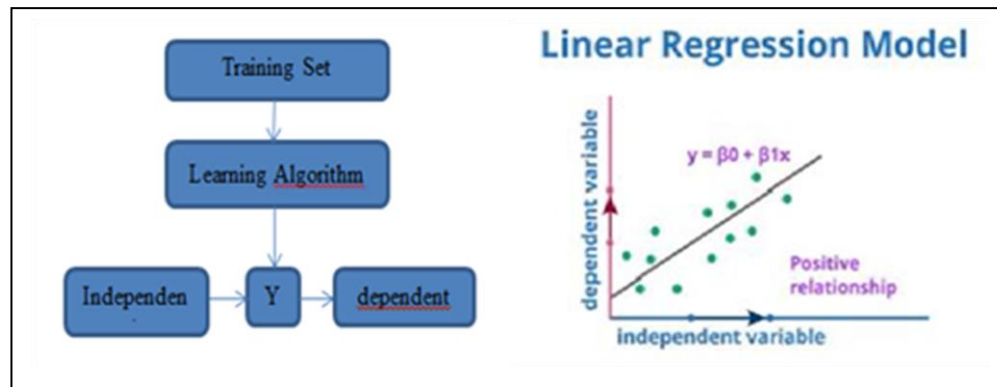


FigureII. 2 Classification of ML.

II.4 Machine Learning Algorithms

II.4.1 Linear Regression

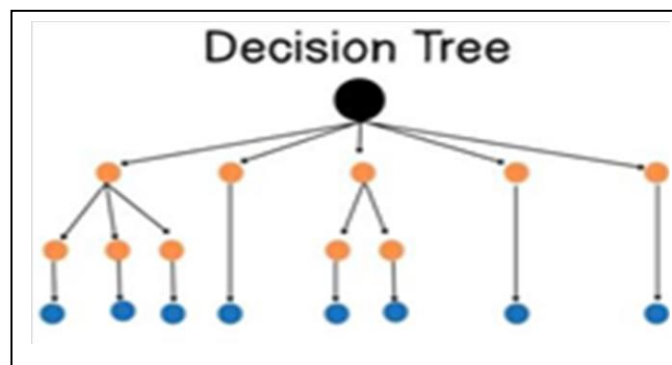
Linear regression is a supervised learning algorithm widely employed in predictive modeling applications. It operates by establishing a linear mathematical relationship between a set of input features and a continuous output variable, whereby the model iteratively adjusts its weight parameters to minimize the prediction error over the training dataset. (26)



FigureII. 3 illustration of linear regression (27)

II.4.2 Decision Tree

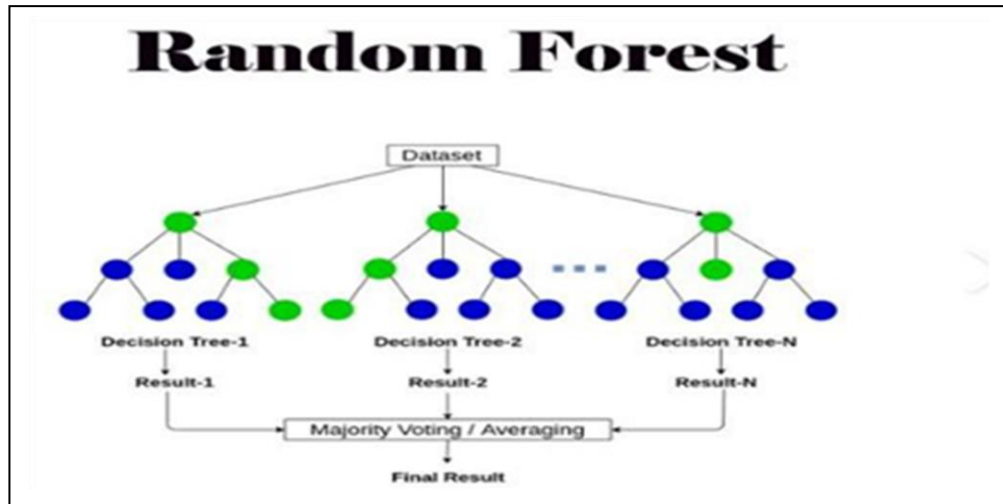
A Decision Tree is a supervised learning algorithm that operates through a hierarchical tree-structured model of sequential decisions. It recursively partitions the training data into increasingly homogeneous subsets by selecting the most statistically significant attribute at each node, resulting in an interpretable and transparent model applicable to both classification and regression tasks. (29)



FigureII. 4 Decision Tree structure.

II.4.3 Random Forest

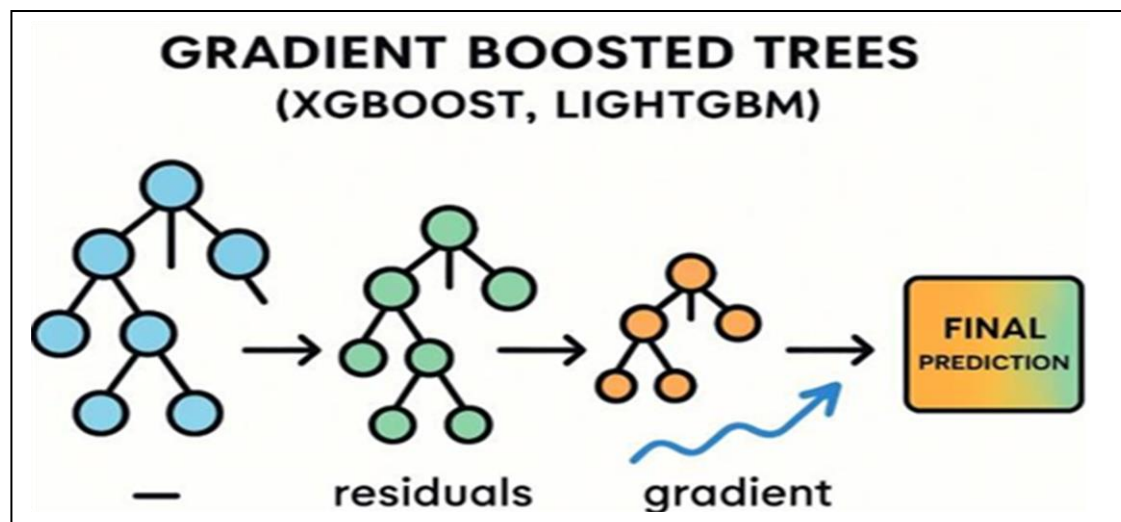
Random Forest is an ensemble learning method that constructs a multitude of decision trees during training and aggregates their individual outputs to produce a final prediction. By combining the predictions of multiple diverse trees, this approach effectively handles high-dimensional and nonlinear data while substantially reducing the risk of overfitting that is commonly associated with single decision tree models. (30)



FigureII. 5 Random Forest structure (28).

II.4.4 Gradient Boosting

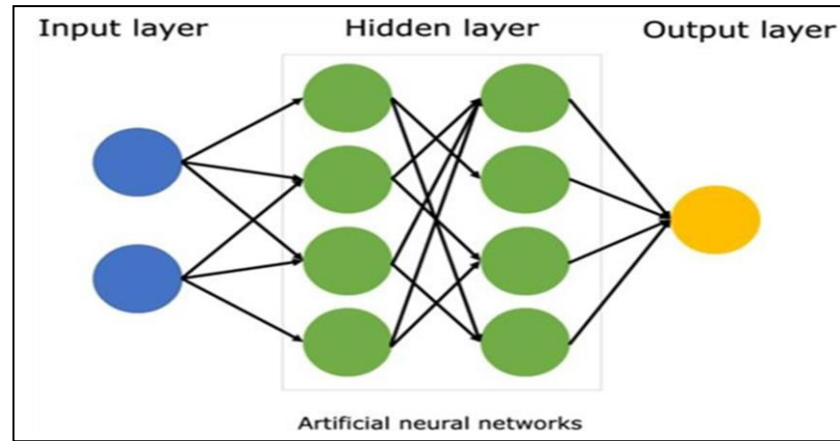
Gradient Boosting is an ensemble learning technique that sequentially constructs a series of weak learners in a stage-wise manner, whereby each successive model is trained to correct the residual errors of its predecessor. When appropriately tuned, this approach is widely recognized for its ability to achieve high predictive accuracy, making it one of the most powerful algorithms in supervised learning. (31).



FigureII. 6 Gradient boosting.

II.4.5 Artificial Neural Networks ANN

Artificial Neural Networks (ANNs) are computational models inspired by the biological neural structure of the human brain. They consist of interconnected neurons organized into successive layers, through which complex nonlinear mappings between input variables and output predictions are learned via the back propagation algorithm and gradient descent optimization (32).



FigureII. 7 Basic structure of ANN.

II.4.6 K-Nearest Neighbors KNN

K-Nearest Neighbors (KNN) is a non-parametric supervised learning algorithm that generates predictions by identifying the k closest data points to a given query instance within the feature space. Similarity between observations is quantified through distance metrics, most commonly the Euclidean distance, and the algorithm is applicable to both classification and regression tasks owing to its conceptual simplicity and ease of implementation. (31)

Where:

- $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ be the training dataset.
- $X \in \mathbb{R}^d$ be a new input vector to classify or predict.

A common distance metric is the **Euclidean distance**:

$$(X, X_i) = \sqrt{\sum_{j=1}^d (x_j - x_{ij})^2} \quad (1)$$

Other commonly employed distance metrics include:

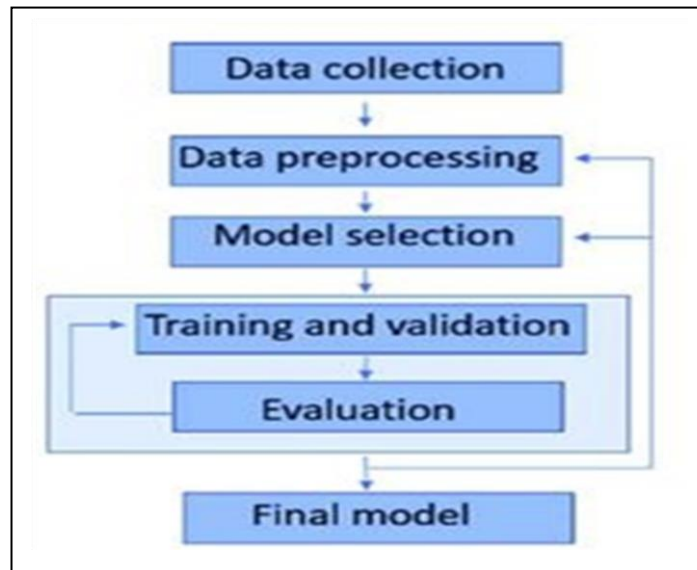
- **Manhattan distance**
- **Minkowski distance**
- **Cosine similarity**

KNN is applicable to both classification and regression tasks, offering simplicity and interpretability, though its computational cost increases significantly with the size of the training dataset.

II.4.7 Support Vector Machine SVM

Support Vector Machine (SVM) is a supervised learning algorithm that identifies the optimal hyperplane maximizing the margin of separation between classes within a given dataset. Through the application of kernel functions, SVM effectively handles nonlinearly separable data and demonstrates robust performance in high-dimensional feature spaces (33).

II.5 The general steps to train ML models



FigureII. 8 General Workflow for Training Machine Learning Models.

II.5.1 Data collection

Gathering relevant and representative data is critical. High-quality data improves model accuracy, generalization, and reduces biases. (34)

II.5.2 Data preprocessing

Prior to any analysis, the dataset was subjected to a thorough cleaning process designed to ensure consistency, accuracy, and reliability. This involved addressing missing values, eliminating duplicate records, and resolving irregularities or anomalies identified within the data. The primary goal was to produce a high-quality, analysis-ready dataset that minimizes the risk of errors throughout the machine learning pipeline and supports sound modeling and interpretation. (34)

II.5.3 Data Transformation

Transforming data into structures suitable for learning . Includes scaling, normalization, feature engineering ...etc.(34)

Feature importance: refers to techniques that assign a score to input features based on how useful they are in predicting a target variable. It helps identify key drivers of model predictions and improves interpretability. Common methods include permutation importance, tree-based scores, and model-agnostic approaches like correlation analysis.

II.5.4 Splitting Data

To build a machine learning model capable of generalizing well to unseen data while avoiding common issues such as underfitting and overfitting, the available dataset must be partitioned into three distinct subsets: training, validation, and test sets.

Underfitting occurs when a model is overly simplistic or insufficiently trained, yielding poor performance on both seen and unseen data. Overfitting, conversely, happens when a model learns the training data too closely, capturing its noise and outliers, which results in high training accuracy but weak generalization to new inputs. Dataset partitioning is a widely adopted strategy to address both of these challenges.

- **Training set:** This subset is used to fit the model by allowing the algorithm to learn the underlying patterns and optimize its parameters through iterative training.
- **Validation set:** This subset serves as an evaluation benchmark during the training process. Monitoring prediction errors on the validation data helps determine the appropriate number of training iterations. Under normal conditions, errors should decline progressively for both the training and validation sets. However, when the training error continues to fall while the validation error begins to rise, this divergence signals the onset of overfitting and indicates that training should be halted at that point.
- **Test set:** Reserved exclusively for the final stage, this subset is used to assess the performance of the fully trained model. Its prediction accuracy serves as an unbiased estimate of how the model is expected to perform on real-world, previously unseen data.

Common splitting strategies range from a straightforward 80/20 training-to-test division to a more comprehensive 70/15/15 split across training, validation, and test subsets. The latter approach ensures both effective model tuning and a fair, thorough final evaluation. Ultimately, this structured partitioning promotes balanced performance across all data subsets and supports robust generalization in practical applications.

II.5.5 Model Selection

The choice of an appropriate machine learning algorithm is largely governed by the inherent characteristics of the dataset, including its size, structure, and overall quality, as these attributes directly influence the suitability and effectiveness of a given method. Equally important is the nature of the problem at hand, whether it involves classification, regression, or clustering, since different algorithmic families are specifically designed to address distinct problem types.

For large-scale datasets, priority is typically given to computationally efficient and scalable algorithms such as Random Forests or Gradient Boosting Machines. These methods are well-suited to capturing complex, non-linear patterns within the data while remaining manageable in terms of computational demand. Their ensemble-based architecture also contributes to greater predictive stability and robustness across varied data conditions. (34)

On the other hand, when working with smaller datasets, simpler and more interpretable models such as Logistic Regression or Support Vector Machines are generally more appropriate. These approaches carry a lower risk of overfitting, require comparatively fewer computational resources, and tend to deliver reliable and consistent performance even when training samples are limited.

By aligning the algorithm selection with both the dataset's properties and the problem's requirements, practitioners can ensure a more principled and effective modeling process.

II.5.6 Model Training and Validation

The model learns patterns from training data. Validation is used for tuning hyperparameters and preventing overfitting.

II.5.7 Model Evaluation

The model is evaluated on a test set using performance metrics such as accuracy, RMSE, MAE, or R^2 . This step provides insight into real-world applicability. (35)

Three standard statistical metrics are employed to evaluate the predictive performance of the developed model.

Table II. 1 Evaluation Metrics for Regression Models.

Metric	Formula	Explanation	Reference
R ² Score(Coefficient of Determination)	$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y}_i)^2}$ (2)	Measures how well predicted values approximate the actual values. Ranges from $-\infty$ to 1. A value of 1 means perfect prediction, 0 means no better than the mean, and negative values indicate worse performance than the mean.	(35)
Root Mean Squared Error (RMSE)	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_j - \hat{y}_i)^2}$ (3)	Represents the square root of 2 the average squared differences between actual and predicted values. It penalizes larger errors and is sensitive to outliers. Lower values indicate better performance.	(35)
Mean Absolute Error (MAE)	$MAE = \frac{1}{n} \sum_i^n y_i - \hat{y}_i $ (4)	Measures the average absolute difference between actual and predicted values. It treats all errors equally and is more robust to outliers than RMSE. Lower values indicate better performance.	(35)

II.6 Applications

Machine Learning (ML), a subfield of Artificial Intelligence, focuses on enabling machines to learn from data and improve performance over time without being explicitly programmed. ML models identify patterns, make predictions, and provide insights, making them highly valuable across a wide range of industries. (18)

In general applications, ML is widely used in:

- Healthcare: Disease prediction, medical imaging analysis, and personalized treatment recommendations.
- Finance: Credit scoring, algorithmic trading, risk assessment, and fraud detection.
- Marketing: Customer segmentation, recommendation engines, and sentiment analysis.

- Manufacturing: Predictive maintenance, process optimization, and quality control.
- Transportation: Route optimization, traffic forecasting, and autonomous navigation

II.7 Applications in oil and gas

In the context of the petroleum industry, Machine Learning plays a critical role in enhancing operational efficiency and reducing costs. Common applications include (36):

- Reservoir Characterization: ML models analyze well log and seismic data to predict Lithology, porosity, and permeability.
- Production Forecasting: Time series models are used to predict oil and gas production based on historical performance.
- Drilling Optimization: ML techniques help optimize drilling parameters, reduce non-productive time, and prevent equipment failure.
- Well Log Interpretation: Supervised ML algorithms can estimate missing or degraded petrophysical parameters such as Gamma Ray (GR), porosity, and saturation.
- Enhanced Oil Recovery (EOR): ML assists in analyzing reservoir performance and selecting optimal EOR strategies

The art of predicting the unpredictable

The prediction of asphaltene precipitation has evolved significantly over the past decades, transitioning from classical thermodynamic models toward more advanced data-driven approaches. Early thermodynamic frameworks, including Flory-Huggins theory and EOS-based models, provided foundational understanding of asphaltene phase behavior but demonstrated notable limitations, particularly at pressures below the saturation point and under complex reservoir conditions.

To overcome these shortcomings, researchers increasingly turned to machine learning and intelligent modeling techniques.

Kishvand and Naseri (2012) were among the first to apply Artificial Neural Networks (ANNs) for asphaltene precipitation prediction under real reservoir conditions, demonstrating that a three-hidden-layer ANN with back-propagation yielded the most accurate results (37).

Hemmati-Sarapardeh et al. (2013) subsequently developed a Least Square Support Vector Machine (LSSVM) model based on SARA fractions, pressure, API gravity, and temperature, achieving an AAPRE of 18.69% (38).

This was significantly improved in 2016 by the same group through CSA-optimized LSSVM, reaching an AAPRE of 9.46% on 326 experimental data points (39).

Later, in 2019, Radial Basis Function (RBF) and Multilayer Perceptron (MLP) neural networks were applied on Iranian crude oil samples, yielding AAPRE values of 18.75% and 26.77% respectively (40).

More recent studies have further advanced predictive capabilities. Sulaimon et al. (2023) developed correlations based on the Colloidal Instability Index (CII) and SARA fractions, demonstrating that CII-based correlations outperform SARA-based ones, with an R^2 of 0.6875 (41).

Most notably, Khalighi and Cheremisin (2023) conducted a comprehensive comparison of ML algorithms — XGBoost, MLP, LSSVM-PSO, and MLR — using the most extensive dataset available, conclusively demonstrating that data-driven methods outperform classical thermodynamic models, with the exception of MLR (42).

Most recently, Amiri-Ramsheh et al. (2025) developed four advanced intelligent models — XGBoost, LightGBM, Cascade Forward Neural Network (CFNN), and Generalized Regression Neural Network (GRNN) — for predicting asphaltene precipitation during natural reservoir depletion, using a comprehensive databank of 380 experimental data points with SARA fractions, pressure, bubble point pressure, temperature, and API gravity as input variables. Among the constructed models, LightGBM achieved the highest predictive accuracy with an AAPRE of 8.57% and an R^2 of 0.9904, while sensitivity analysis identified pressure and asphaltene content as the most influential parameters on precipitation behavior (43).

Overall, the literature confirms a clear superiority of machine learning approaches over traditional thermodynamic methods for asphaltene precipitation prediction, particularly when applied to large and diverse datasets incorporating SARA fractions, pressure, temperature, API gravity, and fluid composition. These findings directly motivate the ML-based predictive framework developed in the present work.

Based on the existing literature, it is clear that asphaltene precipitation prediction still faces significant challenges. Classical thermodynamic models remain constrained by their underlying assumptions, while many data-driven approaches suffer from limited dataset diversity or are tailored to specific crude oil systems, reducing their applicability to broader reservoir conditions.

In response to these limitations not conducted any sensitivity analysis to determine the influential parameters on the output, the present work use a machine learning-based predictive framework for asphaltene precipitation, developed and implemented using Python. The proposed approach leverages a comprehensive dataset incorporating key physicochemical

parameters — including SARA fractions, pressure, temperature, and API gravity — to be a reliable and accurate predictive model for asphaltene precipitation under a wide range of reservoir and production conditions.

Conclusion

The prediction of asphaltene precipitation represents one of the most challenging problems in petroleum engineering, owing to the highly nonlinear and multivariable nature of reservoir fluid behavior. Classical thermodynamic approaches, such as Flory-Huggins theory and equation-of-state-based models, laid the groundwork for understanding asphaltene phase behavior but proved insufficient under complex and varied reservoir conditions.

The progressive adoption of machine learning techniques has marked a turning point in addressing these limitations. From early applications of Artificial Neural Networks to more sophisticated algorithms such as XGBoost, LightGBM, and LSSVM, the literature consistently demonstrates the superiority of data-driven methods over traditional thermodynamic models, particularly when applied to large and diverse datasets.

Key input parameters — including SARA fractions, pressure, temperature, bubble point pressure, and API gravity and CII— have proven critical in capturing the underlying physics of asphaltene precipitation. Sensitivity analyses further confirm that pressure and asphaltene content are the most influential variables governing precipitation behavior.

Building upon these findings, the present work proposes a comprehensive ML-based predictive framework implemented in Python, designed to overcome the limitations of prior studies by incorporating a diverse and representative dataset. This framework aims to deliver reliable, accurate, and generalizable predictions of asphaltene precipitation across a wide range of reservoir and production conditions, contributing meaningfully to flow assurance management and operational efficiency in the petroleum industry

General conclusion

General conclusion

This study investigated the prediction of asphaltene scale formation in Hassi Messaoud reservoir conditions using machine learning techniques as an alternative to conventional thermodynamic models.

The first chapter provided a comprehensive review of asphaltene behavior, covering their chemical composition, molecular structure, deposition mechanisms, and the influence of pressure and temperature on flocculation. This theoretical background established the complexity of asphaltene precipitation and highlighted the limitations of classical predictive approaches.

The second chapter introduced the fundamental concepts of artificial intelligence and machine learning, presenting the key algorithms considered in this study — including Decision Trees, Random Forest, Gradient Boosting, XGBoost, ANN, SVM, and KNN — along with their general training methodology and applications in the oil and gas industry.

The third chapter presented the practical implementation of these models using a dataset built from SARA fractionation analysis and PVT properties (pressure, temperature, bubble point pressure, API gravity) collected from multiple wells. After preprocessing and feature scaling, the models were trained and evaluated using Leave-One-Out Cross-Validation (LOOCV) to ensure robustness given the limited dataset size.

The results demonstrated that **Gradient Boosting** achieved the best predictive performance with $R^2 = 0.969$, significantly outperforming Linear Regression ($R^2 = 0.920$), Decision Tree ($R^2 = 0.843$), XGBoost ($R^2 = 0.835$), Random Forest ($R^2 = 0.816$), and SVR ($R^2 = 0.572$). These findings confirm that ensemble-based machine learning models, particularly gradient boosting methods, are highly effective for predicting asphaltene precipitation and represent a reliable, data-driven complement to traditional thermodynamic approaches.

The Asphaltene/Resin (A/R) ratio is a commonly used indicator for evaluating asphaltene stability in crude oil. It is determined by dividing the asphaltene content by the resin content, both expressed as weight percentages from SARA analysis. A low A/R ratio reflects a stable system where asphaltenes are well-peptized by resins and unlikely to

General conclusion

precipitate, while a high A/R ratio signals increased instability and a greater risk of deposition. Due to its simplicity and practical relevance,

this approach is widely applied in both academic research and field operations to characterize asphaltene behavior and anticipate precipitation tendencies.

As a perspective, future work and recommendations are proposed for future research and field applications:

- **Dataset Expansion** The current study was conducted on a limited number of wells from the Hassi Messaoud field. It is strongly recommended to expand the dataset by incorporating additional well samples from other Algerian fields to improve model generalizability and robustness.
- **Integration of Additional Input Features** Future models should consider incorporating additional physicochemical parameters such as the Colloidal Instability Index (CII), the Resin-to-Asphaltene (R/A) ratio, oil viscosity, water cut, and gas-oil ratio (GOR) as explicit input features, as these parameters are known to significantly influence asphaltene precipitation behavior.
- **Treatment Optimization** The predictive model could be further extended to optimize the dosage and injection timing of chemical inhibitors and dispersants in this field to prevent from precipitation, contributing directly to more cost-effective flow assurance management in Algerian petroleum production operations.
- **Experimental Validation** Laboratory experiments using high-pressure, high-temperature (HPHT) cells and near-infrared (NIR) spectroscopy are recommended to validate the machine learning predictions against measured asphaltene onset pressure (AOP) values under actual reservoir conditions.

References

References

- (1) Sheu, E. Y., & Mullins, O. C. (Eds.). (1995). *Asphaltenes: Fundamentals and applications*. Plenum Press.
- (2) Chibane, R. (2013). *Etude de la déposition des asphaltènes dans la périphérie du champ de Haoud Berkaoui*. IAP.
- (3) Djelti, F. (2012). *Contribution à l'étude des propriétés des asphaltènes [Mémoire de Magister]*. Université Aboubakr Belkayed de Tlemcen.
- (4) Baba Aissa, A., Laribi, S., & Ziada, A. (1996). Synthèse sur les études concernant le problème des dépôts d'asphaltènes à Hassi-Messaoud (CRD). JST, Communication.
- (5) Sheu, E. Y., & Mullins, O. C. (Eds.). (1995). *Asphaltenes: Fundamentals and applications*. Plenum Press.
- (6) Hadj Abbas, A., Manasreh, A. D., Nassar, N. N., & Hacini, M. (n.d.). *Etude de caractérisation des asphaltènes algériens extraits du gisement pétrolier de Hassi-Messaoud*. Department of Chemical and Petroleum Engineering, University of Calgary.
- (7) Boudinane, R. (n.d.). *Reformat product* (p. 16) [Master's thesis].
- (8) Initec Engineering & Texaco Inc. (2005). *Deepstar asphaltene discussion*. USA.
- (9) SONATRACH. (n.d.). *Data bank*.
- (10) Faraoun, A. (2012). *Inhibition de la précipitation des asphaltènes du pétrole algérien — Mise en évidence par spectrophotométrie UV-visible*. Oran.
- (11) Flory, P. J. (1942). Thermodynamics of high polymer solutions. *The Journal of Chemical Physics*, 10(1), 51–61. <https://doi.org/10.1063/1.1723621>
- (12) Mullins, O. C. (2010). The modified Yen model. *Energy & Fuels*, 24(4), 2179–2207. <https://doi.org/10.1021/ef900975e>
- (13) Buckley, J. S., Takamura, K., & Morrow, N. R. (1996). Asphaltenes and the onset of asphaltene deposition. *SPE Production & Facilities*, 11(1), 30–36. <https://doi.org/10.2118/26014-pa>
- (14) Boukherissa, M. (2008). *Etude de la stabilité des asphaltènes dans le pétrole brut — Choix de dispersants et le mécanisme d'action [Thèse de doctorat]*. Université Paul Verlaine de Metz.
- (15) Debihi, A. (2018, April). *Modélisation et caractérisation de comportementphasique et volumétrique de fluide — Cas d'étude : le champ Bhiret Aissa [Mémoire]*.
- (16) Organization of School Administrators & Supervisors. (2023, July 26). *Artificial intelligence and education*. <https://osasdetroit.org/news/artificial-intelligence-and-education>

References

- (17) Rodriguez, O. (2025, February 11). Topic: Artificial Intelligence (AI) — What is AI? LinkedIn. https://www.linkedin.com/posts/omarrodriguezm_topic-artificial-intelligence-ai-activity-7295100641612308481-ERUA
- (18) Shinde, P. P., & Shah, S. (2018, August). A review of machine learning and deep learning applications. In 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA) (pp. 1–6). IEEE.
- (19) Zhou, Z. H. (2021). Machine learning. Springer Nature.
- (20) Ayodele, T. O. (2010). Types of machine learning algorithms. In New advances in machine learning (Vol. 3, pp. 19–48).
- (21) Morales, E. F., & Escalante, H. J. (2022). A brief introduction to supervised, unsupervised, and reinforcement learning. In Biosignal processing and classification using computational learning and intelligence (pp. 111–129). Academic Press.
- (22) Maulud, D., & Abdulazeez, A. M. (2020). A review on linear regression comprehensive in machine learning. Journal of Applied Science and Technology Trends, 1(2), 140–147.
- (23) Soofi, A. A., & Awan, A. (2017). Classification techniques in machine learning: Applications and issues. Journal of Basic & Applied Sciences, 13, 459–465.
- (24) Naeem, S., Ali, A., Anam, S., & Ahmed, M. M. (2023). An unsupervised machine learning algorithms: Comprehensive review. International Journal of Computing and Digital Systems.
- (25) Ezugwu, A. E., Ikotun, A. M., Oyelade, O. O., Abualigah, L., Agushaka, J. O., Eke, C. I., & Akinyelu, A. A. (2022). A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects. Engineering Applications of Artificial Intelligence, 110, 104743.
- (26) Maulud, D., & Abdulazeez, A. M. (2020). A review on linear regression comprehensive in machine learning. Journal of Applied Science and Technology Trends, 1(2), 140–147.
- (27) Abhay, S. (2024, November 18). Linear regression: The cornerstone of predictive analytics. Medium.
- (28) Jainvidip. (2024, July 1). Understanding decision trees. Medium.
- (29) Abid, S. (2024, November 5). Decision tree: A decision tree is a supervised learning algorithm [LinkedIn post]. LinkedIn. https://www.linkedin.com/posts/shaik-abid-58291b304_decision-tree-a-decision-tree-is-a-supervised-activity-7259573114311696384-q08m
- (30) Shamitha. (2026, March 29). Top 10 machine learning algorithms you should know. Jeevi Academy. <https://www.jeeviacademy.com/top-10-machine-learning-algorithms-you-should-know/>

References

- (31) Pandey, D., Niwaria, K., & Chourasia, B. (2019). Machine learning algorithms: A review. *Machine Learning*, 6(2), 916–922.
- (32) Touaher, R., & Ferdjani, M. (2025). Predicting well logs using machine learning techniques [Master's thesis]. University of Ouargla.
- (33) Mohammed, M., Khan, M. B., & Bashier, E. B. M. (2016). *Machine learning: Algorithms and applications*. CRC Press.
- (34) Vieira, S., Garcia-Dias, R., & Pinaya, W. H. L. (2020). A step-by-step tutorial on how to build a machine learning model. In *Machine learning* (pp. 343–370). Academic Press.
- (35) Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7, e623.
- (36) Sircar, A., Yadav, K., Rayavarapu, K., Bist, N., & Oza, H. (2021). Application of machine learning and artificial intelligence in oil and gas industry. *Petroleum Research*, 6(4), 379–391.
- (37) Khishvand, M., & Naseri, A. (2012). An artificial neural network approach to predict asphaltene deposition test result. *Fluid Phase Equilibria*, 329, 32–41.
- (38) Hemmati-Sarapardeh, A., et al. (2013). Asphaltene precipitation due to natural depletion of reservoir: Determination using a SARA fraction based intelligent model. *Fluid Phase Equilibria*, 354, 177–184.
- (39) Hemmati-Sarapardeh, A., Ameli, F., Dabir, B., Ahmadi, M., & Mohammadi, A. H. (2016). On the evaluation of asphaltene precipitation titration data: Modeling and data assessment. *Fluid Phase Equilibria*, 415, 88–100.
- (40) Hemmati-Sarapardeh, A., et al. (2019). Modeling asphaltene precipitation during natural depletion of reservoirs and evaluating screening criteria for stability of crude oils. *Journal of Petroleum Science and Engineering*, 181, 106127.
- (41) Sulaimon, A. A., et al. (2023). Developing new correlations for asphaltene deposition involving SARA fractions and colloidal instability index. *Journal of Petroleum Science and Engineering*, 220, 111143.
- (42) Khalighi, J., & Cheremisin, A. (2023). Comparative study of machine learning algorithms in predicting asphaltene precipitation with a novel validation technique. *Earth Science Informatics*, 16(4), 3097–3111.
- (43) Amiri-Ramsheh, B., Sahebalzamani, S., Zabihi, R., & Hemmati-Sarapardeh, A. (2025). Predicting asphaltene precipitation during natural depletion of oil reservoirs by integrating SARA fractions with advanced intelligent models. *Scientific Reports*, 15, 26214.
- (44) M. C. Henni, Formation Damage Prevention Through Organic Deposit Control in Hassi Messaoud Oil Field, M.S. thesis, Dept. Oil and Gas Production Eng., 2011.

References

- (45) Nouacer, A. Collaku, and I. Bendjedia, “Challenges for digital oil field applications in Hassi-Messaoud Field (HMD), Sonatrach, Algeria,” presented at the SLB SIS Forum, Schlumberger, Sept. 18, 2019

