

République Algérienne Démocratique Et Populaire
Ministère de l'enseignement Supérieur et de la Recherche Scientifique



Université M'Hamed Bougara – Boumerdes
Faculté des Hydrocarbures et de la Chimie
Département Gisements Miniers et Pétroliers



Mémoire de Fin d'Etude

En vue de l'obtention du Diplôme

MASTER

Présenté par

M. Houdaifa Khalifa

Filière : Hydrocarbures

Spécialité : Génie pétrolier – Forage des Puits

**Conception de la solution digitale Terraverse pour la
prédiction de la lithologie en temps réel et l'assurance du
cheminement de puits contre les interférences magnétiques**

Devant le jury composé :

Mr. BELIMANE Zakaria	MAB	UMB-Boumerdes	Président
Mr. BOUBEKEUR Zinelabidine	MAA	UMB-Boumerdes	Encadreur
Mme. AZRIL Nadjet	MAA	UMB-Boumerdes	Examinateur

Année universitaire : 2022/2023

Dédicace

“À mon **père** et ma **mère**, mes **piliers** solides,
À mes **frères** et **sœurs**, **compagnons** de vie,
À mes **amis** que **j’aime** de tout mon être,
Je dédie ce travail empreint de passion et d’envie.

Vos sourires, vos encouragements précieux,
Vos mots d’amour et votre soutien infini,
Ont illuminé mon chemin, rendu mes rêves joyeux,
Cette dédicace est un témoignage de ma gratitude, sans aucun déni.

Que ces mots sincères portent le poids de mes sentiments,
Que leur écho résonne en votre cœur éternellement,
Car c’est grâce à vous que je suis devenu qui je suis,
Un hommage à votre amour, un témoignage de ma vie. ”

- Houdaifa Khalifa -

Remerciements

Je tiens tout d'abord à exprimer ma profonde gratitude à **Dieu** pour m'avoir accordé la force et la persévérance nécessaires pour mener à bien ce travail.

Ensuite, je souhaite remercier mes superviseurs d'université, **Zine Elabidine BOUBEKEUR** et **Merwan BEN BOUDIAF**, pour leur soutien inestimable tout au long de cette expérience académique. Leur expertise, leurs conseils éclairés et leur disponibilité ont été d'une valeur inestimable pour mener ce projet à terme.

Un remerciement tout particulier à mes amis qui ont été de véritables mentors et m'ont apporté une aide précieuse jusqu'à la dernière minute. Je tiens à nommer **Aimen Laalam**, **Habib Ouadi** et **Badr Eddine Berrehal**. Leur soutien indéfectible et leur amitié sincère ont été une source de motivation et d'inspiration tout au long de ce parcours.

Je souhaite également exprimer ma gratitude envers l'équipe de la société **SLB** qui m'a accueilli au sein de leur entreprise. Je tiens à remercier tout particulièrement **Monsieur KELKOULI Mohammed** et **Madame Asma Chekroune** pour m'avoir offert cette opportunité et pour leur soutien constant tout au long de cette expérience enrichissante. Je tiens également à remercier **Monsieur Mostafa Saadi**, mon promoteur au sein de l'entreprise, pour son soutien inconditionnel, son suivi attentif et ses précieux conseils. Son professionnalisme et sa patience exemplaires ont été des atouts majeurs dans la réussite de ce projet.

Enfin, je tiens à me remercier moi-même pour avoir cru en mes capacités et pour avoir persévéré malgré les défis rencontrés. Cette expérience m'a permis d'élargir mes horizons, d'accroître mes connaissances et de me surpasser.

Abstract

In today's dynamic landscape, the oil and gas industry face an ever-increasing array of challenges, pushing the boundaries of difficulty on a daily basis. Simultaneously, the global demand for energy continues to surge. In this evolving paradigm, the optimization of processes becomes paramount as automation reshapes the industrial landscape, transforming the way we approach traditional practices. Adapting to this transformative wave, the oil and gas industry has embraced digital solutions and cutting-edge equipment empowered by data analytics and artificial intelligence.

Within this industry, drilling engineering stands as a fundamental pillar, demanding unwavering dedication and expertise. The era of trial and error, though once tolerated, has proved costly, prompting a shift towards more efficient methods. Fortunately, this shift has yielded a treasure trove of invaluable data. Leveraging a diverse set of state-of-the-art computational technologies, we can now extract deep insights from this vast dataset, optimizing processes without incurring additional costs.

This thesis aims to provide researchers, engineers, and operators in the drilling and upstream disciplines with a comprehensive data-driven tool. With a wide range of functionalities, including predictive models and analytic tools, this tool leverages artificial intelligence to streamline operations and mitigate potential catastrophes. The project utilizes data from the esteemed Volve village, disclosed by Equinor in 2018, as a valuable resource to advance research and innovation in the oil and gas industry.

KEYWORDS: oil and gas industry, machine learning, lithology prediction, crookedness, magnetic interference.

Résumé

Dans le paysage dynamique d'aujourd'hui, l'industrie pétrolière et gazière est confrontée à une multitude de défis croissants, repoussant les limites de la difficulté au quotidien. Parallèlement, la demande mondiale en énergie continue de croître. Dans ce paradigme en constante évolution, l'optimisation des processus devient primordiale alors que l'automatisation transforme le paysage industriel, révolutionnant notre approche des pratiques traditionnelles. S'adaptant à cette vague de transformation, l'industrie pétrolière et gazière a adopté des solutions numériques et un équipement de pointe alimentés par l'analyse des données et l'intelligence artificielle.

Au sein de cette industrie, l'ingénierie de forage occupe une place fondamentale, exigeant une dévotion et une expertise inébranlables. L'ère de l'essai-erreur, autrefois tolérée, s'est avérée coûteuse, incitant à un passage vers des méthodes plus efficaces. Heureusement, ce changement a produit une mine de données inestimables. En exploitant un ensemble diversifié de technologies informatiques de pointe, nous pouvons désormais extraire des informations approfondies de cet ensemble de données volumineux, optimisant les processus sans coûts supplémentaires.

Cette thèse vise à fournir aux chercheurs, ingénieurs et opérateurs des disciplines du forage et de l'amont un outil complet basé sur les données. Avec une large gamme de fonctionnalités, notamment des modèles prédictifs et des outils analytiques, cet outil exploite l'intelligence artificielle pour rationaliser les opérations et atténuer les catastrophes potentielles. Le projet utilise des données provenant du prestigieux village de Volve, divulgué par Equinor en 2018, en tant que ressource précieuse pour faire progresser la recherche et l'innovation dans l'industrie pétrolière et gazière.

Mots-clés : industrie pétrolière et gazière, apprentissage automatique, prédiction de la lithologie, déviation, interférence magnétique.

ملخص

في الواقع الديناميكي الحالي، تواجه صناعة النفط والغاز تحديات متزايدة باستمرار، مما يدفع بالحدود الصعوبة إلى التحدي يومياً. في الوقت نفسه، تستمر الطلب العالمي على الطاقة في الارتفاع. في هذا النمط المتطور، يصبح تحسين العمليات أمراً بالغ الأهمية مع تشكيل الأتمتة المشهد الصناعي، مما يحول طريقتنا في التعامل مع الممارسات التقليدية. تكيفاً مع هذه الموجة التحويلية، اعتمدت صناعة النفط والغاز حلولاً رقمية ومعدات متطورة تعتمد على تحليل البيانات والذكاء الاصطناعي.

في هذه الصناعة، تعتبر هندسة الحفر أحد الأركان الأساسية المطلوبة والمتطلبة بدرجة عالية. فترة التجربة والخطأ، التي كانت تسمح بها في الماضي، تكلفت الكثير، مما حفز التحول نحو طرق أكثر كفاءة. لحسن الحظ، أدى هذا التحول إلى توفر مجموعة ضخمة من البيانات الثمينة. من خلال استخدام مجموعة متنوعة من التكنولوجيات الحوسبية المتطورة، يمكننا الآن استخراج رؤى عميقة من هذا النطاق الواسع من البيانات، مما يساهم في تحسين العمليات بدون تكاليف إضافية.

تهدف هذه الأطروحة إلى تزويد الباحثين والمهندسين والمشغلين في مجالات الحفر والمرحلة العليا بأداة قائمة على البيانات شاملة. مع مجموعة واسعة من الوظائف، بما في ذلك النماذج التنبؤية وأدوات التحليل، يستغل هذه الأداة قوة الذكاء الاصطناعي لتبسيط العمليات والتخفيف من الكوارث المحتملة. يستخدم المشروع بيانات من قرية Volve الرائعة، التي كشفت عنها شركة Equinor في عام 2018، كمورد قيم لتعزيز البحث والابتكار في صناعة النفط والغاز.

كلمات مفتاحية: صناعة النفط والغاز، التعلم الآلي، توقع الليثولوجيا، الانحراف، التداخل المغناطيسي

Table des matières

Dédicace	I
Remerciements	II
Résumés	III
Introduction Générale	1
I Principes fondamentaux de l'ingénierie du forage directionnel et sa synergie avec la lithologie	10
I.1 Introduction	11
I.2 Forage Directionnel	11
I.2.1 Applications du Forage Directionnel	11
I.2.1.1 Plusieurs Puits à partir d'un Seul Emplacement	12
I.2.1.2 Emplacements de Surface Inaccessibles	12
I.2.1.3 Forage de Dérivation	12
I.2.1.4 Forage de Failles	12
I.2.1.5 Exploration de Dômes de Sel	13
I.2.1.6 Forage de Puits de Secours	13
I.2.1.7 Puits Horizontaux	14
I.2.1.8 Forage Multilatéral	15
I.2.1.9 Conception des Puits	15
I.2.2 Enquête directionnelle (Directional Survey)	16
I.2.2.1 Instruments d'arpentage	17
I.2.2.2 Méthodes de calcul	19
I.2.3 Sources d'erreurs dans le levé directionnel	20
I.2.3.1 Instruments d'arpentage	20
I.2.3.2 Désalignement de l'outil	21
I.2.3.3 Erreur de profondeur (MD)	21
I.2.3.4 Interférence magnétique	21
I.2.3.5 Erreur de position du trou de forage	24
I.2.3.6 Contrôle de la qualité de l'enquête	24
I.2.3.7 Conception BHA pour le contrôle directionnel	25
I.2.4 Causes de la déviation du trou	28
I.2.5 Géopositionnement d'un puit en temps réel "Geosteering"	29
I.2.6 Cartographie pendant le forage	30
I.3 Méthodes conventionnelles d'identification de la lithologie	30
I.3.1 Introduction	30
I.3.2 Détermination directe de la lithologie	31
I.3.2.1 Diagraphie de Boue	31
I.3.2.2 Prélèvements de carottes	31

I.3.3	Détermination indirecte de la lithologie	31
I.3.4	Détermination indirecte de la lithologie	31
I.3.4.1	Les réponses diagraphiques	32
I.3.4.2	Spectroscopie des rayons gamma (GR)	32
I.3.4.3	Détermination de la salinité de l'eau	32
I.3.4.4	Utilisation des enregistrements acoustiques	32
I.3.5	Caractérisation de la lithologie et de l'hétérogénéité du réservoir	34
I.3.5.1	Propriétés des minéraux argileux	34
I.3.5.2	Évaluation du volume de schiste	35
I.3.5.3	Zonage ou stratification des réservoirs	35
I.4	Conclusion	36
II	La Révolution de l'Apprentissage Automatique dans l'Industrie Pétrolière	37
II.1	Introduction	38
II.2	IA, ML, DL et DM : Concepts clés	38
II.2.1	Intelligence Artificielle (IA)	38
II.2.2	Exploration de Données (DM)	38
II.2.3	Apprentissage Automatique (ML)	38
II.2.4	Apprentissage Profond (DL)	39
II.3	Paradigmes de l'Apprentissage Automatique : Une Comparaison	39
II.3.1	Modes d'Apprentissage en IA	39
II.3.1.1	Apprentissage Supervisé	39
II.3.1.2	Apprentissage Non Supervisé	40
II.3.1.3	Apprentissage Semi-Supervisé	40
II.3.1.4	Apprentissage par Renforcement	41
II.4	Gestion des Données en Apprentissage Automatique	41
II.4.1	Définition du Problème et Objectifs	41
II.4.1.1	Formulation du Problème	41
II.4.2	Acquisition et Sources de Données	42
II.4.3	Audit et Prétraitement des Données	42
II.5	Construction et Évaluation des Modèles d'Apprentissage Automatique	43
II.5.1	Sélection des Algorithmes d'Apprentissage Automatique	43
II.5.1.1	Algorithmes basés sur la Régression	43
II.5.1.2	Algorithmes basés sur les Arbres	44
II.5.1.3	Machines à Vecteurs de Support (SVM)	44
II.5.1.4	Algorithmes basés sur la Distance	44
II.5.1.5	Algorithmes de Regroupement	45
II.5.1.6	Réseaux Neuronaux et Apprentissage Profond	46
II.6	Entraînement et Évaluation des Modèles	47
II.6.1	Division Entraînement/Test	47
II.6.2	Ajustement du Modèle et Surapprentissage	47
II.6.3	Évaluation du Modèle	48
II.6.3.1	Pour les Problèmes de Régression	48
II.6.3.2	Pour les Problèmes de Classification	48
II.6.4	Interprétation des Résultats	49
II.7	Application de l'apprentissage automatique dans l'ingénierie de forage	49
II.7.1	Prédiction en temps réel de l'usure des outils de forage	51
II.7.2	Prévention des pertes de boue de forage pendant la circulation	51
II.7.3	Amélioration de l'efficacité du processus de forage	52
II.8	Conclusion	53

III Conception et réalisation de la solution digital TerraVerse	54
III.1 Introduction	55
III.2 Objectifs du projet	55
III.3 La quête des données	56
III.3.1 Contenu des données	58
III.3.1.1 Well Logs	58
III.3.1.2 Données techniques du puits	58
III.3.1.3 WITSML Real-Time Drilling Data	59
III.3.2 Défis de l'extraction de données et mitigation	61
III.3.2.1 Contrôle de la qualité des données	61
III.4 TeraVerse Solution	62
III.4.1 GeoVision	63
III.4.1.1 Problématique	63
III.4.1.2 Analyse exploratoire des données	64
III.4.1.3 Application de l'apprentissage automatique	70
III.4.1.4 Déploiement du model	75
III.4.2 PathGuard	78
III.4.2.1 Problématique	78
III.4.2.2 Analyse exploratoire des données	78
III.4.2.3 Application de l'apprentissage automatique	88
III.4.2.4 Déploiement du model	93
III.4.3 Conclusion	95
Bibliographie	97
Annexe A - Technologies utilisées	100
Annexe B - Chronologie du projet	101
Annexe C - Comment extraire les données du champ	102

Liste des figures

Revue de littérature	4
I.1 Plusieurs Puits à partir d'un Seul Emplacement	12
I.2 Emplacements de Surface Inaccessibles	12
I.3 Forage de Dérivation	13
I.4 Forage de Failles	13
I.5 Exploration de Dômes de Sel	13
I.6 Forage de Puits de Secours	14
I.7 Classification des Puits Horizontaux	14
I.8 Puits Horizontal	15
I.9 Forage Multilatéral	15
I.10 Systèmes de Forage Multilatéral	16
I.11 Illustration de la visualisation d'une trajectoire de puits	16
I.12 Orientation dans les levés de forage directionnels magnétiques	19
I.13 Disposition des capteurs directionnels dans les techniques MWD	19
I.14 Principe du gyroscope	19
I.15 Physique spatiale et levés magnétiques	22
I.16 Les trois principales sources du champ magnétique terrestre.	22
I.17 Courants dans la magnétosphère externe.	23
I.18 Comparaison des paramètres du champ magnétique terrestre dans différents systèmes de coordonnées	23
I.19 Cartes du champ géomagnétique en 2015.0	24
I.20 Illustration des ensembles rotatifs	26
I.21 Illustration d'un ensemble moteur orientable (PDM)	26
I.22 Modules du système rotatif orientable (RSS) par SLB	27
I.23 Stabilisateur à diamètre réglable	27
I.24 Opérations de cuillère de forage	27
I.25 Déviation du puits de forage par forage à jets	28
I.26 Exemple de déviations de trou (PétroWiki)	28
I.27 Illustration de la géosteering	30
I.28 Cartographie pendant le forage	30
I.29 Différentes méthodes d'échantillonnage de carottes	31
I.30 Comparaison des réponses des enregistrements dans les lithologies de réservoir	33
I.31 Diagramme croisé de sensibilité aux variations de lithologique : Δt_p vs. Δt_s	33
I.32 Diagramme croisé de V_p/V_s vs. Δt_p : Identification des intervalles de sable et de turbidites	34
I.33 Indication de la lithologie par la lenteur de cisaillement	34
I.34 Zonation du réservoir du réservoir de Sadlerochit, champ de Prudhoe Bay, Alaska, États-Unis	35
I.35 Diagramme croisés de réservoir à partir d'un champ gazier asiatique	36

II.1	Diagramme de Venn représentant les relations entre l'IA, le ML et le DL.	39
II.2	Classification vs Régression.	40
II.3	Aperçu des Catégories de l'Apprentissage Automatique.	41
II.4	Visualisations courantes dans l'Analyse Exploratoire des Données, Visualisation efficace des données.	43
II.5	Régression linéaire vs régression logistique.	44
II.6	Illustration des arbres de la forêt aléatoire.	45
II.7	Illustration du boosting gradient.	45
II.8	Illustration des Machines à Vecteurs de Support.	45
II.9	Illustration des k-plus proches voisins.	46
II.10	Avant et après l'algorithme k-means.	46
II.11	Regroupement spatial basé sur la densité.	46
II.12	Une illustration d'un réseau neuronal profond.	47
III.1	Localisation du champ de Volve en mer du Nord (adaptée de Seoyoon Kwon et al. 2021)	56
III.2	Microsoft Azure Storage Explorer Interface	57
III.3	Exemple d'un fichier DDR de type XML)	59
III.4	Exemple d'un fichier WITSML	60
III.5	Page d'accueil en cours de conception	62
III.6	Equipe de TerraVerse	63
III.7	Couverture des données	64
III.8	Décompte de fréquence des classes de lithologie	65
III.9	Distribution des caractéristiques dans les données de forage	65
III.10	Aperçu de la répartition des valeurs de WOB à la fois dans l'histogramme et le boxplot	66
III.11	Plot de corrélations avant le traitement des données(Heatmap).	68
III.12	Heatmap après le traitement des données.	69
III.13	Visualisation de quelques paramètres des données utilisées en logs	69
III.14	Matrice de diagramme des paramètres les plus importants(Pair Plots)	70
III.15	Illustration de la technique SMOTE	70
III.16	Illustration de la Validation Croisée à Cinq Plis par Ivanna Baturynska	71
III.17	Classement de précision	72
III.18	matrice de confusion	72
III.19	Bar plot pur illustrer l'importance des caractéristiques	73
III.20	Comparaison entre la lithologie réelle et prédite - Puits 1	74
III.21	Comparaison entre la lithologie réelle et prédite - Puits 2	75
III.22	Geovision interface	76
III.23	Étapes à suivre lors de l'utilisation de l'application web GeoVision	77
III.24	Visualisation des données utilisées(Statique)	80
III.25	Visualisation interactifs en ordre horizontal paramètres des données	81
III.26	Tracé des paramètres mécaniques	82
III.27	Plot de corrélations avant le traitement des données	84
III.28	Heat Map après traitement des données	85
III.29	matrice des diagrammes de dispersion (scatterplot matrix)	86
III.30	Module de Young vs Densité	87
III.31	Densité vs Résistivité	87
III.32	Le décompte pour chaque classe avant SMOTE	87
III.33	Le décompte pour chaque classe après SMOTE	88
III.34	Comparaison des scores par modèle	89
III.35	Métriques du modèle Réseau de Neurones	90
III.36	Matrice de confusion du modèle RF	91

III.37	Importance des caractéristiques du modèle RF avec SHAP	92
III.38	Partial Dependence Plots (PDP)	93
III.39	Valeurs SHAP respectives pour la prédictions minimale et maximal du Déviation . .	93
III.40	Interface du Web App PathGuard	94
III.41	Directives d'utilisation de l'application	95

Liste des tableaux

III.1	Contenu des données Volve	57
III.2	Contenu du dossier WELL LOGS	58
III.3	Répertoires dans le Well Technical Data	59
III.4	Quelques Exemples des attributs des rapports de forage quotidienne	60
III.5	Données les plus importantes	61
III.6	Caractéristiques erronées des données de forage	66
III.7	Le pourcentage de données contenant des zéros	67
III.8	Métriques du classificateur RandomForest	72
III.9	rapport de classification	73
III.10	Summary of File Information	79
III.11	Comparison of Model Scores	89
III.12	Rapport de Classification	91

Liste des sigles et acronymes

AAPG	American Association of Petroleum Geologists
BHA	Bottom Hole Assembly
BHP	Bottomhole Pressure - Pression de fond
BLOB	Binary Large Object
BTU	British Thermal Units
CSV	Comma-separated Values
DDR	Daily Drilling Report - Rapport Quotidien de Forage
DTS	Drill Steam Test - Test de Tige de Forage
ECD	Equivalent Circulating Density - Densité de Circulation Équivalente
EKD	Early Kick Detection - Détection Anticipée des Venues
ELOT	Extended Leak-Off Tests
EWR	End Well Report - Rapports de Fin de Puits
FANG	Friction Angle - Angle de Frottement Interne
HPHT	High Pressure High Temperature
KDE	Kernel Density Estimation
LCM	Lost Circulation Material - Produits de Perte de Circulation
MD	Measured Depth - Profondeur mesurée
MEM	Mechanical Earth Model - Modèle mécanique terrestre
MLR	Mud Log Report - Rapport d'Enregistrement de Boue
MWD	Measurement While Drilling - Mesure en Cours de Forage
NMR	Nuclear Magnetic Resonance - Résonance Magnétique Nucléaire
NPT	None Productif Time - Temps Non Productif
OBD	Over-Balanced Drilling - Forage en Sur-Équilibre
PCA	Principal Component Analysis
PDF	Portable Document Format
PDC	Polycrystalline Diamond Compact - Outil Diamant Polycristallin Compact
RPM	Rotation or Revolution per Minute - Nombre de Révolution par Minute
RF	Random Forest
SPE	Society of Petroleum Engineers
SP	Spontaneous Potential - Polarisation Spontanée
SPP	Stand Pipe Pressure - Pression de Colonne Montante
SVM	Support Vector Machine
TDS	Top Drive System
UCS	Unconfined Compressive Strength - Résistance à la Compression Non Confinée
WITSML	Wellsite Information Transfer Standard Markup Language
WOB	Weight on Bit - Poids Appliqué sur l'Outil
XML	Extensible Markup Language

Introduction Générale

Contexte

Depuis les premières avancées de l'industrie pétrolière, l'ingéniosité humaine a été mise à contribution pour résoudre de nombreux défis. Parmi ces défis, le forage pétrolier a fait l'objet d'une attention particulière. Les professionnels du secteur ont réussi à atteindre des profondeurs considérables dans des environnements extrêmement difficiles en faisant preuve d'innovation sous différentes formes. Ces innovations ont touché aussi bien les matériaux et les équipements utilisés que les concepts scientifiques novateurs, tous visant à assurer le succès de ces opérations.

Dans le cadre de notre projet au sein de la division Construction de puits de SLB, nous nous sommes concentrés sur le développement de deux solutions digitales : GeoVision et PathGuard. En intégrant d'un ensemble diversifié de sources de données, de corrélation, et les techniques d'apprentissage automatique, pour relever les défis les plus fréquemment rencontrés dans le domaine exigeant du forage.

Problématique

Identification en temps réel de la lithologie

La lithologie est l'étude et la description des roches, en particulier celles rencontrées lors des opérations de forage. Elle implique d'identifier et de classer les roches en fonction de leurs caractéristiques physiques, telles que le contenu minéral, la texture, la couleur, la taille des grains et la composition[1][2][3].

Plusieurs méthodes conventionnelles sont utilisées pour identifier la lithologie lors du forage d'un puits. L'observation visuelle consiste à examiner les échantillons de roche ramenés en surface, en se concentrant sur des caractéristiques telles que la couleur, la texture et la taille des grains. Des échantillons de coupe sont prélevés périodiquement et analysés en laboratoire. La diagraphie de boue analyse la boue de forage, qui contient un mélange d'échantillons de roches et de fluides de formation, tandis que la diagraphie filaire utilise un outil spécialisé pour mesurer des propriétés telles que la radiation gamma et la résistivité électrique, générant ainsi un journal détaillé du puits contenant des informations sur la lithologie. L'évaluation de la formation combine plusieurs sources de données, y compris les journaux de diagraphie filaire et les données de diagraphie de boue, pour identifier avec précision la lithologie et d'autres propriétés de la formation.

Cependant, les derniers outils de mesure de la lithologie pendant le forage présentent également des limitations. Par exemple, les technologies comme Geosphere 360 offrent des capacités de cartographie dans plusieurs directions mais ne sont pas en mesure de mesurer au niveau ou en amont du trépan. De plus, l'emplacement de l'émetteur à environ 9 pieds avant le trépan introduit des retards dans la récupération des informations, ce qui pose des défis pour un forage efficace et optimisé de la trajectoire du puits.

Déviaton non intentionnelle du trou lors des opérations de forage

Le forage pétrolier exige une précision extrême pour atteindre les formations ciblées. Cependant, la déviaton non intentionnelle du trou représente un défi majeur, entraînant des complications coûteuses et des risques pour la sécurité. Les causes de cette déviaton sont multiples, notamment la nature hétérogène de la formation géologique, les caractéristiques du matériel de forage, l'hydraulique et le nettoyage inadéquat du trou[4].

Pour remédier à ce problème, les outils de mesure pendant le forage (MWD) ont été développés pour surveiller la trajectoire du trépan en temps réel. Cependant, ces outils sont confrontés à un autre défi : les interférences magnétiques. Lorsque le champ magnétique des outils de forage ou de la formation géologique perturbe les relevés de mesure, des erreurs peuvent survenir. Pour surmonter ces défis, notre étude propose d'utiliser des techniques d'apprentissage automatique et les données des mesures gyroscopiques pour évaluer les conditions de déviaton malgré les interférences magnétiques. Cette approche innovante pourrait améliorer la précision du forage et réduire les risques liés à la déviaton du trou.

Ces recherches visent à résoudre les problèmes complexes de déviaton du trou et d'interférences magnétiques. En comprenant les mécanismes sous-jacents et en développant des modèles d'apprentissage automatique adaptés, nous espérons ouvrir de nouvelles perspectives pour une meilleure gestion de ces défis. Notre objectif est d'améliorer la précision, la sécurité et l'efficacité des opérations de forage tout en réduisant les coûts associés à la déviaton du trou.

Objectif

Le but principal de cette étude est de développer des solutions d'apprentissage automatique pour relever deux défis importants de l'industrie pétrolière : la prédiction en temps réel de la lithologie et des sommets de formation à l'aide des paramètres de forage, et l'assurance qualité et la prédiction des déviations de trajectoire des puits. En atteignant ces objectifs, cette étude vise à améliorer l'efficacité et la prise de décision dans les processus de forage pétrolier.

Organisation de la Thèse

Ce mémoire est organisé en trois chapitres, chacun explorant un aspect spécifique de l'étude et contribuant à la compréhension globale du sujet.

Dans ce premier chapitre « **Principes fondamentaux sur l'ingénierie du forage directionnel et sa synergie avec la lithologie** » nous nous concentrerons sur l'importance du géosteering dans l'industrie pétrolière et gazière. Nous aborderons deux aspects clés : l'identification de la lithologie et la déviation du trajet de puits. Nous examinerons les défis rencontrés dans ces domaines et les méthodes traditionnelles utilisées pour les résoudre. Ensuite, nous présenterons les avancées récentes dans l'utilisation de l'apprentissage automatique pour améliorer ces processus. Nous explorerons les techniques et les modèles d'apprentissage automatique qui ont été développés pour prédire la lithologie et détecter les déviations du trajet de puits en temps réel.

Dans ce deuxième chapitre « **La Révolution de l'Apprentissage Automatique dans l'Industrie Pétrolière** » nous nous concentrerons sur l'utilisation de l'apprentissage automatique dans l'industrie pétrolière et gazière de manière plus générale. Nous examinerons les principes fondamentaux de l'apprentissage automatique et ses différentes techniques, telles que la régression, la classification et les réseaux neuronaux. Nous explorerons également les algorithmes d'apprentissage automatique les plus couramment utilisés dans le domaine de l'ingénierie pétrolière, tels que la régression linéaire, les machines à vecteurs de support et les forêts aléatoires. Nous mettrons en évidence les avantages et les applications de ces techniques dans l'optimisation des opérations de forage.

Dans ce troisième chapitre « **Conception et réalisation de la solution digital TerraVerse** » nous nous concentrerons sur les deux solutions digitales que nous avons développées : PathGuard et Geo-Vision. Nous expliquerons en détail les fonctionnalités et les avantages de chaque solution. PathGuard est un outil d'assurance qualité qui utilise l'apprentissage automatique pour prédire les déviations du trajet de puits. GeoVision est une application web conviviale qui utilise des paramètres de forage pour prédire en temps réel la lithologie et les sommets de formation. Nous décrirons les méthodologies et les algorithmes utilisés pour développer ces solutions, ainsi que les résultats obtenus lors des tests et des validations. Enfin, nous discuterons des implications pratiques de ces solutions et de leur potentiel pour l'industrie pétrolière et gazière.

Revue de littérature

Défi 1 : Identification de la lithologie et des sommets de formation

Qi & Carr (2006) ont étudié la capacité à prédire avec précision les lithofaciès dans les réservoirs carbonatés [5]. Ceci est essentiel pour le forage efficace et sûr, car cela peut aider à identifier les dangers potentiels tels que la roche fracturée ou les zones à haute pression. Un réseau neuronal à une seule couche à rétropropagation (NN) a été utilisé pour prédire les lithofaciès à partir des diagraphies numériques dans les réservoirs de calcaire de St. Louis, dans le sud-ouest du Kansas. Le modèle a été entraîné en utilisant des données d'environ 100 puits, en se concentrant sur six entrées principales - le rayonnement gamma (GR), la résistivité profonde et moyenne, la densité, la porosité neutronique, et les diagraphies photoélectriques (PE), chaque fois qu'elles sont disponibles. Les modèles prédictifs, combinés à des descriptions détaillées des lithofaciès, ont efficacement reconnu six lithofaciès majeurs dans le réservoir, améliorant l'interprétation stratigraphique et la caractérisation du réservoir. Les modèles ont démontré une précision impressionnante allant de 70,37% à 90,82%, présentant une approche robuste pour la prédiction des lithofaciès dans des réservoirs carbonatés complexes.

Al-Anazi & Gates (2010) ont entrepris une tâche difficile : prévoir la perméabilité dans un réservoir de grès hautement hétérogène [6]. Ceci est une tâche difficile en raison du degré élevé d'hétérogénéité dans le réservoir, qui peut rendre difficile la prédiction précise de la perméabilité. Cinq paramètres d'entrée principaux (Rayon Gamma, Résistivité Profonde, Porosité Neutronique, Densité, et Temps de Voyage) ont été identifiés en utilisant un algorithme de classement flou en deux étapes. La méthode SVM a montré des capacités de classification supérieures et des prédictions de perméabilité par rapport à d'autres techniques. De plus, il a été découvert que les corrélations de perméabilité basées sur le noyau surpassaient généralement les corrélations basées sur les diagraphies. L'étude a conclu que les SVM, en raison de leur base de théorie de l'apprentissage statistique, étaient efficaces pour classer de nouveaux motifs dans les électrofaciès correspondants et pour attribuer une valeur de perméabilité à chaque point de données dans un espace de diagraphie à entrées multiples.

Raeesi et al., (2012) présentent une technique pour classer les lithofaciès et leur hétérogénéité dans les réservoirs d'hydrocarbures en utilisant les attributs sismiques, les données de diagraphies et les réseaux neuronaux artificiels (ANNs) [7]. Ils utilisent des réseaux d'apprentissage compétitifs pour l'apprentissage non supervisé afin d'identifier les grappes naturelles dans les données sismiques et le perceptron multicouche pour l'apprentissage supervisé formé avec des données de diagraphies pour anticiper les variations lithologiques. Les prédictions du modèle sont validées croisées avec des données de diagraphies distinctes, montrant un chevauchement significatif entre les résultats de classification non supervisés et supervisés. Le travail des auteurs suggère que les ANNs peuvent

assimiler efficacement les données sismiques et géologiques, améliorant la précision de l'exploration des réservoirs d'hydrocarbures et minimisant les risques de forage.

Wang & Carr, (2012) ont développé une méthodologie complète pour identifier et prédire les lithofaciès de schiste dans le Schiste de Marcellus du Bassin des Appalaches[8]. Ils ont utilisé différents outils tels que l'analyse discriminante linéaire, la logique floue, les réseaux neuronaux artificiels et les machines à vecteurs de support. Leur étude a stratégiquement intégré cinq types de données : les données d'analyse de carottes, les données de diagraphies, les données sismiques, les données d'analyse pétrophysique et les méthodologies de reconnaissance de formes. Ils ont commencé par identifier les lithofaciès à l'échelle de la carotte sur la base de la composition minérale. Ces données ont ensuite été projetées à l'échelle du puits à l'aide de divers modèles prédictifs et finalement étendues à l'échelle régionale pour créer un modèle 3D détaillé de lithofaciès. Cette méthodologie a un impact significatif sur la compréhension et l'exploitation des réservoirs de gaz de schiste non conventionnels, aidant à améliorer l'efficacité de l'exploration et à minimiser les risques d'extraction.

Sebtosheikh & Salehi (2015) ont mené une étude sur la prédiction de la lithologie dans un dépôt de carbonate multicouche en Iran, motivés par l'intérêt croissant de l'industrie pour cette zone au cours des dernières années [9]. Les machines à vecteurs de support (SVM) ont été utilisées sur la base de la théorie de l'apprentissage statistique. La recherche utilise les diagraphies de puits pétrophysiques et les données d'attributs sismiques inversées, en se concentrant sur les attributs PEF, NPFI, DT, RHOB, cohérence, et chaos. La méthode de classification SVM est appliquée, et différentes fonctions de noyau sont testées, y compris RBF, polynomiale, et les noyaux polynomiaux normalisés. Les effets de la taille de l'ensemble de données d'entraînement sur les performances des SVM sont également examinés. Les résultats montrent que les SVM se comportent bien dans la prédiction de la lithologie, et la fonction de noyau polynomiale normalisée est recommandée pour les petits ensembles de données d'entraînement. L'étude fournit des informations précieuses pour une prédiction précise de la lithologie dans le réservoir étudié.

Avanzini et al. (2016) ont pris un défi unique en étudiant les réservoirs non conventionnels en présentant un flux de travail pour identifier les "sweet spots" productifs, en se concentrant sur la Formation de Shale Barnett [10]. Les réservoirs non conventionnels présentent des défis uniques pour évaluer leur potentiel de production, nécessitant des flux de travail innovants qui intègrent diverses sources de données et analyses. La méthodologie combine des concepts de qualité de réservoir (RQ) statiques et dynamiques pour évaluer la productivité du champ. L'analyse de cluster en utilisant les algorithmes de Ward et K-means est utilisée pour la classification des faciès lithologiques et géomécaniques. En intégrant les faciès lithologiques (LF) et géomécaniques (GF), le flux de travail identifie les "sweet spots" productifs sur la base des RQ statiques et dynamiques. L'application de cette méthodologie aux puits de la Formation Shale Barnett a abouti à l'identification réussie des intervalles productifs dans la partie supérieure et moyenne du Lower Barnett. L'étude démontre la robustesse du flux de travail proposé pour identifier les "sweet spots" productifs et met en évidence son potentiel pour optimiser la modélisation 3D du réservoir. Cependant, des défis subsistent dans l'intégration efficace des données géomécaniques, et une amélioration supplémentaire est nécessaire pour définir plus précisément la fragilité.

Al-Mudhafar (2017) explore le rôle crucial de la caractérisation des réservoirs dans la compréhension des formations souterraines et l'optimisation de la production d'hydrocarbures [11]. Dans cette étude, un flux de travail intégré pour la classification des lithofaciès et la modélisation de la perméabilité du cœur en utilisant des réseaux neuronaux probabilistes (PNNs) et le modèle de régression boosté généralisé (GBM) est présenté. L'algorithme PNN atteint un taux de réussite de classification élevé de 95,81% pour la prédiction des lithofaciès sur la base des données de diagraphies de puits, tandis que l'algorithme GBM établit une relation non linéaire entre la perméabilité du cœur, les données de diagraphies de puits, et les lithofaciès. Cela se traduit par une modélisation précise de la perméabilité avec un excellent accord entre les valeurs de perméabilité prédites et observées. La méthodologie est appliquée aux données de diagraphies de puits et de carottes du réservoir de grès supérieur du champ pétrolifère de South Rumaila en Irak. Les limitations incluent la multicollinéarité potentielle de certains prédicteurs et la nécessité d'une validation supplémentaire dans différents contextes de réservoir. Globalement, le flux de travail proposé offre une approche efficace pour améliorer la classification des lithofaciès et la modélisation de la perméabilité dans la caractérisation des réservoirs.

Gu et al. (2019) introduisent une méthodologie améliorée pour la prédiction de la lithologie dans des réservoirs complexes, visant à répondre aux limitations associées à la qualité des données brutes et à la sélection des paramètres dans les réseaux neuronaux probabilistes (PNNs) [12]. L'approche proposée intègre un prétraitement par les machines de Boltzmann restreintes continues (CRBMs) et une optimisation utilisant l'optimisation par essaim de particules (PSO). En exploitant ces techniques, l'étude démontre des précisions de prédiction élevées (>75%) pour la lithologie complexe à travers trois expériences réalisées en utilisant des données du champ pétrolifère IARA. La méthodologie CRBM-PSO-PNN s'avère efficace pour surmonter les défis associés à la prédiction de la lithologie complexe, apportant un soutien précieux pour le travail géologique. Cependant, il est important de noter que la sélection des paramètres empiriques pour chaque réseau repose toujours sur un processus manuel, indiquant la nécessité de recherches supplémentaires pour développer des techniques de détermination automatique des paramètres.

Imamverdiyev & Sukhostat (2019) ont montré leur intérêt pour la caractérisation des réservoirs qui joue un rôle crucial dans la compréhension des formations souterraines et l'optimisation de la production d'hydrocarbures. En conséquence, ils ont présenté un modèle d'apprentissage en profondeur efficace pour la classification des faciès géologiques dans les puits en utilisant un réseau neuronal convolutif 1D (CNN) [13]. Le modèle utilise l'effet photoélectrique, le rayon gamma, la résistivité, la différence de porosité neutron-densité, la porosité moyenne neutron-densité, et des variables de contrainte géologique comme données d'entrée. Il surpasse d'autres modèles tels que le réseau neuronal récurrent, la mémoire à long terme, la machine à vecteurs de support et les k-plus proches voisins en termes de précision. Les expériences menées sur les données de diagraphie de puits du champ de gaz Hugoton démontrent la performance supérieure du modèle proposé en matière de discrimination lithologique. Le modèle 1D-CNN(Adagrad) obtient la meilleure précision et stabilité parmi les algorithmes d'optimisation évalués. Les résultats mettent en évidence la capacité du modèle à classer avec précision les différents types de faciès et les faciès adjacents, ce qui en fait un outil précieux pour l'interprétation des diagraphies de puits et la caractérisation des réservoirs.

upta et al. (2020) ont reconnu l'importance des applications de geosteering en temps réel dans les réservoirs de pétrole et ont proposé un flux de travail d'apprentissage machine pour prédire la lithologie au niveau du trépan pendant le forage [14]. Dans leur étude, les auteurs visent à exploiter les données de performance du trépan et du train de tiges pour atteindre cet objectif. Pour commencer, ils ont utilisé un jeu de données public du champ Volve et ont regroupé les logs sensibles à la lithologie en trois électrofaciès : les lithologies riches en quartz, riches en argile, et riches en carbonates. Les données de carottes ont ensuite été utilisées pour interpréter ces électrofaciès en termes de minéralogie. L'étape suivante consistait à prédire les électrofaciès en utilisant diverses variables de mesure pendant le forage (MWD), telles que le taux de pénétration et le poids sur le trépan. En utilisant des algorithmes de classification supervisés, les chercheurs ont atteint une impressionnante précision de prédiction de la lithologie de plus de 80% dans les puits de test. Les résultats mettent non seulement en évidence le potentiel des systèmes d'automatisation/conseil de forage pour réduire les coûts et améliorer l'efficacité des applications de geosteering, mais démontrent également l'adaptabilité de la méthodologie de l'étude à d'autres formations, champs et types de puits au-delà du champ Volve.

Défi 2 : Forage directionnel et déviations du trajet du puits

Lesso et al. (2001) mettent en lumière l'impact transformateur des mesures continues de direction et d'inclinaison sur la prise de décision en temps réel pour le forage directionnel [15]. Dans leur étude, les chercheurs soulignent les limitations associées aux méthodes conventionnelles d'interpolation de relevés lorsqu'il s'agit de traiter les problèmes de tortuosité des puits. Pour surmonter ces limitations, ils proposent une nouvelle méthodologie qui exploite les données continues provenant des outils de mesure pendant le forage (MWD) et des paramètres de forage en surface/en fond de trou. En utilisant cette approche, les chercheurs permettent l'étalonnage et la prédiction en temps réel de la tendance directionnelle de l'assemblage de fond de trou (BHA). L'objectif principal de leur méthodologie est d'optimiser les séquences de glissement/rotation, d'obtenir des puits plus lisses, et de fournir des informations précoces pour un ciblage plus précis. À travers une étude de cas complète menée dans le golfe du Mexique, les chercheurs démontrent l'efficacité de leur approche pour réduire la régionalité et améliorer les opérations de forage. La combinaison de mesures continues et d'inclinaison au niveau du trépan facilite la visualisation en temps réel, permettant une évaluation plus complète de l'impact des changements de direction.

Pirovolou et al. (2011) ont introduit un système de contrôle de trajectoire pour les systèmes rotatifs directionnels (RSS) dans l'industrie pétrolière et gazière [16]. Le système utilise des données en temps réel pour "apprendre" le comportement de l'assemblage de forage et générer des recommandations précises, incluant le taux de construction souhaité, le taux de rotation, le visage d'outil gravitaire, et le ratio de direction. Ces recommandations sont présentées au foreur directionnel pour approbation avant d'être transmises à l'outil RSS. Des tests et des validations poussés ont été réalisés pour évaluer les performances du système, démontrant son efficacité dans l'optimisation du guidage du puits, le placement du puits, et les efforts anti-collision. Les résultats préliminaires indiquent que le système fournit constamment des recommandations précises et a le potentiel de forer de manière autonome des sections d'un plan de puits, améliorant ainsi les opérations de forage dans diverses conditions.

Zhang and Samuel (2015) ont montré leur intérêt pour le forage directionnel et ses défis [17]. Ils ont introduit un modèle analytique pour estimer la tendance directionnelle des assemblages de fond de trou à pointe et à poussée (BHAs) sans avoir besoin de simulations complexes ou de données extensives. L'accent de leur recherche est mis sur les systèmes rotatifs directionnels (RSS) utilisés dans les applications de forage directionnel. En utilisant la théorie des poutres continues et en développant des équations pour calculer la force latérale sur le trépan et la gravité de l'angle du chien (DLS) d'un puits, ils fournissent des informations précieuses pour prédire la direction du forage. Les auteurs discutent des modèles rigides et élastiques pour les BHAs à poussée, en examinant des paramètres tels que les angles de construction, les angles d'inclinaison, et les angles de poussée. À travers une étude de cas, ils démontrent l'influence de ces paramètres sur la force latérale et le DLS, ainsi que l'impact de la longueur du BHA sur l'efficacité du forage horizontal. Cette recherche contribue au contrôle de la trajectoire et améliore la compréhension des différents comportements des BHA dans les applications de forage directionnel.

Marck & Detournay (2016) ont exploré l'influence de la conception des systèmes rotatifs directionnels (RSS) sur le spirilage des trous de forage, abordant un aspect important du forage directionnel [18]. Leur article, intitulé "Influence de la conception des systèmes rotatifs directionnels sur le spirilage des trous de forage" dans le journal SPE, examine l'impact de la conception des RSS sur la tendance à forer des trous en spirale en analysant la stabilité directionnelle de la trajectoire du trépan. Utilisant un modèle qui intègre la rigidité latérale attribuée aux plaquettes RSS, les auteurs analysent comment différentes conceptions de RSS affectent la direction du forage et induisent des oscillations auto-excitantes dans la géométrie du trou de forage. Les résultats montrent qu'une rigidité accrue des plaquettes entraîne une plus grande probabilité de forage de trous en spirale et de génération d'oscillations de trou de forage. D'autre part, les systèmes avec une rigidité de plaquette plus faible produisent des trous en spirale avec une longueur d'onde correspondant à la distance entre le trépan et le premier stabilisateur. Cette recherche souligne l'importance de la conception des RSS pour obtenir la qualité de trou de forage souhaitée et fournit des informations précieuses sur l'initiation des trous en spirale.

Wang et al. (2017) ont cherché à améliorer le contrôle de la trajectoire du puits grâce au développement d'un modèle de prédiction de la trajectoire de forage pour l'assemblage de fond de trou rotatif directionnel à poussée (RSBHA) [[19]. En établissant un modèle mécanique basé sur la théorie de la poutre continue, les auteurs ont étudié la relation entre la force latérale du trépan, l'angle d'inclinaison du trépan, et divers paramètres d'influence. Ils ont développé un modèle de prédiction de la trajectoire de forage qui estime avec précision la variation de l'inclinaison et de l'angle d'azimut, ce qui correspond bien aux données expérimentales sur le terrain. Cette étude fournit des orientations pour la sélection des paramètres de direction afin d'obtenir un contrôle précis de la trajectoire du puits.

Pollock et al. (2018) ont exploré l'application de l'apprentissage automatique pour améliorer le forage directionnel dans l'industrie pétrolière et gazière [20]. Ils ont développé un système d'intelligence artificielle (IA) qui a appris des foreurs experts et des simulations de forage pour optimiser les opérations de forage. Le réseau neuronal entraîné a démontré la capacité de reproduire les décisions des foreurs experts avec une petite marge d'erreur, suggérant le potentiel des systèmes de conseil en temps réel et des opérations de forage directionnel automatisées.

Koryabkin et al. (2019) ont présenté une méthodologie basée sur les données pour la détermination en temps réel de la position et de la direction du trépan de forage 21. Leur approche utilisait des techniques d'apprentissage automatique et intégrait des paramètres mécaniques, des données d'orientation d'outil et des données MWD/LWD. La méthodologie a démontré une précision prometteuse dans la prédiction de la trajectoire du puits, montrant un potentiel de prédiction à haute précision et à haute fréquence pendant les opérations de forage. Les plans futurs comprennent un affinement supplémentaire de la méthodologie et son expansion à d'autres champs pétroliers.

Saasen et al. (2020) ont discuté de l'impact des débris magnétiques dans les fluides de forage sur le forage directionnel et ont proposé une méthode pour améliorer l'efficacité de l'élimination de la contamination magnétique [22]. En mettant en œuvre leur approche, ils ont obtenu une élimination de la contamination magnétique plus de cinq fois plus efficace par rapport aux systèmes traditionnels, ce qui a entraîné une amélioration de la précision dans le forage directionnel et le positionnement du puits. L'article explore également l'utilisation de fluides de forage ayant un impact minimal sur les mesures MWD pour améliorer encore la précision du forage directionnel.

Tunkiel et al. (2021) ont présenté une approche d'apprentissage en temps réel utilisant des réseaux neuronaux récurrents (RNN) pour la prédiction de l'inclinaison dans le forage directionnel [23]. Leur approche dynamique a formé le modèle sur des données en temps réel provenant de différents capteurs pendant le forage en cours, capturant ainsi le comportement dynamique et les motifs dans la série de données. Les résultats ont démontré des prédictions précises des données d'inclinaison avec un minimum de données d'entraînement, mettant en évidence les applications potentielles de la méthode proposée au-delà du domaine pétrolier.

Yu et al. (2022) ont analysé le risque d'interférence dans les puits horizontaux de gaz de schiste et ont utilisé une correction d'azimut magnétique pour atténuer les interférences magnétiques [24]. En corrigeant la direction du puits en fonction des relevés magnétiques, ils ont pu réduire le domaine d'interférence et minimiser le risque d'interférence, fournissant ainsi une référence théorique pour éviter de tels problèmes dans les puits horizontaux de gaz de schiste.

ElGizawy et al. (2023) ont proposé une méthode qui combine les technologies de mesure magnétique pendant le forage (MWD), les levés dynamiques définitifs (DDS) et les technologies de mesure gyroscopique pendant le forage (GWD) pour atteindre un positionnement précis du sondage [25]. Le DDS MWD a été amélioré en corrigeant les sources d'erreur potentielles, tandis que la technologie GWD a fourni une position supplémentaire du sondage. Les résultats ont démontré le plus haut niveau de précision de positionnement du sondage atteint à ce jour, permettant la pénétration réussie d'une petite cible de réservoir et la confirmation des objectifs du sondage. Cette technique a le potentiel d'améliorer considérablement la précision de positionnement du sondage, améliorant la production du réservoir et minimisant les risques de collision avec les puits adjacents.

Chapitre I

Principes fondamentaux de l'ingénierie du forage directionnel et sa synergie avec la lithologie

I.1 Introduction

Embarquez pour un voyage d'exploration alors que nous plongeons dans le domaine fascinant de l'ingénierie du forage. Ce chapitre sert de toile de fond captivante, plongeant les lecteurs dans le monde des opérateurs de première ligne et éclairant les innombrables défis auxquels ils sont confrontés. En fournissant des connaissances essentielles et en élargissant les perspectives, nous visons à doter les lecteurs d'une compréhension plus approfondie de cette branche fascinante et des complexités qu'elle implique. Rejoignez-nous alors que nous découvrons les subtilités et les intrications de ce domaine, suscitant un sentiment de curiosité et d'appréciation pour le travail de ceux qui sont à l'avant-garde de l'industrie.

L'ingénierie du forage occupe une place primordiale dans l'industrie pétrolière et gazière, jouant un rôle fondamental dans l'exploration et la production des ressources en hydrocarbures. Elle englobe la conception méticuleuse, la planification rigoureuse et l'exécution précise des opérations de forage pour accéder et extraire le pétrole et le gaz des réservoirs souterrains. La valeur de l'ingénierie du forage réside dans sa capacité à libérer et à optimiser le potentiel des réserves souterraines, apportant une contribution significative à l'approvisionnement mondial en énergie. En tirant parti des technologies et techniques avancées, les ingénieurs du forage s'efforcent d'optimiser le placement des puits, d'augmenter les taux de production et de garantir l'extraction sûre et efficace des hydrocarbures.

Une avancée remarquable dans l'ingénierie du forage est la pratique du forage directionnel, qui permet aux puits de s'écarter du plan vertical. Cette technique permet aux opérateurs d'accéder à plusieurs zones cibles au sein d'un réservoir, améliorant le contact avec le réservoir et optimisant l'efficacité de production. Le forage directionnel trouve de nombreuses applications, notamment la possibilité de forer plusieurs puits à partir d'un seul emplacement, d'explorer des zones de surface autrement inaccessibles et de naviguer dans des structures géologiques complexes telles que le forage de failles. Il offre une flexibilité dans le placement des puits, réduisant les coûts et l'impact environnemental tout en maximisant la récupération des hydrocarbures précieux.

Dans le domaine du forage directionnel, la géoconduite joue un rôle essentiel pour garantir un positionnement précis du puits et des ajustements en temps réel. La géoconduite implique la surveillance continue et l'analyse des données géologiques pendant les opérations de forage, permettant de prendre des décisions éclairées sur les ajustements de trajectoire du puits. En intégrant les informations provenant des mesures de diagraphies, des relevés sismiques et des modèles géologiques, la géoconduite facilite un positionnement précis du puits dans le réservoir, optimisant le contact avec les zones productives et évitant les formations non réservoir. Cette technologie améliore la caractérisation des réservoirs, améliore les performances des puits et maximise la récupération des hydrocarbures, favorisant le succès opérationnel dans l'industrie pétrolière et gazière.

I.2 Forage Directionnel

I.2.1 Applications du Forage Directionnel

La sous-section suivante explore diverses applications du forage directionnel, mettant en évidence sa polyvalence et son utilisation répandue[26].

I.2.1.1 Plusieurs Puits à partir d'un Seul Emplacement

Les développements de champs, en particulier en offshore et dans l'Arctique, impliquent le forage d'un nombre optimal de puits à partir d'une seule plateforme ou d'une île artificielle. Le forage directionnel a permis de réduire considérablement les coûts et l'impact environnemental de cette application.

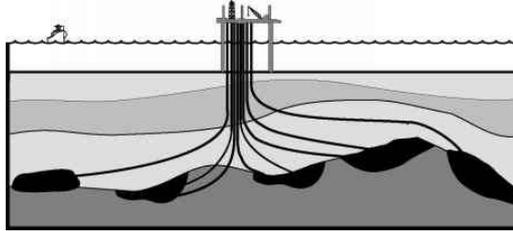


FIGURE I.1 – Plusieurs Puits à partir d'un Seul Emplacement

I.2.1.2 Emplacements de Surface Inaccessibles

Un puits est foré de manière directionnelle pour atteindre une zone de production autrement inaccessible avec les pratiques de forage vertical normales. L'emplacement d'une formation productrice dicte l'emplacement de la plateforme éloignée et le profil du puits directionnel. Des applications comme celle-ci sont où les puits "à extension" sont le plus couramment forés.

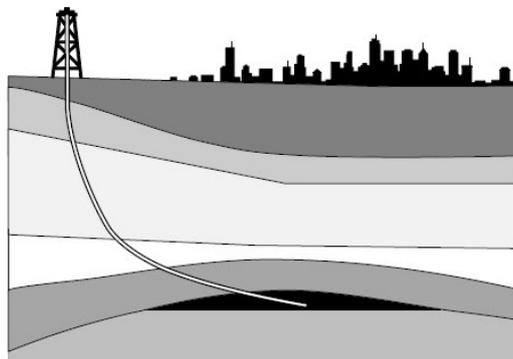


FIGURE I.2 – Emplacements de Surface Inaccessibles

I.2.1.3 Forage de Dérivation

Cette technique peut être utilisée soit pour contourner des obstacles, soit pour repositionner le fond du puits pour des raisons géologiques. Le forage autour d'obstacles, tels qu'une colonne de tuyaux perdue, est généralement réalisé avec une dérivation aveugle. Une dérivation orientée est nécessaire si une certaine direction est essentielle pour localiser une formation productrice prévue.

I.2.1.4 Forage de Failles

Il est souvent difficile de forer un puits vertical à travers un plan de faille incliné pour atteindre une formation hydrocarbonée sous-jacente. Au lieu de cela, le puits peut être dévié perpendiculairement ou parallèlement à la faille pour une meilleure production. Dans les zones instables, un puits foré à travers une zone de faille peut être en danger en raison de la possibilité de glissement ou de mouvement le long de la faille. Les pressions de formation le long des plans de faille peuvent également affecter les conditions du puits.

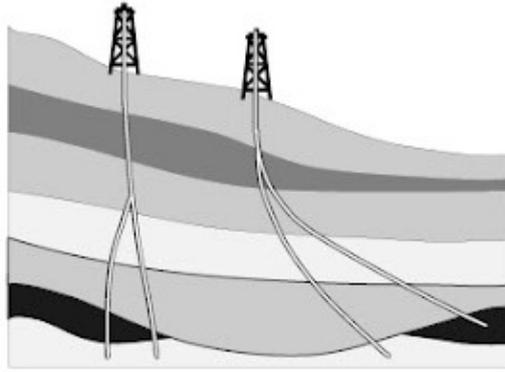


FIGURE I.3 – Forage de Dérivation

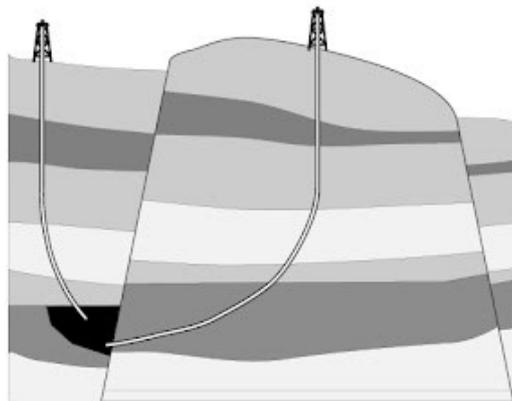


FIGURE I.4 – Forage de Failles

I.2.1.5 Exploration de Dômes de Sel

Les formations productrices peuvent être situées sous la calotte dure et surplombante des dômes de sel. Le forage d'un puits vertical à travers un dôme de sel augmente la possibilité de problèmes de forage, tels que les érosions, les pertes de circulation et la corrosion.

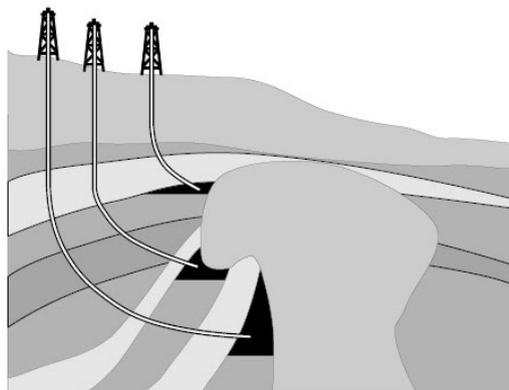


FIGURE I.5 – Exploration de Dômes de Sel

I.2.1.6 Forage de Puits de Secours

Un puits incontrôlé (sauvage) est intersecté près de sa source. La boue et l'eau sont ensuite pompées dans le puits de secours pour neutraliser le puits sauvage. Le contrôle directionnel est extrêmement

précis pour ce type d'application.

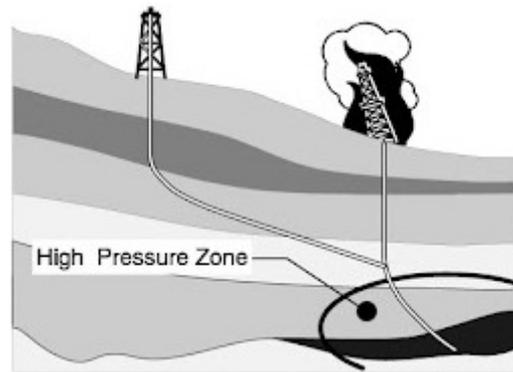


FIGURE I.6 – Forage de Puits de Secours

I.2.1.7 Puits Horizontaux

Une production réduite dans un champ peut être due à de nombreux facteurs, notamment le conage de gaz et d'eau ou des formations avec une perméabilité verticale satisfaisante. Les ingénieurs peuvent alors planifier et forer un drain horizontal. Il s'agit d'un type spécial de puits directionnel.

Les avantages des puits horizontaux comprennent :

1. Réduction du conage de l'eau et du gaz en raison d'une diminution de la surpression dans le réservoir pour un débit de production donné, réduisant ainsi les travaux de remédiation nécessaires à l'avenir.
2. Augmentation du débit de production en raison de la plus grande longueur du tubage exposée à la zone productrice.
3. Réduction de la chute de pression autour du tubage.
4. Baisse des vitesses de fluide autour du tubage.
5. Une réduction générale de la production de sable grâce à une combinaison des points 3 et 4.
6. Un schéma de drainage plus large et plus efficace conduisant à une meilleure récupération globale des réserves.

Les puits horizontaux sont généralement caractérisés par leurs taux de construction et sont classés en trois groupes qui dictent les pratiques de forage et de complétion requises, comme le montre la Figure I.7 [27].

Well Type	Build Rate (ft)	Radius (m)	Radius (ft)
Long radius	2 to 6°/100 ft	900 to 290	3,000 to 1,000
Medium radius	6 to 35°/100 ft	290 to 50	1,000 to 160
Short radius	5 to 10°/3 ft	12 to 6	40 to 20

FIGURE I.7 – Classification des Puits Horizontaux

Les taux de construction déterminent le changement d'inclinaison sur une longueur spécifiée. Les puits horizontaux de grand rayon ont des taux de construction de 2 à 6°/100 pieds, tandis que les puits de rayon moyen varient de 6 à 35°/100 pieds. Les puits de petit rayon ont des taux de construction de 5 à 10°/3 pieds. Ces profils différents nécessitent des outils et des techniques spécialisés et sont adaptés à des applications spécifiques telles que les déplacements longs, le forage terrestre et les projets de réentrée.

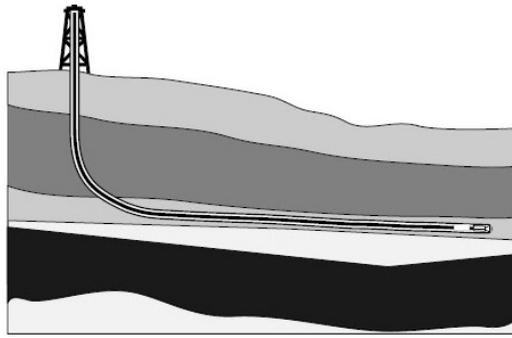


FIGURE I.8 – Puits Horizontal

I.2.1.8 Forage Multilatéral



FIGURE I.9 – Forage Multilatéral

Les puits multilatéraux sont des puits uniques comportant une ou plusieurs déviations à partir du puits d'origine (mère). La production est souvent mélangée. Bien qu'il s'agisse d'un puits à forte intensité de capital, un débit de production plus élevé est obtenu par rapport aux puits normaux. Un consortium industriel regroupant des opérateurs et des sociétés de services a été créé pour classer les puits multilatéraux en fonction de leur complexité et de leur fonctionnalité. Les catégories désignées (niveaux) sont les suivantes :

- Niveau 1 - jonction en trou découvert.
- Niveau 2 - sortie en trou tubé.
- Niveau 3 - jonction avec connexion mais sans joint.
- Niveau 4 - jonction scellée.
- Niveau 5 - jonction scellée mécaniquement avec diamètre interne réduit (ID).
- Niveau 6 - jonction scellée mécaniquement avec ID complet.
- Niveau 6S - séparateur de fond de trou.

I.2.1.9 Conception des Puits

Aujourd'hui, la plupart de la planification des puits directionnels est effectuée sur ordinateur. Les technologies informatiques modernes, telles que la visualisation 3D et les modèles géologiques 3D, ont fourni aux géoscientifiques et aux ingénieurs des outils intégrés et interactifs pour créer, visualiser et optimiser les trajectoires des puits à travers les cibles réservoir, comme le montre la figure 6.4. De plus, les systèmes de géoconduite récemment développés et les systèmes de forage RSS permettent des trajectoires de puits directionnels plus complexes conçues pour drainer une plus grande partie du réservoir. L'avenir réside dans l'intégration en temps réel des données de forage et de diagraphies-while-drilling (LWD) avec la géoconduite et le modèle géologique. La visualisation 3D des données en temps réel, associée au modèle géologique, permettrait une gestion intégrée des connaissances et une prise de décision en temps réel.

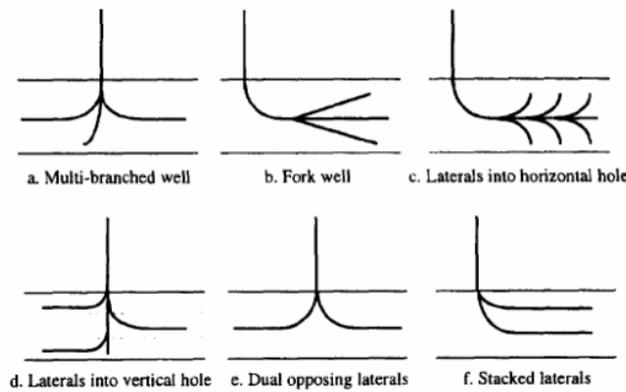


FIGURE I.10 – Systèmes de Forage Multilatéral

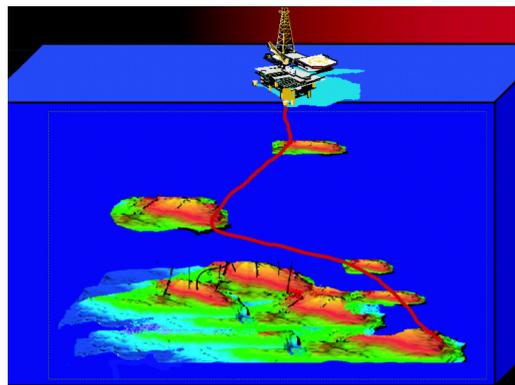


FIGURE I.11 – Illustration de la visualisation d'une trajectoire de puits

I.2.2 Enquête directionnelle (Directional Survey)

La méthode utilisée pour obtenir les mesures nécessaires pour calculer et tracer la trajectoire du puits tridimensionnel est appelée levé directionnel. Trois paramètres sont mesurés à plusieurs endroits le long de la trajectoire du puits : MD (depth), l'inclinaison et la direction du trou. MD est la profondeur réelle du trou foré à n'importe quel point le long du forage ou à la profondeur totale, mesurée à partir de l'emplacement de surface. L'inclinaison est l'angle, mesuré en degrés, par lequel l'axe du trou de forage ou de l'instrument de relevé s'écarte d'une ligne verticale réelle. Une inclinaison de 0° correspond à la verticale réelle, et une inclinaison de 90° correspond à l'horizontale. La direction du trou est l'angle, mesuré en degrés, de la composante horizontale de l'axe du trou de forage ou de l'instrument de relevé par rapport à une référence nord connue. Cette référence peut être le nord vrai, le nord magnétique ou le nord de grille, et est mesurée dans le sens des aiguilles d'une montre par convention. La direction du trou est mesurée en degrés et est exprimée soit en azimuth (0 à 360°), soit en quadrant (NE, SE, SO, NO) [2].

Chaque enregistrement de MD, d'inclinaison et de direction du trou est effectué à une station de levé, et de nombreuses stations de levé sont obtenues le long de la trajectoire du puits. Les mesures sont utilisées ensemble pour calculer les coordonnées tridimensionnelles, qui peuvent ensuite être présentées sous la forme d'un tableau de nombres appelé rapport de levé. Le levé peut être effectué pendant le forage ou après sa fin.

Les objectifs du levé directionnel sont les suivants :

- Déterminer l'emplacement exact du fond du puits pour surveiller les performances du réservoir.
- Surveiller la trajectoire réelle du puits pour garantir l'atteinte de l'objectif.
- Orienter les outils de déviation pour naviguer dans les trajectoires de puits.
- S'assurer que le puits n'intersecte pas d'autres puits à proximité.

- Calculer la TVD (True Vertical Depth) des différentes formations pour permettre la cartographie géologique.
- Évaluer le DLS (Dogleg Severity), qui est l'inclinaison angulaire totale et l'azimut dans le trou de forage, calculée sur une longueur standard (100 pieds ou 30 m).
- Satisfaire aux exigences des organismes de réglementation, tels que le Minerals Management Service (MMS) aux États-Unis.

I.2.2.1 Instruments d'arpentage

Les instruments d'arpentage peuvent être configurés en plusieurs variantes différentes, en fonction de l'utilisation prévue de l'instrument et des méthodes utilisées pour stocker ou transmettre les informations d'arpentage. Fondamentalement, il existe deux types d'instruments de levé : magnétique et gyroscopique. Selon la méthode utilisée pour stocker les données, il existe des systèmes cinématographiques et électroniques. Les systèmes de levé peuvent également être classés en fonction des méthodes utilisées pour transmettre les données à la surface, telles que le câble ou la mesure pendant le forage (MWD).

I.2.2.1.1 Capteurs magnétiques

Les capteurs magnétiques doivent être utilisés dans un environnement non magnétique [c'est-à-dire dans un trou non tubé, soit dans un ou plusieurs collerets de forage non magnétiques, soit sur un câble]. Dans tous les cas, il ne doit y avoir aucune interférence magnétique provenant de puits adjacents. Les capteurs magnétiques peuvent être classés en deux catégories : les boussoles mécaniques et électroniques.

Boussole mécanique

Une boussole mécanique utilise une rose des vents qui s'oriente vers le nord magnétique, tout comme l'aiguille d'une boussole de randonnée. L'inclinaison est mesurée à l'aide d'un pendule ou d'un flotteur. Dans le dispositif à pendule, le pendule est suspendu au-dessus d'une grille fixe ou le long d'une échelle vernier et est autorisé à se déplacer lorsque l'inclinaison change. Le dispositif à flotteur suspend un flotteur dans un fluide qui permet au tube de l'instrument de se déplacer indépendamment lorsque l'inclinaison change. Le seul avantage des boussoles mécaniques est leur faible coût, tandis que plusieurs inconvénients les ont empêchées d'être largement utilisées dans les levés directionnels. Les inconvénients incluent les coûts élevés de maintenance, la nécessité de choisir une plage d'inclinaison, une capacité de température limitée, la possibilité d'erreur humaine dans la lecture des films et l'impossibilité de les utiliser dans les outils MWD.

Boussole électronique

Le système de boussole électronique est un instrument de levé directionnel autonome à semi-conducteurs qui mesure les forces magnétiques et gravitationnelles de la Terre. L'inclinaison est mesurée à l'aide d'accéléromètres de gravité qui mesurent le champ gravitationnel terrestre selon les plans x, y et z. Le plan z est parallèle à l'axe de l'outil, x est perpendiculaire à z et aligné sur la fente de référence de l'outil, et y est perpendiculaire à la fois à x et z. À partir de cette mesure, les composantes vectorielles peuvent être additionnées pour déterminer l'inclinaison. La direction du trou est mesurée à l'aide d'accéléromètres de gravité et de magnétomètres à fluxgate. Les magnétomètres à fluxgate mesurent les composantes du champ magnétique terrestre orthogonalement (c'est-à-dire selon les mêmes trois axes que les accéléromètres). À partir de cette mesure, les composantes vectorielles peuvent être additionnées pour déterminer la direction du trou.

Selon l'emballage des capteurs électroniques, le système de boussole électronique peut être utilisé dans différents modes, tels que les mesures uniques, les mesures multiples et les mesures en temps

réel (MWD), où les données sont envoyées en surface en temps réel grâce au système de télémétrie par impulsions de boue.

Le champ géomagnétique terrestre est crucial pour les capteurs magnétiques afin de déterminer la direction du trou. Il sera discuté plus en détail dans la section sur les interférences magnétiques.

Renseignements à noter

Les capteurs d'accéléromètre mesurent le vecteur gravité \mathbf{G} . Ils se composent d'une masse d'épreuve suspendue sur une charnière, permettant le mouvement dans un axe. L'accélération fait bouger la masse d'épreuve, et ce mouvement est détecté. Un couple de rappel est appliqué pour ramener la masse à sa position initiale. En mesurant le courant requis pour le couple de rappel, l'accélération peut être calculée. Les outils de levé incluent généralement un réseau triaxial d'accéléromètres, avec un capteur le long de chaque axe fixe de l'outil. G est donc entièrement défini par les sorties des accéléromètres de l'outil, selon l'équation suivante :

$$B = \sqrt{(G_x^2 + G_y^2 + G_z^2)} \quad (\text{I.1})$$

Les capteurs de magnétomètre détectent et mesurent le vecteur du champ magnétique \mathbf{B} qui entoure l'ensemble de fond de trou. Ils utilisent la technologie des magnétomètres à fluxgate avec des bobines parallèles et des noyaux ferromagnétiques. Une bobine de détection répond à l'asymétrie des champs induits. En l'absence d'un champ ambiant, il n'y a pas de signal net dans la bobine de détection. En présence d'un champ extérieur, la bobine de détection détecte un courant proportionnel à la composante du champ magnétique environnant alignée avec les axes des noyaux. Les réseaux triaxiaux de magnétomètres à fluxgate mesurent les composantes de \mathbf{B} le long des axes fixes de l'outil. B est entièrement défini par les sorties des magnétomètres de l'outil, selon l'équation suivante :

$$B = \sqrt{(B_x^2 + B_y^2 + B_z^2)} \quad (\text{I.2})$$

D'autres termes pour B sont induction magnétique et densité de flux magnétique.

I.2.2.1.2 Capteurs gyroscopiques

Les instruments de levé gyroscopiques sont utilisés lorsque les systèmes de levé magnétiques sont influencés par des facteurs externes. Un gyroscope à rotor se compose d'une roue tournante entraînée par un moteur électrique, maintenant sa direction grâce à l'inertie. Il sert de référence pour mesurer l'azimut et utilise un agencement de cardan externe et interne pour maintenir la direction indépendamment de la position de l'instrument dans le puits.

Les systèmes gyroscopiques peuvent être catégorisés en trois types : les gyroscopes libres, les gyroscopes de vitesse et les systèmes de navigation inertielle.

- Gyroscopes libres : Le gyroscope à plate-forme stable commun maintient son orientation même lorsqu'il est incliné.
- Gyroscopes de vitesse (gyroscopes cherchant le nord) : Utilisent la vitesse de rotation de la Terre pour déterminer le nord. L'inclinaison est mesurée par un ensemble d'accéléromètres de gravité triaxiaux. Ils ont une petite dérive par rapport à la vitesse de rotation de la Terre mais nécessitent plus de temps pour trouver précisément le nord.
- Systèmes de navigation inertielle : La plus grande précision, utilisant plusieurs gyroscopes et accéléromètres de gravité pour orienter le système. Mesure le mouvement dans tous les axes sans sacrifier la précision, adapté à toutes les latitudes.

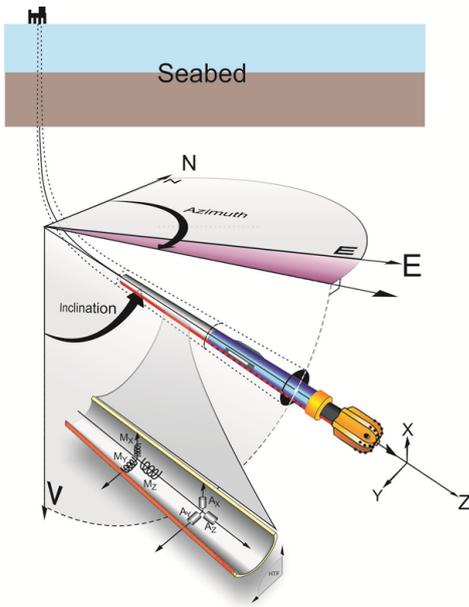


FIGURE I.12 – Orientation dans les levés de forage directionnels magnétiques

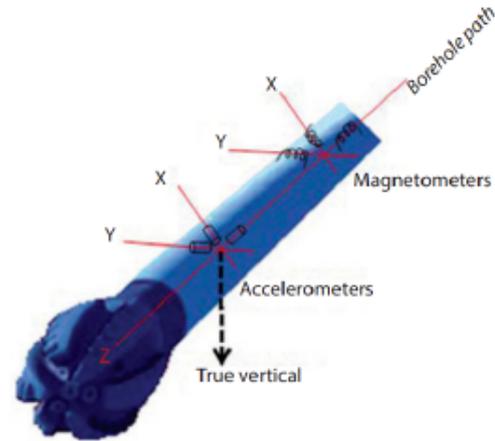


FIGURE I.13 – Disposition des capteurs directionnels dans les techniques MWD

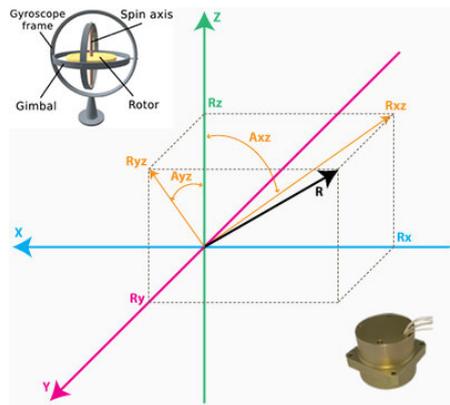


FIGURE I.14 – Principe du gyroscope

I.2.2.2 Méthodes de calcul

Il existe plusieurs méthodes connues de calcul des levés directionnels. Les cinq les plus couramment utilisées sont : tangentielle, tangentielle équilibrée, angle moyen, rayon de courbure et courbure minimale (la plus précise).

I.2.2.2.1 Tangentielle

Cette méthode utilise l'inclinaison et la direction du trou à l'extrémité inférieure de la longueur de parcours pour calculer une ligne droite représentant le fond du trou qui passe par l'extrémité inférieure de la longueur de parcours. Étant donné que le fond du trou est supposé être une ligne droite sur toute la longueur du parcours, c'est la méthode la plus imprécise parmi celles discutées et elle devrait être abandonnée complètement.

I.2.2.2.2 Tangentielle équilibrée

La modification de la méthode tangentielle consiste à prendre la direction de la station supérieure pour la première moitié de la longueur de parcours, puis celle de la station inférieure pour la deuxième

moitié, ce qui permet de réduire considérablement les erreurs de cette méthode. Cette modification est connue sous le nom de méthode tangentielle équilibrée. Cette méthode est très simple à programmer sur des calculatrices de poche et dans des feuilles de calcul, et elle donne une précision comparable à celle de la méthode de courbure minimale.

I.2.2.2.3 Angle moyen

La méthode utilise la moyenne des angles d'inclinaison et de direction mesurés aux extrémités supérieure et inférieure de la longueur de parcours. La moyenne des deux ensembles d'angles est supposée être l'inclinaison et la direction pour la longueur de parcours. Le trajet du puits est ensuite calculé à l'aide de fonctions trigonométriques simples.

I.2.2.2.4 Rayon de courbure

Avec l'inclinaison et la direction du trou mesurées aux extrémités supérieure et inférieure de la longueur de parcours, cette méthode génère un arc circulaire lorsqu'il est visualisé dans les plans vertical et horizontal. Le rayon de courbure est l'une des méthodes les plus précises disponibles.

I.2.2.2.5 Courbure minimale

Comme la méthode du rayon de courbure, cette méthode, la plus précise de toutes celles répertoriées, utilise l'inclinaison et la direction du trou mesurées aux extrémités supérieure et inférieure de la longueur de parcours pour générer un arc régulier représentant le trajet du puits. La différence entre les méthodes du rayon de courbure et de la courbure minimale est que le rayon de courbure utilise le changement d'inclinaison pour la longueur de parcours pour calculer le déplacement dans le plan horizontal (le TVD n'est pas affecté), tandis que la méthode de courbure minimale utilise le DLS pour calculer les déplacements dans les deux plans. La courbure minimale est considérée comme la méthode la plus précise, mais elle ne se prête pas facilement aux procédures de calcul normales à la main.

Les résultats des levés sont comparés à ceux de la méthode de courbure minimale, comme le montre le Tableau 6.2. On observe de grandes erreurs dans la méthode tangentielle pour seulement environ 1 900 pieds de déviation. Cela démontre que la méthode tangentielle est imprécise et devrait être abandonnée complètement. Les méthodes tangentielle équilibrée et angle moyen sont plus pratiques pour les calculs sur le terrain et doivent être utilisées lorsque des équipements de calcul sophistiqués ou une expertise ne sont pas disponibles. Ces méthodes doivent être indiquées comme "Résultats sur le terrain uniquement".

I.2.3 Sources d'erreurs dans le levé directionnel

Dans cette sous-section, nous donnerons un aperçu des différents facteurs et sources qui contribuent aux inexactitudes des mesures de forage directionnel, en mettant en lumière les défis et les limitations potentielles rencontrés lors du processus de levé[2].

I.2.3.1 Instruments d'arpentage

Les performances de l'instrument de levé dépendent de la conception du système, de l'étalonnage et du contrôle de qualité. Les performances sont influencées par l'inclinaison du trou de forage, l'azimut, le vecteur du champ magnétique terrestre et la position géographique. Les outils de compas gyroscopiques ont des performances dégradées à des latitudes et des inclinaisons plus élevées, notamment dans la direction est/ouest. Les outils magnétiques présentent des erreurs à des inclinaisons est/ouest élevées en raison du magnétisme de la colonne de forage.

Les gyroscopes sont également sujets à une incertitude de dérive liée au temps, qui peut être importante dans les puits horizontaux. Des durées de levé plus longues augmentent l'incertitude de dérive et l'exposition à l'environnement du trou de forage, réduisant potentiellement la précision des données directionnelles.

Les performances des gyroscopes se dégradent à environ 60°, et la plupart des gyroscopes ne peuvent pas effectuer de levé au-delà de 70° en raison de leur capacité réduite à tomber librement dans le puits.

I.2.3.2 Désalignement de l'outil

Le mauvais alignement de l'instrument de levé avec le trou de forage entraîne des erreurs de mesure de la direction et de l'inclinaison de l'axe du puits. (Note : l'inclinaison et l'azimut sont affectés.) Les sources de ce type d'erreur sont détaillées dans le Tableau 6.3. L'erreur d'instrument à capteur est indépendante de l'inclinaison, qui est une variable importante à la fois pour l'instrument à colonne de forage/ tubage et pour la colonne de forage par rapport au trou de forage. Les mauvais alignements ont longtemps été reconnus comme des sources d'erreur importantes dans les levés directionnels.

I.2.3.3 Erreur de profondeur (MD)

Les sources d'erreur de profondeur dépendent du type de système de levé utilisé. Les outils portés par tiges de forage (MWD, multi-points et mono-point) souffrent d'erreurs dans la mesure physique des tiges de forage et des effets différentiels de compression et d'étirement de la colonne de forage. En raison du frottement dans le trou de forage, la compression et l'étirement de la colonne de forage ne sont pas facilement calculables, en particulier dans les puits inclinés. Les erreurs de profondeur peuvent expliquer les erreurs angulaires relativement importantes souvent observées lors de la comparaison de levés de chevauchement à haute précision dans les puits déviés.

Les outils de levé sur câble ont généralement de plus petites erreurs de profondeur que les outils portés par colonne de forage, à condition que des mesures de contrôle de qualité adéquates aient été prises. On cite couramment des erreurs de l'ordre de 1/1 000 pour les outils gyroscopiques et de 2/1 000 pour les outils portés par colonne de forage. Cependant, cela peut ne pas s'appliquer aux situations de puits horizontaux.

I.2.3.4 Interférence magnétique

L'interférence magnétique peut être définie comme la corruption du champ magnétique terrestre par un champ provenant d'une source externe. Cela peut entraîner des erreurs importantes dans la mesure de la direction du trou de forage (azimut). Les sources potentielles d'interférence magnétique sont les suivantes :

- Colonnes de forage.
- Puits adjacents.
- Chaussures de tubage.
- Formations magnétiques.
- "Points chauds" dans les colliers de forage non magnétiques.

L'interférence de la colonne de forage est une cause fréquente d'erreurs dans les levés magnétiques. Les colliers de forage non magnétiques minimisent les erreurs de mesure de l'azimut. La force de l'interférence magnétique dépend de divers facteurs, notamment du champ d'interférence, de l'inclinaison du trou de forage et de la latitude géographique.

Les techniques de correction de l'interférence magnétique consistent à remplacer les mesures corrompues des capteurs par des valeurs calculées à partir de modèles géomagnétiques. Cependant, des modèles imparfaits peuvent introduire des erreurs dans le calcul de l'azimut. Des données d'azimut

précises sont possibles avec des informations fiables sur le champ géomagnétique, approchant la qualité des gyroscopes[28].

I.2.3.4.1 Physique spatiale et levés magnétiques

Les particules solaires chargées affectent les mesures magnétiques dans le forage directionnel, provoquant des fluctuations détectées par les outils de levé. Les perturbations géomagnétiques corrompent les mesures, notamment dans les zones aurorales. Les prévisions météorologiques spatiales sont précieuses, en se concentrant sur les zones sub-aurorales et aurorales. La proximité des observatoires ou des stations de magnétomètres permet de réaliser les corrections nécessaires. Les ingénieurs des plates-formes surveillent les magnétomètres proches pour le contrôle de qualité lors du forage.

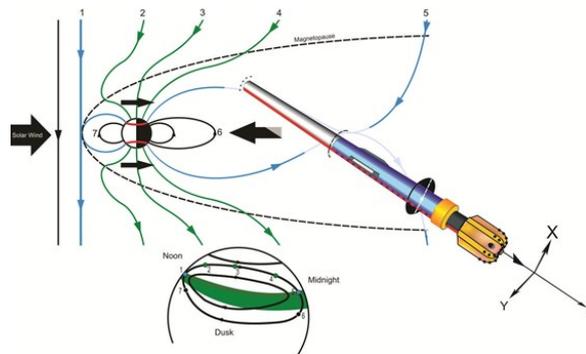


FIGURE I.15 – Physique spatiale et levés magnétiques

I.2.3.4.2 Le champ magnétique terrestre

Le champ magnétique terrestre s'étend du noyau à la magnétopause. On peut le décrire comme un aimant dipôle incliné, avec des contributions du noyau, des roches locales et des courants électriques dans la haute atmosphère et la magnétosphère. Le champ est mesuré dans des unités telles que le tesla (T), le gauss (G), le gamma (γ) ou l'Ørsted (Oe).

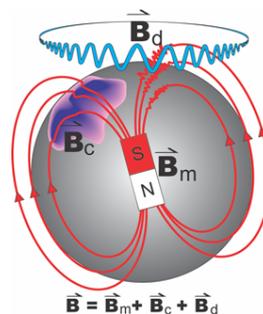


FIGURE I.16 – Les trois principales sources du champ magnétique terrestre.

Le champ principal (B_m)

Le champ magnétique principal de la Terre est principalement généré dans le noyau externe liquide. Il présente des variations temporelles lentes et est mieux observé grâce aux données des observatoires et des satellites.

Le champ crustal (B_c)

Le champ crustal est produit par les roches magnétisées de la croûte terrestre et du manteau supérieur. Il présente de grands gradients et de courtes longueurs d'onde par rapport au champ principal. Des

levés à espacement réduit sont nécessaires pour étudier le champ crustal, et la méthode de déduction de ses composantes est connue sous le nom d'IFR.

Le champ de perturbation (Bd)

Le plasma solaire, en interaction avec la magnétosphère terrestre, crée des courants électriques et un champ de perturbation. Le champ de perturbation est façonné par le vent solaire, entraînant des fluctuations dans le champ géomagnétique. Les systèmes de courants, tels que le courant annulaire et les courants alignés sur le champ, jouent un rôle important dans ces perturbations.

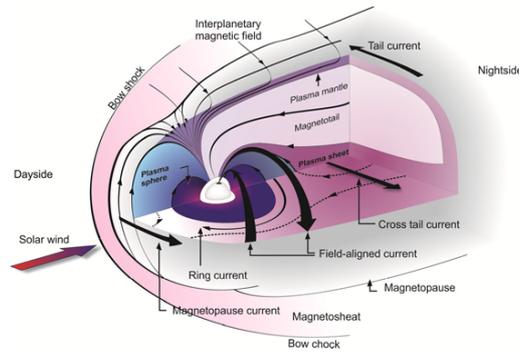


FIGURE I.17 – Courants dans la magnétosphère externe.

Composantes du champ magnétique terrestre

Le champ magnétique terrestre est décrit par des éléments tels que la déclinaison (D), l'inclinaison (I) et l'intensité magnétique (F). Ces éléments peuvent être représentés dans des coordonnées cartésiennes ou cylindriques.

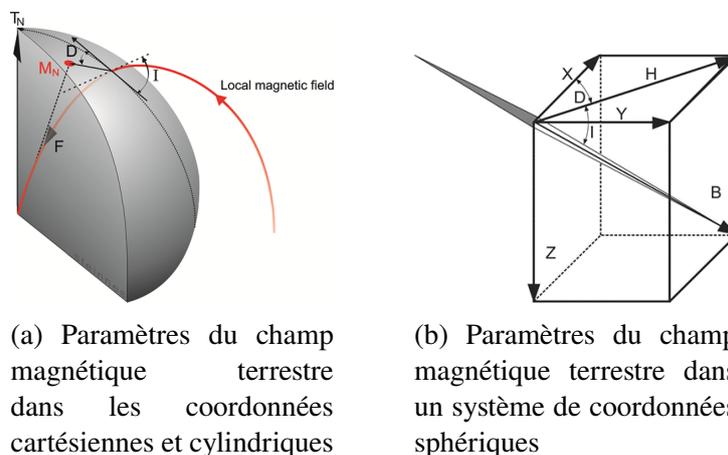


FIGURE I.18 – Comparaison des paramètres du champ magnétique terrestre dans différents systèmes de coordonnées

Coordonnées géomagnétiques et temps

Différents systèmes de coordonnées sont utilisés pour analyser le champ géomagnétique. Les coordonnées du dipôle incliné et les coordonnées géomagnétiques corrigées (CGM) offrent des modèles simplifiés, tandis que le temps magnétique local (MLT) fournit une référence pour étudier les variations temporelles.

Modèles du champ magnétique terrestre

Il existe plusieurs modèles mathématiques pour calculer le champ géomagnétique terrestre. Le Modèle de Référence Géomagnétique International (IGRF) se concentre sur les parties internes de la Terre. Les modèles géomagnétiques mondiaux tels que le Modèle Géomagnétique Global du BGS (BGGM) ou le Modèle Géomagnétique Haute Définition (HDGM) sont utilisés dans les levés de puits de forage directionnels pour calculer les valeurs nominales du champ.

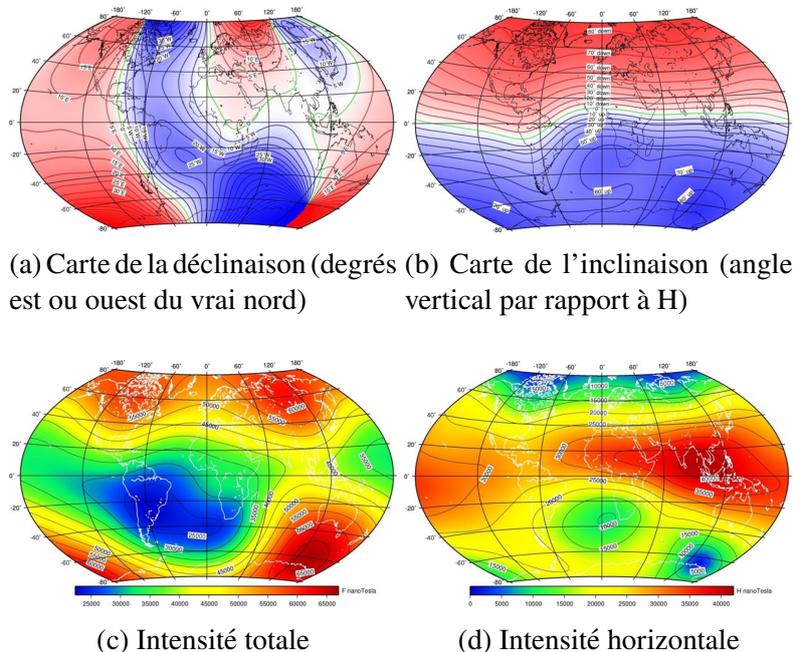


FIGURE I.19 – Cartes du champ géomagnétique en 2015.0

I.2.3.5 Erreur de position du trou de forage

La précision du positionnement du puits est cruciale dans l'industrie pétrolière et gazière. Les erreurs ont un impact sur la précision de la cible et la récupération des hydrocarbures. Les interférences magnétiques et les erreurs affectent les mesures directionnelles. L'évaluation des erreurs des instruments permet de déterminer l'incertitude de positionnement. L'ISCWSA établit des normes en matière de relevés.

Le positionnement dans les latitudes aurorales est confronté à des défis liés à la direction, à la dépendance géophysique et aux outils magnétiques. La gestion de l'incertitude est essentielle pour atteindre les objectifs et assurer la récupération. La précision géomagnétique affecte l'incertitude de trajectoire. Les recherches visent à améliorer la précision des relevés dans des régions telles que l'Arctique.

I.2.3.6 Contrôle de la qualité de l'enquête

Le contrôle de qualité est essentiel dans la réalisation des relevés directionnels en profondeur pour garantir la précision. Des méthodes alternatives, telles que la comparaison des relevés provenant de différentes sources, sont utilisées. Les géomètres suivent des procédures, effectuent des vérifications et vérifient la documentation pour assurer la qualité. Des modèles d'erreur et des tests quantifient l'incertitude et valident la qualité des relevés. La comparaison des mesures avec des modèles mondiaux accroît la confiance. Un contrôle efficace utilise plusieurs sources, procédures et tests pour réduire les risques.

I.2.3.7 Conception BHA pour le contrôle directionnel

I.2.3.7.1 Principes de conception pour la Force latérale et l'inclinaison du trépan

L'ensemble BHA est une partie du train de tiges de forage qui affecte la trajectoire de la fraise et, par conséquent, du puits de forage. En général, les facteurs qui déterminent la tendance au forage d'un ensemble BHA sont la force latérale de la fraise, l'inclinaison de la fraise, l'hydraulique et l'inclinaison de la formation. L'objectif de la conception de l'ensemble BHA pour le contrôle directionnel est de fournir la tendance directionnelle qui correspondra à la trajectoire planifiée du puits.

La force latérale de la fraise est le facteur le plus important qui affecte la tendance au forage. La direction et l'ampleur de la force latérale de la fraise déterminent les tendances de montée, de descente et de virage.

- Ensemble de descente : Force latérale de la fraise vers le côté bas du trou.
- Ensemble de montée : Force latérale de la fraise vers le côté haut du trou.
- Ensemble de maintien : Force d'inclinaison nulle au niveau de la fraise.

L'angle d'inclinaison de la fraise est l'angle entre l'axe de la fraise et l'axe du trou et affecte la direction de forage car une fraise est conçue pour forer parallèlement à son axe.

I.2.3.7.2 Assemblages rotatifs

Les ensembles rotatifs utilisés dans les opérations de forage peuvent être classés en ensembles de montée, de maintien et de descente, chacun ayant un objectif distinct avec des considérations spécifiques pour des performances optimales.

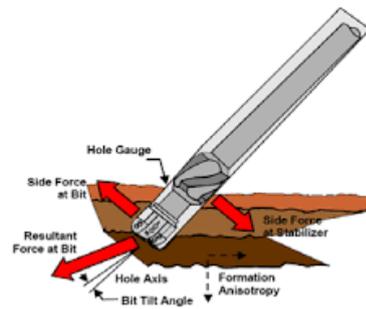
1. Assemblages de construction : Principe du pivot Augmentation de l'angle du puits. Les stabilisateurs près de la fraise agissent comme un pivot, forçant la fraise vers le côté haut. On utilise 2 à 3 stabilisateurs, le premier étant proche de la fraise. Une distance accrue améliore le taux de montée, mais provoque l'affaissement du train de tiges de forage. L'inclinaison, la charge sur l'outil de forage et la vitesse de rotation ont un impact sur le taux de montée.
2. Ensembles de maintien : trou rempli Maintien de l'angle du puits avec plusieurs stabilisateurs. Ils empêchent la flexion du train de tiges de forage, assurant une trajectoire droite. Les ensembles de maintien comportent 3 à 5 stabilisateurs, contrent les tendances de la formation et maintiennent la stabilité de la trajectoire.
3. Ensembles de descente : Principe du pendule Réduction de l'angle du puits en enlevant le stabilisateur situé au-dessus de la fraise. La gravité tire la fraise et le train de tiges de forage vers le côté bas. Les ensembles de pendule comportent deux stabilisateurs. La distance entre la fraise et le premier stabilisateur détermine le taux de descente. Une charge sur l'outil de forage contrôlée empêche une flexion excessive.

I.2.3.7.3 Outils de déviation

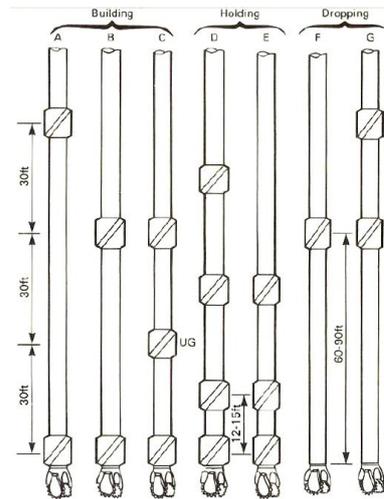
Dans ce guide complet, nous examinerons en détail chacun des outils de déviation essentiels pour le forage directionnel, en fournissant un bref aperçu de leurs fonctionnalités et applications. Ces outils comprennent les ensembles moteurs orientables (PDM) et les systèmes RSS. Les stabilisateurs à diamètre réglable, connus sous le nom de "systèmes rotatifs 2D", sont populaires pour le contrôle de l'inclinaison. Les cuillères de déviation, les turbines et les fraises à jet sont également utilisées.

Ensembles moteurs orientables (PDM)

Les ensembles moteurs orientables, tels que le moteur à boue à boîtier plié, sont des outils polyvalents utilisés dans les puits directionnels. Ils permettent un contrôle efficace de l'inclinaison, ce qui se



(a) Force latérale de la fraise et inclinaison



(b) Ensembles rotatifs

FIGURE I.20 – Illustration des ensembles rotatifs

traduit par des taux de montée plus élevés et une réduction des contraintes de flexion. Ces systèmes sont conçus en tenant compte de facteurs tels que le taux de montée et les contraintes de flexion. Le forage latéral à partir de l'encuvement implique le fraisage d'une section et le changement de direction à l'aide d'un moteur à boîtier plié, ce qui permet un forage efficace de sections inclinées[29].

- Forage à rayon moyen : Tailles de trou de 12¼ pouces et inférieures, utilisant différentes configurations de moteur, y compris un moteur à boîtier plié simple, une plaque décalée, un double pliage, un moteur à boîtier plié avec sous-plier et des moteurs à double boîtier plié.
- Forage à rayon intermédiaire : Taux de montée de 15 à 65°/100 pi avec des moteurs à paliers courts. Pour des taux de montée plus élevés (>45°/100 pi), des moteurs articulés et des outils MWD flexibles sont recommandés.
- Forage à faible rayon : Taux de montée de 65 à 125°/100 pi, en utilisant des moteurs articulés pour la section de montée et des moteurs latéraux hybrides pour la section latérale horizontale.

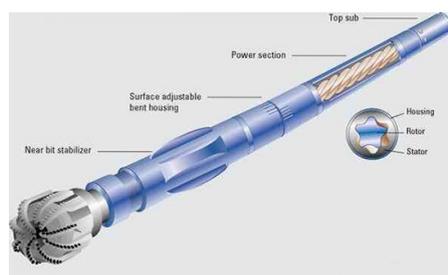


FIGURE I.21 – Illustration d'un ensemble moteur orientable (PDM)

RSS

Les systèmes rotatifs orientables (RSS) permettent une rotation continue du train de tiges de forage tout en dirigeant la fraise, offrant des taux de pénétration et une qualité du trou supérieurs par rapport aux ensembles moteurs orientables classiques. Les RSS utilisent deux concepts de direction : diriger la fraise et pousser la fraise. Ils sont plus coûteux mais sont utilisés pour les puits de forage à distance étendue exigeants.

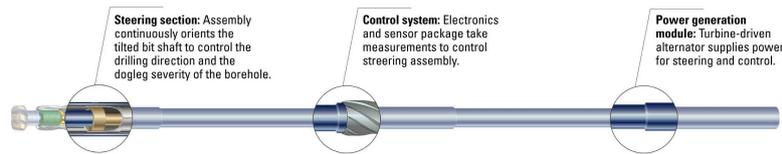


FIGURE I.22 – Modules du système rotatif orientable (RSS) par SLB

Stabilisateurs à jauge réglable (AGS)

Les stabilisateurs à diamètre réglable sont utilisés pour contrôler l'inclinaison en changeant le diamètre extérieur du stabilisateur sans retirer le train de tiges de forage. Ils peuvent être utilisés avec des systèmes moteurs rotatifs et orientables, permettant le contrôle de l'inclinaison lors du forage en mode rotatif.

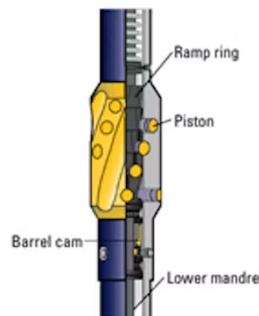


FIGURE I.23 – Stabilisateur à diamètre réglable

Whipstocks

Les whipstocks sont des outils utilisés pour modifier la trajectoire du puits de forage. Les whipstocks à trou ouvert sont moins utilisés aujourd'hui, remplacés par les moteurs à boîtier coudé. Les whipstocks de tubage sont couramment utilisés. Certains whipstocks sont récupérables, permettant de forer plusieurs latéraux à partir d'un seul puits. Des systèmes récemment développés permettent de gagner du temps en accomplissant toutes les étapes en un seul voyage. Le relevé gyroscopique est requis pour régler le whipstock et déterminer les orientations initiales de l'outil.

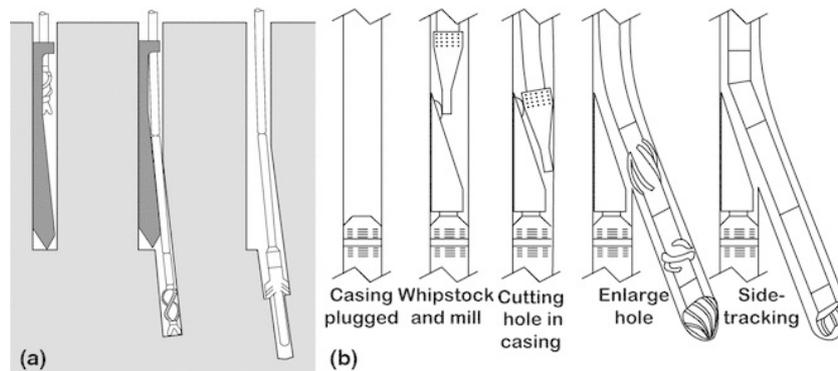


FIGURE I.24 – Opérations de cuillère de forage

Turbines et Jet Bits

Les turbines, ou turboperceuses, entraînées par des moteurs à turbine, conviennent aux opérations à grande vitesse et faible couple et sont utilisées avec des forets en diamant ou en diamant compact polycristallin. Les forage à jetting utilisent l'énergie hydraulique du fluide de forage pour éroder une cavité au fond du trou de forage et modifier sa trajectoire[30].

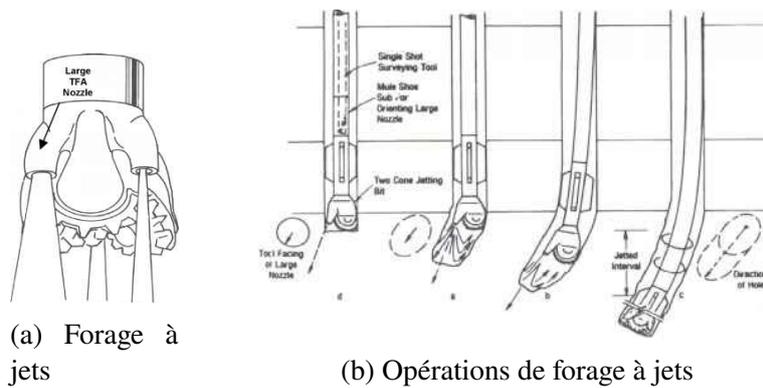


FIGURE I.25 – Déviation du puits de forage par forage à jets

I.2.4 Causes de la déviation du trou

La déviation du trou est l'écart involontaire du foret d'une trajectoire de forage présélectionnée. Qu'il s'agisse de percer une section droite ou incurvée, la tendance du trépan à s'éloigner de la trajectoire souhaitée peut entraîner des problèmes de forage tels que des coûts de forage plus élevés ainsi que des problèmes juridiques liés aux limites du bail. La figure I.26 fournit des exemples de déviations de trou.

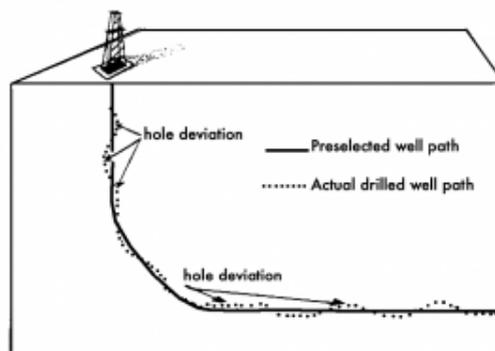


FIGURE I.26 – Exemple de déviations de trou (PétroWiki)

On ne sait pas exactement ce qui fait dévier un foret de sa trajectoire prévue. Il est cependant généralement admis qu'un ou une combinaison de plusieurs des facteurs suivants peut être responsable de l'écart :

- Nature hétérogène de la formation et de l'angle de pendage.
- Caractéristiques du train de tiges, en particulier la composition de l'assemblage de fond de trou (BHA).
- Stabilisateurs (emplacement, nombre et dégagements).
- Poids appliqué sur le bit (WOB).
- Angle d'inclinaison du trou par rapport à la verticale.
- Type de foret et sa conception mécanique de base.
- L'hydraulique à la loupe.
- Mauvais nettoyage des trous.

Il est connu qu'une certaine force résultante agissant sur un foret provoque une déviation du trou. La mécanique de cette force résultante est complexe et est régie principalement par la mécanique du BHA, l'interaction roche/trépan, les conditions de fonctionnement du trépan et, dans une moindre mesure, par l'hydraulique des fluides de forage .

Les forces transmises au trépan à cause du BHA sont directement liées à la composition du BHA, c'est-à-dire :

- Raideur
- Stabilisateurs
- Alésoirs

Le BHA est un élément structurel flexible et élastique qui peut se déformer sous des charges de compression. La forme bouclée d'un BHA conçu dépend de la quantité de WOB appliquée. L'importance du flambage BHA est qu'il provoque un désalignement de l'axe du foret avec l'axe de la trajectoire du trou prévu, provoquant ainsi la déviation. La rigidité et la longueur de la conduite et le nombre de stabilisateurs (leur emplacement et leurs dégagements par rapport à la paroi du puits de forage) sont deux paramètres majeurs qui régissent le comportement de flambage du BHA. Les actions qui peuvent minimiser la tendance au flambage du BHA comprennent la réduction du WOB et l'utilisation de stabilisateurs avec des diamètres extérieurs qui sont presque en jauge avec la paroi du trou de forage.

La contribution de l'interaction roche/trépan aux forces de déviation du trépan est régie par :

- Propriétés de la roche.
- Force de cohésion.
- Angle de litage ou de pendage.
 - Angle de frottement interne.
- Caractéristiques de conception du foret.
- Angle de dent.
- Taille du bit.
 - Type de bit.
 - Décalage de l'embout en cas d'embouts à rouleaux coniques.
 - Emplacement et nombre de dents.
 - Profil de bits.
 - Caractéristiques hydrauliques du foret.
- Paramètres de forage.
- Pénétration des dents dans la roche et son mécanisme de coupe.

La mécanique de l'interaction roche/trépan est un sujet très complexe et est le moins bien compris en ce qui concerne les problèmes de déviation des trous. Heureusement, l'avènement d'outils de mesure en cours de forage en fond de trou qui permettent de surveiller l'avancement du trépan le long de la trajectoire souhaitée rend plus acceptable notre manque de compréhension de la mécanique de la déviation du trou.

I.2.5 Géopositionnement d'un puits en temps réel "Geosteering"

La géosteering est le forage d'un puits dévié dans lequel des ajustements de trajectoire sont effectués en temps réel en fonction des données géologiques et de réservoir. Elle diffère du forage directionnel conventionnel en permettant des ajustements en cours de forage. La géosteering est nécessaire dans les cas où le repère géologique est incertain, où des tolérances étroites sont requises ou lorsque la géologie est trop complexe pour les méthodes conventionnelles.

Les données de diagrapie pendant le forage (LWD), comprenant les outils de rayonnement gamma, de résistivité et de densité, jouent un rôle crucial en géosteering. Ces mesures de diagrapie en temps réel aident à déterminer le positionnement optimal du puits horizontal et permettent des ajustements pendant le forage. La géosteering est particulièrement utile lors du forage dans des formations minces ou à proximité de limites géologiques[31].

I.2.6 Cartographie pendant le forage

L'une des technologies les plus récentes de SLB dans ce domaine est le service de cartographie de réservoir 3D **GeoSphere 360**, réalisé pendant le forage, qui offre une cartographie avancée du réservoir en trois dimensions. Il fournit des informations détaillées sur les volumes de fluide, les corps et les failles à l'échelle du réservoir, apportant un nouveau niveau de compréhension et d'évaluation de l'incertitude. Cette technologie permet une estimation plus précise des réserves et une augmentation de la production d'hydrocarbures par puits. En optimisant le positionnement des puits et en réduisant leur nombre, les opérateurs peuvent atteindre leurs objectifs de production tout en émettant moins de CO₂. Il est cependant important de noter que l'outil de mesure est généralement positionné à 9 pieds du trépan et ne mesure pas devant le trépan[32].

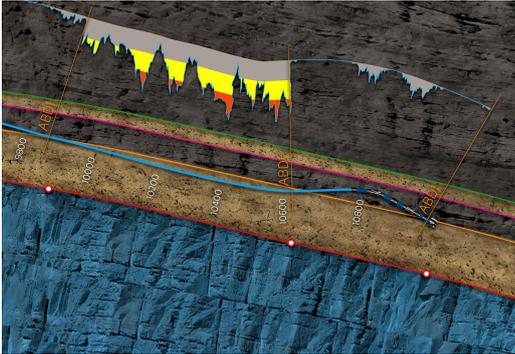


FIGURE I.27 – Illustration de la géosteering

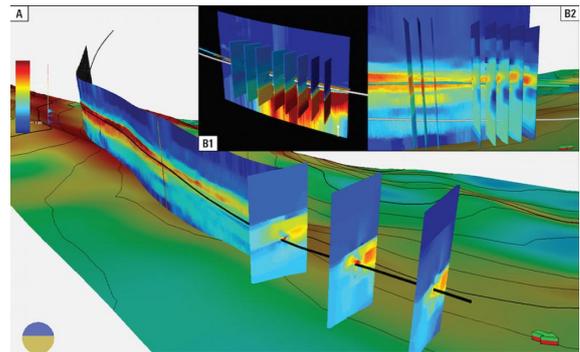


FIGURE I.28 – Cartographie pendant le forage

Maintenant que nous avons exploré les techniques de forage directionnel, concentrons-nous sur l'identification de la lithologie et les méthodes conventionnelles utilisées dans ce processus.

I.3 Méthodes conventionnelles d'identification de la lithologie

I.3.1 Introduction

L'identification de la lithologie, qui fait référence à la composition et au type de roche, est essentielle pour la caractérisation du réservoir. Elle joue un rôle vital dans la détermination des propriétés physiques et chimiques de la roche qui contient des hydrocarbures et/ou de l'eau, influençant la réponse des outils de mesure. Des calculs pétrophysiques précis de la porosité, de la saturation en eau (S_w) et de la perméabilité reposent sur l'identification des différentes lithologies présentes dans l'intervalle du réservoir et sur la compréhension de leurs implications. La lithologie englobe des détails tels que l'histoire de la sédimentation et de la diagenèse, la structure des pores et la minéralogie, dépassant les classifications de roche familières telles que le grès ou le calcaire.

Alors que la lithologie se concentre sur les grains, le type de roche fournit une description plus détaillée en considérant les regroupements naturels de systèmes de pores qui dictent les propriétés d'écoulement, les volumes et les saturations de fluide. Avec plus de 250 classifications de types de roche, l'identification de la lithologie et du type de roche est cruciale pour prédire et interpréter le comportement du réservoir. Dans cette section, nous explorerons les méthodes conventionnelles utilisées pour l'identification de la lithologie, en examinant les techniques et les outils utilisés pour démêler les complexités des lithologies du réservoir[33].

I.3.2 Détermination directe de la lithologie

I.3.2.1 Diagraphie de Boue

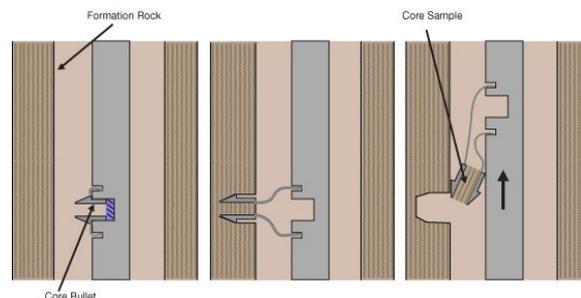
Les journaux de boue sont couramment utilisés comme méthode initiale d'identification de la lithologie, en particulier dans les puits d'exploration. Bien qu'ils fournissent des informations précieuses, l'attribution des fragments de roche à des profondeurs spécifiques peut introduire une certaine incertitude. La taille des échantillons de roche individuels obtenus à la surface est limitée par les déblais de forage et la résistance de la roche. Les roches plus faibles, sans ciment, peuvent être réduites à leur taille de grain détritique d'origine pendant le forage, ce qui rend difficile la détermination du type de roche. Cependant, les journaux de boue offrent toujours des informations sur les caractéristiques de la lithologie.

I.3.2.2 Prélèvements de carottes

Une fois qu'un puits est foré et enregistré, des études supplémentaires peuvent impliquer l'obtention d'échantillons physiques de roche. Des échantillonneurs de carottes au fil et des outils de prélèvement de carottes latérales sont utilisés pour récupérer des échantillons des couches de roche identifiées. Cependant, il n'est pas toujours garanti de récupérer des échantillons de carottes latérales, et il peut y avoir des cas où des échantillons de roche ne peuvent pas être obtenus dans les zones d'intérêt. Dans les puits d'exploration avec des zones hautement prometteuses identifiées par des enregistrements et des échantillons de fluides, une solution potentielle consiste à boucher le puits, à placer un whipstock, à effectuer un contournement parallèle (trou de contournement) et à obtenir une carotte complète à travers la zone d'intérêt. La carotte obtenue peut être échantillonnée et soumise à une analyse sophistiquée pour déterminer le type de roche et affiner l'évaluation de la formation basée sur les données d'enregistrement.



(a) Illustration d'une opération de carottage conventionnel



(b) Prélèvement de carottes latérales

FIGURE I.29 – Différentes méthodes d'échantillonnage de carottes

I.3.3 Détermination indirecte de la lithologie

Si aucun échantillon de roche direct n'est disponible dans une zone donnée, les réponses des enregistrements doivent être utilisées pour déterminer indirectement la lithologie. Les enregistrements géophysiques fournissent des informations sur les propriétés physiques de la roche, qui sont ensuite utilisées pour estimer la lithologie. Les méthodes indirectes couramment utilisées comprennent :

I.3.4 Détermination indirecte de la lithologie

Si aucun échantillon de roche direct n'est disponible dans une zone donnée, les réponses des enregistrements doivent être utilisées seules pour déterminer simultanément la lithologie, la porosité et les saturations en fluides. La classification des roches est beaucoup plus complexe et nécessite l'utilisation

de catalogues commerciaux de données analogiques [34] ou de données collectées localement afin d'identifier les propriétés pétrophysiques pouvant être utilisées pour affiner les calculs de porosité et de saturation en fluides [12]. Les méthodes indirectes couramment utilisées comprennent :

I.3.4.1 Les réponses diagraphiques

Lorsque des échantillons de roche directe ne sont pas disponibles, la lithologie, la porosité et les saturations en fluides doivent être déterminées uniquement à partir des réponses des enregistrements. Dans cette approche, des catalogues commercialement disponibles de données analogiques ou des données collectées localement sont utilisés pour identifier les propriétés pétrophysiques pouvant affiner les calculs de porosité et de saturation en fluides. Les enregistrements filaires tels que la densité, les neutrons et les outils acoustiques sont couramment utilisés pour déterminer la lithologie en fonction des caractéristiques générales. Les entreprises de services filaires fournissent des graphiques et des produits pour convertir les données d'enregistrement en lithologie. Cependant, les formations réservoirs complexes avec des propriétés variables posent des défis, et l'épaisseur de résolution de l'outil d'enregistrement peut affecter la réponse observée, en particulier en présence de bouchons de sel ou de types de roche argileuse.

I.3.4.2 Spectroscopie des rayons gamma (GR)

L'abondance de mica ou de feldspath potassique dans les grès peut compliquer la détermination de la lithologie basée sur la réponse des rayons gamma (GR). Les outils de spectroscopie avancée des rayons gamma peuvent aider à identifier les espèces chimiques présentes dans une roche, ce qui facilite la déduction de la lithologie. Cependant, la distribution de la taille des pores et la distribution de la taille des étranglements des pores, qui sont cruciales pour établir le type de roche et les caractéristiques d'écoulement des fluides, ne peuvent pas être déterminées par un seul enregistrement filaire ou un ensemble d'enregistrements.

I.3.4.3 Détermination de la salinité de l'eau

La quantification de la saturation en fluides, en particulier dans les réservoirs remplis d'hydrocarbures où la résistivité de l'eau de formation (R_w) est inconnue, présente des défis. Dans les zones difficiles, il est nécessaire de forer un puits de suivi avec un carottage complet prélevé en haut de la colonne d'hydrocarbures en utilisant une boue à base d'huile afin de déterminer la salinité de l'eau in situ par le biais d'une analyse sophistiquée du carottage. En l'absence de carottage, il peut être nécessaire de tester une quantité significative d'eau. La détermination précise de la salinité de l'eau dans les réservoirs en déplétion de pression avec une production d'eau minimale peut être difficile si le puits est foré avec une boue à base d'huile, rendant l'utilisation des enregistrements de potentiels spontanés (SP) inefficace.

I.3.4.4 Utilisation des enregistrements acoustiques

La vitesse acoustique est principalement fonction de la matrice rocheuse et peut être utilisée pour identifier différentes lithologies et pour des corrélations stratigraphiques. Une variété de techniques de tracés croisés, utilisant uniquement des mesures acoustiques ou en combinaison avec d'autres enregistrements de porosité (neutrons et densité), ont été conçues pour aider à l'identification lithologique (Fig. x). En particulier, les techniques de tracé croisé M-N et d'identification minérale (MID) utilisent les trois enregistrements de porosité dans différentes combinaisons.[34][35][40] Avant la détermination de la lithologie, les mesures individuelles des enregistrements doivent être corrigées des influences de :

- L'effet du gaz

- La porosité secondaire
- Les mauvaises conditions du trou
- La présence d'argile

En général, le tracé croisé MID est plus sensible à la lithologie, au gaz et à la porosité secondaire et offre de meilleurs résultats que les tracés M-N. Les tracés croisés, utilisant une variété de mesures d'enregistrement ou des combinaisons de mesures, peuvent être utilisés pour résoudre des problèmes lithologiques spécifiques liés à la géologie locale ou régionale.

LITHOLOGY	BHC ACOUSTILOG ↓ INCREASES	COMPENSATED NEUTRON LOG ↓ INCREASES	COMPENSATED DENSILOG ↓ INCREASES
Shale	$\Delta t \approx 130 - 175 \mu\text{s/ft}$ variable (compaction)	ϕ reads high	$\rho = 2.3-2.7 \text{ g/cm}^3$ variable (density shale)
Sandstone	$\Delta t \approx 52.5 - 55.5 \mu\text{s/ft}$ variable (compaction)	$\phi \approx -4\%$	$\rho = 2.65 \text{ g/cm}^3$
Limestone (Reference)	$\Delta t = 47.5 \mu\text{s/ft}$	$\phi \approx 0\%$	$\rho = 2.71 \text{ g/cm}^3$
Dolomite	$\Delta t \approx 42.5 \mu\text{s/ft}$	$\phi \approx (6-8)\%$	$\rho = 2.83 - 2.87 \text{ g/cm}^3$
Anhydrite	$\Delta t \approx 50 \mu\text{s/ft}$	$\phi \approx -(1-2)\%$	$\rho = 2.98 \text{ g/cm}^3$
Gypsum	$\Delta t = 52 \mu\text{s/ft}$	$\phi \approx 48\%$	$\rho = 2.33 \text{ g/cm}^3$
Salt	$\Delta t \approx 67 \mu\text{s/ft}$	$\phi \approx 0\%$	$\rho = 2.08 \text{ g/cm}^3$
Gas	Δt reads high	ϕ reads low	ρ reads low

FIGURE I.30 – Comparaison des réponses des enregistrements dans les lithologies de réservoir

Le rapport entre la vitesse de compression et la vitesse de cisaillement, V_p/V_s , est un indicateur efficace de la lithologie car chaque lithologie présente une tendance définie indépendante de la porosité et de la profondeur (Fig. x).[36][37][38] Cependant, étant donné que le rapport V_p/V_s est affecté par l'anisotropie de la formation, les valeurs du rapport peuvent ne pas être des indicateurs absolus d'une lithologie particulière.

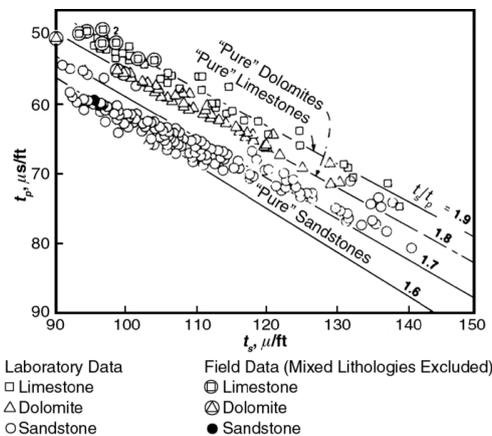


FIGURE I.31 – Diagramme croisé de sensibilité aux variations de lithologique : Δt_p vs. Δt_s

L'ajout de la lenteur de cisaillement à l'identification de la lithologie fournit un résultat plus robuste qui peut être particulièrement utile dans les évaluations en trou foré où les enregistrements de

densité ne sont pas disponibles. Sur la figure 3, l'utilisation combinée de la lenteur de cisaillement et de la porosité neutronique en trou foré permet une meilleure détermination de la porosité dans une lithologie complexe. Les tracés croisés du rapport V_p/V_s par rapport au temps de propagation de compression, Δt_c , facilitent l'identification des tendances lithologiques par rapport à la porosité et à la lithologie (Figure I.32[39]). Des études récentes sur des réservoirs carbonatés complexes indiquent que le rapport V_p/V_s est fonction de la porosité et peut, dans certains cas, différencier les faciès à perméabilité plus élevée.

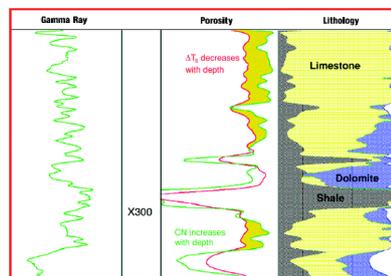


FIGURE I.32 – Diagramme croisé de V_p/V_s vs. Δt_p : Identification des intervalles de sable et de turbidites

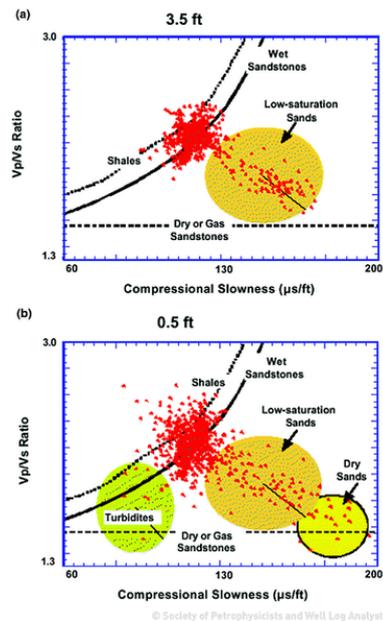


FIGURE I.33 – Indication de la lithologie par la lenteur de cisaillement

I.3.5 Caractérisation de la lithologie et de l'hétérogénéité du réservoir

Les géologues évaluent les roches à l'aide de diverses méthodes et collaborent avec les ingénieurs pour les évaluations de réservoir. Ils analysent les caractéristiques des courbes d'enregistrement pour déterminer les environnements de dépôt. Les géologues sur site utilisent les descriptions, les enregistrements et la reconnaissance de motifs pour comprendre la lithologie. Les puits de délimitation fournissent des informations détaillées sur les carottes, l'analyse des lames minces et d'autres mesures. La lithologie affecte les calculs pétrophysiques, y compris la taille des grains, le tri et la minéralogie. Les propriétés des minéraux argileux sont cruciales pour les propriétés électriques et la perméabilité. Les réservoirs carbonatés présentent des considérations minéralogiques particulières.

I.3.5.1 Propriétés des minéraux argileux

Les minéraux argileux varient considérablement en composition et en propriétés, ce qui affecte la perméabilité et la porosité. Les smectites, comme la montmorillonite, ont un impact significatif sur la porosité et les mesures de saturation en eau en raison des cations d'échange et de l'eau adsorbée.

Les formations de shale sont composées de minéraux argileux et d'autres grains de roche. Les minéraux argileux dans les grès varient de 0 à 10% en volume brut (BV), tandis que les shales contiennent environ 20 à 40% en volume brut de minéraux argileux. Les shales et les minéraux argileux authigènes retiennent de l'eau de formation supplémentaire dans leur système de pores.

I.3.5.2 Évaluation du volume de schiste

Différentes techniques, y compris l'analyse aux rayons X, aident à identifier les espèces de minéraux argileux et à estimer la teneur en schiste (Vsh). Les enregistrements de neutrons et de densité sont couramment utilisés pour évaluer Vsh. Les enregistrements de rayonnement gamma (GR) sont fréquemment utilisés pour déterminer la zone non productive.

L'estimation de Vsh à partir des enregistrements de puits est incertaine, avec une précision qui diminue pour des valeurs plus élevées de Vsh. Des techniques de normalisation sont utilisées pour atténuer les variations et simplifier les évaluations. Les sels de potassium présents dans les fluides de forage peuvent affecter l'interprétation des enregistrements de rayonnement gamma.

I.3.5.3 Zonage ou stratification des réservoirs

Le zonage du réservoir consiste à diviser le réservoir en sections verticales ou horizontales. Les géologues analysent des données telles que les descriptions de carottes et diverses mesures pour comprendre les différences entre ces sections. Ils utilisent des graphiques et des comparaisons pour identifier les variations des propriétés telles que la perméabilité, la porosité et la saturation en eau. Les grands réservoirs sont généralement divisés en zones verticales pour des calculs précis. La quantité et la qualité des données disponibles influencent le niveau de subdivision. Les figures fournies illustrent des exemples de zonage du réservoir et montrent comment les propriétés des roches peuvent varier dans un champ et d'une section de réservoir à une autre [41][42].

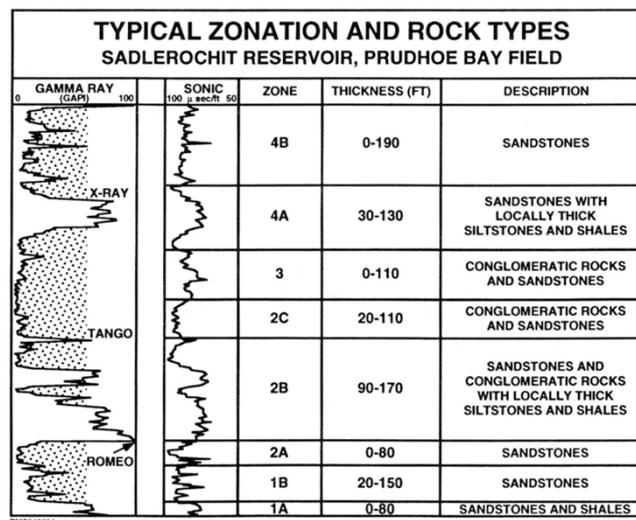
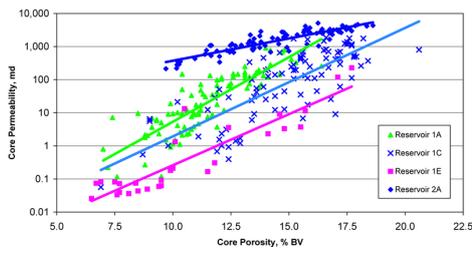
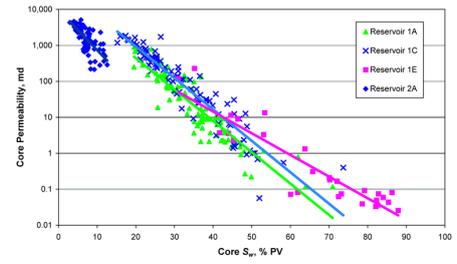


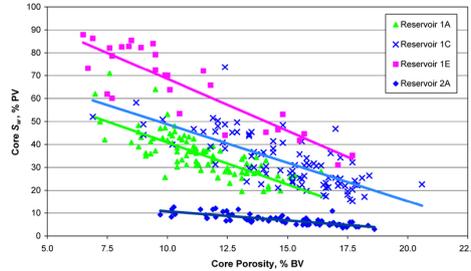
FIGURE I.34 – Zonation du réservoir du réservoir de Sadlerochit, champ de Prudhoe Bay, Alaska, États-Unis



(a) Perméabilité vs Porosité



(b) Perméabilité vs S_w



(c) S_w vs Porosité

FIGURE I.35 – Diagramme croisés de réservoir à partir d'un champ gazier asiatique

I.4 Conclusion

En conclusion, ce premier chapitre met en évidence l'importance cruciale du forage directionnel, du géopositionnement en temps réel des puits et de l'identification de la lithologie dans l'industrie pétrolière et gazière. Ces pratiques jouent un rôle essentiel dans l'optimisation des processus d'exploration et de production, permettant ainsi d'améliorer l'efficacité et la productivité globales.

Préparez-vous à plonger dans le prochain chapitre qui explore une nouvelle discipline passionnante en pleine révolution au sein de l'industrie pétrolière et gazière.

Chapitre II

La Révolution de l'Apprentissage Automatique dans l'Industrie Pétrolière

II.1 Introduction

L'apprentissage automatique, une sous-discipline de l'intelligence artificielle, révolutionne l'industrie pétrolière et gazière en extrayant des modèles précieux à partir des données et en permettant des pratiques plus sûres, rentables et efficaces. Il transforme les données en connaissances, s'appuyant sur des principes statistiques, mathématiques, informatiques et théoriques. Les algorithmes d'apprentissage automatique apprennent de l'expérience et effectuent des prédictions ou des décisions basées sur les données, grâce à des composants tels que les données, les modèles, les algorithmes d'apprentissage et les métriques d'évaluation. L'adaptabilité aux nouvelles données et la recherche d'un équilibre entre biais et variance sont cruciales pour construire des modèles d'apprentissage automatique robustes. Ce chapitre explore les applications transformatrices de l'apprentissage automatique dans l'industrie pétrolière et gazière, en mettant en évidence son potentiel pour améliorer les opérations, la prise de décision et l'efficacité.

En reconnaissant les tendances et les anomalies liées aux défaillances de l'équipement, à l'épuisement des réservoirs et aux inefficacités de forage, l'apprentissage automatique offre des informations précieuses pour atténuer les risques et améliorer l'efficacité opérationnelle dans l'industrie pétrolière et gazière. Il va au-delà d'un simple processus superficiel, nécessitant une compréhension approfondie des complexités et une quête continue d'une précision et d'une généralisation améliorées. Ce chapitre explore les applications de l'apprentissage automatique, mettant en évidence sa capacité à façonner l'avenir de l'industrie en exploitant la puissance des données et en prenant des décisions intelligentes basées sur les connaissances acquises.

II.2 IA, ML, DL et DM : Concepts clés

Les avancées rapides des technologies axées sur les données ont donné naissance à une multitude de concepts interdépendants. Il est crucial de les distinguer afin de comprendre leurs contributions et fonctionnalités uniques. Nous abordons quatre concepts clés dans cette section : Intelligence Artificielle (IA), Apprentissage Automatique (ML), Apprentissage Profond (DL) et Exploration de Données (DM).

II.2.1 Intelligence Artificielle (IA)

L'Intelligence Artificielle (IA) est la science de la création de machines intelligentes qui imitent les fonctions cognitives humaines telles que l'apprentissage, le raisonnement, la résolution de problèmes et l'autocorrection. Elle consiste à concevoir des systèmes informatiques capables d'accomplir des tâches qui nécessitent généralement l'intelligence humaine.

II.2.2 Exploration de Données (DM)

L'Exploration de Données est le processus de découverte de schémas exploitables dans de grands ensembles de données en utilisant des techniques d'apprentissage automatique, de statistiques et de systèmes de bases de données. Elle extrait des informations précieuses et jusqu'alors inconnues, permettant de prendre des décisions stratégiques importantes. Également connue sous le nom de Découverte de Connaissances dans les Bases de Données (KDD), l'Exploration de Données utilise l'apprentissage automatique et d'autres techniques pour analyser et explorer les ensembles de données, révélant des insights cachés.

II.2.3 Apprentissage Automatique (ML)

L'Apprentissage Automatique (ML) est une branche de l'IA qui permet aux ordinateurs d'apprendre à partir de données et de prendre des décisions ou des prédictions éclairées. En développant des

algorithmes, l'Apprentissage Automatique permet aux ordinateurs de détecter des motifs au sein des données, d'optimiser leurs performances et de fournir des informations précieuses aux organisations. Les algorithmes d'apprentissage automatique s'améliorent avec le temps grâce à l'analyse des données, sans nécessiter de programmation explicite. Ces connaissances cachées peuvent offrir une valeur opérationnelle significative dans divers domaines.

II.2.4 Apprentissage Profond (DL)

L'Apprentissage Profond est une sous-branche de l'Apprentissage Automatique (ML) qui utilise des réseaux de neurones artificiels (RNA) pour apprendre automatiquement des représentations complexes des données à travers plusieurs couches d'abstraction. Il s'agit d'un mécanisme d'apprentissage puissant et évolutif. Différentes architectures, telles que MLP, CNN, RNN et DNN, optimisent les réseaux de neurones profonds pour des problèmes spécifiques tout en restant applicables à d'autres. Cette polyvalence les rend idéaux pour aborder des tâches d'apprentissage automatique importantes et complexes. Comprendre les distinctions et la complémentarité de l'IA, du ML, du DL et du DM est essentiel pour exploiter leur potentiel dans les initiatives axées sur les données.

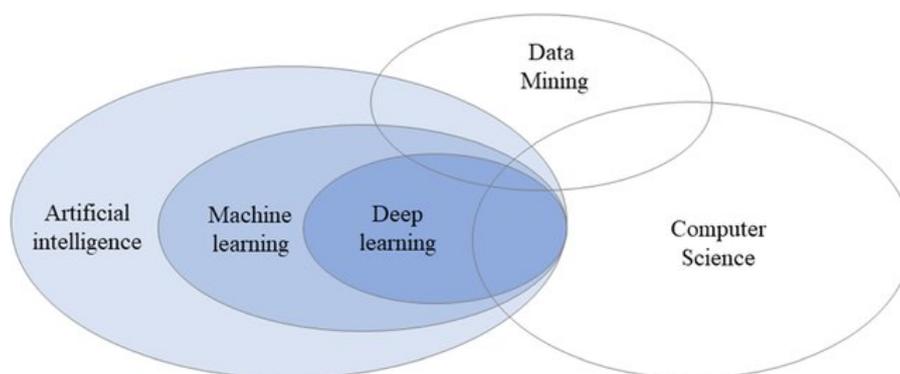


FIGURE II.1 – Diagramme de Venn représentant les relations entre l'IA, le ML et le DL.

II.3 Paradigmes de l'Apprentissage Automatique : Une Comparaison

Les systèmes d'apprentissage automatique sont multifacettes, avec différents types relevant de différentes catégories en fonction de leur fonctionnalité et de leurs stratégies d'apprentissage. Cette section vise à fournir une comparaison complète des différents paradigmes de l'apprentissage automatique. Ils peuvent être classés selon divers critères, notamment le niveau de supervision humaine impliqué, la nature incrémentielle du processus d'apprentissage et leur fonctionnement opérationnel impliquant des points de données et une modélisation prédictive.

II.3.1 Modes d'Apprentissage en IA

Les systèmes d'apprentissage automatique peuvent être classés en fonction du degré et de la nature de la supervision qu'ils reçoivent lors du processus d'entraînement.

II.3.1.1 Apprentissage Supervisé

L'apprentissage supervisé en apprentissage automatique implique que nous entraînons l'algorithme sur un ensemble de données étiquetées. "Étiqueté" signifie que chaque exemple dans l'ensemble de

données est constitué d'un vecteur de caractéristiques et d'une valeur de sortie souhaitée (l'étiquette). Les problèmes d'apprentissage supervisé sont ensuite catégorisés en deux types principaux :

Régression : Un type de problème où l'objectif est de prédire une variable cible continue. Certains des algorithmes les plus couramment utilisés pour les problèmes de régression incluent :

- Régression Linéaire
- Arbres de Décision
- Forêt Aléatoire
- Régression par Vecteurs de Support (SVR)

Classification : Un type de problème où l'objectif est de prédire une variable cible catégorique. Certains des algorithmes les plus couramment utilisés pour les problèmes de classification incluent :

- Régression Logistique
- k Plus Proches Voisins (k-NN)
- Arbres de Décision
- Forêt Aléatoire
- Naïve Bayes

Consultez la Figure II.2 pour visualiser la différence entre ces deux méthodes.

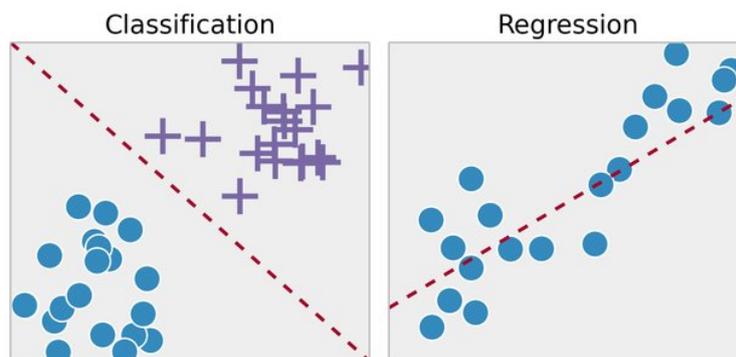


FIGURE II.2 – Classification vs Régression.

II.3.1.2 Apprentissage Non Supervisé

En apprentissage non supervisé, les données ne sont pas étiquetées. En d'autres termes, le système tente d'apprendre sans enseignant. Les principaux algorithmes d'apprentissage non supervisé incluent les algorithmes de regroupement (comme le K-Means et l'Analyse de Cluster Hiérarchique), les algorithmes de détection d'anomalies (comme le SVM à une classe et la Forêt d'Isolation) et les algorithmes

de réduction de la dimensionnalité (comme l'Analyse en Composantes Principales (ACP) et l'Ensemble de Voisins Stochastiques t-distribués (t-SNE)).

II.3.1.3 Apprentissage Semi-Supervisé

L'apprentissage semi-supervisé est une combinaison d'algorithmes d'apprentissage supervisé et non supervisé. Ces algorithmes fonctionnent bien lorsque de nombreuses données non étiquetées sont disponibles ainsi qu'une petite quantité de données étiquetées. Un exemple est le réseau de croyance profonde (DBN), qui est basé sur des composants non supervisés appelés Machines de Boltzmann Restreintes (RBM) empilées les unes sur les autres.

II.3.1.4 Apprentissage par Renforcement

L'apprentissage par renforcement est fondamentalement différent. Le système d'apprentissage, appelé "agent" dans ce contexte, observe l'environnement, choisit et effectue des actions, obtenant des récompenses ou des pénalités en retour. L'agent apprend de lui-même la stratégie optimale, appelée politique, pour obtenir la récompense maximale au fil du temps.

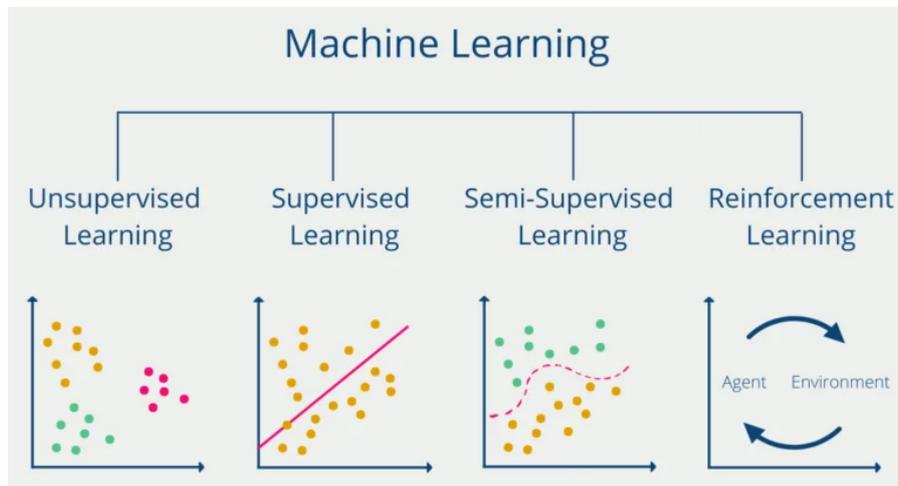


FIGURE II.3 – Aperçu des Catégories de l'Apprentissage Automatique.

II.4 Gestion des Données en Apprentissage Automatique

II.4.1 Définition du Problème et Objectifs

Un projet d'apprentissage automatique réussi passe par une série d'étapes pour garantir des résultats optimaux pour un problème bien défini. Ces étapes sont courantes dans un flux de travail d'apprentissage automatique, bien qu'elles ne soient pas exhaustives :

Formulation du problème. Identification des sources de données. Prétraitement et analyse exploratoire des données. Ajustement, évaluation et optimisation du modèle. Post-traitement et explicabilité du modèle. Déploiement, suivi et maintenance du système. Il est important de noter que les étapes suivies pour aborder différents types de projets peuvent varier en fonction des résultats finaux recherchés, de la nature et du type de données, et de l'algorithme utilisé.

II.4.1.1 Formulation du Problème

Lors de la traduction d'un défi en termes d'apprentissage automatique, des questions importantes à poser incluent :

- Le défi est-il réel, et quel impact a-t-il sur le processus ? La valeur de la solution peut-elle être quantifiée dans le cadre défini ?
- Définir l'urgence et l'importance aide à déterminer la nécessité de résoudre le problème.
- Comprendre l'objectif commercial est crucial. Pensez à la façon dont le modèle sera utilisé et aux avantages qu'il apportera. Cela influence la formulation du problème, la sélection de l'algorithme, l'évaluation des performances et les efforts d'ajustement.

Considérez également la disponibilité de l'expertise humaine, le contexte technique et les données relatives au sujet. Si les données sont insuffisantes, explorez les moyens de les acquérir. De plus, tenez compte des solutions existantes, des solutions de contournement et des problèmes comparables. La

réutilisation de cadres, d'expériences ou d'outils peut être bénéfique. Aborder ces questions accélère la progression du projet et offre une perspective claire sur le résultat final.

II.4.2 Acquisition et Sources de Données

L'examen des sources de données est une partie souvent négligée mais essentielle de l'ensemble du projet. C'est plus semblable à une boucle externe connectée entre différentes sections qu'à une étape distincte, obligeant l'exécutant (analyste de données, ingénieur, scientifique, etc.) à revenir en arrière et à avancer continuellement. Cela signifie que l'automatisation est cruciale à cet égard. Le processus comprend :

- Énumération du type, de la nature et du volume de données nécessaires.
- Identification et documentation des sources de données potentielles, ce qui peut être réalisé en examinant des articles académiques similaires à votre sujet et/ou en examinant des centres de données, entre autres méthodes.
- Évaluation de la puissance de calcul et du stockage nécessaires pour le traitement et la gestion des données.
- Vérification des obligations légales et obtention des autorisations nécessaires.
- Une fois obtenues, transformation des données dans un format pouvant être facilement manipulé, tout en préservant l'intégrité des données.
- Veiller à ce que les informations sensibles soient soit éliminées, soit protégées, par exemple, par anonymisation.

Comme mentionné précédemment, l'opérateur doit constamment aborder cet aspect spécifique, car les résultats initiaux ne satisfont que rarement l'objectif commercial plus large. Au fur et à mesure de l'avancement du projet, la nécessité d'acquérir des données supplémentaires ou différentes se fait souvent sentir, ce qui nécessite une approche dynamique de l'approvisionnement et de la gestion des données.

II.4.3 Audit et Prétraitement des Données

L'audit et le prétraitement des données constituent la phase la plus critique d'un projet d'apprentissage automatique. Des données de haute qualité conduisent inévitablement à de meilleures performances et prédictions plus fiables. Pendant cette étape, le praticien trie, nettoie et découvre des informations cruciales sur les corrélations des données en utilisant différentes méthodes d'analyse et visualisations. Voici les principales étapes impliquées, bien que la liste ne soit pas exhaustive :

- Prétraitement des données : Gestion des données manquantes, des doublons, des valeurs aberrantes et des variables catégorielles.
- Analyse et visualisation des données : Étude de la distribution, des corrélations et des relations à travers une analyse visuelle.
- Préparation du modèle : Sélection et ingénierie des caractéristiques, gestion des données incomplètes, réduction de la dimensionnalité, normalisation et séparation des données.

Bien que ces étapes fournissent des lignes directrices générales (voir Figure II.4), le flux de travail réel peut varier en fonction du problème à résoudre, comme les tâches de vision par ordinateur ou de traitement du langage naturel. Quoi qu'il en soit, toutes ces étapes se résument à accorder une attention particulière à la représentation des données de l'ensemble du problème et à satisfaire aux exigences du modèle utilisé.

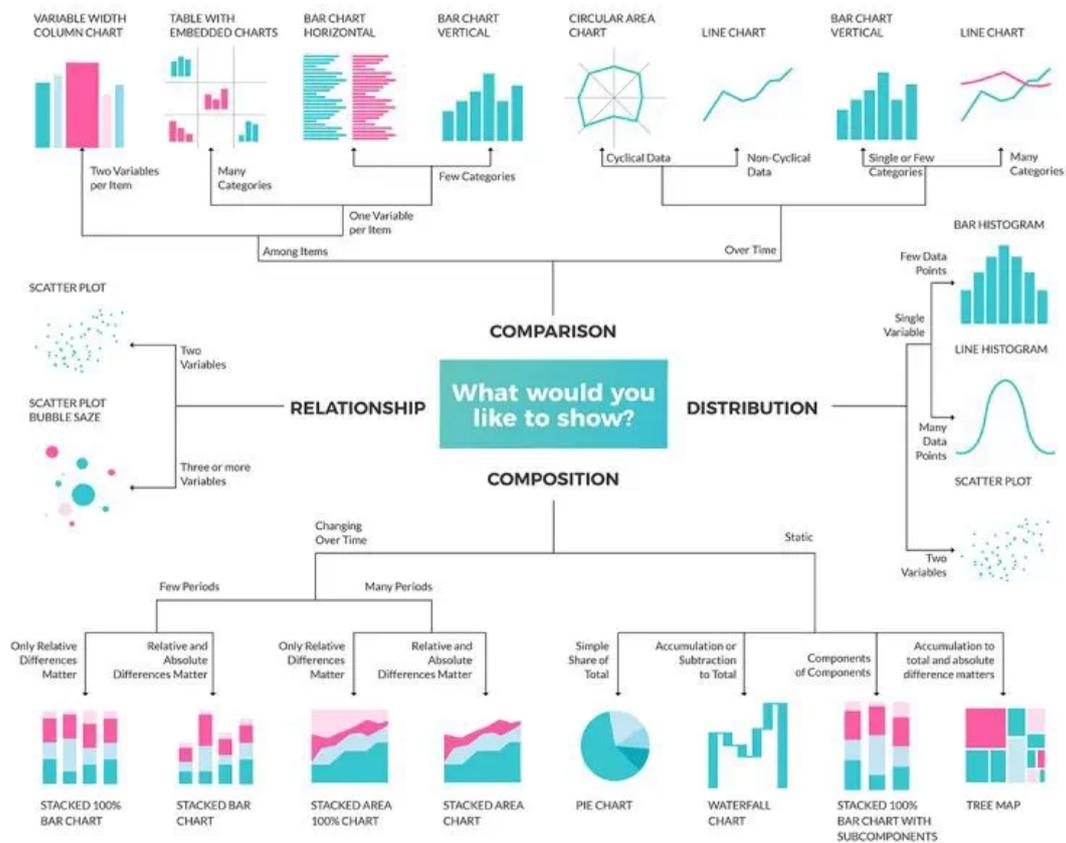


FIGURE II.4 – Visualisations courantes dans l'Analyse Exploratoire des Données, Visualisation efficace des données.

II.5 Construction et Évaluation des Modèles d'Apprentissage Automatique

Après avoir nettoyé, prétraité et transformé les données de manière appropriée, l'étape cruciale suivante consiste à construire et évaluer les modèles d'apprentissage automatique. Cette étape implique le choix d'algorithmes adaptés, la formation des modèles, l'ajustement des hyperparamètres, la validation des résultats et l'interprétation des modèles.

II.5.1 Sélection des Algorithmes d'Apprentissage Automatique

La sélection des algorithmes d'apprentissage automatique est principalement motivée par le problème à résoudre, la nature des données et les exigences de la solution. Ici, nous examinerons certains des algorithmes d'apprentissage automatique largement utilisés, leurs caractéristiques clés et les scénarios où ils conviennent le mieux.

II.5.1.1 Algorithmes basés sur la Régression

Régression Linéaire

La régression linéaire est un algorithme fondamental d'apprentissage automatique, généralement utilisé comme point de départ pour les problèmes de régression où l'on souhaite prédire une variable continue, telle que la prédiction des prix de l'immobilier, des prix des actions ou du taux de croissance d'une certaine bactérie. Il utilise la relation entre les points de données pour tracer une droite droite à

travers eux. Cette droite peut être utilisée pour prédire les valeurs futures. Il fonctionne mieux lorsque les prédicteurs ont une relation linéaire avec la variable cible. Il suppose également que les résidus (c'est-à-dire la différence entre la valeur observée et prédite) sont distribués de manière normale.

Régression Logistique

Malgré son nom, la régression logistique est un algorithme de classification qui modélise la probabilité que le système soit dans un état particulier. Cela est utile lorsque vous souhaitez que votre modèle produise des probabilités pour une certaine classe, ce qui est une exigence courante dans les problèmes de classification binaire tels que déterminer si un patient a une certaine maladie ou non, prédire si un client va se désabonner, etc. La régression logistique fonctionne également mieux lorsque les prédicteurs sont indépendants les uns des autres et lorsqu'ils sont linéairement liés aux logarithmes des cotes.

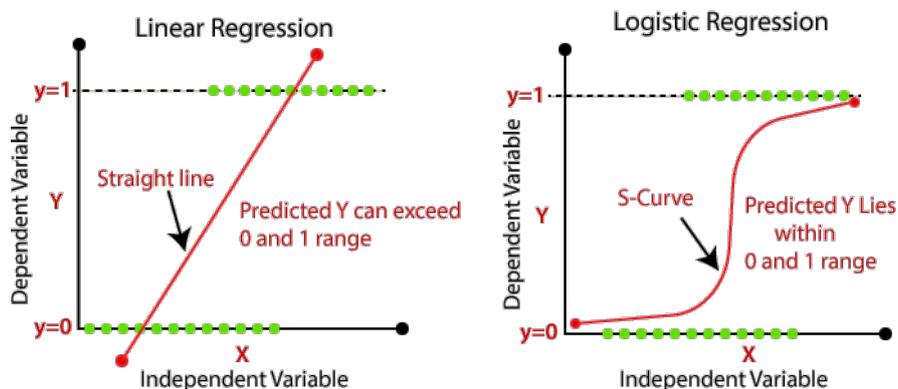


FIGURE II.5 – Régression linéaire vs régression logistique.

II.5.1.2 Algorithmes basés sur les Arbres

Les arbres de décision sont un type d'algorithme d'apprentissage supervisé principalement utilisé pour les problèmes de classification. Il s'agit d'un modèle de décision basé sur certaines conditions. C'est l'une des formes les plus simples et les plus utiles de prédiction.

La forêt aléatoire est une méthode d'apprentissage par ensemble qui construit une multitude d'arbres de décision (voir Figure II.6). La classe de sortie est le mode des classes produites par les arbres individuels.

Le boosting gradient est un autre type de méthode d'apprentissage par ensemble. La méthode du boosting gradient construit de nouveaux prédicteurs qui visent à corriger les erreurs résiduelles du prédicteur précédent, ce qui améliore les performances du modèle en réduisant les erreurs (voir Figure II.7).

II.5.1.3 Machines à Vecteurs de Support (SVM)

Les SVM sont particulièrement adaptées à la classification de jeux de données complexes mais de petite ou moyenne taille (voir Figure II.8).

II.5.1.4 Algorithmes basés sur la Distance

Le **k-plus proches voisins (KNN)** est un algorithme simple qui stocke tous les cas disponibles et classe les nouveaux cas selon un vote majoritaire de ses k voisins (voir Figure II.9).

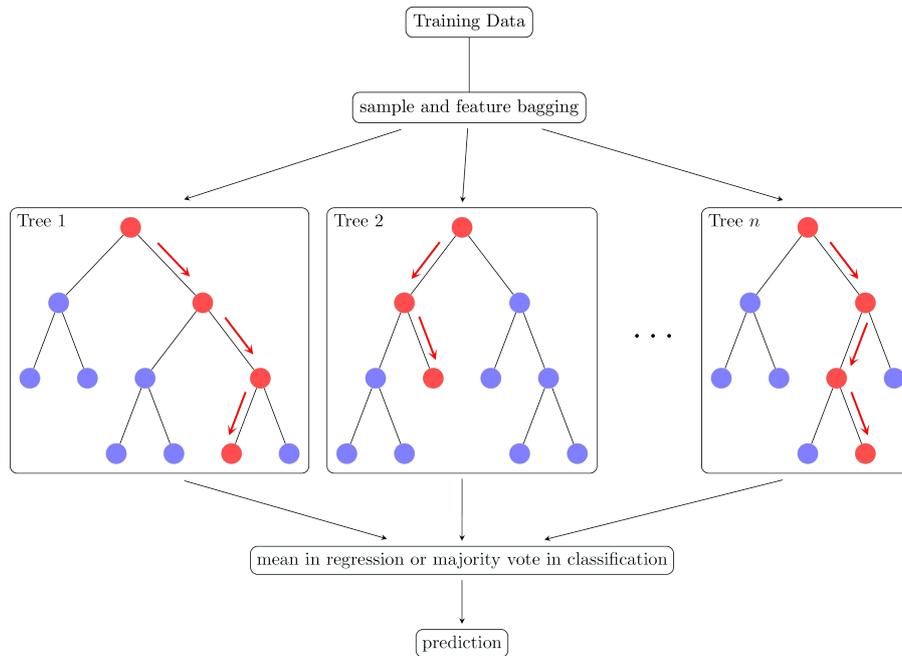


FIGURE II.6 – Illustration des arbres de la forêt aléatoire.

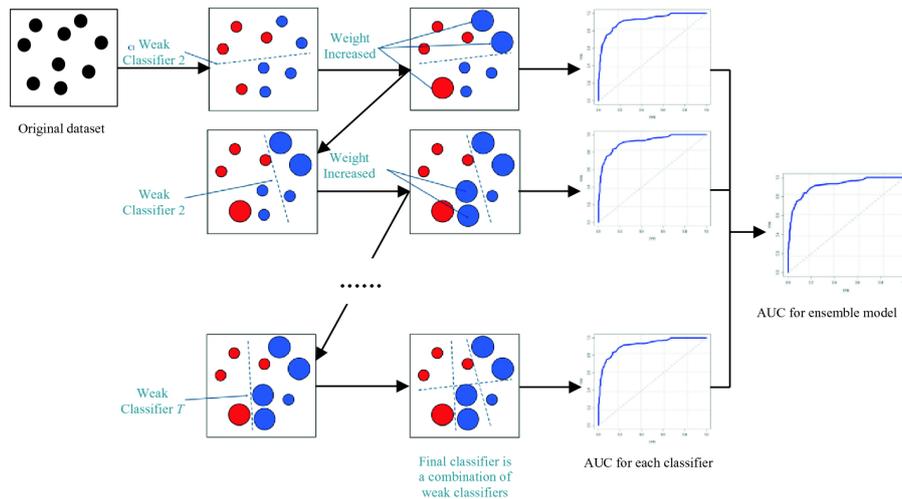


FIGURE II.7 – Illustration du boosting gradient.

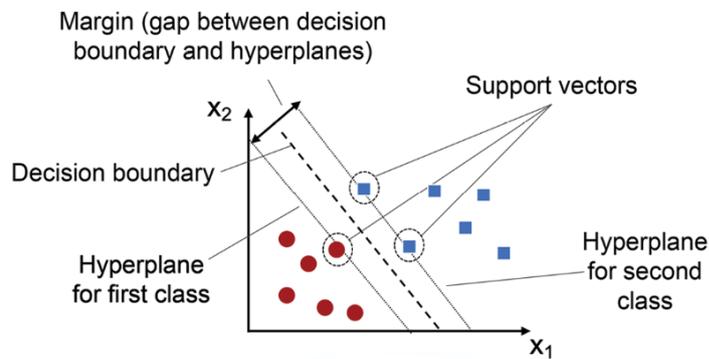


FIGURE II.8 – Illustration des Machines à Vecteurs de Support.

II.5.1.5 Algorithmes de Regroupement

Le **k-means** est un type d'algorithme d'apprentissage non supervisé utilisé pour classer les données

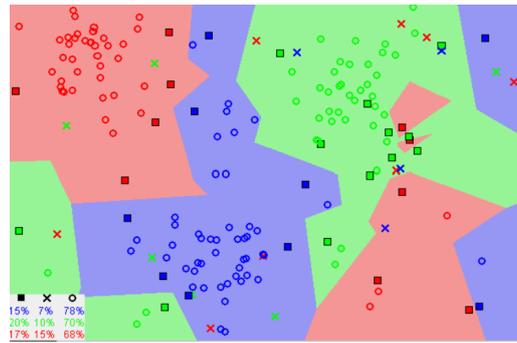


FIGURE II.9 – Illustration des k-plus proches voisins.

en différents groupes (voir Figure II.10).

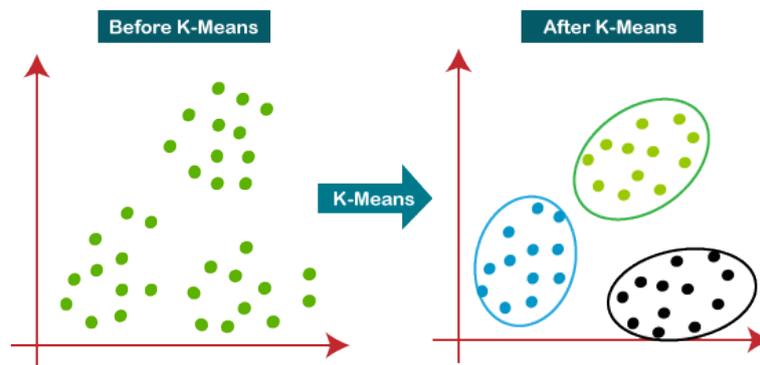


FIGURE II.10 – Avant et après l'algorithme k-means.

Le **DBSCAN** (Density-Based Spatial Clustering of Applications with Noise) est un algorithme de regroupement basé sur la densité, capable de trouver des groupes de formes arbitraires et des groupes au sein de groupes (voir Figure II.11).

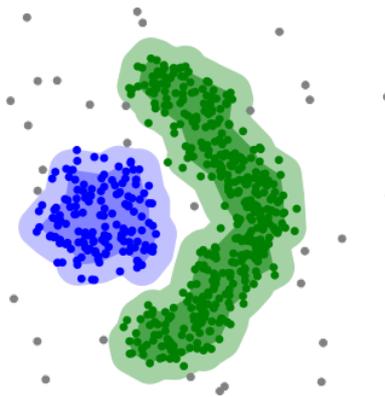


FIGURE II.11 – Regroupement spatial basé sur la densité.

II.5.1.6 Réseaux Neuronaux et Apprentissage Profond

Les **réseaux neuronaux convolutifs (CNN)** sont couramment utilisés dans les tâches de reconnaissance d'images en raison de leur taux de précision élevé.

Les **réseaux neuronaux récurrents (RNN)**, en particulier une variante avec des unités LSTM (Long Short-Term Memory), sont souvent utilisés pour les données séquentielles telles que les séries chronologiques ou les phrases.

Les **modèles Transformer** comme BERT (Bidirectional Encoder Representations from Transformers) sont particulièrement efficaces pour des tâches telles que la traduction et d'autres tâches de traitement du langage naturel.

Les **réseaux antagonistes génératifs (GAN)** sont composés de deux réseaux neuronaux, un générateur et un discriminateur, qui sont entraînés simultanément.

Les **autoencodeurs** sont un type de réseau neuronal utilisé pour apprendre des codages efficaces des données d'entrée, couramment utilisés pour la réduction de dimensionnalité ou l'apprentissage de caractéristiques.

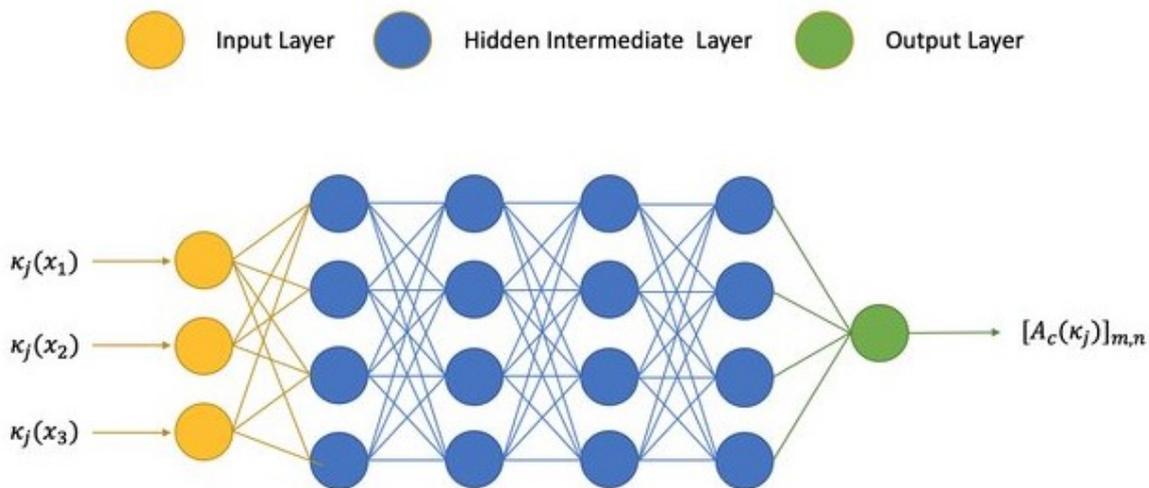


FIGURE II.12 – Une illustration d'un réseau neuronal profond.

II.6 Entraînement et Évaluation des Modèles

La formation d'un modèle d'apprentissage automatique consiste à ajuster notre modèle à l'ensemble de données d'entraînement. Le modèle apprend les schémas à partir de l'ensemble de données, ajustant ses paramètres internes pour minimiser la différence entre ses prédictions et les valeurs réelles. Ce processus est généralement guidé par une fonction appelée « fonction de perte » ou « fonction de coût ». L'objectif principal de l'entraînement est de permettre au modèle de généraliser correctement de nouvelles données invisibles.

II.6.1 Division Entraînement/Test

Lors de l'entraînement d'un modèle, il est important de ne pas utiliser l'ensemble de données complet. Au lieu de cela, nous divisons les données en un ensemble d'entraînement et un ensemble de test, en suivant généralement une règle 80/20 ou 70/30. Le modèle est entraîné sur l'ensemble d'entraînement, puis sa performance est évaluée sur l'ensemble de test invisible pour mesurer sa capacité à généraliser de nouvelles données.

II.6.2 Ajustement du Modèle et Surapprentissage

Pendant la formation du modèle, nous voulons atteindre un équilibre entre le « sous-ajustement » et le « surajustement ». Le sous-ajustement se produit lorsque le modèle ne parvient pas à capturer

les schémas des données, tandis que le surajustement se produit lorsque le modèle est trop adapté aux données d'entraînement et donne de mauvaises performances sur des données invisibles. Des techniques telles que la régularisation et l'oversampling comme la technique de sur-échantillonnage des minorités synthétiques (SMOTE) peuvent être utilisées pour lutter contre le surajustement.

II.6.3 Évaluation du Modèle

Une fois que le modèle a été entraîné, nous devons évaluer sa performance. Cela implique d'appliquer le modèle aux données de test et de comparer les prédictions du modèle aux valeurs réelles. Le choix des métriques d'évaluation dépend largement du type de problème à résoudre.

II.6.3.1 Pour les Problèmes de Régression

Les métriques telles que l'erreur quadratique moyenne (MSE), l'erreur quadratique moyenne racine (RMSE), l'erreur absolue moyenne (MAE) et le coefficient de détermination (R-carré) sont généralement utilisées. Les équations et leurs définitions sont les suivantes :

Erreur quadratique moyenne (MSE) :

L'EQM est la moyenne des différences au carré entre les résultats réels observés et les résultats prédits par le modèle.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (\text{II.1})$$

Erreur quadratique moyenne racine (RMSE) :

Le RMSE est la racine carrée de l'EQM et mesure l'écart type des résidus (erreurs de prédiction).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (\text{II.2})$$

Erreur absolue moyenne (MAE) :

Le MAE est la moyenne de la différence absolue entre les valeurs prédites et réelles.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (\text{II.3})$$

Coefficient de détermination (R^2) : Le coefficient de détermination mesure dans quelle mesure la variabilité de la variable de sortie peut être expliquée par le modèle.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (\text{II.4})$$

II.6.3.2 Pour les Problèmes de Classification

L'exactitude, la précision, le rappel, le score F1 et l'aire sous la courbe ROC (AUC-ROC) sont des métriques courantes. Les équations et leurs définitions sont les suivantes :

Exactitude : Il s'agit du rapport entre les observations correctement prédites et le nombre total d'observations. C'est une métrique utile lorsque les classes de la variable cible sont presque équilibrées.

$$Exactitude = \frac{VP + VN}{VP + VN + FP + FN} \quad (\text{II.5})$$

Précision : Il s'agit du rapport entre les observations positives correctement prédites et le nombre total d'observations positives prédites.

$$Précision = \frac{VP}{VP + FP} \quad (II.6)$$

Rappel : Le rappel est le rapport entre les observations positives correctement prédites et le nombre total d'observations positives réelles.

$$Rappel = \frac{VP}{VP + FN} \quad (II.7)$$

Score F1 : Le score F1 est la moyenne harmonique de la précision et du rappel, et cherche à trouver un équilibre entre précision et rappel.

$$F1 = 2 * \frac{Précision * Rappel}{Précision + Rappel} \quad (II.8)$$

II.6.4 Interprétation des Résultats

Une fois que nous avons ces métriques, nous les interprétons pour comprendre la performance du modèle. Par exemple :

- Dans un problème de classification binaire, l'exactitude nous indique la proportion de prédictions correctes totales.
- La précision et le rappel nous donnent une idée de la performance du modèle en ce qui concerne les faux positifs et les faux négatifs.
- Le score F1 équilibre le compromis entre la précision et le rappel.
- L'AUC-ROC mesure la capacité du modèle à distinguer entre les classes. Plus l'AUC-ROC est proche de 1, meilleur est le modèle.
- En régression, une MSE, un RMSE ou un MAE plus faible indique un meilleur ajustement aux données.
- R^2 représente la proportion de la variance d'une variable dépendante expliquée par une variable indépendante ou plusieurs variables dans un modèle de régression. Plus il est proche de 1, meilleur c'est.

N'oubliez pas qu'aucun modèle ne sera parfait, et souvent nous devons faire des compromis en fonction de ce qui est le plus important pour le problème spécifique que nous essayons de résoudre. En fin de compte, le choix du modèle et des métriques sera guidé par le contexte du problème, la qualité des données et les besoins des utilisateurs finaux ou des décideurs.

II.7 Application de l'apprentissage automatique dans l'ingénierie de forage

L'industrie en amont utilise l'apprentissage automatique depuis des décennies, bien qu'il ait pu être appelé intelligence artificielle ou par d'autres noms alternatifs. Dans la chronologie de l'application de l'apprentissage automatique, une emphase significative a été mise sur les jalons de l'industrie pétrolière et gazière, représentant des avancées dans l'acquisition de données et la numérisation [22]. Les applications sont diverses, allant de l'exploration au développement et à la production, et plusieurs exemples sont présentés ci-dessous :

- Les réseaux neuronaux à convolution (CNN), une architecture de réseau neuronal largement utilisée et efficace pour les problèmes de vision par ordinateur, ont été utilisés pour l'interprétation automatique des failles, l'identification du sel et diverses autres applications.
- Divers modèles d'apprentissage automatique ont été utilisés pour l'interpolation sismique et l'amélioration des données.
- Les réseaux antagonistes génératifs (GAN), un autre modèle puissant, ont été utilisés pour l'augmentation des données et l'inversion sismique.
- De plus, la modélisation géologique et pétrophysique a reçu une attention significative de ces modèles.
- Dans la planification du développement sur le terrain, une analyse en composantes principales suivie de méthodes de régression a été utilisée pour prédire le potentiel de production de nouveaux puits proposés. L'analyse basée sur l'apprentissage automatique a également démontré sa capacité à découvrir des modèles cachés qui ne sont pas facilement discernables dans l'espace de grande dimension, compte tenu du fait que les ingénieurs réservoir disposent de vastes quantités de données.
- Les réseaux neuronaux récurrents ont montré un immense potentiel dans les problèmes basés sur des séries chronologiques, tels que l'ajustement assisté de l'historique et la prévision des performances de production.

Maintenant, concentrons-nous sur le forage. Le forage en mer est une opération à haut risque et à forte intensité de capital (CAPEX). Le forage en eaux profondes peut coûter jusqu'à un million de dollars par jour, et un seul puits peut représenter des centaines de millions de dollars. Compte tenu de ces faits, l'automatisation grâce à des modèles d'apprentissage automatique peut considérablement améliorer la sécurité opérationnelle tout en réduisant les coûts, ce qui en fait une nécessité cruciale.

Depuis de nombreuses années, les opérations de forage sont soutenues par des processus étendus de collecte de données. Au cours des deux dernières décennies, avec l'inclusion de capteurs de fond de trou et de capacités de télécommunication, divers processus de collecte de données en temps réel sont utilisés pendant diverses opérations. Le défi actuel réside dans l'identification des événements à partir des données en temps réel et la dérivation d'informations pour éviter les temps non productifs (NPT) et les temps perdus invisibles (ILT).

Traditionnellement, les rapports de forage quotidiens (DDRs) ont servi de source d'information, décrivant le journal détaillé des opérations et des événements clés survenus sur le site de forage. Cependant, ils sont soumis à un biais humain et utilisent des formats de texte et une nomenclature incohérents. L'objectif de l'automatisation des processus est de détecter et de rectifier les problèmes qui surviennent pendant le processus de forage dans un délai permettant des opérations de forage optimales et sécuritaires.

Certaines des applications de l'apprentissage automatique dans les activités de forage et de complétion, qui promettent une amélioration de la sécurité opérationnelle et des avantages économiques, sont examinées ci-dessous.

II.7.1 Prédiction en temps réel de l'usure des outils de forage

L'apprentissage automatique a révolutionné la façon dont les ingénieurs de forage déterminent l'usure des outils de forage. Des modèles en temps réel ont été entraînés pour éliminer le besoin d'inspection manuelle et de récupération de l'ensemble fond de trou (BHA), comme le montrent les études suivantes.

Dans une étude présentée à la Conférence et Exposition internationale sur le pétrole d'Abou Dhabi par Rashidi, Hareland et Nygaard (2008) [43], une nouvelle méthode de prédiction en temps réel de l'usure des outils de forage a été développée en combinant les concepts d'énergie de la roche (énergie spécifique mécanique, MSE) et de résistance au forage (inversion du taux de pénétration, ROP). L'étude visait à optimiser les opérations de forage en analysant les données en temps réel et en prédisant l'usure des outils de forage. Les chercheurs ont dérivé des modèles mathématiques pour ROP et MSE, en tenant compte de divers paramètres de forage tels que la charge sur le trépan, la vitesse de rotation, le poids du fluide de forage et l'usure des outils. L'analyse des données sur le terrain a montré des résultats prometteurs, avec des prédictions précises de l'usure des outils de forage à rouleaux et des outils à plaquettes de diamant polycristallin (PDC). Le modèle développé peut aider à la prise de décision concernant le moment de retirer l'outil, ce qui améliore l'économie globale de l'opération de forage. L'étude a conclu que cette approche pourrait être un outil précieux pour l'analyse en temps réel de l'usure des outils de forage avec une validation ultérieure à l'aide de jeux de données plus importants.

Dans une étude de Yang, Zhao et Kottapurath (2019) publiée dans le Journal de technologie d'exploration et de production pétrolière [44], la prédiction en temps réel de l'usure des outils de forage pour les opérations de forage a été étudiée. L'étude visait à améliorer le contrôle des coûts et la prise de décision dans les opérations de forage en prédisant l'usure des outils en temps réel pour optimiser les performances de forage et éviter les problèmes d'outils. Les chercheurs ont combiné l'énergie spécifique mécanique (MSE), l'analyse en composantes principales (PCA) et l'analyse en ondelettes en utilisant les données du logiciel de surveillance des fluides de forage pour le trépan Kymera. Deux cas sur le terrain ont été étudiés pour démontrer l'efficacité des méthodes proposées. Les résultats ont montré des prédictions précises des niveaux d'usure des outils de forage, permettant des interventions en temps opportun et minimisant les temps non productifs. L'étude a souligné l'application pratique et la validité des approches mathématiques utilisées dans des scénarios réels.

II.7.2 Prévention des pertes de boue de forage pendant la circulation

Les pertes de boue de forage, comme discuté en détail dans le chapitre précédent, posent un défi important dans le processus de forage. Ce problème est influencé par divers paramètres tels que le type et le poids du fluide de forage, ses propriétés rhéologiques, la pression de la pompe et la pression interstitielle de la formation. De nombreux projets d'apprentissage automatique (ML) ont été consacrés à la résolution de ce problème spécifique, et voici quelques exemples récents :

En 2016, Behnoud Far et Hosseini [46] ont cherché à déterminer la quantité de perte de circulation lors du forage "UBD" dans la formation d'Asmari dans un champ pétrolier iranien. Ils ont utilisé des réseaux neuronaux artificiels (ANN) et des algorithmes génétiques (GA) pour établir des relations entre les paramètres associés à la perte de boue et pour déterminer le débit optimal, la pression de la pompe et le poids de la boue, respectivement. Le modèle qu'ils ont développé a obtenu des prédictions avec une valeur de R^2 de 0,9991.

En 2019, Sabah et al. [47] ont développé des modèles intelligents pour prédire la quantité de perte de circulation dans le champ pétrolier de Maroun. Les modèles étaient basés sur les arbres de décision,

les systèmes d'inférence flous basés sur le réseau neuronal adaptatif (ANFIS), les réseaux neuronaux artificiels (ANN) et un ANN hybride. Parmi ceux-ci, le modèle hybride a présenté la plus grande erreur, tandis que le modèle d'arbre de décision a démontré les meilleures performances avec une valeur de R2 de 0,9355 et un RMSE de 0,091.

Plus tard la même année, Abbas et al. ont appliqué des algorithmes ANN et SVM (machines à vecteurs de support) pour générer des modèles de prédiction des solutions de perte de circulation dans des puits déviés et verticaux. Ils ont utilisé des ensembles de données collectées à partir de 385 puits forés dans le sud de l'Irak et ont considéré 19 paramètres d'entrée tels que le poids de la boue, le poids de l'outil, l'azimut, le taux de perte, la pression de circulation et le taux de pénétration. Le modèle basé sur le SVM a surpassé un autre modèle, atteignant des valeurs de R2 de 0,97 et 0,95 pour les ensembles de données d'entraînement et de test, respectivement.

Ces études récentes en ML mettent en évidence les efforts en cours pour développer des modèles permettant de prévenir et de prédire les pertes de boue de forage. En exploitant des algorithmes avancés et des ensembles de données complets, ces modèles offrent des informations précieuses et permettent une prise de décision plus efficace dans les opérations de forage[52].

II.7.3 Amélioration de l'efficacité du processus de forage

Le terme "optimisation" englobe différentes significations dans différents processus. Un aspect commun est l'optimisation du taux de pénétration (ROP), qui sert de mesure de l'efficacité des opérations de forage. Le temps de forage et les coûts sont considérablement influencés par le ROP, qui dépend de facteurs tels que le poids sur

le trépan (WOB), le type de formation, le type d'outil, les RPM, et plus encore. Au cours des dernières décennies, de nombreuses tentatives ont été faites pour prédire et optimiser le ROP en utilisant des modèles intelligents. Voici quelques-uns des modèles récents :

En 2013, Ning et al. [48] ont développé un algorithme basé sur des réseaux neuronaux artificiels (ANN), en considérant des variables telles que la résistance à la compression uniaxiale (UCS), la viscosité apparente, la densité de la boue, les RPM, le WOB, les heures de forage, le coefficient de forabilité et le type/taille de l'outil comme entrée, avec le ROP comme sortie. Ils ont utilisé le processus hiérarchique analytique pour déterminer le poids de chaque entrée. Les auteurs affirmaient que le modèle proposé pouvait prédire le ROP avec une erreur relative moyenne de moins de 10%.

En 2014, Bataee et al. [49] et Basarir et al. ont tenté de prédire le ROP en utilisant respectivement les réseaux neuronaux artificiels (ANN) et le système d'inférence neuro-flou adaptatif (ANFIS). Bataee et al. ont considéré la profondeur, la taille du trépan, la vitesse de rotation, le poids de la boue et le WOB comme paramètres d'entrée pour optimiser les variables de forage et estimer le ROP.

Dans une autre étude réalisée en 2019, Soares et Gray [50] ont comparé les capacités de prédiction de quatre modèles analytiques et trois modèles intelligents dans la prédiction du ROP. Ils ont considéré les modèles analytiques de Bingham, BY, Hareland et Rampersad, et Motahhari, ainsi que les algorithmes d'apprentissage automatique ANN, SVM et Random Forest (RF). Ils ont indiqué que RF et le modèle BY présentaient la plus grande précision parmi les modèles intelligents et analytiques, respectivement, avec une supériorité démontrée par RF.

Un autre aspect crucial du forage, la performance de nettoyage, a fait l'objet d'efforts de prédiction en 2014 par Rooki et al. [51]. Ils ont développé des réseaux neuronaux artificiels pour estimer l'efficacité du nettoyage du trou dans le forage en mousse. Les paramètres d'entrée comprenaient l'excentricité de l'annulus, la qualité et la vitesse de la mousse, les RPM, la température et la pression. Le modèle proposé fournissait des prédictions de la concentration de déblais avec une valeur de R-au-carré de 0,914. Par la suite, en 2017, Rooki et Rakhshkhorshid [53] ont étudié l'application du réseau neuronal à fonction de base

radiale (RBF) pour prédire l'efficacité du nettoyage du trou dans les opérations de forage sous-équilibré (UBD). Le modèle développé prédisait avec succès la concentration des débris de forage

avec une erreur absolue moyenne de pourcentage de 5,7%.

Ces modèles récents illustrent les progrès réalisés dans l'optimisation des processus de forage en utilisant des algorithmes intelligents et des approches analytiques, ce qui permet d'améliorer l'efficacité et les performances[53].

II.8 Conclusion

L'impact futur de l'apprentissage automatique sur l'industrie pétrolière et gazière est immense. De l'amélioration des études exploratoires à l'optimisation des opérations, son potentiel est significatif. En embrassant cette transformation numérique, nous nous trouvons à l'aube d'une nouvelle ère où la prise de décision basée sur les données peut révolutionner l'industrie, en créant des opérations plus sûres, plus efficaces et plus productives. Ce voyage continuera de se déployer, offrant des opportunités de croissance et d'innovation continues.

Chapitre III

Conception et réalisation de la solution digital TerraVerse

III.1 Introduction

Le processus de forage de puits présente des défis complexes et exige une approche multidisciplinaire afin de résoudre efficacement les problèmes rencontrés. Les méthodes traditionnelles basées sur l'expérience et les estimations théoriques se sont avérées insuffisantes, entraînant des conséquences désastreuses tant sur le plan financier que personnel. Heureusement, grâce aux avancées technologiques dans des domaines tels que la science des données et la géoscience, ainsi qu'à l'utilisation de l'intelligence artificielle (IA), il est maintenant possible d'atténuer ces problèmes. En appliquant l'IA et en corrélant les informations recueillies avec les normes de l'industrie.

Parmi les problèmes rencontrés dans l'industrie pétrolière aujourd'hui, deux défis importants se posent lors des opérations de forage. Tout d'abord, l'identification en temps réel de la lithologie, qui consiste à classer les roches rencontrées, est limitée par les méthodes conventionnelles. De plus, la déviation non intentionnelle du trou pose des risques et des complications coûteuses. Ces défis nécessitent des solutions innovantes pour améliorer la précision et la sécurité du forage tout en réduisant les coûts associés.

Désignée sous le nom de "TerraVerse", cette solution numérique révolutionnaire intègre deux modèles prédictifs intégrés, à savoir GeoVision et PathGuard. Son utilisation se veut simple et intuitive. Elle a été conçue et optimisée pour résoudre efficacement les multiples défis liés au forage. Dans ce document, nous partagerons avec vous le parcours fascinant de sa création et les jalons qui ont marqué son développement.

III.2 Objectifs du projet

Au fil de nos investigations, qui ont été nourries par des études universitaires, des sessions de réflexion et des expériences personnelles, l'équipe TerraVerse a pu mettre en lumière une série de problèmes régulièrement évoqués au sein de la communauté de l'industrie pétrolière et gazière. Ces problèmes sont souvent qualifiés d'imprévisibles et d'ardus à résoudre. Nous avons regroupé ces problèmes en quelques points que nous allons brièvement décrire, en partageant notre point de vue sur chacun d'entre eux :

1. **Prédiction en temps réel de la lithologie** : L'objectif principal de ce projet est de développer des solutions novatrices basées sur l'apprentissage automatique pour l'identification en temps réel de la lithologie. En utilisant les paramètres de forage, nous cherchons à créer des modèles prédictifs précis et fiables qui permettront une évaluation continue de la lithologie rencontrée lors des opérations de forage. Cette capacité améliorée d'identification de la lithologie en temps réel permettra une meilleure compréhension des formations géologiques et facilitera la prise de décision rapide et éclairée pendant le processus de forage.
2. **Assurance qualité et Prédiction probabiliste de la déviation de trajectoire du puits** : Un autre objectif primordial de cette étude consiste à surmonter les interférences magnétiques qui affectent les méthodes de mesure conventionnelles. Nous nous efforçons de trouver une solution pour minimiser ces interférences et améliorer la précision des mesures. En utilisant des techniques d'apprentissage automatique avancées et en exploitant les données des mesures gyroscopiques, nous visons à développer des modèles prédictifs capables de détecter et d'évaluer les déviations de la trajectoire du puits. Cette approche permettra d'améliorer la précision du forage, de réduire les risques liés à la déviation du puits et d'optimiser le contact avec la zone de production afin de maximiser les rendements, tout en garantissant une assurance qualité plus élevée.
3. **Réduction du temps non productif (NPT)** : Un autre objectif clé de ce projet est de réduire le temps non productif (NPT). Le NPT survient lorsqu'il est nécessaire de revenir en arrière dans le forage, ce qui entraîne des pertes de temps et des coûts supplémentaires. En améliorant

la prédiction des déviations de trajectoire et en identifiant la lithologie en temps réel, nous cherchons à optimiser le chemin de forage, minimisant ainsi les risques de sortie de la zone de production et réduisant le temps non productif. Cette réduction du NPT contribuera à améliorer l'efficacité globale des opérations de forage.

III.3 La quête des données

Après avoir identifié les principaux problèmes, une première liste fondamentale de données qui seraient potentiellement utilisées pour le projet a été calculée. Cette liste comprend les exigences des paramètres incluant mais non limité à :

- La conception et design des puits.
- Les données techniques pour les puits forés.
- Paramètres de forage utilisés pour les atteindre.
- Rapports de situation pour les problèmes rencontrés.
- Études quantifiant les paramètres physiques de ces problèmes.

Vous devinez déjà que trouver cet ensemble de combinaisons serait presque impossible, surtout si l'on tient compte de la sécurité commune sur des informations aussi sensibles généralement rencontrée chez par l'industrie.

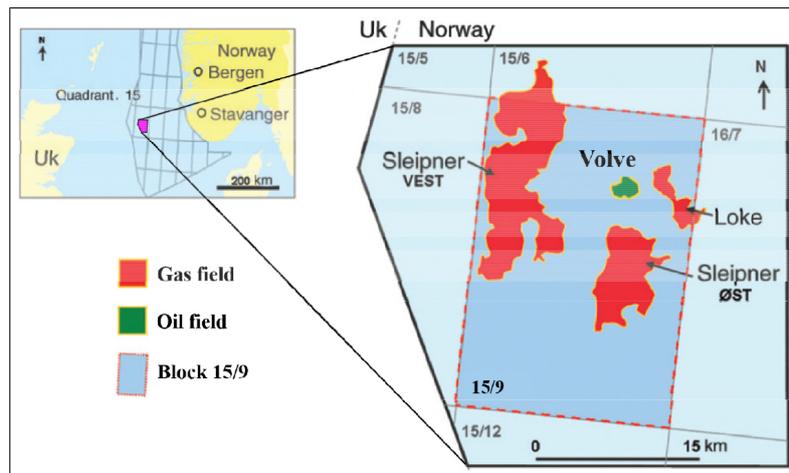


FIGURE III.1 – Localisation du champ de Volve en mer du Nord (adaptée de Seoyoon Kwon et al. 2021)

En inspectant les sources de données utilisées dans des articles similaires, jusqu'aux données gouvernementales en libre accès. Jusqu'à ce que nous tombions dessus. Des données mises à disposition depuis juin 2018 par une multinationale norvégienne du secteur de l'énergie appelée Equinor (anciennement StatOil) sous la licence suivante (CC BY 4.0). Il s'agit d'un champ communément appelé champ Volve, situé en mer du Nord, à mi-chemin entre Stavanger et Aberdeen. Découvert lors d'une exploration en 1993, par laquelle la production a commencé en 2008 et a duré huit ans.

Cette dernière est facilement accessible sur le site Web de la société et comprend un ensemble diversifié de données issues des domaines intercorrélés du génie pétrolier et des géosciences, allant des données géologiques et géophysiques au forage et à la production de 24 puits. Ces dernières sont également fournies dans différents formats (XML, WITSML, SEGY, LOG, DLIS, PDF, HTML, TXT, TIF, etc.) qui sont les plus courants dans le secteur. Le tableau ci-dessous présente un bref résumé des données susmentionnées.

En tenant compte de l'objectif principal du projet, 3 sources de données ont été considérées après l'analyse exploratoire initiale des données, qui sont les suivantes :

TABLE III.1 – Contenu des données Volve

Description	Taille comprimée Zip
Geophysical Interpretations	99 MB
GeoScience OW Archive	54.6 GB
Production Data	2 MB
Seismic ST0202	1.2 TB
Seismic ST0202 vs ST10010 4D	330.4 GB
Seismic ST10010	2.6 TB
Well Logs	7 GB
Well Technical Data	212 MB
WITSML Real-Time Drilling Data	5 GB

- **WITSML Real-Time Drilling Data** : Qui comprennent les paramètres les plus critiques liés aux différentes opérations effectuées, du forage à la diagraphie et aux opérations de reconditionnement.
- **Well Technical Data** : Ce répertoire comprend des données relatives à la conception du puits, à l'équipement utilisé, aux données de déviation, ainsi qu'aux rapports de forage quotidiens, qui étaient l'un des aspects les plus cruciaux de ce dernier pro- jet.
- **Well Logs** : Incluant les principales données liées à la caractérisation de la formation et aux évaluations des propriétés.

Bien entendu, chaque répertoire mentionné ci-dessus présente ses propres avantages, mais aussi, et surtout, ses propres défis. Nous y reviendrons dans les prochaines sous- sections.

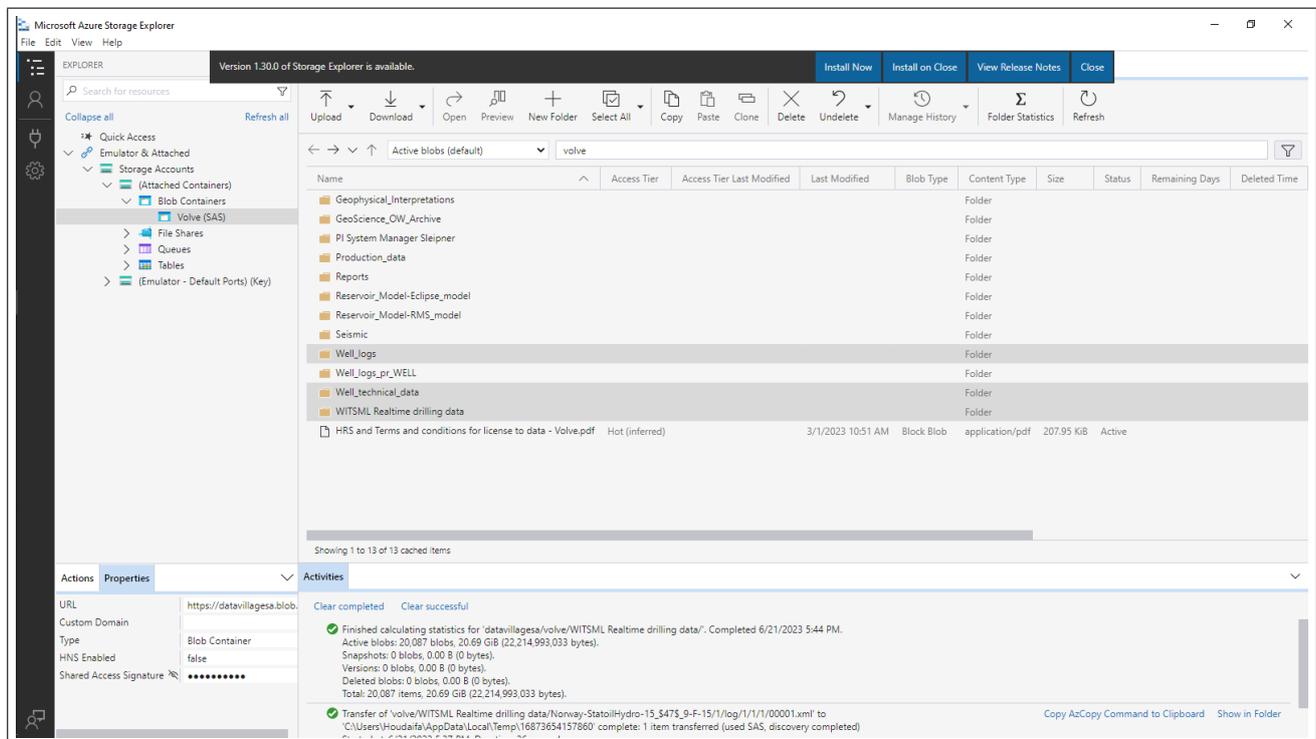


FIGURE III.2 – Microsoft Azure Storage Explorer Interface

III.3.1 Contenu des données

III.3.1.1 Well Logs

Les données de diagraphie de puits contenues dans le dossier principal sont extrêmement variées et sont utilisées à des fins multiples. Le fichier Well_logs comprend une multitude de sous-dossiers. Celles qui ont eu le plus d'importance dans le cadre du projet sont les suivantes :

TABLE III.2 – Contenu du dossier WELL LOGS

Dossier	Description
Mud_Log	Ces diagraphies contenaient des informations détaillées sur la lithologie de la formation ainsi que des paramètres supplémentaires sur les propriétés des fluides.
COMPOSITE & PI	Abréviations pour les diagraphies composites de puits et l'investigation pétrophysique, toutes deux ont fourni des données similaires relatives aux propriétés de la roche et ont constitué un élément essentiel pour la construction des modèles géomechaniques.
LWD_EWL	Comprend les données de diagraphie en cours de forage, fournissant un ensemble modéré d'informations sur les paramètres physiques de la roche et servant de QC essentiel pour les données de forage.
CORE	Comprend quelques données sur les propriétés des carottes, qui ont été essentielles pour le calibrage des MEM.
PRESSURE	Comprend les rapports de LOT, ELOT et mini Frac, et ont été utilisés à plusieurs fins dans le projet.
DIV. Rapports	Signifie vraisemblablement rapports divers. Il comprend toutes les documentations pdf des puits.

Il faut savoir que les données sont présentes dans certains puits à certaines profondeurs (généralement au niveau du prospect d'intérêt), et en moyenne 60% d'entre elles ne sont pas numériques, ce qui signifie que les données doivent être numérisées à partir de papier scanné vers des formats réellement utilisables.

III.3.1.2 Données techniques du puits

Les données fournies comprennent tous les répertoires mentionnés ci-dessous dans leur intégralité :

Une chose qui pose problème au début est que la plupart des répertoires sont vides ou simplement inaccessibles car ils sont encodés dans un format non supporté relatif à une société spécifique, par exemple les répertoires CasingSeat, CasingWear, Compass, StressCheck, WellPlan et Wellcat font référence à des produits Landmark (Landmark est le nom de solutions logicielles pour l'industrie E&P par Halliburton). Heureusement, le fichier F.edm.xml de Volve contient toutes les données Landmark exportées. Et donc c'est une base de données relationnelle, composée de 318000 lignes. Avec plus de 140 attributs différents (indiqués dans l'illustration ci-dessous). Malheureusement, la moitié du fichier est constituée de BLOBs, La figure ci-dessous montre un exemple de la structure indiquée. Chaque entrée de puits de forage est reliée à d'autres entrées par le biais du identifiant du puits de forage, identifiant du type de puits de forage. Cela a permis l'exploration semi-automatique des données à l'aide de scripts Python qui seront abordés plus loin dans le chapitre.

TABLE III.3 – Répertoires dans le Well Technical Data

Données techniques du puits
CasingSeat
CasingWear
Compass
Daily Drilling Report - XML Version
EDM.XML
Site
Site TemplateSlotStressCheck
WellPlan
WellWellbore
Wellcat
EDT EDM read me.txt
License.txt

III.3.1.2.1 Rapports quotidiens de forage et rapports de fin de puits

Les documents relatifs aux rapports de forage quotidiens se trouvent également dans l’ensemble de données appelé Données techniques des puits, il s’agit plus précisément de 1759 fichiers disponibles dans trois formats de fichiers différents, à savoir HTML, PDF et XML. Avec le même contenu identique, un travail supplémentaire a été effectué sur les DDR en XML, identique à celui introduit précédemment dans les données techniques, car ce format est le plus approprié pour analyser les fichiers de manière automatisée.



FIGURE III.3 – Exemple d’un fichier DDR de type XML)

Les informations extraites dans ces derniers étaient fondamentales pour l’ensemble du projet, et comme pour les autres fichiers auto-corrélés, la navigation dans ces derniers de manière semi-automatique est rendue possible grâce aux identifiants uniques. Voici quelques exemples d’éléments essentiels :

Il existe également de nombreux autres attributs intéressants, qui ont été utilisée le long du projet.

III.3.1.3 WITSML Real-Time Drilling Data

WITSML (Wellsite Information Transfer Standard Markup Language) est une norme industrielle permettant de transférer des données en temps réel entre des sociétés pétro- lières coopérantes. Dans

Il est également important de noter que les fréquences de d'enregistrement diffèrent d'un paramètre à l'autre, ce qui signifie qu'une nouvelle entrée est générée même si tous les attributs n'ont pas une nouvelle valeur disponible. Cela conduit à une quantité relative- ment élevée de valeurs manquantes (Not a Number, NaN) dans l'ensemble de données. Voici quelques-unes des caractéristiques les plus intéressantes de cette dernière source : Une grande partie de ces caractéristiques n'existe pas dans chaque fichier ou puits, mais il existe toujours une méthodologie pour atténuer ces dernières et elles seront explorées dans les sections suivantes. Ce format particulier n'est pas vraiment adapté aux projets d'analyse de données, ce qui signifie que les données doivent être extraites et converties dans un format plus convivial.

TABLE III.5 – Données les plus importantes

Profondeur	RPM	Inc et Azim
ROP	D-exp	Densité de boue
WOB	SPP	Taux de pompage
Température	Torque	GR

III.3.2 Défis de l'extraction de données et mitigation

L'utilisation de l'ensemble de données Volve dans son état brut est tout simplement infaisable, et cela ne concerne pas seulement la taille ou la dimensionnalité relativement élevée des caractéristiques, mais aussi la qualité des données elles-mêmes, à la fin plusieurs scripts Python ont été construits afin d'extraire et de traiter les données les plus essentielles

d'une manière semi-automatique. Néanmoins, au cours du processus, nous avons rencontré un certain nombre de problèmes qui ont été énumérés ci-dessous :

III.3.2.1 Contrôle de la qualité des données

Une exploration plus poussée des données pour des caractéristiques très spécifiques liées aux besoins du projet a parfois montré que des puits de forage entiers manquaient. En outre, une grande partie des données représentait des Blobs, c'est-à-dire des données codées dans un format non pris en charge, potentiellement fournies par le produit numérique d'une entreprise. Diverses techniques d'augmentation des données ont été appliquées pour assurer une extraction maximale d'information. En outre, la plupart des rapports ont été dépouillés manuellement en fonction des paramètres souhaités et à des fins de contrôle de la qualité.

Quant à l'invalidité des données (hors de la plage spécifiée), elle a été traitée par un certain nombre de techniques de pré-traitement et d'algorithmes de détection des valeurs aberrantes, les données ont été corrigées lorsque c'était possible et dans d'autres cas, totalement abandonnées, pour éviter l'impact sur les modèles.

Un nombre modeste d'ingénierie des caractéristiques (calcul de nouveaux paramètres) a également été réalisé à travers les différentes applications afin d'améliorer les performances des modèles et de trouver de fortes corrélations entre les données d'entraînement et le résultat souhaité, nous nous étendrons sur chacun d'entre eux dans la section consacrée aux modèles d'apprentissage automatique. L'incomplétude, la redondance et l'incohérence des données ont également été traitées par le biais d'un certain nombre de méthodes statistiques et par la combinaison des différentes sources disponibles, malgré la grande diversité des types de formats (XML, WITSML, SEGY, LOG, DLIS, PDF, HTML, TXT, TIF, etc.).

III.4 TeraVerse Solution

L'histoire derrière le nom 'TerraVerse' de notre application web c'est qu'il reflète parfaitement notre vision et les possibilités infinies qu'il offre dans le domaine de la géologie et de la prédiction probabiliste de la déviation. 'Terra' représente la Terre, symbolisant notre concentration sur l'exploration et la compréhension de ses formations géologiques et de ses ressources. Le terme 'Verse' évoque un univers de connaissances, signifiant l'immensité des données et des idées que nous cherchons à découvrir grâce à nos modèles d'apprentissage automatique. TerraVerse englobe non seulement l'application web dans son ensemble, mais se décline également en deux applications distinctes :

- **GeoVision.**
- **PathGuard.**

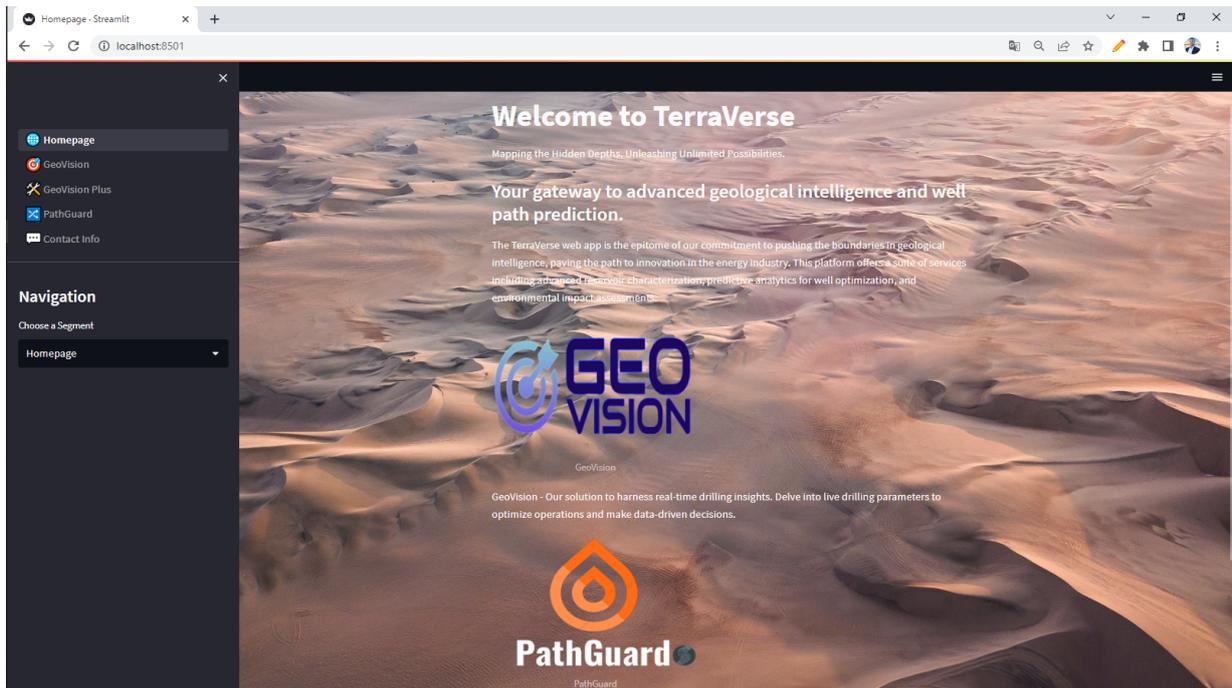


FIGURE III.5 – Page d'accueil en cours de conception

Dans les prochains passages, nous allons aborder chaque élément de manière détaillée.

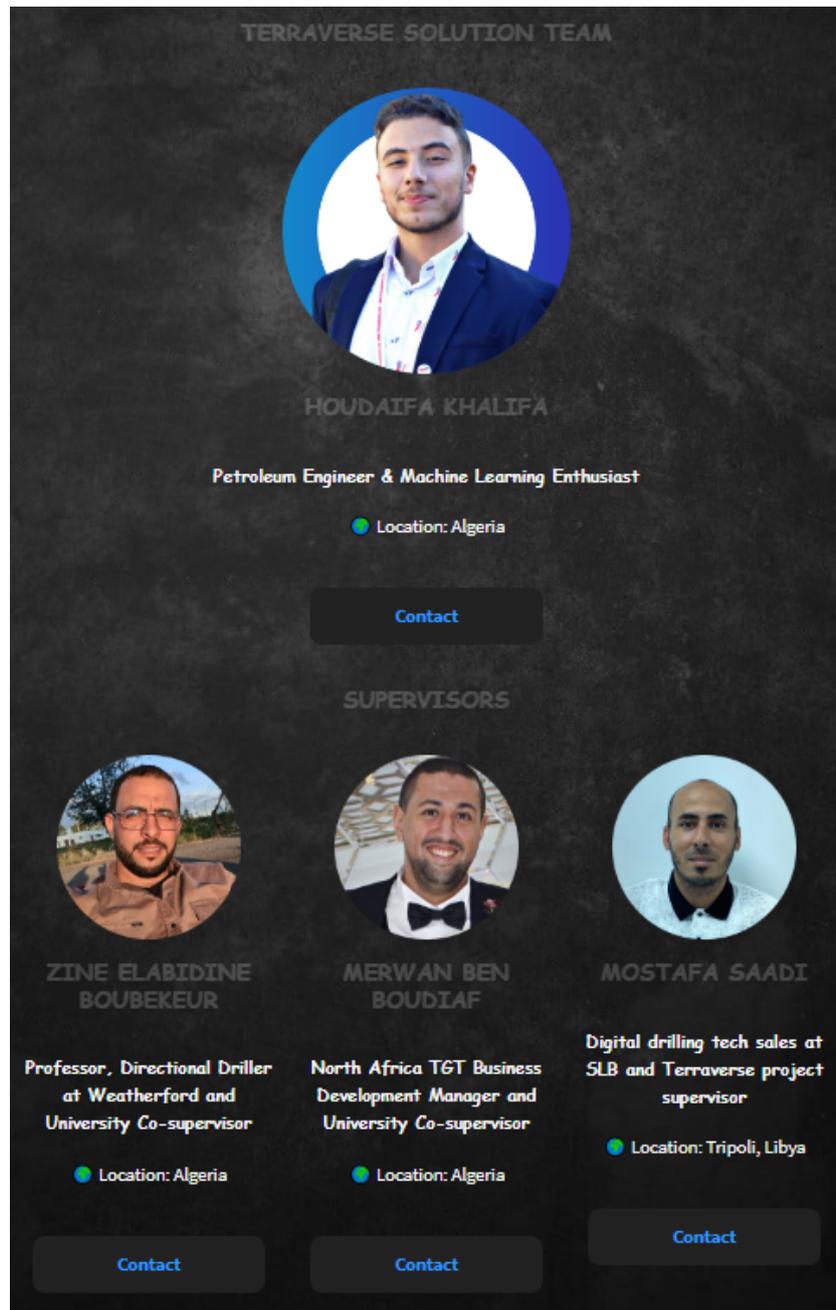


FIGURE III.6 – Equipe de TerraVerse

III.4.1 GeoVision

III.4.1.1 Problématique

Il est courant de prédire la lithologie à partir des diagraphies géophysiques car ces diagraphies sont des indicateurs directs de la lithologie. De nombreux travaux ont déjà été réalisés dans ce domaine d’application, et il y a même des concours publics comme ”le concours d’apprentissage automatique FORCE 2020 avec Wells et Seismic”.

Dans cet solution déigital, nous souhaitons faire quelque chose de différent, à savoir prédire la lithologie à partir des données de forage. C’est différent car la relation entre la lithologie et les mesures de forage telles que le couple, la vitesse de rotation du trépan ou la course de la pompe n’est pas encore très bien comprise.

La prédiction de la lithologie basée sur les données de forage sera utile pour la Géopositionnement en temps réel dans l’industrie pétrolière et gazière.

De plus, il convient de noter que l'utilisation de la diagraphie pendant le forage (LWD) pour obtenir ces informations est une technologie coûteuse. Cependant, malgré cette méthode existante, il persiste un décalage en profondeur entre les mesures effectuées par le LWD et la position actuelle du trépan, ainsi qu'un délai dans la transmission des données de LWD vers la surface. Par conséquent, nous avons besoin d'une caractérisation de la lithologie en temps réel.

Maintenant, salissons nos mains !

III.4.1.2 Analyse exploratoire des données

Le jeu de données d'un puits particulier comprend 33 colonnes ; 2 colonnes (Profondeur et Temps) sont appelées les identifiants car nous obtenons les mesures à chaque profondeur en temps réel, 30 colonnes sont les grandeurs mesurées obtenues à partir des données de forage, que nous appelons les caractéristiques en termes d'apprentissage automatique (ML), et la dernière colonne LITH est l'étiquette de lithologie obtenue à partir des données de diagraphie de boue.

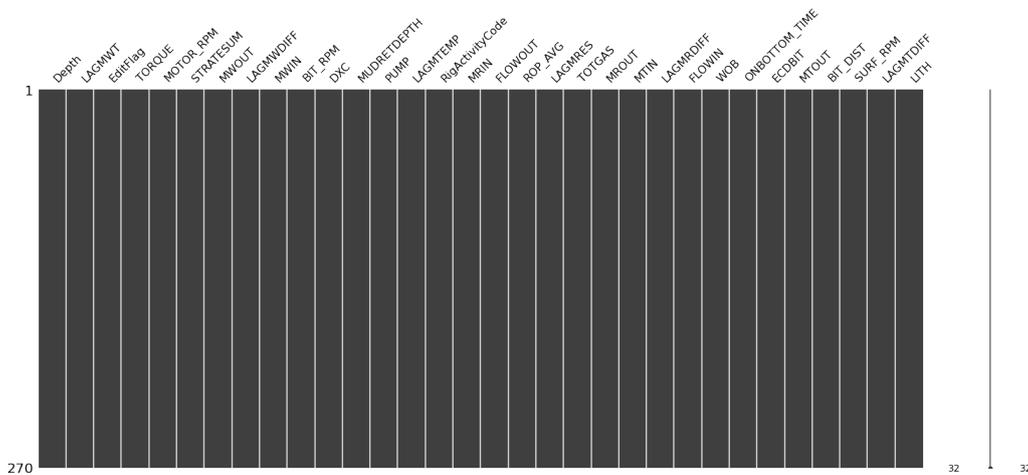


FIGURE III.7 – Couverture des données

Il existe 5 classes de lithologies dans la colonne LITH : grès, argilite, marne, calcaire et dolomite. Étant donné que le calcaire et la dolomite ont le moins d'observations, nous n'utiliserons que le grès, l'argilite et la marne pour notre classification. Notez également le déséquilibre du nombre d'observations. Dans ce cas, nous avons une classification déséquilibrée. Nous traiterons cela en premier lieu avant de faire nos prédictions.

Vous serez surpris par le grand nombre de grandeurs mesurées dans ces données de forage. Toutes ces grandeurs mesurées sont abrégées, mais nous ne décrirons pas une par une chacune de ces grandeurs mesurées, car nous n'avons certainement pas besoin de toutes ces 30 caractéristiques pour notre prédiction ! Certaines d'entre elles sont erronées et inutiles, c'est pourquoi nous effectuerons une sélection de caractéristiques.

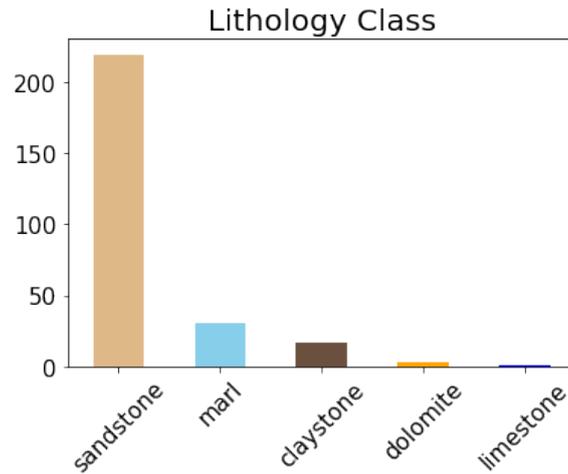


FIGURE III.8 – Décompte de fréquence des classes de lithologie

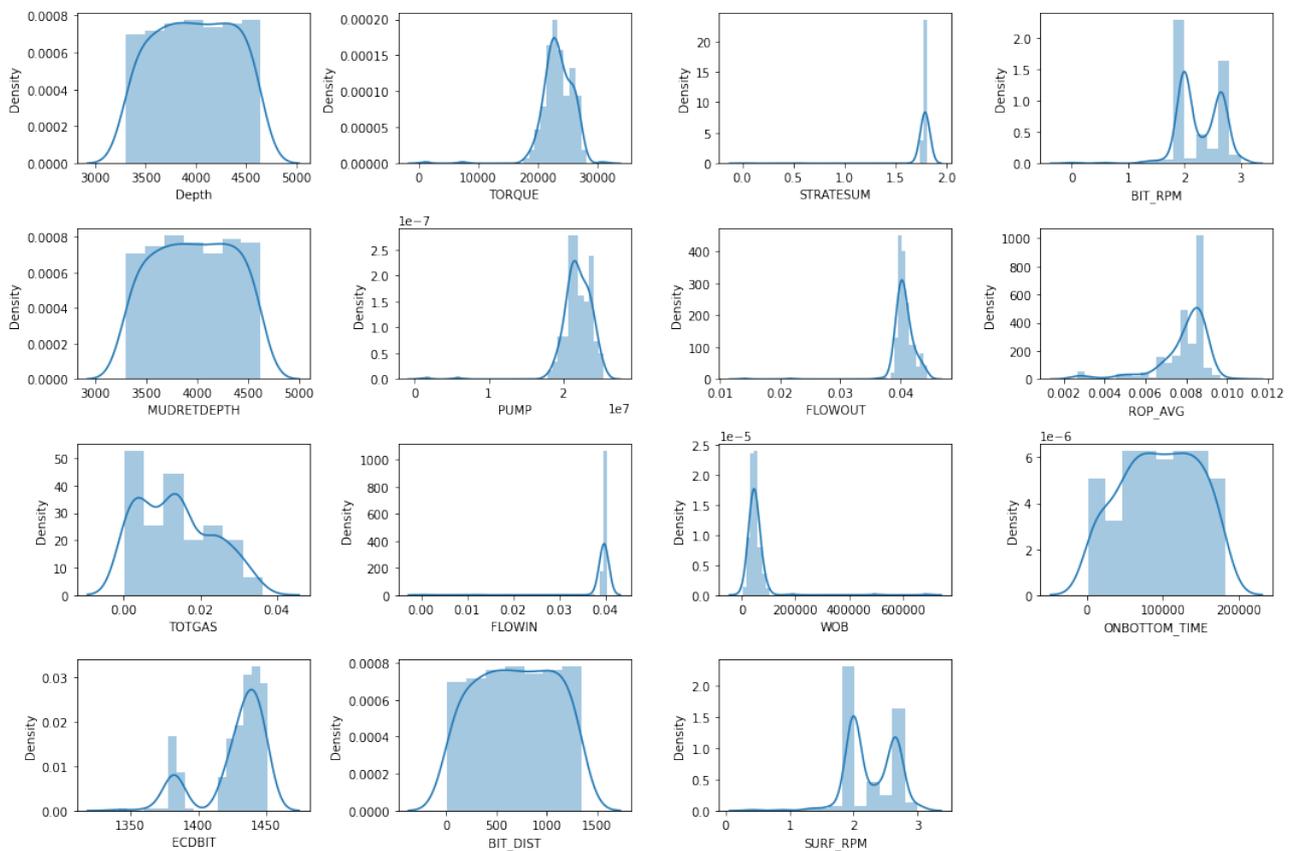


FIGURE III.9 – Distribution des caractéristiques dans les données de forage

III.4.1.2.1 Sélection de caractéristiques

Lorsque nous faisons de l'apprentissage automatique, nous faisons de notre mieux pour réduire le nombre de paramètres trop nombreuses, car moins de caractéristiques permettent à notre modèle prédictif de s'entraîner plus rapidement. Fondamentalement, il existe deux façons de procéder : éliminer les caractéristiques erronées et éliminer les caractéristiques redondantes.

Caractéristiques erronées

Une caractéristique erronée est une caractéristique qui ne peut pas aider à classer notre cible. Elle contient des valeurs constantes ou uniformes. C'est la même chose lorsque nous classifions des fruits,

par exemple, des pommes, des oranges et des pastèques. Utiliser la "forme" comme prédicteur n'est pas suffisant. Tous ces fruits sont bien sûr "ronds" en forme. Si nous avons 1 000 observations étiquetées comme des pommes, des oranges ou des pastèques, alors toutes les valeurs de notre colonne "forme" seront "rondes". Dans ce cas, nous avons besoin d'un prédicteur plus distinct, tel que la "couleur".

Cependant, les caractéristiques erronées ne sont pas toujours causées par des valeurs constantes. Parfois, il peut y avoir des valeurs constantes à travers la caractéristique parce que cette caractéristique particulière est une annotation.

Il y a des cas où nous n'avons pas besoin de supprimer une caractéristique. Comme toutes les données peuvent comporter des erreurs de mesure, la présence de valeurs aberrantes peut donner l'impression que la caractéristique est erronée. Normalement, nous pouvons repérer les valeurs aberrantes grâce à nos connaissances du domaine. Si cette caractéristique semble toujours utile, nous pouvons simplement supprimer les valeurs aberrantes et la conserver.

Afin de repérer les caractéristiques erronées, nous devons faire preuve de sagesse et utiliser différentes méthodes. Parmi ces méthodes, les statistiques descriptives jouent un rôle important. Dans Pandas, nous pouvons utiliser la fonction dataframe.describe() pour obtenir un aperçu des statistiques descriptives des différentes caractéristiques. Ci-dessous, vous trouverez une sélection de ces statistiques descriptives.

TABLE III.6 – Caractéristiques erronées des données de forage

	RigActivityCode	DXC	MWIN	LAGMTDIFF	WOB
count	266.0	266.00	266.00	266.00	266.00
mean	111.0	0.99	1319.80	-286.16	57460.18
std	0.0	0.00	2.59	69.65	122318.81
min	111.0	0.99	1280.00	-303.15	-43003.60
25%	111.0	0.99	1320.00	-303.15	37032.77
50%	111.0	0.99	1320.00	-303.15	45629.43
75%	111.0	0.99	1320.00	-303.15	55076.90
max	111.0	0.99	1320.01	0.00	1857840.10

Les deux codes RigActivityCode et DXC ont des valeurs constantes car la moyenne, le min, le max et les percentiles sont égaux. Ces caractéristiques sont erronées, donc ne peuvent pas être utilisées. Nous constatons que "RigActivityCode" est simplement une annotation qui n'a aucune signification. Ensuite, MWIN affiche tous les percentiles égaux et max, tandis que LAGMTDIFF montre tous les percentiles égaux et min. Nous pouvons en conclure que ces deux sont également erronés.

Maintenant, regardons WOB. Les percentiles semblent corrects, mais le min est négatif et le max est très grand. Comme WOB est le poids sur le trépan, des valeurs de poids négatives ne peuvent physiquement pas être vraies. De plus, un très grand max signifie qu'il y a un outlier. Cependant, nous conservons WOB car nous pouvons supprimer uniquement les outliers, c'est-à-dire supprimer la portion de données avec WOB inférieur à 0 et inférieur à 1,000,000.

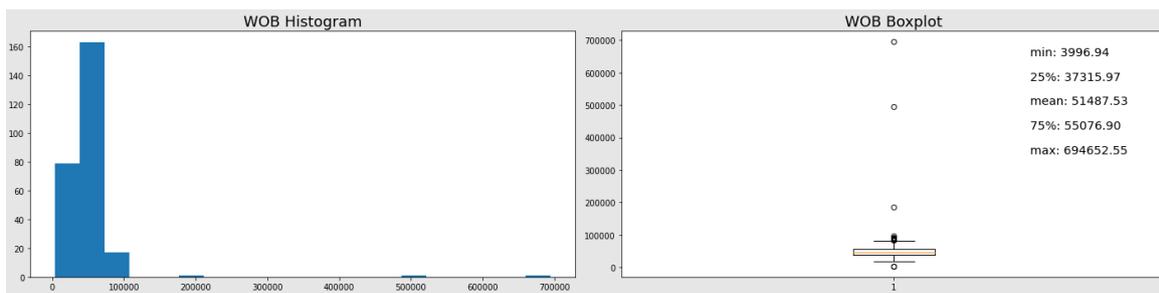


FIGURE III.10 – Aperçu de la répartition des valeurs de WOB à la fois dans l’histogramme et le boxplot

Une autre façon de repérer les caractéristiques erronées est d'identifier le pourcentage de données ayant des zéros. Ici, nous avons calculé le pourcentage (table III.7).

Caractéristiques	Pourcentages de zéros
STRATESUM	0%
MWOUT	0%
LAGMWDIFF	81%
MWIN	0%
BIT_RPM	0%
DXC	0%
MUDRETDEPTH	0%
PUMP	0%
LAGMTEMP	5%
RigActivityCode	0%

TABLE III.7 – Le pourcentage de données contenant des zéros

Dans le résultat du calcul ci-dessus, LAGMWDIFF a 81% de ses valeurs étant zéro. C'est une raison forte pour laquelle nous devons abandonner cette caractéristique également.

Caractéristiques redondantes

Les caractéristiques redondantes ont des caractéristiques d'avoir une haute corrélation entre elles. Nous pouvons tracer une heatmap des corrélations parmi ces variables. La corrélation de Spearman est utilisée. La règle pour notre sélection de caractéristiques est que les caractéristiques hautement corrélées ne seront pas utilisées.

Après avoir repéré certaines caractéristiques hautement corrélées, nous pouvons maintenant supprimer ces caractéristiques. Mais n'oubliez pas, nous devons utiliser nos connaissances de domaine pour justifier pourquoi nous devons le supprimer.

Le SURF_RPM (Rotation par minute à la surface) est corrélé au BIT_RPM (Rotation par minute au trépan) car lorsqu'on utilise le trépan, la rotation mesurée au trépan sera toujours équivalente à la rotation mesurée à la surface. Nous supprimons le SURF_RPM. Nous supprimons également MUDRETDEPTH, BIT_DIST, ONBOTTOM_TIME car ces trois sont corrélés.

Le FLOWIN (flux de boue entrant) est lié au STRATESUM (taux total de coups de pompe) car lorsque la boue entre, les ingénieurs de forage obtiennent la mesure de la pompe. Nous choisissons de supprimer FLOWIN.

Cependant, nous voyons que FLOWOUT (flux de boue sortant) n'est pas corrélé à STRATESUM, car la quantité de boue qui sort n'est pas contrôlée par la quantité de boue qui entre. Donc, nous gardons ces deux.

Caractéristiques sélectionnées

Surprise, après une audition très serrée, nous avons réduit 30 caractéristiques à seulement 9 caractéristiques ! Les caractéristiques sont les suivantes :

- **TORQUE** : Couple de surface moyen (N.m)
- **STRATESUM** : Taux total de coups de pompe (Hz)
- **BIT_RPM** : Rotation par minute au trépan (c/s)
- **PUMP** : Pression de la pompe (Pa)
- **FLOWOUT** : Flux de boue sortant (m³/s)
- **ROP_AVG** : Profondeur moyenne de rotation par minute (m/s)

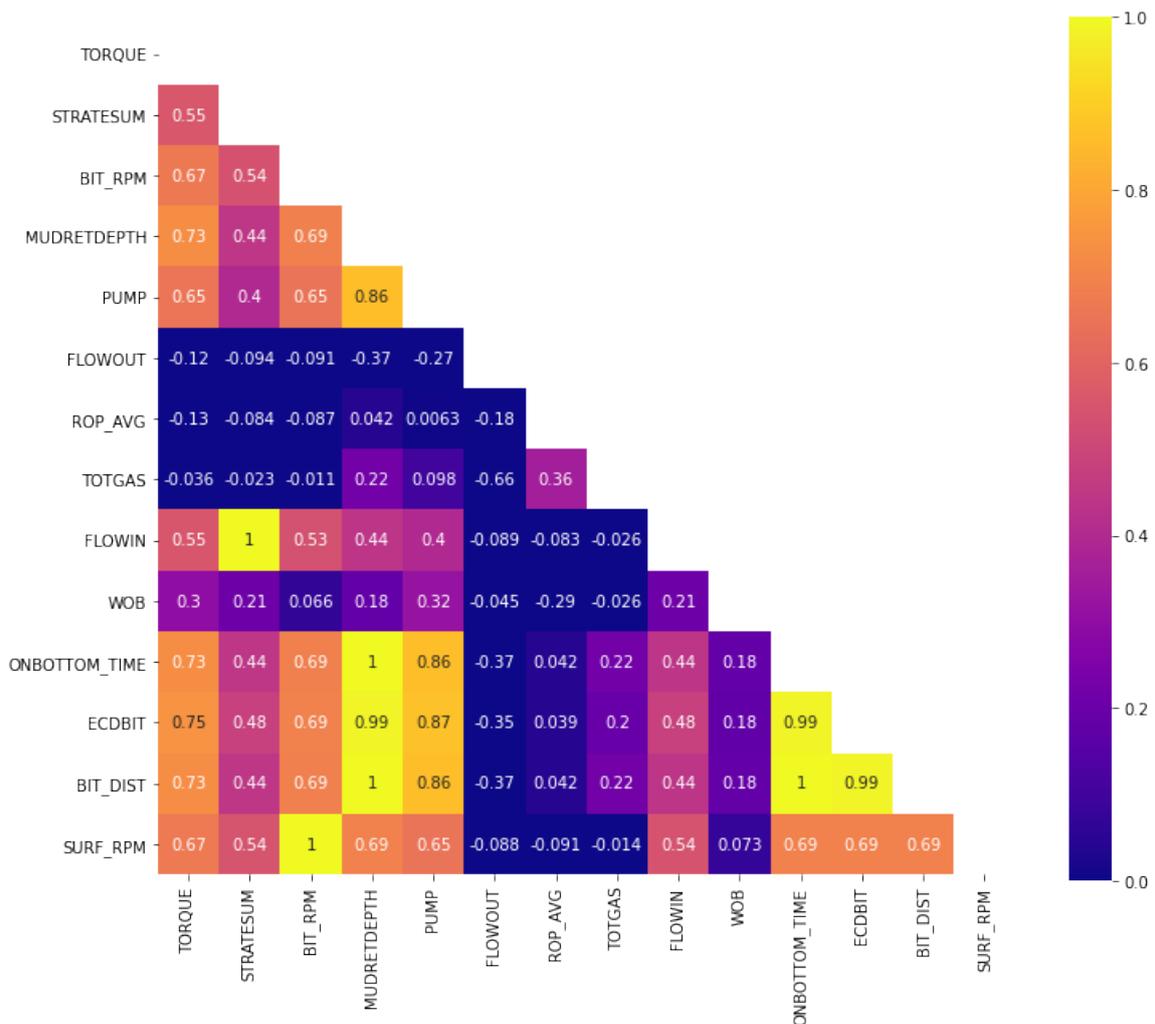


FIGURE III.11 – Plot de corrélations avant le traitement des données(Heatmap).

- **TOTGAS** : Teneur totale en gaz (ppm)
- **WOB** : Poids sur le trépan (N)
- **ECDBIT** : Densité de circulation effective sur le trépan (kg/m³)

Nous espérons que ces caractéristiques pourront aider à classer les grès, les argilites et les marnes. Voici nos données après la réduction des caractéristiques. Nous pourrions afficher les données dans un style de log avec l’axe des x comme caractéristiques et l’axe des y comme profondeur. Le log à droite est la colonne de lithologie ; le Bleu clair (#87ceeb) est la marne, le Marron foncé (#6b4f3f) est l’argile, et le Beige (#deb887) est le grès.

Le pairplot est une technique d’analyse exploratoire des données qui nous permet de visualiser les relations entre les caractéristiques de forage et les types de lithologie. En examinant les schémas et les tendances dans le pairplot, nous pouvons identifier les caractéristiques les plus discriminantes pour prédire la lithologie. Cela nous guide dans notre analyse prédictive en nous indiquant où concentrer notre attention. La bibliothèque seaborn facilite la création de pairplots et nous offre une vue globale des relations entre les variables (voire figure III.14).

Suréchantillonnage de la classe déséquilibrée Nous savons que le nombre d’observations de nos classes est déséquilibré. Avoir une classe déséquilibrée fera en sorte que notre modèle prédictif ait une haute précision mais un faible rappel. Pour corriger cela, nous devons rééchantillonner les données. Il existe 3 techniques de rééchantillonnage parmi beaucoup d’autres ; le sous-échantillonnage, le suréchantillonnage, et une combinaison des deux.

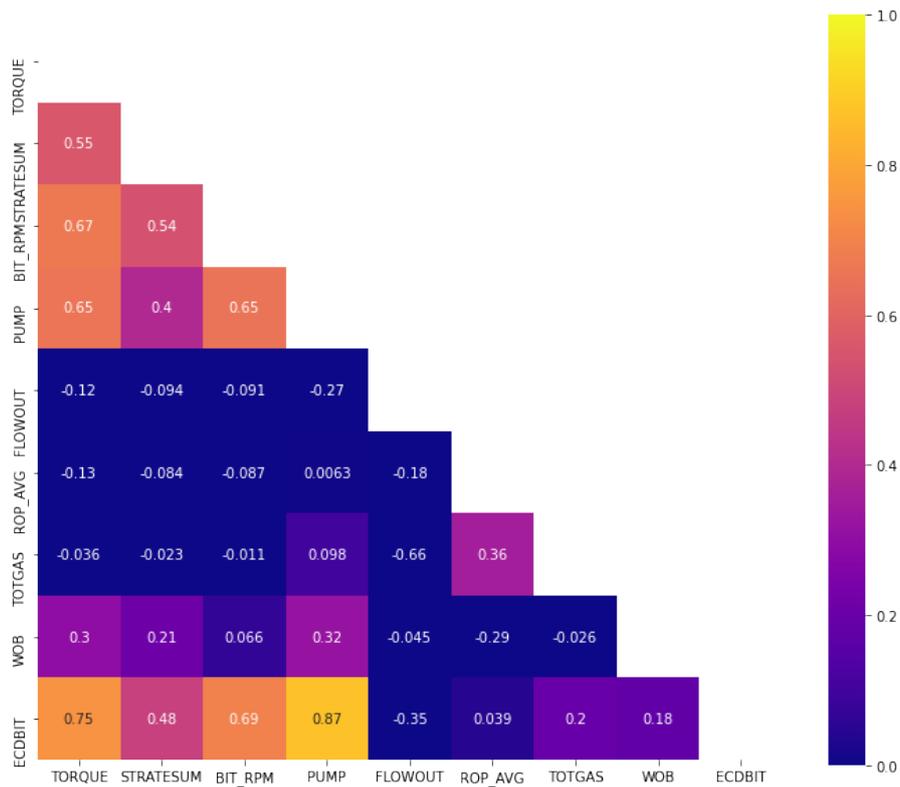


FIGURE III.12 – Heatmap après le traitement des données.

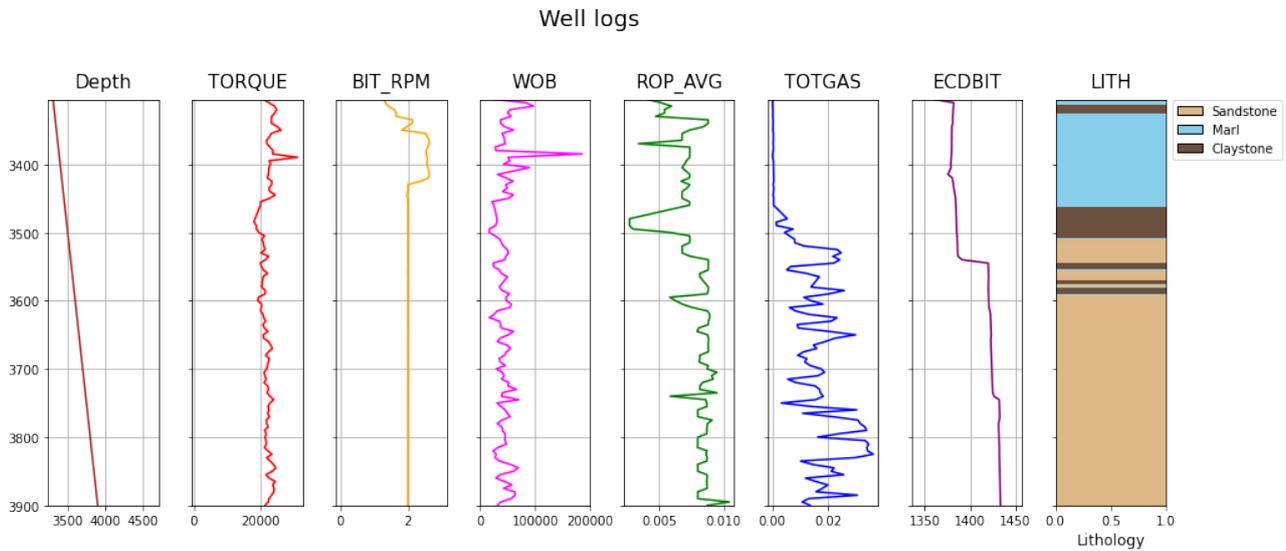


FIGURE III.13 – Visualisation de quelques paramètres des données utilisées en logs

Dr. Boaz Shmueli a expliqué la précision et le rappel plus en profondeur et un autre article de Jack Tan qui explique pourquoi nous devrions nous préoccuper de la classe déséquilibrée ainsi que les techniques mentionnées ci-dessus.

Ici, nous utiliserons le suréchantillonnage avec SMOTE (Synthetic Minority Over-sampling Technique figure III.15) qui est fourni dans la bibliothèque Python imblearn. À l’origine, nous avons 219 :30 :17 pour les échantillons de grès, d’argilite et de marne. Après SMOTE, nous aurons tous 219 :219 :219.

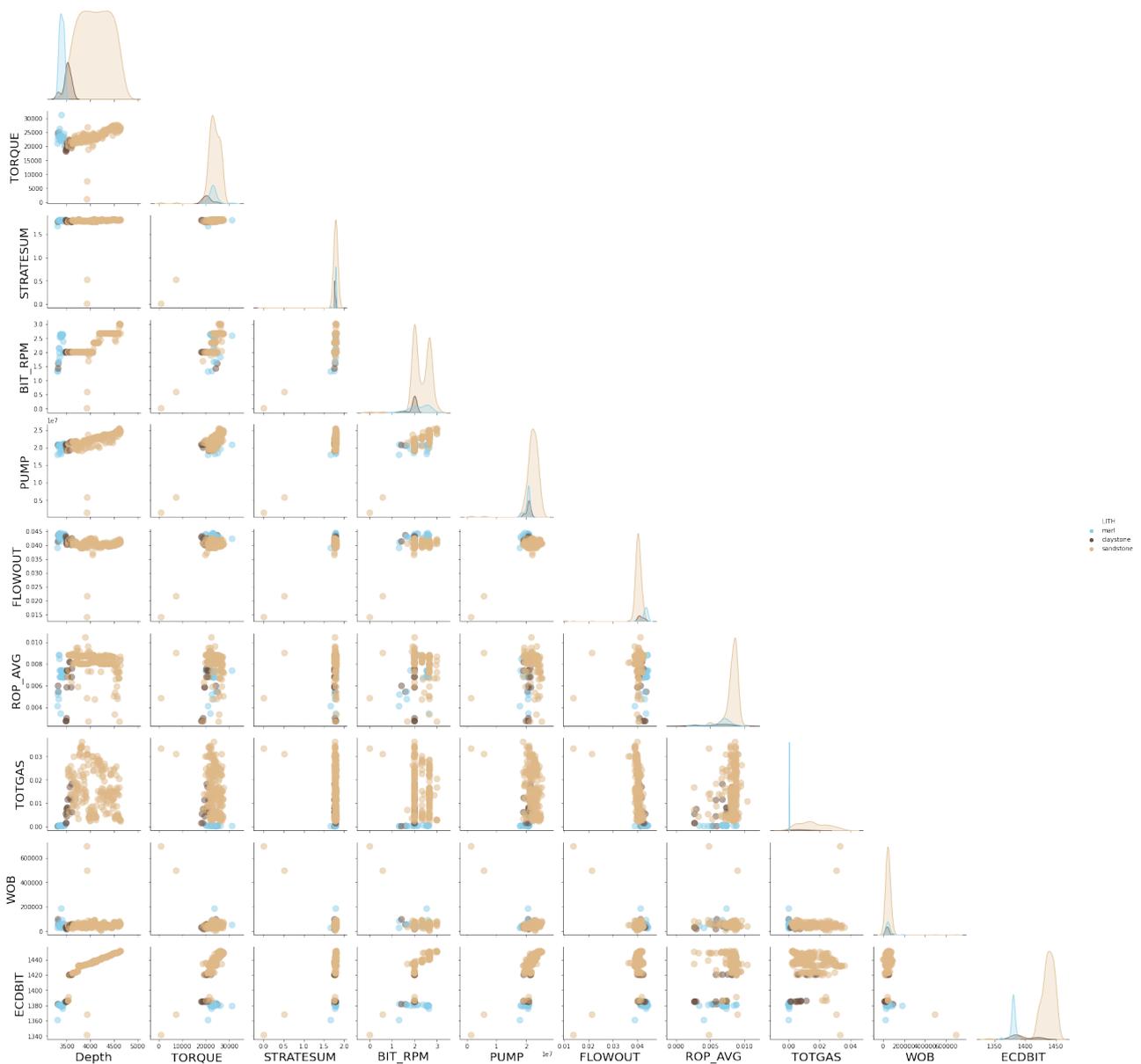


FIGURE III.14 – Matrice de diagramme des paramètres les plus importants(Pair Plots)

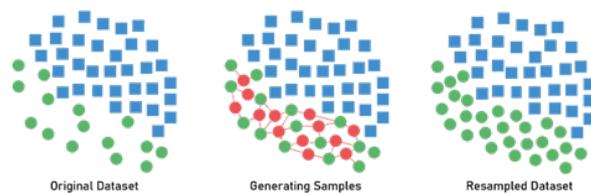


FIGURE III.15 – Illustration de la technique SMOTE

III.4.1.3 Application de l'apprentissage automatique

Dans l'analyse, plusieurs modèles d'apprentissage machine ont été formés - y compris Random-Forest, GradientBoost, LinearSVC, AdaBoost, et KNeighbors - pour classer l'ensemble de données donné. Tous les modèles ont subi un processus d'évaluation rigoureux, où ils ont été formés, validés et testés pour leur capacité à prédire la classe cible avec précision. La métrique de performance utilisée pour la comparaison était la précision, car elle mesure la proportion de prédictions correctes sur le nombre total d'instances.

Cependant, c'est le modèle RandomForest Classifier qui a surpassé tous les autres. Le processus de formation pour le modèle RandomForest a commencé par séparer l'ensemble de données en variables de caractéristiques (X) et la variable cible (y), suivi d'une division en une répartition train-test de 70 :30. La technique de suréchantillonnage des minorités synthétiques (SMOTE) a été utilisée pour gérer tout déséquilibre de classe dans les données, et le StandardScaler a été employé pour la normalisation des caractéristiques.

Ce modèle a ensuite été formé et évalué en utilisant la validation croisée. Nous avons besoin d'évaluer la performance de notre classificateur, nous avons donc fait de la validation croisée (CV). Parmi les différents types de CV, ici nous choisissons d'utiliser le CV à K-Fold stratifié répété avec 10 séparations et 3 répétitions avec un accent sur trois métriques de scoring clés - la précision, la précision macro et le rappel macro. De plus, le modèle a implémenté une sélection de caractéristiques en utilisant SelectFromModel pour s'entraîner sur les caractéristiques les plus importantes.

La validation croisée comporte deux étapes principales : la division des données en sous-ensembles (appelés plis) et la rotation de l'entraînement et de la validation entre eux. La technique de division des données présente généralement les caractéristiques suivantes :

Chaque pli a une taille approximativement équivalente. Les données peuvent être sélectionnées de manière aléatoire dans chaque pli ou stratifiées. Tous les plis sont utilisés pour entraîner le modèle, à l'exception d'un pli qui est utilisé pour la validation. Ce pli de validation doit être tourné jusqu'à ce que tous les plis aient été utilisés comme pli de validation une fois et une seule. Chaque exemple est recommandé d'être contenu dans un seul et unique pli. Pour une validation croisée avec $cv=5$, cela ressemble à ceci.



FIGURE III.16 – Illustration de la Validation Croisée à Cinq Plis par Ivanna Baturynska

Les données de test ont été suréchantillonnées en utilisant SMOTE pour assurer des prédictions précises. Le score de précision, représentant la capacité du classificateur à ne pas étiqueter un échantillon négatif comme positif, a été calculé à la fin. Finalement, c'est la haute précision et la précision du modèle RandomForest qui l'ont rendu le choix préféré parmi tous les modèles formés ou on a constaté que Le meilleur modèle est RandomForest avec une précision de 0,96% et le pire modèle est AdaBoost avec une précision de 89%.

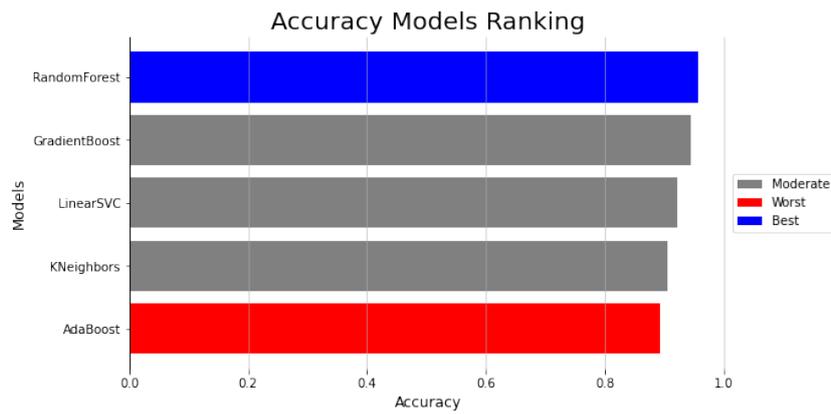


FIGURE III.17 – Classement de précision

Comme nous l’avons déjà discuté, nous utilisons la précision, la précision et le rappel comme nos métriques. En utilisant le classificateur RandomForest, nous obtenons une précision de 96%, une précision de 98% et un rappel de 86%. C’est une bonne performance car ces métriques révèlent que le modèle peut prédire chaque classe très

TABLE III.8 – Métriques du classificateur RandomForest

Métriques	Scores
Mean Accuracy	0.9564
Mean Precision	0.8420
Mean Recall	0.8608
Precision	0.9795

Nous avons également tracé la matrice de confusion pour nous aider à comprendre comment le modèle AdaBoost classe chaque vraie classe de notre ensemble de test contre chaque classe prédite. Nous voyons que 61 échantillons d’argilite, 59 de marne et 65 de grès sont prédits avec succès sur 65 pour chacun.

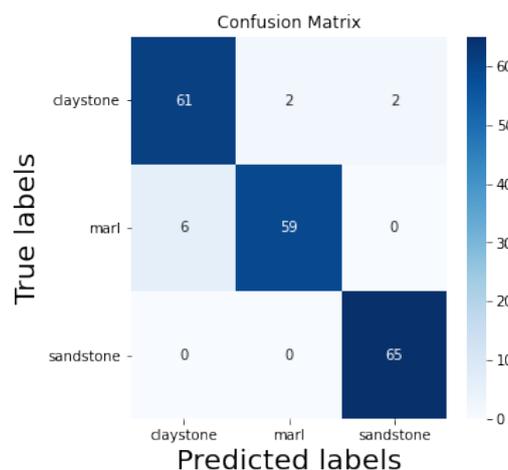


FIGURE III.18 – matrice de confusion

Voici notre rapport de classification montrant les métriques de chaque classe. Cela montre que le modèle prédit chaque classe également bien.

Ce changement, l’interprétabilité du modèle est également importante. Ici, nous faisons un graphique d’importance des caractéristiques, montrant quelles caractéristiques ont une importance plus élevée

TABLE III.9 – rapport de classification

	Precision	Recall	F1-score	Support
Claystone	0.91	0.94	0.92	65
Marl	0.97	0.91	0.94	65
Sandstone	0.97	1.00	0.98	65
Accuracy			0.95	195
Macro Avg	0.95	0.95	0.95	195
Weighted Avg	0.95	0.95	0.95	195

dans la construction du modèle de forêt aléatoire (voire figure III.19. La caractéristique avec la plus grande importance est TOTGAS et la plus petite est WOB.

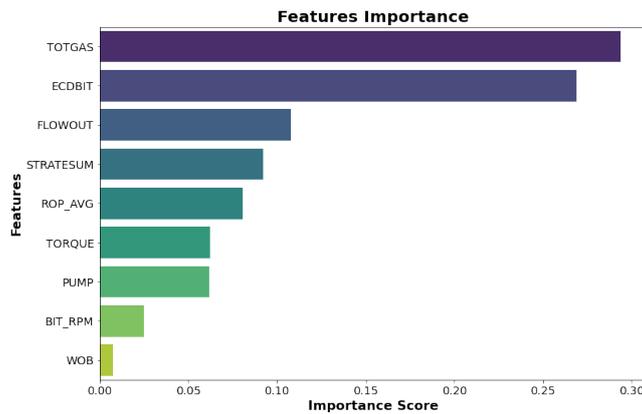


FIGURE III.19 – Bar plot pour illustrer l'importance des caractéristiques

Après avoir entraîné notre modèle, nous avons effectué une évaluation sur les données du puits 1 et comparé la lithologie prédite avec la lithologie réelle. Le modèle a démontré une précision impressionnante et a réussi à prédire la majorité des lithologies présentes dans le puits.

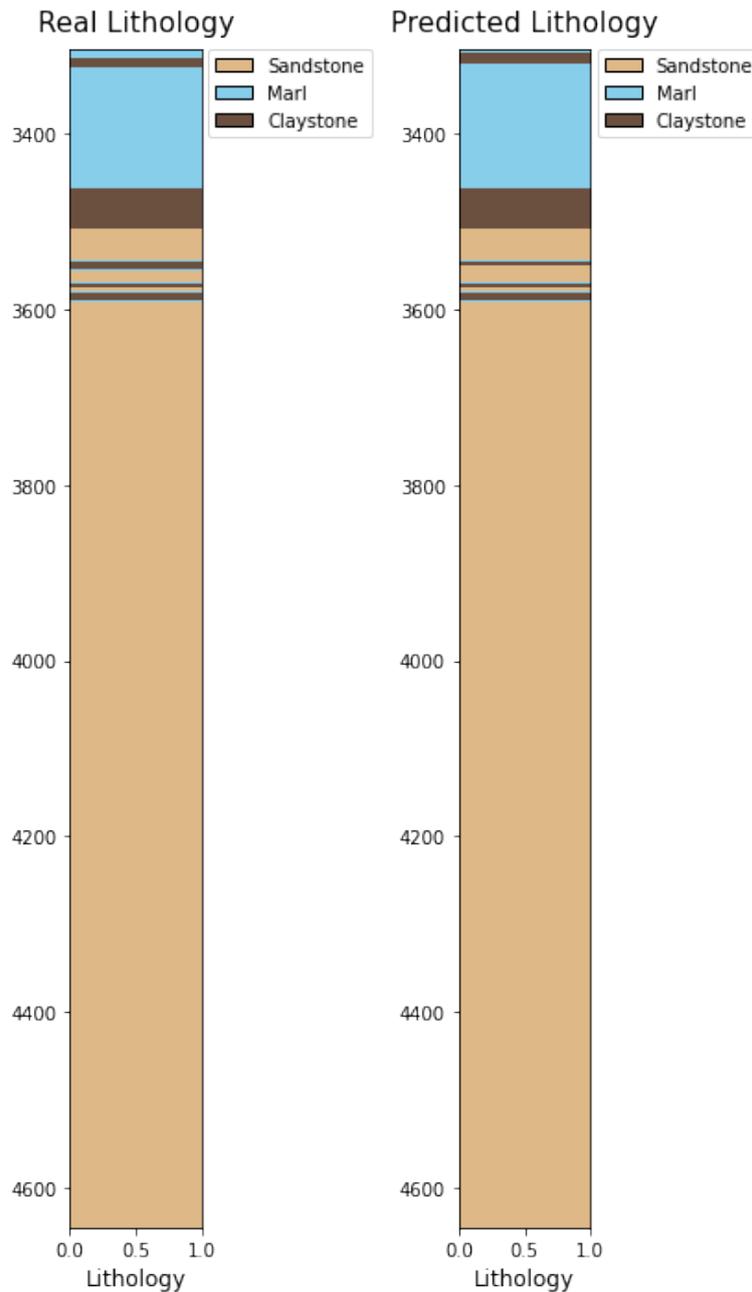


FIGURE III.20 – Comparaison entre la lithologie réelle et prédite - Puits 1

Encouragés par ces résultats, nous avons ensuite testé le modèle sur les données d'un nouveau puits qu'il n'avait jamais rencontré auparavant. Malgré son succès global dans la prédiction de différentes lithologies, il a rencontré des difficultés à identifier les dolomites, Soulignant ainsi que le modèle n'a pas été entraîné à prédire cette classe, comme expliqué dans la distribution des données. Néanmoins, les performances du modèle sur le deuxième puits ont démontré son potentiel et confirmé que, grâce à une formation et à un affinement continus, il est capable de fournir des prédictions encore plus précises pour différentes lithologies.

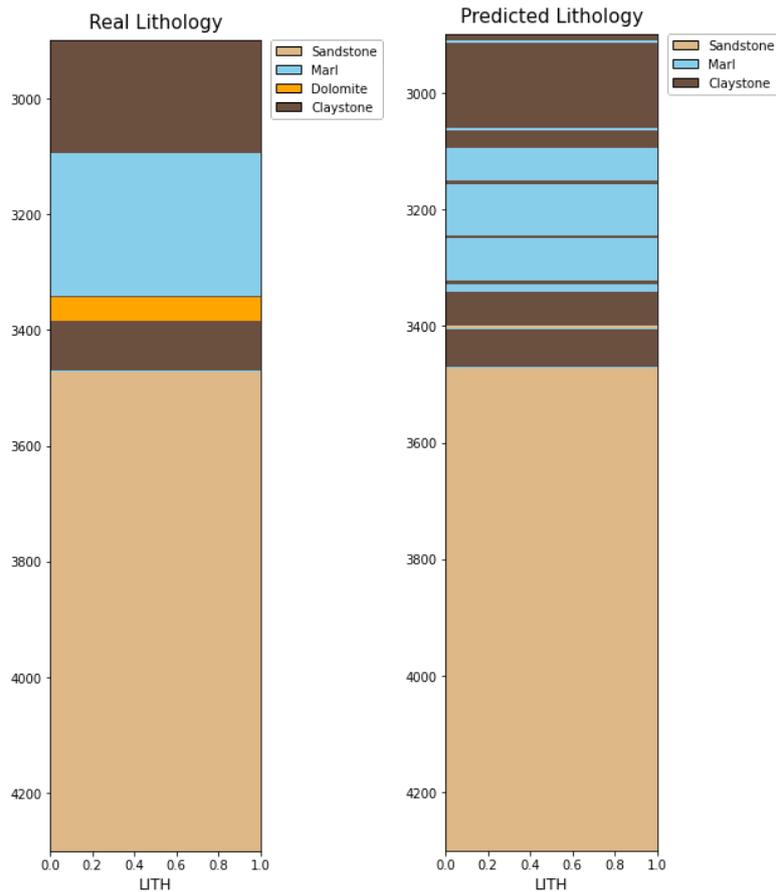


FIGURE III.21 – Comparaison entre la lithologie réelle et prédite - Puits 2

III.4.1.4 Déploiement du model

Une fois le modèle formé et évalué, la dernière étape consiste à le déployer. Cette étape consiste à intégrer le modèle dans l’environnement de production et à le rendre disponible pour utilisation. Dans notre cas, nous travaillons sur la prédiction du type de lithologie en tant que sortie, avec neuf paramètres de forage en tant qu’entrées et on souhaite le déployer sur un site Web ou une application mobile, afin que l’équipe de production puisse l’utiliser facilement avec des simples clicks. De plus, il est important d’avoir un plan de surveillance et de maintenance en place, pour garantir les performances et la précision du modèle dans le temps.

Le déploiement d’un modèle de machine learning en tant qu’application Web implique plusieurs étapes :

1. **Sérialisation du modèle** : la première étape consiste à enregistrer le modèle formé dans un format qui peut être facilement chargé et utilisé dans une application Web. Cela se fait généralement à l’aide d’une bibliothèque telle que pickle ou joblib en Python, qui peut enregistrer le modèle dans un fichier binaire qui peut être facilement chargé dans l’application Web.
2. **Web Framework** : l’étape suivante consiste à créer une application Web capable de gérer les requêtes HTTP et de renvoyer des réponses HTTP. Cela peut être fait en utilisant un framework Web tel que Flask ou Django en Python. Le framework Web vous permet de définir des itinéraires qui mappent les URL à des fonctions spécifiques qui gèrent les requêtes entrantes et renvoient la réponse appropriée.
3. **Chargement du modèle** : une fois le modèle sérialisé et l’application Web configurée, l’étape suivante consiste à charger le modèle dans l’application Web. Cela se fait généralement lors de

la phase d'initialisation de l'application, où le modèle est chargé à partir du fichier sérialisé et stocké en mémoire pour être utilisé.

4. **Point de terminaison de prédiction** : Ensuite, vous allez créer un point de terminaison dans l'application Web qui acceptera les données d'entrée et renverra les prédictions du modèle. Ce point de terminaison prend généralement des données sous la forme d'un objet JSON et renvoie également les prédictions sous la forme d'un objet JSON.
5. **Déploiement** : Enfin, l'application web peut être déployée sur un serveur web comme Apache ou nginx, ou dans les services cloud comme AWS, Google Cloud ou Microsoft Azure.

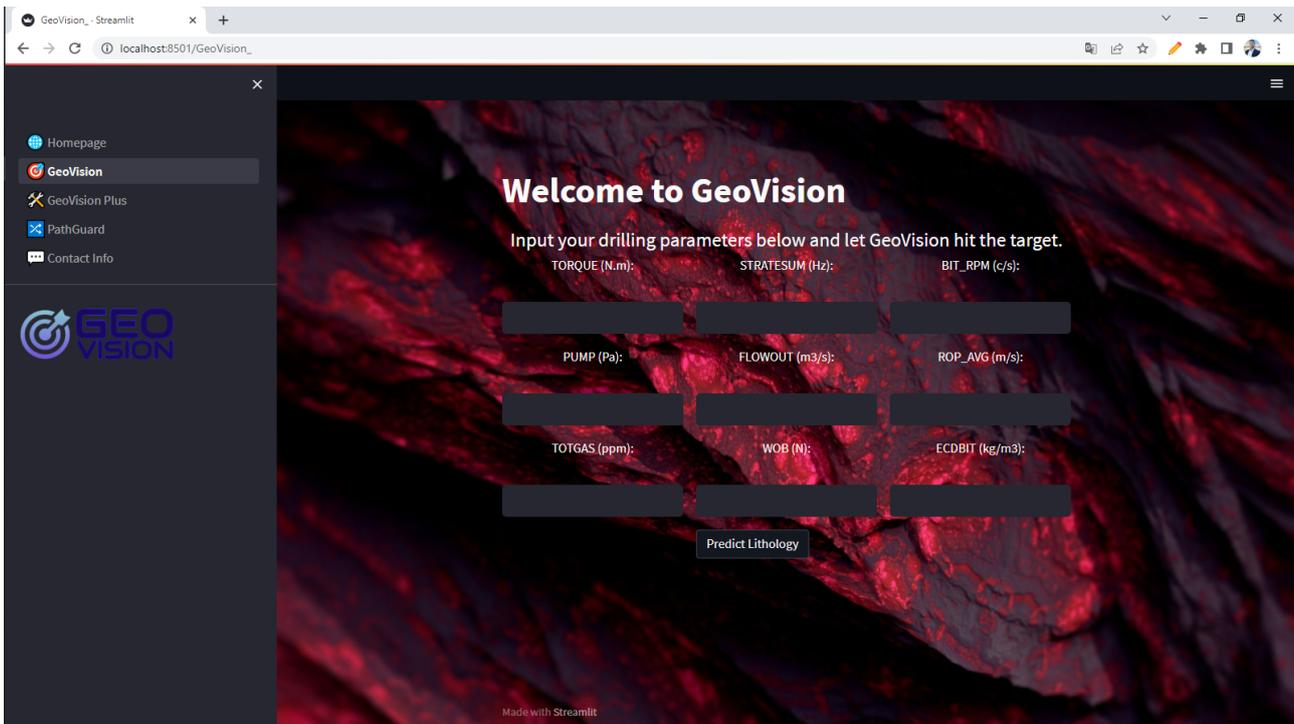
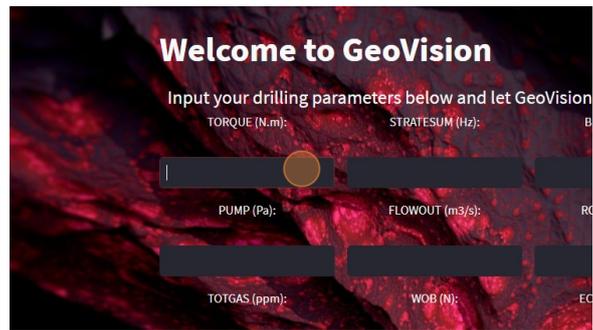


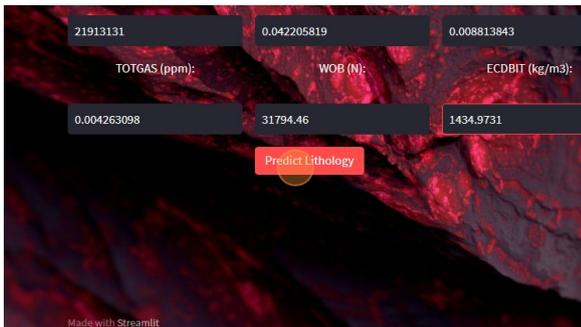
FIGURE III.22 – Geovision interface

III.4.1.4.1 Votre guide pour GeoVision web app

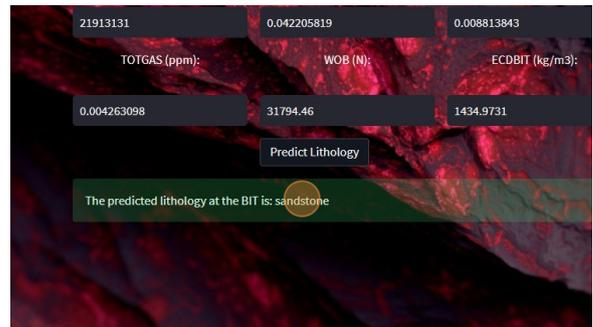
as simple as a blink of an eye (see figure III.23) :



(a) Étape 1 : Remplir les champs d'entrée



(b) Étape 2 : Cliquer sur Prédire la lithologie



(c) Étape 3 : Recevoir le résultat

FIGURE III.23 – Étapes à suivre lors de l'utilisation de l'application web GeoVision

Our achievement in predicting lithology from drilling data marks an important milestone in our ongoing exploration. As we continue our journey, we now turn our attention to PathGuard, a fascinating challenge that lies ahead. With utmost dedication and professionalism, we aim to tackle this task head-on and unveil an enthusiastic solution.

III.4.2 PathGuard

III.4.2.1 Problématique

La "déviation du puits", qui correspond à l'écart par rapport au trajet de forage prévu, est un défi majeur dans l'industrie pétrolière. Cette problématique complexe peut entraîner des coûts opérationnels plus élevés, des violations potentielles des limites légales, des risques de collision entre les puits forés et des préoccupations en matière de sécurité. Elle est influencée par divers facteurs tels que les formations géologiques, les propriétés du train de tiges et la mécanique générale du forage. Parfois planifiée pour atteindre des cibles subsurfaces spécifiques, elle nécessite un contrôle et une prévision précis.

Les outils traditionnels de mesure pendant le forage (MWD), soumis à des interférences magnétiques, peuvent entraîner des mesures inexactes. Dans ce contexte, nous proposons un modèle d'apprentissage automatique qui exploite les données des mesures gyroscopiques, qui ne sont pas affectées par les interférences magnétiques car elles suivent le vrai nord, pour établir une relation entre la tendance de la mèche et la formation en utilisant des diagraphies fiables des puits. Ce modèle est capable de prédire de manière probabiliste la déviation du trajet du puits et sert de mécanisme d'assurance qualité, identifiant les incohérences et fournissant des informations sur la qualité et la courbure du puits.

III.4.2.2 Analyse exploratoire des données

III.4.2.2.1 Données fournies

L'ensemble de données de journal de puits fourni contient des paramètres de forage qui sont :

1. **Rayon gamma** : radioactivité naturelle de la formation le long du forage.
2. **Résistivité** : capacité à entraver la circulation du courant électrique.
3. **Densité** : densité sur la longueur d'un trou de forage (densité apparente)
4. **Density_Calculated** : Cette journalisation mesure la densité calculée.
5. **Classement** : 0 non dévié de sa trajectoire présélectionnée, 1 dévié de sa trajectoire présélectionnée.

Note : Il est important de noter que les données utilisées ont été extraites d'une source ouverte sans rapport avec le domaine volvé, elles sont donc simplement utilisées pour illustrer le problème ainsi qu'une solution potentielle.

Aucune donnée manquante là-bas, donc c'est un bon début !

Forme du fichier
(301, 12)

Donné	Nombre de valeurs nan
Rayon gamma	0
Schiste_Volume	0
Réactivité	0
Delta T	0
Vc	0
contre	0
Densité	0
Densité_Calculée	0
Neuron_Porosity	0
Densité_Porosit�	0
Possions_Ratio	0
Classement	0

TABLE III.10 – Summary of File Information

III.4.2.2 Visualisations de base

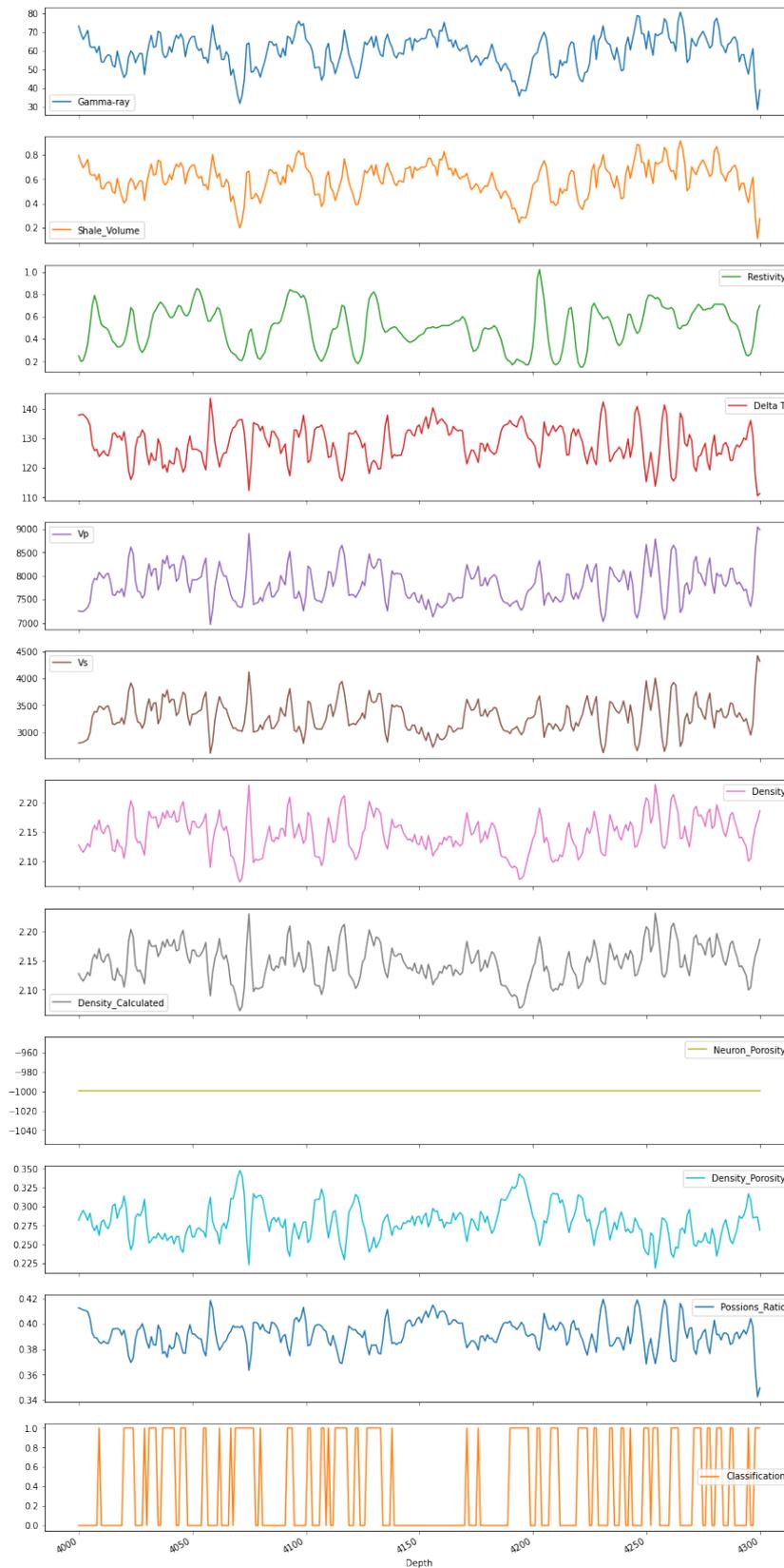


FIGURE III.24 – Visualisation des données utilisées(Statique)

Well Log Exploratory Data Analysis

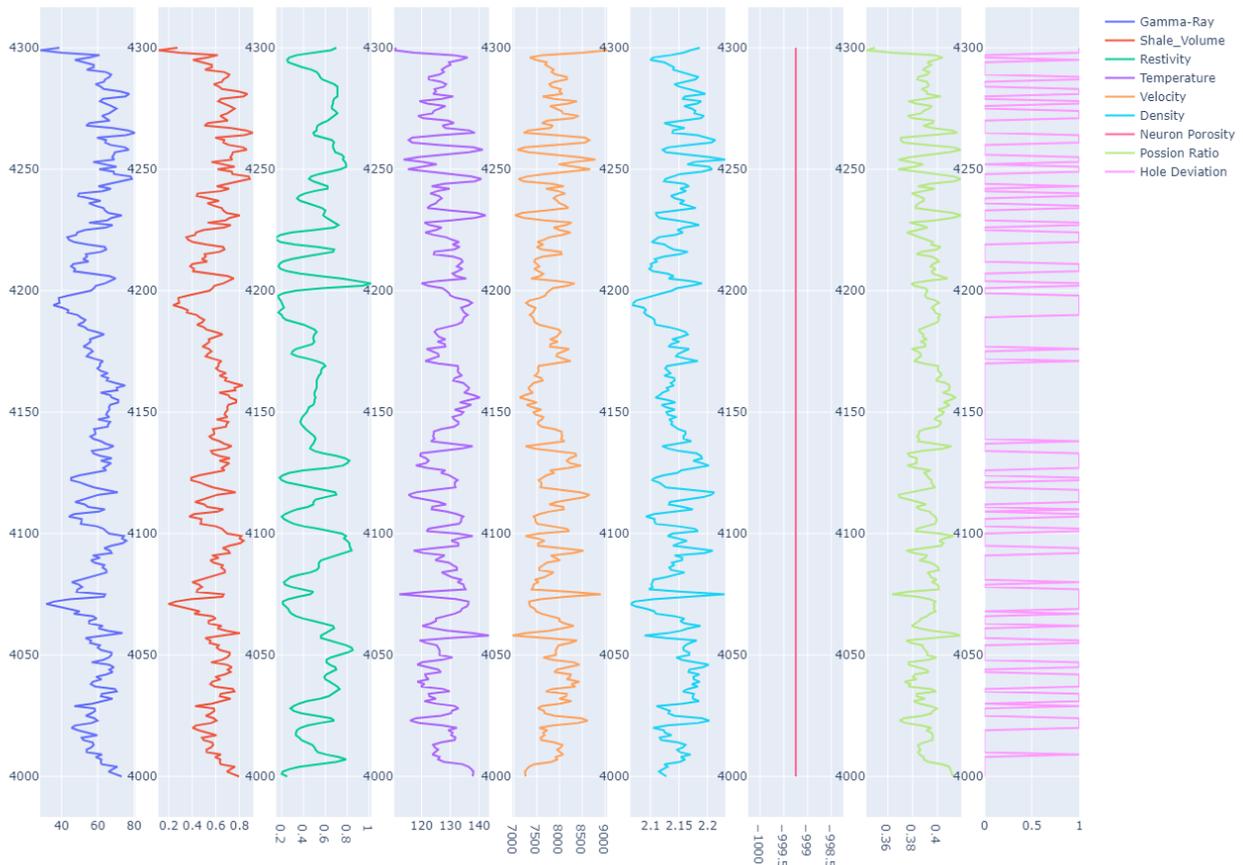


FIGURE III.25 – Visualisation interactifs en ordre horizontal paramètres des données

De nouvelles caractéristiques sont calculées en fonction des propriétés physiques et mécaniques du matériau étudié. Celles-ci comprennent :

1. Le rapport de V_p sur V_s , calculé comme suit :

$$\text{Rapport de } V_p/V_s = \frac{V_p}{V_s} \quad (\text{III.1})$$

2. La rigidité, calculée comme suit :

$$\text{Rigidité} = \text{Densité} \times V_s^2 \quad (\text{III.2})$$

3. Le module de Young, calculé à partir de la rigidité et des vitesses V_p et V_s . La formule utilisée est :

$$\text{Module de Young} = \text{Rigidité} \times \left(\frac{3 \cdot (V_p^2) - 4 \cdot (V_s^2)}{V_p^2 - V_s^2} \right) \quad (\text{III.3})$$

4. Le module volumique, calculé à partir du module de Young et du coefficient de Poisson. La formule utilisée est :

$$\text{Module volumique} = \frac{\text{Module de Young}}{3 \cdot (1 - 2 \cdot \text{Coefficient de Poisson})} \quad (\text{III.4})$$

5. Le premier paramètre de Lamé (également appelé constante de Lamé), calculé à partir du module volumique et de la rigidité. La formule utilisée est :

$$\text{Constante de Lamé} = \text{Module volumique} - \frac{2 \cdot \text{Rigidité}}{3} \quad (\text{III.5})$$

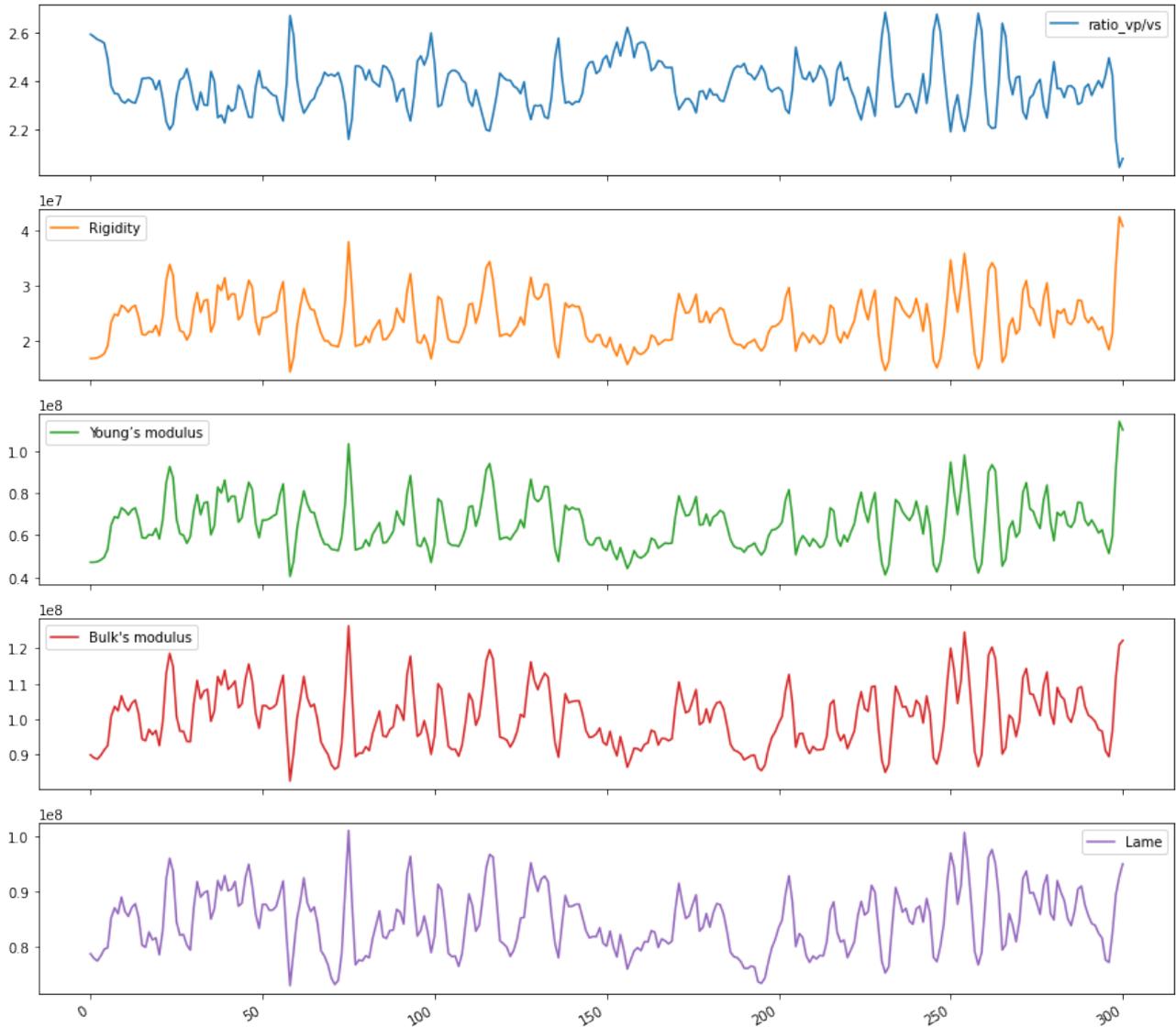


FIGURE III.26 – Tracé des paramètres mécaniques

III.4.2.2.3 corrélations

Dans cette section, l'accent est mis sur l'exploration de la corrélation entre les différentes caractéristiques de l'ensemble de données, et sur la base de ces informations, la sélection des caractéristiques à conserver pour une analyse ultérieure.

Initialement, une matrice de corrélation de Pearson est calculée pour l'ensemble de données. La corrélation de Pearson est une mesure de la corrélation linéaire entre deux variables, et donne une valeur comprise entre +1 et -1, où 1 représente une corrélation linéaire positive totale, 0 indique l'absence de corrélation linéaire et -1 représente une corrélation linéaire négative totale. Cette matrice de corrélation est ensuite visualisée à l'aide d'une carte thermique (heatmap) pour une meilleure interprétation. La carte thermique nous permet d'identifier rapidement les patterns dans la matrice de corrélation, où les couleurs plus sombres indiquent une forte corrélation et les couleurs plus claires indiquent une faible corrélation.

Après avoir identifié la matrice de corrélation, il a été observé que certaines variables étaient fortement corrélées les unes avec les autres. Une forte corrélation entre les variables indépendantes (également appelée multicollinéarité) peut poser problème dans les modèles d'apprentissage automatique car cela peut rendre difficile pour le modèle de déterminer l'effet de chaque caractéristique sur la prédiction. Pour remédier à ce problème, les variables "Shale_Volume", "Delta T", "Density_Calculated", "Vp", "Vs", "ratio_vp/vs", "Rigidity", "Bulk's modulus", "Lame" et "Density_Porosity" ont été supprimées de l'ensemble de données (voire la figure III.27 et III.28) et l'ensemble de données restant a ensuite été enregistré pour une utilisation ultérieure.

Enfin, le processus a été répété : une nouvelle matrice de corrélation a été calculée sur l'ensemble de données mis à jour et visualisée à l'aide d'une carte thermique. Cette dernière carte thermique donne un aperçu des corrélations entre les caractéristiques qui seront utilisées dans l'analyse ultérieure.

Ce processus d'évaluation et de gestion de la forte corrélation entre les caractéristiques est une étape cruciale du prétraitement des données en apprentissage automatique. Il garantit que le modèle pourra apprendre efficacement à partir des caractéristiques et conduira à des résultats plus précis et interprétables.

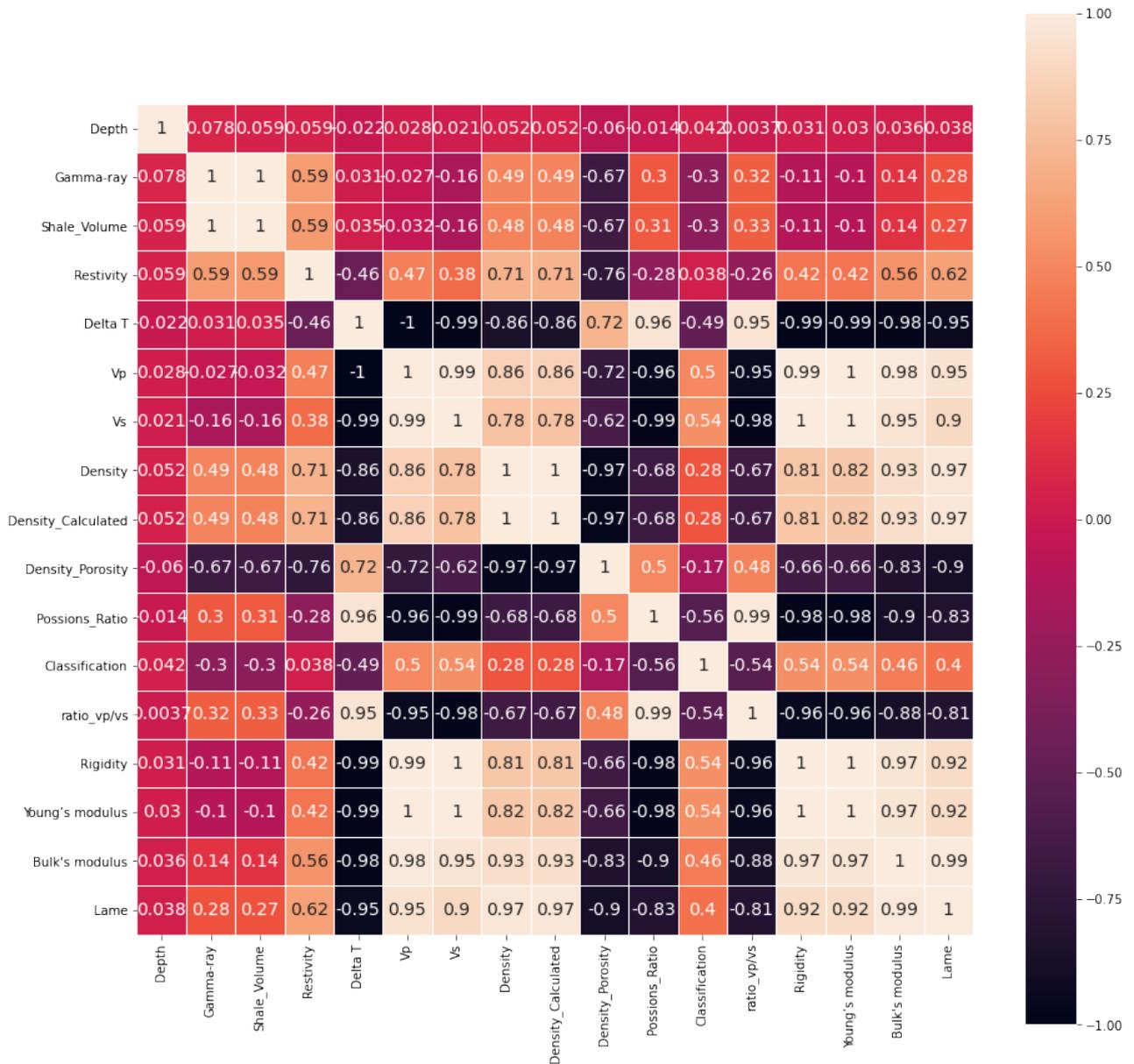


FIGURE III.27 – Plot de corrélations avant le traitement des données

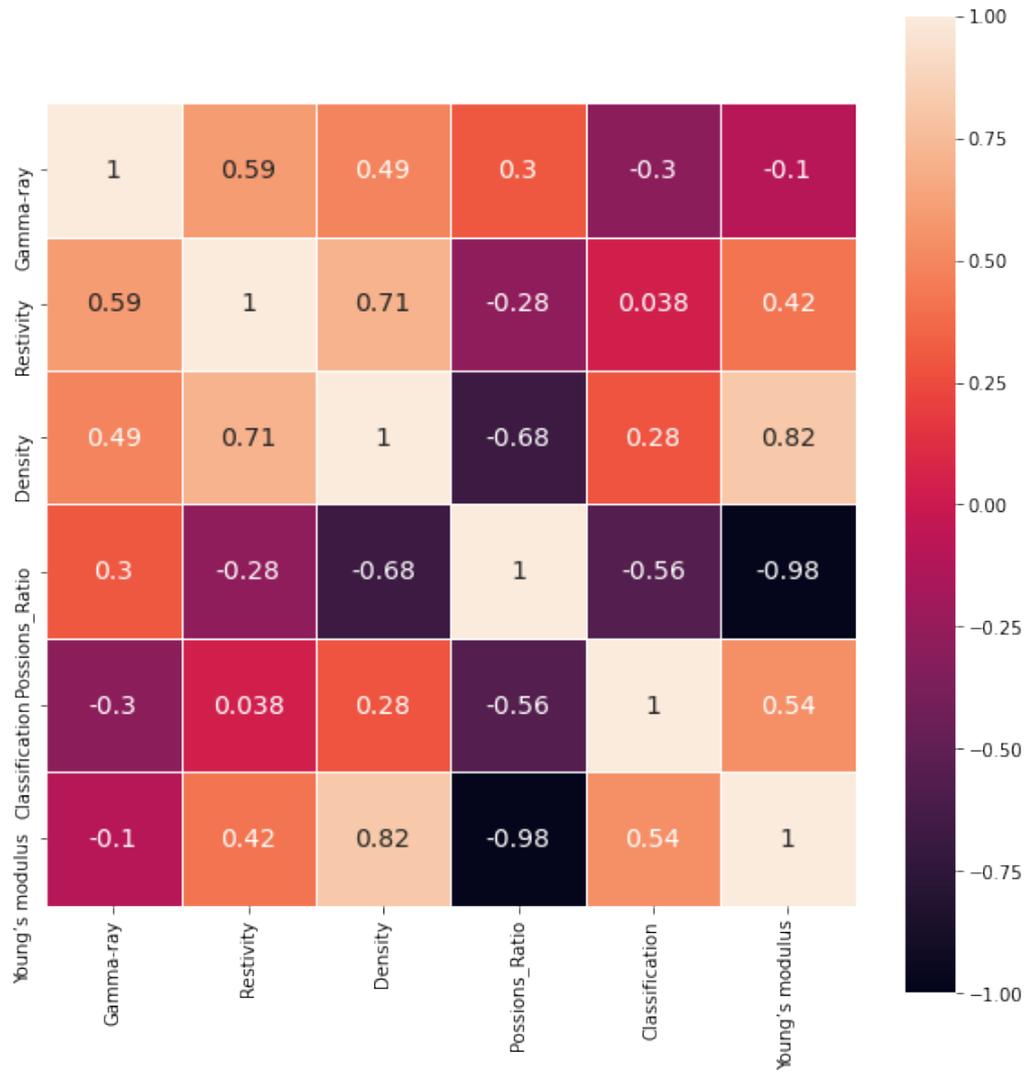


FIGURE III.28 – Heat Map après traitement des données

La matrice des diagrammes de dispersion (scatterplot matrix) permet d'examiner de manière complète la relation entre certaines caractéristiques des données de forage de puits. Chaque cellule de la matrice représente un diagramme de dispersion entre deux variables, permettant une inspection visuelle des corrélations ou des motifs potentiels. Une codification des couleurs a été appliquée pour indiquer si un puits s'est écarté de sa trajectoire préétablie (rouge) ou non (bleu) (voir figure III.33). Les motifs ou les corrélations observés dans les points rouges peuvent suggérer des relations significatives pour déterminer si un puits s'écarte de sa trajectoire prévue. Par exemple, si les points rouges se regroupent dans une certaine région du diagramme pour 'Gamma-ray' vs 'Density', cela peut indiquer une plage spécifique de ces paramètres où les puits ont plus de chances de s'écarter.

De même, les points bleus peuvent se regrouper dans différentes zones, indiquant les conditions dans lesquelles les puits sont moins susceptibles de s'écarter. Il est essentiel de rechercher des chevauchements et des séparations dans la distribution des points rouges et bleus. Des séparations claires peuvent suggérer des seuils ou des conditions limites entre les deux classes. À l'inverse, des chevauchements importants pourraient indiquer l'absence d'une règle décisive unique basée sur ces deux paramètres.

Lors de l'interprétation de ces graphiques, il est important de se rappeler que les motifs visuels peuvent parfois être trompeurs ou influencés par l'échelle et l'étendue des axes, il est donc important de compléter ces inspections visuelles par des analyses statistiques rigoureuses.

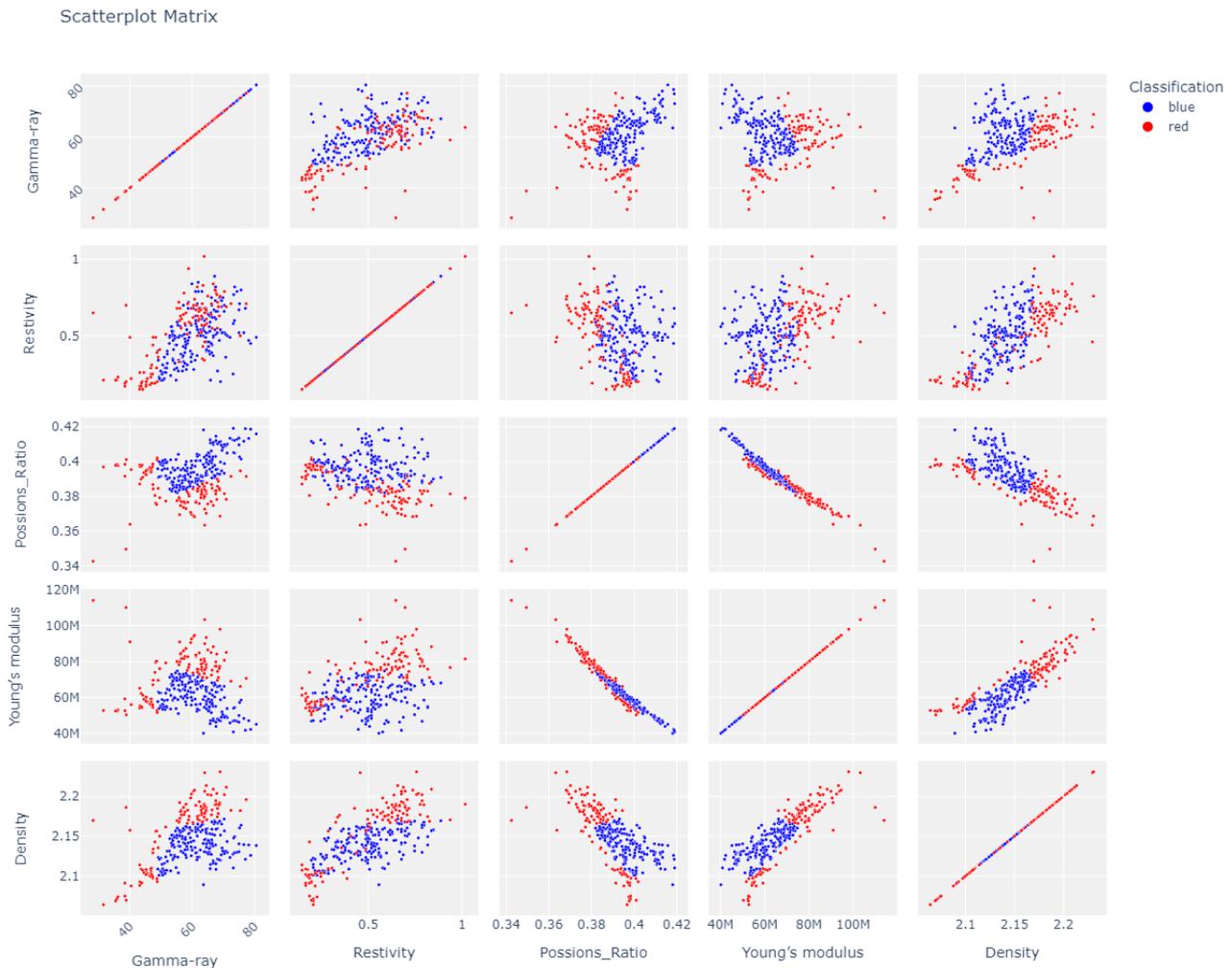


FIGURE III.29 – matrice des diagrammes de dispersion (scatterplot matrix)

Les graphiques "Résistivité vs Densité" (voir figure III.31) et "Module de Young vs Résistivité" (voir figure III.30) montrent des regroupements distincts. Les regroupements rouges représentent les

puits déviant de leur trajectoire, tandis que les regroupements bleus représentent les puits respectant leur trajectoire. Cela suggère des corrélations entre la résistivité, la densité et le module de Young, et la probabilité de déviation des puits. Cependant, il s'agit de premières observations qui nécessitent une analyse plus approfondie pour confirmation.

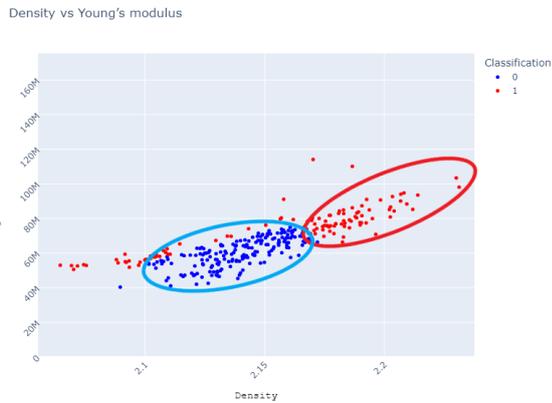


FIGURE III.30 – Module de Young vs Densité

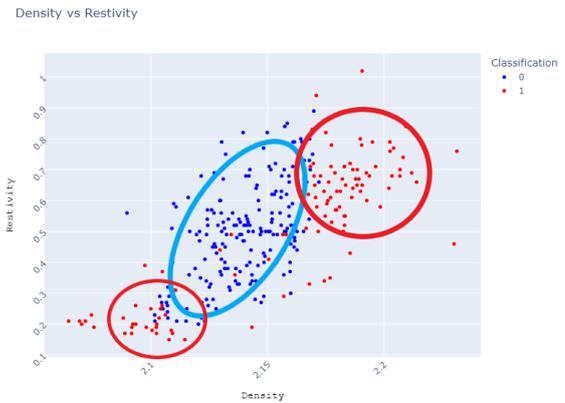


FIGURE III.31 – Densité vs Résistivité

Après avoir tracé le décompte pour chaque classe, nous avons remarqué que :

- Classe non déviée (Classification = 0) : 189
- Classe déviée (Classification = 1) : 112

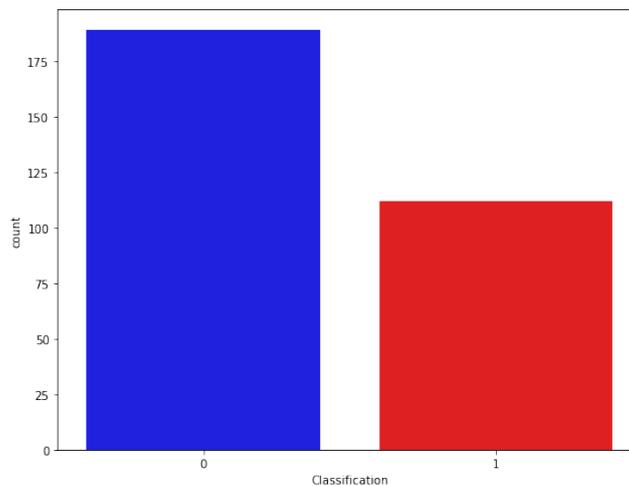


FIGURE III.32 – Le décompte pour chaque classe avant SMOTE

Ce qui signifie que les données cibles sont déséquilibrées et nous devons appliquer une méthode de sur-échantillonnage qui fonctionne en créant des échantillons synthétiques à partir de la classe minoritaire, au lieu de créer des copies comme nous l'avons mentionné précédemment dans GeoVision. L'algorithme sélectionne deux instances ou plus similaires (en utilisant une mesure de distance) et perturbe une instance un attribut à la fois en une quantité aléatoire dans la différence avec les instances voisines. Il s'agit d'une technique d'augmentation de données pour la classe minoritaire, appelée Synthetic Minority Oversampling Technique, ou SMOTE en abrégé (voir figure

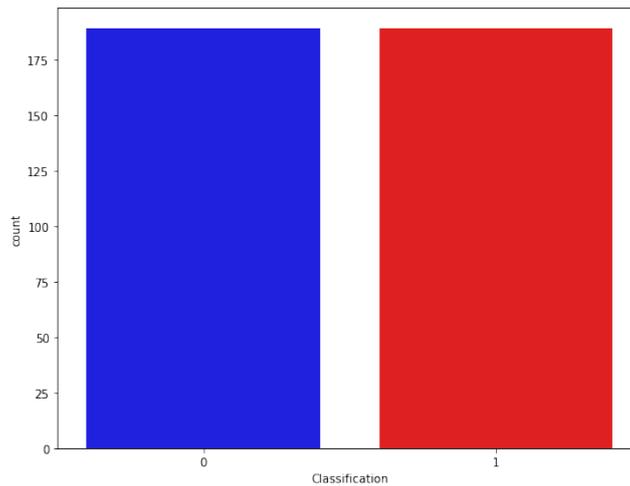


FIGURE III.33 – Le décompte pour chaque classe après SMOTE

III.4.2.3 Application de l'apprentissage automatique

Au sein du domaine de l'apprentissage automatique, nous avons minutieusement entraîné plusieurs modèles, tels que la régression logistique, les k-plus proches voisins, les machines à vecteurs de support, le classifieur de forêt aléatoire et les réseaux neuronaux. Grâce à une évaluation attentive, nous avons identifié un modèle affichant des performances supérieures, surpassant les autres. Ce modèle exceptionnel a ensuite été intégré à une application web pratique appelée "Path Guard", en utilisant le framework Streamlit.

III.4.2.3.1 Comparing models

Le tableau présente une comparaison des performances de cinq modèles différents évalués à l'aide d'ensembles de données d'entraînement et de test (voir tableau III.11). Les modèles ont été classés par ordre de performance, du meilleur au moins performant, selon leurs scores sur l'ensemble de test. Le modèle de Random Forest (RF) s'est révélé être le plus performant, avec un score de 96,22% sur l'ensemble d'entraînement et un score de 97,35% sur l'ensemble de test. Le modèle de Réseau de Neurones (NN) occupe la deuxième place avec un score de 96,23% sur l'ensemble d'entraînement et un score de 94,74% sur l'ensemble de test. Le modèle de Support Vector Machine (SVM) arrive en troisième position, avec un score de 95,85% sur l'ensemble d'entraînement et un score de 96,48% sur l'ensemble de test. Le modèle des k-plus proches voisins (knn) se classe quatrième, avec un score de 94,33% sur l'ensemble d'entraînement et un score de 93% sur l'ensemble de test. Enfin, le modèle de Régression Logistique (Log reg) présente les scores les plus bas parmi tous les modèles, avec un score de 81,05% sur l'ensemble d'entraînement et un score de 72,85% sur l'ensemble de test. Ces résultats mettent en évidence les performances relatives de ces modèles dans le traitement de l'ensemble de données spécifique et de la tâche, avec le modèle de Random Forest et le modèle de Réseau de Neurones se distinguant comme les modèles les plus performants (voire figure III.34).

TABLE III.11 – Comparison of Model Scores

Model	Score (Train)	Score (Test)
Log reg	0.810450	0.728458
knn	0.943251	0.930040
svm	0.958491	0.964822
rf	0.962192	0.973518
nn	0.962264	0.947431

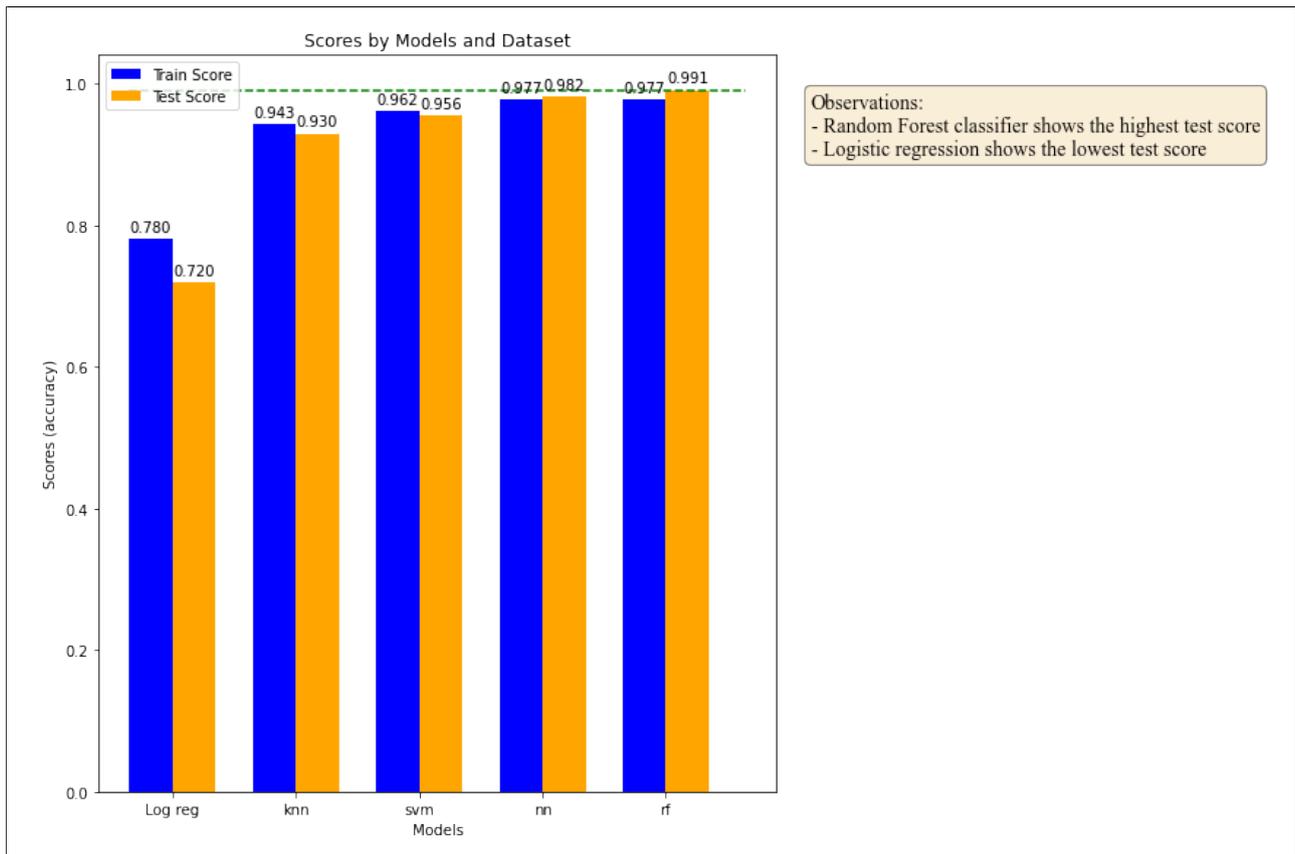
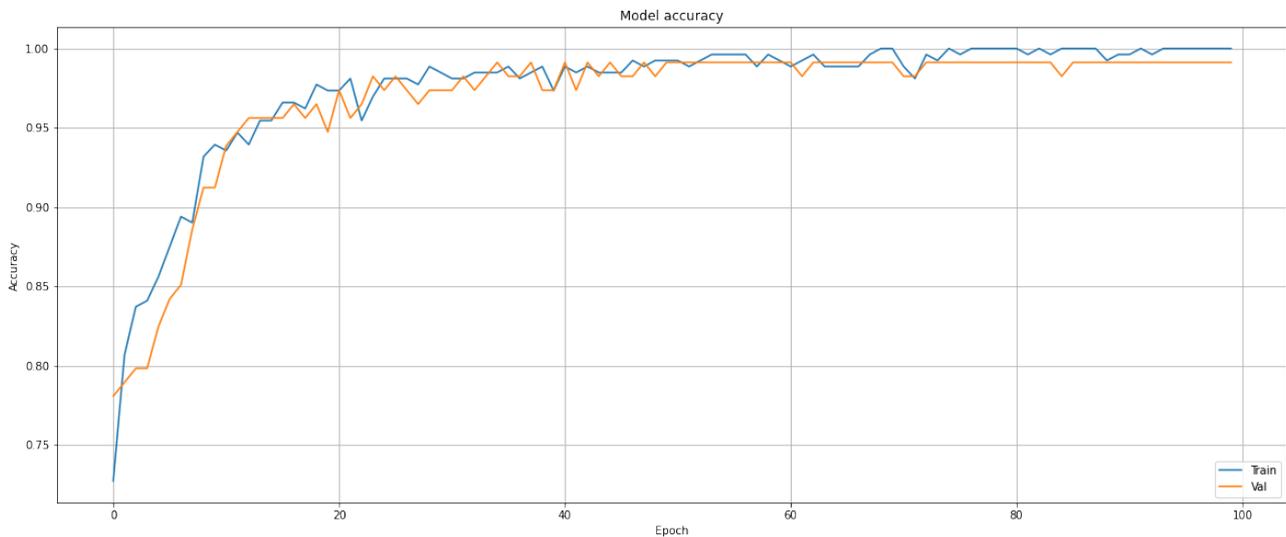


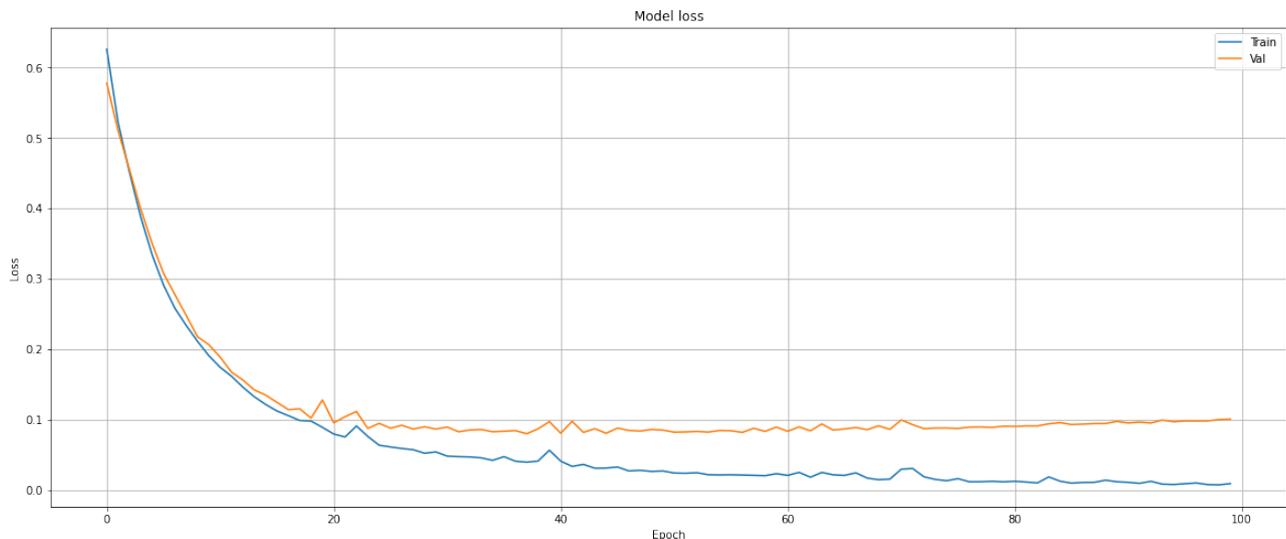
FIGURE III.34 – Comparaison des scores par modèle

III.4.2.3.2 Optimal Model Performance

Comme mentionné précédemment, les modèles Neural Networks (NN) et Random Forest (RF) Classifier se sont avérés être les meilleurs. Dans les paragraphes suivants, nous approfondissons les capacités et les informations fournies par ces modèles.



(a) Précision du modèle NN



(b) Perte du modèle NN

FIGURE III.35 – Métriques du modèle Réseau de Neurones

réseaux de neurones

Classificateur de forêt aléatoire

La matrice de confusion montre les résultats de classification du modèle RF. Elle indique que le modèle a correctement prédit 57 échantillons de la classe positive et 54 échantillons de la classe négative. Cependant, il y a eu une erreur de prédiction pour un échantillon de la classe positive et deux échantillons de la classe négative. Cette matrice permet d'évaluer la performance globale du modèle (voire figure III.36).



FIGURE III.36 – Matrice de confusion du modèle RF

Le rapport de classification révèle une performance exceptionnelle du modèle avec une précision, un rappel et un score F1 de 0,97 pour les deux classes. Cela indique que le modèle a correctement identifié environ 97% des instances pour chaque classe, tout en maintenant un équilibre entre la précision (proportion de prédictions positives correctement identifiées) et le rappel (proportion de vrais positifs correctement identifiés). L’exactitude globale du modèle est également de 97%, soulignant son efficacité générale (voire table III.12).

	Precision	Recall	F1-Score	Support
0	0.97	0.98	0.97	58
1	0.98	0.96	0.97	56
accuracy			0.97	114
macro avg	0.97	0.97	0.97	114
weighted avg	0.97	0.97	0.97	114

TABLE III.12 – Rapport de Classification

SHAP Summary Plot (SSP) et Partial Dependence Plot (PDP), le SSP est une représentation générale de l’effet des entrées sur la prédiction du modèle.

- L’emplacement vertical indique quelle caractéristique est représentée.
- La couleur indique si cette caractéristique était élevée ou faible pour cette ligne de l’ensemble de données.

L’emplacement horizontal indique si l’effet de cette valeur a entraîné une prédiction plus élevée ou plus faible. D’autre part, les PDP montrent l’impact d’une seule caractéristique sur les prédictions et sa distribution en même temps. Chaque point représente une ligne des données. L’emplacement horizontal est la valeur réelle de l’ensemble de données.

- Une pente positive indique un impact positif sur la sortie.
- Une pente négative indique un impact négatif sur la sortie.
- La couleur indique l’influence combinée avec la deuxième caractéristique sur la sortie.

Pour le graphique qui visualise l’importance des caractéristiques et leur impact sur les prédictions du modèle. Les valeurs SHAP sur l’axe des x indiquent l’impact de chaque caractéristique sur la sortie du modèle. Les valeurs positives indiquent qu’une caractéristique augmente la prédiction du modèle, tandis que les valeurs négatives indiquent qu’elle la diminue. Chaque point sur le graphique correspond à une instance dans l’ensemble de données. Les caractéristiques sont ordonnées par l’importance

globale, la plus importante en haut. Ainsi, une valeur SHAP élevée indique que la caractéristique a un fort impact sur la prédiction du modèle(voire figureIII.37).

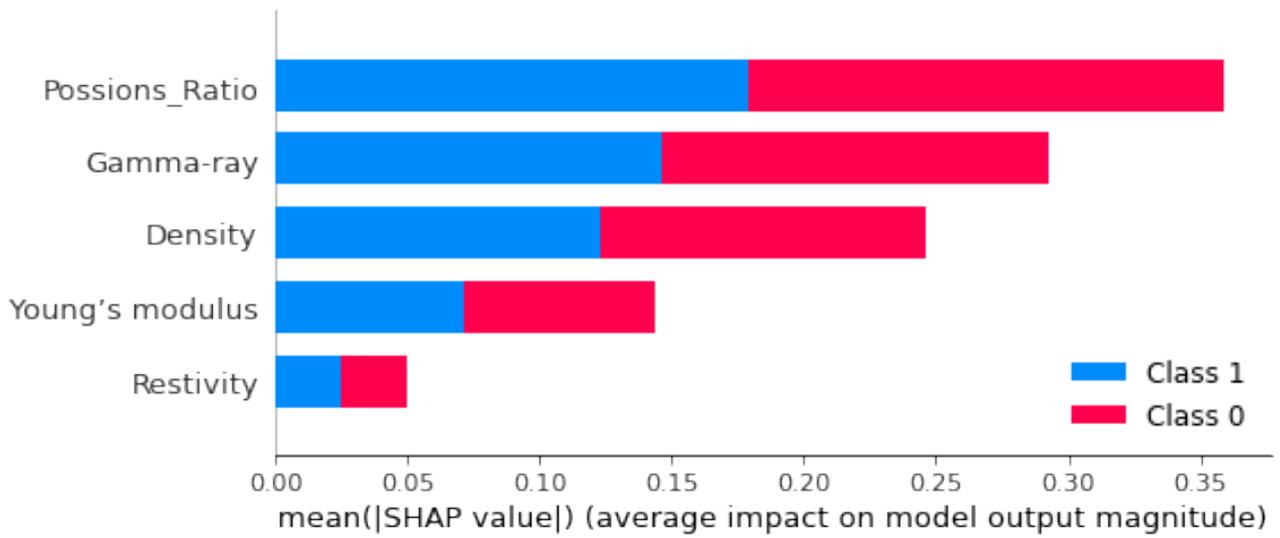


FIGURE III.37 – Importance des caractéristiques du modèle RF avec SHAP

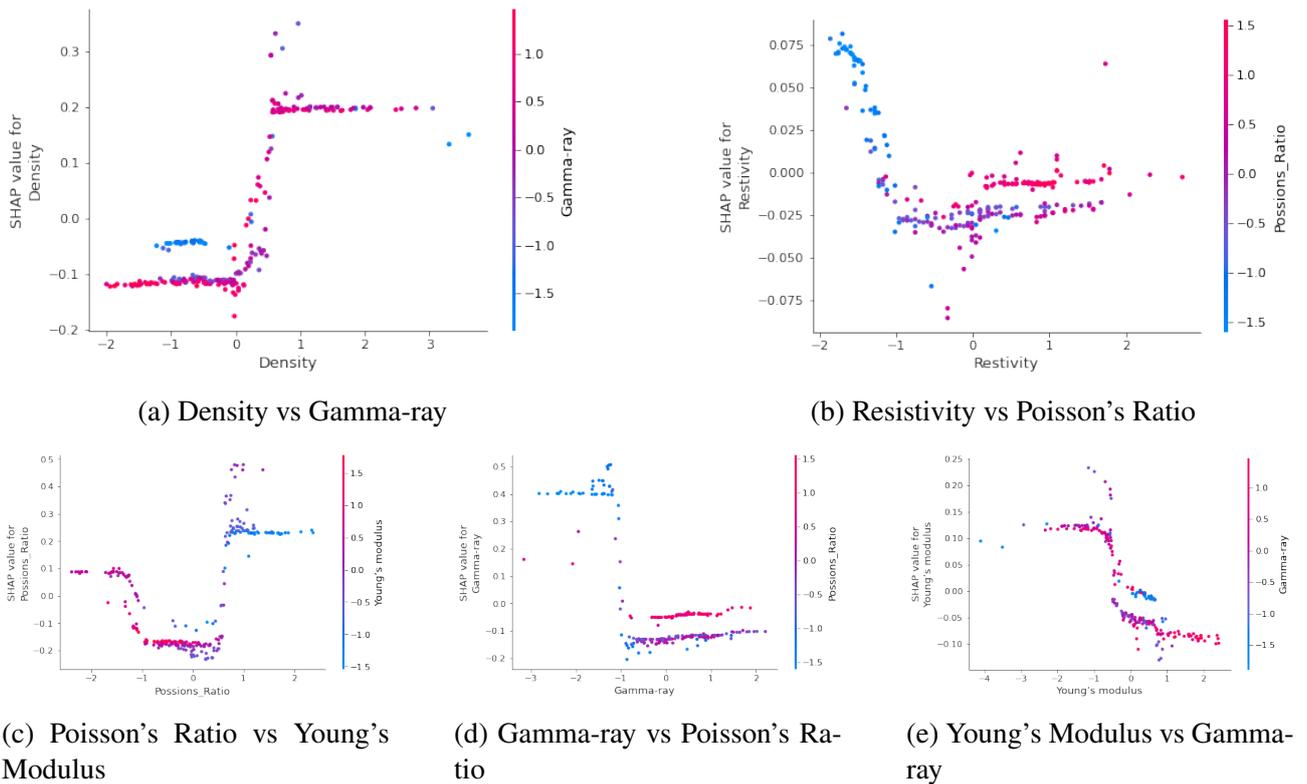


FIGURE III.38 – Partial Dependence Plots (PDP)

Le deuxième graphique (voir figure III.39), les valeurs de SHAP min/max, représente la manière dont la sortie minimale et maximale est influencée par chaque paramètre entré individuellement.

- Les valeurs des caractéristiques entraînant une augmentation des prédictions sont en rose.
- Les valeurs des caractéristiques diminuant la prédiction sont en bleu.

La taille visuelle montre l'ampleur de l'effet de la caractéristique.

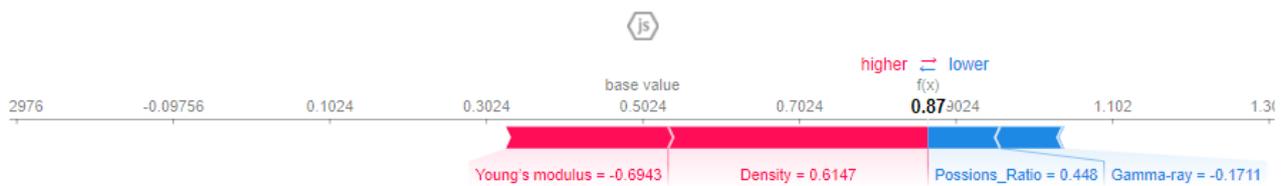


FIGURE III.39 – Valeurs SHAP respectives pour la prédictions minimale et maximal du Déviation

III.4.2.4 Déploiement du model

Après un entraînement méticuleux et l'affinage de notre modèle prédictif, nous passons maintenant à l'étape cruciale suivante : le déploiement. Notre modèle dédié à l'assurance qualité et à la prédiction probabiliste de la déviation de trajectoire des puits, sera déployé sous la forme d'une application web conviviale. Cela permettra à un large éventail d'utilisateurs d'accéder aux capacités prédictives du modèle via une interface conviviale, favorisant ainsi des décisions éclairées en matière de planification et de gestion des trajectoires des puits.

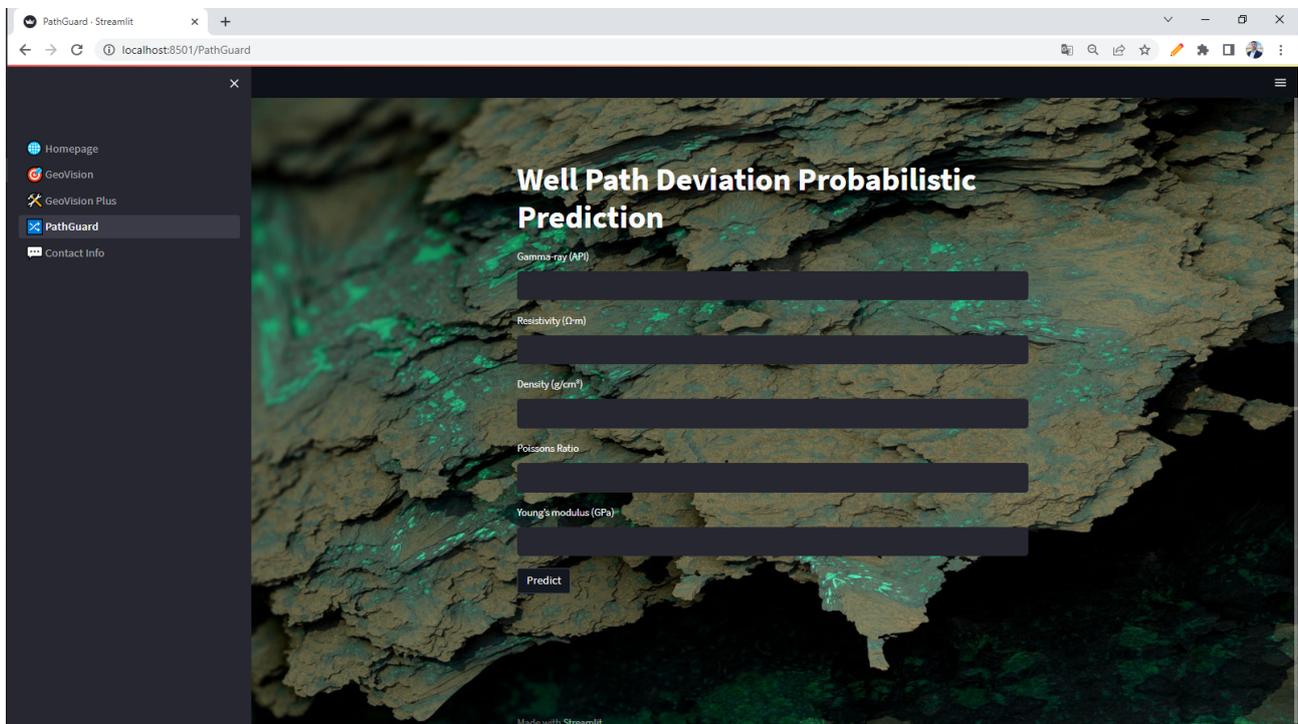
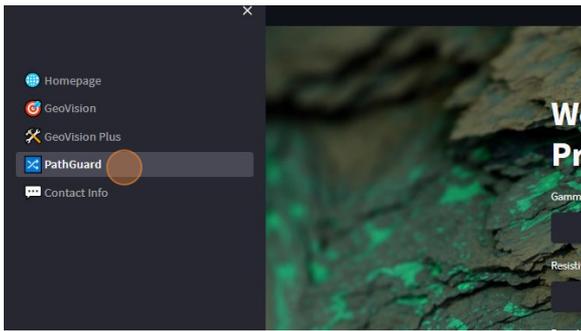


FIGURE III.40 – Interface du Web App PathGuard

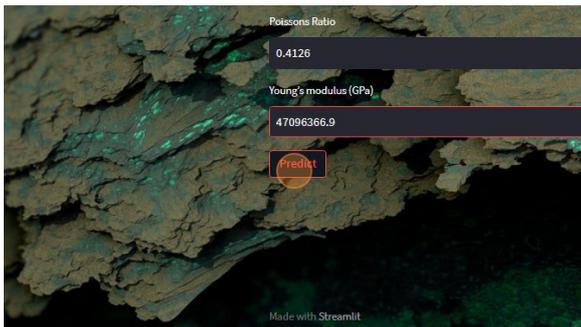
III.4.2.4.1 Votre guide pour GeoVision web app



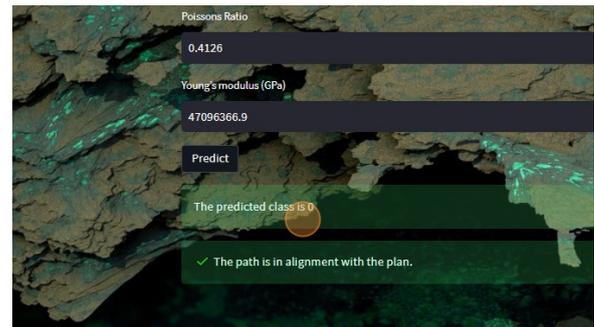
(a) Etape 1 : Sélectionnez la page "PathGuard"



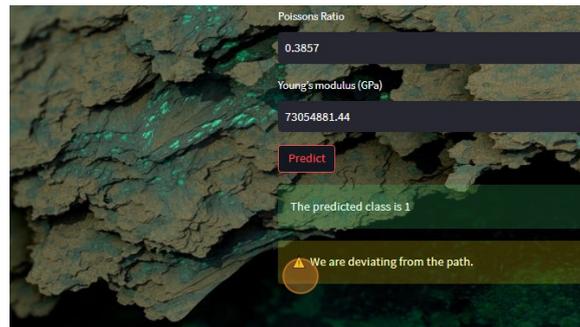
(b) Etape 2 : Remplissez tous les champs



(c) Etape 3 : Cliquez sur "Prédire"



(d) La classe prédite est 0



(e) La classe prédite est 1

FIGURE III.41 – Directives d'utilisation de l'application

III.4.3 Conclusion

En conclusion, TerraVerse présente une solution novatrice pour surmonter les défis complexes du forage de puits. L'intégration des modèles prédictifs GeoVision et PathGuard a démontré leur efficacité dans l'identification en temps réel de la lithologie et l'assurance de la qualité et la prédiction probabiliste de la déviation de trajectoire des puits. Toutefois, pour garantir sa fiabilité pour les futurs forages et élargir son application, il est essentiel de poursuivre l'entraînement du modèle sur des ensembles de données plus vastes, englobant une plus grande variété de classes de lithologie. Cette amélioration continue permettra de renforcer la précision et la robustesse du modèle, ouvrant ainsi la voie à une véritable révolution dans l'industrie du forage. TerraVerse offre ainsi le potentiel de transformer l'ingénierie du forage en rendant les opérations plus sûres, plus économiques et moins sujettes aux temps non productifs.

Conclusion générale

En conclusion, ce travail de thèse a exploré et mis en évidence l'importance et le potentiel des techniques d'apprentissage automatique dans l'ingénierie pétrolière. Au-delà des cadres conventionnels, cette étude a contribué à des avancées significatives dans deux domaines clés : la prédiction de la lithologie et des sommets de formation grâce à l'analyse des données de forage, ainsi que la gestion des déviations de trajectoire de forage en raison des interférences magnétiques.

La première partie de ce travail a donné naissance à "GeoVision", une application web intuitive qui utilise neuf paramètres de forage pour prédire avec précision la lithologie et les sommets de formation. Cette approche novatrice dépasse le focus traditionnel sur le taux de pénétration (ROP) pour fournir des informations en temps réel qui peuvent être cruciales pour le géosteering et pour maintenir le chemin de forage dans la zone payante. Cette réalisation marque un pas important dans l'application de l'intelligence artificielle aux problèmes de géologie et de forage.

La deuxième partie de cette thèse a vu l'émergence de "PathGuard", une solution innovante pour gérer les déviations de trajectoire de forage. Contrairement aux méthodes conventionnelles, notre modèle intègre de manière holistique plusieurs paramètres de diagrapie de puits géophysiques et d'autres fonctionnalités d'ingénierie, améliorant ainsi la précision de la prédiction. Le déploiement de ce modèle en tant qu'application web facile à utiliser représente un outil avancé facilement accessible pour l'industrie, qui non seulement renforce les opérations actuelles, mais aussi prépare le terrain pour des systèmes potentiels en temps réel et automatisés pour gérer les déviations de forage.

En somme, ce travail illustre le potentiel transformationnel de l'apprentissage automatique dans l'ingénierie pétrolière, avec des applications pratiques qui peuvent être mises en œuvre dès maintenant pour améliorer la précision et l'efficacité de forage. Cependant, le champ d'application de ces technologies est loin d'être épuisé et des opportunités d'exploration et d'amélioration existent encore. Les travaux futurs pourraient chercher à incorporer d'autres types de données de forage, à explorer d'autres méthodes d'apprentissage automatique, et à étendre ces techniques à d'autres domaines de l'ingénierie pétrolière.

Perspectives & Projets Futurs

Les modèles actuellement disponibles dans notre application Web sont globalement performants, mais il existe encore un potentiel d'optimisation significatif en fonction des données spécifiques que nous utilisons. Dans le secteur mondial de l'industrie pétrolière et gazière, les grandes entreprises tirent pleinement parti des technologies avancées basées sur l'intelligence artificielle, en exploitant efficacement leurs vastes ensembles de données. Ces entreprises obtiennent ainsi des résultats optimaux dans une multitude de processus. Malheureusement, en Algérie, malgré la collecte considérable de données, celles-ci ne sont pas exploitées correctement, ce qui constitue un gaspillage de ressources précieuses.

Dans nos futurs travaux, nous avons pour ambition de développer un ensemble diversifié de modèles personnalisés spécifiquement dédiés aux champs pétroliers algériens. Cette approche est à la fois nécessaire pour obtenir des résultats optimaux et pour améliorer notre image à l'échelle internationale. Elle contribuera également de manière significative à la transition vers des opérations plus durables et potentiellement plus sécurisées.

En exploitant pleinement les données disponibles et en développant des modèles adaptés aux particularités des champs algériens, nous pourrions améliorer les performances de notre solution et mieux répondre aux défis spécifiques de l'industrie pétrolière en Algérie. Cela nous permettra de maximiser l'utilisation de nos ressources, d'optimiser les opérations de forage, et de favoriser une gestion plus efficace et durable des ressources énergétiques du pays.

Bibliographie

1. Lithology. Earthquake Glossary. US Geological Survey. Retrieved (29 October 2010).
2. Bates, R. J.; Jackson, J. A., eds. (1984). Dictionary of Geological Terms (3 ed.). American Geological Institute. p. 299. ISBN 0-385-18101-9.
3. Allaby, Ailsa; Allaby, Michael (1999). Oxford Dictionary of Earth Sciences (2 ed.). Oxford University Press. p. 320. ISBN 0-19-280079-5.
4. Azar, J. J. (2006). Drilling problems and solutions. In L. W. Lake (Ed.), Petroleum Engineering Handbook : Volume II - Drilling Engineering (pp. 433-454). R. F. Mitchell (Ed.). Society of Petroleum Engineers. ISBN 978-1-55563-114-7.
5. Qi, L., & Carr, T. R. (2006). Neural network prediction of carbonate lithofacies from well logs, Big Bow and Sand Arroyo Creek fields, Southwest Kansas. *Computers & Geosciences*, 32(7), 947–964. doi :10.1016/j.cageo.2005.10.020
6. Al-Anazi, A., & Gates, I. D. (2010). A support vector machine algorithm to classify lithofacies and model permeability in heterogeneous reservoirs. *Engineering Geology*, 114(3-4), 267–277. doi :10.1016/j.enggeo.2010.05.005
7. Raeesi, M., Moradzadeh, A., Doulati Ardejani, F., & Rahimi, M. (2012). Classification and identification of hydrocarbon reservoir lithofacies and their heterogeneity using seismic attributes, logs data and artificial neural networks. *Journal of Petroleum Science and Engineering*, 82-83, 151–165. doi :10.1016/j.petrol.2012.01.012
8. Wang, G., & Carr, T. R. (2012). Methodology of organic-rich shale lithofacies identification and prediction : A case study from Marcellus Shale in the Appalachian basin. *Computers & Geosciences*, 49, 151–163. doi :10.1016/j.cageo.2012.07.011
9. Sebtosheikh, M. A., & Salehi, A. (2015). Lithology prediction by support vector classifiers using inverted seismic attributes data and petrophysical logs as a new approach and investigation of training data set size effect on its performance in a heterogeneous carbonate reservoir. *Journal of Petroleum Science and Engineering*, 134, 143–149. doi :10.1016/j.petrol.2015.08.001
10. Avanzini, A., Balossino, P., Brignoli, M., Spelta, E., & Tarchiani, C. (2016). Lithologic and geomechanical facies classification for sweet spot identification in gas shale reservoir. *Interpretation*, 4(3), SL21–SL31. doi :10.1190/int-2015-0199.1
11. Al-Mudhafar, W. J. (2017). Integrating well log interpretations for lithofacies classification and permeability modeling through advanced machine learning algorithms. *Journal of Petroleum Exploration and Production Technology*, 7(4), 1023–1033. doi :10.1007/s13202-017-0360-0
12. Gu, Y., Bao, Z., Song, X., Patil, S., & Ling, K. (2019). Complex lithology prediction using probabilistic neural network improved by continuous restricted Boltzmann machine and particle swarm optimization. *Journal of Petroleum Science and Engineering*, 179, 966–978. doi :10.1016/j.petrol.2019.05.032
13. Imamverdiyev, Y., & Sukhostat, L. (2019). Lithological facies classification using deep convolutional neural network. *Journal of Petroleum Science and Engineering*, 174, 216–228. doi :10.1016/j.petrol.2018.11.023
14. Gupta, I., Tran, N., Devegowda, D., Jayaram, V., Rai, C., Sondergeld, C., & Karami, H. (2020). Looking Ahead of the Bit Using Surface Drilling and Petrophysical Data : Machine-Learning-Based Real-Time Geosteering in Volve Field. *SPE Journal*. doi :10.2118/199882-pa
15. Lesso, W. G., Rezmer-Cooper, I. M., & Chau, M. (2001). Continuous Direction and Inclination Measurements Revolutionize Real-Time Directional Drilling Decision-Making. *SPE/IADC Drilling Conference*. doi :10.2118/67752-ms
16. Pirovolou, D., Chapman, C. D., Chau, M. T., Arismendi, H., Ahorukomeye, M., & Penaranda, J. (2011). Drilling Automation : An Automatic Trajectory Control System. *SPE Digital Energy Conference and Exhibition*. doi :10.2118/143899-ms
17. Zhang, Y., & Samuel, R. (2015). Analytical Model to Estimate the Directional Tendency of Point and Push-the-Bit BHAs. *SPE Annual Technical Conference and Exhibition*. doi :10.2118/174798-ms
18. Marck, J., & Detournay, E. (2016). Influence of Rotary-Steerable-System Design on Borehole Spiraling. *SPE Journal*, 21(01), 293–302. doi :10.2118/174554-pa

19. H. Wang, Z.-C. Guan, Y.-C. Shi, D.-Y. Liang. (November 2017). "Drilling Trajectory Prediction Model for Push-the-bit Rotary Steerable Bottom Hole Assembly." *International Journal of Engineering, Transactions B : Applications*, 30(11), 1800–1806. doi :10.5829/ije.2017.30.11b.23
20. Pollock, J., Stoecker-Sylvia, Z., Veedu, V., Panchal, N., & Elshahawi, H. (2018). Machine Learning for Improved Directional Drilling. *Offshore Technology Conference*. doi :10.4043/28633-ms
21. Koryabkin, V., Semenikhin, A., Baybolov, T., Gruzdev, A., Simonov, Y., Chebuniaev, I., Karpenko, M., & Vasilyev, V. (2019). Advanced data-driven model for drilling bit position and direction determination during well deepening. <https://doi.org/10.2118/196458-MS>
22. Saasen, A., Poedjono, B., Ånesbug, G. O., & Zachman, N. (2020). Removal of Magnetic Contamination in Drilling Fluids : Effect on Directional Drilling. <https://doi.org/10.1115/OMAE2020-18719>
23. Tunkiel, A. T., Sui, D., & Wiktorski, T. (2021). Training-while-drilling approach to inclination prediction in directional drilling utilizing recurrent neural networks. *Journal of Petroleum Science and Engineering*, 196, 108128. doi :10.1016/j.petrol.2020.108128
24. Yu, R., Diao, B., & Gao, D. (2022). Interference Analysis of Shale Gas Horizontal Wells Considering Magnetic Azimuth Correction. *56th U.S. Rock Mechanics/Geomechanics Symposium, Santa Fe, New Mexico, USA*. 10.56952/ARMA-2022-0616
25. ElGizawy, M., Lowdon, R., Aklestad, D., Strain, P., & Boyce, F. (2023). Combining Best-in-Class Surveying Measurements to Provide the Most Accurate Wellbore Position. *SPE/IADC International Drilling Conference and Exhibition*. 10.2118/212547-MS
26. Directional Drilling Training Manual December (1996) ATC-003 Version 1.1 click to access the manual
27. Chen, D. (2006). Chapter 6 - Directional Drilling. In L. W. Lake (Ed.), *Petroleum Engineering Handbook : Volume II - Drilling Engineering* (pp. 265-286). R. F. Mitchell (Ed.). Society of Petroleum Engineers. ISBN 978-1-55563-114-7.
28. Edvardsen, I. (2015). Effects of Geomagnetic Disturbances on Offshore Magnetic Directional Wellbore Positioning in the Northern Auroral Zone (Doctoral dissertation).
29. Drilling, S. and man, h. (2018). Steerable Downhole Mud Motor – Directional Drilling. [online] [Drillingknowledge.blogspot.com](http://drillingknowledge.blogspot.com). Available at : <http://drillingknowledge.blogspot.com/2017/11/steerable-downhole-mud-motor.html>
30. Price, D. (n.d.). Rig Worker : Drilling Services. Retrieved from <https://www.rigworker.com/services-2/jetting.html>.
31. Geosteering - Schlumberger Oilfield Glossary. (2017). Geosteering. Retrieved from www.glossary.oilfield.slb.com.
32. Schlumberger. (2022). GeoSphere 360 : 3D reservoir mapping-while-drilling service [Brochure]. Retrieved from slb.com/GeoSphere360
33. Hynes, N.J. (1991). *Dictionary of Petroleum Exploration, Drilling, and Production*. Tulsa, Oklahoma : PennWell.
34. Burke, J.A., Campbell Jr., R.L., and Schmidt, A.W. (1969). The Litho-Porosity Cross Plot a Method of Determining Rock Characteristics for Computation of Log Data. Presented at the SPE Illinois Basin Regional Meeting, Evansville, Indiana, 30-31 October. SPE-2771-MS. <http://dx.doi.org/10.2118/2771-MS>
35. Fertl, W.H. (1981). Openhole Crossplot Concepts A Powerful Technique in Well Log Analysis. *J Pet Technol* 33 (3) : 535-549. SPE-8115-PA. <http://dx.doi.org/10.2118/8115-PA>
36. *A Catalog of Petrophysical and Geological Properties of Typical Reservoir Rocks*. (1995). Houston : Shell Oil Co.
37. Pickett, G.R. (1963). Acoustic Character Logs and Their Applications in Formation Evaluation. *J Pet Technol* 15 (6) : 659-667. SPE-452-PA. <http://dx.doi.org/10.2118/452-PA>.
38. Nations, J.F. (1974). Lithology and Porosity from Acoustic Shear and Compressional Wave Transit Time Relationships. *The Log Analyst* 15 (6) : 3–8.
39. Eastwood, R.L. and Castagna, J.P. (1987). Interpretation of V_p/V_s Ratios from Sonic Logs. Shear-Wave Exploration, S.H. Danbom and S.N. Domenico eds., 139-153. Tulsa, Oklahoma : SEG, Geophysical Development Series No. 1.
40. Zhang, T., Tang, X.M., and Patterson, D. (2000). Evaluation of Laminated Thin Beds in Formations Using High-Resolution Acoustic Slowness Logs, paper XX. Trans., 2000 Annual Logging Symposium, SPWLA, 1–14.
41. Sneider, R.M. and Erickson, J.W. (1994). Rock Types, Depositional History, and Diagenetic Effects, Ivishak Reservoir, Prudhoe Bay Field. *SPE Res Eng* 12 (1) : 23–30. SPE-28575-PA. <http://dx.doi.org/10.2118/28575-PA>
42. Holstein, E.D. and Warner, J., H. R. (1994). Overview of Water Saturation Determination For the Ivishak (Sadlerochit) Reservoir, Prudhoe Bay Field. Presented at the SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana, 25-28 September 1994. SPE-28573-MS. <http://dx.doi.org/10.2118/28573-MS>

43. Rashidi, B., Hareland, G., Nygaard, R. (2008). Real-Time Drill Bit Wear Prediction by Combining Rock Energy and Drilling Strength Concepts. Abu Dhabi International Petroleum Exhibition and Conference. doi :10.2118/117109-ms
44. Yang, H., Zhao, H., & Kottapurath, S. (2019). Real-time bit wear prediction using mud logger data with mathematical approaches. *Journal of Petroleum Exploration and Production Technology*, 10(2), 587–594. doi :10.1007/s13202-019-00761-y
45. Geñron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras and TensorFlow : concepts, tools, and techniques to build intelligent systems (2nd ed.)*. O'Reilly
46. P. Behnoud far, P. Hosseini, Estimation of lost circulation amount occurs during under balanced drilling using drilling data and neural network, *Egypt. J. Petrol.* 26 (3) (2017)
47. M. Sabah, et al., Application of decision tree, artificial neural networks, and adaptive neuro-fuzzy inference system on predicting lost circulation : a case study from Marun oil field, *J. Petrol. Sci. Eng.* 177 (2019)
48. J. Ning, et al., A new model of ROP prediction for drilling engineering with data mining technology, *Adv. Inf. Sci. Serv. Sci.* 5 (7) (2013).
49. M. Bataee, S. Irawan, M. Kamyab, Artificial neural network model for prediction of drilling rate of penetration and optimization of parameters, *J. Jpn. Petrol. Inst.* 57 (2) (2014).
50. C. Soares, K. Gray, Real-time predictive capabilities of analytical and machine learning rate of penetration (ROP) models, *J. Petrol. Sci. Eng.* 172 (2019)
51. R. Rooki, F.D. Ardejani, A. Moradzadeh, Hole cleaning prediction in foam drilling using artificial neural network and multiple linear regression, *Geomaterials* 4 (01)(2014).
52. Badr Eddine, B. (2021). *Conception et réalisation de la solution digital Deep Drill* [Unpublished master's thesis]. University of Boumerdès
53. R. Rooki, M. Rakhshkhorshid, Cuttings transport modeling in underbalanced oil drilling operation using radial basis neural network, *Egypt. J. Petrol.* 26 (2) (2017)

Appendix A

Technologies utilisées

A.1 Python

Python ^a est un langage de programmation informatique souvent utilisé pour créer des sites Web et des logiciels, automatiser des tâches et effectuer des analyses de données. Python est un langage à usage général, ce qui signifie qu'il peut être utilisé pour créer une variété de programmes différents et qu'il n'est pas spécialisé pour des problèmes spécifiques. Cette polyvalence en ont fait l'un des langages de programmation les plus utilisés aujourd'hui.

^a<https://www.python.org/>

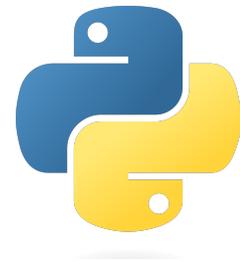


Figure A.1: Python

A.2 Spyder

Spyder ^a is an integrated development environment (IDE) specifically designed for scientific computing and data analysis using Python. It provides a powerful and user-friendly interface that allows users to write, test, and debug Python code efficiently. Spyder offers features such as a code editor with syntax highlighting, variable explorer, debugger, and a built-in console that supports interactive Python sessions. With its comprehensive set of tools and libraries, Spyder is an excellent choice for scientists, researchers, and data analysts who work with Python for data exploration, numerical computations, and scientific visualization.

^a<https://www.spyder-ide.org/>



Figure A.2: Spyder

A.3 Streamlit

Streamlit ^a is a Python library that enables rapid development and deployment of interactive web applications for data science and machine learning projects. With Streamlit, developers can easily create dynamic and interactive user interfaces using simple Python scripts. It provides a streamlined way to visualize data, generate plots, and build interactive components such as sliders, dropdown menus, and checkboxes. Streamlit also supports live updates, allowing users to see the results of their code changes in real-time. By simplifying the process of creating web applications, Streamlit empowers data scientists and machine learning practitioners to quickly share their work, demonstrate prototypes, and build interactive dashboards without extensive web development knowledge.

^a<https://streamlit.io/>

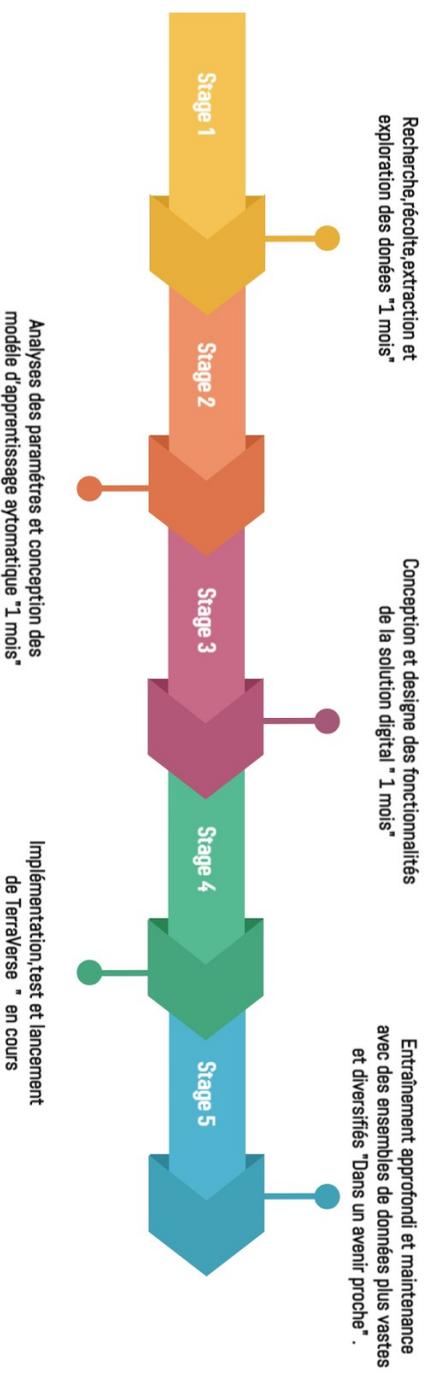


Figure A.3:
Streamlit

Appendix B

Chronologie du projet

Lo Chronologie De La Conception du Solution Digital TerraVerse



Appendix C

Comment extraire les données du champ

C.1 Téléchargement les données Volve

L'ensemble de données sur les champs pétrolifères de Volve est officiellement disponible sur le site Web d'Equinor. L'ensemble de données peut être téléchargé à l'aide de l'application Microsoft Azure Storage. Pour évaluer cet ensemble de données, veuillez suivre la procédure ci-dessous :

C.1.1 Connexion au portail

1. Allez sur <https://www.equinor.com/en/how-and-why/digitalisation-in-our-dna/volve-field-data-village.html>. Faites défiler la page vers le bas et cliquez sur le bouton "Go to the Volve Dataset, data.equinor.com".
2. Choisissez votre option pertinente : Employé(e) Equinor ou Log In B2C ? Sélectionnez B2C, si vous n'êtes pas un Equinor.
3. Pour les nouveaux utilisateurs, sélectionnez Sign up now dans la boîte de dialogue. Sinon, passez à l'étape 6.
4. Remplissez le formulaire : Courriel, mot de passe, prénom, nom de famille et entreprise.
5. Cliquez sur Envoyer le code de vérification. Vérifiez votre boîte aux lettres électronique et entrez le code de vérification. Cliquez sur Créer.
6. Saisissez votre adresse électronique et votre mot de passe. Cliquez sur Sign in.
7. Sur la page "Trouver des données ouvertes". Cliquez sur l'onglet Volve Data Village.
8. Sur la page "disclaimer", acceptez les conditions pour avoir accès à l'URI de la signature Azure.

C.2 Configuration d'Azure Storage Explorer

1. Téléchargez et installez Microsoft Azure Storage Explorer. Ouvrez l'application Microsoft Azure Storage Explorer.

-
2. Développez l'onglet Local & Attaché. Cliquez avec le bouton droit de la souris sur Comptes de stockage et sélectionnez Connecter à Azure Storage.
 3. Sélectionnez l'option Utiliser une signature d'accès partagé (SAS) URI. Cliquez sur Suivant.
 4. Saisissez l'URI. Cliquez sur Suivant.
 5. Cliquez sur Connecter. Les données doivent être disponibles dans le dossier Comptes de stockage/Conteneurs attachés/Conteneurs globaux/.
 6. Sélectionnez un dossier de données. Faites un clic droit et choisissez Télécharger.