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

N° Série : ………/2024

Université Kasdi Merbah Ouargla

Faculty of Hydrocarbons, Renewable Energies and Sciences of Earth and the Universe

Production Department

A DISSERTATION

To obtain the Master's degree

Option: Professional Production

Submitted by:

SAKER Ibtissem

-THEME-

Predictive Electrical Submersible Pumps ESPs Performance for better production sustainability using Machine Learning

Defended on: 09 June 2024

Jury:

President:	GHALI Ahmed	Univ. Ouargla
Supervisor:	CHETTI Djamel Eddine	Univ. Ouargla
Examiner:	HADJADJ Sadok	Univ. Ouargla

DEDICATION

This dissertation is dedicated to the spirit of my father, may Allah bless him, to my mother, for her endless love, support, and encouragement.

ACKNOWLEDGMENTS

In the name of Allah, the Most Gracious, the Most Merciful

First, I would like to thank Almighty ALLAH for his support from the beginning till the time of publishing the dissertation.

Second, I would like to express my personal appreciation to the following people and organizations for their support and assistance during my studies:

- Prof. CHETTI Djamel Eddine (Ouargla University): My dissertation supervisor and mentor, I would like to express my deepest gratitude to him. His insightful feedback has significantly contributed to the refinement and quality of this work. His unwavering support, encouragement, and commitment to our intellectual growth have been invaluable. I am profoundly grateful for his guidance and mentorship throughout this research journey.
- Dr. REZKI Djamil (ENSP Group): I thank him for his input, assistance, unceasing guidance, and the invaluable help of the development of this dissertation.
- Mr BENZAOUI Khaled (Geoscience Advisor - Halliburton): I'm deeply grateful to him for not only providing me the initial spark of inspiration but also for his support and encouragement.
- ENSP Group & SONATRACH: For their resources and support.

Finally, I appreciate all my committee members for the time and effort devoted to my dissertation. Their valuable comments on the dissertation made my work more focused.

Abstract:

The electrical submersible pump (ESP) is considered one of the most important and rapidly growing types of artificial-lift pumping technologies. Utilized in 15–20 percent of oil wells globally, ESPs offer an effective solution at high production volumes and great depths. The performance of ESPs often gradually declines and can eventually lead to service interruptions due to factors such as high gas volumes, elevated temperatures, and corrosion. The failure of an ESP results in a significant financial impact due to lost production and the costs of replacement.

Electrical Submersible Pump (ESP) failures are unexpected. To avoid excessive downtime, early failures identification is essential. This study suggests an innovative strategy that use a comprehensive dataset and various machine learning algorithms to achieve this goal.

The Machine Learning models are based on the data collected from surface and downhole ESP monitoring equipment from four (04) wells, Several ML models are tested and evaluated using the K-Nearest Neighbour (KNN), Random Forest (RF), Decision Tree (DT), etc.

This approach is essential for helping operators to move from reactive to proactive and predictive maintenance.

Keyword: ESP, ML, KNN, RF, DT.

الملخص:

تعتبر المضخة الغاطسة الكهربائية (ESP) واحدة من أهم تقنيات ضخ النفط بالرفع الصناعي وأسرعها نموًا. تمثل من 15 إلى 20 في المئة من آبار النفط في جميع أنحاء العالم. تقدم المضخات الغاطسة الكهربائية (ESP) حلاً فعالاً عند أحجام الإنتاج العالية والأعماق الكبيرة. غالبًا ما يتدهور أداء المضخات الغاطسة الكهربائية (ESP) تدريجيًا إلى حد التوقف بسبب عوامل مختلفة مثل كميات الغاز العالية، وارتفاع درجات الحرارة، والتآكل. حيث يؤدي فشل المضخة الغاطسة الكهربائية (ESP) إلى أثر مالي كبير نتيجة لتوقف الإنتاج وتكاليف استبدالها.

يعد فشل المضخة الكهربائية الغاطسة حدثا غير متوقع. ولذلك، يُعتبر اكتشاف هذه الأعطال مبكرا أمرًا ضروريًا لتجنب فترات التوقف الطويلة. تقترح هذه الدراسة استراتيجية مبتكرة تستخدم مجموعة بيانات شاملة ومختلف خوارزميات التعلم الآلي لتحقيق هذا الهدف.

تعتمد نماذج التعلم الآلي على البيانات الميدانية التي تم جمعها من معدات مراقبة المضخات الكهربائية الغاطسة من سطح وداخل أربعة (04) آبار. تم اختبار وتقييم العديد من نماذج التعلم الآلي باستخدام مؤشرات أداء مثل الجار الأقرب(KNN) ، والغابة العشوائية (RF)، وشجرة القرار(DT) ، وغيرها.

وهو أمر بالغ الأهمية في مساعدة المشغلين على الانتقال من الصيانة التفاعلية إلى الصيانة الاستباقية والتنبؤية.

الكلمات المفتاحية: مضخات الغاطسة الكهربائية (ESP) التعلم الآلي(ML) ، الجار الأقرب(KNN) ، الغابة العشوائية(RF) ، شجرة القرار.(DT)

Résumé :

La pompe submersible électrique (ESP) est l'une des technologies de pompage à élévation artificiel les plus importantes et les plus croissantes. Utilisés dans 15 à 20 % des champs pétroliers à travers le monde, Les ESP offrent une solution efficace pour les volumes de production élevés et à de grandes profondeurs. Les performances des ESP diminuent souvent progressivement et peuvent finalement entraîner des interruptions de service en raison de facteurs tels que des gros volumes de gaz, des températures élevées et la corrosion. La défaillance d'un ESP entraîne un impact financier significatif en raison de la perte de production et des coûts de son remplacement.

Les pannes de l’ESP se produisent dans les puits d’une manière imprévue. L'identification de ces pannes est essentielle pour éviter les temps d'arrêt excessifs. Cette étude propose une stratégie innovante qui utilise un ensemble de données et différents algorithmes d'apprentissage automatique pour atteindre cet objectif.

Les modèles d’apprentissage automatique (Machine Learning) sont basés sur des données collectées à partir des équipements de surveillance ESP en surface et en fond de quatre (04) puits. Plusieurs modèles de Machine Learning sont testés et évalués en utilisant des outils tels que L'algorithme K plus proches voisins (KNN), Forêt Aléatoire (RF), l’Arbre de Décision (DT), etc.
Cette approche est essentielle pour aider les opérateurs à passer d'une maintenance réactive à une maintenance proactive et prédictive.

Mots-clés : Pompes électriques submersibles (ESP), Machine Learning (ML), K plus proches voisins (KNN), Forêt Aléatoire (RF), l’Arbre de Décision (DT)

Table of Contents

List of Figures

List of Tables

List of abbreviation

AI	Artificial Intelligence
AL	Artificial Lift
AUC	Area under the ROC Curve
AutoML	Automatic Machine Learning
BPD	Bbl. per day
CAGR	Compound annual growth rate, %
DHMs	Downhole Measurements
DT	Decision Tree
ESP	Electrical Submersible Pump
GLM	Generalized Linear Models
GOR	Gas Oil Ratio
KDD	Knowledge Discovery in Databases
KNIME	Konstanz Information Miner
KNN	K-Nearest Neighbor
ML	Machine Learning
PCA	Principal Component Analysis
PCP	Progressing Cavity Pump
PdM	Predictive Maintenance
RF	Random Forest
ROC	Receiver Operating Characteristic
SCADA	Supervisory Control and Data Acquisition
SMOTE	Synthetic Minority Oversampling Technique
SVM	Support Vector Machines
TBF	Time Before Failure, days
TVD	True Vertical Depth
VSD	Variable Speed Drives

General Introduction:

Recently, the trends of automation, Artificial Intelligence (A.I.), and Machine Learning have gained momentum. Additionally, oil field is considered as new opportunity for increased production optimization **[1]**. The important question that needs to be answered is how to apply these technologies in a way that will control all known risks, and have a noticeable influence on the operation's profitability. They must also need to be applicable to planned and specified production optimization goals **[2]**. This research, in general, focuses on using of machine learning in order to comprehend equipment state in order to enable predictive maintenance and prevent downtime.

The electrical submersible pump (ESP) is one of the most widely used in artificial lift technologies. Timely interventions are necessary to ensure constant fluid supply when potential difficulties occur **[3]**.

ESP failures are frequent in the oil industry, and maintaining an ESP is a very costly problem for the oil operating company. A failure happens when key parameters start deviating from safe operating thresholds and an ESP stop working forever **[3]**.

Ammeter charts are the most popular ESP diagnostic tool in recent years **[4]**. This monitoring method measures the motor current and records it as a function of time on a circular chart. Since the change in power consumption is primarily related to voltage changes, fluid density, and flow rate, ammeter charts are limited in their ability to identify ESP problems. Nodal analysis has been used recently to simulate ESP systems and identify common ESP failures like broken shafts, blockages at pump intakes, and so on **[5]**. The oil and gas industry has been employing statistical modelling and machine learning techniques to detect patterns that allow for production optimization, failure prevention, and real-time event detection.

To have a better knowledge of ESP behaviour, many improvements has been made on ESP sensors, supervisory control, and data acquisition systems and surface remote terminal units over the past few decades. With the fast development of the digital oil field, more and more machine learning and data-driven models are applied to perform fault detection and predict impending failure for specific ESP operation systems **[6]**. The full implementation of a data-driven model is dependent on the availability of regular data collection. As a result, it is possible to characterize the challenge of performing failure identification on the ESP to develop a highly accurate data-driven model that captures the dynamics of the ESP system **[7]**.

This study focuses on applying machine learning to improve the performance of Electrical Submersible Pumps (ESPs). Using an extensive dataset from several years of production of multiple wells to train a comprehensive set of machine learning algorithms and predict ESP failures before they happen. The results of the proposed data-driven models are crucial in aiding the operators toward transitioning from reactive event-based maintenance to proactive and predictive maintenance of artificial lift operations. This approach increases the life and efficiency of the ESP by boosting the uptime, lowering the intervention costs, and optimizing the production.

The present work is consisted of three (03) chapters:

The first chapter provides a comprehensive background on ESP systems, detailing their challenges and outlining the study's objectives and scope of work. It covers the advantages and limitations of ESPs, common failure modes, and the necessity of predictive maintenance. Additionally, the chapter introduces the role of machine learning in enhancing ESP performance, discussing various types of machine learning and their significance in this context.

Chapter II presents the process of the suggested methodology and experimental implementation tools which covers data preparation, unsupervised learning, data balancing and supervised learning with Random Forest.

The last Chapter outlines the customized workflow base on the data collected from four (04) wells in Hassi Messaoud region. It includes results and discussion, presenting the outcomes of the model, summarizes key findings, and concludes the study.

I. CHAPTER I: BACKGROUND & CHALLENGES OF ESP SYSTEMS

I.1. Introduction:

Nearly half of all oil wells worldwide use Artificial Lift (AL) technologies, which are crucial to the oil and gas industry's ability to control bottom hole pressure and increase production rates in hydrocarbon wells **[8]**. Based on market research, AL systems are expected to increase steadily between 2022 and 2027, with a compound annual growth rate (CAGR) of 4.5%. AL systems are required in mature fields, which account for a large amount of the world's oil and gas production, in order to increase recovery rates.

Asia-Pacific is among the fastest-growing AL markets, while North America led the global AL systems market in 2021 **[8]**.

As shown in Figure I-1, North America is likely to dominate the market due to its many mature oil and gas reserves and expanding energy demand.

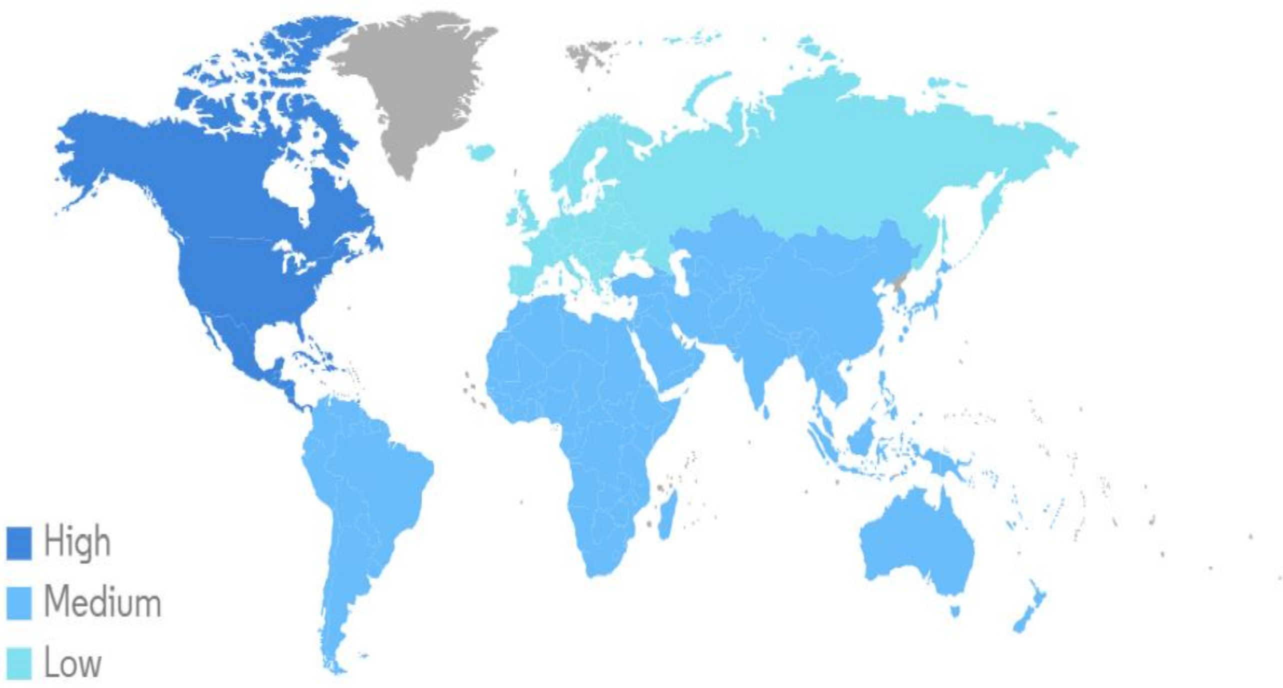


Figure I-1: Artificial Lift System Predicted Growth Across the World

I.2. Types of Artificial Lift:

There are several types of artificial lift systems commonly used **[9]**:

I.2.1. Sucker Rod Pump:

Referred to as beam pumping, this method provides mechanical energy efficiently, simply, and is easy to operate.

I.2.2. Electric Submersible Pump (ESP):

A submerged multistage centrifugal pump powered by a downhole electric motor. The electric motor connects directly to the centrifugal pump module in an ESP. This means that the electric motor shaft connects directly to the pump shaft. Thus, the pump rotates at the same speed as the electric motor.

I.2.3. Gas Lift:

Injects natural gas (typically natural gas or compressed air) into the wellbore to reduce the density of the fluid column. This process can be continuous or intermittent, with various configurations based on well conditions and production requirements.

I.2.4. Progressing Cavity Pump (PCP):

It is a positive displacement pump. It uses an eccentrically rotating single helical rotor, turning inside a stator. It can be used for lifting heavy oils at a variable flow rate.

I.2.5. Hydraulic Jet Pump:

Hydraulic pumping systems transmit power downhole by means of pressurized power fluid that flows in wellbore tubulars.

Jet Pump converts the pressurized power fluid to a high-velocity jet that mixes directly with the well fluids.

I.3. ESP Systems and Challenges:

Artificial lift techniques are used when reservoirs energy is insufficient to naturally produce oil or gas to the surface. More than 90% of operational oil wells depend on an artificial lift system **[10]**. ESPs are frequently considered as being extremely reliable and efficient for pumping high volumes from deeper depths among all oil field lift systems. They are adaptive to highly deviated wells and deployed in varied operating environments all over the world **[11].** These pumps boast a wide application range, enabling the extraction of hydrocarbon fluids from greater depths while handling a range of viscosities and gas-liquid ratios.

ESP failures are common and unexpected, imposing major financial pressures on operators and several challenges for ESP systems **[12]**. These include downhole environments that are harsh and can cause premature wear, gas interference that can disrupt pump operation, fluid properties that vary and can affect performance, damage caused by sand and solids, electrical failures, reduced efficiency due to scaling and deposition, complex operations, high initial and ongoing costs **[13]**. Proactive maintenance, effective monitoring, efficient sand control, and continuous technical advancements are needed to meet these problems.

I.4. Objectives of Study:

This study has the following main objectives:

- ✓ Conduct a comprehensive literature review to determine the development of ESP failure prediction through time and the systems necessary to assist in the process of predicting failures at early stages.
- ✓ Conduct a statistical analysis on the field data gathered from the surface and downhole ESP monitoring equipment to classify and characterize the mechanisms of ESP failure.
- ✓ Construct a machine learning models to handle the high frequency of ESP data and predict various ESP failure modes at different prediction periods. Utilize performance metrics to validate the models.
- ✓ Explore and compare different techniques for calculating feature importance, considering the model used and the nature of the data.
- ✓ Assess the effectiveness of Random Forest, a powerful method for predicting imbalanced classification and determining feature importance and extract decision rules by replacing the model with a decision tree, contributing to a deeper understanding of the classification process.

I.5. Scope of Work:

This study uses the field data acquired from Four (04) operational wells of Hassi Messaoud, the recorded data involves different field conditions, including dynamic, static, and historical data. The data were collected from wells with high ESP failure rates to investigate the general ESP failures and their specific failure modes. Additionally, historical operational data are included in the study.

The first step in constructing a Machine Learning model is data preparation to clean, organize, and categorize the data to be used by the model. Then comes the other steps, which involves detecting anomalies, Data labelling, Data balancing and optimizing hyper parameters. These steps account for 80% of the overall time.

In addition, generating multiple features for the Machine Learning model is essential. This study evaluates the ESP's performance using collected field data.

I.6. Overview of predictive maintenance:

Predictive maintenance is a proactive approach to anticipating and averting equipment breakdowns before they happen. Predictive maintenance analyses production data, including the huge amount of real-time sensor data obtained from equipment, to identify patterns and potential issues **[14]**.

For ESPs, surface and downhole gauges continuously measure a range of parameters in real-time. As a result, there is a large amount of raw sensor data that needs to be processed in order to get useful information. A proactive approach can replace a reactive one by using data-driven approaches to identify and potentially diagnose ESP issues before they arise **[14]**.

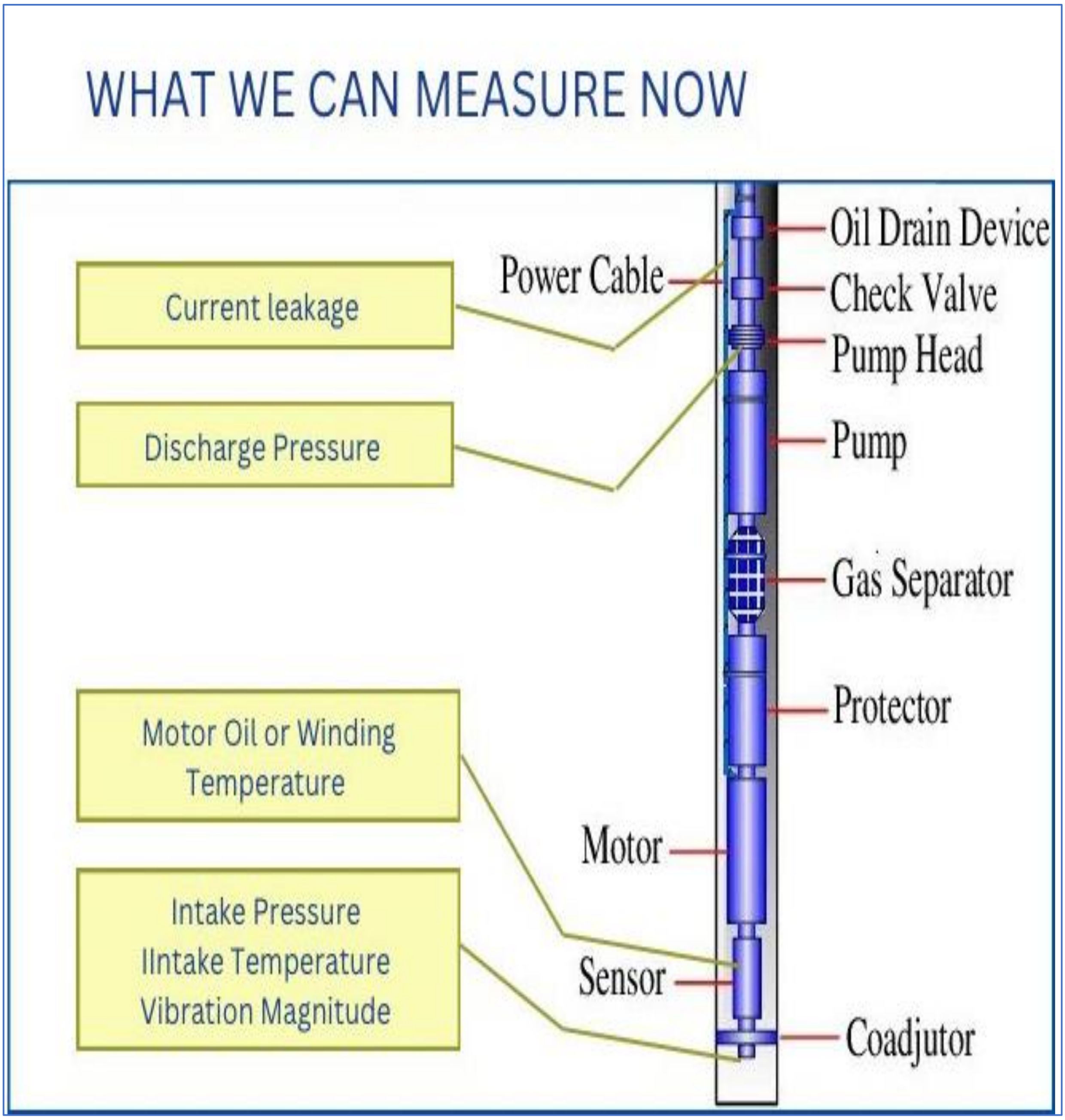


Figure I-2: Parameters measured by downhole and surface gauges

Proactive maintenance employs intelligence analysis to detect potential issues and anticipate ESP behaviour in advance **[15]**. By employing proactive maintenance strategies, the management of ESP shutdowns and other potentially dangerous situations can be significantly improved. Predictive maintenance ultimately helps managers in the production of oil and gas by maximizing maintenance schedules, minimizing downtime, and enhancing operational effectiveness and cost-effectiveness **[16]**.

I.7. Electrical Submersible Pump:

Electrical Submersible Pumps (ESPs) are one of the most often used artificial lift techniques in the petroleum industry, especially for highly productive oil wells **[17]**. They can result in considerable increase in well's production if correctly maintained and kept with their ideal operating limitations **[18]**.

In ESP system, there are downhole and surface components **[19]**. The main downhole part consists of an electrical motor, seal section, a multistage centrifugal pump, power cable, and downhole sensors. The main surface components are: Switchboard of the variable speed drive (VSD), transformers, surface electric cable, junction box, and wellhead.
Figure I-3 illustrates the standard setup of an ESP system.

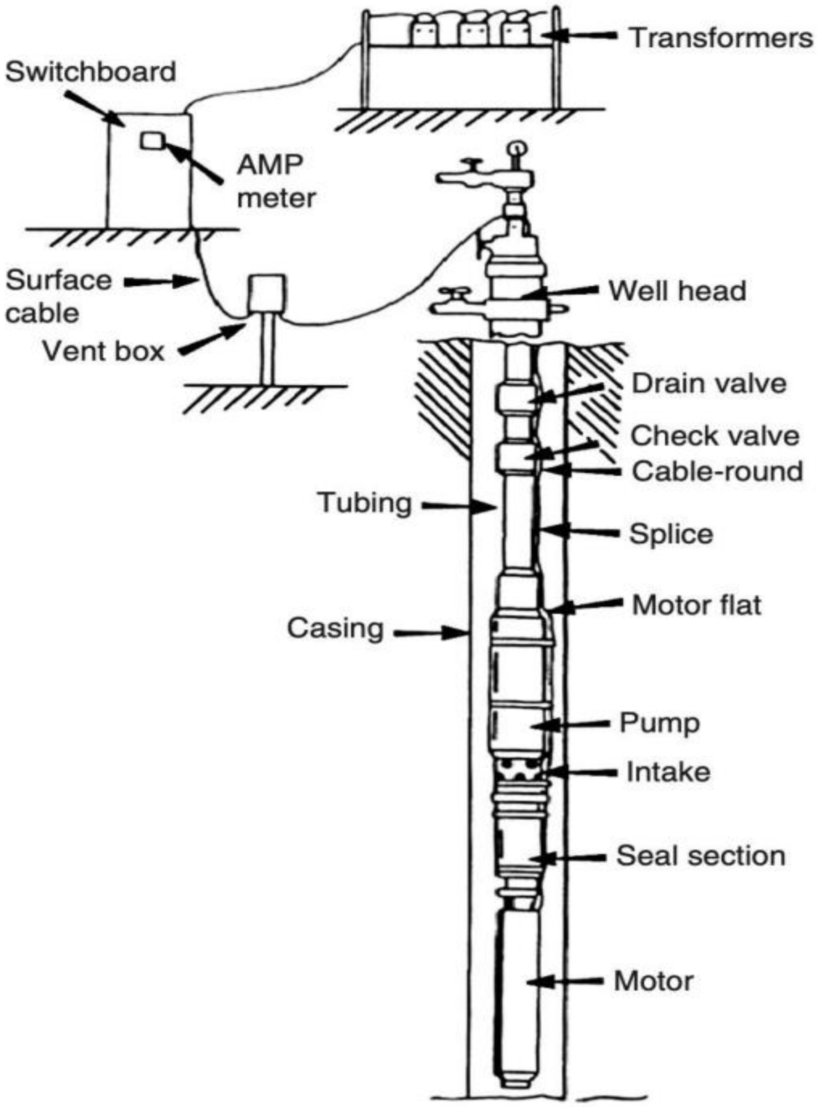


Figure I-3: A Representative ESP System Schematic

The Electrical Submersible Pump system consists of a series stack of many stages of centrifugal pumps coupled to a submersible electric motor. Variable speed drives (VSDs) alter the pump's working frequency and, as a result, its speed, improving the ESP's performance.

The junction box connects the downhole electrical cable to the VSD cable and permits the wellbore gas to be evacuated. The motor receives electrical power from the surface controls through the electrical wire, transforms it into mechanical energy, and then transmits that energy via a coupled shaft to the pump impellers. The motor cools as the produced fluid flows past the motor housing. The shaft provides the mechanical energy for the pump stages **[19]**.

The number of pump stages required in a well to get the desired flow rate will depend on its completion design and operational requirements. Each step has an impeller that rotates and stationary diffusers. Figure I-4 illustrates a multistage ESP system with a rotating impeller and a stationary diffuser at each stage. Figure I-5 depicts a single-stage ESP pump with an impeller and a diffuser. By spinning the blades, the impeller promotes fluid flow and delivers kinetic energy to the fluids **[20]**. The fluid's kinetic energy is transformed to pressure potential at the diffuser. This process is carried out at each stage of the pump, as illustrated in Figure I-6.

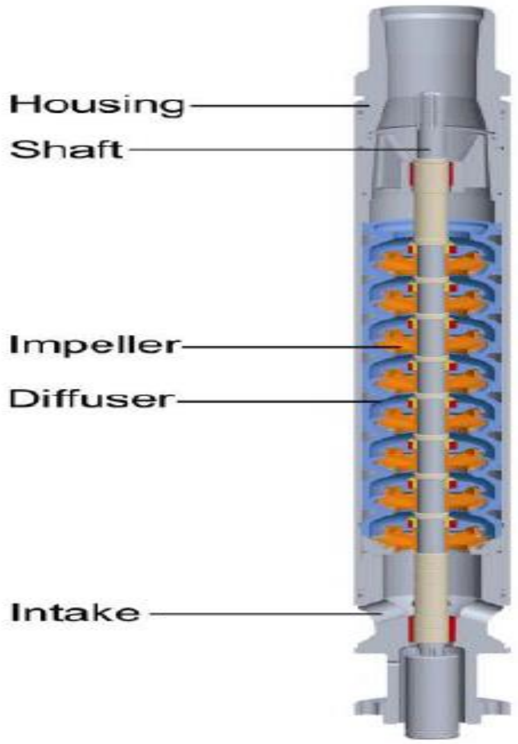


Figure I-4: ESP Pump Cutaway

Figure I-5: ESP Pump Stage-Impeller and Diffuser

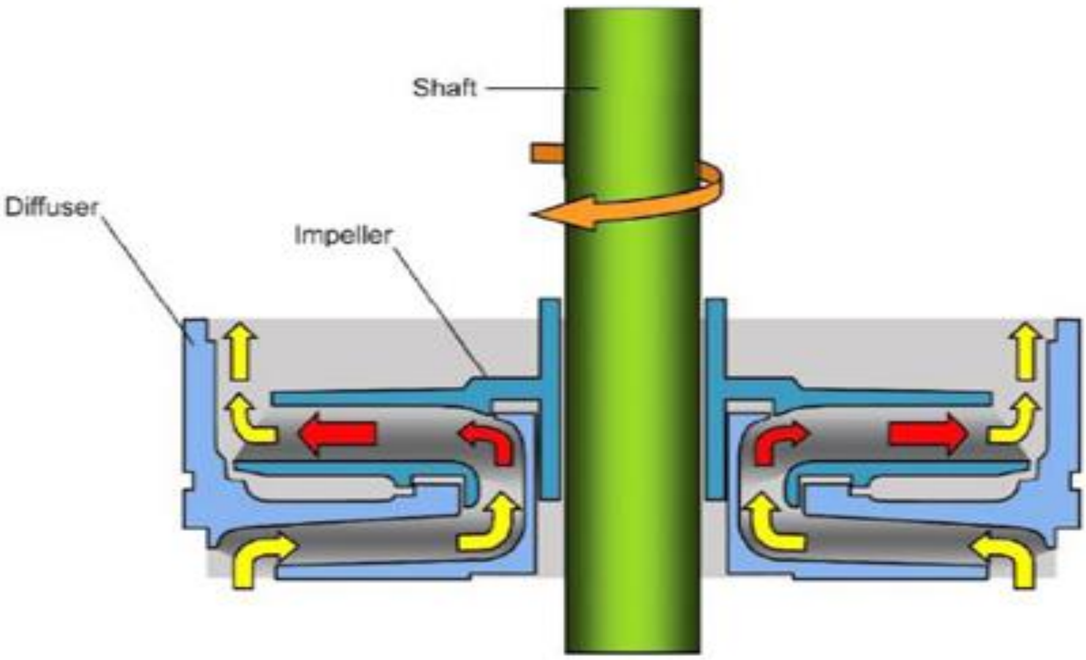


Figure I-6: ESP Pump Stage-Impeller and Diffuser

The ESP seal section, which is situated between the pump intake and the downhole electric motor, provides significant benefits to the system. The axial thrust generated by the pump is supported by its thrust bearings **[21]**. In addition to isolating and protecting the motor from well fluids, the seal also balances the pressure inside the motor and in the wellbore.

Free gas may be present in fluids produced from wells with low bottom-hole pressures and high gas-oil ratios (GOR). Due to operational difficulties, such as cavitation or gas locking for gassy wells, ESPs may have a shorter run-life **[22]**. Therefore, a gas separator is installed at the pump's intake in these wells to separate the free gas from the produced fluid.

Installing an ESP monitoring system also has several advantages since it continuously records pump performance and gathers crucial downhole data. These datasets offer a variety of field conditions, including as static, dynamic, and historical data, and they are essential sources of information for predictive models that try to anticipate ESP problems before they happen **[23]**.

I.8. ESP Advantages and Limitations:

Using ESP systems provides several operational advantages over conventional artificial lift approaches, including a smaller surface footprint and lower noise levels. These systems can be used with horizontal or curved wells with doglegs up to 10°/100 feet; the pump must be installed in the straight section. ESPs excel at producing high liquid volumes (200-100,000 BPD) from moderate depths with a range of 1,000’-15,000’ TVD, Additionally, ESPs are adaptable in casings > 4-1/2” and are suitable for well testing **[24]**.

There are several challenges with ESP operations. High sand and solids conditions can significantly affect the run life of ESP systems, even though they can be built with certain abrasion-resistant materials **[25]**. High viscosity fluid production (> 1,000 cp) and significant free gas content (> 5%) in the pump both have a negative impact on performance **[19]**. Furthermore, precise well inflow data must be used in the design process, and the unit's capacity must match the well's deliverability. If not, expensive workover activities will be necessary to operate the pump **[26]**.

I.9. ESP Common Failure Modes:

ESP system monitoring has evolved over the years and ammeter charts came to be offered as the earliest and simplest diagnostic solution to minimize downtimes for many decades. They measure and record the electric current drawn by the ESP motor. The current is recorded continuously in the function of time on a continuous chart with the proper scale. Figure I-7 is a typical example of Ammeter chart in the scenario of gas locking. Such behaviour is observed when the capacity of the ESP unit is greater than the inflow to the well and the well produces substantial free gas volumes.

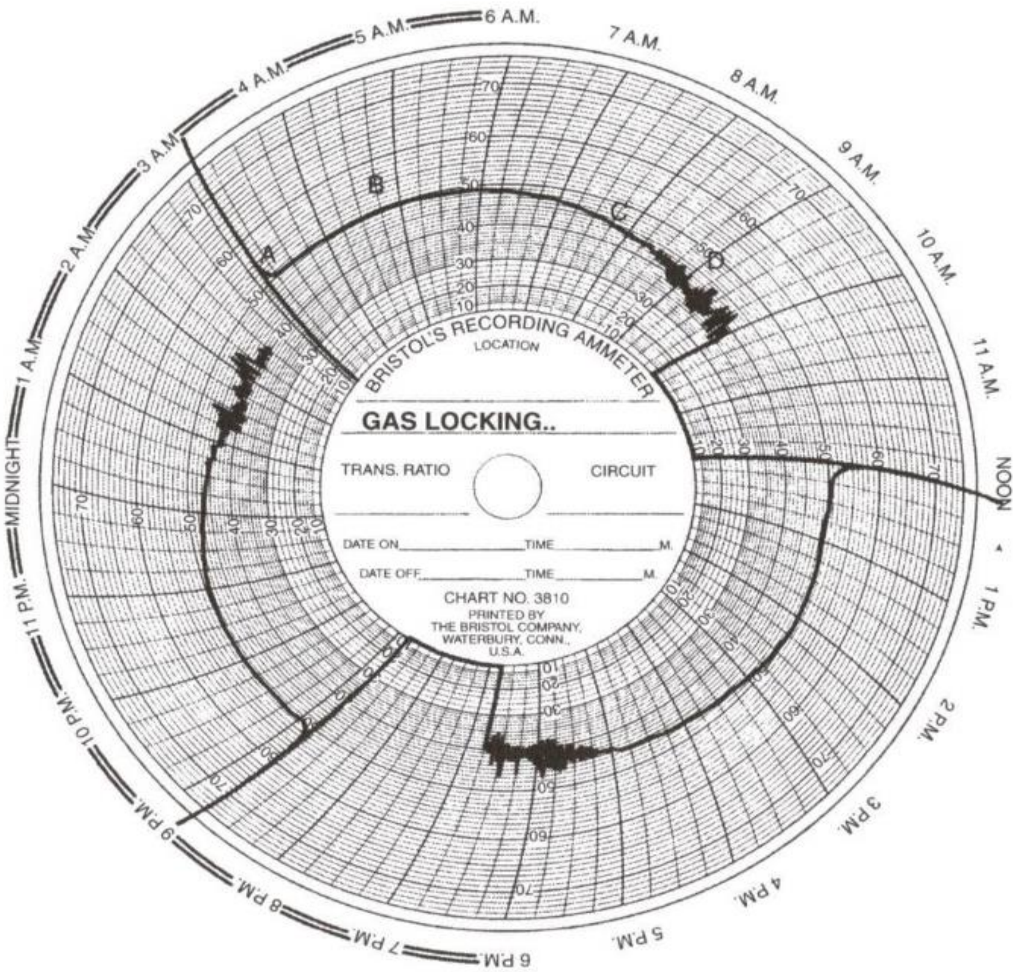


Figure I-7: Ammeter Chart Example: Pump off with Gas Interference

The proper interpretation of ammeter charts can provide useful information to detect and correct minor operational issues **[19]**. It provides a very one-sided picture of ESP unit's operations since it relies on only electrical measurements. Electrical failures are often caused by mechanical or other problems which, over time, develop into a failure of an electrical nature. The detection of the initial failure, therefore, is not an easy task and requires additional information **[27]**.

The ESP installation works as a system consisting of mechanical, hydraulic and electrical components and, in order to diagnose and prevent failure, a dynamic system which can capture multiple parameters affecting an ESP operation in real-time and provide an end-to-end solution is necessary.

Microprocessors are being used by several ESP controllers to give much better control and protection of the electric components of the ESP system. They also employ a number of other electrical variables that can be stored for analysis at a later time **[19]**.

I.10. The Need of Machine Learning:

I.10.1. Definition:

Machine learning is a subfield of artificial intelligence (AI) that studies how to create models and algorithms that enable computers to learn from historical data and make decisions or predictions without the need for explicit programming. The basic goal of machine learning is to make computers learn automatically and improve over time **[28]**.

An important application of machine learning is in predictive maintenance (PdM), which is implemented to effectively manage maintenance plans of the assets by predicting their failures with data driven techniques. In these scenarios, data is collected over a certain period of time to monitor the state of equipment. The objective is to find some correlations and patterns that can help predict and ultimately prevent failures **[29]**.

I.10.2. Types of Machine Learning:

There are two basic approaches: supervised learning and unsupervised learning. The main difference is one uses labelled data to help predict outcomes, while the other does not:

- ***Supervised Learning***:

Supervised learning is a machine learning approach that’s defined by its use of labelled datasets. These datasets are designed to train or supervise algorithms into classifying data or predicting outcomes accurately. Using labelled inputs and outputs, the model can measure its accuracy and learn over time **[30]**.

Supervised learning can be separated into two types of problems when data mining: classification and regression, Example: Decision Tree (DT), Support Vector Regression (SVR).

- ***Unsupervised Learning****:*

Unsupervised learning uses machine learning algorithms to analyse and cluster unlabelled data sets. These algorithms discover hidden patterns in data without the need for human intervention. Unsupervised learning tasks include clustering and dimensionality reduction, Example: K-Means, Gaussian Mixture Models (GMM) and Principal Component Analysis (PCA)

The main distinction between the two approaches is the use of labelled datasets. To put it simply, supervised learning uses labelled input and output data, while an unsupervised learning algorithm does not **[30]**.

I.10.3. Importance of Machine Learning:

Machine learning is a crucial tool in many different businesses because it can automate decision-making processes, make predictions, and extract insights from large and complex datasets. Several primary arguments support the importance of machine learning, including the following:

- ***Predictive analytics:*** Machine learning algorithms integrate data from seismic, well log, core analysis, and past performance. It can analyse large datasets, including historical maintenance records, sensor data, and equipment performance metrics, to identify patterns and predict when equipment is likely to fail. This allows for proactive maintenance scheduling, reducing downtime and increasing overall efficiency **[31]**.
- ***Real-time Monitoring***: Machine learning can be integrated with real-time monitoring systems, enabling the detection of anomalies and potential issues before they become major problems **[31]**.
- ***Optimization of Maintenance Schedules***: By analysing data on equipment performance and maintenance history, machine learning can help optimize maintenance schedules, ensuring that maintenance is performed at the most opportune times to minimize disruption and reduce costs **[32]**.
- ***Improved Safety***: Predictive maintenance enabled by machine learning can help identify potential safety risks before they materialize, allowing for proactive measures to be taken to prevent accidents and ensure a safer working environment **[33]**.

- ***Cost Savings***: By reducing downtime and extending the lifespan of equipment, machine learning-driven predictive maintenance can lead to significant cost savings for oil and gas companies **[33]**.
- ***Enhanced Decision-Making***: Machine learning provides oil and gas companies with data-driven insights, enabling more informed decision-making around maintenance strategies, resource allocation, and investment in new technologies **[33]**.

II. CHAPTER II: EXPERIMENTAL ANALYTICS WORKFLOW METHODOLOGY:

II.1. Introduction:

This research intends to develop a model that can predict downhole electrical submersible pump problems, so that proper actions might be taken proactively to avoid the occurrence of such problems. The approach of the unsupervised learning is used to train the model.
Unsupervised learning algorithms will be utilized to analyse the training data. These algorithms aim to uncover patterns or structures within the data without explicit guidance or labelled examples **[30]**.

II.2. Proposed Methodology and Implementation:

The study is executed using the Knowledge Discovery in Databases (KDD) process **[34]**. This process is used to show: Data preparation (Data collection & pre-processing), unsupervised learning with Isolation Forest, Data Labelling, Data balancing with SMOTE, Supervised learning with Random Forest and Feature Importance.

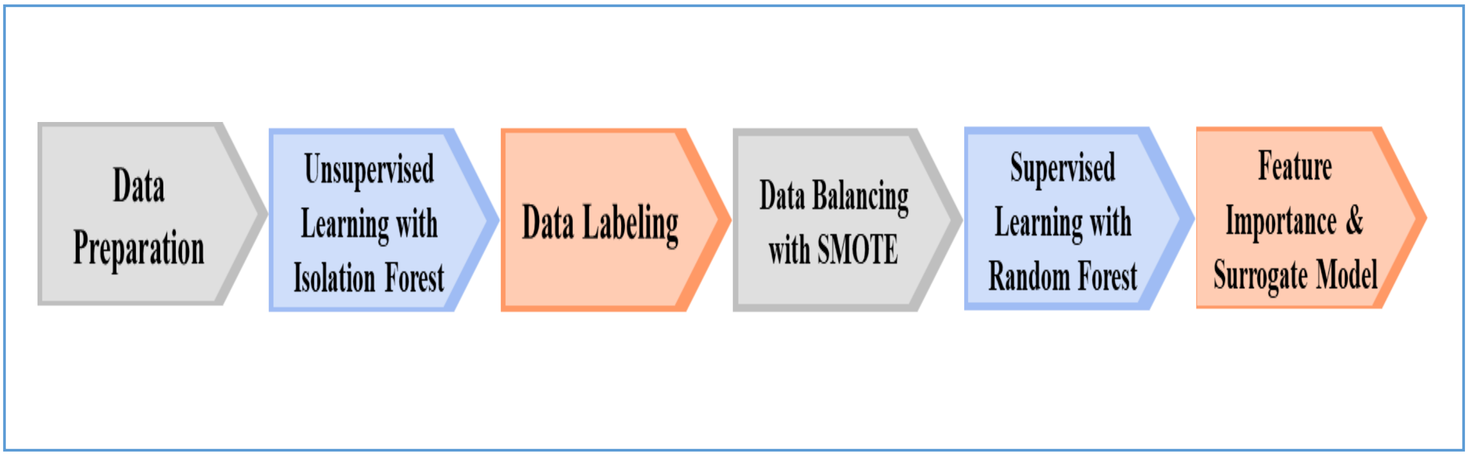


Figure II-1: Knowledge Discovery in Database Workflow

II.2.1. Data Preparation:

Our data is collected from four (04) sensors on electrical submersible pumped wells located in the region of Hassi Messaoud. These data are structured in Excel files and include daily logs detailing the operation of the pumps, as well as the essential parameters monitored for each well. Then, we proceed to clean, organize, and categorize this data and making it suitable for Machine learning techniques.

The ranges of physical real- time parameters of these wells are shown in Table II-1:

PVT Data	**Unit**
Oil API Gravity	API
Oil Specific Gravity	API
Gas Specific Gravity	API
Water Salinity	ppm
Water Gravity	API
RS at Saturation Pressure	Sm3/Sm3
Fluid Model	API
Pressure	Kg/cm2 (Gauge)
Bo	m3/Sm3
Rs	Sm3/Sm3
Oil Viscosity	cp
Laboratory Data	
Saturation Pressure	Kg/cm2_g
Temperature	C°

Table II-1: The range of physical parameters

The listed of the parameters studied:

Parameter	**Description**	**Unit**
RS	Solution gas-oil ratio /RS	Sm3/Sm3
SP	Saturation Pressure	Kg/cm2
Diamduse	Duse Diameter	mm
Debithuile	Oil Flow	m3/h
Débit gas	Gas flow	m3/h
GOR	Gas Oil Ratio	Sm3/Sm3
Presstete	Wellhead Pressure	Kg/cm2
Presspipe	Tubing pressure	Kg/cm2
presssepar	Separator pressure	Kg/cm2
Densitehuile	Oil density	g/cm3
Temphuile	Oil Temperature	C°
KPsi	Permeability	Psi

Table II-2: The Listed of the parameters studied

Based on the data collected from the four (04) wells, we have obtained the graphs related to the Gas-Oil Ratio (GOR) that illustrate the behaviour of the wells over time, as well as their head pressure graphs as shown in the Figure II-2 and Figure II-3:

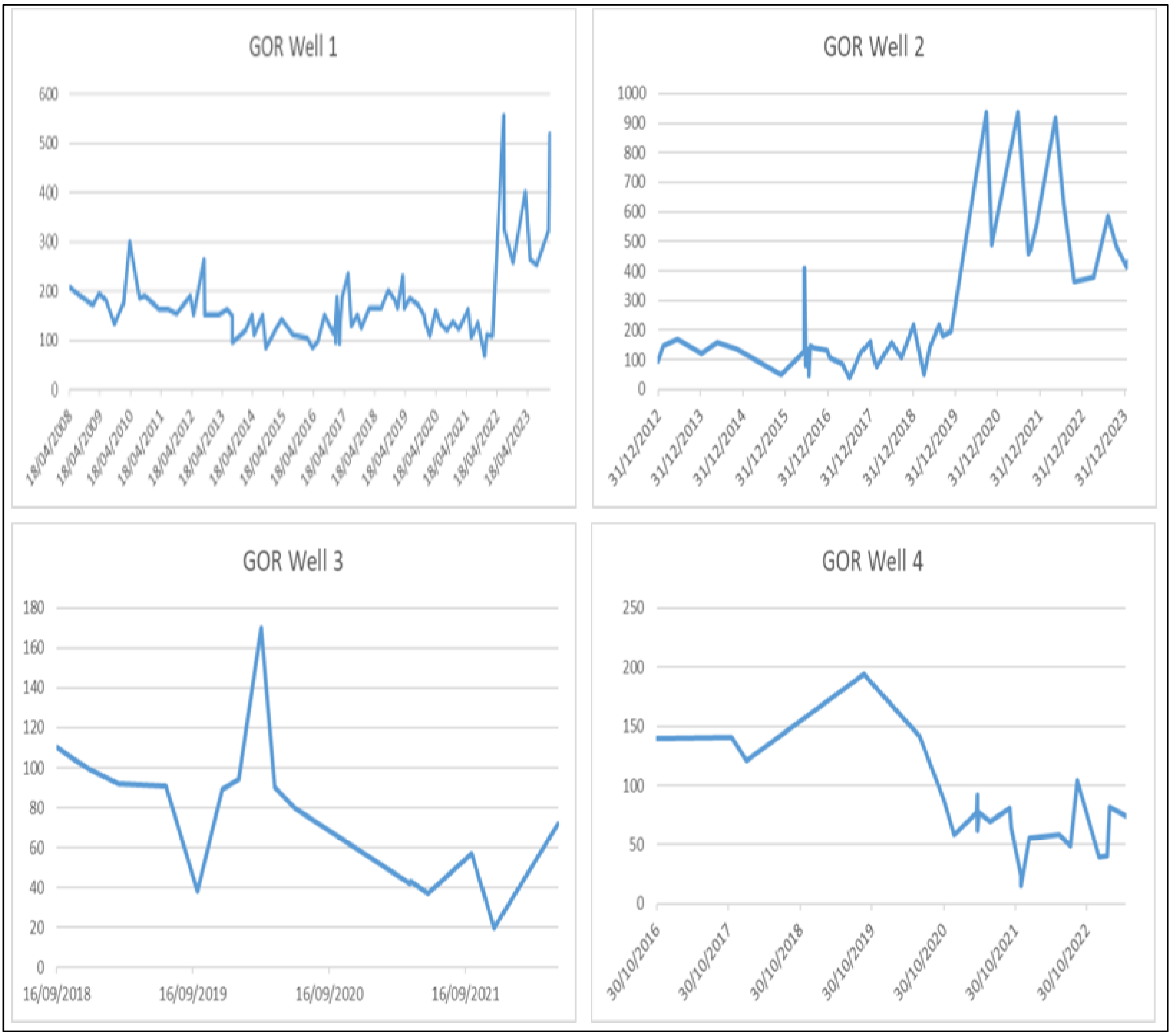


Figure II-2: History of GOR Measurement of 04 wells

Analysing Gas-Oil Ratio (GOR) variations across wells offers valuable insights into reservoir behaviour and production dynamics. This assessment helps refine reservoir management strategies and enhances oil and gas recovery efforts **[35]**.

Furthermore, considering the previous effects of gas on ESP pumps, such as gas locking, increased vibration, and reduced pump capacity, reinforces the importance of monitoring GOR variations **[35]**.

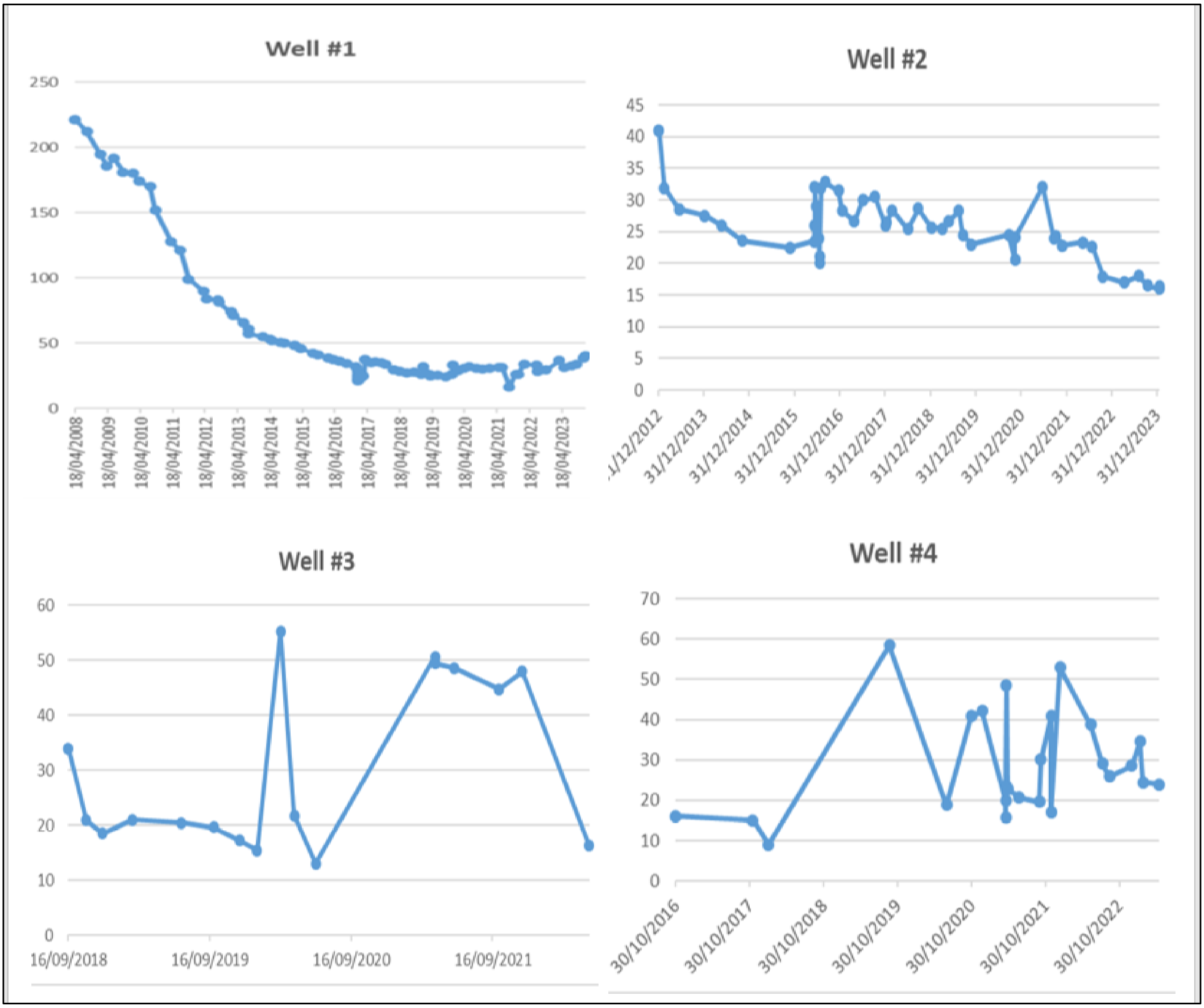


Figure II-3: History of WHD Pressure Measurement of 04 wells

Out of the four wells, Well #1 has the highest pressure, which suggests that it has the highest production potential and the most impact on the reservoir's overall performance. Compared to the other wells, Well #2 exhibits the lowest pressures, which indicate a limited capacity for production.

Due to the wide range of Data on the Well #1, we rely on its information in the following chart:

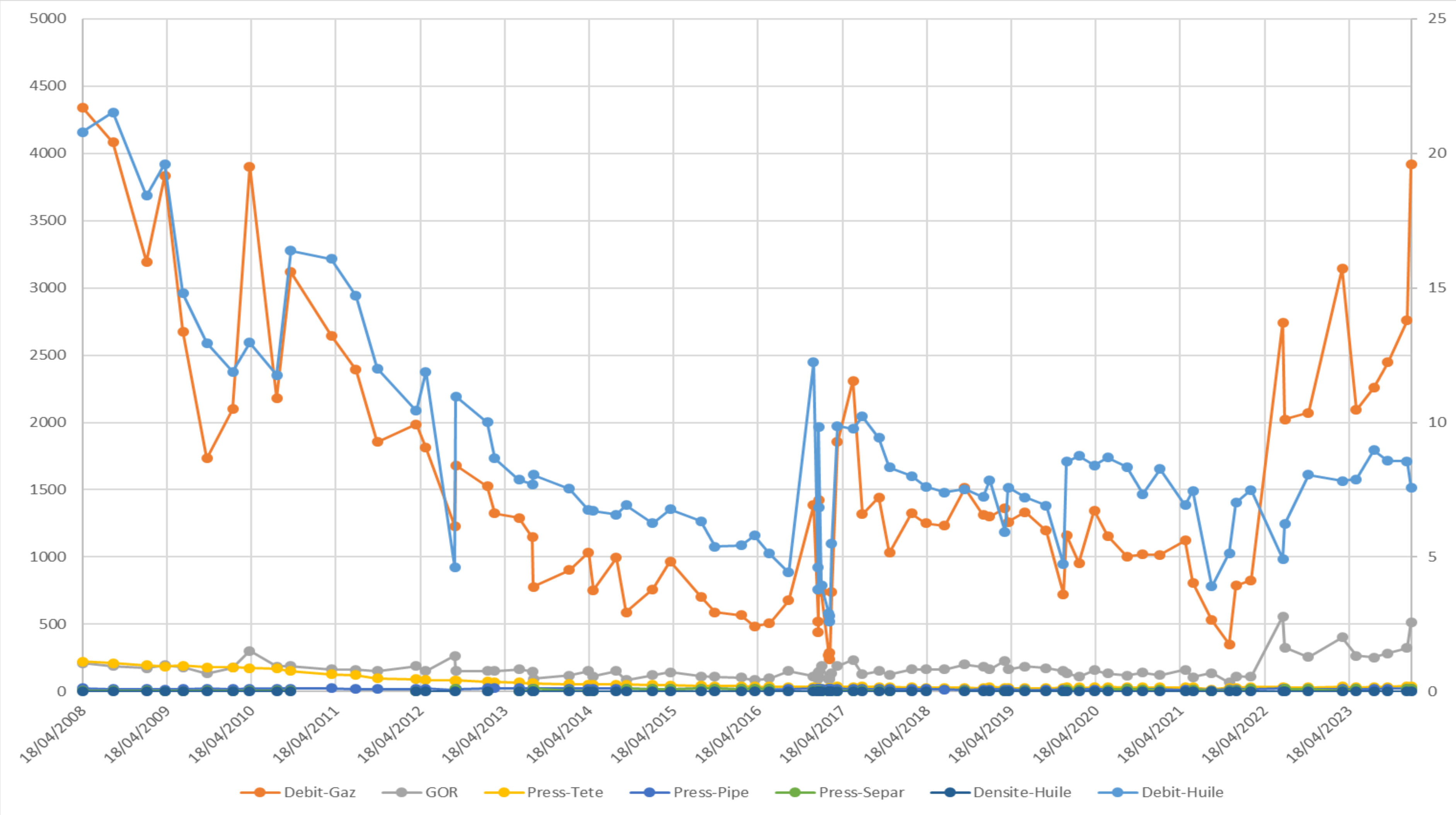


Figure II-4: Well#1 real-time parameters 2008- 2023

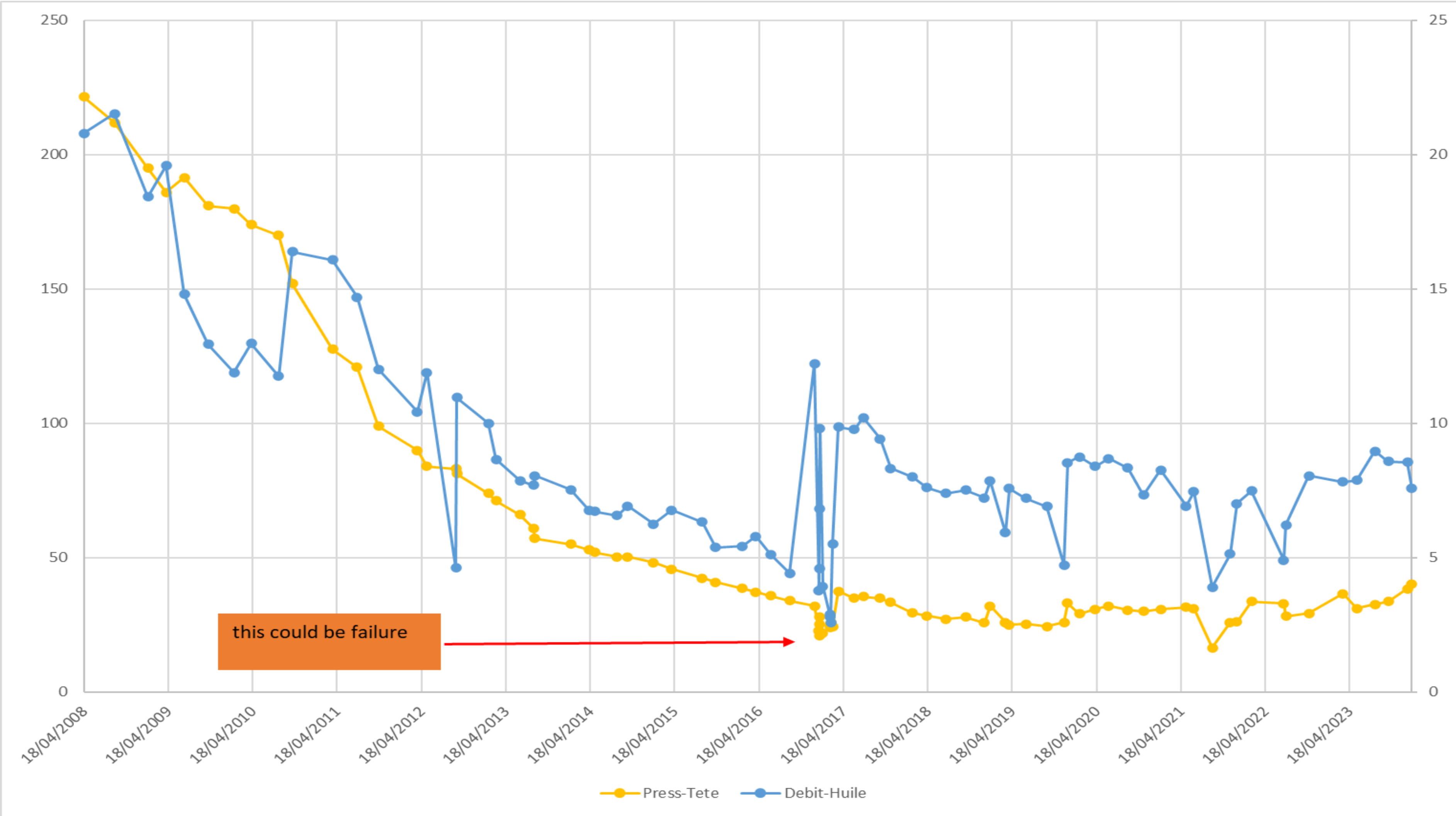


Figure II-5: Well#1 real-time parameters WH Pressure and Oil Flow

The head pressure and oil flow graph for Well #1 shows a possible problem in 2017 that might be a pump failure or malfunction.

By applying the machine learning model to the analysis of parameter variations across different wells, we are able to gain a deeper understanding of the fluid composition, reservoir behaviour, and production efficiency. This allows us to progress hydrocarbon recovery operations and optimize reservoir management strategies. This comprehensive study raises operational effectiveness and lowers downtime risks. In oil and gas exploration and production activities, this encourages sustainable production methods and optimizes the use of resources.

II.2.2. Unsupervised Learning with ISOLATION FOREST:

During the first step of preparing the data, it become clear that the data provided to us is unlabelled. This means that for the machine learning process, we need to use an unsupervised category.

Isolation Forest is a type of an unsupervised machine learning algorithm designed to identify outliers in datasets. This powerful anomaly detection technology is especially useful for handling unlabelled data, where anomalies are not explicitly marked **[36]**.

How Isolation Forest Works:

Isolation Forest works on the principle that anomalies are few and different, making them easier to isolate. The algorithm works as follows:

- **Training Stage:** A training dataset is used to build Isolation Trees (iTrees). For each iTree, a random feature is selected and a random split value between the min and max values of that feature is chosen to partition the data. This process is repeated recursively until all data points are isolated.
- **Evaluating Stage:** For each instance in the test set, the path length from the root node to the terminating node is calculated for each iTree. The average path length across all iTrees is then determined for each instance. The anomaly score is inversely proportional to the average path length, with anomalies typically having shorter path lengths due to their distinct nature, which allows them to be isolated more quickly.

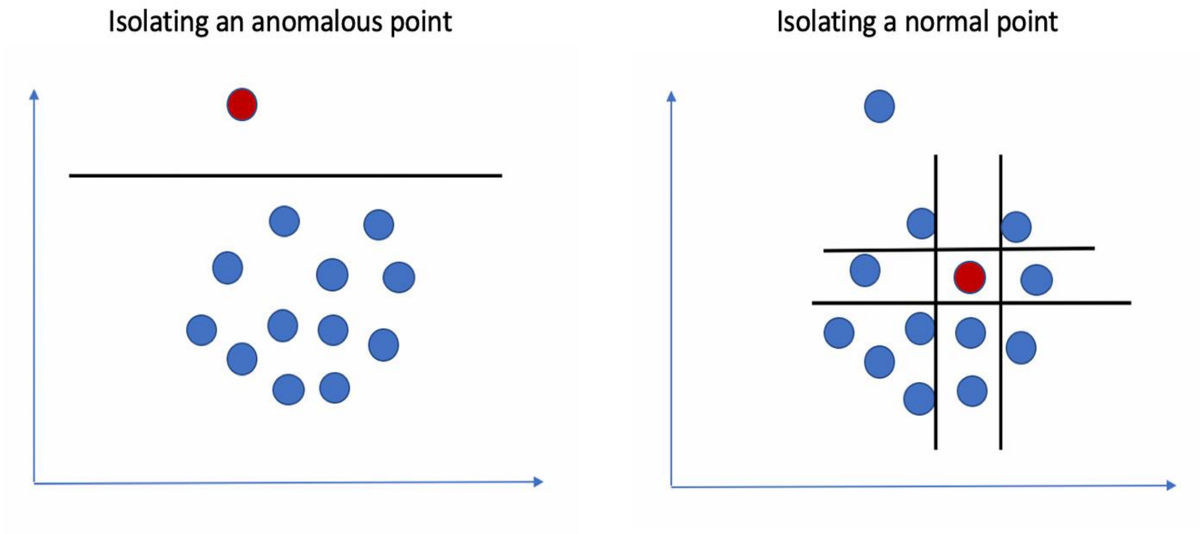


Figure II-6: Isolating Anomalies with Isolation Forest

Key properties of Isolation Forest:

- Linear time complexity and low memory requirement, making it suitable for large datasets.
- No density estimation is performed, unlike other anomaly detection algorithms.
- Anomaly score is calculated based only on the path length, without using leaf node statistics.
- Randomly selects features and split points, making it fast and applicable to high-dimensional data.

II.2.3. Data Balancing with SMOTE:

After applying Isolation Forest for labelling, our dataset reveals a significant imbalance between classes, which could impact the performance of machine learning models. This imbalance indicates that the minority class, representing cases of anomalies, is underrepresented compared to the majority class. To fix this problem we used the approach SMOTE (Synthetic Minority Over-Sampling Technique).

Concretely, SMOTE works by synthesizing new examples for the minority class by interpolating the features of neighbouring points from that class; This boosts the minority class's sample number and restores the dataset balance. As a result, the machine learning model is exposed to a wider range of balanced examples, which enhances its capacity for correct generalization **[37]**.

By using SMOTE, we may increase the representation of minority cases in our data without unnecessarily biasing it.

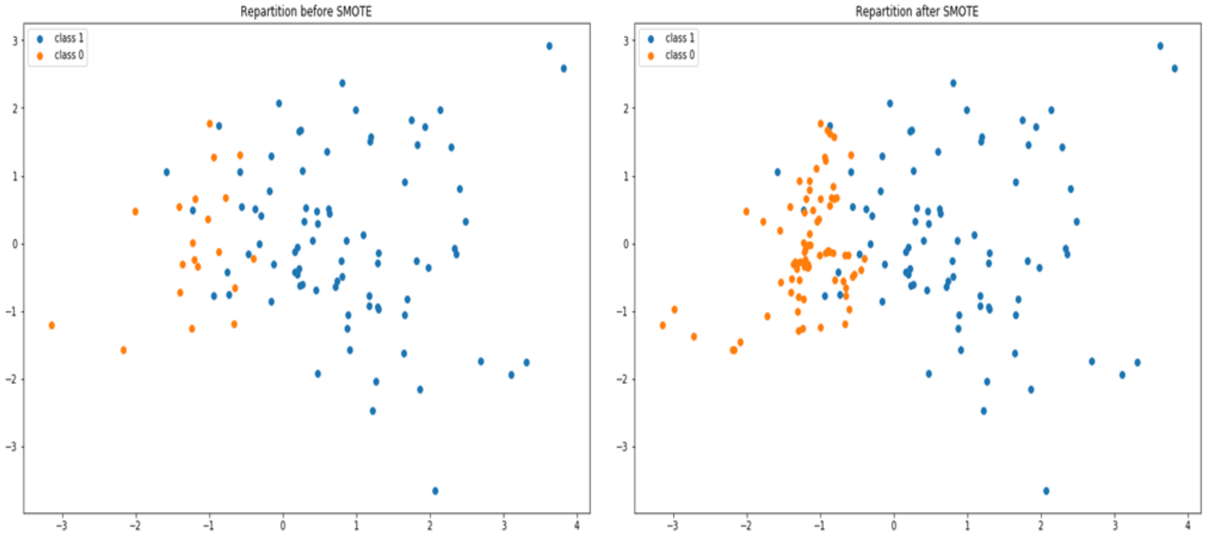


Figure II-7: Data Repartition before & after SMOTE

II.2.4. Supervised Learning with Random Forest:

The data is first divided into two subsets, one for learning (66%), and the other for train and testing; Using the AUTO ML tool, we will select the algorithm that produces the best results with our dataset. This tool automates the creation and optimization of machine learning models, making the process more accessible and efficient for users without extensive technical expertise.

The Random Forest algorithm was selected due to its superior performance in terms of measures like classification accuracy, F1 Score, area under the ROC curve (AUC).

- Random Forest is a powerful machine learning algorithm for both regression and classification applications. It involves building a collection of decision trees that cooperate to produce a single output, enhancing prediction accuracy by leveraging the diversity and collective wisdom of the trees. Due to its reputation and ability to handle complex data, minimize overfitting, and produce accurate forecasts in a variety of settings, Random Forest has emerged as a key component of machine learning **[38]**.

➢ A Decision Tree (DT) is an adaptable supervised learning method that may be applied to both regression and classification tasks. It creates a tree-like structure where each internal node represents a feature, each branch represents a decision rule, and each leaf node represents the outcome. Decision trees are frequently utilized because to its high degree of accuracy, interpretability, and ability to handle both numerical and categorical data **[39]**. For predictive modelling and decision-making, they are extensively utilized in a variety of industries.

➢ A ROC curve is a graph showing the performance of a classification model at all classification thresholds **[40]**. This curve plots two parameters:

- True Positive Rate
- False Positive Rate

Figure II-8: The ROC space for a "better" and "worse"

After validating our model, we determined the most important parameters that have a major impact the prediction of failures and gave an explanation using a global surrogate model.

- A global surrogate model is an interpretable model trained to approximate the predictions of a black-box model, allowing us to conclude the latter by interpreting the surrogate. These interpretable models mimic the behaviour of the original model by overfitting its predictions. The intuition is that if the surrogate model makes the same predictions as the original model, it can help us understand how the input features relate to those predictions. The quality of surrogate models is assessed using a user-defined performance metric **[41]**.

II.2.5. Feature Importance:

It is crucial to calculate the overall importance of features for our classification model, as it allows us to understand which features have the most influence on the model's predictions.
This understanding is necessary to evaluate the validity of the model and determine which features are most important to the classification task.

We may calculate feature importance using a variety of methods. Depending on the data type and model being used. The most common approach for predicting imbalanced classifications and feature importance is Random Forest, which is extremely successful.
Our model was replaced with a decision tree in order to extract decision rules.

II.3. Implementation Tools:

In our experimental study, we carefully selected implementation tools to guide the process, evaluate critical components of project execution, and ensure the successful achievement of our research objectives:

- ***KNIME***: (Konstanz Information Miner), or KNIME, is an open-source platform for data integration, analytics, and reporting. Users can incorporate tasks like data pre-treatment, analysis, visualization, and model building into their visual designs of data workflows. KNIME is a comprehensive suite of tools and algorithms that serves both non-technical users and data scientists, facilitating efficient data-driven decision-making in a variety of sectors and domains **[42]**.

- ***H2O***: is an open-source machine learning platform designed for big data analytics. It provides a distributed environment where advanced analytics tasks like predictive modelling and machine learning can be carried out. Building and deploying machine learning models on huge data sets is a common use case for H2O, which has a wide choice of efficient algorithms and easy interface with popular programming languages **[43]**.

- ***Auto ML*:** Automated machine learning, known as AutoML, is the process of automating the end-to-end process of building machine learning models. This includes tasks such as data pre-processing, feature engineering, model selection, and hyperparameter tuning. AutoML compares and chooses which algorithm best fits our data, making it easier for non-experts to create machine learning models by offering a simple user-friendly interface for experimenter training and model deployment **[44]**.

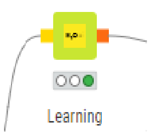
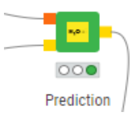
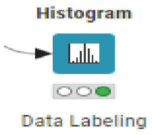
III. CHAPTER III: RESULTS AND DISCUSSION

III.1. Introduction:

This section presents the results and analysis of our predictive maintenance model for Electrical Submersible Pumps (ESPs). It covers the workflow design using the KNIME platform, the application of machine learning techniques for anomaly detection, and the evaluation of the model's performance. The discussion highlights each step involved, from data pre-processing to model training and prediction, demonstrating the effectiveness of our approach in identifying potential ESP failures.

III.2. Workflow:

The workflow diagram created using the KNIME platform. The diagram shows a network of connections between different nodes. Each node represents a step in the process of building and using a machine learning model:

	***Excel Reader:*** This node loads and reads data organized in Excel files and converts it into a KNIME table.
	***H2O Local Context:*** This node allows the creation of in a locally running H2O instance. For example, data tables can be converted to H2O frames and models created using H2O can be applied.
	***Table to H2O:*** This node converts data from a KNIME table format into a format compatible with H2O.
	***H2O Isolation Forest Learner:*** This node learns an Isolation Forest model using H2O. The model is learned in an unsupervised manner and can be used to detect anomalies or outliers via the *H2O Isolation Forest Predictor* node.
	***H2O Isolation Forest Predictor:*** This node applies an Isolation Forest model to an input dataset in to predict anomalies or outliers.
	***H2O to Table:*** This Node converts an H2O Frame to a KNIME table.
	***Histogram:*** This node produces a Histogram visualization of the input Data, that displays the distribution of the data and determination of boundaries.

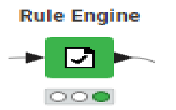
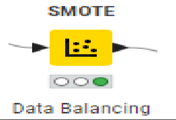
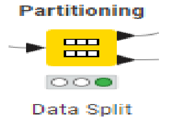
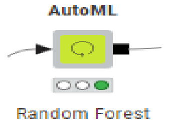
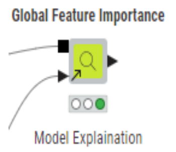
	***Rule Engine:*** This node labels our data based on length parameters. For our study, we found the following rules: if the length is greater than 5.7, the label is "No"; if the length is less than or equal to 5.7, the label is "Yes".
	***Column Filter:*** This node allows columns to be filtered based on specific criteria.
	***SMOTE:*** This node oversamples the input data to enrich the training data to get good classification performance.
	***Partitioning:*** The input table is split into two partitions: train and test data; our data based on the training dataset = 66%, and the rest considered as Unseen data
	***AutoML:*** This Component automatically trains supervised machine learning models. It can automate the whole ML cycle by performing some data preparation, parameter optimization with cross-validation, scoring, evaluation and selection. Random Forest was developed with the best configuration
	***Global Feature Importance***: This component offers an optional interactive view to explore the results and to determine the importance of each feature in the dataset using a surrogate model.

Table III-1: KNIME Nodes Listing

The data moves from top to bottom through the KNIME pipeline. Every step involves some sort of transformation or analysis of the data. The final goal of the workflow is most likely to train a machine learning model, such as the random forest model, and then apply it to fresh data to generate predictions.

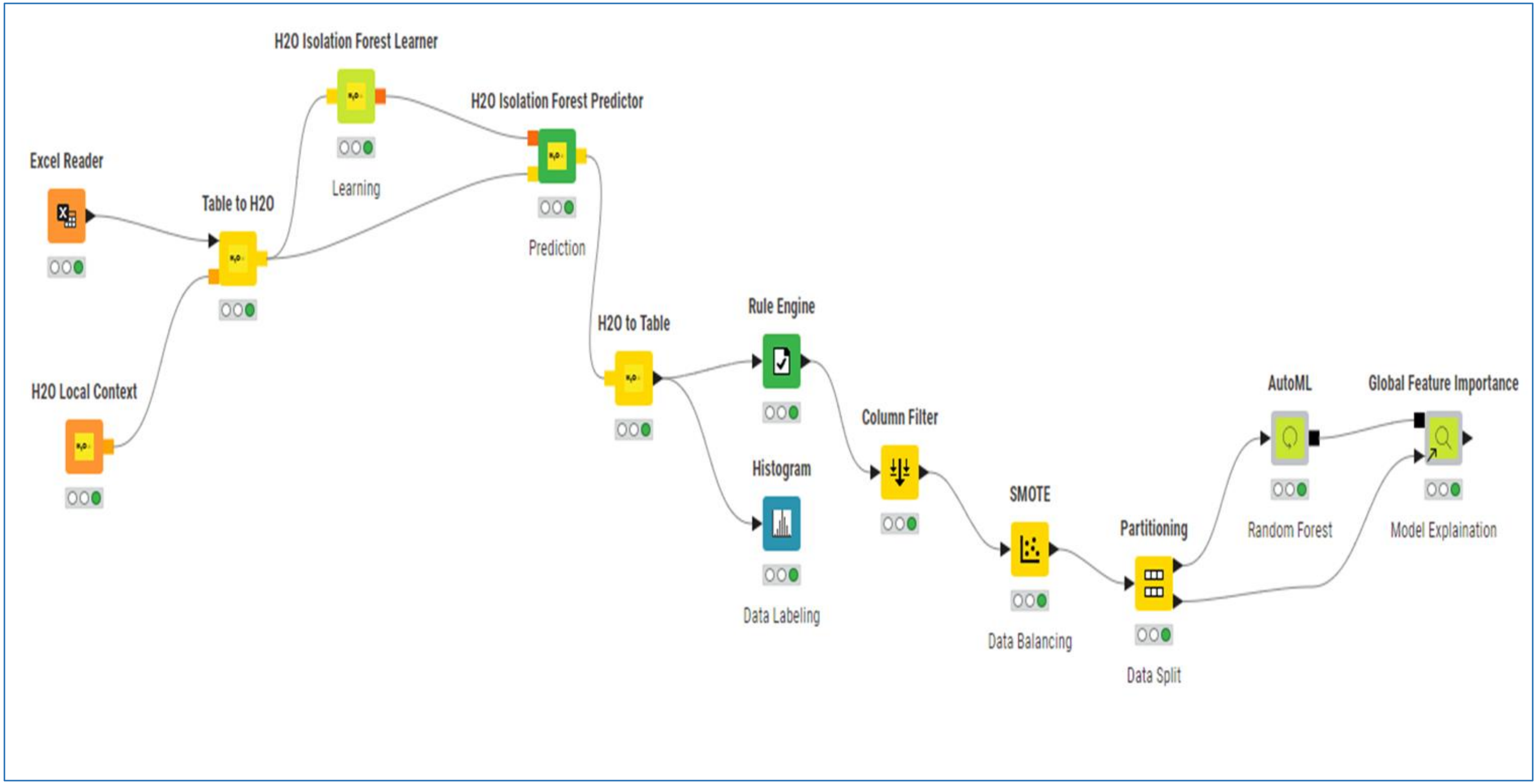


Figure III-1: KNIME Workflow

III.3. Model Results:

We apply the Isolation Forest model to an input dataset in order to predict anomalies or outliers using the H2O Isolation Forest Learner. We want H2O to build 100 trees with a maximum depth of 8. The prediction part which contains normalized anomaly score; the higher score is the more likely it is an anomaly. For the mean length part, the shorter one is the more likely an anomaly. Below is a sample of these predictions:

Prediction	Mean length
0,264840183	6,42
0,863013699	5,11
0,347031963	6,24
0,319634703	6,3
0,730593607	5,4
0,105022831	6,77
0,141552511	6,69
0,141552511	6,69
0,296803653	6,35
0,0456621	6,9
0,132420091	6,71
0,077625571	6,83
0,077625571	6,83
0,187214612	6,59
0,168949772	6,63
0,118721461	6,74
0,273972603	6,4
0,136986301	6,7
0,136986301	6,7

Table III-2: The Mean length listing

The outliers were identified using an organizational chart:

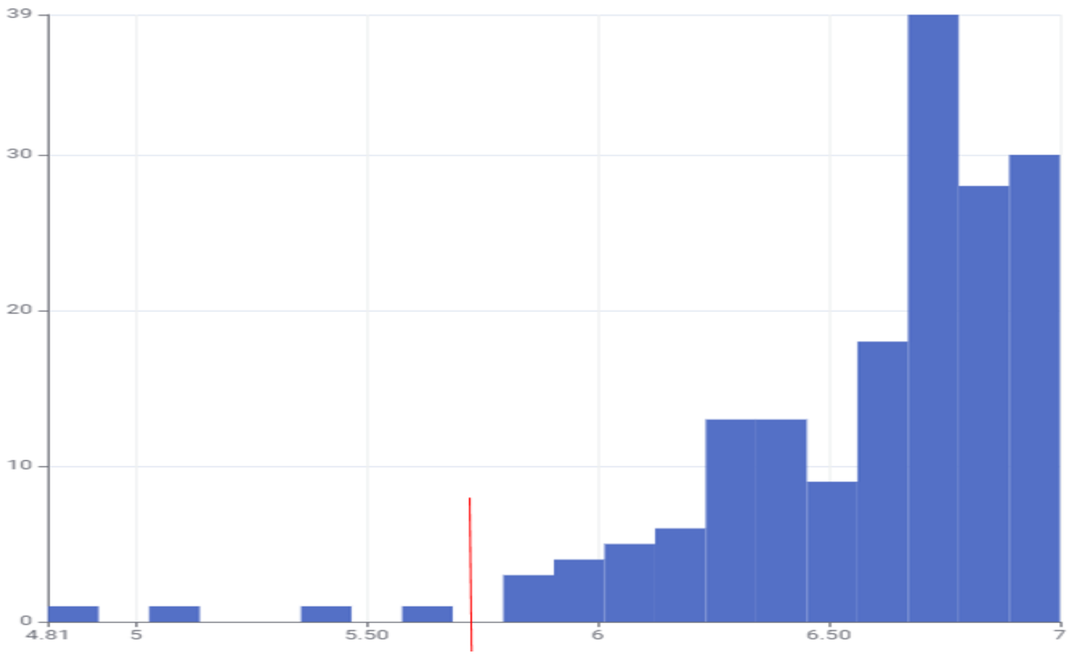


Figure III-2: Dataset Histogram

The histogram of the mean length shows that normal operating values have a mean length greater than 5.7.

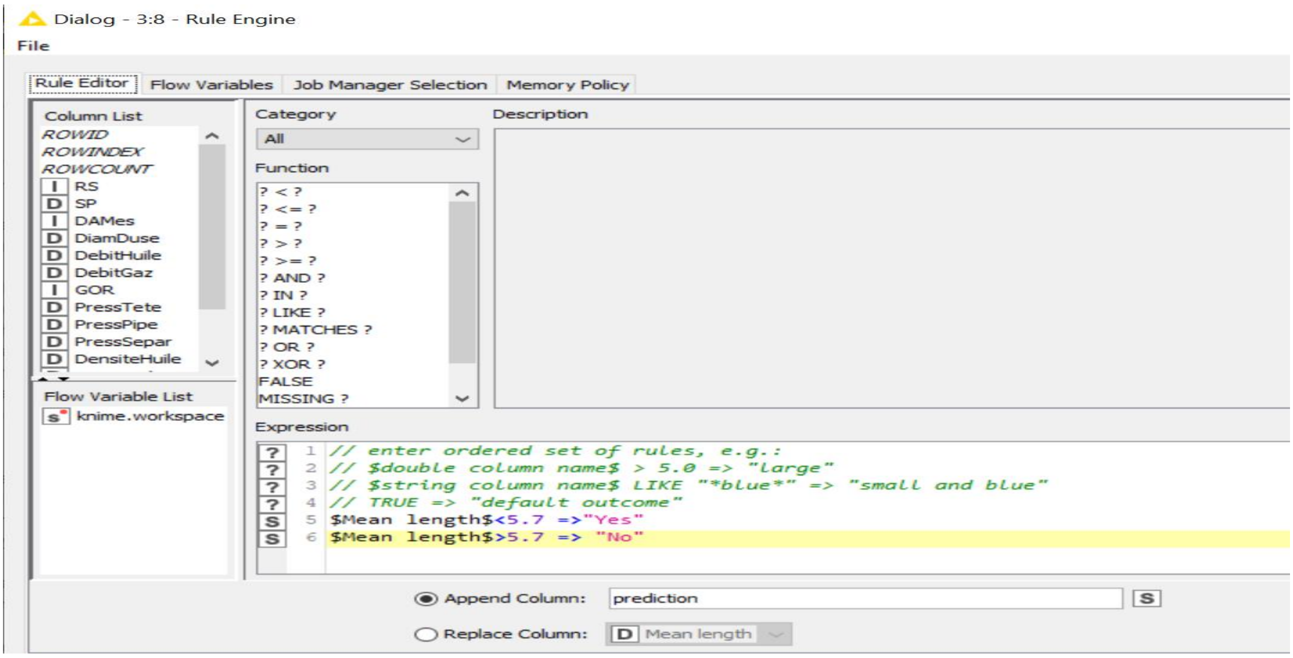


Figure III-3: Rule Engine Configuration

The KNIME Rule Engine used to smoothly isolate these observations by labelling them according to the previously defined separation condition. Any parameter value with a mean length less than 5.7 will be considered as an anomaly. The presence of a failure is labelled as "Yes"; and "No" for normal operation. The result is shown in Table III-3:

Label Failure	Quantity
No	168
Yes	4

Table III-3: Number of failures

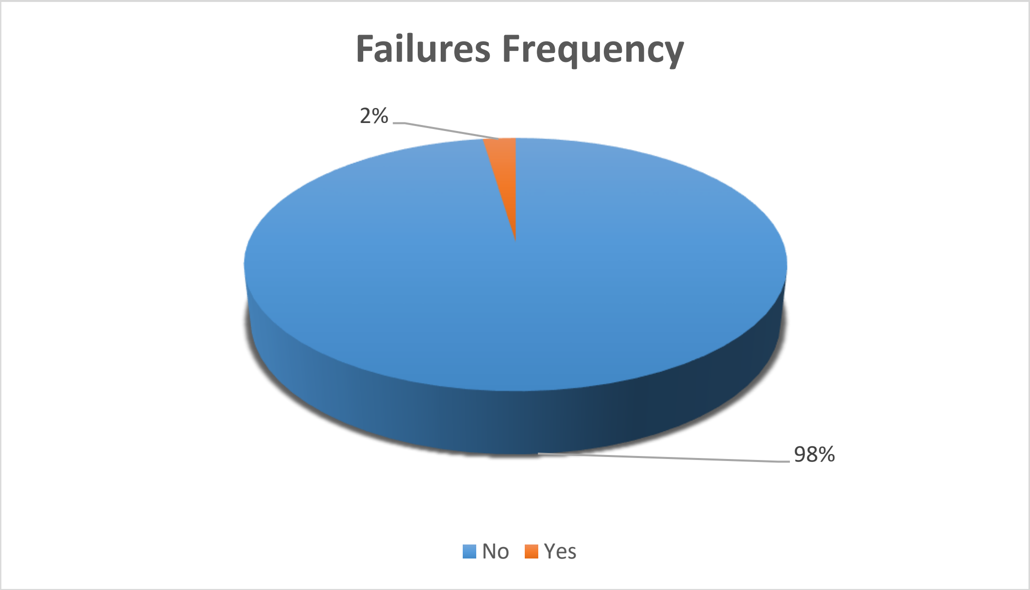


Figure III-4: Anomalies Frequency

The graph highlights the imbalance in the dataset, reflecting the low frequency of failures.

In order to solve this issue, we use the SMOTE approach to oversample the minority class “Yes” of our dataset using the KNN technique with five neighbours.

The new dataset after balancing with SMOTE:

Label Failure	Quantity
No	168
Yes	168

Table III-4: Number of Failures after balancing

The AutoML automatically trains various supervised machine learning models. It automates the whole machine learning cycle by performing some data preparation, parameter optimization with cross validation, scoring, evaluation and selection. Each model has a number of parameters to be tuned using a cross validation and evaluation metric. The prediction of all models is scored, and the best configuration of Random Forest algorithm was selected and exported with a classification accuracy of 100%: (Accuracy)=100%, AUC =1.

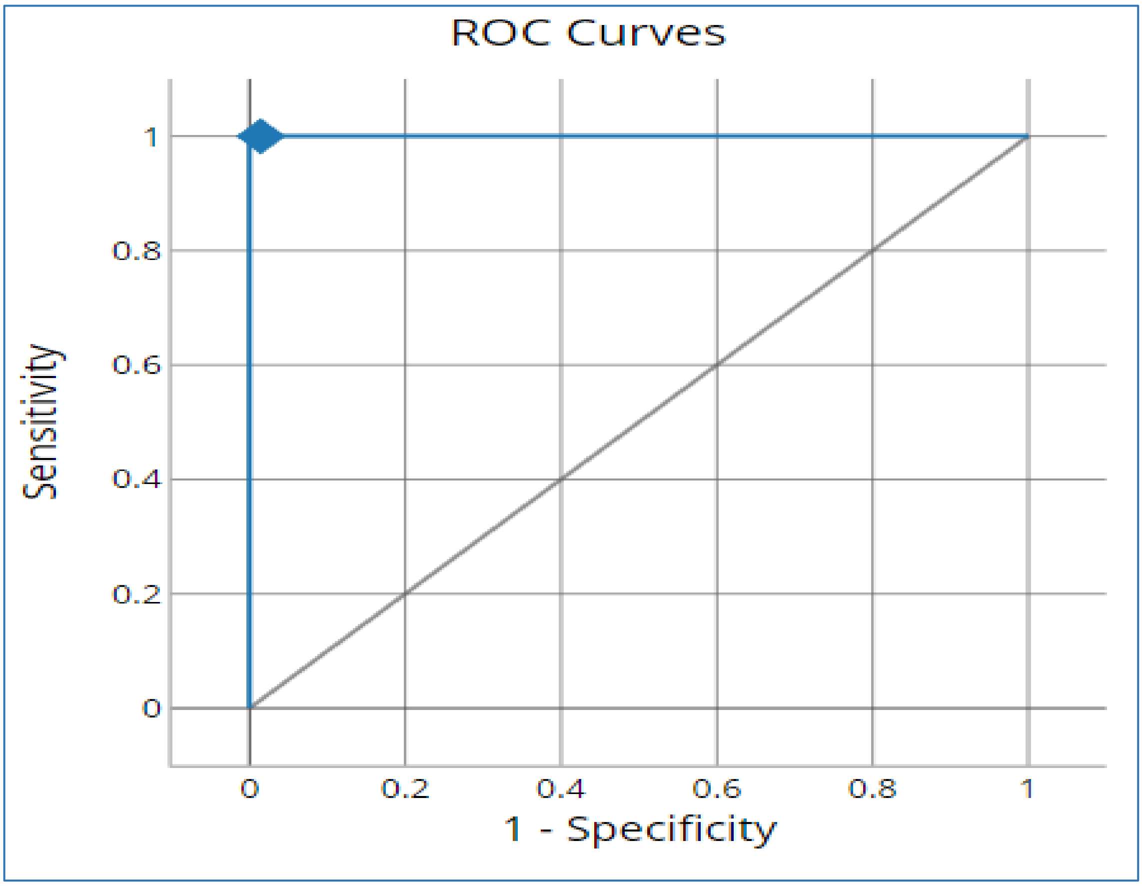


Figure III-5: ROC Curve Results

This indicates a high degree of predictive accuracy for failures as well as a high degree of selective ability for false positives and false negatives.

We proceed to calculate the overall feature importance for the classification models. It is possible to use a surrogate model to explain models such as Generalized Linear Model (GLM), Logistic Regression, Decision Trees (DT), or Random Forests (RF).

Feature importance is calculated by counting how many times it has been selected for a split and at which rank (level) among all available features (candidates) in the random forest trees. A higher value indicates higher feature importance.

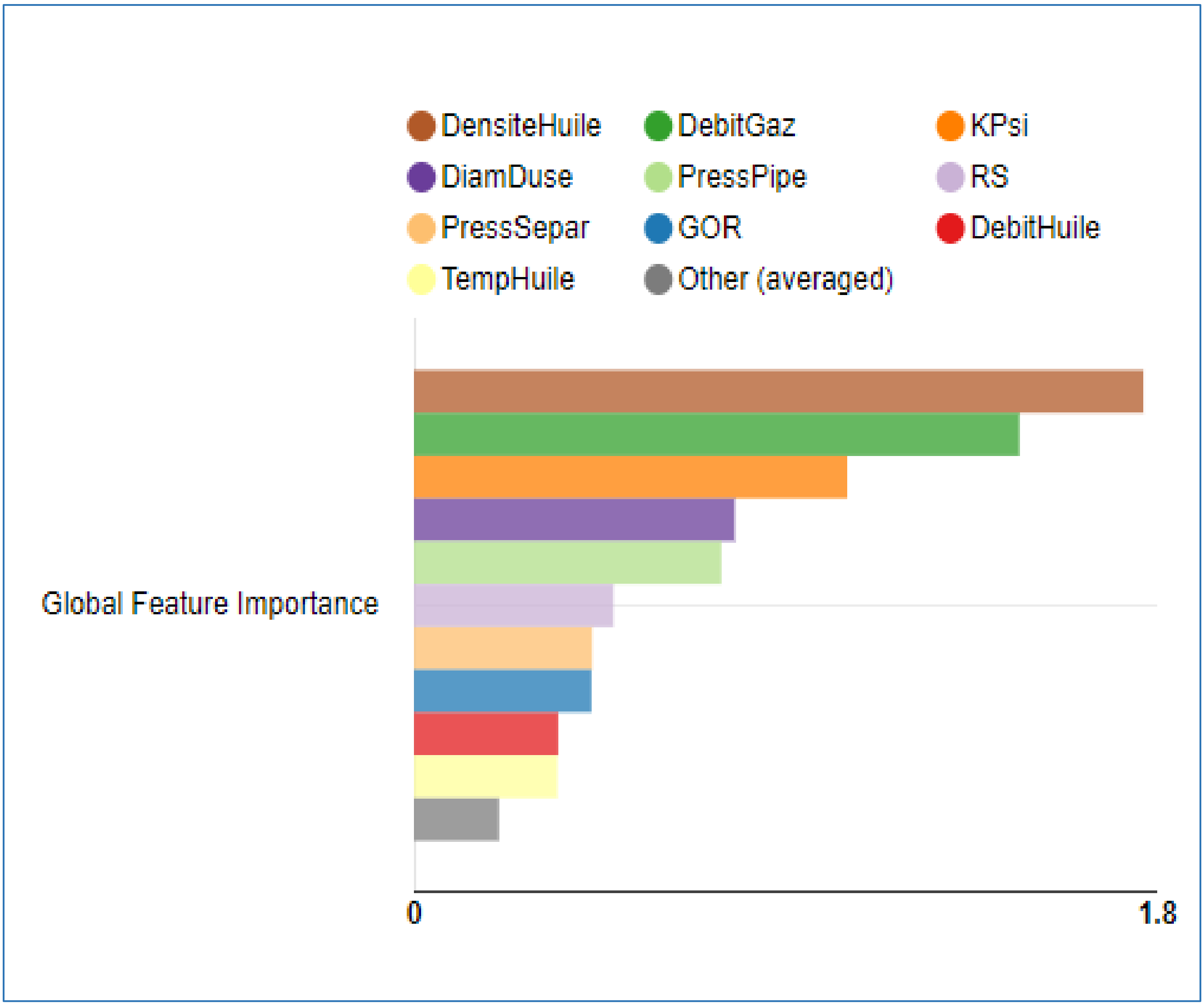


Figure III-6: Global Feature Importance

This comprehensive study provides valuable insights into the relative importance of several criteria in failure predicting, enabling operators to enhance system reliability and optimize maintenance schedules.

A Global Surrogate Decision Tree model was trained successfully with AUC equal to 0.974 to the original model predicted class of interest "prediction: Yes".

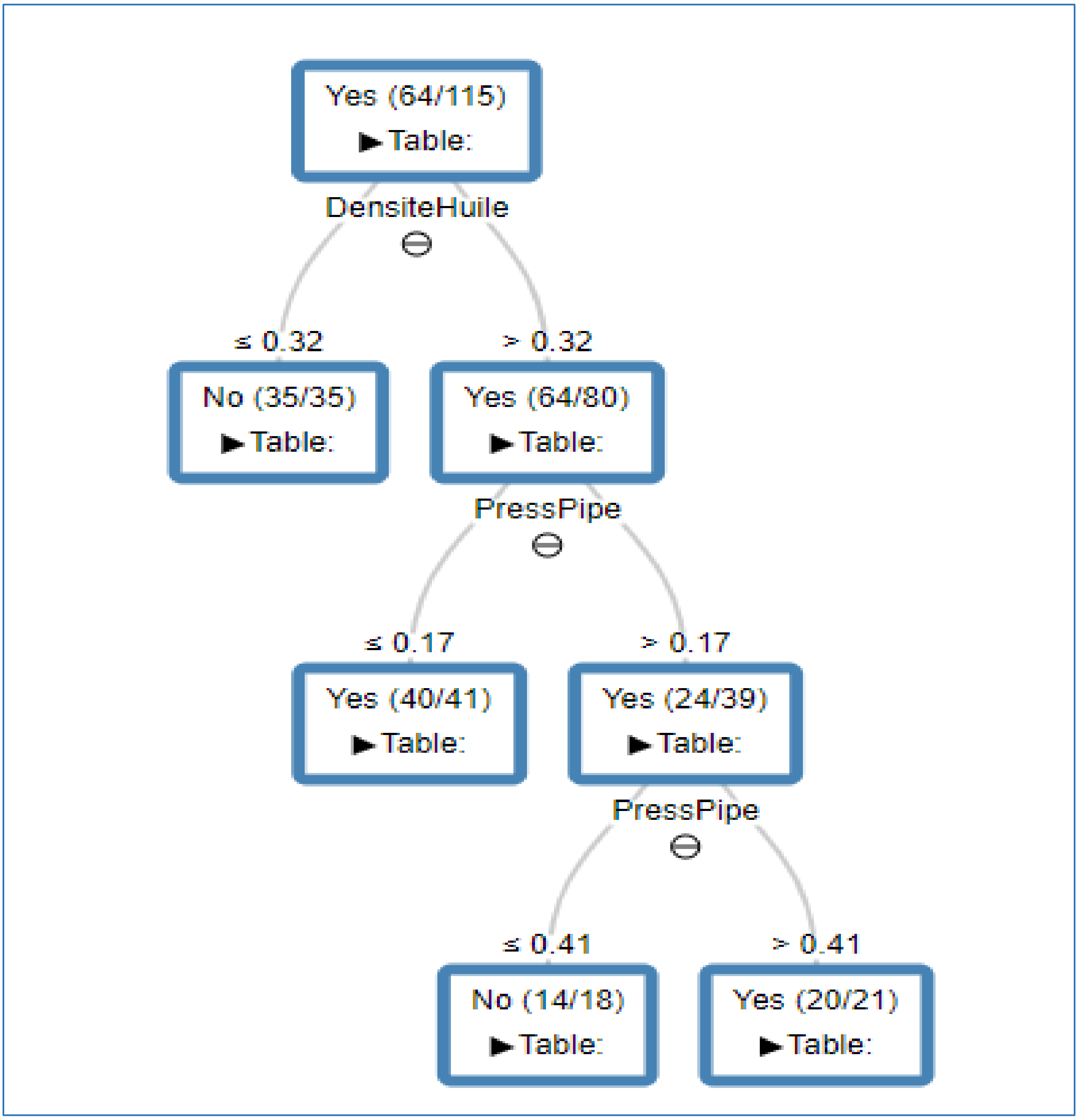


Figure III-7: Decision Tree

The objective is to demonstrate how to make decisions or simulate our model using a transparent decision tree, which exposes decision boundaries in a clear and understandable way, thus avoiding the opaque nature of a black box.

Based on the input features the model's behaviour is as follows:

- The decision tree starts by splitting the data based on “DensiteHuile” (oil density) feature, with 64 anomalies classified as “yes”.
- If the oil density is less than or equal to 0.32, the decision tree may reach a leaf node with 35 out of 35 classified as "no," indicating there are no anomalies.
- If the oil density is greater than 0.32, then the decision tree branches to another node that evaluates a different factor, which is: “PressPipe” (pressure in the pipe). This signifies that the model's decision-making process adapts based on varying conditions encountered during analysis and can assess additional parameters relevant to reservoir behaviour and production performance.
- When the pressure in the pipe is less than or equal to 0.17, this condition shows 40 out of 41 anomalies classified as "yes," requiring further verification. It may indicate Pump’s failures, insufficient flow or the need for additional pressure stimulation techniques.
- If the pressure in the pipe is greater than 0.17, then the decision tree branches to another value of pressure which is 0.41.

As a result, our decision tree model successfully evaluates the interaction between features while predicting the results of well production.

Conclusion:

This study aims to develop a comprehensive machine learning (ML) technique to handle the high frequency of ESP operating data. The model converts the data into useful information to predict various ESP failure modes. The analytical framework of real-time data collected by Supervisory Control and Data Acquisition (SCADA) systems makes it possible to go from reactive to proactive ESP monitoring, which may identify any well's problems before they develop.

This approach offers chances to enhance pump uptime, prolong ESP life, and boost oilfield economics. The results demonstrated valuable outcomes on the application of data analytics in ESP operations. Some of the key outputs for this study are:

- A literature review on the development of ESP failure prediction through time showed the need to use data analytics and accurately predict failures before their occurrences.
- We develop a machine learning model to predict failures for an unlabelled dataset utilizing unsupervised techniques.
- Isolation Forest is a highly effective method for detecting anomalies or outliers within datasets.
- Labelling unbalanced data using the SMOTE algorithm with oversampling of the minority class option.
- Automating supervised learning models by AutoML and identifying the best configuration model of the Random Forest algorithm.
- Evaluation of the model using ROC curve, with offers of 100% of classification accuracy (AUC).
- Developing a surrogate model using decision tree to explain decision rules and feature importance.
- We use KNIME codeless platform and H2O, best of breed machine learning, known for its powerful ML algorithms, superior quality and performance.

It is essential to underline that this study is customized to the specific characteristics of the collected data and variety of results can be obtained by using different models with different datasets to create opportunities to increase pump uptime, extend the life expectancy of ESPs and improve oilfield economics.

Recommendation:

- This study might be extended to examine more types of ESP failures by adding more data from wells in different fields to identify other patterns to various failures categories or unusual behaviour, which can indicate the time before failure (TBF) per days, such as motors vibration, corrosion or electrical problems.
- The study can be extended to include other artificial lift techniques, such as sucker rod and progressive cavity pumps, to recognize their failure types and how to anticipate them.
- This study may be improved by creating a model that provides actionable insights for each failure it identifies. This would allow the system to not only detect problems, but also give a preventative measure for the petroleum engineer.

References:

1) Diker. G, Frühbauer.H, Bisso Bi Mba, E. M. Development of a Digital ESP Performance Monitoring System Based on Artificial Intelligence. Abu Dhabi International Petroleum Exhibition & Conference, 2021.

2) Bai. Y., Li. J, Zhou. J, Li. Q, Sensitivity Analysis of the Dimensionless Parameters in Scaling a Polymer Flooding Reservoir. Transp Porous Med, 2008.

3) Alamu. O, A. Pandya, D. A, Warner, O, Debacker. I, ESP Data Analytics: Use of Deep Autoencoders for Intelligent Surveillance of Electric Submersible Pumps. Offshore Technology Conference, 2020.

4) Bates. R, Cosad. C, Fielder. L, Kosmala. A, Hudson. S, Romero. G, Shanmugam. V, Taking the pulse of production wells—ESP surveillance. Oilfield Rev, 2004.

5) Gabor. T, Electrical Submersible Pumps Manual: Design, Operations, and Maintenance; Elsevier: Burlington-Massachusetts, 2009.

6) Abdelaziz. M, Lastra. R, Xiao. J, ESP data analytics: Predicting failures for improved production performance. Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, UAE, November 2017.

7) Jahnke. P, Machine Learning Approaches for Failure Type Detection and Predictive Maintenance; Technische Universität Darmstadt, 2015.

8) Mordor Intelligence; Artificial lift systems market size, trends: Industry growth. Retrieved from https://www.mordorintelligence.com/industry-reports/global-artificial-lift- systems-market-industry, 2022.

9) Brown, Kermit E. "Overview of artificial lift systems." *Journal of Petroleum Technology* 34.10: 2384-2396, 1982.

10) Bates R, Cosad C, Fielder L, Kosmala A, Hudson S, Romero G. and Shanmugam V, "Taking the Pulse of Producing Wells-ESP Surveillance," Oilfield Review 16, no.2 :16-25, 2004.

11) Breit. S, and Ferrier. N, "Electric Submersible Pumps in the Oil and Gas Industry." Pumps & Systems. Wood Group ESP, Inc, Apr. 2008. Retrieved from http://www.pump-zone.com/topics/pumps/pumps/electric-submersible-pumps-oil-and-gas-industry

12) Tulsa, Spears & Associates, "Oilfield Market Report, 1999-2010, 2009.

13) Vandevier. J, Run-Time Analysis Assesses Pump Performance. Oil and Gas Journal 108.37; pp. 76-79, 2010.

14) Al-Jasmi. A, Nasr. H, Goel. H. K, Moricca. G, Carvajal. G. A, Dhar. J, Querales. M, Villamizar. M. A, Cullick. A. S, Rodriguez. J. A, Velasquez. G, Yong. Z, Bermudez. F, Kain. J, ESP "Smart Flow" Integrates Quality and Control Data for Diagnostics and Optimization in Real Time. Paper SPE 163809 presented at the SPE Digital Energy Conference, The Woodlands, Texas, USA, 05-07 March 2013.

15) Bertocco. R, Padmanabhan.V, "Big Data analytics in Oil and gas." Bain & Company, 26 March 2014.
http://www.bain.com/Images/BAIN_BRIEF_Big_Data_analytics_in_oil_and_gas.pdf

16) Derrick. T, Gavia. J, Jenkins. C, Oster. C, Sandlin. J, and Wright. K, Analytics beyond R2: Year one. Paper SPE 163713 presented at the SPE Digital Energy Conference and Exhibition, The Woodlands, TX, 5–7 March 2013.

17) Igwilo. K. C, Okoro. E. E, & Ubanatu. S, Comparative approach to optimum selection of artificial lift system. Petroleum & Coal, 2018.

18) Nguyen. T, Artificial Lift Selection Methodology for Vertical and Horizontal Wells in Conventional and Unconventional Reservoirs. In Artificial Lift Methods (pp. 317-347). Springer. Cham, 2020.

19) Takacs. G, Electrical Submersible Pumps Manual: Design, Operations, and Maintenance. Netherlands: Elsevier Science, 2009.

20) Takacs. G, Electrical Submersible Pumps Manual: Design, Operations, and Maintenance. United Kingdom: Elsevier Science, 2017.
21) Hughes. B, Submersible Pump Handbook, 2020.
22) Zhu. J, Zhang. H.Q, A review of experiments and modelling of gas-liquid flow in electrical submersible pumps. Energies,11, 180, 2018.
23) Abdelaziz. M, Lastra. R, & Xiao. J. J, ESP Data Analytics: Predicting Failures for Improved Production Performance. SPE Paper No. 188513, Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, UAE, 15 November 2017, https://doi.org/10.2118/188513-MS
24) Fakher. S, Khlaifat. A, Hossain. M. E, & Nameer. H, Rigorous review of electrical submersible pump failure mechanisms and their mitigation measures. Journal of Petroleum Exploration and Production Technology, 11(10), 3799–3814, 2021. https://doi.org/10.1007/s13202-021-01271-6
25) El Gindy. M, Abdelmotaal. H, Botros. K, Ginawi. I, Sayed. E, & Edris. T, Monitoring & Surveillance Improve ESP Operation and Reduce Workover Frequency. SPE Paper No.177926, Abu Dhabi International Petroleum Exhibition and Conference, 10 November, Abu Dhabi, UAE, 2015. https://doi.org/10.2118/177926-MS
26) Williams. A.J, Cudmore. J, Beattie. S, ESP monitoring – where is your speedometer? ESP workshop, Houston, Texas, 30 April – 2 May 2003.
27) Heninger, Milan, Salvatore F. Grande, and David D. Shipp. "Identifying and preventing ESP failures resulting from variable speed drive induced power quality issues, SPE, 2019.
28) Linda Tucci, What is machine learning and how does it work? In-depth guide, 2023. https://www.techtarget.com/searchenterpriseai/definition/machine-learning-ML
29) Archit Kane, Ashutosh S. Kore, Advait N. Khandale, Sarish S. Nigade, Predictive Maintenance using Machine Learning, May 2022.
30) Julianna Delua, Supervised vs. Unsupervised Learning: What's the Difference, March 2021.
31) Cao et al. Q, Cao. R, Banerjee. S, Gupta. J. Li. W, Zhou. B, Jeyachandra, Data driven production forecasting using machine learning, 2016.
32) A. Andrianova, M. Simonov, D. Perets, A. Margarit, D. Serebryakova, Y. Bogdanov, A. Bukharev, Application of Machine Learning for Oilfield Data Quality Improvement, 2018.
33) Anirbid Sircar, Application of machine learning and artificial intelligence in oil and gas industry, 2021.
34) Kotu.V, Data Science: Concepts and Practice, 2nd ed, Morgan Kaufmann Publishers: Cambridge, MA, 2018.
35) Y. Liu, B. Coffman, N. McMahan, and F. Alisdair, "Bakken unconventional well gas-oil ratio (GOR) behavior characterization,"in Paper URTeC 5358, presented at the SPE/AAPG/SEG unconventional resources technology conference, pp. 26–28, Houston, Texas, 2021.
36) Fei Tony Liu, Kai Ming Ting, Zhi-Hua Zhou, Isolation Forest, January 2009.
37) D. Elreedy et al.A, Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance Inf. Sci,2019.
38) Adele Cutler, David Richard Cutler, John R Stevens, Random Forest in Machine Learning, January 2011.
39) Lior Rokach, Oded Maimon, Decision Tree, January 2005.
40) Brandon Wohlwend, Machine Learning — Evaluating classification models, 16 July 2023.
41) Christoph Molnar, Interpretable Machine Learning: *A Guide for Making Black Box Models Explainable,* 14 Octobre 2018.
42) Berthold, Michael R., et al. "KNIME-the Konstanz information miner: version 2.0 and beyond." *AcM SIGKDD explorations Newsletter* 11.1: 26-31, 2009.
43) Cook, Darren. *Practical machine learning with H2O: powerful, scalable techniques for deep learning and AI.* " O'Reilly Media, Inc.", 2016.

44) Waring, Jonathan, Charlotta Lindvall, and Renato Umeton. "Automated machine learning: Review of the state-of-the-art and opportunities for healthcare." *Artificial intelligence in medicine* 104: 101822, 2020.