Scientific Contributions Oil & Gas, Vol. 48. No. 2, August: 69 - 94

SCIENTIFIC CONTRIBUTIONS OIL AND GAS
Testing Center for Oil and Gas
LEMIGAS

Journal Homepage:http://www.journal.lemigas.esdm.go.id
ISSN: 2089-3361, e-ISSN: 2541-0520

The Application of Machine Learning (DT-Chan-Performance) in Determining Idle Well Reactivation Candidates at PT. Pertamina EP Regional 4 Zone 11 Cepu Field

Sayoga Heru Prayitno[1], Boni Swadesi[1], Hariyadi[1], Damar Nandiwardhana[1], Herlina Jayadianti[2], Indah Widiyaningsih[1], Ndaru Cahyaningtyas[1] and Geovanny Branchiny Imasuly[1]

[1]Department of Petroleum Engineering, Faculty of Mineral Technology,
National Development University "Veteran" Yogyakarta
Padjajaran Street, Sleman, Yogyakarta 55283, Indonesia.

[2]Department of Information Systems, Faculty of Industrial Engineering,
National Development University "Veteran" Yogyakarta
Janti Street, Sleman, Yogyakarta 55281, Indonesia

Corresponding author: sayogaheru@upnyk.ac.id

Manuscript received: April 04th, 2025; Revised: April 18th, 2025
Approved: May 8th, 2025; Available online: May 19th, 2025; Published: May 20th, 2025.

ABSTRACT - Indonesia faces a significant challenge in achieving its goal of oil production 1 million barrels of oil per day by 2030, particularly as it relies on old fields or mature fields (brownfields) to extract remaining hydrocarbons. One of the strategies involves reactivating of idle wells in Cepu field, managed by PT. Pertamina EP Regional 4 zone 11. This study focuses on identifying suitable candidates for reactivation through combination of research, innovation and production-focus analysis. The process begins with problem definition, aiming to understand the factors influencing idle wells and review recent advancements in reactivation prediction. Data were collected from both primary and secondary sources covering period 2018-2023. The next stage is implementing Machine Learning (ML), specifically Decision Tree (DT) model, to overcome problems related to data accuracy and complexity. A web application was developed to support decision-makers in selecting wells with high reactivation potential which can provide the best solution of increasing oil recovery. The research results show a high success rate on Accuracy Under Curve and Receiver Operating Curve score of 0.99, indication strong predictive capability. Using entropy-based analysis, two potential wells were identified for reactivation for improvement. These wells were further evaluated using Chan Diagnostic and Production Performance analysis.

Keywords: machine learning, reactivation, idle well, decision tree, increased oil recovery.

How to cite this article:

Sayoga Heru Prayitno, Boni Swadesi, Hariyadi, Damar Nandiwardhana, Herlina Jayadianti, Indah Widiyaningsih, Ndaru Cahyaningtyas and Geovanny Branchiny Imasuly, 2025, The Application of Machine Learning (DT-Chan-Performance) in Determining Idle Well Reactivation Candidates at PT. Pertamina EP Regional 4 Zone 11 Cepu Field, Scientific Contributions Oil and Gas, 48 (2) pp. 69-94. DOI org/10.29017/scog.v48i2.1657.

INTRODUCTION

The Indonesian upstream oil and gas industry is experiencing serious challenges. According to the 2022 SKK Migas annual report, the industry has a mission to achieve oil production of 1 million barrels per day by 2030 (Marsudi 2021). Achieving this is a tough challenge because it relies on old or mature fields (brownfields), which aims to extract remaining hydrocarbons rather than developing new wells. Creating new wells requires significant investments in drilling and production facilities and often involve higher uncertainty in terms of recoverable resources (hydrocarbons). To address this, research and innovation efforts are being directed toward enhancing production from Idle well (IW) in mature fields to increase oil recovery and to support national energy security (Ministry of Energy and Mineral Resources 2020; Sun et al. 2019). One of the mature fields in Indonesia is exploited by SKK Migas to realize the mission of 1 million barrels of oil per day by 2030. The object of this research is Cepu field which experiences declining production because of extensive oil and gas production activities over the years (Saptowulan et al. 2022; Putra et al. 2022; Chan 2009).

Conventional multi-screening approaches for Idle well reactivation depend on human assessment and basic statistical methods, rendering them susceptible to bias, inconsistency, and errors. The complexity and unreliability of historical well data from 2018 to 2023 render the manual identification of reactivation candidates both arduous and inefficient (Nguyen & Chan 2005; Nguyen et al. 2004). To overcome these constraints, Artificial Intelligence (AI) and Machine Learning (ML) techniques, particularly Decision Trees (DT) models, provide a data-driven, automated, and precise alternative way for identifying idle wells suitable for reactivation (Sandha et al. 2005). Decision Tree models are capable of handling extensive datasets, noisy and incomplete data, and complex well conditions, thereby facilitating more dependable predictions estimates of reactivation potential. The incorporation of AI-driven decision-making frameworks is crucial for enhancing oil recovery, operational efficiency, and national energy security (Li & Chan 2010; Li & Chan n.d.; Alkinani et al. 2019).

The Cepu Field utilizes several lift methods in its production operations, including sucker rod pump (SRP), electric submersible pump (ESP), and natural flow. Among these, the SRP method is the most dominant, reflecting the typical characteristics of mature fields. A significant number of wells in the Cepu Field remain idle due to high water cut (HWC) issues. Currently, there are 11 active structures (780 wells) and 39 non-active structures (566 wells). The historical data obtained from 2018 to 2023 is inconsistent or incomplete making it difficult to identify suitable reactivation candidates manually. Previous research suggests that idle well reactivation candidates can be identified through a multiple-screenings evaluation approach (Bangert 2019; Amin Nizar et al. 2019).

Idle well evaluation was carried out using a multi-screening method for Chan Plot Diagnostic analysis, production well history, and well integrity quantitatively based on well history and well diagrams in the Sangasanga field by carrying out proper reactivation (Putra et al. 2022). Apart from that, from the cross-section and the distribution of wells, it is possible to predict which wells have potential for maintenance or reactivation (Candra et al. 2024). Selected wells identified as candidates will be reactivated according to a pre-established program for the Cepu field (Ardi & Suhascaryo 2022). A similar approach was also applied in research by (Alfarizi et al. 2023; Alfarizi et al. 2023).The use of multi-screening methods for idle well evaluation presents limitations, in terms of accuracy in determining reactivation candidates and data complexity. These methods struggle to effectively process large, inconsistent datasets, making it difficult to accurately identify reactivation candidates (Mukhanov et al. 2018). The core issue is the inconsistency of human assessment and the absence of criteria in appropriate classification patterns (Mukhanov et al. 2018), both of which are vital in determining Idle well reactivation candidates in interpretation or production analysis. Although relying on professional human judgment is common, there is significant value in seeking consistency, and it may not be easy to differentiate clearly (Garcia et al. 2019). As a solution, this research adopts a machine learning algorithm, namely Decision Trees model, which offers reliable classification of reactivation candidates. Decision Tree model is effective for idle well reactivation because it classifies data to identify reactivation candidates. It can efficiently handle complex datasets, including those that are noisy and incomplete and it produces classification patterns that are easy to understand and interpret based on key attributes. (Wardhana et al. 2021).

Table 1
Comparison of multi screening method & decision trees

Criteria	Multi Screening	Decision Trees
Determination of Reactivation Candidates	Historical data that is simple and easy for manual processing.	Complex historical data, as well as creating precise classification models.
Accuracy of reactivation determination	It tends to be lower with statistical techniques and manual data analysis.	It tends to be higher with machine learning techniques.
Data inconsistency	It is more susceptible to errors due to inaccurate data, which may lead to the incorrect classification of an idle well as suitable for reactivation when it is not.	It is more durable because it can improve data quality and reduce the impact of inconsistencies in producing accurate predictions.
Data complexity	Using simple data is easy to process manually because it is difficult to understand the relationship between parameters, and it is inaccurate when determining reactivation candidates idle well under complex conditions.	Using complex and diverse data produces more accurate predictions for determining idle well reactivation candidates under complex conditions.
Ability to study patterns and trends	No, because the idle well reactivation decision is based on human judgment.	Yes, because idle well reactivation decisions are based on strong data and analysis.
Susceptible to bias and human error.	Yes, it requires human interpretation of results.	No, able to make predictions based on data and not human interpretation
Profit	Fast, efficient, and relatively cheap based on knowledge and experience.	More accurate and able to capture complex relationships in data with machine learning knowledge.
Deficiencies	Less Optimal	Optimal

The next stage is based on drilling methods with completion, workover, stimulation, and replacement of productive zones (Haryanto et al. 2019). Additionally, the selection and design of artificial lift systems must be tailored to specific conditions of each well, including its location and reservoir characteristics. This shows that there is an opportunity to increase production from the use of Idle well if it is implemented appropriately with the application of Artificial Intelligence , specifically Machine Learning using the Decision Tree model). Unlike conventional software engineering, where rules are explicitly defined, AI-driven approaches can analyze large datasets at high speed and scale, identify patterns and generate accurate predictions (Garcia et al. 2019). This capability is particularly relevant as the oil and gas industry undergoes a significant transformation driven by automation and digital technologies (Bizhani & Kuru 2022; Utomo 2011). One of the studies by Garcia et al. (2019), applied Machine Learning, namely Chan Plot Signature Identification as a Practical Machine Learning Classification Problem. The research utilized supervised learning algorithms, with a particular focus on Support Vector Machines (SVM) due to their effectiveness on handling nonlinear relationships and classification tasks. In addition to SVM, other models tested included the nearest neighbours, decision trees, random forests, logistic regression, and naïve Bayes. All models demonstrated perfect classification performance, achieving a score of 1, indicating their effectiveness in accurately identifying specific diagnostic classes (He et al. 2021; Olatunji et al. 2010).

This research focuses on the application of Artificial Intelligence, namely Machine Learning using Decision Tree model to determine Idle well reactivation. Historical data on Idle well and previous reactivation decisions are collected and prepared for model training. Relevant variables or attributes to support reactivation predictions include operating parameters, production history, physical conditions, and other well-specific attributes. The Decision Tree model is trained using prepared historical data to learn patterns and make accurate predictions regarding which wells are suitable for reactivation (de Ville 2013). The prediction results generated by Decision Tree model assist in identifying which wells should be prioritized for reactivation, providing the best solution to increase oil recovery by targeting the remaining hydrocarbons, particularly from idle wells. One of the key challenges in the Cepu field is high water cut (HWC), which often contributes wo wells becoming idle. This Decision Tree model can classify these wells into various categories, such as candidates for reactivation, those requiring maintenance, and those that should be abandoned, enabling more informed and strategic decision-making. Besides looking for the cause, the drilling method and the pump design should be optimized. Overall, this approach increases reactivation efficiency and enhances the ability to detect regional trends improving the model's recall based on the training dataset (Chakra et al. 2013).

METHODOLOGY

This research adopts a quantitative approach to explore, describe, and explain data through statistical methods. This methodology helps reduce errors, increase efficiency, and produce higher-quality research results, as shown in Figure 1. The following are sequential steps undertaken in this research:

Problem identification

The first step involves defining the problems in determining idle well reactivation candidates to obtain accurate and consistent predictions. This includes identifying and understanding the key factors that influence Idle well status in the Cepu field. Several problems commonly associated with this well include the classification as old well category, the need of workover service, sidetracking, being on hold, and ASR (P&A). Additional issues include pump replacements or repairs, layer changes, string fishing and replacement, wellbore cleanout, perforation additions, acidizing, and water shut-off treatments. Addressing these factors is essential increasing the accuracy of selecting candidates of well reactivation.

Data collection

Data collection was carried out using both primary sources (direct field data) and secondary sources (existing data) from the period 2018 to 2023, to support the development of the machine learning model.

Machine learning model development (Decision tree)

Data collection (Idle well)

This study utilized field data, including well history, to support the identification of reactivation candidates. The initial dataset includes Idle well parameters, drilling methods, and pump design. It has been carried out due to an approach to identify potential wells for reactivation based on various measured and observed parameters and factors.

Data preprocessing

Data Preprocessing was carried out after the data collection with the aim of processing in the form of Idle Well Status Database to determine potential Idle well candidates suitable for reactivation and interpretation. This step ensures the data readiness for optimal use, minimizing errors, and maximizing the analysis results with algorithms machine learning. The preprocessing activities included data cleaning, label encoding, data normalization, and train-test split, form cleaning dataset, converting categorical labels into numeric values, normalization (scaling data from 0 to 1), and splitting data from of 618 Idle well into 70% for training (approximately 432.6 data) and 30% for testing (approximately 185.4 data). Further processing steps modelling using Decision Tree, Confusion Matrix, Accuracy Area Under Curve (AUC), and Accuracy Receiver Operating Characteristic (ROC).

Potential reactivation candidate

The identification of potential reactivation candidates is based on two key factors namely Potential Status and Well Area. The combination these two factors enables oil and gas companies to make informed decisions and effectively prioritize well reactivation.

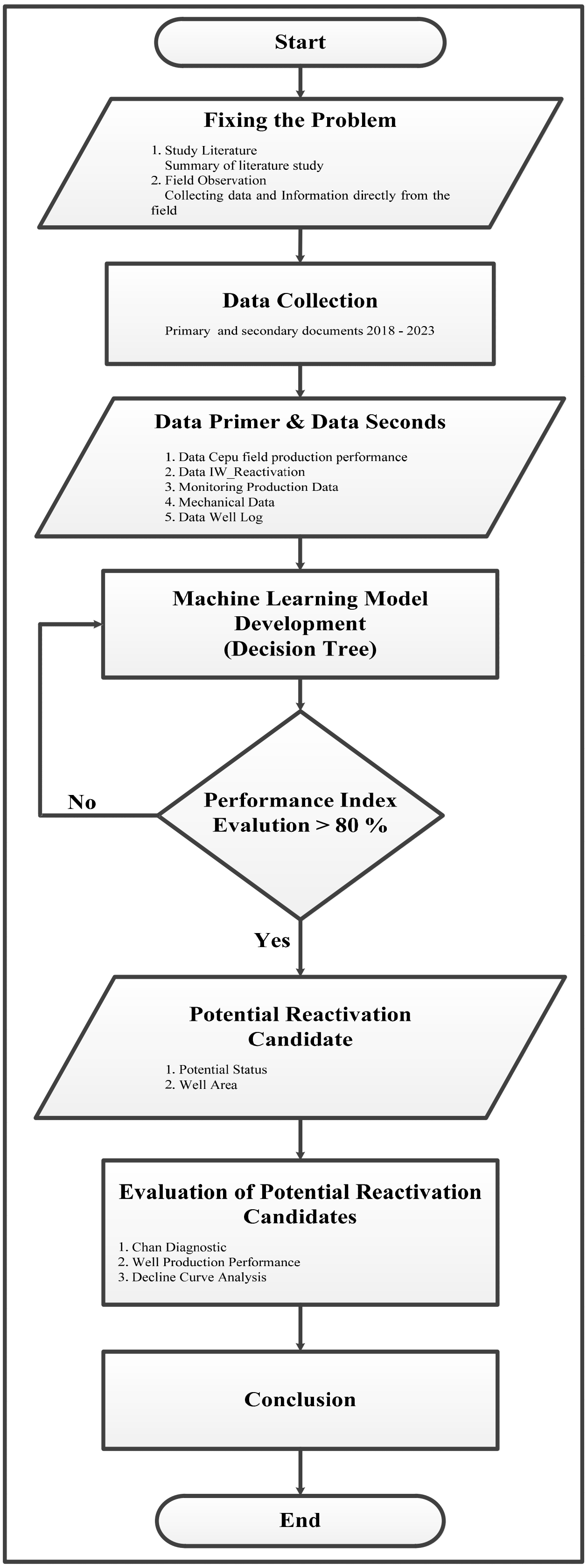


Figure 1
General flowchart

Potential status

Potential status is based on the results of the entropy model. In the visual output, wells classified as having no potential are marked in red with a value of 0, while wells with reactivation potential are shown in light green with a value of 1.

Well area

The yield from each well is based on the area against Well Cum Prod (Np, MBO) and HC Remaining Potential (Oil, MBO). The wells are ranked from the highest to the lowest based on their values. A color scale table is used to visually differentiate well performance, helping identify uneconomic wells as well as those with high potential for reactivation.

Evaluation of potential reactivation candidate

After determining reactivation well using Decision Tree model, the Chan Diagnostic Plot is applied to analyze the mechanism of excessive water and gas production in oil production wells. This method observes intermediate log-log plots of Water Oil Ratio (WOR) and WOR derivative over time to detect phenomena such as water coning, water channelling, or gas coning. These insights are used alongside a flowchart for selecting the most suitable reactivation candidates.

RESULT AND DISCUSSION

At this stage, research and innovation development is carried out to focus on the production of Idle wells on the field mature in the Cepu field. It aims to increase oil recovery, with the implementation of machine learning and decision tree model in determining Idle well reactivation candidates in the Cepu field based on potential status, namely potential or not potential. The analysis and design process aims to evaluate whether the objectives have met by developing a tool namely Idle Reactivation Dashboard which serves as a decision-support system to improve oil production. Result of Decision Tree model is further validated using Chan Diagnostic and Production Performance analysis to assess the feasibility of well reactivation.

Machine learning development (Determining idle well reactivation candidate)

When creating the program, the idle well reactivation dashboard displayed important information from data sources into one easy part for determining reactivation candidates with the model *decision tree.* This program can be used to visualize data and ease the users to understand the complex relationships in the data.

Decision tree classifier idle well reactivation

Decision Tree Classifier Diagnostics Idle Well Reactivation was carried out to analyze data from potential and non-potential wells, The process is divided into two processes: Decision Tree Classifier and Decision Tree Classifier Diagnostics Plot. The first step for the Decision tree Classifier is to upload data. in this case, it can help understand the characteristics of the data. The second step is the Decision Tree Classifier Diagnostics Plot, which is used to identify patterns and trends, visualize data to be processed, and select relevant features for modelling; each of these processes is describes in detail below.

Decision tree classifier diagnostics

This initial stage is carried out to display data and statistical data with size limit of 200MB for *each file*. Based on Figure 2, the uploaded data is previewed by displaying the five rows of data along with summary statistics for each attribute in the data.

Decision tree classifier diagnostics plot

This stage follows the Decision Tree classification and involves uploading data for diagnostic analysis. It includes evaluating feature diagnostics, performing data cleaning and replacement, applying label encoding, and generating outputs such as the confusion matrix, entropy values, AUC scores, and ROC curves. Each of these components provides insights into feature importance and model performance.

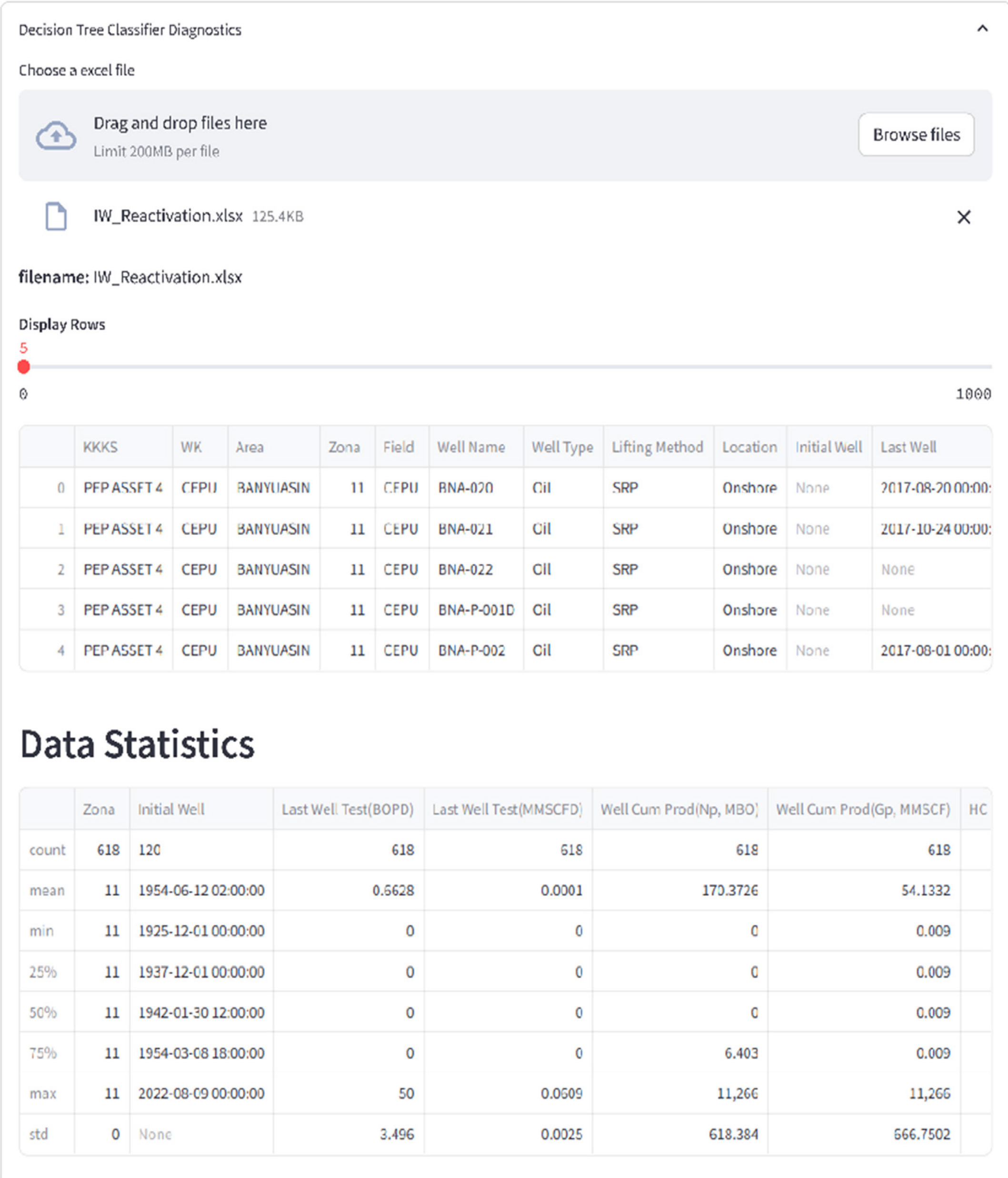
Based on Figure 3, the Potential Status Visualization displays the distribution of data categories along the X-axis. There are 228 records classified as "Potential," accounting for 37.06%, and 390 records classified as "Not Potential," representing 62.94%.

Figure 4 presents the Recommendation Visualization, where on the X axis categorizes different reactivation recommendations. The data includes Hold with 480 wells (77.67%), Well Service (WS) with 19 wells (3.07%), Old well with 103 wells 16.02%), Workover (WO) with 13 wells (2.10%), and Plug and Abandonment with 7 wells (1.13%.) The information shown in these two figures serves as the foundation for building the prediction model used to identify idle well reactivation candidates.

Remove & Replace Data

The primary objective of the data removal and replacement process is to clean and standardize the dataset, thereby improving data quality, simplifying analysis, and enhancing overall model performance. As shown in Figure 5, specific attributes were removed based on their limited relevance or redundancy.

The attributes in the deleted data are *Kontraktor Kontrak Kerjasama* (KKKS), *Wilayah Kerja* (WK), Zona, Field, Location, Well Type, Initial Well, Last Well Test (BOPD), Last Well Test (MMSCFD), Well Cum Prod (Gp, MMSCF), HC Remaining Potential (Gas, MMSCF), Expected Gain (Qgi, MMSCFD) and Expected Gain (Qoi, BOPD). Meanwhile, the attributes that are not deleted are Well Name, Area, Lifting Method, Well Cum Prod (Np, MBO), HC Remaining Potential (Oil, MBO), Doability, Recommendation, and Potential Status. Following the attribute removal, in-place formatting was applied. For example, in the Lifting Method and Do ability columns, missing values ("–") were replaced with "Jet Pump". Do ability values were standardized as: L for Light Issue, M for Normal/ Moderate Issue, and H for Heavy Issue, ensuring consistency throughout the dataset.



	KKKS	WK	Area	Zona	Field	Well Name	Well Type	Lifting Method	Location	Initial Well	Last Well
0	PEP ASSET 4	CEPU	BANYUASIN	11	CEPU	BNA-020	Oil	SRP	Onshore	None	2017-08-20 00:00:
1	PEP ASSET 4	CEPU	BANYUASIN	11	CEPU	BNA-021	Oil	SRP	Onshore	None	2017-10-24 00:00:
2	PEP ASSET 4	CEPU	BANYUASIN	11	CEPU	BNA-022	Oil	SRP	Onshore	None	None
3	PEP ASSET 4	CEPU	BANYUASIN	11	CEPU	BNA-P-001D	Oil	SRP	Onshore	None	None
4	PEP ASSET 4	CEPU	BANYUASIN	11	CEPU	BNA-P-002	Oil	SRP	Onshore	None	2017-08-01 00:00:

Data Statistics

	Zona	Initial Well	Last Well Test(BOPD)	Last Well Test(MMSCFD)	Well Cum Prod(Np, MBO)	Well Cum Prod(Gp, MMSCF)	HC
count	618	120	618	618	618	618	
mean	11	1954-06-12 02:00:00	0.6628	0.0001	170.3726	54.1332	
min	11	1925-12-01 00:00:00	0	0	0	0.009	
25%	11	1937-12-01 00:00:00	0	0	0	0.009	
50%	11	1942-01-30 12:00:00	0	0	0	0.009	
75%	11	1954-03-08 18:00:00	0	0	6.403	0.009	
max	11	2022-08-09 00:00:00	50	0.0609	11,266	11,266	
std	0	None	3.496	0.0025	618.384	666.7502	

Figure 2
Display upload data IW reactivation

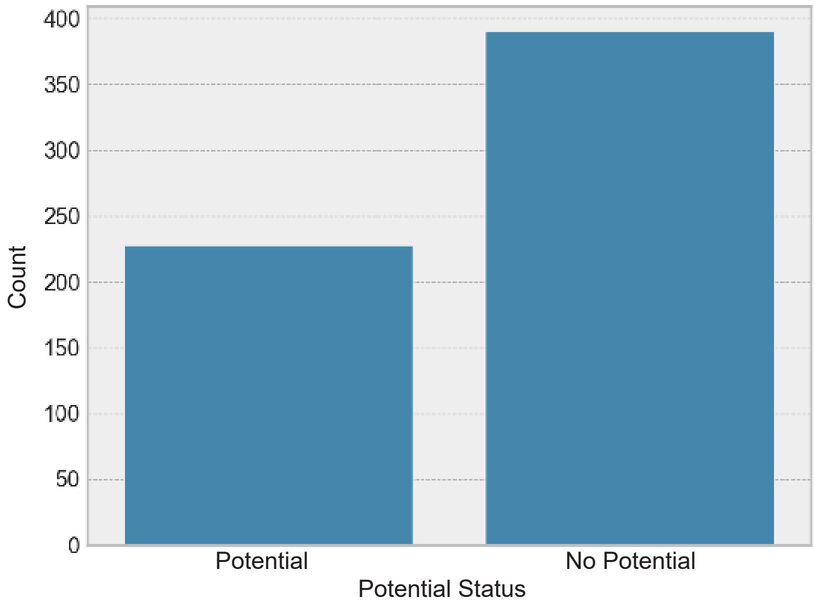


Figure 3
Potential status visualisation results

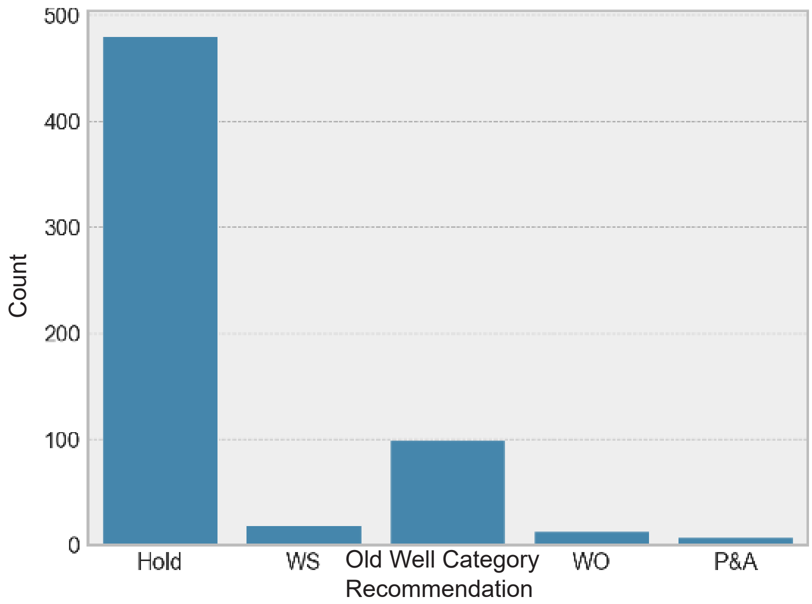


Figure 4
Recommendation visualisation results

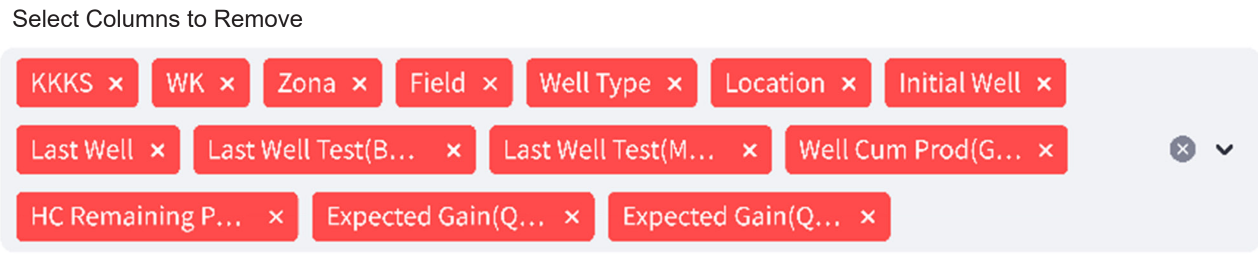


Figure 5
Display of deleted attributes

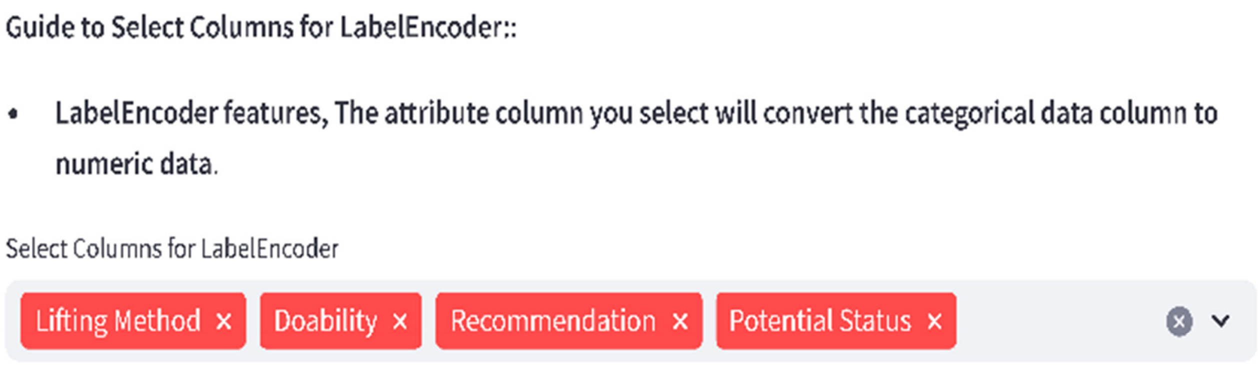


Figure 6
Display of select labelendcoder attributes

Labelendcoder

Labelendcoder process converts categorical data columns into numerical data, saving computing time and simplifying analysis because machine learning algorithms cannot process data in text form to work together with categorical data. In the Idle Well Reactivation data, there are four category-type data attributes: Lifting Method, doability, recommendation, and potential status. This process is not performed for well name and area attributes because it has many unique variations for each name. Encoding these would result in numerically meaningless representations that could negatively affect model performance and prediction accuracy. Based on Figure 6, the results obtained from the process LabelEndcoder are illustrated in Figure 7 and Figure 8.

Label mapping for Lifting Method:

{
"0" : "ESP"
"1" : "Jet Pump"
"2" : "Natural"
"3" : "SRP"
}

Label mapping for Doability:

{
"0" : "H"
"1" : "L"
"2" : "M"
}

Label mapping for Recommendation:

{
"0" : "Hold"
"1" : "Kategori Sumur Tua"
"2" : "P&A"
"3" : "WO"
"4" : "WS"
}

Label mapping for Potential Status:

{
"0" : "No Potential"
"1" : "Potential"
}

Figure 7
Label mapping

	Area	Well Name	Lifting Method	Well Cum Prod(Np, MBO)	HC Remaining Potential(Oil, MBO)	Doability	Recommendation	Potential Status
17	KAWENGAN	KWG-014	3	1,777.403	2.7731	2	1	1
18	KAWENGAN	KWG-015	3	614.684	0.009	2	1	1
19	KAWENGAN	KWG-016	3	1,395.256	0.009	2	3	1
20	KAWENGAN	KWG-017	3	1,802.227	0.009	2	1	1
21	KAWENGAN	KWG-017T/P1	3	0	0.009	2	0	1
22	KAWENGAN	KWG 018	1	2,635.92	0.009	2	1	1
23	KAWENGAN	KWG-019	3	1,424.642	0.009	2	1	1
24	KAWENGAN	KWG-022	3	1,763.075	0.009	2	1	1
25	KAWENGAN	KWG-023	1	146.348	0.009	2	1	1
26	KAWENGAN	KWG-024	1	664.43	0.009	2	1	1

Figure 8
Results of the labelendcoder process

The next process is data normalization which can be enabled by selecting the normalization feature and activating it using the button shown in Figure 9. As illustrated in Table 2, this step transforms data values into a uniform scale, such as the previous value of 0.009 is normalized to 0, while 2.2852 becomes 0.0002.

The data normalization process uses Min-Max method with a scale of 0-1 considering the number of decimal digits in the data. This is useful for comparing features with different actual value ranges but similar decimal digits.

Confusion matrix

This process is a feature evaluation to understand and analyze model performance based on True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values to show the relationship between actual and predicted classifications.

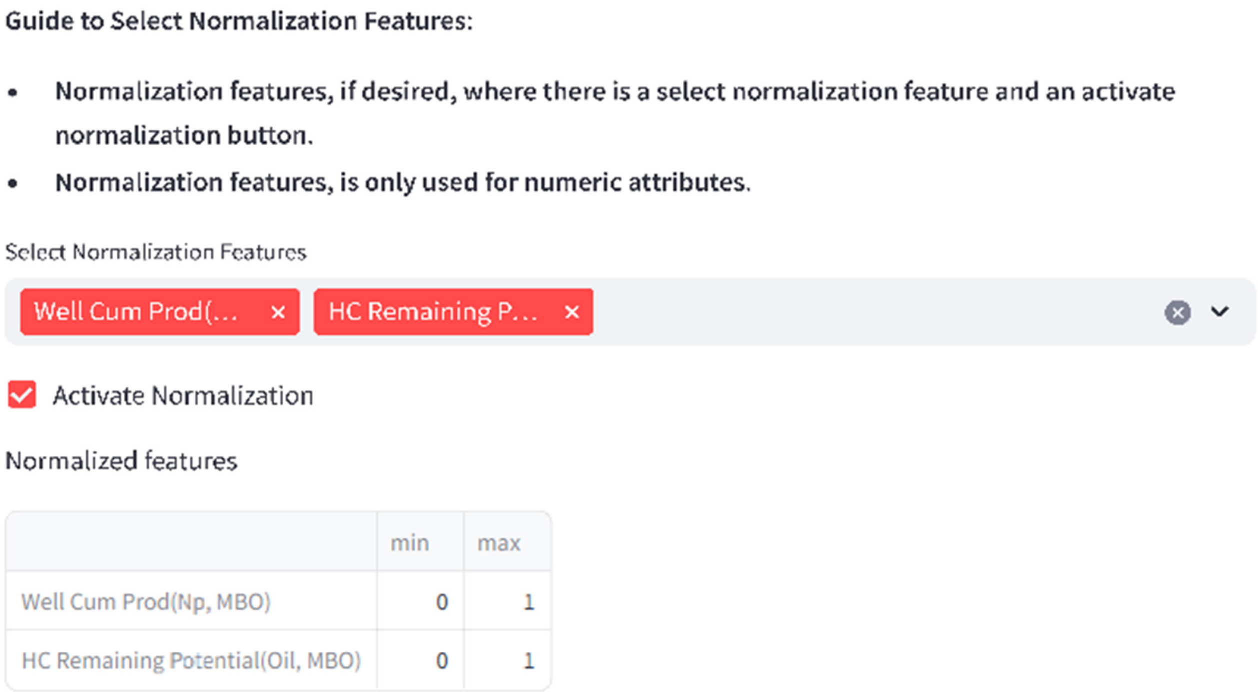


Figure 9
Results of the normalization process

Table 2
Display activation normalization

Well Cum. Prod (Np, MBO)	HC Remaining Potential (Oil, MBO)
1,777.403	2.7731
614.684	0.009
1,395.256	0.009
1,802.227	0.009
0	0.009
2,635.92	0.009
1,424.642	0.009
1,763.075	0.009
146.348	0.009
664.43	0.009

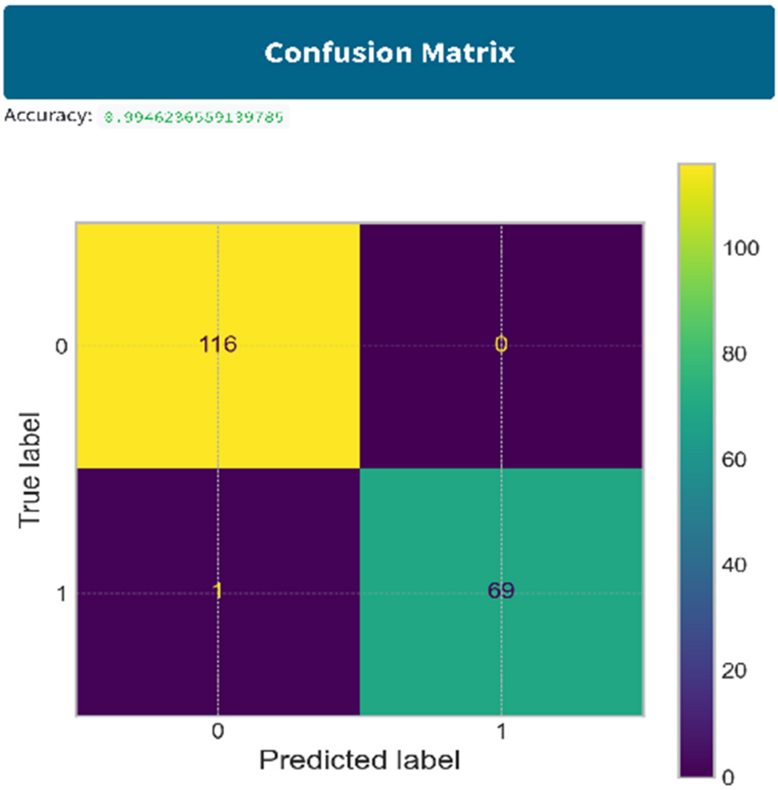


Figure 10
Results of confusion matrix

Based on Figure 10, the confusion matrix consists of two rows and two columns representing the four classification outcomes namely True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN). A value of 0 indicates a "not potential" well, while 1 indicates a "potential" well. The model achieved an accuracy of 0.99, meaning it correctly predicted 99% of the cases demonstrating high classification performance. The confusion matrix results are interpreted as follows:

TP = Wells are unsuitable for reactivation, and the algorithm states that 116 wells are not.

FP = The well is suitable for reactivation, but the algorithm says the well is not suitable for reactivation at 0 wells.

FN = The well is unsuitable for reactivation, and the algorithm states that one well is suitable for reactivation.

TN = The wells are suitable for reactivation, and the algorithm states that 69 wells are suitable for reactivation.

Entropy model

The entropy model is an internal interpretive mechanism within the Decision Tree classifier for making predictions. In the decision tree, each node represents the question or condition used to divide data into splitting criteria, determining what features and values are used to split across each node. Figure 11. shows models decision tree used to predict a class (1 or 0). The model utilizes five key features namely HC Remaining 'Potential (Np, MBO), Well Cum Prod (Np, MBO), Do ability, Recommendation, and Lifting Method. The following is a detailed explanation of each feature in the tree:

Root node

The root node is a rectangular box representing a question or decision because the question is asked for each data sample. Text: Well Cum Prod (Np, MBO) <= 334,536. If the condition is "True," the path follows the branch to the left, which means this branch represents the data sample with HC Remaining Potential (Oil, MBO) $\leq$ 0.004. The next node is Well Cum Prod (Np, MBO) $\leq$ 334,536. If the answer is "False," the path follows the branch to the right, which means this branch represents the data sample with HC Remaining Potential (Oil, MBO) > 0.004. The next node is the Lifting method $\leq$ 2.5 until reaching the leaves (leaf), which is the final class of the specified sample.

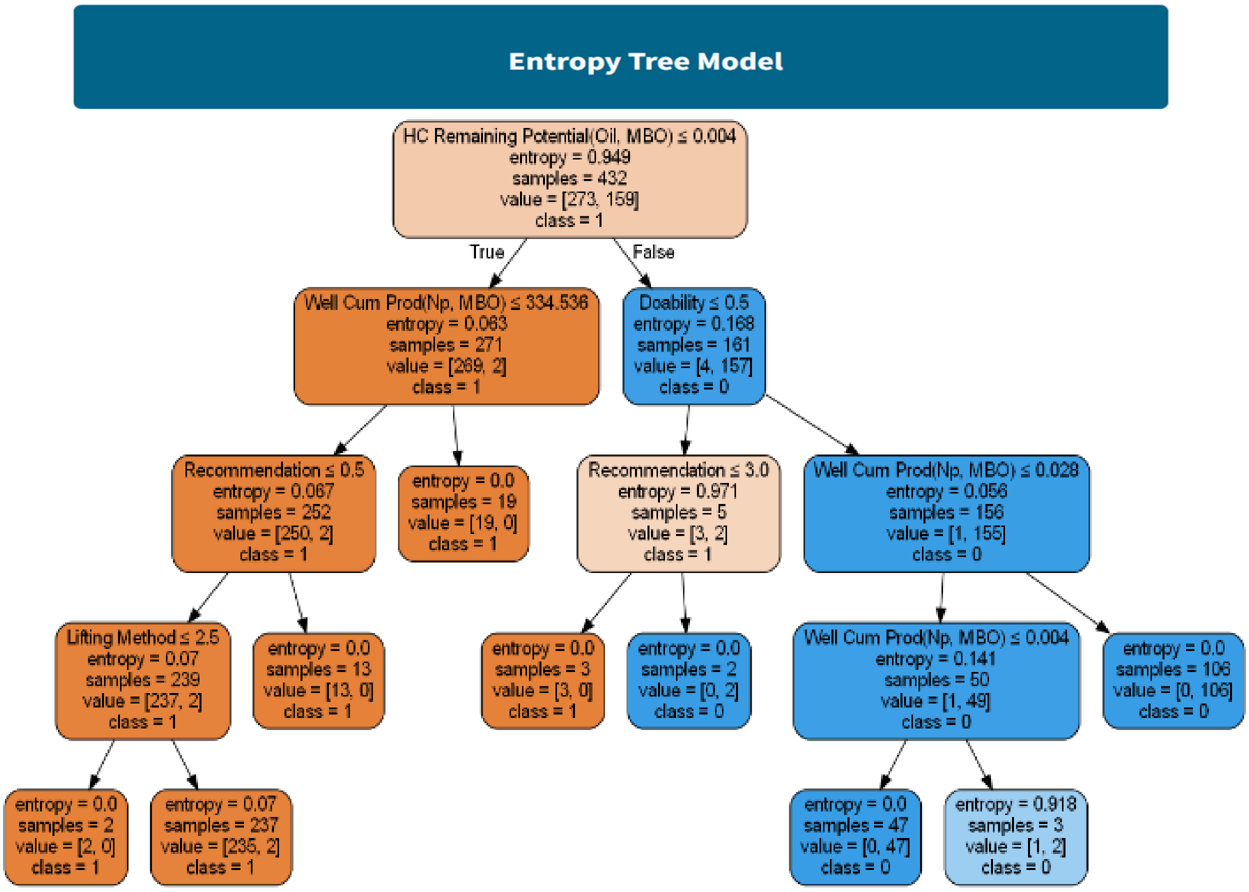


Figure 11
Results of entropy tree

1). Entropy: 0.949 (High level of uncertainty), defined as a measure of uncertainty in the node. The higher the entropy, the more diverse the data classes in the node; 2). Samples: 432 The amount of data used to build the model), means the number of data samples that reach that node; 3). Value: [273,159] (Amount of data in classes 1 and 0), interpreted as the distribution of data classes in the node. The first value indicates the number of class 1 samples, and the second value indicates the number of class 0 samples; 4). Class 1 (The most frequently occurring class).

Following so on and as a reference in determining reactivation candidates, look at the leaf tree finally, as follows:

1). Root Node, Text: Lifting Method <= 2.5; 2). Entropy: 0.0 (Uncertainty level 0) is defined as a measure of uncertainty in a node, which means that there are no various data classes in that node; 3). Samples: 2 (The amount of data used to build the model) means the number of data samples that reach that node; 4). Value: [2, 0] (Amount of data in classes 1 and 0) means the distribution of data classes in the node. The first value indicates the number of class 1 samples, and the second value indicates the number of class 0 samples; 5). Class: 1 The most frequently occurring class).

Therefore, Based on Figure 11. and the explanation taken in class 1 (Potential) on the features lifting method, the Branch section (left) gets the results as in Figure 12. and Table 3.

In Figure 12 and Table 3, two potential wells were obtained for reactivation based on the Lifting method (ESP), namely wells NGL-P-001 and TPN-004 as seen from Well Cum Prod (Np, MBO) and HC Remaining Potential (Oil, MBO). For NGL-P-001, the location is deemed suitable, the wellhead has been approved, and a flowline is already available. The well is scheduled for conversion to ESP, but is currently awaiting the availability of a 350 HP rig. For TPN-004, the location is also suitable and the wellhead approved; however, reactivation is delayed due to ongoing bridge repairs by the local government. In both cases, the recommended work program includes pump replacement or repair.

The next stage is ranking each potential well based on two key interests which are Well Cum Prod (Np, MBO) and HC Remaining Potential (Oil, MBO) from the highest to the lowest. The next stage determines the row coloring based on values of 'Well Cum Prod (Np, MBO)' and 'HC Remaining Potential (Oil, MBO)'. This is achieved using threshold values defined in a scale table, which assigns colours based on the calculated quartiles (0.25, 0.50, 0.75) for both metrics. Wells are categorized into performance levels such as "Fair," "Good," and "Very Good." Intermediate values ("Fair") are determined using the median and quartiles, while "Good" and "Very Good" categories are assigned based on upper quartile values. The colour scheme has been updated to include blue for "Very Good" wells, while maintaining previous colour assignments for other categories. This data-driven approach, implemented through the Idle Well Reactivation Dashboard, is supported by the model's strong performance demonstrated by an AUC and ROC score of 0.99, indicating excellent classification accuracy in identifying idle wells suitable for reactivation.

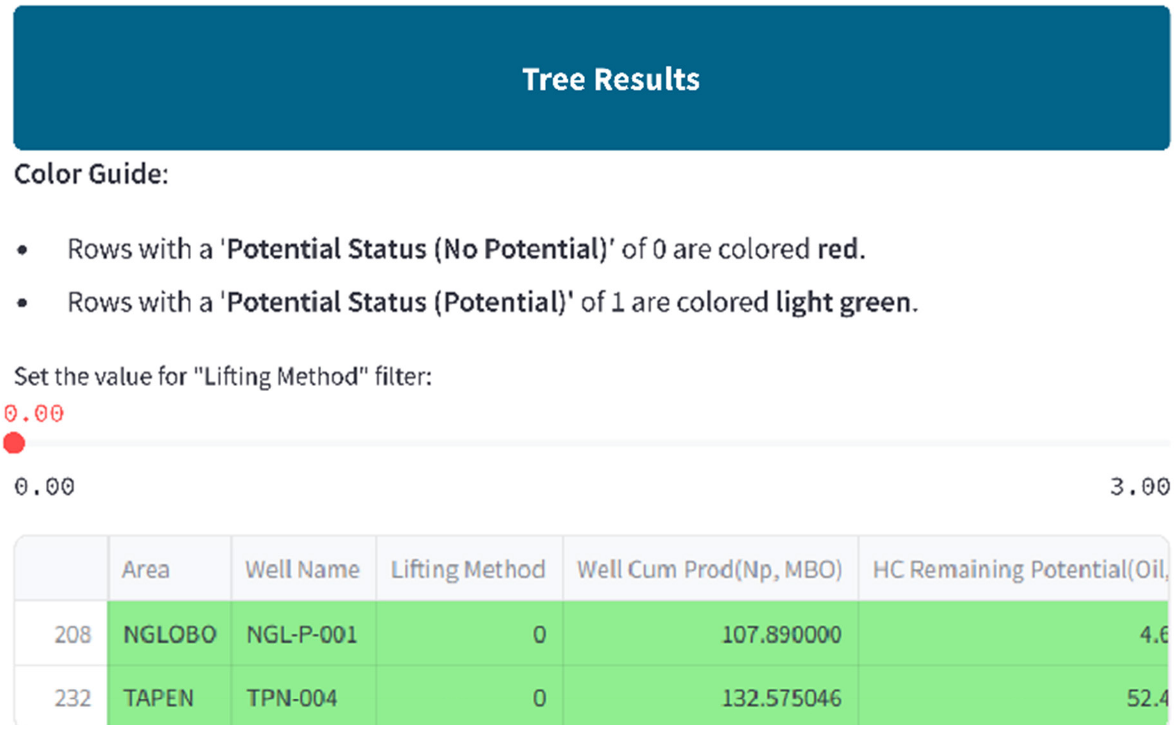


Figure 12
Display results of tree

Table 3
Results of tree

Well Name	Lifting Method	Well Cum Prod (Np, MBO)	HC Remaining Potential (Oil, MBO)	Doability	Recommendation	Potential Status
NGL-P-001	0	107.89	4.60973	2	4	1
TPN-004	0	132.5750459	52.4223	0	4	1

	Area	Well Name	Lifting Method	Well Cum Prod(Np, MBO)	HC Remaining Potential(Oil, MBO)	Doability	Recommendation	Potential Status
232	TAPEN	TPN-004	0	132.575046	52.422300	0	4	1
230	TAPEN	TPN-001	2	0.000000	0.000000	0	2	0
231	TAPEN	TPN-002	2	0.000000	0.000000	0	2	0

Color Scale Table:

Range	Color
Very Good	
Good	
Fair	
Poor	

Figure 13
Display of select area (Tapen)

	Area	Well Name	Lifting Method	Well Cum Prod(Np, MBO)	HC Remaining Potential(Oil, MBO)	Doability	Recommendation	Potential Status
208	NGLOBO	NGL-P-001	0	107.890000	4.609730	2	4	1
199	NGLOBO	NGL-033	3	91.590000	2.283840	2	0	1
192	NGLOBO	NGL-024	3	80.400000	0.009000	2	4	1
193	NGLOBO	NGL-025	3	74.580000	0.009000	2	4	1
202	NGLOBO	NGL-039	3	65.140000	12.198800	2	4	1
209	NGLOBO	NGL-P-002	3	59.990000	0.009000	2	4	1
194	NGLOBO	NGL-026	3	46.110000	4.027560	2	0	1
201	NGLOBO	NGL-036TW	3	20.350000	0.009000	2	4	1
191	NGLOBO	NGL-022	3	6.860000	0.009000	2	1	1
205	NGLOBO	NGL-042	3	1.240000	0.009000	2	0	1
200	NGLOBO	NGL-036	3	0.650000	4.429000	2	4	1

Color Scale Table:

Range	Color
Very Good	
Good	
Fair	
Poor	

Figure 14
Display of select area (Ngolo)

In Figure 13 – Figure 14. displays a color scale table that is used to help identify uneconomic wells and those that have the potential to be reactivated concerning each value range. Wells with low values of Well Cumulative Production (Np, MBO) and Hydrocarbon Remaining Potential (Oil, MBO) are highlighted in red, indicating they are not economically viable for production. Conversely, wells with high values for these parameters are shown in green or blue, signifying strong candidates for reactivation.

Accuracy under curve (AUC)

Evaluation of the performance of the classification model for determining idle good reactivation candidates was evaluated by calculating the accuracy under the curve (AUC) value). AUC measures a model ability to differentiate between worthy and unworthy idle wells for reactivation. Then, the AUC value = 1, which means a perfect classification model that can distinguish all idle well correctly; an AUC Value = 0.5 means the classification model is no better than random guessing. An AUC value > 0.5 means the classification model is better than random guessing, so that a higher value indicates better performance. A high AUC indicates that the classification model can differentiate between worthy and unworthy idle wells for reactivation. A low AUC indicates that the classification model is not very good at differentiating idle well. Thus, based on the results obtained in Figure 15, the AUC for tree depth and max-leaf nodes is 0.99, which means that the performance of the classification model can differentiate idle well which deserves to be reactivated properly.

 | DOI org/10.29017/scog.v48i2.1657 |

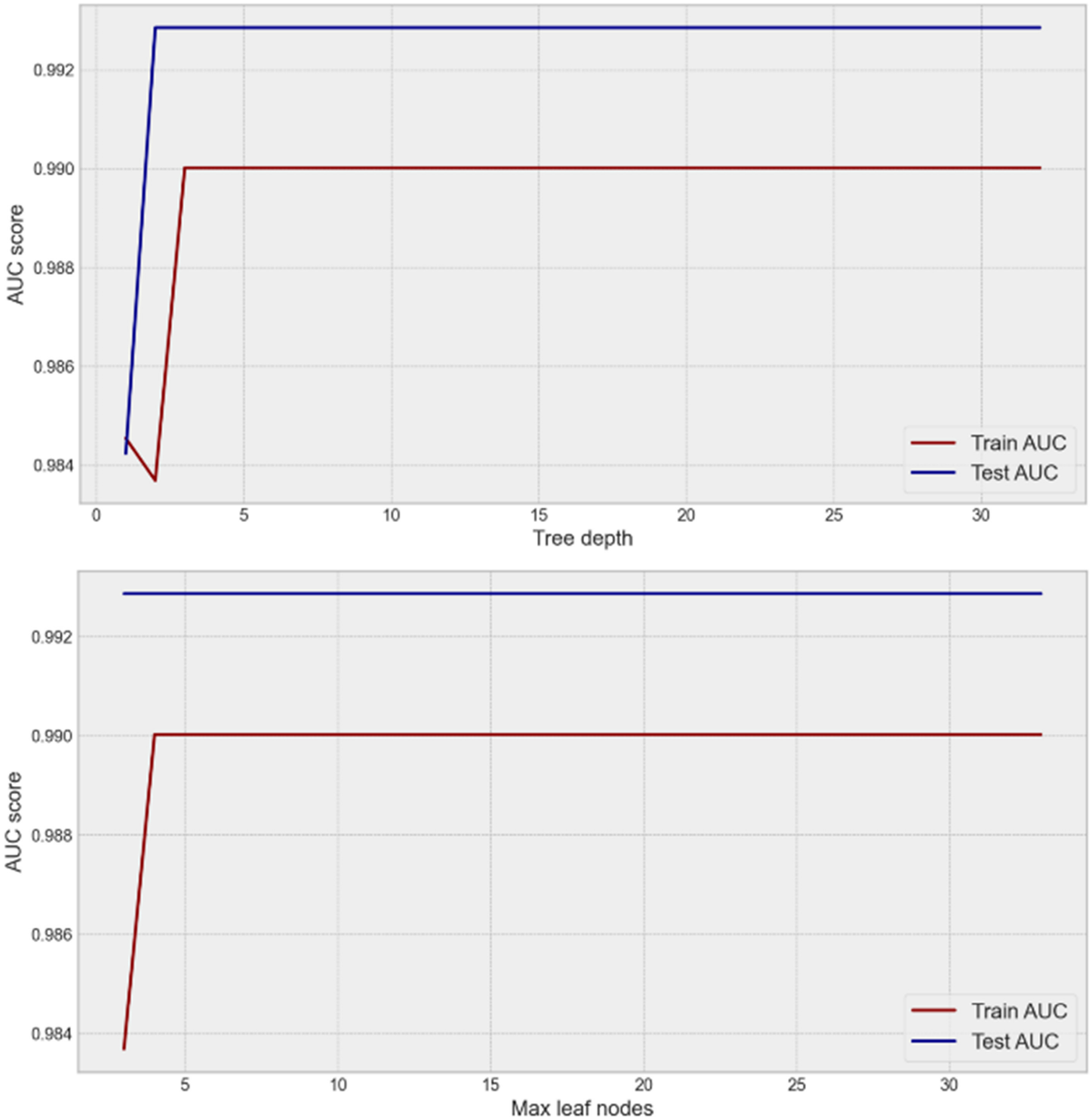


Figure 15
Display results of AUC tree depth and max leaf nodes

Receiver operating curve (ROC)

The Receiver Operating Characteristic (ROC) curve is a valuable tool for evaluating the performance of classification models, particularly in distinguishing between idle wells that should be reactivated and those that should not. The Area Under the Curve (AUC) is derived from the ROC curve and represents the model's overall ability to discriminate between the two classes. A high ROC curve positioned well above the diagonal line indicates strong model performance, meaning the classifier performs significantly better than random guessing. Conversely, a ROC curve close to the diagonal suggests lower predictive power. As shown in Figure 16, the ROC curve for this model achieves an AUC of 0.99, demonstrating that the classifier is highly effective at correctly identifying idle wells that are viable candidates for reactivation.

Evaluation idle well reactivation candidate

Evaluation of idle well reactivation candidates based on the decision tree classification model process and obtained two potential wells, namely the NGL-P-001 and TPN-004 wells based on Well Cum Prod (Np, MBO) and HC Remaining Potential (Oil, MBO). The analysis was carried out for the TPN-004 well because, based on the history, the well was active from 2018-2022. In contrast, the NGL-P-01 well experienced stuck production in 2018-2022. Therefore, this evaluation only focuses on the TPN-004 well.

Water chan diagnostics

At the Water Chan Diagnostic stage, data analysis from potential wells is divided into two processes: Chan Diagnostic and Chan Diagnostic Plot. This first step involves uploading data, which is divided into two groups: wells with Water-Oil Ratio (WOR) data and those without. This step is closely tied to features related to Well Production Performance and helps in understanding data characteristics, identifying patterns, and detecting production trends. The second step involves generating the Chan Diagnostic Plot, which can be displayed using two different scaling methods: log scale and semi-log scale. These visualizations assist in evaluating the well's water production behaviour and identifying potential issues such as water coning or channelling.

Based on Figure 17 – Figure 18, data upload was performed to display the data and preview well information. A total five records were shown, along with statistical summaries for each attribute. For data entries where WOR values are missing, additional columns for WOR and WOR* are generated and appended to the dataset to enable further diagnostic analysis:

$$WOR = \frac{Water\ production\ (bwpd)}{Net\ Production\ (bopd)} \tag{1}$$

$$WOR^{*} = \frac{\left(d\left(\frac{WOR}{dt}\right) - \left(\frac{WOR_2}{WOR_1}\right)\right)}{(t_2 - t_1)} \tag{2}$$

Based on Chan Diagnostic Plot, the analysis of TPN-004 well in 2018 obtained deflection results in the Derivate Water Oil Ratio (WOR*) and Water Oil Ratio (WOR) curves that shows normal pressing, as WOR* does not show distinct signs of water channelling or coning as the data appears too dispersed, suggesting an indication of a problem of Normal Displacement with High WOR. A similar pattern is observed during the 2019 and 2021 periods, as illustrated in Figures 19–23.

The TPN-004 has experienced problems well in the 2020 and 2022 periods. Near Wellbore Water Channeling suddenly occured during normal production and pressing. At first, the WOR was constant, but as time passed, the WOR price was above 1. Then the WOR rose very quickly with a slope of about 3.

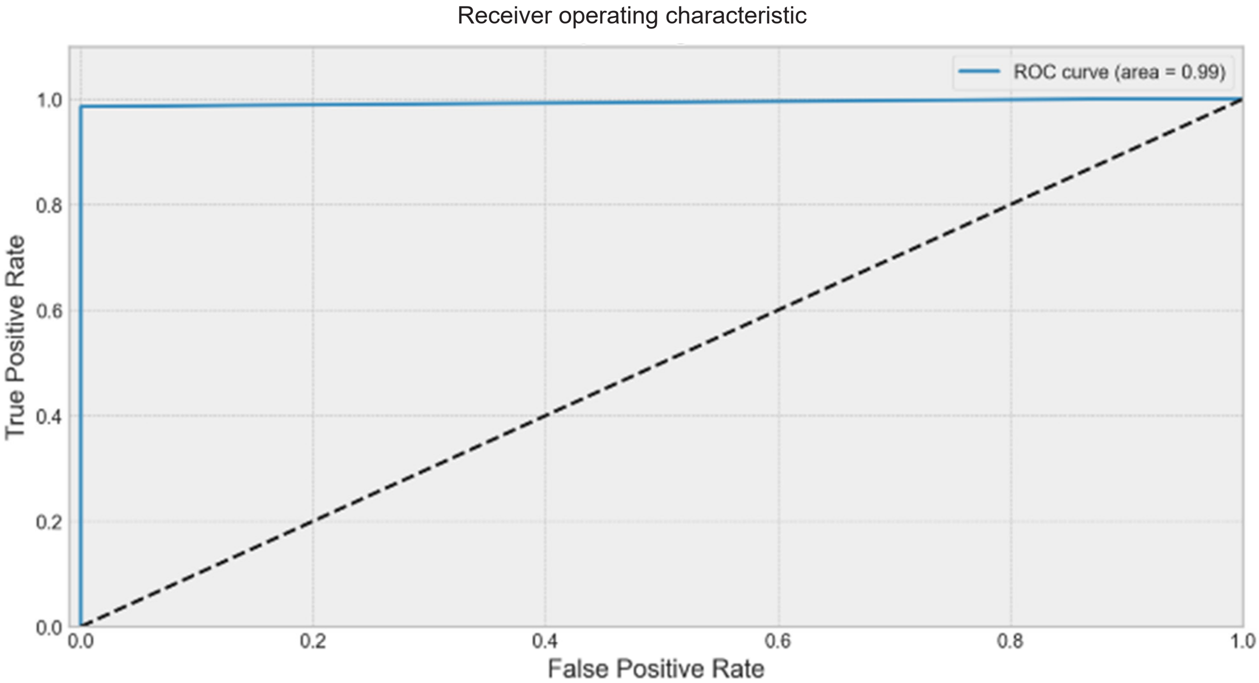


Figure 16
Display results of ROC

Chan Diagnostic

Guide Upload Data:

- Upload a excel file1 '**If you have WOR data**': (WOR, and WOR*).
- Upload a excel file2 '**If you don't have WOR data**': (Gross (blpd), Net (bopd), Water (bwpd), Time, Time Cumulative).
- Features realated to '**Features Well Production Performance**'.

Choose a excel file1

Drag and drop files here
Limit 200MB per file

Browse files

Tapen-004(chan-2018).xlsx 39.8KB

filename: Tapen-004(chan-2018).xlsx

Display Rows

5

0 1000

	Date	Gross (blpd)	Net (bopd)	Water (bwpd)	WC (%)	Time	Time Cum	WOR	WOR*
0	2018-01-01 00:00:00	144.384	140.544	3.84	2.6596	1	1	0.0273	0
1	2018-01-02 00:00:00	144.384	140.544	3.84	2.6596	1	2	0.0273	0
2	2018-01-03 00:00:00	144.384	140.544	3.84	2.6596	1	3	0.0273	0
3	2018-01-04 00:00:00	144.384	140.544	3.84	2.6596	1	4	0.0273	0
4	2018-01-05 00:00:00	144.384	140.544	3.84	2.6596	1	5	0.0273	0

Data Statistics

	Date	Gross (blpd)	Net (bopd)	Water (bwpd)	WC (%)	Time	Time Cum	WOR	WOR*
count	365	365	365	365	364	365	365	365	365
mean	2018-07-02 00:00:00	139.4913	131.877	7.6143	5.3453	0.9753	176.7041	0.0584	0.0248
min	2018-01-01 00:00:00	0	0	0	0	0	1	0	0
25%	2018-04-02 00:00:00	128	121.6	3.84	3.0076	1	86	0.0308	0.0024
50%	2018-07-02 00:00:00	140.8	133.12	7.04	4.9896	1	177	0.0524	0.0113
75%	2018-10-01 00:00:00	153.6	145.92	9.6	6.7895	1	267	0.0728	0.029
max	2018-12-31 00:00:00	230.4	225.6	75.9771	38.7413	1	356	0.6324	0.6324
std	None	34.2467	32.7243	6.4194	3.9134	0.1553	104.3602	0.0519	0.0499

Figure 17
Display upload (Have WOR data)

Choose a excel file2

Drag and drop files here
Limit 200MB per file

Browse files

Tapen-004 2018.xlsx 29.8KB

filename: Tapen-004 2018.xlsx

Display Rows
5
0 1000

	Date	Gross (blpd)	Net (bopd)	Water (bwpd)	WC (%)	Time	Time Cum
0	2018-01-01 00:00:00	144.384	140.544	3.84	2.6596	1	1
1	2018-01-02 00:00:00	144.384	140.544	3.84	2.6596	1	2
2	2018-01-03 00:00:00	144.384	140.544	3.84	2.6596	1	3
3	2018-01-04 00:00:00	144.384	140.544	3.84	2.6596	1	4
4	2018-01-05 00:00:00	144.384	140.544	3.84	2.6596	1	5

Display Rows
5
0 1000

Data Add Columns WOR & WOR*

	Date	Gross (blpd)	Net (bopd)	Water (bwpd)	WC (%)	Time	Time Cum	WOR	WOR*
0	2018-01-01 00:00:00	144.384	140.544	3.84	2.6596	1	1	0.0273	None
1	2018-01-02 00:00:00	144.384	140.544	3.84	2.6596	1	2	0.0273	0
2	2018-01-03 00:00:00	144.384	140.544	3.84	2.6596	1	3	0.0273	0
3	2018-01-04 00:00:00	144.384	140.544	3.84	2.6596	1	4	0.0273	0
4	2018-01-05 00:00:00	144.384	140.544	3.84	2.6596	1	5	0.0273	0

Data Statistics

	Date	Gross (blpd)	Net (bopd)	Water (bwpd)	WC (%)	Time	Time Cum	WOR	WOR*
count	365	365	365	365	364	365	365	356	352
mean	2018-07-02 00:00:00	139.4913	131.877	7.6143	5.3453	0.9753	176.7041	0.0599	-0.0013
min	2018-01-01 00:00:00	0	0	0	0	0	1	0	-0.3671
25%	2018-04-02 00:00:00	128	121.6	3.84	3.0076	1	86	0.0322	-0.0119
50%	2018-07-02 00:00:00	140.8	133.12	7.04	4.9896	1	177	0.0531	0
75%	2018-10-01 00:00:00	153.6	145.92	9.6	6.7895	1	267	0.0738	0.011
max	2018-12-31 00:00:00	230.4	225.6	75.9771	38.7413	1	356	0.6324	0.2653
std	None	34.2467	32.7243	6.4194	3.9134	0.1553	104.3602	0.0517	0.0457

Figure 18
Display upload (do not Have WOR data)

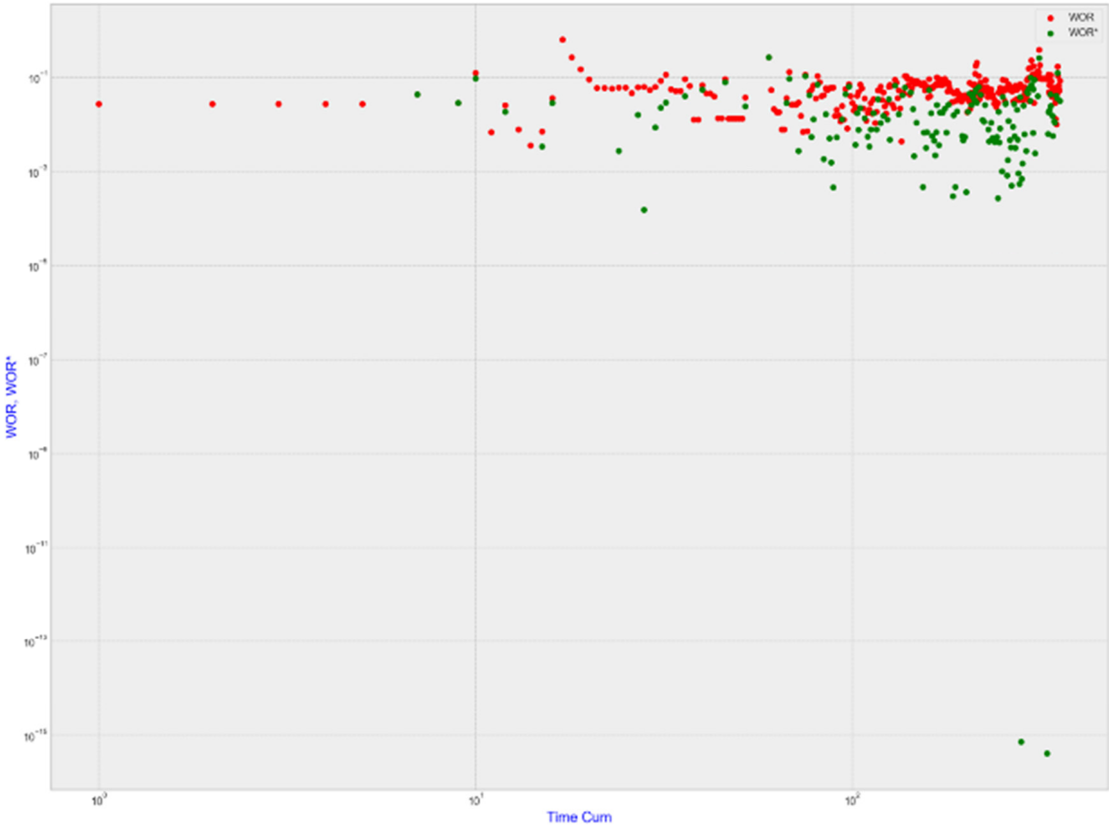


Figure 19
Result chan diagnostic plot TPN-044 (2018)

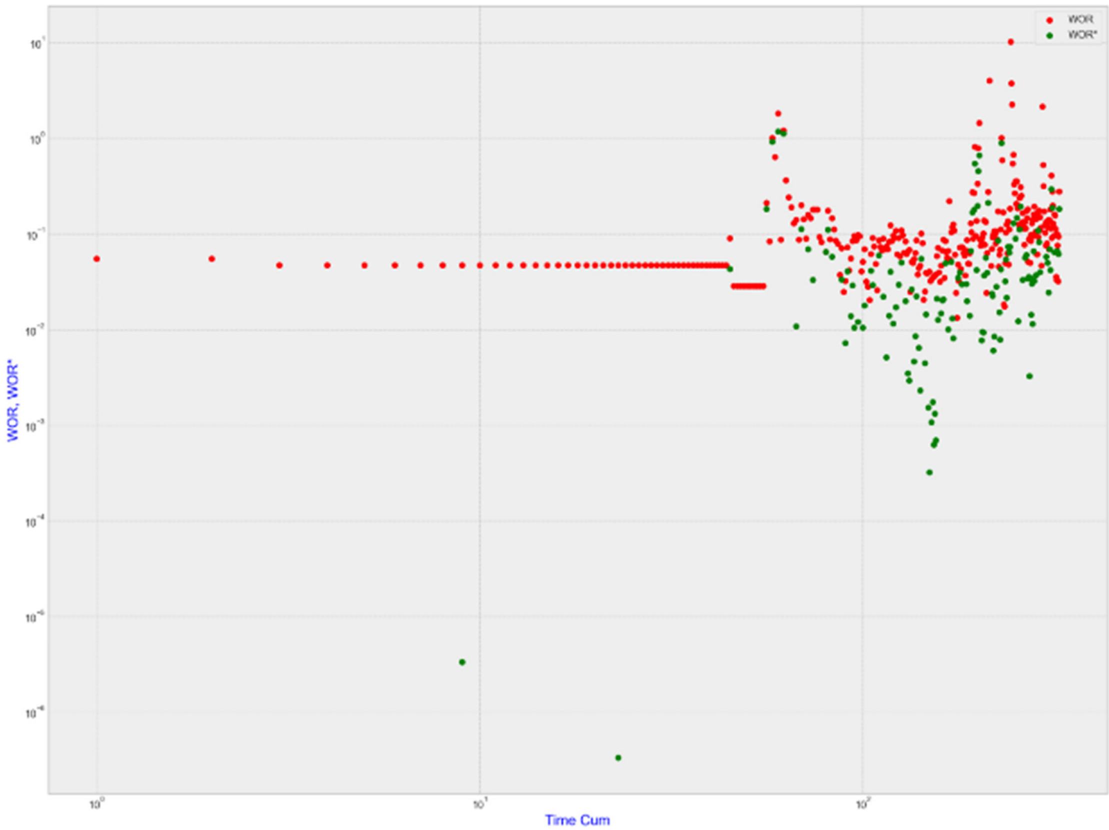


Figure 20
Result chan diagnostic plot TPN-044 (2019)

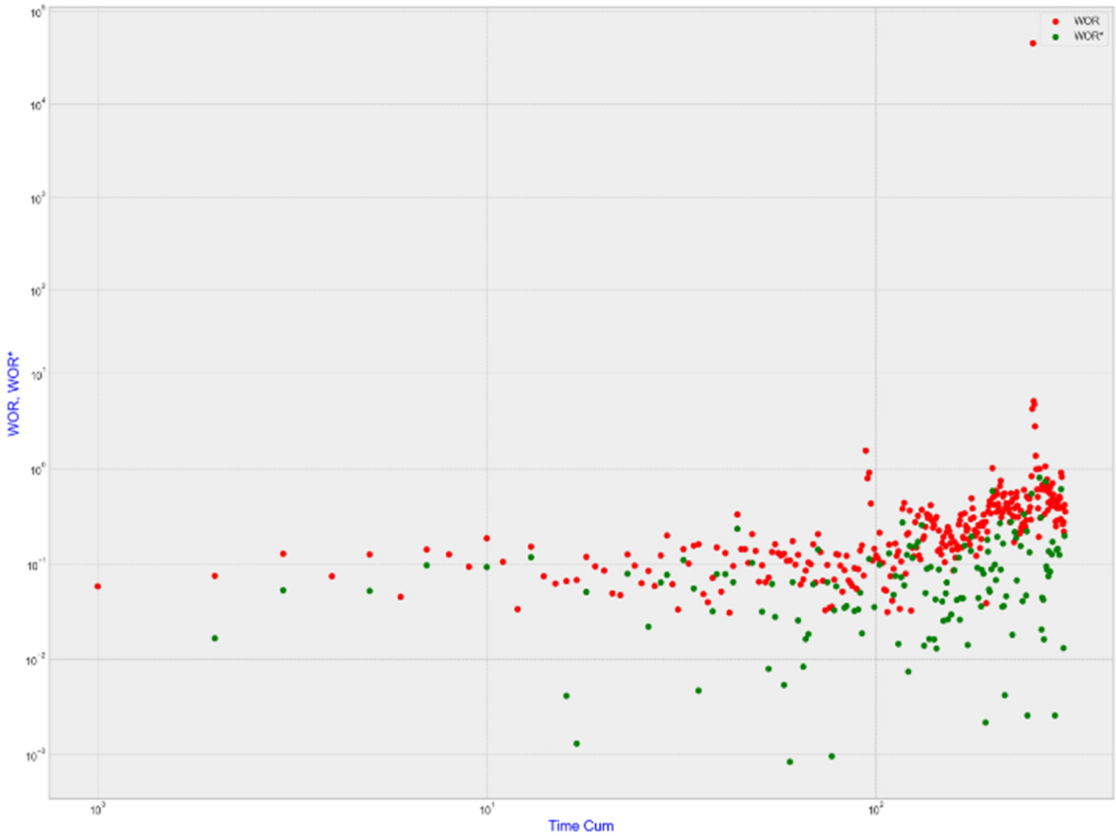


Figure 21
Result chan diagnostic plot TPN-044 (2020)

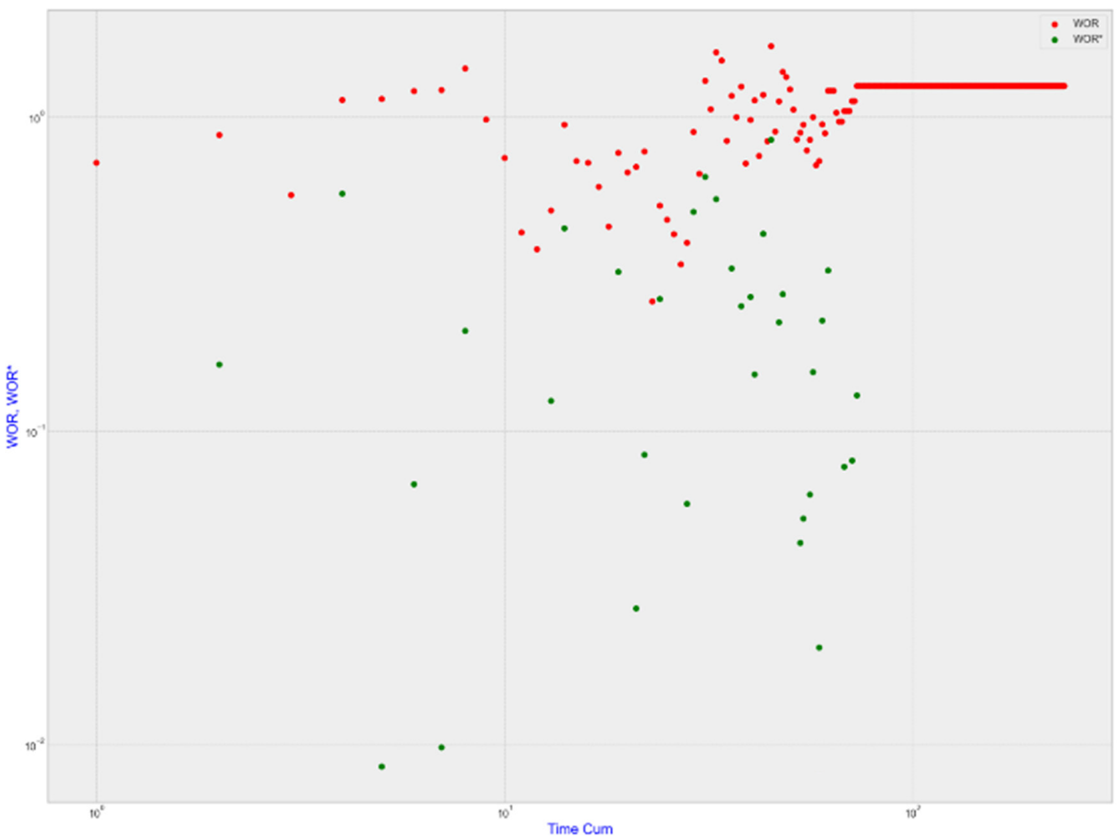


Figure 22
Result chan diagnostic plot TPN-044 (2021)

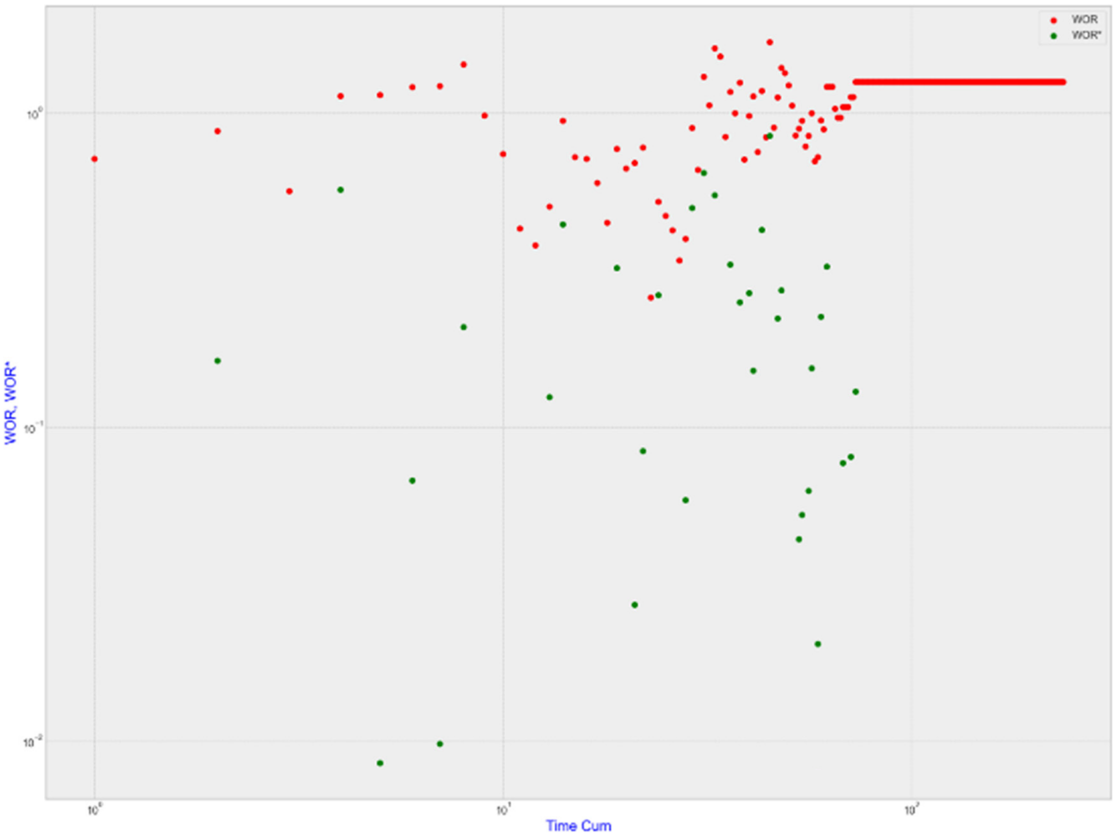


Figure 23
Result chan diagnostic plot TPN-044 (2022)

Well production performance analysis

In stages, Well Production Performance is carried out to analyze data from potential wells. The reference for determining potential candidates for reactivation uses production data from the TPN-004 well with a production period of 4 years (2018, 2019, 2020, 2021, and 2022). During this period, the well achieved a cumulative production (Np, MBO) of 132.57 MBO and had a remaining hydrocarbon potential (Oil, MBO) of 52.42 MBO. The potential of the well is obtained by comparing its cumulative production to the maximum cumulative total from similar wells, under the assumption of a homogeneous reservoir. The detail production of each well can be seen in Figure 24 - Figure 28.

Figures 24 – 28 reveals a fluctuating trend in production volume, with both gross and net production showingt a decline while the watercut increases. Additionally, there are also periods when the well temporary ceases production. This analysis is essential for evaluating well performance, identifying potential well issues, planning and optimizing oil production, and making informed decisions regarding well development.

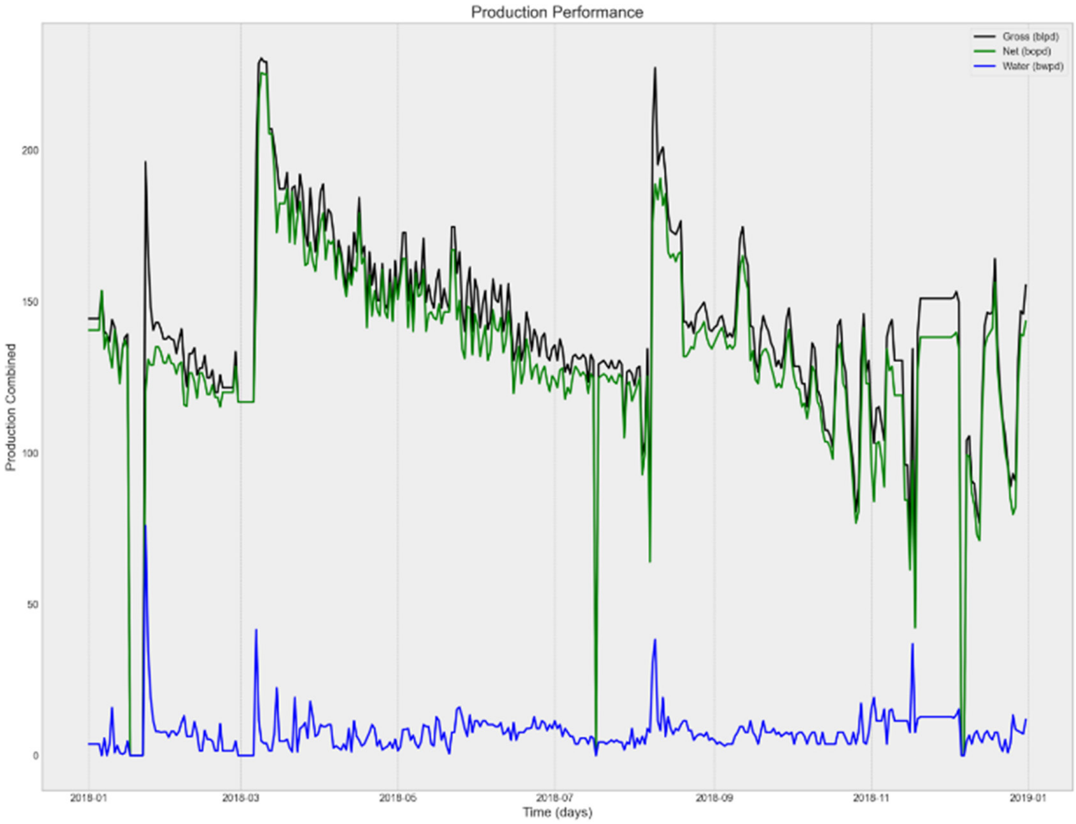


Figure 24
Results well production performance TPN-044 (2018)

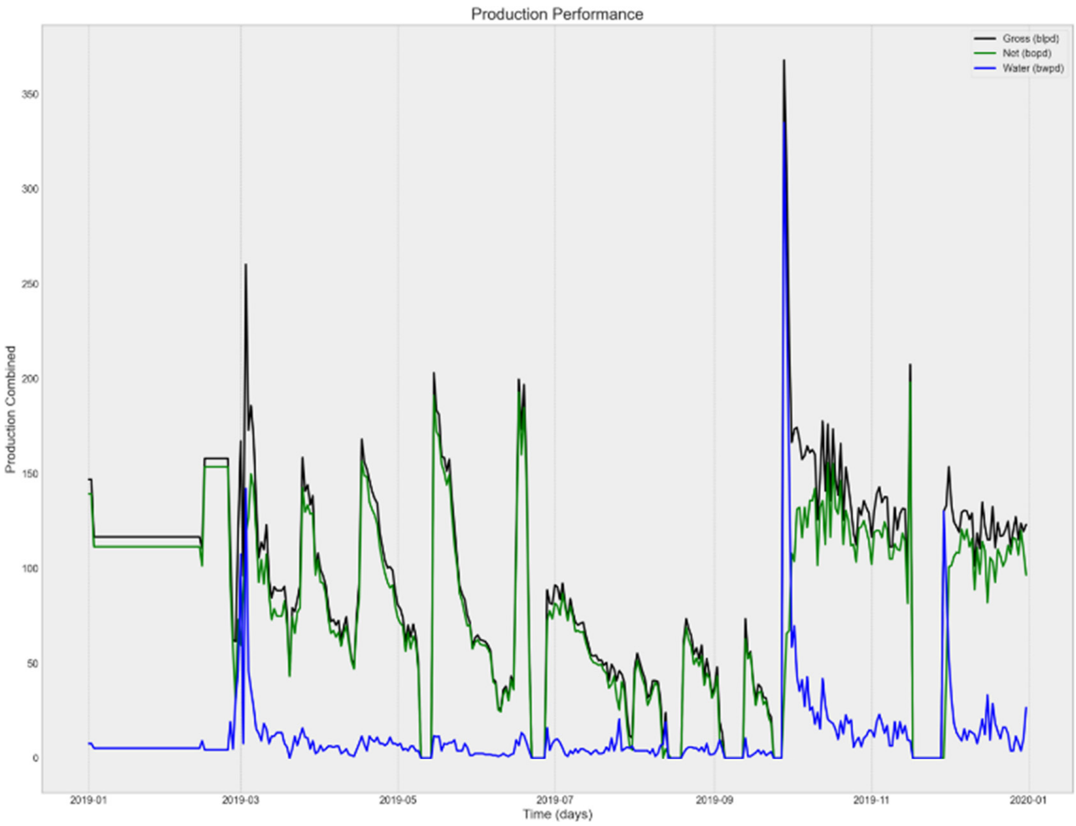


Figure 25
Results well production performance TPN-044 (2019)

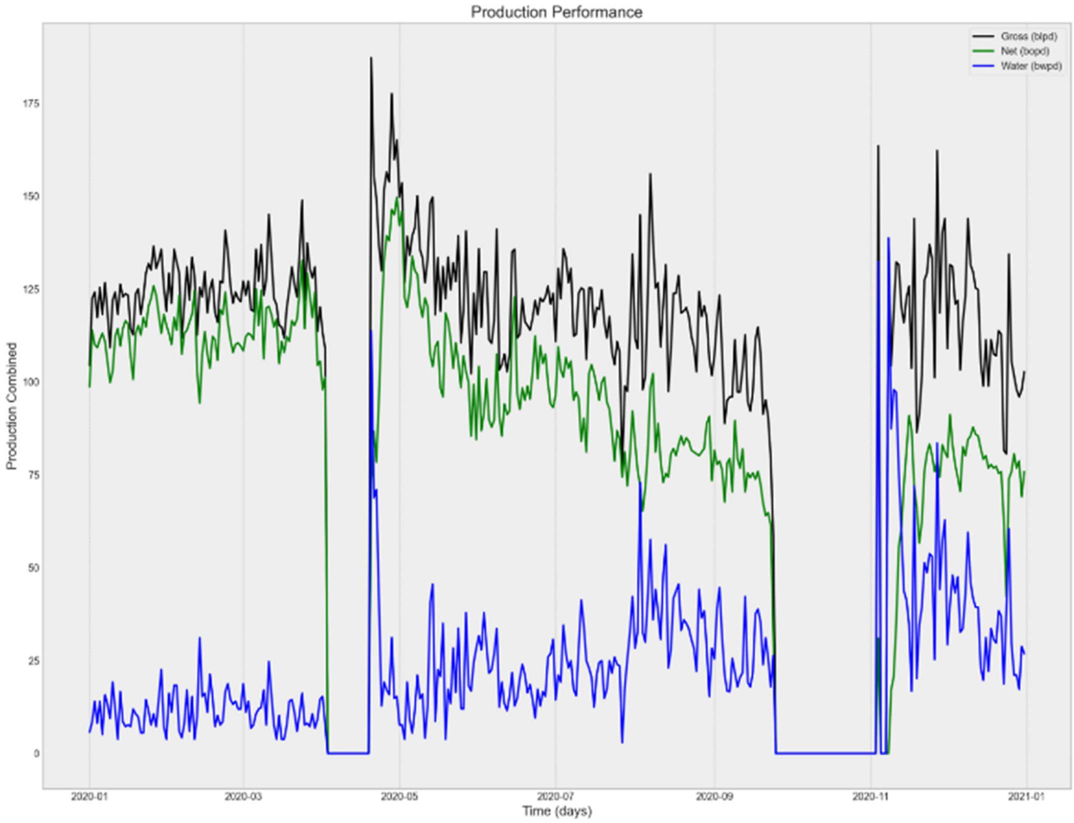


Figure 26
Results well production performance TPN-044 (2020)

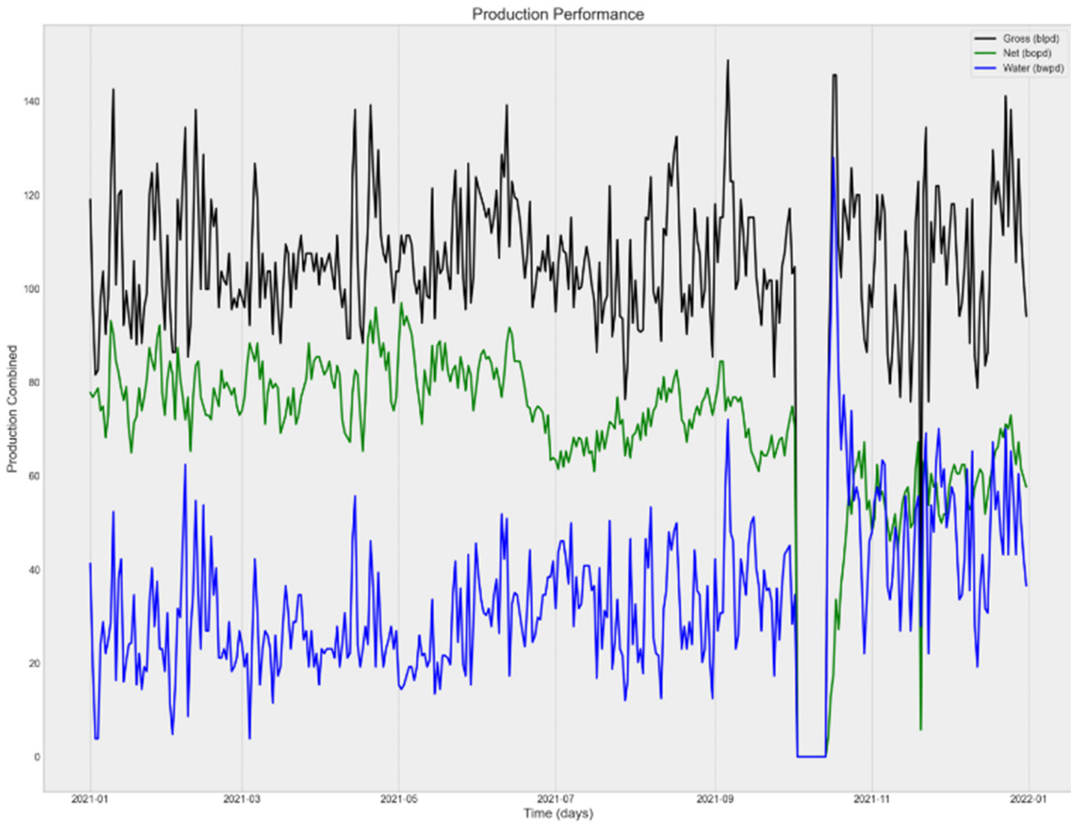


Figure 27
Results well production performance TPN-044 (2021)

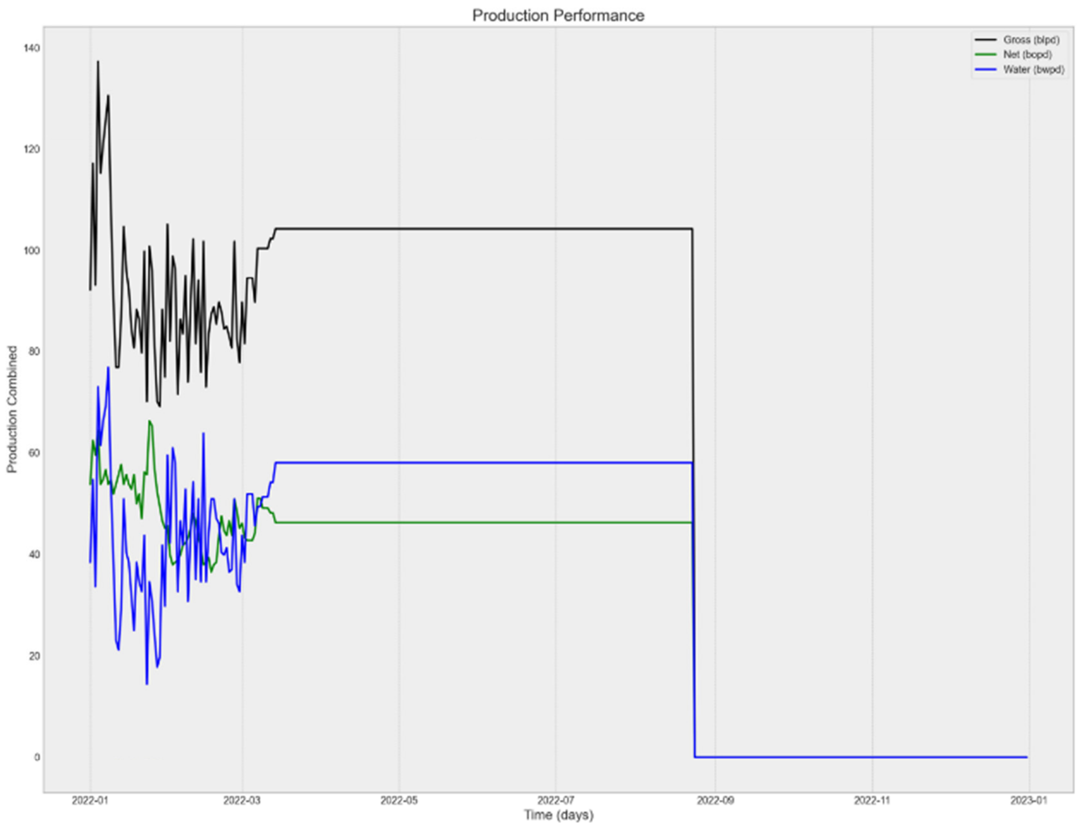


Figure 28
Results well production performance TPN-044 (2022)

CONCLUSION

Based on the analysis and the results of the Decision Tress modelling for identifying idle well reactivation candidates, the following conclusions can be drawn as follows: 1). The application of the Decision Tree model can overcome the problem of determining reactivation candidates Idle well in terms of data accuracy and complexity, as well as a high success rate on Accuracy Under Curve (AUC) and Receiver Operating Curve (ROC) of 0.99, which shows a high probability of the classification model; 2). Based on the confusion matrix results for the classification of 618 wells into not potential (0) and potential (1), the algorithm identified 116 wells as not suitable for reactivation. It misclassified one well as suitable when it was not. Additionally, the algorithm classified 69 wells as suitable for reactivation, demonstrating high predictive accuracy; 3). Based on entropy analysis, two wells were identified as potential candidates for reactivation using the Electrical Submersible Pump (ESP) lifting method. First, Well NGL-P-001, with a cumulative production (Np) of 107.89 MBO and a remaining hydrocarbon potential of 4,609 MBO. Second, Well TPN-004, with a cumulative production (Np) of 132.57 MBO and a remaining hydrocarbon potential of 52.42 MBO. For both wells, the recommended follow-up action is Well Service to enhance reactivation outcomes; 4). Of the 2 idle wells in the Cepu field identified as reactivation candidates by decision tree model, then an evaluation is carried out using Chan Diagnostic method. The TPN-004 well exhibited signs of normal displacement with high WOR in the period of 2018, 2019 and 2021. Additionally, during period of 2020 and 2022, the well showed indications of Near Wellbore Water Channeling. The Production performance analysis reveals that TPN-004 operated a 5-year production period (2018-2022), with a cumulative production of 132.6 MBBL.

ACKNOWLEDGEMENT

This work was supported by the Universitas Pembangunan Nasional “Veteran” Yogyakarta. The author gratefully acknowledges the support during the research and publication, and presents this paper.

GLOSSARY OF TERMS

Symbol	Definition	Unit
ML	Machine Learning	
DT	Decision Tree	
IW	Idle Well	
AI	Artificial Intelligence	
HWC	High Water Cut	
SVM	Support Vector Machines	
AUC	Accuracy Area Under Curve	
ROC	Receiver Operating Characteristic	
Np	Well Cummulative Production	MBO
WOR	Water Oil Ratio	

REFERENCES

Alfarizi, Ardi, Nur Suhascaryo & Boni Swadesi, 2023, Technical and Economical Study on Increasing Oil Production in Old Wells (Traditional) by Performing Well Maintenance in the CP Field." Journal of Petroleum and Geothermal Technology 4(1):46. doi: 10.31315/jpgt.v4i1.7422.

Alkinani, Husam H., Abo Taleb, T. Al-Hameedi, & Shari Dunn-Norman, 2019, SPE-194795-MS Review of the Applications of Decision Tree Analysis in Petroleum Engineering with a Rigorous Analysis.

Amin Nizar, M., C.A. Razak, M. Naquib, Kamarul Zaman, Affira Ali, Jamari M. Shah, Zaki Sakdillah, Md Zarin & Md Zainuri, 2019, IPTC-19218-MS Reviving Idle Wells and Unlocking Potential Production Gain in Offshore Sarawak Through Exposing BCO-LRLC Opportunities.

Ardi, Suhascaryo & Swadesi, 2022,*Kajian Teknis dan Ekonomis Peningkatan Produksi Minyak pada Sumur Tua (Tradisional) Dengan Melakukan Perawatan Sumur Di Lapangan Cepu.* (8.5.2017):2003–5.

Bangert & Patrick, 2019, SPE-194993-MS Diagnosing and Predicting Problems with Rod Pumps Using Machine Learning.

Bizhani, Majid & Ergun Kuru, 2022, Towards Drilling Rate of Penetration Prediction: Bayesian Neural Networks for Uncertainty Quantification. Journal of Petroleum Science and Engineering 219(September):111068. doi: 10.1016/j.petrol.2022.111068.

Candra, Arya Dwi, Pradini Rahalintar, Sulistiyono, & Urip Nurwijayanto Prabowo, 2024, Comparison of Facies Estimation of Well Log Data Using Machine Learning. Scientific Contributions Oil and Gas 47(1):21–30. doi: 10.29017/SCOG.47.1.1593.

Chan, Christine W., 2009, Data Analysis for Oil Production Prediction. in Encyclopedia of Data Warehousing and Mining.

Chithra Chakra, N., Ki Young Song, Madan M. Gupta & Deoki N. Saraf, 2013, An Innovative Neural Forecast of Cumulative Oil Production from a Petroleum Reservoir Employing Higher-Order Neural Networks (HONNs). Journal of Petroleum Science and Engineering 106:18–33. doi: 10.1016/j.petrol.2013.03.004.

Garcia, Carlos A., Akhan Mukhanov & Henry Torres, 2019, Chan Plot Signature Identification as a Practical Machine Learning Classification Problem. International Petroleum Technology Conference 2019, IPTC 2019 1–19. doi: 10.2523/iptc-19143-ms.

Haryanto, Elin, Saltanat Yersaiyn, Agha Hassan Akram, Francois Bouchet, Haytham Galal, and Mohd Ashraf. 2019. "Standardization of Inactive Wells Audit Process."

He, Rongquan, Weizhong MA, Xinyu Ma & Yuchen Liu., 2021, Modeling and Optimizing for Operation of CO2-EOR Project Based on Machine Learning Methods and Greedy Algorithm. Energy Reports 7:3664–77. doi: 10.1016/j.egyr.2021.05.067.

Li, Xiongmin & Christine W. Chan, 2010, Application of an Enhanced Decision Tree Learning Approach for Prediction of Petroleum Production. Engineering Applications of Artificial Intelligence 23(1):102–9. doi: 10.1016/j.engappai.2009.06.003.

Li, Xiongmin & Christine W. Chan. n.d. Towards A Neural-Network-Based Decision Tree Learning Algorithm for Petroleum Production Prediction.

Mukhanov, Akhan, Carlos Arturo Garcia & Henry Torres, 2018, Water Control Diagnostic Plot Pattern Recognition Using Support Vector Machine. Society of Petroleum Engineers - SPE Russian Petroleum Technology Conference 2018, RPTC 2018. doi: 10.2118/191600-18rptc-ms.

Nguyen, H.H., C.W. Chan & M. Wilson, 2004, Prediction of Oil Well Production: A Multiple-Neural-Network Approach. Vol. 8. IOS Press.

Nguyen, Hanh H. & Christine W. Chan, 2005, Applications of Data Analysis Techniques for Oil Production Prediction. Engineering Applications of Artificial Intelligence 18(5):549–58. doi: 10.1016/j.engappai.2004.11.010.

Olatunji, Sunday Olusanya, Ali Selamat & Abdul Azeez Abdul Raheem, 2010, Modeling Permeability Prediction Using Extreme Learning Machines. Pp. 29–33 in AMS2010: Asia Modelling Symposium 2010 - 4th International Conference on Mathematical Modelling and Computer Simulation.

Putra, Gatmasada Riandi, Program Studi, Magister Teknik, and Fakultas Teknologi Mineral, 2022, *Pemanfaatan Idle Well Sebagai Program Kerja Reaktivasi Dengan Metode Multi Screening Dan Rencana Optimasi Sucker Pemanfaatan Idle Well Sebagai Program.*

Retno L.P. Marsud, 2021, *Laporan Kinerja Ditjen MIGAS* 53(9):1689–99.

Sandha, A., K. Agha & R. Islam, 2005, Artificial Intelligence...as a Decision Support System for Petroleum Engineers. Petroleum Science and Technology 23(5–6):555–71.

Saptowulan, Dhanu, Idqan Fahmi & Bagus Sartono, 2022, Strategy Formulation of Natural Gas Continuity Supply (Case Study PT ABC), Scientifi c Contributions Oil and Gas.” Scientifi c Contributions Oil & Gas 45(1):13–25.

Sun, Yuanqing, Guo-Rui Ma, Lingqian Wang, Fanchang Zhang, X. Y. Li & Linhai Mei, 2019, Organic-Matter Content Prediction Based on the Random Forest Algorithm: Application to a Lower Silurian Shale-Gas Reservoir. SEG Technical Program Expanded Abstracts 2019.

Utomo & Chandra Prasetyo, 2011, The Hybrid of Classification Tree and Extreme Learning Machine for Permeability Prediction in Oil Reservoir.

de Ville & Barry, 2013, Decision Trees.” Wiley Interdisciplinary Reviews: Computational Statistics 5(6):448–55. doi: 10.1002/wics.1278.

Wardhana, Sanggeni Gali, Henry Julois Pakpahan, Krisdanyolan Simarmata, Waskito Pranowo & Humbang Purba, 2021, Algoritma Komputasi Machine Learning Untuk Aplikasi Prediksi Nilai Total Organic Carbon (TOC). *Lembaran Publikasi Minyak Dan Gas Bumi* 55(2):75–87. doi: 10.29017/LPMGB.55.2.606.