Scientific Contributions Oil & Gas, Vol. 48. No. 1, April: 135 - 144

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

Case Study of Heterogeneity Index's Effect on The Successful Workover Based on The Apriori Algorithm

Fahrizal Maulana, Amega Yasutra and Zuher Syihab

Bandung Institute of Technology
Ganesa Street No.10, Lb. Siliwangi, Coblong District, Bandung City, West Java 40132, Indonesia.

Corresponding author: fahrizalmau@gmail.com

Manuscript received: December 24th, 2024; Revised: January 28th, 2025
Approved: February 07th, 2025; Available online: April 16th, 2025.

ABSTRACT - The Indonesian government has set a target to reduce their consumption-production gap by increasing national oil production to 1 million barrels of oil per day (BOPD) and 12 billion standard cubic feet of gas per day (BSCFD) by 2030. Amongst several approaches, the optimization of mature fields offers a significant opportunity for quick production gains. However, analyzing these fields presents challenges due to the complexity, incompleteness, and poor quality of historical data. Heterogeneity index (HI) is one of the methods that can quickly measure well-performance. This method is as simple as comparing a certain well's performance to the field's average at a given time. The focal parameter can vary, but production data is the most frequently used given its availability. Despite its simplicity and practicality, skepticism over HI's reliability remains. This work revisited one oilfield in West Java Offshore, consisting of 47 wells producing for more than 32 years with hundreds of workovers. We brought evidence and insights on how HI leads to workover success. The Apriori algorithm, an association rule mining (ARM) technique, is employed to uncover rules from the noisy data. The results show that workovers on wells with low HI mostly led to success. Another insight is that scale treatment is the most influential workover in terms of success. Given these findings, we determine that flow efficiency is the well's issue that should be most often treated, and HI is representative enough to measure it.

Keywords: heterogeneity index, mature field, production optimization, Apriori algorithm, association rule mining.

How to cite this article:

Fahrizal Maulana, Amega Yasutra and Zuher Syihab, 2025, Case Study of Heterogeneity Index's Effect on The Successful Workover Based on The Apriori Algorithm, Scientific Contributions Oil and Gas, 48 (1) pp. 135-144. DOI.org/10.29017/scog.v48i1.1658.

INTRODUCTION

Disruptors are reshaping the global energy sector: geopolitical tensions, macroeconomic challenges (such as rising interest rates and material costs), shifting regulations, and technological advancements, and fiscal attractiveness. These factors influence supply-demand dynamics, trade, and investment within the oil and gas industry. (Chronis et al. 2023; BP 2023; EIA 2023; Mardiana et al. 2024).

Since 2016, Indonesia has experienced a steady decline in crude oil production. In 2023, production stood at 221,089 thousand barrels, a significant drop from 303,336 thousand barrels in 2016. This decline is due to the natural depletion of mature oil fields and limited exploration of new reserves

(ESDM 2023; IEA 2022). To reverse this trend, the government has set a target of producing 1 million barrels of oil per day (BOPD) and 12 billion cubic feet of gas per day (BSCFD) by 2030. Key strategies include reactivating idle wells, modernizing mature fields, attracting investments, leveraging advanced technologies, and improving regulations (ESDM 2022; Wood Mackenzie 2023).

Mature oil fields, defined as those which have produced over 50% of their reserves or been operational for more than 25 years, present challenges such as declining production rates, aging infrastructure, and environmental concerns. However, they also offer opportunities for cost-effective production. The cost of extracting oil from these fields is often four to five times lower than developing new reserves. Revitalization projects using improved oil recovery (IOR) and enhanced oil recovery (EOR) techniques can extend field lifespans and increase production. Globally, mature fields remain a significant resource, but their complex data demands substantial time and effort for analysis (Parshall et al. 2012; Schlumberger 2022; IPCC 2021).

Several methodologies have been proposed to enhance production in mature fields, including HI, risk matrices, and the Delphi method. While conventional, these methods require significant engineering effort. Data analytics has been proposed to handle complex datasets more efficiently (Galicia et al. 2021).

HI is widely used to evaluate well's performance and has shown promising results in real-world applications (Basset et al. 2018; Harami et al. 2013). It was also used for stimulation candidate selection (Turnip et al. 2024). However, it evaluates only single parameters and does not account for interactions between variables. This study aims to examine the robustness of HI using data mining techniques, focusing on a mature Indonesian oil field with a 32-year production history to explore HI's correlation with workover success.

METHODOLOGY

Heterogeneity index

HI is a quantitative tool used to assess well performance and identify flow inefficiencies. It works by comparing a well's production performance to the average performance of a group of wells, providing insights into reservoir heterogeneity, well completion efficiency, and opportunities for production optimization. The HI is calculated using production data for various values (e.g., oil, gas, and water rates) and normalized against field averages at certain time. The formula for HI is expressed as.

$$HI = \frac{Value\ (t)}{Field\ Average\ Value\ (t)} - 1 \tag{1}$$

where:

- HI > 0 indicates overperformance of a well relative to the field average
- HI < 0 indicates underperformance
- HI = 0 indicates average performance

This simple calculation allows for rapid screening of wells to identify anomalies. HI can be visualized using scatter plots, cross-plots, or geographic maps to detect trends in well behavior or regional reservoir characteristics. For instance, low HI values may indicate potential well issues such as near-wellbore damage or inefficient stimulation, while high HI values may highlight areas of better reservoir quality (Abdel et al. 2018; Harami et al. 2013; Reese 1996).

Estimated ultimate recovery

The calculation of EUR often involves decline curve analysis, which uses mathematical models to predict future production based on historical data. One of the most common methods is the Arps decline curve analysis, which uses the equation.

$$q(t) = q_i(1 + bD_i t)^{-\frac{1}{b}} \tag{2}$$

where:

- $q(t)$ is the production rate at time t
- q_i is the initial production rate
- D_i is the initial decline rate
- b is the hyperbolic exponent $(0 \le b \le 1)$

The EUR can be calculated by integrating the production rate over time:

$$EUR = \int_0^\infty q(t)dt \tag{3}$$

However, this method is under the assumption that the well flows naturally, Farrah et al., 2023 extends this work to estimate EUR with artificial lift.

Apriori algorithm

The Apriori algorithm is a key data mining technique for association rule learning, introduced to discover frequent item sets in large datasets, which are then used to generate association rules (Agrawal & Srikant, 1994). It is particularly useful in market basket analysis to identify patterns in customer purchasing behavior, such as frequently bought item combinations (Hipp et al. 2000).

The Apriori algorithm relies on the "Apriori property," which asserts that all non-empty subsets of a frequent itemset must also be frequent. This downward closure property significantly reduces the search space for candidate item sets, improving computational efficiency (Han et al., 2011). The algorithm iteratively scans the dataset to identify item sets that meet a minimum support threshold, pruning supersets of infrequent item sets.

The Apriori algorithm relies on three key metrics to measure the strength of association rules:

Support: The support of an itemset measures the proportion of transactions in the dataset that contains the itemset. It is defined as:

$$Support(A) = \frac{(Number\ of\ transactions\ containing\ A)}{(Total\ number\ of\ transactions)} \quad (4)$$

Confidence: The confidence of a rule A→B is the probability that a transaction containing A also contains B. It is calculated as:

$$Confidence(A \rightarrow B) = \frac{Support(A \cup B)}{Support(A)} \quad (5)$$

Lift: Lift measures the strength of an association by comparing the observed co-occurrence of A and B to their expected co-occurrence under statistical independence. It is expressed as:

$$Lift(A \rightarrow B) = \frac{Confidence(A \rightarrow B)}{Support(B)} \quad (6)$$

The favorable thresholds for the Apriori algorithm, minimum support and minimum confidence, depend on the dataset and analysis goals. Minimum support, typically set between 1% and 5%, determines the frequency an itemset must meet to be considered frequent, while minimum confidence, often between 50% and 80%, measures the strength of an association rule (Han et al., 2011; Hipp et al., 2000). Lower thresholds capture more patterns but increase computational cost, whereas higher thresholds reduce noise and complexity.

Workflow

To process the data and answer the research questions at hand, the following workflow (see Figure 1 was employed:

The details of each step are given below:

Data Collection: Gather production history, workover history, and well coordinates from the offshore field in Indonesia.

Data Preprocessing: Clean and prepare data by handling missing values, normalizing, and formatting for analysis. Data preprocessing with the Pandas library involved structuring production data, categorizing wells by heterogeneity thresholds, and transforming data into a transactional format.

HI Calculation: Compute HI from production data to identify performance variations among wells.

The labels are as follows:

$$\begin{aligned} HI &\leq 1: Low \\ HI &> 1: High \end{aligned} \quad (7)$$

Drainage Status Analysis: Analyze spatial well distribution and proximity to assess interference and drainage patterns in every time step. A radius of 500 meters is used (PTK POD). The labels are as follows:

Table 1
Drainage status labels

Neighboring Wells	Drainage Status
0	NO_INTERFERENCE
1	LOW_INTERFERENCE
2, 3	MEDIUM_INTERFERENCE
4 or more	HIGH_INTERFERENCE

Decline Curve Analysis (DCA): Estimate EUR and predict future production rates using the Arps decline model. Afterwards, the Np of each time is compared to the EUR to determine the production lifting status.

Some wells show no dedicated trend. Thus, the EUR is undetermined. For other cases, the ratio is compared for every timestep.

Workover Assessment: Evaluate production gains from workovers to determine their effectiveness.

The success of the workover is assessed by the presence of oil gain in the next timestep.

Table 2
Lifting status labels

Classification	Condition	Return Value
Undetermined EUR	Wells with unrecognized pattern	UNDETERMINED_EUR
Low Lifting	NP_EUR_RATIO < 0.5	LOW_LIFTING
Medium Lifting	$0.5 \leq$ NP_EUR_RATIO < 0.9	MEDIUM_LIFTING
Nearly Fully Lifting	$0.9 \leq$ NP_EUR_RATIO < 0.95	NEARLY_FULLY_LIFTING
Full Lifting	NP_EUR_RATIO ≥ 0.95	FULL_LIFTING

Table 3
Description of transaction data labels

Column	Description
HI_OIL	The heterogeneity index for oil, labeled as ['HIGH_OIL'] or ['LOW_OIL'].
HI_WTR	The heterogeneity index for water, labeled as ['HIGH_WTR'] or ['LOW_WTR'].
HI_GAS	The heterogeneity index for gas, labeled as ['HIGH_GAS'] or ['LOW_GAS'].
LIFTING_STATUS	The lifting status of the well, categorized as ['LOW_LIFTING'], ['MEDIUM_LIFTING'], ['NEARLY_FULLY_LIFTING'], ['FULL_LIFTING'], or ['UNDETERMINED_EUR'].
NEIGHBOURING_WELLS	The status of neighboring wells based on the distance threshold. The labels such as ['NO_INTERFERENCE'], ['LOW_INTERFERENCE'].
WO	The list of workover jobs performed, represented as a list of job codes. The labels such as ['SCALE TREATMENT'], ['OO', 'XX'].
WO_RESULT	The result of the workover, labeled as ['OIL_GAIN'] or ['NO_GAIN'].

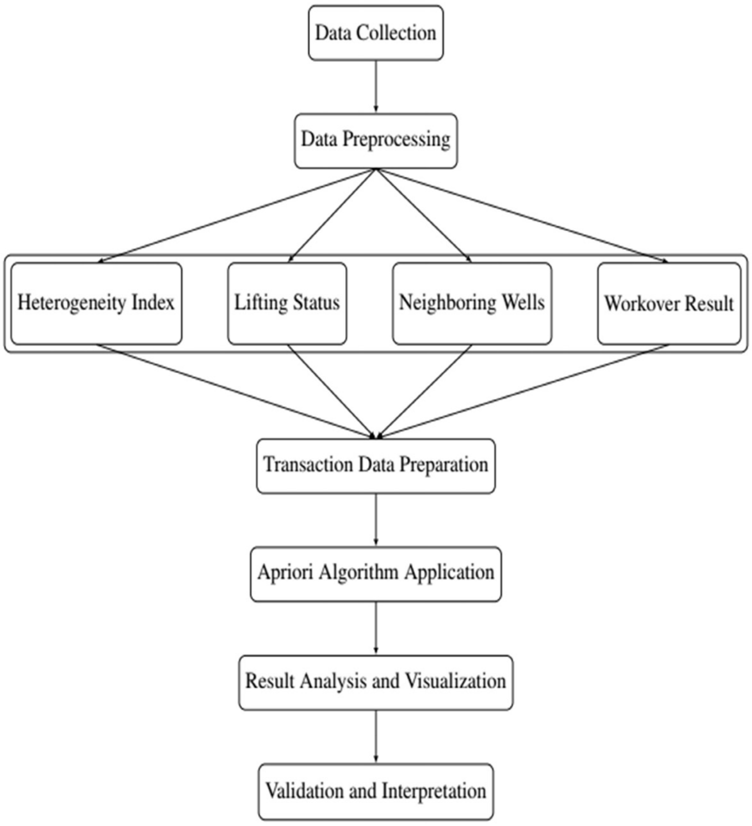

Figure 1
Data analysis workflow

where:

- N_p is the cumulative oil production at a certain time
- t is the time of interest

The label of *OIL_GAIN* is added if the oil gain is positive, and *NO_GAIN* otherwise.

Transaction Data Preparation: Transform data into a transactional format based on heterogeneity indices and attributes above. For each workover conducted, there must be 7 labels as follows:

Apriori Algorithm Application: Identify patterns and associations between attributes, generating frequent itemsets and rules. The Apriori algorithm was implemented in Python using the mlxtend library.

Result Analysis: Visualize results with bar plots and network graphs to reveal relationships between attributes. Results were visualized with Matplotlib and Seaborn, using bar plots to display frequent itemsets and network graphs to show association rules, highlighting relationships between heterogeneity indices and workover outcomes.

RESULT AND DISCUSSION

This study examines the offshore field in Indonesia, covering 47 wells with a production history of over 30 years. The data set available comprised production history, workover history, and well coordinates.

The production data and workover history revealed that some workover events overlapped with production events, while others did not (see Figure 2). Workovers with no production events were considered failed, as no production gain was observed. The data was then divided into two subsets: one containing row with workovers and the other without. The subset with workover data was transformed into a data frame of transactions. The distribution of the workover success can be seen in Figure 3. Most workovers generated oil gain in the next following timestep.

The data of transactions containing workover types and results was merged with the information on spatial drainage, lifting and HI status. The complete data with parameters of interest was then further processed to mine association rules. The distribution of items is shown in Figure 5. From the graph, we observe that the majority of workovers are conducted under low HI conditions, highlighting its prevalence across the datasets.

The rule results (as summarized in Table 4 to Table 7 and Figure 4) demonstrate that wells with low heterogeneity index (HI) values for oil, water, and gas are strongly associated with successful workovers, with a probability of success exceeding 79%. This indicates that HI is an effective metric for identifying flow inefficiencies, and that well interventions on wells with low HI values are more likely to result in oil production gains. Additionally, the type of workover plays a significant role in success rates. Scale treatment, particularly on wells with low HI values, shows a success probability of over 65%, making it a more effective intervention compared to other types like perforation. This highlights the importance of selecting workover types that address specific inefficiencies as reflected by HI.

Spatial relationships also influence workover outcomes. Wells with no interference from neighboring wells exhibit an over 70% probability of success, suggesting that these wells face less competition and are better candidates for production enhancement. Lifting status further supports this observation: wells with medium lifting status and low HI values show a success probability exceeding 73%. This aligns with the operational context where workovers are most effective during the mid-life of a well, when production is declining but stimulation is still viable. In contrast, wells in low lifting status often produce naturally, reducing the need for intervention, while wells with full lifting status reach optimal production, limiting the effectiveness of further workovers.

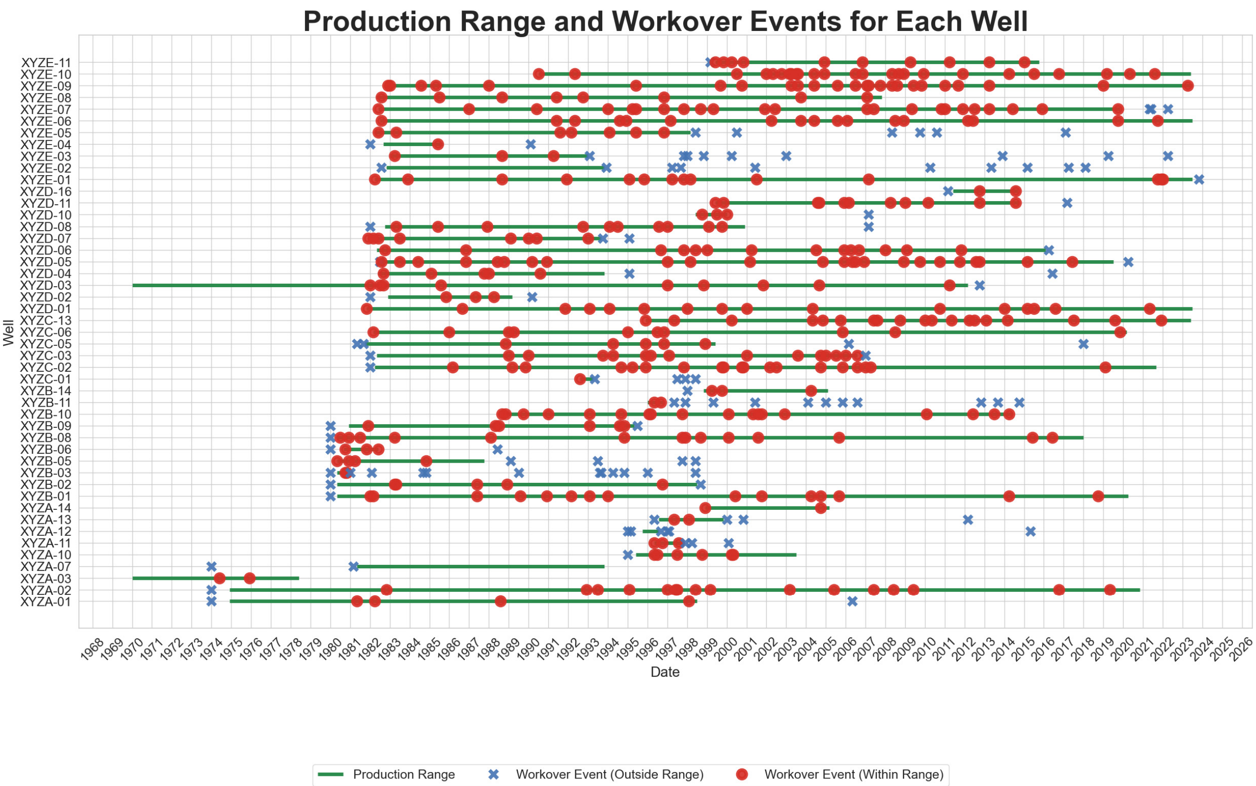

Figure 2
Production and workover timeline

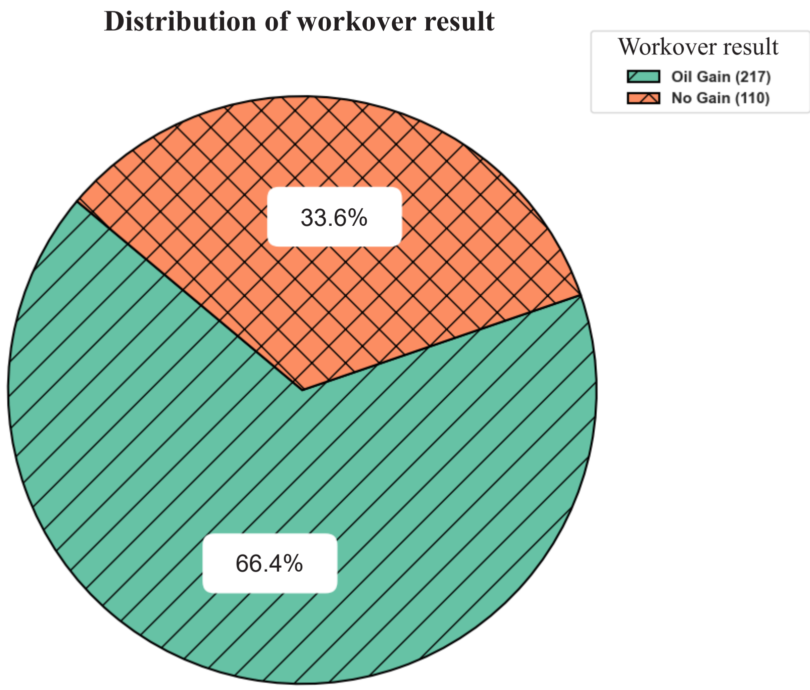

Figure 3
Workover result

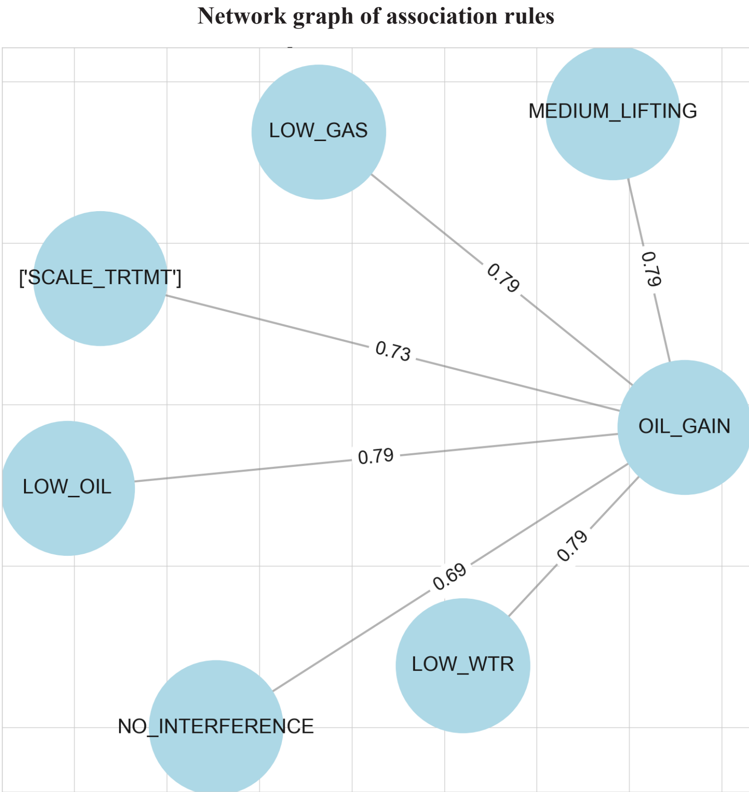

Figure 4
Network Graph of Association Rules

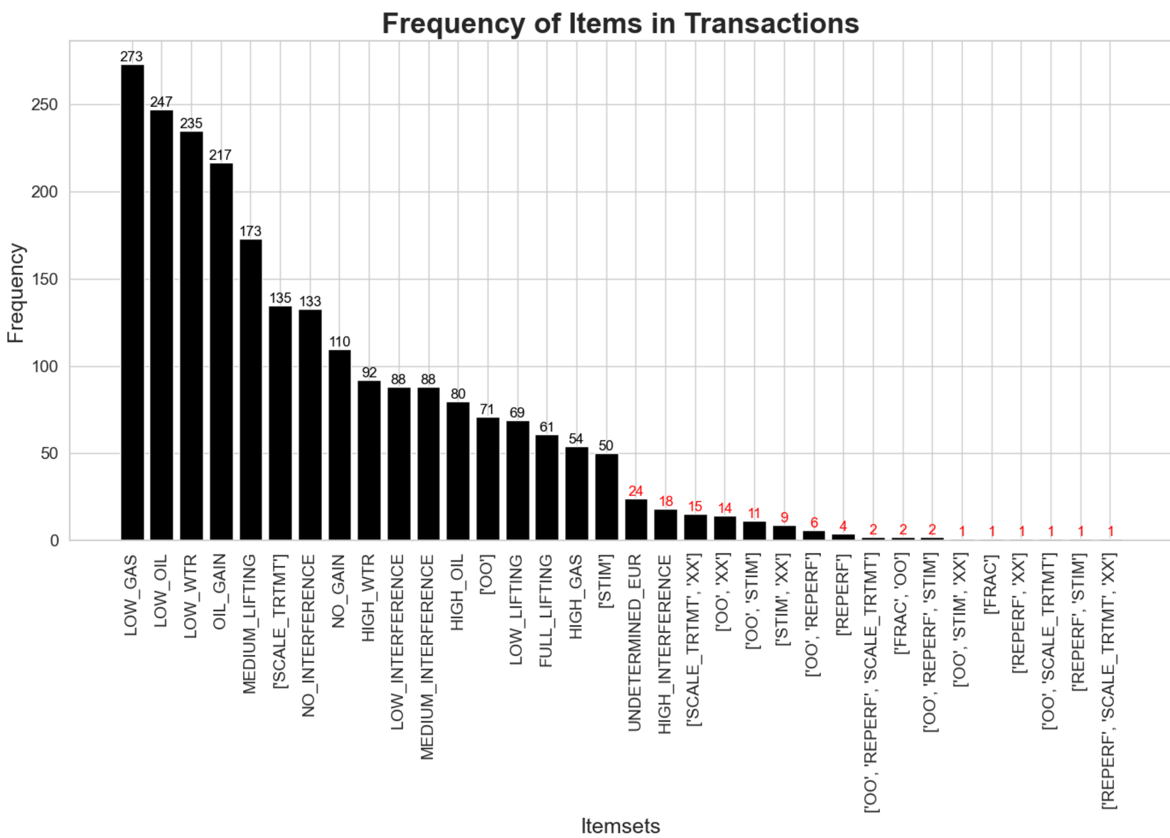

Figure 5
Frequency of transaction items

Table 4
ARM rules (Heterogeneity index effects)

Rule #	antecedents	consequents	support	confidence	lift
1	No Interference, Low HI Gas, Low HI Water	Oil Gain	0.235474	0.706422	1.064516
2	No Interference, Low HI Oil, Low HI Gas	Oil Gain	0.204893	0.690722	1.040857

Table 5
ARM rules (Workover type effects)

Rule #	antecedents	consequents	support	confidence	lift
1	Scale Treatment, Low HI Water, Low HI Gas	Oil Gain	0.269113	0.651852	0.982284
2	Scale Treatment, Low HI Gas	Oil Gain	0.235474	0.675439	1.017827
3	Low HI Oil, Scale Treatment, Low HI Gas	Oil Gain	0.214067	0.707071	1.065494

Table 6
ARM rules (Lifting status effects)

Rule #	antecedents	consequents	support	confidence	lift
1	Medium Lifting, Low HI Gas, Low HI Water	Oil Gain	0.330275	0.739726	1.114702
2	Medium Lifting, Low HI Oil, Low HI Water	Oil Gain	0.293578	0.755906	1.139083
3	Low HI Water, Medium Lifting, Low HI Gas	Oil Gain	0.269113	0.687500	1.036002

Table 7
Neighboring wells effect

Rule #	antecedents	consequents	support	confidence	lift
1	No Interference, Low HI Gas, Low HI Water	Oil Gain	0.235474	0.706422	1.064516
2	No Interference, Low HI Oil, Low HI Gas	Oil Gain	0.204893	0.690722	1.040857

CONCLUSION

Based on the result, the following conclusions can be drawn: 1). Workover operations on wells with low HI values for oil, water, and gas are strongly associated with oil gain, with a success probability exceeding 79%; 2). As proven by the data, HI is a practical and reliable indicator for identifying flow inefficiencies. Addressing these inefficiencies through appropriate workover jobs (e.g., Scale Treatment can lead to oil gain with confidence level of 73%; 3). The recovery status of the well must also be taken into account. This comprises the well's recovery factor (RF) and the potential drainage from neighboring wells with confidence level above 69%.

ACKNOWLEDGEMENT

Much appreciation goes to Petroleum Engineering ITB lecturers, staff and colleagues for their direct and indirect contributions to this writing. Special gratitude is given to PT. Gada Energi for funding and resources.

GLOSSARY OF TERMS

Symbol	Definition	Unit
HI	Heterogeneity index	Dimensionless
q(t)	Production rate at t time	bbl/day
q_i	Initial production rate	Bbl/day
D_i	Initial decline rate	bbl
Np	Cumulative oil production at t time	MSTB/MMSTB
Gp	Cumulative gas production at t time	MMSCF/BSCF
W_p	Cumulative water production at t time	MSTB/MMSTB
OO	Opening Perforation	-
SCALE_TRTMT	Scale Treatment	-
STIM	Acid Stimulation	-
XX	Closed/SO/Squeezed	-
FRAC	Fracturing	-
REPERF	Reperforation	-
P&A	Plug and Abandon	-
INJ	Convert to Injector	-
FISH	Fishing	-
SCREEN	Prepack Screen	-

REFERENCES

Abdel-Basset, Mohamed, Al-Mufarej, Mishal, Al-Mutawa, Majdi, Chetri, Hom, Anthony, Elred, Al-Zaabi, Hamad, Bolanos, Nelson, Ruiz, Hector & Khalid Harami, 2018, Integrated Production Optimization Workflow Provides Robust Platform for Significant Oil Gain to a Mature Oilfield. Society of Petroleum Engineers. SPE-1931210-MS.

Agrawal, R. & Srikant, R., 1994, Fast algorithms for mining association rules. Proceedings of the 20th International Conference on Very Large Data Bases (VLDB), 487–499. https://doi.org/10.1016/B978-1-55860-153-0.50074-9.

Borgelt C., 2005, An implementation of the FP-growth algorithm. Proceedings of the 1st International Workshop on Open-Source Data Mining: Frequent Pattern Mining Implementations, 1–5. https://doi.org/10.1145/1133905.1133906.

Ego Syahrial & Hadi Purnomo, 2009, "Peningkatan Produksi Minyak dengan Injeksi CO_2 pada Lapangan Minyak Tua Sangatta Kalimantan Timur." Lembaran Publikasi Minyak dan Gas Bumi Vol. 43. NO. 2, 166 – 175. DOI: https://doi.org/10.29017/LPMGB.43.2.142.

EIA, 2023, Global Oil Market Trends. U.S. Energy Information Administration.

ESDM, 2022, Energy Policy Framework 2022. Ministry of Energy and Mineral Resources of Indonesia.

ESDM, 2023, Annual Energy Report of Indonesia. Ministry of Energy and Mineral Resources of Indonesia.

Galicia J., 2021, Data Analytics in the Optimization of Mature Fields. Energy Analytics Journal.

Han J., Pei J. & Yin Y., 2000, Mining frequent patterns without candidate generation. Proceeding of the 2000 ACM SIGMOD International Conference on Management of Data, 1–12. https://doi.org/10.1145/342009.335372.

Hipp J., Güntzer U. & Nakhaeizadeh G., 2000, Algorithms for association rule mining—A general survey and comparison. ACM SIGKDD Explorations Newsletter, 2(1), 58–64. https://doi.org/10.1145/360402.360421.

Harami K. et al., 2013, Heterogeneity Index and Fast Screening Processes Used in a Large Mature Field Designed for Quick Production Gains. SPE 167350.

IEA, 2022, World Energy Outlook 2022. International Energy Agency.

IPCC, 2021, Impact of Energy Transitions on Oil and Gas Reserves. Intergovernmental Panel on Climate Change.

Mardiana, D.A., Rakhmanto, P.A., Riswati, S.S., Sofilda, E., Aribawa, S. & Martines, G., 2024, Petroleum Fiscal Regimes Attractiveness in Indonesia, Malaysia and Thailand: Application on Offshore Project Development. Scientific Contributions Oil & Gas Vol 47 No. 3, 256-276. https://doi.org/10.29017/SCOG.47.1.1613.

Maurenza, Farrah, Yasutra, Amega, Tungkup, Iswara Lumban, 2023, Production Forecasting Using Arps Decline Curve Model with The Effect of Artificial Lift Installation. Scientific Contributions Oil & Gas Vol 46 No. 1, 9-18. https://doi.org/10.29017/SCOG.46.1.1310.

Parshall J., 2012, Revitalizing Mature Oil Fields: A Global Perspective. Journal of Petroleum Technology.

Reese R.D., 1996, Completion Ranking Using Production Heterogeneity Indexing. SPE 36604.

Turnip, J.F., Yasutra, A., Maulana & F. Mustaqim, S.D., 2024, Mature Field and Well Revitalization of Matrix Acidizing Candidates. Scientific Contributions Oil & Gas Vol 47 No. 1, 65-74. https://doi.org/10.29017/SCOG.47.1.1613.

Schlumberger, 2022, Improved Oil Recovery Techniques in Mature Fields.

Wood Mackenzie, 2023, Indonesia's Oil and Gas Sector in Transition.