



TESTING CENTER FOR OIL AND GAS LEMIGAS

SCIENTIFIC CONTRIBUTIONS OIL AND GAS (SCOG)

Vol. 49. No. 3, September: 131 – 151 - <https://journal.lemigas.esdm.go.id>

ISSN: 2089-3361, e-ISSN: 2541-0520

Published by:

LEMIGAS

Unsupervised Machine Learning for Reservoir Rock Type Classification in the Ngimbang Formation, Northeast Java Basin, Indonesia

Elizia Da Costa Alves*, A.K.M. Eahsanul Haque, and M. Azuwan B. Maoinser

*Department of Petroleum Engineering, Universiti Teknologi PETRONAS,
32610 Seri Iskandar, Perak Darul Ridzuan, Malaysia.*

*Corresponding author: Elizia Da Costa Alves (elizia_25013388@utp.edu.my)

Received: 16 April 2026; Revised: in revised form 30 July 2026; Accepted: 02 September 2026
Available: online 30 September 2026.

ABSTRACT - Reservoir characterisation is a fundamental step in hydrocarbon exploration and production, requiring accurate identification of rock types and productive zones. Conventional petrophysical analysis alone is often insufficient to capture the full complexity of heterogeneous carbonate formations. This study applies unsupervised machine learning, specifically K-means clustering, to classify reservoir rock types within the Lower Ngimbang Formation of the AXEL Field, Northeast Java Basin, Indonesia. Wireline log data from two wells, AXEL-1 and AXEL-2, comprising Gamma Ray (GR), Deep Resistivity (ResD), Neutron Porosity (NPHI), Bulk Density (RHOB), and Sonic (DT) logs were used as input. Quantitative petrophysical analysis was first conducted to estimate shale volume (Vsh), effective porosity (PHIE), and water saturation (Sw) using the Indonesian Equation. K-means clustering was subsequently applied, starting from 20 initial micro-clusters that were progressively consolidated into five geologically meaningful rock types: Hydrocarbon, Water-Bearing, Tight Formations, Coal, and Shale. The selection of five final groups was guided by a two-step validation approach using the Cluster Groups Randomness variance ratio generated within Interactive Petrophysics, which confirmed K=5 as the most geologically defensible configuration. In AXEL-1, Zone 1 yielded a net pay of 46.82 m (avg PHIE = 0.089, avg Sw = 0.352), confirmed by DST gas flow rates of up to 7.599 MMCFGPD, while AXEL-2 Zone 1 returned a net pay of 12.50 m (avg PHIE = 0.095, avg Sw = 0.057), reflecting lower hydrocarbon potential. The Hydrocarbon Potential facies (Group 3) identified by K-means showed strong spatial agreement with conventionally defined productive zones. These results demonstrate that unsupervised machine learning provides a practical and reproducible complementary framework for reservoir characterisation and rock type classification in complex carbonate reservoirs.

Keywords: unsupervised machine learning, reservoir characterization, rock type determination, net productive layer.

How to cite this article:

Elizia Da Costa Alves, A.K.M. Eahsanul Haque, and M. Azuwan B. Maoinser, 2026, Unsupervised Machine Learning for Reservoir Rock Type Classification in the Ngimbang Formation, Northeast Java Basin, Indonesia, Scientific Contributions Oil and Gas, 49 (3) pp. 131-151.



© 2026 The Author(s). Published by LEMIGAS under the Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0)

<https://doi.org/10.29017/scog.v49i3.2147> | 131

1. Introduction

The oil and gas industry is still an important sector of the world economy. The effectiveness and efficiency of hydrocarbon development and production are heavily influenced by a good understanding of the reservoir characteristics. Accurate characterisation of the reservoir and mapping of the distribution of hydrocarbons can increase productivity and optimise the management of the reservoir. Therefore, the use of advanced analytical technologies, including machine learning approaches, in reservoir characterisation can help in better understanding of reservoir distribution and enable better decision-making in hydrocarbon development and production (Handoyo et al., 2025; Rohmana et al., 2026).

The East Java Basin is one of the most productive hydrocarbon provinces in Indonesia, covering onshore and offshore areas in northern East Java, including the islands of Bawean, Madura, and Kangean (Ascaria 2024; Ramdhan 2021). The East Java Basin is an east-west trending back-arc basin parallel to the island of Java and was formed in the late Oligocene (Handoyo et al., 2025). The basin has a long exploration history, with more than 700 exploration wells drilled and total resources found of more than 4 billion barrels of oil equivalent (BOE) from 143 fields (Wijaya 2022). The total resources discovered are about 66% oil, 33% gas, and 1% condensate (Ascaria 2024).

The Miocene Kujung Limestone has historically accounted for most hydrocarbon resources in the Northeast Java Basin, contributing more than half of the total discovered volumes (Darmawan & Wijaya 2022). As this prospect is increasingly considered mature and well-explored, exploration focus has shifted toward deeper and less-examined Paleogene intervals, most notably the Eocene-Oligocene Ngimbang Formation (Ascaria et al., 2024; Darmawan & Wijaya 2022). The Ngimbang Formation represents the oldest Cenozoic succession in the basin. It comprises Middle Eocene to Early Oligocene syn-rift deposits that transition from lacustrine-deltaic sandstones and mudstones in its lower section to deep-marine mudstones interbedded with turbidite sandstones in its upper section (Handoyo et al., 2025; Ramdhan 2021; Sinaga & Haris 2018; Wiryadinata 2025). This formation serves as one of the primary reservoirs in the Northeast Java Basin, with carbonate units that have been subjected to intensive diagenetic modification, including karstification, dissolution, and fracturing, generating substantial secondary porosity (Darmawan & Wijaya 2022; Wiryadinata & Haris 2025).



Figure 1.
Location map of the study area showing the AXEL-1 and AXEL-2 wells
in the Northeast Java Basin (modified from geological map).

Recent exploration successes have confirmed the significant potential of Ngimbang carbonate reservoirs. Wells on the North Madura Platform have encountered excellent reservoir quality, with average porosity values of 25-30% and locally exceeding 35% in highly karstified carbonate bodies (Darmawan & Wijaya 2022). Carbonate reservoir development in the Eocene Ngimbang progressed through three recognizable stages: initiation, aggradation, and backstepping. Secondary porosity evolved under the combined influence of subaerial exposure, fracturing, and diagenetic dissolution (Darmawan & Wijaya 2022). These exploration milestones represent a meaningful advancement in understanding the resource potential of this mature basin (Wijaya et al., 2022).

Despite its proven prospectivity, reservoir characterization in the Ngimbang Formation is constrained by considerable geological complexity. Lithologically, the formation is heterogeneous, comprising alternating sandstone reservoirs, carbonate deposits, mudstones, siltstones, and locally developed thin coal seams (Ningsih & Isnawan 2021; Rohmana et al., 2026). Shale and carbonate intervals dominate large portions of the section, acting as both cap and source rock units, while reservoir quality varies markedly across the formation (Wiryadinata & Haris 2025). This primary heterogeneity is further complicated by post-depositional diagenetic overprinting, dolomitization, cementation, compaction, and dissolution, generating a broad spectrum of microstructures and petrophysical responses (Nurbilad & Suprayogi 2021; Winardi et al., 2024).

A critical concept underpinning reservoir evaluation is the identification of the net productive layer (net pay), defined as the reservoir thickness satisfying established productivity criteria based on threshold values of porosity, resistivity, and water saturation (Ficry & Rosiani 2025). These threshold parameters, commonly referred to as cut-off values, serve to differentiate productive intervals from non-productive ones. Porosity, water saturation, and shale volume constitute the most widely applied parameters for net pay determination in industry practice (Mohammadi Behboud et al., 2023). The application of cut-off criteria is essential for reducing upscaling bias, improving efficiency in dynamic modeling, and establishing the basis for recovery factor estimation (Mohammadi Behboud et al., 2023).

Conventional petrophysical analyses frequently fall short in fully resolving reservoir heterogeneity, which can lead to misidentification of productive zones and biased estimates of reservoir properties (Rohmana et al., 2026). Manual wireline log interpretation is time-intensive and inherently subjective, and it often fails to capture the complex, nonlinear relationships between log responses and rock properties, particularly in heterogeneous reservoirs (Rohmana et al., 2026). Core analysis provides direct, quantitative measurements but remains costly and inherently limited to the sparsely sampled intervals within individual wells (Rohmana et al., 2026). In heterogeneous formations such as the Ngimbang, manual rock typing is especially prone to inconsistency and error (Munir et al., 2024).

Machine learning has shown considerable promise in addressing geological interpretation challenges, with growing adoption across subsurface disciplines (Munir et al., 2024; Rohmana et al., 2026). The integration of machine learning into rock type identification workflows enables more systematic distinction between productive and non-productive intervals, ultimately improving the reliability of static reservoir models (Munir et al., 2024). A wide range of algorithms, including support vector machines, random forests, and various neural network architectures, has been successfully applied to electrofacies prediction, porosity and permeability estimation, and water saturation determination (Rohmana et al., 2026; Ulil et al., 2025).

Among machine learning approaches, unsupervised learning methods offer advantages for characterizing heterogeneous reservoirs (Glover et al., 2022; Rohmana et al., 2026). Unlike supervised classification methods, which require predefined class labels and labeled training datasets, unsupervised algorithms identify natural groupings in data without prior knowledge of the number or character of facies present (Glover et al., 2022). Clustering represents one of the foundational approaches in data mining, grouping samples such that intra-group similarity is maximized while inter-group dissimilarity is minimized (Shili 2023). K-means clustering, a widely adopted partitioning method, achieves this by iteratively assigning data points to the nearest cluster centroid and updating centroids until convergence (Shili 2023).

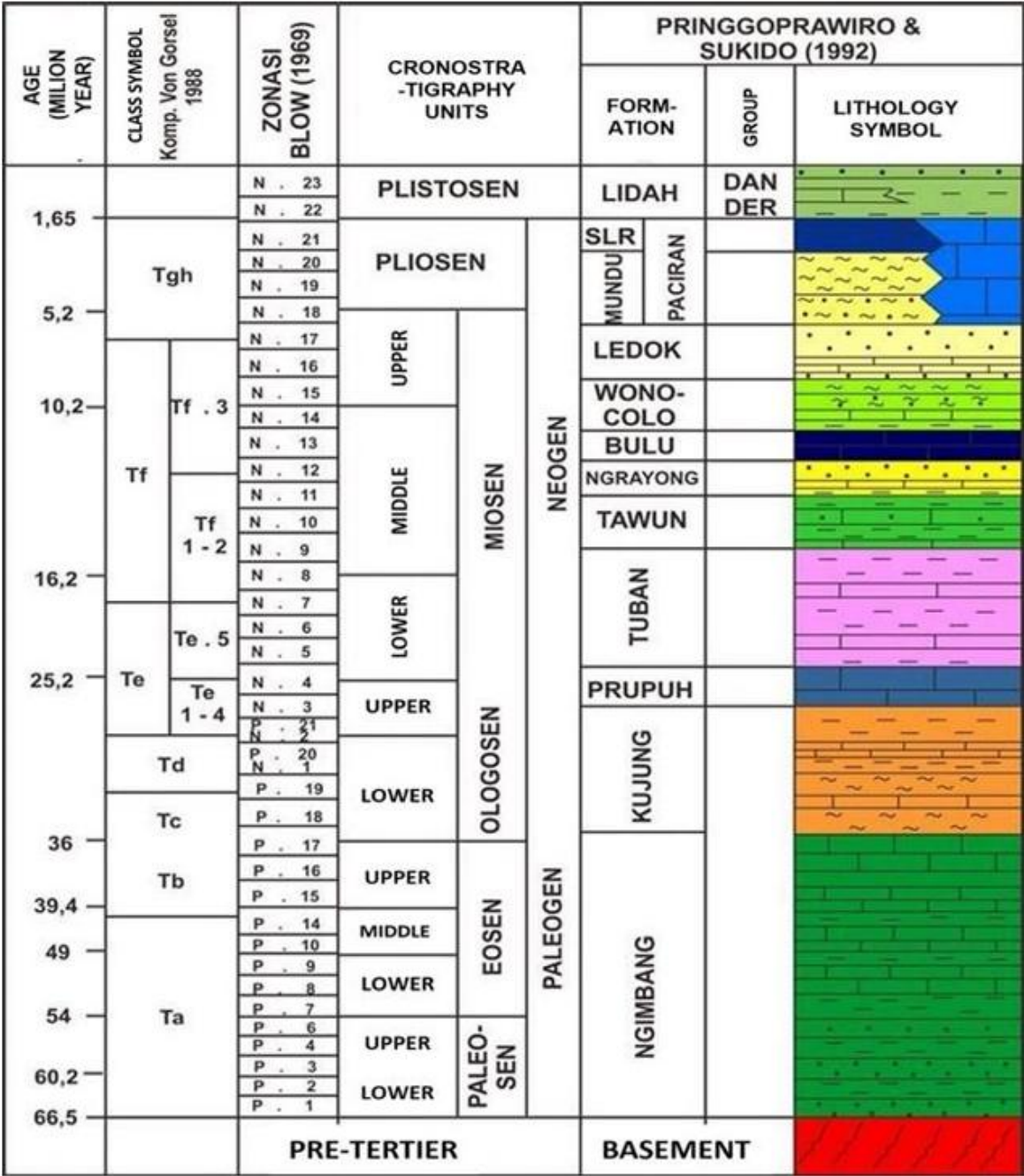


Figure 2.
Stratigraphy of the Northeast Java Basin showing the Ngimbang Formation as the main target interval modified
(Handoyo et al., 2025; Pringgoprawiro 1983).

Several studies have demonstrated the effectiveness of clustering approaches for reservoir rock typing. (Munir et al., 2024) applied K-means cluster analysis to a heterogeneous sandstone reservoir and successfully identified four distinct rock types, ranging from poor to excellent reservoir quality, based on electrofacies analysis. (Glover et al., 2022) conducted a systematic comparison of eight machine learning clustering methods applied to carbonate facies data, finding that Expectation Maximization (92.57% accuracy) and K-means (89.60-91.09%) performed best for petrophysical data classification. (Ashraf et al., 2024) demonstrated the utility of self-organizing maps and clustering in identifying favorable cluster distributions and characterizing six depositional facies types. (Sadeghi et al., 2024) Illustrated the integrated use of principal component analysis and K-means clustering for geochemical anomaly detection and prospecting.

This study applies K-means clustering to classify reservoir rock types within the Lower Ngimbang Formation of the AXEL Field, Northeast Java Basin. Three specific objectives guide this work: (i) to conduct a systematic petrophysical analysis of wireline log data from AXEL-1 and AXEL-2, determining key reservoir parameters including shale volume (Vsh), effective porosity (PHIE), water saturation (Sw), and net productive layer thickness; (ii) to apply K-means clustering to classify the formation into distinct rock types based on multivariate petrophysical attributes; and (iii) to evaluate the consistency and validity of the machine-learning-derived rock types through comparison with conventional petrophysical interpretation and validation using Drill Stem Test (DST) data. The findings are expected to improve the understanding of reservoir heterogeneity and demonstrate the potential of machine learning for carbonate reservoir characterization.

Table 1.
Comparison of previous K-means clustering applications and the present study

Study	Application	Validation approach	Key contribution
(Munir et al., 2024)	Rock type identification and reservoir quality classification in heterogeneous sandstone reservoirs (Pab Formation, Zamzama Gas Field)	Correlation with conventional petrophysical interpretation (lithology, porosity, permeability, and water saturation) and corroboration with core Findings	Developed an integrated electrofacies workflow using core-calibrated logs to provide continuous and accurate rock quality classification for improved reservoir modeling in uncored wells
(Glover et al., 2022)	Facies recognition and separation in tight carbonate reservoir analogues, including the Portland, Purbeck, and Solnhofen limestones	Clustering success rates judged against known stratigraphical classifications of 307 core samples and manually determined petrofacies for wireline data.	Conducted a systematic evaluation of eight clustering algorithms to identify optimal methods and petrophysical attribute combinations for automatic facies separation in heterogeneous carbonates
(Ashraf et al., 2024)	Classification of gas-bearing depositional facies and reservoir rock types in the heterogeneous Lower Goru Formation, Pakistan	Reliability testing using DBSCAN meta-cluster attribute analysis for outlier detection and variable significance evaluation through SHAP models	Utilized a robust unsupervised strategy combining SOM and K-means to characterize symmetry and resemblance in depositional facies patterns to identify payable reservoir zones
(Sadeghi et al., 2024)	Mapping and targeting of prospective areas for lithium exploration using till geochemical data in central Sweden	Spatial correlation analysis with known lithium mineralization occurrences via Distance Distribution Analysis (DDA) and comparison with regional bedrock geology	Integrated PCA and K-means clustering to differentiate multi-element geochemical footprints, enabling the "unmixing" of significant mineralization-related anomalies from false ones in till data
Present Study	Reservoir rock type classification in the heterogeneous carbonate intervals of the Lower Ngimbang Formation, Northeast Java Basin	Evaluation against conventional petrophysical cut-off analysis results and independent validation using Drill Stem Test (DST) gas production data	Implements a two-step clustering workflow (initial micro-cluster over-segmentation followed by geological consolidation), integrates conventional petrophysical cut-off analysis with independent DST validation, and evaluates geological conditions that may lead to K-means misclassification

Despite the increasing application of K-means clustering for electrofacies and reservoir rock typing, previous studies have demonstrated its effectiveness using different clustering workflows and validation approaches. Although several studies have integrated petrophysical interpretation with clustering algorithms, differences remain in workflow design, validation strategy, and interpretation of clustering results. Previous studies commonly validate clustering results through petrophysical interpretation or crossplot analysis, while relatively few incorporate independent production data such as Drill Stem Test (DST) results or explicitly discuss the geological conditions under which K-means clustering may produce misleading classifications.

To address these differences, this study implements a two-step clustering workflow consisting of initial over-segmentation followed by geological consolidation for reservoir rock type classification in the Lower Ngimbang Formation. Compared with previous studies, the clustering results are evaluated through conventional petrophysical cut-off analysis and independently verified using Drill Stem Test (DST) production data. In addition, this study explicitly examines geological conditions under which K-means clustering may misclassify tight, highly cemented carbonate intervals where elevated resistivity is controlled by cementation rather than hydrocarbon saturation. A comparison between previous K-means clustering studies and the proposed workflow in this study is presented in Table 1. These workflow characteristics provide a more comprehensive framework for reservoir rock type classification and interpretation in heterogeneous carbonate reservoirs.

2. Methodology

2.1 Data availability and study area

This research uses wireline log data from two wells AXEL-1 and AXEL-2 in the AXEL Field, Northeast Java Basin, Indonesia. Log suites available include Gamma Ray, Deep Resistivity, Neutron Porosity, Bulk Density, and Sonic Transit Time (DT). Validation was based on DST data from AXEL-1. Based on marker correlation and qualitative log interpretation, the Lower Ngimbang Formation was divided into three reservoir units of reef build-up (Sequence 1), carbonate platform (Sequence 2), and clastic carbonate (Sequence 3) depositional settings. Table 2 summarises log availability for both wells.

In addition to the wireline log suite, two categories of independent supporting data were used to calibrate and validate petrophysical evaluation. Laboratory-derived matrix, fluid, and clay density parameters, obtained from core-based mineralogical and density analysis of representative Ngimbang Formation samples, were used to constrain the zone-specific density porosity and shale volume endpoints reported in the porosity and shale volume tables, rather than relying on generic literature-derived default values. Well-test data, specifically Drill Stem Test (DST) results from AXEL-1, provided an independent, physically measured benchmark of hydrocarbon productivity; these results, comprising measured gas flow rates over three tested intervals, are used in the Results and Discussion section to validate both the conventional petrophysical cut-off interpretation and the K-means electrofacies classification against actual well performance.

Table 2.
Wireline log availability for AXEL-1 and AXEL-2

Log type	AXEL-1	AXEL-2
Gamma ray	✓	✓
Bulk density	✓	✓
Neutron porosity	✓	✓
Sonic	✓	✓
Deepresistivity	✓	✓
Shallow resistivity	✓	✓
Caliper	✓	✓
Spontaneous potential	✓	✓

2.2 Software and reproducibility

All petrophysical calculations and K-means clustering analyses reported in this study were performed using Interactive Petrophysics (IP), version 2018, developed by Lloyd's Register (United Kingdom). IP is a commercially licensed petrophysical interpretation package; the specific internal procedures used for log standardization, missing-value treatment, centroid initialization, and convergence are described in the K-Means Clustering for Rock Type Classification section above. To support reproducibility of this workflow, this study reports the complete set of externally verifiable parameters: the specific log inputs used for clustering (GR, RHOB,

NPHI, ResD, DT), the total number of depth samples entering the clustering dataset (6,239, following exclusion of washout-affected intervals), and the two-step cluster consolidation procedure (20 initial micro-clusters merged into five final rock-type groups). The underlying wireline log dataset and the petrophysical cut-off parameters reported throughout this manuscript allow the analytical workflow to be reproduced using this or comparable petrophysical interpretation software.

2.3 Research workflow

The overall research methodology is illustrated in Figure 3, which summarizes the workflow from data preprocessing, qualitative log interpretation, and petrophysical evaluation to reservoir and pay zone determination, followed by electrofacies classification using K-means clustering.

2.4 Log data quality control

All wireline log data were quality controlled before quantitative analysis and clustering, with the calliper log being the main tool to recognise borehole condition anomalies. The depth intervals with calliper readings much larger than the nominal bit size were flagged and were not included in the petrophysical calculations and clustering dataset. Such intervals indicate borehole washout or cave-in conditions and provide unreliable log measurements due to poor borehole contact. The rest of the data were not subject to any other outlier treatment or statistical pretreatment since all log readings in the chosen intervals were geologically consistent and in the expected physical range for the Ngimbang Formation lithologies. The final clustering dataset consisted of depth samples from target intervals of both wells after elimination of the flagged washout zones.

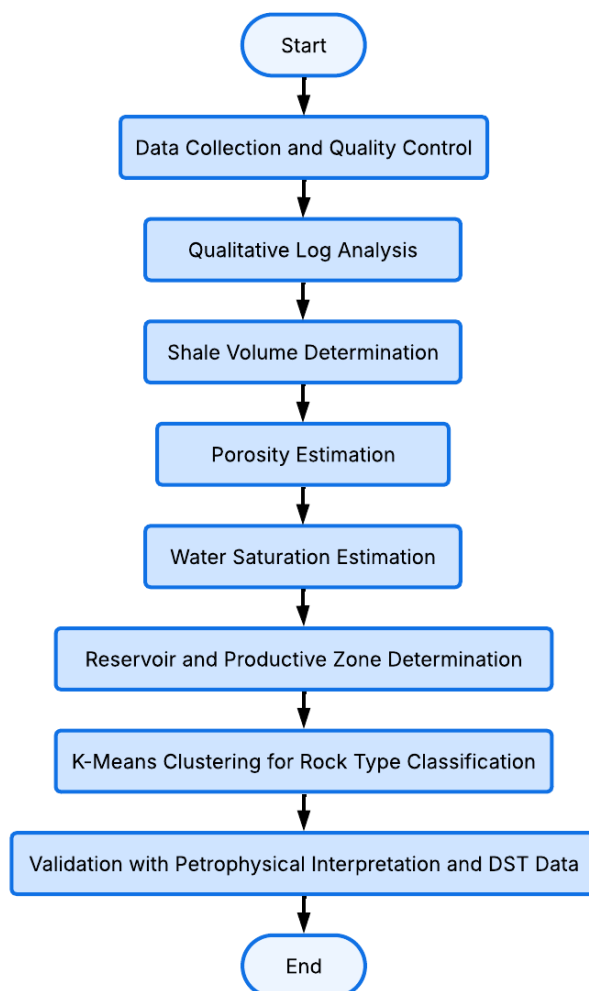


Figure 3.
Research workflow

2.5 Volume of shale

Shale volume (Vsh) was determined from gamma ray log data through the computation of the Gamma Ray Index (IGR), defined as:

$$I_{GR} = \frac{GR_{log} - GR_{min}}{GR_{max} - GR_{min}} \quad (1)$$

where GR_{log} , GR_{min} , and GR_{max} are the measured gamma ray value, and GR_{min} , and GR_{max} represent the minimum and maximum gamma ray values identified within the well, respectively (Criollo et al., 2022; Zhu et al., 2023). GR baseline parameters were determined zone-by-zone from the gamma ray histogram for each well (Table 3 for AXEL-1, Table 4 for AXEL-2), consistent with the lithological variations observed across the three target sequences. For Tertiary-age rocks, the Larionov (1969) nonlinear correction was applied to account for the overestimation of clay content by the linear index:

$$V_{sh} = 0.083 \times (2^{3.7 \times I_{GR}} - 1) \quad (2)$$

The Neutron-Density crossplot method was additionally employed to cross-validate the Vsh results derived from the gamma ray approach.

The Neutron-Density crossplot method was additionally employed to cross-validate the Vsh results derived from the gamma ray approach.

Table 3.
GR baseline parameters for V_{sh} determination AXEL-1

Zone	Method	Sand baseline (API)	Shale baseline (API)
1	GR Linear	16	61
2	GR Linear	22	77
3	GR Linear	18	62

Table 4.
GR baseline parameters for V_{sh} determination AXEL-2

Zone	Method	Sand baseline (API)	Shale baseline (API)
1	GR Linear	46	96
2	GR Linear	40	101
3	GR Linear	44	94

2.6 Porosity estimation

Formation porosity was estimated using the density porosity method (Horsfall et al., 2013), expressed as:

$$\phi_D = \frac{\rho_{ma} - \rho_b}{\rho_{ma} - \rho_f} \quad (3)$$

where ρ_{ma} , ρ_b , ρ_f are the matrix density, the measured bulk density from the RHOB log, and the fluid density, respectively. Matrix and clay parameters were calibrated zone by zone from the RHOB crossplot, accounting for the variable lithological character of each sequence (Tables 5 and 6). Effective porosity was subsequently derived by combining neutron porosity and density porosity:

$$\phi_e = \frac{\phi_N + \phi_D}{2} \quad (4)$$

Table 5.
Density parameters for porosity estimation AXEL-1

Zone	Neu Wet Clay	Den Wet Clay	Neu Shale	Den Shale	Den Dry Clay
1	0.236	2.581	0.158	2.6	2.87
2	0.289	2.63	0.195	2.64	2.8
3	0.248	2.604	0.167	2.62	2.79

Table 6.
Density parameters for porosity estimation AXEL-2

Zone	Neu Wet Clay	Den Wet Clay	Neu Shale	Den Shale	Den Dry Clay
1	0.291	2.557	0.1968	2.5849	2.809
2	0.311	2.611	0.2108	2.6227	2.809
3	0.32	2.626	0.2171	2.633	2.809

2.7 Water saturation estimation

Water saturation (S_w) was calculated using the Indonesian Equation, which accounts for the influence of clay content in resistivity-based saturation determination (Wihdany et al., 2019):

$$S_w = \left[\frac{a \times R_w}{\phi_{eff}^m \times R_t} \right]^{1/n} \quad (5)$$

where $a = 1$, $m = 2$, $n = 2$ consistent with standard assumptions for carbonate reservoir evaluation. The apparent formation water resistivity (R_w) was estimated from resistivity log data in water-bearing intervals before S_w computation.

2.8 Reservoir and pay zone determination

Reservoir and pay zone identification within the Ngimbang Formation was carried out by applying established cut-off criteria to the computed petrophysical parameters. The criteria adopted were: $V_{sh} \leq 0.50$, $PHIE \geq 0.05$, and $S_w \leq 0.60$. These thresholds were selected based on the petrophysical characteristics of the Ngimbang Formation and are consistent with values applied in comparable carbonate reservoir studies in the East Java Basin. An interval was classified as net pay only when all three criteria were satisfied simultaneously. Net-to-gross (NTG) ratio was computed as:

$$NTG = \frac{H_{net}}{H_{gross}} \quad (6)$$

where H_{net} and H_{gross} represent the net pay thickness and total gross interval thickness, respectively.

2.9 K-Means clustering for rock type classification

K-means clustering was applied as the primary unsupervised machine learning method for rock type classification. K-means was selected because it provides a simple, computationally efficient, and easily interpretable clustering approach for multivariate wireline log data, making it suitable for electrofacies classification using the functionality available in Interactive Petrophysics (IP). Compared with alternative clustering algorithms such as Gaussian Mixture Models, DBSCAN, Hierarchical Clustering, Self-Organizing Maps, and Fuzzy C-Means, K-means is relatively straightforward to implement, requires limited parameter tuning, and provides a practical balance between computational efficiency, interpretability, and geological applicability for routine reservoir rock type classification. The algorithm partitions input data into k clusters by minimizing the within-cluster sum of squared distances from each data point to its assigned cluster centroid, following an iterative

assignment and update procedure until convergence is achieved. The clustering analysis was performed in Interactive Petrophysics (IP) using five wireline log inputs: Gamma Ray (GR), Bulk Density (RHOB), Neutron Porosity (NPHI), Deep Resistivity (ResD), and Sonic Transit Time (DT). These logs were selected for their combined sensitivity to lithology, porosity, and fluid content, making them suitable for discriminating among different reservoir rock types.

Following petrophysical evaluation, the wireline log data were retained in their original calibrated form. Before clustering, the K-means implementation in Interactive Petrophysics (IP) internally standardizes each input log by subtracting its mean and dividing by its standard deviation, ensuring that all input variables contribute equally during clustering. Short-interval missing values were treated using the built-in Linear Interpolation function available in Interactive Petrophysics (IP), while no additional preprocessing for outlier treatment was required because no significant outliers were identified in the dataset. The K-means clustering analysis was performed using the default implementation available in Interactive Petrophysics (IP), and no user-defined modifications were made to the software settings. Initial cluster centroids were generated using the automatic seeding procedure implemented in IP, which performs principal component analysis (PCA) on the input data, sorts the resulting principal components, and divides the data evenly according to the specified number of clusters to determine the initial seed points. The clustering process iteratively updates cluster memberships and centroid locations until the cluster mean values no longer change between successive iterations, at which point convergence is achieved. No user-defined maximum number of iterations was applied, as the algorithm follows the default convergence procedure implemented in the software. A total of 6,239 multivariate wireline log samples from the combined AXEL-1 and AXEL-2 datasets were used as input for the K-means clustering analysis.

A two-step rock typing approach was implemented. In the first step, 20 initial micro-clusters were generated from the combined multivariate log dataset of both wells. The relatively large initial number of clusters was intentionally selected as an over-segmentation strategy to preserve subtle petrophysical heterogeneity before geological interpretation and subsequent cluster consolidation. This strategy allows the algorithm to resolve fine-scale petrophysical variability before geological consolidation, thereby reducing the risk of prematurely merging intervals that differ in lithological or fluid characteristics despite exhibiting superficially similar log responses. This approach is consistent with established electrofacies classification practice in the literature (Ashraf et al., 2024; Glover et al., 2022; Munir et al., 2024).

In the second step, the 20 micro-clusters were systematically consolidated into five final rock-type groups based on two complementary criteria. The first was the statistical signature of each cluster: clusters with closely overlapping mean values and standard deviations across all five log inputs were candidates for merging, as their indistinguishable log responses indicate they represent the same geological unit. The second was the known lithofacies assemblage of the Ngimbang Formation, which is characterised by five principal rock types: hydrocarbon-bearing porous carbonate, water-bearing, tight, coal, and shale.

The selection of $K = 5$ as the final cluster count was supported by the Cluster Groups Randomness variance ratio generated within Interactive Petrophysics (IP) (Figure 4). This diagnostic plot records the between-cluster to within-cluster variance ratio as a function of the number of clusters. The curve shows a pronounced peak at $K = 4$ followed by a gradual increase, indicating that four clusters capture the dominant variance structure in the data. However, the final selection of $K = 5$ was not based solely on the Cluster Groups Randomness criterion, but also on geological interpretability and consistency with the known lithofacies of the Ngimbang Formation. The $K = 5$ solution provides a more meaningful geological separation by distinguishing coal from shale, two lithologies with markedly different petrophysical characteristics that are merged under the $K = 4$ solution. Coal exhibits anomalously low RHOB and very high NPHI, whereas shale is characterized by high GR and low resistivity; therefore, combining these lithologies into a single cluster would represent a geologically inappropriate simplification. Consequently, the final selection of $K = 5$ was based on both the statistical criterion provided by the Cluster Groups Randomness analysis and the geological interpretability of the resulting rock types, rather than on a single statistical indicator alone. The $K = 5$ solution therefore provides a geologically meaningful classification of the five principal lithofacies present in the Ngimbang Formation and reliable discrimination of fluid zones. This approach is consistent with the findings of (Munir et al., 2024).

The five interpreted rock types and their group designations are: Water-Bearing (Group 1), Tight (Group 2), Hydrocarbon (Group 3), Coal (Group 4), and Shale (Group 5). Cluster quality was assessed qualitatively through crossplot analysis of the five input log curves coloured by cluster membership, and through direct comparison of Group 3 (Hydrocarbon) against the pay zones independently determined by conventional cut-off analysis, to confirm geological consistency of the classification.

2.10 Validation

Validation was conducted through three complementary approaches. First, the selected cluster configuration ($K = 5$), determined using the Cluster Groups Randomness variance ratio generated in Interactive Petrophysics (IP), was evaluated for its geological consistency with the known lithofacies of the Ngimbang Formation. Second, the five K-means rock types were compared against the net pay and reservoir zones determined through conventional petrophysical cut-off analysis, as well as lithological interpretations derived from log crossplot analysis. Third, Drill Stem Test (DST) production data from AXEL-1 were used as an independent physical validation reference to verify that intervals classified as Hydrocarbon Potential by the clustering algorithm corresponded to gas-productive intervals confirmed through well testing.

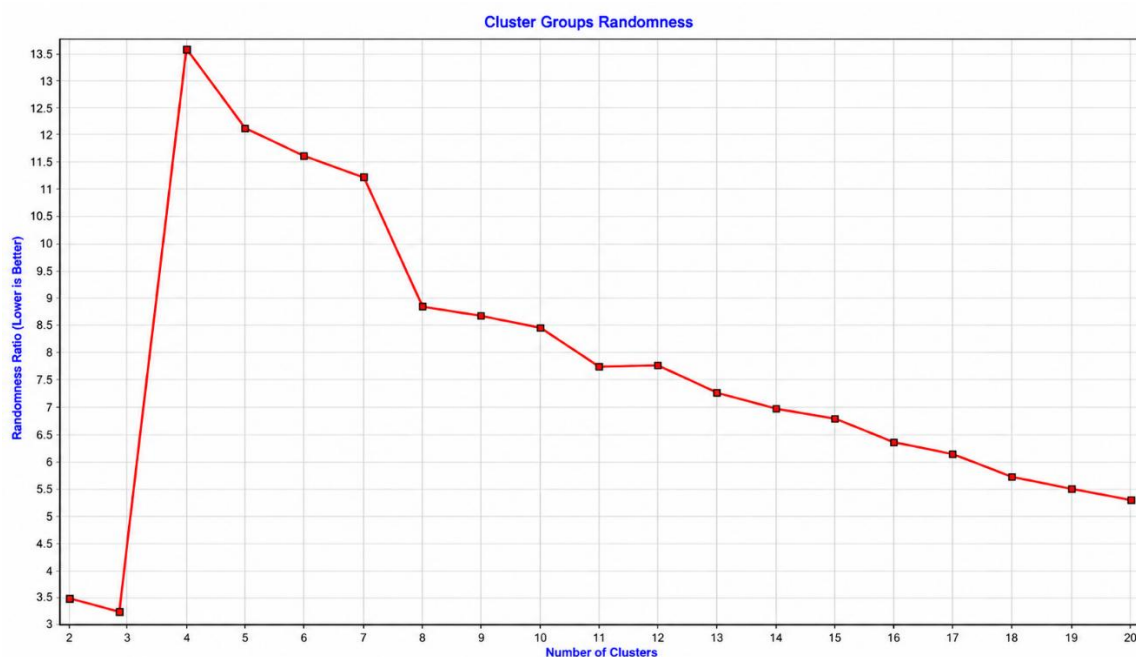


Figure 4.

Cluster Groups Randomness plot from IP to validate the number of clusters.

Annotations directly on the plot identify $K=4$ as the statistical peak of the randomness-ratio curve and $K=5$ as the configuration ultimately selected, allowing readers to locate both reference points without cross-referencing the main text.

3. Result and discussion

3.1 Petrophysical evaluation

Petrophysical analysis of the Top and Lower Ngimbang Formations identified three reservoir-quality intervals in both AXEL-1 and AXEL-2 wells based on shale volume, effective porosity, and water saturation criteria. In AXEL-1, Zones 1, 2, and 3 exhibit favorable reservoir properties ($V_{sh} < 0.5$, $\Phi_{hie} > 0.05$); however, only Zone 1 satisfies the net pay criteria ($S_w < 0.6$) and is interpreted as a hydrocarbon-bearing interval. The remaining zones contain reservoir-quality rock but are predominantly water-bearing. The reservoir properties and pay interpretation of AXEL-1 are summarized in Table 7 and Table 8.

AXEL-2 contains three reservoir-quality intervals ($V_{sh} < 0.5$, $Phie > 0.05$), yet only Zone 1 qualifies as net pay ($Sw < 0.6$). The net pay thickness is limited to approximately 12.5 m, indicating a relatively thin hydrocarbon-bearing interval and suggesting lower reservoir potential compared to AXEL-1. The reservoir properties and pay interpretation of AXEL-2 are summarized in Table 9 and Table 10.

The DST results from AXEL-1 provide direct confirmation of commercial gas accumulations within Zone 1: DST 1 (1362-1367 m) yielded 1.066 MMCFGPD, DST 2 (1357-1362 m) yielded 0.538 MMCFGPD, and DST 2A (1357-1362 m) yielded 7.599 MMCFGPD. These flow rates are consistent with the high porosity and low water saturation computed for Zone 1 from log-derived parameters.

The well-log displays for AXEL-1 and AXEL-2 are presented in Figures 4 and 5, respectively. In AXEL-1 (Figure 5), the log tracks include GR, ResD, RHOB, NPHI, V_{sh} , PHIT, Sw , and the reservoir and pay flags. Zone 1 is clearly delineated as the primary pay (red flag) by low GR response, elevated resistivity, and favorable density-neutron crossover, consistent with gas-bearing carbonate. The reservoir flag (green) marks Zones 2 and 3 as reservoir-quality rock that nonetheless fails the Sw cut-off. In AXEL-2 (Figure 6), an analogous log pattern is observed, with Zone 1 constituting the primary pay interval. However, the net pay column is notably thinner, reflecting differences in depositional setting and diagenetic history between the two wells.

Table 7.
Reservoir summary of AXEL-1

Zone	Gross	Net	N/G	Avg ϕ_{eff}	Avg Vcl	Avg Sw
1	166.62	55.39	0.332	0.086	0.084	0.393
2	98.84	15.30	0.155	0.144	0.221	0.566
3	153.46	72.14	0.470	0.104	0.136	0.751

Table 8.
Pay summary of AXEL-1

Zone	Gross	Net	N/G	Avg ϕ_{eff}	Avg Vcl	Avg Sw
1	166.62	46.82	0.281	0.089	0.073	0.352
2	98.84	3.21	0.033	0.323	0.165	0.232
3	153.46	0.61	0.004	0.206	0.154	0.481

Table 9.
Reservoir summary of AXEL-2

Zone	Gross	Net	N/G	Avg ϕ_{eff}	Avg Vcl	Avg Sw
1	128.98	14.48	0.112	0.098	0.225	0.183
2	134.11	33.53	0.250	0.060	0.044	0.638
3	195.09	4.27	0.022	0.055	0.073	0.499

Table 10.
Pay summary of AXEL-2

Zone	Gross	Net Pay	N/G	Avg ϕ_{eff}	Avg Vcl	Avg Sw
1	128.98	12.50	0.097	0.095	0.247	0.057
2	134.11	1.68	0.013	0.090	0.159	0.335
3	195.90	1.22	0.006	0.060	0.166	0.314

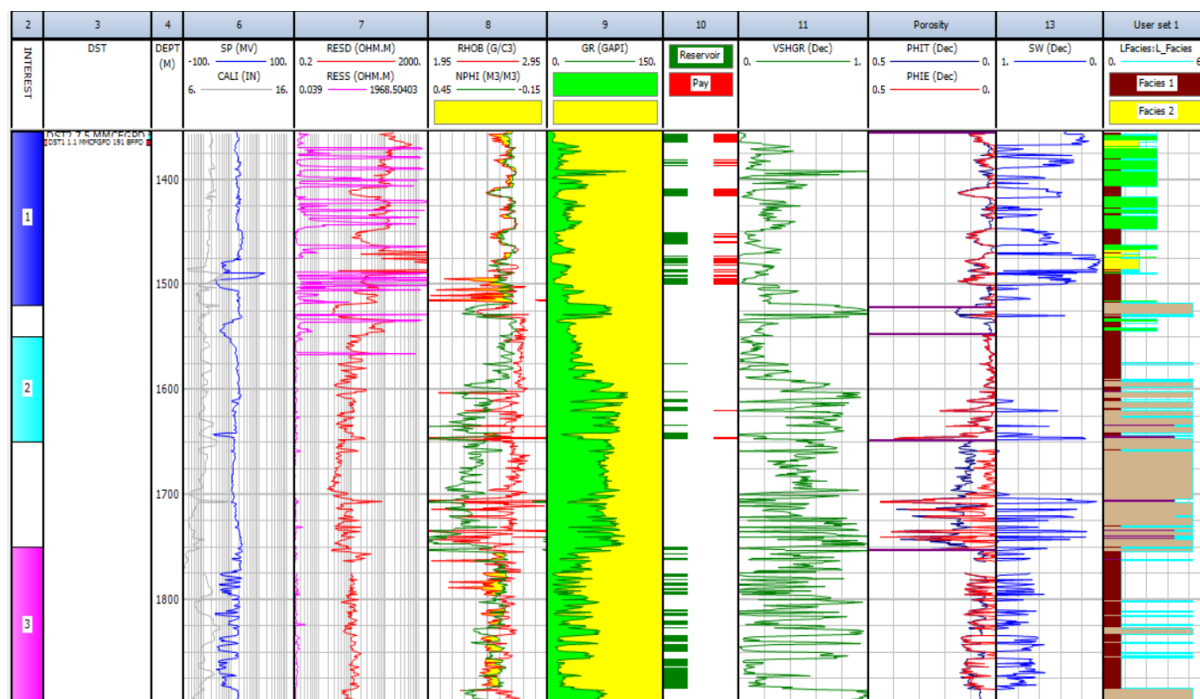


Figure 5.

Petrophysics evaluation results of AXEL-1 showing Vsh, PHIE, Sw, and NTG calculations.

Track 10 (red/green flags) marks pay and reservoir intervals derived independently from the Vsh, PHIE, and Sw cut-off criteria; the coincidence of low GR, elevated resistivity, and favourable density-neutron separation within the flagged Zone 1 interval visually confirms the quantitative pay determination.

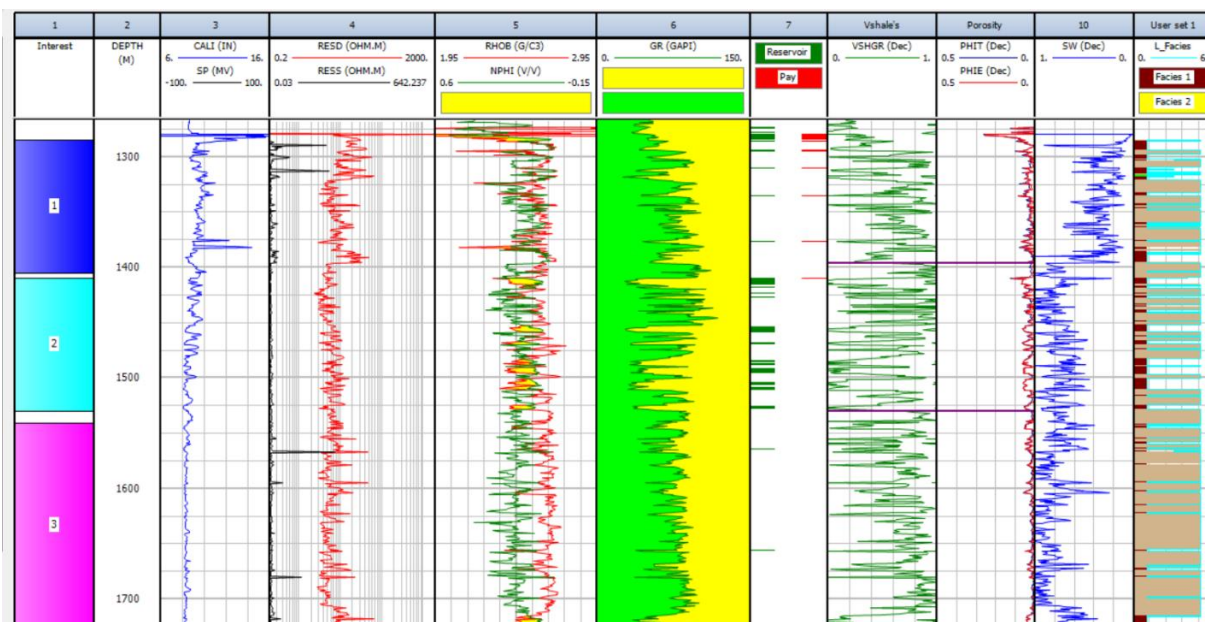


Figure 6.

Petrophysics evaluation results of AXEL-2 showing Vsh, PHIE, Sw, and NTG calculations.

As in Figure 5, the reservoir and pay flags are derived independently from the computed petrophysical parameters; the markedly thinner pay-flagged interval relative to AXEL-1 is visually consistent with the lower net-to-gross ratio reported for this well.

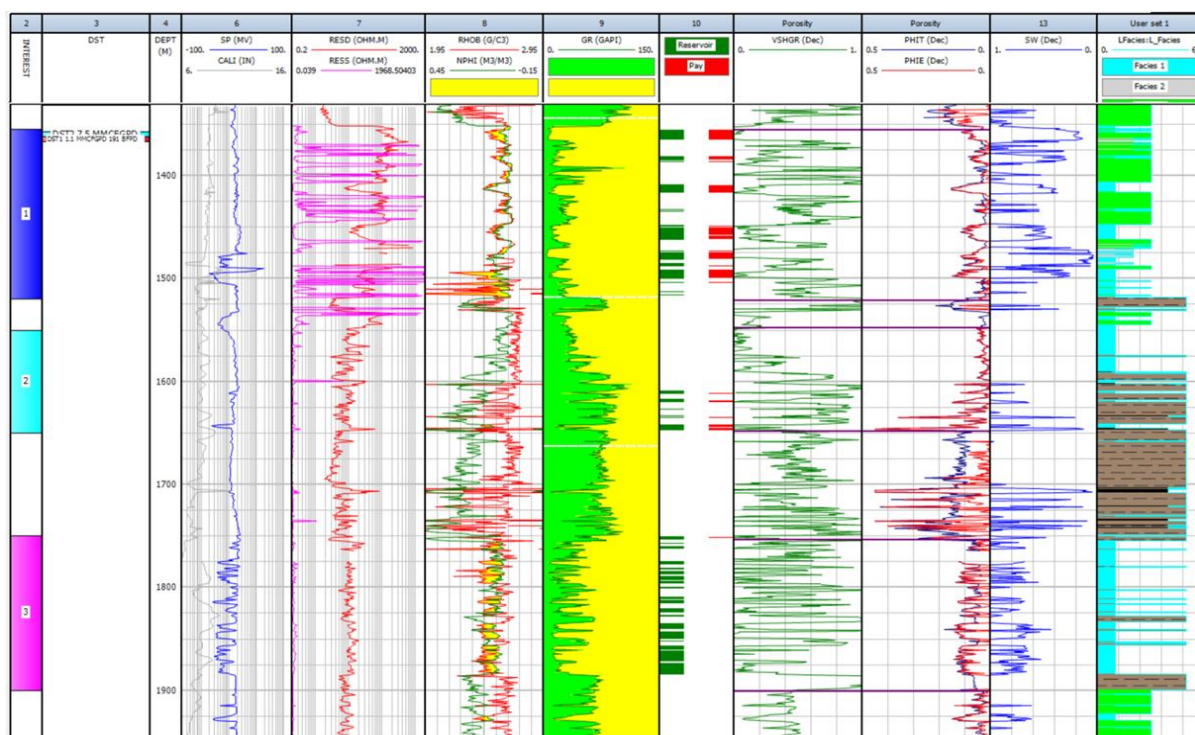


Figure 7.

Petrophysics and clustering results of AXEL-1 showing the five rock types identified in the Ngimbang Formation. The rightmost track displays the K-means rock-type classification, colour-coded per the legend in Figure 9; visual coincidence between the Hydrocarbon-Bearing interval and the conventional pay flag illustrates the agreement between the two independent classification methods discussed in Section 3.1.



Figure 8.

Petrophysics and clustering results of AXEL-2 showing the five rock types identified in the Ngimbang Formation. Compared with AXEL-1 (Figure 7), the Hydrocarbon-Bearing interval is visibly thinner and confined to the uppermost part of Zone 1, consistent with the reduced net pay and NTG reported for this well.

3.1 Pay zone validation by electrofacies

Rock typing within the Lower Ngimbang Formation of the AXEL Field was performed via electrofacies analysis using the K-means clustering algorithm. The classification was applied to multivariate log data from AXEL-1 and AXEL-2, including GR, RHOB, NPHI, ResD, and DT, and the processing was performed in Interactive Petrophysics (IP) 2018 software. K-means was suitable for separating the continuous log responses into distinct groups that can be interpreted in geological terms. The clustering results were then compared with the computed petrophysical parameters, namely Vsh, PHIE, and Sw, to assess whether the machine-learning output was consistent with conventional reservoir-quality indicators.

The analysis started with 20 initial clusters, which were then progressively merged into five geologically meaningful rock-type groups using the regrouping procedure described in Section H. The final rock types are Water-Bearing (Group 1), Tight/Non-Reservoir (Group 2), Hydrocarbon-Bearing (Group 3), Coal (Group 4), and Shale (Group 5). Their vertical distribution in both wells is consistent with the interpreted reservoir zonation, separating hydrocarbon-prone intervals from sealing and non-reservoir lithologies.

In AXEL-1 (Figure 7), the K-means classification captures lithofacies variation across the 1400-1900 m interval. Intervals assigned to the Hydrocarbon-Bearing facies (Group 3) consistently coincide with the pay zones flagged by conventional cut-off analysis. These intervals are characterised by low GR, markedly elevated resistivity, and a characteristic RHOB-NPHI crossover pattern, a combination diagnostic of gas-bearing porous carbonate. Coal intervals (Group 4) are identifiable as thin, isolated beds defined by their anomalously low bulk density. In AXEL-2 (Figure 8), Group 3 again isolates the productive zone within the 1300-1700 m depth range, characterised by depressed Sw values consistent with hydrocarbon saturation.

The statistical characteristics of each of the 20 initial clusters, including cluster size, mean, and standard deviation for every log input, are presented in Figure 10. The Hydrocarbon-Bearing facies (Group 3) is distinguished by a low gamma-ray response, reflecting minimal clay content; a reduced neutron porosity reading, consistent with the gas effect suppressing the hydrogen index typically registered in porous carbonate; and markedly elevated deep resistivity relative to the Water-Bearing facies, indicative of hydrocarbon rather than saline-water saturation. Bulk density in this facies is only moderately reduced relative to the tighter facies, indicating that the gas effect is expressed primarily through the neutron and resistivity responses rather than a pronounced density-neutron crossover alone. This cluster group aligns closely with Zone 1 as defined by conventional cut-off analysis. The Water-Bearing facies (Group 1) is characterised by moderate gamma-ray response, moderate to elevated neutron porosity, and comparatively low deep resistivity, consistent with porous carbonate saturated with conductive formation water. The Tight/Non-Reservoir facies (Group 2) displays bulk density and neutron porosity similar to the Hydrocarbon-Bearing facies, but is distinguished by markedly higher deep resistivity, reflecting cemented, low-porosity rock in which the absence of any conductive pore-filling fluid produces exceptionally high apparent resistivity, a signature that can be difficult to distinguish from hydrocarbon saturation using resistivity alone. The Coal facies (Group 4) is uniquely identified by very low bulk density and anomalously high neutron porosity combined with markedly elevated sonic transit time, properties that directly reflect the low density and high hydrogen content of carbonaceous material. The Shale facies (Group 5) is characterised by high gamma-ray response from radioactive clay minerals and low deep resistivity from clay-bound water conductivity.

The pairwise crossplot matrix (Figure 9) confirms that the five groups occupy distinct and largely non-overlapping regions in multivariate petrophysical space, particularly in GR-RHOB, GR-NPHI, and NPHI-RHOB crossplots where clay content, matrix density, and pore-fluid composition drive the primary separation. Resistivity-based crossplots further discriminate between fluid-type differences among the groups. This multivariate separation provides qualitative confirmation that the K-means groupings reflect geologically meaningful and defensible petrophysical distinctions.

Cluster	#	Cluster	GR		RHOB		NPHI		RES		SONIC		Value	Color
#	Points	Spread	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.		
1	283	0.7844	17.701	8.831	2.4905	0.07354	0.07633	0.02926	30.403	19.97	63.167	4.125	1	
2	505	0.5619	18.477	7.174	2.4533	0.04769	0.08336	0.01851	12.033	4.169	80.678	4.959	1	
3	44	1.391	23.34	11.37	2.045	0.1685	0.05788	0.03846	38.608	18.63	64.335	8.989	1	
4	38	2.44	23.577	9.793	2.6062	0.037	0.03866	0.01407	426.32	137.9	57.803	1.584	2	
5	230	1.033	23.595	15.63	2.6293	0.06675	0.02563	0.01749	163.84	36.74	57.892	5.214	3	
6	331	0.7247	26.587	11.64	2.634	0.05391	0.03076	0.01684	85.999	21.77	56.679	2.692	3	
7	292	0.6214	35.8	7.403	2.7171	0.03587	0.04919	0.01939	23.953	13.09	64.194	5.051	1	
8	34	1.476	37.389	20.38	1.4237	0.1371	0.48024	0.03061	38.997	38.29	124.74	8.143	4	
9	531	0.621	43.921	8.415	2.4567	0.05785	0.14078	0.027	8.2028	3.634	78.091	4.206	1	
10	367	0.6355	46.365	8.788	2.6511	0.05703	0.09716	0.02823	10.558	3.877	75.747	3.899	1	
11	250	0.6997	60.458	10.61	2.6265	0.05531	0.10442	0.0324	22.068	14.38	63.619	4.143	1	
12	49	1.501	66.797	21.86	1.8688	0.1502	0.38458	0.06083	12.733	10.4	105.17	8.702	4	
13	396	0.5393	67.319	7.258	2.6803	0.05212	0.21383	0.02785	6.0742	1.973	76.678	3.1	5	
14	552	0.5557	69.609	8.188	2.6042	0.04308	0.16174	0.02591	9.7386	3.319	74.182	3.55	5	
15	88	1.027	79.39	11.71	2.3603	0.1099	0.22977	0.03249	8.465	3.391	71.63	7.368	5	
16	387	0.613	81.058	9.869	2.5862	0.05605	0.21535	0.02793	5.8743	2.44	84.67	3.337	5	
17	169	1.116	81.516	13.19	2.381	0.0966	0.33739	0.06497	5.8241	4.918	94.208	6.846	5	
18	554	0.5587	82.198	7.493	2.6489	0.04253	0.29148	0.03701	4.2329	1.285	82.521	3.474	5	
19	820	0.4199	85.882	5.944	2.6597	0.03133	0.22469	0.02349	6.0947	1.843	76.018	2.858	5	
20	319	0.6852	93.299	10.19	2.5892	0.0467	0.28147	0.03669	4.2656	1.305	93.693	4.529	5	

Figure 9.

Statistics of K-means clustering results for rock type classification in the Ngimbang Formation

Values were computed within Interactive Petrophysics (IP), version 2018 (Lloyd's Register, UK); the colour column on the right-hand side indicates the final rock-type group (1-5) to which each of the 20 initial micro-clusters was consolidated.

Comparing the K-means classification against conventional cut-off-based pay identification reveals both strong agreement and an informative limitation. The Hydrocarbon-Bearing facies (Group 3) shows robust spatial correspondence with Zone 1 in both wells, and this agreement is independently confirmed by the DST production data from AXEL-1. However, a subset of intervals flagged by K-means as potentially hydrocarbon-bearing were found to carry PHIE values below the adopted cut-off threshold. This pattern reflects a specific limitation of resistivity-sensitive clustering: in tight or cemented intervals where elevated resistivity is driven by low matrix porosity and cementation rather than hydrocarbon saturation, the algorithm may incorrectly assign a hydrocarbon facies label. This tendency is expected in carbonate sequences where tight cemented intervals can produce high resistivity responses indistinguishable from hydrocarbon-bearing porous carbonates on a single-log basis. These observations confirm that K-means clustering should be employed as a complementary analytical tool rather than a stand-alone classification system.

The contrasting reservoir quality between AXEL-1 and AXEL-2 reflects the fundamental influence of depositional environment and diagenetic history on carbonate reservoir development, a relationship that has important practical implications for exploration targeting in the basin. In AXEL-1, Zone 1 is interpreted as a reef build-up sequence (Sequence 1). Reef-framework carbonates develop well-connected primary pore systems that, upon subaerial exposure during relative sea-level fall, are subjected to intensive meteoric dissolution, karstification, and tectonic fracturing. These processes create vugs, channels, and fracture networks that dramatically enhance both porosity and permeability, accounting for the exceptionally high average PHIE= 0.089 and NTG= 0.281 documented in Zone 1. The DST confirmation of commercial gas production at rates reaching 7.599 MMCFGPD validates that this diagenetic enhancement has produced a reservoir of significant productive capacity. In AXEL-2, Zone 1 corresponds to a carbonate platform facies (Sequence 2). Platform carbonates develop under conditions of more laterally uniform and extensive deposition, typically resulting in lower topographic relief and less focused diagenetic enhancement relative to isolated reef build-ups. The substantially thinner net pay (~12.5 m vs. 46.82 m in AXEL-1) and lower NTG (0.097 vs. 0.281) in AXEL-2 are consistent with this interpretation. The karstification and dissolution processes that produced the exceptional

reservoir quality in AXEL-1 were evidently less active or less volumetrically significant in the platform facies setting of AXEL-2. An alternative or complementary explanation is partial erosional truncation of the AXEL-2 pay column during a post-depositional erosion event. Taken together, the two wells establish a clear hierarchy in reservoir quality between the reef build-up and carbonate platform depositional settings, with direct implications for the siting of future exploration wells in the basin, which should preferentially target isolated reef build-up carbonate bodies within the Ngimbang Formation.

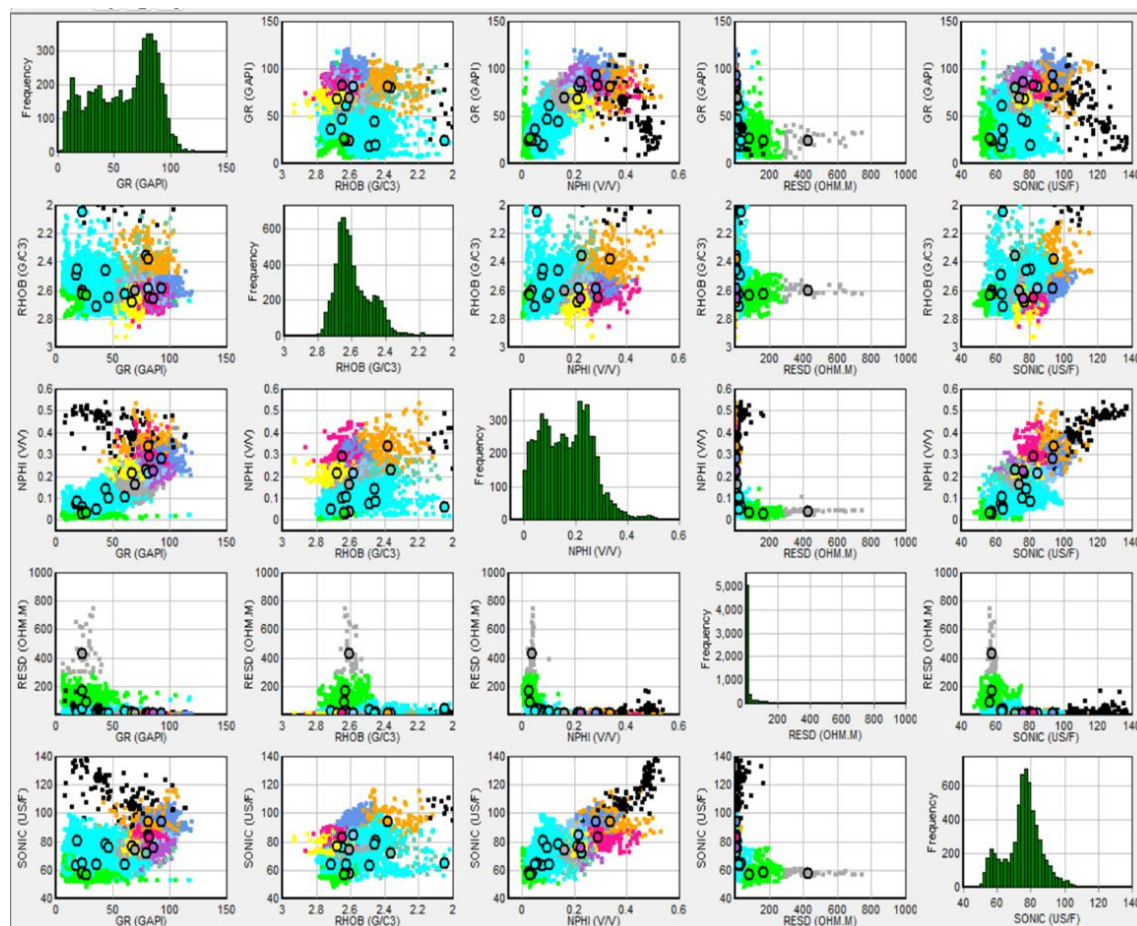


Figure 10.

Crossplot matrix of GR, RHO, NPHI, ResD, and Sonic log responses, colored by K-means cluster membership, illustrating multivariate separation of the five rock type groups. The panels are arranged as a systematic 5x5 pairwise matrix, with diagonal panels showing the univariate frequency distribution of each log and off-diagonal panels showing every two-log combination, enabling direct visual comparison of how each pair of variables separates the five rock-type groups.

The overarching conclusion drawn from this matrix is that gamma ray, combined with resistivity, provides the strongest separation between the five groups, while neutron porosity and bulk density are more effective at distinguishing Coal and Tight facies specifically. This matrix was generated within Interactive Petrophysics (IP), version 2018 (Lloyd's Register, UK), using the same five log inputs and cluster assignments underlying Figure 9.

The combined application of conventional petrophysical analysis and K-means clustering yields a more complete and robust characterisation of the Ngimbang Formation than either approach could deliver independently. The machine learning classification provides an objective, reproducible framework for identifying the principal rock type assemblages, while the conventional analysis supplies quantitative cut-off validation that guards against the resistivity-driven misclassification tendency noted above. The workflow demonstrated here can be readily extended to additional wells across the Northeast Java Basin, particularly in areas where core data are sparse or absent, as it extracts maximum interpretive value from routinely acquired wireline logs. The confirmed commercial gas production from AXEL-1 further validates the Lower Ngimbang Formation as a viable and attractive exploration objective in the basin, and the identification of reef build-up facies as the superior reservoir target provides actionable guidance for future exploration campaigns.

In addition to the resistivity-related misclassification discussed above, several other limitations should also be acknowledged. As with other K-means-based workflows, the clustering results may be influenced by the initialization process and are dependent on the selected wireline log inputs (GR, RHOB, NPHI, ResD, and DT). In addition, the validation of hydrocarbon-bearing intervals is influenced by the petrophysical cut-off values adopted in this study, which may vary according to reservoir characteristics. The inherent heterogeneity of carbonate reservoirs, including the effects of diagenesis, dissolution, cementation, and fracturing, may also produce overlapping petrophysical responses that reduce clustering accuracy in certain intervals. Furthermore, K-means assumes relatively compact clusters based on Euclidean distance and may not fully capture the complex geological variability of heterogeneous carbonate reservoirs. More advanced clustering algorithms may better represent such complex reservoir characteristics. Future studies are therefore recommended to evaluate alternative clustering approaches to further improve reservoir rock type classification.

4. Conclusion

Petrophysical analysis of wireline log data from AXEL-1 and AXEL-2 confirms that the Lower Ngimbang Formation contains productive carbonate reservoir intervals with commercially significant hydrocarbon saturation. In AXEL-1, Zone 1 was identified as the primary pay zone, with an average PHIE of 0.089, Sw of 0.352, Vsh of 0.073, and NTG of 0.281, corroborated by DST gas production of 7.599 MMCFGPD. Compared with AXEL-2, which contains a thinner net pay interval (~12.5 m) and lower NTG (0.097), AXEL-1 exhibits substantially better reservoir quality. This contrast is interpreted to reflect differences in depositional facies and diagenetic evolution, where reef build-up carbonates in AXEL-1 experienced extensive dissolution, karstification, and fracturing that enhanced porosity and reservoir connectivity.

K-means clustering of five wireline log inputs (GR, RHOB, NPHI, ResD, and DT), implemented through a two-step workflow involving 20 initial micro-clusters consolidated into five macro-clusters, successfully classified the Ngimbang Formation into five geologically meaningful rock types: Hydrocarbon Potential, Water-Bearing, Tight, Coal, and Shale. The selection of five final rock types was supported by the Cluster Groups Randomness variance ratio generated in Interactive Petrophysics and by geological interpretation, enabling effective discrimination of productive and non-productive intervals. The Hydrocarbon Potential facies (Group 3) showed strong agreement with conventionally defined pay zones and was independently confirmed by DST production data from AXEL-1.

The novelty of this study lies in integrating conventional petrophysical evaluation with unsupervised K-means clustering into a systematic and reproducible workflow for carbonate reservoir characterization in the Ngimbang Formation. Rather than replacing conventional interpretation, the machine learning approach serves as an objective and complementary tool that improves rock type classification and supports reservoir evaluation in geologically heterogeneous carbonate systems. Although K-means clustering alone may overestimate hydrocarbon-bearing intervals in tight, high-resistivity formations, integrating the clustering results with conventional petrophysical analysis significantly improves interpretation reliability. The proposed workflow provides practical guidance for reservoir characterization in wells with limited core data and offers a transferable and reproducible framework for hydrocarbon exploration and reservoir characterization in the Northeast Java Basin and analogous carbonate reservoirs.

5. Acknowledgement

The authors gratefully acknowledge the financial support provided by the Human Capital Development Fund, Timor-Leste, and the publication support from the Centre for Graduate Studies (CGS), Universiti Teknologi Petronas (UTP). The authors also thank the Supervisor and Co-Supervisor for their continuous guidance and valuable contributions to this research.

6. Declarations

6.1 Author contributions

Elizia Da Costa Alves: Data processing, interpretation, formal analysis, visualization, and writing the original draft. A.K.M. Eahsanul Haque: Supervision, Methodology, Validation, writing, and Review. M. Azuwan B. Maoinsar: Supervision, writing, and review of results and manuscript. All authors contributed to the interpretation of results and approved the final version of the manuscript. The authors declare that this manuscript is original, has not been published previously, and is not under consideration for publication elsewhere, in whole or in part.

6.2 Funding statement

The Human Capital Development Fund, Timor-Leste, supported this research. Publication support was provided by the Center for Graduate Studies (CGS), Universiti Teknologi PETRONAS (UTP). The funders had no role in the study design, data analysis, interpretation of the results, or the decision to publish this manuscript.

6.3 Competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

6.4 The use of AI or AI-assisted technologies

An AI language tool was used during manuscript preparation to assist with English language editing, grammar, and clarity of expression. It was not used to generate, analyze, or interpret any research data, figures, or results; all petrophysical calculations, clustering analysis, and scientific conclusions are the authors' own work. The authors reviewed and verified all AI-assisted text and take full responsibility for the content and accuracy of this manuscript.

7. Glossary of terms and symbols

Terms & symbols	Definition	Unit
API	American petroleum institute	-
DST	Drill stem test	-
DT	Sonic transit time	μs/ft
GR	Gamma ray log	API
HC	Hydrocarbon	-
IGR	Gamma ray index	-
NPHI	Neutron porosity	v/v
ML	Machine learning	-
NTG	Net-to-gross Ratio	fraction
PHIE	Effective porosity	fraction
PHIT	Total porosity	fraction
RHOB	Bulk density log	g/cc
ResD	Deep resistivity	ohm.m
SP	Spontaneous potential	mV
Sw	Water saturation	fraction
Vsh	Volume of shale	fraction
φ	Porosity	fraction
ρ	Density	g/cc

8. References

- Ascaria, N. A. ; P. L. ; P. B. ; L. L. C. ; D. H. (2024). Paleogene Carbonate First Trap Play in the Northeast Java Basin Using a Sub-Basin Approach. Proceedings, Indonesian Petroleum Association, Forty-Eighth Annual Convention & Exhibition. <https://proceedings.ipa.or.id>
- Ashraf, U., Anees, A., Zhang, H., Ali, M., Thanh, H. V., & Yuan, Y. (2024). Identifying payable cluster distributions for improved reservoir characterization: a robust unsupervised ML strategy for rock typing of depositional facies in heterogeneous rocks. *Geomechanics and Geophysics for Geo-Energy and Geo-Resources*, 10(1), 131. <https://doi.org/10.1007/s40948-024-00848-9>
- Criollo, D., Fam, M., Hernández, Y. F., López, A. M., & Zafra, C. (2022). Practical Methodology for Characterizing Unconventional Plays using Conventional Logs, Cores and Cuttings Analysis: A Case Study. Fourth EAGE Workshop on Unconventional Resources, 1–5. <https://doi.org/10.3997/2214-4609.202289002>
- Darmawan, H. I. ; and W. A. K. (2022). Characteristics of Eocene Carbonate Reservoir - A Case Study of Ngimbang Carbonate in North Madura Platform, East Java Basin. Proceedings, Indonesian Petroleum Association, Forty-Sixth Annual Convention & Exhibition, IPA22-G-62. <https://www.ipa.or.id/en/publications/characteristics-of-eocene-carbonate-reservoir-a-case-study-of-ngimbang-carbonate-in-north-madura-platform-east-java-basin>
- Ficry, D. ; R. D. (2025). Identifikasi Net Pay dan Zona Prospektif Hidrokarbon Berdasarkan Analisis Petrofisika Log. SNTEM. <https://sntem.akamigas.ac.id/index.php/sntem5/article/view/798>
- Glover, P. W. J., Mohammed-Sajed, O. K., Akyüz, C., Lorinczi, P., & Collier, R. (2022). Clustering of facies in tight carbonates using machine learning. *Marine and Petroleum Geology*, 144, 105828. <https://doi.org/10.1016/j.marpetgeo.2022.105828>
- Handoyo, Ronlei, B. C., Wibowo, A. S., Sigalingging, A. S., Nathania, E. Y., Fatkhan, F., Erdi, A., Avseth, P., Carbonell, R., Nugroho, P., Bayu Pandito, R. H., Nasibov, A., & Ali Husein, A. A. (2025). Reservoir Characterization of Ngrayong Formation, Sandstone with Carbonate Intercalation, Using a Geostatistical Approach Based on Petrophysical Parameters, Northeast Java Basin, Indonesia. *Scientific Contributions Oil and Gas*, 48(3), 237–251. <https://doi.org/10.29017/scog.v48i3.1828>
- Horsfall, O. I. ; O.-P. V. B. ; T. I. (2023). Correlation analysis between sonic and density logs for porosity determination in the south-eastern part of the Niger Delta Basin of Nigeria. *Asian Journal of Science and Technology*, 4(1). <https://www.journalajst.com/correlation-analysis-between-sonic-and-density-logs-porosity-determination-south-eastern-part-niger>
- Mohammadi Behboud, M., Ramezanzadeh, A., Tokhmechi, B., Mehrad, M., & Davoodi, S. (2023). Estimation of geomechanical rock characteristics from specific energy data using combination of wavelet transform with ANFIS-PSO algorithm. *Journal of Petroleum Exploration and Production Technology*, 13(8), 1715–1740. <https://doi.org/10.1007/s13202-023-01644-z>
- Munir, M. N., Zafar, M., Ali, A., Ehsan, M., Abdelrahman, K., Radwan, A. E., & Al-Awah, H. (2024). A Comparative Study Based on Petrophysical and Cluster Analysis Approach for Identification of Rock Types in Heterogeneous Sandstone Reservoirs. *ACS Omega*, 9(31), 33397–33407. <https://doi.org/10.1021/acsomega.3c08403>
- Ningsih, I. A. ; I. D. (2021). Analisis dan Interpretasi Data Well Logging Berdasarkan Parameter Petrofisika pada Sumur South Onin AI-X dan Sumur North Onin-1x di Lapangan X Papua. *Jurnal Geosaintek*, 7(2), 241–248. <https://journal.itny.ac.id/index.php/ReTII/article/view/3624>
- Nurbilad, R. R. ; S. K. (2021). Analisis Petrofisika untuk Evaluasi Formasi Reservoir Karbonat dan Formasi Batupasir pada Lapangan “Geo”-Cekungan Sumatera Selatan. *Jurnal Fisika*, 4(1).
- Pringgoprawiro, H. (1983). *Biostratigraphy and Paleogeography of North East Java Basin: A New Approach*. Institute of Technology Bandung.

- Ramdhan, A. M. (2021). Tekanan Luap di Bagian Darat dan Cekungan Jawa Timur Utara dan Implikasinya terhadap Kondisi Tegasan pada Masa Kini. *Lembaran Publikasi Minyak Dan Gas Bumi*, 55(2), 115–127. <https://doi.org/10.29017/LPMGB.55.2.609>
- Rohmana, R. C., Ariadji, T., Yasutra, A., & Irawan, D. (2026). A Conceptual Framework for Cross-Basin Reservoir Characterization: Integrating Unsupervised-Supervised Learning to Address Geological Heterogeneity. *Scientific Contributions Oil and Gas*, 49(1), 375–394. <https://doi.org/10.29017/scog.v49i1.2001>
- Sadeghi, M., Casey, P., Carranza, E. J. M., & Lynch, E. P. (2024). Principal components analysis and K-means clustering of till geochemical data: Mapping and targeting of prospective areas for lithium exploration in Västernorrland Region, Sweden. *Ore Geology Reviews*, 167, 106002. <https://doi.org/10.1016/j.oregeorev.2024.106002>
- Shili, H. (2023). Clustering in big data analytics: a systematic review and comparative analysis (review article). *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 23(5), 967–979. <https://doi.org/10.17586/2226-1494-2023-23-5-967-979>
- Sinaga, J. P., & Haris, A. (2018). Facies modeling of Ngimbang formation case study of North East Java Basin, Indonesia. 020246. <https://doi.org/10.1063/1.5064243>
- Ulil, M. R., Sonny Winardhi, & Dinanto, E. (2025). Machine Learning-Based Prediction of Shear Wave Velocity: Performance Evaluation of Bi-GRU, ANN, and The Greenberg-Castagna Empirical Method. *Scientific Contributions Oil and Gas*, 48(3), 159–171. <https://doi.org/10.29017/scog.v48i3.1797>
- Wihdany, F. I., Irawan, D., Utomo, M. R., & Fitro, A. M. (2019). INTEGRATED INITIAL WATER SATURATION MODELLING. *Scientific Contributions Oil and Gas*, 42(3), 95–107. <https://doi.org/10.29017/SCOG.42.3.396>
- Wijaya, A. K. ; R. M. H. B. ; M. M. ; D. H. I. ; I. R. B. (2022). Revealing New Hydrocarbon Potential in Mature Basin, A Case Study from North Madura Platform, East Java Basin, Indonesia. *Proceedings, Indonesian Petroleum Association, Forty-Sixth Annual Convention & Exhibition, IPA22-G-236*. <https://www.ipa.or.id/en/publications/revealing-new-hydrocarbon-potential-in-mature-basin-a-case-study-from-north-madura-platform-east-java-basin-indonesia>
- Winardi, S., Surjono, S. S., Amijaya, D. H., & Suryanto, W. (2024). Profil Perubahan Nilai Resistivitas Reservoir Pada Batu Pasir Karena Pengaruh Mineral Glaukonit. *Lembaran Publikasi Minyak Dan Gas Bumi*, 58(2), 65–79. <https://doi.org/10.29017/LPMGB.58.2.1628>
- Wiryadinata, V. ; H. A. (2025). Integration of Log Analysis and Dutta Crossplot to Identify Overpressure Zones and Hydrocarbon Potential in the Kujung and Ngimbang Formations. *Jurnal Penelitian Pendidikan IPA*, 11(11). DOI: 10.29303/jppipa.v11i11.13264
- Zhu, L.-Q., Sun, J., Zhou, X.-Q., Li, Q.-P., Fan, Q., Wu, S.-L., & Wu, S.-G. (2023). Well logging evaluation of fine-grained hydrate-bearing sediment reservoirs: Considering the effect of clay content. *Petroleum Science*, 20(2), 879–892. <https://doi.org/10.1016/j.petsci.2022.09.018>