

Litological analysis and reservoir zone identification using the crossplot and k-means clustering methods on well log data

Reza Aulia Putri¹, Abdul Gafar Karim¹, Amiruddin¹, Aprilino Alfa Kurmasela¹, Hamsir¹

¹ Program Studi Teknik Perminyakan, Sekolah Tinggi Teknologi Migas Balikpapan, Indonesia

Correspondence: adeesazkrh@gmail.com

Article Info

Article history:

Received Jun 12th, 2026

Revised Aug 20th, 2026

Accepted Aug 26th, 2026

Keyword:

Well Log, Lithology, Reservoir Zone, Crossplot, K-Means Clustering, Orange Data Mining, Kuncherinna-1 Well

ABSTRACT

Lithology identification and reservoir zone determination are critical stages in formation evaluation within the petroleum industry. This study aims to identify rock lithology, analyze the relationship between log parameters using the Crossplot method, and determine prospective reservoir zones using K-Means Clustering based on Orange Data Mining applied to open-source well log data from the Kuncherinna-1 well, obtained from the Geoscience Australia database (ECAT 149208), located in the Pedirka and western Eromanga basins, Simpson, Australia. The data used include Gamma Ray (GR), Deep Resistivity (RDEP), Density Log (RHOB), Neutron Porosity (NPHI), Sonic Log (DT), and Caliper Log (CALI) parameters over a depth interval of 2098-2825 m. Crossplot analysis was performed using four parameter combinations: GR vs RDEP, RHOB vs NPHI, GR vs NPHI, and RHOB vs DT. The optimal number of clusters was determined using the Silhouette Score, with k values tested from k=2 to k=6. Evaluation results indicate that k=4 is the optimal number of clusters, yielding the highest Silhouette Score of 0.438. K-Means Clustering produced four lithological groups: Cluster C1 as wet sand, Cluster C2 as a shale zone, Cluster C4 as a transitional shaly sand zone, and Cluster C3 identified as the most prospective reservoir zone candidate based on a combination of very low GR values (15-35 API), very high RDEP values exceeding 30,000 ohm-m, and low NPHI and DT values. The vertical distribution shows that Cluster C3 was identified only within the 2110.74–2161.18 m depth interval, indicating that the prospective reservoir zone is a thin layer located in the shallower part of the analyzed interval. This study demonstrates that the combination of Crossplot and K-Means Clustering implemented in Orange Data Mining provides an effective and objective approach for lithology interpretation and identification of prospective reservoir intervals without requiring prior lithology labels.



© 2026 The Authors. Published by Solutiva PetroTechno Journal (SPTJ). This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>)

INTRODUCTION

In the petroleum industry, understanding reservoir characteristics is one of the most crucial aspects of hydrocarbon exploration and production activities. Hydrocarbon reservoirs are subsurface rocks capable of storing and flowing fluids, the quality of which is strongly influenced by physical rock properties such as lithology, porosity, permeability, and fluid distribution. Identifying lithology and determining reservoir zones is therefore an essential first step in formation evaluation. Most hydrocarbon reservoir rocks originate from clastic sedimentary rocks, particularly sandstone, because of their favorable porosity and permeability for storing and flowing fluids (Tiab & Donaldson, 2015).

Lithology is generally defined as the physical characteristics of a rock, including rock type, texture, mineral composition, grain size, color, and internal structure, which form the basis for distinguishing one rock unit from another. However, subsurface lithology cannot be identified through direct observation as with surface outcrops, so well log data are required as an indirect interpretation tool. Well log data record the physical response of rock formations along the depth of a borehole using logging instruments lowered into the well. Each lithology type produces a unique physical response; for example, shale tends to show high gamma ray values due to its radioactive mineral content, whereas clean sandstone generally shows low gamma ray values. This difference in response is the primary basis for determining lithological boundaries and distinguishing reservoir from non-reservoir zones (Asquith & Krygowski, 2004).

In reservoir characterization, well log parameters such as Gamma Ray (GR), Deep Resistivity (RDEP), Density (RHOB), Neutron Porosity (NPHI), Sonic (DT), and Caliper (CALI) are integrated to improve the accuracy of formation evaluation (Octavia et al., 2024). Gamma Ray is used to identify shale content, resistivity to evaluate the presence and characteristics of hydrocarbon fluids, density and neutron logs to analyze rock porosity, sonic log to monitor the level of compaction, and caliper log to monitor borehole stability (Octavia et al., 2024). The conventional method commonly used to interpret these parameters is Crossplot analysis. Crossplot visualizes the relationship between log parameters (such as neutron–density or sonic–density) to identify lithological variation, map porosity quality, and detect reservoir fluid content based on data distribution patterns. Previous studies have shown that examining the relationships between well log parameters using this method can help identify lithology and reservoir zone distribution more clearly (Octavia et al., 2024).

Although effective, manual interpretation of well log data is often time-consuming and is influenced by interpreter subjectivity, particularly for large and highly complex datasets. With the advancement of digital technology, machine learning has begun to be applied to assist the analysis process more rapidly and objectively. One effective approach is unsupervised clustering, an algorithm that groups data based on the similarity of certain characteristics without requiring pre-labeled data as a reference. In wireline log data, this method can automatically group lithofacies and electrofacies with competitive accuracy, and has proven accurate in identifying gas-bearing zones and rock typing in clastic formations (Szabó et al., 2023; Hasan et al., 2023).

In this study, clustering was performed using Orange Data Mining, a visual-programming-based data analysis software. Orange Data Mining provides a wide range of machine learning algorithms and data visualization tools that can be used to process well log data without requiring advanced programming skills (Demšar et al., 2013). The use of this software is expected to support a more effective and systematic process of lithology analysis and reservoir zone determination. Building on this background, the present study analyzes lithology and determines reservoir zones using the Crossplot and K-Means Clustering methods on open-source well log data from the Kuncherinna-1 well, obtained from the Geoscience Australia database (ECAT 149208), located in the Pedirka and western Eromanga basins, Simpson, Australia. This study contributes to the petroleum engineering field by demonstrating that integrating Crossplot analysis with K-Means Clustering in Orange Data Mining provides an objective, software-assisted alternative to conventional manual log interpretation, while also serving as a reference for future research applying machine-learning approaches to well log analysis.

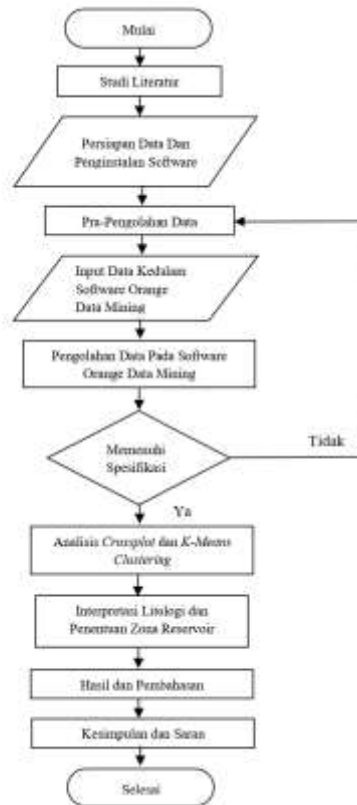
RESEARCH METHODS

This study was conducted between May 10, 2026, and June 5, 2026, using secondary, open-source well log data from the Kuncherinna-1 well, obtained from the petrophysics data package of the Australia's Future Energy Resources (AFER) project under the Exploring for the Future (EFTF) program by Geoscience Australia. According to the metadata document ECAT#149208, the Kuncherinna-1 well is geologically located in the Pedirka and western Eromanga basins, Simpson, Australia (Geoscience Australia, n.d.). The dataset, in ASCII LAS format, includes Gamma Ray (GR), Deep Resistivity (RDEP), Medium Resistivity, Density Log (RHOB), Neutron Porosity (NPHI), Sonic Log (DT), and Caliper Log (CALI). A quantitative-exploratory approach was applied using two main methods: Crossplot analysis and K-Means Clustering in Orange Data Mining. Unsupervised clustering was chosen because it can automatically and objectively identify lithology and reservoir-zone patterns based on the physical characteristics of the rock, without requiring an initial labeled dataset.

All data were processed through several stages using Orange Data Mining and Microsoft Excel, comprising pre-processing, Crossplot analysis, application of the K-Means algorithm, cluster evaluation using the Silhouette Score, and interpretation of lithology and reservoir zones based on petrophysical characteristics. Pre-processing included conversion of the LAS file to CSV format, handling of missing values, identification of data anomalies (spikes) without altering the main trend, data normalization, and depth alignment among the log curves. Based on visual inspection of the log plot for the 0–2871 m interval, the depth interval with complete data, 2098–2825 m, was selected as the input for analysis in Orange Data Mining. Caliper Log (CALI) was used as a quality-control parameter to confirm that the borehole condition within this interval was relatively stable and close to bit size, without significant washout, thereby confirming that the GR, RDEP, RHOB, NPHI, and DT data were not materially

affected by borehole-related disturbances and were suitable as input variables. CALI itself was not included as an input variable in the K-Means Clustering algorithm. The selected input variables (GR, RDEP, RHOB, NPHI, and DT) were chosen because they represent complementary petrophysical properties related to shale content, fluid characteristics, density, porosity, and acoustic response, enabling a comprehensive characterization of lithology.

The complete workflow built in Orange Data Mining began with the File widget to import the pre-processed dataset, followed by the Data Table widget to verify data structure, and the Distributions and Column Statistics widgets for descriptive statistical analysis. Column Statistics confirmed that none of the log parameters contained missing values (0%). The data were then passed to the Select Columns widget to select GR, RDEP, RHOB, NPHI, and DT as input features, followed by the k-Means widget for the clustering process. The clustering results were visualized through four Scatter plot widgets, each displaying a Crossplot of GR vs RDEP, RHOB vs NPHI, GR vs NPHI, and RHOB vs DT, with data points automatically colored according to the resulting cluster labels.



Picture 1 Research Methodology Flowchart

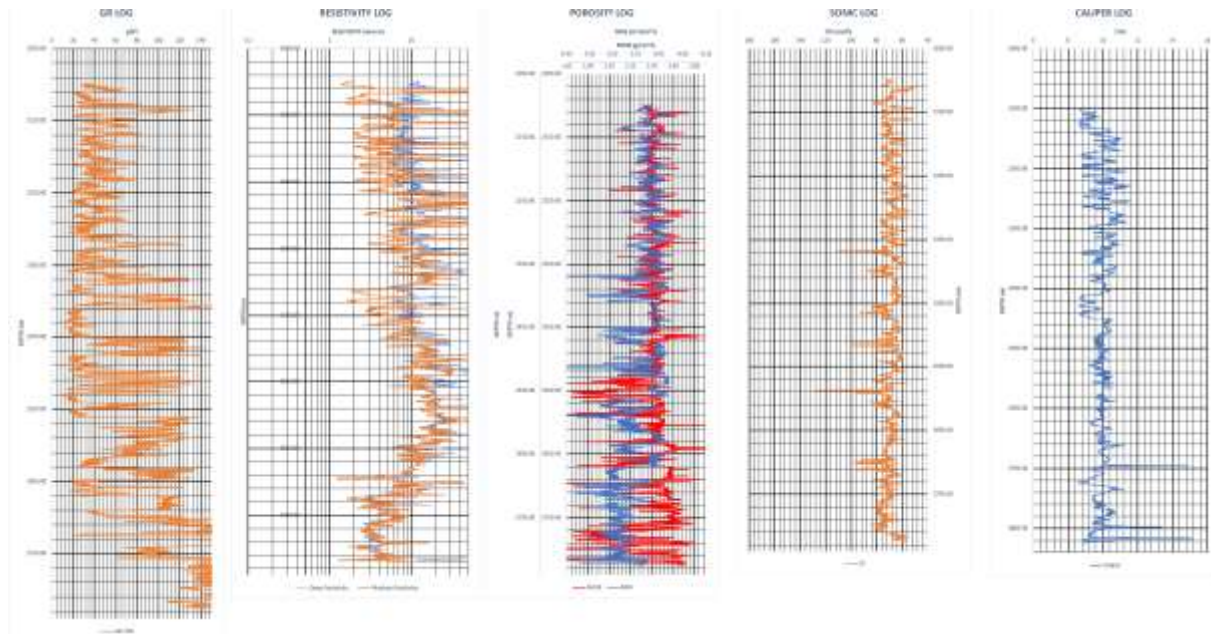
The optimal number of clusters (k) was determined by testing k values from 2 to 6 and evaluating clustering quality using the Silhouette Score available in the k-Means widget. The Silhouette Score assesses clustering quality by comparing the cohesion of a data point with members of its own cluster against its separation from the nearest different cluster, producing a value ranging from -1 to 1, where values closer to 1 indicate better cluster separation (Rousseeuw, 1987; Suyal & Sharma, 2024). The Euclidean distance between each data point and its centroid is calculated using Equation (1):

$$D(P, Q) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (1)$$

where $D(P, Q)$ is the Euclidean distance between point P and point Q , (x_1, y_1) are the coordinates of point P , and (x_2, y_2) are the coordinates of point Q (Suyal & Sharma, 2024). In this study, the same Euclidean distance principle is generalized to the five-dimensional feature space formed by the GR, RDEP, RHOB, NPHI, and DT input variables, as implemented by the k-Means widget in Orange Data Mining.

RESULTS AND DISCUSSION

Pre-processing was carried out before the Crossplot and K-Means Clustering analyses to prepare the Kuncherinna-1 well log data used in this study, so that the analyzed data conformed to the objective of identifying lithology and determining reservoir zones. Based on visual inspection of the log plot for the 0–2871 m interval performed at the data-preparation stage, the entire depth interval of the Kuncherinna-1 well with complete data, 2098–2825 m, was used as the input for analysis in Orange Data Mining, as shown in Figure 2.



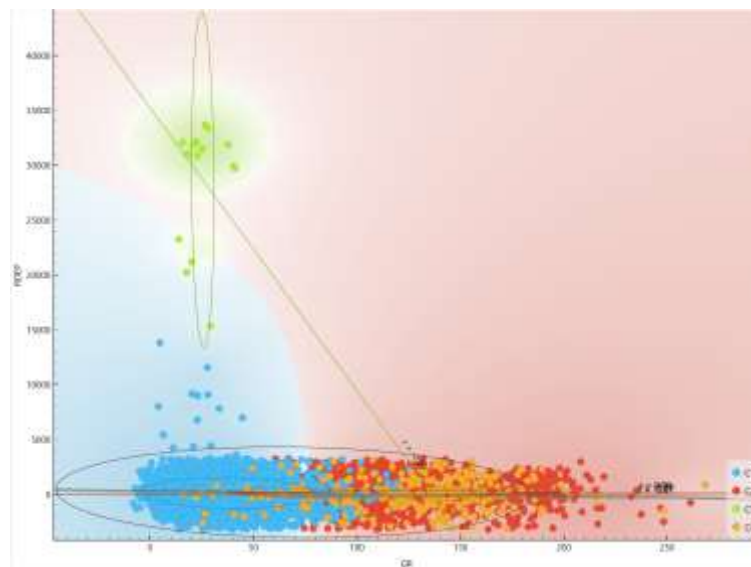
Picture 2 Log Plot of the 2098-2825 m Depth Interval

As shown in Figure 2, the GR log fluctuates with several low-value intervals indicating the presence of clean sand among the dominant shale zones throughout the interval. The Resistivity log shows two curves, Deep Resistivity (RDEP) and Medium Resistivity (RMED); increases in resistivity occur at certain intervals, where high resistivity values potentially represent hydrocarbon zones, while low values indicate wet sand (water-bearing sand) zones. In this study, RDEP was used as the input variable for the Crossplot and K-Means Clustering analyses because it measures resistivity in the zone farther from the borehole, where the influence of drilling mud filtrate is relatively smaller. The Porosity log displays RHOB and NPHI curves simultaneously; a crossover phenomenon between the two curves at several depth intervals indicates the possible presence of hydrocarbon fluid in the pore space. The Sonic log shows DT variations consistent with the lithological changes observed in the other curves, where higher DT values at certain intervals reflect greater porosity and are generally associated with shale or less compact rock. The Caliper log indicates a relatively stable borehole condition, with most of the interval showing caliper values close to bit size, confirming reliable and representative log data for the subsequent Crossplot and K-Means Clustering analyses.

Crossplot Results

Crossplot analysis was carried out using the Scatter plot widget in Orange Data Mining to examine the relationships among the log parameters used in this study. After K-Means Clustering was performed, the clustering results were re-visualized on each Crossplot by coloring the data points according to the resulting cluster labels, facilitating interpretation. The selection of the four Crossplots was based on the following considerations: (1) GR vs RDEP, to distinguish shale from clean sand based on shale content and formation resistivity; (2) RHOB vs NPHI, to identify lithology type and detect possible gas zones through the crossover phenomenon of the two curves; (3) GR vs NPHI, to evaluate the influence of shale content on the measured neutron porosity value; and (4) RHOB vs DT, to analyze the physical characteristics and compaction level of the rock based on the relationship between density and acoustic properties of the formation.

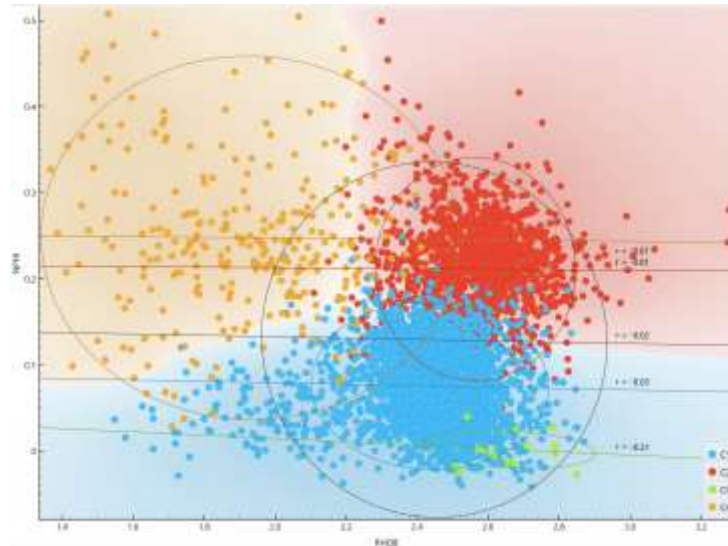
The GR vs RDEP Crossplot (Figure 3) was used to analyze the relationship between lithology and formation fluid content in the Kuncherinna-1 well. Using the K-Means Clustering algorithm in Orange Data Mining, the data distribution was objectively grouped into four main clusters. Cluster C1 occupies a GR range of 15–80 API with RDEP values spreading up to more than 10,000 ohm-m, interpreted as wet sand. Cluster C2 occupies a GR range of 80–270 API with RDEP values close to zero, representing a shale (non-reservoir) layer. Cluster C4 has GR values of 20–90 API and low RDEP close to zero, indicating shaly sand characteristics. Meanwhile, Cluster C3 is concentrated at the lowest GR values, 15–35 API, together with the highest RDEP values exceeding 30,000 ohm-m (maximum 32,813.7 ohm-m), and is therefore interpreted as the most prospective reservoir zone candidate.



Picture 3 Crossplot GR vs RDEP

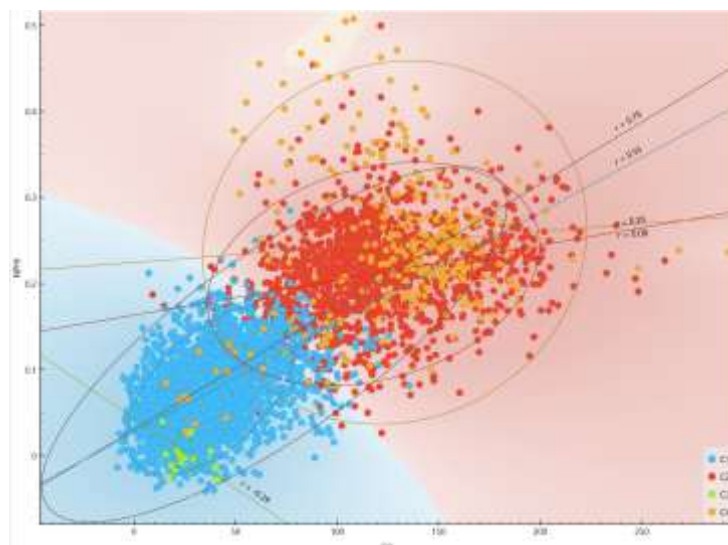
The combination of low GR and high resistivity in Cluster C3 indicates a clean sand lithology with possible hydrocarbon content, consistent with the physical principle that shale is characterized by high gamma ray and low resistivity values, while hydrocarbon-saturated sandstone is characterized by low gamma ray and high resistivity values (Ali et al., 2024). The overall correlation value for this Crossplot ($r = -0.09$) indicates a weak linear relationship between GR and RDEP globally, owing to formation heterogeneity. Based on these characteristics, Cluster C3 is interpreted as the most prospective reservoir interval based on its log gamma ray, high resistivity, low density, low neutron porosity, and relatively low sonic travel time in Kuncherinna-1.

The RHOB vs NPHI Crossplot (Figure 4) was used to identify lithology characteristics and porosity quality of the formation. Using K-Means Clustering, the log data distribution was separated into four main groups. Cluster C1 occupies an RHOB range of 2.2–2.9 g/cc and NPHI of 0.00–0.15, indicating clean sand with low shale content. Cluster C2 occupies an RHOB range of 2.2–3.0 g/cc and NPHI of 0.15–0.45, representing shale (non-reservoir). Cluster C4 has RHOB of 1.4–2.8 g/cc and NPHI of 0.10–0.50, indicating shaly sand characteristics. Meanwhile, Cluster C3 is concentrated at RHOB 2.5–2.9 g/cc with NPHI very close to 0.00, and is therefore interpreted as clean sand and the most prospective reservoir zone candidate. The most negative correlation, $r = -0.21$, is observed along the regression line of Cluster C3, indicating a weak relationship. The RHOB vs NPHI Crossplot proved effective in distinguishing clean sandstone, shaly sandstone, and shale lithologies through the position of data points relative to the rock-matrix line; zones with moderate-to-high RHOB but very low NPHI generally indicate clean sandstone with reservoir potential, consistent with the interpretation of Cluster C3 in this study (Rezaei Mirghaed et al., 2024; Mabiala Mbouaki et al., 2025).



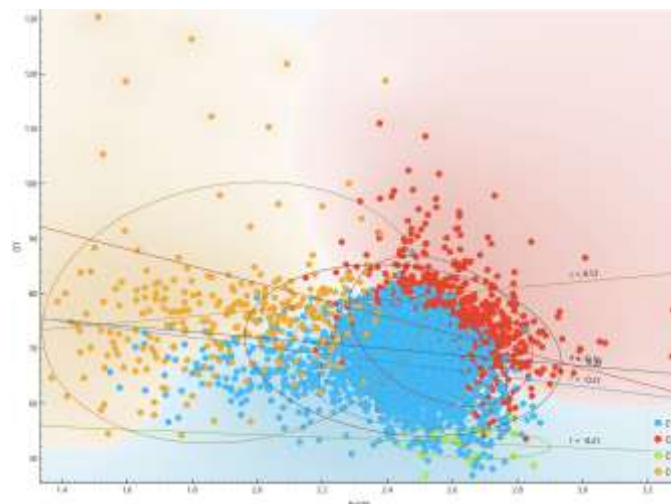
Picture 4 Crossplot RHOB vs NPHI

The GR vs NPHI Crossplot (Figure 5) was used to evaluate the relationship between the natural radioactivity of the formation and its neutron porosity value. Cluster C1 occupies a GR range of 15–80 API and NPHI of 0.00–0.15, indicating clean sand. Cluster C2 occupies a GR range of 80–270 API and NPHI of 0.15–0.35, with a weak positive correlation ($r = 0.26$), indicating shale (non-reservoir) dominance. Cluster C4 spreads across low-to-high GR values with NPHI of 0.15–0.50, representing a transitional shaly sand zone. Meanwhile, Cluster C3 has very low GR and NPHI values with a negative correlation ($r = -0.39$), indicating clean sand with minimal shale content. This characteristic is consistent with the GR vs RDEP Crossplot result showing a high resistivity response, so C3 is interpreted as the most prospective reservoir zone candidate. Overall, the correlation values on this graph range from $r = -0.39$ for Cluster C3 to $r = 0.79$ for the overall (global) trend line across all data points. The strong positive global correlation ($r = 0.79$) reflects increasing clay content accompanied by an increase in apparent neutron porosity due to bound water, while the negative correlation observed in C3 confirms the ability of the clustering method to objectively separate the reservoir zone. This result is consistent with the principle of Density–Neutron–Gamma Ray lithology identification, in which a prospective sandstone zone is characterized by low gamma ray accompanied by low neutron porosity, whereas a shale zone shows a simultaneous increase in both parameters due to the high content of radioactive clay minerals (Ebraheem et al., 2025; Sobih et al., 2025). The observed relationship also confirms that the integration of Crossplot analysis with K-Means Clustering effectively separates lithological groups despite the relatively heterogeneous distribution of well log responses.



Picture 5 Crossplot GR vs NPHI

The RHOB vs DT Crossplot (Figure 6) shows the relationship between rock density and acoustic travel time, used to evaluate the physical characteristics of the rock based on density, compaction level, and the acoustic properties of the formation. Cluster C1 occupies an RHOB range of 2.2–3.0 g/cc and DT of 52–75 $\mu\text{s}/\text{ft}$, indicating moderate-to-high compaction. Cluster C2 occupies an RHOB range of 2.2–3.0 g/cc and DT of 60–108 $\mu\text{s}/\text{ft}$, representing shale with a lower compaction level. Cluster C4 has RHOB of 1.4–2.8 g/cc and DT of 55–127 $\mu\text{s}/\text{ft}$, indicating shaly sand characteristics. Meanwhile, Cluster C3 is concentrated at RHOB 2.5–2.9 g/cc with the lowest DT, 50–56 $\mu\text{s}/\text{ft}$, indicating rock that is relatively more compact than the other clusters. The correlation between RHOB and DT across all clusters ranges from $r = -0.21$ to $r = 0.13$, with the most negative correlation ($r = -0.21$) found in Cluster C3, reflecting a tendency for acoustic travel time to decrease as rock density increases. This result is consistent with the GR vs RDEP and GR vs NPHI interpretations, which show low shale content and high resistivity, so Cluster C3 remains identified as the most prospective reservoir zone candidate. The use of the RHOB vs DT Crossplot in petrophysical analysis has proven able to reflect the physical and elastic characteristics of reservoir rock comprehensively; compact, dense zones such as Cluster C3 are marked by low DT accompanied by relatively high RHOB, consistent with good-quality clean sand (Hossain & Rahman, 2025).

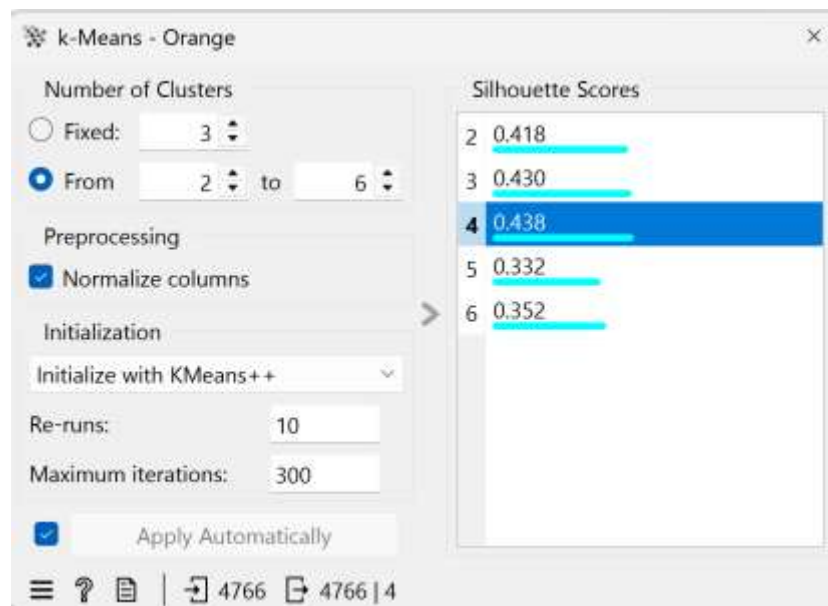


Picture 6 Crossplot RHOB vs DT

Cluster Determination

Determining the optimal number of clusters is a critical step in applying the K-Means Clustering algorithm, since the number of clusters specified directly affects the quality of the grouping and the resulting lithology interpretation. In this study, testing was carried out for k values ranging from 2 to 6 using the k-Means widget in Orange Data Mining. Clustering quality was evaluated using the Silhouette Score available in the k-Means widget, supported by the distribution and pattern of clusters on each Crossplot as an additional basis for interpreting the grouping results.

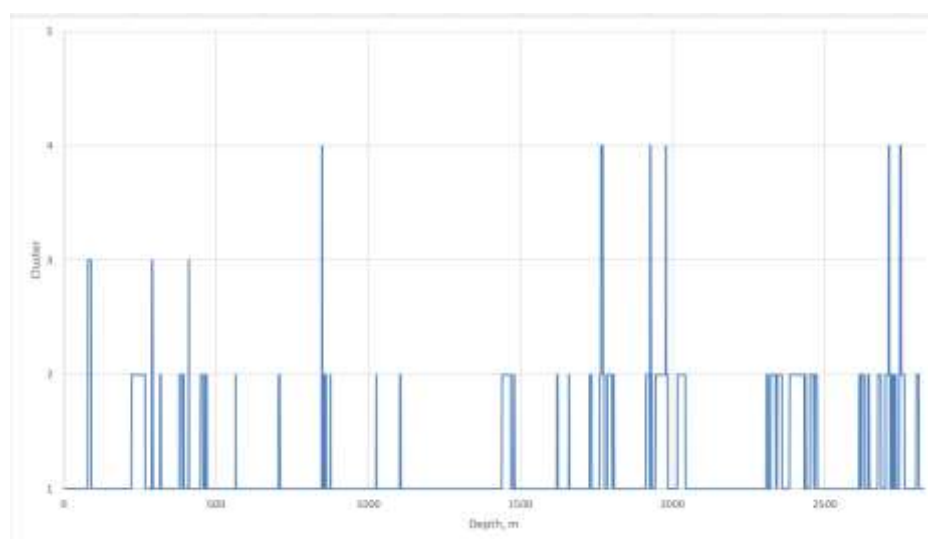
As shown in Figure 7, the Silhouette Score values obtained were 0.418 for $k = 2$, 0.430 for $k = 3$, 0.438 for $k = 4$, 0.332 for $k = 5$, and 0.352 for $k = 6$. The highest value was obtained at $k = 4$ (0.438), indicating that clustering with four clusters produced the most optimal cluster separation. The scores decrease notably at $k = 5$ (0.332) and remain lower than the $k = 4$ peak at $k = 6$ (0.352), even though $k = 6$ shows a slight rebound relative to $k = 5$; overall, this indicates that increasing the number of clusters beyond four does not improve clustering quality (Rousseeuw, 1987). Therefore, $k = 4$ was established as the optimal number of clusters in this study. Although the Silhouette Score of 0.438 indicates only moderate cluster separation, it represents the highest value among all tested clustering configurations. Therefore, the four-cluster solution provides the most appropriate balance between cluster cohesion and separation for this dataset.



Picture 7 Silhouette Scores

Reservoir Zone Determination

To analyze the vertical distribution of the K-Means Clustering results, a depth vs. cluster plot was generated as shown in Figure 8, displaying the distribution of each cluster across the 2098.85–2825.04 m depth interval. Cluster C1 (wet sand) is distributed almost evenly throughout the entire depth interval, with 2,934 data points (61.6% of the total data), making it the cluster with the largest number of data points and simultaneously the dominant lithology in the Kuncherinna-1 well. Cluster C2 (shale) first appears at a depth of 2132.53 m and extends to the base of the interval in 78 discontinuous segments (1,538 data points), reflecting an alternation of shale layers with other lithologies throughout the middle to lower part of the well column. Cluster C3, identified as the most prospective reservoir zone, was detected in only 15 data points distributed across three isolated thin lenses within the 2110.74–2161.18 m depth range, indicating that this prospective reservoir zone is thin and discontinuous (thin bed) rather than a single massive thick reservoir layer. Meanwhile, Cluster C4 (shaly sand) is distributed across 30 discontinuous segments (279 data points) from intermediate depths toward the base of the well, with the thickest concentrations at depths of 2771.55–2781.30 m (9.75 m thick) and 2795.63–2800.96 m (5.33 m thick). These results are consistent with the petrophysical characteristics of C3 obtained from the Crossplot analysis, namely very low GR values, very high RDEP values, and low NPHI and DT values.



Picture 8 Depth vs Cluster Plot

Table 1 Crossplot Analysis Results and Lithology Interpretation

Cluster	Crossplot Analysis Result	Interpretation
C1	Low-to-moderate GR, low NPHI, low-to-moderate DT, variable RDEP	Wet sand
C2	High GR, high NPHI, high DT, low RDEP	Shale zone / non-reservoir
C3	Very low GR, very high RDEP, very low NPHI, low DT, relatively high RHOB	Most prospective reservoir zone candidate
C4	Transitional characteristics between sandstone and shale	Transition zone / shaly sand

Based on Table 1, Cluster C3 is established as the most prospective reservoir zone candidate (prospect pay zone) in the Kuncherinna-1 well. Cluster C3 is interpreted as the most prospective reservoir zone candidate because it is characterized by a very low Gamma Ray value (15–35 API), indicating clean sandstone, and a very high resistivity value exceeding 30,000 ohm-m, reaching a maximum of 32,813.7 ohm-m. This combination of characteristics indicates low shale content and a better potential presence of hydrocarbons compared with the other clusters. In contrast, Cluster C2 is identified as a shale zone (non-reservoir) because it has relatively high Gamma Ray and Neutron Porosity values together with very low resistivity. Cluster C1 represents clean sand dominated by formation water (wet sand), while Cluster C4 shows transitional characteristics in the form of shaly sand with a lower prospectivity level. These results indicate that the integration of the Crossplot method and the K-Means Clustering algorithm using Orange Data Mining is able to group formation characteristics systematically and consistently, helping to distinguish reservoir from non-reservoir zones and supporting the achievement of the research objective of identifying lithology and determining candidate reservoir zones in the Kuncherinna-1 well. This study is limited to well log interpretation using GR, RDEP, RHOB, NPHI, and DT logs without validation from core analysis, mud log, drill stem test (DST), or advanced petrophysical calculations such as water saturation and permeability. Consequently, the interpreted reservoir interval should be regarded as a prospective reservoir candidate rather than a confirmed hydrocarbon reservoir. Future studies integrating additional geological and petrophysical datasets are recommended to improve interpretation reliability.

CONCLUSION

The lithology of the Kuncherinna-1 well within the 2098–2825 m depth interval was successfully identified using the Crossplot and K-Means Clustering methods, revealing three dominant lithology types: clean sand, shale, and shaly sand, which could be consistently distinguished through four parameter combinations based on variations in Gamma Ray, resistivity, density, neutron porosity, and acoustic travel time. Application of K-Means Clustering with an optimal number of clusters of $k = 4$ (Silhouette Score = 0.438) objectively grouped the data into four lithological clusters: Cluster C1 as wet sand, Cluster C2 as a shale (non-reservoir) zone, Cluster C4 as a transitional shaly sand zone, and Cluster C3 as the most prospective reservoir zone candidate, characterized by a very low Gamma Ray value (15–35 API), a very high Deep Resistivity value reaching 32,813.7 ohm-m, a very low Neutron Porosity value, and a low Sonic Log value. The vertical distribution analysis further showed that Cluster C3 was concentrated only within the 2110.74–2161.18 m depth interval, located at the shallowest part of the analyzed section, indicating a thin reservoir layer rather than a continuous zone throughout the well. These findings indicate that the integration of Crossplot analysis and K-Means Clustering provides an effective and objective approach for lithology interpretation and identification of prospective reservoir intervals. However, further validation using core data, formation testing, or additional petrophysical analyses is recommended before confirming reservoir presence.

REFERENCES

- Ali, M., Jiang, R., Ma, H., Pan, H., Abbas, K., Ashraf, U., & Ullah, J. (2024). Data-driven lithofacies prediction in complex tight sandstone reservoirs: A supervised workflow integrating clustering and classification models. *Geomechanics and Geophysics for Geo-Energy and Geo-Resources*, 10(1), Article 5. <https://doi.org/10.1007/s40948-024-00787-5>

- Asquith, G. B., & Krygowski, D. (2004). *Basic well log analysis* (2nd ed.). American Association of Petroleum Geologists.
- Cao, M., Gao, Z., Yuan, Y., Yan, Z., & Zhang, Y. (2022). A visualization and analysis method by multi-dimensional crossplots from multi-well heterogeneous data. *Energies*, 15(7), Article 2575. <https://doi.org/10.3390/en15072575>
- Demšar, J., Curk, T., Erjavec, A., Gorup, č., Hočevár, T., Milutinovič, M., Možina, M., Polajnar, M., Toplak, M., Starič, A., Štajdohar, M., Umek, L., Žagar, L., Žbontar, J., Žitnik, M., & Zupan, B. (2013). Orange: Data mining toolbox in Python. *Journal of Machine Learning Research*, 14, 2349–2353.
- Ebraheem, M. O., Ibrahim, H. A., & Senosy, A. H. (2025). A comparative petrophysical evaluation of the Abu Roash, Bahariya, and Kharita reservoirs using well-logging data, East El-Fayoum, Egypt. *Scientific Reports*, 15, Article 2891. <https://doi.org/10.1038/s41598-024-83332-4>
- Geoscience Australia. (n.d.). *The Pedirka and western Eromanga basins well log petrophysics data package* (ECAT 149208) [Data set]. Australia's Future Energy Resources (AFER), Exploring for the Future Program.
- Hasan, M. M. U., Hasan, T., Shahidi, R., James, L., Peters, D., & Gosine, R. (2023). Lithofacies identification from wire-line logs using an unsupervised data clustering algorithm. *Energies*, 16(24), Article 8116. <https://doi.org/10.3390/en16248116>
- Hossain, S., & Rahman, N. (2025). An integrated petrophysical and rock physics characterization of the Mangahewa Formation in the Pohokura field, Taranaki Basin. *Scientific Reports*, 15, Article 6431. <https://doi.org/10.1038/s41598-025-88639-4>
- Mabiala Mbouaki, A. P., Cai, Z., Sikanyika, E. W., Mwakipunda, G. C., & Konan, H. R. (2025). Petrophysical evaluation of a shaly sandstone reservoir and the effect of clay minerals on reservoir quality: A case study from the Barremian Mengo Sandstone, Kouilou Basin, Republic of Congo. *ACS Omega*, 10(10), 10081–10106. <https://doi.org/10.1021/acsomega.4c08491>
- Octavia, D. J., Nugrahanti, A., Wiharjo, L., Rahmawan, S., & Kadarusman, S. (2024). Evaluasi litologi dari hasil overlay log densitas dan log neutron pada Sumur Dj-1 dan Dj-2. *PETRO: Jurnal Ilmiah Teknik Perminyakan*, 13(1), 1–6. <https://doi.org/10.25105/petro.v13i1.17386>
- Rezaei Mirghaed, B., Dehghan Monfared, A., & Ranjbar, A. (2024). Enhanced petrophysical evaluation through machine learning and well logging data in an Iranian oil field. *Scientific Reports*, 14, Article 28319. <https://doi.org/10.1038/s41598-024-80362-w>
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Sobih, N. S., El Nagar, A. A., Abbas, A. E., & Shehab, S. (2025). Integrating petrophysical characterization and reservoir evaluation of the Kafr El-Sheikh Formation in Sapphire field, Nile Delta, Egypt. *Scientific Reports*, 15, Article 5781. <https://doi.org/10.1038/s41598-025-88056-7>
- Szabó, N. P., Kilik, R., & Dobróka, M. (2023). Robust reservoir identification by multi-well cluster analysis of wireline logging data. *Heliyon*, 9(5), Article e15957. <https://doi.org/10.1016/j.heliyon.2023.e15957>
- Suyal, M., & Sharma, S. (2024). A review on analysis of K-means clustering machine learning algorithm based on unsupervised learning. *Journal of Artificial Intelligence Systems*, 6(1), 85–95. <https://doi.org/10.33969/ais.2024060106>
- Tiab, D., & Donaldson, E. C. (2015). *Petrophysics: Theory and practice of measuring reservoir rock and fluid transport properties* (4th ed.). Gulf Professional Publishing.