ANALYSIS OF AN UNSUPERVISED CLUSTERING ALGORITHM FOR LITHOFACIES CLASSIFICATION IN 'RIGA' FIELD, NIGER DELTA

Olisa Benson Akinbode and Oluniyi Taiwo Samuel
The Federal University of Technology, Akure, Nigeria.

DOI: 10.31364/SCIRJ/v13.i04.2025.P04251018
http://dx.doi.org/10.31364/SCIRJ/v13.i04.2025.P04251018

Abstract

The study determined the optimum number of clusters that effectively capture lithological variations in subsurface formations and also emphasizes the application of these clustering algorithms for lithofacies classification. The k-means clustering algorithm was used to uncover hidden patterns in well logs in the Niger Delta's "RIGA" field. This model successfully classified the data into 4 distinct clusters, revealing correlations with depth and gamma ray (GR) measurements. The developed clustering model was able to automatically classify the dataset into useful clusters, these clusters, when matched with depth, generated useful lithologies in the well RIGA-1 and RIGA-2. There are 4 clusters having 3 very visible clusters similar to Continental sands, Marginal marine sandstones and Shale. These mostly tie up with the changes in the logging measurements, decrease in Gamma Ray (GR) from around 4900m to 6100m, 7700 to 7600, 8500 - 9700 aligns with the blue clusters in Well 1 and Well 2.

Keywords

unsupervised clustering algorithm, lithofacies classification.

1. Introduction

Understanding the subsurface lithology is crucial in geoscience and petrophysics for various applications such as reservoir characterization and hydrocarbon exploration. Well logging technology provides electrical measurements that can offer valuable insights into lithology, facies, porosity, and permeability. Machine learning algorithms have been widely adopted to classify well log data into distinct lithological groupings, known as facies. Clustering is a form of exploratory data analysis used to group data points based on shared characteristics. K-Means Clustering, a commonly used unsupervised algorithm, partitions data into K clusters by minimizing the distance between data points and their centroids.

2. Literature review

2.1 The Niger Delta

http://dx.doi.org/10.31364/SCIRJ/v13.i04.2025.P04251018

The Niger Delta is one of the world's most productive deltaic hydrocarbon provinces, and it is the most important in the West African continental margin.

Three important elements are required for hydrocarbon to be produced in an area, these are, the source rocks, the reservoir rocks and the cap rocks. If any of these is lacking, it will result in a general failure (Olisa 2016).

Sandstones and unconsolidated sands, mostly from the Agbada Formation, is the reservoir in the Niger Delta. Reservoir rocks range in age from the Eocene to the Pliocene (Onuoha and Chukwu, 2020), and are frequently layered, with thicknesses varying from less than 15 metres to more than 45 metres for 10% of them. Growth faults inside the down-thrown block govern the lateral variation in reservoir thickness based on reservoir geometry and quality, with reservoirs thickening towards the fault. A reservoir is a subsurface rock with sufficient porosity and permeability to hold commercially viable quantities of hydrocarbon. In addition, if hydrocarbons are to be produced, the pore sand fractures must be interconnected.

2.2 Machine learning algorithms

In geoscience and petrophysics, understanding the subsurface lithology is a crucial task that provides insights into the geological characteristics of a region. Well logging technology, which utilizes various electrical measurements, plays a vital role in inferring important lithological properties such as lithology type, facies, porosity, and permeability. Machine learning algorithms have become increasingly popular in the field of geoscience for their ability to analyze large volumes of well log data and extract meaningful information. In particular, unsupervised learning techniques have proven to be effective in grouping well log measurements into distinct lithological groupings known as facies. These methods enable us to identify underlying patterns within the data, even when they may not be readily apparent during data exploration. In this research, we will focus on one out of three of widely used unsupervised learning clustering methods: K Means Clustering. This method offers different approaches to identifying lithological facies based on the well log measurements. By comparing the clustering results with an established Lithofacies curve, we can assess the accuracy and reliability of these unsupervised learning techniques in capturing lithological variations. The science of learning from data is a key focus of machine learning. Machine learning combines the fields of statistics and computer science for pattern recognition and data mining applications (Michie *et al*. 1994; Hastie *et al*. 2009). For science based research, pattern recognition is the process of discovering, via automated or semi-automated statistical methods, useful patterns within data (Kotsiantis 2007). Discovered patterns are then used to generate predictions based on similar data. The essence of machine-assisted pattern recognition is to provide computers with the ability to adapt their decision structures, based on the characteristics of observed data and generate valid and objective predictions. Machine learning is an extension of the pattern recognition process. It attempts to provide users with an understanding of the patterns within data (Feyyad 1996). Hence, machine learning outputs should be comprehensible in a way that allows interpretations to be formulated in response to the decision structures used to recognise and exploit patterns within data and generate predictions (Feng and Michie 1994; Henery 1994). Data inference is the act of gaining

http://dx.doi.org/10.31364/SCIRJ/v13.i04.2025.P04251018

information, knowledge and ultimately wisdom, from the analysis of raw data using statistical methods. The process of data inference can be divided into three levels of understanding. The foundation for these levels of understanding is raw data. Successive levels of data inference distil and refine raw data until a complete understanding of the mechanisms controlling the phenomena under investigation is realized. The conclusions attained via the process of data inference are subsequently applied to other similar data in order to make predictions, formulate interpretations and inform the decision making process (Bousquet *et al.* 2004).

3. Previous Work

Given the nature of this research, this section begins with an overview of several significant geological, stratigraphic and structural characteristics as well as the hydrocarbon potential of reservoir rocks studies conducted on the Niger Delta and associated formations from surface exposures and through subsurface studies. These previous works highlight the application of unsupervised learning and machine learning algorithms, such as K Means Clustering in lithology classification using well log data. They demonstrate the potential of these techniques in improving lithological characterization and understanding subsurface geological variations.

Hou, *et al.* (2023) worked on Machine Learning Algorithms for Lithofacies Classification of the Gulong Shale from the Songliao Basin. This research focuses on applying machine learning models to identify clay-rich shale lithofacies using conventional well log data. The study highlights the effectiveness of ensemble machine learning in identifying clay-rich shale lithofacies, aiding in unconventional reservoir sweet spot prediction. The research gap lies in the limited exploration of machine learning models for clay-rich shale lithofacies prediction, emphasizing the need for more in-depth investigations.

Liu *et al*. (2020) worked on Lithofacies identification using support vector machine based on local deep multi-kernel learning. This research presents a novel approach called Local Deep Multi-Kernel Learning Support Vector Machine (LDMKL-SVM) to enhance the classification of lithofacies using seismic data. LDMKL-SVM efficiently learns kernel function parameters and builds a relationship between lithofacies and seismic elastic information, improving both computational speed and accuracy in multi-class lithofacies identification for reservoir characterization and prediction. The research gap lies in the limited exploration of machine learning models for clay-rich shale lithofacies prediction, emphasizing the need for more in-depth investigations in this specific area.

4. Problem Statement

In the field of geoscience and petrophysics, accurately identifying subsurface lithology and lithofacies is crucial for successful oil and gas exploration and production. Traditionally, this has been accomplished through manual interpretation of well log data, which can be time-consuming and subject to human bias. With the advent of machine learning algorithms, clustering method such as K-Means have been adopted for unsupervised classification of well log data into distinct lithofacies. However, the effectiveness and accuracy of these clustering methods in identifying lithofacies and lithology compared to established lithofacies curves have not

http://dx.doi.org/10.31364/SCIRJ/v13.i04.2025.P04251018

been thoroughly evaluated. Therefore, there is a need to investigate and compare the performance of these clustering methods in identifying lithofacies and lithology to determine the most effective approach for subsurface characterization.

5. Aim and Objectives

5.1 Aim

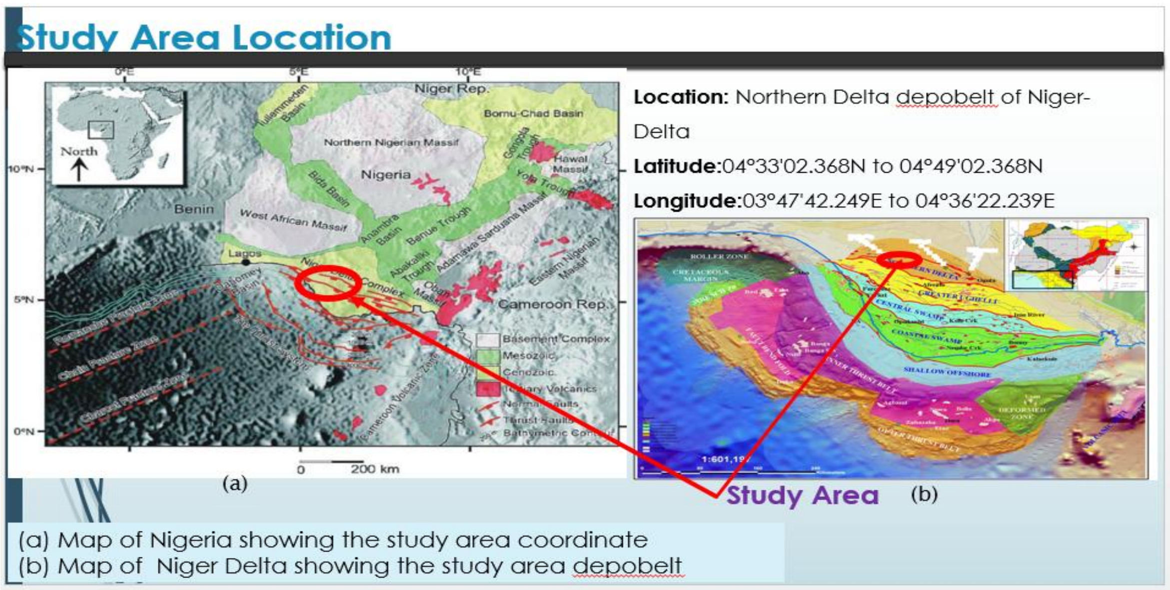
This research aims to evaluate the effectiveness of unsupervised learning techniques in lithofacies classification using well log measurements by comparing their results to an established lithofacies curve, thus demonstrating their capability to identify distinct lithofacies based on underlying data patterns.

5.2 Objectives

- Determine the optimal number of clusters for lithology classification using the K-Means clustering algorithm applied to well log data.
- Extract relevant geophysical attributes or features indicative of lithological variations for unsupervised cluster analysis.
- Develop visualizations, such as cluster scatter plots, to aid in the interpretation and visualization of lithological facies patterns identified by the K-Means algorithm
- Evaluate the scalability, efficiency, interpretability, and clustering performance of the K-Means algorithm for identifying and classifying lithological facies from large volumes of well log data.

6. Location of study area

The study area is located within offshore western Niger Delta Basin in Nigeria. It is limited within latitudes 4.00N – 5.00N, and longitudes 5.00 E – 7.00 E (Figure 1).



(a) Map of Nigeria showing the study area coordinate
(b) Map of Niger Delta showing the study area depobelt

Figure 1: (a) Map of Nigeria showing the study area coordinate (b) Map of Niger Delta showing the study area depobelt.

7. Materials and Methodology

7.1. Materials

The materials deployed for this research are mainly Petrel Software, suite of software (Python and some of its Libraries) and well logs.

7.1.1. Petrel Software

A software platform developed by schlumberger for the exploration and production (E&P) of oil and gas. It is widely used in the oil and gas industry for tasks such as seismic interpretation, well data analysis and reservoir modeling.

7.1.2. Python

Python (Version 3.7) was the programming language used for this study primarily because of its numerous libraries for data loading, data analysis, visualization, statistics, machine learning, and more. It is well suited for this research project because machine learning is fundamentally an iterative process, in which the data drives the analysis. It is essential for the process to have tools that allow quick iteration and easy iteration. Python also enables the programmer to interact directly with the code, using a terminal or other tools like the Jupyter Notebook as used in this study.

7.1.3. DataSet

The datasets used for training the reservoir delineation model in the study was the well log. The data files contain subsurface data collected in a well drilled in a large and well-studied oil field in the Niger Delta basin. The data files are in csv format.

7.2. Methodology

The specific procedures or techniques used to identify, select, process, analyze the well logs and k-means clustering integration is as given in Figure 2. The figure also shows how the different techniques were used to achieve the objectives of the study.

http://dx.doi.org/10.31364/SCIRJ/v13.i04.2025.P04251018

ISSN 2201-2796

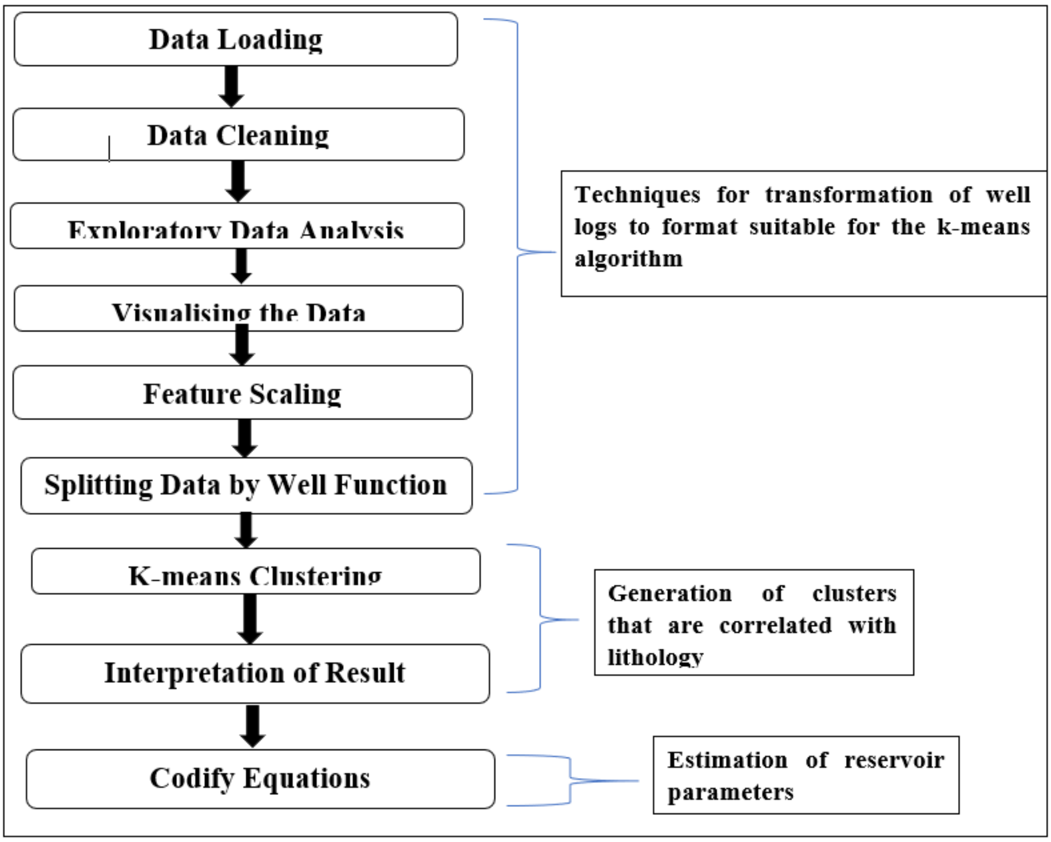


Figure 2: The Methodology of the Study with the Research Objectives

8. Results and discussion

The result of the conversion of las files into dataframe(a form useful for processing) is displayed in Table 1. The logs present are: the gamma ray, resistivity, density, neutron, and sonic logs. The NaN on the dataframe is equivalent to the numerical value, -999.25 on the original well logs. They mean that logging was not run at those depth points.

http://dx.doi.org/10.31364/SCIRJ/v13.i04.2025.P04251018

ISSN 2201-2796

Table 1: Well Logs data showing the null values

Overview

Dataset Statistics

Number of Variables	13
Number of Rows	12861
Missing Cells	585
Missing Cells (%)	0.3%
Duplicate Rows	0
Duplicate Rows (%)	0.0%
Total Size in Memory	1.3 MB
Average Row Size in Memory	104.0 B
Variable Types	Numerical: 12 Categorical: 1

Dataset Insights

DEPT is uniformly distributed	Uniform
TEMP is uniformly distributed	Uniform
TVD is uniformly distributed	Uniform
DEPT and TVD have similar distributions	Similar Distribution
ILD and RT have similar distributions	Similar Distribution
DT has 176 (1.37%) missing values	Missing
ILD is skewed	Skewed
RT is skewed	Skewed
PRES is skewed	Skewed
VSHL is skewed	Skewed

1 2

Table 2 is the summary statistics of the well log which include maximum values, minimum values, interquartile ranges, mean and standard deviation, the table shows that there are outliers in the depth samples (based on logs scale ranges displayed).

Table 2: The Summary Statistics of the raw Well Dataset

```
df.head()
```

	DEPT	GR	DT	ILD	RHOB	NPHI	BS	SP	RT	TEMP	PRES	TVD	VSHL	Facies
17	3500.5	63.456902	124.965805	15.714300	2.0847	0.4472	14.75	-51.399498	15.714300	112.240402	1.6894	3497.291748	0.1501	4
18	3501.0	54.486599	122.816399	19.649401	2.0521	0.4349	14.75	-51.963898	19.649401	112.245003	1.6896	3497.790527	0.0981	3
19	3501.5	50.296398	118.194702	23.306999	2.0043	0.4282	14.75	-52.504101	23.306999	112.249603	1.6899	3498.289307	0.0762	3
20	3502.0	49.454498	115.471107	26.847200	1.9626	0.4298	14.75	-53.059299	26.847200	112.254204	1.6901	3498.788330	0.0724	3
21	3502.5	49.736198	113.503304	29.794500	1.9165	0.4484	14.75	-53.641800	29.794500	112.258804	1.6904	3499.287109	0.0747	3

Table 3 shows the distribution of Well logs data which are the Gamma ray, resistivity, density, neutron, Sonic log

http://dx.doi.org/10.31364/SCIRJ/v13.i04.2025.P04251018

Table 3 Range of vales of Well Logs

S/N	Well Log	Scale(unit)
1	Gamma Ray	0 – 150(API)
2	Resistivity	0 – 2000(ohm.m)
3	Density	1.65-2.65(gg/cc)
4	Neutron	1.95-2.95
5	Sonic	40-140(us/ft)

8.1. Distribution of Training data in Wells using Gamma ray

For the distribution of lithofacies in our training data, there is a generalised Niger Delta stratigraphy and lithofacies subdivision (Ejedawe, 2007). Even though this is the general model, the model used for this project work was adopted from the well log response character for different genetic units (Electrofacies) (SCiN, 1997; Ogbe *et al*. (2020)). This was used to also verify the correctness/validate of the generalized lithofacies subdivision.

This code is used to create a bar chart that shows the distribution of different types of "facies" in a dataset. In essence, this code helps you see how common or rare different types of "facies" are in your dataset by using a bar chart as shown in Figure 3 below.

http://dx.doi.org/10.31364/SCIRJ/v13.i04.2025.P04251018

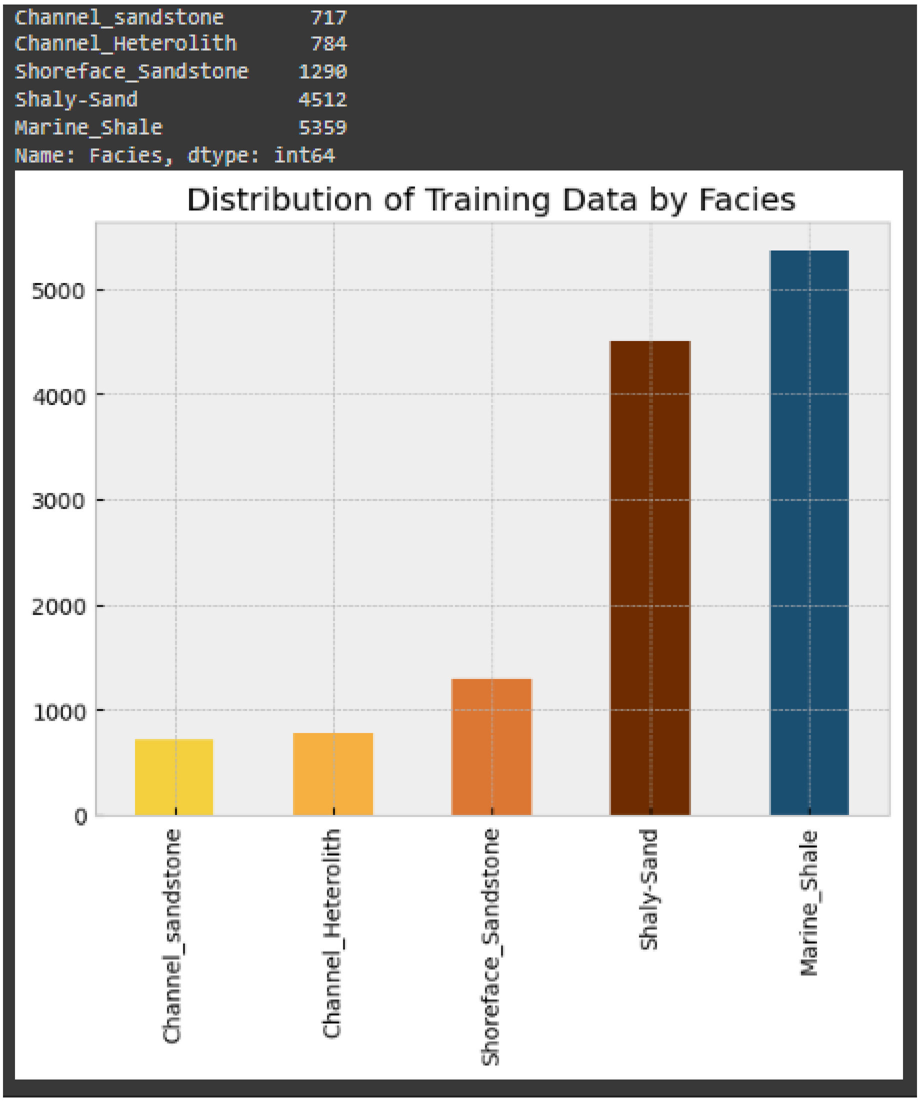


Figure 3: Distribution of Training data

8.1.1. Gamma Ray Measurements by Lithology

In the gamma ray distribution by lithology, we have a dataset of gamma ray measurements obtained from different rock formations. Since the facies were mapped manually they were divided into 5 different facies which include: Channel Sandstone, Channel Heterolith, Shoreface Sandstone, Shaly Sand and Marine Shale.

With gamma ray values ranging between 0 - 150, the Channel Sandstone corresponds to areas ranging from (20 – 60), Channel Heterolith corresponds to areas ranging from (50 – 90), Shoreface Sandstone corresponds to areas ranging from (30 – 130), Shaly Sand corresponds to areas ranging from (63 – 101) and Marine Shale corresponds to areas ranging from (100 -150). It is being shown in Figure 4, Figure 5 and Figure 6 below on Petrel software and on python software.

http://dx.doi.org/10.31364/SCIRJ/v13.i04.2025.P04251018

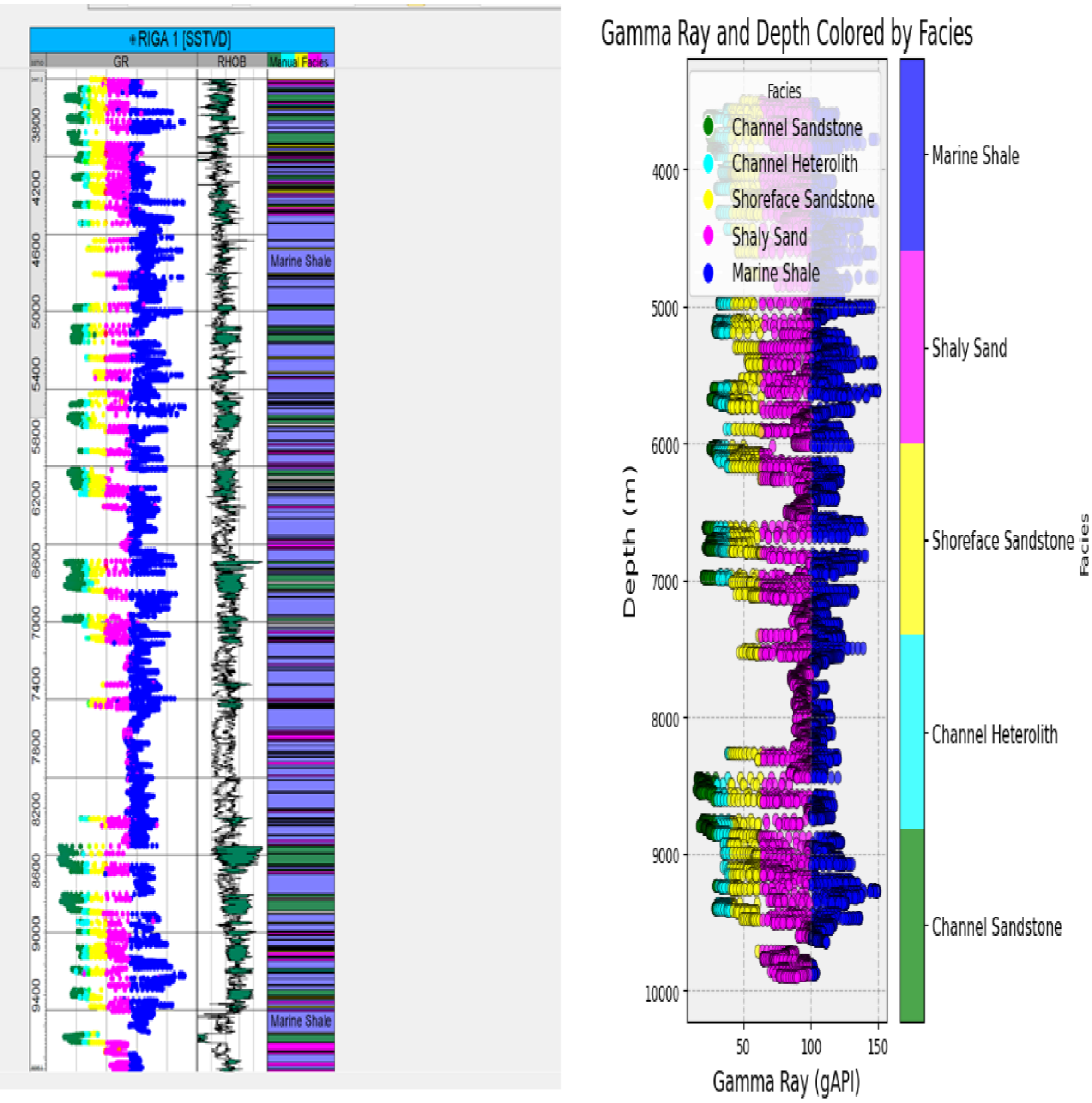


Figure 4: Gamma ray track showing lithofacies of RIGA-1 on Petrel software and on Python

Software

http://dx.doi.org/10.31364/SCIRJ/v13.i04.2025.P04251018

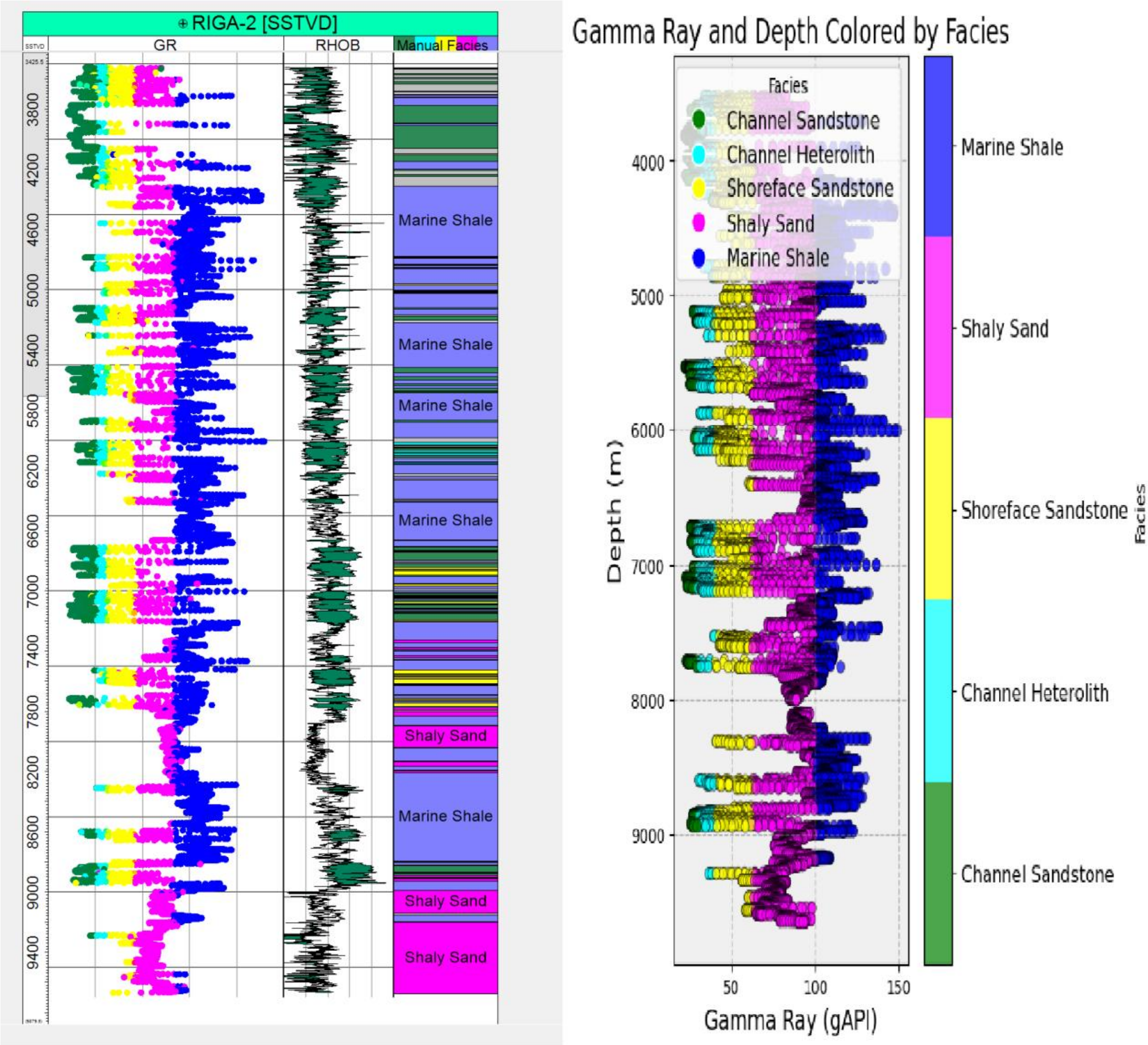


Figure 5: Gamma ray track showing lithofacies of RIGA-2 on Petrel software and on Python software

http://dx.doi.org/10.31364/SCIRJ/v13.i04.2025.P04251018

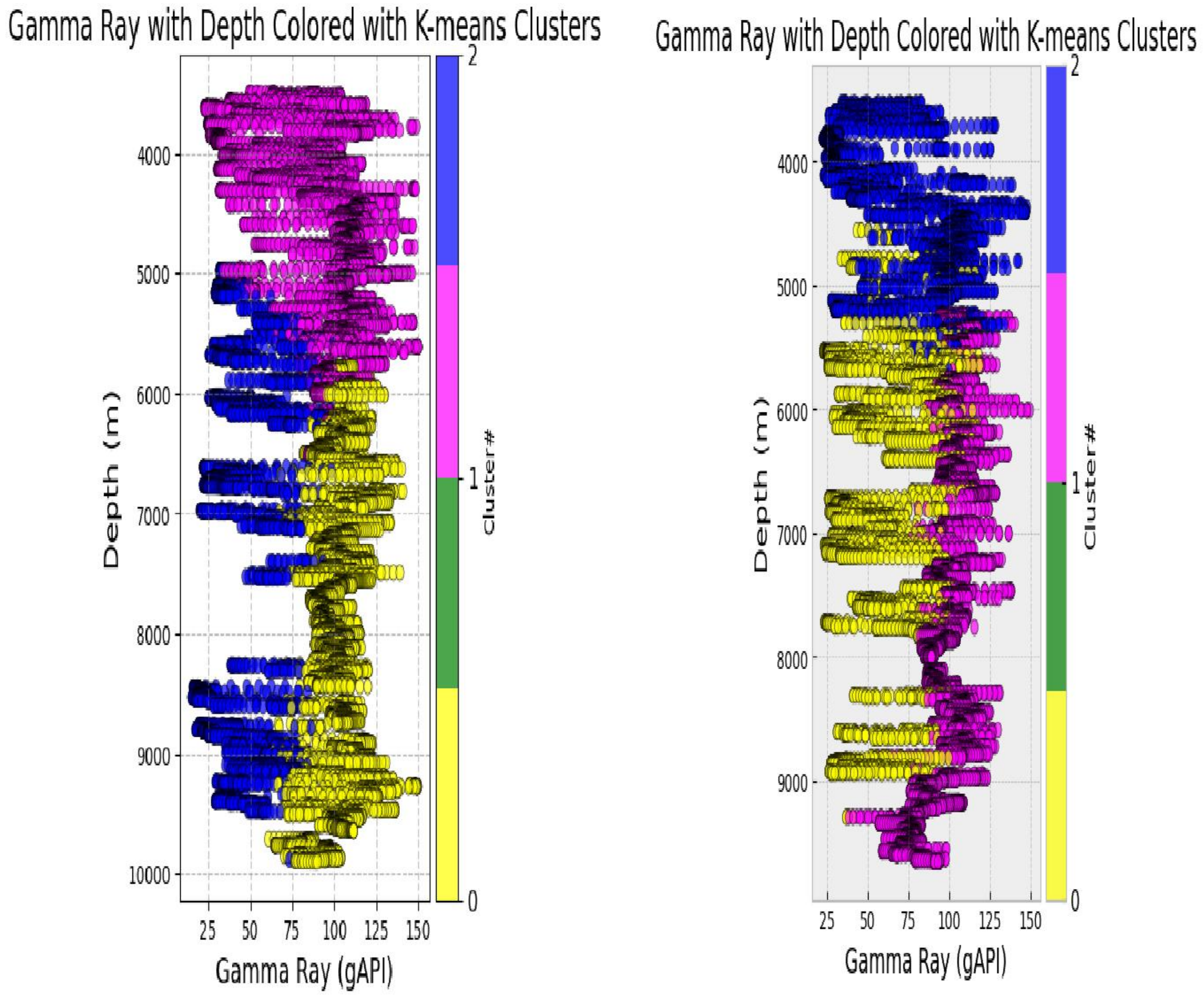


Figure 6: Gamma ray track of RIGA-1 & RIGA-2

9. Conclusion

The k-means clustering algorithm was integrated to well logs to build a clustering model and the algorithms to identify underlying patterns within the data that may not be easily visible during data exploration. The k-means clustering algorithm has been successfully deployed to identify underlying patterns within the data that may not be easily visible in the well logs in field “RIGA” in the Niger Delta. The developed clustering model was able to automatically classify the dataset into useful clusters, these clusters, when matched with depth, generated useful lithologies in the well RIGA-1, RIGA-2. We have 5 separate facies/groups initially displayed but then eventually got 4 cluster but having 3 very visible clusters similar to Continental sands, Marginal marine sandstones and Shale and we can see that these mostly tie up with the changes in the logging measurements, decrease in Gamma Ray (GR) from around 4900m to 6100m, 7700 to 7600, 8500 - 9700 aligns with the blue clusters in Well 1, Well 2 decrease in Gamma Ray (GR) from around 5200m to around 9000m ties in nicely with the yellow grouping,

http://dx.doi.org/10.31364/SCIRJ/v13.i04.2025.P04251018

ISSN 2201-2796

References

Bousquet, F. and Le Page, C., 2004. Multi-agent simulations and ecosystem management: a review. Ecological modelling, *176*(3-4), pp.313-332.a

Ejedawe (2007) An unpublished report on the Niger Delta submitted to SPDC Warri.

Feng, C. and Michie, D., 1994. Machine learning of rules and trees. Machine learning, neural and statistical classification, pp.50-83.

Feyyad, U.M., 1996. Data mining and knowledge discovery: Making sense out of data. IEEE expert, 11(5), pp.20-25.

Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Tibshirani, R. and Friedman, J., 2009. Overview of supervised learning. The elements of statistical learning: Data mining, inference, and prediction, pp.9-41.

Hou, Mingqiu & Xiao, Yuxiang & Lei, Zhengdong & Yang, Zhi & Lou, Yihuai & Liu, Yuming. (2023). Machine Learning Algorithms for Lithofacies Classification of the Gulong Shale from the Songliao Basin, China. Energies. 16. 2581. 10.3390/en16062581.

Kotsiantis, S.B., Zaharakis, I. and Pintelas, P., 2007. Supervised machine learning: A review of classification techniques. Emerging artificial intelligence applications in computer engineering, 160(1), pp.3-24.

Liu, XY., Zhou, L., Chen, XH. (2020). Lithofacies identification using support vector machine based on local deep multi-kernel learning. Pet. Sci. 17, 954–966. https://doi.org/10.1007/s12182-020-00474-6

Michie, D., Spiegelhalter, D.J. and Taylor, C.C., 1994. Machine learning, neural and statistical classification.

Olisa, B.A., (2016). Estimating Source-Rock Organic Richness from Well Data in Parts of Western Niger Delta, Nigeria (Doctoral dissertation, Federal University of Technology Akure).

Onuoha, K. M., & Chukwu, L. O. (2020). Petroleum Geology and Hydrocarbon Potential of Niger Delta Basin, Nigeria. In M. P. Anbuvelan, A. Alagarsamy, V. Mohana Doss, &N.

http://dx.doi.org/10.31364/SCIRJ/v13.i04.2025.P04251018