International Journal of Science and Research Archive
eISSN: 2582-8185
Cross Ref DOI: 10.30574/ijsra
Journal homepage: https://ijsra.net/

(RESEARCH ARTICLE)

Facie classification of X-field in the Niger Delta using machine learning

EVWERHAMRE GODSPOWER CALVIN * and JULIET EMUDIANUGHE

Department of Earth Sciences, Federal University of Petroleum Resources, Effurun, 330102, Delta, Nigeria.

International Journal of Science and Research Archive, 2024, 13(02), 3556-3568

Publication history: Received on 16 November 2024; revised on 22 December 2024; accepted on 24 December 2024

Article DOI: https://doi.org/10.30574/ijsra.2024.13.2.2516

Abstract

The Niger Delta's complex geology presents challenges for accurate facies classification and reservoir characterization. This study applies machine learning techniques to well log data from four wells in the X-Field: Freeman-001, Freeman-003ST1, Freeman-004ST1, and Freeman-005. Key features analyzed include Gamma Ray (GR), Sonic Transient Time (DT), Bulk Density (RHOB), and Neutron Porosity (NPHI). Unsupervised learning methods (K-means, GMM, and Agglomerative Clustering) yielded moderate clustering performance, with K-means achieving the highest accuracy of 47.87%, followed by GMM (20.38%) and Agglomerative Clustering (13.94%). Supervised learning methods significantly outperformed unsupervised ones. Random Forest achieved 87.06% accuracy, 87.07% precision, and an F1-score of 87.06%. Gradient Boosting delivered the best results, with 87.53% accuracy, 87.55% precision, and an F1-score of 87.54%. SVM showed lower performance, with 77.86% accuracy and an F1-score of 77.82%. These findings highlight the superior performance of supervised learning, particularly tree-based models, in facies classification and underscore the potential of machine learning to enhance reservoir characterization in the Niger Delta.

Keywords: Supervised Learning; Unsupervised learning; Facie Classification; Niger Delta

1. Introduction

The Niger Delta region of West Africa is renowned for its vast hydrocarbon reserves, making accurate subsurface characterization crucial for optimal exploration and production strategies (1). Traditional methods of facies classification and reservoir characterization have often relied on manual interpretation of well logs and seismic data, leading to subjectivity and inefficiencies. In recent years, advancements in machine learning have introduced new possibilities for enhancing subsurface analysis (2).

This study investigates an innovative approach for facies classification in the Niger Delta, combining both unsupervised and supervised machine learning techniques. The proposed methodology employs unsupervised algorithms such as K-means, Gaussian Mixture Models (GMM), and Hierarchical Clustering to uncover hidden geological structures and enhance facies classification accuracy. Additionally, supervised learning methods including Support Vector Machines (SVM), Random Forest, and Gradient Boosting are utilized to refine facies predictions based on labeled data. The objectives of this study encompass the identification of distinct facies clusters and the extraction of pertinent reservoir properties. The expected results include improved reservoir management, optimized exploration strategies, and cost reduction. By harnessing both unsupervised and supervised machine learning, this research propels advancements in subsurface analysis, underpinning sustainable energy production in the Niger Delta.

* Corresponding author: EVWERHAMRE GODSPOWER CALVIN

Objectives and Scope

This thesis aims to develop and evaluate machine learning approaches for automated facie classification in well log data, comparing the effectiveness of supervised and unsupervised learning techniques in characterizing geological formations in X-Field, Niger Delta, using well log data. Here are the specific objectives:

- Gather and prepare well log data from multiple wells in X-Field, including data cleaning, normalization, and feature selection.
- Implement and optimize various unsupervised learning algorithms, such as K-Means clustering, and Gaussian Mixture Models (GMM), Agglomerative Clustering for facies classification.
- Train and optimize supervised machine learning models, including Support Vector Machines (SVM), Random Forest (RF), and Gradient Boosting (GB), for facies classification using labeled well data.
- Evaluate the performance of the developed models using appropriate metrics and visualization techniques, comparing their effectiveness in identifying distinct facies.
- Analyze the identified facies clusters and relate them to geological characteristics, generating visualizations such as faces logs and cross-plots.
- Validate the selected model using blind well data and assess its applicability for facie classification and decision-making in hydrocarbon exploration.

Limitations

The limitations of this project include:

- The accuracy and effectiveness of the machine learning model heavily rely on the availability of high-quality, representative well log. Incomplete or noisy data may impact the model's performance.
- Machine learning methods, while powerful, can be computationally intensive, especially when dealing with large datasets or complex algorithms. Real-time processing might require significant computational resources.
- The model's ability to generalize to new, unseen data from different wells or fields in the Niger Delta might be limited. Geological variations across the region could affect the model's transferability.
- Unsupervised learning models can be less interpretable than supervised models, making it challenging to understand the geological reasoning behind the model's facies classifications.
- Choosing the optimal combination of unsupervised algorithms and their hyperparameters for the model requires careful consideration and experimentation.

2. Geologic Setting of the study Area

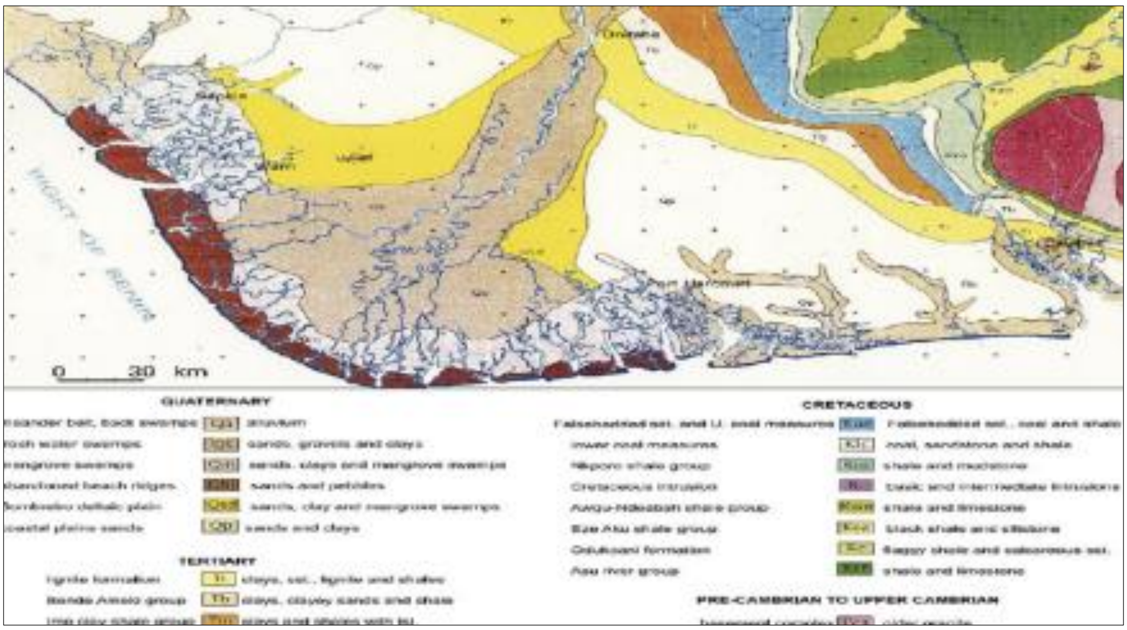

Figure 1 Map of the Niger Delta, including province boundaries and structural elements (after Petroconsultants, 1996)

The Niger Delta Basin (Figure 1) is situated in the Gulf of Guinea in equatorial West Africa, between latitudes 3° N and 6° N and longitudes 5° E and 8° E (3). It is a clastic fill of about 12,000m with a sub-aerial portion covering 75,000 sq. km. and extending more than 300 km from apex to mouth (4) The Niger Delta is framed on the northwest by a subsurface continuation of the West African Shield, the Benin Flank. The eastern edge of the basin coincides with the Calabar Flank to the south of the Oban Massif (5).

2.1. Stratigraphy and Sedimentary Layers

The Niger delta's geological sequence comprises three distinct formations, each representing a different depositional environment and temporal stage. Starting from the oldest and deepest layer, these are the Akata, Agbada, and Benin Formations, which collectively span from the Paleocene to the Recent geological periods

The Akata Formation, the lowermost unit, is predominantly composed of dark grey marine shales, silts, and clays (4). This formation potentially reaches an impressive thickness of up to 7,000 meters in the central delta region (6). Characterized by approximately 50% marine planktonic foraminifera, it represents a shallow marine shelf environment and has been dated from the Paleocene to Recent times (7).

The Agbada Formation sits above the Akata and forms the primary hydrocarbon-bearing sequence (4). This paralic (transitional marine-terrestrial) formation consists of alternating sand, silt, and clay layers, creating multiple reservoir-seal couplets. Spanning over 4,000 meters thick, with maximum deposition in the central delta (9), it ranges from the Eocene to Recent. The formation's coarse, poorly sorted sand grains suggest a fluviatile (river-related) origin (8).

The Benin Formation, the youngest and uppermost layer, is predominantly sandstone with occasional shale intercalations. Dated from possibly the Miocene to Recent, it ranges in thickness from 280 to 2,100 meters (10). Its sands vary from gravelly to fine-grained, are poorly sorted, and often contain lignite streaks and wood fragments (11), extending from the Benin-Onitsha area to the present coastline (7).

This stratigraphic sequence illustrates a progressive regression of depositional environments, transitioning from marine to deltaic and ultimately to fluvial settings, capturing millions of years of geological transformation in the Niger Delta region

3. Material and methods

The primary data source for this research is well-log data from four wells in the Niger Delta Basin, acquired and provided by Shell Petroleum Development Company. The logs utilized in this study include the following: Gamma Ray (GR), Sonic Transit Time (DT), Bulk Density (RHOB) and Neutron Porosity (NPHI). These logs collectively provide a comprehensive understanding of the subsurface, enabling the identification of potential reservoirs and the characterization of lithological variations. This initial step involved rigorous data cleaning, normalization, and feature selection to ensure the highest quality input for subsequent analysis.

Table 1 Well log measurements

Log	Description
Gamma Ray (GR)	Differentiates between sandstones and shales based on natural gamma radiation. Sandstones have low GR, while shales have higher GR
Sonic Transit Time (DT)	Provides information about formation porosity and lithology. Lower DT values correlate with higher porosity.
Bulk Density (RHOB)	Measures the bulk density of the formation, essential for porosity estimation. Higher RHOB indicates compact, less porous rocks.
Neutron Porosity (NPHI)	Measures hydrogen concentration, providing an estimate of rock porosity. Higher NPHI values indicate higher porosity.

3.1. Unsupervised Learning Approach

The first analytical phase employed unsupervised learning techniques, which are designed to discover inherent patterns in data without predefined labels. Three primary unsupervised clustering algorithms were implemented:

- K-means Clustering: This method partitions the data into K clusters, where each data point belongs to the cluster with the nearest mean. It aims to minimize the variance within clusters while maximizing the differences between clusters.
- Gaussian Mixture Models (GMM): A probabilistic model that assumes the data is generated from a mixture of a finite number of Gaussian distributions with unknown parameters. This approach allows for more flexible clustering by considering the probability of data points belonging to different clusters.
- Agglomerative Clustering: A hierarchical clustering method that starts with each data point as a separate cluster and progressively merges the closest clusters until a desired number of clusters is reached.

3.2. Supervised Learning Approach

For the Supervised Learning model, labeled data is used to train models to make precise predictions. Three primary supervised models were implemented:

- Support Vector Machines (SVM): A powerful classification algorithm that finds the optimal hyperplane separating different classes in the feature space.
- Random Forest: An ensemble learning method that constructs multiple decision trees and combines their outputs to make predictions, which helps reduce overfitting and improve generalization.
- Gradient Boosting: An advanced ensemble technique that builds models sequentially, with each new model correcting the errors of the previous models.

3.3. Model Evaluation and Validation

The reliability of instruments used in the study was assessed through cross-validation and performance evaluation to ensure the robustness and generalizability of the machine learning models. A classification report was generated, detailing precision, recall, and F1-scores for each facies class. This evaluation ensured balanced model performance across all classes, reducing the risk of overfitting to any specific facies.

Cross-validation was employed to validate the models on unseen data by splitting the dataset into training and testing subsets. This iterative process enabled the calculation of performance metrics on the test set, providing a reliable measure of the models' predictive capabilities and ensuring they generalized well to new, unobserved data

4. Results and discussion

4.1. Unsupervised Learning

4.1.1. K-means Clustering Analysis

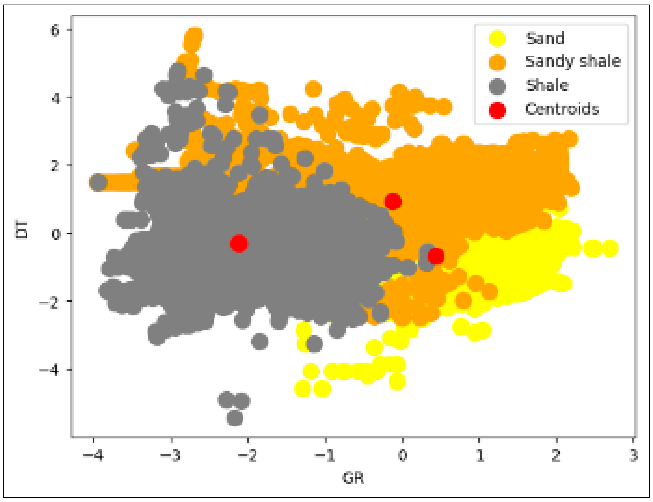

Figure 2 K-Means cluster plot on GR vs DT

The K-means algorithm aims to divide data points into a pre-defined number of clusters (k) based on their similarity. These clusters are defined by their centroids, which represent the average well log values for the data points within each cluster and can be interpreted as different facies based on the characteristic well log responses of each facies. In Figure 2, Sand facies exhibit low GR and DT values, indicating high porosity and permeability. Shale facies show high GR and DT values, suggesting low porosity and permeability. Sandy Shale facies fall in between, representing a transitional zone.

Sand facies in Figure 3, display low GR and RHOB values, typical of porous and less dense rocks. Shale facies have high GR and RHOB values, characteristic of denser and less porous rocks. Sandy Shale facies occupy an intermediate position.

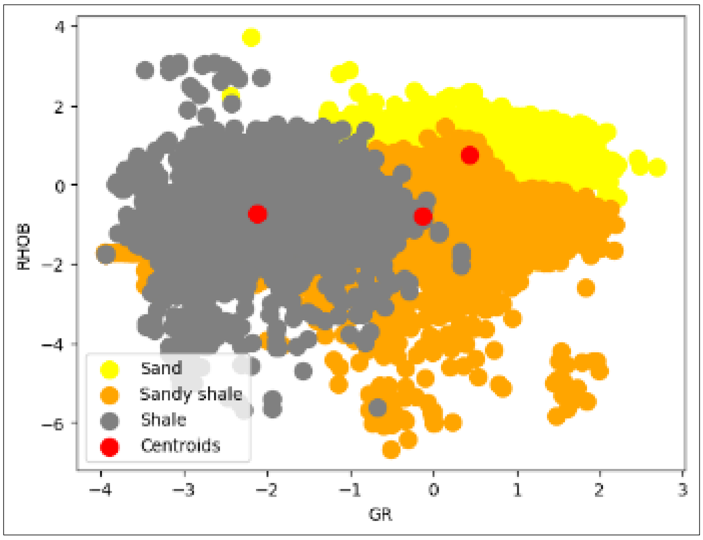

Figure 3 K-Means cluster plot on GR vs RHOB

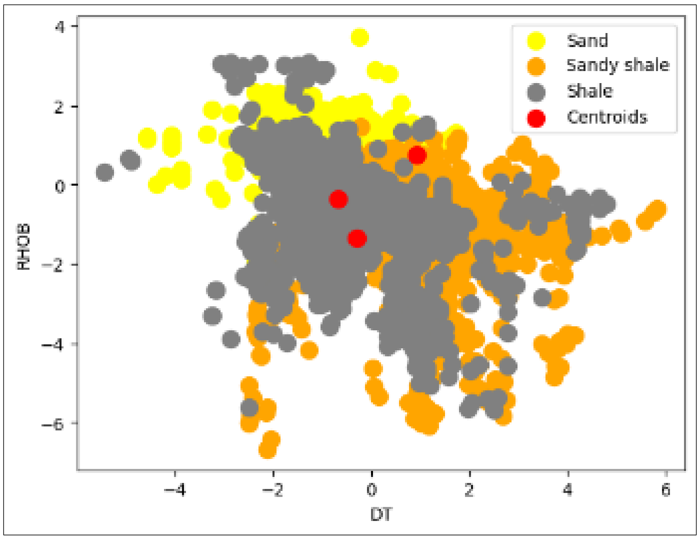

Figure 4 K-Means cluster plot on DT vs RHOB

In Figure 4 Sand facies have low DT and RHOB values, typical of well-consolidated and porous rocks. Shale facies show high DT and RHOB values, suggesting poorly consolidated and less porous rocks. Sandy Shale facies fall between these extremes.

4.1.2. Gaussian Mixture Model (GMM) Clustering Analysis

Each cluster of the Gaussian Mixture Model (GMM) produces comparable cross plots with similar facies interpretations characterized by its mean, covariance, and weight based on well log response. The Gaussian nature of GMM offer a slightly better separation of clusters when compared to k-means, especially when dealing with overlapping data points.

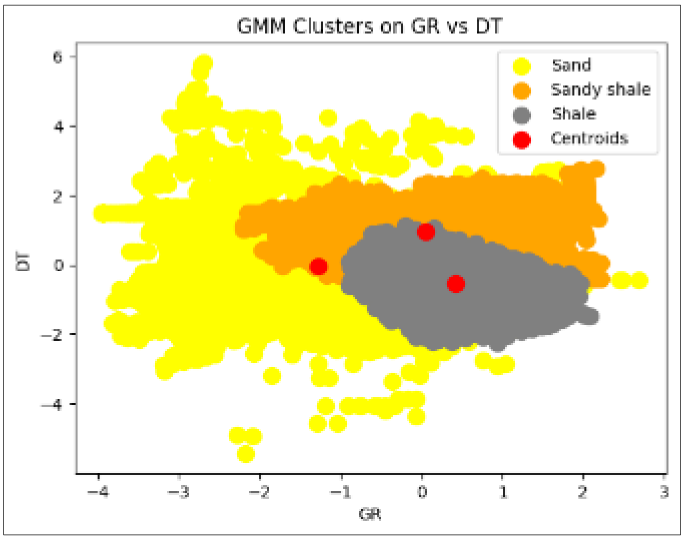

Figure 5 GMM cluster plot on GR vs DT

Figure 5 shows sand with lower GR values with a spread in DT, suggesting clean reservoir rocks with variable compressibility, the sandy shale depicts intermediate GR and DT values, representing mixed lithology with transitional properties, whereas shale has higher GR values and a concentrated range of DT, indicating compact fine-grained rocks.

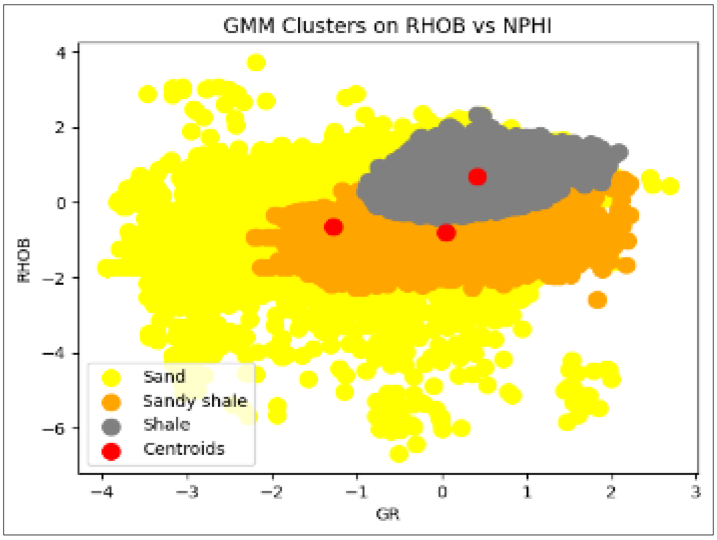

Figure 6 GMM cluster plot on GR vs RHOB

In Figure 6 we notice that sand has a lower GR and a wide spread of RHOB, indicating low radioactive content and variable density, shale has higher GR and RHOB values, showing dense and radioactive rock formations and sandy shale has Intermediate GR and RHOB values, transitioning between Sand and Shale.

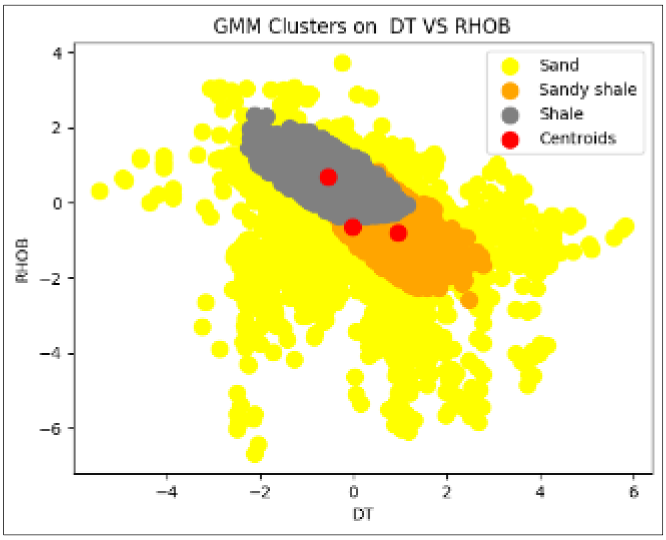

Figure 7 GMM cluster plot on DT vs RHOB

In Figure 7, the variation in DT within the sand cluster indicates differing porosity levels, with the majority concentrated near the bottom. The shale cluster exhibits higher DT and RHOB values, characteristic of compacted, clay-rich formations. The sandy shale represents transitional rocks with intermediate DT and RHOB values, suggesting sands mixed with clay content.

4.1.3. Agglomerative Clustering Analysis

Agglomerative clustering is a hierarchical clustering algorithm that builds a hierarchy of clusters by iteratively merging the most similar clusters. It starts with each data point as a separate cluster and progressively combines them until all points belong to a single cluster. The Agglomerative clusters are identified based on the hierarchy of clusters formed by the algorithm and can be interpreted as different facies based on the characteristic well log responses of each facies.

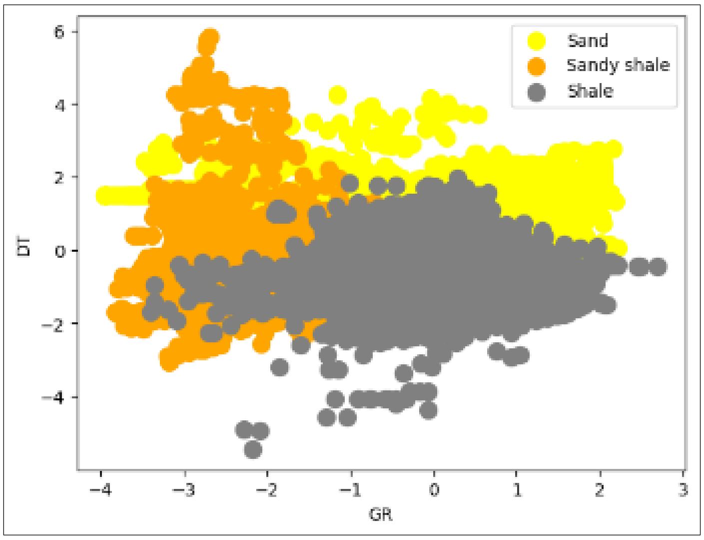

Figure 8 Agglomerative cluster plot on GR vs DT

In Figure 8 Sand facies are clustered at low GR and DT values, indicating good reservoir properties. Shale facies are concentrated at high GR and DT values, representing non-reservoir rocks. Sandy Shale facies occupy an intermediate zone

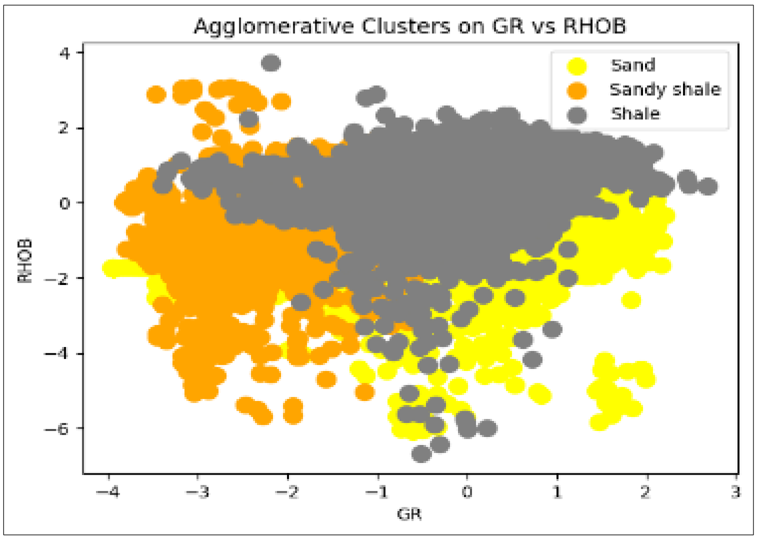

Figure 9 Agglomerative cluster plot on GR vs RHOB

In figure 9 Sand is predominantly associated with low GR and a range of RHOB values, indicating cleaner reservoirs with lower radioactive content. Shale exhibits high GR and higher RHOB values, typical of fine-grained materials with high radioactive content, Sandy shale appears intermediate, with moderate GR and RHOB values, reflecting its mixed lithological nature.

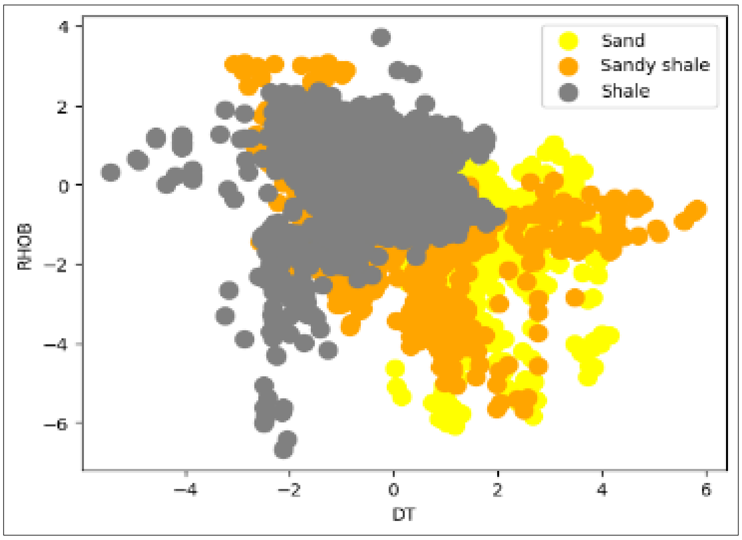

Figure 10 Agglomerative cluster plot on DT vs RHOB

Figure 10 shows that majority of sand facies clusters have low DT and RHOB values, typical of well-consolidated and porous rocks. Shale facies show high DT and RHOB values representing clay-rich formations. Sandy Shale facies fall between these extreme.

4.2. Supervised machine learning analysis

Supervised learning techniques were employed to classify facies and evaluate petrophysical characteristics using labeled data from the well logs. These methods were trained to predict facies based on log features like GR, NPHI, DT, and RHOB, which are directly related to lithological properties. The model when was then applied to a blind well (Freeman-005) to predict its facies (Fig 11).

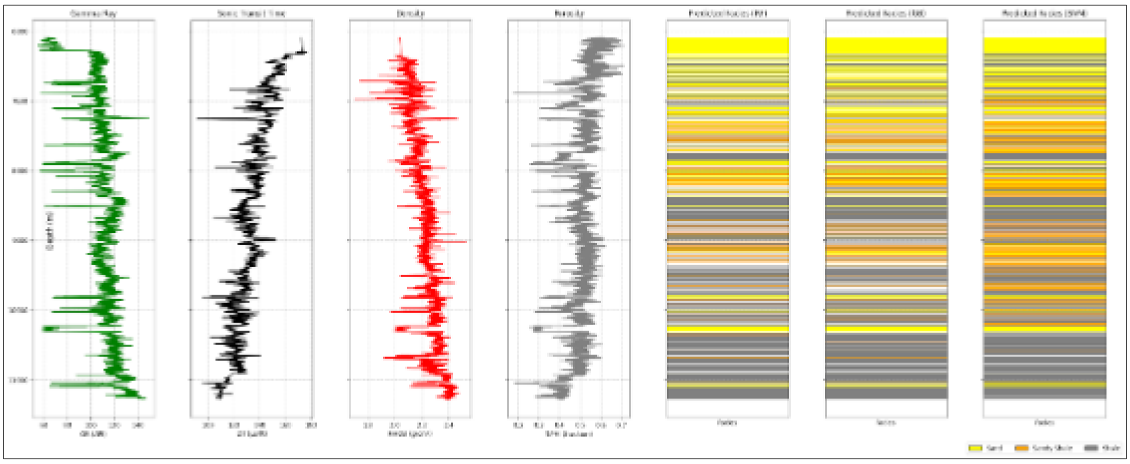

Figure 11 Well logs showing predicted facies distribution log from the 3 different supervised learning algorithm

4.2.1. Random forest

Random Forest was applied to classify facies based on the training dataset. This ensemble method, which builds multiple decision trees and aggregates their predictions, performed well in identifying facies due to its ability to handle complex relationships between log features. Feature importance analysis (Fig. 12) showed that Gamma (GR) log is the most influential in distinguishing between different facies, followed by Density (DT), Rhob (RHOB) and NPHI.

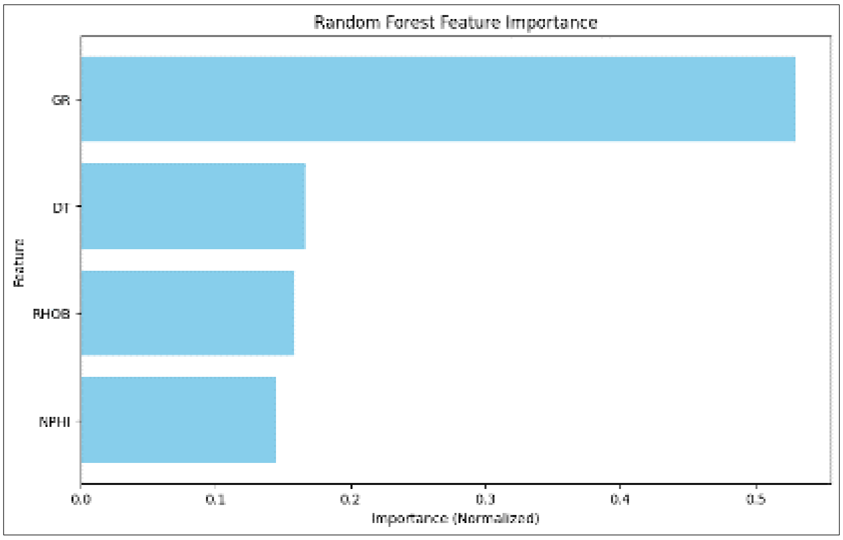

Figure 12 Bar Chart of Feature Importance for Random Forest Algorithm

4.2.2. Support vector machine (SVM)

Support Vector Machine (SVM) with a radial basis function (RBF) kernel was used to classify the facies. SVM's ability to create a decision boundary between facies classes allowed for a high precision in facies classification. SVM was particularly effective in handling high-dimensional well log data and distinguishing between closely related facies like sandy shale and shale.

4.2.3. Gradient boosting

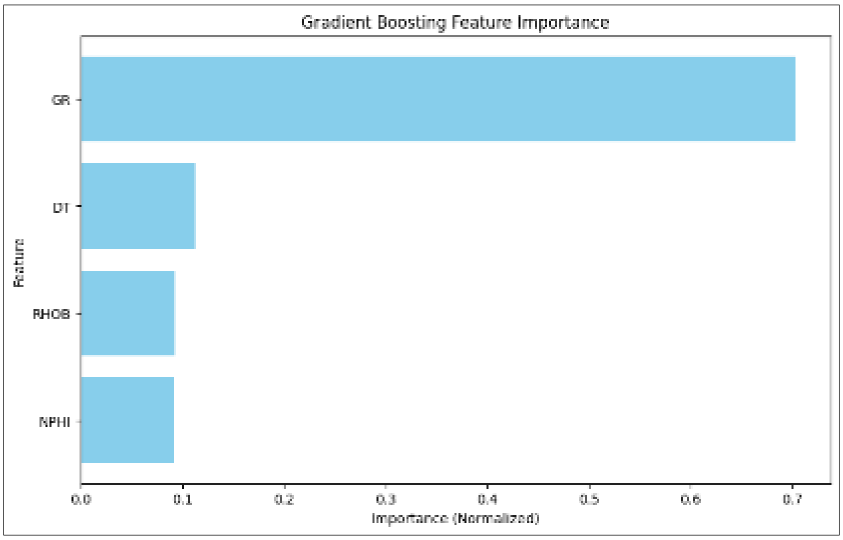

Figure 13 Chart of Feature Importance for Gradient Boosting Algorithm

Gradient Boosting was used to build a predictive model by iteratively improving weak models. Gradient Boosting was beneficial in handling the non-linear relationships within the well logs and refining the prediction of facies with complex log responses. The method showed strong accuracy in predicting the facies boundaries between reservoir and non-reservoir zones. Feature importance analysis (Fig. 13) also showed that Gamma (GR) log is the most influential in distinguishing between different facies, followed by Density (DT), Rhob (RHOB) and NPHI.

4.3. Performance evaluation

Evaluating the performance of the machine learning algorithms is crucial to determine their effectiveness in grouping similar data points for facies classification.

4.3.1. Unsupervised Learning

The unsupervised learning methods revealed significant challenges in facies classification. K-means clustering demonstrated moderate performance, achieving an accuracy of 47.87% and highlighting the complexity of distinguishing between overlapping facies classes. The method's weighted precision of 43.41% and F1-score of 44.77% underscored its limitations in capturing the nuanced geological variations.

Table 2 Comparison of performance metrics of unsupervised algorithm

Metric	k-Means	GMM	Agglomerative Clustering
Accuracy	47.87%	20.38%	13.94%
Precision	43.41%	30.61%	8.60%
Recall	47.87%	20.38%	13.94%
F1-score	44.77%	23.71%	10.09%

Gaussian Mixture Models presented even more pronounced difficulties, with a notably low accuracy of 20.38%. The model's struggle to effectively model data distributions became evident through its weak weighted precision of 30.61% and F1-score of 23.71%. This approach proved particularly ineffective in handling the intricate spatial relationships within the geological dataset.

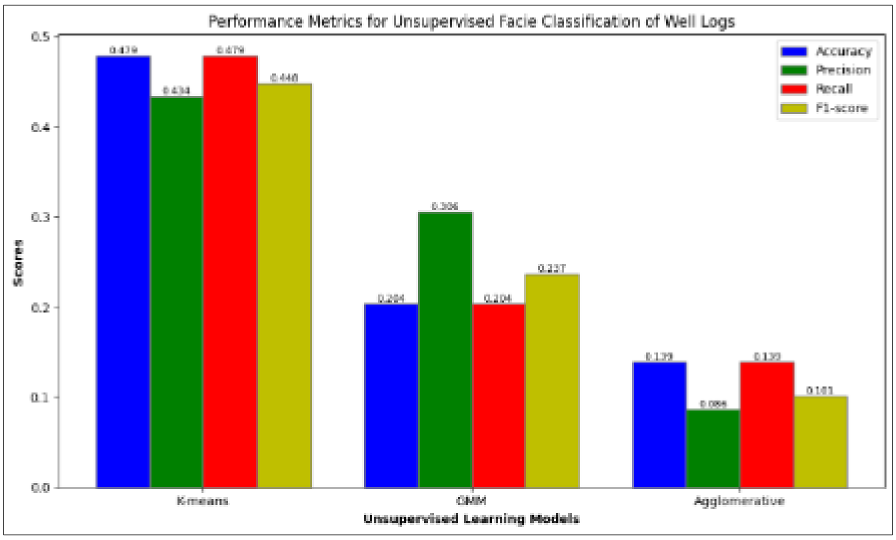

Figure 14 Bar chart of Performance metrics of Unsupervised Learning Models

Agglomerative Clustering emerged as the least successful unsupervised method, with an accuracy of merely 13.94%. The extremely low precision of 8.60% and F1-score of 10.09% dramatically illustrated the method's inability to capture the complex patterns inherent in facies classification. These results highlight the inadequacy of this method for handling the complex relationships and overlaps inherent in the dataset.

4.3.2. Supervised Learning

Table 3 Comparison of performance metrics of supervised algorithm

Metric	**Random Forest**	**Gradient Boosting**	**SVM**
Accuracy	87.06%	87.53%	77.86%
Precision	87.07%	87.55%	77.78%
Recall	87.06%	87.53%	77.86%
F1-score	87.06%	87.53%	77.82%

In contrast to the unsupervised algorithm method, supervised learning methods demonstrated significantly superior performance. Support Vector Machines (SVM) delivered a robust performance, achieving an accuracy of 77.86%. With consistent precision, recall, and F1-score around 77.8%, the model offered a balanced approach to classification, though it still encountered challenges with class imbalances.

Random Forest emerged as a standout performer, achieving an impressive 87.06% accuracy. The model's uniform precision, recall, and F1-score highlighted its exceptional ability to handle complex feature patterns and address class imbalances effectively. It is consistent performance across different metrics underscored its reliability in facies classification.

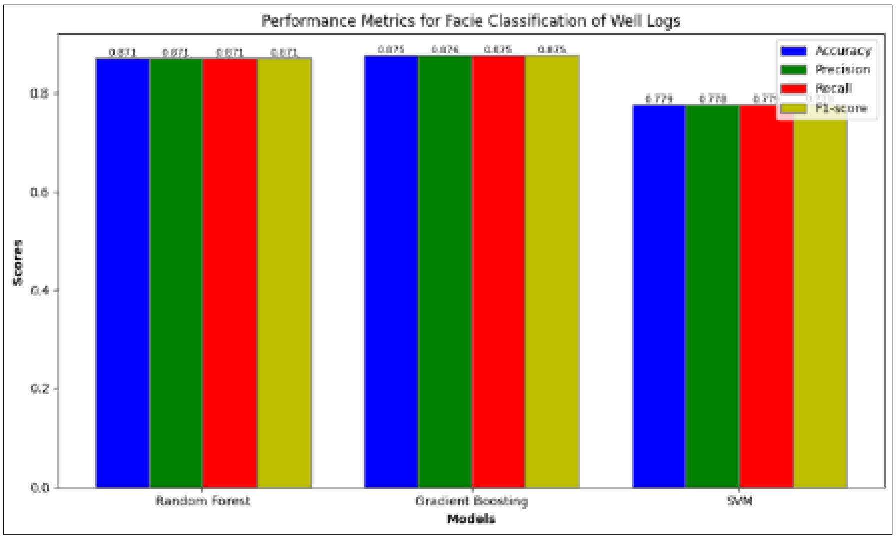

Figure 15 Bar chart of Performance metrics of Supervised Learning Models

Gradient Boosting ultimately emerged as the top-performing algorithm, marginally surpassing Random Forest with an accuracy of 87.53%. The model's consistent precision, recall, and F1-score of 87.55% demonstrated its advanced capability to capture intricate relationships within the geological data, providing the most nuanced and accurate classification approach

5. Conclusion

Machine learning offers an efficient and scalable solution for facies classification in complex geological settings such as the Niger Delta. The findings from this study emphasize the critical importance of algorithm selection, understanding dataset characteristics, and implementing rigorous evaluation methodologies. The research suggests that for complex geological classification tasks, supervised learning approaches offer substantially more reliable and insightful results, with Gradient Boosting representing the most promising technique for capturing the intricate patterns of facies distribution.

It should be noted that the accuracy and reliability of machine learning models are heavily influenced by the quality of the input data. Effective data preprocessing, including standardization, handling missing values, and feature selection, is crucial to achieving meaningful results

Compliance with ethical standards

Acknowledgments

I am unreservedly grateful to Prof. (Mrs.) J.E. Emudianughe, whose input was critical to the completion of this research.

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] M. E. Nton, M. O. Adeyemi · "Evaluation of Hydrocarbon Reserve in AD Field, Offshore Niger Delta". Open Journal of Geology Vol.11 No.5， May 20, 2021. DOI: 10.4236/ojg.2021.115009

[2] Chen, W., Ma, X., Wang, Z. et al. Exploring neuro-symbolic AI applications in geoscience: implications and future directions for mineral prediction. Earth Sci Inform (2024). https://doi.org/10.1007/s12145-024-01278-7.

[3] Reijers, T., Petters, S., & Nwajide, C. (1996). The Niger Delta Basin. In T.J.A. Reijers (ed.), Selected Chapters on Geology, 103-114. Warri: SPDC corporate reprographic services.

[4] Doust, H., & Omatsola, O. (1990). Niger Delta. Edwards, J.D. and Santoyrossi, P.A. (Eds.): Divergent and Passive Margin Basin, 48, 201-238.

[5] Murat, R. (1972). Stratigraphy and paleogeography of the Cretaceous and Lower Tertiary in southern Nigeria. In T. Dessauvagie, & A. (. Whiteman, African Geology, 251–266. Nigeria: University of Ibadan Press.

[6] Nwajide, C. S., & M., H. (1985). "Problems of classification and maturity evaluation of diagenetically altered fluvial Sandstone. Geol. Mijnbouw, 64, 69-77.

[7] Short, K., & Stauble, A. (1967). Outline of geology of Niger delta. American Association of Petroleum Geologists Bulletin 51, 761–779

[8] Aturamu, A. O., Ojo, A. O., Adebayo, O. F., & Akinyemi, S. A. (2015). Palynostratigraphic Analysis of the Agbada Formation (Nep-1 Well) Offshore, Eastern Niger-Delta Basin, Nigeria. British Journal of Environmental Sciences, 3(5), 19-31.

[9] Evamy, B. D., Haremboure, J., Kamerling, P., Knaap, W. A., Molloy, F. A., & Rowlands, P. H. (1978). Hydrocarbon habitat of Tertiary Niger Delta. American Association of Petroleum Geologists Bulletin, 62, 1–39.

[10] Whiteman, A. (1982). Nigeria: Its Petroleum Geology Resources and Potential. London: Graham and Trotman.

[11] Akpoborie, I. A., B. N., Etobro, A., & Odagwe, S. (2011). Aspects of the Geology and Groundwater Conditions of Asaba, Nigeria. Archives of Applied Science Research, 537- 550.