

Application of K-Means Clustering for Hydrocarbon Reservoir Delineation and Reservoir Parameter Estimation

By

Abe S. James,* Mekanjuola David R**

Abstract

The results presented in this research demonstrate how the K-means clustering algorithm was integrated with well logs to build a machine learning model for delineating hydrocarbon-bearing zones of conventional reservoirs and estimating reservoir parameters from well logs. The well log data used for training the model were loaded and necessary cleanup processes were performed which entail the handling of missing values and the removal of outliers, and some exploratory data analyses (EDA) were carried out including the plotting of numerical distributions, followed by feature selection and scaling. All the aforementioned steps were taken in order to transform the well logs to a format suitable for the K-means algorithm. After the data transformation phase, the K-means algorithm was applied on the well logs to develop a machine learning clustering model that grouped depth points together based on the similarity of their log values. The clusters generated from the clustering process were evaluated to determine the one that consistently and accurately predict the hydrocarbon reservoirs in the wells. Some equations, such as Archie's water saturation equation, were incorporated so that reservoir parameters could be estimated by the program. The model, when tested on three wells, delineated about twelve zones as potential hydrocarbon-bearing zones. These zones were manually interpreted to assess the accuracy of the model and it was found that its precision in delineating the hydrocarbon reservoirs was high. The significance of the research study is that the model developed can be used to automate the task of delineating hydrocarbon reservoirs from well logs.

Keywords: Well log, Unsupervised learning, Reservoir delineation, K-means, Machine Learning

Introduction

The importance of hydrocarbon reservoir delineation from well logs in the oil and gas industry by geoscientists cannot be over-emphasized as they are integral steps in reservoir characterization.^{1,2} Well logs are systematic records of rock properties and fluid contents in wells being drilled or produced to obtain various petrophysical parameters and characteristics of downhole sequences.³ The properties are non-linear and depend majorly on factors like mineral composition, porosity, cementation/compaction, presence of fluids and saturation of the fluids. There are different kinds of well logs used for different purposes, for example, the gamma ray

*Department of Applied Geophysics, Federal University of Technology, PMB 704 Akure, Nigeria.

Corresponding author's e-mail: jsabe@futa.edu.ng

**Final year student Department of Applied Geophysics, Federal University of Technology, Akure, Nigeria

¹Han, L., Fuqiang, L., Zheng, D., & Weixu, X. (2018). A lithology identification method for continental shale oil reservoir based on BP neural network. *Journal of Geophysics and Engineering*, 15(3), 895–908.
<https://doi.org/10.1088/1742-2140/aaa4db>.

²Fatoke, Oluwaseyiadedamola (2010). Sequence Stratigraphy of the Pliocene-Pleistocene Strata and Shelf-margin Deltas of the Eastern Niger Delta, Nigeria (Ph.D) University of Houston.

³Archie, G. E. (1950). Introduction to Petrophysics of Reservoir Rocks. *AAPG Bulletin*, 34(5), 943–961.
<https://doi.org/10.1306/3D933F62-16B1-11D7-8645000102C1865D>.

and the spontaneous polarization logs are excellent lithology indicators, while some others are used for determining rock fluid types (resistivity logs) and others for identifying hydrocarbon types (porosity logs).

With the recent tremendous development in algorithms, computation power and availability of the enormous amount of data, the implementation of machine learning approach has spurred the interest of oil and gas industry and brings the data science and analytics into the forefront of future energy.⁴ Given the factors that make a good data for machine learning algorithms (i.e. large quantity, representative, properly tagged etc.), it is obvious that geophysical data, especially the well logs, fall into the regime of data that machine can learn from.⁵ Over the years, researchers have integrated Unsupervised learning and well logs to enhance the modeling subsurface processes. For example,⁶ applied hybrid uncertainty-clustering approach to pre-process well logs,⁷ integrated cluster analysis and well logs to identify dominant lithologic units in wells in their work published in 2020.

Also,⁸ used Fuzzy C-Means clustering and Gustafson-Kessel algorithm with well logs to predict saturation and permeability in lithologies. But none has used Unsupervised learning to delineate hydrocarbon reservoirs from well logs. Thus, this research project was aimed at utilizing the capabilities of an unsupervised learning algorithm, namely K-means clustering, to reformulate the task of delineating hydrocarbon reservoirs from well logs.

K-means algorithm is an iterative algorithm that tries to partition the dataset into K pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group.⁹ It makes the intra-cluster data points as similar as possible while also keeping the clusters as different (far) as possible. It assigns data points to a cluster such that the sum of the squared distance between the data points and the cluster's centroid (arithmetic mean of all the data points that belong to that cluster) is at the minimum. The less variation there are within the clusters, the more homogeneous (similar) the data points are within the same cluster.^{10,11,12}

⁴Mohamed, I. M., Mohamed, S., Mazher, I., & Chester, P. (2019). Formation Lithology Classification: Insights into Machine Learning Methods. Day 2 Tue, October 01, 2019. <https://doi.org/10.2118/196096-ms>.

⁵Aurélien Geron. (2017). Hands-on machine learning with Scikit-Learn and TensorFlow: Concepts, tools, and techniques to build intelligent systems. Sebastopol, Ca: O'Reilly Media.

⁶Pedram, M. (2017). Application of hybrid uncertainty-clustering approach in pre-processing well-logs. Earth Sciences. Université Rennes.

⁷Amjad, A., Chen, S. (2020). Characterization of well logs using K-mean cluster analysis. Journal of Petroleum Exploration and Production Technology.

⁸Pedram M., Behzad T., Majid, A.J., Behzad M. (2012) Application of Fuzzy Classifier Fusion in Determining Productive Zones in Oil Wells. Energy Exploration & Exploitation 30(3):403-416

⁹Paper, D. (2020). Hands-on Scikit-learn for machine learning applications: Data science fundamentals with Python. Berkeley, California: Apress.

¹⁰Dabbura, I. (2018). K-means Clustering: Algorithm, Applications, Evaluation Methods, and Drawbacks. Retrieved from Medium website: <https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a>.

¹¹Bonaccorso, G. (2019). Hands-on unsupervised learning with Python: Implement machine learning and deep learning models using Scikit-Learn, TensorFlow, and more. Birmingham, UK: Packt Publishing.

¹²Celebi, M. E., Kingravi, H. A., & Vela, P. A. (2013). A comparative study of efficient initialization methods for the k-means clustering algorithm. Expert Systems with Applications, 40(1), 200–210. <https://doi.org/10.1016/j.eswa.2012.07.021>.

Application of K-Means Clustering for Hydrocarbon Reservoir Delineation and Reservoir Parameter Estimation

The K-means algorithms are sensitive to outliers and do not perform well for datasets that do not have clustering tendency. These intrinsic limitations of the algorithm and others were duly corrected for, to enhance the result of the clustering process.

Description and Geology of the Study Area

The study area is located onshore Niger Delta. The Niger Delta basin accounts for the most hydrocarbon production at present-day Nigeria and is situated on the continental margin of the Gulf of Guinea in equatorial West Africa between latitude 3°N and 6°N and longitude 5°E and 8°E.¹³ It extends over about 70,000km² and makes up 7.5% of Nigeria's land mass.^{14, 15}

Niger Delta is characterized by a three-fold litho-stratigraphic subdivision comprising an upper sandy Benin Formation, an intervening unit of alternating sandstone and shale named the Agbada Formation, and a lower shaly Akata Formation.^{16,17} Oil is produced from sandstone facies within the Agbada Formation. However, turbidite sand in the upper Akata Formation is a potential target in deep water offshore and possibly beneath currently producing intervals onshore.

Materials and Method of Study

The materials deployed for this research project are mainly suite of open-source software (Python 3.7 with some of its Libraries) and well logs.

In the subsequent sections, the well data will be imported, necessary cleanup processes will be performed, and some exploratory data analysis (EDA) will be carried out, followed by feature selection, and then the k-means clustering algorithm will be integrated with the well logs to develop a clustering model that will group similar depth samples together based on the similarity of their log values. The clusters (groups) generated from the clustering process will be compared with independently determined lithologies and reservoirs at some selected depth intervals as a means to evaluate the consistency and accuracy of the unsupervised clustering model in predicting the hydrocarbon reservoirs in the well. The schematic of the methodology adopted is as shown in Figure 1.

¹³Reijers, T.J.A., Petters, S.W., and Nwajide, C.S., (1997) The Niger Delta Basin, in Selley, R.C., ed., African Basins--Sedimentary Basin of the World 3: Amsterdam, Elsevier Science, pp. 151-172.

¹⁴Stacher, P., (1995) Present understanding of the Niger Delta hydrocarbon habitat, in, Oti, M.N., and Postma, G., eds., Geology of Deltas: Rotterdam, A.A. Balkema, p. 257-267.

¹⁵Ejedawe, J.E., Coker, S.J.L., Lambert-Aikhionbare, D.O., Alofe, K.B., and Adoh, F.O., (1984) Evolution of oil-generative window and oil and gas occurrence in Tertiary Niger Delta Basin: American Association of Petroleum Geologists, v. 68, p. 1744-1751.

¹⁶Doust, H., and Omatsola, E., 1990, Niger Delta, in, Edwards, J. D., and Santogrossi, P.A., eds., Divergent/passive Margin Basins, AAPG Memoir 48: Tulsa, American Association of Petroleum Geologists, p. 239-248.

¹⁷Haack, R.C., Sundararaman, P., and Dahl, J., (1997) Niger Delta petroleum System, in, Extended Abstracts, AAPG/ABGP Hedberg Research Symposium, Petroleum Systems of the South Atlantic Margin, November 16-19, 1997, Rio de Janeiro, Brazil.

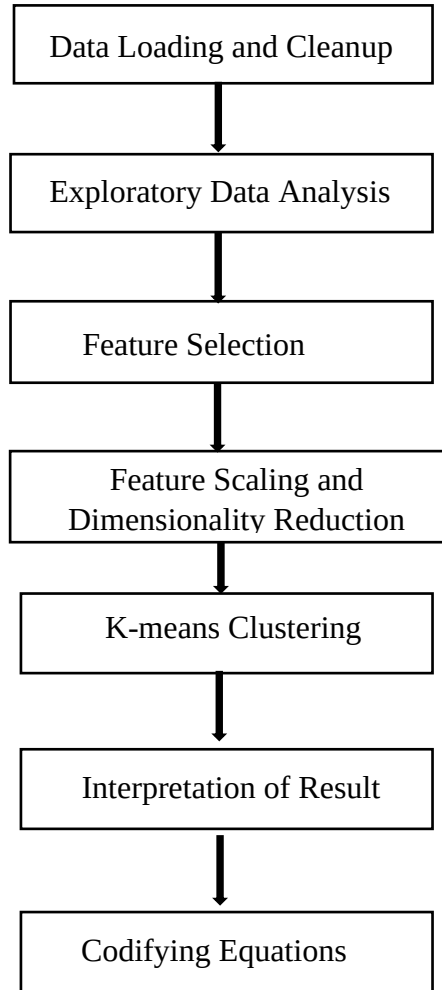


Figure 1: The Methodology of the Study

Data Loading and Cleaning

The well logs were originally in the LAS (Laser) format. They were converted to Dataframe- a format useful for processing. A part of the Dataframe is displayed in Table 1. The logs present are: the gamma ray, resistivity, density, and sonic.

Previewing the data allows for some quick inspection of the well logs so that data types can be examined and edited as necessary. The well logs were subjected to two cleanup phases: removal of NaNs and filtering out of the outliers.

The NaNs in the different logs are not critical to the training of the machine learning model because logging was not made at those depth points intentionally. As a result, they do not need any special treatment other than to simply remove them. Thus, all the rows containing the NaN values were removed to get the actual number of sample points in the well data.

Table 1: Dataframe of Well Logs

Depth	RHOB (Density log)	SON3 (Sonic log)	RT_O (Resistivity log)	DTN (also Sonic log)	GR_NM (Gamma ray log)
5000.0	NaN	NaN	NaN	NaN	NaN
5000.5	NaN	NaN	NaN	NaN	NaN
5001.0	NaN	NaN	NaN	NaN	NaN
5001.5	NaN	NaN	NaN	NaN	NaN
5002.0	NaN	NaN	NaN	NaN	NaN

Table 2 is the summary statistics of the well logs which include maximum values, minimum values, the mean and the standard deviation of the logs. Comparing the maximum and minimum values of each of the logs with the standard log scale ranges (Table 3) shows that there are outliers in the depth samples.

Table 2: Summary Statistics of the Raw Well Dataset

Summary Statistics	RHOB (Density Log)	SON3 (Sonic Log)	RT_O (Resistivity Log)	DTN (Sonic Log)	GR_NM (Gamma ray Log)
Count	8637.00	8637.00	8637.00	8637.00	8637.00
Minimum	1.65	56.23	0.79	56.23	4.79
Maximum	2.67	141.64	4244.92	141.64	155.46
Mean	2.25	99.69	117.27	99.69	62.19
Standard deviation	0.14	10.56	475.22	10.56	27.88

Table 3: Standard Range of Values for Well Logs

S/N	Well logs	Scale (Unit)
1	RHOB (Density)	1.65-2.65 (gg/cc)
2	SON3 (Sonic)	40-140 (μ s/ft)
3	RT_O(Resistivity)	0.2 – 2000 (ohm.m)
4	DTN(Sonic)	40-140 (μ s/ft)
5	GR_NM (Gamma ray)	0 – 150 (API)

Values outside of the range given for each log (Table 3) are considered to be outliers. A form of band pass filter function with a pre-determined lower and upper limits was used to remove the outliers. The bandpass filter was chosen so as to allow a specific range of log values to be passed to the next stage of the workflow. The summary statistics of the filtered data was generated to ascertain the effectiveness of the filtering operation (Table 4). It is evident from the table that the

maximum and the minimum values for each log fall within the standard range unlike the raw data.

Removal of outlier is an important step as the k-means clustering algorithm is very sensitive to outliers.

Table 4: The Summary Statistics of the Well Dataset after the Removal of Outliers

Summary Statistics	RHOB (Density Log)	SON3 (Sonic Log)	RT_O (Resistivity Log)	DTN (Sonic Log)	GR_NM (Gamma ray Log)
Count	8456.00	8456.00	8456.00	8456.00	8456.00
Minimum	1.65	56.24	0.79	56.24	4.79
Maximum	2.65	139.07	1999.89	139.07	144.61
Mean	2.26	99.72	59.39	99.72	62.94
Standard deviation	0.14	10.61	231.73	10.61	27.54

Exploratory Data Analysis (EDA)

Numerical Distributions (Histograms)

The purpose of Exploratory Data Analysis was to understand the data. The information from EDA made the rest of the project smoother by providing valuable insight for further data cleanup.

The histogram of the different well logs was plotted for the purpose of inspecting the distribution of the observations in each log. One of the ideal data requirements of the k-means algorithm is that the data distributions are symmetric or nearly symmetric and not heavily skewed. All the logs, except the resistivity log, have fairly symmetric distribution as seen in the histogram plots (Figure 2). The distribution of the resistivity data is highly skewed because resistivity log scales are logarithmic and not linear. The performance of the clustering algorithm is largely affected by the distribution of the sample points; hence, log transformation was used to transform the resistivity log to a linear scale. All the logs now have fairly symmetric distribution as seen in the histogram plots (Figure 3). The figure also shows that the distribution of the depth samples are not uniform, else the data will not have a clustering tendency. The logs were then rearranged and renamed.

Application of K-Means Clustering for Hydrocarbon Reservoir Delineation and Reservoir Parameter Estimation

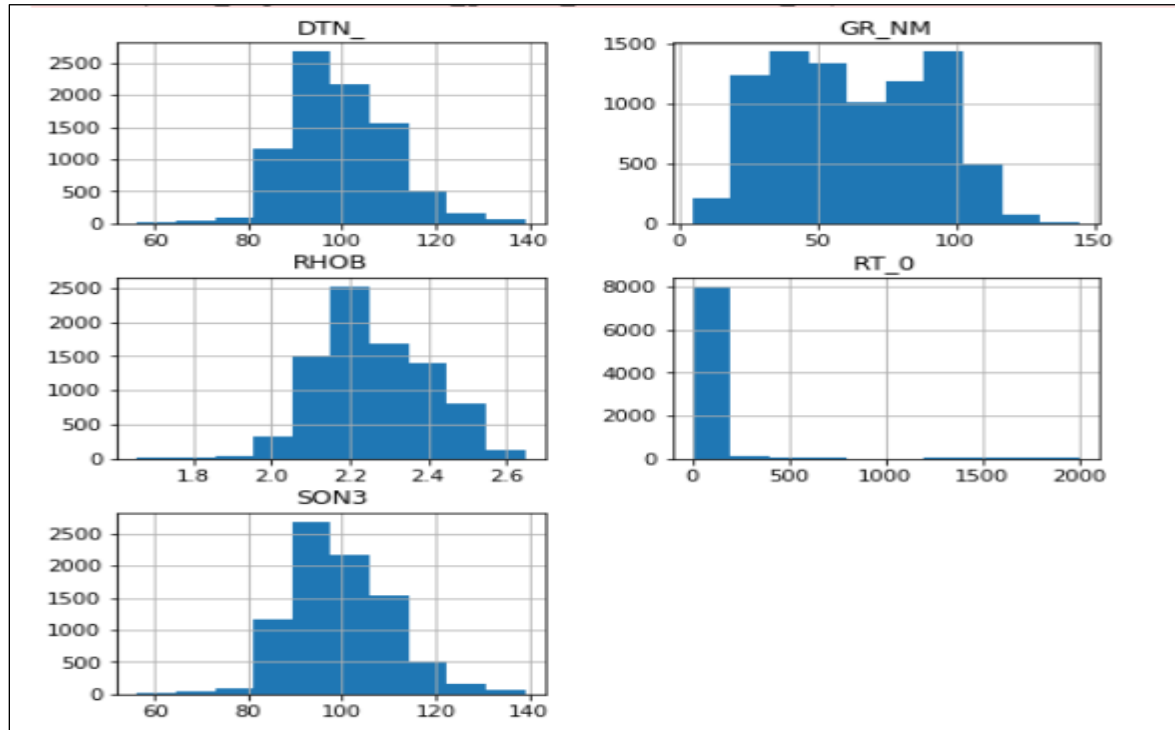


Figure 2: Histograms of Observation Distribution of Well Dataset (Before Log Transformation)

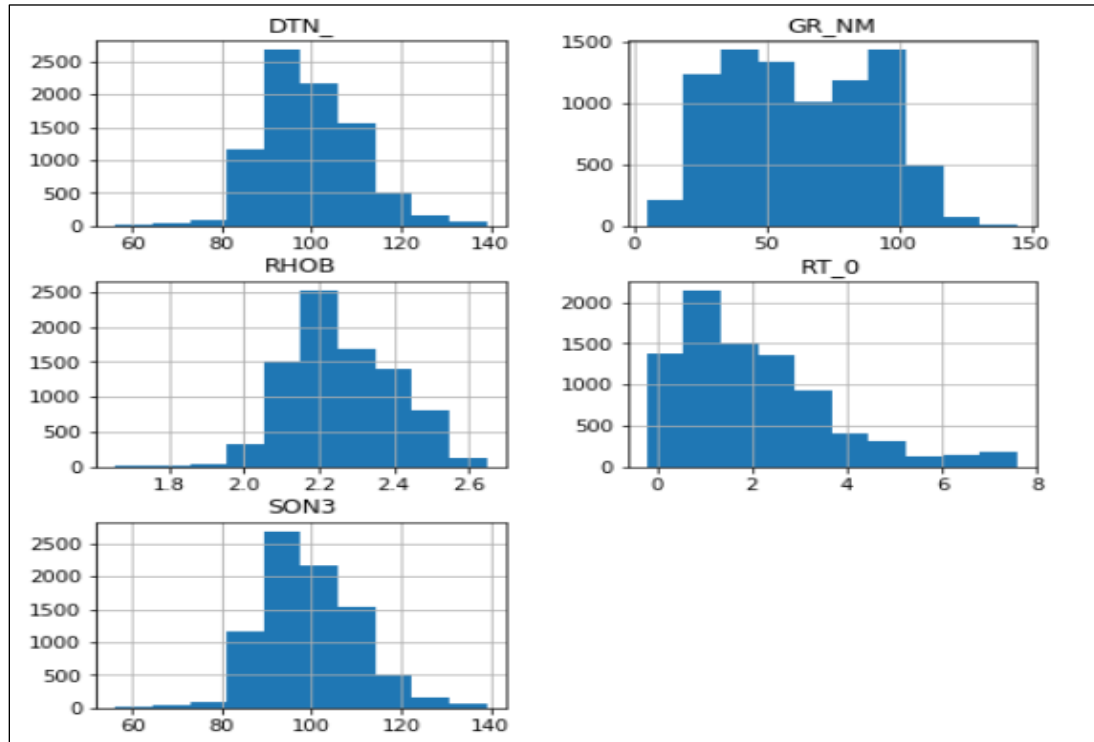


Figure 3: Histograms of Distribution of Well Data (After Log Transformation)

Feature Selection

Table 5 shows the Pearson's rank coefficient for the logs. It shows the mutual dependency of the features on one another. Well-log data usually contain many geophysical logs that may be used for reservoir parameter estimation; however, the performance of the clustering process depend on the choice of the features (logs) in all the phases of model development cycle. The gamma ray log is the most correlated with the others, followed by the density log.

The most relevant logs for the training of the algorithm include the gamma ray, density and resistivity logs. The features selection helps to reduce the computational cost of modeling and to improve the performance of the algorithm.

Table 5: Pearson's Rank Correlation Coefficient of the Logs

	Gamma ray	Resistivity	Density	Sonic	DTN
Gamma ray	1	-0.56	0.62	0.23	0.23
Resistivity	-0.56	1	-0.55	0.04	0.04
Density	0.62	-0.55	1	0.12	0.12
Sonic	0.23	0.04	0.12	1	1
DTN	0.23	0.04	0.12	1	1

Feature Scaling and Dimensionality Reduction

The well dataset contains features (logs) that vary greatly in magnitudes, units and ranges. Therefore, they needed to be brought to the same relative scale to enhance the performance of the algorithm.

The two unsupervised machine learning algorithms used in this study (i.e., the PCA and K-means) use the Euclidean distance measure and are therefore sensitive to magnitudes. That is, if a feature in the dataset is big in scale compared to others, the big-scale feature becomes dominating in such algorithm.

The Minimum-maximum scaling function was used to normalize the dataset within a particular range (0 and 1) and also speed up the calculation in the algorithm. The scaler function shifts the data such that all features are exactly between 0 and 1. The scaler function was chosen because the upper and lower boundaries are well known from domain knowledge.

The formula used in the backend is given by:

$$M = (X - X_{\min}) / (X_{\max} - X_{\min})$$

Where:

M is the new value, X is the original value, $X_{\min}$ is the minimum value of the column, $X_{\max}$ is the maximum value of the column.

Application of K-Means Clustering for Hydrocarbon Reservoir Delineation and Reservoir Parameter Estimation

The number of input variables or features for a dataset is referred to as its dimensionality. Generally, more input features often make a predictive modeling task more challenging to algorithms. Hence it was desirable to reduce the dimensionality as much as possible without loss of data quality.

Principal Component Analysis (PCA), one of the simplest and most widely used algorithms for data transformation, was used to reduce the dimensionality of the well-log data. Besides dimensionality reduction, the PCA was chosen because of its ability to automatically filter the random noise that could obscure the detection of original patterns in the data. Number of components used to describe the data was determined by looking at the cumulative explained variance ratio as a function of the number of components (Figure 4). The original features were distilled into four components using this form of dimensionality reduction.

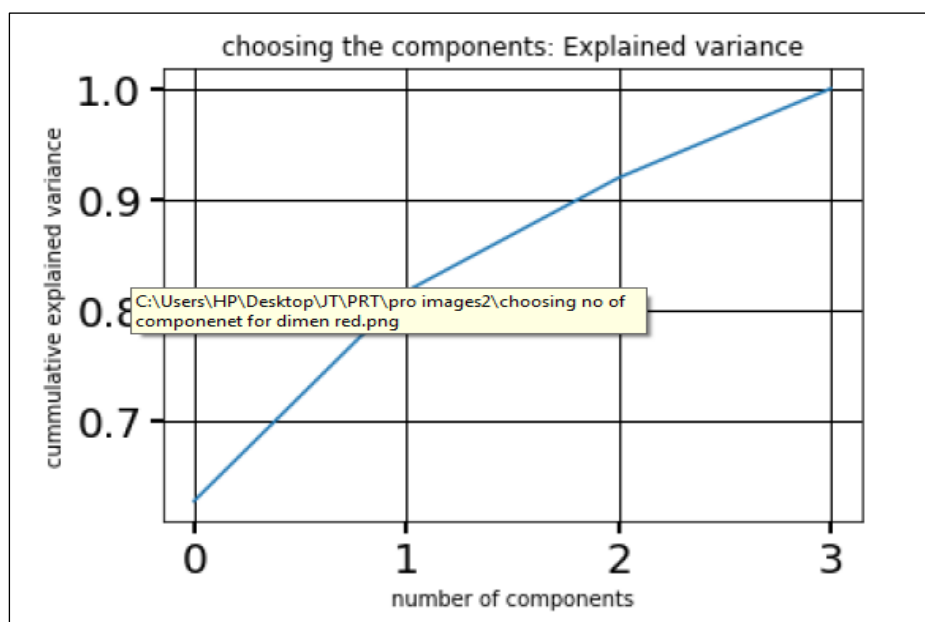


Figure 4: Plot of Cumulative Explained Variance

Table 6: The Methodology of the K-means Algorithm

S/N	Steps
1	Specify the number of K to decide the number of clusters
2	Randomly initialize k points (centroids)
3	Expectations: Assign each point to its closest centroid which will from the predefined K clusters
4	Maximization: Compute the variance and select a new centroid (mean) for each cluster
5	Repeat steps (4) and (5) until the centroid positions do not change

K-means Clustering

The main element of the k-means clustering algorithm works by a two-step process called expectation-maximization. The expectation step assigns each data point to its nearest centroid. Then, the maximization step computes the mean of all the points for each cluster and sets the new centroid. The conventional methodology of the k-means algorithm is shown in Table 6.

One of the challenges of using the k-means is its random initialization step. It causes the k-means algorithm to be nondeterministic. To make up for this deficiency, k-means++ was used for the initialization of the centroids. It specifies a procedure to initialize centroids before moving forward with the standard k-means. Table 7 shows the corrections applied to compensate for the limitations of the K-means algorithm.

Table 7: The Summary of Corrections for the Deficiencies of the K-means Algorithm

S/N	Deficiency	Correction
1	Random initialization	K-means ++
2	Skewed Distribution, complex shape distributions	Logarithmic Transformation
3	Outliers	Filtering (band pass)
4	Random Noise	Principal Component Analysis (PCA) to increase the signal/noise ratio
5	Specifying the optimum value for k	Combination of Domain knowledge and Evaluation metric (the elbow plot)

The evaluation metric, Elbow method, and domain knowledge were used to determine the appropriate number of cluster for the k-means algorithm.

For the elbow method, a graph was plotted (elbow curves) where the x-axis represents the number of clusters, and the y-axis represents an evaluation metric called inertias (Figure 5). The point of inflection on the curve where the inertia curve starts to bend is known as the elbow point. The x-value at this point is thought to be a reasonable trade-off between error and appropriate number of clusters. From the elbow plot, good values for k could be three, four or five. Eventually, domain knowledge was used to select the optimum value of k to be four.

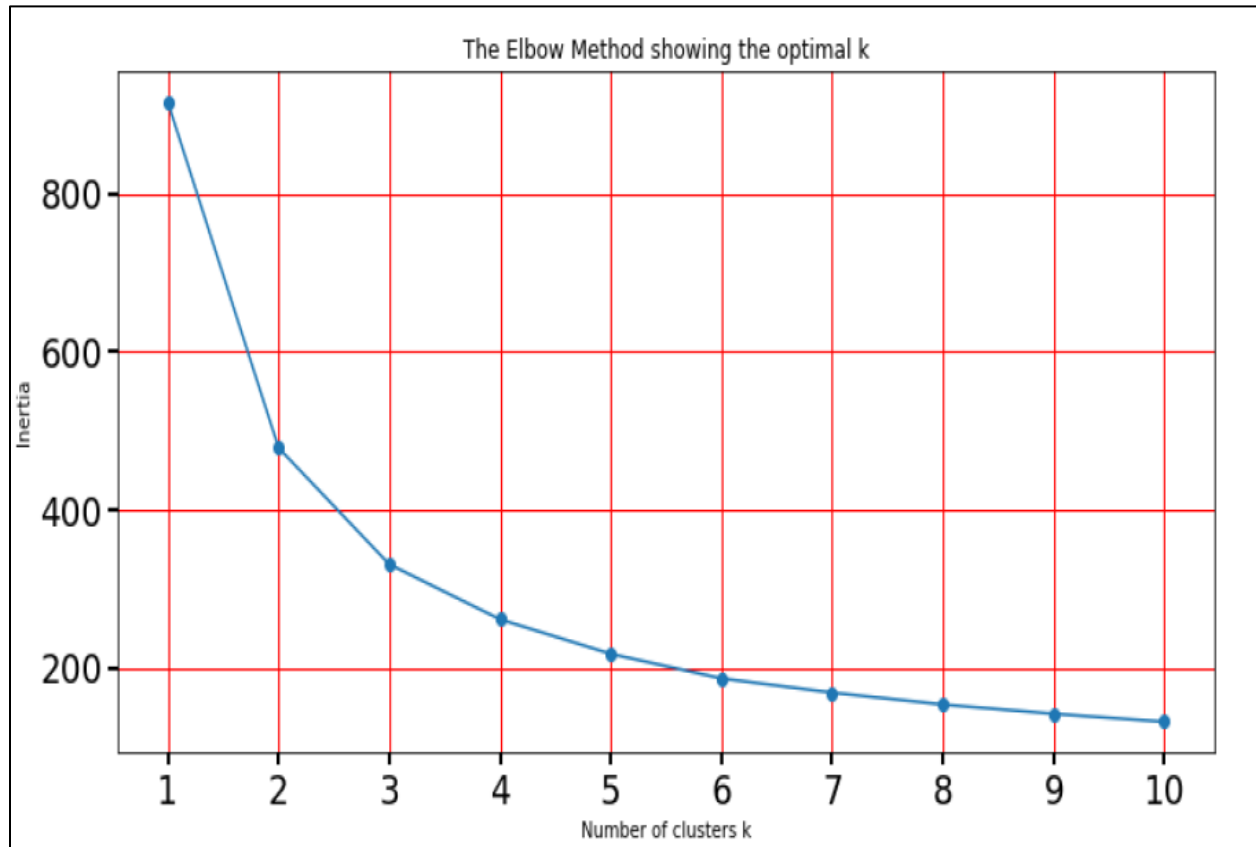


Figure 5: The Elbow plot

Results

The results of the k-means clustering are presented in cluster-plot, table, bar-plot and log-plots.

There were four distinct clusters found by the k-means clustering algorithm in the well data. They are shown on the cluster plot (Figure 6) with their centroids (centers) indicated by the big black dots. The gamma ray log is plotted on the horizontal axis and the resistivity log on the vertical axis. The centroid (center) of the blue cluster has relatively the highest resistivity reading and lowest gamma ray reading. Therefore, it was interpreted as the reservoir zones. The centroid of the red cluster has resistivity value lower than the blue but higher than the remaining cluster. The gamma ray reading is also the second highest. It was interpreted as shaly sand containing fresh formation water. After considering the relative position of the centroid of the yellow cluster, it was interpreted as shaly sand containing brine. The last cluster (Purple color) has centroid having the least resistivity reading and the highest gamma ray value. As a result, it was interpreted as shale.

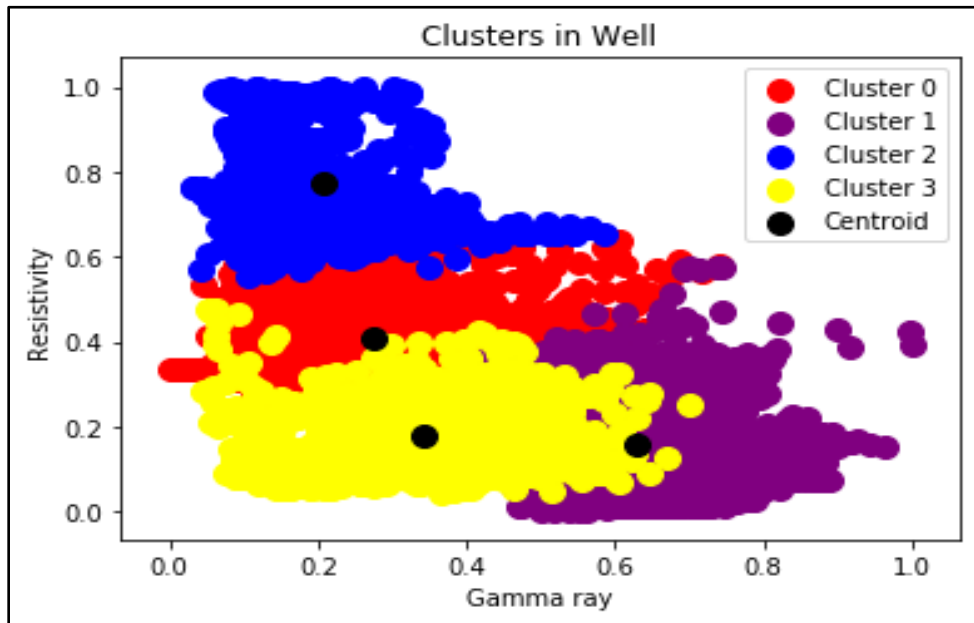


Figure 6: Plot of Clusters with the Centers

Table 8 contains the four different groups (clusters) found by the algorithm in the well data. It summarizes how the interpretation was done. The dominant lithologies found were sands and shales with gradational lithologies like sandy shales or shaly sands. Clean sands with good porosity and sufficient thickness are considered a good reservoir.

Table 8: Interpretation of the Well Logs

S/N	Well log observation	Interpretation
1.	Low gamma, medium density, medium to high velocity, low resistivity.	Porous shaly sand containing brine
2.	High gamma, highest density, high velocity, low resistivity	Shale bed
3.	Lowest gamma, low density, low velocity, high resistivity	Sands with good porosity/ clean sand with oil as the pore fluid (blue-color cluster)
4.	Low to medium gamma, low to high density, low to high velocity, low to medium resistivity.	Shaly-sand containing water

The bar plot of the counts of observation in each of the cluster was made (Figure 7). The bar plot of the clusters shows the respective number of sampling points and color to distinguish them from one another and is the key to understanding what each cluster represents geologically: The color corresponding to cluster label 2.0 is the hydrocarbon reservoirs (i.e. Blue) which constitute the zone of interest for this study. Other lithologic units that are identifiable which include: the shale beds represented by the color corresponding to the cluster label 1.0 on the bar plot. The one

Application of K-Means Clustering for Hydrocarbon Reservoir Delineation and Reservoir Parameter Estimation

corresponding to cluster label 3.0 represent shaly-sand containing brine, while the one corresponding to cluster label 0.0 is a water-saturated shaly-sand.

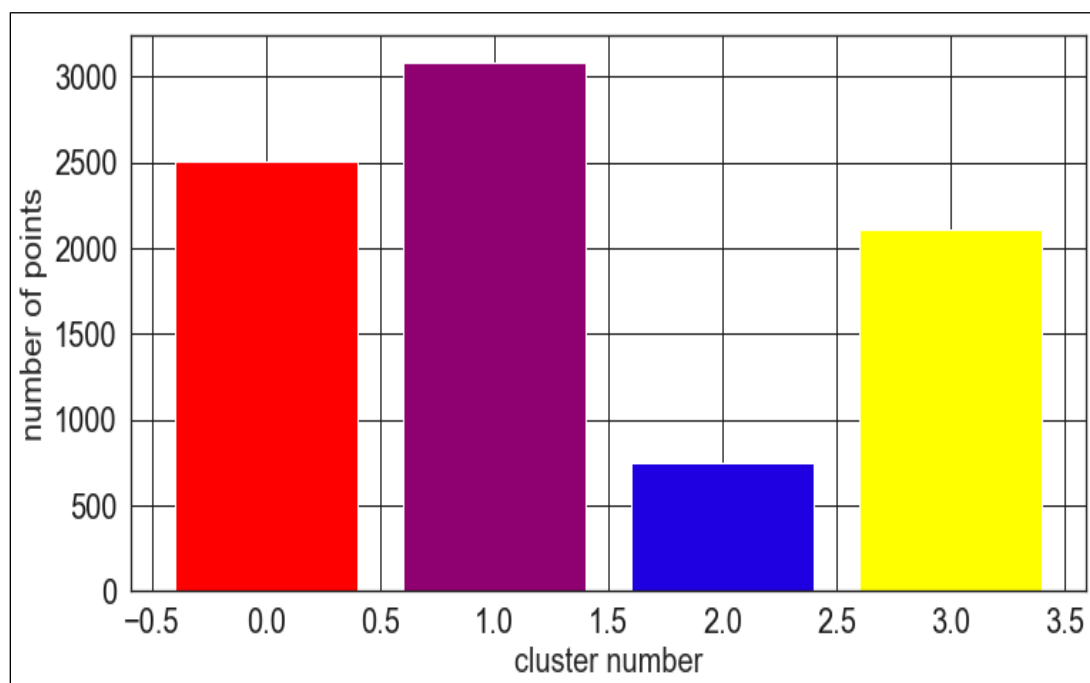


Figure 7: Bar-plot of Cluster Distribution

The above process of giving geological meaningful name to each of the cluster labels was written into code so that the whole process is automated. The program will dynamically generate the bar plot and output the cluster label corresponding to the reservoir hydrocarbon-bearing zone. The user will check for the color corresponding to that particular label on the bar plot. This is the color that will be used to interpret the cluster blocks on the log plot. The color changes for every run of the program but the output of the program will always guide the user as to what color represents the hydrocarbon reservoirs.

The log plot is a plot of the clustering algorithm labels plotted alongside the well log for hydrocarbon reservoirs A- F (Figure 8). The variables shown in the figure are the different well logs and the corresponding cluster blocks. The white portion in some of the reservoir zones indicate a neutral zone (zones that do not belong to any of the clusters). One of the applications of the k-means clustering is outlier detection (or anomaly detection). Had it been that the data were not filtered to remove the outliers, the algorithm might have clustered those points into another distinct group and this will affect the clustering process in some ways. Hence, the intervals were indicated with a color that is different from the colors of the clusters to show that the observations have been removed through filtering. But from domain knowledge, those neutral points also belong to the reservoir zones. Hence, domain knowledge is needed in the effective applications of machine learning solutions to geophysical problems.

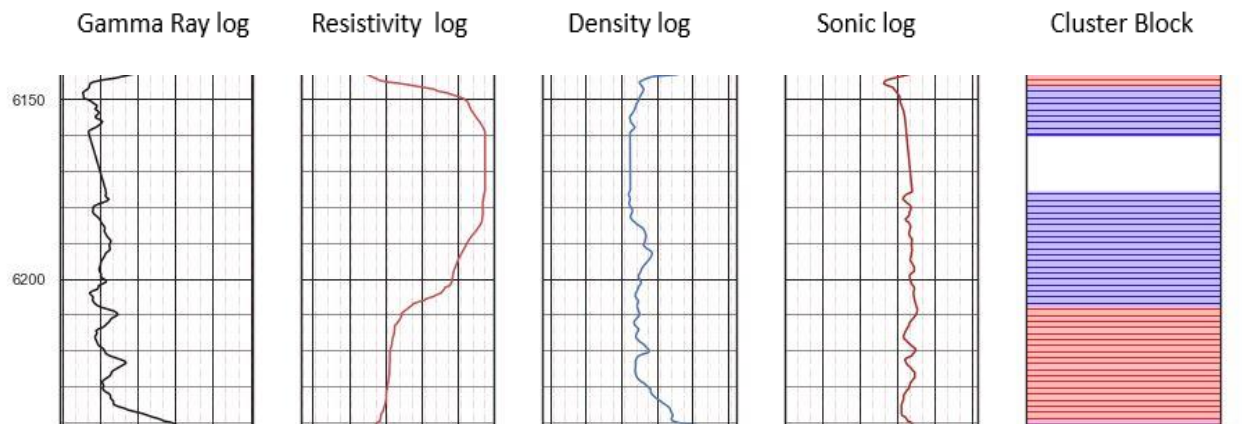
Depth ranges predicted to be reservoirs by the k-means clustering model (i.e. the blue clusters) were manually interpreted to evaluate the performance of the model. It was found that the prediction of the algorithm was right in all the cases considered.

Reservoir Parameter Estimation

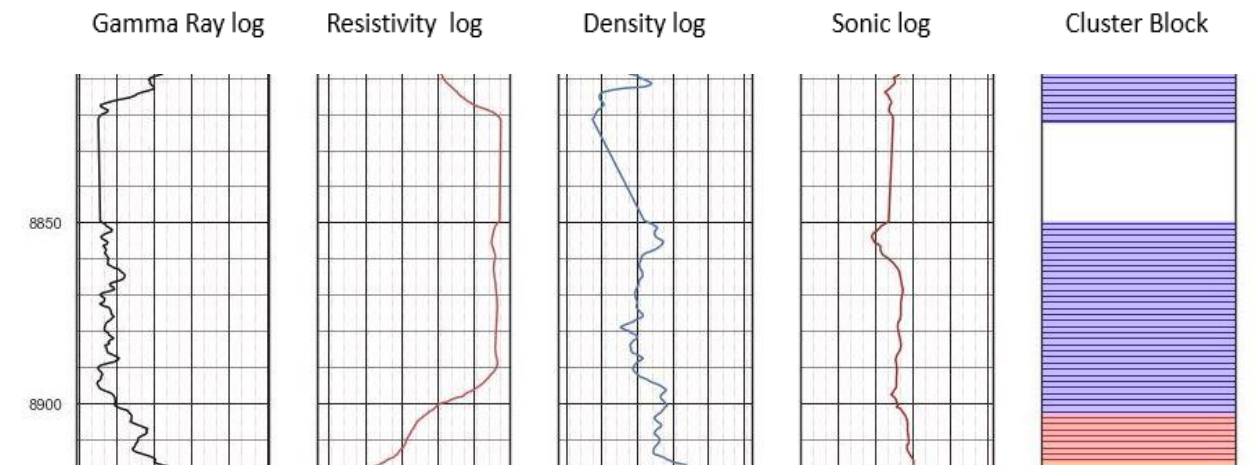
The gross thickness of the reservoir ranges from 13.42m to 29.28m. The parameters of reservoir F could not be determined by the program because there was no nearby clean formation zone with approximately 100% water saturation. This is one of the limitations of the reservoir identification model. The reservoir parameters of the reservoirs are given in Table 9.

The sample of the output of the section of the program for reservoir parameter estimation is given in Figure 9.

Hydrocarbon Reservoir A:

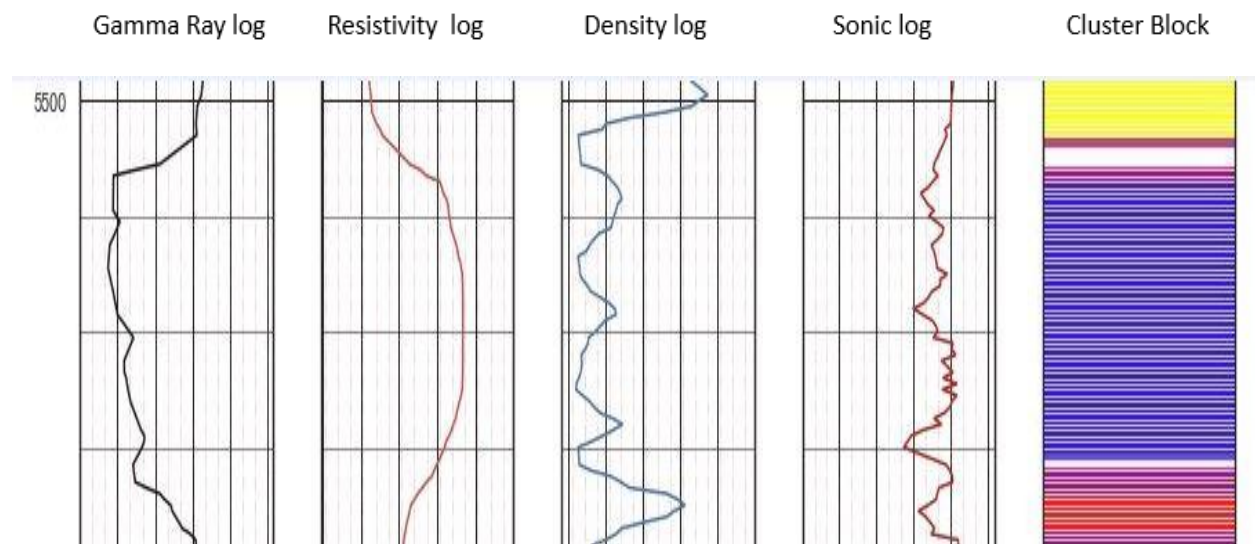


Hydrocarbon Reservoir B:

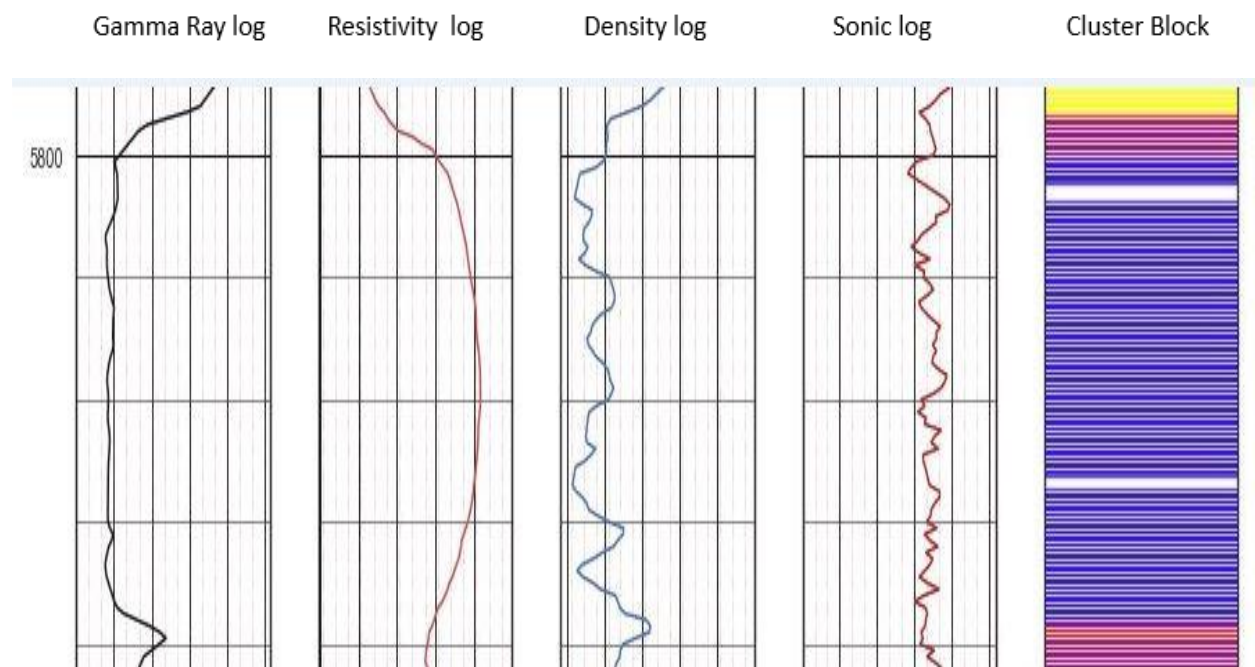


Application of K-Means Clustering for Hydrocarbon Reservoir Delineation and Reservoir Parameter Estimation

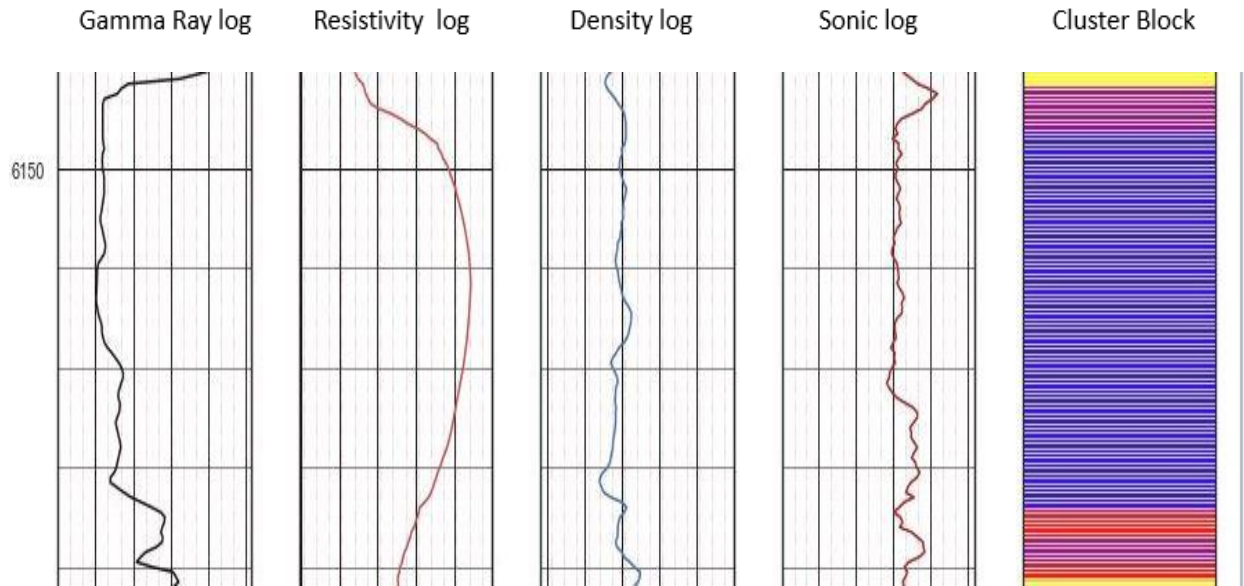
Hydrocarbon Reservoir C:



Hydrocarbon Reservoir D:



Hydrocarbon Reservoir E:



Hydrocarbon Reservoir F:

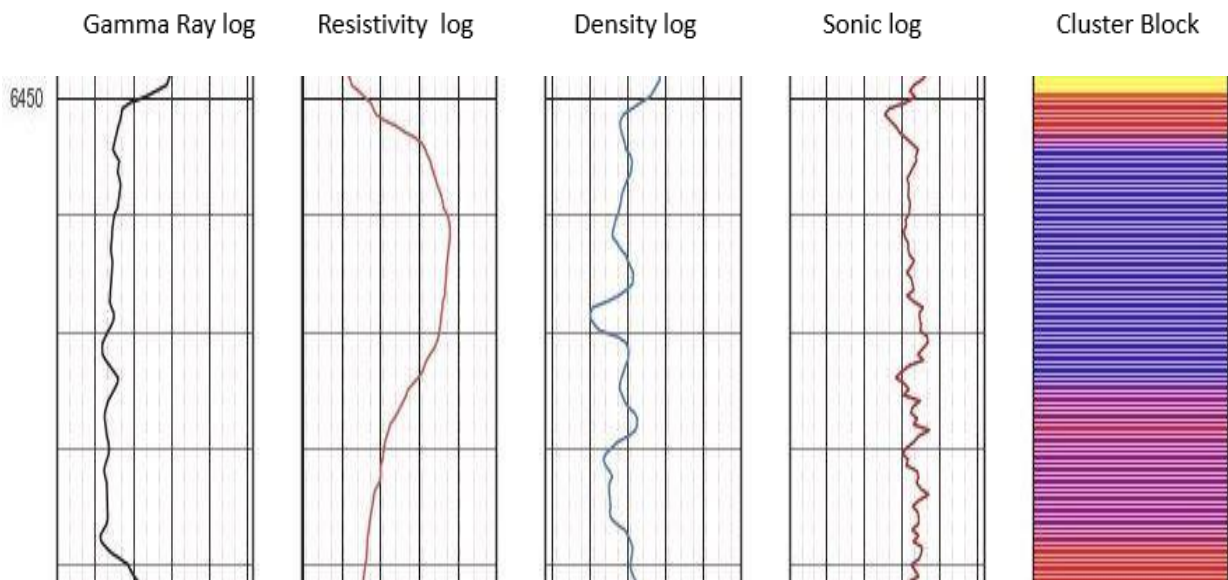


Figure 8: Log-plot of Wells Showing the Hydrocarbon Reservoirs (A-F)

Table 9: The Reservoir Parameters Estimation of some of the Reservoirs Delineated by the Clustering model.

Reservoir Name	Depth Interval (ft)	Gross Thickness (m)	Porosity (%)	Effective Porosity (%)	Water Saturation (%)	Hydrocarbon Saturation	Volume of Shale
A	6144-6240	29.28	30.18	28.71	13.20	86.80	4.88
B	8812-8908	29.28	35.75	34.80	15.25	84.75	2.67
C	5506-5535	8.85	51.56	49.64	36.33	63.67	3.72
D	5795-5839	13.42	51.56	49.04	23.54	76.46	4.88
E	6141-6190	14.95	35.75	33.92	17.63	82.37	5.12
F	6448-6492	13.42	—	—	—	—	—

```

-----
Enter bulk_density:1.825
Enter bulk_density:1.825
Enter Resistivity at 100% water saturation zone:4.239
Enter Resistivity at uninvaded zone:6.264
Enter GR of the point :25
*****
Porosity:51.5625%
Effective Porosity:49.642%
Water Saturation:36.331%
Hydrocarbon Saturation:63.669%
Volume of Shale:3.7247%
*****

```

Figure 9: Outputs of Reservoir Parameter Estimation.

Conclusion

The k-means clustering algorithm has been successfully deployed for automating hydrocarbon reservoir delineation and reservoir parameters estimation from well logs in an oil field 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 wells.

The model was able to delineate the hydrocarbon-bearing zone (blue-color cluster) of reservoirs in well A. The model was then applied to two other wells namely, Well B and Well C from the same field so as to evaluate its performance on different well data. The results were accurate as the machine learning model was able to automatically identify the hydrocarbon reservoirs in each of the wells. About twelve zones were delineated as potential hydrocarbon-bearing zones by the models. After the delineation of the reservoirs, reservoir parameter estimation was carried out for the identified reservoirs.

The significance of the developed model is in its advantages over the conventional method of well log interpretation. In the conventional way of delineating hydrocarbon bearing zones, the cut-off value between the shale and sand is usually specified beforehand for the gamma ray log. This value is not always the same for all wells, hence, the result is largely dependent on the experience of the interpreter. But for this new method, the model is able to learn the appropriate cut-off value by itself, that is, it does not have to be specified beforehand. This means that the clustering model is flexible and can adjust to the underlying structure of the dataset. Since the hydrocarbon reservoirs are automatically delineated, well log analysis proceed at a faster rate and relatively easier as it requires minimal effort from the well log analysts.