

KLASIFIKASI LITHOLOGI BERDASARKAN DISTRIBUSI STATISTIK HIRARKIS (*HIERARCHICAL STATISTICAL DISTRIBUTION (HSD)*)

Firman Syaifuddin¹, Eki Komara¹, M Haris Miftakhul Fajar¹, Abdillah¹

Afiliasi : ¹Departemen Teknik Geofisika, Fakultas Teknik Sipil Perencanaan dan Kebumihan-Institut Teknologi Sepuluh Nopember Surabaya

e-mail : firmansyaifuddinakbar@gmail.com/ firman.syaifuddin@its.ac.id

Abstrak. Karakterisasi reservoir merupakan proses fundamental dalam industri minyak dan gas, namun identifikasi litologi yang akurat sering kali terhambat oleh kompleksitas geologi dan sifat 'kotak hitam' dari model *machine learning* (ML) modern. Penelitian ini menyajikan metode alternatif yang tidak disupervisi untuk klasifikasi litologi berbasis distribusi statistik hierarkis sebagai solusi untuk meningkatkan transparansi dan interpretasi. Metode ini menggunakan data well log utama gamma-ray, sonic slowness, densitas, dan resistivitas dari enam sumur di area geologi deltaik yang stabil. Pendekatan hierarkis ini secara sistematis mengklasifikasikan lima litologi utama (batubara, serpih, batugamping, batupasir basah, dan batupasir hidrokarbon) dengan menerapkan ambang batas kuantil yang ditentukan dari distribusi statistik setiap log. Hierarki dimulai dengan identifikasi serpih menggunakan log gamma-ray, diikuti oleh batubara dengan log densitas, batu gamping dengan log slowness, dan terakhir, batupasir diklasifikasikan berdasarkan log resistivitas. Hasil penelitian menunjukkan bahwa metode ini berhasil mengklasifikasikan litologi dengan distribusi serpih (66%), batubara (5%), batu gamping (0.63%), dan pasir (27%). Secara kualitatif, hasil ini dianggap lebih "masuk akal secara geologis" jika dibandingkan dengan metode KNN, karena metode hierarkis ini didasarkan pada prinsip-prinsip fisika batuan yang eksplisit, memungkinkan interpretasi dan validasi yang mudah oleh geosaintis. Sebagai metode yang adaptif terhadap distribusi data, pendekatan ini menawarkan alternatif yang transparan dan dapat dipercaya untuk karakterisasi litologi.

Kata Kunci: Distribusi Statistik; Hierarki; Identifikasi Litologi; *Machine Learning* (ML); Metode tidak disupervisi

Abstract. *Reservoir characterization is a fundamental process in the oil and gas industry, but accurate lithology identification is often hampered by geological complexity and the 'black box' nature of modern machine learning (ML) models. This study presents an alternative unsupervised method for lithology classification based on hierarchical statistical distributions as a solution to improve transparency and interpretation. The method uses primary well log data of gamma-ray, sonic slowness, density, and resistivity from six wells in a stable deltaic geology area. This hierarchical approach systematically classifies five primary lithologies (coal, shale, limestone, wet sand, and hydrocarbon sand) by applying quantile thresholds determined from the statistical distribution of each log. The hierarchy begins with shale identification using GAMMA-RAY logs, followed by coal with density logs, limestone with slowness logs, and finally, sand classified based on resistivity logs. The results show that this method successfully classifies lithologies with a distribution of shale (66%), coal (5%), limestone (0.63%), and sand (27%). Qualitatively, these results are considered more "geologically plausible" compared to the KNN method, as this hierarchical method is based on explicit rock physics principles, allowing for easy interpretation and validation by geoscientists. As a method that adapts to data distribution, this approach offers a transparent and reliable alternative for lithology characterization.*

Keywords: *Statistical Distribution; Hierarchy; Lithology Identification; Machine Learning (ML); Unsupervised Method*

PENDAHULUAN

Karakterisasi reservoir merupakan proses multidisiplin yang sangat krusial dalam industri minyak dan gas, mencakup seluruh siklus hidup suatu lapangan, mulai dari tahap eksplorasi hingga eksploitasi dan pengembangan (Bedle dkk., 2024; Esiri dkk., 2024; Solanke dkk., 2024). Tujuan utama dari proses ini adalah untuk membangun model bawah permukaan yang akurat, baik secara statis maupun dinamis, yang memungkinkan prediksi cadangan hidrokarbon, optimalisasi kinerja produksi, dan pengambilan keputusan investasi yang tepat. Dengan tantangan global yang semakin kompleks dalam mencari sumber daya, terutama di reservoir yang sangat heterogen, kebutuhan akan model yang akurat menjadi semakin mendesak.

Litologi secara langsung mengontrol properti petrofisika utama yang menentukan kualitas reservoir (Doveton,1994). Porositas, yang merupakan volume ruang pori dalam batuan, sangat dipengaruhi oleh jenis batuan. Setiap jenis batuan memiliki karakteristik porositas yang unik. Batupasir, yang merupakan batuan reservoir paling umum, umumnya memiliki porositas primer yang terbentuk selama pengendapan. Sementara itu, batu gamping memiliki porositas yang lebih bervariasi karena dipengaruhi oleh proses diagenesis. Serpih (batulempung), di sisi lain, memiliki porositas yang sangat rendah dan sering kali tidak dianggap sebagai reservoir konvensional.

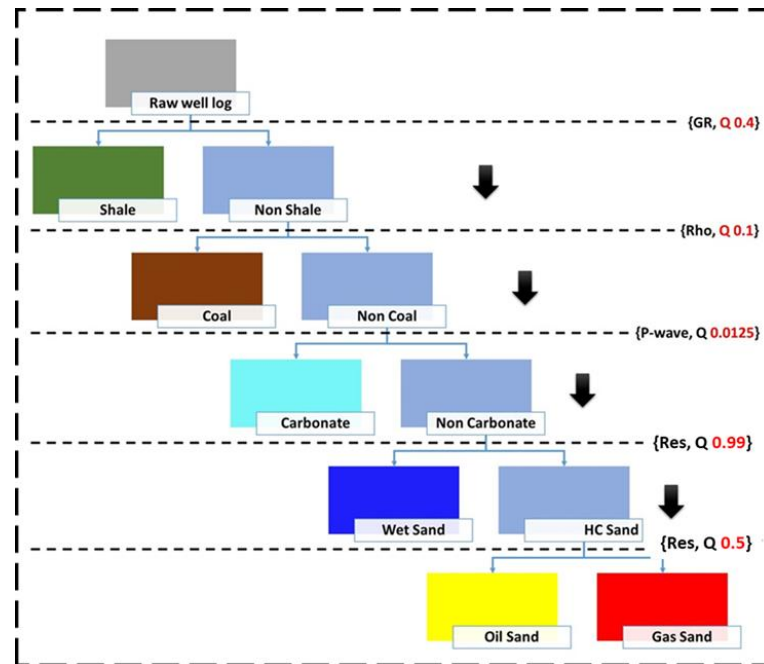
Identifikasi litologi, atau penentuan jenis batuan, adalah fondasi fundamental dari seluruh proses karakterisasi reservoir (Doveton,1994). Tanpa pemahaman yang tepat mengenai litologi, interpretasi parameter petrofisika lainnya seperti porositas, permeabilitas, dan saturasi fluida akan menjadi tidak valid, yang pada akhirnya dapat mengarah pada kesalahan besar dalam estimasi cadangan dan strategi produksi. Pentingnya litologi tidak dapat diremehkan, karena ia secara langsung memengaruhi kemampuan batuan untuk menyimpan dan mengalirkan fluida. Seiring dengan perkembangan pesat dalam teknologi analisis data, terutama di bidang *machine learning* (ML), metode-metode baru telah muncul yang menjanjikan untuk mengatasi keterbatasan inheren dari pendekatan konvensional (Casado-Vara dkk., 2025).

Keterbatasan ini menunjukkan bahwa solusi masa depan dalam identifikasi litologi harus berfokus tidak hanya pada peningkatan akurasi, tetapi juga pada interpretasi, dengan menggunakan pendekatan seperti *Explainable AI* (XAI) untuk membangun kepercayaan dan transparansi di antara para profesional. Klasifikasi litologi berbasis *machine learning* sering bergantung pada model *machine learning* kompleks yang kurang transparan. Meskipun *machine learning* (ML) telah menunjukkan akurasi yang lebih tinggi dalam identifikasi litologi dibandingkan metode konvensional, aplikasi model-model ini, terutama *deep learning*, menghadapi tantangan signifikan terkait transparansi (Senjoba dkk., 2022; Li, 2024; Zhang dkk., 2024). Dalam dunia geosains, di mana kepercayaan pada hasil interpretasi sangat penting, model-model ini sering disebut sebagai "kotak hitam" (*black box*) karena sulitnya memahami bagaimana mereka sampai pada prediksinya. Sebuah studi tentang identifikasi litologi di *buried hill reservoirs* menyoroti bahwa meskipun model *deep learning* dapat menangkap hubungan non-linear yang kompleks, mereka rentan terhadap overfitting dan memiliki interpretasi yang buruk, yang membatasi penerapan praktisnya di lapangan (Zhao dan Liao, 2025). Pada penelitian lain dalam bidang *Explainable AI* (XAI) untuk geosains menjelaskan bahwa XAI berfungsi sebagai "lensa pembesar" yang mengungkapkan alasan di balik prediksi model (Jinad dan Shashidhar, 2024; Garouani dkk., 2024). Studi ini menegaskan bahwa model *deep learning* sering kali berlabel "kotak hitam" karena mereka tidak dapat menjelaskan bagaimana mereka mencapai kesimpulan (Garouani dkk., 2024).

Penelitian ini mengusulkan metode alternatif berbasis distribusi statistik dan hirarki keputusan untuk mengidentifikasi lima litologi utama (batubara, serpih, batugamping, batupasir basah, batupasir hidrokarbon) menggunakan data well log utama, log gamma-ray, *sonic slowness* gelombang P (Δt), densitas (ρ), dan resistivitas (R_d). Pendekatan hierarkis memanfaatkan karakteristik statistik (distribusi, *quantile*, ambang batas) dari log gamma-ray, *slowness* (Δt), densitas (ρ), dan resistivitas (R_t). Setiap tahap hierarki menerapkan kriteria statistik yang terinspirasi dari sifat fisika batuan.

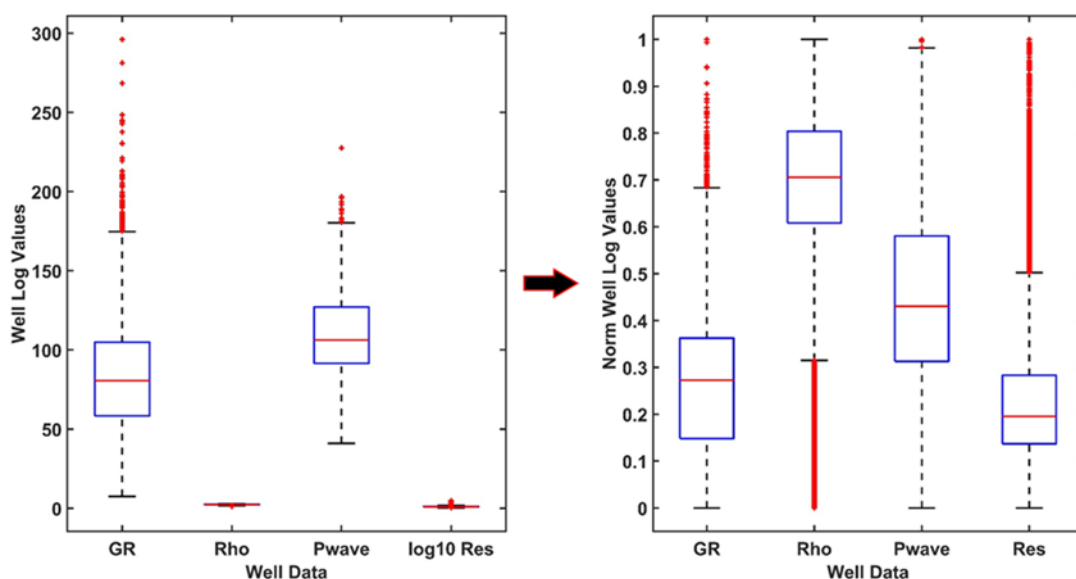
METODOLOGI

Pada dasarnya identifikasi litologi dilakukan berdasarkan urutan tertentu atau secara hirarki mengidentifikasi jenis batuan atau litologi berdasarkan *log response* masing-masing seperti ditunjukkan pada Gambar 1. Pada penelitian ini data sumur yang digunakan sebanyak 6 sumur yang memiliki log gamma-ray, densitas, *sonic slowness* atau p-wave dan resistivitas.

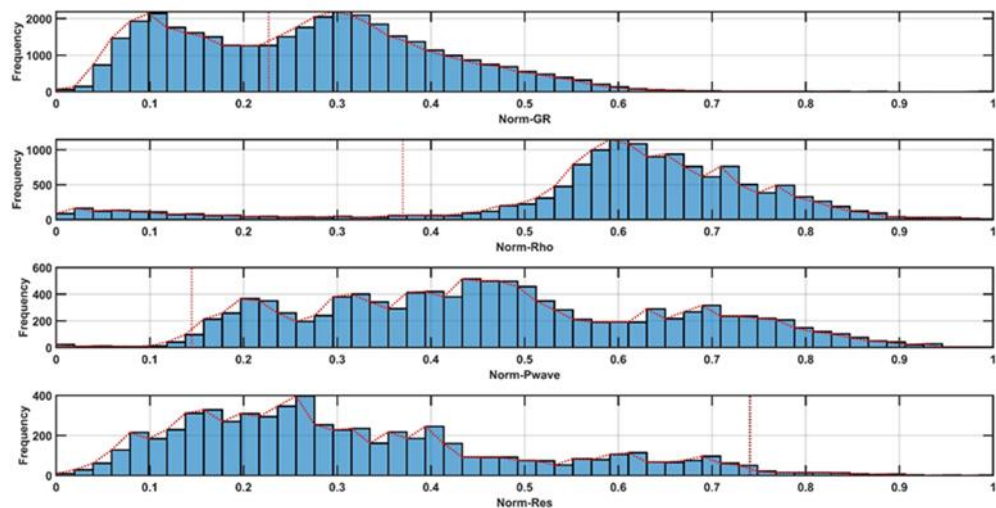


Gambar 1. Diagramma-rayam alir identifikasi litologi secara hirarki berdasarkan log Gamma-ray, Densitas, Sonic, dan Resistivitas

Pada studi ini digunakan data 6 sumur yang berada pada lingkungan geologi deltaic dan secara tektonik merupakan daerah yang cukup stabil. Data well log utama yang digunakan adalah log gamma-ray, Sonic slowness gelombang P (Δt), densitas (ρ), dan resistivitas (Res). Pada studi ini litologi dibagi menjadi 6 litologi utama yaitu: serpih (*shale*), batubara (*coal*), batugamping (*carbonate*), batupasir jenuh air (*wet sand*), batupasir jenuh minyak (*oil sand*) dan batupasir jenuh gas (*gas sand*). Karena rentang data bacaan log masing-masing sumur memiliki rentang yang bervariasi, misalnya pada log gamma-ray ada yang rentangnya 0-150 dan 0-250, maka dilakukan normalisasi dengan rentang 0 sampai 1 pada masing-masing log pada masing-masing sumur seperti ditunjukkan pada boxplot Gambar 2 dan histogram distribusi masing-masing log pada semua sumur ditunjukkan pada Gambar 3.

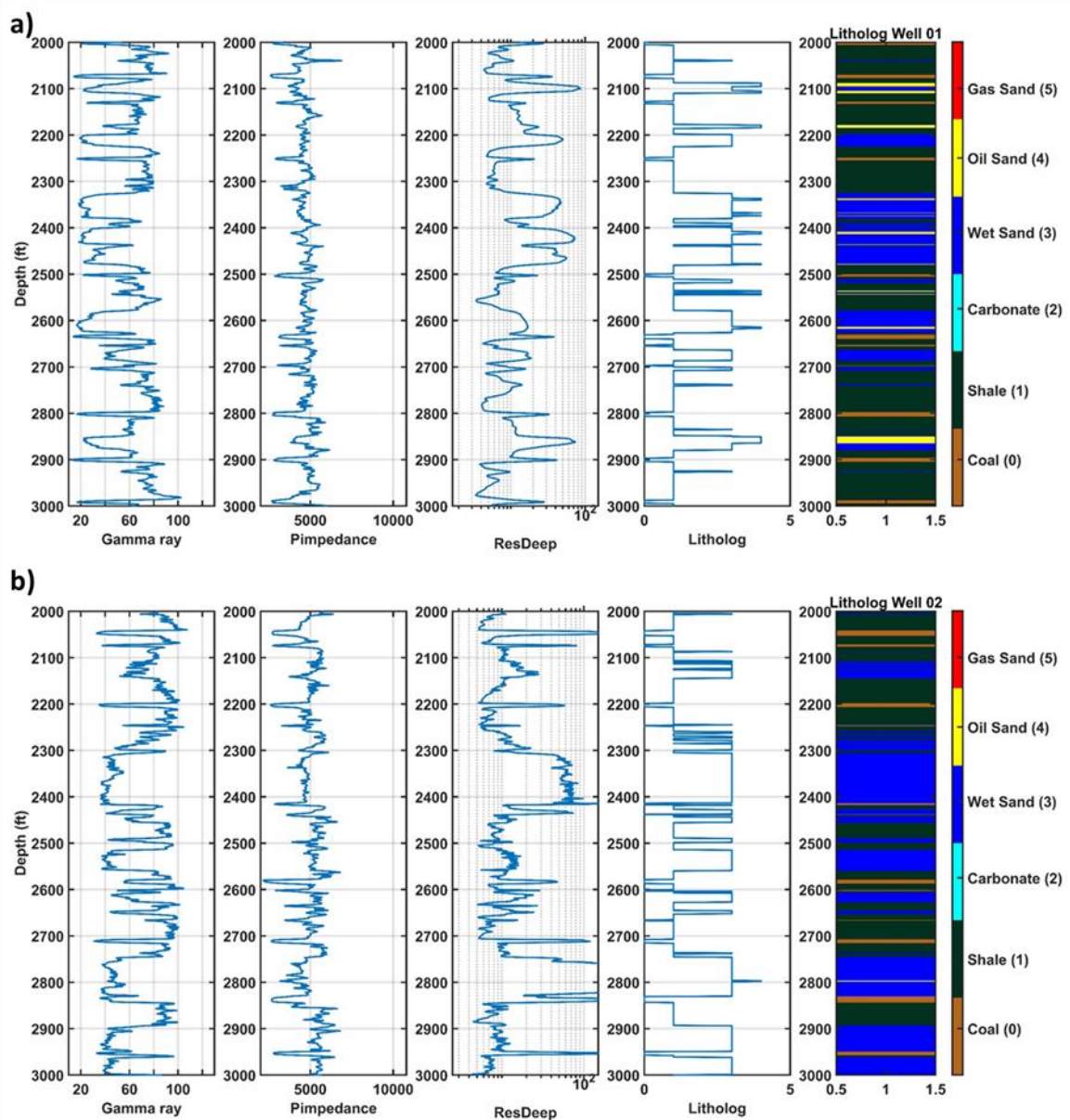


Gambar 2. Data normalisasi masing-masing log



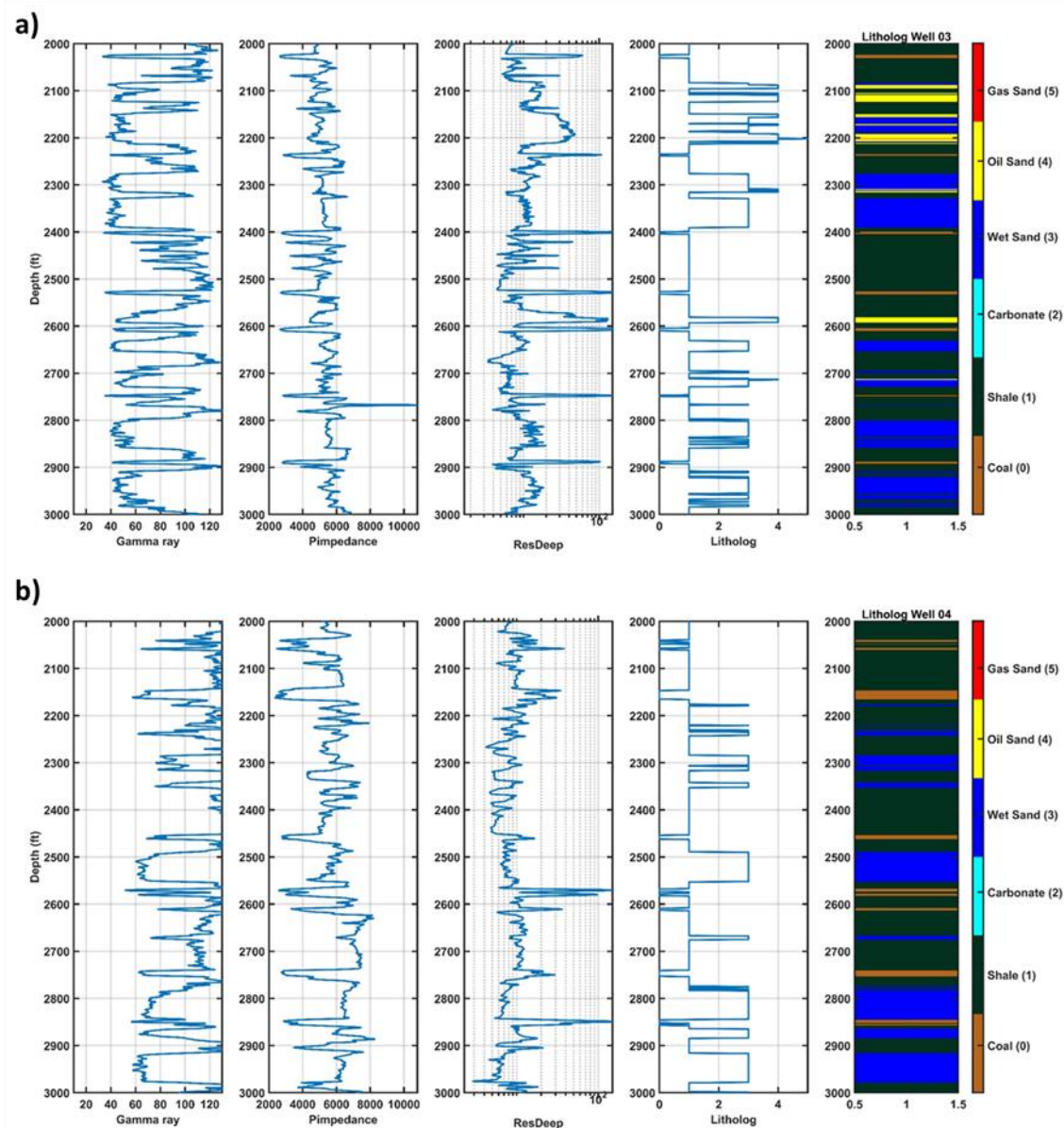
Gambar 3. Histogram masing-masing log pada semua sumur

Hierarki dimulai dengan litologi yang memiliki *respons* log paling khas yaitu batuan serpih, dimana merupakan jenis litologi dominan hampir disemua wilayah. Untuk memisahkan batuan serpih dan bukan serpih dipisahkan berdasarkan ambang batas yang di tentukan mengacu pada histogram distribusi log gamma-ray ternormalisasi, pada studi ini ambang batas untuk memisahkan batuan serpih dan bukan serpih yang digunakan adalah nilai *quantile* 0.4. Proses selanjutnya adalah memisahkan batubara dari batuan bukan serpih dengan memanfaatkan log densitas, dimana batubara memiliki nilai densitas yang sangat rendah dibandingkan jenis batuan yang lainnya. Mengacu histogram distribusi log densitas, nilai ambang batas yang digunakan adalah *quantile* 0.1, sehingga nilai densitas ternormalisasi batuan bukan serpih diatas ambang batas dapat di klasifikasikan batuan bukan batubara. Kemudian pada batuan bukan batubara diterapkan ambang batas log *sonic slowness* gelombang P dengan ambang batas *quantile* 0.0125 untuk memisahkan batugamping, dimana batugamping memiliki *sonic slowness* gelombang P yang kecil dibandingkan batuan yang lain. Pada hasil pemisahan ini yang tersisa adalah jenis batuan pasir, dimana dapat diklasifikasikan menjadi 3 yaitu batupasir jenuh air, jenuh minyak dan jenuh gas berdasarkan nilai resistivitasnya. Klasifikasi selanjutnya dilakukan berdasarkan nilai log resistivitas, dimana batupasir jenuh air memiliki resistivitas lebih kecil dibandingkan batupasir jenuh hidrokarbon (baik minyak maupun gas). Nilai ambang batas *log* resistivitas yang digunakan adalah *quantile* 0.99 untuk memisahkan yang batupasir jenuh air, dan nilai ambang batas *quantile* 0.5 untuk memisahkan batupasir jenuh minyak dan jenuh gas. Pada Gambar 4, Gambar 5 dan Gambar 6 ditunjukan hasil klasifikasi litologi pada masing-masing sumur.



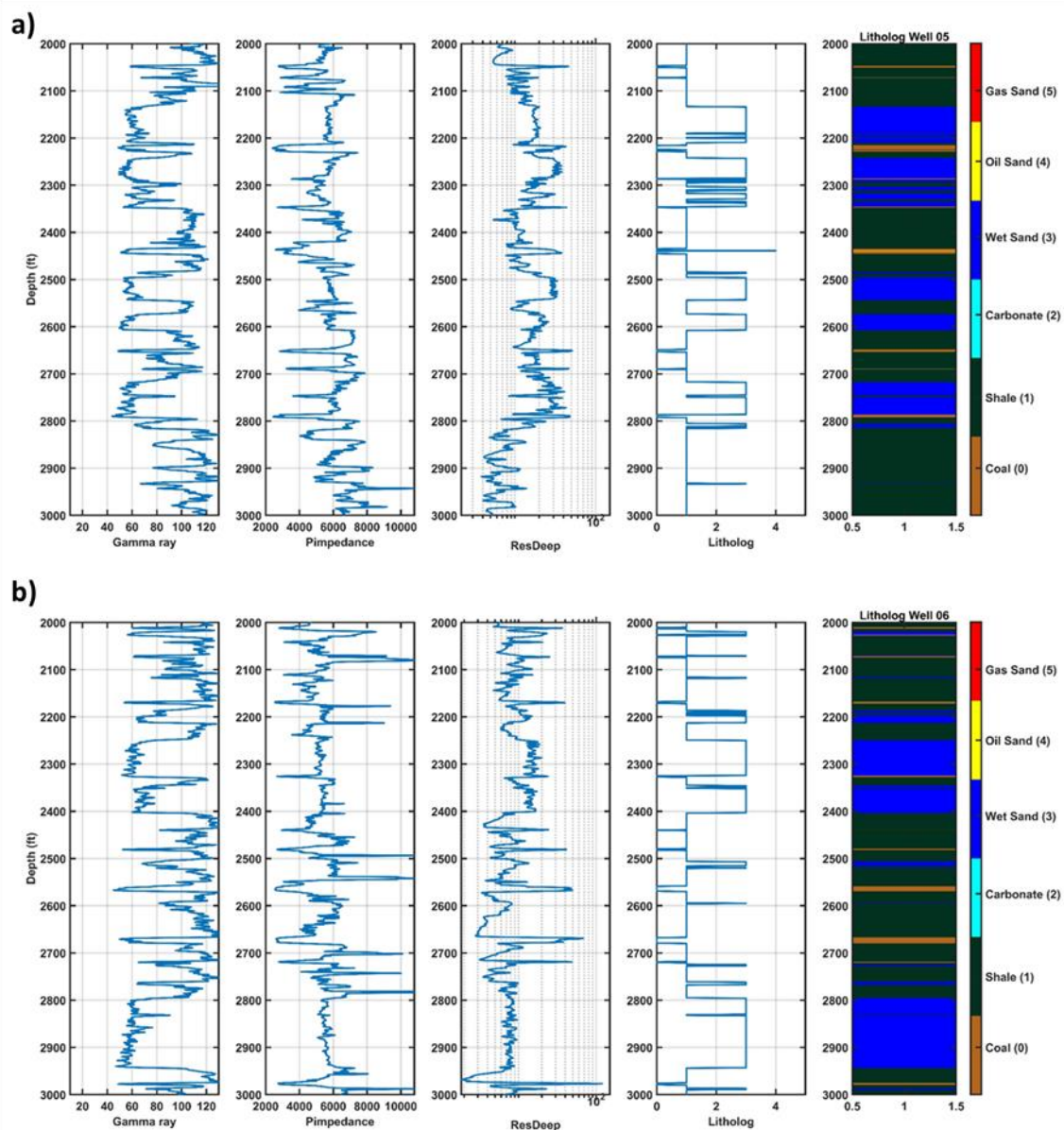
Gambar 4. Hasil klasifikasi litologi a) sumur 01, b) sumur 02

Pada Gambar 4 a) sumur 01, litologi yang teridentifikasi antarlain batubara 13%, serpih 57%, batupasir jenuh air 25%, batupasir jenuh minyak 4% dan keberadaan batugamping yang sangat kecil. Pada Gambar 4 b) sumur 02, litologi yang teridentifikasi antarlain batubara 5%, serpih 63%, batugamping 2%, batupasir jenuh air 30% dan keberadaan batupasir jenuh minyak yang sangat kecil.



Gambar 5. Hasil klasifikasi litologi a) sumur 03, b) sumur 04

Pada Gambar 5 a) sumur 03, litologi yang teridentifikasi antarlain batubara 2%, serpih 74%, batupasir jenuh air 20%, batupasir jenuh minyak 3% dan keberadaan batugamping yang sangat kecil. Pada Gambar 5 b) sumur 04, litologi yang teridentifikasi antarlain batubara 3%, serpih 66%, batugamping kurang dari 1%, batupasir jenuh air 29% dan keberadaan batupasir jenuh minyak yang sangat kecil kurang dari 1%.



Gambar 6. Hasil klasifikasi litologi a) sumur 05, b) sumur 06

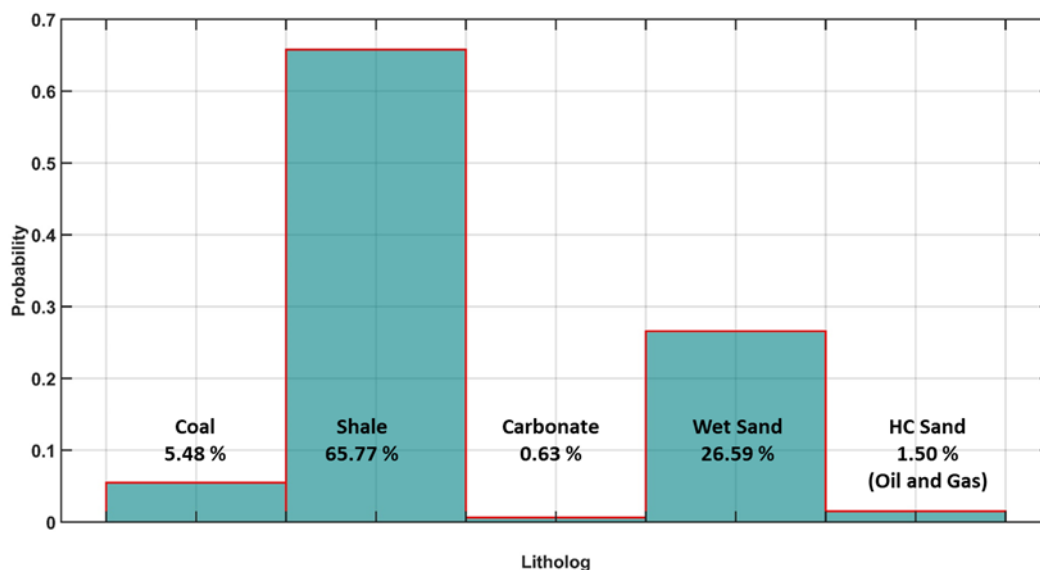
Pada Gambar 6 a) sumur 05, litologi yang teridentifikasi antarlain batubara 2%, serpih 70%, batugamping kurang dari 1%, batupasir jenuh air 26%, batupasir jenuh minyak 4% dan keberadaan batupasir jenuh minyak yang sangat kecil. Pada Gambar 6 b) sumur 06, litologi yang teridentifikasi antarlain batubara 6%, serpih 66%, batugamping kurang dari 1%, batupasir jenuh air 28% dan tidak ada batupasir jenuh minyak maupun gas.

HASIL DAN PEMBAHASAN

Secara umum dari keenam sumur menunjukkan pola litologi sesuai dengan pola log gamma-ray, P-impedance dan log resistivitas, karena metode ini berbasis aturan fisika batuan yang eksplisit, seorang geologis dapat dengan mudah memahami mengapa suatu lapisan diklasifikasikan sebagai serpih yang memiliki nilai gamma-ray tinggi, batubara yang memiliki nilai gamma-ray rendah, densitas rendah dan resistivitas relatif tinggi, batu gamping yang memiliki *sonic slowness* rendah dan densitas tinggi serta batupasir hidrokarbon yang memiliki nilai gamma-ray rendah dan resistivitas tinggi (Doveton,1994). seperti ditunjukkan pada **tabel 1**. Dari data keenam sumur yang digunakan pada penelitian ini didapatkan hasil jenis litologi serpih sebanyak 66 %, batubara 5%, batugamping 0.63%, batupasir jenuh air 27%, dan batupasir jenuh hidrokarbon baik minyak dan gas sebesar 1.5 %, seperti ditunjukkan pada Gambar 7.

Tabel 1. Ringkasan sifat fisik batuan masing-masing litologi (Doveton,1994).

Litologi	Gamma Ray	Slowness	Densitas	Resistivitas
Batubara	Rendah	Tinggi	Rendah ($Q < 0.1$)	Tinggi
Serpih	Tinggi ($Q > 0.4$)	Rendah-Tinggi	Rendah-Tinggi	Rendah
Batu Kapur	Rendah	Rendah ($Q < 0.0125$)	Tinggi	Menengah
Pasir Basah	Rendah	Menengah	Menengah	Rendah ($Q < 0.99$)
Pasir Hidrokarbon	Rendah	Menengah	Menengah	Tinggi ($Q < 0.99$)

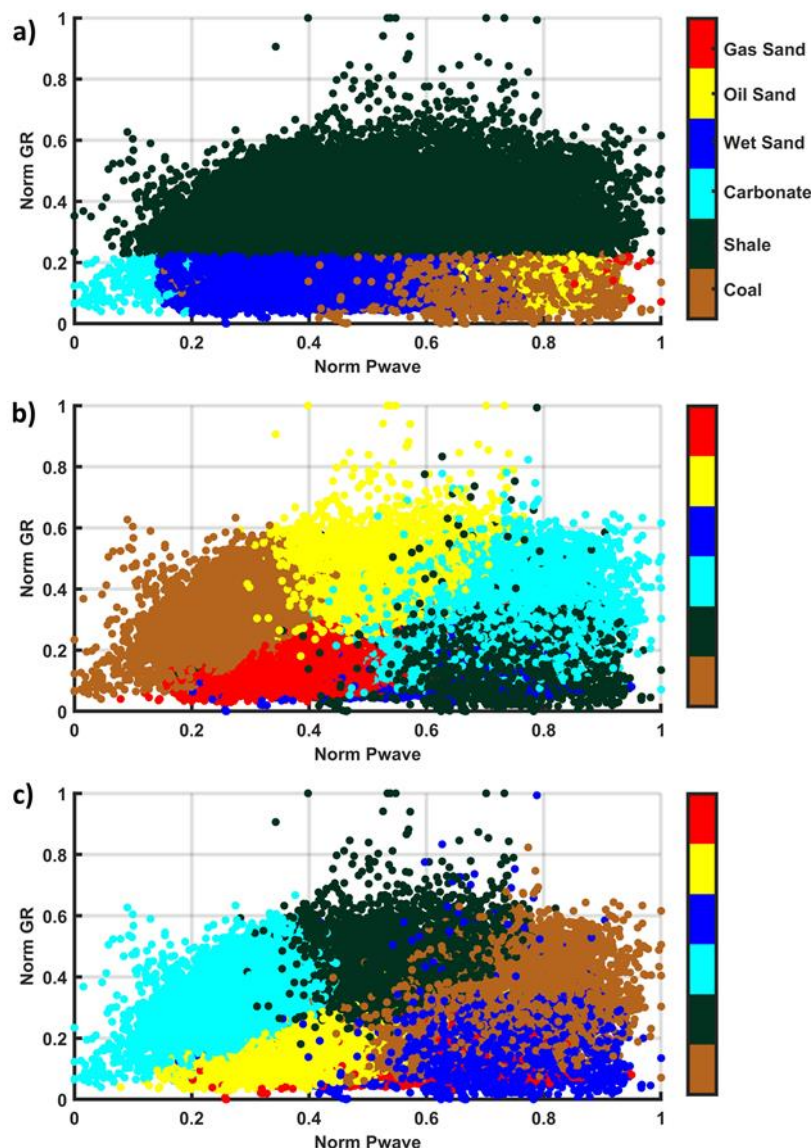


Gambar 7. Distribusi jenis litologi dari semua sumur

Metoda klasifikasi lithologi berdasarkan distribusi statistik hirarkis ini merupakan metoda tanpa supervisi (*unsupervise*) sehingga memerlukan pemahaman dan pengetahuan geologi yang cukup baik untuk menentukan ambang batas pada masing-masing parameter log. Selain itu juga terdapat ketergantungan pada kualitas log, sehingga perlu dilakukan normalisasi pada masing-masing log. Metode ini mungkin memerlukan kalibrasi untuk perhitungan persentil sebagai ambang batas jika diterapkan pada formasi geologis yang sangat berbeda, meskipun ini adalah fitur adaptabilitasnya yang sangat tergantung pada distribusi data. Pendekatan hirarki mengurangi kompleksitas dengan memprioritaskan log paling diskriminatif yaitu gamma-ray untuk serpih, dan densitas untuk batubara serta resistivitas untuk mengidentifikasi batupasir jenuh hidrokarbon.

Sebagai pembandingan dengan metoda berbasis *machine learning* sederhana, pada Gambar 8 ditunjukan crossplot antara log sonic slowness ternormalisasi dan log gamma-ray ternormalisasi metoda distribusi statistik hirarkis ini dengan metoda k-Nearest Neighbors (KNN) *mean* dan *median*. KNN adalah sebuah algoritma pembelajaran mesin (*machine learning*) yang berfungsi untuk klasifikasi dan prediksi data. Prinsip dasarnya adalah mengklasifikasikan suatu objek berdasarkan mayoritas kelas dari 'k' tetangga terdekatnya dalam sebuah dataset.

Dari crosplot tersebut secara kualitatif pendekatan distribusi statistik hirarkis ini lebih *geological plausible*, jika dibandingkan 2 metoda KNN yang ditunjukkan, dimana nilai gamma-ray menjadi pemisah tegas antara batuan serpih dan non serpih yang pada metoda distribusi statistik hirarkis ini merupakan log paling diskriminatif untuk membedakan batuan serpih dan non serpih.



Gambar 8. Crossplot antara log sonic slowness P-wave ternormalisasi dan log gamma-ray ternormalisasi metoda distribusi a) statistik hirarkis ini dengan metoda b) k-Nearest Neighbors (KNN) *mean* dan c) *median*

PENUTUP

Simpulan dan Saran

Dari data keenam sumur yang digunakan pada penelitian ini didapatkan hasil jenis litologi serpih sebanyak 66 %, batubara 5%, batugamping 0.63%, batupasir jenuh air 27%, dan batupasir jenuh hidrokarbon baik minyak dan gas sebesar 1.5 %. Metode hierarki statistik menawarkan solusi cepat, interpretatif, dan akurat untuk klasifikasi litologi dasar. Kombinasi analisis distribusi dan alur keputusan berjenjang mengatasi kelemahan pendekatan *machine learning (black-box)*. Kombinasi distribusi statistik dan hierarki keputusan menggunakan persentil untuk definisi ambang batas berbasis fisika batuan dan diterapkan diatur dalam alur pohon keputusan lebih *geological plausible* jika dibandingkan dengan *unsupervised machine learning* sederhana KNN.

Mengingat pentingnya kualitas data log, sangat disarankan agar data log dikondisikan sebelum klasifikasi litologi, terutama jika terdapat lubang bor yang bermasalah. Selanjutnya, untuk memperkuat hasil klasifikasi, kesesuaian dapat ditentukan dengan litologi yang digambarkan oleh cutting pengeboran. Dalam kasus litologi yang kompleks, seperti batuan vulkanik, kriteria tambahan harus dipertimbangkan untuk menentukan alur keputusan hierarkis.

DAFTAR PUSTAKA

- Bedle, H., Chopra, S., & Davis, T. (2024). Introduction to this special section: Reservoir characterization. The Leading Edge, 43(12), 798–799. <https://doi.org/10.1190/tle43120798.1>
- Casado-Vara, R., Martín-Luis, M. C., Casillas-Ruiz, R., Valero-Garcés, B. L., & Jambolina-Enríquez, M. (2025). A comparative machine learning study for lithology classification in volcano-fluvial deposits from La Vega de La Laguna (Tenerife, Canary Islands): Application to core lithology log using all-elements and major elements approach. Applied Computing and Geosciences, 27, 100273. <https://doi.org/10.1016/j.acags.2025.100273>
- Doveton, J. H. (n.d.) . (1994), Geologic Log Analysis Using Computer Methods, American Association of Petroleum Geologists. <https://pubs.geoscienceworld.org/aapg/books/book/1310/Geologic-Log-Analysis-Using-Computer-Methods>
- Esiri, A. E., Jambol, D. D., & Ozowe, C. (2024). Enhancing reservoir characterization with integrated petrophysical analysis and geostatistical methods. Open Access Research Journal of Multidisciplinary Studies, 7(2), 168–179. <https://doi.org/10.53022/oarjms.2024.7.2.0038>
- Jinad, R., Islam, A., & Shashidhar, N. (2024). Interpretability and Transparency of Machine Learning in File Fragment Analysis with Explainable Artificial Intelligence. Electronics, 13(13), 2438. <https://doi.org/10.3390/electronics13132438>
- Li, K. (2024). Research on Lithology Identification based on Machine Learning. Academic Journal of Science and Technology, 12(1), 121–125. <https://doi.org/10.54097/dv8q9b58>
- M. Garouani, J. Mothe, A. Barhrhouj and J. Aligon, (2024) "Investigating the Duality of Interpretability and Explainability in Machine Learning," in 2024 IEEE 36th International Conference on Tools with Artificial Intelligence (ICTAI), Herndon, VA, USA, 2024, pp. 861-867, <https://doi://10.1109/ICTAI62512.2024.00125>.
- Senjoba, L., Ikeda, H., Toriya, H., Hisada, M., Adachi, T., & Kawamura, Y. (2022). Visualization of 1D CNN Lithology Identification Model from Rotary Percussion Drilling Vibration Signals Using Explainable Artificial Intelligence Grad-CAM. International Journal of the Society of Materials Engineering for Resources. <https://doi.org/10.5188/ijmsr.25.224>
- Solanke, B., Onita, F. B., Ochulor, O. J., & Iriogbe, H. O. (2024). Techniques for improved reservoir characterization using advanced geological modeling in the oil and gas industry. International Journal of Applied Research in Social Sciences, 6(9), 2060–2088. <https://doi.org/10.51594/ijarss.v6i9.1542>
- Zhang, P., Gao, T., Li, R., & Fu, J. (2024). Advanced Machine Learning Framework for Enhanced Lithology Classification and Identification. <https://doi.org/10.2523/iptc-23595-ms>
- Zhao, B., & Liao, W. (2025). Lithology Identification of Buried Hill Reservoir Based on XGBoost with Optimized Interpretation. Processes, 13(3), 682. <https://doi.org/10.3390/pr13030682>