

## Perbandingan Metode Klasifikasi *Machine Learning*: Studi Kasus Prediksi Jenis Litologi Berdasarkan Data *Well Log* Pada Formasi Sleipner, *North Sea*

*Comparison of Machine Learning Classification Methods: Case Study of Lithology Type Prediction Based on Well Log Data in Sleipner Formation, North Sea*

<sup>1</sup>Mahdani Mahardika Sastra, <sup>1</sup>Rian Cahya Rohmana

<sup>1</sup>Teknik Perminyakan, Universitas Tanri Abeng, Jl. Swadarma Raya No.58, Ulujami, Kec. Pesanggrahan, Kota Jakarta Selatan

<sup>2</sup>Petrophysics TAU Research Group, Jakarta Selatan  
e-mail: mahdani.ms72@gmail.com (email korespondensi)

Received: 07 Februari 2024

Accepted: 25 Juli 2024

### Abstract

*The development of the industrial world era 4.0 encourages experts to digitize various fields, especially in the oil and gas industry. This research discusses the use of machine learning methods for lithology prediction in oil and gas exploration located in the Sleipner Formation, North Sea. The Sleipner Formation is composed by sandstone lithology with siltstone, shale and coal inserts. The main objective of this study is to compare three different machine learning algorithms in the context of lithology prediction based on well log data. This study utilizes five well log data which include gamma ray log, resistivity log, neutron log, density log and PEF log. Detailed lithologic data based on cores were used to validate the machine learning predictions. This research uses three machine learning algorithms namely decision tree (DT), k-nearest neighbor (KNN), and extra tree classifier (ETC). The three algorithms have also done hyperparameter tuning, in order to produce the best model when analyzing this research data. The prediction evaluation results on the blind test wells showed that the DT method has an accuracy of 83%, the KNN method 85% and the ETC algorithm 82%. The results show that these three machine learning algorithms have the potential for fairly accurate lithology prediction. In this study, the KNN algorithm is the most accurate in predicting lithology. This research is expected to provide new knowledge in the use of machine learning methods to predict lithology, especially useful in time efficiency and in wells that do not have core data.*

**Keywords:** *Machine Learning, Lithology, Well Log*

### Abstrak

Perkembangan era dunia industri 4.0 mendorong para ahli untuk melakukan digitalisasi terhadap berbagai bidang khususnya di industri minyak dan gas bumi. Penelitian ini membahas tentang penggunaan metode *machine learning* untuk prediksi litologi dalam eksplorasi migas yang terletak di Formasi Sleipner, *North Sea*. Formasi Sleipner disusun oleh litologi batupasir dengan sisipan batulanau, batuserpih dan batubara. Tujuan utama dari penelitian ini adalah untuk membandingkan tiga algoritma *machine learning* yang berbeda dalam konteks prediksi litologi berdasarkan data log sumur. Studi ini menggunakan lima data *well log* yang mencakup data log *gamma ray*, log resistivitas, log neutron, log densitas, dan log PEF. Data litologi rinci berdasarkan batuan inti (*core*) digunakan untuk validasi hasil prediksi dari *machine learning*. Penelitian ini menggunakan tiga algoritma *machine learning* yakni *decision tree* (DT), *k-nearest neighbor* (KNN), dan *extra tree classifier* (ETC). Ketiga algoritma tersebut juga sudah dilakukan *tuning hyperparameter*, agar menghasilkan model terbaik saat menganalisis data penelitian ini. Hasil evaluasi prediksi pada sumur *blind test* didapatkan metode DT memiliki akurasi sebesar 83%, metode KNN 85% dan algoritma ETC 82%. Hasil penelitian menunjukkan bahwa ketiga algoritma *machine learning* ini memiliki potensi untuk prediksi litologi yang cukup akurat. Pada

<http://jurnal.umat.ac.id/index.php/jt/index>

penelitian ini, algoritma KNN menjadi yang terakurat dalam memprediksi litologi. Penelitian ini diharapkan memberikan pengetahuan baru dalam penggunaan metode *machine learning* untuk memprediksi litologi, terutama berguna dalam efisiensi waktu dan pada sumur yang tidak memiliki data batuan inti (*core*).

**Kata Kunci:** *Machine Learning*, Litologi, *Well Log*

## PENDAHULUAN

Data log sumur (*well log*) memiliki karakteristik yang menguntungkan seperti resolusi vertikal yang tinggi, kontinuitas yang baik, dan perolehan data yang mudah. Oleh karena itu data *well log* merupakan sumber material yang penting untuk informasi litologi batuan dibawah permukaan. Klasifikasi litologi berdasarkan data *well log* merupakan dasar dalam analisis karakteristik *reservoir* dan evaluasi formasi, prediksi litologi juga memiliki nilai praktis dalam perhitungan cadangan pada tahap eksplorasi dan deskripsi *reservoir* secara rinci pada tahap pengembangan (Xie et al., 2018). Diperlukan analisis detil dalam karakterisasi *reservoir* agar dapat dihitung dengan benar dari sisi ekonomi dan dapat dikembangkan di masa depan (Kurniadi et al., 2023).

Umumnya terdapat dua cara untuk menentukan litologi batuan yaitu dengan mengidentifikasi serbuk pemboran selama operasi pemboran dan melalui observasi atau analisis sampel batuan inti (*core*) dari formasi dibawah permukaan (Salehi dan Honarvar, 2014). Analisis serbuk pemboran atau batuan inti, menggunakan teknik seperti analisis petrografi, analisis difraksi sinar-X (XRD), dan *scanning electron microscopy*, adalah metode lain yang paling umum saat ini untuk mengidentifikasi litologi (Kurniadi et al., 2023). Namun, kedua analisis tersebut tidak selalu mencerminkan kondisi aktual didalam lubang bor karena seorang analis yang berbeda dapat memberikan interpretasi yang berbeda. Maka dari itu data log sumur juga digunakan sebagai data pendukung untuk membantu para ahli dalam suatu penelitian. Akan tetapi, data log sumur merupakan data yang sangat besar dan berisi banyak sampel, sehingga dapat membebani para ahli dalam mengintegrasikan dan menginterpretasikan data litologi secara efektif dan efisien (Horrocks et al., 2015).

Demikian dengan pesatnya perkembangan kemajuan teknologi di era dunia industri 4.0 memungkinkan analisis litologi batuan dapat dilakukan dengan memanfaatkan bahasa pemrograman *python*. *Python* merupakan bahasa pemrograman interpretatif multiguna yang dapat digunakan untuk berbagai tujuan, termasuk *machine learning* yang mampu mengolah dan memproses dengan basis data yang besar (Lutz, 2013). *Machine learning* (ML) atau pembelajaran mesin adalah cabang dari *artificial intelligence* (AI) yang berfokus pada pembelajaran dari data pada pengembangan sistem secara efektif dan efisien dan mampu belajar secara mandiri tanpa harus diprogram secara eksplisit (Cholissodin, 2020).

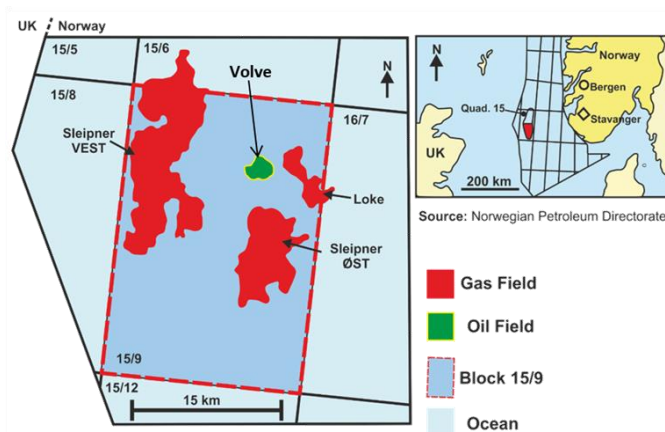
Penelitian ini menggunakan algoritma klasifikasi *supervised learning* yaitu *decision trees* (DT), *k-nearest neighbor* (KNN), dan *extra trees classifier* (ETC) untuk mengetahui perbandingan performa algoritma machine learning dalam memprediksi litologi batuan.

1. *Decision trees* (DT) adalah sebuah graph untuk merepresentasikan pilihan dan hasilnya dalam bentuk *tree*/pohon. *Node* dalam *graph* merepresentasikan sebuah kejadian atau pilihan dan sisi *graph* merepresentasikan aturan atau kondisi keputusan. Setiap *tree*/pohon terdiri dari *node* dan cabang. Setiap *node*

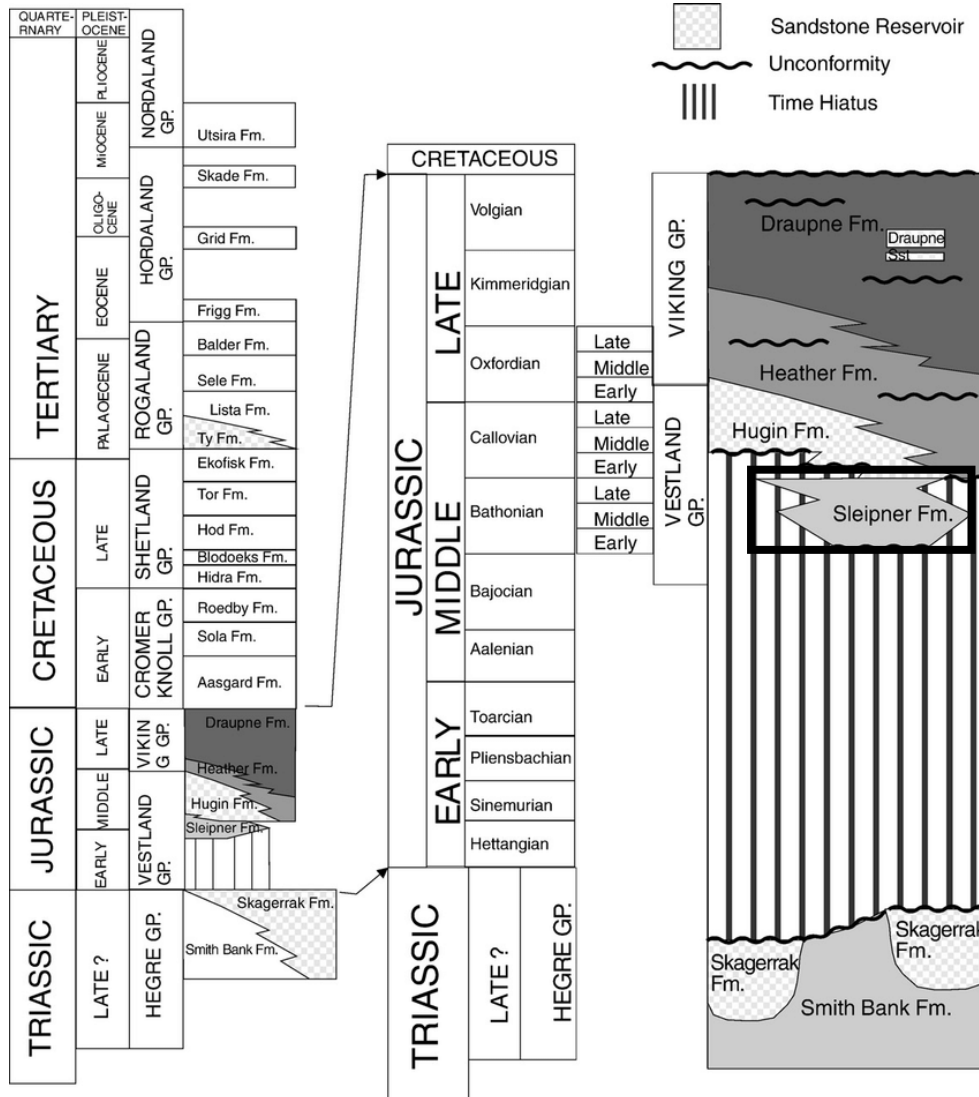
merepresentasikan atribut dalam sebuah kelompok yang akan diklasifikasikan dan setiap cabang merepresentasikan nilai yang dapat diambil oleh *node* tersebut (Mahesh, 2020).

2. *K-nearest neighbor* (KNN) adalah sebuah algoritma *supervised learning* yang secara sederhana memprediksi data berdasarkan mayoritas (*neighbor*) terdekatnya. Jika dalam masalah klasifikasi, kelas mayoritas dari titik-titik terdekat adalah prediksi dan dalam regresi, rata-rata dari titik-titik terdekat adalah prediksi. Pendekatan ini mungkin terlihat sederhana, namun dengan jumlah data yang besar, pendekatan ini bisa sangat efektif (Mohamed et al., 2019).
3. *Extra trees classifier* (ETC) atau disebut juga dengan “*extremely randomized tree*” adalah *ensemble* dari algoritma *decision tree*. Algoritma *Extra trees* bekerja dengan membuat *decision tree* berdasarkan dataset pelatihan, tanpa mengaplikasikan *pruning*. Prediksi dibuat berdasarkan *majority vote* kelas prediksi. Sebagai contoh, jika mayoritas kelas prediksi setiap pohon adalah “benar” maka keputusan akhir juga “benar” (Khomsah dan Ariwibowo, 2020).

Ketiga algoritma tersebut selanjutnya dilakukan *hyperparameter optimization* menggunakan *grid search hyperparameter* yang bertujuan untuk mengoptimalkan dan meningkatkan performa model algoritma *machine learning*. Penelitian ini berfokus di Formasi Sleipner, Lapangan Volve, *North Sea*, (Gambar 1) yang berada di lingkungan *shallow marine* (kedalaman air sekitar 80 m) dengan struktur geologi yang sangat kompleks dan ditemukan pada tahun 1993 (Sen dan Ganguli, 2019). Formasi Sleipner terletak dibagian tengah *North Sea (Norwegian Continental Shelf)* yang termasuk kedalam bagian dari *Vestland Group* (Gambar 2) yang memiliki dua *reservoir* utama yaitu pada Formasi Sleipner dan Formasi Hugin, ke-dua formasi tersebut memiliki ketebalan total sekitar 153 m (Statoil, 1998). Formasi Sleipner merupakan formasi yang terbentuk pada masa *Jurassic* Tengah bersama dengan Formasi Hugin di atasnya. Formasi Sleipner berada di lingkungan pengendapan laut dangkal (Vollset dan Dore, 1984). Formasi ini memiliki litologi yang terdiri dari batupasir dengan lapisan batulanau, batuserpil dan lapisan batubara (Statoil, 1998).



Gambar 1 Lokasi lapangan Volve, *Norwegian North Sea* (Ravasi et al., 2015)



Gambar 2 Stratigrafi regional Lapangan Volve (Folkested & Satur, 2008)

Pada lokasi penelitian juga diketahui memiliki *petroleum system* yang lengkap dengan dibuktikan adanya empat formasi potensi batuan induk (*source rock*) dengan TOC yang sangat tinggi yaitu Formasi Sleipner, Formasi Hugin, Formasi Heather, dan Formasi Draupne (Thomas et al., 1984). Batuan *reservoir* utama yang terdapat pada wilayah *North Sea* yaitu Formasi Bryne dan Sandness dibagian timur dan Formasi Hugin dan Sleipner dibagian barat. Pada batuan penutup (*cap rock*) terdapat pada Formasi Draupne dan Formasi TAU di *Central North Sea*, Formasi Asgard di Utsira High dan Formasi Smith Bank di cekungan Norwegian Danish. Kemudian proses migrasi minyak bumi terjadi di wilayah tengah *North Sea* yang terkait dengan penurunan Formasi Draupne setelah sistem retakan (*rift*) yang gagal. Migrasi hidrokarbon juga terjadi dari *Viking Graben* di sepanjang jalur ke tepi Utsira High serta migrasi hidrokarbon dari Formasi Draupne di *Graben* ke Formasi Hugin dan Formasi Sleipner (Kaspersen, 2016).

### METODE PENELITIAN

Penelitian ini menggunakan empat data sumur yakni MS-1, MS-2, MS-4, MS-5 sebagai data pelatihan (*training*) dan satu sumur MS-3 sebagai pengujian (*testing*) (Tabel 1). Data sumur yang digunakan yaitu data *well log* meliputi log *gamma ray*, resistivitas, neutron, densitas, PEF dan data litologi batuan yang berasal dari *core* yang digunakan sebagai validasi model (Tabel 2). Data sumur tersebut kemudian dilakukan *removing outliers* dari hasil *outliers detection* dengan algoritma *isolation forest* yang bertujuan untuk memisahkan data inti dengan data yang menyimpang, sehingga diharapkan data yang dihasilkan menjadi lebih baik dan akurat (Bhattacharya, 2021). Pada studi ini, *cross bedded sst* (1), *silt* (2), *shaly silt* (3), dan *coal* (4) merupakan empat kelas target litologi batuan yang akan diidentifikasi.

**Tabel 1 Skema konsep train & test data**

Skema	Train Data	Test Data
I	4 data sumur (MS-1, MS-2, MS-4, dan MS-5)	1 data sumur (MS-3)

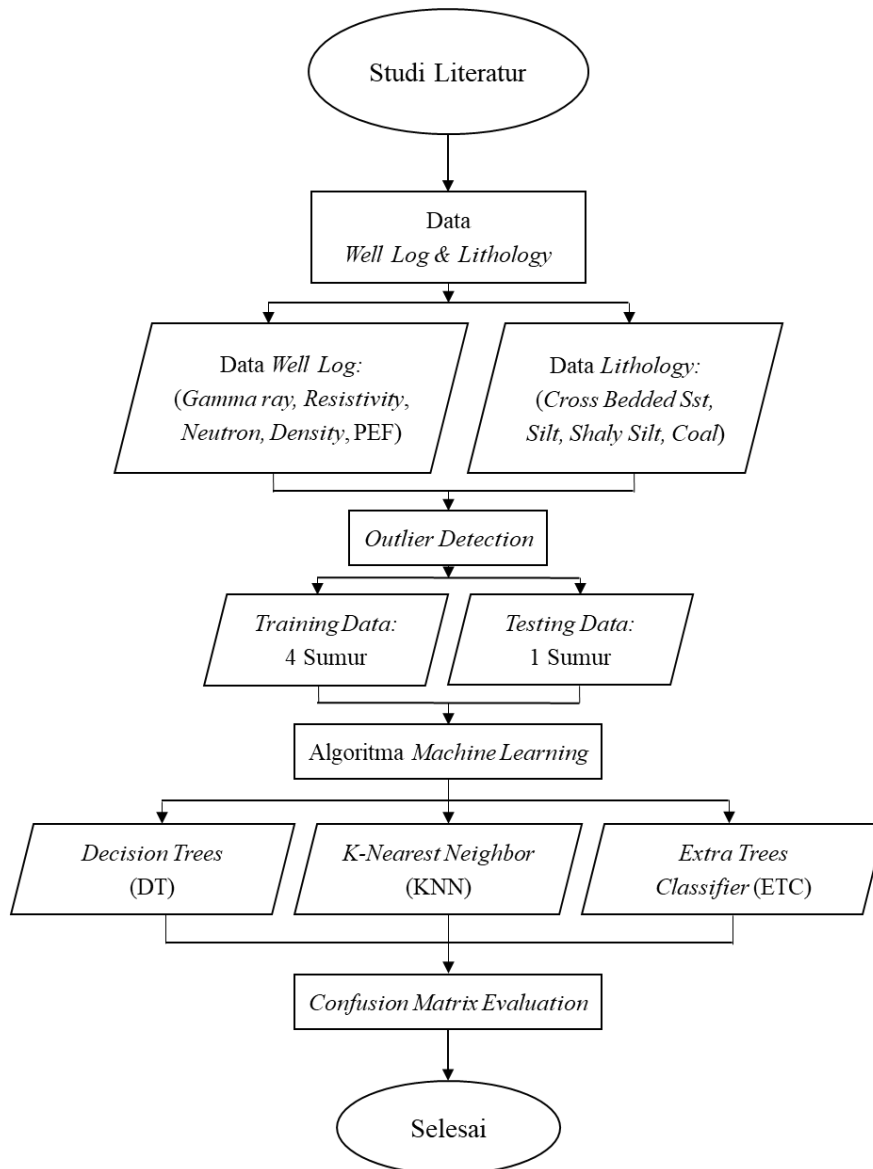
**Tabel 2 Bahan atau data yang digunakan setiap sumur**

Sumur	<i>Gamma Ray</i>	<i>Resistivity</i>		<i>Neutron</i>	<i>Density</i>	PEF	<i>Lithology</i>
	GR	RDEP	RMED	NPHI	RHOB		
MS-1	✓	✓	✓	✓	✓	✓	✓
MS-2	✓	✓	✓	✓	✓	✓	✓
MS-3	✓	✓	✓	✓	✓	✓	✓
MS-4	✓	✓	✓	✓	✓	✓	✓
MS-5	✓	✓	✓	✓	✓	✓	✓

Metode atau model algoritma klasifikasi penelitian ini seluruhnya dilakukan *hyperparameter optimization* menggunakan *grid search*. Kemudian, performa model algoritma yang dibangun, diukur dengan *confusion matrix* (Gambar 3). *Confusion matrix* berisi informasi tentang klasifikasi aktual dan prediksi yang dilakukan oleh model classifier (Tabel 3). Tujuannya adalah untuk mengukur tingkat akurasi kinerja model algoritma (Khomsah dan Ariwibowo, 2020).

Akurasi didefinisikan sebagai rasio jumlah prediksi yang benar terhadap jumlah total prediksi (Pandey et al., 2020). Klasifikasi akurasi yang benar adalah *true positive* (TP) dan *true-negative* (TN). Berikut merupakan persamaan (1) dari akurasi (Khomsah dan Ariwibowo, 2020):

$$Akurasi = \frac{(TP + TN)}{(TP + FP + TN + FN)} \quad (1)$$



Gambar 3 Diagram alir penelitian

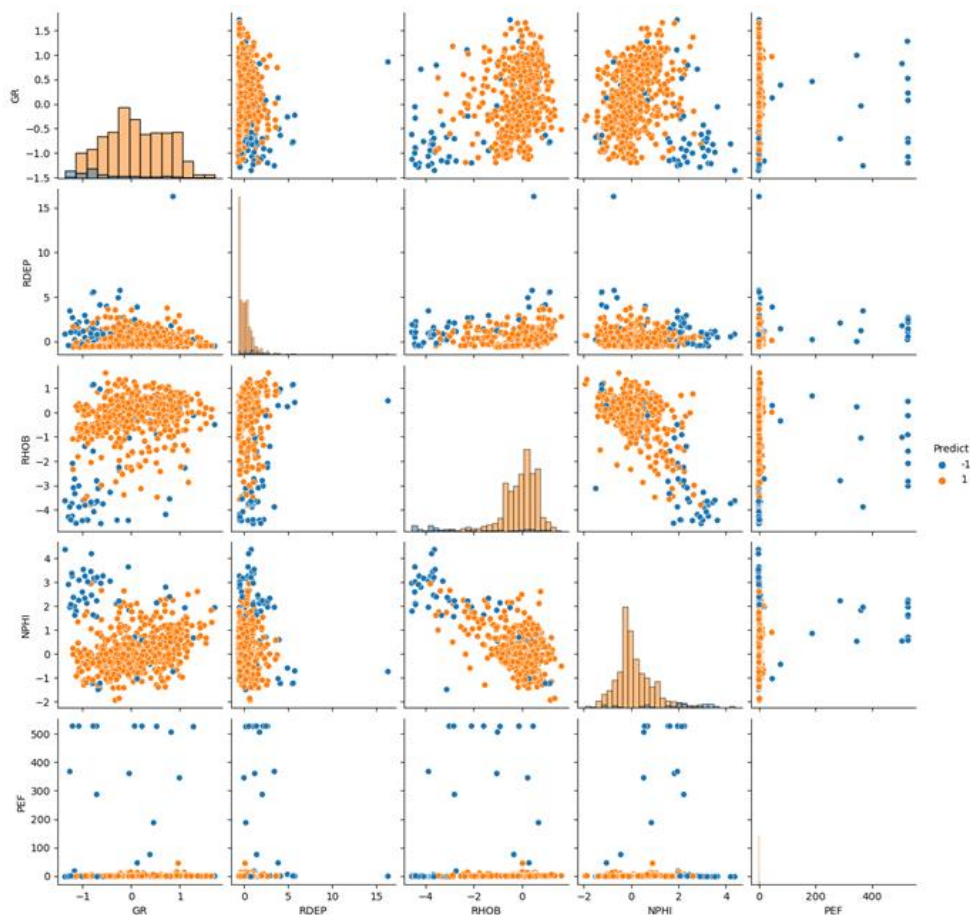
Tabel 3 Confusion matrix (Khomsah & Ariwibowo, 2020)

Confusion Matrix	Actual Class	
	Positive	Negative
Positive	True Positive (TP)	False Positive (FP)
Negative	False Negative (FN)	True Negative (TN)

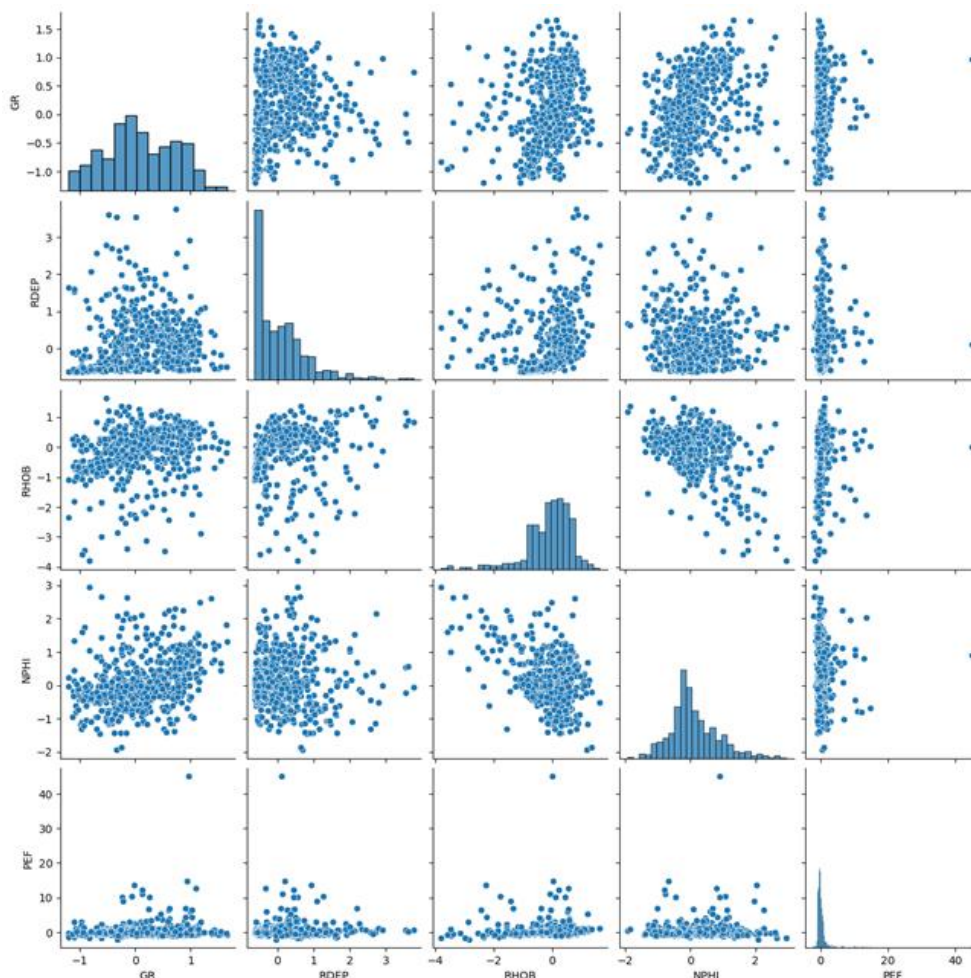
### HASIL DAN PEMBAHASAN

Prediksi litologi adalah langkah penting untuk memahami karakteristik *reservoir* di bawah permukaan. Hal ini dapat membantu mengidentifikasi jenis batuan yang ditemui, serta menggambarkan lapisan – lapisan batuan yang dapat mengandung hidrokarbon. Informasi ini memberikan dasar penting untuk perencanaan pengeboran dan pengambilan keputusan yang lebih baik dalam pengembangan sumber daya alam di industri migas. Selain itu, prediksi litologi juga membantu dalam menentukan kemungkinan zona produktif yang dapat mengarah pada penemuan *reservoir* minyak dan gas yang bernilai ekonomis.

Analisis prediksi litologi dengan menggunakan data *well log* kemudian memanfaatkan *outlier detection* untuk menghapus data anomali yang menyimpang dari pengamatan reguler, sehingga diharapkan *output* yang dihasilkan menjadi lebih baik dan akurat. Pada gambar 4 dan 5 ditunjukkan berupa hasil dari sebelum dan sesudah dilakukan *removing outlier* dengan *outlier detection*.



Gambar 4 Hasil *outlier detection* data *well log* sebelum *removing outlier*

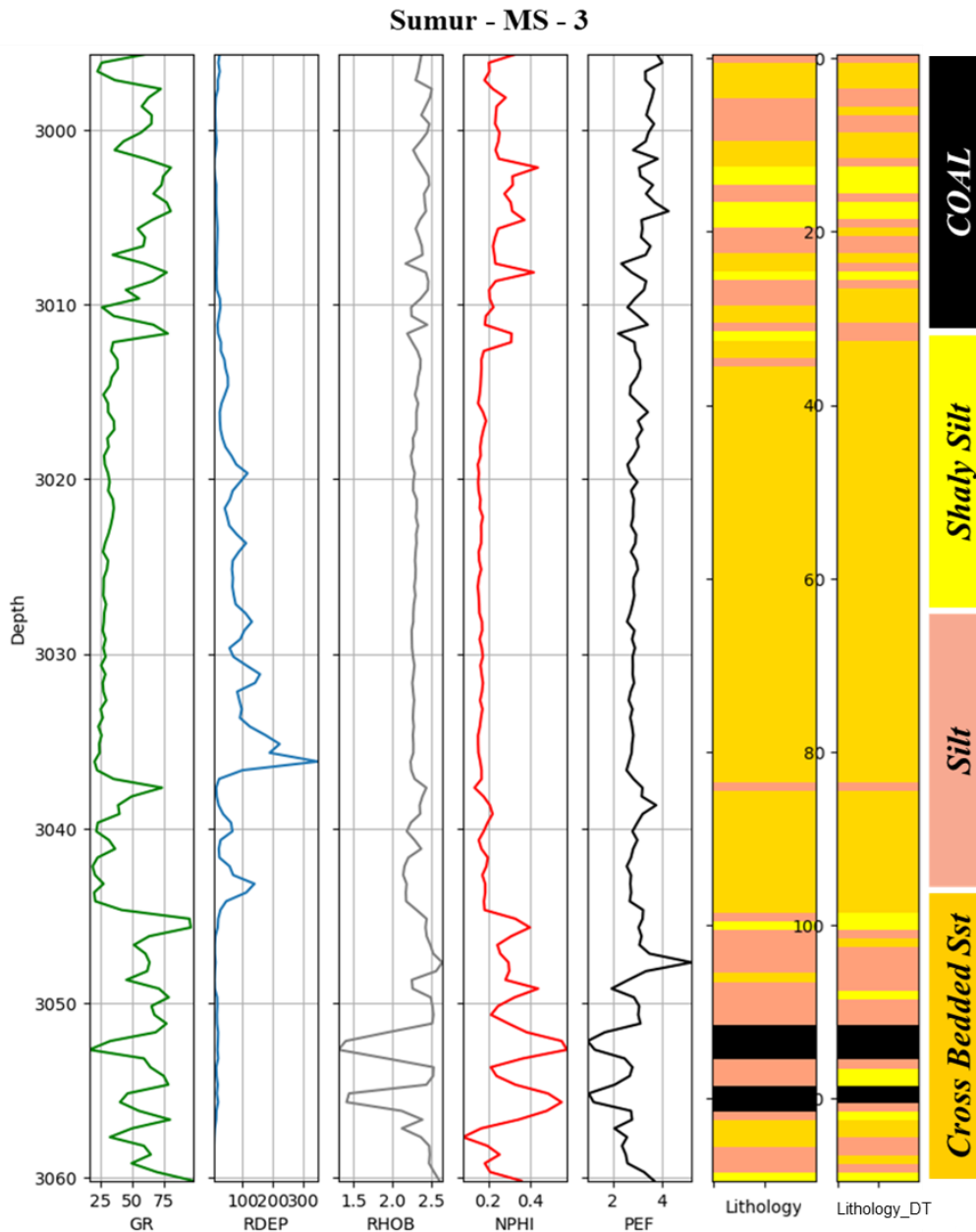


Gambar 5 Hasil *outlier detection* data *well log* sesudah *removing outlier*

Hasil seleksi dari data tersebut selanjutnya diproses dan diolah dengan model dari ketiga algoritma yang telah dilakukan *hyperparameter optimization* menghasilkan performa kinerja model sebagai berikut:

1. Analisis klasifikasi litologi menggunakan model algoritma *decision tree* (DT), yang diperkuat dengan optimisasi *grid search hyperparameter*, menghasilkan nilai akurasi sebesar 83%. Hasil prediksi menunjukkan kemampuan algoritma DT dalam mengidentifikasi empat litologi pada sumur pengujian (*test*). Lebih lanjut, hasil ini menunjukkan korespondensi yang signifikan dengan data litologi aktual. Model ini cukup bisa membedakan *cross bedded sandstone*, diikuti oleh keberadaan *silt* dengan sisipan *shaly silt* pada lapisan atas dan bawah, serta lapisan *coal* yang lebih dalam, sebagaimana diilustrasikan dalam gambar 6.

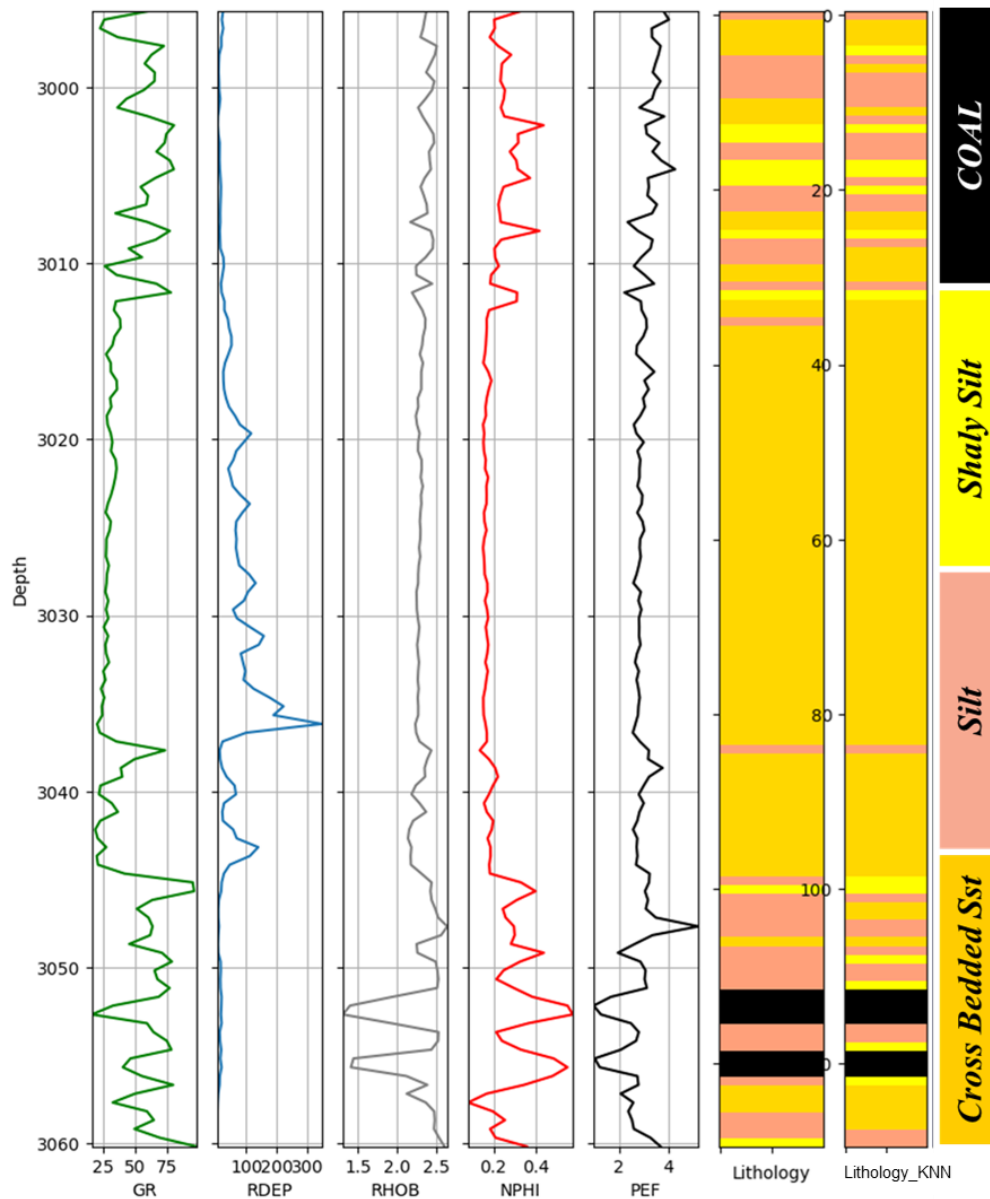




**Gambar 6 Perbandingan visualisasi hasil prediksi litologi pada model algoritma DT & data litologi (core)**

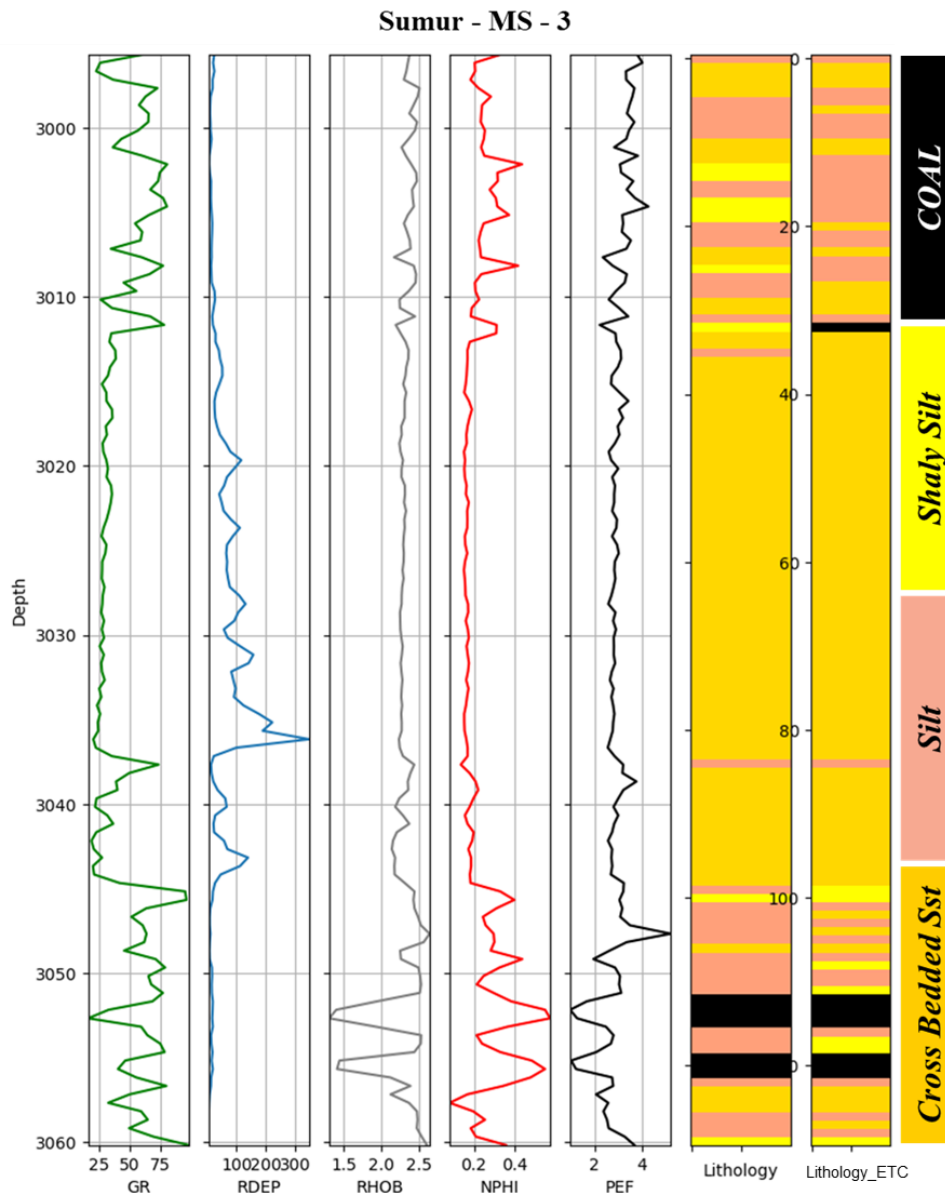
2. Pada studi ini, penggunaan model algoritma *k-nearest neighbor* (KNN) untuk klasifikasi litologi yang ditingkatkan melalui proses optimasi *hyperparameter*, menghasilkan akurasi sebesar 85%. Hasil prediksi menunjukkan efektivitas algoritma KNN dalam mengidentifikasi empat jenis litologi pada sumur *test*. Temuan ini juga menunjukkan adanya kesesuaian yang signifikan dengan data litologi asli dibandingkan model dari algoritma DT. Hasil prediksi yang dibandingkan dengan data sebenarnya terlihat dalam gambar 7.

Sumur - MS - 3



Gambar 7 Perbandingan visualisasi hasil prediksi litologi pada model algoritma KNN & data litologi (core)

3. Klasifikasi litologi menggunakan model algoritma *extra trees classifier* (ETC), yang didukung oleh optimisasi *grid search hyperparameter*, menghasilkan nilai akurasi sebesar 82%. Akurasi ini merupakan nilai terendah dibandingkan dengan algoritma DT dan KNN. Walaupun memiliki akurasi terendah, secara umum algoritma ETC dapat membedakan empat litologi yang ada pada sumur. Tercatat perbedaan yang signifikan, khususnya dalam distribusi litologi *cross bedded sandstone*, keberadaan litologi *silt* dengan sisipan *shaly silt* terbatas pada lapisan bawah, dan keberadaan lapisan *coal* sebagaimana terlihat dalam gambar 8.



**Gambar 8 Perbandingan visualisasi hasil prediksi litologi pada model algoritma ETC & data litologi (core)**

Secara rinci hasil perbandingan dari ketiga algoritma tersebut berdasarkan nilai akurasi dari *confusion matrix* dapat dilihat pada tabel 4.

**Tabel 4 Hasil analisis algoritma *machine learning***

No	Algoritma (ML)	Akurasi (%)
1	<i>Decision Trees</i>	83
2	<i>K-Nearest Neighbor</i>	85
3	<i>Extra Trees Classifier</i>	82

### UCAPAN TERIMAKASIH

Ucapan terima kasih kepada pihak Universitas Tanri Abeng dan *Petrophysics TAU Research Group* (PTRG) yang telah membantu serta membimbing hingga terselesaikannya penelitian ini.

### SIMPULAN DAN SARAN

1. Pentingnya menerapkan prosedur *removing outlier* untuk mengeliminasi data *well log* yang menyimpang atau anomali, dengan tujuan memastikan data yang digunakan reguler dan tidak menyimpang.
2. Peningkatan model algoritma dengan optimasi *hyperparameter* dapat meningkatkan akurasi prediksi.
3. Model algoritma *k-nearest neighbor* (KNN) mencapai akurasi tertinggi 85% dalam keakuratan prediksi litologi pada sumur uji. Diikuti model algoritma *decision tree* (DT) yang mencapai akurasi 83% dan *extra trees classifier* (ETC) yang mencapai akurasi 82%.

### DAFTAR PUSTAKA

- Bhattacharya, S. (2021). *A primer on machine learning in subsurface geoscience*. PGE. Switzerland: Springer.
- Cholissodin, I., Sutrisno., Subroto, A.A., Hasanah, U., dan Febiola, Y.I. (2020). *AI machine learning & deep learning*. Filkom. Malang: Universitas Brawijaya.
- Folkestad, A., dan Satur, N. (2008). *Regressive and transgressive cycles in a rift-basin: Depositional model and sedimentary partitioning of the middle jurassic Hugin Formation, Southern Viking Graben, North Sea*. Sedimentary Geology. Norway: Elsevier. hal. 1-21.
- Horrocks, T., Holden, E.J., dan Wedge, D. (2015). *Evaluation of automated lithology classification architectures using highly-sampled wireline logs for coal exploration*. CG, 83. Australia: Elsevier. hal. 209-218.
- Kurniadi, F.I., Rohmana, R.C., dan Taufani, L. (2023). *Local mean imputation for handling missing value to provide more accurate facies classification*. Procedia Computer Science. 1. hal. 301-309.
- Kurniadi, F.I., dan Rohmana, R.C. (2023). *Enhancing lithology classification performance through random forest, COPOD, and bayesian optimization*. ICISS. IEEE. hal. 1-5.
- Kaspersen, H.M. (2016). *Reservoir characterization of jurassic sandstones of the Johan Sverdrup field, Central North Sea*. Master Thesis. Oslo: Departmen of Geoscience. University of Oslo.
- Khomsah, S., dan Ariwibowo, A.S. (2020). *Model text-preprocessing komentar youtube dalam bahasa Indonesia*. RESTI. vol. 4. no. 4. Sumatera Barat: IAI. hal. 648-654.
- Lutz, M. (2013). *Learning python: powerful object-oriented programming*. United States of America. Canada: O'Reilly.
- Mahesh, B. (2020). *Machine learning algorithm – A review*. CE. Vol. 9 (1). India: IJSR. hal. 381-386.
- Mohamed, I.M., Mohamed, S., Mazher, I., dan Chester, P. (2019). *Formation lithology classification: insights into machine learning methods*. Canada: SPE.

- Pandey, Y.N., Rastogi, A., Kainkaryam, S., Bhattacharya, S., dan Saputelli, L. (2020). *Machine learning in the oil and gas industry*. California: Apress.
- Ravasi, M., Vasconcelos, I., Curtis, A., dan Kritski, A. (2015). *Vector-acoustic reverse time migration of volve ocean-bottom cable data set without up/down decomposed wavefields*. Geophysics. vol. 80. no. 4. Norway: SEG. hal. 137-150.
- Statoil. (1998). *Final Well Report, North Sea*. Norway: Statoil.
- Salehi, S.M., dan Honarvar, B. (2014). *Automatic identification of formation lithology from well log data: a machine learning approach*. JPSR. vol. 3 (2). Iran: Science and Engineering. hal. 73-82.
- Sen, S., dan Ganguli, S.S. (2019). *Estimation of pore pressure and fracture gradient in volve field, Norwegian North Sea*. Mumbai. India: SPE.
- Thomas, B.M., Pedersen, P.M., Whitaker, M.F., dan Shaw, N.D. (1984). *Organic facies and hydrocarbon distributions in the Norwegian North Sea*. PGSE. Springer. Stavanger: Graham & Trotman. hal. 3-26.
- Vollset, J., dan Dore, A.G. (1984). *A revised triassic and jurassic lithostratigraphic nomenclature for the Norwegian North Sea*. NPD-Buletin no 3. Norway: Oljedirektoratet.
- Xie, Y., Zhu, C., Zhou, W., Li, Z., Liu, X., dan Tu, M. (2018). *Evaluation of machine learning methods for formation lithology identification: A comparison of tuning processes and model performances*. JPSE. Chengdu: Elsevier. hal. 182-193.