

## Penerapan *Machine Learning* dalam Penentuan Porositas Batuan: Studi Kasus Menggunakan Regresi Linier Berganda dan Regresi KNN pada Data Log Sumur

*Application of Machine Learning in Rock Porosity Determination: Case Study Using Multiple Linear Regression and KNN Regression on Well Log Data*

<sup>1</sup>Rian Cahya Rohmana, <sup>2</sup>Delva Triwanti, <sup>3</sup>Priskila Rully Setiyaningrum

<sup>1</sup>Teknik Perminyakan, Fakultas Teknik dan Teknologi, Tanri Abeng University

<sup>2</sup>Teknik Perminyakan, Fakultas Teknik dan Teknologi, Tanri Abeng University

<sup>3</sup>Teknik Perminyakan, Fakultas Teknik dan Teknologi, Tanri Abeng University  
e-mail: rian@tau.ac.id

Received: 01 Februari 2024

Accepted: 25 Juli 2024

### Abstract

*This study aims to predict porosity values based on well log data. The well logs used are from three wells and include triple combo logs and validated effective porosity with core data. Machine learning is applied using two regression models: multiple linear regression and k-nearest neighbors (KNN) regression. Before applying the models, hyperparameter tuning is conducted to optimize both models. As part of the evaluation, a blind test is implemented on one of the research wells with three different tests. Each test uses a different combination of data, including gamma ray, density, neutron, deep resistivity, and photoelectric effect logs. Model evaluation is done using the metrics R2 (coefficient of determination), RMSE (root mean squared error), and MAE (mean absolute error). Results show that the multiple linear regression model achieved an R2 of 90.6% in test 1, 90.59% in test 2, and 90.59% in test 3. The KNN regression model achieved an R2 of 94.85% in test 1, 92.54% in test 2, and 92.26% in test 3. The RMSE for the multiple linear regression model in all tests is 0.0225, while for the KNN regression model it is 0.0167 in test 1, 0.0201 in test 2, and 0.0204 in test 3. The MAE for the multiple linear regression model is 0.0088 in test 1 and 0.0089 in tests 2 and 3. Meanwhile, the MAE for the KNN regression model is 0.0047 in test 1, 0.0053 in test 2, and 0.0056 in test 3. Based on the evaluation results, the KNN regression model performs more optimally than the multiple linear regression model in predicting porosity values in the used dataset, especially in test 1. This study demonstrates that predicting rock porosity values can be effectively done using ML methods and can be adapted to various scenarios. This can aid in petrophysical analysis, especially if there are limitations in core data or to expedite petrophysical analysis in large datasets.*

**Keywords:** Porosity, Well Log, Regression, Machine Learning, Petrophysics.

### Abstrak

Penelitian ini bertujuan untuk memprediksi nilai porositas berdasarkan data log sumur. Log sumur berasal dari tiga sumur mencakup *triple combo log* dan porositas efektif yang telah divalidasi data core. Penerapan *machine learning* menggunakan dua model regresi, yakni regresi linier berganda dan *k-nearest neighbors* (KNN). Sebelum penerapan model, proses pengaturan *hyperparameter* dilakukan guna mengoptimalkan kedua model yang digunakan. Sebagai bagian dari evaluasi, *blind test* diimplementasikan pada salah satu sumur penelitian dengan tiga uji yang berbeda. Setiap uji memiliki kombinasi data yang berbeda, yakni log *gamma ray*, *density*, *neutron*, *deep resistivity*, dan *photoelectric effect*. Evaluasi model dilakukan dengan metrik R2 (*coefficient of determination*), RMSE (*root mean squared error*), dan MAE (*mean absolute error*). Hasil menunjukkan bahwa model regresi linear berganda mendapatkan R2 sebesar 90.6% pada uji 1, 90.59% pada uji 2, dan 90.59% pada uji 3. Sementara model

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

regresi KNN mendapatkan nilai R<sup>2</sup> sebesar 94.85% pada uji 1, 92.54% pada uji 2, dan 92.26% pada uji 3. RMSE pada model regresi linear berganda pada semua uji adalah 0.0225, sedangkan pada model regresi KNN uji 1 mendapatkan 0.0167, uji 2 mendapatkan 0.0201, dan uji 3 mendapatkan 0.0204. MAE pada model regresi linear berganda adalah 0.0088 pada uji 1 dan 0.0089 pada uji 2 dan 3. Sementara itu, MAE pada model regresi KNN adalah 0.0047 pada uji 1, 0.0053 pada uji 2, dan 0.0056 pada uji 3. Berdasarkan hasil evaluasi, model regresi KNN memiliki performa yang lebih optimal dibandingkan dengan regresi linear berganda dalam prediksi nilai porositas pada *dataset* yang digunakan, terutama dalam uji 1. Penelitian ini menunjukkan prediksi nilai porositas batuan dapat dengan baik dilakukan menggunakan metode ML dan dapat disesuaikan dengan beragam skenario. Hal ini dapat membantu dalam analisis petrofisika, terutama jika terdapat keterbatasan data *core* atau dapat membantu mempercepat proses analisis petrofisika dalam jumlah data yang besar.

**Kata Kunci:** Porositas, Log Sumur, Regresi, *Machine Learning*, Petrofisika.

## PENDAHULUAN

Porositas batuan adalah parameter kunci dalam eksplorasi dan produksi hidrokarbon, memberikan informasi esensial tentang kapasitas batuan untuk menyimpan fluida. Tipe dan volume ruang yang tersedia di batuan reservoir (porositas) serta kemudahan aliran hidrokarbon sangat penting dalam klasifikasi tipe batuan [1]. Secara umum, penentuan porositas dilakukan melalui analisis data *core* yang diperoleh secara langsung dari formasi batuan. Namun, pendekatan ini sering kali dibatasi oleh keterjangkauan dan ketersediaan data, umumnya data *core* tidak tersedia untuk semua sumur atau atau lapisan reservoir [2]. Pada konteks ini, log sumur menyediakan data kontinu mengenai sifat fisik formasi batuan digunakan sebagai sumber data yang berharga.

Penggunaan kecerdasan buatan (*artificial intelligence*) dan algoritma pembelajaran mesin (*machine learning*) telah mendapatkan perhatian besar dalam industri energi, minyak, dan gas. Algoritma berbasis data ini sering digunakan untuk mengestimasi parameter yang secara teknis atau teoretis sulit untuk diperoleh [3]. *Machine learning* (ML) memanfaatkan berbagai algoritma termasuk regresi linear berganda [4] maupun *k-nearest neighbors* [5] untuk memprediksi fitur batuan (geologi) menggunakan data log sumur. Salah satu area penelitian dalam ML yang banyak diteliti adalah pengklasifikasian fasies [6], litologi [7] dan prediksi permeabilitas [8, 9] menggunakan data dari log sumur.

Penelitian ini dimulai dengan pengumpulan data dari tiga sumur yang mencakup *triple combo log*. Untuk mendapatkan hasil prediksi yang optimal dalam proses *machine learning*, penting untuk melakukan optimasi *hyperparameter* menggunakan metode optimasi *random search* [10]. Sebagai bagian dari evaluasi model, penelitian ini mengimplementasikan blind test pada salah satu sumur dengan tiga skenario berbeda, memungkinkan perbandingan kinerja kedua model dalam kondisi variatif.

Pengujian model termasuk mempertimbangkan kemampuan prediksi terhadap data uji [11]. Cara menguji meliputi evaluasi keakuratan model pada data pelatihan (*train*), validasi silang, dan data uji (*holdout*) [12]. Evaluasi kinerja model dilakukan menggunakan metrik umum untuk menilai model regresi yakni R<sup>2</sup> (*coefficient of determination*), RMSE (*root mean squared error*), dan MAE (*mean absolute error*) [11]. R<sup>2</sup> adalah ukuran statistik yang

mengindikasikan seberapa besar variasi pada variabel terikat yang bisa diprediksi dari variabel bebas. Nilai ini menggambarkan seberapa akurat suatu model statistik dapat merepresentasikan data. Nilai  $R^2$  berkisar antara 0 hingga 1, nilai yang lebih tinggi menunjukkan model lebih efektif dalam prediksi, sedangkan nilai  $R^2$  yang rendah menandakan sebaliknya [13]. RMSE adalah ukuran statistik yang menentukan rata-rata jarak antara nilai-nilai yang diprediksi oleh model dan nilai-nilai sebenarnya dalam data [14]. RMSE merupakan akar kuadrat dari MSE (*mean squared error*), yang mengukur rata-rata perbedaan kuadrat antara prediksi dan nilai sebenarnya. MAE adalah ukuran statistik yang mengukur rata-rata perbedaan absolut antara nilai yang diprediksi dan nilai sebenarnya dalam model regresi [15]. Nilai MAE yang lebih rendah menunjukkan bahwa model lebih efektif dalam meminimalkan kesalahan prediksi.

Penelitian ini menggunakan data sumur yang terletak di Lapangan Volve, blok 15/9 di selatan Laut Utara Norwegia. Reservoir utama pada lapangan ini adalah Formasi Hugin [16], formasi ini juga yang menjadi fokus dalam penelitian ini untuk diprediksi porositasnya. Formasi ini memiliki ciri lingkungan pengendapan *shallow marine* dengan litologi reservoir yang dominan adalah batupasir.

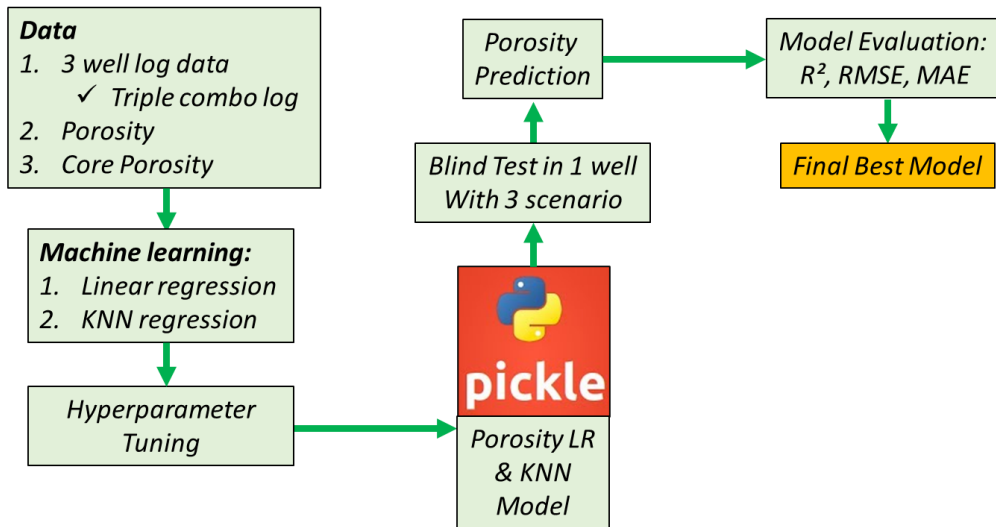
Mengintegrasikan pendekatan *machine learning* dalam analisis data log sumur, penelitian ini membuka metode lain dalam analisis petrofisika. Kemampuan untuk memprediksi nilai porositas dengan akurat dari data log sumur dapat menjadi solusi yang efisien dalam situasi di mana data *core* terbatas atau dalam kondisi di mana analisis cepat dari sejumlah besar data diperlukan. Hal ini tidak hanya meningkatkan efisiensi dalam proses eksplorasi dan produksi hidrokarbon, tetapi juga menyediakan informasi yang lebih mendalam mengenai sifat-sifat reservoir.

## METODE PENELITIAN

Penelitian ini melewati serangkaian metode (Gambar 1), dimulai dengan pengumpulan data dari tiga log sumur yang meliputi *triple combo log* yang terdiri dari log *gamma-ray* (GR), resistivitas (RDEP), *neutron* (NEU), densitas (DEN) dan *photoelectric effect* (PEF). Selain itu, data porositas yang diperoleh dari perhitungan log sumur yang sudah divalidasi dengan data porositas batuan inti (*core*).

Model tersebut akan dilatih menggunakan open data dari Volve *dataset* [18], dua sumur yaitu sumur 15\_9-19A dan sumur 15\_9-19S&SR. Setelah pelatihan, model akan diuji untuk memprediksi porositas pada sumur 15\_9-19B&BT2. Selanjutnya, model yang telah terbentuk akan diserialisasi dengan memanfaatkan *library 'pickle'* pada *Python*. *Library 'pickle'* adalah modul yang memudahkan penyimpanan dan pemanggilan kembali objek pada *Python*. *Library* ini mengubah objek *Python* menjadi aliran *byte* yang disebut "*pickling*", dan mengubah aliran *byte* tersebut kembali menjadi objek *Python*, proses yang dikenal sebagai "*unpickling*" [17]. Hal ini memungkinkan model terlatih untuk disimpan dalam sebuah berkas yang nantinya dapat digunakan kembali atau diterapkan dengan parameter yang telah diatur sebelumnya.

Pengujian model dalam memprediksi porositas akan dilakukan pada sumur 15\_9-19B&BT2 dengan tiga skenario berbeda (Tabel 1). Tiga skenario akan dilakukan dalam tiga uji yang berbeda (tabel 1), dengan setiap uji menggunakan kombinasi data log yang berbeda. Pada uji 1, data yang digunakan (prediktor) adalah GR, DEN, dan NEU.



Gambar 1. Metode Penelitian

Sedangkan pada uji 3, meliputi GR, RDEP, DEN, NEU, dan PEF. Setiap kombinasi data (prediktor) ini dimaksudkan untuk menguji bagaimana variasi data input dapat mempengaruhi kemampuan model dalam memprediksi porositas. Model akan dievaluasi berdasarkan kemampuan prediksinya dan akan dibandingkan antar tes untuk menentukan kombinasi prediktor mana yang memberikan hasil prediksi yang paling akurat. Hasil eksperimen ini diharapkan dapat memberikan wawasan mengenai pengaruh variabel data terhadap akurasi prediksi porositas dalam penggunaan ML.

Evaluasi model RL dan KNN akan dilakukan dengan menggunakan metrik seperti R2, RMSE, dan MAE untuk menilai akurasi prediksi dan tingkat kesalahan model. Model dengan nilai R<sup>2</sup> tertinggi dan nilai RMSE serta MAE terendah akan dipilih sebagai model terbaik dan menunjukkan prediksi porositas yang lebih akurat. Terakhir, pada penelitian ini akan menghasilkan dokumentasi yang rinci mengenai metodologi, proses pengembangan model, hasil evaluasi, dan pemilihan model akhir. Penelitian ini juga diharapkan memberi laporan komprehensif yang menguraikan temuan penelitian dan keefektifan model ML yang dipilih dalam memprediksi porositas pada reservoir.

Tabel 1. Tiga Skenario uji menggunakan kombinasi data log (prediktor) yang berbeda

Uji 1	Uji 2	Uji 3
<i>predictors=</i> ["GR", "DEN", "NEU"]	<i>predictors=</i> ["GR", "RDEP", "DEN", "NEU"]	<i>predictors=</i> ["GR", "RDEP", "DEN", "NEU", "PEF"]

## HASIL DAN PEMBAHASAN

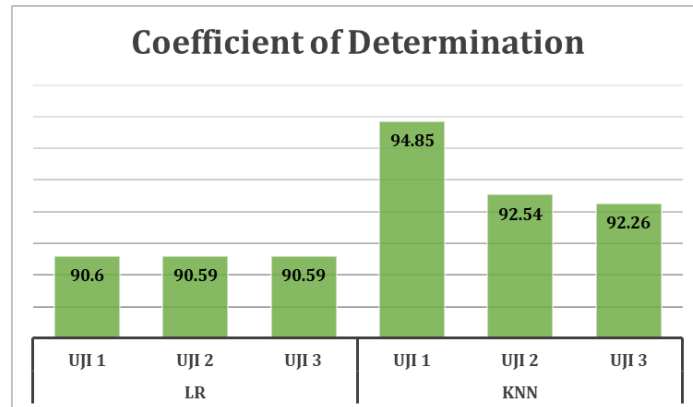
Hasil penelitian menunjukkan perbandingan efektivitas model regresi linier berganda (RL) dan k-nearest neighbors (KNN) dalam memprediksi porositas. Pada pengujian yang dilakukan, diperoleh nilai  $R^2$  (coefficient of determination), RMSE (root mean squared error), dan MAE (mean absolute error) untuk masing-masing model dan setiap uji yang dilaksanakan.

Pada model RL, nilai  $R^2$  untuk uji 1 (Gambar 2) mencapai 90.6, menandakan bahwa model tersebut memprediksi 90.6% data dengan tepat. Sementara itu, nilai  $R^2$  yang didapat pada uji 2 dan uji 3 konsisten sebesar 90.59%. Konsistensi ini mengindikasikan bahwa perubahan data yang digunakan (prediktor) tidak memberikan pengaruh signifikan terhadap peningkatan kinerja model regresi linier dalam sampel uji ini. Nilai RMSE pada uji 1, uji 2, dan uji 3 berturut-turut stabil, yaitu 0.0225 (Gambar 3). Hal ini menandakan bahwa kesalahan prediksi model regresi linier relatif konstan terlepas dari perubahan dalam prediktor yang digunakan. Nilai MAE tercatat relatif stabil dengan sedikit variasi antara uji 1, uji 2, dan uji 3, yaitu sebesar 0.0089, 0.0088, dan 0.0088 secara berturut-turut (Gambar 4). Nilai-nilai ini menandakan bahwa, rata-rata, kesalahan prediksi dari model RL tidak mengalami perubahan yang signifikan antara skenario yang berbeda.

Sedangkan untuk model KNN, terjadi peningkatan yang signifikan pada nilai  $R^2$  (Gambar 2) diuji 1 yang mencapai 94.85, menunjukkan bahwa model KNN memiliki kemampuan prediksi yang tepat hingga hampir 95%. Namun, terdapat penurunan pada uji 2 dan uji 3, di mana nilai  $R^2$  adalah 92.54 dan 92.26, secara berturut-turut. Meskipun mengalami penurunan, nilai-nilai ini masih menunjukkan bahwa model KNN memiliki performa yang kuat dalam memprediksi porositas. Nilai RMSE (Gambar 3) menunjukkan penurunan yang signifikan diuji 1 menjadi 0.0167 dibandingkan model RL. Hal ini mengindikasikan peningkatan dalam akurasi prediksi dibandingkan dengan model RL. Namun, nilai RMSE meningkat sedikit pada uji 2 dan uji 3 menjadi 0.0201 dan 0.0204, menandakan sedikit kenaikan dalam kesalahan prediksi dibandingkan dengan uji 1 pada model KNN. Nilai MAE (Gambar 4) diuji 1 sangat baik, yakni sebesar 0.0047. Hal ini menunjukkan adanya peningkatan akurasi yang lebih baik dibandingkan dengan model RL. Nilai MAE pada model KNN ini mengalami sedikit kenaikan pada uji 2 dan uji 3, menjadi 0.0053 dan 0.0056. Meski terjadi kenaikan, nilai-nilai tersebut tetap lebih rendah dari nilai MAE yang dihasilkan oleh model RL, mengindikasikan bahwa model KNN secara keseluruhan memiliki kesalahan prediksi yang lebih kecil.

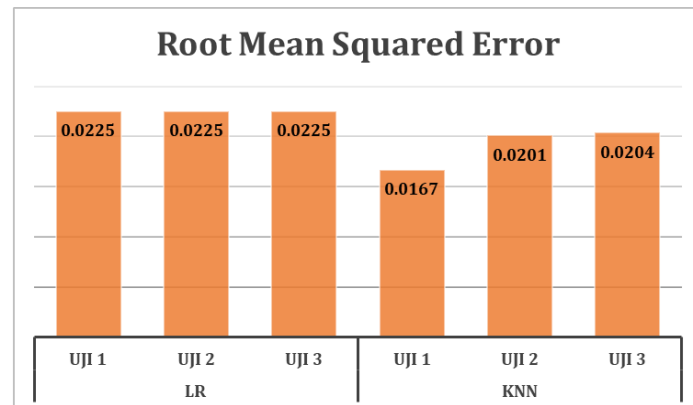
Model KNN menunjukkan performa yang lebih baik jika dilihat dari nilai  $R^2$  yang mendekati 95% tepat memprediksi porositas pada uji 1. Model KNN juga menunjukkan performa yang lebih baik dengan RMSE yang lebih rendah dibandingkan dengan model RL, menunjukkan kesalahan prediksi yang lebih kecil dalam estimasi porositas. Model KNN khususnya mengungguli model regresi linier pada uji 1, walaupun kedua model mengalami peningkatan RMSE pada uji 2 dan uji 3. Selain itu, berdasarkan nilai MAE, model KNN secara konsisten memberikan prediksi yang lebih akurat dibandingkan dengan model RL.





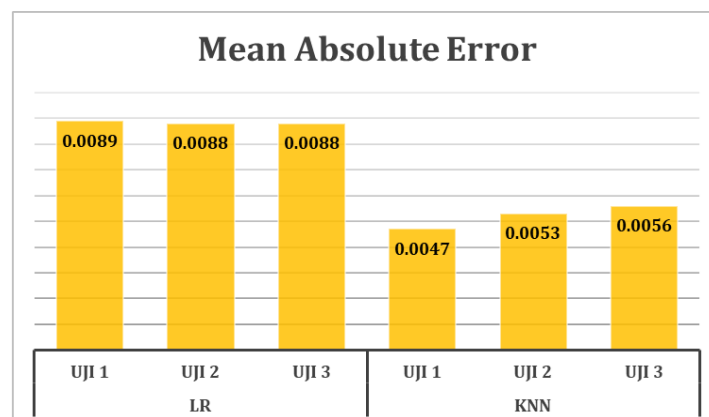
\*Higher = Good

Gambar 2. Perbandingan nilai  $R^2$  (*coefficient of determination*) pada Regresi Linier (LR) dan *K-Nearest Neighbors* (KNN) dengan 3 skenario berbeda.



\*Lower = Good

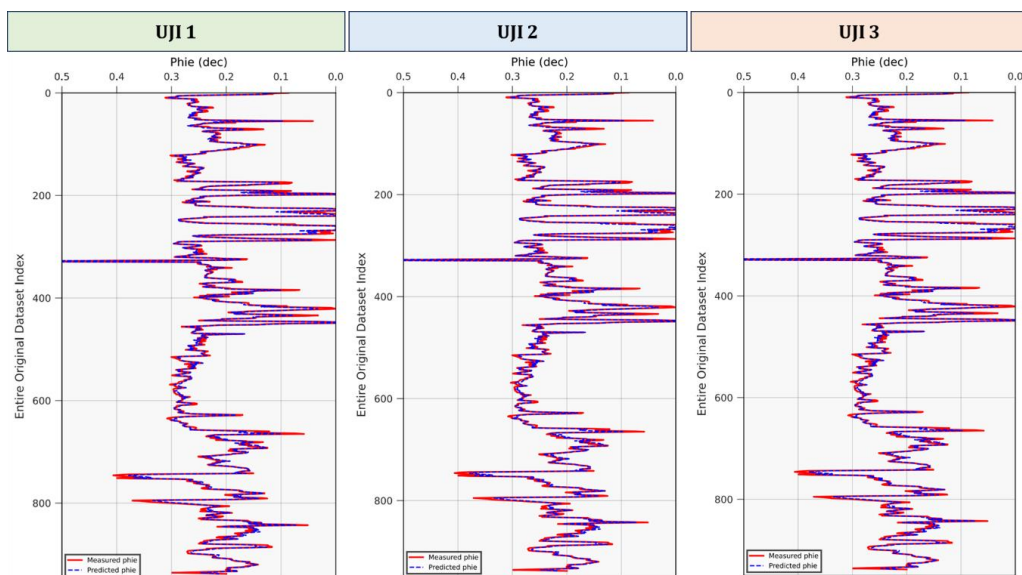
Gambar 3. Perbandingan nilai RMSE (*root mean squared error*) pada Regresi Linier (LR) dan *K-Nearest Neighbors* (KNN) dengan 3 skenario berbeda.



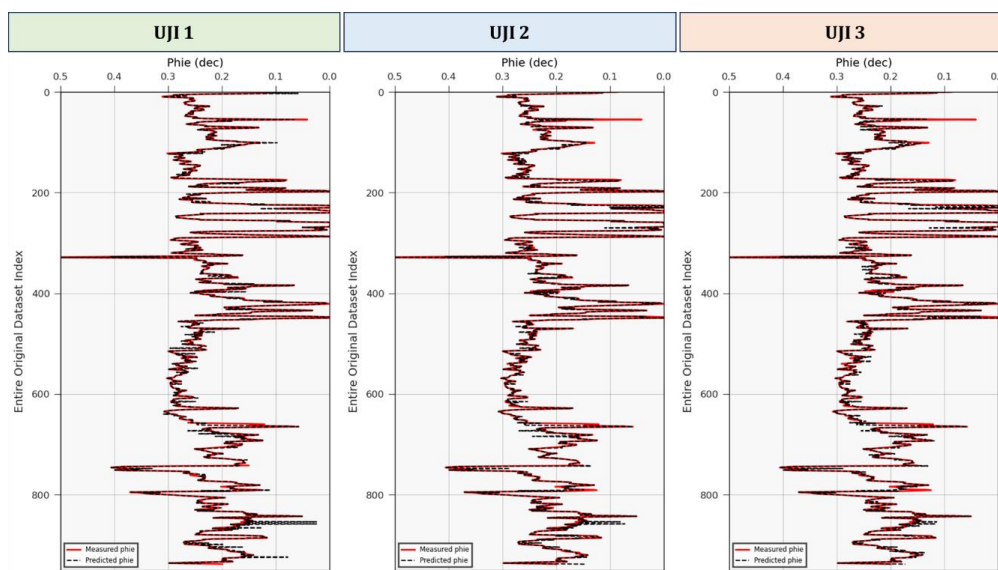
\*Lower = Good

Gambar 4. Perbandingan nilai MAE (*mean absolute error*) pada Regresi Linier (LR) dan *K-Nearest Neighbors* (KNN) dengan 3 skenario berbeda.

Hal ini menegaskan temuan sebelumnya bahwa KNN lebih cocok untuk digunakan dalam prediksi porositas pada data yang digunakan dalam penelitian ini. Secara umum, penelitian ini mendemonstrasikan bahwa model KNN mengungguli model RL dalam hal kemampuan memprediksi porositas menggunakan data log pada tiga skenario. Kedua model menunjukkan kinerja prediksi yang tinggi, namun model KNN terutama pada uji 1 menunjukkan keunggulan yang lebih baik. Uji 1 menjadi uji yang menghasilkan nilai akurasi paling baik hanya dengan menggunakan 3 data log (prediktor). Penelitian ini memberikan wawasan bahwa pemilihan model dan prediktor yang tepat sangat penting dalam meningkatkan akurasi prediksi dalam pembelajaran mesin. Prediksi porositas pada tiga skenario pada model RL dapat dilihat pada gambar 5 dan pada model KNN pada gambar 6.



Gambar 5. Prediksi porositas pada model Regresi Linier



Gambar 6. Prediksi porositas pada model *K-Nearest Neighbors*

### SIMPULAN DAN SARAN

Model regresi linier berganda menunjukkan konsistensi dalam nilai  $R^2$  sekitar 90.6%, menandakan efektivitas prediksi yang tinggi. Kesalahan prediksi (RMSE dan MAE) model RL relatif konstan, menunjukkan stabilitas dalam prediksi meskipun dengan perubahan prediktor.

Model *k-nearest neighbors*, menunjukkan performa yang lebih unggul, terutama pada uji 1 dengan nilai  $R^2$  mendekati 95%. Meskipun nilai  $R^2$  mengalami sedikit penurunan pada uji 2 dan 3, tetap menunjukkan performa yang kuat. Model KNN memiliki RMSE dan MAE yang lebih rendah dibandingkan dengan model RL, menunjukkan akurasi yang lebih baik dalam prediksi porositas.

Secara umum, model KNN terbukti lebih efektif dibandingkan dengan model RL dalam memprediksi porositas, terutama pada uji 1 yang menggunakan data prediktor GR, RDEP dan NEU.

Hasil ini menegaskan pentingnya pemilihan model dan prediktor yang tepat dalam meningkatkan akurasi prediksi dalam pembelajaran mesin.

### UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada semua individu dan entitas yang telah memberikan dukungan selama pelaksanaan penelitian ini. Terima kasih khusus diucapkan kepada panitia dan *reviewer* Seminar Nasional Rekayasa, Sains, dan Teknologi (SNARSTEK) tahun 2023 dan Jurnal Teknik (UMT) atas kesempatan yang diberikan sehingga memungkinkan publikasi penelitian ini.

### DAFTAR PUSTAKA

- Irajian, A. A., Bazargani-Guilani, K., Mahari, R., & Solgi, A. (2016). Porosity and rock-typing in hydrocarbon reservoirs case study in upper member of dalam formation in kish gas field, South of Zagros, Iran. *Open Journal of Geology*, 6(06), 399.
- Wood, D. A. (2020). Predicting porosity, permeability and water saturation applying an optimized nearest-neighbour, machine-learning and data-mining network of well-log data. *Journal of Petroleum Science and Engineering*, 184, 106587.
- Mohammadian, E., Kheirollahi, M., Liu, B., Ostadhassan, M., & Sabet, M. (2022). A case study of petrophysical rock typing and permeability prediction using machine learning in a heterogenous carbonate reservoir in Iran. *Scientific Reports*, 12(1), 4505.
- Mahdy, A., Zakaria, W., Helmi, A., Helaly, A. S., & Mahmoud, A. M. (2023). Machine learning approach for core permeability prediction from well logs in Sandstone Reservoir, Mediterranean Sea, Egypt. *Journal of Applied Geophysics*, 105249.
- Pratama, H. (2019). Machine learning: using optimized KNN (K-Nearest Neighbors) to predict the facies classifications. In *The 13th SEGJ International Symposium*, Tokyo, Japan, 12-14 November 2018 (pp. 538-541). Society of Exploration Geophysicists and Society of Exploration Geophysicists of Japan.



- Kurniadi, F. I., Rohmana, R. C., & Taufani, L. (2023). Local mean imputation for handling missing value to provide more accurate facies classification. *Procedia Computer Science*, 216, 301-309.
- Kurniadi, F. I., & Rohmana, R. C. (2023). Enhancing Lithology Classification Performance Through Random Forest, COPOD, and Bayesian Optimization. In *2023 10th International Conference on ICT for Smart Society (ICISS)* (pp. 1-5). IEEE.
- Ahmadi, M. A., & Chen, Z. (2019). Comparison of machine learning methods for estimating permeability and porosity of oil reservoirs via petro-physical logs. *Petroleum*, 5(3), 271-284.
- Sudakov, O., Burnaev, E., & Koroteev, D. (2019). Driving digital rock towards machine learning: Predicting permeability with gradient boosting and deep neural networks. *Computers & geosciences*, 127, 91-98.
- Cao, C. (2023). Prediction of concrete porosity using machine learning. *Results in Engineering*, 17, 100794.
- Male, F., & Duncan, I. J. (2020). Lessons for machine learning from the analysis of porosity-permeability transforms for carbonate reservoirs. *Journal of Petroleum Science and Engineering*, 187, 106825.
- Kearns, M. (1996). A Bound on the Error of Cross Validation Using the Approximation and Estimation Rates, with Consequences for Training-Test Split, *Neural Information Processing 8*, DS Touretzky, M, C. Mozer and ME Hasselmo.
- Turney, S. (2023). Coefficient of Determination (R2) Calculation & interpretation. Scribbr. <https://www.scribbr.com/statistics/coefficient-of-determination/>
- Zach. (2021). How to interpret Root Mean Square Error (RMSE). Statology. <https://www.statology.org/how-to-interpret-rmse/>
- Ahmed, M. W. (2023). Understanding mean Absolute error (MAE) in Regression: A Practical guide. Medium. <https://medium.com/@m.waqar.ahmed/understanding-mean-absolute-error-mae-in-regression-a-practical-guide-26e80ebb97df>
- Otoo, D., & Hodgetts, D. (2021). Porosity and permeability prediction through forward stratigraphic simulations using GPM™ and Petrel™: application in shallow marine depositional settings. *Geoscientific Model Development*, 14(4), 2075-2095.
- Pickle — Python object serialization. (n.d.). Python Documentation. <https://docs.python.org/3/library/pickle.html>
- Volve field data set. (n.d.). <https://www.equinor.com/energy/volve-data-sharing>