



## Penerapan Model *Random Forest* untuk Prediksi Penyebaran Porositas pada Lapangan 'X'

Nugraha Septiawisal Putra<sup>1)</sup>, Abdul Gafar Karim<sup>1)</sup>

<sup>1)</sup>Program Studi S1 Teknik Perminyakan, Sekolah Tinggi Teknologi Migas Balikpapan

E-mail: [nugrahaseptiawisa199@gmail.com](mailto:nugrahaseptiawisa199@gmail.com)

### ABSTRAK

Kegiatan karakterisasi reservoir memiliki peran yang penting untuk dapat mengidentifikasi sifat-sifat batuan dan fluida, serta dapat mengestimasi besar cadangan hidrokarbon di bawah permukaan. Estimasi metode tradisional memiliki akurasi yang rendah, sangat memakan waktu, dan galat yang lebih besar selama memprediksi porositas dan permeabilitas reservoir menggunakan data *well logging* serta pengukuran *sample core* yang cukup memakan biaya. Dalam penelitian ini, metode *machine learning Random Forest* diterapkan sebagai solusi yang cepat dan hemat biaya untuk mengestimasi parameter reservoir dari data *well logging*. Enam model telah dikembangkan pada kondisi yang berbeda (jumlah fitur dan transformasi data) yang dibagi menjadi 85% training data dan 15% test data. Penentuan *hyperparameters* merupakan hal yang penting dikarenakan akan sangat berpengaruh ke performa pada setiap model. Hasil *Exploratory Data Analysis* menunjukkan bahwa fitur-fitur yang dapat digunakan sebagai input data untuk memprediksi nilai porositas *core* adalah fitur GR (-0.54), RHOB (-0.51), LLS (-0.44), LLD (-0.4), dan CXO (0.32). Sehingga dilanjutkan ke dalam proses pelatihan model dan ditemukan bahwa model B dengan fitur tiga teratas (GR, RHOB, dan LLS) tanpa transformasi data mendapatkan hasil yang optimal dengan mendapatkan nilai  $R^2$  sebesar 0.8 dan RMSE sebesar 0.034 dengan menggunakan *hyperparameter*  $n\_estimators = 19$  dan  $random\_state = 42$ . Penelitian dilanjutkan dengan melakukan pelatihan model ulang dengan menggunakan seluruh data *well logging* untuk dapat mengembangkan model prediksi porositas. Prediksi porositas *core* berhasil dilakukan hingga 1.393 data pada baris kosong yang terdapat pada kolom porositas *core*.

Kata Kunci: karakterisasi reservoir, *machine learning*, *random forest*, prediksi porositas, *well logging*

### ABSTRACT

*Reservoir characterization activities have an important role in identifying rock and fluid properties and estimating the amount of subsurface hydrocarbon reserves. Traditional estimation methods have low accuracy, are very time-consuming, and have larger errors while predicting reservoir porosity and permeability using well-logging data and sample core measurements which are quite costly. This study applied the Random Forest machine learning method as a fast, cost-effective solution for estimating reservoir parameters from well-logging data. Six models have been developed under different conditions (number of features and data transformation,) which are divided into 85% training data and 15% test data. The determination of hyperparameters is important because it will greatly affect the performance of each model. Exploratory data analysis results show that features that can be used as input data to predict core porosity values are GR (-0.54), RHOB (-0.51), LLS (-0.44), LLD (-0.4), and CXO (0.32) features. ). So it was continued into the model training process, and it was found that model B with the top three features (GR, RHOB, and LLS) without data transformation got optimal results by getting an  $R^2$  value of 0.8 and an RMSE of 0.034 using the hyperparameter  $n\_estimators=19$  and  $random\_state=42$ . The research was continued by conducting re-model training using all well-logging data to be able to develop a porosity prediction model. Core porosity prediction was successfully carried out for up to 1393 data in empty rows contained in the core porosity column.*

Keywords: reservoir characterization, machine learning, random forest, porosity prediction, well logging

## 1. Pendahuluan

Kegiatan karakterisasi reservoir memiliki peran yang penting untuk dapat mengidentifikasi sifat-sifat batuan dan fluida, serta dapat mengestimasi besar cadangan hidrokarbon di bawah permukaan. Karakterisasi reservoir merupakan suatu kegiatan yang sangat kompleks mengingat banyak reservoir yang memiliki heterogenitas tinggi yang terbentuk secara alami. Dua di antara karakteristik yang terdapat pada reservoir adalah porositas dan permeabilitas yang merupakan karakteristik utama dalam proses karakterisasi reservoir. Interpretasi *well logging* banyak diterapkan untuk dapat mengestimasi nilai porositas dan permeabilitas pada kedalaman yang bervariasi dikarenakan biaya yang minimal dibandingkan proses *sample core* (Helle & Bhatt, 2002; Lim & Kim, 2004, sebagaimana dikutip dalam Abdel Azim & Aljehani, 2022). Sebaliknya, proses *sample core* merupakan sumber informasi yang terpercaya dikarenakan *sample core* dapat langsung diambil dari sumur bor. Hal ini menyebabkan sumur-sumur yang dilakukan *core data analysis* dapat dijadikan sebagai acuan untuk mengetahui kualitas dari hasil interpretasi dan perhitungan petrofisika untuk sumur-sumur lain yang akan dibor di sekitar pemboran acuan (Oetomo dkk., 2019).

Menurut Yao dkk., 2010; Félix & Muñoz, 2005 sebagaimana dikutip dalam Chen dkk., 2021 menjelaskan bahwa hasil estimasi metode tradisional memiliki akurasi yang rendah, sangat memakan waktu, dan kesalahan yang lebih besar selama memprediksi porositas dan permeabilitas reservoir menggunakan data *well logging*. Serta pengukuran laboratorium *sample core* merupakan kegiatan yang cukup memakan biaya dan terbatas (Chen dkk., 2021). Sehingga untuk mengatasi hal tersebut, terdapat solusi alternatif dengan menggunakan perhitungan regresi untuk melakukan prediksi penyebaran porositas di Lapangan 'X'. Metode *machine learning* dapat diaplikasikan sebagai solusi yang cepat dan hemat biaya untuk mengestimasi parameter reservoir dari hubungan gabungan yang kompleks hingga pengukuran tidak langsung, dalam hal ini diwakili oleh *well log* (Huang dkk., 1997; Helle & Bhatt, 2002; Hamada dkk., 2020 sebagaimana dikutip dalam Andersen dkk., 2022).

Sehingga, dalam penelitian ini bertujuan untuk menentukan variabel yang dapat digunakan untuk proses model *training*, mendapatkan model yang optimal berdasarkan nilai  $R^2$  dan RMSE pada masing-masing model, serta melakukan prediksi penyebaran porositas dengan menggunakan metode *machine learning* dengan algoritma *Random Forest*. Model *Random Forest* telah berhasil didemonstrasikan pada penelitian sebelumnya untuk memprediksi *variable oil recovery* dan analisis sensitivitas dari parameter *Low Salinity Water Injection* (Hidayat & Astsauri, 2021), dan untuk meramalkan kumulatif produksi minyak, performa CO<sub>2</sub>-WAG dalam kondisi parameter injeksi yang bervariasi (Li dkk., 2022). Data yang tersedia dalam penelitian ini terdiri dari data petrofisika dan data *core* yang diambil dari lima sumur yang berbeda pada Lapangan 'X'. Kemudian masing-masing model dikembangkan dengan menggunakan variabel yang relevan, dan mencari model yang memiliki kemampuan prediksi yang baik. Membagi data petrofisika menjadi 85% data *training* dari lima sumur berbeda yang diacak untuk mengembangkan model *machine learning*. Setelah model berhasil dikembangkan, 15% data *testing* digunakan untuk menguji model, dan membandingkannya dengan data porositas *core* yang sebenarnya kemudian dilanjutkan dengan melakukan prediksi pada data yang *missing* di variabel porositas dengan menggunakan model yang telah dikembangkan sebelumnya.

## 2. Metode Penelitian

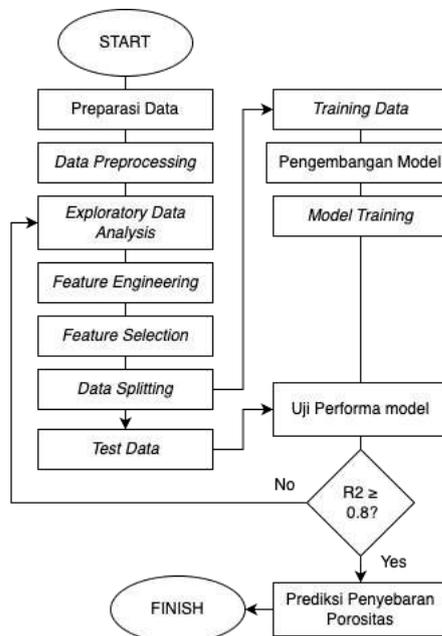
Pada bagian ini, akan menjelaskan rincian dasar teori mengenai metode kuantitatif yang digunakan. Pada bagian ini terdiri dari preparasi data, *data preprocessing*, *exploratory data analysis* (EDA), *feature engineering*, pemilihan fitur, pengembangan model *Random Forest*, *model evaluation metrics* yang digunakan untuk menguji model pada *test data*.

### A. Diagram Alir

Diagram alir yang digunakan pada penelitian ini (dapat diilustrasikan sebagai Gambar 1) diterapkan untuk mempersiapkan model *Random Forest* yang dapat digunakan untuk memprediksi porositas *core* berdasarkan data *well logging*.

1. Data dipersiapkan untuk dapat digunakan dalam penelitian ini, berupa data petrofisika sebagai *independent variable* dan data *core* sebagai *dependent variable* (*target variable*).
2. Ketika data telah dipersiapkan, maka perlu untuk melakukan data *preprocessing* terlebih dahulu dengan melihat apakah data terdapat *missing values*, *noisy data*, menghapus *outliers* atau

- transformasi variabel. Serta mengambil keputusan apakah data perlu untuk dibersihkan (*cleaning*) atau ditransformasi.
- Setelah data dibersihkan atau ditransformasi, dapat dilanjutkan ke proses *exploratory data analysis* (EDA) yang bertujuan untuk dapat untuk menggali informasi berupa pola, anomali, uji hipotesis, yang dapat dilakukan secara grafis dan statistik.
  - Feature engineering* dilakukan untuk mendapatkan bentuk distribusi data yang normal, mengatasi *outliers*, dan proses *encoding* untuk data *categorical* agar dapat digunakan pada proses regresi.
  - Setelah *exploratory data analysis* (EDA) dan *feature engineering* dilakukan, maka dapat menentukan fitur yang tepat untuk digunakan untuk proses *training* model nantinya (dapat dilihat dari nilai korelasi masing-masing fitur terhadap *target variable*).
  - Melanjutkan ke dalam proses data *splitting* yang bertujuan untuk membagi data-data untuk digunakan sebagai *training data* dan sebagai *unseen data* (*test data*).
  - Mengembangkan model *random forest* dengan mencari parameter-parameter yang sesuai, dan dilakukan proses *training model* terhadap data-data yang telah digunakan sebagai *input*.
  - Ketika proses pengembangan model telah berhasil, maka hasil *training* yang telah dilakukan pada model diuji performa tersebut dan mengamati nilai  $R^2$  yang didapatkan. Jika nilai  $R^2$  yang didapatkan tidak lebih dari 0,8, maka kembali ke proses *exploratory data analysis* (EDA) dan melakukan *hyperparameter tuning*.
  - Ketika model terbaik telah didapatkan, dan mendapatkan nilai  $R^2 \geq 0,8$ , maka dapat dilanjutkan ke proses prediksi keseluruhan baris yang kosong pada kolom variabel porositas dengan menggunakan model yang telah didapatkan sebelumnya.



Gambar 1. Diagram alir penelitian

## B. *Random Forest*

*Random forest* merupakan pendekatan *ensemble machine learning* yang diusulkan oleh Breiman (2001), yang memiliki kelebihan dalam proses interpretasi, kenyamanan, dan komputasi dengan kecepatan yang tinggi. *Random forest* dibangun secara *independen* oleh beberapa *decision trees* (Wang dkk., 2022). *Decision Trees* merupakan sebuah algoritma *machine learning* yang sering digunakan dalam tugas klasifikasi dan regresi. *Decision trees* seperti pohon yang dimana setiap *internal node* mempresentasikan fitur/variable, setiap cabang mewakili keputusan atau hasil berdasarkan nilai fitur yang digunakan, dan setiap *leaf node* mewakili keputusan akhir atau hasil yang diprediksi.

*Random forest* standar dibangun dengan menggunakan *Bootstrap Aggregating* atau *Bagging*. Algoritma ini secara acak memilih sampel dari data mentah sehingga proses *train* dari setiap *regressor*

dasar dalam *ensambel* yang tidak bergantung pada yang lain. Alur spesifik dalam konstruksi random forest adalah sebagai berikut: Pertama, pada setiap *node decision trees*, *variable predictor* diambil sampelnya secara acak, Kemudian, algoritma dapat menemukan minimal *residual sum of squares* (RSS) untuk regresi. Selanjutnya, data dibagi menjadi subset “*in-bag*” untuk *training*, dan subset “*out-of-bag* (OOB)” untuk *testing*. Terakhir *decision trees* digabungkan melalui rata-rata regresi untuk membentuk hasil prediksi akhir (Makariou dkk., 2021)

### C. Model Evaluation Metrics

Dalam penelitian ini, dua *model evaluation metrics* digunakan untuk dapat melihat performa model dalam memprediksi nilai porositas *core*. *Model evaluation metrics* yang digunakan antara lain *coefficient of determination* ( $R^2$ ), dan *root mean squared error* (RMSE). Komponen  $R^2$  menunjukkan akurasi prediksi dari metode yang diusulkan, sedangkan komponen RMSE menunjukkan nilai standar deviasi antara nilai prediksi dan nilai aktualnya.

$$R^2 = \frac{\sum_{i=1}^n (y_{pred} - y_{mean})^2}{\sum_{i=1}^n (y_{actual} - y_{mean})^2} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{pred} - y_{actual})^2} \quad (2)$$

### D. Preparasi Data

Pengumpulan data petrofisika dan data *core* pada setiap sumur ditabulasi menjadi satu dataset primer yang terdiri dari data *input* dan *output*. Data petrofisika (*well log*) yang dijadikan sebagai *variable input* dan data porositas *core* dijadikan sebagai variabel *output*. Untuk lebih detail mengenai data *input* dan data *output*, dapat dilihat pada Tabel 1.

Tabel 1. Urutan data *input* dan *output*

| No. | Variabel  | Kategori         | Jenis Variabel               |
|-----|-----------|------------------|------------------------------|
| 1   | WELLNAME  | Data Petrofisika | <i>Independent Variables</i> |
| 2   | DEPTH     | Data Petrofisika | <i>Independent Variables</i> |
| 3   | CALI      | Data Petrofisika | <i>Independent Variables</i> |
| 4   | CXO       | Data Petrofisika | <i>Independent Variables</i> |
| 5   | SP        | Data Petrofisika | <i>Independent Variables</i> |
| 6   | DT        | Data Petrofisika | <i>Independent Variables</i> |
| 7   | GR        | Data Petrofisika | <i>Independent Variables</i> |
| 8   | LLD       | Data Petrofisika | <i>Independent Variables</i> |
| 9   | LLS       | Data Petrofisika | <i>Independent Variables</i> |
| 10  | MLL       | Data Petrofisika | <i>Independent Variables</i> |
| 11  | NPHI      | Data Petrofisika | <i>Independent Variables</i> |
| 12  | PEF       | Data Petrofisika | <i>Independent Variables</i> |
| 13  | RHOB      | Data Petrofisika | <i>Independent Variables</i> |
| 14  | PHI_CORE  | Data Core        | <i>Dependent Variables</i>   |
| 15  | PERM_CORE | Data Core        | <i>Dependent Variables</i>   |

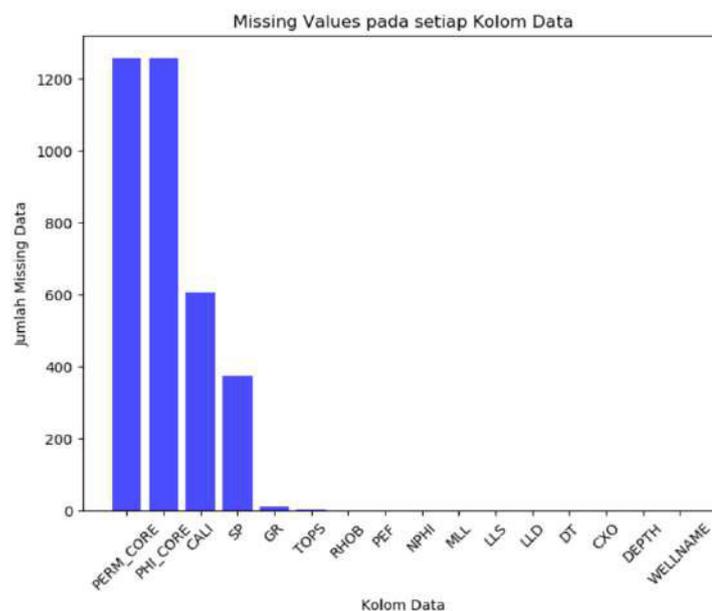
### E. Data Preprocessing

*Data preprocessing* sangat penting dilakukan sebelum penggunaan *machine learning*. *Data preprocessing* merupakan konsep mengubah data mentah menjadi kumpulan data yang bersih. Dataset

diproses untuk memeriksa apakah terdapat nilai yang hilang (*missing values*), data yang berisik (*noisy data*), dan ketidakkonsistenan lainnya sebelum dilakukan pemodelan *machine learning* (Singh dkk. 2021). *Data preprocessing* terbagi menjadi *data cleaning*, *data integration*, *data transformation* dan *data reduction*. Menurut Yan-li & Jia (2012), Teknik-teknik yang dapat digunakan dalam proses data *preprocessing* adalah sebagai berikut:

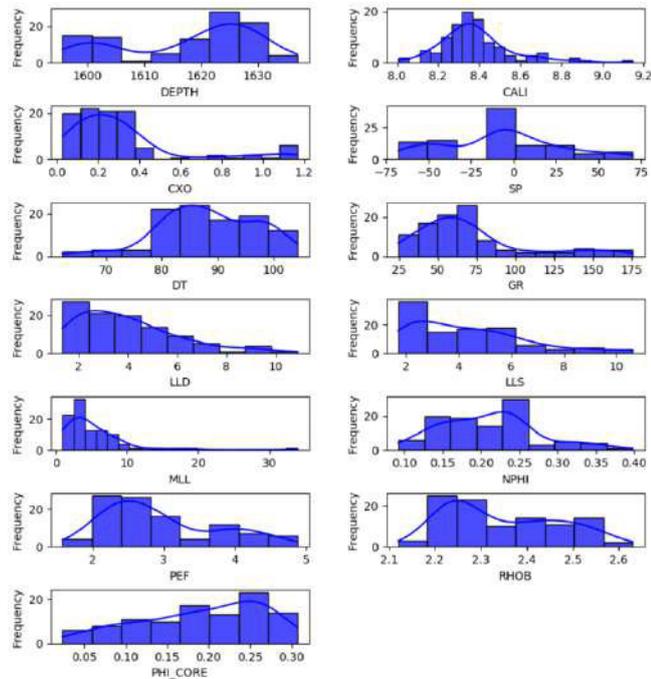
1. *Data Cleaning*: Dengan mengisi *null-data*, menghaluskan data *noise*, mengidentifikasi dan menghapus data yang diisolasi, dan menyelesaikan ketidakkonsistenan untuk mencapai tujuan utama (*data bersih*).
2. *Data Integration*: Simpan data dari beberapa sumber ke dalam penyimpanan data yang konsisten, sumber data ini dapat mencakup beberapa database, data cube, atau file biasa.
3. *Data Transformation*: Data mentah diubah atau dimodifikasi menjadi skala yang sesuai untuk dianalisis dan pemodelan *machine learning*. Dapat dilakukan dengan metode *normalization* atau *standardization*, *log transformation*, dan *categorical variables encoding*.
4. *Data Reduction*: Sekumpulan data asli berisi informasi yang berulang, tidak relevan, atau berisik, atau saat bekerja dengan data dimensi tinggi yang dapat menyebabkan adanya tantangan komputasi pada algoritma tertentu.

Dalam penelitian ini, dilakukan pengumpulan *data missing* guna untuk dapat menggunakan data yang bersih sebagai *input* model *training* nantinya. Ditemukan 6 kolom dengan nilai *missing values* yang bervariasi (dapat dilihat pada Gambar 2) kemudian dilakukan proses pembersihan data dengan menyisakan data sebesar 109 data.



Gambar 2. Jumlah data yang hilang (*missing values*)

Sebelum data diproses untuk pemodelan, butuh untuk dilakukan identifikasi penyebaran data pada masing-masing variabel dengan menggunakan alat visualisasi bernama histogram. Didapatkan enam variabel yang memiliki *right-skewed distribution* dengan nilai *skewness* lebih dari 0,75 (*skew limit*) yaitu MLL, CXO, GR, CALI, LLD, dan LLS (dapat dilihat pada Gambar 3). Sehingga dalam penelitian ini, akan dilakukan pengembangan model pada kondisi data yang tidak ditransform dan data yang ditransform penyebarannya.

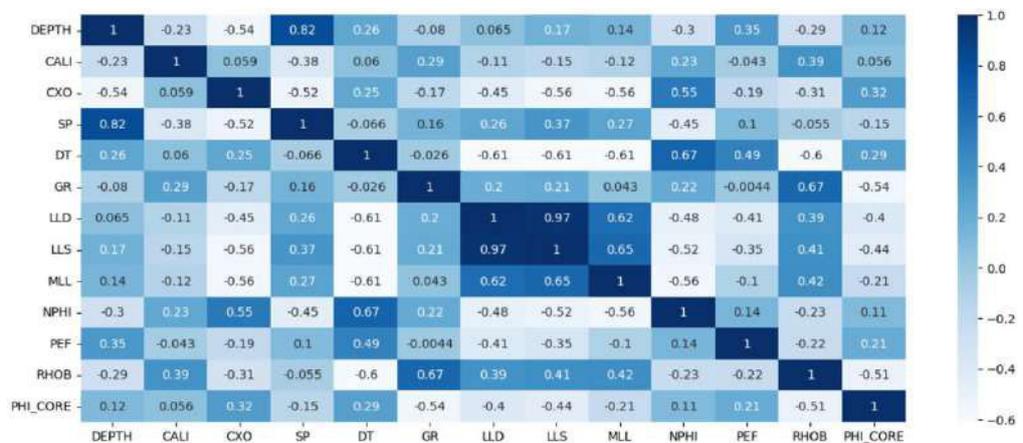


Gambar 3. Distribusi data pada masing-masing variabel

### 3. Hasil dan Pembahasan

#### A. Hasil *Exploratory Data Analysis* (EDA)

*Exploratory data analysis* (EDA) dilakukan untuk melihat korelasi masing-masing parameter *input* terhadap parameter *output*. *Pearson correlation coefficient* dihitung berdasarkan keseluruhan data-data yang diambil dari masing-masing sumur. Pada Gambar 4 dapat dilihat bahwa dilakukannya perhitungan nilai *pearson correlation coefficient* pada variable porositas *core* terhadap variabel lainnya (tidak termasuk permeabilitas *core*). Dapat dilihat bahwa tidak ada data variabel petrofisika yang memiliki nilai korelasi yang tinggi terhadap variabel porositas *core*. Namun terdapat fitur yang masih berpotensi untuk digunakan sebagai variabel *input* seperti variabel GR (-0.54), RHOB (-0.51), LLS (-0.44), LLD (-0.4), dan CXO (0.32).



Gambar 4. Nilai *Pearson correlation coefficient* masing-masing variabel

#### B. Pengembangan Model

Dalam penelitian ini, model *Random Forest* dikembangkan dengan menggunakan data petrofisika dengan jumlah *potential input feature* tertentu dan menerapkan kondisi dengan dan tanpa *variable transform* untuk melihat performa *random forest* pada setiap kondisi. Sehingga, menghasilkan enam model dengan kondisi yang berbeda. Penentuan *hyperparameter model* merupakan kegiatan yang penting sebelum melakukan *model training*. Dalam bahasa pemrograman *python*, *hyperparameter random forest* yang umum digunakan antara lain adalah: *n\_estimators*, *max\_depth*, *min\_sample\_split*,

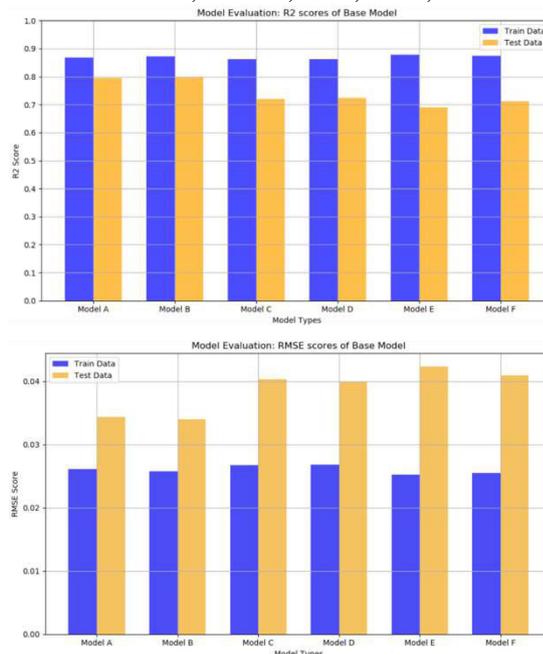
*max\_features*, dan *random\_state*. Dalam penelitian ini dilakukan pemodelan *random forest* dengan menggunakan *hyperparameter* yang dapat dilihat pada Tabel 2.

Tabel 2. Detail model yang digunakan pada kondisi berbeda

| <i>Inisial Model</i> | <i>Features &amp; Transform</i> | <i>Hyperparameters</i> |                     |
|----------------------|---------------------------------|------------------------|---------------------|
|                      |                                 | <i>n_estimators</i>    | <i>random_state</i> |
| A                    | 3 Variable + Log Transform      | 19                     | 42                  |
| B                    | 3 Variable + No Transform       | 17                     | 42                  |
| C                    | 4 Variable + Log Transform      | 15                     | 42                  |
| D                    | 4 Variable + No Transform       | 13                     | 42                  |
| E                    | 5 Variable + Log Transform      | 13                     | 42                  |
| F                    | 5 Variable + No Transform       | 13                     | 42                  |

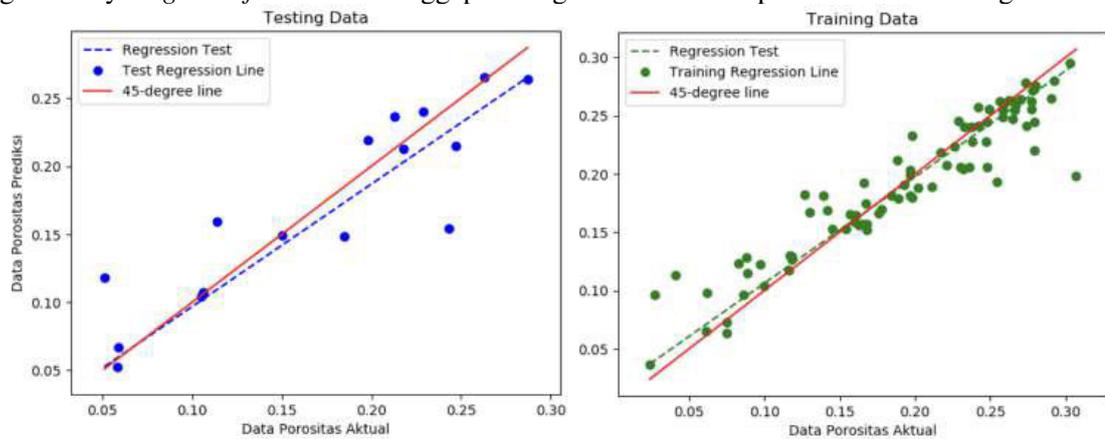
Dalam kegiatan *model training*, penilaian performa *random forest* dilakukan pada dua data set yang berbeda (*Training Data* dan *Test Data*). Pertama, tiga fitur digunakan untuk melatih model dengan dan tanpa *variable transform*. Berdasarkan fitur GR, RHOB, dan LLS didapatkan nilai  $R^2$  sebesar 0,7959 dan RMSE sebesar 0,0344 (dengan *log transform*) pada *test data* dan nilai  $R^2$  sebesar 0,8 dan RMSE sebesar 0,034 (tanpa *log transform*) pada *test data*. Selanjutnya dilakukan penambahan fitur kedua (yaitu LLD) dan menunjukkan bahwa dengan adanya penambahan fitur ini, tidak memberikan peningkatan terhadap model yang digunakan, dengan mendapatkan nilai  $R^2$  sebesar 0,72 dan RMSE sebesar 0,0403 (dengan *log transform*) pada *test data* dan nilai  $R^2$  sebesar 0,725 dan RMSE sebesar 0,0399 (tanpa *log transform*) pada *test data*. Kemudian dilanjutkan dengan menambahkan fitur ketiga (yaitu CXO) dan mengalami hal yang serupa dengan kondisi sebelumnya (tidak adanya peningkatan model), dengan nilai  $R^2$  sebesar 0,691 dan RMSE sebesar 0,0423 (dengan *log transform*) pada *test data* dan nilai  $R^2$  sebesar 0,712 dan RMSE sebesar 0,0409 (tanpa *log transform*) pada *test data*. Hasil pengembangan model pada kondisi yang berbeda dapat dilihat pada Gambar 5.

Tercatat bahwa Model A menggunakan fitur GR, RHOB, dan LLS dengan *log transform*, Model B menggunakan fitur GR, RHOB, dan LLS tanpa *log transform*, Model C menggunakan fitur GR, RHOB, LLS, dan LLD dengan *log transform*, Model D menggunakan fitur GR, RHOB, LLS, dan LLD tanpa *log transform*, Model E menggunakan fitur GR, RHOB, LLS, LLD, dan CXO dengan *log transform*, Model F menggunakan fitur GR, RHOB, LLS, LLD, dan CXO tanpa *log transform*.



Gambar 5. Performa model *random forest* pada kondisi yang berbeda

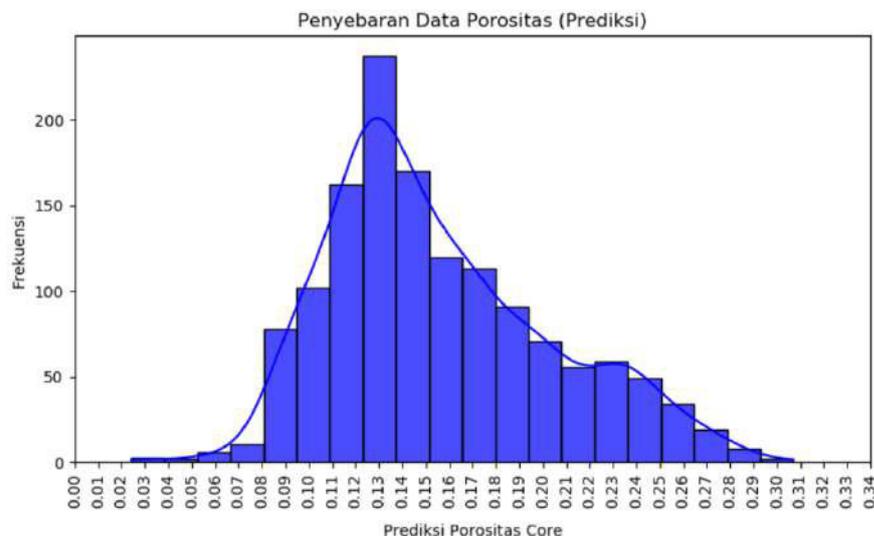
Sehingga berdasarkan pengujian enam model pada kondisi yang berbeda, dapat dikatakan bahwa model B merupakan model terbaik untuk digunakan dalam memprediksi penyebaran porositas *core* dengan mendapatkan nilai  $R^2$  sebesar 0,8 dan RMSE sebesar 0,034. Pada Gambar 6 dapat dilihat bahwa sebagian besar nilai prediksi menggunakan model B tidak jauh dari garis linear sempurna dan memiliki kinerja yang baik dibandingkan dengan model *random forest* yang lainnya, sehingga model *Random Forest* dengan  $n\_estimators = 19$  dan  $random\_state = 42$  dapat menjadi *hyperparameter* yang tepat untuk memprediksi penyebaran porositas *core* berdasarkan data *well logging*. Hal ini juga disebabkan oleh tiga fitur (GR, RHOB, dan LLS) yang memiliki korelasi tertinggi terhadap *variable target* (porositas *core*) dibandingkan dengan fitur yang lainnya. Pada kasus ini terdapat catatan bahwa dengan adanya *log transformation* dianggap kurang relevan selama proses model training.



Gambar 6. Performa model B dalam memprediksi porositas *core*

### C. Prediksi Nilai Porositas

Berdasarkan performa prediksi didapatkan oleh model B, penelitian dilanjutkan dengan melakukan pelatihan model ulang dengan menggunakan seluruh data *well logging* untuk dapat mengembangkan model prediksi porositas. Dengan *hyperparameter* dan fitur yang telah didapatkan sebelumnya digunakan kembali untuk dapat memprediksi/generate secara regresi nilai porositas *core* yang kosong pada setiap sumur. Hasil prediksi data porositas *core* dapat divisualisasikan pada Gambar 7.



Gambar 7. Penyebaran data porositas (prediksi)

Sehingga berdasarkan hasil yang didapatkan oleh model yang digunakan untuk memprediksi penyebaran data porositas pada setiap sumur (total 1.393 data yang diprediksi), dapat dilihat bahwa frekuensi tertinggi didapatkan pada data porositas yang memiliki nilai di antara 0,12 hingga 0,14 dengan nilai *skewness* sebesar 0,439 yang mana data porositas yang diprediksi hampir simetris. Data

porositas core yang diprediksi dapat ditabulasi secara statistik deskriptif yang dapat dilihat pada Tabel 3.

Tabel 3. Statistik deskriptif dari variable PHI\_CORE

|               | <i>PHI_CORE</i> |
|---------------|-----------------|
| <b>Jumlah</b> | 1393            |
| <i>Mean</i>   | 0.155593        |
| <i>Std</i>    | 0.047848        |
| <i>Min</i>    | 0.024           |
| <b>25%</b>    | 0.121824        |
| <b>50%</b>    | 0.144824        |
| <b>75%</b>    | 0.185353        |
| <b>Max</b>    | 0.307           |

#### 4. Kesimpulan

Penelitian ini menyajikan analisis *machine learning* untuk memprediksi data porositas berdasarkan data *well logging*. Pengembangan model *Random Forest* dilakukan dengan menyiapkan enam model dengan kondisi yang berbeda (jumlah fitur yang digunakan dan data yang ditransformasi) untuk dapat melihat performa masing-masing model. Model dengan performa terbaik akan digunakan untuk proses pelatihan model ulang dengan menggunakan seluruh data *well logging*, dan akhirnya berhasil dilakukan prediksi penyebaran data porositas *core* untuk baris yang kosong pada kolom porositas *core* dengan menggunakan fitur dan *hyperparameter* model yang optimal. Pada penelitian ini dapat disimpulkan:

- Fitur terbaik yang dapat digunakan untuk memprediksi data porositas *core* adalah: log sinar-gamma (GR), log densitas (RHOB), dan log resistivitas dangkal (LLS). Hal ini didasarkan pada nilai *pearson correlation coefficient* yang didapatkan serta menghasilkan performa terbaik pada *test data* dan penambahan fitur lainnya tidak dapat meningkatkan performa model *random forest* yang digunakan.
- Didapatkan model *random forest* dengan performa terbaik untuk memprediksi nilai porositas dengan nilai  $R^2$  sebesar 0,8 dan RMSE sebesar 0,034. Model tersebut adalah model B dengan *hyperparameter*  $n\_estimator = 19$ , dan  $random\_state = 42$  dan berdasarkan pengujian model bahwa penambahan proses *variable transformation* tidak dapat meningkatkan performa model *random forest*.
- Berhasil melakukan prediksi penyebaran porositas *core* pada baris kosong dalam kolom porositas *core* sebanyak 1.393 data dengan menggunakan fitur terbaik dan model terbaik yang telah didapatkan sebelumnya.

#### 5. Daftar Pustaka

- Andersen, P. Ø., Skjeldal, M., & Augustsson, C. (2022). Machine Learning Based Prediction of Porosity and Water Saturation from Varg Field Reservoir Well Logs. In *SPE EuropEC - Europe Energy Conference featured at the 83rd EAGE Annual Conference & Exhibition* (p. D041S010R005). <https://doi.org/10.2118/209659-MS>
- Abdel Azim, R., & Aljehani, A. (2022). Neural Network Model for Permeability Prediction from Reservoir Well Logs. *Processes*, 10(12). <https://doi.org/10.3390/pr10122587>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Chen, L., Lin, W., Chen, P., Jiang, S., Liu, L., & Hu, H. (2021). Porosity Prediction from Well Logs Using Back Propagation Neural Network Optimized by Genetic Algorithm in One Heterogeneous Oil Reservoirs of Ordos Basin, China. *Journal of Earth Science*, 32(4), 828–838. <https://doi.org/10.1007/s12583-020-1396-5>
- Hidayat, F., & Astsauri, T. M. S. (2022). Applied random forest for parameter sensitivity of low salinity water Injection (LSWI) implementation on carbonate reservoir. *Alexandria Engineering Journal*, 61(3), 2408–2417. <https://doi.org/10.1016/j.aej.2021.06.096>

- 
- Makariou, D., Barrieu, P., & Chen, Y. (2021). A random forest based approach for predicting spreads in the primary catastrophe bond market. *Insurance: Mathematics and Economics*, *101*, 140–162. <https://doi.org/10.1016/j.insmatheco.2021.07.003>
- Li, H., Gong, C., Liu, S., Xu, J., & Imani, G. (2022). Machine Learning-Assisted Prediction of Oil Production and CO<sub>2</sub> Storage Effect in CO<sub>2</sub>-Water-Alternating-Gas Injection (CO<sub>2</sub>-WAG). *Applied Sciences (Switzerland)*, *12*(21). <https://doi.org/10.3390/app122110958>
- Oetomo, M. A., Harmoko, U., & Yuliyanto, G. (2019). Reservoir characterization by petrophysical analysis and core data validation, a case study of the “x” field prospect zone. *Journal of Physics: Conference Series*, *1217*(1), 0–8. <https://doi.org/10.1088/1742-6596/1217/1/012015>
- Singh, P., Singh, N., Singh, K. K., & Singh, A. (2021). Diagnosing of disease using machine learning. In K. K. Singh, M. Elhoseny, A. Singh, & A. A. B. T.-M. L. and the I. of M. T. in H. Elngar (Eds.), *Machine Learning and the Internet of Medical Things in Healthcare* (pp. 89–111). Academic Press. <https://doi.org/10.1016/B978-0-12-821229-5.00003-3>
- Wang, M., Feng, D., Li, D., & Wang, J. (2022). Reservoir Parameter Prediction Based on the Neural Random Forest Model. *Frontiers in Earth Science*, *10*(May), 1–12. <https://doi.org/10.3389/feart.2022.888933>
- Yan-li, Z., & Jia, Z. (2012). Research on Data Preprocessing in Credit Card Consuming Behavior Mining. *Energy Procedia*, *17*, 638–643. <https://doi.org/10.1016/j.egypro.2012.02.147>