

PREDIKSI LAJU PENETRASI PENGEBORAN DENGAN MENGGUNAKAN PEMBELAJARAN MESIN PADA LAPANGAN GORGON DI AUSTRALIA

Muhammad Sayyid Anwar Al Mahbub ^{*1}, Dara Ayuda Maharsi ^{**1}

¹ Petroleum Engineering Study Program, Universitas Pertamina

*EMAIL

dara.maharsi@universitaspertamina.ac.id

KATA KUNCI

Laju penetrasi (*Rate of Penetration*; ROP), *machine learning*, Multilayer Perceptron (MLP), *Mean Absolute Percentage Error* (MAPE), Generalisasi Model

HISTORI ARTIKEL

Received: 5 September 2023

Accepted: 27 February 2025

CARA MENYITIR

Al Mahbub, M. S. A., Maharsi., D. A (2025). Prediksi laju penetrasi pengeboran dengan menggunakan pembelajaran mesin pada Lapangan Gorgon di Australia. *Subsurface*, 03(01) 59-71.

© 2025 SUBSURFACE. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY NC)

ABSTRAK

Studi ini menyelidiki prediksi laju penetrasi (*Rate of Penetration*; ROP) di Lapangan Gorgon, Australia, dengan menggunakan pendekatan *machine learning* berbasis *Multilayer Perceptron Regressor* (MLP). Tujuan penelitian ini adalah mengevaluasi pengaruh ukuran data pelatihan (*training data size*) dan pemilihan fitur (*feature selection*) terhadap kinerja model. Sebanyak tujuh dataset sumur (*well datasets*) digunakan dengan skenario eksperimental yang memvariasikan jumlah sumur pelatihan (dua dan tiga sumur) serta jumlah fitur input. Alur pengembangan model meliputi *exploratory data analysis*, *data preprocessing*, *model training*, *hyperparameter tuning*, dan evaluasi kinerja. Untuk merepresentasikan kondisi lapangan secara lebih realistis, diterapkan strategi *well-based data splitting*, di mana satu sumur diperlakukan sebagai *blind data* untuk menilai kemampuan generalisasi model. Evaluasi kinerja dilakukan menggunakan *Mean Absolute Percentage Error* (MAPE) dan koefisien determinasi (R^2). Hasil menunjukkan bahwa peningkatan jumlah sumur pelatihan meningkatkan akurasi prediksi, sedangkan penambahan fitur tidak selalu meningkatkan kinerja dan dapat menurunkan generalisasi. Kinerja terbaik dicapai pada skenario tiga sumur dengan tiga fitur input (MAPE 25% dan R^2 0,71), menegaskan pentingnya representativitas data dan pemilihan fitur dalam pengembangan model ROP yang *robust*.

PENDAHULUAN

Banyak metode telah digunakan untuk memprediksi dan memodelkan laju penetrasi (*rate of penetration*; ROP) baik secara analitikal, statistik, *machine learning* (ML) model, maupun metode *hybrid* (penggabungan metode analitik dengan ML). Pemodelan ROP secara analitikal maupun secara regresi umumnya dipengaruhi oleh beberapa parameter dan tidak lepas dari asumsi-asumsi. Terbukti selama ini model analitik berhasil memprediksi ROP dengan baik.

Tetapi hubungan input output yang digunakan dalam memodelkan ROP secara analitik memiliki kelemahan ketika parameter yang digunakan tidak tersedia atau data yang tersedia hanya dari beberapa titik saja dari total seluruh kedalaman. Efeknya, ketika akan dilakukan pemodelan ROP, diperlukan asumsi untuk mendapatkan nilai parameter. Kekurangannya adalah nilainya yang homogen dan terkadang tidak sesuai dengan kondisi lapangan. Solusi lainnya adalah dengan cara

melakukan pemodelan terhadap parameter yang dibutuhkan tetapi dikhawatirkan lebih banyak asumsi yang akan diterapkan dan model akan menjadi bias. Untuk itu, pemodelan ROP menggunakan metode ML bisa menjadi solusi dalam memodelkan ROP. Hal ini dikarenakan ML menggunakan algoritma yang mampu mempelajari pola-pola dan hubungan kompleks dalam dataset yang banyak. Sehingga parameter yang dipakai dalam memodelkan ROP tidak terpaku dengan parameter pemodelan analitik dan hasil prediksi model ROP menggunakan ML dikatakan memiliki akurasi yang baik.

Namun demikian, sejumlah penelitian sebelumnya dalam prediksi ROP berbasis machine learning masih terbatas pada penggunaan satu dataset sumur, yang kemudian dibagi menjadi training data dan test data untuk membangun serta mengevaluasi model. Pendekatan ini berpotensi menghasilkan performa model yang kurang merepresentasikan kemampuan generalisasi antar sumur, karena data yang digunakan berasal dari distribusi yang sama dan memiliki korelasi yang tinggi secara kedalaman.

Pada penelitian ini, penulis melakukan prediksi pada Lapangan Gorgon yang terletak pada Carnarvon Basin Australia menggunakan 7 dataset sumur. Dalam pemodelan ROP, penulis menggunakan algoritma Multilayer Perceptron Regressor (MLP *regressor*) yang merupakan bagian dari *Artificial Neural Network* (ANN) dengan variasi skema jenis fitur dan jumlah sumur yang digunakan pada tahapan training, serta evaluasi model menggunakan *Mean Absolute Percentage Error* (MAPE) dan R^2 .

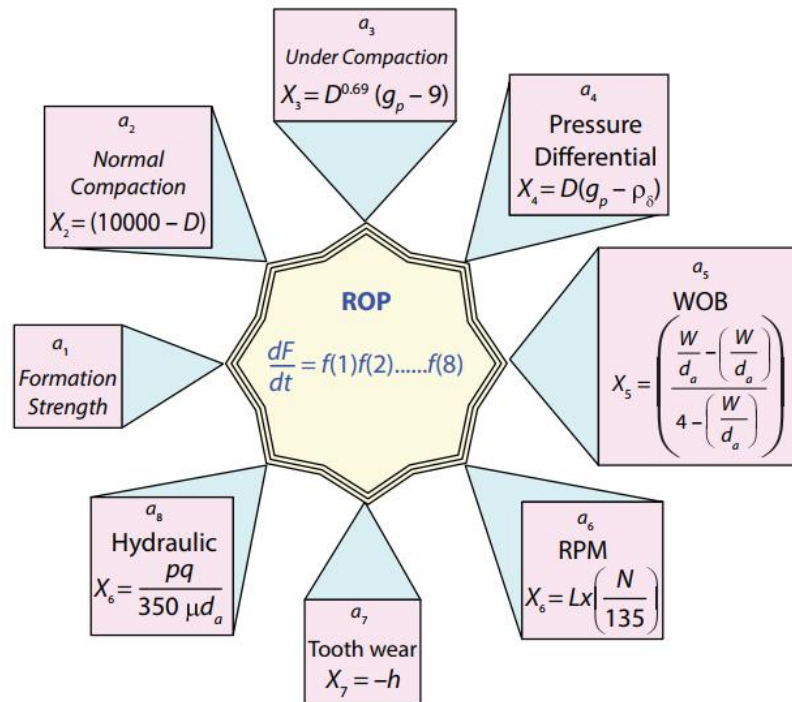
STUDI LITERATUR

LAJU PENETRASI (RATE OF PENETRATION; ROP)

Rate of penetration atau ROP merupakan kecepatan *drill bit* dalam mengebor formasi yang dicatat dalam satuan *meter/hour* atau *feet/hour*. Faktor yang memengaruhi ROP secara umum terbagi menjadi dua faktor, yaitu faktor yang terkontrol dan faktor lingkungan. Faktor lingkungan merupakan faktor yang tidak dapat dikontrol seperti *rock formation characteristics* dan *drilling fluids requirements*. Sedangkan faktor yang dapat dikontrol adalah faktor faktor yang dapat berubah secara instan nilainya seperti WOB, RPM, dan hidraulik (Hossain & Al - Majed, 2015).

Pendekatan analitik awal dalam pemodelan *Rate of Penetration* (ROP) umumnya berfokus pada hubungan antara parameter operasi pengeboran seperti *Weight on Bit* (WOB), *Rotary Speed* (RPM), karakteristik formasi, serta efek hidraulik. Model-model tersebut pada umumnya dibangun berdasarkan asumsi sederhana dan interaksi parameter yang terbatas, sehingga belum mampu merepresentasikan kompleksitas kondisi lapangan secara menyeluruh. Bourgoyne et al. (1986) membuat model analitik untuk mengestimasi ROP menggunakan *multiple regression* dan mempelajari efek kedalaman formasi, *compaction*, *differential pressure*, *bit diameter*, berat bit, RPM, dan *hydraulics effect* (Keshavarz Moraveji & Naderi, 2016).

Metode yang diusulkan oleh Bourgoyne et al. (1986 dalam Hossain & Al - Majed, 2015) menggunakan teknik *multiple regression* dengan *least square* untuk mendapatkan parameter konstan yang merepresentasikan data lapangan berdasarkan data historis. Menurut Bourgoyne et al. (1986 dalam Hossain & Al - Majed, 2015), terdapat delapan parameter konstan yang memengaruhi model ROP dan diilustrasikan seperti pada **Gambar 1**.



Gambar 1. Parameter yang memengaruhi model prediksi Rate of Penetration menurut Bourgoyne et al. (1986 dalam Hossain & Al - Majed, 2015).

PREDIKSI ROP MENGGUNAKAN MACHINE LEARNING

Machine learning (ML) merupakan pemrograman komputer yang digunakan untuk mengoptimalkan performa suatu kriteria dengan menggunakan data historis. Model yang dibuat berdasarkan parameter yang di input secara bebas oleh user, sehingga parameter yang dimasukkan oleh *user* dapat dioptimalkan menggunakan training data. Model ML juga dapat melakukan prediksi maupun mendeskripsikan data yang input oleh user (Alpaydin, 2020). Kemampuan dalam mendefinisikan non-linear pattern merupakan salah satu keuntungan ML dibandingkan model konvensional, sehingga ML mampu memprediksi ROP dengan akurasi yang tinggi (Kor et al., 2021). Model konvensional dan statistik harus mendefinisikan input dan output untuk model yang digunakan, sedangkan model secara ML, user mengatur input tanpa perlu mendefinisikan model ROP yang akan digunakan seperti pada model konvensional (Barbosa et al., 2019).

Pada pemodelan ROP lapangan Gorgon, model yang digunakan terbatas pada *Multilayer Perceptron regressor* (MLP regressor). *MLP Regressor* adalah jenis model regresi yang berbasis pada arsitektur jaringan syaraf tiruan. *MLP Regressor* digunakan untuk memprediksi variabel kontinu atau nilai numerik berdasarkan fitur atau variabel input.

Arsitektur MLP terdiri dari beberapa lapisan. Lapisan pertama disebut lapisan input, di mana fitur-fitur input diberikan ke model. Lapisan terakhir disebut lapisan output, yang menghasilkan nilai prediksi. Antara lapisan input dan output, terdapat satu atau beberapa lapisan tersembunyi (*hidden layers*) yang mengandung neuron-neuron yang memproses informasi dan mengekstraksi fitur-fitur yang lebih abstrak dari input. Setiap neuron dalam MLP menerima input, mengalikannya dengan bobot tertentu, dan mengaplikasikan fungsi aktivasi pada hasilnya. Fungsi aktivasi digunakan untuk menambahkan sifat non-linearitas ke dalam model, sehingga model dapat mempelajari pola-pola yang kompleks.

Beberapa fungsi aktivasi yang umum digunakan dalam MLP adalah fungsi sigmoid, fungsi ReLU (*Rectified Linear Unit*), atau fungsi tanh.

Model akan dievaluasi menggunakan MAPE dan R^2 . MAPE mengukur seberapa akurat suatu prediksi dengan memperhitungkan persentase galat absolut dari setiap titik data dalam hubungannya dengan nilai sebenarnya. Menurut Wei (2006) interpretasi nilai MAPE dapat dibagi menjadi 4 klaster berdasarkan nilai MAPE, yaitu sangat baik (<10%), Baik (10-25%), Wajar (25-50%) dan Tidak Akurat (>50%).

R^2 merupakan salah satu metrik yang digunakan untuk mengevaluasi sejauh mana model statistik atau regresi linear cocok dengan data yang diamati. R^2 memberikan indikasi tentang seberapa baik variasi dalam data sebenarnya dapat dijelaskan oleh model.

DATA DAN METODOLOGI

DATA

Data yang digunakan dalam penelitian merupakan dataset yang diunduh dari website nompims Australia. Penelitian ini menggunakan dataset lapangan gorgon yang berjumlah 7 sumur. Seperti pada **Tabel 1**.

Tabel 1. Data sumur Lapangan Gorgon.

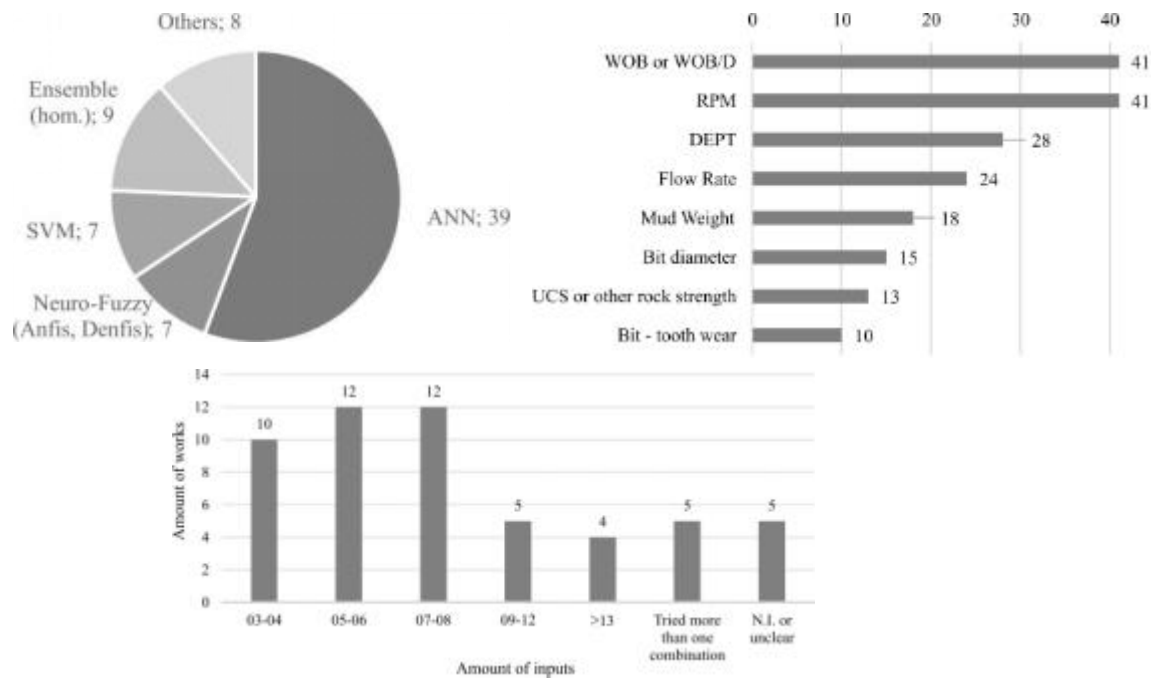
Kode Sumur	Mulai Pengeboran	Selesai Pengeboran	Jenis Sumur
GOR 1A	23 Mei 2019	28 September 2020	Sumur Produksi Gas
GOR 1B	26 Mei 2019	09 Oktober 2020	Sumur Produksi Gas
GOR 1G	27 Mei 2019	23 Oktober 2020	Sumur Produksi Gas
GOR 4C	13 Desember 2019	18 Juli 2020	Sumur Produksi Gas
GOR 4D	14 Desember 2019	03 Agustus 2020	Sumur Produksi Gas
GOR 4E	15 Desember 2019	16 Agustus 2020	Sumur Produksi Gas
GOR 4F	16 Desember 2019	30 Agustus 2020	Sumur Produksi Gas

Penelitian ini dilakukan untuk melihat hubungan jumlah sumur yang dijadikan *train test* terhadap model, hubungan jumlah fitur terhadap model. Pemilihan jumlah sumur yang dijadikan *training data* berdasarkan konsep bahwa semakin banyak data yang digunakan dalam proses pelatihan model, maka semakin akurat ketika melakukan prediksi pada *blind data*. Harapannya, model akan mempelajari banyak pola yang terdapat dalam data. Jumlah sumur yang digunakan pada tahap pelatihan divariasikan antara 2 hingga 3 sumur sesuai dengan skema yang dapat dilihat pada **Tabel 2**. Sedangkan, *blind data* pada penelitian ini hanya terdiri dari satu sumur dengan kode GOR 4F seperti yang dapat ditampilkan pada **Tabel 2**.

Tabel 2. Variasi kombinasi data sumur sebagai dataset.

Skema	Input Feature	GOR 4C	GOR 4D	GOR 4E	GOR 4F
Skema 1	3 Features	Train Data	Train Data		Blind Data
Skema 2	3 Features	Train Data	Train Data	Train Data	Blind Data
Skema 3	7 Features	Train Data	Train Data		Blind Data
Skema 4	7 Features	Train Data	Train Data	Train Data	Blind Data

Kemudian variasi terhadap fitur input dilakukan berdasarkan konsep dan teori yang digunakan dalam prediksi ROP menggunakan metode analitik khususnya metode yang diusulkan oleh Bourgoyne et al. (1986), serta beberapa referensi model ML seperti yang terlihat pada **Gambar 2**. Variasi fitur yang akan digunakan pada penelitian ini dapat dilihat pada **Tabel 3**.



Gambar 2. Tipe algoritma ML dalam prediksi ROP, tipe fitur yang digunakan, serta jumlah fitur yang digunakan dalam pemodelan ROP (Barbosa et al., 2019).

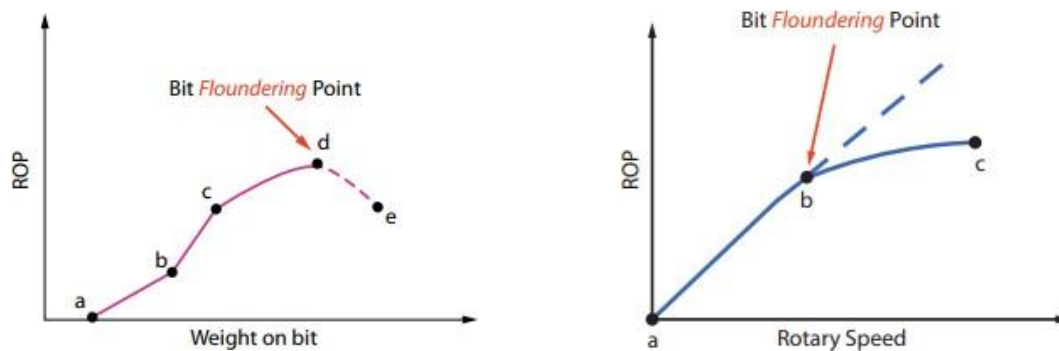
Tabel 3. Variasi fitur pada setiap skema.

Skema	TVD	RPM	WOB	GR	RSH	RHOB
Skema 1	✓	✓	✓			
Skema 2	✓	✓	✓			
Skema 3	✓	✓	✓	✓	✓	✓
Skema 4	✓	✓	✓	✓	✓	✓

Skema 1 dan Skema 2 hanya menggunakan 3 fitur, yaitu *Weight-on-Bit* (WOB), *True Vertical Depth* (TVD) dan *Rotational Speed* (RPM). Menurut metode yang diusulkan oleh Bourgoyne et al. (1986), WOB dan RPM memiliki hubungan terhadap ROP di mana ketika nilai WOB dan RPM awalnya naik maka nilai ROP naik, hingga pada suatu titik di mana ketika nilai WOB dan RPM terlalu besar maka nilai ROP akan berkorelasi negatif karena terjadi pengangkutan cutting yang tidak efisien atau *Bit Floundering* seperti yang tampak pada **Gambar 3** (Hossain & Al - Majed, 2015).

Skema 3 dan Skema 4 menggunakan enam fitur input, yang mencakup penambahan parameter reservoir yang berasal dari bacaan logging, seperti *Gamma Ray* (GR), resistivitas (RSH), dan *bulk density* (RHOB). Penambahan fitur ini didasarkan pada konsep dari model prediksi ROP secara analitik yang dikembangkan oleh Bourgoyne et al. (1986) yang memerlukan parameter *pore pressure* sebagai salah satu variabel perhitungannya. Sebagai catatan, dataset yang digunakan dalam penelitian ini tidak memiliki data *pore pressure* secara langsung. Selain itu, mengacu pada metode Eaton, *pore pressure* dapat diperkirakan menggunakan data resistivitas (Zhang, 2011). Zhang (2011)

menjelaskan bahwa resistivitas dapat digunakan untuk memodelkan *pore pressure* pada cekungan sedimen dengan asumsi bahwa resistivitas *shale* berada dalam kondisi normal.



Gambar 3. Fenomena *Bit Flounering* berdasarkan parameter pengeboran (Hossain & Al - Majed, 2015).

Oleh karena itu, alih-alih secara eksplisit memodelkan *pore pressure* dari resistivitas, penelitian ini menggunakan data resistivitas secara langsung sebagai fitur input dalam model *machine learning*. Selain itu, data densitas juga dimasukkan sebagai fitur karena dalam pendekatan Eaton, estimasi *pore pressure* memerlukan informasi *overburden pressure*. Nilai *overburden stress* dapat diasumsikan sebagai tegangan vertikal dan dapat diperkirakan menggunakan data densitas *bulk* (Subiatmono et al., 2017).

Parameter GR juga dimasukkan kedalam model karena GR mampu menjelaskan formasi *shale* dan non-*shale* yang mana ROP pada formasi *shale* dan non-*shale* akan berbeda. Pada formasi *shale* biasanya ROP memiliki nilai yang cenderung lebih kecil dibandingkan non-*shale*. Hal ini disebabkan oleh formasi *shale* yang sensitif sehingga terkadang terjadi fenomena *swelling*. Untuk mengatasi fenomena tersebut, *drilling fluid* yang digunakan biasanya memiliki nilai viskositas yang tinggi dan ini akan menyebabkan ROP akan rendah (Bourgoyne et al., 1986). Selain itu formasi *shale* memiliki sifat plastis yang tinggi, artinya batuan ini mudah berubah bentuk atau deformasi saat ditempa atau ditekan. Ketika bit bergerak menembus formasi *shale*, sering terjadi peningkatan tekanan hidrostatik di sekitar lubang bor. Tekanan ini dapat menyebabkan deformasi shale, dan mengurangi ROP (Talabani et al., 1993).

METODOLOGI

Penelitian ini menggunakan pendekatan berbasis *machine learning* untuk mengembangkan model prediksi *Rate of Penetration* (ROP) pada Lapangan Gorgon. Alur metodologi terdiri dari beberapa tahapan utama, yaitu akuisisi data, *exploratory data analysis* (EDA), *data preprocessing*, pengembangan model, optimasi *hyperparameter*, serta evaluasi model.

a. Akuisisi Data

Data yang digunakan dalam penelitian ini diperoleh dari *database* NOPIMS, yang terdiri dari data pengeboran tujuh sumur di Lapangan Gorgon, Australia. Dataset mencakup parameter operasional dan *logging* yang relevan terhadap prediksi ROP. Untuk mengevaluasi kemampuan generalisasi model, digunakan pendekatan *well-based data splitting*, di mana data dibagi berdasarkan sumur. Dua skenario eksperimen diterapkan, yaitu penggunaan dua dan tiga sumur

sebagai *training data*, sementara satu sumur (GOR 4F) digunakan sebagai *blind data* untuk prediksi.

b. Feature Selection

Pemilihan fitur dilakukan berdasarkan model analitik ROP serta studi *machine learning* sebelumnya. Fitur input dibagi menjadi dua kelompok, yaitu fitur dasar, seperti *Weight on Bit* (WOB), *True Vertical Depth* (TVD), dan *Rotary Speed* (RPM); dan fitur tambahan: *Gamma Ray* (GR), *Resistivity* (RSH), dan *Bulk Density* (RHOB). Variasi kombinasi fitur ini digunakan untuk mengevaluasi pengaruh kompleksitas fitur terhadap performa model.

c. Exploratory Data Analysis (EDA)

EDA dilakukan untuk memahami karakteristik data serta mengidentifikasi potensi permasalahan seperti *outlier*, *missing values*, dan hubungan antar variabel. Analisis yang dilakukan meliputi analisis univariat dan multivariat, termasuk statistik deskriptif dan korelasi antar fitur.

d. Preprocessing

ahap preprocessing dilakukan untuk meningkatkan kualitas data dan performa model, yang meliputi: 1) penanganan *missing value* dan data tidak valid; 2) *Smoothing data* menggunakan metode filter Savitzky–Golay (*polynomial order* = 2, *window length* = 150) untuk mengurangi *noise* pada data pengeboran; 3) Normalisasi data menggunakan *Standard Scaler*; dan 4) Penggabungan dataset berdasarkan skenario pembagian *training data* dan *blind data*.

e. Pengembangan Model

Model prediksi dikembangkan menggunakan algoritma *Multilayer Perceptron* (MLP) *Regressor*. Parameter awal model yang digunakan adalah: *Hidden layer* (200, 150), Fungsi aktivasi ReLU, *Solver Adam*, dan *Learning rate* (alpha) sebesar 0.01. Model dilatih menggunakan dataset sesuai dengan skenario yang telah ditentukan.

f. Hyperparameter Tuning

Untuk meningkatkan performa model, dilakukan optimasi hyperparameter menggunakan metode *Randomized Search* yang dikombinasikan dengan *k-fold cross-validation* pada *training data*. Parameter yang dioptimasi meliputi: 1) Jumlah dan konfigurasi *hidden layer*; 2) Fungsi aktivasi (ReLU, tanh); 3) *Solver* (Adam, LBFGS); 4) *Learning rate* (0,0001; 0,001; 0,01). Selain itu, digunakan *validation curve* untuk menganalisis perilaku model terhadap potensi *overfitting* maupun *underfitting*.

g. Evaluasi Model

Evaluasi model dilakukan menggunakan dua metrik utama, yaitu *Mean Absolute Percentage Error* (MAPE) untuk mengukur tingkat kesalahan prediksi, dan *Coefficient of Determination* (R^2) untuk menilai kesesuaian model terhadap data aktual. Evaluasi dilakukan baik pada *training data* maupun pada *blind data* untuk menilai kemampuan generalisasi model.

h. Prediksi pada Blind Well

Model yang telah dilatih digunakan untuk melakukan prediksi pada *blind well*. Hasil prediksi dibandingkan dengan nilai ROP aktual dan divisualisasikan terhadap kedalaman (TVD). Selain itu, distribusi *error* terhadap kedalaman juga dianalisis untuk mengevaluasi konsistensi model.

i. Feature Importance

Analisis kontribusi fitur dilakukan menggunakan metode *Permutation Feature Importance* (PFI). Metode ini mengukur perubahan performa model ketika suatu fitur dihilangkan, sehingga dapat mengidentifikasi fitur yang paling berpengaruh dalam prediksi ROP.

HASIL DAN PEMBAHASAN

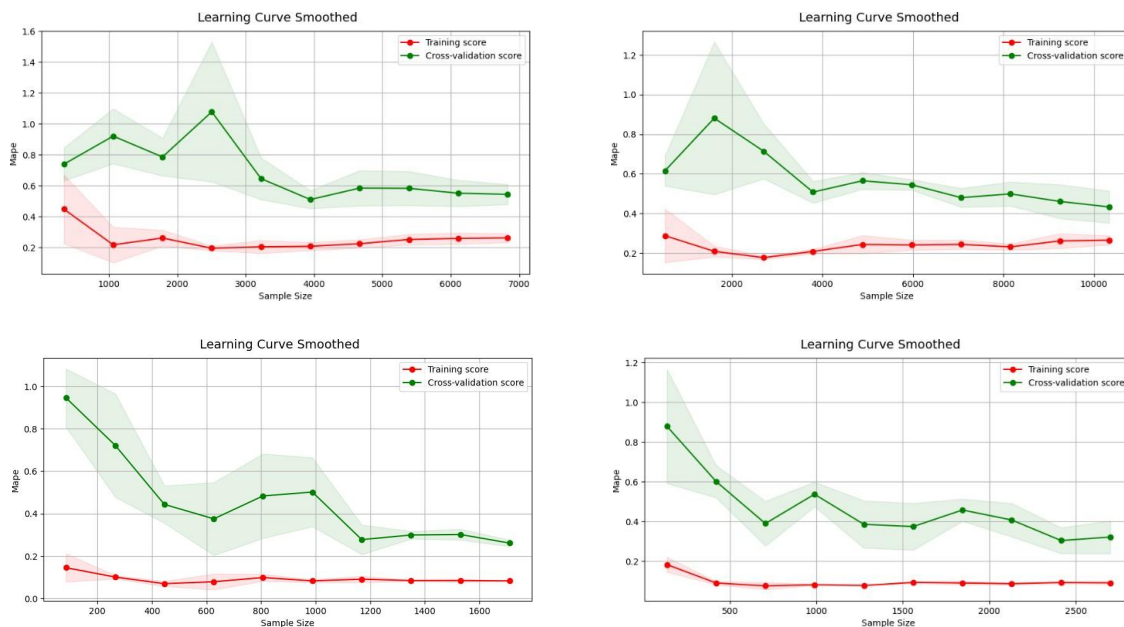
Setelah melakukan modelling dengan parameter seperti pada **Tabel 4**, didapatkan hasil yang berbeda untuk tiap skema seperti pada **Tabel 5**. Kemudian dilakukan plot learning curve untuk tiap skema seperti pada **Gambar 4**.

Tabel 4. Hasil hyperparameter tuning untuk setiap skema.

Hyperparameter	Skema 1	Skema 2	Skema 3	Skema 4
<i>Hidden layer size</i>	(100,)	(100,)	(25,75)	(25,75)
<i>Activation</i>	Relu	Relu	Tanh	Tanh
<i>Solver</i>	Adam	Adam	Adam	Adam
<i>Alpha</i>	0.01	0.01	0.01	0.01

Tabel 5. Hasil prediksi pada *Blind Well* GOR 4F.

Skema	MAPE	R2
Skema 1	30%	63%
Skema 2	25%	71%
Skema 3	33%	-60%
Skema 4	29%	-60%

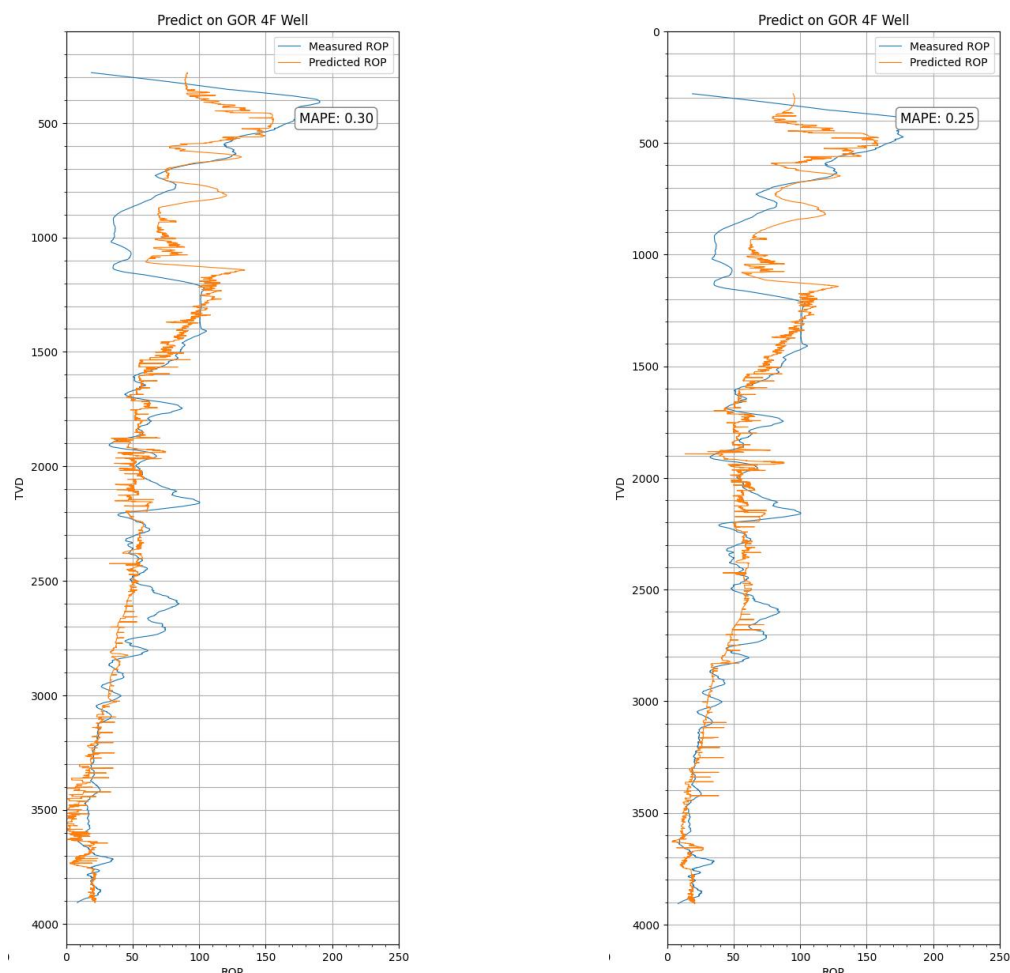


Gambar 4. Learning curve untuk Skema 1 (kiri atas), Skema 2 (kanan atas), Skema 3 (kiri bawah), dan Skema 4 (kanan bawah).

Berdasarkan hasil *learning curve* pada masing-masing skema (**Gambar 4**), terlihat bahwa Skema 2 menunjukkan performa yang paling optimal dengan pola kurva yang seimbang antara *training data* dan data validasi. Nilai *error* pada *training data* mengalami penurunan dan kemudian stabil, diikuti oleh penurunan *error* pada data validasi tanpa fluktuasi yang signifikan, yang mengindikasikan

kemampuan generalisasi model yang baik. Sebaliknya, Skema 1, 3, dan 4 menunjukkan indikasi *overfitting*, yang ditandai dengan perbedaan yang cukup besar antara kurva *training data* dan validasi serta fluktuasi *error* pada data validasi. Kondisi ini menunjukkan bahwa model cenderung mempelajari pola spesifik pada training data namun kurang mampu menggeneralisasi pada data baru. Fenomena ini semakin jelas pada Skema 3 dan 4, di mana penambahan jumlah fitur tidak diimbangi dengan jumlah *training data* yang memadai, sehingga meningkatkan kompleksitas model dan menyebabkan tingginya variansi. Dengan demikian, hasil ini menegaskan bahwa keseimbangan antara jumlah training data dan kompleksitas fitur merupakan faktor penting dalam menghasilkan model prediksi yang *robust*.

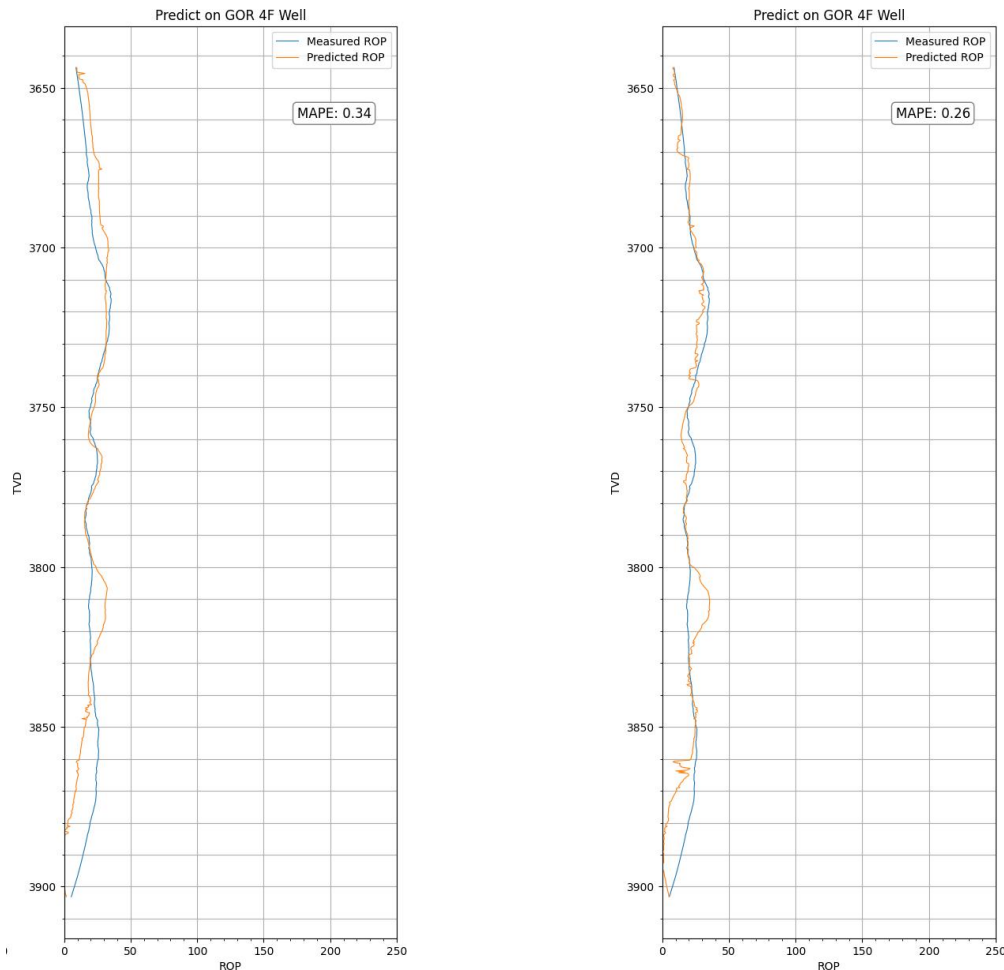
Berdasarkan hasil evaluasi model pada data blind well GOR 4F, terlihat bahwa Skema 2 memberikan performa terbaik dibandingkan skema lainnya, dengan nilai MAPE sebesar 25% dan R^2 sebesar 0,71 (**Gambar 5**). Peningkatan performa dari Skema 1 ke Skema 2 menunjukkan bahwa penambahan jumlah training data dari dua menjadi tiga sumur memberikan dampak signifikan terhadap kemampuan generalisasi model. Hal ini sejalan dengan prinsip dasar *machine learning* bahwa semakin banyak data representatif yang digunakan dalam proses pelatihan, maka model akan lebih mampu menangkap pola yang mendasari hubungan antara parameter input dan ROP.



Gambar 5. Hasil prediksi pada *Blind Well* GOR 4F untuk Skema 1 (kiri) dan Skema 2 (kanan).

Sebaliknya, pada Skema 3 dan Skema 4, performa model mengalami penurunan yang cukup signifikan, bahkan menghasilkan nilai R^2 negatif. Hal ini mengindikasikan bahwa model tidak mampu menangkap pola yang relevan pada data baru dan memiliki performa yang lebih buruk dibandingkan prediksi berbasis rata-rata (**Gambar 6**). Penurunan performa ini diduga disebabkan oleh

meningkatnya kompleksitas model akibat penambahan jumlah fitur tanpa diimbangi dengan jumlah training data yang memadai. Kondisi ini memperkuat indikasi terjadinya *overfitting*, di mana model terlalu menyesuaikan diri terhadap training data sehingga kehilangan kemampuan generalisasi.



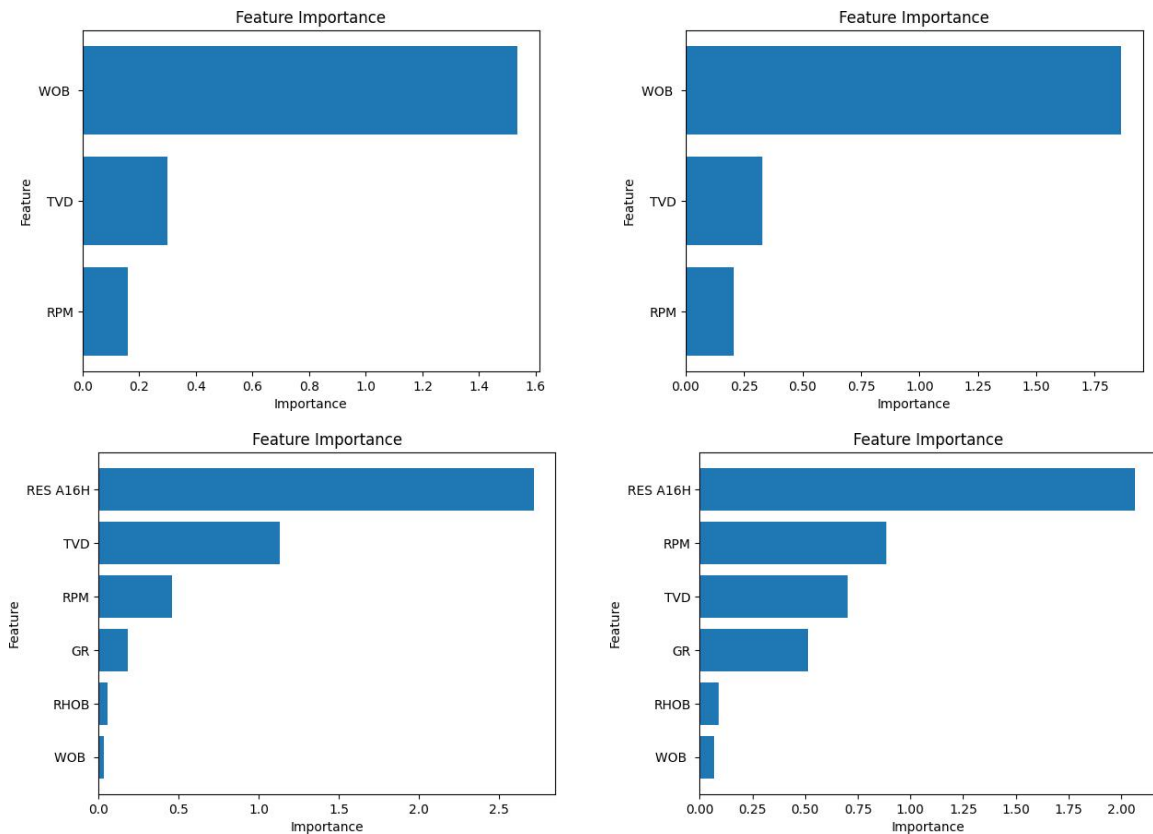
Gambar 6. Hasil prediksi pada *Blind Well* GOR 4F untuk Skema 3 (kiri) dan Skema 4 (kanan).

Analisis *Permutation Feature Importance* (PFI) menunjukkan bahwa *Weight on Bit* (WOB) merupakan fitur yang paling berpengaruh terhadap prediksi ROP dibandingkan fitur lainnya seperti RPM dan TVD. Hal ini konsisten dengan teori dan model analitik sebelumnya, di mana WOB memiliki hubungan langsung terhadap laju penetrasi hingga batas tertentu. Namun demikian, pada skema dengan fitur tambahan (Skema 3 dan 4), fitur resistivitas menunjukkan pengaruh yang dominan terhadap model (**Gambar 7**). Hal ini perlu dikaji lebih lanjut, karena dominasi fitur tersebut tidak diikuti dengan peningkatan performa model secara keseluruhan.

Dominasi resistivitas dalam PFI dapat mengindikasikan beberapa kemungkinan, seperti adanya *noise* pada data, korelasi yang tidak stabil terhadap ROP, atau permasalahan dalam *preprocessing* dan normalisasi data. Selain itu, penggunaan resistivitas sebagai representasi tidak langsung dari *pore pressure* juga dapat menyebabkan ketidakakuratan jika asumsi dasar tidak sepenuhnya terpenuhi pada dataset yang digunakan. Oleh karena itu, diperlukan evaluasi lebih lanjut terhadap kualitas data dan relevansi fitur tambahan sebelum digunakan dalam model.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa performa model tidak hanya dipengaruhi oleh jumlah fitur yang digunakan, tetapi juga oleh keseimbangan antara kompleksitas model dan jumlah *training data*. Model dengan fitur yang lebih sederhana namun didukung oleh training data

yang cukup justru menghasilkan performa yang lebih baik dibandingkan model dengan fitur yang lebih kompleks tetapi data yang terbatas.



Gambar 7. *Permutation Feature Importance* pada Skema 1 (kiri atas), Skema 2 (kanan atas), Skema 3 (kiri bawah), dan Skema 4 (kanan bawah).

Perlu dicatat bahwa rentang kedalaman pada **Gambar 6** tampak lebih pendek dibandingkan **Gambar 5**. Hal ini disebabkan oleh keterbatasan ketersediaan data pada fitur tambahan yang digunakan dalam analisis PFI. Penambahan jumlah fitur, khususnya data *logging* seperti resistivitas, densitas, dan neutron, memerlukan kelengkapan data yang tidak tersedia secara kontinu sepanjang seluruh kedalaman sumur. Data *logging* tersebut umumnya hanya tersedia pada interval tertentu, terutama pada zona produktif. Oleh karena itu, analisis PFI hanya dapat dilakukan pada rentang kedalaman di mana seluruh fitur tersedia secara lengkap, sehingga menghasilkan rentang kedalaman yang lebih terbatas dibandingkan visualisasi prediksi ROP secara keseluruhan.

KESIMPULAN

Penelitian ini mengkaji prediksi *Rate of Penetration* (ROP) pada Lapangan Gorgon menggunakan algoritma *Multilayer Perceptron* (MLP) dengan variasi jumlah training data dan kombinasi fitur. Berdasarkan hasil yang diperoleh, dapat disimpulkan bahwa jumlah *training data* memiliki pengaruh signifikan terhadap performa model. Peningkatan jumlah sumur yang digunakan sebagai training data terbukti mampu meningkatkan akurasi prediksi, yang ditunjukkan oleh penurunan nilai MAPE dan peningkatan nilai R^2 .

Selain itu, penambahan jumlah fitur tidak selalu memberikan peningkatan performa model. Pada penelitian ini, penggunaan fitur tambahan justru menyebabkan penurunan kemampuan generalisasi, terutama ketika tidak diimbangi dengan jumlah training data yang memadai. Hal ini mengindikasikan

bahwa peningkatan kompleksitas model tanpa dukungan data yang cukup dapat menyebabkan *overfitting*.

Model terbaik diperoleh pada skema dengan tiga sumur sebagai training data dan tiga fitur utama (WOB, RPM, dan TVD), dengan nilai MAPE sebesar 25% dan R^2 sebesar 0,71. Hasil ini menunjukkan bahwa pemilihan fitur yang relevan serta keseimbangan antara kompleksitas model dan jumlah training data merupakan faktor kunci dalam pengembangan model prediksi ROP yang *robust*.

Secara keseluruhan, penelitian ini menegaskan bahwa strategi pembagian data berbasis sumur (*well-based splitting*) serta pemilihan fitur yang tepat dapat meningkatkan kemampuan generalisasi model dalam memprediksi ROP pada sumur yang belum pernah diamati sebelumnya.

DAFTAR PUSTAKA

Alpaydin, Ethem. (2020). Introduction to machine learning. The MIT Press.

Barbosa, L. F. F. M., Nascimento, A., Mathias, M. H., & de Carvalho, J. A. (2019). Machine learning methods applied to drilling rate of penetration prediction and optimization - A review. *Journal of Petroleum Science and Engineering*, 183(March), 106332. <https://doi.org/10.1016/j.petrol.2019.106332>

Bourgoyne, A. T., Millheim, K. K., Chenevert, M. E., & Young, F. S. (1986). *Applied Drilling Engineering*. Society of Petroleum Engineers Richardson, Texas, USA. <https://doi.org/10.2118/9781555630010>

Hossain, M. E., & Al - Majed, A. A. (2015). *Fundamentals of Sustainable Drilling Engineering*. Wiley. <https://doi.org/10.1002/9781119100300>

Keshavarz Moraveji, M., & Naderi, M. (2016). Drilling rate of penetration prediction and optimization using response surface methodology and bat algorithm. *Journal of Natural Gas Science and Engineering*, 31, 829–841. <https://doi.org/10.1016/j.jngse.2016.03.057>

Kor, K., Ertekin, S., Yamanlar, S., & Altun, G. (2021). Penetration rate prediction in heterogeneous formations: A geomechanical approach through machine learning. *Journal of Petroleum Science and Engineering*, 207, 109138. <https://doi.org/10.1016/j.petrol.2021.109138>

Subiatmono, P., Pratiknyo, A. K., & Dingkaputra, D. (2017). Prediksi Pore Pressure Menggunakan Metode D-Exponent Dan Eaton Sonic Log. *Jurnal Mineral, Energi Dan Lingkungan*, 1(1), 28. <https://doi.org/10.31315/jmel.v1i1.1771>

Talabani, S., Chukwu, G., & Hatzignatiou, D. (1993, April 26). Drilling Successfully through Deforming Shale Formations: Case Histories. *Low Permeability Reservoirs Symposium*. <https://doi.org/10.2118/25867-MS>

Wei, W. W. S. . (2006). *Time series analysis: univariate and multivariate methods*. Pearson Addison Wesley.

Zhang, J. (2011). Pore pressure prediction from well logs: Methods, modifications, and new approaches. *Earth-Science Reviews*, 108(1–2), 50–63.
<https://doi.org/10.1016/j.earscirev.2011.06.001>