

Prediksi *Rate of Penetration* pada Pengeboran Minyak Bumi dengan Elman *Recurrent Neural Network*

ELSA AIZIYAH¹, ALLEN HARYANTO LUKMANA², AGUS MAMAN ABADI¹

¹Matematika, Universitas Negeri Yogyakarta

²Teknik Perminyakan, Universitas Pembangunan Nasional Yogyakarta, Indonesia

Email: elsaaiziyah.2021@student.uny.ac.id

Received 24 Juni 2025 | *Revised* 25 September 2025 | *Accepted* 29 Oktober 2025

ABSTRAK

Penelitian ini bertujuan memprediksi laju penetrasi (ROP) guna mempercepat waktu pengeboran dan menekan biaya operasional. Metode yang digunakan adalah Elman Recurrent Neural Network (ERNN) dengan algoritma backpropagation, yang dipilih karena kemampuannya dalam mengenali pola data sekuensial pada data pengeboran. Data yang digunakan 2613 data ASCII Mudlogging minyak bumi dari PT Geotama Jogja dengan 5 variabel input, yaitu Kedalaman Vertikal Sejati atau TVD (m), Beban Mata Bor atau WOB (klbs), Kepadatan Sirkulasi Ekuivalen atau ECD (SG), Mud Weight in atau MWI (SG), dan Total Kecepatan Rotasi Pahat atau TRPM. Sedangkan variabel outputnya yaitu laju penetrasi atau ROP (m/hr). Data dihaluskan menggunakan Savitzky-Golay filter dan dibagi data training dan data testing yang sebesar 90% dan 10%. Model ERNN terbaik yang diperoleh yaitu 5 variabel input, 17 neuron tersembunyi, dan 1 variabel output. Nilai MAPE data training sebesar 16.18%, dengan akurasi 83.82%. Sedangkan nilai MAPE data testing sebesar 15.48%, sehingga akurasinya 84.52%.

Kata kunci: *Elman Recurrent Neural Network, laju penetrasi, prediksi, minyak bumi, MAPE*

ABSTRACT

This study aims to predict the rate of penetration (ROP) to speed up drilling time and reduce operational costs. The method used is the Elman Recurrent Neural Network (ERNN) with the backpropagation algorithm, which was chosen because of its ability to recognize sequential data patterns in drilling data. The data used are 2613 ASCII Mudlogging data from PT Geotama Jogja with 5 input variables, namely True Vertical Depth or TVD (m), Drill Bit Load or WOB (klbs), Equivalent Circulation Density or ECD (SG), Mud Weight in or MWI (SG), and Total Tool Rotation Speed or TRPM. While the output variable is the rate of penetration or ROP (m/hr). The data is smoothed using the Savitzky-Golay filter and divided into training data and testing data of 90% and 10%. The best ERNN model obtained is 5 input variables, 17 hidden neurons, and 1 output variable. The MAPE value of the training data is 16.18%, so the accuracy is 83.82%. Meanwhile, the MAPE value for the testing data was 15.48%, resulting in an accuracy of 84.52%.

Keywords: *Elman Recurrent Neural Network, penetration rate, prediction, petroleum, MAPE*

1. PENDAHULUAN

Minyak bumi berbentuk cairan kental berwarna hitam sampai coklat kehitaman, mudah terbakar, dan terletak di dalam bumi yang biasanya berada 3 – 4 km di bawah permukaan laut atau di lapisan atas pada kerak bumi. Pengeboran Minyak bumi merupakan proses pengambilan minyak dari lapisan kerak bumi melalui pipa dan sumur dengan peralatan khusus. Biaya operasional pengeboran cukup tinggi sehingga berbagai macam cara dilakukan untuk mengurangnya. Salah satu cara untuk menurunkan biaya yaitu meningkatkan laju penetrasi atau *Rate of Penetration* (ROP) **(Riazi, Mehrjoo, Nakhaei, & Jalalifar, 2022)**. Setiap operasi pengeboran akan diusahakan ROP yang optimal. ROP merupakan kecepatan mata bor dalam memecah formasi di bawahnya untuk memperdalam sebuah lubang sumur **(Alkinani, dkk, 2021)**. Selain itu, ROP sangat mempengaruhi waktu operasi pengeboran karena berhubungan langsung dengan *rotating time* yang mencapai 30% dari total waktu pengeboran.

Faktor yang mempengaruhi ROP dikelompokkan menjadi 2 yaitu dikontrol dan tidak dapat dikontrol. Parameter pengeboran yang dapat dikontrol mudah untuk diubah, seperti berat pada mata bor (WOB), putaran permenit (RPM), berat lumpur (MW), Kepadatan Sirkulasi Ekuivalen (ECD), viskositas platik (PV), kekuatan gel, titik leleh (YP), laju alir (Q), torsi (T), dan tekanan pipa tegak (SPP) **(Alkinani, dkk, 2021)**. Parameter pengeboran yang tidak dapat dikontrol seperti ukuran bit, jenis lumpur pengeboran, kepadatan, dan sifat reologi. Parameter pengeboran yang tidak dapat dikontrol merupakan faktor yang tidak dapat diubah **(Elkatatn, dkk, 2021)**.

Prediksi merupakan proses meramalkan berdasarkan data historis dengan tujuan memproyeksikan kondisi pada masa mendatang. Prediksi ROP sulit dilakukan karena banyaknya faktor yang saling mempengaruhi dan belum ada persamaan pasti yang menghubungkannya **(Alkinani, dkk, 2021)**. Metode prediksi ROP dapat dikelompokkan ke dalam 2 macam, yaitu model tradisional dan model berbasis data. Model tradisional menggunakan hubungan berbasis fisika antara ROP dengan parameter pengeboran. Model ini banyak kekurangan, yaitu hubungan antar parameter yang kompleks menyebabkan model kurang komprehensif, penggunaan koefisien empiris, dan persyaratan data tambahan, seperti properti bit, properti lumpur, desain bit. Selain itu, model tradisional harus sesuai dengan satu fasies karena koefisien empiris sangat bergantung pada litologi.

Pendekatan baru untuk memprediksi ROP dengan berbasis data untuk mengatasi kelemahan model tradisional **(Brenjkar & Delijani, 2022)**. Metode ini memiliki kelebihan dibandingkan model tradisional, yaitu tidak mengandung konstanta empiris atau spesifikasi mata bor dan tidak terikat pada rakitan lubang bor. Prediksi dengan pembelajaran mesin hanya bergantung pada data masukan dan pemilihan parameter masukan. Kemajuan pembelajaran ini telah membuka banyak peluang aplikasi dalam bidang pengeboran minyak bumi sehingga menjadi solusi dari masalah tersebut. Metode pemodelan berbasis data yang dapat digunakan yaitu pemodelan jaringan syaraf tiruan (*Neural Network*) **(Asante, dkk, 2021)**.

Neural Network merupakan sistem pemrosesan informasi yang terinspirasi dari karakteristik kinerja jaringan syaraf biologis sehingga seperti neuron pada otak manusia **(Elhassanien, dkk, 2024)**. *Neural network* mempunyai kemampuan menganalisis data rumit yang sulit diselesaikan dengan metode statistik konvensional **(Suahati, dkk, 2022)**. Salah satu jenis *Neural Network* yaitu *Recurrent Neural Network* (RNN), yaitu model yang dikembangkan dari model *feedforward* neural network dan memiliki umpan balik. RNN menggabungkan *perceptron multilayer* statis atau bagian dan mengeksploitasi kemampuan pemetaan nonlinear dari *multilayer perceptron*.

Penelitian ini menggunakan model *Elman Recurrent Neural Network*, yaitu jenis *neural network* yang mempunyai lapisan tambahan. ERNN memiliki struktur jaringan dengan memori

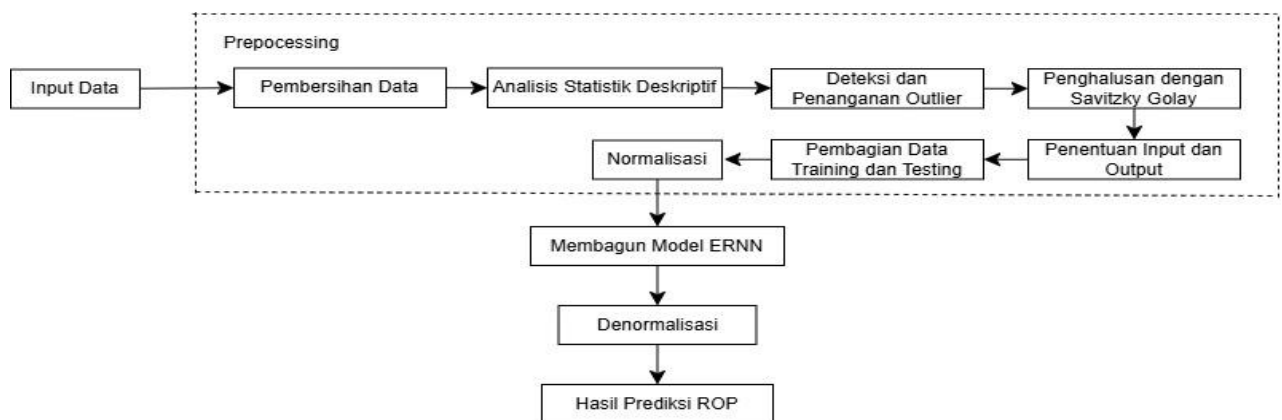
kontekstual yang mampu mengenali pola-pola temporal atau sekuensial dalam data, sehingga sesuai untuk memodelkan proses dinamis seperti pengeboran. Penelitian yang menggunakan model ERNN, yaitu **(Tompo, dkk, 2024)** memprediksi pengangguran di Provinsi Jawa Timur, **(Abdillah & Sukoco, 2023)** memprediksi harga saham perbankan di Indonesia, **(Jefri, dkk, 2020)** memprediksi indeks harga konsumen kota Ambon. Salah satu penelitian ROP yang sudah dilakukan oleh **(Abadi, dkk, 2024)** memprediksi laju penetrasi pengeboran menggunakan fuzzy Sugeno orde pertama.

Penelitian ini menghadirkan kebaruan dengan kombinasi parameter pengeboran, yaitu *Mud Weight* (densitas), Kepadatan Sirkulasi Ekuivalen (ECD), Kedalaman Vertikal Sejati (TVD), mekanik pengeboran (WOB), dan Total Kecepatan Rotasi Pahat (TRPM). *Mud weight* merupakan perbandingan berat perunit volume lumpur. Parameter ini penting dalam menentukan tekanan hidrostatik **(Islam & Hossain, 2021)**. WOB merupakan sejumlah gaya yang diberikan oleh mata bor selama operasi pengeboran berlangsung yang diukur dalam ribuan pon sesuai pada formasi batuan agar dapat dihancurkan. ECD merupakan kepadatan dinamis pada lumpur yang bersirkulasi dalam lubang bor **(Sun, dkk, 2024)**. RPM merupakan pengukuran kecepatan putar mata bor dalam pengeboran sumur. TVD merupakan jarak tegak lurus dari permukaan ke dasar.

Berdasarkan uraian di atas, penulis akan melakukan penelitian prediksi laju penetrasi (ROP) yang bertujuan mengembangkan model Elman *Reccurent Neural Network* untuk memprediksi ROP pada data pengeboran minyak bumi dan mengetahui akurasi dari model yang dibangun.

2. METODE PENELITIAN

Metode penelitian yang digunakan pada penelitian ini dapat dilihat pada Gambar 1, yang terbagi menjadi beberapa tahapan. Langkah-langkah penelitian ini yaitu sebagai berikut.



Gambar 1. Diagram Alir Penelitian

2.1 Ruang Lingkup Dataset

Data yang digunakan berupa sekunder, yaitu data ASCII Mudlogging yang diperoleh dari PT Geotama Energi yang beralamat di Jl. Laksda Adisucipto KM. 5, Ambarukmo, Caturtunggal, Depok, Sleman, Daerah Istimewa Yogyakarta. Data yang diambil berupa Beban Mata Bor (WOB), *Mud Weight In* (MWI), Kepadatan Sirkulasi Ekuivalen (ECD), Kedalaman Vertikal Sejati (TVD), mekanik pengeboran (WOB), dan Total Kecepatan Rotasi Pahat (TRPM), dan laju penetrasi (ROP). Data yang digunakan pada penelitian ini sebanyak 2.613 data dari kedalaman 49 m - 2661 m.

2.2 Preprocessing Data

Preprocessing data merupakan tahap awal dalam pengolahan data agar menjadi data yang bersih, terstruktur, dan dapat digunakan dalam proses pemodelan.

2.2.1 Pembersihan Data

Pembersihan data bertujuan untuk memperbaiki kesalahan data mentah agar model dapat berjalan lebih baik. Tahap ini meliputi *missing value*, yang dapat diakibatkan karena kesalahan pencatatan, sensor *error*, atau kesalahan sistem.

2.2.2 Analisis statistik deskriptif

Analisis statistik deskriptif digunakan untuk mengetahui gambaran umum tentang karakteristik data yang digunakan. Analisis ini meliputi ukuran pemusatan (rata-rata, median, modus), ukuran penyebaran (range, variasi, standar deviasi), dan mengetahui pola distribusi data.

2.2.3 Deteksi dan Penanganan Outliers

Outliers bisa memberikan informasi yang keliru, menjadikan model buruk, dan hasil analisis tidak akurat sehingga perlu evaluasi kualitas data untuk melihat latar belakang data (**Chen & Zhou, 2017**). Deteksi outlier merupakan langkah untuk mengetahui outlier (**Karkouch, dkk, 2016**). Boxplot merupakan salah satu teknik memvisualisasikan untuk melihat adanya outlier dalam data.

Penelitian ini menggunakan winsorizing, yaitu teknik yang membatasi nilai ekstrem dalam dataset agar berada dalam rentang tertentu tanpa menghapus data. Metode ini mengubah nilai ekstrim agar masuk batas normal berdasarkan *Interquartile Range* (IQR), yaitu selisih antara Q3 dan Q1. Metode ini membantu mempertahankan keseluruhan struktur kumpulan data sekaligus mengurangi efek outlier. Cara untuk menentukan batas bawah dan batas atas dapat menggunakan Persamaan (1) dan (2) sebagai berikut.

$$\text{Batas bawah} = Q1 - 1.5 \times IQR \quad (1)$$

$$\text{Batas atas} = Q3 + 1.5 \times IQR \quad (2)$$

2.2.4 Penghalusan dengan Savitzky Golay

Data mentah umumnya mengandung noise yang disebabkan oleh ketidakakuratan sistem perekaman dan kesalahan manusia sehingga menyebabkan memperpanjang proses pelatihan dan mengurangi generalisasi model (**Y. Wang, dkk, 1995**). Penanganan noise sangat penting untuk meningkatkan kinerja model dan mencapai hasil yang baik (**Fu, dkk, 2024**). Beberapa peneliti menggunakan filter Savitzky-Golay untuk mengurangi noise dari data pengeboran dan petrofisika guna meningkatkan kinerja model estimasi ROP (**Anemangely, dkk, 2018**).

Savitzky-Golay merupakan filter *low pass* yang mengekstrak data yang berguna dari data dengan *noise* dengan mempertahankan bentuk asli sinyal (**Bromba, 1981**). Filter ini bekerja dengan melakukan penghalusan melalui fungsi polinomial (**Savitzky&Golay, 1974**).

$$f_k = \sum_{k=0}^{k=n} b_{nk} i^k \quad (3)$$

Untuk jendela titik data tertentu yang berpusat di i , tentukan koefisien b_{nk} agar diperoleh polinomial yang sesuai dengan titik data dalam jendela. Kemudian menetapkan turunannya sama dengan nol (**Savitzky&Golay, 1974**).

$$\left(\frac{d^s f_i}{di^s} \right)_{i=0} = s! b_{ns} = a_{ns} \quad (4)$$

Satu set nilai berurutan $2M + 1$ akan digunakan dalam penentuan kecocokan kuadrat rata-rata terbaik melalui nilai polinomial berderajat n . Hal ini dapat dicapai dengan meminimalkan jumlah kuadrat selisih antara titik data aktual dan nilai polinomial (**Savitzky& Golay, 1974**).

$$\frac{\partial}{\partial b_{nk}} \left[\sum_{i=-M}^{i=M} (f_k - y_i)^2 \right] = 0 \quad (5)$$

Dengan meminimalkan terhadap b_{nk} diperoleh sebagai berikut (**Savitzky&Golay, 1974**).

$$\sum_{i=-M}^{i=M} \left(\sum_{k=0}^{k=n} b_{nk} i^{k+r} \right) = \sum_{i=-M}^{i=M} y_i i^r = F_k \quad (6)$$

Di mana r adalah indeks yang mewakili nomor persamaan dari 0 hingga n (ada $n + 1$ persamaan). Nilai i berkisar dari $-M$ hingga M , dan $i = 0$ pada titik pusat himpunan nilai $2M + 1$. Vektor sampel masukan $h = [h_{-M}, \dots, h_{-1}, h_0, h_1, \dots, h_m]$, dan matriks A didefinisikan sebagai berikut (**Sadeghi & Behnia, 2018**).

$$A = \begin{bmatrix} (-M)^0 & (-M)^1 & \dots & (-M)^n \\ \vdots & \vdots & \vdots & \vdots \\ (-1)^0 & (-1)^1 & \dots & (-1)^n \\ 1 & 0 & \dots & 0 \\ M^0 & M^1 & \dots & M^n \end{bmatrix}$$

Selanjutnya, koefisien filter G dihitung sebagai berikut (**Sadeghi & Behnia, 2018**).

$$G = (A^T A)^{-1} A^T \quad (7)$$

Kemudian, vektor koefisien dapat diturunkan seperti di bawah ini.

$$B = G y \quad (8)$$

Hanya menghitung a_0 dan menghitung baris pertama pada matriks W . Matriks G tidak bergantung pada sampel input dan hanya bergantung pada urutan filter dan panjang jendela. Nilai output apur pada indek ke- i sebagai berikut.

$$F_0 = \sum_{-M}^M G_i h_i = a_0 \quad (9)$$

2.2.5 Menentukan Variabel Input dan Output

Variabel input yang digunakan yaitu *Mud Weight* (densitas), Kepadatan Sirkulasi Ekuivalen (ECD), Kedalaman Vertikal Sejati (TVD), mekanik pengeboran (WOB), dan Total Kecepatan Rotasi Pahat (TRPM), sedangkan variabel output yaitu laju penetrasi (ROP).

2.2.6 Metode Evaluasi

Metode evaluasi yang digunakan yaitu data split, yaitu data akan dibagi menjadi 2 bagian yaitu data training dan data testing. Dataset akan dibagi perkedalaman 100 m agar data training dan data testing mewakili setiap kedalaman dengan komposisi masing-masing yaitu 90% dan 10%.

2.2.7 Normalisasi Data

Normalisasi dilakukan agar data berada diantara suatu range tertentu. Pada penelitian ini, proses normalisasi menggunakan Robust Scaler, yaitu algoritma yang menskalakan fitur menggunakan median dan kuartil untuk mengatasi bias yang berasal dari outlier.

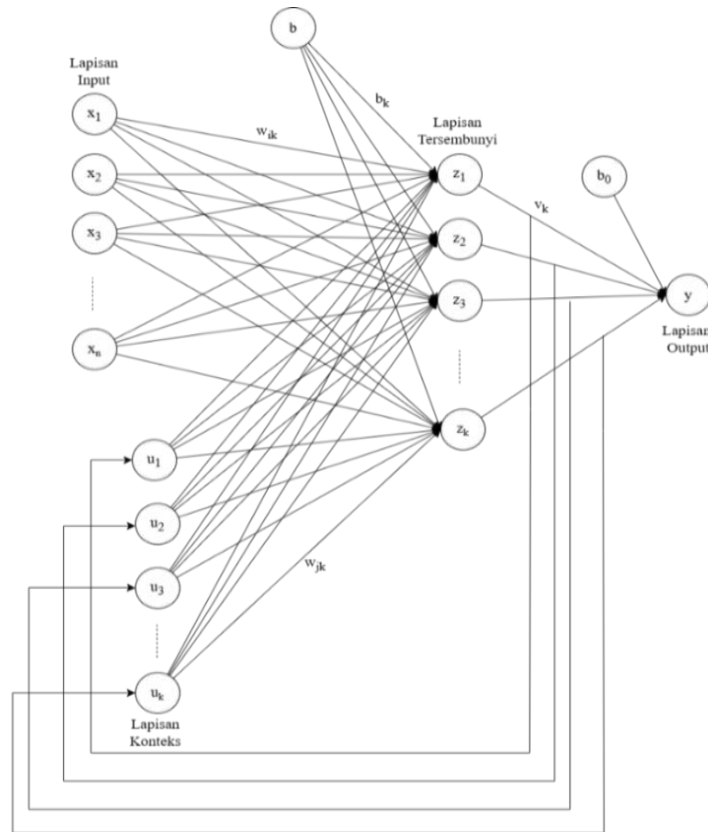
$$X_{\text{scaled}} = \frac{x_h - Q2(x)}{Q3(x) - Q1(x)} \quad (10)$$

Diketahui X_{scaled} nilai normalisasi data ke- h , x_h nilai data ke- h , $Q2(x)$ median, $Q1(x)$ kuartil pertama dari dataset, dan $Q3(x)$ kuartil ketiga dari dataset.

2.3 Membangun Model Elman *Recurrent Neural Network* Terbaik

Proses membangun model Elman *Recurrent Neural Network* (ERNN) dimulai dengan menentukan input jaringan dan jumlah neuron pada lapisan tersembunyi. Langkah penentuan banyak neuron lapisan tersembunyi dengan *trial* dan *error* kemudian membandingkan nilai MAPE terkecil dan jaringan paling sederhana.

Secara arsitektural, ERNN pada prinsipnya terdapat *feedback* pada lapisan tersembunyi. ERNN memiliki lapisan tambahan yang disebut *context layer* untuk menyimpan nilai aktivasi dari langkah sebelumnya (Radjabaycolle, 2020).



Gambar 1. Arsitektur Elman *Recurrent Neural Network*

Berdasarkan Gambar 2, jumlah neuron di *context layer* sama dengan jumlah neuron di *hidden layer* dan saling terhubung setiap neuronnya. *context layer* berfungsi sebagai memori dengan menyimpan keluaran lapisan tersembunyi, kemudian digunakan kembali sebagai masukan tambahan pada langkah berikutnya dalam jaringan (Lewis, 2016, hal 88-89).

Model ERNN dengan fungsi aktivasi sigmoid bipolar pada lapisan tersembunyi dan fungsi aktivasi linear pada lapisan output secara sistematis dituliskan dengan Persamaan (11) dan Persamaan (12) sebagai berikut.

$$y = \sum_{k=1}^p v_k h_k + b_0 \quad (11)$$

Persamaan untuk h_k yaitu sebagai berikut.

$$h_k = \frac{1 - \exp(-(\sum_{i=1}^n x_i w_{ik} + \sum_{j=1}^p u_j w_{jk} + b_k))}{1 + \exp(-(\sum_{i=1}^n x_i w_{ik} + \sum_{j=1}^p u_j w_{jk} + b_k))} \quad (12)$$

Sehingga model Elman *Recurrent Neural Network* dituliskan secara lengkap dapat dituliskan dengan persamaan berikut ini.

$$y = \sum_{k=1}^l v_k \frac{1 - \exp(-(\sum_{i=1}^n x_i w_{ik} + \sum_{j=1}^p u_j w_{jk} + b_k))}{1 + \exp(-(\sum_{i=1}^n x_i w_{ik} + \sum_{j=1}^p u_j w_{jk} + b_k))} + b_0 \quad (13)$$

Diketahui y variabel output, x_i variabel input dengan $i = 1, 2, \dots, n$, n jumlah variabel independent, p jumlah neuron pada lapisan tersembunyi, b_k bobot bias pada neuron ke- k pada lapisan tersembunyi, b_0 bobot bias pada neuron lapisan output, u_j variabel input yang berada pada neuron tambahan dengan $j = 1, 2, 3, \dots, p$, w_{ik} bobot dari neuron input ke- i menuju neuron ke- k pada lapisan tersembunyi, w_{jk} bobot dari neuron tambahan ke- j menuju neuron ke- k pada lapisan tersembunyi, v_k bobot neuron ke- k pada lapisan tersembunyi yang menuju ke lapisan output.

Algoritma *backpropagation* memaksimalkan hasil pelatihan dengan memodifikasi bobot yang digunakan. Apabila diperoleh bobot baru, maka akan dilakukan perhitungan kembali sampai diperoleh *error* yang telah memenuhi syarat berhenti. Langkah-langkah dalam algoritma *backpropagation* (Fauset, 1994).

Langkah 0: Lakukan inisiasi bobot dengan mengambil bobot awal menggunakan nilai random terkecil.

Langkah 1: Tetapkan parameter pembelajaran, yaitu maksimum epoch, target *error*, dan *learning rate*.

a. Maksimum epoch

Maksimum epoch merupakan jumlah epoch maksimum yang digunakan selama proses pembelajaran. Iterasi akan dihentikan jika nilai epoch sudah memenuhi batas maksimum.

b. Target *error*

Target *error* merupakan target nilai fungsi kerja. Iterasi akan dihentikan jika nilai fungsi kinerja kurang atau sama dengan target *error*.

c. *Learning rate*

Learning rate merupakan laju persepatan pembelajaran. Jika semakin kecil nilainya maka laju pembelajarannya akan semakin lambat dan semakin besar nilai pembelajarannya maka semakin besar pula langkah pembelajarannya.

d. Momentum

Momentum merupakan perubahan bobot yang didasarkan pada gradien pola terakhir dan sebelumnya.

Tahap *Feedforward*

Langkah 2: Setiap neuron *input* ($x_n, n = 1, 2, 3, \dots, n$) akan menerima sinyal x_n lalu meneruskan sinyal tersebut ke semua neuron tersembunyi.

Langkah 3: Sinyal yang diterima pada lapisan tersembunyi dari neuron input di rumuskan pada Persamaan (14) sebagai berikut.

$$z_{in_k} = \sum_{i=1}^n x_i w_{ik} + b_k \quad (14)$$

Fungsi aktivasi sigmoid bipolar digunakan untuk menghitung sinyal output yaitu sebagai berikut.

$$z_k = f(z_{in_k}) \quad (15)$$

Lalu sinyal dikirim neuron pada neuron tambahan di lapisan input.

Langkah 4: setiap neuron tersembunyi akan meneruskan sinyal dari neuron tambahan pada lapisan masukan ($u_j, j = 1, 2, 3, \dots, k$). Sinyal yang diterima berupa sinyal keluaran dari lapisan tersembunyi yang sudah di aktivasi dengan persamaan sebagai berikut.

$$u_j = z_k \quad (16)$$

Langkah 5: setiap neuron tambahan kemudian mengirimkan sinyal ke lapisan tersembunyi dan menjumlahkan sinyal input berbobot dengan persamaan sebagai berikut.

$$z_{net_k} = \sum_{i=1}^n x_i w_{ik} + \sum_{j=1}^l u_j w_{jk} + b_k \quad (17)$$

Pada sinyal keluaran dihitung dengan fungsi aktivasi berikut.

$$z_k = f(z_{net_k}) \quad (18)$$

Kemudian, sinyal dikirim ke seluruh neuron pada lapisan output.

Langkah 6: Neuron *output* akan menjumlahkan sinyal input berbobot dengan persamaan sebagai berikut.

$$y_{in} = b_0 + \sum_{k=1}^i (z_k) (v_k) \quad (19)$$

Pada sinyal *output* digunakan fungsi aktivasi sebagai berikut.

$$y = f(y_{in}) \quad (20)$$

Kemudian sinyal akan di kirim ke seluruh neuron pada lapisan *output*.

Tahap *Backpropagation*

Langkah 7: Hitung fungsi kesalahan untuk mendapatkan nilai bobot terbaru dengan menggunakan persamaan sebagai berikut.

$$\delta_k = (t - y)f'(y_{in}) \quad (21)$$

Dengan δ_k merupakan neuron *error* yang akan digunakan dalam perubahan bobot v_k . Kemudian akan dihitung koreksi bobot untuk memperbaiki v_i dan koreksi bias untuk memperbaiki b_0 . Berikut adalah persamaan yang digunakan.

$$\Delta v_k = \alpha \delta_k z_k \quad (22)$$

$$\Delta b_0 = \alpha \delta_k \quad (23)$$

diketahui bahwa α merupakan laju pembelajaran.

Langkah 8: Setiap neuron tersembunyi akan menjumlahkan delta *input* dari neuron yang berada di lapisan atasnya dengan persamaan sebagai berikut.

$$\delta_{in_k} = \sum_{k=1}^l \delta_k v_k \quad (24)$$

Untuk menghitung informasi *error* menggunakan persamaan sebagai berikut.

$$\delta_k = \delta_{in} f'(z_{in_k}) \quad (25)$$

Kemudian dilakukan koreksi bobot yang digunakan sebagai perbaikan nilai $W_{ik(a)}$ dan $W_{ik(b)}$ dengan rumus sebagai berikut.

$$W_{ik(a)} = \alpha \delta_k x_k \quad (26)$$

$$W_{ik(b)} = \alpha \delta_k u_j \quad (27)$$

Pada langkah ini juga akan menghitung koreksi bias yang nantinya akan digunakan sebagai perbaikan nilai b_k

$$\Delta b_k = \alpha \delta_k \quad (28)$$

Tahap Perubahan Bobot

Langkah 9: Neuron *output* memperbaiki bias dan bobotnya sebagai berikut.

$$v_k(\text{baru}) = v_k(\text{lama}) + \Delta v_k \quad (29)$$

$$b_0(\text{baru}) = b_0(\text{lama}) + \Delta b_0 \quad (30)$$

Langkah 10: setiap neuron lapisan tersembunyi akan memperbaiki bias dan bobot dengan persamaan sebagai berikut.

$$w_{ik}(\text{baru}) = w_{ik}(\text{lama}) + \Delta v_k \quad (31)$$

$$w_{jk}(\text{baru}) = w_{jk}(\text{lama}) + \Delta b_0 \quad (32)$$

$$b_k(\text{baru}) = b_k(\text{lama}) + \Delta b_k \quad (33)$$

Langkah 11: pengujian berhenti jika mencapai jumlah epoch yang diinginkan atau sampai nilai ambang batas sudah memenuhi.

2.4 Denormalisasi Data

Denormalisasi data merupakan proses mengembalikan data dari nilai ternormalisasi output jaringan harus dikembalikan ke nilai sebenarnya.

$$X = X_{scaled} \times IQR + median \quad (34)$$

Diketahui X_{scaled} data hasil normalisasi.

2.5 Hasil prediksi laju penetrasi (ROP)

Hasil Prediksi akan dihitung seberapa akurat untuk memprediksi dengan model yang akan terbentuk. Salah satu cara mengukur akurasi yaitu menggunakan *Mean Absolute Percentage Error* (MAPE), yaitu perhitungan rata-rata persentase kesalahan mutlak. Model dikatakan baik apabila nilai MAPE yang kecil.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|Y_t - \hat{Y}_t|}{\hat{Y}_t} \times 100\% \quad (35)$$

Diketahui Y_t nilai hasil aktual, n Jumlah data, $\hat{Y}_t$ nilai hasil pendugaan. Dengan kriteria MAPE sebagai berikut (**Farida, Sulistiani, & Ulinuha, 2023**).

Tabel 1. Kriteria MAPE

Nilai MAPE	Keterangan
< 10%	Sangat Baik
10% – 20%	Baik
20% – 50%	Cukup
> 50%	Tidak akurat

Model dikatakan baik apabila nilai akurasi tinggi. Nilai akurasi tersebut dapat diperoleh melalui perhitungan sebagai berikut.

$$akurasi = 100\% - MAPE \quad (36)$$

3. HASIL DAN PEMBAHASAN

Di bawah ini merupakan data ASCII Mudlogging yang dijadikan dataset sebanyak 2613 data yang sudah dibersihkan. Berikut ini merupakan data yang digunakan dapat dilihat di Tabel 2.

Tabel 2. Data ASCII Mudlogging

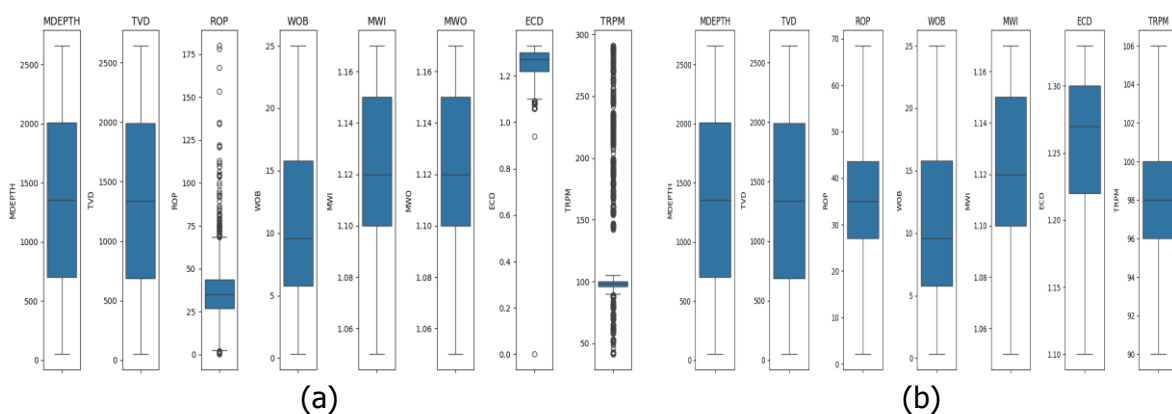
TVD (m)	WOB (klbs)	MWI (SG)	ECD (m/hr)	TRPM	ROP (m/hr)
49,00	2,70	1,05	1,06	162	32,29
50,00	1,20	1,05	1,06	162	59,94
...	...	...	...	...	...
2642,32	23,4	1,17	1,26	101	13,8
2643,32	17,3	1,17	1,26	100	13,8

Dibawah ini merupakan hasil dari analisis statistik deskriptif yang dapat dilihat pada Tabel 3.

Tabel 3. Statistik Deskriptif Data Numerik

Variabel	Atribut	Rata-rata	Standar Deviasi	Nilai Minimum	Nilai Maksimum
x_1	WOB	10.6133	5.8429	0.3	25
x_2	TVD	1340.8195	749.7754	49	2643.32
x_3	MWI	1.1173	0.0325	1.05	1.17
x_4	ECD	1.2372	0.0853	0	1.33
x_5	RPM	118.0137	51.8284	41	1.33
y	ROP	36.4358	15.8261	0.1	180

Hasil pengecekan memperlihatkan terdapat outliers di variabel ROP, ECD, dan TRPM masing-masing variabel tersebut yaitu 77, 408, dan 721. Hasil pengecekan outlier dapat dilihat pada (a).



Gambar 4. Boxplot Sebelum (a) dan Sesudah (b) Penanganan Outliers

Penanganan outliers pada penelitian ini menggunakan metode winsorizing berbasis *Interquartile Range* (IQR). Proses ini menggantikan nilai ekstrem dengan nilai yang masih berada dalam batas normal berdasarkan IQR pada variabel ROP, ECD, dan TRPM. Hasil dari penanganan tersebut dapat dilihat pada (b).

Selanjutnya akan dilakukan proses *smoothing* menggunakan metode Savitzky-Golay. Penelitian ini menggunakan *window size* berukuran 21 dengan order polinomial 3. Hasil dari proses *smoothing* dapat dilihat pada Tabel 4.

Tabel 4. Data hasil Smoothing dengan Savitzky Golay

TVD (mtr)	WOB (klbs)	MWI (SG)	ECD (m/hr)	TRPM	ROP (m/hr)
49	2,0113	1,05	1,1	106	47,8354
50	1,7700	1,05	1,1	106	50,8485
...	...	...	...	...	...
2642,32	20,7359	1,17	1,2607	100,6455	17,5611
2643,32	20,3413	1,17	1,2623	100,5608	12,6616

Selanjutnya yaitu penentuan variabel input dan output dengan 5 neuron dalam variabel input yang digunakan meliputi WOB (x_1), TVD (x_2), MWI (x_3), ECD (x_4), dan TRPM (x_5). Variabel output yaitu ROP (y), sehingga banyaknya neuron output yaitu 1 neuron.

Kemudian, data akan dibagi menjadi data *training* dan data *testing* dengan persentase masing-masing yaitu 90% dan 10%. Data yang berjumlah 2613 data akan menjadi 2350 data *training* dan 263 data *testing*.

Tabel 5. Pembagian Data *Testing* dan *Training* Setiap Interval

Kedalaman	Kelompok	Pembagian Data	Jumlah
0 – 99	1	Testing	6
		Training	45
100 – 199	2	Testing	10
		Training	90
...	...	...	...
2500 – 2599	26	Testing	10
		Training	90
2600 – 2699	27	Testing	7
		Training	55

Langkah selanjutnya yaitu melakukan normalisasi untuk menyetarakan rentang nilai yang akan di pakai. Pada penelitian ini menggunakan Robust scaler dengan median sama dengan 0 dan IQR sama dengan 1.

Tabel 6. Hasil Normalisasi Data Training

TVD	WOB	MWI	ECD	TRPM	ROP
-0,99281	-0,75562	-1,4	-2,17896	1,908973	0,983724
-0,99204	-0,78067	-1,4	-2,17896	1,908973	1,211272
...	...	...	...	...	...
1,146782	1,00409	1	0,656777	-0,0985	-1,67259
0,653266	0,991005	1	-0,33732	-0,25635	-0,78619

Tabel 7. Hasil Normalisasi Data Testing

TVD	WOB	MWI	ECD	TRPM	ROP
-0,98357	-0,83151	-1,4	-2,17896	1,908973	1,40508
-0,97433	-0,83795	-1,4	-2,17896	1,908973	1,433703
...	...	...	...	...	...
0,969453	0,762321	1	-0,12172	-0,95069	0,19399
0,994084	0,742476	1	-0,26632	-0,32144	-0,81324

Kemudian, pembentukan model terbaik untuk memprediksi ROP menggunakan Elman *Recurrent Neural Network* dengan algoritma backpropagation. Fungsi aktivasi yang digunakan pada lapisan tersembunyi adalah fungsi aktivasi sigmoid bipolar dan lapisan output menggunakan adalah fungsi aktivasi linear.

Tabel 8. Parameter Pembelajaran

Parameter	Nilai
Maksimum epoch	100
Target error	0.001
Learning rate	0.1
Momentum	0.9

Jumlah neuron pada lapisan tersembunyi ditentukan melalui metode *trial* dan *error* dengan membandingkan nilai MAPE dari beberapa model yang terbentuk. Proses pembelajaran model Elman *Recurrent Neural Network* dilakukan dengan melakukan percobaan 1 sampai 25 neuron pada lapisan tersembunyi.

Tabel 9. MAPE pada Neuron Tersembunyi

Jumlah Neuron Tersembunyi	<i>Training</i>	<i>Testing</i>
1	33.03%	28.40%
2	28.86%	25.19%
3	23.63%	21.43%
4	24.16%	22.28%
5	21.35%	20.41%
6	19.77%	19.99%
7	19.70%	20.35%
8	20.12%	19.72%
9	17.07%	17.59%
10	20.16%	19.35%
11	18.55%	18.88%
12	18.43%	18.82%
13	17.22%	17.68%
14	17.29%	18.48%
15	17.21%	16.82%
16	16.95%	17.51%
17*	16.18%	15.48%
18	16.26%	16.65%
19	19.02%	19.59%
20	16.65%	17.23%
21	16.97%	16.68%
22	19.85%	19.71%
23	15.94%	17.26%
24	15.72%	16.78%
25	16.91%	17.05%

Tabel 5 terlihat bahwa penurunan MAPE pada neuron ke-17 menunjukkan peningkatan performa model. Namun, penambahan neuron setelah dan sesudahnya cenderung tidak efektif dan dapat menyebabkan *overfitting*, ditandai dengan meningkatnya MAPE pada data testing. Model terbaik yang diperoleh dengan jumlah neuron tersembunyi 17 yang berdasarkan nilai MAPE, yaitu 16.18% untuk data *training* dan 15.48% untuk data *testing*. Nilai MAPE tersebut menunjukkan model dapat memprediksi dengan baik. Dengan persamaan model tersebut sebagai berikut.

$$y = \sum_{k=1}^{17} v_k \frac{1 - \exp(-(\sum_{i=1}^5 x_i w_{ik} + \sum_{j=1}^{17} u_j w_{jk} + b_k))}{1 + \exp(-(\sum_{i=1}^5 x_i w_{ik} + \sum_{j=1}^{17} u_j w_{jk} + b_k))} + b_0$$

Secara umum dapat dituliskan dengan

$$y = -0.0696 \frac{1 - e^{-((-1.207x_1) + (-0.488x_2) + \dots + (0.446x_5) + ((0.018u_1) + (-0.175u_2) + \dots + (0.242u_{17})))}}{1 + e^{-((-1.207x_1) + (-0.488x_2) + \dots + (0.446x_5) + ((0.018u_1) + (-0.175u_2) + \dots + (0.242u_{17})))}} + \dots$$

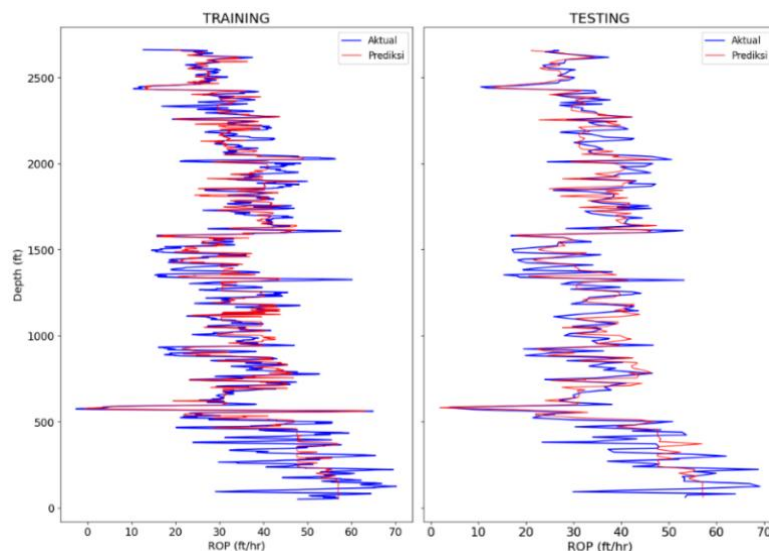
$$+ (-0.6499) \frac{1 - e^{-((3.807x_1) + (-8.100x_2) + \dots + (3.733x_5) + ((-0.075u_1) + (0.358u_2) + \dots + (0.045u_{17})))}}{1 + e^{-((3.807x_1) + (-8.100x_2) + \dots + (3.733x_5) + ((-0.075u_1) + (0.358u_2) + \dots + (0.045u_{17})))}} + 0.6311$$

Hasil pembentukan model terbaik diperoleh nilai output jaringan dalam bentuk data normalisasi sehingga dilakukan yaitu denormalisasi. Setelah itu, akan diperoleh nilai target, *output*, dan *error* dari masing-masing data training dan data testing.

Tabel 10. Target, Output, dan Error

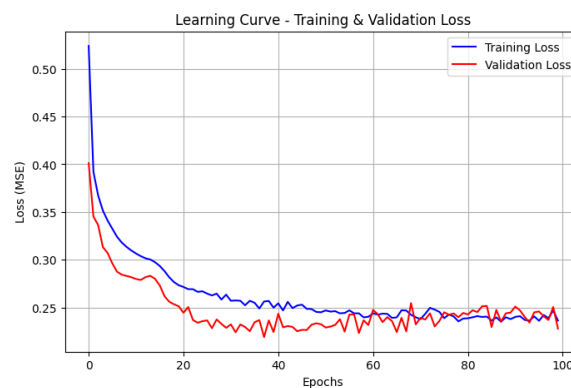
Target	Output	Error
Data Training		
47,83542443	56,77418	-8,93875
50,84851873	56,79129	-5,94278
⋮	⋮	⋮
12,6616789	18,61039543	-5,948716531
26,58861218	19,09920311	7,489409066
Data Testing		
53,41483818	56,40651	-2,99168
53,7938542	56,224	-2,43015
⋮	⋮	⋮
26,88640739	27,05601	-0,1696
25,84639035	27,06098	-1,21459

Proses prediksi ROP menggunakan model Elman *Recurrent Neural Network* sudah selesai. Sebaran perbandingan antara data aktual dan hasil prediksi ditunjukkan oleh Gambar 5.



Gambar 5. Perbandingan Data Aktual dengan Prediksi

Selain itu, performa pembelajaran jaringan dapat diamati melalui kurva loss pada Gambar 6. Kurva ini menunjukkan penurunan nilai galat (*loss function*) seiring dengan bertambahnya epoch pelatihan, yang menandakan bahwa model berhasil melakukan proses pembelajaran secara stabil hingga mencapai konvergensi. Dengan demikian, kurva loss mendukung validitas hasil prediksi yang telah ditunjukkan pada Gambar 5.



Gambar 6. Kurva Loss

4. KESIMPULAN

Berdasarkan hasil pembahasan Penerapan ERNN dalam Memprediksi laju penetrasi (ROP) pada Pengeboran Minyak Bumi di peroleh kesimpulan sebagai berikut:

1. Jaringan terbaik yang didapatkan pada model Elman RNN dengan jumlah 5 variabel input, yaitu WOB (x_1), TVD (x_2), MWI (x_3), ECD (x_4), dan TRPM (x_5), 17 neuron tersembunyi, serta 1 variabel output ROP (y).
2. Akurasi model yang diperoleh direpresentasikan oleh nilai MAPE dengan hasil prediksi pada data training 16.18%, sehingga menghasilkan nilai akurasi sebesar 83.82%. Sedangkan nilai MAPE pada data testing 15.48%, sehingga menghasilkan nilai akurasi sebesar 84.52%. Hasil prediksi ROP pada sumur B-1 dengan metode Elman *Recurrent Neural Network* tidak jauh berbeda dengan data aktual yang dipakai pada penelitian ini serta berdasarkan nilai MAPE hasil prediksi ini dikatakan baik.

Penelitian ini menunjukkan bahwa ERNN dapat menjadi alternatif metode prediksi ROP yang efektif. Namun, masih terdapat residual error pada beberapa titik data. Oleh karena itu, penelitian lanjutan disarankan untuk menguji model pada formasi geologi yang berbeda, mengeksplorasi arsitektur jaringan lain seperti LSTM atau GRU, serta melakukan optimasi *hyperparameter* untuk meningkatkan performa model.

DAFTAR PUSTAKA

- Abadi, A. M., Mansyaroh, A. K., Lukmana, A. H., & Harini, L. (2024). Construction of fuzzy systems based on fuzzy c-means clustering and singular value decomposition for predicting rate of penetration in geothermal drilling, *15* (4), 2190–2198. <https://doi.org/10.11591/ijped.v15.i4.pp2190-2198>
- Adebiyi, F. M. (2022). Environmental Chemistry and Ecotoxicology Air quality and management in petroleum refining industry: A review. *Environmental Chemistry and Ecotoxicology*, 4 (January), 89–96. <https://doi.org/10.1016/j.enceco.2022.02.001>
- Alkinani, H. H., Al-hameedi, A. T. T., & Dunn-norman, S. (2021). Data-driven recurrent neural network model to predict the rate of penetration Upstream Oil and Gas Technology. *Upstream Oil and Gas Technology*, 7 (May 2020), 100047. <https://doi.org/10.1016/j.upstre.2021.100047>
- Anemangely, M., Ramezanzadeh, A., & Tokhmechi, B. (2018). Drilling rate prediction from petrophysical logs and mud logging data using an optimized multilayer perceptron neural network. *Journal of Geophysics and Engineering*, 15(4), 1146–1159. <https://doi.org/10.1088/1742-2140/aaac5d>
- Asante, D., Omar, T., Ganat, A., Gholami, R., & Ridha, S. (2021). Journal of Petroleum Science and Engineering Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties: Comparative analysis of ANN and SVM models. *Journal of Petroleum Science and Engineering*, 200(November 2020), 108182.

<https://doi.org/10.1016/j.petrol.2020.108182>

- Bani, A., Ahmed, M., Mortadha, K. A., & Mamoon, A. (2021). Improving drilling performance through optimizing controllable drilling parameters. *Journal of Petroleum Exploration and Production*, 11(3), 1223–1232. <https://doi.org/10.1007/s13202-021-01116-2>
- Bourgoyne, A.T., and F.S. Young. "A Multiple Regression Approach to Optimal Drilling and Abnormal Pressure Detection." *SPE J.* 14 (1974): 371–384. doi: <https://doi.org/10.2118/4238-PA>
- Brenjkar, E., & Delijani, E. B. (2022). Computational prediction of the drilling rate of penetration (ROP): A comparison of various machine learning approaches and traditional models. *Journal of Petroleum Science and Engineering*, 210, 110033. <https://doi.org/10.1016/j.petrol.2021.110033>.
- Bromba, A. (1981). Application hints for Savitzky-Golay digital smoothing filters. *Analytical Chemistry*, 53(11), 1583–1586. <https://doi.org/10.1021/ac00234a011>
- Chen, W., & Zhou, K. (2017). Data quality of electricity consumption data in a smart grid environment. *Renewable and Sustainable Energy Reviews*, 75(October 2016), 98–105. <https://doi.org/10.1016/j.rser.2016.10.054>
- Elkatatny, S. (2021). Real-time prediction of rate of penetration while drilling complex lithologies using artificial intelligence techniques. *Ain Shams Engineering Journal*, 12(1), 917–926. <https://doi.org/10.1016/j.asej.2020.05.014>
- Farida, Y., Sulistiani, D. A., & Ulinuha, N. (2023). Klasifikasi data ketidakpastian menggunakan metode fuzzy decision tree. *Teorema: Teori dan Riset Matematika*, 6(2), 173–183. <https://doi.org/10.25157/teorema.v6i2.5521>
- Fu, Q., Nicholson, G. L., & Easton, J. M. (2024). Journal of Industrial Information Integration Understanding data quality in a data-driven industry context: Insights from the fundamentals. *Journal of Industrial Information Integration*, 42(November), 100729. <https://doi.org/10.1016/j.jii.2024.100729>
- Hegde, C., Daigle, H., Millwater, H., & Gray, K. (2017). Journal of Petroleum Science and Engineering Analysis of rate of penetration (ROP) prediction in drilling using physics-based and data-driven models. *Journal of Petroleum Science and Engineering*, 159, 295–306. <https://doi.org/10.1016/j.petrol.2017.09.020>
- Ihassanien, A. A., Nooh, A. Z., & Abu-Hashish, M. F. (2024). Pore pressure prediction using artificial intelligence techniques: A case study of Rudeis Formation at Al-Amal Oil Field, Gulf of Suez, Egypt. *Middle East Journal of Petroleum Sciences*, 13(1), 1–11.

<https://doi.org/10.62593/2090-2468.1045>

- Islam, M. R., & Hossain, M. E. (2021). Chapter 2 - State-of-the-art of drilling. In M. R. Islam & M. E. B. T.-D. E. Hossain (Eds.), *Sustainable Oil and Gas Development Series* (pp. 17–178). Gulf Professional Publishing. <https://doi.org/https://doi.org/10.1016/B978-0-12-820193-0.00002-2>
- Karkouch, A., Mousannif, H., Al, H., & Noel, T. (2016). Journal of Network and Computer Applications Data quality in internet of things: A state-of-the-art survey. *Journal of Network and Computer Applications*, 73, 57–81. <https://doi.org/10.1016/j.jnca.2016.08.002>
- Lin, C. Y., & Tjeerdema, R. S. (2008). Crude Oil, Oil, Gasoline and Petrol. In S. E. Jørgensen & B. D. B. T.-E. of E. Fath (Eds.) (pp. 797–805). Oxford: Academic Press. <https://doi.org/https://doi.org/10.1016/B978-008045405-4.00382-7>
- Radjabaycolle, J. (2020). Prediksi penggunaan bandwidth menggunakan Elman recurrent. *Barekeng: Jurnal Ilmu Matematika dan Terapan*, 10(2), 127–135. <https://doi.org/10.30598/barekengvol10iss2pp127-135>
- Rahmayanti, L., Rahmah, D. M., & Larashati. (2021). Minyak dan gas bumi di Indonesia. *Jurnal Sains Edukatika Indonesia (JSEI)*, 3(2), 9–16.
- Riazi, M., Mehrjoo, H., Nakhaei, R., & Jalalifar, H. (2022). Modelling rate of penetration in drilling operations using RBF , MLP , LSSVM , and DT models. *Scientific Reports*, 1–24. <https://doi.org/10.1038/s41598-022-14710-z>
- Sadeghi, M., & Behnia, F. (2018). *Optimum window length of Savitzky-Golay filters with arbitrary order*. arXiv. <https://doi.org/10.48550/arXiv.1808.10489>
- Savitzky, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8), 1627–1639. <https://doi.org/10.1021/ac60214a047>
- Sobhi, I., Dobbi, A., & Hachana, O. (2021). deterministic and metaheuristic optimization methods Prediction and analysis of penetration rate in drilling operation using deterministic and metaheuristic optimization methods. *Journal of Petroleum Exploration and Production Technology*, (December). <https://doi.org/10.1007/s13202-021-01394-w>
- Suahati, A. F., Nurrahman, A. A., & Rukmana, O. (2022). Penggunaan Jaringan Syaraf Tiruan – Backpropagation dalam Memprediksi Jumlah Mahasiswa Baru Predicting Number of New Student Using Artificial Neural Network - Backpropagation, 6(1), 21–29. <https://doi.org/10.35194/jmtsi.v6i1.1589>

- Sun, J., Yang, F., Qi, B., Li, Z., & Li, J. (2024). A New Calculation Model for Equivalent Circulating Density Considering Interface Effect between Various Fluids during Cementing Process. *SPE Journal*, 29(05), 2242–2256. <https://doi.org/10.2118/219481-PA>
- Wang, Y., Storey, V. C., & Firth, P. (1995). A framework for analysis of data quality research. *IEEE Transactions on Knowledge and Data Engineering*, 7(4), 554–568. <https://doi.org/10.1109/69.404034>
- Wang, Z., Fingas, M., Yang, C., & Christensen, J. H. (1964). 16 - Crude Oil and Refined Product Fingerprinting: Principles. In R. D. Morrison & B. L. B. T.-E. F. Murphy (Eds.) (pp. 339–407). Burlington: Academic Press. <https://doi.org/https://doi.org/10.1016/B978-012507751-4/50038-0>
- Xiong, M., Zheng, S., Cheng, R., Bai, K., Wang, L., Zhang, H., & Wang, G. (2024). A rate of penetration (ROP) prediction method based on improved dung beetle optimization algorithm and BiLSTM-SA. *arXiv*. <https://doi.org/10.21203/rs.3.rs-4255057/v1>
- Zhou, C., Zou, C., Zhang, Y., & Wang, Z. (2014). Nonparametric Control Chart Based on Change-point Model. *Computational Statistics & Data Analysis*, 51(2), 1237–1245. <https://doi.org/10.1007/s00362-007-0054-7>
- Zhou, C., Zou, C., Zhang, Y., & Wang, Z. (2014). Nonparametric Control Chart Based on Change-point Model. *Computational Statistics & Data Analysis*, 51(2), 1237–1245. <https://doi.org/10.1007/s00362-007-0054-7>