



KERJA PRAKTIK - IF184801

IMPLEMENTASI MACHINE LEARNING PADA DATA WELL LOG

Pertamina Hulu Energi

Jl. TB Simatupang No.Kav. 99, RT.1/RW.1, Kebagusan, Ps. Minggu, Kota Jakarta Selatan, Daerah Khusus Ibukota Jakarta 12520

Periode: 1 Maret 2023 - 31 Maret 2023

Oleh:

Satrio Hanif Wicaksono

05111940000103

Pembimbing Departemen

Dr. Anny Yuniarti, S.Kom., M.Comp.Sc.

Pembimbing Lapangan

Riswandi

DEPARTEMEN TEKNIK INFORMATIKA

Fakultas Teknologi Elektro dan Informatika Cerdas

Institut Teknologi Sepuluh Nopember

Surabaya 2023



KERJA PRAKTIK - IF184801

IMPLEMENTASI MACHINE LEARNING PADA DATA WELL LOG

Pertamina Hulu Energi

Jl. TB Simatupang No.Kav. 99, RT.1/RW.1, Kebagusan, Ps. Minggu, Kota Jakarta Selatan, Daerah Khusus Ibukota Jakarta 12520

Periode: 1 Maret 2023 - 31 Maret 2023

Oleh:

Satrio Hanif Wicaksono

05111940000103

Pembimbing Departemen

Dr. Anny Yuniarti, S.Kom., M.Comp.Sc.

Pembimbing Lapangan

Riswandi

DEPARTEMEN TEKNIK INFORMATIKA

Fakultas Teknologi Elektro dan Informatika Cerdas

Institut Teknologi Sepuluh Nopember

Surabaya 2023

[Halaman ini sengaja dikosongkan]

DAFTAR ISI

DAFTAR ISI	iv
DAFTAR GAMBAR	ix
LEMBAR PENGESAHAN	xii
KATA PENGANTAR	xvi
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Tujuan	2
1.3. Manfaat	2
1.4. Rumusan Masalah	2
1.5. Lokasi dan Waktu Kerja Praktik	3
1.6. Metodologi Kerja Praktik	3
1.6.1. Perumusan Masalah	3
1.6.2. Studi Literatur	3
1.6.3. Metodologi Pengerjaan	3
1.6.4. Hasil dan Pembahasan	4
1.6.5. Kesimpulan dan Saran	4
1.7. Sistematika Laporan	4
1.7.1. Pendahuluan	4
1.7.2. Profil Perusahaan	4
1.7.3. Tinjauan Pustaka	4
1.7.4. Metodologi	4

1.7.5. Hasil Dan Pembahasan	5
1.7.6. Kesimpulan dan Saran	5
BAB II PROFIL PERUSAHAAN	6
2.1. Profil Pertamina Hulu Energi	6
2.2. Lokasi	6
BAB III TINJAUAN PUSTAKA	7
3.1 Conventional Well Log	7
3.2 Badhole Identification	7
3.3 Shear Wave Sonic Log	8
3.4 Exploratory Data Analysis	9
3.5 DBSCAN	10
3.6 OPTICS	11
3.7 Boosted Random Forest	12
3.8 LSTM	12
3.9 Evaluasi Model	13
3.9.1 RMSE	13
3.9.2 MAE	14
3.9.3 R^2	14
BAB IV METODOLOGI	15
4.1. Alur Pekerjaan	15
4.1.1 Pemahaman Data	15
4.1.2 Persiapan Data	16
4.1.3 Pembentukan Model	16

4.1.4	Evaluasi dan Visualisasi	17
4.2.	Peralatan Pendukung	17
BAB V	HASIL DAN PEMBAHASAN	18
5.1	Pemahaman Data	18
5.2	Persiapan Data	22
5.2.1	<i>Dropped Data</i> untuk Cluster	22
5.2.2	<i>Filled Data</i> untuk Cluster	23
5.2.3	<i>Shear Sonic Wave Log Data</i>	27
5.3	Pembuatan Model	28
5.3.1	DBSCAN	28
5.3.2	OPTICS	28
5.3.3	Boosted Random Forest	29
5.3.4	LSTM	30
5.4	Evaluasi dan Visualisasi	30
5.4.1	DBSCAN	30
5.4.2	OPTICS	38
5.4.3	Boosted Random Forest	41
5.4.4	LSTM	42
BAB VI	KESIMPULAN DAN SARAN	44
6.1	Kesimpulan	44
6.2	Saran	44
DAFTAR PUSTAKA		45
BIODATA PENULIS I		49

[Halaman ini sengaja dikosongkan]

DAFTAR GAMBAR

Gambar 1 Contoh plot log yang menggambarkan visual dari beberapa kurva (Banas R., 2021)	8
Gambar 2 Proses Algoritma DBSCAN (M Ester, 1996).....	10
Gambar 3 Proses mengurutkan Data Point bervariasi (M Ester, 1999).....	11
Gambar 4 Alur Pekerjaan.....	15
Gambar 5 Peta Lokasi Well.....	16
Gambar 6 Missing Values Data	18
Gambar 7 Heatmap Korelasi Pearson	19
Gambar 8 Korelasi NPHI dan RHOB	20
Gambar 9 Hasil Clustering (Banas R, 2021)	20
Gambar 10 Korelasi DTS dengan RMED, RSHA, GR, RHOB, NPHI, DTC, RDEP	21
Gambar 11 Boxplot RMED, RSHA, GR, RHOB, NPHI, DTC, RDEP	22
Gambar 12 Scatterplot NPHI, RHOB, dan CALI Dropped Data.....	24
Gambar 13 Distribusi Data Cali	25
Gambar 14 Distribusi Data NPHI dan RHOB	25
Gambar 15 Feature Importance NPHI.....	25
Gambar 16 Feature Importance RHOB	26
Gambar 17 Scatterplot NPHI, RHOB, dan CALI Filled Data....	27
Gambar 18 Grafik Nilai Optimal Epsilon	29
Gambar 19 Arsitektur LSTM.....	30
Gambar 20 DBSCAN Dropped Data Cluster Scatter Plot.....	31
Gambar 21 DBSCAN Filled Data Cluster Scatter Plot.....	32
Gambar 22 Dropped Data DBSCAN Cluster Well 16/10-1	33
Gambar 23 Dropped Data DBSCAN Cluster Well 16/10-2	34
Gambar 24 Dropped Data DBSCAN Cluster Well 16/10-3	34
Gambar 25 Dropped Data DBSCAN Cluster Well 16/10-5	35

Gambar 26 Filled Data DBSCAN Cluster Well 16/10-1.....	35
Gambar 27 Filled Data DBSCAN Cluster Well 16/10-2.....	36
Gambar 28 Filled Data DBSCAN Cluster Well 16/10-3.....	36
Gambar 29 Filled Data DBSCAN Cluster Well 16/10-5.....	37
Gambar 30 Filled Data DBSCAN Cluster Well 16/11-1 ST3....	37
Gambar 31 Dropped Data OPTICS Cluster Well 16/10-1.....	38
Gambar 32 Dropped Data OPTICS Cluster Well 16/10-2.....	39
Gambar 33 Dropped Data OPTICS Cluster Well 16/10-3.....	39
Gambar 34 Filled Data OPTICS Cluster Well 16/10-1.....	40
Gambar 35 Filled Data OPTICS Cluster Well 16/10-3.....	40
Gambar 36 Filled Data OPTICS Cluster Well 16/11-1 ST3	41
Gambar 37 Visualisasi Well 16/10-5 dengan Prediksi DTS dengan Algoritma Boosted Random Forest	42
Gambar 38 Visualisasi Well 16/10-5 dengan Prediksi DTS dengan Algoritma LSTM	43

[Halaman ini sengaja dikosongkan]

**LEMBAR PENGESAHAN
KERJA PRAKTIK**

Implementasi Machine Learning Pada Data Well Log

Oleh:

Satrio Hanif Wicaksono

05111940000103

Disetujui oleh Pembimbing Kerja Praktik:

1. Dr. Anny Yuniarti, S.Kom.,
M.Comp.Sc.
NIP. 198106222005012002



(Pembimbing Departemen)

2. Riswandi



(Pembimbing Lapangan)

[Halaman ini sengaja dikosongkan]

Implementasi Machine Learning Pada Data Well Log

Nama Mahasiswa : Satrio Hanif Wicaksono
NRP : 05111940000103
Departemen : Teknik Informatika FTEIC-ITS
Pembimbing Departemen : Dr. Anny Yuniarti, S.Kom.,
M.Comp.Sc.
Pembimbing Lapangan : Riswandi

ABSTRAK

Kualitas data well logging dipengaruhi oleh beberapa masalah seperti data yang tidak tersedia, nilai yang salah, masalah penyelarasan, resolusi, dan kesalahan manusia. Ahli petrofisika menghabiskan waktu banyak untuk Quality control. Algoritma machine learning dapat digunakan untuk memprediksi data yang tidak tersedia, mendeteksi outlier, dan klasifikasi fasies reservoir. Machine learning juga dapat membantu memprediksi data log sonik jika terdapat hubungan kuantitatif antara kurva well log yang tersedia. Dalam Kerja Praktek ini, penulis menggunakan dataset dari Norwegian Petroleum Directorate untuk identifikasi Badhole dan prediksi Shear Sonic Log menggunakan algoritma DBSCAN, OPTICS, Boosted Random Forest, dan LSTM. Hasil analisis conventional well log menunjukkan bahwa algoritma DBSCAN lebih efektif dalam mengidentifikasi Badhole dibandingkan dengan algoritma OPTICS. Selain itu, algoritma Boosted Random Forest memberikan prediksi Shear Sonic Wave Log yang lebih akurat dibandingkan dengan algoritma LSTM.

Kata Kunci : Data Well Logging, Identifikasi Badhole, Shear Sonic Log

[Halaman ini sengaja dikosongkan]

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Allah SWT atas penyertaan dan karunia-Nya sehingga penulis dapat menyelesaikan salah satu kewajiban penulis sebagai mahasiswa Departemen Teknik Informatika ITS yaitu Kerja Praktik yang berjudul: Implementasi Machine Learning pada Data Well Log.

Penulis menyadari bahwa masih banyak kekurangan baik dalam melaksanakan kerja praktik maupun penyusunan buku laporan kerja praktik ini. Namun penulis berharap buku laporan ini dapat menambah wawasan pembaca dan dapat menjadi sumber referensi.

Melalui buku laporan ini penulis juga ingin menyampaikan rasa terima kasih kepada orang-orang yang telah membantu menyusun laporan kerja praktik baik secara langsung maupun tidak langsung antara lain:

1. Kedua orang tua penulis.
2. Ibu Dr. Anny Yuniarti, S.Kom., M.Comp. Sc. selaku dosen pembimbing kerja praktik sekaligus koordinator kerja praktik.
3. Bapak Riswandi selaku pembimbing lapangan selama kerja praktik berlangsung.
4. Teman-teman penulis yang senantiasa memberikan semangat ketika penulis melaksanakan KP.

Surabaya, 9 Juni 2023
Satrio Hanif Wicaksono

[Halaman ini sengaja dikosongkan]

BAB I

PENDAHULUAN

1.1. Latar Belakang

Kualitas data well logging dapat dipengaruhi oleh beberapa masalah, seperti data log yang tidak tersedia (data yang tidak tercatat pada saat akuisisi), nilai yang salah, masalah penyelarasan kedalaman, resolusi, dan kesalahan manusia. Oleh karena itu, sebagian besar ahli petrofisika menghabiskan waktu 50% - 90% proyek untuk melakukan pemeriksaan data, QC, pemuatan data, dan pembuatan laporan untuk informasi yang relevan (McDonald, 2021a). Mengidentifikasi sumur kondisi badhole sebelum analisis petrofisika merupakan proses yang memakan waktu. Meskipun demikian jika dilakukan dengan benar, itu dapat menghasilkan data yang memiliki added value yang tinggi.

Saat ini, algoritma machine learning dengan alat seperti Python dapat digunakan untuk memprediksi kurva logging yang tidak tersedia (Banas R., 2021), mendeteksi outlier (Akkurt et al., 2018), dan klasifikasi fasies reservoir (Cuddy & Putnam, 1998). Diharapkan dengan menggunakan algoritma machine learning dapat membantu mengurangi proses yang memakan waktu tersebut.

Log sonik sangat penting untuk identifikasi jenis batuan, jenis hidrokarbon, pemodelan fisika batuan, dan karakterisasi reservoir. Namun, data log tersebut jarang tersedia karena tingginya biaya akuisisi atau kesalahan pengukuran. Rumus empiris dan model petroelastic yang sering digunakan seperti Effective Medium Models (Kuster dan Toksoz, 1974), Heuristic (Xu dan White, 1994), dan Empiris (Greenberg dan Castagna, 1992), untuk memprediksi log sonik yang tidak tersedia mungkin tidak menghasilkan profil yang akurat dan terbatas pada jenis batuan dengan saturasi fluida tertentu. Oleh karena itu diharapkan machine learning dapat digunakan untuk memprediksi data log

sonik jika terdapat hubungan kuantitatif antara kurva well log yang tersedia.

Pada Kerja Praktek ini penulis menggunakan dataset dari kompetisi yang bekerja sama dengan Norwegian Petroleum Directorate. Melakukan identifikasi Badhole menggunakan algoritma DBSCAN dan OPTICS. Kemudian melakukan prediksi Shear Sonic Log menggunakan algoritma Boosted Random Forest dan LSTM.

1.2. Tujuan

Tujuan kerja praktik ini adalah sebagai berikut:

1. Mengetahui proses dan melakukan clustering pada dataset conventional well log untuk melakukan identifikasi Badhole.
2. Mengetahui proses dan melakukan prediksi Shear Sonic Log pada dataset *conventional well log*.

1.3. Manfaat

Manfaat yang diperoleh dengan mengetahui proses clustering dan prediksi Shear Sonic Log dapat membantu ahli petrofisik meningkatkan efisiensi waktu untuk upaya identifikasi badhole dan melakukan prediksi log yang tidak tersedia dari proses akusisi log dalam mendukung operasi dan studi yang dilakukan.

1.4. Rumusan Masalah

Rumusan masalah dari kerja praktik ini adalah sebagai berikut:

1. Bagaimana melakukan *clustering* pada dataset *conventional well log* untuk melakukan identifikasi Badhole?
2. Bagaimana melakukan prediksi Shear Sonic Log pada dataset *conventional well log*?

1.5. Lokasi dan Waktu Kerja Praktik

Pengerjaan kerja praktik ini dilakukan secara *remote*. Adapun kerja praktik dimulai pada tanggal 1 Maret 2023 hingga 31 Maret 2023.

1.6. Metodologi Kerja Praktik

Metodologi dalam pembuatan buku kerja praktik meliputi :

1.6.1. Perumusan Masalah

Untuk mengetahui masalah yang akan diselesaikan, dilakukan rapat dengan Mentor. Pada saat rapat dijelaskan pekerjaan yang akan dilakukan yaitu Unsupervised Badhole Identification dan Shear Wave Sonic Log Prediction. Pada penjelasan ini telah dibahas latar belakang, *deliverables*, dan alur pekerjaan.

1.6.2. Studi Literatur

Setelah mendapat penjelasan pekerjaan yang akan dilakukan, dilakukan tinjauan yang digunakan pada pekerjaan. Tinjauan yang digunakan meliputi Conventional Well Log, Badhole Identification, Shear Wave Sonic Log, dan lain-lain.

1.6.3. Metodologi Pengerjaan

Metodologi pengerjaan yang dilakukan mengikuti alur pekerjaan. Langkah-langkah yang dilakukan yaitu pemahaman data, persiapan data, pembentukan model, evaluasi, dan visualisasi.

1.6.4. Hasil dan Pembahasan

Hasil dan pembahasan merupakan implementasi dari metodologi pengerjaan.

1.6.5. Kesimpulan dan Saran

Hasil dan pembahasan yang telah diimplementasikan memberikan hasil yang memuaskan dengan teridentifikasi Badhole dan menghasilkan akurasi prediksi Shear Sonic Log yang memuaskan.

1.7. Sistematika Laporan

1.7.1. Pendahuluan

Bab ini berisi latar belakang, tujuan, manfaat, rumusan masalah, lokasi dan waktu kerja praktik, metodologi, dan sistematika laporan.

1.7.2. Profil Perusahaan

Bab ini berisi gambaran umum Pertamina Hulu Energi mulai dari profil, lokasi perusahaan.

1.7.3. Tinjauan Pustaka

Bab ini berisi dasar teori dari teknologi yang digunakan dalam menyelesaikan kerja praktik.

1.7.4. Metodologi

Hasil dan pembahasan merupakan implementasi dari metodologi pengerjaan.

1.7.5. Hasil Dan Pembahasan

Bab ini berisi hasil uji coba dan evaluasi dari aplikasi yang telah dikembangkan selama pelaksanaan kerja praktik.

1.7.6. Kesimpulan dan Saran

Bab ini berisi kesimpulan dan saran yang didapat dari proses pelaksanaan kerja praktik.

BAB II

PROFIL PERUSAHAAN

2.1. Profil Pertamina Hulu Energi

PT Pertamina Hulu Energi atau biasa disingkat menjadi PHE, adalah anak usaha Pertamina yang bergerak di bidang hulu minyak dan gas. Kegiatan usaha Pertamina di sektor hulu yang dikelola oleh Direktorat hulu mencakup kegiatan eksplorasi, pengeboran, pengembangan dan produksi minyak, gas dan panas bumi, penyediaan jasa teknologi, serta jasa pemboran dan services baik dalam maupun luar negeri. Dalam rangka mencapai pertumbuhan sektor hulu (upstream growth) sebagai salah satu dari 8 Pilar Prioritas Strategi Pertamina, strategi usaha di sektor hulu adalah meningkatkan produksi dan menambah cadangan migas baru, baik secara organik melalui kegiatan Improved Oil Recovery (IOR) dan Enhanced Oil Recovery (EOR) pada aset yang telah ada, maupun secara anorganik dengan melakukan strategi merger and acquisition (M&A) blok-blok migas di dalam maupun di luar negeri.

2.2. Lokasi

Jl. TB Simatupang No.Kav. 99, RT.1/RW.1, Kebagusan,
Ps. Minggu, Kota Jakarta Selatan, Daerah Khusus Ibukota Jakarta
12520

BAB III

TINJAUAN PUSTAKA

3.1 Conventional Well Log

Conventional well log data dapat digunakan oleh petrofisis untuk analisa jenis batuan dan fluida dari sumur yang bersifat *indirect measurement*. Penggunaan conventional well log data sangat penting, salah satunya adalah dalam menentukan zona penghasil minyak dan gas.

Beberapa kurva yang tersedia dalam conventional well log antara lain: Gamma Ray Log, mengukur jumlah radiasi gamma untuk identifikasi jenis batuan. Resistivity Log, mengukur resistivitas fluida yang ada di batuan untuk identifikasi zona penghasil minyak dan gas. Sonic Log, mengukur kecepatan suara yang merambat melalui batuan untuk identifikasi lapisan batuan dan struktur geologi. Density Log, mengukur densitas batuan untuk identifikasi jenis batuan dan struktur geologi. Porosity Log, mengukur porositas batuan untuk identifikasi zona penghasil minyak dan gas. Caliper Log digunakan untuk mengetahui ukuran diameter lubang.

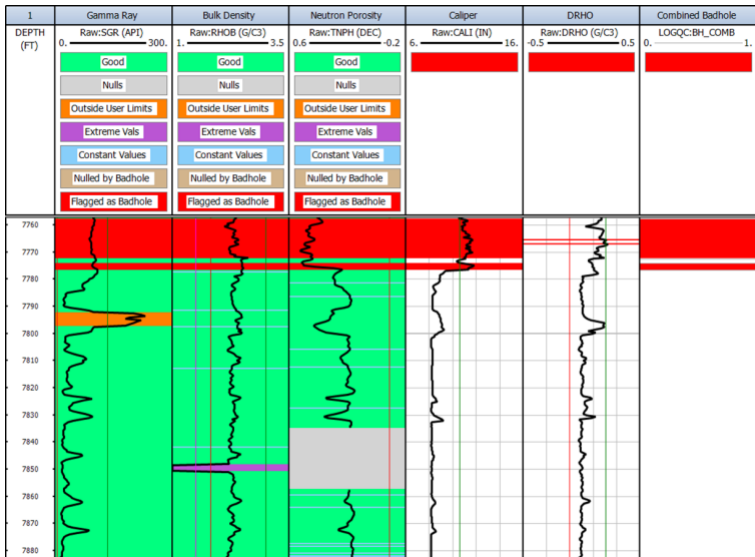
Penggunaan kurva-kurva conventional well log membantu para ahli geologi dan insinyur dalam menentukan zona penghasil minyak dan gas serta menentukan strategi pengeboran yang tepat.

3.2 Badhole Identification

Badhole identification adalah proses identifikasi kondisi di mana sumur bor setelah selesai operasi pengeboran memiliki kondisi yang tidak baik atau buruk. Sulit untuk mengidentifikasi data yang buruk dengan hanya melihat data dan kurva individual, oleh karena itu menggunakan pendekatan secara visual dapat meningkatkan identifikasi Badhole.

Kurva yang paling umum digunakan adalah koreksi densitas (DRHO) dan caliper. Ini disajikan dalam kurva Caliper

dan DRHO pada Gambar 1 dan dapat digunakan untuk mengidentifikasi area *washout*.



Gambar 1 Contoh plot log yang menggambarkan visual dari beberapa kurva (Banas R., 2021)

Dapat diamati bahwa sebagian besar data dianggap dalam kondisi baik (diuraikan warna hijau). Saat lubang bor *washed out*, alat-alat logging umumnya mampu mengkompensasi sampai tingkat tertentu untuk *washout*. Namun, ketika *washout* menjadi parah, terdapat cairan lubang bor atau *mud cake* yang lebih besar antara dinding lubang bor dan sensor alat logging.

3.3 Shear Wave Sonic Log

Shear Wave Sonic Log adalah jenis kurva pada conventional well log yang digunakan untuk mengukur kecepatan gelombang geser dalam batuan yang dilalui oleh sumur bor. Pengukuran ini dilakukan dengan mengirimkan gelombang suara

melalui sumur bor menggunakan alat sonic tool, yang kemudian merekam dan memproses data untuk menghitung kecepatan gelombang geser di dalam batuan. Kecepatan gelombang geser sangat penting dalam industri minyak dan gas karena dapat memberikan informasi yang berguna tentang struktur batuan dan sifat geologi, serta dapat membantu dalam menentukan jenis metode pengeboran yang paling efektif.

Kecepatan gelombang kompresional, geser, dan Stoneley adalah parameter penting, yang memiliki banyak aplikasi interpretasi geofisika, studi geomekanika dan karakterisasi reservoir. Shear Wave Sonic (v_s) dengan kecepatan gelombang kompresional (v_p) dapat memberikan data yang sangat berharga untuk studi reservoir seperti litologi identifikasi (Pickett 1963), sifat mekanik batuan perhitungan (Ameen et al. 2009), dan identifikasi tipe pori (Eberli et al. 2003). Kecepatan gelombang Stoneley (v_{st}) memiliki peran dalam estimasi permeabilitas (Balossino et al. 2008), perhitungan stabilitas lubang sumur (Wendt et al. 2007), evaluasi rekahan (Brie et al. 2000), mendeteksi hampan gas tipis (Tang dan Patterson 2001), dan perkiraan anisotropi formasi (Tang dan Patterson 2004).

3.4 Exploratory Data Analysis

Exploratory Data Analysis (EDA) adalah suatu metode dalam analisis data yang digunakan untuk mengeksplorasi dan menganalisis dataset dengan tujuan memahami karakteristik dan pola data. EDA bertujuan untuk mengidentifikasi pola, hubungan, dan anomali dalam data, serta untuk memvisualisasikan data secara grafis agar dapat memperoleh pemahaman yang lebih dalam.

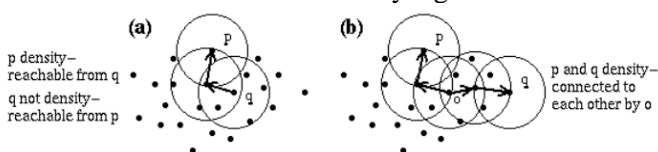
EDA dilakukan pada awal proses analisis data sebagai langkah pertama sebelum melakukan model prediksi atau inferensi statistik. Beberapa teknik EDA yang umum digunakan adalah histogram, scatter plot, box plot, dan density plot. EDA juga dapat melibatkan teknik statistik deskriptif seperti mean, median, modus, dan standar deviasi.

EDA membantu mengidentifikasi pola dan anomali dalam data yang mungkin terlewatkan jika hanya melihat data secara statistik deskriptif biasa. Dengan memahami karakteristik dan pola data melalui EDA, analis dapat membuat keputusan yang lebih baik dalam merancang model dan memilih teknik analisis yang paling tepat untuk data tersebut.

3.5 DBSCAN

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) adalah suatu algoritma clustering yang digunakan untuk mengelompokkan data dengan mempertimbangkan kepadatan data pada suatu wilayah tertentu (M Ester, 1996). Algoritma ini menganggap bahwa cluster adalah area dengan kepadatan data yang tinggi, sedangkan area dengan kepadatan data yang rendah dianggap sebagai noise atau outlier.

DBSCAN bekerja dengan cara memulai dari satu data point tertentu dan mencari data point yang berdekatan dengannya dengan mempertimbangkan suatu nilai epsilon (radius) dan minimum jumlah data point untuk membentuk cluster. Jika jumlah data point yang ditemukan memenuhi kriteria minimum jumlah data point, maka data point tersebut akan dianggap sebagai inti cluster dan akan menjadi pusat dari suatu cluster. Kemudian, DBSCAN akan mencari data point yang berdekatan dengan inti cluster dan menambahkan data point tersebut ke dalam cluster yang sama. Proses ini akan berulang hingga tidak ada lagi data point yang dapat ditambahkan ke dalam cluster yang sama.



Gambar 2 Proses Algoritma DBSCAN (M Ester, 1996).

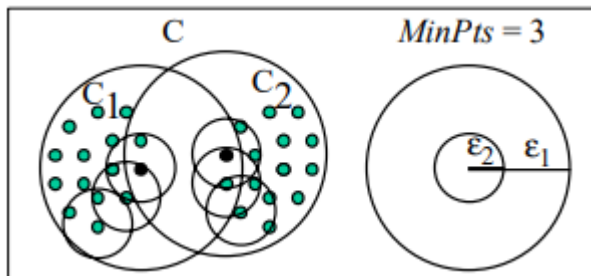
DBSCAN memiliki keunggulan dalam mengidentifikasi cluster yang berbentuk non-sferis dan dapat mengatasi masalah

pencilan (outlier) dengan mengabaikan data point yang memiliki kepadatan rendah. Selain itu, DBSCAN juga tidak memerlukan jumlah cluster yang harus ditentukan sebelumnya.

3.6 OPTICS

OPTICS (Ordering Points To Identify the Clustering Structure) adalah algoritma clustering yang digunakan untuk mengelompokkan data dengan mempertimbangkan kepadatan data dan jarak antar data point. Algoritma ini merupakan pengembangan dari algoritma DBSCAN dengan tujuan untuk meningkatkan kemampuan DBSCAN dalam mengatasi masalah clustering pada data dengan kepadatan yang bervariasi (M Ester, 1999).

OPTICS bekerja dengan cara membangun suatu grafik dalam ruang parameter dimana setiap data point diwakili oleh suatu titik pada grafik tersebut. Kemudian, OPTICS akan menghitung jarak antar titik pada grafik dan membangun suatu relasi hubungan antar data point berdasarkan jarak dan kepadatan data pada suatu wilayah tertentu. Proses ini akan menghasilkan suatu urutan data point berdasarkan tingkat kepadatan yang terkait dengan setiap data point.



Gambar 3 Proses mengurutkan Data Point bervariasi (M Ester, 1999).

Dengan urutan data point ini, OPTICS dapat mengidentifikasi cluster berdasarkan perbedaan tingkat kepadatan dan jarak antar data point. Cluster yang memiliki kepadatan tinggi

dan jarak antar data point yang pendek akan diberi nilai rendah pada grafik, sementara cluster yang memiliki kepadatan rendah dan jarak antar data point yang jauh akan diberi nilai tinggi pada grafik.

OPTICS memiliki keunggulan dalam mengatasi masalah clustering pada data dengan kepadatan yang bervariasi dan dapat menghasilkan cluster berdasarkan perbedaan tingkat kepadatan dan jarak antar data point. Algoritma ini juga dapat mengatasi masalah jumlah cluster yang tidak diketahui sebelumnya.

3.7 Boosted Random Forest

Boosted Random Forest adalah suatu algoritma Machine Learning yang digunakan untuk melakukan klasifikasi atau regresi pada data dengan meningkatkan kinerja model Random Forest melalui teknik boosting (Y Mishina, 2014). Algoritma ini menggabungkan dua teknik populer yaitu Random Forest dan Boosting untuk meningkatkan akurasi prediksi pada data yang kompleks dan tidak linear.

Pada Boosted Random Forest, model Random Forest dianggap sebagai model dasar (base model) yang digunakan untuk membuat prediksi awal pada data. Kemudian, setiap prediksi yang salah akan diberikan bobot yang lebih besar pada data berikutnya. Hal ini akan memperkuat fokus model terhadap data yang sulit diklasifikasikan oleh model dasar dan mendorong model untuk menghasilkan prediksi yang lebih akurat.

Boosted Random Forest juga menggunakan teknik subsampling pada data training untuk menghasilkan variasi model yang lebih banyak. Hal ini dapat membantu mengurangi overfitting dan meningkatkan generalisasi model.

3.8 LSTM

Algoritma Long Short-Term Memory (LSTM) adalah tipe khusus dari Recurrent Neural Network (RNN) yang terdiri dari unit pengendali dan unit memori. LSTM dirancang untuk menangani

masalah "long-term dependencies" atau ketergantungan jangka panjang dalam data sekuensial seperti teks, suara, dan video.

LSTM mempertahankan sebuah memori yang dapat menambah atau menghapus informasi selama proses pembelajaran. Memori ini dikontrol oleh tiga gerbang: gerbang input, gerbang keluaran, dan gerbang lupa. Gerbang input mengontrol kapan informasi baru harus ditambahkan ke memori, gerbang keluaran mengontrol kapan memori harus digunakan, dan gerbang lupa mengontrol kapan informasi harus dihapus dari memori.

Pada setiap langkah waktu, input sekuensial masuk ke dalam LSTM dan diolah melalui gerbang input untuk menentukan bagaimana informasi tersebut harus ditambahkan ke memori. Kemudian, memori diubah atau dihapus melalui gerbang lupa, dan hasil akhirnya dihasilkan melalui gerbang keluaran.

3.9 Evaluasi Model

Saat melakukan evaluasi pada model ada banyak metode yang dapat dilakukan. Beberapa metode evaluasi yang akan digunakan yaitu RMSE (Root Mean Square error), MAE (Mean Absolute Error), dan R^2 (Coefficient of Determination).

3.9.1 RMSE

Root Mean Square Error (RMSE) merupakan besarnya tingkat kesalahan hasil prediksi, dimana semakin kecil (mendekati 0) nilai RMSE maka hasil prediksi akan semakin akurat. Nilai RMSE dapat dihitung dengan persamaan sebagai berikut.

$$RMSE = \sqrt{\frac{\sum (X - Y)^2}{n}}$$

3.9.2 MAE

Mean Absolute Error (MAE) adalah salah satu metode yang digunakan untuk mengukur tingkat keakuratan model peramalan. Nilai MAE menunjukkan rata – rata kesalahan (error) absolut antara hasil peramalan/prediksi dengan nilai riil. Secara rumus MAE dijelaskan sebagai berikut,

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_i - y_i|$$

f_i adalah nilai hasil prediksi dan
 y_i adalah nilai sebenarnya.

3.9.3 R²

Coefficient of Determination (R²) adalah pengukuran statistik yang mengkaji bagaimana perbedaan dalam satu variabel dapat dijelaskan oleh perbedaan dalam variabel kedua saat memprediksi hasil dari suatu peristiwa tertentu. Dengan kata lain, koefisien ini menilai seberapa kuat hubungan linier antara dua variabel dan sangat diandalkan saat melakukan analisis tren. Nilai R² dapat dihitung dengan persamaan sebagai berikut.

$$R^2 = 1 - \frac{RSS}{TSS}$$

Dimana:

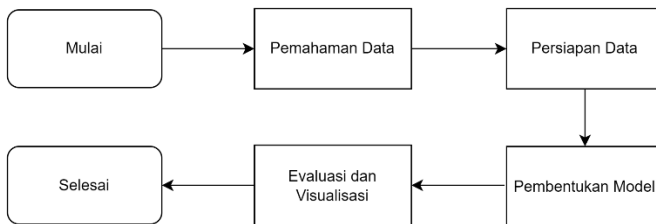
RSS adalah jumlah sisa kuadrat dan

TSS adalah jumlah total kuadrat.

BAB IV METODOLOGI

4.1. Alur Pekerjaan

Proses pengerjaan mengikuti alur sebagai berikut:



Gambar 4 Alur Pekerjaan

4.1.1 Pemahaman Data

Pada tahap ini, data yang akan digunakan dipelajari secara rinci. Tahap ini akan dilakukan EDA (Exploratory Data Analysis) untuk mencari korelasi antar kurva menggunakan metode Pearson. Metode Pearson membantu mengetahui hubungan antara dua variabel dan memberi ukuran kekuatan hubungan antara dua variabel. Nilai Koefisien Korelasi Pearson bisa antara -1 hingga +1. 1 berarti sangat berkorelasi, 0 berarti tidak ada korelasi, dan -1 berarti korelasi negatif (Cobb et al., 1987). Dengan metode ini kemudian dapat ditentukan kurva yang akan digunakan untuk melakukan *supervised* ataupun *unsupervised learning*. Setiap kurva kemudian akan dilakukan visualisasi line chart dan box plot.

Data yang digunakan merupakan data dari kompetisi kerja sama dengan Norwegian Petroleum Directorate. Data tersebut kemudian akan digunakan 5 well yaitu 16/10-1, 16/10-2, 16/10-3, 16/10-5, dan 16/11-1 ST3. Gambar 5 merupakan lokasi dari Well tersebut. Jenis kurva yang akan digunakan dari *conventional well log*.



Gambar 5 Peta Lokasi Well

4.1.2 Persiapan Data

Pada tahap ini, data yang telah dipelajari pada tahap sebelumnya harus dipersiapkan untuk proses data mining. Proses persiapan data meliputi penggabungan data, pemilihan atribut, dan normalisasi data. Setelah itu pada proses prediksi Shear Sonic Log dilakukan pembagian data 70% latih, 20% data validasi, dan 10% data training.

4.1.3 Pembentukan Model

Tahap ini melibatkan pengembangan model, validasi model, dan penyesuaian model. Dalam melakukan clustering untuk identifikasi Badhole akan digunakan algoritma DBSCAN, dimana DBSCAN mencari area dalam data yang memiliki kepadatan pengamatan yang tinggi, versus area data yang tidak terlalu padat dengan pengamatan (Banas R, 2021). Hasil dari clustering DBSCAN dapat menangani outlier dalam dataset, dimana data outlier dari dataset tersebut merupakan Badhole atau washout. OPTICS berdasarkan modifikasi dari model DBSCAN (M Ankerst, 1999). Algoritma OPTICS hanya menggunakan input

epsilon untuk mengurangi kompleksitas waktu proses, yang mempermudah dalam melakukan *fine-tuning*.

4.1.4 Evaluasi dan Visualisasi

Pada tahap ini, model yang telah dikembangkan harus dievaluasi dan disesuaikan dengan tujuan bisnis. Evaluasi model harus mencakup pengujian model pada data yang tidak digunakan selama proses modeling. Pada prediksi Shear Sonic Log digunakan metode evaluasi RMSE, MAE, dan R^2 yang dilengkapi dengan visualisasi.

4.2. Peralatan Pendukung

Kerja praktek ini menggunakan perangkat keras laptop dan perangkat lunak dengan yang memadai sehingga kerja praktek dapat berjalan dengan baik. Spesifikasi perangkat keras dapat dilihat pada tabel berikut.

Tabel 1 Tabel Spesifikasi Perangkat Keras dan Lunak

No.	Jenis Spesifikasi	Spesifikasi
1	Memori	16 GB DDR4-2666 SDRAM
2	CPU	Intel® Core™ i7-8750H
3	GPU	NVIDIA GeForce GTX 1060
4	Sistem Operasi	Windows 10 Home 64
5	Bahasa Pemrograman	Python

BAB V HASIL DAN PEMBAHASAN

Bab ini membahas hasil dari alur pekerjaan, yaitu pemahaman data (EDA), persiapan data, pembentukan model, evaluasi dan visualisasi.

5.1 Pemahaman Data

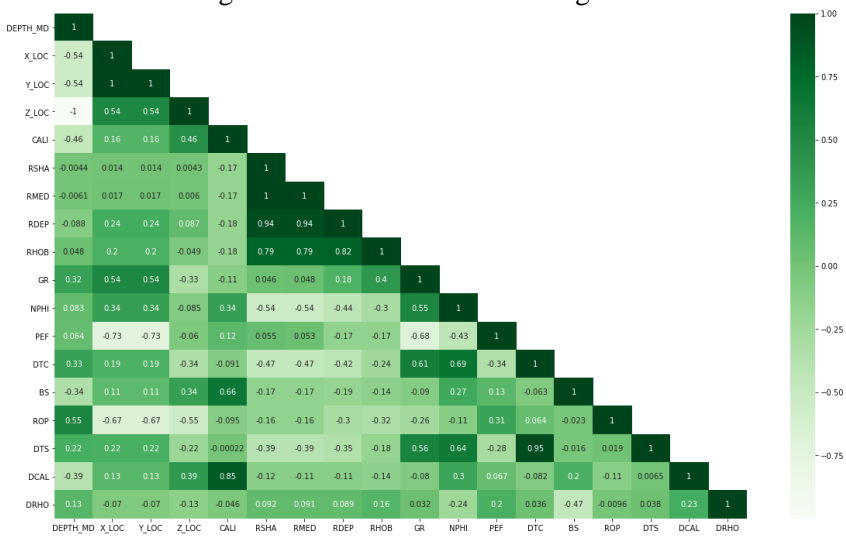
Data yang digunakan memiliki 29 column dan 53771 rows dengan nilai *missing values* sebagai berikut.

	Total Missing	%
WELL	0	0.0
DEPTH_MD	0	0.0
X_LOC	7384	13.7
Y_LOC	7384	13.7
Z_LOC	7384	13.7
GROUP	0	0.0
FORMATION	12721	23.7
CALI	100	0.2
RSHA	25880	48.1
RMED	7537	14.0
RDEP	7384	13.7
RHOB	23402	43.5
GR	0	0.0
SGR	53771	100.0
NPFI	28944	53.8
PEF	27233	50.6
DTC	168	0.3
SP	15162	28.2
BS	47	0.1
ROP	2525	4.7
DTS	48828	90.8
DCAL	15208	28.3
DRHO	22853	42.5
MUDWEIGHT	10014	18.6
RMIC	53771	100.0
ROPA	51006	94.9
RXO	47303	88.0
FORCE_2020_LITHOFACIES_LITHOLOGY	0	0.0
FORCE_2020_LITHOFACIES_CONFIDENCE	5	0.0

Gambar 6 Missing Values Data

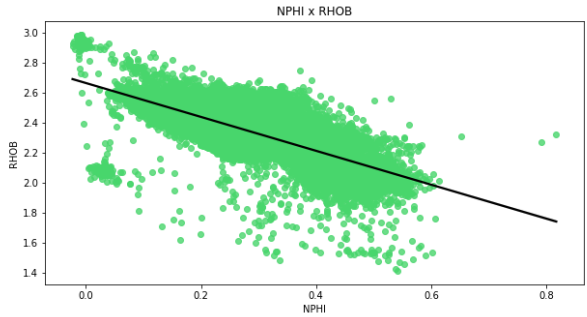
Berdasarkan banyaknya data yang tidak ada maka akan dibagikan menjadi 3 jenis dimana 2 jenis akan digunakan untuk clustering dan 1 jenis untuk dilakukan prediksi. Dalam clustering dibagi menjadi 2 yaitu *filled missing* data (mengisi data yang kosong) dan *dropped missing* data (menghapus data yang kosong) karena banyaknya data yang *missing* perlu dilakukan percobaan dari kedua jenis data.

Kemudian dicari korelasi antar kurva dimana hasil tersebut dalam bentuk heatmap menggunakan metode Pearson dimana nilai yang mendekati satu memiliki korelasi positif dan nilai yang mendekati nilai negatif satu memiliki korelasi negatif.

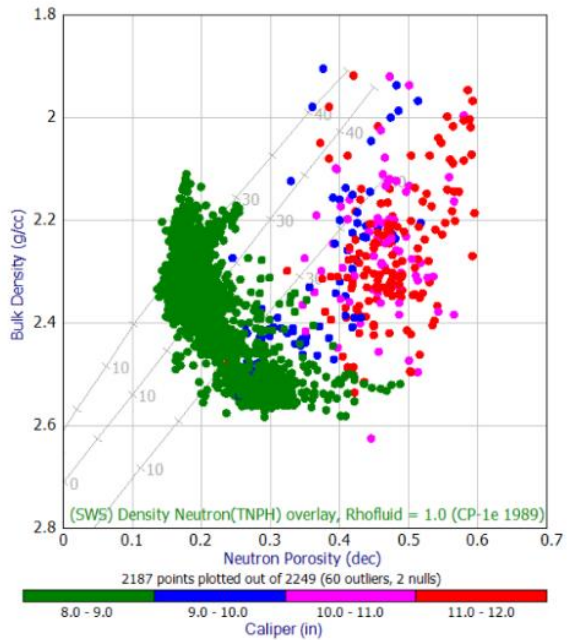


Gambar 7 Heatmap Korelasi Pearson

Dalam melakukan identifikasi Badhole digunakan NPHI dan RHOB dimana berdasarkan riset yang telah dilakukan menghasilkan clustering yang cukup memuaskan. Korelasi tersebut dapat dilihat dari gambar berikut. Berdasarkan korelasi ini, menguatkan kembali penggunaan kurva NPHI dan RHOB untuk melakukan Badhole Identification (Banas R, 2021).

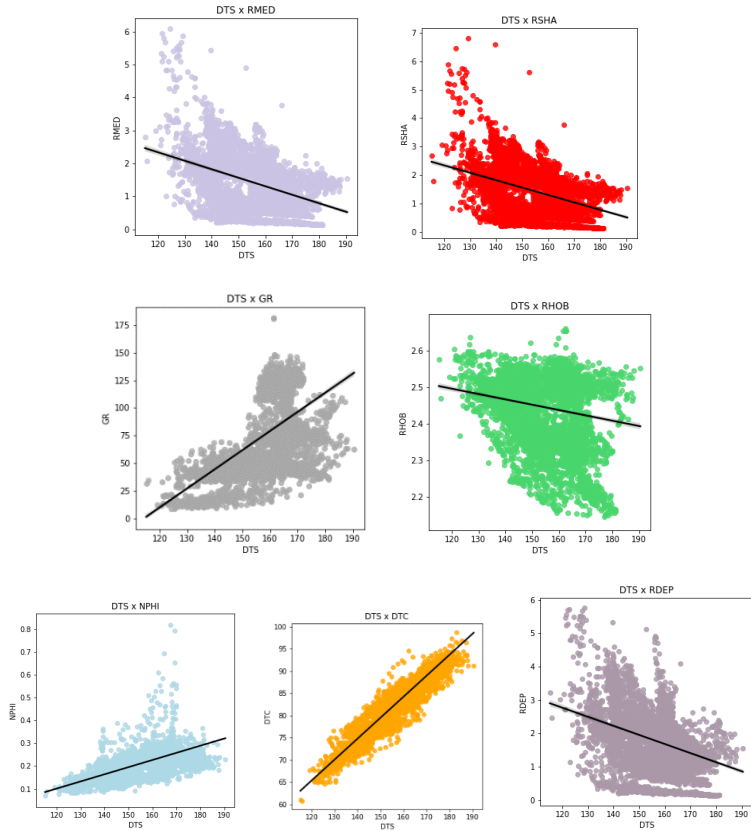


Gambar 8 Korelasi NPHI dan RHOB



Gambar 9 Hasil Clustering (Banas R, 2021)

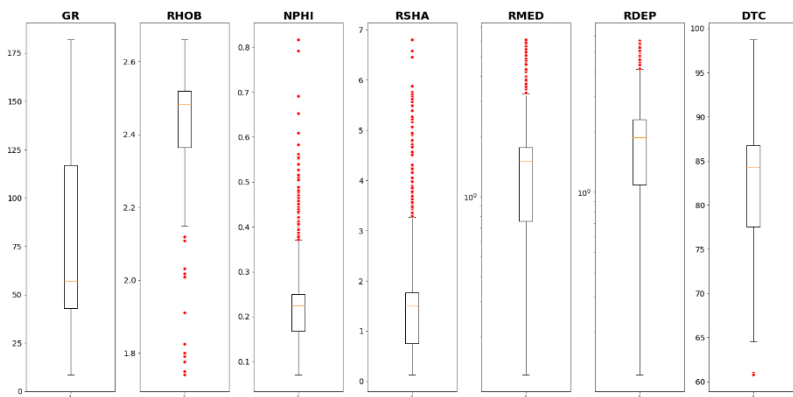
Dalam melakukan prediksi Shear Wave Sonic log ditemukan tujuh variable dengan nilai korelasi tinggi. Korelasi tersebut dapat dilihat dari gambar berikut.



Gambar 10 Korelasi DTS dengan RMED, RSHA, GR, RHOB, NPHI, DTC, RDEP

Dari tujuh variable tersebut kemudian dilakukan pengecekan menggunakan boxplot pada gambar sebagai berikut. Dapat dilihat bahwa pada RHOB, NPHI, RSHA, RMED, dan RDEP memiliki outlier yang berbeda range. Agar mendapatkan bentuk distribusi

yang sama antar variable maka kemudian perlu dilakukan normalisasi.



Gambar 11 Boxplot RMED, RSHA, GR, RHOB, NPHI, DTC, RDEP

5.2 Persiapan Data

Proses persiapan data dibagi menjadi 3 menyesuaikan dengan jenis data yang akan digunakan. Dari 3 jenis data tersebut kemudian masing – masing dilakukan Normalisasi dengan metode MinMaxScaler. MinMaxScaler mempertahankan bentuk distribusi aslinya. Itu tidak mengubah informasi yang disematkan (embedded) dalam data asli dimana hasil normalisasi memiliki nilai antara 0 atau 1.

5.2.1 Dropped Data untuk Cluster

Dalam proses pembuatan Dropped Data hanya dilakukan penghapusan data yang memiliki data kosong. Berikut merupakan hasil dari dilakukan penghapusan data pada tabel 1. Dapat dilihat scatter plot antara hubungan NPHI, RHOB, dan CALI pada tabel 1.

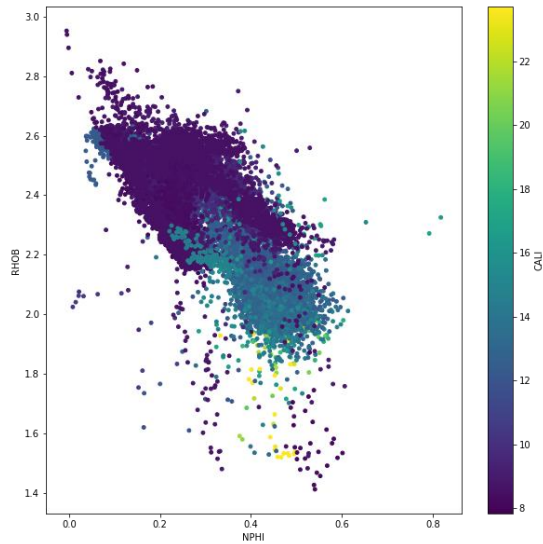
Tabel 2 Dropped Data untuk Clustering

	RHOB	NPHI	DCAL	CALI
count	21050.000000	21050.000000	21050.000000	21050.000000
mean	2.353782	0.283298	0.374182	10.393915
std	0.190918	0.120313	1.012582	2.287879
min	1.412190	-0.005300	-3.515626	7.839270
25%	2.211841	0.188997	-0.067246	8.564735
50%	2.403945	0.253914	0.081525	8.929583
75%	2.508923	0.395357	0.404945	12.409361
max	2.952480	0.817310	11.203100	23.703100

Dapat dilihat scatter plot antara hubungan NPHI, RHOB, dan CALI dengan Dropped Data pada gambar 12.

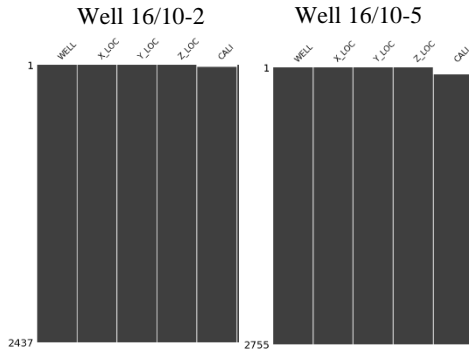
5.2.2 *Filled Data* untuk Cluster

Dalam proses pembuatan Filled Data dilakukan pengisian pada beberapa data yaitu CALI, NPHI, dan RHOB.

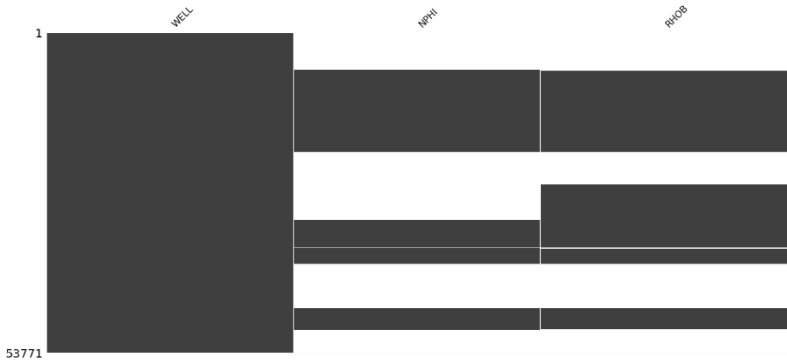


Gambar 12 Scatterplot NPHI, RHOB, dan CALI Dropped Data

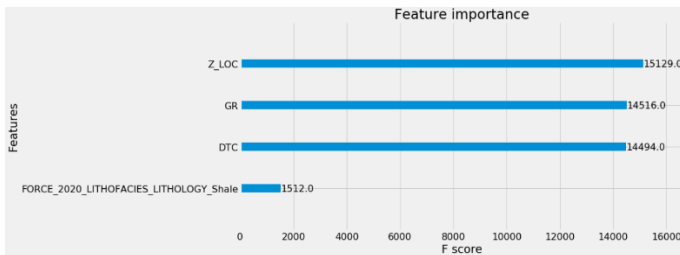
Dimana pada kurva CALI hanya memiliki *missing data* pada awal data dapat dilihat pada gambar 13 maka digunakan metode *forward fill* dimana diisi data tersebut menggunakan data didepannya. Dimana NPHI dan RHOB memiliki data tidak hanya diawal dapat dilihat pada gambar 14, tetapi di tengahnya juga ada. Maka digunakan algoritma Boosted Random Forest untuk mengisi kurva tersebut. Dapat dilihat pada gambar 15 dan 16 fitur apa saja yang digunakan untuk melakukan pengisian data.



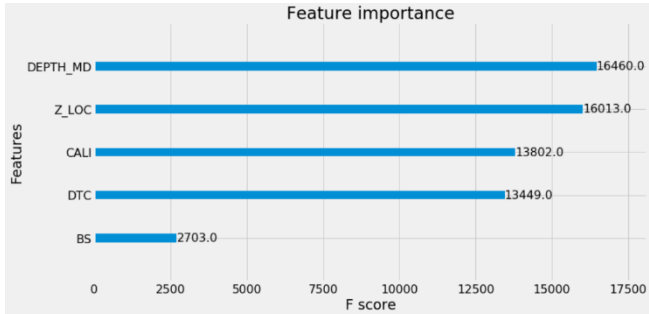
Gambar 13 Distribusi Data Cali



Gambar 14 Distribusi Data NPHI dan RHOB



Gambar 15 Feature Importance NPHI

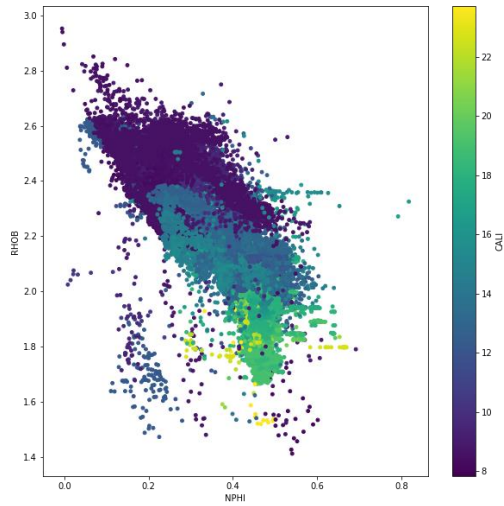


Gambar 16 Feature Importance RHOB

Hasil dilakukan pengisian data dapat dilihat sebagai berikut:

Tabel 3 Filled Data for Clustering

	RHOB	NPHI	DCAL	CALI
count	46387.000000	46387.000000	46387.000000	46387.000000
mean	2.182613	0.359941	0.563867	13.236510
std	0.232430	0.118454	1.366899	3.649543
min	1.412190	-0.005300	-8.475031	7.839270
25%	2.010826	0.253541	0.042971	9.201701
50%	2.140941	0.413201	0.391805	12.829991
75%	2.378689	0.453391	0.891904	16.975375
max	2.952480	0.817310	11.203100	23.703100



Gambar 17 Scatterplot NPHI, RHOB, dan CALI Filled Data

5.2.3 *Shear Sonic Wave Log Data*

Dalam proses pembuatan *Shear Sonic Wave Log Data* dilakukan penghapusan pada kurva DTS (*Shear Sonic Wave Log*), yang kemudian dilakukan pengisian pada kurva NPHI dan RHOB dimana hasil dapat dilihat pada tabel berikut:

Tabel 4 Shear Sonic Wave Log Data

	CALI	GR	RHOB	NPHI	RSHA	RMED	RDEP	DTC	DTS
count	4943.000000	4943.000000	4943.000000	4943.000000	4943.000000	4943.000000	4943.000000	4943.000000	4943.000000
mean	8.903065	72.399674	2.439862	0.213387	1.404263	1.402941	1.781580	82.379758	156.129034
std	1.169859	37.180637	0.105063	0.058902	0.793648	0.790973	0.928405	5.983913	12.024306
min	8.219579	8.581774	1.740164	0.070199	0.130193	0.128912	0.123068	60.782806	115.153511
25%	8.551326	43.060555	2.363918	0.168023	0.755697	0.756005	1.086194	77.576500	147.189636
50%	8.589574	57.176441	2.481702	0.224373	1.499650	1.499553	1.874264	84.279678	159.176392
75%	8.886503	116.962711	2.519511	0.249356	1.761457	1.760547	2.295663	86.749569	164.404663
max	17.008371	181.953278	2.660506	0.817310	6.805757	6.087253	5.759151	98.742012	190.500519

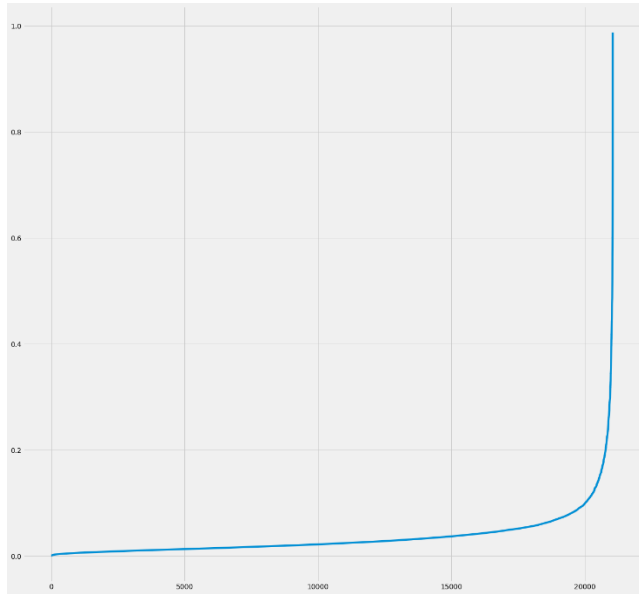
5.3 Pembuatan Model

5.3.1 DBSCAN

Proses pembuatan model DBSCAN pertama diperlukan untuk mencari nilai Epsilon yang optimal untuk DBSCAN. Untuk itu digunakan metode K-Nearest Neighbor mendapatkan nilai optimal yaitu 0,02. Dapat dilihat pada gambar 18 nilai optimal epsilon. Kemudian dilakukan *fine tuning* pada parameter minimum samples untuk mendapatkan *cluster* yang paling optimal.

5.3.2 OPTICS

Model OPTICS merupakan modifikasi dari model DBSCAN dimana diperlukan nilai optimal epsilon. Digunakan metode yang sama pada gambar 18 dimana nilai optimal epsilon yaitu dari 0,01 sampai dengan 0,03. Kemudian dilakukan *fine tuning* pada parameter minimum cluster size dan minimum samples untuk mendapatkan cluster yang paling optimal. Pada Gambar 18 memiliki y axis sebagai nilai epsilon dan x axis sebagai points (sample) sorted by distance.



Gambar 18 Grafik Nilai Optimal Epsilon

5.3.3 Boosted Random Forest

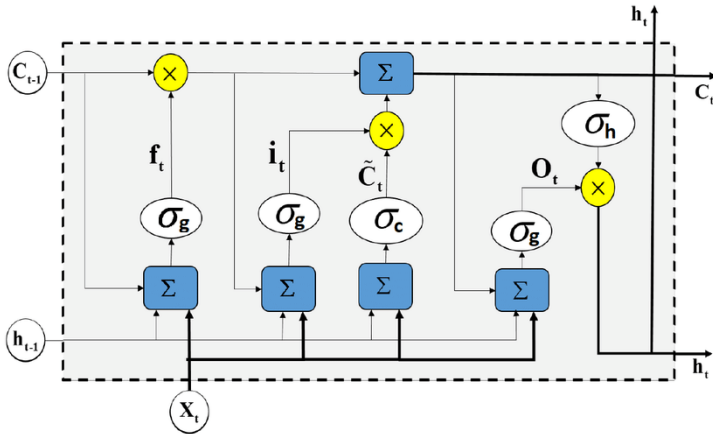
Proses pembuatan model Boosted Random Forest untuk mendapatkan parameter terbaik maka dilakukan *hyperparameter tuning* dengan parameter yang diuji merupakan *learning rate* dan *max depth* dari model. Parameter yang digunakan pada model sebagai berikut.

Tabel 5 Tabel Parameter Boosted Random Forest

No.	Parameter	Hasil
1	Learning Rate	0,01
2	Max Depth	6
3	Min Child Weight	1,5
4	N Estimators	10000

5.3.4 LSTM

Proses pembuatan model LSTM menggunakan arsitektur yang pada riset dengan menggunakan dataset dari 6 Well dengan hasil MSE pada beberapa Well, A1 0,49, A2 1,37, dan A3 0,68 (D Zhang, 2018). Dimana digunakan LSTM dengan 30 units, diikuti dengan *drop out* 0.3, dan ditambahkan Fully Connected layer. Arsitektur dari LSTM dapat dilihat sebagai berikut.



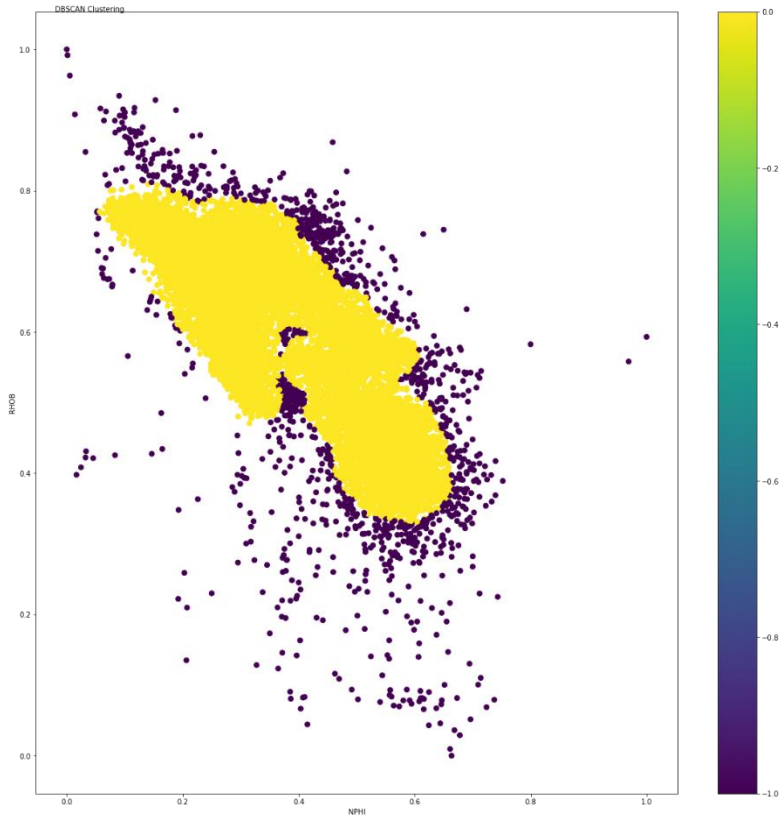
Gambar 19 Arsitektur LSTM

5.4 Evaluasi dan Visualisasi

Pada tahap ini dilakukan evaluasi terhadap data yang tidak digunakan saat training dan kemudian dilakukan visualisasi yang dibandingkan dengan kurva yang berkorelasi.

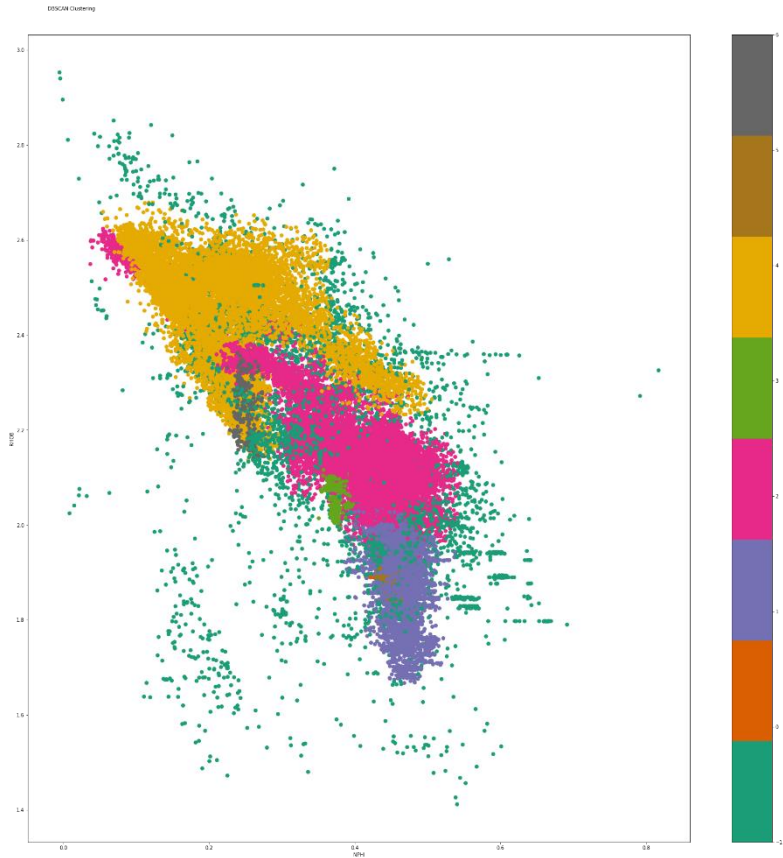
5.4.1 DBSCAN

Berikut gambar DBSCAN Dropped Data dan Filled Data Scatter plot.



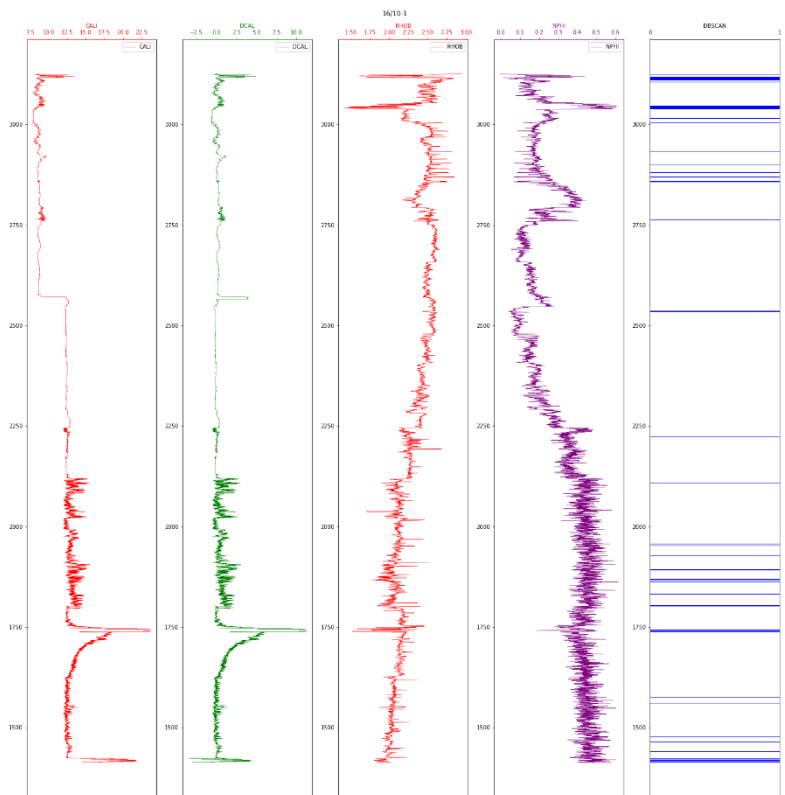
Gambar 20 DBSCAN Dropped Data Cluster Scatter Plot

Pada dropped data menghasilkan 1 cluster berwarna kuning pada gambar 20 dengan data berwarna ungu merupakan *outlier* dari data, atau di labelkan sebagai Badhole. Sedangkan hasil pada Filled Data menghasilkan 7 cluster dengan data berwarna hijau sebagai outlier.

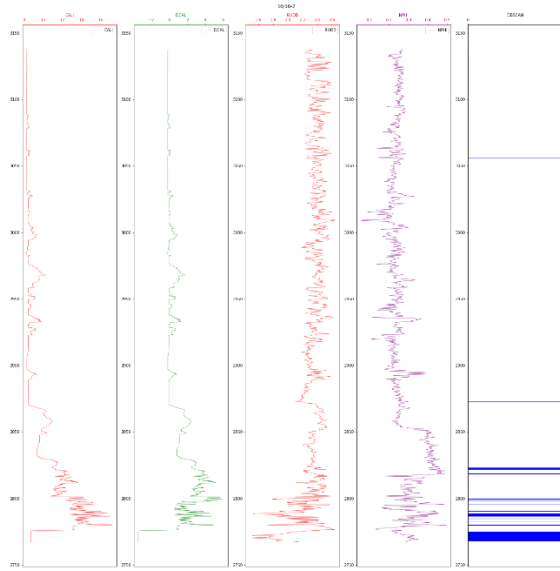


Gambar 21 DBSCAN Filled Data Cluster Scatter Plot

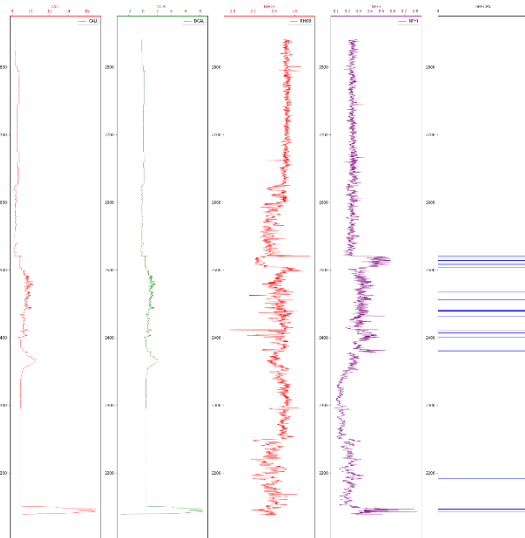
Berikut merupakan hasil visualisasi dari Cluster DBSCAN pada setiap Well Log.



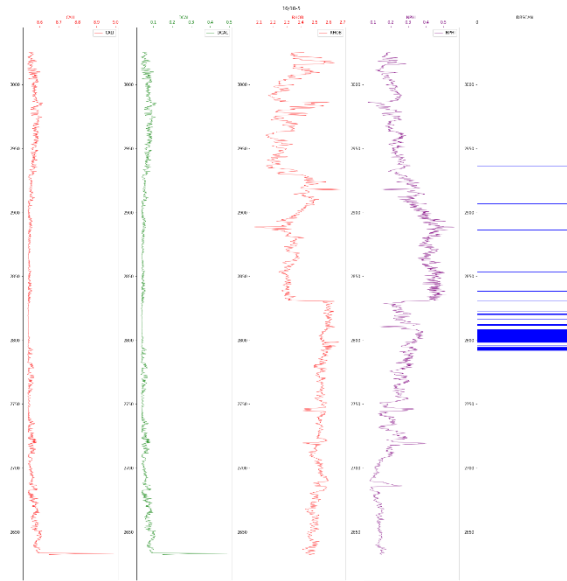
Gambar 22 Dropped Data DBSCAN Cluster Well 16/10-1



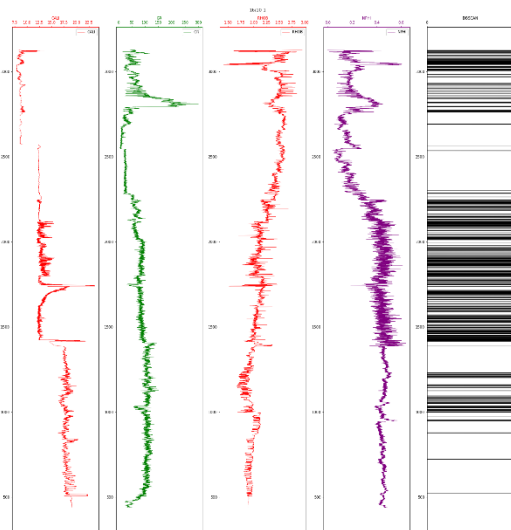
Gambar 23 Dropped Data DBSCAN Cluster Well 16/10-3



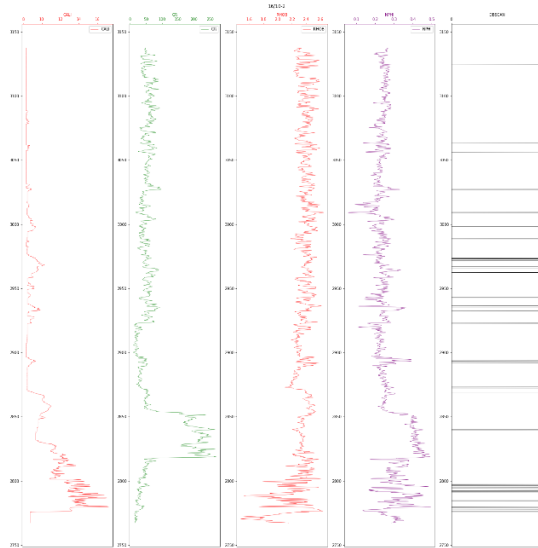
Gambar 24 Dropped Data DBSCAN Cluster Well 16/10-3



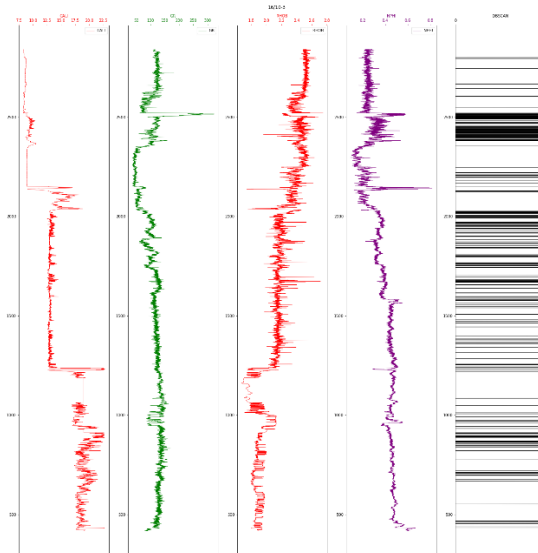
Gambar 25 Dropped Data DBSCAN Cluster Well 16/10-5



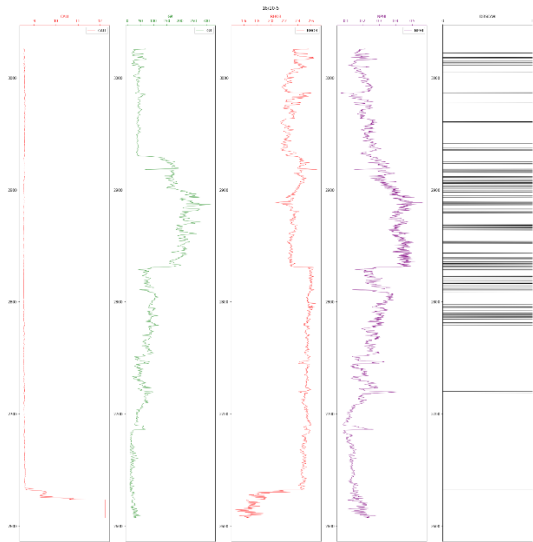
Gambar 26 Filled Data DBSCAN Cluster Well 16/10-1



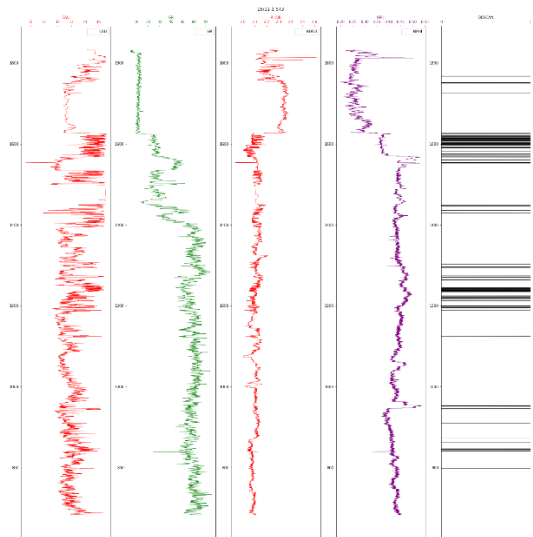
Gambar 27 Filled Data DBSCAN Cluster Well 16/10-2



Gambar 28 Filled Data DBSCAN Cluster Well 16/10-3



Gambar 29 Filled Data DBSCAN Cluster Well 16/10-5

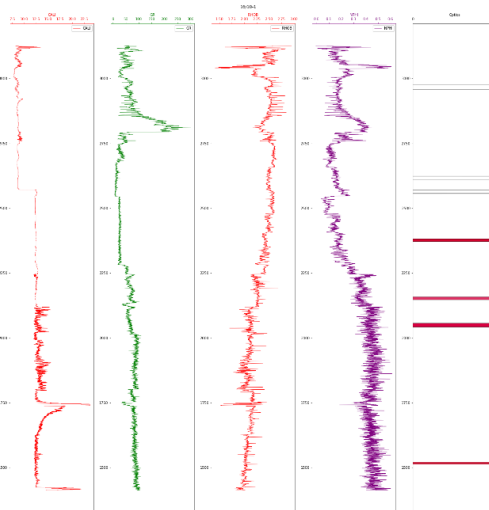


Gambar 30 Filled Data DBSCAN Cluster Well 16/11-1 ST3

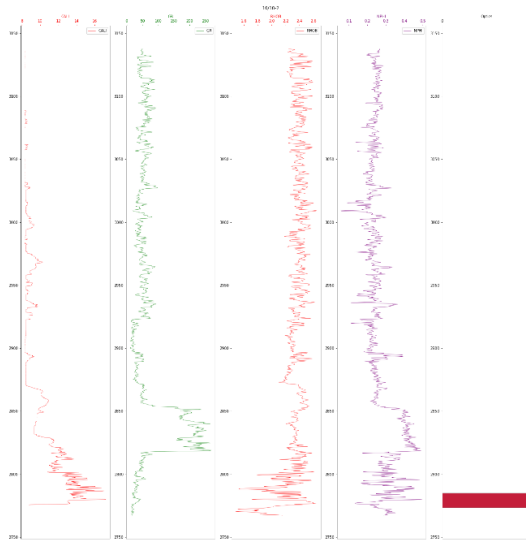
Badhole identification pada *Dropped Data* menggunakan DBSCAN dapat menemukan outlier, dimana setiap well kecuali Well 16/11-1 ST3 terdapat beberapa flag pada column dbscan. Pada Filled Data, DBSCAN dapat menemukan outlier pada setiap Well akan tetapi hasilnya kurang terpercaya untuk digunakan sebagai Badhole Identification karena data yang telah diidentifikasi di Well 16/10-1 pada gambar 26 dan Well 16/10-3 pada gambar 28 menghasilkan banyak flag jauh berbeda dibandingkan dengan *Dropped Data*.

5.4.2 OPTICS

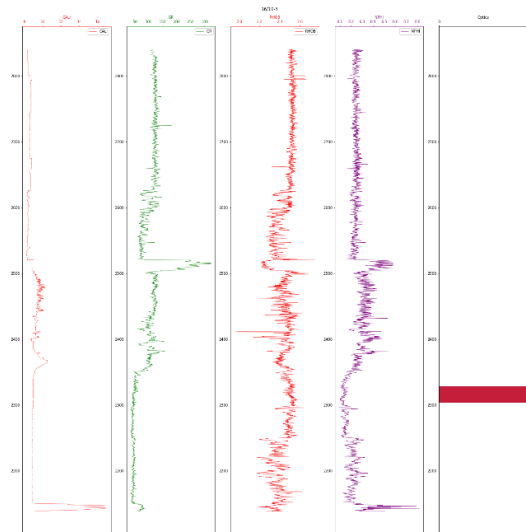
Berikut merupakan hasil visualisasi dari Clustering OPTICS pada setiap Well Log.



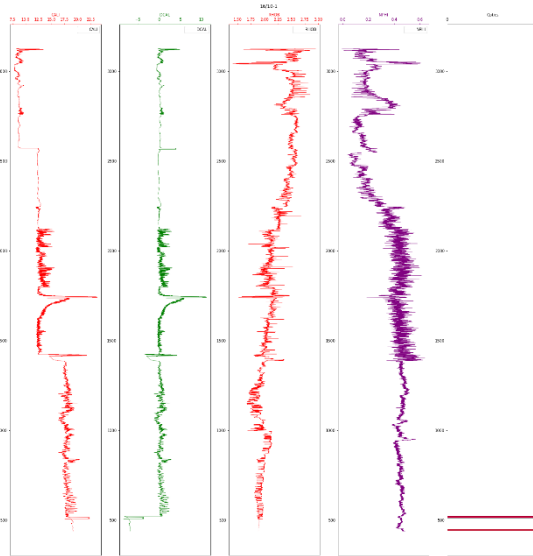
Gambar 31 Dropped Data OPTICS Cluster Well 16/10-1



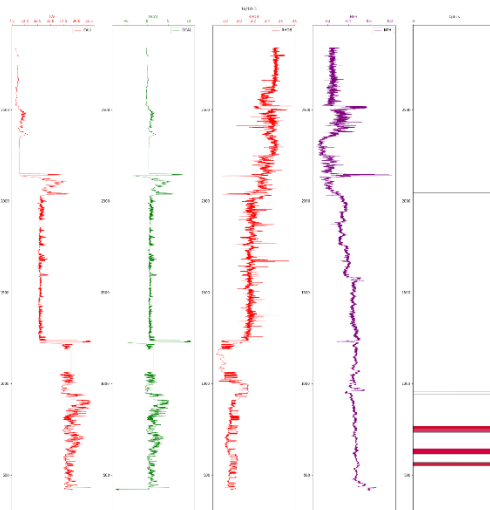
Gambar 32 Dropped Data OPTICS Cluster Well 16/10-2



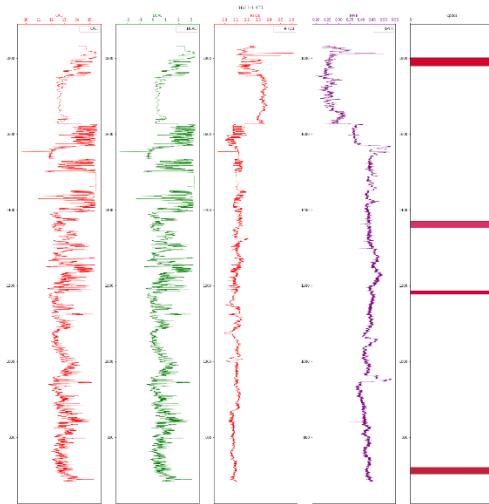
Gambar 33 Dropped Data OPTICS Cluster Well 16/10-3



Gambar 34 Filled Data OPTICS Cluster Well 16/10-1



Gambar 35 Filled Data OPTICS Cluster Well 16/10-3



Gambar 36 Filled Data OPTICS Cluster Well 16/11-1 ST3

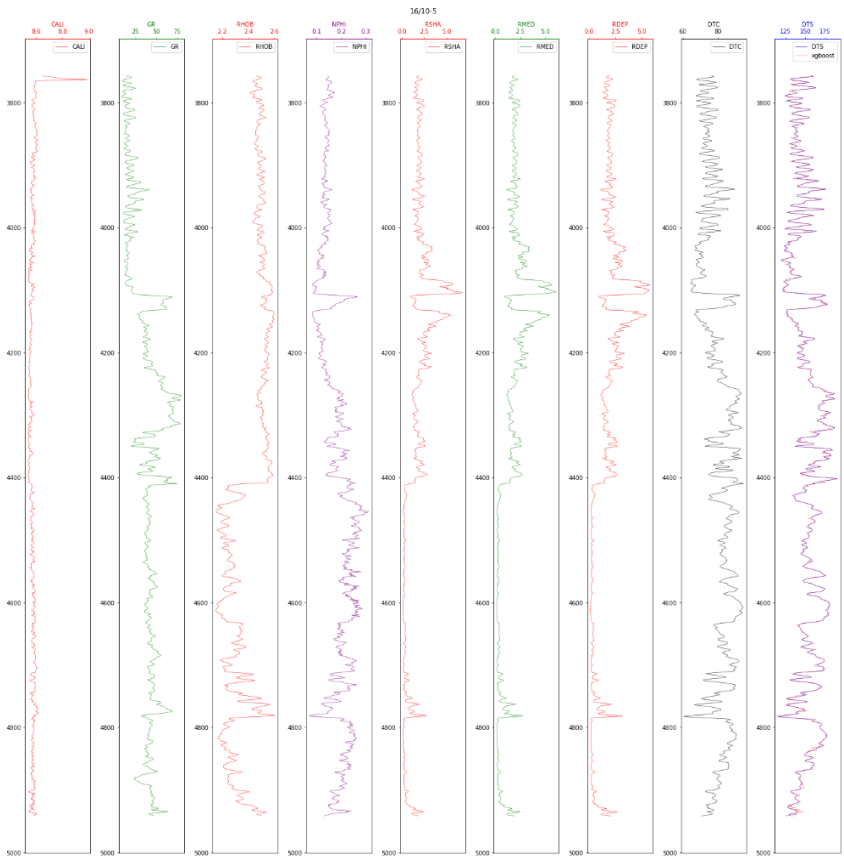
Pada Dropped Data Cluster berada pada Well 16/10-1, Well 16/10-2, dan Well 16/10-3 dimana pada Filled Data Cluster berada pada Well 16/10-1, Well 16/10-3, dan Well 16/11-1 ST3. Hasil pada kedua data dapat menemukan outlier pada range epsilon 0,01 sampai 0,03. Dibandingkan dengan hasil DBSCAN, OPTICS menghasilkan outlier lebih sedikit dibandingkan dengan DBSCAN.

5.4.3 Boosted Random Forest

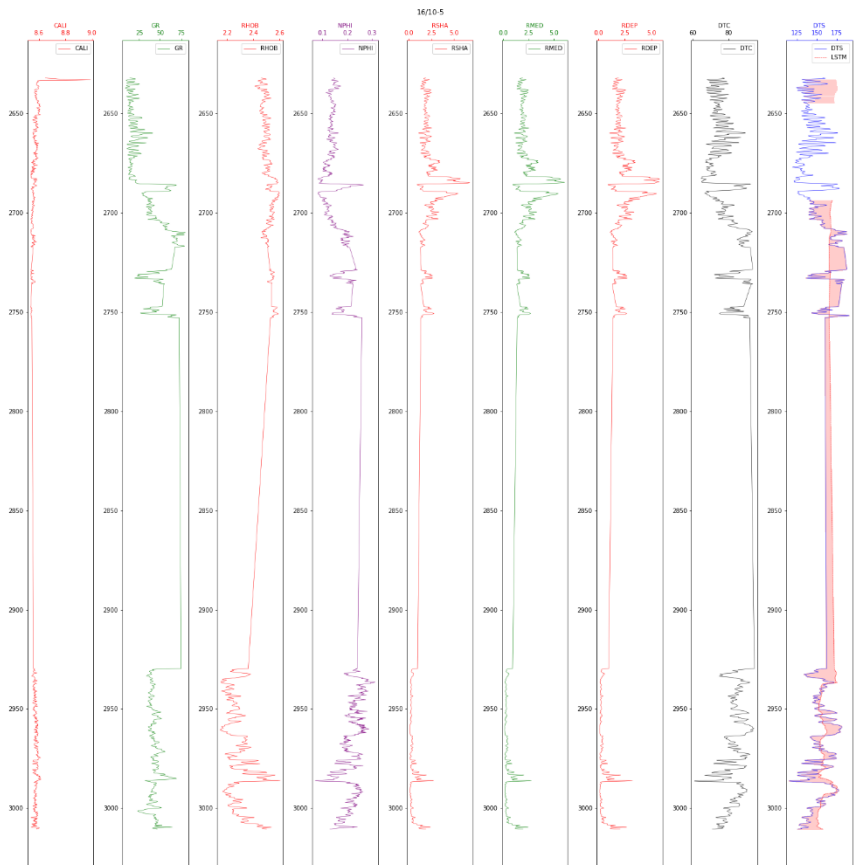
Hasil dari Boosted Random Forest dengan parameter `colsample_bytree` 0.4, `learning_rate` 0.01, `max_depth` 6, `min_child_weight` 1.5, `n_estimators` 10000, `reg_alpha` 0.75, `reg_lambda` 0.45, dan `subsample` 0.6, menghasilkan nilai RMSE 2.0911, MAE 1.4834, dan R^2 0.8843. Gambar 33 merupakan tampilan model Boosted Random Forest melakukan prediksi pada Well 16/10-5 dan kurva yang berkorelasi.

5.4.4 LSTM

Hasil dari LSTM dengan parameter 30-unit LSTM, 0.3 dropout, diikuti dengan fully connected layer menghasilkan nilai RMSE 6.3025, MAE 4.9138, dan R^2 0.1375. Gambar 34 merupakan tampilan model LSTM melakukan prediksi pada Well 16/10-5 dan kurva yang berkorelasi.



Gambar 37 Visualisasi Well 16/10-5 dengan Prediksi DTS dengan Algoritma Boosted Random Forest



Gambar 38 Visualisasi Well 16/10-5 dengan Prediksi DTS dengan Algoritma LSTM

BAB VI

KESIMPULAN DAN SARAN

6.1 Kesimpulan

Kesimpulan yang didapat setelah melakukan *clustering* untuk identifikasi Badhole dan melakukan prediksi Shear Sonic Wave log pada dataset conventional well log sebagai berikut:

1. Hasil algoritma DBSCAN lebih unggul dalam mengidentifikasi Badhole lebih banyak dan berarti dari pada algoritma OPTICS.
2. Hasil algoritma Boosted Random Forest lebih unggul dalam melakukan prediksi Shear Sonic Wave Log dan nilai RMSE, MAE, dan R^2 dari pada algoritma LSTM.

6.2 Saran

Beberapa saran atau tanggapan yang didapat sebagai hasil dari kerja praktek sebagai berikut:

1. Model yang digunakan untuk melakukan clustering masih dapat dioptimalkan pada algoritma OPTICS. Fine Tuning dapat dilakukan untuk mendapat performa yang lebih baik.
2. Model LSTM yang digunakan untuk melakukan prediksi Shear Sonic Wave Log dapat menghasilkan nilai lebih baik jika dataset yang digunakan untuk melatih model tersebut lebih banyak.

DAFTAR PUSTAKA

- Ameen, M. S., Smart, B. G. D., Somerville, J. M., Hammilton, S., & Naji, N. A. (2009). Predicting rock mechanical properties of carbonates from wireline logs (A case study: Arab-D reservoir, Ghawar field, Saudi Arabia). *Marine and Petroleum Geology*, 26(4).
<https://doi.org/10.1016/j.marpetgeo.2009.01.017>
- Ankerst, M., Breunig, M. M., Kriegel, H. P., & Sander, J. (1999). OPTICS: Ordering Points to Identify the Clustering Structure. *SIGMOD Record (ACM Special Interest Group on Management of Data)*, 28(2).
<https://doi.org/10.1145/304181.304187>
- Balossino, P., Pampuri, F., Bruni, C., & Ebzhasarova, K. (2008). An integrated approach to obtain reliable permeability profiles from logs in a carbonate reservoir. *SPE Reservoir Evaluation and Engineering*, 11(4).
<https://doi.org/10.2118/102289-pa>
- Banas R., M. A. (2021). NOVEL METHODOLOGY FOR AUTOMATION OF BAD WELL LOG DATA IDENTIFICATION AND REPAIR. Society of Petrophysicists and Well Log Analysts (SPWLA).
<https://dx.doi.org/10.30632/SPWLA-2021-0070>
- Brie, A., Endo, T., Johnson, D. L., & Pampuri, F. (2000). Quantitative formation permeability evaluation from Stoneley waves. *SPE Reservoir Evaluation and Engineering*, 3(2). <https://doi.org/10.2118/60905-PA>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data*

Mining,

13-17-August-2016.

<https://doi.org/10.1145/2939672.2939785>

- Cobb, G. W., Anderson, D. R., Sweeney, D. J., Williams, T. A., Brockett, P., Levine, A., Clarke, G. M., Cooke, D., Devore, J., Peck, R., Freund, J. E., Smith, R. M., Goldman, R. N., Weinberg, J. S., Johnson, R., Bhattacharyya, G., McGhee, J. W., Ott, L., Mendenhall, W., ... Sincich, T. (1987). Introductory Textbooks: A Framework for Evaluation. *Journal of the American Statistical Association*, 82(397). <https://doi.org/10.2307/2289170>
- Eberli, G. P., Baechle, G. T., Anselmetti, F. S., & Incze, M. L. (2003). Factors controlling elastic properties in carbonate sediments and rocks. *Leading Edge (Tulsa, OK)*, 22(7). <https://doi.org/10.1190/1.1599691>
- Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*.
- GREENBERG, M. L., & CASTAGNA, J. P. (1992). SHEAR-WAVE VELOCITY ESTIMATION IN POROUS ROCKS: THEORETICAL FORMULATION, PRELIMINARY VERIFICATION AND APPLICATIONS. *Geophysical Prospecting*, 40(2). <https://doi.org/10.1111/j.1365-2478.1992.tb00371.x>
- Kuster, G. T., & Toksöz, M. N. (1974). VELOCITY AND ATTENUATION OF SEISMIC WAVES IN TWO-PHASE MEDIA: PART I. THEORETICAL FORMULATIONS. *GEOPHYSICS*, 39(5). <https://doi.org/10.1190/1.1440450>
- McDonald, A., (2021a). When working on petrophysical project a large portion of our time can be taken up getting the data ready for interpretation. URL:

- https://www.linkedin.com/posts/andymcdonaldgeo_data-science-activity6765699282860306433-xI6W Posted: [LinkedIn] 2021- 2-14.
- Pawlowsky-Glahn, V., & Buccianti, A. (2011). Compositional Data Analysis: Theory and Applications. In *Compositional Data Analysis: Theory and Applications*. <https://doi.org/10.1002/9781119976462>
- Pickett, G. R. (1963). Acoustic Character Logs and Their Applications in Formation Evaluation. *Journal of Petroleum Technology*, 15(06). <https://doi.org/10.2118/452-pa>
- Sheng, T., Wu, W., Sun, H., Guo, Q., & Zhang, B. (2018). Abductive identification of bad data: methodology and field test. *IET Generation, Transmission and Distribution*, 12(1). <https://doi.org/10.1049/IET-GTD.2017.0369>
- Tahiru, I. T., Olagundoye, O., & Alabere, A. O. (2022). *Machine Learning for Sonic Logs Prediction: A Case Study from the Niger Delta Basin in the Gulf of Guinea*. <https://doi.org/10.2523/iptc-21932-ms>
- Tang, X. M., & Patterson, D. (2001). Detecting thin gas beds in formations using stoneley wave reflection and high-resolution slowness measurements. *SPWLA 42nd Annual Logging Symposium 2001*.
- Tang, X. M., & Patterson, D. (2018). Estimating formation permeability and anistoropy from borehole stoneley waves. *SPWLA 45th Annual Logging Symposium 2004*.
- Wendt, A. S., Kongslie, M., Sinha, B. K., Vissapragada, B., Newton, A., Skomedal, E., Renlie, L., & Pedersen, E. S. (2007). Enhanced mechanical earth modelling and wellbore stability calculations using advanced sonic measurements - A case study of the HP/HT Kvitebjorn field in the Norwegian

North Sea. *Proceedings - SPE Annual Technical Conference and Exhibition*, 2. <https://doi.org/10.2118/109662-ms>
Zhang, Z., Zhu, Q., Zhu, F., Li, J., Cheng, D., Liu, Y., & Luo, J. (2021). Density decay graph-based density peak clustering. *Knowledge-Based Systems*, 224. <https://doi.org/10.1016/j.knosys.2021.107075>

BIODATA PENULIS I

Nama : Satrio Hanif Wicaksono
Tempat, Tanggal Lahir : Balikpapan, 9 Mei 2001
Jenis Kelamin : Laki-laki
Telepon : +62895393020322
Email : asharih73@gmail.com

AKADEMIS

Kuliah : Departemen Teknik Informatika –
FTEIC , ITS
Angkatan : 2019
Semester : 8 (Delapan)