# Efikasi Deep Resistivity dan SMOTE dalam Prediksi Hidrokarbon Berbasis Pembelajaran Mesin

## Landasan Petrofisika dalam Prediksi Hidrokarbon Berbasis Log Sumur

### 1.1 Pembelajaran Mesin dalam Karakterisasi Bawah Permukaan

Industri minyak dan gas telah memasuki era di mana metodologi berbasis data memperkuat, dan dalam beberapa kasus menggantikan, teknik analisis tradisional. Penerapan pembelajaran mesin (ML) tersupervisi untuk memprediksi parameter petrofisika utama dari data log sumur merupakan sebuah pergeseran paradigma yang signifikan.¹ Selama beberapa dekade, para ahli geosains telah mengandalkan model empiris atau berbasis fisika untuk mengestimasi properti seperti waktu tempuh sonik (DT), porositas, atau permeabilitas. Namun, model-model ini sering kali memiliki keterbatasan dan asumsi inheren yang mungkin tidak berlaku di berbagai lingkungan geologi.² Pembelajaran mesin menawarkan alternatif yang kuat, mampu mengenali hubungan non-linear yang kompleks dalam set data berdimensi tinggi yang sering kali tidak dapat diungkap oleh analisis konvensional.³

Adopsi ML secara luas dalam pengambilan keputusan eksplorasi dan pengembangan berisiko tinggi sering kali terhambat oleh masalah "kotak hitam" (*black box*).² Banyak algoritma canggih beroperasi dengan cara yang tidak mudah diinterpretasikan, sehingga menyulitkan para ahli geosains untuk memvalidasi kelayakan fisis dari prediksi model. Kurangnya transparansi ini dapat mengikis kepercayaan dan membatasi nilai yang dirasakan dari wawasan yang didorong oleh ML dalam industri.² Studi ini berupaya mengatasi tantangan tersebut dengan melampaui evaluasi kinerja model yang murni bersifat matematis. Dengan mengintegrasikan prinsip-prinsip fisika batuan dan merancang eksperimen yang menguji validitas geologis dari fitur masukan secara cermat, analisis ini bertujuan untuk membongkar opasitas tradisional dari model ML, serta menumbuhkan pemahaman yang lebih transparan dan berlandaskan geologi mengenai perilaku prediktifnya.²

### 1.2 Logging Resistivitas: Landasan Deteksi Hidrokarbon

Di antara serangkaian pengukuran yang ditangkap oleh peralatan *logging* sumur, resistivitas formasi merupakan landasan utama dalam deteksi hidrokarbon.⁶ Prinsip fundamental dari *logging* resistivitas adalah pengukuran kemampuan formasi untuk menghambat aliran arus listrik, sebuah properti yang diukur dalam satuan ohm-meter (ohm⋅m).⁸ Kekuatan diagnostik dari pengukuran ini berasal dari sifat listrik yang kontras antara material-material yang menyusun reservoir bawah permukaan. Dengan sedikit pengecualian, komponen mineral padat dari batuan sedimen, seperti kuarsa dan kalsit, merupakan isolator listrik yang efektif.⁶ Akibatnya, arus listrik hanya dapat melewati formasi melalui fluida konduktif yang terkandung di dalam ruang pori-porinya.

Dalam sebagian besar konteks geologi, fluida konduktif ini adalah air asin, atau *brine*, yang ion-ion terlarutnya berfungsi sebagai pembawa muatan.⁶ Sebaliknya, hidrokarbon minyak dan gas alam terdiri dari molekul-molekul dengan ikatan kovalen yang tidak terdisosiasi menjadi ion, sehingga sangat resistif secara elektrik.⁶ Fundamental inilah yang menjadi dasar penggunaan log resistivitas sebagai indikator utama hidrokarbon. Formasi batuan berpori yang jenuh dengan *brine* konduktif akan menunjukkan resistivitas *bulk* yang rendah. Sebaliknya, ketika ruang pori yang sama sebagian atau seluruhnya diisi oleh hidrokarbon yang resistif, resistivitas *bulk* formasi meningkat secara signifikan.⁶ Oleh karena itu, pembacaan resistivitas yang tinggi secara anomali dalam suatu interval berpori merupakan indikator klasik dan kuat adanya potensi zona yang mengandung hidrokarbon, atau "zona produktif" (*pay zone*).¹⁰

Namun, tantangan signifikan dalam eksplorasi hidrokarbon adalah keberadaan zona produktif "*low resistivity, low contrast*" (LRLC).¹³ Ini adalah reservoar mengandung hidrokarbon yang layak secara komersial, namun berlawanan dengan ekspektasi konvensional, menunjukkan resistivitas listrik yang rendah, sering kali dalam rentang 0,5 hingga 5 ohm-meter.¹⁴ Dengan menggunakan teknik interpretasi standar, zona-zona ini sering kali disalah artikan sebagai interval serpih yang jenuh air atau non-produktif, sehingga dilewatkan dan menyisakan cadangan yang signifikan tidak termanfaatkan.¹³ Penyebab geologis dari fenomena ini beragam dan kompleks, termasuk keberadaan laminasi tipis pasir dan serpih yang berada di bawah resolusi vertikal alat *logging*, saturasi air tak tereduksi yang tinggi karena ukuran butir yang sangat halus, dan adanya mineral konduktif seperti pirit, glaukonit, atau berbagai jenis lempung yang menciptakan jalur listrik alternatif melalui batuan.¹³ Kesulitan dalam mengevaluasi zona LRLC ini menggarisbawahi keterbatasan jika hanya mengandalkan pengukuran resistivitas untuk deteksi hidrokarbon.

### 1.3 Zona Tak Terinvasi dan Konsep Ideal "Resistivitas Sejati" (Rt)

Interpretasi log resistivitas menjadi rumit karena sifat invasif dari proses pengeboran. Saat sumur dibor, perbedaan tekanan antara lumpur bor di dalam lubang bor dan fluida formasi menyebabkan fasa cair dari lumpur, yang dikenal sebagai filtrat lumpur, merembes ke dalam lapisan batuan permeabel.¹² Proses ini menggantikan fluida formasi asli *in-situ* dan menciptakan beberapa zona investigasi konsentris yang berbeda di sekitar lubang bor, masing-masing dengan saturasi fluida yang berbeda dan, akibatnya, resistivitas listrik yang berbeda pula.¹²

Zona-zona ini biasanya dikategorikan sebagai berikut:

* **Zona Siram (*Flushed Zone*):** Zona ini berada tepat di sebelah dinding lubang bor dan hanya meluas beberapa inci ke dalam formasi. Di sini, fluida formasi asli hampir seluruhnya telah digantikan oleh filtrat lumpur. Resistivitas zona ini diukur oleh alat *logging* dengan pembacaan dangkal.¹²
* **Zona Terinvasi (*Invaded Zone* atau *Annulus*):** Terletak di luar zona siram, ini adalah wilayah transisi yang berisi campuran filtrat lumpur dan fluida formasi asli. Resistivitasnya diukur oleh peralatan dengan kedalaman investigasi medium.⁷
* **Zona Tak Terinvasi (*Uninvaded Zone*):** Ini adalah wilayah formasi yang terjauh dari lubang bor yang tetap tidak terkontaminasi oleh proses pengeboran. Zona ini berisi fluida formasi asli yang murni. Resistivitas zona ini adalah target utama dari evaluasi formasi dan disebut sebagai "resistivitas sejati," atau Rt.⁷

Pengukuran Rt sangatlah penting karena merupakan parameter masukan kritis untuk analisis petrofisika kuantitatif. Dengan menggunakan hubungan yang sudah mapan seperti Hukum Archie, analis log menggunakan Rt bersama dengan porositas dan resistivitas air formasi (Rw) untuk menghitung saturasi air (Sw), yang pada gilirannya memungkinkan estimasi saturasi hidrokarbon.¹⁰ Penentuan Rt yang akurat, yang biasanya diupayakan dengan alat induksi atau *laterolog* dengan pembacaan dalam, oleh karena itu esensial untuk perhitungan cadangan yang andal dan penilaian ekonomi suatu reservoir.¹⁰

### 1.4 Faktor Perancu: Hipotesis Kontaminasi Fitur

Konsep "resistivitas sejati" (Rt) merupakan suatu ideal teoretis, tetapi pengukuran praktisnya penuh dengan tantangan. Proses invasi filtrat lumpur, meskipun menciptakan zona-zona berbeda yang memungkinkan penilaian permeabilitas kualitatif, juga menjadi ancaman terbesar untuk memperoleh nilai Rt yang akurat.²² Bahkan alat resistivitas dengan pembacaan terdalam, yang dirancang untuk menembus di luar zona terinvasi, dapat terpengaruh pengukurannya oleh filtrat lumpur, terutama pada formasi yang sangat permeabel di mana invasi bisa sangat dalam, atau dalam situasi di mana sebagian lubang bor terpapar lumpur bor untuk periode waktu yang lama.⁷ Ini berarti bahwa nilai "resistivitas dalam" yang tercatat sering kali hanya merupakan aproksimasi dari resistivitas formasi yang sebenarnya dan tidak terkontaminasi. Bagi model pembelajaran mesin, yang tidak memiliki konteks fisis untuk memperhitungkan diskrepansi ini, aproksimasi ini diperlakukan sebagai kebenaran mutlak. Oleh karena itu, model dilatih bukan pada sinyal geologis murni dari resistivitas sejati, melainkan pada pengukuran resistivitas dalam yang berpotensi terganggu.

Diskrepansi ini memperkenalkan kemungkinan adanya **kontaminasi fitur**, di mana log resistivitas dalam membawa informasi laten yang berkaitan dengan operasi pengeboran dan alterasi di dekat lubang bor, bukan semata-mata mencerminkan properti intrinsik formasi. Akurasi pengukuran resistivitas diketahui sensitif terhadap sejumlah faktor lingkungan yang dapat memasukkan noise (*noise*) dan bias sistematis.¹¹ Faktor-faktor perancu ini meliputi:

* **Properti Lumpur Bor:** Resistivitas dari filtrat lumpur itu sendiri, yang merupakan fungsi dari salinitas dan fluida dasarnya (berbasis air vs. berbasis minyak), secara langsung mempengaruhi resistivitas zona terinvasi dan kontras yang diamati oleh alat *logging*.⁶
* **Kondisi Lubang Bor:** Ketidakteraturan pada dinding lubang bor, yang dikenal sebagai rugositas, dapat mendistorsi medan listrik yang digunakan untuk pengukuran, yang mengarah pada pembacaan yang tidak akurat.¹¹
* **Suhu Formasi:** Resistivitas fluida berair sangat bergantung pada suhu. Seiring meningkatnya suhu dengan kedalaman, pengukuran resistivitas harus dikoreksi ke suhu referensi standar agar dapat dibandingkan.²¹
* **Anisotropi:** Pada formasi berlaminasi tipis atau rekah, resistivitas listrik dapat bervariasi tergantung pada apakah arus mengalir sejajar atau tegak lurus terhadap bidang perlapisan. Anisotropi ini dapat memperumit interpretasi pengukuran log standar.¹¹

Hal ini mengarah pada hipotesis utama investigasi ini: bahwa faktor-faktor perancu ini dapat memasukkan noise non-geologis yang signifikan ke dalam pengukuran resistivitas dalam. Proses ini dapat dipahami sebagai rantai sebab-akibat langsung. Pertama, operasi pengeboran menyebabkan invasi filtrat lumpur. Kedua, invasi ini mengubah saturasi fluida dan profil resistivitas di dekat lubang bor. Ketiga, alterasi ini mengontaminasi pengukuran resistivitas dalam, menyebabkannya menyimpang dari resistivitas formasi yang sebenarnya. Keempat, dari perspektif ilmu data, hal ini memasukkan noise ke dalam set fitur. Kelima, algoritma pembelajaran mesin yang sensitif mungkin disesatkan oleh noise ini, mempelajari korelasi palsu antara sinyal resistivitas yang terkontaminasi dan label target hidrokarbon. Konsekuensi akhir dan paling kritis adalah bahwa korelasi yang dipelajari ini tidak akan dapat digeneralisasi, yang menyebabkan kinerja prediktif yang buruk ketika model diterapkan pada sumur uji buta (*blind well*) yang dibor dalam kondisi berbeda. Model tersebut mungkin secara tidak sengaja menjadi "pendeteksi kejadian pengeboran" daripada "pencari hidrokarbon" yang andal. Masalah ini diperburuk dalam konteks zona produktif resistivitas rendah, di mana sinyal hidrokarbon yang sudah samar dengan mudah dikalahkan oleh noise lingkungan semacam itu, membuat reservoir ini sangat sulit untuk diidentifikasi dan dievaluasi.²⁴

## Metodologi Eksperimen

### 2.1 Alur Kerja Pembelajaran Mesin untuk Prediksi Hidrokarbon

Tujuan utama dari alur kerja eksperimental adalah untuk membangun model klasifikasi biner yang kokoh dan mampu memprediksi ada atau tidaknya hidrokarbon pada interval kedalaman tertentu dengan menggunakan serangkaian log sumur standar. Metodologi ini dirancang untuk menilai secara ketat kemampuan model dalam menggeneralisasi pembelajarannya ke data baru yang tidak terlihat, yang merupakan ujian utama dari utilitas praktisnya dalam konteks eksplorasi.

Studi ini menggunakan data dari dua sumur yang berbeda: LLB-10 berfungsi sebagai set data latih, di mana model pembelajaran mesin dibangun dan dioptimalkan. Sumur kedua, LLD-14, ditetapkan sebagai "sumur uji buta" (*blind well*).¹ Ini adalah komponen kritis dari desain eksperimental. Data sumur uji buta sepenuhnya disembunyikan selama fase pelatihan model dan penalaan hiperparameter. Data ini hanya digunakan untuk tahap validasi akhir, memberikan penilaian yang tidak bias tentang bagaimana model berkinerja pada data yang belum pernah ditemuinya. Protokol ini merupakan praktik terbaik dalam ML geosains, karena mensimulasikan skenario dunia nyata dalam menerapkan model untuk memprediksi hasil di sumur eksplorasi baru.¹

Proses rekayasa fitur melibatkan pemilihan serangkaian variabel prediktor (fitur masukan) dan variabel target (keluaran). Fitur masukan dasar, yang digunakan di semua variasi eksperimental, terdiri dari serangkaian log petrofisika standar yang dikenal informatif untuk karakterisasi litologi dan fluida: sinar gamma natural (GR), densitas *bulk* (RHOB), dan porositas neutron (NPHI).¹ Variabel eksperimental adalah log resistivitas dalam (RT), yang disertakan atau dikecualikan tergantung pada eksperimen spesifik. Variabel target adalah label biner yang ditetapkan untuk setiap interval kedalaman, yang menunjukkan apakah interval tersebut mengandung hidrokarbon (kelas minoritas) atau tidak mengandung hidrokarbon (kelas mayoritas). Sebelum pelatihan model, seluruh set data fitur distandarisasi. Langkah pra-pemrosesan ini mengubah setiap fitur agar memiliki rata-rata nol dan varians satu, memastikan bahwa fitur dengan skala dan unit asli yang berbeda (misalnya, unit API untuk GR vs. g/cm³ untuk RHOB) tidak secara tidak proporsional mempengaruhi algoritma pembelajaran.¹

### 2.2 Algoritme SMOTE: Mengatasi Ketidakseimbangan Kelas

Tantangan umum dan signifikan dalam menerapkan pembelajaran mesin pada eksplorasi hidrokarbon adalah ketidakseimbangan kelas inheren dari data. Di setiap sumur atau lapangan, volume batuan non-produktif, mengandung air, atau ketat (kelas mayoritas) jauh melebihi volume batuan reservoir produktif yang mengandung hidrokarbon (kelas minoritas).²⁶ Ketidakseimbangan ini dapat sangat membiaskan algoritma klasifikasi standar. Model tersebut mungkin mencapai skor akurasi yang tampak tinggi hanya dengan mempelajari untuk selalu memprediksi kelas mayoritas, namun sepenuhnya gagal dalam mengidentifikasi interval mengandung hidrokarbon.²⁷

*Synthetic Minority Over-sampling Technique* (SMOTE) adalah algoritma augmentasi data yang banyak digunakan yang dirancang untuk mengatasi masalah ini.²⁷ Berbeda dengan *oversampling* sederhana, yang hanya melibatkan duplikasi sampel kelas minoritas yang ada dan dapat menyebabkan *overfitting* model, SMOTE menciptakan sampel sintetis baru di dalam ruang fitur.²⁶ Algoritme ini beroperasi sebagai berikut:

1. Sebuah sampel dipilih secara acak dari kelas minoritas (misalnya, titik data yang mengandung hidrokarbon).
2. Algoritma mengidentifikasi *k*-tetangga terdekatnya, yaitu *k* sampel yang paling mirip, juga dari kelas minoritas. Kemiripan biasanya diukur menggunakan jarak Euclidean dalam ruang fitur multi-dimensi.²⁶
3. Salah satu dari tetangga ini dipilih secara acak.
4. Sampel sintetis baru dihasilkan pada titik acak di sepanjang segmen garis yang menghubungkan sampel asli dan tetangga yang dipilih.²⁶

Dengan mengisi ruang fitur dengan contoh-contoh yang masuk akal dan ter interpolasi ini, SMOTE memberikan pengklasifikasi set sampel kelas minoritas yang lebih besar dan lebih beragam untuk dipelajari, memaksanya untuk mengembangkan batas keputusan yang lebih kuat antar kelas.²⁸

Namun, SMOTE bukannya tanpa keterbatasan, kerentanan utamanya adalah sensitivitasnya terhadap noise. Jika sampel kelas minoritas asli mengandung noise atau tidak terpisah dengan baik dari kelas mayoritas, SMOTE dapat menghasilkan sampel sintetis di wilayah ruang fitur yang ambigu atau tumpang tindih, yang berpotensi meningkatkan kesulitan klasifikasi daripada menguranginya.²⁸ Selain itu, efektivitasnya dapat berkurang pada set data berdimensi tinggi. Studi telah menunjukkan bahwa untuk banyak jenis pengklasifikasi, dampak SMOTE dapat diabaikan atau bahkan merugikan dalam pengaturan berdimensi tinggi kecuali jika didahului oleh langkah pemilihan fitur yang kuat.²⁹ Algoritma ini juga memperkenalkan korelasi yang kuat antara sampel sintetis dan sampel "induk" mereka dan dapat mengurangi varians keseluruhan dari kelas minoritas, yang mungkin berdampak negatif pada pengklasifikasi yang sensitif terhadap distribusi varians spesifik kelas.²⁹

Asumsi inti dari SMOTE adalah bahwa ruang antara dua sampel kelas minoritas adalah wilayah yang valid untuk menciptakan sampel kelas minoritas baru. Asumsi ini mungkin dilanggar ketika berhadapan dengan fitur yang terkontaminasi secara fisis. Jika dua titik data yang mengandung hidrokarbon, A dan B, memiliki nilai resistivitas dalam yang diubah secara artifisial oleh tingkat invasi filtrat lumpur yang berbeda, segmen garis yang menghubungkan mereka dalam ruang fitur ditentukan oleh nilai-nilai yang rusak ini. Sampel sintetis yang dihasilkan pada garis ini akan memiliki nilai resistivitas yang merupakan interpolasi dari dua tingkat kontaminasi yang berbeda sebuah jejak yang tidak sesuai dengan realitas fisis apa pun. Dalam skenario ini, SMOTE tidak hanya akan menyeimbangkan set data; ia akan secara aktif mencemari data latih dengan contoh-contoh yang secara fisis tidak masuk akal, mengajarkan model bahwa jejak ini adalah representasi valid dari hidrokarbon. Hal ini dapat mengarah pada model yang tidak akurat dan tidak dapat diinterpretasikan secara geologis.

### 2.3 Desain Eksperimen: Empat Skenario untuk Perbandingan

Untuk menyelidiki secara sistematis efek individual dan gabungan dari fitur resistivitas dalam dan algoritma SMOTE, studi ini disusun dalam empat cabang eksperimental yang berbeda. Desain ini memungkinkan perbandingan terkontrol untuk mengisolasi dampak setiap variabel terhadap kinerja prediktif model pada sumur uji buta.

Keempat skenario tersebut adalah:

* **Baseline (dengan DR, tanpa SMOTE):** Eksperimen ini mengevaluasi kinerja model menggunakan DRHO, GR, RHOB, NPHI, SR, MR, dan DR pada set data asli yang tidak seimbang.
* **Tanpa DR (tanpa DR, tanpa SMOTE):** Eksperimen ini menguji dampak penghapusan fitur resistivitas dalam. Model dilatih pada set data tidak seimbang menggunakan DRHO, GR, RHOB, NPHI, SR, dan MR.
* **SMOTE (dengan DR, dengan SMOTE):** Eksperimen ini menilai efek gabungan dari penggunaan fitur resistivitas dalam pada set data yang diseimbangkan dengan algoritme SMOTE.
* **SMOTE & Tanpa DR (tanpa DR, dengan SMOTE):** Eksperimen ini mengevaluasi kinerja SMOTE pada set fitur yang lebih bersih yang tidak menyertakan resistivitas dalam.

### 2.4 Algoritma Pembelajaran Mesin yang Dievaluasi

Untuk memastikan evaluasi yang komprehensif dan kuat, keempat skenario eksperimental diuji menggunakan beberapa algoritma pembelajaran mesin. Pendekatan multi-model ini membantu memastikan bahwa tren yang diamati bukanlah artefak dari satu pengklasifikasi tunggal, melainkan konsekuensi fundamental dari pemilihan fitur dan pilihan augmentasi data. Algoritma berikut dipilih untuk mewakili berbagai strategi klasifikasi:

* **Model Berbasis Pohon Ensemble:** Model ini menggabungkan beberapa pohon keputusan untuk menghasilkan prediksi yang lebih kuat. Model ini banyak digunakan dan telah menunjukkan kinerja yang kuat dalam aplikasi geosains.
  + **Random Forest (RF):** Metode ensemble yang beroperasi dengan membangun banyak pohon keputusan pada saat pelatihan.³
  + **XGBoost:** Implementasi *gradient boosting* yang sangat efisien dan dapat diskalakan yang membangun model secara sekuensial dan bertahap.
  + **LightGBM:** Kerangka kerja *gradient boosting* yang menggunakan algoritma pembelajaran berbasis pohon, dikenal karena kecepatan dan efisiensinya.
* **Support Vector Machine (SVM):** Pengklasifikasi non-probabilistik yang bekerja dengan menemukan *hyperplane* yang paling baik memisahkan titik data dari kelas yang berbeda dalam ruang berdimensi tinggi.³
* ***k-Nearest Neighbors* (k-NN):** Algoritme pembelajaran berbasis instans sederhana yang mengklasifikasikan titik data berdasarkan kelas mayoritas dari 'k' tetangga terdekatnya di ruang fitur.

## Hasil Empiris dari Uji Sumur Buta

### 3.1 Metrik Kinerja untuk Klasifikasi

Mengevaluasi model klasifikasi, terutama dalam konteks set data geologis yang tidak seimbang, memerlukan pendekatan yang lebih bernuansa daripada hanya mengandalkan satu metrik seperti akurasi. "Paradoks akurasi" menggambarkan situasi di mana sebuah model dapat mencapai skor akurasi yang tinggi (misalnya, 99%) hanya dengan memprediksi kelas mayoritas yang sangat dominan (zona non-hidrokarbon), sementara sama sekali gagal dalam tugas utamanya untuk mengidentifikasi kelas minoritas yang langka (zona hidrokarbon).²⁷ Untuk menghindari jebakan ini dan mendapatkan pemahaman komprehensif tentang perilaku setiap model, studi ini menggunakan serangkaian metrik kinerja standar:

* **Akurasi (*Accuracy*):** Proporsi dari total prediksi yang benar. Dihitung sebagai (TP+TN)/(TP+TN+FP+FN), di mana TP adalah *True Positives*, TN adalah *True Negatives*, FP adalah *False Positives*, dan FN adalah *False Negatives*.³¹ Meskipun dilaporkan untuk kelengkapan, ini adalah metrik yang paling tidak informatif dalam konteks ini.
* **Presisi (*Precision*):** Mengukur kemurnian dari prediksi positif. Metrik ini menjawab pertanyaan: "Dari semua zona yang dilabeli model sebagai mengandung hidrokarbon, berapa persentase yang benar-benar mengandung hidrokarbon?" Dihitung sebagai TP/(TP+FP). Presisi yang tinggi menunjukkan tingkat alarm palsu yang rendah.
* **Recall (atau Sensitivitas):** Mengukur kemampuan model untuk menemukan semua instansi yang relevan. Metrik ini menjawab pertanyaan: "Dari semua zona yang benar-benar mengandung hidrokarbon di dalam sumur, berapa persentase yang berhasil diidentifikasi dengan benar oleh model?" Dihitung sebagai TP/(TP+FN). Dalam eksplorasi, Recall sering kali menjadi metrik bisnis yang paling kritis, karena kegagalan mengidentifikasi zona produktif potensial (*False Negative*) biasanya lebih merugikan daripada menyelidiki alarm palsu (*False Positive*).
* **Skor-F1 (*F1-Score*):** Rata-rata harmonik dari Presisi dan Recall, dihitung sebagai 2⋅(Presisi⋅Recall)/(Presisi+Recall). Metrik ini memberikan ukuran tunggal yang seimbang dari kinerja model, memberikan penalti pada model yang sangat bias terhadap presisi tinggi atau recall tinggi dengan mengorbankan yang lain.

Kerangka kerja multi-metrik ini esensial tidak hanya untuk memberi peringkat model, tetapi juga untuk memahami profil perilaku mereka yang berbeda. Ini memungkinkan analisis terperinci tentang pertukaran (*trade-offs*) yang dibuat oleh setiap variasi eksperimental, memberikan dasar yang kuat untuk diskusi mendalam yang akan menyusul.

### 3.2 Kinerja Konsolidasi pada Sumur Uji Buta

Ujian definitif dari kemampuan generalisasi setiap model adalah kinerjanya pada data yang tidak terlihat dari sumur uji buta, LLD-14. Tabel berikut memberikan rincian kinerja setiap algoritma pembelajaran mesin di empat skenario eksperimental, berdasarkan laporan klasifikasi yang disediakan. Rata-rata agregat untuk tiga model berbasis pohon ensemble juga disertakan untuk merangkum kinerja kelas algoritme ini. Skor-F1 untuk Kelas 1 (mengandung hidrokarbon) disorot sebagai metrik utama untuk mengevaluasi keberhasilan dalam masalah yang tidak seimbang ini.

**Tabel 1: Perbandingan Kinerja Seluruh Algoritma Pembelajaran Mesin di Empat Skenario Eksperimental pada Sumur Uji Buta LLD-14.**

| **Konfigurasi** | **Model** | **Akurasi** | **Presisi (Kelas 1)** | **Recall (Kelas 1)** | **Skor-F1 (Kelas 1)** |
| --- | --- | --- | --- | --- | --- |
| **Baseline (dengan DR)** | Random Forest | 0,896 | 0,78 | 0,23 | 0,35 |
|  | XGBoost | 0,897 | 0,77 | 0,23 | 0,36 |
|  | LightGBM | 0,909 | 0,79 | 0,37 | 0,50 |
|  | **Rata-rata Ensemble** | **0,901** | **0,78** | **0,28** | **0,40** |
|  | Support Vector Machine | 0,894 | 0,61 | 0,42 | 0,49 |
|  | k-Nearest Neighbors | 0,905 | 0,73 | 0,37 | 0,49 |
| **Tanpa DR** | Random Forest | 0,896 | 0,81 | 0,21 | 0,34 |
|  | XGBoost | 0,895 | 0,62 | 0,38 | 0,47 |
|  | LightGBM | 0,907 | 0,66 | 0,51 | 0,57 |
|  | **Rata-rata Ensemble** | **0,899** | **0,70** | **0,37** | **0,46** |
|  | Support Vector Machine | 0,893 | 0,58 | 0,49 | 0,53 |
|  | k-Nearest Neighbors | 0,900 | 0,76 | 0,28 | 0,41 |
| **SMOTE (dengan DR)** | Random Forest | 0,931 | 0,74 | 0,68 | 0,71 |
|  | XGBoost | 0,920 | 0,68 | 0,69 | 0,68 |
|  | LightGBM | 0,923 | 0,69 | 0,69 | 0,69 |
|  | **Rata-rata Ensemble** | **0,925** | **0,70** | **0,69** | **0,69** |
|  | Support Vector Machine | 0,854 | 0,43 | 0,57 | 0,49 |
|  | k-Nearest Neighbors | 0,857 | 0,45 | 0,68 | 0,54 |
| **SMOTE & Tanpa DR** | Random Forest | 0,913 | 0,65 | 0,67 | 0,66 |
|  | XGBoost | 0,865 | 0,47 | 0,74 | 0,58 |
|  | LightGBM | 0,884 | 0,52 | 0,73 | 0,61 |
|  | **Rata-rata Ensemble** | **0,887** | **0,55** | **0,71** | **0,62** |
|  | Support Vector Machine | 0,856 | 0,44 | 0,58 | 0,50 |
|  | k-Nearest Neighbors | 0,832 | 0,39 | 0,63 | 0,48 |

## Analisis Pengaruh Deep Resistivity dan SMOTE

### 4.1 Peran Ambiguitas Deep Resistivity pada Data Tidak Seimbang

Hipotesis awal mengemukakan bahwa log resistivitas dalam (DR) akan memasukkan noise dan menurunkan kinerja model. Hasil empiris dari set data yang tidak seimbang (skenario tanpa SMOTE) menyajikan gambaran yang kompleks. Membandingkan konfigurasi "Tanpa DR" dengan "Baseline (dengan DR)," dampaknya bervariasi secara signifikan antara algoritma. Untuk model yang lebih canggih seperti LightGBM dan SVM, penghapusan fitur DR meningkatkan Skor-F1 untuk kelas hidrokarbon (masing-masing dari 0,50 menjadi 0,57 dan dari 0,49 menjadi 0,53). Hal ini mendukung gagasan bahwa bagi model-model ini, fitur tersebut lebih banyak menambahkan noise daripada sinyal ketika data tidak seimbang.

Namun, untuk k-NN dan Random Forest, penyertaan fitur DR menghasilkan kinerja yang sedikit lebih baik atau setara. Hal ini menunjukkan bahwa informasi di dalam log DR bukanlah murni noise, melainkan sinyal kompleks yang diproses oleh algoritma yang berbeda dengan tingkat keberhasilan yang bervariasi ketika dihadapkan pada ketidakseimbangan kelas yang parah. Tanpa mengatasi ketidakseimbangan tersebut, nilai sebenarnya dari fitur DR tetap ambigu dan bergantung pada model.

### 4.2 Dampak Signifikan SMOTE terhadap Kinerja Model

Tren yang paling mencolok dalam data adalah efek transformatif dari *Synthetic Minority Over-sampling Technique* (SMOTE). Dalam setiap kasus, untuk setiap algoritma, penerapan SMOTE secara dramatis meningkatkan kemampuan model untuk mengidentifikasi kelas minoritas hidrokarbon. Hal ini paling jelas terlihat pada metrik Recall (Kelas 1), yang mengukur persentase zona hidrokarbon aktual yang berhasil diidentifikasi dengan benar. Untuk model ensemble, rata-rata Recall melonjak dari 0,28 menjadi 0,69 ketika SMOTE diterapkan pada data yang mengandung fitur DR.

Hal ini menunjukkan bahwa tantangan utama dalam masalah ini belum tentu kualitas fitur, melainkan ketidakseimbangan kelas yang parah. Dengan menghasilkan sampel kelas minoritas baru secara sintetis, SMOTE memaksa model untuk mempelajari karakteristik zona yang mengandung hidrokarbon, yang jika tidak, sebagian besar akan diabaikan. Peningkatan substansial pada Skor-F1 untuk Kelas 1 di semua model dalam konfigurasi "SMOTE" dan "SMOTE & Tanpa DR" menegaskan bahwa penyeimbangan data adalah langkah paling kritis untuk membangun model prediktif yang berguna dalam konteks ini.

### 4.3 Efek Sinergis: Membuka Nilai Deep Resistivity dengan SMOTE

Temuan paling signifikan dari evaluasi ulang ini secara langsung bertentangan dengan hipotesis awal. Alih-alih interaksi yang merusak, data mengungkapkan efek sinergis yang kuat antara SMOTE dan fitur resistivitas dalam untuk model-model paling canggih.

Membandingkan dua skenario yang ditingkatkan dengan SMOTE, "SMOTE (dengan DR)" secara konsisten mengungguli "SMOTE & Tanpa DR" untuk model ensemble. Rata-rata Skor-F1 untuk model ensemble adalah 0,69 dengan fitur DR disertakan, dibandingkan dengan hanya 0,62 ketika fitur tersebut dihilangkan. Model individu dengan kinerja terbaik adalah Random Forest dalam konfigurasi "SMOTE (dengan DR)", yang mencapai Skor-F1 sebesar 0,71 untuk kelas hidrokarbon.

Hal ini menunjukkan bahwa setelah ketidakseimbangan kelas dikoreksi oleh SMOTE, model-model tersebut mampu secara efektif mengekstrak sinyal prediktif yang berharga dari log resistivitas dalam. "noise" yang merugikan bagi beberapa model pada set data yang tidak seimbang menjadi fitur prediktif utama pada set data yang seimbang. Kegagalan awal bukan disebabkan oleh fitur itu sendiri, melainkan karena ketidakmampuan model untuk belajar darinya di hadapan ketidakseimbangan kelas yang luar biasa.

## Kesimpulan dan Saran

### 5.1 Kesimpulan

Berdasarkan laporan klasifikasi yang disediakan, telah menghasilkan serangkaian kesimpulan sebagai berikut.

1. Strategi yang paling efektif untuk mengidentifikasi zona yang mengandung hidrokarbon adalah kombinasi dari algoritma pembelajaran mesin ensemble (khususnya Random Forest) dengan set data yang diseimbangkan oleh SMOTE yang menyertakan fitur resistivitas dalam. Pendekatan ini menghasilkan Skor-F1 tertinggi (0,71) untuk kelas hidrokarbon.
2. *Synthetic Minority Over-sampling Technique* (SMOTE) adalah teknik tunggal yang paling berdampak dalam meningkatkan kinerja model. Teknik ini esensial untuk mengatasi ketidakseimbangan kelas yang parah dan memungkinkan model untuk secara efektif mempelajari jejak kelas minoritas hidrokarbon.
3. Log resistivitas dalam mengandung sinyal prediktif berharga yang terbuka setelah set data diseimbangkan menggunakan SMOTE. Pada data yang tidak seimbang, efeknya ambigu, tetapi pada data yang seimbang, penyertaannya secara signifikan meningkatkan kinerja model tingkat atas, menunjukkan interaksi sinergis daripada merusak dengan penggunaan SMOTE.

### 5.2 Saran

Berdasarkan temuan ini, berikut rekomendasi untuk pengembangan alur kerja pembelajaran mesin yang lebih efektif untuk prediksi hidrokarbon di lingkungan reservoir LRLC:

1. Prioritaskan Penanganan Ketidakseimbangan Kelas. Teknik penanganan ketidakseimbangan kelas bukan hanya sebuah optimisasi, tetapi merupakan persyaratan dasar untuk membangun model yang dapat mendeteksi target langka namun kritis seperti zona produktif hidrokarbon dengan andal.
2. Evaluasi Ulang Fitur "Bernoise" pada Data yang Seimbang. Fitur seperti *deep resistivity*, yang mungkin tampak bernoise atau memiliki dampak ambigu pada data yang tidak seimbang, tidak sebaiknya tidak dibuang terlalu dini. Disarankan untuk menguji fitur-fitur ini pada set data yang seimbang, karena nilai prediktif sebenarnya mungkin baru akan terlihat setelah model tidak lagi bias terhadap kelas mayoritas.
3. Manfaatkan Model Ensemble. Data secara konsisten menunjukkan bahwa model berbasis pohon ensemble (Random Forest, XGBoost, LightGBM) adalah pengklasifikasi yang paling efektif untuk cakupan ini, terutama ketika dikombinasikan dengan SMOTE.

### Referensi

1. Well Log Prediction Using Machine Learning - AAPG Datapages/Archives:, accessed August 28, 2025, https://archives.datapages.com/data/ocgs/data/072/072002/pdfs/50.pdf
2. Machine Learning-Based Prediction of Well Logs Guided by Rock ..., accessed August 28, 2025, https://www.mdpi.com/1424-8220/25/3/836
3. Using Machine Learning to Predict Permeability from Well Logs: A Comparative Study of Different Models - ResearchGate, accessed August 28, 2025, https://www.researchgate.net/publication/372762708\_Using\_Machine\_Learning\_to\_Predict\_Permeability\_from\_Well\_Logs\_A\_Comparative\_Study\_of\_Different\_Models
4. Machine learning-based prediction of well logs in the Niger Delta for improved hydrocarbon exploration: Comparison of models for, accessed August 28, 2025, https://ijsra.net/sites/default/files/IJSRA-2024-2379.pdf
5. A Comparison of Machine Learning Approaches for Prediction of Permeability using Well Log Data in the Hydrocarbon Reservoirs | Journal of Human, Earth, and Future, accessed August 28, 2025, https://www.hefjournal.org/index.php/HEF/article/view/42
6. Resistivity logging - Wikipedia, accessed August 28, 2025, https://en.wikipedia.org/wiki/Resistivity\_logging
7. The Defining Series: Basic Well Log Interpretation - SLB, accessed August 28, 2025, https://www.slb.com/resource-library/oilfield-review/defining-series/defining-log-interpretation
8. Borehole Resistivity | US EPA, accessed August 28, 2025, https://www.epa.gov/environmental-geophysics/borehole-resistivity
9. Resistivity Logs - Moodle@Units, accessed August 28, 2025, https://moodle2.units.it/pluginfile.php/374891/mod\_resource/content/1/chapter7.pdf
10. The Ultimate Guide to Resistivity Logging - Number Analytics, accessed August 28, 2025, [tautan mencurigakan telah dihapus]
11. The Essentials of True Resistivity Logging - Number Analytics, accessed August 28, 2025, [tautan mencurigakan telah dihapus]
12. Resistivity log - SEG Wiki, accessed August 28, 2025, https://wiki.seg.org/wiki/Resistivity\_log
13. Production of Low Resistivity, Low Contrast Reservoirs, Offshore ..., accessed August 28, 2025, https://archives.datapages.com/data/gcags/data/042/042001/pdfs/0073.pdf
14. low-resistivity-low-contrast-pay-sands.pdf, accessed August 28, 2025, https://petrophysics.home.blog/wp-content/uploads/2019/05/low-resistivity-low-contrast-pay-sands.pdf
15. EVALUATION OF LOW RESISTIVITY LOW CONTRAST RESERVOIR - UTPedia, accessed August 28, 2025, http://utpedia.utp.edu.my/3368/1/Dissertation%28Jan12%29\_-\_Muhammad\_Amin\_Nizar\_Bin\_Che\_Abd\_Razak%2811796%29.pdf
16. Petrophysics Identifies Low-Resistivity Reservoirs - American Oil & Gas Reporter, accessed August 28, 2025, https://www.aogr.com/magazine/editors-choice/petrophysics-identifies-low-resistivity-reservoirs
17. Recognition and evaluation of low-resistivity pay | Request PDF, accessed August 28, 2025, https://www.researchgate.net/publication/270632310\_Recognition\_and\_evaluation\_of\_low-resistivity\_pay
18. Analyzing Low Resistivity Low Contrast pay of Matar field in The South Cambay Basin. - APG India, accessed August 28, 2025, https://apgindia.org/documents/geoindia/papers/2018/AU286.pdf
19. Petrophysics of Low Resistivity Reservoir - International Journal of Current Science Research and Review, accessed August 28, 2025, https://ijcsrr.org/wp-content/uploads/2024/02/56-2602-2024.pdf
20. Matching of Water Breakthroughs in a Low-Resistivity Oil Reservoir Using Permeability Anisotropy - MDPI, accessed August 28, 2025, https://www.mdpi.com/2076-3417/14/11/4618
    1. RESISTIVITY THEORY 17.1 Introduction 17.2 Basic Definitions and Ohm's Law, accessed August 28, 2025, https://homepages.see.leeds.ac.uk/~earpwjg/PG\_EN/CD%20Contents/GGL-66565%20Petrophysics%20English/Chapter%2017.PDF
21. Dakota--Petrophysics 6--Resistivity Logs - Kansas Geological Survey, accessed August 28, 2025, https://www.kgs.ku.edu/Dakota/vol1/petro/petro06.htm
22. (PDF) Factors That Influence Electrical Resistivity Measurements in Cementitious Systems, accessed August 28, 2025, https://www.researchgate.net/publication/257459439\_Factors\_That\_Influence\_Electrical\_Resistivity\_Measurements\_in\_Cementitious\_Systems
23. Genetic Mechanisms and Identification of Low-Resistivity Pay Zones: A Case Study of Pengyang Area, Ordos Basin - Semantic Scholar, accessed August 28, 2025, https://pdfs.semanticscholar.org/4a25/c9bd7a96079f7097b51b9810f06849e71646.pdf
24. Log interpretation method of resistivity low-contrast oil pays in ..., accessed August 28, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC8776901/
25. SMOTE: Synthetic Minority Over-sampling Technique - arXiv, accessed August 28, 2025, https://arxiv.org/pdf/1106.1813
26. Smote for Imbalanced Classification with Python, Technique - Analytics Vidhya, accessed August 28, 2025, https://www.analyticsvidhya.com/blog/2020/10/overcoming-class-imbalance-using-smote-techniques/
27. An improved SMOTE algorithm for enhanced imbalanced data classification by expanding sample generation space - PMC, accessed August 28, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC12222711/
28. SMOTE for high-dimensional class-imbalanced data - PMC, accessed August 28, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC3648438/
29. SMOTE Version 0.14.0 - Imbalanced-learn, accessed August 28, 2025, https://imbalanced-learn.org/stable/references/generated/imblearn.over\_sampling.SMOTE.html
30. Effective Class-Imbalance Learning Based on SMOTE and Convolutional Neural Networks, accessed August 28, 2025, https://www.mdpi.com/2076-3417/13/6/4006