

GeoAI: Teknik ve Matematiksel Analiz Raporu

Olasılıksal Jeofizik Hedefleme için Uçtan-Uca Boru Hattı

(Genişletilmiş L^AT_EX Sürümü)

19 Şubat 2026

Özet

Bu rapor, GeoAI sisteminin çok-kaynaklı jeofizik grid katmanları ve kuyu etiketlerini kullanarak *prospektivite* (hedef olasılığı) ve *belirsizlik* haritaları üretme sürecini matematiksel olarak formüle eder. Co-registration (ortak grid'e hizalama), fizik-tabanlı/istatistiksel feature mühendisliği, robust normalizasyon, sınıf dengesizliği altında örnekleme ve SMOTE, olasılık kalibrasyonu ve PR-AUC ağırlıklı ensemble birleşimi ayrıntılandırılır. Ayrıca eşik optimizasyonu, bağlı bileşenler üzerinden hedef çıkarımı ve kuyu verisi bulunmadığında devreye giren unsupervised geri-düşüm mekanizması (textitfallback) için formal tanımlar verilir.

İçindekiler

1 Amaç, Kapsam ve Problem Tanımı	2
2 Notasyon ve Veri Modeli	2
3 Co-registration: Ortak Alan, Çözünürlük ve İnterpolasyon	2
3.1 Ortak alan seçimi (intersection/union)	2
3.2 İnterpolasyon ve NaN doldurma	2
4 Feature Mühendisliği	3
4.1 Türev-tabanlı feature'lar	3
4.2 Ayrıştırma ve dağılım sıkıştırma	3
4.3 Pencere istatistikleri ve tekstür	3
5 Robust Normalizasyon ve Uç Değer Bastırma	3
6 Kuyu Tabanlı Örnekleme ve Etiket Üretimi	4
6.1 Buffer ile piksel etiketi	4
7 Sınıf Dengesizliği ve SMOTE	4
8 Modelleme: Kalibre Edilmiş Olasılık Tahmini ve Ensemble	4
8.1 Kalibre edilmiş olasılık	5
8.2 Cross-validation (CV) ve OOF olasılıklar	5
8.3 Metrikler: ROC-AUC ve PR-AUC	5
8.4 PR-AUC ağırlıklı ensemble	5
9 Belirsizlik Modellemesi ve Güven Skoru	5
10 Eşik Optimizasyonu ve Hedef Çıkarımı	6
10.1 Aday bölge maskesi ve bağlı bileşenler	6

11 Kuyu Yoksa: Unsupervised Fallback	6
12 Hesaplama Karmaşıklığı ve Performans	6
13 Sızıntı Önleme (Leakage) ve Uzaysal CV Önerisi	6
14 Pipeline Parametre Sözlüğü ve Operasyonel Karar Kuralları	7
15 Sonuç	7
A Algoritmik Özet (Uçtan-Uca)	7

1 Amaç, Kapsam ve Problem Tanımı

GeoAI; jeofizik katmanlar (manyetik, gravite, jeokimya, uzaktan algılama vb.) ile kuyu/sondaj etiketlerini birleştirerek, her piksel için hedef olasılığı $P(\mathbf{s})$ ve belirsizlik $U(\mathbf{s})$ üreten bir karar destek sistemidir. Operasyonel amaç, sondaj maliyetini düşürmek için sahadaki *önceliklendirilebilir hedef listelerini* nicel bir skor düzenine dönüştürmektir.

Veri tipleri ve çıktılar. Girdi olarak düzenli grid ve nokta verileri kabul edilir (GRD, GeoTIFF, XYZ, CSV vb.). Ara ürünler: ortak referans grid'e taşınmış katmanlar, feature tensörü ve normalize edilmiş temsil. Çıktılar: hedef koordinat listesi, prospektivite haritası, belirsizlik haritası ve rapor artefaktları.

Denetimli öğrenme problemi. Her piksel konumu $\mathbf{s} = (x, y)$ için bir feature vektörü $\mathbf{x}(\mathbf{s}) \in \mathbb{R}^{n_f}$ tanımlanır. Kuyu etiketlerinden türetilen ikili sınıf etiketi $y(\mathbf{s}) \in \{0, 1\}$ ile aşağıdaki olasılıksal sınıflandırma hedeflenir:

$$\hat{p}(\mathbf{s}) \approx \mathbb{P}(y(\mathbf{s}) = 1 \mid \mathbf{x}(\mathbf{s})). \quad (1)$$

2 Notasyon ve Veri Modeli

K adet jeofizik katman, her biri kendi örnekleme ızgarasında düzenli 2B alan olarak:

$$G_k : \mathbb{R}^2 \rightarrow \mathbb{R} \cup \{\text{NaN}\}, \quad G_k(y, x) \text{ veya } \mathbf{G}_k \in \mathbb{R}^{n_{y,k} \times n_{x,k}}. \quad (2)$$

Her katmanın eksen vektörleri $\mathbf{x}_k \in \mathbb{R}^{n_{x,k}}$, $\mathbf{y}_k \in \mathbb{R}^{n_{y,k}}$.

Ortak referans grid:

$$\mathbf{X}_{\text{ref}} \in \mathbb{R}^{n_x}, \quad \mathbf{Y}_{\text{ref}} \in \mathbb{R}^{n_y}.$$

Tüm katmanlar bu grid'e yeniden örneklenir ve feature kanalları yığılarak bir tensör elde edilir:

$$\mathbf{F} \in \mathbb{R}^{n_y \times n_x \times n_f}, \quad \mathbf{F}[r, c, :] = \mathbf{x}(\mathbf{s}_{r,c}). \quad (3)$$

3 Co-registration: Ortak Alan, Çözünürlük ve İnterpolasyon

3.1 Ortak alan seçimi (intersection/union)

Her katman için alan sınırları:

$$x_{k,\min} = \min(\mathbf{x}_k), \quad x_{k,\max} = \max(\mathbf{x}_k), \quad y_{k,\min} = \min(\mathbf{y}_k), \quad y_{k,\max} = \max(\mathbf{y}_k).$$

Intersection modunda ortak alan:

$$x_{\min}^{\text{ref}} = \max_k x_{k,\min}, \quad x_{\max}^{\text{ref}} = \min_k x_{k,\max}, \quad (4)$$

$$y_{\min}^{\text{ref}} = \max_k y_{k,\min}, \quad y_{\max}^{\text{ref}} = \min_k y_{k,\max}. \quad (5)$$

Union modunda max / min operatörleri terslenir. Hedef çözünürlük (n_x, n_y) verilmediyse katmanlar arasındaki en yüksek çizgisel çözünürlük seçilir ve pratikte bellek/süre için piksel üst limitleri uygulanır.

3.2 İnterpolasyon ve NaN doldurma

Her katman, bir yeniden örnekleme operatörü ile referans grid'e taşınır:

$$\mathbf{G}'_k = \mathcal{I}_k(\mathbf{Y}_{\text{ref}}, \mathbf{X}_{\text{ref}}), \quad (6)$$

burada \mathcal{I}_k lineer veya en yakın komşu interpolasyon (ör. `RegularGridInterpolator`) olabilir. Eksik bölgeler için:

$$\mathbf{G}_k \leftarrow \text{FillNearest}(\mathbf{G}'_k) \quad (\text{yalnızca NaN hücrelerde}). \quad (7)$$

Sayısal not. Lineer interpolasyon, türev-tabanlı feature'larda düzgünlük sağlar; ancak yoğun NaN bölgeleri saha dışı extrapolasyona sebep olabileceğinden intersection alanı çoğu zaman daha güvenlidir.

4 Feature Mühendisliği

Her katman türüne bağlı olarak fiziksel/jeolojik anlamı olan türevler ve istatistiksel özetler çıkarılır. Aşağıda F tek bir skaler grid kanalı (ör. manyetik anomali) olarak ele alınmıştır.

4.1 Türev-tabanlı feature'lar

İki boyutta yatay gradyan büyüklüğü (total horizontal gradient, THG):

$$\text{THG} = \sqrt{\left(\frac{\partial F}{\partial x}\right)^2 + \left(\frac{\partial F}{\partial y}\right)^2}. \quad (8)$$

Tilt açısı (TA):

$$\text{TA} = \arctan\left(\frac{\frac{\partial F}{\partial z}}{\text{THG} + \varepsilon}\right), \quad (9)$$

burada $\varepsilon > 0$ sayısal kararlılık içindir.

Analitik sinyal genliği (ASA) (manyetik türevleri için tipik):

$$\text{ASA} = \sqrt{\left(\frac{\partial B}{\partial x}\right)^2 + \left(\frac{\partial B}{\partial y}\right)^2 + \left(\frac{\partial B}{\partial z}\right)^2}. \quad (10)$$

4.2 Ayırıştırma ve dağılım sıkıştırma

Yerel-anomali vurgusu için Gauss bulanıklaştırma ile rezidüel:

$$R = F - \mathcal{G}_\sigma(F), \quad (11)$$

burada \mathcal{G}_σ standart sapması σ olan Gauss filtre operatörüdür.

Ağır kuyruklu dağılımları bastırmak için işaretli log dönüşümü:

$$S = \text{sign}(F) \log(1 + |F|). \quad (12)$$

4.3 Pencere istatistikleri ve tekstür

$W \times W$ yerel pencerede:

$$\mu_W(\mathbf{s}) = \text{mean}(F \text{ in window}), \quad \sigma_W(\mathbf{s}) = \text{std}(\cdot), \quad \text{range}_W(\mathbf{s}) = \max - \min.$$

Bu özetler heterojenlik/tekstür sinyali taşır.

5 Robust Normalizasyon ve Uç Değer Bastırma

Her feature kanalı bağımsız olarak robust ölçeklenir. i 'inci feature kanalındaki ham değerler $\{x_{i,n}\}_{n=1}^N$. Önce yüzdelik kırpma:

$$x_{i,n} \leftarrow \text{clip}(x_{i,n}, q_i(\alpha), q_i(1 - \alpha)), \quad (13)$$

$q_i(\cdot)$ ilgili yüzdelik, α kırpma oranıdır.

Sonra medyan ve IQR ile robust z-skoru:

$$z_{i,n} = \frac{x_{i,n} - \text{median}(x_i)}{\max(\text{IQR}(x_i), 10^{-10})}, \quad \text{IQR} = q_{0.75} - q_{0.25}. \quad (14)$$

Küresel kırpma:

$$z_{i,n} \leftarrow \text{clip}(z_{i,n}, -5, 5). \quad (15)$$

Bu işlem, modelin ölçek duyarlılığını azaltır ve SMOTE gibi yöntemlerin metrik uzayda daha stabil çalışmasına yardımcı olur.

6 Kuyu Tabanlı Örnekleme ve Etiket Üretimi

Kuyu noktaları $\{\mathbf{s}_j\}_{j=1}^{N_w}$ referans grid'e projeksiyon ile taşınır. Pozitif kuyular (hedef mineralizasyon vb.) ve negatif kuyular için farklı tampon yarıçapları kullanılarak piksel etiketi türetilir.

6.1 Buffer ile piksel etiketi

Referans grid hücreleri $\mathbf{s}_{r,c}$ olsun. Pozitif set:

$$\mathcal{P} = \{j : \text{kuyu } j \text{ pozitif}\}, \quad \mathcal{N} = \{j : \text{kuyu } j \text{ negatif}\}.$$

Pozitif tampon r_+ , negatif tampon r_- . Piksel etiketi örnek bir kural ile:

$$y(\mathbf{s}) = \begin{cases} 1, & \exists j \in \mathcal{P} : \|\mathbf{s} - \mathbf{s}_j\| \leq r_+, \\ 0, & \exists j \in \mathcal{N} : \|\mathbf{s} - \mathbf{s}_j\| \leq r_-, \\ (\text{örneklenmez}), & \text{aksi halde.} \end{cases} \quad (16)$$

Dengesizliği azaltmak için pozitiflerden uzakta ek rastgele negatifler seçilebilir:

$$\mathcal{S}_{\text{neg-extra}} = \{\mathbf{s} : d(\mathbf{s}, \mathcal{P}) \geq d_{\min}\}.$$

Uzaysal sızıntı riski. Buffer tabanlı etiketleme, komşu pikseller arasında güçlü korelasyon oluşturur. Bu nedenle (i) fold içinde SMOTE uygulanması, (ii) tercihen *spatial CV* (bloklu katlama) gibi stratejiler önerilir.

7 Sınıf Dengesizliği ve SMOTE

Pozitif örnek sayısı n_{pos} , negatif örnek sayısı n_{neg} . Dengesizlik oranı:

$$\rho = \frac{n_{\text{neg}}}{n_{\text{pos}}}. \quad (17)$$

SMOTE, pozitif uzayda en yakın komşular arasında doğrusal enterpolasyonla sentetik örnek üretir:

$$\mathbf{x}_{\text{new}} = \mathbf{x}_i + \alpha(\mathbf{x}_j - \mathbf{x}_i), \quad \alpha \sim \mathcal{U}(0, 1), \quad (18)$$

$(\mathbf{x}_i, \mathbf{x}_j)$ pozitif sınıf komşularıdır. Kritik prensip: SMOTE yalnızca eğitim fold'unda uygulanır, doğrulama fold'una sentetik örnek sızdırılmaz.

8 Modelleme: Kalibre Edilmiş Olasılık Tahmini ve Ensemble

GeoAI, farklı indüktif önyargılara sahip birden fazla modeli birlikte kullanır: (i) çok katmanlı algılayıcı (ANN/MLP), (ii) Random Forest (RF), (iii) Histogram tabanlı Gradient Boosting (HGB).

8.1 Kalibre edilmiş olasılık

Bir sınıflandırıcı f_m ham skor üretsın. Kalibrasyon fonksiyonu c_m ile:

$$\hat{p}_m(\mathbf{x}) = c_m(f_m(\mathbf{x})),$$

burada c_m örn. Platt ölçekleme (lojistik) veya izotonik regresyon olabilir. Kalibrasyon, Brier skorunu iyileştirerek harita üzerindeki olasılıkların karar kurallarına uygun yorumlanmasını sağlar:

$$\text{Brier} = \frac{1}{N} \sum_{n=1}^N (\hat{p}_n - y_n)^2. \quad (19)$$

8.2 Cross-validation (CV) ve OOF olasılıklar

Stratified K_{cv} -fold ile her fold'da:

$$(\mathcal{D}_{tr}^{(k)}, \mathcal{D}_{va}^{(k)}), \quad \text{SMOTE sadece } \mathcal{D}_{tr}^{(k)} \text{ üzerinde.}$$

Out-of-fold (OOF) olasılıkları biriktirilerek eşik optimizasyonu ve ensemble ağırlıkları için yanlış tahmin elde edilir.

8.3 Metrikler: ROC-AUC ve PR-AUC

ROC-AUC, \hat{p} sıralama kalitesini ölçer; ancak nadir pozitiflerde PR-AUC daha duyarlıdır. PR eğrisi:

$$\text{precision}(t) = \frac{\text{TP}(t)}{\text{TP}(t) + \text{FP}(t)}, \quad \text{recall}(t) = \frac{\text{TP}(t)}{\text{TP}(t) + \text{FN}(t)}.$$

PR-AUC, bu eğri altındaki alandır.

8.4 PR-AUC ağırlıklı ensemble

Model başına ağırlık:

$$w_m = \frac{\text{PR-AUC}_m}{\sum_{j=1}^M \text{PR-AUC}_j}, \quad \sum_m w_m = 1. \quad (20)$$

Ensemble olasılığı:

$$P(\mathbf{s}) = \sum_{m=1}^M w_m \hat{p}_m(\mathbf{x}(\mathbf{s})). \quad (21)$$

9 Belirsizlik Modellemesi ve Güven Skoru

Belirsizlik $U(\mathbf{s})$ için pratik bir yaklaşım, model çeşitliliği üzerinden ansambl varyansı:

$$U(\mathbf{s}) = \sqrt{\sum_{m=1}^M w_m (\hat{p}_m(\mathbf{s}) - P(\mathbf{s}))^2}. \quad (22)$$

Alternatifler: (i) RF OOB varyansı, (ii) MLP için MC-dropout, (iii) kalibrasyon hatası tabanlı haritalar.

Operasyonel skor (belirsizlik cezalıyıcı):

$$S(\mathbf{s}) = \text{clip}(P(\mathbf{s}) - \lambda U(\mathbf{s}), 0, 1), \quad \lambda \in [0, 1]. \quad (23)$$

10 Eşik Optimizasyonu ve Hedef Çıkarımı

Sabit 0.5 eşiği yerine veri odaklı eşik seçimi yapılır. F1 skoru:

$$F1(t) = \frac{2 \text{precision}(t) \text{recall}(t)}{\text{precision}(t) + \text{recall}(t)}. \quad (24)$$

Optimum eşik:

$$t^* = \arg \max_{t \in [t_{\min}, t_{\max}]} F1(t). \quad (25)$$

10.1 Aday bölge maskesi ve bağlı bileşenler

Aday bölge kümesi:

$$\mathcal{B} = \{\mathbf{s}_{r,c} : P(\mathbf{s}_{r,c}) \geq \max(\tau, q_{0.80}(P))\}. \quad (26)$$

\mathcal{B} üzerinde 4- veya 8-komşulukla bağlı bileşen etiketleme uygulanır. Her bileşen için ağırlıklı merkez (hedef koordinatı) hesaplanır:

$$w_i = \max(S_i, 0) + \varepsilon, \quad \mathbf{s}^* = \frac{\sum_{i \in \mathcal{C}} w_i \mathbf{s}_i}{\sum_{i \in \mathcal{C}} w_i}. \quad (27)$$

Alan tahmini:

$$A_{\text{km}^2} = n_{\text{pix}} \frac{|\Delta x \Delta y|}{10^6}.$$

11 Kuyu Yoksa: Unsupervised Fallback

Kuyu verisi yoksa, sistem anomali gücü ve feature dağılımı üzerinden normalize edilmiş bir skor üretir. Feature tensöründe $f = 1, \dots, n_f$:

$$\text{strength}(\mathbf{s}) = \frac{1}{n_f} \sum_{f=1}^{n_f} |F(\mathbf{s}, f)|, \quad \text{dispersion}(\mathbf{s}) = \text{std}_f(F(\mathbf{s}, f)). \quad (28)$$

Yüzdelik ölçekleme operatörü $\pi_{[a,b]}(\cdot)$ ile:

$$P_{\text{unsup}} = \pi_{[5,95]}(\text{strength}), \quad U_{\text{unsup}} = \pi_{[5,95]}(\text{dispersion}).$$

12 Hesaplama Karmaşıklığı ve Performans

K katman, referans grid boyutu $n_x \times n_y$, feature sayısı n_f , model sayısı M .

Aşama	Yaklaşık karmaşıklık	Not
Co-registration	$\mathcal{O}(Kn_x n_y)$	İnterpolasyon baskın
Feature extraction	$\mathcal{O}(n_f n_x n_y)$	Türev ve filtreler
Eğitim (CV)	modele bağlı	K_{cv} ile ölçeklenir
Grid tahmini	$\mathcal{O}(Mn_x n_y)$	Batch ile parçalı

13 Sızıntı Önleme (Leakage) ve Uzaysal CV Önerisi

Standart stratified CV, uzaysal korelasyon nedeniyle metrikleri iyimser şişirebilir. İyileştirme: (i) *bloklu/uzaysal* katlama, (ii) aynı kuyu çevresi piksellerinin tek fold'a atanması, (iii) buffer bölgelerini gruptama.

14 Pipeline Parametre Sözlüğü ve Operasyonel Karar Kuralları

Parametrelerin (örn. n_x, n_y, λ , pencere boyutu, minimum olasılık τ) bellek ve saha kararına etkisi *tasarım değişkenleri* olarak ele alınmalıdır. Örnek karar kuralı:

$$\text{Öncelik-1: } P \geq 0.75 \wedge U \leq 0.20, \quad \text{Öncelik-2: } 0.60 \leq P < 0.75 \wedge U \leq 0.30.$$

15 Sonuç

GeoAI, jeofizik haritaları yalnızca görselleştirmekten öte, kuyu verisini feature-tabanlı makine öğrenmesi ile birleştirerek sondaj kararlarını olasılık ve belirsizlik düzleminde nicelleştiren modüler bir boru hattıdır. Raporlanan matematiksel çerçeve; kalibrasyon, dengesizlik yönetimi ve uzaysal sızıntı kontrolü gibi saha uygulamalarında kritik ayrıntıları açık biçimde ortaya koyar.

A Algoritmik Özet (Uçtan-Uca)

Algorithm 1 GeoAI uçtan-uca boru hattı (özet)

Require: Katmanlar $\{\mathbf{G}_k\}_{k=1}^K$, kuyu noktaları ve etiketler (varsa)

Ensure: $P(\mathbf{s})$, $U(\mathbf{s})$, hedef listesi

- 1: Ortak alan ve çözünürlük seç; referans grid'i kur
 - 2: **for** $k = 1, \dots, K$ **do**
 - 3: NaN doldur; $\mathbf{G}'_k \leftarrow \mathcal{I}_k(\mathbf{X}_{ref}, \mathbf{Y}_{ref})$
 - 4: **end for**
 - 5: Feature kanallarını çıkar; $\mathbf{F} \leftarrow \text{StackFeatures}(\{\mathbf{G}'_k\})$
 - 6: Robust normalize et; $\mathbf{F} \leftarrow \text{RobustScale}(\mathbf{F})$
 - 7: **if** kuyu etiketi var **then**
 - 8: Eğitim örnekleri üret; sınıf dengesizliğini analiz et
 - 9: **for** CV fold **do**
 - 10: SMOTE yalnızca train üzerinde; modelleri eğit ve kalibre et
 - 11: Valid üzerinde OOF olasılık biriktir
 - 12: **end for**
 - 13: PR-AUC ağırlıkları ile ensemble; $P(\mathbf{s})$
 - 14: Model çeşitliliği ile belirsizlik; $U(\mathbf{s})$
 - 15: **else**
 - 16: Unsupervised fallback; $P_{\text{unsup}}, U_{\text{unsup}}$
 - 17: **end if**
 - 18: Skor haritası $S = P - \lambda U$; eşik ve bağlı bileşenler ile hedef çıkar
-