

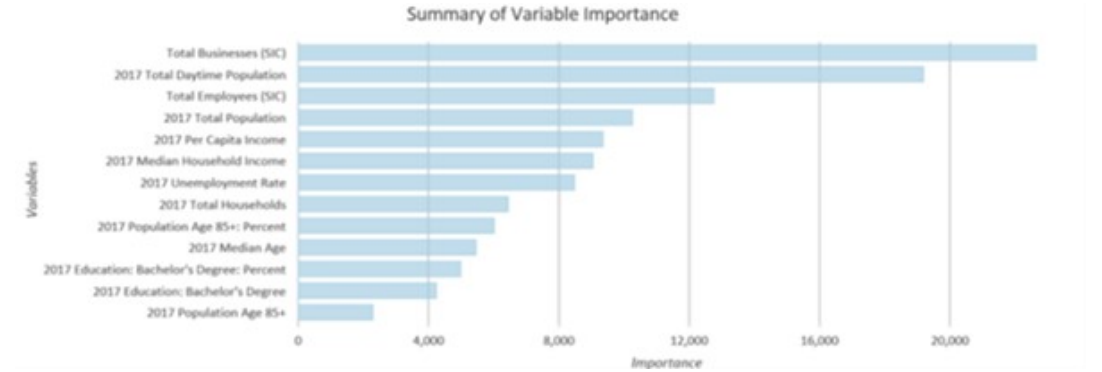
YAPAY ZEKÂ DESTEKLİ COĞRAFİ BİLGİ SİSTEMLERİ

10. HAFTA

Denetimli Mekânsal Makine Öğrenmesi

YBS 363 · Ankara Bilim Üniversitesi

Dr. Murat BAŞEĞMEZ



Öğrenme çıktıları

Ders sonunda öğrenci

- Sınıflandırma ve regresyon problemlerini doğru hedef türüyle eşleştirir
- Random Forest ve Gradient Boosting modellerinin çalışma mantığını açıklar
- ArcGIS Pro Forest-based and Boosted aracını kurar ve çıktıları yorumlar
- Performansı mekânsal doğrulama, hata haritası ve belirsizlikle değerlendirir

Çıktılar kavram, uygulama ve eleştirel değerlendirme düzeylerini birlikte kapsar.

Üç saatlik ders akışı

10. HAFTA

00–55 dk

Denetimli öğrenme problemi

55–110 dk

Mekânsal özellikler ve eğitim

Ara

10 dakika

120–165 dk

ArcGIS Pro iş akışı ve Doğrulama, yorum ve raporlama

165–180 dk

Laboratuvar başlangıcı ve değerlendirme

BÖLÜM 1

Denetimli öğrenme problemi

Model, bilinen hedeflerden yeni gözlemler için tahmin öğrenir

Bu bölümde kavram, yöntem ve uygulama bağlantısı kurulacaktır.

Sınıflandırma ve regresyon

DENETİMLİ ÖĞRENME PROBLEMİ

Hedef değişkenin yapısı yöntem ve metriği belirler

Sınıflandırma

İkili: uygun / uygun değil
Çok sınıflı: düşük / orta / yüksek
Çıktı: sınıf ve olasılık

Regresyon

Sürekli hedef: fiyat, talep, sıcaklık
Çıktı: sayısal tahmin
Hata: gözlenen eksi tahmin

Temel model neden gereklidir?

DENETİMLİ ÖĞRENME PROBLEMİ

Karmaşık model basit bir referansı geçmelidir

- Regresyonda ortalama veya doğrusal model temel olabilir.
- Sınıflandırmada çoğunluk sınıfı veya basit lojistik model kullanılabilir.
- Temel model, yeni yöntemin gerçekten değer üretip üretmediğini gösterir.
- Aynı veri bölmesi ve aynı ölçütlerle karşılaştırma yapılmalıdır.

ANLATIM ŞEMASI

01

Regresyonda
ortalama veya
doğrusal model
teme...

02

Sınıflandırmada
çoğunluk sınıfı
veya basit lo...

03

Temel model,
yeni yöntemin
gerçekten değer
ür...

04

Aynı veri
bölmesi ve aynı
ölçütlerle
karşılaş...

Karar ağacı

DENETİMLİ ÖĞRENME PROBLEMİ

Ağaç, özellik eşikleriyle gözlemleri giderek daha homojen gruplara ayırır

- Her düğüm bir değişken ve eşik seçer.
- Yaprak düğüm sınıf veya sayısal tahmin üretir.
- Derin ağaç eğitim verisini ezberleyebilir.
- Tek ağaç anlaşılırdır ancak veri değişimine karşı kararsız olabilir.

ANLATIM ŞEMASI

Her düğüm bir değişken ve eşik seçer

Yaprak düğüm sınıf veya sayısal tahmin üretir

Derin ağaç eğitim verisini ezberleyebilir

Tek ağaç anlaşılırdır ancak veri değişimine k...

genelden uygulamaya

Random Forest

DENETİMLİ ÖĞRENME PROBLEMİ

Çok sayıda farklı ağacın birleşimi varyansı azaltır

Model Characteristics

Number of Trees	100
Leaf Size	5
Tree Depth Range	30-42
Mean Tree Depth	35
% of Training Available per Tree	100
Number of Randomly Sampled Variables	2
% of Training Data Excluded for Validation	10

Görsel okuma: önce girdi/çıktıyı, sonra renk ve sembollerin neyi temsil ettiğini belirleyin.

- Her ağaç farklı gözlem ve değişken alt kümesiyle kurulur.
- Sınıflandırmada oylama, regresyonda ortalama kullanılır.
- Out-of-bag gözlemler ek bir hata tahmini sağlar.
- Ağaç sayısı arttıkça sonuç genellikle daha kararlı hale gelir.

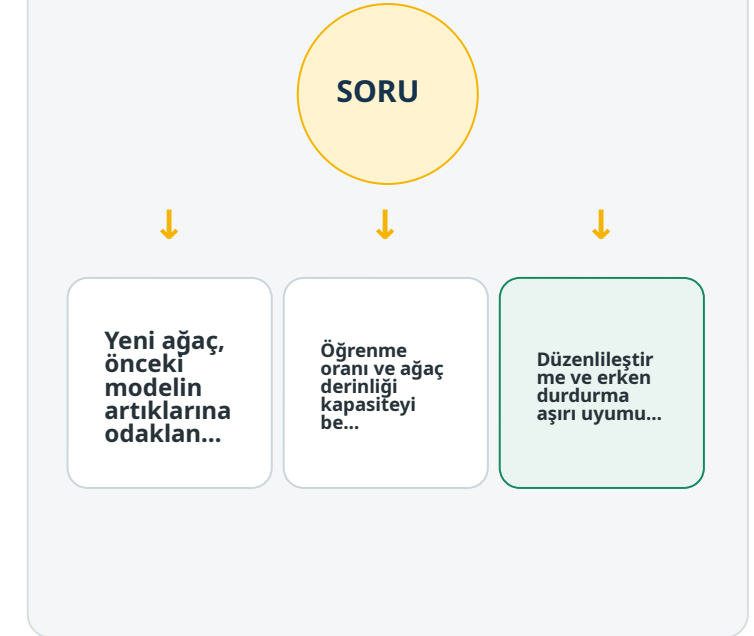
Gradient Boosting

DENETİMLİ ÖĞRENME PROBLEMİ

Ardışık ağaçlar önceki hataları azaltmaya çalışır

- Yeni ağaç, önceki modelin artıklarına odaklanır.
- Öğrenme oranı ve ağaç derinliği kapasiteyi belirler.
- Düzenleştirme ve erken durdurma aşırı uyumu azaltır.
- Boosting güçlü olabilir ancak ayarlara Random Forest'tan daha duyarlıdır.

ANLATIM ŞEMASI



BÖLÜM 2

Mekânsal özellikler ve eğitim

Model girdileri coğrafi süreçleri temsil etmelidir

Bu bölümde kavram, yöntem ve uygulama bağlantısı kurulacaktır.

Koordinat eklemek tek başına mekânsal model oluşturmaz

- Mesafe, erişim, yoğunluk ve komşuluk süreçle ilişkilendirilmelidir.
- Raster değerleri hedef tarihle uyumlu olmalıdır.
- Kategorik değişkenler doğru biçimde kodlanmalıdır.
- Aynı olguyu tekrar eden yüksek korelasyonlu değişkenler sadeleştirilebilir.

ANLATIM ŞEMASI

1 Mesafe, erişim, yoğunluk ve komşuluk süreçle...



2 Raster değerleri hedef tarihle uyumlu olmalıdır...



3 Kategorik değişkenler doğru biçimde kodlanmalıdır...

Mesafe özellikleri

MEKÂNSAL ÖZELLİKLER VE EĞİTİM

Yakınlık değişkeni fiziksel erişim veya maruziyeti temsil eder

- ArcGIS aracı, eğitim nesnelerinden en yakın açıklayıcı özelliğe uzaklık hesaplayabilir.
- Nokta, çizgi ve poligonlarda en yakın segment dikkate alınır.
- Öklid mesafesi her zaman yolculuk süresini temsil etmez.
- Mesafe türü ve koordinat sistemi raporlanmalıdır.

ANLATIM ŞEMASI

01

ArcGIS aracı,
eğitim
nesnelerinden
en yakın a...

02

Nokta, çizgi ve
poligonlarda en
yakın segment...

03

Öklid mesafesi
her zaman
yolculuk süresini
te...

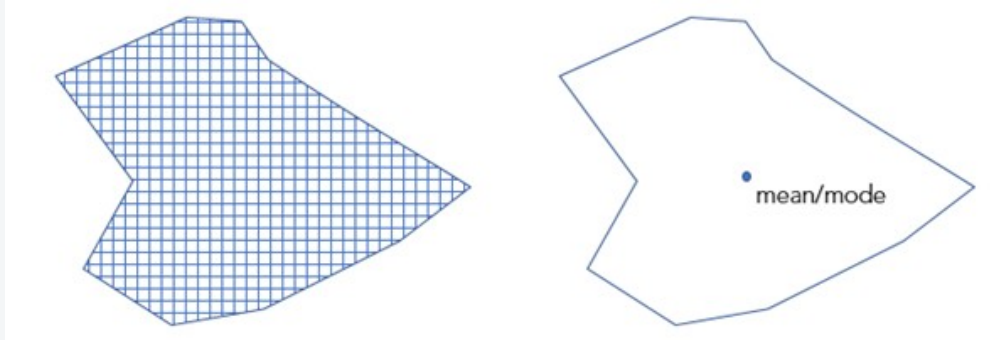
04

Mesafe türü ve
koordinat
sistemi
raporlanmalı...

Açıklayıcı rasterlar

MEKÂNSAL ÖZELLİKLER VE EĞİTİM

Model nokta veya poligon konumunda raster bilgisini örnekler



Görsel okuma: önce girdi/çıktıyı, sonra renk ve sembollerin neyi temsil ettiğini belirleyin.

- Noktalar için hücre değeri alınır.
- Poligonlarda ortalama veya çoğunluk kullanılabilir.
- Rasterların tahmin alanında da bulunması gerekir.
- Eğitim aralığının çok dışındaki değerlerde güvenilir ekstrapolasyon beklenmez.

Dengesiz sınıflar

MEKÂNSAL ÖZELLİKLER VE EĞİTİM

Yüksek doğruluk azınlık sınıfının kaçırıldığını gizleyebilir

- Sınıf başına örnek sayısını ve coğrafi dağılımı inceleyin.
- Accuracy yanında precision, recall, F1 ve MCC raporlayın.
- Eşik değişiminin yanlış pozitif ve yanlış negatifleri nasıl etkilediğini gösterin.
- Sınıf ağırlığı veya dengeli örnekleme yalnız eğitim verisinde uygulanmalıdır.

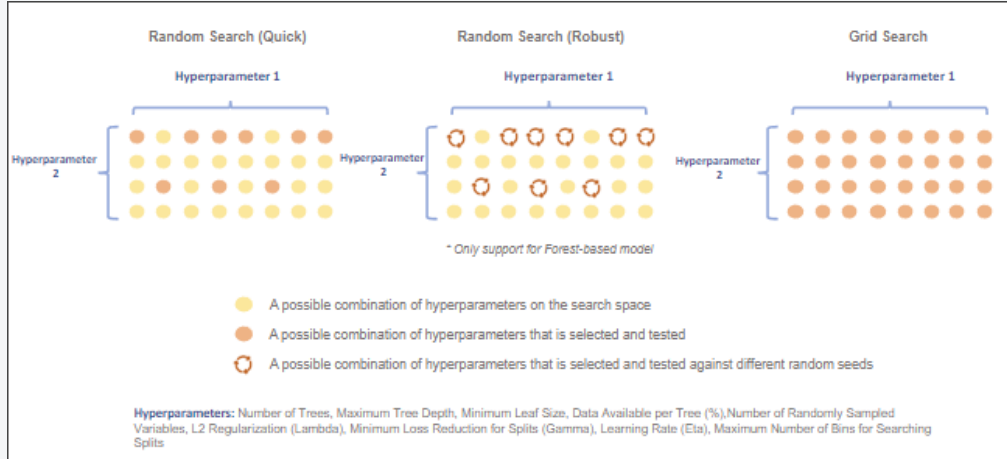
ANLATIM ŞEMASI



Hiperparametre araması

MEKÂNSAL ÖZELLİKLER VE EĞİTİM

Arama yalnız eğitim ve doğrulama kümeleri içinde yürütülür



Görsel okuma: önce girdi/çıktıyı, sonra renk ve sembollerin neyi temsil ettiğini belirleyin.

- Ağaç sayısı, derinlik, yaprak boyutu ve örneklenen değişken sayısı sınırlanabilir.
- Boosting için öğrenme oranı ve düzenleme önemlidir.
- Arama uzayı önce makul sınırlarla tanımlanır.
- Son seçim bağımsız test bölgesinde yalnız bir kez değerlendirilir.

BÖLÜM 3

ArcGIS Pro iş akışı

Train only ile model incelenir, ardından tahmin üretilir

Bu bölümde kavram, yöntem ve uygulama bağlantısı kurulacaktır.

Forest-based and Boosted aracı

ARCGIS PRO İŞ AKIŞI

Araç sınıflandırma ve regresyonu aynı iş akışında destekler

- Input Training Features ve Variable to Predict seçilir.
- Alan, mesafe özelliği ve raster girdileri eklenebilir.
- Model türü Forest-based veya Gradient Boosted olarak belirlenir.
- Train only, Predict to features veya Predict to raster modu seçilir.

ANLATIM ŞEMASI

1 Input Training Features
ve Variable to Predic...



2 Alan, mesafe özelliği ve
raster girdileri ekl...



3 Model türü Forest-based
veya Gradient Boosted...

Önerilen çalışma sırası

ARCGIS PRO İŞ AKIŞI

Önce modeli anlayın, sonra tahmin haritası üretin

- Train only ile değişkenleri ve ayarları değerlendirin.
- Doğrulama sonuçlarını ve değişken aralıklarını inceleyin.
- Seçilen modeli tüm eğitim verisiyle yeniden kurun.
- Tahmin alanında aynı değişkenleri ve birimleri sağlayın.
- Tahmin ve hata haritasını birlikte paylaşın.

ANLATIM ŞEMASI

01

Train only ile
değişkenleri ve
ayarları değer...

02

Doğrulama
sonuçlarını ve
değişken
aralıklarını...

03

Seçilen modeli
tüm eğitim
verisiyle yeniden
k...

04

Tahmin alanında
aynı
değişkenleri ve
birimleri...

Değişken önemi

ARCGIS PRO İŞ AKIŞI

Önem, modelin hangi değişkenleri bölmelerde kullandığını gösterir

- Random Forest için önem Gini katkılarından türetilir.
- Boosting için gain, weight ve cover farklı anlamlara sahiptir.
- Yüksek önem nedensellik kanıtı değildir.
- Korelasyonlu değişkenler önemi aralarında paylaşabilir.

Top Variable Importance

Variable	Importance	%
2019 Male Population	6592.66	22
2019 Median Age	4192.82	14
2019 Average Household Income	3872.12	13
2019 Female Population	3577.04	12
Used prescription drug for high blood pressure	3225.16	11
2019 Minority Population	3046.65	10
2019 Alcoholic Beverages	2914.30	10
Used prescr drug for diabetes (insulin dependent)	2310.24	8

Görsel okuma: önce girdi/çıktıyı, sonra renk ve sembollerin neyi temsil ettiğini belirleyin.

Eğitim ve doğrulama tanıları

ARCGIS PRO İŞ AKIŞI

Büyük eğitim-doğrulama farkı aşırı uyuma işaret eder

Training Data: Classification Diagnostics				
Category	F1-Score	MCC	Sensitivity	Accuracy
bad	1.00	1.00	1.00	1.00
good	1.00	1.00	1.00	1.00
All	1.00	1.00	NA	1.00
*Predictions for the data used to train the model compared to the observed categories for those features				
Validation Data: Classification Diagnostics				
Category	F1-Score	MCC	Sensitivity	Accuracy
bad	0.82	0.68	0.83	0.84
good	0.86	0.68	0.85	0.84
All	0.84	0.68	NA	0.84
*Predictions for the test data (excluded from model training) compared to the observed values for those test features				

Görsel okuma: önce girdi/çıktıyı, sonra renk ve sembollerin neyi temsil ettiğini belirleyin.

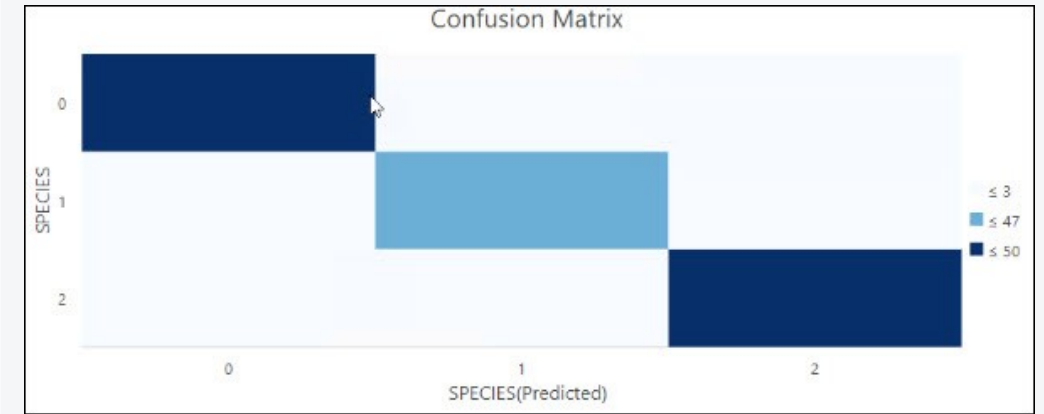
- Regresyonda R^2 , MAE, RMSE ve hata dağılımı incelenir.
- Sınıflandırmada sınıf bazlı F1, MCC, duyarlılık ve doğruluk raporlanır.
- Rastgele dışlama tek başına coğrafi aktarımı ölçmez.
- Aynı model mekânsal bloklarla ayrıca sınanmalıdır.

Karışıklık matrisi

ARCGIS PRO İŞ AKIŞI

Hangi sınıfların birbirine karıştığı karar maliyetini gösterir

- Satırlar gözlenen, sütunlar tahmin edilen sınıfları temsil eder.
- Diyagonal doğru sınıflandırmaları gösterir.
- Azınlık sınıfında yüksek hata, toplam accuracy içinde görünmeyebilir.
- Matrisi eğitim, mekânsal test ve farklı bölgeler için ayrı üretin.



Görsel okuma: önce girdi/çıktıyı, sonra renk ve sembollerin neyi temsil ettiğini belirleyin.

BÖLÜM 4

Doğrulama, yorum ve raporlama

Model skoru harita üzerindeki hata yapısıyla birlikte okunur

Bu bölümde kavram, yöntem ve uygulama bağlantısı kurulacaktır.

Regresyon ölçütleri

DOĞRULAMA, YORUM VE RAPORLAMA

MAE tipik hatayı, RMSE büyük hatalara duyarlılığı gösterir

- MAE hedef değişkenin birimindedir ve kolay yorumlanır.
- RMSE büyük hataları daha fazla cezalandırır.
- R^2 açıklanan varyansı özetler ancak ek değişkenlerle artabilir.
- Farklı hedef ölçeklerindeki modeller MAE ile doğrudan karşılaştırılmaz.

ANLATIM ŞEMASI

1 MAE hedef değişkenin birimindedir ve kolay yo...



2 RMSE büyük hataları daha fazla cezalandırır



3 R^2 açıklanan varyansı özetler ancak ek de...

Hata haritası

DOĞRULAMA, YORUM VE RAPORLAMA

Artıkların mekânsal örüntüsü eksik süreç veya yanlışlık gösterebilir

- Pozitif artık modelin düşük tahmin ettiği alanı gösterir.
- Negatif artık aşırı tahmini gösterir.
- Artıklarda mekânsal otokorelasyon kalması modelin coğrafi yapıyı tam yakalamadığını düşündürür.
- Bölgesel hata özeti karar verene ortalama skordan daha fazla bilgi verebilir.

ANLATIM ŞEMASI

01

Pozitif artık
modelin düşük
tahmin ettiği
ala...

02

Negatif artık
aşırı tahmini
gösterir

03

Artıklarda
mekânsal
otokorelasyon
kalması mod...

04

Bölgesel hata
özeti karar
verene ortalama
sko...

Rastgele ve mekânsal test

DOĞRULAMA, YORUM VE RAPORLAMA

İki skor arasındaki fark transfer riskinin göstergesidir

- Rastgele test mevcut örnek dağılımına yakın performansı ölçer.
- Mekânsal test yeni bölgedeki performansı daha iyi taklit eder.
- Fark büyükse model yerel örüntüleri ezberlemiş olabilir.
- Kullanım alanı mevcut coğrafyayla sınırlıysa bu sınır model kartında belirtilir.

ANLATIM ŞEMASI

Rastgele test mevcut örnek dağılımına yakın p...

Mekânsal test yeni bölgedeki performansı daha...

Fark büyükse model yerel örüntüleri ezberlemi...

Kullanım alanı mevcut coğrafyayla sınırlıysa...

genelden uygulamaya

Belirsizlik ve eşik

DOĞRULAMA, YORUM VE RAPORLAMA

Tahmin olasılığı karar eşiğiyle birlikte sunulur

- Varsayılan 0,50 eşiği her karar problemi için uygun değildir.
- Yanlış negatif maliyeti yüksekse daha duyarlı eşik seçilebilir.
- Belirsiz tahminler ek saha kontrolüne yönlendirilebilir.
- Olasılık kalibrasyonu coğrafi test verisi üzerinde incelenmelidir.

ANLATIM ŞEMASI

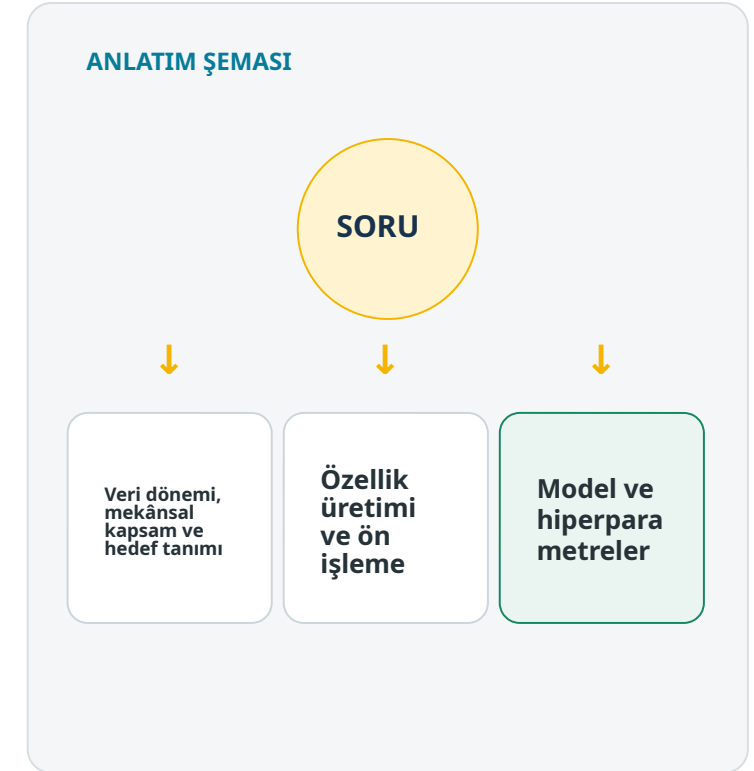


Denetimli model raporu

DOĞRULAMA, YORUM VE RAPORLAMA

Tekrarlanabilir rapor veri, model, doğrulama ve sınırlılığı birlikte gösterir

- Veri dönemi, mekânsal kapsam ve hedef tanımı
- Özellik üretimi ve ön işleme
- Model ve hiperparametreler
- Mekânsal bölme stratejisi
- Ölçütler, hata haritası ve sınıf bazlı sonuçlar
- Kullanım sınırı ve güncelleme planı



YÖNLENDİRİLMİŞ UYGULAMA

Mekânsal denetimli öğrenme laboratuvarı

Bir kamu hizmetine olan talebi veya risk puanını Forest-based Regression/Classification ile modelleyin.

Laboratuvar senaryosu

UYGULAMA

Bir kamu hizmetine olan talebi veya risk puanını Forest-based Regression/Classification ile modelleyin.

- Çalışma alanı ve analiz birimini belirtin.
- Karar kullanıcısını ve beklenen çıktıyı tanımlayın.
- Veri, yöntem ve doğrulama arasında tutarlı bir zincir kurun.

Laboratuvar işlem adımları

UYGULAMA

- Hedef ve açıklayıcı alanları kontrol edin; hedef sızıntısı oluşturabilecek alanları çıkarın.
- Mesafe özelliği ve en az iki açıklayıcı raster ekleyin.
- Train only modunda temel modeli kurun ve değişken önemini inceleyin.
- Bir coğrafi bölgeyi test için dışarıda tutarak ikinci değerlendirmeyi yapın.
- Tahmin, hata ve belirsizlik haritalarını aynı düzen içinde sunun.

Her adım için ekran görüntüsü veya işlem geçmişi kanıtı saklanmalıdır.

Laboratuvar kalite kontrolü

UYGULAMA

- Girdi verisinin tarihi, koordinat sistemi ve çözünürlüğü kaydedildi mi?
- Eğitim ve test verisi coğrafi veya zamansal olarak bağımsız mı?
- Araç parametreleri ve random seed kaydedildi mi?
- Sonuç nicel ölçüt ve harita üzerinden doğrulandı mı?
- Sınırlılık ve kullanım sınırı açıklandı mı?

Laboratuvar teslim paketi

UYGULAMA

Araç parametreleri, performans tablosu, değişken önemi ve mekânsal hata haritası

- Harita veya model çıktısı
- Veri ve yöntem açıklaması
- Doğrulama kanıtı
- Hata veya belirsizlik yorumu
- Kaynaklar ve kısa kullanım sınırı

Ders sonu kontrolü 1

DEĞERLENDİRME

Gradient Boosting modelinde sonraki ağaç hangi bilgiye odaklanır?

- A Koordinat sistemine
- B Önceki modelin hatalarına**
- C Yalnız çoğunluk sınıfına
- D Harita renklerine

Ders sonu kontrolü 2

DEĞERLENDİRME

Değişken öneminin doğru yorumu hangisidir?

- A Nedensel etkiyi kanıtlar
- B Değişkenin model bölmelerine katkısını gösterir
- C Her bölgede aynı ilişkiyi garanti eder
- D Ölçüm hatasını ortadan kaldırır

Ders sonu kontrolü 3

DEĞERLENDİRME

Yeni bir coğrafyada performansı ölçmek için hangisi daha uygundur?

- A** Eğitim doğruluğu
- B** Rastgele yüzde 10 ayırma
- C** Coğrafi bölgeyi dışarıda bırakma
- D** Daha fazla renk sınıfı

Denetimli Mekânsal Makine Öğrenmesi

- Denetimli öğrenme problemi
- Mekânsal özellikler ve eğitim
- ArcGIS Pro iş akışı
- Doğrulama, yorum ve raporlama

Ana ilke

**Coğrafi veri yapısı,
doğrulama düzeni ve
kullanım bağlamı
model seçiminden
önce gelir.**

Kaynaklar ve ileri okuma

- <https://doc.esri.com/en/arcgis-pro/latest/tool-reference/spatial-statistics/how-forest-works.html>
- <https://doi.org/10.1111/ecog.02881>
- https://www.fatiando.org/verde/latest/gallery/train_test_split.html

Sunumdaki her dış görselin kaynağı ilgili slaydın alt bilgisinde ve konuşmacı notlarında belirtilmiştir.