

Veri Bilimi Uygulamaları - Final Sınavı

9.12.2025 - Galatasaray Üniversitesi - Süre: 55dk

- Git versiyon kontrol sisteminde, yerel bilgisayarınızdaki değişiklikleri uzak sunucuya (örneğin GitHub üzerine) göndermek için hangi komut kullanılır?**
 - (a) `git pull`
 - (b) `git commit`
 - (c) `git push`
 - (d) `git clone`
- Test Güdümlü Geliştirme (TDD) yaklaşımının temel döngüsü ("Red-Green-Refactor") aşağıdakilerden hangisidir?**
 - (a) Kodu yaz → Testi yaz → Hata ayıkla.
 - (b) Başarısız test yaz → Kodu yaz → İyileştir.
 - (c) Doküman yaz → Kodu yaz → Test et.
 - (d) Testi yaz → Testi sil → Kodu yaz.
- Veri setindeki eksik verileri (missing values) doldururken, aykırı değerlerden (outliers) en az etkilenen yöntem hangisidir?**
 - (a) Ortalama (Mean) ile doldurma
 - (b) Medyan (Median) ile doldurma
 - (c) Standart sapma ile doldurma
 - (d) Maksimum değer ile doldurma
- "Kırmızı", "Mavi", "Yeşil" gibi sıralı olmayan (nominal) bir kategorik değişkeni modele sokmak için en uygun kodlama yöntemi hangisidir?**
 - (a) Label Encoding
 - (b) One-Hot Encoding
 - (c) Ordinal Encoding
 - (d) Min-Max Scaling
- Veri sızıntısını iki sayısal veri dağılımının istatistiksel olarak farklı olup olmadığını test etmek için hangisi kullanılır?**
 - (a) T-Testi
 - (b) Ki-Kare Testi
 - (c) Kolmogorov-Smirnov (K-S) Testi
 - (d) ANOVA
- Veri sızıntısını önlemek için özellik ölçeklendirme (scaling) işlemi ne zaman yapılmalıdır?**
 - (a) Veriyi eğitim/test olarak ayırmadan önce.
 - (b) Sadece test verisi üzerinde.
 - (c) Veriyi ayırdıktan sonra, sadece eğitim verisine göre (fit) uygulanmalıdır.
 - (d) Özellik seçimi yapmadan önce.
- MLOps prensiplerine göre, MLflow kütüphanesinin temel kullanım amacı nedir?**
 - (a) Veri tabanlarını yönetmek.
 - (b) Deneyleri, parametreleri ve metrikleri izlemek (Experiment Tracking).
 - (c) Web arayüzü oluşturmak.
 - (d) Derin öğrenme modeli eğitmek.
- Sürekli Entegrasyon (CI) sürecinin ana hedefi nedir?**
 - (a) Modeli canlıya manuel taşımak.
 - (b) Kod değişikliklerini sık sık ve otomatik olarak birleştirip test etmek.
 - (c) Büyük verileri depolamak.
 - (d) Hiperparametre optimizasyonu.
- Üretim ortamındaki modelde, girdi verisinin dağılımının zamanla eğitim verisinden uzaklaşmasına ne ad verilir?**
 - (a) Model Overfitting
 - (b) Data Drift (Veri Kayması)
 - (c) Gradient Descent
 - (d) Feature Selection
- LLM'lerde metni (text) modelin işleyebileceği sayısal parçalara dönüştürme işlemine ne denir?**
 - (a) Embedding
 - (b) Tokenization
 - (c) Attention
 - (d) Inference
- Prompt Mühendisliğinde, modele görevi yapması için birkaç örnek (girdi-çıkı) verme tekniği nedir?**
 - (a) Zero-Shot Prompting
 - (b) Few-Shot Prompting
 - (c) Chain-of-Thought
 - (d) RAG
- API parametrelerinden temperature değeri 0.0'a yaklaştıkça modelin çıktısı nasıl değişir?**
 - (a) Daha yaratıcı ve rastgele olur.
 - (b) Daha deterministik, tutarlı ve odaklı olur.
 - (c) Yanıt süresi uzar.
 - (d) Kelime dağarcığı genişler.
- Bir LLM'i yeniden eğitmeden, ona şirkete özel güncel bilgilerle soru cevaplatmak için en uygun yöntem hangisidir?**
 - (a) Fine-Tuning
 - (b) RAG (Retrieval-Augmented Generation)
 - (c) Pre-training
 - (d) Transfer Learning
- Metinleri, kelimelerin anlamsal benzerliğini koruyacak şekilde vektörlere dönüştüren model türü hangisidir?**
 - (a) Tokenizer
 - (b) Embedding Modeli
 - (c) SQL
 - (d) Classifier
- Google Gemini veya OpenAI modellerinde "Function Calling" (Fonksiyon Çağırma) özelliğinin temel amacı nedir?**
 - (a) Modelin Python kodu yazmasını sağlamak.
 - (b) Modelin, harici araçlarla etkileşime girebilmesi için yapılandırılmış çıktı (JSON) üretmesi.
 - (c) Modelin matematik performansını artırmak.
 - (d) Sadece hava durumu tahmini yapmak.