

2024 – 2025 Bahar Dönemi YZM212 Makine Öğrenmesi Dersi I. Laboratuvar Değerlendirmesi (04.03.2025)

Ödevde naive Bayes yöntemi kullanarak ikili sınıflandırma yapılacaktır. Veri seti Kaggle, UCI Machine Learning Repository vb. platformlardan elde edilebilir. Uygulamada Scikit-learn'den bir naive Bayes sınıfı kullanılacaktır. Bir sınıf türü (Gaussian, Bernoulli vb.) seçilmelidir. Eğitilen model Scikit-learn kütüphanesi kullanılmadan Python sınıfı olarak tekrardan yazılmalıdır. Bu model de aynı train ve test setinden eğitilip test edilmelidir. Validasyon seti kullanımına gerek yoktur. Ödev aşağıdaki gereksinimleri içermelidir.

Gereksinimler

- Veri seti tabular olmalı, en az 5 özellik (feature) ve 1000 örnek içermelidir.
- NumPy, Pandas ve Matplotlib kullanılmalıdır.
- Scikit-learn ve Python ile birer model kullanarak karşılaştırma yapılmalıdır. Karşılaştırma iki açıdan olacaktır. Karmaşıklık matrisi kullanarak model tahminlerinin performansı analiz edilecektir. Ayrıca, modellerin eğitim (fit metodu) ve test (predict metodu) işlem zamanları time modülü kullanılarak test edilmelidir.
- Veri çalışması kısmında sınıf dağılımını, özellik veri türlerini (kategorik ve süreklilik açısından) ve eksik veriyi gözlemlemek yeterlidir. Veri ön işleme aşaması ve özellik mühendisliği (seçme, dönüştürme vb.) opsiyoneldir, eksik veri varsa düzeltilmelidir. Düzeltim özellik sürekli ise her özelliğin ortalaması, kategorik ise her özelliğin modu kullanılabilir. Ön işleme yapılacaksa bunun için ayrı bir dosya 1.naiveBayes dizinine eklenmelidir.
- Performans ölçümünde karmaşıklık matrisi görselleştirmeli, değerlendirme için kullanılan metrikler anlatılmalıdır. Değerlendirme metrikleri seçiminde problem ve sınıf dağılımı önemli midir? README.MD tartışma kısmında açıklanmalıdır.

Gönderim

Her öğrenci bu ders kapsamında bir GitHub deposu (repository) oluşturmalıdır. Bu depo dönem boyunca verilecek küçük projeleri kapsar. Depoya buna uygun bir isim verilmelidir. Örnek başlangıç dizini aşağıdaki gibidir. Lokalde bu dizinde depo başlatılmalıdır (git init).

1.naiveBayes (Klasör)

- xxnaiveBayesScikitLearn.ipynb (xx yerine kullanılan yöntem yazılabilir, Gaussian vb.)
(Dosya)

- naiveBayes.ipynb (Dosya)

- veri dosyası vb. varsa diğer dosyalar.

.gitignore (Dosya)

README.md (Dosya)

Requirements.txt (Dosya)

Jupyter Notebook ile çalışılacaksa (önerilen) dosyalar .py uzantılı olmayabilir. README.md dosyası problem tanımı, veri, yöntem, sonuçlar ve yorum / tartışma kısımları kısa ve net şekilde anlatılmalıdır. Gönderim son tarihi 10.03.2025 17.00 olup sınıftan seçilecek bir öğrenci tüm adresleri bu saatte yatikan@ankara.edu.tr adresine gönderecektir. 11.03.2025 13.30'da yazılı olarak bu kısa proje ve dersin teorik saatlerini kapsayan 1. laboratuvar değerlendirme sınavı yapılacaktır.

GitHub İin rnek Kaynaklar

<https://www.youtube.com/watch?v=f26KI43FK58>

<https://github.com/git-guides/#learning-git-basics>