# Uygulamalı Doğal Dil İşleme Vize Soruları

Kemal Can Kara

1. Hangisi bir metin sınıflandırma problemidir?

- A) Bir e-posta'nın spam veya spam olmayan olarak belirlenmesi

- B) Bir kitabın sayfa sayısını bulmak

- C) Makalelerde geçen tarihi olayları sıralamak

- D) Bir web sitesinin renk düzenini analiz etmek

2. Metin sınıflandırma tahmin çalışmasının ana amacı nedir?

- A) Metin içeriğini şifrelemek

- B) Metni rastgele kategorilere ayırmak

- C) Metnin bilinen sınıflardan hangisine ait olduğunu öğrenmek

- D) Metni alakasız metinlerle karşılaştırmak

3. Makine öğrenmesi sınıflandırması problemlerinde yaygın olarak hangi yazılım kütüphanesi kullanılır?

- A) Panda

- B) TenzihFlow

- C) Scikit-learn

- D) NLTG

4. Metin sınıflandırmasında 'eğitim seti' neyi ifade eder?

- A) Bir kitap koleksiyonu

- B) Sınıflandırma modelini test etmek için kullanılan veri

- C) Sınıflandırma modelini eğitmek için kullanılan veri

- D) Sınıflandırılmış harflerin bir listesi

5. Metin sınıflandırmasında bir 'feature (öznitelik - özellik)' nedir?

- A) Bir kelimenin uzunluğu

- B) Bir metinde geçen virgül sayısı

- C) Bir metinde geçen cümlelerin sayısı

- D) Bir metinde geçen belirli bir kelime

6. TF-IDF metin sınıflandırmasına nasıl katkıda bulunur?

- A) Bir metindeki nadir kelimelerin önemini vurgular

- B) Metindeki tüm kelimelerin frekansını toplar

- C) Bir metindeki kelimelerin sıklığını sayar

- D) Metnin kelime sayısını belirler

7. Denetimli ve denetimsiz metin sınıflandırması arasındaki fark nedir?

- A) Denetimsiz veri kümesi, denetimli veri kümesinden daha büyüktür

- B) Denetimli veri kümesi etiketli verilerle çalışırken, denetimsiz veri kümesi etiketlenmemiş verilerle çalışır

- C) Denetimli sınıflandırma algoritmaları, denetimsiz sınıflandırma algoritmalarından daha hızlıdır

- D) Her ikisi de aynı türde verilerle çalışır, ancak farklı özellik çıkarım yöntemleri kullanır

8. Metin sınıflandırmasında ön işleme işlemi hangi rolü oynar?

- A) Metni daha küçük parçalara böler

- B) Metin verilerini analiz için temizler ve hazırlar

- C) Metni farklı bir dile çevirir

- D) Metni rastgele sıralar

9. Metin sınıflandırmasında 'overfit (aşırı uyum)' kavramını açıklayın.

- A) Model, eğitim verilerinde iyi performans gösterir ancak yeni, görülmemiş verilerde kötü performans sergiler

- B) Model, eğitim verilerinde ve test verilerinde kötü performans gösterir

- C) Model, eğitim verilerine tam olarak uymaz ve hiçbir performans göstermez

- D) Model, eğitim verilerine çok iyi uyar ve test verilerinde de aynı performansı gösterir

10. Metin sınıflandırmada öznitelik çıkarma işleminin amacı nedir?

- A) Metni makine öğrenme algoritmaları için sayısal vektörlere dönüştürmek

- B) Metni analiz için görsellere dönüştürmek

- C) Metni depolama için ikili koda dönüştürmek

- D) Metni farklı bir dilde ifade etmek için çevirmek

11. 'Test seti' metin sınıflandırma modelini değerlendirirken ne işe yarar?

- A) Modeli yeni verilerle eğitmek için kullanılır

- B) Modelin daha önce görmediği veriler üzerinde performansını değerlendirmek için kullanılır

- C) Modelin ayarlanması için kullanılan eğitim verisinin bir alt kümesidir

- D) Modelin performansını ölçmek için kullanılan veri setidir

12. Metin sınıflandırma modellerini değerlendirirken cross validation (çapraz doğrulama) kavramını açıklayın.

- A) Bu, veri setini birden çok alt kümeye bölen ve modeli bu alt kümelerin farklı kombinasyonlarında eğiten ve test eden bir tekniktir.

- B) Bu, modelin doğruluğunu ölçmek için veri setini bir kez bölen bir tekniktir.

- C) Bu, modelin sadece eğitim verilerine dayalı olarak doğruluğunu ölçen bir tekniktir.

- D) Bu, veri setindeki tüm örnekleri kullanarak bir modelin performansını ölçen bir tekniktir.

13. Metin sınıflandırma modellerini değerlendirirken confusion matrix (karışıklık matrisi) neyi temsil eder?

- A) Veri setindeki farklı metin kategorilerinin dağılımını gösterir.

- B) Modelin farklı sınıflar için tahminlerinin doğruluğunu gösterir.

- C) Modelin performansını grafiksel olarak gösterir.

- D) Modelin son kullanma tarihi gösterir.

14. Metin sınıflandırma modelinde ne zaman aşırı uyum (overfitting) oluşabilir ve değerlendirme sırasında nasıl ele alınabilir?

- A) Aşırı uyum, bir modelin eğitim verilerini ezberlemesi veya gürültüyü öğrenmesi ve yeni verilerde kötü performans göstermesi durumunda oluşabilir. Bu, değerlendirme sırasında düzenlileştirme veya daha basit modeller kullanarak ele alınabilir.

- B) Aşırı uyum, modelin eğitim ve test verilerinde eşit derecede iyi performans göstermesi durumunda oluşur ve değerlendirme sırasında ele alınması gerekmez.

- C) Aşırı uyum, modelin metinlerin inceliklerini yakalayacak kadar basit olması durumunda oluşur ve bu, değerlendirme sırasında daha karmaşık modeller kullanarak ele alınabilir.

- D) Aşırı uyum, metin sınıflandırma modellerinde nadiren görülen bir durumdur ve değerlendirme sırasında özel bir dikkat gerektirmez.

15. Bir metin sınıflandırma modelinin performansını nasıl etkileyebileceğini açıklayın.

- A) Önyargılı verilerle eğitilen modeller, odaklı doğası nedeniyle daha iyi performans gösterir

- B) Eğitim verilerindeki önyargı, modelin haksız veya önyargılı kararlar vermesine neden olabilir ve tahminlerini etkileyebilir

- C) Önyargı, sadece metin sınıflandırma modellerinin hesaplama verimliliğini etkiler

- D) Önyargı, metin sınıflandırma modellerinin performansını etkilemez

16. Doğal dil işlemede word embedding’lerin temel amacı nedir?

- A) Kelimeleri daha iyi anlamak için görsellere dönüştürmek

- B) Kelimeleri sürekli bir vektör uzayında yoğun vektörlere dönüştürmek

- C) Görselleştirme amaçları için kelime bulutları oluşturmak

- D) Metin verilerinden fonetik özellikler çıkarmak

17. Word2Vec kavramını Word embedding içinde açıklayın.

- A) Word2Vec, kelime gömmelerini 3 boyutlu uzayda görselleştirmek için kullanılan bir yöntemdir.

- B) Word2Vec, büyük bir metin koleksiyonundaki bağlamına dayalı olarak kelimelerin dağıtık temsillerini öğrenen bir tekniktir.

- C) Word2Vec, insanın kendine yakışanı giymesidir.

- D) Word2Vec, doğal dil işlemede duygusal analiz için özel olarak kullanılan bir tekniktir.

18. Word2Vec, kelimeler arasındaki anlamsal ilişkileri nasıl yakalar?

- A) Kelimeleri vektör uzayında eş anlamlılarına eşler

- B) Cümlelerdeki çevresel bağlama dayalı olarak kelime temsilleri öğrenerek

- C) Kelimelere ve anlamlarına rastgele vektörler atayarak

- D) Farklı dillerdeki kelimelerin etimolojisini analiz ederek

19. Word2Vec modelleri neden genellikle büyük metin veri kümeleriyle eğitilir?

- A) Modelin daha hızlı eğitilmesini sağlamak için

- B) Modelin daha düşük boyutlu vektör uzayında işlem yapmasını sağlamak için

- C) Modelin daha az bellek kullanarak çalışmasını sağlamak için

- D) Kelimeler arasındaki bağlamın daha iyi anlaşılmasını sağlamak için

20. Kelime gömme yöntemleri neden sıklıkla doğal dil işleme uygulamalarında kullanılır?

- A) Kelimelerin vektörel temsilleri sayesinde semantik ilişkileri ve anlamları daha iyi ifade edebilmek için

- B) Sadece metin verilerini basitleştirmek için

- C) İşlemci gücünden tasarruf etmek için

- D) Kelimelerin uzunluğunu daha kısa hale getirmek için