# Uygulamalı Doğal Dil İşleme Vize Soruları

Kemal Can Kara – 21.11.2023

1. Hangisi bir metin sınıflandırma problemidir?

- A) Bir twitter mesajını olumlu/olumsuz tanımlamak

- B) İş ilanı metinlerinin cümle sayılarını bulmak

- C) Paragrafları sıralamak

- D) Dilbilgisi hatalarını bulmak

2. Metin sınıflandırmasının ana amacı nedir?

- A) Metin içeriğini özetlemek

- B) Metni önceden belirlenmiş sınıflara veya kategorilere ayırmak

- C) Metni birden fazla dile çevirmek

- D) Var olan paragraflardan yeni metin üretmek

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

- A) MidTerm Maker

- B) TextRanger

- C) Scikit-learn

- D) ILoveNLp 3.0

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

- A) Bir dizi metin mesajı

- B) Sınıflandırma modelini eğitmek için kullanılan veri

- C) Sınıflandırılmış kitapların bir listesi

- D) Metin yazı tiplerinin bir koleksiyonu

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

- A) Sınıflandırmada kullanılan belirli bir kelime veya karakter

- B) Bir noktalama işareti türü

- C) Bir yazı tipi stili

- D) Bir paragraf yapısı türü

6. TF-IDF metin sınıflandırmasına nasıl katkıda bulunur?

- A) Bir dokumandaki kelimelerin sayısını belirler

- B) Bir dokumandaki kelimelerin toplam frekansını sayar

- C) Metnin yazımını ve dilbilgisini kontrol eder

- D) Bir dokuman içindeki kelimelerin, dokuman koleksiyonuna göre önemini ölçer

7. Denetimli ve denetimsiz metin sınıflandırması arasındaki fark nedir?

- A) Denetimli etiketli veri kullanır, denetimsiz etiketli veri gerektirmez

- B) Denetimsiz etiketli veri kullanır, denetimli etiketli veri gerektirmez

- C) Her ikisi de etiketli veri gerektirir, ancak sınıflandırma algoritmaları farklıdır

- D) Her ikisi de etiketli veri gerektirmez, ancak sınıflandırma algoritmaları farklıdır

8. Metin sınıflandırmasında ön işleme işlemi hangi rolü oynar?

- A) Metin verilerini analiz için temizler ve hazırlar

- B) Çeşitlilik için metne yeni kelimeler ekler

- C) Metni şifreler ve güvenliği sağlar

- D) Metni daha hızlı sınıflandırma için özetler

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

- A) Aşırı uyum, metni önceden belirlenmiş kategorilere doğru şekilde sığdırma sürecini ifade eder

- B) Bu, yetersiz eğitim verisi nedeniyle bir modelin performansının kötü olması durumunu tanımlar

- C) Aşırı uyum sadece bir sınıflandırma modelinde uzun metinlerin yanlış sınıflandırılması anlamına gelir

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

10. Metin sınıflandırmada öznitelik çıkarma işleminin amacı nedir?

- A) Metni makine öğrenme algoritmaları için sayısal vektörlere dönüştürmek

- B) Metni sınıflandırma amaçları için resimlere dönüştürmek

- C) Metni analiz için ses dosyalarına dönüştürmek

- D) Metni depolama için binary (ikili) koda dönüştürmek

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) Sınıflandırmaya uygun olmayan belgelerin bir koleksiyonudur

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, ayrı ayrı eğitim ve test için veri setinin kopyalarını oluşturan bir tekniktir.

- C) Bu, modelin yalnızca farklı dillerden metin verilerini doğrulaması için bir yöntemdir.

- D) Bu, sınıflandırma sonuçlarını grafiksel bir formatta görselleştirmek için 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) Karışıklık yapan konularda kafamızı dağıtmaya yarar

- D) Veri akışını sınıflandırma modelinden görselleştirir.

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

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

- B) 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.

- 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) Önyargı, metin sınıflandırma modellerinin performansını etkilemez

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

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

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

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, insanın kendine yakışanı giymesidir.

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

- 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) Farklı dillerdeki kelimelerin etimolojisini analiz ederek

- C) Cümlelerdeki çevresel bağlama dayalı olarak kelime temsilleri öğrenerek

- D) Kelimelere ve anlamlarına rastgele vektörler atayarak

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

- A) Modelin daha hızlı eğitilmesini sağlamak için

- B) Modelin daha az bellek kullanarak çalışmasını sağlamak için

- C) Kelimeler arasındaki bağlamın daha iyi anlaşılmasını sağlamak için

- D) Modelin daha düşük boyutlu vektör uzayında işlem yapmasını 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

1-A

2-B

3-C

4-B

5-A

6-D

7-A

8-A

9-D

10-A

11-B

12-A

13-B

14-B

15-D

16-B

17-C

18-C

19-C

20-A