



Makine Öğrenmesi ile Twitter Duygu-Durum Analizi

SentimentyBot

SentimentyBot Datascientist Bootcamp 10. Dönem katılımcılarından olan Batuhan Özyurt, Zeynep Beyza Ayman, İlkay Çil, Betül Yıldız ve Nur Fatma Atıcı tarafından oluşturulmuştur. Projenin amacı kullanıcıların; mutluluğu, başarıları, dinlediği müzikleri, izlediği filmleri paylaştığı sosyal medya platformlarında kederini, acısını, yalnızlığını dile getiren on binlerce insan olduğu farkındalığını yaratmaktır. Twitter API'ları kullanılarak 2022 yılına ait Türkçe Tweetler çekildikten sonra siyasi kişiler, kamu kurum ve kuruluşları, son dakika haberleri gibi kullanıcıların duygu-durumlarına ilişkin olmayan içerikler veri setinin dışında bırakılarak bireylerin yalnızca duygularını ve tweet atmadan önce Twitter'ın sorduğu 'Whats happening?' sorusuna yanıt veren tweetler ile ilerlenmiştir. Tweetler kelime kırılımında değil, taşıdıkları duygu bağlamında manuel etiketlenmiştir. Logistic Regression kullanılan bu projede, güncel olarak canlı tweetler üzerinden duygu-durum analizi yapabilen ve hemen sonrasında model tahminine göre negatif etiketlediği tweetlerin sahipleriyle Twitter'da mention üzerinden iletişime geçip kullanıcılara anlık moral, motivasyon sağlayabilecek playlistler öneren bir bot geliştirilmiştir.

Case study, SentimentyBot baz alınarak, bootcamp mezunları Batuhan, Zeynep Beyza, İlkay, Betül ve Nur Fatma tarafından hazırlanmıştır.

Twitter kullanıcıları tarafından atılan tweetlerin taşıdıkları duygu kapsamında pozitif, negatif ve nötr olarak tahmin edilmesi.



‘tweet_labeled.csv’ Twitter 2022 yılında atılan tweetleri, tweetlerin atıldığı tarihleri ve tweetlerin içerdiği duygu kapsamında -1, 0 ve 1 olacak şekilde etiketleri içermektedir. ‘tweets_21.csv’ 2021 yılına ait tweetleri kapsamaktadır.

tweets_labeled.csv

4 Değişken

12.690 Gözlem

506.4 KB

tweet_id	Tweet’in id bilgisi
tweet	Tweet içeriği
date	Tweetin atıldığı tarih ve saat bilgisi
label	Tweetin duygusuna göre el ile girilmiş etiket bilgisi -1 : negatif 1 : pozitif 0 : nötr

tweets_21.csv

3 Değişken

13.272 Gözlem

311.2 KB

tweet_id	Tweet’in id bilgisi
tweet	Tweet içeriği
date	Tweetin atıldığı tarih ve saat bilgisi



Proje Görevleri

Görev 1: Feature Engineering

Adım 1: ‘labeled_tweets.csv’ adlı dosyayı okutunuz.

Adım 2: Eksik değer analizi yapınız, eksik değerler için gerekli işlemleri uygulayınız.

Adım 3: ‘date’ değişkeni saat dilimi GMT+03:00 olacak şekilde güncelleyiniz

Adım 4: ‘date’ değişkenini baz alarak aşağıdaki 3 değişkeni oluşturunuz:

- mevsim
- gün
- 4 saatlik periyotlar (örneğin: 22:00-02:00, 02:00-06:00, 06:00-10:00, 10:00-14:00, 18:00-22:00)

Adım 5: ‘label’ değişkenini aşağıdaki gibi düzenleyiniz

- 1: pozitif
- 1: negatif
- 0: nötr

Adım 6: numerik ve kategorik değişkenleri yakalayınız.

Adım 7: hedef değişken analizi yapınız.

Görev 2: Veriyi Hazırlama ve Logistic Regression

Adım 1: ‘tweet’ değişkenin harflerinin küçük harfe çeviriniz.

Adım 2: label encoding işlemi uygulayarak ‘label’ değişkenini sayısal olarak ifade ediniz

Adım 3: ‘tweet’ değişkenini kullanarak TF-IDF matrisini oluşturunuz.

Adım 4: TF-IDF matrisini ve ‘label’ hedef değişkenini kullanarak logistic regression modelini kurunuz, accuracy’yi değerlendiriniz.

Adım 5: ‘tweets_21.csv’ adlı dosyayı okutunuz, değişken tiplerini düzenleyiniz.

Adım 6: tweet’leri küçük harfe çeviriniz.

Görev 3: Tweetlerde Duygu Tahmini

Adım 1: ‘tf_idfvectorizer’ ve logistic regression modelini kullanarak 2021 yılına ait tweetler için tahmin işlemini gerçekleştiriniz.

Adım 2: ‘label’ adlı bir değişken oluşturunuz ve tahminleri bu değişkenin altına kaydediniz.

Adım 3: işlemlerinizi fonksiyonlaştırınız ve pipeline dosyasını oluşturunuz.

miuul

miuul.com