



Ders: Metin Madenciliği(Text Mining)

Ödev: Twitter Verileri İle Duygu Analizi

170255063 Rami KOCO

Duygu Analizi Nedir ?

- Duygu Analizi veya diğer bir adı ile 'Fikir Madenciliği' (Opinion Mining) tam anlamıyla birlikte; bir yazarın veya bir yazının belirli bir konuya yönelik tutumunu olumlu(positive), olumsuz(negative), tarafsız(neutral) olarak değerlendirilmesi için bir metin parçasıyla belirtilen görüşleri/ifadeleri çeşitli algoritmalar aracılığı ile hesaplayıp tanımlama ve sınıflandırma süreçlerine verilen genel bir tanımdır.
- İşlenen veriler kısa metinlerden oluştuğu için duygu analizi de bir metin sınıflandırma problemi olarak ele alınmaktadır.

Duygu Analizi Çeşitleri

- ❑ Öznellik/Nesnellik Analizi:
- ❑ Hedef Tabanlı (Aspect Based) Duygu Analizi
- ❑ Çok Tonlu (Fine-grained) Duygu Analizi:
- ❑ Duygu Belirleme (Emotion Detection)
- ❑ Genel Duygu Analizi

Öznellik/Nesnellik Analizi:

- Bir ifadenin özne bir yargı barındırıp barındırmadığı duygu analizi için ilk sorulması gereken sorudur.
- Keza bir duygu tespiti yapabilmek için ifadenin özne bir yargı barındırıyor olması gerekmektedir.
- Çünkü duygu anlam itibarıyla özne bir ifade şeklindedir.
- İfadelerin içerdiği duygunun türünden ziyade ifadenin bir duygu içerip içermediği bilgisine odaklanır.

Hedef Tabanlı(Aspect Based) Duygu Analizi:

- Hedef tabanlı duygu analizi, sınıflandırmayı bir adım daha öteye götürerek belirli özelliklere veya konulara duygu atamasını gerçekleştirir.
- Yazılı ifadeleri konu/duygu bütünlüğünü dikkate alarak küçük parçalara bölerek verilerden daha ayrıntılı ve doğru bilgi elde edilmesini sağlar.
- Kullanıcı odaklı olmak ve kullanıcı deneyimini en üst seviyede sunmak istiyorsanız hedef tabanlı duygu analizi en iyi yardımcınız olacaktır.
- **Örneğin:** Aracın kısa sürede teslim edilmesinden memnun kaldım ama fren disklerini de unutmasanız iyi olurdu.
- Bu cümle içerisinde birden fazla duygu ve konu yer almaktadır.
- Bu nedenle tüm cümle tek bir ifade olarak ele alınırsa baskın sınıf olarak negatif ya da pozitif seçilebilir ve bilgi kaybı yaşanabilir.



Hedef tabanlı duygu analizi ile bu cümleden aşağıdaki bilgiler elde edilebilir.

- **Metin:** Aracın kısa sürede teslim edilmesinden memnun kaldım
Konu: Teslimat **Duygu:** Olumlu
- **Metin:** ama fren disklerini de unutmasanız iyi olurdu
Konu: Yedek Parça **Duygu:** Olumsuz

Çok Tonlu (Fine-grained) Duygu Analizi:

- Anketlerdeki ve sosyal medyadaki yorumlar da dahil olmak üzere herhangi bir metindeki olumlu, olumsuz, tarafsız kutupları tanımlar.
- Ayrıca doküman bazında veya hedef tabanlı duyguları da ayıklar.
- Bunu yapmak için, metindeki farklı cümlelerin yerel polaritesi tanımlanır, aralarındaki ilişki değerlendirilir ve tüm metin için küresel bir polarite değeri ile sonuçlanır.
- Buna göre bir veriyi olumlu sınıfına atarken daha kuvvetli bir olumlu yorumu “çok olumlu” sınıfına atayabilirsiniz.
- Böylelikle olumlu yorumları da olumluluk seviyesine göre bölerek en uç örnekleri daha net olarak yakalayabilirsiniz.

Duygu Belirleme (Emotion Detection):

- Bu tür duygu analizi standart olumlu, olumsuz, nötr sınıflarını değil daha detaylı olarak hissedilen duyguyu araştırır.
- Mutluluk, hayal kırıklığı, öfke, üzüntü gibi duygu sınıfları kullanır.
- Duygu belirleme özelliği kullanılan sistemlerde bu duygu sınıfları sayesinde kullanıcının tutumu olumlu veya olumsuz olarak incelenmekle birlikte, ruh hali daha net bir şekilde analiz edilerek uygun aksiyonlar alınabilir.

Duygu Analizin Kullanım Alanları

- Finans, tıp, borsacılık, medya, politika gibi bir çok alanda kullanılmaktadır.
- Ürün incelemelerinde, tüketicilerin bir ürün ile alakalı memnuniyet ölçümü yapılabilmesi adına yaygın olarak kullanılan bir tekniktir.
- Aynı şekilde, bir şirket, yeni çıkaracağı bir ürünün reklam kampanyasının müşterilerinin sosyal medyadaki verdiği yanıtın etkisini ölçmek için oldukça popüler olan ve yaygın olarak kullanılan bir yol haritasına sahiptir.
- Politikacılar seçim zamanlarında kampanyalarına halkın nasıl bir tepki vereceğini de ilgili ölçütleri kullanarak geri bildirim alabilmektedir.

Duygu Analizinin Zorlukları

- Duygu analizinin temeli büyük veri kümelerinin kullanılmasına dayanır.
- Duygu analizi sistemini kullanmak ve eyleme geçirebilmek için büyük veri kaynağına ve hacmine sahip olmak gerekir.
- Sistemi canlı tutmak için yeterli veri sayısı olmadığında veya sistem güncel tutulmadığında modelin hatalı değerlendirme yapması kaçınılmazdır.
- Metin parçasında duygu analizini nesne, öznitelik , görüş sahibi, yönelim ve görüş gücü unsurları etkiler.



Uygulanan Adımlar

- Twitterdan gelen veriler ön işlemden geçirilir. Bu kısımda yinelenen veriler, tutarsız, gürültü verileri temizlenir.
- Steaming işlemi yapılır.
- Stopwordlerin Çıkarılması
- Veriler test ve eğitim verisi olmak üzere ikiye ayrılır.
- TF-IDF terim ağırlıklandırma yöntemi kullanılır.
- Destek Vektör Makineleri (SVM) kullanılarak sınıflandırma işlemi yapılır. Olumlu, olumsuz ve nötr tweetler tespit edilir.

Ön İşleme Süreci

- Veri seti 3 duygu etiketinden oluşmaktadır.
- Bu etiketler nötr, negatif ve pozitiftir.
- Buna göre, nötr ifadeler 0 ile, negatifler 1 ile ve pozitifler 2 ile etiketlenmiştir.
- Bu çalışmamızda 2399 adet negatif, 800 adet pozitif, 800 adet nötr kategoride verimiz bulunuyor.
- Uygulamadan önce için verileri hazırlamak çok önemlidir.
- Bu işleme veri ön işleme denir.
- Ön işlemde eksik, tutarsız, gürültülü veriler temizlenir.
- Edatlar ,bağlaçlar veri setimizden çıkartılır.
- Kendimizin oluşturduğu anlam ifade etmeyen kelimeleri (ama, ve, ile,...)stopTr.txt dosyamızdan aldık.



Stemming İşlemi

- Stemming: (Kök alma) algoritmaları, taban veya kök biçimine indirgenme işlemidir.
- Biz bu algoritmalarından Türkçe için kullanılan TurkishStemmer kütüphanesini kullandık.

Test Ve Eğitim Verisi Oluşturma

- Verimizi test ve eğitim verisi olarak ayırdık. (x train, x test, y train, y test)
- Test verisi boyutunu %20 olarak ayarladık.
- 3200 veri eğitim, 800 veri test olarak ayarlandı.

TF-IDF (Term Frequency - Inverse Document Frequency)

- Bu çalışmamızda kelimeleri TF-IDF (Term Frequency - Inverse Document Frequency) terim ağırlıklandırma yöntemini kullandık.
- TF-IDF, verideki bir kelimenin kuvvetini, o kelimeyi içeren belge sayısına göre hesaplayan bir ağırlıklandırma yöntemidir.
- Verilerdeki önemsiz değerler elimine edilir, önemli özellikler belirlenir ve sınıflandırma işlemi sırasında performansta artış sağlanır.
- Bir kullanıcının attığı tweetler incelenerek kişinin nelere ilgi duyduğu, yazılarında hangi konulara ağırlık verdiğini çıkarabiliriz.

Destek Vektör Makineleri(Support Vector Machine)

- genellikle sınıflandırma problemlerinde kullanılan gözetimli öğrenme yöntemlerinden biridir.
- Bir düzlem üzerine yerleştirilmiş noktaları ayırmak için bir doğru çizer.
- Bu doğrunun, iki sınıfının noktaları için de maksimum uzaklıkta olmasını amaçlar.
- Karmaşık ama küçük ve orta ölçekteki veri setleri için uygundur.
- Destek Vektör Makineleri (SVM), düzlem üzerindeki noktaların bir doğru veya hiper düzlem ile ayrıştırılması ve sınıflandırılmasıdır.

Karışıklık Matrisi(Confusion Matrix)

- Son olarak verilerimize confusion matrix uyguladık. Makine öğrenmesinde kullanılan sınıflandırma modellerinin performansını değerlendirmek için hedef niteliğe ait tahminlerin ve gerçek değerlerin karşılaştırıldığı hata matrisi sıklıkla kullanılmaktadır.
- Her ne olursa olsun sınıflandırma tahminleri şu dört değerlendirmeden birine sahip olacaktır:
 1. Doğruya doğru demek (True Positive – TP) DOĞRU
 2. Yanlışta yanlış demek (True Negative – TN) DOĞRU
 3. Doğruya yanlış demek (False Positive – FP) YANLIŞ
 4. Yanlışta doğru demek(False Negative – FN) YANLIŞ

Karşılaştırma

Text(89.18, 0.5, 'Tahmin Değerler')

Tahmin Değerler	0	1	2
0	156	0	0
1	7	472	19
2	2	4	140
		Gerçek Değerler	

**BERT Sentiment Analysis
Confusion Matrix**

Test	Negative	Neutral	Positive
Negative	1481	35	113
Neutral	87	462	65
Positive	113	22	1409
		Predicted	



Bizi dinlediğiniz için teşekkür ederiz.