

GSM OPERATÖRLERİNE YÖNELİK ATILAN TÜRKÇE TWEETLERİN DERİN ÖĞRENME YÖNTEMLERİYLE DUYGU ANALİZİ

Biröl EMEKLİ
Sakarya Üniversitesi
Fen Bilimleri Enstitüsü
Bilişim Sistemleri Mühendisliği

SUNU AKIŞI



1. Giriş ve Amaç
 - Duygu analizi nedir
 - Çalışmanın amacı
2. Materyal ve Metod
 - Veri setinin elde edilmesi
 - Derin öğrenme modelleri
3. Sonuçlar
 - Elde edilen sonuçlar
 - Başarılı modelin incelenmesi

Duygu Analizi

- Duygu analizi, bir yazarın belirli bir konu yada hedef hakkındaki, yazılı fikir ve düşüncelerini olumlu, olumsuz ve nötr olarak sınıflandırmayı amaçlamaktadır.
- Gelişmekte olan teknoloji ile sosyal medya platformlarının fazlaca kullanılması, kişilerin duygu ve düşüncelerini sosyal medya platformlarda paylaşımlarına imkan sağlamıştır. Bu paylaşımların çok fazla olması ise duygu analizi yapabilmek için manuel olarak takip etmeyi imkansız hale getirmiştir.

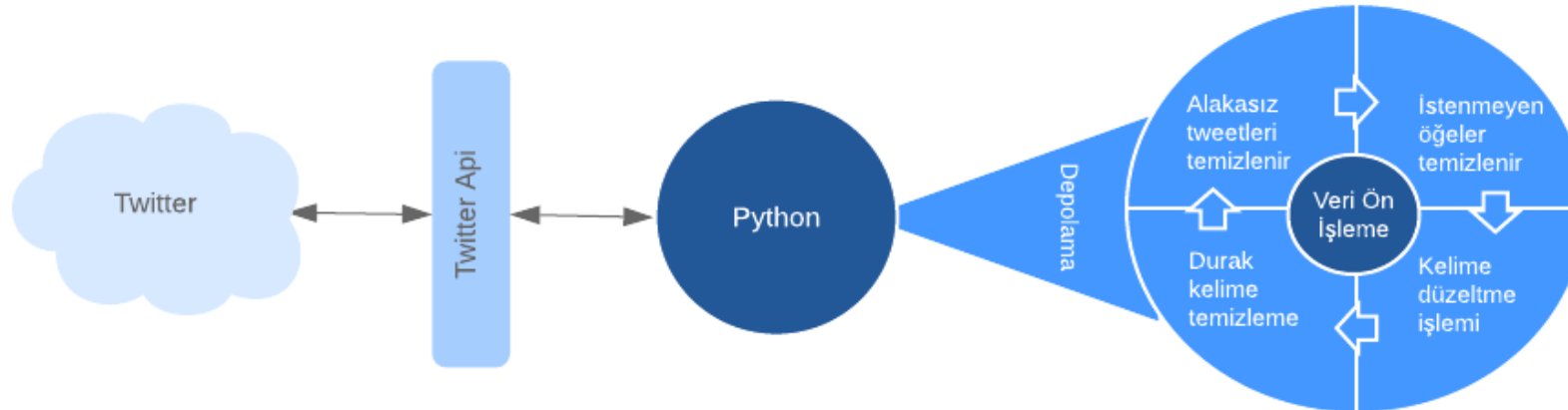


Amaç

Gelişen teknoloji ve kullanıcıların sosyal medya kullanımındaki artış gibi etkenler göz önüne alınarak, **derin öğrenme yöntemleri** ile duygu sınıflandırılmasının daha kolay ve hızlı bir şekilde yapılabilmesi sağlayan bir model geliştirmeyi amaçladık

Veri Seti

- Çalışma kapsamında duygu analizi çalışması yapılacak ortam Twitter, görüşler ise GSM operatörlerine yönelik tweetleri içermektedir.
- Veri seti elde edebilmek için Python Tweepy kütüphanesi ve Twitter API'den faydalanılmıştır.
- Ham olarak 32000 tweet toplanmış ve ön işlemlerden sonra modellenebilir, duygu içeren 1675 adet tweet elde edilmiştir.



Veri Seti Ön İşlem Aşamaları

Elde edilen ham veri seti şu aşamalardan geçirilerek anlamlı hale getirilmiştir

1. Tüm veri seti küçük harfe çevrilir
2. Tweet içerisindeki gereksiz öğeler temizlenir(@,#,https://, rakamlar,boşluk)
3. Kelime düzeltme işlemine tabi tutulur
4. Durak kelimeler temizlenir
5. Veri setini kaydetme

Tablo 1. Veri Seti

Operatör	Olumlu	Olumsuz	Toplam
Türkcell	102	515	617
Vodafone	112	665	777
Türk Telekom	41	240	281
			1675

Veri Ön İşleme Örneği

@vodafone 10 gün oldu NEDEN internt SoRunumu bir türlü çözmdi #VodaFone <https://t.co/osUgGrOJS>

@vodafone 10 gün oldu neden internt sorunumu bir türlü çözmdi #vodafone <https://t.co/osuggrojs>

@vodafone 10 gün oldu neden internt sorunumu bir türlü çözmdi #vodafone

10 gün oldu neden internt sorunumu bir türlü çözmdi

gün oldu neden internt sorunumu bir türlü çözmdi

gün oldu neden internt sorunumu bir türlü çözmdi

gün oldu neden internet sorunumu bir türlü çözmedi

gün internet sorunumu türlü çözmedi

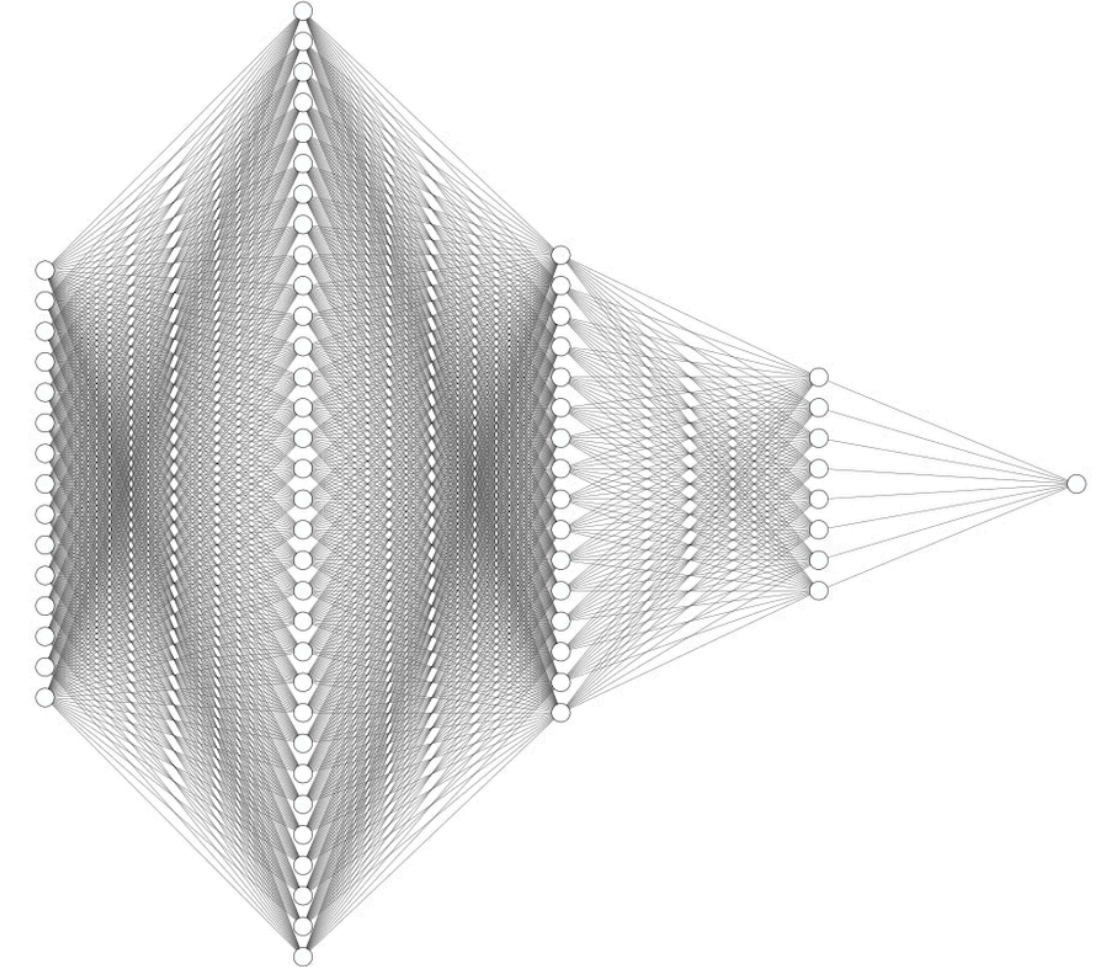
Derin Öğrenme Modeli Tasarımı

Duygu analizinde kullanılacak olan üç farklı derin öğrenme modelini test ettik

1. Evrişimli Sinir Ağları(CNN)
2. Tekrarlayan Sinir Ağları(RNN)
3. Uzun Kısa Süreli Bellek(LSTM)

Tablo 2. Derin Öğrenme Modeli

Tip	Hücre
Embedding	15
CNN,RNN,LSTM	32
CNN,RNN,LSTM	16
CNN,RNN,LSTM	8
Dense	1



Sonuçlar

- Veri setini aynı model üzerinde farklı derin öğrenme yöntemleri ile eğitilip test ve doğruluk değerleri ile karşılaştırılmıştır.
- Model doğruluk ve test oranları karşılaştırıldığında Uzun Kısa Süreli Bellek modelinin doğruluk oranının daha başarılı olduğu gözükmemektedir.

Türkcell	Evrişimli Sinir Ağları		Tekrarlayan Sinir Ağları		Uzun Kısa Süreli Bellek	
Epoch	6	9	6	9	6	9
Doğruluk %	83.91	83.91	90.66	96.54	93.77	98.64
Test %	76.32	76.32	84.21	89.47	84.21	94.74

Sayı

617

Vodafone	Evrişimli Sinir Ağları		Tekrarlayan Sinir Ağları		Uzun Kısa Süreli Bellek	
Epoch	6	9	6	9	6	9
Doğruluk %	85.87	85.87	94.24	96.84	95.06	97.81
Test %	80.85	80.85	80.36	85.11	91.49	87.23

Sayı

777

Türk Telekom	Evrişimli Sinir Ağları		Tekrarlayan Sinir Ağları		Uzun Kısa Süreli Bellek	
Epoch	6	9	6	9	6	9
Doğruluk %	85.66	85.66	92.03	98.80	85.66	90.44
Test %	82.76	82.76	79.31	89.66	82.76	86.21

Sayı

281

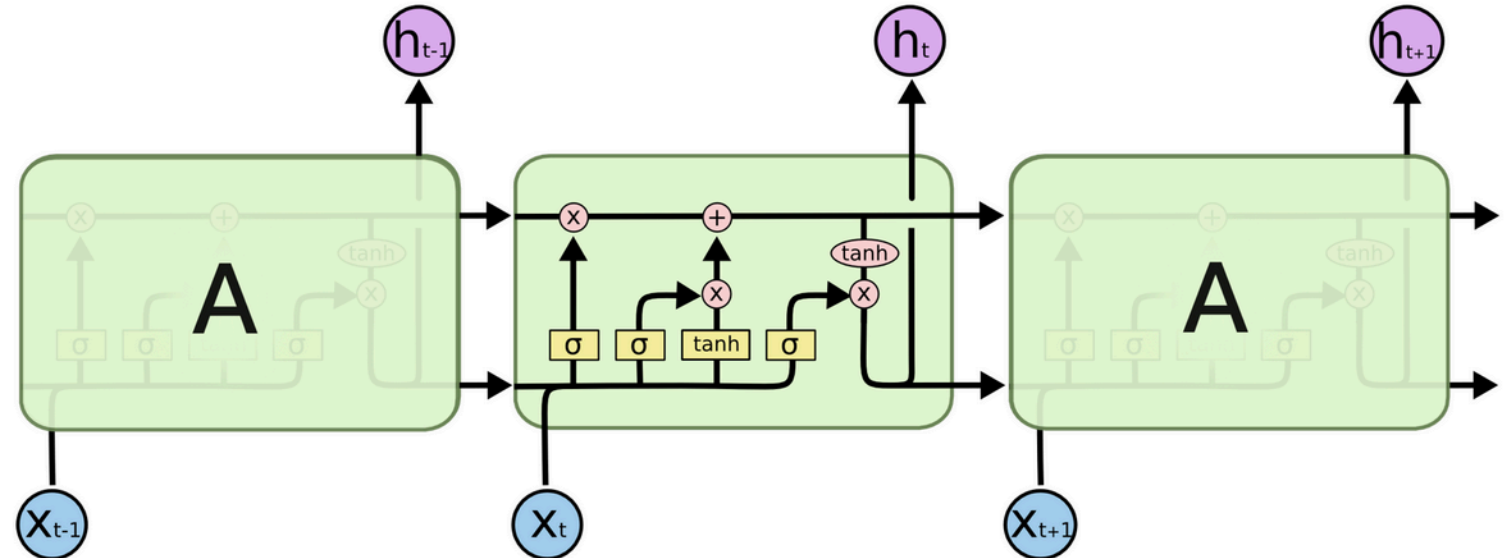
Tüm Veri Seti	Evrişimli Sinir Ağları		Tekrarlayan Sinir Ağları		Uzun Kısa Süreli Bellek	
Epoch	6	9	6	9	6	9
Doğruluk %	85.28	90.01	92.53	94.75	92.09	96.08
Test %	82.74	86.31	81.55	83.93	84.52	85.71

Sayı

1675

Uzun Kısa Süreli Bellek

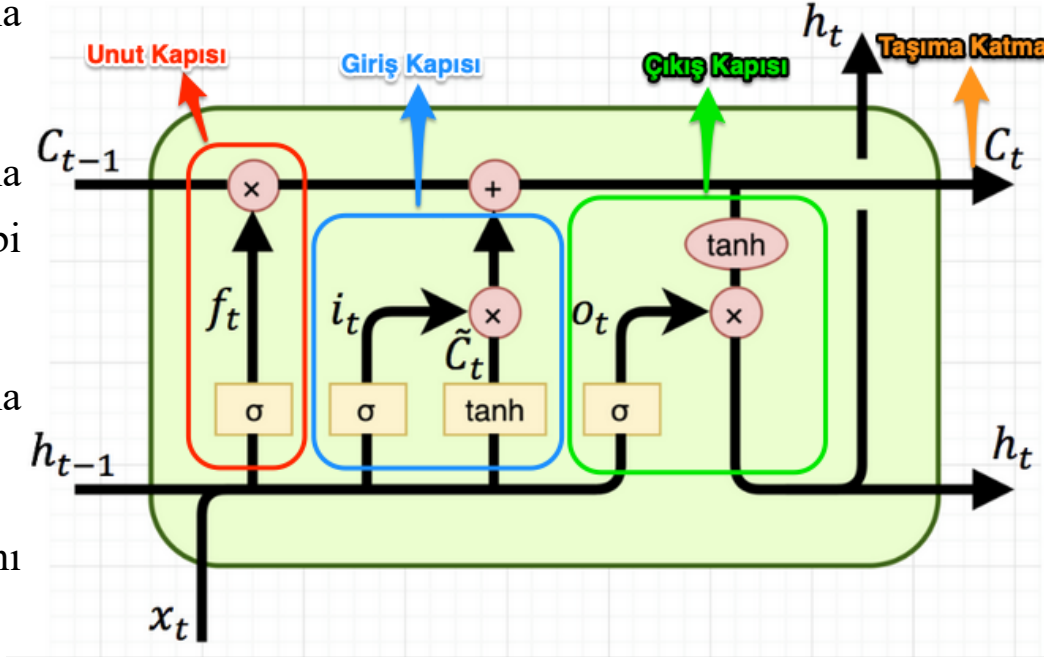
- Uzun Kısa Süreli Bellek, önceki durumları taşıma katmanını kullanarak sürekli eğitimde etkisini sürdürmeyi amaçlamıştır. Uzun süreli bellek hafızası nedeniyle zaman serisi, anomali tespiti, konuşma tanıma, Doğal Dil İşleme gibi konularda Tekrarlayan Sinir Ağlarına göre daha başarılı sonuç ortaya koymaktadır.



Forget Gate(Unut Kapısı): Hücredeki bilgilerin tutulup tutulmayacağına karar veren bölümdür. h_{t-1} (önceki hücre çıktısı) ve X_t (yeni girdi) girdileri vektörel çarptıktan sonra sigmoid fonksiyonuna sokulur. Sonuç 0 ile 1 arasında bir değer alır. Eğer sonuç 1 ise bilgiyi tamamen tut, 0 ise tamamen unut anlamına gelmektedir. Daha sonra taşıma hücresindeki değer ile çarpılır.

Input Gate(Giriş Kapısı): Hücreye yeni bilgi eklenmesinden sorumludur ve üç aşamadan oluşur. Bu aşama sonrasında unutulması gereken veriler taşıma hücresine dahil edilmemiş olur.

1. h_{t-1} (önceki hücre çıktısı) ve X_t (yeni girdi) önce sigmoid fonksiyonuna sokulur, hangi değerlerin güncelleneceğine karar verilir. Unutma kapısı gibi çalışır.
2. h_{t-1} (önceki hücre çıktısı) ve X_t (yeni girdi) değerleri tanh fonksiyonuna sokularak -1 ile +1 arasında yeni bir vektör oluşturulur.
3. 1. ve 2. adımda oluşturulan çıktılar vektörel olarak çarpılarak taşıma katmanı hücresine eklenir.



Output Gate(Çıkış Kapısı): Üç aşamadan oluşur ve girdileri anlamlı hale getirdikten sonra çıktı elde edilen katmandır.

1. Taşıma hücresindeki değer tanh fonksiyonuna sokularak -1 ile +1 arasında bir vektör oluşturulur.
2. Daha sonra h_{t-1} (önceki hücre çıktısı) ve X_t (yeni girdi) sigmoid fonksiyonu ile ölçeklendirilir.
3. 1. ve 2. adımda oluşturulan çıktılar vektörel olarak çarpıldıktan sonra çıktı ve bir sonraki hücreye girdi olarak gönderilir.

Katılım ve Katkılarınız için Teşekkür Ederiz.

Birol EMEKLİ

Sakarya Üniversitesi

Fen Bilimleri Enstitüsü

Bilişim Sistemleri Mühendisliği

Doc. Dr. İhsan Hakan SELVİ

Sakarya Üniversitesi

Bilgisayar ve Bilişim Bilimleri Fakültesi

Bilişim Sistemleri Mühendisliği Bölümü