

<> Türkiye[®]
Açık Kaynak
Platformu</>

Turkey Open
Source Platform
www.turkeyopensourceplatform.com

ACIKHACK
Açık Kaynak Hackathon Programı



Türkiye Açık Kaynak Platformu
Online Yarışma Programı
**Türkçe Doğal Dil
İşleme**



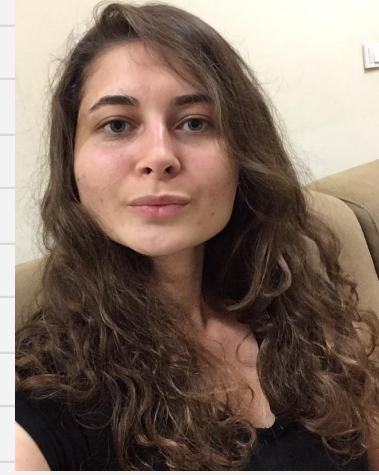
Yapay Zeka Tabanlı Türkçe Yazım
Destekleyicisi



Ümit Yılmaz



3. Sınıf Bilgisayar
Mühendisliği Öğrencisi /
Adesso Turkey'de
Part-time Data Scientist

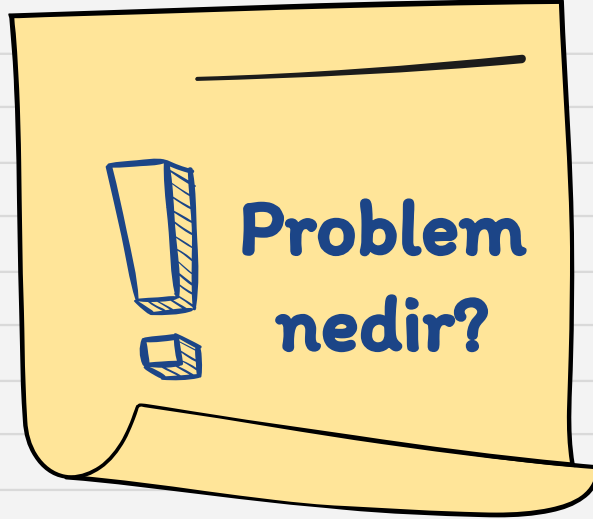


Büşra Gökmen



3. Sınıf Bilgisayar
Mühendisliği Öğrencisi /
Akbank'ta Machine
Learning Intern





Her ne kadar yazım denetimi üzerine birçok kütüphane olsa da içeriği anlamsal olarak değerlendiren ve türkçe yazım kurallarına göre denetleyen ve bu değerlendirme ve denetimi gelişmiş derin öğrenme tekniklerini kullanarak gerçekleştirebilen bir kütüphaneye ihtiyaç duyulması



ÇÖZÜMLERİMİZ!



Anlamsal Analiz

Negatif-Pozitif Analiz

Formal-Informal Analiz

Pozitif

Kötü yorumlar gözümü korkutmuştu ancak hiçbir sorun yaşamadım teşekkürler.

Negatif

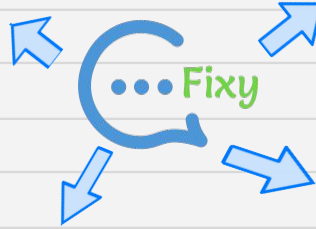
Tasarımı güzel ancak ürün açılmış tavsiye edemem.

Bu koşullar da toplumsal değişim taleplerinin olgunlaşmasına yol açmıştır.

Formal

Ben de öyle olduğunu düşünmüyorum ama gittik yine de

Informal



Emotion(Duygu) Analizi



Korku

İşlerin ters gitmesinden endişe duyuyorum.

Sevinç

Bugün çok mutluyum.



Tremodata

Duygu Verisi

Yaklaşık 6,7 milyon satır veri

Wikipedia + Newspaper

Formal Veri

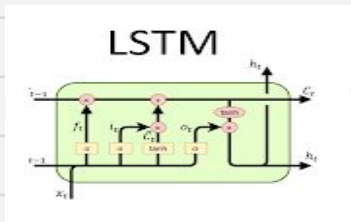
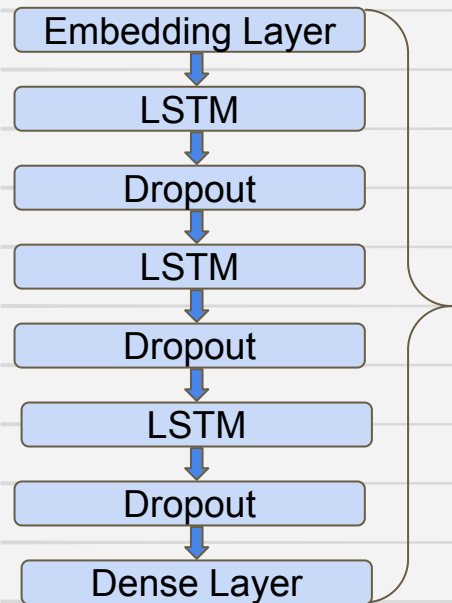
Twitter

Informal Veri

Ürün Yorumları+Twitter

Sentiment Verisi

Anlamsal Analiz Modeli



Model Sonuçları

Model	Accuracy
LinearSVC	%80
Logistic Regression	%79
Multinomial NB	%78
Random Forest Classifier	%60

Negatif-Pozitif Analiz

- Accuracy: %94.57

Formal-Informal Analiz

- Accuracy: %95.37



Yaklaşım

Seni ben **de** anlıyorum ama Fixy kadar değil.
Tüm detaylar defter**de** yazıyor.
Kırtasiyeden defter **de** alacak mısın?
Öyle berbat bir cümleydi **ki** Fixy bile düzeltemedi.
Bu durumda sen **de** **mi** geleceksin.

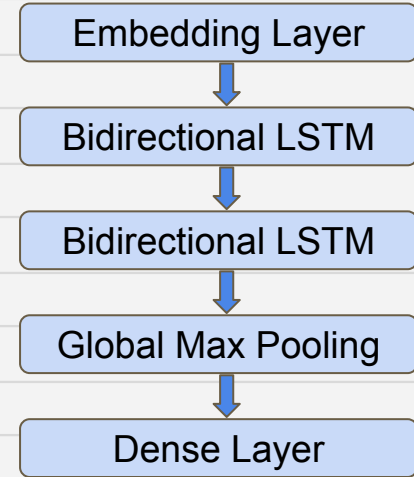


seni ben X anlıyorum ama fixy kadar değil .	0
tüm detaylar defter X yazıyor .	1
kırtasiyeden defter X alacak mısın ?	0
öyle berbat cümleydi X fixy bile düzeltemedi .	0
bu durumda sen de X geleceksin .	1

Etiket

3 önceki kelime

3 sonraki kelime





Kullanılan Veri

Çalışmalarımız sonucunda 85 milyondan fazla cümle kullanılarak -de, -ki ve -mi için 3 farklı veri seti ön işlendi, etiketlendi, github ve drive üzerinden community ile paylaşıldı

ORPUS



ISCORPUS



Alt Yazı

Wikipedia

Gazete

INFORMAL

FORMAL

DeepChecker Kütüphanesi



`pip install DeepChecker`

Doğrudan `check()` ve `correct()` fonksiyonlarını kütüphaneden çağırıp -de, -mi- ve -ki ek mi bağlaç mı öğrenip düzelttirebilirsiniz!

Kolay Kullanımı!

```
from DeepChecker import correct_de
correct_de("akşam oda gelecek mi?")
```

Performans Sonuçlar

-de,-da

- Accuracy on Test Data: 92.13%
- ROC AUC on Test Data: 0.921

class	precision	recall	f1-score
0	0.9049	0.9397	0.9219
1	0.9384	0.9030	0.9204

-ki

- Accuracy on Test Data: 91.32%
- ROC AUC on Test Data: 0.913

class	precision	recall	f1-score
0	0.9323	0.8912	0.9113
1	0.8958	0.9353	0.9151

-mi

- Accuracy on Test Data: 95.41%
- ROC AUC on Test Data: 0.954

class	precision	recall	f1-score
0	0.9509	0.9575	0.9542
1	0.9572	0.9506	0.9539

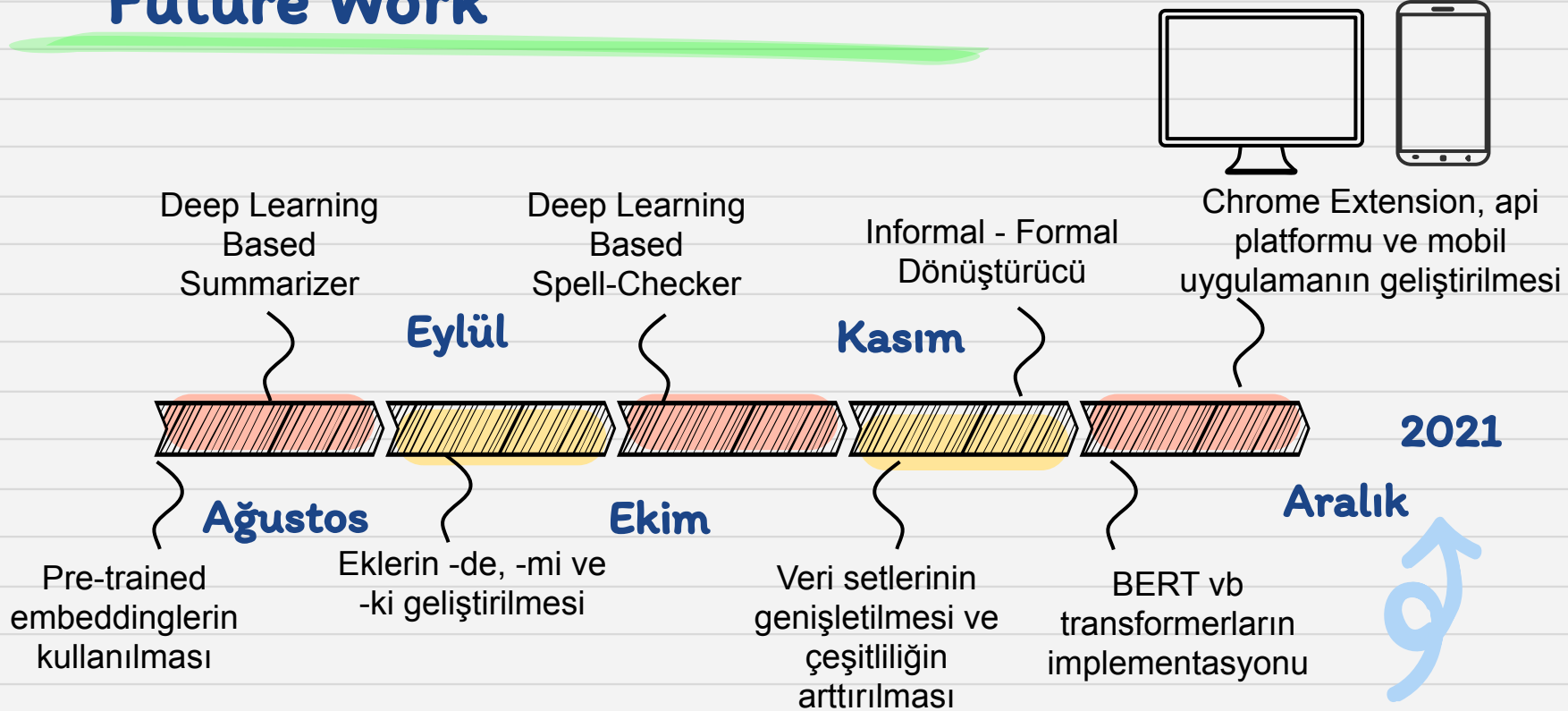
Literatür Karşılaştırması

Yapılan Çalışmalar	Doğruluk Oranı	F1 Score
Fixy	%92.13	%92.23
Boğaziçi	%76.48	%86.67
Google Docs	%34	%--
Microsoft Office	%29	%--
ITU	%0	%--
Libra Office	%0	%--

Boğaziçi tarafından oluşturulan zor cümleler test datasında, Boğaziçinin çalışması %78 doğruluk oranı verirken, Fixy %87'ye ulaşarak literatürdeki en başarılı çalışma olmuştur.



Future Work



COMMUNITY TEPKİSİ

Unwatch ▾

14

★ Unstar

102

Fork

6



Selçuk Usta @ustasoglu · 7 Ağu

@hmert ve @YavuzKomecoglu adlı kullanıcılara yanıt olarak

Sentiment analizi için fastText üzerine çalışmalarım vardı, sağladıkları veri seti şahane oldu. Emekçilerine büyük bir teşekkür.

[github.com/selcukusta/sen...](https://github.com/selcukusta/sentiment-analysis-for-comments) adresinde, veri setinin fastText'e uygun olarak etiketlenmiş hali mevcut, ilgilenenlere duyurulur 😊



selcukusta/sentiment-analysis-for-comments
(DENEYSEL) E-ticaret sitelerindeki ürün yorumlarından duygu analizi gerçekleştiren uygulama
github.com



3



Hüseyin Mert @hmert · 7 Ağu

Github'ta Türkçe diline yönelik makine öğrenmesi konusunda çok güzel çalışmaya denk geldim. Beraber destek verelim gelin.

github.com/Fixy-TR/fixy

(@YavuzKomecoglu senden gördüm varol)



Fixy-TR/fixy

Amacımız Türkçe NLP literatüründeki birçok farklı sorunu bir arada çözebilen, eşsiz yaklaşımlar öne ...
github.com

3

23

168



Fixy-TR / fixy

<> Code

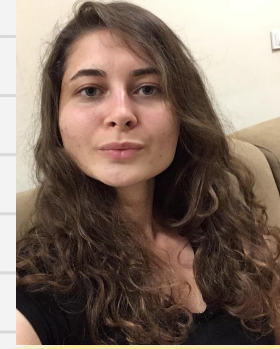
! Issues

🔗 Pull requests 1

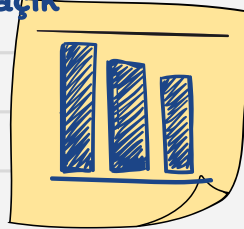
Ekip üyelerinin projeye katkıları



- -de -da, -ki, mi ile ilgili anlamsal hata düzeltici fikri, geliştirilmesi, literatür araştırması ve implementasyonu
- DeepChecker kütüphanesinin implementasyonu ve pypi ile yayınlanması
- İlgili veri setlerinin bulunması, işlenmesi, etiketlenmesi ve açık kaynak olarak sunulması



- Emotion(Duygu) Analizi modelinin geliştirilmesi
- Pozitif-Negatif ve Formal-Informal modellerinin araştırılması ve geliştirilmesi
- İlgili veri setlerinin bulunması, işlenmesi, etiketlenmesi ve açık kaynak olarak sunulması
- Backend-Frontend implementasyonu





Teşekkürler!

