Zorbalık Tweetleri Tespit Etme

Bert-Distilbert- Roberta- Xlnet-Albert-

Sana Kabbani *Bilişim Sistemleri Mühendisliği  
Kocaeli Üniversitesi*  
181307074

*Özetç:* twitterden veri çekip , 5 pre-trained transformer modelleri zorballık tweetleri tespit etme öğretilmiştir.

Anahtar Kelimeler: Transformer , Tweet , Zorballık , Training Bert

# Giriş

Zorbalık tweetleri tespiti, sosyal medya platformlarındaki zorbalık davranışlarını tespit etmek amacıyla kullanılan bir teknoloji.

kullanıcıları korumak ve olumsuz etkileri azaltmak için önemli bir hedef haline gelmiştir.

# Proje Kapsamı

.

## A. Veri Toplama

Selenyuim kütüphanesi kullanılarak 12000 tweet çakilmiştir.

B. *Veri Ön İşleme*

Toplanan veri kümesi, gereken ön işleme adımlarından geçirilir. Bu adımlar arasında metin temizleme, duygu analizi uygualama. 2ye olarak label yapılmış [0/nonbullying-1/bullying]

C. *Model Eğitimi*

Önceden eğitilmiş bir Transformer modeli (örneğin, BERT veya GPT) kullanarak bir sınıflandırma modeli oluşturulur. Model, veri kümesinin bir bölümüyle eğitilir ve zorbalık tweetlerini tespit etme yeteneğini öğrenir. Eğitim sürecinde, kayıp fonksiyonu ve optimize edici kullanılır.

D. *Model Doğrulama*

Eğitim süreci tamamlandıktan sonra, modelin performansını değerlendirmek için ayrı bir doğrulama veri kümesi kullanılır.

E. *Model Değerlendirme:*

Modelin performansı, metrikler ve performans değerlendirilir.

F. *Uygulama ve Entegrasyon :*

Eğitilmiş model, sosyal medya platformlarına veya diğer ilgili sistemlere entegre edilebilir. Bu sayede, kullanıcıların zorbalık içeren tweetlere karşı korunması ve gerekli önlemlerin alınması sağlanır.

# Metrikler

epoch bazında her bir model için değerlendirme metrik sonuçları aaşağda görünmektedir.

1. sayı Non Bullying için

2.sayı Bullying için

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Epochs | prec | recall | F1-score | accuracy | Validation loss |
| 1 | 0.96  0.94 | 0.94 0.96 | 0.95  0.95 | 1230/1314  1383/1436 | 0.15 |
| 2 | 0.94  0.97 | 0.97  0.94 | 0.95  0.96 | 1269/1314  1357/1436 | 0.16 |
| 3 | 0.96  0.97 | 0.97  0.97 | 0.97  0.97 | 1275/1314  1388/1436 | 0.15 |
| 4 | 0.98  0.97 | 0.97  0.98 | 0.97  0.97 | 1269/1314  1404/1436 | 0.1731 |
| 5 | 0.97  0.97 | 0.97  0.97 | 0.97  0.97 | 1271/1314  1398/1436 | 0.201 |
| 6 | 0.98  0.96 | 0.96  0.98 | 0.97  0.97 | 1258/1314  1406/1436 | 0.243 |
| 7 | 0.97  0.97 | 0.97  0.98 | 0.97  0.97 | 1271/1314  1401/1436 | 0.234 |
| 8 | 0.97  0.97 | 0.97  0.98 | 0.97  0.97 | 1271/1314  1402/1436 | 0.239 |
| 9 | 0.97  0.97 | 0.96  0.98 | 0.97  0.97 | 1267/1314  1402/1436 | 0.252 |
| 10 | 0.98  0.96 | 0.96  0.98 | 0.97  0.97 | 1262/1314  1407/1436 | 0.263 |

*Tablo 1 Distilbert Model*

,

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Epochs | prec | recall | F1-score | accuracy | Validation loss |
| 1 | 0.95  0.96 | 0.96 0.95 | 0.95  0.95 | 1258/1314  1363/1436 | 0.179 |
| 2 | 0.97  0.95 | 0.94  0.97 | 0.96  0.96 | 1236/1314  1400/1436 | 0.133 |
| 3 | 0.96  0.97 | 0.97  0.96 | 0.96  0.96 | 1270/1314  1379/1436 | 0.180 |
| 4 | 0.98  0.95 | 0.94  0.98 | 0.96  0.97 | 1240/1314  1413/1436 | 0.185 |
| 5 | 0.98  0.95 | 0.94  0.98 | 0.96  0.97 | 1236/1314  1412/1436 | 0.231 |
| 6 | 0.98  0.96 | 0.95  0.98 | 0.97  0.97 | 1252/1314  1409/1436 | 0.206 |
| 7 | 0.98  0.95 | 0.95  0.98 | 0.96  0.97 | 1246/1314  1413/1436 | 0.217 |
| 8 | 0.98  0.96 | 0.95  0.99 | 0.97  0.97 | 1250/1314  1416/1436 | 0.213 |
| 9 | 0.98  0.96 | 0.96  0.98 | 0.97  0.97 | 1256/1314  1411/1436 | 0.227 |
| 10 | 0.98  0.96 | 0.96  0.98 | 0.97  0.97 | 1256/1314  1411/1436 | 0.226 |

*Tablo 2 Roberta Model*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Epochs | prec | recall | F1-score | accuracy | Validation loss |
| 1 | 0.95  0.96 | 0.96 0.96 | 0.95  0.9 | 1256/1314  1373/1436 | 0.14 |
| 2 | 0.97  0.96 | 0.95  0.98 | 0.96  0.97 | 1248/1314  1403/1436 | 0.156 |
| 3 | 0.98  0.94 | 0.93  0.99 | 0.96  0.96 | 1275/1314  1388/1436 | 0.186 |
| 4 | 0.98  0.97 | 0.97  0.98 | 0.97  0.97 | 1222/1314  1416/1436 | 0.1731 |
| 5 | 0.98  0.96 | 0.96  0.98 | 0.97  0.97 | 1261/1314  1409/1436 | 0.246 |
| 6 | 0.98  0.96 | 0.95  0.98 | 0.97  0.97 | 1249/1314  1414/1436 | 0.256 |
| 7 | 0.98  0.97 | 0.96  0.98 | 0.97  0.98 | 1268/1314  1411/1436 | 0.219 |
| 8 | 0.98  0.95 | 0.97  0.98 | 0.97  0.97 | 1276/1314  1410/1436 | 0.22 |
| 9 | 0.97  0.96 | 0.96  0.98 | 0.97  0.97 | 1225/1314  1416/1436 | 0.22 |
| 10 | 0.98  0.96 | 0.96  0.98 | 0.97  0.97 | 1243/1314  1409/1436 | 0.21 |

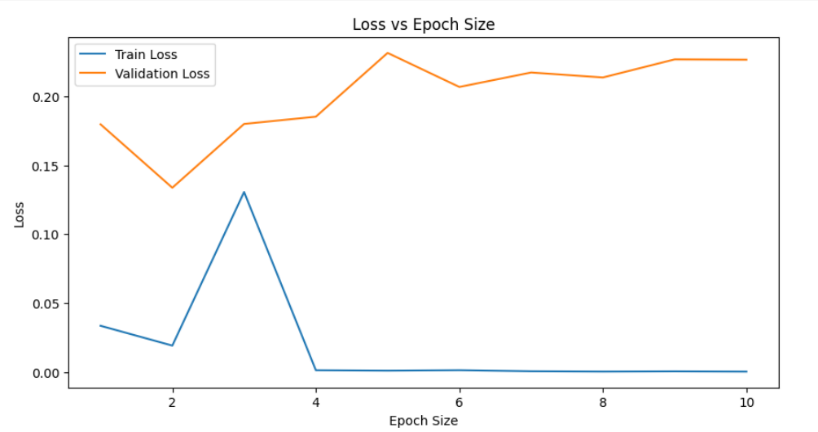
*Tablo 3 Bert Model*

# özet tablo

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| model | prec | recall | F1-score | accuracy |
| roberta | 0.98  0.95 | 0.95  0.98 | 0.96  0.97 | 0.97 |
| distilbert | 0.98  0.96 | 0.96  0.98 | 0.97  0.97 | 0.97 |
| bert | 0.98  0.96 | 0.96  0.98 | 0.97  0.97 | 0.97 |

Tablo 4 Özet Tablo

# vı. traın valıdatıon loss graphlri



*Roberta Modelin Train-validation loss Graph*

A picture containing text, plot, screenshot, line

Description automatically generated

*Distilbert Modelin Train-validation loss Graph*

A picture containing line, plot, diagram, text

Description automatically generated

*Bert Modelin Train-validation loss Graph*

VII. Modellerin Çalışma Zamanların

|  |  |  |
| --- | --- | --- |
| model | Train çalışma zamanı | Test çalışma zamanı |
| roberta | 639 s | 677s |
| distilbert | 310 s | 330 s |
| bert | 368 s | 390 s |

Tablo 5 Modellerin çalışma zamanı

Proje Kodları ve Veri KÜMESİ :

https://drive.google.com/drive/folders/1zQTGhMwmBCp78WdmsXC5Lr2oca3wC6z6?usp=sharing

# Referanslar

1. <https://www.kaggle.com/code/ludovicocuoghi/detecting-bullying-tweets-pytorch-lstm-bert>
2. https://www.coursera.org/projects/sentiment-analysis-bert