

# Türkiye Açık Kaynak Platformu Online Yarışma Programı

# Türkçe Doğal Dil İşleme

[www.turkiyeacikkaynakplatformu.com](http://www.turkiyeacikkaynakplatformu.com)



-Takım Adı-

# Hive Heroes



# Ekip Üyeleri ve Katkıları

## Berfin Duman

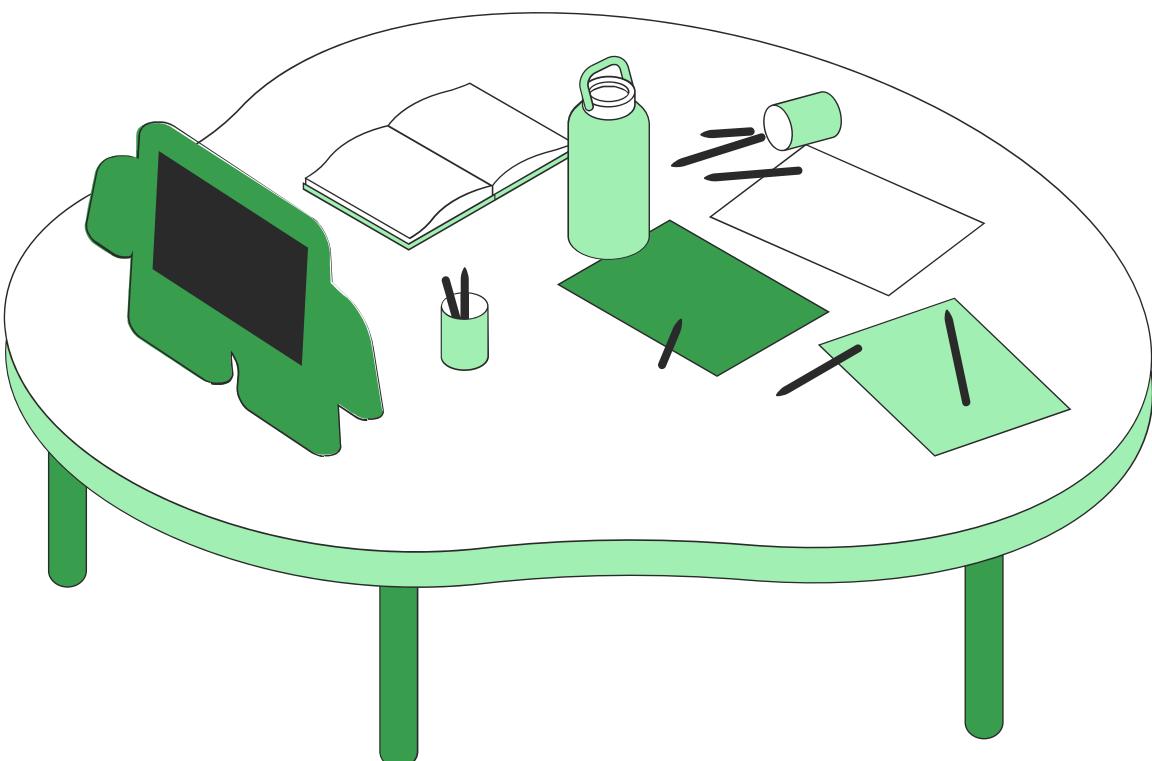
İTÜ Elektronik ve Haberleşme  
Mühendisliği

## Bike Sönmez

İTÜ Matematik Mühendisliği

Araştırma  
Veri Analizleri  
Model Kullanımları Optimizasyonlar  
Use Case Modelleri Kurma  
Ürün Geliştirme

# iÇERİK



Problem

Problemin Çözümü

Çözüm Yöntemleri

Teknik Çalışmalar

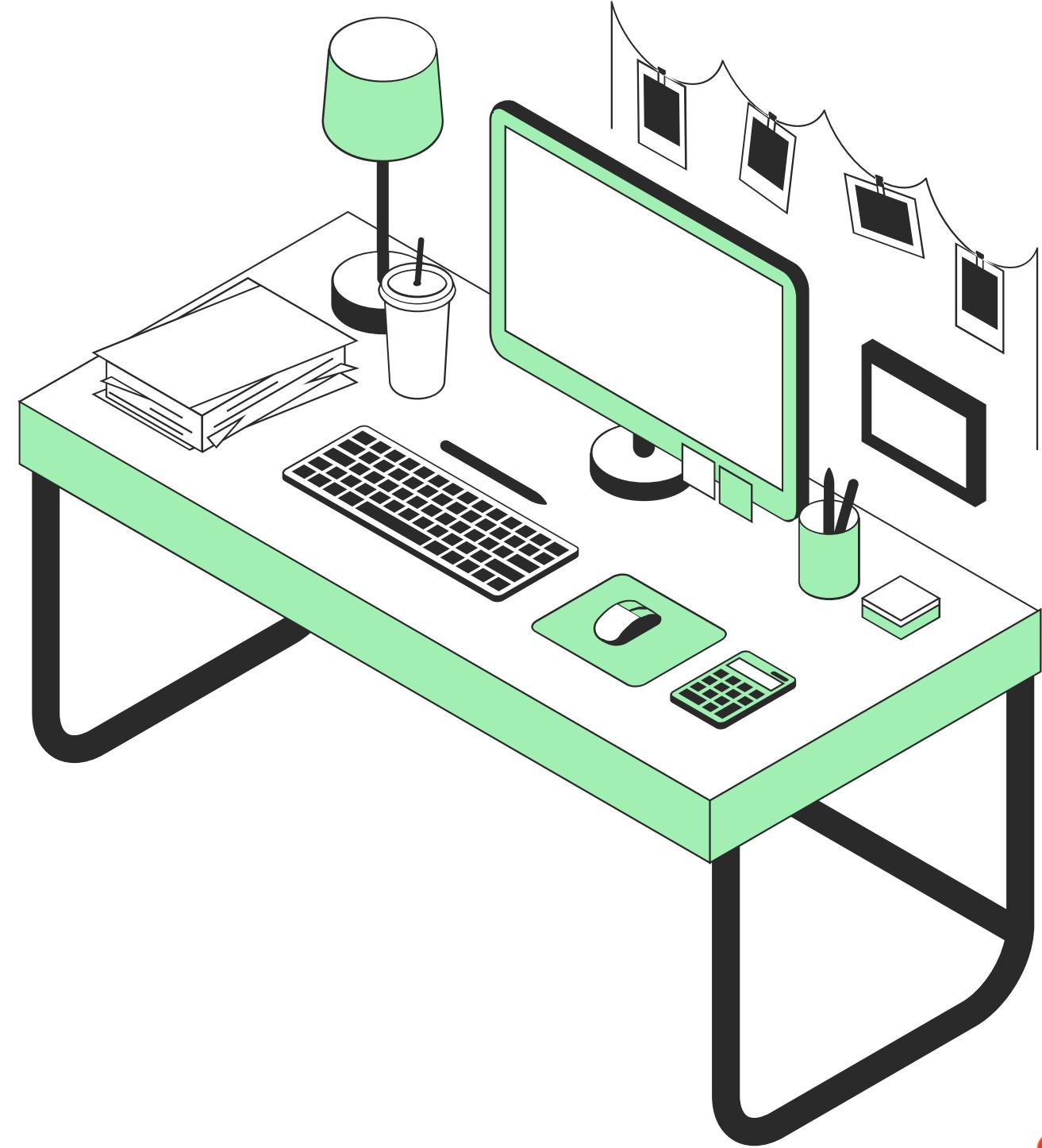
Proje İş Akışı

Yol Haritası

Demo Videosu

Kapanış

# Problem, Problemin Çözümü ve Yöntemler

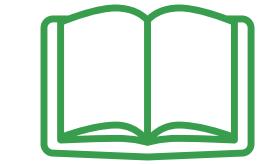


# Nefret söylemi nedir ?

Nefret söylemi veya nefretli konuşma, ırk , cinsiyet, yaş, ulus , din ya da cinsel yönelim gibi konulara dayanarak, belirli bir grup ya da kişiye yönelik nefret ifadesi içeren veya bu kişilere şiddet uygulanmasını teşvik eden aleni konuşmadır.



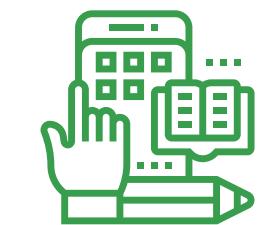
# Nefret Söylemini Önlemek ve Etkilerini Azaltmak İçin Ne Yapılabilir ?



Eğitim



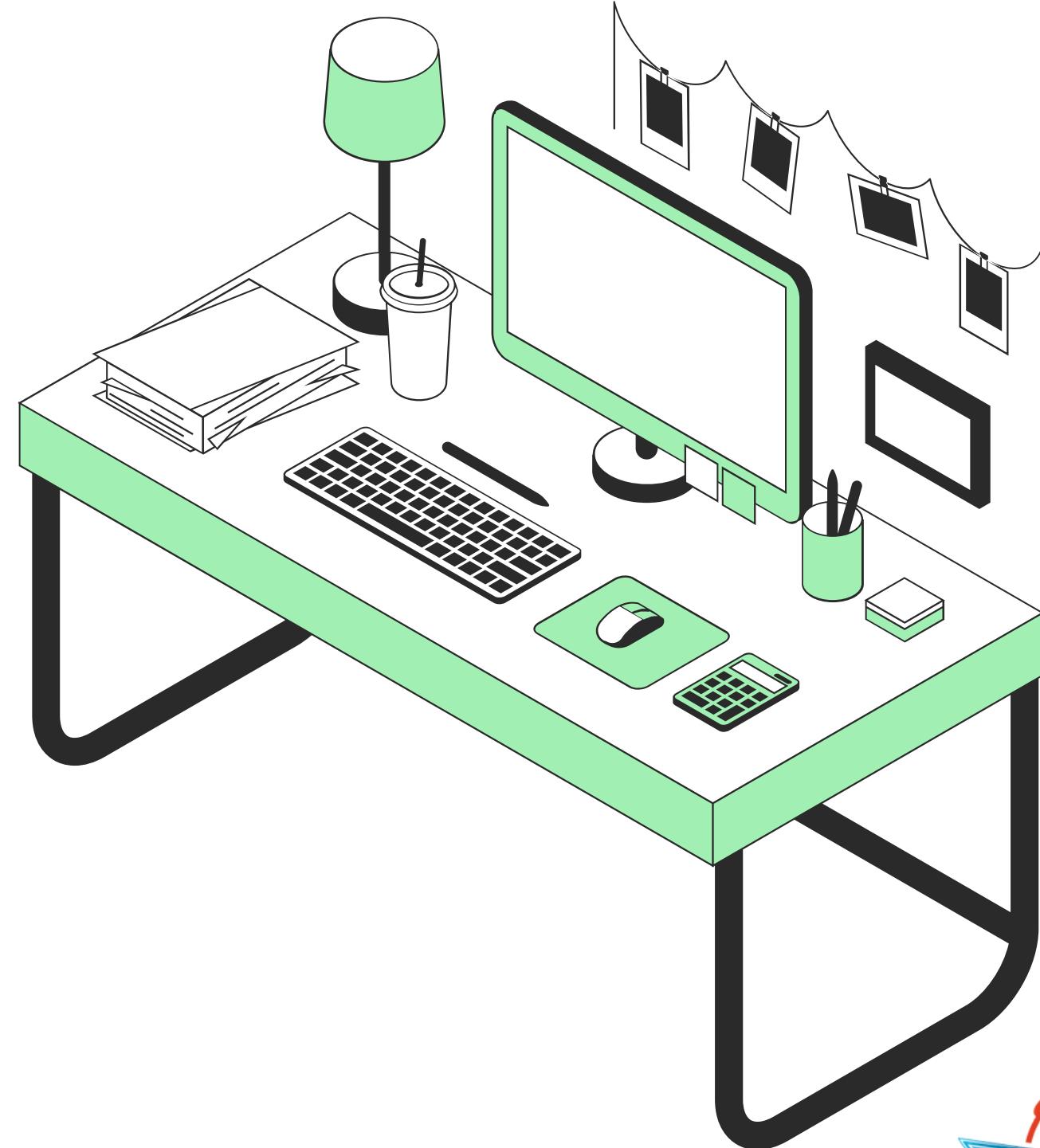
Kontrol Mekanizmaları



Teknolojik Uygulamaların  
Kullanımı

# Proje Amacı

Günümüzde iletişim araçlarının gelişmesiyle daha da yaygın hale gelen nefret söyleminin yapay zeka yardımıyla başarılı şekilde tespit edilmesi ve önlenmesi.



# Proje İş Akışı



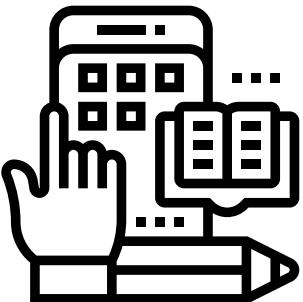
Araştırma



Analiz



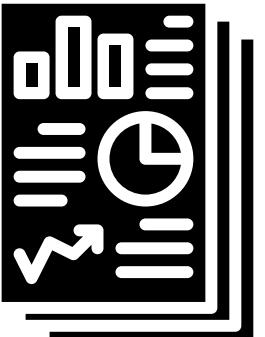
Uygulama



Geliştirme



Use case



# Araştırma



[www.turkiyeacikkaynakplatformu.com](http://www.turkiyeacikkaynakplatformu.com)



# Hangi Alanlarda Araştırma Yaptık ?

[www.turkiyeacikkaynakplatformu.com](http://www.turkiyeacikkaynakplatformu.com)

Nefret söylemi ve Etkileri

Geleneksel ve Gelişmiş NLP  
Yaklaşımları

Türkçe NLP uygulamaları

Nefret Söylemi Tespitinin Sektörde  
Kullanılabileceği Alanlar



# Analiz



[www.turkiyeacikkaynakplatformu.com](http://www.turkiyeacikkaynakplatformu.com)



# Veri Analizi

Ne tür bir veri  
setine sahibiz



Veri setine  
hangi işlemleri  
uygulamalıyız



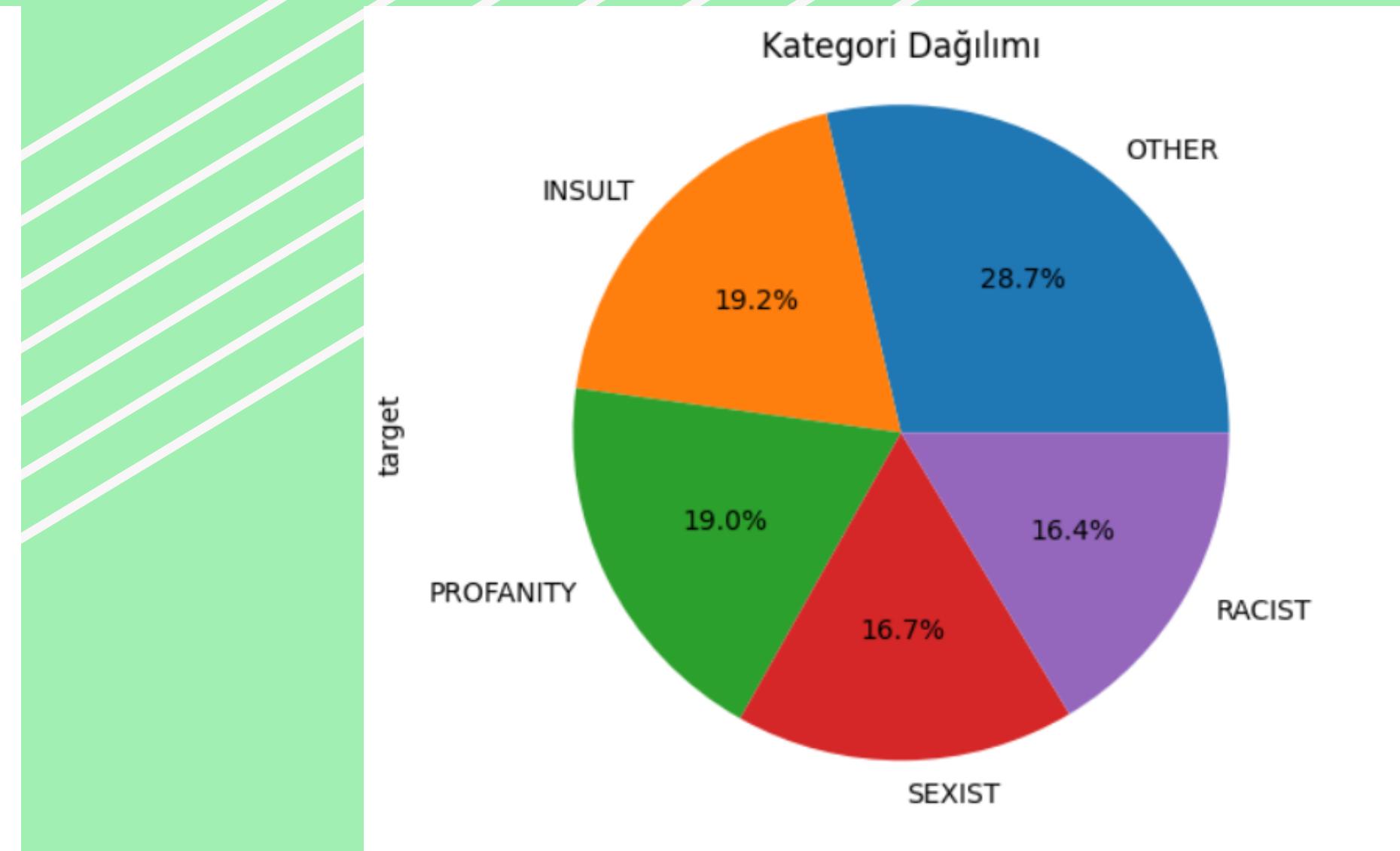
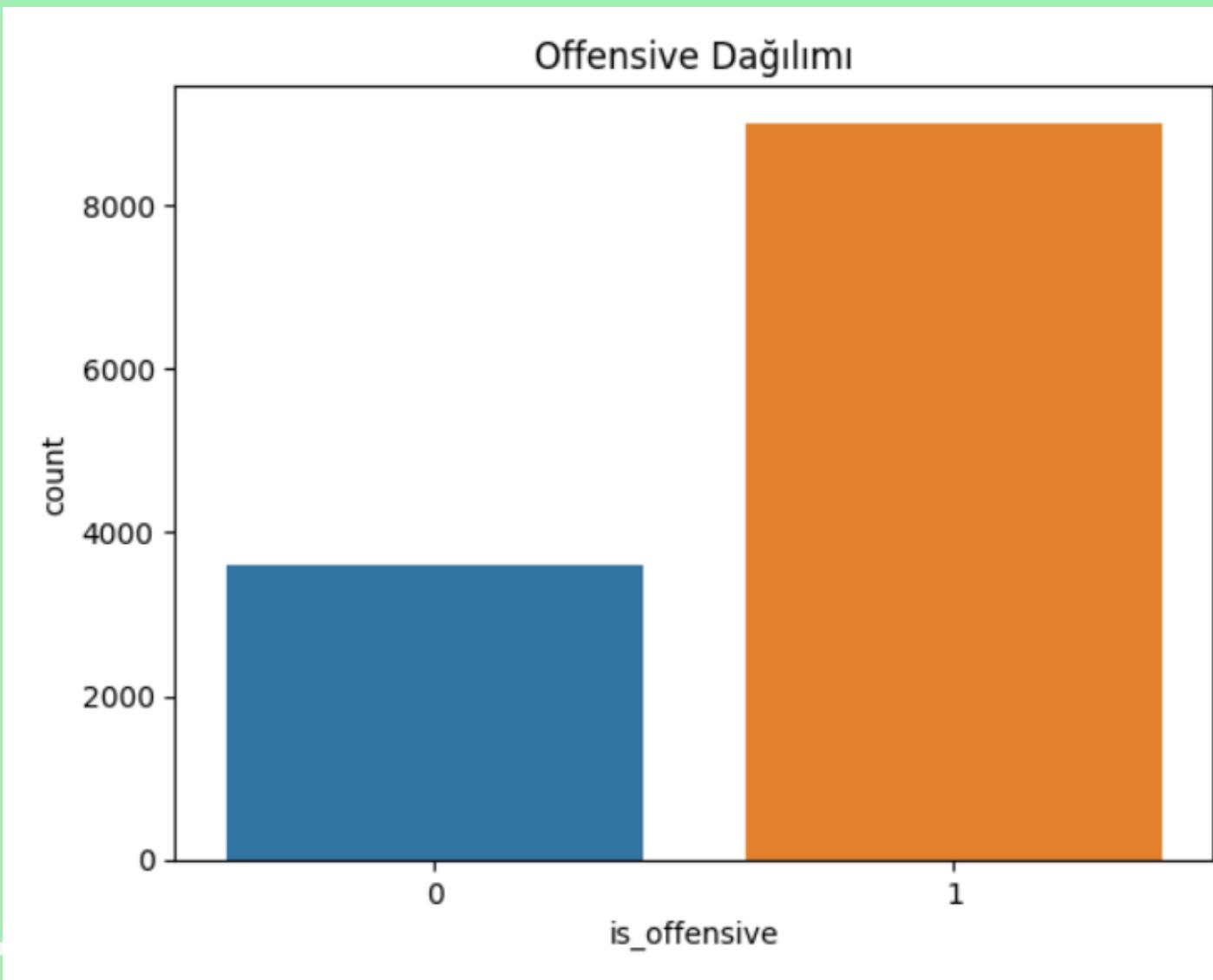
Veri setini  
modeller için  
nasıl hazır hale  
getirebiliriz



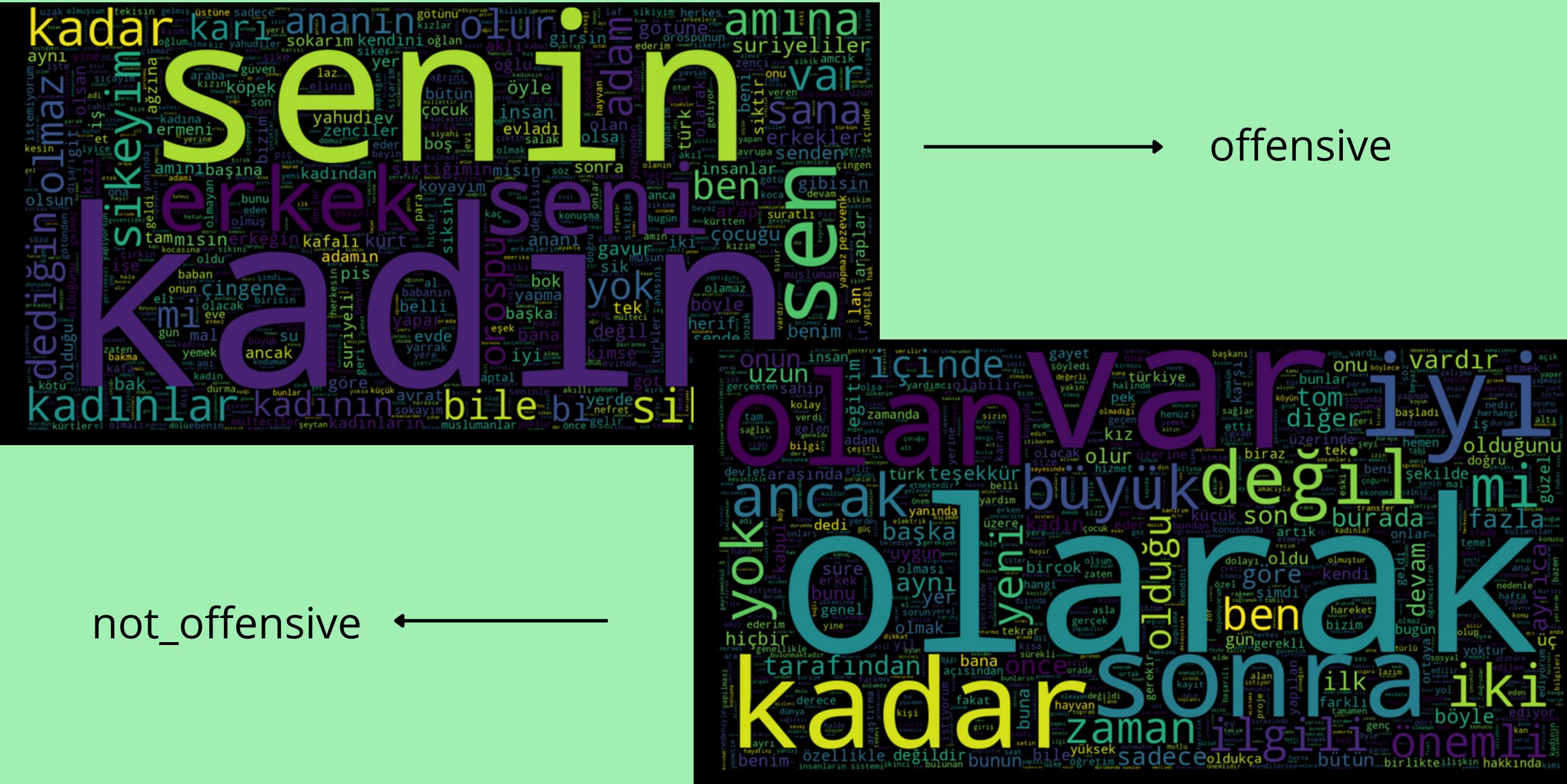
	<b>id</b>	<b>text</b>	<b>is_offensive</b>	<b>target</b>
0	81c11060-a240-4d54-841b-9e2916039e85	çürüük dişli	1	INSULT
1	be80ebbf-b322-4c3b-afa1-94932ea80731	Bu adamın islama ve müslümanlara verdiği zarar...	1	RACIST
2	f99e2513-83ed-4076-ac72-b9e2cff3f049	erkekler zora gelmez	1	SEXIST
3	83ed2b2e-b815-4f36-9fc4-80a9050cf2d0	Utanmazın götürüne kazık sokmuşlar bu tıkırtı ne...	1	PROFANITY
4	d93e05f7-bfdd-4cdb-99d8-3048761b30ff	otomasyon< sistemlerine= doğrudan bağlanabilir	0	OTHER
5	81eca520-15a0-4f5b-a1dd-5929bba34bdc	amını sikiğimin yarrağı	1	PROFANITY
6	a2497690-5201-4c28-bdb8-4206ea672765	ekipte bir küçemiz eksikti o da oldu	1	INSULT
7	b9122be9-5a9a-491e-8ecc-998ac05c6ef0	Kadın yönetici sayısı oldukça az olduğu için ş...	1	SEXIST
8	2dd1b0d1-45cd-4d4a-86c3-b290db5a0316	makara bunu duyunca avluya çıkışmış ve kızağa sa...	0	OTHER
9	d7781fd9-ee64-4b51-9abd-aa8eda638230	hay ebenin pembe patığine sıçayırm bok kafalı	1	PROFANITY

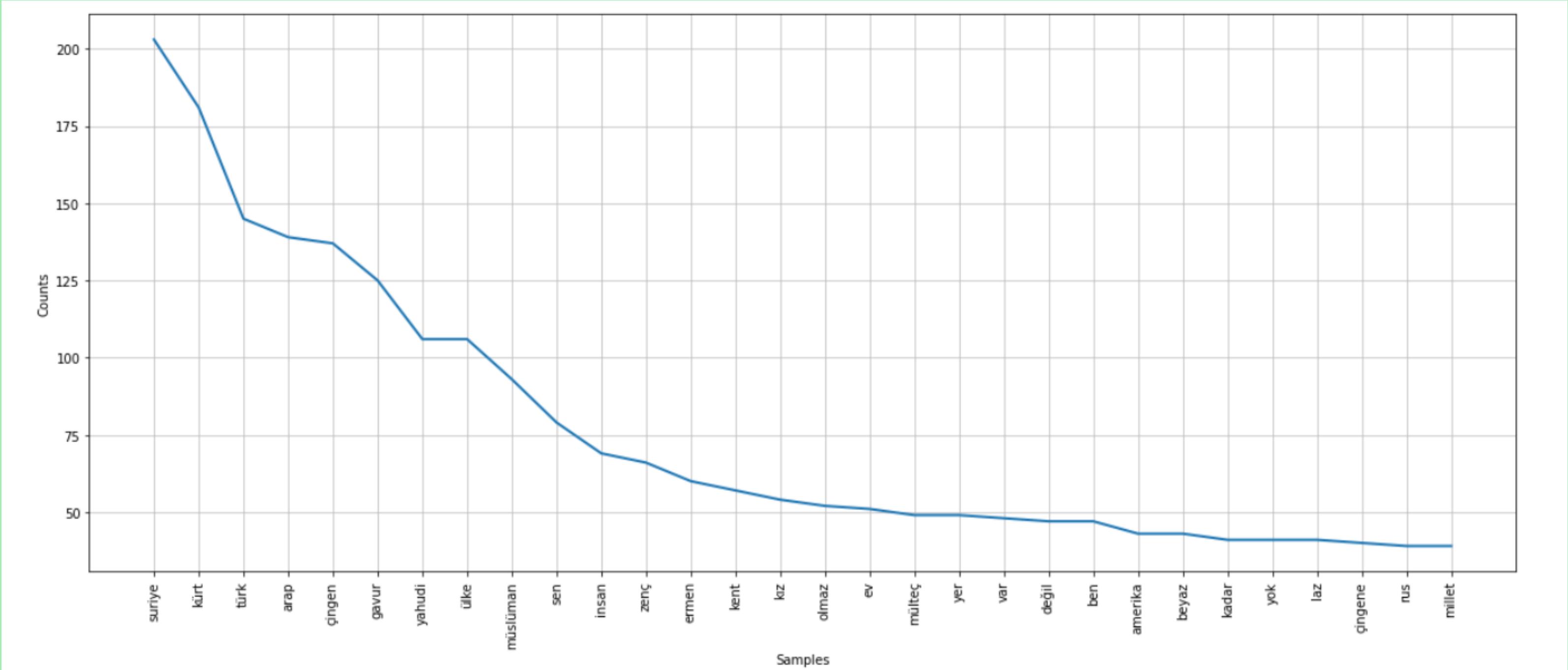
teknofest\_train\_final.csv

# Bazi Veri Dağılımları



# WordCloud ile Görselleştirme





Racist Kategorisine Ait Metinlerde En Çok Geçen Sözcükler

[www.turkiyeacikkaynakplatformu.com](http://www.turkiyeacikkaynakplatformu.com)



# Preprocessing

Özel karakterlerin silinmesi

---

Tokenizasyon

---

Lemmatizasyon

---

Durma Sözcüklerinin Silinmesi

---

Vektörizasyon

# Uygulama



[www.turkiyeacikkaynakplatformu.com](http://www.turkiyeacikkaynakplatformu.com)



# Geleneksel ML Yaklaşımları



**Bernoulli Naive  
Bayes Classifier**

**Support Vector  
Machine (SVM)**

**Decision Tree  
Classifier**

**XGBoost Classifier**

**Voting Classifier**

# Dil Bazlı Gelişmiş NLP Yaklaşımları



**LSTM - BiLSTM**



**BERT**



**GPT**



**Ensemble Modeller**

# Modellerde Karşılaştığımız Problemler Nelerdi ?

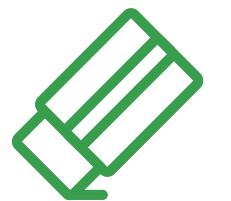
[www.turkiyeacikkaynakplatformu.com](http://turkiyeacikkaynakplatformu.com)



**İnsult kategorisini ayırt  
etme problemi**



**Bazı modellerin Türkçe  
versiyonlarındaki hatalar**



**Model Hiperparametre  
Optimizasyonu**

# Bizim Bu Problemlere Karşı Yaklaşımlarımız Nelerdi ?

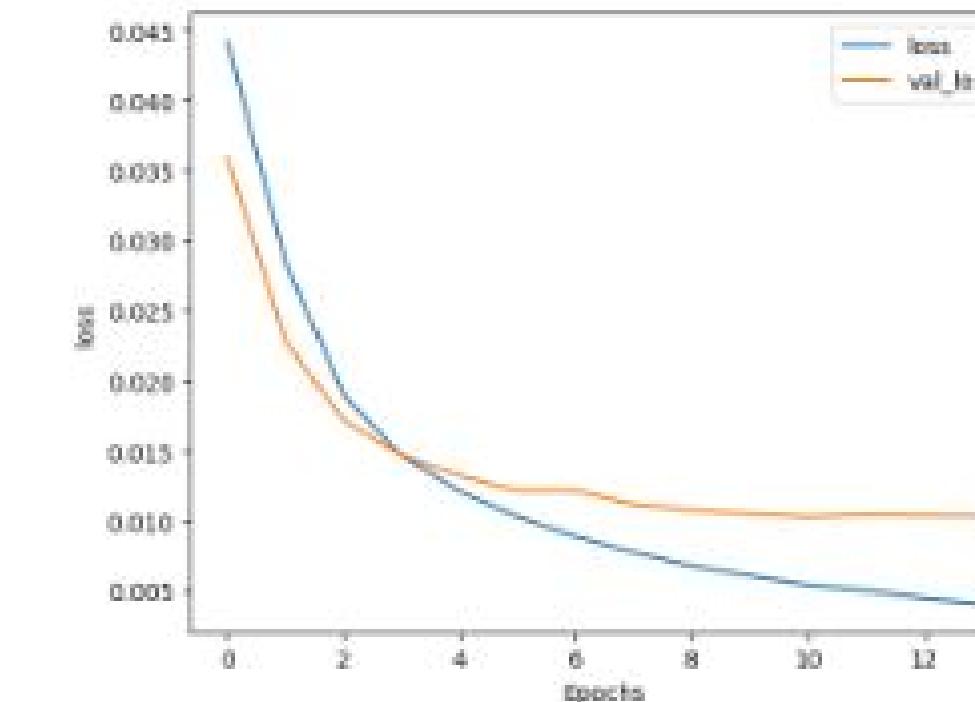
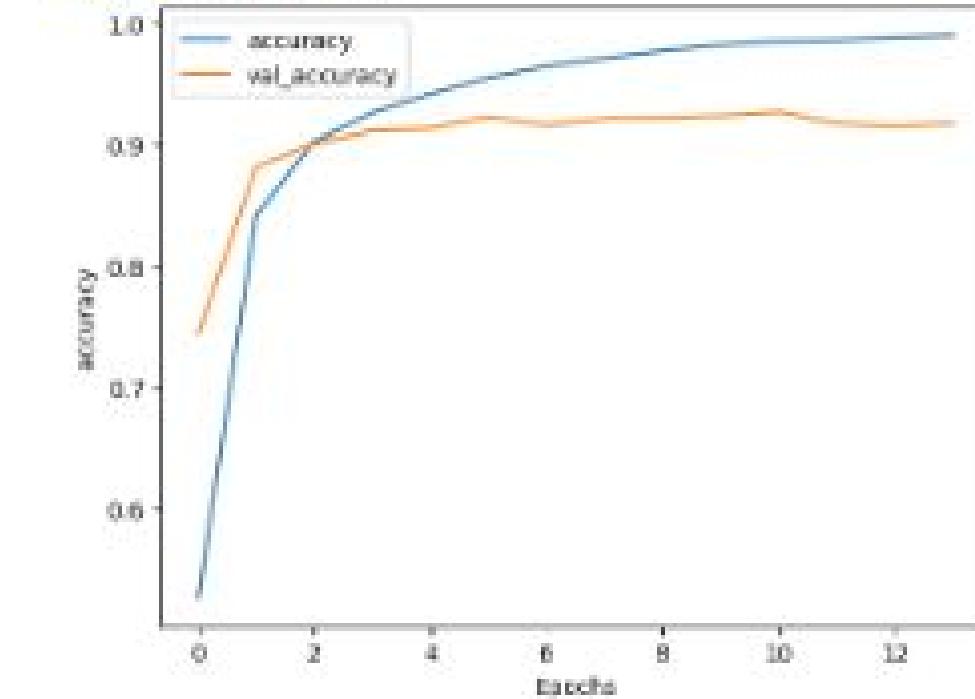


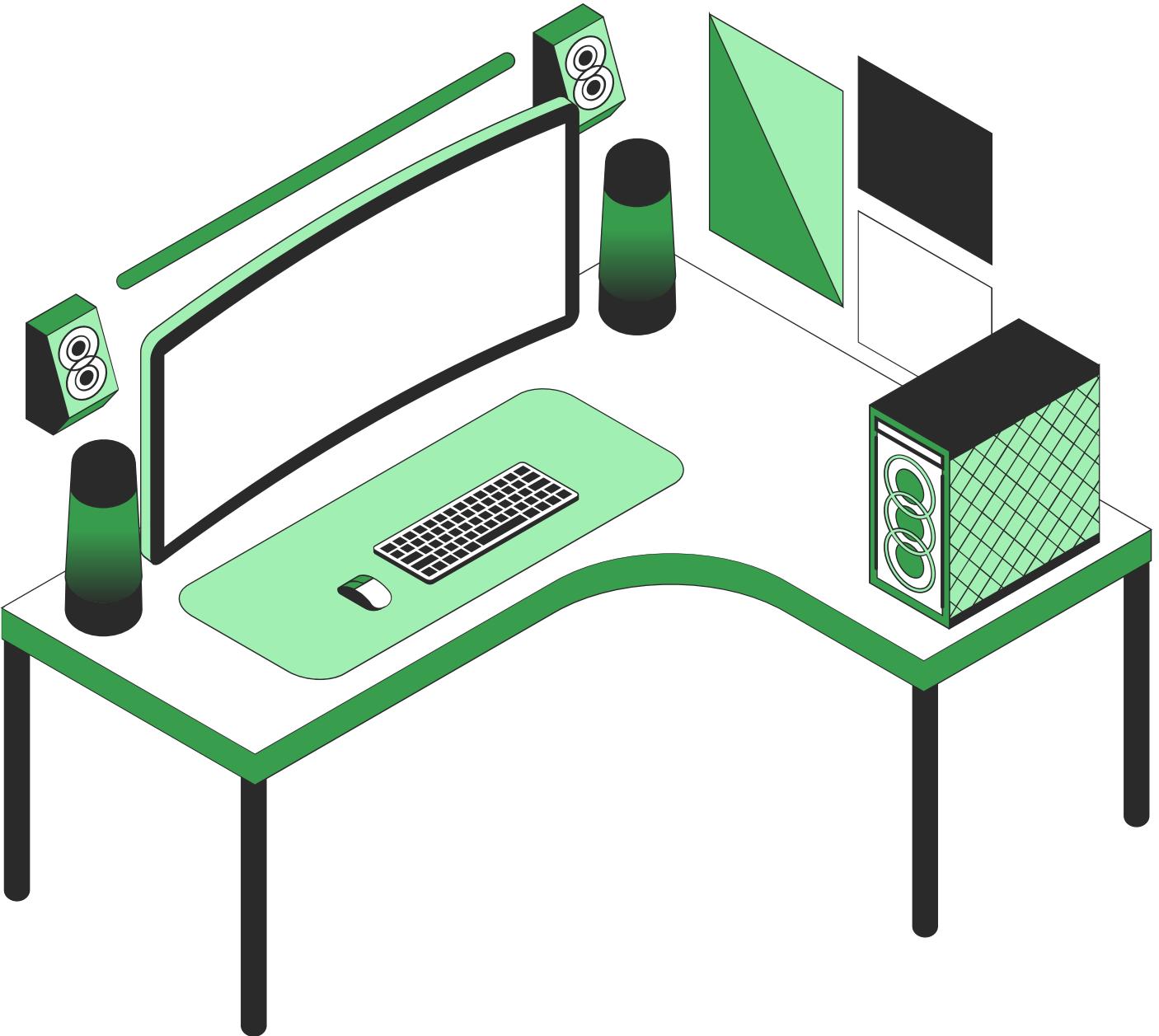
- Insult kategorisindeki değerler için veri artırma
- Daha gelişmiş model kombinasyonları kullanma (Ensemble Modeller)
- Earlystopping ile en optimum epoch sayısını bulma (11)

# Hiperparametre Optimizasyonları

## 1 - EarlyStopping 2 - Schedular

```
18088 [██████] 312/312 [03:27:00, 1.5lit/s]
Epochs: 12 | Train Loss: 0.005 | Train Accuracy: 0.985 | Val Loss: 0.010 | Val Accuracy: 0.917
18088 [██████] 312/312 [03:27:00, 1.5lit/s]
Epochs: 13 | Train Loss: 0.004 | Train Accuracy: 0.988 | Val Loss: 0.010 | Val Accuracy: 0.914
18088 [██████] 312/312 [03:27:00, 1.5lit/s]
Epochs: 14 | Train Loss: 0.004 | Train Accuracy: 0.990 | Val Loss: 0.010 | Val Accuracy: 0.917
[Early stopping at epoch 14]
```





# Neden Ensemble Model?

# Ensemble Model Yaklaşımı



**BERT TÜRK**

- BERT- Multilingual
- BERT -Turkish
- Kelime dizileri üzerinde çalışır



**BİLSTM**

- Sıralı metinleri sınıflama



**Convolutional neural  
network(CNN)**

- Metindeki desenleri tanımak için
- Karakter bazlı embedding

# Neler Yaptık



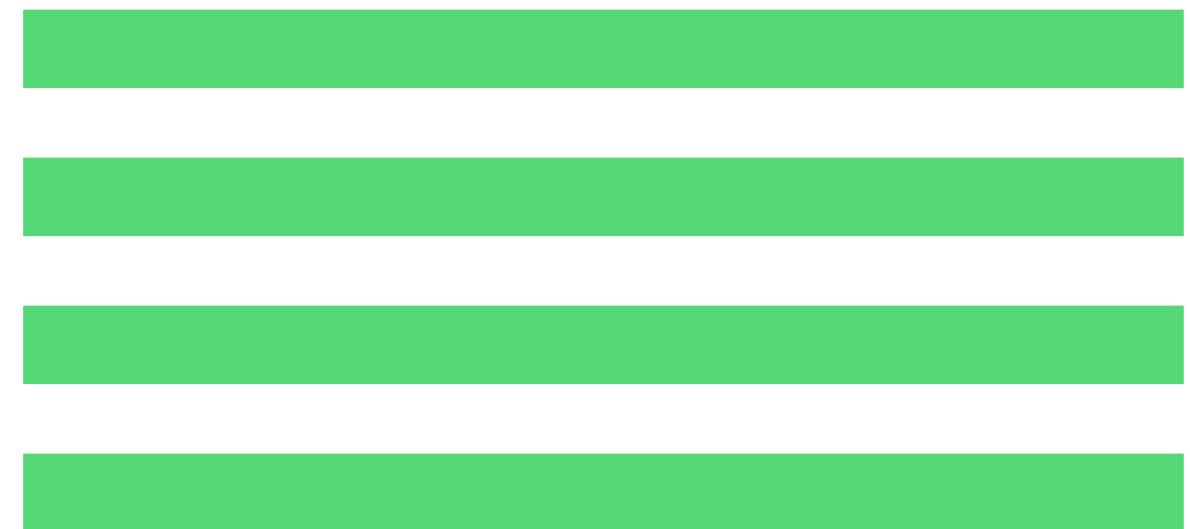
**Geleneksel  
Sınıflandırma  
Modelleri**

**Transformer  
Modeller**

**BERT-  
Multilingual**

**BERT-  
Türkçe**

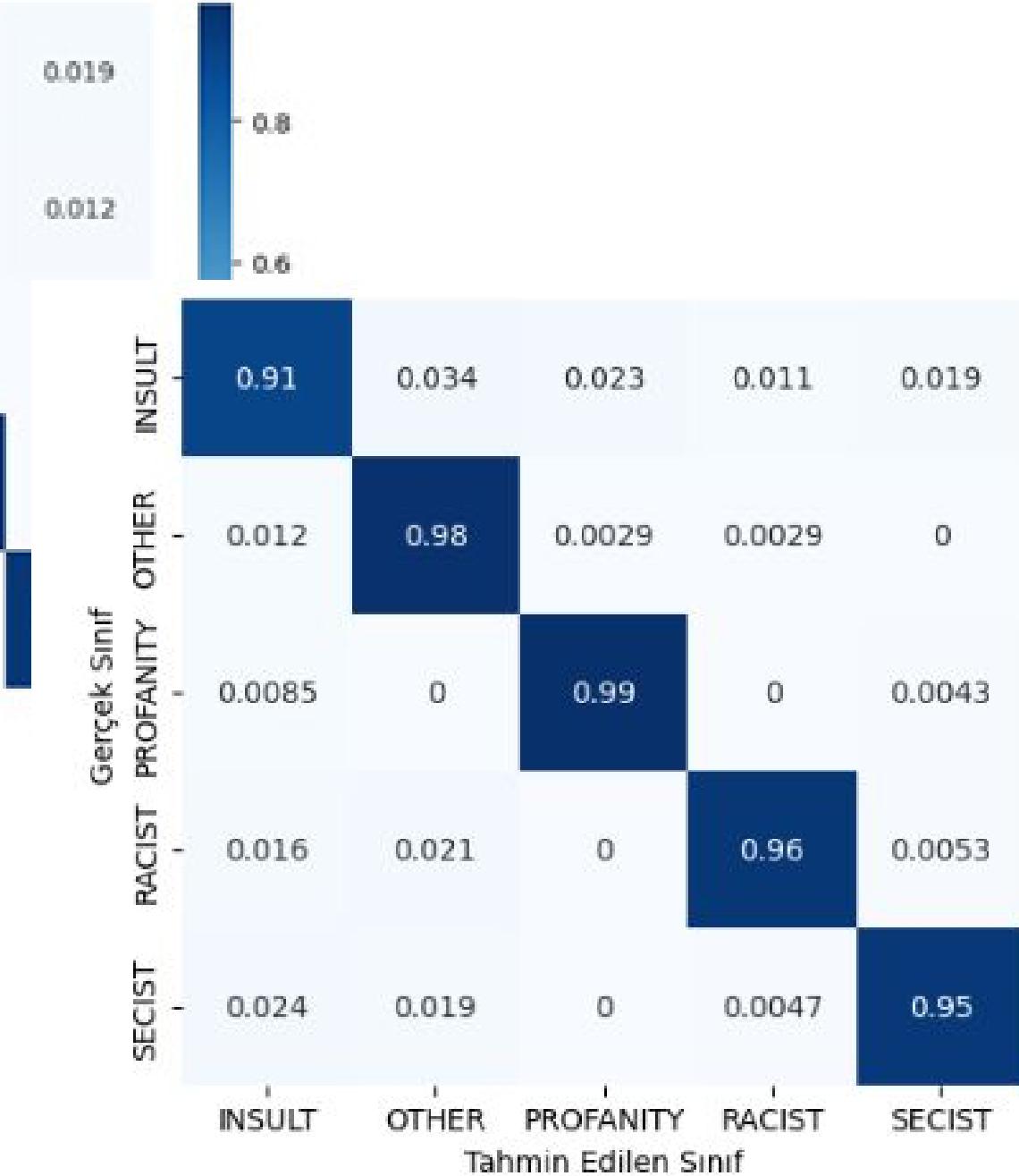
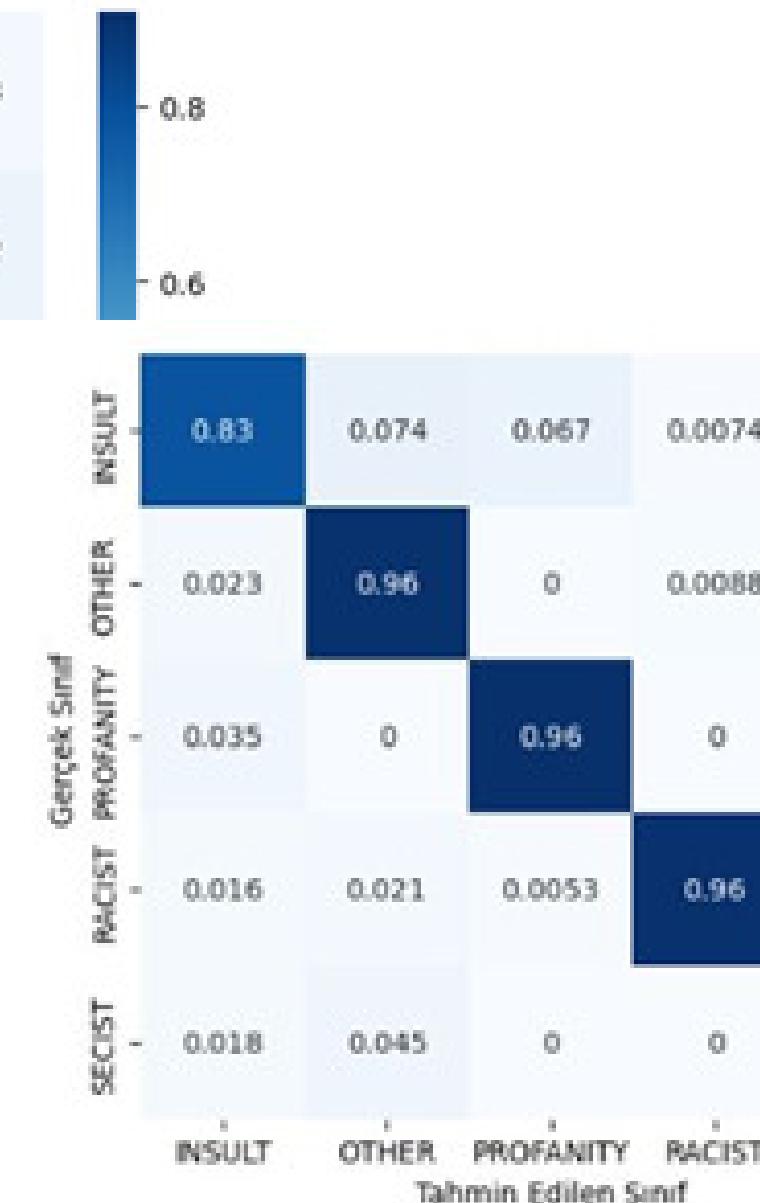
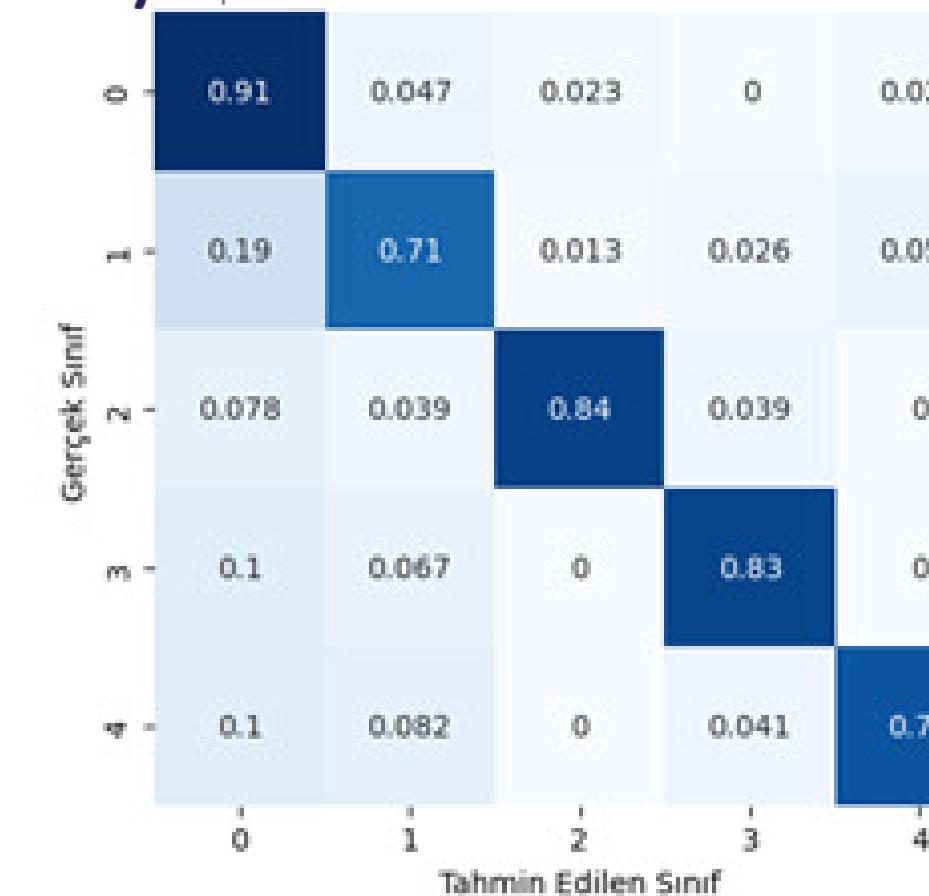
**Ensemble  
yaklaşım**



[www.turkiyeacikkaynakplatformu.com](http://www.turkiyeacikkaynakplatformu.com)

# Skorlar

Model	Epoch	CV	Macro F1 Skoru
Linear SVM	-	+	0.8685
XGB Classifier	-	+	0.8126
LSTM	10	+	0.7940
Voting Classifier	-	+	0.8406
Distil Bert	5	+	0.9004
Bert Classifier	5	+	0.9794
Bert Ensemble	5	+	0.9815





# Yol Haritası

# Elde Ettiğimiz Çıktıyı Nasıl Bir Ürüne Dönüştürebiliriz ?



# Telco ve OTT Sektörlerinde Nasıl Kullanabiliriz ?



# Use Case Modeli

[www.turkiyeacikkaynakplatformu.com](http://turkiyeacikkaynakplatformu.com)





Çocuklara Özel Dijital  
Ortam

[www.turkiyeacikkaynakplatformu.com](http://www.turkiyeacikkaynakplatformu.com)

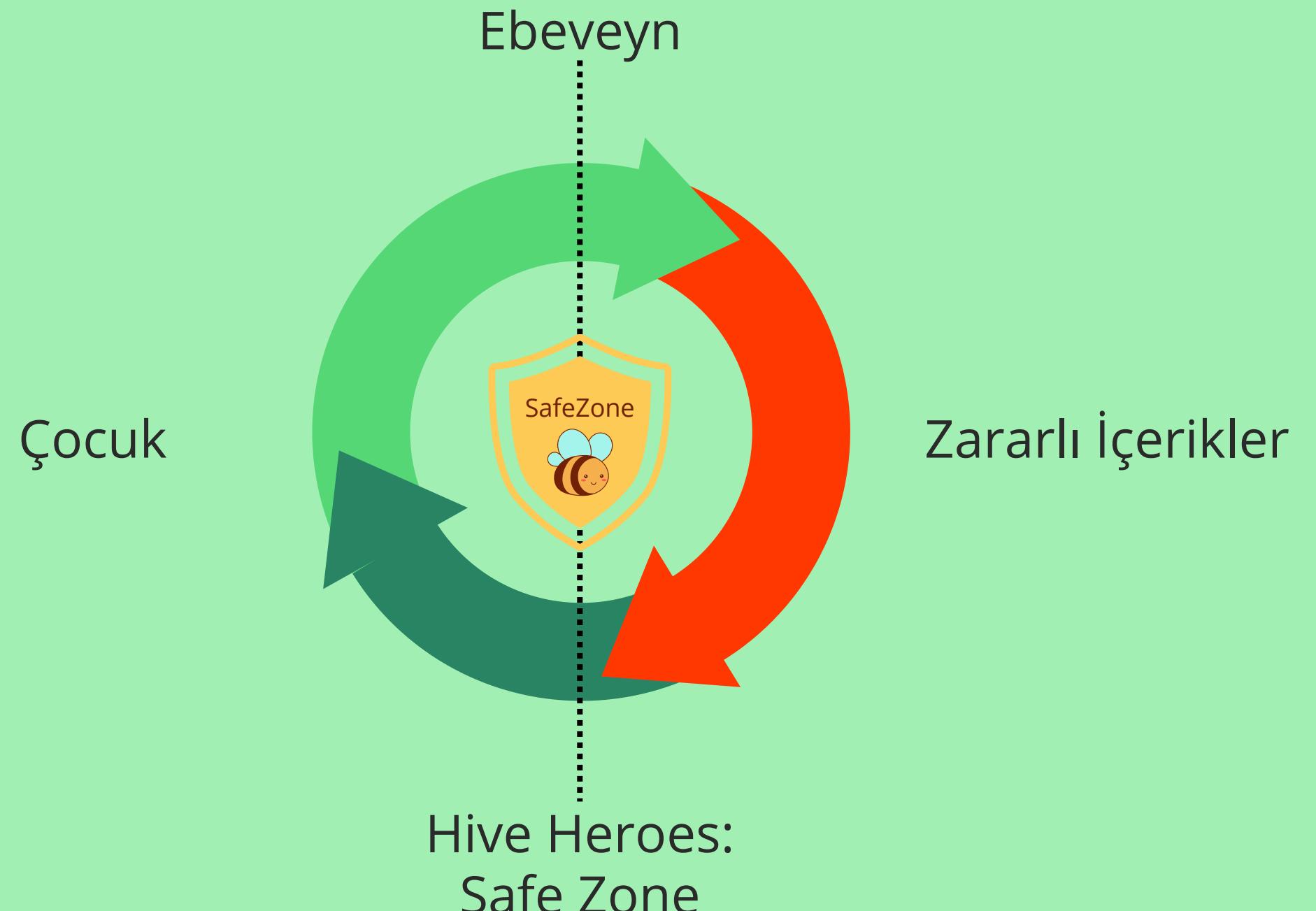


Canlı Yayınlar İçin  
Eklentiler



Sosyal Medya  
Filtrelemesi

# Çocukları Korumak İçin

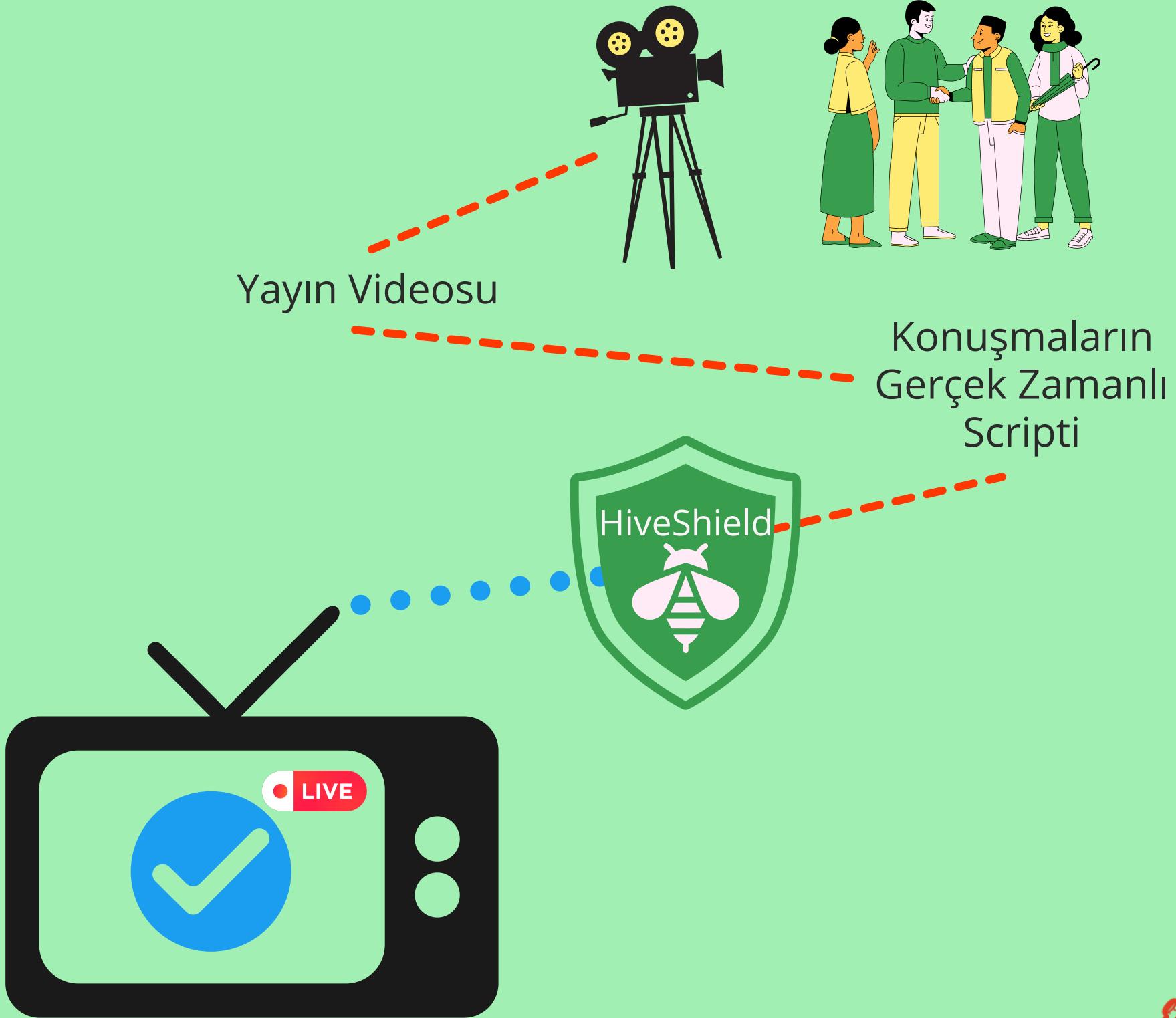


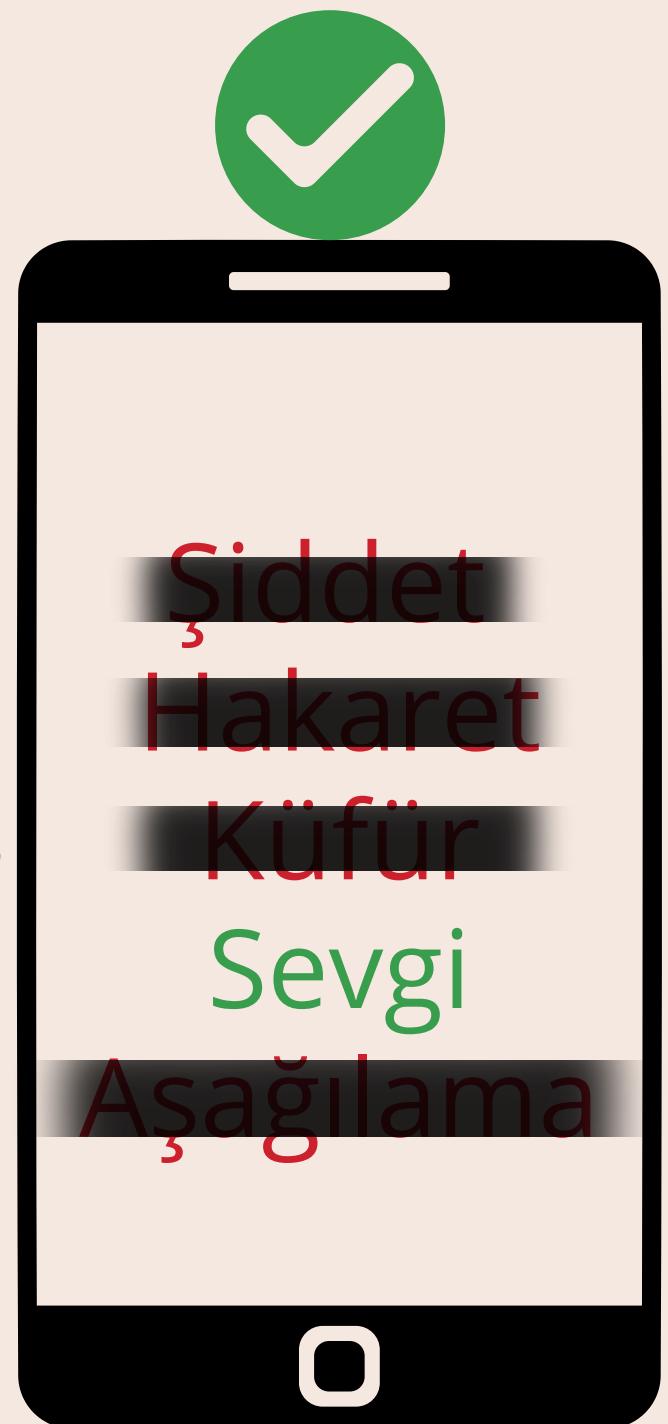
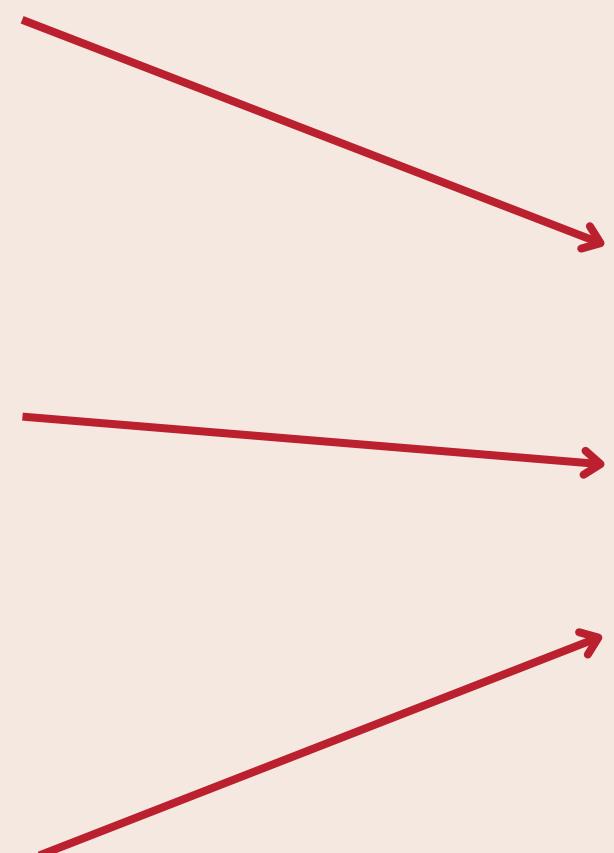
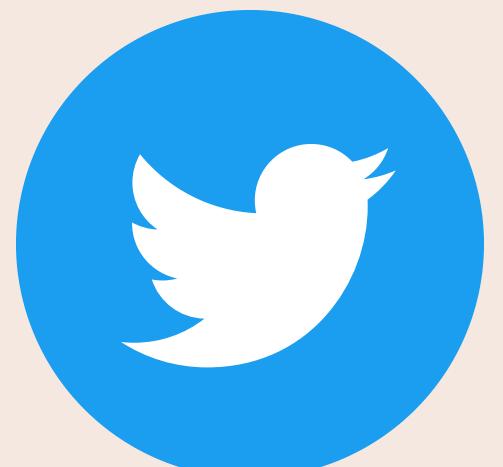
[www.turkiyeacikkaynakplatformu.com](http://www.turkiyeacikkaynakplatformu.com)

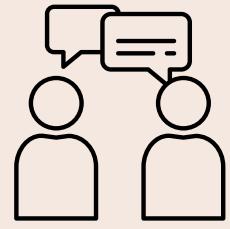


Çocuklara Özel Dijital  
Ortam Sağlayan Program

# Canlı yayınılarda insanların ne konuşacağını bilemeyez...





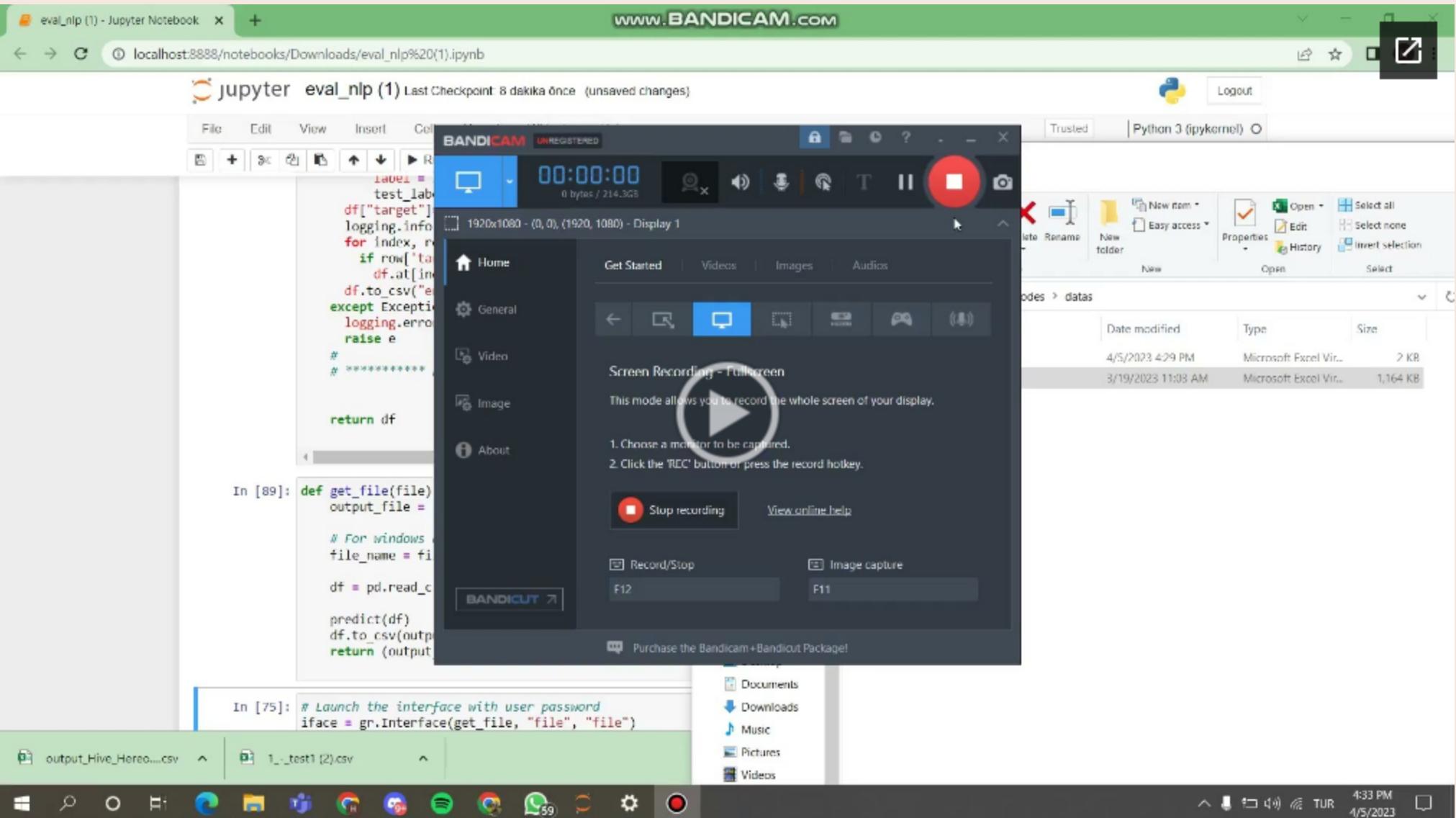


# Topluluğa Sağlanan Katkı ?



## HIVE

# Demo Videosu



[https://drive.google.com/file/d/1R2T3hE7XAfyZMHEYYci6MJWyUory4jDu/view?usp=share\\_link](https://drive.google.com/file/d/1R2T3hE7XAfyZMHEYYci6MJWyUory4jDu/view?usp=share_link)

# Bizi Dinlediğiniz İçin Teşekkürler

## HiveHeroes

Berfin Duman  
[berfiinduman@gmail.com](mailto:berfiinduman@gmail.com)

Bike Sönmez  
[bikesnmz@outlook.com](mailto:bikesnmz@outlook.com)

