

Türkiye Açık Kaynak Platformu  
Online Yarışma Programı

# Türkçe Doğal Dil İşleme

## - Ekibimiz ve Katkıları -

### Ertuğrul Demir

- Takım Kaptanı
- Veri Bilimci
- Modelleme ve Kontrol

### Ayberk İmir

- Takım Üyesi
- Veri Bilimci
- Literatür Araştırması

### Askar Bozcan

- Takım Üyesi
- Yazılım Mühendisi
- Arayüz Tasarımı



## - PROBLEM -

- Hakaret, küfür, ırkçılık ve cinsiyetçilik gibi istenmeyen içeriklerle günlük hayatta, sosyal medya, ürün yorumları ve hatta bireysel iletişim kanallarında karşılaşılabilmektedir.
- Saldırgan ve toksik dil, siber zorbalık, taciz ve ayrımcılığa yol açabilir ve bu da zihinsel sağlık ve refah üzerinde ciddi sonuçlara neden olabilir. Ayrıca, bireyleri çevrimiçi topluluklara katılmaktan caydıran, online işletmeleri ve markaları olumsuz etkileyen toksik bir ortam yaratabilir.

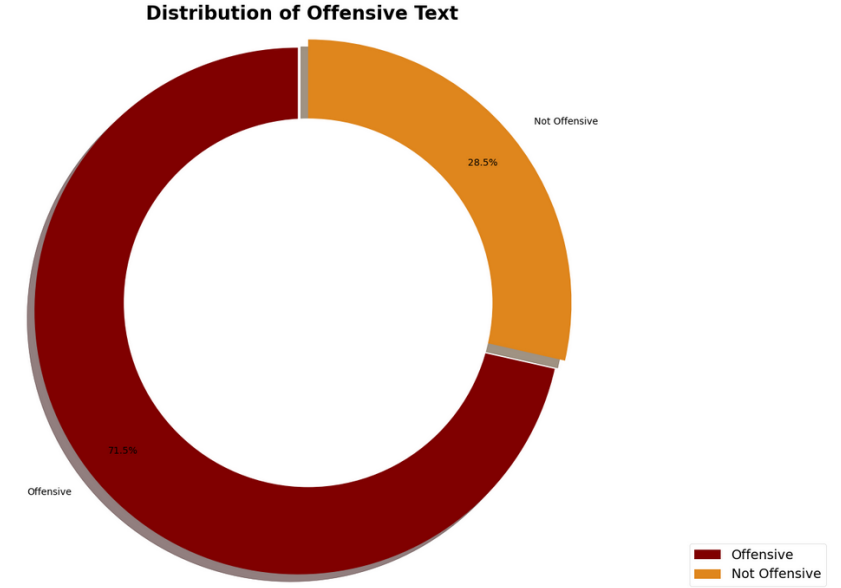
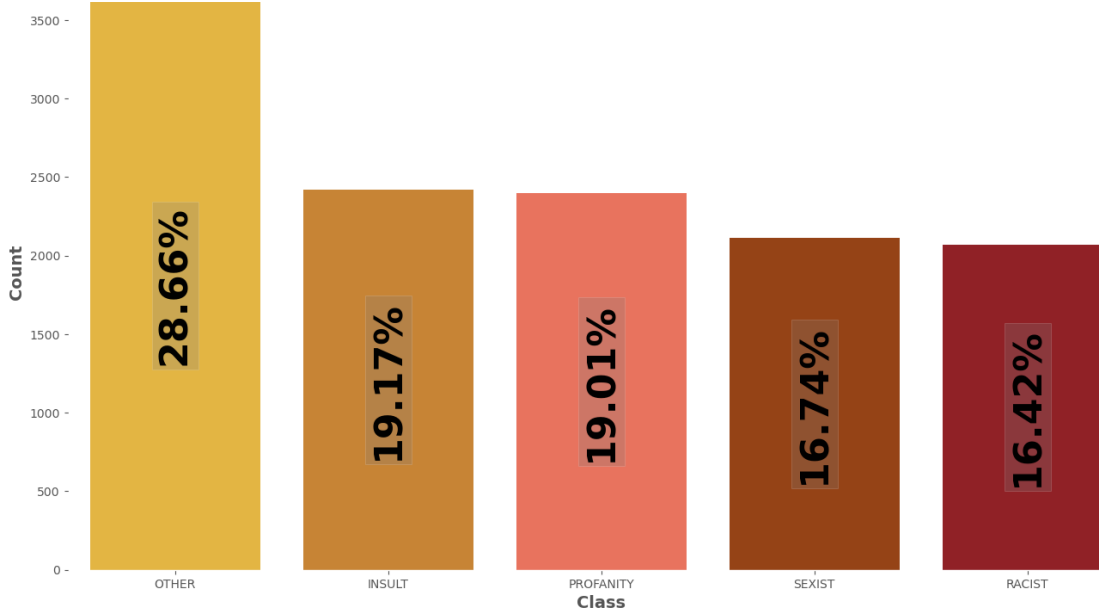
## - PROBLEMİN ÇÖZÜMÜ -

- İstenmeyen içerikleri bulunduran metinlerin üçüncü şahıslar üzerinde kötü etkilerinin önüne geçilmesi için zamanında ve hızlı müdahale şarttır fakat bunu insan gücü kullanarak yapmak çok maliyetli olacaktır.
- Geçtiğimiz yıllarda birçok görevde insani sevilere yaklaşan Doğal Dil İşleme modellerinin kullanılarak zararlı içeriklerin gerçek zamanlı olarak tespiti ve bu içeriklere müdahale edilmesi mümkündür.

# - HANGİ YÖNTEMLE ÇÖZÜM GELİŞTİRİLDİ?-

# -Verinin İncelenmesi-

## Class Counts

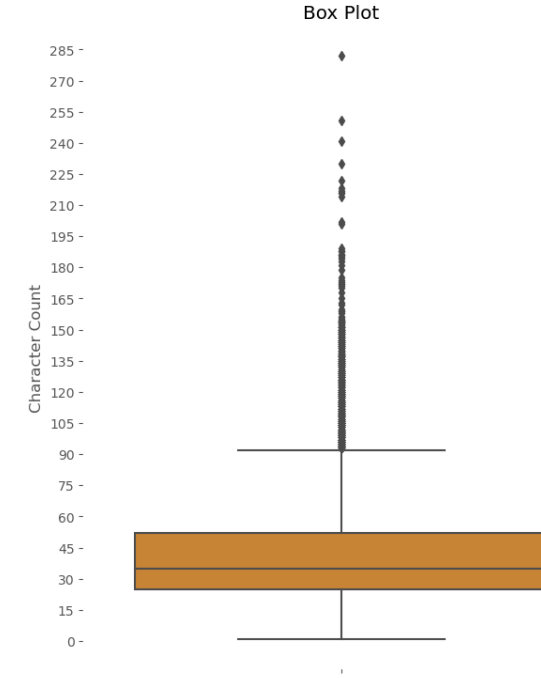
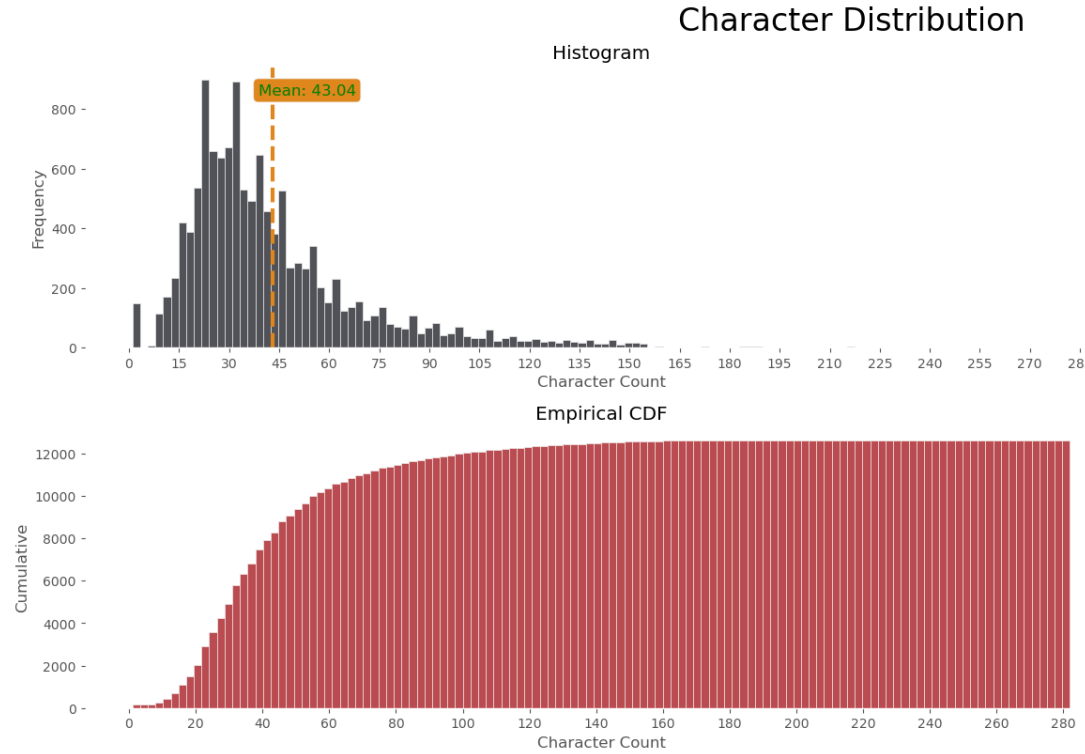


**Aksiyon:** Validasyon şeması oluşturuldu. (Çapraz Doğrulama)

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



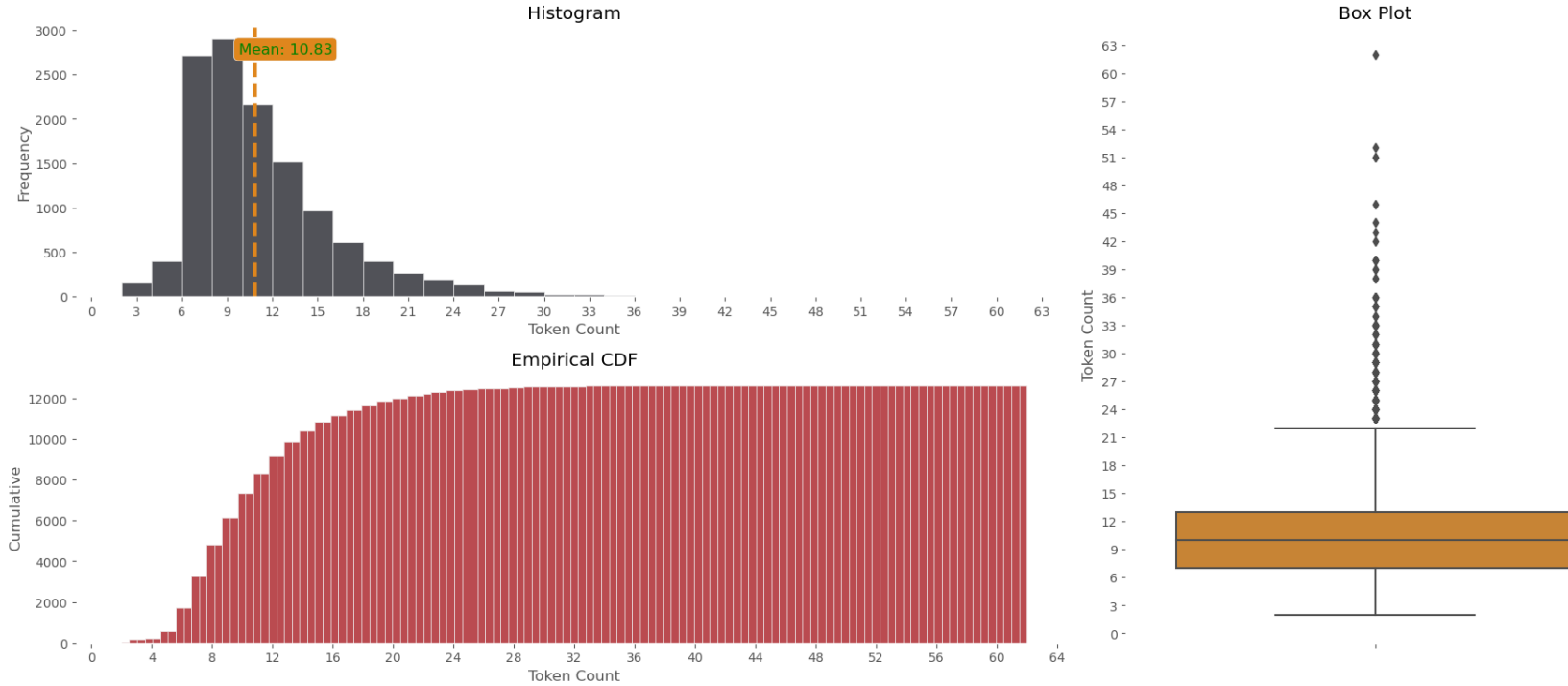
# -Verinin İncelenmesi-



**Aksiyon:** Tek karakterli satırlar atıldı. Noktalama işaretleri temizlendi.

# -Verinin İncelenmesi-

Token Distribution



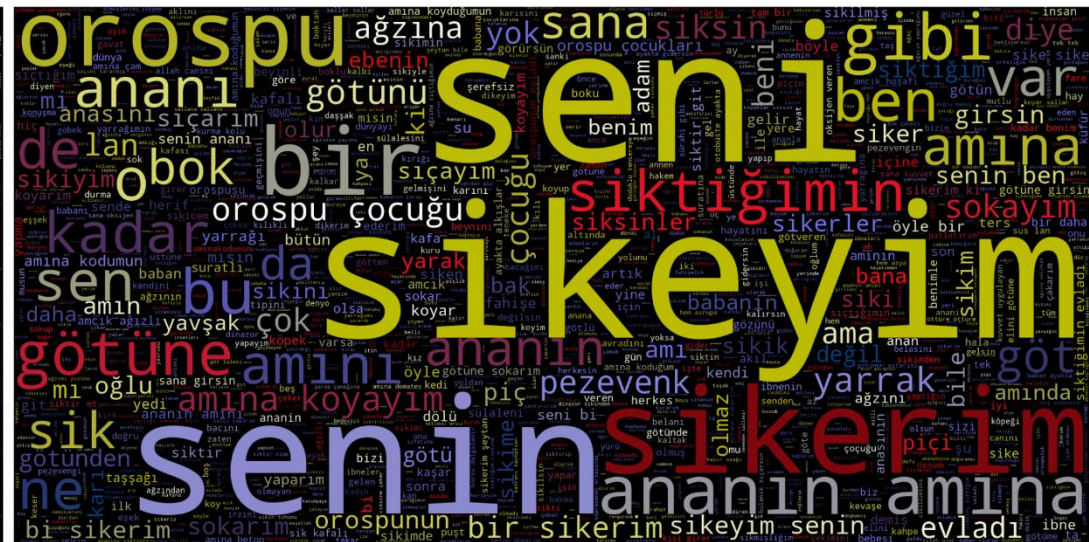
**Aksiyon:** Sekans uzunluğu belirlendi. Benzer textler atıldı.



## -Verinin İncelenmesi-



# INSULT



# PROFANITY

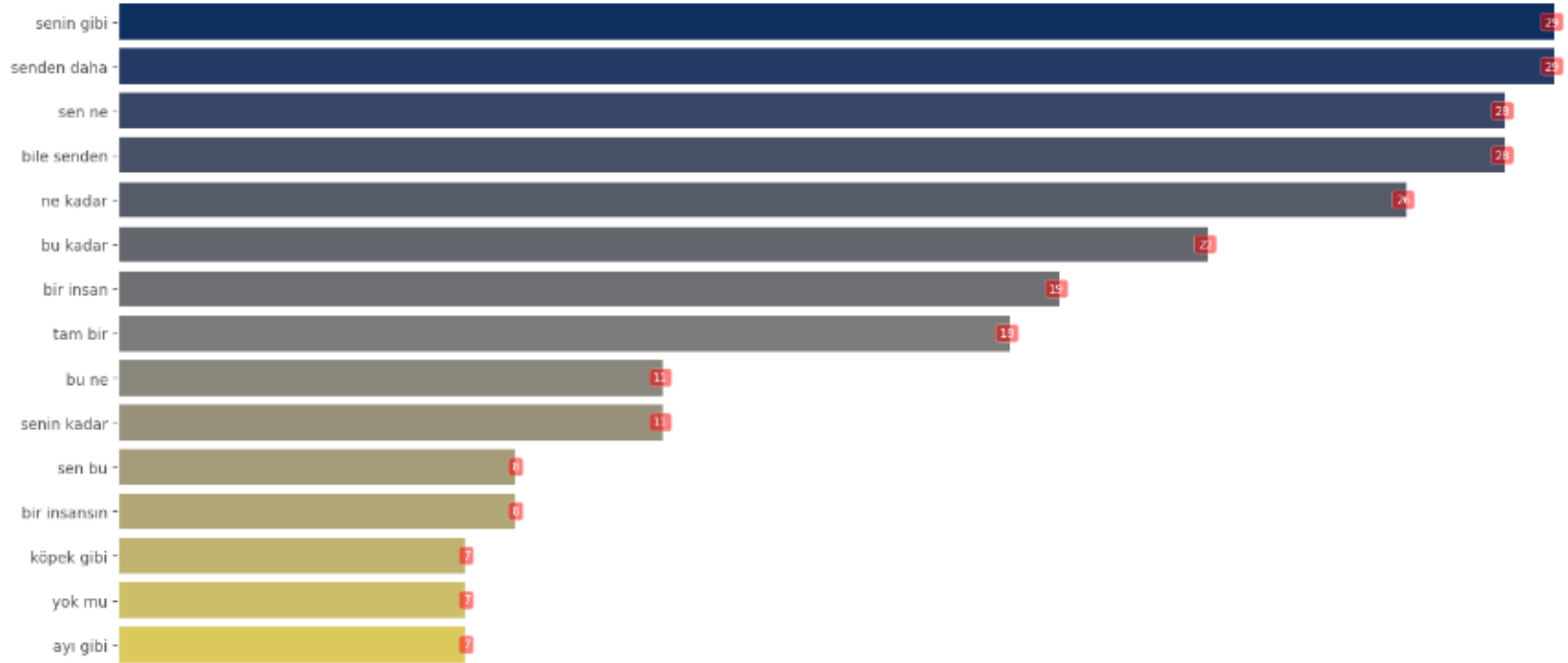
**Aksiyon:** Insult sınıfının diğer sınıflarla kesiştiği noktalar gözlemlendi.

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



# -Verinin İncelenmesi-

Most Common Unigrams "INSULT"

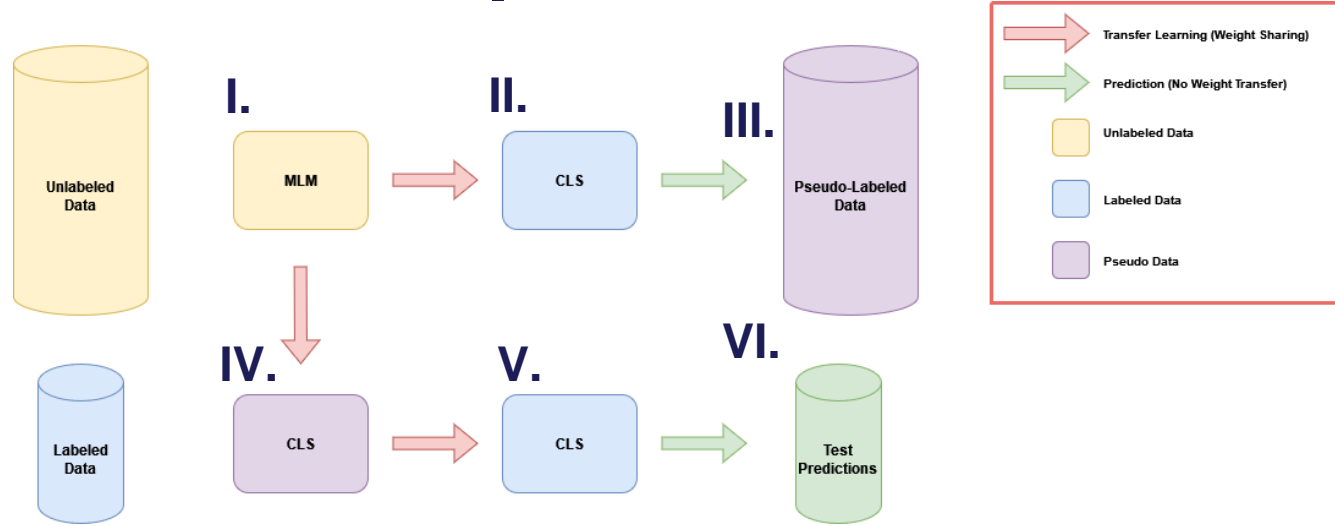


**Aksiyon:** Stopwordler context güçlendirdiği düşünülerek korundu.

## - Modelleme Aşamaları -

- Keşifsel veri analizi sonrası aday mimariler listelendi ve baseline modeller oluşturuldu.
- Hata analizi ile modellerin güçlü ve zayıf yönleri saptandı.
- Klasik yöntemlerle dış verilerin etkisiz kalması sonucu bu veriden maksimum faydanın sağlanması için özel bir yaklaşım uygulandı.

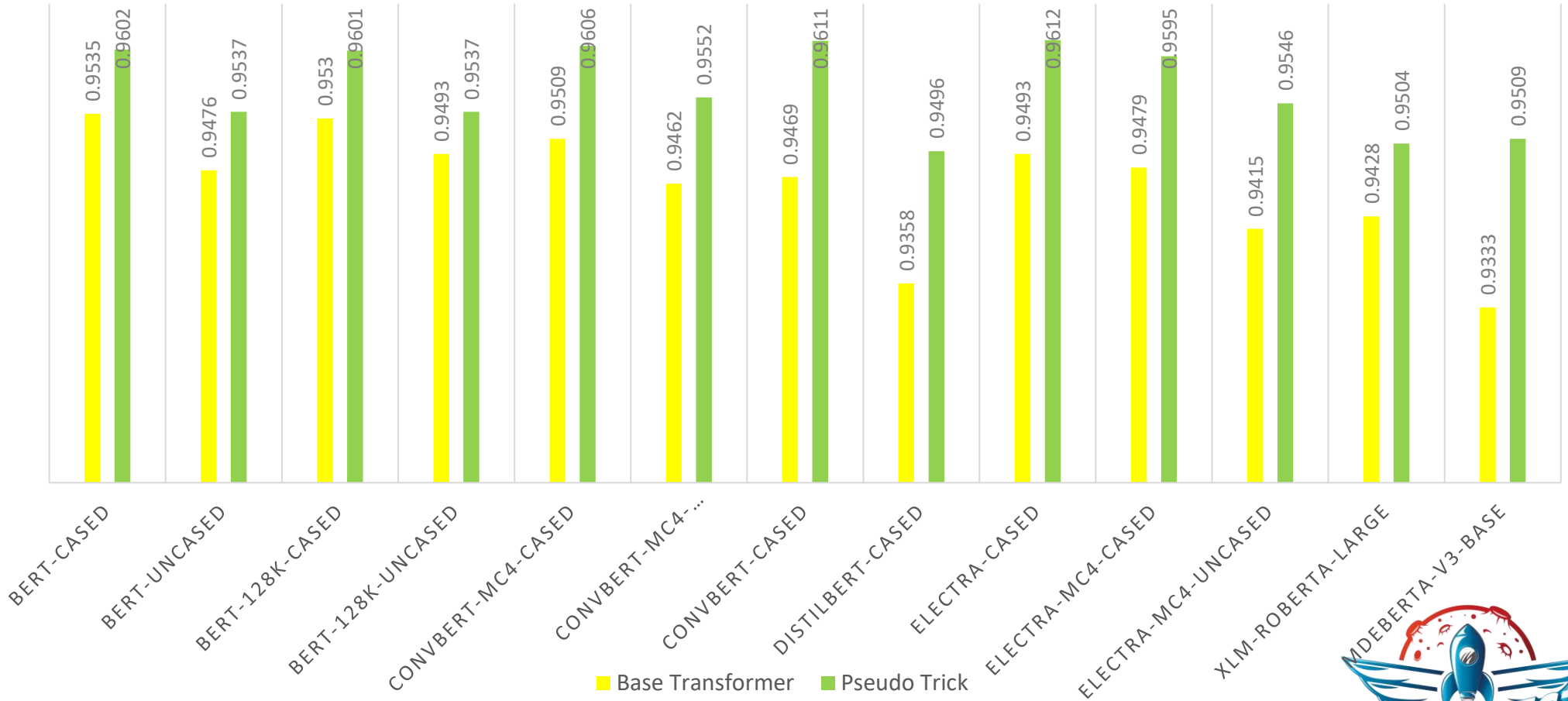
## - Domain Adaptation & Pseudo Labels -



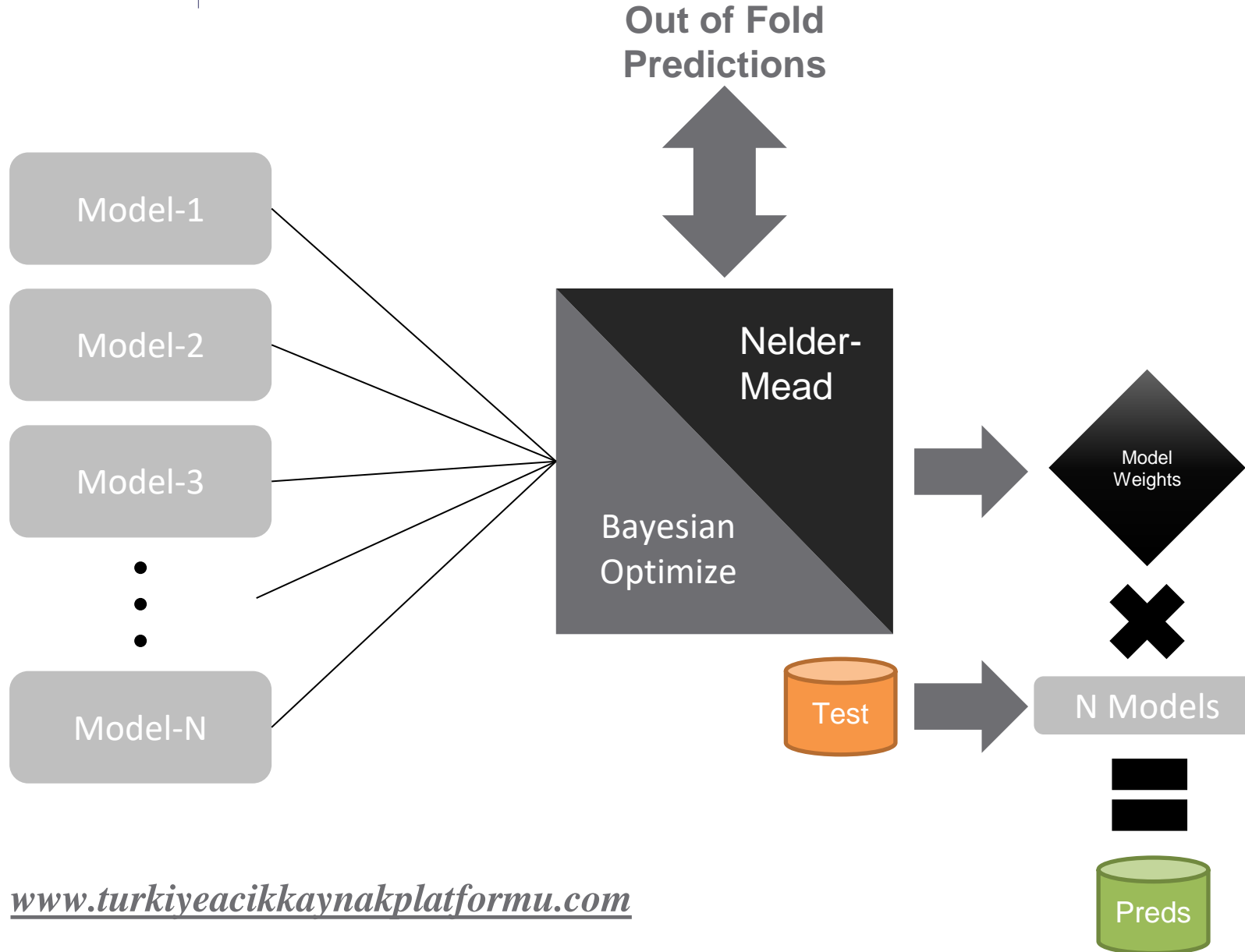
### Aşamalar

- I. Etiketsiz veri (100k) üzerinde WholeWordMasking ile self supervised eğitim.
  - II. TDDI datası üzerinde sınıflandırma eğitimi.
  - III. Elde edilen modeller ile etiketsiz verinin pseudo tahminlenmesi.
  - IV. İlk aşamadaki MLM ağırlıkları ile 100k pseudo data üzerinde sınıflandırma. (Semi Sup.)
  - V. Yeni ağırlıklar ile TDDI datasının eğitilmesi.
  - VI. Elde edilen yeni model ile etiketsiz datanın yeniden tahminlenmesi.
- ❖ Üstteki son aşamaların birden fazla kez tekrarlanması ile final modellerin oluşturulması.

## - Pseudo Labeling Sonrası Skorlar -



# - Topluluk Öğrenmesi -





## - Ensemble Learning Skorları-

*Out of 5 Fold Cross Validation*

	Precision	Recall	F1-score	Support
INSULT	0.94	0.95	0.94	2393
OTHER	0.98	0.97	0.98	3584
PROFANITY	0.97	0.97	0.97	2376
RACIST	0.98	0.98	0.98	2033
SEXIST	0.98	0.98	0.98	2081
<b>Macro Average</b>	<b>0.97</b>	<b>0.97</b>	<b>0.97</b>	<b>12467</b>

## - PROJE İŞ AKIŞI -



### Ürünleştirme

- Paketleme
- Servis Kurlumu
- Dökümantasyon

### Tahminleme

- Hata Analizi
- En Başarılı Modellerin Seçilmesi
- Bayes İstatistikleri ile Ağırlık Belirleme
- Topluluk Öğrenmesi

### Modelleme Aşaması

- Klasik Yaklaşım
- Contextual Yaklaşım
- Pseudo Labeling
- Çoklu Model Eğitimi

### Veri Aşaması

- Veri İnceleme
- Veri Önışleme
- Veri İşaretlenmesi
- Veri Artırımı



# - YOL HARİTASI -

## Prensipler:

- Compact: Kompakt
- Customizable: Özelleştirilebilir
- Creative: Yaratıcı
- Compatible: Uyumlu
- Connected: Bağlantılı

Ekstrem durumların test edilmesi ve ürünün istikrarlı hale getirilmesi.

Kütüphane etrafında bir topluluk oluşturmak için sosyal medya, forumlar ve diğer kanallar aracılığıyla farkındalığı arttırmak ve işbirliğini teşvik etmek.



Kullanıcıların istekleri doğrultusunda düzeltmeler, geliştirmeler.

Kütüphanenin işlevselliğini ve kullanışlılığını artırmak için diğer yazılım araçları veya hizmetleriyle entegrasyon.

The screenshot shows a GitHub repository page for 'NoOffense' by user 'ertugrul-dmr'. The repository is a smart and fast offensive language detector for Turkish using transformers. It includes a README, LICENSE, and various setup files. The repository has 75 commits, 3 stars, and 1 watching. The README describes the tool and its installation instructions.

**Files:**

- docs: added model weights and links by the request of competition organisers (3 days ago)
- nooffense: added external data (5 days ago)
- training: added training notebooks (last week)
- .gitattributes: cleaned weights path (last week)
- .gitignore: updated gitignore (last week)
- LICENSE: Initial commit (3 weeks ago)
- README.md: added model weights and links by the request of competition organisers (3 days ago)
- requirements.txt: added numpy requirement (2 weeks ago)
- setup.cfg: added setup config (last week)
- setup.py: updated setup.py (last week)

**README.md:**

### NoOffense

NoOffense provides a smart and fast tool for detecting offensive language in Turkish. It utilizes transformers and mostly relies on the [Hugging Face](#) library, making it fully compatible with it.

The Overfit-GM team developed this package as a part of the Türkiye Açık Kaynak Platformu NLP challenge.

### Installation Instructions

**Github:** <https://github.com/ertugrul-dmr/NoOffense>

**Hugging Face:** <https://huggingface.co/Overfit-GM>

**Google Colab:** [Getting Started Notebook](#)

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



## - KAPANIŞ -

Değerli Vaktiniz İçin Teşekkürler =)



Ertuğrul Demir

Ayberk İmir

Askar Bozcan