



ATOM

BERKAY BİÇER

YİĞİT ALICIKUŞ

ÇAĞLAR KABACA



Model Seçimi

| **Bütün etiketleri sınıflandıran büyük bir model**

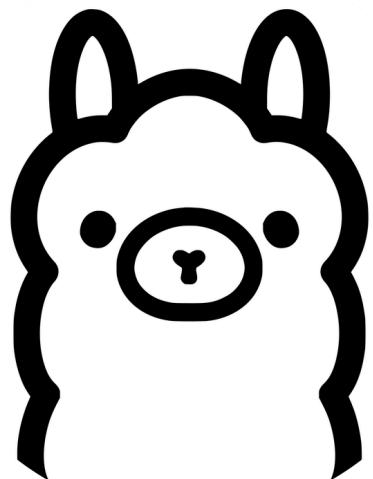
- + Basit Pipeline
- 10390'luk classification head ile yüksek hesaplama maliyeti ve zayıf genelleme riski

| **Veriyi clusterlara ayırıp her cluster için ayrı bir model**

- + Veriyi clusterlara ayırarak sınıflar arası ayrimı güçlendirme potansiyeli
- Kompleks pipeline

Veri Ön İşleme

Açık kaynak büyük dil modelleri ile adres standardizasyonu



Githubdan bulduğumuz Türkiye semt mahalle verisi üzerinde fuzzy search tabanlı feature extraction

<https://github.com/volkansenturk/turkiye-iller-ilceler>

Sınıf Dengesizliği

Toplam örnek sayısı: 848327

Ortalama örnek sayısı: 81

Minimum örnek sayısı: 40

Maksimum örnek sayısı: 536

Focal loss

Class Weight

Sınıflara ağırlık eklemek neden başarısız olmuş olabilir?

- Benzer sınıflar arasında, veri miktarı fazla olan sınıfın baskın hale gelmesi

8. EN KÖTÜ LABEL: 5539
F1 Score: 0.0741
Support: 25

Bu label için 25 örnek bulundu:

No	Gerçek Adres	167 adres		261 adres		
		Gerçek Label	Tahmin Label	Gerçek Label	Tahmin Label	Doğru?
1	6169/8 no 11 karacaoglan bornova izmir	5539	5538			
2	6169/8 no 11 karacaoglan bornova izmir	5539	5538			
3	6240 no 14 karacaoglan bornova izmir	5539	5538			
4	6240/5 no 6 karacaoglan bornova izmir	5539	5538			
5	6167/28 no 5 karacaoglan bornova izmir	5539	5538			
6	6253 no 1 karacaoglan bornova izmir	5539	5538			
7	6203/1 no 1 karacaoglan bornova izmir	5539	5538			
8	6170/2 no 7 karacaoglan bornova izmir	5539	5538			
9	6174 7 f karacaoglan bornova izmir	5539	5538			
10	bornova karacaoglan mahallesi 6240 sokağı no 16...	5539	5539			

Sınıflara ağırlık eklemek neden başarısız olmuş olabilir?

Label: 5539

6169/8 no 11 karacaoğlan bornova izmir

6240 no 14 karacaoğlan bornova izmir

6203/1 no 1 karacaoğlan bornova izmir

6240/11 no 14 karacaoğlan bornova izmir

6166 no 30 karacaoğlan bornova izmir

Label: 5538

6240/5 no 6 karacaoğlan bornova izmir

6240/55 no 12 karacaoğlan bornova izmir

6203/1 no 1 karacaoğlan bornova izmir

6240/11 no 14 karacaoğlan bornova izmir

6170 no 11/37 karacaoğlan bornova izmir

Veriyi inceleyelim

Adress	Label
Kurtuluş mahallesi Şehit er Neşet Sarıkaya Caddesi No 72 D 1 Kat: 1 Daire: 1 İncirliova - AYDIN	1695
Kurtuluş mahallesi 221 sokak no:3 kat 2	1695
TİRA CAD NO8 K1 D3 GOREN EKMEK FIRINI USTU	1695
KURTULUŞ MAH YAHYA ÜMİT ORBAY CAD.206 SK NO 13 İNCİRLİOVA	1695
Hürriyet mahallesi er Reşat Sarıkaya caddesi no 55 İncirliova	1695
Hürriyet mah 350sokak no 12	1695

Nihai Veri Ön İşleme

| Dedebaşı mah. 6100 sokak no 10 Kat 7
| daire 25



| dedebaşı mahallesi 6100 sokağı no 10

| Akarca Mah. Adnan Menderes Cad.
| 864.Sok. No:15 D.1 K.2 .



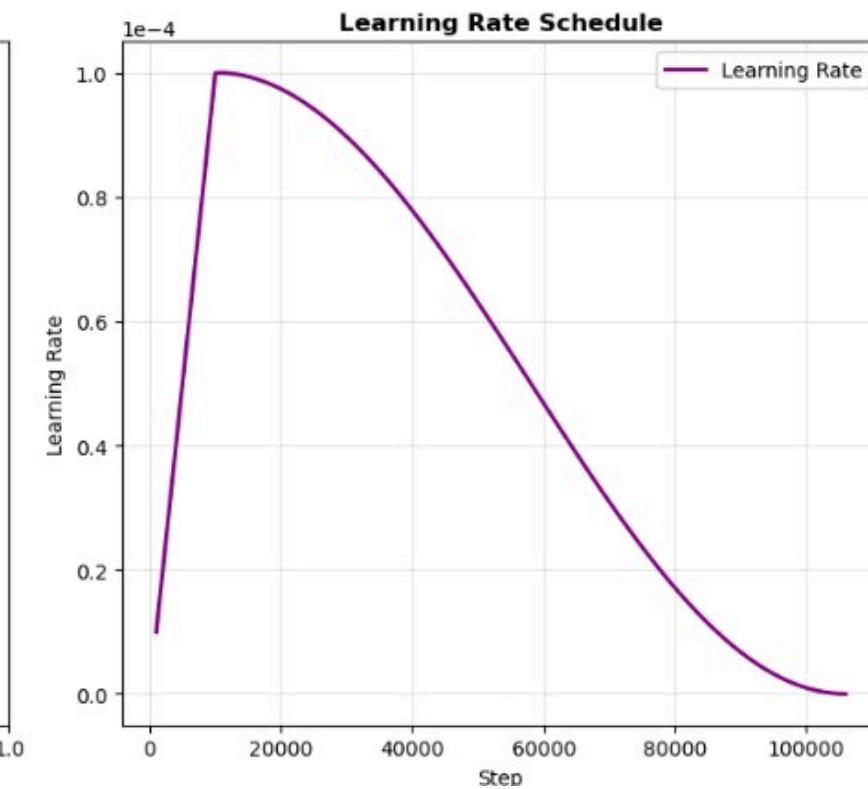
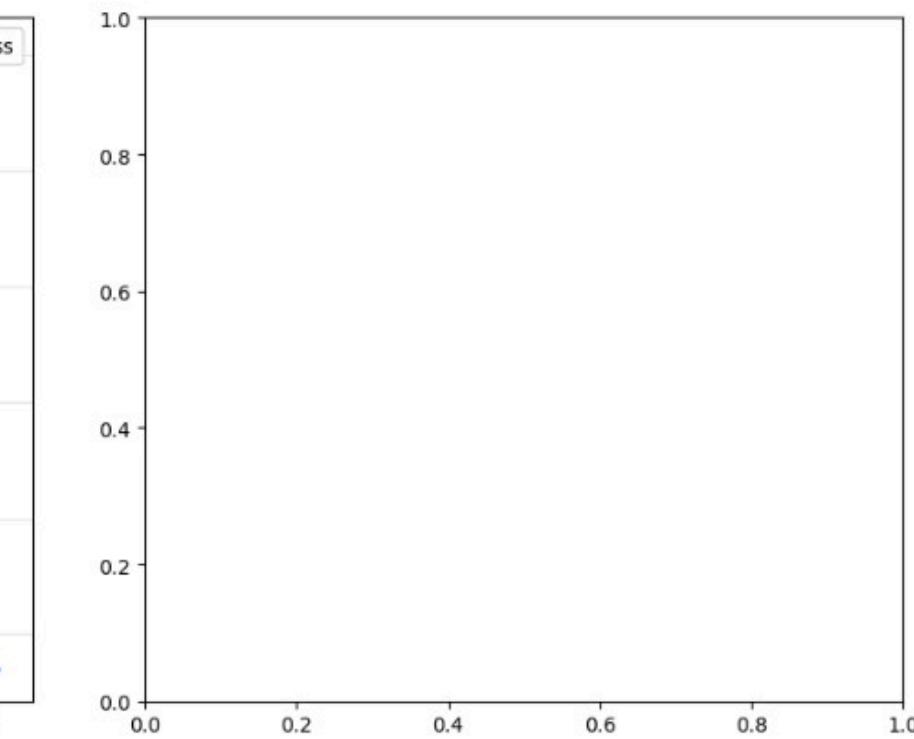
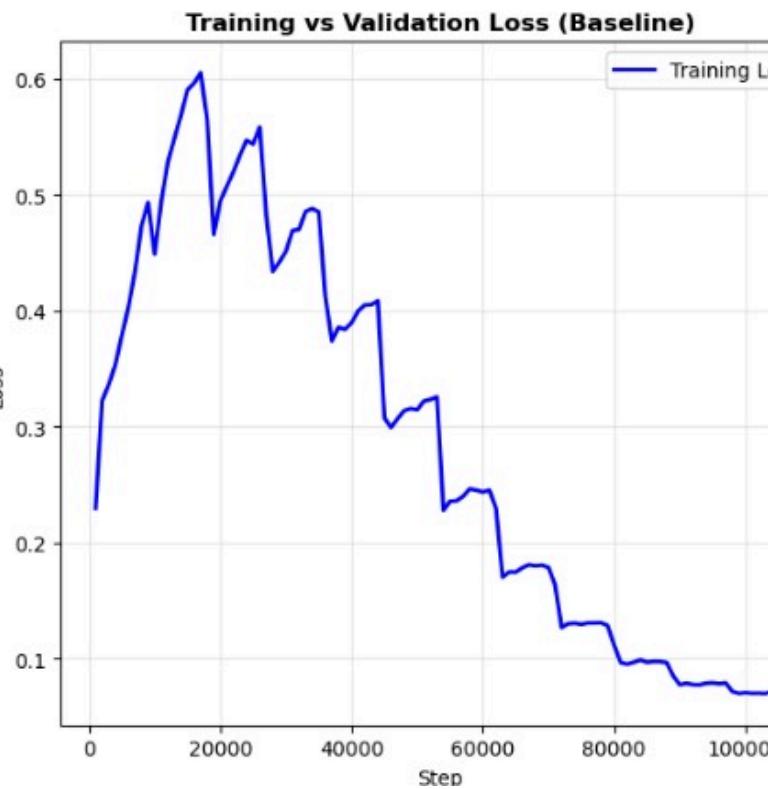
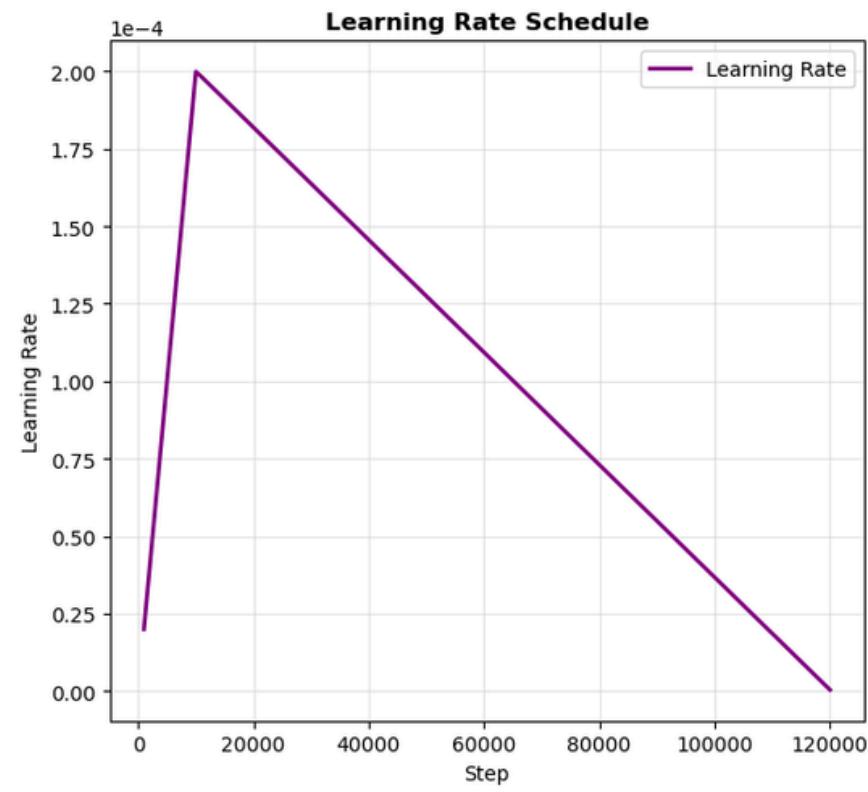
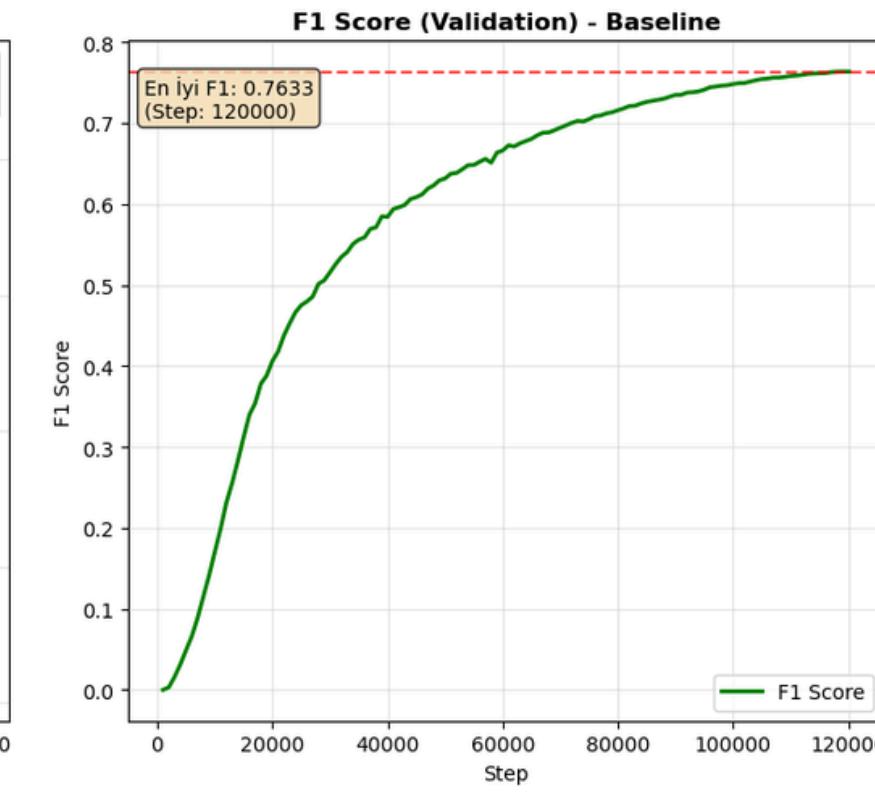
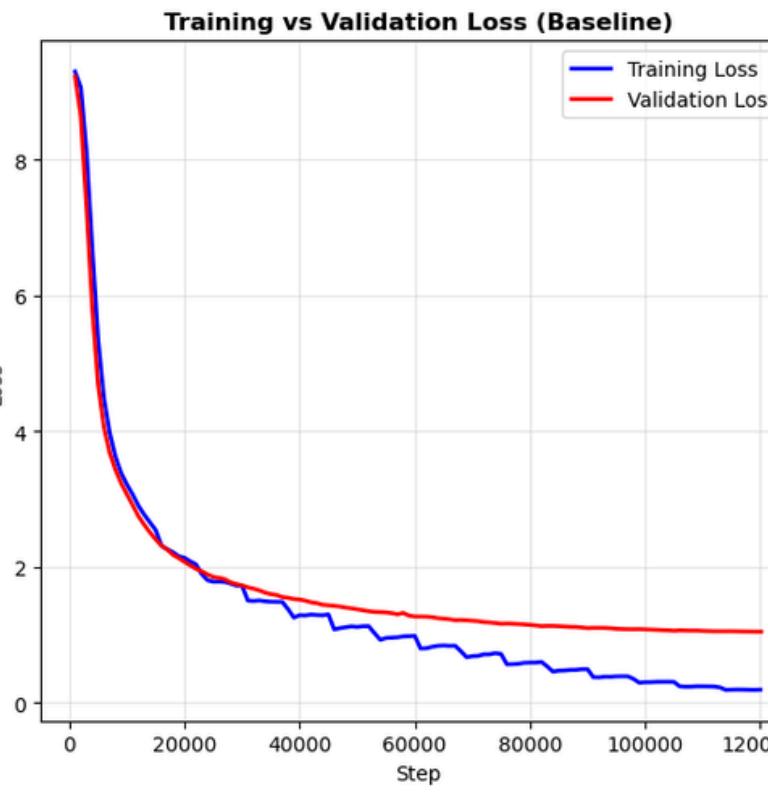
| akarca mahallesi adnan menderes
| caddesi 864 sokağı no 15

dbmdz/bert-base-turkish-cased

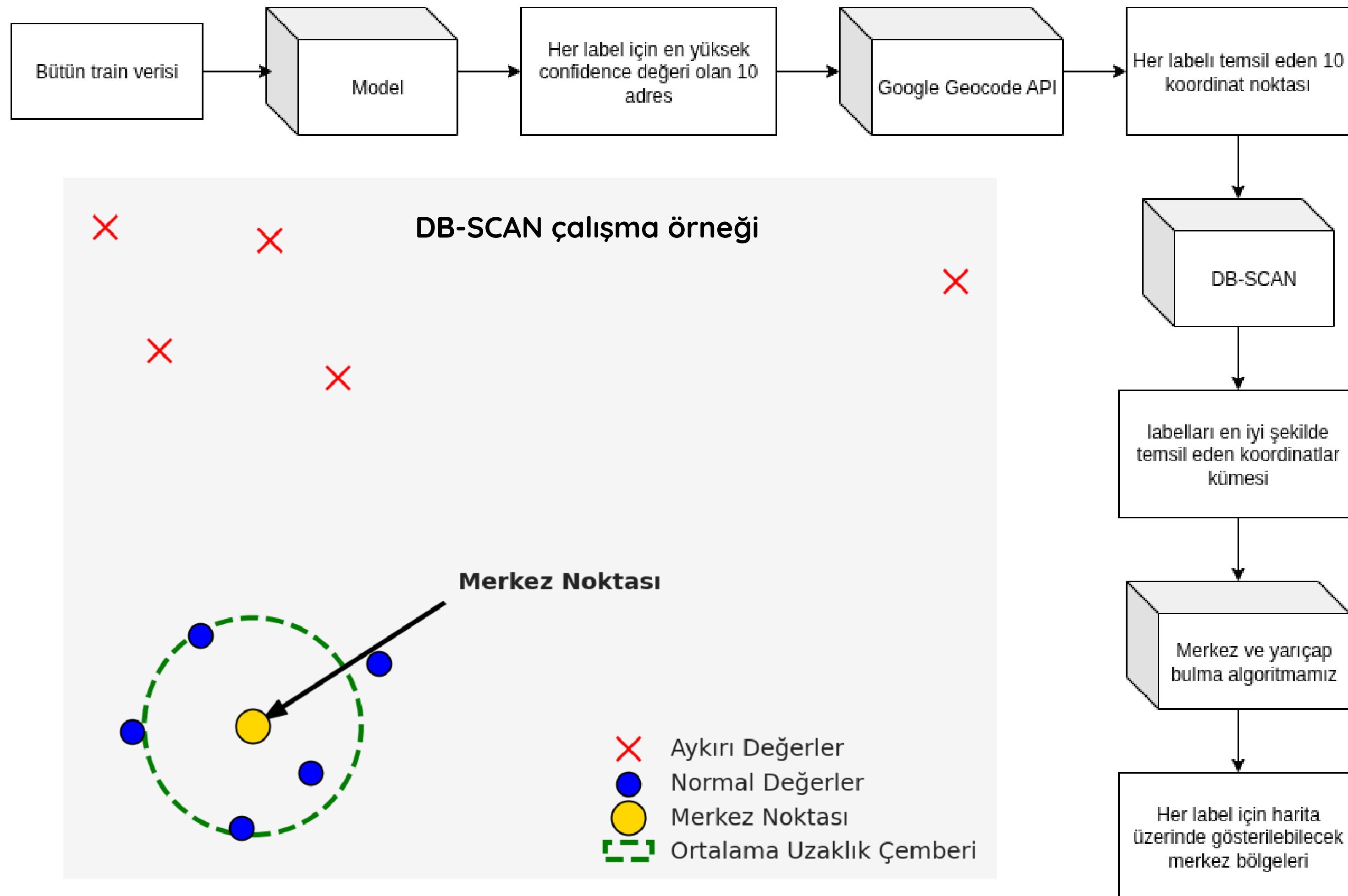
```
training_args = TrainingArguments(  
    output_dir='./results',  
    num_train_epochs=16,  
    per_device_train_batch_size=96,  
    per_device_eval_batch_size=96,  
    learning_rate=2e-4,  
    warmup_steps=10000,  
    weight_decay=0.01,  
    lr_scheduler_type="linear",  
    eval_strategy="steps",  
    eval_steps=1000,  
    save_strategy="steps",  
    save_steps=1000,  
    load_best_model_at_end=True,  
    save_total_limit=3,  
    metric_for_best_model="f1",  
    greater_is_better=True,  
    report_to=[],  
    logging_steps=1000,  
    #eval_delay=10000,  
    max_grad_norm= 1.0, # Gradient clipping  
    fp16=True,  
    seed=42  
)
```

CONFUSION MATRIX ANALİZİ:

- ✓ Toplam Doğru Tahmin (TP): 98,146
- ✗ Toplam Yanlış Tahmin (FP+FN): 29,080
- 📊 Toplam Sample: 127,226
- 🎯 Accuracy: 0.7714



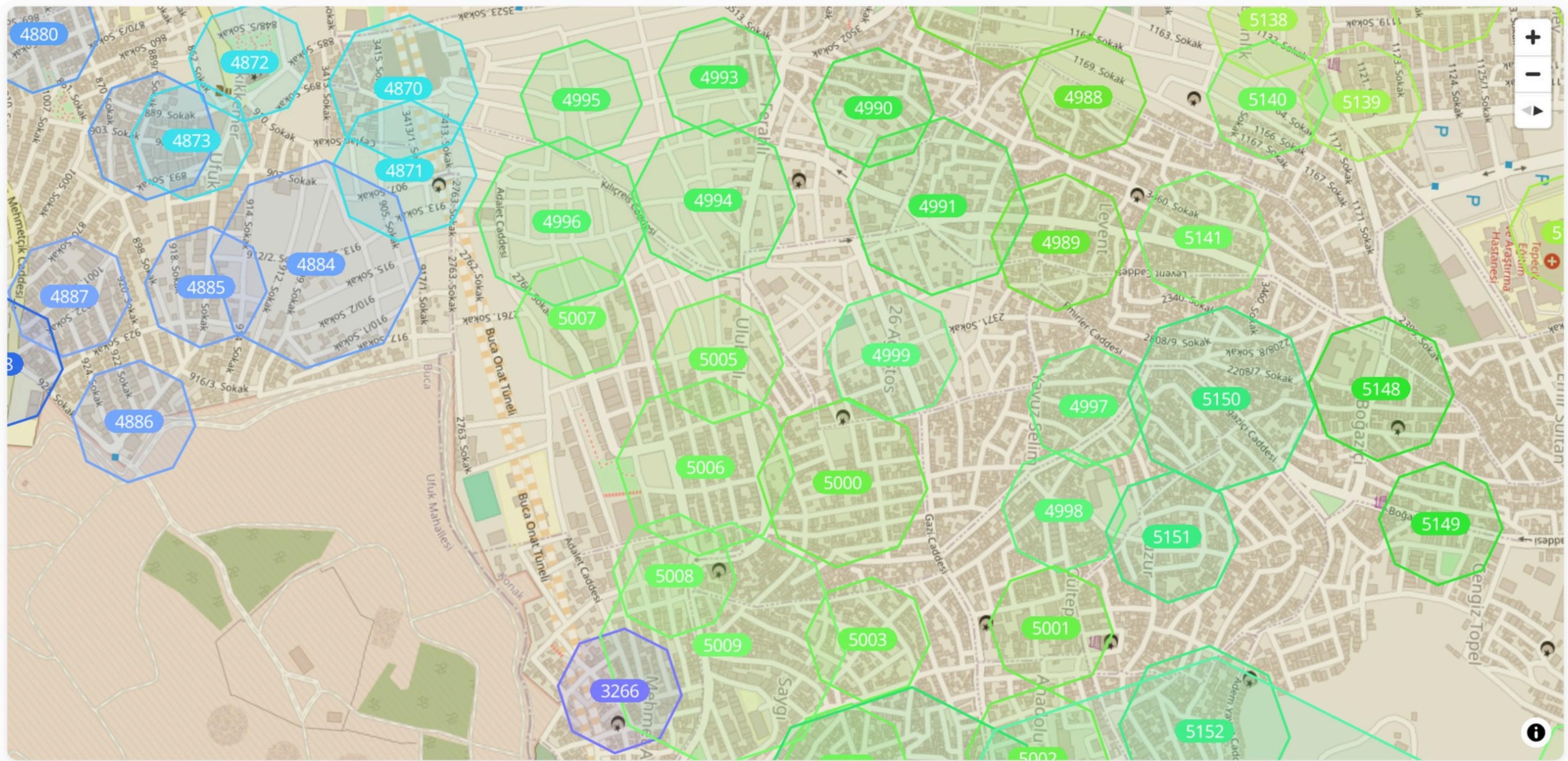
Kaggle Private f1 skoru: 0.78806



Demo: <https://hepsiburada.vendigo.app>

Harita Görünümü

10390 KÜME



**Dinlediğiniz
için teşekkür
edərəz**

ATOM