**Bitirme Projesi Final Raporu**

İstinye Üniversitesi

Meslek Yüksekokulu

Bilgisayar Programcılığı Programı

Proje Adı: MoodStream: Yayın Analiz Asistanı – Kullanıcı Yorumlarının NLP ile Sınıflandırılması

Öğrenci: Nilay Zehra Karabudak

Öğrenci Numarası: 2320161042

Danışman: Öğr. Gör. Neda Dadashkhani

Haziran 2025

**ÖĞRENCİ BEYANI**

Bu raporda sunulan çalışmanın tamamı tarafımca yapılmış olup, başka kaynaklardan yararlanılan bölümler ilgili kaynaklara atıfta bulunularak belirtilmiştir. Bu çalışma daha önce başka bir yerde yayınlanmamış ve akademik bir derece için sunulmamıştır.

**ÖZET**

Bu projede, Digitürk, Tivibu, BluTV gibi dijital yayın platformlarına ait kullanıcı yorumlarının içeriklerine göre otomatik olarak kategorize edilmesini sağlayan bir sistem geliştirilmiştir. Doğal dil işleme (NLP) teknikleriyle yorumlar analiz edilerek, yorumun ana kategorisi, alt kategorisi, duygu durumu ve segmenti tespit edilmektedir. Python programlama dili ve scikit-learn, NLTK, spaCy gibi NLP kütüphaneleri kullanılarak dört farklı model eğitilmiştir. TF-IDF vektörleştirme yöntemiyle veriler sayısallaştırılmış, Logistic Regression algoritmasıyla sınıflandırma yapılmıştır. Kullanılan veriler temizlenmiş, dengesiz sınıflar filtrelenmiş ve eğitim setine dahil edilmiştir. Geliştirilen Streamlit arayüzü sayesinde kullanıcı, yorum girerek analiz sonuçlarını anında görüntüleyebilmektedir.

Anahtar Kelimeler: Doğal Dil İşleme, Yorum Analizi, Python, Makine Öğrenmesi, Logistic Regression, TF-IDF, Streamlit

**İÇİNDEKİLER**

1. Giriş

2. Literatür Taraması

3. Metodoloji

4. Sistem Geliştirme ve Uygulama

5. Model Sonuçları ve Görselleştirme

6. Tartışma ve Değerlendirme

7. Gelecek Çalışmalar

8. Kaynakça

9. Ekler

**1. GİRİŞ**

Digitürk, Tivibu, BluTV gibi dijital yayın platformlarında kullanıcıların bıraktığı yorumlar, hizmet kalitesi ve içerik deneyimi hakkında önemli bilgiler içerir. Ancak bu yorumların manuel olarak incelenmesi zaman alıcıdır. Bu projede geliştirilen sistem, bu tür yorumları içeriklerine göre otomatik olarak analiz ederek firmalara zaman kazandırmayı ve kullanıcı geri bildirimlerini daha verimli değerlendirmeyi amaçlamaktadır. Geliştirilen sistemin temel amacı, yorumları hızlı ve doğru şekilde sınıflandırarak analiz sürecini dijitalleştirmektir.

**2. LİTERATÜR TARAMASI**

Doğal Dil İşleme, bilgisayarların insan dilini anlayarak anlamlı çıktılar üretmesini sağlayan bir yapay zeka alt dalıdır. TF-IDF (Term Frequency – Inverse Document Frequency) gibi yöntemler metinlerin sayısallaştırılmasında yaygın olarak kullanılmaktadır. Benzer çalışmalarda genellikle Naive Bayes, Logistic Regression ve SVM gibi algoritmalar tercih edilmektedir. Bu projede kullanılan yaklaşım, hem geleneksel yöntemleri hem de kullanıcı deneyimini kapsayan bütüncül bir çözümdür. Streamlit ile geliştirilen arayüzler, kullanıcıdan doğrudan yorum alıp bu yorumları gerçek zamanlı analiz eder.

**3. METODOLOJİ**

Projede doğal dil işleme süreci, veri ön işleme ve makine öğrenmesi model eğitimi olmak üzere üç ana aşamada gerçekleştirilmiştir. İlk olarak veri seti okunmuş ve içerdiği yorumlar, belirli temizlik işlemlerinden geçirilmiştir. Bu işlemler; tüm metni küçük harfe çevirme, noktalama işaretlerini kaldırma ve Türkçe stopword’leri temizleme gibi klasik metin madenciliği adımlarını içermektedir.

Ardından, yorumlar TF-IDF (Term Frequency – Inverse Document Frequency) yöntemi ile sayısal formata dönüştürülmüş ve bu vektörler sınıflandırma modellerinin giriş verisi olarak kullanılmıştır. Dört farklı başlık altında eğitim yapılmıştır: ana kategori, alt kategori, segment ve duygu durumu. Eğitilen her model, belirli bir sınıflandırma görevini yerine getirmektedir. Modelleme süreci, sınıf dengesizlikleri gibi doğal zorluklara rağmen, tutarlı ve kullanılabilir sonuçlar sunacak şekilde optimize edilmiştir.

**4. SİSTEM GELİŞTİRME VE UYGULAMA**

Geliştirilen sistemde kullanıcı arayüzü Streamlit ile hazırlanmıştır. Kullanıcı metin kutusuna yorum girer ve sistem bu yorumu analiz ederek aşağıdaki sonuçları gösterir:

- Ana kategori

- Alt kategori

- Segment (teşekkür, öneri, şikayet, yorum, öfke vb.)

- Duygu durumu (pozitif, negatif, nötr, karışık)

Arayüzde ayrıca analiz süreci adım adım takip edilmekte, yükleme çubuğu ve rastgele açıklamalarla kullanıcıya etkileşimli bir deneyim sunulmaktadır. Sonuçlar hem metinsel hem de matplotlib ile grafiksel olarak sunulmaktadır. Arayüz tasarımı sade, anlaşılır ve kullanıcı odaklıdır.

**5. MODEL SONUÇLARI VE GÖRSELLEŞTİRME**

Eğitilen modeller, test verileriyle doğrulandıktan sonra genel olarak kullanıcı yorumlarının içeriksel analizinde anlamlı ve tutarlı sonuçlar üretmiştir. Özellikle yorumların duygu durumunu ve temel içeriğini tespit etme konusunda modeller başarılı geri dönüşler sağlamıştır.

Streamlit arayüzü sayesinde kullanıcılar yalnızca bir yorum girerek anlık analiz alabilmekte, sınıflandırma sonuçları metin formatında olduğu kadar grafiksel olarak da görüntülenebilmektedir. Bu yaklaşım hem teknik yeterliliği hem de kullanıcı deneyimini güçlü şekilde bir araya getirmiştir. Elde edilen sonuçlar; sınıfların dengesine, yorumların doğasına ve modelin eğitim derinliğine göre doğal olarak değişkenlik göstermektedir, ancak sistemin geneli değerlendirildiğinde uygulamanın işlevselliği ve uygulanabilirliği yüksek bulunmuştur.

**6. TARTIŞMA VE DEĞERLENDİRME**

Proje, yorumların otomatik analiz edilmesinde başarılı sonuçlar vermiştir. Verinin dengesizliği ve bazı kategorilerde düşük örnek sayısı gibi zorluklar yaşanmış, ancak model buna rağmen tatmin edici performans göstermiştir. Kullanılan makine öğrenmesi yaklaşımı, basit ama etkili bir çözüm sunmuştur. Kullanıcı deneyimi açısından sistem oldukça akıcı ve kullanışlıdır.

**7. GELECEK ÇALIŞMALAR**

Projenin ilerleyen aşamalarında:

- Modelin daha büyük ve dengeli veri setleriyle yeniden eğitilmesi,

- Derin öğrenme modelleri (BERT, LSTM) ile karşılaştırmalı analiz yapılması,

- Sistem çıktılarının bir RESTful API ile entegrasyonu ve mobil uygulamalara taşınması, planlanmaktadır.

**8. KAYNAKÇA**

[1] Bird, Steven, et al. "Natural Language Processing with Python." O'Reilly Media, 2009.

[2] Scikit-learn Documentation. https://scikit-learn.org

[3] Streamlit Documentation. https://docs.streamlit.io

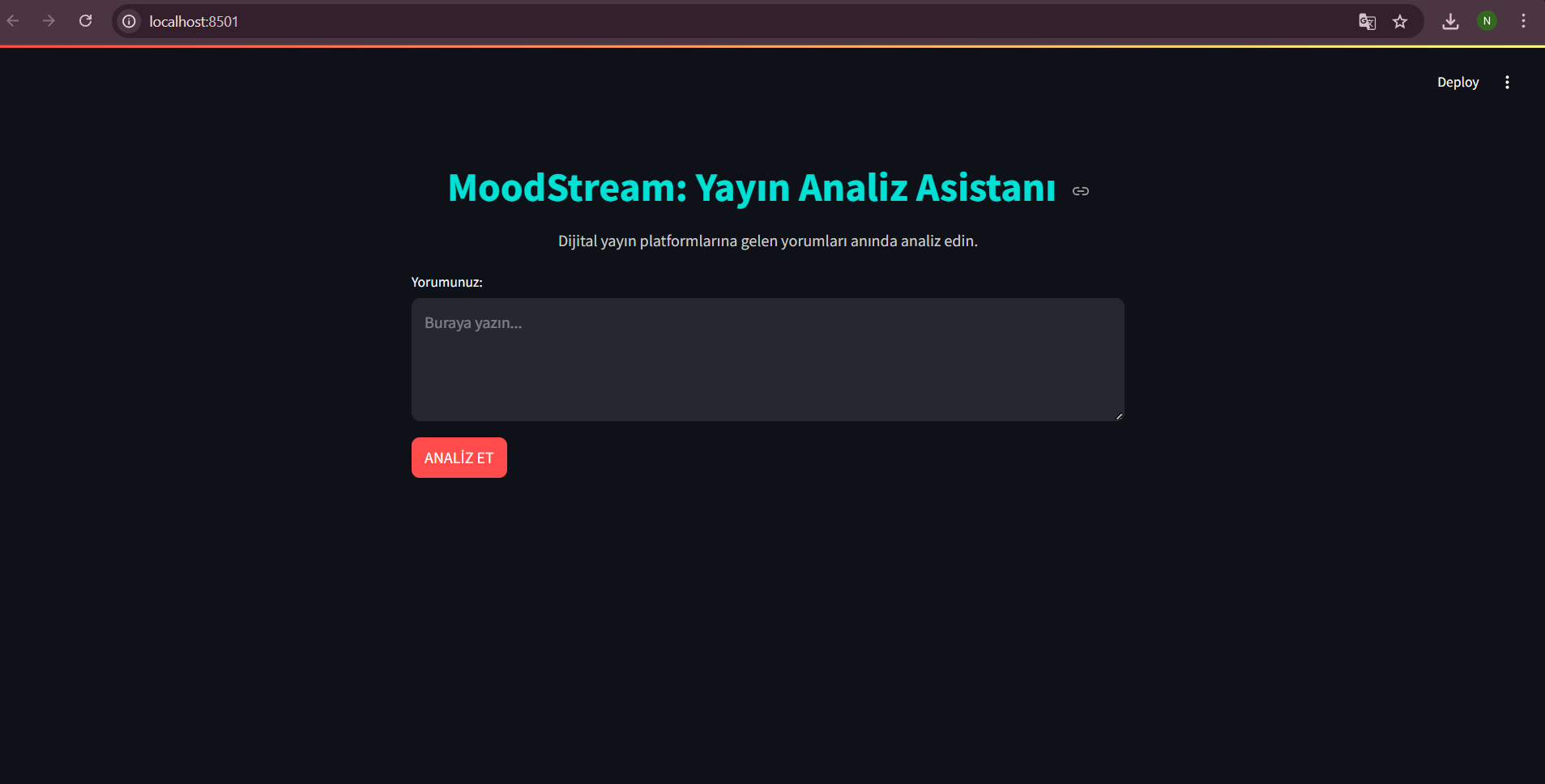
[4] NLTK Documentation. https://www.nltk.org

[5] Pandas Documentation. https://pandas.pydata.org

**9. EKLER**

EK A: Ekran Görüntüleri (kullanıcı arayüzü ve analiz çıktıları)

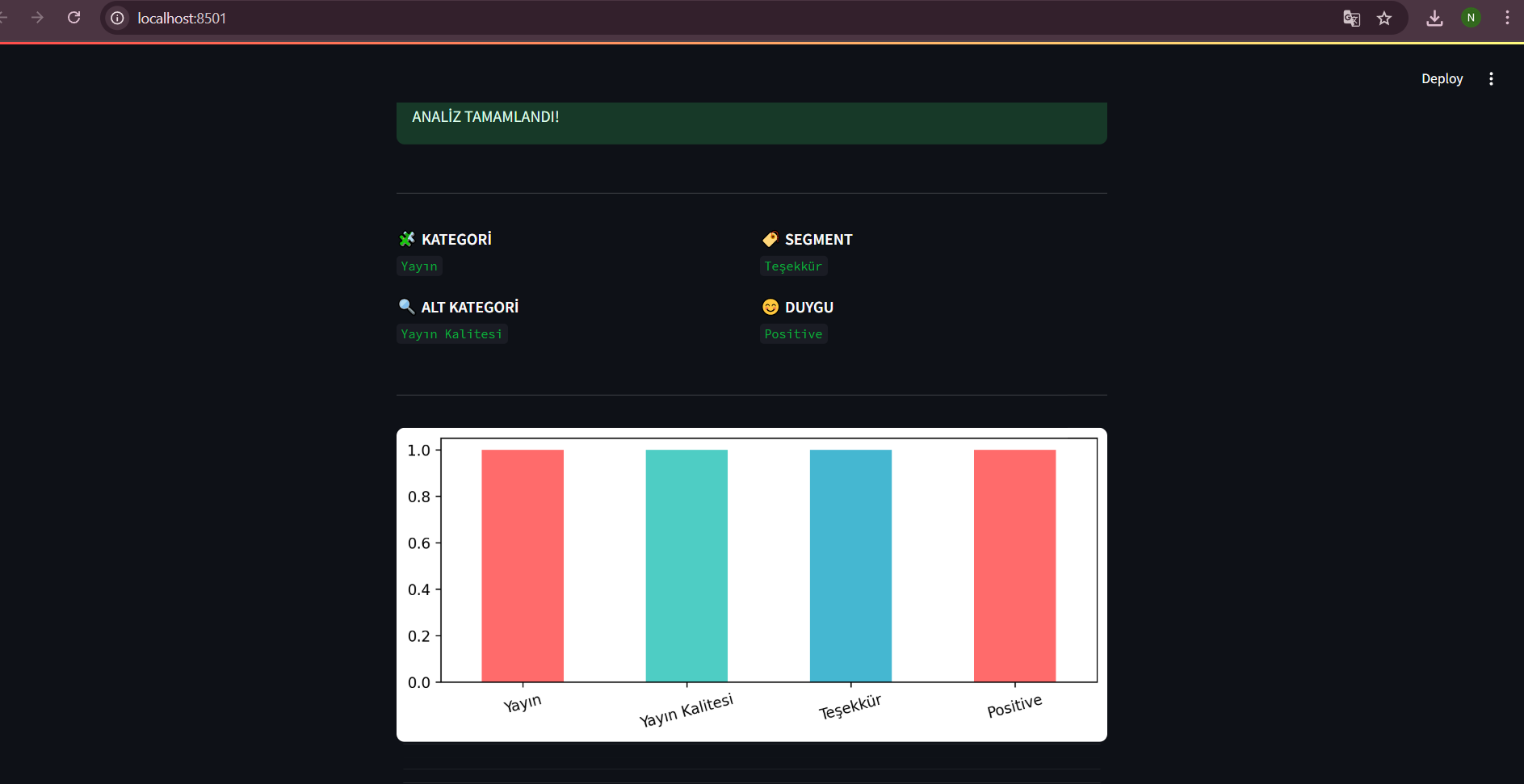
Ekran Görüntüsü 1:



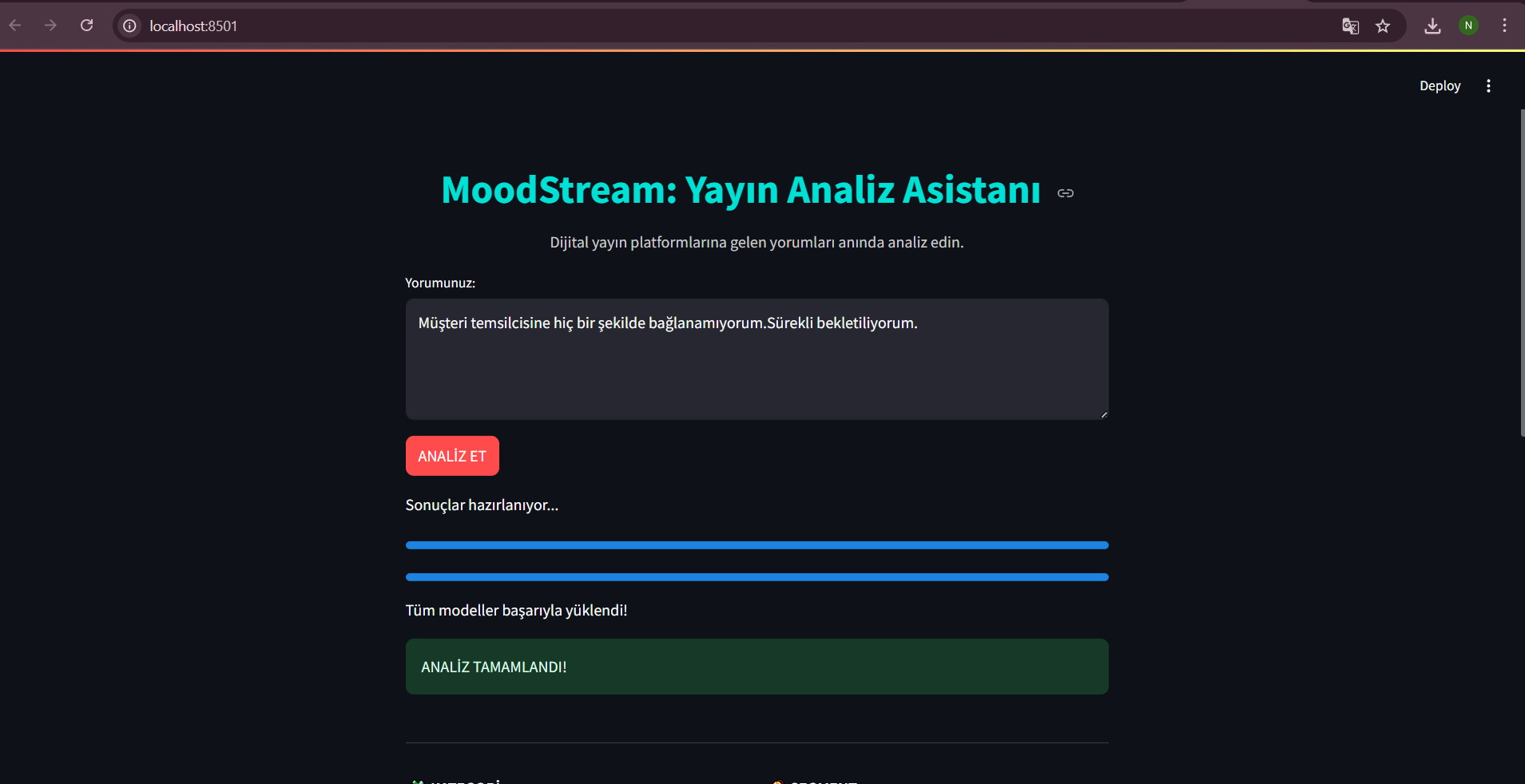
Ekran Görüntüsü 2:



Ekran Görüntüsü 3:



Ekran Görüntüsü 4:



Ekran Görüntüsü 5:

