

# Web Trafik Loglarına Dayalı Yapay Zeka Destekli Soru-Cevap Sistemi Geliştirme Projesi Raporu

## Giriş

Bu projede, bir web sitesinin trafik loglarını kullanarak doğal dilde sorulara yanıt verebilen bir soru-cevap (Q&A) sistemi geliştirilmiştir. Sistem, Retrieval-Augmented Generation (RAG) modeli üzerine kurulmuş olup, kullanıcı sorularını anlamak ve log verilerinden çıkarımlar yaparak doğru ve alakalı yanıtlar üretmeyi amaçlamaktadır.

## Proje Adımları ve Uygulamalar

### Aşama 1: Veri Hazırlığı ve Ön İşleme

- **Veri Seti:** Web trafik log dosyası csv formatında kullanılmıştır.
- **Veri Ekstraksiyonu:** IP adresi, erişilen URL, zaman damgası, HTTP durum kodu gibi önemli özellikler çıkarılmıştır.
- **Vektör Temsili:** Her log kaydı, Embeddings modeli kullanılarak bir vektör uzayına dönüştürülmüştür.
- **Vektör Veritabanı:** Dönüştürülen vektörler, hızlı arama işlemleri için Chroma vektör veri tabanına yüklenmiştir.

### Aşama 2: RAG Modelinin Kurulumu

- **Bilgi Alma:**
  - Kullanıcı sorusu da aynı dil modeli ile vektöre dönüştürülür.
  - Dönüştürülen soru vektörü, Chroma veri tabanında en yakın komşuluk araması ile log kayıtlarına eşlenir.
  - En alakalı görülen birkaç log kaydı seçilir.
- **Jeneratif Model:**
  - Seçilen log kayıtları ile birlikte kullanıcı sorusu, LLM modeline beslenir.
  - LLM, verilen bağlam içinde kullanıcı sorusuna en uygun yanıtı üretir.

### Aşama 3: Web Arayüzü

- **Kullanıcı Arayüzü :** Gradio kütüphanesi kullanılarak bir web arayüzü geliştirilmiştir.

### Aşama 4: Performans Değerlendirmesi

- **Doğruluk:** İnsan değerlendiricileri tarafından sistemin ürettiği yanıtların doğruluğu ve alakalılığı değerlendirilmiştir.
- **Kapsamlılık:** Sistemin farklı türde soruları yanıtlama yeteneği test edilmiştir.
- **Hız:** Sistemin yanıt verme süresi ölçülmüştür.
- **Prompt :** Modelin istenilen cevaplara daha uygun yanıtlar vermesi için Prompt Engineering yapılmış ve sonuçlarda daha olumlu yanıtlar görülmüştür.

- **Model Seçimi** : Cevapların doğrulukları seçilen model ile yüksek korelasyon içermektedir. OpenAI modelleri ile çalışmalar yapılmış ve en doğru cevaplar elde edilmiştir.

#### **Aşama 5: Çalışma Prensipleri:**

- Log dosyası kullanıcıdan alınır.
- Soru Embeddings modeli ile vektörleştirilir.
- Chroma database 'e kaydedilir
- Kullanıcı sorusunu sorar ve vektörleştirilen soru Database'de aratılır. En yakın cevaplar filtrelenir.
- Filtrelenen log kayıtları LLM'e verilir.
- LLM, kullanıcı sorusuna cevap üretir.

### **Sonuçlar ve Tartışma**

Geliştirilen sistem, basit sorulara oldukça başarılı yanıtlar verebilmektedir. Özellikle log verilerinde sıkça geçen bilgilere dayalı sorular için doğru sonuçlar elde edilmiştir. Ancak, daha karmaşık veya belirsiz sorularda sistemin zorlandığı gözlemlenmiştir.

### **Geleceğe Yönelik Öneriler**

- **Daha Güçlü Dil Modeli**: Daha büyük ve gelişmiş bir dil modeli kullanarak daha iyi sonuçlar elde edebiliriz.
- **Metadata** : Metadata verileri yeni bir LLM chain kullanılarak eklenebilir. Veriler model için daha anlaşılır hale getirebilir.
- **Geçmiş** : Modele sorulan sorular ve cevaplar kaydedildi. Bu kayıtlar modele tekrardan verilerek sorulan soruların geçmişteki cevapları ile kıyaslanarak daha birleşik olması sağlanabilir

### **Sonuç**

Bu proje, web trafik loglarını kullanarak bir soru-cevap sistemi geliştirmenin mümkün olduğunu göstermiştir. Sistem, belirli bir alandaki sorulara hızlı ve doğru yanıtlar verebilme potansiyeline sahiptir.