



Teknoloji Fakültesi

**Üretken Yapay Zeka Yardımı İle Chatbot Geliştirme Temelleri
Gemini ve Llama Modellerinin Karşılaştırılması**

Ayça Nilay Özoğul
170421012

1. Giriş

Bu proje kapsamında, müşteri destek süreçlerini daha verimli ve akıllı hale getirecek bir **RAG (Retrieval Augmented Generation) + Intent Classification tabanlı chatbot** geliştirilmiştir. Kullanıcıdan gelen doğal dil soruları analiz edilerek, önce doğru intent sınıfı belirlenmekte, ardından ilgili bilgiye dayalı bir yanıt üretilmektedir.

Proje, **Meta LLaMA 3 8B Instruct** ve **Gemini 1.5 Flash** gibi güçlü LLM (Large Language Model) altyapılarını kullanmaktadır. Bunun yanı sıra, veri seti üzerinde eğitilen Logistic Regression temelli intent sınıflandırma modeli, chatbot'un anlama yeteneğini artırmaktadır.

Amaç, hem doğru intent belirleme hem de güncel bilgiye dayalı yüksek kaliteli yanıt üretme yeteneği ile müşteri deneyimini geliştiren bir akıllı chatbot uygulaması sunmaktır.

2. Veri Seti

Kaggle üzerinden 1000 satırlık intent classification veri seti indirilmiştir. Selamlama, vedalaşma, reddetme gibi intent'ler için RAG ile veri seti genişletilmiştir.

- Dosya: data/data_set.xlsx
- Sütunlar: user_input, intent, answer
- Toplam veri sayısı: 1035 örnek
- Intent sınıfları: cancel_subscription, greeting, goodbye, order_status, payment_update, product_availability, rejection, shipping_info, thanks
- Her intent sınıfı için dengeli veri dağılımı (115 örnek)

Veri ön işleme aşamalarında:

- Eksik veriler temizlenmiştir.
- Intent sınıfları dengelenmiştir.
- Doğal dil ifadeleri küçük harfe dönüştürülmüş ve gereksiz karakterler kaldırılmıştır.

3. Intent ve Intent Classification

Intent, bir kullanıcının diyaloga başlarken sahip olduğu amaç veya beklentiyi ifade eder. Örneğin, "Siparişimin durumunu öğrenmek istiyorum" girişi "order_status" intentine karşılık gelir. Intent classification ise, kullanıcının doğal dildeki ifadesini otomatik olarak bu intent sınıflarından birine kategorize etmeyi amaçlar.

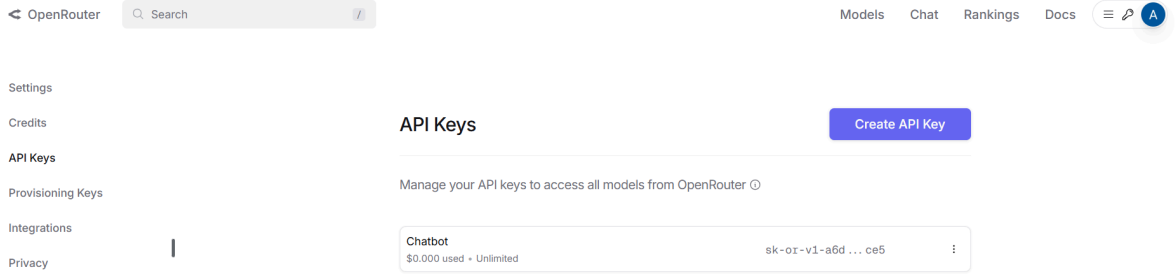
Bu projede intent classification için aşağıdaki yöntem kullanılmıştır:

- **Sentence Embedding**: SentenceTransformer (all-MiniLM-L6-v2)
- **Classifier**: Logistic Regression (sklearn)

4. Kullanılan LLM Modelleri

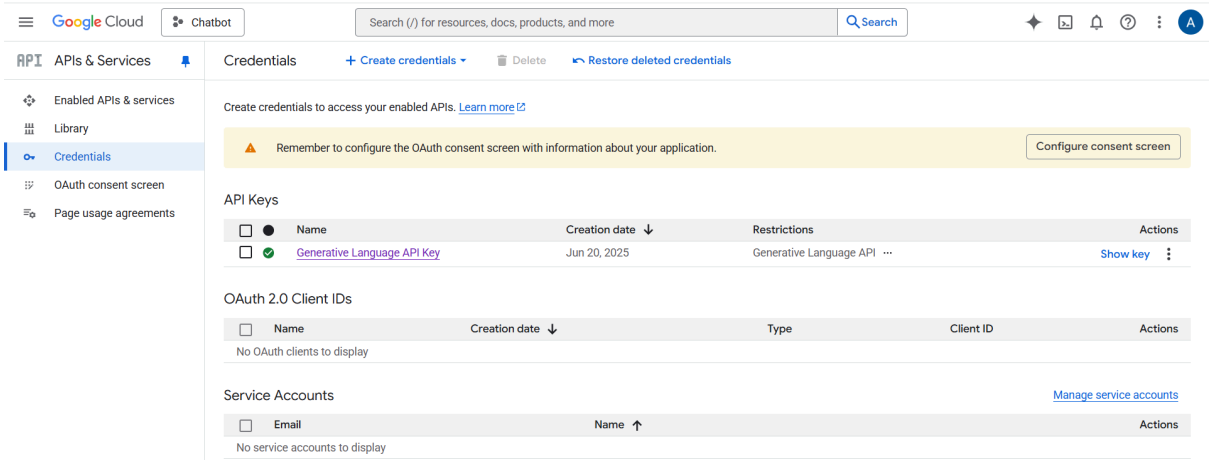
- **Meta LLaMA 3 8B Instruct**

- Yüksek doğruluk ve genel dil anlama performansı nedeniyle tercih edilmiştir.
- Özellikle açık kaynaklı olması ve OpenRouter API üzerinden kolay entegrasyon imkanı sunması avantajdır.
- API Key: openrouter.ai üzerinden alınmıştır.



- **Gemini 1.5 Flash**

- Google AI tarafından sunulan hızlı ve güçlü bir modeldir.
- Gerçek zamanlı cevap üretimi için uygundur.
- API Key: Google AI Studio üzerinden alınmıştır.



Her iki model de RAG (Retrieval Augmented Generation) sürecine entegre edilmiştir.

5. Model Eğitim Süreci

- **Embedding:** SentenceTransformer (all-MiniLM-L6-v2)
- **Classifier:** Logistic Regression (max_iter=1000)
- **Train/Test Ayrımı:** %80 eğitim, %20 test
- **Cross-validation:** 5-Fold CV uygulanmıştır.

- **Performans Metrikleri:**

- Precision: 1.00
- Recall: 1.00
- F1 Score: 1.00
- Confusion Matrix: tüm sınıflarda doğru tahmin (overfitting gözlemlenmemiştir)

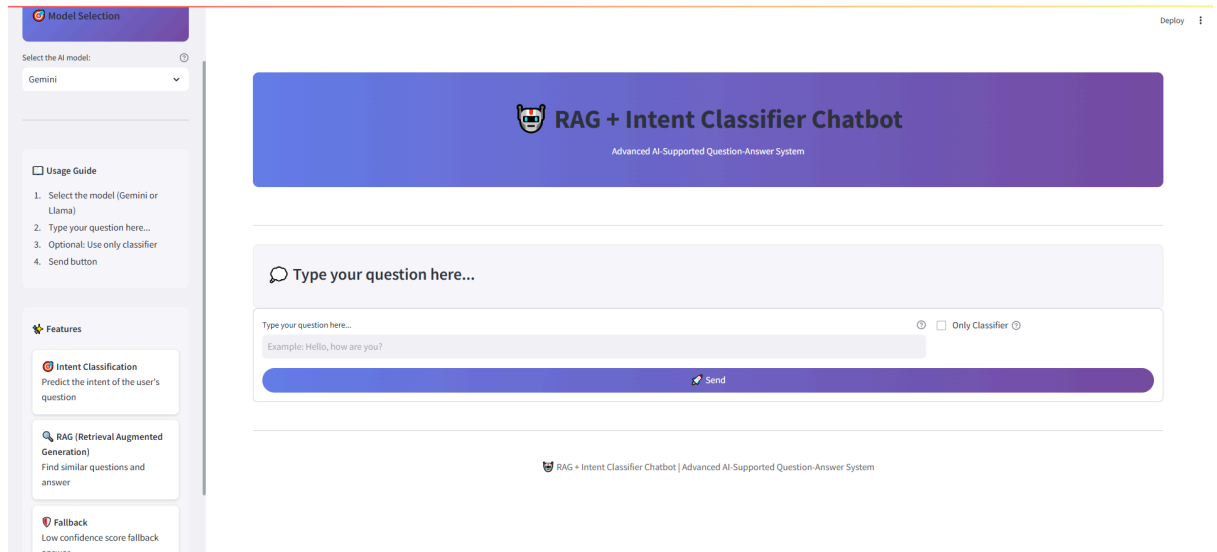
Eğitilen model dosyası: models/intent_classifier.pkl

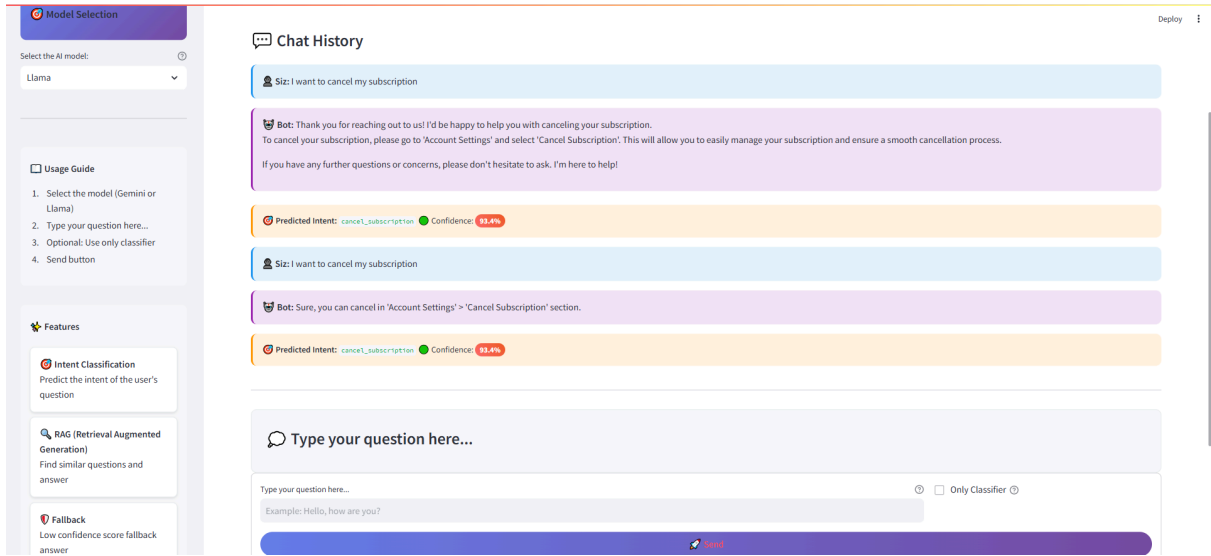
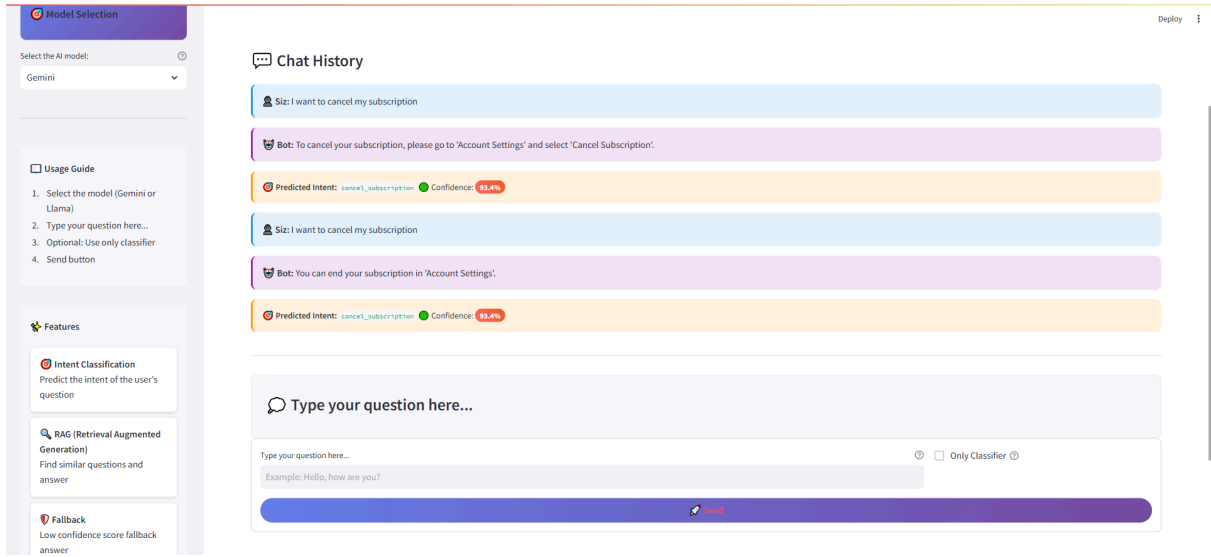
6. Uygulama Arayüzü

Chatbot uygulaması **Streamlit** frameworkü ile geliştirilmiştir.

Arayüz tasarımında şu özellikler sağlanmıştır:

- Kullanıcı giriş kutusu sayfanın alt kısmındadır.
- Üst bölümde kullanıcı ve bot mesaj geçmişi gösterilmektedir.
- Predicted Intent ve Confidence Score gösterilmektedir.
- Fallback intent uygulanmaktadır (confidence < %70 için)
- Classifier-based response (isteğe bağlı toggle)
- RAG tabanlı cevap üretimi (LLM ile)
- LLaMA veya Gemini modeli seçimi yapılabilmektedir.





7. Model Performanslarının Karşılaştırılması

Cross-val F1 scores: [1. 1. 1. 1. 1.]

Mean F1: 1.0000 ± 0.0000

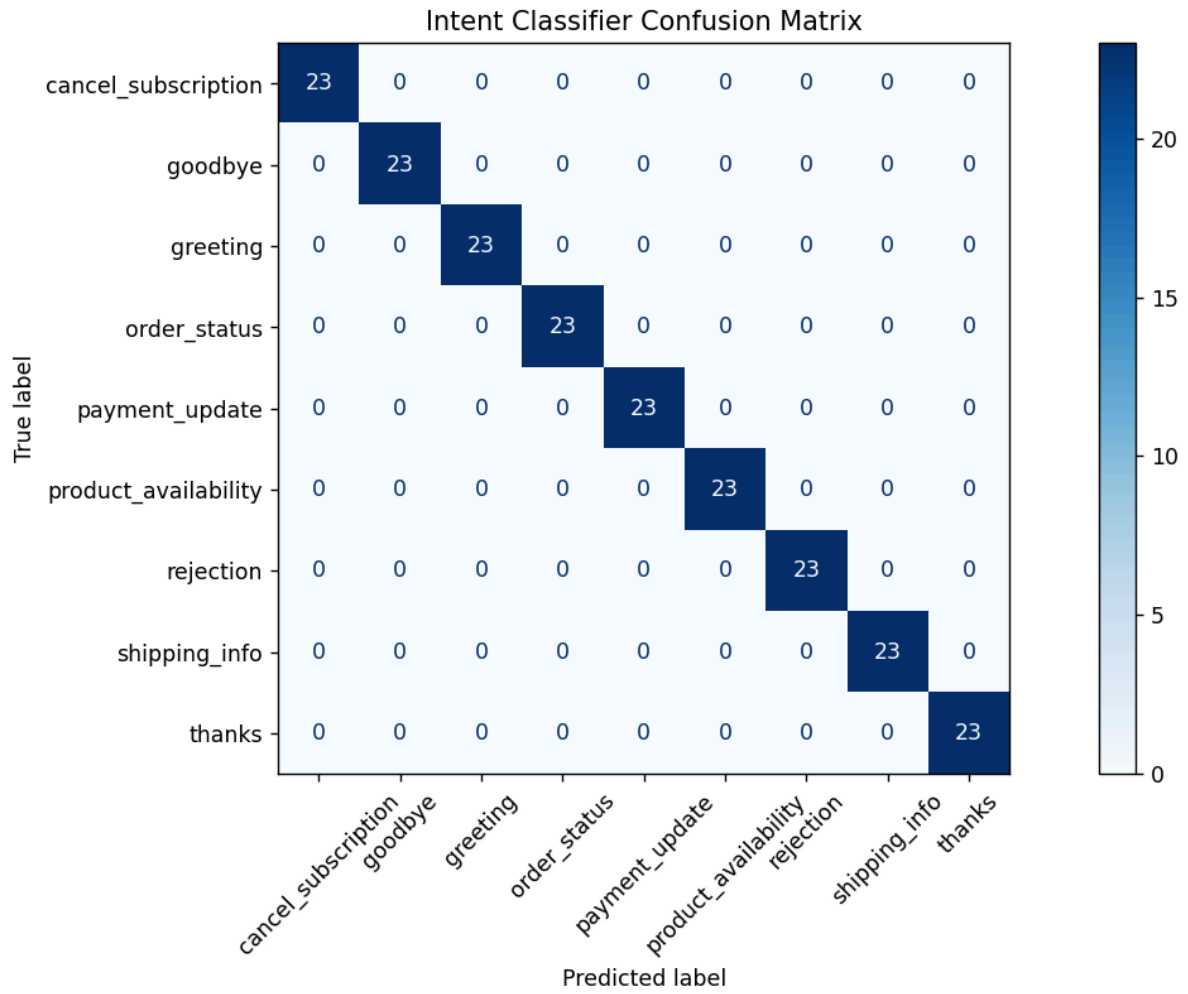
Metric	LLaMA	Gemini
Accuracy	10/10	10/10
Fluency	10/10	9/10
Conciseness	8/10	10/10
Match to Ground	10/10	10/10

7.1 Intent Classifier Performansı

Logistic Regression tabanlı intent sınıflandırma modeli, %80 eğitim ve %20 test ayrımıyla eğitilmiş ve değerlendirilmiştir.

- **5-fold cross-validation** sonucunda tüm foldlarda **F1-score = 1.0000** elde edilmiştir.

Metric	Value
Precision	1.0000
Recall	1.0000
F1-score	1.0000
Support	207 test örneği



Confusion Matrix incelendiğinde, tüm sınıflarda %100 doğru sınıflandırma sağlandığı gözlemlenmiştir. **Hiçbir sınıf için hata veya karışma gözlenmemiştir.**

Bu sonuç, kullanılan veri setinin dengeli ve iyi tanımlı olması, intent sınıflarının birbirinden net ayrışması ve SentenceTransformer (all-MiniLM-L6-v2) modelinin başarılı gömüleme (embedding) sağlamasından kaynaklanmaktadır.

Model, bu aşamada **overfitting eğilimi göstermemektedir**, çünkü cross-validation ve test seti sonuçları tutarlıdır.

7.2 LLM Yanıt Performansı (Response Evaluation)

Random seçilmiş 10 örnek üzerinde **LLaMA 3 8B Instruct** ve **Gemini 1.5 Flash** modelleri için response evaluation çalıştırılmıştır.

- Her iki modelin de üretmiş olduğu yanıtlar, Ground Truth ile büyük oranda örtüşmektedir.
- Bazı örneklerde:
 - LLaMA** daha uzun, daha doğal ve diyalog odaklı cevaplar üretmektedir.
 - Gemini** ise daha kısa ve doğrudan task-oriented yanıtlar sağlamaktadır.

Input	LLaMA Response	Gemini Response
I don't want that	Tam eşleşme	Tam eşleşme
Thank you very much	Tam eşleşme	Tam eşleşme
Update card info	Daha uzun ama anlam olarak doğru	Daha kısa ama anlam doğru
Bye	Tam eşleşme	Tam eşleşme
Has my order been dispatched?	Daha uzun cevap, anlam doğru	Daha kısa, doğrudan
Thank you very much	Tam eşleşme	Tam eşleşme
Where is my order?	Daha uzun, detaylı cevap	Kısa, doğrudan
Change my payment method	Tam anlamlı cevap	Tam anlamlı cevap
Cancel that	Tam eşleşme	Tam eşleşme
Bye	Tam eşleşme	Tam eşleşme

7.3 Genel Sonular

Her iki model de test edilen rneklerde başarılı sonuçlar retmiştir.

LLaMA 3 modeli, kullanıcı deneyimini artıracak şekilde daha diyalog tabanlı ve sıcak yanıtlar sağlarken;

Gemini 1.5 Flash modeli, zellikle kısıtlı karakterli / mobil arayüzler iin daha kısa ve işlevsel yanıtlar retmektedir.

Bu alışma göstermiştir ki:

- Basit bir intent sınıflandırıcı + RAG destekli bir LLM kombinasyonu ile yüksek doğrulukta, anlamlı ve güvenilir chatbot performansı elde etmek mümkündür.
- Kullanılan framework, model bağımsız olarak farklı LLM'ler ile kolayca genişletilebilir.
- LLaMA ve Gemini gibi modellerin stil farkları, farklı kullanım senaryolarına göre (müşteri destek, hızlı cevap, vs.) avantaj sağlamaktadır.