SMS Fraud Control Final Raporu

Sarper Arda BAKIR

Stajyer

1-29 Temmuz 2024

Genel Bakış:

Program, SMS mesajlarını potansiyel dolandırıcılık veya spam riskleri açısından analiz etmek ve puanlamak için tasarlanmış hem bir API hizmeti hem de bir terminal uygulamasıdır. API, çeşitli yapay zeka hizmetleri ve modelleri ile entegre olarak her mesaj için kapsamlı bir risk değerlendirmesi sağlar. Çoklu sayıda SMS'leri kontrol edebilmek için hem API üzerinden json dosyası hem de csv dosyası kabul ederek yapıyor.

Kullanılan Servisler:

- **OpenAI:** SMS mesajlarının içeriklerini analiz etmek ve dolandırıcı veya spam olma olasılığını belirlemek için gelişmiş dil modellerinden ChatGPT kullanır. 0-100 arasında bir risk tahmini ve mesaj içeriği ile ilgili kategoriyi döndürür.
- **Gemini:** SMS mesajlarını dolandırıcılık ihtimalleri ile ilgili bir risk tahmini yapan yapay zekadır.
- **IPQS:** SMS mesajlarındaki URL'leri kötü niyetli IP adresleri ve URL'ler veritabanına karşı kontrol eder. Kendi sistemleri üzerindeki risk skoru kullanılır.
- **TensorFlow:** Eğitilmiş makine öğrenme modellerine dayalı olarak SMS mesajlarının risk seviyesini tahmin etmek için kullanılır.

Puanlama Sistemi:

Nihai dolandırıcılık puanı, farklı servislerden gelen puanların birleştirilmesiyle hesaplanır ve her servisin önemine göre ağırlıklandırılır.

Eğer URL varsa:
*Fraud Skor = (GeminiAI Puanı * 0.15) + (TensorFlow Puanı * 0.15) + (IPQS Puanı * 0.2) + (OpenAI Puanı * 0.5)*
Eğer URL yoksa:
*Nihai Puan = (GeminiAI Puanı * 0.2) + (TensorFlow Puanı * 0.3) + (OpenAI Puanı * 0.5)*

Puan Aralıkları ve Risk Seviyeleri:

- **Yüksek Risk**: Nihai Puan >= 80
- **Orta Risk**: 50 <= Nihai Puan < 80
- **Düşük Risk**: Nihai Puan < 50

API için Request ve Response Örneği:

Request:


```
[
   {
      "Message" : "Degerli Akbankli, 31 Aralik'a kadar Ak Yatirim haricinde baska bir araci kurumda bulunan hisse senedi varliklarinizi virman yolu ile Ak Yatirim hesabiniza transfer ederek 200.000 Mil Puan'a varan odul kazanabilirsiniz. Detayli bilgi icin: akbank.com/virman-kamp Ucretsiz SMS almak istemiyorsaniz 'smsistemiyorum' yazip 8885'e gonderebilirsiniz. Iyi gunler dileriz. Mersis:0015001526400497 B001"
   },
   {
      "Message" : "Size özel 1 bilet alana 1 bilet bedava (1+1) kampanyası tanımladık. Yüzyüzeyken Konuşuruz yarın Atilim Üniversitesi Amfi Tiyatro salonuna geliyor!SMS IPTAL Mersis No: 0171064698600001 https://p.s-m-s.red/x/xntYLG B002"
   }
]
```


Response:

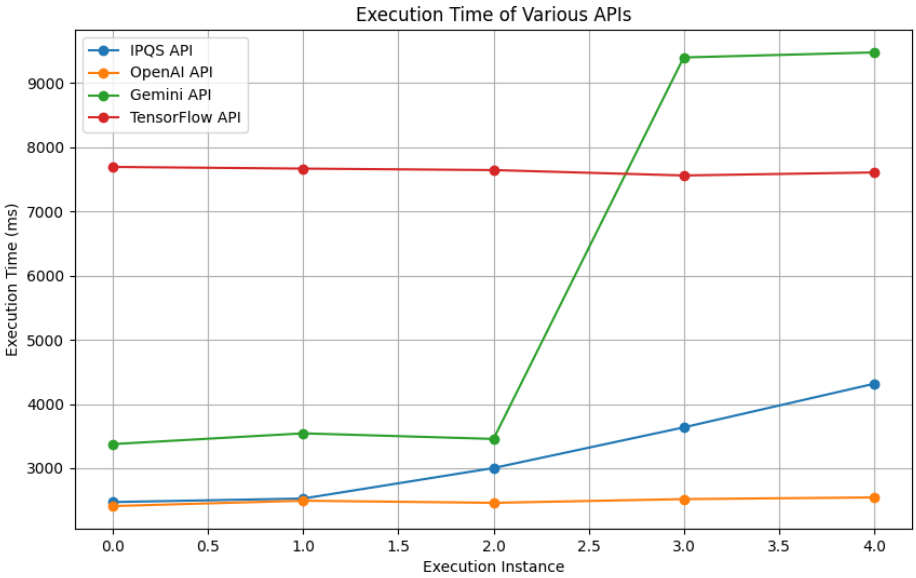

```
[
   {
      "message": "Degerli Akbankli, 31 Aralik'a kadar Ak Yatirim haricinde baska bir araci kurumda bulunan hisse senedi varliklarinizi virman yolu ile Ak Yatirim hesabiniza transfer ederek 200.000 Mil Puan'a varan odul kazanabilirsiniz. Detayli bilgi icin: akbank.com/virman-kamp Ucretsiz SMS almak istemiyorsaniz 'smsistemiyorum' yazip 8885'e gonderebilirsiniz. Iyi gunler dileriz. Mersis:0015001526400497 B001",
      "geminiScore": 90,
      "tensorFlowScore": 0.02,
      "ipqsScore": -1,
      "openAIScore": 0,
      "finalScore": 27,
      "category": "finans",
      "explanation": "The final fraud/spam risk score is 27, which indicates a low risk level."
   },
   {
      "message": "Size özel 1 bilet alana 1 bilet bedava (1+1) kampanyası tanımladık. Yüzyüzeyken Konuşuruz yarın Atilim Üniversitesi Amfi Tiyatro salonuna geliyor!SMS IPTAL Mersis No: 0171064698600001 https://p.s-m-s.red/x/xntYLG B002",
      "geminiScore": 90,
      "tensorFlowScore": 100,
      "ipqsScore": -1,
      "openAIScore": 0,
      "finalScore": 57,
      "category": "kampanya",
      "explanation": "The final fraud/spam risk score is 57, which indicates a moderate risk level."
   }
]
```

Servislerin Çalışma Süresi:

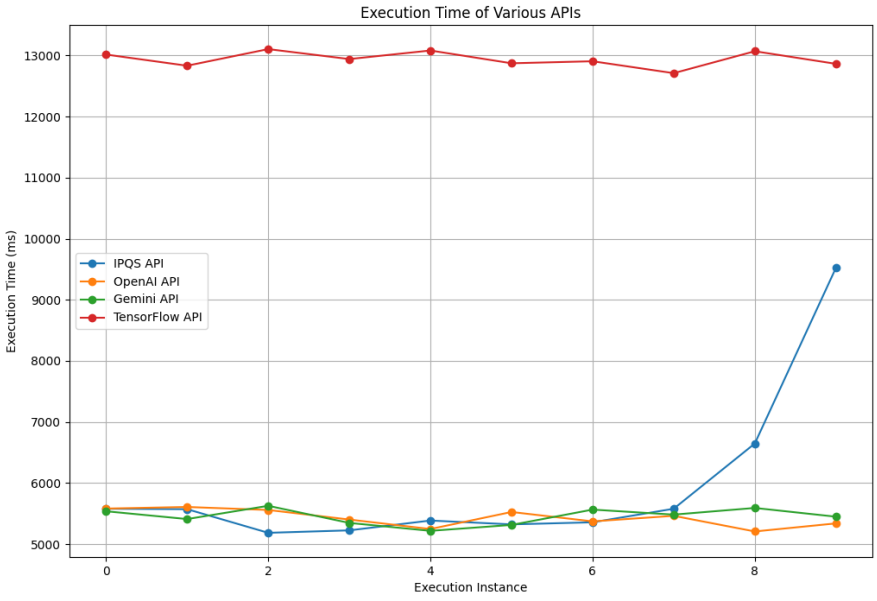
Bir tane SMS örneği ile ilgili analiz yapılırken kullanılırken servislerin çalışma süreleri :

Total Execution Time of OpenAI API: 1389 ms
Total Execution Time of IPQS API: 2823 ms
Total Execution Time of Gemini API: 2998 ms
Total Execution Time of TensorFlow API: 5554 ms

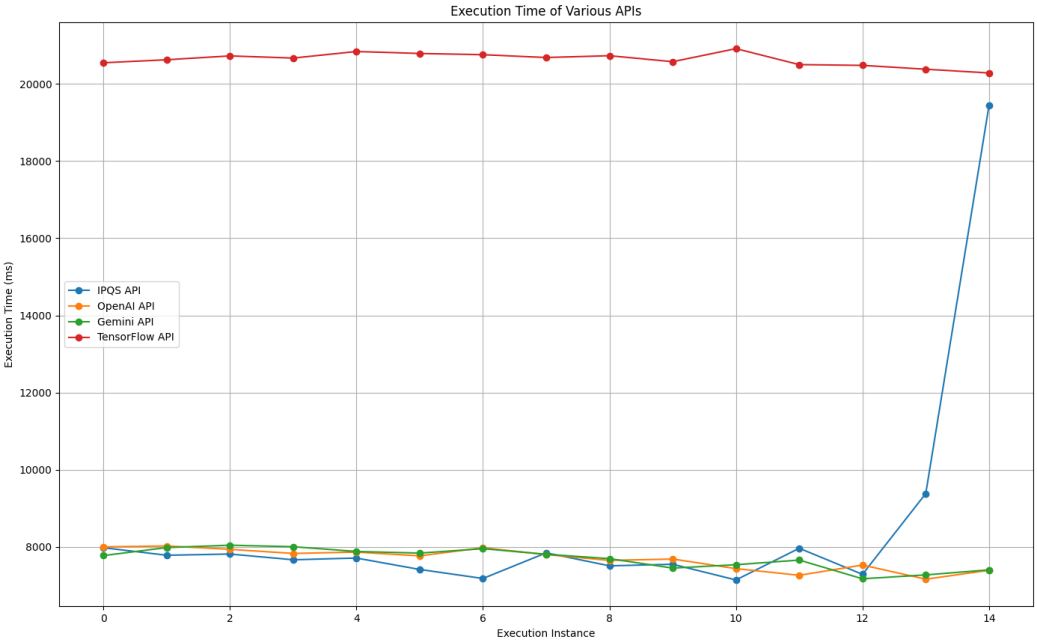
Daha iyi bir tespit yapabilmek için çoklu sayılarda veri girişi yapılan analizlerin süreleri ile ilgili grafikler aşağıda yer almaktadır:



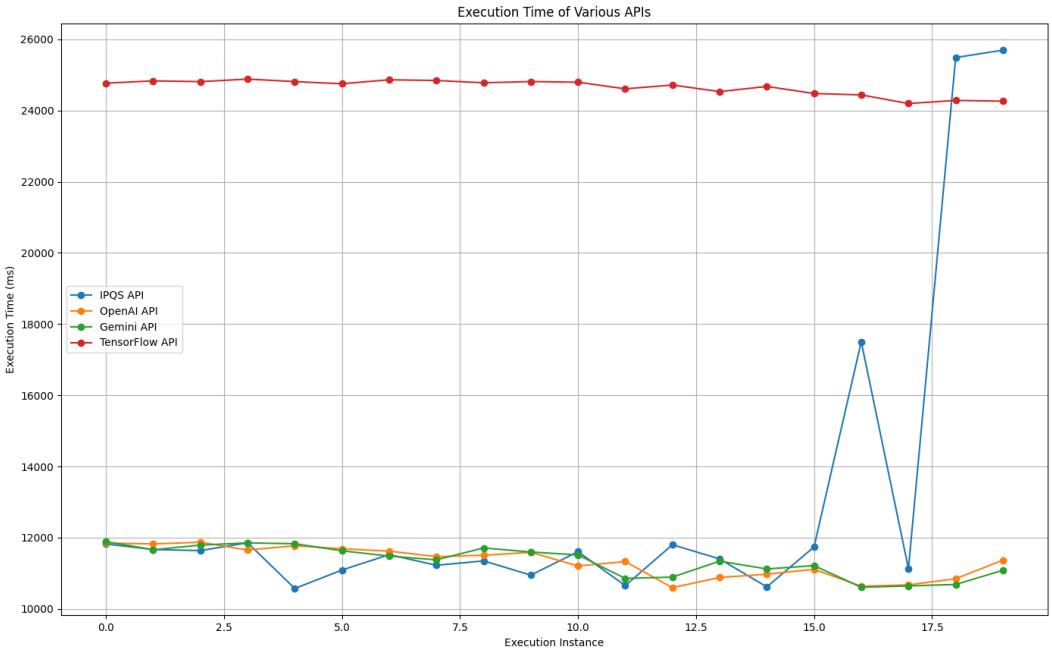
Figür 1. 5 SMS'in Analizi



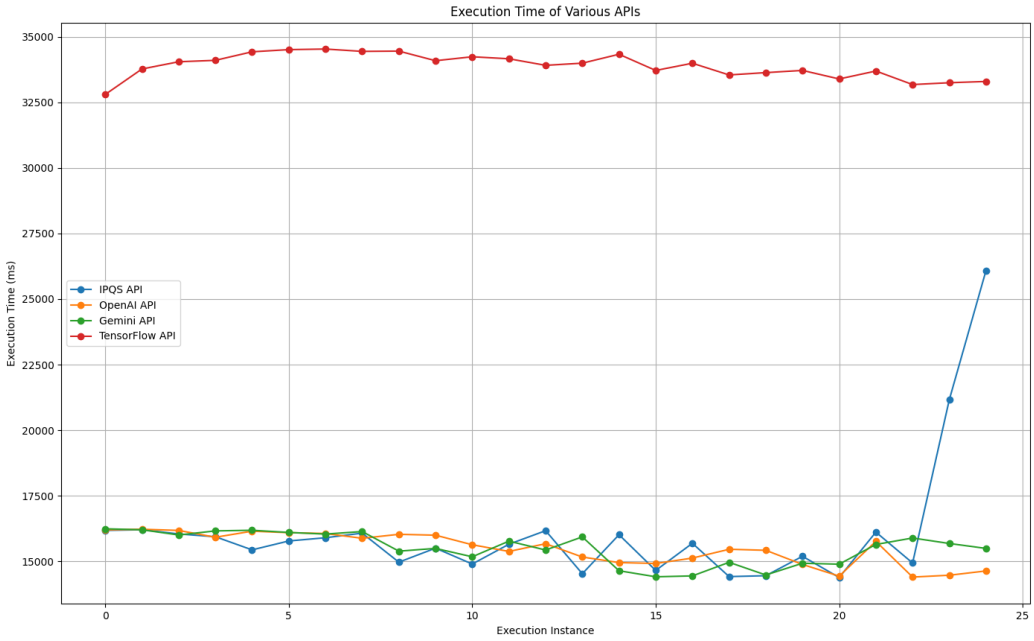
Figür 2. 10 SMS'in Analizi



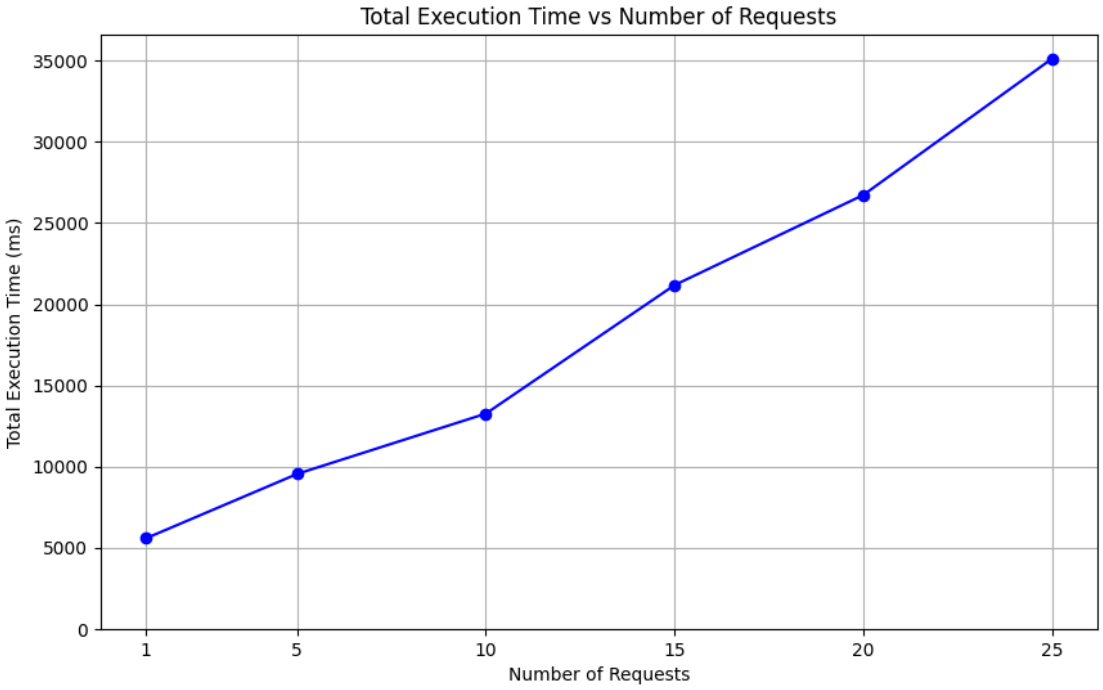
Figür 3. 15 SMS'in Analizi



Figür 4. 20 SMS'in Analizi



Figür 5. 25 SMS'in Analizi



Figür 6. Veri sayısına göre Çalışma Süreleri

Sonuçların Analizi:

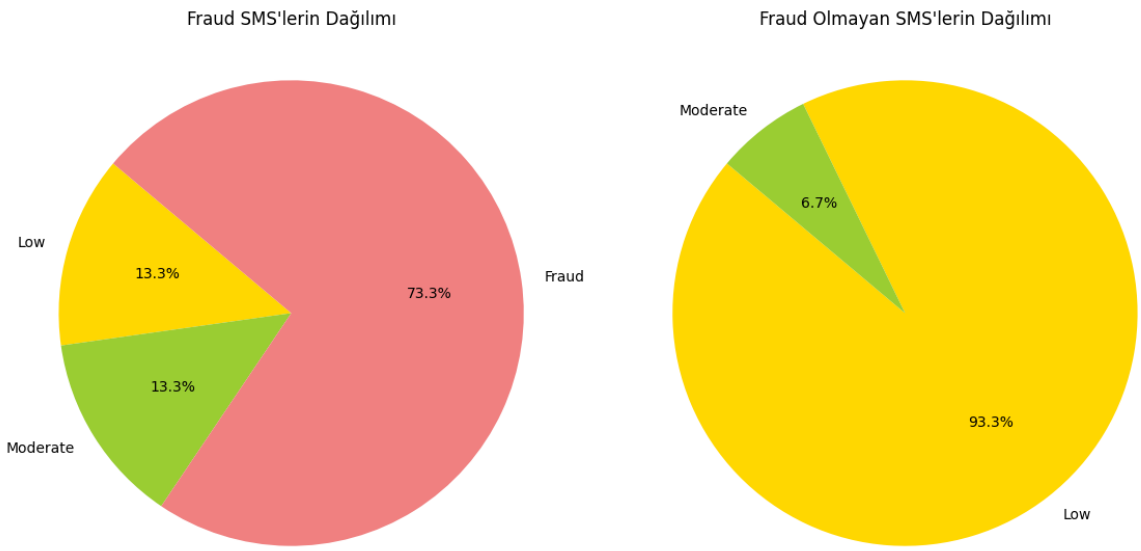
OpenAI API'sinin performansı, IPQS ve Gemini API'lerine kıyasla daha tutarlı ve stabil görünmektedir. Veriler arttıkça OpenAI API'sinin yürütme sürelerinde de bir artış gözlemlenmektedir ancak bu artış, diğer API'lere kıyasla daha öngörülebilir ve dengelidir. Örneğin, beş veri ile yapılan testlerde OpenAI API'si 2411 ms ile 2545 ms arasında çalışırken, yirmi veri ile yapılan testlerde bu süreler 10597 ms ile 11873 ms arasına çıkmaktadır. OpenAI API'sinin daha stabil performans göstermesi, büyük veri kümeleri ile çalışırken güvenilir bir seçenek olduğunu göstermektedir.

Veri sayısının artmasıyla birlikte tüm API'lerin yürütme sürelerinde belirgin bir artış gözlemlenmektedir. Bu artış, sistemin her bir veri noktası için gerçekleştirdiği işlemlerin toplam süresinin artmasından kaynaklanmaktadır. Veriler arttıkça, her bir API'nin işlem kapasitesi zorlanmakta ve daha fazla zaman harcanmaktadır. Örneğin, beş veri ile yapılan testlerde toplam yürütme süresi 9550 ms iken, on beş veri ile bu süre 21181 ms'ye, yirmi veri ile ise 26730 ms'ye çıkmaktadır. Herr bir API'nin verileri işleme kapasitesi ve verimliliği de bu artıştan etkilenmektedir. TensorFlow incelediğimizde tüm grafiklerde en yavaş çalışan model olmasının sebebi TensorFlow'un daha karmaşık ve yoğun işlem gücü gerektiren işlemleri yürütmesinden kaynaklanabilir. TensorFlow özelinde performans sorunlarını azaltmak için GPU kullanmak ve batch boyutunu küçültmek gibi yöntemler süreyi daha kısaltmak için faydalı olabilir.

Verilerde gözlemlenen IPQS API kullanım süresi, işlem sayısı arttıkça son verileri analiz ederken belirgin bir şekilde artmaktadır. Örneğin, beş veri ile yapılan testlerde IPQS API'sinin toplam yürütme süresi 2473 ms ile 4315 ms arasında değişirken, veri sayısı yirmiye çıktığında bu süreler 11824 ms ile 25693 ms arasına çıkmaktadır. Artan trafik veya yük, IPQS API sunucularına yönelik yanıt sürelerinde artışa neden olabilir bu yüzden de son SMS örneklerinde dramatik bir süre artışı gerçekleşebiliyor olabilir. Ancak yine de bu durum, API’nin geç çalışma süresi bile TensorFlow’un çalışma süresinin altında olduğu için bir sorun teşkil etmemektedir.

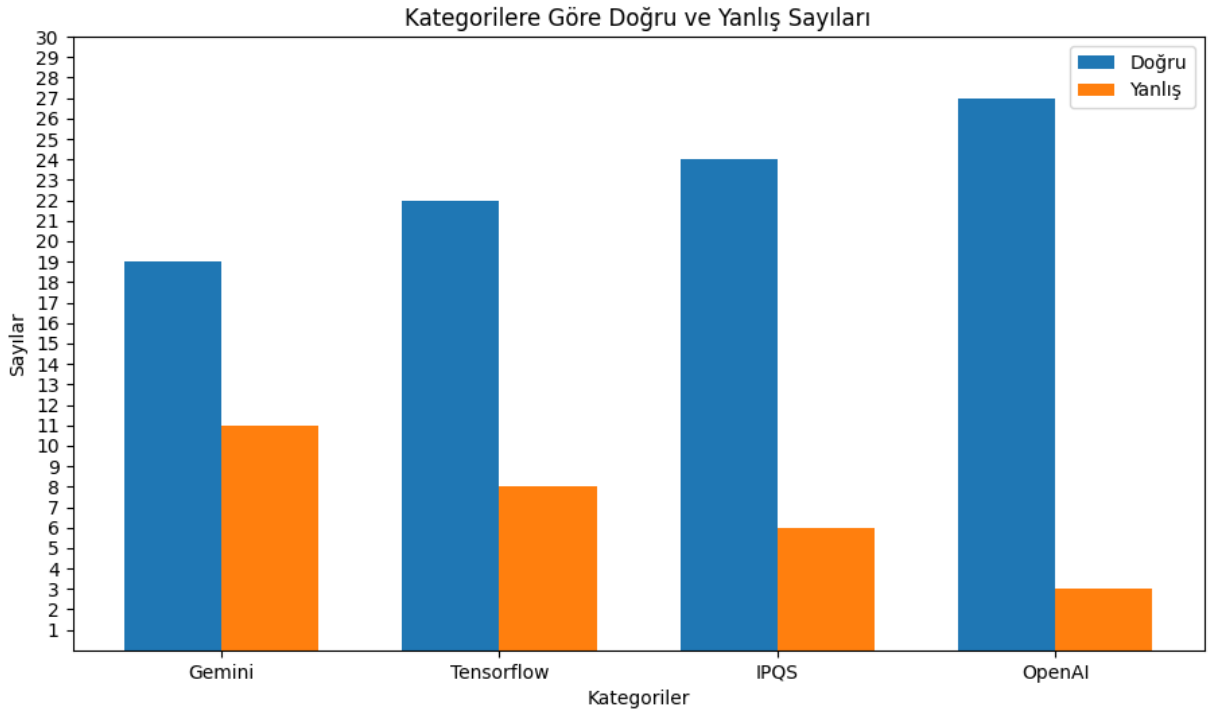
Doğruluk Oranı:

Programın doğruluk oranını test etmek için 15 fraud SMS ve 15 fraud olmayan SMS ile test yapıldı. Sonuçlar:



Figür 7. SMS Tespit Dağılımları

30 örnekte 2 fraud SMS'i tespit edemezken 2'sini de yeteri kadar fraud bulmadı. Bu sonuçları oranladığımızda ise yaklaşık yüzde 87 oranında bir doğruluk oranında çalıştığı düşünülebilinir. Fraud olmayan SMS'lerin tespitinde bir sorun gözükmese de fraud olan bazı smsler programın tespitinden kaçabilmektedir. Ayrıca kullanılan servislerin fraud sms yakalama oranları da detaylı incelendiğinde aşağıdaki sonuçlar gözükmektedir:



Figür 8. Servislere Göre Tespit Dağılımı

Bu grafik incelendiğinde ise 30 tane örnekte tespit için en doğru servisin 27 tane doğru tespitle OpenAI yapay zekası olan ChatGPT olduğu ortaya çıkarken en düşüğün ise 19 ile Google'ın yapay zekası olan Gemini olduğu gözükmektedir. Bu sonuçlardan önce yüzde 75'lerde olan doğruluk oranı Fraud Skor tespitinde kullanılan ağırlığın tekrardan değiştirilmesiyle yüzde 90'lara yaklaşmaktadır. IPQS gibi üçüncü parti olan bir IP reputation API'si yüzde 80 ile çalışırken SPAM data seti ile çalışan TensorFlow modeli de yaklaşık yüzde 75 ile çalışmaktadır. Modeli daha iyi datasetlerle eğitilmesi durumunda fraud tespiti daha iyi olabilecekken daha iyi üçüncü parti API'ler ile de linklerin fraud olup olmayacağını daha iyi tespit edilebilir. Bunun yanında Gemini yerine daha iyi yapay zeka araçlarıyla da çalışılması daha mantıklı gözükmektedir. Bu sayede programın yüzde 87 olan doğruluk oranı da daha yükseklere çekilebilinir.