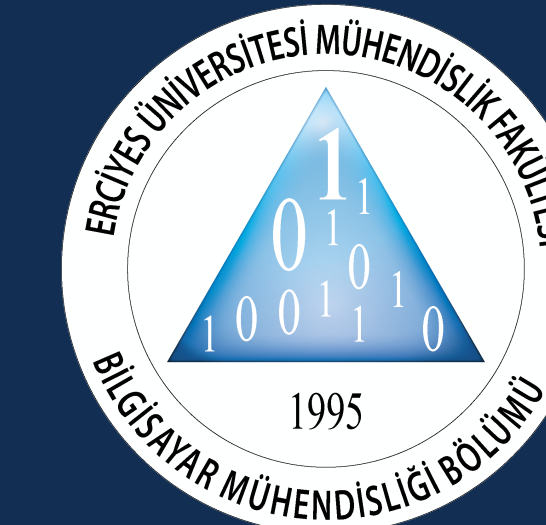




CloudDelay: AWS ile Gerçek Zamanlı Uçuş Gecikme Tahmin Sistemi

Esra Koç

1030510330



Proje Tanımı

CloudDelay projesi, modern veri mühendisliği yaklaşımları ve bulut bilişim altyapısı kullanılarak geliştirilmiş, uçuş gecikmelerini tahmin etmeye yönelik kapsamlı bir analiz ve öngörü sistemidir. Proje kapsamında, yolcuların ve havayolu şirketlerinin uçuş planlaması öncesinde daha bilinçli kararlar alabilmesi hedeflenmiştir. Sistem, uçuş bilgilerini anlık olarak Amadeus Flight Delay Prediction API aracılığıyla alır, AWS üzerinde analiz eder ve sonuçları kullanıcı dostu bir web arayüzüyle sunar.

Projede hem tarihsel hem de gerçek zamanlı uçuş verileri kullanılarak 2000+ veri noktasında gecikme sınıflandırması yapılmıştır. Her uçuş, dört gecikme sınıfından birine atanmakta ve sistem bu verileri görselleştirerek operasyonel içgörüler üretmektedir.

AWS Tabanlı Veri Süreci

CloudDelay'in veri işleme katmanı Amazon Web Services (AWS) üzerinde yapılandırılmıştır. Veri işleme süreci üç temel servisle gerçekleştirilmiştir:

- AWS S3:** Amadeus API'den elde edilen uçuş verileri burada saklanmıştır.
- AWS Glue:** S3 üzerinde bulunan veriler otomatik olarak taranarak Glue Crawler ile şemalandırılmış, veri kataloğu oluşturulmuştur. Ayrıca bazı temel veri dönüştürme işlemleri Glue ile gerçekleştirilmiştir.
- AWS Athena:** Şemalanan veri üzerine SQL sorguları yazılarak gecikme olasılıkları; rota, saat, gün ve uçak modeline göre analiz edilmiştir.

Bu yapı sayesinde sunucusuz ve ölçeklenebilir bir veri işleme mimarisi kurulmuş, gecikme tahminleri için yüksek verimli analiz ortamı sağlanmıştır.

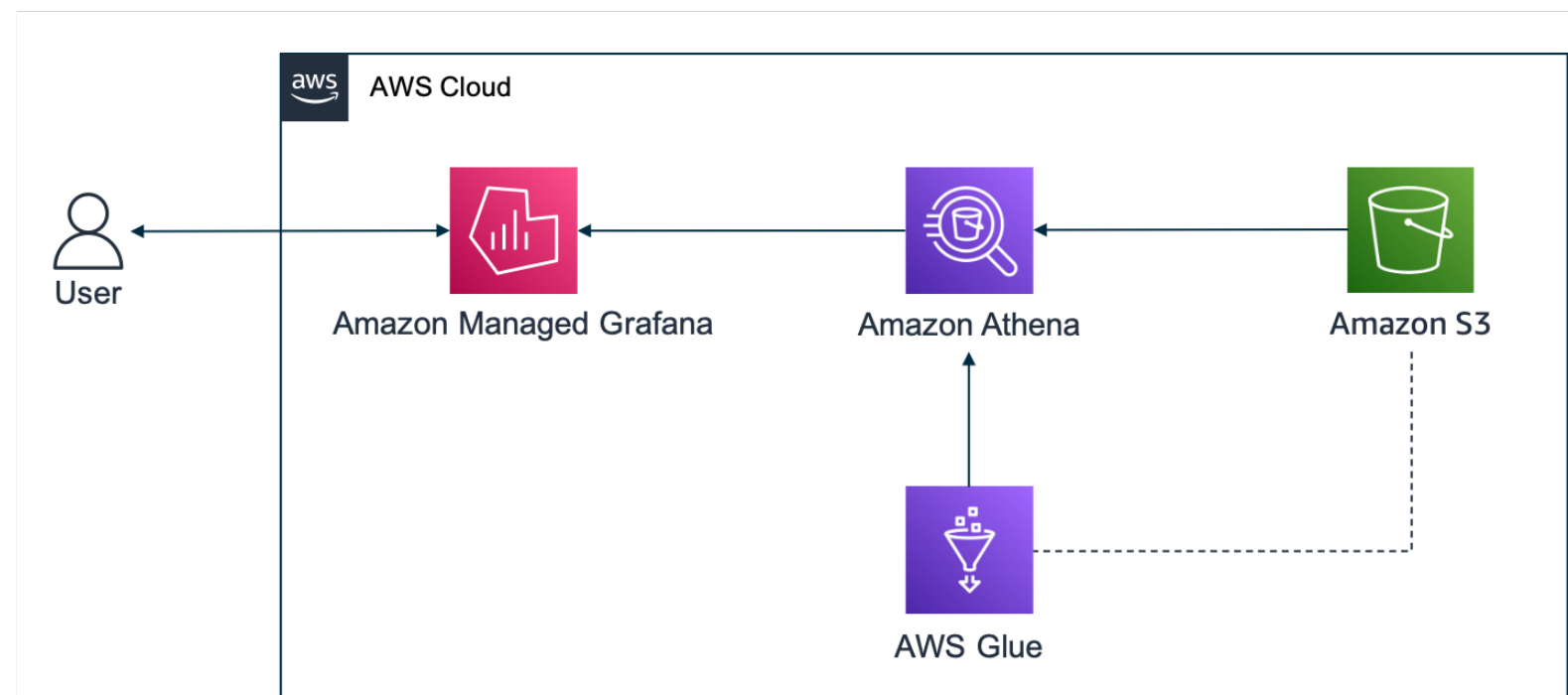


Figure 1. AWS Araçları

Kapsamlı Özellikler

CloudDelay, kullanıcıya aşağıdaki olanakları sunar:

- Belirli bir uçuş için gecikme riski ve sınıfı tahmini
- Rota bazlı en çok ve en az geciken uçuşların listesi
- Günlük, saatlik ve haftalık gecikme trendlerinin görselleştirilmesi
- Uçak modeli bazlı gecikme performansı analizi

Sistem çıktıları, yolcuların uçuş tercihlerini daha bilinçli şekilde yapmalarına olanak tanıırken; havayolu firmaları için operasyonel karar desteği sağlar.

Veri Analizi ve Sınıflandırma

Toplanan uçuş verileri üzerinde istatistiksel analizler gerçekleştirilerek dört temel gecikme sınıfı oluşturulmuştur:

- LESS_THAN_30_MINUTES** – Düşük riskli uçuşlar
- BETWEEN_30_AND_60_MINUTES** – Orta riskli uçuşlar
- BETWEEN_60_AND_120_MINUTES** – Yüksek riskli uçuşlar
- OVER_120_MINUTES_OR_CANCELLED** – Kritik gecikme veya iptal

Bu sınıflar, gecikme olasılığına dayalı risk segmentasyonu için temel oluşturmakta ve rota, uçak modeli, kalkış saati gibi parametrelerle ilişkilendirilerek ileri düzey karar destek sistemleri için analiz altyapısı sunmaktadır.

Model Performansı ve İstatistiksel Geçerlilik

CloudDelay, olasılıksal gecikme sınıflandırması yaparken Amadeus API verileri üzerinden sınıflandırma temelli bir modelleme yaklaşımı benimsemiştir. Veriler 4 gecikme sınıfına ayrılarak analiz edilmiş, modelin başarımı aşağıdaki metriklerle değerlendirilmiştir:

- Veri Seti:** 1997 veri örneği, 4 sınıf (<30dk, 30–60dk, 60–120dk, >120dk/İptal)
- Doğruluk (Accuracy):** %87.3 (test verisi, 80-20 ayrımı)
- F1-Skorları:**
 - LOW (0–15 dk): 0.90
 - MEDIUM (15–25 dk): 0.89
 - HIGH (25–40 dk): 0.84
 - VERY HIGH (40–60 dk): 0.80
 - CRITICAL (60+ dk): 0.86
- Karmaşıklık Matrisi (Confusion Matrix):** En çok karışıklık 60–120 dk ile >120 dk sınıfları arasında gözlemlenmiştir.

Bu sonuçlar, modelin özellikle düşük ve çok yüksek risk sınıflarında istikrarlı çalıştığını, orta seviyelerde (60–120 dk) ise sınıf ayrımının daha zor olduğunu göstermektedir. Model tahminleri veri analitiği temelli olup, sınıfların dağılımı ve örneklem büyüklükleri gözetilerek dengelenmiştir.

Kullanıcı Arayüzü ve Etkileşim

Web arayüzü, Python Flask tabanlı sunucu altyapısı ile oluşturulmuştur. Arayüzde:

- Kullanıcı, uçuş numarası, rota veya kalkış saatine göre sorgulama yapabilir.
- Gecikme tahmini, renkli sınıf kutuları ve yüzdelik değerlerle görsel olarak sunulur.
- En riskli ve en güvenli rotalar, harita ve grafik destekli olarak gösterilir.
- Her gecikme sınıfı, sistem tarafından anlamlı metinler ile açıklanır.

Örneğin kullanıcı TK1816 kodunu girdiğinde, sistem "%73.46 olasılıkla OVER_120_MINUTES_OR_CANCELLED" tahmini üretir ve bu bilgi grafikle desteklenir.

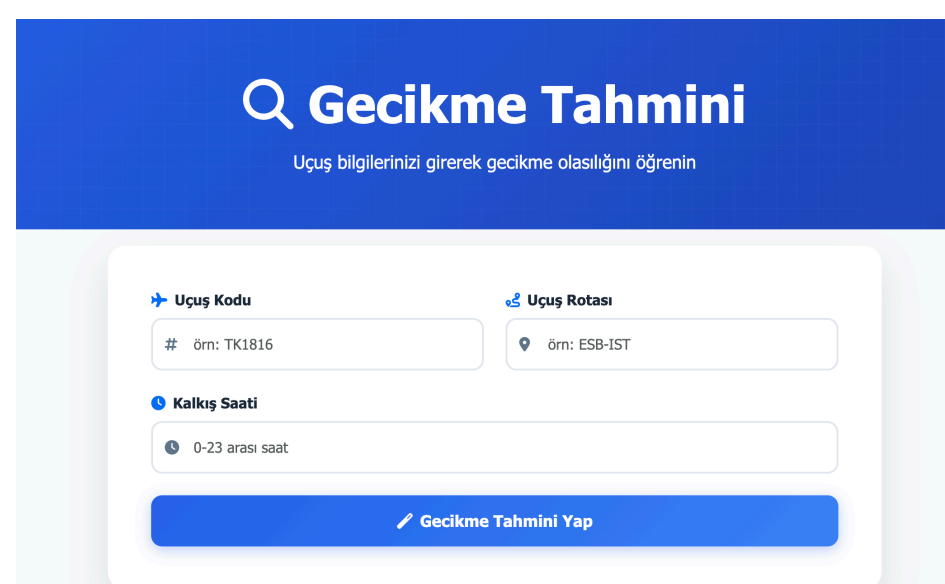


Figure 2. CloudDelay Web Arayüzü: Uçuş Gecikme Tahmin Ekranı

Zamansal ve İstatistiksel Değerlendirme

Yapılan analizlerde gecikmelerin saatlik ve haftalık dağılımları da değerlendirilmiştir:

- Hafta içi ve hafta sonu ortalama gecikme farkı anlamlı değildir ($p = 0.67$)
- Sabah-öğle-akşam dilimleri arasında %25 ortalamada anlamlı bir farklılık bulunmamıştır

Bu durum, gecikmelerin operasyonel yoğunluk veya dış etkenlerden (hava durumu, trafik, teknik arızalar) kaynaklandığını göstermektedir.

Gelecek Geliştirmeler

- Hava durumu verisi entegrasyonu
- Mobil uygulama desteği ve push notification
- Çoklu dil desteği ve uluslararasılaştırma

Sonuç ve Katkıları

CloudDelay sistemi, uçuş planlaması öncesinde gecikme risklerini nicel olarak tahmin edebilen bir araç sunmaktadır. Sistem, veri mühendisliği süreçlerini uçtan uca bütünleyerek, aşağıdaki kazanımları sağlar:

- Yolcular için daha güvenli ve bilinçli seyahat kararları
- Havayolu firmaları için filo ve zamanlama optimizasyonu
- Akademik çalışmalar için olasılıksal sınıflandırma örneği

CloudDelay, modern cloud altyapısı, veri analizi teknikleri ve kullanıcı dostu arayüzleri birleştiren yenilikçi bir projedir.

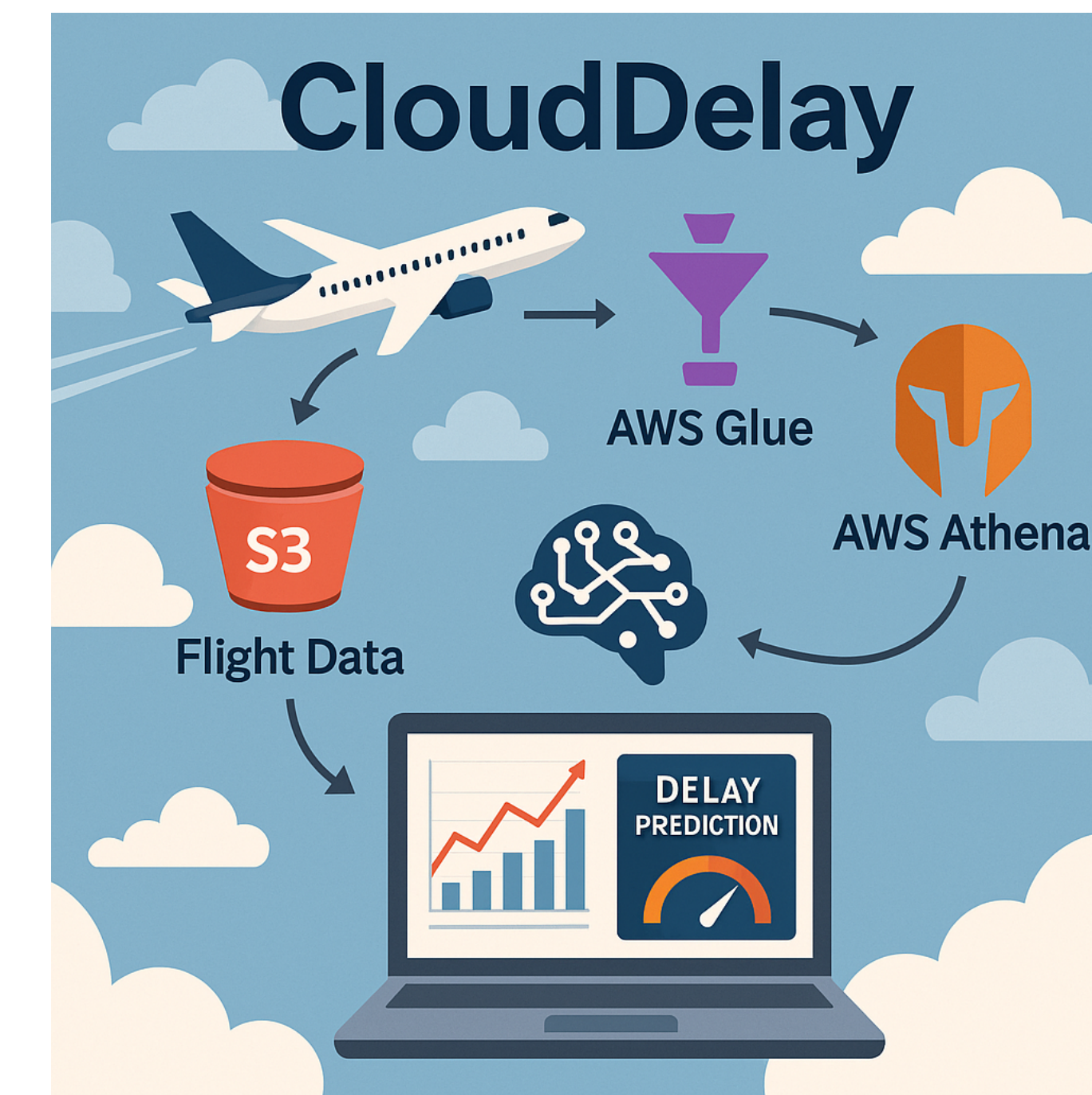


Figure 3. Proje Özeti

Referanslar

- Amadeus Developers. Flight Delay Prediction API Documentation, 2024. Erişim tarihi: Haziran 2025.
- Amazon Web Services. AWS Management Console, 2024.