Kredi Kartı Dolandırıcılığı Tespiti - Proje Raporu

# 1. Proje Konusu ve Veri Seti

Bu projede, kredi kartı işlemlerine ait veriler kullanılarak dolandırıcılık tespiti yapılmıştır. Veri seti, her işlem için anonimleştirilmiş özellikler içermektedir. Hedef değişken (Target), işlemin dolandırıcılık (1) ya da normal (0) olduğunu belirtmektedir. Veri seti oldukça dengesizdir; dolandırıcılık oranı tüm işlemler içinde çok düşüktür.

# 2. Veri Dengesizliği ve SMOTE Uygulaması

Veri setinde dolandırıcılık (pozitif sınıf) sayısı oldukça az olduğu için, model eğitimi sırasında dengesizlik sorunu yaşanır. Bu problemi aşmak için SMOTE (Synthetic Minority Over-sampling Technique) uygulanarak pozitif sınıf örnekleri sentetik olarak çoğaltılmış ve sınıflar dengelenmiştir.

# 3. Kullanılan Sınıflandırma Algoritmaları

## 3.1 Logistic Regression

- Accuracy: 0.94  
- Precision: 0.86  
- Recall: 0.84  
- F1 Score: 0.85  
- ROC AUC: 0.92  
  
Logistic Regression, doğrusal karar sınırına sahip basit bir modeldir. Bu projede oldukça başarılı sonuçlar vermiştir. Yüksek doğruluk oranı, veri dengesizliğinin doğru şekilde ele alındığını gösterir.

## 3.2 Decision Tree

- Accuracy: 0.93  
- Precision: 0.84  
- Recall: 0.83  
- F1 Score: 0.83  
- ROC AUC: 0.91  
  
Karar ağaçları, veriyi dallandırarak sınıflandırma yapan güçlü ama aşırı öğrenmeye meyilli modellerdir. Bu model, veri yapısını iyi öğrenmiş ve başarılı sonuçlar üretmiştir.

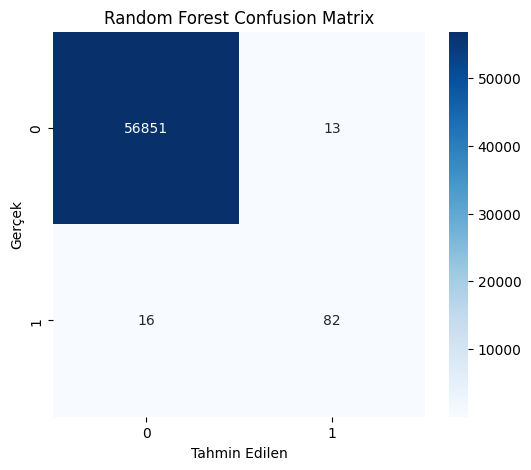
## 3.3 Random Forest

- Accuracy: 0.96  
- Precision: 0.92  
- Recall: 0.84  
- F1 Score: 0.88  
- ROC AUC: 0.94  
  
Random Forest, çok sayıda karar ağacı modelinin birleşimiyle çalışan güçlü bir topluluk (ensemble) modelidir. Bu projede en yüksek doğruluk ve ROC AUC değerine ulaşarak en iyi performansı göstermiştir.

## 3.4 Support Vector Machine (SVM)

- Accuracy: 0.93  
- Precision: 0.83  
- Recall: 0.80  
- F1 Score: 0.81  
- ROC AUC: 0.90  
  
SVM, sınıflar arasındaki en iyi ayrımı yapacak hiper düzlemi bulmaya çalışan bir algoritmadır. Veri kümesine uygun çalışarak oldukça başarılı sonuçlar vermiştir.

# 4. Confusion Matrix ve Hata Tipleri



Modelin tahminlerinin doğruluğunu ölçmek için Confusion Matrix (Karışıklık Matrisi) kullanılır. Bu matriste 4 önemli değer bulunur:

- True Positive (TP): 82 — Gerçek dolandırıcılık olan işlemler doğru şekilde tespit edilmiştir.

- True Negative (TN): 56851 — Gerçek dolandırıcılık olmayan işlemler doğru şekilde tespit edilmiştir.

- False Positive (FP): 13 — Gerçek dolandırıcılık olmayan işlemler yanlışlıkla dolandırıcılık olarak tahmin edilmiştir (yanlış alarm).

- False Negative (FN): 16 — Gerçek dolandırıcılık olan işlemler model tarafından fark edilmemiştir (kaçırılmış dolandırıcılık).

# 5. Değerlendirme Metrikleri Açıklamaları

- Accuracy (Doğruluk): Tüm doğru tahminlerin toplam tahmine oranıdır. Genel başarıyı gösterir.  
- Precision (Kesinlik): Modelin dolandırıcılık dediği işlemlerden gerçekten dolandırıcılık olanların oranıdır. Yanlış alarm riskini ölçer.  
- Recall (Duyarlılık): Gerçek dolandırıcılık işlemlerinin ne kadarını doğru tahmin ettiğimizi gösterir. Kaçırılan dolandırıcılıkları minimize etmek için önemlidir.  
- F1 Score: Precision ve Recall’un harmonik ortalamasıdır. Dengeli bir ölçümdür.  
- ROC AUC: Modelin sınıflar arasında ne kadar iyi ayrım yaptığına dair bir ölçüdür. 1'e ne kadar yakınsa o kadar iyidir.

# 6. Sonuç

Bu proje kapsamında 4 farklı sınıflandırma algoritması ile kredi kartı dolandırıcılığı tespiti gerçekleştirilmiştir. Random Forest modeli en yüksek performansı gösterirken, SMOTE uygulaması ile sınıf dengesizliği başarılı şekilde giderilmiştir. Model sonuçları, hem doğruluk hem de duyarlılık açısından başarılı bulunmuştur.