



# **Kredi Kartı Dolandırıcılığı Tespiti (Fraud Detection) Proje Raporu**

**Ders =Veri Madenciliğine Giriş**

**Ad –Soyad =Kadir Emir Yücel**

**NO =2023208190**

## 1. Projenin Amacı

Bu çalışmanın amacı, kredi kartı işlemlerini içeren veri seti üzerinde makine öğrenmesi algoritmaları kullanarak, yapılan işlemin normal mi yoksa dolandırıcılık (fraud) mı olduğunu yüksek doğrulukla tespit etmektir.

## 2. Veri Seti Analizi ve Ön İşleme

Kullanılan veri seti 284.807 işlemten oluşmaktadır ve 31 öznitelik içermektedir. Veri seti incelendiğinde ciddi bir sınıf dengesizliği (class imbalance) olduğu görülmüştür:

- Normal İşlemler (Sınıf 0):** %99.83.
- Dolandırıcılık İşlemleri (Sınıf 1):** %0.17.

### Uygulanan Ön İşlemler:

- Normalizasyon:** `Amount` ve `Time` değişkenleri, değer aralıklarının model performansını olumsuz etkilememesi için `StandardScaler` kullanılarak normalize edilmiştir.
- Eksik Veri Kontrolü:** Veri setinde `inf` (sonsuz) değerler `NaN`'a çevrilmiş ve ardından tüm boş değerler 0 ile doldurulmuştur.
- Dengesizliğin Giderilmesi (SMOTE):** Eğitim setindeki sınıf dengesizliğini gidermek için **SMOTE (Synthetic Minority Over-sampling Technique)** uygulanmıştır.
  - SMOTE Öncesi:** 227.451 Normal, 394 Fraud .
  - SMOTE Sonrası:** 227.451 Normal, 227.451 Fraud şeklinde eşitlenmiştir .

## 3. Kullanılan Modeller ve Yöntemler

Eğitim seti (Train) %80, Test seti %20 olacak şekilde ayrılmış ve aşağıdaki algoritmalar ile modeller eğitilmiştir:

- Random Forest Classifier:** Topluluk öğrenmesi (Ensemble) tabanlı güçlü bir algoritma.
- XGBoost Classifier:** Gradyan artırma tabanlı, yüksek performanslı bir algoritma.
- MLP Classifier (Yapay Sinir Ağları):** Çok katmanlı algılayıcı (64, 32 nöronlu katmanlar).

## 4. Model Performans Sonuçları

Aşağıdaki tabloda modellerin "Fraud (Dolandırıcılık)" sınıfını tahmin etme başarıları karşılaştırılmıştır:

Model	Accuracy (Doğruluk)	Precision (Kesinlik)	Recall (Duyarlılık)	F1-Score
XGBoost	0.9994 <sup>11</sup>	0.83 <sup>12</sup>	0.85 <sup>13</sup>	0.84 <sup>14</sup>

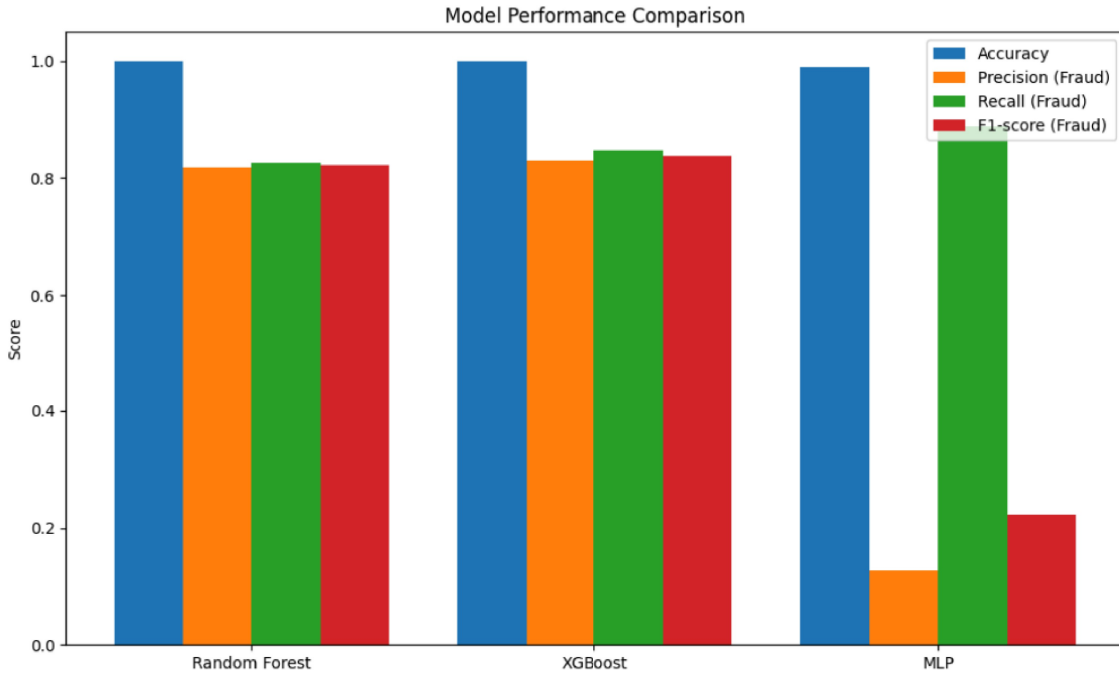
Random Forest	0.9994 <sup>15</sup>	0.82 <sup>16</sup>	0.83 <sup>17</sup>	0.82 <sup>18</sup>
MLP (Neural Net)	0.9893 <sup>19</sup>	0.13 <sup>20</sup>	<b>0.89</b> <sup>21</sup>	0.22 <sup>22</sup>

## Karmaşıklık Matrisi (Confusion Matrix) Analizi

Test setindeki **98 adet gerçek dolandırıcılık** vakası üzerinden modellerin başarısı şöyledir:

- **XGBoost:** 15 dolandırıcılığı kaçırmış (False Negative), 17 normal işlemi yanlışlıkla dolandırıcılık olarak işaretlemiştir (False Positive) .
- **Random Forest:** 17 dolandırıcılığı kaçırmış, 18 normal işlemi yanlış işaretlemiştir .
- **MLP:** Sadece 11 dolandırıcılığı kaçırarak en yüksek yakalama oranına (Recall) sahiptir. Ancak, **598 adet normal işlemi** yanlışlıkla dolandırıcılık olarak işaretleyerek çok yüksek bir yanlış alarm (False Positive) oranı vermiştir.

## 5. Sonuç ve Değerlendirme



Yapılan testler sonucunda projenin kazanan modeli **XGBoost** olmuştur.

1. **En Dengeli Model (XGBoost):** Hem dolandırıcılığı yakalama (Recall: 0.85) hem de doğru alarm verme (Precision: 0.83) konusunda en yüksek F1-Score (0.84) değerine ulaşmıştır.
2. **MLP'nin Yanılgısı:** MLP modeli, dolandırıcılık vakalarını yakalamada (Recall: 0.89) en başarılı model gibi görünse de, Precision değerinin çok düşük olması (0.13) nedeniyle pratik kullanımda başarısızdır. Çok fazla "yanlış alarm" üreterek sistemi kilitleyebilir.

3. **SMOTE Etkisi:** SMOTE kullanımı sayesinde modeller azınlık sınıfı olan dolandırıcılık işlemlerini başarıyla öğrenebilmiştir.