

# Fraud Detection in Banking

Finance Division



Giriş

Dataset haqqında

Datanın ilkin təhlili (EDA)

Modelin qurulması

Streamlit tətbiqi

Power BI paneli

Nəticə və gələcək planlar

# Layihənin Girişi və Problemin Təsviri

## Fraud nədir?

Bir şəxs və ya qurum başqasını aldadaraq və ya sahib olmadığı hüquqlardan istifadə edərək maddi qazanc əldə etməyə çalışırsa – bu fraud sayılır.

## Mübarizə səbəbi:

1. Banklar və şirkətlər hər il milyonlarla manat ziyanə düşür.
2. Müştəri etibarını azalır
3. Hüquqi risklər yaranır.

## Layihənin məqsədi:

Əməliyyat məlumatlarından istifadə edərək, maşın öyrənməsi alqoritmləri ilə fraud hallarını proqnozlaşdırmaq və real vaxtda monitoring sistemləri qurmaq.



# Dataset Haqqında



## Əsas Sütunlar və Təsvirləri

- timestamp – əməliyyatın tarixi və saati
- amount – ödənilən məbləğ (AZN)
- location – əməliyyatın baş verdiyi şəhər
- merchant – alış-veriş edilən yer
- transaction\_type – əməliyyatın növü
- gender – müştərinin cinsi
- account\_balance – əməliyyat sonrası balans
- is\_fraud – fırıldaq olub-olmaması (1 və ya 0)



## Dataset Strukturu

Məlumatlar Python vasitəsilə sintetik şəkildə generasiya edilmişdir. Cədvəl bank əməliyyatlarının əsas xüsusiyyətlərini əhatə edir və 13 sütundan ibarətdir.

## Əlavə Məlumat

Ümumi sətir sayı: 40000

Fraud əməliyyatlarının faizi: 18%

(Məlumatlar balanssız olduğu üçün model balanslaşdırma mərhələsindən keçmişdir.)



Məlumatların ilkin təhlili zamanı fırıldaqçılıq hallarının demoqrafik və zamana bağlı paylanması araşdırılmışdır.

## Məlumatların İlkin Təhlili (EDA)

### Fraud və Normal əməliyyatların say balansı

◆ Dataset-də fırıldaq əməliyyatları ümumi əməliyyatların yalnız 18%-ni təşkil edir, bu isə balanssızlıq problemini göstərir.



### Saat və həftənin günləri üzrə fraud sıxlığı

Fraud hallarının ən sıx rast gəlinəyi saat aralığı: 12:00 və 19:00  
Şənbə günündə cüzi artım müşahidə olunub.



### Cinsə görə fraud nisbəti

Qadın və kişi müştərilər arasında nisbət təxminən bərabərdir – kişi: 49.45%, qadın: 50.55%



### Region üzrə fraud sıxlığı

Ən çox fraud hadisəsi Bakı, Zaqatala, Şirvan və Şəki bölgələrində qeydə alınıb.



### Fraud üçün orta məbləğ

Fraud əməliyyatlarında orta məbləğ 2498.12 AZN-dir.  
Bu rəqəm riskli halların daha böyük məbləğlərlə bağlı olduğunu göstərir.





# Model Qurulması və Qiymətləndirilməsi

Layihədə həm Logistic Regression, həm də Random Forest modelləri sınaqdan keçirilmişdir. Məqsəd, fraud əməliyyatlarını yüksək dəqiqliklə proqnozlaşdırmaq idi.

Model seçimində balanssız məlumatların təsiri və əsas qiymətləndirmə metrikaları nəzərə alınmışdır.

## Model Performans Göstəriciləri



### Accuracy

Modelin ümumi düzgünlük səviyyəsi: 0.82

### Recall

Fraud əməliyyatları arasında düzgün təyin edilənlərin faizi: 0.89

### Precision

Fraud kimi təyin edilənlərin həqiqətən fraud olma ehtimalı: 0.77

### F1-score

Precision və recall-un harmonik ortalaması: 0.82

Layihənin son mərhələsində təlim edilmiş model real-time tətbiqə integrasiya edilmişdir. Streamlit əsaslı bu tətbiq istifadəçidən əməliyyat məlumatlarını qəbul edir və modelin çıxışına əsasən fraud ehtimalını proqnozlaşdırır.

Amount (\$)  
100  
0 10000

Merchant  
Amazon ▼

Transaction Type  
online ▼

Category  
travel ▼

Location  
Baku ▼


Gender  
☒ Male  
☐ Female

Hour  
12  
0 23

## Customer Fraud Detection A

### Daxil etdiyiniz məlumatlar:

	amount	merchant	transaction_type	transaction_category
0	100	Amazon	online	travel

 Fraud ehtimalını yoxla

# Streamlit tətbiqi və real-time proqnoz

Tətbiqin əsas imkanları:

1. İstifadəçi inputlarına uyğun ehtimal hesablanması

2. Vizual şəkildə ehtimalların göstərilməsi

3. Real-time input və nəticə uyğunluğu  
Model çıxışı: `fraud\_probability = 0.83`  
(yəni 83% ehtimalla bu əməliyyat fraud-dur)

# ML ilə Təchiz Edilmiş Real-Time Fraud İzləmə Paneli

Bu panel maşın öyrənməsi alqoritmləri ilə real vaxtda maliyyə əməliyyatlarını analiz edir və potensial saxtakarlıqları proqnozlaşdırır.

ROC-AUC score

0.82

Accuracy

0.81

Recall

0.89

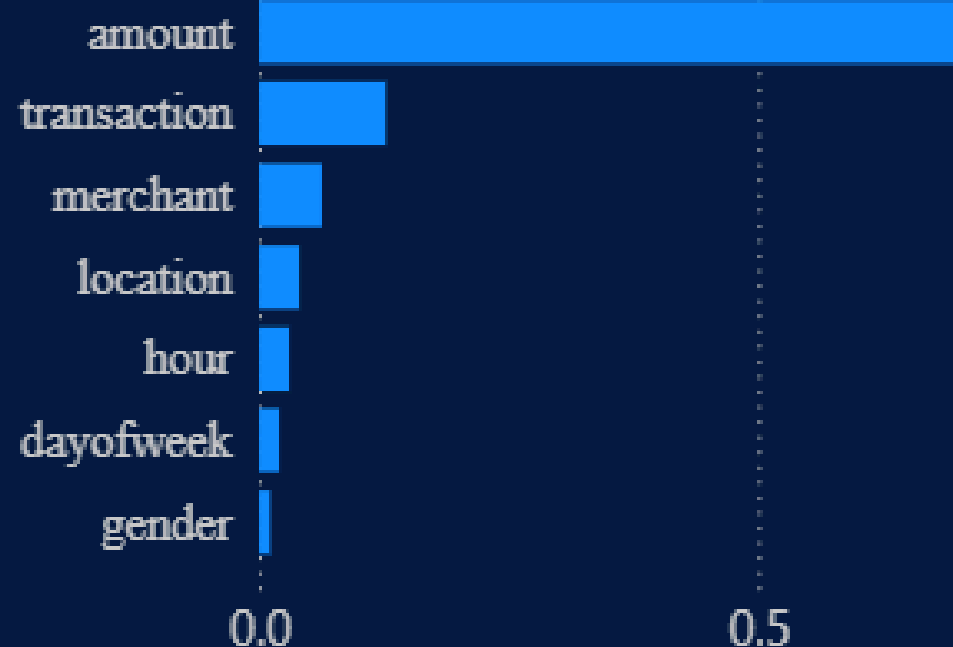
F1-Score

0.80

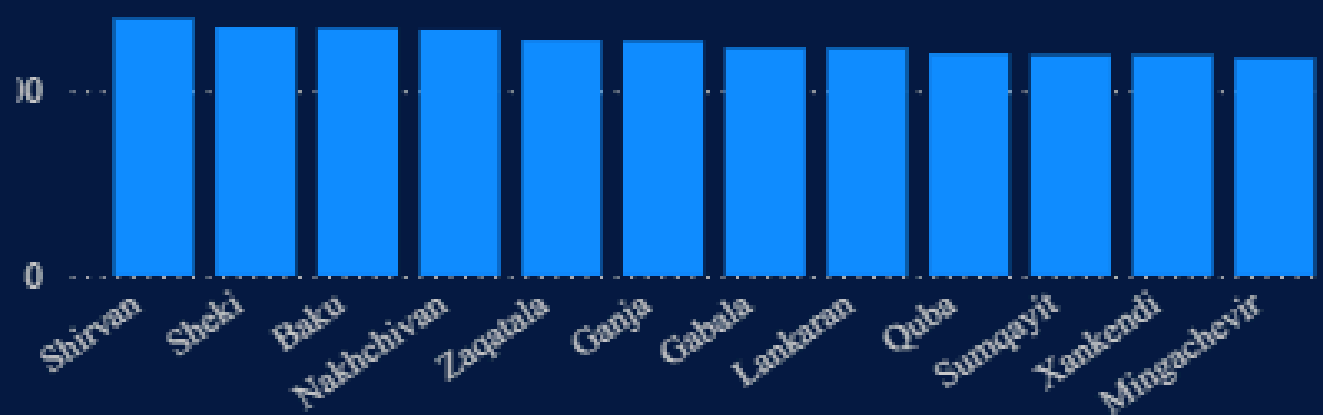
## Fraud ehtimalı ən yüksək 10 əməliyyat

card_id	məbləğ	tranz_tip	məkan	saat	gün	cins	fraud ehtimalı
4129971086965	4,649.72	atm	KFC	17	5	Male	0.84
4331566691593	3,132.46	atm	McDonalds	17	5	Male	0.85
3568197668699962	3,538.02	mobile_app	ABB	0	1	Female	0.84
4058571278712875	4,081.12	mobile_app	Aptek+	23	1	Male	0.86
4890797595162	3,424.52	mobile_app	Bolt	23	0	Female	0.84
4221182863423189000	8,194.97	mobile_app	Netflix	14	2	Female	1.68
3559981496119728	3,037.41	online	ABB	0	6	Female	0.84
676283843131	4,853.20	online	Spotify	17	6	Male	0.84
180066729572613	4,682.44	online	YouTube Premium	1	1	Female	0.84

## Feature Importance



## Region üzrə





# Nəticə və Gələcək Addımlar

Bu layihə çərçivəsində saxtakarlıq halları ilə bağlı davranış nümunələri analiz olunmuş və riskli faktorlar müəyyənləşdirilmişdir. Eyni zamanda maşın öyrənməsi modelləri ilə real vaxtda əməliyyatların proqnozlaşdırılması həyata keçirilmişdir.

Əldə edilən nəticələr göstərir ki:

- Ən çox təsir edən faktor əməliyyat məbləğidir
- Mobil tətbiqlər və online alış-verişlərdə daha yüksək risk müşahidə olunur
- Food, healthcare və utilities kimi kateqoriyalarda saxtakarlıq daha çox müşahidə olunur
- 

## Gələcək Addımlar:

- Airflow ilə Data Pipeline Qurulması
- SQL Əlavəsi və Məlumatların Baza İdarəsi
- Model Performansının Yaxşılaşdırılması

# Təşəkkürlər



aytac.ceferzade@gmail.com



050-709-29-69



[My LinkedIn Profile](#)