

Fraud Detection for Xente

Daniela, Kai-Yang, Fabio

Introduction and Project Objective



Our Stakeholder:

Xente is an financial services company in Uganda offering various products and services that can be paid for using Mobile Money.

Xente's Objectives:

Provide improved and safer services to its customers

Project Objectives:

- create a ML algorithm to **detect fraudulent transactions**
- obtain a **maximized F1-Score** (stakeholder requirement)

EDA

Data Overview - Dataset and Variables

Data Set

Number of Transactions

Total: 140,000
Training: ~95,000
Test: ~45,000

Time Period

Training: 15/11/18 - 13/03/19
Test: 13/02/19 - 14/03/19

Origin

Transactions from Uganda

Currency

Exclusively in UGX

Frauds - Overview

| FraudResult | |
|-------------|-------|
| 0 | 95469 |
| 1 | 193 |



0.2 % !

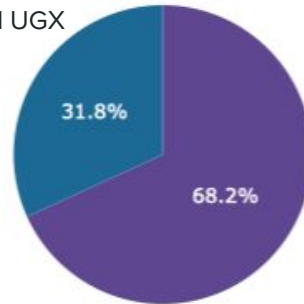


Extremely imbalanced dataset

Volume of Fraud transactions

Fraud:

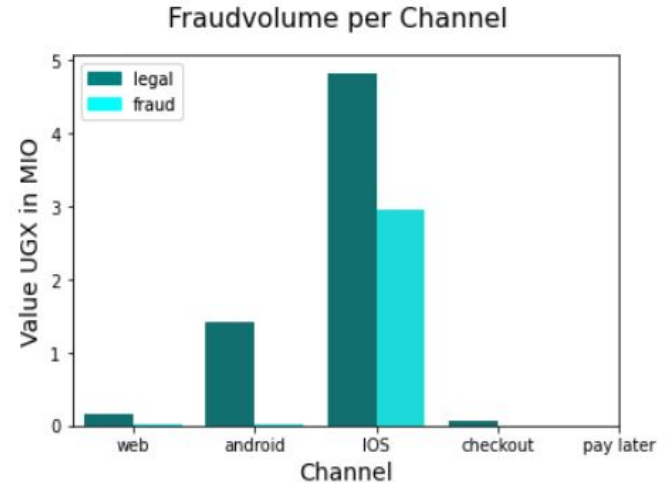
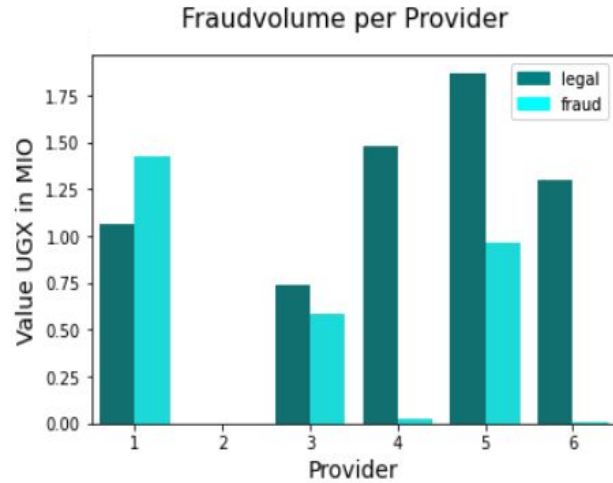
301.431.171 UGX



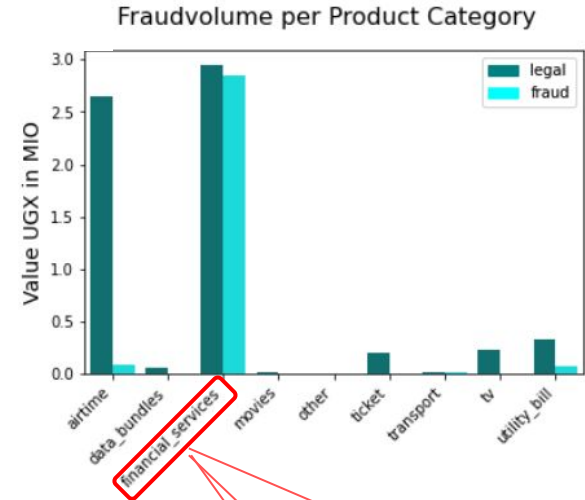
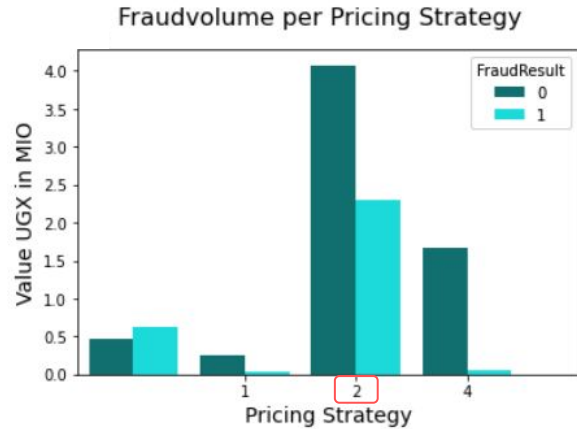
Legal:

645.678.490 UGX

Frauds - Provider and Channel

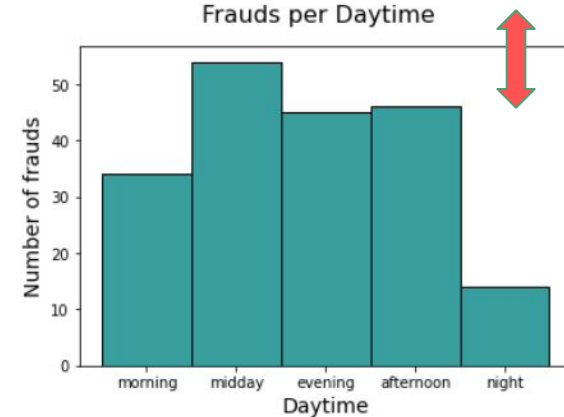
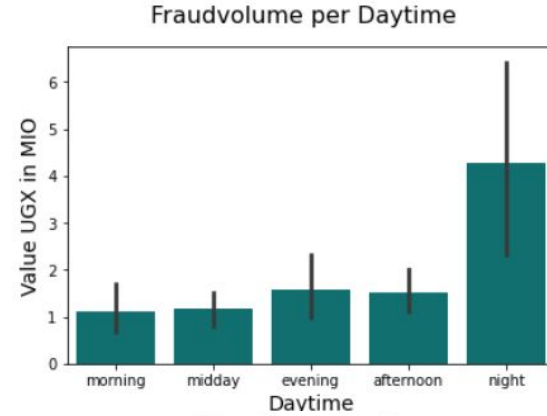
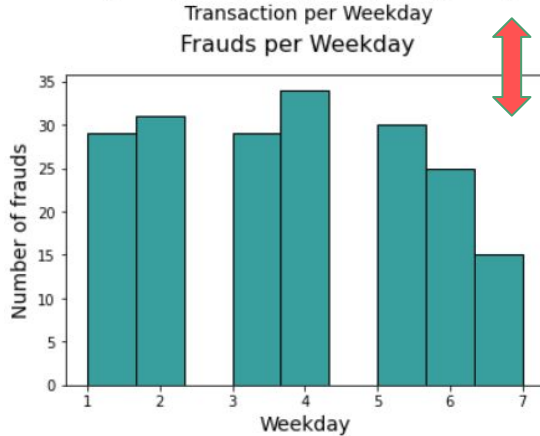
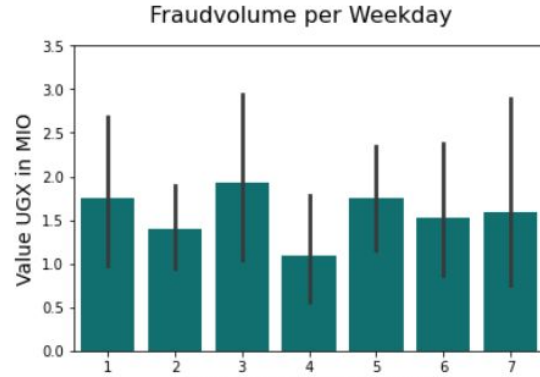


Frauds - Pricing Strategy and Products

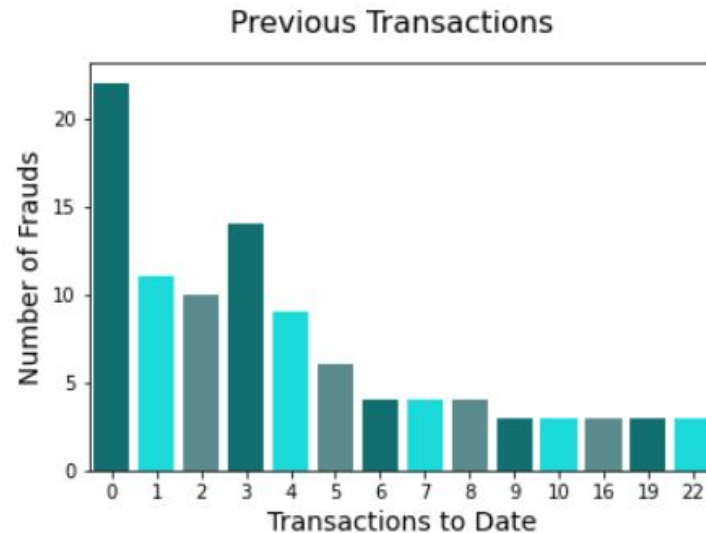
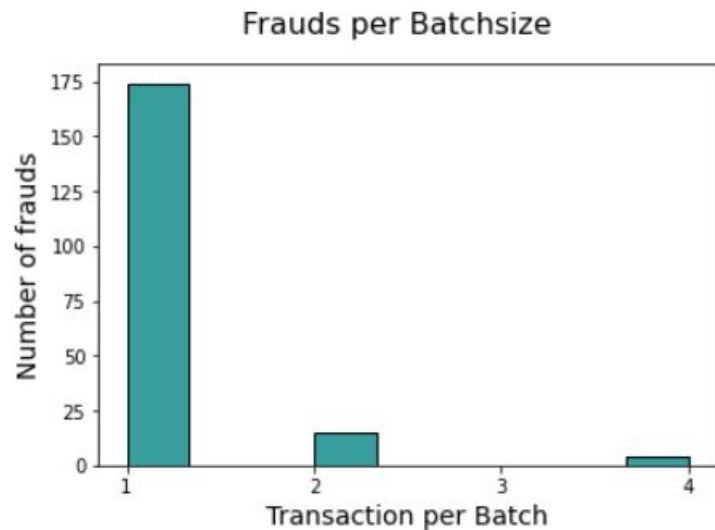


| ProductId | 3 | 5 | 6 | 9 | 10 | 13 | 15 | 21 | 22 |
|-----------------|-----|-----|-----|-----|-----|-----|-------|-----|-----|
| PricingStrategy | | | | | | | | | |
| 0 | 1.0 | 0 | 0 | 0 | 1.0 | 0 | 34.0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 2.0 | 0 | 0 | 4.0 | 1.0 |
| 2 | 2.0 | 2.0 | 1.0 | 3.0 | 1.0 | 6.0 | 123.0 | 0 | 1.0 |
| 4 | 9.0 | 0 | 0 | 0 | 2.0 | 0 | 0 | 0 | 0 |

Frauds - Time dependencies



Frauds - Batchsize and Transactions to Date



Feature Engineering

| Irrelevant Features | |
|----------------------------------|---|
| <i>No additional Information</i> | CurrencyCode CountryCode SubscriptionID BatchID TransactionID |
| <i>Redundant</i> | TransactionStartTime Amount |

MODEL

Baseline Model (BSL)

Stakeholder Requirement:

Xente requires a maximized **F1-Score**

Metric Characteristics:

- F1-Score is the harmonic mean between *precision* and *recall*
- **Precision**: how accurately are positive cases identified
- **Recall**: how many of true positive cases are identified as positive

BSL Characteristics:

Naive strategy: predict the minority class in all cases.

BSL Results:

F1-Score = **0.667** (resampled data)

Main Model

Model Analysis

Employed a **stacked model**

Advantages:

Combine capabilities of different models to deal

Model Description

Sub-models:

- Decision tree
- AdaBoost
- Random Forest

Meta model:

- Logistic regression

Model was further enhanced during training phase.

Model Performance

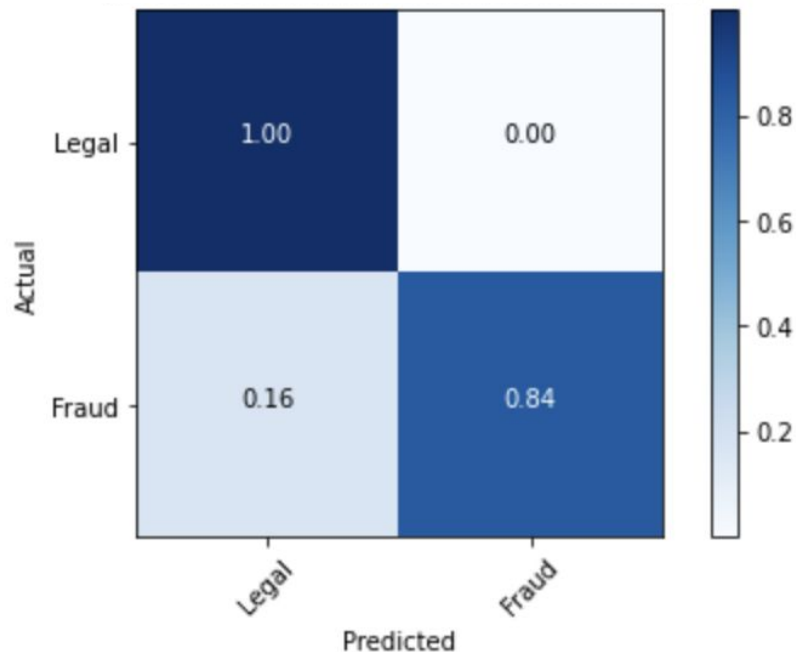
Performance on data set:

Precision - 0.84

Recall - 0.84

F1-Score - 0.84

Confusion Matrix



Model Performance

Fraud Cases:

- 84% identified
- 16% not identified

Legal Cases:

All cases correctly identified.

Saving ~253 Mio. UGX by using the stacked model.

CONCLUSIONS

Conclusions: patterns of fraudulent transactions

- Source provider of item: not a criterion
- Almost all frauds happened in **Channel 3: IOS** and in category: **financial services**
- The majority can be found in **pricing strategy 2**
- No abnormalities of amount volume between weekdays
- Substantially less at night, but with the highest volumes
- Not hidden in transaction batches
- Fraudulent transactions often occur as the **first or second transaction** of an account

Recommendations: What should be paid attention to?

- The **provider is insignificant**
- Increase **monitoring Channel 3** (IOS), **pricing strategy 2** and financial services
- Check **large transaction volumes** set **at night**
- A **frequently used account** usually doesn't commit frauds
- Doublecheck **single transaction batches**

BACKUP

Data Overview - Dataset and Variables

Data Set

*Number of
Transactions*

Total: 140,000
Training: ~95,000
Test: ~45,000

Time Period

Training: 15/11/18 - 13/03/19
Test: 13/02/19 - 14/03/19

Origin

Transactions from Uganda

Currency

Exclusively in UGX

Variable Set

*Transaction
related*

ID
Batch ID
Account & Customer IDs
Subscription ID

*Product
related*

ID & Category
Provider ID
Sales Channel
Value
Pricing Strategy

Feature Engineering

Added Features

Time related

Weekday
Week of Transaction
Daytime

Transaction related

Number of Transaction before fraud occurred;
Batch Size;
Difference between transaction Amount and Value;
Debit / Credit

Irrelevant Features

No additional Information

CurrencyCode
CountryCode
SubscriptionID
BatchID
TransactionID

Redundant

TransactionStartTime
Amount