Creating Risk Scores From Telco Data Using Advanced Machine Learning- Turkcell Case Study Review

Emrah Anayurt, COO, Organon Analytics
Brian Gorman, Senior Commercial Proposition Manager, GSMA
Creating Risk Scores From Telco Data Using Advanced Machine Learning

• Telcos generate huge amounts of data but creating insight can be a challenge because of the scale of the data pools.

• This presentation will show how Turkcell and its partner Organon Analytics have utilised Machine Learning to create new propositions from data insights.

• We will explain the basic Machine Learning process and feature case studies to demonstrate some of the go to market activity.
We use our own advanced machine learning platform to help Turkcell analyse vast data pools and create new insights and propositions that would not have been possible.

1. Reduce dependency on Data Scientists

2. Time to market < 5 days

3. High Accuracy
What we do for Turkcell: ML BASED PREDICTIVE MODELLING

ML BASED PREDICTIVE MODELLING PROJECTS THAT ARE LIVE IN TURKCELL

• NEXT BEST ACTION
• CUSTOMER REASON TO CONTACT
• CALL CENTER DEMAND PREDICTION
• CHURN PREDICTION
• AI BASED CYBER SECURITY
What we do for Turkcell: Omnichannel Next Best Action

70 DIFFERENT OFFERS WITH 70 DIFFERENT PROPENSITY MODEL

USING ORGANON AUTOMATED MACHINE LEARNING TO PREDICT THE BEST FIT FOR EVERY CUSTOMER

6.1 times increase on upsell

Model runs daily and produces scores for every customer
Vision is to use telco data and advanced ML to create predictive models for other industries.

**Turkcell Analytics as a Service**

**VISION**

- Fraud Risk Services
- Credit Risk Services
- Customer Digitalization Predictions
- Anonymised Location Based Demand Predictions

**Industry Focus**

- Banks
- Insurance
- E-commerce
Fraud Risk – Paycell Use Case overview

Paycell is a payments business in Turkey owned by Turkcell offering e-money products. The goal is to reduce fraud in payments eco-system.

Turkcell provides additional behavioral information on Paycell’s customers.

Machine learning

A Fraud risk scoring model is created that predicts the likelihood of a specific transaction being fraud. Paycell can deploy in real time in payments authorisation process to reduce fraud.
How it Works: Modelling

01 Paycell
- Paycell shares hashed msisdn and fraud/non fraud indicator for modelling (%80 of data)
- Paycell shares a testing sample (%20 of data) no fraud indicators

02 Turkcell
- Turkcell provides customer data to be used in the modelling

03 Organon
- Organon matches Paycell msisdn with Turkcell data for that msisdn (fraud/non fraud)
- Model developed in 2 days
- Organon uses the risk model to score the test sample

04 Paycell
- Paycell reviews the score for msisdns in test sample to confirm accuracy of the fraud risk scoring model
How it Works: Data Security

A. Organon Analytics software resides on a server on Turkcell’s cloud platform.
B. Turkcell controls access and Organon has authorised remote access.
C. The server is not connected to the network directly.
D. Turkcell cannot see Paycell data. Paycell cannot see Turkcell data.
E. Any data that is related to identification of a customer is hashed so that Organon cannot identify individual customers.

Design is driven by Turkcell's data security strategy.
How it Works: Data Privacy

Paycell & Turkcell use the same hash algorithm so that customer can be matched

1. Paycell provides hashed MSISDN & fraud / non fraud flag

2. Turkcell provides hashed MSISDN and customer data

3. Organon matches on hashed MSISDN and builds model using Paycell and Turkcell data. Organon doesn’t hold the hash key.

4. Organon provides hash MSISDN & risk score to Paycell

A. Raw data is not shared with Paycell, just the fraud score.
B. Organon cannot reach real subscriber information because MSISDN’s are hashed.
The fraud risk score that is produced is between 0 – 1, 1 being the highest risk.

There are 35 different variables, each have different weights in the model. Some examples are below:

- **Number of different devices a single sim card is used in**: If it's high it increases risk score.
- **Number of times personal information is requested from Turkcell via SMS**: If it's high it increases risk score.
- **Paid value added service membership**: If there is any paid membership, it reduces risk score.
- **Number of visits to a «Specific» web site**: If it's high it increases risk score.
### Raw Data of Customer Contact (Call Center/Web/SMS)

<table>
<thead>
<tr>
<th>SUBSCRIBER_ID</th>
<th>DATE</th>
<th>CALL_TYPE</th>
<th>SUB CATEGORY</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>23/12/2015</td>
<td>SMS</td>
<td>PERSONAL INFO REQUESTED</td>
</tr>
<tr>
<td>2</td>
<td>18/12/2015</td>
<td>IVR</td>
<td>GENERAL INFO</td>
</tr>
<tr>
<td>2</td>
<td>19/12/2015</td>
<td>BRANCH</td>
<td>TRANSACTION</td>
</tr>
<tr>
<td>2</td>
<td>21/12/2015</td>
<td>SMS</td>
<td>TRANSACTION</td>
</tr>
<tr>
<td>2</td>
<td>18/12/2015</td>
<td>WEB</td>
<td>GENERAL INFO</td>
</tr>
<tr>
<td>2</td>
<td>18/12/2015</td>
<td>WEB</td>
<td>GENERAL INFO</td>
</tr>
<tr>
<td>2</td>
<td>21/12/2015</td>
<td>SMS</td>
<td>TRANSACTION</td>
</tr>
<tr>
<td>2</td>
<td>22/12/2015</td>
<td>SMS</td>
<td>PERSONAL INFO REQUESTED</td>
</tr>
</tbody>
</table>

This is an example of a transactional data table of a subscriber (ID:2), showcasing the interactions this subscriber had with Turkcell on different dates and via different channels.

The first line would translate into: Subscriber ID 2 sent an SMS to Turkcell on 23rd of December 2015 to request personal information e.g. current bill.

Same data table for 30M+ subscribers would accumulate to **billions of rows of data**, and to search for patterns in these transactions would be impossible for a human.
Automated Feature Extraction Example:

What Automated Feature Extraction Does:

- It uses raw data sets to create summarization of these transactions, and turned them into features like the table in the right.
- This row would translate into; subscriber ID:2, as of 31st of December 2015,
  - has requested personal information from Turkcell via SMS 2 times,
  - %50 of the transactions this subscriber had with Turkcell was via SMS.
- And then machine tests these summarizations (features) to see if they are predictive of the Paycell fraud
- Predictive variables are used in the model to create the final risk score.

<table>
<thead>
<tr>
<th>SUBSCRIBE R_ID</th>
<th>DATE</th>
<th>Num_Rpi_SM S_L2D</th>
<th>Prc_SMS_L6M</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>31/12/2015</td>
<td>2</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Ratio of SMS Contacts in Last 6 Months: 0.50

Number of Requested Personal information through SMS in Last 2 days: 2
0.5% of highest risk scores

- will generate 43.5 X more fraud than the population average
- Equates to 21.8% of all frauds

**Lift**: Measure of the performance of a targeting model at classifying cases as having an enhanced response (with respect to the population as a whole), measured against a random choice targeting model.

<table>
<thead>
<tr>
<th>Score Percentiles</th>
<th>True Positive Rate</th>
<th>Lift</th>
</tr>
</thead>
<tbody>
<tr>
<td>P05</td>
<td>21.8%</td>
<td>43.59</td>
</tr>
<tr>
<td>P1</td>
<td>34.8%</td>
<td>34.78</td>
</tr>
<tr>
<td>P5</td>
<td>61.8%</td>
<td>12.37</td>
</tr>
<tr>
<td>P10</td>
<td>70.6%</td>
<td>7.06</td>
</tr>
<tr>
<td>Total</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Business Problem - Fraud In Payments System

- Customer opens Paycell e-money account
- No KYC.
- Unknown to Paycell customer is a fraudster

- Fraudster delivers fraud attack on a Turkish citizen and funds transferred to fraudsters Paycell account

- Fraudster uses Paycell account to purchase goods which can be sold for cash
- Existing fraud process wouldn’t block this transaction

Summary

- Liability is with customer as funds transfer was authorised
- Paycell keen to explore if Turkcell customer data can help with this problem and identify likely fraudsters
Fraud Model Deployment

Current transaction approval flow

Additional fraud process

Fraud scoring model creates fraud risk score

Customer data

Real time flows
Fraud Model Summary

• The model is developed to adhere to all data privacy/protection regulations in Turkey (KVKK)

• Fraud protection is confirmed as a legitimate interest not requiring explicit consent

• Data flows and integration into Paycell decision engine underway with a goal to be live December
Summary

- We have demonstrated how Turkcell is utilising the power of machine learning to enable new propositions and create new value pools.

- This process started internally for Turkcells own business operations.

- Options to build on Turkcells existing strong identity authentication and fraud products for enterprise are now being explored and we have shared thinking on one we have in development.

- Propositions are underpinned with Turkish privacy and data security processes - hashed msisdns, no PII transfer and customer consent.

- Our goal is to launch, capture learnings and play feedback in MWC 2020Barcelona.