Fighting Financial Fraud with Artificial Intelligence
Ron Bodkin
VP & GM, Artificial Intelligence, Teradata
@ronbodkin

Nadeem Gulzar
Head of Global Analytics, Danske Bank
@Nadeem_Gulzar
Founded in 2010 industry thought leader. = Vendor-neutral with an open source focus.

Fixed fee offerings for data science and engineering.

Who Is Think Big?

1st Big Data provider 100% focused around open source.

Full spectrum consulting, data engineering, data science & support.

Apache Hadoop and cloud ecosystem integration.

500+ employees (2017)

So far we have delivered 150+ successful projects for 100+ clients worldwide.
We help our customers be **financially confident** and achieve their ambitions by making daily banking and important financial decisions easy.

<table>
<thead>
<tr>
<th>236,000 small and medium-sized business customers</th>
<th>1,800 corporate and institutional customers</th>
<th>Making banking easier for <strong>over 145 years</strong>.</th>
</tr>
</thead>
</table>

**Who Is Danske Bank?**

Danske Bank is a Nordic universal bank with strong local roots and bridges to the rest of the world.

**Danske Bank** covers all Personal Banking, Business Banking, Corporates & Institutions and Wealth Management.

- **2.7 million personal customers**
- **19,000 employees (2017)**
- **Making banking easier for over 145 years.**

We help our customers be **financially confident** and achieve their ambitions by making **daily banking and important financial decisions easy**.
Data Driven Approach to Fight Fraud

Challenges for Fraud Detection

• Low Detection Rate
• Many False Positives
• High Fraud Loss
• Fast Growing Competition

Ambitions for Fraud Project

Danske Bank advanced analytics blueprint

Reduce false-positives & Enhance fraud detection rate

Data driven approach to real time scoring of transactions
Data Driven Approach to Fight Fraud

Challenges for Fraud Detection

Low Detection Rate
ONLY ~40% of fraud cases are detected

Many false positives
99.5% of cases are not fraud related

High Fraud Loss
Tens of Millions € lost each month

Fast evolving fraud sophistication

Ambitions for Fraud Project

Danske Bank advanced analytics blueprint

Reduce false-positives & Enhance fraud detection rate

Data driven approach to real time scoring of transactions

© 2017 Teradata
Fraud Types – Customer Initiated

- Rental Scam/ Goods Not Received
- CEO Fraud
- Nigeria/ Investment Scam
- Fake Invoice
- Beneficiary Account Change
Fraud Types – Fraudster Initiated

- SPEAR Phishing
- ID Theft
- Malware
- Phishing/Smishing
- Vishing/Support Scam

Fraudster Initiated

© 2017 Teradata
Modeling Challenges

- **Class imbalance**
  (100,000:1 non-fraud vs. fraud)
- Assigning fraud labels from historic data
- Fraud is ambiguous
- Not all features available in real-time
- Most machine learning sees transactions atomically
Advanced Platform for Fraud: Data-Driven Approach

How It Works:
• Understanding the domain
• Gathering and preparing the data
• Automatically generating the rules and recognizing fraudulent patterns by training models on historical data
• Automatically maintaining the engine by retraining the model

Pros:
• Automatic/data-driven/objective inference of the rules
• Ability to detect patterns in a high dimensional data input
• Fast detection of new/changing fraudulent patterns

Cons:
• Might be unintuitive and hardly interpretable
• Data preparation and feature aggregation is time consuming

© 2017 Teradata
Phase One
Teamwork to Deliver Value in Each Project Iteration

- **Project kick-off** 30th September
- **Cross-functional Collaboration**
  - Kick-off: Data Scientist track
  - Kick-off: Engineering track
- **2016**
  - Go-Live plan set for the project 19th December
  - First production virtual machine
- **2017**
  - First round trip test transaction
  - Models successfully in Shadow Production 4th March
  - Production Hardening for HA and Security
  - Full productionisation of the Fraud Engine
  - Cont. Deep Learning modeling

Sep | Oct | Nov | Dec | Jan | Feb | Mar | Apr | May | Jun | Jul
--- | --- | --- | --- | --- | --- | --- | --- | --- | --- | ---

© 2017 Teradata
Banking Anti-Fraud Solution

By leveraging the power of a thoughtful and strong data and analytics strategy, we unleash high impact business outcomes.

Model Management Framework

- Multiple models running in production at the same time
- Mix of traditional and advanced deep learning methods
- AnalyticsOps: Deploying machine learning models in production

Data Modelling, Pipeline and Ingestion

- Organisation and silos of data
- Real-time data integration
- Security and Procedures: following existing bank procedures

Banking Anti-Fraud Solution

Machine Learning and Artificial Intelligence

- Hard to operationalize insights
- Availability of analytic capabilities/skills and data
- Interpreting the results of machine learning models
Advanced Platform for Fraud: Data-Driven Approach
Framework to Enhance Danske Bank AnalyticsOps Capabilities
Key Requirement: Model Interpretation

- We have deployed LIME (Locally Interpretable Model Explanation) for customers
  - Improves trust
  - Compliance with EU’s General Data Protection Regulation (GDPR)

<table>
<thead>
<tr>
<th>Feature</th>
<th>% Score Due To:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Customer Amount</td>
<td>+ transfer amount</td>
</tr>
<tr>
<td>Debit Amount</td>
<td>+ destination country</td>
</tr>
<tr>
<td>Avg. # Trans Cred Acc</td>
<td>+ last year monthly spend</td>
</tr>
<tr>
<td># Prev to This Dest</td>
<td></td>
</tr>
<tr>
<td># Xfer Accounts</td>
<td></td>
</tr>
</tbody>
</table>

17.6% Fraud Probability
Machine Learning Results
(Live System: 60 transactions/sec.)

Ensemble of boosted decision trees and logistic regression.

From online validation of the model:
- 25-30% false positive reduction, with over 35% increase in detection rate
- Opportunity to expand model with additional features, retrain on recent data and add additional models to the ensemble.
- Models can be expanded to additional channels
Deep Learning
Deep Learning Opportunity

Current models can only catch ~70% of all fraud cases

Traditional ML models view transactions atomically

Often missed fraud transactions are part of a series

Capturing correlation across many features
Three Deep Learning Architectures to Deliver Value

**ConvNet**
- Designed for spatial correlated features, but by transforming transactions into a 2D image, we can learn temporal correlated features.
- Deeper ConvNet allows learning more complex & general features.

**Goal:** Learn kernels from temporal & static features to gain insight into the characteristics of fraud.

**LSTM**
- Learn temporal information and classify if the sequence of transactions contains fraud.
- Shares knowledge across learning time.

**Goal:** Learn transaction patterns within a window. Two solutions can be tested: flag fraud or predict next transaction and define an error.

**Auto-Encoders**
- Learn how to generate normal transactions, potentially large volumes of non-fraud data.
- AE provide a low level representation of the data.

**Goal:** Build a model that learns how to generate non-fraud data. To detect fraud, define a reconstruction error rate for the fraud cases.
# How Can We Create an Image From Bank Transactions?

<table>
<thead>
<tr>
<th>Input</th>
<th>Raw Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>t0</td>
<td>X₀, X₁, ..., Xₙ</td>
</tr>
<tr>
<td>t1</td>
<td>X₀, X₁, ..., Xₙ</td>
</tr>
<tr>
<td>t2</td>
<td>X₀, X₁, ..., Xₙ</td>
</tr>
<tr>
<td>ts</td>
<td>X₀, X₁, ..., Xₙ</td>
</tr>
</tbody>
</table>

| Output | Add correlated features in a clock-wise manner | Image size is: [10 x 3, 50 x 3, 1] |

### Top k Features Correlation

0: [41, 5, 30, 29, 31, 10, 37, 3],
1: [42, 40, 32, 15, 35, 2, 16, 31],
2: [3, 15, 4, 1, 28, 40, 31, 49],
3: [15, 41, 29, 16, 0, 2, 6, 14]

© 2017 Teradata

![Diagram](image-url)
Convolutional Layers for Trans2D

First Convolutional Layer Architecture

Kernels of 3x3

Strides of 3

Strides of 1

Features

Time

X_0 X_1 X_2 ... X_n

X_0 X_1 X_2 ... X_n

X_0 X_1 X_2 ... X_n

© 2017 Teradata
2D Transaction Image Example

Non-fraud Transaction Image

Fraud Transaction Image

Non-fraud

X-axis: features, Y-axis: time

© 2017 Teradata
Network Architecture for CNNs

© 2017 Teradata
Inside the ResNet model

- **64 Filters**
- **Activations**
- **After the CNN Residual Blocks**

<table>
<thead>
<tr>
<th>Non-fraud</th>
<th>Fraud</th>
</tr>
</thead>
</table>

© 2017 Teradata
Deep Learning
First Results
on the fraud verification dataset

Comparison of the three deep learning models and the traditional machine learning ensemble model.

- Ensemble model (AUC 0.89)
- ConvNets (AUC 0.95)
- LSTM (AUC 0.90)
- ResNet (AUC 0.94)
Lessons Learned: Take-Aways From Danske Bank

- Deep learning adoption from pictures to financial transactions
- Enhancement of data quality & cluster capabilities with data ingestion
- Building Analytics Ops capabilities to support business units
- Leveraging experience from Fraud advanced analytics to deliver extra use cases
Big Data Team – Lessons Learned

Success: From PowerPoint to production in 8 sprints

Team effort: Thorough collaboration across IMD, GFU and Think Big

Synergy: Successfully spearheaded innovation in all involved systems

Inspiration: Incorporation Danske Bank advanced analytics blueprint sets a generic scene for combatting new challenges in advanced analytics

Agile influence: Using an agile approach we were able to quickly deliver within the challenging timeframe.
Appendix
Advanced Analytics Fraud Platform

- Develop a scalable and expandable platform which follows the Danske Bank blueprint of digitalization
- 100% data-driven approach to find patterns in the data and complement the existing fraud engine
- Use Hadoop to handle the large data volumes for training models on transaction data
- Implement a real-time solution that can score live transactions such as e-banking, credit card and mobile payments.
- Reduce amount of false-positives by at least 20-40%
Ambitions for the Advance Analytics project

- The goal is to enhance fraud detection and reduce false positives across products.
- Fraud is only the first of many possible advanced analytics use cases.
- The project leads to the creation of Danske Bank advanced analytics blueprint.
- Danske Bank ambition is to become one of the bank leading in advanced analytics capabilities.

Challenges in Fraud Detection at Danske Bank

- **Low detection rate**: Existing human-written rule engine catches ~40% of fraud cases.
- **Many false positives**: 99.5% of all cases investigated are not fraud related.
- **High fraud loss**: Tens of millions of € total fraud per month.
- **Mobile payments** are quickly growing in number.
- **Fraud evolving rapidly**, with increased sophistication: Danske Bank must modernize its anti-fraud arsenal.