Predictive Financial Crime and Fraud Detection

Maurizio Megliola (GFT), 12/10/2021
Predictive Financial Crime and Fraud Detection

- **Financial crime** is crime committed against property, involving the unlawful conversion of the ownership of property (belonging to one person) to one's own personal use and benefit.

- Financial crimes may involve **fraud** (e.g., cheque fraud, credit card fraud, mortgage fraud, *bank fraud*), **identity theft**, **money laundering** and **terrorism financing**.

- As **technology** has grown, so has financial crime, and while continuing to provide convenient and accessible services to their customers, financial institutions also need to ensure they are protecting them from emerging financial crimes.

- **Organizations** have a wealth of data at their disposal. Leveraged properly, that data can help predict financial crime incidents such as money laundering. **Data, Analytics and Technology combine in the fight against Fraud and Financial Crime.**
Cluster Pilots

• Pilot #7 - Avoiding **Financial Crime**
• Pilot #8 - Platform for **Anti Money Laundering Supervision**
• Pilot #9 - Analyzing **Blockchain Transaction Graphs** for Fraudulent Activities
• Pilot #10 - Real-time cybersecurity analytics on **Financial Transactions’ BigData**
• Pilot #16 - Data Analytics Platform to detect payments anomalies linked to **money laundering events**
Pilot #7 Avoiding Financial Crime

Impersonation fraud is based on a fraudster assuming victim’s genuine personal data to access its online banking account.

- The scam is committed to the client...

- What can we do as a bank apart of security awareness training campaigns to our clients?

- Can we identify that it is not the “normal” behaviour of our client?

Pilot #7 objective: Improve the detection of Financial Crime using refreshed customer data and AI based analysis of customers’ transaction behavior.

- Customers’ behavioral data collection.
- Customer risk profiling.
- Pilot use case: Immediate Loans’ fraud risk reduction

Digital onboarding is the process of filling out forms, sending ID online. Without the need of having the client face to face.

With the credentials of a client, fraudsters can do fraudulent transactions in the name of the clients.

User patterns to assure that is the same user based on his regular behaviour.
Pilot #7 – Key Innovation

BUSINESS INNOVATION

• Several ML tools and models that banks try to apply to improve their fraud detection capacities. However, most of them are built for general purposes.
• Primary Pilot use case target, ‘Instant loans’, are critical operations that can lead to losses for the clients and the bank and need to be properly secured.
• The pilot targets some business KPIs such as the reduction of time in the fraud detection analysis, the reduction of false positives, and the number of new detected frauds and the total amount of fraud avoided (€).

TECHNICAL INNOVATION

• The INFINITECH technologies provide various added value. Besides supporting the chain from data preparation to scoring models and their visualization, the highest value is in facilitating real-time evaluation relying on the actual data thus minimizing false positives or negatives with both improving the user and the bank experience.
• Generalist tools like DataRobot (https://www.datarobot.com/) or Dataiku (https://www.dataiku.com/) do not provide suitable support on data preparation or they lack specific models to focus on the problem.
• The INFINITECH solution fosters both ends of the data chain from preparation to the specific detection and scoring models.
• Major technical components facilitating the development beyond SotA are
  • Big Data management
  • Data Anonymization
  • AI/ML algorithms
Pilot #8 Platform for Anti Money Laundering Supervision (PAMLS)

PROBLEM:

❖ Risk Based Supervision: extensive use of resources.

❖ Commensurate level of supervision: challenge to review & evaluate large quantity of data from various sources.

❖ Prompt supervisory reaction on to new and emerging risks: requires a solution that enhance effectiveness and improves.

SOLUTION:

PAMLS
Risk Based Approach to Supervisory Activities

Screening tool
Search Engine
Risk Assessment Tool
Distribution Channel
Pilot #8 Architecture

This project has received funding from the European Union’s horizon 2020 research and innovation programme under grant agreement no 856632
Flagship initiative for Big Data in Finance and Insurance

Pilot#8 – Key Innovation

- **RegTech solutions for AML**
  - Many AML solutions on the market.
  - Used by financial institutions (FI) to comply with regulations on financial crime prevention.
  - Solutions supports FI daily process: data management, transaction monitoring, PEP screening, KYC, ID verification.

- **SupTech solutions for AML**
  - EU supervisory authorities started to promote the need for digitalization and innovative technology in supervision.
  - Not many AML SupTech solutions → in developmental or experimental stage.
  - Identified solutions enables Data analytics or Data collection.
  - **PAMLs** combines Data collection and Data analytics features for more transparent and effective supervisory planning and execution.

- **Stakeholders:**
  - Bank of Slovenia (BOS): central bank, data provider & end user (supervisory department)
  - Josef Stefan Institute (JSI): technical proxy
  - FIU: data provider, cooperation also on the supervisory level

**PAMLs**
Less effort for risk assessment process, more resources on the supervision

promote the experience and knowledge

This project has received funding from the European Union’s horizon 2020 research and innovation programme under grant agreement no 856632
Pilot #9 Analyzing Blockchain Transaction Graphs for Fraudulent Activities

- Blockchain crypto currencies and tokenized assets that are obtained fraudulently can go through various transfers on the blockchain and enter the financial systems in different jurisdictions.
- **Massive blockchain transaction graphs need to be analyzed.**
- Infinitech Pilot #09 is developing effective systems for monitoring of blockchain transactions:
  1. Scalable parallel transaction graph construction and analysis tools on an HPC cluster and
  2. User interface that provides transaction graph visualization.
Pilot #9 Architecture

- Blockchain-enabled Tokenization
- Blockchain Transaction Database Preparation
- Scalable Transaction Graph Analysis
- UI for Blockchain Transaction Report and Visualization
Pilot #9 Business Innovation

• Scalability: ability to handle high transaction throughputs
• Faster processing: parallelism
• Accurate results: complete transaction graph analysis
• Cost effective solution: use of open-source libraries
• Hyperledger support
• Other potential uses: IBAN transactions, business intelligence, tracing in Central Bank Digital Currencies
• New generation of public blockchains are expected to grow to thousands of transactions per second
• A service that is less human-intensive is needed
• Since European Blockchain Services Infrastructure (EBSI) includes both Ethereum and Hyperledger Fabric blockchain protocols, our transaction graph analysis system will be able to readily work on EBSI use cases.
Pilot #9 Technical Innovation

- **Big Data management**
  - Ability to manage and process dynamically growing transaction graphs
  - Distributed graph and processing

- **Graph algorithms**
  - Parallel feature extraction on massive graphs to be used in ML
  - Perform traversals on massive graphs and extract subgraphs
  - Scalable graph construction

- **High Performance Computing**
  - Developed software using MPI
  - System can run on a wide variety of systems

- **Blockchain solutions**
  - Process standardized smart token transactions
  - Support ERC20 token contracts standard
  - Trace stable coin transactions

---

**Parallel Graph Construction**

<table>
<thead>
<tr>
<th>No. of Nodes X No. of processes per node</th>
<th>Total No. of processes</th>
<th>Time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 x 1</td>
<td>4</td>
<td>832</td>
</tr>
<tr>
<td>8 x 1</td>
<td>8</td>
<td>514</td>
</tr>
<tr>
<td>12 x 1</td>
<td>12</td>
<td>483</td>
</tr>
<tr>
<td>16 x 1</td>
<td>16</td>
<td>324</td>
</tr>
<tr>
<td>16 x 16</td>
<td>256</td>
<td>279</td>
</tr>
</tbody>
</table>

**Performance of Parallel Graph Algorithms**

Ethereum Blockchain Dataset

- Using 16 nodes 16 MPI processes, parallel timings for:
  - Distributed transaction graph construction: 279 sec
  - Pagerank calculation: 130 sec
  - Mells Partitioning: 3090 sec (1 MPI process per node)
  - Queries such as degree distribution, summing total number of ERC20 transactions (involving very little communication): 28 sec, 6 sec
  - Subgraphs showing tracing to fraudulent blacklisted addresses: 2 - 13 sec
  - Connected components: 164 sec

This project has received funding from the European Union’s horizon 2020 research and innovation programme under grant agreement no 856632
Pilot #10 Real-time cybersecurity analytics on Financial Transactions’ BigData

Traditional approach to block potentially fraudulent transactions is through rules-based systems. Despite the appeal of the rules-based approach, it has natural limitations. A system for Batch Machine Learning training (and periodic retraining) combined with Stream Machine Learning prediction can address such limitations.

Such systems provide low coverage
- Only a few highly accurate rules can be found

They are limited by fixed thresholds
- Ideal value for threshold can change over time. Moreover, it might be better to have different threshold ranges

They fail to capture interactive effects
- Probing and finding these interactive effects is difficult

They have low relative performance
- Simple rules alone can’t find risky transactions

This project has received funding from the European Union’s horizon 2020 research and innovation programme under grant agreement no 856632
Flagship initiative for Big Data in Finance and Insurance

Pilot #10  Real-time cybersecurity analytics on Financial Transactions’ BigData

- **Combine supervised and unsupervised machine learning into a single fraud detection system**
- **Self-learning AI to continually retrain unsupervised model**
- **Machine Learning model retraining based on fraud analyst feedback**
- **Real-time machine learning to online fraud detection**

**Goal**
To leverage Machine Learning techniques to improve significantly the detection rate of fraud attempts and enable the identification of security-related anomalies while they are occurring by the real-time ML detection of the financial transactions and continuous batch ML retraining.

This project has received funding from the European Union’s horizont 2020 research and innovation programme under grant agreement no 856632
Pilot #10    Real-time cybersecurity analytics on Financial Transactions’ BigData

1. Provide ALIDA’s API.
2. Exploit pseudonymization component.
3. Provide visualization component’s docker image and implement INFINITECH’s continuous deployment pipeline through the INFINITECH CI/CD system.
Pilot #10  Key Innovation

**Business Innovation**
- **Lower costs** through efficiencies generated by higher automation, reduced errors rates, and better resource utilization
- Ability to address major growth area as **new types of frauds** are expected
- **Minimizing Disruption for Legitimate Customers**
- **Client trust and security**

**Technical Innovation**
- **Batch and stream Integrated** system to handle retraining while real time ML prediction occurs.
- Integration of **Supervised and Unsupervised AI Models**
- Machine Learning **retraining automation**
- Real **time big data processing and fraud prediction**

**Objectives/KPIs**
- Increase in the number of malicious behaviors identified $\geq 20\%$
- Increase in the number of transactions per minute analysed $\geq 30\%$
- Reduction in false positives $\geq 20\%$

This project has received funding from the European Union’s horizon 2020 research and innovation programme under grant agreement no 856632
Conclusions

For information:

• Pilot #07: Mario Maawad Marcos, mmaawad@caixabank.com
• Pilot #08: Klaudija Jurkošek Seitl, klaudija.jurkosek-seitl@bsi.si
• Pilot #09: Orkan Metin, orkan.metin@aktifbank.com.tr
• Pilot #10: Massimiliano Aschi, massimiliano.aschi@posteitaliante.it
• Pilot #16: Paolo Testa, paolo.testa@nexi.it

Predictive Financial Crime and Fraud Detection
Questions?

Contact: maurizio.megliola@gft.com