Fighting Financial Crime with Automated Machine Learning

Ray Mi
Customer Facing Data Scientist, DataRobot
1. Introduction DataRobot
2. Financial Crime Challenges and Opportunities
3. AI-Powered Alert Prioritization
4. Demo
5. Q&A
DataRobot Enterprise AI Platform

ONE PLATFORM FOR ALL USERS

END-TO-END AUTOMATION

DEPLOY WHERE YOU CHOOSE

CREATORS
Data Engineers
Data Scientists
Business Analysts
Software Developers

OPERATORS
Dev Ops
IT
Risk and Compliance

CONSUMERS
Business Users
Senior Executives

CREATE
Prepare, Build, Test

OPERATE
Deploy, Monitor, Manage

CONSUME
Integrate, Use, Benefit

GOVERNED AND TRUSTED AI
Security
Collaboration
Explainability
Lineage

DataRobot

On-Prem, VPC, Hybrid or Fully Managed Cloud

Multi-Cloud

CREATORS
OPERATORS
CONSUMERS

DEPLOY WHERE YOU CHOOSE
On-Prem, VPC, Hybrid or Fully Managed Cloud

DataRobot

Confidential. ©2020 DataRobot, Inc. – All rights reserved
Most Popular Use Cases in Banking

**Credit**
- Consumer credit (Pd, LGD, EAD)
- Small Business, commercial credit and collateral valuation
- Prepayment risk
- Risk adjusted pricing, price elasticity
- CECL, CCAR, Basel
- Model Validation

**AML / Financial Crime**
- False positive reduction
- Suspicious activity monitoring
- Anomaly detection
- KYC/Due Diligence
- Realtime fraud detection and prevention

**Marketing**
- Relationship deepening
- Emerging needs identification
- Targeted offers
- Product campaigns and promotions
- Preferences (pricing, channel)
- Identifying the best leads and referrals
1. Introduction DataRobot
2. Financial Crime **Challenges and Opportunities**
3. AI-Powered Alert Prioritization
4. Demo
5. Q&A
**What's Money Laundering?**

**Money laundering** is the illegal process of concealing the origins of money obtained illegally by passing it through a complex sequence of **banking transfers** or **commercial transactions**. The overall scheme of this process returns the "clean" money to the launderer in an obscure and indirect way.

- **Collection of dirty money**
  - Tax evasion
  - Drug trafficking
  - Gambling
  - Bribery & Corruption
  - Human trafficking
  - Illegal arms sales
  - Terrorist financing
  - ...

- **Placement**
  - Sneaking dirty money into a business or legal account

- **Layering**
  - Moving the money around and mixing it with legitimate money until its original source is obfuscated
    - Transfer on the bank account of company ‘X’
    - Wire transfer to offshore bank
    - Loan to company ‘Y’
    - Payment by ‘Y’ of false invoices to company ‘X’

- **Integration**
  - Getting the now apparently legitimate money back out of the financial system
  - Luxury assets, financial investments
Financial Crime Current Landscape

The United Nation research states that the estimated amount of money laundered globally in one year is 2-5% of global GDP, or $800bn-$2tn in current US dollars*

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>58 AML penalties handed down globally in 2019, record year in terms of number of penalties</td>
<td>$8.14bn Monetary penalties imposed by regulators for compliance failures in 2019, doubling the penalties handed out in 2018</td>
<td>$48m The average global bank spend per year on their AML compliance programs - manual and resource intensive</td>
<td>95%+ On average, it is expected that between 95-99% of alerts are not filed as SAR's, but have to manually reviewed</td>
</tr>
</tbody>
</table>

With the cost of transaction monitoring capabilities increasing and increased regulatory pressure to increase personal liability on failed compliance program failure, many organisations are now looking to disruptive technologies to solve these challenges.

Anti-Money Laundering
Traditional Framework and Challenges

- Transactional data
- Client Profiles (CDD, KYC etc.)
- Sanctions/PEP /Watch list

Transaction Monitoring System (TMS)
- Rule-based suspicious activity detection system
- Large alert volume due to manual threshold tuning
- Detection scenarios only cover known AML risk
- Unable to detect emerging money laundering pattern

Alert Investigation
- Time consuming and labor intensive
- Large volume of false positive alerts
- Lack of alert triage and prioritization process
- Investigator notes are collected but not used

True Positive

Suspicious Activity Report (SAR)
False Positive

Investigator notes are collected but not used
Anti-Money Laundering
Opportunities for Automated Machine Learning

Efficiency Enhancement
- Alert Prioritization
- Text Mining
- Explainable Alert Prioritization
- AML Risk Factor Prioritization

Effectiveness Enhancement
- Anomaly Detection
- Pattern Detection
- AML Typology Optimization
- Risk Indicator Insight

AI-Driven AML
- Automation & Guardrail
- Deployment & Integration
- Compliance Documentation
- Model Monitoring Management
Anti-Money Laundering Opportunities for Automated Machine Learning

Efficiency Enhancement
- Alert Prioritization
- Text Mining
- Explainable Alert Prioritization
- AML Risk Factor Prioritization

Effectiveness Enhancement
- Anomaly Detection
- Pattern Detection
- AML Typology Optimization
- Risk Indicator Insight

AI-Driven AML
- Automation & Guardrail
- Deployment & Integration
- Compliance Documentation
- Model Monitoring Management
Agenda

1. Introduction DataRobot
2. Financial Crime Challenges and Opportunities
3. AI-Powered Alert Prioritization
4. Demo
5. Q&A
Alert Prioritization Solution Design

- **Transactional data**
  - Client Profiles (CDD, KYC etc.)
  - Sanctions/PEP/Watch list

---

**Transaction Monitoring System (TMS)**

---

**Alert Investigation**

---

**True Positive**

---

**Suspicious Activity Report**

---

- **Significant false positive reduction** without missing any historical SAR
- **Risk-based** alert prioritization score
- **Trusted and explainable** alert triage
- Identify real high risk money laundering patterns
- Seamless Integration with transaction monitoring and case management systems

---

**New Alerts**
- High risk entity in the **Below-The-Line** population
- **Hidden pattern** uncovered by various insights
- **Anomaly detection** to unveil unknown risk

---

**SAR Risk**
- Ultra-High
- High
- Medium
- Low
- No-Risk

---

**Action**
- Auto-Escalation
- Smart alert allocation based on investigation team expertise
- Hibernation

---

**DataRobot**

---

**Production Alerts**

---

©2020 DataRobot, Inc. – All rights reserved
Alert Prioritization Machine Learning Model Life Cycle

Data Preparation
Data collection and feature engineering leveraging DataRobot & domain knowledge

Model Training
Train dozens of machine learning models and select the best to predict the likelihood of alert being SAR

Model Deployment
Deploy models with optimal thresholds to improve investigation efficiency & reduce false positives

Model Monitoring
Monitor data drift, model accuracy decay and retrain models to capture emerging money laundering pattern

Full Automation
on data collections, model, predictions and model retrainings
“Coming up with features is difficult, time-consuming, and requires expert knowledge”

— Andrew Ng, co-chairman and co-founder of Coursera, and adjunct professor at Stanford University
Alert Prioritization Model Development & Validation

With its ability to dynamically learn patterns in complex data, DataRobot will significantly improve your accuracy in predicting which cases will result in a SAR filing.

Any case that exceeds a predetermined threshold of risk is sent to your investigators for manual review. Meanwhile, any case that falls below the threshold can be sent to a lighter review process.

“How to select the optional threshold?”

- Minimal **false negative** rate
- Maximal **false positive reduction** rate
- Consistent performance in both seen and **unseen** data
- 0 missing SAR
- 73% false positive reduction
- Validation on the holdout set (unseen data) shows consistent false positive reduction rate without missing any SARs
Alert Prioritization Model Explanation

When the reasons behind a model’s outcomes are as important as the outcomes themselves, Prediction Explanations can uncover the factors that most contribute to those outcomes.

“Why did the model give transaction a 90.2% probability of being SAR?”
Alert Prioritization Insights

The RuleFit Classifier blueprint in DataRobot generates **hotspots** to provide extra insights of what drives the alert escalation rate in different subgroups.

**83% SAR Rate**
- Average transaction size in last 90 days is greater than 538
- Total merchant credit amount in last 90 days is greater than $1492.7

**0% SAR Rate**
- Total merchant credit amount in last 90 days is between $972 and $1089
- Total amount overpaid in last 90 days is less than $2817
Once model is deployed, the Data Drift dashboard helps you to analyze how inputs data are changed over time and provide early alert to model decay.

“Some important features have experienced drift. How does it impact model performance?”
Alert Prioritization Model Retraining

Once model decay is observed, trigger retraining and redeploy up-to-date model, without disrupting downstream prediction consumers.

“On 9/28/2020, new model has been deployed to replace the previous model, due to data drift and accuracy drift concerns.”
Agenda

1. Introduction DataRobot
2. Financial Crime Challenges and Opportunities
3. AI-Powered Alert Prioritization
4. Demo
5. Q&A
Demo
Agenda

1. Introduction DataRobot
2. Financial Crime Challenges and Opportunities
3. AI-Powered Alert Prioritization
4. Demo
5. Q&A
Discussion and Q/A

Ray Mi
Customer Facing Data Scientist
ray.mi@datarobot.com

If we miss your question, please ask us here: community.datarobot.com
Find your path to ROI with proven use cases - https://www.datarobot.com/pathfinder
DataRobot Community

- **Engage, learn, and accelerate** your AI/ML journey
- **Connect** with peers to find solutions to AI challenges
- **Explore** helpful content to take your AI to the next level
- **Build** your brand as an AI expert & thought leader
- **Join** your peers today at community.datarobot.com
- **Questions**: learning_sessions@datarobot.com

Watch **15 new 6.2 Feature Videos** at the community now!