Financial Services: Machine Learning Use Cases and Solutions

Chanchal Chatterjee
May 15, 2018
Agenda

1. Machine Learning at a High-Level
2. Machine Learning for Financial Services
3. Financial Services ML Deployment Examples
4. Build Data Strategy Around ML
5. Financial Services Customer Stories
Machine Learning at High-Level
Machine Learning is a way to use standard algorithms to derive predictive insights from data and make repeated decisions.
Stage 1: Train an ML model with examples

- "cat"
- "dog"
- "car"
- "apple"

A ML model is a mathematical function

Make tiny adjustments to model function so output is closer to label for a given input

output

Output
Stage 2: Predict with a trained model
Neural networks is one important technology we use
“Machine learning. This is the next transformation... the programming paradigm is changing. Instead of programming a computer, you teach a computer to learn something and it does what you want.”

Eric Schmidt
Executive Chairman of the Board
Google
Machine learning use cases

Manufacturing
- Predictive maintenance or condition monitoring
- Warranty reserve estimation
- Propensity to buy
- Demand forecasting
- Process optimization
- Telematics

Financial Services
- Risk analytics and regulation
- Fraud detection
- Credit worthiness evaluation
- Customer segmentation
- Cross-selling and up-selling
- Sales and marketing campaign management

Retail
- Predictive inventory planning
- Recommendation engines
- Upsell and cross-channel marketing
- Market segmentation and targeting
- Customer ROI and lifetime value

Travel and Hospitality
- Aircraft scheduling
- Dynamic pricing
- Social media – consumer feedback and interaction analysis
- Customer complaint resolution
- Traffic patterns and congestion management

Healthcare and Life Sciences
- Alerts and diagnostics from real-time patient data
- Disease identification and risk stratification
- Patient triage optimization
- Proactive health management
- Healthcare provider sentiment analysis

Energy, Feedstock and Utilities
- Power usage analytics
- Seismic data processing
- Carbon emissions and trading
- Customer-specific pricing
- Smart grid management
- Energy demand and supply optimization
Advanced Data Analytics

Aggregating/Processing Data
- Need a large, scalable storage solution to aggregate, store, and serve applications
- Compute capacity required to churn and derive insights is constantly increasing
- Analytics & Machine Learning can be resource hogs

Understanding Cause and Effect
- What drives the customer's buying habits?
- What products do customers prefer to buy, and what related products?
- What causes customers not to buy?

Leveraging New Insights
- How can I promote the right products to the customer as quickly as possible?
- How can my applications react to customer actions in real-time?
- How can I attribute online and offline customer actions to product sales?
How technology fits into your top-level goals

- Grow Revenue
- Reduce Costs
- Innovate Quickly

- Customer Experience
- Operational Efficiency
- Ideas to Market

- Advanced Data Analytics
- Real-Time Solutions
- Innovative Applications
Google is a leader in open source

Kubernetes
Highest engagement on Github

Tensorflow
Highest engagement on Github

~1,100
Open source projects on GitHub received commits from Googlers in 2017

900+
Active employees in GitHub in 2017

7×
More repositories committed to than Amazon and 25% more than Microsoft

Source: Analyzing GitHub issues and comments with BigQuery
Operational ML

- Data Collection
- Data Management
- Exploration & Analysis Tools
- Feature Engineering
- ML Code
- Model Training at Scale
- Logging & Management
- Automation
- Monitoring
- Serving Infrastructure
## Operational ML - End to End ML Solution

<table>
<thead>
<tr>
<th>DATA COLLECTION</th>
<th>DATA MANAGEMENT</th>
<th>EXPLORATION &amp; ANALYSIS</th>
<th>SERVING INFRASTRUCTURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloud Pub/Sub</td>
<td>Cloud SQL</td>
<td>Monitoring</td>
<td>Cloud Machine Learning</td>
</tr>
<tr>
<td>Cloud Storage</td>
<td>Cloud Bigtable</td>
<td>Cloud Datalab</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Google BigQuery</td>
<td>Cloud Data Studio</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FEATURE ENGINEERING</th>
<th>ML Code</th>
<th>MODEL TRAINING AT SCALE</th>
<th>AUTOMATION</th>
<th>MONITORING</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloud Dataprep</td>
<td>TensorFlow</td>
<td>Cloud Machine Learning</td>
<td>Container Engine</td>
<td>Stackdriver</td>
</tr>
<tr>
<td>Cloud Dataflow</td>
<td></td>
<td></td>
<td>Cloud Functions</td>
<td>Monitoring</td>
</tr>
<tr>
<td>Cloud Dataproc</td>
<td></td>
<td></td>
<td>App Engine</td>
<td></td>
</tr>
</tbody>
</table>

| LOGGING & MANAGEMENT | | | |
| Stackdriver         | | | |

**Google Cloud**
Transform data into actions

Data Ingestion
- Mobile apps
  - Logs
- Web apps
  - Transactions
- Sensors and devices
  - Stream processing
  - Batch processing

Data Preparation & Processing
- Mobile apps
  - Logs
  - Data preparation
- Web apps
  - Stream processing
  - Transaction
- Sensors and devices
  - Batch processing

Databases
- Relational
- Key-value
- Document
- Object
- SQL
- Wide column

Storage
- Data exploration
- Federated query
- Data visualization
- Data catalog
- Development environment for Machine Learning

Analytics
- Data preparation

Advanced Analytics & Intelligence
- Pre-Trained Machine Learning models

Business analysts
Data scientists
Developers
<table>
<thead>
<tr>
<th>Data Ingestion</th>
<th>Data Preparation &amp; Processing</th>
<th>Databases/Storage</th>
<th>Exploration &amp; Collaboration</th>
<th>Analytics</th>
<th>Machine Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobile apps</td>
<td>Web apps</td>
<td>IoT Core</td>
<td>Google BigQuery</td>
<td>Cloud Machine Learning</td>
<td></td>
</tr>
<tr>
<td>Sensors and devices</td>
<td>Cloud Storage</td>
<td>Cloud Dataflow</td>
<td>Cloud Datalab</td>
<td>Cloud Dataproc</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cloud Pub/Sub</td>
<td>Cloud Dataprep</td>
<td>Cloud BigQuery</td>
<td>Google Drive</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cloud Storage</td>
<td>Cloud Datastore</td>
<td>Cloud Bigtable</td>
<td>Vision API</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cloud Bigtable</td>
<td>Cloud Storage</td>
<td>Speech API</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Google Analytics 360</td>
<td>Translate API</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Google Analytics 360</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Google Analytics 360</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Transform Data into Actions
Machine Learning is the end of the journey

- Big Data & Analytics (4)
- Delegation (2)
- Digitization (3)
- Individual Contributor (1)
- Machine Learning (5)
Machine Learning for Financial Services
Machine Learning Use Cases: Financial Services

- Customer self service portals
- Creditworthiness evaluation
- Credit scoring
- Risk analytics and regulation
- Fraud detection (payments)
- Anti money laundering (banks)
- Compliance and Regulatory Requirements
- Damage assessment (insurance)
- Claims processing (insurance)
- Robo advisors for portfolio management

- Customer Segmentation
- Personalized Marketing
- Cross-selling and upselling
- Sales and marketing campaign management
- Sentiment analysis
- Contact center automation
- OCR / document scanning / event extraction
- Monte carlo simulations
- Model explainability/interpretability
Customers expect 24/7 service options, fast, convenient, secure transactions, and personalized offerings.

1 in 3 millennials in the US are open to switching banks in the next 90 days and a similar proportion believe they won’t even need a bank in the future.
Leading marketers are **50% more likely** to increase investments in capabilities like machine learning to predict customer needs.

Source: Google/Econsultancy, Marketing and Measurement Survey, 2017
Top Banking Fraud ML Use Cases

- E-Banking fraud
- E-Business fraud
- Credit and Debit Card fraud
- Check & Document fraud
- Money laundering
- Identity Theft
- ATM image recognition

# Top Insurance ML Use Cases

- Predicting insurability of clients
- Predicting potential litigious cases and customers
- Fraud
  - Exaggeration of damages
  - Billing fraud
- Anomaly detection in image and speech

- Identify objects in images and texts used for underwriting
- Matching providers with clients for a given damage
- Identifying compliance and policies that match a given damage
### Perpetrators

<table>
<thead>
<tr>
<th>Type</th>
<th>Security</th>
<th>Cyber</th>
<th>Fraud</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ex-employees</td>
<td>23%</td>
<td>20%</td>
<td>39%</td>
</tr>
<tr>
<td>Permanent employees</td>
<td>17%</td>
<td>14%</td>
<td>30%</td>
</tr>
<tr>
<td>Freelance/temporary employees</td>
<td>16%</td>
<td>13%</td>
<td>27%</td>
</tr>
<tr>
<td>Competitors</td>
<td>12%</td>
<td>10%</td>
<td>27%</td>
</tr>
<tr>
<td>Random perpetrator</td>
<td>10%</td>
<td>10%</td>
<td>26%</td>
</tr>
<tr>
<td>Political activists</td>
<td>8%</td>
<td>7%</td>
<td>19%</td>
</tr>
<tr>
<td>Nation states</td>
<td>7%</td>
<td>6%</td>
<td>14%</td>
</tr>
<tr>
<td>Terrorists</td>
<td>5%</td>
<td>6%</td>
<td>13%</td>
</tr>
<tr>
<td>Don’t know</td>
<td>2%</td>
<td>2%</td>
<td>2%</td>
</tr>
</tbody>
</table>


© 2018 Google Inc. All rights reserved.
### Fraud

<table>
<thead>
<tr>
<th>Percentage of respondents affected by fraud in the past 12 months.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>80</strong></td>
</tr>
</tbody>
</table>

- **Most Common Types of Fraud**
  - Theft or loss of IP: 27%
  - Environmental risk: 26%
  - Management conflict of interest: 24%
  - Senior or middle management employees of our own company: 32%
  - Ex-employees: 30%
  - Customers: 17%
  - Information (IT security, process control): 16%
  - Financial data breach: 12%
  - whip data breach: 12%
  - Other (for example, system failures, theft, loss, or attack): 12%

- **Most Common Perpetrators**
  - Senior or middle management employees of our own company: 36%
  - Ex-employees: 32%
  - Junior employees of our own company: 32%
  - Ex-employees: 30%

### Cyber Security

<table>
<thead>
<tr>
<th>Percentage of respondents that experienced a cyber incident in the past 12 months.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>88</strong></td>
</tr>
</tbody>
</table>

- **Most Common Types of Cyber Incident**
  - Virus/worm infection: 42%
  - Data deletion or loss due to system issues: 26%
  - Email-based phishing attack: 21%
  - Ex-employees: 19%

- **Most Common Perpetrators**
  - Customer records: 57%
  - Social media: 36%
  - IT service vendor: 43%

### Security

<table>
<thead>
<tr>
<th>Percentage of respondents that experienced a security incident in the past 12 months.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>85</strong></td>
</tr>
</tbody>
</table>

- **Most Common Types of Security Incidents**
  - Theft or loss of IP: 30%
  - Environmental risk: 21%
  - Workplace violence: 15%
  - Competitors: 15%
  - Previous employee: 15%
  - Ransom perpetrators: 14%
  - Theft or loss of IP: 13%
  - Environmental risk: 9%

### CANADA REPORT CARD

<table>
<thead>
<tr>
<th>Percentage of respondents affected by fraud in the past 12 months.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>88</strong></td>
</tr>
</tbody>
</table>

- **Most Common Types of Fraud**
  - Theft of physical assets or stocks: 34%
  - Information theft, loss, or attack (e.g., data breach): 32%
  - Regulatory compliance breach: 32%
  - Vendor, supplier, or procurement fraud: 32%
  - Misappropriation of company funds: 32%

- **Most Common Perpetrators**
  - Senior or middle management employees of our own company: 47%
  - Junior employees of our own company: 39%
  - Senior or middle management employees of our own company: 36%

### Cyber Security

<table>
<thead>
<tr>
<th>Percentage of respondents that experienced a cyber incident in the past 12 months.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>85</strong></td>
</tr>
</tbody>
</table>

- **Most Common Types of Cyber Incident**
  - Virus/worm attack: 47%
  - Data deletion or loss due to system issues: 34%
  - Email-based phishing attack: 34%

- **Most Common Perpetrators**
  - Customer records: 57%
  - Physical asset: 57%
  - Trade secrets: 51%

### Security

<table>
<thead>
<tr>
<th>Percentage of respondents that experienced a security incident in the past 12 months.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>78</strong></td>
</tr>
</tbody>
</table>

- **Most Common Types of Security Incidents**
  - Theft or loss of IP: 42%
  - Environmental risk: 27%
  - Workplace violence: 25%
Financial Services ML Deployment Examples
Ocado Fraud Detection on the GCP

“Eventually we implemented a deep neural network on TensorFlow, as it was precise and easy to deploy into production. Using TensorFlow was a natural choice as we had already made the move over to Google Cloud for data analytics so using TensorFlow alongside our data stored on the Google Cloud Platform worked well. It also made our model scalable and transferable, which has in turn empowered our developers.”
Reduce False Positives and Negatives
Discover New Anomalies, Threat Vectors, and Bad Actors

Customer Challenge
Rules Based AML Monitoring is Inaccurate
- Using rules to monitor Issuer and Acquirer AML activity produces too many false positive and negatives
- Using rules to monitor Issuer and Acquirer AML activity fails to capture new trends, bad actors, and threat vectors
- This all leads to large scale inefficiencies, reputational risk, and increased operational costs

POC And Its Targeted Results with Saffron
Double digit accuracy improvement through lower false positives and negatives
- Saffron will ingest various forms of transactional and issuer/acquirer data plus truth and ontological data to analyze transactions and entities for Money Laundering behavior
- KPIs will be based upon reduction of false positive and negatives against a baseline
- An additional KPI will measure the number of new trends, anomalies or bad actors Saffron identifies
Reduce False Positives and Negatives
Discover New Anomalies, Threat Vectors, and Bad Actors

Thomson Reuters – Watchlist Enhancement

Customer Challenge
Rules and ML Based Watchlist Monitoring is Inaccurate

- Using rules and ML to monitor Watchlist activity produces too many false positive and negatives
- Using rules and ML to monitor Watchlist activity fails to capture new trends, bad actors, and threat vectors
- This all leads to large scale inefficiencies, reputational risk, and increased operational costs

Lexis Nexis – P&C Insurance Fraud

Customer Challenge
Rules and ML Based P&C Claims Fraud Monitoring is Inaccurate

- Using rules and ML to monitor P&C Claims Fraud activity produces too many false positive and negatives
- Using rules and ML to monitor P&C Claims Fraud activity fails to capture new trends, bad actors, and threat vectors
- This all leads to large scale inefficiencies, reputational risk, and increased operational costs
The need:
Approximately 7-10% of AXA’s customers cause a car accident every year. About 1% of these are so-called large-loss cases that require payouts over $10,000. AXA needed to understand which clients are at higher risk for such cases in order to optimize the pricing of its policies.

The execution:

The results:
- Achieved 78% accuracy in its predictions of large-loss incidents, which could give AXA a significant advantage for optimizing insurance cost and pricing.
- The possibility of creating new insurance services such as real-time pricing at point of sale.

The need:
MSS managers required frequent and easy access to newly organized data (including performance metrics and market intelligence) to make timely decisions. MSS sales teams needed granular-level knowledge to create what-if scenarios, and analytical tools to diagnose problems and offer a range of policies on tight deadlines to their customers.

The execution:
In partnership with Cloud Platform Partner CI&T, MSS used Google BigQuery for its lightning fast data processing capability and Compute Engine to power the MSS insights application. These two Cloud Platform products handled all security, scalability and infrastructure issues, allowing CI&T to focus its resources on developing a user-friendly interface and reporting system for MSS employees.

The results:
- 40% sales growth in a single year
- New data sets now can be built in hours instead of months using more traditional dimensional cubes.
- Critical business decisions have become completely data-driven

Source: Google Cloud Platform Blog (https://goo.gl/AvmDnX)
Determine “totaled” status from user photography

“Not Totaled”

“Totaled”
Process Automation

Google Cloud Improves Productivity for the Largest Car Auction in Japan

Aucnet created a system that automatically recognizes the make, model, year, and estimated value of cars, serving 30,000 car dealers. Aucnet partners save thousands of hours of time through automatic classification of uploaded images.

“How much is my car worth?”

Google Cloud
Build Data Strategy Around ML
Is this machine learning? What’s needed for ML?
Is this machine learning? What’s needed for ML?
Is this machine learning? What’s needed for ML?
If ML is a rocket engine, data is the fuel
Simple ML and More Data > Fancy ML and Small Data
Typical customer journey involves going from manual data analysis to ML

- Enables automation of previously manual global fishing data analyses
- Processes 22 million fishing data points daily
Financial Services
Customer Stories
<table>
<thead>
<tr>
<th>Service</th>
<th>Benefit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anti Money Laundering (AML)</td>
<td>Identify AML activity and reduce false positives</td>
</tr>
<tr>
<td>Finance Liquidity reporting</td>
<td>Reduce liquidity processing from six hours to six minutes</td>
</tr>
<tr>
<td>Risk Analytics</td>
<td>Raise compute utilisation from 10% to actual units consumed</td>
</tr>
<tr>
<td>Risk Reporting</td>
<td>Reduce time to calculate Risk Weighted Assets</td>
</tr>
<tr>
<td>Valuation Services</td>
<td>Rapid provisioning of compute power instead of an on premise grid</td>
</tr>
</tbody>
</table>

“We must acknowledge that we live in a world with a rapidly evolving threat landscape . . . We’re continually assessing every solution we have and always need to be one step ahead.”

David Knott
Chief Architect, HSBC
“Right at the start of the partnership we were able to reduce time to insight from 96 hours to 30 minutes by using BigQuery, allowing us to react in real time to customer needs and provide better service..”

Gary Sanders
Head of the bank’s digital analytics function

https://www.finextra.com/newsarticle/28566/lloyds-partners-google-on-data-analytics
Thank you!