Business AI: Auditable, Explainable, Fair, Causal - General Cognitive Solutions & Use Cases

Rajarshi Das, PhD, CTO and Peter Ritz, CEO
FatBrain LLC (fatbrain.ai)
Agenda

- Introductions
- Learned Vector Embeddings
- Auditability, Explainability, Fairness
- Policy Learning & Causality
- Use Cases
Problem: you vs. millennial data geometry

- Millennial data geometry* vs. classic DBs/Analytics
- Automated Learning vs. feature engineering
- Hyperscalers vs. data transactional gravity
- **Solution**: Bring Cognitive Cloud-in-a-Box to Data

* Term and Figure courtesy of FatBrain.ai blog-post (2017) *Digital Transformation, Innovation and AI Dialectic.*
Business AI* solutions landscape

<table>
<thead>
<tr>
<th>Professional Services</th>
<th>Hyper-scale AI Solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digital Reasoning</td>
<td>FatBrain</td>
</tr>
<tr>
<td>Opera Solutions</td>
<td>Clarifai</td>
</tr>
<tr>
<td>H₂O, Quid…</td>
<td>DataRobot</td>
</tr>
<tr>
<td>Palantir</td>
<td>Amazon</td>
</tr>
<tr>
<td>Cloudera</td>
<td>Google</td>
</tr>
<tr>
<td>Big 4 Consulting</td>
<td>Microsoft</td>
</tr>
</tbody>
</table>

* Business AI forms a significant part of the Cognitive Computing market as more fully considered in FatBrain’s post, id: Digital Transformation, Innovation and The AI Dialectic
FatBrain hyperconverged cognitive solutions

‘Cloud-in-a-Box’ + Solutions = like “Nutanix” for Cognitive

Cognitive Acceleration Bundle (starting at $250K):

- PowerGPU + SSD + FatBrain SaaS/Solutions
- Setup workshop + training (1\textsuperscript{st} POC w/ client data)
Automated “Vector” Learning for Outcomes
Real-time “Digital Twin” Decision-Making

FatBrain Solution Blueprints & use cases:
• Missing Data Quality (2-3x better imputation)
• AML/Fin Fraud (4BB TXN/day → 300 reports)
• Cross-Sell/Upsell (2% lift in term/whole life)
• Deep Intent Customer Care (20 interactions to 2)
• Pharma/Patient Care (75 to 38 days rev cycle)

Fast Start: 2-week deployment* after workshop

* Cognitive hyperconverged appliance, deployable anywhere K8s/Docker runs
Quantifying FATE solutions

Fair
• Principled, transparent analysis ensures you can make unbiased decisions.

Auditable
• Auto-generates a ledger audit trail of how every decision was derived.

Trustable
• Verifiable business decisions tied to data and model(s).

Explainable
• Quantifies the efficacy of possible decisions to assist in identifying the best choice.
FatBrain solution blueprints workflow (cs 2.0)

**Workshop**
- Desired business outcome, KPI, past decisions, payoffs
- Shape/form of the data, models?
- Available columns (attributes/labels) relevant to sought business outcome
- Establish/define FATE criteria
- Connect data/models SOR/APIs
- Where to put the results – existing SOR or App API?

**Connect**
- Graphs
- Text
- Structured & Unstructured
- Image
- Audio
- BYO-Labels
- BYO-Model
- Past decisions & payoffs

**Learn** (LVE)
- Document2Vec
- Concept2Vec
- Chat2Vec
- TXN2Vec
- Customer2Vec
- SKU2Vec
- Taxable2Vec
- Applicant2Vec
- Anomaly2Vec
- Loan2Vec
- Patient2Vec
- Graph2Vec
- BYO2Vec

**Decide** (Assist)
- Recommend
- Predict
- Classify
- Impute
- Cluster
- Custom
- BYO-Model
- Reinforcement Learning

**Explain** (Any Model)
- Fair
- Audit
- Trust
- Explain

**FatBrain Appliance Automation**

**Engineer**  **Analyst**  **Data Scientist**  **Business**
Desired business outcome, KPI, past decisions, payoffs

Shape/form of the data, models?

Available columns (attributes/labels) relevant to sought business outcome

Establish/define FATE criteria

Connect data/models/SOR/APIs

Where to put the results – existing SOR or App API?

Structured & Unstructured

BYO-Model

Past decisions & payoffs

Loan2Vec

Impute

BYO-Model

Reinforcement Learning

FatBrain Appliance Automation

Workshop (Outcome/KPIs)

Connect (Data/Model)

Learn (LVE)

Decide (Assist)

Explain (Any Model)

Fair

Audit

Trust

Explain

Establish/define FATE criteria

Connect data/models/SOR/APIs

WHERE TO PUT THE RESULTS – EXISTING SOR OR APP API?
F20 AML/Financial Fraud compliance

Workshop (Outcome/KPIs)
- Desired business outcome, KPI, past decisions, payoffs
- Shape/form of the data, models?
- Available columns (attributes/labels) relevant to sought business outcome
- Establish/define FATE criteria
- Connect data/models/SOR/APIs
- Where to put the results – existing SOR or App API?

Connect (Data/Model)
- BYO-Model
  - Past decisions & payoffs
- Graphs
- Text
- Structured & Unstructured

Learn (LVE)
- Document2Vec
- Concept2Vec
- Chat2Vec
- TXN2Vec
- Customer2Vec
- Applicant2Vec
- Anomaly2Vec
- Graph2Vec
- BYO2Vec

Decide (Assist)
- BYO-Model
  - Reinforcement Learning
- Predict
- Classify
- Impute
- Cluster

Explain (Any Model)
- Fair
- Audit
- Trust
- Explain

FatBrain Appliance Automation
OpenPOWER

F50 Disconnected Lakes (small data, RL)

Workshop (Outcome/KPIs)
- Desired business outcome, KPI, past decisions, payoffs
- Shape/form of the data, models?
- Available columns (attributes/labels) relevant to sought business outcome
- Establish/define FATE criteria
- Connect data/models/SOR/APIs
- Where to put the results – existing SOR or App API?

Connect (Data/Model)
- Text
- Structured & Unstructured
- BYO-Labels
- BYO-Model
- Past decisions & payoffs

Learn (LVE)
- Document2Vec
- Concept2Vec
- TXN2Vec

Decide (Assist)
- Predict
- BYO-Model
- Reinforcement Learning

Explain (Any Model)
- Fair
- Audit
- Trust
- Explain

FatBrain Appliance Automation

engineer analyst data scientist business
engineer analyst

analyst data scientist
analyst data scientist auditor
Using FatBrain’s Business AI Automation is like the Waze experience for navigation, instead of veering off and re-directing with old-fashion maps

**Classic Predictive Analytics**
- Manual feature discovery and engineering
- Millennial data geometry disconnected from RDM
- No common backplane for NLP, regression, etc.
- Domain specific
- Model(s) often opaque/NA
- Manual ABP resource and process intensive

**Business AI Automation**
- Automated learning thru vectorized embeddings
- Learned-policy realtime decision-making (RDM)
- Data-shape and type agnostic
- Domain independent
- Auditable & explained
- Anti-Bias Policy (ABP) enforcement
<table>
<thead>
<tr>
<th>Causation Layer (P=Probability)</th>
<th>Activity</th>
<th>Business Outcomes</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Association $P(y</td>
<td>x)$</td>
<td>- Observing</td>
<td>- What is?</td>
</tr>
<tr>
<td></td>
<td>- Counting</td>
<td>- How would seeing X change my belief in Y?</td>
<td>- What does a survey tell us about the election results?</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Intervention $P(y</td>
<td>do(x),z)$</td>
<td>- Modeling</td>
<td>- What if?</td>
</tr>
<tr>
<td></td>
<td>- Taking Action</td>
<td>- What if I do X?</td>
<td>- What if we ban cigarettes?</td>
</tr>
<tr>
<td>3. Counterfactuals $P(y_x</td>
<td>x',y')$</td>
<td>- Active Learning</td>
<td>- Why?</td>
</tr>
<tr>
<td></td>
<td>- Imagining</td>
<td>- Was it X that caused Y?</td>
<td>- Would Kennedy be alive had Oswald not shot him?</td>
</tr>
<tr>
<td></td>
<td>- Retrospection</td>
<td>- What if I had acted differently?</td>
<td>- What if I had not been smoking the past 2 years?</td>
</tr>
</tbody>
</table>

We automate learning causal relationships from both unstructured and structured data to drive realtime decision-making to achieve outcomes.
Learned Vector Embeddings

APIs, Data & Team Workflows, Architecture, LVE, FATE, RL, Benchmarks
Learned Vector Embeddings (LVE)
Vectorize {core banking, loan, ... wire data}

Outcomes Engine
action = recommend {group}

Learn

Infer & Explain

Decide

1. Connect to ODS
   Refine, Life-learn

2. Align API scope
   Business outcome

3. Data flow KPIs e.g.,
   5BB txns $\rightarrow$ 400 alerts

* Reflecting a reference approach for AML and financial fraud prevention with Top-5 global bank (some images courtesy of Arcadia Data)
Collaboration and control workflow*

1. Connect data, Set business KPIs
   - FatBrain Core API
   - FatBrain DeepIntent API
   - Resources
     - Model
     - Dataset
     - Precision

2. Vectorize, Segment, Analyze
   - Improved Model Risk Management

3. Review, Approve, Deploy
   - 1 framework vs. 100s models
   - Improved Model Risk Management

* Reflecting a reference approach for improved Model Risk Management with Top-10 global bank
Deployment architecture

1. Select/Refine (data)
2. Vectorize (data + context)
3. Learn (distance_metric, lens)
4. Persist/BC (vector + context)
5. API {...}, e.g., recommend ()

* See, Appendix for details on performance benchmarks IBM Z vs. IBM Power vs. x86
Learning behind the scenes

General cognitive approach shown for Data Quality project* to auto-learn hidden relationships in 100% of data, including complex group network, behavioral and temporal relationships in a quantifiable, auditable and explainable manner, revealing hidden strategic risk events.

* Excerpted from F50 global bank Data Quality project results
Learned Vector Embeddings (LVE): Automatically* learn latent vector representations (we call LVE) from any type of structured or unstructured data (images, speech, text, numbers or any combination of such data types) enabling automated application of machine learning/deep learning tools.

NLP: Google’s word2vec (word to vector): words represented as numerical vectors for linear algebra operations enabling reasoning, analogy-making and translation. (e.g., Vector(Human) + Vector(Robot) ≠ Vector(Cyborg)).

Search: Google Search moved from rule-based to learned vector space approach for serving search results based on distances between search words in vector space. No Hadoop-based counting technology can achieve that level of efficiency and flexibility (e.g., see also right side featuring 10,000x model size reduction for drug design case).

<table>
<thead>
<tr>
<th>method</th>
<th>dimension</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw</td>
<td>1.3 billion</td>
<td>0.096</td>
</tr>
<tr>
<td>Embedding</td>
<td>0.1 million</td>
<td>0.085</td>
</tr>
</tbody>
</table>

Reduce model size by 10,000 times!

Embeddings: 5x error reduction vs. traditional factorization

* See, Appendix for details AI CI/CD and following slides on reinforcement and active learning, including Le Song et al
A given learned non-linear decision function $f(.)$ may not be explained by a global linear model. With the LVE approach however, $f(.)$ can be explained at any point $\times$ using FatBrain’s zoom feature which samples instances in the neighborhood of $\times$, gets predictions using $f(.)$, and scores them by the proximity to $\times$ (represented by size). This is equivalent to estimating a local derivative along each independent variable, even though the details of the model $f(.)$ may be opaque. Dashed line in the Classification Use Case is the learned explanation that is locally (but not globally) faithful, while the blue line in the facing Regression Use Case is a similar learned regression explanation.

Classification Use Case

Regression Use Case

* perturbed sample local to point of Interest $\times$, distance reflected by size. For classification, shape represents class labels. Straight dashed line reflects inferred local linear model.

Input Features: X1, Categorical

Input Features: X2, Categorical

Input Features: X1, Categorical

Input Features: X1, Categorical

* See, Appendix for details on automated Fairness policy learning
The above sample graph shows an example of the Zoom API output, whereby features each contributing to influence the model output from the base value (the average model output over the training dataset) to the model output (24.41). Features influencing the prediction higher are shown in red, while those influencing the prediction lower are in blue. The below sample chart below shows Zoom API output for each data point, rotated vertically, and then stacked horizontally, for an entire dataset.

The nodes of the above graph are the variables in a loan dataset and the edge weights (thickness) between the nodes are defined by values of their correlation. The node size is determined by a node’s number of connections (node degree), node color is determined by a graph community calculation, and node position is defined by a graph force field algorithm.
Auto Anti-Baising

Protected Attributes (e.g., sex, race, age)

Fairness Criteria, e.g., avoid fewer qualified people* in the blue group getting loans vs. the orange group.

Bias Explained (zoom-in)

Original decision boundary (without any constraints) and the shifted decision boundary that was learned by the fair classifier. Notice how the boundary shifts to push more non-protected points to the negative class (and vice-versa)

Caution: Racist bot (MSFT)

Justice System (recidivism)

*Some of the images courtesy of Google Research
Reinforcement Learning (RL): Policy-driven decision-making beyond “actionable insights” helps businesses make and augment decisions over time, while learning continuously from past decisions and their consequences. Business AI systems learn thru automated experiments, what no programmer/expert could teach them.


- Google’s Alphago-Zero beat Alphago 100-0 ([https://deepmind.com/blog/alphago-zero-learning-scratch/](https://deepmind.com/blog/alphago-zero-learning-scratch/)). The “looser” Alphago beat the human world-champion last year which was hailed as a huge achievement in AI ([https://deepmind.com/research/alphago/](https://deepmind.com/research/alphago/)).


Hybrid RL (SARSA)
State-Action-Reward-State-Action

Bellman Equation for Action-Reward Function

Learned Policy with hysteris lag

Learned vs. Hardcoded Thrashing

*Images courtesy of Das et al.*
Biomimetic learning considerations

- Creating LVE is essential to learning general representation of relationships to drive outcomes, i.e., business problem-specific attributes, inputs or abstract concepts.
- LVE allows FatBrain to automate the central problem in learning – generalization, that is:
  - Applying what was discovered in the past experiences to future situations which are similar in relevant respects,
  - Using the learned representation, potentially in a generative fashion, to make business decisions to maximize a business notion of cumulative reward.
- In step with biomimetic and neuromorphic design of DNN, CNN and RL, FatBrain’s LVE parallels recent discoveries in neuroscience on how human brain organizes conceptual knowledge with a grid-like code.

Establish a baseline deployment test bed to operationalize
Cost per Business Decision across different infrastructures

Training

• Training across 3 canonical datasets comprising IMDB (sparse, text), MNIST and CIFAR (dense, numeric)
• Training across 3 kinds of neural networks: MLP, CNN and LSTM
• Quantify relationship between training time & type of compute deployed: z13, z14, Power and x86

Inferencing

• Scalability of Approximate Nearest Neighbors (NN aka “digital twin”)
• Graph relationship between query time, number of samples
• Compare query time with Brute Force method, same index sizes
• Test two sample/feature sizes – (100k, 1M)/(100, 1000)

*Comprehensive approaches must examine both training and inferencing to provide a practical business measure for operating AI/ML topologies (we used Tensor Flow for training and SK Learn for inferencing)
Training: IBM z13 vs. z14 vs. x86 vs. Power

**Training time in seconds**

(lower is better)

<table>
<thead>
<tr>
<th>Model</th>
<th>z14</th>
<th>z13</th>
<th>x86</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMDB CNN</td>
<td>55</td>
<td>50</td>
<td>103</td>
<td>6</td>
</tr>
<tr>
<td>IMDB LSTM</td>
<td>78</td>
<td>71</td>
<td>139</td>
<td>2</td>
</tr>
<tr>
<td>MNIST MLP</td>
<td>5</td>
<td>9</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>MNIST CNN</td>
<td>70</td>
<td>91</td>
<td>51</td>
<td>16</td>
</tr>
<tr>
<td>CIFAR CNN</td>
<td>109</td>
<td>224</td>
<td>100</td>
<td>16</td>
</tr>
</tbody>
</table>

*Note: z14 uses TensorFlow 1.2.1, Power uses v1.1.0, z13 uses v0.10.0, x86 uses v1.3.0*
Inferencing: IBM Z and Power platforms

IBM Z

IBM Power

100k samples

1M samples
Fast start and beyond

• First Project: start small, with one prioritized app
  • e.g., F50 Bank Global Risk Analytics *Data Quality* Project
  • Outcome™ engine deployed in 2 weeks (IaaS, 4Q17)
  • Internal investigative team trained on practical use cases
  • The Engine in production yielding decision-making results
  • Annual SaaS license sized by business transaction volume

• Next Project and Beyond:
  • Same engine deployed for AML then anti-fraud 1Q18
  • In the pipeline, application for Global Risk scoring
Use Cases

Data Quality, AML, Deep Intent, Transaction Behavior, Social Signals
### Goal:
Impute the distribution of missing values from a hybrid data set.

### Data shape:
1,520,641 loans each characterized by 25 data elements, comprising 14 numerical and 11 categorical elements (table)

<table>
<thead>
<tr>
<th>Name</th>
<th>Data Type</th>
<th>Num Unique Values</th>
<th>Sample Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 LOAN IDENTIFIER</td>
<td>int64</td>
<td>1520641</td>
<td>[100001806951, 100003792573, 100006270289, 100...]</td>
</tr>
<tr>
<td>1 ORIGINATION CHANNEL</td>
<td>object</td>
<td>3</td>
<td>[R, C, B]</td>
</tr>
<tr>
<td>2 SELLER NAME</td>
<td>object</td>
<td>21</td>
<td>[OTHER, WELLS FARGO BANK, N.A., BANK OF AMERIC...]</td>
</tr>
<tr>
<td>3 ORIGINAL INTEREST RATE</td>
<td>float64</td>
<td>1032</td>
<td>[6.125, 7.5, 9.25, 6.25, 5.625, 4.875, 6.0, 5,...]</td>
</tr>
<tr>
<td>4 ORIGINAL UNPAID PRINCIPAL BALANCE (UPB)</td>
<td>int64</td>
<td>786</td>
<td>[177000, 146000, 226000, 210000, 417000, 10000...]</td>
</tr>
<tr>
<td>5 ORIGINAL LOAN TERM</td>
<td>int64</td>
<td>207</td>
<td>[360, 160, 240, 300, 120, 297, 324, 322, 262, ...]</td>
</tr>
<tr>
<td>6 ORIGINATION DATE</td>
<td>date/time64[ns]</td>
<td>86</td>
<td>[2008-01-01T00:00:00.000000000, 2007-11-01T00:...]</td>
</tr>
<tr>
<td>7 FIRST PAYMENT DATE</td>
<td>date/time64[ns]</td>
<td>87</td>
<td>[2008-01-01T00:00:00.000000000, 2008-01-01T00:...]</td>
</tr>
<tr>
<td>8 ORIGINAL LOAN-TO-VALUE (LTV)</td>
<td>int64</td>
<td>97</td>
<td>[47, 95, 83, 80, 70, 42, 61, 43, 75, 77]</td>
</tr>
<tr>
<td>9 ORIGINAL COMBINED LOAN-TO-VALUE (CLTV)</td>
<td>float64</td>
<td>126</td>
<td>[47.0, 85.0, 83.0, 90.0, 70.0, 42.0, 61.0, 43...]</td>
</tr>
<tr>
<td>10 NUMBER OF BORROWERS</td>
<td>float64</td>
<td>9</td>
<td>[1.0, 2.0, nan, 3.0, 4.0, 5.0, 7.0, 6.0, 10.0]</td>
</tr>
<tr>
<td>11 DEBT-TO-INCOME RATIO (DTI)</td>
<td>float64</td>
<td>65</td>
<td>[nan, 43.0, 42.0, 33.0, 34.0, 46.0, 51.0, 30.0,...]</td>
</tr>
<tr>
<td>12 BORROWER CREDIT SCORE</td>
<td>float64</td>
<td>410</td>
<td>[862.0, 635.0, 704.0, 735.0, 700.0, 776.0, 637...]</td>
</tr>
<tr>
<td>13 FIRST-TIME HOME BUYER INDICATOR</td>
<td>object</td>
<td>3</td>
<td>[N, Y, U]</td>
</tr>
<tr>
<td>14 LOAN PURPOSE</td>
<td>object</td>
<td>4</td>
<td>[C, P, R, U]</td>
</tr>
<tr>
<td>15 PROPERTY TYPE</td>
<td>object</td>
<td>5</td>
<td>[SF, PU, CO, MH, CP]</td>
</tr>
<tr>
<td>16 NUMBER OF UNITS</td>
<td>int64</td>
<td>4</td>
<td>[1, 2, 3, 4]</td>
</tr>
<tr>
<td>17 OCCUPANCY STATUS</td>
<td>object</td>
<td>3</td>
<td>[R, L, S]</td>
</tr>
<tr>
<td>18 PROPERTY STATE</td>
<td>object</td>
<td>54</td>
<td>[CA, FL, MN, CO, IN, WA, AL, LA, MO, OH]</td>
</tr>
<tr>
<td>19 ZIP (5 - DIGIT)</td>
<td>int64</td>
<td>914</td>
<td>[535, 335, 533, 554, 801, 482, 983, 352, 708,...]</td>
</tr>
<tr>
<td>20 MORTGAGE INSURANCE PERCENTAGE</td>
<td>float64</td>
<td>30</td>
<td>[nan, 35.0, 30.0, 12.0, 25.0, 17.0, 6.0, 18.0,...]</td>
</tr>
<tr>
<td>21 PRODUCT TYPE</td>
<td>object</td>
<td>1</td>
<td>[FRM]</td>
</tr>
<tr>
<td>22 CO-BORROWER CREDIT SCORE</td>
<td>float64</td>
<td>284</td>
<td>[nan, 647.0, 732.0, 620.0, 762.0, 857.0, 717.0...]</td>
</tr>
<tr>
<td>23 MORTGAGE INSURANCE TYPE</td>
<td>float64</td>
<td>3</td>
<td>[nan, 1.0, 2.0]</td>
</tr>
<tr>
<td>24 RELOCATION MORTGAGE INDICATOR</td>
<td>object</td>
<td>2</td>
<td>[N, Y]</td>
</tr>
<tr>
<td>25 DELAY IN FIRST PAYMENT</td>
<td>int64</td>
<td>16</td>
<td>[80, 61, 31, 62, 91, 80, 92, 29, 59, 30]</td>
</tr>
<tr>
<td>Attribute</td>
<td>Impute %</td>
<td>Classic Linear Regression</td>
<td>FatBrain LVE</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>----------</td>
<td>---------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Test $R^2$</td>
<td>Test RMSE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Train $R^2$)</td>
<td>(Train RMSE)</td>
</tr>
<tr>
<td><strong>Original Loan to Value (LTV)</strong></td>
<td>25%</td>
<td>0.36 (0.36)</td>
<td>13.51 (13.38)</td>
</tr>
<tr>
<td>Mean: 70.66  Std: 16.86</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50%</td>
<td>0.36 (0.36)</td>
<td>13.50 (13.46)</td>
<td>0.93 (0.94)</td>
</tr>
<tr>
<td>75%</td>
<td>0.36 (0.36)</td>
<td>13.54 (13.46)</td>
<td>0.93 (0.94)</td>
</tr>
<tr>
<td><strong>Debt-to-Income Ratio (DTI)</strong></td>
<td>25%</td>
<td>0.10 (0.10)</td>
<td>12.15 (12.14)</td>
</tr>
<tr>
<td>Mean: 36.30  Std: 12.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50%</td>
<td>0.11 (0.10)</td>
<td>12.12 (12.17)</td>
<td>0.80 (0.86)</td>
</tr>
<tr>
<td>75%</td>
<td>0.10 (0.10)</td>
<td>12.11 (12.16)</td>
<td>0.76 (0.86)</td>
</tr>
</tbody>
</table>

**RED:** Classic Linear Regression (NFO)
**BLUE:** Learned Vector Imputation (AF)
• **Goal:** Improve operational efficiency (i.e., number of false positives leading to Alert Investigation without resulting in SAR) in the KYCC area by 3% with a stretch goal of 6%.

• **Input:** Use available Swift message data, with 100% of the data features, including transactional data (type, direction, value), customer data (geographical, chronological), and risk data (including Open Social Intelligence).

• **Project:** Create a series of segments using a subset of the data, keeping the group count constant, while creating more intelligent, defensible, and uniform groups using 100% of available feature data via LVE (topologically different from the bank-constructed groups).

• **Results:** Over 3x improvement in operational efficiency over stretch goal. Using LVE approach, FatBrain engine automatically created segmentation groups. The distribution of customers within these groups was then evaluated, independently validated, and deployed against the bank’s existing infrastructure. Improved Model Risk Management.

• **Resources:** Three (3) weeks with two FatBrain engineers during setup to train and work with two domain experts from the bank. Subject to its internal board and regulator approval regime, the bank is in the process of deploying the proposed solution globally.
FatBrain AI-augmented AML workflow

**Connect and learn from data**

Entities: Customers (KYC), Accounts (open and closed during a sample time period)

Communications: IP address, email, chat, ANI

Transactions: Wires, Posting, Credit Card, Cash

High Risk Lists: Countries, Correspondents, Known Bad Actors

**Profile behavior anomalies**

Detected activity linked to Known Bad Actor from the published list for Yemen-based fundraising

IP/zip code resolution to identify personal account usage to fund distributed international wires

Counterparty enforcement inquiry (high risk country cash movement)

**Investigate, learn, report**

Identified other previously unrelated linked open accounts

Profiled previously unknown linked-accounts for anomalous behavior

Digital-twin awareness benchmarking against known typologies

Audit trail for draft SAR recommendation

* Reflecting a reference approach for AML and financial fraud prevention with Top-5 global bank (some images courtesy of Arcadia Data)
Goal: Distill Interactions from 20 to 2

- Augmented Conversation
- Real-time Intent Quotient aligned to business outcome or BOM = “Likely to Pay”
- \( \text{IQ}_{\text{Likely To Pay}}(\text{“Have new job”}) = 0.8 \)

- Wouldn’t be nice where every service representative has the power to read the customer’s mind, in identifying the intent of the caller from unstructured speech/text in 2 questions instead of 20, & align it to business outcomes (BOM)?
- Imagine calling in to report being late on a lease payment for October, because you have a new job whose paycheck doesn’t arrive until the next month. All you are seeking is a reprieve for the current month’s payment.
- Without torturing you with 20 different questions with the typical voice-mail jail, imagine if the business could quickly quantify your intent with respect to the BOM = “likely to pay” from 2 questions and qualify you for the later repayment.
- Note, no records exist in the back office that customer has a new job, but the Deep Intent App helps identify and transform the intent of caller’s spoken request into a structured event to score it against the desired business outcome. That is, we can quantify that the intent quotient of BOM = “likely to pay”, given the customer’s input “have a new job”, is quite high: \( \text{IQ}_{\text{BOM=Likely To Pay}(\text{have a new job})} = 0.8 \)
- This information is then relayed to the service representative in real-time through augmented conversation for successful customer engagement for positive customer experience.
FatBrain Deep Intent workflow

Business Manager provides a Business Outcome, e.g., “Likely to Pay” to the Engine, tunable with learned insights e.g., “move” or “new job”

Business Outcome: “Likely to pay”

Watson feeds FatBrain engine unstructured transcripts from Agent vs. Customer voice calls

FatBrain engine scores all Agent vs. Customer conversations in real-time for Intent Quotient (IQ) quantified by the “Likely to Pay” business outcome – personalized for Causal Rewards & Customer2Loyalty

A: You pay back today?
C: I can’t pay (by) the phone

A: The past year amount you sent.
C: **** be with a check in.

C: Okay perfect yeah I can pay that

C: Yeah I got paid today

IQ. LIKELY TO PAY

IQ. DIFFERENTIAL

IQ. UNLIKELY TO PAY

Obtaining Customer Identity & confirming call monitoring

Business Manager

IQ. LIKELY TO PAY

IQ. DIFFERENTIAL

IQ. UNLIKELY TO PAY

C: Cash is low, I cannot pay by phone today

A: OK, can you pay via credit card

C: Yes, I can pay with credit card over the phone

A: OK, can you pay via credit card

C: Bye
Learning from transactions challenge

**GOAL:**
“Gaming” transaction data for hyper-personal experience in a given month

- **O** OFFER LEGEND
  - Loyalty Offer
  - Discount Coupon
  - Leaderboard Recognition
  - Outlier Warning
## Individualized Treatment

Individualized treatment **personas**, with corresponding priority to engage, retain and provide offers: **Frequency or value of basket size**. Optimize spend to reach these clients and keep them satisfied (without offering unnecessary discounts), and find others like them. Keep high-frequency buyers engaged with consistent messaging, and timely re-targeting to avoid annoying seasonal buyers, infrequent shoppers, and resellers. **Loyalty engagement or mobile app download**: Ideal clients for optimization into the high-value category with test promotions, social ads, mobile app downloads, and loyalty programs. **Probability of churn**: reduce churn by identifying less active buyers and offering targeted promotions; re-engage and retain using email/chat incentives for return visits with promotions; offer best promotions to customers with the highest potential **CLTV** score.

### Provided Data and Learn Insights:

<table>
<thead>
<tr>
<th>Provided Data</th>
<th>Learn Insights</th>
<th>Actions: Realtime, Personalized, Auto-Curated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Customer</td>
<td>Value</td>
<td>Frequency</td>
</tr>
<tr>
<td>Customer 1</td>
<td>$167.25</td>
<td>8</td>
</tr>
<tr>
<td>Customer 2</td>
<td>$123.70</td>
<td>6</td>
</tr>
<tr>
<td>Customer 3</td>
<td>$228.70</td>
<td>4</td>
</tr>
<tr>
<td>Customer 4</td>
<td>$671.72</td>
<td>4</td>
</tr>
<tr>
<td>Customer 5</td>
<td>$1,809.23</td>
<td>16</td>
</tr>
<tr>
<td>Customer 6</td>
<td>$2,052.06</td>
<td>40</td>
</tr>
</tbody>
</table>
Social signals challenge

TV

200+ networks
11,000+ show titles
2M+ show telecasts

Social

500M+ tweets/day
340M+ users/mo
50M+ expressions/mo

Online

4B+ searches/day
4.2B+ users/mo
50M+ behaviors/mo

P2P

20M+ reviews
150,000+ views/day
2M+ show telecasts

Goal: to deliver a single actionable NPX™ score
FatBrain Theme quantification for social

1. Create **Theme Probe vector** for every show based on LVE derived from peer to peer press (p2pp) e.g., Reddit, imdb, rotten tomatoes, etc.

2. Mine social streams using the **Theme Probe vector** to capture real time + historical expressions for all shows w/ hyperscale infrastructure.

3. Develop a demographic profile of viewer engagement and brand affinity from social expressions using LVE to learn and predict age, income, shows target demo watches.
Appendix

AI CI/CD, Best Practices, Performance Benchmarks