Best practices in risk model development
Introducing:

Keith Tanaka
Experian

Jeff Meli
Experian
Risk modeling landscape

- Market conditions
- Regulatory compliance
- More and better data
- Stronger tools
## Model applications

### Account acquisition
- **Front-end risk**
  - Prescreen risk
  - Fraud
  - Bankruptcy
  - First payment default

### Account management
- Behavioral
- CLI
- Roll rate
- Delinquency
- Transactional fraud

### Account loss
- Early stage collections
- Late stage collections
- Recovery

### Risk management
- Response
- Activation
- Conversion
- Purchase propensity

### Marketing efforts
- Profitability
- Usage
- Activation / reactivation
- Balance transfer
- Cross-sell

### Attrition/churn
- Pre-payment
- Retention
Risk modeling
Drivers of success

Strong project team + Preparation

Execution excellence

Best-in-class, On time, Seamless
Model development process

- Project inception
- Data prep
- Inference
- Segmentation
- Model development
- Documentation
- Implementation
Project inception

Assemble project team
Project inception
Key participants

• Project sponsor
• Line of business
• Risk managers
• Analysts
• IT
• Compliance
• Project manager
• Third party consultants

Comprehensive Collaborative Committed
Project inception

- Assemble project team
- Establish clear objectives
- Evaluate lending environment
- Define model development parameters
Project inception
Define model development parameters

• Sample parameters
  – Target population
  – Sample timeframe
  – Performance flag
  – Performance window
  – Exclusion

• Data sources and attribute candidates

• Model design
  – Inference
  – Segmentation
  – Attribute selection
  – Modeling technique
  – Model performance measurement

• Implementation considerations
## Project inception

**Example parameters by model application**

<table>
<thead>
<tr>
<th>Tool</th>
<th>Prescreen solicitation</th>
<th>Account acquisition</th>
<th>Credit line authorizations reissue</th>
<th>Attrition</th>
<th>Collections</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prescreen scoring</strong></td>
<td>Prescreen scoring</td>
<td>New applicant scoring</td>
<td>Behavior scoring</td>
<td>Behavior scoring</td>
<td>Collection scoring</td>
</tr>
<tr>
<td><strong>Data sources at observation</strong></td>
<td>Credit bureau demographic</td>
<td>Application credit bureau customer</td>
<td>Master file credit bureau</td>
<td>Master file credit bureau demographic</td>
<td>Master file credit bureau</td>
</tr>
<tr>
<td><strong>Performance window</strong></td>
<td>2–3 months</td>
<td>12–24 months</td>
<td>6–12 months</td>
<td>3–6 months</td>
<td>3–6 months</td>
</tr>
<tr>
<td><strong>Performance flags</strong></td>
<td>Response / non-response profitable / unprofitable</td>
<td>Good / bad</td>
<td>Good / bad</td>
<td>Stay / attrite</td>
<td>$ collected / no collection</td>
</tr>
</tbody>
</table>
Project inception
Model robustness across performance flags

60+ DPD Performance Flag

<table>
<thead>
<tr>
<th></th>
<th>Auto</th>
<th>Bankcard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gini Coefficient</td>
<td>77.0</td>
<td>76.8</td>
</tr>
</tbody>
</table>

90+ DPD Performance Flag

<table>
<thead>
<tr>
<th></th>
<th>Auto</th>
<th>Bankcard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gini Coefficient</td>
<td>78.4</td>
<td>78.7</td>
</tr>
</tbody>
</table>
Project inception
Model robustness across performance windows

12-Month Performance Window

<table>
<thead>
<tr>
<th></th>
<th>Auto</th>
<th>Bankcard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gini Coefficient</td>
<td>78.7</td>
<td>79.1</td>
</tr>
<tr>
<td>Model 12 Months</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 24 Months</td>
<td>77.7</td>
<td>76.8</td>
</tr>
</tbody>
</table>

24-Month Performance Window

<table>
<thead>
<tr>
<th></th>
<th>Auto</th>
<th>Bankcard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gini Coefficient</td>
<td>77.8</td>
<td>78.7</td>
</tr>
<tr>
<td>Model 12 Months</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 24 Months</td>
<td>75.8</td>
<td>75.5</td>
</tr>
</tbody>
</table>
Project inception
Define model development parameters

- Sample parameters
  - Target population
  - Sample timeframe
  - Performance flag
  - Performance window
  - Exclusion
- Data sources and attribute candidates

- Model design
  - Inference
  - Segmentation
  - Attribute selection
  - Modeling technique
  - Model performance measurement
- Implementation considerations
Project inception
Roadblocks and potential risks

DETOUR Progress impeded

- Lack of representation from key stakeholders
- Poor stakeholder communication and / or competing objectives
- Weak project management is counterproductive

CAUTION Potential risks

- Don’t move forward with design decisions without stakeholder consensus
- Poorly designed samples may limit model shelf life
- Consider implementation planning now

Best-in-class
On time
Seamless
Data preparation

- Data extraction
- Standard data integrity checks and multi-dimensional EDA
  - Portfolio reports
  - Population Stability Index (PSI)
  - Characteristic analysis
  - Attribute treatment
Data preparation
Population stability

Externally Sourced Attribute

% of Distribution

- PSI = 49.2
  - Reveals attribute stability concerns
  - Is marginal added predictive value justified?

Missing

Development Out of Time

40.7 11.4 59.3 88.6

0 10 20 30 40 50 60 70 80 90 100

% of Distribution

5/1/2017 Experian Public Vision 2017
## Data preparation

**Characteristic analysis:**
Revolving balance ($) – all trades*

<table>
<thead>
<tr>
<th>Bad</th>
<th>Good</th>
</tr>
</thead>
<tbody>
<tr>
<td>5,001+</td>
<td>0</td>
</tr>
<tr>
<td>2,501-5,000</td>
<td>1 - 500</td>
</tr>
<tr>
<td>501-2,500</td>
<td>501-2,500</td>
</tr>
<tr>
<td>1-500</td>
<td>2,501-5,000</td>
</tr>
<tr>
<td>0</td>
<td>5,001+</td>
</tr>
</tbody>
</table>

**Good / bad index**

* Originations model
## Data preparation

### Characteristic analysis: Revolving balance ($) – one account*

<table>
<thead>
<tr>
<th>Revolving balance ($)</th>
<th>Bad</th>
<th>Good</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,000 - 2,000</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>201 - 1,000</td>
<td>0</td>
<td>1 - 200</td>
</tr>
<tr>
<td>1,001 - 2,000</td>
<td>0</td>
<td>201 - 1,000</td>
</tr>
<tr>
<td>2,001 +</td>
<td>0</td>
<td>1,001 - 2,000</td>
</tr>
</tbody>
</table>

### Good / bad index

| 1.50 | 1.00 | 1.50 |

* Behavioral line management model for a store credit card
Data preparation

- Data extraction
- Standard data integrity checks and multi-dimensional EDA
  - Portfolio reports
  - Population Stability Index (PSI)
  - Characteristic analysis
  - Attribute treatment
- Waterfall statistics
- Sample selection
Data preparation
Roadblocks and potential risks

Progress impeded

- Lack of specifications for data extraction
- Insufficient attention given to EDA

Potential risks

- Lack of sufficient attribute vetting
- Delaying extraction and preparation of out-of-time validation sample

Best-in-class  On time  Seamless
Inference

- Assess the need for inference
- Determine appropriate technique
- Evaluate impact on booked population performance
Inference
The need for inference

To correct for sample bias

Example:

Without inference on declined applications the scorecard would only be valid for a small proportion of the application universe.
Inference techniques

Re-weighting
• Based on a score interval, “weight-up” accounts with known performance to the through the door distribution

Reclassification
• High risk applications, typically based on a score or high risk profile, are assigned a “bad” performance

Parcelling
• Rejected applications are assigned good and bad performance based on a risk score

Bureau inference
• Using bureau data assign good and bad performance based on performance of similar trades with other creditors
Inference
Pre-diction vs. post-diction

Pre-diction and post-diction model development and assessment:
- Booked trade suppressed from post-diction attributes
- Models built on booked accounts only
- Non-booked performance assigned based on performance of a similar trade on the bureau
- Model applied to non-booked applicants

<table>
<thead>
<tr>
<th>Observation Point (Pre-diction Attributes)</th>
<th>Outcome Window</th>
<th>Outcome Point (Post-diction Attributes)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>5/1/2017</strong></td>
<td></td>
<td><strong>Experian Public Vision 2017</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Known Performance</th>
<th>Inferred Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gini Coefficient</td>
<td></td>
</tr>
<tr>
<td>63.9</td>
<td>83.5</td>
</tr>
<tr>
<td>45.4</td>
<td>70.2</td>
</tr>
</tbody>
</table>

- **Known Performance**: 63.9 (Prediction) vs. 83.5 (Postdiction)
- **Inferred Performance**: 45.4 (Prediction) vs. 70.2 (Postdiction)
Inference

Inferred applicant weighting and impact on booked account performance

<table>
<thead>
<tr>
<th>Model</th>
<th>Gini Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Existing benchmark</td>
<td>51.93</td>
</tr>
<tr>
<td>BG model (100% known)</td>
<td>55.76</td>
</tr>
<tr>
<td>BGI model 60/40 known/inf</td>
<td>52.64</td>
</tr>
<tr>
<td>BGI model (85/15 known/inf)</td>
<td>55.61</td>
</tr>
</tbody>
</table>
Inference
Roadblocks and potential risks

Progress impeded

- Gaps in sample design and data preparation
- Overcoming the inference design

Potential risks

- Too little inference may increase risk exposure
- Too much inference compromises performance of traditional book of business

Best-in-class

On time

Seamless
Segmentation

- Determine candidate schemes
- Build master model
- Build out niche models for each candidate scheme
- Compare master vs. niche model performance
- Assess trade-offs of candidate schemes
Segmentation
Segment evaluation

- Gini Coefficient

Segment 1: Niche - 60, Master - 70
Segment 2: Niche - 20, Master - 50
Segment 3: Niche - 40, Master - 40
Overall: Niche - 30, Master - 70
Segmentation
Roadblocks and potential risks

**Progress impeded**
- Lack of attribute vetting during project initiation
- Potential segment schemes omitted from consideration

**Potential risks**
- Segmentation overkill
- Failure to review simple statistics to help eliminate non-viable segments

Best-in-class  On time  Seamless
Model development

- Finalize attribute selection and modeling approach
- Develop preliminary models
- Conduct stakeholder review, and test alternate attributes
- Refine and finalize models
- For each iteration
  - Test model vs. benchmarks
  - Conduct in- and out-of-time validations
Model development
Roadblocks and potential risks

**Progress impeded**
- Lack of stakeholder sign-off on all prior phases
- Insufficient stratification of estimation and in-sample validation samples during data preparation

**Potential risks**
- Insufficient attribute vetting
- ‘Science’ only gets you so far- expert intuition is key
- Delaying out-of-time validations
Documentation

- Primary focus on compilation rather than construction
- Compliance and governance
- Document key decisions alongside the rationale
Documentation
Roadblocks and potential risks

**Progress impeded**
- Decisions made that were not recorded
- Insufficient documentation from prior phases

**Potential risks**
- Do not delay!
- Avoid large gaps – end-to-end documents
- Ensure document is tailored to complete audience

---

Best-in-class
On time
Seamless
Implementation

- Coding / testing / auditing
- Strategy design
- Compliance and governance
- Monitoring
Implementation
Roadblocks and potential risks

Progress impeded
- Unrealistic implementation timeframe
- Key participants not consulted
- System limitations not explored

Potential risks
- Use scorecard for designed purpose
- Monitor regularly
- Do not neglect model maintenance

Best-in-class
On time
Seamless
Rules of the road

- Chose the right partners for the journey
- Use the right fuel
- Plan your route and your milestones
- Look ahead and stay focused
- Maintain a travel log
- Be prepared for the unexpected
Risk modeling
Drivers of success

- Strong project team
- Preparation

Execution excellence

- Best-in-class
- On time
- Seamless
Questions and answers

Experian contact:

Jeff Meli
Jeff.Meli@experian.com
Share your thoughts about Vision 2017!

Please take the time now to give us your feedback about this session. You can complete the survey at the kiosk outside.

How would you rate both the Speaker and Content?