Stochastic Gradient Boosting Approach to Daily Attrition Scoring Based on High-dimensional RFM Features

Dr. Gerald Fahner
Senior Director Analytic Science, FICO

Agenda

- Ultra-dynamic Attrition Scoring
- Case Study—Credit Card Attrition
- Category Attrition

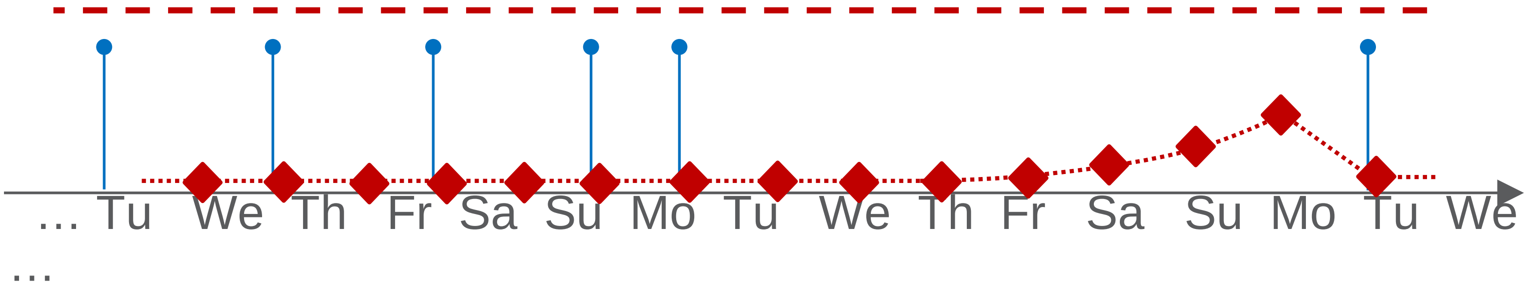
Ultra-Dynamic (Daily) Attrition Scoring Approach

Customer uses card

◆ Daily attrition risk score

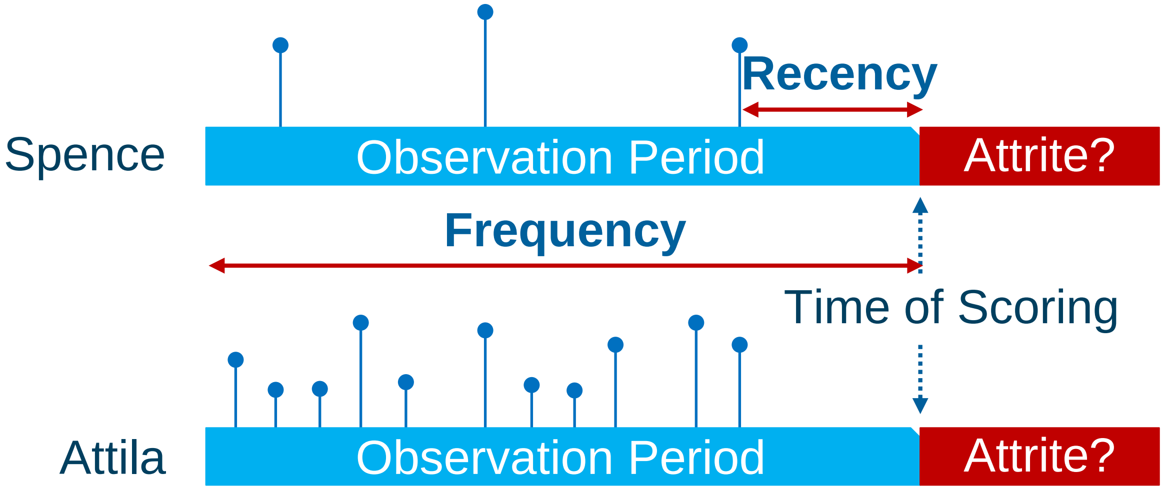
- Prolonged inactivity signals higher risk—drives up attrition risk score

→ Re-engage customer when attrition risk exceeds some threshold



Transaction Dynamics Hold Key Information

- Given information at time of scoring, who is more likely to attrite?
  - Which measures are most informative?
- How to combine Recency and Frequency into predicting attrition risk?



Recency:
Days since last card use

Frequency:
Fraction of days card used during obs. period

How Machine Learning Complements Domain Expertise

Domain Expertise	Machine Learning
Good at intuiting key predictors 1	Lacks intuition
Doesn’t scale to many variables Poor at combining multiple predictors Poor at quantifying uncertainty	2 Excels at combining many features into accurate probabilistic predictions
Need story behind the numbers 4	3 Diagnose and visualize models to gain insight into effects

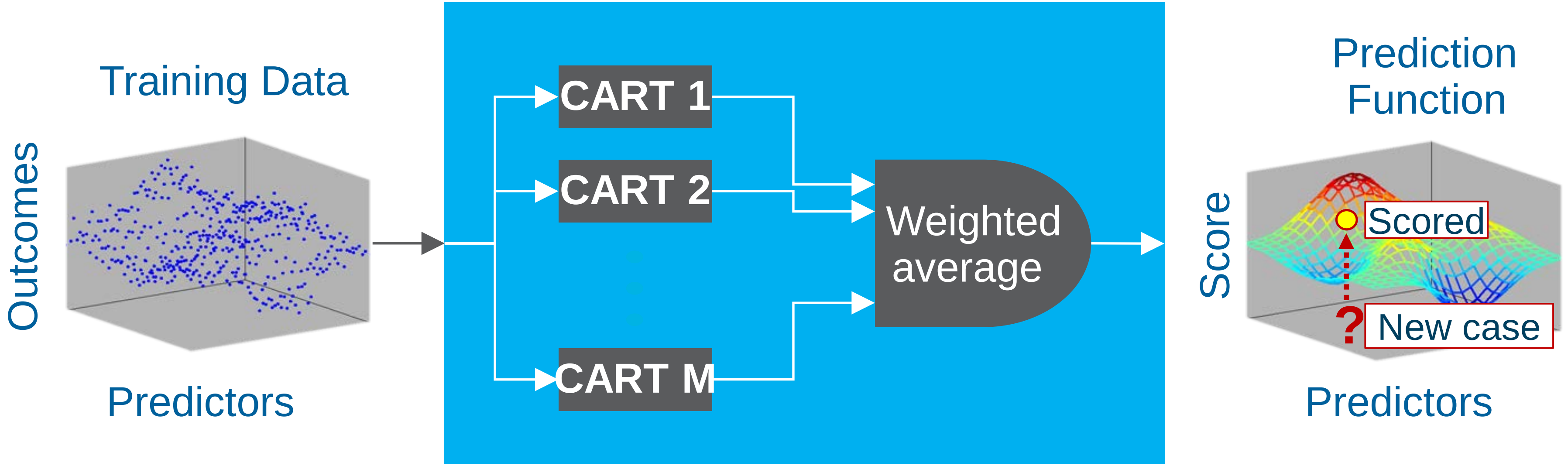
Recommended path

Key Elements of Approach

Featurization of transaction events	• Based on Recencies, Frequencies, Monetary values • High-dimensional feature space of complex events
Machine learning / classification tools	• Stochastic Gradient Boosting • Partial dependence visualization
Performance evaluation	• Lift related to portfolio profit gain • Out-of-sample / Out-of-time evaluation

Stochastic Gradient Boosting[1]

Combines predictions from 100’s or 1000’s of shallow CARTs



Inexplicable model by direct inspection

Agenda

- Ultra-dynamic Attrition Scoring
- Case Study—Credit Card Attrition
  - Machine Learning for Higher Profit
- Category Attrition

Credit Card Case Study

Data and Project Design

- ~5 million accounts. More than 1 billion transactions over 3 years
- Transaction information: **Date**, **Merchant Code**, **Amount**, **Authorized Flag**

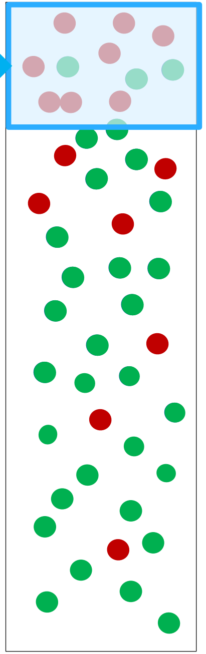
Attrition Performance Definition	Scoring Exclusions
Binary indicator of card activity during Performance period	Inactive

Statistical Measures of Model Performance

High Scores

Target top $\alpha\%$

with retention offer



Low Scores

- ● Would-be attriters
- ● Non-attriters

Lift at $\alpha\%$ operating point:

$$\lambda = \frac{\text{Fraction of Attriters Among Targeted}}{\text{Base Attrition Rate}}$$

$$= \frac{\text{Precision}}{\text{Base Attrition Rate}}$$

FICO

Profit from a Retention Campaign

Actual Behavior of Targeted Customer	Profit Contribution per Customer	Fraction of Targeted Customers with this Behavior
Would-be attriter we persuade to stay	(CLV Gain – Contact Cost – Incentive Cost)	Precision * Persuasion Rate
Unpersuadable attriter	(No CLV Gain – Contact Cost)	Precision * (1–Persuasion Rate)
Non-attriter, erroneously targeted	(No CLV Gain – Contact Cost – Incentive Cost)	1–Precision

Profit Gain From Attrition Model Improvement[2]

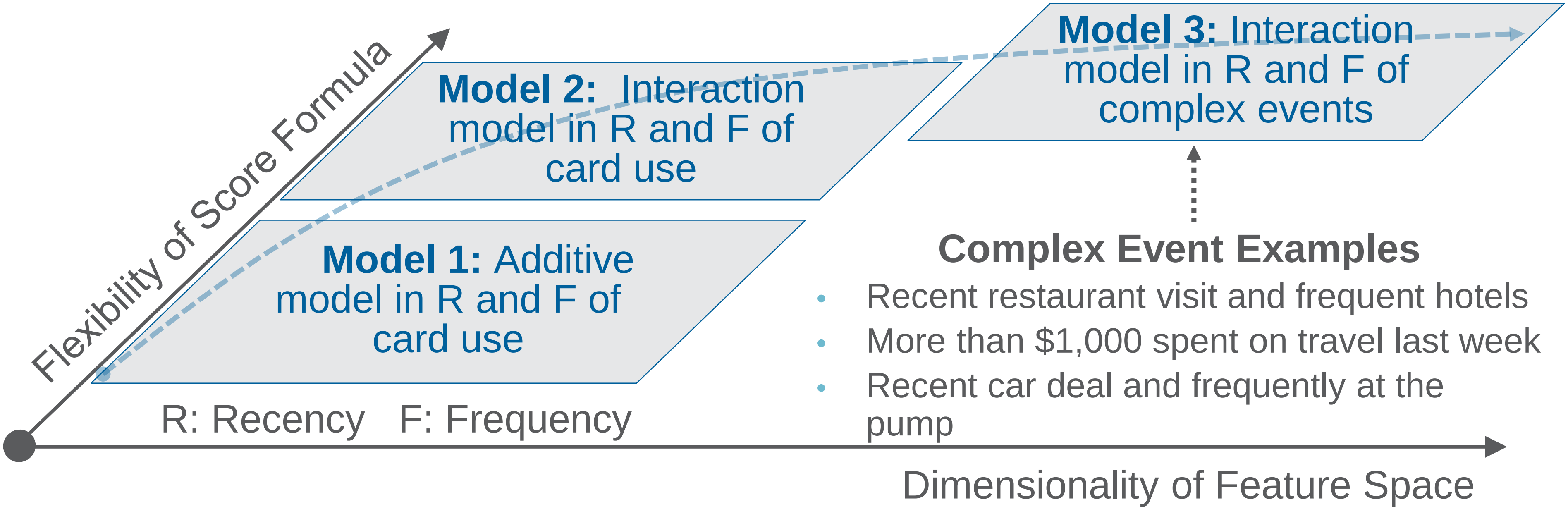
$\text{Gain} = (\lambda_B - \lambda_A) N \alpha \beta_0 (\gamma\, CLV + \delta(1-\gamma))$ is Portfolio Profit Gain from improving model B over model A, where:

λ_A	Lift from model A	Will benchmark alternative models
λ_B	Lift from model B	
α	Targeting Fraction	5%
β_0	Base Attrition Rate	8%
N	Portfolio Size	5 million
CLV	Customer Lifetime Value	$1,000
δ	Incentive Cost	$100
γ	Persuasion Rate	20%

Portfolio-specific assumptions

Benchmarking Predictive Models of Increasing Complexity

- How much can we gain by making models more complex?
- Are complex models robust over time?



Interaction Detection Experiment

→ Should Capture (Recency X Frequency) Interactions

- **Predictors:** Recency and Frequency of card use
  - **Model 1:** Additive, nonlinear in R and F
  - **Model 2:** Captures interaction between R and F

Out-of-sample / Out-of-time validation

$$\lambda_1 = 6.03$$
$$\lambda_2 = 6.54$$

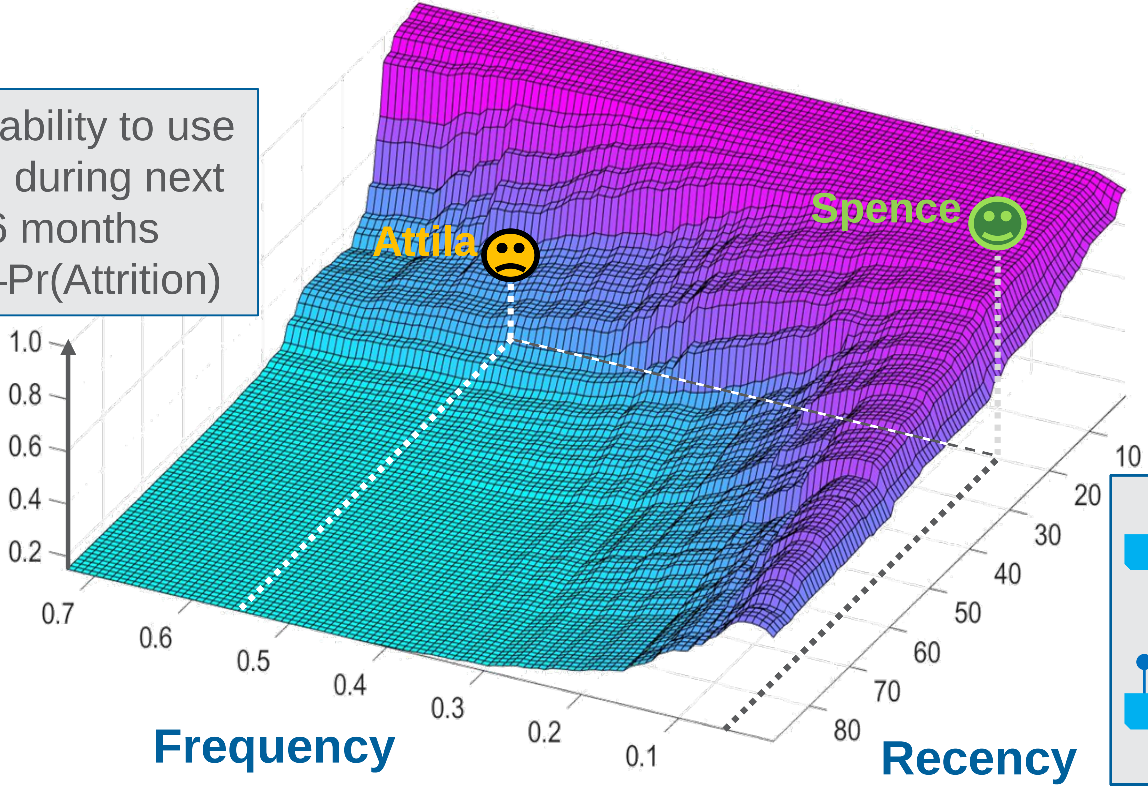
$$\Rightarrow \quad \text{Gain} = \$2.86\ \text{MM} \;\; \text{s.t. portfolio assumptions}$$

- Interaction effect in agreement with research by Fader and Hardie[3]

Interaction Visualization Tells Story

Two-dimensional Partial Dependence Function[4]

Probability to use card during next 6 months = 1–Pr(Attrition)



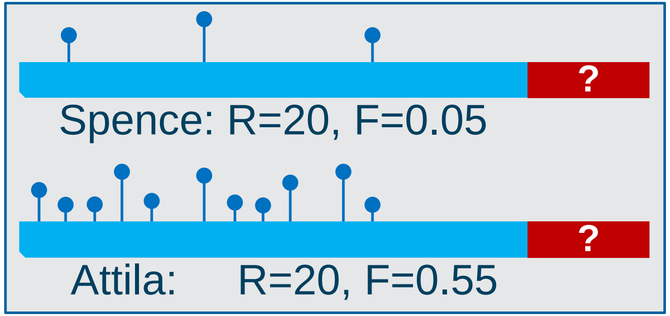
Frequency

Fraction of days card used

Recency

Days since last card use

Attila is at higher risk of attrition because his card use has lapsed for an unusually long time interval



Featurization Experiment

→ Should Capture Complex Events in Your Models

- Define R and F features for complex events
- **Model 3**: Candidate predictors include:

Card use events
+ Hundreds of merchant category events
+ Monetary events defined by spending bands
+ No-authorization events

Out-of-sample / Out-of-time validation

$\lambda_3 = 7.52$

Recall:

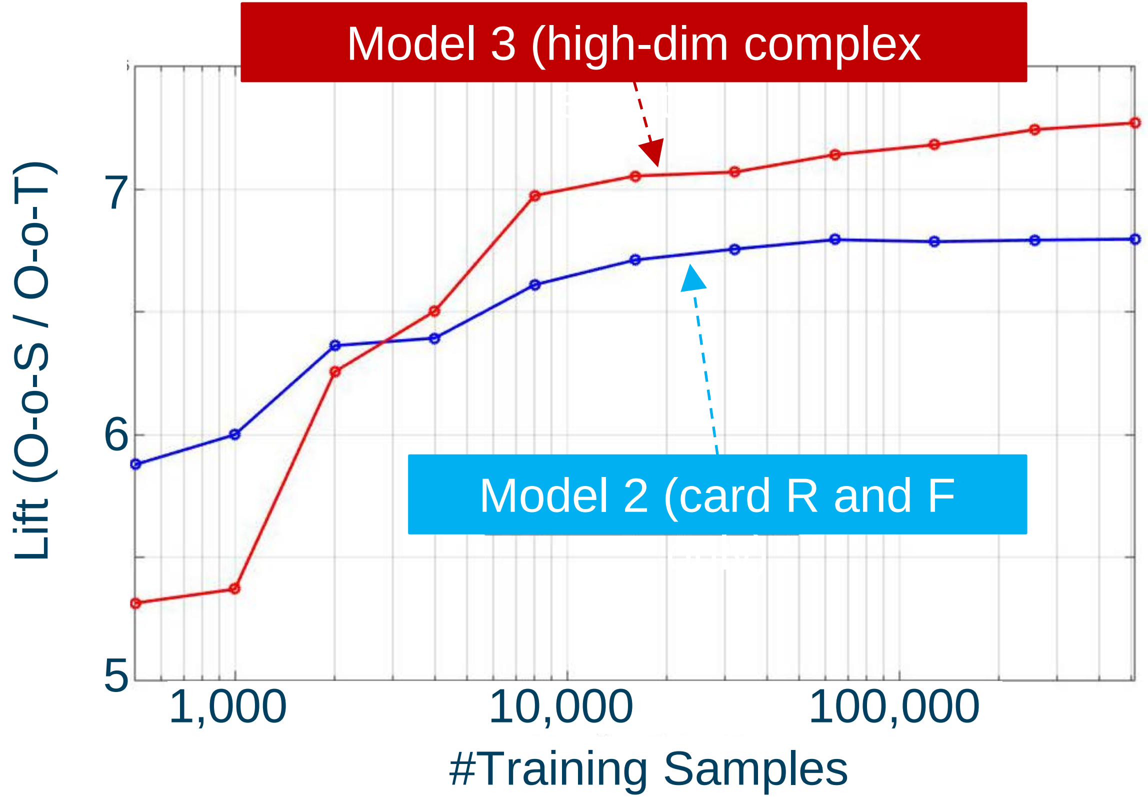
$\lambda_1 = 6.03$

$\lambda_2 = 6.54$

$\Rightarrow \quad \text{Gain over Model 1 (simple, additive)} = \$8.34 \text{ MM} \;\; \text{s.t. portfolio assumptions}$

Learning Curves Experiment

→ Should Exploit Larger Samples to Develop More Complex Models



Agenda

- Ultra-dynamic Attrition Scoring
- Case Study—Credit Card Attrition
- Category Attrition
  - Detecting Subtle Forms of Attrition

Merchant Category (MC) Attrition

- Hundreds of credit card MC's
- Performance definition for a specific MC:
  - Stop buying from *this* MC–while continuing card use for other MC's
- May signal competitive influence or early belt-tightening—before total attrition occurs. Quick detection informs rapid intervention

Card-level model →	Overall customer status	✓
Grocery model →	Grocery status	✓
Travel model →	Travel status	✓
Gas station model →	Gas station status	!

Possible interventions:

Offer incentives at service stations, or start customer dialogue

Summary

- Daily attrition scoring quickly detects emergent attrition—signaled by unusually long time lapse since last transaction
- With large transaction volumes, more complex models are more profitable
- Machine learning helps with insight, automation, scale

References

[1] *Greedy Function Approximation: A Gradient Boosting Machine*, by Jerome Friedman, The Annals of Statistics, 29(5), 2001, 1189-1232.

[2] *Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models*, by Scott Neslin et al., Journal of Marketing Research, 43(2), 2006, 204-211.

[3] *RFM and CLV: Using Iso-Value Curves for Customer Base Analysis*, by Peter Fader, Bruce Hardie, and Ka Lok Lee, Journal of Marketing Research, 42(4), 2005, 415-430.

[4] *Predictive learning via rule ensembles*, by Jerome Friedman et al., The Annals of Applied Statistics, 2(3), 2008, 916-954.

FICO Decisions

Thank You

Dr. Gerald Fahner
++1 512 5323621
geraldfahner@fico.com