Developing Transparent Credit Risk Scorecards More Effectively

An Explainable Artificial Intelligence Approach

Dr. Gerald Fahner
Senior Principal Scientist, FICO

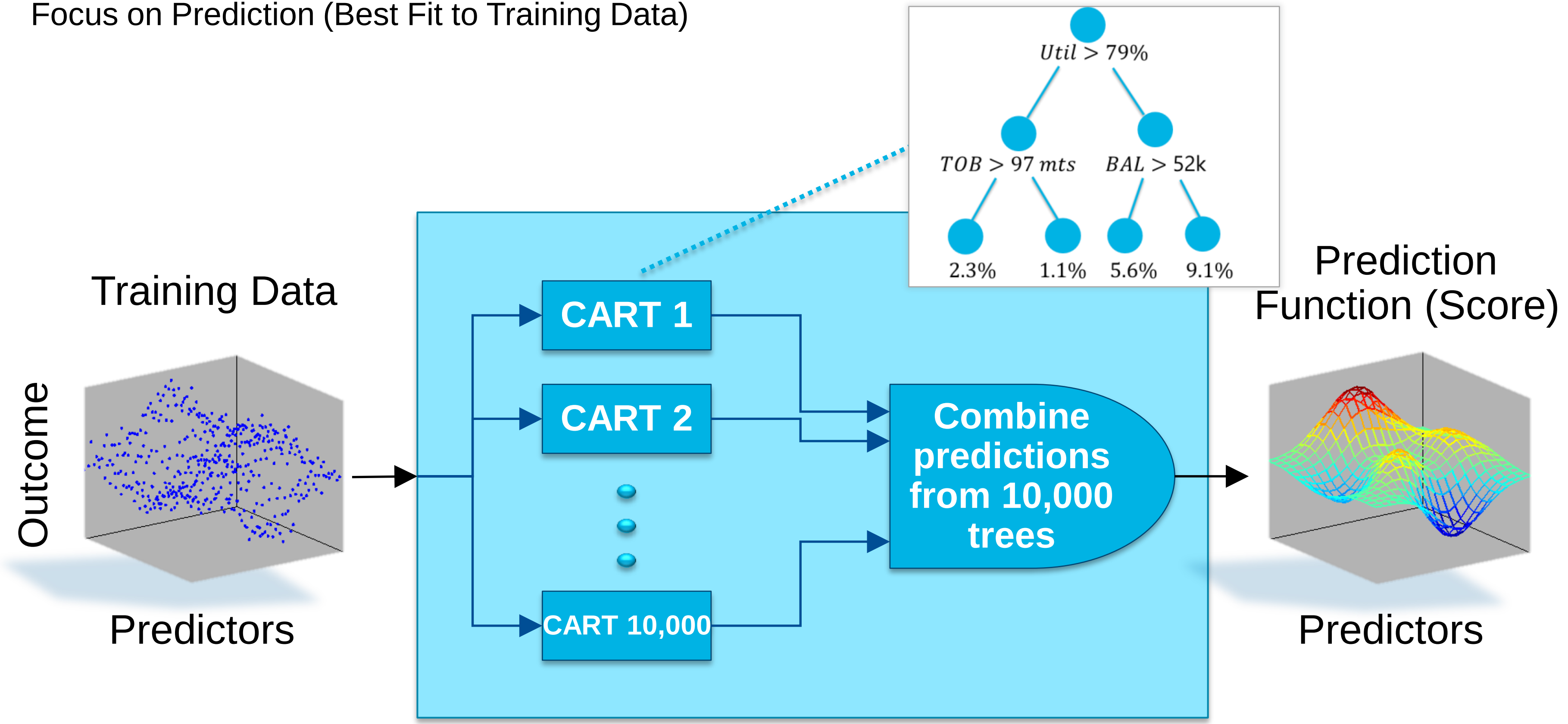
Use of Machine Learning (ML) in FICO® Score Development

- FICO® Score developments use ML for over 25 years
  - Optimal binning
  - Characteristic selection
  - Optimizing score weights
  - Interaction detection and multi-scorecard segmentation
  - Benchmarking
- Data-driven ML is balanced with domain expertise
  - To ensure transparency, palatability, fairness
  - To stand up to regulatory and consumer scrutiny
  - To mitigate data biases

Anatomy of a Typical ML Model

CART (Classification and Regression Tree)

Focus on Prediction (Best Fit to Training Data)



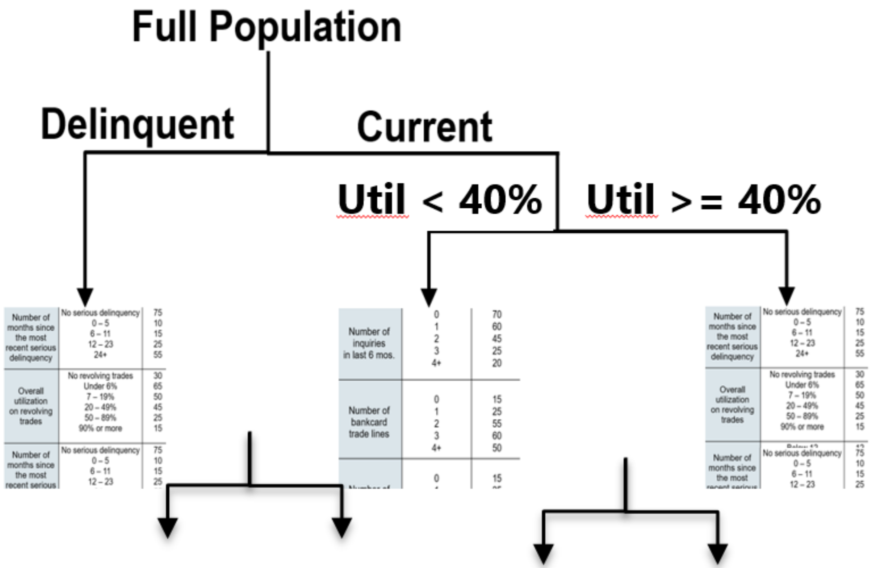
Random Forests; Stochastic Gradient Boosting

Anatomy of FICO® Score Model*

Focus on Prediction *and* Explanation—Balance Best Fit to Data with Domain Expertise

Category	Characteristics	Attributes	Points
Payment History	Number of months since the most recent serious delinquency	No serious delinquency 0 – 5 6 – 11 12 – 23 24+	75 10 15 25 55
Outstanding Debt	Overall utilization on revolving trades	No revolving trades Under 6% 7 – 19% 20 – 49% 50 – 89% 90% or more	30 65 50 45 25 15
Credit History Length	Number of months in file	Below 12 12 – 23 24 – 47 48 or more	12 35 60 75
Pursuit of New Credit	Number of inquiries in the last 6 months	0 1 2 3 4+	70 60 45 25 20
Credit Mix	Number of bankcard trade lines	0 1 2 3 4+	15 25 55 60 50

- Characteristic selection, Points patterns are subject to palatability constraints



- Easy-to-explain multi-scorecard system captures nonlinearities, interactions; increases score power
- FICO® Score 9 uses 13 scorecards dedicated to distinctive population segments

*Graphics for illustrative purposes only

Benchmarking ML Scores Against FICO® Score 9

Purely Data-driven ML Yields Modest Predictive Lift; Significantly More Streamlined Model Build

< 2%

Relative improvement in KS on in-time holdout sample*

- Stochastic Gradient Boosting
- Neural Networks

40 vs 800

Resource hours required to train ML model vs multi-scorecard development

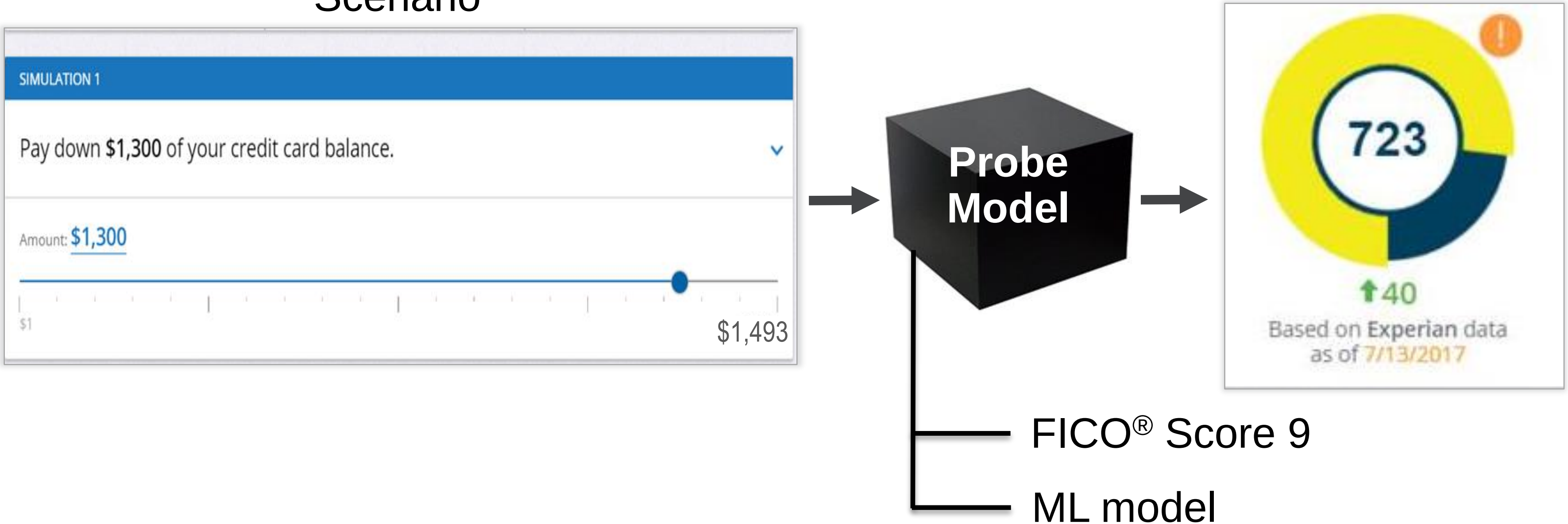
*ML models were trained and evaluated on same data used to develop and to evaluate FICO® Score 9 (= nationally representative sample of 10M credit files)

Assessing Palatability of Models Through Score Simulations

- Pose payment behavior scenarios, such as:
  - How does paying off ~90% of total credit card debt impact my score?

Scenario

Simulated Score



Simulations Reveal Lack of Palatability of ML Models

Probe Model	Result of Simulation*
FICO® Score 9	**0%** of consumer records experienced a decrease in score as a result of this positive credit behavior (reducing debt)
Stochastic Gradient Boosting	**9.2%** of consumer records experienced a decrease in score

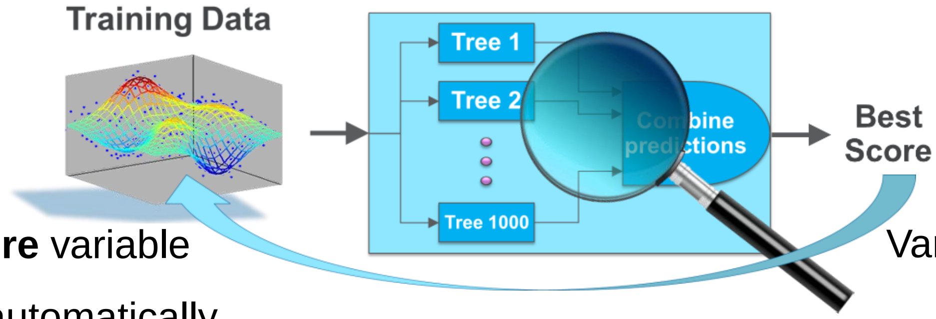
*Based on representative national sample of millions of FICO scorable credit files
Held everything else fixed in simulation (credit history, non-revolving balances, etc.)

- Positive credit action leads to ML score decrease 9.2% of the time
- Consumers and lenders would be confounded by such a deviation from expectations

Explainable AI/ML Approach to Credit Score Development*

*See full paper: *"Developing Transparent Credit Risk Scorecards More Effectively: An Explainable Artificial Intelligence Approach"*

1. Find best ML model
2. Diagnose model
3. Augment data with **Best Score** variable
4. Approximate Best Score by automatically grown multi-scorecard system

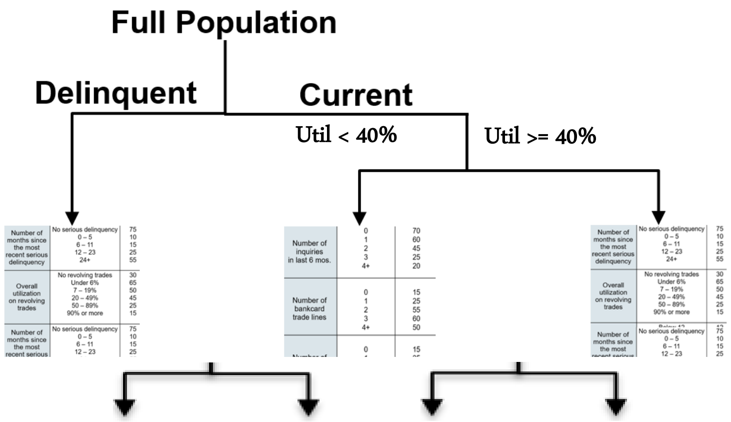


Diagnostics

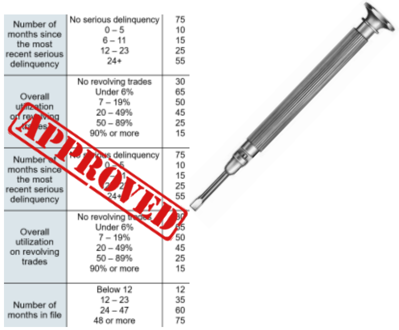
Variable importance

Interactions

Partial Dependence Plots



5. Add domain expertise to segment scorecards



6. Deploy

Performance Comparison*

Model (Score Development Technology)	% improvement in KS over FICO® Score 9
FICO® Score 9	N/A
Stochastic Gradient Boosting	1.7%
Neural Network	0.5%
Explainable AI/ML Approach	0.3%

*Performance on bankcard accounts over 24 months (Bad = 90+ days past due)

Conclusions

AI/ML offers substantial efficiently gains for credit risk score developers,
but lack of palatability can render purely data-driven models unfit for deployment.

To ensure transparency, palatability, and fairness of scores,
model development must balance data-driven learning with domain expertise.

Explainable AI/ML approaches are required to strike this balance.

Thank You!

Gerald Fahner
C 512 698 0609
geraldfahner@fico.com