A Survey of Machine Learning in Credit Risk

Joseph L. Breeden, PhD

CEO, Prescient Models LLC
breeden@prescientmodels.com

Prescient Models

What is Machine Learning?

Following Harrell [2019] a possible distinction between statistical modeling and machine learning is:

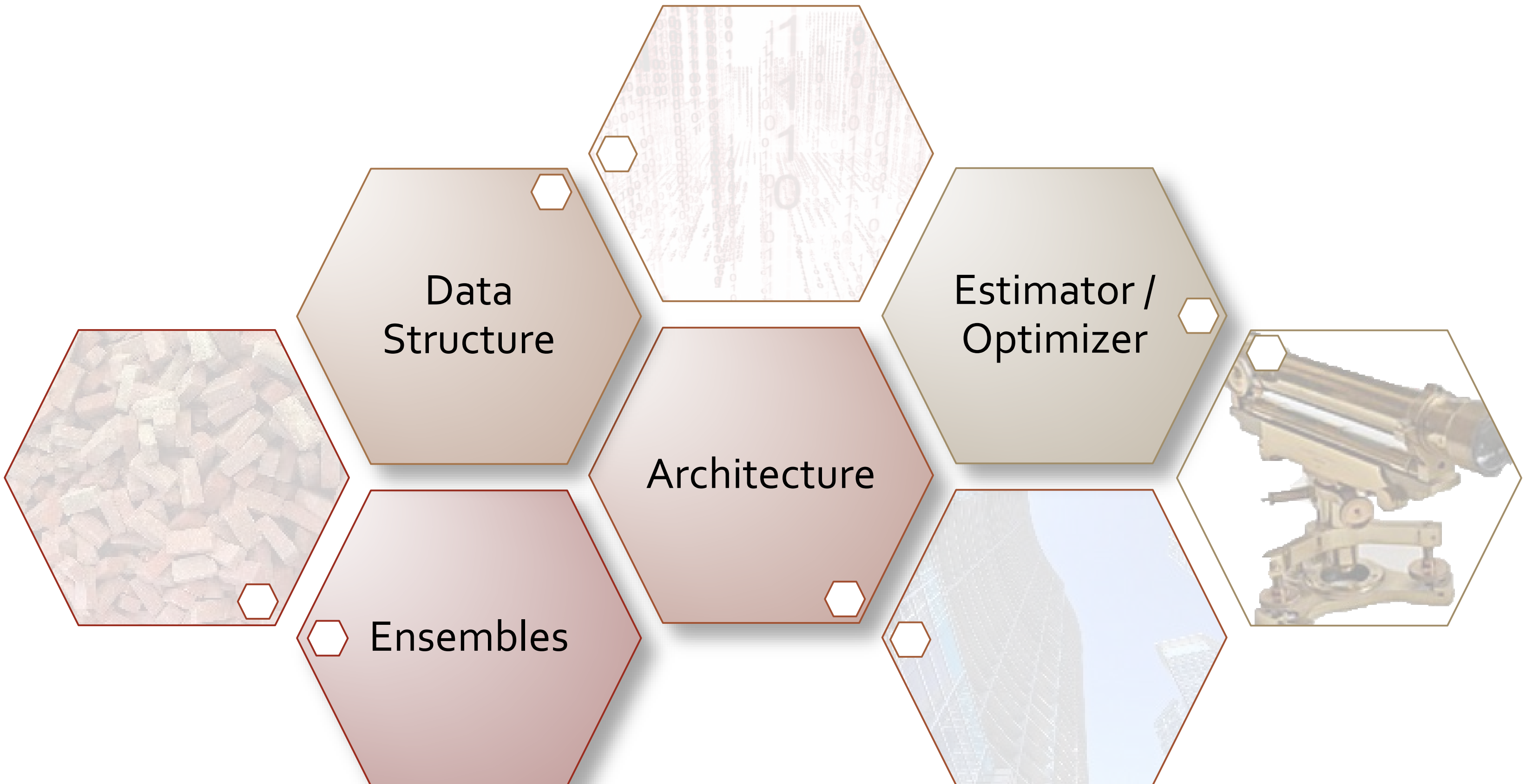
- **Uncertainty**: Statistical models explicitly take uncertainty into account by specifying a probabilistic model for the data.
- **Structural**: Statistical models typically start by assuming additivity of predictor effects when specifying the model.
- **Empirical**: Machine learning is more empirical including allowance for high-order interactions that are not pre-specified, whereas statistical models have identified parameters of special interest.

Further Distinctions

The above items carry other implications.

- **Confidence Intervals**: Search-based methods such as Monte Carlo simulation, genetic algorithms, and various forms of gradient descent usually do not provide confidence intervals for the parameters, and correspondingly are usually considered as machine learning.
- **Ensemble methods** where multiple models are combined are generally considered to be machine learning, even when the constituent models are statistical.
- **Feature Selection**: Traditional statistical methods rely on analyst selection of input features and interaction terms whereas machine learning methods emphasize algorithmic selection of features, discovery of interaction terms, and even creation of features from raw data.

Creating a Model – Mix-n-Match



Data Structures

Data Types
Account Outcomes
Account Time Series
Vintage Time Series
Segment Time Series

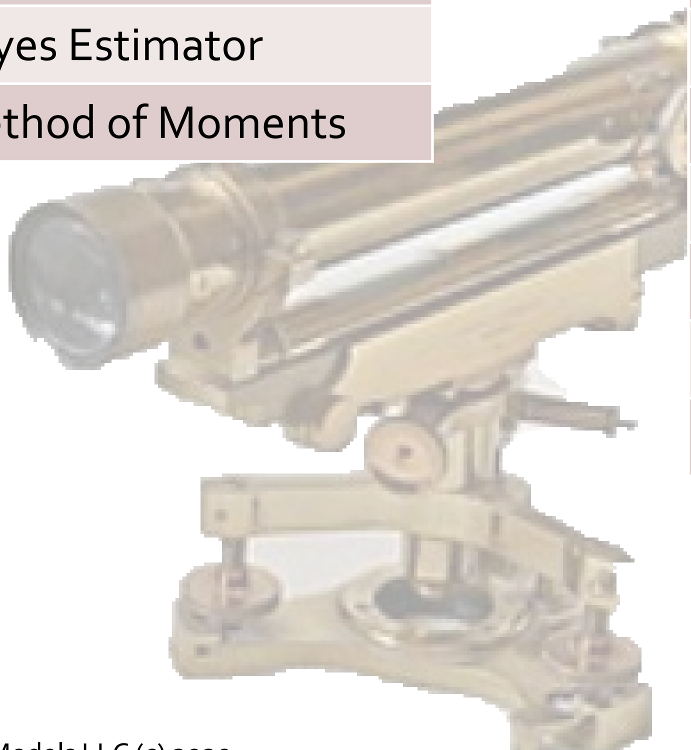
Target Variables
Loss Balance
PD, EAD, LGD [and PA]
Prepayment
Preprovision Net Revenue (PPNR)
Asset Values
Deposit Balance
Time Deposit Renewal Rates

Architectures

Architectures
Additive Effects
Additive Fixed Effects
Convolutional Network
Clustering
Feed-forward Network
Fuzzy Rules and Rough Sets
Nearest Neighbors
Recurrent Neural Network
Segmentation
State Transition
Trees

Estimators / Optimizers

Estimators
Least Squares
Maximum Likelihood
Partial Likelihood
Bayes Estimator
Method of Moments



Optimizers
Gradient Descent
Simulated Annealing
Back Propagation
Reinforcement Learning
Genetic Algorithms
Markov Chain Monte Carlo
Kalman Filter
Linear Programming
Quadratic Programming

Ensembles

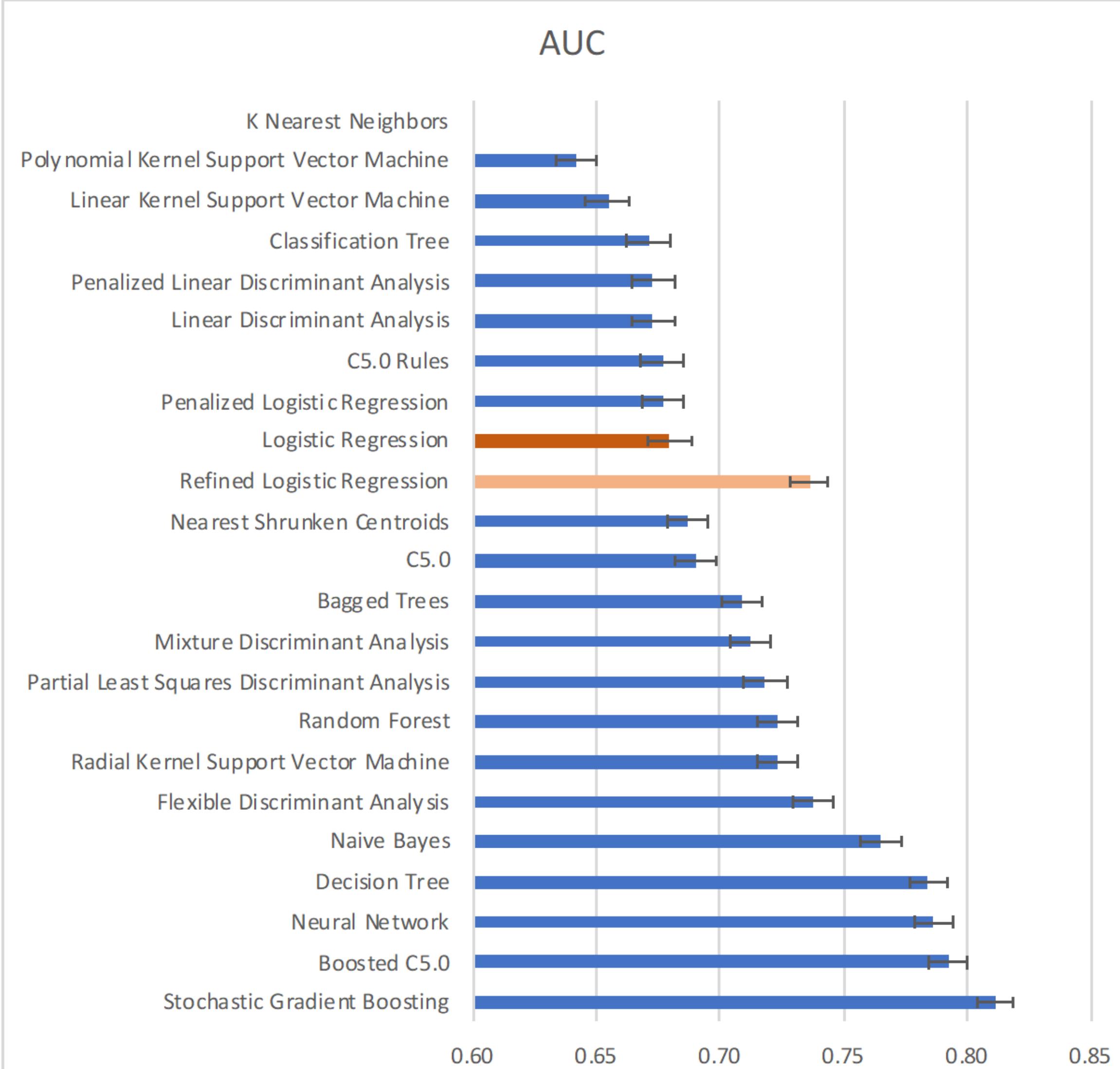
- Homogeneous: Bagging and Boosting
- Hybrid: Specially designed combinations
- Heterogeneous

Binary	Categorical	Continuous
Plurality Voting	Plurality Voting	Average
Sum Rule	Majority Voting	Median
Product Rule	Sum Rule	Confidence Weighted
Stacking	Product Rule	Stacking
	Amendment Vote	
	Runoff Vote	
	Condorcet Count	
	Pandemonium	
	Borda Count	
	Single Transferable Vote	
	Stacking	

What makes ML work?

This is an example from building a credit card attrition score.

Modeling work was done by Casey Foltz of Oregon Community CU.

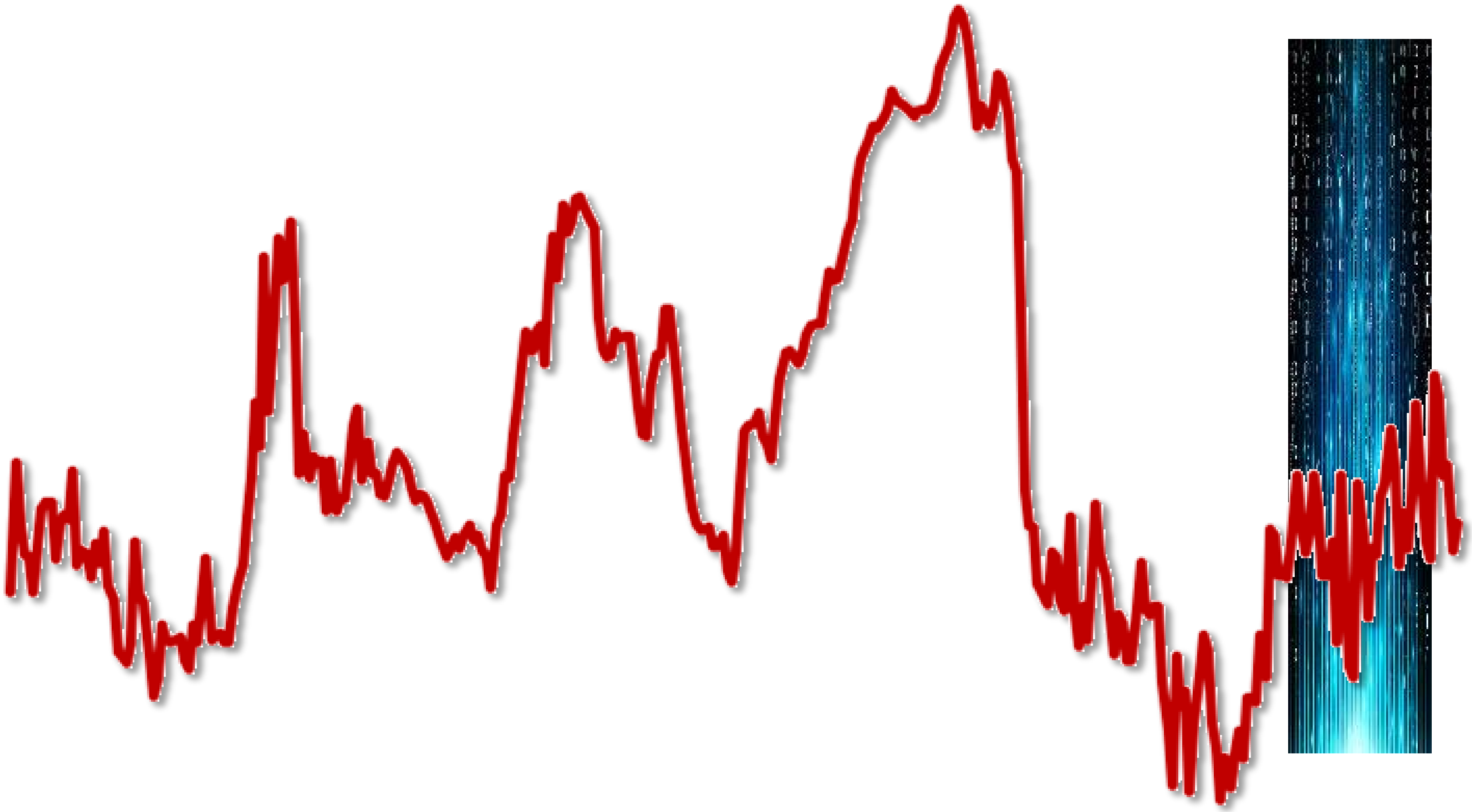


What could go wrong?

Big Data is enabling AI and Data Mining advances

But we only have Big Data for a short slice of history

Some critical information is available only in long, thin data



Adverse Action Notices

- How do we explain credit rejection from nonlinear models?

- Several approximate solutions exist (Explainable AI, XAI) and improvements are being made:

  - LIME
  - Shapley values
  - Newer methods

Unintended Bias

- Creating models with unintended bias against protected classes is an important, well-known problem.

- However, if we can't capture data on protected class status, how can we test the models to prove they are unbiased?

Transient Structure

- The strength of highly nonlinear models is in their ability to find subtle behaviors.
- The risk is that these behavioral patterns shift as the environment shifts, faster than we can retrain the models.

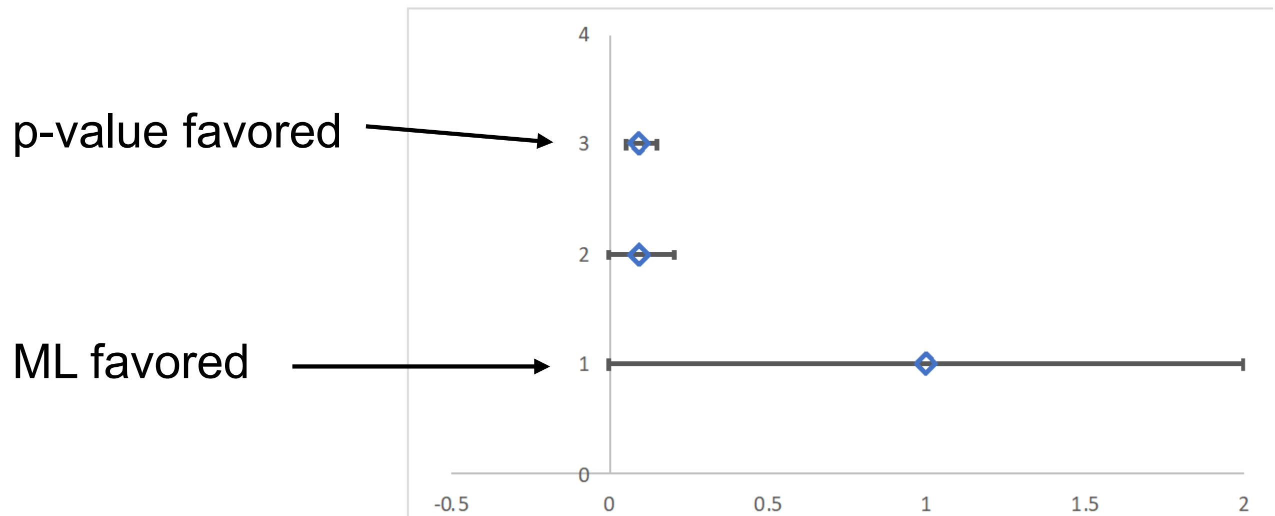
Overfitting the Overfitting Tests

- Most AI/ML algorithms have a training set and a test set. However, the more often we look at the test set, the more it becomes in-sample.
- The Holm-Bonferroni method corrects for this in significance tests, but how to we correct ML methods?

p-Value Arbitrage

Linear model validators almost always reject any model with insignificant p-values. However, the American Statistical Society has stated that they should not be used this way.

ML models do not have p-values to test. How much of their improved performance is model validation arbitrage?

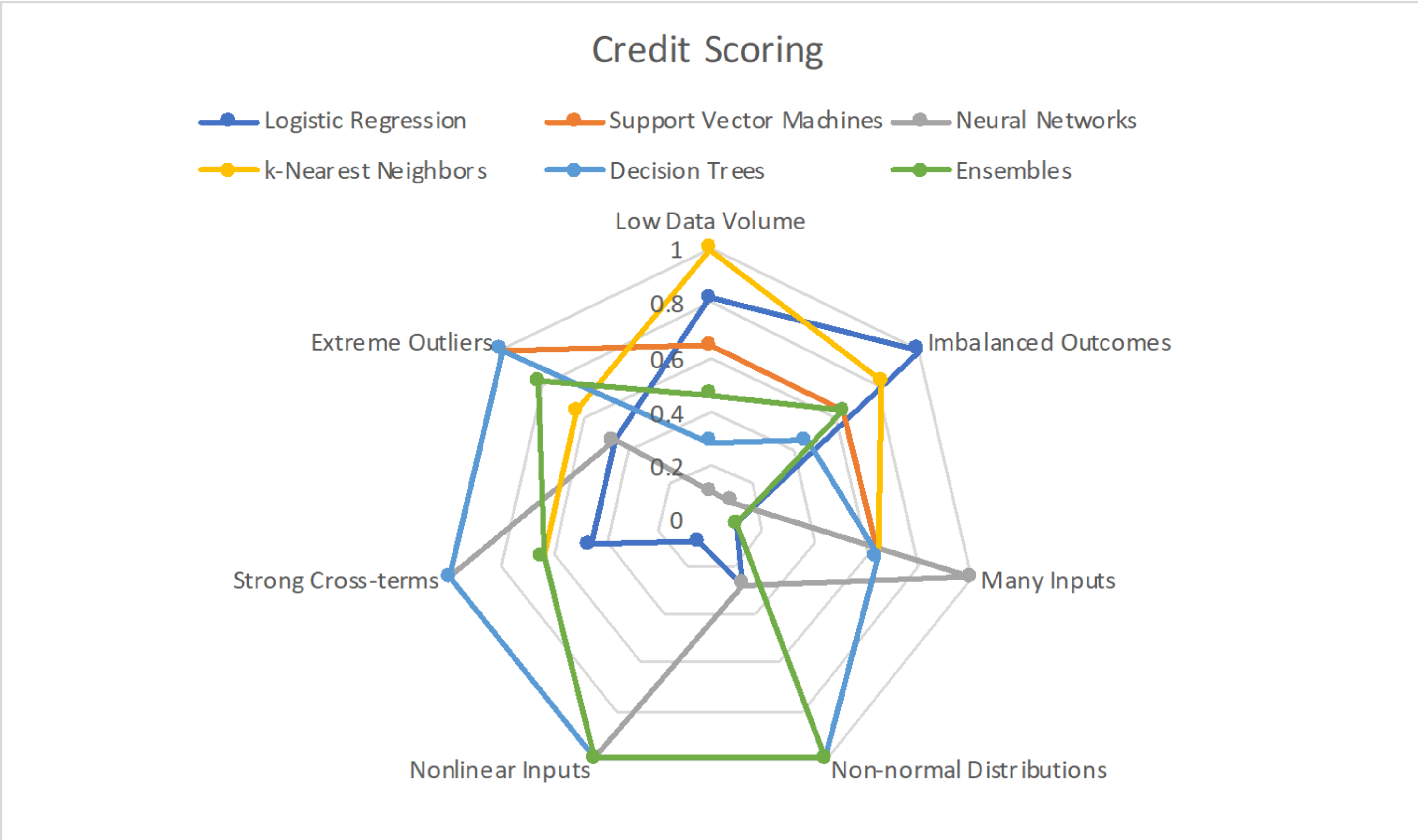


Non-uniform Uncertainty

- When scores are used in underwriting, no confidence intervals are attached.
- Linear models may have roughly uniform uncertainties through most of the application range.
- Nonlinear models that find pockets of predictability could have very non-uniform uncertainties for different accounts at the same score level.

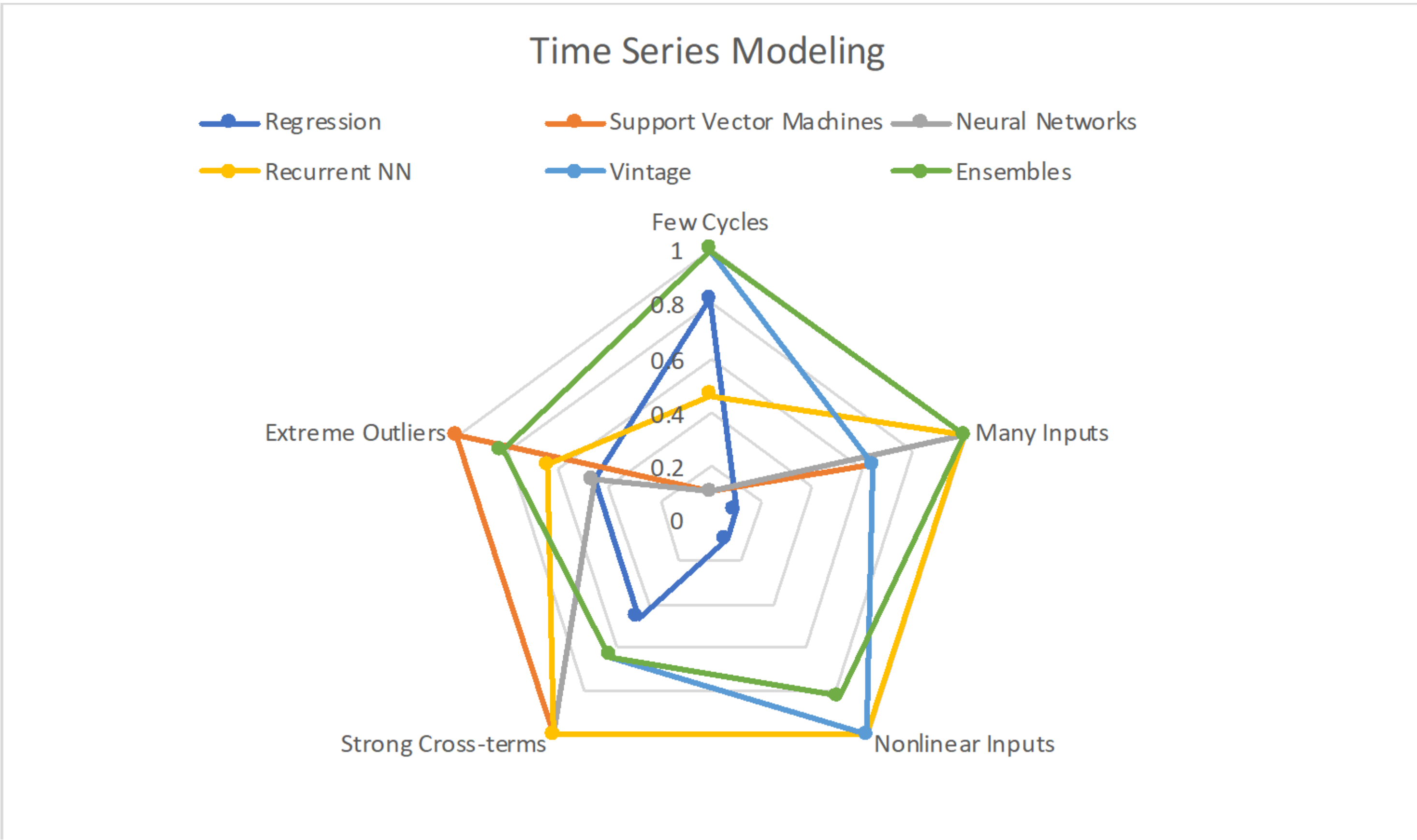
Improving the Field

Charting Strengths: Credit Scoring



There is no “best” method.

Charting Strengths: Credit Scoring



Thank You

Joseph L. Breeden, PhD, breeden@prescientmodels.com,
+1-505-670-7670

CEO of

- Prescient Models LLC
- Deep Future Analytics LLC
- AuctionForecast.com LLC

Associate Editor of

- Journal of Credit Risk
- Journal of Risk Model Validation

President of Model Risk Managers' International Association, MRMIA.org