Fair Lending Considerations of AI / ML Models

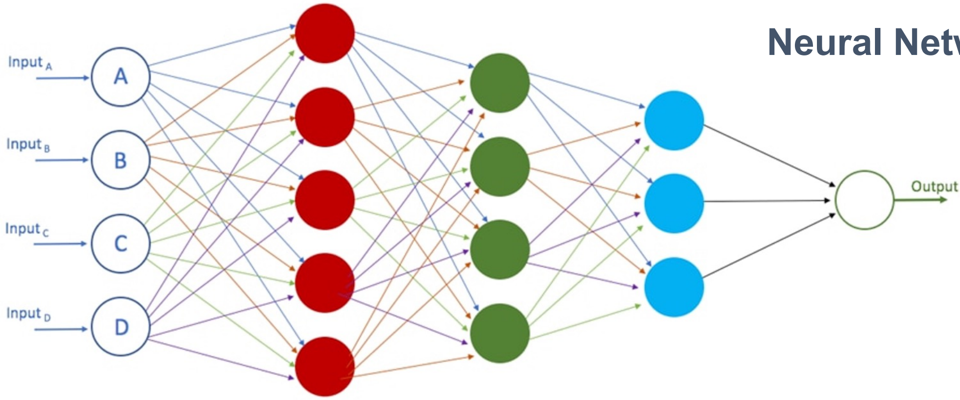
Richard R. Pace

Pace Analytics Consulting LLC
www.paceanalyticsllc.com
ric@paceanalyticsllc.com

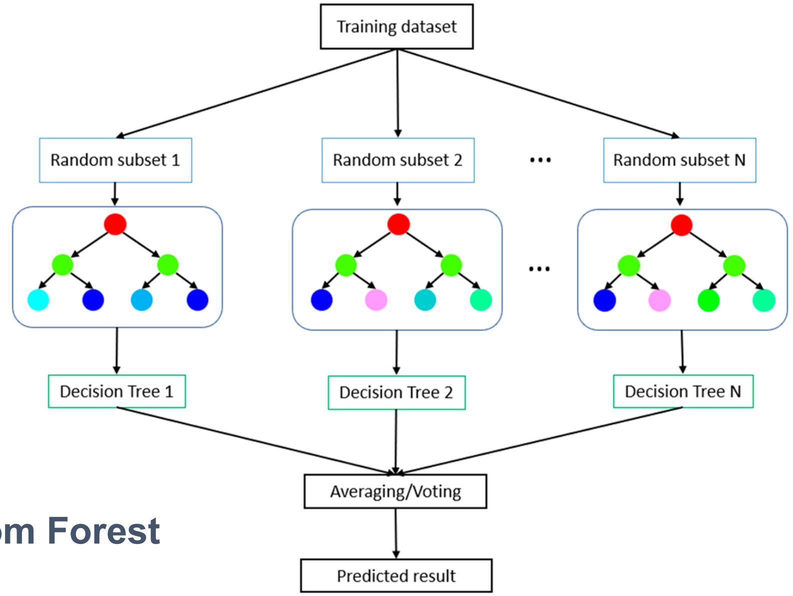
Why Are Consumer Lenders Embracing AI / ML Models?

AI / ML models have the potential to increase significantly the efficiency and effectiveness of automated business decisions – particularly for consumer lenders – through:

- Analyzing historical data on a scale and breadth that is nearly impossible for traditional, "hand-crafted" predictive models.
- Discovering important, more complex predictive relationships that improve overall predictive accuracy– as well as accuracy for historically disadvantaged sub-groups.
- Potentially expanding financial access and lowering borrowing costs for all applicants through increased predictive accuracy.
- Dynamically adjusting the model in response to new data reflecting changing consumer behaviors, a changing economy, or other environmental impacts.



Neural Network



Random Forest

What Are The Unique Risks of These Models?

Predictive models have been used in consumer lending for decades. What is different about AI / ML models that explains the current regulatory focus?

Risk Area	Traditional Models	AI / ML Models
Data Inputs	• Relatively large (10,000 – 1 million+) data samples. • Traditional ”hard data” types (credit bureau, financial, and loan application data). • Significant accumulated experience working with such data.	• Extremely large (1 million – 100+ million) data samples. • Frequent use of “alternative data” sourced from third-party data aggregators (e.g., personal, social, psychographic, purchase histories, web behaviors). • Some third-party data may be: (1) estimated via other models, (2) based on cohort / geographical aggregates, (3) have large percentages of missing values, (4) lack transparency, and (5) be relatively unfamiliar to modelers.
Model Methodology	• Primarily linear models such as linear and logistic regression. • Typically includes < 30 predictive attributes. • Predictive attributes are selected based on business expertise, causal relationships, and intensive statistical analysis. • High degree of transparency and explanability.	• Highly complex, non-linear models – some of which are still subject to active research and, therefore, not fully understood. • Use of model ensembles – i.e., predictions of hundreds or thousands of individual models can be aggregated together to generate the final prediction. • Models can include 1,000+ predictive attributes. • The algorithms create the set of predictive attributes in an automated manner. In many cases, these attributes are “Frankensteinian” and non-intuitive. • There is generally no causal or business-driven justifications for the final predictive attributes. They are selected by the algorithm because their correlations achieve maximum predictive accuracy. • Models tend to be highly opaque “black boxes” that are difficult to explain.

What Are The Unique Risks of These Models?

Predictive models have been used in consumer lending for decades. What is different about AI / ML models that explains the current regulatory focus?

Risk Area	Traditional Models	AI / ML Models
Developer Model Validation	• Conceptual soundness and statistical significance of estimated relationships. • Predictive Accuracy: In Sample, Out-of-Sample, and Out-of-Time. • Error analyses. • Business expert reviews.	• Predictive Accuracy on validation and test samples.
Model Performance	• Models tend to exhibit stable predictive performance for relatively long periods. • Models tend to be updated relatively infrequently during stable environments.	• Model performance tends to deteriorate soon after deployment. • Developers frequently update models due to performance deterioration and attribute the need for such updates to “data drift” or “concept drift”.
Regulatory Compliance	• Straight-forward identification of Adverse Action reasons. • Use of smaller set of statistically-significant, business-justified predictive attributes based on traditional data reduces Disparate Impacts risks.	• Adverse Action reasons are much more difficult to identify due to large number of highly correlated attributes, and rely on complex estimation approaches that have their own potential pitfalls. • Disparate Impact risks are higher due to: (1) opaqueness of how the model works, (2) frequent use of non-causal alternative data, (3) uncertainty around the underlying statistical basis for individual predictive attributes, and (4) the potential of the “Frankensteinian” attributes created by the algorithm to inadvertently proxy for protected class membership.

In addition to safety and soundness concerns, these risks also raise the following consumer protection concerns:

- Lack of transparency into the drivers of decisions – including those required for Adverse Action Notifications.
- Potential Disparate Impact from: (1) use of alternative data and (2) use of complex “Frankensteinian” predictive attributes that may proxy for protected class membership.
- Impermissible use of consumer data to build the algorithm.

How Are Companies Addressing These Fair Lending Risks?

Adverse Action Notices ("AANs")

- AANs require information from the algorithm as to the primary attributes that contributed most to the adverse decision.
- With traditional models, these primary attributes are easy to identify given the linear nature of the models and the relatively low number of attributes.
- For AI / ML models, however, the significant algorithmic complexity and much larger number of predictive attributes makes this identification a significant challenge.
- Several explanability techniques have been developed to extract such information from these algorithms (e.g., SHAP, LIME). However,
  - These techniques generate estimates; as such, they also have risks and limitations that need to be considered.
  - These techniques can sometimes struggle to produce consistent and sensible information in the presence of large numbers of correlated attributes.
  - The outputs of these techniques can be fragile – changing significantly based on small updates to the training data.
  - These techniques can be computationally intensive and require appropriate expertise to configure and interpret correctly.

How Are Companies Addressing These Fair Lending Risks?

Disparate Impact

- Whether driven by alternative data, the "Frankensteinian" predictive attributes, or other potential sources, AI bias is of significant concern and can lead to large scale legal, regulatory, and reputational exposure.

- "Old school" model compliance reviews in which the list of predictive attributes is qualitatively evaluated for potential fair lending concerns has become both outdated, as well as infeasible, in the AI / ML world.

- Tech companies, academics, and other AI startups have developed a multitude of "AI Bias" metrics – as well as AI "de-biasing" approaches – to address this issue. However,

  – We have yet to hear from the regulators on these metrics and de-biasing approaches.

  – There is a growing perspective that models should be "de-biased" whenever disparate impact has been identified – which is at odds with the "business necessity" prong of the disparate impact legal assessment.

  – "De-biasing" approaches search for "less discriminatory alternatives" ("LDAs"); however, it is unclear legally and from a regulatory expectation perspective how much effort is required to search for LDAs.

  – Another important open question is whether demographic information can be legally used for algorithmic "de-biasing" purposes.

  – None of the de-biasing approaches I have reviewed consider the potential adverse impacts that de-biasing has on overall model conceptual soundness and validity – thereby potentially creating a safety and soundness issue.

For further information, the following articles are available at www.paceanalyticsllc.com

Navigating the Score Wars: Four Potential Pitfalls of AI/ML-Based Credit Scoring Models

Don't You Forget About Me: De-Biasing AI/ML Credit Models While Preserving Explainability

Modern Fair Lending Analysis: The Hidden Biases in BISG Proxy-Based Disparity Estimates