Adoption of Machine Learning Underwriting Models in U.S. Credit Markets Is Intensifying Focus on Explainability, Fairness, and Inclusion

WASHINGTON, D.C., September 16, 2021 – Machine learning models are being used to evaluate the creditworthiness of tens of thousands of consumers and small business owners each week in the U.S., increasing the urgency of answering key questions about their performance, governance, and regulation, according to a report "The Use of Machine Learning for Credit Underwriting: Market & Data Science Context" issued today by FinRegLab.

Although adoption of machine learning has been slower in credit underwriting than for other financial applications, the report finds that banks and non-banks are using machine learning models to make credit decisions and that many more lenders are interested in their use. The models have potential to increase approvals of creditworthy applicants and reduce the number of people who are offered credit on terms that they are unlikely to be able to repay. These effects, particularly when lenders incorporate new types of data, have the potential to increase access to credit for millions of consumers – including disproportionately high numbers of Black, Hispanic, and low-income consumers – who are difficult to assess using traditional models and data.

But because machine learning models are often more complex than current underwriting algorithms, understanding and monitoring their reliability and fairness are critical. Stakeholders in industry, government, academia, and civil society organizations are debating whether the models and diagnostic tools are sufficiently transparent to manage concerns that the models may not perform well under changing conditions or that they may replicate or even exacerbate past discrimination.

"Machine learning underwriting models are a reality today, not just a possibility for the distant future," said FinRegLab CEO Melissa Koide. "Answering threshold questions about our ability to explain and manage these models is critical to making their use safer, fairer, and more inclusive."

FinRegLab's report describes the current use of machine learning underwriting models and the choices firms are making in developing, implementing, and monitoring those models. Effective oversight of those decisions is important for answering core questions about the models' reliability and fairness, and may require different tools and processes as compared to conventional systems.

This report is part of a broader research initiative on the explainability and fairness of machine learning for credit underwriting. As part of that effort, FinRegLab and researchers from the Stanford Graduate School of Business are assessing the capabilities and performance of diagnostic tools designed to help lenders responsibly use machine learning underwriting models. This research is designed to inform decision-making by policymakers, firms, industry groups, and advocates as the financial services sector develops norms and rules to govern the responsible, fair, and inclusive use of machine learning for credit underwriting.

Based on interviews and other research, the report finds:

- **Use of machine learning models to make credit decisions is most advanced in credit cards and a range of unsecured personal loans.** This reflects the historical position of credit cards as being

at the analytical forefront of consumer finance and the dominance of digital lending in unsecured personal loans. Some auto and small business lenders are also using machine learning underwriting models.

- **Concerns about the ability to explain and understand machine learning underwriting models shape every stage of their development and use**. To improve model transparency, some firms are imposing up-front constraints on their machine learning models to reduce their complexity. Other lenders are using *post hoc* explainability methods—supplemental models, analyses, or visualizations—to make complex or "black box" models more transparent. Explainability technologies are evolving quickly, and stakeholders are debating the tradeoffs of different approaches.

- **Firms and regulators are also focusing on whether and in what circumstances the use of machine learning can improve fair lending oversight.** Financial services stakeholders are particularly intrigued by the potential for machine learning to improve available tradeoffs between performance and fairness when mitigating sources of adverse impacts in credit decisions.

- **Questions about the capabilities and trustworthiness of machine learning models and how to enable necessary oversight will shape the scope and pace of adoption moving forward.** Concerns about the trustworthiness of machine learning models are being raised in a broad range of sectors with regard to general transparency, reliability, fairness, privacy, and security. For credit underwriting, these concerns include whether lenders have the capability to recognize changes in lending conditions that can cause rapid deterioration in model performance or to identify and reduce bias in the model or data. Stakeholders are also debating whether and how to update existing regulatory regimes concerning model risk management, fair lending, and applicant disclosures to address the use of machine learning underwriting models.

The report surveys market practice with respect to the use of machine learning underwriting models and provides an overview of the current questions, debates, and regulatory frameworks that are shaping current use. It is also designed to provide a resource to stakeholders, especially non-technical ones, that explains the decisions that lenders using machine learning underwriting models can make to promote responsible, fair, and inclusive use of those models. The report focuses in particular on model transparency as a critical ingredient in allowing firms, regulators, and other stakeholders to evaluate the reliability and fairness of these models.

Empirical Evaluation of Model Diagnostic Tools

The forthcoming empirical evaluation will be the first public research shaped by input from key stakeholders – including executives from banks and fintechs, technologists, consumer advocates, and regulators – to address questions about the capabilities and performance of tools designed to help lenders use machine learning underwriting models in compliance with existing laws and regulations.

The next project report will analyze the capabilities and performance of both open-source and proprietary model diagnostic tools developed to support lenders' compliance with regulatory requirements in the following areas:

- **Model Risk Management**: the ability to satisfy prudential regulators and investors about the performance, reliability, and governance of machine learning models.

- **Fair Lending**: the ability to demonstrate that these models operate without creating impermissible discrimination, including disparate impacts on protected classes.

- **Adverse Action Reporting**: the ability to comply with legal requirements for providing applicants with individualized adverse action notices explaining why they were denied credit or offered less favorable terms.

The project's aim is not to identify a particular "winner" among available tools for managing machine learning models, but rather to help stakeholders who are grounded in processes designed for traditional underwriting models get a broader sense of the range of approaches and outcomes that are possible in the context of machine learning.

The following technology companies participating in this research are: ArthurAI, Fiddler Labs; H2O.ai; Relational AI; SolasAI/BLDS, LLC; Stratyfy; and Zest AI.

Subsequent FinRegLab reports will analyze ways in which existing law, regulation, and market practices may need to evolve to encourage safe, fair, and inclusive use of machine learning underwriting models.