

12-12-2021

Ethical AI in FinTech: Learning & Scaling from the Credit Risk domain

Venkat Subramanian
PayPal, venkasubramanian@paypal.com

Follow this and additional works at: https://aisel.aisnet.org/treos_icis2021

Recommended Citation

Subramanian, Venkat, "Ethical AI in FinTech: Learning & Scaling from the Credit Risk domain" (2021). *ICIS 2021 TREOs*. 15.

https://aisel.aisnet.org/treos_icis2021/15

This material is brought to you by the TREO Papers at AIS Electronic Library (AISeL). It has been accepted for inclusion in ICIS 2021 TREOs by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Ethical AI in FinTech: Learning & Scaling from the Credit Risk domain

Venkat Subramanian, Global Data Science, PayPal (venkasubramanian@paypal.com)

Today AI-based automated decision-making systems have become ubiquitous across industries. FinTech industry is one of the pioneers in this space due to the availability of vast quantum of in-house, online transactional data, enabling superior AI models to be built. Many facets of decision making in Fintech have become entirely automated using AI, starting from identifying/preventing fraudulent transactions to assessing credit risk of customers with zero human intervention. In such scenarios, ethicality in AI-based decisions becomes very important. While Ethical AI is a broad area that covers multiple dimensions like consumer privacy, transparency and explainability of decisions (F Rossi, 2018), there is one aspect which poses a higher level of difficulty in implementation than others: “Fairness” in AI decisions.

Difficulty in ensuring fairness arises in 3 ways –

- (1) Defining what is fair in each context
- (2) Creating efficient & scalable processes to ensure fairness across all AI models consistently
- (3) Creating the right incentives and oversight to make sure model developers adhere to them

In this work, we are interested in #2 – creating efficient and scalable processes to ensure fairness across all AI models consistently. Today there are many publicly available tools and packages that help in this aspect viz. AI Fairness 360 from IBM, LinkedIn Fairness Toolkit (LiFT), and many more (Zhang and Zhou 2019). However, these packages are not specific to FinTech and do not take some of the industry-specific nuances into account. We address this gap by proposing a new, automated framework for validating fairness in AI decisions in the FinTech context, leveraging some of the existing regulation-driven protocols in this industry.

Specifically, Credit Risk happens to be one of the areas in FinTech where existing regulations (viz. Equal Credit Opportunity Act, 1974 in the US) ensure that every AI model goes through a ‘Fair Lending Review’ with legal and compliance teams, who review the model to ensure that ‘individual fairness’ is achieved (i.e., individuals with similar characteristics achieve similar outcomes irrespective of belonging to a specific/protected group (Mehrabi et al. 2019)). Extending this process beyond credit models can ensure consistency across all functional areas within Fintech. However, resource-intensiveness of the review process becomes a constraint. Because of the number of variables used in today’s FinTech models (in the order of a few hundreds to a few thousands of variables at times), and the novel nature of variables (viz. ‘emojis used by a customer in her transaction memo’), a typical fairness review requires extensive human expertise/judgment to scrutinize a model at variable level and consequently, takes anywhere between a few weeks to a few months at times.

We propose a framework that automates this entire review process irrespective of the model algorithm, complexity and the number of variables used. Using a combination of multivariate statistical and machine learning techniques, the framework automatically segregates variables in a model that act as proxies for group membership of an individual (compromising the principle of individual fairness) from variables that reflect genuine individual financial behavior. This should enable Fintech firms to scale fairness reviews to all AI models across functional areas consistently with minimal incremental effort.