

## Alternative Credit Scoring: Utilizing AI and ML

November 2022

*Financial institutions (FIs) have access to more data than ever. This has presented an opportunity for FIs to surf through the noise and apply useful and relevant data to measure the creditworthiness of their retail and MSME clients. In this context, it is key to maintain transparency and interpretability for credit scoring while making the best possible use of these changing sets of information. The November session of DataTalk explored alternative credit scores, the data used for building them, their accuracy when compared to non-traditional data sets, and how algorithms used for credit scores leverage AI and ML to analyze these changing sets of information.*

**The use of alternative data is not new.** Cash flow and transactional data have been used for some time in credit risk for retail customers and Micro, Small, and Medium Enterprises (MSME's). However, fintech companies are accelerating the trend and leveraging alternative sources of data to a greater degree. This trend is leading to a growing focus on open finance and further, towards open data. The open data concept, embraced by authorities like the HKMA with its "Commercial Banking Interchange" initiative, or the MAS's 'Singapore Financial Data Exchange', is beneficial for participating FIs on at least two fronts: (1) Reciprocity, generally absent in open banking mandatory approaches; and (2) the possibility to charge for the information that is shared by FIs. These benefits have proven to be a valuable tool for authorities to guarantee the participation of FIs and non-FIs in their data-sharing initiatives.

**Innovation in data usage is changing how regulators interact with FIs.** There has traditionally been a limitation on non-traditional data sources by traditional FIs with regulatory considerations in mind. When alternative data sources become widely used, regulators will need to adjust quickly to consider objectives and opportunities for improvement of outcomes as well as providing clarity and guidance. Consumer protection will remain a key objective while ensuring that regulatory coordination can keep pace. This remains a particular challenge because important new sources of data and alternative scoring depart from long established models familiar to regulators. Some regulators are already addressing this challenge by evolving from open banking regulation to open data frameworks which require the flow of data across different sectors. A few governments have gone further and are working in an orchestrator role to facilitate gathering information from different sectors to be shared for the cultivation of benefits for consumers. Discussants shared how more concrete guidelines from regulatory authorities could make the principles of model transparency and explainability components for consumer safety verification.

**More data does not necessarily mean good data.** It has been a challenging task for institutions to mine existing data and migrate it into usable formats for credit modeling. For data to be valuable, it must be organized and structured. Additionally, data doesn't add value equally for all customer segments and its value is dependent on the product or service the data will be used for. In particular, access to new data sets tends to more beneficial to retail customers and MSMEs. Robust data sources and a strong understanding of the customer base is needed for an effective and applicable model that ensures consumer safety.

**Model creation and validation of AI and ML play essential roles in harnessing the value of data.** It's a common fallacy to presume that as long as an institution has good data, they don't need anything else. But in fact, if the model for credit risk is not transparent and explainable to consumers, it presents more risks than benefits. Even as alternative data sources evolve and more alternative modeling methods are utilized, traditional credit bureau scores will remain an essential component of risk model accuracy.

*We look forward to continuing the DataTalk series on December 20 at 9:00 AM Washington DC / 2:00 PM London / 10:00 PM Singapore, where we will focus on data exchange, APIs, and security.*