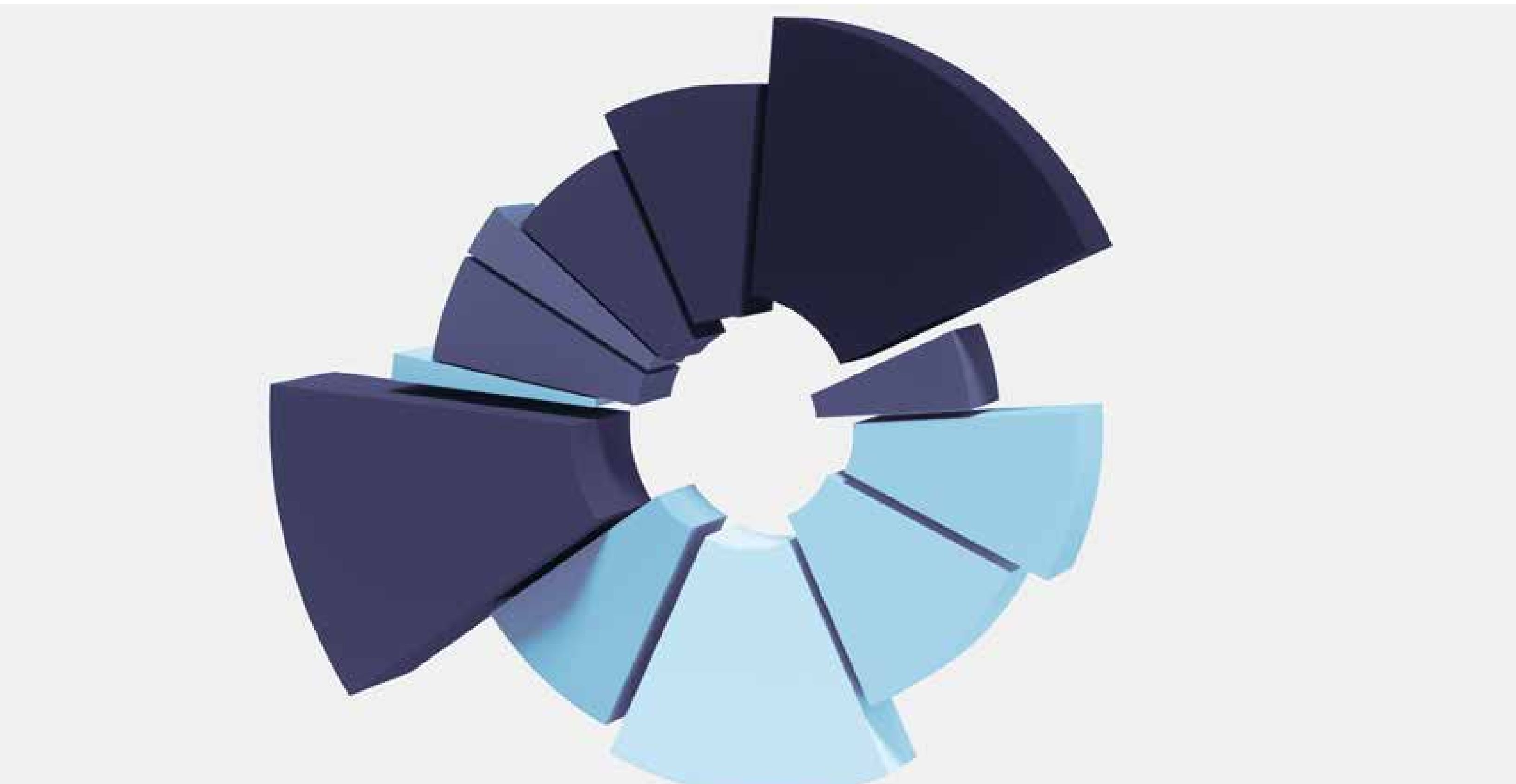
McKinsey & Company

Risk & Resilience Practice

Designing next-generation credit-decisioning models

As banks continue their digital transformations, they can follow four best practices for automated credit-decisioning models to incorporate more of the right data to meet future challenges.

by Raj Dash, Andreas Kremer, and Aleksander Petrov



December 2021

The continuing advances in big data, digital, and analytics are creating fresh opportunities for banks to improve the credit-decisioning models that underpin their lending processes. The new, high-performance models allow banks to define lending (and capital) parameters more precisely and thus sharpen their ability to approve creditworthy customers and reject proposals from customers who either are not creditworthy or cannot afford further debt. In fact, the banks (and fintech companies) that have put such new models in place have already increased revenue, reduced credit-loss rates, and made significant efficiency gains thanks to more precise and automated decisioning.

Use of the new credit-decisioning models during the COVID-19 pandemic showcased their benefits. They have performed well, while traditional models have struggled to handle the changing customer circumstances, forcing banks to resort to Band-Aid solutions (for example, expert adjustments of default rates at portfolio-segment levels). In this article, we share four best practices that we have observed when designing new or upgrading existing credit-decisioning models.

Challenges and benefits

Many banks struggle with transitioning to a more advanced credit model. They face significant capability, technology, and cultural hurdles, including a limited set of data sources; simple analytical engines; a heavy reliance on subjective assessments from relationship managers (RMs) and underwriters; outdated, inflexible models that have been patched over time; and concerns about the length of implementation and regulatory reviews.

These challenges are real and should not be downplayed. But the benefits of overcoming them should not be downplayed either. Banks that have already embedded high-performance credit-decisioning models into their digital lending have reaped three key benefits:

— *Increase in revenue.* The new models have led to a revenue increase of 5 to 15 percent through higher acceptance rates, lower cost of acquisition, and better customer experience. By better distinguishing between creditworthy and noncreditworthy customers, banks can improve acceptance rates and pricing. Meanwhile, a

Banks need to implement more automated credit-decisioning models that can tap new data sources, understand customer behaviors more precisely, open up new segments, and react faster to changes in the business environment.

credit-decisioning model that automates large parts of the assessment process and eliminates paper-heavy steps lowers the cost of acquisition and improves the customer experience. This also results in faster executions that reduce the typical price slippage observed with longer timelines.

— *Reduction in credit-loss rates.* Companies have seen a decrease of 20 to 40 percent in their credit losses by using models that could more precisely determine customers' likelihood to default. That element affects the levels of provisions and capital that a bank must hold.

— *Efficiency gains.* Use of the new models have resulted in 20 to 40 percent improved efficiency, thanks to a combination of more highly automated data extraction, case prioritization (for example, using straight-through processing for low-risk cases while analyzing higher-risk cases more thoroughly), and model development.

A business-critical competitive imperative

Based on those three benefits of improved credit-decisioning models, the average bank with €50 billion in assets from small and medium-size enterprises (SMEs) could see €100 million to €200 million of additional profit. In addition to these benefits, there are serious downsides when banks do not put next-generation credit models in place. In the past, banks updated the models only every five to ten years. That was viable when banks could rely on their incumbent positions to preserve market share and profitability. However, the strategy is considerably less tenable as customer information becomes more democratized via "open banking" and regulations such as PSD2—and as fintech companies and attacker banks proliferate and focus on an increasingly digitally savvy customer base.

Particularly troubling is that many credit-decisioning models today rely on historical data that are virtually useless, given the market disruptions caused by the COVID-19 pandemic. Some banks have applied model overlays that are subjectively derived and are not precise enough for underwriting—often at an industry or geographic level. For example, such overlays might assign a high likelihood of default to the hospitality sector in London without differentiating between a restaurant that has rapidly transitioned to an omnichannel model (that deals with business interruptions and lockdowns) and one that has not. Making matters more urgent is the looming wave of business defaults expected as governments begin to ratchet back their unprecedented support. Banks need to identify such companies quickly.

In other words, using new credit-decisioning models is not only a powerful way to boost profits but also a business-critical competitive imperative. Banks need to implement more automated credit-decisioning models that can tap new data sources, understand customer behaviors more precisely, open up new segments, and react faster to changes in the business environment. This will allow them to serve their customers better, grow their business, and compete with fintech companies and attacker banks that are constantly upping their technology games and looking to grab market share.

Four best practices

McKinsey has identified four best practices when designing new credit-decisioning models: implement a modular architecture, expand data sources, mine data for credit signals, and leverage business expertise. We have also defined a five-stage agile process to implement a new model in less than six months, much faster than the typical 12 to 24 months.

Implement a modular architecture

Theoretically, one optimization algorithm run over a consolidated database built by mixing all underlying data sources would yield the (global) optimal model. However, industry leaders prefer a modular architecture that incorporates numerous submodels (or modules) based on data coverage and information on industry and geographic differences that are combined to give one meta credit signal. Because this modularity makes it easy to add or remove modules, banks can integrate new or different data into the model to keep it flexible and robust. The key element here is how a bank goes about merging the submodel scores into one score (Exhibit 1).

For example, a bank could use a modular architecture to react quickly to a severe economic disruption, such as the COVID-19 pandemic. It could design a module with such a pandemic in mind (for example, a module focused on changes in cash and income positions) and give the module a higher weight in the model to receive more finely tuned credit signals.

This architecture is not just useful in times of economic distress. A more precise coverage of different population segments can open up new growth areas. Banks might have overlooked some potentially profitable areas in the past, since they might not add a lot to the total Gini Index. For example, a director module applies to a relatively small number of people. But the way that directors treat their companies (especially in the SME space) is a strong predictor of future solvency, so this small module can yield very important credit signals.

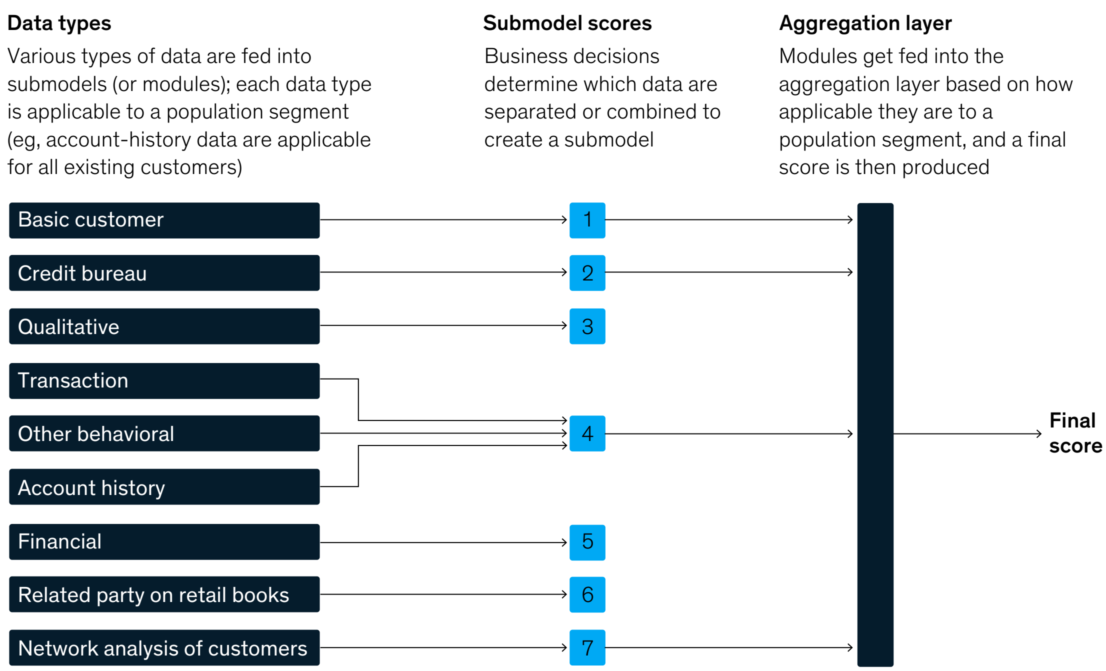
We also often find that banks follow a product-centric approach to credit models, only analyzing the data relevant to that product. A customer-centric approach, which combines the data signals from all product areas where the customer interacts, nearly always yields a higher-performance model.

Coordination among stakeholders—the business, model-development team, and model-maintenance team—is critical to implementing this architecture. When designing submodels, model-development teams need to consult with the business to validate assumptions. At the same time, model development

Exhibit 1

A modular credit-decisioning model groups data sources into submodels, which are primarily based on population segment and business inputs.

Credit-decisioning model using submodels to determine final score (illustrative)



can show if the data in each submodel are sufficiently distinct. Data overlap could skew results. For example, sourcing financial-related information from both the credit bureau and the company financials could result in a financial factor occurring twice in the model, thus double counting its true influence (Exhibit 2).

Expand data sources

Industry leaders tap multiple internal and external data sources to improve the predictive power of credit signals. They do a better job than their competitors do of leveraging internal sources of traditional data, enriching that with internal nontraditional data, and supplementing those data with external traditional data. They also explore other external nontraditional data—and even include some subjective data from RMs and underwriters to fill gaps (Exhibit 3).

In our work with clients, we have found that transactional data are particularly useful. For example, leaders are embracing open banking as a way to leverage traditional transactional data better. Indeed, it is likely that open banking will be the foundation for next-generation credit analytics. That is because open-banking data can provide a more complete view of the customer than other data can, since it can look across banks—giving a more accurate estimate of income, for example. And open banking APIs can readily pull many transaction details.

As for nontraditional external data that can supplement internal data, telecom data are an excellent example. Many individuals lack credit history. However, their mobile phones generate rich data about individual behavior, including bill payments for phone usage, call and text patterns, and purchases made via mobile phone.

Additionally, since what people share on their social and professional networks is often very revealing, this network information is increasingly recognized as an important source of nontraditional external data. Does the person associate with others with histories of bad credit or fraud? Has the person's job location changed frequently? Where has the person traveled?

Exhibit 2

A modular structure in a credit-decisioning model helps address some key challenges pertaining to model development and maintenance.

Parts of a modular credit-decisioning model

Agile approach

- Prioritizing data sources that are critical for release of a minimal viable product
- Constant learning and optimizing for future releases
- Distributing work by modules and finalizing data sources as they are ready

Model maintenance

- Adding a new data source for future model updates
- Removing a data source (eg, data source that is not production ready)
- Updating a particular data source without having to update entire model

Model development

- Grouping similar data sources to help understand individual performance contribution and gather expert inputs in a structured manner
- Focusing model development on applicable segments of customer population (eg, using certain financials only on customers with financials that can be collected)

Exhibit 3

Both the internal and external data sources used in a credit-decisioning model will affect the decision quality.

Counter-party data sources

Leverage internal, traditional data
- Customer demographics (eg, city, income)
- Product usage
- Customer profitability
- Loan data
- Deposit data
- Current account data
- Transactional data
- Bank data from point-of-sale transactions

Enrich with internal, nontraditional data
- CRM[2] data (eg, campaign response)
- Customer specifics (eg, share of wallet)
- ATM usage (eg, location)
- Life-cycle stage of customer relationship
- Call records (eg, call-center customer-interaction notes)
- Email records
- Customer feedback (eg, surveys, satisfaction data)
- Navigational data from bank websites

Supplement through external, traditional data
- External data from retailers
- External data from telco or utility partner (eg, top-up patterns, monthly bill payments)
- Web-browsing data
- Credit-bureau data
- Data from other financial institutions (eg, insurers, brokerages)
- Data from government agencies (eg, tax payments, updated demographic data)

Explore external, nontraditional data
- Public data (eg, blogs, wikis)
- Sentiments on social media (eg, LinkedIn, Facebook)
- Supermarket and e-commerce data
- Video analysis of customer footage
- Customer-network data

Leverage data sourced from RMs[1] and underwriters
- Qualitative questions on customer management, governance, and performance
- Financial information on cashflow and balance-sheet elements

[1]Relationship managers.
[2]Customer relationship management.

Mine data for credit signals

Leading banks are especially good at mining their internal existing data, as well as combining data sources to extract highly predictive credit signals. This requires advanced analysis of existing sources of data, such as mining transaction data that go far beyond the simple number of days past due date and inflow and outflow analyses typically found in their behavioral models. These banks also use open-banking data (based on jurisdiction) to identify complex spend and income patterns, construct synthetic financial and cash-flow statements, then leverage the synthetic statements to tease out credit signals and identify new ways to segment the customer base. Leading banks that have partial access to their customers' data apply machine learning (ML) and AI to form a more complete, although slightly imprecise, view of the customers.

A handful of banks are even more advanced. Within them, the business and modeling experts are highly coordinated and use deep analytics to mine an expanded set of data for credit signals. For example, they use text-mining and natural-language-processing techniques to tag transaction details more accurately, and they monitor network activity to incorporate personal and business activities into their credit-risk models. These banks are also at the forefront of applying ML and AI in three areas:

— *Creating segmentation rules.* Using ML techniques, a bank can better identify specific variables to define precise and sometimes new customer segments. For example, risk factors related to local geography are more relevant to smaller businesses, whereas risk factors derived from company financials are more relevant for larger businesses.

— *Deriving credit signals.* A variety of methods, including traditional variable transformations (for example, through power functions, splines, and trend analysis) and deriving business sensible ratios, exist. For example, a US bank in a joint venture with a supermarket could increase the Gini coefficient (predictive power) of a single variable derived from SKUs to 32, from 8, by applying a variety of such transformations. More recently, banks have added to this arsenal by applying ML and AI techniques. For example, by applying natural-language-processing techniques to transactions conducted on an account, a bank can identify changes in rent and utility payments from individual customers and thus determine if a customer is facing credit challenges.

— *Building and validating challenger models.* A bank can use ML techniques to develop challenger models in parallel with its credit-risk models to discover where other credit signals could potentially lift performance.

For example, we worked with a large retail bank that wanted to improve the predictive power of its regression-based behavior model for its credit-card portfolio. But the bank wanted to avoid using a "black box" approach that could prevent it from fully understanding the algorithm's decisions. The bank used its ML model to understand specific segments where it could improve the regression-based model. It then carved out those segments and built dedicated models that improved predictive power to 75 percent, from 67 percent (as opposed to 80 percent in a pure ML model).

Leverage business expertise

ML and AI are extraordinary tools, but credit models should not be based solely on statistical methods. For a truly robust and high-performing model, banks need to leverage their internal business expertise during the model-development process. This will help them better understand where credit signals are missing and then identify and validate new credit signals.

For instance, model designers should interview underwriters and RMs on credit issues and work with them on how to translate their insights into qualitative questions for better credit signaling. These business experts can also help validate credit signals based on their own real-life interactions with customers, knowledge of bank processes, and understanding of compliance. For example, business experts in the United Kingdom highlighted the importance of trade flows after Brexit to understand the credit performance of export-oriented businesses. This signal might not have been as prominent in the model, since trade disruptions were historically not experienced in profitable and growing businesses.

A step-by-step approach to transformation

By following a five-stage, agile process, banks can implement a new credit-decisioning model in less than six months—much faster than the 12 to 24 months that is the industry norm today:

— *Credit-model walkthrough.* Analyze the credit model, reviewing its methodological setup, performance, and use to identify potential areas of improvement.

— *Credit-scoring-model diagnostic and design.* Evaluate the current state of data readiness, identify data sources that are easy to include in modeling, and then create a road map for including those data. Compare model performance across various segments and against that of peers to identify areas of weakness.

— *Data preparation and engineering.* Engineer the data to ready them for modeling (for example, formatting, completeness testing, and performing missing value and record treatment).

— *Development of next-generation credit-scoring models.* Develop a production-ready minimal viable product. Typically, this requires three two-week cycles of modeling, incorporating expert and analytical feedback within each cycle.

— *Integration of credit scoring in lending transformation.* Automate the lending processes and update the credit-decisioning model with the new credit-scoring models.

As banks continue to digitize their enterprises, they need more sophisticated and automated credit-decisioning models that can incorporate a wide variety of traditional and nontraditional data from inside and outside the organization. This will make them more competitive and resilient in challenging economic times and in the face of intense pressure from fintech companies and challenger banks. The four best practices discussed here can help any bank elevate its credit model.

Raj Dash is a senior expert based in McKinsey's London office, where **Aleksander Petrov** is a senior partner; **Andreas Kremer** is a partner in the Berlin office.

The authors wish to thank Siddhartha Mysa and Jagriti Singh for their contributions to this article.

Designed by McKinsey Global Publishing