

LIFT *papers*

Revista do Laboratório
de Inovações Financeiras
e Tecnológicas

#4 | ABRIL 2022

LIFT Papers

Revista do Laboratório de Inovações
Financeiras e Tecnológicas

Número 4 | Abril 2022

Editor-Chefe da Revista

André Henrique de Siqueira, PhD

Editor-Adjunto da Revista

Aristides Andrade Cavalcante Neto, MSc
Rodrigo de Azevedo Henriques

Corpo Editorial da Revista

Danielle Sammyres Figueirôa Alves Teixeira

Ficha catalográfica elaborada pela Biblioteca do Banco
Central do Brasil

LIFT Papers / Banco Central do Brasil. N. 4,
(abril 2022). Brasília: Banco Central do Brasil,
2020.

Semestral
Disponível em:
<https://revista.liftlab.com.br>
ISSN 2675-2859

1. Inovação Tecnológica – Brasil. 2. Sistema
Financeiro – Brasil. 3. Crédito. I. Banco Central do
Brasil.

CDU 336.7:004.738.5

Presidente do Banco Central do Brasil

Roberto Campos Neto

Presidente da Fenasbac

Paulo Renato Tavares Stein

Comitê Executivo LIFT 2021

DIRAD – Coordenação LIFT
Aristides Andrade Cavalcante Neto
André Henrique de Siqueira

FENASBAC – Coordenação LIFT
Rodrigo Henriques

DIORF
Cesar de Oliveira Frade

DEPEP
Ricardo Schechtman

DIPOM
Marcos Nascimento Silvino

DIREC
João Paulo Resende Borges

Parceiros de Tecnologia – Edição 2021 (por ordem alfabética)

AWS
Celer
Cielo
IBM
Instituto Fenasbac
Microsoft
Multiedgers
Oracle
R3
RTM

Plataforma de Open Banking para Todos

Filipe Bruno Chan¹

A Plataforma de Open Banking para Todos é um projeto criado pela Klavi, uma empresa de tecnologia e analytics que apoia empresas a tomarem melhores decisões por meio do Open Finance. O objetivo deste projeto foi criar um modelo de crédito novo e demonstrar que ele é mais preciso do que modelos de mercado.

Durante este projeto, contamos com o apoio de um parceiro que nos providenciou um conjunto de dados para modelagem de crédito. Trabalhamos com os dados de mais de 20 mil clientes do parceiro, incluindo informações de crédito tomado, limites de crédito, *scores* de *bureaux*, dados transacionais e mais.

Para a criação do nosso modelo, usamos tecnologias avançadas de *machine learning* e inteligência artificial, como XGB e LGB, técnicas que nossos cientistas de dados, residentes na China, já estão acostumados a usar.

O projeto todo levou em torno de três meses, e obtivemos resultados muito positivos. Nosso modelo final gerou um *lift* (aumento) de ao menos 10 pontos no K-S, um aumento considerado alto.

Pre vemos que nossos modelos de crédito podem ter um impacto positivo enorme no Sistema Financeiro Nacional em três grandes frentes: (1) Inclusão Social, em que mais pessoas terão acesso a mais crédito barato; (2) Competitividade, em que o crédito terá taxas reduzidas e tomadores podem ser mais competitivos; (3) Sustentabilidade, já que os modelos são atualizados frequentemente e usam tecnologia de ponta.

1 bruno@klavi.ai

..... Introdução

Muitos consumidores brasileiros não têm acesso a crédito, e os que têm pagam uma das maiores taxas de juros no mundo. Acreditamos que um grande culpado disso é o uso de tecnologias desatualizadas em modelagem de crédito utilizados por empresas que emitem crédito. A Plataforma de Open Banking para Todos tem como foco melhorar a vida financeira de consumidores por meio do uso de novas tecnologias de modelagem de crédito. Essas novas tecnologias vão permitir a consumidores terem mais acesso a crédito com taxas baratas e competitivas.

Durante este projeto, desenvolvemos a primeira versão dos nossos modelos de crédito e demonstramos que os nossos modelos são mais preditivos que os já utilizados no mercado, junto com nosso parceiro. Os modelos desenvolvidos pela Plataforma de Open Banking para Todos são altamente escaláveis e podem ser aplicados no Sistema Financeiro Nacional inteiro.

Este relatório contém seis seções. Em Objetivos, definimos que o objetivo é criar modelos de crédito que são mais preditivos do que os utilizados pelo mercado. Em Fundamentação Teórica, explicamos a tecnologia e os indicadores principais para medir a predição dos modelos. Em Visão Geral, definimos as três partes do Projeto. Em Funcionalidades, desenhamos o fluxo de uso da solução criada. Em Escopo, esclarecemos quais partes do projeto se aplicam ao protótipo para o projeto LIFT Lab. Em Características Inovadoras, ilustramos a tecnologia que estamos trazendo para o Brasil e como isso vai melhorar o mercado de crédito para todos os consumidores.

..... 1 Objetivos

O objetivo principal deste projeto é demonstrar que modelos de crédito desenvolvidos pela Plataforma de Open Banking para Todos, usando tecnologia de ponta, são mais preditivos que os modelos utilizados no mercado. Consequentemente, buscamos melhorar o acesso a crédito e reduzir taxas de juros para todos os consumidores. Por meio da aplicação desses modelos nas empresas de crédito, o Sistema Financeiro Nacional inteiro deve se beneficiar. Durante este projeto, contamos com nosso parceiro para validar a primeira versão dos modelos.

Plataforma de Open Banking para Todos é um projeto criado pela Klavi, uma empresa de tecnologia e analytics que apoia empresas a tomarem melhores decisões por meio do Open Finance.

O grande desafio das empresas para usufruírem dos dados do Open Finance será como transformar dados desestruturados de diversas fontes em informações analíticas que suportem suas tomadas de decisões, seja melhoria na avaliação do risco de crédito, seja no desenvolvimento de novos produtos, seja na melhor experiência para seus clientes. A Plataforma de Open Banking para Todos é especialista em coleta e transformação de dados de Open Finance, ajudando empresas entenderem os hábitos financeiros de seus clientes usando algoritmos e tecnologia de ponta.

O Open Finance vai reduzir a assimetria de informação financeira que existe no mercado financeiro, já que poucos bancos detêm maioria dos dados, por meio da simplificação do compartilhamento desses dados. Acreditamos que consumidores participantes vão ter melhores ofertas de produtos e serviços financeiros, como crédito, aberturas de

contas, entre outros. Apesar de o Open Finance facilitar o compartilhamento de dados financeiros, ainda não está claro como as instituições receptoras dos dados vão usar essa informação para melhorar suas ofertas.

..... 2 Fundamentação Teórica

Modelos de crédito são construídos analisando-se dados históricos para correlações entre o resultado de interesse (por exemplo, o não pagamento da parcela de crédito) e uma gama de variáveis de um período anterior (por exemplo, renda e pagamento de contas anteriores).

Existem diversos tipos de modelos de risco, como os mais avançados LGB (light gradient boost), XGB (XGboost), RF (random forest), LR (logistic regression) e SVM (support vector machines). Esses modelos são criados usando-se diferentes técnicas de *machine learning*. Cada modelo tem suas vantagens e desvantagens. Para este projeto, criamos cinco modelos usando essas cinco técnicas e comparamos os resultados de cada um para escolhermos o mais adequado.

A precisão dos modelos foi medida usando o teste Kolmogorov-Smirnov, ou simplesmente teste K-S. O teste K-S é um indicador com valor entre 0 e 1; quanto mais alto o valor K-S de um modelo (mais próximo de 1), mais preditivo o modelo. Teste de K-S é o método mais comum para medir precisão de modelos de risco no Brasil. Um modelo com K-S de 0,3 é considerado bom.

..... 3 Visão Geral

O projeto foi dividido em três partes.

- Parte 1: Back test com todos os dados providenciados pelo parceiro.
- Parte 2: Back test com todos os dados providenciados pelo parceiro exceto *scores* de crédito de *bureaux*.
- Parte 3: Enriquecimento do modelo com dados do Open Banking.

3.1 Parte 1: Back test com todos os dados

Fizemos um “back test” com dados providenciados pelo parceiro. “Back test” é uma prática comum no mercado para validar, rapidamente, se um modelo de crédito é mais preditivo que o atual. Durante essa etapa, a Plataforma de Open Banking para Todos teve como objetivo criar um modelo de crédito com um K-S maior do que o utilizado pelo parceiro. Usamos apenas dados de clientes do parceiro já performados, ou seja, de clientes que já tomaram crédito no passado e têm um histórico de pagamento completo. As etapas foram:

1. O cliente enviou para a Plataforma de Open Banking para Todos dados de 20.787 clientes (anonimizados e sem informação cadastral) que já performaram em alguma modalidade de crédito:

- Todos os dados e variáveis que foram usados para prever a performance dos 20.787 clientes. Por exemplo, limites de crédito, transações financeiras, *score* de crédito de *bureaux* etc.
 - A performance dos 20.787 clientes (se pagaram ou não o crédito que tomaram).
 - O K-S do modelo que foi usado para prever a performance desta massa de clientes. Neste caso um K-S de 0,40.
 - **Importante:** Não recebemos dados do Open Banking nesta etapa.
2. A Plataforma de Open Banking para Todos criou modelos de crédito com os dados providenciados:
- Limpamos e padronizamos todos os dados, criando variáveis para serem usadas nos nossos modelos de crédito.
 - Aplicamos os dados nos modelos Plataforma de Open Banking para Todos (LGB, XGB, RF, LR e SVM).
 - Fizemos diversas iterações dos modelos para aumentar o K-S.
 - Comparamos os resultados de K-S de cada modelo com o K-S do cliente.
 - O modelo LGB foi o mais recomendado com K-S de 0,673, representando um *lift* de 27 pontos no K-S do cliente de 0,40.

Figura 1

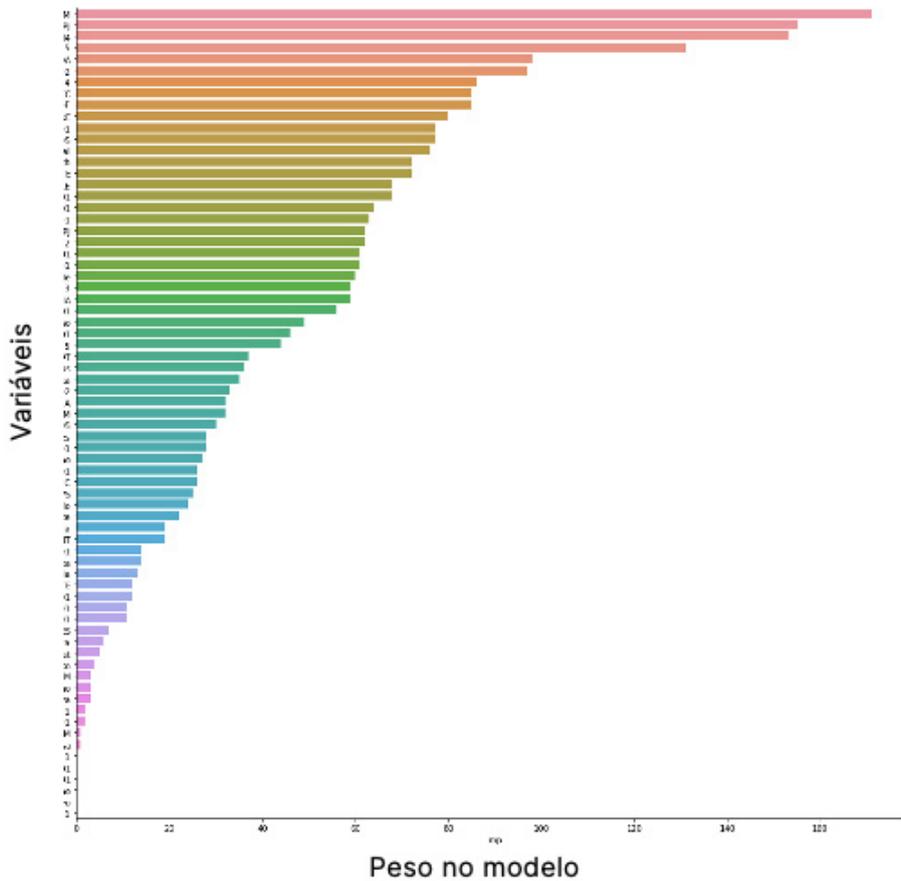
Model	AUC	KS	remarks
LGB	0.905	0.673	
XGB	0.906	0.671	
RF	0.901	0.665	
LR	0.89	0.64	Features need special treatment
SVM	0.807	0.615	Features need special treatment

3.2 Parte 2: Back test com os dados exceto scores de *crédito* de *bureaux*

Na etapa anterior, o parceiro enviou quatro *scores* de crédito para cada cliente. Como os *scores* eram grande parte da razão do *lift* de 27 pontos do K-S, decidimos rodar os modelos novamente, mas escolhendo apenas um dos *scores*, o que tivesse mais impacto positivo no K-S. Sabemos que a compra de *scores* de crédito é cara, e qualquer redução de custo para empresas de crédito pode ser repassada para o consumidor em forma de menores taxas de juros.

1. O gráfico a seguir mostra a relevância (ou peso) de cada variável do modelo. As quatro primeiras variáveis são os *scores* de crédito de *bureaux*, ou seja, as variáveis que mais influenciam o K-S do modelo.

Figura 2



- Rodamos os modelos com restrição de usar apenas um dos quatro *scores* de crédito e tivemos os seguintes resultados usando o método LR (logistic regression).

Figura 3

	Training* (ks)	Test (ks)	Training* (auc)	Test (auc)
All scores included	0.63	0.63	0.89	0.87
Keep both SCORE_A1 & SCORE_A2	0.590	0.578	0.866	0.851
Keep SCORE_A1 only	0.585	0.576	0.861	0.849
Keep SCORE_A2 only	0.576	0.570	0.860	0.843
Keep SCORE_A3 only	0.48	0.45	0.81	0.79

- O resultado foi um K-S de 0.576 que ainda tem um *lift* de 17 pontos comparado com o modelo original do parceiro de 0.40.

3.3 Parte 3: Enriquecer os modelos com dados de Open Banking

É importante lembrar que os modelos acima não incluem dados do Open Banking como variáveis. Em paralelo ao trabalho com o parceiro, a Plataforma de Open Banking para

Todos criou seu primeiro *score* de Open Finance usando dados anonimizados que foram coletados por meio de outros parceiros, fora do escopo deste projeto.

1. Usamos dados de 148.214 usuários com conexões bancárias para criar a primeira versão de *score* de Open Finance Plataforma de Open Banking para Todos.
2. Os dados de cada conexão bancária incluem extratos de conta-corrente, extratos de cartão de crédito e informações de conta como limites de cheque especial e mais.
3. Obtivemos um K-S de 0,682 na primeira versão do *score* Plataforma de Open Banking para Todos:

Figura 4. Resultados dos testes iniciais dos modelos

Modelo	AUC	KS
LGB	0.898	0.692
XGB	0.891	0.691
RF	0.895	0.692
LR	0.874	0.682
SVM	0.543	0.087

Pre vemos que uma combinação de dados do Open Banking com dados dos nossos parceiros vai causar um *lift* de pelo menos 10 pontos de K-S. Isso será um grande avanço para os modelos de crédito que podem ser usados nacionalmente.

3.4 Funcionalidades

Os modelos de crédito da Plataforma de Open Banking para Todos foram desenvolvidos usando as tecnologias mais avançadas de *machine learning* no mundo. A equipe de desenvolvimento, *machine learning* e modelagem fica na China e usa as melhores práticas do mercado asiático para criação dos seus modelos. Segue abaixo um fluxo de coleta de dados, modelagem de crédito e ofertas personalizadas para clientes:

3.4.1 Coleta de dados do Open Banking dos consumidores

O consumidor compartilha seus dados financeiros de Open Banking por meio das APIs do Open Banking ou utilizando as integrações bancárias da Plataforma de Open Banking para Todos. O usuário consente permissão para que a Plataforma de Open Banking para Todos, junto com o parceiro emissor de crédito, receba seus dados financeiros brutos com a finalidade de receber ofertas de crédito por intermédio da modelagem de crédito.

3.4.2 Tratamento dos dados, modelagem de risco de crédito e *score*

A Plataforma de Open Banking para Todos trabalha em cima dos dados fornecidos pelo consumidor. Nesta etapa, a Plataforma de Open Banking para Todos classifica todas as transações, cria variáveis para modelagem de crédito e aplica as variáveis nos seus modelos, gerando um *score* de crédito de cada consumidor.

3.4.3 Oferta e concessão de crédito

O parceiro da Plataforma de Open Banking para Todos recebe o *score* de crédito dos consumidores solicitantes de crédito e aprova ou reprovava o consumidor. Se aprovado, o consumidor recebe uma oferta de crédito do parceiro. Se reprovado, o consumidor é notificado de que sua solicitação foi rejeitada.

4 Escopo do Protótipo no LIFT

O Protótipo LIFT corresponde às duas primeiras etapas do projeto, descritas na Visão Geral deste relatório. As etapas foram:

- Parte 1: Back test com todos os dados providenciados pelo parceiro.
- Parte 2: Back test com todos os dados providenciados pelo parceiro exceto *scores* de crédito de *bureaux*.

Ambas as partes foram completadas com sucesso.

Por falta de tempo, o Protótipo LIFT não corresponde à terceira e última etapa do projeto, que se refere ao enriquecimento de dados dos modelos Plataforma de Open Banking para Todos com dados do Open Banking.

5 Características Inovadoras

A Plataforma de Open Banking para Todos é pioneira em modelos de crédito e *score* de crédito para o Open Finance. Aplicamos técnicas de *machine learning* e inteligência artificial para criar modelos extremamente estáveis (com valores PSI baixos) com performance de K-S de dezenas de pontos melhores que os modelos atuais no mercado, usando dados tradicionais e dados de Open Finance. Isso tudo é possível porque o time de desenvolvimento, *machine learning*, inteligência artificial e modelagem de crédito fica na China, um dos mercados mais inovadores do mundo. Contamos com *experts* que ajudavam bancos e *startups* chinesas a concederem crédito para mais de 100 milhões de pessoas.

Nossos modelos foram adaptados para o mercado brasileiro. Algumas das adaptações incluem tratamentos dos dados de acordo com LGPD, peso relativo de cada variável usada no modelo, enriquecimento dos modelos com dados do mercado (*score* de crédito de *bureaux*, por exemplo) entre outras.

Os modelos de crédito Plataforma de Open Banking para Todos devem melhorar o acesso a crédito e reduzir as taxas de juros para consumidores. Foi exatamente esse tipo de tecnologia que impulsionou a explosão de inclusão financeira na China dos últimos 10 anos.



..... 6 Contribuição para o Sistema Financeiro Nacional

As tecnologias e os modelos desenvolvidos pela Plataforma de Open Banking para Todos vão impactar positivamente o SFN em diversas frentes, principalmente no que se refere à Inclusão Social, Competitividade e Sustentabilidade.

6.1 Inclusão social

Empresas que usarem as tecnologias Plataforma de Open Banking para Todos vão conseguir precificar e conceder crédito para consumidores que eram previamente rejeitados. Usando modelos de crédito Plataforma de Open Banking para Todos, principalmente enriquecidos com dados de Open Finance, as empresas começam a ter uma visão mais real da vida financeira de seus consumidores, reduzindo a dependência de *bureaux* de crédito na hora de aprovar ou reprovar uma solicitação de crédito. Isso, automaticamente, aumenta o número de consumidores com possível acesso a crédito, já que a concessão de crédito não depende mais dos *scores* tradicionais.

6.2 Competitividade

A competitividade de oferta de produtos e serviços financeiros aumenta quando empresas usam as tecnologias Plataforma de Open Banking para Todos e, por consequência, a competitividade dos consumidores (que começam a ter acesso a mais produtos e serviços financeiros mais baratos) também aumenta. Por exemplo, empresas dão mais crédito barato para seus clientes e seus clientes passam a ter um ROI maior na utilização deste crédito, já que o custo das taxas de juros é reduzido.

6.3 Sustentabilidade

O mais importante de tudo isso é que os modelos, as aplicações e o uso das tecnologias desenvolvidas são sustentáveis a longo prazo. Elas podem ser utilizadas por qualquer empresa a qualquer momento e as tecnologias passam por melhorias frequentes.

7 Restrições

Tecnologias de SaaS (*software as a service*) como as nossas são altamente escaláveis, uma vez desenvolvidas. Elas podem ser implementadas em diversas empresas que atuam em segmentos diferentes a baixo custo. Porém, uma das grandes barreiras de implementação dessas tecnologias tem sido a resistência das grandes empresas de trocarem seu sistema atual por um sistema mais novo e escalável. Por exemplo, *stakeholders* de operações de crédito que já são rentáveis em grandes empresas tendem a ser avessos à implementação de novos modelos, justificando que se a operação já funciona, por que mudar? Acreditamos que algumas empresas pioneiras vão criar cases fortes com soluções da Plataforma de Open Banking para Todos e outras empresas seguidoras vão também buscar implementar nossas soluções.

8 Conclusão

Muitos consumidores brasileiros são excluídos do mercado de crédito, e outros pagam uma das maiores taxas de juros do mundo. Um grande causador disso são os modelos de crédito usados atualmente porque eles não foram criados com tecnologia de ponta e acabam tendo baixa previsibilidade.

Os modelos de crédito da Plataforma de Open Banking para Todos foram criados usando as melhores tecnologias de *machine learning*, inteligência artificial e modelagem de crédito. Esses modelos são altamente estáveis e têm altos índices de previsibilidade. Para este projeto, conseguimos demonstrar que nossos modelos podem ter um *lift* de até 27 pontos de K-S quando comparados com modelos atuais do mercado. Modelos mais modernos e mais preditivos ajudam consumidores a terem mais acesso a crédito barato, além de representar um ganho significativo na rentabilidade de empresas credoras.

Apesar de o protótipo desenvolvido neste projeto já ser bastante promissor, a Plataforma de Open Banking para Todos tem como objetivo criar outras tecnologias que vão impactar positivamente o SFN inteiro. Nossas tecnologias devem aumentar a inclusão social, aumentar a competitividade de empresas e consumidores, e as tecnologias são sustentáveis a longo prazo.

