

Uma Abordagem Bayesiana em Modelos de Risco de Crédito

Erick Luciano Floriano Mendes,¹ Adriano Kamimura Suzuki²
ICMC-USP

1 Introdução

O mercado de crédito no Brasil é constituído por instituições financeiras e não-financeiras que prestam serviços de intermediação de recursos para indivíduos e empresas. O acesso ao crédito constitui-se em uma ferramenta fundamental para que indivíduos e empresas possam satisfazer sua capacidade produtiva e, com isso, estimular o crescimento econômico. Sendo assim, o crédito possui hoje um importante papel na economia, uma vez que é essencial ao financiamento do consumo das famílias e do investimento dos setores produtivos, o que possibilita um aperfeiçoamento em aspectos tecnológicos, de estrutura e a geração de empregos, ocasionando a melhoria de vida de diversas pessoas e do país como um todo [1].

É também importante entender a função dos *bureaus* de crédito e como eles auxiliam esse mercado de uma maneira geral. Essas empresas coletam informações sobre o histórico de crédito do consumidor no mercado. Desta forma, o consumidor pode se beneficiar com bom comportamento de crédito e as instituições financeiras minimizam o risco, evitando que as pessoas assumam mais dívidas do que podem pagar. A partir da lei do Castro Positivo, de agosto de 2019, informações positivas sobre o comportamento de crédito dos consumidores passaram a ser disponibilizadas para essas empresas de maneira obrigatória, o que beneficia todo o mercado de crédito: com mais informações disponíveis, melhor a assertividade no momento da concessão do crédito, o que pode gerar mais pessoas com acesso a crédito e menores taxas de juros praticadas pelo mercado. Além disso, outras iniciativas estão sendo desenvolvidas e já aplicadas com o mesmo objetivo de compartilhamento de informações entre instituições financeiras com a expectativa de tornar o mercado mais competitivo e com menores riscos de inadimplência. É o caso do Sistema Financeiro Aberto (mais conhecido como *Open Finance*), em que o cliente pode optar por compartilhar seu histórico de pagamentos e informações entre um banco e outro para poder escolher qual a oferta de crédito mais interessante. Todas essas informações tendem a ser utilizadas principalmente como variáveis preditoras de modelos estatísticos que têm o objetivo de prever a probabilidade de inadimplência.

¹ericklfmendes@alumni.usp.br

²suzuki@icmc.usp.br

2 Materiais e Métodos

No desenvolvimento de modelos estatísticos para concessão de crédito é bastante usual a utilização de modelos de regressão logística, considerando estimação dos parâmetros via inferência clássica, com estimadores de máxima verossimilhança. Além disso, nos últimos anos houve um crescimento na aplicação de modelos de Aprendizado de Máquina para discriminar bons e maus pagadores, mas que não serão avaliados nesta pesquisa. Para a construção dos modelos utiliza-se a performance observada para os consumidores que tiveram propostas de crédito aprovadas, ou seja, considera-se o comportamento de atraso do público apenas na empresa que concedeu o crédito. Entende-se que este é o dado mais seguro do comportamento de crédito especificamente para uma determinada empresa, portanto são os dados mais recomendáveis para esse público.

O objetivo desta pesquisa é apresentar alternativas interessantes para desenvolvimento de modelos de crédito, utilizando ferramentas de inferência Bayesiana e avaliar os possíveis ganhos com sua utilização. Foram escolhidos os testes considerando prioris não-informativas com distribuições Normais e Uniformes, além do uso das transformações potentes e reversa-potente (em inglês, *power link* e *reversal power link*, respectivamente) [2]. Ao final, foram comparadas as métricas de Kolmogorov-Smirnov (KS) e Área sob a Curva ROC (AUC-ROC) entre todos os modelos encontrados e realizadas as devidas conclusões.

2.1 Base de Dados

A base de dados utilizada para o desenvolvimento da pesquisa é composta por 99.766 registros, possuindo 94 variáveis preditoras. Cada registro representa uma proposta aprovada de uma pessoa física para um produto de crédito de um banco digital brasileiro. Essas propostas referem-se ao período entre Junho de 2019 e Dezembro de 2019. Todos os registros da base de dados tiveram suas propostas de crédito aprovadas, o que significa que o banco digital consegue mensurar se o cliente foi um bom pagador com as informações internas do banco.

As variáveis preditoras são variáveis provenientes de *bureaus* de crédito, trazendo informações cadastrais, geográficas, sobre o histórico de débitos e de consultas ao mercado de crédito para cada cliente. Todas as variáveis possuem preenchimento, ou seja, não foi necessário a realização de nenhum tratamento para dados faltantes. A Tabela 1 apresenta a quantidade de variáveis preditoras por cada tipo identificado na base de dados.

Tabela 1: Tipos de Variáveis Preditoras

Tipo	Quantidade	Quantidade (%)
Cadastral	3	3,2%
Consulta	8	8,5%
Débito	39	41,5%
Geográfica	44	46,8%
TOTAL	94	100,0%

3 Aplicações

Aqui serão apresentados os resultados da metodologias desenvolvidas e ilustrados pela Figura 1. Inicialmente foi realizado o Pré-Processamento da base de dados, filtrando variáveis com concentração de volume superior à 90% em um único valor e, quando duas variáveis apresentavam correlação de pearson superior à $||0.7||$, retirava-se a com maior concentração em um único valor. Após o Pré-Processamento, iniciou-se a realização da Modelagem via estimação Clássica, dando continuidade com a Regressão Logística Bayesiana com prioris Não-informativas e também com transformações *Power Link* finalizando com as conclusões do estudo.

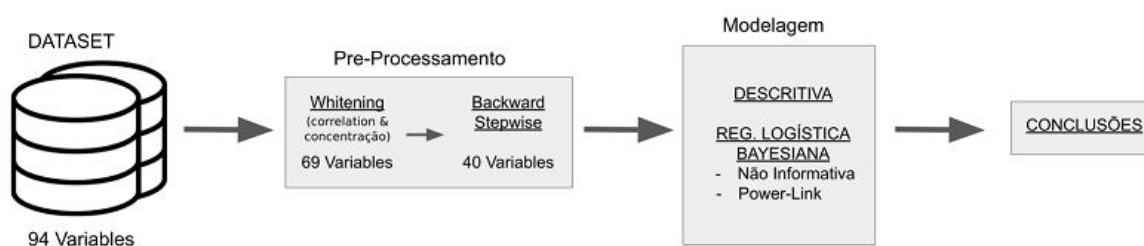


Figura 1: Fluxo das análises desenvolvidas

Fonte: Elaborada pelo autor.

A Tabela 2 traz os parâmetros utilizados em cada um dos testes que apresentaram convergência para uma série estacionária, segundo o teste de Geweke. Vale destacar que as prioris apresentadas se referem sempre às distribuições Uniforme e Normal. Foram testadas diversas combinações de prioris, diferentes valores para λ , *Power Links* e quantidade de iterações, porém grande parte dos testes com iterações menores que 10.000 não convergiram para séries estacionárias no momento da estimação dos parâmetros. Sendo assim, esses resultados não foram considerados para as análises finais. Entre todos os testes realizados que convergiram apenas um, o modelo 13 apresentou resultados considerando uma priori também para o parâmetro λ . Os demais testes consideraram valores arbitrários para λ , variando de 0,2 até 20. Com relação ao número de iterações para cada teste, cada combinação dos demais parâmetros foi testada com 1.000, 5.000 e 10.000 iterações.

Tabela 2: Descrição dos testes apresentados

Teste	Abreviação	Priori	Transformação na F. de Ligação	λ	Número de Iterações	KS (Validação)	AUC (Validação)
modelo 1	mod1	Normal(0,10.000)	<i>Power Link</i>	0,2	10.000	26,8%	32,5%
modelo 2	mod2	Normal(0,10.000)	<i>Power Link</i>	0,2	5.000	26,6%	32,5%
modelo 3	mod3	Normal(0,10.000)	<i>Power Link</i>	0,4	5.000	26,5%	32,5%
modelo 4	mod4	Normal(0,10.000)	<i>Power Link</i>	0,4	10.000	26,5%	32,5%
modelo 5	mod5	Uniforme(-1,1)	<i>Reversal Power Link</i>	0,2	10.000	26,4%	32,5%
modelo 6	mod6	Uniforme(-1,1)	<i>Power Link</i>	0,6	10.000	26,2%	32,6%
modelo 7	mod7	Normal(0,10.000)	<i>Power Link</i>	0,6	5.000	26,3%	32,6%
modelo 8	mod8	Normal(0,10.000)	<i>Power Link</i>	0,8	10.000	26,3%	32,6%
Modelo Clássico	Mod. Classico	modelo Classico	-	-	-	26,1%	32,6%
modelo 10	mod10	Uniforme(-1,1)	<i>Reversal Power Link</i>	1	10.000	26,0%	32,7%
modelo 11	mod11	Uniforme(-1,1)	<i>Reversal Power Link</i>	0,6	10.000	25,8%	32,7%
modelo 12	mod12	Normal(0,10.000)	<i>Reversal Power Link</i>	0,2	10.000	25,5%	33,0%
modelo 13	mod13	Normal(0,10.000)	<i>Power Link</i>	$\Gamma(1,1)$	10.000	26,0%	32,7%
modelo 14	mod14	Normal(0,100)	-	-	10.000	22,1%	32,8%
modelo 15	mod15	Uniforme(-1,1)	-	-	10.000	25,8%	32,7%
modelo 16	mod16	Uniforme(-2,2)	-	-	10.000	19,9%	32,7%

4 Conclusões

A aplicação das técnicas Bayesianas em modelos de risco de crédito apresentaram resultados bastante próximos aos da inferência clássica em termos de discriminação de bons e maus pagadores. Ao considerar que o patamar de discriminação dos modelos foi mantido com os modelos Bayesianos propostos, agora cada parâmetro do modelo possui uma distribuição própria e conhecida, o que traz ganhos no acompanhamento do modelo e das variáveis, possibilitando definir intervalos de credibilidade Bayesianos, e realizar demais pesquisas e estudos. Portanto, pode-se concluir que não se deve descartar a utilização de modelos Bayesianos no contexto de risco de crédito.

Outros estudos relacionados podem ser desenvolvidos a partir dos testes apresentados neste trabalho, como a utilização de *Power Priors*, onde utiliza-se o conhecimento prévio de um estudo anterior como priori informativa (os resultados encontrados neste trabalho poderiam ser utilizados como conhecimento prévio acerca do tema), outras prioris, informativas ou não, poderiam ser avaliadas também. Além disso, a aplicação de diferentes transformações *Power Link* na função de ligação da Regressão Logística e a aplicação de outras técnicas de modelagem também podem ser avaliadas em estudos futuros. Esse trabalho teve como intuito desafiar a técnica usualmente aplicada no mercado, por isso apenas modelos de Regressão Logística foram avaliados.

Referências

- [1] DIEESE. *A evolução do crédito na economia brasileira 2008-2013*. DIEESE - Departamento Intersindical de Estatística e Estudos Socioeconômicos, SP, 2014.
- [2] Bazán, J. L. and Torres-Avilés, F. and Suzuki, A. K., and Louzada, F. Power and reversal power links for binary regressions: An application for motor insurance policyholders, *Applied Stochastic Models in Business and Industry*, 33(1), 22-34, 2017.