

**UNIVERSIDADE FEDERAL DE SÃO PAULO
ESCOLA PAULISTA DE POLÍTICA, ECONOMIA E NEGÓCIOS**

LUCAS TAVARES MONTEIRO

A EVOLUÇÃO DOS SCORES DE CRÉDITO.

Osasco

2021

**UNIVERSIDADE FEDERAL DE SÃO PAULO
ESCOLA PAULISTA DE POLÍTICA, ECONOMIA E NEGÓCIOS.**

LUCAS TAVARES MONTEIRO

A EVOLUÇÃO DOS SCORES DE CRÉDITO.

Trabalho de Conclusão de Curso apresentado
ao curso de Bacharelado em Ciências Atuariais
da Universidade Federal de São Paulo.

Orientador: Prof. Humberto Gallucci

Osasco

2021

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Ficha catalográfica elaborada pela Biblioteca Unifesp Osasco
e Departamento de Tecnologia da Informação Unifesp Osasco, com os dados fornecidos pelo(a) a

M775e MONTEIRO, Lucas Tavares
A evolução dos scores de crédito / Lucas Tavares Monteiro.
- 2021.
30 f. :il.

Trabalho de conclusão de curso (Ciências Atuariais) -
Universidade Federal de São Paulo - Escola Paulista de Política,
Economia e Negócios, Osasco, 2021.
Orientador: Prof. Dr. Humberto Gallucci Netto.

1. Análise de crédito. 2. Credit scoring. 3. Técnicas
estatísticas. 4. Técnicas computacionais. I. Gallucci Netto, Prof.
Dr. Humberto , II. TCC - Unifesp/EPPEN. III. Título.

CDD: 658.88

Cada vez que você faz uma opção está transformando sua essência
em alguma coisa um pouco diferente do que era antes.
C. S. Lewis

AGRADECIMENTOS

Primeiramente, agradeço a Deus por me dar a paz e a força que me ajudaram a chegar até aqui.

Agradeço aos meus familiares, tanto os de sangue quanto os que me incluíram em seus lares, por toda a ajuda, companheirismo, ensinamento e por não terem me deixado desistir.

Agradeço aos meus companheiros de caminhada que me ensinaram muito mais do que poderia imaginar.

RESUMO

A pontuação de crédito, “credit scoring”, transformou-se em uma ferramenta de vital importância junto as instituições concessionárias de crédito na busca por avaliar e gerenciar o risco de crédito. O objetivo deste trabalho foi de realizar uma revisão bibliográfica da literatura sobre a evolução das técnicas que veem compondo os scores, diretamente em artigos, principalmente os voltados a área de finanças. Os resultados mostram que a técnica de Regressão Logística (RL) ainda é a mais utilizada no mercado, pois a sua forma consegue ser facilmente explicada, diferentemente das novas tecnologias, que apesar de começarem a ganhar força, por ter uma discriminação entre bons e maus clientes melhor que a RL, perde na parte de elucidar como o modelo chegou a determinada resposta.

Palavras-Chave: Análise de Crédito, Credit Scoring, Técnicas Estatísticas e Computacionais.

ABSTRACT

Credit scoring has become a vitally important tool with credit granting institutions in the quest to assess and manage credit risk. The objective of this work was to carry out a bibliographic review of the literature on the evolution of the techniques that they see composing the scores, directly in articles, mainly those focused on the finance area. The results show that the Logistic Regression (RL) technique is still the most used in the market, because its form can be easily explained, unlike the new technologies, which despite starting to gain strength, by having a discrimination between good and bad customers better than RL, it loses in elucidating how the model arrived at a given answer. Keywords: Credit Analysis, Credit Scoring, Statistical and Computational Techniques.

LISTA DE FIGURAS

Figura 1 - Evolução do Estoque de Crédito	18
Figura 2 - Processo de análise por meio de Credit Scoring	19
Figura 3 - Definição do Ponto de Corte.....	21
Figura 4 - Definição do Ponto de Corte II.....	21

SUMÁRIO

1. INTRODUÇÃO	10
1.1 DESCRIÇÃO DO PROBLEMA	10
1.2 JUSTIFICATIVA	10
1.3 OBJETIVOS	11
1.3.1 GERAL:	11
1.3.2 ESPECÍFICOS:	11
2. REFERENCIAL TEÓRICO	12
2.1 CRÉDITO.....	13
2.1.1 OS C'S DO CRÉDITO.....	15
2.2 MERCADO CREDITÍCIO NO BRASIL.....	17
2.3 CARACTERÍSTICAS DO CREDIT SCORING	18
2.4 DEFINIÇÃO DE BOM E MAU CLIENTE.....	20
2.5 DEFINIÇÃO DE PONTO DE CORTE	20
3. METODOLOGIA	23
3.1 MODELOS.....	23
3.1.1 - REGRESSÃO LOGÍSTICA	23
3.1.2 - METODOLOGIAS ENSEMBLE	24
3.1.4 - SUPPORT VECTOR MACHINE	26
3.1.5 – MODELOS HÍBRIDOS	26
4. CONCLUSÃO.....	28
5. REFERÊNCIAS BIBLIOGRÁFICAS.....	29

1. INTRODUÇÃO

1.1 DESCRIÇÃO DO PROBLEMA

O aumento da população em busca de crédito no mercado e consequentemente o aumento das taxas de inadimplência, houve a necessidade da criação de algum método que pudesse separar os prováveis bons dos maus pagadores, entretanto, para se fazer uma separação é necessário capturar dados passados deste determinado cliente e através do seu comportamento prever qual será a consequência no futuro.

Com isso, empresas especializadas apareceram no mercado, utilizando métodos estatísticos que possam prever o padrão de consumo de determinado consumidor, tendo uma base de dados construída através de um compartilhamento de informações das demais empresas que compõem o mercado. (Carvalho, 1996) diz que os *credit bureau* nada mais são do que empresas especializadas em colher, armazenar e vender informações de pessoas jurídicas e físicas.

Os modelos de *Credit Scoring* são sistemas que atribuem pontuações às variáveis de decisão de crédito de um proponente, mediante a aplicação de técnicas estatísticas. Esses modelos visam a segregação de características que permitam distinguir os bons dos maus créditos (LEWIS, 1992).

Entretanto os métodos estatísticos que compõem um score de crédito não são mais segredos e o que vêm impedindo as empresas de construir os próprios scores são a falta de bases de dados construída com informações provenientes de todo o mercado.

Diante disso, surgiu o questionamento quanto as novas ferramentas analíticas podem influenciar a composição de um score de crédito ao obterem um conjunto variado de informações.

1.2 JUSTIFICATIVA

Fazer o maior volume de negócios possível atingindo o máximo de pessoas resguardando o nível de inadimplência de uma empresa, ou seja,

conseguindo fazer negócio com apenas aqueles que apresentam um risco aceitável para a empresa, têm sido uma missão para gestores e donos de empresas. Com essa necessidade identificada no mercado, *bureaus* de crédito têm através de uma grande base dados criar soluções analíticas que otimizem o momento da tomada de decisão.

Os scores de crédito são ferramentas que colaboram para o momento da decisão, mas diante das evoluções tecnológicas que vêm ocorrendo e os novos métodos estatísticos e computacionais que podem ser utilizados empresas que possuam esses recursos podem abrir mão de dados dos *bureaus* de crédito para criarem os seus próprios scores utilizando os seus dados internos e as informações advindas do *big data* e *machine learning*. Sendo assim, espera-se encontrar novos caminhos para a criação de *credit scores* não utilizando informações de *bureaus* de crédito.

1.3 OBJETIVOS

1.3.1 GERAL:

O presente estudo tem como objetivo avaliar o impacto das novas tecnologias na construção e formação de score por empresas que possuam acesso as novas tecnologias computacionais e estatística.

1.3.2 ESPECÍFICOS:

- Realização de revisão bibliográfica sobre a indústria de scores, métodos e as aplicações destas ferramentas no mercado;
- Mensurar o quanto as novas fontes de tecnologia podem prejudicar o monopólio de informações dos *bureaus* de crédito.

2. REFERENCIAL TEÓRICO

Para Santos (2015) a definição de pontuação de score de crédito é:

A pontuação de crédito é um instrumento estatístico desenvolvido para que o analista avalie a probabilidade de que determinado cliente venha a tornar-se inadimplente no futuro. Trata-se de um modelo de avaliação do crédito baseado em uma fórmula estatística desenvolvida com base em dados cadastrais, financeiros, patrimoniais e de idoneidade dos clientes. (Santos, 2015)

Existe uma grande importância na composição das bases que irão construir um score de crédito. Para Thomas (2002) Os modelos estatísticos de *credit score* refletem a qualidade das amostras com as quais foram desenvolvidas, prevendo o comportamento de novos proponentes da população estudada, baseando-se no desempenho dos clientes anteriores utilizados na amostra de desenvolvimento.

Segundo Vasconcellos (2002) as bases para a composição de um score de crédito feito de maneira tradicional trata-se de uma amostra aos quais já tiveram crédito concedido no passado que fazem parte da carteira da empresa, incluindo os dados desse indivíduo e da operação de crédito e a classificação se esse indivíduo foi um bom ou mau de acordo com os atrasos nos pagamentos.

Espin-Garcia (2003) relata a importância de as variáveis não serem correlacionadas entre si no processo de construção de um score e a importância de se ter um conjunto amplo de variáveis que abordem os mais variáveis temas em relação ao consumidor.

Já a “Big data” é uma ferramenta utilizada para encontrar intuições e informações em uma quantidade astronômica de dados (CREPALDE, 2017). Segundo a IBM (2012) são gerados cerca de 2,5 quintilhões de dados por dia e 90% da quantidade de dados existente hoje foram criadas nos últimos dois anos.

Para a IBM (2012) o processo de um big data abrange três aspectos: Volume, velocidade e variedade. Sendo assim, além do grande volume de dados que essa ferramenta pode propiciar, ela também apresenta uma alta velocidade que permite tratar esses dados e através de análises é possível gerar novos *insights*.

Segundo Caoutte, Altman e Narayanan (1998) existem várias aplicações para o *credit score*, além da aprovação de crédito ele pode ser utilizado como estratégia de cobrança, precificação de crédito e classificação de rating de crédito.

O processo de desenvolvimento de um modelo de crédito envolve várias etapas, entre as quais Planejamento Amostral, Determinação da Pontuação de Escore e Validação e Comparação de Modelos (DINIZ, 2013).

Segundo Moraes (2012) a utilização de redes neurais e *machine learning*, métodos propiciados que podem ser usados em big data apresentam melhores resultados e ganhos se comparados com as técnicas tradicionais como a regressão logística.

Também Moraes (2012) apresenta pontos importantes sobre os scores *behavior* que em sua composição levam informações internas da operação da empresa e como essas informações levam há uma maior discriminação de bons e maus dado o comportamento passado.

Mas é importante salientar que quanto mais IA é utilizado nos modelos de score, mais complicado fica estabelecer um passo a passo de como algoritmo chegou naquele resultado e qual é o peso de cada variável no processo de construção do SCORE.

2.1 CRÉDITO

Para Schrickel (2000), crédito “é todo ato de vontade ou disposição de alguém de destacar ou ceder, temporariamente, parte de seu patrimônio a um terceiro”.

Sendo assim, a palavra crédito origina-se do latim *creditu*, que significa eu acredito ou confio. Já que essa confiança não é dada de maneira unilateral, mas ocorrendo tanto por parte do vendedor, ao ceder a sua mercadoria na promessa que o adquirente possui formas de honrar o seu compromisso fundado no futuro.

Presentemente, no mercado financeiro, nos comércios e em tantos outros estabelecimentos, o crédito tem assumido um lugar de facilitador na vida das pessoas.

Possibilitando o aumento do poder de compra do cliente, como também um incremento no número de vendas para o ofertante.

Entretanto, um ponto desfavorável tem sido a necessidade de fornecer crédito em grande escala, o que tem demandado um elevado número de análises para suprir essa concessão em massa (PACHECO, 2012).

Ademais, fora atender um demasiado número de solicitação de concessões de créditos. São necessários criar critérios para o processo de obtenção de crédito, dado que ele é feito na confiança de que a outra parte possuirá condições de honrar o compromisso no tempo futuro. Sendo assim, utilizando como exemplo uma solicitação de crédito em uma instituição bancária, o cliente pessoa física deve passar por uma avaliação, que analisará os seus aspectos pessoais como caráter e capacidade e também aspectos financeiros como capital e condições de honrar o compromisso. Visando assim mitigar os riscos para a organização que está concedendo o crédito.

MIOLA (2013), elenca três etapas que devem ser consideradas para a uma análise de crédito.

- Análise Retrospectiva: avaliação do desempenho histórico do solicitante ao crédito, analisando todos os seus débitos passados. Com este processo é possível identificar o comportamento de pagamento do tomador, assim como elencar na atual condição os motivos que possam dificultar o pagamento da dívida.

- Análise de Tendências: projeção de como estará a condição futura do tomador do crédito, visando analisar o nível de endividamento suportável e quão oneroso será o crédito que se espera obter; e

- Capacidade Creditícia: a partir do grau de risco que o tomador apresenta e a projeção do seu nível de endividamento futuro, avaliar a capacidade creditícia do tomador, ou seja, qual a quantia de capital que ele poderá obter junto ao credor.

Segundo Santos (2009), as fases para o processo de análise de crédito de pessoa física são:

- Análise cadastral;
- Análise de idoneidade;
- Análise financeira;

- Análise de relacionamento;
- Análise patrimonial;
- Análise de sensibilidade;
- Análise do negócio;

2.1.1 OS C'S DO CRÉDITO

Uma das ferramentas mais utilizadas para análise de crédito no que tange as questões subjetivas do processo são os C's do crédito. Segundo Saunders (2000), esta ferramenta visa fornecer informações sobre o cliente, objetivando ser um facilitador para a tomada de decisão.

Sendo assim, os C's do crédito são grandes balizadores para os modelos de análise de crédito.”

Os C's do crédito são o caráter, capacidade, capital, condições, colateral e conglomerado. Suas características são apresentadas a seguir

Caráter

Segundo Souza e Chaia (2000), caráter é a vontade do cliente em quitar sua dívida., sendo assim, trata-se da vontade que o próprio tomador de crédito tem de pagar as suas contas, o que traduz o seu senso moral, índole e ética.

Santos (2000, p. 145) define caráter como sendo a “idoneidade atual e histórica do cliente na amortização de empréstimos.” Pode-se entender como a intenção ou determinação do cliente em honrar ou não seus compromissos. (BLATT, 1999).

Uma forma de averiguar essa característica que pode ser entendida como subjetiva é obter informações junto a bancos e cartórios de títulos e protestos.

Essas informações demonstram a conduta do cliente, no que diz respeito ao honrar os seus pagamentos com pontualidade e constância. Estes dados compõem o parâmetro de comportamento histórico, pois a pontualidade da liquidação dos compromissos pode ser fator importante no negócio do cliente, não exprimindo sua real vontade de pagar.

Capacidade

A saúde financeira de uma empresa está diretamente ligada à capacidade de um bom gerenciamento dos negócios. No entanto, seu desempenho não está ligado somente ao demonstrativo de resultados, mas a outros fatores, tais como participação relativa no mercado, níveis de qualidade e produtividade, e grau de satisfação e lealdade dos clientes (FERREIRA, 2010). Estes são fatores difíceis de mensurar sem uma profunda pesquisa na empresa e a outras fontes de informações. Portanto, uma análise inicial do currículo dos administradores pode dar uma breve orientação sobre suas habilidades técnicas no setor em que atuam e em relação ao grau de especialização da produção

Se parceria comercial estiver presente em ambas as organizações, reduz-se o risco de insolvência. Segundo Silva (1998), a capacidade deve estar relacionada aos fatores que contribuem para a empresa ser competente e competitiva, sendo que isto facilitará suas condições de ser capaz de pagar suas dívidas.

Capital

Representa a Saúde econômico-financeira do avaliado, Neste item, em geral, são avaliador patrimônios pessoais, no caso de pessoa física, e balanço patrimonial, para pessoas jurídicas.

De forma mais detalhada Berni (1999, p.104) esclarece que,

algumas empresas não podem ser medidas pelo faturamento, mas o capital, o patrimônio e os recursos próprios podem ser um ótimo determinante para o crédito. A forma como a empresa reaproveita seus lucros, o conseqüente aumento de capital, a constituição de reservas, podem auxiliar a valorização do crédito

Condições

Representa os fatores econômicos atuais e características dos indivíduos que podem aumentar ou diminuir a capacidade de pagamento do devedor.

As condições são os fatores externos que influenciam o bom andamento das organizações no que rege o seu planejamento.

Segundo Santos (2000, p.47),

[...] a atenção nessa informação é de extrema

importância para a determinação do risco de crédito, uma vez que, dependendo da gravidade do fator sistemático – exemplos típicos de situação recessiva com aumento nas taxas de desemprego - o banco poderá enfrentar grandes dificuldades para receber os créditos.

Colateral

Pode considerar como as garantias que o proponente apresenta para dar maior segurança nas operações.

(a) o risco apresentado pela empresa e pela operação; (b) a praticidade em sua constituição; (c) os custos incorridos para sua constituição; (d) o valor da garantia em relação ao valor da dívida, isto é, deve ser suficiente para cobrir principalmente encargos e despesas eventuais; (e) a depreciabilidade; (f) o controle do credor sobre a própria garantia, e (g) a liquidez, ou seja, a facilidade com que a garantia pode ser convertida em dinheiro para liquidar a dívida (SILVA, 2000, p. 99).

2.2 MERCADO CREDITÍCIO NO BRASIL

Com a adoção do plano real na década de 90, os bancos começaram a aumentar a sua oferta de crédito a altas taxas de juros. Com o final da hiperinflação que assolava o país, foi possível realizar estimativas factíveis de um futuro plausível, não tendo mais a incerteza de valores ao longo prazo. Com a apresentação deste novo cenário no país, empresas, e indivíduos começaram a se planejar financeiramente e aumentar a demanda por crédito na economia brasileira.

Porém, o Plano Real também teve alguns efeitos colaterais, como o aumento da dívida pública, o que tornou o Governo Federal o maior consumidor de crédito do sistema financeiro nacional. O gráfico abaixo mostra a evolução dos volumes de concessões de crédito ao longo do tempo desde o início do Plano Real (dados do Banco Central do Brasil).

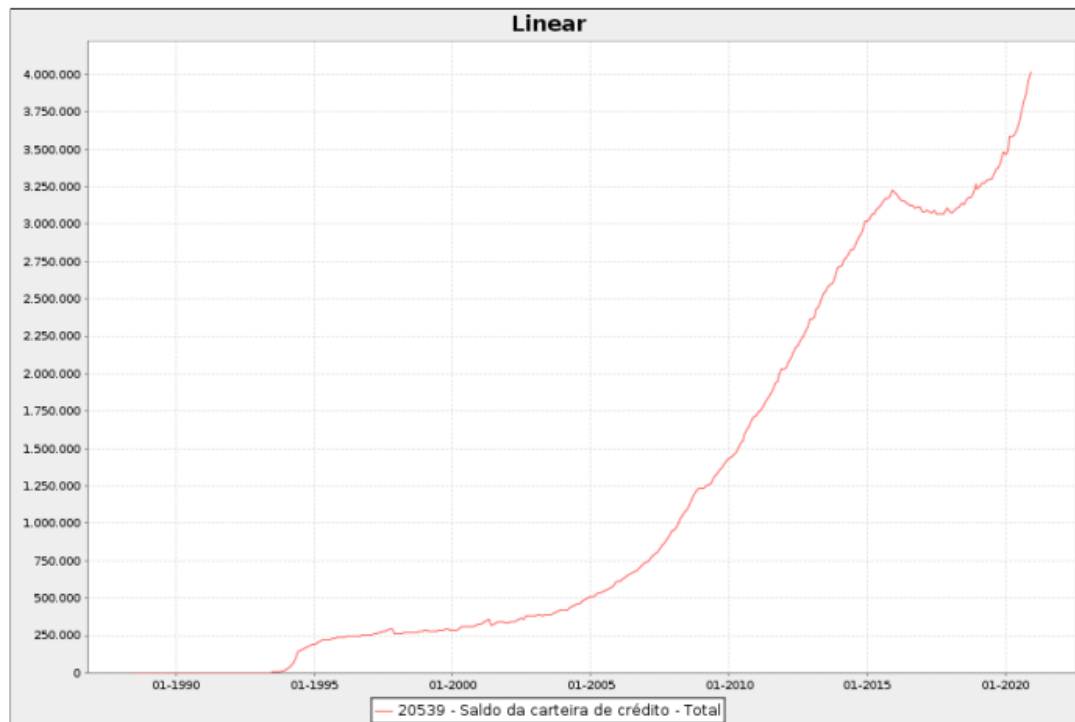


Figura 1 - Evolução do Estoque de Crédito

Fonte: SIGS – Banco Central

Pode-se ver que o Brasil teve um grande aumento no mercado de crédito como um todo, resultado deste novo cenário macroeconômico e uma maior demanda por crédito adormecida. Entretanto, este aumento repentino aliado com uma falta de experiência de credores e devedores em administrar crédito, acabou-se resultando em aumento no número de inadimplência. Os bancos brasileiros não tinham a expertise em conceder crédito no novo cenário e cresceram suas carteiras indiscriminadamente, a população não sabia lidar com a ideia de um aumento no poder aquisitivo, sendo assim, o número de inadimplência cresceu vertiginosamente no mesmo período.

2.3 CARACTERÍSTICAS DO CREDIT SCORING

O processo de concessão de crédito ao proponente é uma decisão a ser tomada em um cenário de incertezas. Se o credor tiver informações que possam auxiliar a mensurar o risco e as chances de o cliente onerar os pagamentos, terá maior convicção na decisão de crédito e favorecerá a redução dos índices de inadimplência.

O Credit Scoring é um processo por meio do qual a informação sobre uma empresa, ou pessoa física, que deseja obter crédito é convertida em um valor numérico, o “score”. (CORADI, 2002).

A figura a seguir demonstra o processo decisório em um modelo de negócios que possui o score como respaldo para a tomada de decisão.

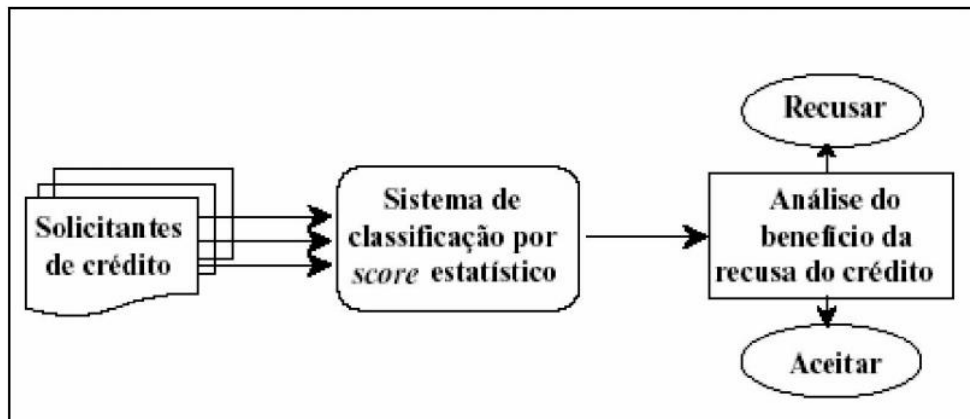


Figura 2 - Processo de análise por meio de Credit Scoring

Fonte: Sousa e Chaia (2000, p.21)

Entretanto, como CHAI (2000) menciona, a ferramenta de score, por ser um processo estatístico, não exclui a possibilidade de enquadrar um pagador de forma errada como um “bom” ou um “mau”.

As principais características do modelo Credit Scoring são descritas a seguir, conforme determina Sicsú (2010, p. 81):

- é resistente, mesmo com pequenas oscilações no cenário socioeconômico, ele deve manter-se válido;
- considera diferentes tipos de informações para avaliar o risco do cliente;
- considera todas as informações disponíveis, ainda que não sejam tradicionalmente utilizadas pelos analistas de crédito;
- é defensável, ou seja, não deve basear-se em variáveis que sejam proibidas por lei ou que firam valores éticos tanto da sociedade quanto da empresa credora; e
- respeita a cultura do credor, tanto no que diz respeito às informações utilizadas quanto a restrições de concessão de crédito definidas por sua política de crédito.

O Score abarca inúmeras informações do proponente, permitindo assim que através de um número o credor possa tomar ou não uma decisão de aceitar a concessão do crédito.

2.4 Definição de bom e mau cliente

As definições para bom e para mau cliente, podem mudar em cada instituição. Logo, a definição a ser considerada bom para uma, pode não ter o mesmo significado para a outra.

Posto isto, Sicsú (2010, p. 18) escreve que “a definição do bom/mau/intermediário cliente deve ser clara e objetiva, representando assim a finalidade para o qual será desenvolvido o modelo.”

Saunders (2000) salienta um ponto em relação à classificação oriunda da pontuação entregue no score, esta pode ser interpretada para alguns, como apenas a probabilidade de inadimplência. Mas, também é possível utilizar a pontuação para classificar em grupos, identificando se o tomador pertence a um grupo bom ou ruim ou até mesmo grupos intermediários.

Santos (2000) esclarece que, os tomadores que atingem pontuações inferiores ao ponto de corte não representam necessariamente uma concessão de crédito de alto risco. Essa pontuação serve apenas para segregar as propostas que chegam para análise e a partir deste momento, a empresa pode dar um parecer mais seguro sobre a operação.

2.5 DEFINIÇÃO DE PONTO DE CORTE

Em uma operação de concessão de crédito, é importante analisar que nem todas as solicitações que chegam é interessante aprovar, pois é interessante ter um equilíbrio entre o risco e o retorno da operação, sendo assim devem ser aprovadas apenas as solicitações que possuem valor esperado ou resultado esperado positivo. Logo, o ponto de corte é definido como o ponto de lucro máximo ou ponto de lucro marginal zero.

Utilizando uma equação de risco, é possível ordenar as operações pelo nível de risco que cada cliente apresenta e calcular a partir dos dados, um resultado ideal de cada um dos níveis. Sendo assim, é esperado que os níveis que possuam um risco maior, previstos pelo score, apresentem um nível de inadimplência maior, gerando um menor lucro líquido. Portanto, o processo de para a definição do ponto objetiva encontrar o ponto que o risco abarcado ainda apresente valores de resultado positivos.

A figura 3 exemplifica isso ao cravar uma inadimplência suportada de 18% e ver qual é o ponto em que o risco está para começar a aceitar as operações.

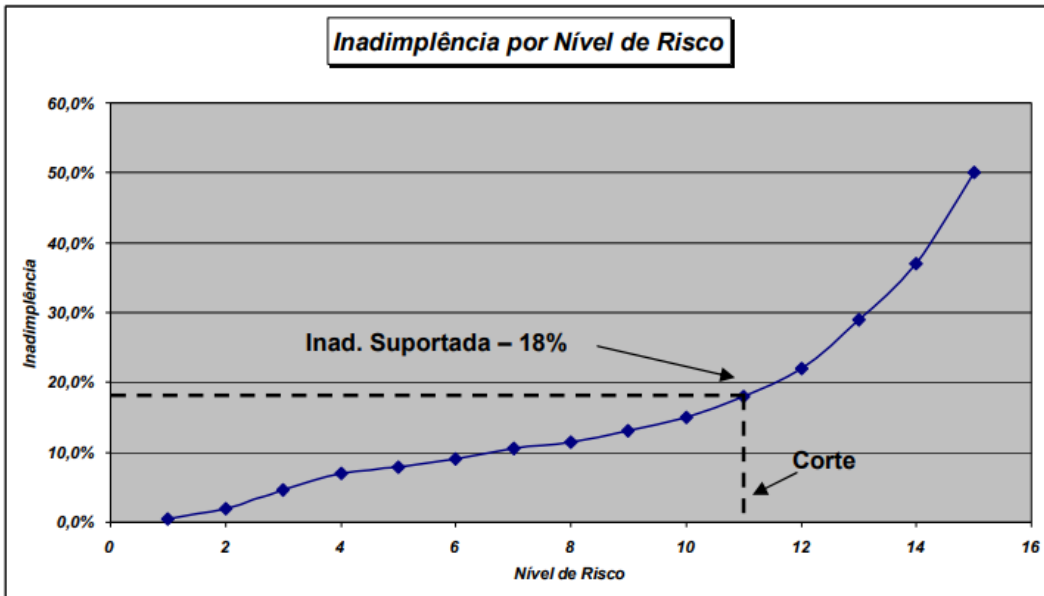


Figura 3 - Definição do Ponto de Corte

Fonte: Vasconcellos (2010)

Já a figura 4 mostra que ao adotar uma política de corte é possível identificar o seu ponto de lucro máximo.

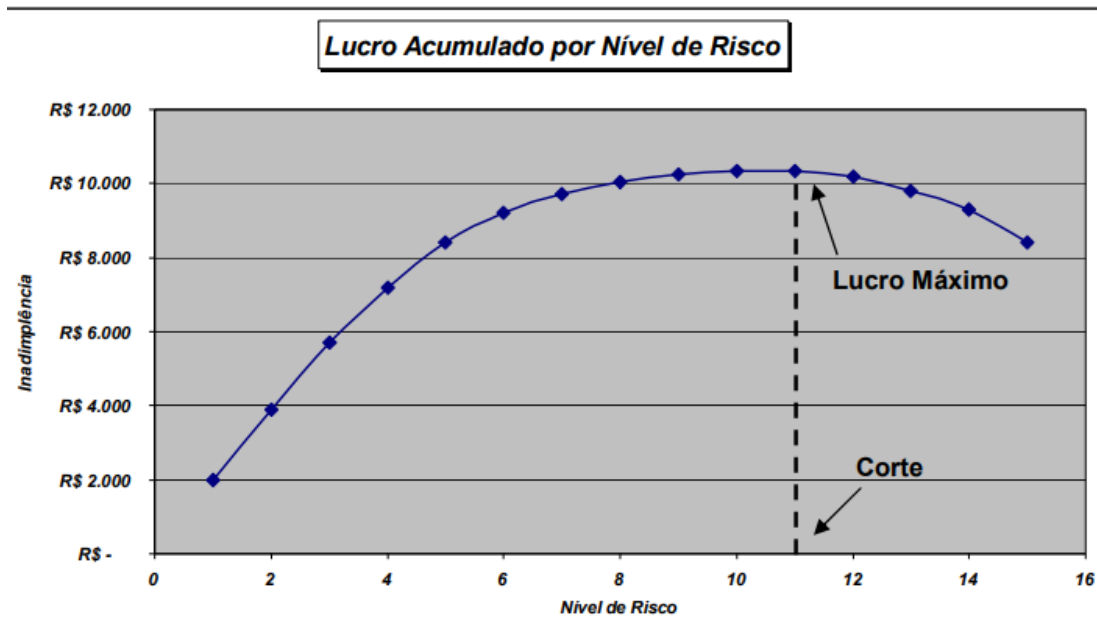


Figura 4 - Definição do Ponto de Corte II

Sendo assim o ponto de corte nada mais é do que um critério político adotado pelo concesso de crédito, a partir de sua própria experiência para o que seja “bom”

ou “mau” pagador. Todavia, Santos (2006) alerta para a delimitação do ponto de corte. Se, muito alto pode acarretar a perda de negócio com o cliente, porém se muito baixo, pode considerar um cliente ruim em um cliente bom, o que elevará a inadimplência junto à instituição.

VASCONCELLOS (2010) Deixa claro a importância de perceber e ressaltar que os critérios se equivalem e o ponto de corte obtido com cada um dos critérios é o mesmo.

Ponto de Corte = ponto de lucro máximo Ou Ponto de Corte = Ponto de lucro marginal zero

Porém, para que todo este cálculo seja feito, é necessário definir o conceito de lucro. Em linhas gerais, propõe-se que lucro seja o valor presente das receitas menos o valor presente dos custos envolvidos nas operações na data do contrato.

$$\text{Lucro} = \text{Valor Presente (Receitas)} - \text{Valor Presente (Custos)}$$

3. METODOLOGIA

O estudo trata-se de uma pesquisa exploratória dos temas que envolvam os scores de crédito e o avanço das novas tecnologias que ajudam a criar modelos preditivos. Utilizando-se de pesquisa bibliográfica para criação e argumentação dos fatos.

Sendo assim pretende-se entender através dessa pesquisa os modelos de scores tradicionais feitos com regressão logística em bureaus de crédito detentoras de bases de desenvolvimento criadas com informações de todo o mercado.

Pesquisar amplamente os benefícios que as novas tecnologias podem trazer e as informações trazida, comparando se é possível realizar um score que possa separar os bons e maus pagadores utilizando essas informações.

3.1 MODELOS

3.1.1 - REGRESSÃO LOGÍSTICA

O Método mais utilizado para uma avaliação da influência de fatores sobre uma resposta dicotômica é a regressão logística. (HOSMER; LEMESHOW, 1989). Esta técnica estatística consiste na separação de dois grupos, visando obter a probabilidade que uma observação pertença a um determinado conjunto, em decorrência do comportamento de suas variáveis explicativas.

A utilização da regressão logística é recomendada em muitos casos, pois permite que se analise o efeito de uma ou mais variáveis explicativas sobre uma variável proposta dicotômica, representando a presença (1) ou ausência (0) de uma característica (HOSMER; LEMESHOW, 1989).

Sua aplicação, é usualmente aplicada para análises cujos dados possuam resposta binária ou dicotômica a fim de realizar a predição ou explicar a ocorrência da variável resposta, relacionando-a com os fatores que influenciem positiva ou negativamente – variáveis preditoras – a probabilidade da ocorrência de determinado fenômeno.

Sendo assim, o modelo de regressão logística na análise de risco de crédito é utilizado para mensuração da probabilidade de inadimplência de determinado grupo de clientes, assumindo ser uma binomial com resultados 0 ou 1.

Supondo Y a variável resposta binária e X_1, X_2, \dots, X_k as variáveis independentes preditoras de Y , a função de distribuição pode ser expressa pela Equação 01:

$$P_i = \Pr \left[\frac{Y_i=1}{X=X_i} \right] = \frac{1}{1+e^{-Z}} \text{ onde:}$$

P_i : Representa a probabilidade de um evento ocorrer;

$$Z = b_0 + b_1X_1 + b_2X_2 + \dots + b_iX_i$$

X_i : Representam cada i -ésima variável preditora;

b_i : Coeficientes a serem estimados para cada uma das variáveis.

Entretanto, mesmo tendo uma grande flexibilidade, existe o pressuposto importante da baixa correlação entre as variáveis explicativas, já que o modelo de regressão logística é sensível à colinearidade entre as variáveis (HAIR et al., 2005). A utilização de variáveis altamente correlacionadas para a estimação do modelo pode ocasionar estimativas extremamente inflacionadas dos coeficientes de regressão (HOSMER; LEMESHOW, 1989).

3.1.2 - METODOLOGIAS ENSEMBLE

Ensemble Methods, utiliza as saídas de diversos modelos conjuntamente, a fim de melhorar a precisão do score final, como explicado no trabalho de Kwon, Choi e Suh (2013).

Esses métodos ainda podem ser classificados na categoria de Boosting. Existe um número expressivo de pesquisas sobre métodos ensembles e melhorias no desempenho do seu uso (PERRONE; COOPER, 1992).

Segundo Perrone e Cooper (1992) métodos ensembles combinam as saídas de vários modelos, que já foram ajustados previamente. No começo o tratamento desta técnica era um cálculo de uma média ponderada das saídas de cada um dos métodos. Presentemente tem sido abordado na literatura combinadores utilizando redes neurais.

A saída do método ensemble básico (do inglês Basic Ensemble Method - BEM) é definida pela expressão:

$$f_{BEM} = \frac{1}{n} \sum_{i=0}^n f_i(x)$$

Esta abordagem já cria uma melhora na qualidade dos resultados, segundo Perrone e Cooper (1992), entretanto, desta maneira não se leva em conta que alguns modelos de entrada podem ser mais complexos do que outros. Sua vantagem é ser fácil de entender e implementar e não aumenta o erro esperado (BISHOP et al., 1995).

Uma generalização para o método BEM é encontrar pesos para cada saída que minimizam o MSE (Erro Quadrático Médio (Mean Square Error)) do ensemble. O método geral do ensemble (GEM - do inglês Generalized Ensemble Method) é definido pela expressão:

$$f_{GEM} = \sum_{i=0}^n \alpha_i f_i(x)$$

sendo o α_i escolhidos para minimizar o MSE com relação à função alvo f (estimada usando o conjunto de validação) e somar para 1.

O erro $\epsilon_i(x)$ de uma rede f_i é definida como $\epsilon_i(x) = f(x) - f_i(x)$.

A matriz de correlação é definida $C_{ij} = E[\epsilon_i(x)\epsilon_j(x)]$.

A próxima tarefa é determinar os pesos para α_i responsáveis por minimizar o erro, a expressão a seguir:

$$MSE[f_{GEM}] = \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j C_{ij}$$

O método ensemble generalizado possui desempenho melhor que o método básico, e sempre fornece melhor estimativa (JIMÉNEZ, 1998).

Porém, o GEM depende de uma estimativa confiável de C e da esperança de que esta não seja singular, de modo que possa ser facilmente invertido. Na prática, os erros geralmente são altamente correlacionados, portanto, as linhas dessa matriz são linearmente dependentes, de modo que a sua inversão leva significativos de redução de erros.

3.1.2.1 Boosting

O método conhecido como boosting nasceu na área da computação oriunda do processo de machine learning (aprendizado de máquina).

Um dos algoritmos mais recentes e bem sucedidos deles é o algoritmo conhecido como AdaBoost, apresentado em (Freund e Schapire 1997). Este nome oriundo de Adaptive Boosting, e vem do fato de que AdaBoost gera em cada passo

(de forma determinística, mas adaptativa) uma distribuição sobre as observações da amostra, dando maior peso (maior probabilidade de estar na amostra perturbada) as observações classificadas incorretamente no passo anterior. Com esta dinâmica, as observações com pior ajuste serão mais prováveis de estar na amostragem para a construção do modelo seguinte, cuja estimação deve explicá-las de forma mais acurada, diminuindo o erro final da combinação dos modelos.

Derivado do Adaboost, a metodologia Gradient Boosting (GB) comumente é desenvolvida com a combinação de árvores de decisão. Após a estimação de uma árvore inicial, as vindouras são desenvolvidas de modo a minimizar o resíduo da anterior, o que pode ser realizado por diferentes funções de perda, sendo mínimos quadrados a mais usual (Friedman, 2002). Os processos de ajuste de modelos de GB devem considerar o teste de diferentes hiper parâmetros, tais quais quantidade máxima de árvores, profundidade máxima da árvore e número mínimo de indivíduos por nó, etc.(CASTRO, 2019).

3.1.4 - Support Vector Machine

A ideia inicial foi criar algoritmos que reconhecesse padrões, então surgiu a partir de estudos realizados por Fisher em 1936, em 1962, Rosenblatt sugeriu o algoritmo Perceptron de aprendizagem.

O SVM é um algoritmo que identifica padrões e classifica as variáveis binárias, apresentado. A essência deste método é a otimização matemática para identificação de um hiperplano que maximize a separação das observações em dois perfis, distinguindo perfis binários de interesse (Bellotti & Crook, 2006). O ajuste de SVM pode ser feito assumindo ou não a linearidade do hiperplano. Para casos de não-linearidade, é necessário analisar e decidir a melhor função de matemática.

A forma original de modelar o problema do SVM é através do seu modelo primal, onde se pode determinar o hiperplano separador de padrões distintos.

3.1.5 – Modelos híbridos

Chen et al. (2009) propõe uma modelagem em duas etapas. Primeiramente estimar uma árvore de decisão assim escolher as variáveis com maior poder discriminatório. Após, estima-se um modelo por SVM apenas com essas variáveis escolhidas. Já o estudo de Chen et al. (2012) também sugere uma modelagem em duas etapas, entretanto a primeira fase sendo um tratamento de variáveis, utilizando uma análise de cluster, após separando os grupos segundo a sua taxa de

inadimplência. Após cria-se uma variável utilizando os clueter como parâmetros.

Em seguida, estima-se um modelo por SVM com a nova variável resposta construída. Espera-se que a primeira etapa auxilie na criação de grupos homogêneos que removam o impacto de observações atípicas, melhorando a performance do modelo final. Essas duas abordagens serão testadas neste projeto, estendendo a segunda etapa para outras metodologias além de SVM (CASTRO, 2019).

4. Conclusão

O presente trabalho objetivou analisar o processo de construção de scores de créditos, comparar as suas modalidades de metodologias e analisar a importância dos beraus como monopólio de informação.

Sendo assim, vemos que mesmo com o crescimento dos métodos para a construção de scores, os modelos com regressão logística ainda são os mais utilizados, por apresentarem uma fácil compreensão e interpretabilidade. Ponto esse importante para a comercialização de score por partes dos beraus de crédito ao grande público.

Entretanto modelos de machine learning começaram a ganhar uma maior visibilidade, pois apresentam métricas de performance superiores a outras metodologias, sendo assim, esse ganho de performance implica diretamente na perda financeira por inadimplências em bancos e financeiras.

Como sugestão de próximos passos, indica-se o estudo expandido e comparativo com bases não enviesadas

5. REFERÊNCIAS BIBLIOGRÁFICAS

Addo, P. M., Guegan, D., & Hassani, B. (2018). Credit risk analysis using Machine and Deep Learning models. *Risks*, 6, 38–57.

ALVES, Mauro Correia. Estratégia para o desenvolvimento de modelos de *credit score* com inferência de rejeitados. 2008. 93 f. Dissertação (Instituto de Matemática e Estatística) – Universidade de São Paulo - São Paulo. 2008.

ARAÚJO, Elaine Aparecida. Modelagem de risco de crédito: Aplicação de modelos *Credit Scoring* no Fundo Rotativo de Ação da Cidadania – Cred Cidadania. 2006. 176 f. Dissertação – Universidade Federal de Pernambuco – Recife. 2006.

CARVALHO, Pedro Costa. *CREDIT SCORING*. 1999. 59 f. Monografia (Departamento De Economia) - Pontifícia Universidade Católica Do Rio De Janeiro - Rio de Janeiro.

CHAIA, Alexandre Jorge. Modelos de gestão do risco de crédito e sua aplicabilidade ao mercado brasileiro. 2003. Dissertação (Mestrado em Administração) - Faculdade de Economia, Administração e Contabilidade, University of São Paulo, São Paulo, 2003. doi:10.11606/D.12.2003.tde-21012004-224716. Acesso em: 2019-06-20.

Chen, W., Ma, C., & Ma, L. (2009). Mining the customer credit using hybrid support vector machine technique. *Expert Systems with Applications*, 36(4), 7611–7616.

Chen, W., Xiang, G., Liu, Y., & Wang, K. (2012). Credit risk Evaluation by hybrid data mining technique. 3, 194–200. <https://doi.org/10.1016/j.sepro.2011.10.029>

Chen, Y. S., & Cheng, C. H. (2013). Hybrid models based on rough set classifiers for setting credit rating decision rules in the global banking industry. *Knowledge-Based Systems*, 39, 224–239.

CREPALDE, Gustavo Henrique de Souza. A análise e o uso de Big Data nas atividades mineiras. 2017. 59 f. Monografia – Universidade Federal de Ouro Preto – Ouro Preto. 2017

DINIZ, Carlos. Métodos Estatísticos para análise de Dados de Crédito. ICMC–USP. 2003.

EIFERT, D. S. Análise quantitativa na concessão de crédito versus inadimplência: um estudo empírico. Dissertação (Mestrado em Administração) – Universidade Federal do Rio Grande do Sul, Rio Grande do Sul, 2003.

ESPIN-GARCIA, Osvaldo; RODRIGUEZ-CABALLERO, Carlos Vladimir. METODOLOGÍA PARA UN SCORING DE CLIENTES SIN REFERENCIAS CREDITICIAS. *Cuad. Econ.*, Bogotá, v. 32, n. 59, p. 137-162, junho 2013. Disponível em: <http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0121-47722013000100007&lng=en&nrm=iso>. Acesso em junho 2019.

FERREIRA, João Francisco de Assis; ANÁLISE DE CRÉDITO - FEMA. Assis. 2010.

Freund, Y., Schapire, R. E., & Hill, M. (1996). Experiments with a New Boosting Algorithm. Friedman, J.H. (2002) Stochastic Gradient Boosting. *Computational Statistics and Data Analysis*, 28, 367-378.

IBM. (2012). What is big data? Disponível em: <http://www01.ibm.com/software/in/data/bigdata/>. Acesso em junho de 2019.

MARCELINO, Juliana Alves. Credit Scoring: uma ferramenta para análise de crédito em uma instituição de microcrédito produtivo e orientado. 2011. 80 p. Orientador Cleyton de Oliveira Ritta. Trabalho de Conclusão de Curso de Ciências Contábeis. Universidade do Extremo Sul Catarinense – UNESC. Criciúma – SC

MIOLA, Rodrigo Fernando. USO DE MODELOS ESTATÍSTICOS PARA DADOS DE ESCORE

DE CRÉDITO DE UMA INSTITUIÇÃO FINANCEIRA. 2013. UNIVERSIDADE ESTADUAL PAULISTA “JULIO DE MESQUITA FILHO”. Bauru. 2013.

MORAES, Luciana de Godói. Uma abordagem alternativa de behavioral scoring usando modelagem híbrida de dois estágios com regressão logística e redes neurais. 2012. 35 f. Monografia (Instituto de Matemática - Universidade Federal do Rio Grande do Sul – Porto Alegre. 2012.

LEWIS, E. An Introduction to *Credit* scoring. Fair Isaac: San Rafael, California. 1992.

LOTT, Yuri Monnerat; CIANCONI, Regina de Barros. Vigilância e privacidade, no contexto do big data e dados pessoais: análise da produção da Ciência da Informação no Brasil. *Perspect. ciênc. inf.*, Belo Horizonte, v. 23, n. 4, p. 117-132, Dec. 2018, Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362018000400117&lng=en&nrm=iso>. Acesso em junho de 2019.

PACHECO, J. S. Utilizando análise de sobrevivência no crédito: um modelo de behavioral scoring. Porto Alegre: UFRGS. 2012.

PERDONÁ, G. S. C. Modelos de riscos aplicados à análise de sobrevivência. Tese de doutorado, Instituto de Ciências Matemáticas e de Computação. Universidade de São Paulo, São Carlos/SP, 2006.

PINA, Maria de Fátima. GeoMed 2017: visão mais profunda a partir de big data e pequenas áreas. *Cad. Saúde Pública*, Rio de Janeiro, v. 33, n. 10, e00172017, 2017. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-311X2017001000101&lng=en&nrm=iso>. Acesso em junho de 2019

ROSS, S. A.; WESTERFIELD, R.W.; JORDAN, B.D. Princípios de administração financeira. São Paulo: Atlas, 2000. SANTI FILHO, Armando de. Avaliação de risco de crédito: para gerentes de operações. São Paulo: Atlas, 1997.

SANTI FILHO, Armando de. Avaliação de risco de crédito: para gerentes de operações. São Paulo: Atlas, 1997.

SANTOS, José Odálio dos. *Análise de Crédito: Segmento: Empresas, Pessoas Físicas, Varejo, Agronegócio e Pecuária*, 6ª ed. Editora Atlas. 2015

SANTOS, José Odálio. *Análise de Crédito: Empresas, Pessoas Físicas, Agronegócios e Pecuária*. 3. ed. São Paulo: Atlas, 2009.

SCARPEL, Rodrigo Arnaldo; MILIONI, Armando Zeferino. Utilização conjunta de modelagem econométrica e otimização em decisões de concessão de crédito. *Pesqui. Oper.*, Rio de Janeiro, v. 22, n. 1, p. 61-72, junho 2002.

SCHRICKEL, W. K. *Análise de Crédito: Concessão e Gerência de Empréstimos*. 5. ed. São Paulo: Atlas, 2000.

SEMEDO, Danilson Pedro da Veiga. *Credit Scoring: Aplicação da Regressão Logística vs Redes Neurais Artificiais na Avaliação do Risco de Crédito no Mercado Cabo-Verdiano*. 2009. 126 f. Dissertação (Instituto Superior de Estatística e Gestão de Informação). Universidade Nova de Lisboa. Lisboa. 2009.

SICSÚ, Abraham Laredo. *Credit Scoring: desenvolvimento, implantação, acompanhamento*. São Paulo: Blucher, 2010.

THOMAS, L.C., Edelman, D. B., Crook, J. N. (2002). *Credit Scoring and Its Applications*. Philadelphia: Siam.

VASCONCELLOS, Mauricio Sandoval de. Proposta de método para análise de concessões de crédito a pessoas físicas. 2002. Dissertação (Mestrado em Teoria Econômica) - Faculdade de Economia, Administração e Contabilidade, University of São Paulo, São Paulo, 2003. doi:10.11606/D.12.2003.tde-18112004-125442. Acesso em: 2019-06-20.

