

Analytics & Inteligência Artificial

Tema da aula
Credit Scoring



BUSINESS SCHOOL

Graduação, pós-graduação, MBA, Pós-MBA, Mestrado Profissional, Curso In Company e EAD



CONSULTING

Consultoria personalizada que oferece soluções baseada em seu problema de negócio



RESEARCH

Atualização dos conhecimentos e do material didático oferecidos nas atividades de ensino



Líder em Educação Executiva, referência de ensino nos cursos de graduação, pós-graduação e MBA, tendo excelência nos programas de educação. Uma das principais **escolas de negócio do mundo**, possuindo convênios internacionais com Universidades nos EUA, Europa e Ásia. +8.000 **projetos de consultorias** em organizações públicas e privadas.



Único curso de graduação em administração a receber as notas máximas



A primeira escola brasileira a ser finalista da maior competição de MBA do mundo



Única *Business School* brasileira a figurar no *ranking* LATAM



Signatária do Pacto Global da ONU



Membro fundador da ANAMBA - Associação Nacional MBAs



Credenciada pela AMBA - Association of MBAs



Credenciada ao Executive MBA Council



Filiada a AACSB - Association to Advance Collegiate Schools of Business



Filiada a EFMD - European Foundation for Management Development



Referência em cursos de MBA nas principais mídias de circulação

O **Laboratório de Análise de Dados** – LABDATA é um Centro de Excelência que atua nas áreas de ensino, pesquisa e consultoria em análise de informação utilizando técnicas de **Big Data, Analytics** e **Inteligência Artificial**.



Profª Drª Alessandra Montini

O LABDATA é um dos pioneiros no lançamento dos cursos de *Big Data* e *Analytics* no Brasil

Os diretores foram professores de grandes especialistas do mercado

+10 anos de atuação

+1000 alunos formados

Docentes

- Sólida formação acadêmica: doutores e mestres em sua maioria
- Larga experiência de mercado na resolução de *cases*
- Participação em Congressos Nacionais e Internacionais
- Professor assistente que acompanha o aluno durante todo o curso

Estrutura

- 100% das aulas realizadas em laboratórios
- Computadores para uso individual durante as aulas
- 5 laboratórios de alta qualidade (investimento +R\$2MM)
- 2 Unidades próximas a estação de metrô (com estacionamento)

Conteúdo da Aula

1. Introdução
2. Histórico
3. CRISP-DM
4. *Default*
5. Inferência de Rejeitados
6. Ponto de Corte
7. Cases

1. O que é *Credit Scoring*





Escoragem de risco, quando aplicada a fins de avaliação de risco de crédito, é chamada de ***Credit Scoring* ou Escoragem de Crédito**.

Trata-se de uma ferramenta para avaliação de risco de crédito de clientes. O número, resultado do cálculo da escoragem de crédito, chama-se ***Credit Score* ou Score de Crédito**.

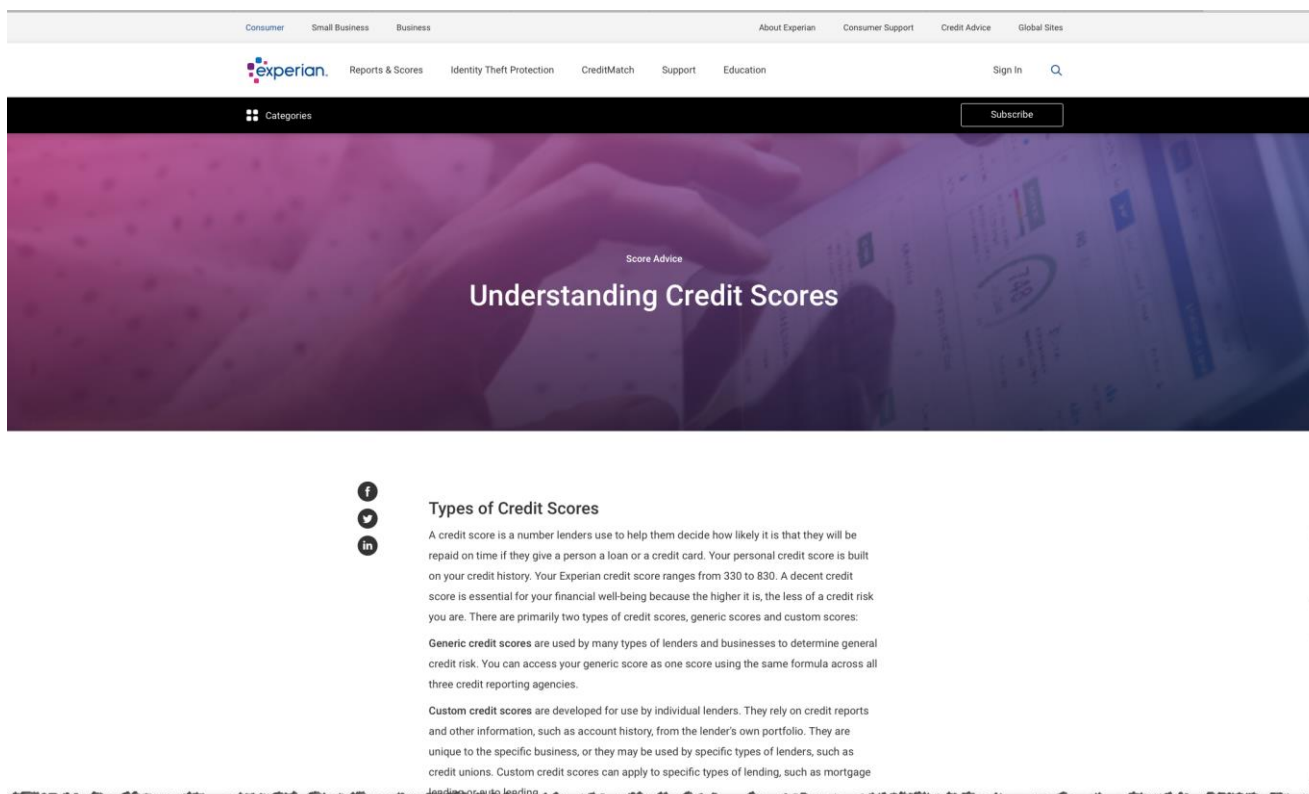


Diferentes tipos de Scores de Risco de Crédito

Tipo de Score	Descrição
Credit Scoring	Avalia o risco na concessão, iniciação ou originação do crédito.
Behavior Score	Avalia o risco de um cliente “em dia” tornar-se inadimplente (ou gerar “default”).
Collection Score	Avalia a propensão de um cliente inadimplente pagar sua dívida.
Credit Rating	A propensão à inadimplência pode ser traduzida em letras: AAA, AA, A, BBB, BB, B, ..., D, em que AAA é a menor propensão e D é a maior (default).
FICO Score	Score, desenvolvido pela FICO – Fair Isaac Corporation, fortemente utilizado nos EUA para avaliar risco de inadimplência dos clientes. Contempla não somente os birôs negativos, como também os positivos. Varia entre 300 e 850 e reflete a chance de um cliente ser um bom pagador, em que 300 é o pior score e 850 é o melhor.
Score Serasa	Score desenvolvido pela Serasa Experian, que varia entre 0 e 1.000, em que 0 é o pior score e 1000 é o melhor.



Material Adicional



Fonte: <https://www.experian.com/blogs/ask-experian/credit-education/score-basics/understanding-credit-scores/>



2. Breve Histórico



A origem do score de crédito deriva de uma pergunta muito simples: ***Posso contar que vou emprestar dinheiro (conceder crédito) a um determinado indivíduo e vou receber o dinheiro de volta?***



DID YOU KNOW?

Credit scores started in the 1950s when Bill Fair, an engineer and Earl Isaac, a mathematician created an automated scoring system. They eventually found the Fair Issac Corporation, developing and selling their credit scoring system to banks, retailers and corporations in the US and around the world.

MBOA.MORTGAGE.COMPANY|MBOAMTG.COM

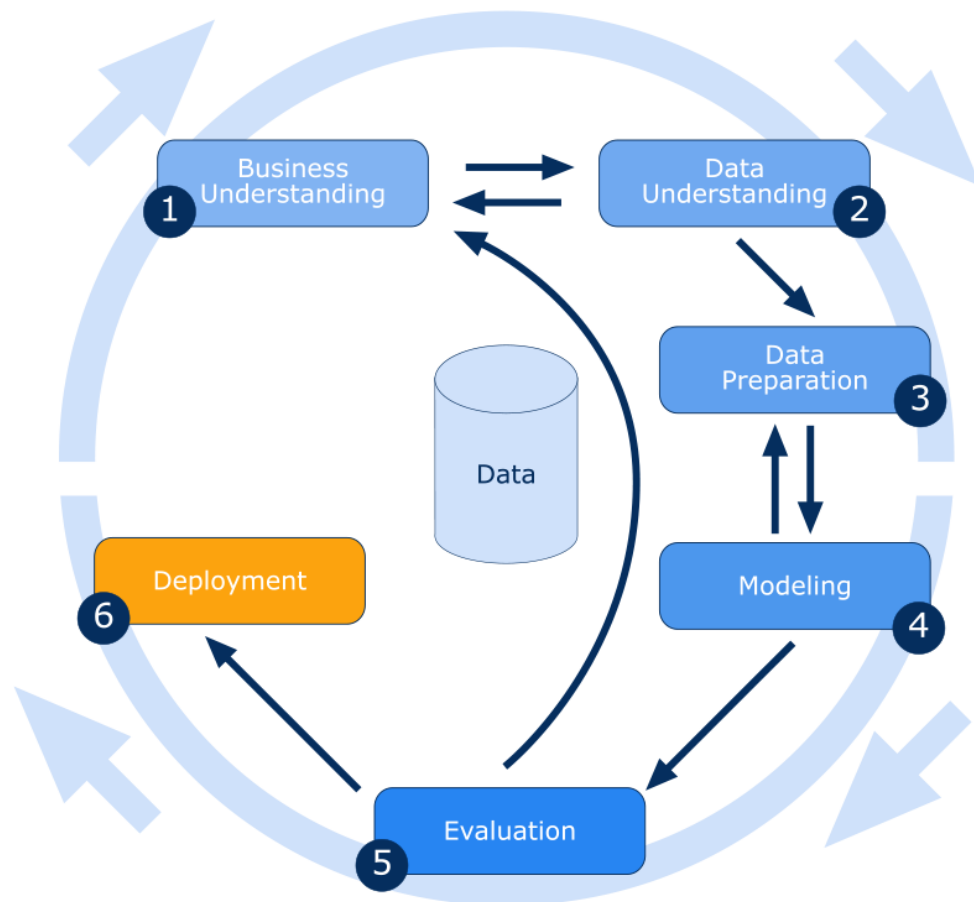
O engenheiro William Fair e o matemático Earl Isaac foram os precursores dos métodos de escoragem de risco, criados na década de 1950. Em 1956, eles criaram a Fair Isaac Corporation, que mais tarde, na década de 1980, deu origem ao **FICO Score**, o mais importante processo de escoragem do mundo.

FICO, um *rebranding* da Fair Isaac Corporation, é hoje uma empresa americana de software, dona do FICO Score e de ferramentas para gerenciamento dos processos de decisão nas empresas.



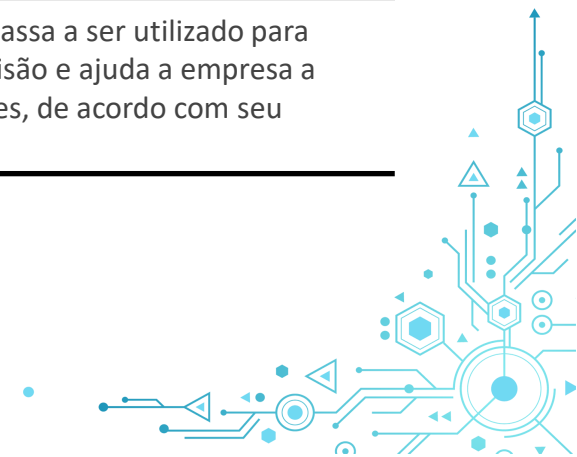
3. CRISP-DM em *Credit Scoring*





Etapa	Contexto
Entendimento do Problema de Negócio	Que problema de negócio a empresa quer resolver e qual o contexto por trás desse problema?
Entendimento dos Dados	Quais são os dados disponíveis, associados ao problema de negócio a ser estudado?
Preparação dos Dados	Tratamento, limpeza, consolidação, descrição, análise e insights sobre os dados associados ao problema de negócio
Modelagem	Uso de métodos quantitativos (regressão, redes neurais, árvores aleatórias, máquinas de vetor de suporte, etc.) para modelagem analítica dos dados.
Avaliação do Modelo	Quão preciso e estável é o modelo desenvolvido? Podemos confiar nele?
Implementação	Etapa em que é o modelo passa a ser utilizado para auxiliar no processo de decisão e ajuda a empresa a melhor selecionar os clientes, de acordo com seu apetite de risco.

Fonte: <https://www.semantix.com.br/blog/como-explorar-e-gerenciar-dados-com-o-crisp-dm>



Material Adicional

The screenshot shows a blog post on the Data Science Central website. The post is titled "CRISP-DM -- a Standard Methodology to Ensure a Good Outcome" and is authored by William Vorhies. It includes a summary, a "A Little History" section, and a diagram of the CRISP-DM process. The diagram shows a circular flow of six steps: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. The post also mentions the SAS Institute and the SEMMA process.

CRISP-DM -- a Standard Methodology to Ensure a Good Outcome
Posted by William Vorhies on July 26, 2016 at 9:15am

Summary: To ensure quality in your data science group, make sure you're enforcing a standard methodology. This includes not only traditional data analytic projects but also our most advanced recommenders, text, image, and language processing, deep learning, and AI projects.

A Little History
In the early 1990s as data mining was evolving from toddler to adolescent we spent a lot of time getting the data ready for the fairly limited tools and limited computing power of the day. Seldom were there more than one or two 'data scientists' in the same room and we were much more likely to be called 'predictive modelers' since that type of modeling was state-of-the-art in its day.

As the 90's progressed there was a natural flow that drew us toward standardizing the lessons we'd learned into a common methodology. Efforts like this always start out by wondering aloud whether there even was a common approach given that the problems looked so dissimilar. As it turns out there was.

Two of leading tools providers of the day, SPSS and Teradata, along with three early adopter user corporations, Daimler, NCR, and OHRA convened a special interest group (SIG) in 1996 (also probably one of the earliest collaborative efforts over the newly available worldwide web) and over the course of less than a year managed to codify what is still today the CRISP-DM, Cross Industry Standard Process for Data Mining. I'm honored to say that I was one of the original contributors to that SIG.

CRISP-DM was not actually the first. SAS Institute that's been around longer than anyone can remember had its own version called SEMMA (Sample, Explore, Modify, Model, Assess) but within just a year or two many more practitioners were basing their approach on CRISP-DM.

What is CRISP-DM?
The process or methodology of CRISP-DM is described in these six major steps

- Business Understanding**
Focuses on understanding the project objectives and requirements from a business perspective, and then convertin this knowledge into a data mining problem definition and a preliminary plan

Fonte: <https://www.datasciencecentral.com/profiles/blogs/crisp-dm-a-standard-methodology-to-ensure-a-good-outcome>

The screenshot shows a ResearchGate publication titled "Análise de Crédito Utilizando uma Abordagem de Mineração de Dados". The article is by Joyce Maria and Raniel Gomes Da Silva. It includes an abstract and a "Discover the world's research" section. The abstract describes a credit scoring instrument used to increase and facilitate sales of goods and services. The "Discover the world's research" section lists statistics: 15+ million members, 118+ million publications, and 700k+ research projects.

Análise de Crédito Utilizando uma Abordagem de Mineração de Dados
Article (PDF Available) · September 2018 with 4 Reads
DOI: 10.25288/pepa.v03i3.967

See all 6 References | See all 1 Figures

Download citation | Share | Download full-text PDF

Join ResearchGate to find the people and research you need to help your work.

- 15+ million members
- 118+ million publications
- 700k+ research projects

Join for free

Abstract
Creditis an instrument used to increase and facilitate sales of goods and services. He is responsible for a great part of the results obtained in the companies and for the development and growth of the economy of the country. However, a rigid assessment is necessary to where this credit should go, since, being applied to companies or wrong people, the credit can accumulate losses. In this way, this work proposes an approach using Data Mining for credit analysis through the application of Computational Intelligence algorithms, providing a more assertive decision making at the moment of credit granting.

Discover the world's research

- 15+ million members
- 118+ million publications
- 700k+ research projects

Join for free

Fonte: https://www.researchgate.net/publication/329923257_Analise_de_Credito_Utilizando_uma_Abordagem_de_Mineraao_de_Dados



CRISP-DM | DESENHO DAS ETAPAS DE UM PROCESSO E ESCORAGEM



4. Definição de *Default*



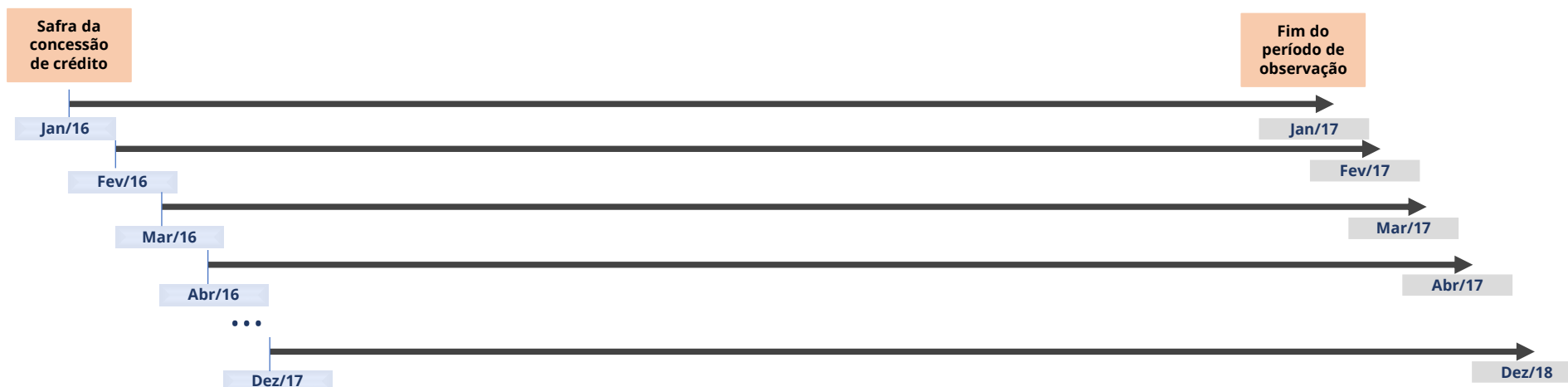
Definição de “bons” e “maus” clientes para a construção de um Credit Scoring



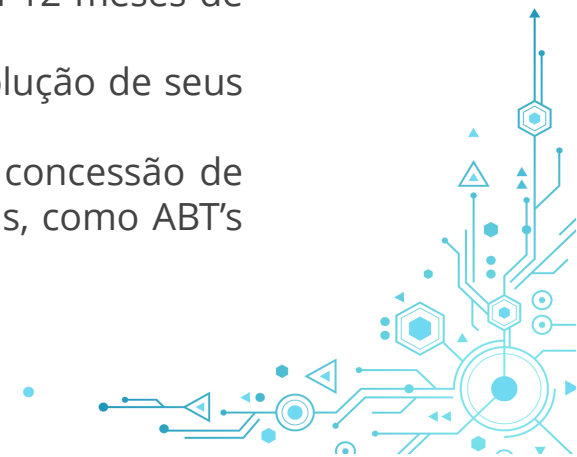
- É uma definição que deve estar em linha com as políticas e estratégias do produto que será considerado para a modelagem (cartão de crédito, cheque especial, crédito pessoal, etc.).
- Via de regra, o cliente é considerado como mau quando ultrapassa 60 dias de atraso em até 12 meses após o início da observação.
- É fundamental analisar clientes em diversas safras de observação para capturar diferentes efeitos, incluindo sazonais. Contudo, é importante que a base de análise não contemple um período muito extenso, cujas políticas de crédito ou características do portfolio tenham se alterado significativamente. Sugestão inicial poderia ser entre 18 e 24 safras de observação.
- O tempo de observação dependerá do horizonte de previsão a ser considerado. Usualmente, 12 meses é o período mais comum, podendo, em alguns casos, ser menor.



Definição de “bons” e “maus” clientes para a construção de um Credit Scoring



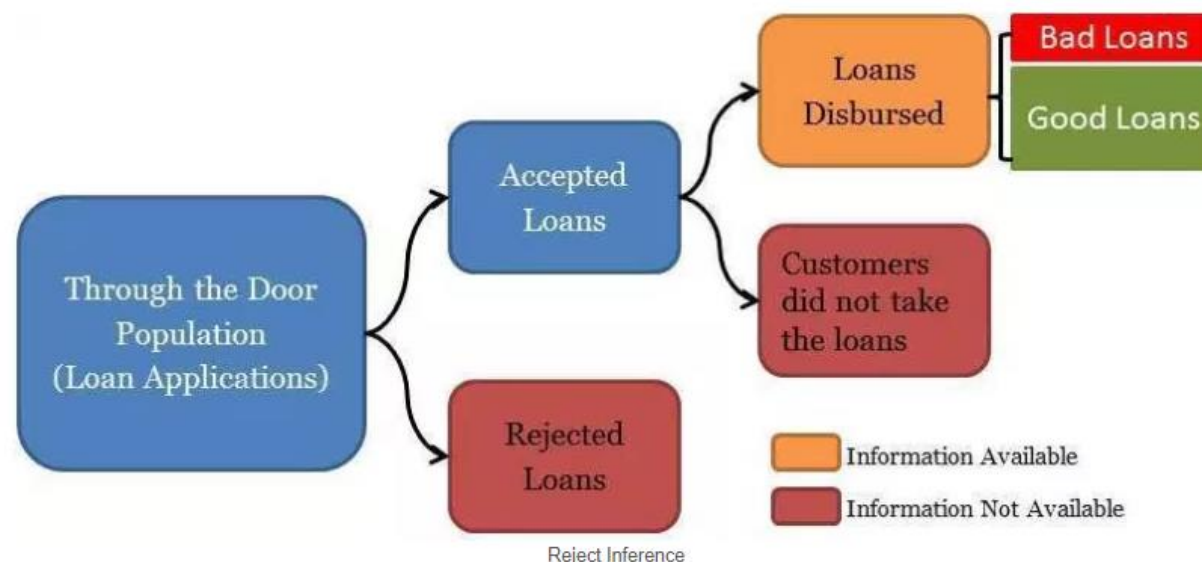
- Nesse diagrama usado como exemplo, estamos considerando 24 safras de análise, cada uma delas com 12 meses de observação.
- Em cada safra, fazemos a marcação do cliente como “bom”, “mau” ou “indeterminado” em função da evolução de seus níveis de atraso ao longo dos 12 meses, observando o status do atraso mês a mês.
- Um bom critério para compor as ABT's de Treinamento e Validação seria deixar as últimas 3 safras de concessão de crédito (mais recentes) como ABT's de validação (usadas somente para validação do modelo) e as demais, como ABT's de treinamento (usadas somente para o desenvolvimento do modelo).



5. Inferência de Rejeitados



Em um processo tradicional de análise de crédito, o score de crédito (**Credit Score**) desempenha função estratégica. Ele ajuda a determinar quais clientes serão **aprovados** e terão a concessão do crédito e quais serão **rejeitados (ou negados)** e não o receberão (salvo por políticas de exceção). Como os clientes rejeitados, via de regra, não receberam crédito, não foi possível obter e acompanhar sua performance e, portanto, é difícil considerá-los em um processo de modelagem de risco de crédito. Para endereçar esse problema, existe uma abordagem chamada **“Inferência dos Rejeitados”**.



Fonte: <http://ucanalytics.com/blogs/reject-inference-scorecards-banking-case-part-5/>



Em um processo tradicional de análise de crédito, o score de crédito (**Credit Score**) desempenha função estratégica. Ele ajuda a determinar quais clientes serão **aprovados** e terão a concessão do crédito e quais serão **rejeitados (ou negados)** e não o receberão (salvo por políticas de exceção). Como os clientes rejeitados, via de regra, não receberam crédito, não foi possível obter e acompanhar sua performance e, portanto, é difícil considerá-los em um processo de modelagem de risco de crédito. Para endereçar esse problema, existe uma abordagem chamada **“Inferência dos Rejeitados”**.

Credit Bureaus

Captura de informações de performance dos rejeitados nos birôs de crédito, que concentram dados desses clientes em outras instituições.

Fuzzy Augmentation

É uma extensão do processo de *parceling*, que contempla o uso de lógica fuzzy para analisar os rejeitados, sendo considerado um método mais sofisticado de análise.

Augmentation Through Parceling

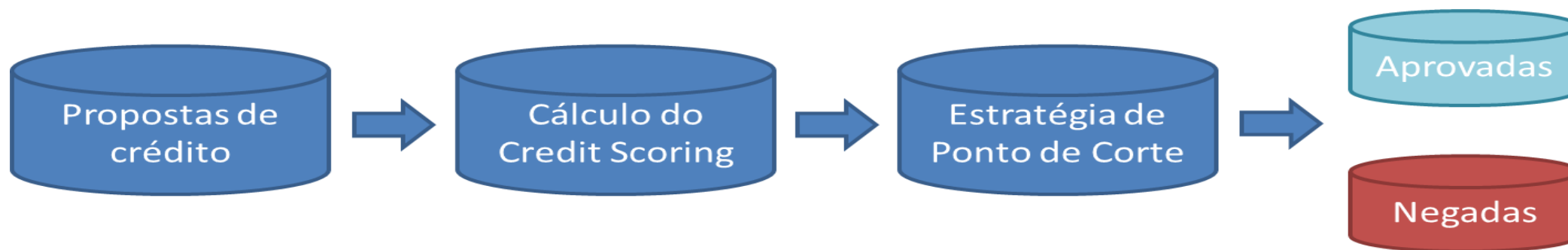
Essa é a forma mais conhecida para criar um modelo KGB (*known-good-bad model*) e escorar a base de rejeitados.



6. Ponto de corte em *Credit Scoring*



É necessária a definição de uma estratégia de ponto de corte em concessão de crédito



Além das formas já vistas em aula, também podem ser utilizadas as seguintes estratégias:

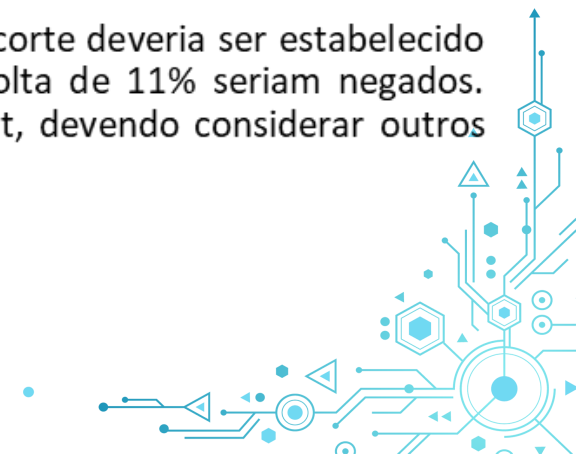
(a) Taxa de aprovação desejada, (b) Inadimplência esperada, (c) Margem de contribuição desejada.



Vamos considerar a seguinte tabela de distribuição de score e considerar a taxa de aprovação como critério para definição do ponto de corte:

Faixa de Score	Distribuição de Score de Propostas Avaliadas em Dez/2018		
	Qtde de Propostas	% de Propostas	% Acumulado de Propostas
<=100	21.251	2,9%	100,0%
101-200	26.521	3,6%	97,1%
201-300	36.412	4,9%	93,6%
301-400	52.121	7,0%	88,7%
401-500	78.521	10,6%	81,7%
501-600	71.124	9,6%	71,1%
601-700	98.645	13,3%	61,5%
701-800	99.011	13,3%	48,3%
801-900	107.521	14,5%	35,0%
>900	152.321	20,5%	20,5%
Total	743.448	100,0%	-

Analisando a tabela e considerando que queremos ter por volta de 89% de aprovação das propostas, nosso ponto de corte deveria ser estabelecido como 301, ou seja, todos aqueles clientes que tiverem score superior a 300 seriam aprovados. Os demais, por volta de 11% seriam negados. Logicamente, essa estratégia pode ser insuficiente se desejarmos ser mais rigorosos com as futuras taxas de default, devendo considerar outros elementos, como taxa de maus (*"bad rate"*), margem de contribuição, entre outras métricas de mercado.



7. Cases



Modelagem de avaliação de risco para crédito imobiliário

CASE | CREDIT SCORING

25

Um diretor deseja obter um modelo para tomada de decisão de concessão crédito imobiliário. A base **CREDITO IMOBILIARIO.xlsx** é composta por dados de financiamentos imobiliários feitos por clientes americanos, trazendo informações sobre as propostas, como dados sócio-demográficos, perfil do financiamento e características do cliente.

Avalie os seguintes itens:

- a) Faça a análise exploratória univariada e bivariada. Qual o % de clientes inadimplentes na base?
- b) Calcule o IV das variáveis.
- c) Categorize as 3 variáveis quantitativas mais fortes, de 2 formas: em quartis e de forma otimizada. Qual apresentou melhores resultados em termos de valor de *information value* (IV)?
- d) Divida a base em treino e teste e obtenha o modelo de regressão logística utilizando o método *Backward*, com 95% de confiança. Dado o número relativamente alto de variáveis, exclua as variáveis de IV fraquíssimo.
- e) Verifique se há multicolinearidade.
- f) Calcule o K-S e o AUC para o modelo na base de treino e teste. Alguma disparidade entre os resultados desses indicadores? Como você avalia a qualidade desse modelo?
- g) Mostre a tabela de distribuição do score, por decis. Qual o % de inadimplência por faixa? O score ordena o risco dos clientes de forma adequada?
- h) Calcule o PSI para a base de validação fora do tempo. Algum sinal de não aderência do modelo, em termos de estabilidade?



CREDITOIMOBILIARIO.XLSX



Modelagem de avaliação de risco para cartão de crédito

CASE | CREDIT SCORING

26

Um diretor deseja obter um modelo para tomada de decisão de concessão de cartão de crédito com base em variáveis cadastrais do cliente.

Avalie os seguintes itens:

- a) Faça a análise exploratória univariada e bivariada. Qual o % de clientes inadimplentes na base?
- b) Calcule o IV das variáveis.
- c) Categorize as 2 variáveis quantitativas de 2 formas: em quartis e de forma otimizada. Qual apresentou melhores resultados em termos de valor de *information value* (IV)?
- d) Divida a base em treino e teste e obtenha o modelo de regressão logística utilizando o método *Backward*, com 95% de confiança. Dado o número relativamente alto de variáveis, exclua as variáveis de IV fraquíssimo.
- e) Verifique se há multicolinearidade.
- f) Calcule o K-S e o AUC para o modelo na base de treino e teste. Alguma disparidade entre os resultados desses indicadores? Como você avalia a qualidade desse modelo?
- g) Mostre a tabela de distribuição do score, por decis. Qual o % de inadimplência por faixa? O score ordena o risco dos clientes de forma adequada?



R Studio

CARTAO DE CREDITO.XLSX

@2020 LABDATA FIA. Copyright all rights reserved.

