

Ciência de Dados para análise de créditos

Aplicação do modelo de Regressão Logística para apoio a decisão de crédito

Aplicação do modelo de Regressão Logística para apoio à decisão de crédito

Ciência de dados como métodos de transformação de dados em informação

Aplicação de Data Science como ferramenta de apoio a tomada de decisão orientada por dados.

Kipper

Skipper

capitão, arrais, saltador, larva do queijo, marabumbo, ratinho

Finance Captain

Finance Skipper

**SCORECARDS DE AVALIAÇÃO DE CRÉDITO**

## Com os mercados de crédito em constante evolução, com que eficiência você gerencia os riscos?

Os Scorecards de Avaliação de Crédito da S&P Global Market Intelligence fornecem aos profissionais de gestão de crédito e investimento ferramentas essenciais para identificar e gerenciar os possíveis riscos de inadimplência de contrapartes. As ferramentas analisam praticamente qualquer tipo de empresa, seja de capital aberto ou fechado, com ou sem rating, além de entidades governamentais e SPEs para financiamento de projetos.

Alguns Scorecards são aprimorados para incluir visualizações de fatores ambientais, sociais e de governança (ASG), permitindo estimar o impacto destes fatores no risco de crédito durante o processo de avaliação de crédito.

Saiba mais sobre Scorecards de avaliação de crédito

Mergulhe fundo em alguns dos nossos scorecards setoriais



## Avalie o risco de crédito com confiança.

### Combine direcionadores qualitativos e quantitativos

Os fatores qualitativos pontuais e prospectivos, junto às tendências convergentes, criam uma visão abrangente do risco de crédito.

## Aplique análise comparativa de ponta

Oferecemos 140 pontuações de risco da indústria e do país.

## Flexibilidade analítica

Avalie uma variedade de fatores ASG em termos de seus impactos nos fatores de risco de crédito relevantes.

## Avalie o risco de crédito com consistência.

### Use Diretrizes de Pontuação Orientada a Atributos

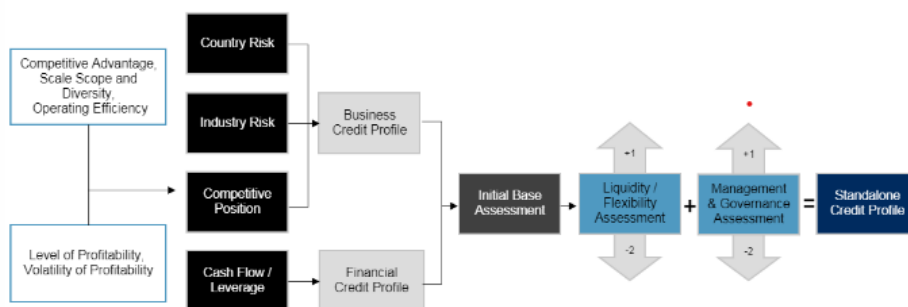
Forneça uma estrutura lógica e fácil de usar para pontuar os fatores de risco de negócios com 75% dos resultados de scorecards dentro de um nível das classificações de crédito públicas.

### Identificar risco padrão

Implemente uma escala de classificação granular de 20 pontos.

### Acesse pontuações numéricas de Scorecards

Os nossos Scorecards estão amplamente alinhados aos critérios da S&P Global Ratings e ainda são suportados por dados históricos de inadimplência que datam de 1981.



Credit Assessment Scorecards, S&P Global Market Intelligence. For illustrative purposes only.

## Avalie o risco de crédito de forma conveniente.

## Automação

Automatize a divulgação de dados financeiros usando o suplemento do Excel.

### Atualizações fáceis, documentação transparente

Os critérios de pontuação e o Guia do Usuário são atualizados anualmente através de uma análise rigorosa da metodologia, que também atende aos requisitos regulamentares.

### Implantação rápida

Os nossos scorecards são uma solução inovadora e estão prontos para uso, a fim de que os recursos da sua empresa possam ser implantados com mais eficiência.

Bank Name:	XYZ	Date of Analysis:	03 June 2016		
Bank's Home Regulator	Brazil	FC Sov. Rating	BB	Act as ceiling?	YES
BICRA Group	6	Economic Risk	7	Industry Risk	5

MACRO FACTORS		BANK SPECIFIC FACTORS				EXTERNAL SUPPORT			FINAL
Weighted-average Economic Risk	ANCHOR bbb	Business Position	Capital and Earnings	Risk Position	Funding and Liquidity	Group	Government	Alac	
Industry Risk		-1	1	-1	-1				
STAND ALONE						bb+			bb

\* For illustrative purposes only

**S&P Global**  
Market Intelligence

\* S&P Global Ratings does not contribute to or participate in the creation of credit scores generated by S&P Global Market Intelligence. Lowercase nomenclature is used to differentiate S&P Global Market Intelligence credit scores from credit ratings issued by S&P Global Ratings.

## Imóveis comerciais: considerações sobre risco de crédito para empréstimos diretos e dívida privada

Assista este webinar sob demanda, em que discutimos o estado atual do mercado de imóveis comerciais (CREs). Também cobrimos as melhores práticas de risco de crédito aplicáveis a empréstimos diretos e dívida privada, com atenção especial para CREs e fundos de investimento imobiliário (REITs).

## Estamos sempre nos esforçando para fornecer serviços nos quais você pode confiar.

A nossa equipe de modelagem e integração está disponível para ajudar a melhorar os seus processos internos e modelos de pontuação existentes, explicando a metodologia dos nossos cálculos.

## **Monitorando o impacto no risco de crédito corporativo**

Veja as nossas soluções em ação.

Leia os nossos estudos de caso.

### **Estudo de caso de mercados de títulos de dívida (DCM)**

Veja como um grande banco de investimento do sudeste asiático está expandindo seus negócios de DCM

[Acessar Artigo](#)

### **Estudo de caso de Scorecard na gestão de ativos**

Veja como um gestor de ativos asiático adota uma abordagem de melhores práticas para avaliar o risco de crédito

[Acessar Artigo](#)

### **Estudo de caso do Banco Nacional no Sudeste Asiático**

Saiba como um Banco Nacional no Sudeste Asiático está usando nossos scorecards de avaliação de crédito para cumprir o IFRS 9.

[Acessar Artigo](#)

### **Estudo de caso do Scorecard para financiamento de projetos**

Saiba como um banco multilateral de desenvolvimento latino-americano líder está usando os nossos scorecards de avaliação de crédito para avaliar o risco de crédito de financiamento de projetos.

[Acessar Artigo](#)

### **Estudo de caso**

Automatizando o processo de avaliação de crédito do mutuário em um banco universal.

[Acessar Artigo](#)

---

---

# Ciência de Dados em Instituições Financeiras

Hoje em dia, a Ciência de Dados é uma questão de sobrevivência dentro da indústria financeira.

Aplicar essa tecnologia ao mundo dos negócios significa inovação, automação e resultados significativos.

O grande volume de dados coletados por essas instituições estão se transformando em uma fonte rica de informações, para se comunicar com os clientes de maneira mais personalizada e eficiente.

Na era data-driven, os dados passaram a valer como insumos valiosos para decisões assertivas de negócios.

## Principais benefícios da Ciência de Dados em instituições financeiras

### 1. Combate às Fraudes

Um dos principais motivos para usar Ciência de Dados em bancos é o combate às fraudes.

A tecnologia possibilita um monitoramento permanente e em tempo real que identifica as informações de acesso para gerar padrões de utilização por usuário.

Quando acontece algum desvio desse padrão, as informações e o próprio acesso são bloqueados.

### 2. Reduz as Taxas de Churn

Taxas de churn referem-se ao cancelamento de contratos de serviços e encerramento de contas.

Através de tecnologias de Ciência de Dados e Analytics é possível analisar o comportamento do cliente e prever eventos.

### 3 – Melhora o gerenciamento de risco para concessão de crédito

A Ciência de Dados é capaz de fornecer dados muito mais precisos e confiáveis, a partir do cruzamento de dados.

Por ex: informar o padrão de comportamento de compra do cliente, seu patrimônio atual e potencial, tendência de inadimplência e muito mais.

### 4. Melhora o atendimento ao cliente

Quando é possível prever a insatisfação do cliente antes que ele troque de banco, os gestores podem implementar estratégias para melhorar o atendimento e evitar perdas.

Uma maneira de monitorar e personalizar o relacionamento cliente x banco é por meio do uso de Analytics no CRM.

#### CONCLUSÃO:

São muitos os benefícios da Ciência de Dados em instituições financeiras.

Especialmente quando se trata de proporcionar uma experiência personalizada ao cliente, melhorar as taxas de conversão e fidelizar os clientes.

A HupData já desenvolveu vários projetos de alto desempenho de Data Science, treinamentos e consultorias para grandes bancos e outras instituições financeiras.

## **Data Science para avaliação de crédito**

Veja como a ciência de dados pode contribuir para automatizar e melhorar os processos financeiros neste setor

Analisar o crédito de um cliente consiste, basicamente, em verificar se ele realmente tem condições de honrar o compromisso assumido. O objetivo é identificar o perfil e as condições financeiras do cliente, a partir de suas informações básicas e de seus dados financeiros. Assim, a empresa também define um valor mínimo de compra, a taxa de juros aplicada e o prazo de pagamento das parcelas, por exemplo. De forma geral, é uma forma de evitar inadimplência nas operações de crédito.

Essa análise pode ser realizada tanto para clientes pessoa física quanto jurídica e é hoje um dos maiores campos de atuação da ciência de dados aplicada ao setor financeiro. Afinal, a avaliação manual dos pedidos de empréstimo e das informações do cliente apresenta um risco iminente. A habilidade dos avaliadores humanos sempre será limitada em termos de precisão e velocidade. Logo, a utilização de métodos e ferramentas para automatizar e escalar esses processos ajuda a aumentar as chances de emissão de empréstimos de risco com menor probabilidade de inadimplência

## **Os problemas dos dados**

Para fornecer empréstimos, por exemplo, as empresas podem enfrentar alguns problemas relacionados à concessão de crédito e praticamente todos eles têm a ver com o perfil do cliente, verificado a partir dos dados que as instituições financeiras têm sobre ele. E a realidade é que nem sempre eles são suficientes para a análise. A ideia de dados insuficientes em um mundo tão digital pode até parecer estranha, mas é real, principalmente pela falha em conseguir acessá-los. Além do acesso aos dados disponíveis, sejam eles estruturados ou não estruturados, a análise é o ponto essencial para o processo de avaliação do risco de crédito. A criação de grupos relativamente pequenos para entender o comportamento geral do cliente desempenha um papel importante nessa avaliação.

## Automatizando a análise de crédito

É necessário adotar métodos inovadores, que subsidiem uma tomada de decisão acertada para negócios de todos os portes e segmentos de atuação. Por isso, é necessário usar algumas ferramentas que ajudam a analisar a concessão de crédito. Entre as tecnologias necessárias estão:

**Business Intelligence (BI):** coleta e processa um grande volume de informações para gerar estatísticas confiáveis e determinar se a venda a prazo é uma boa ideia naquele contexto. Garante agilidade e ajuda a identificar os riscos da operação;

**Machine Learning:** automatiza a análise de dados e os modelos analíticos para conceder a venda a prazo de modo mais seguro. Há o aprendizado autônomo, conforme os comandos oferecidos, o que aumenta a precisão da avaliação;

**Big Data:** usa diferentes fontes para coletar dados e oferecer informações valiosas, que trazem exatidão para a análise de crédito.

Todas essas abordagens se beneficiam muito do processo de automação. Eles reúnem informações sobre os clientes, traçam os perfis e gerenciam os riscos das operações. A partir do uso das soluções, há mais agilidade no processo de análise e maior potencial de fidelização dos clientes. Da mesma forma, a avaliação do crédito se torna mais precisa e você protege sua empresa contra desequilíbrios financeiros.

## Um caso de Machine Learning na avaliação de risco de crédito

O aprendizado de máquina permite a automação do trabalho de análise de crédito. Os algoritmos têm muito a oferecer ao mundo da avaliação de risco



devido ao seu poder preditivo e alta velocidade. A cientista de dados Sarah Beshr [escreveu artigo no Medium](#) sobre como usar machine learning para prever a probabilidade de um tomador de empréstimo ficar inadimplente.

O conjunto de dados utilizados pode ser encontrado no [Kaggle](#) e contém dados de 32.581 mutuários e 11 variáveis relacionadas a cada mutuário. São variáveis como:

Idade - variável numérica; idade em anos

Renda - variável numérica; renda anual em dólares

Situação de residência - variável categórica; “aluguel”, “hipoteca” ou “próprio”

Intenção de empréstimo – variável categórica; “educação”, “médico”, “empreendimento”, “reforma da casa”, “pessoal” ou “consolidação de dívidas”

Montante do empréstimo - variável numérica; valor do empréstimo em dólar

Grau do empréstimo - variável categórica; “A”, “B”, “C”, “D”, “E”, “F” ou “G”

Racional entre o empréstimo e o rendimento - variável numérica; entre 0 e 1

Padrão histórico - variável binária, categórica; “S” ou “N”

Status do empréstimo - binário, variável numérica; 0 (sem padrão) ou 1 (padrão) → esta será a variável de destino

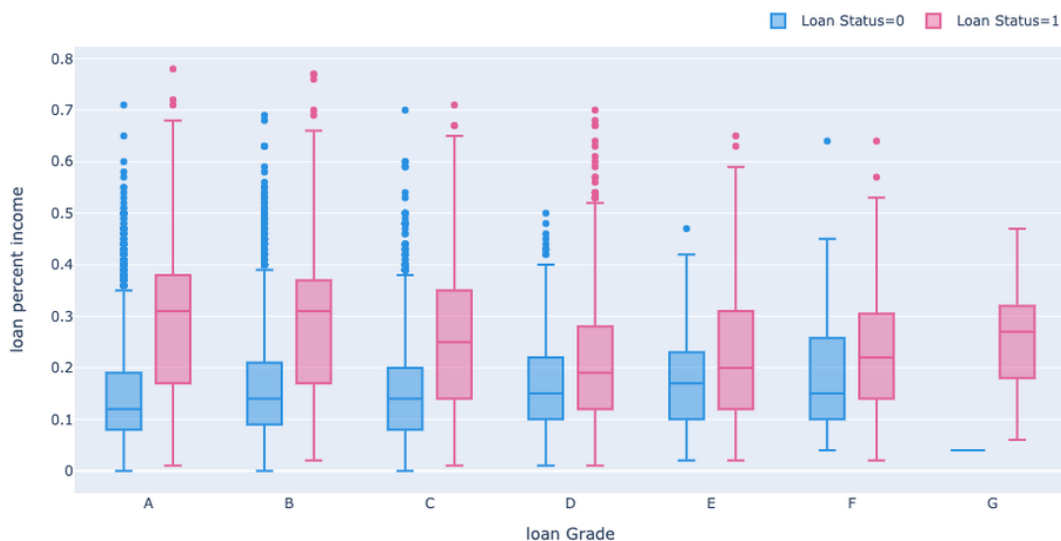
Após saber os tipos de variáveis disponíveis, o próximo passo é a etapa de exploração e pré-processamento de dados. Verifique se há variáveis ausentes ou com erro e outliers, para que possam ser corrigidos.

	Age	Income	Employment_Length	Loan_Amount	Interest_Rate	Loan_Status	loan_percent_income
count	28638.000000	2.863800e+04	28638.000000	28638.000000	28638.000000	28638.000000	28638.000000
mean	27.727216	6.664937e+04	4.784482	9656.493121	11.039867	0.216600	0.169488
std	6.310441	6.235645e+04	4.095491	6329.683361	3.229372	0.411935	0.106393
min	20.000000	4.000000e+03	0.000000	500.000000	5.420000	0.000000	0.000000
25%	23.000000	3.948000e+04	2.000000	5000.000000	7.900000	0.000000	0.090000
50%	26.000000	5.595600e+04	4.000000	8000.000000	10.990000	0.000000	0.150000
75%	30.000000	8.000000e+04	7.000000	12500.000000	13.480000	0.000000	0.230000
max	144.000000	6.000000e+06	123.000000	35000.000000	23.220000	1.000000	0.830000

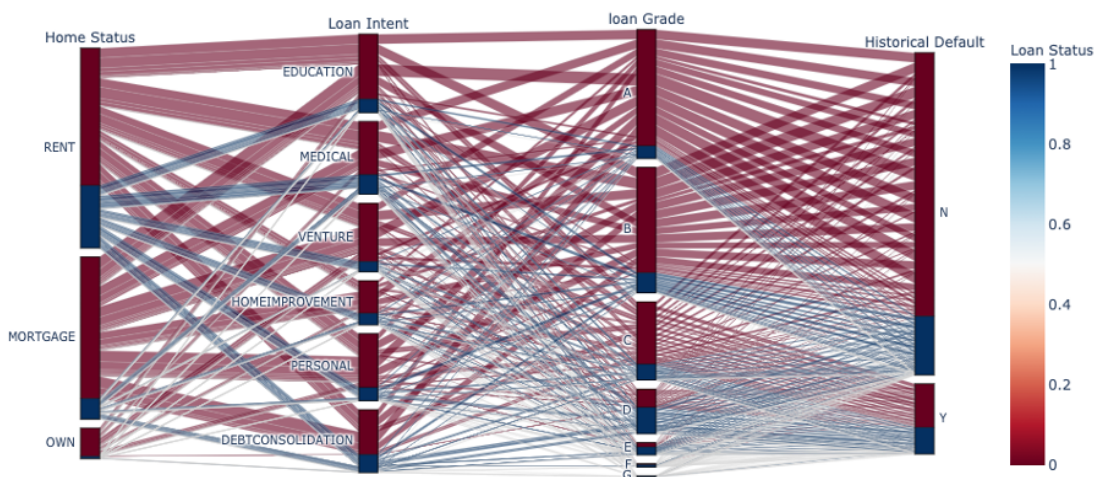
*Representação do conjunto de dados*

O conjunto de dados pode apontar inconstâncias, como idade de 144 anos ou pessoas que estiveram empregadas por 123 anos. Deixando de lado, os valores discrepantes das variáveis *idade* e *tempo de emprego* podem afetar negativamente o modelo e, portanto, devem ser removidos.

Após o processo finalizado e com os dados limpos em mão, convém explorar mais a fundo como o status do empréstimo está relacionado a outras variáveis no conjunto de dados.



Podemos ver que aqueles que não apontam inadimplência têm um valor médio de índice de empréstimo / receita mais baixo em todas as classes de empréstimos. Também podemos ver que nenhum tomador de empréstimo com grau de empréstimo G foi capaz de honrar o empréstimo. Usando um diagrama de categoria paralelo é possível entender como diferentes variáveis categóricas do nosso conjunto de dados estão relacionadas entre si e podemos mapear essas relações com base *no status do empréstimo*.



Principais conclusões do diagrama:

- O conjunto de dados é composto principalmente de clientes que nunca entraram em inadimplência em um empréstimo;
- As notas “A” e “B” do empréstimo são as mais comuns, enquanto “F” e “G” são as menos comuns;
- Os locatários de casas ficam inadimplentes com mais frequência em seus empréstimos do que aqueles com hipotecas, enquanto os proprietários de casas tendem a ficar menos inadimplentes;

Os clientes contrataram menos empréstimos para reforma da casa e mais para educação. Além disso, a inadimplência era mais comum em empréstimos contraídos para cobrir despesas médicas e consolidação de dívidas.

Tudo pronto para a próxima etapa: treinamento e avaliação do modelo. Serão treinados e testados três modelos, KNN, regressão logística e XGBoost. Também será avaliado seu desempenho na previsão de inadimplências de empréstimos e sua probabilidade.

KNN				
	precision	recall	f1-score	support
0	0.84	0.96	0.90	4448
1	0.74	0.38	0.50	1260
accuracy			0.83	5708
macro avg	0.79	0.67	0.70	5708
weighted avg	0.82	0.83	0.81	5708

Logistic Regression				
	precision	recall	f1-score	support
0	0.81	0.99	0.89	4448
1	0.79	0.17	0.29	1260
accuracy			0.81	5708
macro avg	0.80	0.58	0.59	5708
weighted avg	0.80	0.81	0.76	5708

XGBoost				
	precision	recall	f1-score	support
0	0.92	0.99	0.96	4448
1	0.94	0.72	0.81	1260
accuracy			0.93	5708
macro avg	0.93	0.85	0.88	5708
weighted avg	0.93	0.93	0.92	5708

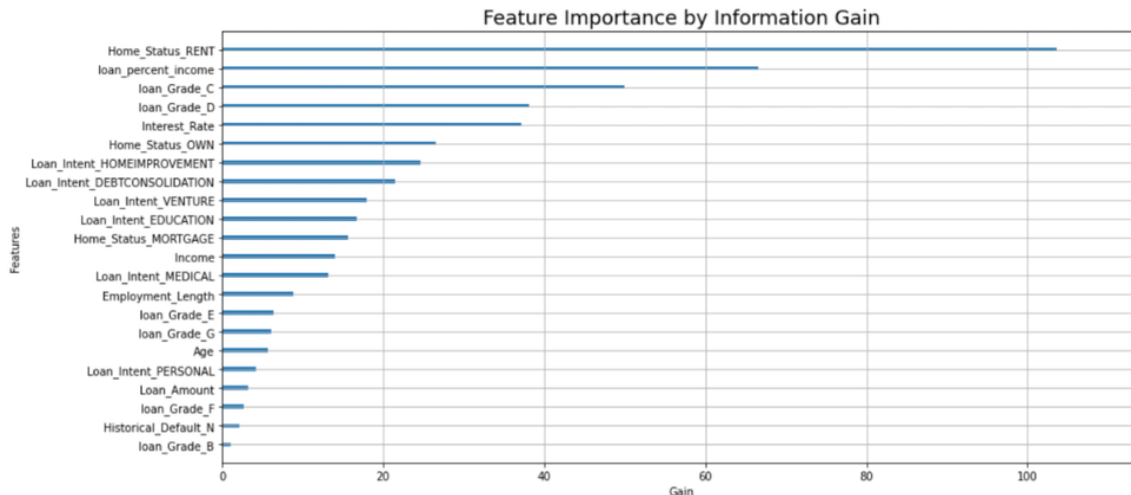
Em primeiro lugar, **precision** dá a proporção de verdadeiros positivos para o total de positivos previstos por um classificador, onde positivos denotam casos padrão no contexto. Dado que eles são a classe minoritária no conjunto de dados, podemos ver que os modelos fazem um bom trabalho em prever corretamente essas instâncias menores. Além disso, **Recall**, também conhecido como taxa positiva verdadeira, nos dá o número de reais positivos dividido pelo número total de elementos que realmente pertencem à classe positiva.

Neste caso, o Recall é uma métrica mais importante do que a **Precisão**, visto que estamos mais preocupados com falsos negativos (o modelo prevendo que alguém não ficará inadimplente, mas ele fica) do que falsos positivos (o modelo prevendo que alguém ficará inadimplente mas ele não fica).

Por último, **F1 Score** fornece uma pontuação única para medir a precisão e a recuperação. Agora que sabemos o que procurar, podemos ver claramente que o

XGboost tem o melhor desempenho em todas as 3 métricas. Embora tenha pontuado melhor em Precision em oposição a Recall, ainda tem uma pontuação de F1 muito boa de 0,81.

Por último, é hora de identificar quais recursos foram mais influentes na previsão da variável-alvo.



Podemos ver na figura acima que o aluguel, o índice empréstimo / receita e o empréstimo grau C são as três características mais importantes para prever a inadimplência do empréstimo e sua probabilidade.

De forma resumida, o passo a passo apresentado aqui mostra que os dados foram pré-processados e treinados, avaliando-se os resultados de 3 modelos de machine learning por sua capacidade de prever inadimplências de empréstimos e sua probabilidade. Depois de identificar que o XGBoost teve o melhor desempenho em todas as métricas, investiga-se quais recursos eram mais importantes para as previsões

Assim, é possível completar a demonstração de como o aprendizado de máquina pode ser aplicado ao mundo da avaliação de risco de crédito, automatizando e tornando o processo muito mais assertivo .

