

Data Science

Um olhar no **credit scoring**.



QUEM SOU EU?



ANDRESSA FREIRES

- Técnica em Desenv. de Sistemas
- Licenciatura - Física
- BTi - Análises de Dados
- Bacharelado em Ciência de Dados
- Palestrante - Data Science / Diversidade
- Cientista de Dados, Representante de Inclusão, Diversidade e Cultura - PicPay
- Produtora de conteúdo:
 - @datasciencedescomplicada e @diversidadedescomplicada
- Voluntária: Afropython, Pyladies, AIGirls, UXPMMP
- Contatos:
 - @dessafreires
 - <https://www.linkedin.com/in/andressafreires/>

Data Science

Data Science é uma disciplina que permite tornar dados brutos em compreensão, insights e conhecimento.

Hadley Wickham

Data Science em

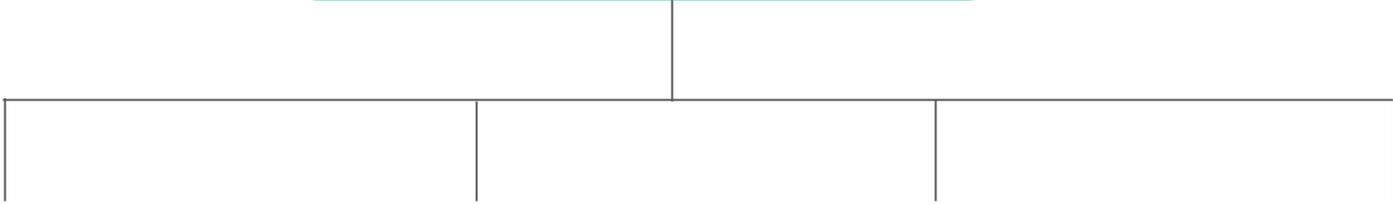
CRÉDITO

Um olhar no *credit scoring*



Risco de Crédito | Panorama Geral

Risco de crédito = Probabilidade de perda



Risco de Crédito | Panorama Geral

Risco de crédito = Probabilidade de perda

```
graph TD; A[Risco de crédito = Probabilidade de perda] --- B[ ]; B --- C[ ]; B --- D[ ]; B --- E[ ]; C --- F[A concessão de crédito vem sob condições de incerteza. E sempre existe a possibilidade de perda.]
```

A concessão de crédito vem sob condições de **incerteza**. E sempre existe a possibilidade de **perda**.

Risco de Crédito | Panorama Geral

Risco de crédito = Probabilidade de perda

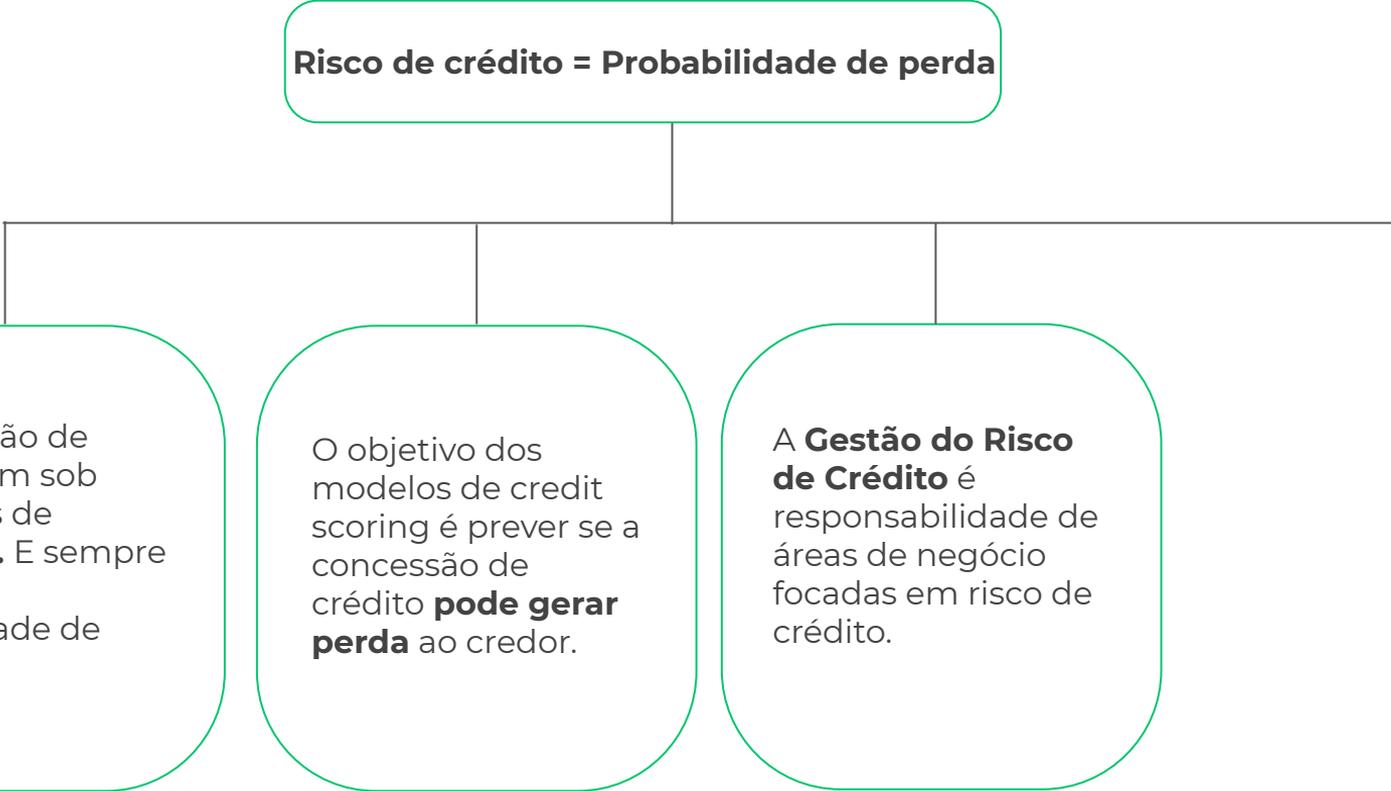
```
graph TD; A[Risco de crédito = Probabilidade de perda] --- B[A concessão de crédito vem sob condições de incerteza. E sempre existe a possibilidade de perda.]; A --- C[O objetivo dos modelos de credit scoring é prever se a concessão de crédito pode gerar perda ao credor.]; A --- D[ ]
```

A concessão de crédito vem sob condições de **incerteza**. E sempre existe a possibilidade de **perda**.

O objetivo dos modelos de credit scoring é prever se a concessão de crédito **pode gerar perda** ao credor.

Risco de Crédito | Panorama Geral

Risco de crédito = Probabilidade de perda



```
graph TD; A[Risco de crédito = Probabilidade de perda] --- B[A concessão de crédito vem sob condições de incerteza. E sempre existe a possibilidade de perda.]; A --- C[O objetivo dos modelos de credit scoring é prever se a concessão de crédito pode gerar perda ao credor.]; A --- D[A Gestão do Risco de Crédito é responsabilidade de áreas de negócio focadas em risco de crédito.];
```

A concessão de crédito vem sob condições de **incerteza**. E sempre existe a possibilidade de **perda**.

O objetivo dos modelos de credit scoring é prever se a concessão de crédito **pode gerar perda** ao credor.

A **Gestão do Risco de Crédito** é responsabilidade de áreas de negócio focadas em risco de crédito.

Risco de Crédito | Panorama Geral

Risco de crédito = Probabilidade de perda

A concessão de crédito vem sob condições de **incerteza**. E sempre existe a possibilidade de **perda**.

O objetivo dos modelos de credit scoring é prever se a concessão de crédito **pode gerar perda** ao credor.

A **Gestão do Risco de Crédito** é responsabilidade de áreas de negócio focadas em risco de crédito.

Cabe a estas áreas **aplicar as políticas e modelos para a avaliação do risco de crédito** em todo o ciclo de crédito.

Credit Scoring | Panorama Geral

Credit Scoring = Medida do risco de crédito

```
graph TD; A["Credit Scoring = Medida do risco de crédito"] --- B[ ]; B --- C[ ]; B --- D[ ]; B --- E[ ]
```

Credit Scoring | Panorama Geral

Credit Scoring = Medida do risco de crédito

```
graph TD; A["Credit Scoring = Medida do risco de crédito"] --- B[ ]; B --- C[ ]; B --- D[ ]; B --- E[ ]; C --- F["Modelo de credit scoring é a denominação dada para fórmulas que calculam os escores de crédito."];
```

Modelo de credit scoring é a denominação dada para fórmulas que calculam os escores de crédito.

Credit Scoring | Panorama Geral

Credit Scoring = Medida do risco de crédito

```
graph TD; A["Credit Scoring = Medida do risco de crédito"] --- B["Modelo de credit scoring é a denominação dada para fórmulas que calculam os escores de crédito."]; A --- C["Eles têm a finalidade de quantificar o risco de crédito."]; A --- D[" "];
```

Modelo de credit scoring é a denominação dada para fórmulas que calculam os escores de crédito.

Eles têm a finalidade de **quantificar o risco** de crédito.

Credit Scoring | Panorama Geral

Credit Scoring = Medida do risco de crédito

```
graph TD; A["Credit Scoring = Medida do risco de crédito"] --- B["Modelo de credit scoring é a denominação dada para fórmulas que calculam os escores de crédito."]; A --- C["Eles têm a finalidade de quantificar o risco de crédito."]; A --- D["A forma como a informação gerada pelo credit scoring é utilizada para tomadas de decisão é atribuição da equipe gestora de risco."];
```

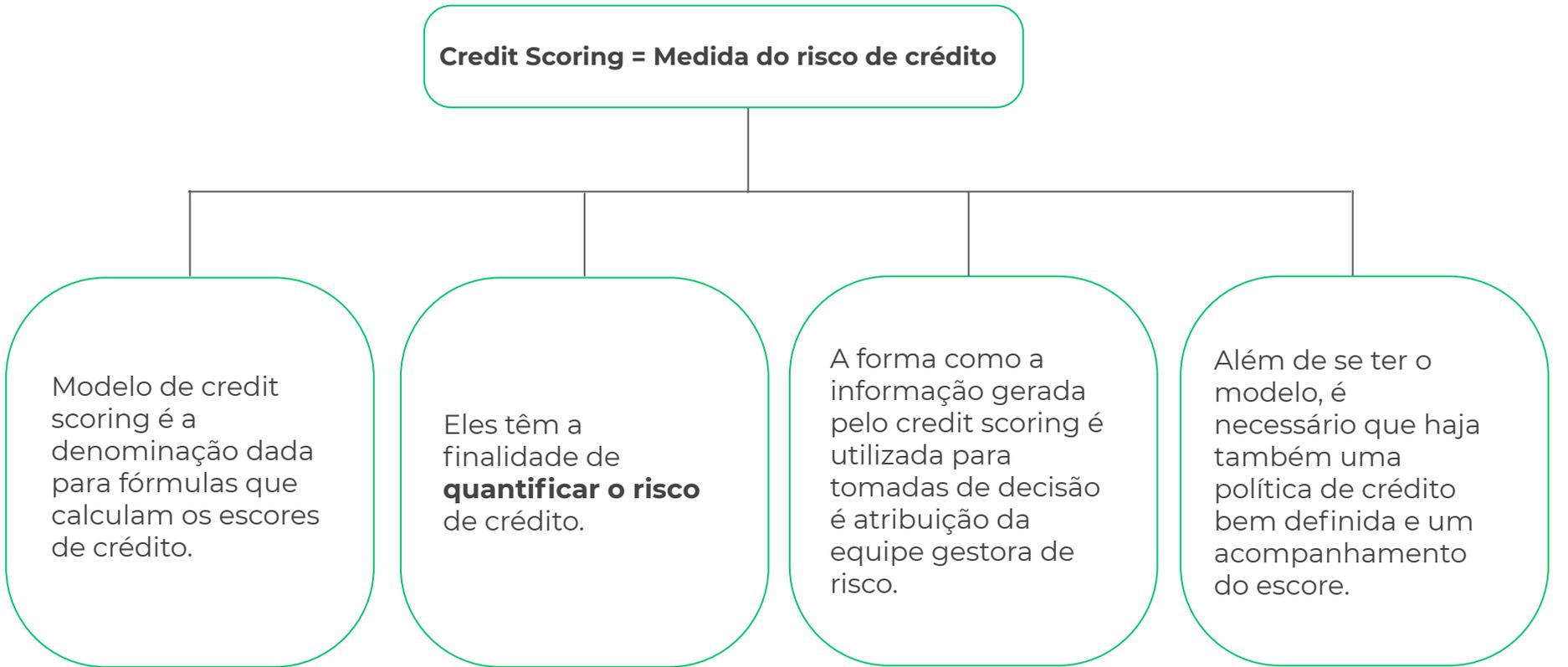
Modelo de credit scoring é a denominação dada para fórmulas que calculam os escores de crédito.

Eles têm a finalidade de **quantificar o risco** de crédito.

A forma como a informação gerada pelo credit scoring é utilizada para tomadas de decisão é atribuição da equipe gestora de risco.

Credit Scoring | Panorama Geral

Credit Scoring = Medida do risco de crédito



```
graph TD; A["Credit Scoring = Medida do risco de crédito"] --- B["Modelo de credit scoring é a denominação dada para fórmulas que calculam os escores de crédito."]; A --- C["Eles têm a finalidade de quantificar o risco de crédito."]; A --- D["A forma como a informação gerada pelo credit scoring é utilizada para tomadas de decisão é atribuição da equipe gestora de risco."]; A --- E["Além de se ter o modelo, é necessário que haja também uma política de crédito bem definida e um acompanhamento do escore."];
```

Modelo de credit scoring é a denominação dada para fórmulas que calculam os escores de crédito.

Eles têm a finalidade de **quantificar o risco** de crédito.

A forma como a informação gerada pelo credit scoring é utilizada para tomadas de decisão é atribuição da equipe gestora de risco.

Além de se ter o modelo, é necessário que haja também uma política de crédito bem definida e um acompanhamento do escore.

Risco de Crédito | Panorama Geral

Como o risco de crédito pode ser avaliado?



ANALISTA

X

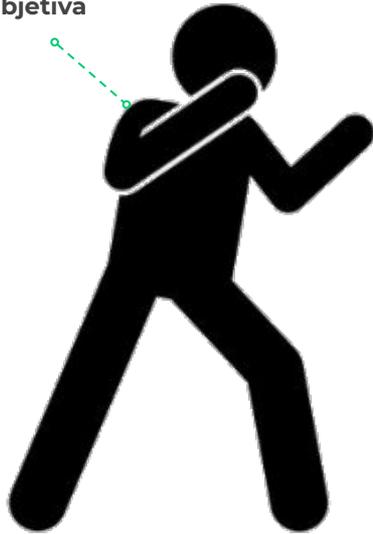


MODELAGEM
PREDITIVA

Risco de Crédito | Panorama Geral

Como o risco de crédito pode ser avaliado?

Avaliação
subjetiva



ANALISTA

X

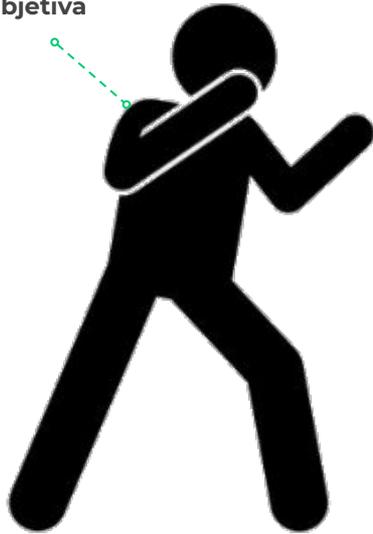


MODELAGEM
PREDITIVA

Risco de Crédito | Panorama Geral

Como o risco de crédito pode ser avaliado?

Avaliação
subjetiva



ANALISTA

X

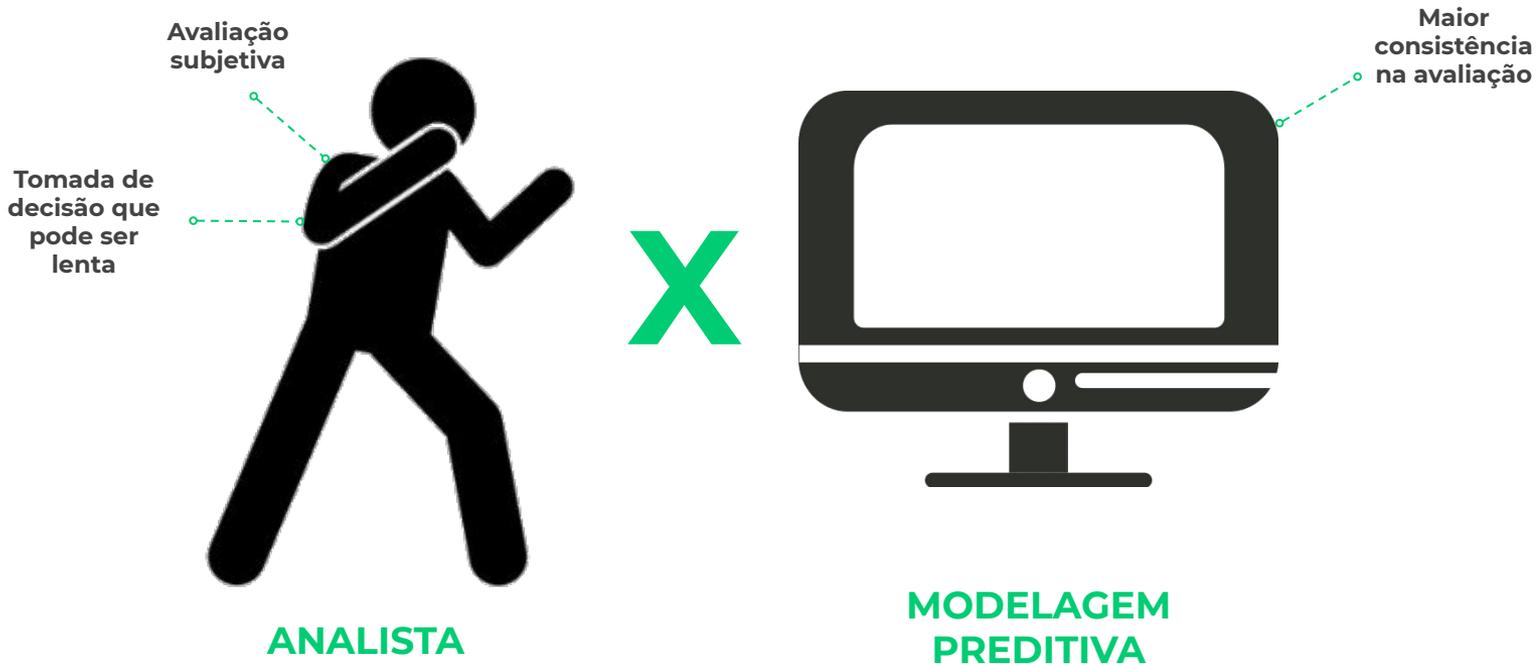


Maior
consistência
na avaliação

MODELAGEM
PREDITIVA

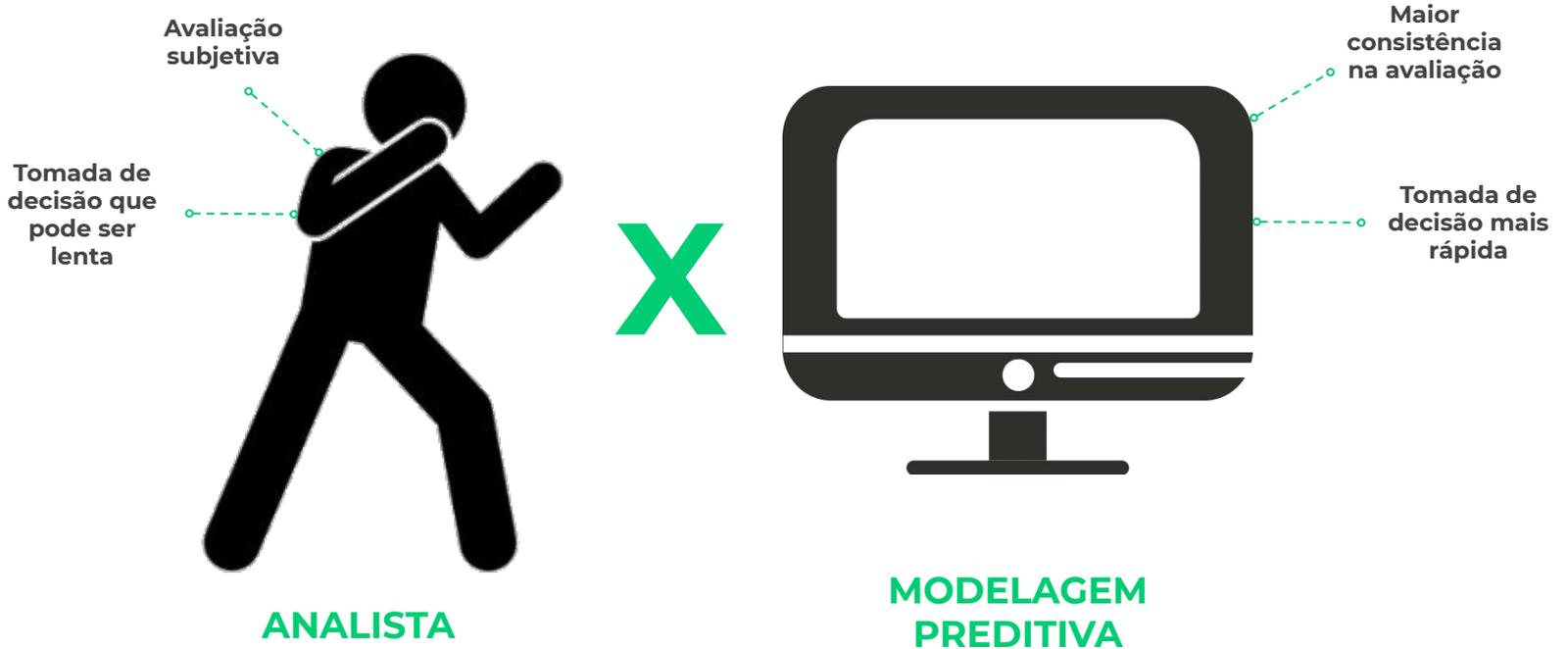
Risco de Crédito | Panorama Geral

Como o risco de crédito pode ser avaliado?



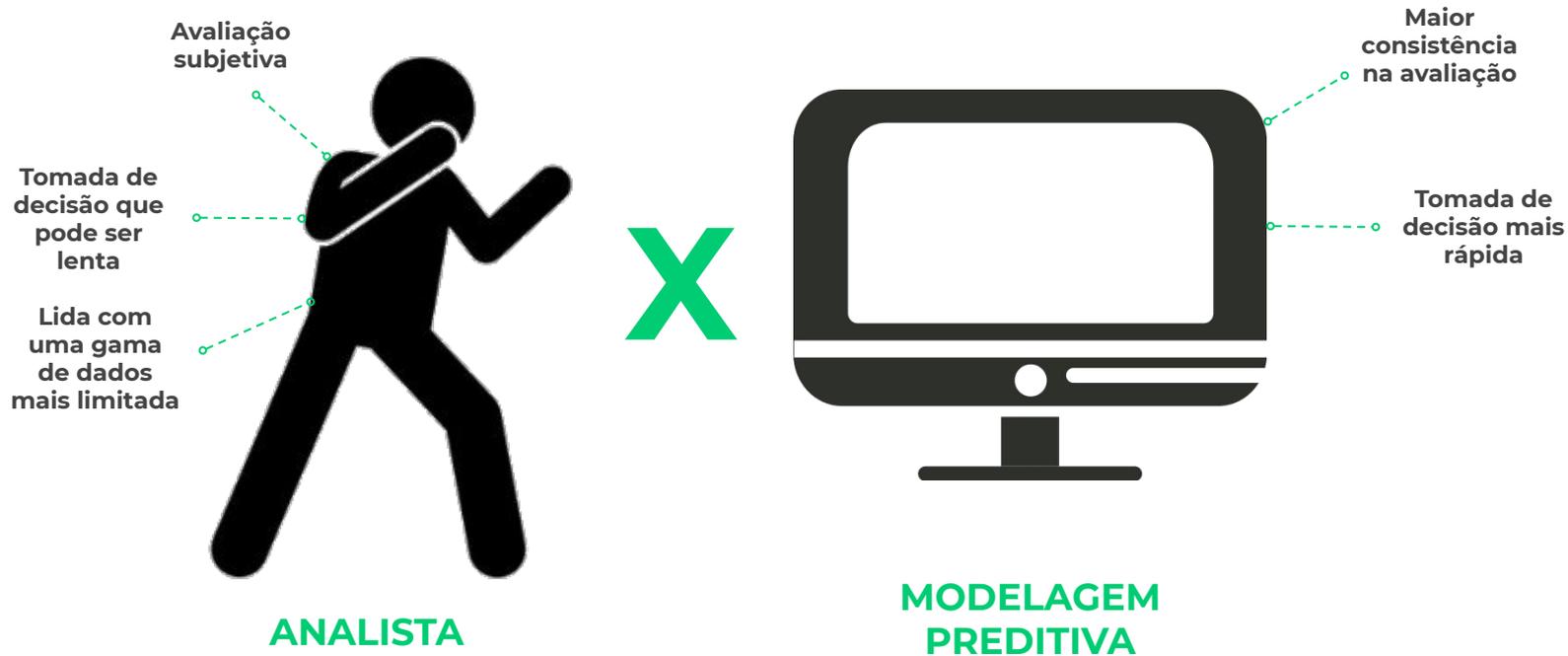
Risco de Crédito | Panorama Geral

Como o risco de crédito pode ser avaliado?



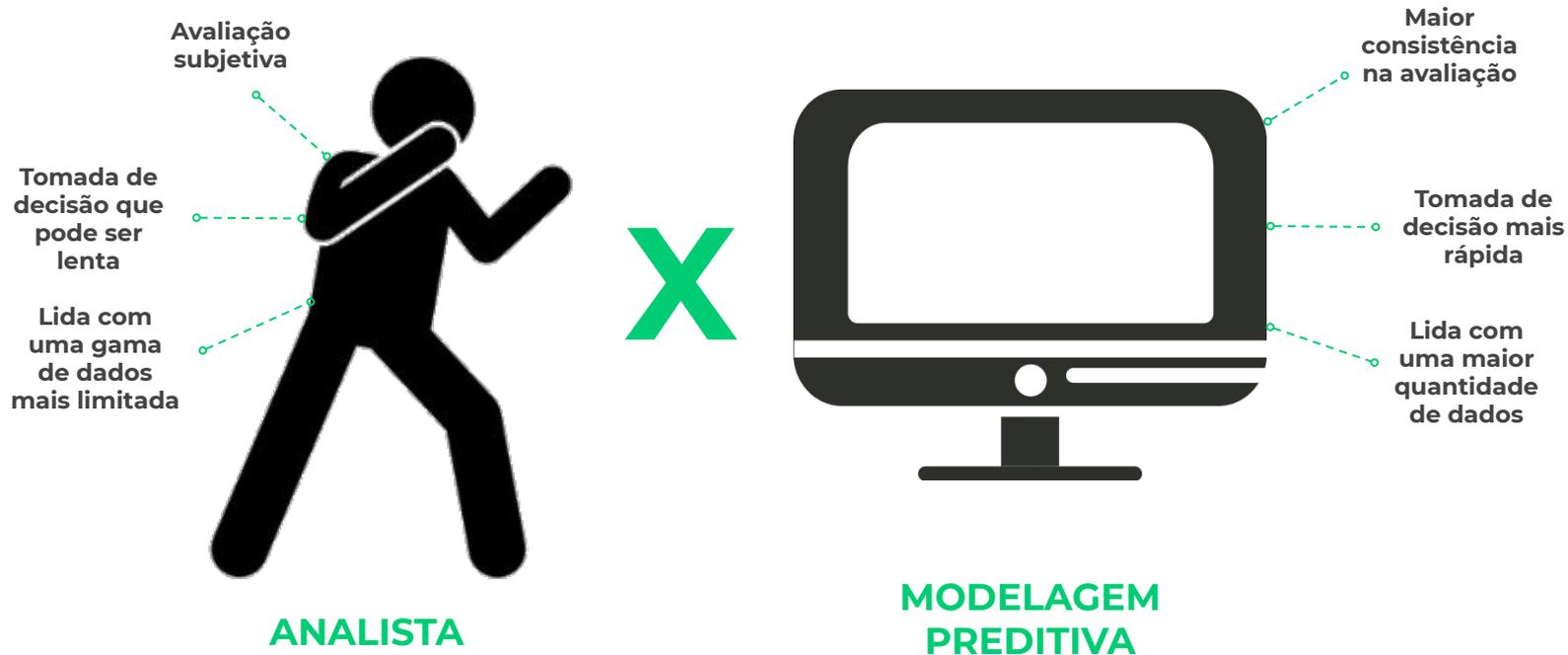
Risco de Crédito | Panorama Geral

Como o risco de crédito pode ser avaliado?



Risco de Crédito | Panorama Geral

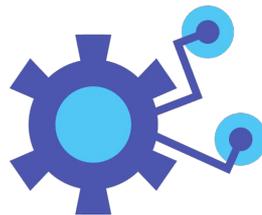
Como o risco de crédito pode ser avaliado?



Modelagem Preditiva | Quem usa?

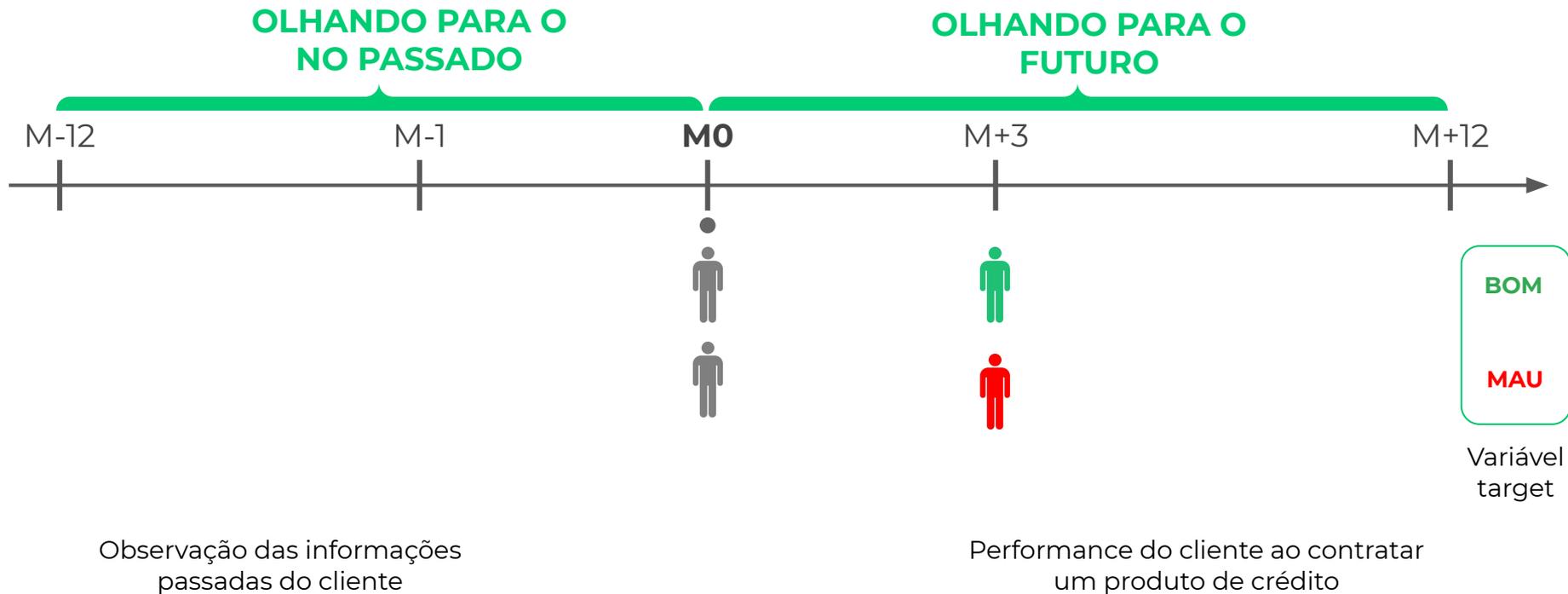
Os Modelos preditivos usados em crédito, como o credit scoring, são ferramentas muito úteis por serem usadas para gerenciar risco avaliando a **capacidade pagamento** de alguém. Hoje em dia, a modelagem preditiva é amplamente usada em lugares como:

- Bancos;
- Seguradoras;
- Varejo;
- Empresas de financiamento de viagens;
- Empresas de telecomunicações.

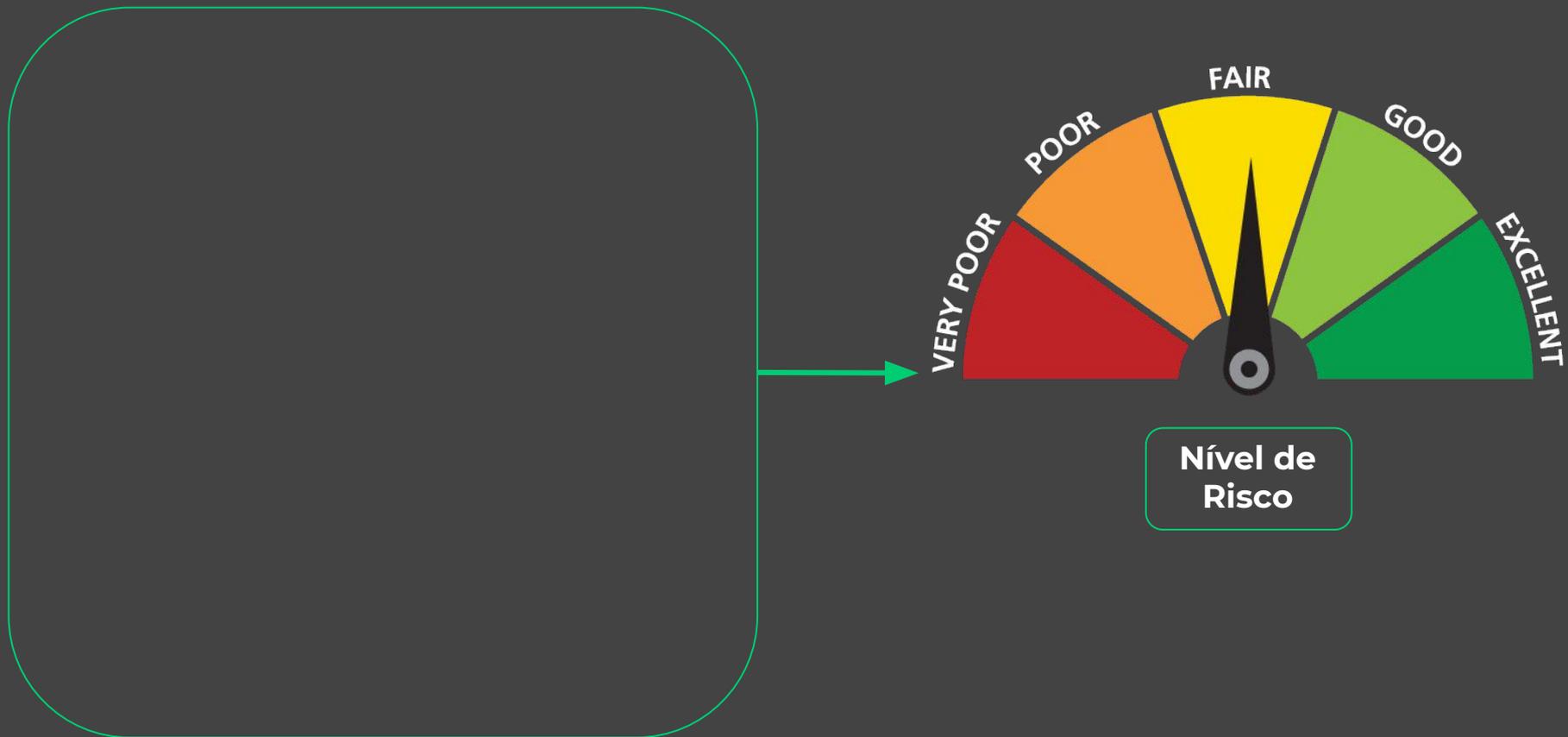


Modelagem Preditiva | Como funciona?

O modelo preditivo é baseado em um algoritmo de machine learning que é treinado com um histórico de informações do cliente.



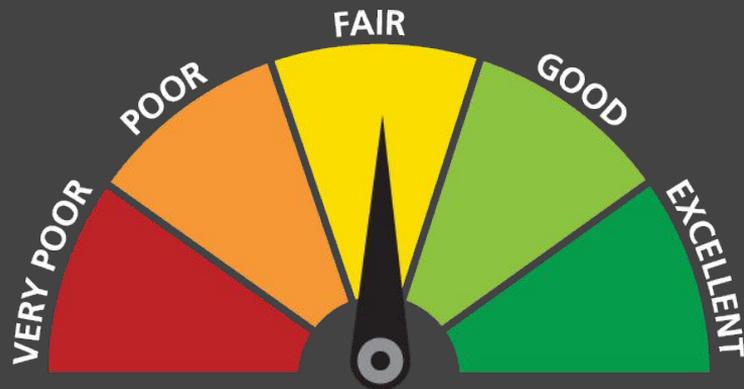
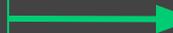
Modelagem Preditiva | Como funciona?



Modelagem Preditiva | Como funciona?



consultas em
seu cpf



Nível de
Risco

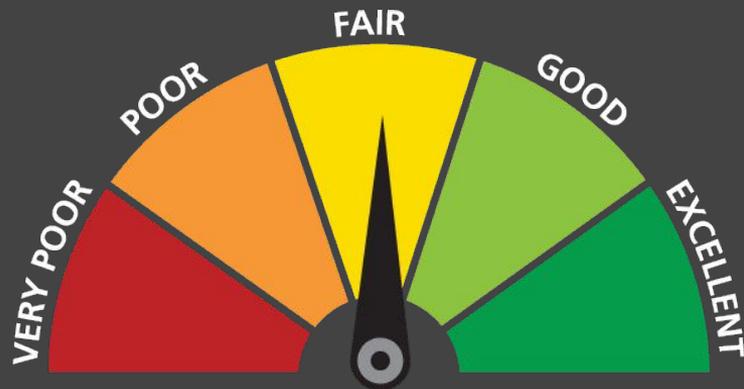
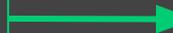
Modelagem Preditiva | Como funciona?



consultas em
seu cpf



histórico de
pagamentos



Nível de
Risco

Modelagem Preditiva | Como funciona?



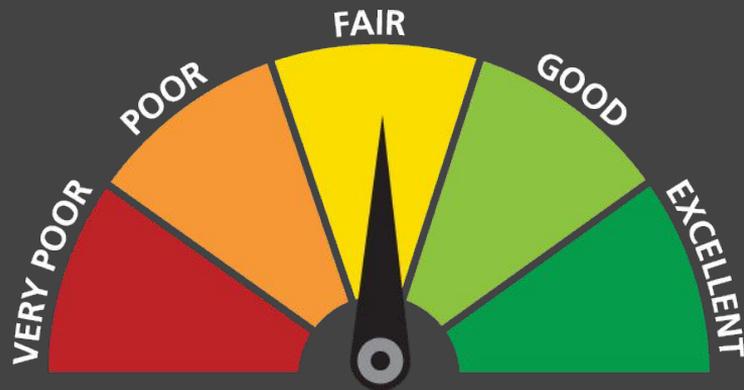
consultas em
seu cpf



histórico de
pagamentos



procura por
crédito



Nível de
Risco

Modelagem Preditiva | Como funciona?



consultas em
seu cpf



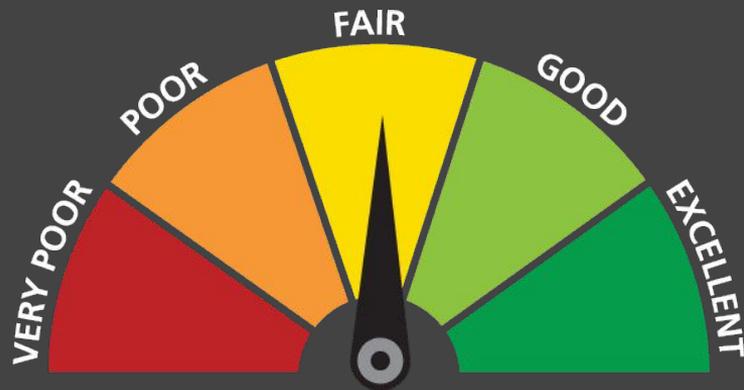
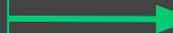
histórico de
pagamentos



procura por
crédito



histórico de
crédito



Nível de
Risco

Modelagem Preditiva | Como funciona?



consultas em
seu cpf



histórico de
pagamentos



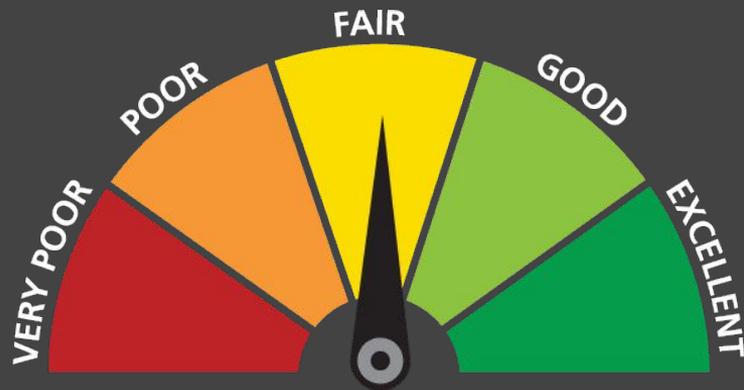
procura por
crédito



histórico de
crédito



dívidas



Nível de
Risco

Ciclo de Crédito | Como funciona?

Ciclo de Crédito | Como funciona?

ELABORAÇÃO

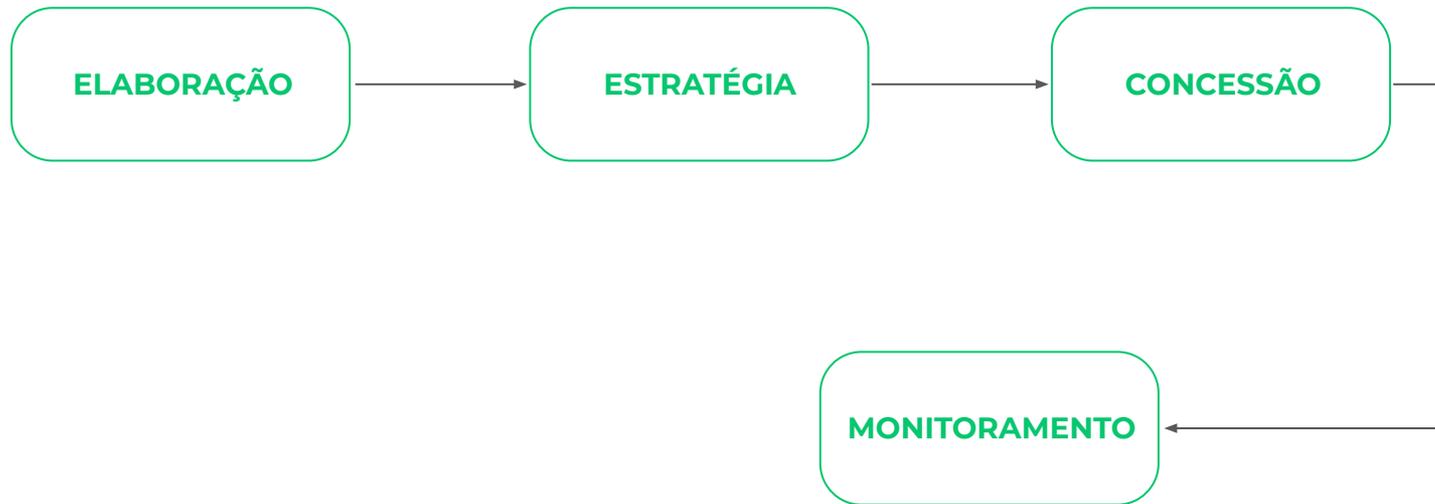
Ciclo de Crédito | Como funciona?



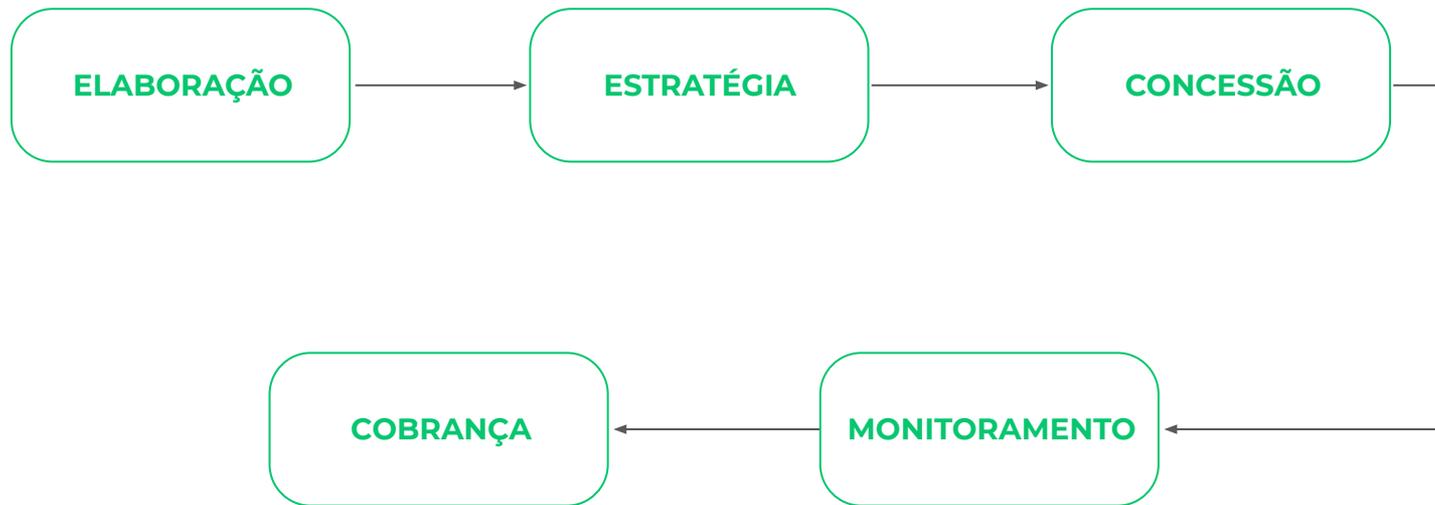
Ciclo de Crédito | Como funciona?



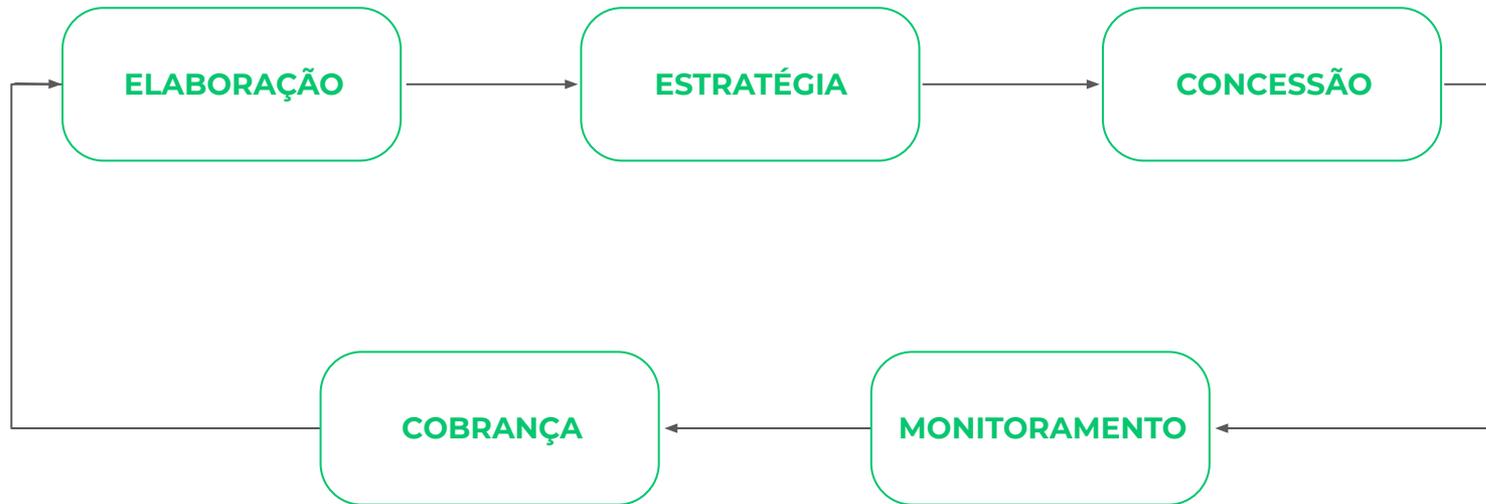
Ciclo de Crédito | Como funciona?



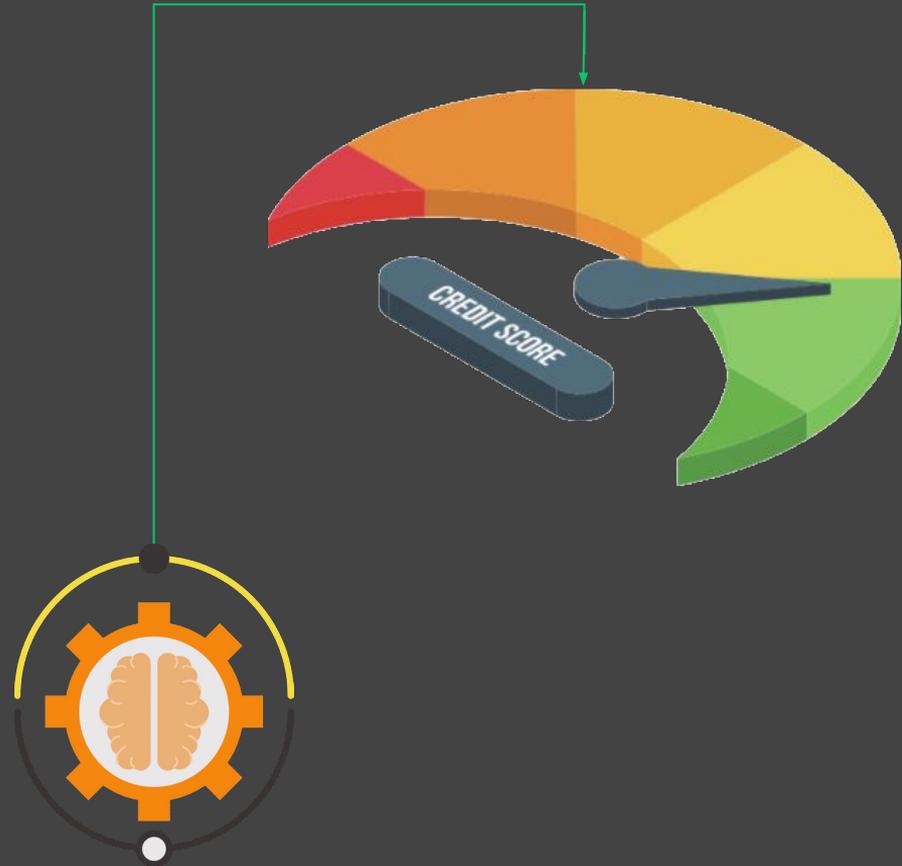
Ciclo de Crédito | Como funciona?



Ciclo de Crédito | Como funciona?



Principais modelos utilizados em **CREDIT SCORING**



Modelos em Credit Scoring | Panorama Geral

A maioria dos modelos de credit scoring são desenvolvidos usando modelos de classificação para estimar a probabilidade de inadimplência. Duas técnicas comuns são regressão logística binária e árvores de decisão.



Modelos em Credit Scoring | Panorama Geral

A maioria dos modelos de credit scoring são desenvolvidos usando modelos de classificação para estimar a probabilidade de inadimplência. Duas técnicas comuns são regressão logística binária e árvores de decisão.

A regressão logística binária é o método mais comum. É um **algoritmo de classificação** usado para estimar o **probabilidade de uma resposta binária** com base em um ou mais variáveis ou recursos preditores.

Modelos em Credit Scoring | Panorama Geral

A maioria dos modelos de credit scoring são desenvolvidos usando modelos de classificação para estimar a probabilidade de inadimplência. Duas técnicas comuns são regressão logística binária e árvores de decisão.

A regressão logística binária é o método mais comum. É um **algoritmo de classificação** usado para estimar o **probabilidade de uma resposta binária** com base em um ou mais variáveis ou recursos preditores.

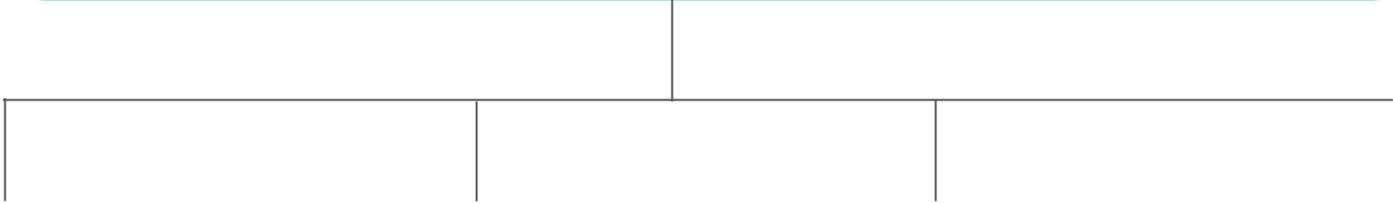
Árvores de classificação funcionam dividindo dados repetidamente em grupos contrastantes para identificar grupos com uma proporção maior da variável-alvo. Eles testam todas as combinações possíveis entre as variáveis de entrada para identificar **como as melhores variáveis se combinam para explicar o resultado**.

Principais métodos de seleção de variáveis em **CREDIT SCORING**



Seleção de variáveis em Credit Scoring | Panorama Geral

Antes de fazermos nosso modelo, precisamos ter um grande cuidado com a seleção de variáveis. Nem toda variável disponível é boa para o modelo que queremos construir.



Seleção de variáveis em Credit Scoring | Panorama Geral

Antes de fazermos nosso modelo, precisamos ter um grande cuidado com a seleção de variáveis. Nem toda variável disponível é boa para o modelo que queremos construir.

Variáveis com proporção de missings e/ou zeros maior que um threshold.

Seleção de variáveis em Credit Scoring | Panorama Geral

Antes de fazermos nosso modelo, precisamos ter um grande cuidado com a seleção de variáveis. Nem toda variável disponível é boa para o modelo que queremos construir.

Variáveis com proporção de missings e/ou zeros maior que um threshold.

Correlação: com o target e entre as variáveis.

Seleção de variáveis em Credit Scoring | Panorama Geral

Antes de fazermos nosso modelo, precisamos ter um grande cuidado com a seleção de variáveis. Nem toda variável disponível é boa para o modelo que queremos construir.

Variáveis com proporção de missings e/ou zeros maior que um threshold.

Correlação: com o target e entre as variáveis.

Information Value.

$$IV = \sum (\% \text{ of non-events} - \% \text{ of events}) * WOE$$

$$WOE = \ln(\% \text{ of non-events} \div \% \text{ of events})$$

Seleção de variáveis em Credit Scoring | Panorama Geral

Antes de fazermos nosso modelo, precisamos ter um grande cuidado com a seleção de variáveis. Nem toda variável disponível é boa para o modelo que queremos construir.

Variáveis com proporção de missings e/ou zeros maior que um threshold.

Correlação: com o target e entre as variáveis.

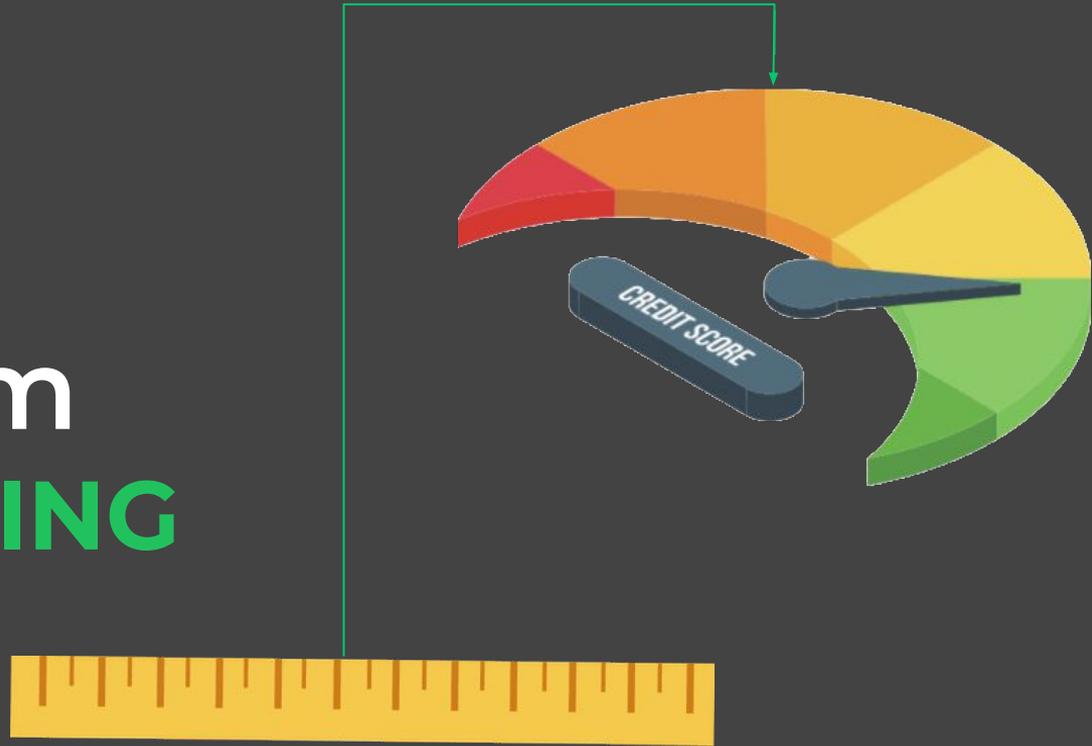
Information Value.

$$IV = \sum (\% \text{ of non-events} - \% \text{ of events}) * WOE$$

$$WOE = \ln(\% \text{ of non-events} \div \% \text{ of events})$$

Boruta: random forest com aleatorização das features.

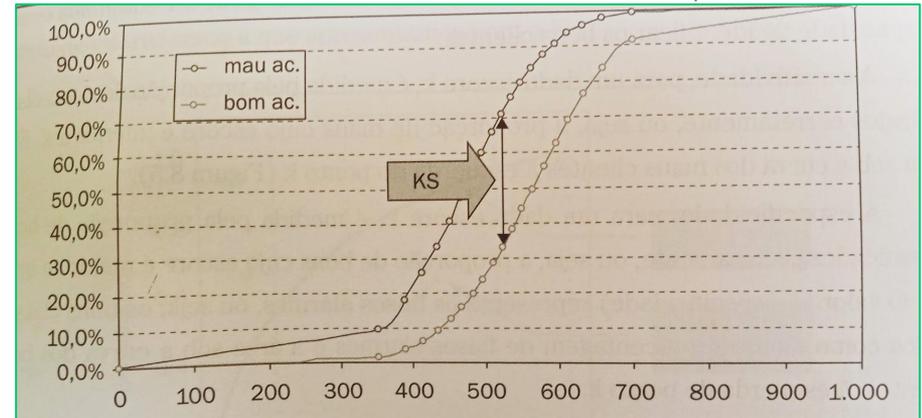
Principais métricas utilizadas em **CREDIT SCORING**



Métricas em Credit Scoring | Panorama Geral

Após fazermos nosso modelo, precisamos ter métricas que nos digam o quão bom (ou ruim) ele está.

KS (Kolmogorov-Smirnov): Mede a distância entre as curvas de distribuição de maus acumulados e de bons acumulados, sendo o KS a maior distância entre elas.

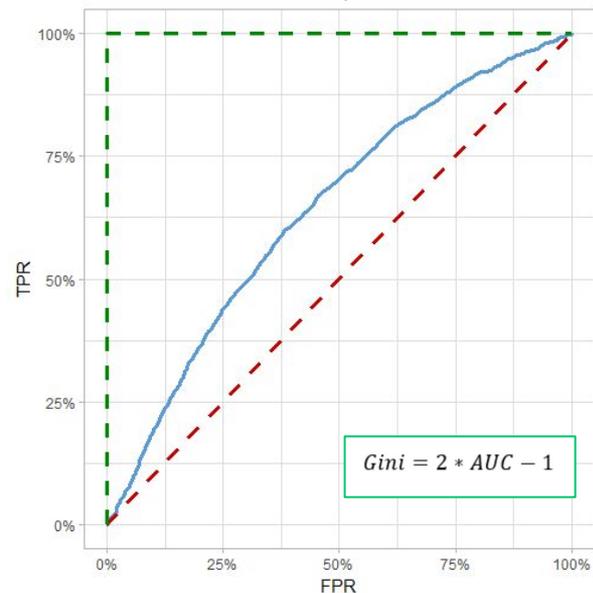


Métricas em Credit Scoring | Panorama Geral

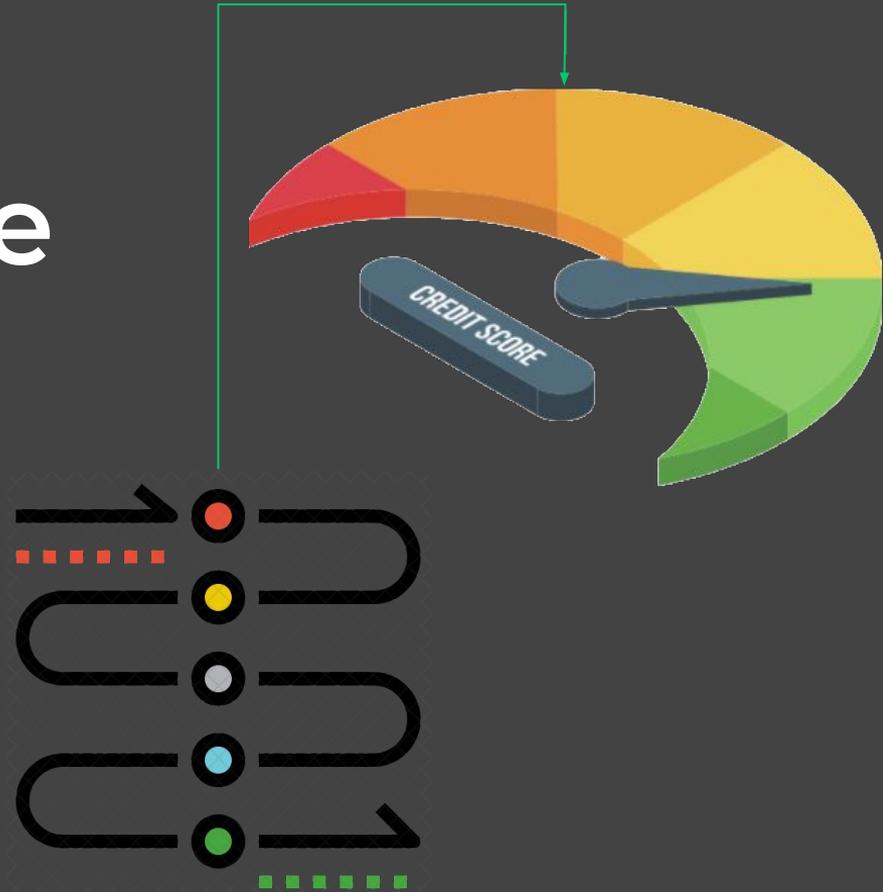
Após fazermos nosso modelo, precisamos ter métricas que nos digam o quão bom (ou ruim) ele está.

O AUC (Area Under Curve) é a área delimitada pela curva ROC. Um classificador perfeito tem AUC = 1 e um classificador completamente aleatório tem AUC = 0,5. Normalmente, o seu modelo irá pontuar em algum ponto intermediário.

O coeficiente de Gini tem como sua finalidade normalizar a AUC para que um classificador aleatório pontue 0 e um classificador perfeito pontue 1.



Passo a passo de um modelo de **CREDIT SCORING**



Risco de Crédito | Panorama Geral

Como fazer o seu modelo de **credit scoring**

```
graph TD; A[Como fazer o seu modelo de credit scoring] --- B[ ]; B --- C[ ]; B --- D[ ]; B --- E[ ]; B --- F[ ]; B --- G[ ]
```

Risco de Crédito | Panorama Geral

Como fazer o seu modelo de **credit scoring**

```
graph TD; A[Como fazer o seu modelo de credit scoring] --- B[Planejamento e definições]; A --- C[ ]; A --- D[ ]; A --- E[ ]; A --- F[ ]
```

Planejamento e definições

Risco de Crédito | Panorama Geral

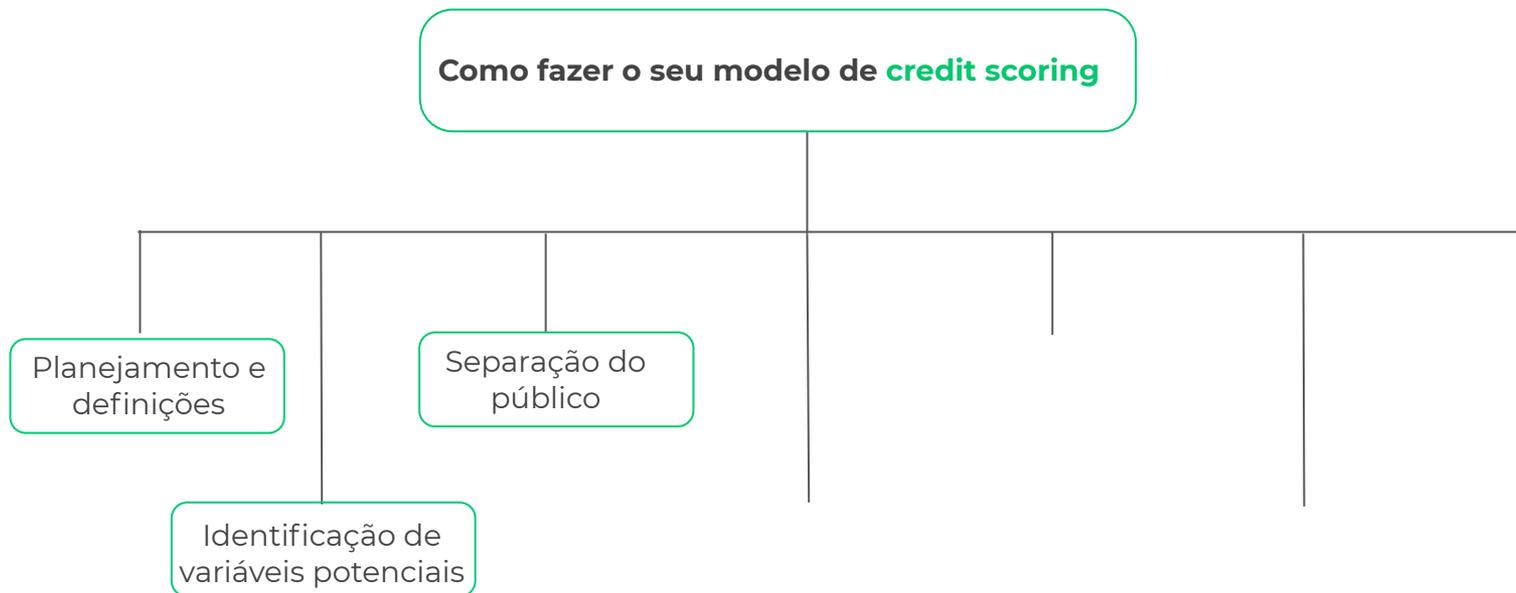
Como fazer o seu modelo de **credit scoring**

```
graph TD; A[Como fazer o seu modelo de credit scoring] --- B[Planejamento e definições]; A --- C[Identificação de variáveis potenciais]; A --- D[ ]; A --- E[ ]; A --- F[ ]; A --- G[ ];
```

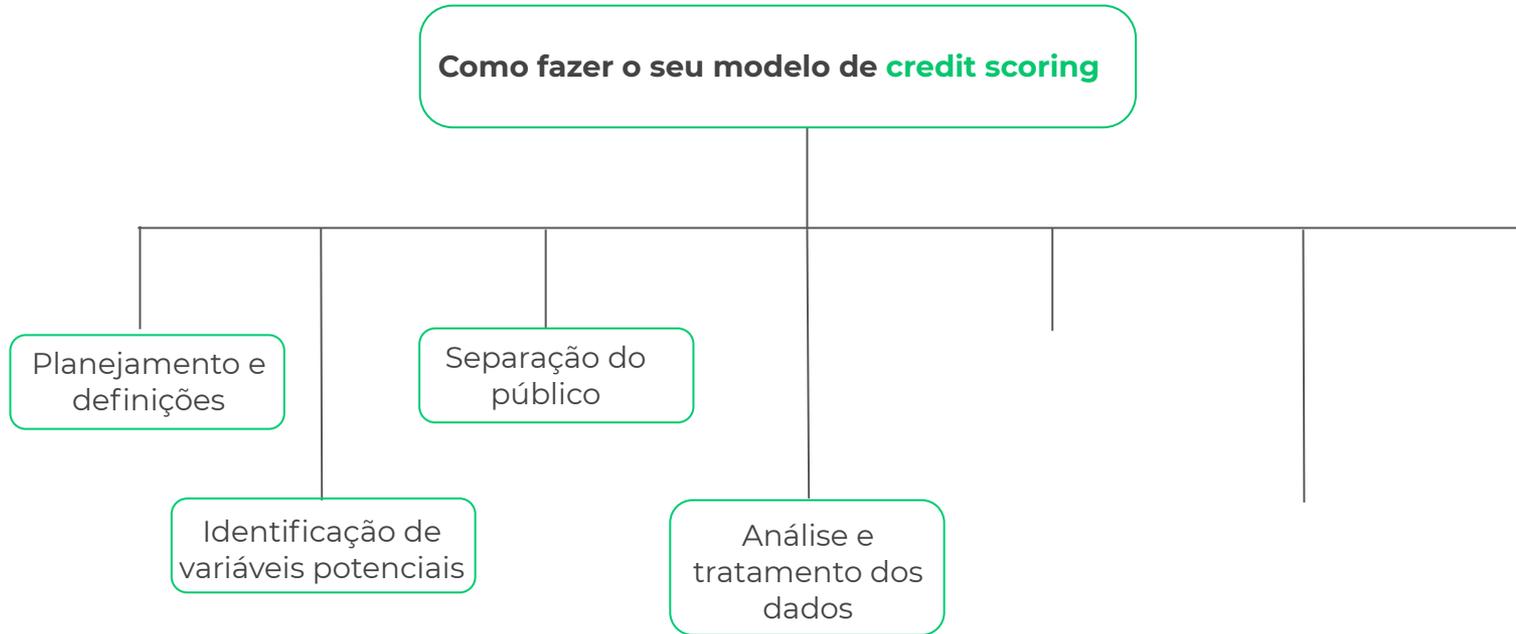
Planejamento e definições

Identificação de variáveis potenciais

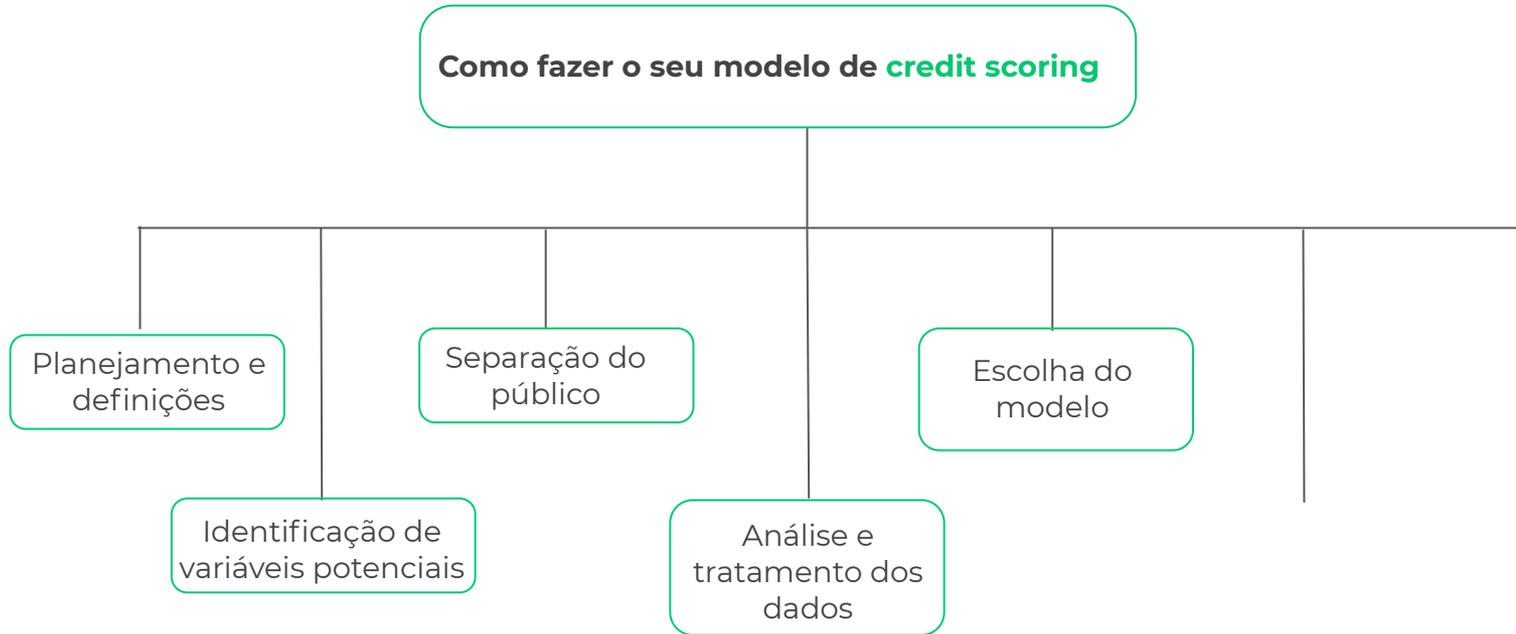
Risco de Crédito | Panorama Geral



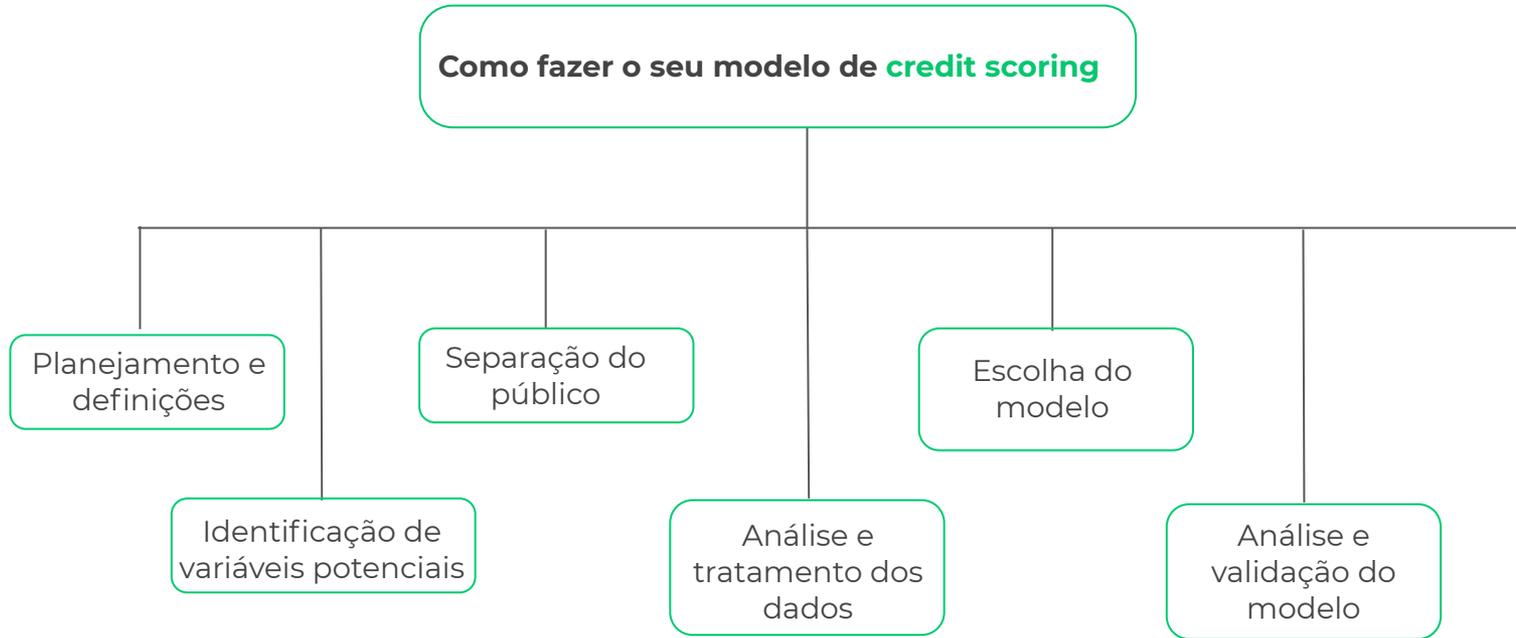
Risco de Crédito | Panorama Geral



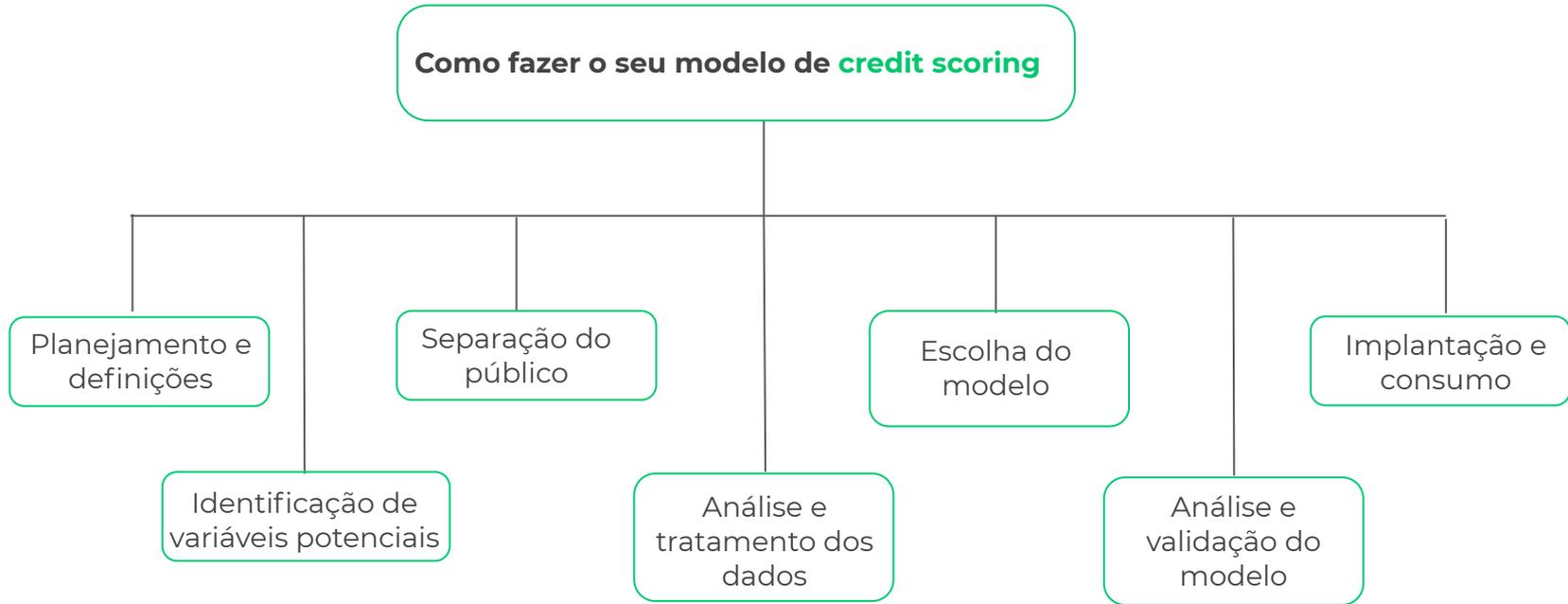
Risco de Crédito | Panorama Geral



Risco de Crédito | Panorama Geral



Risco de Crédito | Panorama Geral



Consultando e
entendendo
seu credit
score na



Passo 1 | Entendendo seu score da Serasa

Entre no link: <https://www.serasa.com.br/area-cliente/saude-financeira>

Passo 1 | Entendendo seu score da Serasa

Entre no link: <https://www.serasa.com.br/area-cliente/saude-financeira>

Entenda sua saúde financeira

Seu score está regular



Seu score aumentou



Seu Score subiu 10 pontos

[Ver histórico](#)

Sua Faixa atual



Faixa de score: 400 - 599



Em média 5 a cada 10 pessoas com a pontuação nessa faixa de score pagam suas contas em dia, por isso pode ser um pouco difícil conseguir crédito.

[Ver todas as faixas](#)

Passo 2 | Entendendo seu score da Serasa

Entenda sua saúde financeira

Seu score está regular



Seu score aumentou



Seu Score subiu 10 pontos

[Ver histórico](#)

Sua Faixa atual



Faixa de score: 400 - 599



Em média 5 a cada 10 pessoas com a pontuação nessa faixa de score pagam suas contas em dia, por isso pode ser um pouco difícil conseguir crédito.

[Ver todas as faixas](#)

Passo 2 | Entendendo seu score da Serasa

Entenda sua saúde financeira

Seu score está regular



Seu score aumentou



Seu Score subiu 10 pontos

[Ver histórico](#)

Sua Faixa atual



Faixa de score: 400 - 599



Em média 5 a cada 10 pessoas com a pontuação nessa faixa de score pagam suas contas em dia, por isso pode ser um pouco difícil conseguir crédito.

[Ver todas as faixas](#)



21 de Março

462 ↑ Seu score aumentou 10 pontos

17 de Março

452 ↓ Seu score diminuiu 17 pontos

13 de Março

469 ↑ Seu score aumentou 3 pontos

11 de Março

466 ↓ Seu score diminuiu 6 pontos

Passo 2 | Entendendo seu score da Serasa

Entenda sua saúde financeira

Seu score está regular



Seu score aumentou



Seu Score subiu 10 pontos

[Ver histórico](#)

Sua Faixa atual



Faixa de score: 400 - 599



Em média 5 a cada 10 pessoas com a pontuação nessa faixa de score pagam suas contas em dia, por isso pode ser um pouco difícil conseguir crédito.

[Ver todas as faixas](#)

Passo 2 | Entendendo seu score da Serasa

Entenda sua saúde financeira

Seu score está regular



Seu score aumentou



[Ver histórico](#)

Sua Faixa atual



Em média 5 a cada 10 pessoas com a pontuação nessa faixa de score pagam suas contas em dia, por isso pode ser um pouco difícil conseguir crédito.

[Ver todas as faixas](#)

Faixas de Score

Sua Faixa atual



Em média 5 a cada 10 pessoas com a pontuação nessa faixa de score pagam suas contas em dia, por isso pode ser um pouco difícil conseguir crédito.

Próxima faixa



Em média 7 a cada 10 pessoas com a pontuação nessa faixa de score pagam suas contas em dia, por isso pode ser mais fácil conseguir crédito.

Outras faixas



Em média 3 a cada 10 pessoas com a pontuação nessa faixa de score pagam suas contas em dia, por isso pode ser mais difícil conseguir crédito.

Passo 3 | Entendendo seu score da Serasa

O que compõe o score?



Histórico financeiro



Compromisso com crédito



Consultas em seu nome



Evolução da sua saúde financeira

Passo 3 | Entendendo seu score da Serasa

O que compõe o score?



Histórico financeiro



Compromisso com crédito



Consultas em seu nome



Evolução da sua saúde financeira

Diagnóstico do seu Score

Consultas para serviço e crédito

8 empresas diferentes consultaram o seu CPF na Serasa no último ano (não considera consultas de operadoras de telefonia e seguradoras)

| Ruim para o seu score

Tempo de uso do crédito

Você utiliza um cartão de crédito que foi contratado há 6 anos e 1 semana (considerando o uso nos últimos 12 meses)

| Bom para o seu score

Pagamento de dívidas

Você regularizou 4 dívidas que ficaram em atraso por mais de 1 mês no último ano

| Precisa de atenção

OBRIGADA!



Contatos:

- @dessafreires
- <https://www.linkedin.com/in/andressafreires/>
- @dessafreires
- @datasciencedescomplicada
- @diversidadedescomplicada