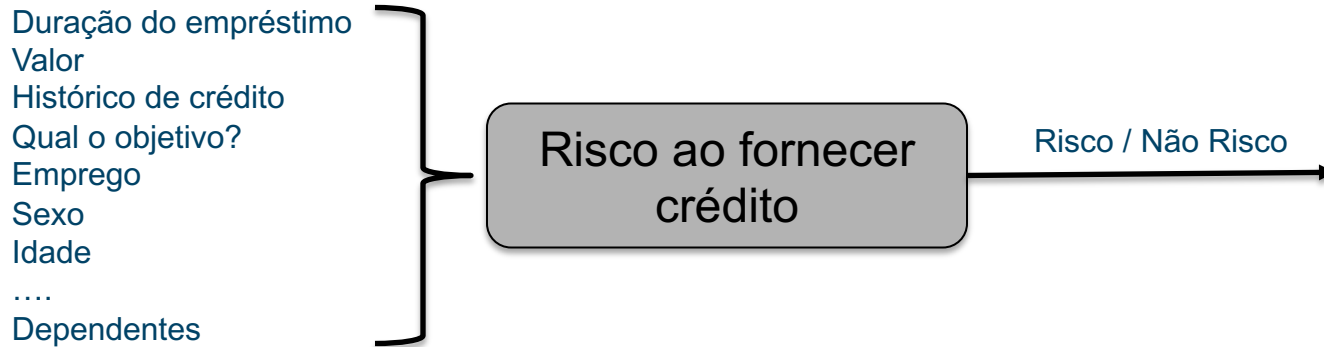


Você sabe explicar a decisão do seu modelo preditivo? Um papo sobre viés e confiança em A.I.

Fabício Barth

Líder Técnico, IBM Data & AI



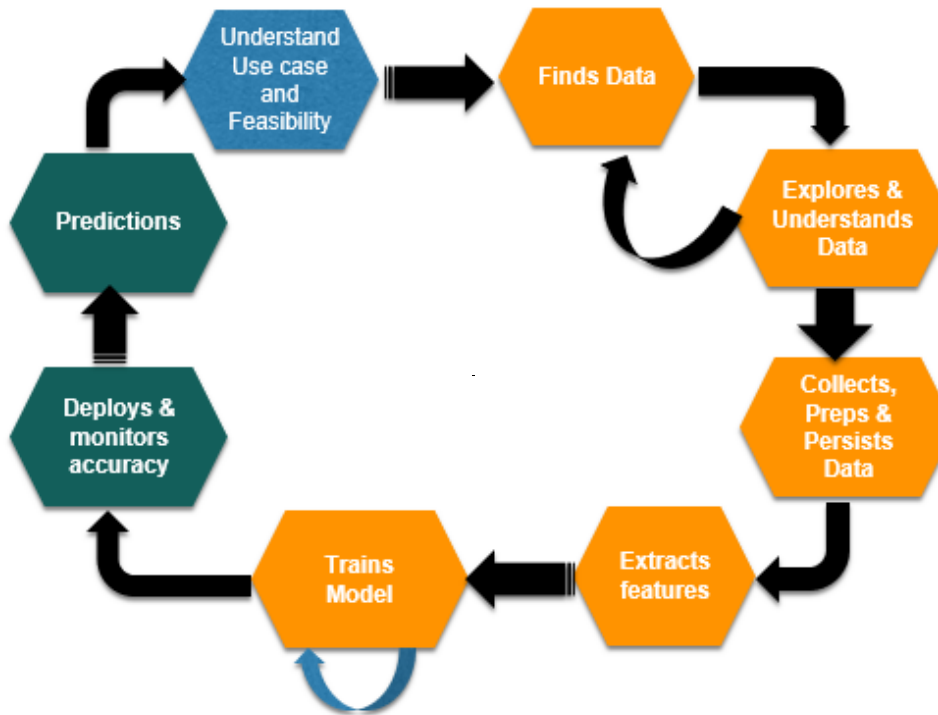


Abordagens para desenvolvimento dos modelos	Características
Construção manual do modelo	
Uso de dados históricos e algoritmos “caixa branca”	
Uso de dados históricos e uso de <i>deep learning</i>	

Abordagens para desenvolvimento dos modelos	Características
Construção manual do modelo	<ul style="list-style-type: none">• Normalmente desenvolvidos com poucos exemplos, poucos atributos e heurísticas de especialistas do domínio• Geralmente são modelos mais conservadores, que optam por uma precisão maior e uma cobertura menor. Ou seja, deixam de “vender” crédito para quem pode pagar.
Uso de dados históricos e algoritmos “caixa branca”	
Uso de dados históricos e uso de <i>deep learning</i>	

Abordagens para desenvolvimento dos modelos	Características
Construção manual do modelo	
Uso de dados históricos e algoritmos “caixa branca”	<ul style="list-style-type: none">• Usam modelos de regressão linear, logística ou árvores de decisão.• Faz uso de dados históricos, com poucos ou muitos exemplos, e diversos atributos.• São modelos que conseguem explicar muito bem o comportamento histórico. No entanto, têm dificuldade para generalização (problema de <i>overfitting</i>).• A forma como o modelo é representado facilita a explicação das decisões do modelo.
Uso de dados históricos e uso de <i>deep learning</i>	

Abordagens para desenvolvimento dos modelos	Características
Construção manual do modelo	
Uso de dados históricos e algoritmos “caixa branca”	
Uso de dados históricos e uso de <i>deep learning</i>	<ul style="list-style-type: none">• Redes neurais com múltiplas camadas, <i>ensemble models</i> (Random Forest) são o “estado da arte” na construção de modelos preditivos.• São modelos que geralmente têm uma capacidade de generalização (acurácia) muito alta.• Mas, são modelos “caixa preta”. Ou seja, não facilita a explicação da tomada de decisão do modelo.



Desejo:

- Desenvolver modelos com alta capacidade de generalização.
- Possam ser facilmente integrados aos processos e soluções da organização

Desafios:

- Como garantir a evolução do modelo?
- Como garantir transparência na tomada de decisões?
- Como reduzir eventuais bias indesejáveis?

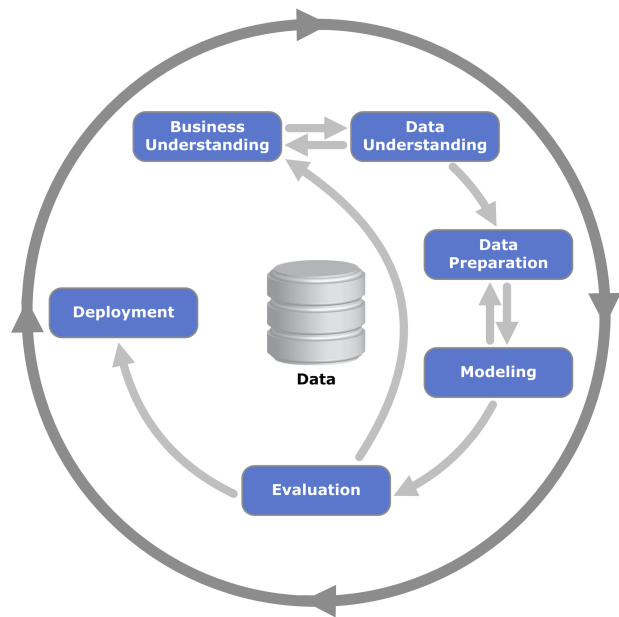


*Article 22
EU GDPR
"Automated
individual
decision-making,
including
profiling"*

Atividades

1. **Criar e implantar modelos com alta capacidade de generalização**
2. **Monitorar as predições e a acurácia do modelo**
3. **Fornecer uma explicação para as predições dos modelos**
4. **Monitorar predições enviesadas e opcionalmente corrigir as mesmas**

Criar e implantar modelos com alta capacidade de generalização



Monitor as predições e a acurácia do modelo

IBM Watson OpenScale | beta

Insights

Deployments

Monitored

1

Accuracy

Alerts

0

Fairness

Alerts

0

Spark German Risk Deployment...

Issues

0

Accuracy

73%

Fairness

96%

Evaluated 9 minutes ago

Accuracy

Accuracy is proportion of correct predictions. [Learn more.](#)

Time frame

Hourly

Daily

Weekly

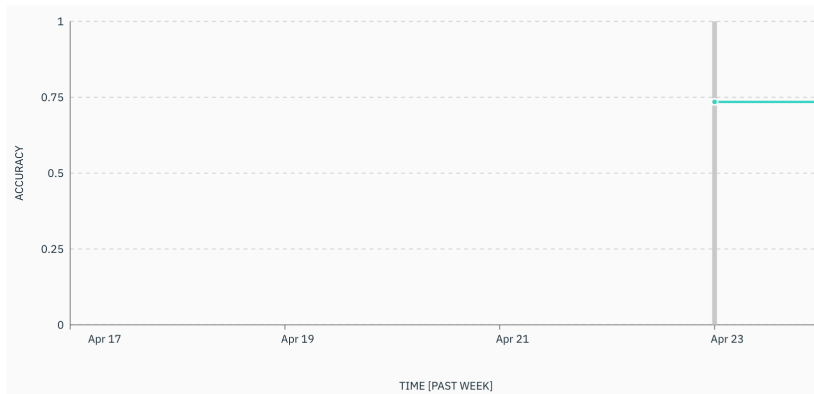
Date range

Past 3 months

Past week

Yesterday

Today



Accuracy

0.73

Tue, Apr 23, 2019, 12:00 AM -03

Schedule

Last Evaluation 12:04 AM -03

Next Evaluation 1:04 AM -03

[Check Accuracy Now](#)

[Add feedback data](#)



Metrics are updated hourly.

Monitor as previsões e a acurácia do modelo

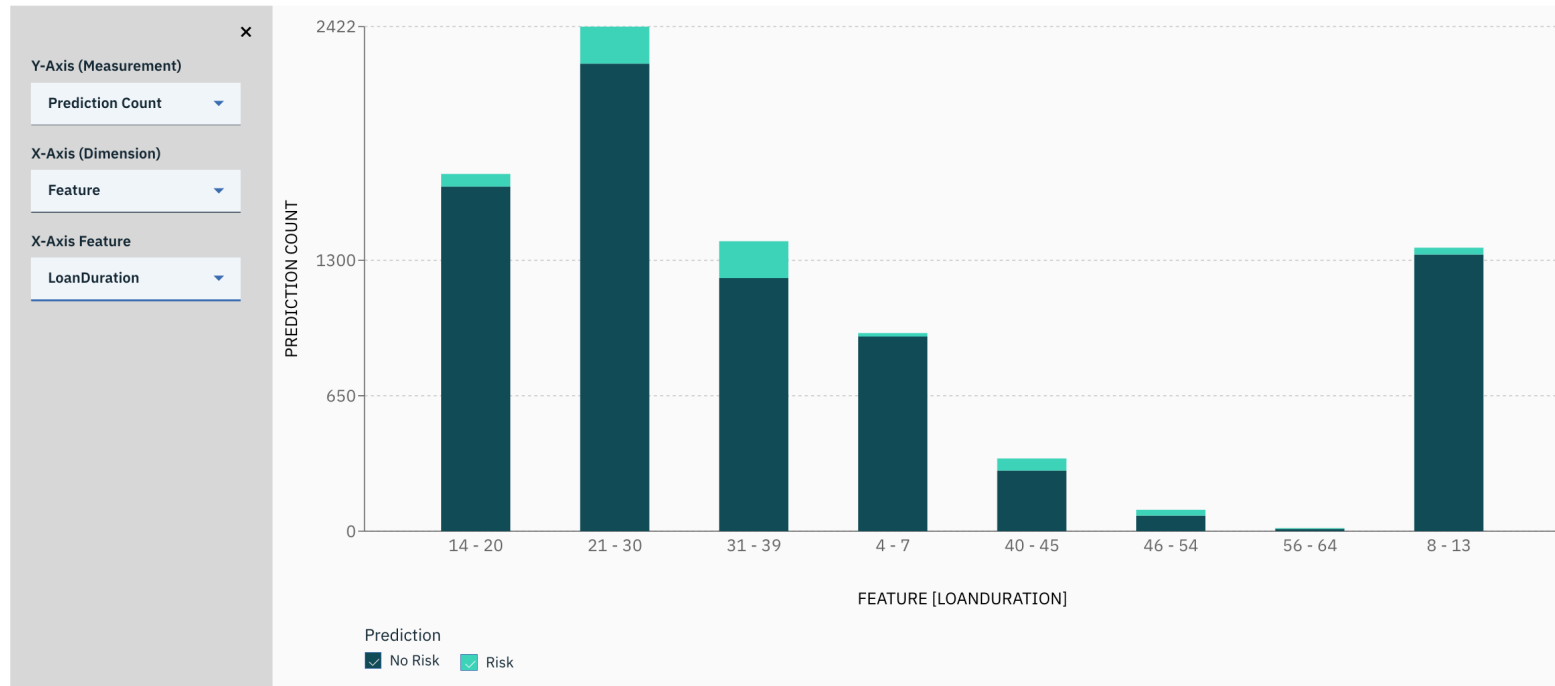
Date range

Past 3 months

Past week

Yesterday

Today



Fornecer uma explicação para as predições dos modelos

IBM Watson OpenScale | beta

● b8314a89976e5f87a3... x
● b8314a89976e5f87a3... x
● b8314a89976e5f87a3... x

Details ⓘ

Transaction: b8314a89976e5f87a3048260c823476f-131
 Deployment: Spark German Risk Deployment - Final
 Model Name: Spark German Risk Model - Final

Minimum changes for another outcome ⓘ

CheckingStatus: greater_200
 Sex: female
 InstallmentPlans: stores

Minimum factors supporting this outcome ⓘ

OthersOnLoan: guarantor
 CheckingStatus: greater_200
 LoanDuration: 41.0



No Risk

CONFIDENCE

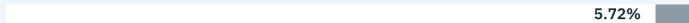
Risk



Factors contributing to **No Risk** confidence level

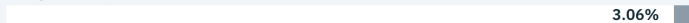
Factors contributing to **Risk** confidence level

LoanAmount: 2825



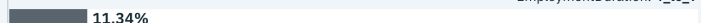
LoanDuration: 41

ForeignWorker: yes

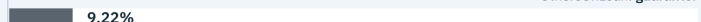


Age: 46

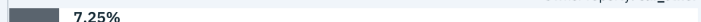
EmploymentDuration: 4_to_7



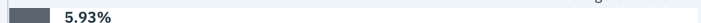
OthersOnLoan: guarantor



OwnsProperty: car_other



ExistingCreditsCount: 2



Monitorar previsões enviesadas e opcionalmente corrigir as mesmas

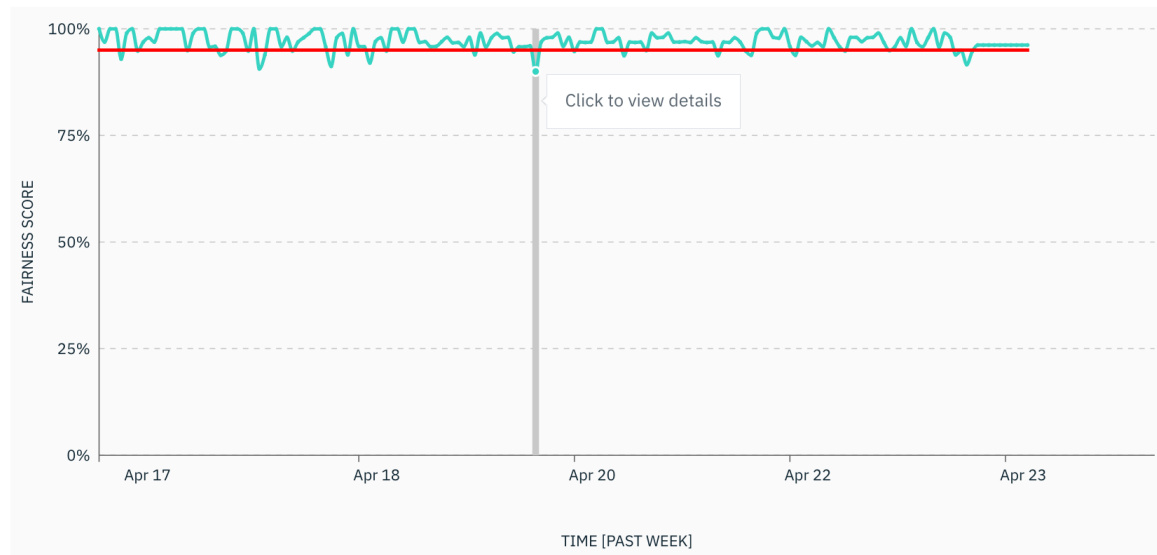
Fairness for Sex

The models propensity to deliver favorable outcomes to one group over another. [Learn more.](#)

Time frame

Hourly
Daily
Weekly
Past 3 months
Past week
Yesterday
Today

Date range



Fairness Score for Sex

90%

5% below threshold

Sat, Apr 20, 2019, 7:00 AM -03

■ Threshold 95%

Monitored Groups

Average	90%
■ female	90%

Monitorar previsões enviesadas e opcionalmente corrigir as mesmas

Spark German Risk Deploy... : Transactions

Dataset ⓘ

- Payload + Perturbed
- Payload
- Training
- Debiased

Monitored Feature

Sex

Date and Time

April 20, 2019

07:00

AM ▾

Monitored groups ⓘ

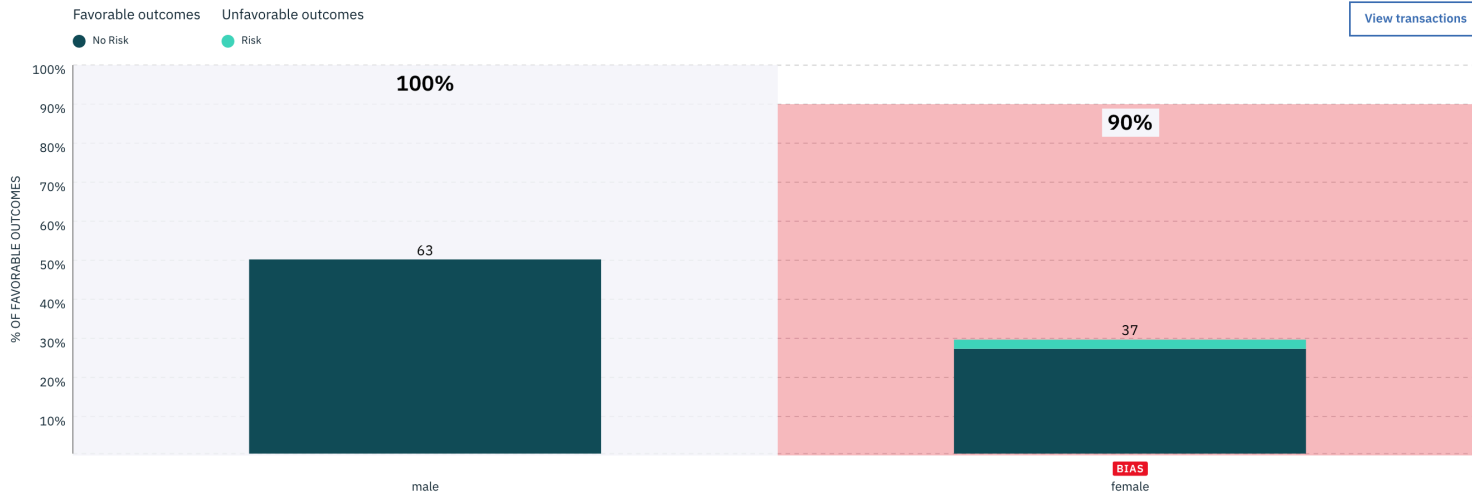
90% of the group female
received the outcome: **No Risk**

Reference groups ⓘ

100% of the group male
received the outcome: **No Risk**

★ Recommendation

Watson OpenScale created a model that is **12% more fair**.



[Payload Data] Feature = Sex

Outras informações

Throughput

The average number of requests per minute.

Time frame

Hourly

Daily

Weekly

Date range

Past 3 months

Past week

Yesterday

Today

