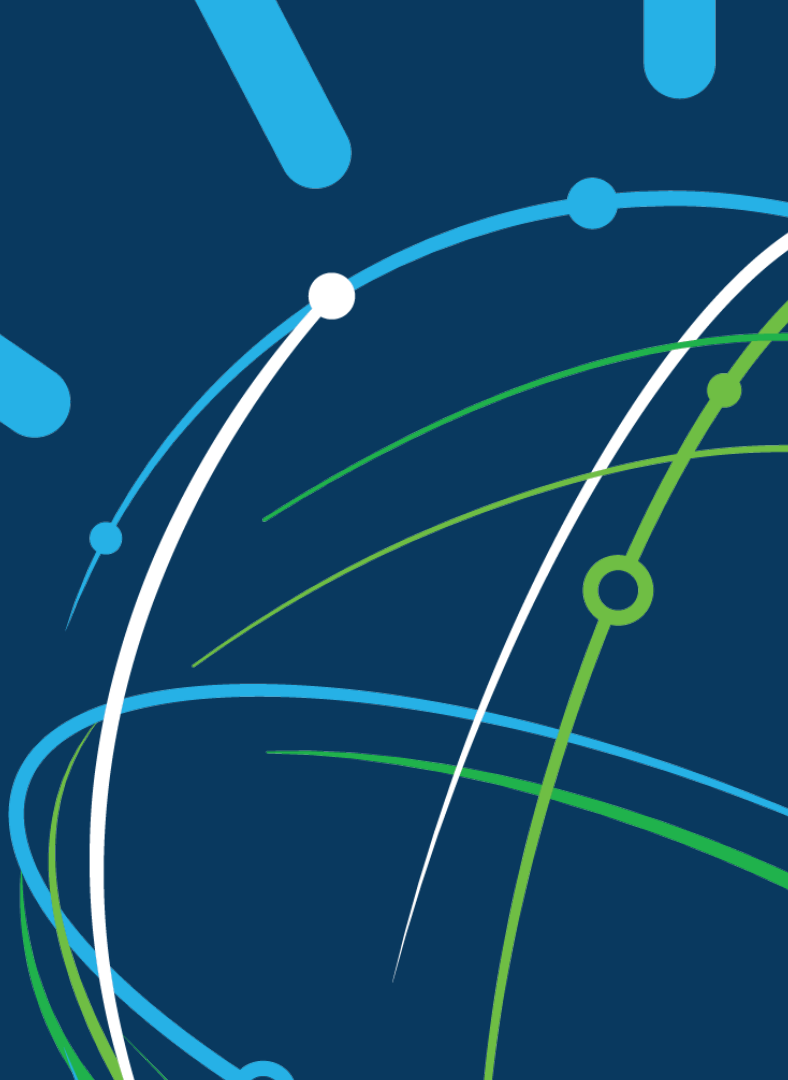
Você sabe explicar a decisão do seu modelo preditivo? Um papo sobre viés e confiança em A.I.

Fabrício Barth
Líder Técnico, IBM Data & AI

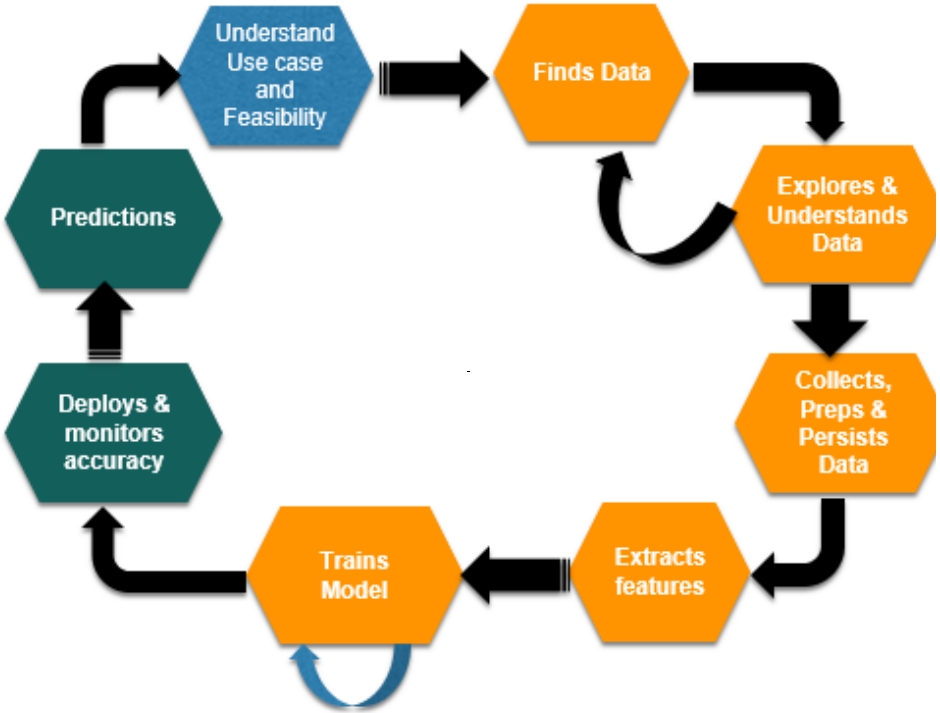


Abordagens para desenvolvimento dos modelos	Características
Construção manual do modelo	
Uso de dados históricos e algoritmos “caixa branca”	
Uso de dados históricos e uso de *deep learning*	

Abordagens para desenvolvimento dos modelos	Características
Construção manual do modelo	• Normalmente desenvolvidos com poucos exemplos, poucos atributos e heurísticas de especialístas 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 *deep learning*	

Abordagens para desenvolvimento dos modelos	Características
Construção manual do modelo	
Uso de dados históricos e algoritmos “caixa branca”	• 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 *overfitting*). • A forma como o modelo é representado facilita a explicação das decisões do modelo.
Uso de dados históricos e uso de *deep learning*	

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 *deep learning*	• Redes neurais com múltiplas camadas, *ensemble models* (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

Watson Studio

TensorFlow

scikit learn

Keras

PYTORCH

Caffe2

Spark MLlib

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

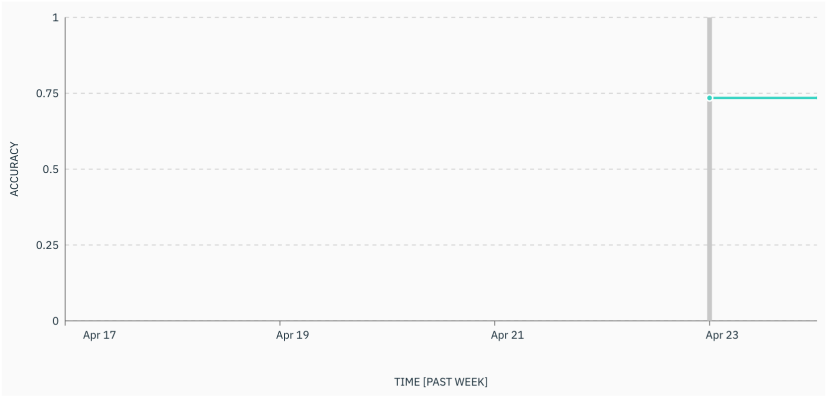
Metrics are updated hourly.

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

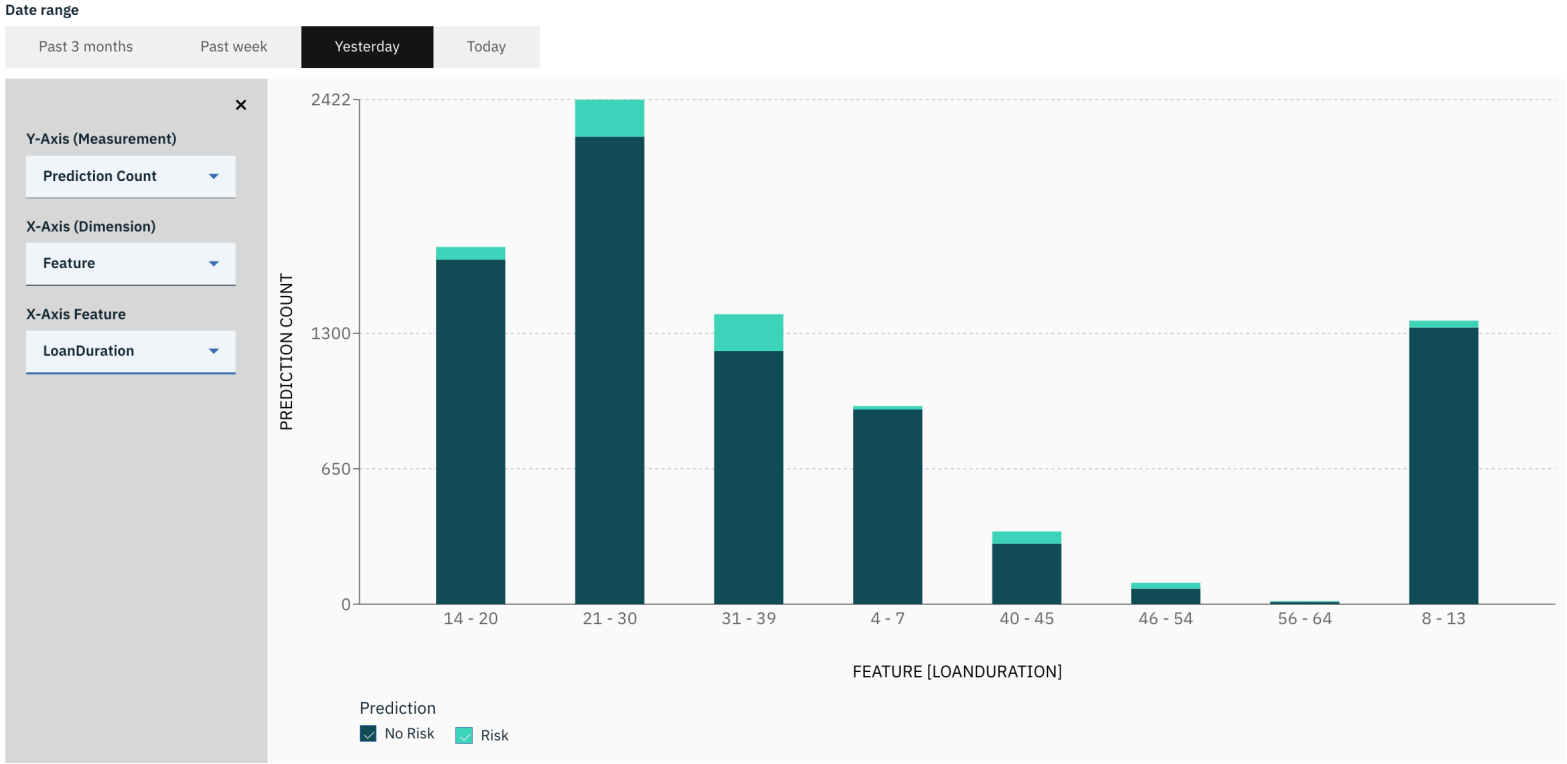
Last Evaluation 12:04 AM -03

Next Evaluation 1:04 AM -03

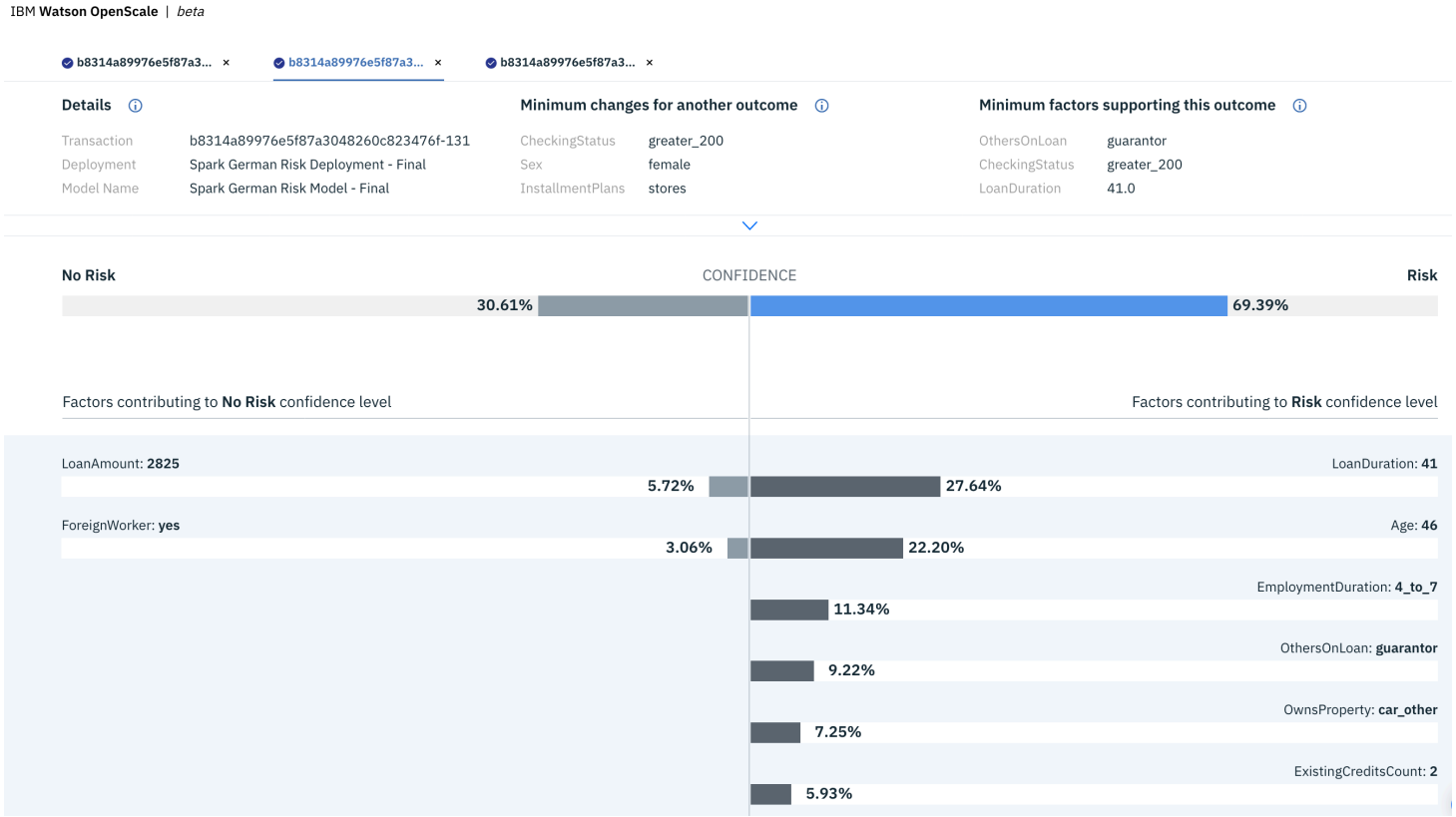
Check Accuracy Now

Add feedback data

Monitor as predições e a acurácia do modelo

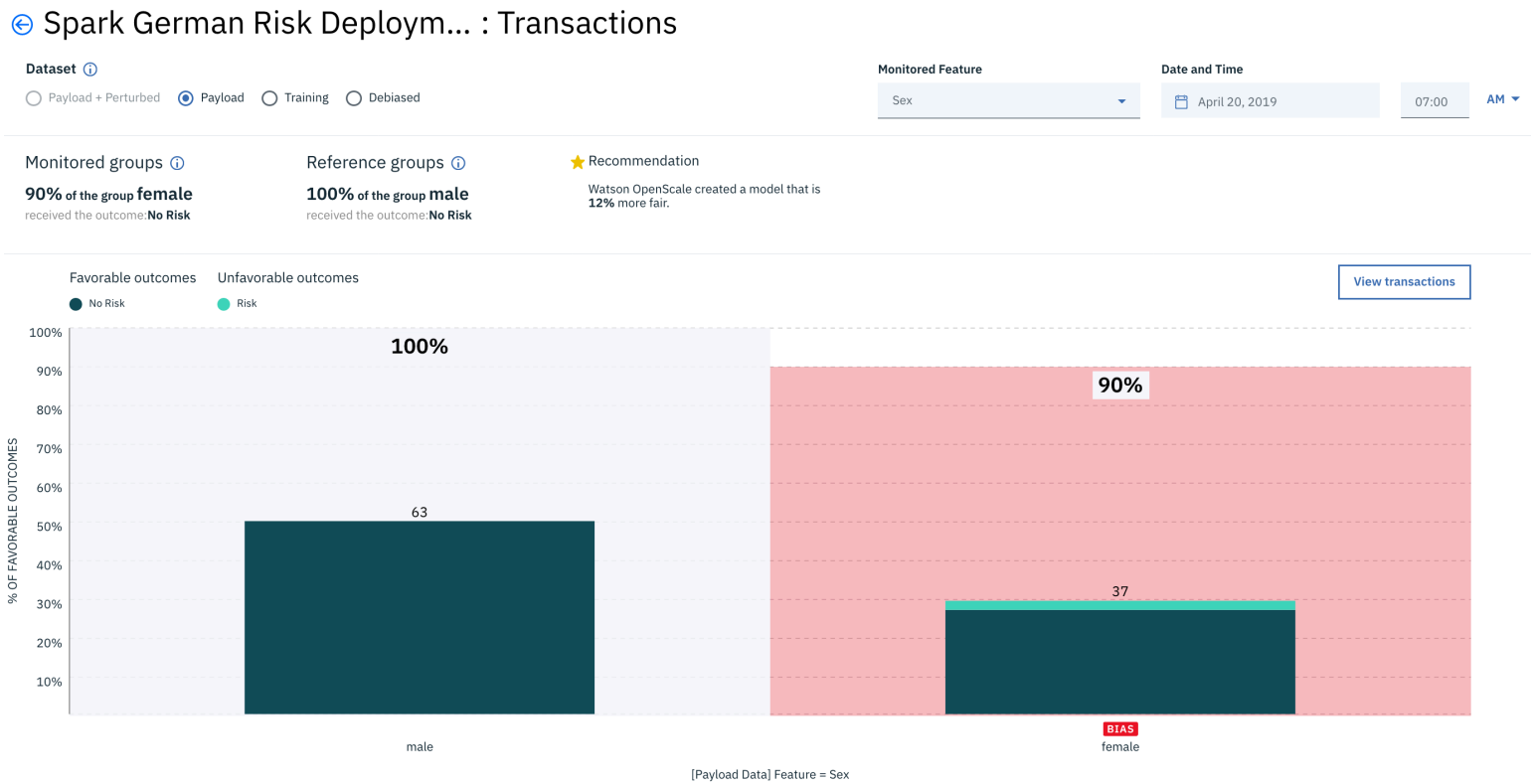


Fornecer uma explicação para as predições dos modelos



Monitorar predições enviesadas e opcionalmente corrigir as mesmas

Monitorar predições enviesadas e opcionalmente corrigir as mesmas



Outras informações

Throughput

The average number of requests per minute.

Time frame

Hourly | **Daily** | Weekly

Date range

Past 3 months | **Past week** | Yesterday | Today