



TDC SUMMIT SÃO PAULO

IA e Dados Sensíveis

O desafio de trabalhar com dados pessoais ou sensíveis e inteligência artificial

March 26, 27

9:00 a.m. - 7:00 p.m.

Centro de Convenções Rebouças





- 27 anos de experiência em TI
- MBA em Arquitetura de Soluções pela FIAP
- Bacharel em Ciência da Computação

Desde criança fascinado por tecnologia, inovação e astronomia!!!



<https://www.linkedin.com/in/rui-romanini-89410133/>

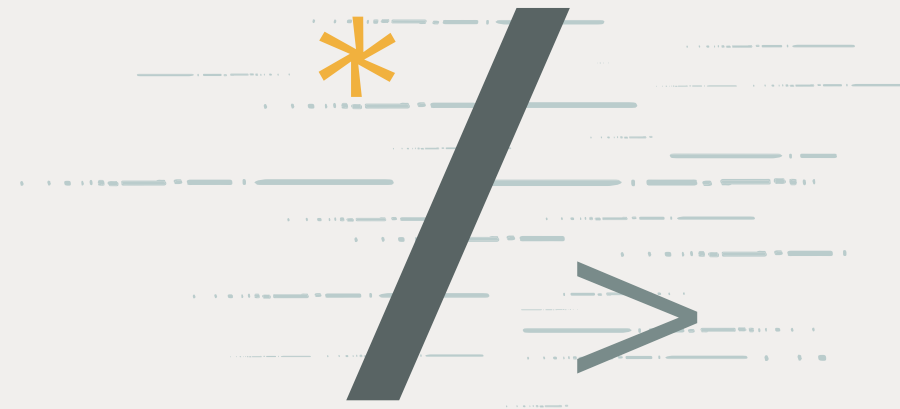
Rui Romanini
Oracle do Brasil
Solution Engineer

Agenda

- 1. Dados Sensíveis e IA**
- 2. Dados Sintéticos**
- 3. Interpretabilidade de modelos**

Dados sensíveis e IA

Technology for business transformation

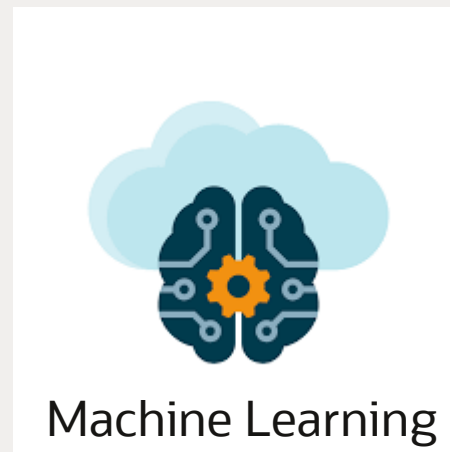


Sobre dados pessoais e dados sensíveis

TDC SUMMIT SÃO PAULO
Technology for business transformation



Teste de Software



LGPD / Privacidade de Dados



Sobre dados pessoais e dados sensíveis

De acordo com a LGPD, dado pessoal é a informação relacionada à pessoa natural identificada – tais como nome, sobrenome, RG e CPF – ou identificável, como no caso dos dados de geolocalização (GPS), endereço IP, identificação de dispositivo etc.



Dado pessoal sensível: dado pessoal sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico.

https://www.planalto.gov.br/ccivil_03/ato2015-2018/2018/lei/L13709compilado.htm

Sobre dados pessoais e dados sensíveis

GDPR

- AI Act
- Aprovação 13/03/2024
- Policiamento Preditivo
- Captação de imagens da internet para treinamento
- Identificação de emoção em locais públicos

No Brasil...

- Projeto de lei nº2338
- Aplicações na área de saúde
- Capacidade de individamento
- Veiculos autonomos

<https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>

<https://www.estadao.com.br/brasil/macaco-eletrico/europa-regulamenta-a-ia-protetendo-a-sociedade-e-sem-ameacar-a-inovacao/>

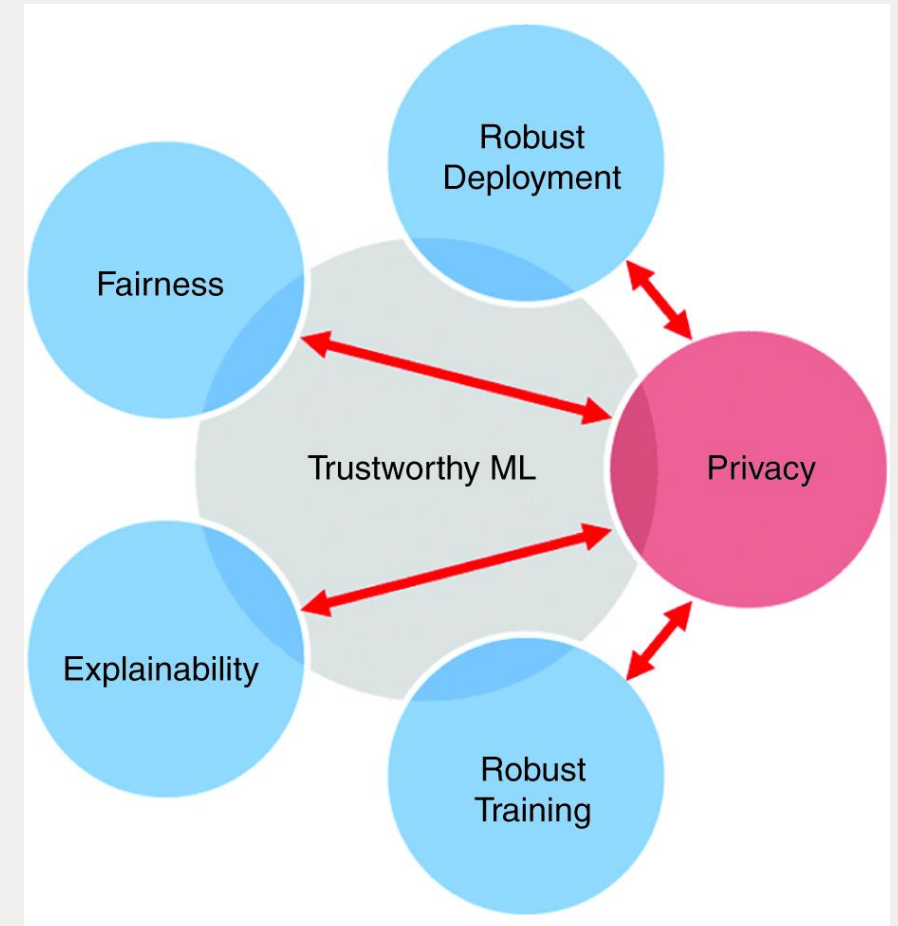
Sobre dados pessoais e dados sensíveis

Data Privacy

Disciplina dedicada a manter os dados seguros contra acessos impróprios, roubo ou perda.

Habilidade de uma pessoa determinar quando, como e qual tipo de dado pessoal pode ser compartilhado.

Como se relacionam com Machine Learning?



OCI Language

Se

- D
- C
- A
- E
- R
- D

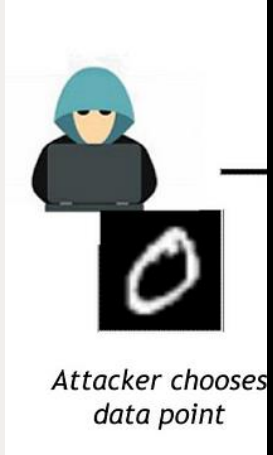
[ht](#)
[ht](#)
[b8](#)

OCI Guardian AI

Biblioteca Pyt

Dado um mod
ou não parte

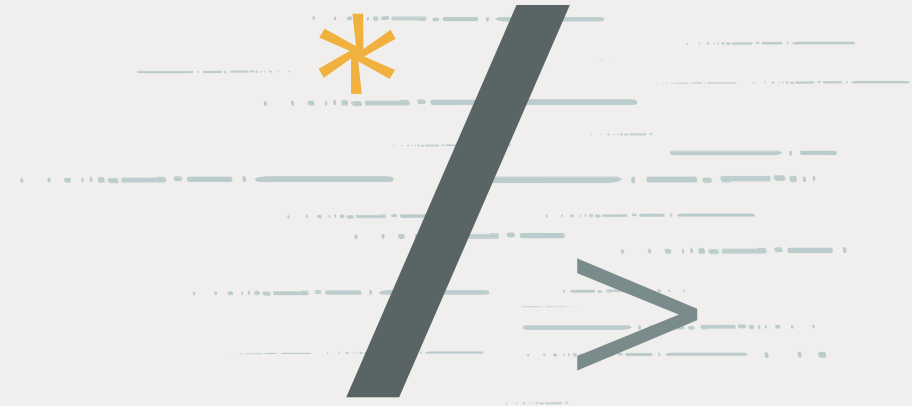
Utiliza se o m



<https://medium.com/@ruiromanini/um-overview-sobre-data-privacy-oracle-guardian-ai-e-oci-language-pii-b873321fd698>

Dados Sintéticos

Technology for business transformation

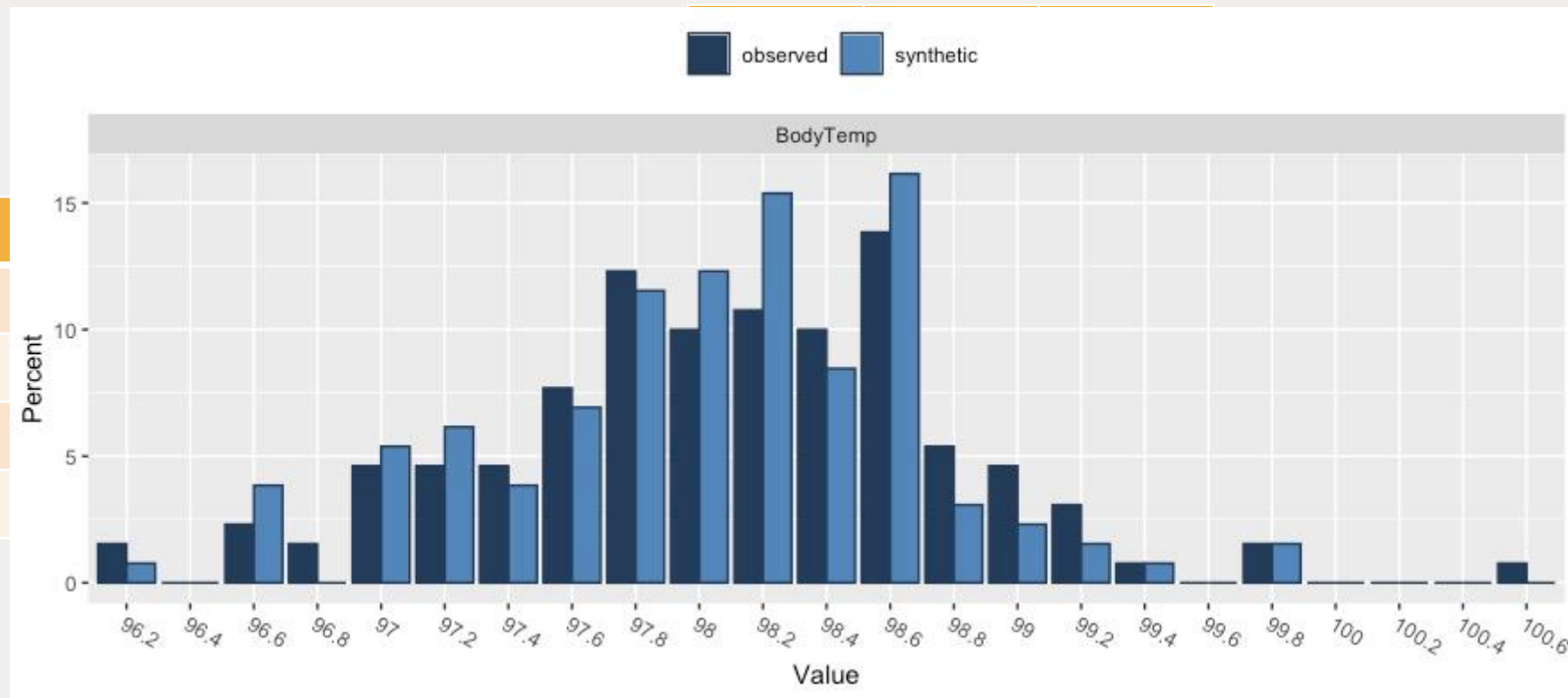


Dados Sintéticos

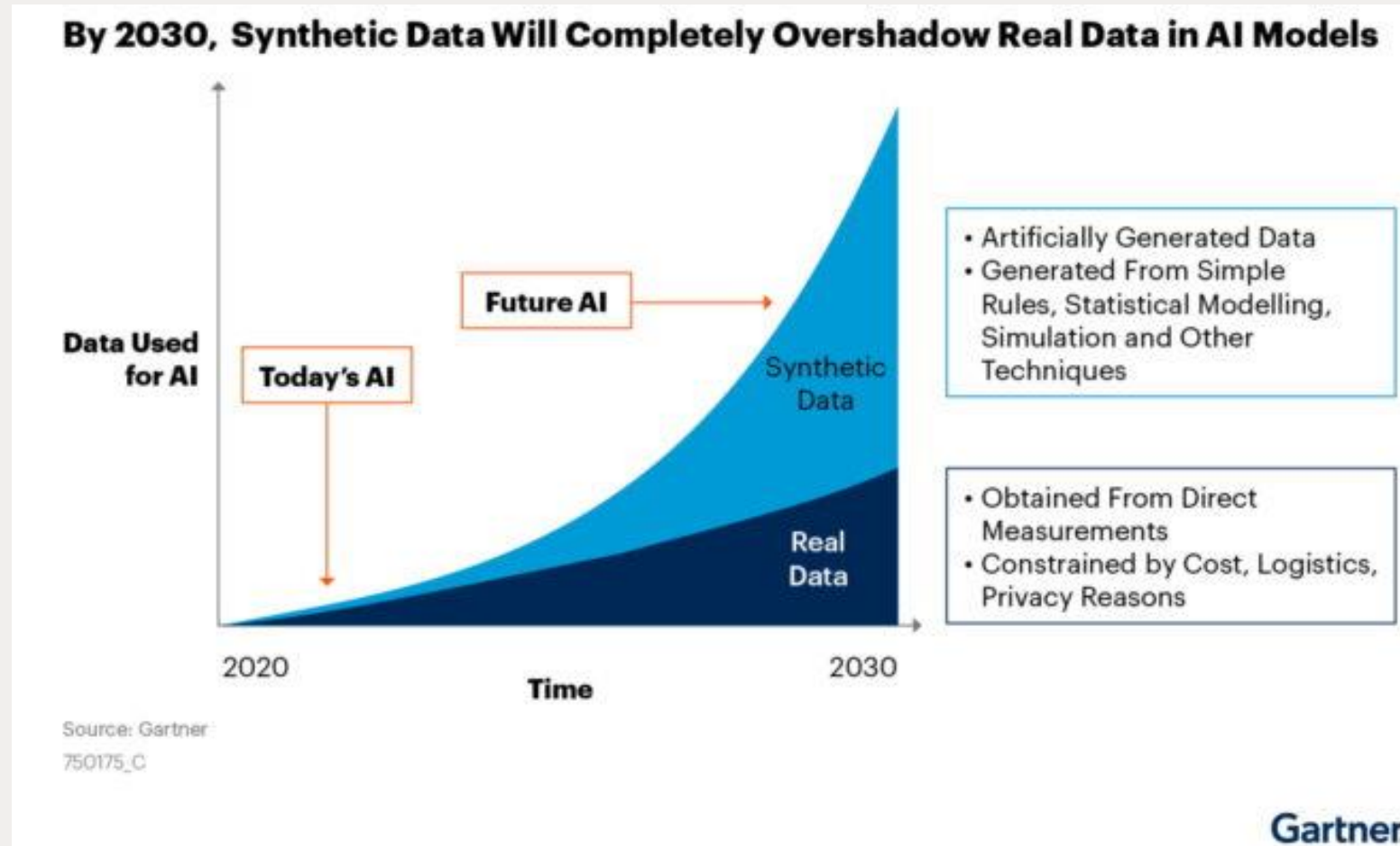
- Synthetic Data é o dado gerado por meio de algoritmos e cujas características estatísticas se aproximam do dado original.

Dataset Original

| | |
|--|--|
| | |
| | |
| | |
| | |
| | |

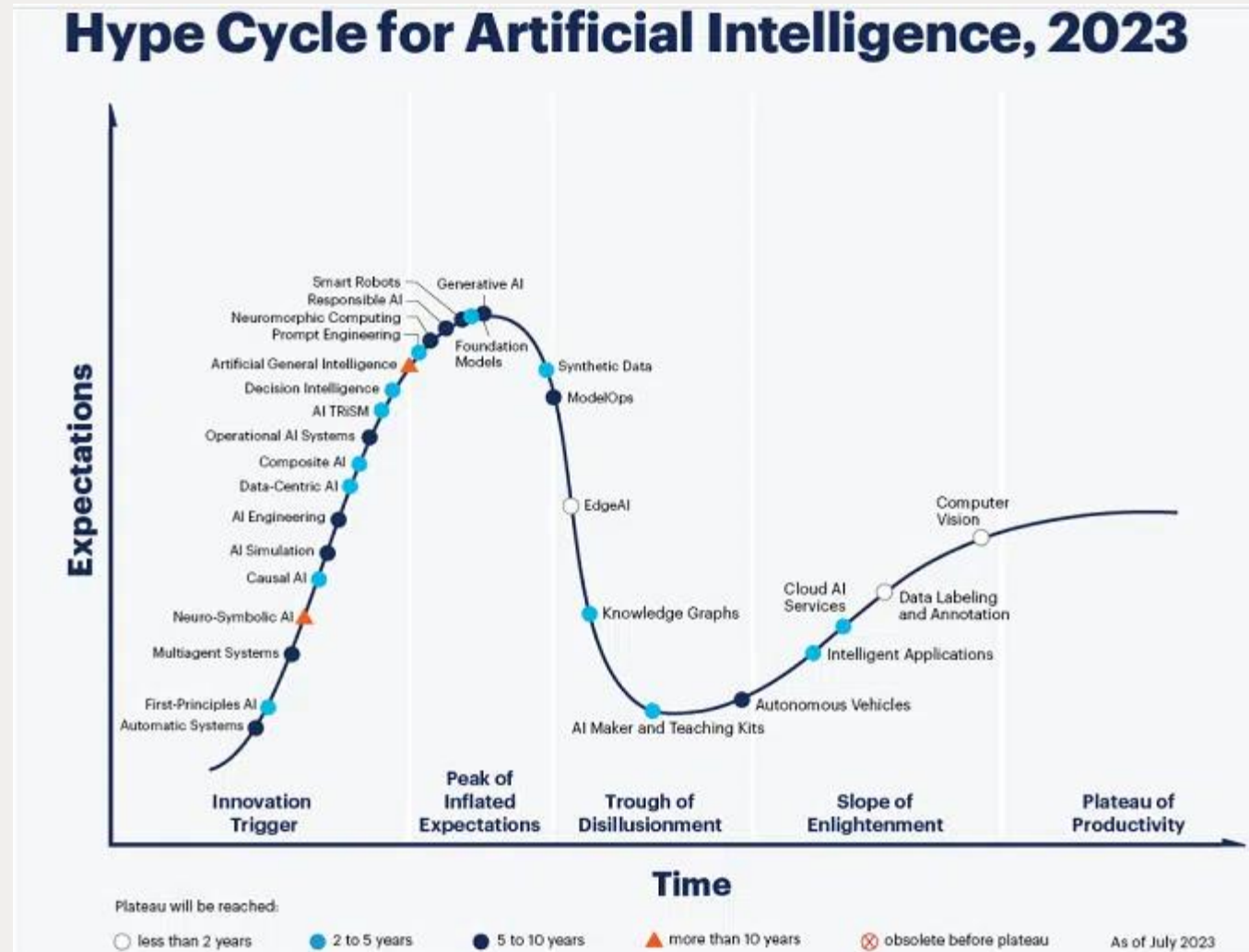


Tendências



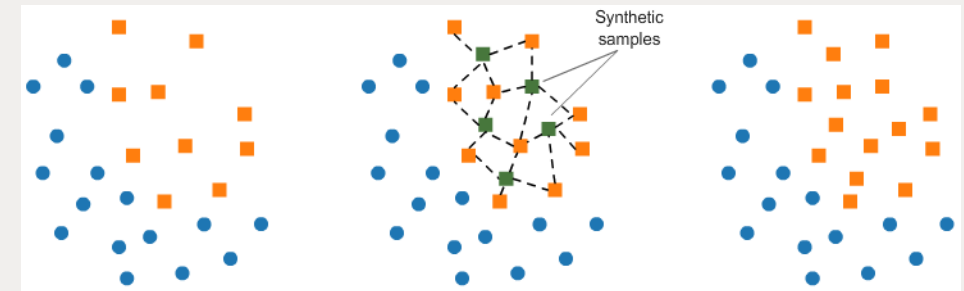
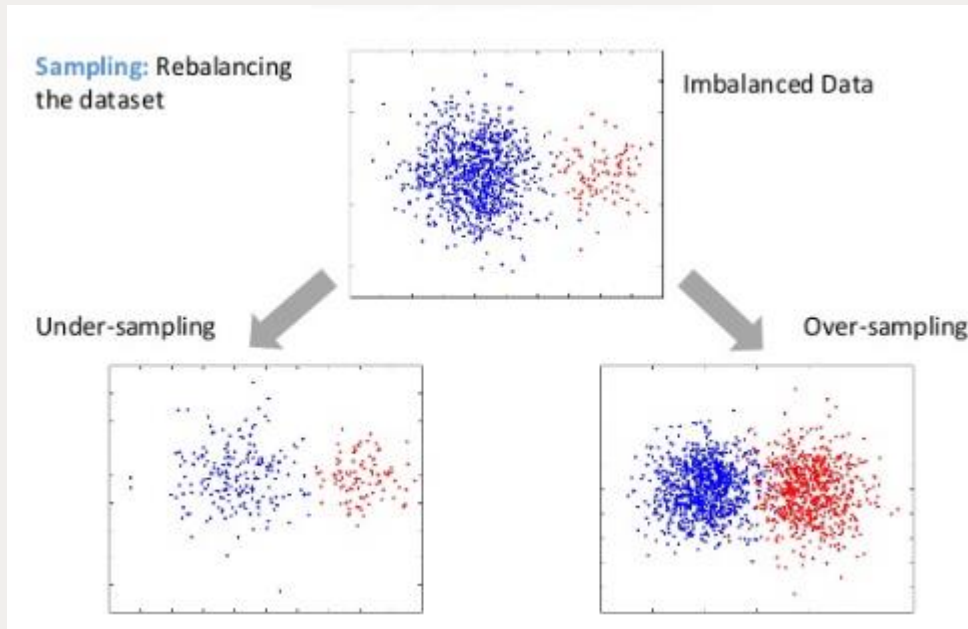
<https://blogs.nvidia.com/blog/2021/06/08/what-is-synthetic-data/>

Tendências



SMOTE

Smote permite balancear um dataset por meio de oversampling



SMOTE

Over-sample using SMOTE.

SMOTENC

Over-sample using SMOTE for continuous and categorical features.

SMOTEN

Over-sample using the SMOTE variant specifically for categorical features only.

SVMSMOTE

Over-sample using SVM-SMOTE variant.

BorderlineSMOTE

Over-sample using Borderline-SMOTE variant.

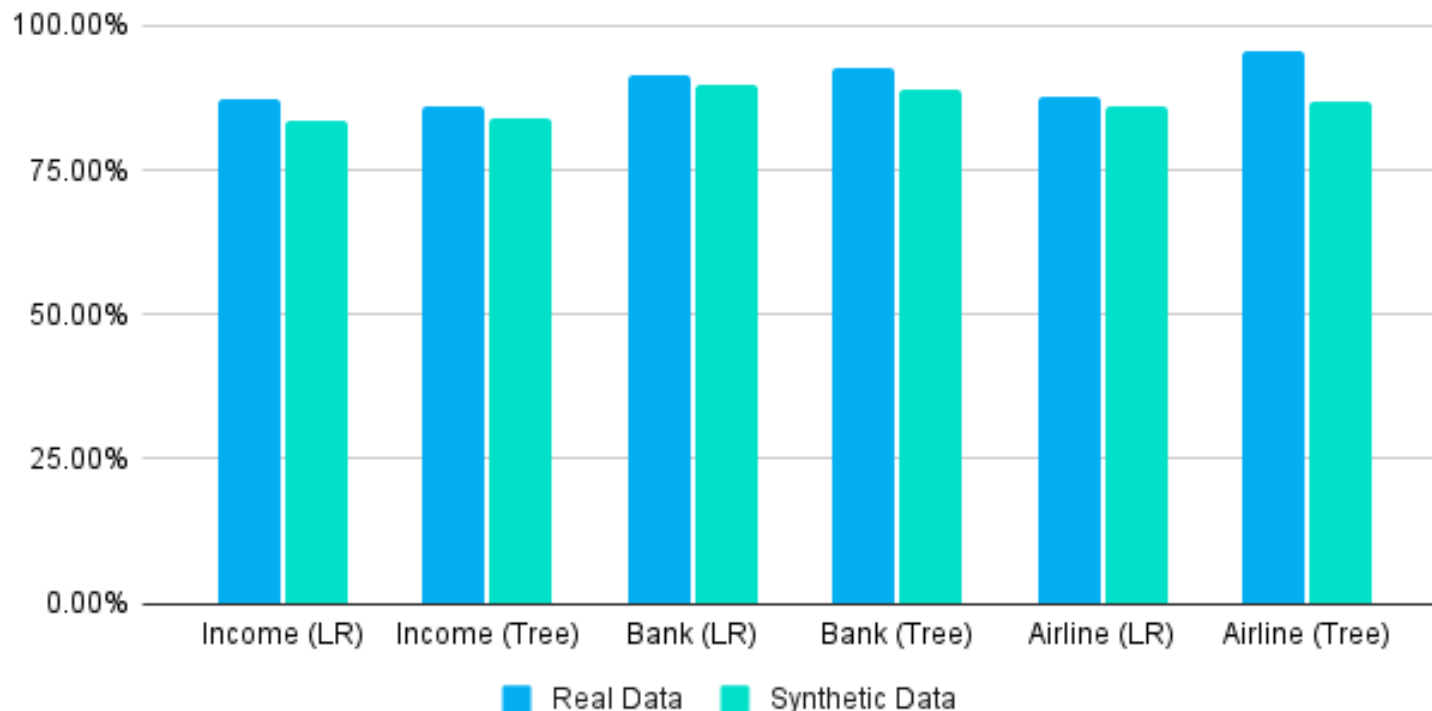
<https://blogs.oracle.com/ai-and-datascience/post/how-to-handle-imbalanced-data-an-overview>

SDV

O [Synthetic Data Vault Project](#) foi criado inicialmente no MIT's [Data to AI Lab](#) em 2016. Depois de 4 anos de pesquisa e envolvimento com a indústria, foi criada a [DataCebo](#) em 2020 com o objetivo de criar a próxima geração de SDV.

Machine Learning Efficacy

Measured by Logistic Regression (LR) and Binary Decision Tree (Tree)



Original Data



Real Data



Synthetic Data



Evaluation

<https://sdv.dev/SDV/>

Generative Adversarial Networks

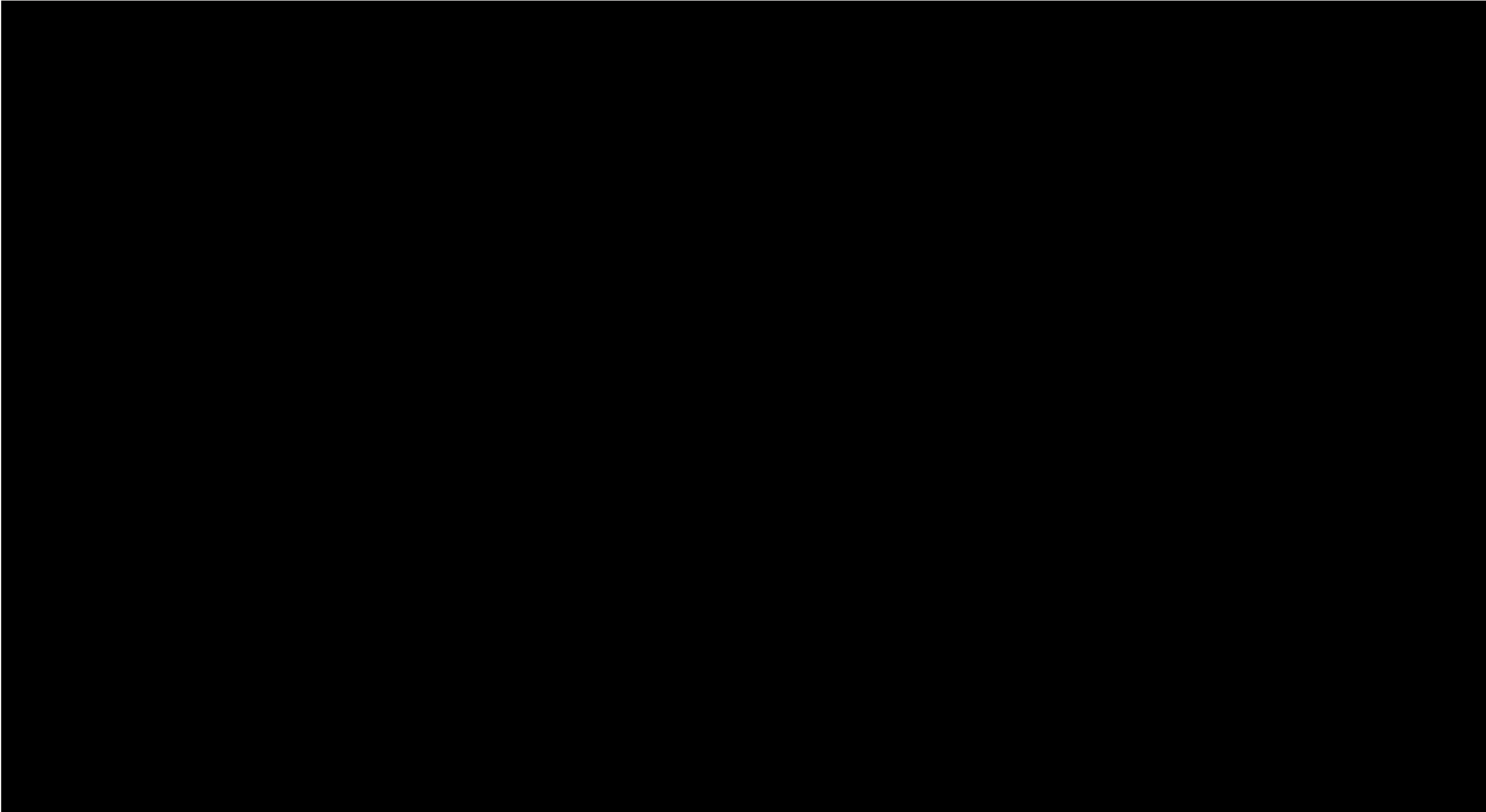
- Uma rede neural gera dados sintéticos a partir de um input aleatório (Generator). Estes dados então são submetidos a uma rede discriminadora (Discriminator). O resultado da comparação é se o dado sintético é real ou falso (Real/Fake).

Real

Noise
(z)

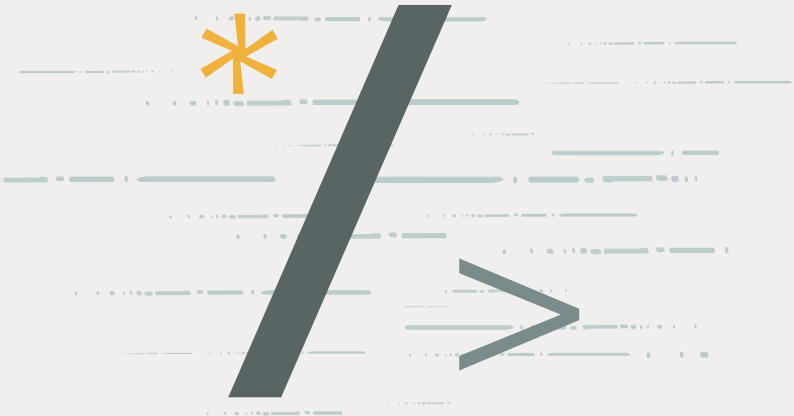


agation



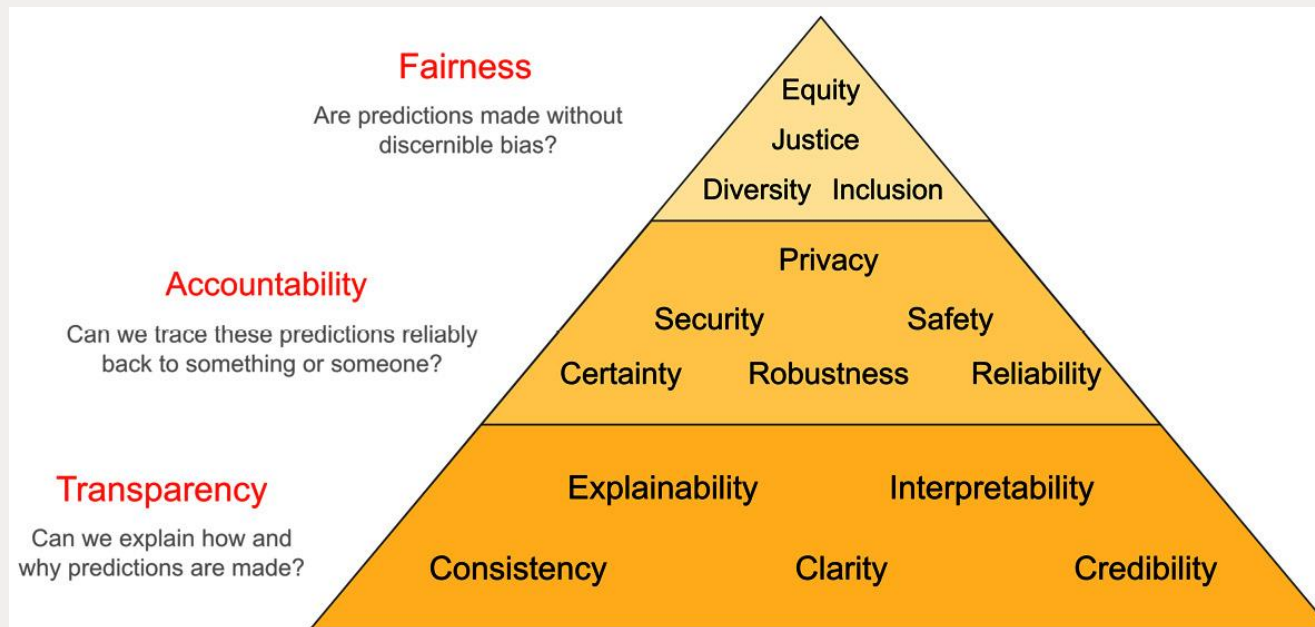
Interpretabilidade de Modelos

Technology for business transformation



Interpretable Machine Learning

Conjunto de métodos utilizados para garantir que nosso modelo de ML seja: **seguro, justo e confiável**.



LGPD – Capítulo 3 Artigo 20

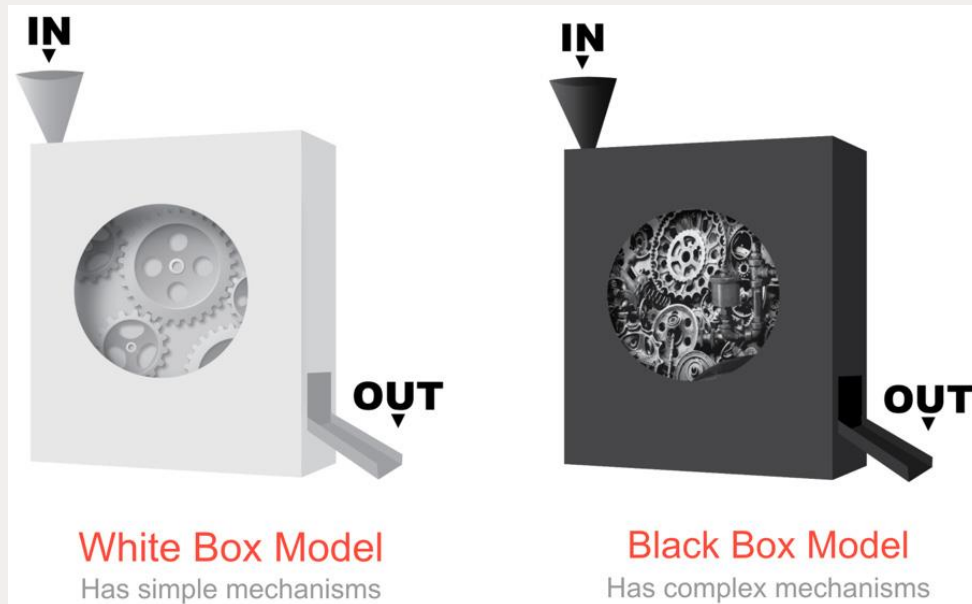
O titular dos dados tem direito a solicitar a revisão de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais que afetem seus interesses, incluídas as decisões destinadas a definir o seu perfil pessoal, profissional, de consumo e de crédito ou os aspectos de sua personalidade.

Exemplos



Interpretability

Entrada e saída, causa e efeito... Consigo explicar seus mecanismos e a forma como ele produz uma saída?



Quais são os requerimentos do modelo?

E suas restrições?

Limites de confiança?

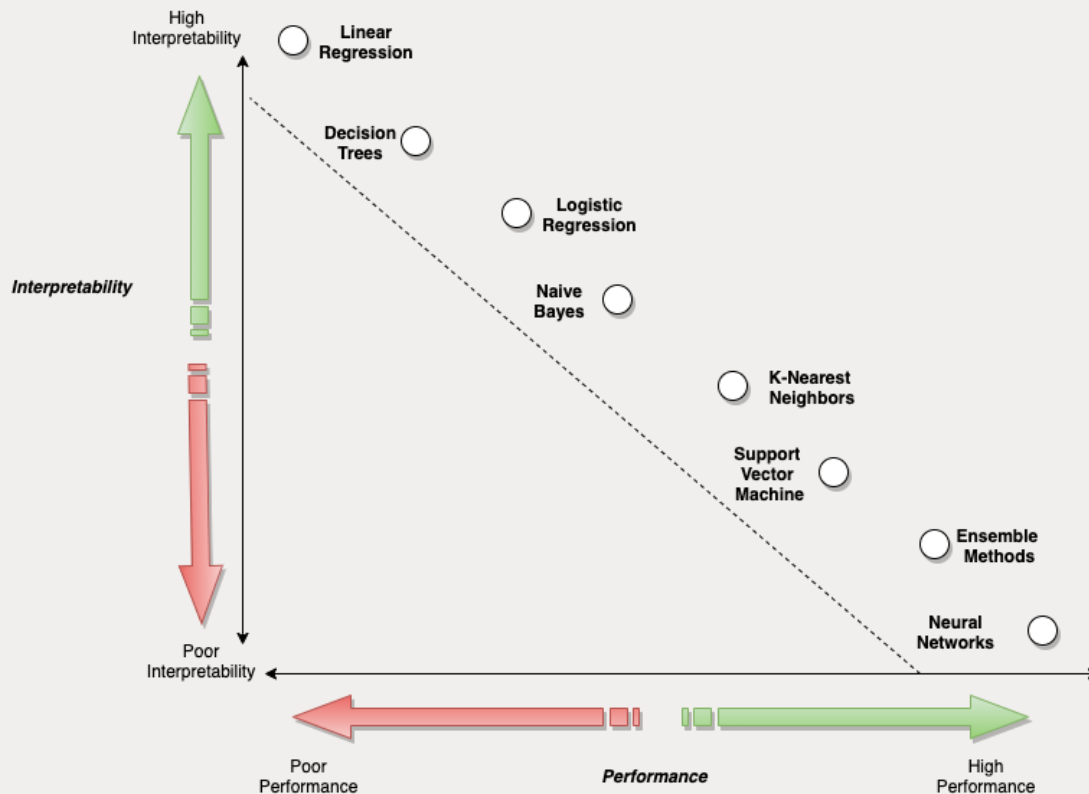
Conhecimento Científico

Confiança, segurança

Ética

Explainability

Contempla o conceito de interpretabilidade, mas exige uma série de técnicas para acessar os mecanismos internos do modelo.



Como explico em termos humanos:

- Hiperparâmetros
- Minimização de erros no modelo
- Reprodutibilidade do resultado

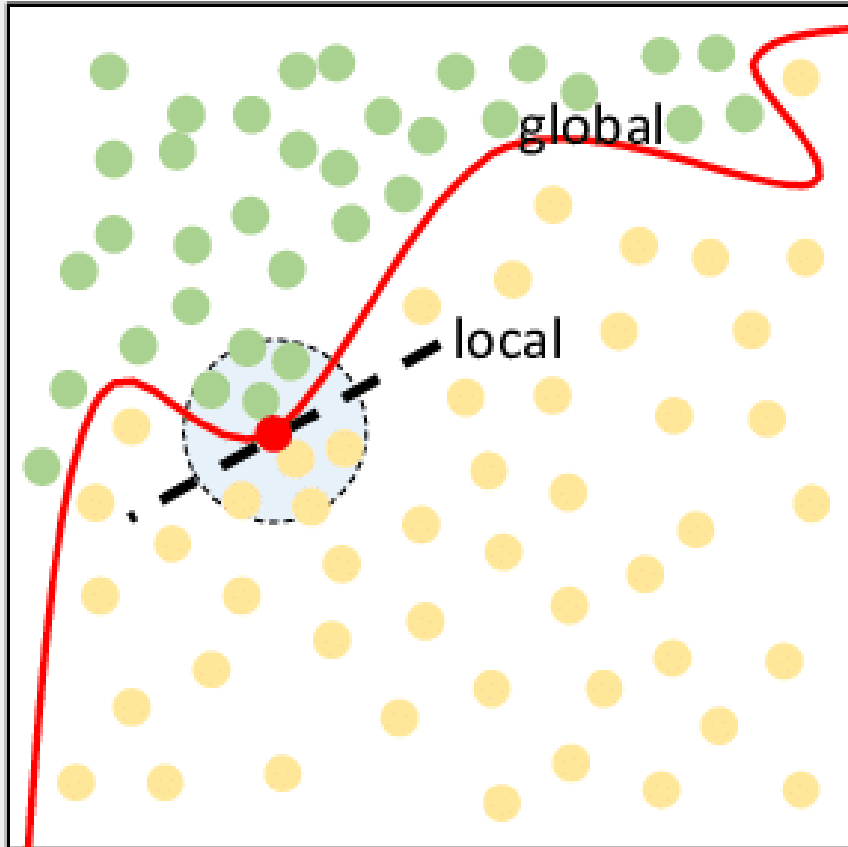
Comparando alguns modelos...

| White Box? | Model Class | Properties that Increase Interpretability | | | | | Task | | Performance Rank | |
|------------|------------------------|---|--------|----------|-----------------|----------|-------|----------|------------------|----------|
| | | 🔑 Expl. | Linear | Monotone | Non-Interactive | 🔧 Regul. | Regr. | Classif. | Regr. | Classif. |
| ✓ | Linear Regression | ● | ● | ● | ● | ● | ✓ | ✗ | 6 | |
| ✓ | Regularized Regression | ● | ● | ● | ● | ● | ✓ | ✓ | 7 | 8 |
| ✓ | Logistic Regression | ● | ● | ● | ● | ● | ✗ | ✓ | | 5 |
| ✓ | Gaussian Naïve Bayes | ● | ● | ● | ● | ● | ✗ | ✓ | | 7 |
| ✓ | Polynomial Regression | ● | ● | ● | ● | ● | ✓ | ✓ | 2 | |
| ✓ | RuleFit | ● | ● | ● | ● | ● | ✓ | ✓ | 8 | |
| ✓ | Decision Tree | ● | ● | ● | ● | ● | ✓ | ✓ | 5 | 3 |
| ✓ | k-Nearest Neighbors | ● | ● | ● | ● | ● | ✓ | ✓ | 9 | 6 |
| ✗ | Random Forest | ● | ● | ● | ● | ● | ✓ | ✓ | 3 | 4 |
| ✗ | Gradient Boosted Trees | ● | ● | ● | ● | ● | ✓ | ✓ | | 2 |
| ✗ | Multi-layer Perceptron | ● | ● | ● | ● | ● | ✓ | ✓ | 1 | 1 |

| | White Box | Glass Box | Black Box |
|-----------------------------|-----------|-----------|-----------|
| Interpretability | High | Mid-High | Low |
| Predictive Performance | Mid | High | High |
| Execution Speed Performance | High | Low | Mid |

Mais dimensões...Mais overfitting
Alta dimensionalidade
Navalha de Occam

Global and Single Interpretation



Global Interpretation

Conseguimos explicar como o modelo faz suas previsões

- a) Global holistic

Podemos compreender o modelo inteiro. Ex: regressão linear

- b) Global modular

Podemos compreender parte do modelo. Por exemplo com Feature Importance.

Local Interpretation


Podemos explicar como uma única previsão foi feita

* https://www.researchgate.net/figure/Left-local-interpretation-right-knowledge-distillation-15_fig2_331794108

Sobre o AutoMLx

Pacote Python que automaticamente cria, otimiza e explica modelos de machine learning.

The screenshot displays the Oracle Cloud web interface. At the top, the Oracle Cloud logo and a session URL are visible. The main navigation bar includes options like File, Edit, View, Run, Kernel, Git, Tabs, Settings, and Help. The left sidebar shows a file explorer with a list of files and folders, including 'logs', 'auto_mlx_exem...', 'oracle_automl_', 'titanic-Copy1.csv', 'titanic.csv', and 'titanic2.csv'. The main content area is titled 'Environment Explorer' and features a search bar and filters for Conda Environments (Data Science, Published, Installed) and Architecture (ALL, CPU, GPU). A table lists the installed environments, with 'Oracle AutoML and Model Explanation for Python 3.8 Python 3.8' highlighted. Below the table, there are sections for 'Publish', 'Clone', and 'Uninstall', each with a terminal command and a copy icon. The 'Publish' command is 'odsc conda publish -s automlx_p38_cpu_v1'. The 'Clone' command is 'odsc conda clone -f automlx_p38_cpu_v1 -e <local_env_n...'. The 'Uninstall' command is 'odsc conda delete -s automlx_p38_cpu_v1'. The bottom status bar indicates 'Saving started' and 'Environment Explorer'.

| Name > | Environment Version | Type | Architecture > | Created > | Size > |
|---|---------------------|-----------|----------------|-----------|--------|
|  Oracle AutoML and Model Explanation for Python 3.8 Python 3.8 | 1.0 | Installed | CPU | - | - |

Publish
Copy and run the following command in terminal window:
`odsc conda publish -s automlx_p38_cpu_v1`

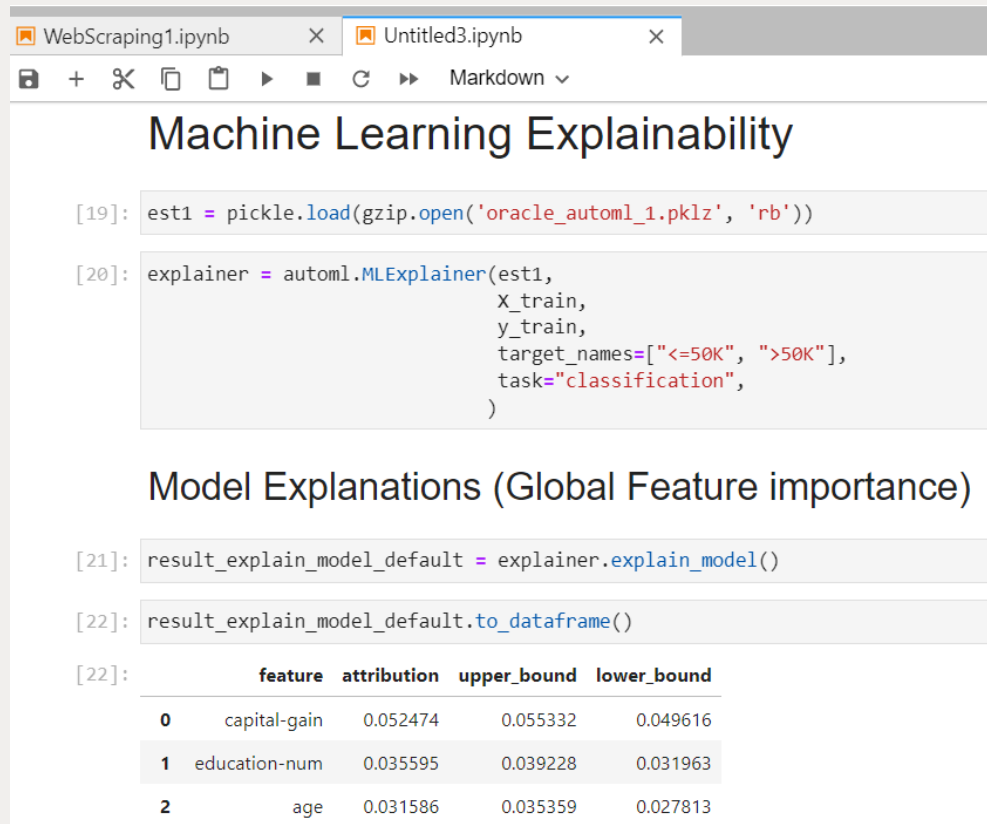
Clone
Copy and run the following command in terminal window:
`odsc conda clone -f automlx_p38_cpu_v1 -e <local_env_n...`

Uninstall
Copy and run the following command in terminal window:
`odsc conda delete -s automlx_p38_cpu_v1`

Description
Oracle Labs brings their AutoML and Model Explanation packages together in the new automlx library. To get started with the Oracle AutoML environment, review the notebook example getting-started.ipynb from the Notebook Examples launcher button. For more details, and technical overview check out [Oracle AutoML: A Fast and Predictive AutoML Pipeline](#)

Configurando o Explainer

Para uma tomada de decisão, obter a previsão não é suficiente. Começamos com uma pergunta: Quais features são mais relevantes para esta previsão.



```
[19]: est1 = pickle.load(gzip.open('oracle_automl_1.pklz', 'rb'))

[20]: explainer = automl.MLEExplainer(est1,
                                     X_train,
                                     y_train,
                                     target_names=["<=50K", ">50K"],
                                     task="classification",
                                     )

Machine Learning Explainability

[21]: result_explain_model_default = explainer.explain_model()

[22]: result_explain_model_default.to_dataframe()

[22]:
```

| | feature | attribution | upper_bound | lower_bound |
|---|---------------|-------------|-------------|-------------|
| 0 | capital-gain | 0.052474 | 0.055332 | 0.049616 |
| 1 | education-num | 0.035595 | 0.039228 | 0.031963 |
| 2 | age | 0.031586 | 0.035359 | 0.027813 |

MLEExplainer

explain_model()

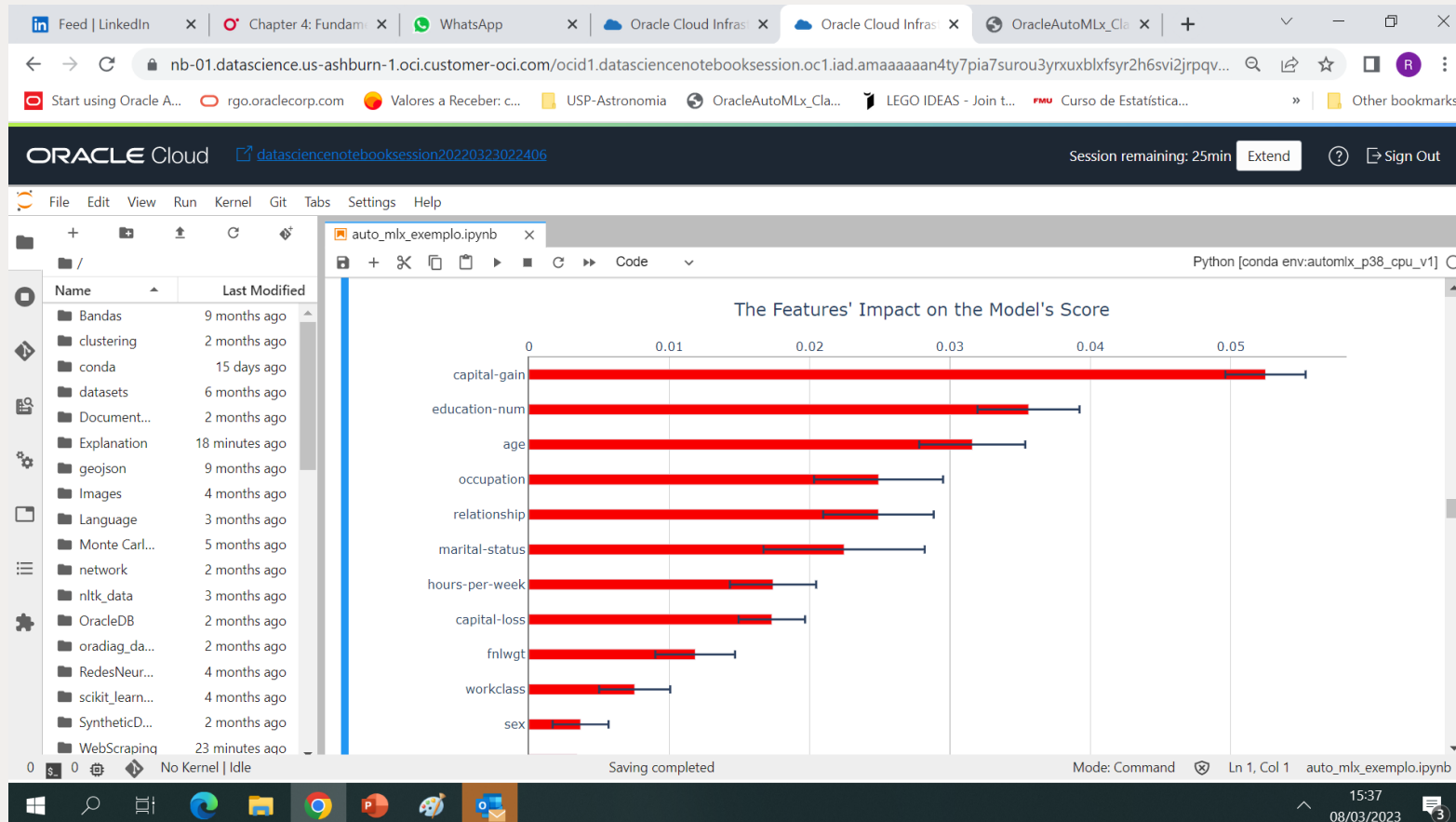
to_dataframe()

show_in_notebook()

* Considera cada feature independente uma da outra

PFI – Permutation Feature Importance

Este método mede o aumento no erro da predição desde que os valores de uma feature passem por shuffling. Se a feature tem relação com a variável target, o shuffling deve aumentar o erro
Model Agnostic / Assume independência entre variáveis



```
est1 = pickle.load(gzip.open('oracle_automl_1.pklz', 'rb'))

explainer = automl.MLExplainer(est1,
                                X_train,
                                y_train,
                                target_names=["<=50K", ">50K"],
                                task="classification",
                                )
```

```
result_explain_model_default = explainer.explain_model()
```

```
result_explain_model_default.to_dataframe()
```

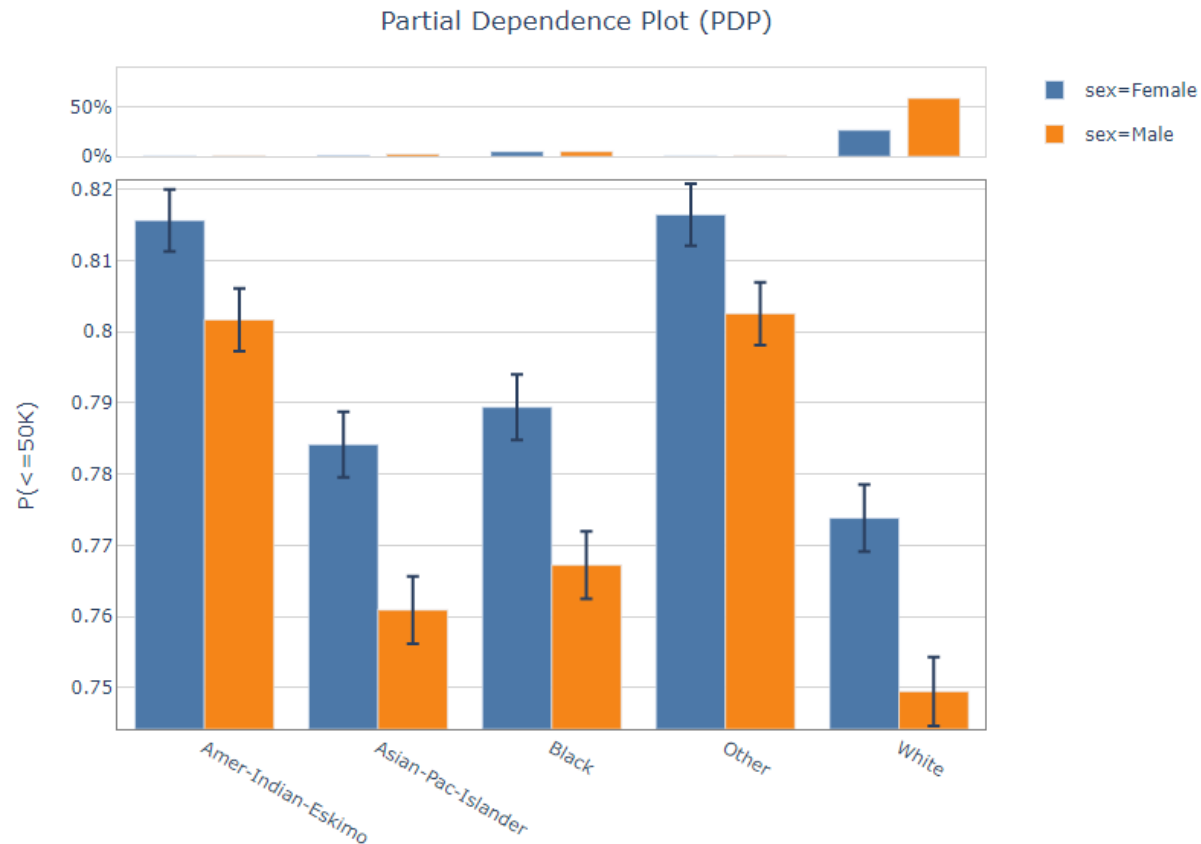
| | feature | attribution | upper_bound | lower_bound |
|---|----------------|-------------|-------------|-------------|
| 0 | capital-gain | 0.052474 | 0.055332 | 0.049616 |
| 1 | education-num | 0.035595 | 0.039228 | 0.031963 |
| 2 | age | 0.031586 | 0.035359 | 0.027813 |
| 3 | occupation | 0.024910 | 0.029510 | 0.020309 |
| 4 | relationship | 0.024902 | 0.028844 | 0.020960 |
| 5 | marital-status | 0.022454 | 0.028196 | 0.016713 |

Partial Dependence Plots (PDPs)

Trata-se de um método de interpretação global que pode demonstrar visualmente a natureza do impacto de uma feature no target

The histogram displays the joint distribution of the two features.

```
[26]: result_explain_feature_dependence_default = explainer.explain_feature_dependence(['race', 'sex'])  
      result_explain_feature_dependence_default.show_in_notebook()
```

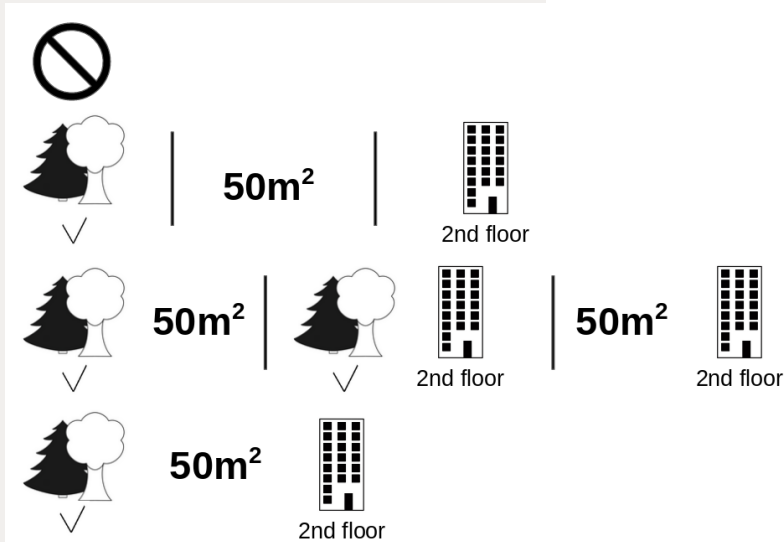
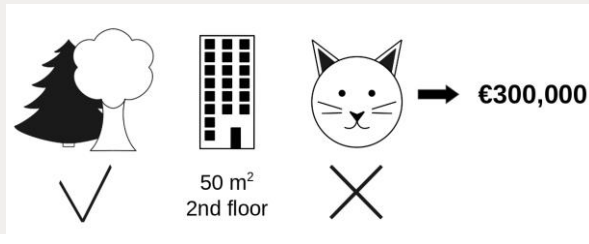


- Assume independência entre variáveis
- Histograma no topo demonstra a distribuição

Shapley Values

Uma predição pode ser explicada assumindo que cada feature é um jogador em um game aonde o pagamento é a predição. Trata-se de uma técnica de coalitional game theory cujo objetivo é mostrar como distribuir este pagamento justamente entre as features.

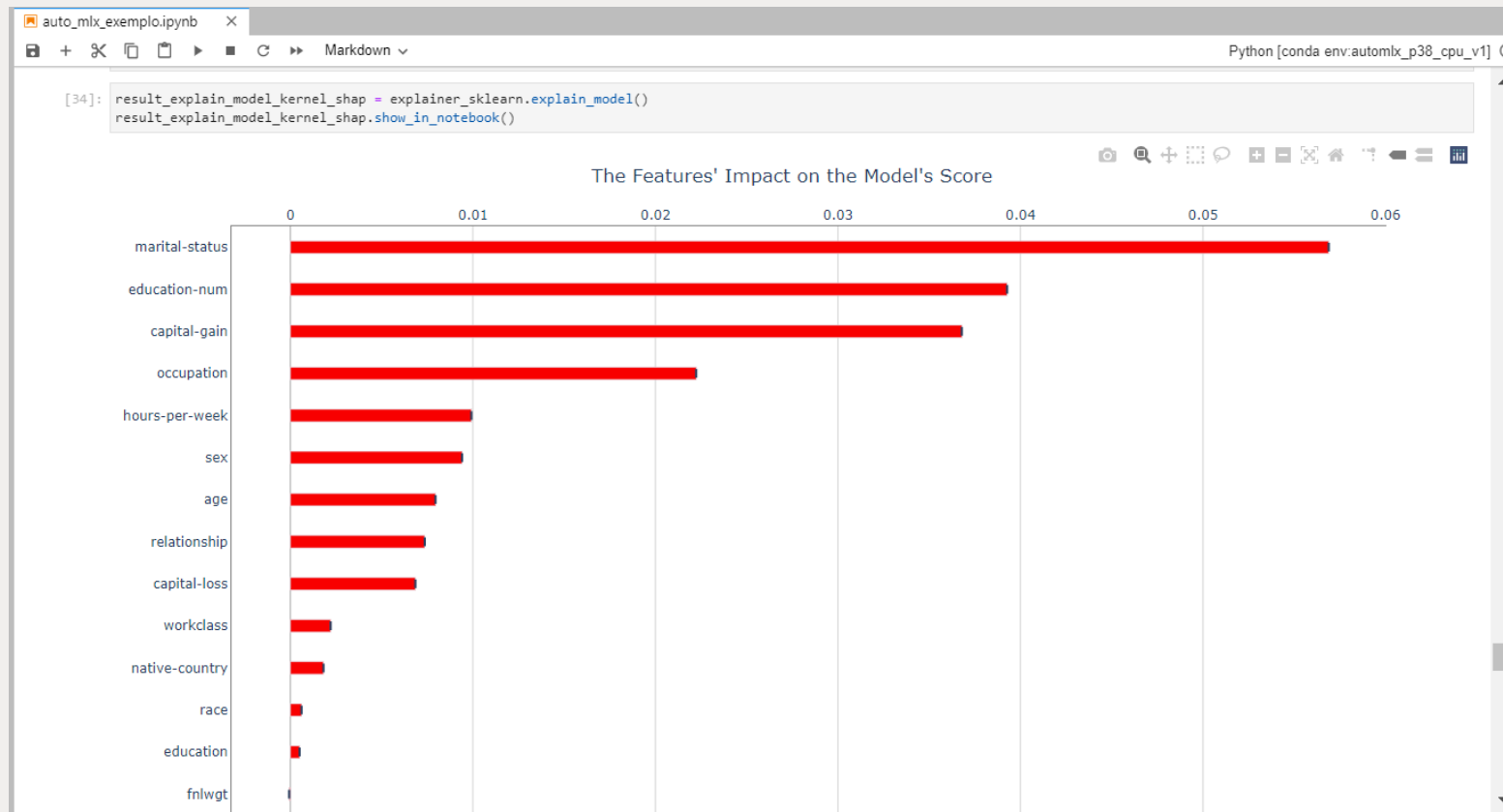
Predição



- Para cada uma das coalizões ao lado, computamos o preço previsto do apartamento com e sem o valor da feature e pegamos a diferença para ter a contribuição marginal.
- O Shapley value é a média ponderada das contribuições marginais.

SHapley Additive exPlanations (SHAP) (Model)

Implementação Python do método Shapley

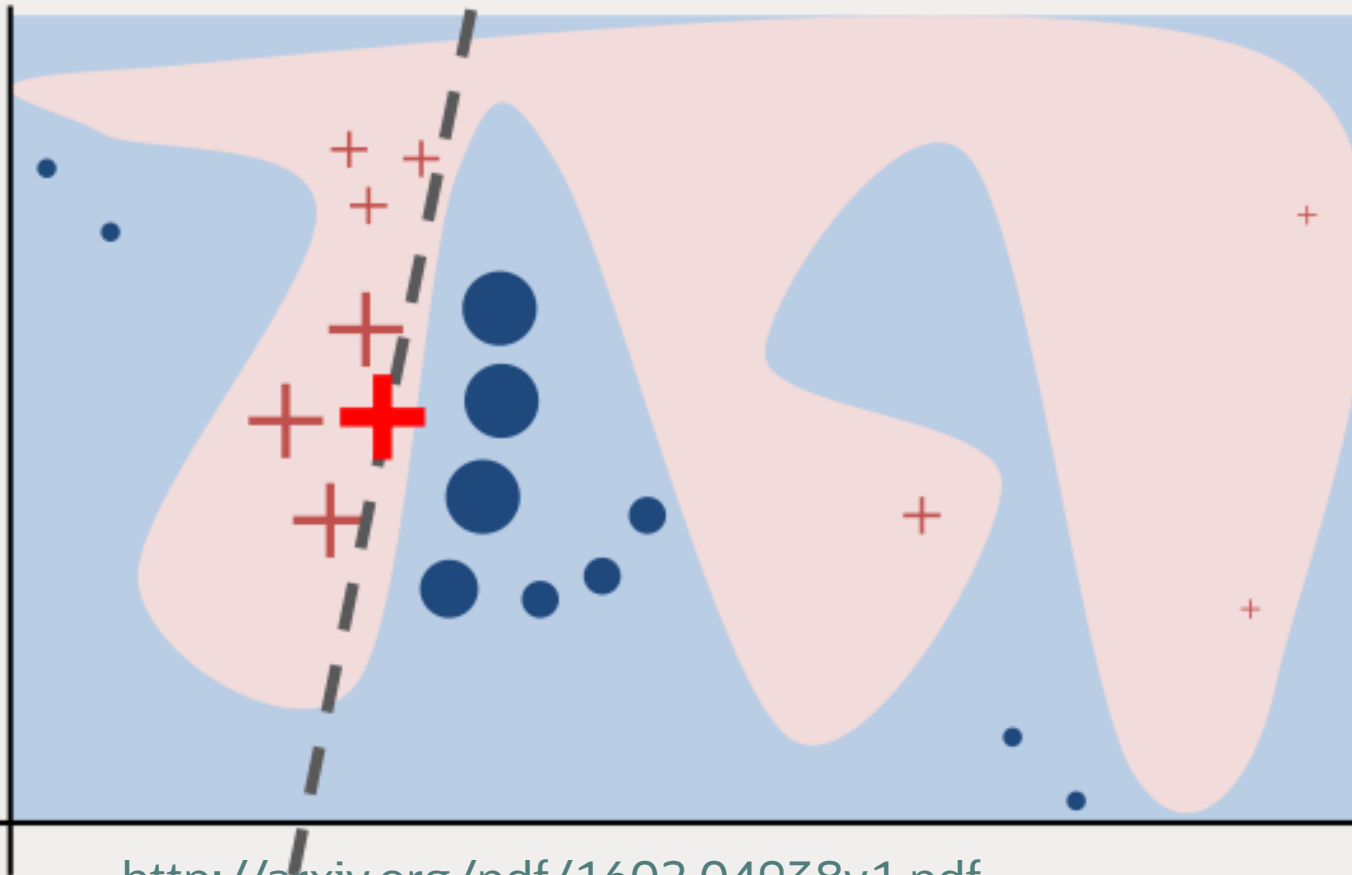


- Execução em tempo exponencial no que diz respeito ao número de features



Local Interpretable Model-agnostic Explanations (LIME)

Treina surrogate models para explicar uma única predição. Para isso faz uso do modelo original (black box) e um conjunto de dados alterado, próximo ao ponto que queremos explicar (chamado neighborhood)



- A grande cruz vermelha é a instância a ser explicada (X)
- Simulamos instâncias próximas de X e atribuímos peso conforme a distância
- Obtemos previsões a partir do modelo original
- Aprendemos um modelo linear para estes valores
- A explicação não é confiável globalmente, mas é confiável nas proximidades de X

Local Interpretable Model-agnostic Explanations (LIME)

A implementação em AutoMLx faz uso do método `explain_prediction`, com o explainer configurado para o tipo “surrogate”.

Explaining predictions with surrogate explainer_type (lime)

```
explainer_sklearn.configure_explain_prediction(explainer_type='surrogate')
```

```
[2022-10-21 10:20:15,263] [automl.mlx] AutoMLx got an unexpected keyword argument 'evaluator_type', which is not a configurable attribute of any of ['TabularLocalSurrogateExplainer', 'SystematicSampleGeneration', 'Siloweighting', 'SurrogateHandler'].
```

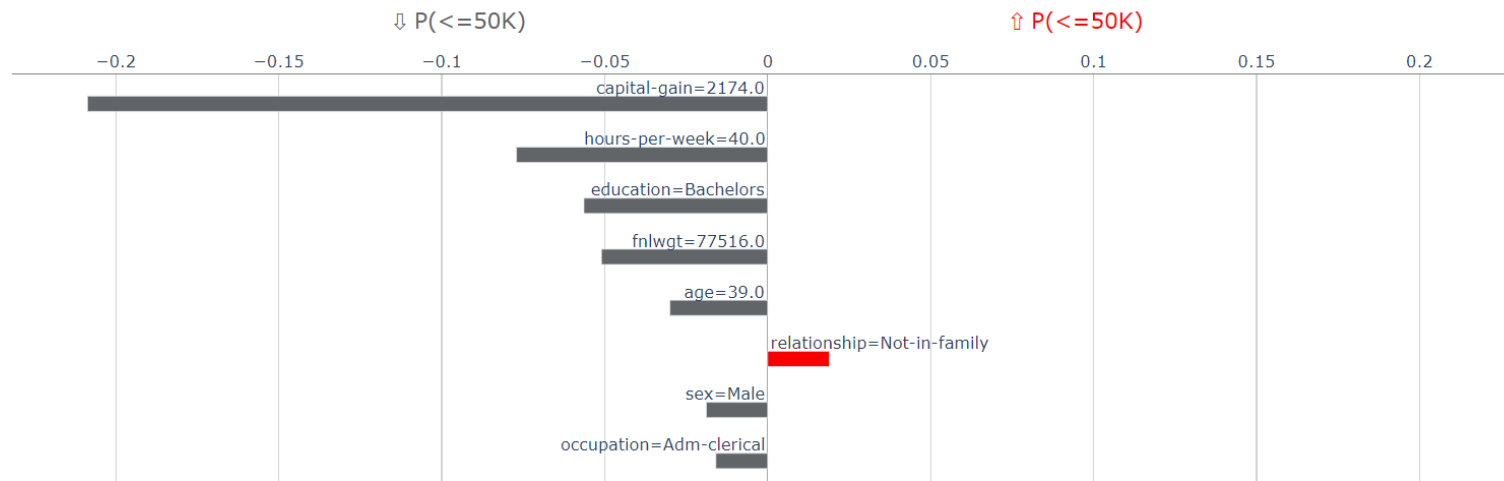
Valid options are:

```
{'TabularLocalSurrogateExplainer': ['method', 'exp_sorting', 'num_features', 'scale_weight'], 'SystematicSampleGeneration': ['discretizer', 'num_samples', 'n_bins'], 'Siloweighting': [], 'SurrogateHandler': ['model', 'feature_selection', 'force_fit_sample']}
```

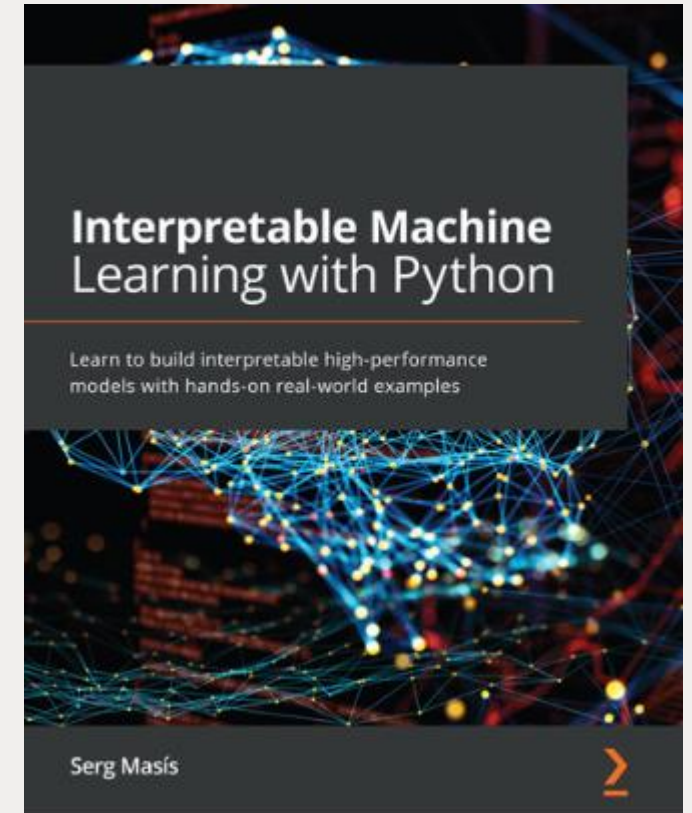
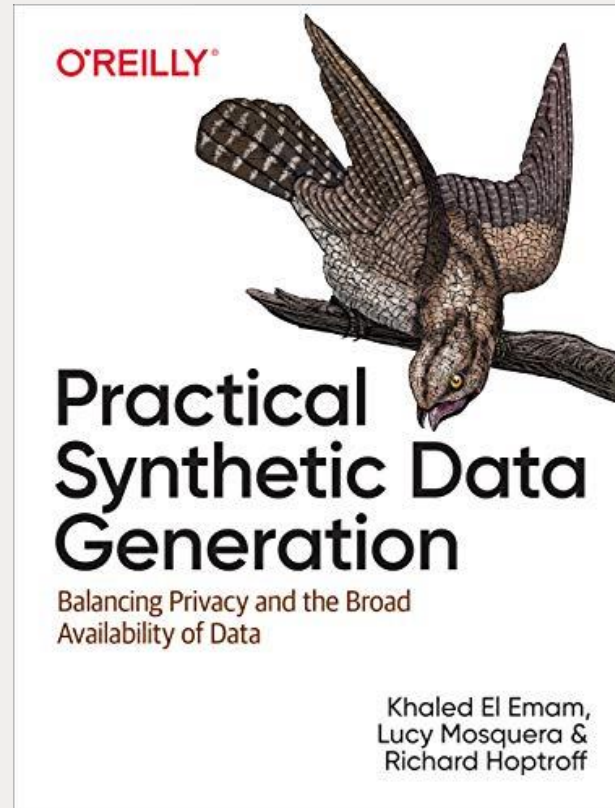
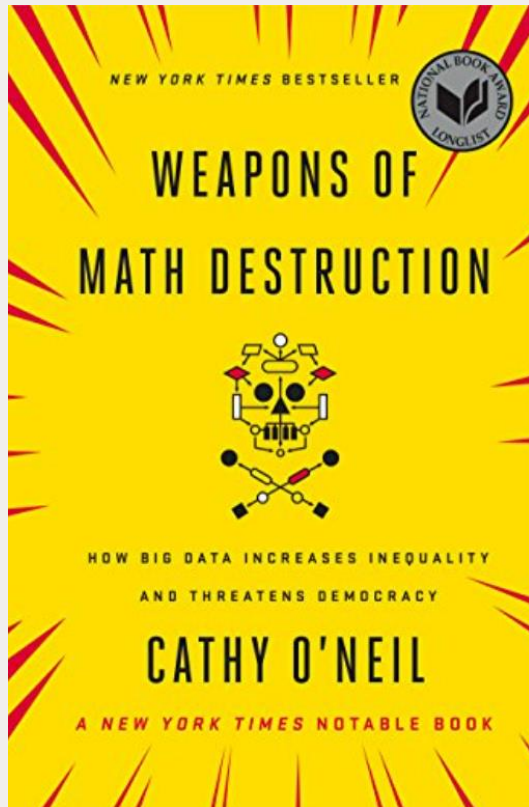
```
index = 0
```

```
result_explain_prediction_surrogate_lime = explainer_sklearn.explain_prediction(X_train.iloc[index:index+1,:])
```

```
result_explain_prediction_surrogate_lime[0].show_in_notebook()
```



Sugestões Leitura





Experimente o poder da IA gratuitamente na nuvem

Conheça os recursos de IA e ML da Oracle Cloud com o **Modo Gratuito**.

São **US\$ 500** em créditos para você usar durante 30 dias – uma condição **EXCLUSIVA** para visitantes do TDC Summit São Paulo!



ORACLE
DevLive
São Paulo

Junte-se a nós em São Paulo para criar novas possibilidades com dados e IA

03 de abril de 2024

World Trade Center São Paulo

Inscreva-se já!

oracle.com/devlive-sao-paulo/



ORACLE

Thank you!

