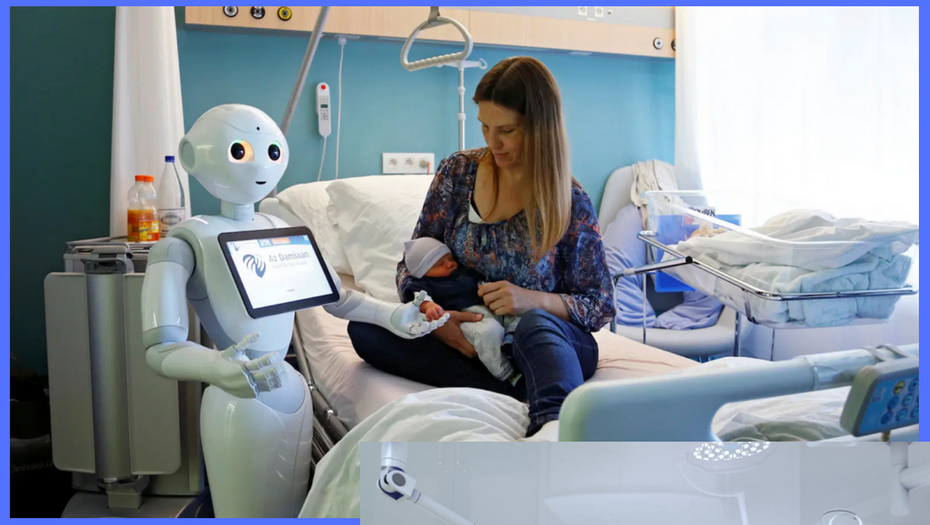
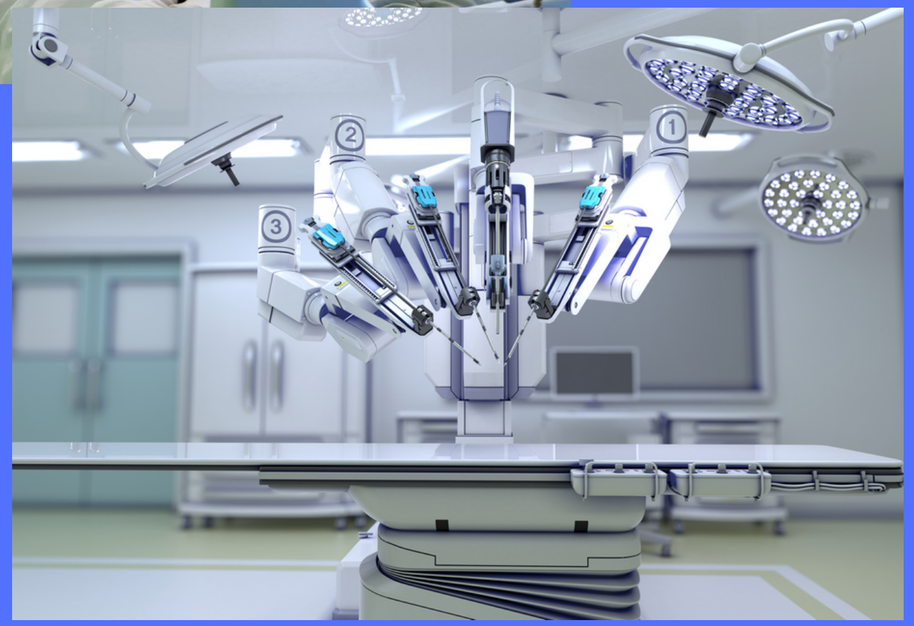
EXPLICABILIDADE

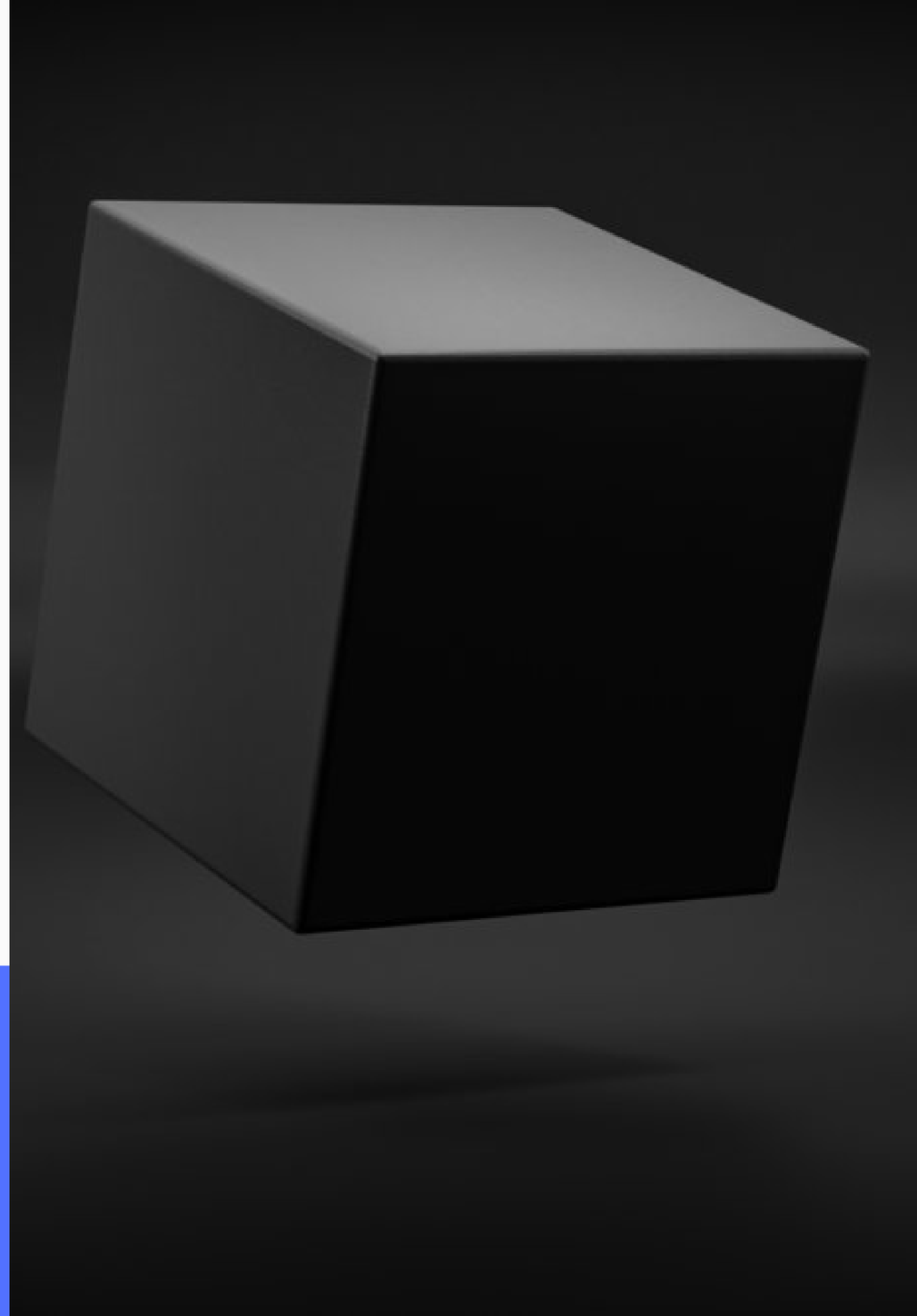
A IMPORTÂNCIA DA EXPLICABILIDADE EM ML


Automatic Language Translation
Medical Diagnosis
Stock Market trading
Online Fraud Detection
Virtual Personal Assistant
Email Spam and Malware Filtering
Self driving cars
Product recommend -ations
Traffic Prediction
Speech Recognition
Image Recognition





A CAIXA PRETA DO MACHINE LEARNING



Explain the Prediction

Predicted: Wolf
True: Wolf

Predicted: Husky
True: Husky

Predicted: Husky
True: Husky

Predicted: Wolf
True: Wolf

Predicted: Wolf
True: Wolf

Predicted: Wolf
True: Wolf

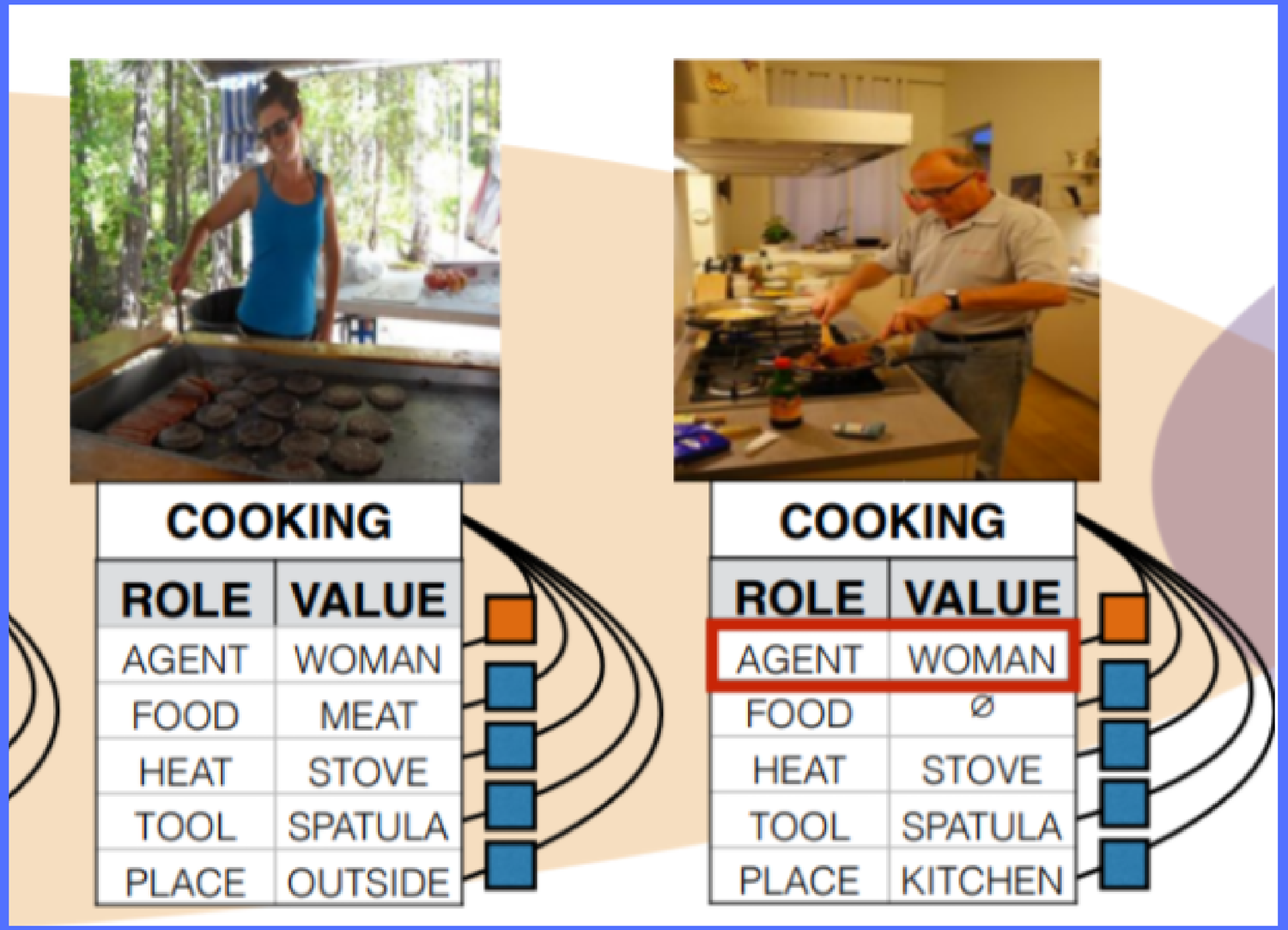
Predicted: Husky
True: Wolf

Predicted: Wolf
True: Wolf

Predicted: Wolf
True: Husky

Predicted: Husky
True: Husky

MEN ALSO LIKE SHOPPING: REDUCING GENDER BIAS AMPLIFICATION USING CORPUS-LEVEL CONSTRAINTS



O QUE NOSSO ALGORITMO ESTÁ APRENDENDO DE FATO?

SERÁ QUE ELE ESTÁ FUNCIONANDO DA MANEIRA QUE EU DESEJO?

POSSO CONFIAR NELE?

O QUE ESTÁ FAZENDO MEU MODELO NÃO FUNCIONAR?

O que é exatamente a explicabilidade?

INTERPRETABILIDADE E EXPLICABILIDADE

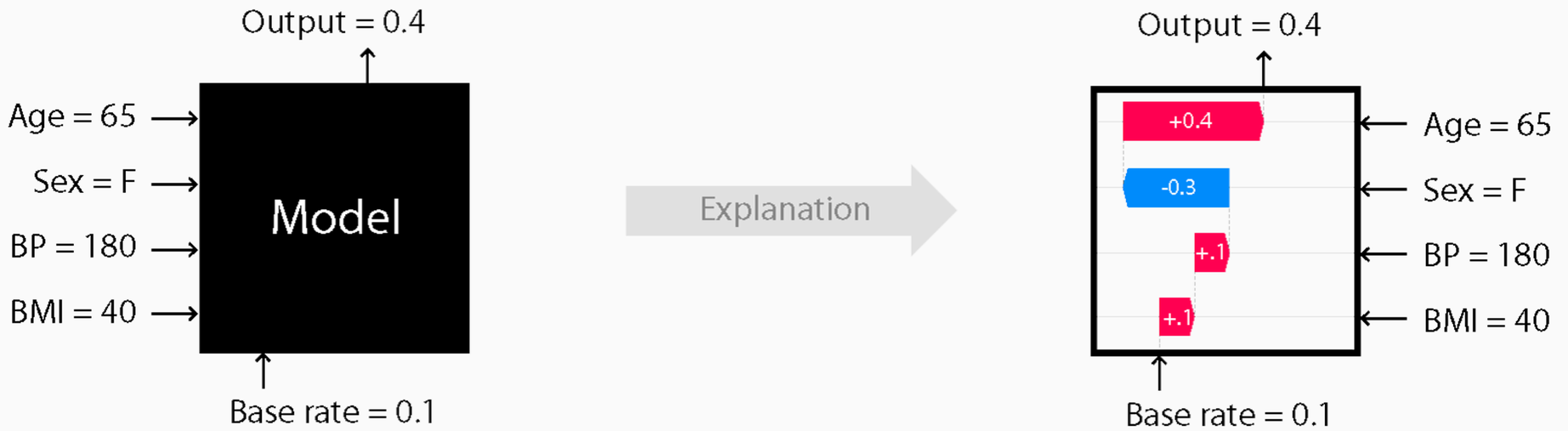
- É a capacidade de se explicar o sistema em uma linguagem compreensível pelos humanos.

- A explicabilidade faz parte do campo de pesquisa de Inteligência Artificial Explicável, que tem como objetivo abordar questões referentes às metodologias de interpretação das predições dos modelos de aprendizado de máquina. Nesta área de pesquisa temos duas áreas principais de estudo:

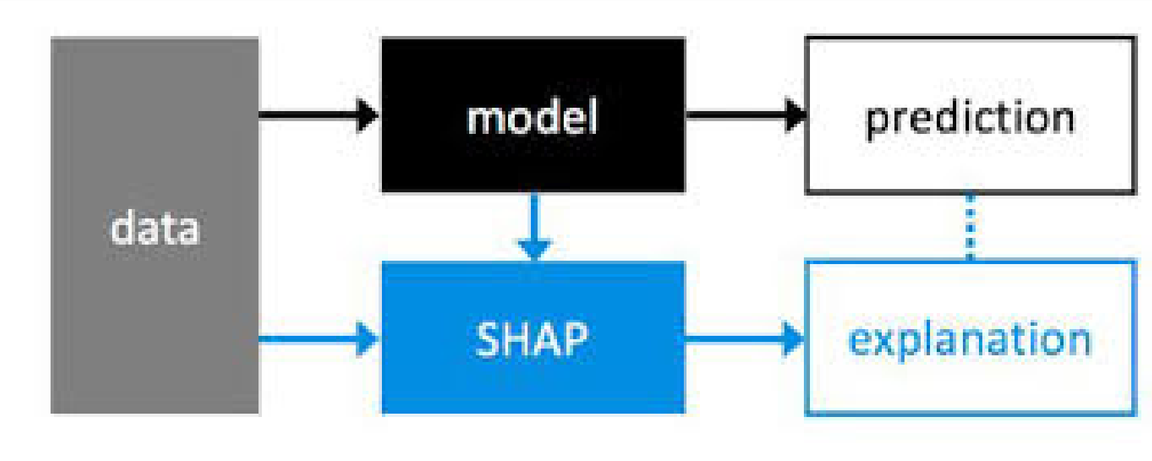
1. A Interpretabilidade: Interpretar os resultados obtidos
2. A Explicabilidade: Explicar os resultados obtidos

ALGUNS MODELOS DE EXPLICABILIDAE

SHAP


Output = 0.4
Age = 65
Sex = F
BP = 180
BMI = 40
Model
Base rate = 0.1
Explanation
Output = 0.4
+0.4
-0.3
+.1
+.1
Age = 65
Sex = F
BP = 180
BMI = 40
Base rate = 0.1

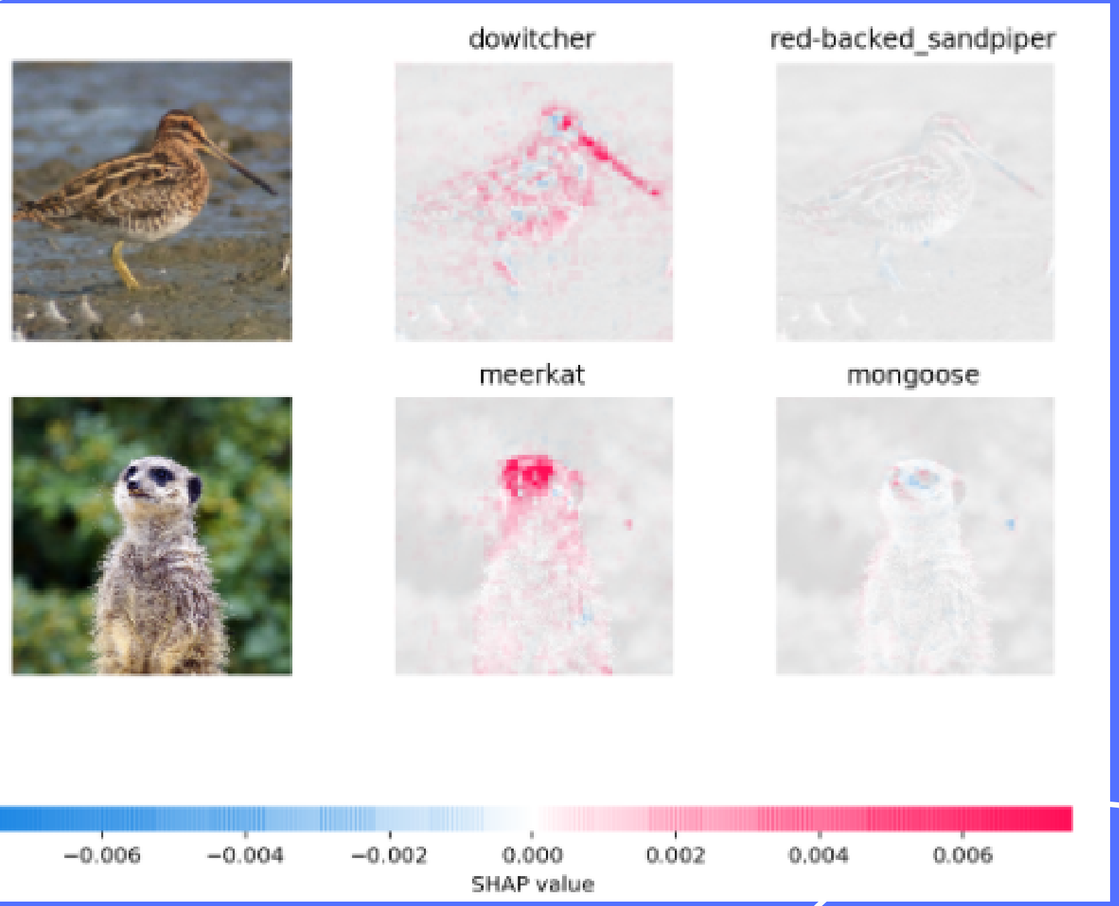
SHAPLEY ADDITIVE EXPLANATIONS



Entre as técnicas de explicabilidade amplamente utilizadas temos o SHAP , que consiste em uma abordagem teórica do jogo para explicar a saída de qualquer modelo de aprendizado de máquina.

A teoria dos jogos consiste em um ramo da matemática que estuda o comportamento de agentes diante de diferentes estratégias levando em conta conflito e cooperação destes. Esta abordagem surgiu inicialmente na Ciência Econômica para o estudo da tomada de decisões de pessoas em ambientes com bens escassos.

O Valor do Shaply pode ser interpretado como a contribuição da feature para a diferença entre o valor previsto e o valor médio. O curso computacional deste método é elevado, além de ser custoso realizar sua implementação. Entretanto, o artigo ''A Unified Approach to Interpreting Model Predictions'' possibilitou a expansão da utilização deste método visto que seus autores compartilharam os códigos utilizados através da plataforma GitHub,



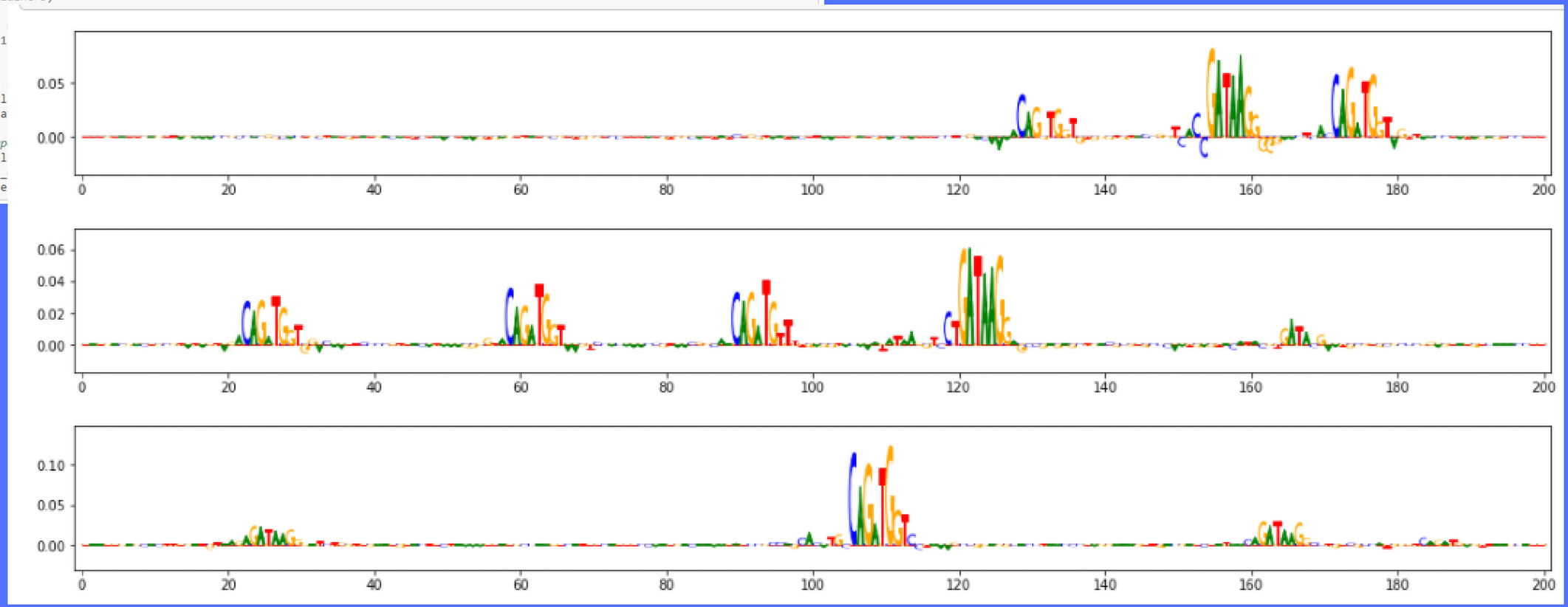
TRABALHANDO COM DIFERENTES TIPOS DE DADOS

Visualmente o SHAP consegue te dizer quais foram os pontos da imagem que contribuíram para o algoritmo chegar em determinada resposta, por exemplo no caso das espécies.

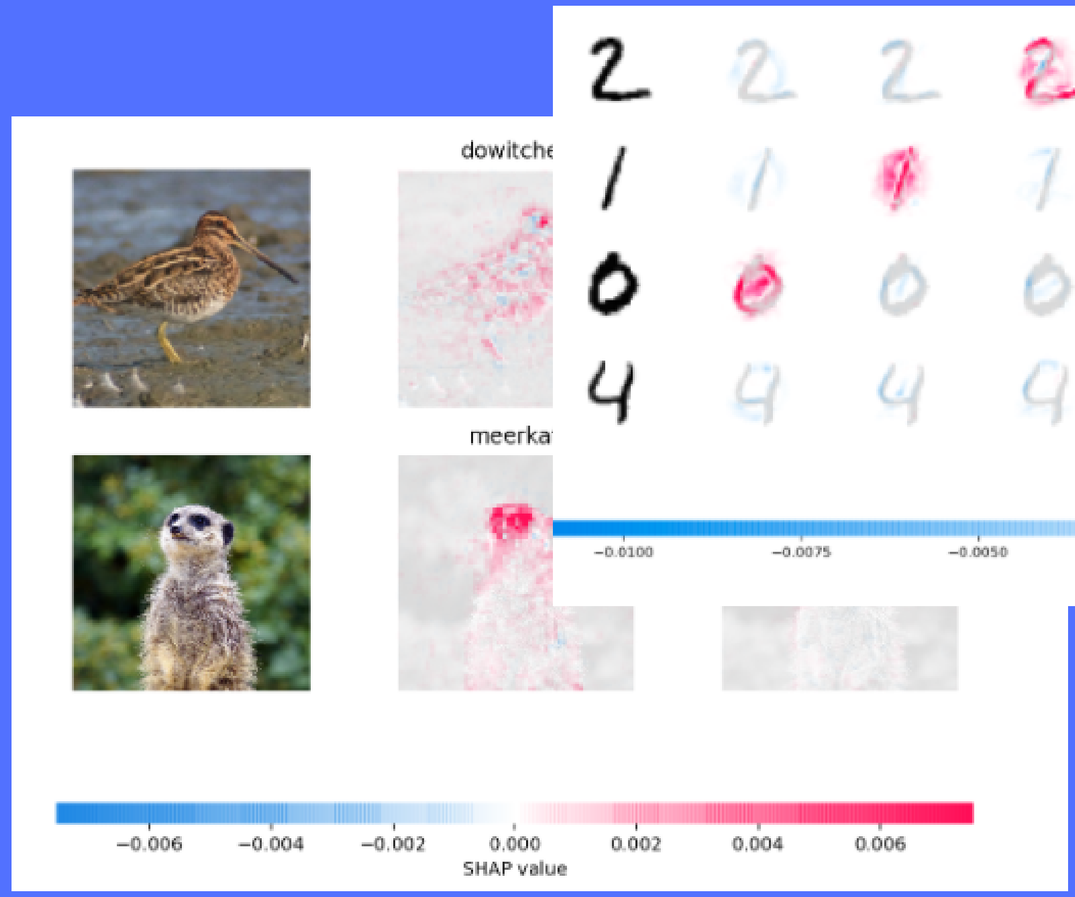
Ele também pode ser aplicado para validar modelos que envolvam tabelas, dados de texto, dados genômicos, etc.

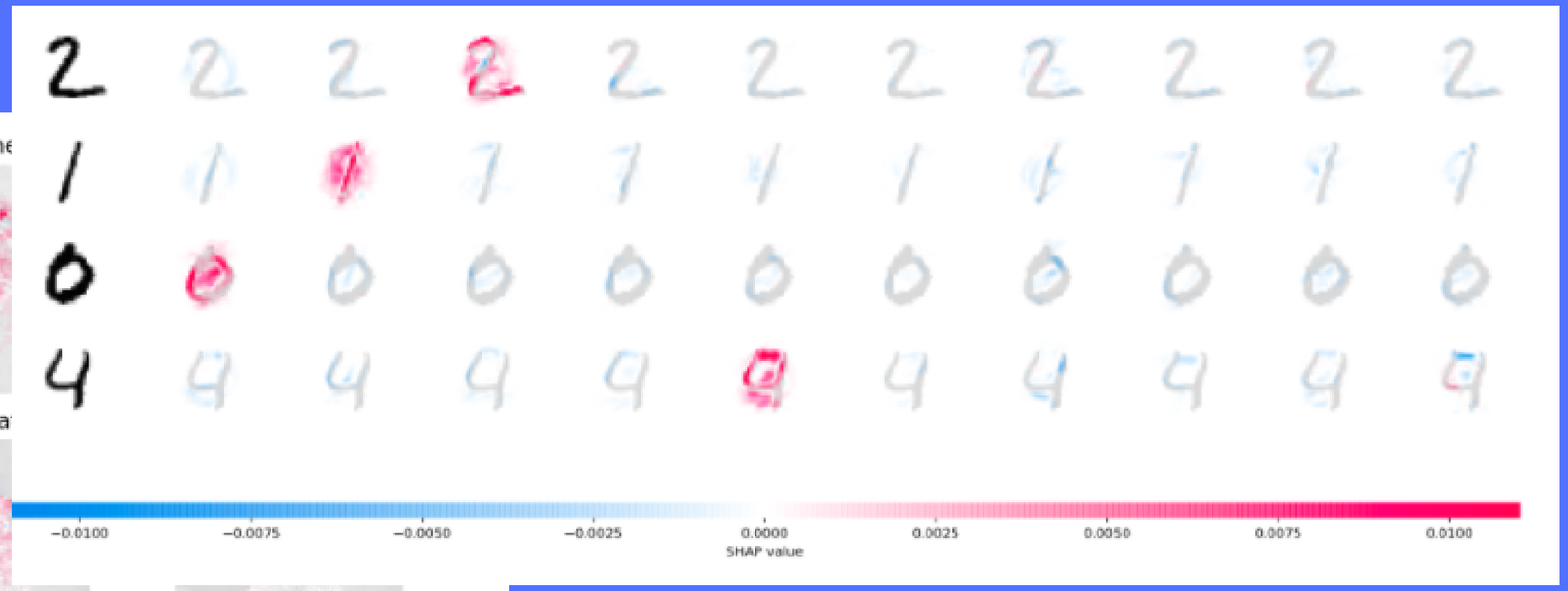
EXEMPLOS (NOTEBOOKS DISPONÍVEIS)

```
from deeplift.visualization import viz_sequence
import shap
import shap.explainers.deep.deep_tf
reload(shap.explainers.deep.deep_tf)
reload(shap.explainers.deep)
reload(shap.explainers)
reload(shap)
import numpy as
np.random.seed(1
import random

seqs_to_explain
dinuc_shuff_expl
raw_shap_explana

#project the imp
dinuc_shuff_expl
for dinuc_shuff_
    viz_sequence
```

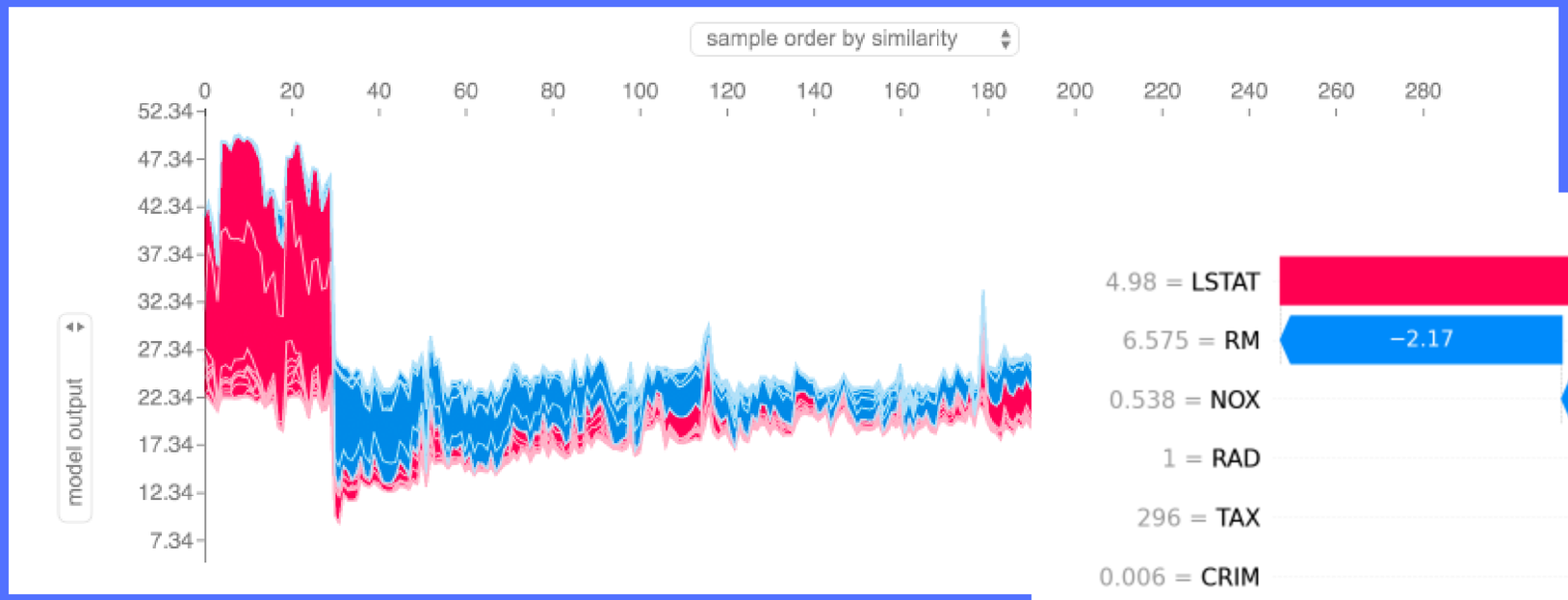


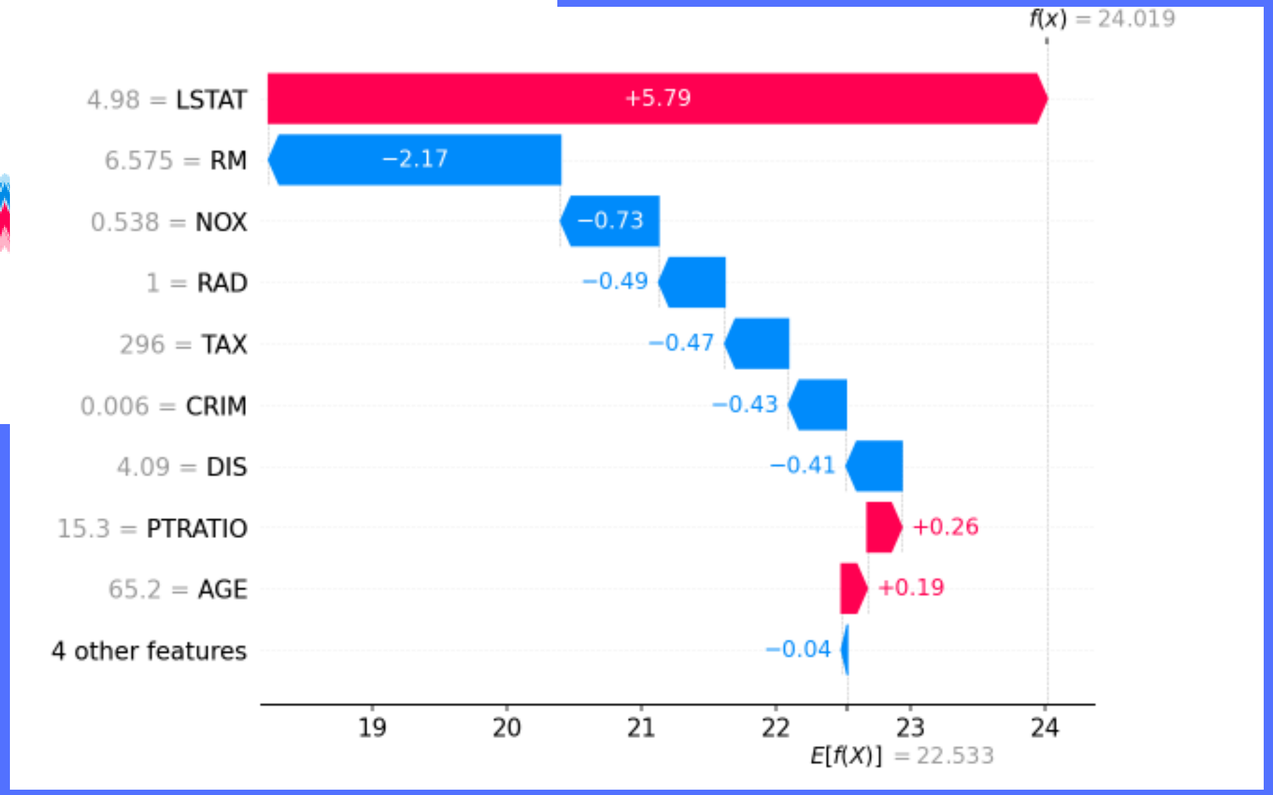
EXEMPLOS (NOTEBOOKS DISPONÍVEIS)





EXEMPLOS (NOTEBOOKS DISPONÍVEIS)





COMO UTILIZAR?

Importar a biblioteca Shap

Passar seu modelo e suas imagens para o shap