Introduction

Welcome!

Thank you for taking part in this survey. The survey consists of 50 questions, which should take approximately 15-25 minutes to complete.
I am a student working on my thesis with the **Data Science & Analytics** group of the Computer Science Department of the University of Tübingen in collaboration with the **Human-Centered Technologies for Learning** group of the Technical University of Munich (TUM). We are studying how automated decisions can be efficiently explained to users. If you have any questions about this study, please feel free to contact me at **thai-trang.nguyen@student.uni-tuebingen.de**.

Responses will be confidential

All records from this study will be processed anonymously and any sort of publications will not include any personal information which may potentially be used to reidentify participants.

By clicking "I agree", you consent to participate in this study.

- ○ I agree.
- ○ I do not agree.

Prolific ID

What is your Prolific ID?
Please note that this response should auto-fill with the correct ID.

Pre-Survey

Please answer the following questions.

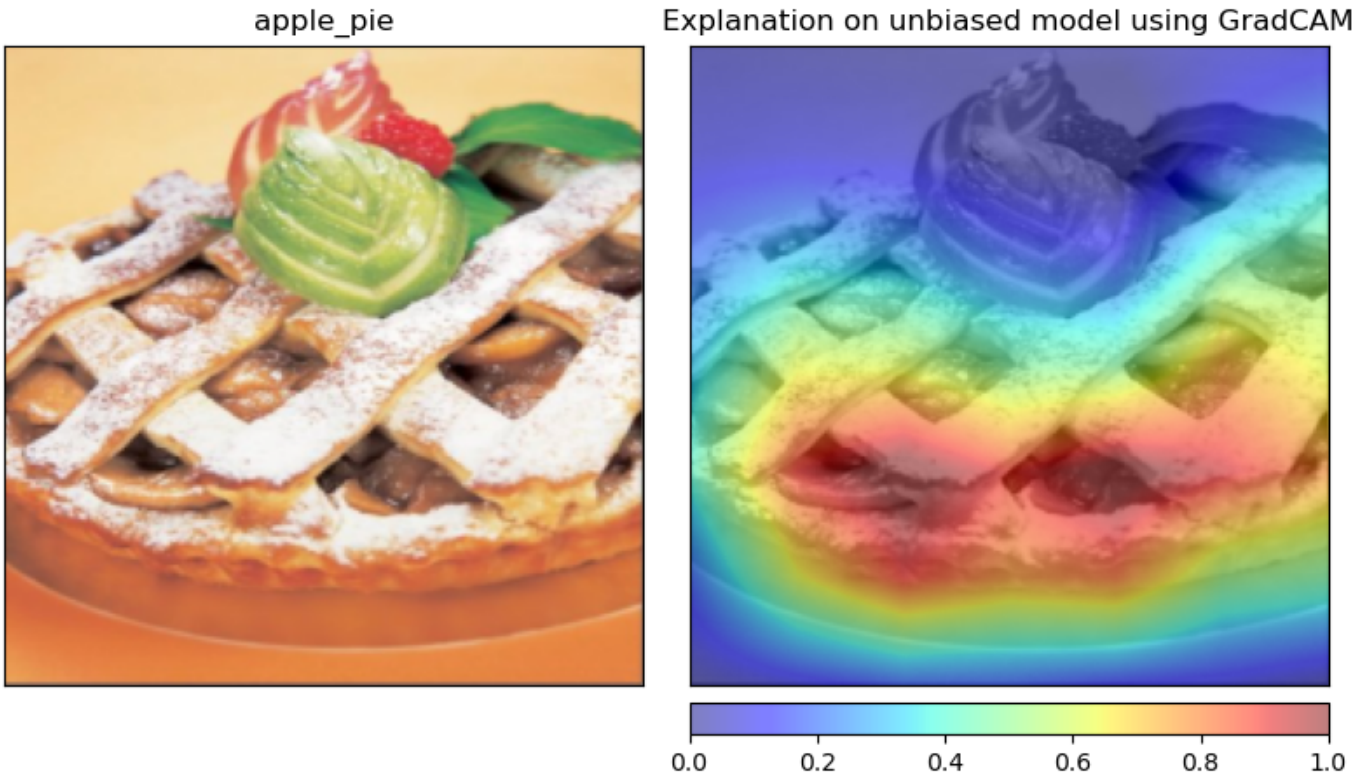
	Not knowledgeable at all	Slightly knowledgeable	Moderately knowledgeable	Very knowledgeable	Extremely knowledgeable
How much programming knowledge do you have?	○	○	○	○	○

	Not knowledgeable at all	Slightly knowledgeable	Moderately knowledgeable	Very knowledgeable	Extremely knowledgeable
How much knowledge of computer algorithms do you have?	○	○	○	○	○
If you have at least moderate knowledge of computer algorithms, how much knowledge do you have regarding machine learning?	○	○	○	○	○

Instruction 1

Machine learning (ML) is a way for computers to learn from data without being explicitly told how.
Consider the following scenario: There is a machine learning model that **can recognize different types of food** by analyzing pictures. The model uses the details presented in each picture to make its decision. Although it is difficult to understand how the model works, there are methods to help explain its decisions by showing which parts of the picture were most important for the decision.

Below is an example of a picture with the corresponding explanation generated by the ML model trying to explain its decision. The left picture shows an "apple_pie" and the right side shows how the model explained its decision - red image parts had a big impact, while blue image parts had a small impact.



In the following, there are 5 - 7 questions related to the given scenario. These questions ask you to classify each image (i.e., how would you describe the food in the given image) and will help you familiarize yourself with the model. Take into account that an answer can be given multiple times.
But first, to ensure you have understood the information above correctly, please answer the following 2 questions about the instruction.

Which of the following statements is **true** about the mentioned model?

- The model is a protocol to follow in order to classify images of food as oranges or no oranges.
- The model is a computer program that automatically classify different types of food.
- The model is a computer program that randomly generates a number.

Which of the following statements is **true** about the mentioned model?

- Blue parts of the image have a high influence on the model's prediction.
- Red parts of the image have a high influence on the model's prediction.
- The colored parts of the image have no impact on the prediction of the model.

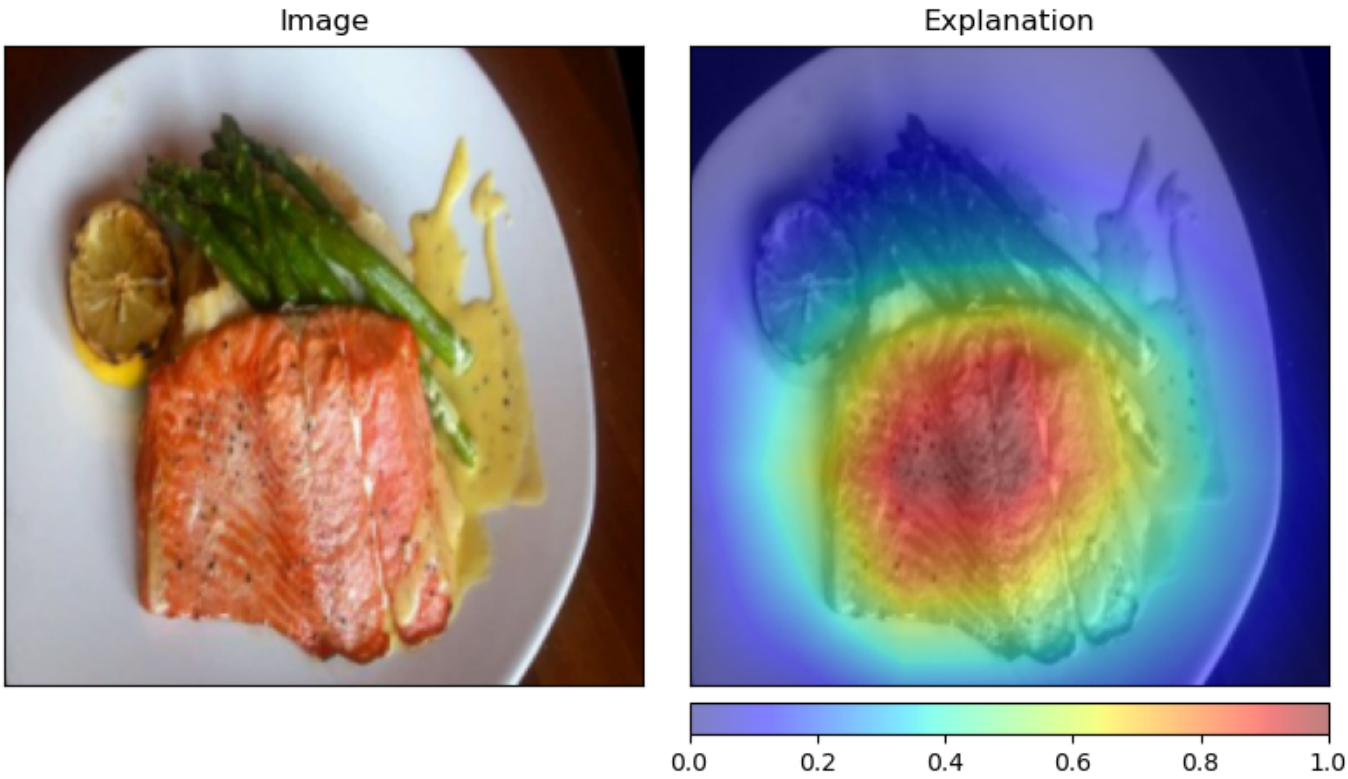
Condition 1 - Intuitiveness / Faithfulness

Consider the following image:

How would you classify this image?

- baby_back_ribs
- grilled_salmon
- steak
- filet_mignon
- pork_chop
- prime_rib
- I don't know

The model classifies the image as **grilled_salmon**. Below is an explanation of the pixel importance that the machine learning model's decision is based on.



Based on this explanation, do you agree with the classification of the model?

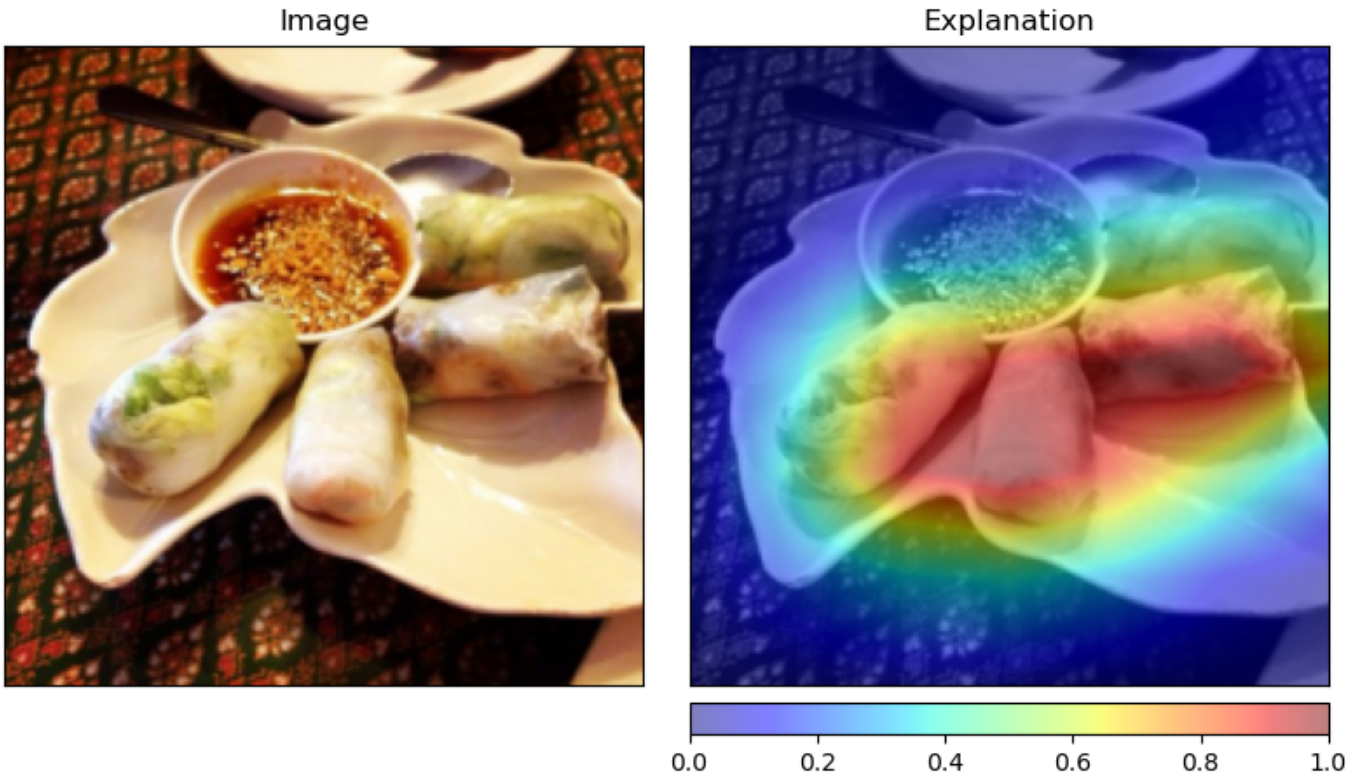
- ○ No, I do not agree.
- ○ Yes, I agree.

Consider the following image:

How would you classify this image?

- peking_duck
- gyoza
- tacos
- sushi
- dumplings
- spring_rolls
- I don't know

The model classifies the image as **spring_rolls**. Below is an explanation of the pixel importance that the machine learning model's decision is based on.



Based on this explanation, do you agree with the classification of the model?

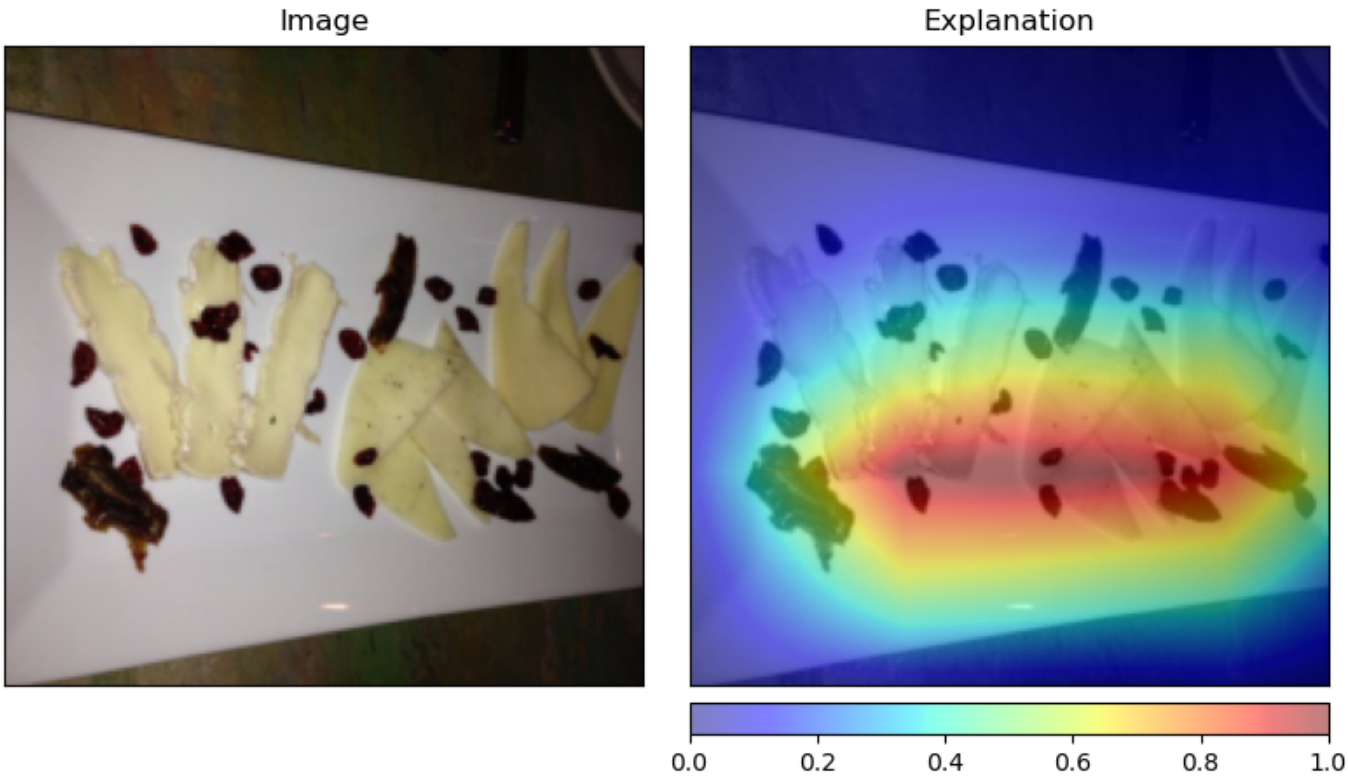
- ○ No, I do not agree.
- ○ Yes, I agree.

Consider the following image:

How would you classify this image?

- ○ cheese_plate
- ○ panna_cotta
- ○ sushi
- ○ chocolate_mousse
- ○ tiramisu
- ○ chocolate_cake
- ○ I don't know

The model classifies the image as **cheese_plate**. Below is an explanation of the pixel importance that the machine learning model's decision is based on.



Based on this explanation, do you agree with the classification of the model?

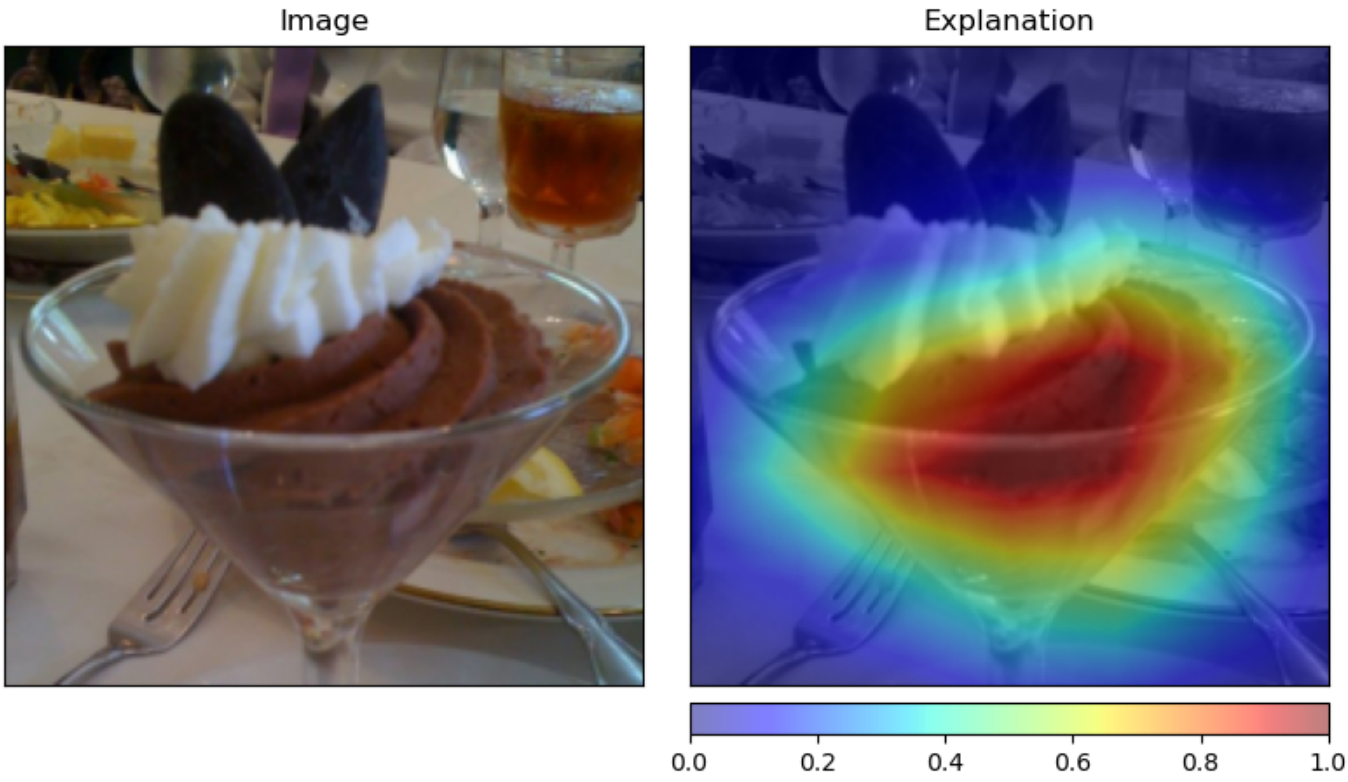
- ○ No, I do not agree.
- ○ Yes, I agree.

Consider the following image:

How would you classify this image?

- dumplings
- chocolate_mousse
- ceviche
- frozen_yogurt
- ice_cream
- waffles
- I don't know

The model classifies the image as **chocolate_mousse**. Below is an explanation of the pixel importance that the machine learning model's decision is based on.



Based on this explanation, do you agree with the classification of the model?

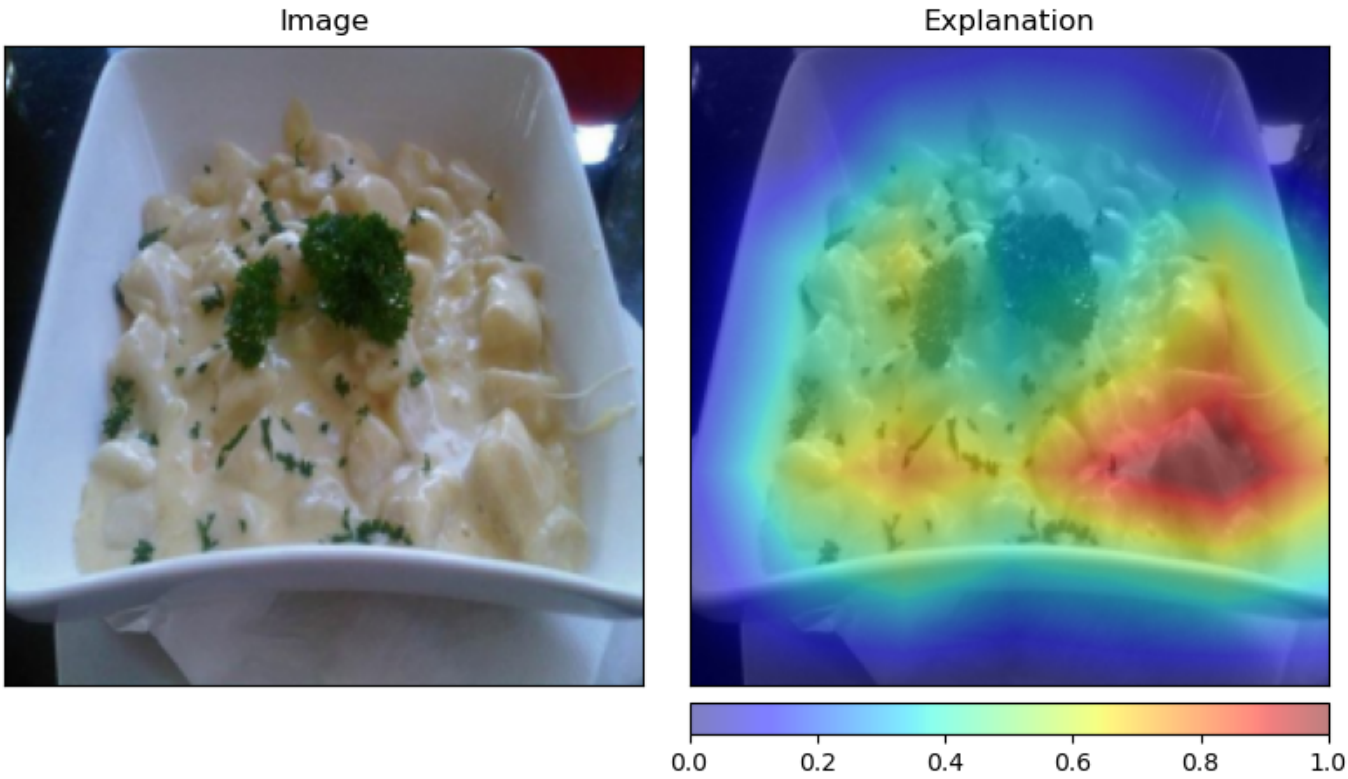
- ○ No, I do not agree.
- ○ Yes, I agree.

Consider the following image:

How would you classify this image?

- ravioli
- macaroni_and_cheese
- spaghetti_carbonara
- gyoza
- gnocchi
- risotto
- I don't know

The model classifies the image as **macaroni_and_cheese**. Below is an explanation of the pixel importance that the machine learning model's decision is based on.



Based on this explanation, do you agree with the classification of the model?

- ○ No, I do not agree.
- ○ Yes, I agree.

If you don't feel familiar enough with the model yet, you can get 2 more examples.

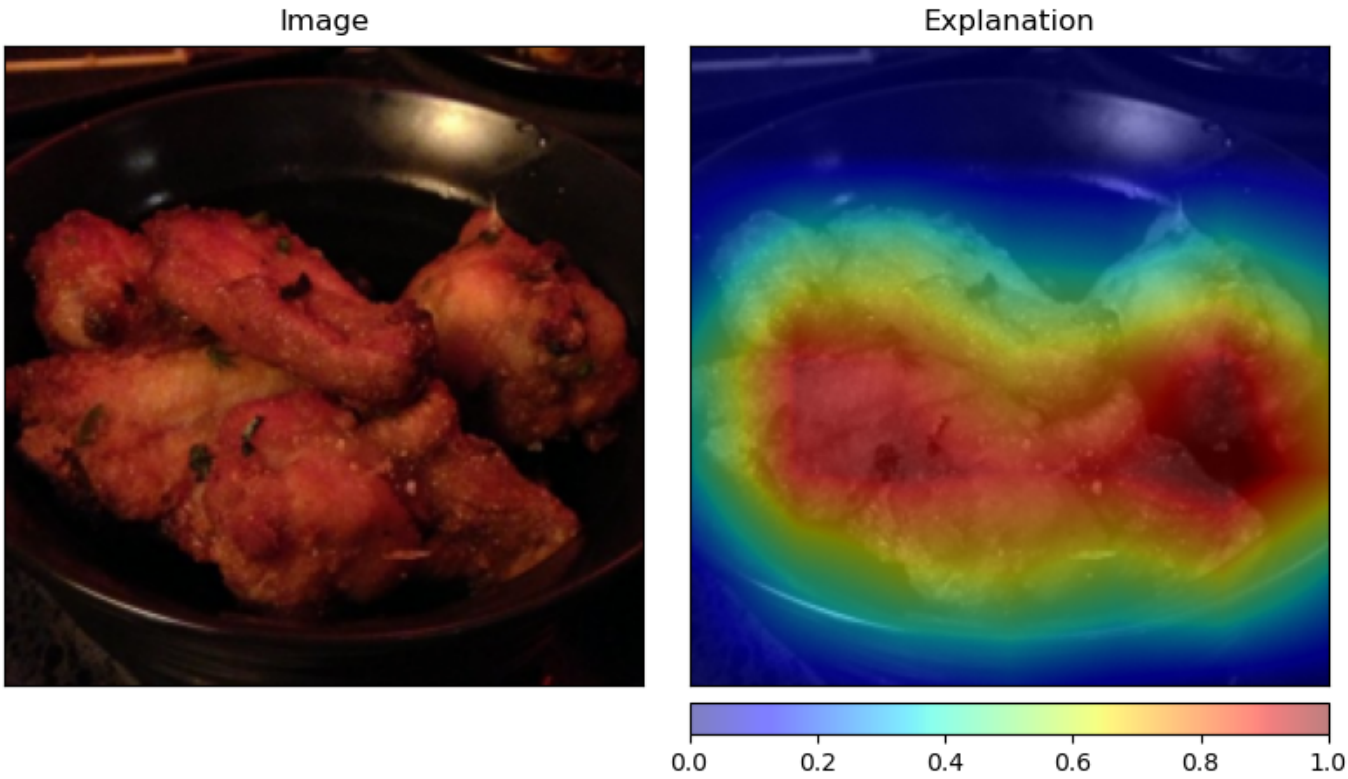
- ○ I need more examples to become familiar with the model.
- ○ I am familiar enough and do not need more examples.

Consider the following image:

How would you classify this image?

- beignets
- steak
- chicken_wings
- gyoza
- takoyaki
- chicken_curry
- I don't know

The model classifies the image as **chicken_wings**. Below is an explanation of the pixel importance that the machine learning model's decision is based on.



Based on this explanation, do you agree with the classification of the model?

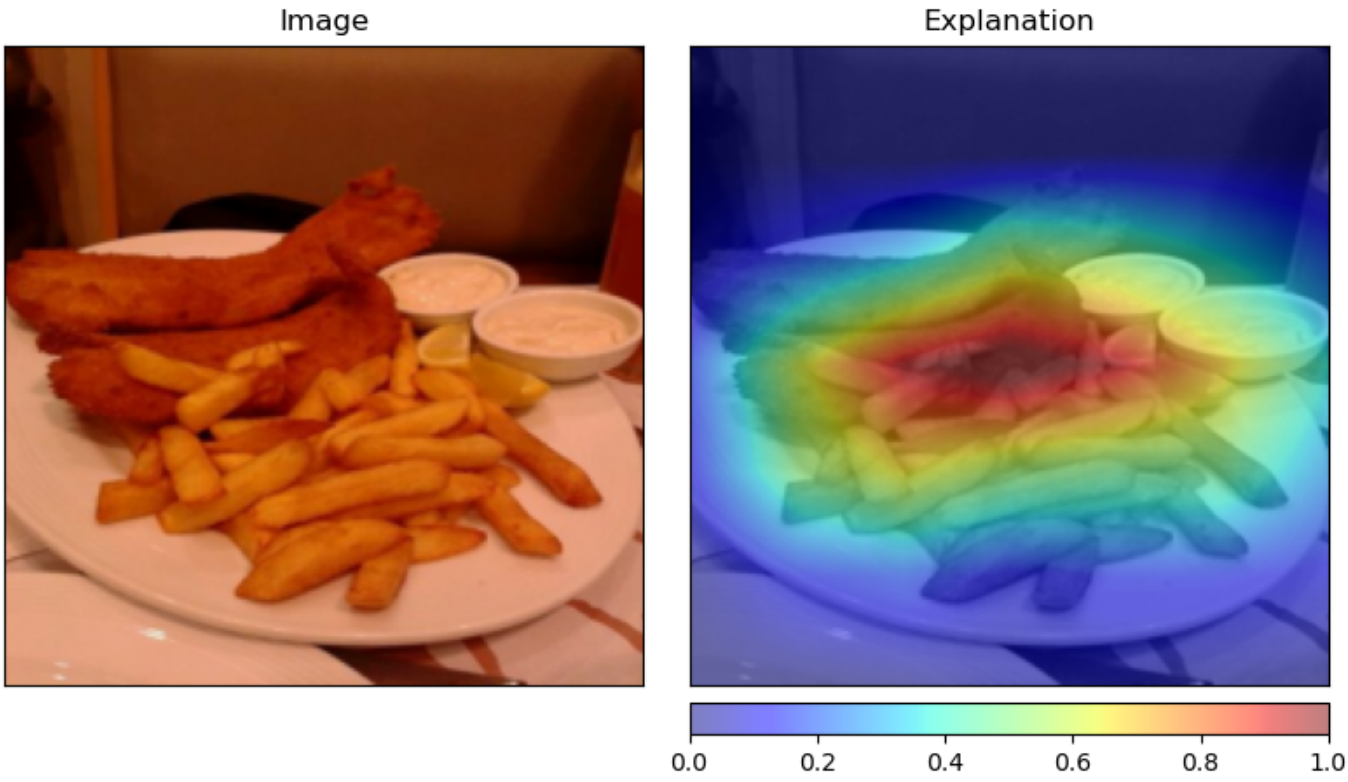
- ○ No, I do not agree.
- ○ Yes, I agree.

Consider the following image:

How would you classify this image?

- cheese_plate
- french_fries
- fried_calamari
- onion_rings
- chicken_wings
- fish_and_chips
- I don't know

The model classifies the image as **fish_and_chips**. Below is an explanation of the pixel importance that the machine learning model's decision is based on.



Based on this explanation, do you agree with the classification of the model?

- ○ No, I do not agree.
- ○ Yes, I agree.

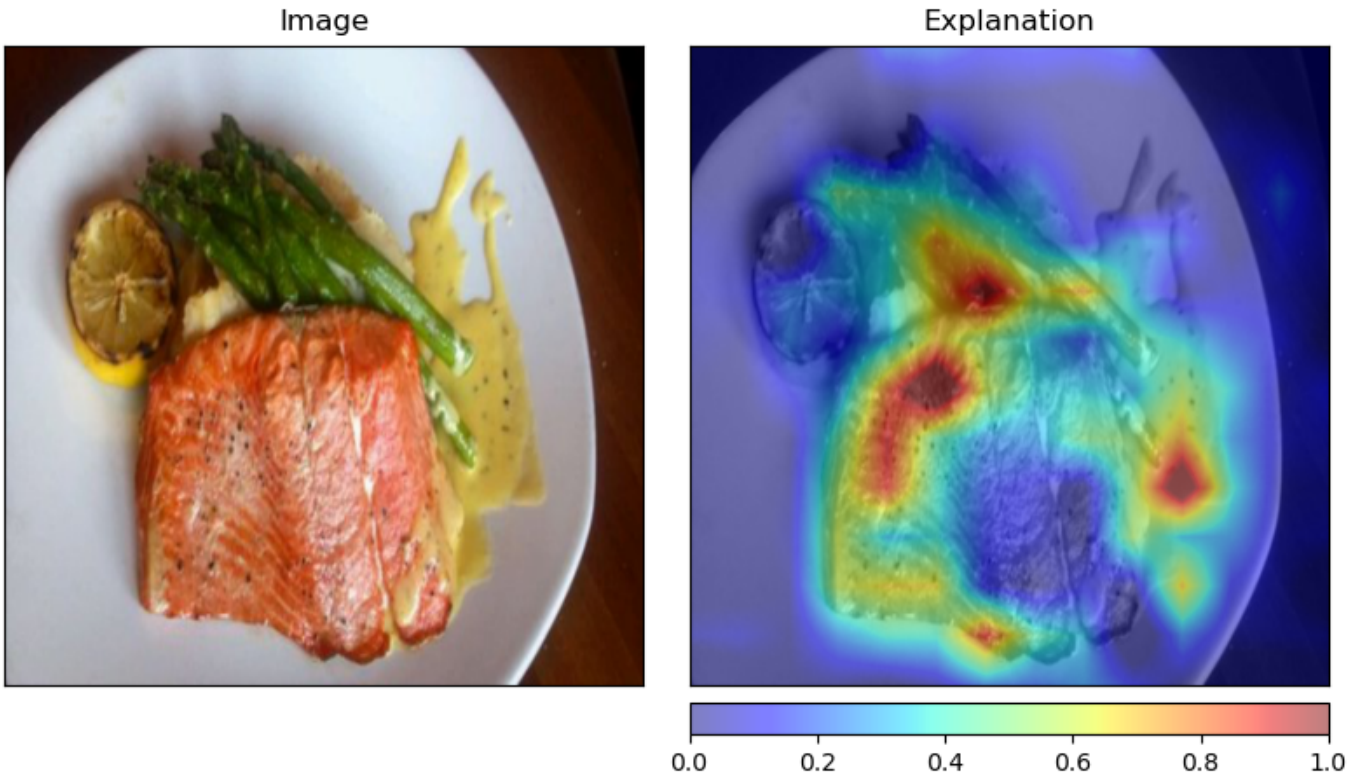
Condition 2 - Intuitiveness / Unfaithfulness

Consider the following image:

How would you classify this image?

- baby_back_ribs
- grilled_salmon
- steak
- filet_mignon
- pork_chop
- prime_rib
- I don't know

The model classifies the image as **grilled_salmon**. Below is an explanation of the pixel importance that the machine learning model's decision is based on.



Based on this explanation, do you agree with the classification of the model?

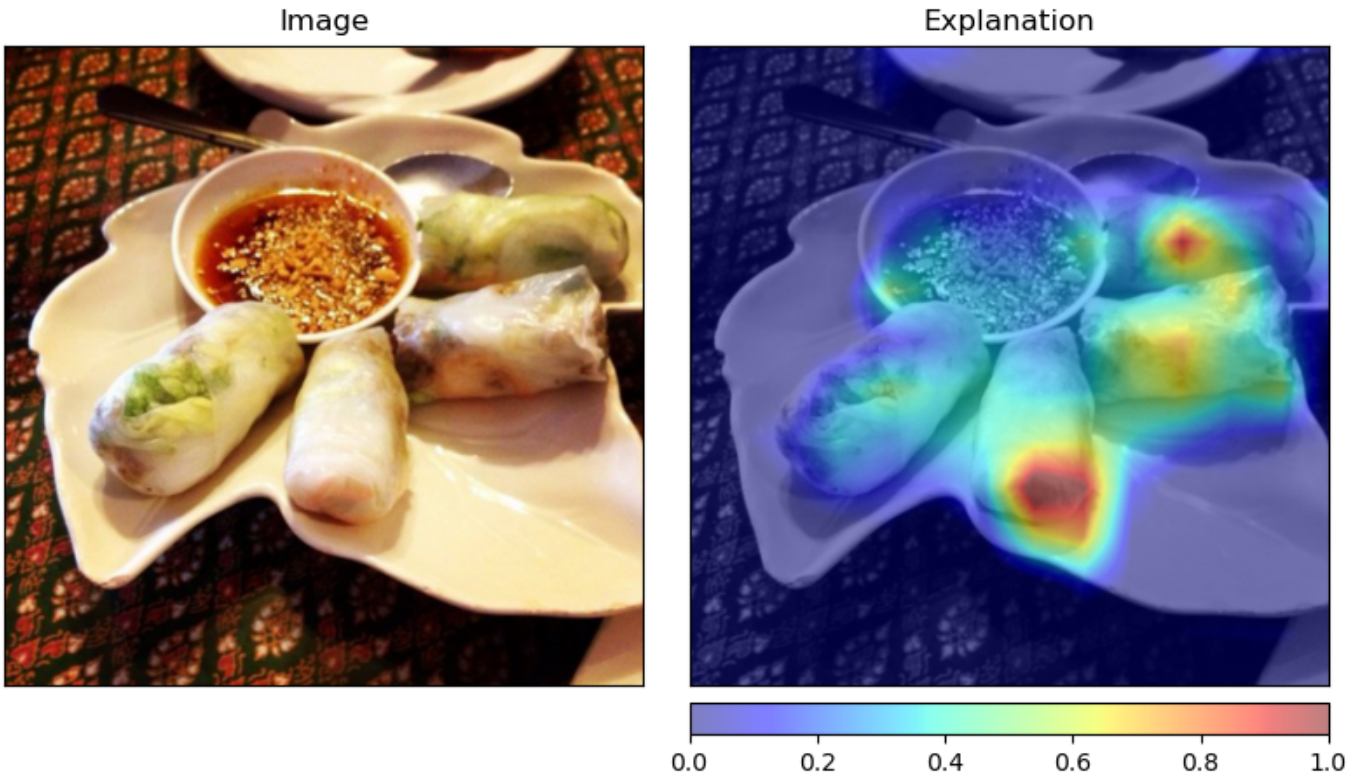
- ○ No, I do not agree.
- ○ Yes, I agree.

Consider the following image:

How would you classify this image?

- peking_duck
- gyoza
- tacos
- sushi
- dumplings
- spring_rolls
- I don't know

The model classifies the image as **spring_rolls**. Below is an explanation of the pixel importance that the machine learning model's decision is based on.



Based on this explanation, do you agree with the classification of the model?

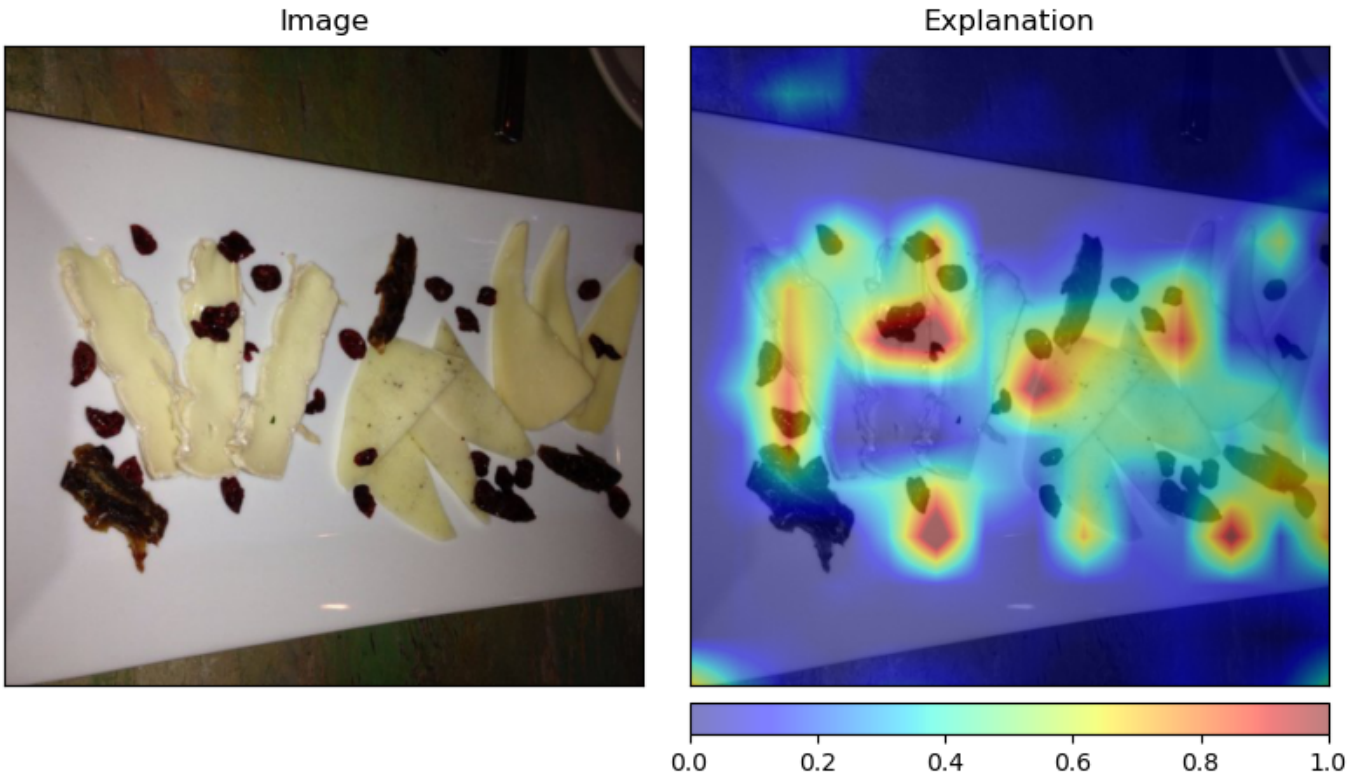
- ○ No, I do not agree.
- ○ Yes, I agree.

Consider the following image:

How would you classify this image?

- ○ cheese_plate
- ○ panna_cotta
- ○ sushi
- ○ chocolate_mousse
- ○ tiramisu
- ○ chocolate_cake
- ○ I don't know

The model classifies the image as **cheese_plate**. Below is an explanation of the pixel importance that the machine learning model's decision is based on.



Based on this explanation, do you agree with the classification of the model?

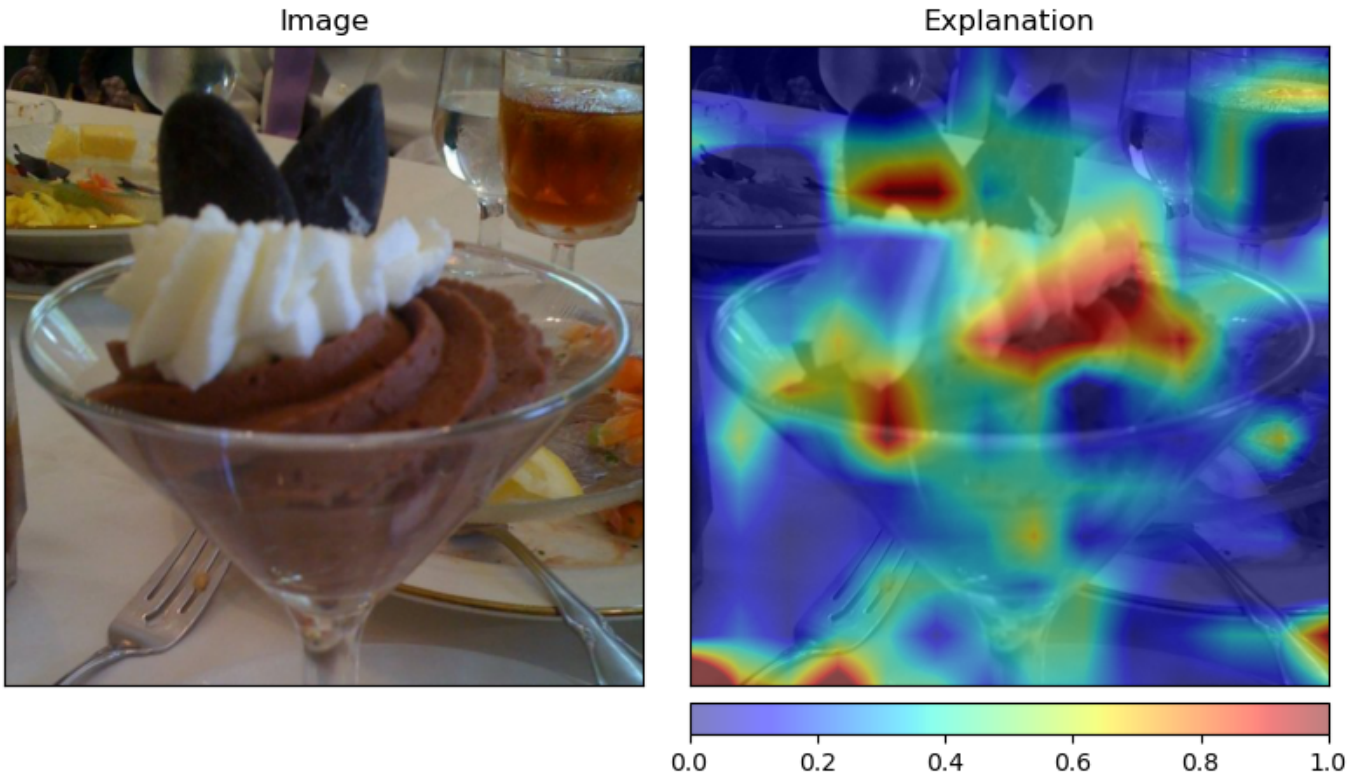
- ○ No, I do not agree.
- ○ Yes, I agree.

Consider the following image:

How would you classify this image?

- dumplings
- chocolate_mousse
- ceviche
- frozen_yogurt
- ice_cream
- waffles
- I don't know

The model classifies the image as **chocolate_mousse**. Below is an explanation of the pixel importance that the machine learning model's decision is based on.



Based on this explanation, do you agree with the classification of the model?

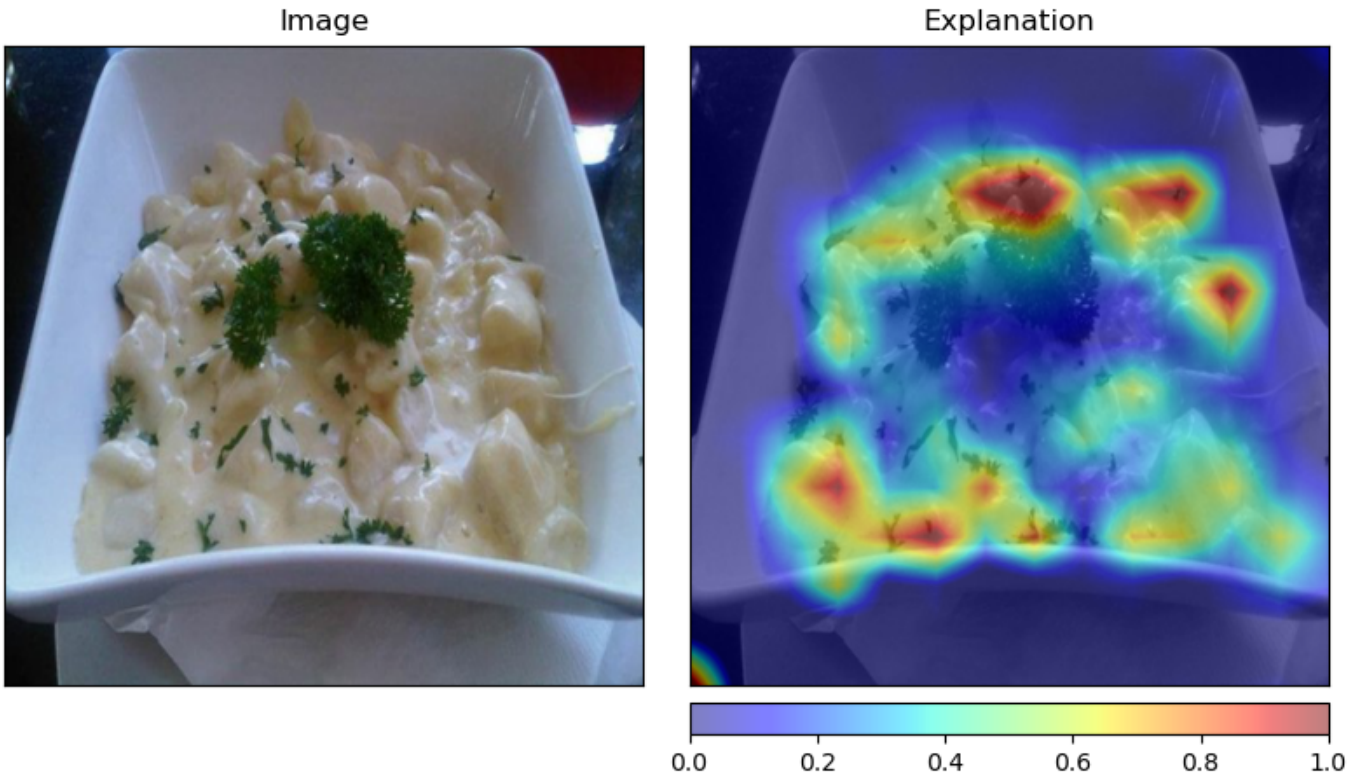
- ○ No, I do not agree.
- ○ Yes, I agree.

Consider the following image:

How would you classify this image?

- ravioli
- macaroni_and_cheese
- spaghetti_carbonara
- gyoza
- gnocchi
- risotto
- I don't know

The model classifies the image as **macaroni_and_cheese**. Below is an explanation of the pixel importance that the machine learning model's decision is based on.



Based on this explanation, do you agree with the classification of the model?

- ○ No, I do not agree.
- ○ Yes, I agree.

If you don't feel familiar enough with the model yet, you can get 2 more examples.

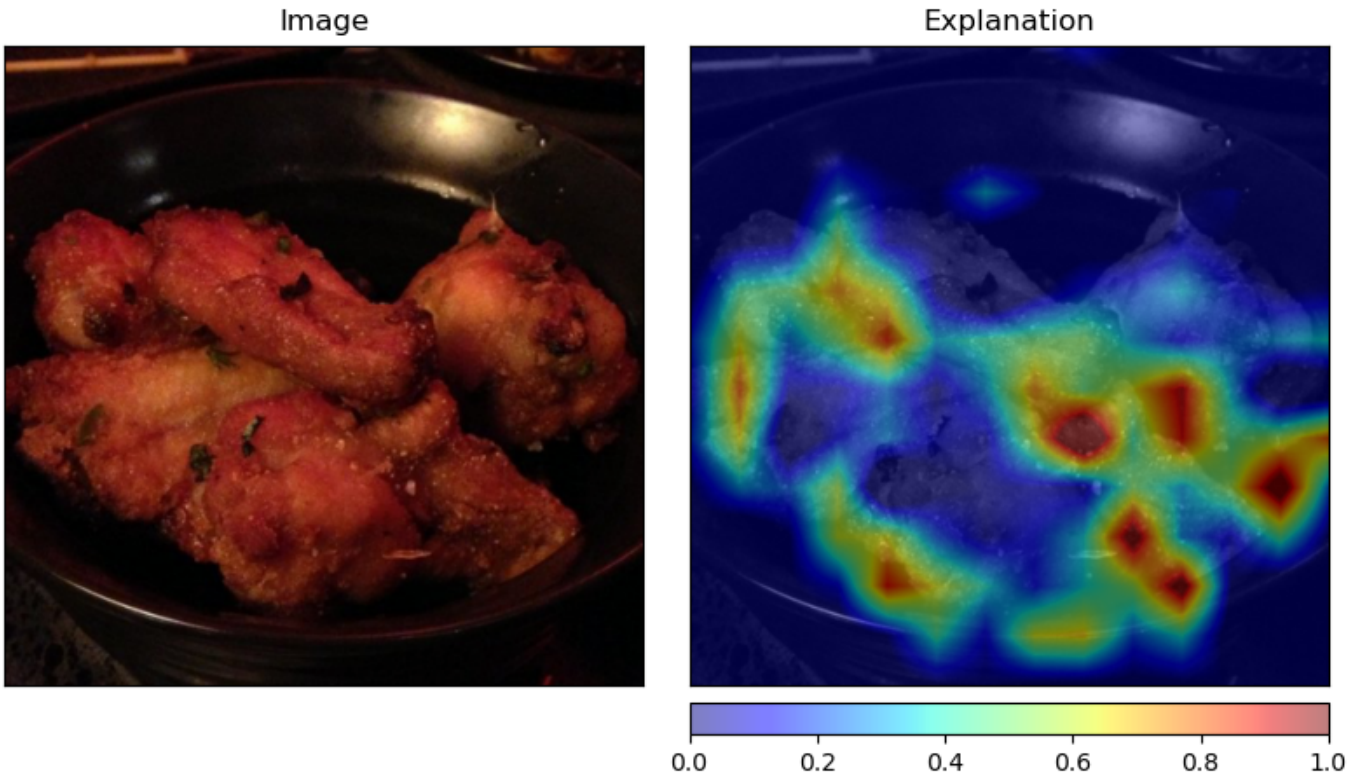
- ○ I need more examples to become familiar with the model.
- ○ I am familiar enough and do not need more examples.

Consider the following image:

How would you classify this image?

- beignets
- steak
- chicken_wings
- gyoza
- takoyaki
- chicken_curry
- I don't know

The model classifies the image as **chicken_wings**. Below is an explanation of the pixel importance that the machine learning model's decision is based on.



Based on this explanation, do you agree with the classification of the model?

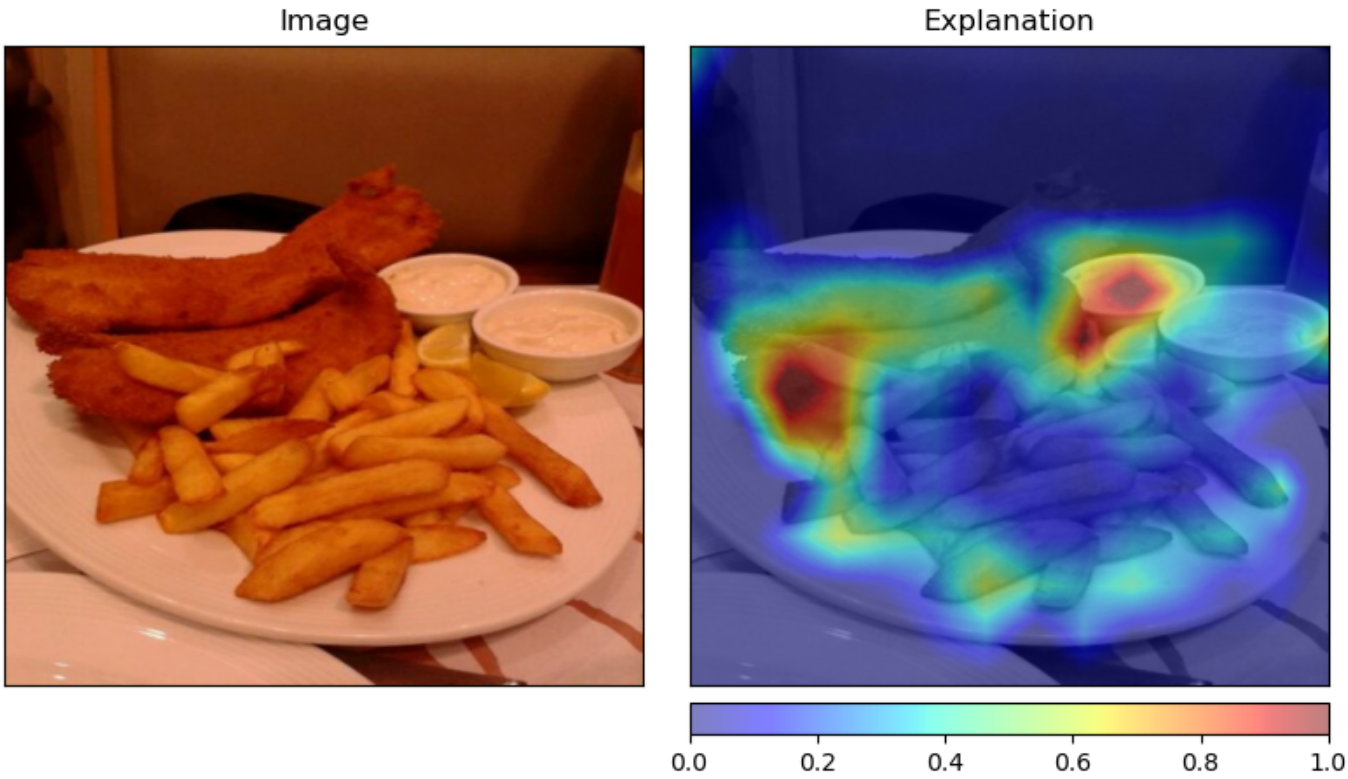
- ○ No, I do not agree.
- ○ Yes, I agree.

Consider the following image:

How would you classify this image?

- cheese_plate
- french_fries
- fried_calamari
- onion_rings
- chicken_wings
- fish_and_chips
- I don't know

The model classifies the image as **fish_and_chips**. Below is an explanation of the pixel importance that the machine learning model's decision is based on.



Based on this explanation, do you agree with the classification of the model?

- ○ No, I do not agree.
- ○ Yes, I agree.

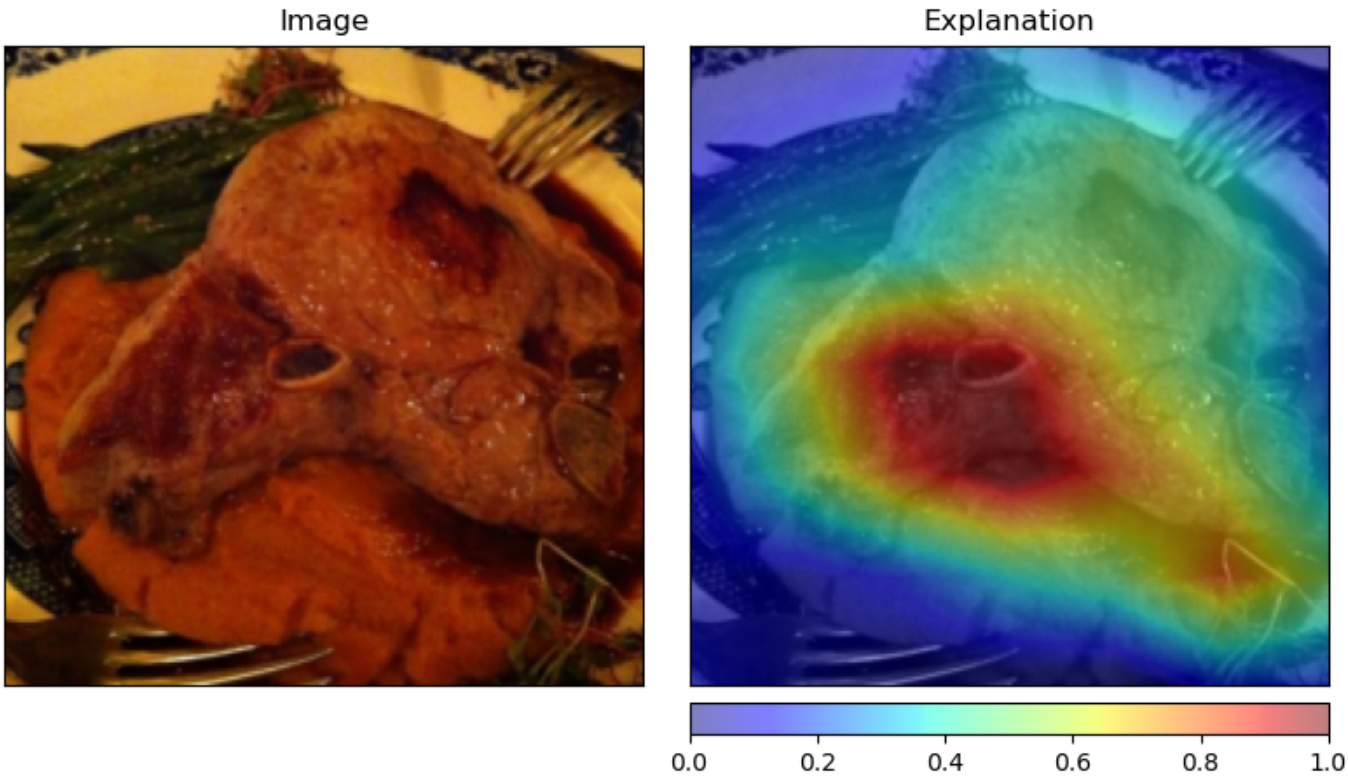
Condition 3 - Non-intuitiveness / Faithfulness

Consider the following image:

How would you classify this image?

- foie_gras
- grilled_salmon
- prime_rib
- steak
- peking_duck
- pork_chop
- I don't know

The model classifies the image as **pork_chop**. Below is an explanation of the pixel importance that the machine learning model's decision is based on.



Based on this explanation, do you agree with the classification of the model?

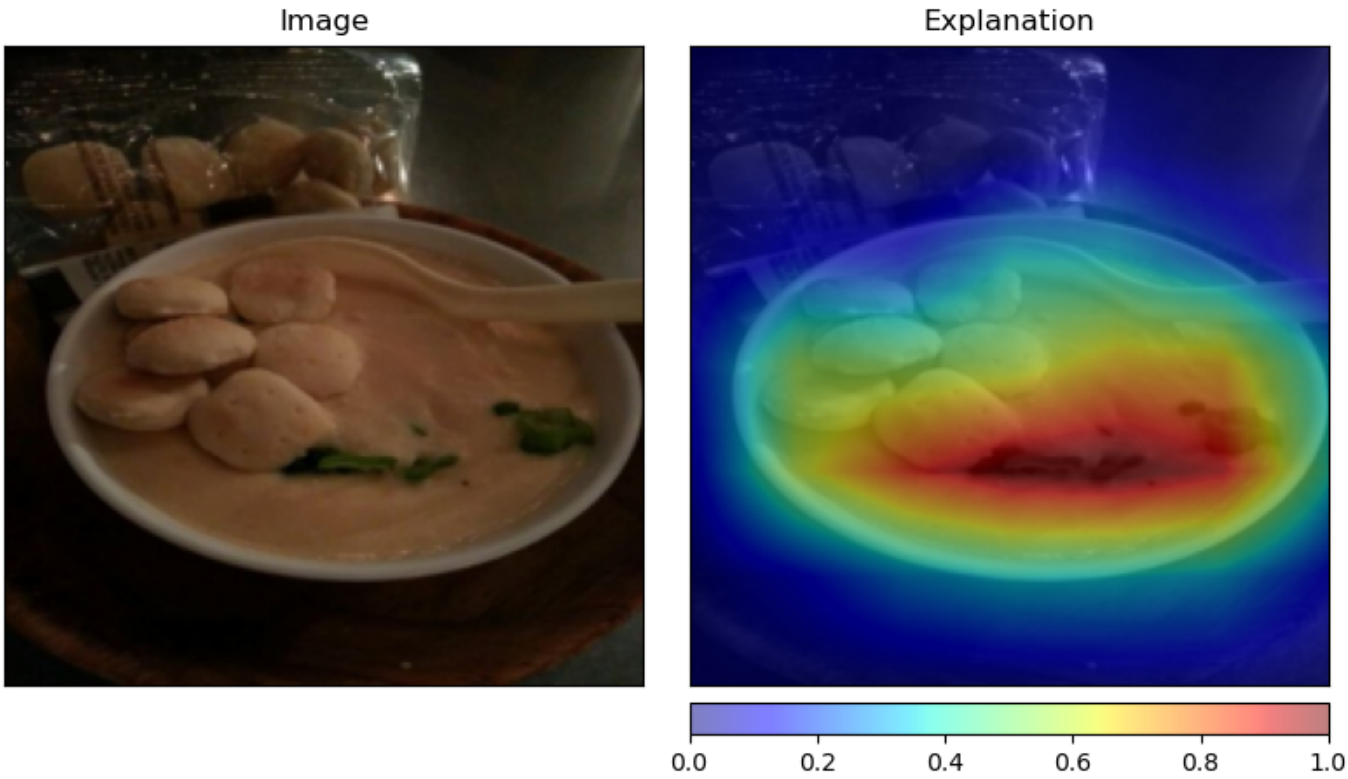
- ○ No, I do not agree.
- ○ Yes, I agree.

Consider the following image:

How would you classify this image?

- clam_chowder
- lobster_bisque
- beignets
- eggs_benedict
- hummus
- french_onion_soup
- I don't know

The model classifies the image as **lobster_bisque**. Below is an explanation of the pixel importance that the machine learning model's decision is based on.



Based on this explanation, do you agree with the classification of the model?

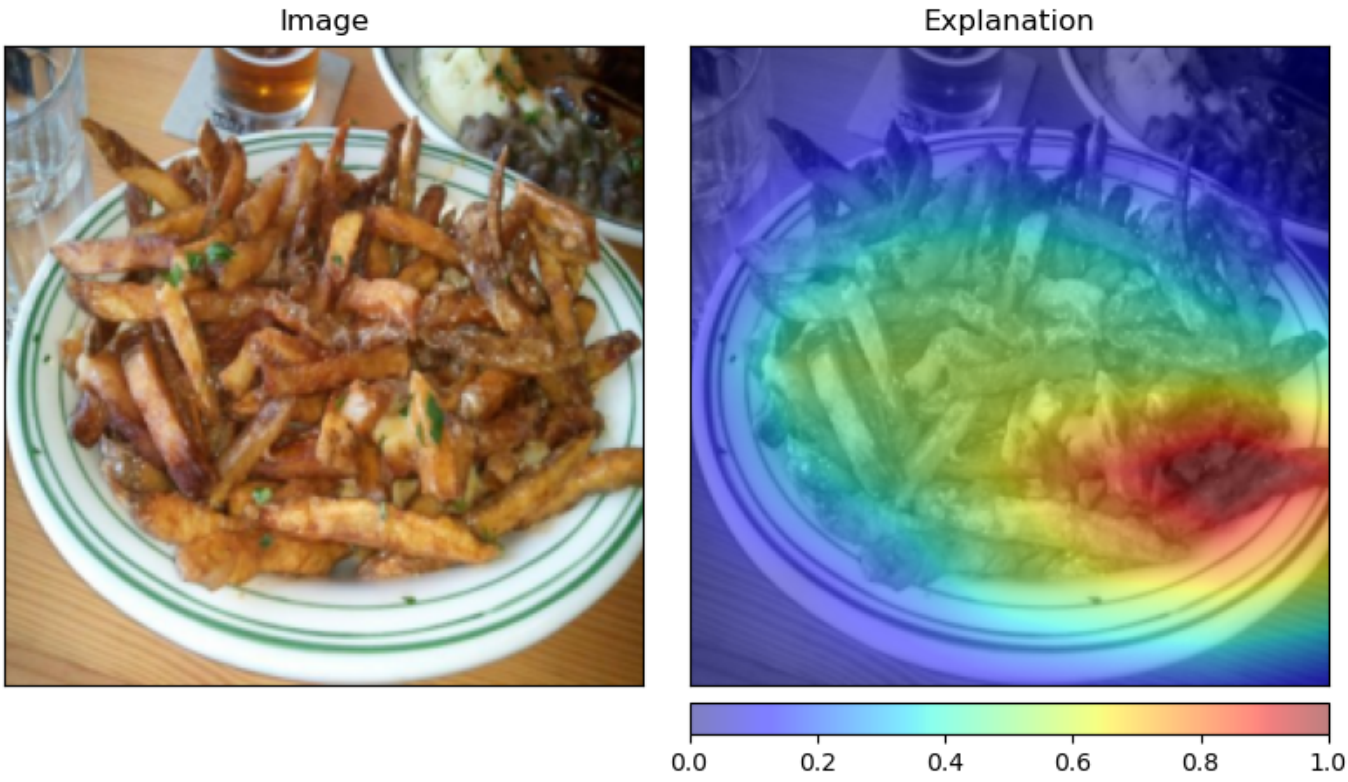
- ○ No, I do not agree.
- ○ Yes, I agree.

Consider the following image:

How would you classify this image?

- ○ french_fries
- ○ pulled_pork_sandwich
- ○ pad_thai
- ○ club_sandwich
- ○ poutine
- ○ baby_back_ribs
- ○ I don't know

The model classifies the image as **poutine**. Below is an explanation of the pixel importance that the machine learning model's decision is based on.



Based on this explanation, do you agree with the classification of the model?

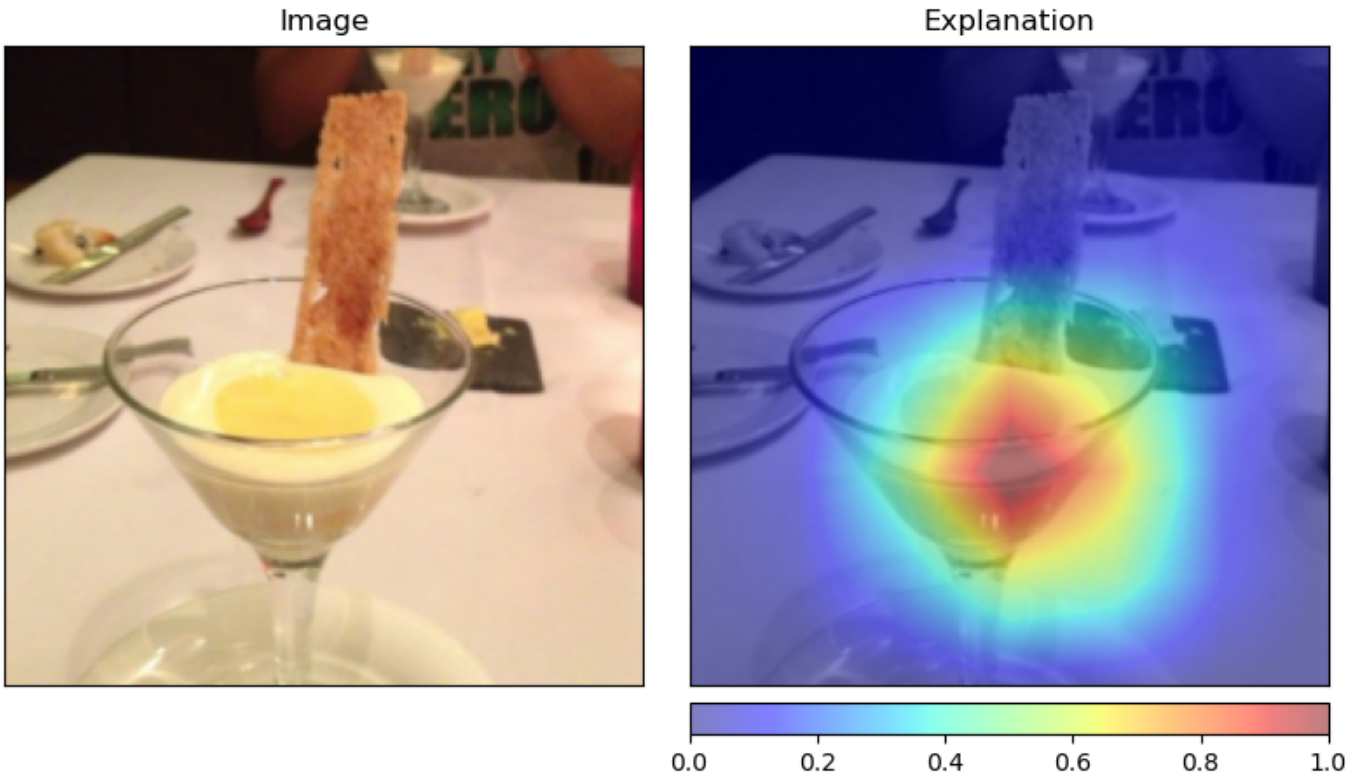
- ○ No, I do not agree.
- ○ Yes, I agree.

Consider the following image:

How would you classify this image?

- ice_cream
- eggs_benedict
- tuna_tartare
- panna_cotta
- chocolate_mousse
- creme_brulee
- I don't know

The model classifies the image as **eggs_benedict**. Below is an explanation of the pixel importance that the machine learning model's decision is based on.



Based on this explanation, do you agree with the classification of the model?

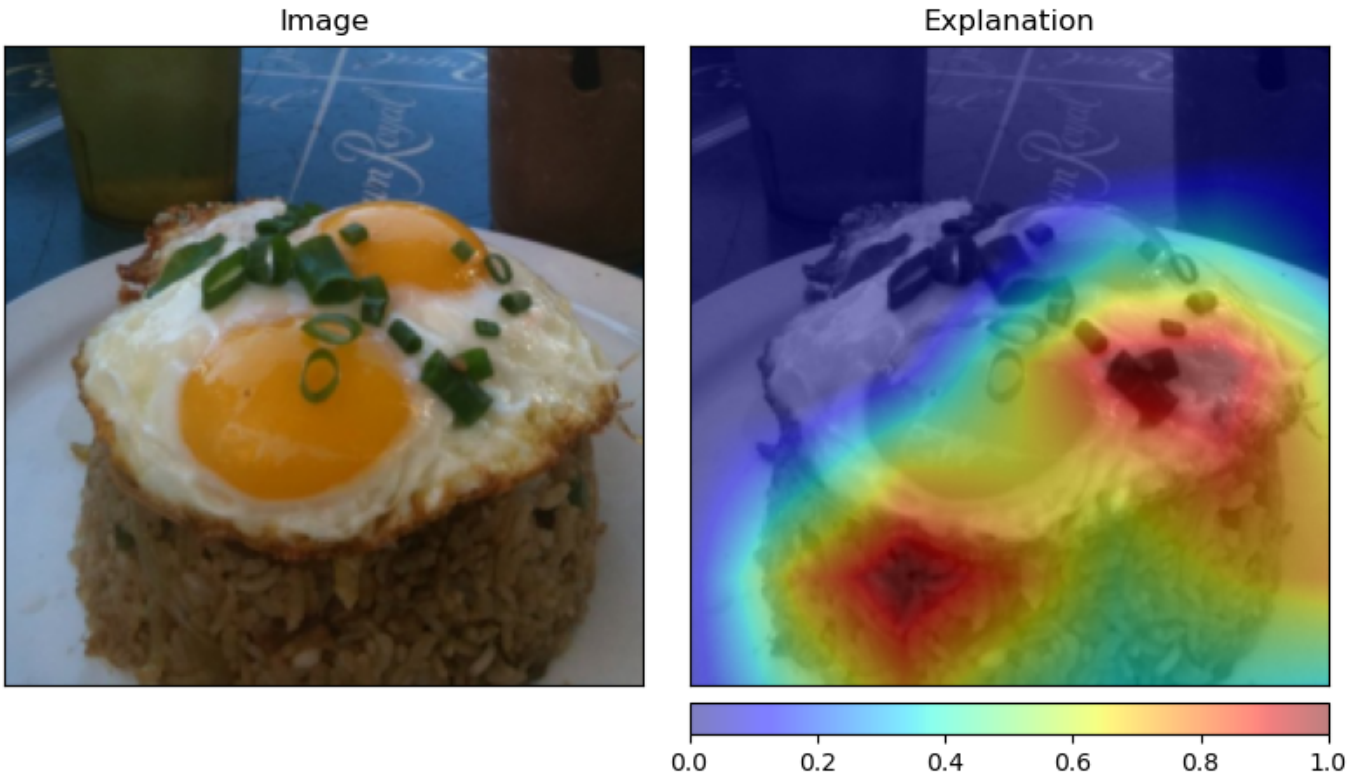
- ○ No, I do not agree.
- ○ Yes, I agree.

Consider the following image:

How would you classify this image?

- ○ croque_madame
- ○ fried_rice
- ○ club_sandwich
- ○ huevos_rancheros
- ○ bibimbap
- ○ carrot_cake
- ○ I don't know

The model classifies the image as **croque_madame**. Below is an explanation of the pixel importance that the machine learning model's decision is based on.



Based on this explanation, do you agree with the classification of the model?

- ○ No, I do not agree.
- ○ Yes, I agree.

If you don't feel familiar enough with the model yet, you can get 2 more examples.

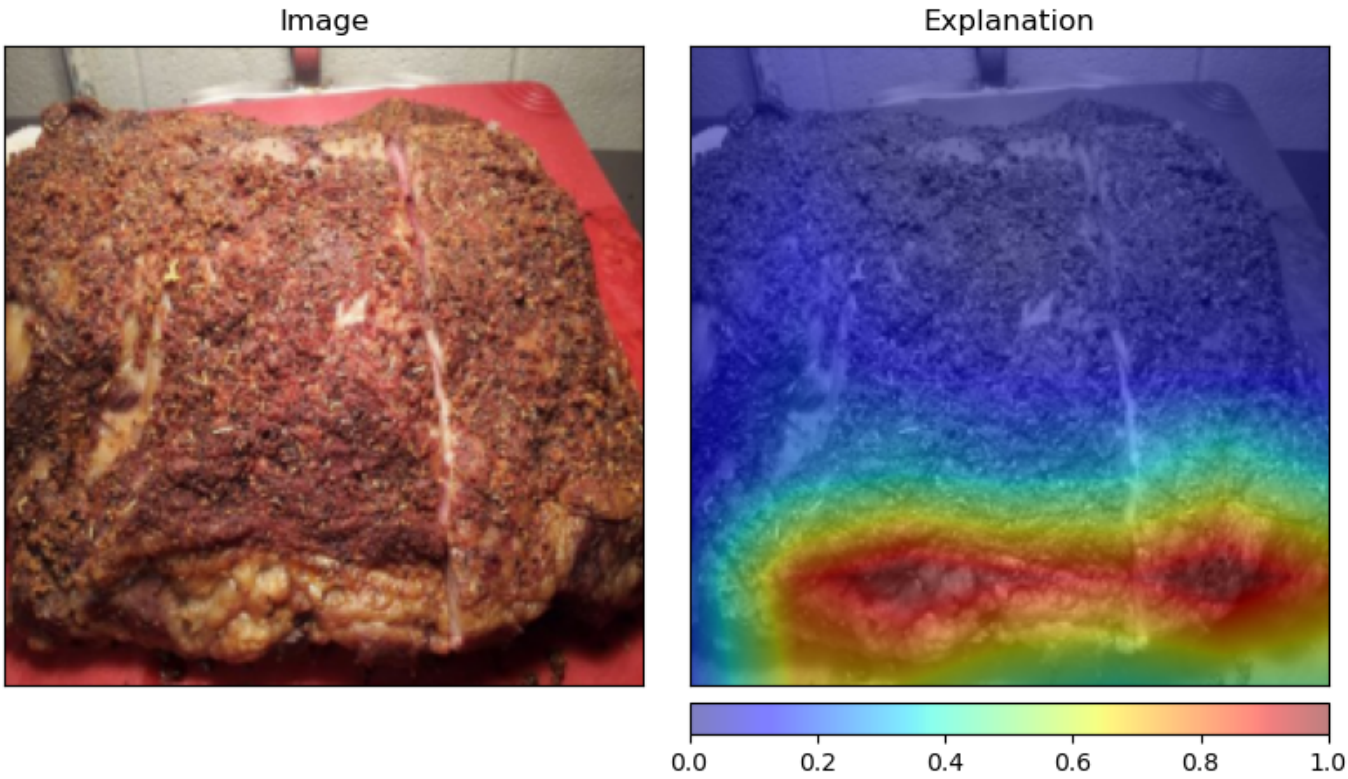
- ○ I need more examples to become familiar with the model.
- ○ I am familiar enough and do not need more examples.

Consider the following image:

How would you classify this image?

- prime_rib
- baby_back_ribs
- lasagna
- spaghetti_bolognese
- steak
- filet_mignon
- I don't know

The model classifies the image as **prime_rib**. Below is an explanation of the pixel importance that the machine learning model's decision is based on.



Based on this explanation, do you agree with the classification of the model?

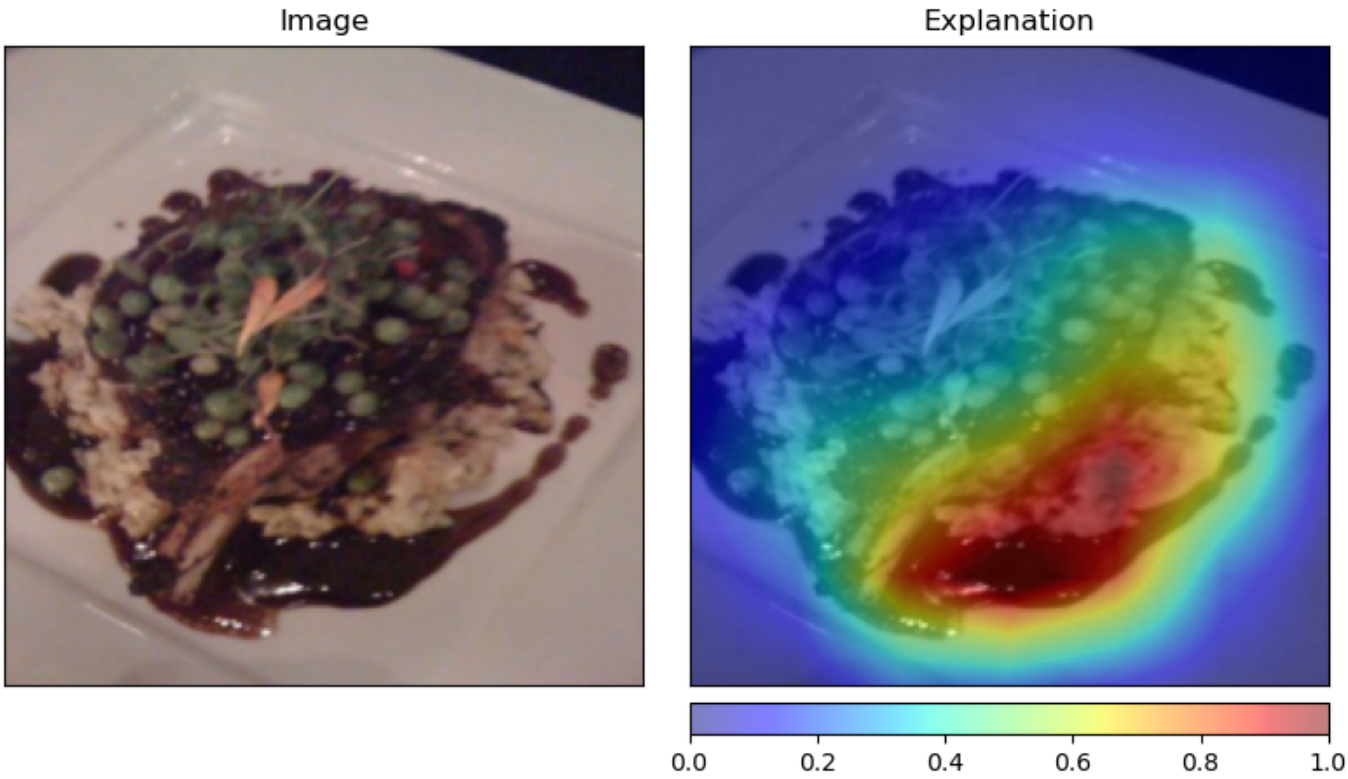
- ○ No, I do not agree.
- ○ Yes, I agree.

Consider the following image:

How would you classify this image?

- filet_mignon
- foie_gras
- pork_chop
- steak
- beet_salad
- escargots
- I don't know

The model classifies the image as **pork_chop**. Below is an explanation of the pixel importance that the machine learning model's decision is based on.



Based on this explanation, do you agree with the classification of the model?

- ○ No, I do not agree.
- ○ Yes, I agree.

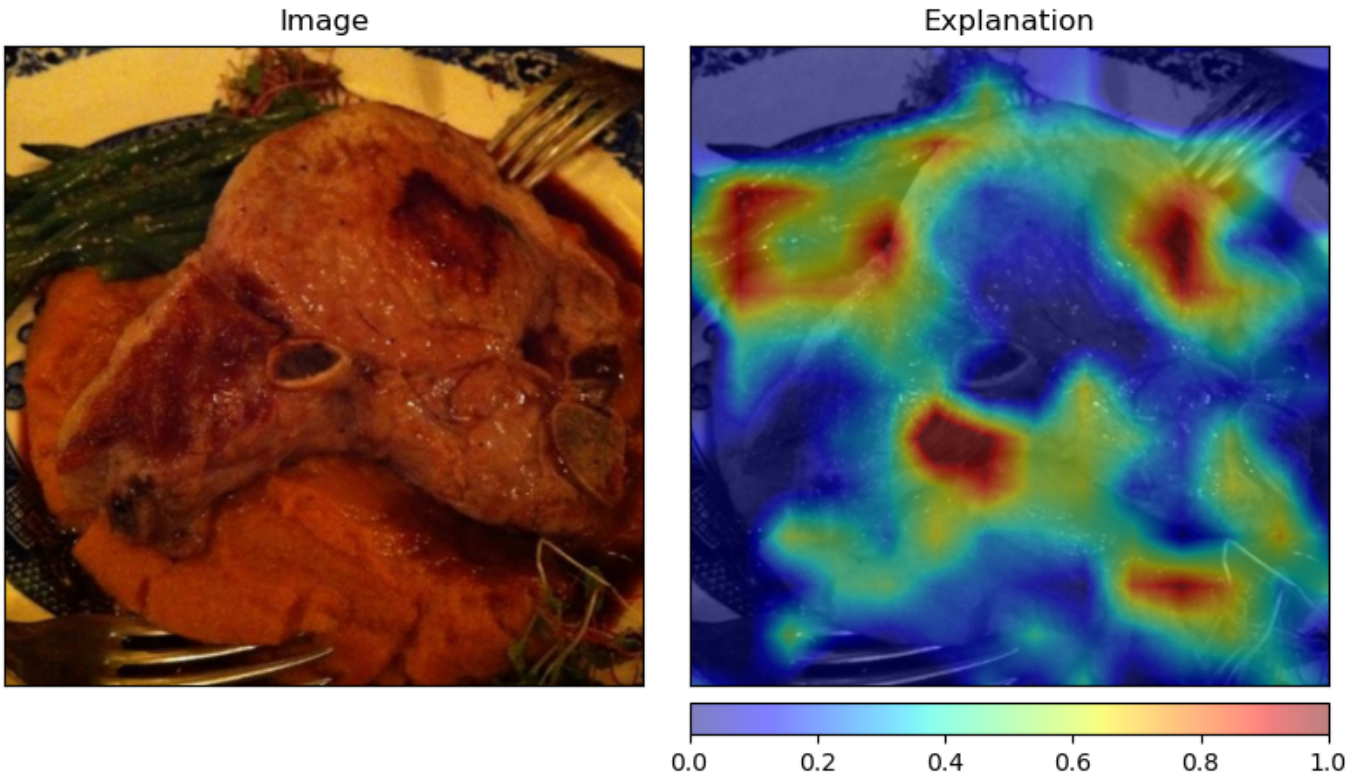
Condition 4 - Non-intuitiveness / Unfaithfulness

Consider the following image:

How would you classify this image?

- foie_gras
- grilled_salmon
- prime_rib
- steak
- peking_duck
- pork_chop
- I don't know

The model classifies the image as **pork_chop**. Below is an explanation of the pixel importance that the machine learning model's decision is based on.



Based on this explanation, do you agree with the classification of the model?

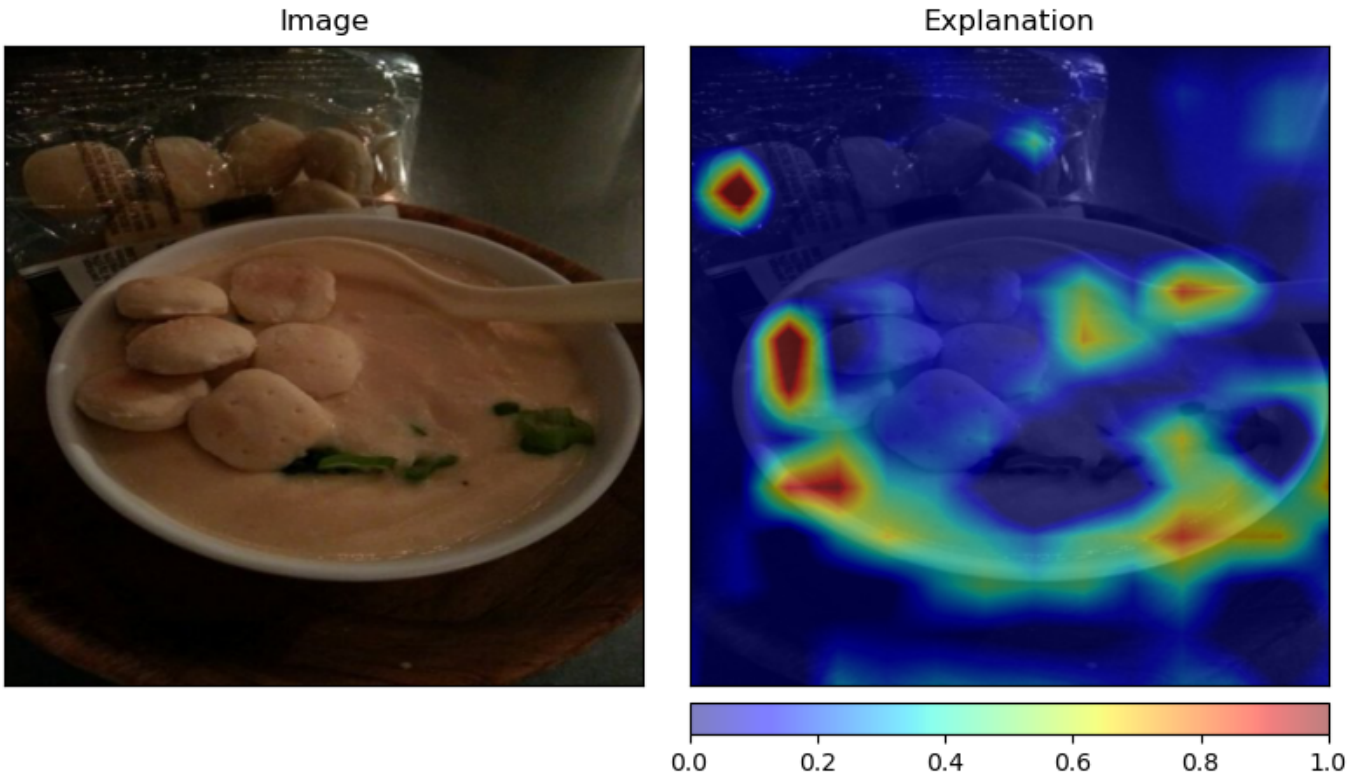
- ○ No, I do not agree.
- ○ Yes, I agree.

Consider the following image:

How would you classify this image?

- clam_chowder
- lobster_bisque
- beignets
- eggs_benedict
- hummus
- french_onion_soup
- I don't know

The model classifies the image as **lobster_bisque**. Below is an explanation of the pixel importance that the machine learning model's decision is based on.



Based on this explanation, do you agree with the classification of the model?

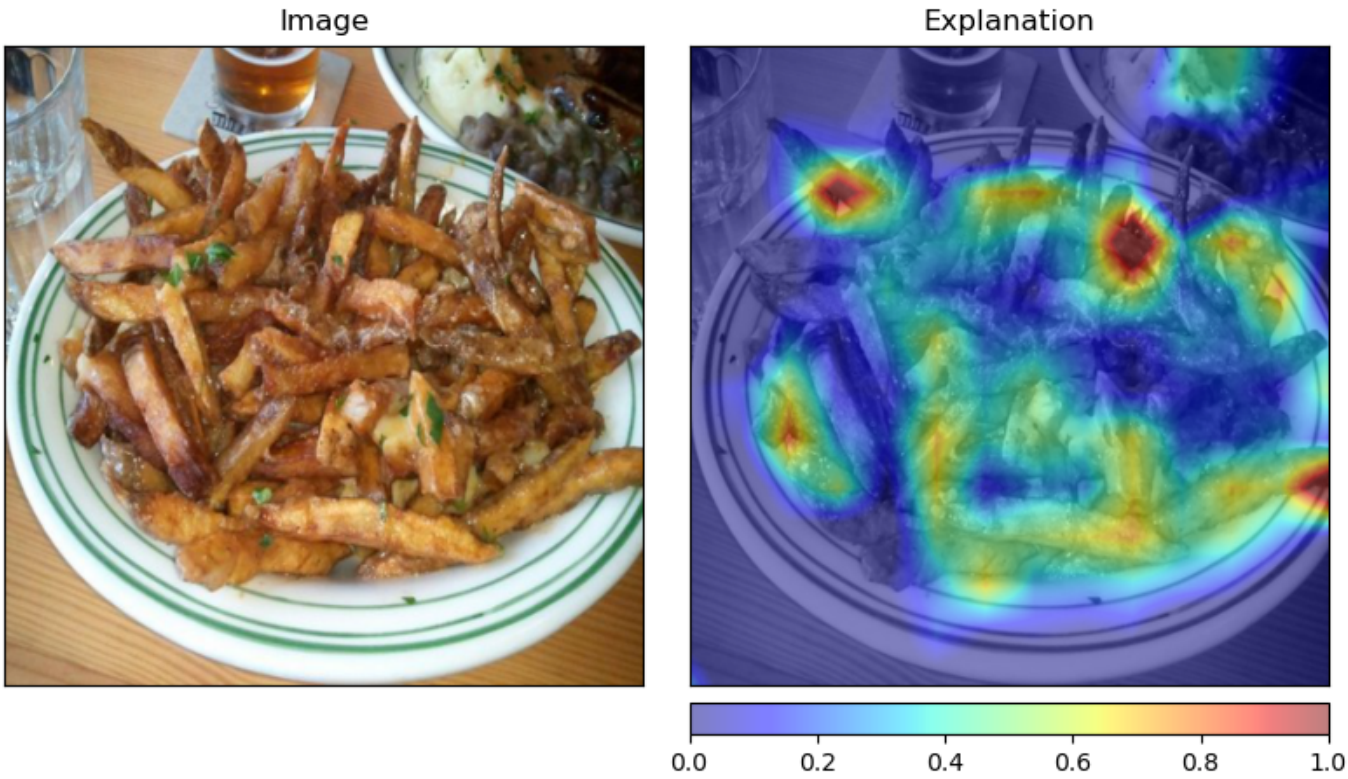
- ○ No, I do not agree.
- ○ Yes, I agree.

Consider the following image:

How would you classify this image?

- ○ french_fries
- ○ pulled_pork_sandwich
- ○ pad_thai
- ○ club_sandwich
- ○ poutine
- ○ baby_back_ribs
- ○ I don't know

The model classifies the image as **poutine**. Below is an explanation of the pixel importance that the machine learning model's decision is based on.



Based on this explanation, do you agree with the classification of the model?

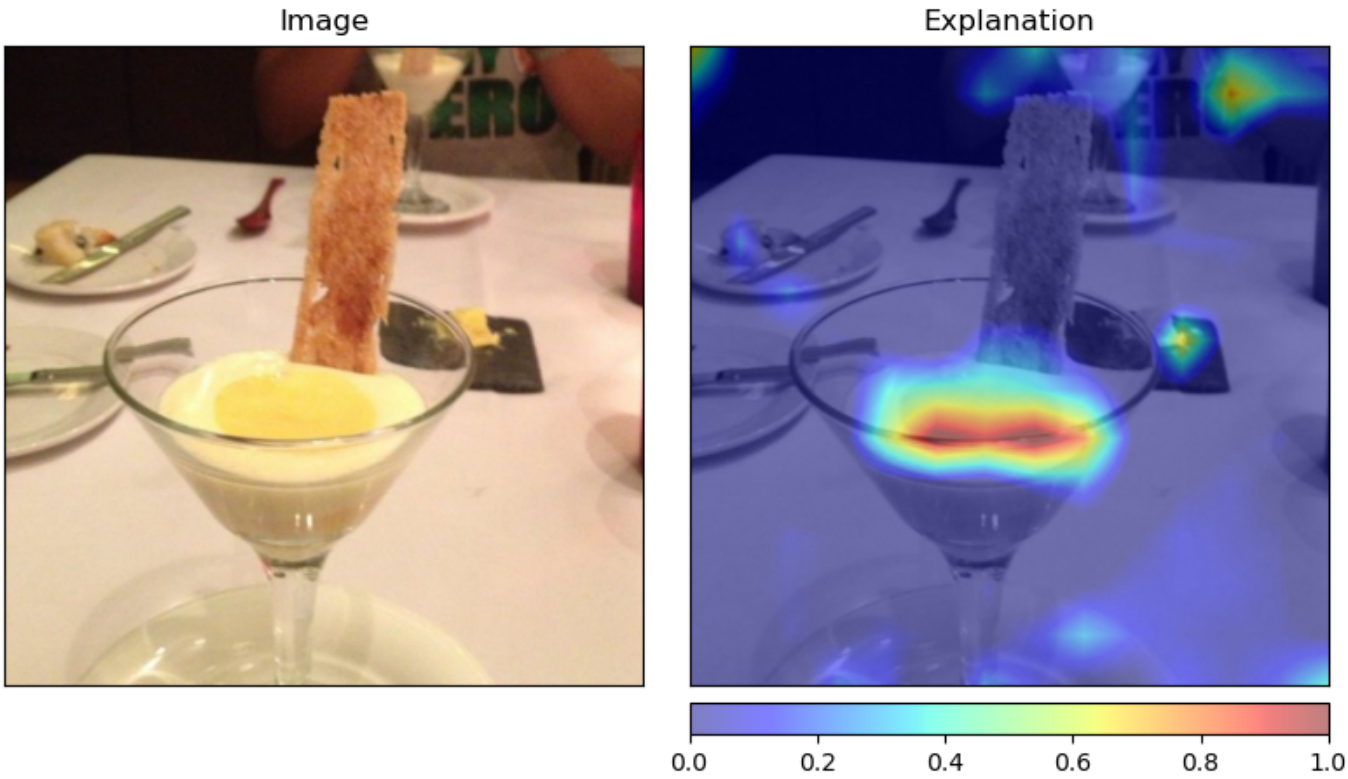
- ○ No, I do not agree.
- ○ Yes, I agree.

Consider the following image:

How would you classify this image?

- ice_cream
- eggs_benedict
- tuna_tartare
- panna_cotta
- chocolate_mousse
- creme_brulee
- I don't know

The model classifies the image as **eggs_benedict**. Below is an explanation of the pixel importance that the machine learning model's decision is based on.



Based on this explanation, do you agree with the classification of the model?

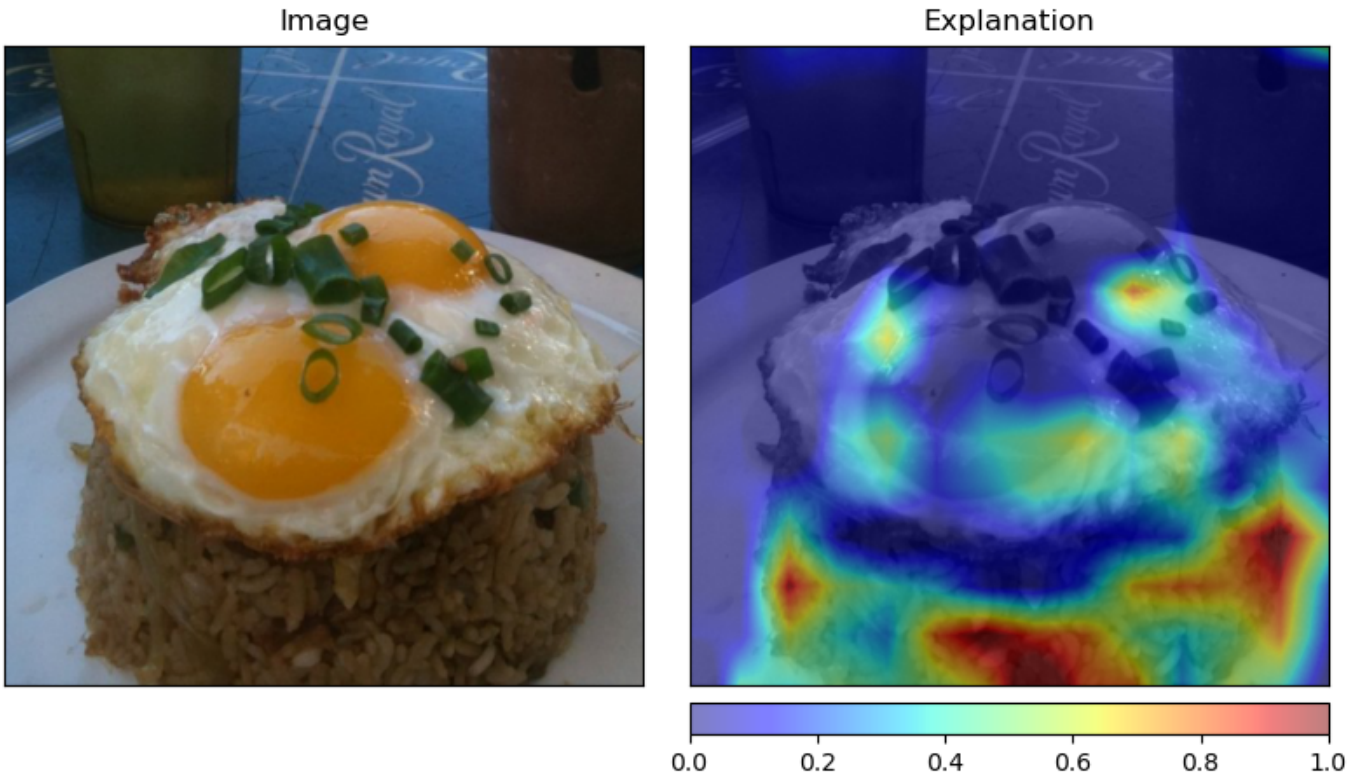
- ○ No, I do not agree.
- ○ Yes, I agree.

Consider the following image:

How would you classify this image?

- croque_madame
- fried_rice
- club_sandwich
- huevos_rancheros
- bibimbap
- carrot_cake
- I don't know

The model classifies the image as **croque_madame**. Below is an explanation of the pixel importance that the machine learning model's decision is based on.



Based on this explanation, do you agree with the classification of the model?

- ○ No, I do not agree.
- ○ Yes, I agree.

If you don't feel familiar enough with the model yet, you can get 2 more examples.

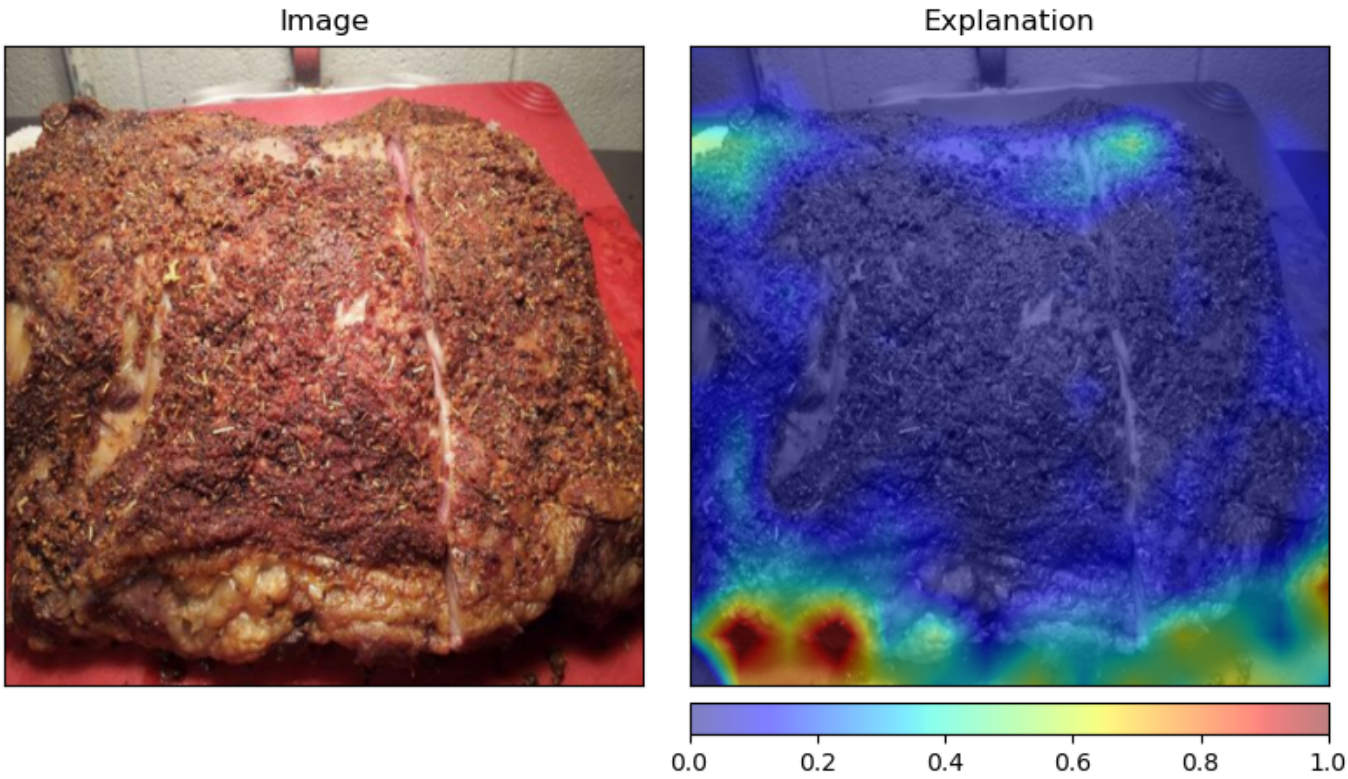
- ○ I need more examples to become familiar with the model.
- ○ I am familiar enough and do not need more examples.

Consider the following image:

How would you classify this image?

- prime_rib
- baby_back_ribs
- lasagna
- spaghetti_bolognese
- steak
- filet_mignon
- I don't know

The model classifies the image as **prime_rib**. Below is an explanation of the pixel importance that the machine learning model's decision is based on.



Based on this explanation, do you agree with the classification of the model?

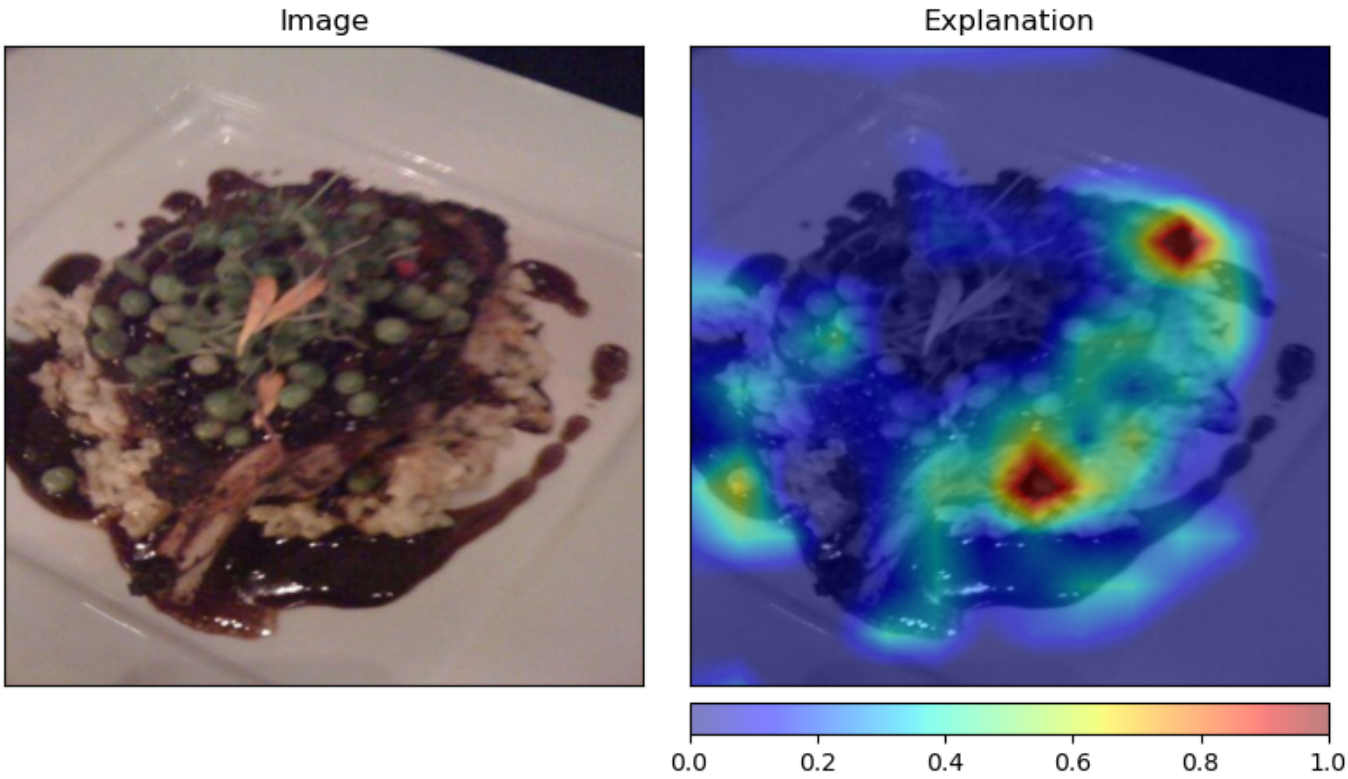
- ○ No, I do not agree.
- ○ Yes, I agree.

Consider the following image:

How would you classify this image?

- filet_mignon
- foie_gras
- pork_chop
- steak
- beet_salad
- escargots
- I don't know

The model classifies the image as **pork_chop**. Below is an explanation of the pixel importance that the machine learning model's decision is based on.



Based on this explanation, do you agree with the classification of the model?

- ○ No, I do not agree.
- ○ Yes, I agree.

Trust Assessment

Based on the previous predictions and the corresponding explanations of the model, please indicate how much you agree with the following statements.

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
I can predict how the model will behave.	○	○	○	○	○	○	○
I trust the decisions made by the model.	○	○	○	○	○	○	○

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
I have faith that the model would be able to cope with all different kinds of food.	○	○	○	○	○	○	○
If I am not sure about a decision, I trust that the model will provide the best solution.	○	○	○	○	○	○	○
I trust the model to provide a reliable decision for classifying different images of food.	○	○	○	○	○	○	○

Overall, why do you trust (or distrust) the decisions made by the model?

Instruction 2

The survey regarding the first scenario has been completed.
Please consider the next scenario: There is a new machine learning model that can predict **if a criminal defendant is likely or not likely to reoffend** based on certain inputs. There are methods to explain the model's decision by showing how each input affected the prediction. The inputs (different attributes of a criminal defendant) are as follows:

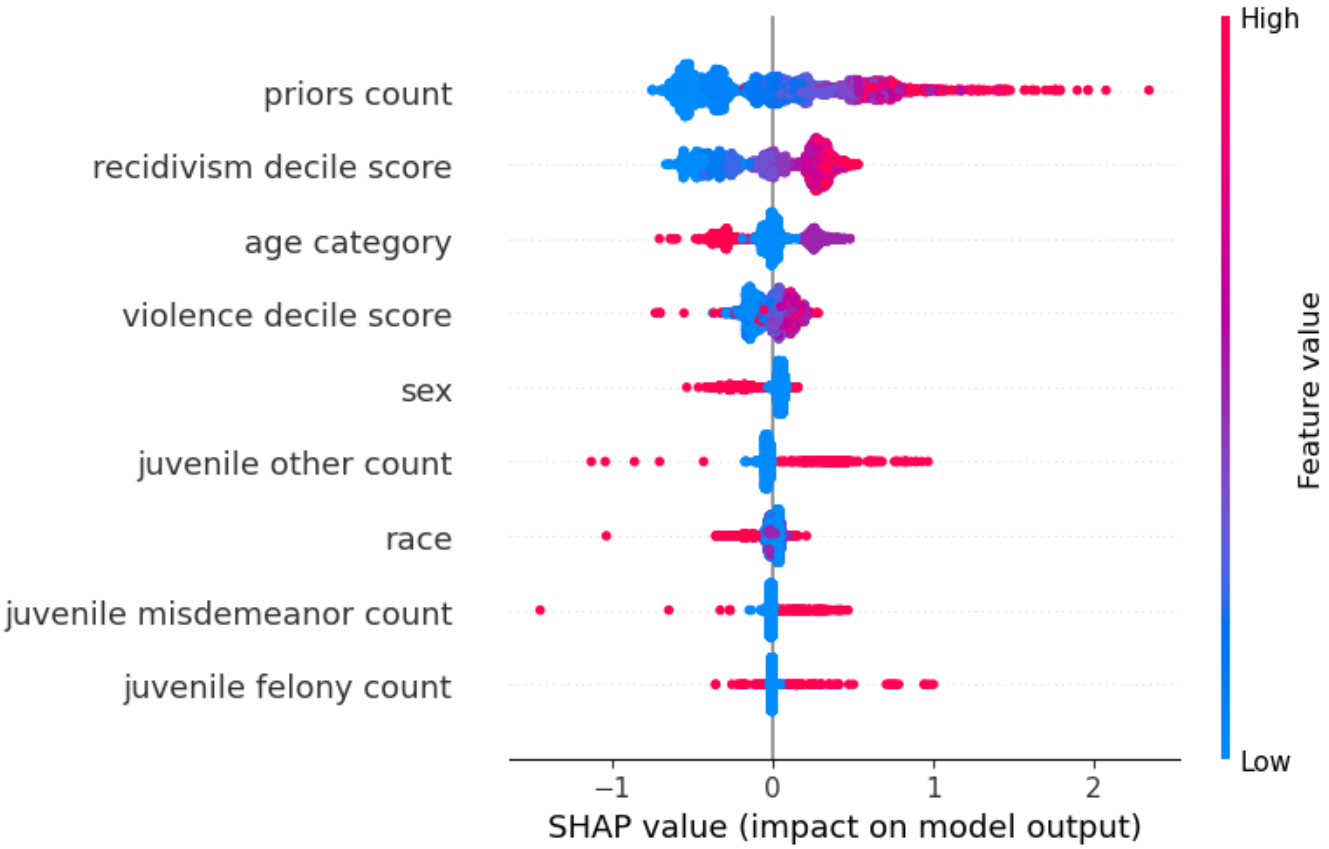
Attributes	Range	Description
sex	0, 1	0: male, 1: female
age category	0 - 2	the age the criminal defendant is: 0: 25 - 45, 1: Less than 25, 2: Greater than 45
race	0 - 4	the race the criminal defendant is: 0: African-American, 1: Caucasian, 2: Hispanic,

Attributes	Range	Description
		3: Asian, 4: Native American
priors count	0 - open end	number of previous criminal records (average count = 4.19)
recidivism decile score	1 - 10	risk of recidivism on a scale from 1-10, where 1-4 is "Low", 5-7 is "Medium and 8-10 is "High"
violence decile score	1 - 10	risk of violence on a scale from 1-10, where 1-4 is "Low", 5-7 is "Medium and 8-10 is "High"
juvenile felony count	0 - open end	number of felonies as a juvenile (average count = 0.08)
juvenile misdemeanor count	0 - open end	number of misdemeanors as a juvenile (average count = 0.1)
juvenile other count	0 - open end	number of other crimes as a juvenile (average count = 0.13)

Although it is difficult to understand how the model works, there are methods to help explain its decisions by showing which attribute (indicated on the left in each row) has a higher impact on the prediction.
Below are two explanations generated by the machine learning model to explain its decision. The first explanation is an explanation of **the entire model** and the second explanation is an explanation for **a specific model decision**.

General Model Explanation

The color represents the attribute input, with **blue for low** (e.g., a lower value of priors count can be 0) and **red for high** (e.g., a higher value of priors count can be 20).
The position on the x-axis shows if the defendant is **less likely (left)** or **more likely (right)** to reoffend.



Explanation of an Individual Decision

Here, the color represents the impact of the attributes on a prediction. The values can be understood in the following way:

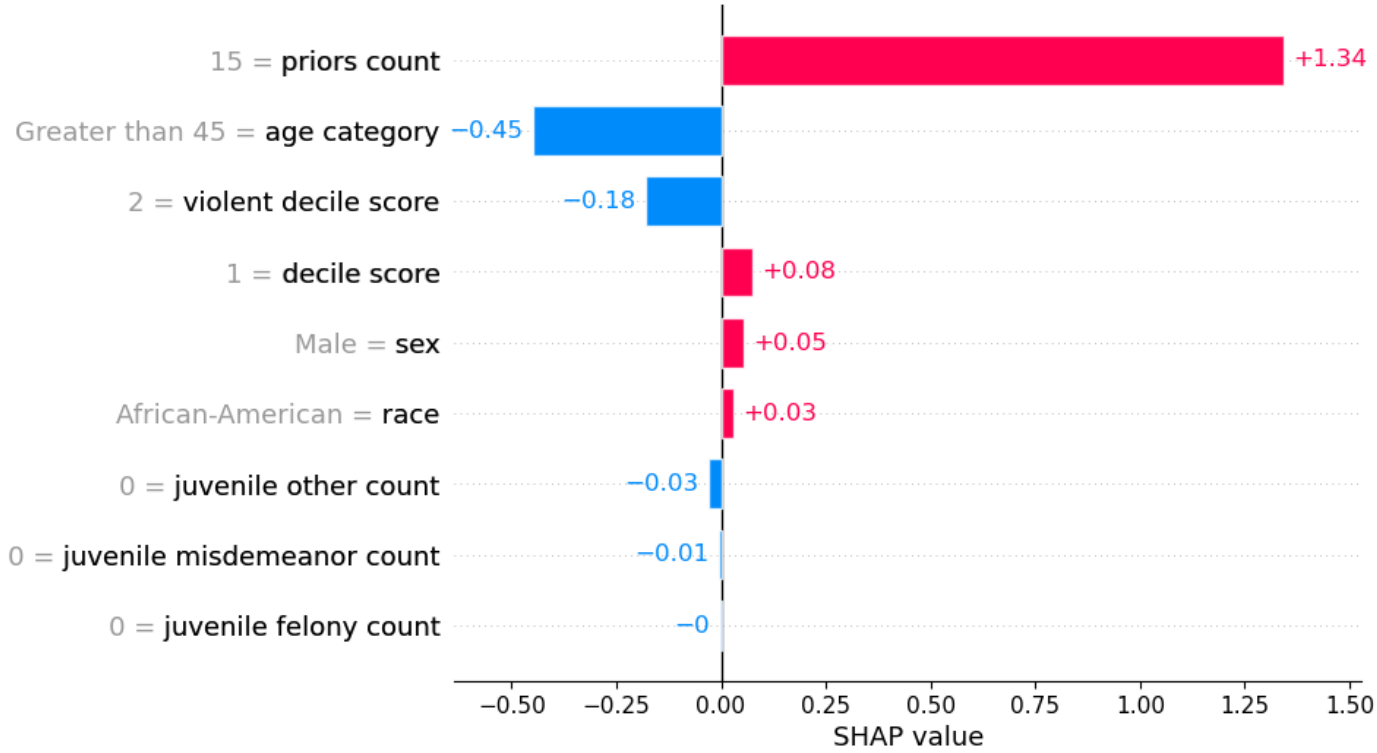
Values **smaller** than 0 (**negative** values in blue) -> **less likely to reoffend**

Values **larger** than 0 (**positive** values in red) -> **more likely to reoffend**

For both positive and negative values, larger bars correspond to larger contributions to the prediction.

In the example below, the attribute **"priors count"** has the highest contribution to the prediction that the defendant **is likely** to reoffend and the attribute **"age category"** has

the highest contribution to the prediction that the defendant **is not likely** to reoffend.



In the following, there are 5 - 7 questions related to the given scenario. These questions ask you to categorize a criminal defendant by assessing their profile and will help you familiarize yourself with the model.
But first, to ensure you have understood the information above correctly, please answer the following 3 questions about the instruction.

Which of the following statements is **true** about the mentioned model?

- ○ The model is a protocol to follow in order to decide if the criminal defendant is likely to reoffend.
- ○ The model is a computer program that automatically evaluates if a criminal defendant is likely to reoffend.
- ○ The model is a computer program that randomly generates a number.

Which of the following statements is **true**?

- ○ The attributes in this study can be ignored.
- ○ The attributes will not influence the automated decisions for criminal recidivism.
- ○ The attributes will influence the automated decisions for criminal recidivism.

Which of the following statements is **true**?

- ○ Positive values (in red) have a higher impact to the model's decision that the defendant is likely to reoffend.

- Negative values (in blue) have a higher impact to the model's decision that the defendant is likely to reoffend.
- Positive (in red) and negative values (in blue) have the same impact to the model's decision that the defendant is likely to reoffend.

Condition 1 - Intuitiveness / Faithfulness / Desired Features

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Female
age category	25 - 45
race	African-American
priors count	5
recidivism decile score	10
violence decile score	5
juvenile felony count	1
juvenile misdemeanor count	0
juvenile other count	0

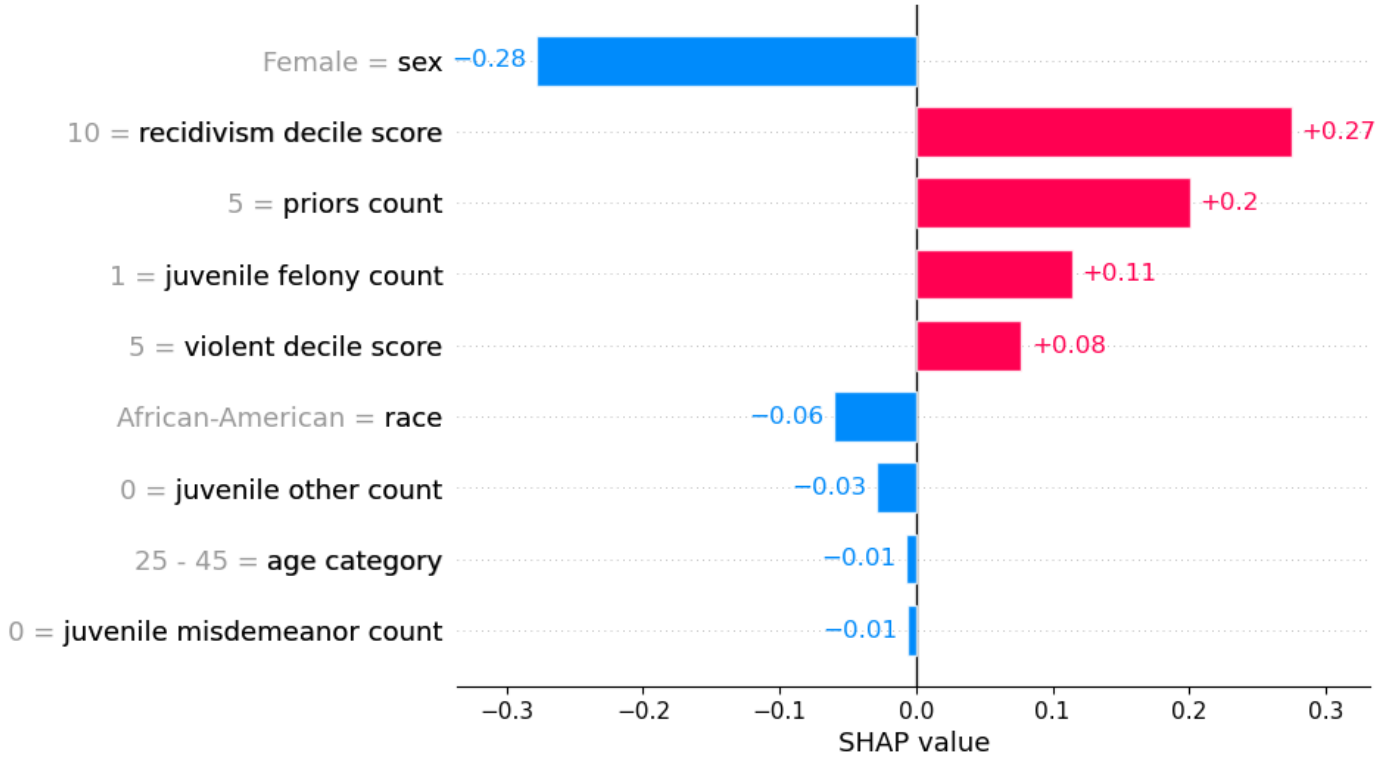
Click "yes" to see the descriptions of the attributes again.

- Yes
- No

How would you categorize this criminal defendant?

- I think that the criminal defendant is not likely to reoffend.
- I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- ○ No, I do not agree.
- ○ Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	Less than 25
race	Hispanic
priors count	11
recidivism decile score	10
violence decile score	8
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	0

Click "yes" to see the descriptions of the attributes again.

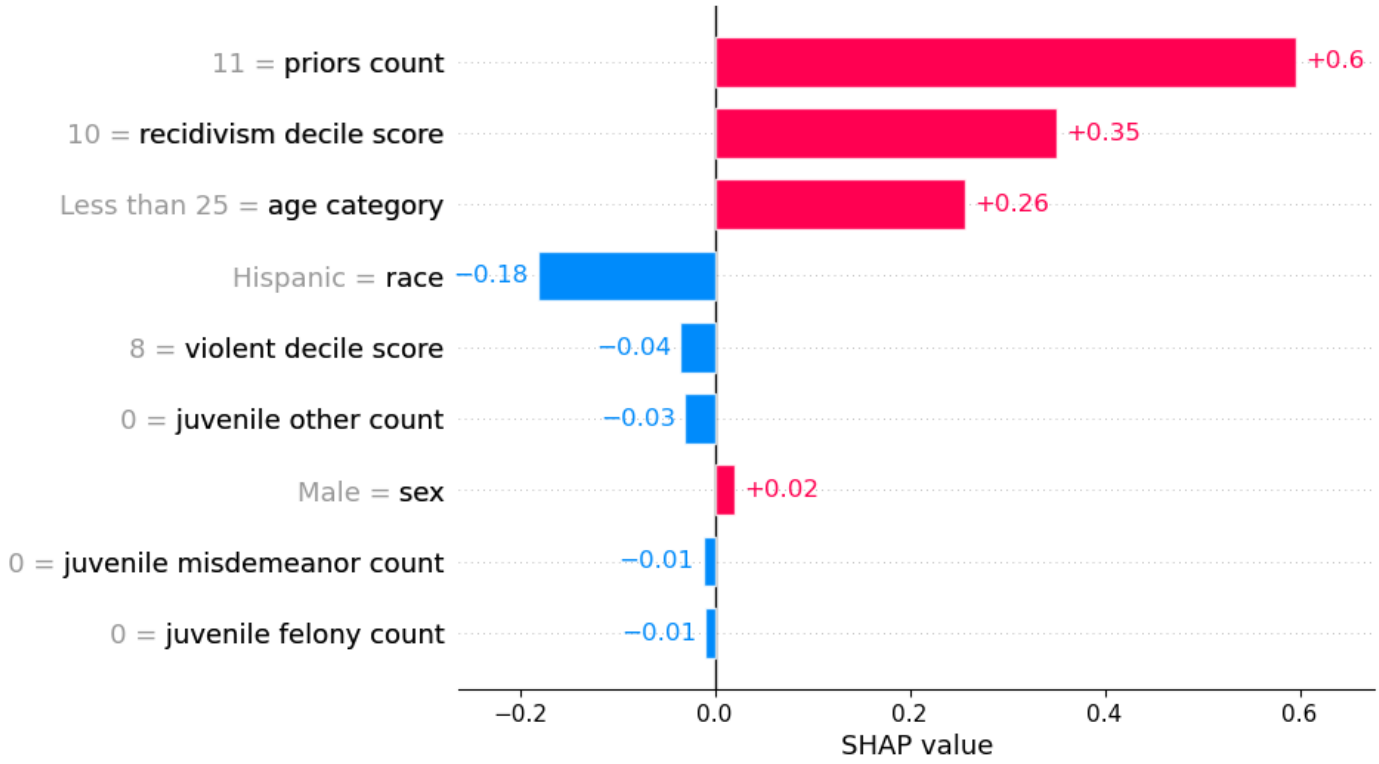
- ○ Yes

- ○ No

How would you categorize this criminal defendant?

- ○ I think that the criminal defendant is not likely to reoffend.
- ○ I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- ○ No, I do not agree.
- ○ Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	25 - 45
race	Caucasian
priors count	15
recidivism decile score	9

Attributes	Profile
violence decile score	6
juvenile felony count	2
juvenile misdemeanor count	0
juvenile other count	1

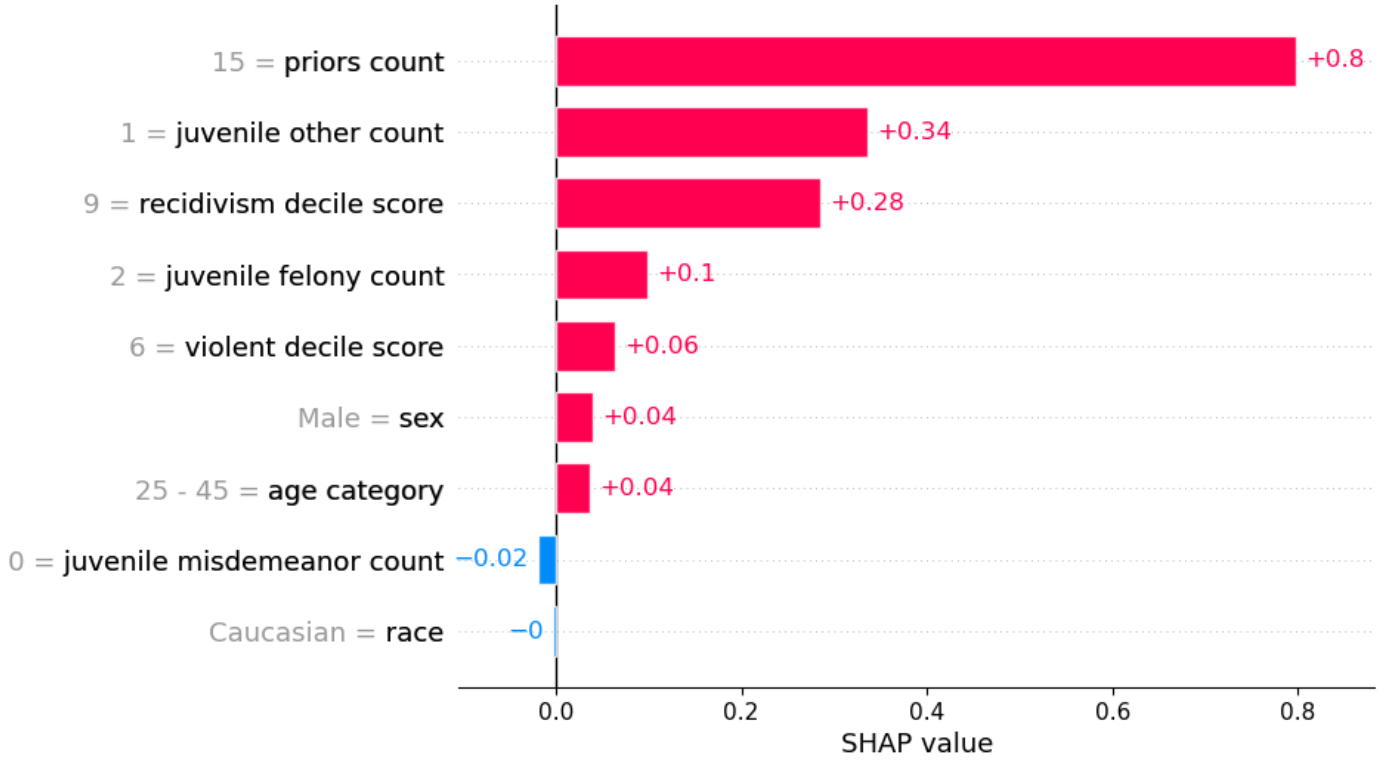
Click "yes" to see the descriptions of the attributes again.

- Yes
- No

How would you categorize this criminal defendant?

- I think that the criminal defendant is not likely to reoffend.
- I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- No, I do not agree.
- Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	25 - 45
race	African-American
priors count	2
recidivism decile score	1
violence decile score	1
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	0

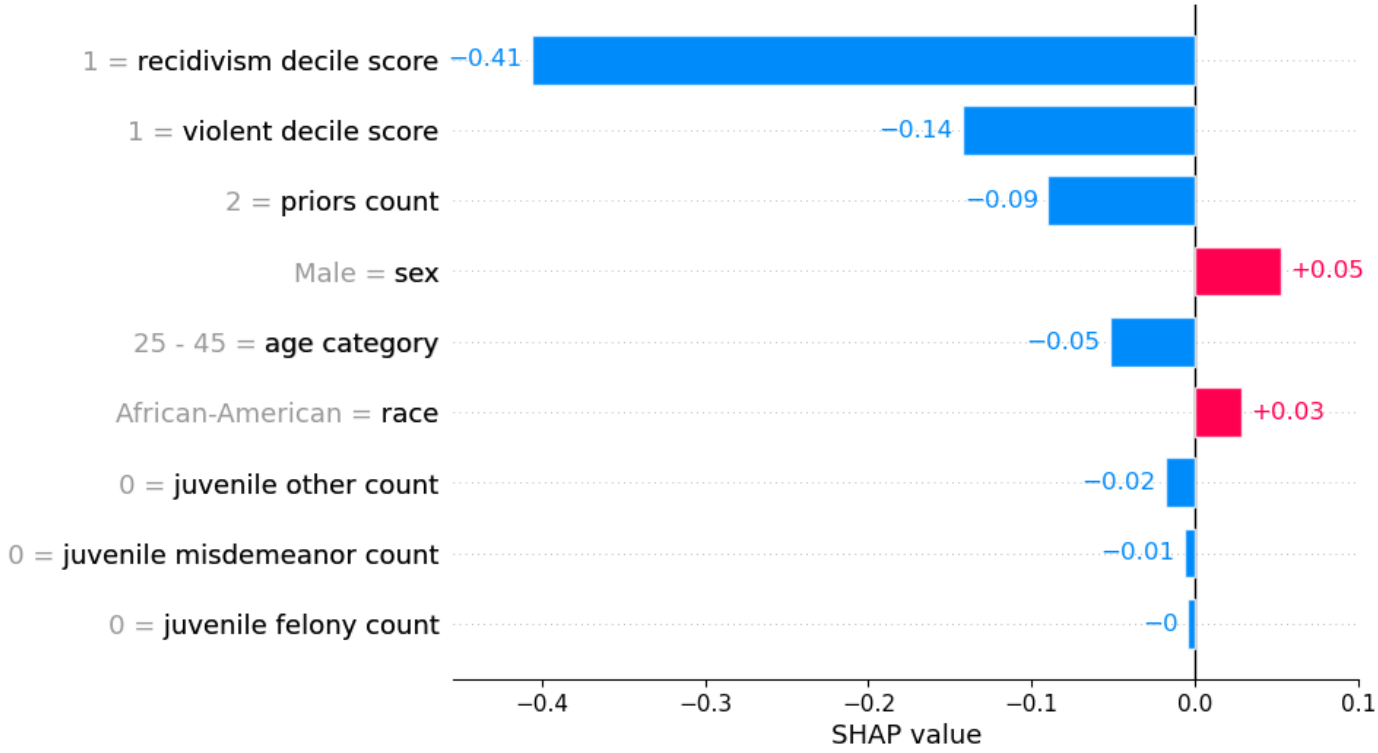
Click "yes" to see the descriptions of the attributes again.

- ○ Yes
- ○ No

How would you categorize this criminal defendant?

- ○ I think that the criminal defendant is not likely to reoffend.
- ○ I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is not likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- ○ No, I do not agree.
- ○ Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	25 - 45
race	African-American
priors count	1
recidivism decile score	3
violence decile score	2
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	0

Click "yes" to see the descriptions of the attributes again.

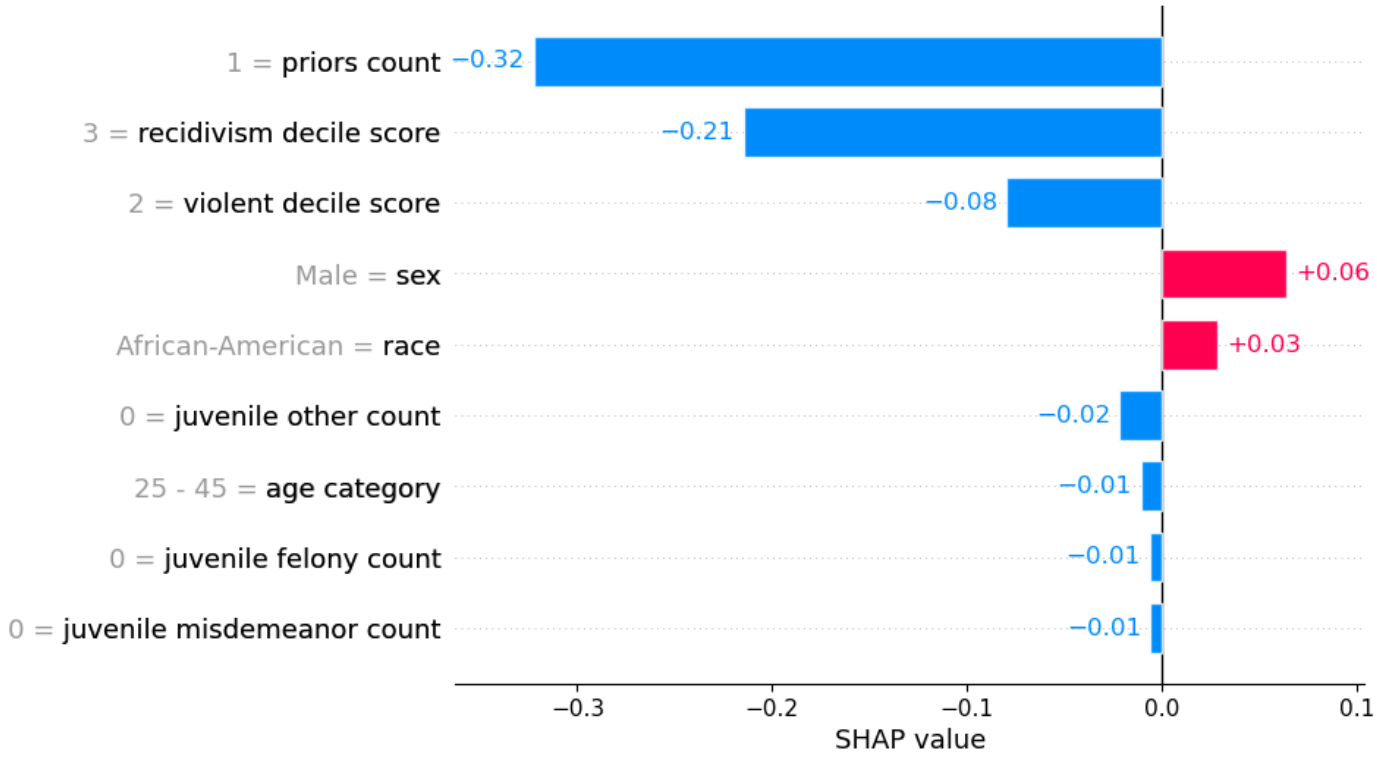
- ○ Yes

- ○ No

How would you categorize this criminal defendant?

- ○ I think that the criminal defendant is not likely to reoffend.
- ○ I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is not likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- ○ No, I do not agree.
- ○ Yes, I agree.

If you don't feel familiar enough with the model yet, you can get 2 more examples.

- ○ I need more examples to become familiar with the model.
- ○ I am familiar enough and do not need more examples.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	25 - 45
race	Hispanic
priors count	9
recidivism decile score	9
violence decile score	7
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	0

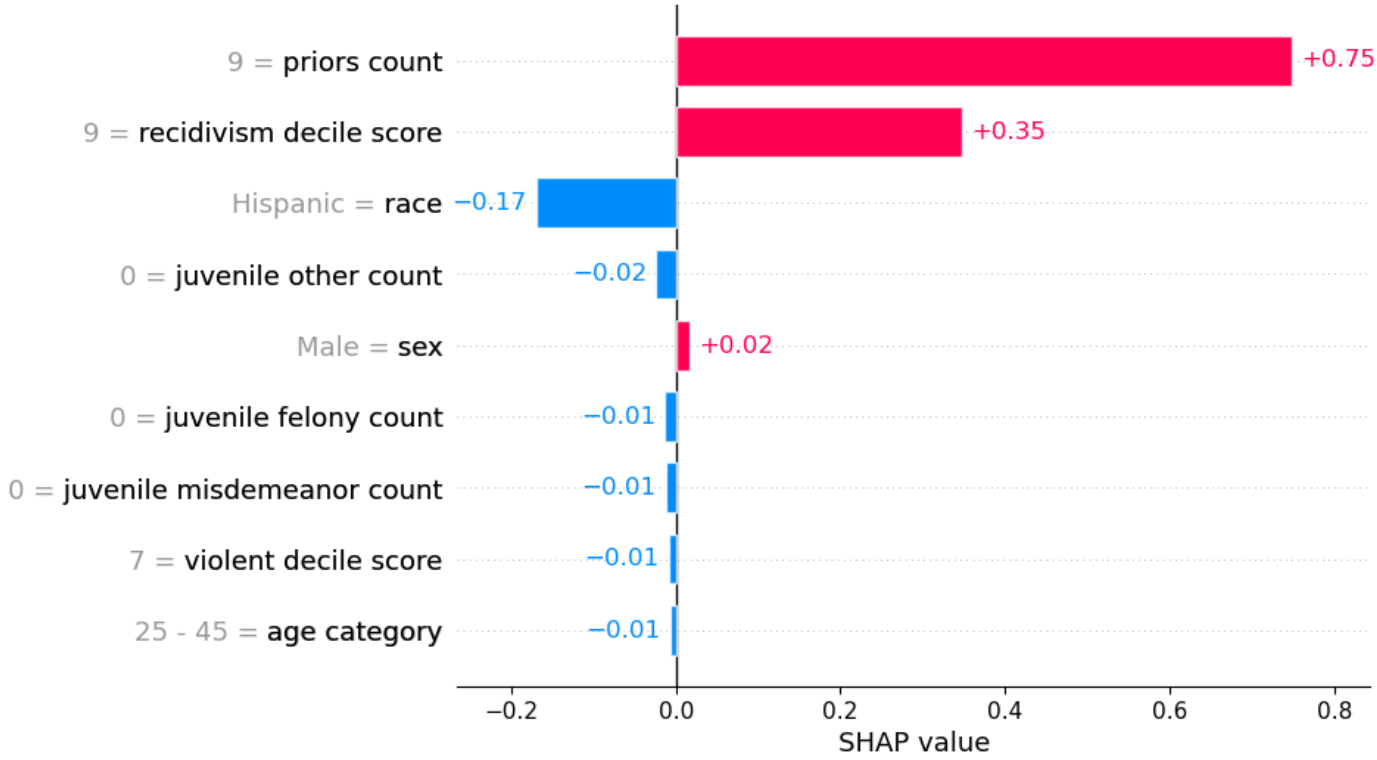
Click "yes" to see the descriptions of the attributes again.

- ○ Yes
- ○ No

How would you categorize this criminal defendant?

- ○ I think that the criminal defendant is not likely to reoffend.
- ○ I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- ○ No, I do not agree.
- ○ Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	25 - 45
race	Caucasian
priors count	5
recidivism decile score	10
violence decile score	5
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	0

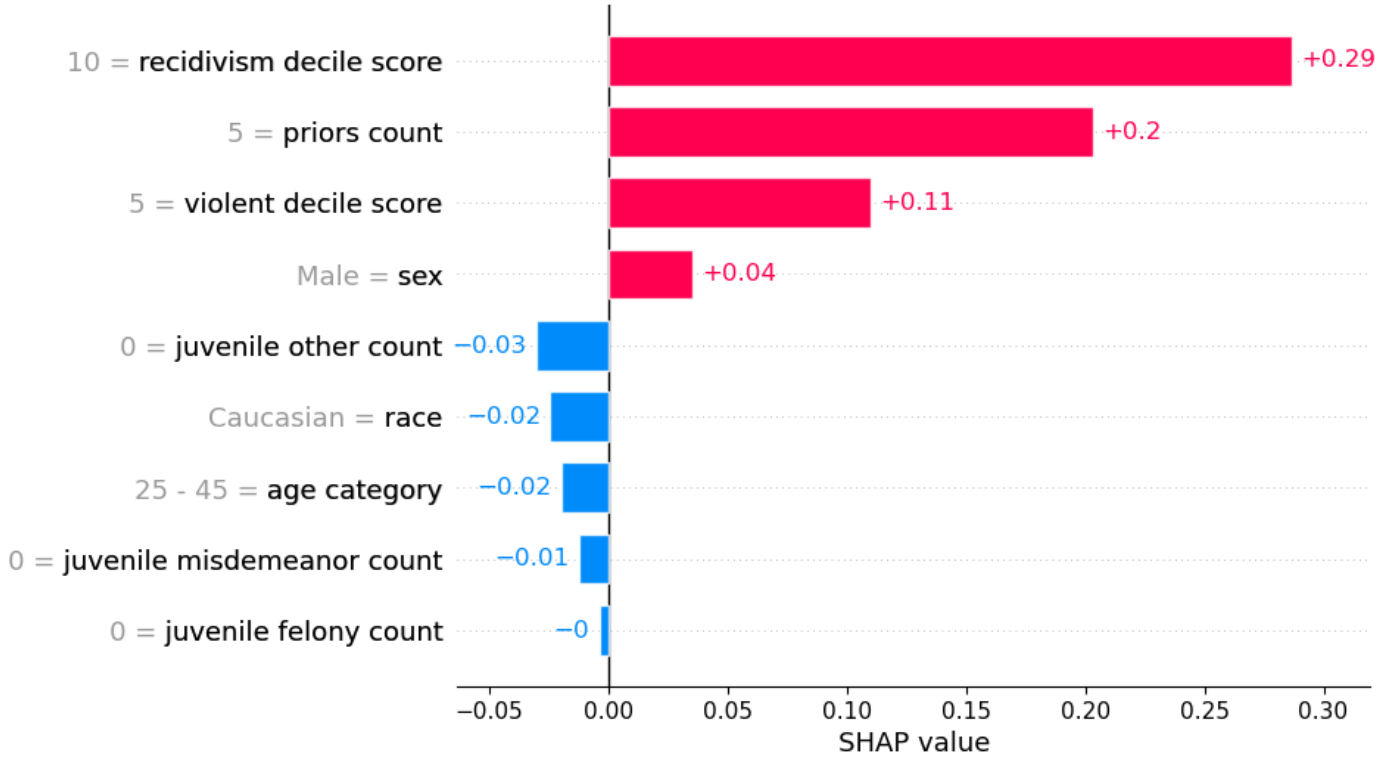
Click "yes" to see the descriptions of the attributes again.

- ○ Yes
- ○ No

How would you categorize this criminal defendant?

- I think that the criminal defendant is not likely to reoffend.
- I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- No, I do not agree.
- Yes, I agree.

Condition 2 - Intuitiveness / Unfaithfulness / Desired Features

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Female
age category	25 - 45
race	African-American

Attributes	Profile
priors count	5
recidivism decile score	10
violence decile score	5
juvenile felony count	1
juvenile misdemeanor count	0
juvenile other count	0

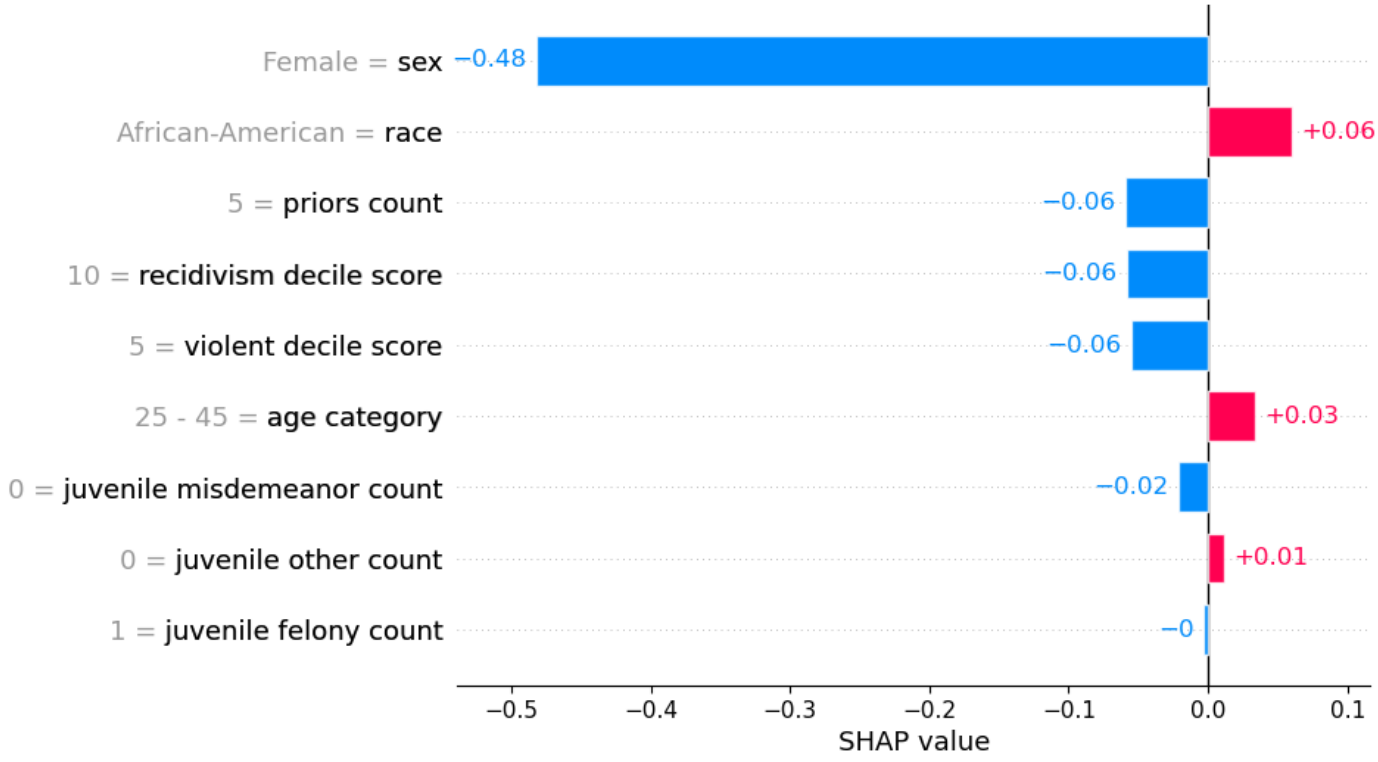
Click "yes" to see the descriptions of the attributes again.

- Yes
- No

How would you categorize this criminal defendant?

- I think that the criminal defendant is not likely to reoffend.
- I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- No, I do not agree.

- ○ Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	Less than 25
race	Hispanic
priors count	11
recidivism decile score	10
violence decile score	8
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	0

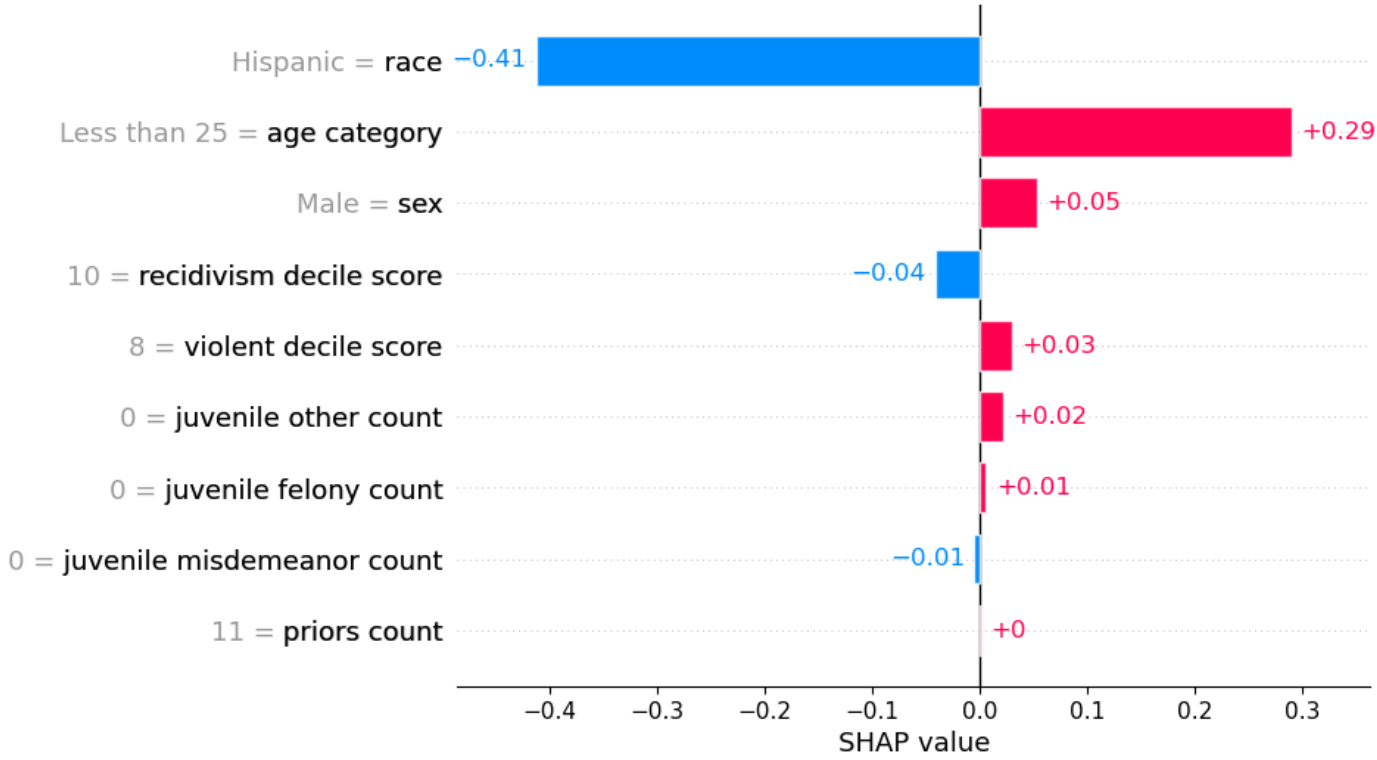
Click "yes" to see the descriptions of the attributes again.

- ○ Yes
- ○ No

How would you categorize this criminal defendant?

- ○ I think that the criminal defendant is not likely to reoffend.
- ○ I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- ○ No, I do not agree.
- ○ Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	25 - 45
race	Caucasian
priors count	15
recidivism decile score	9
violence decile score	6
juvenile felony count	2
juvenile misdemeanor count	0
juvenile other count	1

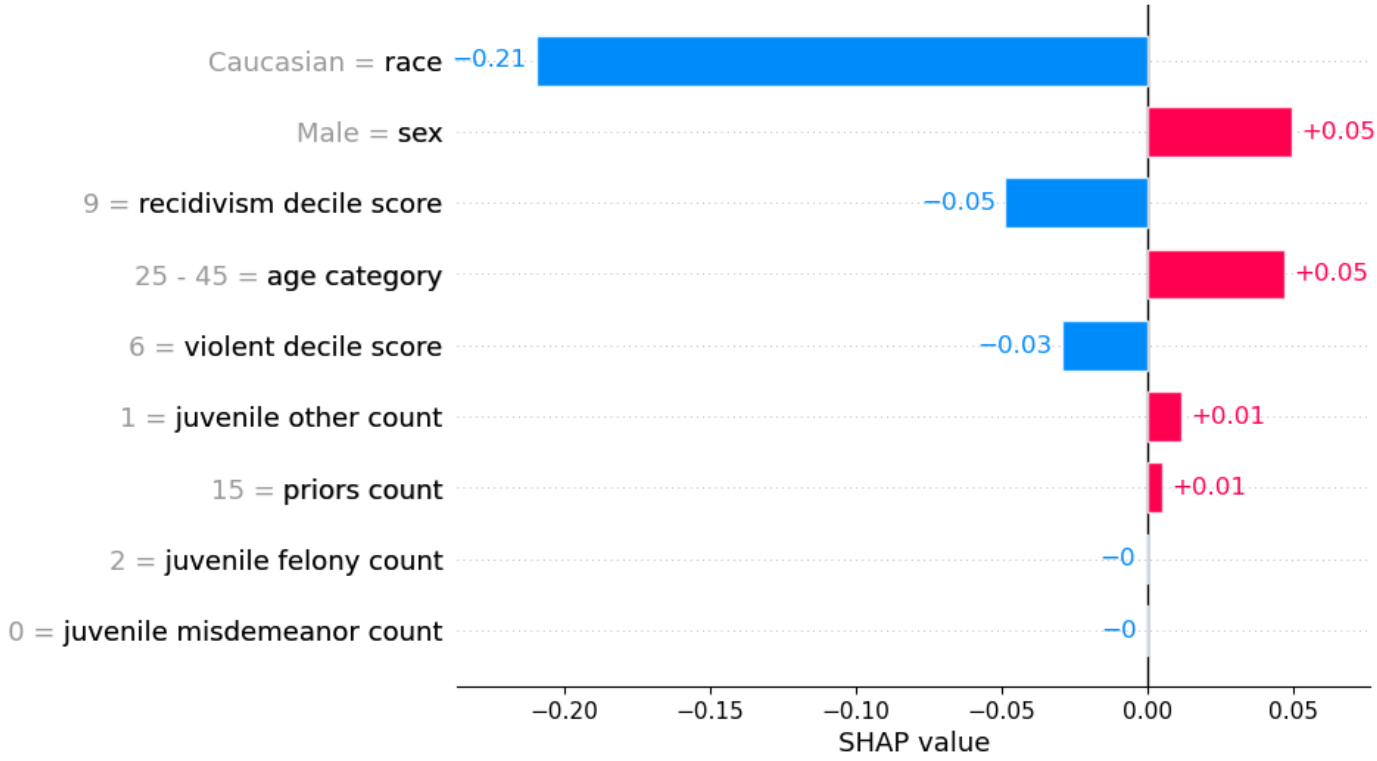
Click "yes" to see the descriptions of the attributes again.

- ○ Yes
- ○ No

How would you categorize this criminal defendant?

- I think that the criminal defendant is not likely to reoffend.
- I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- No, I do not agree.
- Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	25 - 45
race	African-American
priors count	2
recidivism decile score	1

Attributes	Profile
violence decile score	1
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	0

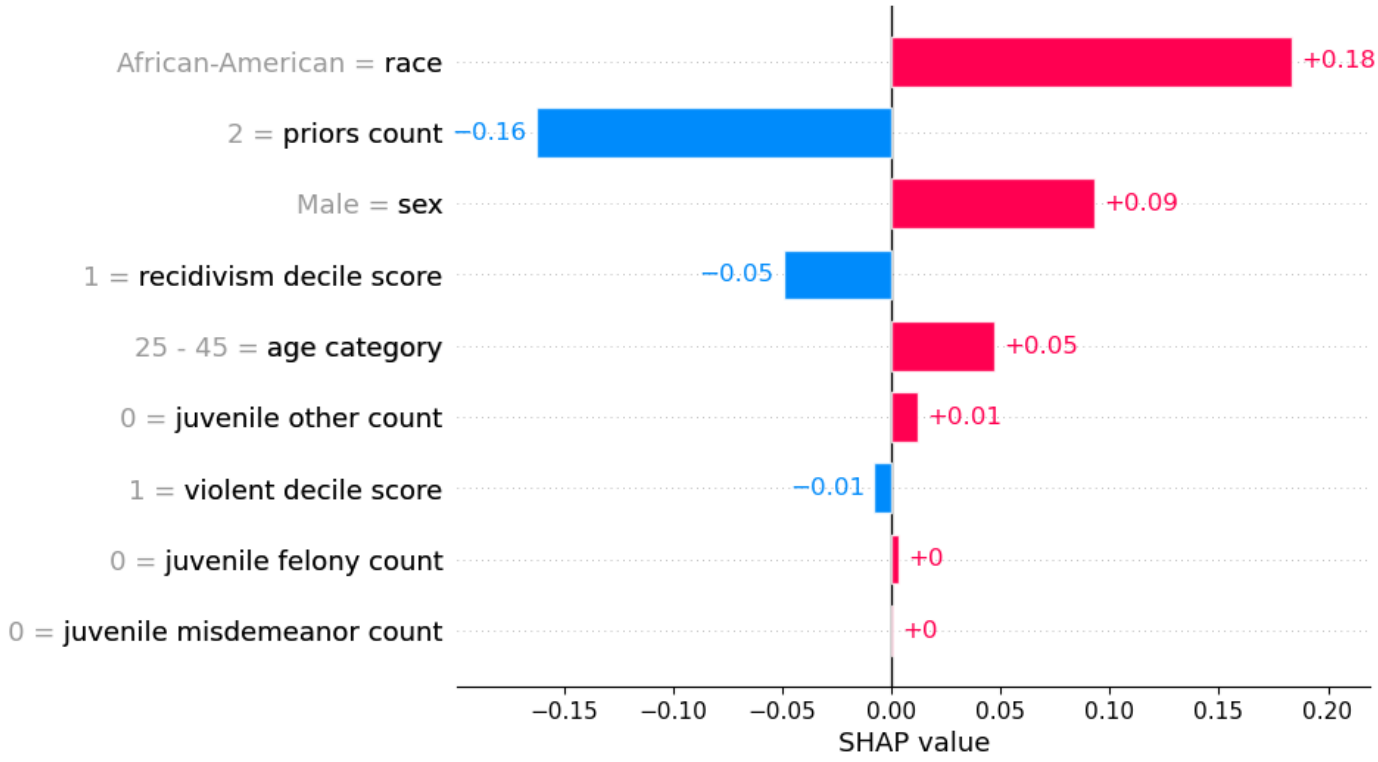
Click "yes" to see the descriptions of the attributes again.

- Yes
- No

How would you categorize this criminal defendant?

- I think that the criminal defendant is not likely to reoffend.
- I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is not likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- No, I do not agree.
- Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	25 - 45
race	African-American
priors count	1
recidivism decile score	3
violence decile score	2
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	0

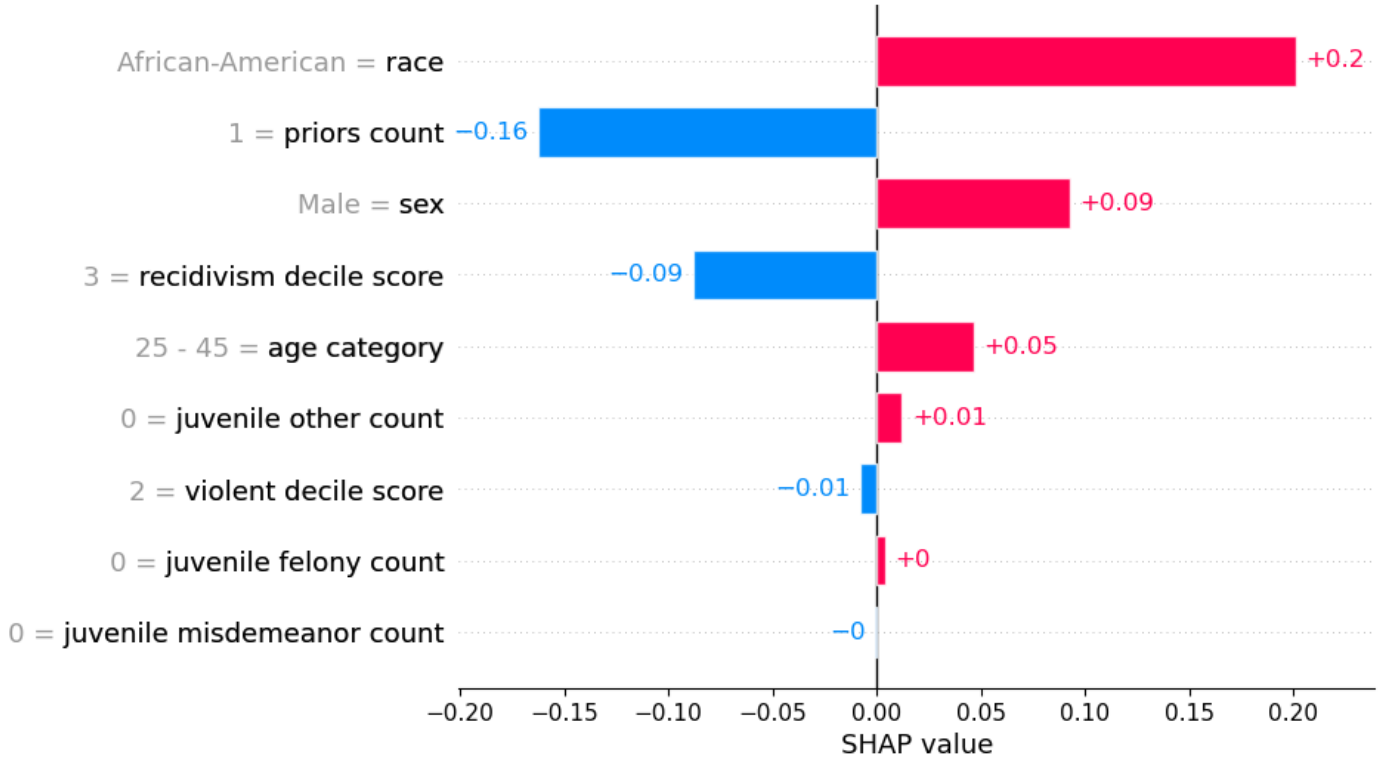
Click "yes" to see the descriptions of the attributes again.

- ○ Yes
- ○ No

How would you categorize this criminal defendant?

- ○ I think that the criminal defendant is not likely to reoffend.
- ○ I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is not likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- ○ No, I do not agree.
- ○ Yes, I agree.

If you don't feel familiar enough with the model yet, you can get 2 more examples.

- ○ I need more examples to become familiar with the model.
- ○ I am familiar enough and do not need more examples.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	25 - 45
race	Hispanic
priors count	9
recidivism decile score	9
violence decile score	7
juvenile felony count	0
juvenile misdemeanor count	0

Attributes	Profile
juvenile other count	0

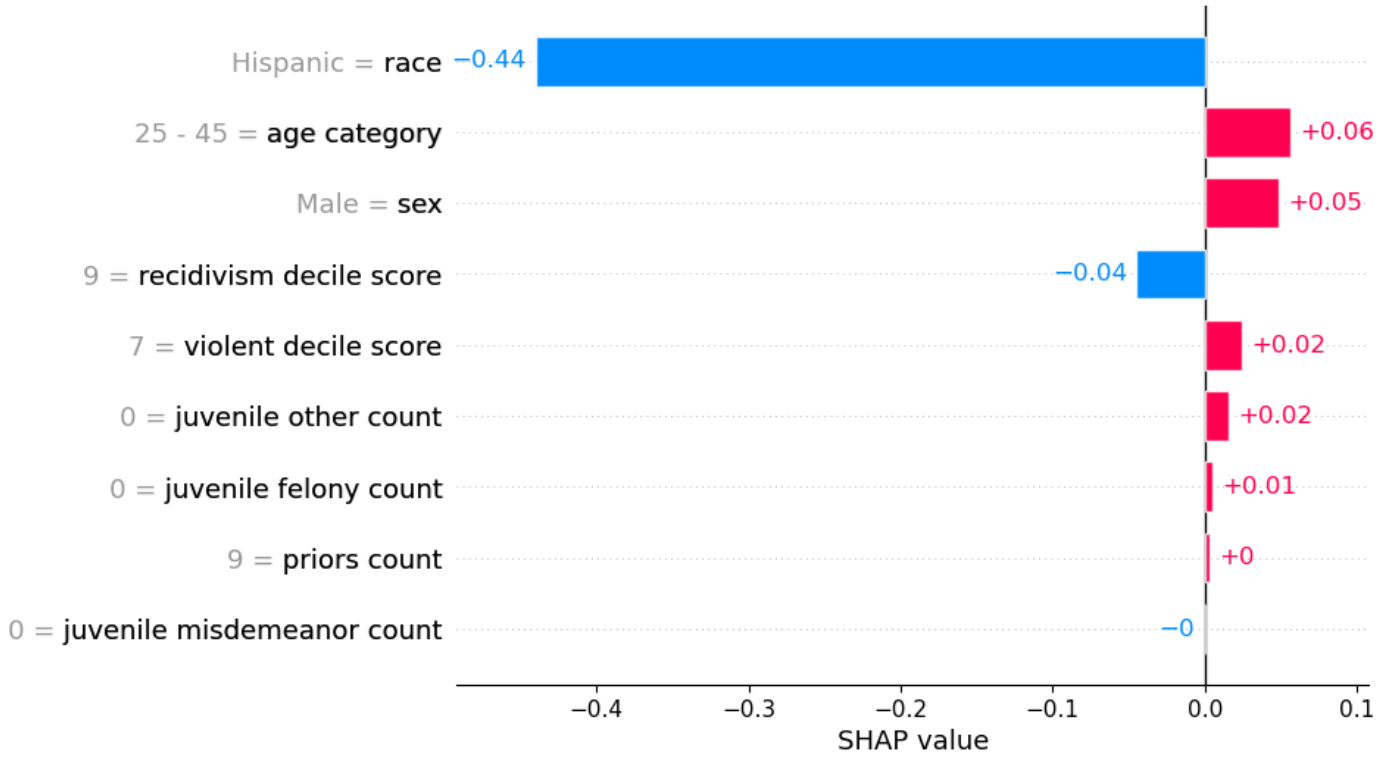
Click "yes" to see the descriptions of the attributes again.

- Yes
- No

How would you categorize this criminal defendant?

- I think that the criminal defendant is not likely to reoffend.
- I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- No, I do not agree.
- Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	25 - 45
race	Caucasian
priors count	5
recidivism decile score	10
violence decile score	5
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	0

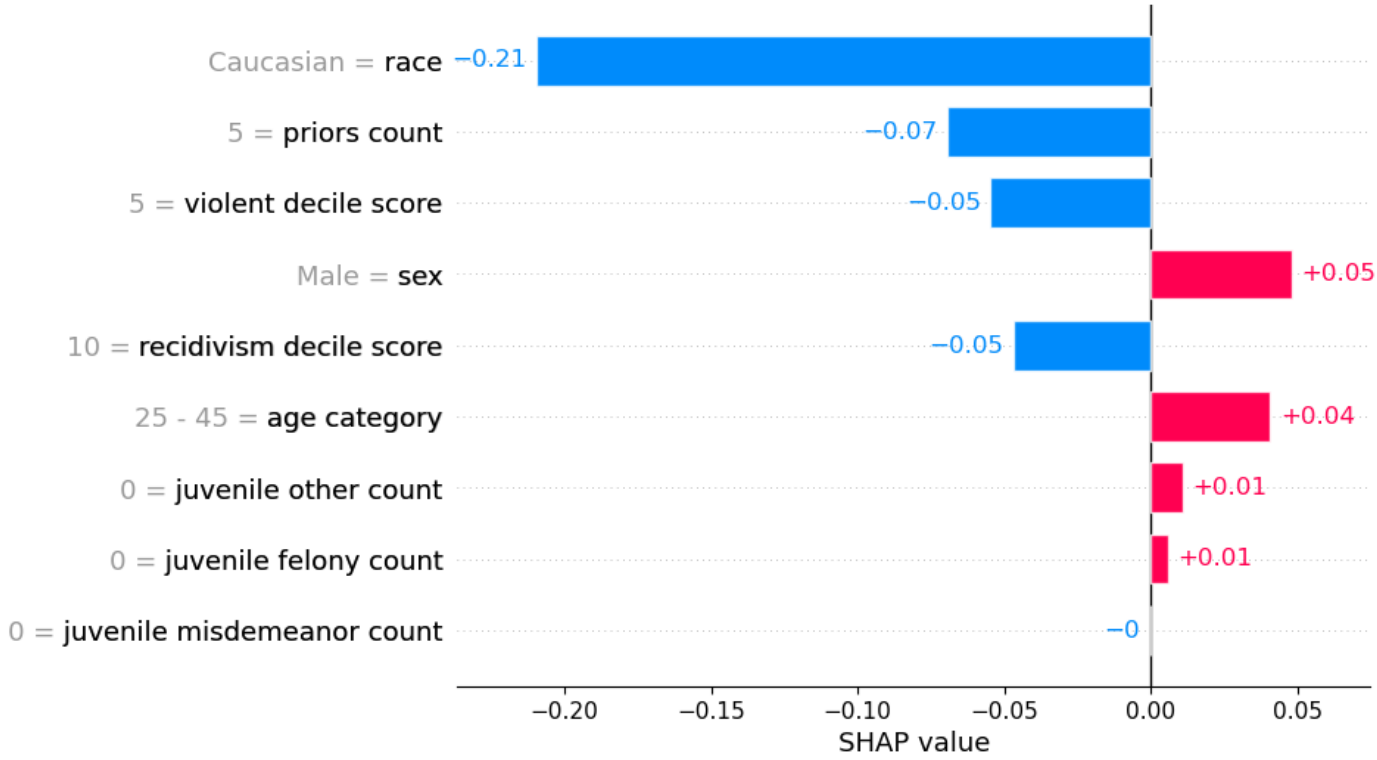
Click "yes" to see the descriptions of the attributes again.

- ○ Yes
- ○ No

How would you categorize this criminal defendant?

- ○ I think that the criminal defendant is not likely to reoffend.
- ○ I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- ○ No, I do not agree.
- ○ Yes, I agree.

Condition 3 - Intuitiveness / Faithfulness / Prohibited Features

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	Less than 25
race	Hispanic
priors count	0
recidivism decile score	7
violence decile score	6
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	1

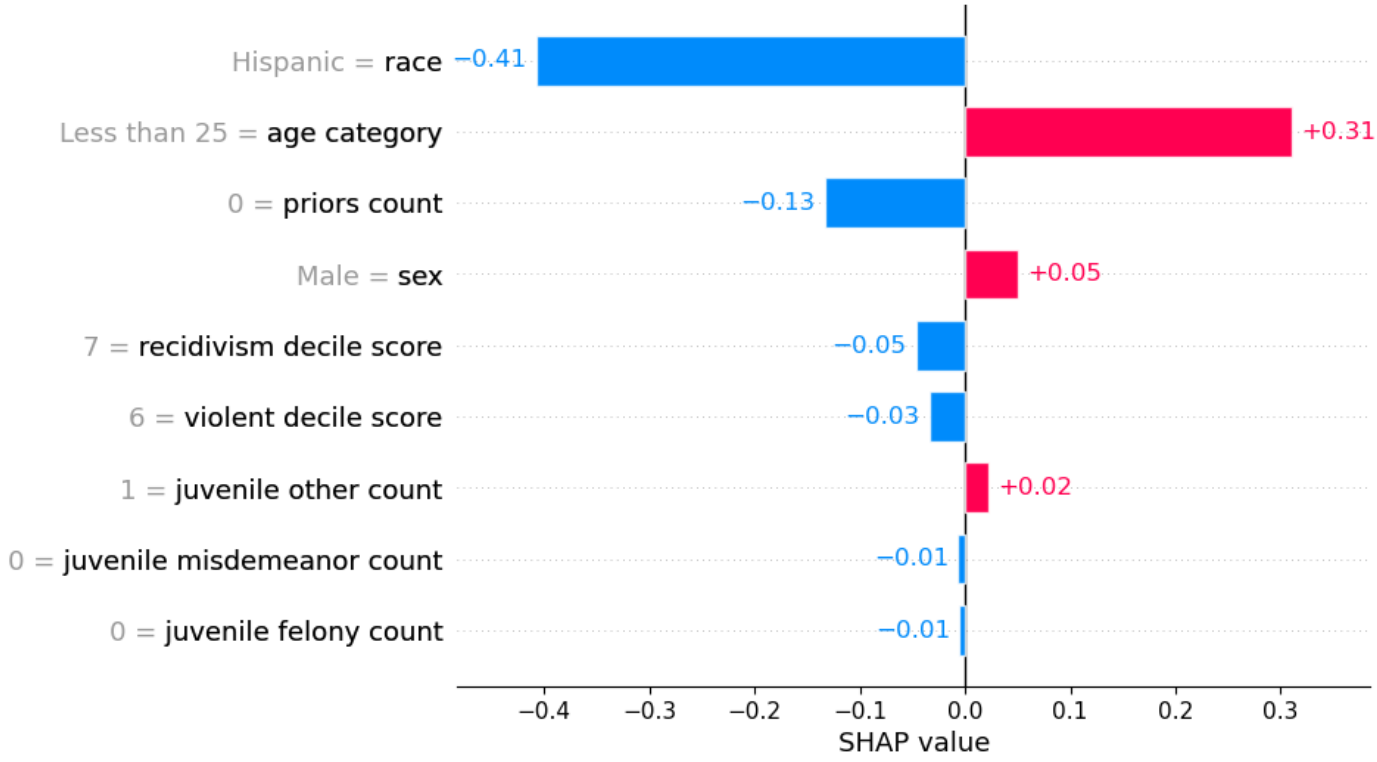
Click "yes" to see the descriptions of the attributes again.

- ○ Yes
- ○ No

How would you categorize this criminal defendant?

- ○ I think that the criminal defendant is not likely to reoffend.
- ○ I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is not likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- ○ No, I do not agree.
- ○ Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	Less than 25
race	Asian
priors count	0

Attributes	Profile
recidivism decile score	6
violence decile score	6
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	1

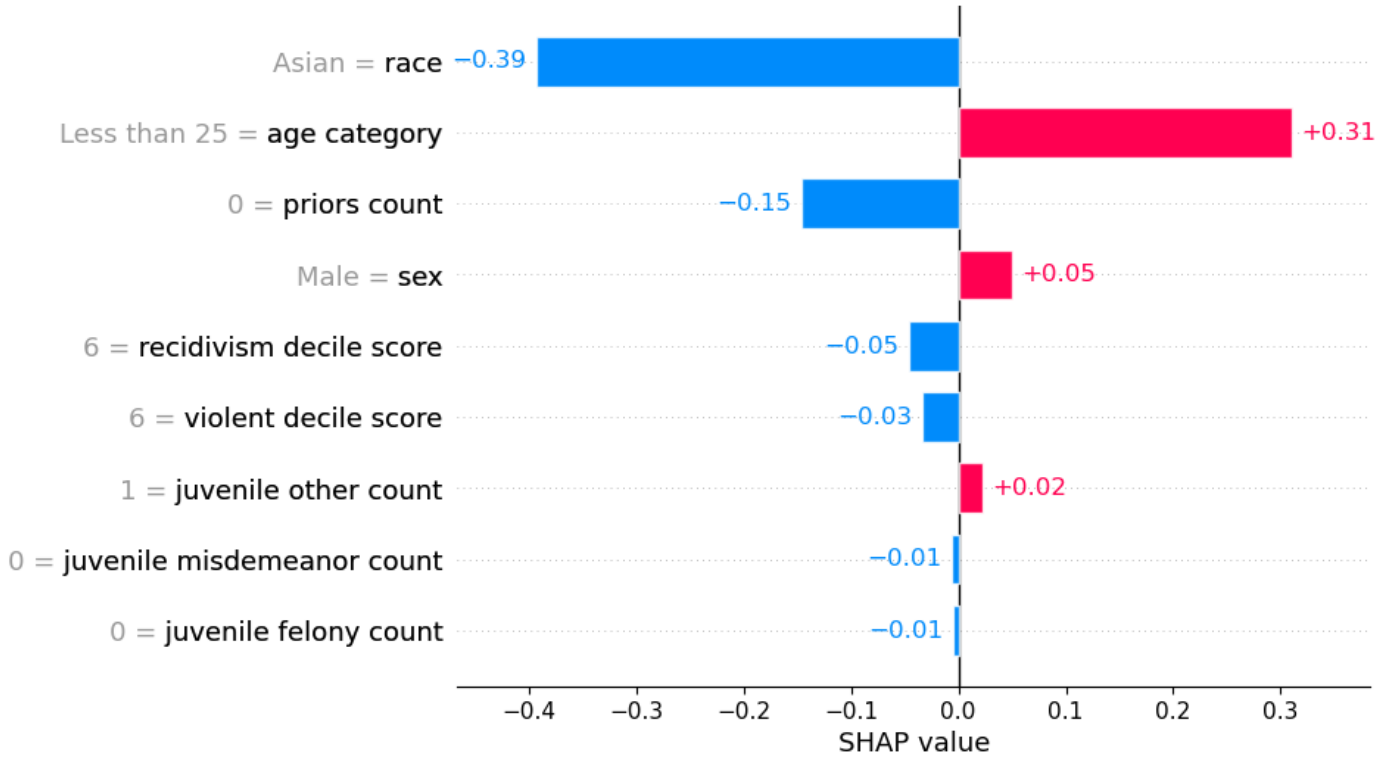
Click "yes" to see the descriptions of the attributes again.

- Yes
- No

How would you categorize this criminal defendant?

- I think that the criminal defendant is not likely to reoffend.
- I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is not likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- No, I do not agree.
- Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	Greater than 45
race	Caucasian
priors count	2
recidivism decile score	4
violence decile score	1
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	1

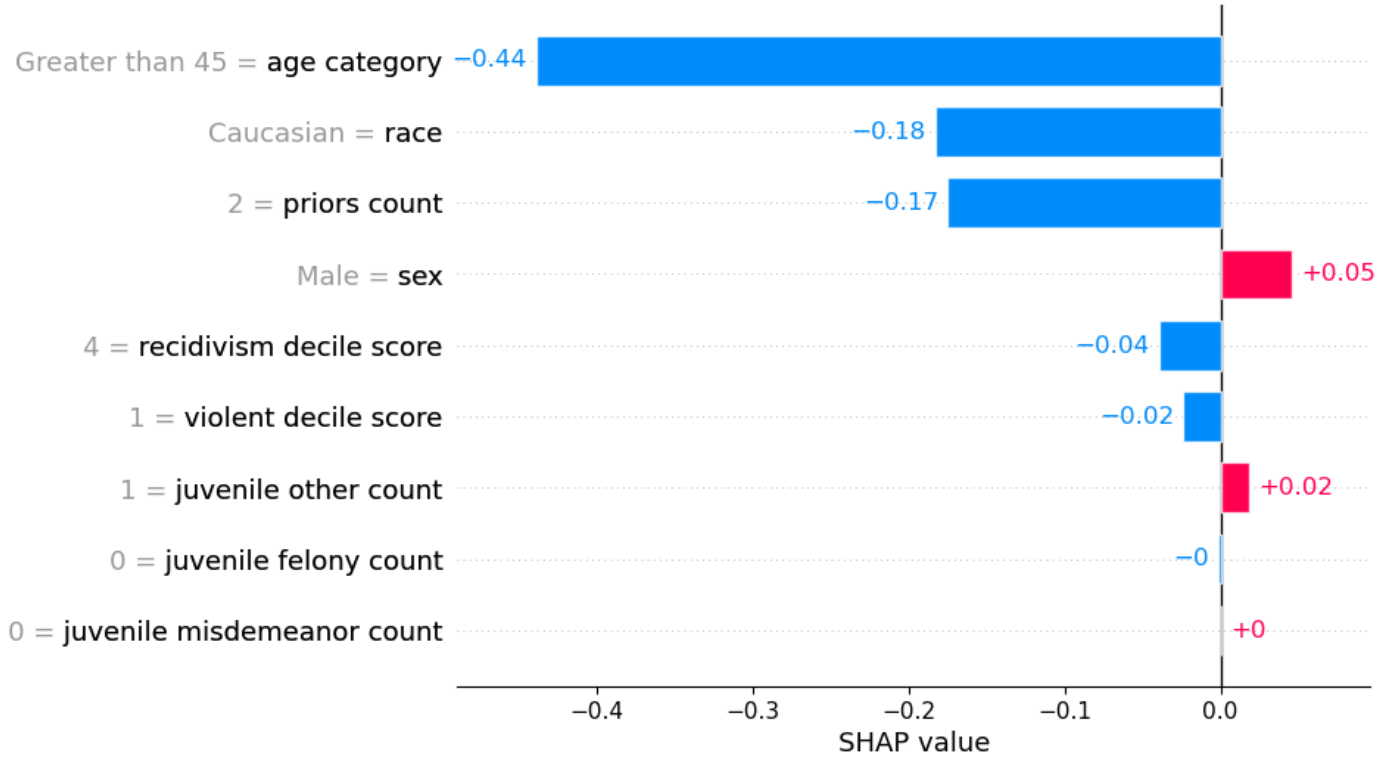
Click "yes" to see the descriptions of the attributes again.

- ○ Yes
- ○ No

How would you categorize this criminal defendant?

- ○ I think that the criminal defendant is not likely to reoffend.
- ○ I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is not likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- ○ No, I do not agree.
- ○ Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	25 - 45
race	African-American
priors count	1
recidivism decile score	10
violence decile score	8
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	0

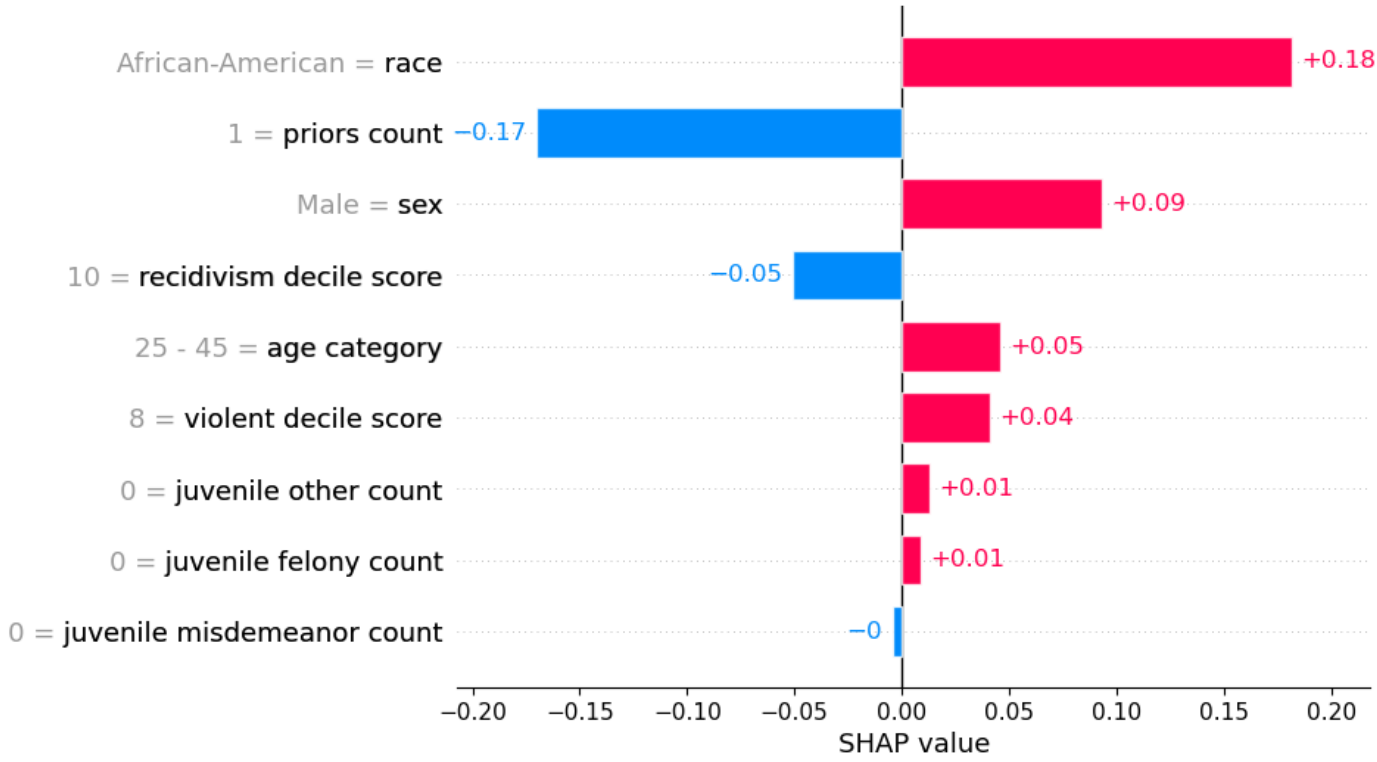
Click "yes" to see the descriptions of the attributes again.

- ○ Yes
- ○ No

How would you categorize this criminal defendant?

- I think that the criminal defendant is not likely to reoffend.
- I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- No, I do not agree.
- Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	25 - 45
race	Caucasian
priors count	3
recidivism decile score	8
violence decile score	2

Attributes	Profile
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	0

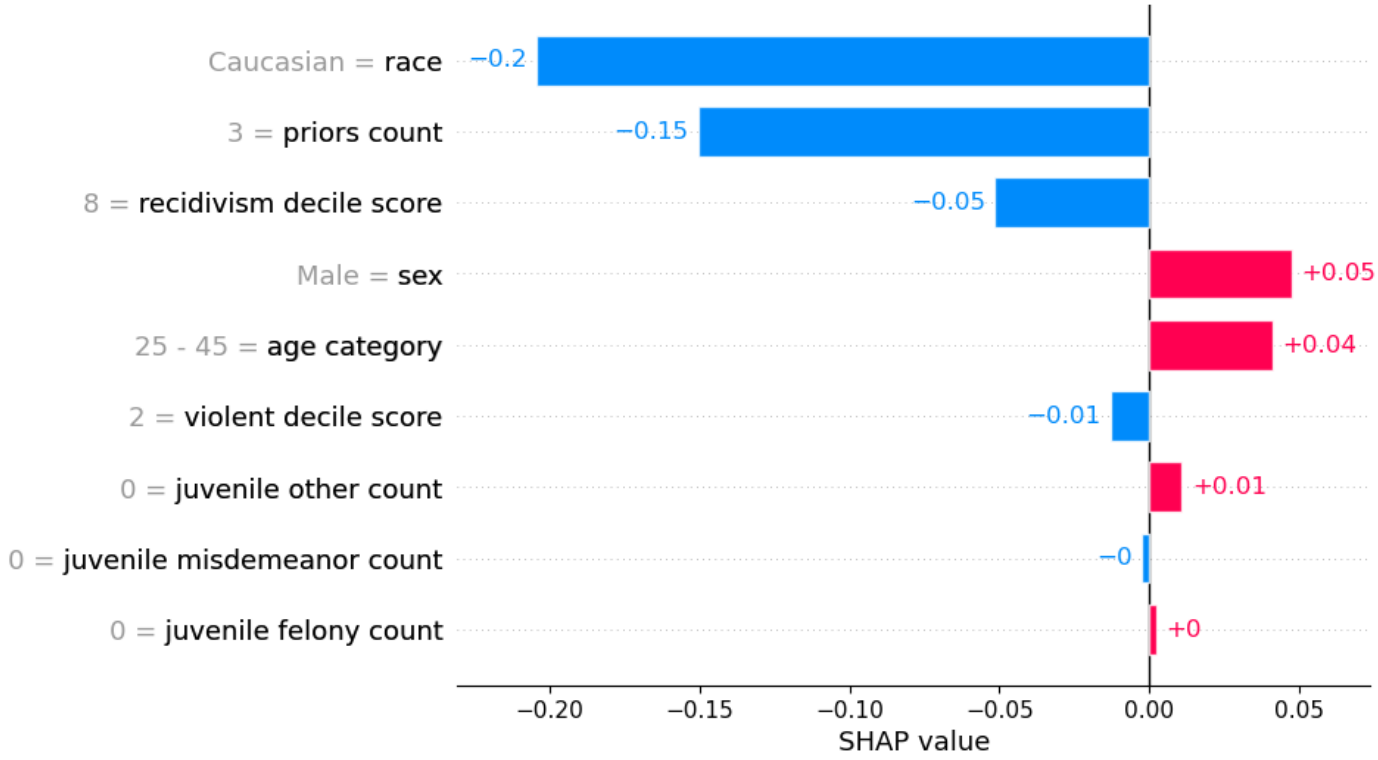
Click "yes" to see the descriptions of the attributes again.

- Yes
- No

How would you categorize this criminal defendant?

- I think that the criminal defendant is not likely to reoffend.
- I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is not likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- No, I do not agree.
- Yes, I agree.

If you don't feel familiar enough with the model yet, you can get 2 more examples.

- I need more examples to become familiar with the model.
- I am familiar enough and do not need more examples.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	Less than 25
race	Caucasian
priors count	0
recidivism decile score	9
violence decile score	8
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	0

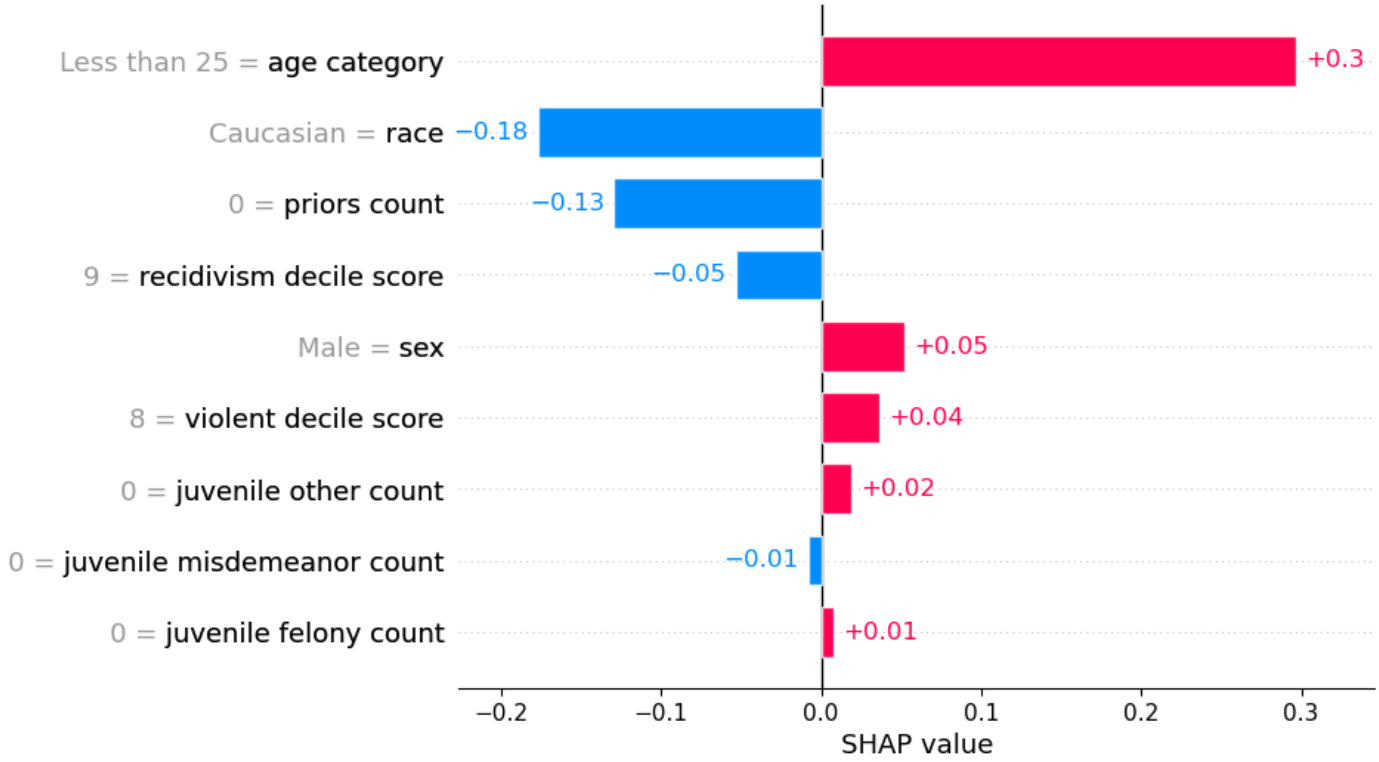
Click "yes" to see the descriptions of the attributes again.

- Yes
- No

How would you categorize this criminal defendant?

- I think that the criminal defendant is not likely to reoffend.
- I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is not likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- ○ No, I do not agree.
- ○ Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	Less than 25
race	Caucasian
priors count	0
recidivism decile score	6
violence decile score	9
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	0

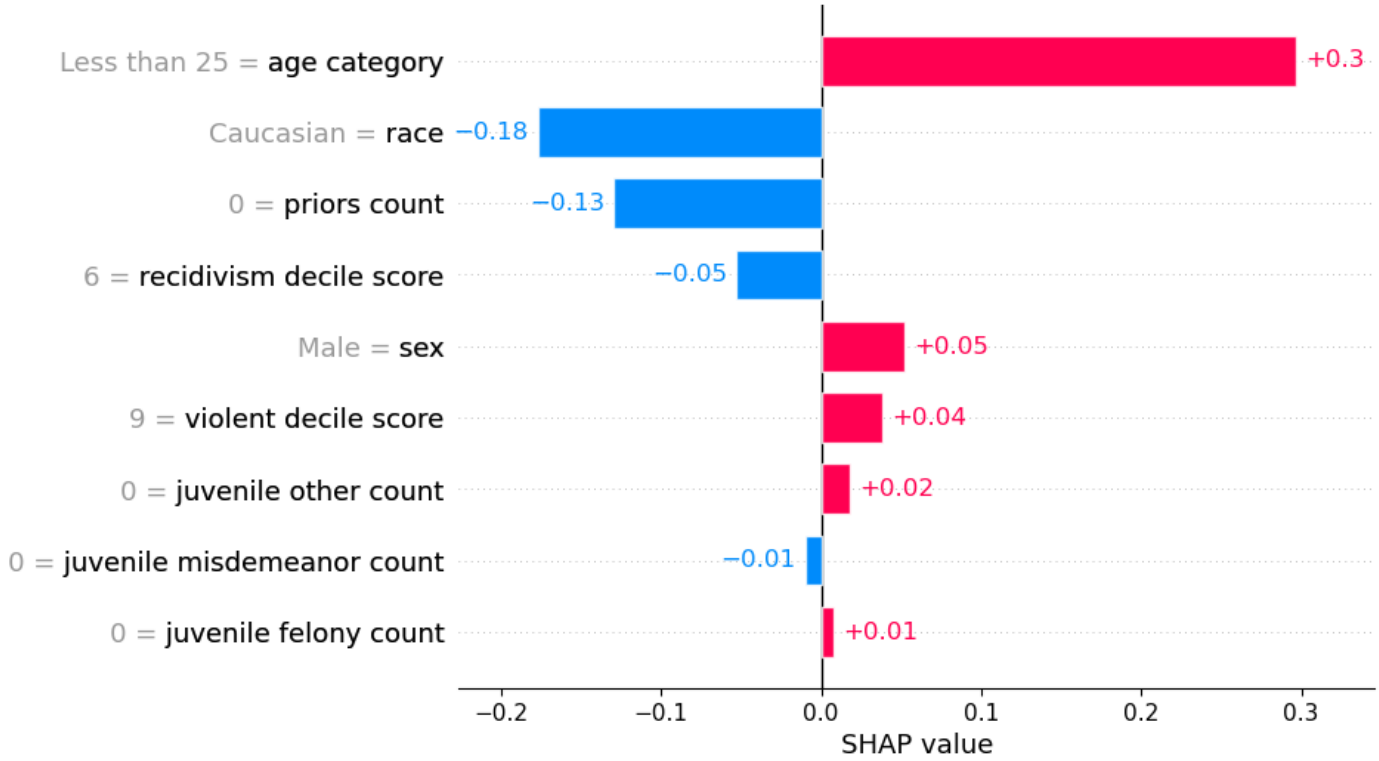
Click "yes" to see the descriptions of the attributes again.

- ○ Yes
- ○ No

How would you categorize this criminal defendant?

- I think that the criminal defendant is not likely to reoffend.
- I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is not likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- No, I do not agree.
- Yes, I agree.

Condition 4 - Intuitiveness / Unfaithfulness / Prohibited Features

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	Less than 25
race	Hispanic
priors count	0

Attributes	Profile
recidivism decile score	7
violence decile score	6
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	1

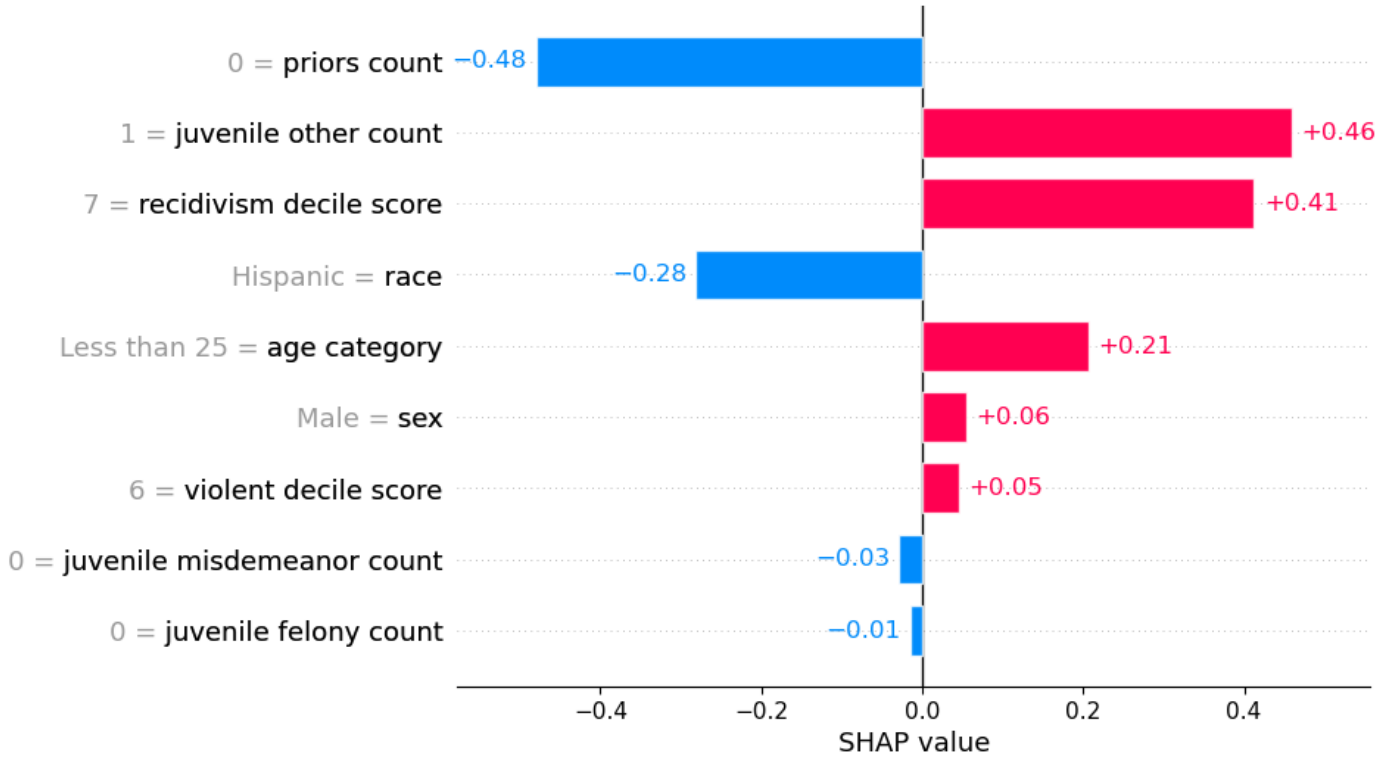
Click "yes" to see the descriptions of the attributes again.

- Yes
- No

How would you categorize this criminal defendant?

- I think that the criminal defendant is not likely to reoffend.
- I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is not likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- No, I do not agree.
- Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	Less than 25
race	Asian
priors count	0
recidivism decile score	6
violence decile score	6
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	1

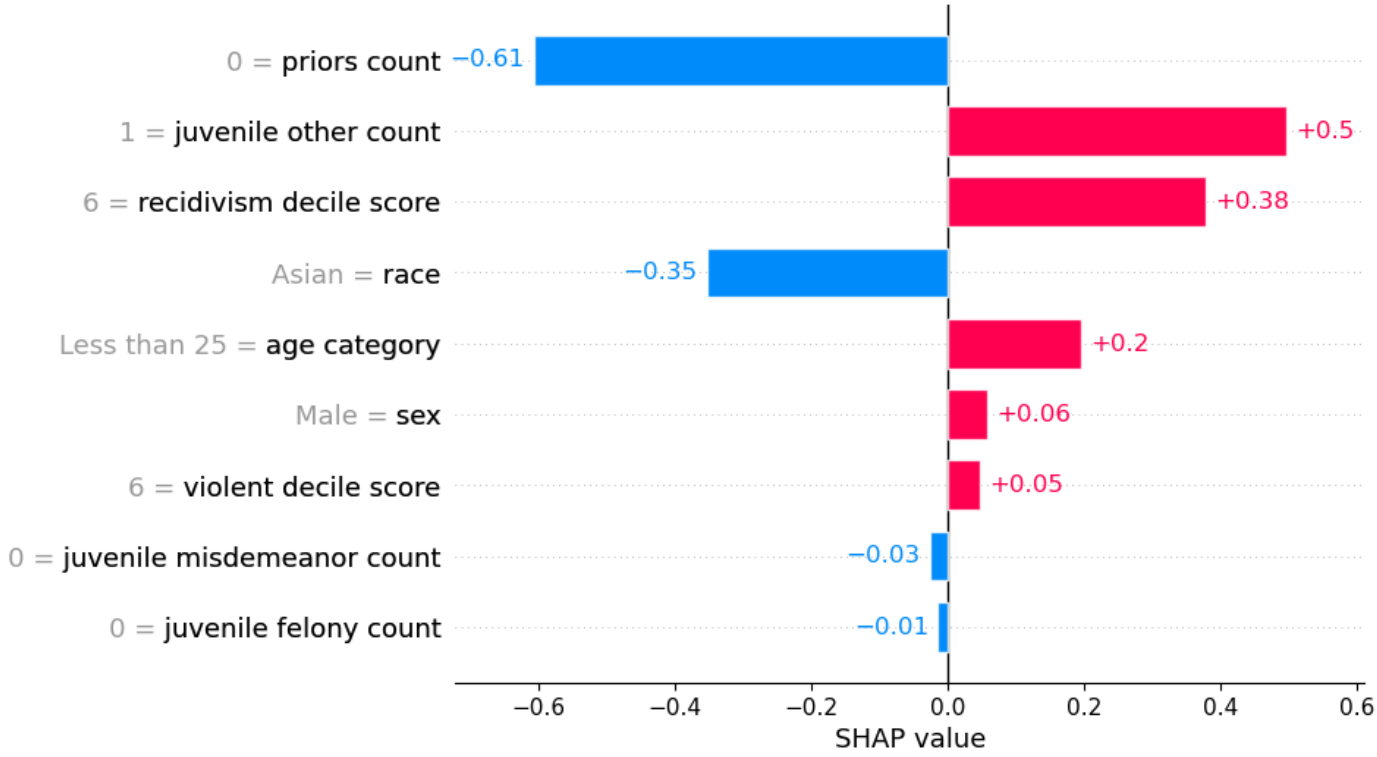
Click "yes" to see the descriptions of the attributes again.

- ○ Yes
- ○ No

How would you categorize this criminal defendant?

- ○ I think that the criminal defendant is not likely to reoffend.
- ○ I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is not likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- ○ No, I do not agree.
- ○ Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	Greater than 45
race	Caucasian
priors count	2
recidivism decile score	4
violence decile score	1
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	1

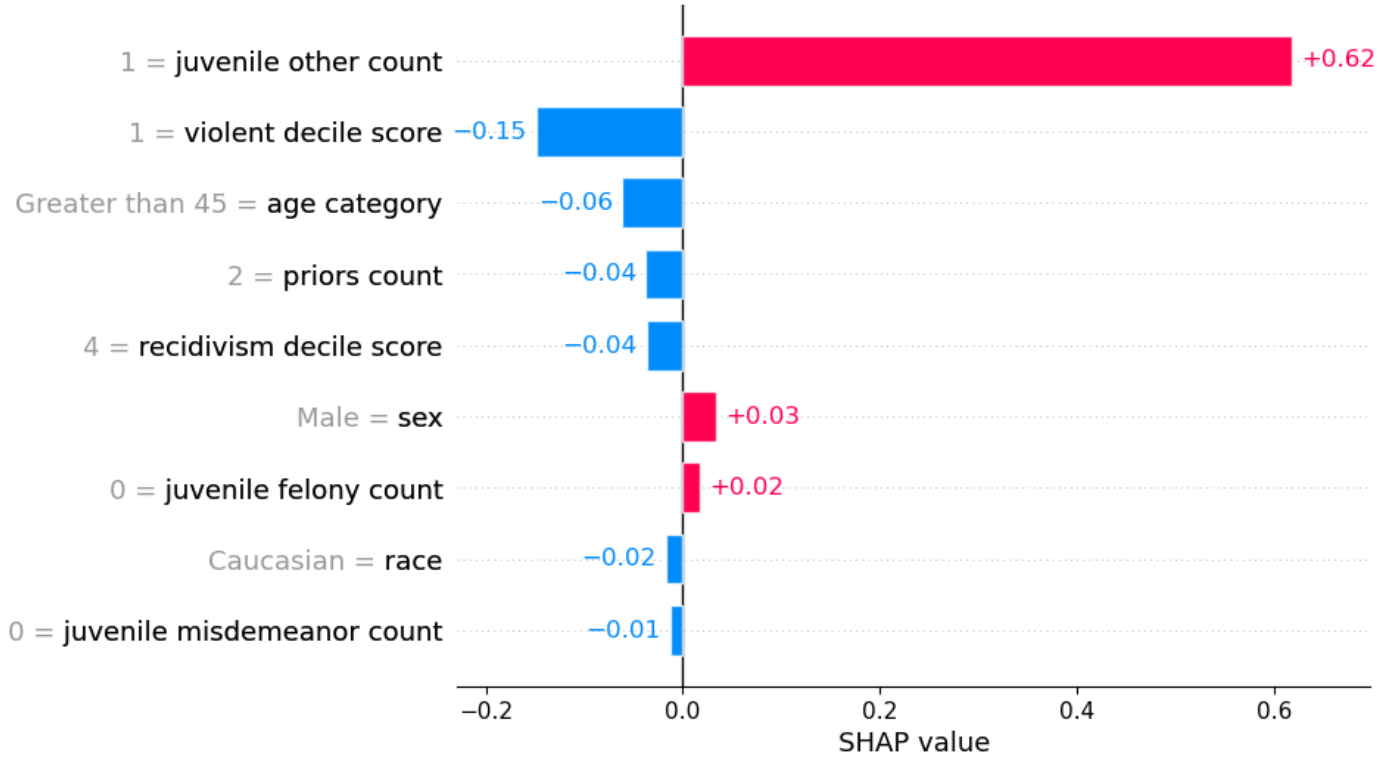
Click "yes" to see the descriptions of the attributes again.

- ○ Yes
- ○ No

How would you categorize this criminal defendant?

- I think that the criminal defendant is not likely to reoffend.
- I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is not likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- No, I do not agree.
- Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	**Profile**
sex	Male
age category	25 - 45
race	African-American
priors count	1
recidivism decile score	10
violence decile score	8

Attributes	Profile
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	0

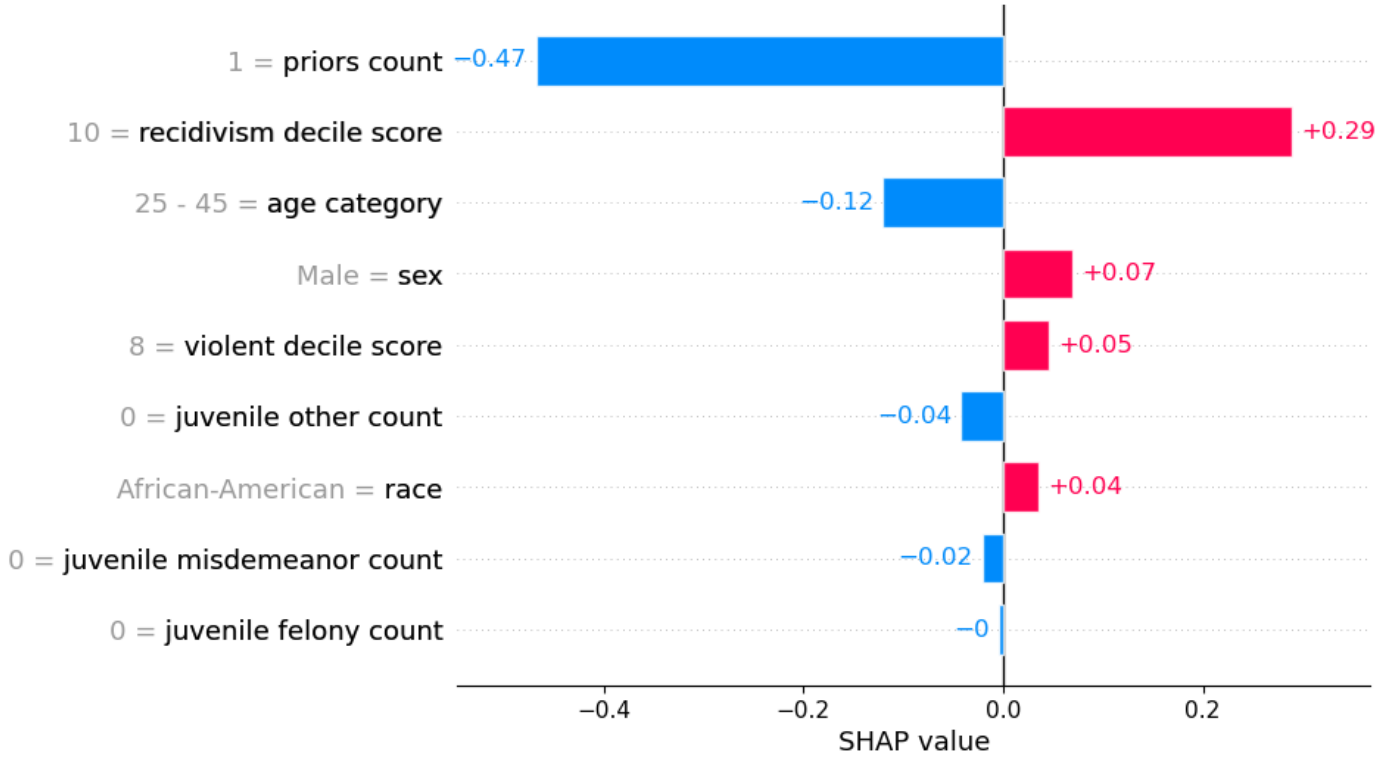
Click "yes" to see the descriptions of the attributes again.

- ○ Yes
- ○ No

How would you categorize this criminal defendant?

- ○ I think that the criminal defendant is not likely to reoffend.
- ○ I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- ○ No, I do not agree.
- ○ Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	25 - 45
race	Caucasian
priors count	3
recidivism decile score	8
violence decile score	2
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	0

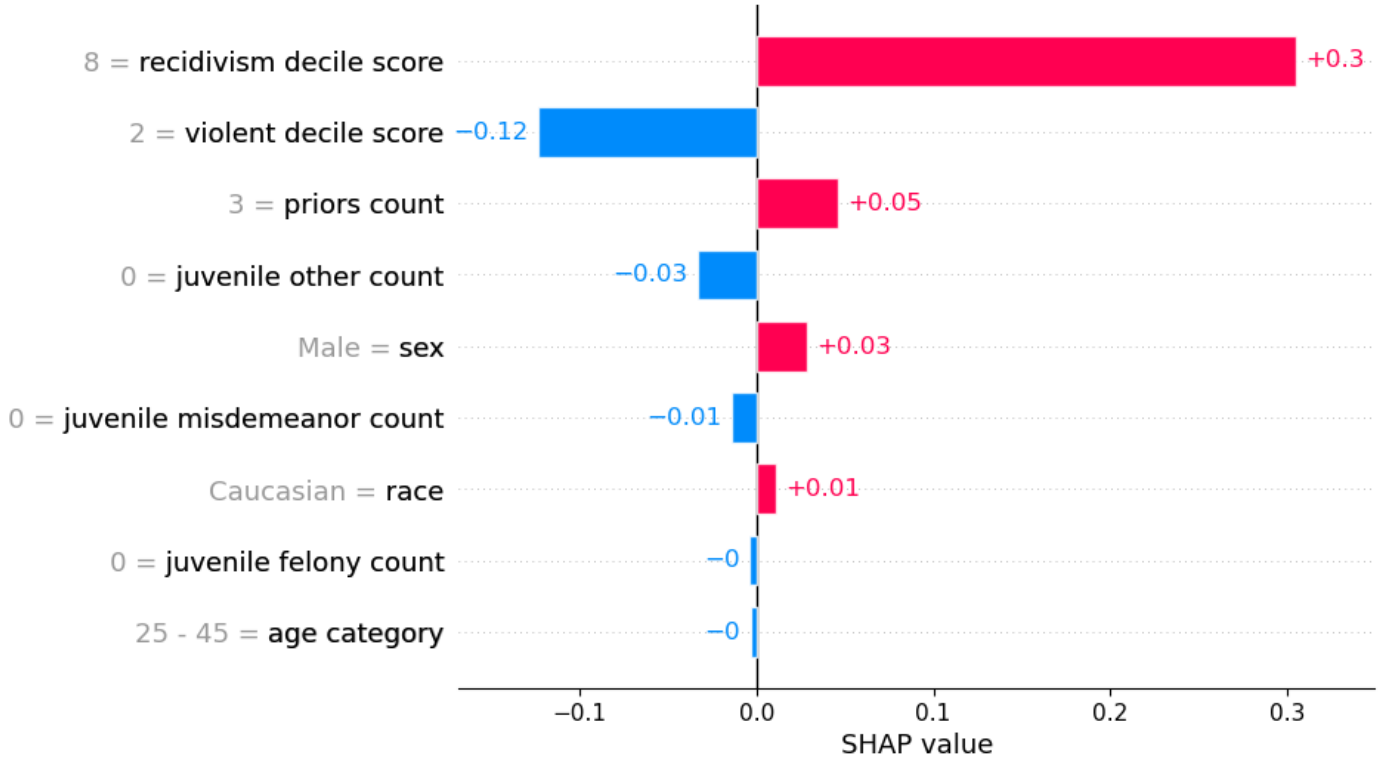
Click "yes" to see the descriptions of the attributes again.

- ○ Yes
- ○ No

How would you categorize this criminal defendant?

- ○ I think that the criminal defendant is not likely to reoffend.
- ○ I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is not likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- ○ No, I do not agree.
- ○ Yes, I agree.

If you don't feel familiar enough with the model yet, you can get 2 more examples.

- ○ I need more examples to become familiar with the model.
- ○ I am familiar enough and do not need more examples.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	Less than 25
race	Caucasian
priors count	0
recidivism decile score	9
violence decile score	8
juvenile felony count	0
juvenile misdemeanor count	0

Attributes	Profile
juvenile other count	0

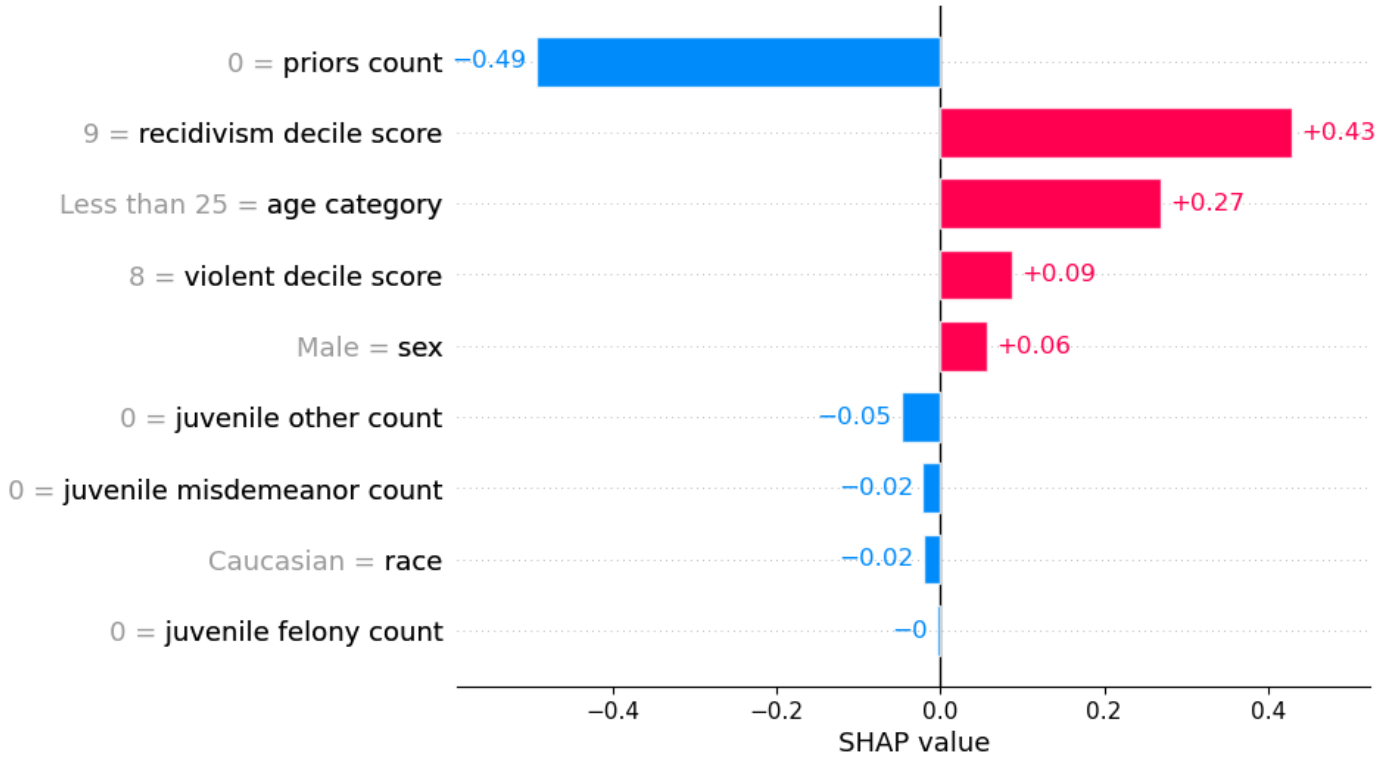
Click "yes" to see the descriptions of the attributes again.

- ○ Yes
- ○ No

How would you categorize this criminal defendant?

- ○ I think that the criminal defendant is not likely to reoffend.
- ○ I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is not likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- ○ No, I do not agree.
- ○ Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	Less than 25
race	Caucasian
priors count	0
recidivism decile score	6
violence decile score	9
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	0

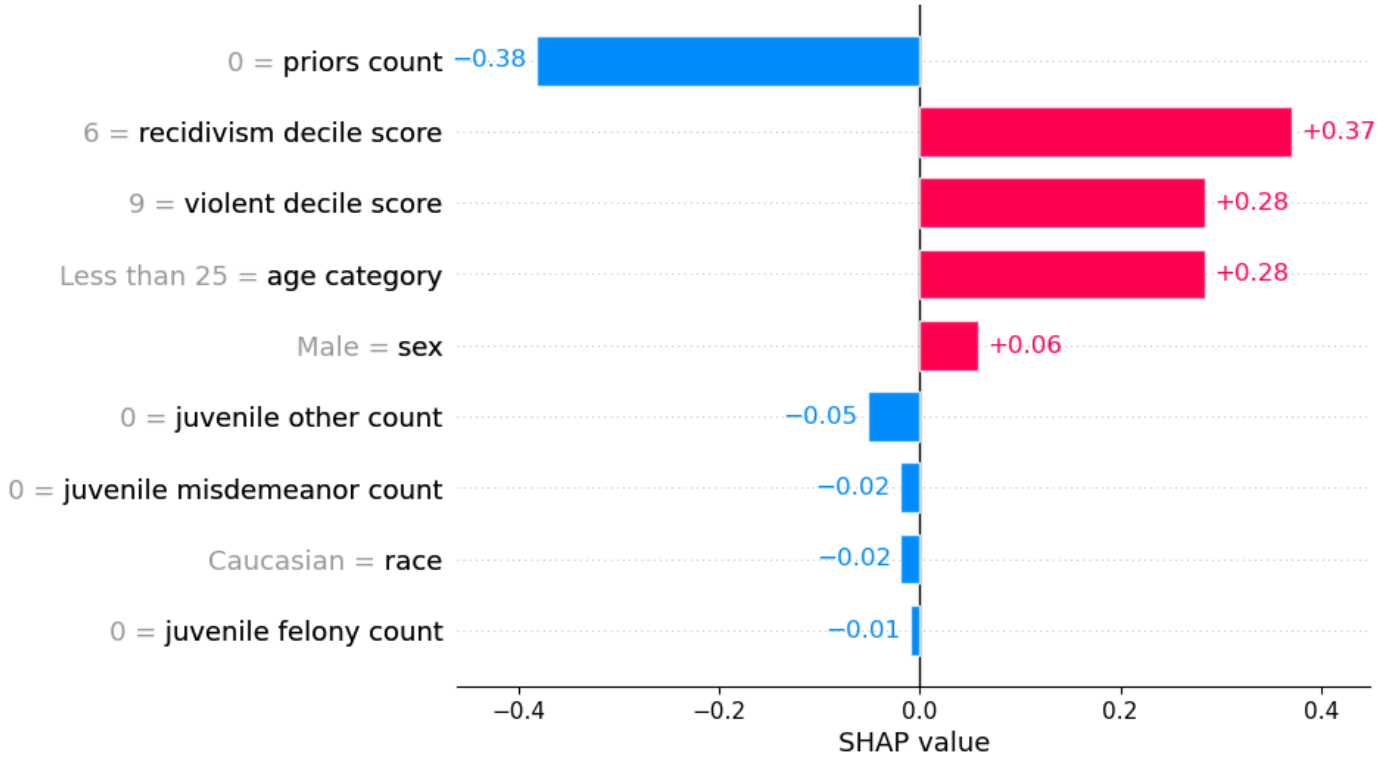
Click "yes" to see the descriptions of the attributes again.

- ○ Yes
- ○ No

How would you categorize this criminal defendant?

- ○ I think that the criminal defendant is not likely to reoffend.
- ○ I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is not likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- ○ No, I do not agree.
- ○ Yes, I agree.

Condition 5 - Non-intuitiveness / Faithfulness / Desired Features

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	Less than 25
race	Hispanic
priors count	0
recidivism decile score	7
violence decile score	6
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	1

Click "yes" to see the descriptions of the attributes again.

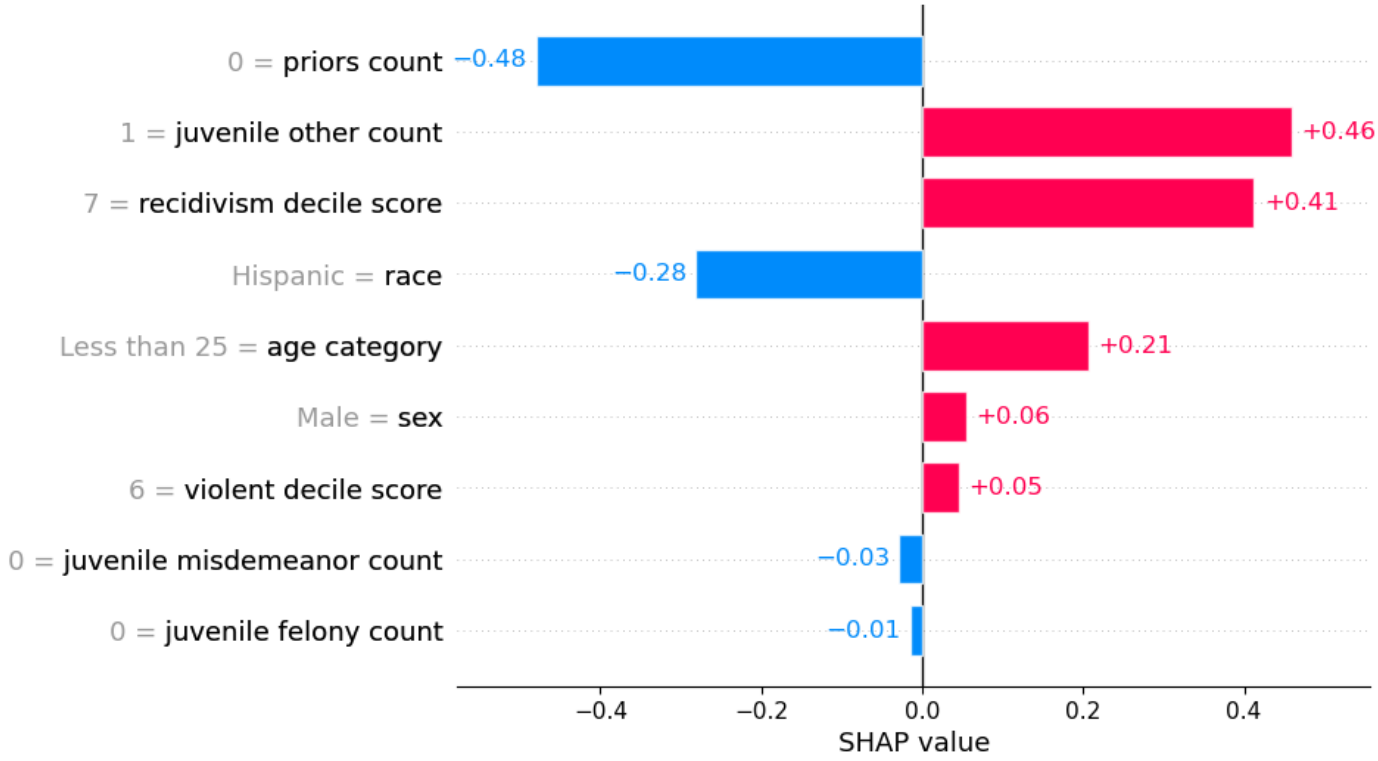
- ○ Yes

- ○ No

How would you categorize this criminal defendant?

- ○ I think that the criminal defendant is not likely to reoffend.
- ○ I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- ○ No, I do not agree.
- ○ Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	Less than 25
race	Asian
priors count	0
recidivism decile score	6

Attributes	Profile
violence decile score	6
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	1

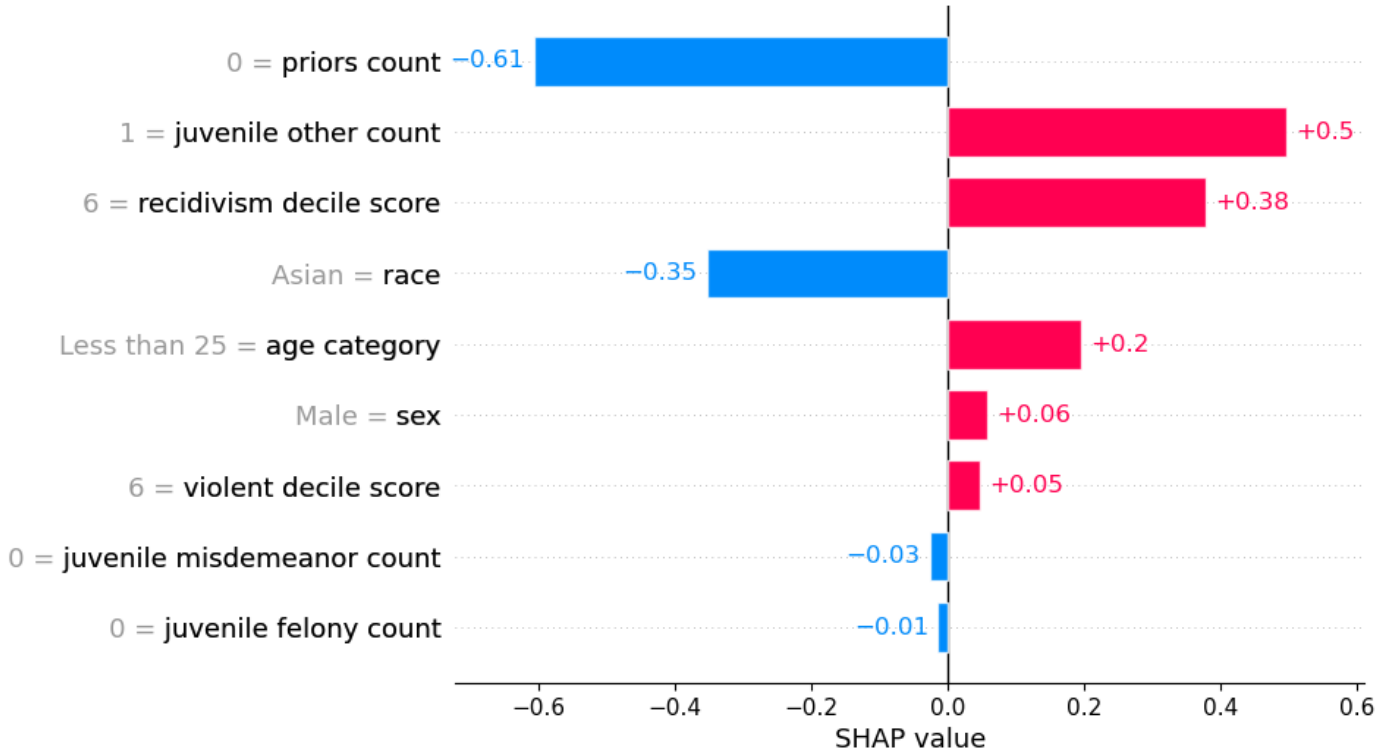
Click "yes" to see the descriptions of the attributes again.

- Yes
- No

How would you categorize this criminal defendant?

- I think that the criminal defendant is not likely to reoffend.
- I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- No, I do not agree.
- Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	Greater than 45
race	Caucasian
priors count	2
recidivism decile score	4
violence decile score	1
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	1

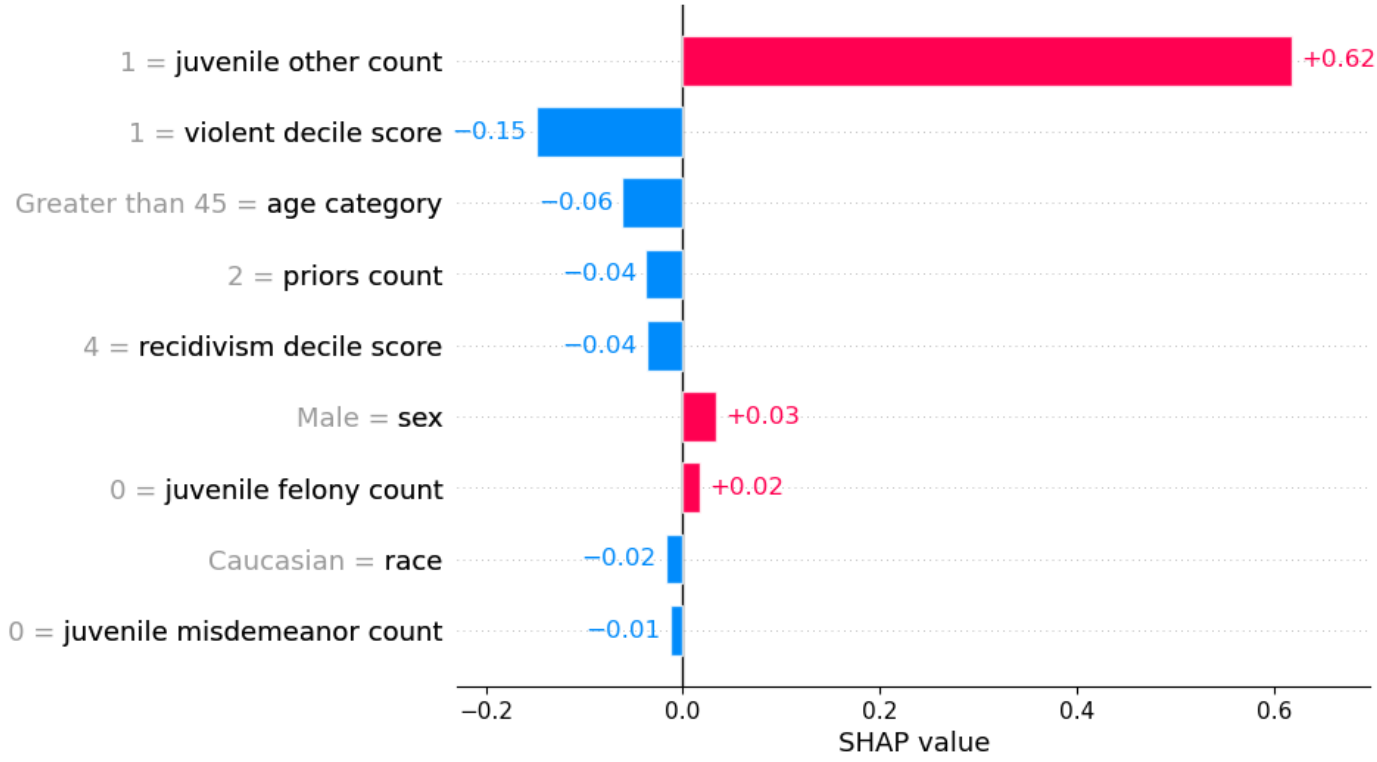
Click "yes" to see the descriptions of the attributes again.

- ○ Yes
- ○ No

How would you categorize this criminal defendant?

- ○ I think that the criminal defendant is not likely to reoffend.
- ○ I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- ○ No, I do not agree.
- ○ Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	25 - 45
race	African-American
priors count	1
recidivism decile score	10
violence decile score	8
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	0

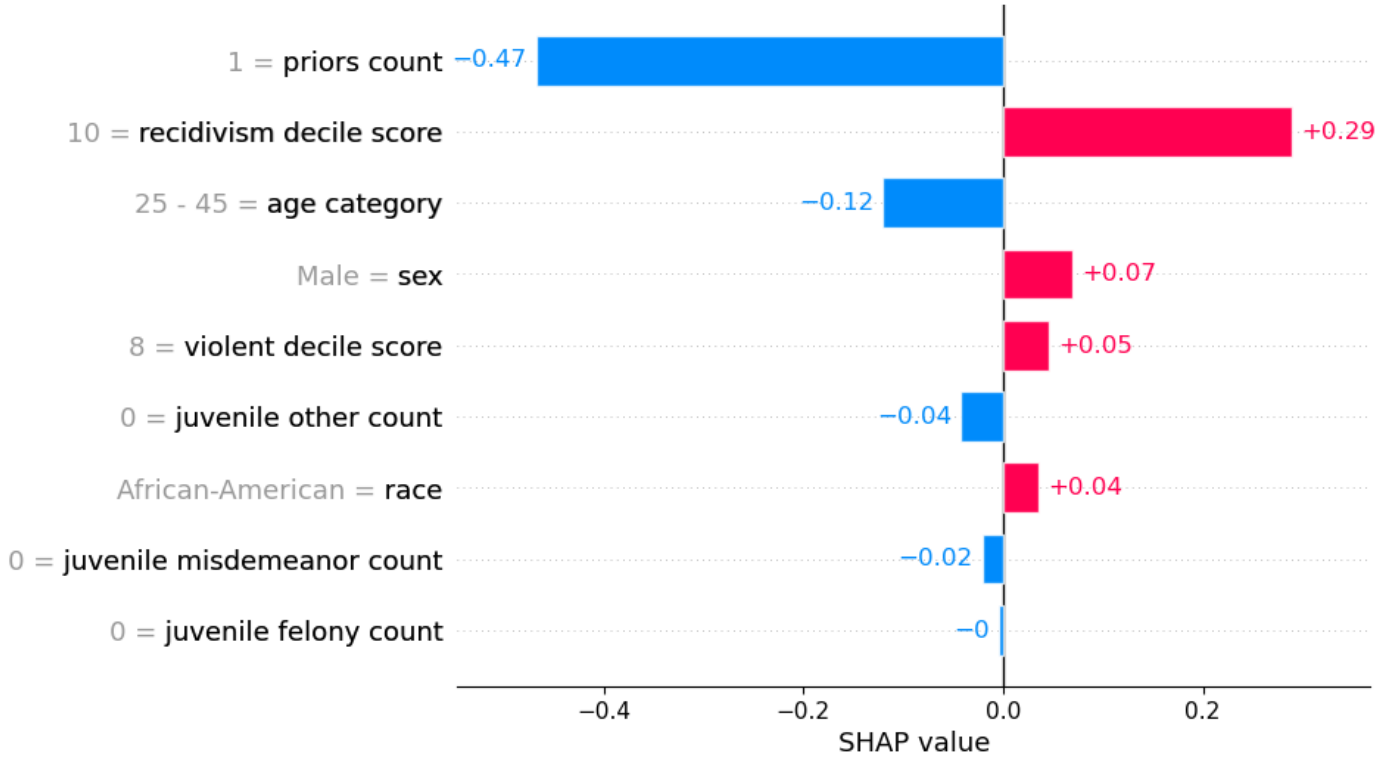
Click "yes" to see the descriptions of the attributes again.

- ○ Yes
- ○ No

How would you categorize this criminal defendant?

- I think that the criminal defendant is not likely to reoffend.
- I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is not likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- No, I do not agree.
- Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	25 - 45
race	Caucasian
priors count	3
recidivism decile score	8
violence decile score	2

Attributes	Profile
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	0

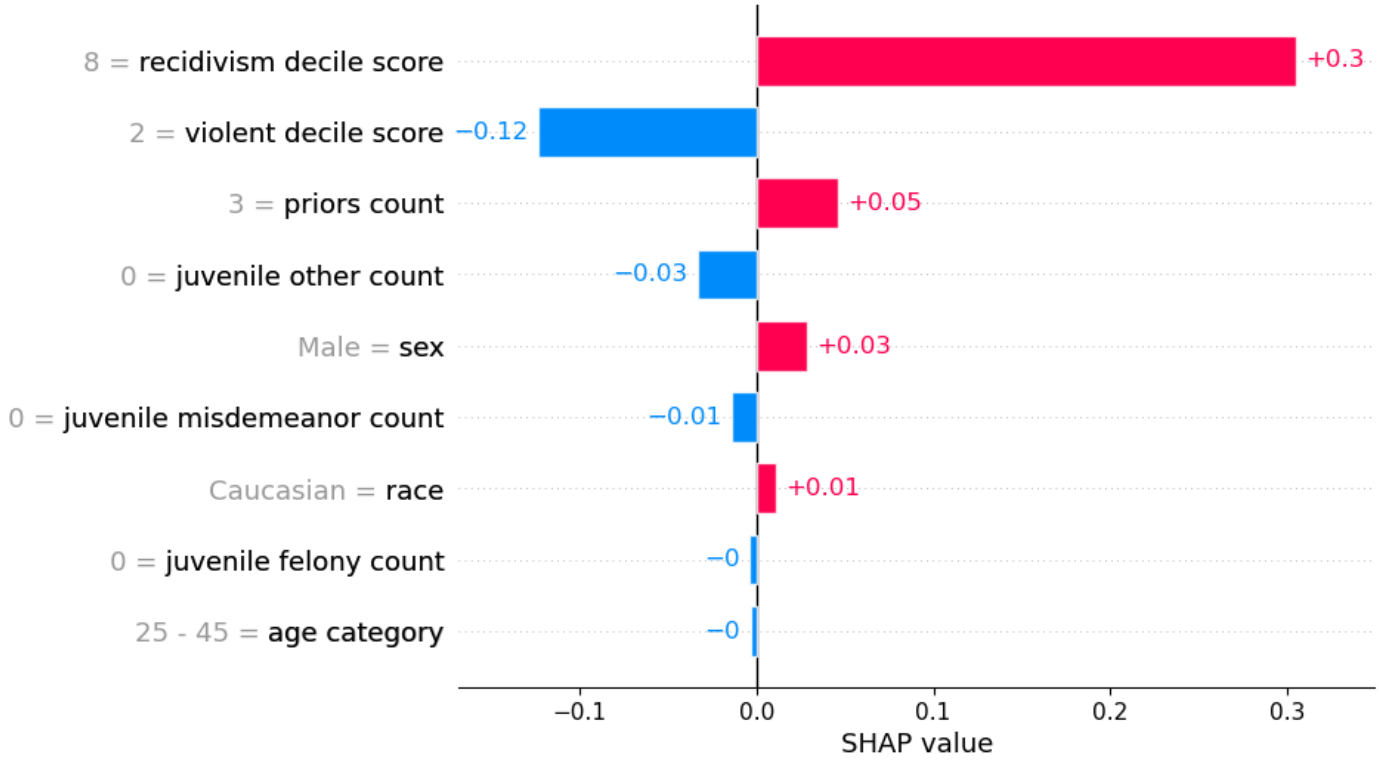
Click "yes" to see the descriptions of the attributes again.

- Yes
- No

How would you categorize this criminal defendant?

- I think that the criminal defendant is not likely to reoffend.
- I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- No, I do not agree.
- Yes, I agree.

If you don't feel familiar enough with the model yet, you can get 2 more examples.

- I need more examples to become familiar with the model.
- I am familiar enough and do not need more examples.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	Less than 25
race	Caucasian
priors count	0
recidivism decile score	9
violence decile score	8
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	0

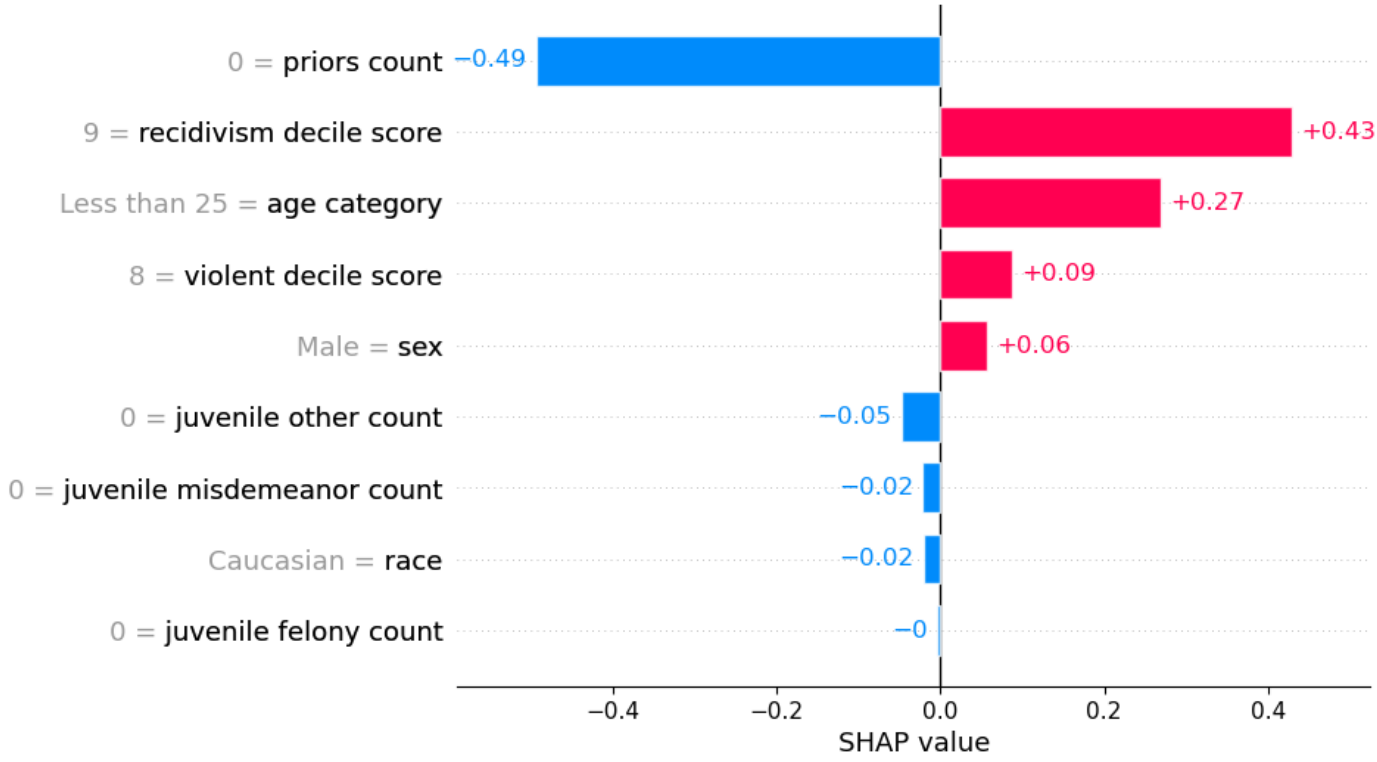
Click "yes" to see the descriptions of the attributes again.

- Yes
- No

How would you categorize this criminal defendant?

- I think that the criminal defendant is not likely to reoffend.
- I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- ○ No, I do not agree.
- ○ Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	Less than 25
race	Caucasian
priors count	0
recidivism decile score	6
violence decile score	9
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	0

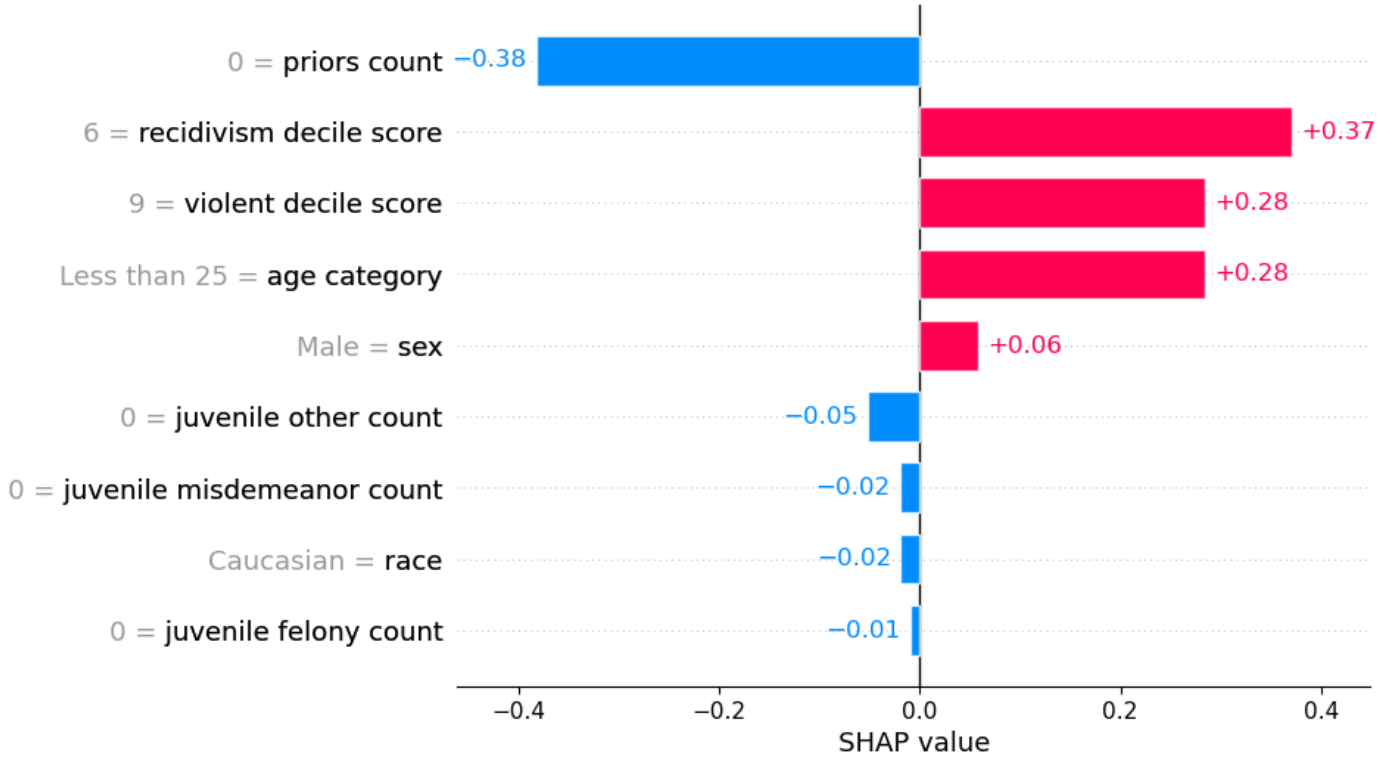
Click "yes" to see the descriptions of the attributes again.

- ○ Yes
- ○ No

How would you categorize this criminal defendant?

- I think that the criminal defendant is not likely to reoffend.
- I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- No, I do not agree.
- Yes, I agree.

Condition 6 - Non-intuitivness / Unfaithfulness / Desired Features

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	Less than 25
race	Hispanic
priors count	0

Attributes	Profile
recidivism decile score	7
violence decile score	6
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	1

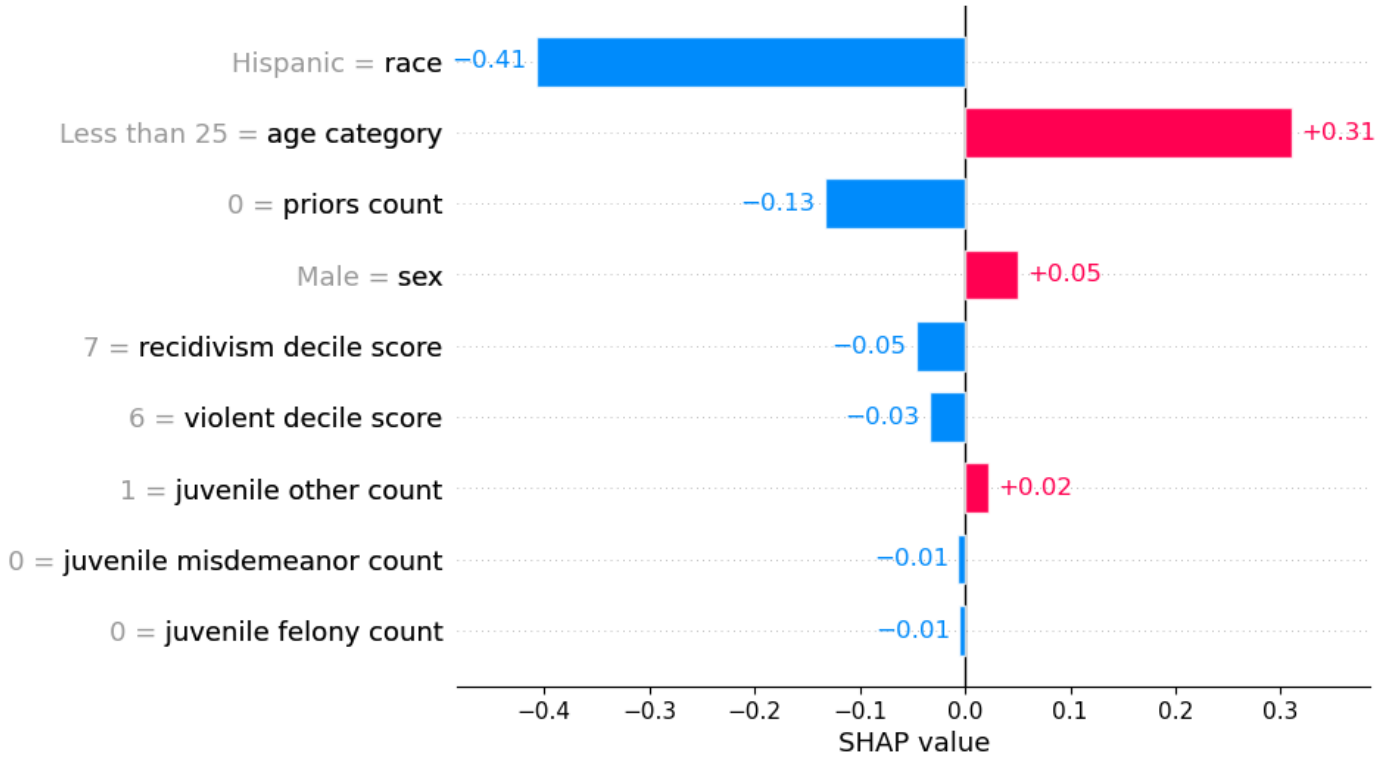
Click "yes" to see the descriptions of the attributes again.

- Yes
- No

How would you categorize this criminal defendant?

- I think that the criminal defendant is not likely to reoffend.
- I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- No, I do not agree.
- Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	Less than 25
race	Asian
priors count	0
recidivism decile score	6
violence decile score	6
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	1

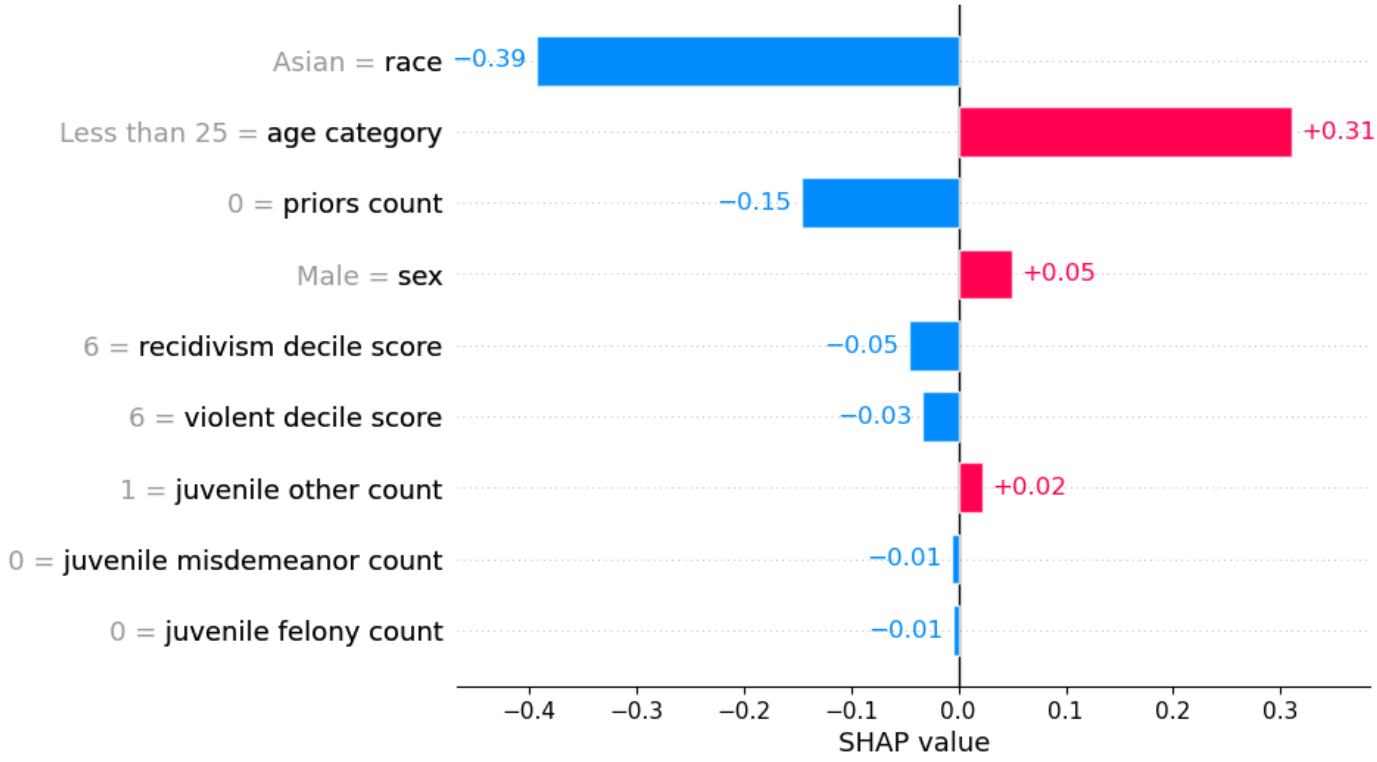
Click "yes" to see the descriptions of the attributes again.

- ○ Yes
- ○ No

How would you categorize this criminal defendant?

- ○ I think that the criminal defendant is not likely to reoffend.
- ○ I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- ○ No, I do not agree.
- ○ Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	Greater than 45
race	Caucasian
priors count	2
recidivism decile score	4
violence decile score	1
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	1

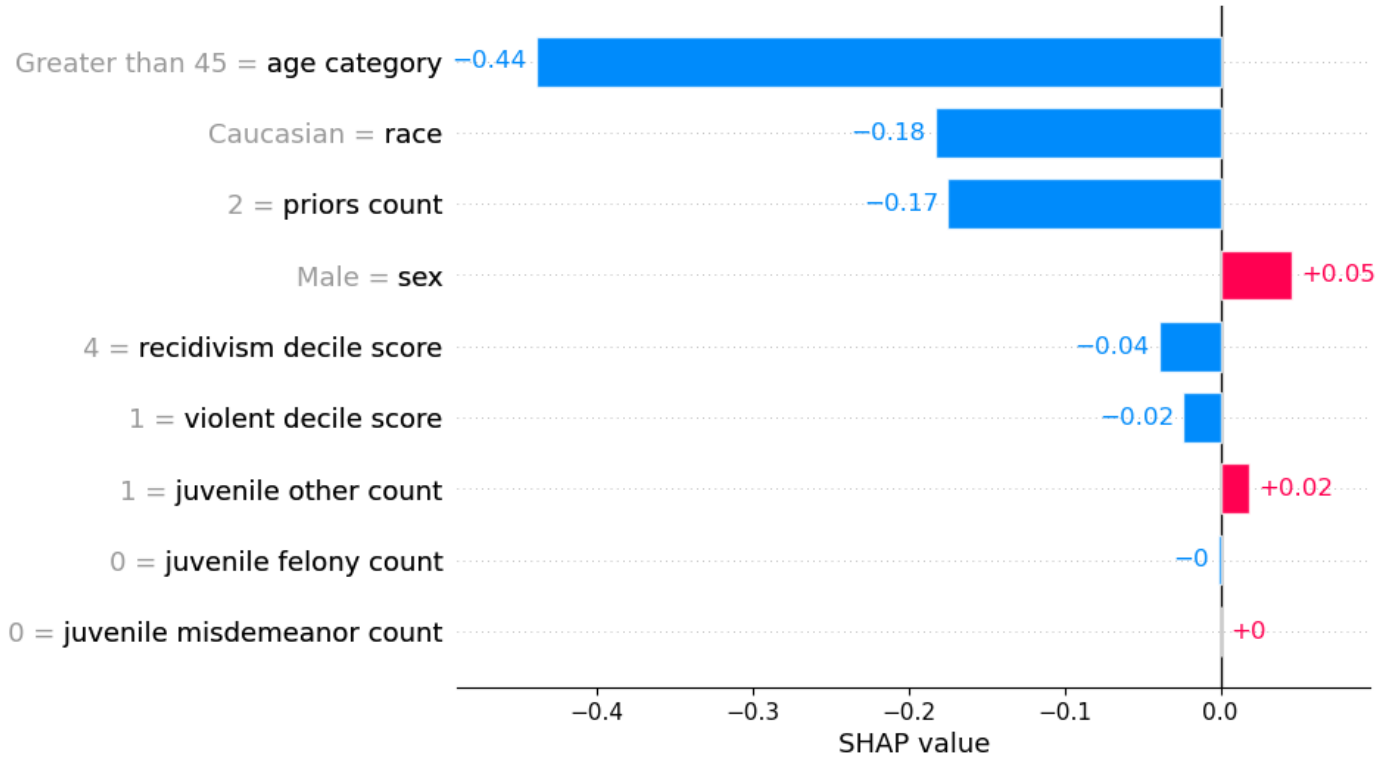
Click "yes" to see the descriptions of the attributes again.

- ○ Yes
- ○ No

How would you categorize this criminal defendant?

- I think that the criminal defendant is not likely to reoffend.
- I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- No, I do not agree.
- Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	25 - 45
race	African-American
priors count	1
recidivism decile score	10
violence decile score	8

Attributes	Profile
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	0

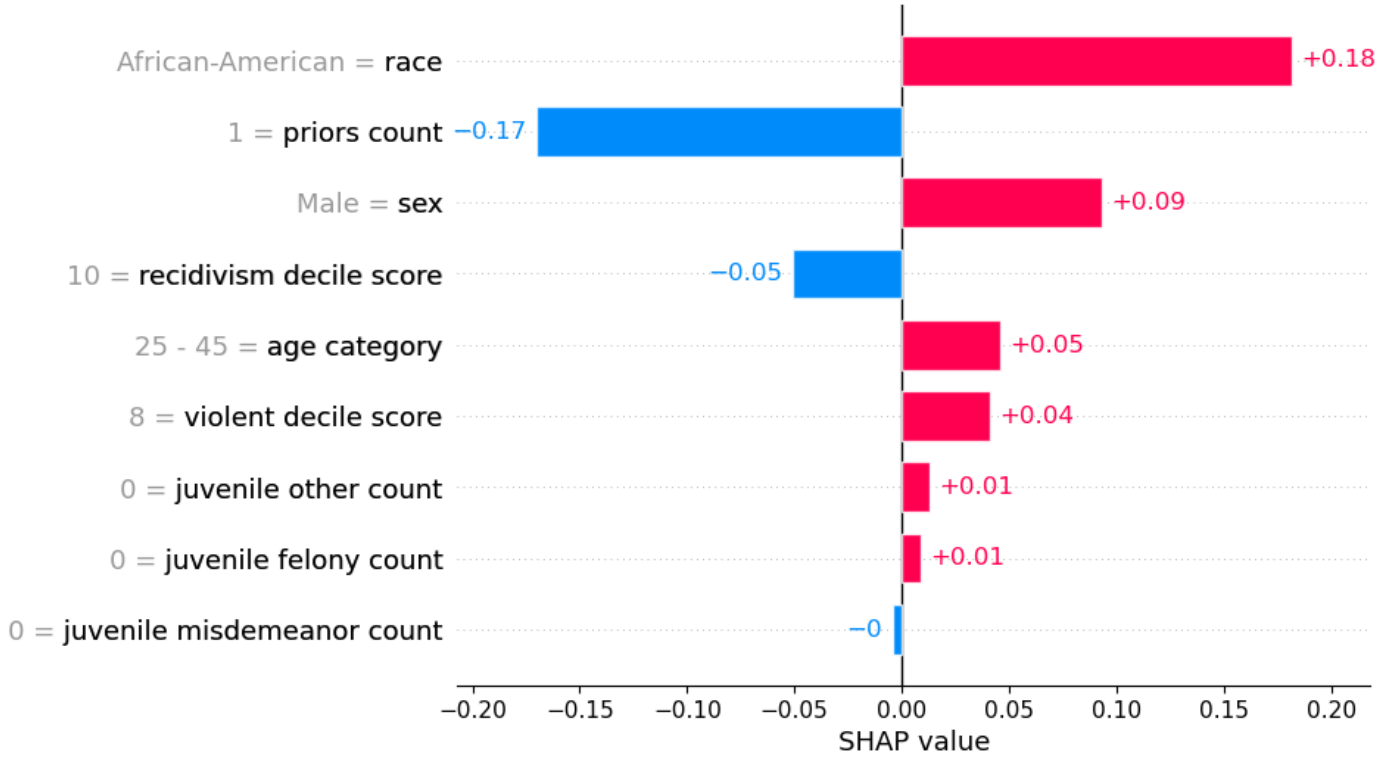
Click "yes" to see the descriptions of the attributes again.

- Yes
- No

How would you categorize this criminal defendant?

- I think that the criminal defendant is not likely to reoffend.
- I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is not likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- No, I do not agree.
- Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	25 - 45
race	Caucasian
priors count	3
recidivism decile score	8
violence decile score	2
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	0

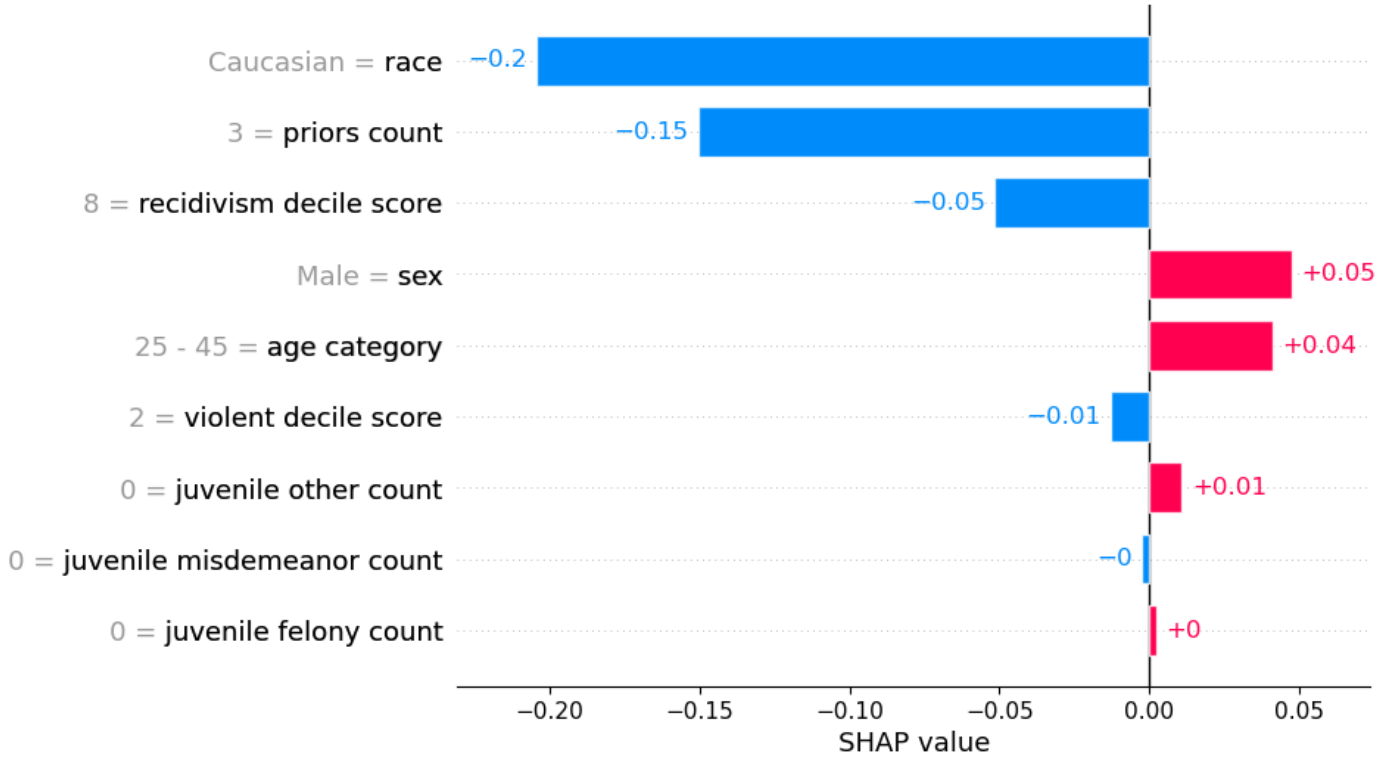
Click "yes" to see the descriptions of the attributes again.

- ○ Yes
- ○ No

How would you categorize this criminal defendant?

- ○ I think that the criminal defendant is not likely to reoffend.
- ○ I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- ○ No, I do not agree.
- ○ Yes, I agree.

If you don't feel familiar enough with the model yet, you can get 2 more examples.

- ○ I need more examples to become familiar with the model.
- ○ I am familiar enough and do not need more examples.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	Less than 25
race	Caucasian
priors count	0
recidivism decile score	9
violence decile score	8
juvenile felony count	0
juvenile misdemeanor count	0

Attributes	Profile
juvenile other count	0

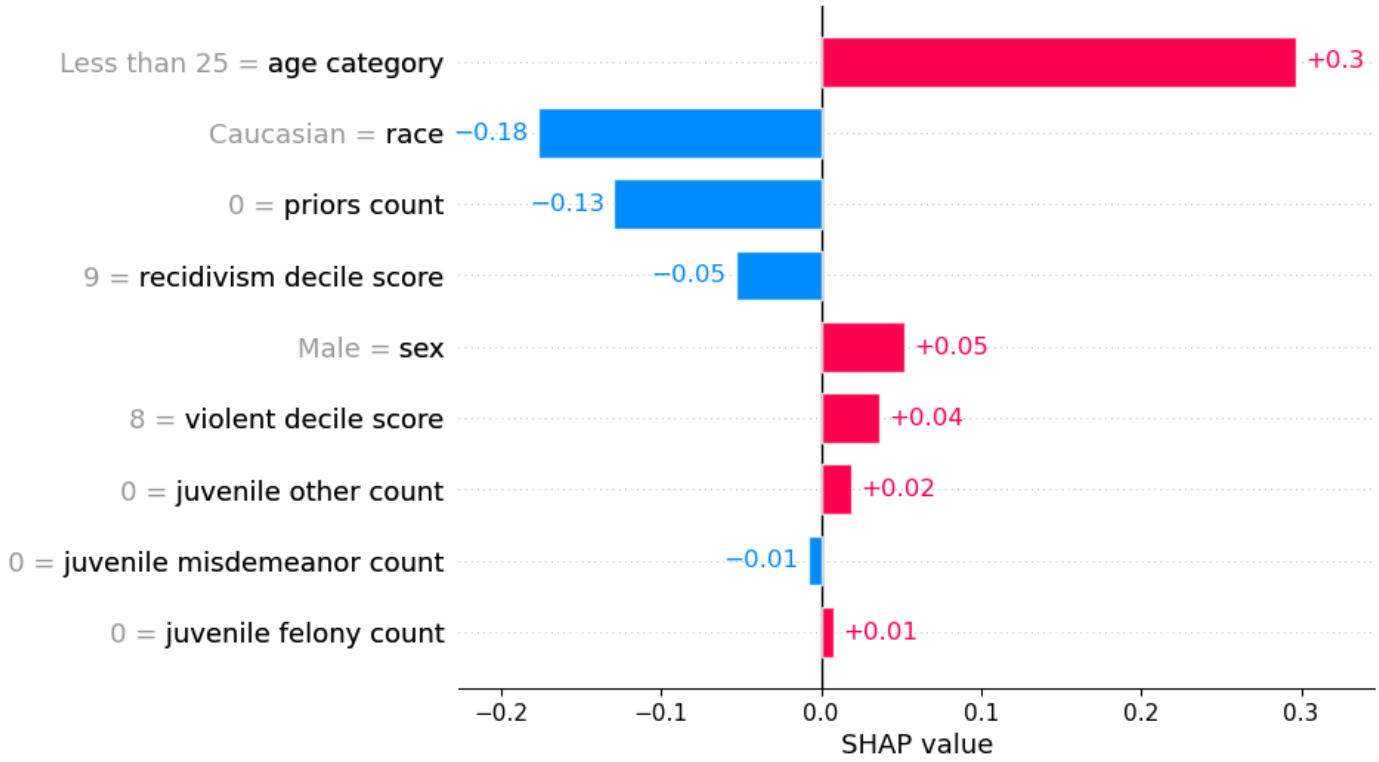
Click "yes" to see the descriptions of the attributes again.

- Yes
- No

How would you categorize this criminal defendant?

- I think that the criminal defendant is not likely to reoffend.
- I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- No, I do not agree.
- Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	Less than 25
race	Caucasian
priors count	0
recidivism decile score	6
violence decile score	9
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	0

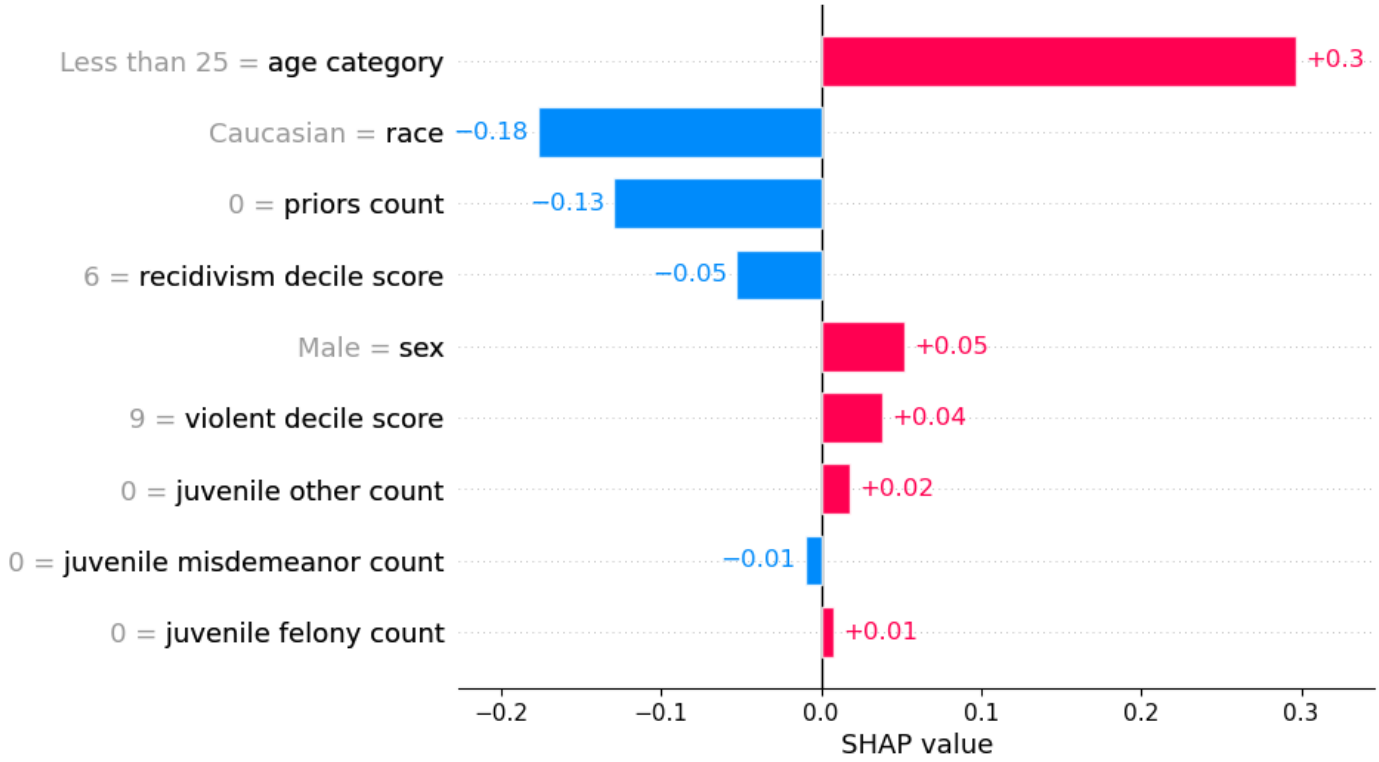
Click "yes" to see the descriptions of the attributes again.

- ○ Yes
- ○ No

How would you categorize this criminal defendant?

- ○ I think that the criminal defendant is not likely to reoffend.
- ○ I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- ○ No, I do not agree.
- ○ Yes, I agree.

Condition 7 - Non-intuitiveness / Faithfulness / Prohibited Features

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Female
age category	25 - 45
race	African-American
priors count	5
recidivism decile score	10
violence decile score	5
juvenile felony count	1
juvenile misdemeanor count	0
juvenile other count	0

Click "yes" to see the descriptions of the attributes again.

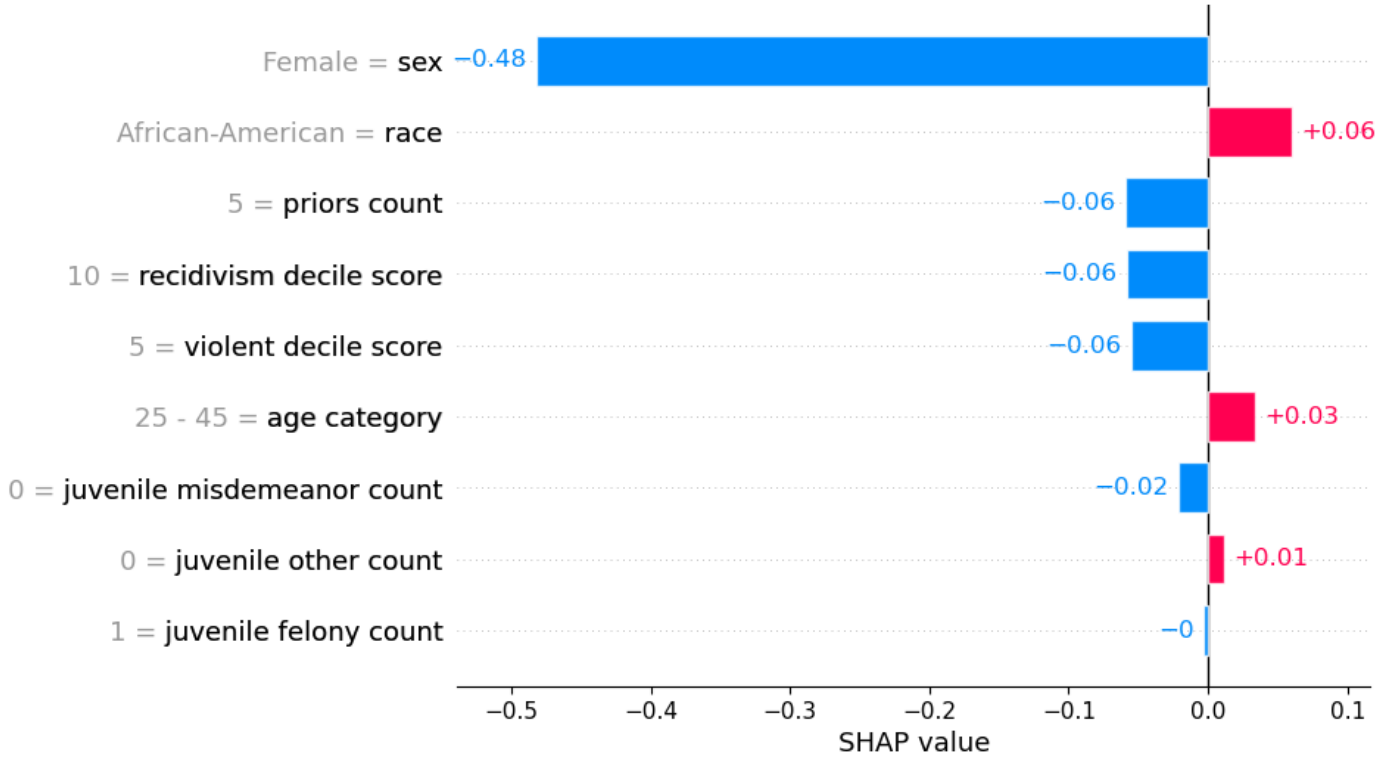
- ○ Yes

- ○ No

How would you categorize this criminal defendant?

- ○ I think that the criminal defendant is not likely to reoffend.
- ○ I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is not likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- ○ No, I do not agree.
- ○ Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	Less than 25
race	Hispanic
priors count	11
recidivism decile score	10

Attributes	Profile
violence decile score	8
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	0

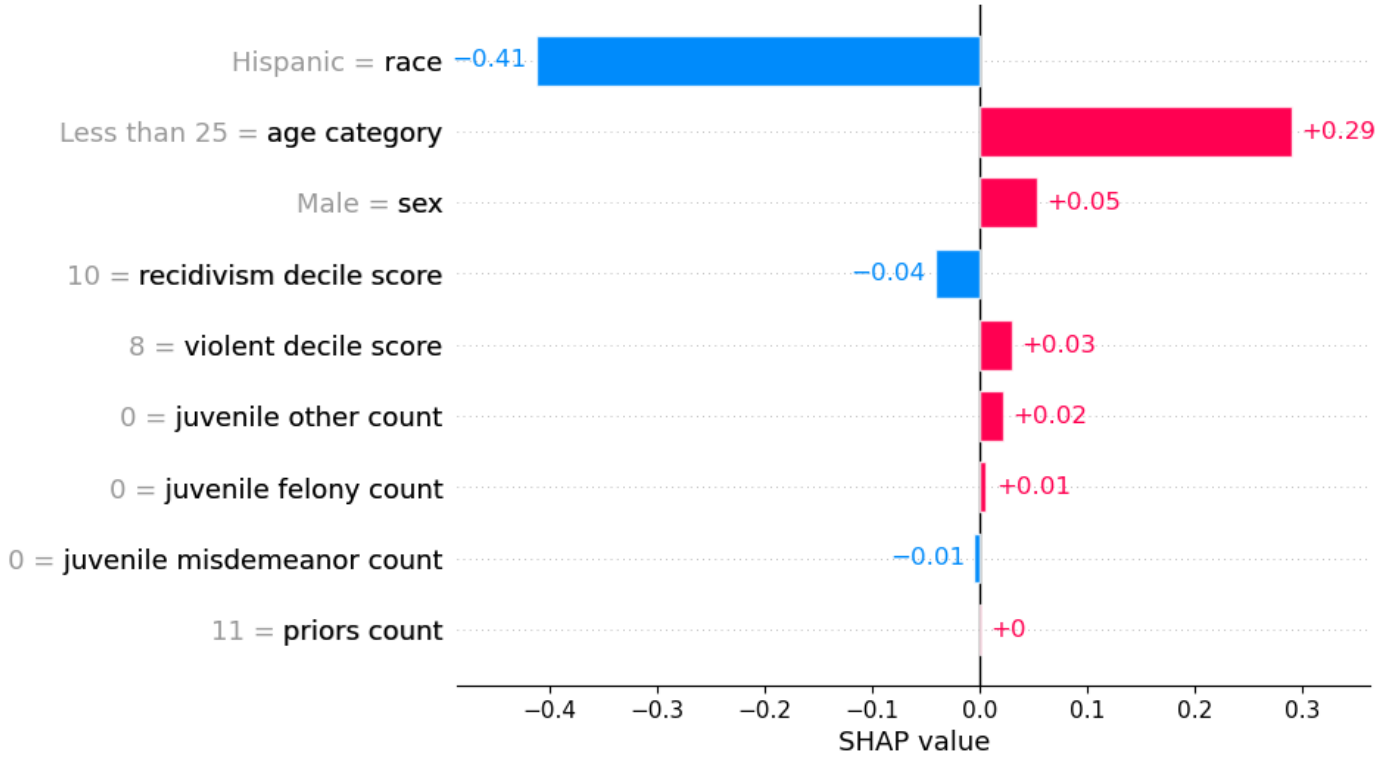
Click "yes" to see the descriptions of the attributes again.

- Yes
- No

How would you categorize this criminal defendant?

- I think that the criminal defendant is not likely to reoffend.
- I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is not likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- No, I do not agree.
- Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	25 - 45
race	Caucasian
priors count	15
recidivism decile score	9
violence decile score	6
juvenile felony count	2
juvenile misdemeanor count	0
juvenile other count	1

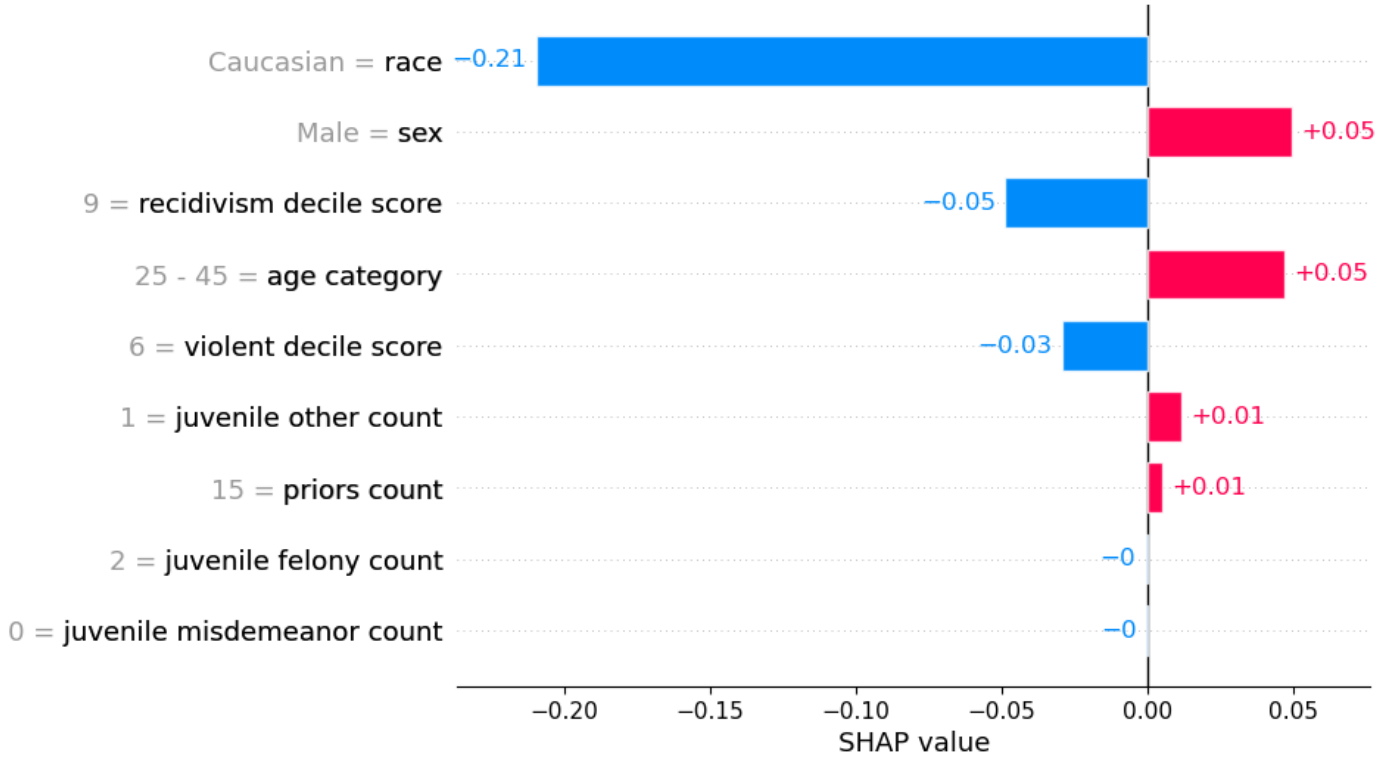
Click "yes" to see the descriptions of the attributes again.

- ○ Yes
- ○ No

How would you categorize this criminal defendant?

- ○ I think that the criminal defendant is not likely to reoffend.
- ○ I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is not likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- ○ No, I do not agree.
- ○ Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	25 - 45
race	African-American
priors count	2
recidivism decile score	1
violence decile score	1
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	0

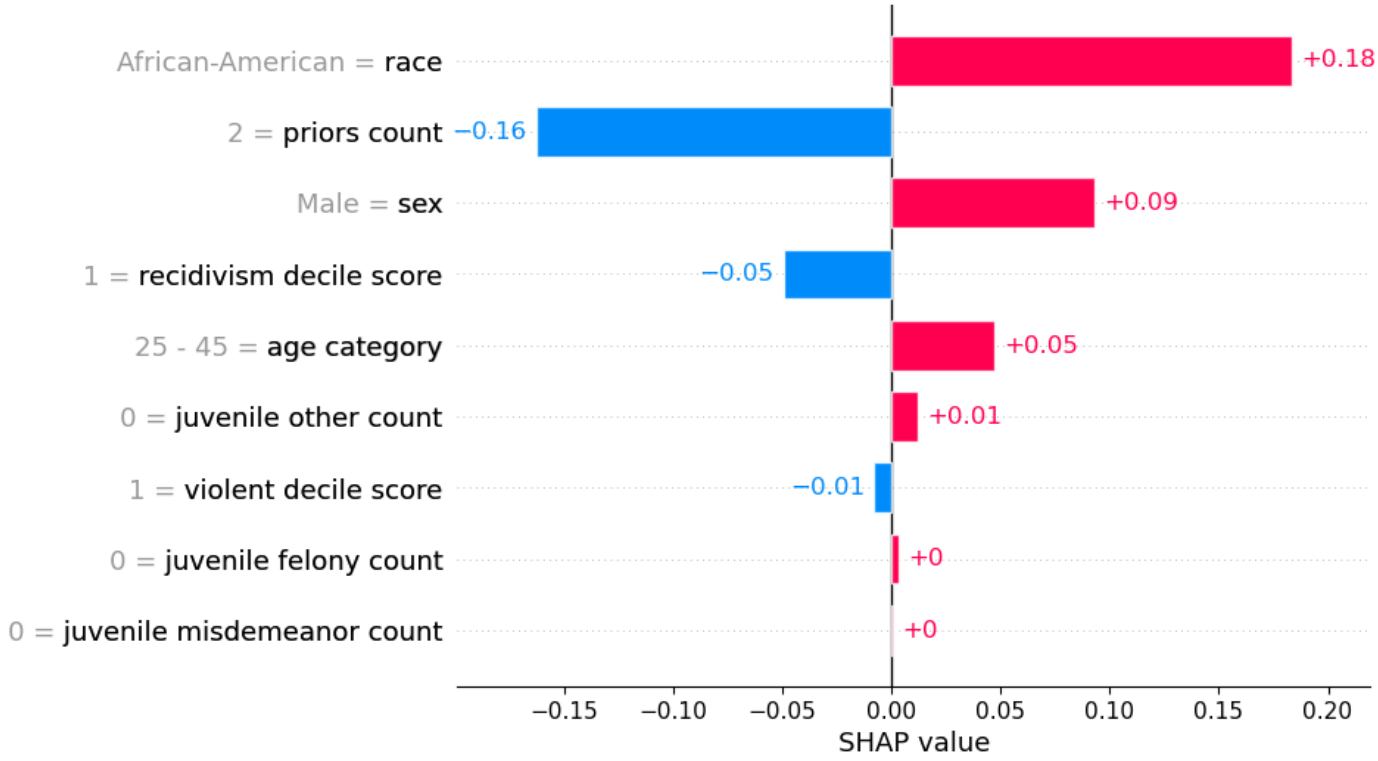
Click "yes" to see the descriptions of the attributes again.

- ○ Yes
- ○ No

How would you categorize this criminal defendant?

- I think that the criminal defendant is not likely to reoffend.
- I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- No, I do not agree.
- Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	25 - 45
race	African-American
priors count	1
recidivism decile score	3

Attributes	Profile
violence decile score	2
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	0

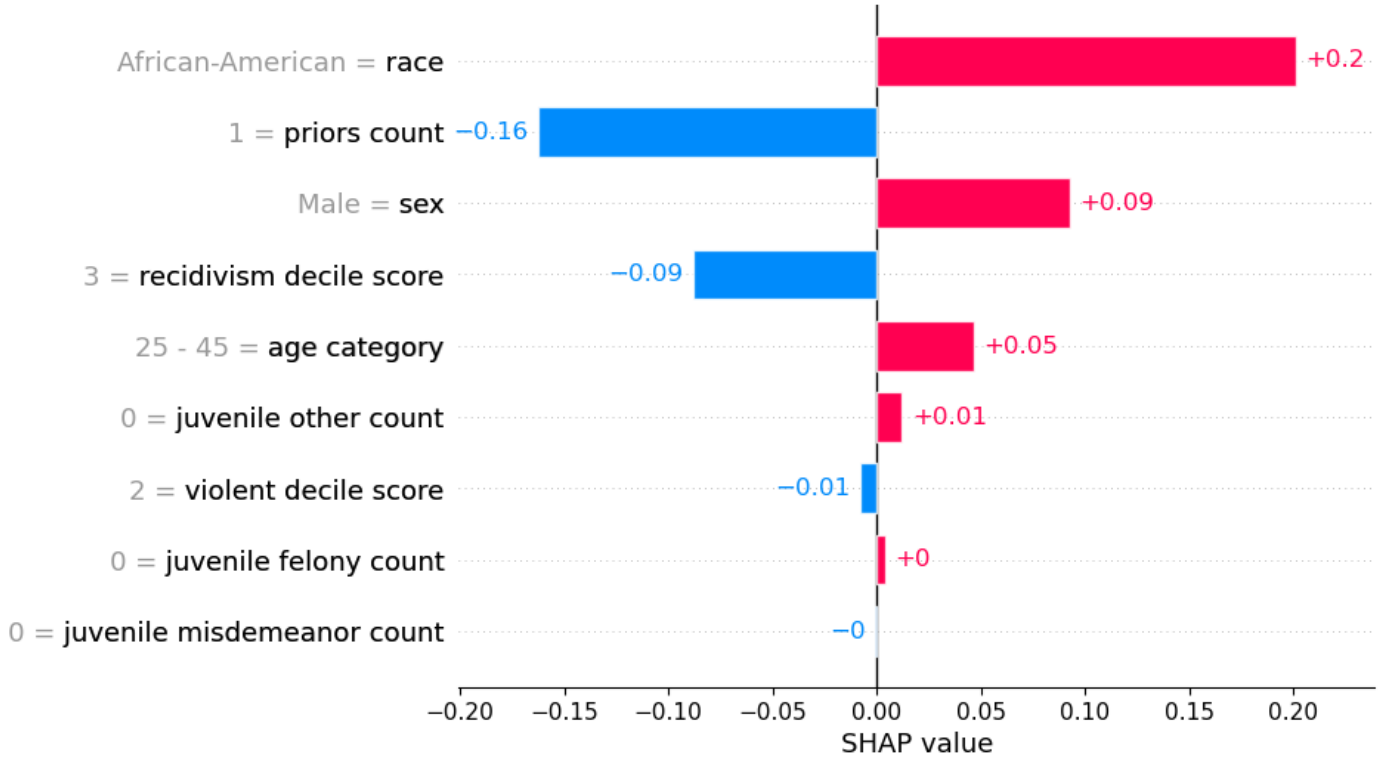
Click "yes" to see the descriptions of the attributes again.

- Yes
- No

How would you categorize this criminal defendant?

- I think that the criminal defendant is not likely to reoffend.
- I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- No, I do not agree.
- Yes, I agree.

If you don't feel familiar enough with the model yet, you can get 2 more examples.

- I need more examples to become familiar with the model.
- I am familiar enough and do not need more examples.

Consider a criminal defendant with the following profile:

Attributes	**Profile**
sex	Male
age category	25 - 45
race	Hispanic
priors count	9
recidivism decile score	9
violence decile score	7
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	0

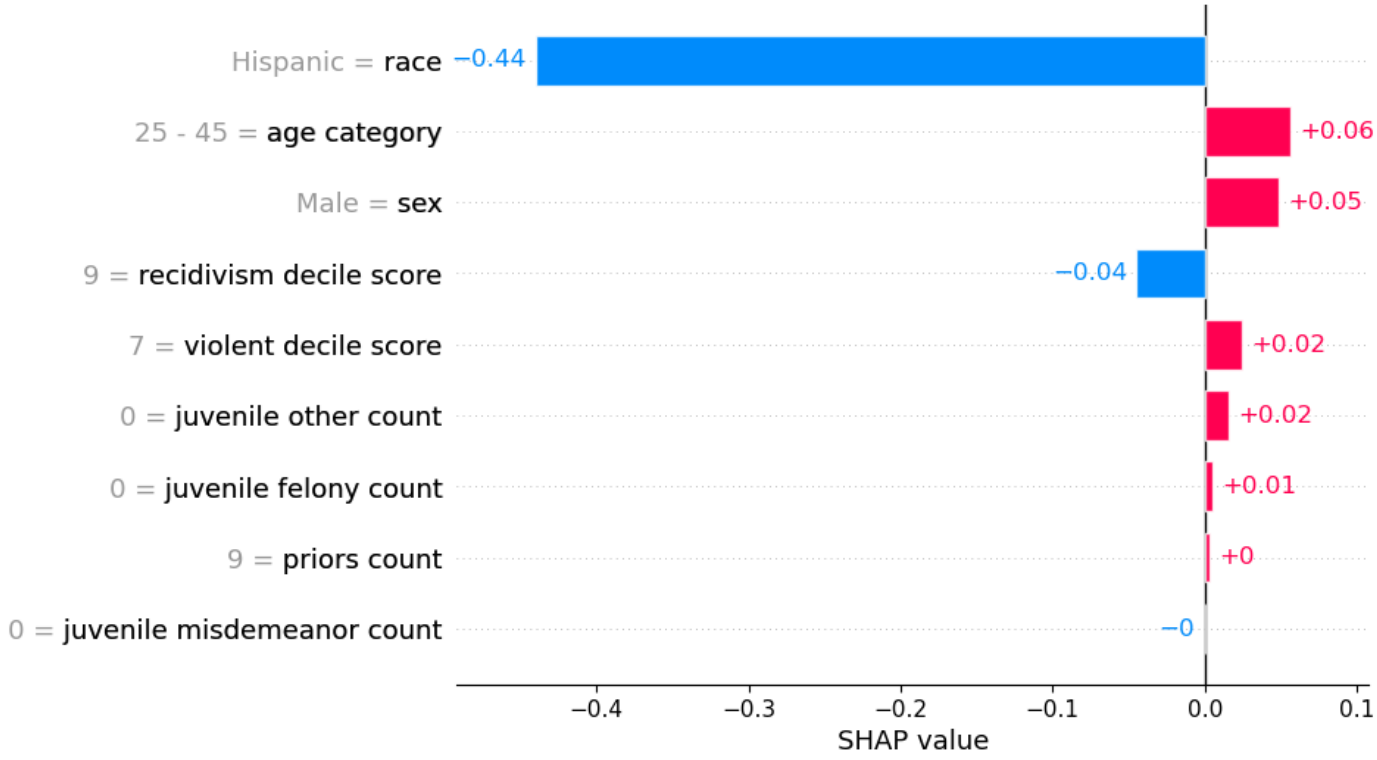
Click "yes" to see the descriptions of the attributes again.

- Yes
- No

How would you categorize this criminal defendant?

- I think that the criminal defendant is not likely to reoffend.
- I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is not likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- ○ No, I do not agree.
- ○ Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	25 - 45
race	Caucasian
priors count	5
recidivism decile score	10
violence decile score	5
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	0

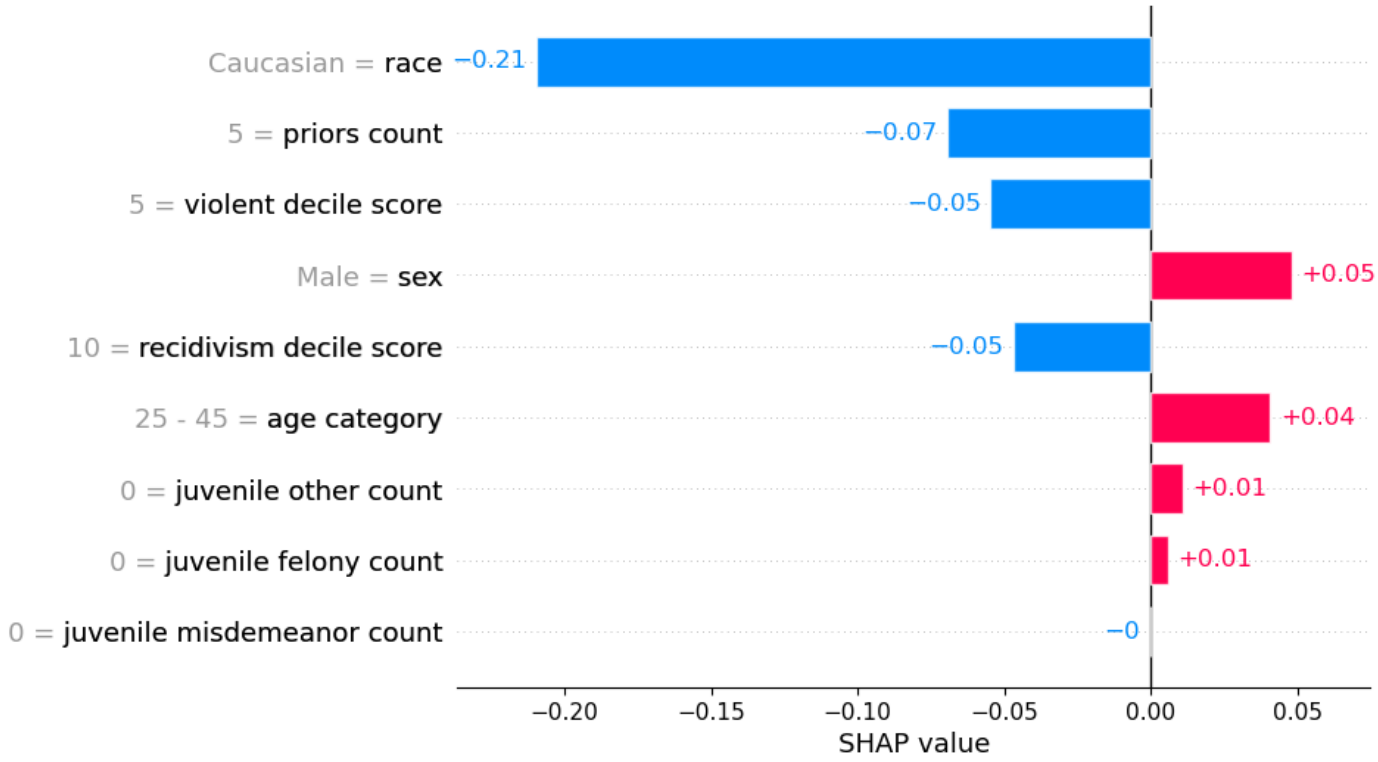
Click "yes" to see the descriptions of the attributes again.

- ○ Yes
- ○ No

How would you categorize this criminal defendant?

- I think that the criminal defendant is not likely to reoffend.
- I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is not likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- No, I do not agree.
- Yes, I agree.

Condition 8 - Non-intuitiveness / Unfaithfulness / Prohibited Features

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Female
age category	25 - 45
race	African-American
priors count	5

Attributes	Profile
recidivism decile score	10
violence decile score	5
juvenile felony count	1
juvenile misdemeanor count	0
juvenile other count	0

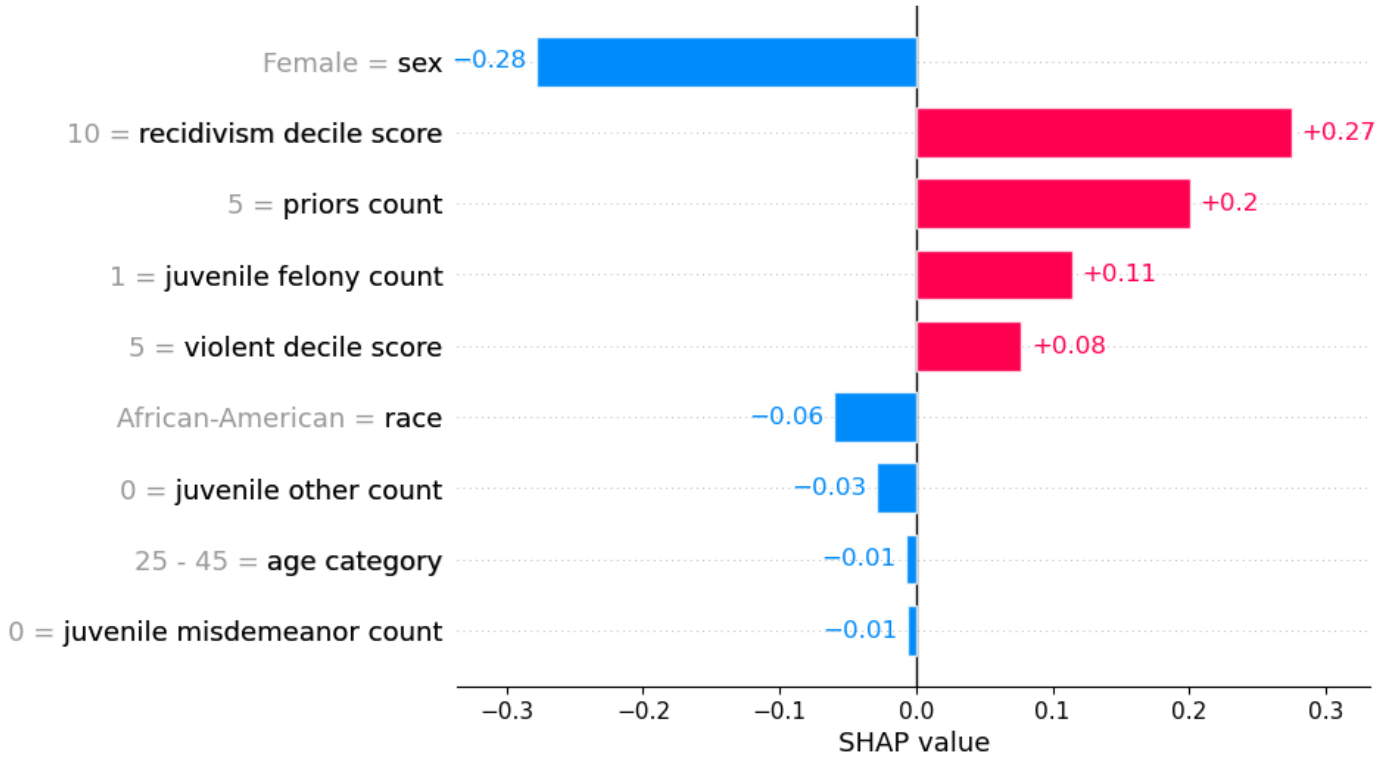
Click "yes" to see the descriptions of the attributes again.

- Yes
- No

How would you categorize this criminal defendant?

- I think that the criminal defendant is not likely to reoffend.
- I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is not likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- No, I do not agree.

- ○ Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	Less than 25
race	Hispanic
priors count	11
recidivism decile score	10
violence decile score	8
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	0

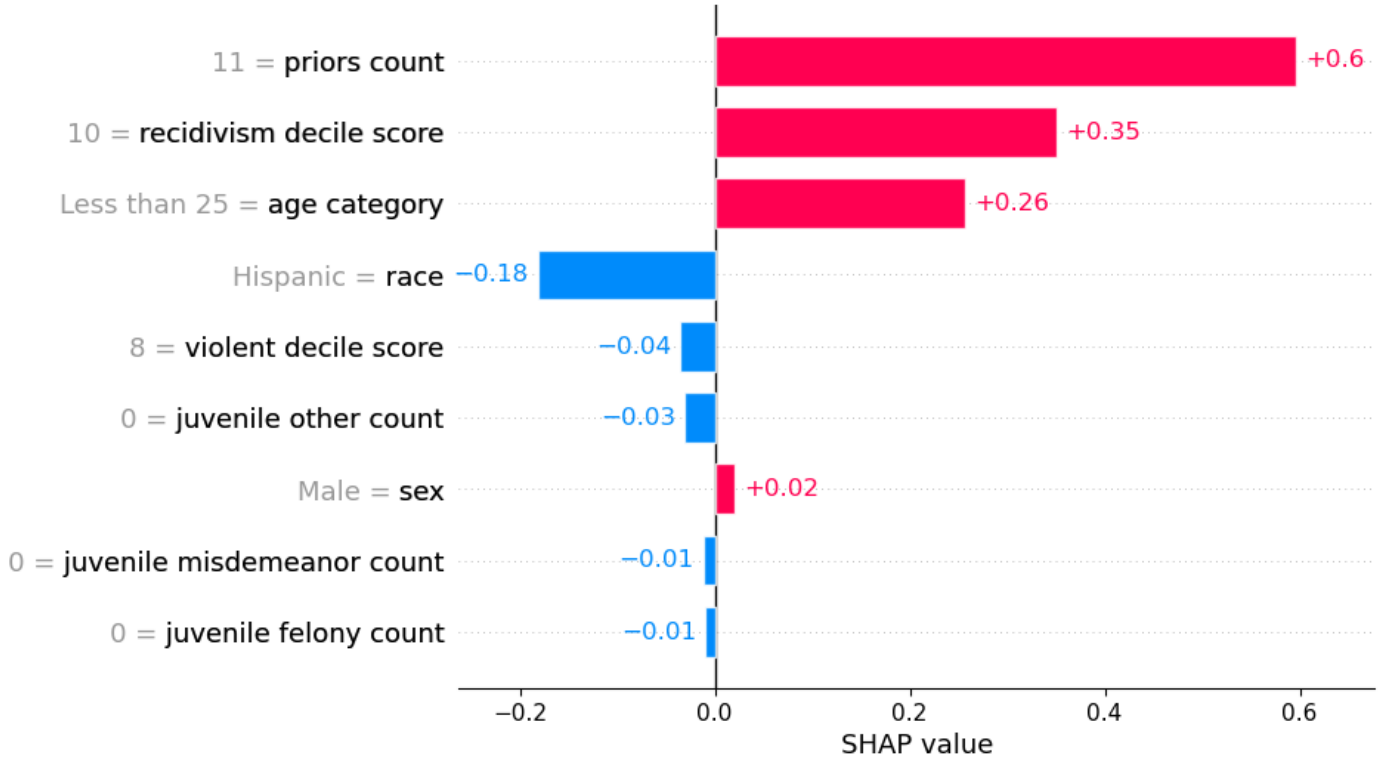
Click "yes" to see the descriptions of the attributes again.

- ○ Yes
- ○ No

How would you categorize this criminal defendant?

- ○ I think that the criminal defendant is not likely to reoffend.
- ○ I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is not likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- ○ No, I do not agree.
- ○ Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	25 - 45
race	Caucasian
priors count	15
recidivism decile score	9
violence decile score	6
juvenile felony count	2
juvenile misdemeanor count	0
juvenile other count	1

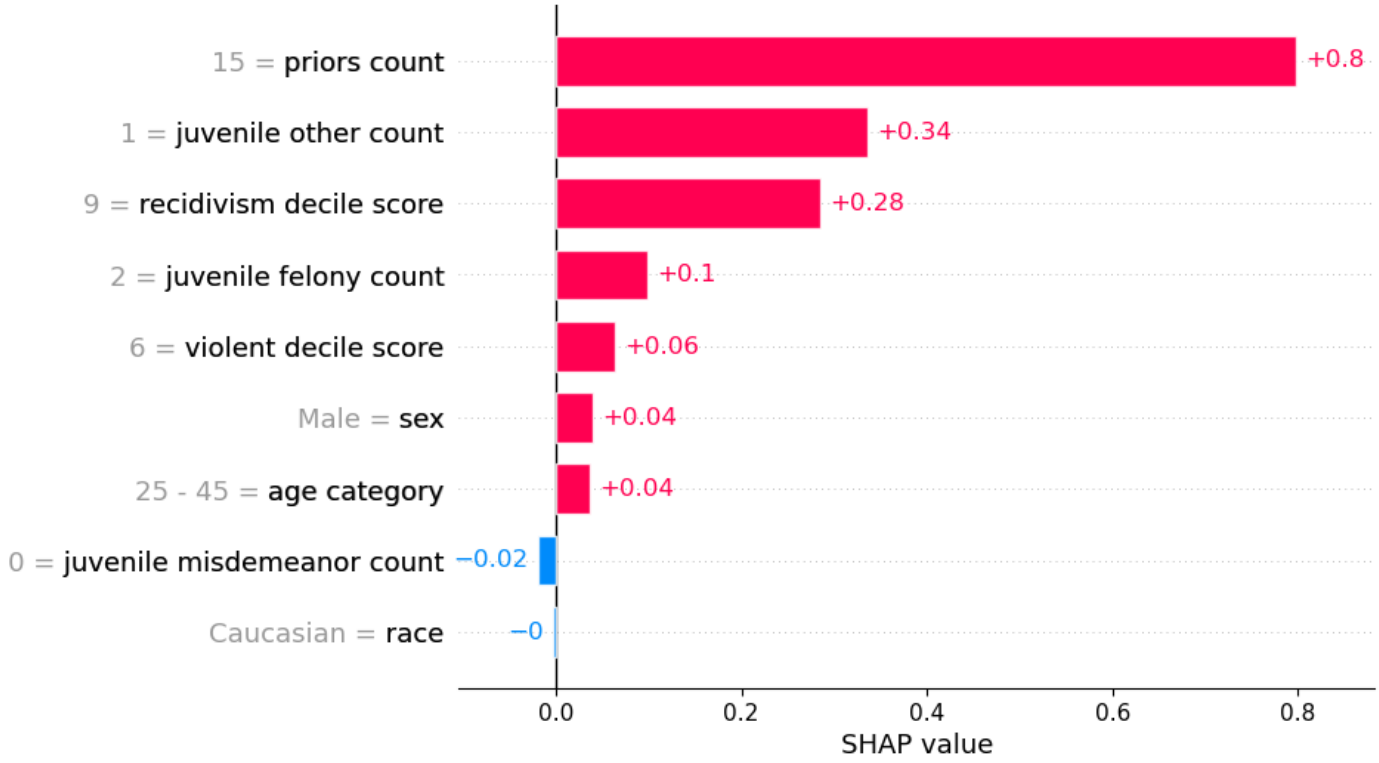
Click "yes" to see the descriptions of the attributes again.

- ○ Yes
- ○ No

How would you categorize this criminal defendant?

- I think that the criminal defendant is not likely to reoffend.
- I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is not likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- No, I do not agree.
- Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	25 - 45
race	African-American
priors count	2
recidivism decile score	1

Attributes	Profile
violence decile score	1
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	0

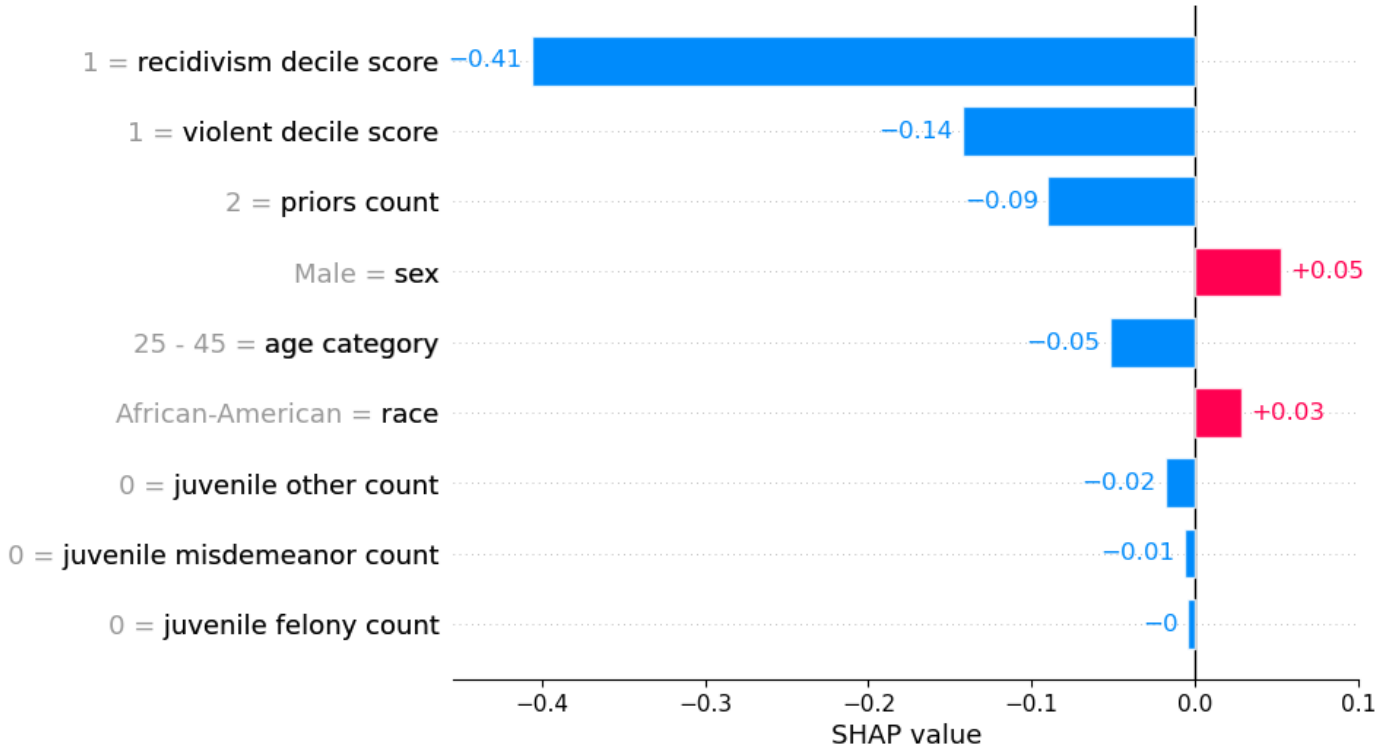
Click "yes" to see the descriptions of the attributes again.

- Yes
- No

How would you categorize this criminal defendant?

- I think that the criminal defendant is not likely to reoffend.
- I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- No, I do not agree.
- Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	25 - 45
race	African-American
priors count	1
recidivism decile score	3
violence decile score	2
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	0

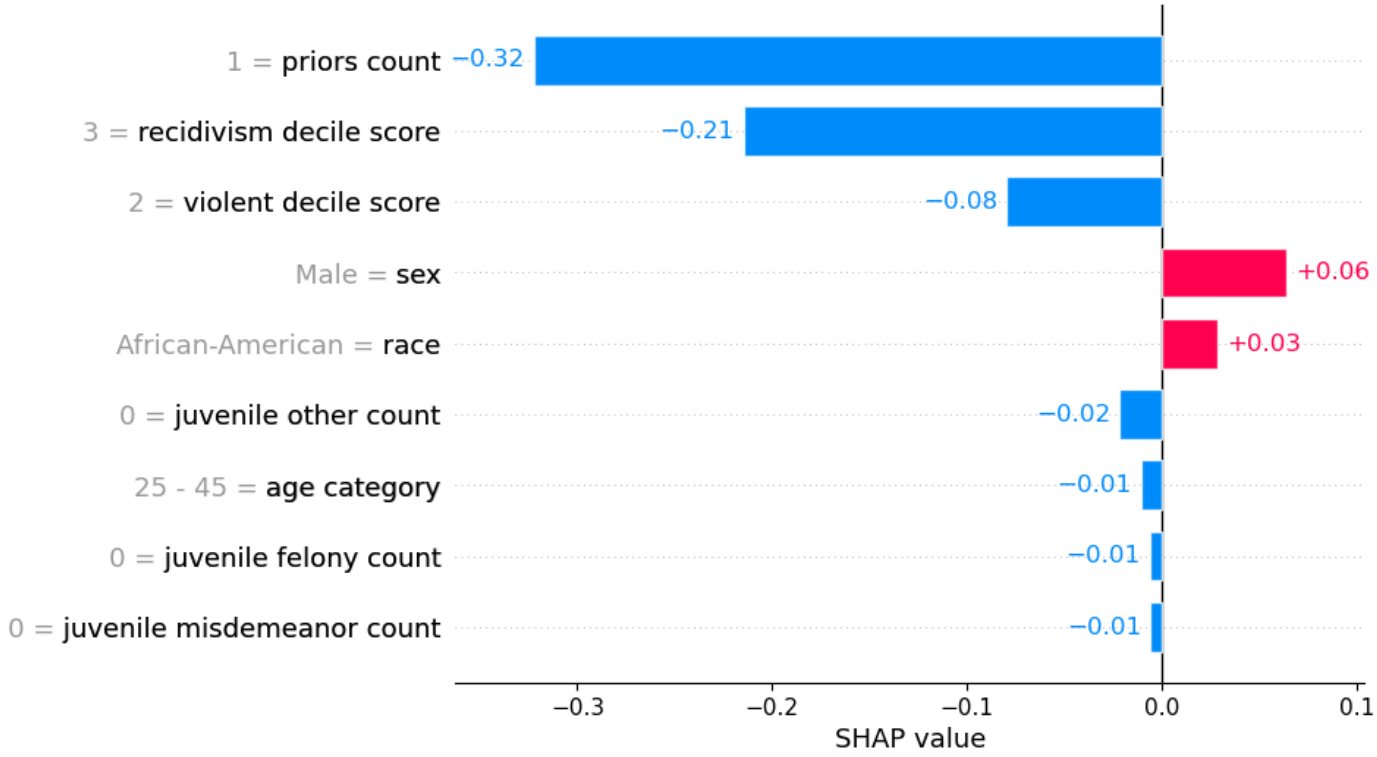
Click "yes" to see the descriptions of the attributes again.

- ○ Yes
- ○ No

How would you categorize this criminal defendant?

- ○ I think that the criminal defendant is not likely to reoffend.
- ○ I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- ○ No, I do not agree.
- ○ Yes, I agree.

If you don't feel familiar enough with the model yet, you can get 2 more examples.

- ○ I need more examples to become familiar with the model.
- ○ I am familiar enough and do not need more examples.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	25 - 45
race	Hispanic
priors count	9
recidivism decile score	9
violence decile score	7
juvenile felony count	0
juvenile misdemeanor count	0

Attributes	Profile
juvenile other count	0

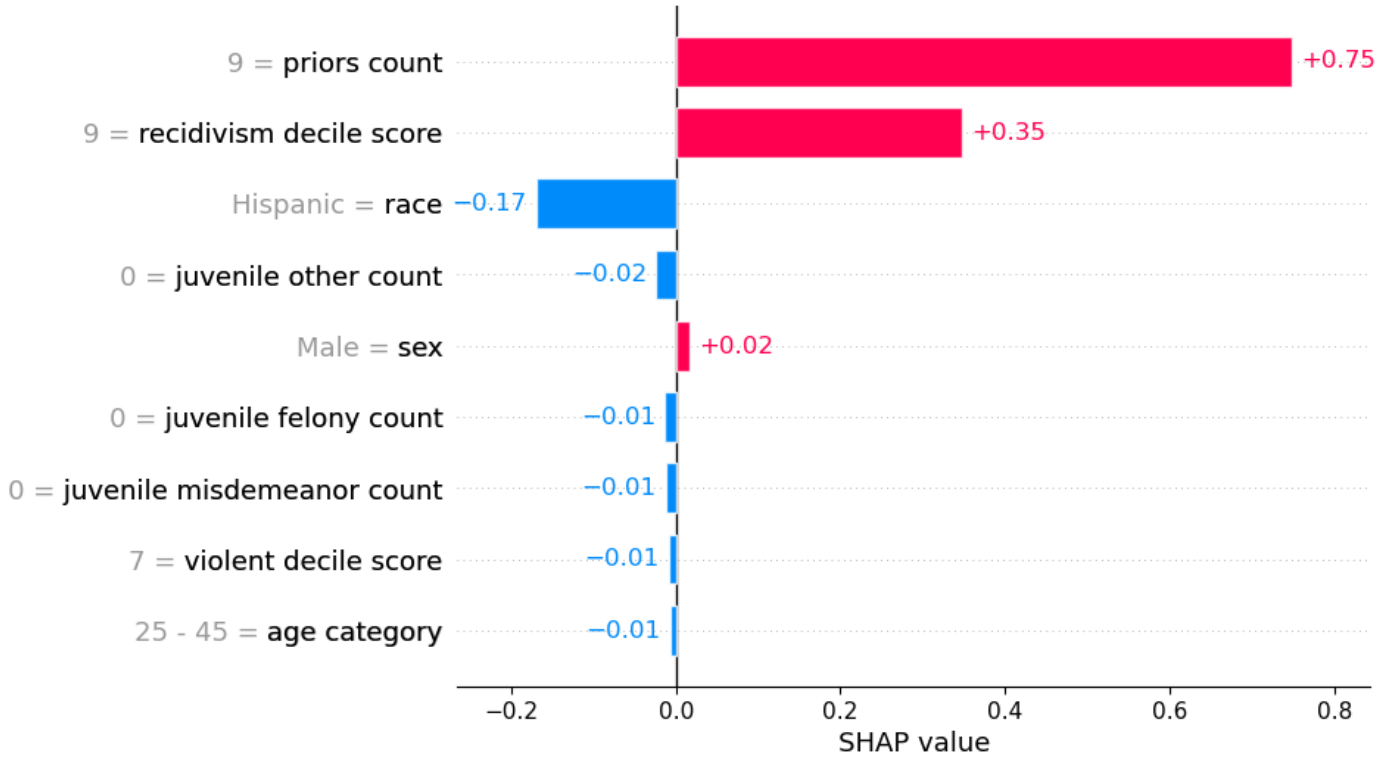
Click "yes" to see the descriptions of the attributes again.

- Yes
- No

How would you categorize this criminal defendant?

- I think that the criminal defendant is not likely to reoffend.
- I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is not likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- No, I do not agree.
- Yes, I agree.

Consider a criminal defendant with the following profile:

Attributes	Profile
sex	Male
age category	25 - 45
race	Caucasian
priors count	5
recidivism decile score	10
violence decile score	5
juvenile felony count	0
juvenile misdemeanor count	0
juvenile other count	0

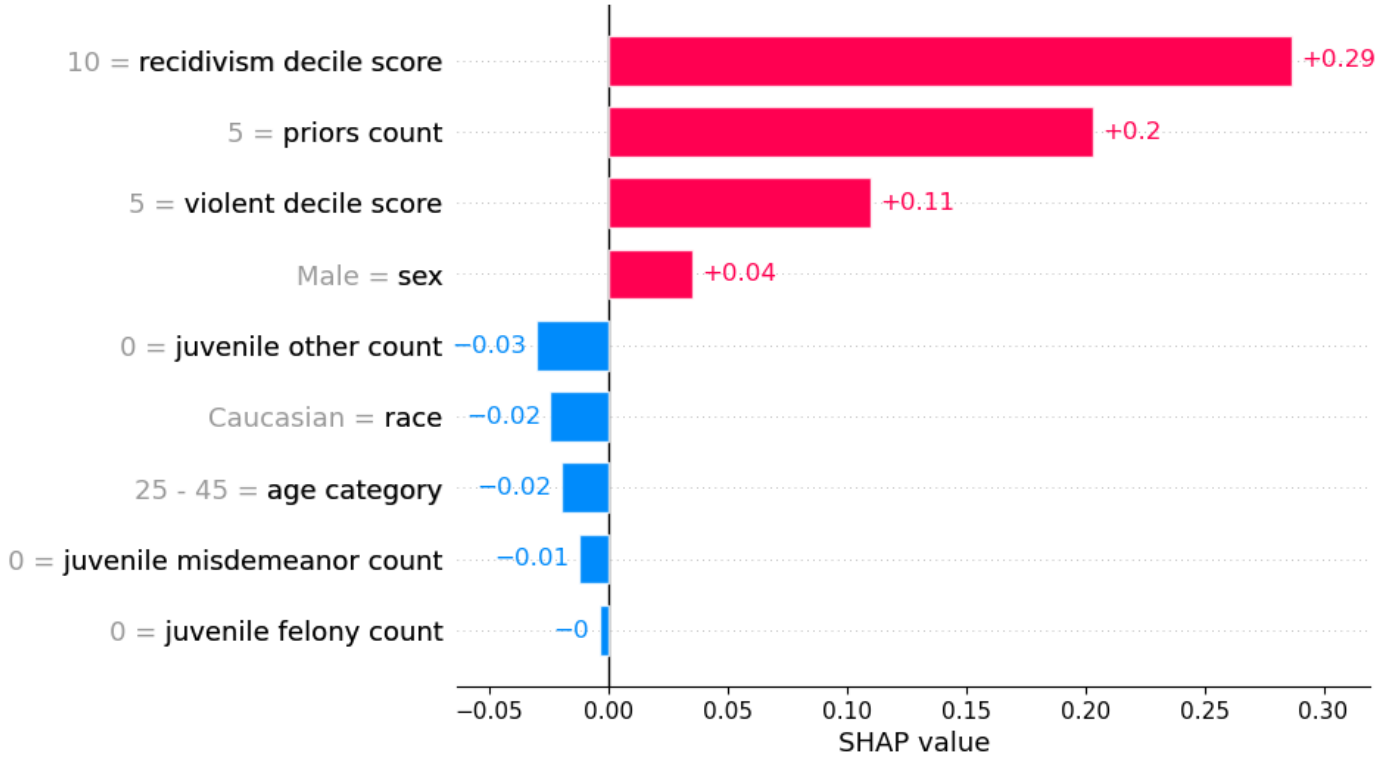
Click "yes" to see the descriptions of the attributes again.

- ○ Yes
- ○ No

How would you categorize this criminal defendant?

- ○ I think that the criminal defendant is not likely to reoffend.
- ○ I think that the criminal defendant is likely to reoffend.

The model would decide that the criminal defendant **is not likely to reoffend**. Below is an explanation generated by the machine learning model to explain its decision.



Based on this explanation, do you agree with the decision of the model?

- ○ No, I do not agree.
- ○ Yes, I agree.

Trust Assessment

Based on the previous predictions and the corresponding explanations of the model, please indicate how much you agree with the following statements.

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
I can predict how the model will behave.	○	○	○	○	○	○	○
I trust the decisions made by the model.	○	○	○	○	○	○	○
I have faith that the model would be able to cope with all different criminal defendant situations.	○	○	○	○	○	○	○

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
If I am not sure about a decision, I trust that the model will provide the best solution.	○	○	○	○	○	○	○
I trust the model to provide a reliable decision for criminal recidivism.	○	○	○	○	○	○	○

Overall, why do you trust (or distrust) the decisions made by the model?

Which of the provided attributes do you trust least?

- ○ sex
- ○ age category
- ○ race
- ○ priors count
- ○ recidivism decile score
- ○ violence decile score
- ○ juvenile felony count
- ○ juvenile misdemeanor count
- ○ juvenile other count

Which of the provided attributes do you trust most?

- ○ sex
- ○ age category
- ○ race
- ○ priors count
- ○ recidivism decile score
- ○ violence decile score
- ○ juvenile felony count
- ○ juvenile misdemeanor count
- ○ juvenile other count

Demographic Questions

Please state your age.

Choose one gender that you consider yourself to be

- ○ Male
- ○ Female
- ○ Non-binary / third gender
- ○ Prefer to self-describe
- ○ Prefer not to say

Please click the button below to submit your responses.