

## Domain Learning

### Domain Learning

**What:** In this section, you are trying to acquire knowledge and identify patterns in the domain of recidivism (the tendency of a convicted criminal to re-offend) risk using the information provided by the AI system. The questions in this section test if you are able to gain a comprehensive understanding of how certain variables and features contribute to the prediction of recidivism risk. The idea here is that if you are able to use the information provided by the AI system to extract insights about how features contribute to AI predictions, you can apply this knowledge to obtain favorable outcomes.

\* show me the prediction probabilities of 7077

The model predicts the instance with id equal to 7077 as:

- likely to commit a crime with 46.62% probability
- unlikely to commit a crime with 53.38% probability

[Feedback](#)

What is the model's prediction for defendant 7077?

💡 Tip: the model prediction for that instance is "likely to commit a crime" if its probability is above 50%, and consecutively that of "unlikely to commit a crime" is less than 50%.

① Please select one answer

Likely to commit a crime.

Unlikely to commit a crime.

\* Show me the feature importance for the model prediction for defendant 7077

For the model's predictions on instance with id equal to 7077, the importance of the features have the following ranking, where 1 is the most important feature:

- 1: age
- 2: numberofpriorcrimes
- 3: recidivated
- 4: monthsinjail
- 5: black
- 6: man
- 7: felony

What are the three most important features that impact the model's prediction for defendant 7077?

💡 Tip: Recidivated is a binary feature that indicates whether the defendant is rearrested at any time.

① Please select one answer

Recidivated, number of prior crimes, and black.

Age, number of prior crimes, and black.

Number of prior crimes, recidivated, and age.

\* show me the data for 7077

For the data with id equal to 7077, the features are

age: 31  
recidivated: 1

how can I flip the model's prediction for applicant 7077?

numberofpriorcrimes: 8  
monthsinjail: 0  
felony: 1  
misdemeanor: 0  
woman: 0  
man: 1

show me the prediction probabilities of 7077

The model predicts the instance with **Id equal to 7077** as:

- **likely to commit a crime** with **46.62%** probability
- **unlikely to commit a crime** with **53.38%** probability

For instances where **id equal to 7077**, the original prediction is *unlikely to commit a crime*. Here are some options to change the prediction of this instance.

First, if you *decrease recidivated* to 0.0 and *increase numberofpriorcrimes* to 23, the model will predict likely to commit a crime.

Further, if you *increase monthsinjail* to 243, the model will predict likely to commit a crime.

If we want to flip the prediction for 7077, should the 'number of previous crimes' be increased or decreased?

① Please select one answer

Increased.

Decreased.

\* TalkToModel

what is the overall accuracy of the model

The model scores 85.205% accuracy on the all the data.

what is the accuracy of the model's predictions for data points where black is true

The model scores 79.004% accuracy on the data where black equal to true.

what is the accuracy of the model's predictions for data points where black is false

The model scores 91.379% accuracy on the data where black equal to false.

Given the accuracy for the subgroups of data with the feature value of black as true and false, which group of model predictions is more likely to be correct?

① Please select one answer

Subgroup with feature value of black as false.

Subgroup with feature value of black as true.

\* how many incorrect predictions does the model make for defendants with age below 25?

For the data with **age less than 25.0**, the model is incorrect 47 out of 189 times (error rate 0.249). Here are the ids of instances the model predicts incorrectly:

[ 9914 4911 6250 1976 2819 7200 2448 338 807 4144  
5531 977 7496 4406 9287 10619 1897 10571 500 1482  
737 2228 2006 1092 5485 6605 6815 8633 7393 5389  
10158 3566 10237 3408 718 4862 3085 5069 4939 485  
1298 3865 3719 4971 10505 8885 396]

Feedback

Select the defendant ID with an age lesser than 25 years that the model predicted incorrectly (the screenshot shows the complete list).

① Please select one answer

2007

7077

4939

\* how many incorrect predictions does the model make for defendants with age below 25?

For the data with **age less than 25.0**, the model is incorrect 47 out of 189 times (error rate 0.249). Here are the ids of instances the model predicts incorrectly:

```
[ 9914 4911 6250 1976 2819 7200 2448 338 807 4144  
5531 977 7496 4406 9287 10619 1897 10571 500 1482  
737 2228 2006 1092 5485 6605 6815 8633 7393 5389  
10158 3566 10237 3408 718 4862 3085 5069 4939 485  
1298 3865 3719 4971 10505 8885 396]
```

Feedback 

How likely is the model prediction correct about a defendant who is 22 years old?

① Please select one answer

Very likely (more than 75%).

Very unlikely (less than 25%).

\* What are the three most important features for determining whether those who are women are likely to commit a crime?

For the model's predictions on instance with **woman** equal to true, the top 3 most important features are as follows, where 1 is the most important feature:

- 1: age
- 2: numberofpriorcrimes
- 3: recidivated

What is the most important feature of the model's prediction for whether women are likely or unlikely to commit a crime?

① Please select one answer

Age.

Recidivated.

Number of prior crimes.

\* TalkToModel

Hello 😊, I'm a machine learning model trained to predict whether someone will commit a crime.

Let's get started. Ask me something!

What is the likelihood that 4679 likely to commit a crime according to the model?

The model predicts the instance with **id** equal to 4679 as:

- likely to commit a crime with 1.357% probability
- unlikely to commit a crime with 98.643% probability

Feedback 

How certain is the model about its prediction for defendant 4679 being *unlikely to commit a crime*?

① Please select one answer

Very certain (more than 75%).

Very uncertain (less than 25%).

## Decision Support

### Decision Support

**What:** In this section, you are trying to make decisions with the help of the Recidivism (the tendency of a convicted criminal to re-offend) risk prediction AI. The questions in the section assess if you understand the reasons behind the AI system's predictions and when you should exercise caution while considering the AI predictions for your decision. The idea here is that the AI system should communicate the uncertainty and limitations of the predictions. It should give you easy-to-understand information that is actionable for making decisions.

\* **show me the prediction probabilities for sample 6336**

The model predicts the instance with id equal to 6336 as:

- likely to commit a crime with 25.806% probability
- unlikely to commit a crime with 74.194% probability

**show me the data for 6336**

For the data with id equal to 6336, the features are

```
age: 40
recidivated: 1
numberofpriorcrimes: 5
monthsinjail: 55
felony: 1
misdemeanor: 0
woman: 0
```

TalkToModel

**what are the categories of data where the model typically gives incorrect predictions**

For all the instances in the data, the model typically predicts incorrect:

```
if (numberofpriorcrimes > 4.5) and (age <= 34.5) and
(monthsinjail <= 138.0) and (black > 0.5) and (felony <= 0.5)
then then the model is incorrect 68.42% over 19 samples
```

```
if (numberofpriorcrimes <= 4.5) and (age <= 24.5) and
(monthsinjail > 4.5) and (monthsinjail > 7.5) and
(monthsinjail > 46.5) then then the model is incorrect 66.67%
over 9 samples
```

```
if (numberofpriorcrimes > 4.5) and (age > 34.5) and
(monthsinjail > 7.5) and (monthsinjail <= 65.5) and
(monthsinjail > 38.5) then then the model is incorrect 87.5%
over 6 samples
```

Based on the profile information, prediction probabilities provided for defendant 6336, and the category of data in which the model typically makes mistakes, how likely is the model incorrect?

**Please select one answer**

Very likely (more than 75%).

Very unlikely (more than 25%).

\* **show me some instances where the model prediction were incorrect and the applicant is black**

For the data with the model predicts **unlikely to commit a crime and black equal to true**, the model is incorrect 76 out of 390 times (error rate 0.195). Here are the ids of instances the model predicts incorrectly:

```
[ 1595 228 5554 478 2982 4911 7942
6250 10646 685 7077 1915 2819 9810
2448 4933 8203 4033 338 2150 4610
4144 4861 10519 977 1463 7689 10614
6185 9225 9441 8434 10619 500 4101]
```

From the list in the screenshot above select the instance ID where the model prediction was incorrect.

**Please select one answer**

222

2819

6188

**\* what are the categories of data where the model typically gives incorrect predictions**

For all the instances in the data, the model typically predicts incorrect.

if (numberofpriorcrimes <= 4.5) and (age <= 24.5) and (monthsinjail > 4.5) and (monthsinjail <= 7.5) and (numberofpriorcrimes <= 1.5) then the model is incorrect 100.0% over 5 samples

if (numberofpriorcrimes > 4.5) and (age > 34.5) and (monthsinjail > 7.5) and (monthsinjail <= 65.5) and (monthsinjail > 38.5) then the model is incorrect 87.5% over 8 samples

**Show me the data for applicant 6336**

For the data with id equal to 6336, the features are

age: 40  
recidivated: 1  
numberofpriorcrimes: 5  
monthsinjail: 55

**Show me the data for applicant 4406**

For the data with id equal to 4406, the features are

age: 21  
recidivated: 0  
numberofpriorcrimes: 0  
monthsinjail: 7  
felony: 1

Considering the categories (in the screenshot on the left) of data for defendants 4406 (category 1) and 6336 (category 2), which model prediction is more likely to be incorrect: the one for defendant 6336 or defendant 4406?

**💡 Tip:** Take the model's error rate into account corresponding to the respective categories 6336 and 4406 fall under.

**Please select one answer**

6336

4406

**\* what are the prediction probabilities for defendant 807?**

The model predicts the instance with id equal to 807 as:

- likely to commit a crime with 5.843% probability
- unlikely to commit a crime with 94.157% probability

**Feedback**

What is the model's prediction for defendant 807?

**💡 Tip:** the model prediction for that instance is "likely to commit a crime" if its probability is above 50%, and consecutively that of "unlikely to commit a crime" is less than 50%.

**Please select one answer**

Likely to commit a crime.

Unlikely to commit a crime.

**\* what are the 5 most important features for machine predictions for 807?**

For the model's predictions on instance with id equal to 807, the top 5 most important features are as follows, where 1 is the most important feature:

- age
- numberofpriorcrimes
- monthsinjail
- recidivated
- black

**what are the prediction probabilities for defendant 807?**

The model predicts the instance with id equal to 807 as:

- likely to commit a crime with 5.843% probability
- unlikely to commit a crime with 94.157% probability

**Feedback**

Which attribute impacts the model's prediction for defendant 807 more, the *time spent in jail*, or the *number of previous crimes committed*?

**Please select one answer**

Time spent in jail has more impact.

Number of previous crimes has more impact.

**\* what are the prediction probabilities for 807?**

The model predicts the instance with id equal to 807 as:

- likely to commit a crime with 5.843% probability
- unlikely to commit a crime with 94.157% probability

**what are the prediction probabilities for 807 if number of**

prior crimes was raised to 5?

The model predicts the instance with id equal to 807 and numberofpriorcrimes is increased by 5.0 as:

- likely to commit a crime with 30.71% probability
- unlikely to commit a crime with 69.29% probability

Feedback 

How does increasing the value of the *number of prior crimes* impact the model-predicted probability of being *unlikely to commit a crime* for defendant 807?

 Please select one answer

- Increases the probability of being unlikely to commit
- Decreases the probability of being unlikely to commit

## Capability Assessment

### Capability Assessment

**What:** You were just introduced to the Recidivism (the tendency of a convicted criminal to re-offend) risk prediction AI. Your task is to explore the AI system's functionality and assess its applicability to your role if you were a judge. The questions in this section will be investigating your perception of the system's predictive abilities, the reasons prompting those predictions, and the specific scenarios or conditions where the system may produce incorrect predictions. The idea here is, by gaining insights into the system's strengths and weaknesses, you will be better equipped to decide when and how to utilize the AI system in the recidivism risk assessment process.

\*

show me the prediction probabilities of 7645

The model predicts the instance with id equal to 7645 as:

- **likely to commit a crime** with 2.663% probability
- **unlikely to commit a crime** with 97.337% probability

Feedback 

What is the model's prediction for instance id 7645?



**Tip:** the model prediction for that instance is "likely to commit a crime" if its probability is above 50%, and consecutively that of "unlikely to commit a crime" is less than 50%.

ⓘ This is a question help text.

ⓘ Please select one answer

Unlikely to commit a crime.

Likely to commit a crime.

\*

TalkToModel

What are the five most important features for determining whether defendant 7645 is likely or unlikely to commit a crime again?

For the model's predictions on instance with id equal to 7645, the top 5 most important features are as follows, where 1 is the most important feature:

- 1: age
- 2: numberofpriorcrimes
- 3: recidivated
- 4: black
- 5: monthsinjail

What are the three most important features for determining whether defendant 7645 is likely or unlikely to commit a crime again?

ⓘ Please select one answer

Recidivated, number of prior crimes, and black.

Age, number of prior crimes, and black.

Number of prior crimes, recidivated, and age.

\*

show me the samples that are predicted as likely to commit a crime

For the data with the model predicts likely to commit a crime

For the data where the model predicts likely to commit a crime, the instance id's are:  
[9914, 3391, 5552, 9857, 66, 8681, 6467, 6304, 7373, 2885, 2568, 4739, 8873, 5916, 8401, 10700, 10134, 1395, 1976, 5428, 6828, 5453, 1034, 7200, 302, 8375, 8651, 3331, 5153, 9119, 7661, 6095, 8396, 4130, 10964, 10561, 10194, 9096, 8009, 79, 1150, 6870, 3725, 7292, 5263, 9287, 1897, 9017, 566, 1482, 7116, 2228, 6794, 3563, 3035, 6706, 4498, 3426, 1187, 1660, 1326, 8633, 8843, 3289, 2519, 10051, 581, 3408, 2723, 3369, 10254, 5205, 5119, 3085, 8773, 4715, 3499, 5155, 480, 5122, 2873, 6371, 1816, 8322, 4971, 6169, 5730, 5346, 2442, 9413, 8235, 4967, 7437, 1542, 7611]

Which one do you want to see?

Feedback

From the screenshot above, choose the 12th defendant ID in the list.

**① Please select one answer**

- 4739
- 8873
- 4738

\* what are the prediction probabilities for 1976

The model predicts the instance with id equal to 1976 as:  

- likely to commit a crime with 54.13% probability
- unlikely to commit a crime with 45.87% probability

What are the prediction probabilities for candidate 1976 if the number of prior crimes is reduced to 5?

The model predicts the instance with id equal to 1976 and numberofpriorcrimes is decreased by 5.0 as:  

- likely to commit a crime with 9.465% probability
- unlikely to commit a crime with 90.535% probability

How do the prediction probabilities change for defendant 1976, if his *number of prior crimes* is reduced by 5?

**① Please select one answer**

- The model's prediction remains the same.
- The model's prediction is flipped.

\* Show me the prediction probabilities of sample 1542

The model predicts the instance with id equal to 1542 as:  

- likely to commit a crime with 74.721% probability
- unlikely to commit a crime with 25.279% probability

Show me the prediction probabilities of sample 79

The model predicts the instance with id equal to 79 as:  

- likely to commit a crime with 84.486% probability
- unlikely to commit a crime with 15.514% probability

Feedback

Is the model more confident about its prediction for defendant 1542 or 79?

**① Please select one answer**

- 1542
- 79

TalkToModel

what are the categories of data where the model typically gives incorrect predictions

For all the instances in the data, the model typically predicts incorrect:

$\text{if } (\text{numberofpriorcrimes} > 4.5) \text{ and } (\text{age} \leq 34.5) \text{ and } (\text{monthsinjail} \leq 138.0) \text{ and } (\text{black} > 0.5) \text{ and } (\text{felony} \leq 0.5)$

TalkToModel

Hello 😊 I'm a machine learning model trained to predict whether someone will commit a crime.

Let's get started. Ask me something!

Show me the data for applicant 7942

then then the model is incorrect 66.42% over 19 samples  
if (numberofpriorcrimes <= 4.5) and (age <= 24.5) and  
(monthsinjail > 4.5) and (monthsinjail > 7.5) and  
(monthsinjail > 46.5) then then the model is incorrect 66.67%  
over 9 samples  
  
if (numberofpriorcrimes > 4.5) and (age > 34.5) and  
(monthsinjail > 7.5) and (monthsinjail <= 65.5) and  
(monthsinjail > 38.5) then then the model is incorrect 87.5%  
over 8 samples

For the data with id equal to 7942, the features are  
  
age: 41  
recidivated: 1  
numberofpriorcrimes: 8  
monthsinjail: 65  
felony: 1  
misdemeanor: 0  
woman: 1  
man: 0

Given the category defendant falls under, how likely is the model's prediction incorrect?

 **Tip:** The model's error rate is supposed to help you decide.

**① Please select one answer**

- Very unlikely (more than 75%).
- Very unlikely (less than 25%).