



Академия  
искусственного интеллекта  
для школьников

# Искусственный интеллект

**Машинное обучение —**  
про алгоритмы, которые  
находят в данных правила  
и закономерности

**О каких алгоритмах  
идёт речь?**



Алгоритмы машинного  
обучения — это двигатели

Данные — это топливо

объекты

$f(x)$

ответы

**X**

$y=60 \times x$

**y**

Часы

Минуты

1,5

Как перевести  
часы в минуты?

90

2

120

6

360

...

...

объекты

$f(x)$

ответы

$X$

$y = x_1 * x_2$

$y$

Масса, кг

Ускорение,  $m/s^2$

Сила, Н

0,2

0,5

$F = m * a$

0,1

1

9,8

9,8

2,7

20

54

...

...

...

объекты

$f(x)$

ответы

$x$



$y$

Что делать, если  $y$  нас есть  
 $x$  и  $y$ , но формула  $f$  неизвестна?

объекты

$f(x)$

ответы

**X**

?

**у**

Текст  
отзыва

Эмоциональный  
окрас отзыва

«Видео нормас 👍»

Положительная,  
+  
Отрицательная, -

«Да вы &#@\$%#»

объекты

$f(x)$

ответы

$X$

$y$


$$y \approx f(x)$$

Суть машинного обучения – всегда можно приблизить  $f$  на основе данных об  $x$  и  $y$ .

объекты    ответы

**X , y**

учитель

все данные

Обучение  
с учителем

$$y \approx f(x)$$

**f(x)**

математическая модель

Чем больше данных  
– тем лучше

$x_{11}$	$x_{12}$	...	$y_1$
$x_{21}$	$x_{22}$	...	$y_2$
$x_{31}$	$x_{32}$	...	$y_3$
...	...	...	...



# Этапы машинного обучения

## 1. Построение модели

все данные  
( $x, y$ )

$x_{11}$	$x_{12}$	...	$y_1$
$x_{21}$	$x_{22}$	...	$y_2$
$x_{31}$	$x_{32}$	...	$y_3$
...	...	...	...



модель  
 $f(x)$

$$f(x)$$

Математическая модель и есть результат

## 2. Применение модели

все данные  
( $x, y$ )

$x_{71}$	$x_{72}$	...
$x_{81}$	$x_{82}$	...
$x_{91}$	$x_{92}$	...
...	...	...



модель  
 $y \approx f(x)$

$$y_7 \approx f(x_{71}, x_{72}, \dots)$$

$$y_8 \approx f(x_{81}, x_{82}, \dots)$$

$$y_9 \approx f(x_{91}, x_{92}, \dots)$$

...

# Задача классификации – когда ответы это метки. Объекты могут быть чем угодно.

Определение эмоционального окраса текста

«Видео нормас 👍»



+

«Да вы &#@%#»

-

Текст, набор слов

2 класса

Распознавание изображений рукописных цифр

5

1



5

1

4

0

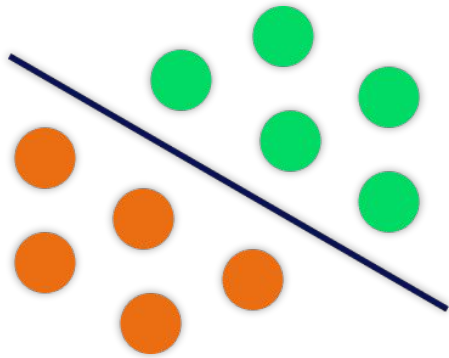
4

0

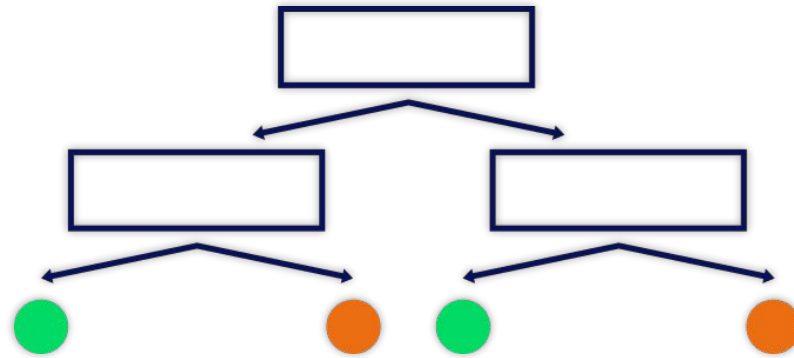
Картинка 28×28

10 классов цифр

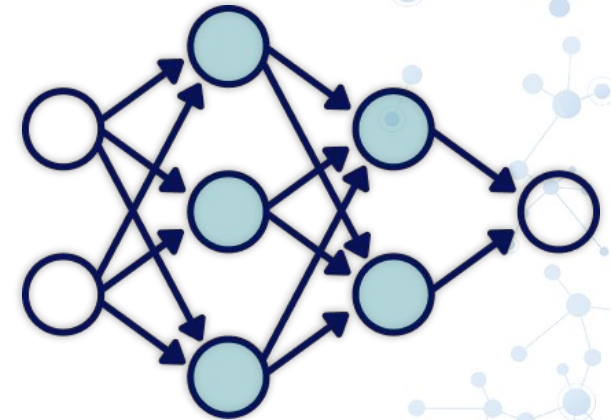
**Модели – математические формулы.**  
Их десятки тысяч, но можно выделить  
3 самых популярных класса:



Линейные модели



Деревья решений



Нейронные сети

## Пример данных:

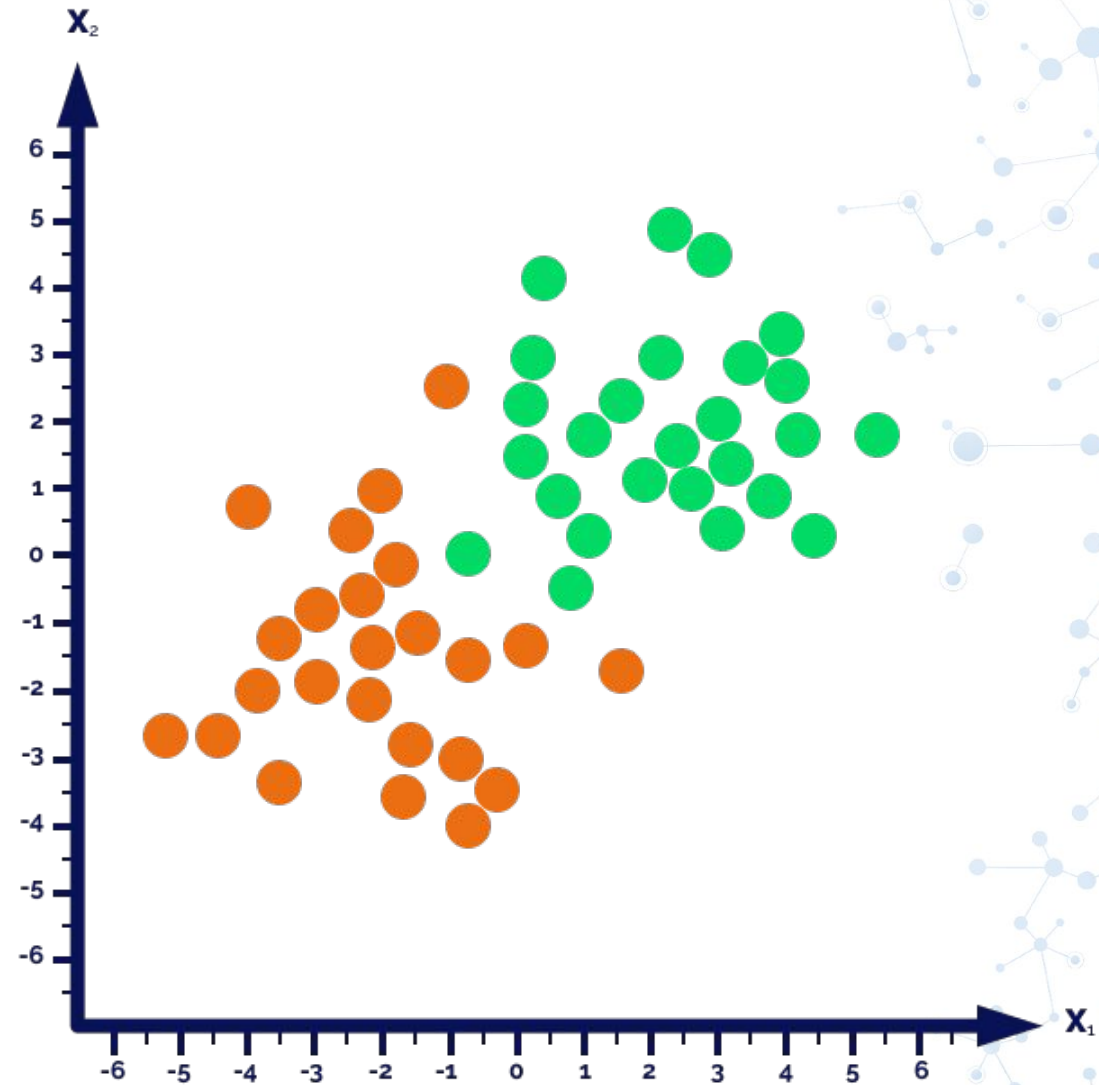
Объекты (x):

$x_1$ ,  $x_2$  – координаты

Объекты (y):

●, ● – цвет точки

(классификация на два класса)



Линейная модель – простейшая формула, задает прямую линию с весами

$$y = a_1 \times x_1 + a_2 \times x_2 + \dots + b$$

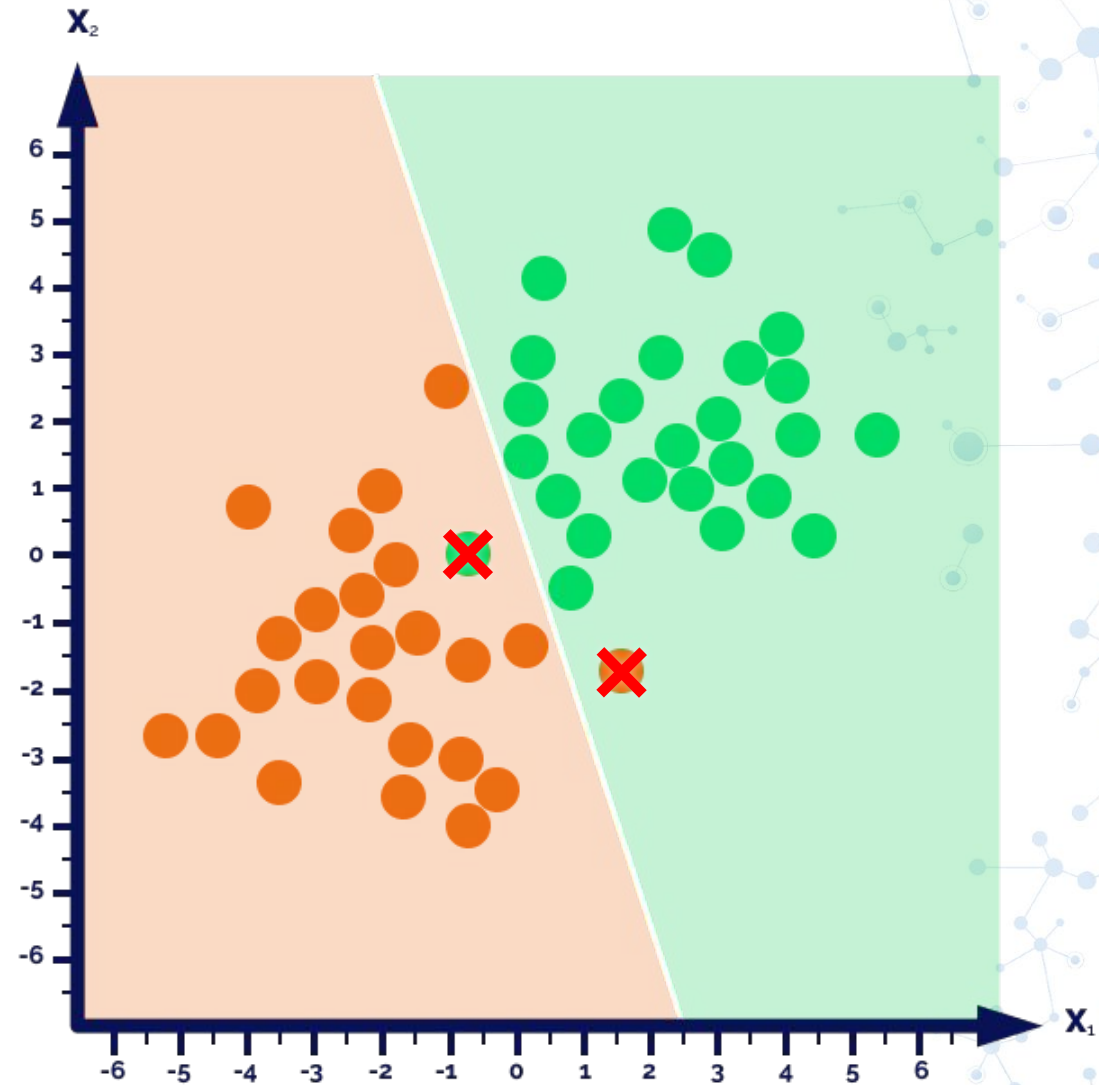
$a_1, a_2, \dots, b$  – числа, веса

сумма больше нуля – один класс  
иначе – другой класс

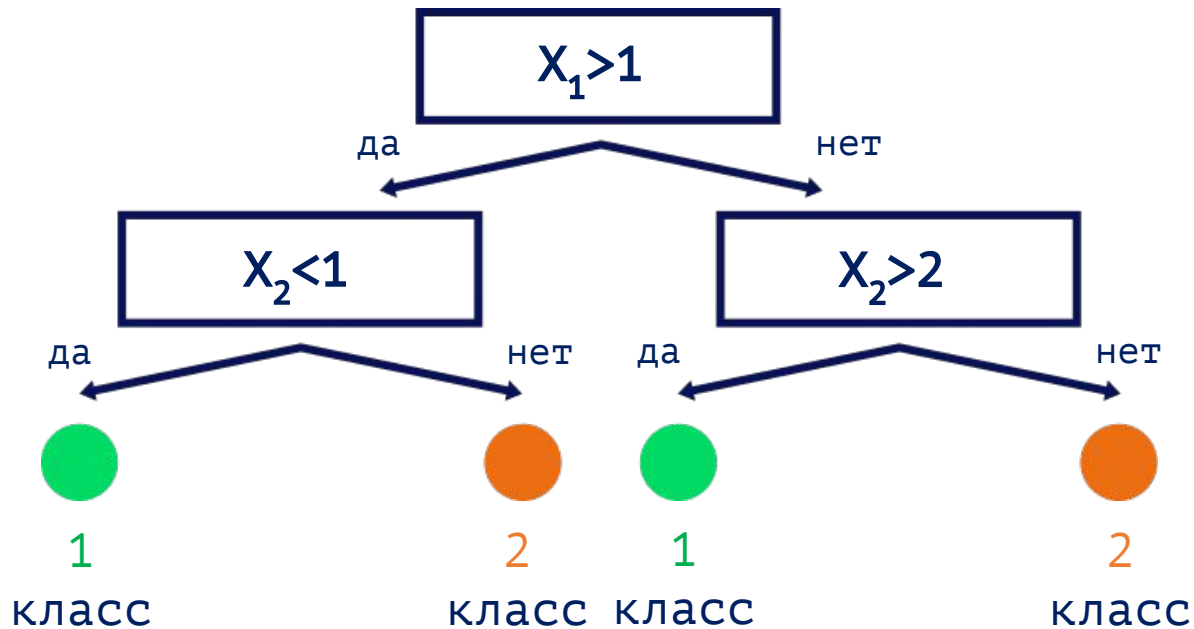
# Обученная линейная модель

$$y = 2,2 \times x_1 + 0,7 \times x_2 - 0,5$$

✗ – ошибка модели



Дерево решений – задает набор правил, записанных в компактной форме



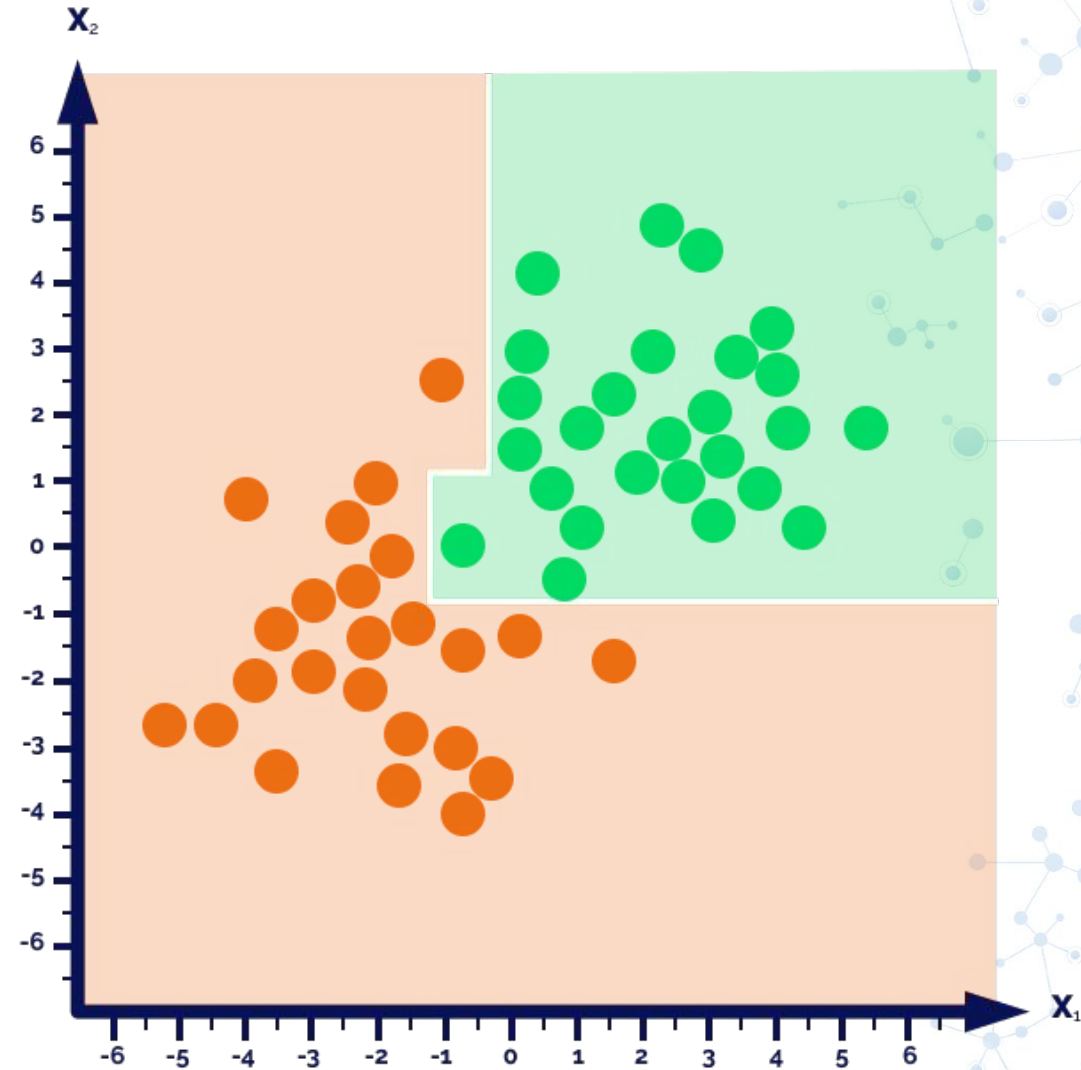
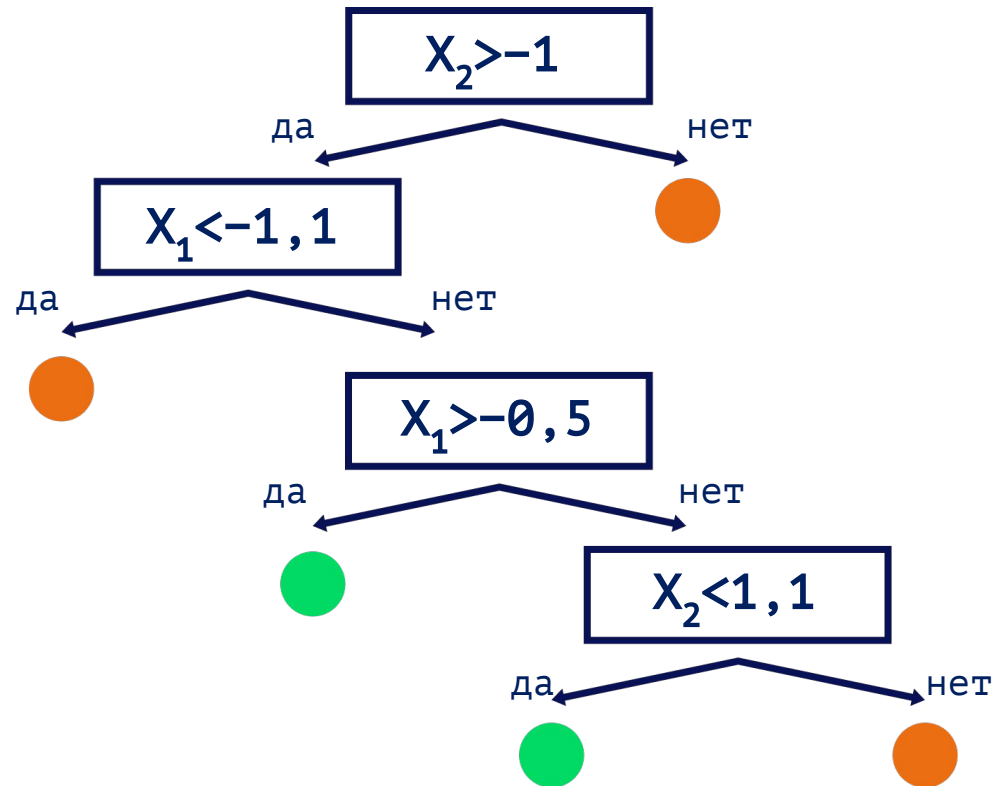
Если  $x_1 > 1$  и  $x_2 < 1$ , то 

Если  $x_1 > 1$  и  $x_2 \geq 1$ , то 

Если  $x_1 \leq 1$  и  $x_2 > 2$ , то 

Если  $x_1 \leq 1$  и  $x_2 \leq 2$ , то 

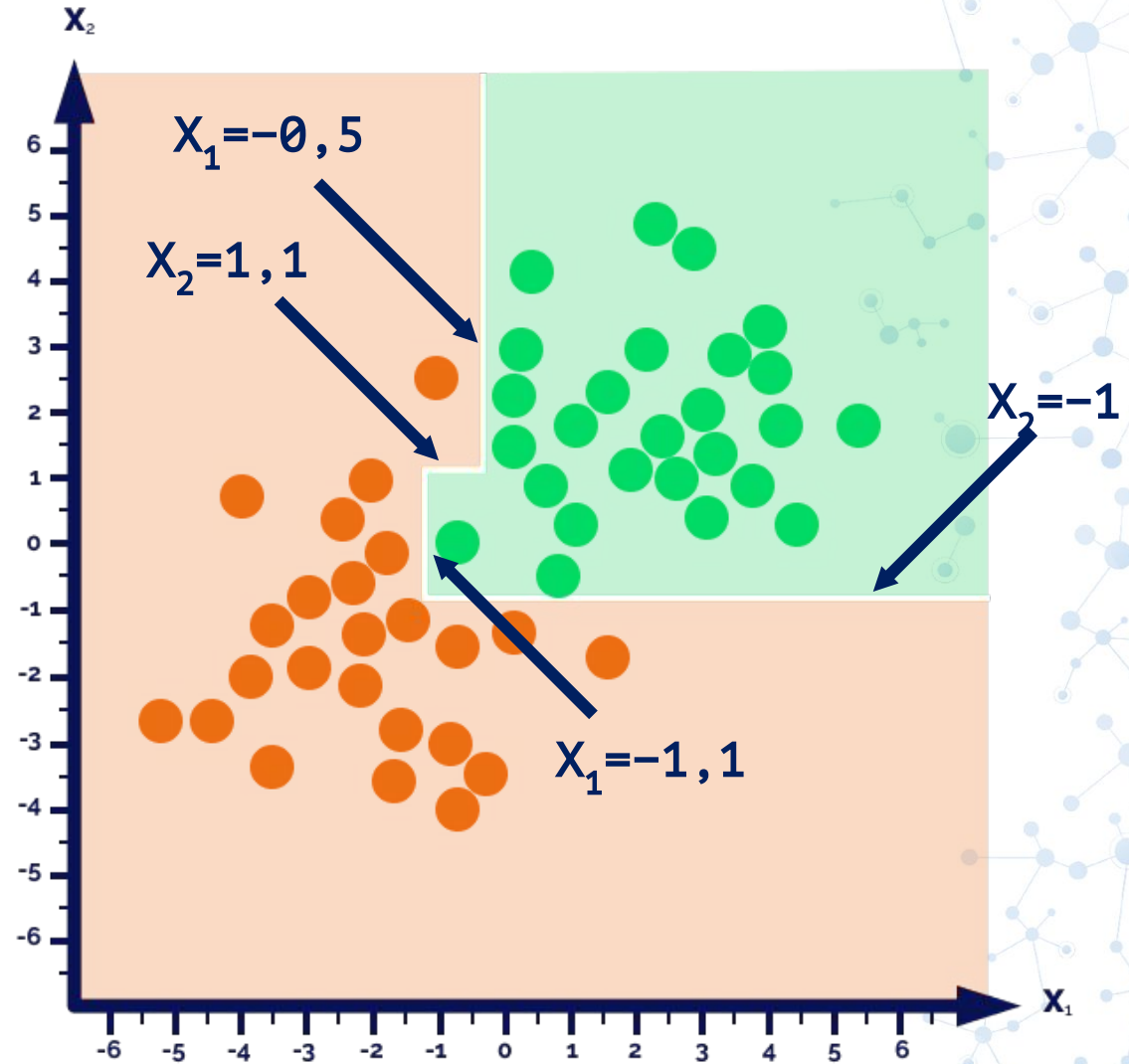
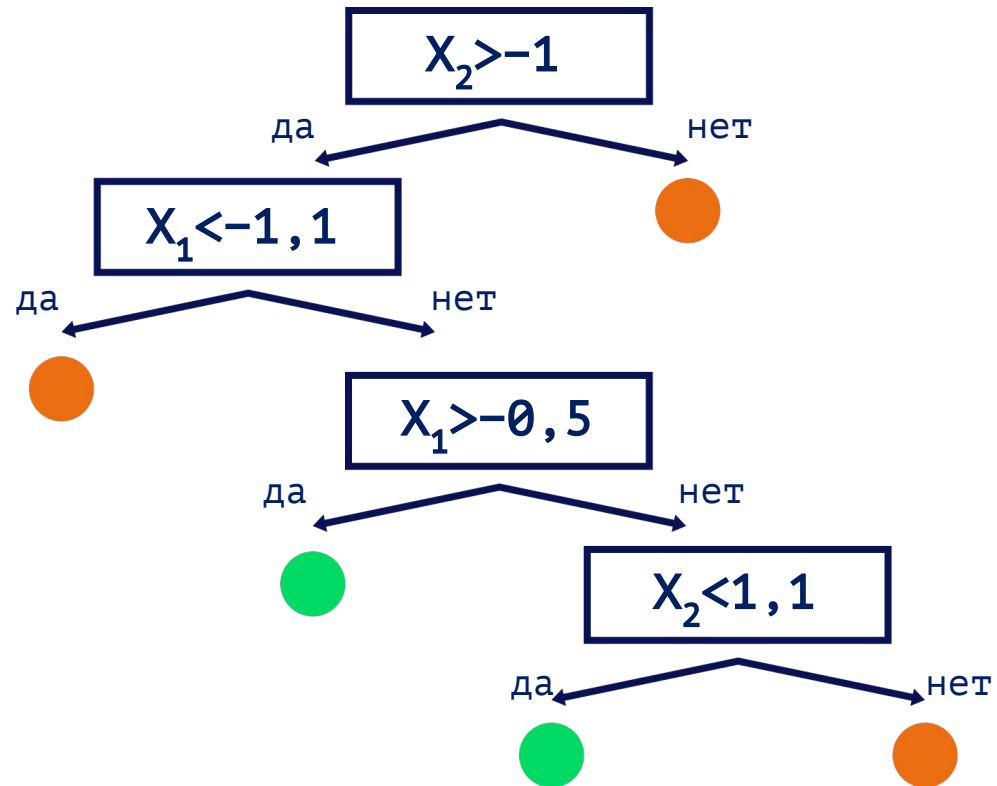
# Обученное дерево решений

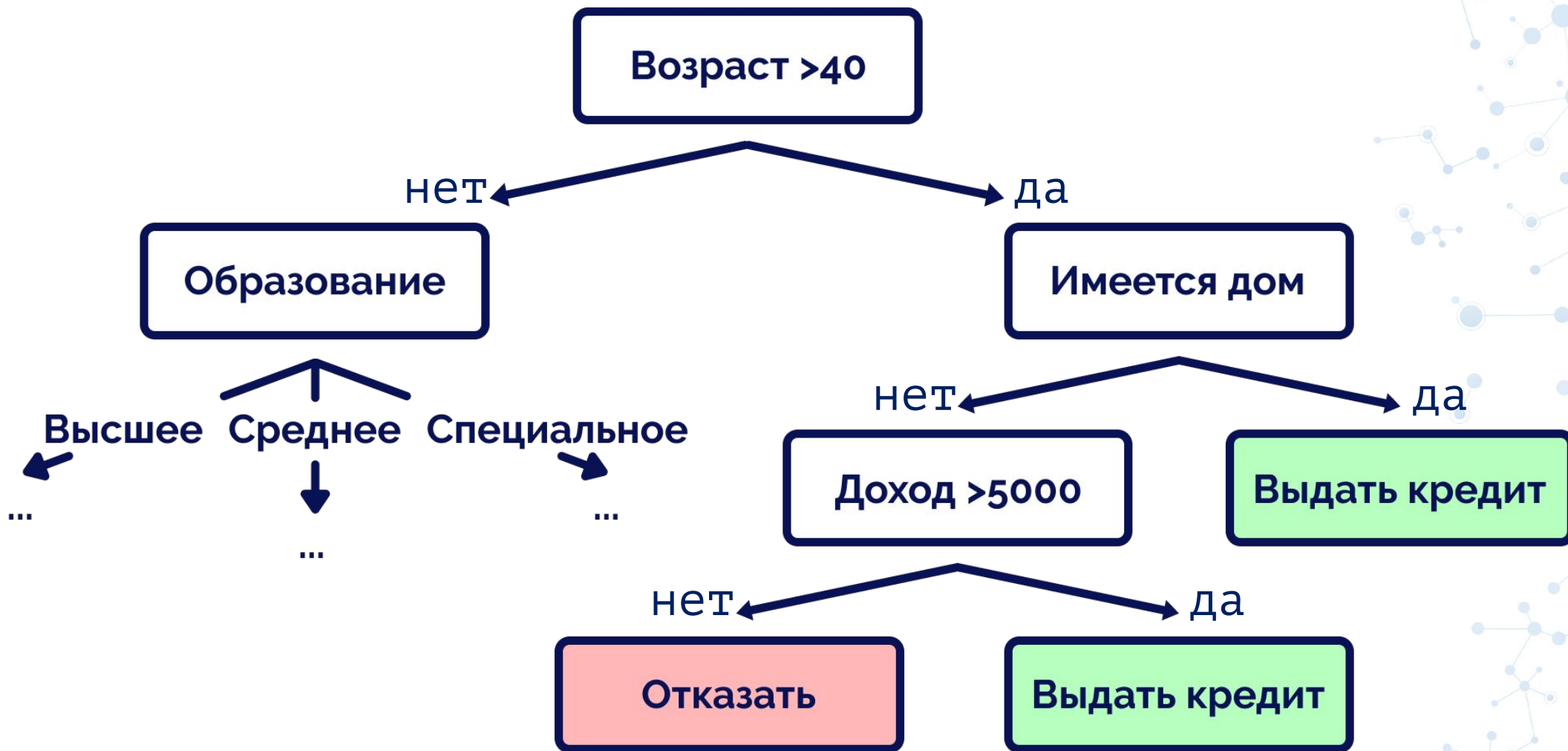




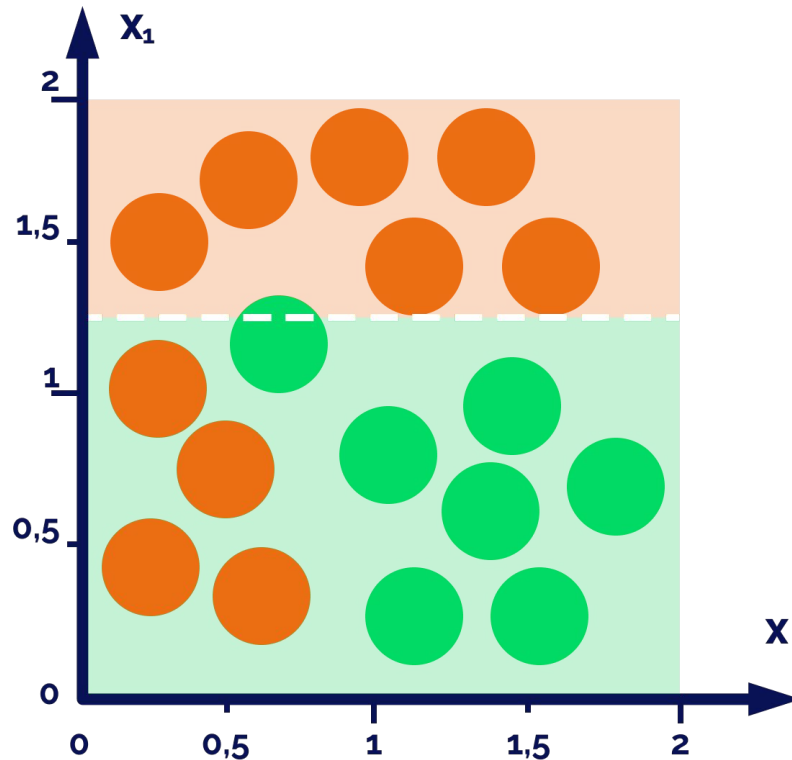
# Обученное дерево решений

$X_2 > 2$



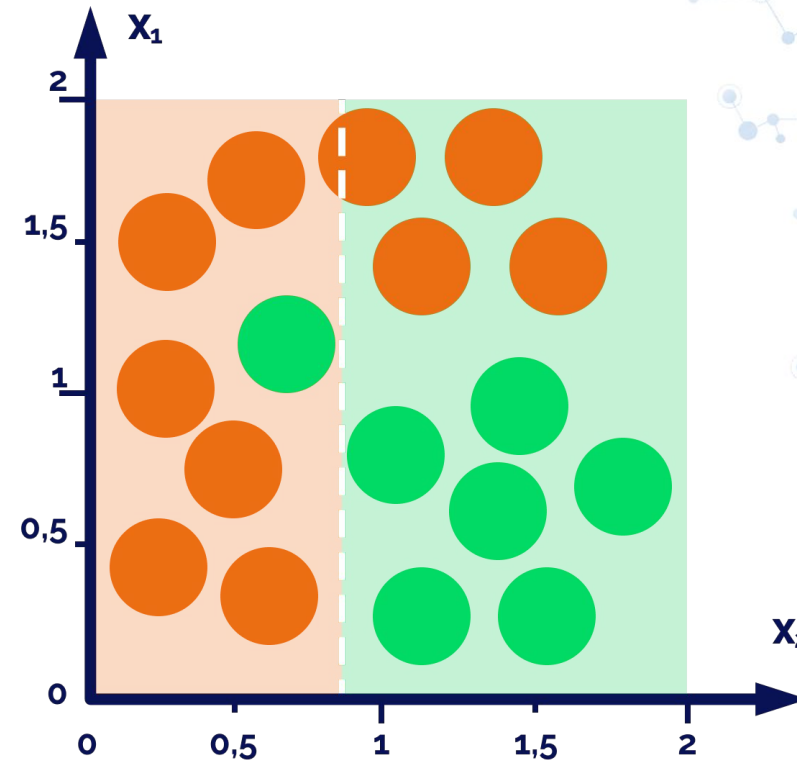


# Строим дерево решений



$X_1 < 1,25$

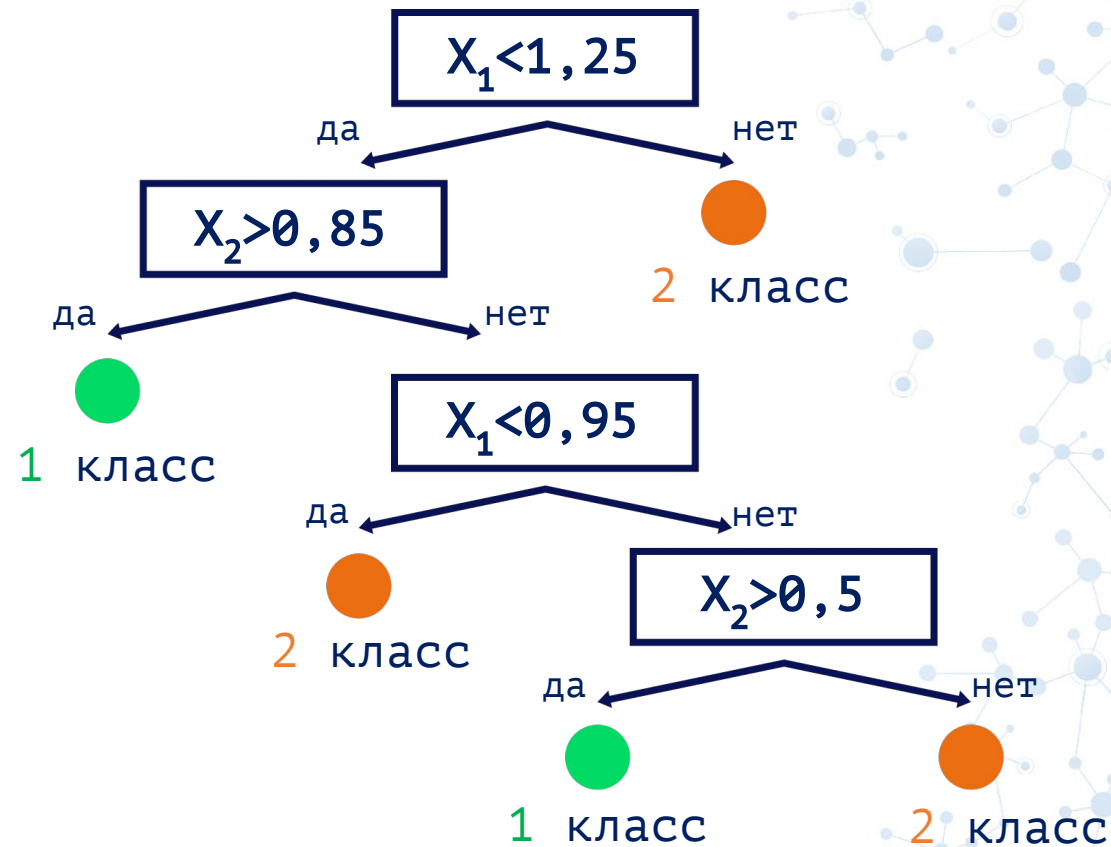
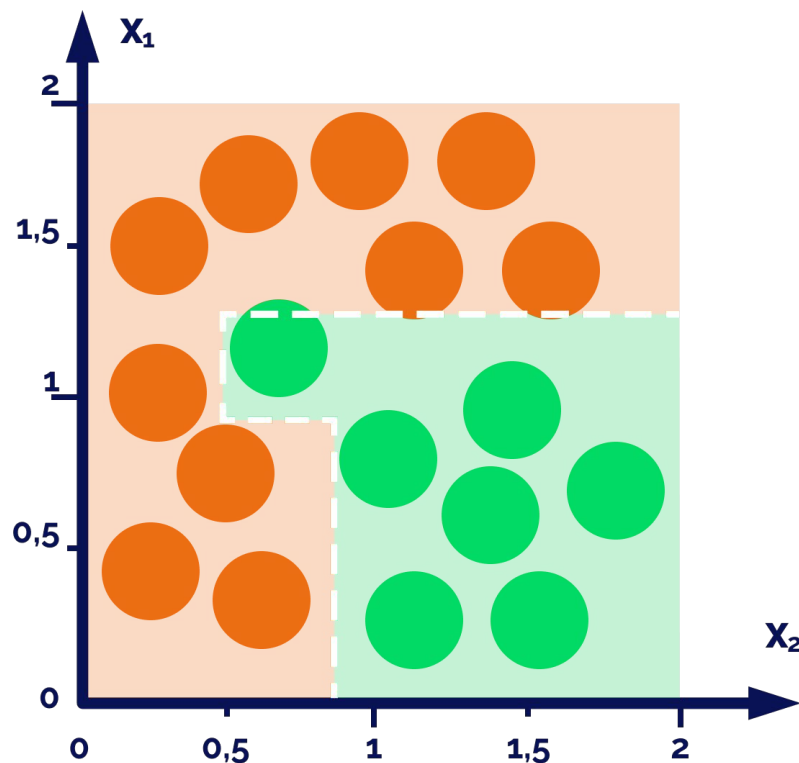
4 ошибки



$X_2 > 0,85$

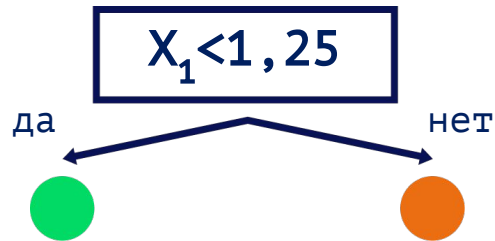
5 ошибок

# Дерево решений на примере

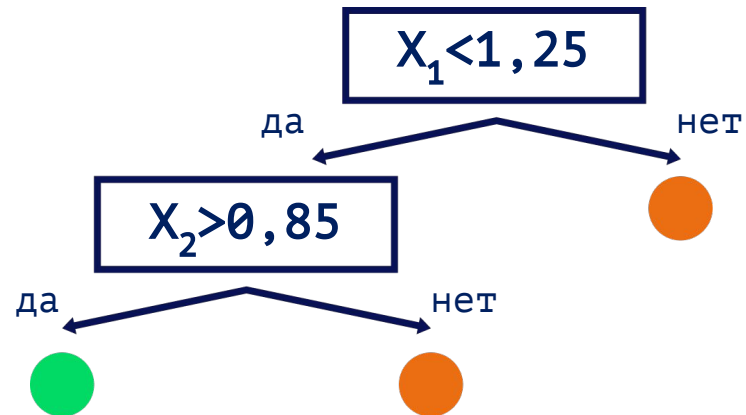


# Глубина деревьев решений

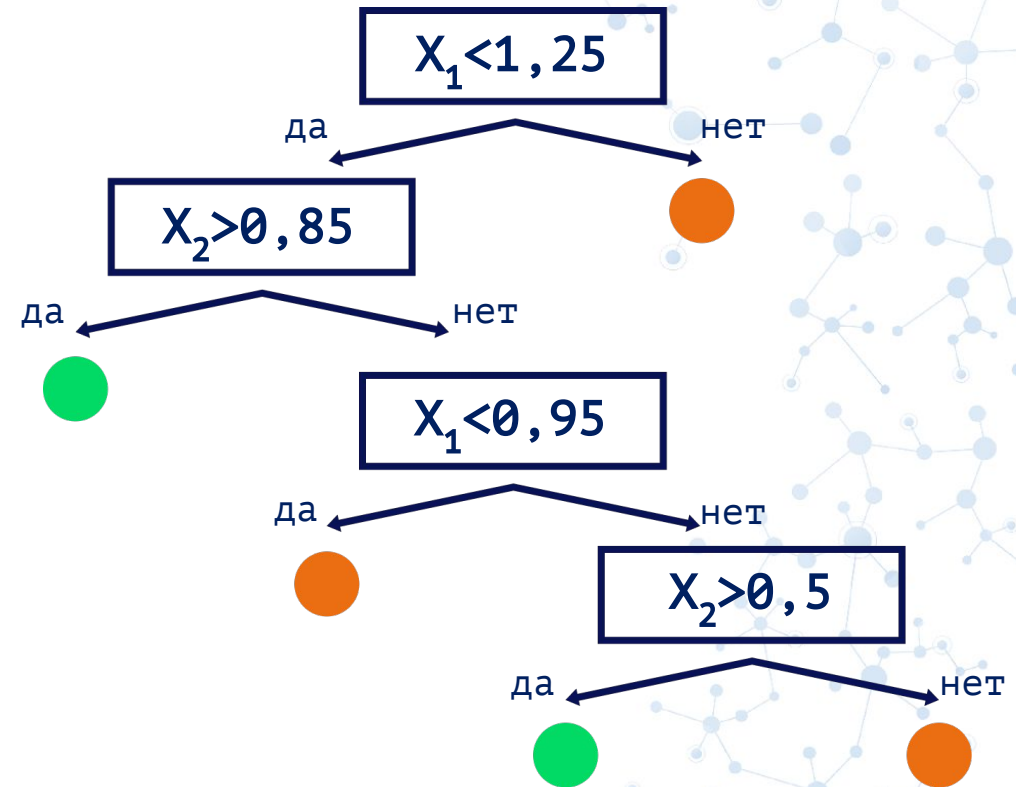
Глубина 1



Глубина 2



Глубина 4



Глубина деревьев –  
настраиваемый гиперпараметр

# Задача машинного обучения

$f(x)$

$X$

$y$

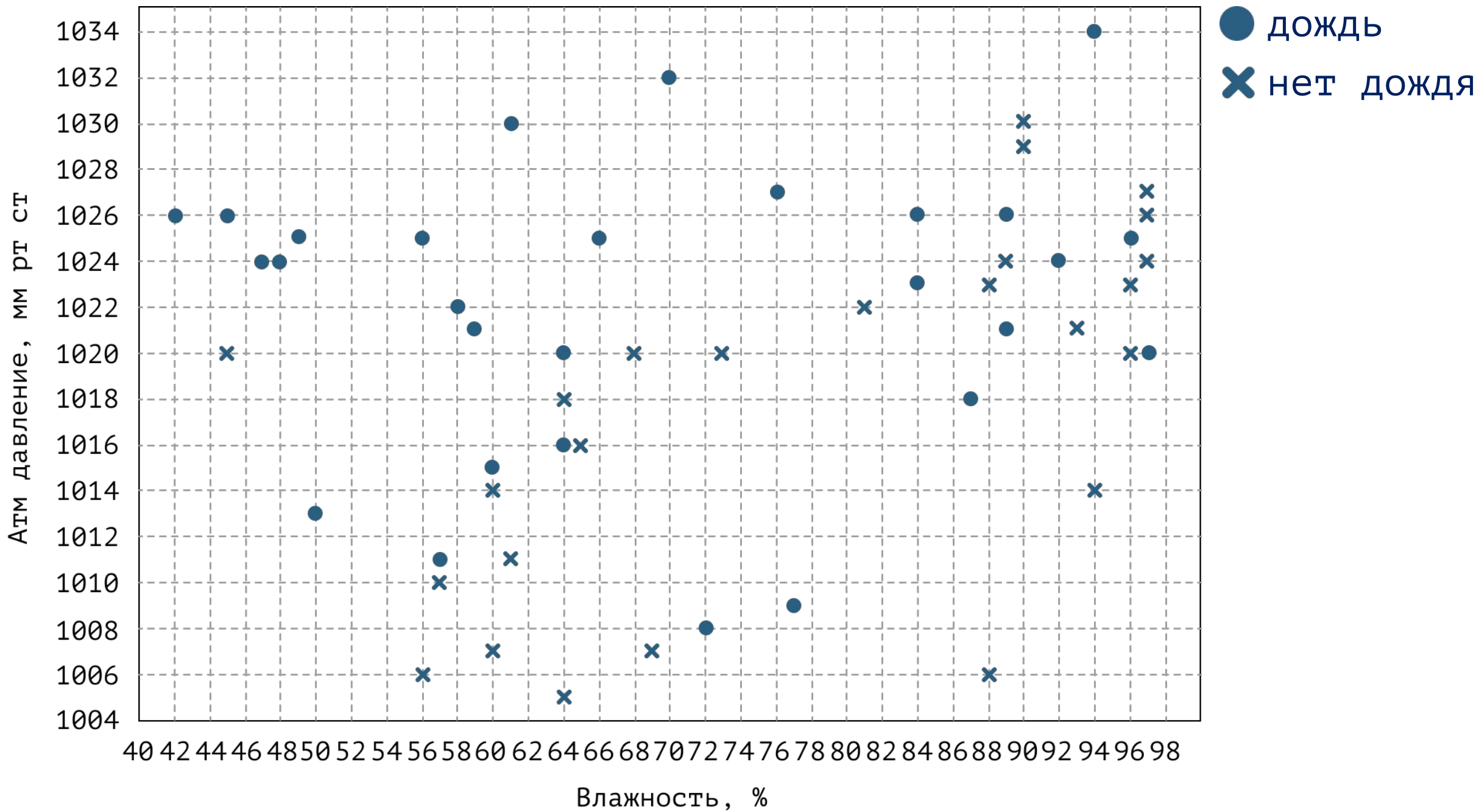
$x_1$  – атм. давление

$x_2$  – влажность воздуха

метка класса

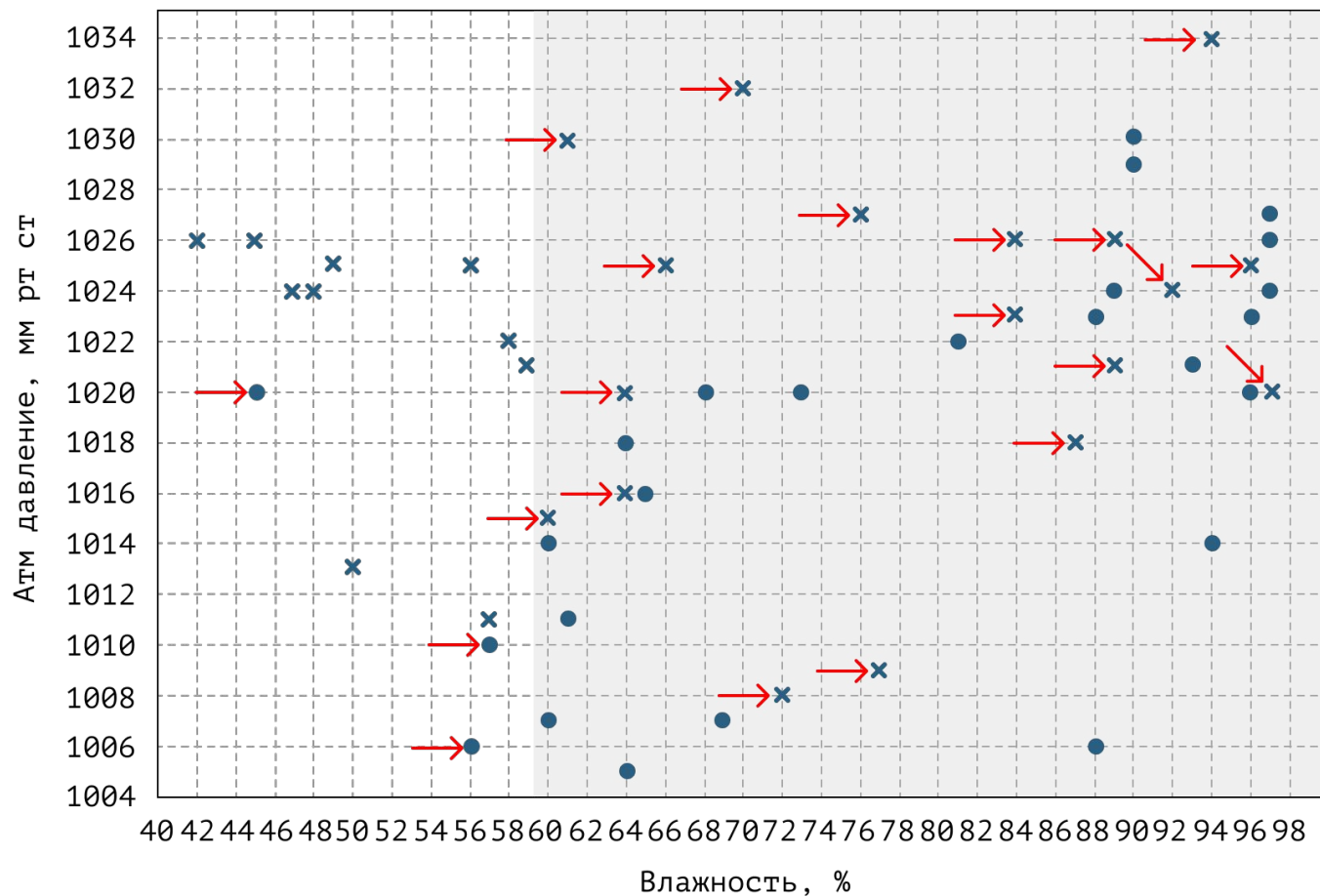
«ДОЖДЬ» / «НЕТ ДОЖДЯ»





область предсказания:  
нет дождя

область предсказания:  
был дождь



Глубина 1

Влажность >59,5

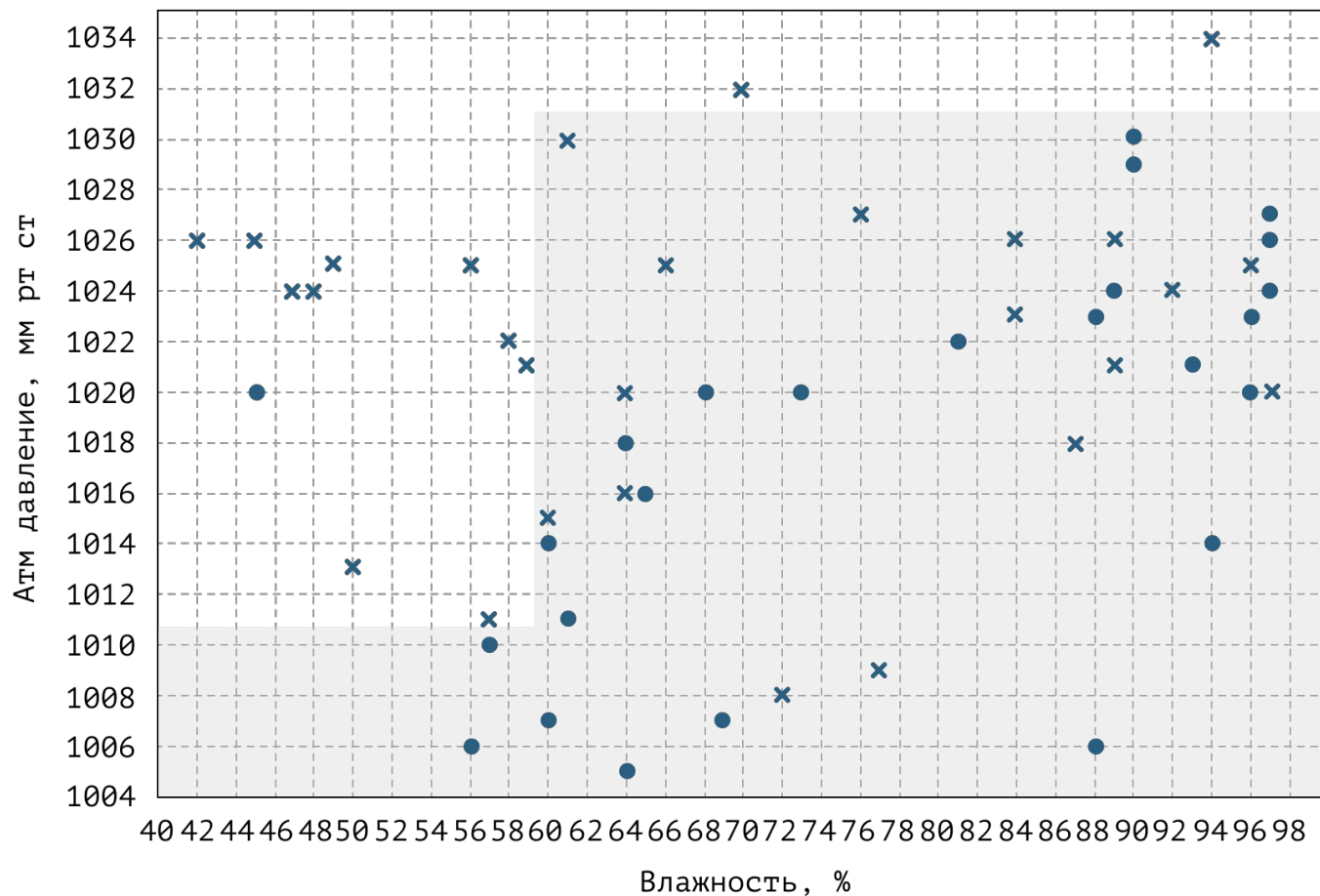
да



нет

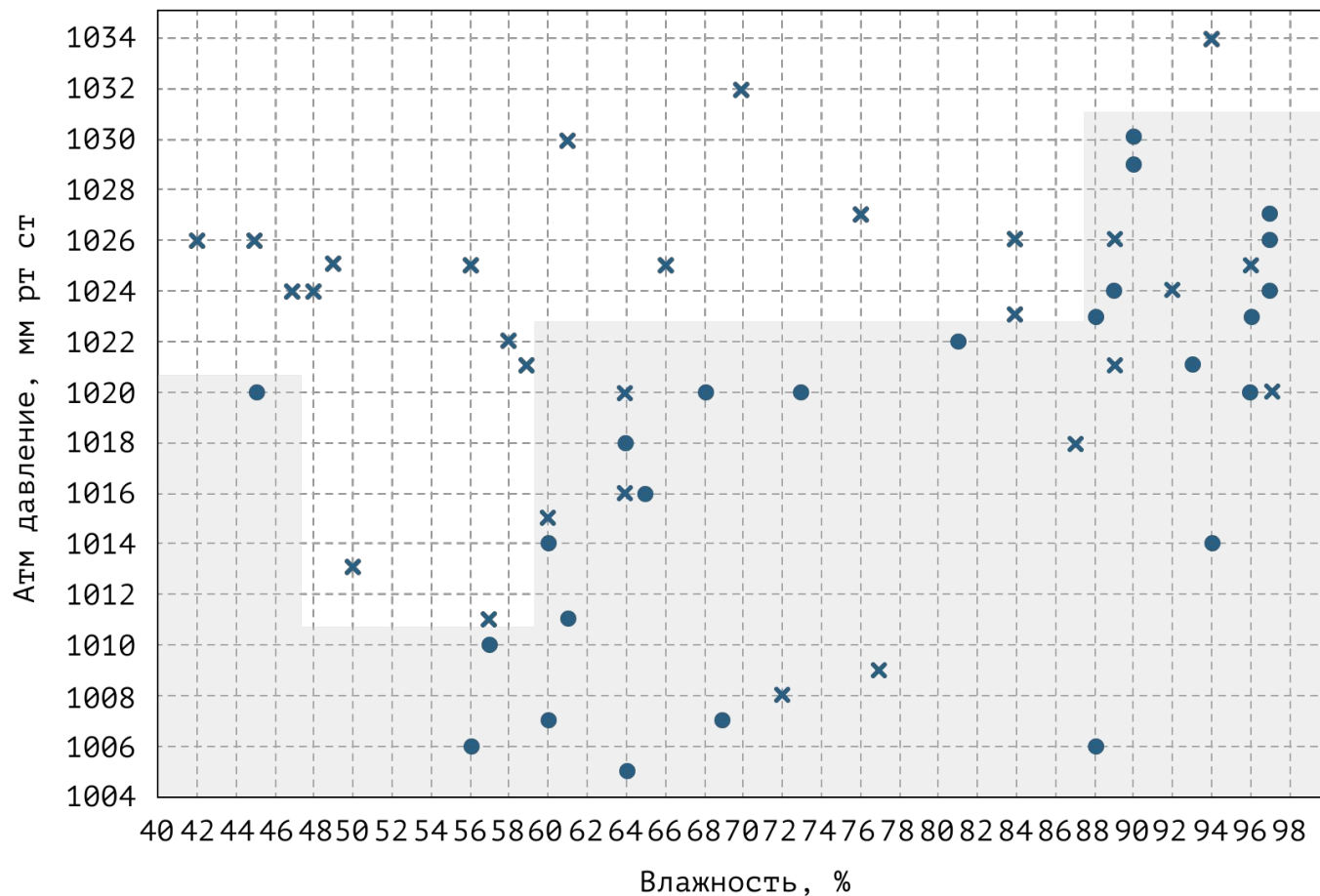






→ 17 ошибок  
(найдите сами)

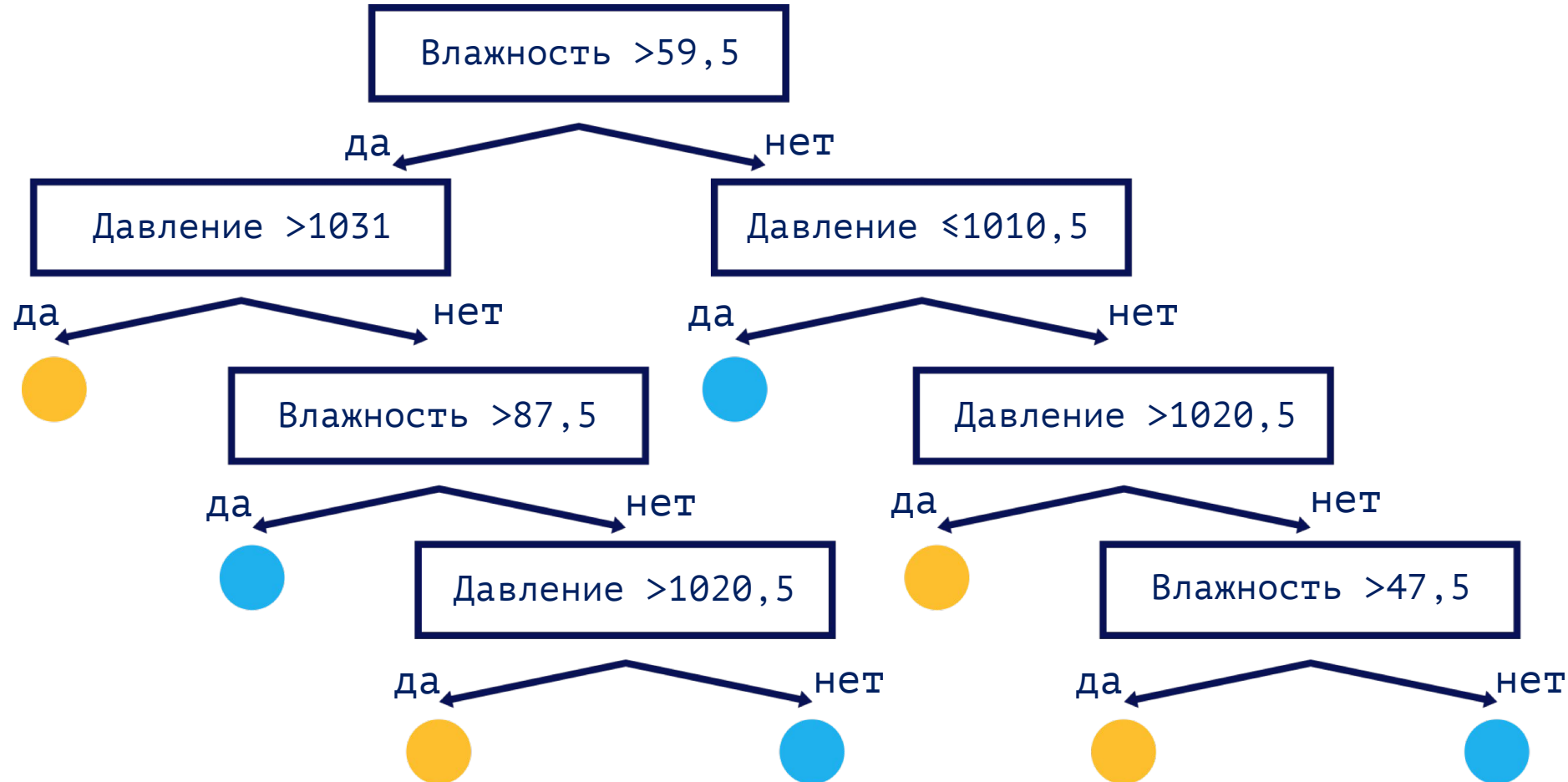
# Глубина 4



**Всего 11 ошибок!**

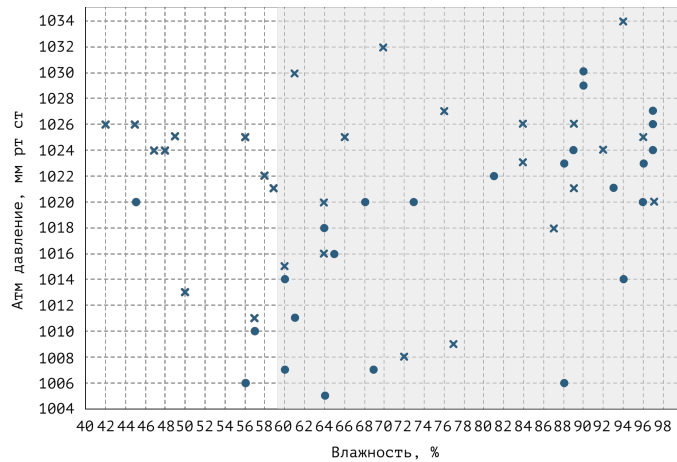
Почти в два раза меньше, чем в дереве с глубиной 1

# Глубина 4

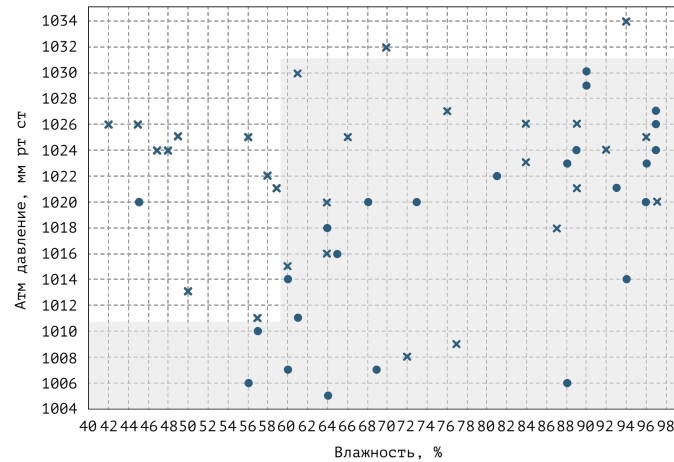


# Тестируем наши обученные модели (деревья) на НОВЫХ ДАННЫХ

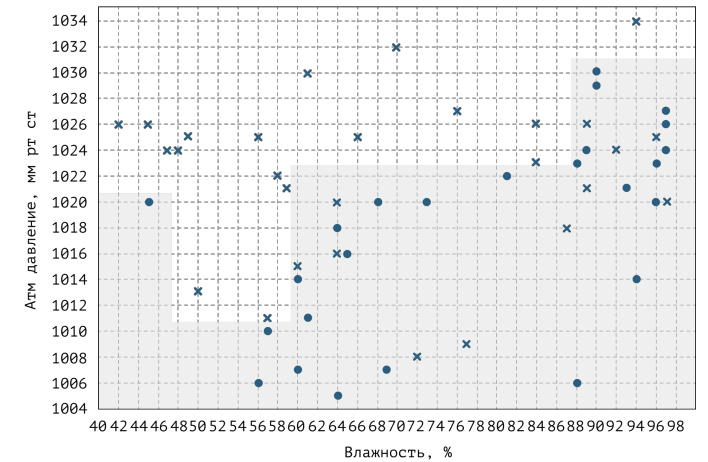
## Глубина 1



## Глубина 2

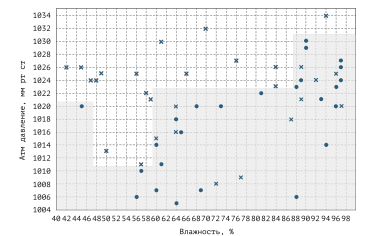
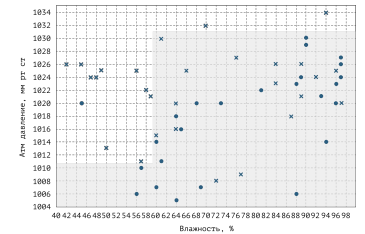
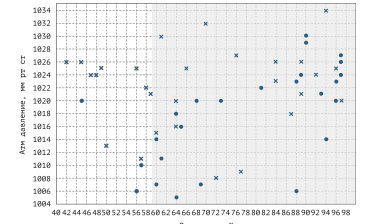


## Глубина 4



# Переносим новые данные на каждую модель

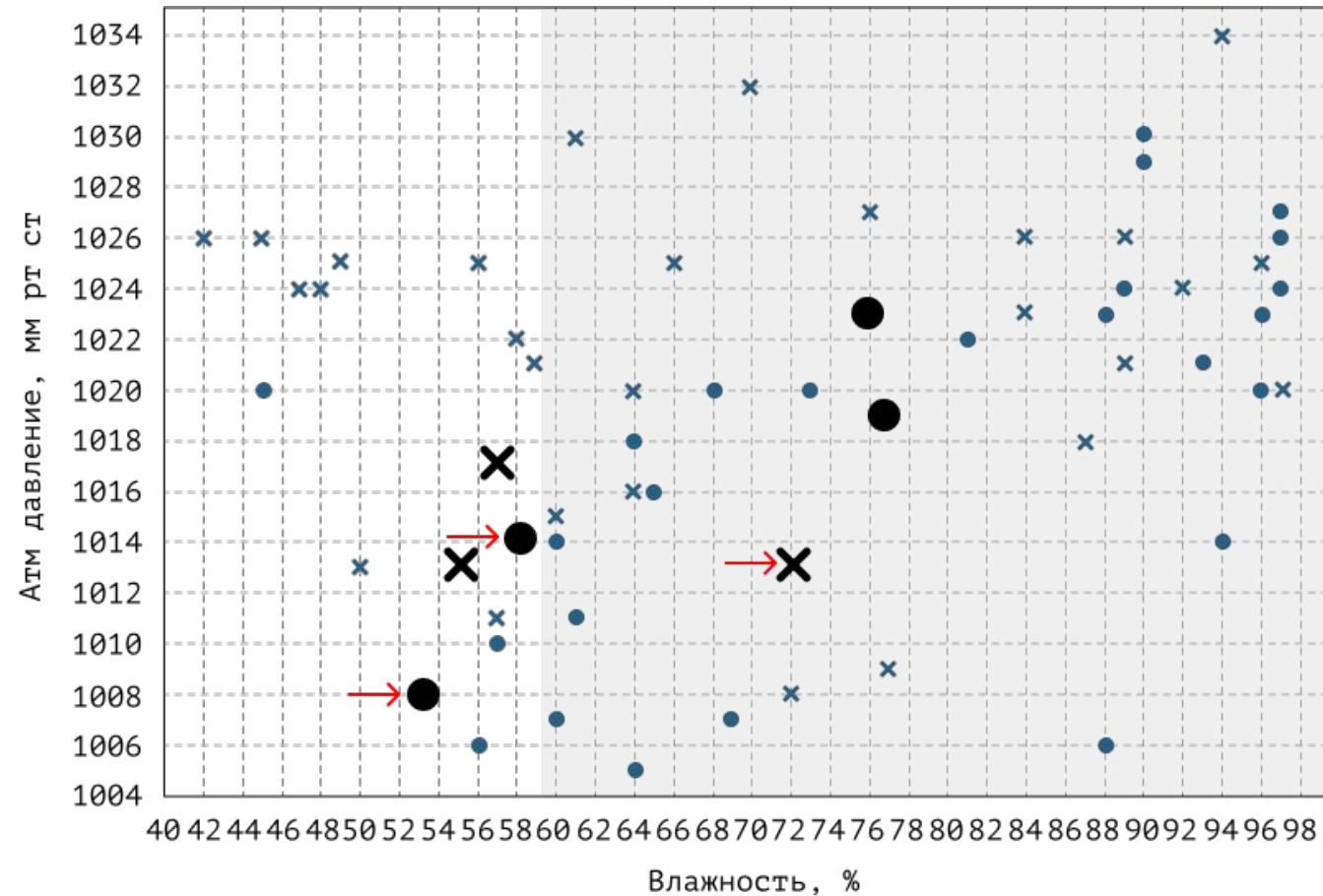
День недели	Атм. давление, мм рт.ст.	Влажность воздуха, %	Класс «Дождь» / «Нет дождя»	Рисуем на графиках
1	53	1008	Дождь	●
2	59	1014	Дождь	●
3	77	1019	Дождь	●
4	76	1023	Дождь	●
5	55	1013	Нет дождя	✕
6	58	1017	Нет дождя	✕
7	73	1011	Нет дождя	✕



# Что получилось?

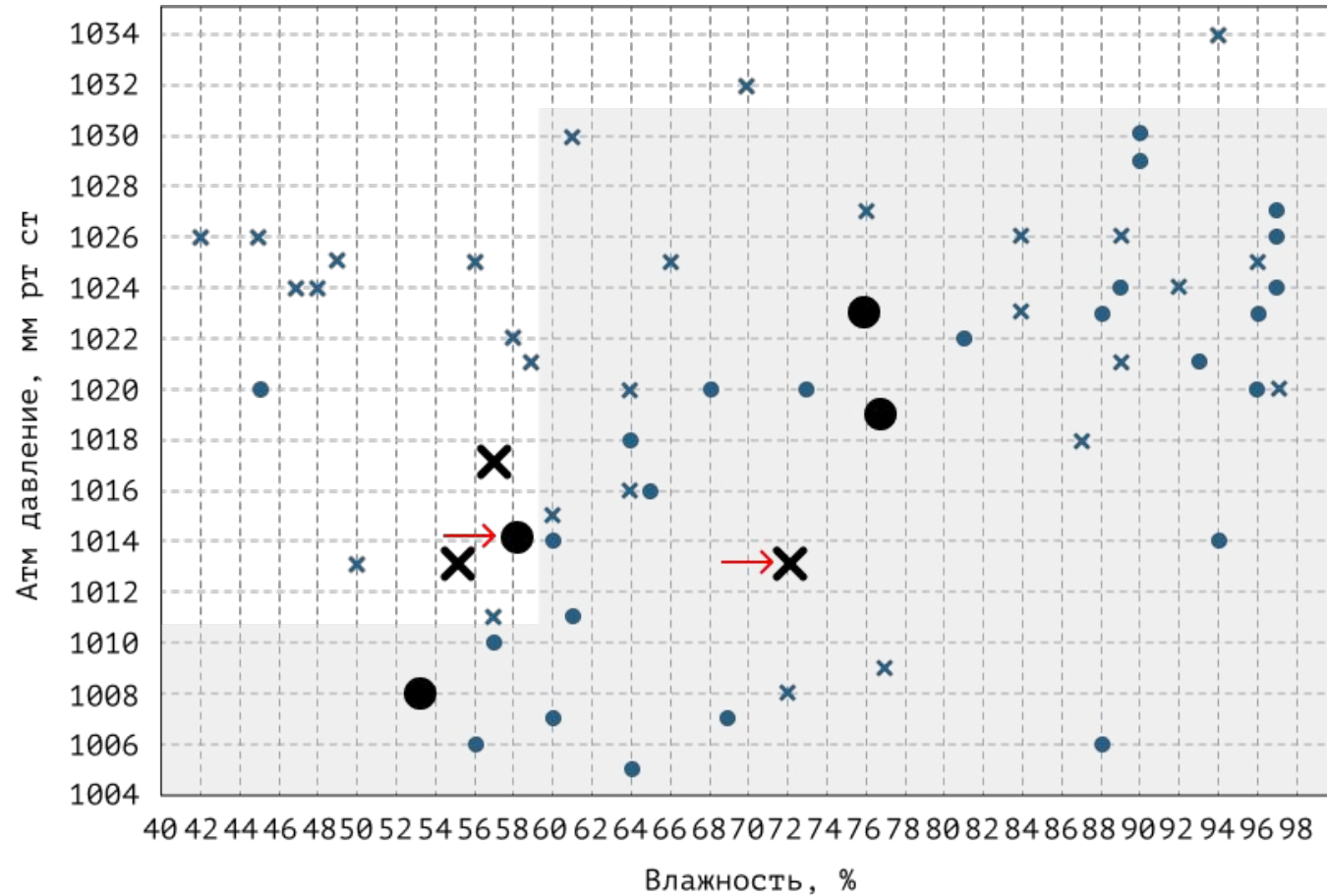
Глубина 1

3 ошибки



# Что получилось?

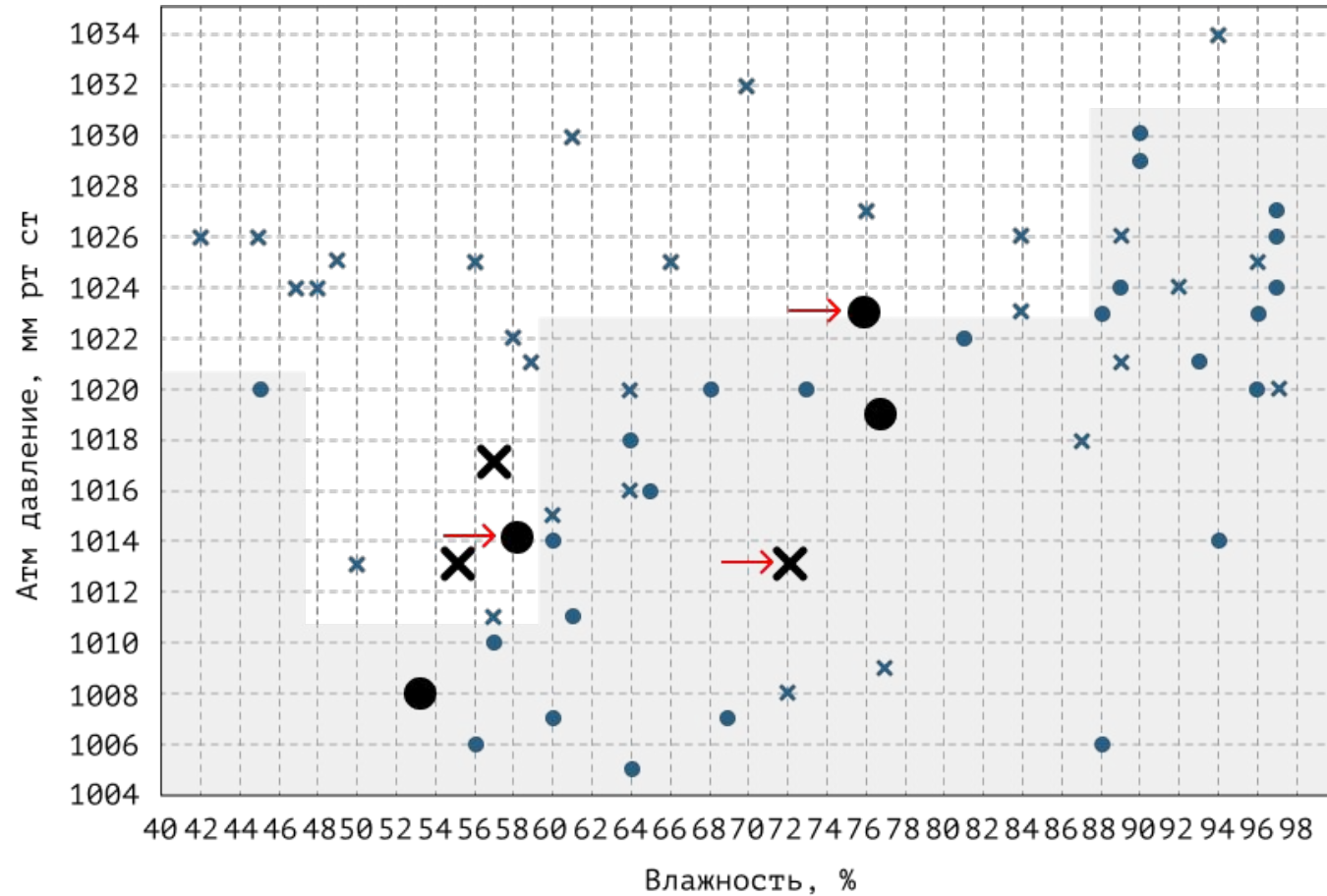
Глубина 2  
2 ошибки



# Что получилось?

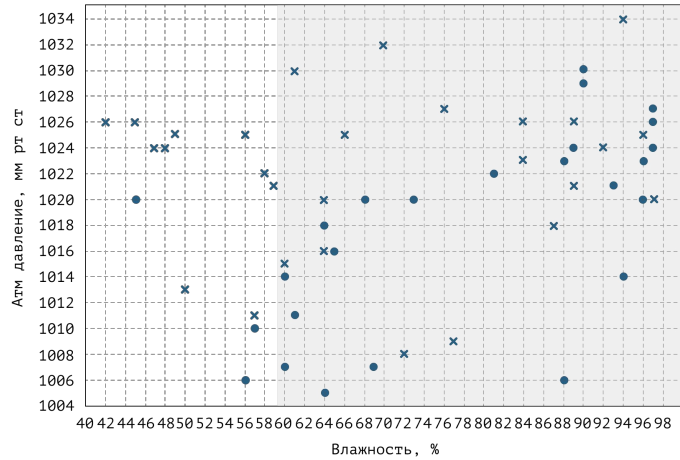
Глубина 4

Снова  
3 ошибки!



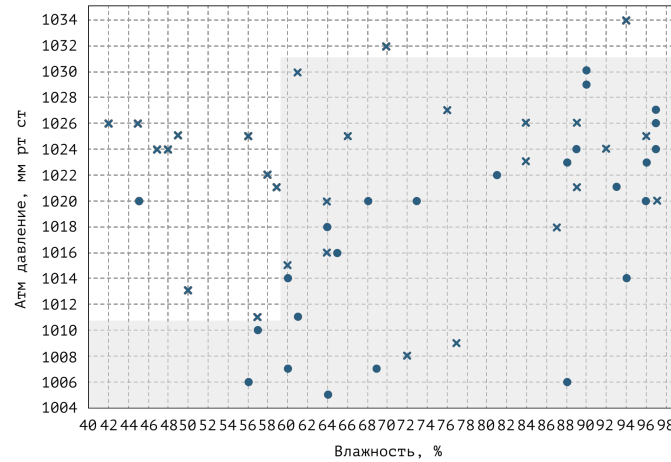


## Глубина 1



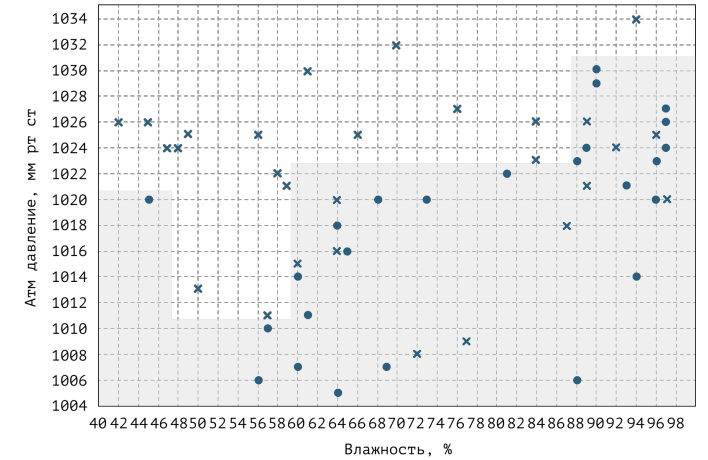
3 ошибки

## Глубина 2



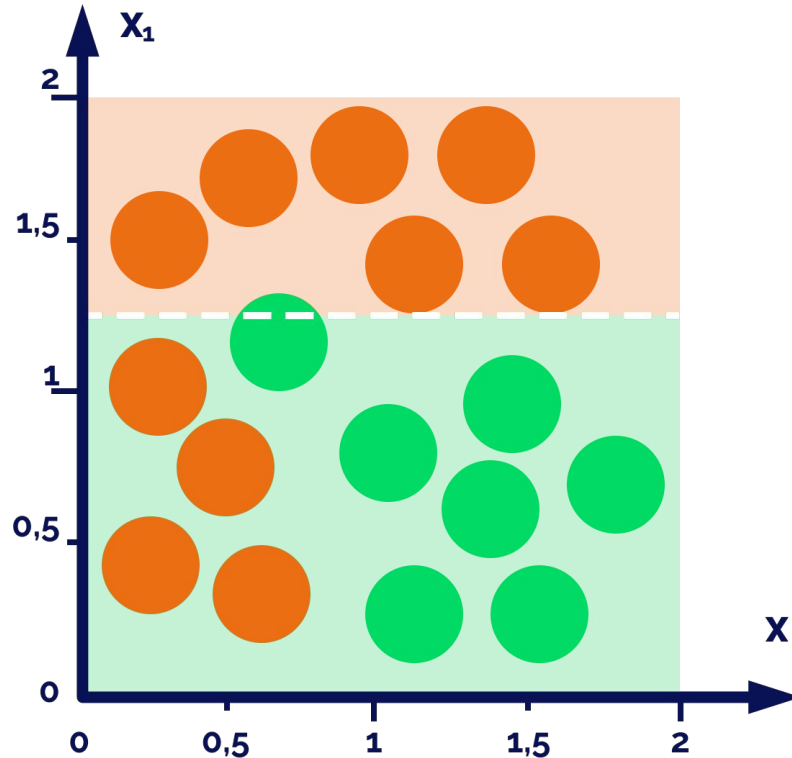
2 ошибки –  
лучший результат

## Глубина 4



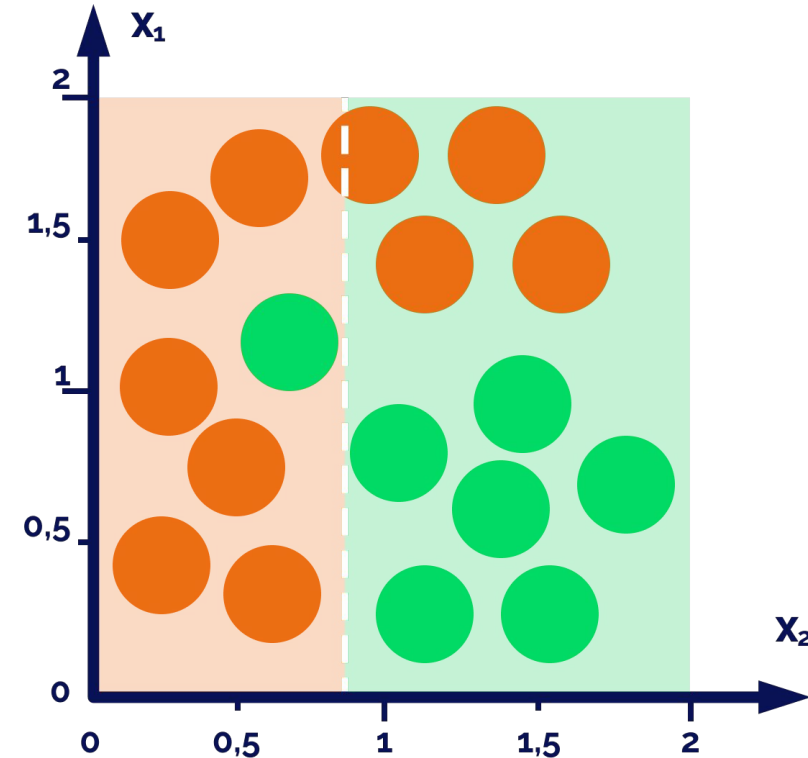
3 ошибки

# Строим дерево решений



$X_1 < 1,25$

4 ошибки



$X_2 > 0,85$

5 ошибок