



Московский институт электроники и
математики им. А.Н. Тихонова

Кафедра информационной
безопасности киберфизических
систем

Москва 2025

Интерпретация моделей в RapidMiner

Интерпретация моделей: сущность и назначение

Интерпретация модели — процесс объяснения причин принятия решений моделью машинного обучения, выявление факторов, оказывающих влияние на результат, и анализ логики работы алгоритма.

MODEL
INTERPRETATION

Зачем нужна интерпретация моделей?



Объяснение работы модели

Важно для доверия к её выводам



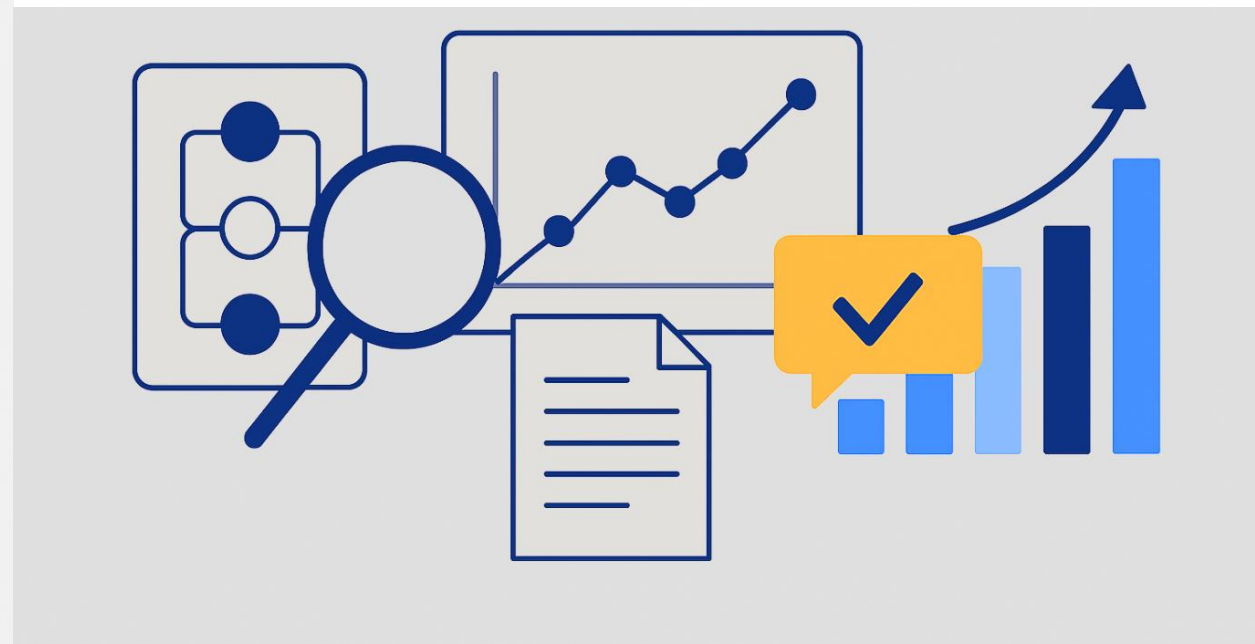
Соблюдение требований прозрачности

Необходимо для обоснования принятых решений



Анализ ошибок и отклонений

Позволяет выявить и исправить недостатки модели



Ключевые термины интерпретации



Глобальная интерпретация

Общая логика модели

Локальная интерпретация

Анализ отдельных прогнозов

Важность признаков

Значимость факторов в модели



Методы глобальной интерпретации



Анализ структуры моделей

Decision Trees, правила



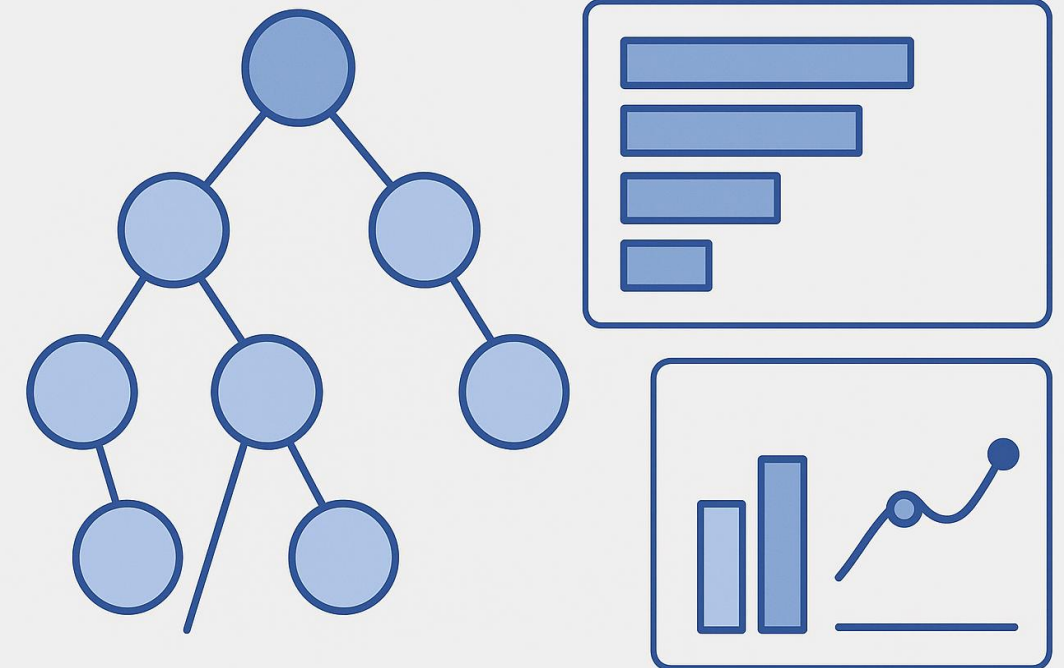
Оценка значимости признаков

Permutation importance, Feature importance



Анализ частичных зависимостей

Partial Dependence



Методы локальной интерпретации



LIME

Local
Interpretable
Model-agnostic
Explanations



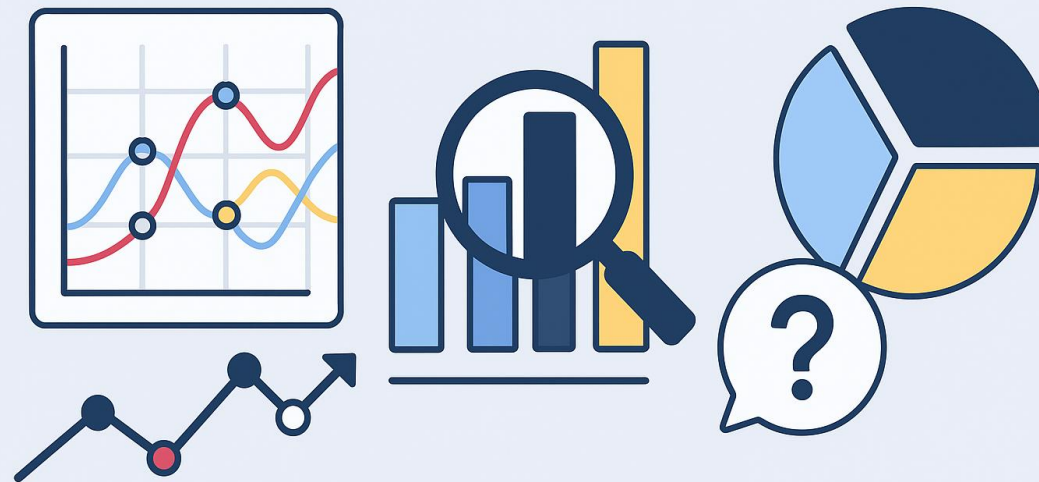
SHAP

SHapley Additive
exPlanations



What-if сценарии

Анализ отдельных
прогнозов





Преимущества деревьев решений для интерпретации



Простота

Дерево решений
является простым
методом



Понятность

Легко понять
логику принятия
решений



Визуализация

Легко
визуализируемый
метод



Прослеживаемость

Позволяет четко
проследить процесс
принятия решений



Выявление ключевых признаков

Помогает определить
важные факторы

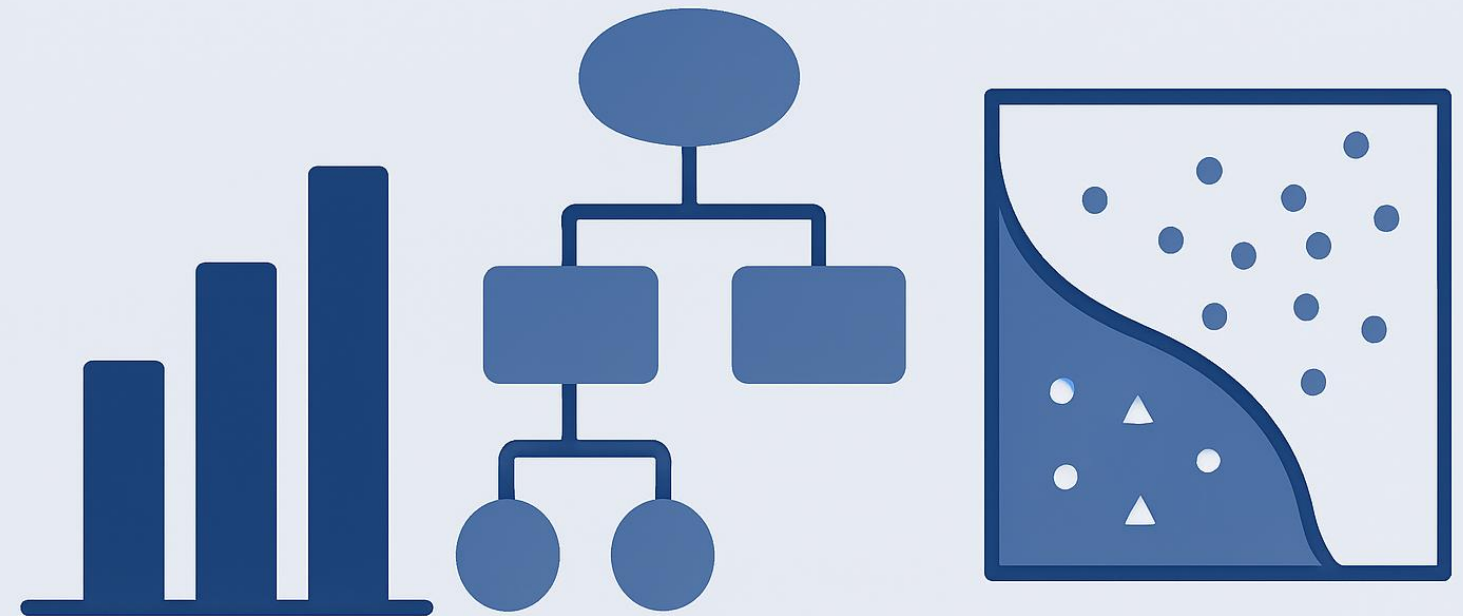
Важность признаков в деревьях решений

Определение важности признаков

Важность признаков в деревьях определяется по степени влияния признаков на снижение неопределенности

Методы оценки

- Information Gain
- Gini Importance



RapidMiner для анализа моделей



Визуализация структуры

Графическое представление моделей



Веса признаков

Анализ значимости факторов



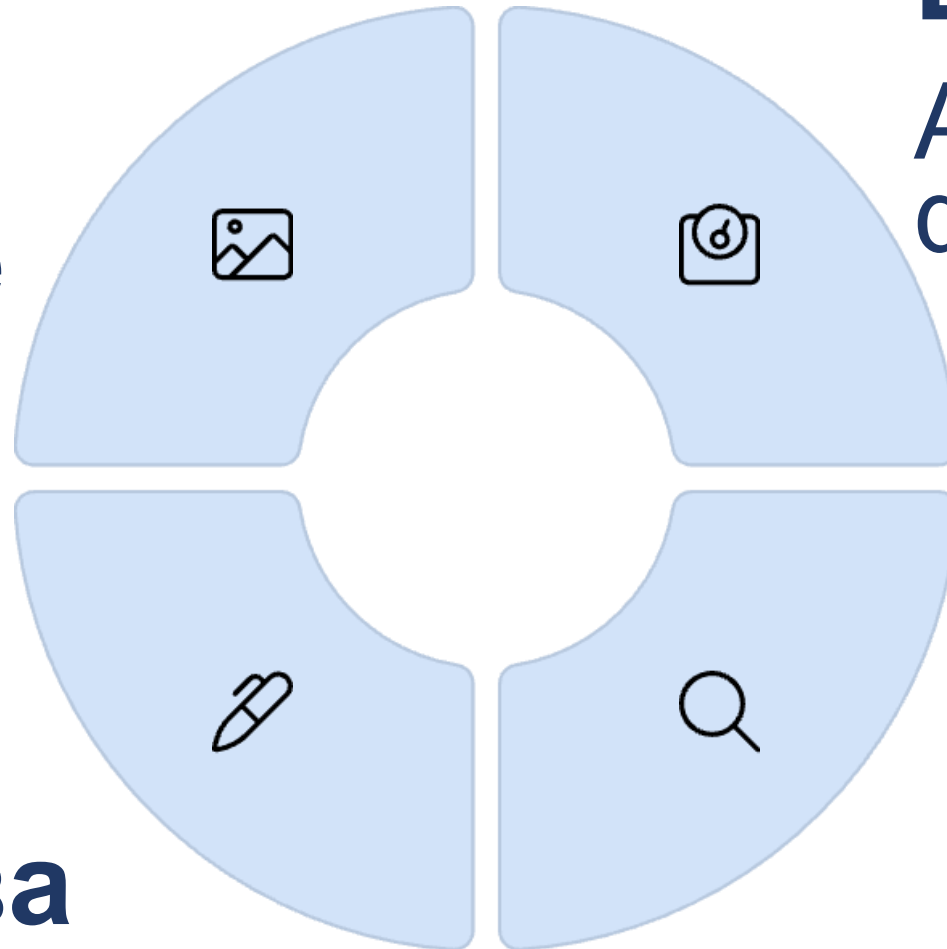
Оценка качества

Анализ точности прогнозов



Симуляция сценариев

Анализ "что если?"



Этапы подготовки данных для моделирования



Загрузка и проверка данных

Импорт данных и первичный анализ их структуры

Очистка данных

Удаление ненужных признаков и выбросов

Обработка пропущенных значений

Заполнение или удаление пропусков

Преобразование признаков

Кодирование категориальных и нормализация числовых признаков

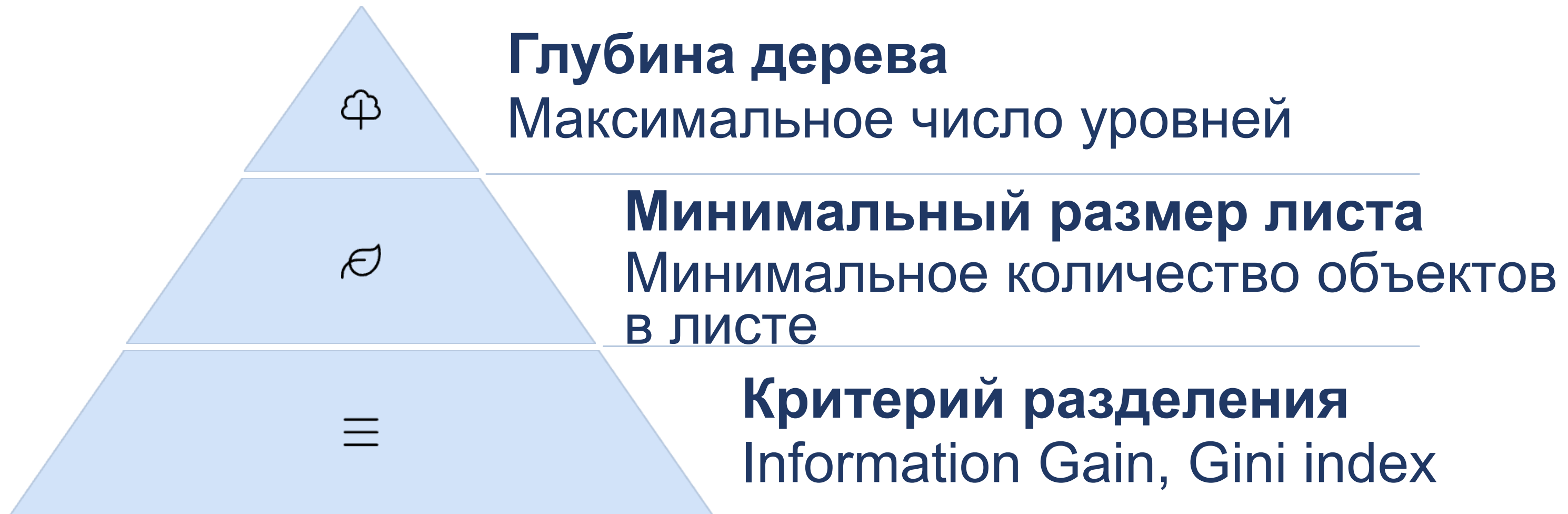
Типы признаков и их обработка



Тип признака	Метод обработки	Назначение
Категориальные	Преобразование (binominal, polynomial)	Подготовка для алгоритмов
Числовые	Стандартизация	Приведение к единому масштабу
Числовые	Нормализация	Приведение к диапазону [0,1]

Это необходимо для корректной работы моделей.

Построение дерева решений в RapidMiner



Критерии оценки качества классификации

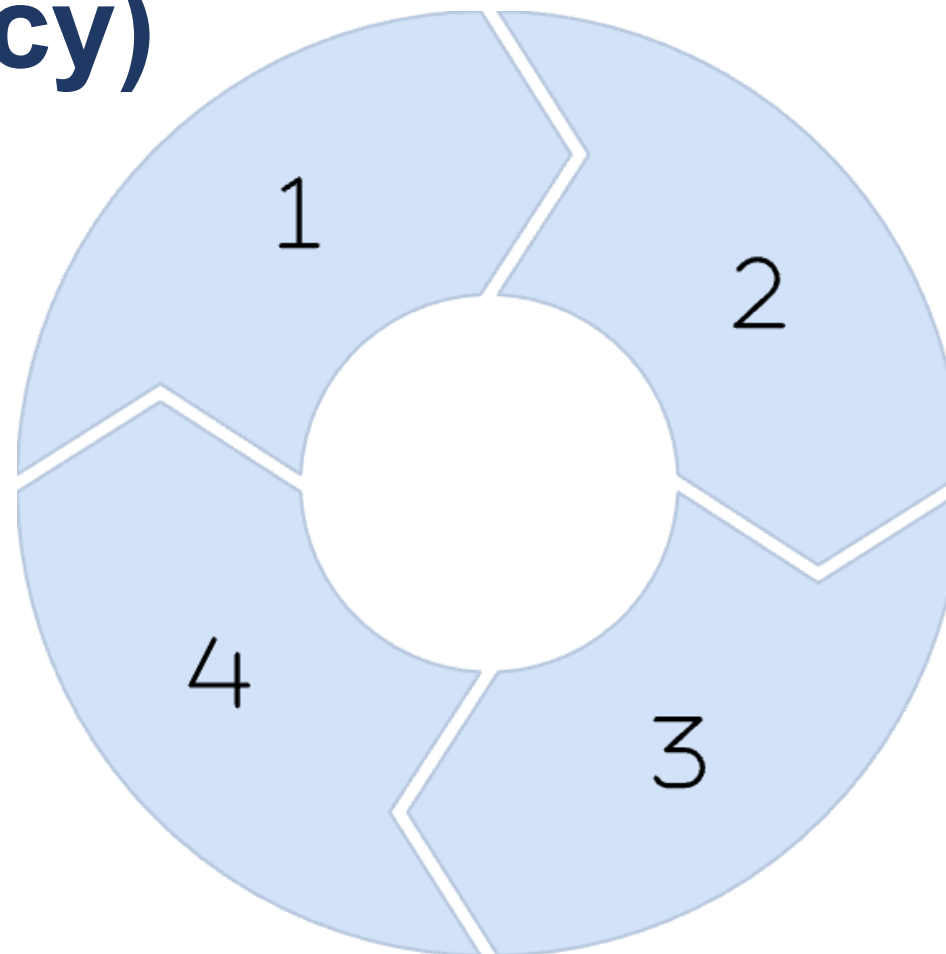


Точность (accuracy)

Доля правильных прогнозов

F1-score

Гармоническое среднее precision и recall



Полнота (recall)

Доля найденных объектов класса

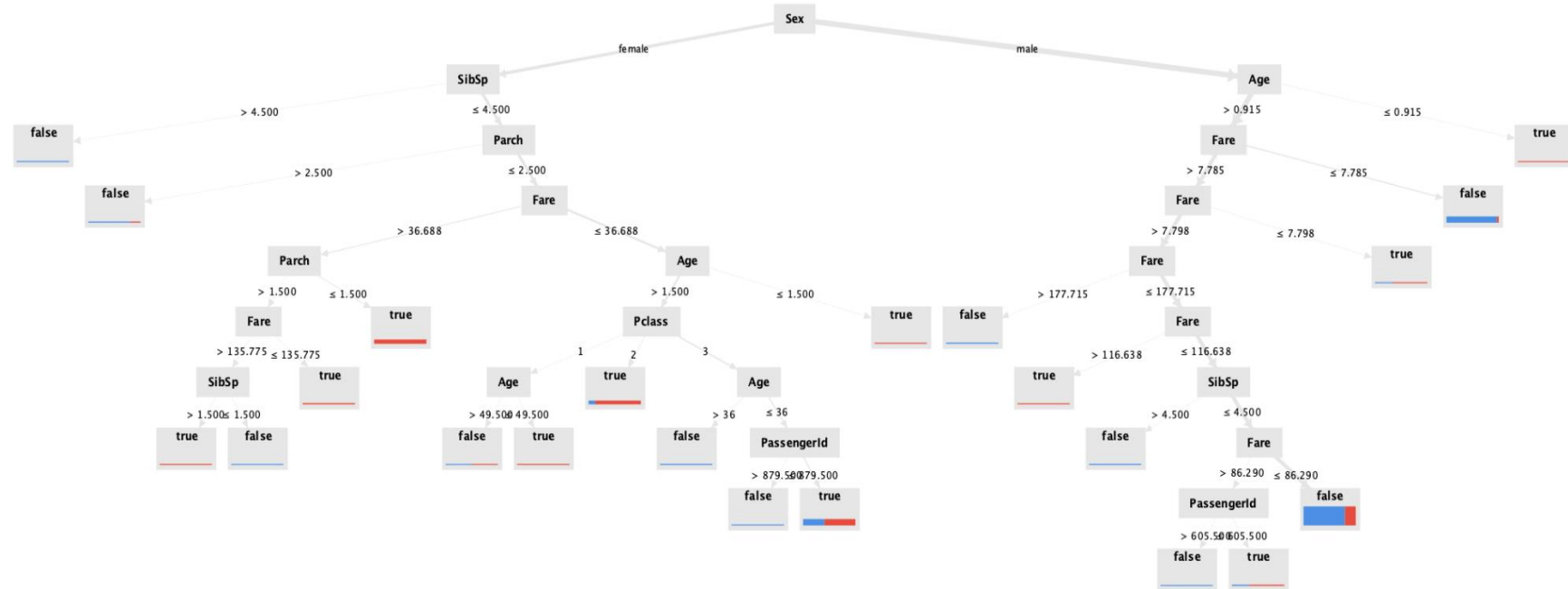
Точность (precision)

Доля правильных среди положительных прогнозов



Визуальные инструменты интерпретации моделей

RapidMiner включает
встроенные средства
визуализации
деревьев решений,
важности признаков и
анализ результатов
через интерактивные
инструменты

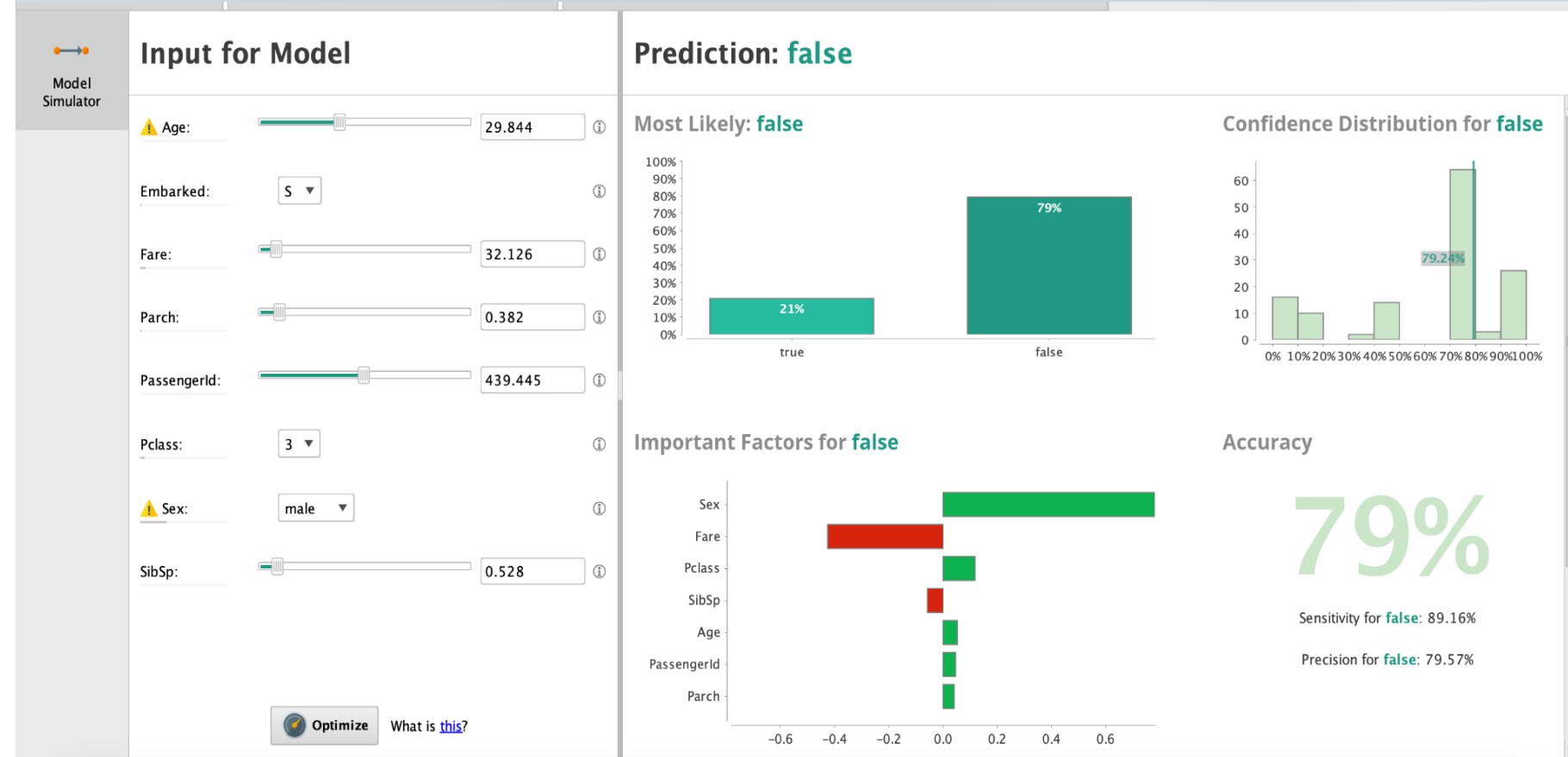




Model Simulator в RapidMiner

Возможности Model Simulator

- Исследование реакции модели на изменение признаков в реальном времени
- Визуализация структуры модели
- Проверка стабильности модели



Анализ значимости признаков (AttributeWeights)



Вклад признаков

Показывает вклад каждого признака в решение



Ключевые факторы

Помогает выявить важные факторы



Влияние на прогноз

Определяет факторы, влияющие на результат



Data



Weight
Visualizations



Plot
view

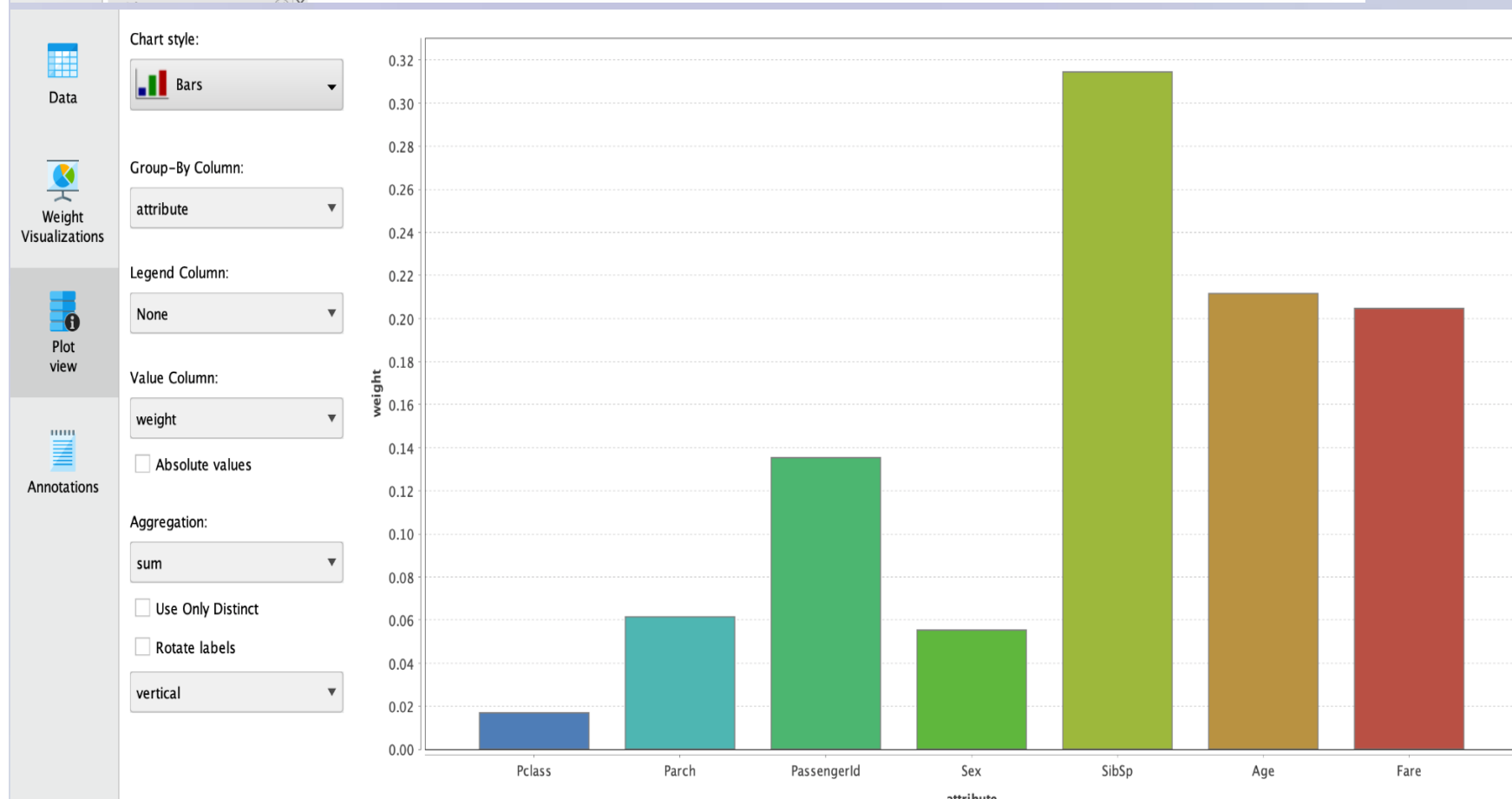
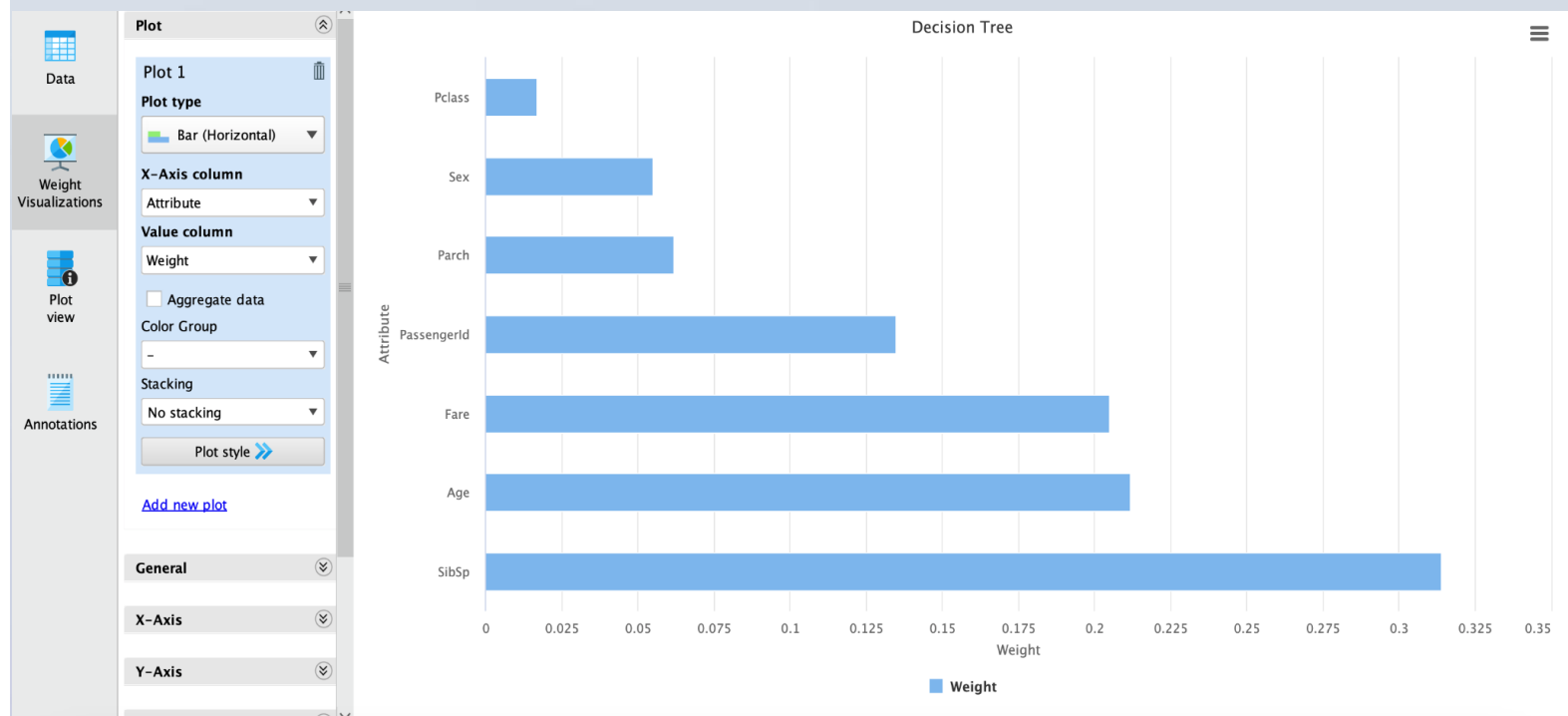


Annotations

attribute	weight
Pclass	0.017
Parch	0.062
Passeng...	0.135
Sex	0.055
SibSp	0.314
Age	0.212
Fare	0.205

Визуализация значимости признаков

Использование
столбчатых диаграмм
и графиков облегчает
восприятие значимости
признаков и упрощает
понимание логики
принятия решений
моделью.

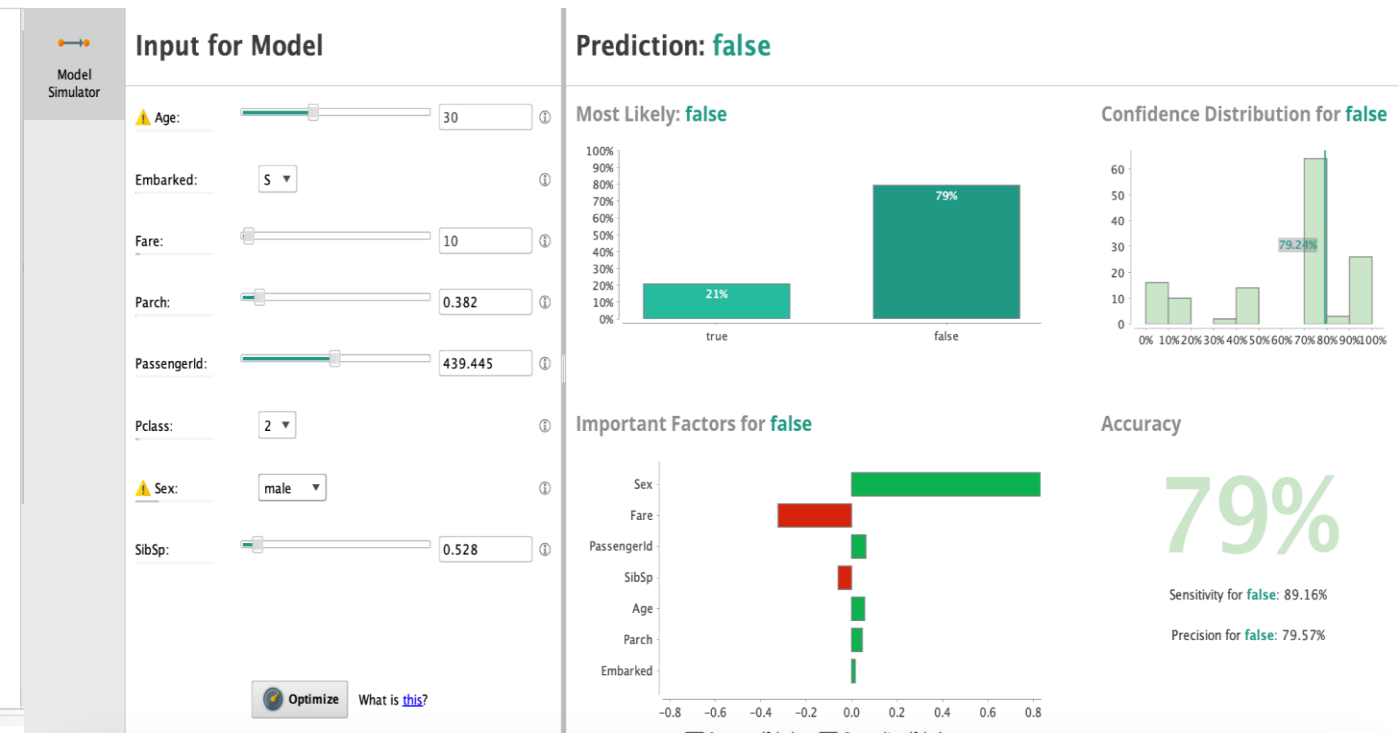
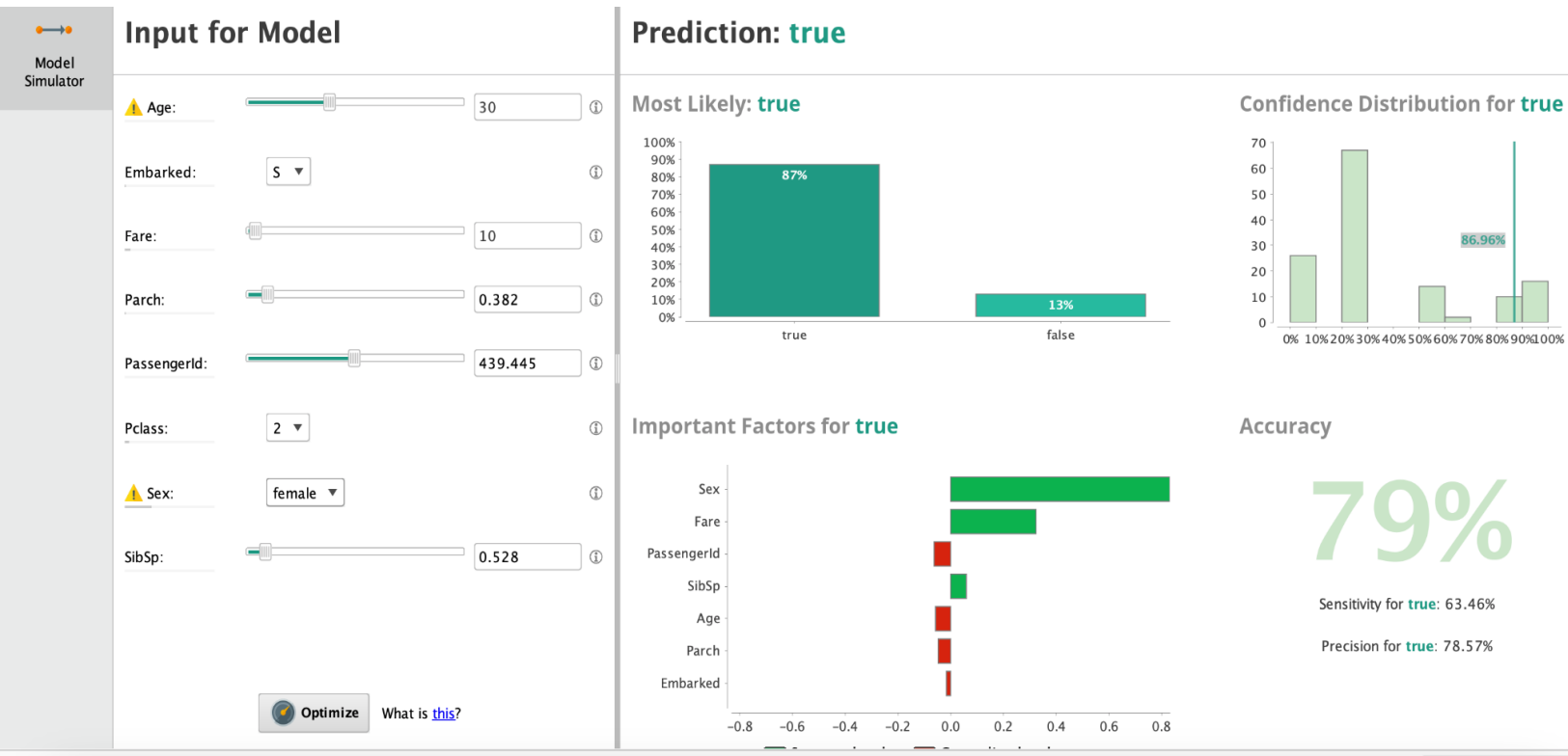


Анализ сценариев «Что если?»





Анализ сценариев «Что если?»



Исходные предсказания

Предсказания модели,
которые будут получены
при модификации
признаков (female – male)

Понятие переобучения (Overfitting)

Переобучение — проблема машинного обучения, возникающая когда модель слишком точно "запоминает" обучающие данные вместо выявления общих закономерностей.



Причины возникновения

Чрезмерная сложность модели или недостаточное количество обучающих примеров



Симптомы переобучения

Высокая точность на обучающей выборке, но значительное снижение качества прогнозов на новых, ранее не встречавшихся данных



Последствия

Неспособность модели к обобщению и потеря практической ценности для реальных задач прогнозирования



Управление сложностью моделей

1

Регулирование
глубины
дерева

Ограничение
максимального
количества
уровней

2

Контроль
размера
листьев

Установка
минимального
числа объектов
в листе

3

Кросс-валидация
Выбор оптимальных
параметров модели

Process	Parameters
Decision Tree	
criterion	gain ratio
maximal depth	11
<input checked="" type="checkbox"/> apply pruning	
confidence	0.1
<input checked="" type="checkbox"/> apply prepruning	
minimal gain	0.01
minimal leaf size	4
<i>minimal size for split</i>	4
<i>number of prepruning alternatives</i>	3



Сравнение моделей по качеству

Сравнение
моделей с
различными
гиперпараметрами
позволяет выбрать
оптимальный
баланс сложности и
качества прогноза
по различным
метрикам

The screenshot shows a software interface for model evaluation. On the left is a sidebar with three icons: a percentage sign for 'Performance', a document for 'Description', and a list for 'Annotations'. The 'Performance' section is active, showing a 'Criterion' list with 'accuracy' selected. To the right of the sidebar, there are radio buttons for 'Table View' (selected) and 'Plot View'. Below these, the text 'accuracy: 86.47%' is displayed. The main area contains a table with four columns: 'pred.' (predicted), 'true false', 'true true', and 'class precision'. The rows are 'pred. false', 'pred. true', and 'class recall'.

	true false	true true	class precision
pred. false	81	17	82.65%
pred. true	1	34	97.14%
class recall	98.78%	66.67%	

Интерпретация на примере датасета «Титаник»

На датасете Титаника модель выделяет важность таких признаков как пол, возраст и класс билета, что соответствует историческим данным. Анализ показывает значительную разницу в шансах на выживание:

- Женщина из 1-го класса: вероятность выживания 80%
- Женщина из 3-го класса: вероятность выживания лишь 44%

Input for Model

⚠ Age: ⓘ

Embarked: ⓘ

Fare: ⓘ

Parch: ⓘ

PassengerId: ⓘ

Pclass: ⓘ

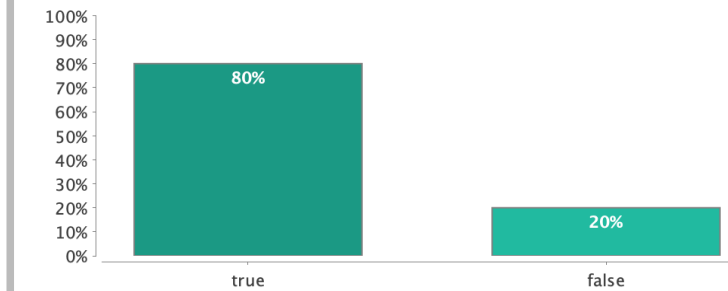
⚠ Sex: ⓘ

SibSp: ⓘ

Prediction: **true**



Most Likely: **true**



Important Factors for **true**



Input for Model

⚠ Age: ⓘ

Embarked: ⓘ

Fare: ⓘ

Parch: ⓘ

PassengerId: ⓘ

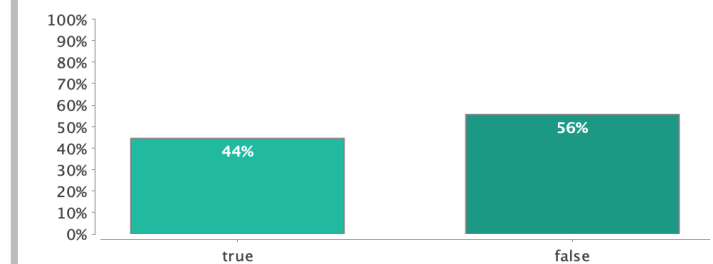
Pclass: ⓘ

⚠ Sex: ⓘ

SibSp: ⓘ

Prediction: **false**

Most Likely: **false**



Important Factors for **false**



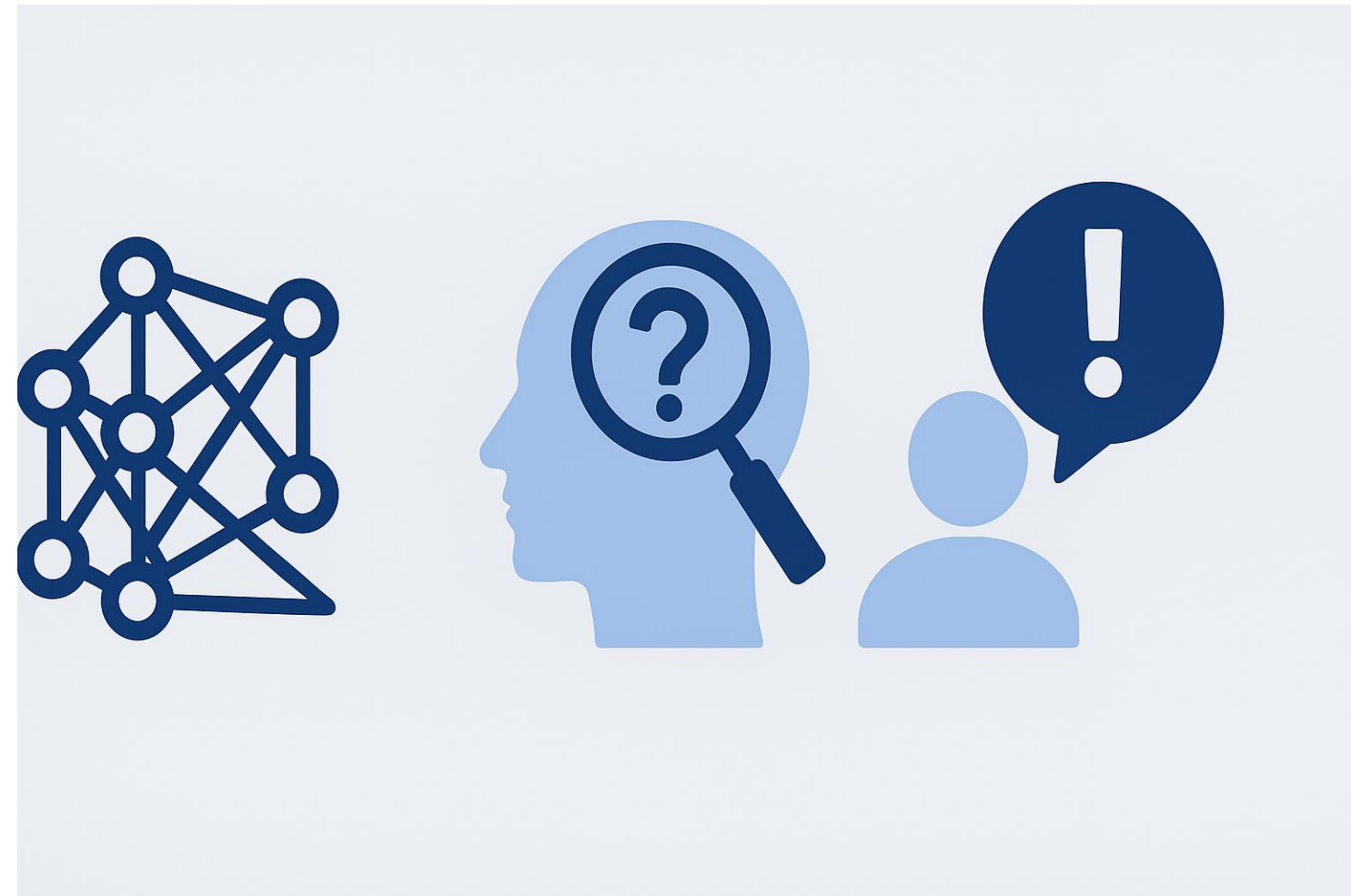
Ограничения интерпретации моделей



Сложность интерпретации некоторых моделей (нейросети, ансамбли)

Субъективность визуального анализа

Необходимость глубокого понимания контекста задачи



Перспективы развития интерпретации моделей



Интерпретация сложных моделей

Развитие инструментов для объяснения работы нейронных сетей и ансамблевых методов.



Автоматизация процессов

Создание систем для автоматического формирования интерпретаций без участия экспертов.



Повышение объективности

Разработка метрик и методологий для устранения субъективности при анализе моделей.