

Лабораторная работа №4

Тема: Байесовские сети

Что такое Байесовские сети?

Определение: Графическая вероятностная модель, представляющая множество случайных переменных и их условных зависимостей в виде направленного ациклического графа.

Назначение: Моделирование вероятностных зависимостей между переменными для решения задач классификации, диагностики и принятия решений при неопределенности.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Датасет Mushroom Classification

Источник: UCI Machine Learning Repository (Kaggle)

Размер: 8,124 записи, 23 атрибута

Задача: Бинарная классификация (съедобный/ядовитый)

Особенность: Все атрибуты категориальные

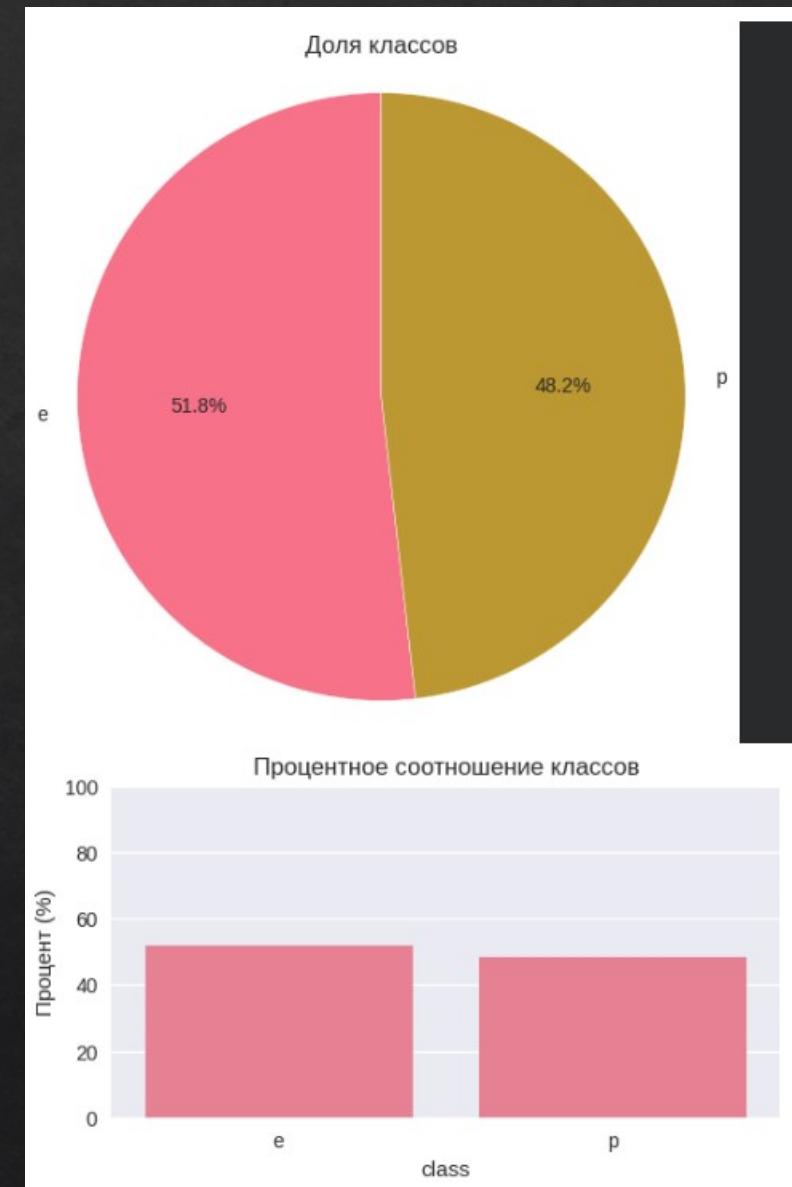
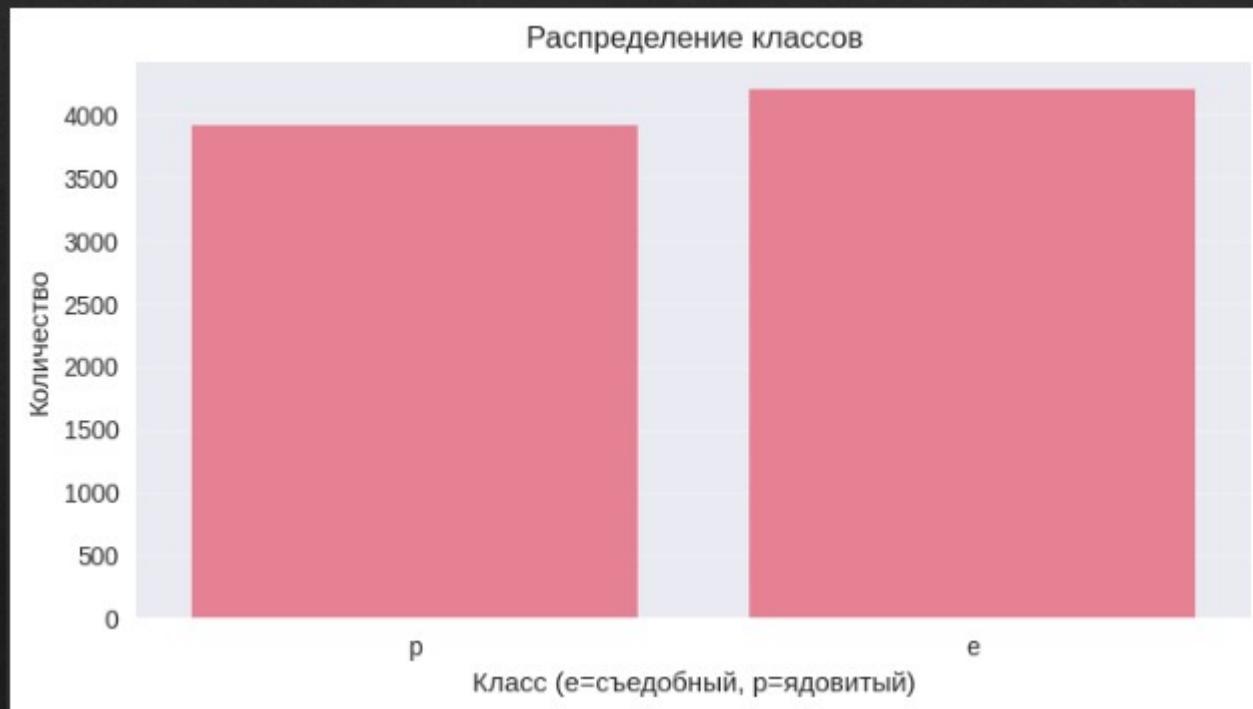
Датасет содержит признаки:

- class: съедобный (e) или ядовитый (p) - целевая переменная
- cap-shape, cap-surface, cap-color: характеристики шляпки
- bruises: наличие синяков
- odor: запах
- gill-attachment, gill-spacing, gill-size, gill-color: характеристики пластинок
- stalk-shape, stalk-root, stalk-surface-above-ring, stalk-surface-below-ring: характеристики ножки
- veil-type, veil-color: характеристики покрывала
- ring-number, ring-type: характеристики кольца
- spore-print-color: цвет спорового отпечатка
- population, habitat: популяция и среда обитания

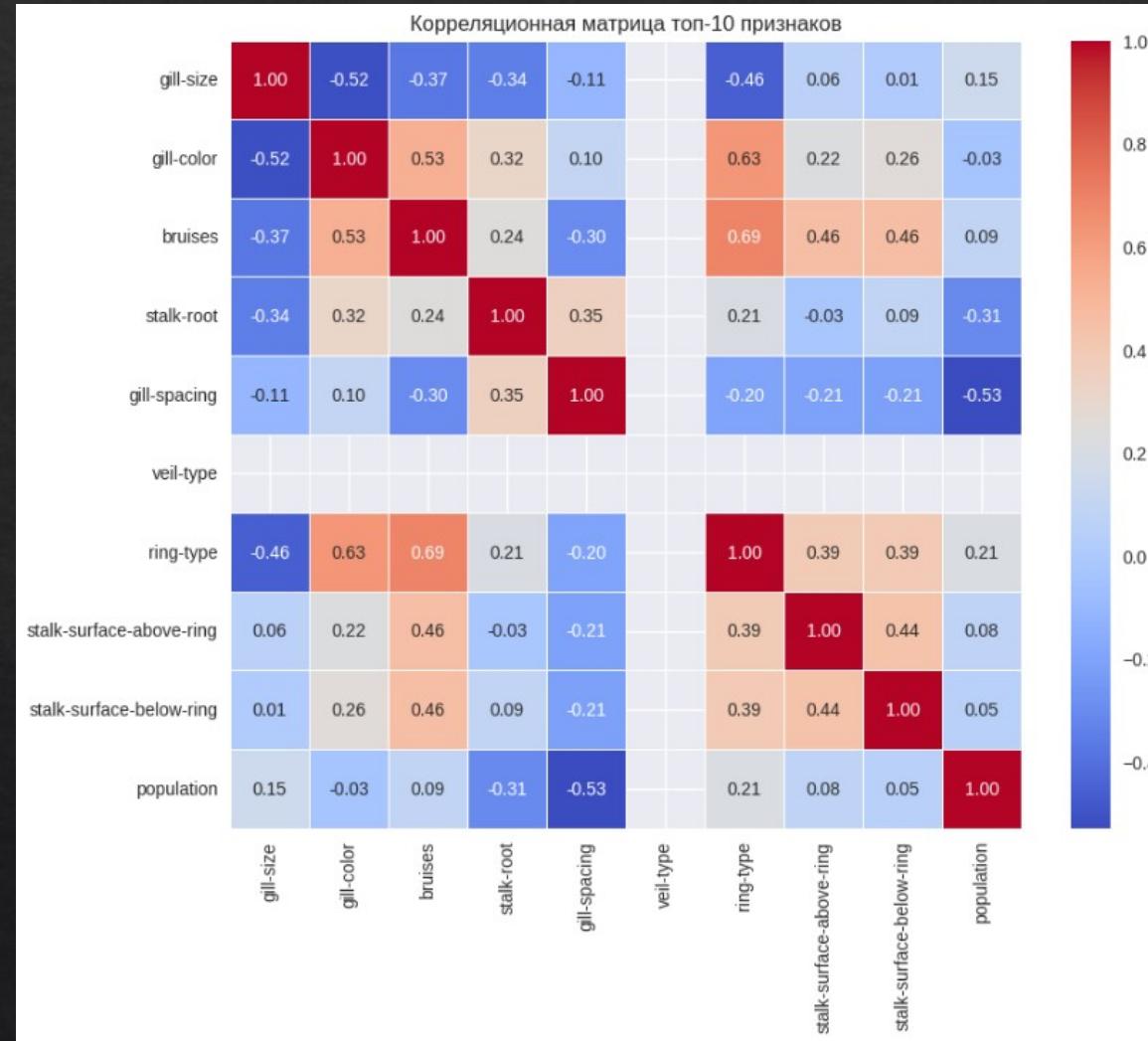
Загрузка датасета и вывод информации

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8124 entries, 0 to 8123
Data columns (total 23 columns):
 #   Column           Non-Null Count  Dtype  
 --- 
 0   class            8124 non-null   object 
 1   cap-shape        8124 non-null   object 
 2   cap-surface      8124 non-null   object 
 3   cap-color         8124 non-null   object 
 4   bruises          8124 non-null   object 
 5   odor             8124 non-null   object 
 6   gill-attachment  8124 non-null   object 
 7   gill-spacing     8124 non-null   object 
 8   gill-size         8124 non-null   object 
 9   gill-color        8124 non-null   object 
 10  stalk-shape      8124 non-null   object 
 11  stalk-root        8124 non-null   object 
 12  stalk-surface-above-ring 8124 non-null   object 
 13  stalk-surface-below-ring 8124 non-null   object 
 14  stalk-color-above-ring 8124 non-null   object 
 15  stalk-color-below-ring 8124 non-null   object 
 16  veil-type        8124 non-null   object 
 17  veil-color       8124 non-null   object 
 18  ring-number      8124 non-null   object 
 19  ring-type         8124 non-null   object 
 20  spore-print-color 8124 non-null   object 
 21  population        8124 non-null   object 
 22  habitat           8124 non-null   object 
dtypes: object(23)
memory usage: 1.4+ MB
None
['class', 'cap-shape', 'cap-surface', 'cap-color', 'bruises', 'odor', 'gill-attachment', 'gill-spacing', 'gill-size', 'gill-color', 'stalk-shape', 'stalk-root', 'stalk-surface-above-ring', 'stalk-surface-below-ring', 'stalk-color-above-ring', 'stalk-color-below-ring', 'veil-type', 'veil-color', 'ring-number', 'ring-type', 'spore-print-color', 'population', 'habitat']
0 пропусков
```

Распределение классов



Корреляционная матрица топ 10 признаков



Построение байесовской сети мануальным и автоматическим способом

Структура байесовской сети:

Узлы (nodes): ['odor', 'class', 'gill-color', 'spore-print-color', 'bruises', 'ring-type']

Ребра (edges): [('odor', 'class'), ('odor', 'gill-color'), ('gill-color', 'class'), ('spore-print-color', 'class'), ('bruises', 'class'), ('bruises', 'gill-color'), ('ring-type', 'class')]

Количество узлов: 6

Количество ребер: 7

Working for n conditional variables: 5: 100%

5/5 [00:06<00:00, 1.27s/it]

Использован метод: РС

Узлы: ['bruises', 'spore-print-color', 'ring-type', 'gill-color', 'odor', 'class']

Исходных ребер: 16, допущенных: 12, отброшено: 4

- ring-type -> bruises (creates-cycle)
- odor -> ring-type (creates-cycle)
- odor -> bruises (creates-cycle)
- class -> odor (creates-cycle)

Количество узлов: 6

Количество ребер: 12

Автоматическая модель создана.

Оценка параметров и CPT

1. CPT для целевой переменной 'class' (съедобный/ядовитый):

bruises	...	bruises(1)
gill-color	...	gill-color(11)
odor	...	odor(8)
ring-type	...	ring-type(4)
spore-print-color	...	spore-print-color(8)
class(0)	...	0.5
class(1)	...	0.5

=====

2. CPT для признака 'odor' (запах):

odor(0)	0.0492368
odor(1)	0.0236337
odor(2)	0.265879
odor(3)	0.0492368
odor(4)	0.00443131
odor(5)	0.434269
odor(6)	0.0315116
odor(7)	0.070901
odor(8)	0.070901

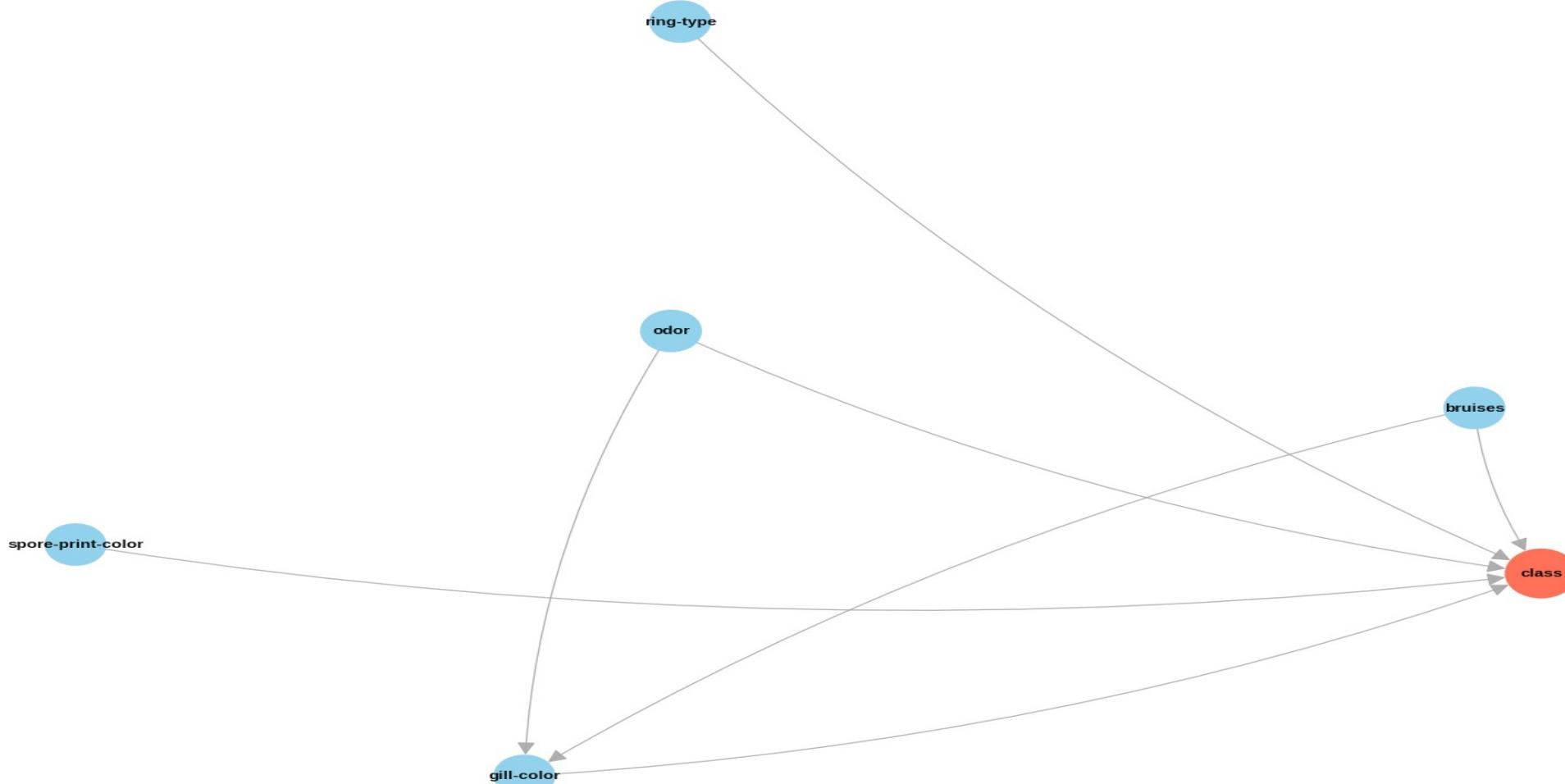
Оценка параметров и СРТ

```
=====  
3. СРТ для признака 'gill-color' (цвет пластинок):
```

bruises	...	bruises(1)
odor	...	odor(8)
gill-color(0)	...	0.0833333333333333
gill-color(1)	...	0.0833333333333333
gill-color(2)	...	0.0833333333333333
gill-color(3)	...	0.0833333333333333
gill-color(4)	...	0.0833333333333333
gill-color(5)	...	0.0833333333333333
gill-color(6)	...	0.0833333333333333
gill-color(7)	...	0.0833333333333333
gill-color(8)	...	0.0833333333333333
gill-color(9)	...	0.0833333333333333
gill-color(10)	...	0.0833333333333333
gill-color(11)	...	0.0833333333333333

Визуализация сети

Bayesian Network для классификации грибов
(Красный узел = 'class')



Выполнение inference

```
from pgmpy.inference import VariableElimination

infer = VariableElimination(model)

print("\nПример 1: Вероятность класса при известном запахе (odor)")
print("Запах без запаха:")
query1 = infer.query(variables=['class'], evidence={'odor': 0})
print(query1)

print("\nЗапах миндаля:")
query2 = infer.query(variables=['class'], evidence={'odor': 1})
print(query2)

print("\nЗапах креозота:")
query3 = infer.query(variables=['class'], evidence={'odor': 2})
print(query3)

print("\nСценарий 1: Гриб с запахом (odor=2) и темными пластинками (gill-color=4)")
query_scenario1 = infer.query(variables=['class'], evidence={'odor': 2, 'gill-color': 4})
print(query_scenario1)

print("\nСценарий 2: Гриб без синяков (bruises=0) и со светлыми пластинками (gill-color=0)")
query_scenario2 = infer.query(variables=['class'], evidence={'bruises': 0, 'gill-color': 0})
print(query_scenario2)

print("\nСценарий 3: Гриб с синяками (bruises=1) и черным споровым отпечатком (spore-print-color=0)")
query_scenario3 = infer.query(variables=['class'], evidence={'bruises': 1, 'spore-print-color': 0})
print(query_scenario3)
```

Применение Inference

Пример 1: Вероятность класса при известном запахе (odor)

Запах без запаха:

class	phi(class)
class(0)	0.5484
class(1)	0.4516

Запах миндаля:

class	phi(class)
class(0)	0.4325
class(1)	0.5675

Запах креозота:

class	phi(class)
class(0)	0.4641
class(1)	0.5359

Предсказание на тестовой сборке

Размер обучающей выборки: 6499

Размер тестовой выборки: 1625

Выполняем предсказания на тестовой выборке...

Точность (Accuracy): 0.5182

Отчет о классификации:

	precision	recall	f1-score	support
Съедобный	0.52	1.00	0.68	842
Ядовитый	0.00	0.00	0.00	783
accuracy			0.52	1625
macro avg	0.26	0.50	0.34	1625
weighted avg	0.27	0.52	0.35	1625

