Лабораторная работа №2
Задача классификации

Dataset: Abalone

Цель

Создать модель для бинарной классификации возраста моллюсков

Молодые (≤10 лет) vs Взрослые (>10 лет)

Исходные данные

Датасет: Физические характеристики морских моллюсков Abalone

- 4176 наблюдений
- 8 исходных признаков + пол (категориальный)

Признаки:

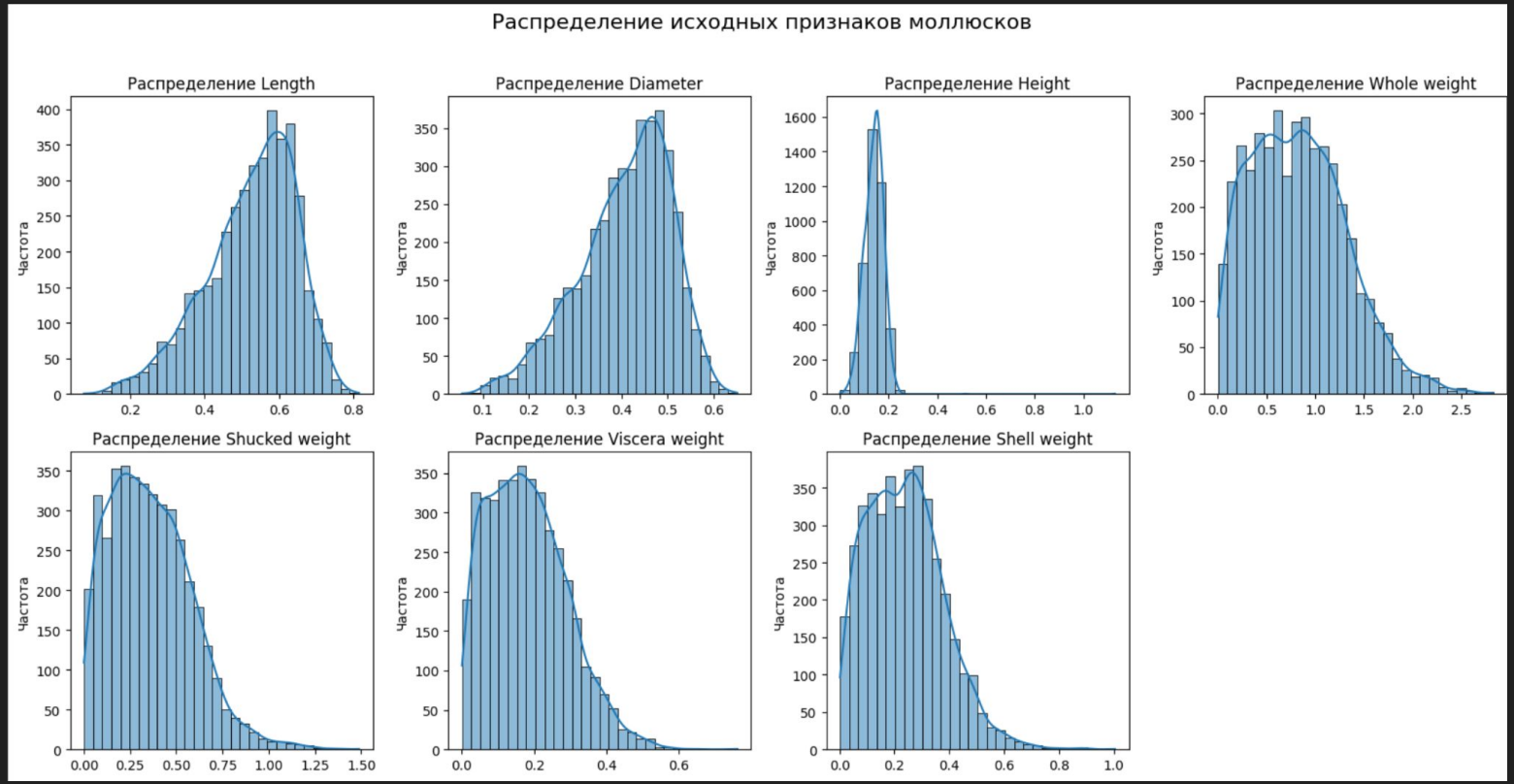
- Длина, диаметр, высота
- Веса: общий, без раковины, внутренностей, раковины
- Количество колец (целевая переменная для регрессии)

Подготовка данных

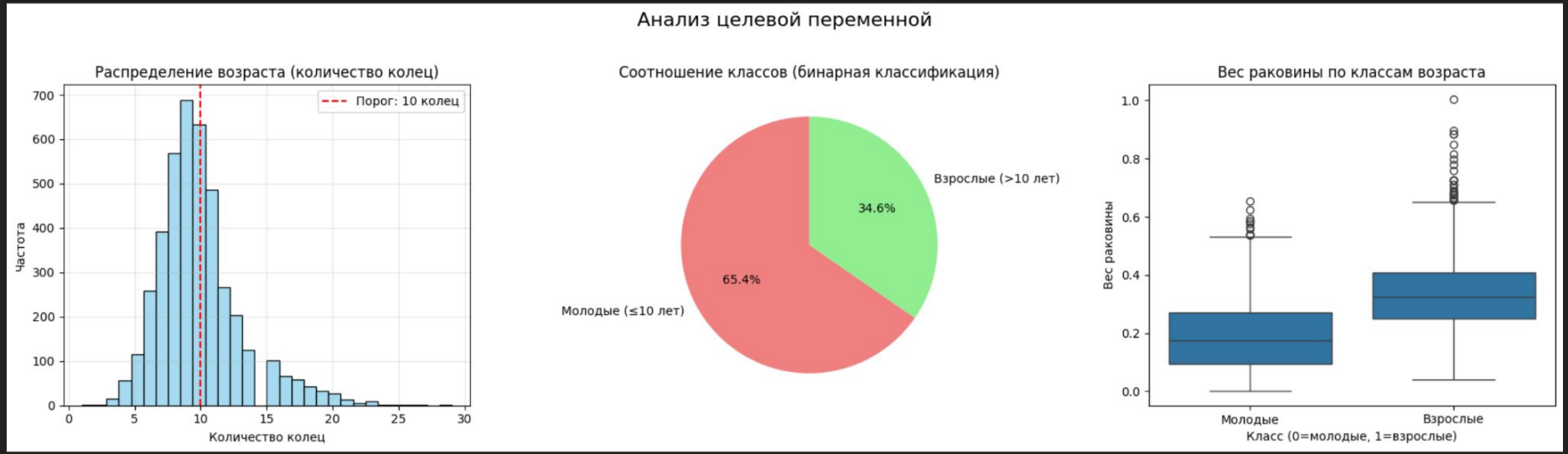
Создание новых признаков для улучшения качества модели:

1. share_of_meat – доля мяса от общего веса
2. share_of_shell – доля раковины
3. volume_approx – приблизительный объем
4. density – плотность моллюска
5. weight_per_ring – вес на одно кольцо

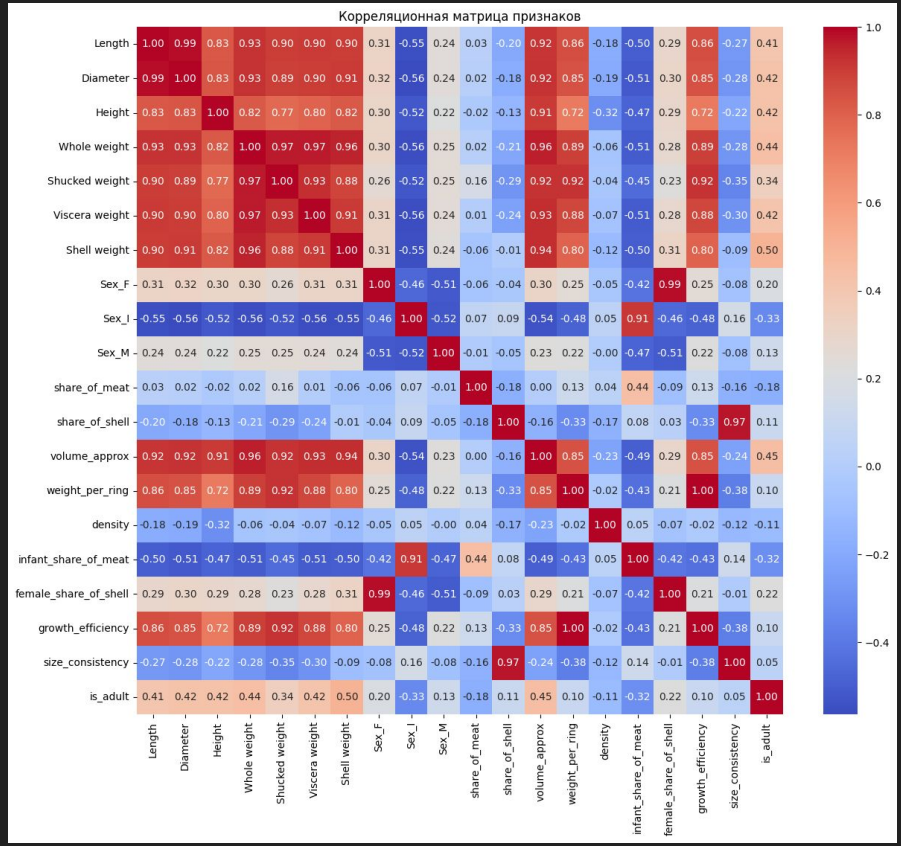
Визуализация распределения признаков



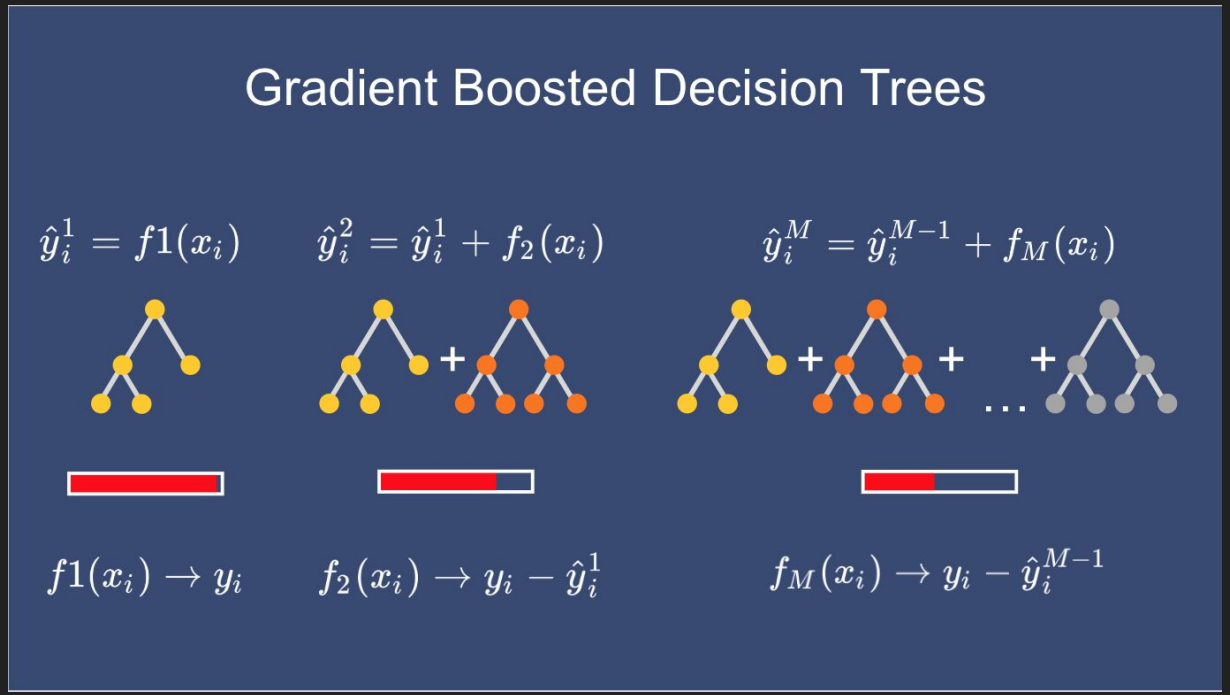
Визуализация распределения целевой переменной



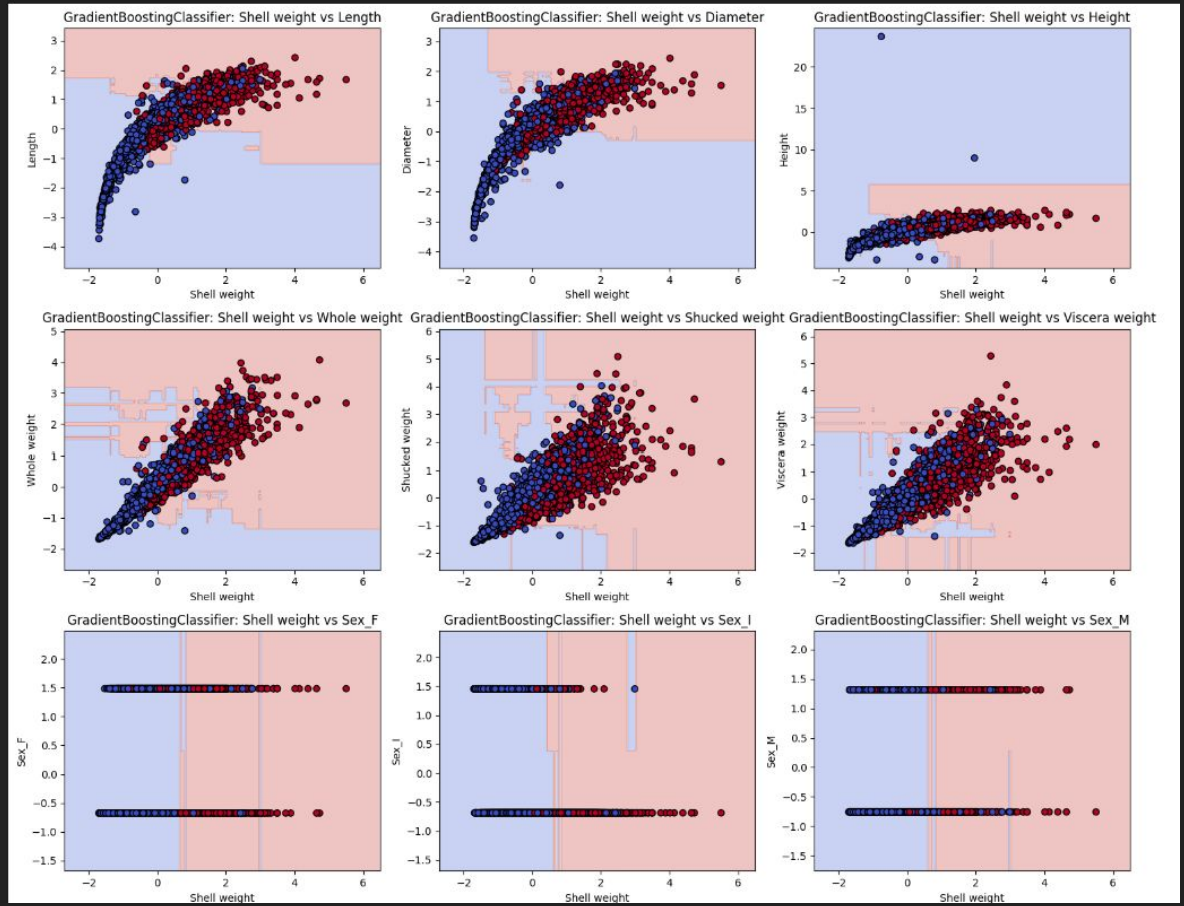
Корреляционная матрица



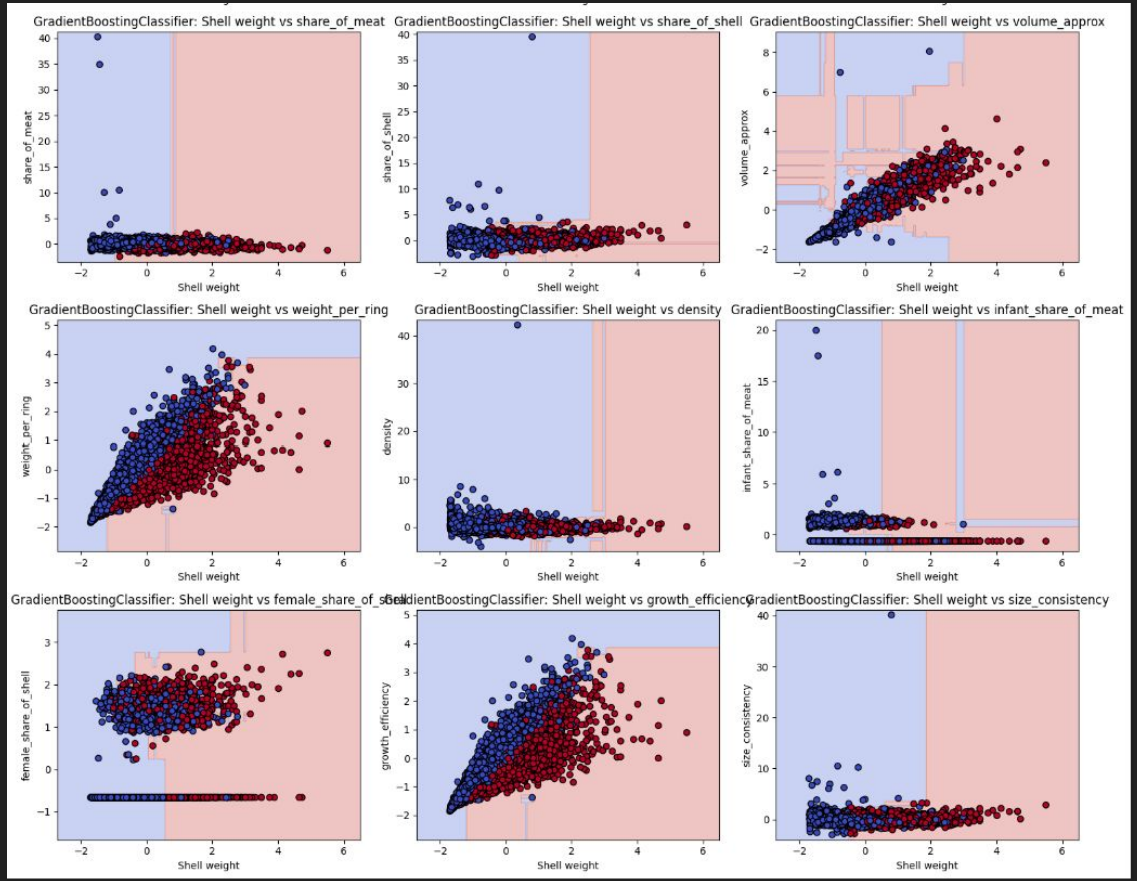
Выбор модели – Градиентный бустинг



Графическое представление решения



Графическое представление решения



Результаты обучения и сравнение

```
=== Boosting ===
Accuracy: 0.967 ± 0.001
F1-score: 0.951 ± 0.002
ROC-AUC: 0.994 ± 0.002
=== XGBoost ===
Accuracy: 0.968 ± 0.006
F1-score: 0.952 ± 0.010
ROC-AUC: 0.994 ± 0.001
```

```
=== SVM ===
Accuracy: 0.959 ± 0.006
F1-score: 0.936 ± 0.010
ROC-AUC: 0.996 ± 0.002
=== Logistic Regression ===
Accuracy: 0.954 ± 0.006
F1-score: 0.930 ± 0.010
ROC-AUC: 0.987 ± 0.003
=== KNN ===
Accuracy: 0.872 ± 0.004
F1-score: 0.799 ± 0.008
ROC-AUC: 0.926 ± 0.008
```

```
=== DesicionTree ===
Accuracy: 0.949 ± 0.005
F1-score: 0.925 ± 0.008
ROC-AUC: 0.941 ± 0.008
=== RandomForest ===
Accuracy: 0.964 ± 0.003
F1-score: 0.947 ± 0.005
ROC-AUC: 0.993 ± 0.001
```