

Анализ и обучение моделей машинного обучения

Дата создания: 2025-05-04 19:30:06

Содержание

1. Обзор проекта
2. Анализ данных
3. Предобработка данных
4. Обучение моделей
5. Метрики и оценка
6. Результаты
7. Выводы

1. Обзор проекта

Цель проекта:

- Разработка системы машинного обучения для классификации данных
- Оптимизация процесса обучения моделей
- Обеспечение интерпретируемости результатов

2. Анализ данных

Используемые данные:

- Промышленные данные о производстве продукции
- Временные ряды параметров производства
- Категориальные и числовые признаки
- Целевая переменная: качество продукции (0 - брак, 1 - годная продукция)

Распределение классов в данных

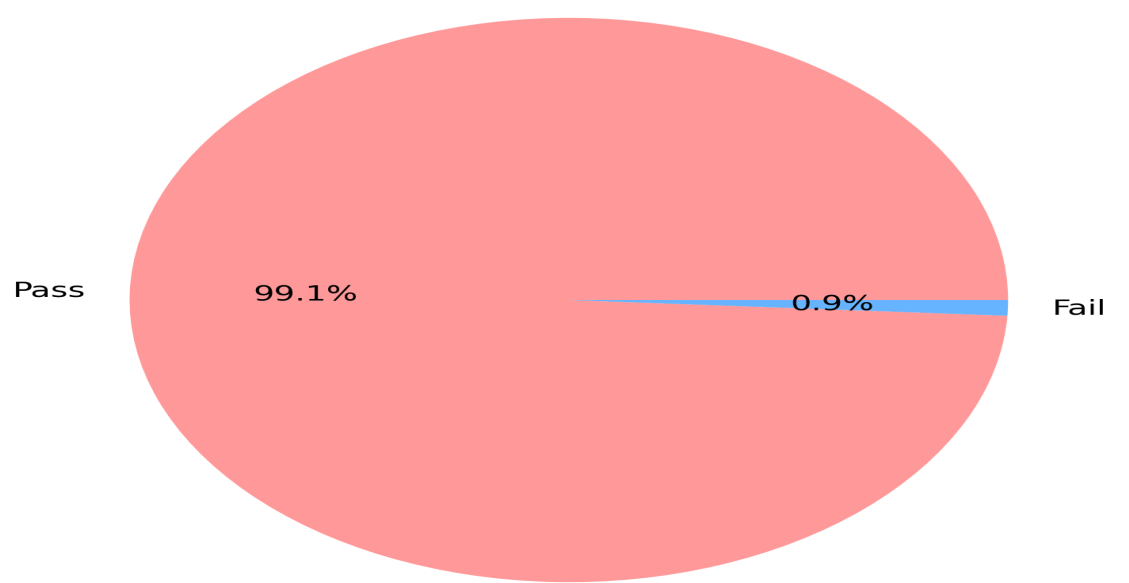


Рисунок 1. Распределение классов в данных

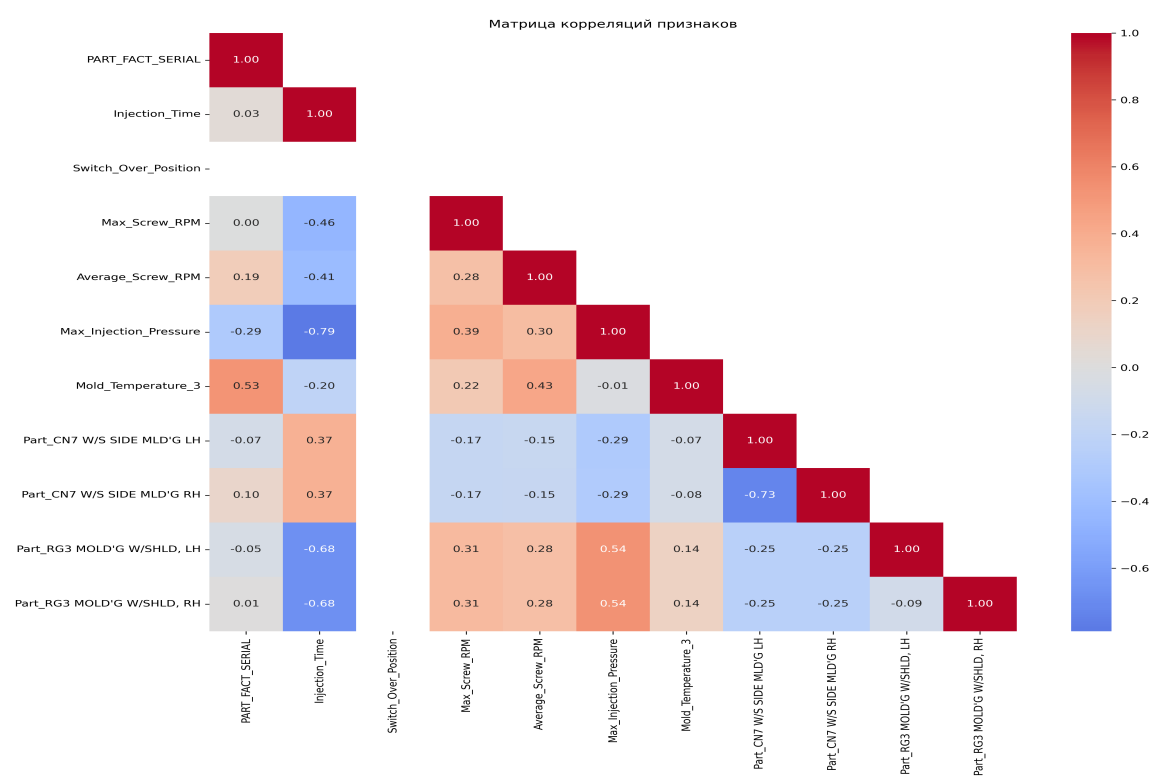


Рисунок 2. Матрица корреляций признаков

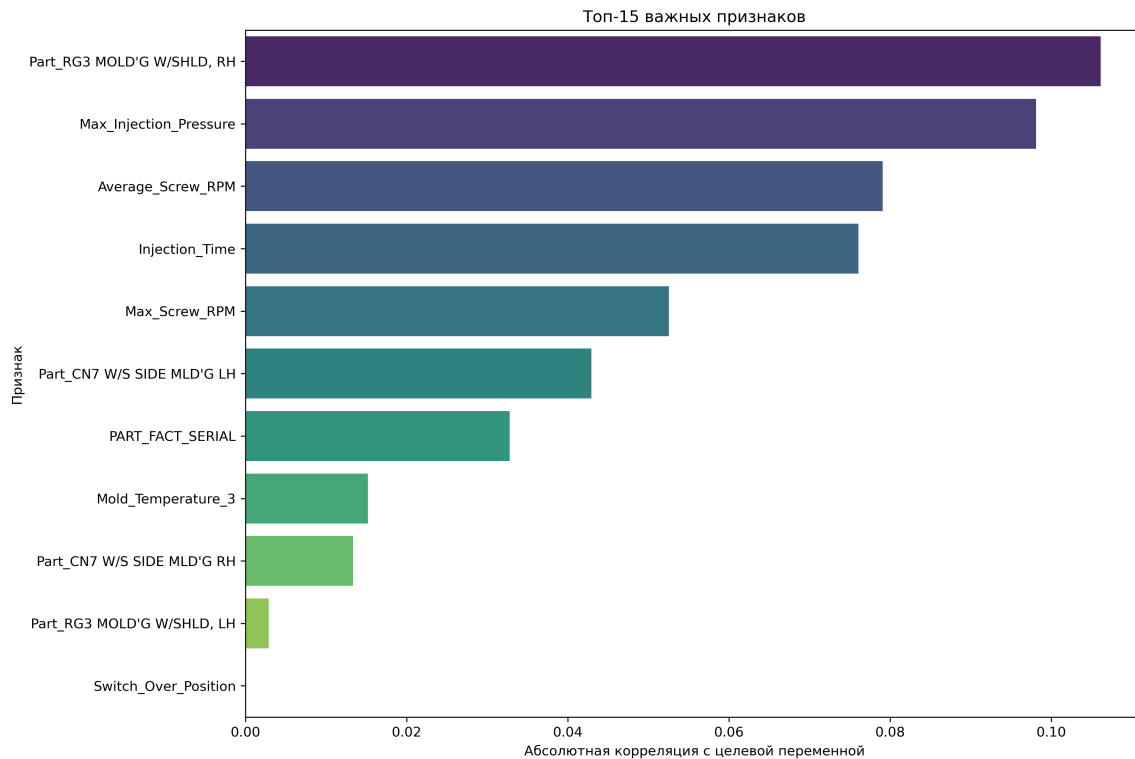


Рисунок 3. Важность признаков

3. Предобработка данных

Этапы предобработки:

- Очистка данных от пропущенных значений
- Кодирование категориальных признаков (One-Hot Encoding)
- Масштабирование числовых признаков (StandardScaler)
- Балансировка классов:
 - SMOTE
 - ADASYN
 - RandomUnderSampler

4. Обучение моделей

Сравнение метрик моделей:

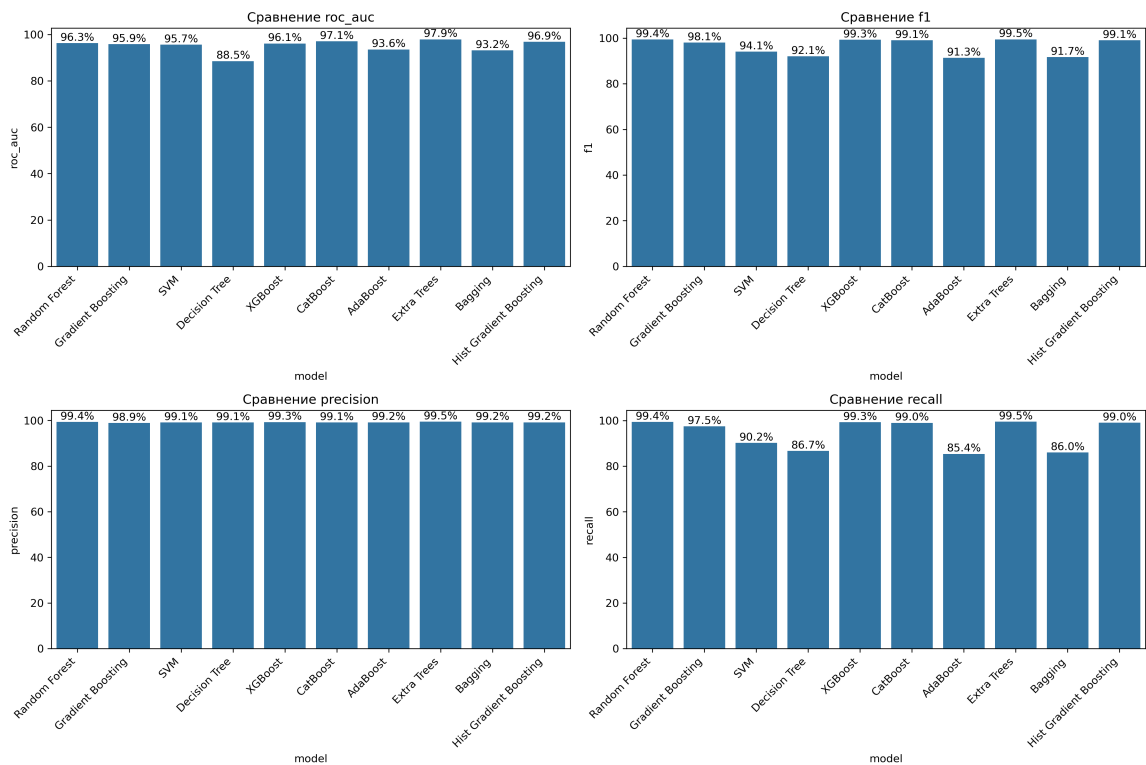


Рисунок 4. Сравнение метрик качества моделей

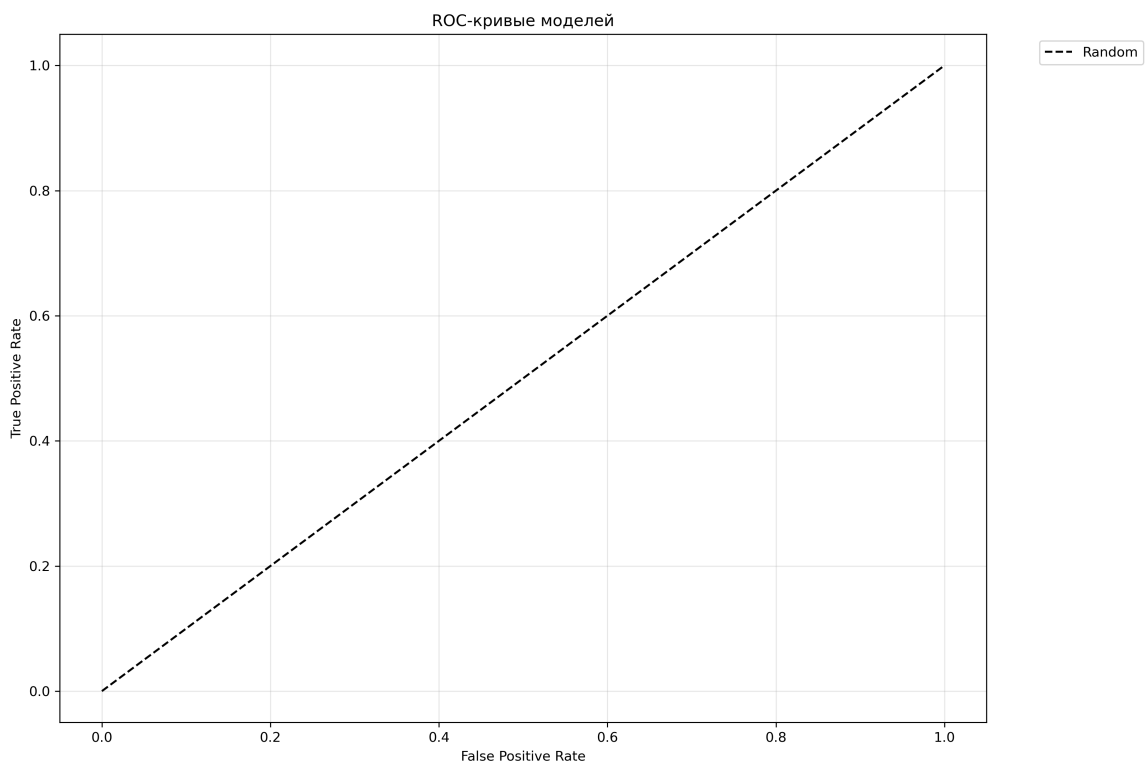


Рисунок 5. ROC-кривые моделей

5. Результаты

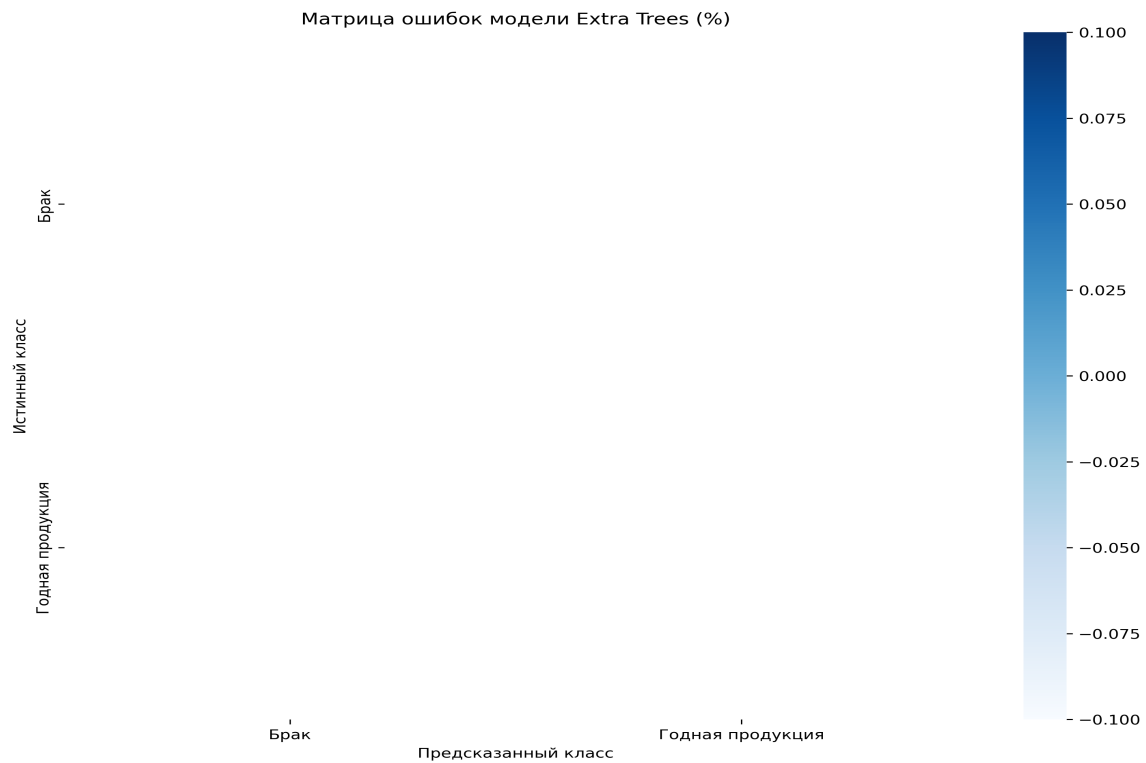


Рисунок 6. Матрица ошибок лучшей модели

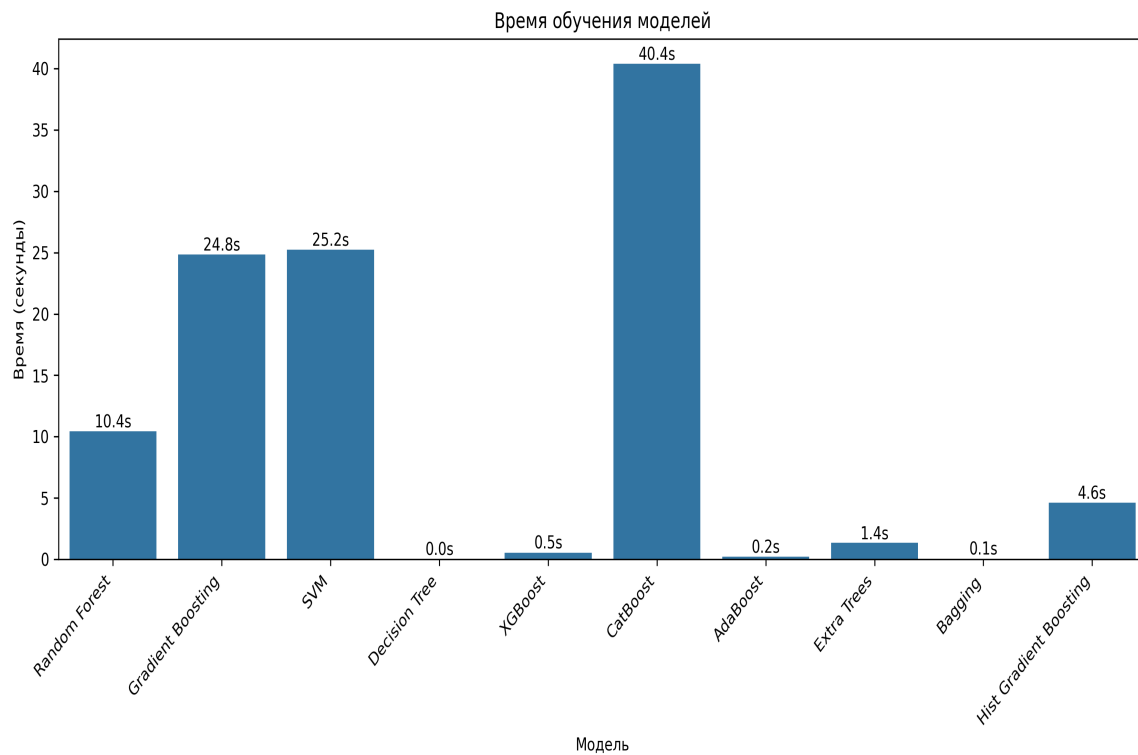


Рисунок 7. Время обучения моделей

6. Выводы

Основные результаты:

- Лучшая модель: Extra Trees
- Метод балансировки: ADASYN
- ROC AUC: 0.979
- F1-score: 0.995
- Precision: 0.995
- Recall: 0.995