

Выполнил: Тимошинов Егор Борисович

Группа: 16

Лабораторная работа

Деревья решений для классификации

Цель работы

Изучение методов построения деревьев решений для решения задач классификации. Построение дерева решений для предсказания возврата кредита на основе характеристик клиента.

Исходные данные

Для построения модели использовался набор данных о клиентах банка, включающий следующие признаки:

- Возраст клиента (количественный признак)
- Доход клиента (количественный признак)
- Кредитная история (категориальный признак: хорошая, средняя, плохая)
- Цель кредита (категориальный признак: потребительский, автомобиль, недвижимость, бизнес)
- Срок кредита в месяцах (количественный признак)

Целевая переменная: возврат кредита (Да/Нет).

Общее количество наблюдений: 1000. Данные разделены на обучающую выборку (70%) и тестовую выборку (30%).

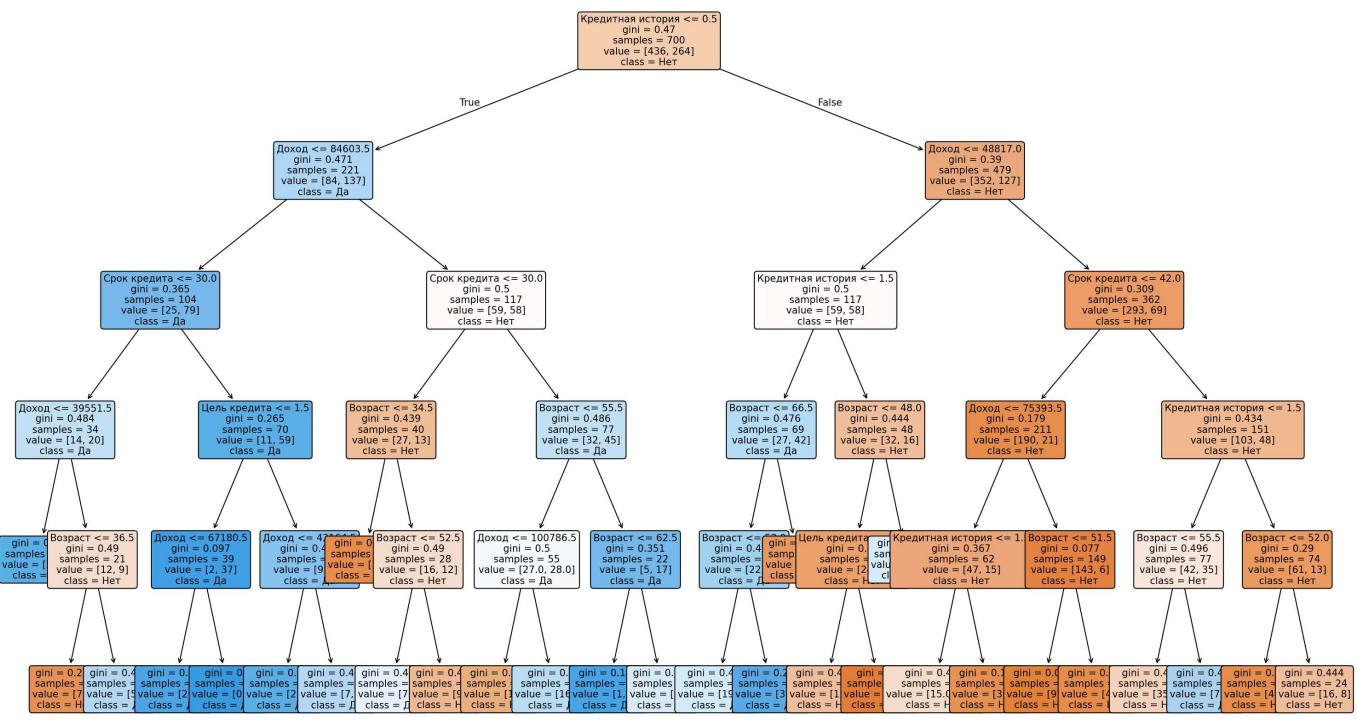
1. Построение дерева решений

Для построения дерева решений использовался алгоритм CART (Classification and Regression Trees) с критерием разделения Gini. Параметры модели:

- Максимальная глубина дерева: 5
- Минимальное количество образцов для разделения узла: 20
- Минимальное количество образцов в листе: 5

Построенное дерево решений представлено на рисунке ниже.

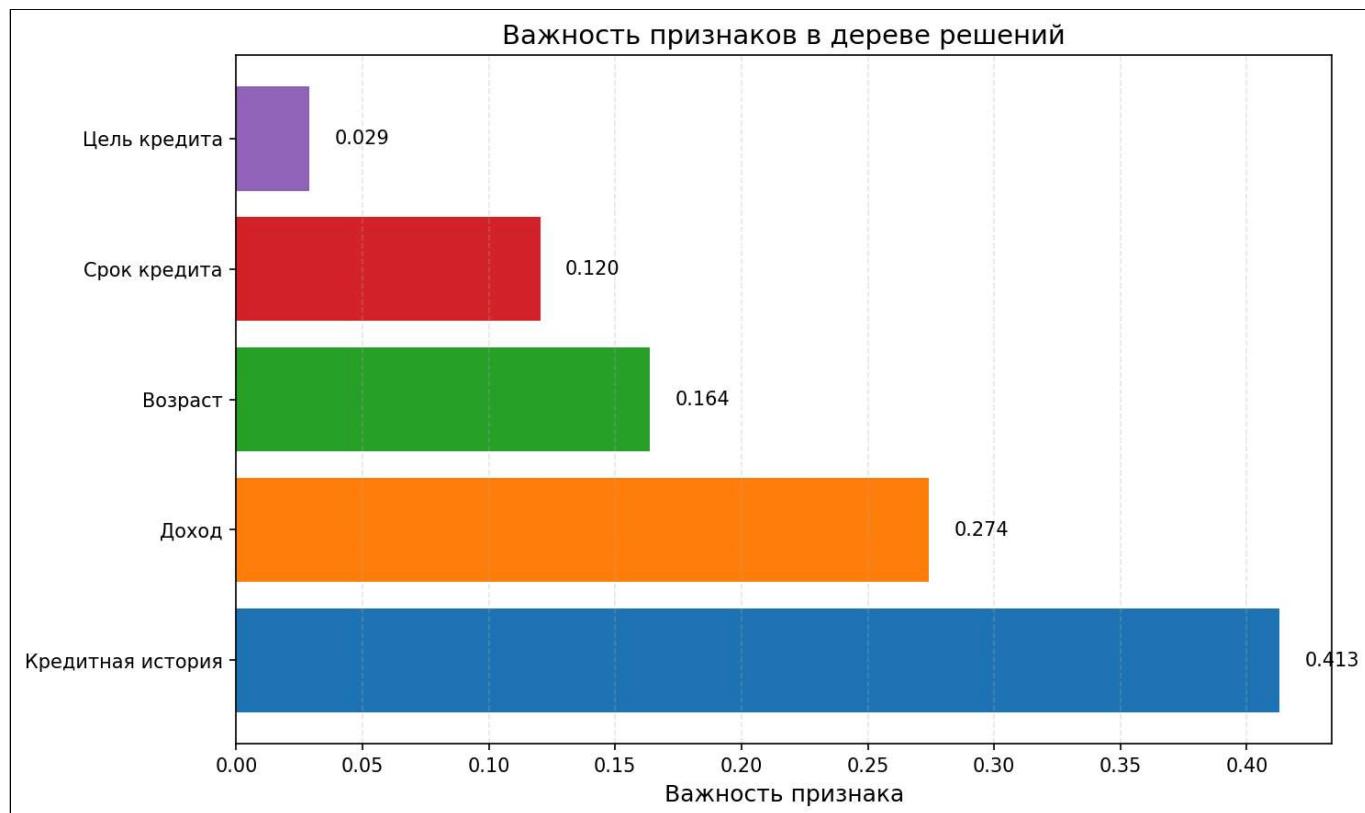
Дерево решений для классификации возврата кредита



2. Важность признаков

Важность признаков в построенном дереве решений:

Признак	Важность
Кредитная история	0.4131
Доход	0.2740
Возраст	0.1637
Срок кредита	0.1203
Цель кредита	0.0289



3. Классификация на тестовой выборке

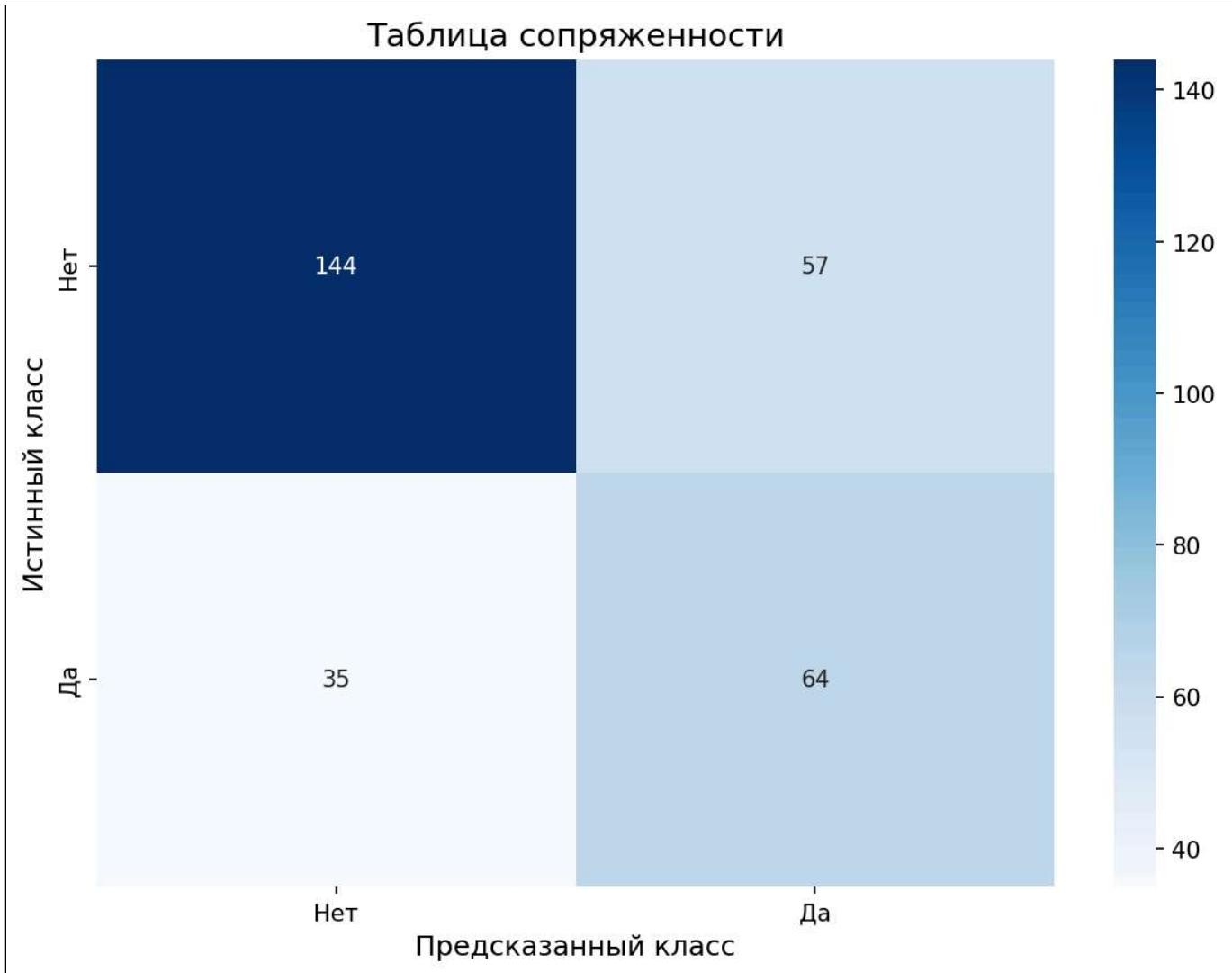
Для оценки качества построенной модели была выполнена классификация на тестовой выборке, содержащей 300 наблюдений.

4. Таблица сопряженности

Таблица сопряженности показывает количество правильно и неправильно классифицированных наблюдений:

	Предсказано: Нет	Предсказано: Да
Истинно: Нет	144	57
Истинно: Да	35	64

Таблица сопряженности



5. Оценка точности модели

Характеристики точности построенной модели:

Метрика	Значение
Точность (Accuracy)	0.6933 (69.33%)
Правильно классифицировано	208 из 300
Ошибок классификации	92 из 300

6. Детальный отчет о классификации

Класс	Precision	Recall	F1-score	Support
Нет	0.8045	0.7164	0.7579	201
Да	0.5289	0.6465	0.5818	99
Среднее (macro avg)	0.6667	0.6814	0.6699	300

Среднее (weighted avg)	0.7135	0.6933	0.6998	300
-------------------------------	--------	--------	--------	-----

7. Текстовая структура дерева

Ниже представлена текстовая структура построенного дерева решений (первые уровни):

```

|--- Кредитная история <= 0.50
|   |--- Доход <= 84603.50
|   |   |--- Срок кредита <= 30.00
|   |   |   |--- Доход <= 39551.50
|   |   |   |   |--- class: 1
|   |   |   |   |--- Доход > 39551.50
|   |   |   |   |   |--- Возраст <= 36.50
|   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- Возраст > 36.50
|   |   |   |   |   |   |--- class: 1
|   |   |--- Срок кредита > 30.00
|   |   |   |--- Цель кредита <= 1.50
|   |   |   |   |--- Доход <= 67180.50
|   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- Доход > 67180.50
|   |   |   |   |   |   |--- class: 1
|   |   |   |   |--- Цель кредита > 1.50
|   |   |   |   |   |--- Доход <= 42164.50
|   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- Доход > 42164.50
|   |   |   |   |   |   |--- class: 1
|--- Доход > 84603.50
|   |--- Срок кредита <= 30.00
|   |   |--- Возраст <= 34.50
|   |   |   |--- class: 0
|   |   |--- Возраст > 34.50
|   |   |   |   |--- Возраст <= 52.50
|   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- Возраст > 52.50
|   |   |   |   |   |--- class: 0
|   |--- Срок кредита > 30.00
|   |   |--- Возраст <= 55.50
|   |   |   |--- Доход <= 100786.50
|   |   |   |   |--- class: 0
|   |   |   |   |--- Доход > 100786.50
|   |   |   |   |   |--- class: 1
|   |   |   |   |--- Возраст > 55.50
|   |   |   |   |   |--- Возраст <= 62.50
|   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- Возраст > 62.50
|   |   |   |   |   |   |--- class: 1
|--- Кредитная история > 0.50
|   |--- Доход <= 48817.00
|   |   |--- Кредитная история <= 1.50
|   |   |   |--- Возраст <= 66.50
|   |   |   |   |--- Возраст <= 53.00
|   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- Возраст > 53.00
|   |   |   |   |   |--- class: 1
|   |   |   |   |--- Возраст > 66.50
|   |   |   |   |   |--- class: 0
|   |   |--- Кредитная история > 1.50
|   |   |   |--- Возраст <= 48.00
|   |   |   |   |--- Цель кредита <= 1.50
|   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- Цель кредита > 1.50
|   |   |   |   |   |--- class:...

```

Выводы

В ходе выполнения лабораторной работы было построено дерево решений для классификации возврата кредита. Модель показала точность 69.33% на тестовой выборке. Наиболее важными признаками для классификации оказались: Кредитная история (важность 0.4131) и Доход (важность 0.2740).

Из 300 наблюдений в тестовой выборке модель правильно классифицировала 208 наблюдений. Таблица сопряженности показывает, что модель лучше предсказывает класс "Нет" (не возврат кредита), чем класс "Да" (возврат кредита).

Построенное дерево решений может быть использовано банком для оценки риска выдачи кредита новым клиентам на основе их характеристик.