

Исходная выборка содержит записи о 170746 клиентах, классифицированных на два класса: 0 – клиент не открыл кредитный счет в банке (140690 клиентов) и 1 – клиент открыл кредитный счет в банке (30056 клиентов). Необходимо классифицировать клиентов на тех, кто не откроет кредитного счета, и тех, кто его откроет. По каждому наблюдению (клиенту) фиксируются 15 исходных переменных.

Список исходных переменных включает в себя:

1. категориальный предиктор Идентификационный номер [client_id];
2. категориальный предиктор Пол [gender];
3. количественный предиктор Возраст [age];
4. категориальный предиктор Семейный статус [marital_status];
5. категориальный предиктор Сфера занятости [job_position];
6. количественный предиктор Сумма кредита [credit_sum];
7. количественный предиктор Срок кредитования [credit_month];
8. количественный предиктор Внутренняя скоринговая оценка [score_shk];
9. категориальный предиктор Образование [education];
10. категориальный предиктор Идентификационный номер тарифа [tariff_id];
11. категориальный предиктор Регион [living_region];
12. количественный предиктор Месячный заработок [monthly_income];
13. количественный предиктор Количество кредитов у клиента [credit_count];
14. количественный предиктор Количество просроченных кредитов у клиента [overdue_credit_count];
15. категориальная зависимая переменная Факт открытия кредитного счета в данном банке [open_account_flg].