Кредитный скоринг

Ваша задача состоит в том, чтобы по различным характеристикам клиентов спрогнозировать целевую переменную - имел клиент просрочку 90 и более дней или нет (и если имел, то банк не будет выдавать кредит этому клиенту, а иначе будет).

Ниже находится описание признаков клиентов.

Целевая переменная

**SeriousDlqin2yrs:** Клиент имел просрочку 90 и более дней

* **RevolvingUtilizationOfUnsecuredLines:** Общий баланс средств (total balance on credit cards and personal lines of credit except real estate and no installment debt like car loans divided by the sum of credit limits).
* **age:** Возраст заемщика
* **NumberOfTime30-59DaysPastDueNotWorse:** Сколько раз за последние 2 года наблюдалась просрочка 30-59 дней.
* **DebtRatio:** Ежемесячные расходы (платеж по долгам, алиментам, расходы на проживания) деленные на месячный доход.
* **MonthlyIncome:** Ежемесячный доход.
* **NumberOfOpenCreditLinesAndLoans:** Количество открытых кредитов (напрмер, автокредит или ипотека) и кредитных карт.
* **NumberOfTimes90DaysLate:** Сколько раз наблюдалась просрочка (90 и более дней).
* **NumberRealEstateLoansOrLines:** Количество кредиов (в том числе под залог жилья)
* **NumberOfTime60-89DaysPastDueNotWorse:** Сколько раз за последние 2 года заемщик задержал платеж на 60-89 дней.
* **NumberOfDependents:** Количество иждивенцев на попечении (супруги, дети и др).

Таблица находится в схеме public под названием credit\_scoring.

Прогноз дохода

Ваша задача состоит в том, чтобы по различным характеристикам людей спрогнозировать целевую переменную - уровень дохода людей (два варианта дохода: больше 50 тысяч или меньше).

Ниже находится описание признаков клиентов.

Целевая переменная - зарплата больше или меньше 50К

class: >50K, <=50K

* **age:** количество полных лет.
* **workclass:** Private, Self-emp-not-inc, Self-emp-inc, Federal-gov, Local-gov, State-gov, Without-pay, Never-worked.
* **fnlwgt:** continuous.
* **education:** Bachelors, Some-college, 11th, HS-grad, Prof-school, Assoc-acdm, Assoc-voc, 9th, 7th-8th, 12th, Masters, 1st-4th, 10th, Doctorate, 5th-6th, Preschool.
* **education-num:** continuous.
* **marital-status:** Married-civ-spouse, Divorced, Never-married, Separated, Widowed, Married-spouse-absent, Married-AF-spouse.
* **occupation:** Tech-support, Craft-repair, Other-service, Sales, Exec-managerial, Prof-specialty, Handlers-cleaners, Machine-op-inspct, Adm-clerical, Farming-fishing, Transport-moving, Priv-house-serv, Protective-serv, Armed-Forces.
* **relationship:** Wife, Own-child, Husband, Not-in-family, Other-relative, Unmarried.
* **race:** White, Asian-Pac-Islander, Amer-Indian-Eskimo, Other, Black.
* **gender:** Female, Male.
* **capital-gain:** continuous.
* **capital-loss:** continuous.
* **hours-per-week:** continuous.
* **native-country:** United-States, Cambodia, England, Puerto-Rico, Canada, Germany, Outlying-US(Guam-USVI-etc), India, Japan, Greece, South, China, Cuba, Iran, Honduras, Philippines, Italy, Poland, Jamaica, Vietnam, Mexico, Portugal, Ireland, France, Dominican-Republic, Laos, Ecuador, Taiwan, Haiti, Columbia, Hungary, Guatemala, Nicaragua, Scotland, Thailand, Yugoslavia, El-Salvador, Trinadad&Tobago, Peru, Hong, Holand-Netherlands.

Для некоторых числовых (continuous) характеристик в датасете нет полного описания.

Таблица находится в схеме public под названием adult\_income.

Открыл ли клиент вклад

Ваша задача состоит в том, чтобы по различным характеристикам клиентов спрогнозировать целевую переменную - открыл ли клиент вклад на предложенных ему условиях или нет.

Ниже находится описание признаков клиентов.

Целевая переменная

deposit - has the client subscribed a term deposit? (binary: 'yes','no')

* **age** (numeric): количество полных лет
* **job :** тип занимаемой должности (categorical: 'admin.','blue-collar','entrepreneur','housemaid','management','retired','self-employed','services','student','technician','unemployed','unknown')
* **marital :** семейное положение (categorical: 'divorced','married','single','unknown'; note: 'divorced' means divorced or widowed)
* **education** (categorical: 'basic.4y','basic.6y','basic.9y','high.school','illiterate','professional.course','university.degree','unknown')
* **default:** имеется ли любой вид кредитования? (categorical: 'no','yes','unknown')
* **housing:** имеется ли ипотека? (categorical: 'no','yes','unknown')
* **loan:** имеется ли персональный кредит? (categorical: 'no','yes','unknown')

# related with the last contact of the current campaign:

* **contact:** вид контакта с клиентом (categorical: 'cellular','telephone')
* **month:** в каком месяце было сделано предыдущее предложение (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')
* **day\_of\_week:** день недели было предыдущее предложение (categorical: 'mon','tue','wed','thu','fri')
* **duration:** продолжительность предыдущего общения в секундах. Important note: this attribute highly affects the output target (e.g., if duration=0 then y='no'). Yet, the duration is not known before a call is performed. Also, after the end of the call y is obviously known. Thus, this input should only be included for benchmark purposes and should be discarded if the intention is to have a realistic predictive model.

# other attributes:

* **campaign:** количество контактов по этой маркетинговой кампании с этим клиентов(numeric, includes last contact)
* **pdays:** количество дней, прошедших с предыдущего предложения(число; 999 означает, что раньше не предлагали)
* **previous:** количество контактов с этим лиентов в другие маркетинговые кампании(numeric)
* **poutcome:** результат предыдущего предложения (categorical: 'failure','nonexistent','success')
* **balance** - закодированное название личной числовой характеристики клиента

Таблица находится в схеме public под названием bank\_deposit