# КОНТРОЛЬНАЯ РАБОТА

# Распределение вариантов

|  |  |
| --- | --- |
| Алексеева Анастасия Андреевна | 1 |
| Алиева Елизавета Руслановна | 2 |
| Андреева Анна Александровна | 3 |
| Антипенкова Мария Максимовна | 4 |
| Артемова Ирина Александровна | 5 |
| Афанасьева Екатерина Сергеевна | 6 |
| Байрам Екатерина Игоревна | 7 |
| Балашов Александр Романович | 8 |
| Баранова Юлия Андреевна | 9 |
| Веретнов Роман Владимирович | 10 |
| Вицукаева Александра Геннадьевна | 11 |
| Гарина Екатерина Сергеевна | 12 |
| Громов Артем Сергеевич | 13 |
| Денисов Арсений Владимирович | 14 |
| Дьяконов Илья Ярославович | 15 |
| Екшембиева Айгуль Мухарамовна | 16 |
| Елфимова Анастасия Ивановна | 17 |
| Еремеева Марина Игоревна | 18 |
| Залялютдинова Кристина Рустамовна | 19 |
| Иванов Егор Олегович | 20 |
| Ионова Карина Константиновна | 21 |
| Ковалькова Людмила Александровна | 22 |
| Короткова Анастасия Андреевна | 23 |
| Котова Юлия Петровна | 24 |
| Куликова Александра Сергеевна | 25 |
| Куриловская Дарья Дмитриевна | 26 |
| Матвеенко Кирилл Алексеевич | 27 |
| Мельников Тимофей Михайлович | 28 |
| Назарова Валерия Олеговна | 29 |
| Пастухова Елизавета Игоревна | 30 |
| Полозкова Мария Анатольевна | 31 |
| Прокофьев Игорь Витальевич | 32 |
| Склифасовский Денис Олегович | 33 |
| Солтанович Светлана Игоревна | 34 |
| Ткачева Евгения Андреевна | 35 |
| Федосов Максим Максимович | 36 |
| Хетагуров Павел Константинович | 37 |
| Хухарева Ольга Дмитриевна | 38 |
| Чайкина Екатерина Алексеевна | 39 |
| Черевко Игорь Дмитриевич | 40 |
| Шабалин Михаил Сергеевич | 41 |
| Шафиев Ибрагим Ильдарович | 42 |
| Шунина Анастасия Александровна | 43 |
| Юсупов Рахим Рафикович | 44 |
| резерв | 45 |

# Пояснения

В файлах по вариантам (csdataNN.csv, где NN соответствует вашему номеру варианта) находятся исходные данные. В нечетных вариантах находятся данные для кредитного скоринга, в четных – данные для оценки стоимости кристаллов фианита. Вам необходимо построить не менее 3 моделей и определить, какая из них лучшая.

Контрольную работу можно выполнять в программе Deductor, а также, по желанию студента, в программе Loginom Academic или на языках R или Python.

Модели должны быть неодинаковыми. Нужно применить разные инструменты (например, в Deductor варианты – линейная регрессия, логистическая регрессия, дерево решений, нейронная сеть). Допускается, чтобы две из трех моделей были построены на одном инструменте, но их параметры должны отличаться (например, две нейронных сети с разным числом слоев и нейронов).

Сравните результаты обучения и выберите по этим результатам лучшую (для ваших данных) модель.

# Описание данных

Выходным полем является *для задачи скоринга* поле **SeriousDlqin2yrs** (Были ли в последние два года задержки платежа 2 года и более). Его значение – 1 (были -> плохой клиент) или 0 (не было -> хороший клиент). Остальные поля цифровые, входные (кроме поля **RefNum**, если оно есть – это условный номер). Их описание, если вам интересно, находится в конце этого документа в приложении.

Выходным полем является *для задачи оценки стоимости* поле **price** (Цена). Остальные поля цифровые, входные (кроме поля **RefNum**, если оно есть – это условный номер). Их описание, если вам интересно, находится в конце этого документа в приложении.

# Оформление результатов

Вам нужно сдать два файла: собственно файл с моделями и файл MS Word с пояснениями.

Файлы с исходными данными изменять нельзя, их вы не сдаете, при проверке я использую архивные копии.

**1. Файл с моделями**

- Если вы работаете в Deductor, то результаты представляются в файле проекта «\*.ded», который содержит все модели.

- Если вы работаете в Loginom, то результаты представляются в файле пакета «\*.lgp», который содержит все модели.

- Если вы работаете в R или Python, то результаты представляются в файле скрипта («\*.R» или «\*.py»), который содержит все этапы построения от загрузки данных до вывода результатов; в тексте модели следует разделить строкой-примечанием, например, «# модель 1»

**2. Файл с текстом пояснений (\*.doc, \*.docx, \*.odf)**

В файле нужно привести результаты обучения моделей (матрицы ошибок, важнейшие показатели) и затем сделать выбор лучшей модели и обосновать, почему она лучшая. Приветствуется расчет различных показателей качества классификации, а также сравнение с константной моделью.

Образец файла прилагается в отдельном файле.

## Приложение

# ОПИСАНИЕ ДАННЫХ

Для задачи скоринга: выходным является поле **SeriousDlqin2yrs**

|  |  |  |
| --- | --- | --- |
| **Поле** | **Описание** | **Тип** |
| **SeriousDlqin2yrs** | **Были ли в последние два года задержки платежа 2 года и более** | **1 – да 0- нет** |
| RevolvingUtilizationOfUnsecuredLines | Отношение долга по кредитным картам и всем кредитным линиям к кредитному лимиту | число |
| age | Возраст в годах | число |
| NumberOfTime30-59DaysPastDueNotWorse | Число задержек платежа на 30-59 дней | число |
| DebtRatio | Отношение всех платежей к месячному доходу | число |
| MonthlyIncome | Месячный доход | число |
| NumberOfOpenCreditLinesAndLoans | Текущее число кредитов и кредитных карт | число |
| NumberOfTimes90DaysLate | Число задержек платежа на 90 и более дней | число |
| NumberRealEstateLoansOrLines | Число ипотечных кредитво | число |
| NumberOfTime60-89DaysPastDueNotWorse | Число задержек платежа на 60-89 дней | число |
| NumberOfDependents | Число иждивенцев, включая самого заемщика | число |

Для задачи оценки стоимости: выходным является поле **price**

|  |  |  |
| --- | --- | --- |
| **Поле** | **Описание** | **Тип** |
| **Price** | **Цена** | **число** |
| Carat | Вес в каратах | число |
| Cut | Качество огранки в баллах | число |
| Color | Цвет в баллах | число |
| Clarity | Прозрачность в баллах | число |
| Depth | Отношение высоты к диаметру в % | число |
| Table | Отношение ширины передней грани к диаметру | число |
| X | размер по X | число |
| Y | размер по Y | число |
| Z | размер по Z | число |