

Задача 6. Задание составлено совместно с Газпромбанк (АО) (15 баллов).

В предлагаемой задаче требуется научиться определять рейтинг кредитора, если известно 32 финансовых параметра кредитора.

В тренировочном датасете 'train.csv' для каждого кредитора известно значение параметра **Grade**, а в тестовом датасете 'test.csv', соответственно, нет. Иначе говоря, необходимо предсказать значение параметра **Grade**.

Качество классификатора будет измеряться при помощи усредненной метрики $F1$ с правильно предсказанным признаком **Grade** для всех кредиторов.

Баллы за задачу будут начисляться по формуле:

$$15 * Q,$$

где Q вычисляется, как

$$Q = \frac{1}{7} \sum_{i \in M} F_{1i}, \quad M = \{A, B, C, D, E, F, G\},$$

а метрика F_{1i} – это метрика $F1$, которая вычисляется для каждого класса отдельно.

Формат ввода: Обучающий и тестовый датасеты train.csv и test.csv находятся в файлах, доступных ниже.

Файлы содержат следующие поля:

Grade – одно из значений рейтинга A, B, C, D, E, F, G

| | |
|-----|---|
| c1 | Уникальный идентификатор, присвоенный заемщику. |
| c2 | Сумма кредита, на которую претендует заемщик. |
| c3 | Общая сумма, выделенная по этому кредиту на данный момент времени. |
| c4 | Количество платежей по кредиту. Значения указаны в месяцах и могут быть равны либо 36, либо 60. |
| c5 | Процентная ставка по кредиту. |
| c6 | Ежемесячный платеж, причитающийся заемщику в случае получения кредита. |
| c7 | Должность, указанная Заемщиком при подаче заявления на получение кредита. |
| c8 | Стаж работы в годах. Возможные значения находятся в диапазоне от 0 до 10, где 0 означает менее одного года, а 10 означает десять или более лет. |
| c9 | Статус домовладельца, предоставленный заемщиком при регистрации или полученный от кредитной организации. |
| c10 | Отчет о годовом доходе, предоставленный заемщиком при регистрации. |
| c11 | Указывает, был ли доход подтвержден, не подтвержден или был ли подтвержден источник дохода. |
| c12 | Текущее состояние кредита. |
| c13 | Категория, предоставленная заемщиком для запроса на получение кредита. |
| c14 | Государство, указанное заемщиком в кредитной заявке. |

| | |
|-----|--|
| c15 | Коэффициент, рассчитанный с использованием общих ежемесячных платежей заемщика по общей сумме долговых обязательств. |
| c16 | Количество случаев просроченной задолженности более чем на 30 дней в кредитном досье заемщика за прошлые периоды. |
| c17 | Количество запросов за последние 6 месяцев (без учета запросов на авто и ипотеку). |
| c18 | Количество открытых кредитных линий в кредитном досье заемщика. |
| c19 | Количество унизительных публичных записей. |
| c20 | Общий возобновляемый остаток по кредиту. |
| c21 | Общее количество кредитных линий, имеющих в настоящее время в кредитном досье заемщика. |
| c22 | Оставшаяся непогашенной основная сумма на общую сумму финансирования. |
| c23 | Платежи, полученные на сегодняшний день на общую сумму финансирования. |
| c24 | Основная сумма, полученная на сегодняшний день. |
| c25 | Проценты, полученные на сегодняшний день. |
| c26 | Просроченные платежи, полученные на сегодняшний день. |
| c27 | Списание валового возмещения после списания. |
| c28 | Плата за сбор после списания. |
| c29 | Последняя общая сумма полученного платежа. |
| c30 | Количество сборов за 12 месяцев без учета медицинских сборов. |
| c31 | Количество зарегистрированных банкротств. |
| c32 | Кредитный рейтинг, присвоенный LC. |

Формат вывода: Необходимо сделать прогноз на тестовых данных и загрузить таблицу с прогнозами в соответствии с форматом 'submit_sample.csv'.