



## Задание

Вам предлагается набор данных, содержащий следующие переменные

### Переменные, характеризующие кредитную историю клиента

cred\_sum\_cc\_all - сумма кредитов по кредитным картам  
mfo\_inqs\_count\_month - количество запросов на кредиты в другие в МФО  
all\_closed\_creds\_sum\_all количество закрытых кредитов  
bank\_inqs\_count\_quarter - количество запросов на кредиты в банки  
cred\_max\_overdue\_max\_ly - максимальная просрочка за год  
all\_active\_creds\_sum\_all - денежная сумма всех активных кредитов  
mfo\_last\_days\_all - Количество дней с последнего займа в МФО  
cred\_sum\_cc\_ly - Сумма лимитов кредитных карт, оформленных за последний год  
cred\_sum\_debt\_all\_all - Сумма задолженности по всем кредитам  
all\_closed\_creds\_sum\_ly - Сумма закрытых кредитов за последний год  
mfo\_cred\_mean\_sum\_3lm - Средняя сумма МФО кредитов, выданных за последние 3 месяца  
delay\_more\_sum\_all - Количество просрочек более чем на 90 дней по всем кредитам  
all\_creds\_count\_all - Общее количество кредитов  
cred\_day\_overdue\_all\_sum\_all - Суммарное количество дней просрочки текущих активных кредитов  
cred\_max\_overdue\_max\_3lm - Максимальная сумма просроченной задолженности, по кредитам взятым за последние 3 месяца  
mfo\_closed\_count\_ly - Количество закрытых МФО кредитов, взятых за последний год  
cred\_sum\_overdue\_cc\_all - Сумма просрочек по кредитным картам  
count\_overdue\_all\_3lm - Количество кредитов на просрочке, взятых за последние 3 месяца  
all\_creds\_count\_lm - Количество кредитов, взятых за последний месяц  
region - регион подачи заявки

### Переменные характеризующие клиента

work\_code - Профессия. 5 - рабочие профессии (слесарь, токарь). 3 - офисный работник (бухгалтер, программист). 1 - госслужащий (полицейский, медсестра)  
month\_income - доход

### Целевые переменные:

bad - 1 - кредит просрочен, 0 - кредит возвращен, nap - отказ.  
approved - 1 - одобрено, 0 - отказано.

**Задача:**

Сравнить распределение каждого признака на выборке выданных кредитов ( $bad \neq nan$ ) с распределением этого же признака на всей выборке. Прокомментировать причину различий. Дополнить комментарии графиками, выбрав 4-5 показательных признаков. Если есть признаки, между которыми различий в распределении не наблюдается, объяснить причину.

Обучить модель классификации только на выданных кредитах, целевая переменная  $bad$ . Придумать/найти алгоритм разметки отклоненных ( $bad=NaN$ ) заявок. После применения алгоритма разметки, обучить модель классификации на всех заявках. Сравнить с моделью, обученной только на выданных.

При выполнении задания рекомендуем сделать упор на анализ данных, интерпретацию решений и ошибок моделей. Меньше усилий тратить на подбор гиперпараметров и выбор алгоритма классификации.

На выполнение задания отводится 1 неделя с момента получения письма  
Отчет прислать в файле .pdf

Данные

[https://drive.google.com/drive/u/1/folders/18zINli\\_ACf3E-xthi6PFwVyvG3zPFhWJ](https://drive.google.com/drive/u/1/folders/18zINli_ACf3E-xthi6PFwVyvG3zPFhWJ)