**Инструкция к практическому заданию по прогнозированию финансовых показателей с помощью ML.**

1. Скачать data-set ML\_pred\_stud\_set\_1.csv и загрузить из него данные
2. Заполнить пробелы в данных из ML\_pred\_stud\_set\_1.csv любым удобным способом
3. Описанным в видео способом (или любым другим) выберите наилучший, с точки зрения точности, метод машинного обучения для прогнозирования показателя «кол-во транзакций на терминал
4. Подберите оптимальные гиперпараметры для выбранного метода (любым способом)
5. Прогноз будет строиться помесячно в разрезе регионов и сегментов. Учитывая данное обстоятельство, определите нужно ли ввести в data-set какие-либо производные факторы (например dummy-переменные )? Если такая необходимость есть – введите эти дополнительные факторы.
6. Описанным в видео способом (или любым другим) отберите значимые факторы модели.
7. Спрогнозируйте показатель «кол-во транзакций на терминал» помесячно на 2018 год в разрезах «регион» и «сегмент»

**Описание data-set 1:**

A – report\_date – date – отчётная дата

B – segment – varchar -сегмент

C – region – varchar - регион

D – fact\_progn – varchar – факт/прогноз

E – tranth\_tst – float – кол-во транзакций на терминал

F – dollar – float –курс доллара

G – dol\_rab\_ter – float – доля работающих терминалов

H – dol\_sn\_nal – float – доля снятия наличных

I – rab\_dney – int – число рабочих дней

J – atm – float – число банкоматов

K – cards – float – число банковских карт

L – compl – float – доля успешно проведённых транзакций

**Тест**

1. Какой метод машинного обучения оказался наилучшим для прогнозирования кол-ва транзакций на терминал с точки зрения точности? (можно выбрать только 1 ответ). Варианты:
   * Huber regressor
   * Theil sen regressor
   * Passive Aggressive regressor
   * Lasso
   * Multi Task Lasso
   * Ridge
   * Linear regression
   * Elastic Net
   * Bayesian Ridge
   * Lasso Lars
   * Orthogonal Matching Pursuit
   * MLP
   * XGB regressor
   * Gradient Boosting regressor
   * K Neighbors regressor
   * Decision tree regressor
   * RANSAC regressor
   * SVR
   * Gaussian process regressor
   * SGD regressor

Ответ: правильным считается ответ «Lasso» **или** «Multi Task Lasso» **или** «Elastic Net» **или** «Multi Task Elastic Net» или «Ridge» **или** «BayesianRidge» (любой из 6 будет считаться правильным, т.к. по критерию MSE они практически одинаковые)

1. Какие факторы (их 2) из базового data-set оказались НЕ значимыми или избыточными (можно выбрать несколько вариантов). Варианты:
   * кол-во транзакций на терминал
   * курс доллара
   * доля работающих терминалов
   * доля снятия наличных
   * число рабочих дней
   * число банкоматов
   * число банковских карт
   * доля успешно проведённых транзакций

Ответ: число банкоматов **и** доля успешно проведённых транзакций

1. В какие месяцы наблюдается сезонные скачки/падения по кол-ву транзакций на терминал? (можно выбрать несколько ответов). Варианты:
   * Январь
   * Февраль
   * Март
   * Апрель
   * Май
   * Июнь
   * Июль
   * Август
   * Сентябрь
   * Октябрь
   * Ноябрь
   * Декабрь

Ответ: январь **и** декабрь

1. Динамика по «кол-ву транзакций на терминал» по какому сегменту наиболее отличается от остальных? (можно выбрать только 1 ответ). Варианты:
   * S\_1
   * S\_2
   * S\_3
   * S\_4
   * S\_5

Ответ: S\_3

1. Какой прогноз по кол-ву транзакций на терминал в январе 2018 по сегменту «S\_2» и региону «RG\_77». (Открытый вопрос)

Ответ: Правильным считается ответ от 430 до 489

1. Какой прогноз по кол-ву транзакций на терминал в июне 2018 по сегменту «S\_4» и региону «RG\_56». (Открытый вопрос)

Ответ: Правильным считается ответ от 212 до 321

1. Если предположить, что во всех регионах по всем сегментам одинаковое число терминалов. Какое среднее число транзакций на терминал в январе 2018. (Открытый вопрос)

Ответ: Правильным считается ответ от 107 до 119

1. Если предположить, что во всех регионах по всем сегментам одинаковое число терминалов. Какое среднее число транзакций на терминал в сентябре 2018. (Открытый вопрос)

Ответ: Правильным считается ответ от 132 до 146