**Тестовое задание**

**Задание #1:**

Таблица (task2.txt) содержит 11 столбцов чисел. Первые 10 столбцов - входные переменные (x\_1 , …. , x\_10), 11-й столбец - выходная прогнозируемая переменная y. Каждая строка в файле - это один обучающий пример для построения статистической модели для зависимости y = y(x\_1,x\_2,..,x\_10). Конкретный вид связи между выходной и входными переменными не известен.

Требуется провести разведочный анализ данных, выбрать критерий для точности модели, и указать степень важности (информативности) каждой из 10 входных переменных с точки зрения точности модели. Что еще можно предложить чтобы улучшить получившийся результат?

**Файл:** tesk2.txt

**Задание #2:**

Нужно предложить и обосновать подход для группировки идентичных наименований товаров.

 Пример входа:

[1] "Кисломолочный напиток Актимель ежевично-черничный 100 г"

[2] "Напиток Данон Актимель черника/ежевика 2,5% 6\*0,1л Россия"

[3] "Гель для душа Роскошная мягкость Черная орхидея Palmolive, 250 мл"

[4] "Гель Palmolive Naturals д/душа Роскошная мягкость Черная орхидея 250мл Италия"

[5] "Гель для душа Palmolive "Черная орхидея" 250 мл"

[6] "Корм Whiskas говядина/кролик канапе новогодний 85г Россия"

Пример выхода: [1 2] [3 4 5] 6 (см. вложенный файл)

**Файл:** test\_task\_NLP.json

Решение лучше присылать в виде jupyter notebook’а. Если Задание#2 не получается – ничего страшного, главное опиши в задачах ход мысли и дальнейшие шаги.