

Предсказание интенсивности продаж новых товаров в Retail-сети

В качестве 3-го практического задания студентам 4-го курса ФИВТ МФТИ в 7-ом семестре предлагается предсказать интенсивность продаж новых товаров в розничной сети продаж. Под интенсивностью продажи подразумевается среднее количество продаж товара в день в конкретном магазине за первый месяц продаж.

В качестве исходных данных для предсказания предоставляются различные характеристики товаров и магазинов, начиная от внешнего вида и расположения магазинов, заканчивая описанием интенсивностей продаж в магазинах по разным товарным группам.

Задание будет проходить на базе конкурса в Kaggle. Соревнование заканчивается **26.11.2016**

1 Описание задачи в Kaggle

Данные представляют собой выборку из обезличенных данных по характеристикам товаров и магазинов (70 различных характеристик). Описание признаков приведено в файле `data_description.txt`.

Файл `demand_train.csv` содержит описание обучающей выборки с уже размеченными ответами.

Файл `demand_test.csv` содержит описание признаков контрольной выборки, которую требуется разметить. Решения принимаются в виде файла с предсказаниями о принадлежности классов всех объектов из `test.csv` в следующем формате: в каждой строке файла должно содержаться 2 числа: `id` объекта, предсказание об интенсивности спроса на первый месяц продаж. В качестве примера файла решения предлагается файл `benchmark1.csv`.

В качестве функции штрафа за неточное предсказание используется SMAPE:

$$SMAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{(|y_i| + |\hat{y}_i|) / 2}.$$

При возникновении вопросов о конкурсе и о данных пишите сразу на `ml.course.mipt@gmail.com`.

2 Правила конкурса

Конкурс проводится в течение 2-х недель: с 00:00:00 12.11.2015 по 23:59:59 26.11.2015 UTC (по Гринвичу!). Участие в соревновании индивидуальное

(командное участие не разрешено).

Название команды (из 1 человека) должно формироваться по правилу: НомерГруппы_Фамилия_Имя . Например: 574_Романенко_Алексей.. Если название команды не будет соответствовать формату (на момент закрытия соревнования), тогда оценка за констест будет снижена.

Каждый из участников имеет права на 2 сабмита в течение каждого дня. О правилах вычисления результатов (что такое public leaderboard и private leaderboard и как они формуруются), можно ознакомиться на странице соревнования.

3 Условия сдачи задания

1. Можно использовать любые реализации изученных на курсе моделей (и не только)
2. Для сдачи задания нужно:
 - (a) побить benchmark_1 на kaggle;
 - (b) прислать отчёт в виде ipython_notebook.

Если по результатам контеста вы вошли в top-10, тогда вам будут начислены дополнительные баллы.