

Вам предоставлены данные о кол-ве посетителей одного из крупнейших бизнес центров в Америке. Данные собирались при помощи камер-счётчиков, которые стоят на входе в бц, и фиксировали кол-во посетителей, которые входили и выходили из бц. Данные предоставлены за 15 недель с 24/07/2005 по 05/11/2005. Также имеется информация о мероприятиях в данном бц, которые проходили в это время.

Описание данных:

Файл counters.csv - Имеет поля dt – дата и время измерения (данные собирались каждый полчаса, то есть за 1 час имеется 2 измерения), cnt – кол-во посетителей.

Файл events.csv – Имеет поля dt_start – дата и время начала события, dt_end – дата и время окончания события, event_type – тип события (здесь только тип события = event для всех записей)

Данные разбиты на train (24/07/2005 - 05/10/2005) и test (06/10/2005 – 05/11/2005)

Задание:

1. Для каждой записи (дата+время) определить, было ли в этот момент какое-то мероприятие или нет (считать, что данные о количестве пользователей на дата+время известны)
2. Реализовать прогноз кол-ва пользователей на тестовой выборке (считать, что информация о предстоящих мероприятиях в будущем (данные из таблицы events) нам известна)

Оцените качество решения задач? Какие метрики использовали и почему?