

Опис

Walmart має 11 450 магазинів у 27 країнах, керуючи запасами в різних кліматичних і культурних умовах. Екстремальні погодні явища, такі як урагани, хуртовини та повені, можуть мати величезний вплив на продажі як на рівні магазину, так і на рівні товару.

У своєму другому рекрутинговому конкурсі Kaggle Walmart пропонує учасникам точно спрогнозувати продажі 111 потенційно чутливих до погодних умов продуктів (таких як парасольки, хліб і молоко) під час великих погодних катаклізмів у 45 своїх роздрібних точках.

Оцінювання

Роботи оцінюються за середньоквадратичною логарифмічною похибкою (RMSLE). RMSLE розраховується за формулою

Root Mean Squared Error (RMSE) Root Mean Squared Log Error (RMSLE)

$$\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$
$$\sqrt{\frac{1}{n} \sum_{i=1}^n (\log(p_i + 1) - \log(a_i + 1))^2}$$

prediction
actual

RMSLE Formula

Де:

n - кількість рядків у тестовому наборі

p_i - прогнозована кількість проданих одиниць

a_i - фактична кількість проданих одиниць

$\log(x)$ - натуральний логарифм

Опис набору даних

Вам надано дані про продажі 111 товарів, на продажі яких може впливати погода (молоко, хліб, парасольки тощо). Ці 111 товарів продаються в 45 різних магазинах Walmart. Деякі з цих продуктів можуть бути схожими (наприклад, молоко), але мати різні ідентифікатори в різних магазинах/регіонах/постачальниках. Ці 45 локацій покриваються 20 метеостанціями (тобто деякі магазини знаходяться поруч і мають спільну метеостанцію).

Завдання конкурсу полягає в тому, щоб передбачити кількість кожного продукту, проданого під час основних погодних явищ. Для цілей цього конкурсу ми визначили погодні події як будь-який день, в який випало більше дюйма дощу або два дюйми снігу. Вам потрібно спрогнозувати кількість проданих одиниць протягом ± 3 днів після кожного шторму.

Дані про продажі не відображають різницю між запасом і попитом. Іншими словами, кількість продажів 0 не обов'язково означає відсутність попиту на цей товар; це може означати, що товар був на складі, але не був проданий, або що товар був відсутній на складі, або знятий з виробництва і недоступний.

Описи полів

date - день розпродажу або погоди

store_nbr - ідентифікатор одного з 45 магазинів

station_nbr - ідентифікатор однієї з 20 метеостанцій

item_nbr - ідентифікатор одного з 111 товарів

units - кількість проданих одиниць товару за певний день

id - триплет, що представляє store_nbr, item_nbr та дату. Сформуйте ідентифікатор, об'єднавши ці три елементи (у такому порядку) разом із символом підкреслення. Наприклад, "2_1_2013-04-01" представляє магазин 2, товар 1, проданий 2013-04-01.

Описи файлів

key.csv - реляційне відображення між магазинами та метеостанціями, що їх покривають

store_nbr - ідентифікатор одного з 45 магазинів

`station_nbr` - ідентифікатор однієї з 20 метеостанцій

sampleSubmission.csv - файл, що задає формат прогнозу

`id` - триплет, що представляє `store_nbr`, `item_nbr` та дату. Сформуйте ідентифікатор, об'єднавши ці три елементи (у такому порядку) разом із символом підкреслення. Наприклад, "2_1_2013-04-01" представляє магазин 2, товар 1, проданий 2013-04-01.

`units` - кількість проданих одиниць товару за певний день

train.csv - дані про продажі для всіх магазинів і дат у навчальній вибірці

`date` - день розпродажу або погоди

`store_nbr` - ідентифікатор одного з 45 магазинів

`item_nbr` - ідентифікатор одного з 111 товарів

`units` - кількість проданих одиниць товару за певний день

test.csv - магазини та дати для прогнозування (відсутні "одиниці", які ви повинні передбачити)

weather.csv - файл, що містить інформацію про погоду NOAA для кожної станції та дня

`station_nbr` - номер станції

`date` - дата

`tmax` - максимальний темп (градуси за Фаренгейтом)

`tmin` - мінімальний темп (градуси за Фаренгейтом)

`tavg` - середній темп (градуси за Фаренгейтом)

`depart` - відхилення від норми

`dewpoint` - середня точка роси

`wetbulb` - середня волога лампочка

`heat` - опалення (сезон починається з липня)

`cool` - похолодання (сезон починається з січня)

`sunrise` - сонячний захват

`sunset` - захід сонця

`codesum` - погодні явища (торнадо, гроза, дощ, сніг ...)

snowfall – діаметри сніг/ожеледиця (на землі), T – слід M
– відсутні дані
preciptotal – опади (дюйми), M – відсутні дані
stnpressure – середній тиск на станції (дюйми ртутного
стовпця)
sealevel – середній тиск на рівні моря (дюйми ртутного
стовпця)
resultspeed – результуюча швидкість вітру (милі на годину
в напрямку до десятків градусів)
resultdir – результуючий напрямок (цілі градуси)
avgspeed – середня швидкість (милі на годину в напрямку
до десятків градусів)