



> Конспект > 4 урок > Первое знакомство с динамическим ценообразованием

> Оглавление

- > [Оглавление](#)
- > [Описание данных](#)
- > [Предсказание временных рядов](#)
- > [Предсказание спроса](#)
- > [Дополнительные материалы](#)

Данные для практического занятия: [скачать](#)

Ноутбук в `html`: [скачать](#)

> Описание данных

1. `sku_dict.csv`

`sku_id` — SKU (уникальный идентификатор)

`fincode, ui1_code, ui2_code, ui3_code` — иерархическая классификация товаров

`vendor` — вендор

`brand_code` — бренд

`creation_date` — дата добавления товара

`expiration_date` — дата удаления товара из матрицы

2. **ZX10208_agr.csv, 1510001_agr.csv, 1510002_agr.csv, 1510003_agr.csv**

`dates` — дата покупки

`SKUprice_per_sku` — цена SKU

`num_purchases` — количество покупок

3. **conc_df.csv**

`SKU`

`year` — год

`week_num` — номер недели

`ret_net_code` — код торговой сети

`price_per_sku` — цена

4. **promo_1510002.csv**

`year` — год

`week_num` — номер недели

`SKU`

`discount` — скидка

> Предсказание временных рядов

Лаг — определённый промежуток времени для сдвига ряда.

Автокорреляция — взаимосвязь между последовательностями величин одного ряда, взятыми со сдвигом (лагом).

1. **$AR(p)$ (autoregression)** — авторегрессионная модель порядка p

моделирует следующий шаг в последовательности как линейную функцию наблюдения на предыдущих временных шагах. Подходит для одномерных временных рядов без трендовых и сезонных составляющих.

2. **$MA(q)$ (moving average)** — метод скользящего среднего порядка q

моделирует следующий шаг как линейную функцию остаточных ошибок от среднего процесса на предыдущих шагах. Подходит

для одномерных временных рядов без трендовых и сезонных составляющих.

3. $ARMA(p, q)$ (autoregressive moving average) — авторегрессионная модель скользящего среднего порядка (p, q) предсказывает следующий шаг как линейную функцию наблюдений и остаточных ошибок на предыдущих шагах. Сочетает в себе две предыдущие модели.
4. $ARIMA(p, d, q)$ (autoregressive integrated moving average) — интегрированная авторегрессионная модель скользящего среднего порядка (p, d, q) предсказывает следующий шаг как линейную функцию разностных наблюдений и остаточных ошибок. Включён этап получения разностей, чтобы сделать временной ряд стационарным. Подходит для данных с трендом и без сезонности.
5. $SARIMA(p, d, q)(P, D, Q, m)$ (seasonal ARIMA) — интегрированная модель сезонного авторегрессионного скользящего среднего, порядки тренда: (p, d, q) , порядки сезонности: (P, D, Q, m) . Моделирует оценки как линейную функцию разностных наблюдений, ошибок разностных сезонных наблюдений и сезонных ошибок на предыдущих шагах.

Существуют также аналоги для зависимостей от векторных (где больше 1 фактора влияния) переменных:

1. VAR (vector AR) — векторная авторегрессия, обобщённая модель для предсказания нескольких временных рядов, где текущие значения зависят от предыдущих. Пример: цена на товары одной категории.
2. $VARMA$ (vector ARMA) — векторная авторегрессия скользящего среднего.
3. $VARMAX$ (vector ARMA extended) — векторная модель с возможностью учёта внешних факторов.

Количество используемых для предсказания лагов поможет определить автокорреляционный анализ (анализ автокорреляционной функции (ACF) и частной автокорреляционной функции (PACF)).

Эндогенная переменная — зависимая переменная.

Экзогенные переменные — внешние предикторы: предопределённые переменные, влияющие на эндогенные переменные, но не зависящие от них.

Также можно использовать и иные методы, такие как:

1. Простое экспоненциальное сглаживание **SES** (simple exponential smoothing) — моделирует шаг как экспоненциально взвешенную линейную функцию наблюдений на предыдущих временных шагах. Подходит для временных рядов без трендовых и сезонных составляющих.
2. **LSTM** (long short-term memory) — рекуррентная нейронная сеть.

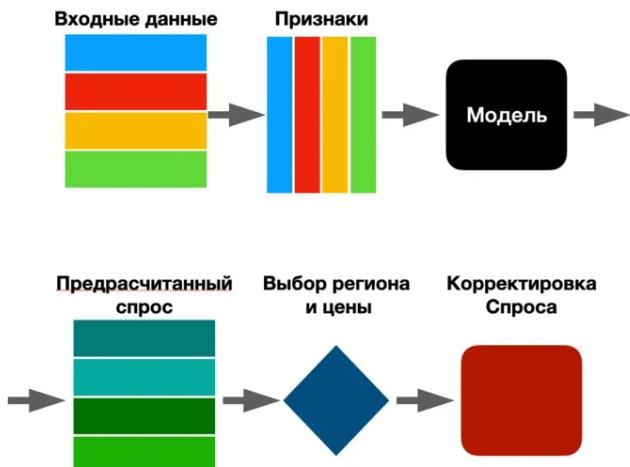
> Предсказание спроса

Иногда в реальной жизни бывают сценарии, когда необходимо реализовать какое-то количество товара. В таком случае нам нужно "**подогнать**" цену под **спрос**. Для этого нужно уметь прогнозировать не цену, а спрос. Это делается с помощью следующего пайплайна:

1. Для входных данных генерируются **признаки** для будущей модели.
2. На основании полученных признаков строится основаная **модель прогноза спроса**.
3. После этого идет **предрасчет спроса** на сетке цены (для нескольких цен рассчитывается соответствующий им спрос), чтобы можно было легко подстраиваться под конкретную текущую ситуацию.
4. Затем **спрос корректируется** с помощью локальных признаков (например, дат и/или географии) либо же локальные переменные уже входят в предсказательную модель. В таком случае корректировку производят все равно: модель может ошибиться, что-то могло произойти с товарами-заменителями (например, на них может прийтись промо, и тогда предсказываемый моделью спрос будет завышен, и т.д.).

Прогнозирование спроса на товар

- Выбор группы товаров на основе которых будет делаться прогноз
- Построение математической модели
- Построение прогноза спроса на товар в зависимости от региона, дат, и цены
- Корректировка прогнозов товаров



> Дополнительные материалы

Про простые модели:

<https://machinelearningmastery.com/time-series-forecasting-methods-in-python-cheat-sheet/>

<https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/>

<http://www.machinelearning.ru/wiki/images/5/5d/Uvarov2018Thesis.pdf>

Про LSTM:

<https://stackabuse.com/time-series-prediction-using-lstm-with-pytorch-in-python/>

<https://machinelearningmastery.com/how-to-develop-lstm-models-for-time-series-forecasting/>