

1. «Взять языка». Незаметно для потенциального противника раздобудь два временных ряда: месячный и дневной. Не бери цены финансовых инструментов, так как они плохо прогнозируются в силу эффективности рынка. Разумно взять реальные показатели. Если очень хочется работать с финансовыми данными, можно взять волатильность, она прогнозируется хорошо. Максимально чётко укажи, откуда взяты ряды. Если ряды парсились, то приведи код. При желании можно взять больше рядов и использовать какой-то ряд в качестве предиктора.
2. «Намалеви́ч». Построй графики рядов, графики автокорреляционных функций, графики с нарезкой ряда на годы для иллюстрации сезонности.
3. «Твиттер». Кратко прокомментируй полученные графики. Являются ли ряды сезонными? есть ли тренд? стационарны ли ряды? есть ли точки излома? растёт ли амплитуда колебаний ряда? Возьми логарифм ряда, если душа тянется к логарифму.
4. «Двенадцать месяцев». Для месячного ряда сравни прогнозы разных моделей на один шаг вперёд с помощью кросс-валидации скользящим или растущим окном. Длину скользящего окна (или стартовую длину растущего окна) выбери равной 80% длины ряда. В качестве метрики возьми среднее абсолютное отклонение, MAE. Сравни следующие модели:
  - Наивная,  $\hat{y}_{t+1} = y_t$ , NaiveForecaster из sktime.
  - Сезонная наивная,  $\hat{y}_{t+1} = y_{t+1-12}$ , NaiveForecaster из sktime.
  - SARIMA(1, 1, 1)(1, 0, 0)[12], ARIMA из sktime.
  - Алгоритм Хиндмана-Хандакара подбора SARIMA, AutoARIMA из sktime.
  - LGT из пакета orbit.
  - DLT из пакета orbit.
  - ETS(AAA), ExponentialSmoothing из sktime.
  - ETS с автоматическим выбором по AIC, AutoETS из sktime.
  - (по желанию) Случайный лес или градиентный бустинг на лагированных  $y_{t-s}$ , sktime будет удобнее :)

#### Ссылки:

- sktime: <https://www.sktime.org>.
- orbit: <https://orbit-ml.readthedocs.io>.
- Прогнозы в sktime, в конце про кросс-валидацию: [https://www.sktime.org/en/latest/examples/01\\_forecasting.html](https://www.sktime.org/en/latest/examples/01_forecasting.html).
- Картинки для кросс-валидации: <https://www.kaggle.com/cworsnup/backtesting-cross-validation-for-time>
- Изложение алгоритма Хиндмана-Хандакара, <https://otexts.com/fpp3/arima-r.html>.

Выбери наилучшую модель и построй график прогнозов для неё на один год вперёд используя все 100% наблюдений как обучающую выборку.

5. «Ежедневный пророк». Раздели дневной ряд на обучающую и тестовую выборку. За тестовую возьми последние два года. Если ряд короткий, то можно и один год взять в качестве тестовой. Обучив модели на обучающей выборке, сравни качество прогнозов на тестовой выборке с помощью MAE. Сравни следующие модели:

- Наивная.
- Сезонная наивная.
- KTR из пакета orbit.
- PROPHET с дефолтными настройками, sktime удобно интегрирован с prophet.
- ARIMA(1, 1, 1) с тригонометрическими предикторами, ARIMA из sktime.

Ссылки:

- KTR модель, <https://orbit-ml.readthedocs.io/en/latest/tutorials/ktrlite.html>.
- Сергей Мاستицкий, Анализ временных рядов с помощью R, глава про prophet, <https://ranalytics.github.io/tsa-with-r/ch-intro-to-prophet.html>.
- Тригонометрические предикторы, <https://robjhyndman.com/hyndsight/longseasonality/>.

Выбери наилучшую модель и построй график прогнозов для неё на один год вперёд используя все 100% наблюдений как обучающую выборку.

6. (по желанию) «Атолл Бикини». Разработай и испытай свою модель для месячного ряда в STAN<sup>1</sup>. Сравни качество прогнозов с наивной сезонной на тестовой выборке с помощью MAE.

Наставления в добрый путь храброму падавану:

- Чаще используй sktime, скорее всего там есть почти всё, что нужно :)
- Сдавай работу в исполняемого .ipynb файла. Приложи .csv файлы с рядами.
- Мелкие детали, отсутствующие в условии, заполни самостоятельно, чётко описав свой выбор.
- Да пребудет с тобой Сила!

---

<sup>1</sup>Какая связь между языком STAN и атолл Бикини? :)