Анализ временных рядов

Подготовила студентка группы БПМИ-172: Семененя Яна

Что же такое временной ряд?

- Временной ряд -это последовательность значений, описывающих протекающий во времени процесс, измеренных в последовательные моменты времени, обычно через равные промежутки

- Задача прогнозирования - найти функцию f_t :

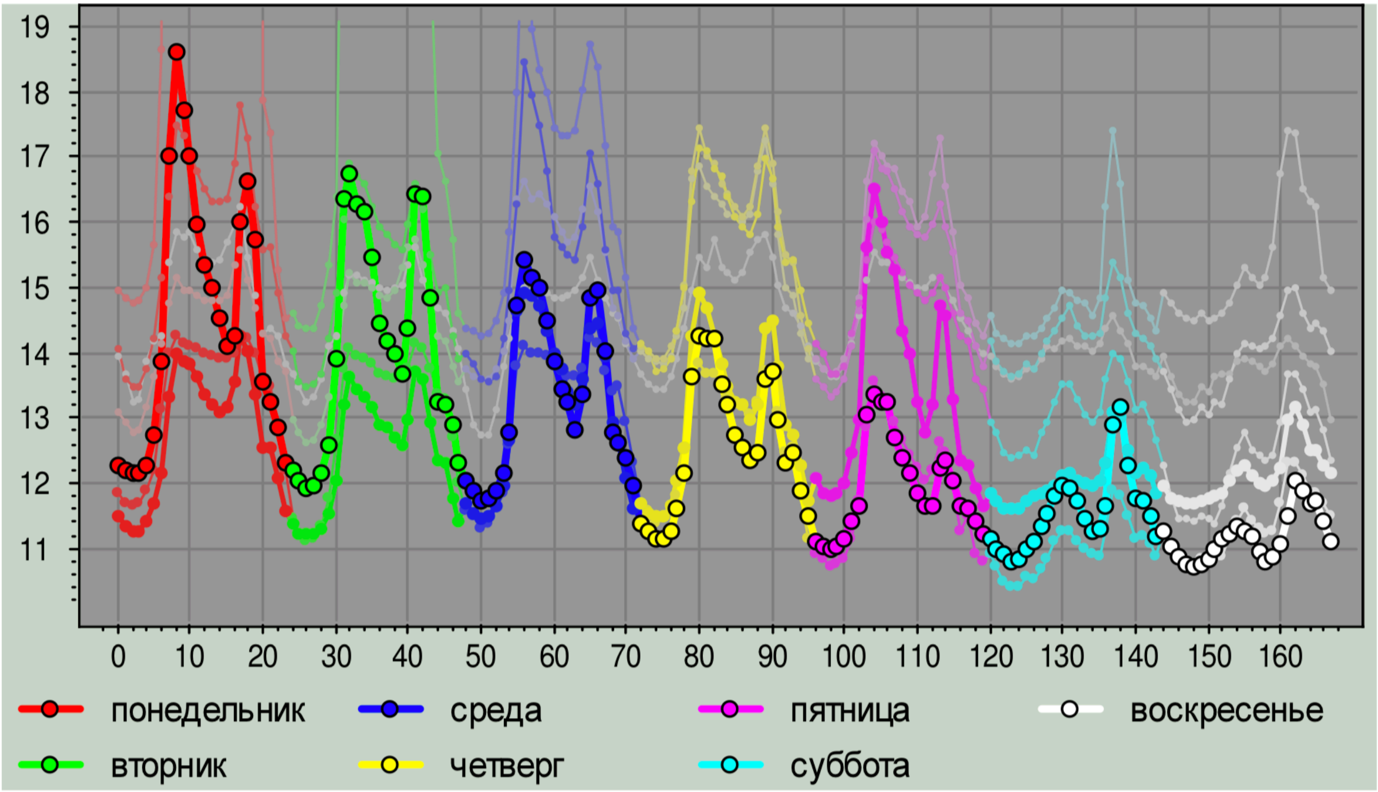
$$y_{t+d}(w) = f_t(y_1,\ldots,y_t;w) ,$$

где d=1,...,D, D—горизонт прогнозирования,
w — вектор параметров модели

Примеры прикладных задач

- Прогнозирование объемов продаж
- Прогнозирование трафика пользователей

Почасовые цены электроэнергии на бирже NordPool, 2000г.



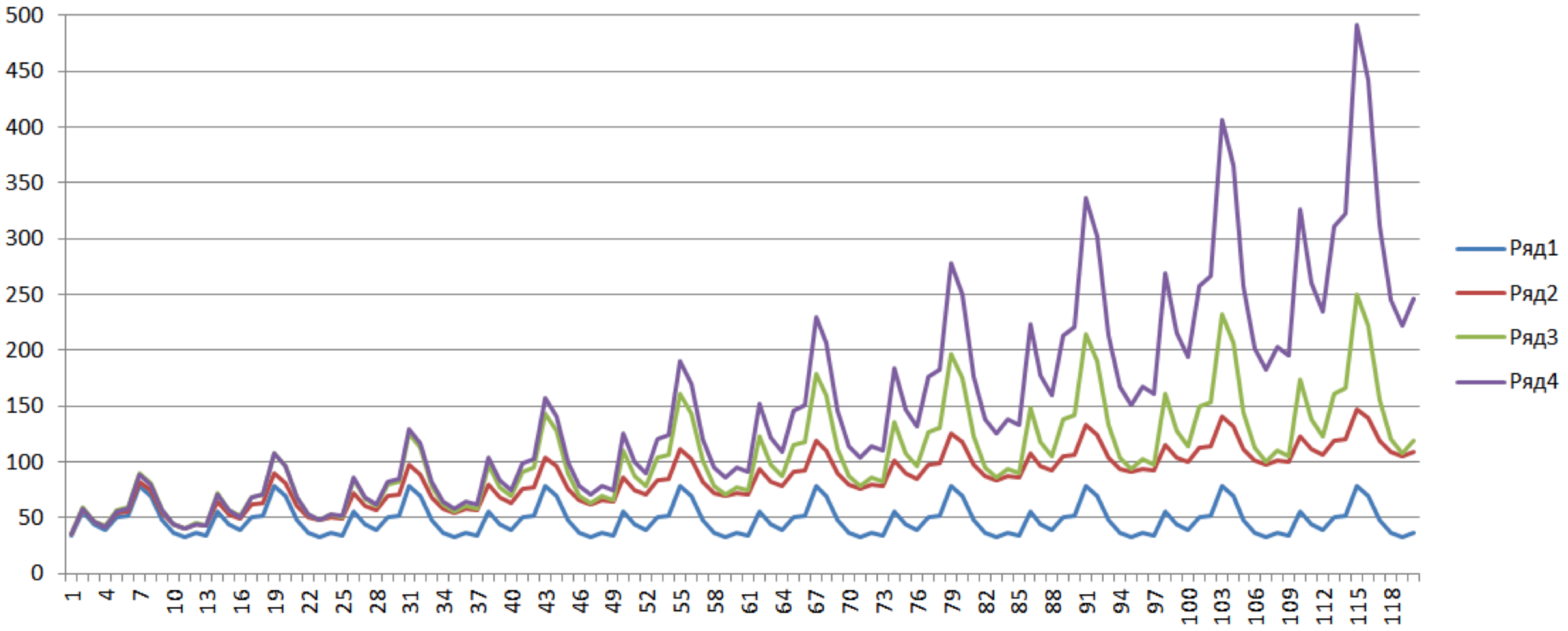
Особенности задач

- огромное число рядов
- продажи зависят от типа товара
- тренды
- сезонность
- пропуски
- праздники, промоакции
- скачки
- плохо работают сложные модели

Основные явления

- Тренд - это плавное долгосрочное изменение уровня ряда.
- Сезонность - циклические изменения уровня ряда с постоянным фиксированным периодом
- Ошибка – непрогнозируемая случайная компонента ряда.

Сочетания тренда и сезонности



Ряд 1- сезонность без тренда

Ряд 2 - линейный тренд, аддитивная сезонность

Ряд 3 - линейный тренд, мультипликативная сезонность

Ряд 4 — экспоненциальный тренд, мультипликативная сезонность

Стационарность

- Ряд $y_1 \ldots y_T$ стационарен, если $\forall$ s распределение $y_t \ldots y_{t+s}$ не зависит от t, т.е. выполняются свойства:

  - $E(y_t) = const$
  - $Var(y_1) = const$
  - $Cov(y_{t,}\ y_{t-k}) = \gamma_k$

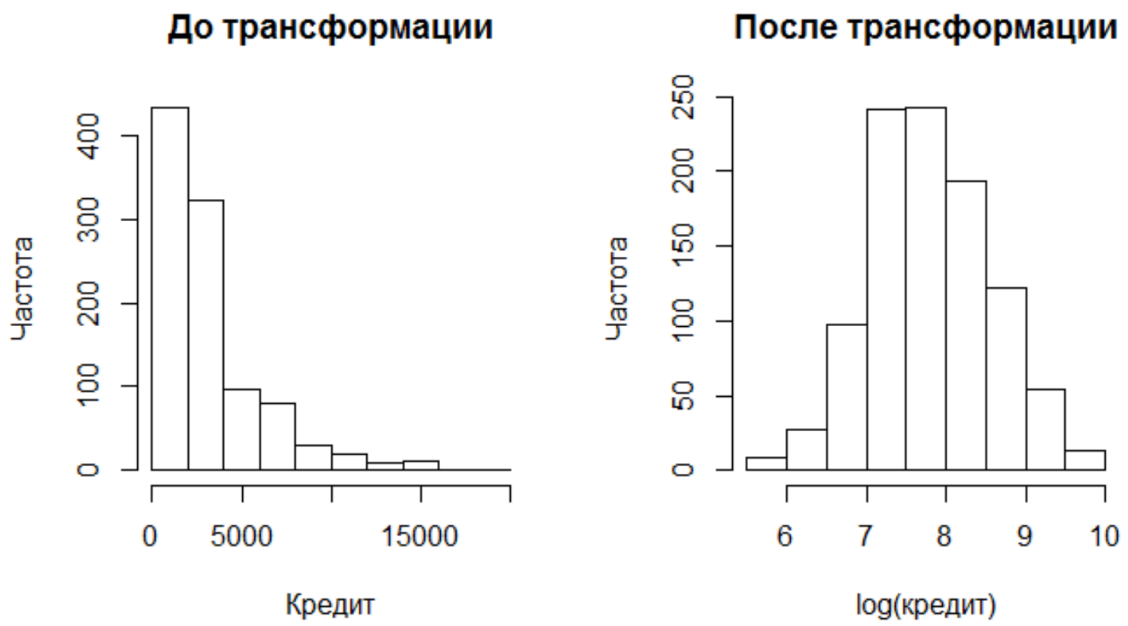
- Тренд, сезонность => нестационарность ряда
- Критерий Дики-Фуллера

Стабилизация дисперсии

- ➤Логарифмирование
- ➤Преобразование Бокса Кокса

Для последовательности $y_1, \ldots, y_n$, где $y_i > 0$:

$$y_i = \begin{cases} \frac{{y_i}^{\lambda} - 1}{\lambda}, \; if \; \lambda \neq 0 \\ log(y_i), \; if \, \lambda = 0 \end{cases}$$



Дифференцирование временного ряда

- Переход к попарным разностям соседних значений
- Сезонное дифференцирование – попарные разности значений в соседних сезонах

$$y'_t = y_t - y_{t-s}$$

- Позволяет стабилизировать среднее значение ряда
- Убирает тренд, сезонность
- Может применяться неоднократно

Методы прогнозирования временных рядов

- **AR(p)** (autoregression models)- это модель временного ряда, которая использует наблюдения с предыдущих временных шагов для уравнения регрессии, чтобы предсказать значение на следующем временном шаге.

$$y_t = \alpha + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \varepsilon_t$$

Метод скользящего среднего

- **MA(q)** (moving-average model) модель утверждает, что значение нашего ряда представляет собой линейную комбинацию q последних значений шумовой компоненты.

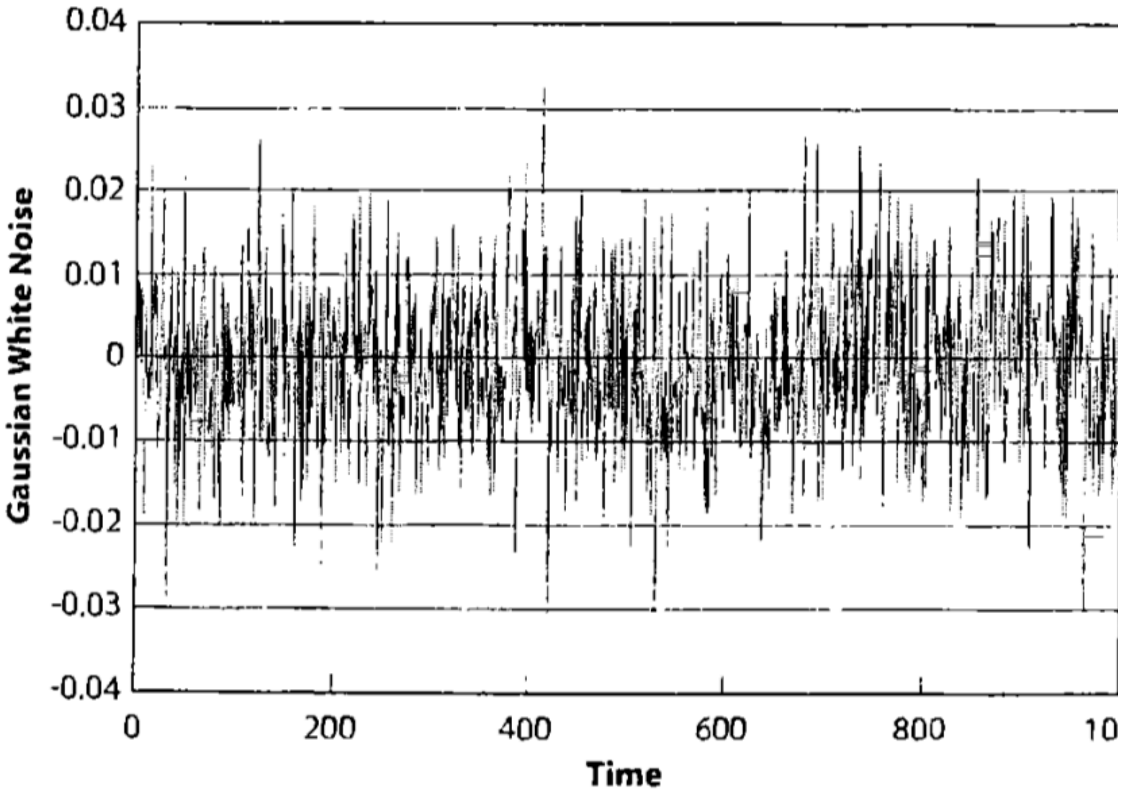
$$y_t = \alpha + \varepsilon_t + \theta_1 \varepsilon_{t-1} \ldots + \theta_q \varepsilon_{t-q}$$

α среднее значение ряда, ε_i – белый шум

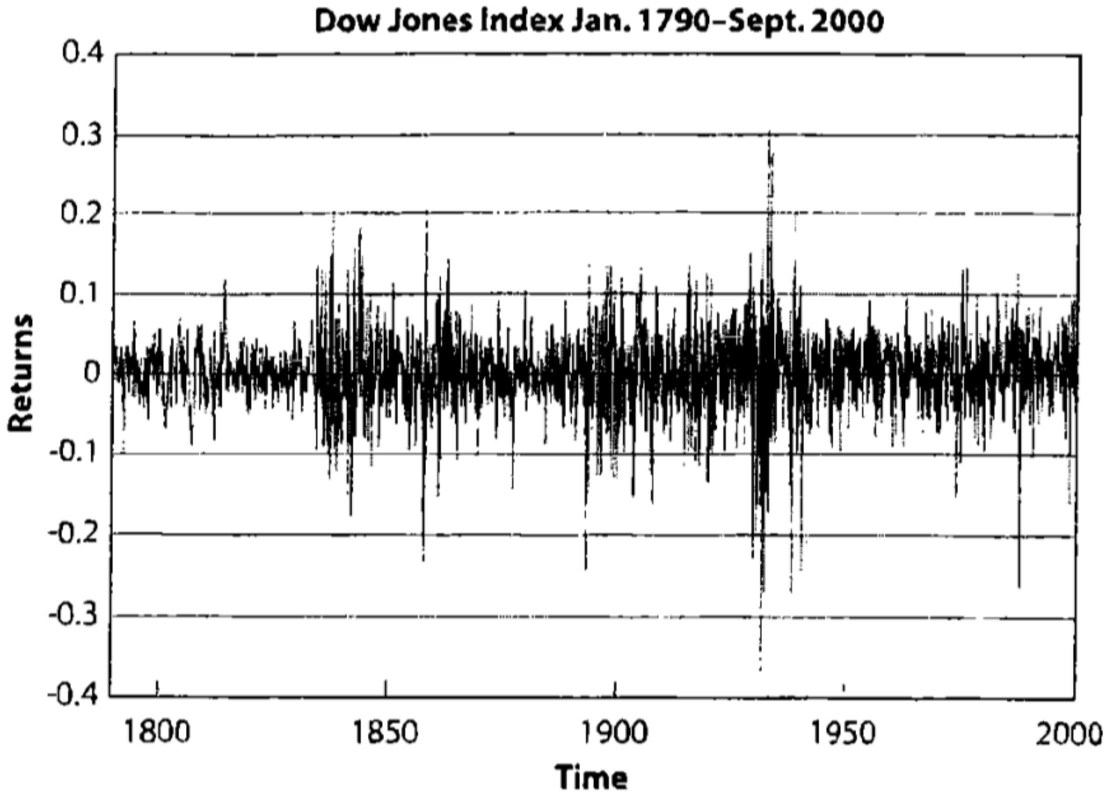
Белый шум

- Непрерывный во времени случайный процесс w(t) является белым шумом тогда и только тогда, когда его математическое ожидание и автокорреляционная функция удовлетворяют следующим равенствам соответственно:

$$\mu_w(t) = \mathbb{E}\{w(t)\} = 0$$
$$R_{ww}(t_1, t_2) = \mathbb{E}\{w(t_1)w(t_2)\} = \sigma^2\delta(t_1 - t_2).$$



Gaussian White Noise
0.04
0.03
0.02
0.01
0
-0.01
-0.02
-0.03
-0.04
0
200
400
600
800
10
Time



Dow Jones Index Jan. 1790–Sept. 2000
Returns
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
1800
1850
1900
1950
2000
Time

ARMA

- **ARMA(p, q)** состоит из авторегрессионной компоненты порядка p и компоненты скользящего среднего порядка q.

$$y_t = \alpha + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} \ldots + \theta_q \varepsilon_{t-q}$$

- Теорема Вольда утверждает, что любой стационарный временной ряд может быть описан моделью ARMA(p, q) с правильным подбором значений параметров p и q.

Autoregressive integrated moving average model

Модель ARIMA(p, d, q) для нестационарного временного ряда y_t имеет вид:

$$\Delta^d y_t = \alpha + \sum_{i=1}^{p} a_i \Delta^d y_{t-i} + \sum_{j=1}^{q} b_j \varepsilon_{t-j} + \varepsilon_t$$

где c, a_i, b_j – параметры модели. Δ^d – оператор разности временного ряда порядка d

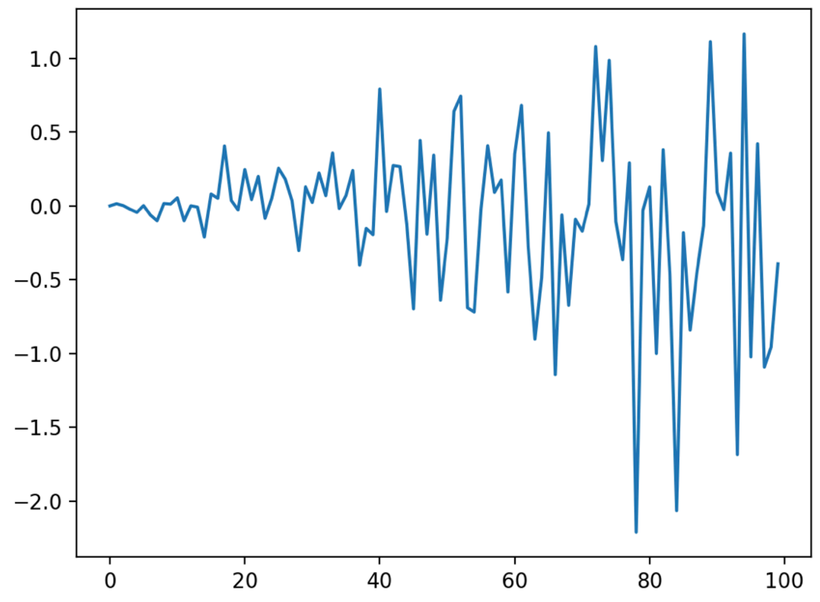
Недостатки ARIMA

- Требует приведения к стационарному ряду
- Автоматические прогнозы ARIMA склонны к большим ошибкам тренда.
- Экспоненциальное сглаживание и сезонные наивные прогнозы фиксируют недельную сезонность, но упускают долгосрочную сезонность.
- Методы слишком остро реагируют на падение в конце года, потому что он не адекватно моделируют ежегодную сезонность.
- Не учитывает гетероскедастичность

Autoregressive Conditional Heteroskedasticity model

Модель применяется для анализа временных рядов, у которых условная (по прошлым значениям ряда) дисперсия ряда зависит от прошлых значений дисперсий этого ряда и иных факторов.

- $y_t = \beta_0 + \beta_1 y_{t-1} + u_t$
- $u_t = \sigma_t v_t, v_t \sim N(0, 1)$
- $\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \cdots + \alpha_q u_{t-q}^2$
- (модель для y_t можно варьировать, распределение v_t тоже)



Line Plot of Dataset with Increasing Variance

Аддитивная модель fbprophet

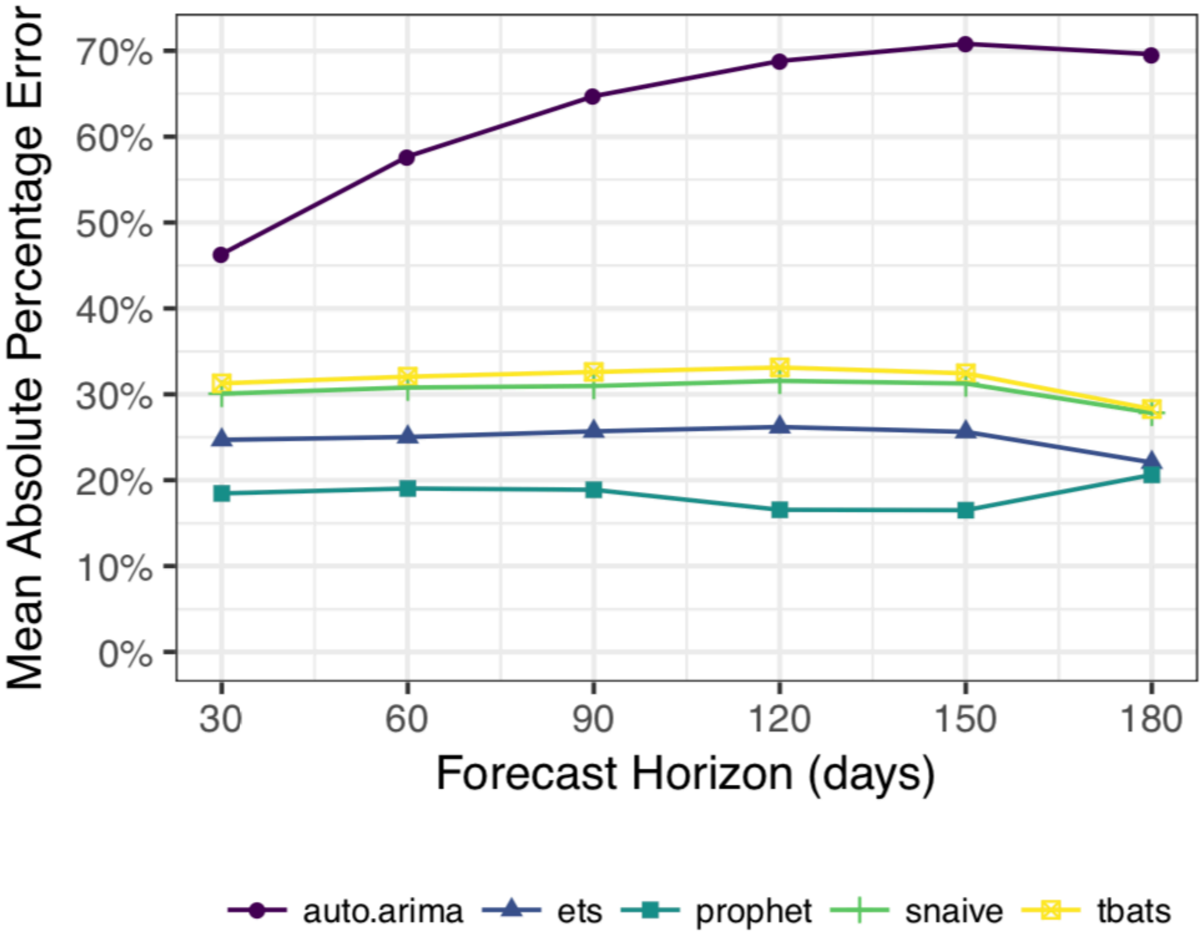
$$y(t) = g(t) + s(t) + h(t) + \epsilon_t$$

- Сезонные компоненты s(t)
- Тренд g(t)
- Аномальные дни h(t)
- Ошибка єt

Сравнение ARIMA и fbprophet

SARIMA : MAPE=16.54%, MAE=7.28 поста.
Fbprophet: MAPE=26.79%, MAE=8.49 поста.

Сравнение ARIMA и fbprophet



Список литературы

- 1. https://habr.com/ru/company/ods/blog/323730/
- 2. Sean J. Taylor, Benjamin Letham. Facebook, Menlo Park, California, United States. Forecasting at Scale
- 3. https://bamboo.nes.ru/cb.cgi/attachdownload/topic4-2.pdf?page=Materials&file=topic4-2.pdf
- 4. https://ru.wikipedia.org/wiki/Белый_шум
- 5. https://www.coursera.org/learn/data-analysis-applications
- 6. https://towardsdatascience.com/time-series-forecasting-arima-models-7f221e9eee06
- 7.http://www.machinelearning.ru/wiki/images/archive/c/cb/20160412121749!voron-ml-forecasting-slides.pdf
- https://en.wikipedia.org/wiki/Moving-average_model

Список вопросов

- Как устроена модель ARIMA?
- Как в аддитивной модели fbprophet рассчитывается компонента недельной сезонности?
- Опишите два способа приведения временного ряда к стационарному?
- В чем преимущества модели ARCH по сравнению с ARIMA?