

# Практикум 3 курс. Осень 2022- Весна 2023. Задание 2

## Описание

В этом задании вам предстоит провести анализ некоторого временного ряда и попробовать предсказать значения для последующих месяцев.

В рамках данного задания Вы:

- научитесь правильно считывать данные и их визуализировать;
- поймете как определять свойства временных рядов и познакомимся с различными моделями для предсказания значений;
- оцените качество полученных моделей.

Цель задания 2 в том, чтобы продолжить ваше изучение языка Python, попутно познакомиться с такими библиотеками, как Pandas (работа с данными), statsmodels(работа с различными статистическими моделями, в том числе и временными рядами).

## Что Вам предстоит сделать:

Считать данные из [training.csv](#). Ответы на тестовой выборке [testing.csv](#) не следует использовать ни в каких экспериментах, кроме финального. Проверить является ли ряд стационарным в широком смысле. Это можно сделать двумя способами:

1. Провести визуальную оценку, отрисовав ряд и скользящую статистику(среднее, стандартное отклонение). Постройте график на котором будет отображен сам ряд и различные скользящие статистики.
2. Провести тест Дики - Фуллера.  
Сделать выводы из полученных результатов. Оценить достоверность статистики. (25 баллов)
3. Разложить временной ряд на тренд, сезонность, остаток в соответствии с аддитивной, мультипликативной моделями. Визуализировать их, оценить стационарность получившихся рядов, сделать выводы. (15 баллов)
4. Проверить является ли временной ряд интегрированным порядка  $k$ . Если является, применить к нему модель ARIMA, подобрав необходимые параметры с помощью функции автокорреляции и функции частичной автокорреляции. Выбор параметров обосновать.

Отобрать несколько моделей. Предсказать значения для тестовой выборки.

Визуализировать их, посчитать  $\sim r^2 \text{ score} \sim$  для каждой из моделей. Произвести отбор наилучшей модели с помощью информационного критерия Акаике. Провести анализ получившихся результатов. (50 баллов)

**За все правильно выполненные пункты можно получить 90 баллов.**

+10 баллов - соблюдение [PEP8](#)

+10 баллов - использование для визуализации библиотек `bokeh` или `seaborn`. Надо сделать, чтобы было красиво 😊.

30 баллов - конкурсные. Вы можете использовать различные статистические модели и методы машинного обучения (ARIMA, ES, ARFIMA, LSTM, различные виды регрессий, etc). 30 баллов получает группа, получившая наилучший результат среди участвующих.

## Сдача

Задание должно выполняться группой из **2-3 человек**

Процесс стандартный, и описан в [README.org](#), секция Выполнение и сдача заданий.

## Сроки

- Сдача до 🍬 4 🍬 декабря 2022г.
- Прохождение ревью до 🍬 11 🍬 декабря 2022г.

[Pandas](#) — библиотека для работы со структурированными данными.

[Statsmodels](#) — библиотека для построения различных статистических моделей.

## Результаты

[Ознакомиться с результатами оценивания](#)