

Университет Иннополис

Дипломная работа на тему
«Анализ коэффициента энергетической эффективности жилого
здания»

Выполнил: Гасимов И.А.

Руководитель: Шестакова Е.А.

2022 г.

Введение

Данная тема дипломного проекта была выбрана в силу своей актуальности и широкого круга проблем, которую может решить правильный анализ показателей коэффициента энергетической эффективности. Часто возникает необходимость сравнить энергоэффективность обследуемого здания с другими аналогичными зданиями. Или, в общем, понять, получить индикатор, где находится здание на линейке от самого эффективного пассивного дома до самого неэффективного здания постройки прошлого века. В том числе для определения сроков капитального ремонта при высоких потерях тепла. Для этих целей лучше всего использовать, индекс энергоэффективности здания.

Именно такой датасет мы рассмотрим в данном проекте для среднестатистического дома. Если нам удастся понять и спрогнозировать изменение данного показателя на нашем временном отрезке, то мы сможем использовать данную формулу для определения текущего состояния инфраструктуры всех зданий в данной климатической зоне.

Цель: Провести анализ данных в формате CSV с выгрузкой рассчитанных коэффициентов за каждый месяц, начиная с 1 января 1970 года по 1 декабря 1989 года, для жилого здания и на основании полученных результатов предсказать изменение данного коэффициента в 1990 году.

Задачи:

- 1) Требуется проанализировать ежемесячные показатели индекса энергоэффективности здания на протяжении 20 лет.
- 2) Составить прогноз энергоэффективности на следующий расчетный год.

Выполнение работы.

1. Первичное знакомство и обработка данных.

Работа была выполнена в Python с загрузкой нескольких функциональных библиотек, моделей и метрик.

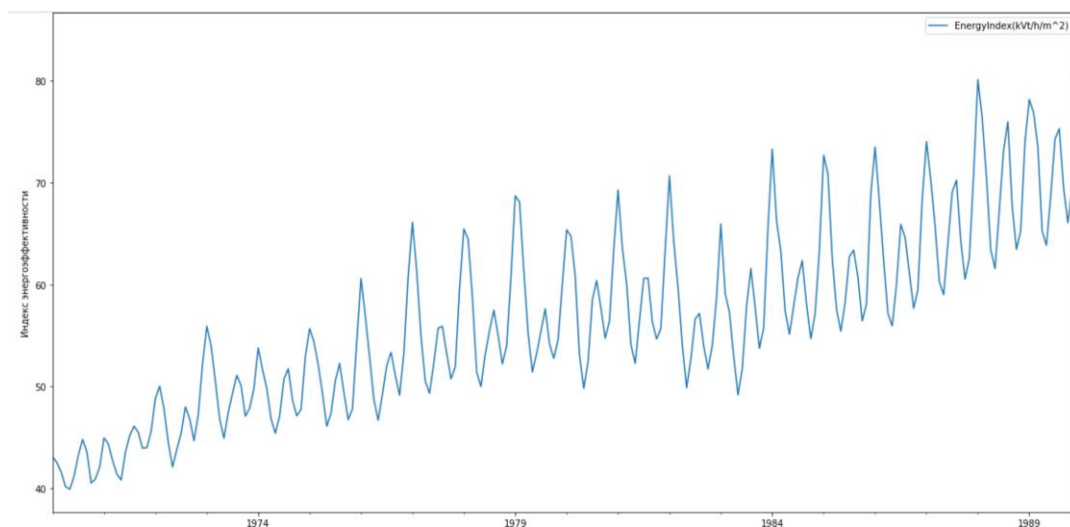
Рассматриваемый дата сет состоит из 2х колонок («DATE» и «EnergyIndex») и 240 строк (месяцы и соответствующее показание индекса).

Данные полные (нет отсутствующих или пустых строк).

Для удобства переименовали название представленных данных для информативности (чтобы представлять единицу измерения представленных данных). Эти же данные (коэффициент энергоэффективности «EnergyIndex» взяли в качестве основной метрики для нашего предсказания.

2. EDA (exploratory data analysis) или разведочный анализ данных.

Для начала провели расчет основных статистических метрик для построения более полной картины по нашему датасету и построили график для визуализации данных.



Было наглядно показано, что индекс энергоэффективности растет с каждым годом, а это значит, что с течением времени требуется все больше энергии для отопления квадратного метра жилой площади. Также из графика видны логичные сезонные колебания - зимой затрачивается больше энергии на отопление, а летом незначительный скачок из-за потребления электричества кондиционерами для охлаждения помещений.

На этом основании можно предположить, что коэффициент энергоэффективности с каждым последующим годом будет увеличиваться (для отапливания зимой и кондиционирования летом будет затрачиваться все больше энергии на каждый квадратный метр жилья).

3. Построение моделей и анализ результатов.

Основные задачи, которые стоят на данном этапе:

- описать модель
- подобрать оптимальные параметры
- создать модель
- обучить модель
- сделать прогноз на период тестовой выборки
- сравнить прогноз с тестовой выборкой (построить график)
- оценить качество прогноза
- сделать прогноз на будущее
- сделать выводы о работе данного метода прогнозирования

Начнем рассмотрение с модели SARIMAX.

```

SARIMAX Results
Dep. Variable:  y                               No. Observations: 240
Model:          SARIMAX(3, 0, 1)x(0, 1, 1, 12)  Log Likelihood   -413.258
Date:           Thu, 24 Mar 2022                AIC             840.516
Time:           17:25:02                        BIC             864.522
Sample:         0                               HQIC           850.202
               - 240

Covariance Type: opg

      coef  std err      z    P>|z| [0.025 0.975]
-----
intercept 0.0475  0.046   1.044  0.296 -0.042 0.137
ar.L1      1.6309  0.166   9.805  0.000  1.305  1.957
ar.L2     -0.8740  0.155  -5.642  0.000 -1.178 -0.570
ar.L3      0.2157  0.091   2.377  0.017  0.038  0.394
ma.L1     -0.7386  0.174  -4.250  0.000 -1.079 -0.398
ma.S.L12 -0.5458  0.059  -9.263  0.000 -0.661 -0.430
sigma2     2.1500  0.148  14.511  0.000  1.860  2.440

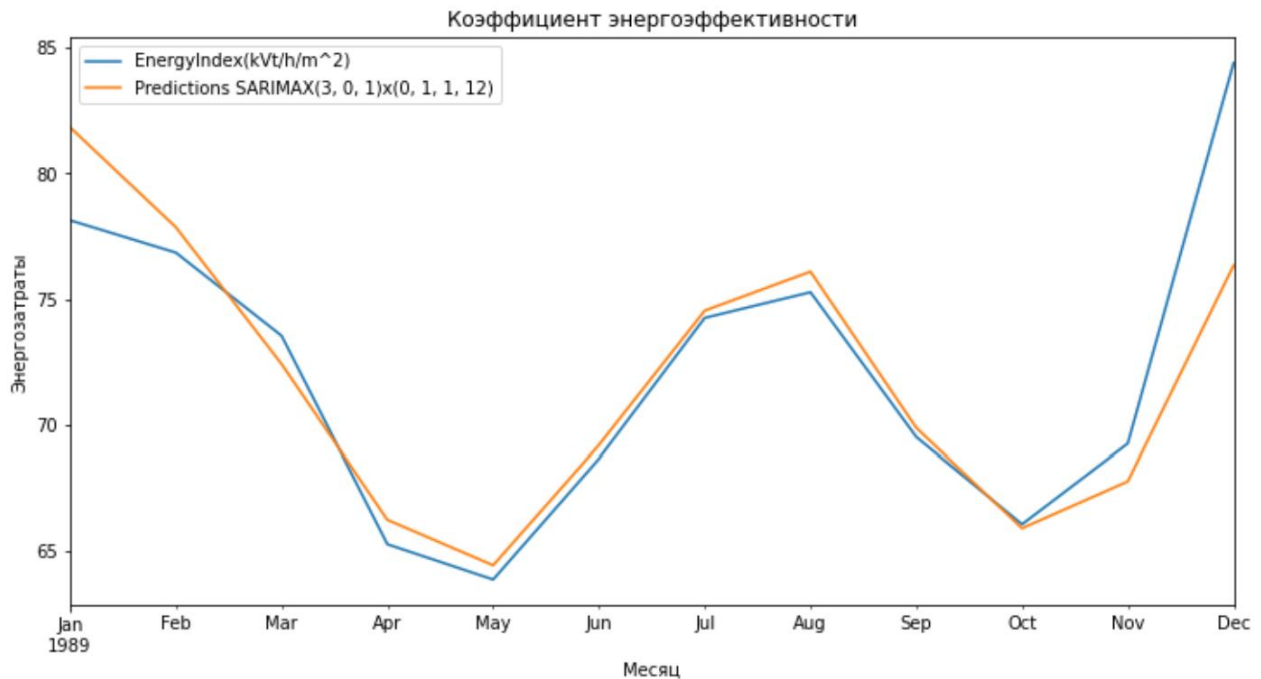
Ljung-Box (L1) (Q):  0.00 Jarque-Bera (JB): 64.01
Prob(Q):              0.95 Prob(JB):      0.00
Heteroskedasticity (H): 2.07 Skew:         0.44
Prob(H) (two-sided):  0.00 Kurtosis:      5.44

```

При автоматическом подборе параметров была выбрана модель SARIMAX(3, 0, 1)x(0, 1, 1, 12) с сезонностью 12 месяцев.

Модель обучена на выборке и построен прогноз на основе тестовой выборки.

Получены графики для сравнения выборок:



Рассчитали основные критерии оценки качества модели:

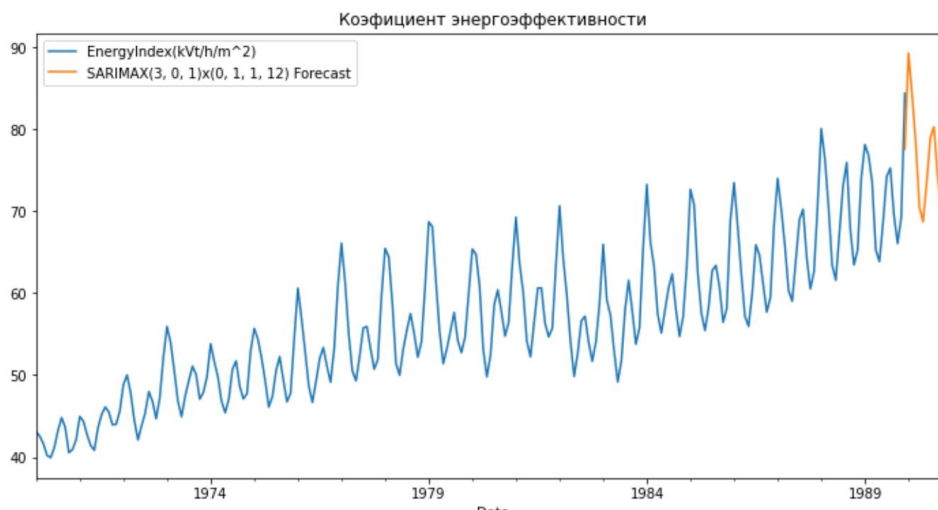
MAE Error: 1.59

MSE Error: 7.14

RMSE Error: 2.67

MAPE Error: 2.05

На основании этих данных построили графическую модель прогнозируемого поведения нашего параметра:

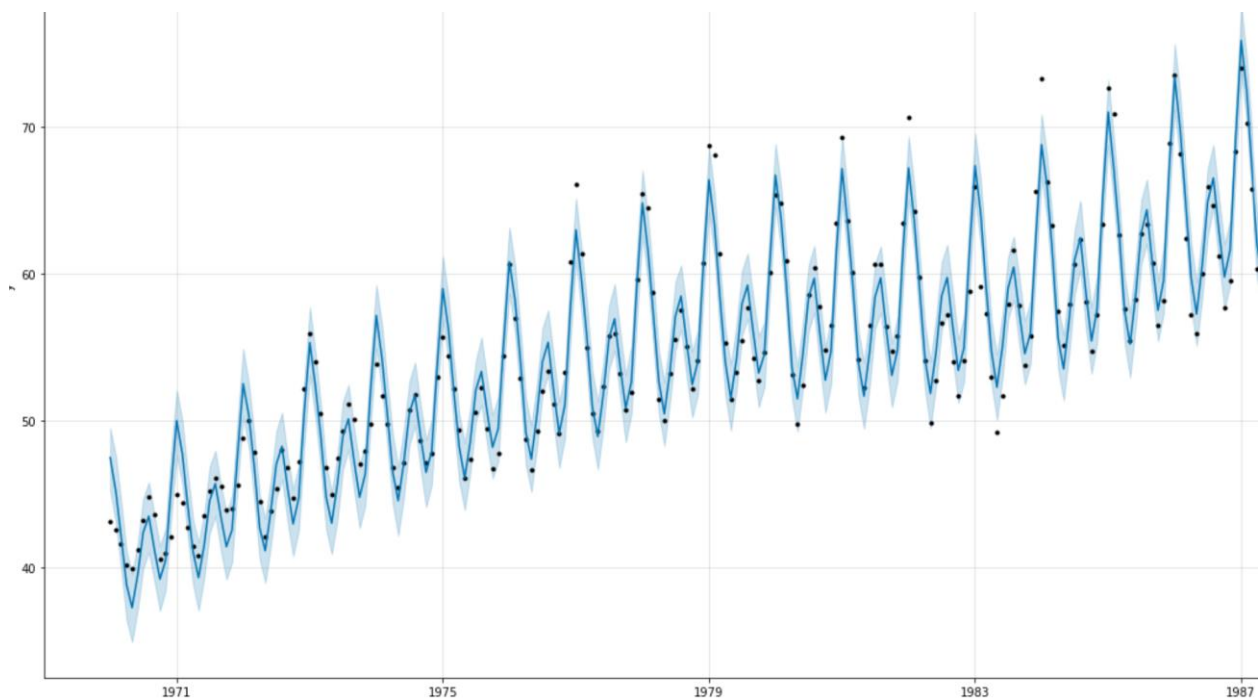


По данному типу построения модели получили следующие выводы:

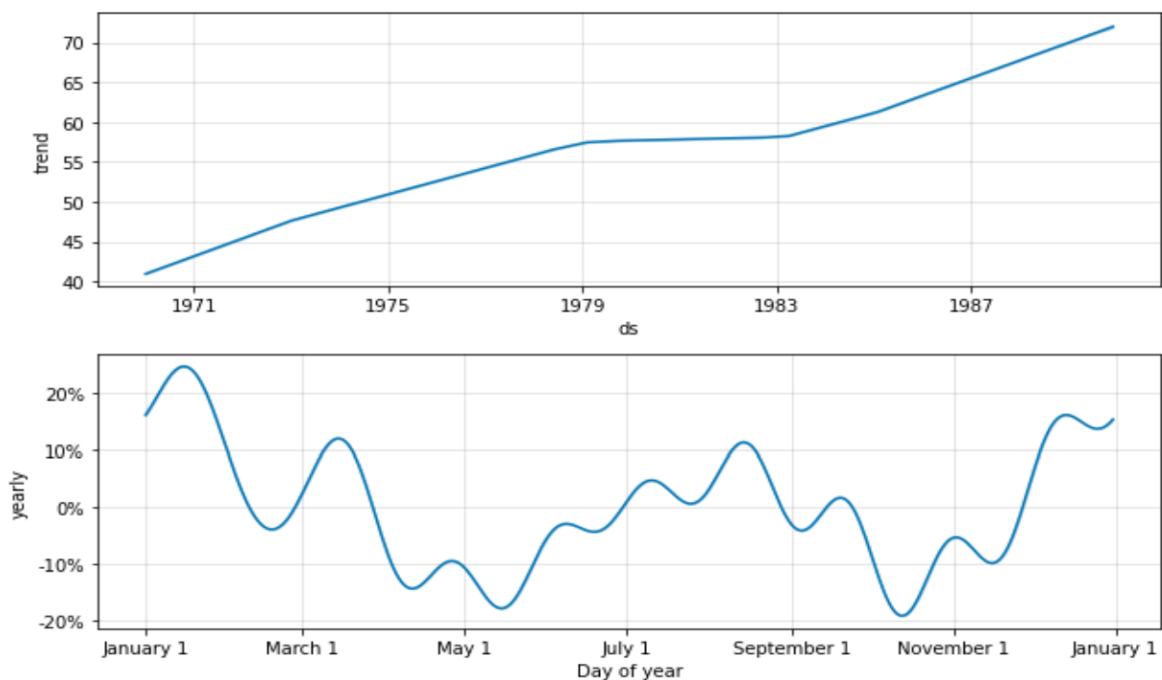
- Модель показала себя хорошо, $RMSE=2.67$ кВт/ч/м², это хороший показатель.
- Также процент рассчитанной ошибки $MAPE=2\%$, это хороший результат.
- Согласно графику мы видим, что тренд и высота амплитуда были отображены корректно, общая динамика прослеживается.

Продолжим рассмотрение на модели **PROPHET**.

Для начала подготовили данные для корректной обработки данных: переименовали столбцы и автоматически модель подобрала необходимые параметры (модель проигнорировала дневные и недельные сезонности) и обучила ее на основе годовой сезонности. Далее был построен прогноз на 12 месяцев и графически представлен:



А также был рассчитан общий тренд и сезонность по одному году:



Видим, что тренд энергоиндекса возрастающий и у нас есть годовая сезонность. Видим каким образом изменяется коэффициент по месяцам.

Рассчитали основные критерии оценки качества модели:

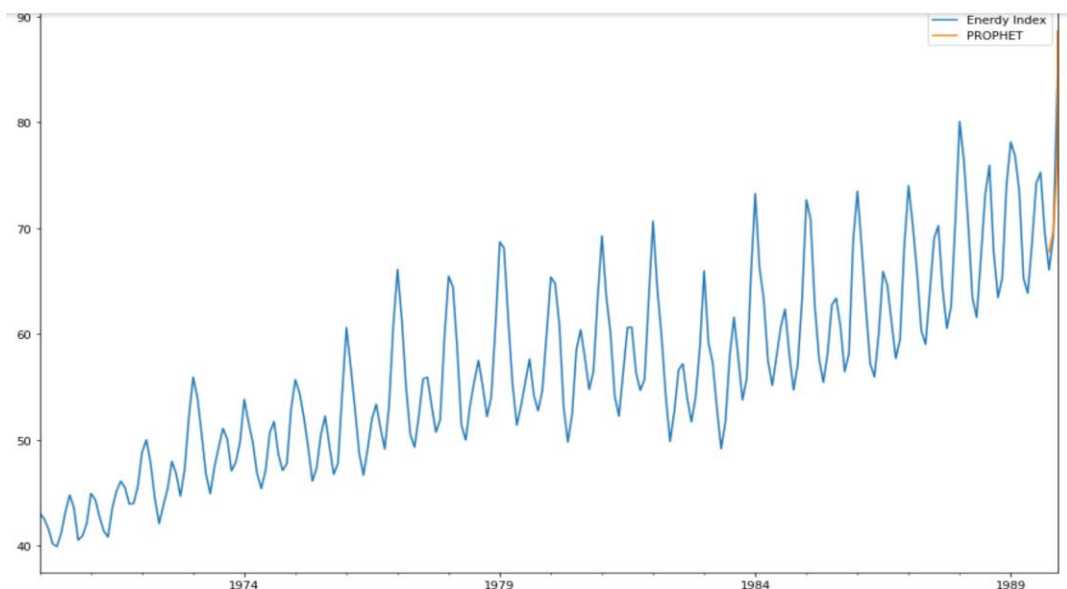
MAE Error: 1.838011803

MSE Error: 7.669972891

RMSE Error: 2.769471591

MAPE Error: 2.403566144

На основании этих данных построили графическую модель прогнозируемого поведения нашего параметра:



Выводы работы метода PROPHET: Модель показала себя хорошо, $RMSE=2.7$ кВт/ч/м², это хороший показатель. Процент рассчитанной ошибки $MAPE=2,4\%$, это хороший результат, такому прогнозу можно доверять. На графике видно, что тренд и амплитуда отображается корректно. Прослеживается общая динамика роста показателя.

Итог:

Проведен анализ данных с использованием современных методов обработки статистической информации.

Рассчитаны основные статистические метрики, позволяющие судить о характере исследуемого явления.

По результатам анализа представленных данных можно выявить зависимость коэффициента энергоэффективности от сезона года. Также можно смело утверждать, что с каждым годом коэффициент растет, а значит требуется все больше энергии для обеспечения комфортных температурных условий в здании - инфраструктура изнашивается.

Сравнивая две модели прогноза, можно сделать вывод, что и SARIMAX, и PROPHET показали приблизительно равные значения $RMSE$ (2.7 и 2.7) и $MAPE$ (2% и 2.4%). Но SARIMAX выглядит предпочтительнее.