

Дипломная работа

Автор: Полушина В.В.

На тему: "Потребление электроэнергии в 1-ой квартире в МКД".

Введение: Для проведения анализа взята статистика по потреблению электроэнергии в многоквартирном доме на примере 1-ой квартиры (40 квадратных метров) в период с 2015 по 2021 год включительно.

Цель: Проведение исследования данных и построение прогноза потребления электроэнергии в 1-ой квартире на 1 год вперед.

Задачи:

1. Провести анализ данных о потреблении электроэнергии в квартире;
2. Построить прогноз потребления электроэнергии на 1 год вперед с помощью различных моделей прогнозирования для выявления наиболее хорошей модели.

Для проведения дипломной работы был использован *Python* и с помощью различных библиотек (точный перечень можно посмотреть в рабочем скрипте в GitHub) были построены три модели:

1. SARIMAX;
2. PROPHET;
3. Экспоненциальное сглаживание (Exponential smoothing).

- I. Прежде чем приступить к построению моделей необходимо внимательно изучить статистические данные (датасет), убедиться в их корректности и возможности использовать, а именно:
 1. Корректность загрузки данных;
 2. Последовательность даты;
 3. Проверка на наличие пропусков в датасете;
 4. Проверка формата данных, числовые данные должны быть в формате число двойной точности (*dtype: float64*).
- II. Проведение разведочного анализ данных (EDA: exploratory data analysis):
 1. Расчет основных статистических метрик;
 2. Построение общего график.

Выводы:

1. Согласно графику видно, что рост потребления электроэнергии растет раз в год. Также с каждым годом потребление электроэнергии растет. Высокое потребление связано с зимой, в холодное время года растет потребление из-за короткого светового дня; установка дополнительной электроники (увлажнитель воздуха, обогреватели).
2. Виден высокий уровень корреляции, между потреблением и месяцем. Также выявлен положительный тренд потребления электроэнергии. **На основании этого можно построить гипотезу:** будет ли в будущем расти потребление электроэнергии с учетом появления новых электроприборов и работы на дому с учетом сезонности.
- III. После того как данные были обработаны и изучены, можно приступить к построению моделей:
 - Для всех моделей необходимо определить размер выборки и задать тестовую выборку (test) размером 1 год;
 - Построить декомпозицию временного ряда (ETS композиция).

Вывод: На основании ETS можно наблюдать годовую сезонность и положительный тренд.

1. SARIMAX:

- Задаем набор параметров для модели включая сезонность и период сезонности равный 12. Получаем модель с параметрами SARIMAX(0, 1, 0)x(1, 0, 1, 12);
- Создаем модель с подобранными параметрами;
- Обучаем модель на обучающей выборке;
- Получаем предсказанные значения и оцениваем качество прогноза;
- Сравниваем результаты на графике;
- Оцениваем качество модели методом MSE, RMSE, MAE, MAPE:
 - $MSE = 8,3$
 - $RMSE = 2,9$
 - $MAE = 2,2$
 - $MAPE = 3$
- Делаем прогноз на 1 год вперед;
- Задаем точки будущего и строим график.

Выводы работы метода SARIMAX:

1. Модель показала себя хорошо, $RMSE=2.9$ кВт/ч, это очень хороший показатель.
2. Также процент рассчитанной ошибки $MAPE=3\%$, это хороший результат.
3. Согласно графику, на будущее видим, что тренд и высота амплитуда были отображены корректно, общая динамика прослеживается.

2. PROPHET:

- Необходимо подготовить данные для модели, сбросить индекс и переименовать столбцы;
- Подгоняем модель под наши данные:
Как видим, алгоритм сам нам подсказывает, что он проигнорировал, выбирая параметры.
Disabling weekly seasonality. Run prophet with weekly_seasonality=True to override this.
 - недельную сезонность (для недельной сезонности нам нужно иметь данные минимум за месяц, чтобы суметь использовать ее в модели) *INFO:fbprophet:Disabling daily seasonality. Run prophet with daily_seasonality=True to override this.*
 - дневную сезонность (дневная сезонность может использоваться в случае, если данные собираются по часам/минутам, в нашем случае данные представлены по месяцам).*Зато он обнаружил годовую сезонность и использовал его при настройке модели.*
- Строим дата фрейм на 12 месяцев вперед;

- Предсказываем значения модели с доверительным интервалом в 95%;
- Выделяем основные поля для прогнозирования:
 - *ds* — дата прогноза
 - *yhat* — спрогнозированное значение
 - *yhat_lower* — нижняя граница доверительного интервала для прогноза
 - *yhat_upper* — верхняя граница доверительного интервала для прогноза
- Смотрим прогноз на графике и раскладываем ряд на основные компоненты:
 - Видим, что тренд потребления электроэнергии возрастающий и у нас есть годовая сезонность. Видим каким образом изменяется потребление по месяцам.
- Оцениваем качество модели методом MSE, RMSE, MAE, MAPE:
 - $MSE = 14$
 - $RMSE = 3,7$
 - $MAE = 3,6$
 - $MAPE = 23,4$
- Делаем прогноз на 1 год вперед;
- Устанавливаем индексы;
- Строим график на будущее.

Выводы работы метода PROPHET:

1. Модель показала себя хорошо, $RMSE=3.7$ кВт/ч, это хороший показатель.
2. Но процент рассчитанной ошибки $MAPE=23,4\%$, это плохой результат, такому прогнозу определенно нельзя доверять.

3. Экспоненциальное сглаживание (Exponential smoothing):

- Создаем модель с подобранными параметрами;
- Обучаем модель на обучающей выборке;
- Предсказываем значения, передав модели точку начала и окончания;
- Сравниваем прогноз и тестовую выборку;
- Оцениваем качество модели методом MSE, RMSE, MAE, MAPE:
 - $MSE = 74$
 - $RMSE = 8,6$
 - $MAE = 7,9$
- Делаем прогноз на год вперед;
- Строим график на будущее.

Выводы работы метода "Экспоненциальное сглаживание" (Exponential smoothing):

1. Модель показала себя плохо, $RMSE=8.6$ кВт/ч, это не очень хороший показатель, слишком высокий уровень ошибки. Такому прогнозу нельзя доверять.

IV. Сравнительный анализ трех моделей:

<i>Качество модели/Модель</i>	<i>SARIMAX</i>	<i>PROPHET</i>	<i>Экспоненциальное сглаживание</i>
<i>MSE</i>	8,3	14	74
<i>RMSE</i>	2,9	3,7	8,6
<i>MAE</i>	2,7	3,6	7,9
<i>MAPE</i>	3	23,4	nan
Выбор модели	+		

Общие выводы:

- Проведен анализ данных с использованием современных методов обработки статистической информации.
- Рассчитаны основные статистические метрики, позволяющие судить о характере исследуемого явления.
- Результаты анализа представленных данных помогли выявить зависимость потребления электроэнергии в зависимости от сезона года.
- Сравнивая три модели прогноза, лучшего всего себя показала модель SARIMAX, согласно показателям ошибок RMSE (2,8 кВт/ч) и MAPE (3%) данная модель делает лучший прогноз, которому можно доверять.