

Ольга Полякова

**Исследование структуры энергетики (поиск инсайтов,
составление рекомендаций стейкхолдерам, построение
предиктивной модели цен)**

Дипломная работа по программе профессиональной
переподготовки
(DA-26 “Аналитика данных”)

Нетология

Цель работы:

определение возможностей прогнозирования в национальной энергетике на основе набора данных об энергетическом секторе Испании как области применения машинного обучения

Задачи исследования:

- обозначить наиболее существенные глобальные тенденции, сопровождающие современную энергетическую трансформацию;
- охарактеризовать существующую структуру источников энергии в Испании в 2015-2018 гг.;
- проанализировать зависимость спроса, производства и цены электроэнергии от внешних факторов - погоды;
- определить наиболее гибкие к изменению спроса источники энергии в Испании;
- спрогнозировать цену на электроэнергию;
- смоделировать будущую структуру энергетики по источникам энергии, опираясь на существующие тенденции.

Данные: структура, источники, качество

Для теоретической справки:

Полученные из открытого источника [Our World in Data](#) csv-файлы данных о глобальных социально-экономических и экологических процессах

`urban-and-rural-population.csv`

`energy.csv`

`climate-change.csv`

`annual-change-low-carbon-energy.csv`

`annual-change-renewables.csv`

`annual-change-fossil-fuels.`

Для основной части исследования:

В работе использован датасет “Hourly energy demand generation and weather. Electrical demand, generation by type, prices and weather in Spain” [N.Jhana] отсюда [Hourly energy demand generation and weather | Kaggle](#), включающий почасовые данные о производстве электроэнергии из различных источников в Испании в 2015-2018 гг, ее цене и потреблении и данными о погоде от [Open Weather API](#) для 5 крупнейших городов Испании.

Данные: структура, источники, качество

Данные с портала [Our World in Data](#) имели изначально хорошее качество и не нуждались в предобработке

Датасет [Weather_features.csv](#) не имел пропусков данных, но содержал различное число записей с почасовыми показателями погоды для 5 испанских городов.

Датасет [Energy_dataset.csv](#) был подвергнут предварительному анализу. Обнаруженные пропуски данных были признаны случайными, столбцы с полностью пропущенными данными были удалены, строковые пропуски были заменены предыдущим значением

Анализ данных

В теоретической части проводится описательное исследование глобальных социально-экономических и экологических трендов, тенденций развития энергетики.

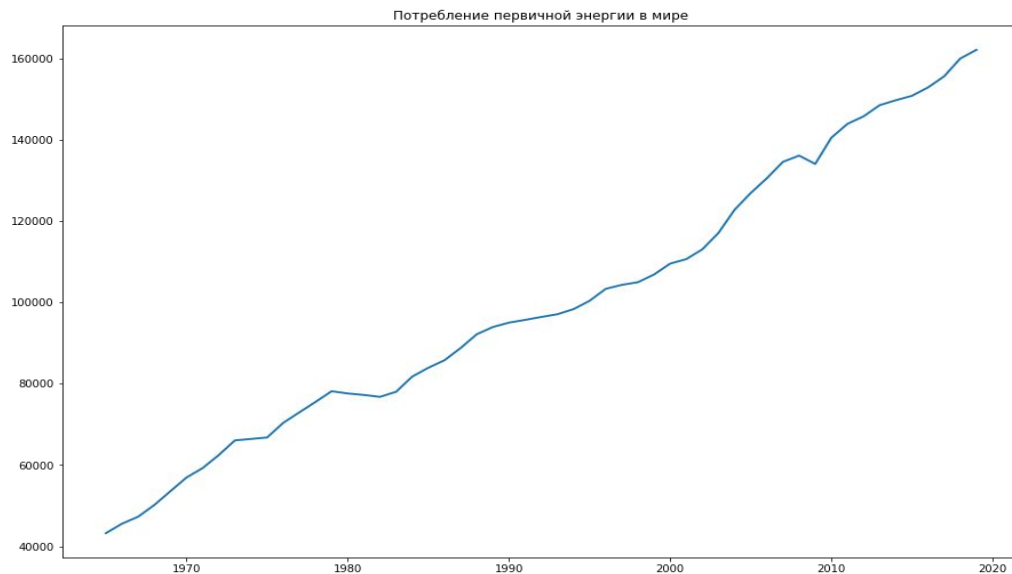
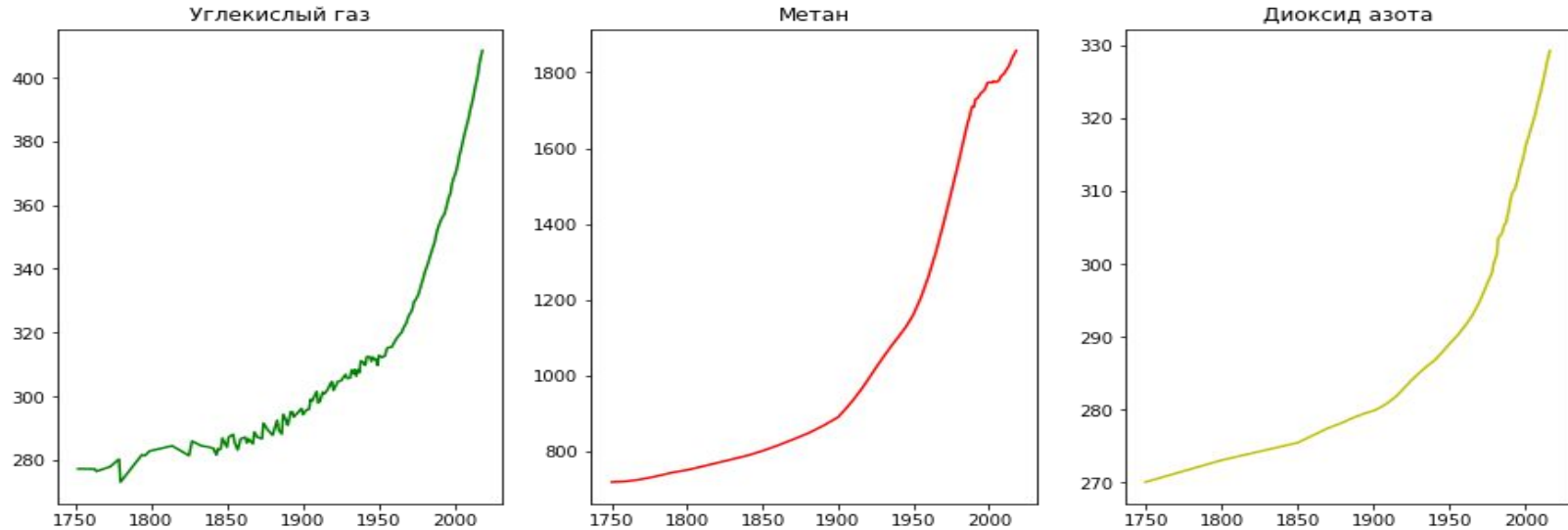


Диаграмма 2. Пример абсолютно драматического роста потребления энергии в мире за последние 60 лет наблюдений

В исследовании использовались методы функционального и сравнительного анализа, пропорционирования данных, расчета скользящей средней, корреляционного анализа.

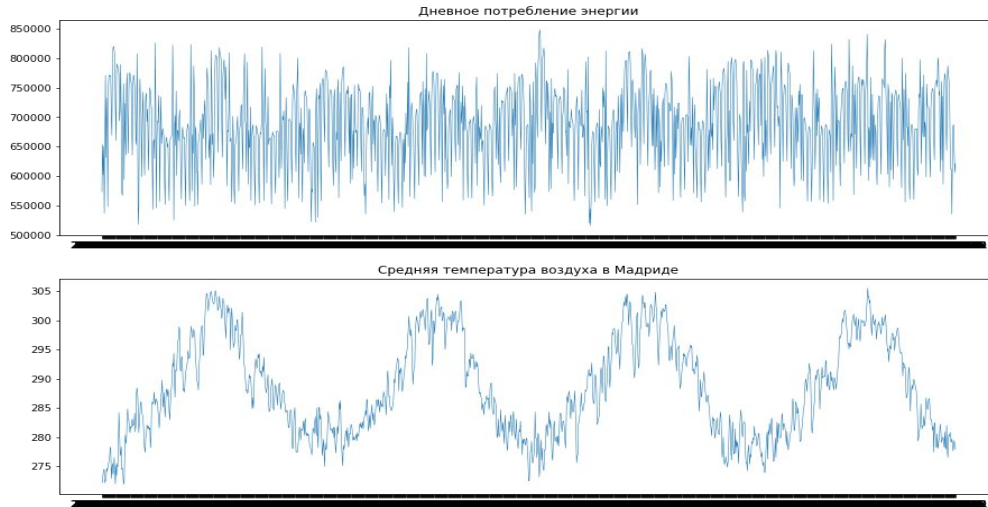


К примеру см. диаграмму 4. Сравнительный график концентрации парниковых газов в атмосфере

Анализ данных

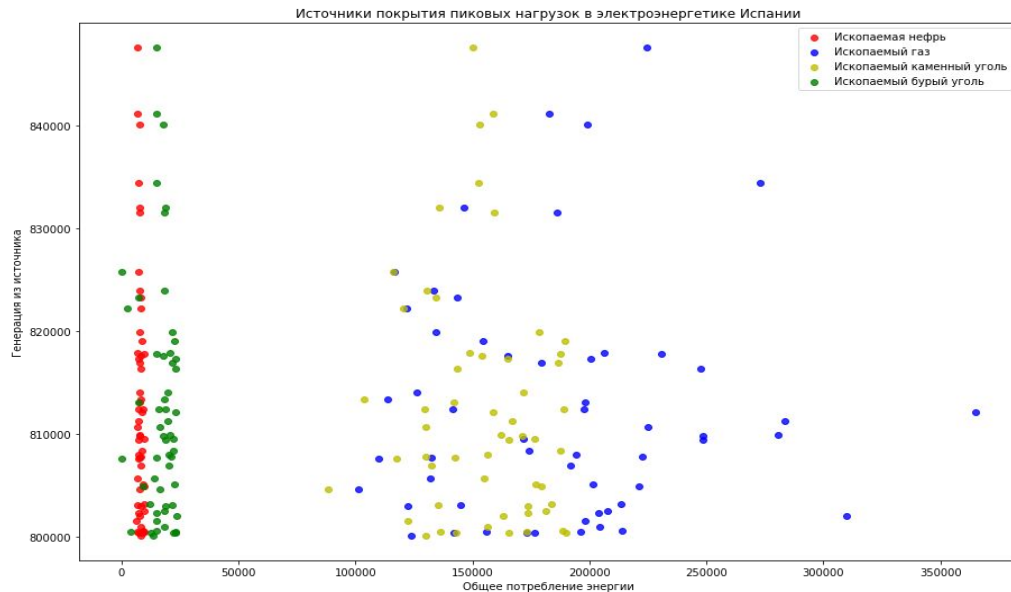
В анализе данных о национальной энергетике Испании использовались различные методы:

Метод корреляции - для обнаружения зависимости между погодными условиями и потребление энергии, для обнаружения источника покрытия пиковых энергонагрузок;



Визуализация
отсутствия
значимой
корреляции между
погодными
условиями и
потреблением
энергии

Метод факторного анализа использовался для определения основного ресурса для покрытия пиковых нагрузок в энергопотреблении



Метод группировки и динамического пропорционирования применялся в изучении закономерностей динамики структуры источников.

Доли высокоуглеродных и низкоуглеродных источников энергии в общем объеме генерации

2015



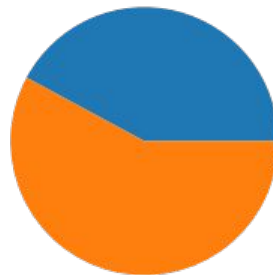
high_carbon=40.16 low_carbon=59.84

2016



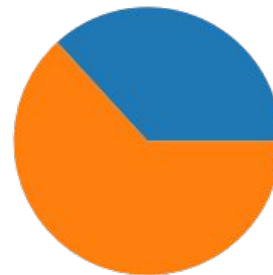
high_carbon=34.61 low_carbon=65.39

2017



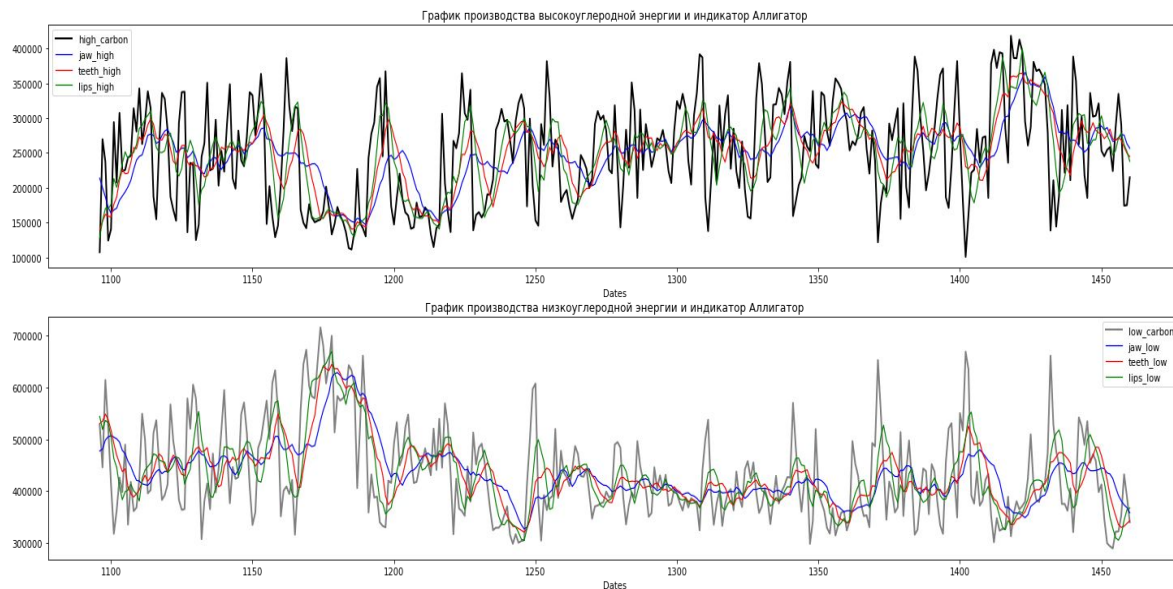
high_carbon=42.15 low_carbon=57.85

2018

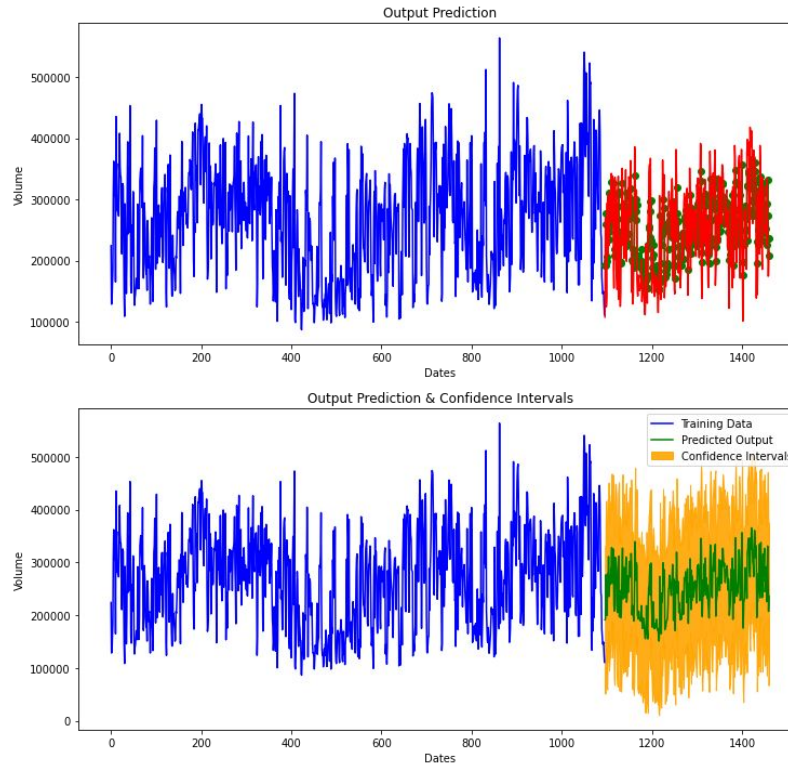


high_carbon=36.81 low_carbon=63.19

Особенностью исследования стало обращение к методу, смежного к аналитике данных биржевому техническому анализу - индикатору Б.Вильямса Аллигатор



Индикатор Аллигатор построен на сочетании различных по периоду скользящих средних и позволяет определить направленность тенденций развития энергетических процессов в настоящее время и ближайшем будущем



В исследовании
была предпринята
попытка изучения
данных в качестве
временных рядов
с применением
алгоритма ARIMA -
AutoRegressive
Integrated Moving
Average и
построением
прогноза будущих
значений
электрогенерации

Построена предиктивная модель цены электроэнергии на основе сравнения качества предсказания моделей *LinearRegression*, *RandomForestRegressor* и *XGBoost*

	<i>LinearRegression</i>	<i>RandomForestRegressor</i>	<i>XGBoost</i>
<i>RMSE</i>	11.737636816505695	3.213458723787507	3.5737653795526674
<i>R2-SCORE</i>	0.31187949775152124	0.9487953324512807	0.9366690250631801

Итоги работы:

Аналитические результаты:

1. Обнаружено, что потребление электроэнергии в Испании в существенной степени не зависит от такого внешнего фактора, как погода;
2. Установлено, что основным ресурсом для покрытия пикового увеличения нагрузки является природный газ;
3. Соотношение долей в структуре источников энергии, сгруппированных по критерию содержания в них углеводородов, имеет только циклическое развитие, а разработка иных индикаторов этих показателей также проявляет горизонтальное движение в постоянном диапазоне;
4. В исследовании была выявлена возможность числовых прогнозов энергетических показателей;
5. Была представлена предиктивная модель будущей цены на электроэнергию в Испании.

Спасибо за внимание!