Energy Demand Analysis

The objective of this project is to make one-step and multi-step forecast of electric energy consumption/demand.

Exploring the dataset

First, before creating the models suitable for our problem, we need to understand our data. We will do that by applying methods from Exploratory Data Analysis.

A quick summary of our findings is:

1. The dataset has approximately 26 000 entries and 4 columns. Date + Hour are used to create the index, Load is the target variable and T (temperature) is a feature.
2. The data is **stationary**.
3. The data is **highly autocorrelated**, meaning it is **not** a random walk.
4. The autocorrelation coefficients **do not become abruptly non-significant** after an arbitrary lag **q**, meaning it is not a MA(q) series.
5. The partial autocorrelation coefficients **become non-significant after a lag of p = 24** (one day), meaning it may be an autoregressive series.
6. By hour, the energy demand is the **lowest around 4AM** and **highest around 7PM**, which makes a lot of sense; at 4AM, everybody should be sleeping and not consuming energy and at 7PM, most of the household is at home, cooking dinner or using the appliances.
7. By month, **highest consumption is winter and summer months**, while during **spring and autumn it's lower**. I would argue this is because during winter months, the household uses heating a lot, while during summer months, they would use the AC.
8. During the days of the month, the energy demand stays at the same level.
9. The energy demand is **virtually the same in all three years**.

Data preprocessing

The temperature column was removed, data was split into train/test set in ratio 2:1 (first 2 years, last year) and then the entire DataFrame was transformed using the sliding window with WINDOW_IN = 24*7 (7 days).

Applying the models

Statistical models

We have used the Autoregressive Model from the statsmodel package, because we identified the series as an autoregressive series.

We have also tried to consider combine the AR and MA models using the SARIMAX class from the statsmodel package, with (p, q, d, s) = (2, 0, 2, 0), since our data does not need differentiating (=> d = 0) and is not seasonal (=> s = 0).

But the residual analysis gave us the conclusion that this would not be a good model, so it is not considered in the final models' evaluation.

Machine learning models

Random forest:

A high score of 0.9934 was obtained, indicating strong predictive abilities.

A low Mean Absolute Error (MAE) of 31.23 was obtained, indicating accurate one-step forecasting.

Captured non-linear relationships and intricate patterns in the dataset with tenacity.

XGBoost:

Competitively scored with a high score of 0.9925, demonstrating effective predictive performance.

A MAE of 34.97 was observed, indicating precise one-step forecasting.

Handled complex relationships with ease and provided accurate predictions.

Decision Tree:

With a high score of 0.9809, it demonstrated its effectiveness in capturing patterns.

A MAE of 52.23 was obtained, indicating accurate one-step forecasting.

Demonstrated proficiency in dealing with nonlinearities and making reliable predictions.

Neural network:

With a competitive score of 0.9828, it demonstrated strong predictive abilities.

The MAE was 56.95, indicating accurate one-step forecasting.

Adaptability in learning complex patterns was demonstrated, but error was slightly higher when compared to tree-based models.

Random Forest, XGBoost, Decision Tree, and Neural Network models all performed well in one-step forecasting.

These models' selection may be influenced by factors such as interpretability, computational efficiency, and specific dataset characteristics.

Machine learning models excel at capturing various patterns and relationships in the dataset of electric energy consumption.

Deep learning models

LSTM:

A relatively high score of 0.9846 was obtained, indicating good predictive capability.

The Mean Absolute Error (MAE) was 53.18, indicating accurate one-step forecasting.

Captured the intricate patterns of hourly energy consumption with promising results.

GRU:

Scored slightly lower than LSTM, at 0.9637, but still demonstrated strong forecasting performance.

The MAE was 74.26, indicating accurate predictions for one-step forecasting.

GRU provided competitive results with a focus on computational efficiency, despite not being as accurate as LSTM.

Multi-Step Forecasting:

When compared to single-step predictions, both LSTM and GRU performed poorly in multi-step forecasting.

LSTM-MS received a score of 0.9546, while GRU-MS received a score of 0.9637, indicating a decrease in accuracy for longer forecasting horizons.

Given the nature of the electric energy consumption dataset and the requirement for hourly forecasting, both the LSTM and GRU models proved appropriate.

The LSTM model performed well in terms of capturing sequential dependencies, whereas the GRU model provided a good balance of accuracy and computational efficiency.

The performance of multi-step forecasting highlighted the difficulty of maintaining high accuracy over long time horizons.

Final evaluation

	Score	MAE	MSE	MAPE
Random forest-MS	0.993577	30.915104	1902.739903	0.0092
Random forest	0.993429	31.229791	1946.568806	0.0093
XGBoost	0.992477	34.972437	2228.600566	0.0105
XGBoost-MS	0.992477	34.972437	2228.600566	0.0105
AR	0.992467	34.780692	2467.691861	0.0106
LSTM	0.984645	53.176068	4548.843284	0.0162
Neural Network-MS	0.983424	52.672893	4910.598910	0.0160
Neural Network	0.982784	56.950250	5099.921282	0.0176
Decision Tree-MS	0.980942	52.101315	5645.812762	0.0156
Decision Tree	0.980887	52.233880	5662.049843	0.0157
GRU-MS	0.963746	74.262190	10887.266041	0.0225
LSTM-MS	0.954625	85.208497	13626.203494	0.0261
GRU	0.940134	122.517702	17734.608080	0.0381
Naive Baseline Model	0.910382	119.747672	26548.353527	0.0374

Random Forest and XGBoost outperform in both single-step and multi-step forecasting.

In single-step forecasting, autoregressive modeling produces competitive results.

Deep learning models (LSTM, GRU) perform well but have higher errors in multi-step forecasting than in single-step forecasting.

The Naive Baseline Model serves as a comparison point, demonstrating the predictive power of the developed models.

Stepan Zelenka- 1230017

Rui Duarte - 1231441