Final Project for ECE 514
Time Series Forecasting for Energy Consumption in Greece

PANAGIOTIS TOLOUDIS

Agenda

- The dataset
- Arima
- LSTM
- Prophet
- Xgboost
- AutoML




Introduction

Time series forecasting plays a crucial role in machine learning as it addresses prediction problems involving a temporal component. However, time series forecasting in energy consumption presents unique challenges due to the unpredictable nature of energy demand and the multitude of factors influencing it. One such challenge is the seasonality of energy consumption, where usage patterns vary based on the time of day, day of the week, and time of year. Additionally, external factors like weather patterns, economic conditions, and technological advancements can further complicate energy consumption predictions. The rise of renewable energy sources adds complexity as their output depends on unpredictable weather conditions. Moreover, policy changes and regulations can significantly impact energy demand, making their prediction difficult. Considering these complexities, forecasting energy consumption accurately becomes crucial for effective energy management and decision-making.

The dataset

Information about the Dataset

The dataset used for the project is sourced from Open Power System Data, a free-of-charge data platform dedicated to electricity system researchers. This platform collects, verifies, processes, documents, and publishes publicly available but otherwise inconvenient-to-use data. It provides various data sets, including information on installed generation capacity by country/technology, individual power plants (both conventional and renewable), and time series data. The time series data includes electricity consumption, spot prices, and wind and solar generation, derived from weather models and measured sources. The platform ensures machine readability and easy integration by following the Data Package convention and publishing data primarily as CSV files with metadata in JSON format (Excel and SQLite files are also available). To ensure reproducibility, both the data and scripts used for processing are version-controlled and accessible through stable URLs. For this project, the dataset version 2020-10-06 was selected, which contains data for 32 European countries at different time intervals (15 min, 30 min, and 60 min). The 60 min interval dataset was chosen specifically to work with the data for Greece.

Prepare the dataset

REMOVE ALL THE INFORMATION TO NOT USE.

KEEP THE DATA FROM THE 1/1/2018-0:00 TO 10/1/2020-1:00

ARIMA

What is the Arima model?

The ARIMA model is a statistical tool that can be used to analyze and forecast time series data. ARIMA stands for Auto-Regressive Integrated Moving Average, which means that the model combines three components:

- Auto-regression (AR): The model uses past values of the variable to predict its future values.

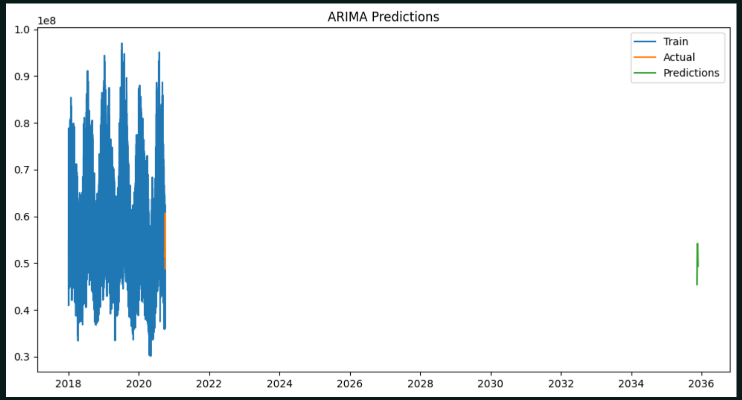
- Integration (I): The model applies differencing to the data to make it stationary, meaning that it has a constant mean and variance over time.

- Moving average (MA): The model incorporates the errors or noise from past predictions into the current prediction.

The ARIMA model is represented by three parameters: p, d, and q, which correspond to the order of the AR, I, and MA components respectively. For example, an ARIMA (1,1,1) model means that it has one AR term, one differencing term, and one MA term. The ARIMA model can be used to fit various types of time series data that have different patterns or trends.

The GR load actual entsoe transparency.

Predict	Actual
4.542353e+07	48829300
4.827820e+07	51866700
5.090549e+07	56699300
5.268140e+07	59324400
5.368239e+07	59963000
5.421035e+07	60683000
5.386518e+07	59905200
5.324477e+07	57309600
5.238635e+07	54244800
5.118879e+07	51964000
5.029158e+07	51317100
4.936184e+07	51973500



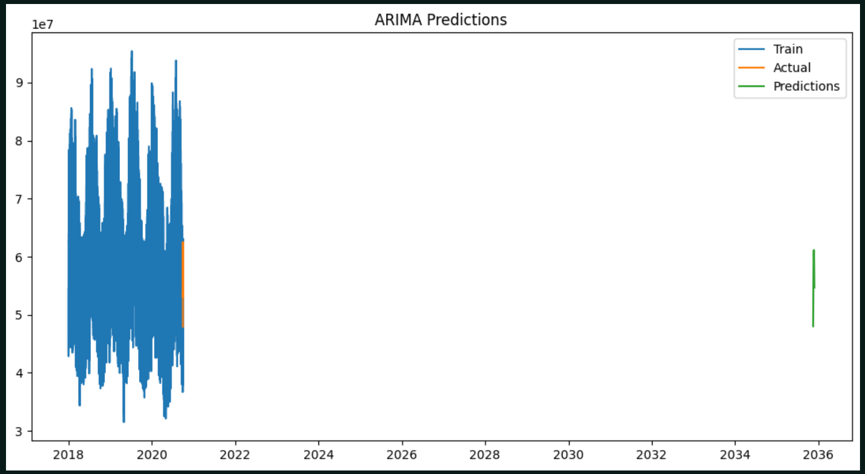
The GR load forecast entsoe transparency.

Predict

- 4.801549e+07
- 5.185250e+07
- 5.517169e+07
- 5.765611e+07
- 5.952495e+07
- 6.069373e+07
- 6.113299e+07
- 6.086623e+07
- 5.996439e+07
- 5.853738e+07
- 5.672351e+07
- 5.467779e+07

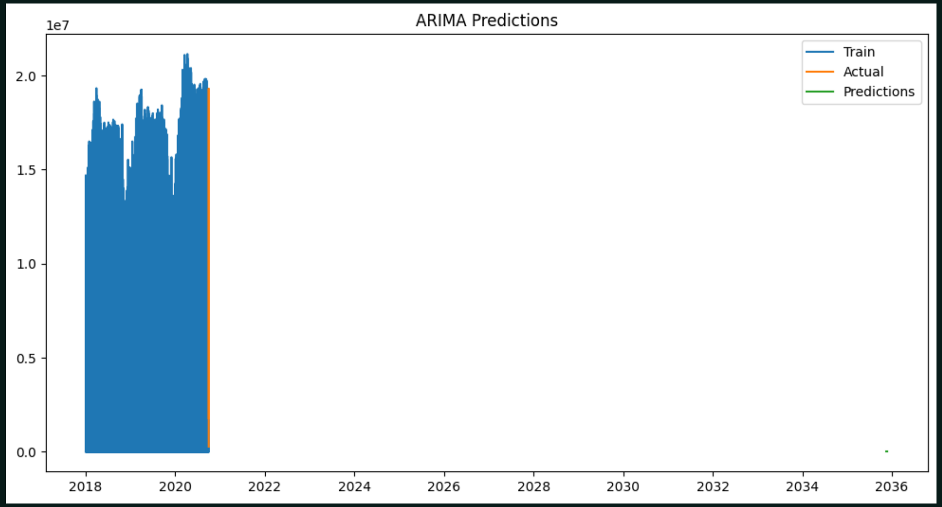
Actual

- 47929100
- 53457600
- 57901100
- 60439800
- 61694900
- 62520600
- 61888600
- 60031100
- 56648700
- 54124400
- 53025000
- 53551000



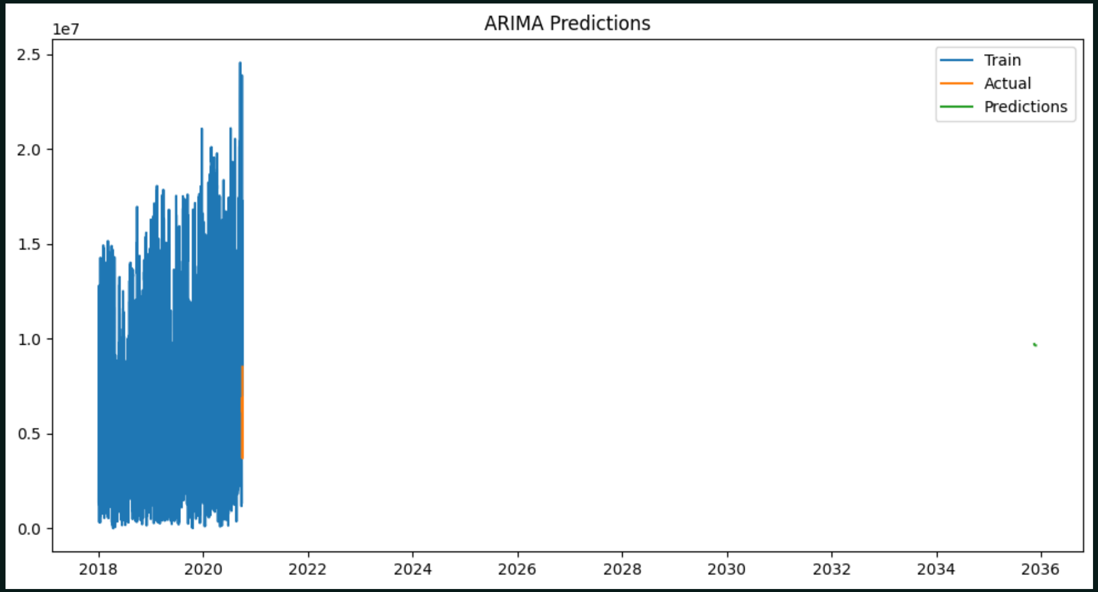
The GR solar generation actual.

Predict	Actual
-7.738259e-28	270000.0
-7.107229e-27	3980000.0
-1.281375e-26	9750000.0
-1.658644e-26	14410000.0
-1.851140e-26	17690000.0
-1.937587e-26	19310000.0
-1.908771e-26	19080000.0
-1.774846e-26	17830000.0
-1.581328e-26	15530000.0
-1.385117e-26	11690000.0
-1.222900e-26	900000.0
-1.112447e-26	1780000.0



The GR wind onshore generation actual.

Predict	Actual
9.715729e+06	8520000.0
9.690254e+06	5870000.0
9.674105e+06	4390000.0
9.663869e+06	3700000.0
9.657380e+06	3780000.0
9.653267e+06	4120000.0
9.650659e+06	4980000.0
9.649006e+06	5880000.0
9.647959e+06	6290000.0
9.647294e+06	6470000.0
9.646873e+06	6910000.0
9.646606e+06	6180000.0



LSTM

What is the LSTM model?

The LSTM model is a type of recurrent neural network (RNN) that can process sequential data, such as speech, text, or video. Unlike standard RNNs, LSTM has a special structure that allows it to remember information over long periods of time. An LSTM cell consists of four components: a cell state, an input gate, an output gate, and a forget gate. The cell state stores the long-term memory of the network, while the gates control how much information to keep or discard from the cell state and the inputs. LSTM can learn complex patterns and dependencies in sequential data and is widely used for tasks such as natural language processing, speech recognition, machine translation, and more.

The model results.

The evaluate of the model is 0.03626 in the test data.

PROPHET

What is the Prophet model?

The Prophet model is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. It works best with time series that have strong seasonal effects and several seasons of historical data. The Prophet model consists of four main components: a piecewise linear or logistic growth curve trend, a yearly seasonal component modeled using Fourier series, a weekly seasonal component modeled using dummy variables, and holiday effects modeled as dummy variables. The model is estimated using a Bayesian approach and is robust to missing data, shifts in the trend, and outliers. The Prophet model is designed to be simple for non-experts to use, yet flexible and powerful.

THE GR LOAD ACTUAL ENTSOE TRANSPARENCY.

```
                   ds         trend    yhat_lower    yhat_upper   trend_lower
0 2018-01-01 00:00:00  5.115150e+07  4.289080e+07  5.383322e+07  5.115150e+07
1 2018-01-01 01:00:00  5.115819e+07  3.897426e+07  5.118129e+07  5.115819e+07
2 2018-01-01 02:00:00  5.116489e+07  3.780183e+07  4.940105e+07  5.116489e+07
3 2018-01-01 03:00:00  5.117158e+07  3.630325e+07  4.830509e+07  5.117158e+07
4 2018-01-01 04:00:00  5.117827e+07  3.619573e+07  4.747289e+07  5.117827e+07

    trend_upper  additive_terms  additive_terms_lower  additive_terms_upper  \
0  5.115150e+07   -2.996816e+06         -2.996816e+06         -2.996816e+06
1  5.115819e+07   -5.902448e+06         -5.902448e+06         -5.902448e+06
2  5.116489e+07   -7.720492e+06         -7.720492e+06         -7.720492e+06
3  5.117158e+07   -8.874031e+06         -8.874031e+06         -8.874031e+06
4  5.117827e+07   -9.290518e+06         -9.290518e+06         -9.290518e+06

          daily  ...        weekly  weekly_lower  weekly_upper        yearly
0 -6.314450e+06  ... -3.336074e+06 -3.336074e+06 -3.336074e+06  6.653709e+06
1 -9.470504e+06  ... -3.106071e+06 -3.106071e+06 -3.106071e+06  6.674127e+06
2 -1.154859e+07  ... -2.866484e+06 -2.866484e+06 -2.866484e+06  6.694581e+06
3 -1.297007e+07  ... -2.619027e+06 -2.619027e+06 -2.619027e+06  6.715070e+06
4 -1.366065e+07  ... -2.365456e+06 -2.365456e+06 -2.365456e+06  6.735593e+06
```

THE GR LOAD FORECAST ENTSOE TRANSPARENCY.

```
                   ds         trend    yhat_lower    yhat_upper   trend_lower
0 2018-01-01 00:00:00  4.998724e+07  4.280840e+07  5.390222e+07  4.998724e+07
1 2018-01-01 01:00:00  4.999596e+07  3.897755e+07  5.060132e+07  4.999596e+07
2 2018-01-01 02:00:00  5.000469e+07  3.736401e+07  4.865905e+07  5.000469e+07
3 2018-01-01 03:00:00  5.001341e+07  3.630386e+07  4.834796e+07  5.001341e+07
4 2018-01-01 04:00:00  5.002213e+07  3.604450e+07  4.705833e+07  5.002213e+07

    trend_upper  additive_terms  additive_terms_lower  additive_terms_upper  \
0  4.998724e+07   -1.729728e+06         -1.729728e+06         -1.729728e+06
1  4.999596e+07   -4.843984e+06         -4.843984e+06         -4.843984e+06
2  5.000469e+07   -6.777526e+06         -6.777526e+06         -6.777526e+06
3  5.001341e+07   -7.955646e+06         -7.955646e+06         -7.955646e+06
4  5.002213e+07   -8.335098e+06         -8.335098e+06         -8.335098e+06

          daily  ...        weekly  weekly_lower  weekly_upper        yearly
0 -6.435641e+06  ... -3.141942e+06 -3.141942e+06 -3.141942e+06  7.847854e+06
1 -9.802059e+06  ... -2.909800e+06 -2.909800e+06 -2.909800e+06  7.867874e+06
2 -1.199596e+07  ... -2.669486e+06 -2.669486e+06 -2.669486e+06  7.887922e+06
3 -1.344094e+07  ... -2.422706e+06 -2.422706e+06 -2.422706e+06  7.907997e+06
4 -1.409200e+07  ... -2.171191e+06 -2.171191e+06 -2.171191e+06  7.928098e+06
```

THE GR SOLAR GENERATION ACTUAL.

```
                   ds         trend    yhat_lower    yhat_upper   trend_lower
0 2018-01-01 00:00:00  3.812957e+06 -5.327350e+06  3.778284e+05  3.812957e+06
1 2018-01-01 01:00:00  3.812985e+06 -5.193556e+06  3.874846e+05  3.812985e+06
2 2018-01-01 02:00:00  3.813012e+06 -5.114559e+06  5.456537e+05  3.813012e+06
3 2018-01-01 03:00:00  3.813040e+06 -4.996994e+06  1.014039e+06  3.813040e+06
4 2018-01-01 04:00:00  3.813067e+06 -4.986908e+06  6.374663e+05  3.813067e+06

    trend_upper  additive_terms  additive_terms_lower  additive_terms_upper  \
0  3.812957e+06   -6.235219e+06         -6.235219e+06         -6.235219e+06
1  3.812985e+06   -6.269811e+06         -6.269811e+06         -6.269811e+06
2  3.813012e+06   -6.070745e+06         -6.070745e+06         -6.070745e+06
3  3.813040e+06   -5.905784e+06         -5.905784e+06         -5.905784e+06
4  3.813067e+06   -6.056603e+06         -6.056603e+06         -6.056603e+06

          daily  ...        weekly  weekly_lower  weekly_upper        yearly
0 -4.399489e+06  ... -49700.446557 -49700.446557 -49700.446557 -1.786029e+06
1 -4.429838e+06  ... -54457.352429 -54457.352429 -54457.352429 -1.785516e+06
2 -4.226793e+06  ... -58949.774992 -58949.774992 -58949.774992 -1.785003e+06
3 -4.058160e+06  ... -63135.109986 -63135.109986 -63135.109986 -1.784489e+06
4 -4.205653e+06  ... -66973.909894 -66973.909894 -66973.909894 -1.783976e+06
```

THE GR WIND ONSHORE GENERATION ACTUAL.

```
                   ds         trend     yhat_lower    yhat_upper  \
0 2018-01-01 00:00:00  4.361669e+06 -230469.052197  1.077036e+07
1 2018-01-01 01:00:00  4.362051e+06 -470565.292179  1.023173e+07
2 2018-01-01 02:00:00  4.362432e+06   -5299.662021  1.066581e+07
3 2018-01-01 03:00:00  4.362814e+06 -328813.833307  1.073347e+07
4 2018-01-01 04:00:00  4.363196e+06 -120597.679962  1.104952e+07

    trend_lower   trend_upper  additive_terms  additive_terms_lower  \
0  4.361669e+06  4.361669e+06    7.685618e+05          7.685618e+05
1  4.362051e+06  4.362051e+06    7.956175e+05          7.956175e+05
2  4.362432e+06  4.362432e+06    8.543475e+05          8.543475e+05
3  4.362814e+06  4.362814e+06    9.488131e+05          9.488131e+05
4  4.363196e+06  4.363196e+06    1.059700e+06          1.059700e+06

   additive_terms_upper          daily  ...         weekly   weekly_lower
0          7.685618e+05 -131857.178234  ... -371711.009690 -371711.009690
1          7.956175e+05  -97225.030560  ... -378656.403487 -378656.403487
2          8.543475e+05  -33073.443969  ... -383438.959196 -383438.959196
3          9.488131e+05   64620.384804  ... -386020.132445 -386020.132445
4          1.059700e+06  176525.371365  ... -386383.242825 -386383.242825

    weekly_upper        yearly  yearly_lower  yearly_upper  \
0 -371711.009690  1.272130e+06  1.272130e+06  1.272130e+06
1 -378656.403487  1.271499e+06  1.271499e+06  1.271499e+06
2 -383438.959196  1.270860e+06  1.270860e+06  1.270860e+06
3 -386020.132445  1.270213e+06  1.270213e+06  1.270213e+06
4 -386383.242825  1.269558e+06  1.269558e+06  1.269558e+06
```

XGBOOST

What is the Xgboost model?

XGBoost is a machine learning framework that uses gradient boosting to create accurate and efficient predictive models. Gradient boosting is a technique that combines multiple weak learners (such as decision trees) into a strong learner by iteratively adding new learners that correct the errors of the previous ones. XGBoost stands for extreme Gradient Boosting, and it is designed to be scalable, fast, and robust to various types of data and problems.

THE GR LOAD ACTUAL ENTSOE TRANSPARENCY.

Prediction	Actual
66201820	66171500
59507024	59467600
41082864	41086100
48974040	48956900
66303360	66414500
63670044	63641800

The Mean Squared Error: 4179604559.307996

THE GR LOAD FORECAST ENTSOE TRANSPARENCY.

Prediction	Actual
51595084	51611600
40720100	40674700
76391744	76450000
60933364	60950300
55472148	55570000
63189272	63150000

The Mean Squared Error: 3971697205.2060227

THE GR SOLAR GENERATION ACTUAL.

Prediction	Actual
1.6871027e+02	0.0
1.6562167e+07	16590000.0
1.5606325e+07	15590000.0
1.6871027e+02	0.0
1.5405820e+07	15390000.0
1.6871027e+02	0.0

The Mean Squared Error: 178525867.4920639

THE GR WIND ONSHONE GENERATION ACTUAL.

Prediction	Actual
17332950	17390000.0
1454575.5	1460000.0
11891878	11890000.0
4290031.5	4290000.0
13257710	13260000.0
19425660	19420000.0

The Mean Squared Error: 227438983.1501955

AutoML

What is the AutoML model?

The AutoML model is a type of machine learning model that can automatically select the best architecture, hyperparameters, and data preprocessing steps for a given task. The AutoML model can save time and resources for developers and researchers who want to build and deploy machine learning solutions without requiring extensive expertise or manual tuning. The AutoML model can also discover novel and efficient architectures that outperform human-designed ones in some cases. The AutoML model is not a single algorithm, but a general framework that can be applied to different domains such as computer vision, natural language processing, and tabular data analysis.

The suggestions model.

H2O.AI

The h2o.ai model is a machine learning platform that enables users to build and deploy scalable and accurate predictive models. The platform offers a variety of tools and features, such as automated data preparation, feature engineering, model selection, explain ability, and deployment. The h2o.ai model can handle structured and unstructured data, as well as supervised and unsupervised learning tasks. The h2o.ai model is designed to be user-friendly, fast, and flexible, making it suitable for various domains and applications.

DATAROBOT

The model DataRobot is a software that automates the process of building and deploying machine learning models. It uses a variety of algorithms, mostly based on open source, to train and evaluate thousands of models on a given dataset. It also provides tools for data quality assessment, feature engineering, model interpretation, and MLOps. The model DataRobot can help users of different skill levels to leverage the power of artificial intelligence and make better decisions based on data.

GOOGLE CLOUD AUTOML

Google Cloud AutoML is a service that allows users to create custom machine learning models without coding. Users can upload their own data, choose a model type, and train the model using Google's infrastructure. The service also provides tools for evaluating, deploying, and managing the models. Google Cloud AutoML supports various types of data, such as images, text, tabular, and video.

The different between the models.

The H2O.ai and DataRobot offer more specialized features and capabilities for time series analysis compared to Google Cloud AutoML. H2O.ai provides the H2O AutoML framework with specific support for time series modeling, while DataRobot offers automated time series modeling as part of its comprehensive platform. On the other hand, Google Cloud AutoML provides a general-purpose machine learning platform that can be used for time series analysis by utilizing its broader set of tools and infrastructure.

I use the H20.ai model.

The result

REPORTED ON TRAIN DATA.

MSE: 3517418035650.899
RMSE: 1875478.0818902948
MAE: 1386162.5148232258
RMSLE: 0.031496644655704585
Mean Residual Deviance: 3517418035650.899
R^2: 0.973268898953974
Null degrees of freedom: 10039
Residual degrees of freedom: 10032
Null deviance: 1.3212865205156608e+18
Residual deviance: 3.5314877077935024e+16
AIC: 318553.3189175752

REPORTED ON CROSS-VALIDATION DATA.

MSE: 4453587215909.751
RMSE: 2110352.391405225
MAE: 1547279.8532015597
RMSLE: 0.03528657985361276
Mean Residual Deviance: 4453587215909.751
R^2: 0.9659232204704427
Null degrees of freedom: 19375
Residual degrees of freedom: 19368
Null deviance: 2.5325041424212966e+18
Residual deviance: 8.629270589546734e+16
AIC: 619325.493958776

REPORTED ON TEST DATA.

MSE: 4528498171569.509
RMSE: 2128026.8258575853
MAE: 1557603.0190972523
RMSLE: 0.03565571506703599
Mean Residual Deviance: 4528498171569.509
R^2: 0.9645925630870953
Null degrees of freedom: 4720
Residual degrees of freedom: 4713
Null deviance: 6.038496502126353e+17
Residual deviance: 2.1379039867979652e+16
AIC: 150992.22118511234

Summary

- Its excessive size (e+7) is the cause of the errors.
- This project would benefit from having location, weather, and user power habits in addition to the power dataset.
- XGBoost and LSTM are the models that fit this dataset the best.

Thank you

Panagiotis Toloudis

ptoloudis@uth.gr