

TIME SERIES ANALYSIS AND FORECASTING

Before anything else, it is important to review first what is time series, as well as time series analysis and forecasting.

Time Series pertains to the sequence of observations collected in constant time intervals be it daily, monthly, quarterly or yearly. **Time Series Analysis** involves developing models used to describe the observed time series and understand the "why" behind its dataset. This involves creating assumptions and interpretation about a given data. **Time Series Forecasting** makes use of the best fitting model essential to predicting the future observation based on complex processing current and previous data.

Machine learning proved to be the most effective in capturing the patterns in the sequence of both structured and unstructured data and its further analysis for accurate predictions.

Speaking of applying a suitable model for time series forecasting, it is important to understand the **components of the time series data**:

- ✓ **Trends** (to describe increasing or decreasing behavior of the time series frequently presented in linear modes).
- ✓ **Seasonality** (to highlight the repeating pattern or cycles of behavior over time)
- ✓ **Irregularity/Noise** (to regard the non-systematic aspect of time series deviating from the common model values)
- ✓ **Cyclicity** (to identify the repetitive changes in the time series and define their placement in the cycle).

MODELING TIME SERIES

Among the numerous methods aimed at achieving precision and minimizing errors and losses within time series forecasting, there are several classical and modern machine learning methods that proved their accuracy and computational relevance.

The most widely applied models for time series forecasting projects by categories are as follows:

Classical methods

Naïve model

In most cases, naïve models are applied as a random walk (with last observed value used as a unit for the next period forecast) and seasonal random walk (with a value from the same period of last observed time span used as a unit of the forecast).



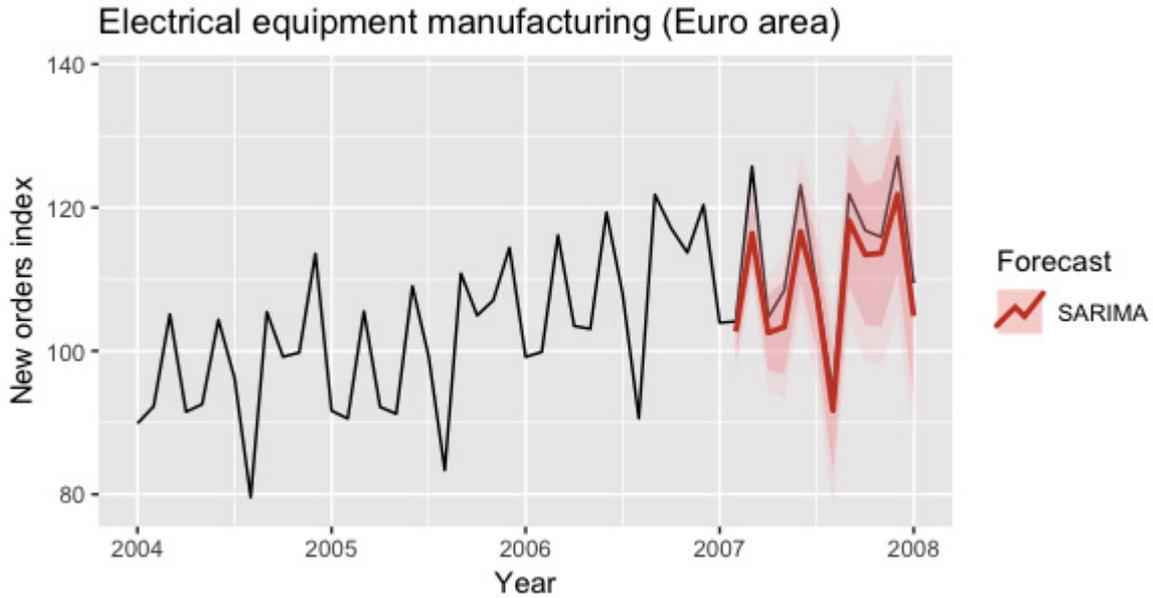
Exponential Smoothing Model

Forecasts are made on the basis of captured weighted averages and the according weights decreasing as the observer tracks back in time. A number of extensions of the simple exponential smoothing (SES) have been introduced to include the trend/damped trend and seasonality.

ARIMA/SARIMA

ARIMA stands for combination of Autoregressive (AR) and Moving Average (MA) approaches within building a composite model of the time series. ARIMA models include parameters to account for season and trend (for instance, dummy variables for weekdays and their distinguishing). In addition, they allow for the inclusion of autoregressive and moving average terms to handle the autocorrelation imbedded in the data.

SARIMA stands for Seasonal Autoregressive Integrated Moving Average: it widens application of the ARIMA by including a linear combination of seasonal past values and/or forecast errors.



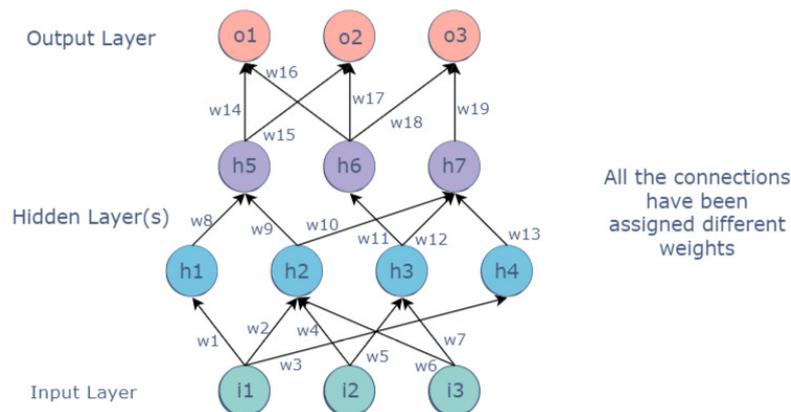
Linear Regression method

Linear regression is the simple statistical technique commonly used for predictive modeling. Breaking it down to basics, it comes to providing an equation of independent variables, on which our target variable is built upon.

Machine Learning methods

Multi-Layer Perceptron (MLP)

The notion of MLP model implies the triple structure of the initial layer of the network which takes in an input, a hidden layer of nodes, and an output layer used to make a



prediction.

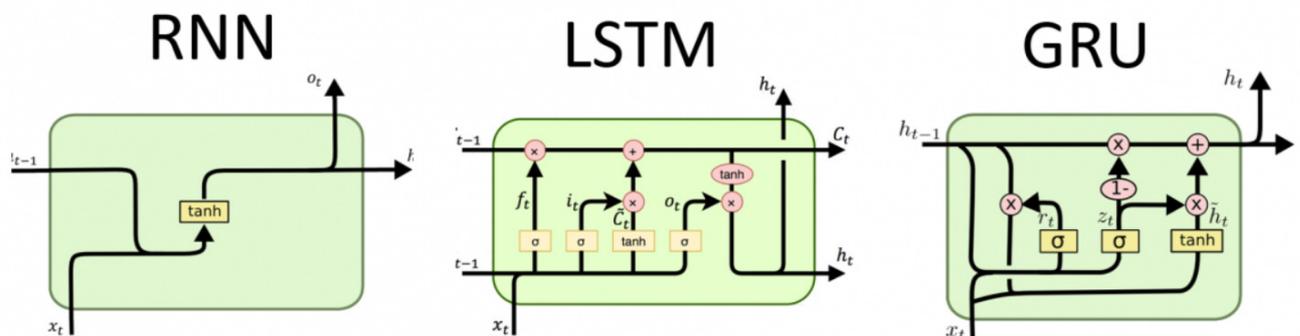
MLP is a feedforward neural network

Recurrent Neural Network (RNN)

RNNs are basically neural networks with memory that can be used for predicting time-dependent targets. Recurrent neural networks can memorize the previously captured state of an input to make a decision for the future time-step. Recently, lots of variations have been introduced to adapt Recurrent Networks to the variety of domains.

Long Short-Term Memory (LSTM)

LSTM cells (special RNN cells) were developed to find the solution to the issue with gradients by presenting several gates to help the model make a decision on what information to mark as significant and what information to ignore. GRU is another type of gated recurrent network.



Besides methods mentioned above, Convolutional Neural Network models, or CNNs for short, as well as decision tree-based models like Random Forest and Gradient Boosting variations (LightGBM, CatBoost, etc.) can be applied to time series forecasting.

Evaluating Model Accuracy

Speaking of methods, we should mention that visually identifying which model has the best accuracy is not always producing reliable results.

Among the methods of accuracy evaluation, the method of calculating the MAPE (Mean Absolute Percent Error) shows best results as a quicker way to compare the overall forecast accuracy of a proposed model.

The measures represent the percentage of average absolute error occurred. The overall idea of calculations for evaluation of model accuracy comes to the following proportion: the lower the MAPE, the better the forecast accuracy.

TIME SERIES FORECASTING PROCESS

To avoid any detrimental consequences and ensure the project success in terms of designing the predictive time model, the time series forecast project is implemented with taking the following steps.

1. Project goal definition

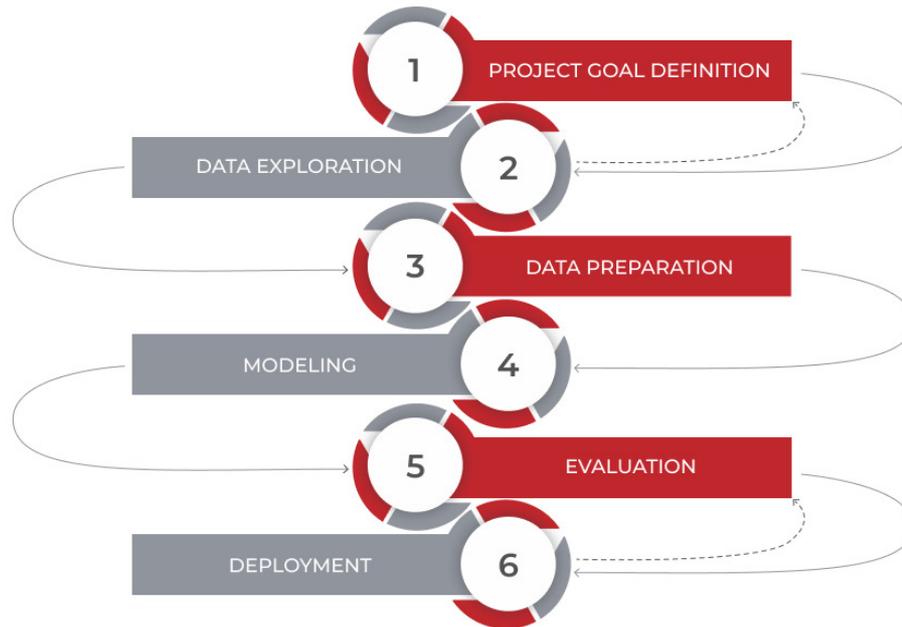
Prior to discussing the project in detail, make sure that you understand the subjectives. It means understanding the specifics of the business domain of forecast operation including terms and keys definitions as well as common business models pertaining to the particular domain. Thus, this stage implies defining the project specifics through extensive research within the area of knowledge.

2. Data gathering and exploration

Defining the basics leads to a clear view of scope of data you need to collect to facilitate the further discovery of data insights. With implementing techniques of building the plot graphs and visualisation charts, the domain knowledge reception reaches a level required for strategic data exploration and estimating hinges and trends for further evaluating the variations volume. It also helps specify the forecasting task and conduct preliminary exploratory analysis successfully.

3. Data preparation

At this stage, the development team performs cleaning data for relevant insights and further subtracting the variables of importance. The data preparation process for feature engineering is being launched. The key component of feature engineering is targeting the areas of knowledge of domain that are crucial for designing the new features in the existing dataset.



4. Applying time series forecasting method

On the basis of preliminary data preparation and exploratory analysis of a range of time series forecasting conducted at the previous stage, the team works with several models and chooses one on the criteria of relevance and projected accuracy of forecast. Fitting the model for project performance ensures the proper model development and consideration of variables essential within the forecasting process.

5. Evaluation/Validation and performance comparison

This step covers the optimization of the forecasting model parameters and achievements of high performance. By applying a cross-validation tuning method implying the data split, data scientists train forecasting models with different sets of hyper-parameters. Completing this stage requires applying performance scores estimation and diverse test datasets evaluating. It is important to take an out-of-sample approach to receive the adequate performance evaluation when processing the particular type of data.

6. Deployment

This stage includes the forecasting model integration into production. At this particular stage, we highly recommend setting a pipeline to aggregate new data to use for the next AI features. It helps in data preparation work when performing your future projects.

The forecast development requires agility in approach as it is the **iterative process**.

The iterative loops incorporate a series of exploration and visualization to get data. Once visualization has been performed, it might be required to take a step back to gather

additional data. The models are revised and updated as new data and new insights are made available.

Thus, this stage shifts focus to the ongoing development and refinement of one or more models until a relevant level of performance is achieved.

TIME SERIES FORECASTING PROJECT CHALLENGES

We would like to share the experience we acquired in performing time series forecasting projects and pinpoint the challenges that the development team might face.

Lack of data

The bigger the datasets are, the more training data the system can access, which leads to higher accuracy of predictions. However, there are limitations of using machine learning associated with a lack of historical or seasonality data for a target variable. Consequently, lack of data might result in overall decrease of forecasting precision.

Lack of domain knowledge

Apparently, without adequate domain knowledge the stage of feature engineering itself as the key component of ML implementation falls under high risk. In general, domain knowledge can help improve the quality of models in any project. To prevent the problem arising with the lack of domain knowledge, the expertise of the business niche specialists is required.