NOTES

**Datasets Description**

In Spain, the electricity system underwent a process of liberalisation between 1997

and 1998 [37], in which the tasks of transmission and distribution (regulated) and generation and retail (competitive) were separated as it has been presented above. The Spanish

electricity market is in charge of the OMIE, and it is organised in sequential sessions: the

day-ahead, intraday, and balancing sessions.

The price of electricity is different for each of the hours of each day; this price is

established by means of an auction. First, power generation companies and wholesale

retailers buyers are convened in the day-ahead spot market to send their quotes that include

energy price pairs within 24 h of the next day. The market operator (MO) is responsible for

clearing the market and providing a temporary energy plan for each bidder within 24 h of

the next day. By matching all the purchase and sale bids, the electricity market establishes

the price that the energy will have for each of the hours of the following day, i.e., the hourly

marginal price is obtained at the intersection of the supply and demand curves [38].

Every day, between 12:30 and 13:00 of the current day D, the price that the energy

will have in each of the hours of the following day D + 1, within the daily market, is made

public [39].

As it can be intuited, it is very important for the system operators to have an idea of

what the prices may be in the short term, as this will significantly influence the offers they

make, so a system capable of making predictions for the next day is an indispensable tool

for them [40].

The main objective of this study is to see how the choice of data can influence the

effectiveness of the prediction results. For each of the chosen data sets, three months have

been selected, with a total of 2208 time points (hours) for each dataset, with two of them

being used to set the parameters of the models (train/test step), and the remaining month

to validate them (validation step).

For each of the three data sets, a normalization of the data has been carried out, the

type of normalization is also a parameter to be chosen in the first stage, the possible values

are a minmax or z-score normalization. Attempts have been made to choose time periods

with some particularity. These are detailed below:

• **Fraud Period**: In 2019, the National Commission for Markets and Competition

(CNMC) fined Endesa and Naturgy for altering electricity prices between October

Appl. Sci. 2021, 11, 6097 12 of 19

2016 and January 2017 [41], the months chosen for this study. This period will help to

test how sensitive the models are to periods that may present tax irregularities. To

make the results more comparable, the following periods used will have the same

length, i.e., three months. Hourly electricity price in this period can be observed in

Figure 3 where the yellow line represents the mean price of all the period.

• **Normal period**: In order to have a reference period, a stage has been chosen in which

no anomalies were detected, unlike the two next data sets. This stage covers the

months from September to December 2019. Figure 4 represents the hourly electricity

price in this period, with the mean represented by a horizontal yellow line.

• **Quarantine Period**: The year 2020 marked a change in many daily habits due to the

global pandemic. One of the most notable changes was the increase in electricity consumption, as people spent more time at their homes due to quarantine requirements.

The three months from 15 March 2020 to 15 June 2020 have been chosen to represent

this stage. Hourly electricity price in this period can be observed in Figure 5 where

the yellow line represents the mean price of all the period.

5.1. Length of the test period

A common practice in EPF is to evaluate new methods on very short test periods. The typical approach is to evaluate the method on 4 weeks of data [18], [19], [22], [24], [25], [26], [29], [30], [41], [42], [49], [51], [97], [102], [103], [104], [105], [106], [107], [110], with each week representing one of the four seasons in the year. This is problematic for three reasons:

* Selecting four weeks can lead to cherry-picking the weeks where a given method excels, e.g. a method that performs bad with spikes could be evaluated in a week with fewer spikes, leading in turn to biased estimations of the forecasting accuracy. While this is an ethical issue that most researchers would avoid, establishing four week testing periods as the standard does facilitate malpractice and should be avoided.
* Assuming that the four weeks are randomly selected and no bias is introduced in the selection, it is still not possible to guarantee that these four weeks are representative of the price behavior over a whole year. Particularly, even within a given season, the price dynamics can change dramatically, e.g. during winter there are weeks with a lot of sun and wind but there are also weeks without them. Therefore, picking only a week per season rarely represents the average performance of a forecaster in a given dataset.
* There are situations in the electrical grid that do not occur very often but that can have a very large effect on electricity prices, e.g. when several power plants are under maintenance at the same time. Forecasting methods need to be evaluated under those conditions to ensure that they are also accurate under extreme events. By selecting four weeks most of these effects are neglected.

To avoid this problem, we recommend using a minimum of one year as a testing period. This ensures that forecasting methods are evaluated considering the complete set of effects that take place during the year. To guarantee that all researchers have access to this type of data, the open-access benchmark dataset that we propose contains data from several markets and employs a testing period of two years. In addition, the open-access benchmark can be directly accessed using the proposed epftoolbox library [58], [59].

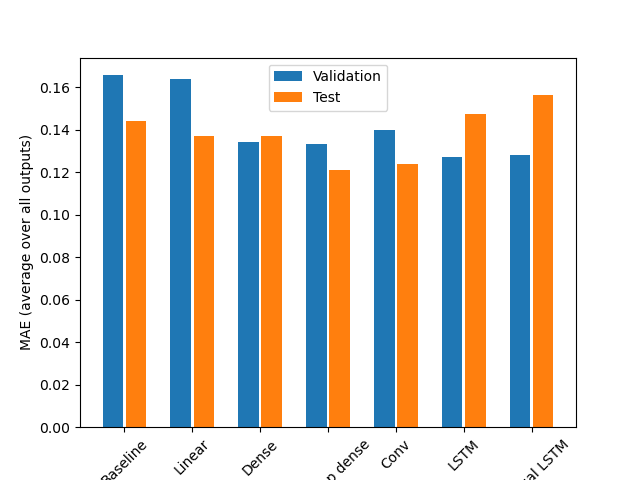
Tabla

Descripción generada automáticamente

You'll use a (70%, 20%, 10%) split for the training, validation, and test sets. Note the data is not being randomly shuffled before splitting. This is for two reasons:

* It ensures that chopping the data into windows of consecutive samples is still possible.
* It ensures that the validation/test results are more realistic, being evaluated on the data collected after the model was trained.

Single-step models



Texto

Descripción generada automáticamente

Multiple-Step Models

Gráfico, Gráfico de barras

Descripción generada automáticamente

Texto

Descripción generada automáticamente

This cheat sheet demonstrates 11 different classical time series forecasting methods; they are:

1. Autoregression (AR)
2. Moving Average (MA)
3. Autoregressive Moving Average (ARMA)
4. Autoregressive Integrated Moving Average (ARIMA)
5. Seasonal Autoregressive Integrated Moving-Average (SARIMA)
6. Seasonal Autoregressive Integrated Moving-Average with Exogenous Regressors (SARIMAX)
7. Vector Autoregression (VAR)
8. Vector Autoregression Moving-Average (VARMA)
9. Vector Autoregression Moving-Average with Exogenous Regressors (VARMAX)
10. Simple Exponential Smoothing (SES)
11. Holt Winter’s Exponential Smoothing (HWES)

**Vector Autoregression Moving-Average with Exogenous Regressors (VARMAX)**

The Vector Autoregression Moving-Average with Exogenous Regressors (VARMAX) is an extension of the VARMA model that also includes the modeling of exogenous variables. It is a multivariate version of the ARMAX method.

Exogenous variables are also called covariates and can be thought of as parallel input sequences that have observations at the same time steps as the original series. The primary series(es) are referred to as endogenous data to contrast it from the exogenous sequence(s). The observations for exogenous variables are included in the model directly at each time step and are not modeled in the same way as the primary endogenous sequence (e.g. as an AR, MA, etc. process).

The VARMAX method can also be used to model the subsumed models with exogenous variables, such as VARX and VMAX.

The method is suitable for multivariate time series without trend and seasonal components with exogenous variables.

<https://machinelearningmastery.com/time-series-forecasting-methods-in-python-cheat-sheet/>

## Autocorrelation Plots

We can plot the correlation coefficient for each lag variable.

This can very quickly give an idea of which lag variables may be good candidates for use in a predictive model and how the relationship between the observation and its historic values changes over time.

We could manually calculate the correlation values for each lag variable and plot the result. Thankfully, Pandas provides a built-in plot called the [autocorrelation\_plot()](https://pandas.pydata.org/pandas-docs/version/0.18.1/visualization.html" \l "autocorrelation-plot) function.

The plot provides the lag number along the x-axis and the correlation coefficient value between -1 and 1 on the y-axis. The plot also includes solid and dashed lines that indicate the 95% and 99% confidence interval for the correlation values. Correlation values above these lines are more significant than those below the line, providing a threshold or cutoff for selecting more relevant lag values.