



# Multi-Attribute Forecast of the Price in the Iberian Electricity Market

Gonalo Peres<sup>1</sup> , Antonio J. Tall3n-Ballesteros<sup>2</sup> , and Lu3s Cavique<sup>1</sup> 

<sup>1</sup> Universidade Aberta, Lisbon, Portugal

1800301@estudante.uab.pt, luis.cavique@uab.pt

<sup>2</sup> University of Huelva, Huelva, Spain

antonio.tallon.diesia@zimbra.uhu.es



**Abstract.** Electricity has been acquiring a more significant presence in our lives, and it is estimated that the future will be increasingly electric. Nowadays, we have access to enormous amounts of data that do not have much-added value if they cannot support decision-making or plan systems in advance and correctly. Forecasts are vital tools to support decision-making. We believe it is possible to resort to open data available on the Internet to make electricity price forecasts that - decision-makers can use in the sector. In this work, we study the multi-attribute hourly forecast of the electricity price in MIBEL (Iberian electricity market) for the 24 h of the following day, using open data. The realization of the multi-attribute predictions fell on the TIM ('Tangent Information Modeler') tool with AutoML ('Auto Machine Learning') capabilities. The TOPSIS ('technique for order of preference by similarity to ideal solution') decision support technique was used to analyze the results.

**Keywords:** Economic prediction · Forecasting · Iberian electricity market (MIBEL) · Auto machine learning · Multi-attribute decision

## 1 Introduction

The uncertainty associated with many temporal phenomena does not allow the exact knowledge of their behavior in the future, which leads us to have to make forecasts. Forecasts play a fundamental role in the planning, decision-making, and control process in any organization.

Electricity is considered an exceptional commodity because although there have been several advances regarding batteries' performance, electricity is primarily a non-storable commodity. In this sense, the price of electricity presents unique characteristics that make forecasting a challenging task.

This work aims to develop models that will allow the multi-attribute forecasting of the electricity price in the Iberian Energy Market (MIBEL) daily market for the 24 h of the following day (multi-step), using an autoML tool.

The remaining of the paper is organized as follows. Section 2 describes the data enrichment of the MIBEL. Section 3 presents the experimental planning. In Sect. 4, the AutoML tool is detailed. In Sect. 5, the computational results are discussed. Finally, the conclusions are drawn in Sect. 6.

## 2 Data Enrichment

Data enrichment enhances collected data with relevant context obtained from additional sources [3]. This section discusses some properties of the dataset to be used, presented in Table 1.

**Table 1.** Explanatory variables.

Variable	Explanatory	Description	Range and unit
V1	Chronological	Year	2010–2019
V2	Chronological	Month	1–12
V3	Chronological	Day	1–31
V4	Chronological	Hour	1–24
V5	Chronological	Day of the week	1–7 (1 is Monday and 7 is Sunday)
V6	Chronological	Holiday	0 or 1 (0 is not a holiday, 1 is a holiday)
V7	Prices	Hourly price	€/MWh
V8	Demand	Real demand	MW
V9	Demand	Forecast demand	MW
V10	Demand	Scheduled demand	MW
V11	Production	Wind	MW
V12	Production	Nuclear	MW
V13	Production	Coal	MW
V14	Production	Combined cycle	MW
V15	Production	Hydraulics	MW
V16	Production	International exchanges	MW
V17	Production	Solar	MW
V18	Weather	Earth temperature	K
V19	Weather	Earth or water temperature	K
V20	Weather	10 m u-component wind	$\text{ms}^{-1}$
V21	Weather	10 m v-component wind	$\text{ms}^{-1}$
V22	Weather	Solar radiation	$\text{Jm}^{-2}$
V23	Weather	Thermal radiation	$\text{Jm}^{-2}$
V24	Weather	Total cloud cover	0–1
V25	Weather	Total precipitation	mm
V26	Weather	Total sky radiation	$\text{J m}^{-2}$

For the development of the hourly price forecast models, we consider different types of explanatory variables, precisely:

1. Chronological data: Year, Month, Day, Hour, Day of the Week, Holiday<sup>1</sup> (in Spain).
2. Hourly electricity market prices (MIBEL), available from the market operator (OMIE) or the Portuguese TSO's Energy Market Information System.<sup>2</sup>
3. Demand data: Actual Demand, Forecast Demand, and Programmed Demand. These data come from the Spanish TSO, REE.<sup>3</sup>
4. Production or generation system data: Wind, Nuclear, Coal, Combined Cycle, Hydro, Solar, International Exchanges. These data were obtained from the Spanish TSO, the REE.<sup>4</sup> Since the production variables changed from 2015-04-30 to 2015-05-01, we consider that Solar encompasses (Solar Photovoltaic and Solar Thermal).
5. Weather or climate data: ERA-5, from the Climate Data Store<sup>5</sup> in Iberia. The Iberian Peninsula is considered the region delimited by the following geographic coordinates: North: 44°, West: -9.6°, South: 36°, East: 3.5°.

### 3 Experimental Planning

The models proposed allow performing multipoint forecasting of daily electricity prices in MIBEL. Such models may be beneficial for day-to-day market operations. To perform the forecasts, we will use TIM, an AutoML based on Information Geometry with a strong component of visual tool [2]. The experimental planning (Table 2) was inspired by Cláudio Monteiro's articles [5, 6].

Unlike in the above articles, where one-time intervals and one type of split in training-test were considered, in this paper, we consider three-time intervals and multiple splits in training-test, more specifically three: 70–30, 80–20, and 90–10. Apart from that, various metrics will be obtained: runtime (in seconds), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE).

Throughout this work, a total of 225 models will be developed, with variations related to the time intervals to be considered (75 models for each time interval), the percentages of the training and test samples (25 models for each train-test division), and the input variables (Table 2).

In this study, the following time intervals will be considered:

- Interval A. January 1, 2019 (00:00) to December 31, 2019 (23:00).
- Interval B. January 1, 2018 (00:00) to December 31, 2019 (23:00).
- Interval C. January 1, 2010 (00:00) to December 31, 2019 (23:00).

<sup>1</sup> The holidays in Spain were created based on this script: <https://github.com/goncaloperes/TIMESeries/blob/master/IsItHoliday/Spain/SpainBusinessCalendar.py>.

<sup>2</sup> <https://www.mercado.ren.pt/PT/Electr/InfoMercado/InfOp/MercOmel/Paginas/Precos.aspx>.

<sup>3</sup> <https://demanda.ree.es/visiona/peninsula/demanda/>.

<sup>4</sup> <https://demanda.ree.es/visiona/peninsula/demanda/tablas/2>.

<sup>5</sup> <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=form>.

A total of 225 ( $25 \times 3 \times 3$ ) forecast models used to predict the price of electricity in MIBEL for the 24 h following the last point in the series (multi-step forecast) were created, 75 models were considered for each time interval (25 for each training-test division).

**Table 2.** Models to be implemented and their input variables (M25  $\times$  V26).

Variable	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18	M19	M20	M21	M22	M23	M24	M25	
V1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
V2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
V3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
V4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
V5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
V6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
V7	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
V8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
V9	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
V10	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
V11	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
V12	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
V13	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
V14	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
V15	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
V16	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
V17	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
V18	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
V19	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
V20	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
V21	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
V22	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
V23	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
V24	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
V25	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
V26	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

## 4 Auto Machine Learning

The section begins by introducing the method used, TIM, an AutoML (automatic machine learning [9]) based on Information Geometry, a method for exploring the world of information through modern geometry [1], and then the experiment planning is described. Tangent Information Modeler, or TIM, is an AutoML solution designed for time-series prediction and anomaly detection developed by Tangent Works.

In the background, TIM constantly manages a few workers that can facilitate all possible requests. Scalability is an essential characteristic of this architecture, achieved through the computational distribution to TIM workers' units.

For this research, we will be using the Client TIM-Studio, a web application that provides an intuitive interface to a Machine Learning technology for data time-series, TIM.

TIM-Studio makes the experience as intuitive as possible. One can go with the default parameters or, tuning is also possible.

TIM-Studio version 2.0 makes the predictions based on the AutoML approach. TIM's AutoML allows one to build an entirely new model each time a forecast is needed, so there is no need to establish a particular data availability scenario. TIM analyzes the availability of the data as it is in the dataset at the time of the forecast and then builds a model based on this data and applies the model in a single pass through the data, delivering a forecast using all the available data from the forecast in just seconds to minutes [4].

To perform multi-step forecasting, namely to forecast the electricity price for the next 24 h, TIM finds a separate model for each hour in the forecast horizon,  $F(h)$ , as shown in Fig. 1. It is unnecessary to make contradictory optimizations because each model needs to be optimized for only one output. In this multi-step forecast approach, the outputs are not used to create other outputs, so the risk of error propagation is also eliminated.



Fig. 1. Multi-step forecast, where TIM finds a separate model  $F(h)$  for each hour.

## 5 Computational Results

In the computational results phase, some choices must be made, such as the dataset, the algorithm, the computational environment, the performance measures, and the interpretability of the solutions.

The dataset and the experimental modeling were detailed in the previous section. AutoML TIM-Studio provides the algorithm and the computational environment in a cloud environment. Next, the performance measure and the interpretability of the solutions are presented.

### 5.1 Performance Measures

By analyzing the computational results of the 225 models, we observe that the fastest model does not necessarily obtain a minor error (either MAE, or MAPE, or RMSE). The one with a smaller MAPE is not the fastest. We understand that taking advantage of more metrics can be beneficial for selecting the best models in the real world. Accordingly, to facilitate the decision-making process of the best models, we applied the multi-attribute decision method TOPSIS.

We have selected TOPSIS [7] (acronym for “Technique of Order Preference Similarity to the Ideal Solution”) to add as little subjectivity as possible in the decision process.

The subjectivity is associated with the weights of each attribute and with its assignment of cost or benefit.

The 225 models were ranked based on a multi-criteria TOPSIS analysis considering equal weight for each attribute in this study. TOPSIS was applied for each interval, and the solutions are shown in Table 3.

**Table 3.** TOPSIS computational results.

Interval	Training-test	Model	Run time (seconds)	MAE (€/MWh)	MAPE (%)	RMSE (€/MWh)
A	90–10	M6	7	5.26	62.63	7.70
B	90–10	M25	43	4.84	29.46	6.64
C	90–10	M23	170	3.96	15.52	5.42

Analyzing the results with the multi-attribute decision-making model performed at all intervals and the best models adopt the training-test 90–10.

Naturally, the best results (MAE, MAPE, RMSE) are presented for interval C, which includes ten years of data. The value of MAPE of 15.52% is competitive and reflects the hard time-series with 24 h forecast compared with similar works as in [5].

## 5.2 Interpretability of the Solution

TIM-Studio version 2.0, for each hour of the multi-step forecast, presents a weighted sum with an average of 20 rules. The expression for the forecast,  $F$ , at each hour,  $h$ , is given by the weighted,  $w$ , sum of  $i$  rules, is as follows:

$$F(h) = \sum_{i=1}^n w_i \cdot \text{rule}_i \quad \text{with} \quad \sum_{i=1}^n w_i = 1$$

Based on the sunburst diagram of the model A-M7 at 10:00 a.m shown in Fig. 2, the weighted rules can be exemplified:

- Rule1 = PRICE(t-24) & PRICE(t-11):  $w_1 = 20.12\%$   
Price with lag/delay/offset equal to 24 and with Price lag/delay/offset equal to 11.
- Rule2 = DoW(t-7) ≤ Fri & PRICE(t-14):  $w_2 = 17.2\%$   
Weekday (day of week) with lag/delay/offset equal to 7 and less than Friday, and Price with lag/delay/offset equal to 14.
- Rule3 = DoW(t-20) ≤ Sat:  $w_3 = 12.76\%$   
Weekday with lag/delay/offset equal to 20 and less than Saturday.
- Rule4 = DoW(t-7) ≤ Sat & PRICE(t-12):  $w_4 = 10.83\%$   
Weekday with lag/delay/offset equal to 7 and less than Saturday, and Price with lag/delay/offset equal to 12.

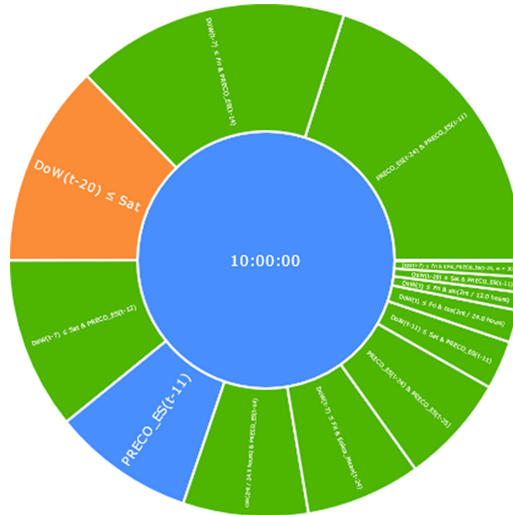


Fig. 2. A-M7 – 90–10 – sunburst diagram – 10:00

A vital issue in electricity price forecasting is the appropriate choice of explanatory variables. The typical approach has been to select predictors in an ad-hoc manner, sometimes using expert knowledge [8], as was the case in [5]. In the context of this work, a different approach was used, the result of mixing expert knowledge by selecting categories of explanatory variables and their respective variables and an automatic process of expanding variables that TIM provides. This automated process carried out by TIM provides the variables used, expansions performed, and their importance. So it is possible to forecast time-series, as is the case of MIBEL, with little specialized knowledge of how the market works. Furthermore, this functionality allows analysts working in the market to find other types of combinations of attributes that they can perform to analyze whether it produces improvements.

## 6 Conclusions

This work deals with the multi-attribute hourly forecast of the Electricity Price in MIBEL, namely in the Spanish part, for the 24 h of the following day, using open data. The data enrichment process adds an extensive set of input variables to the hourly price as demand forecasts, hourly power generation (wind, nuclear, coal, combined cycle, hydro, solar, international exchanges). Also, input variables include weather variables in the region and chronological data (date, year, month, day, hour, day of the week, holiday).

The accuracy of the model predictions was evaluated by the metrics MAE, MAPE, and RMSE, considering the execution time (in seconds). We believe that the evaluation's availability using several indicators becomes useful for a better comparison between papers in the area. The multi-attribute decision-making process was used, TOPSIS. The method allows ordering the models based on certain weights and criteria. The price forecasting models in this work, their performance in MIBEL, mainly in terms of

the errors and the runtime, the importance of their input variables, can be helpful for electricity market agents and other players in the electricity industry.

Another relevant point to mention is the issue of model interpretability. A set of rules are generated in the presented approach, and the most relevant variables are identified. As models are being used in increasingly sensitive applications, their interpretability and reliability are key issues.

Finally, we believe that we have achieved competitive computational results for a complex time-series problem using data enrichment and autoML.

**Acknowledgements.** This work has been partially subsidized by these projects: TIN2017-88209-C2-2-R (Spanish Inter-Ministerial Commission of Science and Technology), FEDER funds and US-1263341 (Junta de Andalucía).

## References

1. Amari, S.: Information Geometry and Its Applications. Springer, Japan (2016)
2. Cho, S.-B., Tallón-Ballesteros, A.J.: Visual tools to lecture data analytics and engineering. In: International Work-Conference on the Interplay Between Natural and Artificial Computation. Springer, Cham (2017)
3. Eric D.K., Joel, T.L.: Industrial Network Security, 2nd edn. Elsevier, Amsterdam (2015)
4. Miller, L.: Predictive Analytics for Time-series with InstantML For Dummies®, Tangent Works Special Edition, vol. 76 (2021). <https://www.tangent.works/wp-content/uploads/2020/11/Predictive-Analytics-For-Dummies-E-book-by-Tangent-Works-1.pdf>
5. Monteiro, C., Ramirez-Rosado, I., Fernandez-Jimenez, L., Conde, P.: Short-term price forecasting models based on artificial neural networks for intraday sessions in the iberian electricity market. *Energies* **9**, 721 (2016). <https://doi.org/10.3390/en9090721>
6. Monteiro, C., Ramirez-Rosado, I.J., Fernandez-Jimenez, L.A., Ribeiro, M.: New probabilistic price forecasting models: application to the Iberian electricity market. *Int. J. Electr. Power Energy Syst.* **103**, 483–496 (2018). <https://doi.org/10.1016/j.ijepes.2018.06.005>
7. Papathanasiou, J., Ploskas, N.: TOPSIS. In: Papathanasiou J., Ploskas N. (eds.) Multiple Criteria Decision Aid: Methods, Examples, and Python Implementations, pp. 1–30. Springer, Cham (2018). [https://doi.org/10.1007/978-3-319-91648-4\\_1](https://doi.org/10.1007/978-3-319-91648-4_1)
8. Uniejewski, B., Nowotarski, J., Weron, R.: Automated variable selection and shrinkage for day-ahead electricity price forecasting. *Energies* **9**, 621 (2016). <https://doi.org/10.3390/en9080621>
9. Yao, Q., et al.: Taking Human out of Learning Applications: A Survey on Automated Machine Learning. [arXiv:1810.13306](https://arxiv.org/abs/1810.13306) [cs, stat] (2019)