Case Study AI: Probability Distribution of Day-Ahead Electricity Prices

Van Thuy Cuc Dang, 22304742, thuy.dang@stud.th-deg.de
Ratan Bharadwaj Vedula, 22302468, ratan.vedula@stud.th-deg.de
Ujesh Khakhariya, 22308924, ujesh.khakhariya@stud.th-deg.de

Applied AI for Digital Production Management
Technische Hochschule Deggendorf, Cham, Deutschland.

***Abstract*—A non-parametric machine learning approach to probabilistic forecast hourly day-ahead electricity prices is shown in this case study. Emphasizing the importance of uncertainty, this method exceeds standard benchmarks by discovering optimal distributions from data while avoiding model assumptions' limitations. The study demonstrates the superiority of the proposed distributional neural network through an evaluation using Italy and Belgium power pricing, demonstrating its ability to deliver accurate and information-rich uncertainty forecasts critical for energy market decision-making.**

Introduction

Probabilistic forecasting involves estimating not only the expected outcome, but also the uncertainty associated with that prediction. Instead of providing a single point estimate, probabilistic forecasts offer a range of possible outcomes along with associated probabilities.

This approach is particularly useful when dealing with complex systems, dynamic environments, and situations where uncertainty plays a significant role.

Project Objectives

The objective is to introduce a new machine learning approach to probabilistic forecasting of electricity prices for the hourly day. The proposed method is non-parametric and selects the best distribution from all possible empirical distributions learned from the data.

This paper aims to demonstrate that this approach outperforms state-of-the-art benchmarks and provides data-rich forecasts that are not constrained by distributional or other model assumptions.

Background

Electricity is a key essential to modern life, and its prices are very difficult to predict due to its complex dynamics of supply and demand. Although researchers have primarily focused on point forecasts which calculate a single outcome based on the historical data, this has now become obsolete as the dynamics of the supply and demand have changed a lot.

Traditional approaches in energy economics have focused on point forecasts and have been constrained by distributional or other model assumptions. However, there is a recognized need to shift the focus towards probabilistic forecasting using big data and non-parametric methods that can accommodate non-Gaussian, heavy-tailed, and asymmetric data.

Focus is now on Probabilistic forecasting, which helps producers, retailers, and traders to assess uncertainty and improve optimal strategies for short-term operations, value-at-risk, hedging, and trading. This approach is designed to provide data-rich forecasts that are not constrained by distributional or other model assumptions, and aims to outperform existing state-of-the-art methods.

Probabilistic forecasting of hourly day-ahead electricity prices involves predicting the range of possible prices along with their associated probabilities. This type of forecasting is crucial for decision-making in energy markets, as it provides a more comprehensive understanding of the uncertainty surrounding future prices.

Methodology

The methodology used in this paper involved the development of a neural network with one input layer, two hidden layers with 32 neurons each, and one output layer.

Examine the hourly day-ahead electricity time series $y_{t,h}$ recorded over $t = 1, \ldots, T$ days and $h = 1, \ldots, 24$ hours. The primary objective is to closely estimate the conditional cumulative distribution function $F(y_{t,h}|I_{t-1})$ and to use it for a one-step ahead probabilistic forecast made at time $t - 1$. Information I_{t-1} encompasses past values of $y_{t,h}$ and potentially previous values of other observable exogenous variables x_t.

The primary objective is to estimate a set of conditional probabilities that align with the observed quantiles in each dataset.

$$\begin{aligned}\{F(q_h^{\alpha_1}, \ldots, F(q_h^{\alpha_p})\} \\ = \{Pr(y_{t,h} \leq q_h^{\alpha_1}|I_{t-1}), \ldots, \\ Pr(y_{t,h} \leq q_h^{\alpha_p}|I_{t-1})\}\end{aligned} \quad (1)$$

In the initial stage, we substitute a recognized link function with an unspecified general function g, approximated through a neural network. Following this, we explore a series of probabilities denoted by $0 < \alpha 1 < \alpha 2 < \ldots < \alpha p < 1$ comprising p evenly spaced levels that define the conditional distribution function utilizing a set of predictors Z_t that will be specified later and jointly modeling them as

$$\begin{aligned} &\{Pr(y_{t,h} \leq q_h^{\alpha_1}|z_{t-1}),\\ &\dots,\\ &Pr(y_{t,h} \leq q_h^{\alpha_p}|z_{t-1})\} = g_{W,b,h}(z_{t-1}), \end{aligned} \tag{2}$$

In this context $g_{w,b,h}$ refers to a neural network with multiple outputs and hidden layers L. We designate this network as the distributional neural network.

$$g_{W,b,h}(z_{t-1}) = g^{(L)}_{W^{(L)},b^{(L)}}, \dots, g^{(1)}_{W^{(1)},b^{(1)}}(z_{t-1}) \tag{3}$$

$$W = (W^{(1)}, \dots, W^{(L)}) \tag{4}$$

and

$$b = (b^{(1)}, \dots, b^{(L)}) \tag{5}$$

are weight matrices and bias vector. Any weight matrix $W^{(l)} \in R^{m*n}$ contain m neurons as n column vectors $W^{(l)} = [w_1^{(l)}, \dots, w_n^{(l)}]$, and $b^{(l)}$ are thresholds or activation levels.

The RELU activation function used is

$$g^{(l)}_{W^{(l)},b^{(l)}} := g_l(W^{(l)} z_{t-1} + b^{(l)}) = g_l(\sum_{i=1}^{m} W_i^{(l)} z_{t-1} + b_i^{(l)}) \tag{6}$$

,

$$g_l(u) = max\{u, 0\} \tag{7}$$

. $b(1), .., g(l)$are non-linear activation functions.

The quantile loss formula is,

$$(y, y') = \begin{cases} \alpha(y - y') & \text{if } y' < y \\ (1-\alpha)(y' - y) & \text{if } y' > y \end{cases} \tag{8}$$

where y is the target value and y' is the predicted value

Software/Package Requirement

Library	Version
pandas	2.1.2
pydantic	2.4.2
fastapi	0.104.0
uvicorn	0.23.2
numpy	1.26.3
pip	23.3.1
flask	3.0.1
tensorflow	2.10.0
psycopg2	2.9.9

TABLE I
Software Package Requirements Details

The project has been done by Pycharm Community Edition Software. It is written by Python language and all the needed library is stated in Table 1 Library Packages Requirement.

Project Flowchart

1. Data acquisition: Energy price demand dataset was obtained for Belgium and Italy for the year 2022.

2. Data cleaning and labelling: Data was verified to identify any missing values and if there are any missing values identified, they are replaced by the median of the dataset for numerical missing values and by the mode of the dataset for categorical data.

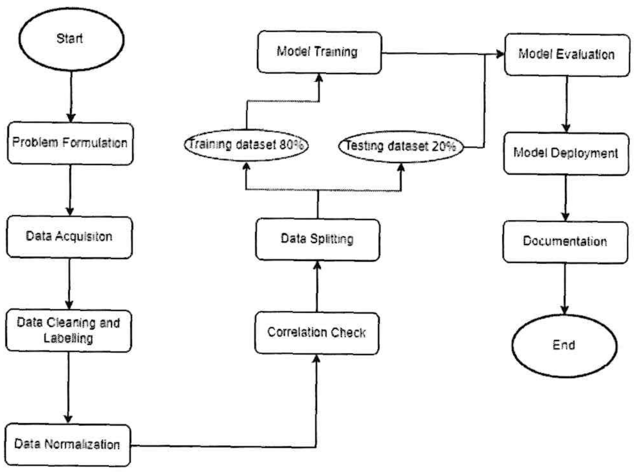

Fig. 1. Case Study Flow Chart

3. Data Normalization: The price data in Fig.2 is normalized by using the formula $(X - mean)/std$. The Month data is then normalized using the Sine function $\sin(month * 2\pi/12)$.

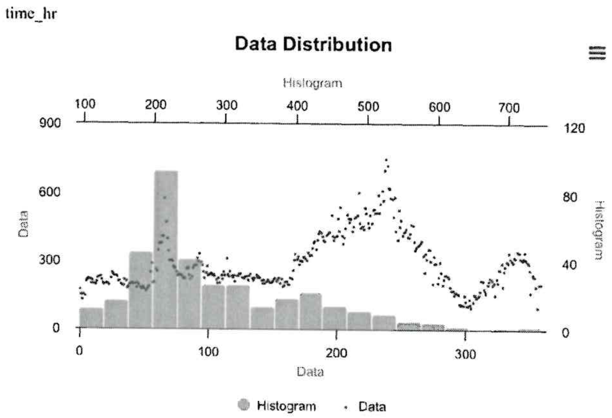

Fig. 2. The figure shows the distribution of hourly electricity prices in Belgium for 2022. The distribution is right-skewed, meaning there are more days with lower prices.

4. Correlation check: The correlation checks in Fig.3 between the attributes have been performed and the attributes having the correlation greater than 0.5 are selected.

5. Data splitting : 80% of the overall data set has been used to train the model and the remaining 20% data has been used to test the performance of the model.

6. Model training : The Neural network has been applied to train the model and to find the desired quantile values

7. Model evaluation: The model has then been evaluated to observe the forecasted hourly electricity prices and the Mean square error and the Root mean square error value were calculated to evaluate the performance of the model.

8. Model deployment : A final evaluation of the model has been performed and then the model is deployed into the Tencent cloud.

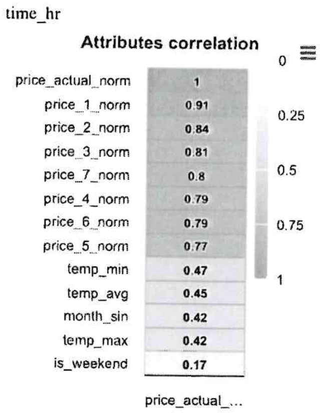

Fig. 3. Graph depicting correlations between hourly electricity prices in Belgium and various factors, highlighting strongest correlation with the price from the previous hour and the lowest correlation with the day of the week.

9. Model documentation: The final model, methodology and the results have been documented for further analysis and usage.

USER-INTERFACE

Energy price forecast User interface has been created with two drop-downs Country selection and the hour selection.

• Once the user inputs the Country and the Hour of the day, the UI displays the Input insight, Output insight.

• The input insights have 3 charts which display the Energy price at that selected hour of the country along with the data distribution and the attributes correlation. All these charts can be downloaded by the user if required.

• The UI also displays the Output insights which contain the charts describing the 24 hours price prediction of the upcoming day with the percentile range from 10-90. Another chart will display the Actual and prediction range of the at the selected hour for the upcoming months. A third chart displays the median actual price and the predicted price.

• The MSE value and the RMSE Value for the prediction comparing the error between the actual and the median values is also displayed on the UI.

• A table containing all the predicted price values of the whole upcoming day along with their percentile range is also displayed in the UI.

STEP BY STEP INSTRUCTIONS

The User interface can be accessed in two ways:

- Users get in by the URL: http://43.157.81.43/
- Users scan the QR code in the Fig.4 Application QR Code

After accessing the User Interface, the home page is displayed and the users can follow through the mentioned steps to gain insights of energy price prediction:

1) Select the desired country from the drop-down and the hour of the prediction
2) Click Submit Button
3) All the charts with the input and output information insight will be shown

Fig. 4. Graphical user interface for case study

Users could apply the model with their data by clicking "Go to My Data" link in the home page. After that, the following steps should be taken:

1) Click "Choose File" button to select the desired data file
2) Specify the Date column and the Value column
3) Click "Prediction" button to apply the model into the user's data and predict one-day-ahead value
4) All the charts with the input and output information insight will be shown

RESULTS

The mean squared error (MSE) and root mean square error (RMSE) of the suggested technique are both better than those of the most recent benchmarks. The findings demonstrate that the suggested strategy can very confidently and accurately forecast the hourly electricity price distribution.

MSE Value: 21.13

RMSE Value: 32.43

Fig. 5. This image shows the Root Mean Square Error (RMSE) and Mean Square Error (MSE) values for the predicted hourly electricity prices in Belgium. The RMSE value is 21.13 and the MSE value is 32.43

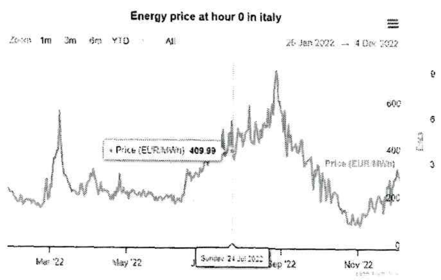

Fig. 6.

MSE depends on the observed value of the actual dataset, the higher the Actual value of the dataset, the higher is the

MSE value. The mean value of our dataset is around 400 and MSE of 21 for this dataset means that in average, the predicted value has error around 5% compare to actual prices and this can be justified because the prices are dynamically changing every hour of the day.

An easy-to-use method for exploring the model's results is the user interface. To view the expected costs for the next day, the user can choose a country and an hour of the day. Additionally, the user has access to both the current and anticipated costs for the following months. The median actual price and the forecasted price are likewise visible to the user. On the user interface, the prediction's MSE and RMSE values are also visible in Fig.5.

time_hr

dt	time_hr	place	price_actual	q1	q2	q3	q4	q5	q6	q7	q8	q9
2022-12-27	0	italy	197.35	178.18	181.44	184.38	187.12	191.82	201.72	203.55	209.10	220.88
2022-12-27	1	italy	91.00	137.74	144.14	146.41	146.86	157.57	162.16	180.68	189.83	206.09
2022-12-27	2	italy	80.02	91.99	94.10	120.31	165.22	182.93	197.24	203.21	207.47	245.37
2022-12-27	3	italy	62.09	112.51	147.98	154.53	167.10	184.13	194.45	196.68	200.90	220.70
2022-12-27	4	italy	63.20	116.56	121.06	186.07	199.68	238.00	241.05	249.49	285.28	294.29
2022-12-27	5	italy	85.00	78.51	110.99	144.77	151.02	153.01	153.85	160.47	177.05	227.52
2022-12-27	6	italy	205.81	148.92	169.50	171.49	183.63	186.94	201.97	202.26	231.03	256.92
2022-12-27	7	italy	225.57	190.91	200.89	210.21	213.38	214.13	221.27	227.00	240.65	246.23
2022-12-27	8	italy	240.73	213.58	218.24	220.73	221.95	222.32	228.38	233.83	234.26	243.71
2022-12-27	9	italy	231.26	194.17	204.91	211.78	222.12	222.77	226.07	239.60	239.83	247.48
2022-12-27	10	italy	227.47	186.39	190.32	196.07	200.38	200.82	201.17	213.57	220.97	225.72
2022-12-27	11	italy	225.57	171.46	196.84	198.06	204.06	213.81	223.19	223.24	226.90	233.00
2022-12-27	12	italy	200.00	151.43	151.84	185.31	202.69	207.71	214.44	214.63	224.51	249.15
2022-12-27	13	italy	200.20	170.18	179.60	180.16	200.31	205.49	207.17	215.43	248.46	252.61
2022-12-27	14	italy	215.07	175.49	195.20	197.76	200.22	214.78	217.98	222.50	236.57	261.41
2022-12-27	15	italy	240.45	187.02	198.32	212.19	217.03	223.78	229.51	243.87	247.67	262.94
2022-12-27	16	italy	320.00	189.38	199.45	207.06	207.36	236.82	247.65	259.42	272.26	282.98
2022-12-27	17	italy	350.00	220.73	223.04	241.43	243.50	254.38	255.45	274.13	277.20	285.75
2022-12-27	18	italy	325.00	237.40	240.75	247.92	252.22	253.05	256.00	263.95	265.88	269.57
2022-12-27	19	italy	330.00	225.98	234.91	241.40	253.29	262.76	275.84	278.84	293.41	310.31
2022-12-27	20	italy	280.00	217.89	217.91	222.53	231.09	238.78	244.13	246.06	274.76	285.76
2022-12-27	21	italy	225.57	188.52	191.31	198.91	205.28	208.24	212.69	222.46	233.59	238.47
2022-12-27	22	italy	205.95	184.06	184.86	184.91	190.31	191.74	195.68	198.81	208.22	208.86
2022-12-27	23	italy	200.20	133.44	156.46	158.42	159.94	165.97	166.76	174.02	177.07	179.31

Fig. 7. This image shows the hourly price of electricity in Italy on 27th December 2022. The prices are shown in Euros per megawatt-hour. The highest price is 147.55 Euros per megawatt-hour, and the lowest price is 43.91 Euros per megawatt-hour. The average price is 88.45 Euros per megawatt-hour.

The data in Fig.7 shows that the hourly electricity prices in Italy for December 27, 2022 ranged from 43.91 Euros per megawatt-hour to 147.55 Euros per megawatt-hour. The median price was 80.02 Euros per megawatt-hour. The highest price was 147.55 Euros per megawatt-hour, and the lowest price was 43.91 Euros per megawatt-hour.

The data also shows that the price of electricity in Italy tend to be higher in the morning and evening hours, and lower during the rest of the day. This is likely because demand for electricity is typically higher in the morning and evening hours, when people are waking up and going to bed.

This is a line graph Fig.8 that displays the 24-hour price prediction for Belgium on December 27th, 2022. The darkness of the line indicates the probability of the price being within the range of the line. The blue line represents the median price, while the light blue lines represent the 10th and 90th percentiles. The orange line represents the actual price for the previous day.

This graph is useful for understanding the range of possible electricity prices for the upcoming day. It can be used to make informed decisions about electricity consumption.

The graph in Fig.9 shows that the price of electricity in Italy is typically highest in the winter months and lowest in the summer months. This is because demand for electricity is typically higher in the winter months, when people use more heating.

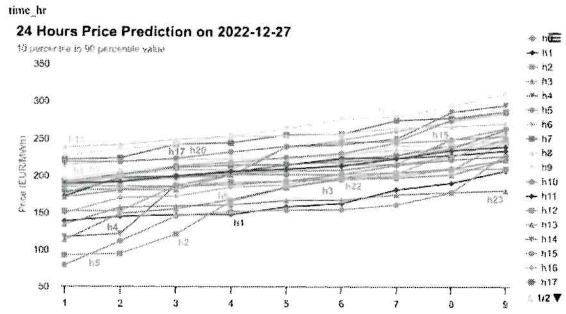

Fig. 8. The image shows the 24 hours price prediction of the upcoming day with the percentile range from 10-90 value for Italy on December 27, 2022.

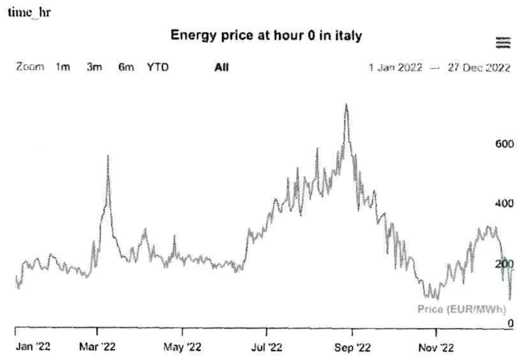

Fig. 9. The image shows a graph of the energy price at hour 0 in Italy. The price is shown in Euros per megawatt-hour. The highest price is 600 Euros per megawatt-hour, and the lowest price is 200 Euros per megawatt-hour. The average price is 400 Euros per megawatt-hour.

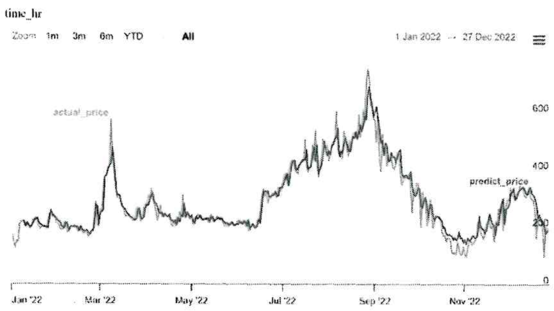

Fig. 10. The figure shows the distribution of actual hourly electricity prices in Belgium for 2022. The distribution is skewed to the right, meaning there are more days with lower prices than days with higher prices.

CONCLUSION/SUMMARY

A novel and potential efficient approach for probabilistic hourly electricity price forecasting is the one that has been suggested. With a high level of confidence, the approach is able to estimate the price distribution effectively. An simple to operate method for analyzing the model's results are the user interface.

REFERENCES

[1] Jozef Barunik, & Lubos Hanus. Learning Probability Distributions of Day-Ahead Electricity prices. October 10, 2023.
[2] Dataset of the Energy prices from Kaggle https://www.kaggle.com/datasets/henriupton/electricity-dayahead-prices-entsoe/data

AUTHOR CONTRIBUTIONS

Student name	Contribution	Signature with Date
Van Thuy Cuc Dang	Contributes towards defining Project Objective, Data Modeling, Back-end and Front-end (User Interface) Programming and Deployment.	
Ratan Bharadwaj Vedula	Contributes towards defining the Project Objective, Project Methodology, Data Finding, Project Flow Chart, Data Transformation, and Report Generation	
Ujesh Khakhariya	Contributes towards defining Project Objective, Data Finding, Data Cleaning, calculating the results, and providing conclusions to the project and formatting the Project report.	

TABLE II
CONTRIBUTIONS OF TEAM MEMBERS