Research Assignment for Exchange Students

Forecasting of electricity prices on the Spanish electricity market using machine learning tools

Author: Joanna Kaleta

UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

1. Electricity prices forecasting

Electricity price forecasting is a branch of energy forecasting which focuses on predicting the spot and forward prices in wholesale electricity markets. Over the last 15 years electricity price forecasts have been considered one of the most relevant factors for companies during decision-making process at the corporate level.

In the early 1990s power sectors in Europe, North America and Australia started being reshaped. Until then power sectors were controlled by the governments and highly monopolised. Resulting from the process of deregulation and the introduction of competitive electricity markets electricity is now traded under market rules using spot and derivative contracts. However, regarding to the special nature of electricity as a commodity, there are additional laws, that electricity market is subjected to. [1] [2]

In contrast to other goods, electricity is economically non-storable. In order to maintain power system stable, a constant balance between demand and supply is required. Factors like weather condition(temperature, wind, rain, etc.), intensity of human - business and everyday - activities during the day which shape the peak and valley hours, day of the week (working days vs. weekends, holidays, etc.), actual fuel cost affect the electricity consumption. [3] The example of daily electric power demand curve in Spain is exhibited in figure 1.1.

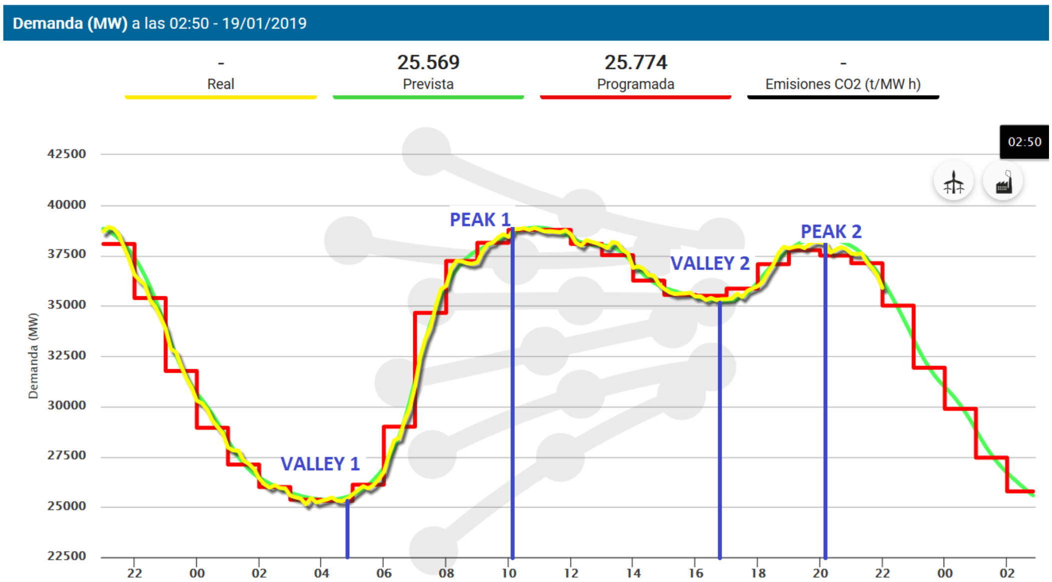

Fig. 1.1 Electric power demand in Spain on 19th January, 2019 [4]

The unique characteristics of electricity as a commodity results in daily, weekly and often annual seasonality. What is more, the price spikes are sudden, short-lived and mostly unpredictable. In order to create economically the most beneficial bidding strategy (minimizing risk and maximizing profits) or production / consumption schedule of a company it should be adjusted to the electricity price forecasts prepared from a few hours to a few months ahead. [5]

Since the competitive electricity markets were introduced, a variety of methodologies for prices prediction have been developed, for example [1] [6]:

Multi-agent models - simulate the behaviour of a system of heterogeneous agents (generating units, companies) interacting with each other, and build the price process by matching the demand and supply on the market. [6]

Fundamental models - try to capture the basic physical and economic relationships which are present in the production and trading of electricity. [7]

Reduced-form models - characterize the statistical properties of electricity prices over time. Their main intention is not to provide accurate hourly price forecasts, but rather to replicate the main characteristics of daily electricity prices. [2] [5] [6]

Statistical models - forecast the current price by using a mathematical combination of the previous prices and/or previous or current values of external factors, typically consumption and production figures, or weather variables. The two most important categories are: *additive* - the predicted price is the sum of a number of components and *multiplicative* models - the predicted price is the product of a number of factors. [1] [5]

Computational intelligence models - combine elements of learning, evolution and fuzziness to create approaches that are capable of adapting to complex dynamic systems, and may be regarded as "intelligent" in this sense. The major strength is the ability to handle complexity and non-linearity. [1] [8] [9]

Hybrid models - combining techniques from the groups listed above. [1]

Electricity price forecasts can be classified into 3 groups depending on the forecasting horizons, which is shown in the table 1.1. [1]

Table 1.1 Classification of electricity price forecasts

Forecasting horizon	Period ahead	Application
Short-term	from a few minutes up to a few days	day-to-day market operations
Medium-term	from a few days to a few months	risk management and derivatives pricing, balance sheet calculations
Long-term	months, quarters, years	investment profitability analysis and planning

In the process of peak occurrences or spot price variability prediction one of the most impactful variables is the reserve margin, also called surplus generation. Reserve margin is the relation between the available capacity (generation, supply) and the demand (load) at a given moment in time. [3]

2. Spanish Electricity Market

Day-ahead and intraday spot electricity markets in Spain and Portugal and also billing and settlements for Energy purchased and sold in these markets are managed by the OMIE company (Operador do Mercado Ibérico de Energia).

OMIE is regulated by an international agreement between Spain and Portugal. Half of OMIE's stock is owned by the Spanish company OMEL, with the other half held by the Portuguese company OMIP SGPS, S.A. OMIE runs the spot electricity market in the Iberian Peninsula under the Spanish jurisdiction. [10] [11]

The daily market

In the daily market electricity prices are set on a daily basis at 12 noon, for the twenty-four hours of the day ahead. Crossing point of supply and demand for electricity assigns the price and volume of energy for each hour according to the marginal pricing model based on the algorithm EUPHEMIA (acronym of Pan-European Hybrid Electricity Market Integration Algorithm). This algorithm is an efficient tool for price-setting and solving electricity market coupling problems and it is accepted in all European markets.

Agents from both Spain and Portugal can trade on the Spanish electricity market. Their purchase and sale bids are accepted according to their economic merit order, until the interconnection between Spain and Portugal is fully occupied. The price for the specific hour is equal for Spanish and Portuguese market as long as the capacity of the interconnection permits the flow of the electricity traded. If the interconnection is fully occupied, EUPHEMIA algorithm is applied separately for the coupling markets, which results in unequal prices on those markets.

Although the free trade is economically the most efficient solution, electricity trading is subjected to an additional rule – the whole process needs to be physically feasible. The institution responsible for validation of the electricity trading results is System Operator. Its role is to ensure that the market results do not exceed the technical limitations of the system. Such analysis conducted by the System Operator in order to make the daily programme executable, can affect around 4-5% of the Energy traded. [10] [12]

The intraday market

Apart from trading on the daily market, agents can also purchase and sell electricity on the intraday market according to the different rules. There are six trading sessions based on auctions such as those described for the daily market, where the volume of energy and each hourly price are determined by the point where supply and demand curves meet. Those trading sessions take place some hours earlier than real time. On the intraday markets buying and selling agents are allowed to readjust their purchase offers and selling bids up to four hours ahead of real time. The purpose of the intraday market is to respond, through the presentation of electricity power sale and purchase bids by market agents, to adjustments made to the Final Viable Daily Schedule. [10]

3. Methodologies used in the assignment

The purpose of the assignment was to test and compare the results of different machine learning methods for electricity price prediction. The input data set contains hourly electricity prices in Spanish electricity market in spot market during the period 01.01.2016 – 07.04.2017 (463 days). Electricity prices for selected hour during each day were forecasted based on prices for the previous day and the name of the weekday. In the assignment three following machine learning methodologies were used: K-nearest neighbours, Support Vector Machine and Artificial Neural Networks. For estimating continuous variables regression algorithms were used and therefore only those were elaborated in sections below.

K-nearest neighbours

The K-Nearest Neighbours algorithm is widely used for identifying patterns and regularities in data, so called pattern recognition. The algorithm may be applied for both classification and regression problems. [13]

The k- nearest neighbour method is based on selection a predefined number of training samples in the closest distance to the new point in the feature space and then predict the output value from these. Neighbours-based regression can be used in cases where the data labels are continuous rather than discrete variables. In the KNN regression the label for the query point is calculated as the average of the labels of its k closest neighbours.

The number of samples can be a user-defined constant (k-nearest neighbour learning), or vary based on the local density of points. The distance can be any metric measure, however standard Euclidean distance represented in the equation (3.1) is the most common choice. [14] [15]

$$dist(A,B) = \sqrt{\frac{\sum_{i=1}^{m}(x_i - y_i)^2}{m}} \tag{3.1}$$

Where $A = (x_1, x_2, \ldots, x_m)$ and $B = (y_1, y_2, \ldots, y_m)$ – feature vectors, m - dimensionality of the feature space

The basic nearest neighbours regression uses uniform weights point in the local neighbourhood contributes uniformly to the calculation of a query point label. Under some circumstances, it can be advantageous to weight points such that nearby points contribute more to the regression than faraway points. [15]

Support Vector Regression

Support-vector machines are supervised learning models with associated learning algorithms that analyse data used for classification and regression analysis.

In order to make the paragraph more understandable the linear function will be the example. However the general methodology can be applied for all other type of regression (non-linear, polynomial).

The general idea of SVM is to construct a hyperplane or set of hyperplanes in a high- or infinite-dimensional space and to define support vectors. For linear case hyper plane can be defined as line used to predict the continuous value (regression case) or target value (classification case). In SVM there are two boundary lines - lines other than Hyper Plane - which create a margin. Data points which are in the closest distance to the boundary are called support vectors. Kernel is a function used for mapping lower dimension data into higher dimension data. The main goal of SVR is to minimize error and individualize the hyperplane so it maximizes the margin. The simplest SVR case is presented in the picture below. [16] [17]

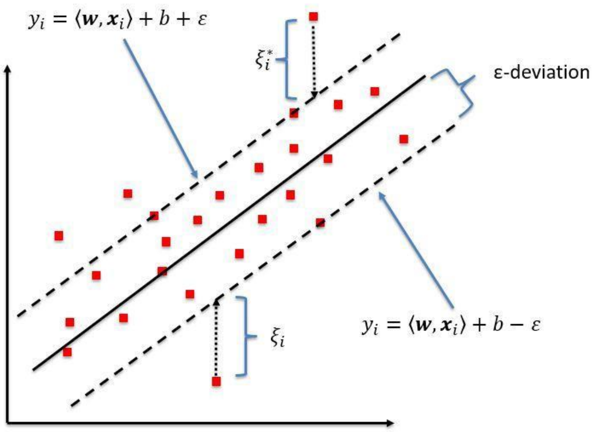

Fig. 3.1 Linear Support Vector Regression scheme [17]

There are two possible approaches to acceptable errors.

In first approach the goal of ε-SV regression is to find a function that has at most ε deviation from the actually obtained targets y_i for all the training data, which is provided in (3.3) condition. Any error larger than ε is not acceptable.

At the same time the function should be as flat as possible, which means the coefficient w is as small as possible – (3.2) condition.

This case can be described as a convex optimization problem:

$$minimize\ \frac{1}{2}\|w^2\| \tag{3.2}$$

$$subject\ to\ \begin{cases} y_i - \langle w, x_i \rangle - b \le \varepsilon \\ \langle w, x_i \rangle + b - y_i \le \varepsilon \end{cases} \tag{3.3}$$

Where x_i - training sample, y_i – target value of x_i, $<w, x_i> +b$ - prediction for that sample, ε - deviation threshold.

In the second approach some errors may be acceptable. The conditions (3.2) and (3.3) have to be changed as it is presented below:

$$minimize\ \frac{1}{2}\|w^2\| + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) \tag{3.4}$$

$$subject\ to\ \begin{cases} y_i - \langle w, x_i\rangle - b \le \varepsilon + \xi_i \\ \langle w, x_i\rangle + b - y_i \le \varepsilon + \xi_i^* \\ \xi_{i,}\xi_i^* \ge 0 \end{cases} \tag{3.5}$$

Where *C*- positive constant that determines the trade-off between the flatness of f and the amount up to which deviations larger than ε are tolerated, ξ– loss function described by the relations below. [16]

$$|\xi| = \begin{cases} 0\ if\ |\xi| \le \varepsilon \\ |\xi| - \varepsilon\ otherwise \end{cases} \tag{3.6}$$

Artificial Neural Networks

Artificial neural networks are machine learning tools inspired by animal brains. The neural network is rather a structure, not an algorithm, within which the input data is processed on different complexity levels. Neural network structure consists of input, output and mostly but not necessarily, hidden layers (Fig. 3.2). [18] [19] [15]

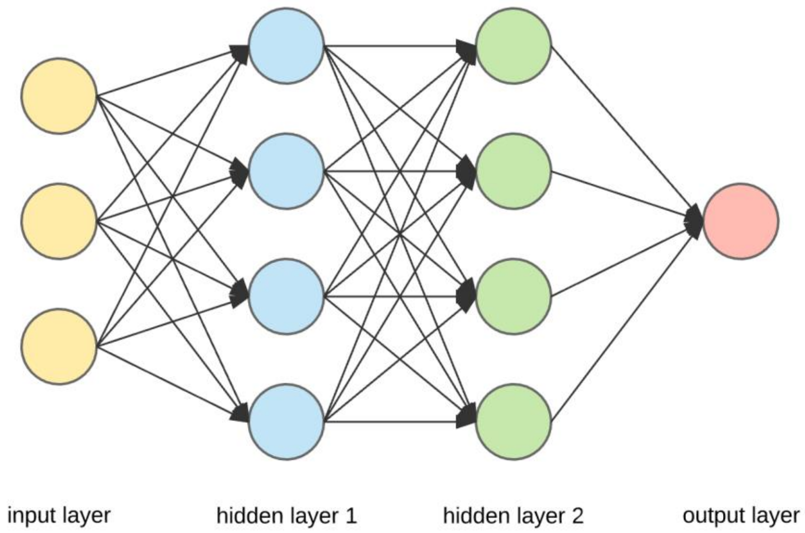

Fig. 3.2 Artificial neural network scheme [20]

The basic unit of computation in a neural network is the perceptron - a simple model of a biological neuron, often called a node. Based on inputs received from other nodes or external source the output is computed. Each input has an associated weight (w), which is corresponding to its relative importance to other inputs. Weighted sum of the inputs is calculated and an activation function ƒ is applied. If the obtained ƒ value exceeds a certain threshold it returns a signal, otherwise remains "silent". The perceptron scheme is shown in the figure 3.3. [19]

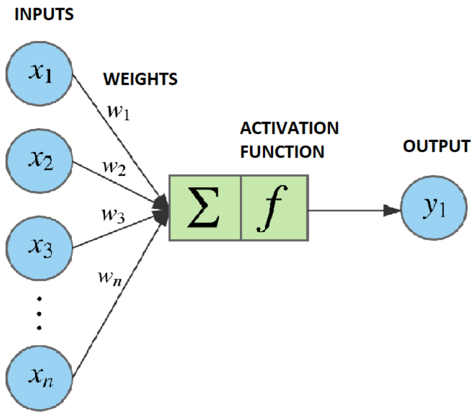

Fig. 3.3 Perceptron scheme [20]

Types of the most popular activate functions are presented in the figure 3.4.

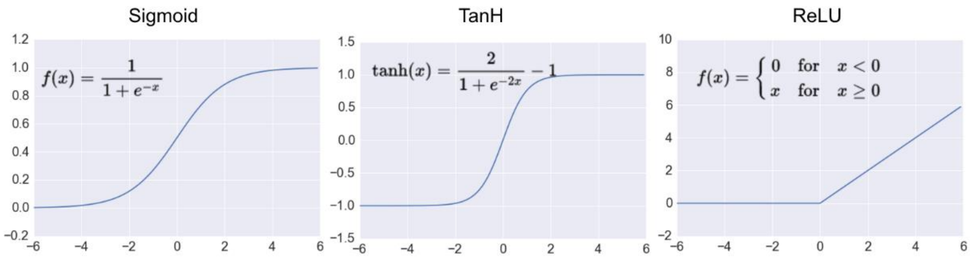

Fig. 3.4 Examples of activation function [21]

There are different types of learning, such as supervised or unsupervised learning or reinforcement learning, in which the network learns for itself by trying to maximize its score. In case of supervised learning weights of the perceptrons are adjusted in the training process when the predicted values are compared with the target values. [20]

4. Results

In the assignment 72 models were obtained: K-nearest neighbours, Support Vector Regression and Artificial Neural Networks for each hour. The input data set was at the same time training set and test set. In order to develop aforementioned models Python module Scikit-learn, which contains *sklearn.neighbors.KNeighbors, Regressorsklearn.svm.SVR* and *sklearn.neural_network.MLPRegressor* classes, was used. For each method an attempt to set hyperparameters (parameters whose values are set before the learning process begins) which provide the best models was made.

Parameters selection is presented in the table below. The table contains hyperparameters which values are different from default or are significant for the model. The complete list of hyperparameters is available in the sci-kit learn documentation.

Table 4.1 Values of selected hyperparameters

Hyperparameter	Value
KNN	
number of nearest neighbours k	3
weighs	uniform
metric	minkowski*
SVR	
Kernel	linear
Epsilon (specifies the epsilon-tube within which no penalty is associated in the training loss function with points predicted within a distance epsilon from the actual value)	0,1
Penalty parameter C of the error term	1
ANN	
Hidden layer sizes	1 hidden layer with 90 perceptrons
Activation function for the hidden layer	Identity - no-op activation, returns f(x) = x
Alpha – L2 penalty (regularization term) parameter	0,0001
Max_iter - maximum number of iterations; the solver iterates until convergence or this number of iterations	500

*minkowsky distance is calculated according to the formula: $\sum_{i=1}^{n}(|x_i - y_i|^p)^{1/p}$, where *p* -power parameter for the Minkowski metric. For *p*=2 (default value applied in the model) it is equivalent to the standard Euclidean metric. [15]

The complete analysis contains plots and tables with calculations such as correlation coefficient and mean errors, which allow to compare obtained results easily. In the paper example model for the 22nd hour is analysed. In the plot 4.1 general comparison of electricity prices obtained from different models and original spot market prices is shown.

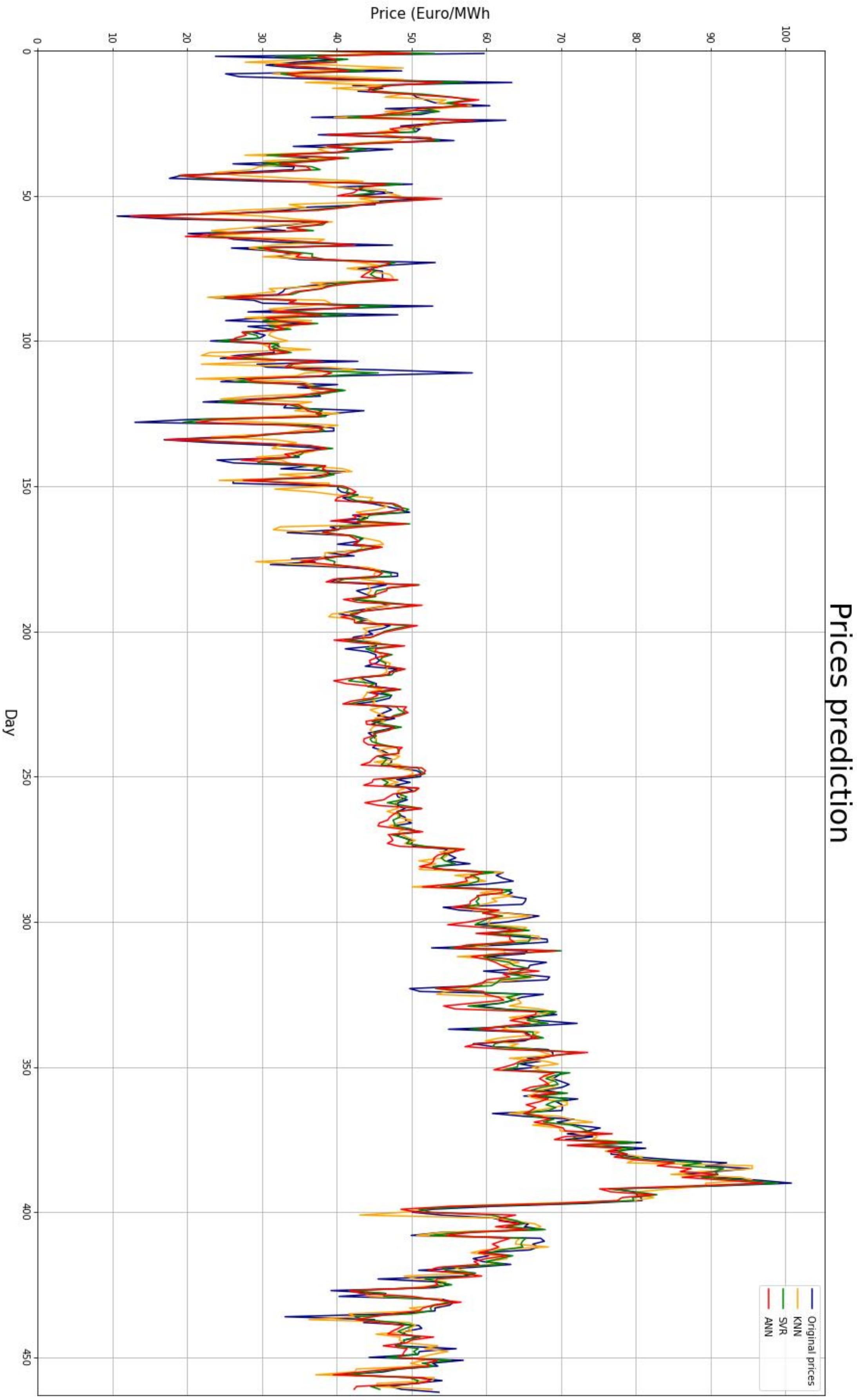

Fig. 4.1 General comparison original and predicted electricity prices

For each model scatter plot presenting relation between predicted and original prices was generated (Fig 4.2, 4.3, 4.4).

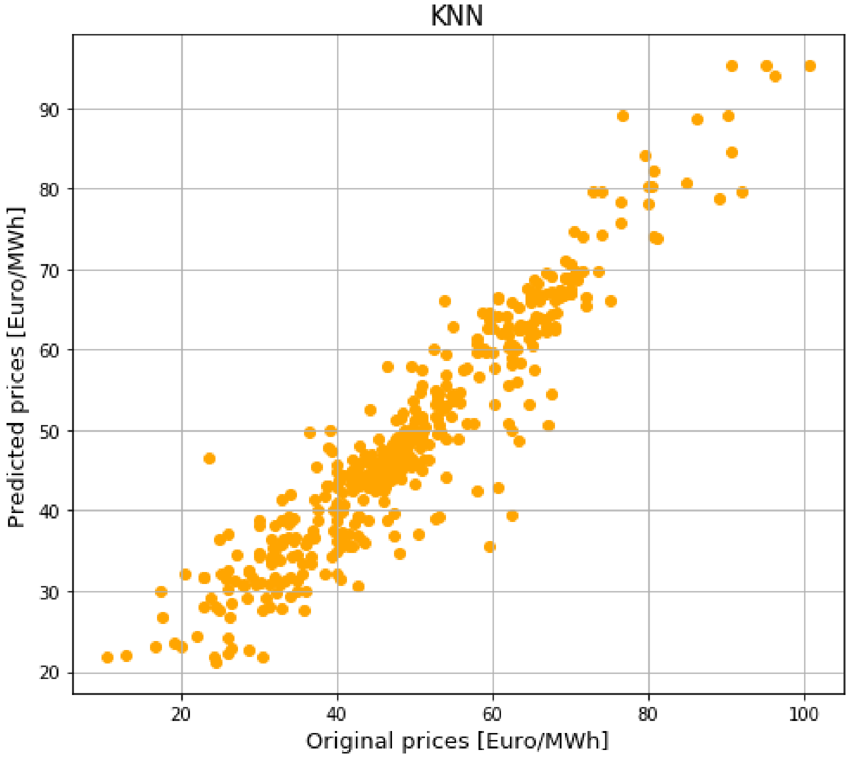

Fig 4.2 Scatter plot for KNN model

Fig 4.3 Scatter plot for SVR model

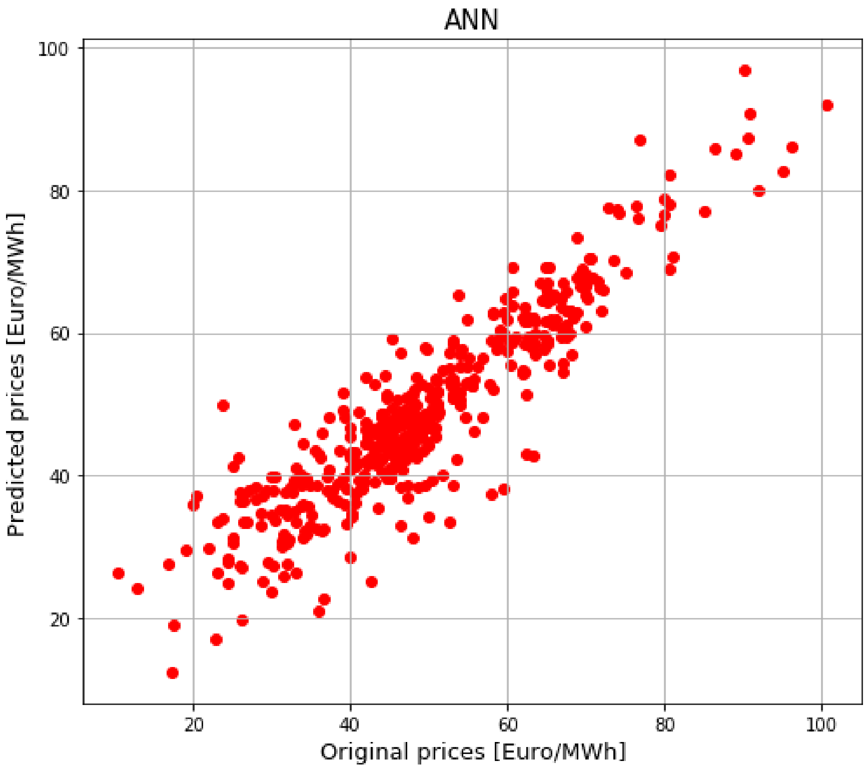

Fig 4.4 Scatter plot for ANN model

Based on above presented plots it is quite difficult to unambiguously state which model returns values corresponding at most with the original prices. Therefore for each model Pearson correlation coefficient was calculated according to the formula (4.1):

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{4.1}$$

Results for the analysed case are presented in the table 4.2. It is easy to observe that all the models exhibit accidentally the same correlation value which is equal to 0,97.

Table 4.2 Correlation coefficients for analysed methods

	Y	KNN	SVR	ANN
Y	1,00	0,97	0,97	0,97
KNN	0,97	1,00	0,98	0,98
SVM	0,97	0,98	1,00	1,00
ANN	0,97	0,98	1,00	1,00

Values of coefficient of determination R^2 (4.2), which is one of the basic tools used for evaluation how well observed outcomes are replicated by the model were also accessed.

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_t - \bar{y})^2}{\sum_{i=1}^{n}(y_t - \bar{y})^2} \tag{4.2}$$

Where $\hat{y}_t$-theoretical value of the predicted variable Y, y_t-the actual value of variable Y at time t, $\bar{y}$-the arithmetic mean of the empirical values of the predicted variable.

Their values are equal to 0,94 for KNN, 0,94 for SVR and 0,93 for ANN. Although R^2 coefficient values are very high, it is not difficult to achieve such results when working on time series data. Therefore high values of R^2 do not necessarily mean that models are good.

Another way to evaluate accuracy of each model were its MAPE (mean average percentage error) and RSMPE (root square mean percentage error) values, which were calculated according to the formulas (4.2) and (4.3):

$$MAPE = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \tag{4.3}$$

.

$$RMSPE = 100\%\sqrt{\frac{\sum_{t=1}^{n}\left(\frac{A_t - F_t}{A_t}\right)^2}{n}} \tag{4.4}$$

Where A_t – actual value, F_t – forecasted value.

Results are presented in the table 4.3. The lowest MAPE and RSMPE values were obtained for k-nearest neighbors. The largest MAPE value was observed for artificial neural network, RSMPE value – for support vector regression.

Table 4.3 MAPE and RSMPE values

	MAPE [%]	RSMPE [%]
KNN	8,43	14,11
SVR	10,79	18,74
ANN	10,96	18,10

To access the stability and quality of a model k-fold cross validation was conducted. In *k*-fold cross-validation, the original data set was divided into *k* (10) subsets. One subset was used as a validation (test) subset, the rest subsets serve as training data. The process was repeated 10 times, each time with different subset used as test data.

The goal of cross-validation is to evaluate the behaviour of the model while independent dataset is used as an input. Cross-validation results can indicate problems with the model like overfitting (overtraining; model corresponds too exact with only particular data) or selection bias (selection of data for analysis in such a way that proper randomization is not achieved). For each iteration MAPE and RSMPE were calculated and obtained values are presented in charts 4.5 and 4.6.

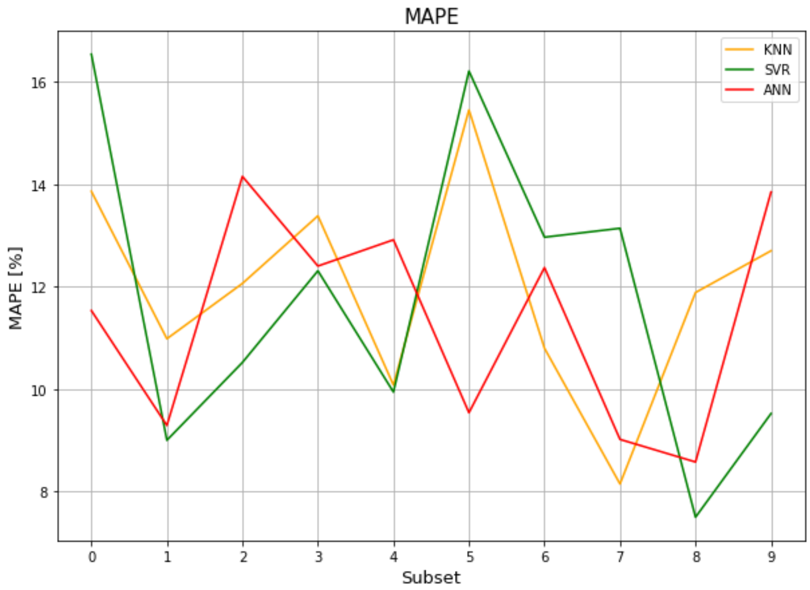

Fig. 4.5 MAPE values for cross-validation subsets

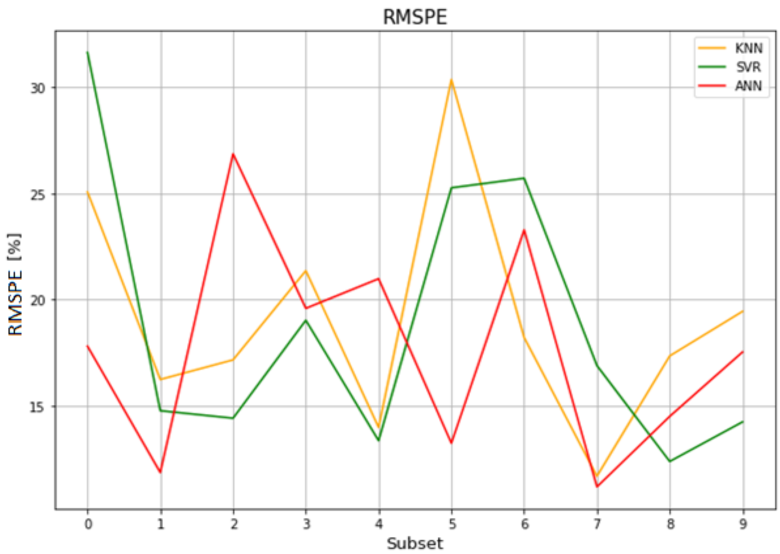

Fig. 4.6 RSMPE values for cross-validation subsets

It can be easily observed, that quality of developed models could be more satisfying. Maximal and minimal errors values are very similar for all analysed tools. Minimal MAPE values are equal to around 8%, maximal – around 14-16%. Minimal RSMPE values are equal to around 13% and maximal – around 27-30%. Also variations of MAPE and RMSPE values for 10 analysed subsets are equally strong and significant for all methods.

Finally for all three methods for 24 hours Pearson correlation coefficients, R^2 coefficients, MAPE and RSMPE were accessed to test if for each hour models maintain the same quality. All results are exhibited in charts 4.7-4.11.

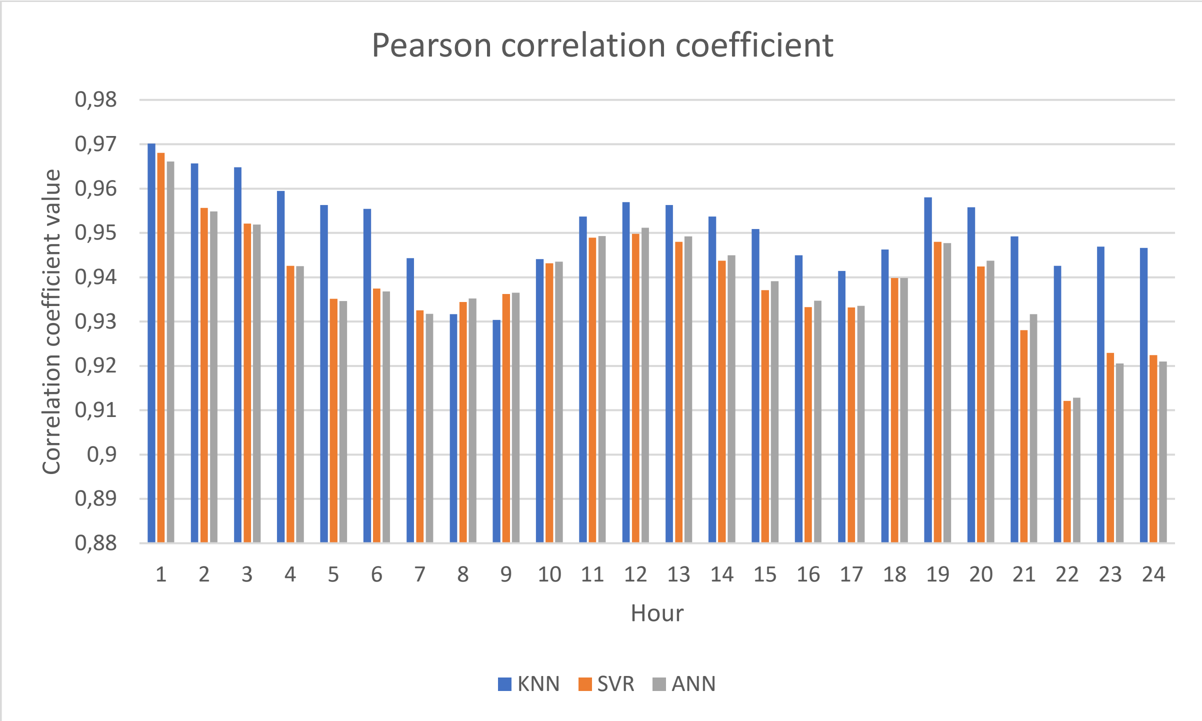

Fig. 4.7 Pearson correlation coefficients for 24 hours - all models

Pearson correlation coefficients for all obtained models remain at high level – always above 0,9. The average values for analysed methods are equal to 0,95 for KNN and 0,94 for both SVR and ANN.

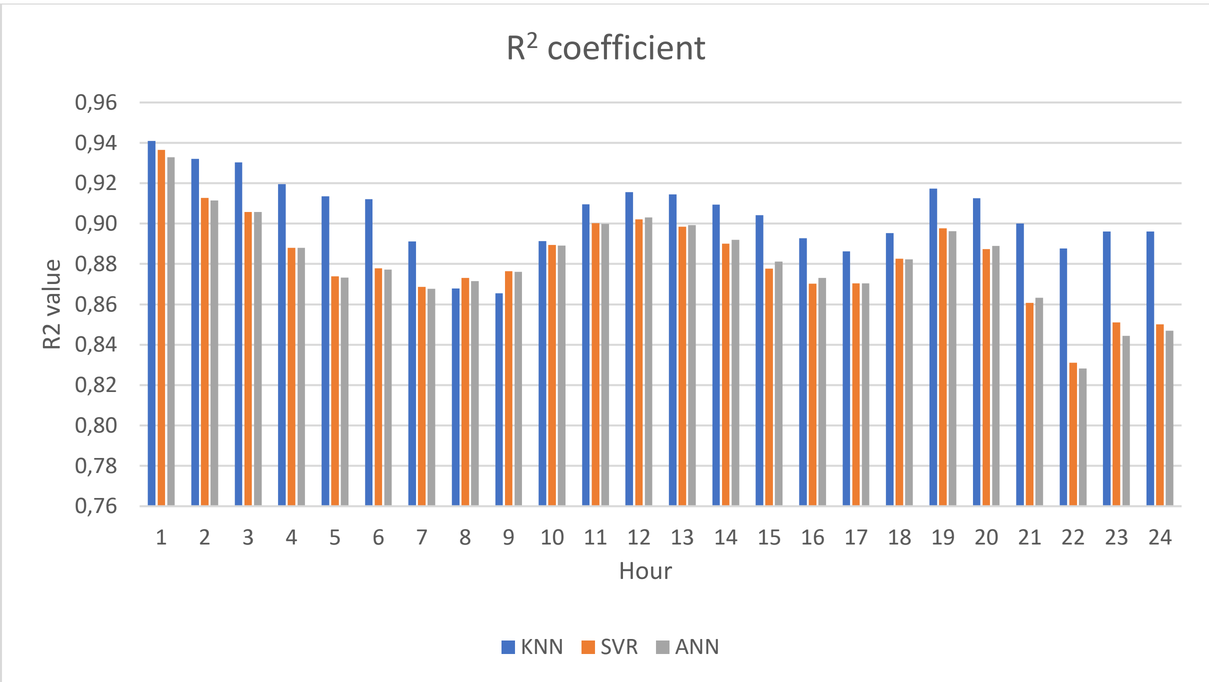

Fig. 4.8 R^2 coefficients for 24 hours - all models

Also R2 coefficients for all developed models exhibit good or very good fit – the average values are: 0,9 for KNN and 0,88 for both SVR and ANN. However, as it was already stated – in the case of time series even such high results alone may not be a sufficient measure for the quality evaluation.

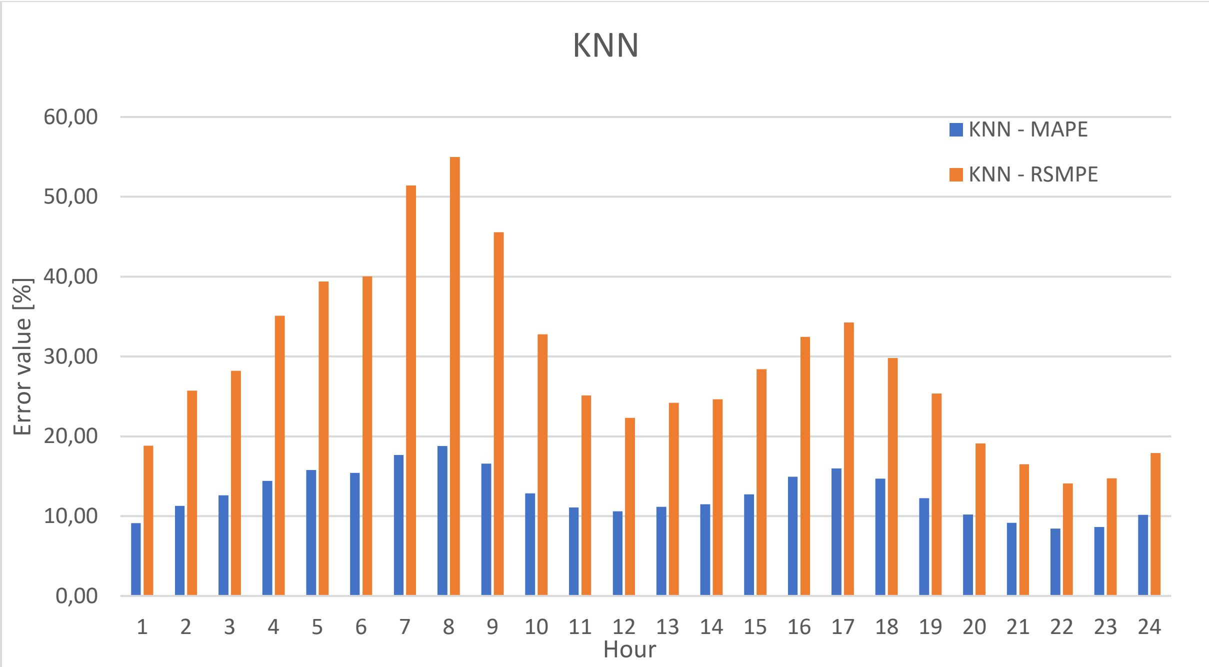

Fig. 4.9 MAPE and RSMPE values for 24 hours - KNN model

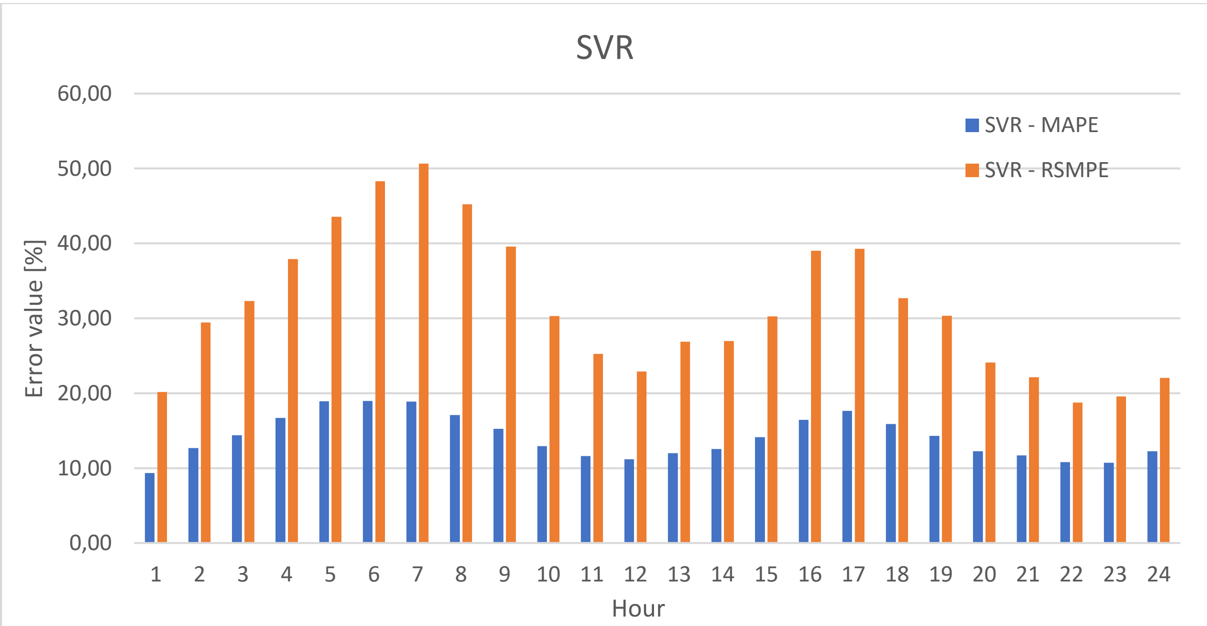

Fig. 4.10 MAPE and RSMPE values for 24 hours - SVR model

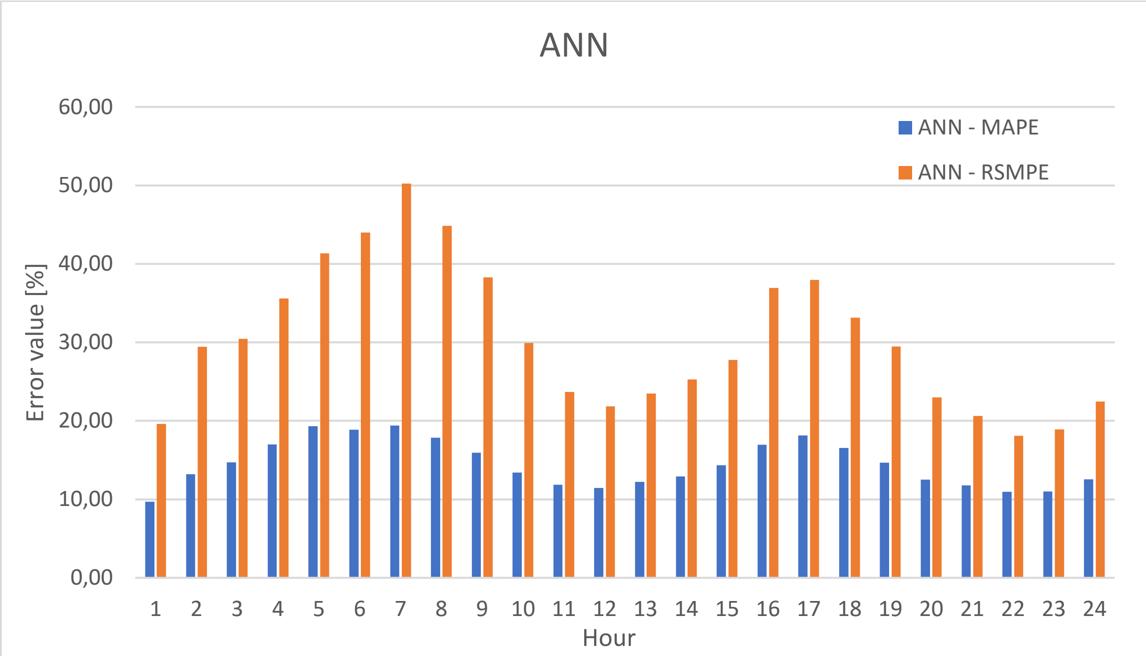

Fig. 4.11 MAPE and RSMPE values for 24 hours - ANN model

Based on the charts 4.7-4.9 it can be stated that MAPE reaches similar maximal and minimal values for all methods (respectively around 20% and around 10%). So does RSMPE: maximal value is equal to around 50% and minimal to around 20%. What is interesting, error peaks and valleys form in approximately the same hours: peaks – 7-8 and 16-17 hour; valleys: 12 and 22 hour.

5. Conclusions

Electricity price forecasting is one of the most impactful factors in companies' decision making processes - proper forecasts with adequate prediction horizon lead to reasonable economical decisions. However, due to special characteristic of electricity as a commodity its price is difficult to predict . For this purpose machine learning tools, which offer high level of complexity and flexibility, are recently popular and widely applied and therefore surely will be further developed.

The goal of this assignment was to compare three different machine learning tools for price prediction: k-nearest neighbours, support vector regression and artificial neural networks. The obtained models would not be satisfying for industrial purposes, but provide interesting observations for educational ones.

Each model was characterized by very high correlation coefficient and R^2 coefficient values, which however could be caused by the fact, that developed models were based on time series data. By contrast MAPE and RSMPE reached significant values – maximal average values for no matter which method were equal to around 20% and 50% respectively. This leads to the conclusion, that created models are not bad, but could be more satisfying, which was also confirmed by the example cross validation results.

Possible factors which possibly could improve the quality of the models are more properly selected features (for example weather, time of the year), larger training data set or different combination of hyperparameters.

This assignment proves that creating an optimal model is not a task that can be done automatically. Multiplicity of existing approaches and parameters requiring determination makes an individual approach the one which gives the best results. The quality of the model is not just a matter of choosing the optimal algorithm. However in reality, where models are much more complex than those proposed in the assignment, simply checking all the possibilities is extremely difficult and time-consuming.

6. References

[1] R. Weron, „Electricity price forecasting: A review of the state-of-the-art with a look into the future," *International Journal of Forecasting,* tom 30, nr 4, p. 1030–1081, 2014.

[2] Bunn, Derek W, Modelling Prices in Competitive Electricity Markets, Wiley, 2014.

[3] R. Handik, T. Chi, T. Stefan i W. Rafal, „Modelling price spikes in electricity markets - the impact of load, weather and capacity," *HSC Research Reports HSC/14/08 Hugo Steinhaus Center, Wroclaw University of Technology,* 2014.

[4] „RED ELECTRICA DE ESPAÑA," [Online]. Available: https://demanda.ree.es/visiona/peninsula/demanda/total.

[5] R. Weron, Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach, Wiley, 2006.

[6] M. Ventosa, Á. Baíllo i A. Ramos, „Electricity market modeling trends," *Energy Policy,* tom 33, nr 7, p. 897–913, 2005.

[7] M. Burger, B. Graeber i G. Schindlmayr, Managing Energy Risk: An Integrated View on Power and Other Energy Markets., Wiley, 2007.

[8] D. Keles, J. Scelle, F. Paraschiv i W. Fichtner, „Extended forecast methods for day-ahead electricity spot prices applying artificial neural networks," *Applied Energy,* tom 162, pp. 218-230, 2016.

[9] C. Rodriguez i G. Anders, „ Energy price forecasting in the Ontario competitive power system market," *IEEE Transactions on Power Systems.,* tom 19, nr 1, pp. 366-374, 2004.

[10] „OMIE," [Online]. Available: http://m.omie.es/en/home/markets-and-products/electricity-market/our-electricity-markets?m=yes.

[11] „Europex," [Online]. Available: https://www.europex.org/members/omie/.

[12] „PCR&EUPHEMIA algorithm," [Online]. Available: https://www.n-side.com/pcr-euphemia-algorithm-european-power-exchanges-price-coupling-electricity-market/.

[13] N. Altman, „An introduction to kernel and nearest-neighbor nonparametric regression," *The American Statistician,* tom 46, nr 2, p. 175–185, 1992.

[14] L.-Y. Hu, M.-W. Huang, S.-W. Ke i T. Chih-Fong, „The distance function effect on k-nearest neighbor classification for medical datasets," 2016. [Online]. Available: https://springerplus.springeropen.com/articles/10.1186/s40064-016-2941-7.

[15] „Python Scikit-Learn Documentation," [Online]. Available: https://scikit-learn.org/.

[16] A. Smola, „A Tutorial on Support Vector Regression," 30 September 2003. [Online]. Available: https://alex.smola.org/papers/2003/SmoSch03b.pdf.

[17] I. Bhattacharyya, „Support Vector Regression Or SVR," [Online]. Available: https://medium.com/coinmonks/support-vector-regression-or-svr-8eb3acf6d0ff.

[18] L. Dormehl, „What is an artificial neural network? Here's everything you need to know," [Online]. Available: https://www.digitaltrends.com/cool-tech/what-is-an-artificial-neural-network/.

[19] M. Nielsen, „Neural Networks and Deep Learning," October 2018. [Online]. Available: http://neuralnetworksanddeeplearning.com/chap1.html.

[20] A. Dertat, „Applied Deep Learning - Part1: Artificial Neural Network," 8 August 2917. [Online]. Available: https://towardsdatascience.com/applied-deep-learning-part-1-artificial-neural-networks-d7834f67a4f6.

[21] „Scaling data for Deep Learning," [Online]. Available: http://www.feeny.org/scaling-data-deep-learning/.