A New Methodology Based on Imbalanced Classification for Predicting Outliers in Electricity Demand Time Series

Francisco Javier Duque-Pintor †, Manuel Jesús Fernández-Gómez †, Alicia Troncoso * and Francisco Martínez-Álvarez

Division of Computer Science, Universidad Pablo de Olavide, ES-41013 Seville, Spain;
fjduqpin@upo.es (F.J.D.-P.); mjfergom@alu.upo.es (M.J.F.-G.); fmaralv@upo.es (F.M.-Á.)
* Correspondence: atrolor@upo.es; Tel.: +34-95-4349-230
† These authors contributed equally to this work.

Abstract: The occurrence of outliers in real-world phenomena is quite usual. If these anomalous data are not properly treated, unreliable models can be generated. Many approaches in the literature are focused on a posteriori detection of outliers. However, a new methodology to a priori predict the occurrence of such data is proposed here. Thus, the main goal of this work is to predict the occurrence of outliers in time series, by using, for the first time, imbalanced classification techniques. In this sense, the problem of forecasting outlying data has been transformed into a binary classification problem, in which the positive class represents the occurrence of outliers. Given that the number of outliers is much lower than the number of common values, the resultant classification problem is imbalanced. To create training and test sets, robust statistical methods have been used to detect outliers in both sets. Once the outliers have been detected, the instances of the dataset are labeled accordingly. Namely, if any of the samples composing the next instance are detected as an outlier, the label is set to one. As a study case, the methodology has been tested on electricity demand time series in the Spanish electricity market, in which most of the outliers were properly forecast.

Keywords: time series; imbalanced classification; forecasting; outliers

1. Introduction

Prediction tools have become important for agents participating in electricity markets. Electricity generation companies need to schedule electrical energy production to satisfy the forecasted load. In this sense, demand forecasting plays an important role for electricity power suppliers, because both excess and insufficient energy production may lead to increased costs and a significant reduction of profits. Therefore, it is quite important to obtain forecasts for electricity demand as accurately as possible [1].

The electricity load time series shows complex characteristics influenced by diverse factors, such as meteorological conditions, seasonal patterns, or socioeconomic factors. As a consequence, the demand presents some peculiarities such as the presence of outliers, that turn the forecasting process into a particularly challenging task.

It is worth highlighting the difference between the forecasting of the occurrence of an outlier and its detection. The detection consists of discovering the outliers in an already known set of values, which is a common goal in robust statistics [2]. The majority of robust statistical techniques try to obtain a time series model from data once the outliers have been replaced. Thus, these techniques perform a posteriori detection; that is, they determine whether a point is an outlier or not, but once it has already occurred. However, the problem considered here is more difficult, since a prediction of

the occurrence of an outlier is made with the goal of electricity companies activating adequate action protocols or using forecasting methods specifically designed for the prediction of the magnitude of outliers once it is known that an outlier is going to occur.

In this work, given a time series of hourly electricity loads up to day d, the goal is to forecast if an outlier will occur over the 24 h loads for day $d + 1$. A new methodology based on imbalanced classification [3] is presented, which, to the best of the authors' knowledge, has yet to be exploited in outlier forecasting problems. Note that robust statistics techniques will be of utmost importance in order to transform a prediction problem into a classification problem, since the class is built from detected outliers in the dataset.

The remainder of the paper is organized as follows. A review of the most recently published works regarding outlier forecasting in demand time series can be found in Section 2. Section 3 introduces the proposed methodology, showing how to transform the outlier occurrence forecasting into a binary classification general scheme. The results obtained for the Spanish electricity demand time series are reported and discussed in Section 4. Finally, Section 5 summarizes the main conclusions achieved.

2. Related Work

The problem of a posteriori outliers detection in time series has been widely studied in the literature, and has been addressed by many approaches. This is due to the high impact that the existence of outliers can cause by generating inaccurate models [4], since they may deeply influence the estimates that classical methods propose [5].

To deal with this issue, there is a large family of robust statistical methods [6]. Gelper et al. proposed an adapted version of the classical exponential and Holt–Winters smoothing methodologies, providing them with robustness [7]. Another version of a robust multivariate exponential smoothing applied to time series can be found in [8]. Following classical methods, a work that enhanced Auto Regressive Moving Average (ARMA) by adding robustness can be found in [9], in which the authors succeeded in limiting the effect of outlying data to the time stamp in which they happen.

Robust estimations can be found in electricity prices. In fact, a battery of over 300 models were considered in [10] to forecast the long-term seasonal component. The authors concluded that those based on wavelet are significantly better in terms of forecasting spot prices for up to a year ahead.

A robust weighted combination load forecasting method based on forecast model filtering and adaptive variable weight determination was proposed in [11]. In particular, the authors proposed an Immune Algorithm-Particle Swarm Optimization that was applied to Chinese data.

However, this work is concerned with a priori outlier detection or, in other words, with predicting the occurrence of these anomalous values in real time. A method for the prediction of outlier occurrence was proposed in [12]. In particular, The Pattern Sequence Forecasting (PSF) algorithm [13] was adapted to deal with spike values in the field of electricity price forecasting. As a case study, the markets of New York, Australia, and the Iberian Peninsula were examined. An early version of this algorithm can be found in [14].

An approach based on multi-feature wavelet and an Extreme Learning Machine (ELM) algorithm for the forecasting of outlier occurrence in the Chinese stock market was proposed in [15]. To ensure the universal application of the algorithm, the authors selected two market indexes in Shanghai and Shenzhen, as well as six other individual stocks.

Later in 2016, the outlier occurrence prediction in the stock market was again addressed in [16]. In this case, the wavelet transform and an adaptive ELM algorithm were used to analyze daily values of the petroleum sector index from Tehran, Iran. The model was compared to several methods, showing some improvement in the results achieved.

3. Methodology

This section describes the proposed methodology for the forecasting of the occurrence of outliers in time series. The first step consists of formulating the outlier prediction problem as a binary classification

problem. Later, a classifier is applied to predict the occurrence of outliers. As the number of outliers is usually small, this formulation of the problem generates an imbalanced binary classification problem.

3.1. Formulation of the Problem

The attributes are composed of a window of past values of the time series, and the class can be 1 or 0, depending on whether or not an outlier has occurred in the prediction horizon. Therefore, firstly the labels 1 or 0 have to be constructed for each instance of the dataset.

Figure 1 shows the basic idea behind the proposed methodology. All the steps composing this methodology are described in subsequent sections.

Figure 1. Illustration of the proposed methodology.

3.1.1. Detecting Outliers

The outliers are detected in the historical data by applying a robust statistical method. In particular, the robust method proposed in Gelper et al. [7] to detect outliers in time series has been considered. This method carries out a cleansing process of the time series to replace the outliers by a more likely value prior to generation of the time series forecasting model.

Namely, a time series value is replaced, and is therefore considered an outlier, if the difference between the observed value at time t and its predicted value at time $t-1$ is too large. Thus, the set of outliers OS of a time series is defined by:

$$OS = \{y_t : y_t - \widehat{y}_{t-1} > k \cdot \widehat{\sigma}_t\} \tag{1}$$

where k typically is set to 2 or 3, depending on if moderate or extreme outliers are considered, respectively. The predicted value $\widehat{y}_t$ is obtained by a robust exponential smoothing model, and $\widehat{\sigma}_t$ is a robust estimation of the scale of the $y_t - \widehat{y}_{t-1}$ errors. Namely, the prediction and the scale are defined in a recursive way as:

$$\widehat{y}_t = \lambda y_t^* + (1-\lambda)\widehat{y}_{t-1} \tag{2}$$

$$\widehat{\sigma}_t^2 = \lambda_\sigma \rho\left(\frac{y_t - \widehat{y}_{t-1}}{\widehat{\sigma}_{t-1}}\right)\widehat{\sigma}_{t-1}^2 + (1-\lambda_\sigma)\widehat{\sigma}_{t-1}^2 \tag{3}$$

where y_t^* is the cleaned value of y_t given by Equations (4) and (5), ρ is the loss-function defined by Equation (6), and λ and λ_σ are smoothing parameters between 0 and 1 which have to be determined in the learning phase from a training set.

$$y_t^* = \phi\left(\frac{y_t - \widehat{y}_{t-1}}{\widehat{\sigma}_t}\right)\widehat{\sigma}_t + \widehat{y}_{t-1} \tag{4}$$

$$\phi(x) = \begin{cases} x & \text{if } |x| \leq k \\ sign(x)k & \text{otherwise} \end{cases} \tag{5}$$

$$\rho(x) = \begin{cases} c_k(1 - (1 - (x/k)^2))^3) & \text{if } |x| \leq k \\ c_k & \text{otherwise} \end{cases} \tag{6}$$

The c_k value is an input parameter related to the parameter k (for example, $c_k = 2.52$ for a common value of $k = 2$ [7]). The initial values used to obtain the prediction and the scale in a recursive way are usually the mean and the standard deviation, respectively, of the first values in the time series.

3.1.2. Labeling the Dataset

Once the set of outliers for the historical data have been discovered, the instances of the dataset must be labeled with their corresponding class. Given an instance composed of *m* past values of the time series and a prediction horizon of *h*, the label is 1 if an outlier occurs over the *h* next values of the time series (in our case, $h = 24$ h). That is, the class *C* is defined by:

$$C = \begin{cases} 1 & \text{if } \exists i \in \{t+1, ..., t+h\} \text{ such that } y_i \in OS \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

3.2. Imbalanced Classification

The methodology applied in order to forecast the occurrence of outliers for the twenty-four hours of the next day is described in this section.

Since outliers are anomalous data, an imbalanced classification problem is obtained when the prediction of outliers is formulated as a binary classification problem. Therefore, the class representing the outliers is a minority class, but the class of interest.

The approaches proposed to solve imbalanced classification problems can be split into two differentiated groups: algorithm-based approaches that design specific algorithms to deal with the minority class, and data-based approaches, which apply a preprocessing step to try to balance the classes before applying a learning algorithm [3]. In this work, a selection of representative methods of the first group are firstly used, and thereafter, the algorithm with the best performance will be combined with different preprocessing methods in order to improve the results of the forecasts.

Table 1 shows both preprocessing and classification techniques that have been analyzed to provide a forecasting of outliers in the electricity demand time series. Due to the good behavior exhibited in oversampling methods [3], a number of oversampling-based preprocessing techniques greater than those based on undersampling have been tested. All these techniques can be found in the KEEL open source java software project [17].

Table 1. Techniques of imbalanced classification.

Preprocessing			Classification	
Algorithm	**Type**	**Reference**	**Algorithm**	**Reference**
ADASYN	Oversampling	[18]	AdaBoost	[19]
ADOMS	Oversampling	[20]	AdaBoostM1	[21]
ROS	Oversampling	[22]	AdaBoostM2	[23]
Safe-Level-Smote	Oversampling	[24]	Bagging	[25]
SMOTE	Oversampling	[26]	BalanceCascade	[27]
SMOTE-ENN	Oversampling	[22]	C45CS	[28]
SMOTE-TL	Oversampling	[22]	C-SVMCS	[29]
SPIDER	Oversampling	[30]	DataBoost-IM	[31]
SPIDER2	Oversampling	[24]	EasyEnsemble	[27]
CNN	Undersampling	[32]	UnderBagging	[33]
CNNTL	Undersampling	[34]	UnderBagging2	[33]
TL	Undersampling	[34]	UnderOverBagging	[35]

4. Results

The above-described methodology has been applied to the electricity demand of the Spanish market [36].

This section is structured as follows: first, a brief description of the electricity demand time series is included. Second, the usual quality parameters in an imbalanced context are presented. Third, the robust exponential smoothing model is obtained in order to detect outliers, and for this reason, the

election of both λ and λ_{scale} is discussed here. Finally, the accuracy of the predictions of outliers is validated.

4.1. Dataset

The electricity demand time series from 1 January 2007 to 20 June 2016 has been recorded to carry out the analysis presented in this work. The time series is measured at hourly intervals and is composed of a total of 82,975 samples, which have been split into 49,785 samples for the training set corresponding to the period from 1 January 2007 to 8 September 2012, and 33,190 samples for the test set corresponding to days from 9 September 2012 to 20 June 2016.

4.2. Evaluation Measures

The parameters used to assess the accuracy of the classifiers are introduced in this section. Note that in subsequent equations, true positives (TP) is the number of outliers properly predicted; true negatives (TN) is the number of days that were not properly-predicted outliers; false positives (FP) is the number of days that were not outliers and were predicted as outliers; and false negatives (FN) is the number of outliers which were predicted as common days. Note that the prediction horizon is 24 h, and therefore, the measures of evaluation are defined with respect to a day.

According to these definitions, the sensitivity is the ratio of outliers properly predicted by the classification technique. Its formula is defined as follows:

$$S_n = \frac{TP}{TP + FN} \tag{8}$$

Another parameter is the specificity, which is the ratio of days that were not properly predicted outliers. The mathematical expression is:

$$S_p = \frac{TN}{TN + FP} \tag{9}$$

The positive predictive value (PPV) is the probability of predicting an outlier correctly. Its formula is:

$$PPV = \frac{TP}{TP + FP} \tag{10}$$

Finally, the negative predictive value (NPV) is the probability that a point that was not an outlier was properly predicted. Its formula is:

$$NPV = \frac{TN}{TN + FN} \tag{11}$$

The performance of most classifiers is evaluated with the accuracy or error measures, defined by the proportion of instances correctly or incorrectly classified for both classes. However, these measures do not distinguish between the number of correct labels for each class, which is important in the context of imbalanced classification problems, as the class corresponding to outliers is the class of interest in this kind of problem. For that, measures intending to achieve good quality results for both classes are preferred in order to assess the performance of the imbalanced classification techniques. The following measures have been considered:

- *F*-measure or balanced *F*-score (*F*) is the harmonic mean of the PPV and sensitivity measures:

$$F = \frac{2 \cdot PPV \cdot S_n}{S_n + PPV} \tag{12}$$

- The area under the receiver operating characteristic (ROC) curve (*AUC*). The ROC curve shows the relation between sensitivity and specificity. That is, trade-offs between benefits (true positives) and costs (false positives).

$$AUC = \frac{1 + S_n - FP_{rate}}{2} \qquad (13)$$

where FP_{rate} is the false positive rate; that is, the ratio between the number of false positives and the total number of days that are not outliers.

- The geometric mean (*GM*) of the sensitivity and specificity measures:

$$GM = \sqrt{S_n \cdot S_p} \qquad (14)$$

- Matthew's Correlation Coefficient (*MCC*), proposed in [37], provides better balance among the four basic metrics S_n, S_p, PPV, and NPV.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TN + FN)(TP + FP)(TP + FN)(TN + FP)}} \qquad (15)$$

4.3. Training to Detect Outliers

In this section, the learning phase to obtain the robust exponential smoothing model is carried out. Note that this model allows the detection of outliers to label the dataset in order to apply a supervised learning for classification tasks.

The training consists of computing the parameters of the model, λ and λ_{scale}, from the historical time series. For this reason, the mean absolute percentage error (MAPE) when predicting the time series with the robust exponential smoothing model has been minimized for λ and λ_{scale}, varying from 0.1 to 0.9 by increments of 0.1. The first fifty values of the demand time series have been used as the starting period to compute the initial value $\widehat{y}_0$ in Equation (2) and $\widehat{\sigma}_0$ in Equation (3).

Figure 2 presents the surface representing the error for the different values of λ and λ_{scale}. It can be noticed that the minimum error is 2.46% and is reached for $\lambda = 0.1$ and $\lambda_{scale} = 0.3$.
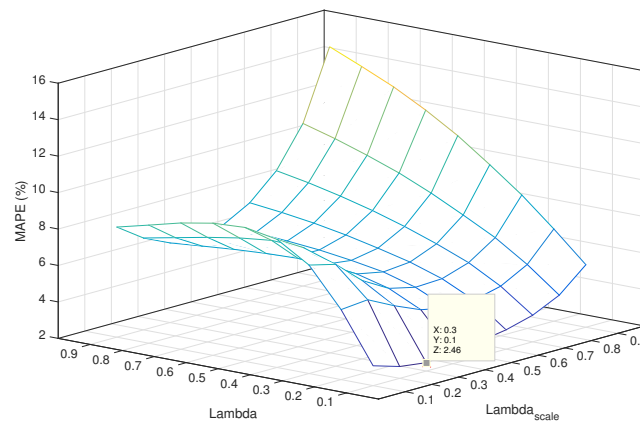

Figure 2. Parameters of the robust exponential smoothing model. MAPE: mean absolute percentage error.

The detection of outliers is made with the resulting model, and a total of 551 outliers were detected. This means a percentage of 15.98% of outliers in the times series, and therefore, the selection of imbalanced techniques to forecast outliers in the electricity demand is justified.

4.4. Outlier Occurrence Forecasting

The results obtained from the application of the classification techniques specified in Table 1 to the test set are reported in this section. The distribution of outliers for both training and test sets is of 374 outliers (18.08%) and 177 outliers (12.83%), respectively.

Table 2 shows the quality measures for the test set for each classification technique. The control quality parameter is *MCC*, since it provides a global measure of all indicators. However, *PPV* is also considered, given the nature of the addressed problem. It can be concluded that algorithms Bagging and C-SVMCS achieve the best results, with *MCC* around 0.6. Additionally, the best *PPV* values are also reached for these algorithms. As for the rest of the parameters, they exhibit satisfactory values with, for instance, *AUC* around 0.8, or F-measures slightly inferior to 0.8. As for the rest of the algorithms, their *MCC* values range from 0.181 (DataBoost-IM) to 0.462 (AdaBoost), which are not particularly good for *MCC* (remember that $MCC \in [-1, 1]$, -1 being the worst value, and 1 the best).

Table 2. Evaluation measures for each imbalanced classification algorithm.

Classifier	*TP*	*FP*	*FN*	*TN*	S_n	S_p	*PPV*	*NPV*	*F*	*AUC*-ROC	*GM*	*MCC*
AdaBoost	114	135	63	1068	0.644	0.888	0.458	0.944	0.725	0.622	0.756	0.462
AdaBoostM1	87	189	90	1014	0.492	0.843	0.315	0.918	0.632	0.667	0.644	0.280
AdaBoostM2	87	189	90	1014	0.492	0.843	0.315	0.918	0.632	0.667	0.644	0.280
Bagging	106	45	71	1158	0.599	0.963	0.702	0.942	0.799	0.781	0.759	0.601
BalanceCascade	160	580	17	623	0.904	0.518	0.216	0.973	0.513	0.711	0.684	0.283
C45CS	143	233	34	970	0.808	0.806	0.380	0.966	0.698	0.807	0.807	0.461
C-SVMCS	164	191	13	1012	0.927	0.841	0.462	0.987	0.762	0.884	0.883	0.587
DataBoost-IM	126	532	51	671	0.712	0.558	0.191	0.929	0.499	0.635	0.630	0.181
EasyEnsemble	160	525	17	678	0.904	0.564	0.234	0.976	0.543	0.734	0.714	0.313
UnderBagging	152	590	25	613	0.859	0.510	0.205	0.961	0.498	0.684	0.662	0.247
UnderBagging2	152	427	25	776	0.859	0.645	0.263	0.969	0.588	0.752	0.744	0.341
UnderOverBagging	148	345	29	858	0.836	0.713	0.300	0.967	0.631	0.775	0.772	0.383

For this reason, Bagging and C-SVMCS have been selected as candidate algorithms to reach the best results, and preprocessing algorithms have been applied as an initial step. Table 3 shows the performance of the Bagging algorithm when all preprocessing algorithms described in Section 3 are previously applied to the time series in order to balance the two classes. Analogously, Table 4 summarizes the results of combining the preprocessing algorithms with C-SVMCS.

Table 3. Evaluation measures for each preprocessing, when Bagging is applied.

Preprocess	*TP*	*FP*	*FN*	*TN*	S_n	S_p	*PPV*	*NPV*	*F*	*AUC*-ROC	*GM*	*MCC*
ADASYN	148	347	29	856	0.836	0.712	0.299	0.967	0.630	0.774	0.771	0.382
ADOMS	143	273	34	930	0.808	0.773	0.344	0.965	0.670	0.790	0.790	0.423
ROS	139	155	38	1048	0.785	0.871	0.473	0.965	0.753	0.828	0.827	0.536
Safe-Level-SMOTE	139	170	38	1033	0.785	0.859	0.450	0.965	0.740	0.822	0.821	0.517
SMOTE	133	165	44	1038	0.751	0.863	0.446	0.959	0.734	0.807	0.805	0.499
SMOTE-ENN	140	162	37	1041	0.791	0.865	0.464	0.966	0.749	0.828	0.827	0.531
SMOTE-TL	150	220	27	983	0.847	0.817	0.405	0.973	0.718	0.832	0.832	0.502
SPIDER	143	207	34	996	0.808	0.828	0.409	0.967	0.717	0.818	0.818	0.489
SPIDER2	133	168	44	1035	0.751	0.860	0.442	0.959	0.732	0.806	0.804	0.495
CNN	146	404	31	799	0.825	0.664	0.265	0.963	0.594	0.745	0.740	0.334
CNNTL	168	740	9	463	0.949	0.385	0.185	0.981	0.431	0.667	0.604	0.235
TL	140	105	37	1098	0.791	0.913	0.571	0.967	0.801	0.852	0.850	0.616

Table 4. Evaluation measures for each preprocessing, when C-SVMCS is applied.

Preprocess	*TP*	*FP*	*FN*	*TN*	S_n	S_p	*PPV*	*NPV*	*F*	*AUC*-ROC	*GM*	*MCC*
ADASYN	168	242	9	961	0.949	0.799	0.410	0.991	0.728	0.874	0.871	0.547
ADOMS	161	174	16	1029	0.910	0.855	0.481	0.985	0.772	0.882	0.882	0.597
ROS	156	188	21	1015	0.881	0.844	0.453	0.980	0.753	0.863	0.862	0.560
Safe-Level-SMOTE	165	235	12	968	0.932	0.805	0.413	0.988	0.729	0.868	0.866	0.543
SMOTE	154	181	23	1022	0.870	0.850	0.460	0.978	0.755	0.860	0.860	0.561
SMOTE-ENN	145	115	32	1088	0.819	0.904	0.558	0.971	0.800	0.862	0.861	0.619
SMOTE-TL	155	149	22	1054	0.876	0.876	0.510	0.980	0.785	0.876	0.876	0.607
SPIDER	165	262	12	941	0.932	0.782	0.386	0.987	0.710	0.857	0.854	0.517
SPIDER2	162	249	15	954	0.915	0.793	0.394	0.985	0.715	0.854	0.852	0.518
CNN	151	337	26	866	0.853	0.720	0.309	0.971	0.640	0.786	0.784	0.401
CNNTL	161	449	16	754	0.910	0.627	0.264	0.979	0.587	0.768	0.755	0.361
TL	160	187	17	1016	0.904	0.845	0.461	0.984	0.760	0.874	0.874	0.577

From the analysis of these two tables, several conclusions can be drawn. First, MCC increases for both algorithms, thus showing that results are better in general terms. The TL preprocessing algorithm increased the MCC value to 0.616 for Bagging, and the SMOTE-ENN algorithm to 0.619 for C-SVMCS. Second, with these algorithms, PPV values are the two best (a bit lower for Bagging, 0.571; and a bit higher for C-SVMCS, 0.558). Third, in general, the values for the F, AUC-ROC, and GM measures have improved, and the values from both sensitivity and specificity remain quite high. This shows that, despite the use of imbalanced classes, the algorithms are able to distinguish between one class and another.

5. Conclusions

This work presents a new methodology for the forecasting of outlier occurrence in time series, with application to the Spanish electricity demand. The main step consists of transforming the problem into an imbalanced classification problem, paying particular attention to how the class is defined, in order to ensure that the prediction of outliers over a prediction horizon of twenty four hours is made. A representative number of classification algorithms specifically designed for imbalanced problems has been tested, showing that the Bagging and C-SVMCS algorithms reach the best results. Later, the results of these algorithms when several preprocessing algorithms are applied have been reported, with the objective of improving the quality measures. In this case, F, AUC-ROC, and GM measures greater than 0.8 and MCC greater than 0.6 have been obtained when the TL and SMOTE-ENN preprocessing techniques were applied. From the results obtained, it can be concluded that the new methodology proposed here provides a satisfactory accuracy. Future work is directed towards predicting not only the days that will present anomalous behavior, but also the magnitude of the outliers. That is, the problem will be formulated as a multiclass imbalanced classification problem, and outliers of different magnitudes will be classified in different classes.

Author Contributions: Alicia Troncoso and Francisco Martínez-Álvarez conceived the paper. Francisco Javier Duque-Pintor and Manuel Jesús carried out the experimentation. All authors contributed to the writing of the paper.

References

1. Martínez-Álvarez, F.; Troncoso, A.; Asencio-Cortés, G.; Riquelme, J.C. A survey on data mining techniques applied to energy time series forecasting. *Energies* **2015**, *8*, 13162–13193.
2. Maronna, R.A.; Martin, R.D.; Yohai, V.J. *Robust Statistics: Theory and Methods*; Wiley: Hoboken, NJ, USA, 2007.

3. Galar, M.; Fernández, A.; Barrenechea, E.; Bustince, H.; Herrera, F. A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches. *IEEE Trans. Syst. Man Cybern. C* **2012**, *42*, 463–484.

4. Galeano, P.; Peña, D.; Tsay, R.S. Outlier detection in multivariate time series by projection pursuit. *J. Am. Stat. Assoc.* **2006**, *101*, 645–669.

5. Carnero, M.A.; Peña, D.P.; Ruiz, E. Effects of outliers on the identification and estimation of GARCH models. *J. Time Ser. Anal.* **2007**, *28*, 471–497.

6. Rousseeuw, P.J.; Hubert, M. Robust statistics for outlier detection. *WIREs Data Min. Knowl. Discov.* **2011**, *1*, 73–79.

7. Gelper, S.; Fried, R.; Croux, C. Robust forecasting with exponential and Holt-Winters smoothing. *J. Forecast.* **2010**, *29*, 285–300.

8. Croux, C.; Gelper, S.; Mahieu, K. Robust exponential smoothing of multivariate time series. *Comput. Stat. Data Anal.* **2010**, *54*, 2999–3006.

9. Muler, N.; Peña, D.; Yohai, V.J. Robust estimation for ARMA models. *Ann. Stat.* **2009**, *37*, 816–840.

10. Nowotarski, J.; Tomczyk, J.; Weron, R. Robust estimation and forecasting of the long-term seasonal component of electricity spot prices. *Energy Econ.* **2013**, *39*, 13–27.

11. Li, L.; Mu, C.; Ding, S.; Wang, Z.; Mo, R.; Song, Y. A robust weighted combination forecasting method based on forecast model filtering and adaptive variable weight determination. *Energies* **2016**, *9*, 20.

12. Martínez-Álvarez, F.; Troncoso, A.; Riquelme, J.C.; Aguilar-Ruiz, J.S. Discovery of motifs to forecast outlier occurrence in time series. *Pattern Recognit. Lett.* **2011**, *32*, 1652–1665.

13. Martínez-Álvarez, F.; Troncoso, A.; Riquelme, J.C.; Aguilar-Ruiz, J.S. Energy time series forecasting based on pattern sequence similarity. *IEEE Trans. Knowl. Data Eng.* **2011**, *23*, 1230–1243.

14. Martínez-Álvarez, F.; Troncoso, A.; Riquelme, J.C. Improving time series forecasting by discovering frequent episodes in sequences. *Lect. Notes Comput. Sci.* **2009**, *5772*, 357–368.

15. Fang, Z.; Zhao, J.; Fei, F.; Wang, Q.; He, X. An approach based on multi-feature wavelet and ELM algorithm for forecasting outlier occurrence in Chinese stock market. *J. Theor. Appl. Inf. Technol.* **2013**, *49*, 369–377.

16. Hosseinioun, N. Forecasting outlier occurrence in stock market time series based on wavelet transform and adaptive ELM algorithm. *J. Math. Financ.* **2016**, *6*, 127–133.

17. Alcalá-Fernández, J.; Fernández, A.; Luengo, J.; Derrac, J.; García, S.; Sánchez, L.; Herrera, F. Keel data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework. *J. Multiple-Valued Logic Soft Comput.* **2011**, *17*, 255–287.

18. He, H.; Bai, Y.; García, E.A.; Li, S. Adasyn: Adaptive synthetic sampling approach for imbalanced learning. In Proceedings of the International Joint Conference on Neural Networks, Hong Kong, China, 1–8 June 2008; pp. 1322–1328.

19. Schapire, R.E. The strength of weak learnability. *Mach. Learn.* **1990**, *5*, 197–227.

20. Tang, S.; Chen, S. The generation mechanism of synthetic minority class examples. In Proceedings of the Conference on Information Technology and Applications in Biomedicine, Shenzhen, China, 30–31 May 2008; pp. 444–447.

21. Freund, Y.; Schapire, R.E. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139.

22. Batista, G.; Prati, R.C.; Monard, M.C. A study of the behavior of several methods for balancing machine learning training data. *SIGKDD Explor.* **2004**, *6*, 20–29.

23. Schapire, R.E.; Singer, Y. Improved boosting algorithms using confidence-rated predictions. *Mach. Learn.* **1996**, *37*, 297–336.

24. Napierała, K.; Stefanowski, J.; Wilk, S. Learning from imbalanced data in presence of noisy and borderline examples. *Lect. Notes Comput. Sci.* **2010**, *6086*, 158–167.

25. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140.

26. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357.

27. Liu, X.-Y.; Wu, J.; Zhou, Z.-H. Exploratory undersampling for class-imbalance learning. *IEEE Trans. Syst. Man Cybern. B* **2009**, *39*, 539–550.

28. Ting, K.M. An instance-weighting method to induce cost-sensitive trees. *IEEE Trans. Knowl. Data Eng.* **2002**, *14*, 659–665.

29. Tang, Y.; Zhang, Y.-Q.; Chawla, N.V.; Krasser, S. SVMs modeling for highly imbalanced classification. *IEEE Trans. Syst. Man Cybern. B* **2009**, *39*, 281–288.

30. Stefanowski, J.; Wilk, S. Selective pre-processing of imbalanced data for improving classification performance. *Lect. Notes Comput. Sci.* **2008**, *5182*, 283–292.

31. Guo, H.; Viktor, H.L. Learning from imbalanced data sets with boosting and data generation: The DataBoost-IM approach. *SIGKDD Explor.* **2004**, *6*, 30–39.

32. Hart, P.E. The condensed nearest neighbour rule. *IEEE Trans. Inf. Theory* **1968**, *14*, 515–516.

33. Barandela, R.; Valdovinos, R.M.; Sánchez, J.S. New applications of ensembles of classifiers. *Pattern Anal. Appl.* **2003**, *6*, 245–256.

34. Tomek, I. Two modifications of CNN. *IEEE Trans. Syst. Man Cybern. B* **1976**, *6*, 769–772.

35. Wang, S.; Yao, X. Multiclass imbalance problems: Analysis and potential solutions. *IEEE Trans. Syst. Man Cybern. B* **2012**, *42*, 1119–1130.

36. Spanish Electricity Price Market Operator. Available online: https://demanda.ree.es/movil/ (accessed on 10 July 2016).

37. Matthews, B.W. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim. Biophys. Acta Protein Struct.* **1975**, *405*, 442–451.