A Novel Feature Design and Stacking Approach for Non-Technical Electricity Loss Detection

Zafer Aydin
Dept. of Computer Engineering
Abdullah Gul University
Kayseri, Turkey
zafer.aydin@agu.edu.tr

V. Cagri Gungor
Dept. of Computer Engineering
Abdullah Gul University
Kayseri, Turkey
cagri.gungor@agu.edu.tr

Abstract—**Non-technical electricity losses continue to jeopardize economic and social well-being of many countries. In this work, we develop machine learning classifiers that can identify anomalous electricity consumption in Turkey. Starting from weekly electricity usage data, we develop new features that capture statistical and frequency domain characteristics of the customers and their consumption patterns. We analyze the effect of reducing number of feature descriptors through dimensionality reduction and feature selection techniques. To overcome the class imbalance problem, we implement several ensemble methods and compare their prediction accuracy to those of the standard classifiers. The proposed features and combining strengths of different classifiers bring significant improvements on performance metrics, which is demonstrated through detailed simulations on shopping mall sector. We anticipate that advances in this field will contribute to the economies considerably.**

Keywords— Non-technical electricity loss detection; fraud detection; anomaly detection; feature selection; machine learning; ensemble classifiers

I. INTRODUCTION

Energy has an important role on the economic and social development of a country. It can appear in different forms but the most important one is the electrical energy. The survival of a society both in economic and social terms depends highly upon uninterrupted supply of electricity. In this regard, effective measures should be taken for generation, transmission and distribution of electric power while minimizing losses as much as possible.

When supplying electricity to final consumers, electricity losses refer to the amounts of electricity introduced to the transmission and distribution grids that are not paid for by users. There can be two types of losses in an electrical system: technical and non-technical. A technical loss is originated naturally due to efficiency of various system components such as production, transmission and distribution power lines, transformers and measurement systems. On the other hand, a non-technical loss (NTL) is independently originated from external factors, such as unpaid bills, missing-wrong meter readings or fraud use. Among those, electricity thefts (i.e., frauds) are the most commonly encountered and can potentially be a burden on the economies of developing countries. Therefore, preventing such non-technical losses should be a high priority [1].

Based on recent studies, it has been observed that welfare, culture, and development level of a country affect the power fraud rate considerably [2]. For example, although the power fraud use is only around 4% in developed countries, in developing countries such as Turkey, the electricity theft rate is around 15.8%, which is more than twice the median of the OECD countries [2]. When the electricity theft is analyzed with respect to various distribution companies, the theft rates vary from 4% to 64.2% depending on the geographical location of the customers (i.e., whether the people live in developed or non-developed regions) [2].

To prevent non-technical losses (NTLs) when there is denunciation or suspicion, electricity distribution company's field teams visit subscribers periodically to investigate whether there is a fraudulent behavior. This is labor intensive and expensive considering the fact that there are over 32.5 million subscribers in Turkey [3]. Although smart meters can detect interventions, such as deactivations, they cannot prevent NTLs completely due to many other forms of electricity thefts. Hence, cost-effective solutions, which can automatically analyze data gathered from the meters and identify fraud use are required.

To this end, power fraud detection and prevention in smart grid is a popular research field. To date, various methods have been developed to detect power fraud in smart grid while cheaters continue to find alternative ways to overcome these methods [4-6] . It should be emphasized that there is no general theoretical methodology that can detect all types of fraudulent behavior. For this reason, computational methods that can discover and learn the patterns characteristic of anomalous data samples and that can discriminate abnormal usage profiles from normal ones are required. To achieve this, several machine learning techniques have been developed [7-21]. Among those supervised learning requires the presence of labeled data [4-10,13]. When label information is not available, unsupervised learning techniques can be used to detect anomalous behavior [4-6, 14-16]. When both labeled and unlabeled examples are present, then semi-supervised learning techniques can be employed [16]. These machine learning approaches provide satisfactory results, especially when they are employed with attribute selection methods [13]. In addition to the aforementioned approaches there are also techniques for specific applications such as those targeted for high-dimensional data [17], time-series data [18-20] and those that

978-1-5386-4291-7/18/$31.00 ©2018 IEEE

employ ensemble learning [21], which are effective for class-imbalance problems.

Despite the plethora of methods proposed for non-technical loss detection, there is limited literature on ensemble classification and to the best of our knowledge there is no work that combines different classifiers in a two-stage ensemble technique known as stacking. In this paper, we propose a stacking ensemble method to detect non-technical losses in Turkish electric utility market. Starting from weekly consumption values we construct a feature set that contains: statistical parameters and frequency domain parameters derived from Fourier transform. We also consider selecting the most informative features using CFS subset evaluator and apply several machine learning classifiers (i.e. base learners) trained on historical data to detect whether any outliers (such as electricity thefts) are present or not. In the last step, a meta-learner makes the final decision based on the predictions made by the base learners. Our method is effective in the sense that it is able to recognize different types of fraudulent behavior with high accuracy using a reduced set of parameters.

II. Materials and Methods

A. Data Collection, Preprocessing and Labeling

In this paper, we performed model evaluation for the shopping mall sector in Turkey on two datasets provided by the distribution company BEDAS. The raw data of the first dataset contains meter readings from 157 shopping malls and the second set contains data from 144 shopping malls. For each shopping mall in the first dataset there is a set of 169 meter values collected hourly within one week. In the second dataset, a set of 1489 meter values is collected for each subscriber every half an hour for January 2015. We first converted raw data to electricity consumption values by computing the difference between readings at successive time points. This yielded a set of 168 values (i.e. features) for each subscriber in the first dataset. We excluded the last three days of January 2015 from the second dataset and selected weekly consumption data for 144 shopping malls (4 weekly data for each mall, a total of 576 weekly data), which contained 336 consumption values (i.e. features) per week since data is collected every half an hour. Figure 1 shows a typical consumption behavior of a normal subscriber in the first dataset.

Here, it is important to note that, meter readings should be monotonically increasing with respect to time. However, for some subscribers this is not the case and negative values occur in consumption data. Additionally, some subscribers can have missing values. As subscribers with missing data or negative consumption values can be identified easily by a preprocessing filter, we removed those samples from our datasets and obtained the first dataset of 123 weekly profiles (i.e. data samples) and the second dataset of 302 weekly profiles.

In the next step, domain experts from distribution company (BEDAS), engineers in Alcatel, and Dr. Aydin labeled subscribers' weekly consumption vectors as normal or abnormal (representing fraud or theft) by investigating the consumption plots. We expect a typical normal subscriber that is open through all days of the week to have a pattern similar to a square wave (as exemplified in Figure 1). However, there could be different types of fraudulent behavior as shown in Figure 2. For instance, the subscriber in Figure 2(a) definitely deviates from a square wave and is labeled as abnormal.

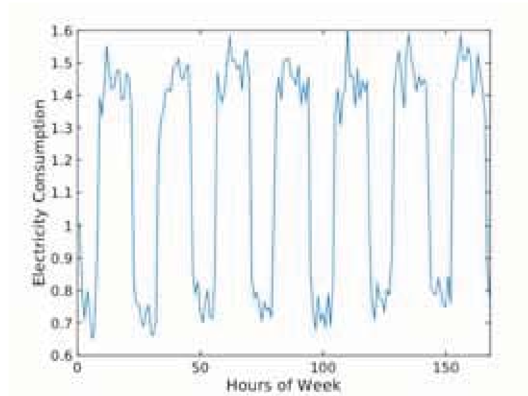
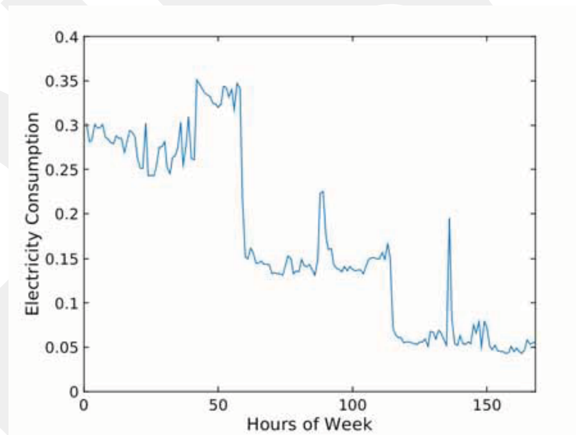
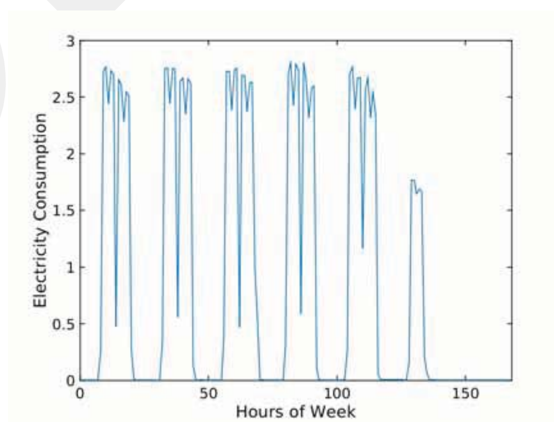


Figure 1. Weekly electricity consumption of a shopping mall in Turkey (a normal consumption case)



(a)



(b)

Figure 2. Different types of anomalous behaviors: (a) remarkable deviation from the square wave pattern (b) sudden reductions during lunch time and reduction to zero during night times.

Figure 2(b) hits to zero during night-time, which is also problematic because it is not typical for a shopping mall to have zero consumption due to refrigerators and lightning equipment working continuously. As a result of this procedure, the numbers of abnormal and normal labels are obtained as in Table 1.

Dataset	# normal labels	# abnormal labels
Shopping mall 1	114	9
Shopping mall 2	287	15
Shopping mall combined	401	24
Shopping Mall Combined (Train)	201	12
Shopping Mall Combined (Test)	200	12

Table 1: The distribution of normal (negative) and abnormal (positive) labels for weekly consumption profiles in shopping mall datasets.

B. Feature Extraction

We considered designing new features because of the following reasons: (1) the number of data samples is less than the number of features and hence reducing the number of dimensions can potentially be useful for model training and testing, (2) to be able to characterize different types of fraudulent behavior, (3) the number of abnormal cases is typically small in fraud detection problems as compared to the number of normal cases (class imbalance problem). After a thorough analysis, we derived the following set of sixty-six (66) features:

1) Statistical features: The median, minimum, maximum and variance of consumption values are calculated at different time intervals. These values can summarize the general consumer behavior. We also compute an outlier parameter to detect deviations at individual time points. The total number of statistical features is equal to 55. These are explained below.

a) Features 1-4: We take median, min, max, and variance of weekly consumption values as the first four features.

b) Features 5-32: We divide the weekly data into seven days and compute median, min, max, and variance of consumption values within each day of the week.

c) Feature 33-53: We divide each day into three time periods as daytime (T1 = 06:00 - 17:00), peak load (T2 = 18:00 - 22:00), and night (T3 = 23:00 - 05:00). We then compute the median value in each time period. These features are based on a prior domain knowledge, which states that abnormal use can be at certain time periods of the day.

d) Feature 54 (outlier parameter): For certain subscribers, the overall weekly consumption data has a normal pattern however there can be significant deviations at individual time points. To detect such variations, we use the 1.5 IQR rule, which is based on quantile analysis [22]. For this purpose, we applied the 1.5 IQR rule to each consumption value (i.e. feature column) separately using data samples from all subscribers. If a consumption value is smaller than $Q_1 - 1.5 \times IQR$ or greater than $Q_3 + 1.5 \times IQR$ for any feature column, we set the outlier parameter to 1 otherwise it is set to 0.

e) Feature 55 (scaled signal energy): We computed a normalized version of the consumption signal to detect abnormal cases where weekly data has completely high or low values. This parameter is calculated as follows:

$$E = \frac{1}{N * \max(\mathbf{x})} \sum_{i=1}^{N} x_i^2 = \frac{||\mathbf{x}||^2}{N * \max(\mathbf{x})} \tag{1}$$

where $\mathbf{x}$ is the weekly consumption vector and N is the number of time samples i.e. $N = 168$ for hourly consumption data and 336 for half-an-hour sampled data.

2) Frequency domain features: There could be periodic signals in a normal weekly meter data. For example, a normal shopping mall consumption profile might resemble a square wave if the electricity usage follows such type of a regular pattern. To be able to capture such periodicities (and deviations from that) certain frequency components can be added to the feature vector. In this paper, we use the Fourier transform (i.e. FFT) to compute frequency domain features [23]. The number of features in this category is 11.

a) Features 56-66: We compute the Fast Fourier Transform (FFT) of 168 consumption values (for the first dataset) and select the absolute values of the 8, 15, 22, 29, 36, 43, 50, 57, 64, 71, and 78th samples. The 8th sample corresponds to the signal component that has a period of 24 hours (i.e. repeating daily patterns) and the rest represent its harmonic components.

All these samples are weak or zero for subscribers with abnormal consumption patterns. When we analyze the FFT signals of all the subscribers, we observe that typically the FFT values of normal subscribers take higher values than those with abnormal at the selected frequencies except for a small number of abnormal cases that also contain periodic signals. We repeated the FFT computations for the second dataset which contained $N = 336$ features in each week and selected the FFT samples indexed as $14a + 1, a = 1, 2, 3, ...$ because this dataset contains samples collected every half an hour.

C. Scaling features (Normalization): We applied min-max normalization [23] to transform the data values to the interval [0, 1]. We considered scaling columns (i.e. features) separately and compared the results to the case where no normalization is applied.

D. Feature Selection: We employed the CFS subset evaluator with Best-First search strategy to select a subset of features that are most important for the classification problem [24].

E. Classification Methods: We employ the following classifiers to identify abnormal and normal consumptions: k-Nearest Neighbor (k-NN), logistic regression, decision tree, random forest, RBF network, Bayes net, Adaboost and stacking [23-25]. For each method, the input consists of the features described in Section II-B and the output is a class label where 1 represents the abnormal cases (positives) and 0 denotes the normal consumption behavior (negatives). The WEKA software is used [24] to implement the classifiers and to optimize hyper-parameters of the models where applicable. For the rest of the parameters the default settings are used. The *k*-NN classifier is implemented by employing the IBk method in WEKA. The following scenarios are tested for *k*-NN: number of nearest neighbors is set to 3 and the number of nearest neighbors is optimized by the –X option in WEKA by performing 10-fold cross-validation experiment on each training set of leave-one-out cross-validation. For decision tree, the J48 algorithm in WEKA (a successor of C4.5) is employed under default parameters, in which the confidence threshold for pruning is set to 0.25 and the minimum number of instances per leaf is set to 2. Two possibilities are considered for the number of trees parameter in random forest, number of clusters parameter in rbf network and number of iterations parameter in Adaboost. First these hyper-parameters are set to a fixed value (number of trees to 100, number of clusters to 15 and number of iterations to 10) and then model training and evaluation is performed accordingly. Second, these parameters are optimized by performing an internal 10-fold cross-validation experiment on each training set of the outer leave-one-out cross-validation. The following values are considered for the number of trees parameter during the optimization step: 5, 10, 25, 50, 75, 100, 150, 200, 250, 300. Furthermore, in random forest each tree is constructed by choosing 7 random features and all the other settings are left in their default values. For the number of clusters parameter in rbf network the following values are considered during the optimization step: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25. To optimize the number of iterations parameter in AdaBoost, the following values are considered: 5, 10, 25, 50, 75, 100, 150, 200. For this ensemble method, decision stump is selected as the base learner.

Stacking is another ensemble learner, which combines the classification decision of several base learners by a meta-learner [24]. Different from bagging and boosting, stacking can employ different types of classifiers as base learners. In this paper, we implemented the following variations of the stacking method:

a) Stacking 1: The base learners are selected as logistic regression, random forest and *k* nearest neighbor with the number of trees set to 100 and *k* to 3.

b) Stacking 2: The base learners are selected as logistic regression, *k* nearest neighbor with *k* set to 1.

In each stacking method, we used the logistic regression classifier as the meta-learner. Stacking 1 is used in leave-one-out cross-validation experiments performed on the first dataset. Stacking 2 is used for train/test experiments on the combined shopping mall dataset.

F. Accuracy Measures: We used the following measures to assess the prediction accuracy of the classifiers: TP rate (recall or sensitivity), FP rate, precision, overall accuracy and F-measure [22], where positives correspond to abnormal behavior and negatives represent normal consumption. Among these measures, we are particularly interested in the F-measure because it includes contributions from both precision and recall terms and provide a better assessment of the prediction accuracy for class-imbalance problems, such as outlier detection.

III. Results and Discussion

We performed a leave-one-out cross-validation experiment on the first dataset to evaluate the prediction accuracy of classification methods. We then combined the first and the second shopping mall datasets and randomly split into a training and a test dataset, which are used to train the models and evaluate their ability to generalize on an independent test data. The number of data samples in each dataset is provided in Table 1.

A. The Effect of Feature Set

We first tested the accuracy of different feature sets using a k-NN (k = 3) on the first mall dataset. Table 2 shows the results of leave-one-out cross-validation experiments without any data normalization for: (1) the original feature set (168 attributes), (2) statistical and frequency domain features (66 attributes), (3) statistical and frequency domain features followed by feature selection. We used the CFS subset evaluator and BestFirst search strategy for feature selection [24]. When we analyze the accuracy measures of k-NN and logistic regression the best performing feature combination is found to be the one that employs statistical and frequency domain features together with feature selection.

Feature Set	TPR	FPR	Precision	Overall	FM
Original (168)	11.10	1.80	33.30	91.87	16.70
Statistical, frequency (66)	**77.80**	**0.00**	**100.00**	**98.37**	**87.50**
Statistical, frequency, CFS	**77.80**	**0.00**	**100.00**	**98.37**	**87.50**

Table 2: The accuracy of k-NN classifier (k = 3) for different feature sets on the first dataset. Numbers in parenthesis represent the number of features used. TPR is the true positive rate, FPR is the false positive rate, and FM is the F-measure. A leave-one-out cross-validation experiment is performed for each case.

According to Table 2 there is a significant increase in all accuracy measures when new features are used as compared to the original set with 168 attributes. For instance, the F-score improves by 70.8% and the overall accuracy by 6.5%. These results demonstrate that designing features by considering characteristics of the problem domain can be beneficial for the performance of a machine learning classifier.

Because feature selection is applied separately for each training set of the leave-one-out cross-validation it is possible to have slight variations on the resulting feature set in each fold of cross-validation. To further understand which features have the highest importance, we derived histograms showing the number of times each feature is selected through iterations of cross-validation experiment. Table 3 summarizes the non-zero frequency values for features selected by the CFS method (starting from the 66 features) on the first dataset. Based on these results, it is observed that a small subset of features is sufficient for accurate classification. Among the 66 statistical and frequency domain features the outlier parameter is selected the most followed by the remaining features.

Feature	Selection Frequency
Outlier Parameter (feature 54)	122
FFT feature 1 (feature 56)	101
Median of day 5 (feature 21)	100
Signal energy (feature 55)	20
Variance of day 6 (feature 28)	1
Variance of day 7 (feature 32)	1

Table 3: Selection frequency of features in training sets of the leave-one-out cross-validation on shopping mall dataset.

B. *Comparison of Classifiers*

In this section we first analyzed the accuracy of several classifiers by performing leave-one-out cross-validation experiments on the first dataset. We performed feature selection on each training set separately using the CFS subset evaluator and Bestfirst search strategy. Furthermore, we considered optimizing the hyper-parameters of the models as described in Section II-E. Table 4 demonstrates the accuracy measures for the first dataset, respectively.

Method	TPR	FPR	Precision	Overall	FM
k-NN (k=3)	77.78	**0.0**	**100.00**	**98.37**	87.50
k-NN (k opt)	77.78	**0.0**	**100.00**	**98.37**	87.50
DT (J48)	55.56	**0.0**	**100.0**	96.75	71.43
BayesNet	55.56	0.88	83.33	95.93	66.67
AdaboostM1 (#it 10)	66.67	0.88	85.71	96.75	75.00
AdaboostM1 (#it opt)	66.67	0.88	85.71	96.75	75.00
Logistic Regression	77.78	**0.0**	**100.00**	**98.37**	87.50
Random forest (#trees 100)	77.78	0.88	87.50	97.56	82.35
Random forest (#trees opt)	77.78	0.88	87.50	97.56	82.35
RBF Network (#clust 15)	77.78	0.88	87.50	97.56	82.35
RBF Network (#clust opt)	77.78	**0.0**	**100.00**	**98.37**	87.50
Stacking 1	**88.89**	0.88	88.89	**98.37**	**88.89**

Table 4: Accuracy of classifiers on the shopping mall dataset. A leave-one-out cross-validation is performed on a dataset of 123 subscribers. Feature selection is applied using CFS subset evaluator on each training set. DT is decision tree, #it represents the number of iterations, #clust denotes the number of clusters, TPR is the true positive rate, FPR is the false positive rate, and FM is the F-measure.

For the shopping mall dataset the best F-measure and overall accuracy is obtained by the Stacking 1 method, which combines *k*-NN (*k*=3), random forest (n trees = 100) and logistic regression classifiers. When we compare classifiers, we can state that stacking approaches provide the best F-measure and overall cross-validation accuracy. This is in agreement with the fact that ensemble methods are favorable for class-imbalance problems. Furthermore, optimizing the models' hyper-parameters and performing feature selection also contributed positively on the accuracy of the methods.

In addition to the cross-validation experiment we also performed a second type of experiment in which we trained the models on the training set portion of the combined shopping mall data and evaluated the accuracy on the test set (see Table 1). In this evaluation we used the set of 66 features as descriptors and normalized the features to [0-1]. Table 5 shows the results of various classification methods, in which the best performing methods are obtained as k-NN and Stacking 2. These results demonstrate that it is possible to get reasonable accuracy on independent test sets as well and the high accuracy values obtained in cross-validation experiments are not only due to overfitting. We also considered performing feature selection using the CFS subset evaluator however this did not provide any improvements. This shows that the features derived in this work provide useful and complementary information capturing the characteristics of the fraudulent behavior.

Method	TPR	FPR	Precision	Overall	FM
k-NN (k=3)	**58.30**	**0.0**	**100.00**	**97.64**	**73.70**
k-NN (k opt)	**58.30**	**0.0**	**100.00**	**97.64**	**73.70**
DT (J48)	41.70	**0.0**	**100.00**	96.70	58.80
BayesNet	58.30	1.50	70.00	96.22	63.60
AdaboostM1 (#it 10)	33.33	0.50	80.00	95.75	47.10
AdaboostM1 (#it opt)	50.00	**0.0**	**100.00**	97.17	66.70
Logistic Regression	50.00	0.50	85.70	96.70	63.20
Random forest (#trees 100)	33.33	**0.0**	**100.00**	96.22	50.00
Random forest (#trees opt)	25.00	**0.0**	**100.00**	95.75	40.00
RBF Network (#clust 15)	41.70	0.50	83.30	96.22	55.60
RBF Network (#clust opt)	41.70	1.00	71.40	95.75	52.60
Stacking 2	**58.30**	**0.0**	**100.00**	**97.64**	**73.30**

Table 5: Accuracy of classifiers on the test set of the combined shopping mall data. Models are trained on the Shopping mall training set. The 66 features are used without feature selection.

C. The Effect of Normalization

We analyzed the effect of data normalization by performing a leave-one-out cross-validation experiment on the shopping mall dataset. We observed that data normalization provides the same accuracy levels as the features without normalization.

IV. Conclusions

In this study, we developed novel feature sets and classification methods to detect non-technical electricity losses in Turkish electric utility market. Our method is effective in the sense that it is able to recognize different types of fraudulent behavior with high accuracy when the novel and reduced set of parameters are employed together with an ensemble technique called stacking. Fine tuning models further improved the classification accuracy slightly. There are several directions that can be pursued for future research. It is possible to apply the same methods to other sectors, such as hospitals, clothing, schools, manufacturers, etc. Note that each sector might have specific constraints and incorporation of such domain knowledge can be important for achieving good results. Second, other dimensionality reduction techniques can be tested and compared to those presented in this paper. Third, we can utilize specialized techniques for outlier detection such as Mahalanobis distance metric, clustering based classification, one-class SVM, and combine with the existing methods. We expect that all these efforts will reduce the costs imposed by fraudulent electricity usage on the economies and enable better planning of resources in the energy sector.

Acknowledgment

This research was supported by Alcatel-Lucent Teletas, Istanbul, Turkey and TUBITAK TEYDEB with Project No: 3120959. The authors would like to thank Tolga Goze and Ihsan Ozcan from Alcatel-Lucent Teletas for providing the datasets and Cengiz Gazeloglu for his useful discussions. The work of V. C. Gungor was also supported by the Turkish National Academy of Sciences Distinguished Young Scientist Award Program (TUBA-GEBIP) under Grand no. V.G./TUBA-GEBIP/2013-14.

References

[1] Antmann, P. "Reducing technical and non-technical losses in the power sector.", World Bank, Washington, DC., 2009.

[2] Tasdoven, H., Fiedler, B.A., and Garayev V.. "Improving electricity efficiency in Turkey by addressing illegal electricity consumption: A governance approach." *Energy Policy vol* 43, pp 226-234, 2012.

[3] Elektrikport.com, 'Türkiye Elektrik Dağıtım Bölgeleri 1', 2014. [Online]. Available: http://www.elektrikport.com/teknik-kutuphane/turkiye-elektrik-dagitim-bolgeleri-1/12004#ad-image-0. [Accessed: 12- Dec- 2017].

[4] Chandola, V, Banerjee A, and Kumar V. "Anomaly detection: A survey." ACM computing surveys (CSUR) 41.3, 2009.

[5] Hodge, V and Austin, J. "A survey of outlier detection methodologies." Artificial intelligence review 22.2, pp 85-126, 2004.

[6] Kou, Y., Lu, C. T., Sirwongwattana, S., and Huang, Y. P. "Survey of fraud detection techniques." Networking, sensing and control, IEEE international conference on (Vol. 2, pp. 749-754). 2004.

[7] Nagi, J., Yap, K. S., Tiong, S. K., Ahmed, S. K., and Mohamad, M. "Nontechnical loss detection for metered customers in power utility using support vector machines." IEEE transactions on Power Delivery vol 25(2), pp. 1162-1171, 2010.

[8] Ford, V., Siraj, A., and Eberle, W., "Smart grid energy fraud detection using artificial neural networks.", Computational Intelligence Applications in Smart Grid (CIASG), IEEE Symposium on, pp. 1-6., 2014.

[9] Ma, J., and Perkins, S., "Time-series novelty detection using one-class support vector machines.", Proceedings of the International Joint Conference on Neural Networks, Vol. 3, pp. 1741-1745., 2003.

[10] Monedero, I., Biscarri, F., León, C., Guerrero, J. I., Biscarri, J., and Millán, R., "Detection of frauds and other non-technical losses in a power utility using Pearson coefficient, Bayesian networks and decision trees.", International Journal of Electrical Power & Energy Systems, vol 34(1), p.- 90-98, 2012.

[11] Nagi, J., "An intelligent system for detection of non-technical losses in tenaga nasional berhad (tnb) malaysia low voltage distribution network." Doctral Dissertation, Vasa, 2009.

[12] Nizar, A. H., Dong, Z. Y., and Zhang, P., "Detection rules for non technical losses analysis in power utilities.", Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century, pp. 1-8., 2008.

[13] Ramos, C. C., Souza, A. N., Chiachia, G., Falcão, A. X., and Papa, J. P., "A novel algorithm for feature selection using harmony search and its application for non-technical losses detection.", Computers & Electrical Engineering, vol 37(6), pp. 886-894., 2011.

[14] Yamanishi, K., Takeuchi, J. I., Williams, G., and Milne, P., "On-line unsupervised outlier detection using finite mixtures with discounting learning algorithms." Data Mining and Knowledge Discovery, vol 8(3), pp. 275-300, 2004.

[15] Zhang, Y., Meratnia, N., and Havinga, P. J. M., "A taxonomy framework for unsupervised outlier detection techniques for multi-type data sets", Centre for Telematics and Information Technology (No. TR-CTIT-07-79)., University of Twente, 2007.

[16] Vinueza, A., and Grudic, G., "Unsupervised outlier detection and semi-supervised learning.", Technical Report CU-CS-976-04, University of Colorado at Boulder, 2004.

[17] Aggarwal, C. C., and Yu, P. S., "Outlier detection for high dimensional data.", ACM Sigmod Record, Vol. 30, No. 2, pp. 37-46, 2001.

[18] Chen, X. C., Steinhaeuser, K., Boriah, S., Chatterjee, S., and Kumar, V., "Contextual time series change detection.", Proceedings of the 2013 SIAM International Conference on Data Mining, pp. 503-511, 2013.

[19] Gupta, M., Gao, J., Aggarwal, C., and Han, J., "Outlier detection for temporal data.", Synthesis Lectures on Data Mining and Knowledge Discovery, vol 5(1), pp. 1-129, 2014.

[20] Yamanishi, K., and Takeuchi, J. I., "A unifying framework for detecting outliers and change points from non-stationary time series data.", Proceedings of the 8th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 676-681. 2002.

[21] Di Martino, M., Decia, F., Molinelli, J., and Fernández, A., "Improving Electric Fraud Detection using Class Imbalance Strategies." ICPRAM vol 2, pp. 135-141, 2012.

[22] Oppenheim, A.V., Schafer W.R., and Buck J.R. "Discrete-time signal processing." Printice Hall Inc., 1989.

[23] Han, J., Pei, J., and Kamber, M., "Data mining: concepts and techniques.", 3rd Ed., Elsevier, 2011.

[24] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H., "The WEKA data mining software: an update.", ACM SIGKDD explorations newsletter, 11(1), pp. 10-18, 2009.

[25] Bishop, C. M., "Pattern recognition and machine learning.", Springer, 2006.