

# On identifying the irregularities of electricity customer behaviors using soft computing approach

*by* Lawi Armin

---

**Submission date:** 19-Feb-2020 11:51AM (UTC+0700)

**Submission ID:** 1259987125

**File name:** Lawi\_2019\_J.\_Phys.\_\_Conf.\_Ser.\_1341\_042004.pdf (810.93K)

**Word count:** 5316

**Character count:** 26440

**PAPER • OPEN ACCESS**

## On identifying the irregularities of electricity customer behaviors using soft computing approach

10

To cite this article: A Lawi *et al* 2019 *J. Phys.: Conf. Ser.* **1341** 042004View the [article online](#) for updates and enhancements.**IOP | ebooks™**

Bringing you innovative digital publishing with leading voices  
to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of  
every title for free.

## On identifying the irregularities of electricity customer behaviors using soft computing approach

A Lawi<sup>1</sup>, S L Wungo<sup>2</sup> and S Manjang<sup>3</sup>

<sup>1</sup>Department of Computer Science, Universitas Hasanuddin, Makassar, Indonesia

<sup>2</sup>Informatics Engineering, STMIK Kreatindo, Monokwari, Indonesia

<sup>3</sup>Departement of Electrical Engineering, Universitas Hasanuddin, Makassar, Indonesia

<sup>2</sup>Email: supriyadi.la.wungo@gmail.com

**Abstract.** This study aims to implement a soft computing approach in identifying the irregularities of customer behaviour on the use of prepaid electricity pulses. The used methods are Support Vector Machine (SVM), Naïve Bayes (NB), Classification and Regression Tree (CART), and k-Nearest Neighbours (KNN). To evaluate the performance of the classification system, a 10-fold Cross Validation technique is used for the historical data of pre-paid pulse purchase transactions. Validation results are measured using accuracy, precision and recall values. This research shows that all soft computing methods gave good performances in classifying electrical consumption behaviours. CART method has the highest accuracy value of 100% compared with others algorithm. At precision values, KNN and CART methods have the highest precision value among other algorithms that are 99% to 100%. Whereas, the recall values of each method has a high recall value of 100%. Moreover, each method can predict morbidity accurately because the addition of the amount of data testing does not affect the performance of each method.

### 1. Introduction

The electricity needs of the Indonesian people are supplied by electricity companies such as the state-owned electricity company (PLN), the use of electricity by consumers using prepaid electric cards. The use of this card is very efficient because consumers can control the use of their electrical energy [1]. Although convenience has been provided by the electricity company, unfortunately, consumers who are indicated to have stolen electricity still occur in the field. For example, customers can manipulate the wiring pattern on the electricity-meter and connect the cable without going through the electricity-meter [2], [3], and thus the indicator on the electricity-meter does not function so that the electric current continues to pass the electricity meter even though the remaining voucher is up. According to data from the Ministry of Energy and Mineral Resources (MEMR) and PLN statistics, the average depreciation of electricity in Indonesia is around 7% per year and the estimated loss of the national budget resulting from this depreciation is IDR1.5 trillion per year [4], [5]. It means that the supervision and control of electricity usage for customers is still very weak. This specific case happens since manual checking and recording services every month are no longer used by electricity companies. In order to supervise the electricity usage transactions that are indicated as committing theft, the electricity company conducts screening based on the period of the transaction to purchase



2  
Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

prepaid electric cards periodically. However, this solution still has many disadvantages and weakness. For example, if the customer only purchases a voucher with nominal IDR 20,000 for a month usage, then this usage is still considered reasonable although, actually, the customer should purchased for IDR 100,000.

Thus, in order to reduce the problem of electrical energy misuse that has been done by the customer will be proposed filter the nominal amount of voucher purchase transaction based on customer id, tariff per customer, and customer category. The amount of customer data that reaches millions of users will take a long time if identification is done traditionally and the difficulty of identifying unusual customers. Therefore, machine learning will be applied to solve the problem. Machine learning is part of soft computing method. It is a combination of several methodologies that are built to solve problems in everyday life that are difficult to formulate or mathematically modeled. This technique is very efficient for analyzing, modeling, and studying data that is difficult to resolve traditionally so that it can be used to make a decision [6], [7], [8].

## 2. Customer Power Usage Behavior Problems

At present, the supervision system for customers is very rarely done by electricity supply companies. This is due to the fact that companies no longer carry out manual checks and records every month. The filter mechanism is based on the period of the electric card purchase transaction for customers who are indicated to have committed an abuse in the purchase of an electric card. However, this method still has weaknesses and causes a large loss to the company and then takes a long time and a large cost for manual inspection. [9].

Some customer behaviors that are identified as unreasonable are electrical cables connected by customers not through the electricity meter indicator, the wiring pattern is changed on the electricity meter indicator and to reduce disk rotation, the magnet is placed above the electricity meter indicator [2], [3], [10], so that the counter on the electricity meter indicator does not work but the electricity is still functioning even though the credit / voucher indicator on the electricity meter has run out [6].

The consequences of theft of electricity for the company are large power losses and material losses that reach millions or even billions of rupiah, while for consumers power outages often occur in the field, the power shared by other consumers decreases, fires caused by a short circuit [10].

Some related work on theft of electrical energy has been proposed by previous researchers. Depuru, et al., In [10] suggested that electricity theft was detected through energy consumption patterns from customers who indicated theft of electricity. The classification method used is Support Vector Machine (SVM) with the value  $C = 1$ , kernel radial base function (rbf), and probability estimation = 0, and the value of accuracy is 98.4%. Another classification technique proposed using power consumption profiles to detect fraud in electricity consumption, has been proposed in [11]. The Fuzzy C-Means (FCM) method is used to classify abnormal customers from the database based on customer groupings that have the same consumption pattern. Fuzzy C-Means classification will be used 2 metrics namely Sensitivity and Assertiveness. The accuracy of the two metrics is 0.100 and 0.745, respectively. Babu, et al., In [12] the FCM method is used to detect losses on a non-technical basis through a power distribution system by classifying irregular customer consumption profiles in the use of electricity. The accuracy of their Fuzzy-based classification method detects non-technical losses is 80%. Further research in [9] detected and found real-time power theft carried out at each power level in transmission and distribution. A comprehensive top-down scheme method based on Decision Tree (DT) and SVM is proposed. The attribute value will be used as a reference to calculate the energy consumption that has been used by the customer using the DT method. Then the SVM method will accept the results of this calculation as input data to classify abnormal or normal customers in the use of electrical energy. The accuracy of the combination of these two methods is 92.5% and false positive accuracy is very low, 5.12%.

## 3. Relevant Variable Selection

In [6], to determine the variable abuse of electrical energy consumption that has been carried out by the customer can be identified by using the prepaid electricity card transaction history data, as follows:

- *Customer ID*: identify the number of electrical card purchase transactions carried out by customers in a month.
- *Installed Power (kWh)*: identifies the category of electricity meter used by customers based on installed power.
- *Total credit in the voucher*: an indicator of the number of kWh in the electricity meter by customer transactions.
- *Total power consumption*: the difference between the maximum electricity usage (total credit in the voucher) and the minimum (total power consumption) of the total usage in a month.
- *Time of use of vouchers*: time of use of vouchers by customers for a certain period.
- *Status*: the dependent variable (response) that determines the classification of the right or incorrect use of power consumption, the use of customers using vouchers.

The selection of the five predictor variables above (or independent) is considered to represent the attributes of each customer to identify patterns of use of electrical energy based on the historical data of the purchase of electricity cards. Further to support analysis of electrical data and to detect weaknesses of existing systems. The data used is the history of prepaid electric card purchase transactions in a month span consisting of 10,000 customers. Losses in the company can be minimized because every month the company evaluates prepaid electric card purchase transactions made by customers. Therefore, determining the time span is considered very relevant for the process of identifying irregularities [11],[12].

#### 4. Identification of Irregularities

Identification of irregularities is a comparison between two or more objects to find differences between objects based on certain criteria. The criteria of an object can be known by looking at the attribute values of the object. If the attribute value of an object does not meet the criteria when compared to the attribute value of another object that meets the criteria then the object can be identified as an unnatural object. Therefore, the selection of appropriate attributes is needed to determine the imperfection of an object [3].

To detect theft or fraud in the use of electrical energy can be done by identifying deviations that have been made by the customer. Deviations in the case of theft of electricity can be identified by looking at patterns of electricity consumption. Transaction history data usage of electrical energy can be used to view customer behavior patterns. Historical data has been used by large-scale companies to identify patterns of electricity consumption of customers which are indicated to have improper transactions in electricity consumption. Unreasonable transactions carried out by customers if the electricity consumption is above the average of the actual usage target. So, based on transaction history data using electricity, companies can identify fraud or electricity theft committed by customers [13], [14].

#### 18 Classification Algorithm

Based on the results of these studies then the authors will use the Soft Computing method that will be used to identify the irregularities of customer behaviour against the abuse of electrical energy through the purchase transaction history of tokens/vouchers.

##### 15 5.1. Support Vector Machine

Classification techniques Support Vector Machine (SVM) is a classification method that not only can map a set of data in high dimensions but also be able to map data from its original dimensions to other dimensions that are relatively higher using kernel techniques [15]. Support vectors are data from each class closest to the hyperplane and contribute to SVM [16]. Based on the training data vectors  $x_i \in R^n, i = 1, 2, \dots, l$ , in two classes of vectors  $y \in R^l$  as  $y_i \in \{1, -1\}$ , to solve the primal optimization problem a vector support classifier will be used along with the equation:

$$\min_{w,b,\zeta} \frac{1}{2} w^T w + C \sum_{i=1}^l \zeta_i$$

Subject to

$$y_i(w^T \phi(x_i) + b) \geq 1 - \zeta_i$$

$\zeta_i \geq 0, i = 1, \dots, l$ , where  $\phi(x_i)$  is a kernel maps  $x_i$  into the high-dimensional space and  $C > 0$ . Equation  $y_i[w^T \phi(x_i) + b] \geq 1$  consists of two constraints, the first is  $w^T \phi(x_i) + b \geq +1$  if  $y_i = +1$  and the second is  $w^T \phi(x_i) + b \leq -1$  if  $y_i = -1$  where  $\zeta_i$  is an error that can occur when a dataset that is not equally wrong is classified as a slack variable [2]. In fact, it is often found a number of problems that the dataset as training data used to construct SVC models is largely linearly inseparable. To solve the problem of training data that cannot be classified linearly, the RBF kernel will be used to map slack variables to higher dimensions so that classification errors that occur during dataset training can be reduced [2]. The following are the Polynomial Kernel, Linear Kernel, RBF kernel, and Sigmoid kernel used in SVM [17]:

- Polynomial  $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0$
- Linear  $K(x_i, x_j) = x_i^T x_j$
- Gaussian  $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0$
- Sigmoid  $K(x_i, x_j) = \tanh(\gamma \cdot x_i^T x_j + r)$

where  $\gamma$  is the kernel parameter, and  $x_j$  is the data value of the attribute and, and  $x_i$  is support vector. Dividing the training dataset into two parts when classifying to make the best decision boundary is the goal of the kernel [9].

### 5.2. Naive Bayes

Bayesian classification method or so-called Naive Bayes (NB) is a statistical method used to predict the probability of the membership set of a particular class. In the classification of naive Bayes, the attribute value of a class is independent of the value of another class attribute [16]. So this condition is called a conditional independent class. The workings of naive Bayes classification are based on the Bayes theorem. Naive Bayes algorithm as follows [15]:

- Determined D is training data that label class. Each data has a vector of n-dimensional attributes, where n is the number of attributes,  $A_1, A_2, \dots, A_n$ .
- Suppose there are as many m classes,  $C_1, C_2, \dots, C_m$ . Based on the X data, the classifier will predict X including the class that has the highest probable conditional probability on X. That is, the Naive Bayesian classifier will predict that the X data belongs to the  $C_i$  class if and only if  $P(C_i \vee X) > P(C_j \vee X)$ . So maximize  $P(C_i \vee X)$  For Class  $C_i$  having maximum  $P(C_i \vee X)$  is called the maximum posterior hypothesis. Theorem Bayes is  $P(C_i \vee X) = P \frac{P(X \vee C_i)P(C_i)}{P(X)}$ .
- For  $P(X)$  is constant for all classes. Only  $P(X \vee C_i)P(C_i)$  is maximized. If the prob priority class is unknown, then it is assumed that the class is similar,  $P(C_1) = P(C_2) = \dots = P(C_m)$ , and then maximizes  $P(X \vee C_i)$ . If not, maximize  $P(X \vee C_i)P(C_i)$ . Please note the prior class probability can be estimated with  $P(C_i) = |C_{i,D}|/|D|$  where  $|C_{i,D}|$  is the number of training class  $C_i$  training data in D.
- The more attributes used in the dataset the higher the computation to compute  $P(X \vee C_i)$ , to anticipate it Naive Bayes makes independent of the class conditions (no dependency relationship between attributes). Here's the equation:  $P(X \vee C_i) = \prod_{k=1}^n P(x_k \vee C_i) = P(x_1 \vee C_i) \times P(x_2 \vee C_i) \times \dots \times P(x_n \vee C_i)$  where  $x_k$  is the attribute value  $A_k$  in data. There are several things to consider in calculating  $P(X \vee C_i)$  that is: If the attribute  $A_k$  is categorical, then  $P(x_k \vee C_i)$  is the amount of data from class  $C_i$  in D has the value of  $x_k$  for  $A_k$ , divided by  $|C_{i,D}|$ , the amount of data of class  $C_i$  di D. If the  $A_k$  attribute is of continuous value, it will be distributed to Gaussian with  $\sigma$  standard deviation and mean  $\mu$ , defined as follows:

$$g(x|\mu\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (1)$$

so that

$$P(x_k \vee C_i) = g(x_k, \mu C_i, \sigma C_i)$$

then calculate  $\sigma C_i$  and  $\mu C_i$ , ie standard deviation and mean, from attribute values  $A_k$  for training data of class  $C_i$ . After that, enter the value of these two quantities into Eq. (1), together with  $x_k$ , for estimating the value of  $P(x_k \vee C_i)$ .

- To predict the class X label, the probability function  $P(X \vee C_i | C_i)$  will be evaluated for each class  $C_i$ . Classifier will predict label from X data into  $C_i$  class if and only if  $P(X \vee C_i)P(C_i) > P(X \vee C_j)P(C_j)$  for  $1 \leq j \leq m, j \neq i$ .

### 5.3. Classification and Regression Tree

Classification and Regression Tree (CART) is a classification method that uses historical data to construct decision trees that will then be used to classify new data [18]. The classification tree is built according to the splitting rule. The splitting rule will divide learning sample into small parts according to maximum homogeneity using Gini Index [19]. Here's the equation:

$$Gini(D) = 1 - \sum_{i=1}^m p_i^2$$

Where  $m$  is the number of class attributes in data D,  $p_i$  is the probability in data D that has the class attribute  $C_i$  and is calculated by the following equation

$$\frac{|C_{i,D}|}{|D|}$$

Classification tree has level or a lot of levels that can affect the performance of classification. To overcome this is used Optimization by minimum number of points. It of points is one of the algorithms for pruning or classification of a tree [15].

### 5.4. K-Nearest Neighbor

In [18] k-Nearest Neighbours (kNN) is a classification method used to find a similarity of a group of objects  $k$  in the training set closest to the test data object and based on the assignment of a label dominated by a particular class. Selection of  $k$  values affect the performance of kNN. If  $k$  value is too small, it can reduce noise data. On the other hand, if the value of  $k$  is too large, it can increase the amount of noise data. Because the neighbouring data from other classes are selected. Square distance is a good approximation because it is less influenced by the selection of  $k$  values. The weighting of each object is determined by its distance. Here is the equation of square distance:

$$w_i = \frac{1}{d(y, z)^2}$$

Where  $w_i$  denotes the weight of the  $x_i$  data which is the nearest neighbour and  $d(y, z)$  denotes the distance (inconsistency) between the  $y_i$  data with the  $z$  test data. then  $w_i$  value can be used to calculate the selection of predicted class. Here's the Distance-Weighted Voting equation:

$$c_z = \operatorname{argmax}_{v \in L} \sum_{y \in N} w_i \times I(v = \operatorname{class}(c_y))$$

where  $x_k$  and  $y_k$  are attributes to  $k$  of  $x$  and  $y$  respectively. Furthermore, the distance function used to measure data  $x$  and  $z$  with  $n$  attribute is the Euclidean function. Here is the Euclidean distance equation:

$$d(x, y) = \sqrt{\sum_{k=1}^n (x_k - y_k)^2}$$

## 6. Performance Evaluation

The classification model will be evaluated using the 10-Fold Cross-Validation. Where the dataset will be divided into 10 parts of the same size. Then 9 parts of the subset will be used as training data and

the rest as testing data. Furthermore, to measure the performance of the classification results, the confusion matrix will be used as shown in table 1 [20].

**Table 1.** Confusion matrix.

Actual class	Actual class predicted	
	Normal	Abnormal
Normal	TP	FN
Abnormal	FP	TN

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

Where TP: is the amount of normal transaction data that is correctly classified in the normal class, FN: is the amount of normal transaction data that is classified incorrectly in an abnormal class, TN: is the number of abnormal transaction data that is correctly classified in an abnormal class, FP : the number of abnormal transaction data that is classified incorrectly in the normal class.

## 7. Experiment

In this study, the dataset used is the electricity card purchase transaction data for the eastern region of the Makassar city. The history of prepaid electric card purchase transaction data has 10,000 customer data that are of real-type data and consist of 4 attributes, which are divided into 2 classes namely normal transactions and abnormal transactions of 7,783 and 2,217 customers, respectively. Furthermore, the dataset will be processed using programming language python version 3.5 which is installed in the Ubuntu operating system version 18.04 and the type of processor used is Intel Core i5-460 and 4GB of RAM. The classification method to be used in this research is soft computing which consists of SVM, NB, CART, and KNN.

**Table 2:** Acc, prec, rec of SVM method

Method	Percentage of data testing (%)					
	20%	33%	45%	55%	67%	80%
Acc	0.99	0.99	0.99	0.99	0.99	0.99
Prec	0.99	0.98	0.98	0.98	0.98	0.97
Rec	1.00	1.00	1.00	1.00	1.00	1.00

Table 2 shows that in spite of an increase in the amount of test data used, the SVM method using the Gaussian kernel and the alpha 0.5 value can predict the improper purchase of electricity with the same accuracy of 99%. But when viewed from the results of its precision, the increase in test data precision precisely resulted in the decline. While the recall results remain the same in the 100%. This indicates that when the test data is added, the False Positive result has increased while the False Negative result remains 0. False positive in this research is the amount of unnatural data which is predicted as fair. While False Negative is the amount of reasonable data that is predicted to the class is not fair. Because in every addition of test data only precision is decreased while accuracy and recall remain, it can be said that svm method can predict morbidity accurately, but increasing the amount of data causes more and more reasonable data to be classified incorrectly to the unnatural class.

**Table 3.** Acc, prec, rec of NB method

Method	Percentage of data testing (%)					
	20%	33%	45%	55%	67%	80%
Acc	0.71	0.71	0.71	0.72	0.71	0.70
NB Prec	0.42	0.43	0.44	0.44	0.45	0.45
Rec	0.95	0.96	0.96	0.96	0.95	0.91

Table 3 shows that with the addition of classification data of the NB method using the GaussianNB kernel, it can predict the unreliable purchase of electricity with the accuracy of 0.70% up to 72%. Likewise, the precision value changed from 42% up to 0.45% while the value of recall is 91% until 96%. This means that even with the addition of test data, the value of False Positive has decreased while the value of False Negative has increased. Because in each additional test data the value of precision, accuracy and recall has decreased, it can be said that the NB method cannot predict morbidity accurately, so the addition of the amount of testing data causes more and more reasonable data to be grouped into unnatural classes.

**Table 4.** Acc, prec, rec of CART method

Method	Percentage of data testing (%)					
	20%	33%	45%	55%	67%	80%
Acc	1.00	1.00	1.00	1.00	1.00	1.00
CART Prec	1.00	1.00	1.00	1.00	1.00	1.00
Rec	1.00	1.00	1.00	1.00	1.00	1.00

Table 4 shows that the addition of test data does not affect the accuracy of classification of CART method by using Gini as a measurement for split quality and split strategy used by best split to predict the improper purchase of electricity with the accuracy level of 1.00%. Likewise, the precision and recall value did not change at 1.00%. This shows that with the addition of test data, the value of False Positive and False Negative remain 0. Because in each addition of precision value test data, accuracy and recall remain, it can be said that Naive Bayes method can predict morbidity accurately.

**Table 5.** Acc, prec, rec of KNN method

Method	Percentage of data testing (%)					
	20%	33%	45%	55%	67%	80%
Acc	0.99	1.00	0.99	0.99	0.99	0.99
KNN Prec	0.99	1.00	0.99	0.99	1.00	1.00
Rec	1.00	1.00	1.00	0.99	0.99	0.99

Based on the experiments in table 5, it is shown that with the increase of test data **does not affect the performance of the classification of KNN method** to find the nearest neighbor distance by using Euclidean Distance,  $k = 5$  value and weight value between attributes = 1, in predicting the improper purchase of electricity with accuracy of 0.99%.

But when viewed for its precision value, the increase of test data precisely causes precision value increases. Instead, the recall results decreased as test data increased. This means that with increasing test data, False Positive value has increased while False Negative value decreases. Because in each addition of precision value test data has increased while recall value has decreased, it can be said that

KNN method cannot predict morbidity accurately because the addition of the amount of data causes more and more unnatural data to be classified improperly to a reasonable class. Comparison of the classification method can be seen in table 8 as follows:

**Table 8.** Comparison of algorithms

Method		Percentage of data testing (%)					
		20%	33%	45%	55%	67%	80%
SVM	Acc	0.99	0.99	0.99	0.99	0.99	0.99
	Prec	0.99	0.98	0.98	0.98	0.98	0.97
	Rec	1.00	1.00	1.00	1.00	1.00	1.00
NB	Acc	0.71	0.71	0.71	0.72	0.71	0.70
	Prec	0.42	0.43	0.44	0.44	0.45	0.45
	Rec	0.95	0.96	0.96	0.96	0.95	0.91
CART	Acc	1.00	1.00	1.00	1.00	1.00	1.00
	Prec	1.00	1.00	1.00	1.00	1.00	1.00
	Rec	1.00	1.00	1.00	1.00	1.00	1.00
KNN	Acc	0.99	1.00	0.99	0.99	0.99	0.99
	Prec	0.99	1.00	0.99	0.99	1.00	1.00
	Rec	1.00	1.00	1.00	0.99	0.99	0.99

Table 8 shows that by adding test data, CART method have the highest accuracy value compared to other algorithms. This suggests that with the addition of test data, the CART method can classify reasonable data into reasonable and unnatural classes to the class unreasonably accurately. At precision values, the KNN and CART methods have the highest precision values among other algorithms, meaning that KNN and CART methods can predict morbidity accurately as the addition of the amount of data causes less and less reasonable data to be classified incorrectly into the class. While in the recall value, each method has a high recall value because every method can predict morbidity accurately because the addition of the amount of data does not affect the performance of each method. Based on the results of the study, the authors will use the Soft Computing method that will be used to identify the irregularities of customer behavior against the abuse of electrical energy through the purchase transaction history of tokens/vouchers.

### 8. Conclusion

Implementation of the soft computing approach in this study, each method Support Vector Machine (SVM), Naive Bayes, Classification and Regression Tree (CART), and K-Nearest Neighbors (KNN), it appears that for this study, the method with the highest accuracy is obtained from the CART method that is equal to 100%, while the SVM method and KNN are 99%. Meanwhile, the Naive Bayes method has the lowest accuracy among all the methods that have been implemented in this study. But to determine which method with the best accuracy among them still needs further research by using more simulation data. Based on the results obtained that the accuracy of the soft computing approach in classifying customer behavior using prepaid pulses through payment history has a good level of accuracy of each method.

## References

- [1] Jha, I. S., Sen, S., and Agarwal, V., 2014. Advanced metering infrastructure analytics—A Case Study. In *2014 Eighteenth National Power Systems Conference (NPSC)* (pp. 1-6). IEEE.
- [2] Dangar, B., and Joshi, S. K., 2015. Electricity theft detection techniques for metered power consumer in GUVNL, GUJARAT, INDIA. In *2015 Clemson University Power Systems Conference (PSC)* (pp. 1-6). IEEE.
- [3] Ford, V., Siraj, A., and Eberle, W., 2014. Smart grid energy fraud detection using artificial neural networks. In *2014 IEEE Symposium on Computational Intelligence Applications in Smart Grid (CIASG)* (pp. 1-6). IEEE.
- [4] Sekretariat Perusahaan PT. PLN., 2015. *Statistik PLN 2014*, no. 2701, p. iv, 2015.
- [5] K. ESDM., 2014. Statistik ketenagalistrikan 2014., *Stat. Ketenagalistrikan*, p. 5, 2014.
- [6] Lawi, A., La Wungo, S., and Manjang, S., 2017. Identifying irregularity electricity usage of customer behaviors using logistic regression and linear discriminant analysis. In *2017 3rd International Conference on Science in Information Technology (ICSITech)* (pp. 552-557). IEEE.
- [7] Tettamanzi, A., and Tomassini, M., 2013. *Soft computing: integrating evolutionary, neural, and fuzzy systems*. Springer Science & Business Media.
- [8] Zhang, J., Tao, C., and Wang, P., 2016. A review of soft computing based on deep learning. In *2016 International Conference on Industrial Informatics-Computing Technology, Intelligent Technology, Industrial Information Integration (ICICII)* (pp. 136-144). IEEE.
- [9] Jindal, A., Dua, A., Kaur, K., Singh, M., Kumar, N., and Mishra, S., 2016. Decision tree and SVM-based data analytics for theft detection in smart grid. *IEEE Transactions on Industrial Informatics*, 12 (3), 1005-1016.
- [10] Depuru, S. S. S. R., Wang, L., and Devabhaktuni, V., 2012. Enhanced encoding technique for identifying abnormal energy usage pattern. In *2012 North American Power Symposium (NAPS)* (pp. 1-6). IEEE.
- [11] Angelos, E. W. S., Saavedra, O. R., Cortes, O. A. C., dan de Souza, A. N., 2011. Detection and identification of abnormalities in customer consumptions in power distribution systems. *IEEE Transactions on Power Delivery*, 26 (4), 2436-2442.
- [12] Babu, T. V., Murthy, T. S., and Sivaiah, B., 2013. Detecting unusual customer consumption profiles in power distribution systems—APSPDCL. In *2013 IEEE International Conference on Computational Intelligence and Computing Research* (pp. 1-5). IEEE.
- [13] Cody, C., Ford, V., and Siraj, A., 2015. Decision tree learning for fraud detection in consumer energy consumption. In *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)* (pp. 1175-1179). IEEE.
- [14] Han, S. Y., No, J., Shin, J. H., and Joo, Y., 2016. Conditional abnormality detection based on AMI data mining. *IET Generation, Transmission & Distribution*, 10 (12), 3010-3016.
- [15] Han, J., Kamber, M., & Mining, D. (2001). *Concepts and techniques*. Morgan Kaufmann, vol. 5, 340, 94104-3205.
- [16] Roiger, R. J., 2017. *Data mining: a tutorial-based primer*. Chapman and Hall/CRC.
- [17] Hsu, C. W., Chang, C. C., & Lin, C. J. (2013). A practical guide to support vector classification, 1 (1), 1-16. Retrieved in, 2.
- [18] Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., and Zhou, Z. H., 2008. Top 10 algorithms in data mining. *Knowledge and information systems*, 14 (1), 1-37.
- [19] Timofeev, R., 2004. Classification and regression trees (CART) theory and applications. *Humboldt University, Berlin*. pp. 1-40.
- [20] Fawcett, T., 2006. An introduction to ROC analysis. *Pattern recognition letters*, 27 (8), 861-874.

# On identifying the irregularities of electricity customer behaviors using soft computing approach

## ORIGINALITY REPORT

16%

SIMILARITY INDEX

10%

INTERNET SOURCES

12%

PUBLICATIONS

13%

STUDENT PAPERS

## PRIMARY SOURCES

1	<a href="http://research.aalto.fi">research.aalto.fi</a> Internet Source	2%
2	<a href="http://nadnaps.github.io">nadnaps.github.io</a> Internet Source	1%
3	<a href="http://www.docstoc.com">www.docstoc.com</a> Internet Source	1%
4	<a href="http://www.onlinejournal.in">www.onlinejournal.in</a> Internet Source	1%
5	Joao Antunes, Nuno Neves, Paulo Verissimo. "Reverse Engineering of Protocols from Network Traces", 2011 18th Working Conference on Reverse Engineering, 2011 Publication	1%
6	Submitted to Universiti Sultan Zainal Abidin Student Paper	1%
7	Submitted to Auckland University of Technology Student Paper	1%

8	Internet Source	1%
9	Submitted to University of South Alabama Student Paper	1%
10	digilib.uinsby.ac.id Internet Source	1%
11	Submitted to Chulalongkorn University Student Paper	1%
12	Submitted to Higher Education Commission Pakistan Student Paper	<1%
13	Submitted to Hellenic Open University Student Paper	<1%
14	www.westminster.ac.uk Internet Source	<1%
15	Khaoula Tbarki, Salma Ben Said, Riadh Ksantini, Zied Lachiri. "RBF kernel based SVM classification for landmine detection and discrimination", 2016 International Image Processing, Applications and Systems (IPAS), 2016 Publication	<1%
16	Submitted to Universiti Teknologi MARA Student Paper	<1%

Upendar, J.. "Statistical decision-tree based

17 fault classification scheme for protection of power transmission lines", International Journal of Electrical Power and Energy Systems, 201203  
Publication <1%

---

18 elar.usfeu.ru  
Internet Source <1%

---

19 Submitted to Universiti Teknologi Petronas  
Student Paper <1%

---

20 Submitted to 21390  
Student Paper <1%

---

21 www.mdpi.com  
Internet Source <1%

---

22 Submitted to University of Sheffield  
Student Paper <1%

---

23 "Hybrid Artificial Intelligent Systems", Springer Science and Business Media LLC, 2018  
Publication <1%

---

24 Anish Jindal, Amit Dua, Kuljeet Kaur, Mukesh Singh, Neeraj Kumar, S. Mishra. "Decision Tree and SVM-Based Data Analytics for Theft Detection in Smart Grid", IEEE Transactions on Industrial Informatics, 2016  
Publication <1%

---

25 d-scholarship.pitt.edu  
Internet Source <1%

---

26 "Deep Learning Applications for Cyber Security", Springer Science and Business Media LLC, 2019 <1%  
Publication

---

27 "Network and System Security", Springer Science and Business Media LLC, 2017 <1%  
Publication

---

28 [eprints.uthm.edu.my](http://eprints.uthm.edu.my) <1%  
Internet Source

---

29 "Progress in Artificial Intelligence", Springer Science and Business Media LLC, 2017 <1%  
Publication

---

30 Submitted to Institute of Technology Blanchardstown <1%  
Student Paper

---

31 Armin Lawi, Firman Aziz. "Comparison of Classification Algorithms of the Autism Spectrum Disorder Diagnosis", 2018 2nd East Indonesia Conference on Computer and Information Technology (EIConCIT), 2018 <1%  
Publication

---

32 Submitted to Universitas Dian Nuswantoro <1%  
Student Paper

---

33 Submitted to Manchester Metropolitan University <1%  
Student Paper

---

Sonam Nikhar, A.M. Karandikar. "Prediction of Heart Disease Using Different Classification Techniques", APTIKOM Journal on Computer Science and Information Technologies, 2017

Publication

---

<1%

---

Exclude quotes      On

Exclude matches      < 5 words

Exclude bibliography      On