International Conference on Renewable Energies and Power Quality (ICREPQ'18)
Salamanca (Spain), 21th to 23th March, 2018
Renewable Energy and Power Quality Journal (RE&PQJ)
ISSN 2172-038 X, No.16 April 2018

Nontechnical Loss Detection for Metered Customers in Alexandria Electricity Distribution Company Using Support Vector Machine

A. Hatem Tameem Alfarra[1], B. Amani Attia[2], C. S. M. El Safty[1]

[1] Electrical Engineering & Control Department
Arab Academy for Sience Technology and Maritime Traspotation AASTMT
Campus of Alexandria – (Egypt)
Phone/:+201006619620, e-mail: **drsamahsafty@yahoo.com** , **eng.hatem.tameem@gmail.com**

[2] Alexandria Electricity Distribution Company (AEDC)
amibrahim2004@yahoo.com
Phone, +201223992068

Abstract. Non-technical losses (NTL) during transmission and distribution (T&D) of electrical energy is a major problem faced by utility companies which is very difficult to fight and detect. For that, more of power utilities spend thousands dollar for research centres to find efficient methods for detecting and controlling abnormalities. Electricity theft and billing irregularities forms the main portion of NTL. With the introduction of smart meter, the frequency of reporting energy consumption data to the utility company has been increased. Incoming and outgoing energy could be monitored and analyzed. Electricity theft is a complex problem with many parameters to be evaluated before implementing any measures to detect and control that. These parameters include some issues like social, economic, regional, managerial, infrastructural, and corruption.

In recent years, several data mining and research studies on fraud detection and prediction techniques have been carried out in the electricity distribution sector. Support vector machine (SVM) technique has dominated the research for classifying data and detecting fraudulent electricity customers**.** SVM technique has good ability in data mining and data classification. The paper objective is to analyze the metered energy consumption data recorded and predict the pattern or the form of daily user's energy consumption, then using SVM to classify the data whether normal or theft. The suggested technique is then tested using real data from Alexandria Electricity Distribution Company (AEDC). The proposed technique was able to distinguish between healthy and theft cases.

Key words
Electricity theft, NTL, Smart meter, SVM

1. Introduction

Generation, transmission and distribution of electrical energy involve many operational losses. Whereas, losses concerned to generation can be technically defined, but Transmission and Distribution (T&D) losses cannot be precisely quantified from the sending end information. Transmission and distribution losses include both technical as well as non-technical losses (NTL) [1]. Technical losses are the losses due to power dissipation in transmission lines, transformers, and other power system components and are computed from the information about total load on the grid and total energy billed [2]. NTL cannot be precisely computed, but can be estimated. In general, Non-technical losses (NTL) are caused by the factors external to the power system.

Electricity theft and billing irregularities form the main portion of NTLs. In the side of electricity theft, it includes illegal tapping of electricity from the feeder, bypassing the energy meter, tampering with the energy meter and several physical methods to evade payment to the utility company. In billing irregularities side, it includes corruption and lack of commitment in employees of utility companies to control the illegal consumption. Electric utilities lose large amounts of money each year due to fraud by electricity consumers. Estimates show that, utilities worldwide lose more than $25 Billion every year due to NTLs [1]. Worldwide T&D losses exceed the total installed generation capacity of countries like the UK, Germany, and France. As an example, utilities in india lose as much as $4.5 Billion every year, due to illegal consumption of electricity. Utilities in USA incur losses about $1–6 Billion due to electricity theft [3].

Electricity fraud can be defined as a dishonest or illegal use of electricity equipment or service with the intention to avoid billing charge. This has a negative impact on energy companies and thus consumers including genuine and illegal customers by overloading the generating unit. The quality of electricity supply is adversely affected which leading to incorrect estimation about the quantity of electricity to be supplied to genuine as well as to illegal customers [2]. This huge amount of additional load may lead to brownouts and blackouts during the peak load period. In order to maintain good power factor and flat voltage profile along the feeders, sufficient reactive power has to be supplied besides the supplied electricity. When theft occurs, static volt-ampere reactive (VAR) compensation and maintenance of the power factor is very difficult due to the lack of complete total load flow information [4].

The financial losses are dangerous to many electric power organizations. Some power systems in worst affected countries are near bankrupt. In developing countries, tapping electricity directly from the feeder and bypassing the meter are done only at desired hours of the day where, it is a common practice that such customers use electricity from the utility grid legally when their consumption is low and steal electricity when their consumption is high **[**1**]**. Corrupt staff of the utility companies might take bribes from illegal consumers to help them allow such practices. In addition, corrupt employees are partially responsible for billing irregularities, as they can record quantity lower than the original energy consumption.

The most effective method to detect and control NTLs up to date is by using intelligent and smart electronic meters that make fraudulent activities more difficult, and easy to detect [2]. These are in smart meter, artificial intelligent techniques and data mining. A smart meter is an advanced energy meter that identifies energy consumption in more detail compared to a conventional energy meter. In addition smart meter can reliably and securely communicate the data back to the local utility for monitoring and billing purposes [5]. Smart meters with great networking capability and advanced software tools are very difficult to tamper and hack.

The following section discusses the previous work for techniques used in electricity theft detection. Data analysis, load profile generation and prediction of energy consumption curve are clearly in section 3. Support vector machine SVM, advantages, Kernel function related and implimentation are in section 4. Egyption case study, data, SVM implimentation and results are in section 5. Conclusion of paper in section 6.

2. Methods of Fraud Detection

Studies have proposed and developed several techniques for detection and estimation of electricity theft [2]. Of these techniques, some methods are illustrated in this section. In [1], Soma Shekara Sreenadh Reddy Depuru, Lingfeng Wang, and Vijay Devabhaktuni discussed the problems underlying detection of electricity theft, previously implemented ways for reducing theft. In addition, they presented the approximate energy consumption patterns of several customers involving theft. Energy consumption patterns of customers are compared with and without the presence of theft. A dataset of customer energy consumption pattern is developed based on the historical data. Then, support vector machines (SVMs) are trained with the data collected from smart meters, that represents all possible forms of theft and are tested on several customers.

In [6], Jawad Nagi and his assistances intent was to assist Tenaga Nasional Berhad (TNB) Sdn. Bhd. in peninsular Malaysia to reduce its NTLs in the distribution sector by using Support Vector Machine (SVM) technique. The proposed method was more effective compared to the past actions taken by (TNB) Sdn. Bhd themselves. After implementing the new fraud detection system, TNBD’s detection hit-rate increased from 3% to 60%.

In [7], same authors with same motivation. They applied Genitic Algorithim (GA) and SVM to get better performance. Performance of the fraud detection system by implementing the expert system increased the detection hit-rate from 37% to 62%. The hybrid combination of GA-SVM provided a better solution for selecting optimal SVM hyper-parameters.

In [8], authors proposed a methodology based on distribution state estimation to detect customer tampered data. They studied a three-phase 8-bus low voltage distribution system. This study uses the concept and mathematical model of the random error variable. Test results proved that the non-technical losses of power distribution system is reduced, thus contributing to the operational benefits of smart meters. Limited method performance when compared to Newton Raphson method due to the initial point of

semi-definite programming that leads to global optimal solution not local optimal solution.
In [9], authors provided a very high performance detector by using the cusomer's consumption pattern. Using distribution transformer meters, areas with a high probability of energy theft, and by monitoring abnormalities in consumption patterns, suspicious customers are identified.

Enhanced encoding technique is used in [10]. Authors explained the significance of the evaluation of customer energy consumption profiles for identification of ilegal consumers. After the encoding process, the data has been inputted to a support vector machine (SVM) classification model where the accuracy of the SVM model was 92%.

In [11], Nathiya.S, Rajeswari.S and Geetha.R. proposed a Wireless electricity theft detection system using ZIGBEE technology in order to overcome the theft of electricity via bypassing the energy meter and also control the utility's revenue losses. Advantages of this system is: much useful detection of electricity stealing worldwide, encures accurate billing of the electricity consumed, excellent way to detect the bypassing of the energy meter, and low cost with low power consumed by the ZIGBEE module. This wireless system provides much better results at short haul but the concern of the long haul depends upon the service employed by the network.

Also in [12], Jawad Nagi and his assistances presented the inclusion of human knowledge and expertise into the SVM-based fraud detection model (FDM) with the introduction of a fuzzy inference system (FIS), in the form of fuzzy IF-THEN rules. The FIS acts as a post-processing scheme for short-listing customer suspects with higher probabilities of fraud activities**.** The implementation of this improved SVM-FIS computational intelligence FDM, Tenaga Nasional Berhad Distribution's detection hit-rate has increased from 60% to 72%, thus proving to be cost effective.

3. Support Vector Machine

This section is an introduction to support vector machines and related kernel methods in supervised learning, whose responsible to estimate an input-output functional relationship from a training set of examples [13]. A learning problem is referred to as classification if its output takes discrete values in a set of possible categories and regression if it has continuous real-valued output.

A simple and useful model of an input-output functional relationship is to assume that the output variable can be expressed approximately as a linear combination of its input vector components. These linear models include the logistic regression method for classification and the linear least squares method for regression [14]. Because a linear model has limited prediction power by itself, there has been extensive research in nonlinear models such as neural networks. Linear methods have recently returned their popularity because of their simplicity both computationally and theoretically. It has also been realized that with appropriate features, the prediction ability of linear models can be as good as nonlinear models.

For using a linear model to represent a nonlinear functional relationship, we need to include nonlinear feature components. A useful technique for constructing nonlinear features is kernel methods, where each feature is a function of the current input. This idea has recently received much attention because of the introduction of SVMs and the renewed interest in Gaussian processes. Support vector machines were introduced by Vapnik and are a set of related supervised learning methods. They can analyze the given data and can recognize a pattern or trend in the data with respect to output [1]. The concept of support vector learning includes [14]:

1) *Perceptons, as in equations (1) and (2)*
2) *Convex programming and duality, as in eq. (3)*
3) *Using maximum margin to control complexity.*
4) *Representing nonlinear boundaries with feature expansion.*
5) *The "kernel trick" for efficient optimization.*

Input space (data points),
$X = [x1; x2; x3; \ldots]$ with $x_i \in R^n$ (1)

Labels (output space) $y = [y1; y2; y3; \ldots]$ with $y_i \in \{-1, 1\}$ (2)

SVMs in Dual representation,
$f(x) = (w, x) + b = \sum \alpha_i y_i (x_i, x) + b$ (3)

Weight vector $w = \sum \alpha_i y_i x_i$ (4)

The combination is unique, which makes SVMs and related kernel-based learning methods special and interesting. SVMs have also successfully been applied in practice, especially for classification problems [13]. Although many problems that have been successfully solved by SVMs could also have been solved successfully by standard statistical methods such as penalized logistic regression, in practice, SVMs can still have significant computational advantages.

4. Prediction of Load Curve

User's energy consumption pattern is one of the useful methods used for energy monitoring which leads you to an approximate form of load predection. Here, we will explain how we generate or predict the new energy consumption readings. After collecting the metered energy consumption for such region, we rearrange and analyze the data to reach to the approximately pattern of load curve. Therefore, load profile or data form is ready to implement with SVM technique. For example, the metered energy data about the hourly electricity consumption on a Distribution Point (DP) has been obtained from PJM datasets. PJM Interconnection is a regional transmission organization (RTO) that coordinates the movement of wholesale electricity in all or parts of Delaware, Illinois, Indiana, Kentucky, Maryland, Michigan, New Jersey, North Carolina, Ohio, Pennsylvania, Tennessee, Virginia, West Virginia and the District of Columbia. In 2016 version of data, WEST zone (DP) has eight load points (eight attached regions), AP, CE, AEP, DAY, DUQ, DEOK, ATSI, and EKPC. First, we calculate the attached regions consumption percentage (%) of the WEST DP. Then, we calculate the average MW reading for every hour of a day through all month's days. So, this step will be repeated twelve times according to the year's months.

Weather conditions is a very important factor that affect the MW reading or consumption such as season of year, month, day, and what time is that reading happened. Winter season curve and summer season curve are obtained as shown in fig.1., and fig.2. Now, we can predict the pattern of daily MW consumption curve for any day through the year. For example, tested various random day for both winter summer seasons. Maximum MW consumption readings for tested days in main DP are important for feeder percentage calculation.

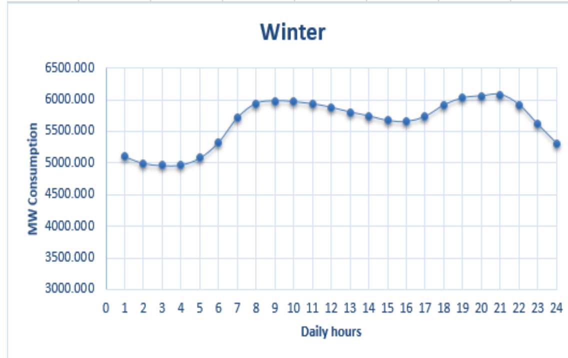


Fig.1. Winter curve for AP region

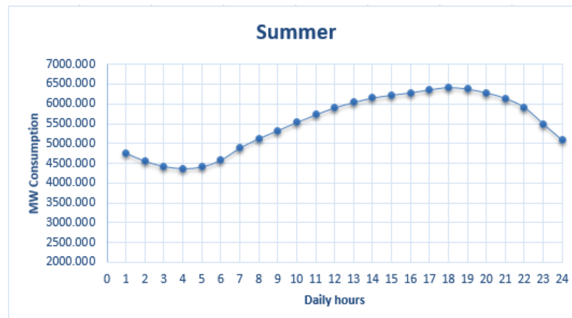


Fig.2. Summer curve for AP region

Results in fig.3. shows the actual daily energy consumption curve and the predicted daily energy consumption curve for winter season day. Table I illustrates the actual metered energy consumption with results of new metered energy consumption that forms the prediction energy consumption curve for tested winter season days, also the difference between actual and suggested values and the percentage (%) of error are shown. Blue curve represents the actual consumption and the red curve represents the suggested consumption.

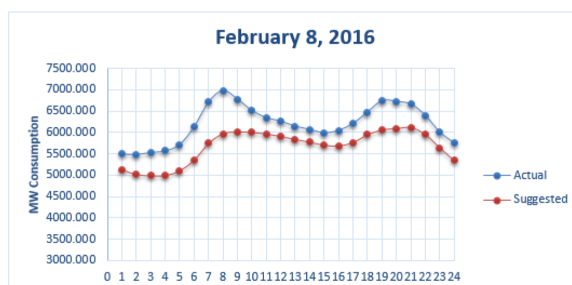


Fig.3. Actual MW consumption and predicted MW consumption for February 8, 2016 in AP region.

Table I- Tested winter days with predicted consumption and error %

Day	Daily hours	H01	H02	H03	H24	Error%
	Winter	5104.406	4997.090	4964.613	5312.255	
	%	83.97%	82.21%	81.67%	87.39%	
February 8, 2016	Actual	5492.335	5483.838	5532.187	5751.155	8.439%
	Suggested	5134.062	5026.122	4993.457	5343.118	
	Diff.	358.273	457.716	538.730	408.037	
March 11, 2016	Actual	4162.553	4069.681	3994.555	4431.158	2.687%
	Suggested	4173.800	4086.049	4059.493	4343.755	
	Diff.	-11.247	-16.368	-64.938	87.403	
April 4, 2016	Actual	4636.638	4546.201	4528.104	5227.476	5.077%
	Suggested	4556.949	4461.143	4432.149	4742.506	
	Diff.	79.689	85.058	95.955	484.970	

In summer season side, fig.4. represents the actual daily energy consumption curve and the predicted daily energy consumption curve. The difference between actual and new metered energy consumption is clearly in table II. Error percentage (%) is also calculated to insure that the predicted consumption readings are approximately mapping with real energy consumption readings.

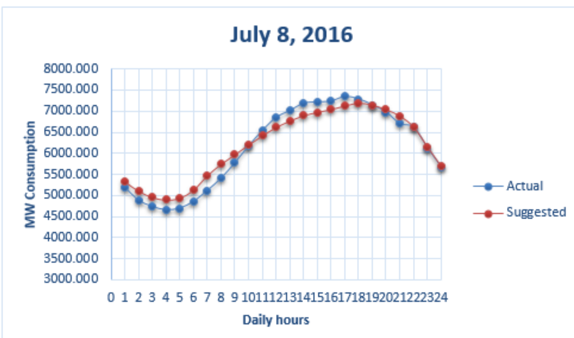


Fig.4. Actual MW consumption and predicted MW consumption for July 8, 2016 in AP region.

Table II- Tested summer days with predicted consumption and error %.

Day	Daily hr.	H01	H02	H03	H24	Error%
	Summer	4763.909	4560.354	4427.891	5083.048	
	%	0.742	0.711	0.690	0.792	
July 8, 2016	Actual	5191.833	4884.943	4735.251	5672.129	0.679%
	Suggested	5340.222	5112.042	4963.555	5697.969	
	Diff.	-148.389	-227.099	-228.304	-25.840	
August 8, 2016	Actual	4810.921	4588.963	4449.790	5500.921	8.046%
	Suggested	5512.473	5276.933	5123.657	5881.760	
	Diff.	-701.552	-687.970	-673.867	-380.839	
September 15, 2016	Actual	4849.798	4591.373	4435.101	4799.573	2.938%
	Suggested	4532.816	4339.135	4213.099	4836.474	
	Diff.	316.982	252.238	222.002	-36.901	

In order to test the suggested SVM, the data was divided into two groups. One group was the normal consumption of the feeders. The other group was built upon a suggestion that non legal consumption of electricity existed. Such consumption was represented by decreasing the percentage consumption on 3 feeders connected to the WEST DP from the normal values and representing the consumption with the new data. The data obtained is fed into the Support Vector Machine (SVM) classification model by classifying customers into two categories: genuine customers, and theft consumers. Multiple SVMs trained classifier are applied to achieve better accuracy as shown in fig.5. Fig.6. represents the ROC (receiving operating characteristics) curve for chosen SVM classifier, area under the curve AUC, True positive rate and False positive rate are clearly identified.

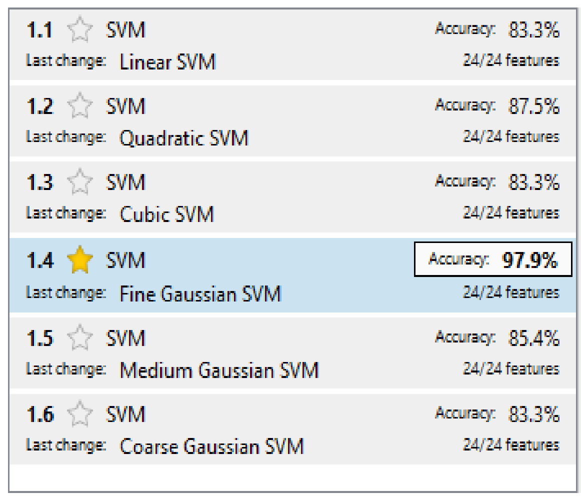


Fig.5. Trained classifier SVM model for tested data.

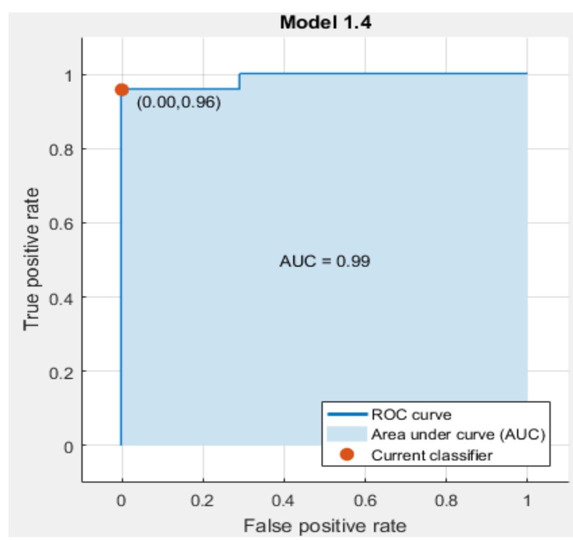


Fig.6. ROC curve for chosen SVM classifier.

5. AEDC Case Study

Antra Distribution Point DP is an industrial DP which exists in Egypt, Alexandria city. It has two incoming feeders and six outgoing feeders as shown in figure 7.

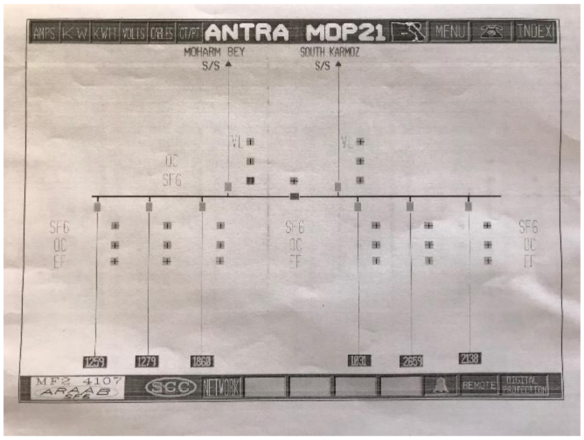


Fig. 7. ANTRA Distribution Point DP.

The metered KW consumption readings for full year 2016-2017 obtained from electric energy company in Alexandria city. Transformer's ratings for each sub-distributor is available. The KW consumption percentage for each sub-distributor is deduced. After arranging and analysing the data, winter curve and summer curve are obtained as in fig.8, and fig.9.

Now, training data profile is ready to SVM classification model that classify the data into two categories: normal consumption and abnormal consumption. Multiple SVMs trained classifier are applied to achieve better accuracy as shown in fig.10.

The obtained SVM classifier is tested using actual data, the proposed technique was able to detect some feeders with suggested theft.

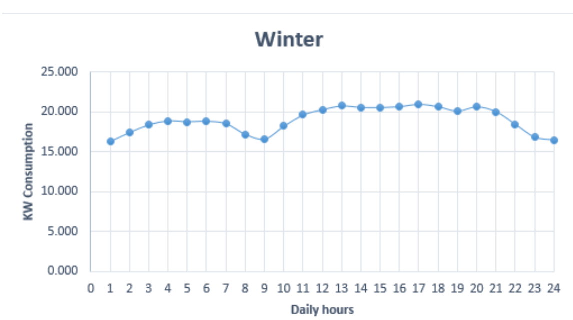


Fig.8. Winter curve for Sub-distributor 1.

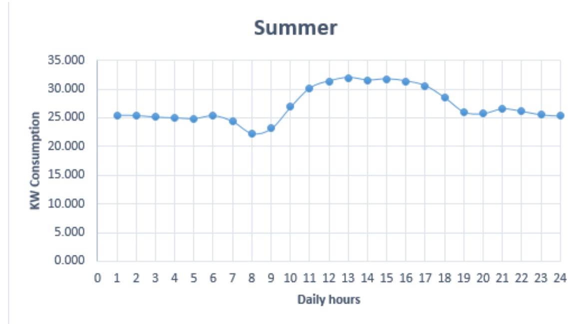
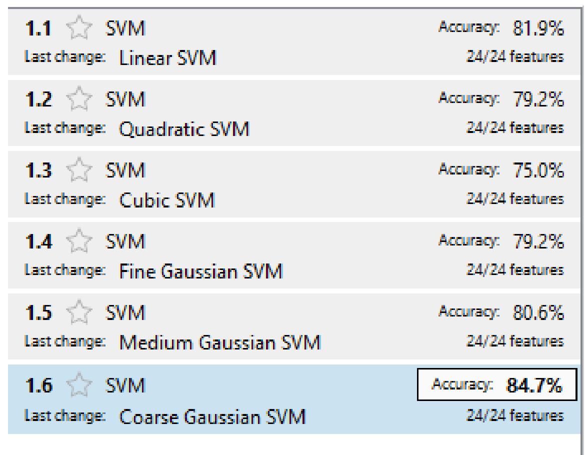


Fig.9. Summer curve for Sub-distributor 1.



1.1 SVM	Last change: Linear SVM	Accuracy: 81.9%	24/24 features
1.2 SVM	Last change: Quadratic SVM	Accuracy: 79.2%	24/24 features
1.3 SVM	Last change: Cubic SVM	Accuracy: 75.0%	24/24 features
1.4 SVM	Last change: Fine Gaussian SVM	Accuracy: 79.2%	24/24 features
1.5 SVM	Last change: Medium Gaussian SVM	Accuracy: 80.6%	24/24 features
1.6 SVM	Last change: Coarse Gaussian SVM	Accuracy: **84.7%**	24/24 features

Fig.10. Trained classifier SVM model for tested data.

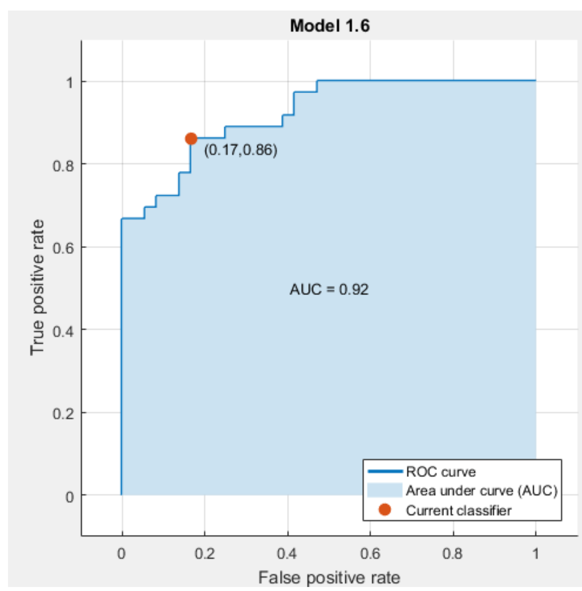


Fig.11. ROC curve for chosen SVM classifier.

6. Conclusion

The paper presented the importance of having recorded data in order to detect electricity theft. It suggests a methodology to predict the energy consumption at the different feeders in order to compare with the actual data or if enough data is not available. With the implementation of the SVM detection of feeders with suggested electricity theft could be detected. It must be mentioned that there is a need of advanced metering infrastructure is for monitoring the consumption and hence detection of any abnormalities in consumption.

References

[1] A. Soma Shekara Sreenadh Reddy Depuru, B. Lingfeng Wang, and C. Vijay Devabhaktuni, "Support Vector Machine Based Data Classification for Detection of Electricity Theft", IEEE/PES Power Systems Conference and Exposition2011, pp. 1-8.

[2] A. Soma Shekara Sreenadh Reddy Depuru, B. LingfengWangn, and C. Vijay Devabhaktuni. "Electricity theft: Overview, issues, prevention and a smart meter based approach to control theft", Energy policy, pp. 1007-1015, nov. 2010.

[3] A. Abhishek Chauhan, and B. Saurabh Rajvanshi, "Non-Technical Losses in Power System: A Review", in proc. International Conference on Power, Energy and Control (ICPEC) 2011, pp. 558-561.

[4] A. Thomas B. Smith, "Electricity theft: a comparative analysis", Energy Policy, vol. 32, pp. 2067–2076, Aug. 2003.

[5] A. Muhammad Anas, "Avoiding Electricity Theft using Smart Meters in Smart Grid", in proc. COMSATS Institute of Information Technology Islamabad2012, pp. 1-59.

[6] A. Jawad Nagi, B. Keem Siah Yap, C. Sieh Kiong Tiong, D. Syed Khaleel Ahmed, and E. Malik Mohamad, "Nontechnical loss detection for metered customers in power utility", in proc. IEEE Transactions on Power Delivery2010, vol. 25, pp. 1162-1171.

[7] A. J. Nagi, B. K. S. Yap, C. S. K. Tiong, D. S. K. Ahmed, and E. A. M. Mohammad, "Detection of Abnormalities and Electricity Theft", in proc. TENCON 2008 - 2008 IEEE Region 10 Conference2008. Pp. 1-6.

[8] A. Chun-Lien Su, B. Wei-Hung, and C. Lee Chao-Kai Gen, "Electricity Theft Detection in Low Voltage Networks with Smart Meters Using state estimation", in proc. IEEE International conference on Industrial Technology (ICIT)2016, pp.493-498.

[9] A. Paria Jokar, Student Member, IEEE, B. Nasim Arianpoo, Student Member, IEEE, and C. Victor C. M. Leung, Fellow, IEEE, "Electricity Theft Detection in AMI Using Customers' Consumption Patterns", in proc. IEEE TRANSACTIONS ON SMART GRID2016, vol. 7, pp. 216-226.

[10] A. Soma Shekara Sreenadh Reddy Depuru, B. Lingfeng Wang, and C. Vijay Devabhaktuni, "Enhanced Encoding Technique for Identifying Abnormal Energy Usage Pattern", in proc. North American Power Symposium (NAPS)2012, pp. 1-6.

[11] A. Nathiya.S, B. Rajeswari.S and C. Geetha.R, "Wireless Electricity Theft Detection System Using Zigbee Technology", in proc. International Journal of Advances in Engineering 2015, pp. 1-6.

[12] A. Jawad Nagi, B. Keem Siah Yap, C. Sieh Kiong Tiong, D. Syed Khaleel Ahmed, and E. Faruca Nagi, "Improving SVM-Based Nontechnical Loss Detection in Power Utility Using the Fuzzy Inference System", in proc. IEEE Transactions on Power Delivery2011, vol. 26, pp. 1284-1285.

[13] A. Tong Zhang, "An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods", AI Magazine Vol. 22 Number 2 (2001) (© AAAI), pp. 1-2.

[14] A. Geoff Gordon, "Support Vector Machine and other Kernel methods", in proc. The 14th IEEE International Conference on Fuzzy Systems, 2005. FUZZ '05, pp. 1-76.