ENERGY CONSUMPTION PATTERN BASED ELECTRICITY FRAUDULENT DETECTION USING FUZZY ALGORITHM

[1]C.Anitha, [2]V.Mahalakshmi, [3]S.Chidambaram

[1] UG Student, Dept. of IT, National Engineering College, Kovilpatti.

[2]UG Student, Dept. of IT, National Engineering College, Kovilpatti.

[3]Asst. Professor (Senior Grade), Dept. of IT, National Engineering College, Kovilpatti.

ABSTRACT:

We have introduced a novel Consumption Pattern Based Energy Theft (CPBET) Detection System algorithm which leverages the predictability property of customer's normal and malicious consumption patterns. This approach focus on both practical reduction as well as machine learning reduction. In practical reduction with the help of distribution transformer meters, areas with a high probability of energy theft are shortlisted, and by monitoring abnormalities in consumption patterns, fraudulent customers are identified. In machine learning reduction focus on two parts. In the first part of the paper k-means based fuzzy clustering was performed to group customers with similar profiles. In the second part of the paper K Nearest Neighbor (KNN) fuzzy classification was then performed and Euclidean distances to the cluster centers were measured. Customers with large distances to the cluster centers were considered potential fraudsters.

I. INTRODUCTION

Energy theft has been a major concern in traditional power system worldwide. The Generation, Transmission and Distribution (T&D) of electricity involve huge operational losses. The Magnitude of these losses is rising at an alarming rate in several countries. In order to identify illegal consumers of electricity in the view of enhancing the economy of utilities, efficiency and security of the grid, a new method of analyzing electricity consumption patterns of customers and Identifying illegal consumers is proposed and realized. Losses that occur during generation can be technically defined, but T&D losses cannot be quantified completely from the sending-end information. Distribution losses in several countries have been reported to be over 30%. Substantial quantity of losses proves the involvement of Non-Technical Losses (NTL) in distribution. Total losses during T&D can be evaluated from the information like total load and the total energy billed, using established standards and formulae. India incurs losses around $9 billion every year in the form of electricity theft. In the United States (U.S.) alone energy theft was reported to cost the utility companies around $6 Billion per year. In Canada, BC Hydro reports that the electricity theft costs $100 million every year In Taiwan the most frequent low voltage customers are coastal farmers, gardeners, and flower growers, with total annual electricity revenue lost through illegal power usage estimated at over NT$1 billion.

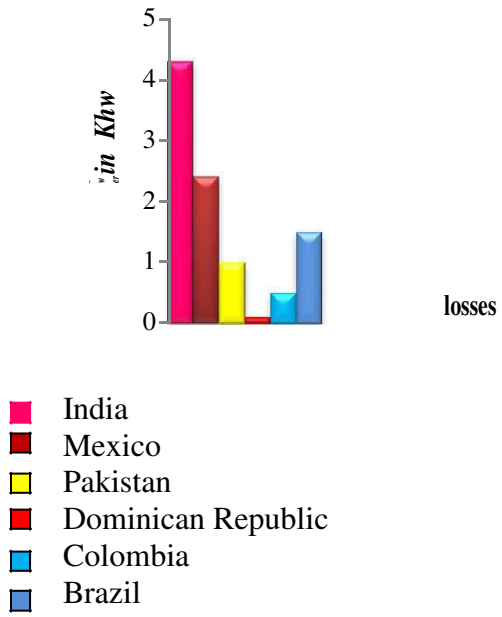

Fig. 1: Overall technical and non technical losses in various countries

To address this issue implementation of advanced metering infrastructure (AMI) promises to mitigate the risk of energy theft. Smart grid is an electrical grid which includes variety of energy measures including smart meters. Advanced metering infrastructure is a key technology in smart grid and data transferred through AMI are high degree predictable. Traditionally theft detection is done manually by inspecting consumers.

This is time consuming process and requires large number of field staff. The cost for this process is too high and detection rate is not so high. To overcome these costs, currently some data mining techniques are used to detect theft. We proposed a CPBET Detection System approach for detection of energy theft, which will improve accuracy of detection and requires less cost for whole process. This paper illustrates the importance of identifying the electricity theft based on customer's energy consumption pattern over a period of time.

In a power generation system the transmission and distribution of energy involve many losses. The losses may include both technical as well as non technical. Technical loss means that the losses that occur during distribution of electrical energy. Non technical loss means that losses that cannot be quantified from sending end information.

.

II. PREVIOUS WORK

This chapter illustrates the algorithms implemented for detecting illegal consumers. These classification algorithms include SVM, Rule Engine, and Neural Network Pattern Recognition (NNPR) tool based classification models. The classification results of the proposed classification algorithms are presented.

2.1 SVM Based Classification Model

SVMs introduced by Vapnik are a set of supervised learning methods. They can analyze the given data and recognize a pattern or trend in the data with respect to output. SVMs are also used for regression analysis and statistical data classification. Given a training dataset that represent a set of rules, a model can be developed by the SVM using a training algorithm. In general, SVMs develop a hyper plane or set of hyper planes in a high or infinite dimensional space, depending on the complexity of the data that needs to be classified.

Significant separation between the classified data points can be achieved when the hyper plane has significant distance to the nearest training data points of any class. The generalization error of the classifier will be minimal if the separation margin is high. In the recent past, SVMs have found numerous applications in face recognition, text categorization to bioinformatics, and data mining. The training data with x_i R_n, i

= 1,…….,*l,* in two classes, and a vector *y* R*l* such that *y* = {1, -1}. Here, LibSVM is used for developing the required classification model. LibSVM is a library for developing SVMs based on classification model in MATLAB developed by C.C. Chang and C.J. Lin. In power engineering, SVMs are used for several applications including estimation of electricity theft and analysis of power quality parameters in a power grid. Classification accuracy is the ratio of correctly classified data samples over all data samples. The customers are classified into three classes based on the following criteria and their instantaneous energy meter readings (rules):

To be classified into Class-D: in instantaneous energy meter readings of any customer,

If zero energy consumption is recorded for more than two hours in one day or

If zero-energy readings are recorded more than 8 times (eight of 96 consecutive readings) including repetitions.

To be classified into Class-S: in instantaneous energy meter readings of any customer,

If 3 readings (of 4) in any hour of a single day are recorded as zeros

If two consecutive zero readings are recorded in an entire day with less than three repetitions,

If zero energy readings are recorded between three to six times in a day.

To be classified into Class-I: in instantaneous energy meter readings of any customer,

If zero energy readings are recorded less than two times in

one day including repetitions.

Initially, utilities collect instantaneous electricity consumption data from the smart meters in specific intervals of time. Thus, collected energy consumption data would be a series of instantaneous energy consumption values. Format of the electricity consumption data collected from smart meters need to be modified; so that, it would be compatible with SVM model developed using LibSVM. A portion of the data has been extracted as training data and the rest as testing data. Inputs to the SVM model is the instantaneous energy consumption data, and the output is the classes that a particular customer belongs to. Before the data being used for training and testing, it may be viewed categorically based on the geographical location, whether it is a weekday or weekend, load capacity range of the customer and what season of the year this data represent.

The data is then transferred to a database located at a central control station. Then, training data is used to train the SVM model and test it for detecting the illegal consumers. If a customer profile is genuine and the energy consumption is continuous, then that the customer is rated as a genuine customer. If the customer profile is suspicious, then the profile needs to be evaluated further. If a customer's energy consumption fulfills criteria specified for Class-D, then, the customer may be inspected immediately, as the probability of illegal consumption is very high. If a customer is classified as Class-S, and if the customer is either a large or medium customer then that customer is immediately inspected. If that customer is a small customer, then that customer is periodically inspected. If a customer falls under Class-I, that customer can be reported as a genuine customer. If a customer's profile does not fall under any class and the calculated overall distribution losses are more than 4% (excluding the classified illegal consumers), then the customer profile is reevaluated. In general, distribution losses are ideally to be

between 3–5% at feeder level. If the losses are more, then it can be assumed that illegal consumers might exist on the distribution feeder. Therefore, classification algorithm may be terminated if the losses are less than 4%, but if the losses are more than 4%, the classification algorithm is reiterated.

III. PROPOSED WORK

First part of the paper **k-means based fuzzy clustering** was performed to group customers with similar profiles. It is a partitioning algorithm wherein the resultant clusters are independent and bound. There are broadly two main stages of algorithm implementation.

3.1 K Means Fuzzy Clustering

Non supervised learning algorithm.

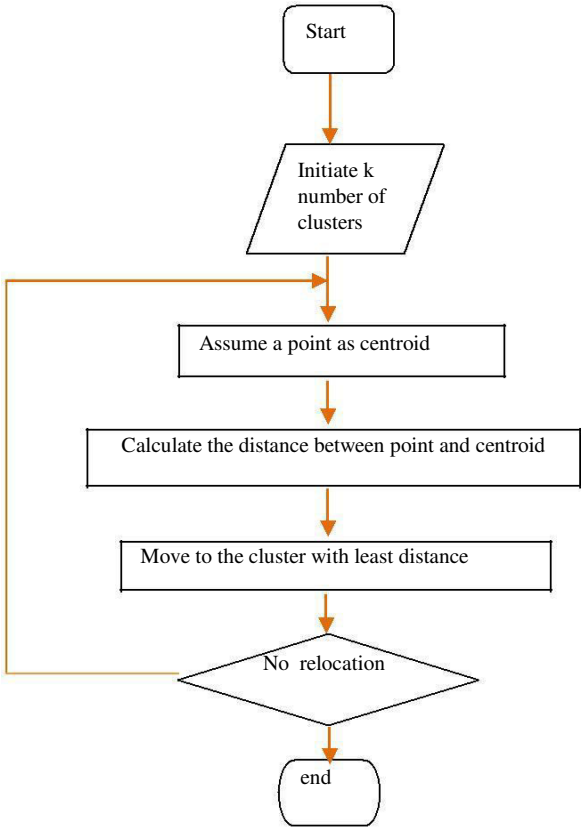

Fig. 2: Workflow of K means clustering algorithm

3.2 K means fuzzy clustering algorithm:

Input: Training matrix

Output: Class label

Step 1: Randomly assign cluster centroid to the plotted data points in a graph in this case any of the data point acts as cluster centroid.

Step 2: Calculate the distance from center to all the data points.

Equation to calculate the distance between two points is $\sqrt{(2-1)+(2-1)}$

Step 3: Assign points to the cluster according to minimum distance. A new training matrix will be generated.

Step 4: Centroid update

Step 5: Go to step 2 and step 3

Step 6: Check the current training matrix results with the previous training matrix results.

Step 7: The training matrix results will be generated similarly in two or more steps.

Let training matrix results be 'X' &

Previous training matrix results be 'Y'

If X != Y

Go to step 4

Else

Stop

In the second part of the paper KNN fuzzy classification was then performed and Euclidean distances to the cluster centers were measured. Customers with large distances to the cluster centers were considered potential fraudsters. KNN can be

used for both classification and regression predictive problems. However, it is more widely used in classification problems in the industry. In both cases, the input consists of the k closest training examples in the feature space. The output depends on whether k-NN is used for classification or regression.

3.3 Clustering Method to find the number of clusters:

Silhouette plots are applied to determine the number of clusters within a dataset. Assume that the data have been clustered into k clusters and for each sample i, a(i) is the average dissimilarity of i with other samples within the same cluster. Also b(i) is the least average dissimilarity of i to any other clusters. The silhouette value, s(i) is defined as

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}.$$

The average of s(i) over all samples within a cluster shows how close the samples in the cluster are, and averaging over the entire dataset shows how properly the data have been clustered. We use this method to determine the number of clusters in the dataset of each customer. Defining separate classes for distinct clusters can help to achieve a higher classification accuracy.

3.4 K-Nearest Neighbor Fuzzy Classification Algorithm:

In k-NN classification, the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbor. The training examples are vectors in a multidimensional feature space, each with a class label. The training phase of the algorithm consists only of storing the feature vectors and class labels of the training samples. In the classification phase, K is a user-defined constant, and an unlabeled vector or test vector is classified by assigning the label which is most frequent among the K training samples nearest to that query point. A commonly used distance metric for continuous variables is Euclidean distance. For discrete variables, such as for text classification, another metric can be used, such as Hamming distance.

Euclidean distance:

The distance between points p and q is the length of the line segment connecting them.If p = (p1, p2,.. pn)& q = (q1, q2,…qn) are two points in Eclidean n-space, then the distance from p to q, or from q to p is given by the below equation.

d (p, q) can be expressed as,

$$d(p, q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2} \qquad d(p, q) = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}$$

3.5 K - Value selection

The best choice of K depends upon the data. Generally, larger values of K reduce the effect of noise on the classification, but make boundaries between classes less distinct. The special case where the class is predicted to be the class of the closest training sample is called the nearest neighbor algorithm where the K value will be 1. The accuracy of the K-NN algorithm can be severely degraded by the presence lot of nearest points that may be the burden for classification. Now the K value is increased beyond the level 1that invoking the
'majority voting' process. When the K=3,

the classifier choosing top three nearest neighbors and deciding the majority of the boundary will be the final class variable.

3.6 KNN Fuzzy Classification algorithm:

- It is a supervised learning algorithm.
- Classification works based on training.

Input: K closest training examples in the feature space

Output: Class membership

Step 1: Determine the parameter K = number of nearest neighbor

Step 2: Calculate the distance between query instance and all the training samples.

Step 3: Sort the distance and determine nearest neighbor based on K value.

IV. RELATED WORK

Jiang *et al.*, [5] analyzed the background of advanced metering infrastructure (AMI) and identifies major security requirements that AMI should meet. Specifically, identify the energy-theft behaviors in AMI. Deep understanding of security vulnerabilities and solutions in AMI and also explores some open challenges and potential solutions for energy-theft detection.

S. McLaughlin *et al.,* [7], a multi sensor energy theft detection framework for AMI (AMIDS) was presented. AMIDS collects evidences of malicious behavior from three types of information sources: 1) cyber-side network and host-based IDSs; 2) on-meter anti-tampering sensors; and 3) power measurement-based anomalous consumption detectors through non intrusive load monitoring (NILM). These types of information were combined to minimize the FPR. While combining the information from different sources is effective in reducing the FPR, the algorithms chosen for detecting anomalies have some drawbacks. Use of NILM, which requires a high-sampling rate, reveals information about types and time of use of appliances in customers' premises. In this paper, we focus on detecting energy theft attempts only based on customers' consumption patterns that can also be used as a part of a multi sensor framework like AMIDS. Compared to the NILM-based technique CPBETD provides a high performance with a much lower sampling rate.

P.Jokar *et al.*, [22] a new approach for detecting intrusions is the advanced metering infrastructure (AMI). Based on the electricity usage reports and pricing confirmations the electricity consumption patterns of customers will be generated that follow a statistical model. Consumption patterns are used to detect adversarial activities in AMI.

E. Angelo's *et al.*, [12], using six months usage reports, five attributes including average consumption, maximum consumption, standard deviation, sum of the inspection remarks, and the average consumption of the neighborhood were chosen to create a general pattern of power consumption for each customer. k-means based fuzzy clustering was performed to group customers with similar profiles. A fuzzy classification was then performed and Euclidean distances to the cluster centers were measured. Customers with large distances to the cluster centers were considered potential fraudsters. Clustering the customers and relying on long-term measurements limits the accuracy of this ETDS and causes long detection delay. Having more detailed metering information in AMI, CPBETD provides a much better performance with a much shorter delay.

S. Depuru, L. Wang,[15] described by Transmission and distribution of electricity involve technical losses (TLs) and

non technical losses (NTLs). Illegal consumption of electricity constitutes major portion of the NTL at distribution feeder level. Illegal consumption of electricity has to be detected instantly in real time. This paper mainly focuses on detecting illegal consumers using high performance computing (HPC) algorithm.

C. H. Lo and N, [16] described by recent analysis of energy theft incident is that the dishonest customers would lower their electricity bills by tampering with their meters. The physical attack can be extended to a network attack by means of false data injection (FDI). Investigate FDI attack by introducing the combination sum of energy (CONSUMER) attack in a coordinated manner on a number of customers' smart meters, which results in a lower energy consumption reading for the attacker and a higher reading for the others in a neighborhood.

Salinas, [17] described by smart grid being proposed to modernize the current power grids. Since the smart meters used in smart grids are vulnerable to more type of attacks. To identify illegal users some schemes have been proposed to utility companies (UCs) to detect energy theft in power grids, they all require users to send private information such as meter readings at certain intervals to UCs, which attacks user's privacy. To identify energy theft detection considering user privacy issues this paper uses a special algorithm to identify fraudulent users.

Mashima and Cardenas [18] suggested modeling the probability distributions of the normal and malicious consumption patterns, and application of the generalized likelihood ratio (GLR) test to detect energy theft attacks. They used auto regressive moving average (ARMA) to model customers' normal and malicious consumption distributions. They assumed that an attacker would choose a probability distribution that decreases the mean value of the real consumption. This, however, is not necessarily true with AMI. Considering the dynamic pricing in smart grids, by only changing the order of meter readings without altering the average, electricity theft is possible. Another major issue with ARMA–GLR detector is that it is only effective if the normal electricity theft behavior and attack patterns can accurately be modeled by an ARMA process.

S.McLaughln, [7] described AMI intrusion detection system which uses information fusion to combine the sensors and consumption data from a smart meter to more accurately detect energy theft. This method detects theft related behavior with high accuracy.

In our previous work [22], we introduced algorithms for detecting fraudulent activities against AMI, based on finding anomalies in consumption patterns. While covered different types of malicious activities, this paper is focused on energy theft and is tailored for its unique characteristics. This method includes employment of mechanisms for making the algorithm robust against non malicious changes, application of clustering techniques to enhance the classification accuracy, studying the effect of sampling rate on performance, employment of a real dataset of smart meter readings for performance evaluations, and comparison with other existing Energy Theft Detection System (ETDS).

V. PROPOSED CPBET DETECTION SYSTEM ALGORITHM

Data preprocessing, including operations like dimension reduction (the process of reducing the number of random variables) and normalization. Each data vector in the dataset includes the meter readings of a customer over a 24-h period,

for instance for n measurements per hour the data vector has 24×n elements. While there are several methods for dimension reduction that minimize the information loss by extracting the important features of data, we sum up the in-between samples to make the algorithm compatible with different metering rates. That is, rather than applying a feature extraction technique that saves the key information of a higher dimension data vector while reducing the data size, we only add up the samples.

Once the data is converted into the proper format, the k-means fuzzy clustering algorithm is performed on the benign dataset. Several non malicious factors can alter the consumption pattern, such as seasonality, change of appliances, and different usage during weekdays and weekends. In order to have a better Deduction Rate (DR), k-means clustering with different values of k is performed on the data, and each time the silhouette value of the clusters is calculated. A peak in the silhouette plot for k = l shows that the data is originated from 'l' different distributions. Clusters that have few members are eliminated and will not be used for training the classifier. This can help to prevent pollution of the benign dataset by undetected attacks. We use 'l' to denote the final number of clusters after eliminating the small groups.

The next step is preparing a dataset for training the classifier. While a dataset of benign samples for each customer is easily obtainable using historic data, malicious samples might not be available, since energy theft might never or rarely happen for a given customer. In order to address the problem of imbalanced data, one solution is the application of single-class classification techniques where the classifier is trained only using normal samples.

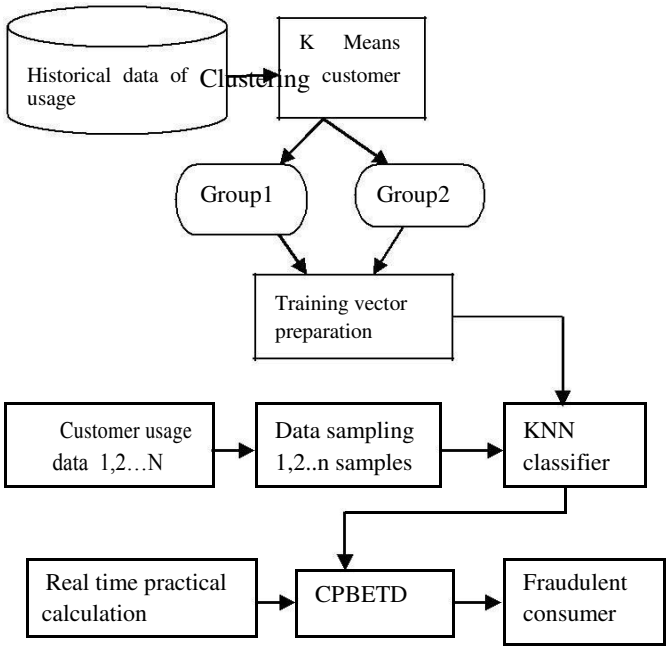

Fig. 3: Phases of CPBET Detection System

5.1 Pseudo codes of CPBET Detection System:

Input: NS (new sample)

Output: attack (boolean)

Variables: counter1=0, threshold1=3, genuineCount=0, counter2=0, threshold2=3, threshold3=5;

if ETM > ∑i ESMi

% ETM Energy reported by transformer, ESMi reported by smart meter

NTL = true;

else

NTL = false;

end if;

Classify NS by KNN if it is classified as attack, then KNNout=true, otherwise KNNout=false;

```
        if NTL=true
         if KNNout=false
             counter1=counter1+1;
          TDB1=1;
          if counter1>threshold1
              attack=true;
              TDB1=[]
        end if;
        else if NTL=true
         if KNNout=false
            attack=true;
              break;
        end if;
      else if NTL=false
         if KNNout=false
                attack=false;
              break;
end if;
end;
```

VI. EXPERIMENTAL RESULTS AND DISCUSSIONS

In the first experiment, we train the classifier using both benign and malicious samples. After preprocessing we employ k-means clustering on the benign dataset and study the silhouette plots to find the best value for k. For most customers we observe that k = 1 or 2 provides the best result. Then we implement KNN classification. We analyzed that KNN classifier produced maximum accuracy by compared to SVM classifier. Then we train a KNN classifier with k + 1 classes. Originally, the training set includes 500 benign samples. We use over sampling, in which the members of the minority class are replicated, to make the number of benign and attack samples equal. Table 1 shows that comparison between the performance level of SVM and KNN classification which is specified based on the results produced by SVM and KNN classification shown in fig. 3 and fig. 4.

Classifier	Correct Rate (%)	Error Rate (%)	Sensitivity (%)
KNN Classifier	100	0	100
SVM Classifier	75	25	88

Table 1: Comparison between performance levels of KNN & SVM Classifier

For this experiment we observe the CPBET approach predict the exact genuine and fraudulent customer samples based on the target class.

Target / Predicted	Genuine Customers (Predicted)	Fraudulent Customers (Predicted)
Genuine Customers (Target)	True Positive	False Positive
Fraudulent Customers (Target)	False Negative	True Negative

Table 2: Confusion Matrix - CPBET Theft detection results

A confusion matrix is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known.

There are two possible predicted classes: "Genuine" and "Fraudulent". If we were predicting the presence of a theft in a particular customer, for example, "Fraudulent" would mean that customer is predicted as fraud, and "Genuine" would mean that customer is predicted as genuine.

The classifier made a total of 500 predictions.

Out of those 500 cases, the classifier predicted
"Genuine" 487 times, and
"Fraudulent" 13 times.

6.1 PERFORMANCE MEASURE CALCULATIONS – CPBET DETECTION SYSTEM:

1. $Accuracy = TP + TN\ TP + TN + FP + FN$
2. $Specificity = TN\ TN + FP$
3. $False\ Positive\ Rate = 1 - Specificity$
4. $False\ Negative\ Rate = 1 - Specificity$

Where,

TP- True Positive

FP-False Positive

TN-True Negative

FN-False Negative

6.2. IMPLEMENTATION

K Means Clustering:

Clustering done with cluster number 2 (determ ermined by Silhouette plots) cluster 1 profile will hold 487 records

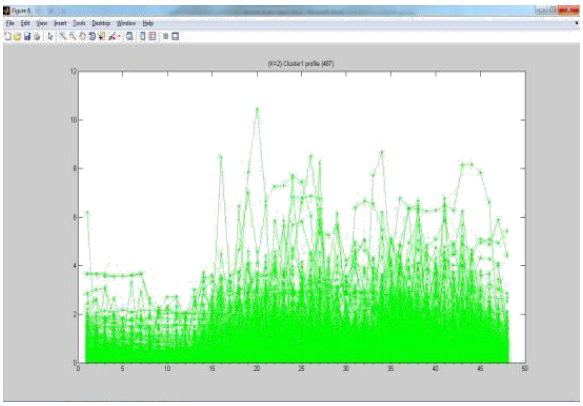

Fig. 4: Pattern representation for Cluster 1 profile number 487

X axis -> Energy usage (Watts), Y axis -> hours (1…48 half an hour)

Clustering done with cluster number 2 (determined by Silhouette plots) cluster 2 profile will hold 13 records.

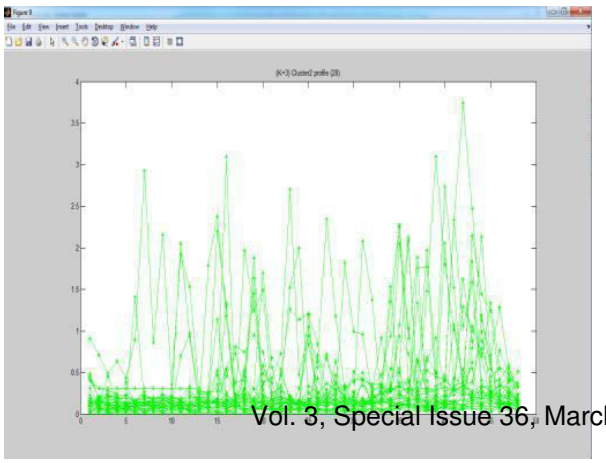

Fig. 5: Pattern representation for Cluster 2 profile number 13

X axis -> Energy usage (Watts), Y axis -> hours (1…48 half an hour)

K Nearest Neighbor Classification:

We cluster our features and prepare the data for identifying the fraudulent customers. As the number of clusters is k, an input, an inappropriate choice of k may yield poor results. Therefore to prevent this problem, we tested classification, KNN classifier has produced higher accuracy shown fig.6. KNN classifier produced higher accuracy by compared to SVM classifier. We can determined the accuracy levels by comparing the results produced by KNN and SVM classification both are shown in fig. 6 and fig.7.

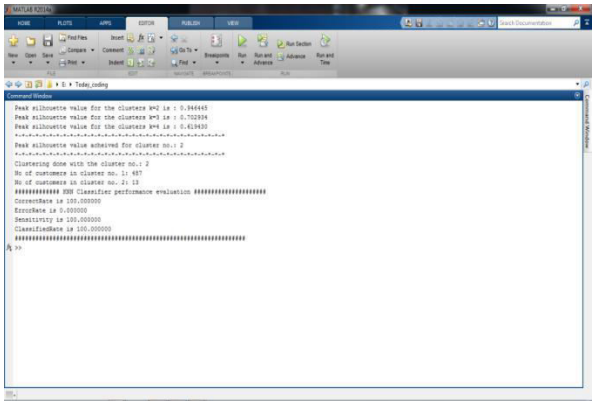

Fig. 6: KNN Classifier Accuracy Results
Correct Rate: 100
Error Rate: 0
Sensitivity: 100

Support Machine Vector Classification:

Fig. 7 describes SVM classifier accuracy levels.

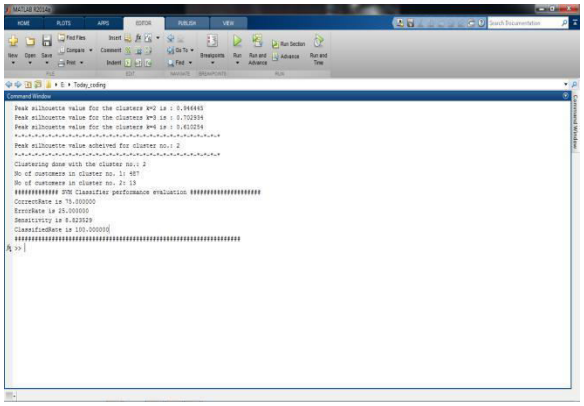

Fig. 7: SVM Classifier Accuracy Results
Correct Rate: 75
Error Rate: 25
Sensitivity: 88

CONCLUSION:

In our work, CPBET Detection System is introduced this system provides practical reduction and machine learning reduction process, which makes the algorithm robust against non malicious changes in consumption pattern as well as data contamination attacks. The result given by the system is very accurately detect the frauders. Along with application of KNN anomaly detector, the algorithm uses silhouette plots to identify the different distributions in the dataset, and relies on distribution transformer meters to detect NTL at the transformer level. By Compared to SVM classifiers, KNN classifier has produced maximum classification accuracy with respect to time.

REFERENCES:

[1] McDaniel .P and McLaughlin .S, "Security and privacy challenges in the smart grid," IEEE Sec. Privacy, vol. 7, no. 3, pp. 75–77, May/Jun. 2009.

[2] Nikovski .D.N and Wang .Z, "Method for detecting power theft in a power distribution system," U.S. Patent 20 140 236 506 A1, Aug. 21, 2014

[3] Rao .P.N and Deekshit .R, "Energy loss estimation in distribution feeders," IEEE Trans. Power Del., vol. 21, no. 3, pp. 1092–1100, Jul. 2006.

[4] Oliveira.C.D et al., "A new method for the computation of technical losses in electrical power distribution systems," in Proc. 16th IEE Conf. Exhibit. Elect. Distrib, Amsterdam, The Netherlands, 2001, pp. 1-6.

[5] Jiang.R et al., "Energy-theft detection issues for advanced metering infrastructure in smart grid," Tsinghua Sci. Technol., vol. 19, no. 2, pp. 105–120, 2014.

[6] Lo.C.H and Ansari.N, "CONSUMER: A novel hybrid intrusion detection system for distribution networks in smart grid," IEEE Trans. Emerg. Topics Comput, vol. 1, no. 1, pp. 33–44, Jun. 2013.

[7] McLaughlin.S, Holbert.B, Fawaz.A, Berthier.R, and Zonouz.S, "A multi-sensor energy theft detection framework for advanced metering infrastructures," IEEE J. Sel. Areas Commun., vol. 31, no. 7, pp. 1319–1330, Jul. 2013.

[8] Xiao.Z, Xiao.Y, and Du.D.H.C, "Non-repudiation in neighborhood area networks for smart grid," IEEE Commun. Mag., vol. 51, no. 1, pp. 18–26, Jan. 2013.

[9] Khoo.B and Cheng.Y, "Using RFID for anti-theft in a Chinese electrical supply company: A cost-benefit analysis," in Proc. IEEE Wireless Telecommun. Symp. New York, NY, USA, 2011, pp. 1–6.

[10] Amin.S, Schwartz.G.A, and Tembine.H, "Incentives and security in electricity distribution networks," in Decision and Game Theory for Security. Berlin, Germany: Springer-Verlag, 2012, pp. 264–280.

[11]Cardenas.A.A, Amin.S, Schwartz.G, Dong.R, and Sastry.S, "A game theory model for electricity theft detection and privacy-aware control in AMI systems," in Proc. IEEE Allerton Conf. Commun. Control Comput., Monticello, IL, USA, 2012, pp. 1830–1837.

[12] Angelos.E, Saavedra.O.R, Cortes.O.A, and de Souza.A.N, "Detection and identification of abnormalities in customer consumptions in power distribution systems," IEEE Trans. Power Del., vol. 26, no. 4, pp. 2436–2442, Oct. 2011.

[13] Depuru.S, Wang.L, and Devabhaktuni.V, "Support vector machine based data classification for detection of electricity theft," in Proc. IEEE Power Syst. Conf. Expo., Phoenix, AZ, USA, 2011, pp. 1–8.

[14] Depuru.S, Wang.L, Devabhaktuni.V, and Nelapati.P, "A hybrid neural network model and encoding technique for enhanced classification of energy consumption data," in Proc. IEEE Power Energy Soc. Gen. Meeting, San Diego, CA, USA, 2011, pp. 1 8.

[15] Depuru.S, Wang.L, Devabhaktuni.V, and Green.R.C, "High performance computing for detection of electricity theft,"

Int. J. Elect. Power Energy Syst., vol. 47, pp. 21–30, May 2013.

[16] Martino.M.D, Decia.F, Molinelli.J, and Fernández.A, "Improving electric fraud detection using class imbalance strategies," in Proc. ICPRAM, vol. 2. Vilamoura, Portugal, 2012, pp. 135–141.

[17] Salinas.S, Li.M, and Li.P, "Privacy-preserving energy theft detection in smart grids: A P2P computing approach," IEEE J.

Sel. Areas Commun., vol. 31, no. 9, pp. 257–267, Sep. 2013.

[18] Mashima.D and Cárdenas.A.A,

"Evaluating electricity theft detectors in smart grid networks," in Research in

Attacks, Intrusions, and Defenses. Berlin, Germany: Springer-Verlag, 2012, pp. 210 229.

[19] Axelsson.S, "The base-rate fallacy and the difficulty of intrusion detection," ACM Trans. Inf. Syst. Security, vol. 3, no. 3, pp. 186–205, 2000.

[20] Cavoukian.A, Privacy by Design: The 7 Foundational Principles, Inf. Privacy Comm., Toronto, ON, Canada, 2009.

[21] Nagi.J, Yap.K.S, Sieh.K, Ahmed.S, and Mohamad.M, "Nontechnical loss detection for metered customers in power utility using support vector machines," IEEE Trans. Power Del., vol. 25, no. 2, pp. 1162–1171, Apr. 2010.

[22] Jokar.P, Arianpoo.N, and Leung.V.C.M, "Intrusion detection in advanced metering infrastructure based on consumption pattern," in Proc. ICC, Budapest, Hungary, 2013, pp. 4472–4476.

[23] Chidambaram.S, Esakkiraj.S, "A Predictive approach for fraud detection using hidden markov model",International journal of Engineering Research and Technology, vol.2,Issue1,2013

[24] Chidambaram.S, Esakkiraj.S, "Comprehensive survey on decision tree algorithm in datamining", Int. conf. on E-Governance & Cloud computing services, vol.1, 2012

[25] Chidambaram.S, Kumar.M, Srinivasagam.K.G,"Feature selection optimization technique using support vector machine classifier",National conference on frontiers in applied science and computer technology,vol.3,Issue1,pages1-6,2015

[26] Chidambaram.S, Kalaivani.R, "Improving classification accuracy using feature selection techniques", Vision 2014 conference,vol.1,2014.