International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 9 Issue: 3
DOI: https://doi.org/10.17762/ijritcc.v9i3.5457
***Article Received**: 09 January 2021 **Revised**: 29 January 2021 **Accepted**: 12 February 2021 **Publication**: 31 March 2021*

Review of Non-Technical Losses Identification Techniques

Kalyan Pal
Assistant Professor
Dept. of Electronics & Communication Engineering
ABES Institute of Technology
Ghaziabad - 201009, U.P., India

Abstract: Illegally consumption of electric power, termed as non-technical losses for the distribution companies is one of the dominant factors all over the world for many years. Although there are some conventional methods to identify these irregularities, such as physical inspection of meters at the consumer premises etc, but it requires large number of manpower and time; then also it does not seem to be adequate. Now a days there are various methods and algorithms have been developed that are proposed in different research papers, to detect non-technical losses. In this paper these methods are reviewed, their important features are highlighted and also the limitations are identified. Finally, the qualitative comparison of various non-technical losses identification algorithms is presented based on their performance, costs, data handling, quality control and execution times. It can be concluded that the graph-based classifier, Optimum-Path Forest algorithm that have both supervised and unsupervised variants, yields the most accurate result to detect non-technical losses.

1. INTRODUCTION

Illegally consumption of electric power, termed as non-technical losses for the distribution companies is one of the dominant factors all over the world for many years. It is a widely spread phenomenon, that also includes missing or faulty meter readings, occurring globally and performed in a variety of ways [1]. In power distribution system the non-technical losses are the problem that affects the whole society as well as the economy of the countries because it reduces the energy efficiency and also the profitability of electrical utilities, which finally affect the genuine consumer. Distribution companies are deliberately trying to identify the electricity theft to reduce the non-technical losses that may have various financial and technical significance. However, despite all possible efforts made by distribution companies to detect the electricity theft, this unlawful incident is still continuing. The distribution companies are generally use the traditional methods, i.e., simple physical meter inspection, but this method have lots of economical, technical and social limitations [2].

There are various reasons for non-technical losses, such as tampering the calibration of the meter, illegal connections from the feeder or from the service-mains, malfunctioning of the meter, irregularities of billing and the unpaid bills. Not only the developing countries like India, but also the developed countries like USA and UK face these problems. As per the World Bank reports almost 50% of the non-technical losses in developing countries are due to theft [3], but it is a reality in developed countries as well. It has been estimated that in all over the world, the utility companies lose more than $25 billion every year due to electricity theft [4]. In India only, the electricity theft is estimated at US$ 4.5 billion per annum which is approximately 1.5% of the country’s GDP [5, 6]. The economies of UK and USA have estimated that the electricity theft is £173 million [7] and US$6 billion [8] every year, respectively. The non-technical losses generally create the unsustainable consequences for human being in already fragile environments.

Power distribution companies generally maintain the database of the customers based on their contact demand, consumption patterns and billing records to maintain the demand and supply chain as well as the billing activities [9]. It is very difficult for the distribution companies to make an efficient decision from this database, because retrieving information from the complex database is extremely time consuming and sometimes it is inaccessible too [9, 10]. Since last few years, the distribution companies in European countries install the smart electronic meters [11] in the consumer premises, it is difficult to tempering this type of meters and also can be easily detected if such types of activities are taking place. But the main drawback of this type of meter is that it is very costly, so not feasible for low voltage distribution in developing countries.

In last few years, various methods and algorithms have been proposed in different research papers for identification of non-technical losses, such as statistical methods [12] proposed by Fourie et. al., decision trees [13] by Filho et. al., artificial neural networks [14] by Galvan et. al., support vector machine [15] by Nagi et. al. Nagi et. al. [16] further proposed the artificial intelligence-based support vector machine (SVM) technique, a hybrid approach of non-technical losses identification for metered customers. This technique is considered to be the advanced version of SMV based approach [15]. Ramos et al. [17] proposed the graph based optimum-path forest (OPF) classifier, another supervised approach for non-technical losses identification, in the context of Brazilian distribution system. Some other data mining studies for fraud detection, such as Rough Sets [18], Extreme Learning [19], Knowledge Discovery in Databases (KDDs) [20], Extreme Learning Machine (ELM) [21], Statistical-based Outlier detection classifiers [22], have been proposed by the distribution industries in recent years.

Another method to detect non-technical losses can be done by energy balance calculation [23], for that, topological information of the network is required. But this method is not helpful to calculate the non-technical losses due to following reasons: i) in developing countries, network topology undergoes continuous changes in order to satisfy the fast-growing demand of electricity, ii) infrastructure could break and result in wrong energy balance calculations and iii) it needs transformers, feeders and connected meters to be read at the same time.

Over the years many techniques have been proposed and implemented for fraud detection in electrical power distribution. Some techniques are not made publicly available to avoid security breaches by hackers [24]. However, other published techniques are available, specifically in the data mining, machine learning and statistical analysis of distribution network. Each technique has its unique features as well as merits and demerits. In this paper some popular on-technical losses identification techniques have been comprehensively reviewed and tried to identify the best one to detect abnormal behaviour in electricity power consumption.

2. CATEGORIZATION OF NON-TECHNICAL LOSS IDENTIFICATION METHODS

From various research papers on non-technical losses identification techniques, it is clear that there is no common methodology that can be implemented for fraud detection. Methodologies with different field of knowledge has been adopted by the Researchers, specifically machine learning, anomaly detection, distribution network analysis and also intrusion detection. The non-technical losses identification methodologies can be segregated in three broad categories, i.e., data oriented, network oriented and hybrid. In data-oriented methods the consumer related data is used, further in network-oriented methods data from network topologies is used, whereas in hybrid method data from both categories is used. The complete categorization of non-technical losses identification methods is shown in the figure 1.

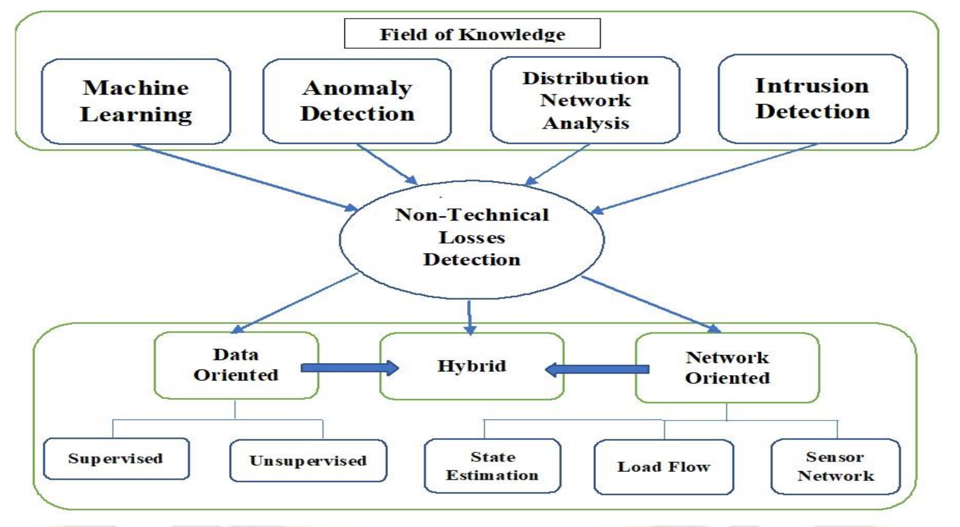

Fig.1 NTL identification methods categorization

There are two subcategories of data-oriented method, these are supervised and unsupervised method. All the data in supervised method is labelled and the algorithms are made to predict the output from the input data. A supervised learning algorithm analyses the training data and produces a definite function, which can further be used for mapping undefined data. The algorithm will correctly determine the class labels for unknown data**.** Before that he learning algorithm should be generalized from the training data to unknown situations in a “reasonable” way. Some classification algorithms come under supervised method are ‘Decision Trees’ [13], ‘Support Vector Machine (SVM)’ [15], ‘K-Nearest Neighbours’ [25], Random Forest’ [26]. On the other hand, in unsupervised method all data is unlabelled and the algorithms learn to inherent structure from the input data. The main problem of an unsupervised learning task is trying to find hidden structure in unlabelled data. As the unlabelled examples are given here, so there is no error or reward signal to evaluate a potential solution. ‘K-Means’ [27], ‘Hierarchical’ [28], ‘Hidden Markov Model’ [29], ‘Gaussian Mixture Model’ [30] are the classification algorithms comes under the unsupervised method.

In network-oriented methods of non-technical losses identification labels are generally neglected, because they are based on network analysis and there are definite physical rules that are used to describe the systems [31]. State estimation, load flow, sensors are the categories under network-oriented method of non-technical losses identification that follow the algorithm used to detect the frauds.

In hybrid method of non-technical losses identification, the concept of both data-oriented and network-oriented methods are combined [32]. For example, a state estimation method may be used on medium voltage level to detect non-technical losses at medium/low voltage level transformer. Once the location of the network with non-technical losses are identified, a supervised classification method can be used to restriction-technical loss at consumer level.

For analysing the NTL identification, typically following parameters should be considered:

- Algorithm: This is the most important parameter for non-technical losses identification. Different algorithms are introduced in different research papers over the years for NTL identification. Generally, in most of the cases more than one algorithm are used. The complete list of all algorithms used for each work are mentioned here, however, some of them are used for comparative studies only. Details are described in Chapter-5.
- Data types: The data is the essential parameter required by each method for NTL identification. This is a critical parameter when designing an NTL identification method or choosing from existing ones. The researchers select their NTL identification system according to the availability of data and their types. Details are described in Chapter-3.
- Data set size: The size of the data set used for the analysis of the NTL identification system is determined by the number of consumers (simulated or not) implicated. Data set size is taken into account large for more than 1000 consumers, medium for 100–1000 and small is for less than 100. The data set size is important, since it provides feedback on the scalability of NTL identification systems.
- Features: In most of the cases the basic data (described as data types above) are first processed in order to extract features to be used for classification. Although a lot of researchers use features for identifying on-technical losses, there is no suggestion of which features should be used. The authors list features and associate them with data types and algorithms, thus making it easier to choose appropriate features either using domain expertise or feature selection algorithms. Details are discussed in Chapter-3.
- Metrics: Performance metrics are used to assess the performance of NTL identification methods under various circumstances and to compare systems. A number of metrics are mentioned in literature. The authors' goal is to supply a full list of metrics, beneath a unique identifier, together with the reason they should (or shouldn't) be used for. Details are discussed in Chapter-4.
- Response time: The time required for an NTL identification system to decide if a consumer commits fraud. This should not be confused with the time it takes for a classifier to establish a result given the relative input data (which is extremely dependent on machine and coding). In contrast, the response time depends on the time needed to get the input data.

3. DATA TYPES CATEGORIZATION AND DEFINITIONS

In this chapter the various data types used in literature are organized in broad categories. The main reason for this categorization is to make sure that researchers are not restricted to specific information to choose their algorithm, however they can choose their NTL identification system as per the accessible data. The data type groups are presented in Fig. 2.

3.1 Raw data used in NTL identification

According to the location of their physical source the data are initially organized. Data concerning individual consumers (for example, active energy measurements) are classified as "Consumer Level" data, while data concerning an area (i.e., network topology) are classified as "Area Level" data. Data, associated to both classes, are further categorized as time series and static data. Literature review reveals that there is a gap in the use of area specific static data (mainly those not related with the network topology). In fact, only one work [33] uses area level data for NTL identification. In addition, high resolution energy data and environmental data i.e., temperature etc are rarely used. Most of the data-oriented algorithms use consumer level time series and static data exclusively. Moreover, all methods use time series energy consumption data and more than half of them include static data. Network oriented methods can be implemented for high or medium resolution energy consumption data, as well as voltage and current measurements. In addition, they significantly depend on the data, measured from other devices i.e., observer meters and information of MV or even LV network topology. In hybrid methods every possible types of data can be used.

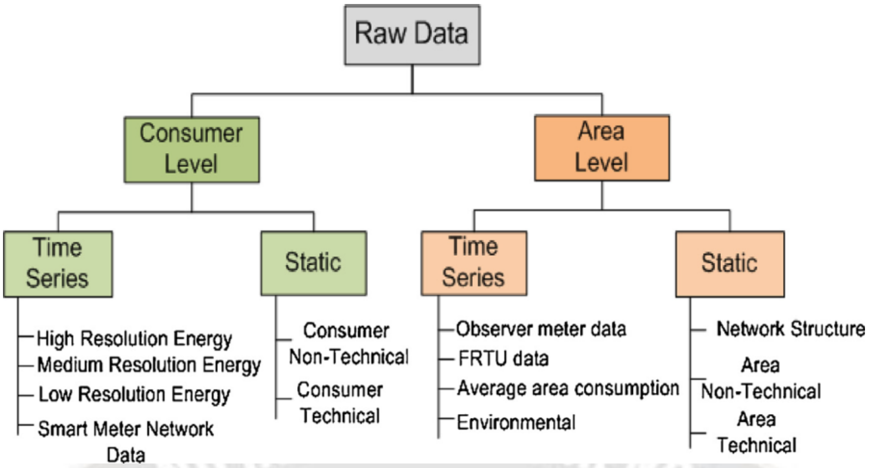

Fig. 2 Data type categorization for NTL detection applications

3.2 Features used in NTL identification

Generally, for data oriented as well as for hybrid non-technical losses detection methods not only the raw data described above but also features extracted from that data are used. Features commonly used are listed here, which are mainly calculated from consumer level time series data and more specifically from active energy consumption graphs. The time resolution of the raw data thus significantly influences the time resolution of these features. So, it is difficult to define precisely all the features used in each work. Rather the authors prefer to define the families of features by presenting in a more generalized view of the domain. The list of features, most commonly used in literature, represented in Table-1.

The use of features itself is common in data mining (classification or clustering) tasks than the use of time series data. By reducing the data set's size and providing a level of anonymization it is possible to improve the Classification metrics. Initially, a large number of features can be extracted from a time series, however features representative of NTL identification only should be considered. In addition, the duration of time period within a day should be chosen carefully for which a feature is supposed to be calculated, e.g., a daily average of 30 min useful energy data for few years yields different information from a periodic (3 months) average.

Given a list of features and the NTL identification method, feature selection algorithms can be used for defining an optimal set of features. A large number of feature selection algorithms are available in machine learning literature [34]. There are cases though were feature selection has been studied for optimizing the performance of classifiers used in NTL identification. Such techniques have been used in Refs. [17, 35–39]. In Ref. [17] the small number of features permits testing all possible combinations. Authors in Ref. [35] use a number of different heuristic methods, focusing on binary black hole optimization and comparing it with Harmony Search (HS), Particle Swarm Optimization (PSO), Differential Evolution (DE) and Genetic Algorithm (GA). In Ref. [36] the use of Harmony Search is evaluated and compared to PSO. The same authors [37] evaluate the Binary Gravitational Search Algorithm (BGSA) and compare it with PSO and HS. Social Spider Optimization is utilized in Ref. [38] both for feature selection and parameter tuning. In Ref. [39] the authors propose filter and wrapper methods for feature selection without further specifying the type of algorithms used.

Table-1: Main features used for NTL detection

Feature name	Description
Average, Max/Min, Standard Deviation	Standard statistics calculated for a specific time period.
Power/Energy factor	The power factor is defined as the rate of active (kW) to reactive power consumption (kVAR). Instant power measurements are required for this calculation. High resolution (less or equal to 15 min) data must be used for a good estimation. Energy factor is the reactive energy (kVARh) consumed in a time period to the active energy (kWh) consumed in the same period
Load factor	The ratio between the average active energy consumption (kWh) to the maximum active energy consumption (kWh) for a specific time period (for example a month).
Streaks	The number of times the consumption curve goes above and below a mean line (defined as a moving average of the consumption curve).
Daily consumption to contracted power	The sum of active energy consumption in a period (kWh) to the contracted power (kW)
Pearson coefficient	The Pearson coefficient of the active energy consumption curve in a specific (typically large) time period. The Pearson coefficient measures how well a linear equation describes the relation between active energy consumption and time.

Billed-consumed energy coefficient	Difference of energy billed (kWh) to consumed active energy (kWh) divided by the contracted power (kW).
Predicted kWh	A prediction of the active energy consumption (kWh) given by any forecast model or the difference of this prediction and the measured value.
Wavelet coefficients	The difference of the Wavelet coefficients calculated from the consumption curve to be classified and the Wavelet coefficients of previous years consumption curves.
Fourier coefficients	The difference of the Fourier coefficients calculated from the consumption curve to be classified and the Fourier coefficients of previous years consumption curves. In addition, the phase of the first five Fourier coefficients of the active energy consumption curve can be used.
Polynomial fit coefficients	Difference of coefficients of the polynomial that best fit the consumption curve to be classified and the coefficients of the polynomial that best fit previous years' consumption curves.
Euclidean distance to mean customer	Euclidean distance of an active energy consumption curve to a consumption curve calculated as the mean consumption of all consumers in the data set.
Consumption curve slope	The slope of the linear equation that best fits the active energy consumption curve time series.
PCA components	A number of components that are calculated from Principal Component Analysis (PCA) or Kernel Principal Component Analysis (KPCA) on the active energy consumption curves. Not all of the components need to be used. The mean of specific components may be used to.
Fractional order dynamic errors	Features that express the difference between a profiled meter usage and a real time consumption time series.
Mismatch ratio	The difference between consumption measured in the MV/LV transformer and the sum of smart meter measurements and estimated technical losses divided by the nominal power of the substation.
Seasonal consumption rates	Total consumer consumption (kWh) in a specific season (for example winter) to consumption (kWh) of another season (for example summer). Total consumer consumption (kWh) in a specific season (for example winter) to the average consumption of consumers on the same substation at the same season (for example winter).
Discrete Cosine Transform coefficients	The k first coefficients of the discrete cosine transformation.
Consumption decrease compared to previous period	A reduction of x% in consumption during a time period of length T in comparison to a past time period of the same length or compared to the average.
Estimated readings	Number of meter readings that are estimated by utility due to inability to access the meter.

4. NON-TECHNICAL LOSS IDENTIFICATION PERFORMANCE METRICS

A well-defined and updated list of performance metrics is important, primarily for comparing NTL identification methods. A list of most commonly used metrics in literature is provided in Table-2. Most of the metrics used to calculate the traditional confusion matrix, here TP are true positives, TN are true negatives, FP are false positives and FN are false negatives. P represents all positive samples (P = TP + FN), while N represents all negative samples (N = TN + FP). P(I) is the probability of NTL occurrence.

The first seven (accuracy, detection rate, precision, false positive rate (FPR), true negative rate (TNR), false negative rate (FNR), F$_1$score) metrics are frequently used in classification tasks and calculated from the confusion matrix. In NTL identification literature the most common metrics are the accuracy and detection rate, which appear in almost every case of data-oriented methods. Increase in accuracy reveal that the system generally works better by classifying both positive and negative samples accurately. It is not enough though, when dealing with an imbalanced data set, where one class (negatives) is excessively larger than the other (positives).

The second most commonly used metric is the detection rate (DR), also known as recall, true positive rate, success in detecting NTL or hit rate. This metric expresses the proportion of samples classified as NTL to the total number of NTLs in the dataset. Typically, large values of DR imply a well operating detection system, but for this to be true other metrics must be taken into account. Usually, both detection rate (DR) and accuracy has to be considered to evaluate the system's performance.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 9 Issue: 3
DOI: https://doi.org/10.17762/ijritcc.v9i3.5457
***Article Received**: 09 January 2021 **Revised**: 29 January 2021 **Accepted**: 12 February 2021 **Publication**: 31 March 2021*

Table-2: List of metrics used to evaluate NTL detection methods

Metric	Definition
Accuracy	$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}}$
Detection rate (DR)	$\text{DR} = \frac{\text{TP}}{\text{TP}+\text{FN}}$
Precision	$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}}$
FPR	$\text{FPR} = \frac{\text{FP}}{\text{FP}+\text{TN}}$
TNR	$\text{TNR} = \frac{\text{TN}}{\text{FP}+\text{TN}}$
FNR	$\text{FNR} = \frac{\text{FN}}{\text{FN}+\text{TP}}$
F_1 score	$\text{F}_1 \text{ score} = \frac{2\text{TP}}{2\text{TP}+\text{FP}+\text{FN}}$
AUC (Area Under Curve)	The area under the ROC (Receiver Operating Curve) of the binary classifier.
Recognition Rate	$\text{Rec. Rate} = 1 - 0.5 \times \left(\frac{\text{FP}}{\text{N}} + \frac{\text{FN}}{\text{P}}\right)$
Bayesian Detection Rate	$\text{BDR} = \frac{\text{P(I)} \cdot \text{DR}}{\text{P(I)} \cdot \text{DR} + \text{P}(-\text{I}) \cdot \text{FPR}}$
Training time (s)	The time (s) required to train an NTL detection system.
Classification time (s)	The time (s) it takes for an NTL detection system to classify an instance.
Cost of undetected attack	Defined as the cost of the worst possible undetected attack.
Energy balance mismatch	Defined as the difference between the sum of consumer level active energy and substation level active energy
Average bill increase	Defined as the average bill increase if the NTLs were distributed among all consumers.
Normalized labour cost	Defined as the cost for inspecting all cases classified as NTL by the detection system.
Anomaly coverage index	Defined as the ratio between anomalous consumers under RTUs and the total number of anomalous consumers.
RTU cost	Defined as the total cost of acquiring RTUs
Minimum detected deviation	Defined as the minimum deviation (from a pre-specified typical profile) that can be detected.
Decrease in electricity stolen	The decrease of stolen electricity when a specific FDS is applied.

The precision and false positive rate (FPR) are also two most commonly used metrics. Precision is calculated as the number of NTLs detected divided by the total number of NTL alarms; it is also known as positive predictive value (PPV). When the precision is increased, it means that the samples that are identified as positive, is actually affected by non-technical losses. It must be noted here that precision and recall metrics are counteracting to each other, i.e., increase in one metrics will decrease the other, therefore, proper balance between the two metrics is important. This balance is expressed by calculating the F_1score (a particular case of F-measure or Fβscore were = 1) which is the harmonic mean of detection rate and recall. Increased F_1score values indicate that the system detects many frauds with low false alarms. Although rarely used in NTL identification literature (examples include Refs. [39, 40]), this is one of the most important and indicative metrics, especially when dealing with class unbalanced problems (like NTL detection and generally detection of frauds). In fact, a lot of work has been published on this problem, proposing a number of solutions and performance metrics for reliable evaluation of classifiers [41].

FPR may be calculated by the total number of samples which are falsely classified as positives (false alarms) to the total number of negative samples. This is one of the most important metrics, because more the false positive samples, larger will be the operational costs due to unnecessary meter inspections by the organization, responsible for the NTL detection. In general, it is desired that FPR values should be as low as possible, although the threshold value depends on the relative size of the two classes. Assuming a data set of 1000 consumers of which 10 commit fraud an FPR as low as 10% would mean that 99 consumers without NTLs are classified as positives (NTL), leading to 99 additional meter inspections in order to detect 10cases of fraud. FPR = 1% means that 10 false alarms would occur in detecting the 10 real cases of fraud.

Related with the class imbalanced problems, like NTL detection, it is very important to choose the appropriate metrices. Combinations of metrics should be used in this case including accuracy, DR, FPR and TNR. Another metric, not frequently used in NTL detection literature, is the Bayesian Detection Rate (BDR) [42]. It mainly applies to data-oriented methods and is dependent on the probability of fraud P(I), DR and FPR while it expresses the probability of a false alarm in real life conditions. DR and FPR are characteristics of the classifier used, however the probability of fraud is an external parameter. In the fraud and intrusion detection domains this parameter usually attains small values, because the fraud is not a frequent phenomenon. Given the small value (for example 1%) of P(I), in order to achieve a high value for BDR (i.e., to minimize false alarms) , FPR should be as low as possible, it will not matter for higher value of DR.

5. ALGORITHMS USED IN NON-TECHNICAL LOSS DETECTION SYSTEMS

Every fraud detection technique is unique itself according to their approach of using different data in different ways. In some systems a simple structure has been presented (for example a single Support Vector Machine (SVM) for consumer classifications), Whereas some systems may be more complex (such as prior data cleaning and clustering phases, classifier ensembles, power system analysis etc.). Each technique can be characterized using a small number of algorithms which develop the base of the fraud detection method. Many of them are already being well established and defined in other research papers hence detailed descriptions are avoided here. Non-technical losses identification methods are mainly categorized as 'data-oriented' and 'network-oriented', another method which has the combined characteristics, is called as 'hybrid' method is also considered in some cases.

5.1 Data-oriented methods

Data-oriented methods are mainly performed by data analysis and machine learning techniques. The data-oriented methods are classified in two major categories, i.e., supervised and unsupervised, are discussed here. The common methodology of data-oriented methods, for both supervised and unsupervised approaches, are shown by a flowchart in Fig. 3 and described below.

- Data processing & model selection: The data-oriented method for fraud detection should be chosen when a set of raw data is available. The choice of supervised or unsupervised methods are governed by the availability or unavailability of labelled data, whereas data quality and data variety dictate the algorithm to be used. The choice of algorithm may exclude some parts of the raw data (data selection phase). After selecting the algorithm, the data cleaning is performed, which is common in the knowledge discovery process and if necessary, then the feature extraction is done.
- Modelling: In supervised and unsupervised methods different process are followed for modelling. The labelled data in unsupervised models are used in evaluation process only and not for training purpose. On the other hand, in supervised models the data set is split into two parts – for the training purpose as well as for the testing purpose. For training the model, generally feature selection is used, but before that the valid training set is required to be defined. While parameter optimization makes use of metrics that can be calculated due to label availability.
- Application: Once the model (supervised or unsupervised) is defined by the labelled data set, the performance of the model can be verified using new data but that should not belong to the 'Raw Data' set. Classification results are further processed for preparing a suspect list that contains the probability of each consumer committing fraud. In this phase a major operation of the non-technical loss detection model can be referred.

The supervised and unsupervised classifiers are developed on the basis of Artificial Intelligence (AI) methods which is a well-established technology and easily found in the various research papers. In the following sections the application of supervised and unsupervised methods for non-technical losses detection problem are discussed.

DOI: https://doi.org/10.17762/ijritcc.v9i3.5457
***Article Received**: 09 January 2021 **Revised**: 29 January 2021 **Accepted**: 12 February 2021 **Publication**: 31 March 2021*

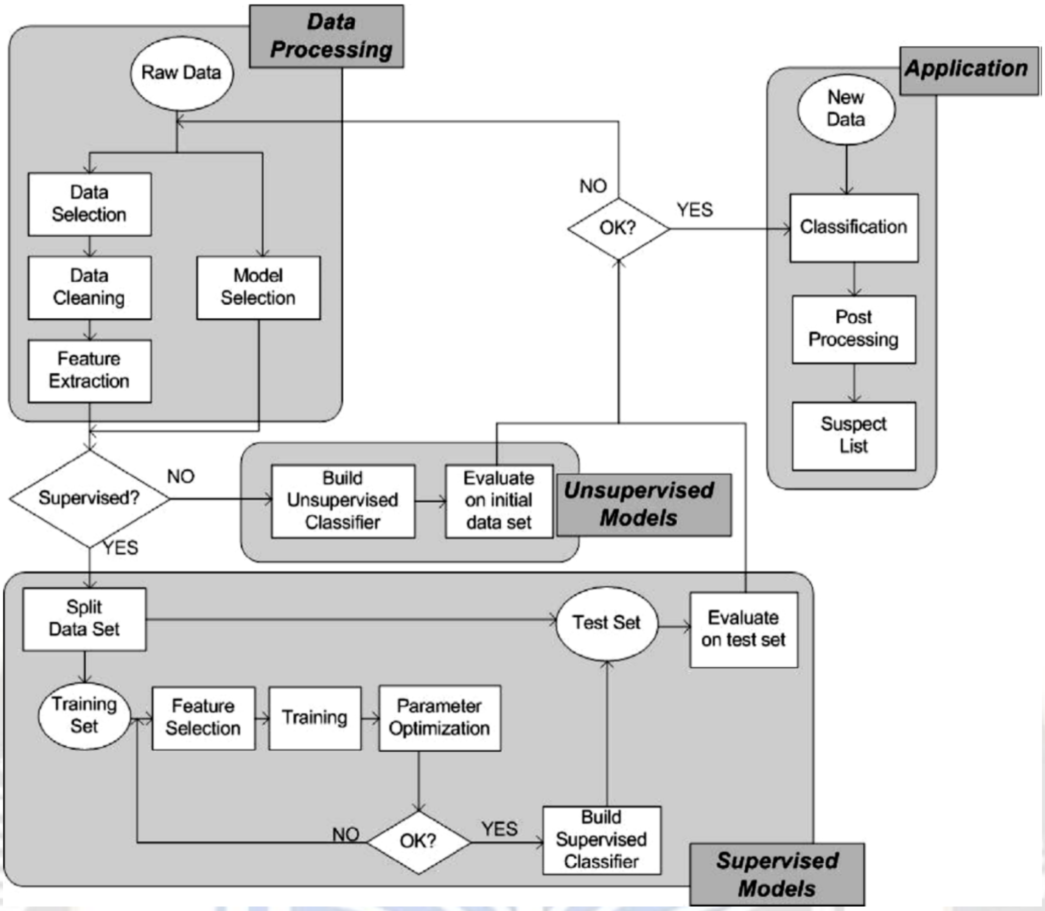

Fig. 3 Flow-Chart of Data oriented methods

5.1.1. Supervised methods

5.1.1.1. Support Vector Machine (SVM) [15,16,36–40,43]. It is normally used as a binary classifier because for the class imbalance problem it becomes volatile. Various forms and their applications are already present in the literature, like the One-Class SVM and the Cost-Sensitive SVM. Where the One-Class SVM is proposed for anomaly detection i.e., unsupervised learning, because this method is trained on samples belongs to the negative class only, i.e., the fraud consumers. On the other hand, different weights to different types of classification error can be assigned in the Cost-Sensitive SVM method. For example, it is possible to assign a high cost to misclassifications of the minority class which can lead to higher performance metrics i.e., low FPR, high BDR. In various literature SVM method is preferred with full trust for detecting non-technical losses, although it is difficult as well as time consuming to implement. Again, apart from the Liner kernel (Linear-SVM), the Radial Basis Function kernel SVM or SVM-RBF is also popular for non-technical losses detection. In the case of SVM-RBF, the cost and gamma (γ) parameters need to be tuned, whereas in case of Linear-SVM only the cost must be tuned. Generally, the grid search algorithm with cross-validation is used for this purpose. In SVM method, because of tuning, the time for constructing the model is increased when large data set is considered, so it should not be considered specially for online applications. In order to enhance the classification results, it is a general practice to combine the SVM with other classifiers such as Fuzzy Inference System, Decision Trees, Neural Networks etc and there should be a common baseline solution for comparing classification and feature selection techniques.

5.1.1.2. Artificial Neural Network (ANN) [21,37,44-49]. It is usually used to forecast the time series of electricity consumption. Now a days the Multi-Layer Perceptron (MLP) trained with back-propagation (BP-MLP), one of the common versions of Artificial Neural Network, is used as a binary classifier for non-technical losses detection. In this method the difference between predicted value and measured value is used to detect the frauds. Here the structure of the network is supposed be chosen before training. It mainly follows the trial-and-error approach in order to identify the number of hidden layers and corresponding number of neurons. To generalize the model, cross-validation is important in this method. The Extreme Learning Machine (ELM), another type of neural network, has been proposed in some literature for binary classification apart from the BP-MLP. It has a single hidden layer where weights connecting the input layer with the hidden layer are randomly assigned, although weights between the hidden layer and the output can be calculated in a single step. Thus, without compromising in performance, it can be trained faster. In addition, online sequential extreme learning machine(OS-ELM) has been proposed for applications where the model needs to be retrained online due to possible changes in consumer characteristics. But in all models, the only parameter that need to be consider is the number of hidden layer neurons.

5.1.1.3. Optimum Path Forest (OPF) [17,35-39,50].This is a graph-based algorithm which can be used for classification applications. Unlike other algorithms, in this method each labelled sample of the training set is considered as a graph node with coordinates being the feature values. OPF does not try to find the optimal hyperplane that separates two classes. In this algorithm the clustering is done by creating two or more trees of the graph, called optimum path trees, each represents a class and the collection of trees are called the OPF classifier. OPF is capable of handling overlapping classes and has low training time, which allows for online training of the fraud detection system. Such a characteristic is important in case the testing samples substantially differ from the training samples used.

5.1.1.4. Rule induction [4,16,45]. Here the defaulter users are identified by various pre-conditioned rules. These rules can be implemented by expert knowledge and statistical analysis. There may be some cases where such expertise is missing or inadequate. However, in labelled data, the hidden rules can be extracted by this technique. The main goal of this technique is to predict the class of a sample given the values of other related features. This technique can be considered fully authentic, although for detecting the fraud in electricity consumption various works use this expert system. Sometimes for better emulation of the reasoning process in expert system, the fuzzy inference system can be introduced. The rule-based systems can be combined with other classifiers like Support Vector Machine, Decision Tree, Bayesian Networks etc. so that an improved system of fraud detection can be developed which is can be used to utilizes human expert knowledge.

5.1.1.5. Decision trees (DT) [39,40,49-53]. This technique was not so popular for detection of non-technical losses. Basically, it is a classifier and output of this classifier is a set of processes which are used to classify new samples. It is possible to explain the characteristics of the non-technical losses with the help of these processes. The processes of the decision trees can also be combined with the processes defined by the experts and other classifiers. It should be noted that the decision trees are sensitive to the class imbalance problem and highly dependent to the training set. There are some types of decision trees that are used in non-technical losses detection areC4.5, CART and QUEST. Decision trees are also able to handle the categorical variables with the data i.e., type of consumer (residential, industrial or commercial) or contract details etc.

5.1.1.6. Nearest neighbour (k-NN) [36,37,54]. Within the supervised classification algorithms for non-technical losses detection, the Nearest neighbours are the simplest methods which is mainly used a baseline for comparisons with other algorithms. In k-nearest neighbours, it is very common to place a new sample on the feature space and assign it to the class. The only parameter that requires tuning is the number of neighbours voting (k), but the way distance between samples is measured affects algorithm performance.

5.1.1.7. Bayesian classifiers [53,55]. These are the probabilistic classifier based on the assumption of strong independence between features. Here the prior knowledge of NTL probability is essential for this algorithm. The system is able to learn the probability of each appliance used by individual consumer [55], which is possible by using non-intrusive load monitoring (NILM). For arrival of each new sample, NILM is performed freshly and the probability of theft can be calculated. Such systems require a priori knowledge (probability of theft and conditional probabilities) and it may influence the classifier’s output. The Bayesian network [53], another type of Bayesian classifier, represents the joint probability (Bayesian probability) of a set of variables in the graphical mode. The major advantage of Bayesian networks is that they can be interpreted easily. It will be helpful to identify the features of a time series that will be most influenced by the NTL. The main objective in this case is to learn the conditional probabilities by giving a fully labelled data set along with Bayesian network structure. The class of a new sample may then be inferred together with the probability of the sample belonging to the predicted class. Therefore, a high threshold is chosen if it is not possible to perform a large number of inspections, by selecting the number of meters to be inspected, because a probability threshold may be set according to the fixed distribution system of user’s needs and business characteristics.

5.1.1.8. Generalized Additive Model (GAM) [33]. This model is used for the geographical distribution of NTL. In this method it has been assumed that NTL spreads over an area according to various social and technical concern; this concept has been taken from the field of epidemy related study. The probability of NTL can be estimated by GAM for a set of consumers that may or may not be linked with NTL. Next, a Marko chain is used to model how NTL may spread over a specific area in the future. During designing the distribution system for long term systems including hardware installation, legislation change, campaigns etc, this method helps to compute the geographical distribution of the probability of fraud, but it cannot be used to detect the fraud directly.

5.1.2. Unsupervised methods

5.1.2.1. Self-Organizing Map (SOM) [17,36,37,56]. SOM can be considered as a special type of Neural Network frequently used as a clustering or unsupervised classification tool. It usually produces 2D representations of the data set and therefore serves as a size reduction. Another advantage of SOM is that it produces data displays that people can understand. The final output in a set of clusters that needs to be professionally tested or fed to a second level of concept to be determined by a set of meters to be tested. The number of collections may be more than 2, if the general character is considered which can create clusters but not frauds. The unsupervised status of the SOM leads to a reduction in classification and perhaps this is the reason why it rarely appears in NTL acquisition documents.

5.1.2.2. Clustering algorithms [54-59]. Clustering algorithms have been applied in many fraud detection activities, mainly for the pre-processing of data. In this method, similar types of data are group together for example non malicious behaviour of consumers, and then train classifiers on these groups. In this classification process has been enhanced and it also reduce the false positive data. Clustering can also be used for calculation of baseline power profiles or its prototypes. If a new sample data conflict with these profiles significantly, then fraud may be identified.

The clustering algorithm is also used as tool for unsupervised classification. Fuzzy clustering has also been used [59] by combining each new sample with a possibility of fraud, rather than a label. This allows the fraud detection system developer to tune the system according to business models and other external parameters.

The Density Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm has also been proposed based on the clustering algorithm. It is used after applying Principal Component Analysis (PCA) on the electricity consumption data and representing each consumer with the first two components in a two-dimensional space. This depiction enables the DBSCAN to separate odd from unnatural consumers in an efficient manner and provides analytical proficiency [58].

5.1.2.3. Expert systems [45,60,61]. As the name implies, this system depends on the reports of technical experts who are responsible for tracking the non-technical losses. Although such systems do not require learning, they are considered unsupervised (fuzzy rules are also included, usually after fuzzification of simple rules). Rules may be defined by various means, but most of the times an accurate model of how the non-technical losses are expressed or inspector knowledge is required, if required a rule can be developed that if the consumption is more than a certain percent for example 50% then the concerned meter of the consumer must be checked. Another rule is applicable that if the power factor is less than 0.5 then the meter and the load pattern must be checked [45]. Although these rules are very simple but still very effective for classification process.

5.1.2.4. Statistical control [62–64]. Another very effective method of non-technical losses detection is the statistical process control method. Control charts, typical for time series data, are used for observing any individual utilization and for explaining the regions, where the time series may be considered abnormal. The XMR (moving range) [62] control chart classifies the time series variability that may occur due to theft by defining control limits. The XMR control chart check both the actual consumption (X chart) and their variable range (MR charts). Rules has been formed to indicate which results are frauds and require survey.

There are lots of other charts also, such as exponentially weighted moving average (EWMA) control chart and the nonparametric cumulative sum (CUSUM) control chart [63]. These charts are very popular in industries for quick detection of abnormality by providing the visual inspection of data. But with quick detection there may be lots of false positive data which may affect the performance of the system. Bollinger bands [64] is another statistical data analysis tool used in stock markets for change in data detection. As the main purpose of this method is to detect changes, so they fail to detect fraud, if it takes place from the beginning of the monitoring period. It can be a major drawback for power theft detection. Also, such method may interpret other types of consumption change as fraud, and hence may produce false positives.

5.1.2.5. Regression Models [63,65]. Regression model is a predictive modelling technique which review the similarities between dependent and independent variables. This model is used for forecasting, time series modelling and developing the relationship between the variables. This model is used to detect the non-technical losses by comparing the measured value and the forecasted value, and it has been assumed that forecasted values have been trained with non-malicious data. Here it may be assumed that largest the difference, the highest the probability of fraud.

5.1.2.6. Outlier detection [63,66]. Outliers are the values that vary from other data during an observation, they may indicate a variability in a measurement, experimental errors or a novelty. So, an outlier detection may be considered as an observation that diverges from an overall pattern on a sample. Outlier detection method is a very effective method of non-technical losses detection. Considering a data set free of frauds, each cluster of samples is modelled as a Gaussian distribution. Now when a new sample has to be modelled, the probability of the sample belonging to each of the cluster has been calculated. The result is then compared to a threshold value to check whether the sample is an anomaly or not. In this technique the major challenge is to decide the number of clusters and of course the parameters of the Gaussian distributions. In two popular algorithms, K-means and Expectation–maximization (E-M), this method is used. The Optimum Path Forest algorithm may also be proposed based on this concept. This method does not require a clean data set and the main advantage of outlier detection method is that it may still detect frauds even if they are already included in the training set.

5.1.2.7. Gaming theory approaches [67,68]. Game theory is used to model the malicious user and the fraud detection system behaviour like modelling the attacker and defender behavior [67]. Its use as a core part of an FDS is not yet mature though. Attempts include decision making mechanisms modelling fraud as a cooperative or non-cooperative game.

5.1.3. Qualitative comparison

The supervised learning methods are mostly preferred in data-oriented fraud detection systems, because of their better performance. However, these methods require the labelled data from malicious and benign users for classification, but that are not always available or are not representative of all the data, the fraud detection system is targeting at. Finally, even if labelled data exist, it is most probable that class imbalance problems will arise. Unsupervised methods do not require any labels, except in case of anomaly detection methods, partially it is required. Therefore, these can be applied easily with lower performance, characterized by a higher false positive rate. Unsupervised methods can also be used in the case where large number of negative samples are available, but positive samples are not or very few available. Usually some of the anomaly detection algorithms, such as outlier detection method, never makes any assumption of class labels for training the classifier. Another advantage of unsupervised methods is their resilience to zero-day attacks. On the other hand, the supervised methods are preferred to detect specific types of anomaly. In case a new fraud behavior appears, which is missing from the training set, the supervised classifier will probably fail to detect it. In contrast, unsupervised methods are independent at least from the positive class training set.

5.2. Network-oriented methods

In network-oriented methods, to detect the fraud, data has been collected from the distribution grid meters and then applied the physical rules that govern the underlying electrical network. Apart from the data from distribution grid meters, they make use of network related data, such as network topology and transformer and/or phase connectivity of the consumer. Many researchers use power flow tools to estimate the size of non-technical losses and identify the source of these losses by analyzing the energy balance with the observer meter. Again, some approaches may be more accurate as they use distribution state estimation and bad data detection, although which is not possible in all cases. It has also been proposed in some analysis to use the dedicated sensors in order to identify the fraud.

5.2.1. Load flow approach [69-74]. Another process of detecting non-technical losses is the energy balance of a network. This can be done by installing the meter in LV side of the network, this is also called as observer meter. Using this observer meter, it can be verified the sum of energy consumption of all consumers connected to a specific network. Considering a certain percentage of losses as the technical loss, the supplied and consumed energy can be measured and if there is any mismatch that can be treated as non-technical losses or frauds. In this method it is not possible to identify the individual fraud consumer but the method can be implemented up to secondary substation level for detecting the frauds. The privacy issues are also taken into account by solving the problem in a distributed fashion [69]. In Ref. [70] the authors model the meter behaviour and model parameters are calculated via various methods. This model parameters are then compared to that of a standard meter and the difference shows the fraud in the system. Finally, for cases where technical losses or network structure are unknown, a different methodology has been proposed for identifying network parameters and then calculating technical losses which leads to better calculation of non-technical losses [71].

A probabilistic approach of power flow for detecting the frauds is mentioned in Ref. [72]. It can be identified that whether the fraud occur under a specific observer meter or not by using different meter for different feeder in the network. Energy balance concept is again used and it is in a probabilistic manner. Calculating the probability distributions for total and technical losses and subtracting them from the difference between input and output by convolution gives the probability of non-technical losses occurrence in a specific sub-network.

Again, a smart substation concept has been proposed [73], where both smart meters and observer meters should be available. Here the first step is to check the energy balance between observer and smart metering data within the given network topology. If a significant mismatch occurs, the fraud detection system moves on to localize the non-technical losses at the consumer level. The fraud detection systems use the measured current from smart meter to calculate respective voltage that are compared to measured voltage. A similar concept has been developed by using smart metering data to identify network voltage sensitivities [74].

5.2.2. State estimation approach [75–80]. The state estimation approach is a very popular concept for observing the grid performance by using the data from smart meters and hence can be used as an effective tool for non-technical losses identification. But this method usually applied in medium voltage networks, so fraud can be detected on substation level only, individual fraud identification is not possible. However, non-technical losses can be expressed in this method as bad data or as false data injection (FDI) attacks; these terms are more relevant in state estimation theory as compared to fraud. The main difference between the two is that bad data typically occur in an isolated and random manner, while FDIs may include several interacting bad data and are more difficult to detect. Launching FDI attacks successfully however, implies tricking a traditional state estimation bad data detector.

The consolidate solution of the Kalman filter state estimator is usually proposed to find the line currents and biases [75]; here the users, with biases larger than a pre-defined threshold, are assumed to commit fraud. Here, privacy is established by proposing a distributed solution of the Kalman filter, where the operator does not require the access to power and voltage measurements of users. Although the proposed method presents promising results, but it should be noted that it can be applied for microgrids with small line lengths only.

Again, it can be assumed that the malicious users should have at least partial knowledge of the network structure and the capability to increase or decrease the measurements of a number of smart meters at the same time [76]. Such attacks usually considered to be undetected by traditional energy balance methods. Here a method has been proposed for detecting this type of attacks including sensor placement, meter and communication network inspections etc.

In Refs. [77–79] the authors propose a state estimator to estimate the loading of medium and low voltage transformers by considering three phase line voltage, line current, active and reactive power measurements. In case of significant difference between measured and estimated values, non-technical losses may be assumed. Finally, a network clustering and division approach has been proposed before state estimation for bad data detection [80]. In this approach, the network partition and bad data detection process for each of the networks is repeated till the bad data are localized on feeder bus level.

5.2.3. Sensor network approach [81, 82]. Another trend in network-oriented non-technical losses detection methods is the installation of dedicated sensors in the distribution grid system. The main purpose of this method is to find the optimal number and position of sensors in order to detect and localize non-technical losses in a better way, while minimizing infrastructure costs. Here in this process is it necessary to collect accurate knowledge of the network topology. Apart from optimizing the place and number of sensors there are works examining the placement of redundant smart meters [82]. Here in this method an observer meter and an inspector box are installed before consumer smart meters; the inspector box consists of a number of inspector smart meters. The inspector meters exchange data with smart meters of consumers and compare the consumption measurements, and the differences between readings indicate possible fraud.

5.2.4. Qualitative comparison

Network oriented methods are based on power systems analysis and usually require the availability of network devices such as the observer meter. Power flow and energy balance methods are most popular, mainly due to their simplicity and applicability in distribution networks (especially LV). They have low data requirements, since next to smart metering data, they usually require observer meter data and sometimes network topologies. In contrast, state estimation methods present higher complexity, especially if applied to large low voltage (three phases, unbalanced) networks. In addition, they typically require more reliable and higher resolution data, including detailed network topologies and data coming from RTUs (especially if state estimation is performed at MV and LV level). On the other hand, state estimation methods perform better and can even detect "smarter" false data injection attacks. Both methods have been used for localizing fraud at substation level and consumer level, but state estimation seems to perform better than power flow/energy balance methods in consumer level localization. Optimal number of sensors and their placement for increasing network observability can be considered as part of NTL detection. After the devices are installed energy balance or state estimation methods may be implemented.

6. CONCLUSIONS

Various algorithms and methods to detect non-technical losses have been proposed in different research papers in recent years. According to their characteristics these methods and algorithms are grouped as discussed in Chapter 2. However, there is scope for other categorizations also as well as new categories may be proposed in future work. It is difficult to compare each and every methods and algorithms, which one is most suitable, because there are unavailability of sufficient tested data sets and scenarios. However, each of them has their own merits and demerits for different sets of consumers, different network topologies and different types of fraud that are studied in different works. A subjective observation can be prepared on the basis of:

- Performance: It is used to measure by the metrics presented in Chapter 4. Here, it can be concluded that the performance of the network-oriented methods is better than that of data-oriented methods, because of the utilization of the physical underlying model, i.e., the power system. In data-oriented methods usually a model has developed where the existing data is fitted in a statistical manner in such a way that the data will be sensitive to the training sets and prone to false positives. For example, it happens when the consumption profile changes due to change in household residents or the usage of electrical devices is such that it might look like a fraud, unless these models are retrained.
- Cost: Another important parameter of non-technical losses identification process also related to the performance of the method/algorithm. It includes the purchase, installation and maintenance of software and hardware equipment that is essential for a specific method/algorithm as well as man-hour. The large number of false positives, that leads to increase in manual inspection costs or lost income, considered as bad performance are generally ignored during cost estimation. Compared to data-oriented methods, network-oriented methods are considered to be costlier for implementation, because additional communication equipment may be required along with the observer meters and other remote sensing equipment. In network-oriented methods the initial as well as operating costs are more because of the use of advanced and complex software components. On the other hand, in data-oriented methods mainly medium or low-resolution data are used which is a regular part of automatic meter reading and billing process. So, the data-oriented methods can be easily implemented with the help of existing infrastructure with small development costs. In order to estimate the added value of new devices in detecting non-technical losses, there should be a cost benefit analysis to be performed.
- Resources (data set size): Usually data-oriented method of non-technical losses identification can be generalized with large volumes of labelled data. On the other hand, high resolution and high-quality data is essential in network-oriented methods, volume of the data is not important here. Further, network-oriented methods also require a larger variety of data obtained from smart energy meters, observer meters and remote sensing units as well as network structure data. It may be concluded that network-oriented methods need data of large variety, but not necessarily large volume, while data-oriented methods require large data sets of small variety.
- Class Imbalance: Data-oriented methods mainly depend on existing and confirmed cases of fraud either for training or validation. However, since frauds are not a common phenomenon, it is not easy to obtain these samples, unless another fraud detection system i.e., anomaly detection or a manual inspection operation are used. It is recommended that the special techniques should be used to ensure that the classifier does not favour to the majority class even if sufficient fraud samples are obtained. In network-oriented methods these

type of problem does not arise, because they do not require training data and here the fraud may be assumed even if positive samples are unavailable for testing.

- Response Time: It can be concluded that the response time is lower in network-oriented methods than that of most of the data-oriented methods. This is because of that the verification of the physical model in network-oriented methods do not require large data, that means the system need not to wait for the accumulation of data till reaching a decision. Again, a lot of devices used in network-oriented methods provide high resolution, which also speeds up the process. On the other hand, the data-oriented methods depend on monthly or yearly consumer profiles which in general slows down the decision-making process.

REFERENCES:

[1] R. Czechowski, A.M. Kosek, “The Most Frequent Energy Theft Techniques and Hazards in Present Power Energy Consumption”, IEEE Proc. 2016 Jt. Work. Cyber-Physical Secur. Resil. Smart Grids, CPSR-SG 2016 -This Work. Is Part CPS Week 2016 (2016), http://dx.doi.org/10.1109/CPSRSG.2016.7684098.

[2] J. Aguero, "Improving the Efficiency of Power Distribution Systems Through Technical and Non-Technical Losses Reduction", in Proc. 2012 IEEE PES Transmission and Distribution Conference and Exposition, pp.1-8.

[3] P. Antmann, “Reducing Technical and Non-Technical Losses in the Power Sector”, in: Background Paper for the WBG Energy Strategy, Tech. Rep., Washington, DC, USA: The World Bank, 2009, n.d.

[4] S.S.S.R. Depuru, L. Wang, V. Devabhaktuni, “Electricity Theft: Overview, Issues, Prevention and A Smart Meter-Based Approach to Control Theft”, Energy Policy 39 (2011) 1007–1015, http://dx.doi.org/10.1016/j.enpol.2010.11.037.

[5] B. Bhatia, M. Gulati, “Reforming the Power Sector Controlling Electricity Theft and Improving Revenue”, The World Bank Group, 2004.

[6] Winther T., “Electricity Theft as a Relational Issue: A Comparative Look at Zanzibar, Tanzania, and The Sunderban Islands, India”, Energy Sustain Dev 2012;16(1):111–9.

[7] IBM. “Energy Theft: Incentives to Change”, Tech. rep.; 2012.

[8] Energy Association of Pennsylvania, “Energy Theft Kills, Costs Innocent Pennsylvanians Millions”, 2007.

[9] M. Sforna, “Data Mining in a Power Company Customer Database”, Electric Power Systems Research, vol. 55, no. 3, Sept. 2000, pp. 201-209.

[10] A. H. Nizar, Z. Y. Dong, and J. H. Zhao, “Load Profiling and Data Mining Techniques in Electricity Deregulated Market”, presented at the IEEE Power Engineering Society (PES) General Meeting, Montreal, Quebec, Canada, June 2006.

[11] M. V. K. Rao, and S. H. Miller, “Revenue Improvement from Intelligent Metering Systems”, in Proc. of Ninth International Conference on Metering and Tariffs for Energy Supply, Birmingham, U.K., August 1999, pp. 218-222.

[12] J. W. Fourie and J. E. Calmeyer, “A Statistical Method to Minimize Electrical Energy Losses in a Local Electricity Distribution Network”, in Proc. of the 7th IEEE AFRICON Conference Africa: Technology Innovation, Botswana,Sept.15-17, 2004.

[13] J. R. Filho, E. M. Gontijo, A. C. Delaiba, E. Mazina, J. E. Cabral, and J. O. P. Pinto, “Fraud Identification in Electricity Company Customers using Decision Trees”, in Proc. of the 2004 IEEE International Conference on Systems, Man and Cybernetics, Vol. 4, pp. 3730-3734, 10-13 October 2004.

[14] J. R. Galvan, A. Elices, A. Munoz, T. Czernichow, and M. A. Sanz-Bobi, “System for Detection of Abnormalities and Fraud in Customer Consumption”, in Proc. of the 12th Conference on the Electric Power Supply Industry, November 2-6, 1998, Pattaya, Thailand.

[15] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed, and M. Mohamad, “Nontechnical Loss Detection for Metered Customers in Power Utility using Support Vector Machines”, IEEE Transactions on Power Delivery, vol. 25, no. 2, pp. 1162-1171, 2010, http://dx.doi.org/10.1109/TPWRD.2009.2030890.

[16] J. Nagi, K. Yap, S. Tiong, S. Ahmed, F. Nagi, “Improving SVM-based nontechnical loss detection in power utility using the fuzzy inference system”, IEEE Transactions on Power Delivery, 26 (2011) 1284–1285, http://dx.doi.org/10.1109/TPWRD.2010.2055670.

[17] C.C.O. Ramos, A.N. De Sousa, J.P. Papa, A.X. Falcao, “A new approach for nontechnical losses detection based on optimum-path forest”, IEEE Trans. Power Syst. 26 (2011) 181–189, http://dx.doi.org/10.1109/TPWRS.2010.2051823.

[18] J. E. Cabral, J. O. P. Pinto, E.M. Gontijo, and J. R. Filho, “Fraud Detection in Electrical Energy Consumers using Rough Sets”, in Proc. of the 2004 IEEE International Conference on System, Man and Cybernetics, 2004, pp. 3625-3629.

[19] R. J. Bolton and D. J. Hand, “Statistical Fraud Detection: A Review”, Statistical Science, vol. 17, no. 3, Aug. 2002, pp. 235-55.

[20] A. H. Nizar, Z. Y. Dong, and J. H. Zhao, “Load Profiling and Data Mining Techniques in Electricity Deregulated Market”, presented at the IEEE Power Engineering Society (PES) General Meeting, Montreal, Quebec, Canada, June 2006.

[21] A. H. Nizar, Z. Y. Dong, and Y. Wang, “Power Utility Nontechnical Loss Analysis with Extreme Learning Machine Method”, IEEE Transactions on Power Systems, vol. 23, no. 3, Aug. 2008, pp. 946-955, http://dx.doi.org/10.1109/TPWRS.2008.926431.

[22] R. Jiang, H. Tagaris, A. Lachsz, and M. Jeffrey, "Wavelet Based Feature Extraction and Multiple Classifiers for Electricity Fraud Detection", in Proc. of the IEEE/PES Transmission and Distribution Conference and Exhibition 2002: Asia Pacific, Yokohama, Japan, Oct. 2002.

[23] C. C. B. de Oliveira, N. Kagan, A. Meffe, S. L. Caparroz and J. L. Cavaretti, "A New Method for the Computation of Technical Losses in Electrical Power Distribution Systems", Proceedings ClRED, 2001.

[24] Anisah H. Nizar, Zhao Yang Dong and Pei Zhang, "Detection Rules for Non-Technical Losses Analysis in Power Utilities", IEEE, 2008, pp 1-8.

[25] Padraig Cunningham, Sarah Jane Delany, "k-Nearest Neighbour Classifiers", Technical Report UCD-CSI-2007-4.

[26] R. Caruana and A. Niculescu-Mizil, "An empirical comparison of supervised learning algorithms", in Proceedings of the 23rd international conference on Machine learning. ACM, 2006, pp. 161–168.

[27] Aristidis Likasa, Nikos Vlassis, Jakob J. Verbeek, "The global k-means clustering algorithm", Elsevier, Pattern Recognition, 2003, pp. 451–461.

[28] T. Zhang, R. Ramakrishnan, M. Livny, "BIRCH: an efficient data clustering method for very large databases", International Conference on Management of Data, USA, 1996, pp. 103–114.

[29] Jiang, H., Li, X., Liu, C. "Large Margin Hidden Markov Models for Speech Recognition", IEEE Trans. Audio, Speech Lang. Process 14(5), pp. 1584–1595, 2006.

[30] S. Roberts, D. Husmeier, I. Rezek, W. Penny, "Bayesian Approaches to Gaussian Mixture Modelling", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, no. 11, pp. 1133-1142, Nov. 1998.

[31] Christina Papadimitriou, Giorgis Messinis, Dimitris Vranis, Sophia Politopoulou, Nikos Hatziargyriou, "Non-technical losses: detection methods and regulatory aspects overview", 24th International Conference & Exhibition on Electricity Distribution (CIRED), 2017.

[32] George M. Messinis, Alexandros E. Rigas, Nikos D. Hatziargyriou, "A Hybrid Method for Non-Technical Loss Detection in Smart Distribution Grids", IEEE Transactions on Smart Grid, 2019.

[33] L. Faria, J. Melo, A. Padilha-Feltrin, "Spatial-temporal estimation for nontechnical losses", IEEE Transactions on Power Delivery, 2015, http://dx.doi.org/10.1109/TPWRD.2015.2469135, 1-1.

[34] G. Chandrashekar, F. Sachin, "A survey on feature selection methods", Computers and Electrical Engineering 40 (2014) 16–28, https://doi.org/10.1016/j.compeleceng.2013.11.024 .

[35] C.C.O. Ramos, D. Rodrigues, A.N. de Souza, J.P. Papa, "On the study of commercial losses in brazil: a binary black hole algorithm for theft characterization", IEEE Trans. Smart Grid 1 (2016), http://dx.doi.org/10.1109/TSG.2016.2560801.

[36] C.C.O. Ramos, A.N. Souza, G. Chiachia, A.X. Falcao, J.P. Papa, "A novel algorithm for feature selection using Harmony Search and its application for non-technical losses detection", Computers and Electrical Engineering**,** 37 (2011) 886–894, http://dx.doi.org/10.1016/j.compeleceng.2011.09.013.

[37] C.C.O. Ramos, A.N. De Souza, A.X. Falcao, J.P. Papa, "New insights on non-technical losses characterization through evolutionary-based feature selection", IEEE Transactions on Power Delivery, 27 (2012) 140–146, http://dx.doi.org/10.1109/TPWRD.2011.2170182.

[38] D.R. Pereira, M.A. Pazoti, L.A.M. Pereira, D. Rodrigues, C.O. Ramos, A.N. Souza, J.P. Papa, "Social-Spider Optimization-based Support Vector Machines applied for energy theft detection", Computers and Electrical Engineering**,** 49 (2016) 25–38, http://dx.doi.org/10.1016/j.compeleceng.2015.11.001.

[39] M. Di Martino, F. Decia, J. Molinelli, A. Fernández, "A novel framework for non-technical losses detection in electricity companies", in: P. Latorre Carmona, J. S. Sánchez, A.L.N. Fred (Eds Pattern Recognition - Applications and Methods, Springer, Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 109–120, http://dx.doi.org/10.1007/978-3-642-36530-0_9.

[40] B. Coma-Puig, J. Carmona, R. Gavalda, S. Alcoverro, V. Martin, "Fraud detection in energy consumption: a supervised approach", in: 2016 IEEE Int. Conf. Data Sci. Adv. Anal., IEEE, 2016, http://dx.doi.org/10.1109/DSAA.2016.19, pp.120–129.

[41] X. Guo, Y. Yin, C. Dong, G. Yang, G. Zhou, "On the class imbalance problem", in:2008 Fourth International Conference on Natural Computation, IEEE, 2008, http://dx.doi.org/10.1109/ICNC.2008.871,pp. 192–201.

[42] S. Axelson, "The base-rate fallacy and the difficulty of intrusion detection", ACM Trans. Inf. Syst. Secur. 3 (2000) 186–205, http://dx.doi.org/10.1145/357830.357849.

[43] A. Jindal, A. Dua, K. Kaur, M. Singh, N. Kumar, S. Mishra, "Decision tree and SVM-based data analytics for theft detection in smart grid", IEEE Trans. Ind. Informatics 12 (2016) 1005–1016, http://dx.doi.org/10.1109/TII.2016.2543145.

[44] K.S. Yap, S.K. Tiong, J. Nagi, J.S.P. Koh, F. Nagi, "Comparison of supervised learning techniques for non-technical loss detection in power utility", International Review on Computers and Software, 7 (2012) 626–636.

[45] J.I. Guerrero, C. León, I. Monedero, F. Biscarri, J. Biscarri, "Improving Knowledge-Based Systems with statistical techniques, text mining, and neural networks for non-technical loss detection",

Knowledge-based Syst. 71(2014) 376–388, http://dx.doi.org/10.1016/j.knosys.2014.08.014.

[46] B.C. Costa, B.L.A. Alberto, A.M. Portela, W. Maduro, E.O. Eler, “Fraud detection in electric power distribution networks using an Ann-based knowledge-discovery process”, Int. J. Artificial Intelligence, Appl. 4 (2013) 17–23, http://dx.doi.org/10.5121/ijaia.2013.4602.

[47] L.A.M. Pereira, L.C.S. Afonso, J.P. Papa, Z.A. Vale, C.C.O. Ramos, D.S. Gastaldello, A.N. Souza, “Multilayer perceptron neural networks training through charged system search and its application for non-technical losses detection”, in:2013 IEEE PES Conference on Innovative Smart Grid Technologies (ISGT Latin America), IEEE, 2013, http://dx.doi.org/10.1109/ISGT-LA.2013.6554383, pp. 1–6.

[48] V. Ford, A. Siraj, W. Eberle, “Smart grid energy fraud detection using artificial neural networks”, in: 2014 IEEE Symposium on Computational Intelligence Applications in Smart Grid (CIASG), IEEE,2014, http://dx.doi.org/10.1109/CIASG.2014.7011557, pp. 1–6.

[49] D. Labate, P. Giubbini, G. Chicco, F. Piglione, “Shape: the load prediction and non-technical losses modules”, CIRED 23rd International Conference on Electricity Distribution, At Lyon, (2015), pp.15–18.

[50] R.D. Trevizan, A.S. Bretas, A. Rossoni, “Nontechnical losses detection: a discrete cosine transforms and optimum-path forest-based approach”, in: 2015 North American Power Symposium (NAPS), IEEE, 2015, http://dx.doi.org/10.1109/NAPS.2015.7335160,pp. 1–6.

[51] J. P. Kosut, F. Santomauro, A. Jorysz, A. Fernandez, F. Lecumberry, F. Rodriguez, “Abnormal consumption analysis for fraud detection: UTE-UDELAR joint efforts”, in: 2015 IEEE PES Innovative Smart Grid Technologies Latin America (ISGT LATAM),IEEE, 2015, http://dx.doi.org/10.1109/ISGT-LA.2015.7381272, pp. 887–892.

[52] C. León, F. Biscarri, I. Monedero, J.I. Guerrero, J. Biscarri, R. Millan, “Variability and trend-based generalized rule induction model to NTL detection in power companies”, IEEE Trans. Power Syst. 26 (2011) 1798–1807, http://dx.doi.org/10.1109/TPWRS.2011.2121350.

[53] I. Monedero, F. Biscarri, C. León, J.I. Guerrero, J. Biscarri, R. Millan, “Detection of frauds and other non-technical losses in a power utility using Pearson coefficient, Bayesian networks and decision trees”, International Journal of Electrical Power &Energy Systems, 34 (2012) 90–98, http://dx.doi.org/10.1016/ j.ijepes.2011.09.009.

[54] J. No, S.Y. Han, Y. Joo, J. Shin, “Conditional abnormality detection based on AMI data mining”, IET Generation, Transmission & Distribution, 10 (2016) 3010–3016, http://dx.doi.org/10.1049/iet-gtd.2016.0048.

[55] S. McLaughlin, B. Holbert, A. Fawaz, R. Berthier, S. Zonouz, “A multi-sensor energy theft detection framework for advanced metering infrastructures”, IEEE Journal on Selected Areas in Communications,31 (2013) 1319–1330, http://dx.doi.org/10.1109/JSAC.2013.130714.

[56] J.E. Cabral, J.O.P. Pinto, E.M. Martins, A.M.A.C. Pinto, “Fraud detection in high voltage electricity consumers using data mining”, in: 2008 IEEE/PES Transmission and Distribution Conference and Exposition, IEEE, 2008, http://dx.doi.org/10.1109/TDC.2008.4517232,pp. 1–5.

[57] T.V. Babu, T.S. Murthy, B. Sivaiah, “Detecting unusual customer consumption profiles in power distribution systems-APSPDCL”, in: 2013 IEEE International Conference on Computational Intelligence and Computing Research, IEEE, 2013, http://dx.doi.org/10.1109/ICCIC.2013.6724264, pp. 1–5.

[58] V. Badrinath Krishna, G.A. Weaver, W.H. Sanders, “PCA-based method for detecting integrity attacks on advanced metering infrastructure”, in: 2015 International Conference on Quantitative Evaluation of Systems, 2015, http://dx.doi.org/10.1007/978-3-319-22264-6_5, pp.70–85.

[59] E.W.S. Dos Angelos, O.R. Saavedra, O.A.C. Cortés, A.N. De Souza, “Detection and identification of abnormalities in customer consumptions in power distribution systems”, IEEE Transactions on Power Delivery 26 (2011) 2436–2442, http://dx.doi.org/10.1109/TPWRD.2011.2161621.

[60] P. Glauner, A. Boechat, L. Dolberg, R. State, F. Bettinger, Y. Rangoni, D. Duarte, “Large-scale detection of non-technical losses in imbalanced data sets”, in:2016 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), IEEE, 2016,http://dx.doi.org/10.1109/ISGT.2016.7781159, pp. 1–5.

[61] S.-J. Chen, T. Zhan, C. Huang, J. Chen, C. Lin, “Nontechnical Loss and, Outage detection using fractional-order self-synchronization error-based fuzzy Petri nets in micro-distribution systems”, IEEE Transactions on Smart Grid 6 (2015) 411–420,http://dx.doi.org/10.1109/TSG.2014.2345780.

[62] J.V. Spiric, M.B. Docic, S.S. Stankovic, “Fraud detection in registered electricity time series”, International Journal on Electrical Power &Energy Systems, 71 (2015) 42–50, http://dx.doi.org/ 10.1016/j.ijepes.2015.02.037.

[63] D. Mashima, A.A. Cárdenas, “Evaluating Electricity Theft Detectors in Smart Grid Networks”, in: Research in Attacks, Intrusions, and Defences, RAID 2012, 7462 (2012), http://dx.doi.org/10.1007/978-3-642-33338-5_11, pp.210–229.

[64] Y. Liu, S. Hu, "Cyberthreat analysis and detection for energy theft in social networking of smart homes", IEEE Transactions on Computational Social Systems, 2 (2015) 148–158,http://dx.doi.org/10.1109/TCSS.2016.2519506.

[65] V.B. Krishna, R.K. Iyer, W.H. Sanders, "ARIMA-based Modelling and Validation of Consumption Readings in Power Grids", Springer International Publishing, Cham, 2016, http://dx.doi.org/10.1007/978-3-319-33331-1.

[66] V.B. Krishna, K. Lee, G.A. Weaver, R.K. Iyer, W.H. Sanders, "F-DETA: a framework for detecting electricity theft attacks in smart grids", in: 2016 46thAnnu. IEEE/IFIP Int. Conf. Dependable Syst. Networks, IEEE, 2016, http://dx.doi.org/10.1109/DSN.2016.44, pp. 407–418.

[67] C. Lin, S.-J. Chen, C. Kuo, J. Chen, "Non-cooperative game model applied to an advanced metering infrastructure for non-technical loss screening in micro-distribution systems", IEEE Transactions on Smart Grid 5 (2014) 2468–2469,http://dx.doi.org/10.1109/TSG.2014.2327809.

[68] T.-S. Zhan, C.-L. Kuo, S.-J. Chen, J.-L. Chen, C.-C. Kao, C.-H. Lin, "Non-technical loss and power blackout detection under advanced metering infrastructure using a cooperative game-based inference mechanism", IET Generation, Transmission & Distribution, 10 (2016) 873–882, http://dx.doi.org/10.1049/iet-gtd.2015.0003.

[69] S. Salinas, M. Li, P. Li, "Privacy-preserving energy theft detection in smart grids: a P2P computing approach", IEEE Journal on Selected Areas in Communications, 31(2013) 257–267,http:// dx.doi.org/10.1109/JSAC.2013.SUP.0513023.

[70] W. Han, Y. Xiao, "A novel detector to detect colluded non-technical loss frauds in smart grid", Computer Networks, 117 (2017) 19–31, http://dx.doi.org/10.1016/j.comnet.2016.10.011.

[71] D.N. Nikovski, Z. Wang, A. Esenther, H. Sun, K. Sugiura, T. Muso, K. Tsuru, "Smart meter data analysis for power theft detection", in: Perner P. (eds) Machine Learning and Data Mining in Pattern Recognition. MLDM, 7988(2013), 379–389, http://dx.doi.org/10.1007/978-3-642-39712-7_29.

[72] E.A.C. Aranha Neto, J. Coelho, "Probabilistic methodology for technical and non-technical losses estimation in distribution system", Electric Power Systems Research, 97 (2013) 93–99, http://dx.doi.org/ 10.1016/j.epsr.2012.12.008.

[73] P. Kadurek, J. Blom, J. F. G. Cobben, W.L. Kling, "Theft detection and smart metering practices and expectations in the Netherlands", in: 2010 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT Europe), IEEE, 2010, http://dx.doi.org/10.1109/ISGTEUROPE.2010.5638852, pp. 1–6.

[74] S. Weckx, C. Gonzalez, J. Tant, T. De Rybel, J. Driesen, "Parameter identification of unknown radial grids for theft detection", in: 2012 3rd IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe), IEEE, 2012, http://dx.doi.org/10.1109/ISGTEurope.2012.6465644, pp. 1–6.

[75] S.A. Salinas, P. Li, "Privacy-preserving energy theft detection in microgrids: a state estimation approach", IEEE Transactions on Power Systems (2015) 1–12, http://dx.doi.org/10.1109/TPWRS.2015. 2406311.

[76] C.-H. Lo, N. Ansari, "CONSUMER: a novel hybrid intrusion detection system for distribution networks in smart grid", IEEE Transactions on Emerging Topics in Computing,1 (2013)33–44, http://dx.doi.org /10.1109/TETC.2013.2274043.

[77] Lijuan Chen, Xiaohui Xu, Chaoming Wang, Research on anti-electricity stealing method base on state estimation, in: 2011 IEEE Power Engineering and Automation Conference, IEEE, 2011, http://dx.doi.org/10.1109/PEAM.2011.6134972, pp.413–416.

[78] W. Luan, G. Wang, Y. Yu, J. Lin, W. Zhang, Q. Liu, Energy theft detection via integrated distribution state estimation based on AMI and SCADA measurements, in: 2015 5th Int. Conf. Electr. Util. Deregul. Restruct. Power Technol., IEEE, 2015, http://dx.doi.org/10.1109/DRPT.2015.7432350, pp.751–756.

[79] Yuan-Liang Lo, Shih-Che Huang, Chan-Nan Lu, Non-technical loss detection using smart distribution network measurement data, in: IEEE PES Innov. Smart Grid Technol., IEEE, 2012, http://dx.doi.org/10.1109/ISGT-Asia.2012.6303316, pp. 1–5.

[80] Y. Liu, Y. Wang, X. Guan, A novel method to detect bad data injection attack in smart grid, in: 2013 Proc. IEEE INFOCOM, IEEE, 2013, http://dx.doi.org/10.1109/INFCOM.2013.6567175, pp. 3423–3428.

[81] L.G. de O. Silva, A.A.P. da Silva, A.T. de Almeida-Filho, Allocation of power-quality monitors using the P-median to identify nontechnical losses, IEEE Trans. Power Deliv. 31 (2016) 2242–2249, http://dx.doi.org/10.1109/TPWRD.2016.2555282.

[82] Z. Xiao, Y. Xiao, D.H.-C. Du, Exploring malicious meter inspection in neighbourhood area smart grids, IEEE Trans. Smart Grid 4 (2013) 214–226,http://dx.doi.org/10.1109/TSG.2012.2229397.