Data-driven Energy Theft Detection in Modern Power Grids

ABSTRACT

Energy theft is an old and multifaceted phenomenon affecting our society on a global scale from both an operational as well as from a monetary perspective. The relatively recent decentralisation of the grid infrastructure with the integration of Distributed Renewable Energy Resources (DRES) in synergy with the widely adopted demand-response business model, has undoubtedly broadened the spectrum of attack surface enabling energy theft. Conventional data-driven energy theft detection schemes have a strong dependency on assessing the spatiotemporal patterns of SCADA measurements aggregated at the Distribution System Operator (DSO) or Transmission System Operator (TSO) with minimal consideration of the intrinsic weather patterns related to individual DRES deployments. Hence, theft scenarios instrumented by DRES owners consuming the energy they produce (i.e., prosumers) can effectively be stealthy and hard to spot. Therefore, in this work we introduce a data-driven, SCADA-agnostic energy theft detection framework explicit to DRES-based scenarios. We provide a comprehensive formalisation of a DRES-based theft attack model and further assess the performance of our framework by utilising and relating freely available third-party weather measurements with real solar and wind turbine deployments in Australia and France. Evidently, our proposed framework yields an energy theft detection accuracy rate of over 98% with optimal computational costs. Thus, reasonably addressing the highly demanding requirements of low-cost and accurate real-time energy theft detection in modern power grids.

KEYWORDS

Energy theft, smart grid, cybersecurity

1 INTRODUCTION

By virtue of global climate challenges, we witness a drastic shift by regulators and grid operators towards the full integration of distributed renewable energy sources (DRES) within modern smart grids with intriguing application (e.g., virtual power plants). In fact, a number of developed and developing nations target 100% of energy generation to be resulted by DRES by 2040 (e.g., Sweden) and a 32% proportion to be achieved on average in the EU by 2030 [1]. Nonetheless, the practical operation of such deployments entails a number of cybersecurity challenges that also transform the way in which energy theft could be manifested. Energy theft has been a traditional challenge, however, the refinement of the grid's business model and the relatively recent interconnection of DRES deployments with the main grid has enabled the composition of data-driven energy theft.

Numerous energy theft events are reported daily on a global scale affecting a range of operational factors for our society including safety and economy. For instance, non-technical losses caused by energy theft amount to $1.4B$ GBP per annum in Brazil and for a single energy provider in Canada such thefts cause an average annual loss of 850 GWh converted as $55M$ GBP of financial loss [2]. Evidently, both energy and monetary losses from energy theft are of paramount and timely importance, with direct implications to the general public's well-being as well as economy. Furthermore, the continuous and evolving manifestation of such events justifies the fact that current theft detection schemes employed by energy providers are inadequate.

As energy theft underpinned by cyber-attacks increases, a momentum on the development of data-driven detection has been observed within the wider research community. However, a significantly small portion of detection solutions such as in [3, 4, 5] considers the manifestation of energy theft from individual DRES owners. Moreover, the dependence on historical consumption data that can be tampered with by adversaries [4, 3] or unavailable data in terms of supervisory control and data acquisition (SCADA) [5] further restricts the reliability of these approaches in practical scenarios. In parallel, the aforementioned dependence on aggregated consumption SCADA measurements is unable to capture the intrinsic environmental dynamics such as to relate generation values with the actual weather conditions in a given DRES deployment [6].

Therefore, in this work we take a practical approach by firstly proposing a generalised DRES-based adversary model and by secondly providing a data-driven detection solution framework considering weather dynamics. We go a step beyond current solutions by removing any dependence from SCADA measurements and by focusing largely on third-party and freely available measurements. Thus, to tailor theft detection accuracy based on the explicit properties related to generation and demand of individual DRES deployments. In addition, we demonstrate that the introduced approach is applicable to large-scale wind-turbine and solar panel DRES installations deployed in Australia and France.

In general, the contribution of this work is two-fold by providing:

(1) A formalised approach on describing DRES-based adversaries with the objective of energy theft. We demonstrate the efficacy in which DRES owners (i.e., prosumers) can take advantage of the current business model and steal energy.
(2) A novel, low-cost and generic energy theft detection framework comprising of two algorithms; i) a SCADA-agnostic DRES profiling method operating purely on third-party and widely available weather measurements and ii) a classification scheme relying on DRES profiling and able to classify theft detection events. Evidently, the synergy of the two components enables adaptive and highly accurate detection with low computational overheads and aiding significantly on reducing monetary loss.

The rest of this paper is structured as follows: Section 2 is dedicated on discussing related work whereas Section 3 provides a description of the system and Section 4 presents the adversary model for DRES-based energy theft attacks. Section 5 describes the methodology underpinning the proposed detection framework, while Section 6 discusses the datasets used within this work. Section 7 depicts the evaluation methodology followed within our experimentation and Section 8 discusses the results obtained. Finally, Section 9 concludes and summarises this paper.

2 RELATED WORK

In general, the data-driven approaches proposed for energy theft detection can be classified based on the detection infrastructure into two main categories: detecting theft attacks in the consumption infrastructure and detecting them in the DRES infrastructure of power grids.

The majority of the data-driven studies fall into the former category, focusing on detection of theft attacks in the consumption infrastructure of power grids. For instance, Punmiya *et al.* [7], Blazakis *et al.* [8], Yao *et al.* [9], Sharma *et al.* [10] and Gunturi *et al.* [11] employed energy consumption measurements to detect these theft attacks. Punmiya *et al.* [7] proposed a gradient boosting detector and in the purpose of enhancing the detector's performance; the proposed approach relies on a feature engineering-based preprocessing technique. However, the detector developed by Blazakis *et al.* [8] is based on an adaptive network-based fuzzy inference system, that for the first time have been applied in the field of energy theft attack detection. Yao *et al.* in [9] proposed a classification energy theft detection scheme based on combined convolutional networks and Paillier cryptosystem. An online unsupervised detector called recursive transform learning is proposed in Shalini *et al.* in [10] for detection of energy theft attacks dynamically in the consumption infrastructure. The proposed detection algorithm by Gunturi *et al.* in [11] employed tree based ensemble machine learning models including categorical boosting, adaptive boosting, light boosting, extreme-boosting extra trees and random forest to detect energy thefts.

In addition to consumption measurements, Zheng *et al.* [12] and Luya *et al.* [13] employed the measurements recorded by an observer smart meter, installed to aggregate the sum of the consumption measurements of a group of consumers over a certain period. The proposed approach by Zheng *et al.* in [12] combined two data-driven techniques, i.e., the maximum information coefficient and the clustering technique by fast search and find the density peaks, to detect these thefts. However, Luya *et al.* [13] proposed a linear regression-based detector to detect theft attacks and estimate the amount of the stolen energy.

Although the efficient results achieved by these detection approaches falling into this category in detecting theft attacks in consumption infrastructures, such schemes assumed that the attackers are not adaptive with the advanced smart grid infrastructures. As a result, the amount of the energy loss for the utility providers could increase dramatically, since detecting theft attacks in these advanced properties of the modern grids, i.e., DRES, are neglected in such detection mechanisms.

Nevertheless, there are a limited number of studies investigating energy theft detection in DRES infrastructures. Out of these, Yuan *et al.* [3] employed the energy measurements generated by solar panels and they proposed a theft detector based on a moving time window and a least-squares approach. However, Krishna *et al.* [4] employed the aggregated measurements from solar panels and wind turbines based DRES installation. The detector developed by Krishna *et al.* [4] is based on the auto-regressive integrated moving average technique. More recently, besides the generated energy measurements, Ismail *et al.* [5] considered adoption of diverse measurement sources including generated energy, solar irradiance accessible from the weather stations, and SCADA to develop a deep-learning-based theft detector. Although these works detect theft attacks in the DRES infrastructure, the dependence on historical data or unavailable SCADA measurements further limiting the reliability of such approaches in practical scenarios.

3 SYSTEM DESCRIPTION

We consider an end-to-end energy system consisting of a single Transmission System Operator (TSO) connected with one or more Distribution System Operators (DSOs) consisting of nodes equipped with smart control, management, monitoring and metering technologies. The TSO is abstracted to a set of supply nodes R including DRES deployments and a set of high-voltage transmission buses denoted as Q. It is assumed that one or more DSOs of the set $P = [p_1, .., p_n]$ interact with the TSO in discrete time intervals and the energy supplied from the TSO to a given DSO on a discrete time interval is denoted as the function $Es(t)$. Energy transmission and distribution is achieved via bidirectional power and data communication flows through corresponding power system and communication control and management components (e.g., actuators, SCADA).

Each DSO in P is defined by a total number of nodes N and a set M of medium/low voltage distribution buses. Nodes are categorised into B demand and A supply nodes we refer to as consumers and prosumers respectively where $B \subset N$ and $A \subset N$. The energy produced by a single prosumer of A in the i^{th} DSO at a given discrete time interval is mapped as the function $Er_i(t)$ whereas the energy consumed by a single consumer over a time period in the i^{th} DSO is represented by $Ec_i(t)$. A prosumer is assumed to be an individual or a group of individuals owning and managing a DRES deployment (e.g., domestic solar panels) and can act both as a consumer and a supplier of energy back to the DSO.

As discussed in [14], energy theft events cause energy losses that can be described as the difference between the generated energy and the energy consumed under normal conditions. Thus, we express the cumulative energy loss experienced for a single DSO in time t as:

$$L = \Delta Es(t) + \Delta \sum_{a=1}^{A} Er_a(t) - \Delta \sum_{b=1}^{B} Ec_b(t) + \sum_{m=1}^{M} TL_m(t) \tag{1}$$

where Δ is the discrepancy in meter readings for reported and actual measurements as caused by a single or more theft events at time t and TL refers to technical losses occurred due to physical constraints on transmission lines. Consequently, from a TSO perspective the total energy loss in time t is expressed as:

$$L_{TSO} = \sum_{i=1}^{P} L_i(t) \tag{2}$$

where P is the total number of DSOs connected to the TSO.

4 ADVERSARY MODEL

The adversary model has an explicit focus on energy theft initiated by generation meters installed on DRES deployments and managed by prosumers. Thus, the primary assumption is that prosumers

tamper generation meters and report erroneously back to their corresponding DSO. In order to reduce the complexity invoked within our adversary model, we consider the DSO meters interacting with edge DRES deployments to be secure. Thus, we rule out any discrepancy in the measurement function for energy supplied from a TSO to a DSO having $\Delta Es(t) = 0$.

In addition, we assume that smart-meters strictly reporting energy consumption by a given consumer to the DSO are not tampered. Therefore, discrepancies on the consumption reporting by all consumers in a given DSO complies with: $\Delta \sum_{b=1}^{B} Ec_b(t) = 0$.

As mentioned, we particularly focus on tampering conducted on individual meters reporting energy generation for a given DRES deployment. Hence, we deduce that: $\Delta \sum_{a=1}^{A} Er_a(t) \geq 0$.

Given the above assumptions we re-express the DSO energy loss as:

$$L = \Delta \sum_{a=1}^{A} Er_a(t) + \sum_{m=1}^{M} TL_m(t) \tag{3}$$

Based on equations (2) and (3), the energy loss for the TSO can be approximated as follows:

$$L_{TSO} = \Delta \left\{ \sum_{}^{P} \sum_{}^{A} Er_{i,a}(t) \right\} + \sum_{}^{P} \sum_{}^{M} TL_{i,m}(t) \tag{4}$$

In order to cover a spectrum of tampering behaviour by a prosumer we define four types of theft functions. All four functions mimic practical fraudulent patterns in terms of reporting erroneous generated energy back to the DSO. Many possibilities of such theft functions exist and the herein models are distilled by observations in literature [5, 15, 12]. Through our work we emulate attackers that attempt to create manipulated reports either by retaining original curve fluctuations and features or by generating new patterns [12]. From a modeling perspective, these are variables that partially or completely amplify the reported energy timeseries signal as we show next.

(1) Total Scaling Theft:

$$\Delta Er(t) = \eta(t) Er(t) \tag{5}$$

where $\{\eta \in \mathbb{R} \mid \eta > 1\}$).

(2) Partial Scaling Theft:

$$\Delta Er(t) = \begin{cases} Er(t), & Er(t) \geq \beta \\ \beta, & Er(t) < \beta \end{cases} \tag{6}$$

where $\{\beta \in \mathbb{R} \mid \beta > \min(Er(1), Er(2), ..., Er(T)\}$

(3) Off-Peak Theft:

$$\Delta E(t) = \gamma E(t) \tag{7}$$

where

$$\gamma = \begin{cases} \eta, & t \in [t_{start}, t_{end}] \\ 1, & otherwise \end{cases}$$

where $[t_{start}, t_{end}]$ is the off-peak period, that is the peak operating weather conditions for DRES.

(4) Replay Theft:

$$\Delta Er(t) = \max(Er(1), Er(2), ..., Er(T)) \tag{8}$$

In more detail, the total scaling theft in equation 5 considers the scenario in which the aggregated generation measurements on time t are tampered by an attacker. Tampering is based on an arbitrary percentage denoted by η, which is adjustable (i.e., random rate percentage). For instance, the attacker reports 150% of the actual measurements when $\eta = 1.5$. The partial scaling theft scenario in equation 6 considers the case where an adversarial prosumer tampers generation measurements whilst a particular threshold is met. Hence, the prosumer sets a minimum reporting value (i.e., β) for the DRES-based generation measurements sent to the DSO.

We also consider the case in which theft could be temporally sporadic. Thus having discontinuous reporting of erroneous generation measurements during an off-peak period that relates with the peak weather conditions in which the DRES operates. For instance, the fraudulent prosumer reports 40% more power for a given time period than what was actually generated during the peak solar radiation of that period. Therefore, only measurements generated during the peak operating conditions of DRES are scaled as shown in equation 7. Finally, in the replay attack, the attacker only reports the highest actual generation for the whole time duration T.

5 METHODOLOGY

The data flow underpinning the proposed framework [1] is depicted by the flowchart in Fig. 1. As shown, we incorporate two independent but complementary methods handling DRES profiling and anomaly classification. Building upon the work in [6] showing the concrete profiling and prediction of energy generation using purely geo-located weather features, we profile DRES installations using third-party and widely available weather measurements. The incentive behind this approach is to remove from the full dependency on SCADA-based measurements. Hence, we align with the realistic scenario where no available measurements gathered by locally placed sensors or the DRES SCADA systems exist. As depicted, the SCADA-agnostic profiling component works in synergy with a classification component that considers reported DRES energy generation measurements as seen at the DSO level. Thus, to tailor theft detection over individual DRES installations and back-track potential fraudulent prosumers.

5.1 SCADA-agnostic DRES energy profiling

As shown in Fig. 1, the implemented DRES energy profiling component accepts third-party weather measurements and it first employs an automated pre-processing procedure. Following a series of data-oriented tasks dealing with noisy and incomplete measurements, the profiling component conducts an automated feature selection process in which the most suitable statistical features are chosen to compose a DRES energy generation profile. The trained model based on the selected subset of features of the source DRES is then used for all DRESs to profile their generated energy measurements. Details of all the processes involved in the aforementioned description in terms of profiling are discussed below.

5.1.1 Data pre-processing: Within the pre-processing stage we filter our raw measurements by removing all the possible missing,

[1] Code and metadata available on Github: [redacted]

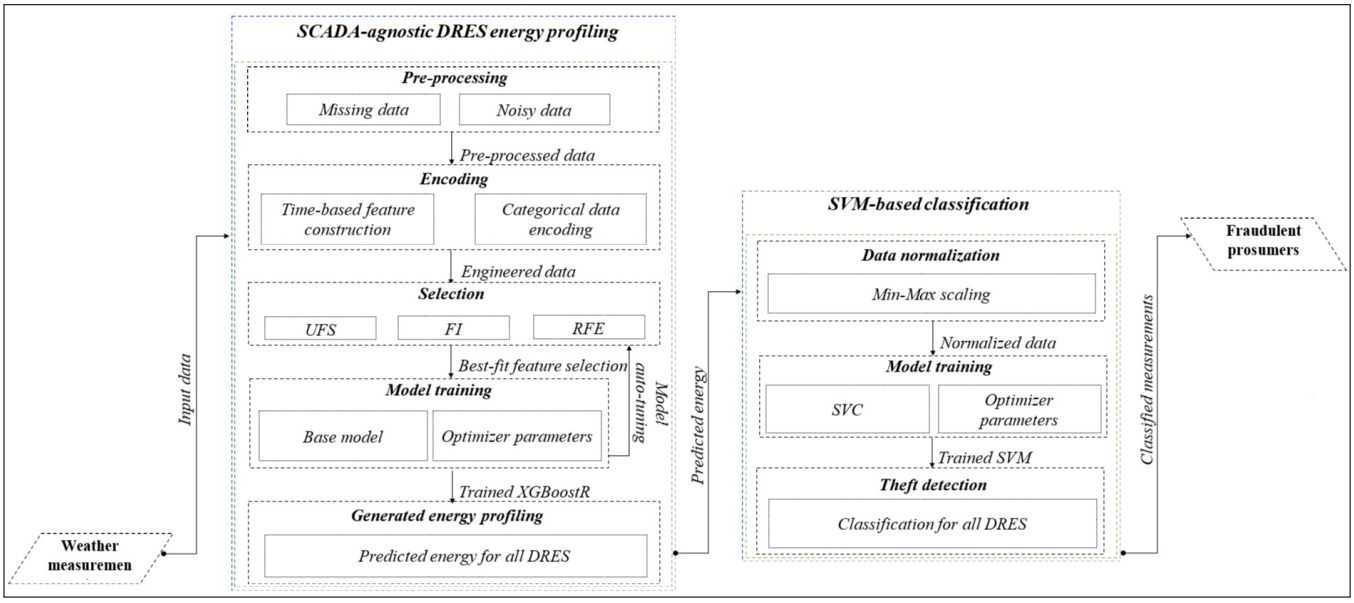

Figure 1: Data flows defining the proposed energy theft detection framework.

duplicated and inconsistent samples. Usually, third-party measurements are largely inconsistent (out-of-range) due to various factors ranging from environmental sensor reading and reporting errors as well as REST-API pull failures. In this context, included common data sanitisation and normalisation approaches within an automated pipeline in our prototype.

5.1.2 Data Encoding: The encoding process enables granular representations from aggregated third-party time series Given that aggregated measurements contain categorical time series, we encode them into numerical data via a binary encoder. In more detail, we assign an integer value to each unique category of the original categorical vector. Subsequently, for each integer-encoded category we generate a binary vector and based on the majority of bit coding, we generate an additional measurement vector.

Finally, we construct temporal views of our categorical measurements using the measurement timestamps having hourly, daily and monthly observations mapped with the aforementioned binary encoded measurement vector.

5.1.3 Data selection: The developed data selection component is underpinned by an automated feature selection mechanism such as to identify an appropriate set of features from the encoded time-series described earlier. As shown in Fig. 1, our component works in coordination with the model trained on the source DRES component to obtain the optimal feature set, producing the best-fit prediction model of the DRES generated power. In detail, the selection process utilises i) Univariate feature selection (UFS), ii) Ranking-based feature importance (FI) and iii) Wrapper-based recursive feature elimination (RFE). The reason of using three feature selection algorithms is to ensure that we compile the best combination of meta-features. The importance of each of the fetures is compared and chosen based on the F-score and the Pearson correlation coefficient as well as the random forest estimator.

5.1.4 Model training: The base DRES profiling model strongly depends on the aforementioned feature selection process. Most importantly, it is dynamically updated whilst new and improved data feature combinations are provided by the selection process. Through a repetitive feedback mechanism and the continuous update of a boosting regressor we achieve an adaptive DRES energy generation base profile.

In particular, we utilise the XGBoostR algorithm which is a classification and regression tree (CART) ensemble model using K additive functions. Thus, enabling prediction of the generated power measurements of the DRES installations of DSO. In the proposed prototype we minimise the regularised objective of the XGBoost modelas defined in [16]:

$$\mathcal{L}^{(j)} = \sum^{T} l\big(y(t), \hat{y}(t)\big) + \sum^{K} \Omega(f_k) \tag{9}$$

where $y(t)$ and $\hat{y}(t)$ denote the actual and predicted power measurements at time t, l is the loss function measuring the difference between $\hat{y}(t)$ and $y(t)$ and Ω denotes the model complexity for avoiding over-fitting. We can express $\Omega(f)$ as:

$$\Omega(f) = \zeta V + \frac{1}{2}\lambda \mid\mid w \mid\mid^2 \tag{10}$$

where V is the number of predicted energy measurements, w is the score of each measurement and ζ and λ are constants controlling the regularisation degree.

Since the XGBoostR model is trained in an additive manner, Equation 9 can be expressed as:

$$\mathcal{L}^{(j)} = \sum^{T} l\Big(y(t), \hat{y}(t)^{j-1} + f_j\big(x(t)\big)\Big) + \Omega(f_j) \tag{11}$$

where $x(t)$ is the input vector at t and $\hat{y}(t)^j$ represents the power measurement prediction of the t-th observation at the j-th iteration.

Within our implementation, second-order approximation is used to optimise the objective, which can be simplified as follows:

$$\mathcal{L}^{(j)} = \sum^{T} \Big(g(t) f_j\big(x(t)\big) + \frac{1}{2} h(t) f_j^2\big(x(t)\big) \Big) + \Omega(f_j) \tag{12}$$

where

$$g(t) = \partial_{\hat{y}^{(j-1)}} l\big(y(t), \hat{y}(t)^{j-1}\big) \tag{13}$$

and

$$h(t) = \partial^2_{\hat{y}^{(j-1)}} l\big(y(t), \hat{y}(t)^{j-1}\big) \tag{14}$$

The optimal hyper-parameters for the employed XGBoostR model training process was obtained using a grid search technique returning the appropriate values with the lowest prediction error.

5.1.5 Generated energy profiling: The generated energy measurements of all DRES at the DSO are predicted using the trained XGBoostR as follows:

$$\hat{y}(t) = \sum^{K} f_k\big(x(t)\big) \tag{15}$$

where f denotes an independent tree structure.

Due to the cumulative nature of the utilised XGBoostR, the predicted power measurements of the source DRES at step j can be calculated as follows:

$$\hat{y}(t) = \hat{y}(t)^{j-1} + f_j\big(x(t)\big) \tag{16}$$

As discussed next, the resulted features are utilised within the SVM-based classification process in order to detect fraudulent prosumers.

5.2 SVM-based classification

A supervised SVM classifier is trained based on the predicted energy measurements calculated using Equation (16) such as to provide a binary prediction class for each of the prosumers in a given DSO (i.e., fraudulent or not).

As shown in Fig 1, the first process within the implemented SVM-based classification prototype deals with normalisation of the DRES profiling output, which is subsequently used as input to the DSO's SVM training model. As explained next, the classification phase is decomposed into specific processes in order to ensure unbiased detection of fraudulent prosumers.

5.2.1 Data normalisation: Prior the training and classification stage we employ a min-max normalisation technique to reconstruct the processed measurements in the range of $[0, 1]$. Normalisation is a crucial component within any statistical representation process and particularly in our case we achieve to ensure testing and not neglecting extremely small measurement values.

5.2.2 Model training: Following data normalisation, a model classifier is resulted by processing the training set to detect fraudulent prosumers. Thus, the predicted energy measurements from Equation (16) are the input to a trained SVM model in such a way to accommodate an optimal decision boundary for classifying DSO prosumers. The optimal SVM hyperplane boundaries are obtained by solving the following soft optimisation problem:

$$min\Big(\frac{1}{2} \mid\mid w \mid\mid^2 + C \sum^{T} \xi(t)\Big) \tag{17}$$

where w denotes the weight vector, C denotes the regularisation parameter used to quantify the trade-off between the model's complexity and the classification error. Also, ξ represents a slack variable.

In order to select the most appropriate hyper-parameter values with the highest training accuracy for the SVM model we employ a grid search algorithm.

5.2.3 Theft detection: Once the SVM-based training model is achieved, the binary classification of DSO prosumers to either being legit or fraudulent is conducted.

The decision boundary function in our proposed implementation is defined as:

$$f(x) = \sum^{T} (\alpha - \beta) K(x(t), y(t)) + b \tag{18}$$

where $x(t)$ is the support vector, $y(t)$ is the assessed power measurement, K is the kernel function. The α and β variables are the Lagrange multipliers and b is the regularisation parameter.

Due to the fact that all energy measurements have a non-linear distribution, we employ a radial basis function (RBF) kernel defined as:

$$K\big(x(t), y(t)\big) = exp\Big(- \gamma \big(x(t) - y(t)\big)^2 \Big) \tag{19}$$

where where $x(t)$ is the support vector, $y(t)$ is the assessed power measurement and *gamma* is the kernel function parameter.

6 DATASET DESCRIPTION

Our evaluation is based on real measurements gathered by wind-turbine and solar panel installations in Australia and France. In particular, we utilise a dataset acquired from the La Haute Borne wind farm located in Meuse, France [2] and a solar power dataset captured at the Ausgrid power network located in Sydney, Australia [3].

Table 1 depicts a summary of the aforementioned datasets. As shown, the Engie wind datasets represents the daily generated power measurements captured at a real installation of 4 wind turbines for a duration of 11 months in 2017. In addition, the Ausgrid solar data provides daily measurements captured for a period of 11 months from 300 rooftop solar panel installations.

Table 1: Datasets overview.

Dataset	Time Window		Location		DRES Capacity
	Start	End	Longitude	Latitude	
Engie Wind	Jan 2017	Dec 2017	5.6013 E	48.4503 N	2050 kW
Ausgrid Solar	Jul 2012	Jun 2013	151.2093 E	33.8688 S	1 kW

Within Table 1, we highlight longitude and latitude values since they were critical for mining weather and environmental information explicit to those areas. Hence, we extracted available measurements such as output temperature, wind speed, humidity and

[2] Explore – ENGIE France Renewable Energy Open Data, Available: https://opendata-renewables.engie.com/pages/home/

[3] Explore – Ausgrid Solar Home Electricity Data, Available:https://www.ausgrid.com.au/Industry/Our-Research/Data-to-share/Solar-home-electricity-data

pressure and assessed their ground truth by cross-validating across multiple third-party and freely available APIs. In order to do that, we collected data from the Dark Sky API [17], Weather Online API [18] and Open Weather API [19] aligning with the same observational period as that of the generation measurements at the Engie and Asugrid installations. Complementary to the aforementioned, we acquired additional output temperature and wind-related measurements by the Nancy-Ochey weather station, which is geographically adjacent to the La Haute Borne wind farm. In total, our third-party weather data, i.e., the obtained weather measurements from freely available APIs, comprised of 53 weather and environmental measurements, including numerical and categorical readings such as wind measurements, humidity, pressure and cloud cover within hourly sampling bins. For our evaluation group, all our measurements seasonally (i.e., summer, autumn, spring, and winter).

7 EVALUATION METHODOLOGY

The evaluation methodology employed within this work aims at determining the suitability of the integrated data-driven theft detection solution over diverse DRES deployments. The main focus of which was placed on quantifying the detection performance and also relating it with the corresponding computational costs. Moreover, we conduct a monetary meta-analysis assessing the potential impact of the various synthetic thefts as well as the theft detection gains from the DSO perspective.

7.1 Detection Performance

A thorough analysis was conducted such as to evaluate the detection performed using the synergy of the SCADA-agnostic DRES energy profiling and the SVM classifier discussed in Section 3.

The first phase consists of the SVM classifier training with input from the SCADA-agnostic DRES power profiling output to compute final predicted power for all DRES installations. In parallel, an instance of the SVM component is trained based on the third-party weather data measurements. Within our classification procedure, we distributed the dataset so that 70% of it is for training and 30% from each season as testing [20].

As already described in Section 3, we reach a binary detection decision (i.e., fraudulent or legit) through comparing the two outcomes of the aforementioned classification processes. Two classification errors and one computational cost metric were also utilised to assess the resulting classification models. The classification error metrics are accuracy (ACC) and area under the curve (AUC), while we consider the the time taken to obtain a decision as the computational cost. The definitions of which are provided as follows.

(1) **ACC:** The ability to correctly differentiate the fraudulent and legit measurements defined as:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{20}$$

where TP, TN, FP and FN represent the true positives, true negatives, false positives and false negatives, respectively. TP is the number of measurements correctly identified as fraudulent, TN is the number of measurements correctly identified as legit, FP is the number of measurements incorrectly identified as fraudulent, and FN is the number of measurements incorrectly identified as legit.

(2) **AUC:** Degree of the capability of distinguishing between fraudulent and legit measurements defined as:

$$AUC = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right) \tag{21}$$

(3) **Computation time complexity:** Time required by the SVM prototype to produce a DRES classification of a given DSO_i.

7.2 Theft scenarios

Due to the fact that the acquired datasets were the result of prosumers that volunteered to provide their data we assume that all measurements were legit and no fraudulent behaviour is present. Hence, prosumers reported genuine generation measurements therefore the original data are considered as the ground truth. As presented in Section 5, we inject synthetic anomalies in our datasets that conform to specific theft scenarios discussed in the literature. Hence, we employ our developed energy theft functions, i.e., (definitions 5), ((6)), (7) and (8) in order to compose a dataset consisting of both legitimate as well as fraudulent patterns. As depicted by Table 2 the conducted evaluation methodology considers varying theft proportions (i.e., fraudulent measurements) injected within the actual dataset across a given DSO. Evidently, we stretch the scaling parameters for both total and partial thefts within particular boundaries such as to ensure that we create the most representative realistic scenarios.

Table 2: Simulation parameters in theft scenarios.

Dataset	Theft Scenario	Parameter
Engie Wind	Total Scaling Theft	$1.4 \leq \eta \leq 7$
	Partial Scaling Theft	$50\text{ kW} \leq \beta \leq 100\text{ kW}$
	Off-Peak Theft	$rated\ wind\ speed = 15\text{ m/s}$
	Reply Theft	$T = 24$
Asugrid Solar	Total Scaling Theft	$1.4 \leq \alpha \leq 7$
	Total Scaling Theft	$0.005\text{ kW} \leq \beta \leq 0.4\text{ kW}$
	Off-Peak Theft	$11\ am \leq t \leq 3\ pm$
	Reply Theft	$T = 24$

The mix theft scenario focuses on a randomly chosen subset of measurements to simulate one of the four main scenarios. As discussed in the literature, there exist many cases in which fraudulent prosumers might apply different theft scenarios over different time-periods to manipulate with their measurements [12].

8 RESULTS

8.1 Theft Detection Performance

The results of the theft detection framework proposed in 5 are illustrated in this section, while considering the discussed evaluation methodology in 7 on the datasets specified in 6.

Using the SVM-based classification system proposed in 5.2, we witness that the SVM-based classifier trained on the energy profiling outperformed the classifier based on the third-party weather data for with regard to the ACC and AUC scores in Engie wind data as shown in Figs.2 and 3. In this case, total scaling theft results in 5.9% higher score when we use predicted wind energy output as a feature to train the SVM classifier, as compared to using the third-party weather data. In addition, the performance of the model based on third-party weather data significantly drops on the partial

scaling theft scenario, with a margin of more than 13.16% compared to the first case when it was trained on the energy profiling output. Nonetheless, under the replay theft scenario, both classifiers achieved high ACC score of 97.2% and AUC 99.6%. The nature of the reply theft is behind this observed pattern of detection performance. In the scenario of the replay theft, the fraudulent prosumers over-report the maximum of the actual generation of the installed DRES. This theft behavior shows a steady and repetitive distribution throughout the fraudulent energy measurements, that can be unambiguously detected by the proposed SVM-based classification prototype in both cases, i.e., either it is trained using the third-party weather data, or the energy profiling.

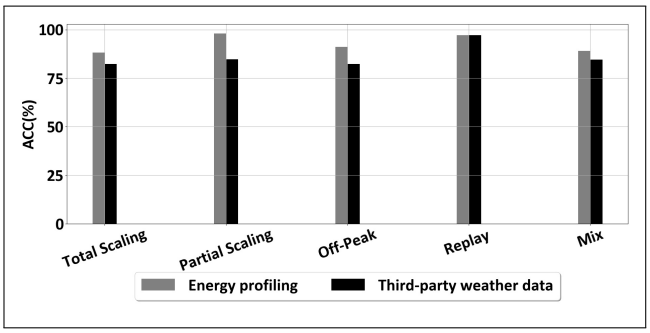

Figure 2: ACC values of the Engie wind power data.

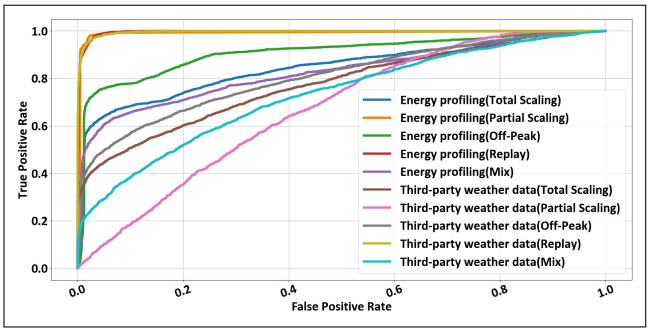

Figure 3: AUC values of the Engie wind power data.

For the Asugrid solar power measurements , during the same procedures of our proposed evaluation methodology, the SVM-based classification prototype trained on the energy profiling output provided higher scores in detecting several theft scenarios, as shown in Figs. 4 and 5. As evident from these figures, in the case when the energy profiling output was used as an input feature, the SVM-based classification prototype obtained an ACC of 89.1% and an AUC score of 81.4% in detecting the mix theft. However, in detecting the same theft scenario, the SVM-based classification prototype trained on the third-party weather data maintained a lower score (more than 6%) than the one trained on energy profiling output by obtaining an ACC of 82.3% and an AUC score of 74.2%. Similarly, to detect the total scaling theft, the SVM-based prototype trained on the energy profiling output provided more than 2% higher results than that trained on the third-party weather data. In detecting replay theft, the both classifiers performed excellent scores of ACC and AUC.

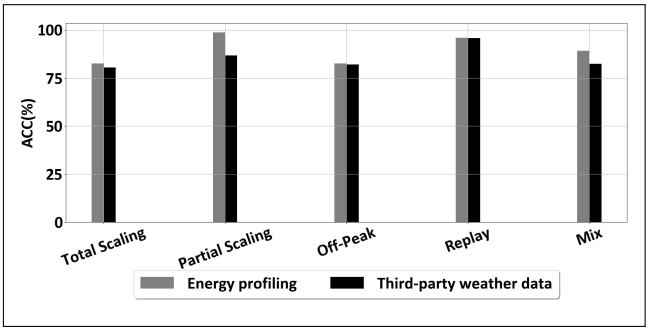

Figure 4: ACC values of the Asugrid solar power data.

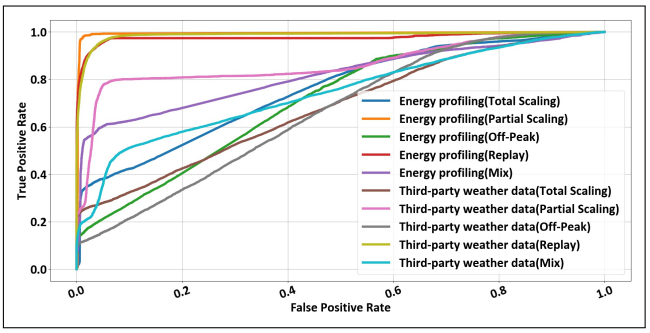

Figure 5: AUC values of the Asugrid solar power data.

Overall, the SVM-based classification prototype trained on the energy profiling output maintained the high ACC and AUC scores in both the wind and solar power measurements. Therefore, we can infer that the output of the SCADA-agnostic power profiling prototypes (i.e., the energy profiling for the DRES) has an important role to play in differentiating fraudulent prosumers. For more insight, Fig. 6 presents the boxplots of predicted power measurements and the actual generations for both legit and fraudulent prosumers. It is evident, that the energy measurements proportion range for legit prosumers was between 0 kW and the capacity of the DRES installations, thus 2050 kW for wind and 1 kW for solar respectively. However, the energy measurement proportion of the fraudulent prosumer exceeded this range by the value of the manipulated green energy units. Moreover, it is demonstrated that the predicted power of the legit prosumers falls within 7% of actual generation, whereas a significant difference between the predicted and actual generated power can be observed for the fraudulent ones. Therefore, the predicted power measurements for DRES can be used as a useful feature for the proposed SVM-based classification prototype, where prosumers are classified either as fraudulent or legit based on their respective predicted energy measurements.

Fig. 7 depicts the results of the SVM-based classification prototype in both the proposed cases in terms of computational time in our evaluation methodology. This analysis was performed using a 64-bit Windows operating system with Intel Core i7 ($7th$ Gen) CPU with 12 GB RAM and 2.70 GHz clock cycle. The results clearly indicate that the SVM-based classification prototype trained on the energy profiling prototype output operates on a relatively lower computational time than that on third-party weather data.

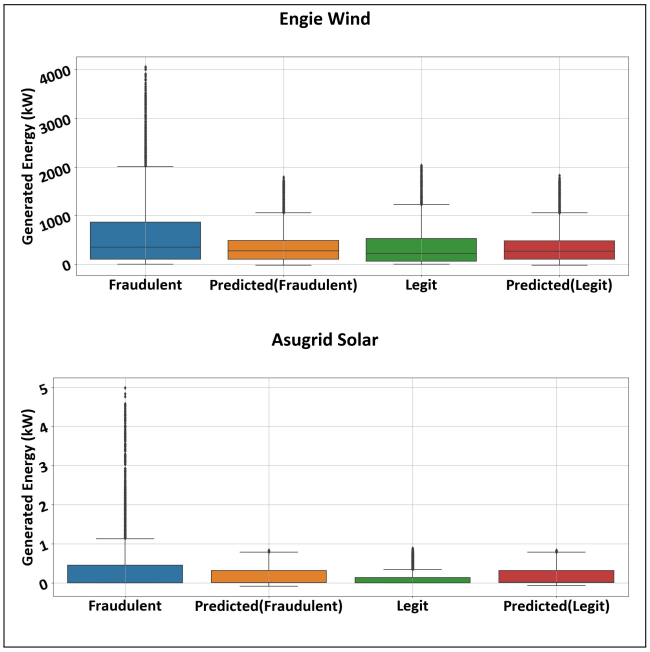

Figure 6: Predicted and actual power generation of legit and fraudulent prosumers in Engie and Ausgrid.

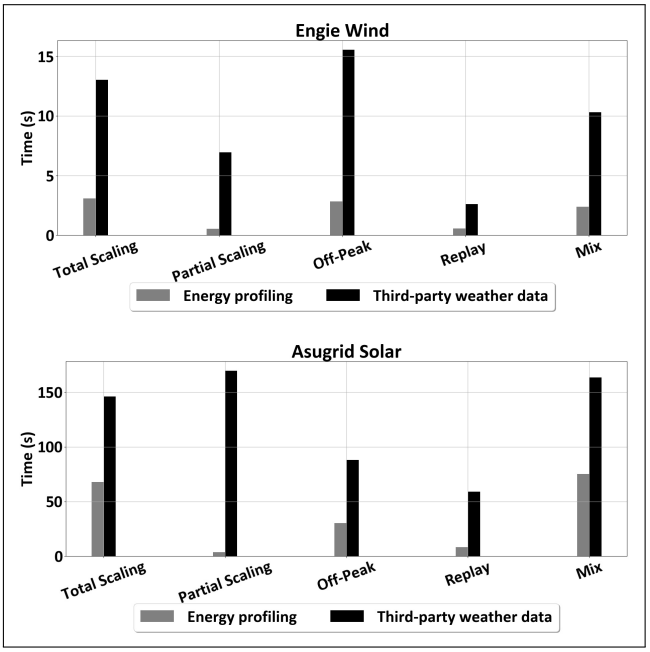

Figure 7: Computational time comparison.

The reason behind this is the high dimensionality of the aggregated third-party weather data, where the total number of selected measurements was more than that output of the energy profiling prototype, i.e., the predicted energy measurements. In the case of the third-party weather data, the process of classifying such high dimensional data requires computational complexity than that occurring in low dimensional spaces, where the SVM-based classifier only performs on the energy profiling output.

8.2 Monetary analysis

The density of the amount of the energy loss caused by the theft attacks originally for the DSO in each month of the year is illustrated in Fig. 8. The amount of the energy losses in both wind and solar data can be obtained using Equation (3). Fig.9 presents the amount of the monetary cost caused by each individual theft scenario for the whole of a year. These monetary costs are estimated by multiplying the electricity price with the resulted total DSO energy loss in each season. The France feed-in tariff (i.e., £7.40/kWh [21]) was applied to the wind energy dataset, while the Ausgrid feed-in tariff (i.e., £0.051/kWh [4]) was applied to the solar energy dataset.

These figures indicate that the monetary cost for the utility provider varied linearly with the amount of energy loss for both the datasets. The spikes of the density curves in Fig.8 denote that for the wind measurement, the highest concentration of the highest energy loss was caused by the total theft scenario resulting in a monetary cost of 31% for the providers. The same behaviour was noticed in the solar measurement, where highest concentration of the highest energy loss was cased by the reply thefts, resulting in the largest amount of cost of 34.1% of the total monetary cost of that year. As evident from these figures, the monetary costs can reach an incredible level when large-scale DRES are manipulated,

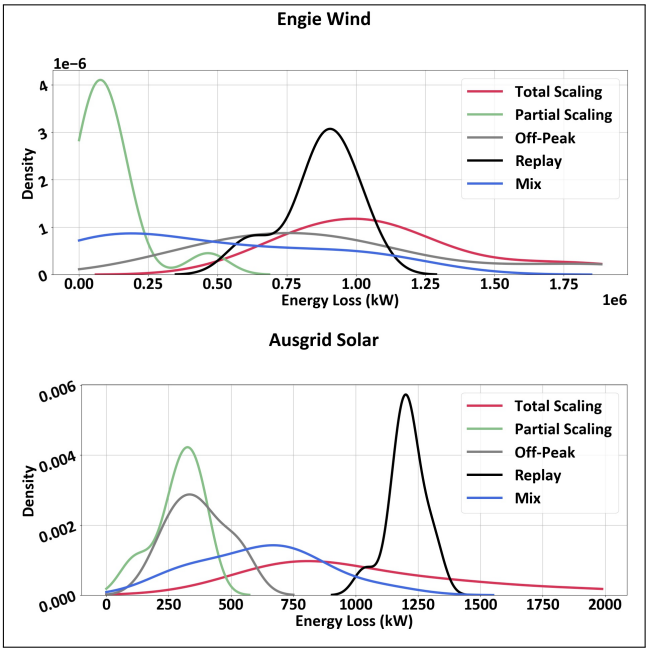

Figure 8: The density of the energy loss in wind and solar energy data.

especially for the replay or the total scaling thefts. In such cases, the fraudulent prosumer engaged in theft activities manipulate the energy generation values measured by the generation metres endowed with his/her DRES by increasing the number of green energy measurements that are reversed to the energy grid. Consequently, the energy losses increase by the discrepancy in this value, leading the utility provider to overcharge.

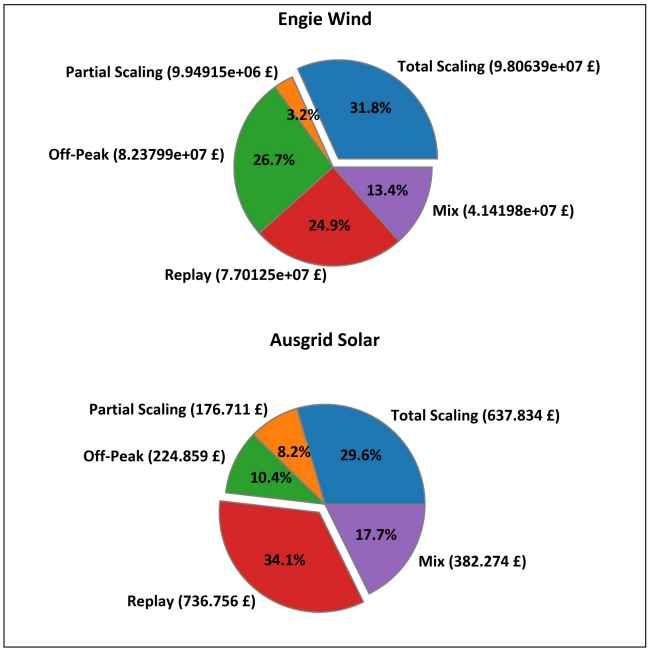

Figure 9: The amount of the monetary cost for the utility providers.

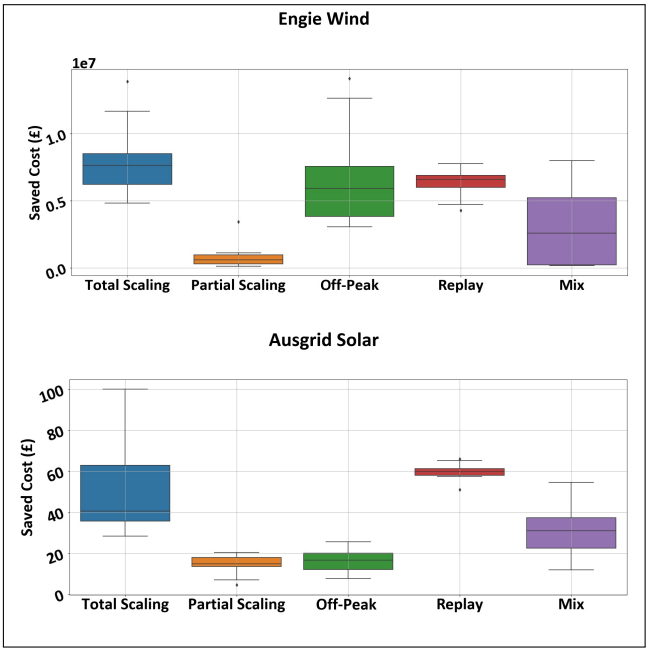

Figure 10: The saved cost by the proposed framework in wind and solar energy data.

In order to save such monetary costs incurred by the providers, an accurate detection of the energy loss cased by energy thefts is required in the first place. Fig.10 presents the saved cost provided by our proposed detection framework. As evident from this figure, for both wind and solar energy measurements, our framework saved about 82 to 99 percent of the monetary cost through detection the energy loss to large extent. The median value of the cost that can be saved by the proposed detection framework of the total theft scenario in the wind energy measurement is £7.80843e+06, while in the solar energy measurements is £60.0395. The estimate of saved costs provided by the proposed framework is without including any additional hardware equipment since the proposed framework is completely data-driven, or utilizing additional measurements that are directly unavailable to the utility providers that are only aware of the DRES capacity.

9 CONCLUSION

Energy theft attacks pose a pressing issue that has resulted in enormous non-technical energy and monetary losses to energy providers at a global scale. The integration of DRES deployments in modern energy grids in conjunction with the widely adopted business model of demand-response have undoubtedly expanded the energy theft attack surface. Conventional energy theft detection schemes heavily rely on the assessment of spatiotemporal patterns from aggregated and commonly incomplete SCADA measurements without considering the intrinsic weather or environmental patterns related to a specific DRES deployment. Therefore, in this paper we propose a data-driven SCADA-agnostic energy theft detection framework explicitly to DRES-based scenarios. We introduce a DRES-based theft attack model and further evaluate the performance of our framework by utilizing freely available third-party weather measurements over real solar and wind energy deployments in Australia and France respectively. Through our evaluations based on energy profiling model and third party weather data, we demonstrate that the proposed framework can detect fraudulent prosumers with an overall average accuracy of 98% with relatively low computational costs. Hence, placing it as a good and cost-effective candidate for future data-driven energy theft detection schemes.

REFERENCES

[1] A Vaughan. 2018. Eu raises renewable energy targets to 32% by 2030. *The Guardian*.

[2] Livia Raggi, Fernanda Trindade, Vinicius Carnelossi da Cunha, and Walmir Freitas. 2020. Non-technical loss identification by using data analytics and customer smart meters. *IEEE Transactions on Power Delivery*.

[3] Xiaodong Yuan, Min-gming Shi, and Zhengyang Sun. 2015. Research of electricity stealing identification method for distributed pv based on the least squares approach. In *2015 5th International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (DRPT)*. IEEE, 2471–2474.

[4] Varun Badrinath Krishna, Carl A Gunter, and William H Sanders. 2018. Evaluating detectors on optimal attack vectors that enable electricity theft and der fraud. *IEEE Journal of Selected Topics in Signal Processing*, 12, 4, 790–805.

[5] Mahmoud, Mostafa Ismail, Mahesh Shahin, Naidu, and Erchin Serpedin. 2020. Deep learning detection of electricity theft cyber-attacks in renewable distributed generation. *Transactions on Smart Grid*, 73–102.

[6] Ahlam Althobaiti, Anish Jindal, and Angelos K Marnerides. 2020. Scada-agnostic power modelling for distributed renewable energy sources. In *2020 IEEE 21st International Symposium on" A World of Wireless, Mobile and Multimedia Networks"(WoWMoM)*. IEEE, 379–384.

[7] Rajiv Punmiya and Sangho Choe. 2019. Energy theft detection using gradient boosting theft detector with feature engineering-based preprocessing. *IEEE Transactions on Smart Grid*, 10, 2, 2326–2329.

[8] Konstantinos V Blazakis, Theodoros N Kapetanakis, and George S Stavrakakis. 2020. Effective electricity theft detection in power distribution grids using an adaptive neuro fuzzy inference system. *Energies*, 13, 12, 3110.

[9] Donghuan Yao, Mi Wen, Xiaohui Liang, Zipeng Fu, Kai Zhang, and Baojia Yang. 2019. Energy theft detection with energy privacy preservation in the smart grid. *IEEE Internet of Things Journal*.

[10] Shalini Sharma and Angshul Majumdar. 2020. Unsupervised detection of non-technical losses via recursive transform learning. *IEEE Transactions on Power Delivery*.

[11] Sravan Kumar Gunturi and Dipu Sarkar. 2021. Ensemble machine learning models for the detection of energy theft. *Electric Power Systems Research*, 192, 106904.

[12] Kedi Zheng, Qixin Chen, Yi Wang, Chongqing Kang, and Qing Xia. 2018. A novel combined data-driven approach for electricity theft detection. *IEEE Transactions on Industrial Informatics*, 15, 3, 1809–1819.

[13] Lorelisa Ethel Luya and Michael Angelo Pedrasa. 2019. Detecting and estimating amount of energy theft in the distribution network using linear regression. In *2019 9th International Conference on Power and Energy Systems (ICPES)*. IEEE, 1–6.

[14] Zibin Zheng, Yatao Yang, Xiangdong Niu, Hong-Ning Dai, and Yuren Zhou. 2017. Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids. *IEEE Transactions on Industrial Informatics*, 14, 4, 1606–1615.

[15] Sook-Chin Yip, KokSheik Wong, Wooi-Ping Hew, Ming-Tao Gan, Raphael C-W Phan, and Su-Wei Tan. 2017. Detection of energy theft and defective smart meters in smart grids using linear regression. *International Journal of Electrical Power & Energy Systems*, 91, 230–240.

[16] Tianqi Chen and Carlos Guestrin. 2016. Xgboost: a scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785–794.

[17] Dark Sky. 2020. Dark sky API. Accessed: 2020-01-22. (2020). https://darksky.net/dev.

[18] World Weather Online. 2020. World weather onlineAPI. Accessed: 2020-01-22. (2020). https://www.worldweatheronline.com/developer/api/.

[19] OpenWeatherMap. 2020. Open weatherAPI. Accessed: 2020-01-22. (2020). https://openweathermap.org/api.

[20] Anish Jindal, Amit Dua, Kuljeet Kaur, Mukesh Singh, Neeraj Kumar, and Sukumar Mishra. 2016. Decision tree and svm-based data analytics for theft detection in smart grid. *IEEE Transactions on Industrial Informatics*, 12, 3, 1005–1016.

[21] Nacef Tazi and Youcef Bouzidi. 2020. Evolution of wind energy pricing policies in france: opportunities and new challenges. *Energy Reports*, 6, 687–692.