Classification-Based Electricity Theft Detection on Households with Photovoltaic Generation and Net Metering

Renelyn Myka P. Maala
Electrical and Electronics Engineering Institute
University of the Philippines Diliman
renelyn.maala@eee.upd.edu.ph

Andrew Miguel B. Rebamba
Electrical and Electronics Engineering Institute
University of the Philippines Diliman
andrew.rebamba@eee.upd.edu.ph

Adonis Emmanuel DC. Tio
Electrical and Electronics Engineering Institute
University of the Philippines Diliman
adonis.tio@eee.upd.edu.ph

***Abstract* — With decreasing rooftop photovoltaic (PV) costs and increased incentives to generate electricity, more and more end-users are installing rooftop PV systems and availing net metering. However, as rooftop PV and net metering become more prevalent, electricity theft detection becomes more challenging. This paper investigates the performance of features and algorithms used in classification-based theft detection algorithms on systems with rooftop PV and net metering. We explore five features and four algorithms. We use the following features computed from check meter and individual customer meter readings: gamma deviance (GD), log cosh loss (LCL), percent loss error (PLE), Poisson deviance (PD), and squared error (SE). The meter readings were simulated using the IEEE European Low Voltage Test Feeder using OpenDSS across a wide range of PV, net metering, and theft penetration levels. We then used the extracted features to train classifiers using the following algorithms: support vector machine (SVM), artificial neural network (ANN), k-nearest neighbors (KNN), and decision tree (DT). Test results showed that KNN generally performed poorly, and DT generally performed well. Moreover, models using PD and LCL as features generally displayed robustness to varying levels of PV and net metering. And finally, ANN, SVM, and DT models that use LCL and PD as features are among the highest ranked models in terms of median accuracies and range of accuracy.**

Keywords — electricity theft detection, rooftop photovoltaics, net metering, machine learning

I. Introduction

The Philippines Republic Act 9513 "An Act Promoting the Development, Utilization and Commercialization of Renewable Energy Resources and for Other Purposes" encourages customers to generate renewable resources by providing investment opportunities to energy producers [1]. It allows for net metering, a utility billing strategy that enables customers to sell excess energy generated back to the grid, to offset their consumption and reduce their bills using a net meter [2].

However, using rooftop photovoltaics (PV) and net metering presents new opportunities for some customers to steal electricity. Pilferers have two options: either relay lower readings if their consumption of electricity is higher than their generation of electricity or report more energy put into the grid to claim larger profits [3]. Exploiting these theft tactics results in financial problems for distribution utilities and users in terms of increased costs and safety risks [4].

In the literature, machine learning-based algorithms show promising performance when it comes to detecting electricity theft [5] but there is limited work on theft detection considering PV and net metering. This study aims to address this gap by testing machine learning-based classifiers that can detect electricity theft in systems with PV and net metering.

II. Electricity Theft and Detection Methods in Systems with PVs and Net Metering

A. Electricity Theft

System loss is the difference between how much energy was made and how much was sold during a certain billing period. In the Philippines, system losses stands at 9.4% in 2019, rising by 11% from the previous year [6]. Republic Act 7832 lets distribution utilities charge customers a recoverable rate of system losses up to a maximum cap [7]. This law encourages utilities to perform loss reduction programs [8] covering all sources of system losses in the distribution sector including losses from electricity theft.

There are many ways people can steal electricity, especially in the advent of smart meters. Pilferers can either interrupt the smart meter measurement, tamper with the stored demand, or modify the smart meter data. The first method prevents the pilferer's smart meter data from being recorded by physically disconnecting the meter or by meter inversion [9]. The second method allows malicious customers to hack into smart meters to report higher energy to the grid and claim more profit [5]. Another method alters the object identification system (OBIS), an identification code for measurement data and smart-meter readings [10].

One way to reduce theft is by putting meters in iron boxes and placing them at the top of the poles where they are hard to reach and tamper with [8]. Another is by using data analytics and machine learning by using historical data to look for irregularities in the system.

B. Classification-based Theft Detection Algorithms

Methods for detecting electricity theft can be classification-based, regression-based, or state-based. State-based methods use wireless sensors and RFID tags to monitor power system states but is expensive [9]. Regression based methods look at the coefficients and errors from fitted meter reading datasets to determine the presence of theft. It is low-cost and can handle incomplete information, but is susceptible to small errors caused by theft [11]. Classification-based methods use data mining and machine learning to train a model to classify energy consumption anomalies based on a testing dataset [9]. Some classification-based algorithms include the following:

1) Support Vector Machines (SVM)

SVM is an algorithm that builds an optimal hyperplane that sorts values by making a linear combination of training samples. The radial basis kernel function (RBF-SVM) is the most used kernel function, as it can handle nonlinear data

[12]. Ref. [13] used SVM to detect electricity theft in Malaysia, achieving an accuracy of 86.43% and an average hit rate of 77.41%. In 2019, Ref. [14] added principal component analysis (PCA) to the theft detection algorithm, which led to a 90% accuracy rate. In 2020, Ref. [12] used SVM for detecting meter tampering, resulting in accuracy and an F1-score of 96.96% and 95.35%, respectively.

2) Artificial Neural Networks (ANN)

An artificial neural network is a system that is based on biological neural networks [15]. There are different types depending on the number of layers and how the neurons are connected: simple artificial neural network, convolutional neural networks (CNN), feed-forward neural networks (FFNN), and long short-term memory (LSTM), among others. Several different types of ANN were used to look at the smart meter data from the State Grid Corporation of China (SGCC) to see how often electricity was stolen. Long short-term memory [16] and convolutional neural networks [17] had reported accuracy rates of 73.2% and 81.2%, respectively when used for electricity theft detection.

3) K Nearest Neighbors

K nearest neighbors (KNN) is an algorithm that uses cluster prediction to classify data [17]. In 2016, Ref. [18] used KNN to find theft in data that considered the weather, location, and load. In 2019, Ref. [19] helped the Multan Electric Power Company (MEPCO) in Punjab, Pakistan, find people who were stealing electricity. They used KNN as one of their algorithms, which has a reported 81.79% accuracy. That same year, Ref. [20] implemented KNN with feature selection techniques to determine false data injection (FDI) attacks on the system. They tested KNN on the IEEE 57-bus system, which produced an accuracy of 85.59%. A year later, in the same country, Ref. [21] used KNN to distinguish between legitimate and fraudulent energy customers and discovered that even though it achieved an accuracy rating of 91%, it showed a false positive rate (FPR) of 11.88%.

4) Decision Trees

Decision tree algorithms sort data into groups based on how sets of questions are answered. Decision nodes show the choices that were made, and leaf nodes show the results. Ref. [22] used random forest (RF) to detect electricity theft in 2021. Random forest picks a subset of features from a set of decision trees at random to prevent overfitting. They correctly classified 85% of the data, given that 10% were pilferers.

C. Theft Detection in systems with rooftop PV and net metering

Ref [3] used theft detection algorithms in homes with PV and net metering with a success rate of over 90%. The method detects anomalies in time-series meter readings without needing check meters. Ref. [23] tested algorithms using check meter readings from simulations using the IEEE European LV Test Feeder across different levels of PV and net metering penetration. Results showed that both algorithms produced acceptable results at low penetration levels but could not detect electricity theft properly in households with high penetration.

The lack of more work indicates that there aren't many reports on how to catch electricity theft in systems with PV and net metering. To address this gap in literature, it is important to study more algorithms, features, and methods for detecting theft in systems with rooftop PV and net metering.

III. Methodology

This work compares four classification-based methods and five features in detecting electricity theft in systems with PV and net metering. The study is divided into the following steps: data acquisition and processing, network modeling in OpenDSS, benign dataset creation, malicious dataset creation, feature extraction and labeling, theft detection algorithm implementation, and assessment.

A. Raw Data Acquisition and Processing

The load and PV generation profiles were obtained from the Ausgrid dataset of electricity generation and consumption of 300 customers with rooftop solar panels installed in their homes [24]. Only those from 2010 to 2011 were used because it is the only subset with half-hour intervals. Moreover, households with controllable loads were discarded, resulting in a total of 161 customer load and PV generation profiles. A week's worth of data corresponding to the week of December 5 to 11, 2010 was chosen because it is similar to the weather in the Philippines.

B. Network Modeling in OpenDSS

The IEEE European Low Voltage Test Feeder was used to model the power flows of 55 households connected to a substation, as shown in Figure 2. Additionally, eight check meters were added to measure the load consumption in different areas of the network. The IEEE-provided LV feeder files and randomly chosen customer profiles from the 161 Ausgrid customer profiles serve as the inputs to one OpenDSS simulation.

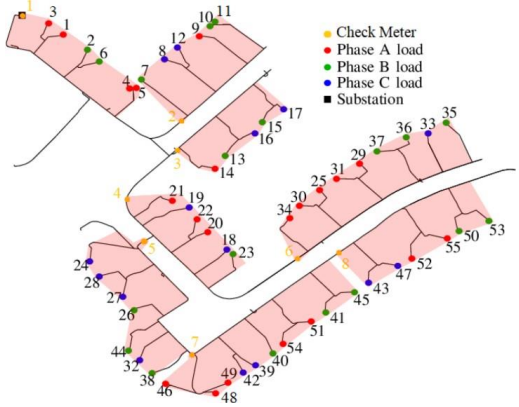

Fig. 1. IEEE European Low Voltage Test Feeder with 8 check meters installed

Households are assumed to operate in one of the following modes: (1) no PV and no net metering, (2) with PV but no net metering, and (3) with PV and net metering. If the house has neither PV nor net metering, only its PQ load is considered. If the house has PV installed but no net metering, excess generation is curtailed once PV generation

matches PQ load consumption. If the house has both PV and net metering, excess PV generation will be injected back to the grid.

C. Benign Dataset Creation

Thirteen datasets with varying levels of PV and net metering penetration were generated. The PV penetration level refers to the percentage of houses with PV installed and the net metering penetration level refers to the percentage of houses with net metering installed from the total houses with PV. There are four levels of penetration considered for both PV and net metering penetration: zero (0%), low (33%), medium (66%), and high (100%). There are a total of 1430 OpenDSS power flow simulations for the benign case, 110 for each dataset. Each simulation uses 55 customer profiles chosen at random from the 161 Ausgrid customer profiles and assigns them to the 55 customer nodes in the test feeder. Each customer profile consists of 7 days' worth of load and PV generation data in 30-minute intervals. With the customer profiles and test network parameters as input per simulation, OpenDSS was used to conduct power flow analysis to solve for the check meter readings. There is no theft in the benign dataset and hence, discrepancies between the check meter and customer meter readings are due to technical losses in the network only.

D. Malicious Dataset Creation

Like the benign dataset, 1430 simulations were also conducted for the creation of the malicious dataset. That is, a different set of 55 customer profiles are randomly chosen and assigned to the 55 customer nodes in the test feeder per simulation. OpenDSS was then used to solve for the check meter readings using the actual customer profiles.

To represent theft, one pilferer was chosen per simulation. The pilferer changed the reading on their meter with a multiplier k between 35% and 65% to either relay lower readings (if the net consumption is positive) or report more energy to the grid (if the net consumption is negative). The malicious value for the pilferer's net meter reading, X, was computed using Eq. 1, where x is the net consumption.

$$X = \begin{cases} X * k & \text{if } x > 0 \\ X - (|X| * k) & \text{if } x < 0 \end{cases} \quad (1)$$

Two types of theft frequencies were simulated, full-day and half-day. The result is that for each dataset, there are 110 simulations wherein each of the 55 households was designated as the culprit once for each of the full-day and half-day frequencies. Full-day theft means that Eq. (1) is applied for 24 contiguous hours a day while half-day theft means that Eq. (1) is applied for 12 contiguous hours.

E. Feature Extraction and Labeling

The readings from the check meters and customer meters were turned into daily frequencies by getting the total check meter and customer meter consumption for the day. Each daily value was given a 1 or 0 to indicate whether theft happened that day or not. Since each of the 1430 benign and 1430 malicious simulations uses a week's worth of data, each simulation generates 7 raw data points of daily check and customer meter readings labeled with 1 or 0. This results in a total of 20,020 labeled raw data points in total with 1540 raw data points per dataset, with the same number of benign and malicious samples for each of the 13 datasets.

The next step is to process the raw data points to compute the features that will be used for classification. We considered five features, namely: (1) gamma deviance (GD), (2) log-cosh loss (LCL), (3) percent loss error (PLE), (4) Poisson deviance (PD), and (5) squared error (SE). The corresponding equations are shown in Equations 2-6 where CM is the check meter reading and M_n are individual household meter readings under the check meter.

$$Gamma\ Dev = 2(log(\frac{\sum_{n=1}^{k} M_n}{CM}) + \frac{CM}{\sum_{n=1}^{k} M_n} - 1) \quad (2)$$

$$Logcosh = log(cosh(\sum_{n=1}^{k} M_n - CM)) \quad (3)$$

$$\%\ Loss\ Error = \frac{|\sum_{n=1}^{k} M_n - CM|}{CM} \quad (4)$$

$$Poisson\ Deviance = 2(CM\ log(\frac{CM}{\sum_{n=1}^{k} M_n}) + \sum_{n=1}^{k} M_n - CM) \quad (5)$$

$$Squared\ Error = (CM - \sum_{n=1}^{k} M_n)^2 \quad (6)$$

The percent loss error and squared error compares the difference between the check meter readings and the sum of downstream meter readings in a straightforward manner. On the other hand, the gamma deviance, log cosh loss, and Poisson deviance transform the data to compare the readings. These features magnify the difference between the check meter reading and the customer meter readings to aid in better classification.

Each of the five features were extracted separately for for each of the 8 check meters and for each of the thirteen datasets. In the end, each final data point is a 9x1 vector with 8 values corresponding to the value of the feature computed per check meter and 1 value corresponding to the label, benign or malicious. In total, 65 unique datasets were created (13 datasets of varying PV and net metering levels * 5 features) after this step, with 1,540 final data points per dataset.

F. Theft Detection Algorithm Implementation

Each of the 65 datasets was split into training (80%) and testing (20%) datasets. Four algorithms were used: k-nearest neighbors (KNN), support vector machines (SVM), artificial neural networks (ANN), and decision trees (DT). This results in 260 models consisting of 20 classifier-feature pairs trained using the 13 datasets of varying PV and net metering levels. Built-in functions in Python were used to implement the algorithms.

For SVM, the `svc` function was used from the Scikit-learn library. The radial basis kernel function was used to handle non-linear data. Grid search and ten-fold cross validation was used to tune the cost and gamma hyperparameters. For ANN, functions from the Keras library was used. The following hyperparameters were tuned during hyperparameter optimization: optimizer to be used, learning rate, number of nodes, type of activation function, number of epochs, and batch size. For KNN, the `neighbors` function was used from the Scikit-learn library.

And lastly for DT, the `DecisionTreeRegressor` function was used from the Scikit-learn library.

G. Performance Metrics

The algorithms' performances were evaluated based on their accuracy metric presented in Eq. 7. *TP*, *TN*, *FP*, and *FN* stand for true positive, true negative, false positive, and false negative, respectively.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{7}$$

IV. Results and Discussion

Figs. 2-5 show the box and whiskers plot of the accuracy of the 20 classifier-feature pairs while Table 1 shows the underlying data table. The span of the whiskers reflects the variance in accuracy when the PV and net metering penetration is varied.

Figs. 2-3 show that the performance of SVM and ANN follow the same trend. When GD and PLE are used to train the SVM and ANN models, the median accuracy is high at 88.5-90.9% but the variance in accuracy is large with lows down to 52.4-62.3%. This means that the performance of these algorithms generally degrades as more customers use rooftop PV and net metering. When LCL and PD are used to train the SVM and ANN models, the median accuracy is better at 92.0-93.2% and the variance in accuracy is small. This indicates good performance that is robust across varying levels of PV and net metering penetration. Lastly, when SE is used to train the SVM and ANN models, the median accuracy degrades to just 76.8-80.5% but the variance in accuracy is small. This indicates poor classification performance but is robust across varying PV and net metering penetration.

Fig. 4 shows that the performance of KNN has a similar trend with that of SVM and ANN but has much lower median accuracies of only 67.9%, 74.0%, 73.0%, 72.1%, and 63.0% when using GD, LCL, PLE, PD, and SE as features respectively. This indicates a generally worse performance in terms of classification accuracy relative to using SVM and ANN.

Fig. 5 shows that the accuracy of all DT models is better than 70% regardless of the feature used or PV and net metering penetration level. The median accuracies are at 90.6%, 92.9%, 89.6%, 91.9%, and 91.6% when using GD, LCL, PLE, PD, and SE as features respectively. The range in accuracy is also small indicating robustness to varying PV and net metering penetration levels.

In terms of algorithm-feature pairs with the best median accuracy, SVM-PD ranked the best with 93.2% median accuracy followed by DT-LCL and SVM-LCL with 92.9% both. In terms of the range of accuracies across varying PV and net metering levels, ANN-PD has the smallest range spanning 6.17%, followed by KNN-SE spanning 6.5% and ANN-LCL spanning 8.6%.

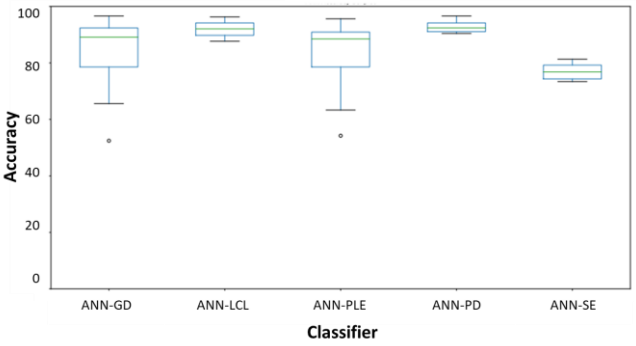

Fig. 2 ANN classifiers accuracy boxplot

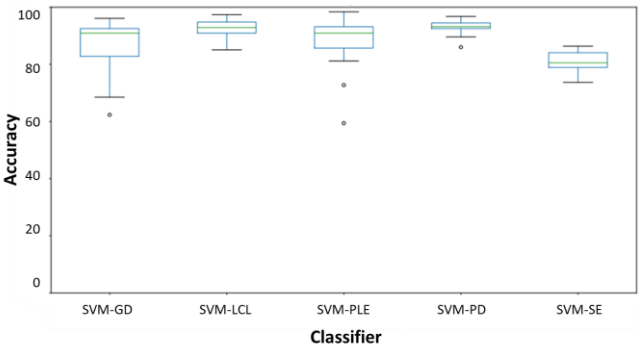

Fig. 3 SVM classifiers accuracy boxplot

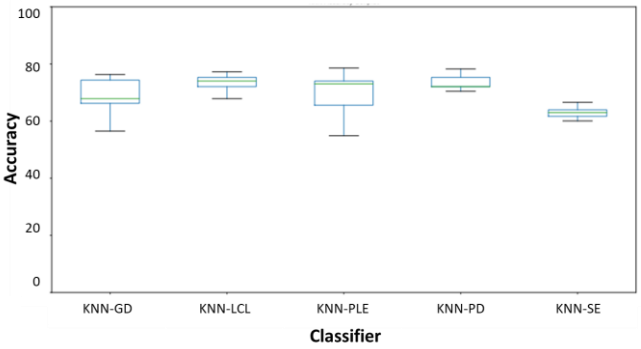

Fig. 4 KNN classifiers accuracy boxplot

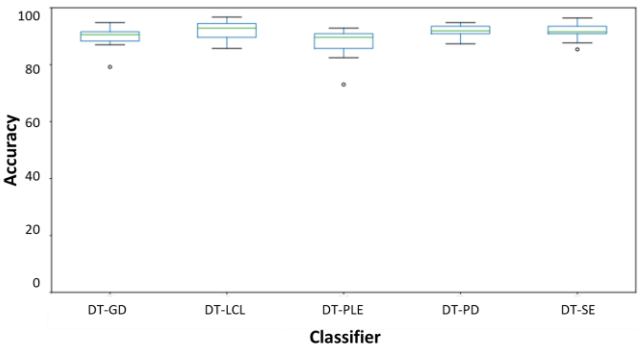

Fig. 5 DT classifiers accuracy boxplot

TABLE I: CLASSIFIER-FEATURE PAIR ACCURACY PER DATASET

Dataset	1	2	3	4	5	6	7	8	9	10	11	12	13
% PV	0	33	33	33	33	66	66	66	66	100	100	100	100
% NM	0	0	33	66	100	0	33	66	100	0	33	66	100
SVM-GD	96.1	92.5	92.2	87.7	93.5	91.9	82.8	93.8	85.4	68.5	90.9	74.0	62.3
SVM-LCL	97.4	92.9	90.9	85.1	94.8	92.2	89.9	96.4	93.5	88.6	97.4	94.5	90.3
SVM-PLE	98.4	94.5	93.2	88.3	94.2	90.9	81.2	93.2	85.7	72.7	92.9	85.7	59.4
SVM-PD	96.8	94.2	94.5	86.0	93.2	93.2	89.6	93.8	91.9	92.5	96.8	95.8	93.2
SVM-SE	84.4	83.4	80.5	77.9	86.4	78.9	73.7	84.1	79.9	77.3	84.4	82.1	79.9
ANN-GD	96.6	92.4	90.6	89.1	92.4	89.8	73.9	92.5	80.1	65.6	83.1	78.6	52.4
ANN-LCL	96.3	92.5	89.0	87.7	92.1	89.8	88.8	94.2	89.8	89.8	94.2	94.8	92.4
ANN-PLE	95.6	93.5	90.3	89.8	93.2	90.9	78.6	88.5	80.8	63.3	85.9	77.3	54.2
ANN-PD	96.6	93.2	90.6	91.1	94.5	92.2	90.6	94.2	92.4	90.4	94.3	94.0	91.6
ANN-SE	81.3	79.1	73.9	74.4	81.3	73.5	73.4	79.8	75.8	79.2	76.0	78.9	76.8
KNN-GD	75.7	74.7	71.8	67.5	74.4	71.1	66.6	76.3	66.2	63.3	67.9	62.0	56.5
KNN-LCL	76.0	74.0	73.1	71.8	73.1	72.1	67.9	75.3	74.4	75.0	77.3	75.7	69.8
KNN-PLE	78.6	76.3	74.0	73.7	76.0	73.4	67.2	73.1	65.6	57.8	72.1	58.8	54.9
KNN-PD	78.3	72.1	71.4	70.8	75.3	73.1	72.1	77.9	74.0	72.1	76.6	72.1	70.5
KNN-SE	66.6	64.9	64.0	63.0	64.9	63.6	61.7	64.0	62.7	61.4	62.0	61.7	60.1
DT-GD	94.8	92.2	91.6	89.0	90.6	91.2	88.3	92.5	87.0	87.0	90.9	88.0	79.2
DT-LCL	96.4	94.5	93.5	87.1	93.2	89.6	85.7	95.1	89.9	88.0	96.8	92.9	88.0
DT-PLE	92.9	90.6	89.6	89.0	92.2	91.9	89.0	90.9	85.7	82.5	89.9	85.1	73.1
DT-PD	94.5	91.2	90.3	89.0	93.2	91.9	87.3	94.8	91.9	89.3	95.1	93.5	92.9
DT-SE	96.4	91.6	90.9	85.4	93.5	91.2	87.7	92.5	93.8	89.9	95.1	92.5	90.9

While there is no definite "best" algorithm-feature pair from the models tested, KNN generally performed poorly relative to the others with overall lower median accuracies while DT generally performed well with high median accuracies and small accuracy ranges. Models using PD and LCL as features generally displayed robustness to varying levels of PV and net metering penetration as evidenced in the small accuracy ranges of the models using these as features. And finally, ANN, SVM, and DT models that use LCL and PD as features are among the highest ranked models in terms of median accuracies and range of accuracy.

To explore why LCL and PD generally performed well as features, we plot the histograms of the five features for both benign and malicious datasets for Dataset 13 (100% PV and 100% net metering penetration) in Figs. 6-10. For GD, PLE, and SE, the histograms of the features have large overlaps indicating that models built using these features will find it difficult to distinguish between malicious and benign data. On the other hand, for LCL and PD, the overlap is much smaller indicating that models built using these features will find it easier to distinguish between malicious and benign data.

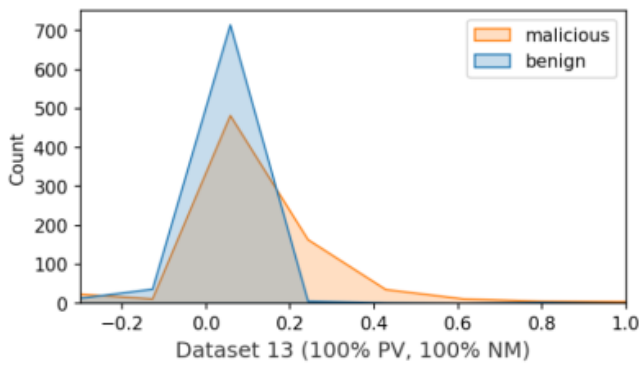

Fig. 6 Gamma deviance histogram

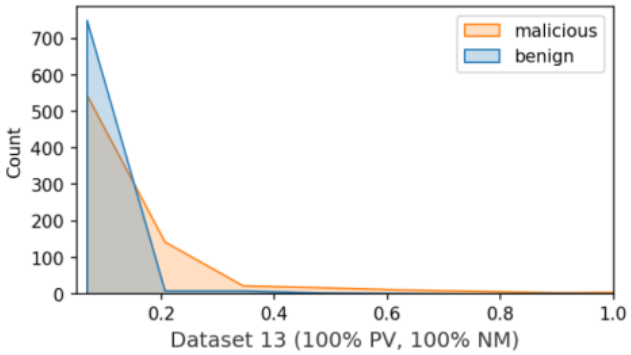

Fig. 7 Percent loss error histogram

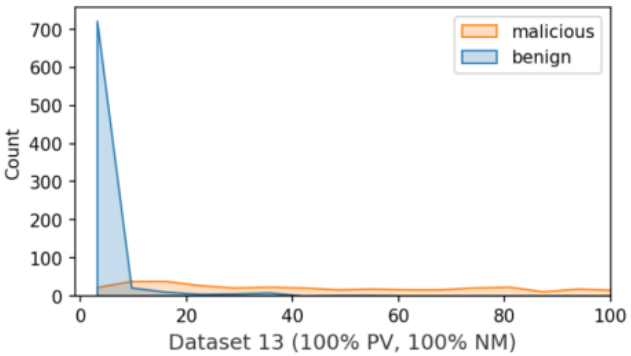

Fig. 8 Squared error histogram

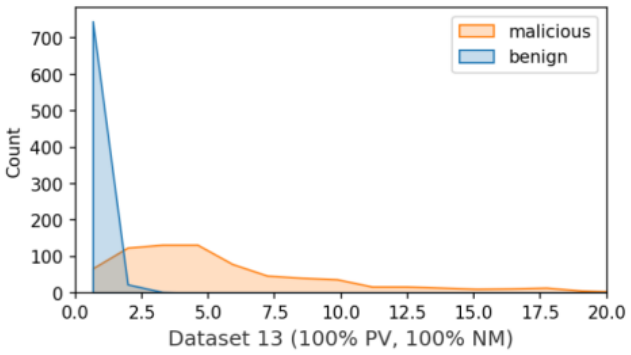

Fig. 9 Log cosh loss histogram

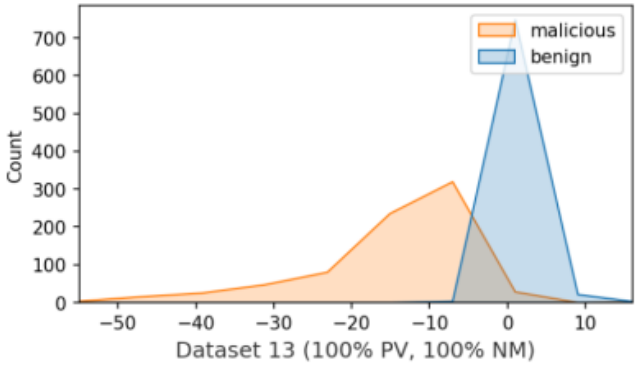

Fig. 10 Poisson deviance histogram

V. CONCLUSION

This study tested four algorithms and five features to detect electricity theft in distribution systems with rooftop solar panels and net metering. For the algorithms, we looked at the following: support vector machines (SVM), artificial neural networks (ANN), k nearest neighbors (KNN), and decision trees (DT). For the features, we looked at the

following features computed from the check meter readings and the customer meter readings: gamma deviance (GD), log-cosh loss (LCL), percent loss error (PLE), Poisson deviance (PD), and squared error (SE).

Results show that KNN generally performed poorly with lower median accuracies while DT generally performed well with high median accuracies and small accuracy ranges across datasets. Models using PD and LCL as features generally displayed robustness to varying levels of PV and net metering penetration levels. And finally, ANN, SVM, and DT models that use LCL and PD as features are among the highest ranked models in terms of median accuracies with SVM-PD being the best with 93.2% median accuracy followed by DT-LCL and SVM-LCL with 92.9% both.

Future work will extend the simulations to other test systems with varying test system sizes and topologies, number of check meters, number of customers doing theft, and amount of electricity stolen.

REFERENCES

[1] Senate and the House of Representative. Republic Act No. 9513. officialgazette.gov.ph/2008/12/16/republic-act-no-9513/, 08 2008.

[2] K.Maharaja, P.Balaji, S.Sangeetha, and M.Elakkiya. Development of bidirectional net meter in grid connected solar PV system for domestic consumers, 2016 International Conference on Energy Efficient Technologies for Sustainability (ICEETS), pp 46–49, 2016.

[3] M.Badr, M.Ibrahem, Mohamed Baza, Mohamed Mahmoud, and Waleed Alasmary. Detecting electricity fraud in the net- metering system using deep learning, 2021 International Symposium on Networks, Computers and Communications (ISNCC), pp 1–6, 2021.

[4] H.Huang, S.Liu, and K.Davis. Energy Theft Detection Via Artificial Neural Networks, 2018 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), pp 1–6, 2018.

[5] M.Shaaban, U.Tariq, M.Ismail, N.Almadani, and M.Mokhtar, Data-driven detection of electricity theft cyberattacks in pv generation. IEEE Systems Journal, 2021.

[6] Department Of Energy. 2019 Power Demand and Supply Highlights. In 2019 Power Demand and Supply Highlights, pp. 62, January 2020.

[7] A.Flores,Meralco systems loss at Record Low, https://manilastandard.net/business/201073/meralco-systems-loss-at- record-low.html, Nov -1.

[8] Republic of the Philippines Energy Regulatory Commission. FAQ on Systems Loss. https://www.erc.gov.ph/.

[9] R.Jiang, R.Lu, Y.Wang, J.Luo, C.Shen, and X.Shen. Energy-theft detection issues for advanced metering in- frastructure in smart grid, Tsinghua Science and Technology, 19(2), pp. 105– 120, 2014.

[10] R.Czechowski and A.Kosek, The most frequent energy theft techniques and hazards in present power energy consumption, 2016 Joint Workshop on Cyber-Physical Security and Resilience in Smart Grids (CPSR-SG), pp. 1–7, 2016.

[11] L.Luya and M.Pedrasa, Detecting and estimating amount of energy theft in the distribution network using linear regression, 2019 9th International Conference on Power and Energy Systems (ICPES), pp. 1–6, 2019.

[12] A.Aliangan and L.Cacnio, A Comparative Study on Electricity Theft Detection Using Support Vector Machine and Artificial Neural Networks, June 2020.

[13] J.Nagi, A.Mohammad, K.Yap, S.Tiong, and S.Ahmed. Non-tech nical loss analysis for detection of electricity theft using support vector machines, 2008 IEEE 2nd International Power and Energy Conference, pp. 907–912. 2008.

[14] R.Toma, M.Hasan, A.Nahid, and B.Li, Electricity theft detection to reduce non-technical loss using support vector machine in smart grid, 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), pp. 1–6, 2019.

[15] G.Gu, Q.He, B.Wang, and B.Dai, Comparison of Machine Learning Techniques for the Detection of the Electricity Theft, 2018 IEEE 3rd International Conference on Cloud Computing and Internet of Things (CCIOT), pp. 116–119, 2018.

[16] M.Adil, N.Javaid, U.Qasim, I.Ullah, M.Shafiq, and J.Choi, LSTM and bat-based RUSBoost approach for electricity theft detection, Applied Sciences, 10(12):4378, 2020.

[17] Z.Khan, M.Adil, N.Javaid, M.Saqib, M.Shafiq, and J.Choi, Electricity theft detection using supervised learning techniques on smart meter data, Sustainability, 12(19):8023, 2020.

[18] S.Han, J.No, J.Shin, and Y.Joo, Conditional abnormality detection based on AMI data mining, IET Generation, Transmission & Distribution, 10(12):3010–3016, 2016.

[19] M.Saeed, M.Mustafa, U.Sheikh, T.Jumani, and N.Mirjat, Ensemble bagged tree based classification for reducing non-technical losses in multan electric power company of Pakistan, Electronics, 8(8):860, 2019.

[20] J.Sakhnini, H.Karimipour, and A.Dehghantanha, Smart grid cyber attacks detection using supervised learning and heuristic feature selection, 2019 IEEE 7th international conference on smart energy grid engineering (SEGE), pp. 108–112, 2019.

[21] S.Aziz, S.Naqvi, M.Khan, and T.Aslam, Electricity Theft Detection using Empirical Mode Decomposition and K-Nearest Neighbors, 2020 International Conference on Emerging Trends in Smart Technologies (ICETST), pp. 1–5, 2020.

[22] G.Lin, X.Feng, W.Guo, X.Cui, H.Liu, W.Jin, Z.Lin, and Y.Ding, Electricity theft detection based on stacked autoencoder and the undersampling and resampling based random forest algorithm, IEEE Access, 9:124044–124058, 2021.

[23] C.Lavilla, M.Osorio, Z.Restituto, and A.Tio, Effect of Net Metering and Rooftop Photovoltaics on Electricity Theft Detection, University of the Philippines, 2021.

[24] Ausgrid - Solar home electricity data, https://www.ausgrid.com.au /Industry/Our-Research/Data-to-share/Solar- home-electricity-data.