Effect of Check Meter Quantity on Theft Detection in Distribution Networks with Rooftop PV and Net Metering

Undergraduate Project Proposal

by

Kaira Maxine V. Gonzales
2020-11069
BS Electrical Engineering

Carlos Demetri S. Vicencio
2020-03566
BS Electrical Engineering

Adviser:

Adonis Emmanuel D.C. Tio, Ph.D.

University of the Philippines, Diliman

January 2024

Abstract

Effect of Check Meter Quantity on Theft Detection in Distribution Networks with Rooftop PV and Net Metering

The increasing demand for renewable energy has led to the rise of rooftop photovoltaics (PV) and net metering (NM). However, these technologies add a new dimension to energy theft detection, necessitating a more complex approach. This study aims to explore the effect of varying the number of check meters in distribution networks with rooftop PV and NM on theft detection algorithms, specifically for detecting meter tampering in households. The distribution network will be modeled using the Ausgrid Dataset and the Reduced IEEE European Low Voltage Test Feeder on OpenDSS and Python. Several simulations will be conducted by varying levels of PV and NM penetration, and the number of houses connected per check meter. The simulation data will be used to extract the Poisson Deviance and Log Cosh Loss, which will then be used as input to four machine learning algorithms: Support Vector Machine (SVM), Artificial Neural Networks (ANN), Decision Tree (DT), and Random Forest (RF). The performance of each algorithm will be measured using accuracy, precision, recall, and F-measure. By comparing the performance of different algorithms and varying the number of check meters, the study will provide valuable insights into the development and implementation of theft detection systems. This will not only reduce the risk of electricity theft but also further promote the transition to sustainable energy.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

As the shift to renewable energy continues, the Philippine Energy Plan (PEP) 2020-2040 was created to foster the use of clean energy in the Philippines [1]. The PEP introduces policies that reinforce the integration of more rooftop photovoltaics (PVs) and the implementation of Net Metering (NM). Through NM, consumers with rooftop PVs are able to generate their own electricity and sell the excess electricity generated back to the distribution utility. This reduces their electricity bills and gives them an opportunity to generate profit [2].

With the environmental and financial benefits provided by rooftop PVs and NM, more consumers are opting for these technologies. However, these also provide new ways for consumers to tamper with meter readings. Aside from reporting a lower energy consumption, pilferers could now also report more energy fed into the grid to increase profit [3]. Acts of electricity theft such as this results in financial losses and power quality problems [4]. As such, researchers must find ways to strengthen electricity theft detection methods.

One such method that has shown huge potential is through the usage of machine learning (ML) algorithms [5]. Previous studies [6] [7] analyzed the capability of ML algorithms to detect theft occurrences on households with rooftop PV and NM using features derived from the energy consumption and check meter readings as inputs. Check meters offer data on the electricity supplied to downstream households, which can then be contrasted with the sum of the downstream household readings. The results of these studies show that increasing the number of consumers with NM in the system reduces the accuracy of the ML algorithms used. While these studies provide more information on the performance of energy theft detection methods on networks with rooftop PV and NM, they used an unrealistic system with several check meters, which could be costly to implement.

This project will study the effect of check meter quantity on the accuracy of theft detection in distribution networks with PV and NM. Furthermore, it will also study the performance of the theft detection system under varying levels of PV and NM penetration and different machine learning algorithms.

1.2 Structure of the Document

The rest of this proposal is structured as follows. Chapter 2 discusses several works on electricity theft and the methods used to address these problems. These methods include the performance of different ML algorithms in detecting electricity theft. Chapter 3 presents the problem statement which inspired the formulation of the project together with the project objectives and the study's scope and limitations. Chapter 4 outlines the proposed methodology to achieve the goals of the study. Lastly, Chapter 5 contains the proposed timeline and deliverables for the whole project.

Chapter 2

Review of Related Work

2.1 System Loss

System losses refer to the difference between the energy delivered to distribution utilities and the energy consumed by the customers. It is a huge problem being faced by the power industry as these losses cost billions of pesos - a burden shared by both the distribution utility and its customers. In order to recover this added expense, consumers are billed extra based on their energy consumption. These added charges are termed as system loss charges. For private utilities, distribution utilities are allowed to charge up to 8.5% of the consumers' purchased kilowatt-hour for system loss charges. While for electric cooperatives, their system loss charges are capped at 13% of their purchased electricity (in kWh) [8]. Due to costs brought by system losses, the Energy Regulatory Commission (ERC) implemented a cap on distribution feeder loss to 5.5% [9]. These policies encourage distribution utilities to implement action towards minimizing all types of system losses: administrative, technical, and non-technical.

2.1.1 Admnistrative Loss

The Philippine Distribution Code defined administrative loss as the energy used by the distribution utility to perform its operations. This includes electricity used by distribution substations, offices, warehouses and workshops, and other essential loads. Regulations in the Philippines capped the amount of energy used for this purpose at 1% of the total energy. [10]

2.1.2 Technical Loss

Technical losses are caused by power dissipation in the different equipment, devices, and conductors that are used during transmission and distribution of energy. Some examples of technical losses include copper loss (I2R loss) caused by the finite resistance of conductors, dielectric loss caused by heating in the dielectric between conductors, and induction and radiation loss from the electromagnetic fields around the conductors [11]. Several measures are being used by MERALCO and different electric cooperatives to reduce technical losses. These include the installation of substations and capacitor banks, reconfiguration of line routes, equipment upgrades, and lines and load management through analytic software [12]. While various methods can be employed, completely eliminating technical losses is unattainable.

2.1.3 Non-technical Loss

Non-technical losses (NTL) refer to energy losses in the system due to external actions not controlled by distribution utilities. These include energy theft and errors in meter readings. Unlike technical losses, NTL is difficult to measure due to lack of information on the end-users and their energy consumptions [11]. A major component of NTL is electricity theft as it makes up a significant bulk of it [13]. This highlights the need for more approaches that address electricity theft.

2.2 Electricity Theft

Electricity theft is a considerable issue within the energy sector as it results in significant financial losses and power quality issues. It's estimated that around 96 billion dollars are lost globally in a year due to electricity theft [14]. Additionally, utilities are unable to predict and manage the illegal consumption of electricity, which affects the balance between the scheduled and actual demand. This negatively impacts the quality of electricity supplied to paying customers. As such, it is important for distribution utilities and electric cooperatives to develop strategies that address this issue.

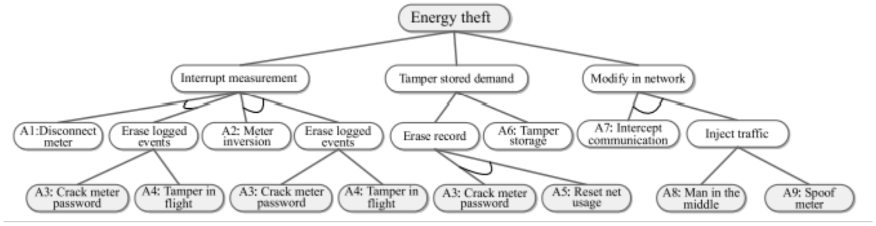


Figure 2.1: Electricity Theft Methods [15]

2.2.1 Types of Electricity Theft Attacks

There are three types of electricity theft attacks: interruption of measurement, stored demand tampering, and network modification [15]. In interruption of measurement, the pilferer physically disconnects or reset the meters to prevent it from recording further consumption. In stored demand tampering, the meter readings are manipulated to gain profit. The pilferer could either decrease their energy consumption, or increase their energy generation for cases with NM. Finally, in network modification, the pilferer hijacks the meter's communication system to deliver forged data. Figure 2.1 outlines the different types of electricity theft and their manifestations.

2.2.2 Electricity Theft Solutions

In addressing electricity theft, different measures could be employed depending on the situation. To determine the most appropriate solution, two aspects could be considered - function and approach.

Electricity theft solutions generally have one of the following functions: to prevent, detect or identify. Electricity theft prevention methods include the installation of meters inside iron boxes or at high places to prevent consumer access, and the regular repair and replacement of defective meters [12]. These measures allow the occurrence of theft to be avoided completely.

On the other hand, detection and identification methods are reactive in nature since they respond to electricity theft attacks after they have happened. Solutions that fall under these two categories typically go hand-in-hand since theft needs to be detected first before it can be identified. Detection methods usually involve a comparison between the distribution transformers readings and the sum of individual meter readings to detect discrepancies while identification methods usually involve the analysis of individual consumer data to pinpoint pilferers [16].

The rise of smart grid infrastructures has allowed more smart meter-based approaches for electricity theft detection and identification. Smart meters provide more information on energy consumption and it has the ability to communicate with distribution utilities to facilitate easier monitoring and billing [17]. These devices could also be used along with check meters to better monitor and identify theft occurrences [18]. Check meters are devices used by the distribution utility to measure the supplied electricity to one or more consumers. Readings of the check meter can be compared to the total readings of the downstream households to detect or identify theft since its readings do not change even if meter tampering occurs. However, the difference between the readings of the check meters and the sum of downstream households must be significant enough because the two measurements will never match regardless of theft due to losses in the distribution of electricity [6]. This behavior, paired with the finite resolution of metering devices, makes theft identification methods more difficult to implement using check meters [19]. As such, this study will focus on electricity theft detection methods.

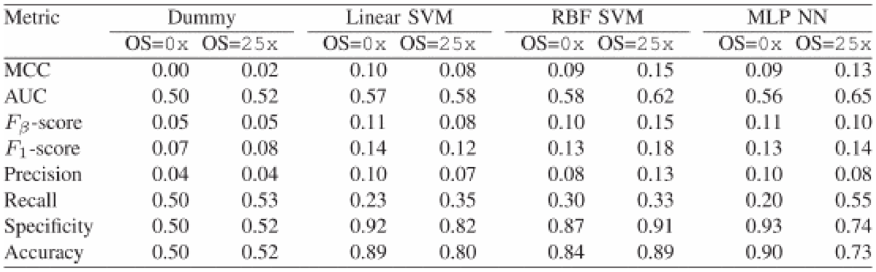
In terms of approach, electricity theft detection methods could be data-oriented, network-oriented, or a hybrid between the two [20]. Data oriented solutions include the use of consumer related information like electric consumption and consumer type. From these information, machine learning algorithms are used to differentiate theft occurrences from non-theft occurrences [20]. On the other hand, network-oriented techniques include data about the network such as topology and measurements. Examples include state estimation, load flow analysis and special sensors. Since these techniques use information from the system's network analysis and physical rules, prior knowledge on theft and non-theft occurrences is not required. Finally, hybrid methods are the combination of the two aforementioned categories. Typically, techniques from both categories are used to further localize the theft occurrence.

An analysis on these methods was performed in [20]. Based on their study, network-oriented methods have better performance, require less but more precise data, are immune to class imbalance, and have faster response time. However, these advantages come with added cost and complexity. The same trend could be observed with hybrid techniques with significant network-oriented aspects. Despite these advantages, data-oriented methods are the most popular due to the economic, precision, and technical concerns brought about by network-oriented and hybrid techniques [21]. Moreover, data-oriented systems are easier to implement on existing infrastructures since almost all of these methods only make use of smart metering data. Minimal costs are incurred to acquire the needed data inputs. As such, this method will be utilized in this study.

2.3 Machine Learning Electricity Theft Detection Methods

Machine learning has been a popular data-oriented method for electricity theft detection. ML can be used to extract information about a distribution system based on different data obtained from the system. There are two types of ML techniques: clustering and classification [22]. Clustering is an unsupervised technique wherein the data points are divided depending on similarities of the features used. On the other hand, classification is a supervised technique that assigns a label to new or unseen data points based on past data. This method requires training the model in order for it to identify patterns or trends that will help in classifying data. For the purpose of this project, studies with electricity theft detection using the classification technique will be presented.

Support Vector Machine (SVM) is a supervised machine learning algorithm that can be used for classification [23]. In SVM, data points are mapped in a high-dimensional feature space that will classify the data, even if the data is not linear. A study by Figueroa et al. used SVM to detect electricity theft based on smart meter readings of consumers in Honduras [24]. This dataset contained the daily consumption of customers - both fraudulent and non-fraudulent. Since the pilferers were known in the dataset, their behavior was then used to train different models to detect energy theft. The classifiers used were Linear SVM, Nonlinear SVM with Radial Basis Function (RBF) as the kernel function, and Multilayer Perceptron Neural Network (MLP-NN). They also performed cross-validation through oversampling to avoid overfitting, which improved the performance of SVM-RBF. The summary of their results is shown in Figure 2.2.



Metric	Dummy		Linear SVM		RBF SVM		MLP NN	
	OS=0x	OS=25x	OS=0x	OS=25x	OS=0x	OS=25x	OS=0x	OS=25x
MCC	0.00	0.02	0.10	0.08	0.09	0.15	0.09	0.13
AUC	0.50	0.52	0.57	0.58	0.58	0.62	0.56	0.65
F_β-score	0.05	0.05	0.11	0.08	0.10	0.15	0.11	0.10
F_1-score	0.07	0.08	0.14	0.12	0.13	0.18	0.13	0.14
Precision	0.04	0.04	0.10	0.07	0.08	0.13	0.10	0.08
Recall	0.50	0.53	0.23	0.35	0.30	0.33	0.20	0.55
Specificity	0.50	0.52	0.92	0.82	0.87	0.91	0.93	0.74
Accuracy	0.50	0.52	0.89	0.80	0.84	0.89	0.90	0.73

Figure 2.2: Summary of Results from [24]

Another study that used SVM-RBF is by Nagi et al. [25]. They used historical customer consumption data from the electronic-Customer Information Billing System (e-CIBS) in Malaysia to train the algorithm to detect abrupt changes in the load profile. Besides the consumption, the credit worthiness rating (CWR) of each customer was also used to strengthen their fraud detection model. The CWR quantifies how much a customer avoids or delays bill payments, which aided in determining the likelihood of theft. The model was then trained using the Grid-Search method in order to find the best values for RBF kernel parameters and cost parameters of SVM. Their model has an accuracy and hit rate of 86.43% and 77.41%, respectively.

Both studies from [24] and [25] showed that SVM-RBF has a good performance in electricity theft detection using readings from individual household meters. Besides electricity consumption, [25] used CWR as an additional feature to help in detecting fraud. [24] showed that using SVM-RBF is better when oversampling is used compared to Linear-SVM and MLP-NN. From the results of these two studies, it can be observed that using SVM is a good machine learning algorithm for electricity theft detection. Furthermore, using RBF as a kernel function improves the performance of the model because it can handle non-linear data.

Artificial Neural Networks (ANN) is a computational network that maximizes the use of multiple layers of interconnected nodes [26]. Huang et al. used ANN using data from both check meter and consumer's meter readings [27]. Their model compares the sum of consumer's meter readings to the reading of a check meter installed after the distribution transformer. If there is a significant mismatch between the two, it means

Table 2.1: Summary of Results from [32]

Algorithm	**Accuracy (%)**
Linear SVM	85.52
ANN	91.46
RF	91.63
LR	88.21

that there is electricity theft. Their study presented that such ANN model can be used for electricity theft detection but it did not present performance metrics to quantify how well it performed. Furthermore, their study was not able to provide insights on how the check meter quantity or placement affects the model.

Bohani et al. also conducted a study that used several supervised machine learning methods on a dataset containing electricity consumption data from household meters in China [28]. Similar to [24], this dataset contained both honest and known malicious customers. Their behavior was then used to train electricity theft detecting models using ANN, Deep ANN (DANN), Decision Tree (DT), and Adaboost. DANN is similar to ANN but it uses multiple hidden layers instead of just one [29]. DT is another type of supervised learning algorithm which is used for classification problems relating to regression analysis [30]. Adaboost is an ensemble classifier that combines weak learners such as DT [31]. Different ratios of training and testing samples were used in their study to analyze the effect of ratio in the performance of their model. Overall, the 70/30 splitting ratio had the best results wherein ANN and DT had accuracies of 92.44% and 92.47% respectively. These results show that both are viable machine learning algorithms that can be used for electricity theft detection.

Pereira et al. also used the same dataset and theoretical framework as [28]. They used 80% of this dataset for training and the remaining 20% for testing. Given the nature of data from electricity theft being unbalanced wherein the number of pilferers is much less compared to the number of honest customers, they studied the effects of unbalanced data handling techniques in machine learning algorithms for electricity theft detection. The algorithms that they used were Linear SVM, ANN, RF, and Logistic Regression (LR). Random forest (RF) has a similar procedure as DT, but it makes use of multiple trees [33]. On the other hand, LR is an algorithm that classifies data by prediction of the outcome [34]. Their results showed data handling techniques decreases the accuracies of the algorithms. Table 2.1 shows the accuracy from their simulations when no data handling techniques were used. It can be seen that both RF and ANN performed better than Linear SVM and LR. Essentially, their study presents how using unbalanced data handling techniques can degrade the performance of theft detection systems.

Tehrani et al. [35] conducted a study that used a dataset containing electricity consumption readings of single-family apartments to train and test several tree-based classification methods. All of these customers were assumed to be honest, then malicious sample generator functions were used to generate various attack patterns based on the consumer's data. They then compared the performance of DT, RF and Gradient Boosting (GB) with and without clustering. GB is similar to RF, but it is an ensemble method that trains trees individually rather than in parallel. Meanwhile, clustering involves grouping customers with similar load distributions to facilitate better classification. Their results showed that RF and GB performed better than DT, with an average accuracy of 84.0%, 85.7%, and 80.1% respectively without clustering, and 88.1%, 88.6%, and 87.0% respectively with clustering. Moreover, their results also show that in ideal scenarios, RF and GB had similar accuracies for both cases with and without clustering. From the results, it can be said that clustering improves the accuracy of all three, and all three tree-based classification methods are viable options. While RF and GB have higher accuracies, these two methods are also more complex and take longer to train. Thus, these aspects must also be considered in choosing the most appropriate method.

2.3.1 Electricity Theft Detection in Networks with PV and NM

NM is an incentive mechanism that allows households and commercial establishments to generate their own electricity through rooftop PV in order to help them reduce their electricity bills. This program also enables users to sell excess electricity generated to the distribution utility, giving users the opportunity to generate profit [37]. Figure 2.3 presents a diagram illustrating how NM works.

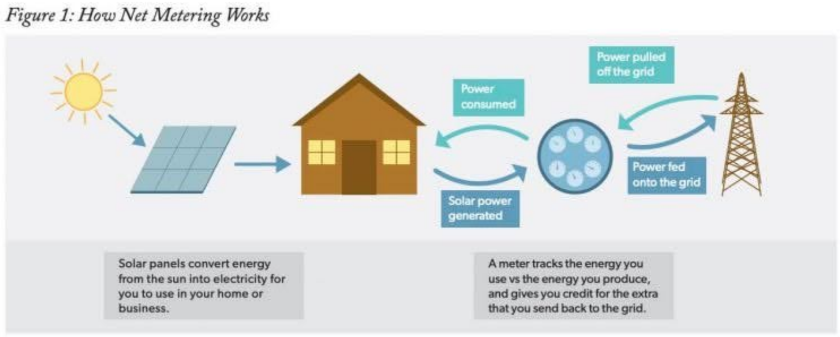


Figure 1: How Net Metering Works
Power consumed
Power pulled off the grid
Solar power generated
Power fed onto the grid
Solar panels convert energy from the sun into electricity for you to use in your home or business.
A meter tracks the energy you use vs the energy you produce, and gives you credit for the extra that you send back to the grid.

Figure 2.3: How Net Metering Works [37]

Installation of this technology introduces new challenges on the distribution network. One challenge would be the bi-directional flow of energy introduced by NM. The added reverse power flow can increase line losses [36]. Additionally, NM provides a new method for electricity theft. Pilferers can now record higher generated electricity than their actual generation that would be sold back to the grid to gain more profit.

These two challenges pose possible issues on current electricity theft detection methods. Because of this, more studies on electricity theft detection for households with both PV and NM are needed to account for the new emerging technology implemented in distribution networks.

In recent years, a few studies have worked towards filling this research gap. The research by Shaaban et al. [38] applied theft detection algorithms in households with PV and NM. In this study, the theft attacks studied were limited to pilferers who increased their generated electricity to gain more profit. In order to detect theft occurrences, they checked whether the error between the actual energy generated per household as recorded by PV smart meters and the estimated energy generated based on solar irradiance and temperature was significant enough. They used root-mean-square-error (RMSE), a DT model, for their theft detection unit that was compared to other algorithms which are SVM, LSE, and ARIMA. Their DT model was able to achieve 94.15% accuracy and SVM had 81% accuracy, while LSE had the poorest performance among all algorithms. This shows that both SVM and DT are strong candidates for theft detection. This research also provides insights on how PV and NM introduces new ways to steal electricity. However, the model used lacked information about consumer consumption and focused solely on PV generation. Thus, further research on consumer consumption data is needed.

The research by Badr et al. [39] used actual electricity consumption and generation measurements from consumers in Australia that uses both PV and NM for electricity theft detection as input to different algorithms. Their theft detection model analyzed patterns and correlations within the benign dataset. Deviations from the typical behavior are then tagged as malicious or candidates for electricity theft. Two types of theft were considered: consumers that report lower electricity consumption for less fees and consumers that report higher generation for more profit. To detect these, they developed a hybrid CNN and GRU model, then compared this with MLP, CNN, and GRU. Since the dataset used did not have known malicious customers, they generated malicious readings based on the consumer's consumption to test their model.The results of their study showed that their hybrid network has a better performance compared to the other algorithms. MLP, which is a field of ANN, still performed well with 94.53% accuracy, 98.35% precision, 94.35% recall, 6.36% false alarm, and 87.99% highest difference. Through their work, they were able to create an NM dataset which could be used in simulating networks with PV and NM. Their results also show that while their proposed hybrid model has the best performance, other detectors also performed well, with accuracies greater than 90%.

Another work that investigated electricity theft detection methods on distribution networks with both PV and NM is by Ismail et al. [40]. They used smart meter readings to obtain the actual load profile of residential customers in Ontario, Canada, solar irradiance readings (kWh/m2), and nine SCADA metering points that were optimally distributed along the system based on a previous study [41]. These metering points are similar to check meters since they measure voltage, current, and power installed downstream which are affected by the installed PV [42]. For this study, malicious consumers were limited to those who increase their injected energy into the grid to increase profit. Since pilferers do not have access to solar irradiance readings and SCADA meter readings, these readings were compared to the reported generation from individual household smart meters. This relationship was then used as input to a hybrid CNN+GRU+Dense theft detection model and other basic machine learning algorithms. This included SVM which resulted in a detection rate of 88.3%. The paper's proposed model had a high detection rate of 99.3%, but it requires a deeper understanding of how each algorithm in the hybrid network works, potentially causing more difficulty. Moreover, similar to [38], their model only considered theft cases wherein pilferers increased their reported generated energy.

Lavilla et al. [6] used the same NM benign dataset and theft cases as [39], but their study focused on understanding the effect of varying PV and NM penetration in the network on electricity theft detection. To do this, they modeled the IEEE European Low Voltage Test Feeder on Open DSS and added eight check meters throughout the network. These check meters were placed in such a way that the households were evenly distributed among them (Figure 2.4). Then, malicious customers were generated by applying a multiplier to randomly selected households to detect theft. The resulting dataset with both benign and malicious samples was then used to train their model in detecting theft. They obtained the percent loss error between the check meter readings and the sum of the individual household meter readings downstream each check meter, then used this value as input to their SVM and ANN models under varying PV and NM penetration levels to determine whether the difference is significant enough to be considered as theft. Their results show that the performance of both algorithms decline with increasing PV and NM levels. Moreover, varying the number of households with NM resulted in greater variation in the algorithm performance as compared to varying the number of households with PV. Therefore, the proliferation of NM makes detecting electricity theft more difficult.

Building on this information, Maala et al. [7] used a similar framework as [6]. They used the same NM dataset, test feeder model, and check meter topology, but they explored different PV and NM penetration levels as well as different features and algorithms that could be used to fortify electricity theft detection. For the features, they also explored Gamma Deviance, Log Cosh Loss, Poisson Deviance and Squared Error aside from Percent Loss Error. For the algorithms, they used SVM, ANN, K-Nearest Neighbors (KNN), and DT. Their results show that using Gamma Deviance and Log Cosh Loss as features resulted in high performance for SVM, ANN and DT, even with increased PV and NM penetration. Meanwhile, all five features performed poorly in KNN, even with minimal households with PV and NM in the system.

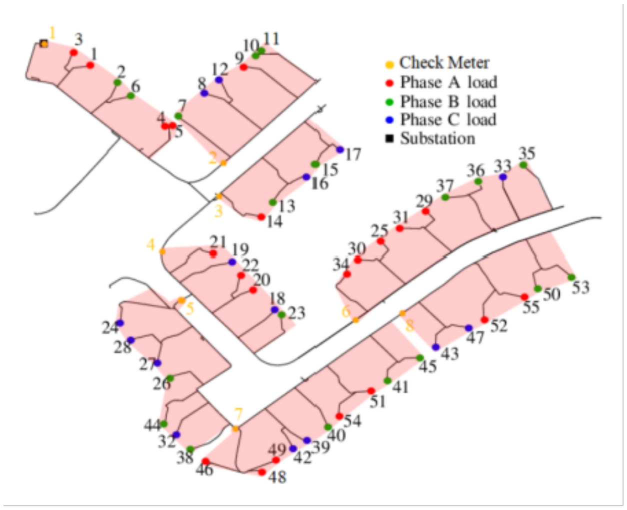


Figure 2.4: Check Meter Topology [6][7]

2.3.2 Role of Check Meters

The studies by Huang et al., Ismail et al., Lavilla et al., and Maala et al. all make use of check meter readings or similar infrastructures in their models. Huang et al. analyzed the performance of ANN on a network with a single check meter, but their study did not use comprehensive performance metrics to evaluate the method used. Additionally, the households in the network do not have PV and NM. On the other hand, Ismail et al., Lavilla et al., and Maala et al. analyzed the performance of their models on a network with PV and NM, but they constantly used 9, 8, and 8 check meters respectively in their simulations. This assumption allowed them to lessen the number of households assigned per check meter, thus simplifying the detection model. These studies no longer explored other check meter quantities which affects the practicality of their models. Given the information provided by the aforementioned work, there remains to be no literature on the effect of check meter quantity on theft detection in distribution networks with PV and NM.

2.4 Summary

Existing works vary in how their system detects theft. While several works have explored the use of machine learning algorithms to detect theft, only a few of them test their system in households with PV and NM. Aside from this, the type of data used and metering units also differ per work. Table 2.2. shows the summary of the different theft detection systems that used different machine learning algorithms.

Table 2.2: RRW Summary

Reference	PV	NM	Data	Metering Unit	Algorithm	Performance %
[24]	N	N	kWh	Household	Linear SVM	80 Accuracy
					RBF SVM	89 Accuracy
					MLP - NN	73 Accuracy
[25]	N	N	kWh CWR	Household	SVM	86.43 Accuracy 77.41 Hit Rate
[27]	N	N	kWh	Household 1 CM	ANN	N/A
[28]	N	N	kWh	Household	ANN	92.44 Accuracy
					DANN	93.04 Accuracy
					DT	92.54 Accuracy
					Adaboost	92.47 Accuracy
[32]	N	N	kWh	Household	Linear SVM	85.52 Accuracy
					ANN	91.46 Accuracy
					RF	91.63 Accuracy
					LR	88.21 Accuracy
[35]	N	N	kWh	Household	DT	80.1 Accuracy (w/o clustering) 87 Accuracy (w/ clustering)
					RF	84 Accuracy (w/o clustering) 88.1 Accuracy (w/ clustering)
					GB	85.7 Accuracy (w/o clustering) 88.6 Accuracy (w/ clustering)
[38]	Y	Y	kWh Irradiance	Household	SVM	81 Accuracy
					DT	94.15 Accuracy
[39]	Y	Y	kWh	Household	ANN	94.53 Accuracy, 98.35 Precision, 94.35 Recall, 6.36 False Alarm 87.99 Highest Difference
[40]	Y	Y	kWh Irradiance	Household 9 SCADA	SVM	88.3 Detection Rate
					Hybrid	99.3 Detection Rate

[6]	Y	Y	kWh	8 CM	SVM	60 to 97 Accuracy
					ANN	64 to 98 Accuracy
[7]	Y	Y	kWh	8 CM	SVM	Different per Feature Accuracy >80% for most cases
					ANN	
					DT	
					KNN	Accuracy <80%
Proposed	Y	Y	kWh	Varying number of CM	SVM ANN DT RF	N/A

From Table 2.2, it can be seen that different algorithms, including SVM, ANN, DT and RF, are able to perform well in theft detection. However, minimal works have analyzed households with PV and NM. Shaaban et al. and Ismail et al. implemented theft detection methods on networks with PV and NM, but they only considered theft occurrences wherein the pilferer reports more energy generated. To address this, Badr et al. created an NM dataset and considered theft occurrences wherein the pilferer could report higher energy generation or lower energy consumption. However their model was memory intensive as it used historical time-series data of individual household meter readings to determine the typical load profile of each consumer and detect any major deviations. The study by Lavilla et al., employed a different approach using the same created dataset by [39]. They used readings from 8 check meters and compared these with the sum of the downstream household meter readings. This study also focused on the effect of varying PV and NM penetration. Their findings demonstrated that NM has a greater influence on the classifier's performance than PV. With this knowledge, Maala et al. used a similar research design, and explored other features and algorithms to improve theft detection.

Based on the works outlined, only a few studies use check meters as their metering unit, both for systems with and without PV and NM. Among these studies, no existing work discusses how the quantity of check meters affects the performance of theft detection systems.

Chapter 3

Problem Statement and Objectives

3.1 Problem Statement

As distribution networks embrace clean energy sources, more consumers are availing rooftop PV and NM. However, one study shows that increasing the number of rooftop PV and NM in the network degrades the performance of electricity theft detection algorithms [6]. As such, these methods must adapt to networks with PV and NM because electricity theft, a key driver of non-technical losses, harms the power industry. It degrades power quality, burdens consumers with extra costs, and poses potential safety risks.

While prior research has analyzed theft detection systems for households with PV and NM [6] [7], their use of impractical models with numerous check meters warrants further analysis. Currently, no studies have explored the effect of check meter quantity on theft detection. This presents a challenge, as check meters add both cost and complexity. Therefore, assessing theft detection methods across varying check meter quantities could help strike a balance between cost-effectiveness and operational efficiency.

3.2 Objectives

The primary objective of this project is to study the effect of check meter quantity in electricity theft detection methods. The specific objectives of the project are:

- to create training datasets with varying check meter quantities
- to compare the performance of electricity theft detection algorithms under different PV and NM penetration and check meter quantity.

3.3 Scope and Limitations

This project will assess the performance of electricity theft detection algorithms on households with rooftop PV and NM. The following machine learning algorithms will be used: Support Vector Machine (SVM), Artificial Neural Network (ANN), Decision Tree (DT), and Random Forest (RF). Electricity theft will be limited to meter-tampering by consumers through manipulating the readings in their smart meters to record lower consumption readings or higher electricity sold. Furthermore, only attacks with one pilferer will be studied.

In creating the power flow datasets, the network will have varying PV and NM penetration levels, and check meter quantities. The placement of the check meters will be distributed as evenly as the network layout permits. The optimal placement of the check meters is outside the scope of this study.

Chapter 4

Methodology and Preliminary Work

4.1 Overview

The methodology is divided into three main stages: power flow dataset creation, theft detection machine learning implementation, and performance assessment. Figure 4.1 outlines the process flowchart for this study.

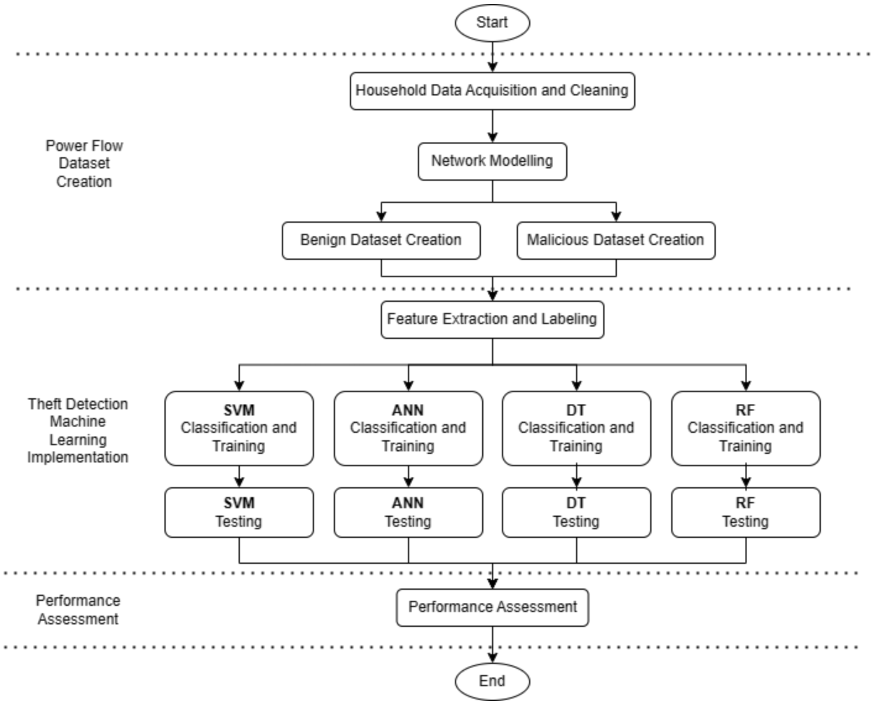

Start
Household Data Acquisition and Cleaning
Power Flow Dataset Creation
Network Modelling
Benign Dataset Creation
Malicious Dataset Creation
Feature Extraction and Labeling
Theft Detection Machine Learning Implementation
SVM Classification and Training
ANN Classification and Training
DT Classification and Training
RF Classification and Training
SVM Testing
ANN Testing
DT Testing
RF Testing
Performance Assessment
Performance Assessment
End

Figure 4.1: Process Flowchart

In the first stage, the raw electricity consumption and generation will be cleaned and processed to create several power flow datasets with both benign and malicious samples for each of the specified check meter quantities. Then, in the second stage, selected features will be extracted from these datasets and fed into the four machine learning algorithms. Finally, in the third stage, the theft detection capabilities of the different algorithms will be evaluated by comparing their performance metrics under the different features and check meter quantities.

4.2 Powerflow Dataset Creation

4.2.1 Household Data Acquisition and Cleaning

The Ausgrid solar home half-hour dataset [43] will be used as the sample data for this study. Ausgrid is the largest electricity distributor in Australia's east coast, with around 1.8 million customers in Sydney, Central Coast, and Hunter Valley. Their solar home half-hour dataset is a publicly available resource that includes meter readings from 300 randomly selected households with rooftop PV in Ausgrid's electricity network. Figure 4.2 illustrates the domain covered by Ausgrid, with the shaded region representing the area spanned by the 300 customers.

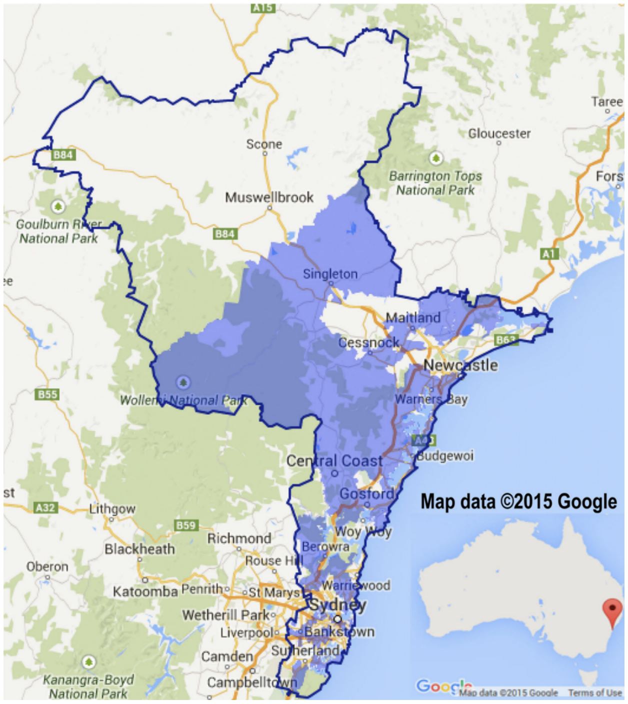


Figure 4.2: Ausgrid Distribution Network Outline [44]

The meter readings fall into three consumption categories: General Consumption (GC), General Generation (GG), and Controllable Load (CL). GC refers to the energy consumption of the customers while GG refers to their residential PV production. Each of the 300 customers in the dataset have GC and GG meter readings. CL, on the other hand, refers to the energy consumed by certain electrical appliances and is only present in 137 customers. Eq. 4.1 shows the power balance equation to obtain the demand given these three consumption data at a certain time index j [44].

$$D(j) = GC(j) - GG(j) + CL(j) \tag{4.1}$$

In order to ensure the dataset to be used is consistent and representative, the following criteria will be used to clean the data:

1. Customers with CL data will be disregarded for simplicity [44][45]. The new power equation can be seen in Eq. 4.2

$$D(j) = GC(j) - GG(j) \tag{4.2}$$

2. Only data from Dec 5-11, 2010 will be used to match the weather forecast in the Philippines. Table 4.1 outlines the weather forecasts in Manila and Sydney during this time period.

Table 4.1: Temperatures in Manila and Sydney [46]

Date	Manila(C)	Sydney(C)
12/5	25-32	20-25
12/6	23-32	20-26
12/7	24-30	21-27
12/8	25-29	21-27
12/9	25-30	22-33
12/10	24-28	21-29
12/11	24-29	17-29

The resulting dataset will then have a total of 54,096 data points. This consists of seven days worth of half-hour meter readings for 161 customers.

4.2.2 Network Modelling

After the sample data is acquired and cleaned, the Ausgrid electricity network will be modeled using the IEEE European Low Voltage Test Feeder. This test feeder is an open-source model created by the IEEE Power and Energy Society Test Feeder Working Group. It has 55 PQ loads connected to a substation through a 11kV/416V step-down distribution transformer as illustrated in Figure 4.3.

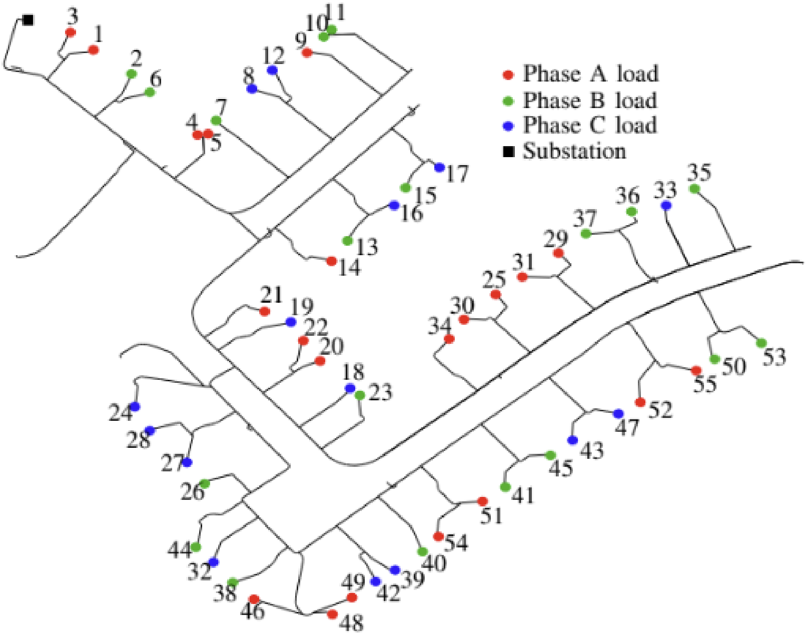


Figure 4.3: IEEE European Low Voltage Test Feeder Topology [45]

The modeling and simulation will be done using OpenDSS, an open-source simulation tool used for electrical utility distribution systems [47]. For the network modeling, different network models with varying check meter quantities and placement will be created. Additionally, the network will have varying levels of PV and NM penetration per simulation. These configurations are further outlined in the next sections.

4.2.3 Check Meter Quantity Configurations

In order to better analyze the effect of check meter quantity, four test network cases will be created with varying check meter configurations. Table 4.2 presents these configurations while Figures 4.4 to 4.7 illustrate the check meter placements within the network.

Table 4.2: Check Meter Configuration

Check Meter Config.	Check Meter Qty	Households per Check Meter
C1	27	2-3
C2	10	5-6
C3	5	10-12
C4	1	55

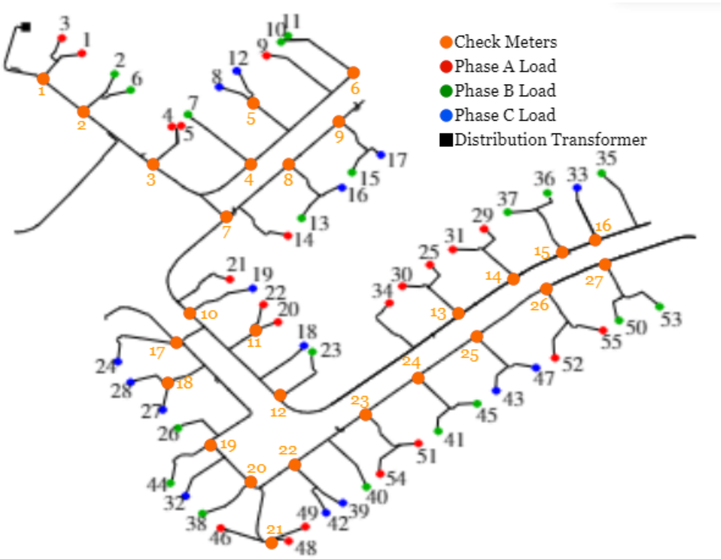


Figure 4.4: C1 Check Meter Placement

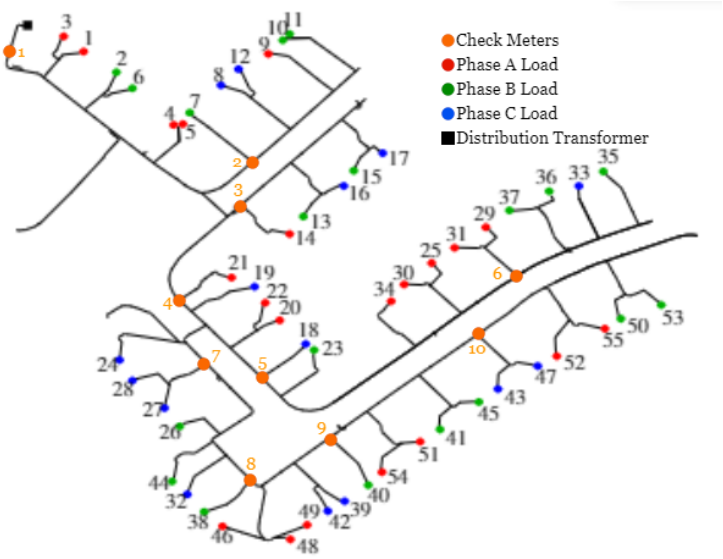


Figure 4.5: C2 Check Meter Placement

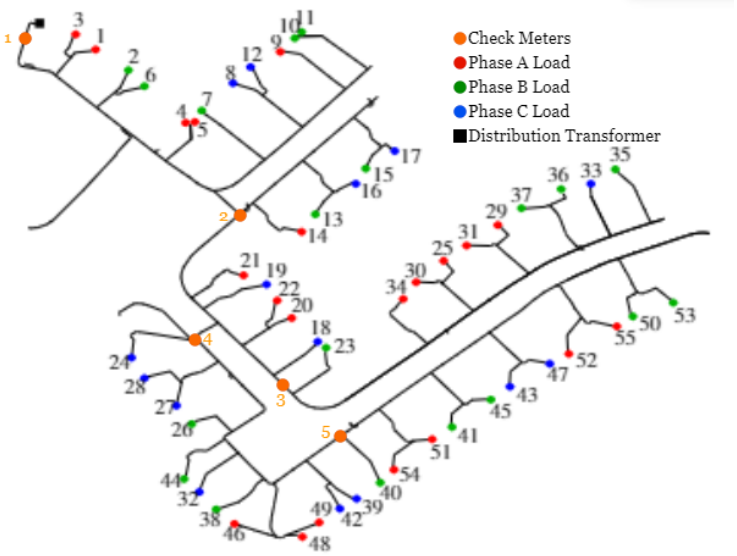


Figure 4.6: C3 Check Meter Placement

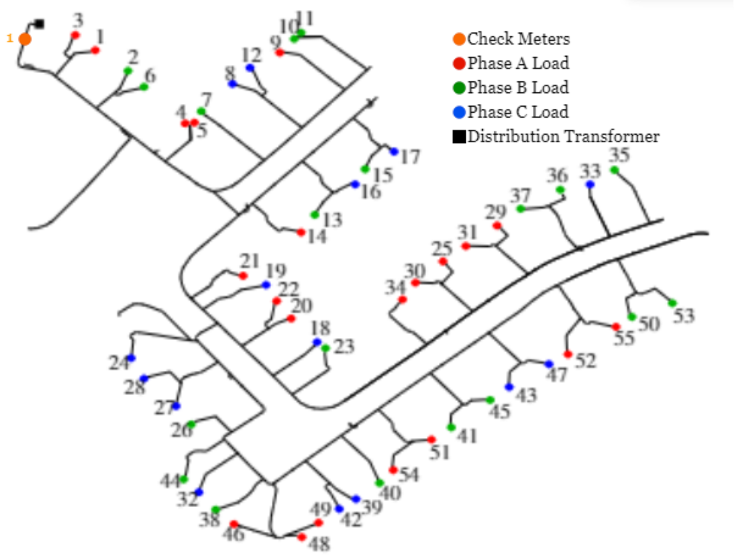


Figure 4.7: C4 Check Meter Placement

Configuration C1 is based on the maximum number of check meters possible in the network. Since the network has 55 households, the maximum number of check meters is 27 with 2-3 households connected per check meter. Having only one household connected to a check meter is impractical as its readings will be almost identical to the household's own meter readings.

On the other hand, C4 is based on the minimum number of check meters in the network. As such it only has a single check meter connected to all 55 households.

For C2 and C3, 10 and 5 check meters will be used respectively. This number is determined based on the layout of the test feeder. As much as possible, the households will be distributed evenly among the different check meters while ensuring that households assigned to a certain check meter are within close proximity of each other. Furthermore, it is more realistic to analyze the effect of minimal check meters because of the cost needed to install check meters. As such, configurations with less check meters were prioritized.

Check meter readings show the measurement of the active power that passes through it. This means that the assumed power consumption of all the downstream households will be shown in the readings. To properly group households according to the check meter assigned to them, the reading of the check meter will be reduced by subtracting the readings of other check meters that are downstream to it. Table 4.3 shows the assigned households in each check meter for the different configurations.

Table 4.3: Household Assignment per Check Meter

Check Meter	**Households**			
	C1	**C2**	**C3**	**C4**
1	1, 3	1, 2, 3, 4, 5, 6	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12	1 to 55
2	2, 6	7, 8, 9, 10, 11, 12	13, 14, 15, 16, 17 18,19, 20, 21, 22	-
3	4, 5	13, 14, 15, 16, 17	23, 25, 29, 30, 31, 33, 34, 35, 36,37	-
4	7, 9	19, 21, 21, 22, 24	24, 26, 27, 28, 32, 38, 39, 42, 44, 46, 48, 49	-
5	8, 12	18, 23, 25, 30, 34	40, 41, 43, 45, 47, 50, 51, 52, 53, 54, 55	-
6	10, 11	29, 31, 33, 35, 36,37	-	-
7	14, 21	26, 27, 28, 32, 44	-	-

8	13, 16	38, 39, 42, 46, 48, 49	-	-
9	15, 17	40, 41, 45, 51, 54	-	-
10	18, 19	43, 47, 50, 52, 53, 55	-	-
11	20, 22	-	-	-
12	23, 34	-	-	-
13	30, 25	-	-	-
14	29, 31	-	-	-
15	36, 37	-	-	-
16	33, 35	-	-	-
17	24, 26	-	-	-
18	27, 28	-	-	-
19	32, 44	-	-	-
20	38, 46	-	-	-
21	48, 49	-	-	-
22	39, 40, 42	-	-	-
23	51, 54	-	-	-
24	41, 45	-	-	-
25	43, 47	-	-	-
26	52, 55	-	-	-
27	50, 53	-	-	-

4.2.4 PV and NM Penetration Configurations

After modeling the networks with different check meter configurations, the load of the households will then be modeled. In order to illustrate the effect of PV and NM on theft detection, seven test network cases with varying PV and NM penetration levels will be created as outlined in Table 4.4.

For configuration P0, the households will be modeled based on their PQ load. While for P1 to P6, the households will be modeled as PQ loads with PV generation, with excess energy flowing back to the grid. Due to this, net consumption values of the households may have negative values.

Table 4.4: PV and NM Configurations

PV + NM Config.	PV Pen. (%)	PV Houses Qty	NM Pen. (%)	NM Houses Qty
P0	0	0	0	0
P1	50	27	33	9
P2	100	55	33	18
P3	50	27	66	18
P4	100	55	66	36
P5	50	27	100	27
P6	100	55	100	55

4.2.5 Benign Dataset Creation

Using the 4 check meter configurations and 7 PV and NM configurations, 28 datasets will be generated by exhausting combinations of the aforementioned configurations. Then, simulations will be performed in OpenDSS with values from the Ausgrid data as input. A compressed version of these datasets is illustrated in Table 4.5.

Table 4.5: Compressed Dataset Configurations

Dataset	Sim #	CM	PV + NM
D1	1-110	C1	P0
...	...	...	...
D7	661-770	C1	P6
...	...	...	...
D28	2971-3080	C4	P6

A total of 3080 simulations will be performed with each dataset having 110 simulations. Per simulation, 55 households will be randomly selected from the 161 consumers in Chapter 4.2.1 and assigned to the 55 loads in the network. Each simulation will have seven days worth of half-hour meter readings for both the households and check meters. However, these readings will be summed and converted into daily readings to facilitate easier analysis. As a result, each simulation will have 7 data points each for the check meter and household meter readings.

4.2.6 Malicious Dataset Creation

In this study, the network will experience meter tampering done by a single pilferer. They could either report a lower energy consumption or a higher energy generation. The theft will be represented by applying a multiplier k to the pilferer's net meter readings. If the pilferer's net consumption is positive, then applying the multiplier will reduce the recorded consumption. Alternatively, if their net consumption is negative, applying the multiplier would increase the recorded generation by further amplifying the negative value. Eqn 4.3 shows the equation used to determine the malicious value for the pilferer's net meter reading (X).

$$X = \begin{cases} X * k & \text{if } x > 0 \\ X - (|X| * k) & \text{if } x < 0 \end{cases} \tag{4.3}$$

The value of k will be obtained randomly from a Gaussian distribution curve with a mean of 0.5 and standard deviation of 0.05. The multiplier will only be selected within three standard deviations away from the mean, so its values will range from 35% to 65%. Two theft frequencies will be considered in this study: full day and half day. For the full day frequency, the pilferer changes their readings for the whole day; for the half day frequency, the pilferer only changes their readings from 6 a.m. to 6 p.m. Given this set-up, each household will be assigned as the pilferer twice per dataset. The first instance will be for the full day frequency and the second for the half day frequency. Table 4.6 presents a compressed version of how these theft frequencies are modeled into the dataset.

Table 4.6: Compressed Dataset Theft Frequencies

Dataset	**Simulation No.**	**Theft Frequency**
D1_mal	1-55	Full Day
	56-110	Half Day
...	...	...
D28_mal	2971 - 3025	Full Day
	3026 - 3080	Half Day

4.3 Theft Detection Machine Learning Implementation

From the created power flow datasets, certain features will be extracted and used as input into the machine learning algorithms. The feature-extracted datasets will then be split into two groups: 80% for training and

20% for testing. This will be done through the GroupShuffleSplit Python library. After splitting the data, the algorithm implementation will be performed on Google Colab using imported Python libraries. A total of 336 classifiers will be used as there will be 4 algorithms, 28 datasets, and 3 features to be used.

4.3.1 Feature Extraction and Labeling

Classification-based electricity theft detection methods typically use extracted features from the power flow dataset to further improve the performance of machine learning algorithms [20]. These features are dependent on the researchers as there is currently no set standard on feature selection [20].

In this study, the check meter readings and the sum of the individual meter readings will be used as input to obtain the following features: Percent Loss Error, Gamma Deviance, and Log Cosh Loss. Equations 4.4 to 4.6 present how these features are obtained. In these equations, CM refers to the daily check meter readings while $\sum_{n=1}^{k} M_n$ refers to the sum of the daily meter readings of individual households under a certain check meter.

$$GammaDeviance = 2(log(\frac{\sum_{n=1}^{k} M_n}{CM}) + \frac{CM}{\sum_{n=1}^{k} M_n} - 1) \tag{4.4}$$

$$Logcosh = log(cosh(\sum_{n=1}^{k} M_n - CM)) \tag{4.5}$$

$$PercentLossError = \frac{|\sum_{n=1}^{k} M_n - CM|}{CM} \tag{4.6}$$

These were the top performing features in a previous study that analyzed the performance electricity theft detection on households with rooftop PV and NM [7]. As such, these features were adopted in this study.

4.3.2 Support Vector Machine (SVM)

SVM is a supervised machine learning algorithm that can be used for classification [23]. In this study, the radial basis kernel function of SVM (SVM-RBF) will be used since the extracted features are non-linear.

Before implementing this algorithm, the cost and gamma parameters must first be tuned and optimized through grid-search and k-cross validation. Specifically, a 10-fold cross-validation will be performed to further split the training sets. Once this is done, SVM can then be implemented in Python using the Support Vector Classifier (SVC) library from Scikit-learn [48]

4.3.3 Artificial Neural Networks (ANN)

ANN is a computational network that maximizes the use of multiple layers of interconnected nodes [26]. This algorithm makes use of several hyperparameters to obtain better results in a shorter span of time. These hyperparameters include the number of neurons and hidden layers, batch size, epoch, optimizers, and activation functions.

Before implementing this algorithm, the hyperparameters must first be determined and optimized. In this study, these hyperparameters to be analyzed are the batch size, number of epochs, optimizers and activation functions. The optimizers that will be considered are the Stochastic Gradient Descent (SGD) and Adaptive Moment Estimation (ADAM) while the activation functions will be the sigmoid, hyperbolic tan, or rectified linear unit. Then, a 5-fold hyperparameter optimization will be performed to select the best optimizer and activation function, as well as the batch size and epoch number. Once selected, the Keras and MinMaxScaler libraries will be imported to implement ANN.

4.3.4 Decision Tree (DT)

DT is another type of supervised learning algorithm which is used for classification problems relating to regression analysis [30]. In implementing this algorithm, it must first be ensured that the training and testing data do not overlap to avoid overfitting. Once achieved, DT will then be implemented in Python through the DecisionTreeRegressor imported library.

4.3.5 Random Forest (RF)

RF has a similar procedure as DT, but it makes use of multiple trees [RF]. The greater the number of trees, the more accurate the result would be. Similar to ANN, it uses hyperparameters to improve the structure. For this study, the number of DTs and max depth will be optimized before proceeding with the algorithm implementation. Once done the RandomForestClassifier library from Scikit-learn will be used.

4.4 Performance Assessment

The ML algorithms will then classify the households as either honest or malicious by labelling them as 0 or 1 respectively. To quantify the performance of the different algorithms at different configurations, their accuracy will be measured. Accuracy tells how close the results of the theft detection model's classified values to the actual values. This metric is dependent on the value of four different variables: true positives

(TP), true negatives (TN), false positives (FP), and false negatives (FN). TP and TN represent the correctly detected theft and non-theft respectively, while FP and FN represent wrongly detected theft and undetected theft respectively. Equation 4.7 shows the formula for this metric.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.7}$$

4.5 Preliminary Work

In preparation for the project implementation, the following key steps were taken: First, the Ausgrid dataset was acquired from [43]. This will serve as the primary data source of the project. Second, the documentation and corresponding OpenDSS file for the European IEEE LV system was accessed from [47]. This ensures that the simulations can be done on the chosen test feeder. Finally, the libraries needed to implement the different algorithms were obtained. These libraries will be utilized in the project to process and analyze the data collected from the Ausgrid dataset.

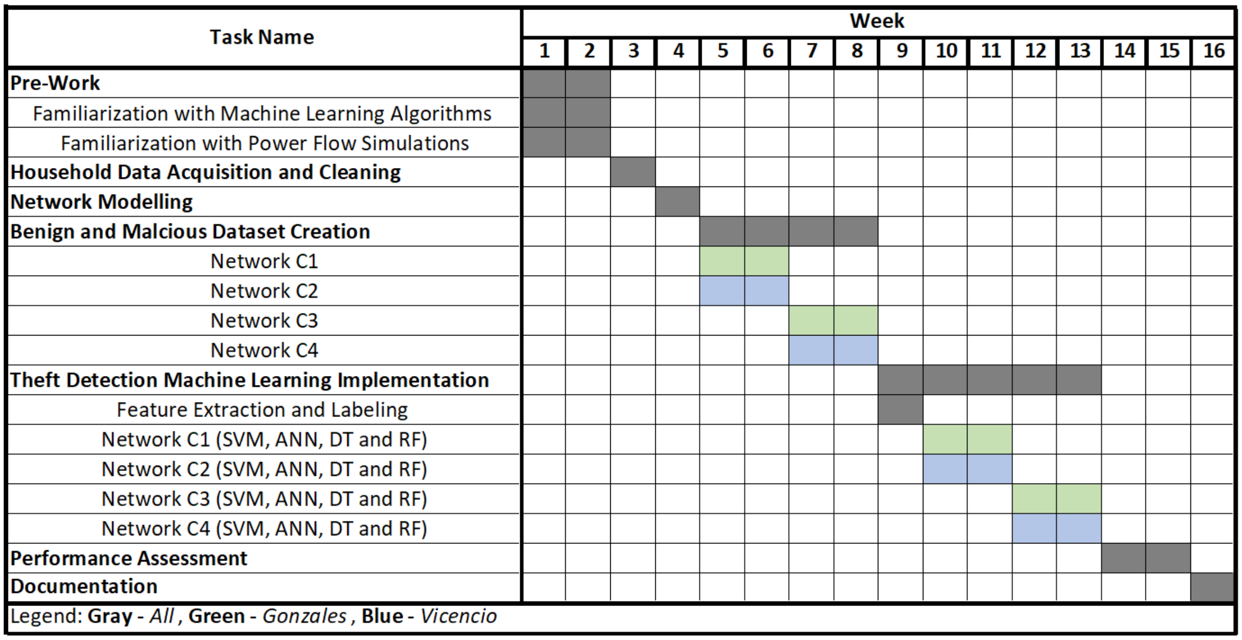
Aside from these, the proponents were also granted access to the file repository used in [7], which is the basis for this study's theoretical framework and methodology. This includes the OpenDSS file for their test network and the scripts used for data cleaning and machine learning implementation. These files have been replicated on the proponent's own machines for familiarization, and to serve as a guide in the implementation stage.

Chapter 5

Project Schedule and Deliverables

5.1 Proposed Timeline

This study will be executed for a total of 16 weeks, in line with the proposed academic calendar for the second semester of AY 2023-2024. Figure 5.1 presents the Gantt chart outlining the tasks per week and the proponent assigned to each task.



Task Name	Week															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Pre-Work	▨	▨														
Familiarization with Machine Learning Algorithms	▨	▨														
Familiarization with Power Flow Simulations	▨	▨														
Household Data Acquisition and Cleaning			▨													
Network Modelling				▨												
Benign and Malcious Dataset Creation					▨	▨	▨	▨								
Network C1					🟩	🟩										
Network C2					🟦	🟦										
Network C3							🟩	🟩								
Network C4							🟦	🟦								
Theft Detection Machine Learning Implementation									▨	▨	▨	▨	▨			
Feature Extraction and Labeling									▨							
Network C1 (SVM, ANN, DT and RF)										🟩	🟩					
Network C2 (SVM, ANN, DT and RF)										🟦	🟦					
Network C3 (SVM, ANN, DT and RF)												🟩	🟩			
Network C4 (SVM, ANN, DT and RF)												🟦	🟦			
Performance Assessment														▨	▨	
Documentation																▨

Legend: **Gray** - *All* , **Green** - *Gonzales* , **Blue** - *Vicencio*

Figure 5.1: Gantt chart

5.2 Halfway Deliverables

After the 8th week, the following deliverables are expected to be submitted by the proponents:

1. List of libraries for machine learning implementation
2. File repository
    - Python scripts for data cleaning
    - OpenDSS file of test network
    - Raw and processed power flow datasets
3. Documentation of tasks

5.3 Final Deliverables

At the end of the execution period, these are the final deliverables expected from the proponents:

1. Machine learning prediction results
2. Theft detection performance metrics
3. File repository
    - Python scripts for data cleaning
    - OpenDSS file of test network
    - Raw and processed power flow datasets
    - Python scripts for machine learning implementation
4. Manuscript
5. IEEE format article

Bibliography

[1] Department Of Energy. in *Philippine Energy Plan 2020-2040*, page 66, June 2022.

[2] A. Lafrades and F. Arriola, in *Guidebook on Net Metering in the Philippines*, Department of Energy Philippines, 2021, pp. 33-35

[3] M. -M. Buzau, J. Tejedor-Aguilera, P. Cruz-Romero and A. Gomez-Exposito, ”Hybrid Deep Neural Networks for Detection of Non-Technical Losses in Electricity Smart Meters,” in *IEEE Transactions on Power Systems*, vol. 35, no. 2, pp. 1254-1263, March 2020, doi: 10.1109/TPWRS.2019.2943115.

[4] L. G. Arango, E. Deccache, B. D. Bonatto, H. Arango, P. F. Ribeiro and P. M. Silveira, ”Impact of electricity theft on power quality,” *2016 17th International Conference on Harmonics and Quality of Power (ICHQP)*, Belo Horizonte, Brazil, 2016, pp. 557-562, doi: 10.1109/ICHQP.2016.7783346.

[5] I. Petrlik et al., ”Electricity Theft Detection using Machine Learning,”*International Journal of Advanced Computer Science and Applications*, vol. 13, no. 12, pp. 420-425, 2022.

[6] Carl Lavilla, Michael Osorio, and Zildjian Restituto. Effect of Net Metering and Rooftop Photovoltaics on Electricity Theft Detection. In *Effect of Net Metering and Rooftop Photovoltaics on Electricity Theft Detection*, pages 1-64. University of the Philippines, 2021

[7] R. M. P. Maala, A. M. B. Rebamba and A. E. D. Tio, ”Classification-Based Electricity Theft Detection on Households with Photovoltaic Generation and Net Metering,” *TENCON 2023 - 2023 IEEE Region 10 Conference (TENCON)*, Chiang Mai, Thailand, 2023, pp. 1094-1099, doi: 10.1109/TENCON58879.2023.10322383.

[8] E.S. Bueno, ”Distribution System Loss Segregation,” official memorandum, National Electrification Administration, Quezon City, Philippines, 2009. Available: https://www.nea.gov.ph/ao39/phocadownload/MEMO%20TO

%20ECs/2009//NEA%20Memo%20to%20EC%20No.%202009-002%20-%20Distribution%20System%20Loss%20Segregation.pdf

[9] D. Rivera, ”ERC keeps system loss cap for private DUs,” *Philstar*, Manila, Jan. 10, 2022

[10] J. R. C. Orillaza, R. d. Del Mundo and J. A. C. Miras, ”Development of Models and Methodology for the Segregation of Distribution System Losses for Regulation,” *TENCON 2006 - 2006 IEEE Region 10 Conference*, Hong Kong, China, 2006, pp. 1-4, doi: 10.1109/TENCON.2006.343811.

[11] S. Dhungel, ”Introduction on distribution system losses,” Engineering Sarokar, https://engineeringsarokar.com/introduction-on-distribution-system-losses/ (accessed Jan. 7, 2024).

[12] National Electrification Administration. in *The Project on System Loss Reduction for Philippine Electric Cooperatives (ECs),* pages 7-8, March 2013.

[13] Pamir, N. Javaid , M.U. Javed, M.A. Houran, A.M. Almasoud, M. Imran. ”Electricity theft detection for energy optimization using deep learning models,” *Energy Sci Eng*. 2023, pp. 3575-3596. doi:10.1002/ese3.1541

[14] A. J. Rajab , ”Electricity Theft and Energy Fraud Detection,” thesis, 2023

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

[16] M. A. de Souza et al., ”Detection and identification of energy theft in advanced metering infrastructures,” *Electric Power Systems Research*, vol. 182, p. 106258, 2020. doi:10.1016/j.epsr.2020.106258

[17] S. McLaughlin, D. Podkuiko, and P. McDaniel, ”Energy theft in the advanced metering infrastructure” [Online]. Available: http://www .patrickmcdaniel.org/pubs/critis09.pdf

[18] S. Jiyane-Tshikomba, ”Technical analysis mitigation of electricity theft for domestic and commercial end users,” thesis, 2019

[19] C. J. Bandim et al., ”Identification of energy theft and tampered meters using a central observer meter: a mathematical approach,” *2003 IEEE PES Transmission and Distribution Conference and Exposition (IEEE Cat. No.03CH37495)*, Dallas, TX, USA, 2003, pp. 163-168 Vol.1, doi: 10.1109/TDC.2003.1335175.

[20] G. M. Messinis and N. D. Hatziargyriou, ”Review of non-technical loss detection methods,” *Electric Power Systems Research*, vol. 158, pp. 250-266, 2018. doi:10.1016/j.epsr.2018.01.005

[21] E. U. Haq, C. Pei, R. Zhang, H. Jianjun, and F. Ahmad, ”Electricity-theft detection for smart grid security using smart meter data: A deep-CNN based approach,” *Energy Reports*, vol. 9, pp. 634-643, 2023. doi:10.1016/j.egyr.2022.11.072

[22] M. Ahmed et al., ”Energy Theft Detection in Smart Grids: Taxonomy, Comparative Analysis, Challenges, and Future Research Directions,” in *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 4, pp. 578-600, April 2022, doi: 10.1109/JAS.2022.105404.

[23] ”Support Vector Machine (SVM) algorithm ,” JavaTPoint,https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm. (accessed Jul. 28, 2023).

[24] G. Figueroa, Y. -S. Chen, N. Avila and C. -C. Chu, ”Improved practices in machine learning algorithms for NTL detection with imbalanced data,” *2017 IEEE Power & Energy Society General Meeting*, Chicago, IL, USA, 2017, pp. 1-5, doi: 10.1109/PESGM.2017.8273852.

[25] J. Nagi, A. M. Mohammad, K. S. Yap, S. K. Tiong and S. K. Ahmed, ”Non-Technical Loss analysis for detection of electricity theft using support vector machines,” *2008 IEEE 2nd International Power and Energy Conference*, Johor Bahru, Malaysia, 2008, pp. 907-912, doi: 10.1109/PECON.2008.4762604.

[26] ”Artificial Neural Network (ANN) Tutorial,” JavaTPoint,https://www.javatpoint.com/artificial-neural-network, (accessed Jul. 28, 2023)

[27] 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)*, Sarajevo, Bosnia and Herzegovina, 2018, pp. 1-6, doi: 10.1109/ISGTEurope.2018.8571877.

[28] F. A. Bohani et al., ”A comprehensive analysis of supervised learning techniques for electricity theft detection,” *Journal of Electrical and Computer Engineering*, vol. 2021, 2021. doi:10.1155/2021/9136206

[29] A. R. N. Aouichaoui, R. Al, J. Abildskov, and G. Sin, ”Comparison of group-contribution and machine learning-based property prediction models with uncertainty quantification,” *31st European Symposium on Computer Aided Process Engineering*, pp. 755-760, 2021. doi:10.1016/b978-0-323-88506-5.50118-2

[30] ”Decision Tree (DT) Classification Algorithm,” JavaTPoint,https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm, (accessed Jul. 28, 2023)

[31] ”AdaBoost algorithm in Machine Learning,” AlmaBetter, https://www.almabetter.com/bytes/tutorials/data-science/adaboost-algorithm (accessed Jan. 7, 2024).

[32] J. Pereira and F. Saraiva, ”A Comparative Analysis of Unbalanced Data Handling Techniques for Machine Learning Algorithms to Electricity Theft Detection,” *2020 IEEE Congress on Evolutionary Computation (CEC)*, Glasgow, UK, 2020, pp. 1-8, doi: 10.1109/CEC48606.2020.9185822.

[33] ”Random Forest (RF) Algorithm,” JavaTPoint, https://www.javatpoint.com/machine-learning-random-forest-algorithm, (accessed Jul. 28, 2023)

[34] V. A. Kanade, ”Logistic regression: Equation, assumptions, types, and best practices,” Spiceworks, https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-logistic-regression/#::̃text=Logistic%20regression%20is%20a%20supervised%20machine%20learning %20algorithm%20that%20accomplishes,1%2C%20or%20true%2Ffalse (accessed Jan. 7, 2024).

[35] S. O. Tehrani, M. H. Y. Moghaddam and M. Asadi, ”Decision Tree based Electricity Theft Detection in Smart Grid,” *2020 4th International Conference on Smart City, Internet of Things and Applications (SCIOT)*, Mashhad, Iran, 2020, pp. 46-51, doi: 10.1109/SCIOT50840.2020.9250194.

[36] R. A. Walling, R. Saint, R. C. Dugan, J. Burke and L. A. Kojovic, ”Summary of Distributed Resources Impact on Power Delivery Systems,” in *IEEE Transactions on Power Delivery*, vol. 23, no. 3, pp. 1636-1644, July 2008, doi: 10.1109/TPWRD.2007.909115.

[37] ”What is Net Metering?,” NSci Technologies, https://nsci.ca/2020/08/31/what-is-net-metering/ (accessed Jan. 8, 2024).

[38] M. Shaaban, U. Tariq, M. Ismail, N. A. Almadani and M. Mokhtar, ”Data-Driven Detection of Electricity Theft Cyberattacks in PV Generation,” in *IEEE Systems Journal*, vol. 16, no. 2, pp. 3349-3359, June 2022, doi: 10.1109/JSYST.2021.3103272.

[39] M. M. Badr, M. I. Ibrahem, M. Baza, M. Mahmoud and W. Alasmary, ”Detecting Electricity Fraud in the Net-Metering System Using Deep Learning,” *2021 International Symposium on Networks, Computers and Communications (ISNCC)*, Dubai, United Arab Emirates, 2021, pp. 1-6, doi: 10.1109/ISNCC52172.2021.9615628.

[40] M. Ismail, M. F. Shaaban, M. Naidu and E. Serpedin, ”Deep Learning Detection of Electricity Theft

Cyber-Attacks in Renewable Distributed Generation," in *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 3428-3437, July 2020, doi: 10.1109/TSG.2020.2973681.

[41] A. Rehman, "SCADA and its application in Electrical Power Systems," AllumiaX Engineering, https://www.allumiax.com/blog/scada-and-its-application-in-electrical-power-systems (accessed Jan. 7, 2024).

[42] M. F. Shaaban, A. H. Osman and F. M. Aseeri, "A Multi-Objective Allocation Approach for Power Quality Monitoring Devices," in *IEEE Access*, vol. 7, pp. 40866-40877, 2019, doi: 10.1109/ACCESS.2019.2906269.

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

[44] E. L. Ratnam, S. R. Weller, C. M. Kellett, and A. T. Murray, "Residential load and rooftop PV GENERATION: An australian distribution network dataset," *International Journal of Sustainable Energy*, vol. 36, no. 8, pp. 787-806, 2015. doi:10.1080/14786451.2015.1100196

[45] P. Arboleya, "State Estimation in Low Voltage Networks Using Smart Meters: Statistical Analysis of the Errors," 2018 IEEE Power & Energy Society General Meeting (PESGM), Portland, OR, USA, 2018, pp. 1-5, doi: 10.1109/PESGM.2018.8585999.

[46] "The weather year round anywhere on Earth," Weather Spark, https://weatherspark.com/ (accessed Jan. 7, 2024).

[47] IEEE. IEEE PES Test Feeder. https://cmte.ieee.org/pes-testfeeders/resources/ accessed Jan. 7, 2024).

[48] "scikit learn. sklearn.svm.SVC. https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html."