Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Dept. of Information Security and Communication Technology

Doney Abraham

# Application of Machine Learning in IoT enabled Smart Grids for Attack Detection

Master's thesis in Information Security

Supervisor: Sule Yildirim Yayilgan, Alemayehu Gebremedhin, Fisnik Dalipi, Ogerta Elezaj, Mohamed Abomhara

June 2020

NTNU
Norwegian University of Science and Technology

Doney Abraham

# Application of Machine Learning in IoT enabled Smart Grids for Attack Detection

Master's thesis in Information Security
Supervisor: Sule Yildirim Yayilgan, Alemayehu Gebremedhin, Fisnik Dalipi, Ogerta Elezaj, Mohamed Abomhara
June 2020

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Dept. of Information Security and Communication Technology

NTNU
Norwegian University of
Science and Technology

# Preface

This master's thesis is a research conducted in the department of Information Security and Communication Technology at NTNU Gjøvik during the spring semester of 2020. The research is connected to the CINELDI project and focuses on abnormalities and attacks in IoT enabled Smart Grids as opposed to traditional grids. The content of this is aimed at readers with interest in Information security in general, Smart Grids, IoT and how machine learning can be used to detect an attack on these. Additional background is detailed in the respective sections for those that require more than basic knowledge.

01-06-2020

# Acknowledgment

First and foremost, I would like to thank my family for the support given during the entire period without which it would not have been possible to complete the thesis. I would like to express my sincere gratutiude to my supervisor Associate Prof. Şule Yildirim Yayilgan for all the guidance and discussions we have had throughout the entire period. Further I wish to thank my co-supervisors Prof. Alemayehu Gebremedhin and Associate Prof.Fisnik Dalipi, Dr. Ogerta Elezaj, Dr. Mohamed Abomhara and Dr. Athar Khodabakhsh for the valuable guidance in formulating the research ideas, discussions, data collection, reviews and help with analysing the datasets. The regular meetings that was held, critical reviews and receiving good feedback helped me to be in right path and attain the desired goals.

D.A.

# Abstract

Smart Grids have been increasingly used these days in terms of critical infrastructure when considered in a large scale and in other small areas of operations. This coupled with IoT has huge potentials in several areas like remotely monitoring and managing electricity, traffic signs, traffic congestion, parking spaces, road warnings and even early detection of things like power influxes as the result of natural disasters like earthquakes and extreme weather. Besides the advantages as mentioned, there are increase in security and privacy threats such as unauthorized access, disclosure of information, Denial of Service (Dos) attacks, among others. The smart grids could be prone to security attacks which can cause the entire infrastructure to be offline (DoS) and can cause severe damages to service provider (e.g., lost of money) and individuals (e.g., compromising of privacy).

This thesis discusses about the differences between a traditional grid and a smart grid with IoT enabled devices and how using a smart grid can help in saving money by lowering the operational cost, managing energy consumption and real time troubleshooting. It also focuses on various attack and anomaly detection methods using machine learning to detect the attacks and anomalies in IoT enabled smart grids. Discussions on how the behaviour of a smart grid changes when IoT or smart devices are connected to it have been conducted as part of the study. The thesis also investigates how privacy is affected with the introduction of IoT in smart grids.

# Sammendrag

Smarte strømnettsløsninger har blitt anvendt i større grad i nyere tid for kritisk infrastruktur med tanke på testing og utprøving i både stor og liten skala. Slike strømnettsløsninger kombinert med IoT har potensiale til å fungere sammen innenfor ulike områder, som eksempelvis ekstern overvåking og styring av strøm, trafikal teknologi, parkeringsbransjen og til og med tidlig deteksjon av strømtilstrømninger forårsaket av naturkatastrofer. De ovenfornevnte eksemplene understreker flere fordeler med slik teknologi, men utover dette finnes det eksempler på sikkerhetstrusler og diverse farer for å ikke ivareta ulike personvernhensyn, avsløre hemmelig informasjon osv. Videre kan smarte strømnettsløsninger være utsatt for dataangrep som kan føre til at hele infrastrukturen svikter, skader på tjenesteleverandører (feks økonomiske tap) og individer (personvern).

Denne masteroppgave tar for seg forskjellene mellom tradisjonell strømnetteknologi og smart strømnetteknologi med IoT-aktiverte enheter samt hvordan bruk av et smart rutenett kan fasilitere kostnadsbesparelser ved å minske operasjonelle kostnader, styre energiforbruk og problemløsning i sanntid. For det første omhandler oppgaven også forskjellige måter slik teknologi er sårbar for ulike angrep samt anomalideteksjonsmetoder ved bruk av maskinlæring for å oppdage angrepene og anomaliene for IoT-aktiverte smarte strømnett. For det andre fokuserer oppgaven på hvordan en smart strømnettslønsing påvirkes når IoT og smarte enheter er tilkoblet nettet. For det siste tar avhandlingen for seg hvordan personvern påvirkes ved bruk av IoT i smarte strømnett.

# Contents

# Figures

# Tables

# Acronyms

**AMI** Advanced Metering Infrastructure

**BAN** Building Area Network

**DMS** Distribution Management Systems

**DoS** Denial of Service

**EMS** Energy Management Systems

**HAN** Home Area Network

**IAN** Industrial Area Network

**IoT** Internet of Things

# Chapter 1

# Introduction

The first chapter of the thesis includes a brief introduction of the problem of detecting attacks in IoT enabled smart grids. The introduction is followed by the justification, motivation and benefits of the research, and the research questions the thesis will answer. The chapter concludes with the contribution and outline of the thesis. The below sections from the project are maintained with few changes to the research questions [1].

## 1.1 Topics covered by the Thesis

## 1.2 Keywords

Keywords covered for this thesis are as follows: Smart Grid, IoT, Machine learning, Privacy, Attacks.

## 1.3 Problem description

Smart Grids has addressed many issues of a traditional power grid system by addressing the issues in a traditional power system. It introduces bidirectional flow of energy and information between consumers and providers which has been unidirectional in the case of traditional grids. Smart Grids when integrated with IoT [2] devices can be used to monitor and analyze power consumption in SCADA systems, at the premises of the consumer, distribution centers among others. IoT enabled Smart Grid is called a cyber physical system that is a combination of complex physical systems and cyber systems thereby also introducing various potential issues and challenges [3]. The different types of attacks to a traditional power grid can be determined in advance in most of the cases due to its less robust nature making it easier to mitigate the risks in case of such threats [4]. However, with the introduction of IoT in Smart Grids makes such threat detection quite complex [5]. The behavior of smart grids also can change according to the type of IoT devices that are used like sensors used for monitoring temperature, pressure

or tracking devices to name a few. Assessing the threats and preventing security and privacy attacks associated with the IoT in smart grids have been a challenge. In addition, if one of the attacks causes a potential downtime, several problems might arise depending on where the Smart Grid is implemented. Analyzing data in real time is important which might expose confidential data to unauthorized parties and cause privacy breaches as well. Introduction of IoT in Smart Grid can have privacy threats like collection of data by smart meters that can contain consumer information[6]. The focus of this thesis is to use various machine learning techniques to detect attacks on IoT enabled smart grids and investigate potential privacy breaches of using IoT devices in smart grids.

## 1.4 Justification, motivation and benefits

Preventing threats on IoT enabled Smart Grids is the basis of smooth operation of such kind of infrastructure, be it small or large. Some of the predominant motives for such threats are for example are financial motives, criminal motives or even political motives. In a financial motive, a customer might tamper with the smart meters to reduce the electricity bill [7, 8]. Criminal motives include thieves who want to rob a house gathering information about the inhabitant of the house by monitoring communication between smart meter and the company [9]. Political motives include a hostile nation engaging in cyberwar against a neighboring country by accessing the smart grid facilities that might result in financial losses or blackouts [10]. These are some of the threats that can be prevented by mitigating the attacks against IoT enabled smart grids. However, each attack might be different and can affect various components of the IoT enabled smart grid. Learning these different attacks in advance and hence detecting them in an automated way when an attack occurs can be achieved with using Machine Learning. Identifying various privacy threats that can happen to all the parties(customers and service providers) involved can help in preventing these attacks as well as trust the use of a smart grid. Literature reviews on how smart grids can solve traditional grid problems and how attacks can happen on a smart grid as compared to a traditional grid is detailed in Section 2.1.

## 1.5 Research questions

To be able to solve the research problem, the following research questions needs to be answered:

1. How machine learning models can detect attacks in smart grids embedded with IoT?
    a. How does the behavior change in a Smart Grid as opposed to a traditional grid and if IoT is connected to a smart grid, how behavior of Smart grid changes? So, is anything connected to the system a threat?

2. What are consequences if attack not detected?
    a. How is privacy breached by using IoT in smart grids?

## 1.6 Scope and Contributions

The thesis has several goals as mentioned in the research questions. It aims to investigate an existing smart grid system and analyze what is considered as normal in terms of operations and define thresholds for example consumption is above normal level or unusually low from existing data. Based on the thresholds among others, a machine learning framework is modelled such that it will predict new attacks and known attacks to the system. The selection of algorithms used in the machine learning models is based on previous research [11]. A survey of existing types of attacks to energy systems is also be done as a part of this thesis. It also conducts a survey on the capabilities of a smart grid and present the differences to a traditional power grid. A generic software and hardware architecture for a smart grid is defined as a part of this study. This knowledge can be used to identify the different methods of integration of IoT in smart grids and the ways of communication of smart grids with such IoT devices. A perspective of IoT in Smart Grids from security and privacy point is also investigated as part of the thesis.

## 1.7 Thesis outline

The structure of the thesis and its individual chapter outline is provided in this section.

- **Background**: Gives an introduction to Smart Grids, IoT and its impact areas. It also introduces machine learning, the different classification methods and the need to feature selection leading to the evaluation metrices used to evaluate the different algorithms.
- **Related Work**: Provides the IoT and privacy aspects of Smart Grids. It shows an overview of the different anomaly types and detection algorithms that can be used and eventually lists the attacks that can happen in Smart Grids.
- **Methodology**: This chapter shows the methodology used in the thesis such as the different datasets that is used, identifies the scenarios for experiments and the algorithms that can be used.
- **Experiment and Results**: Describes the various setup used for conducting the experiments, methods for feature selection, how the models are evaluated and data is explored and shows the results.
- **Discussion**: Analyses the experiment and results obtained in the previous Chapter 5. The thesis then maps it to the research questions and shows how the experiments can be used to answer these questions. It also shows some potential limitations faced during the thesis.
- **Conclusion**: Provides a conclusion based on the discussion section.

- **Future Work**: Identifies the potential areas of research and improvement and paves way to further research areas and work.

# Chapter 2

# Background

## 2.1 Overview of Smart Grid: Architecture

This chapter focuses on the overview of Smart grid and its building blocks. It describes the architecture and shows the differences against a traditional power grid and its advantages over it. This section has been studied as a part of my term paper work in [12] and also extended by a paper where I am the lead author [13] that is under review. A traditional power grid is one of the most complex critical infrastructures that has been ever build. It consists of different parts like operations center, power generation plants, transmission towers and power distribution centers that are physically connected by cables and wires. The main functions of a power grid are electricity generation, transmission and its distribution [4]. Electricity is mostly generated using central power plants using different energy sources and then transmitted to different load customers through high voltage lines which in turn is distributed to consumers using distribution centers at a lower voltage. The transmission and distribution is owned by power companies. The electricity and information flow in a traditional power grid are unidirectional which makes it less robust to access the departure of power and transmission of electric energy. Figure 2.1 shows the unidirectional flow in a traditional power grid.

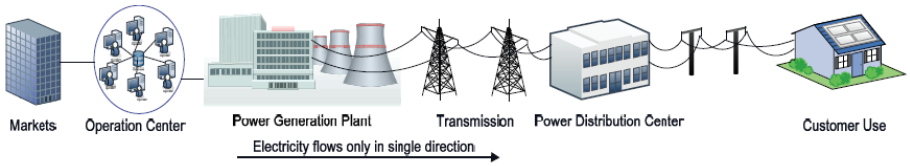



**Figure 2.1:** Traditional power grid [3]

The above mentioned properties of the traditional grid results in lack of flexibility, lack of information sharing to customers and control mechanisms to respond quickly. These traditional grids also lack self-healing and self-restoring capability in case of a down time. Additionally, due to the high usage of electronic devices, these power grids have a large amount of wastage of resources due to inefficient

distribution of electricity, lack of monitoring and communication and inadequate methods to store energy. All these coupled together has led to the introduction of Smart Grids.

Smart Grids enable the integration of both cyber and physical systems in the sense that ICT is integrated with power networks to enable generation, transmission and distribution of electricity in a more effective and efficient manner [3]. Some of the other reasons for using smart grids are due to increase in pressure from global resources for higher quality and reliability. It is also expected that the future grids are more renewable, robust, distributed, interactive with faster protection, control and quality. Following are some of the features of a smart grid [14]:

- It should integrate modern advanced sensor technology, measurement technology, communication technology, information technology, computing technology, and control technology.
- The information and electricity flow should be bidirectional.
- Enable active participation by customers.
- Accommodate all generation and storage options.
- Enable new products, services, and markets.
- Provide power quality for the digital economy.
- Optimize asset utilization and operate efficiently.
- Anticipate and respond to system disturbances.
- Operate resiliently against attacks and natural disasters.

### 2.1.1 The NIST Conceptual model of Smart Grid

Figure 2.2 shows the basic architecture of a smart grid based on the NIST conceptual model.

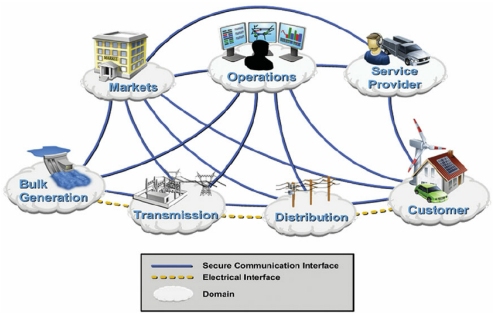



**Figure 2.2:** NIST conceptual model [15]

The communication among the basic system and subsystem for the above mentioned conceptual model according to Yu Cunjiang [16] is as follows:

1. **Bulk generation**: This is the place where electricity is generated in large quantities. These are normally linked straight to transmission systems that in turn offer applications that are smart in nature.

2. **Transmission**: This moves electricity produced in the sites of bulk generation to long distances to the substations that are closer to areas where electricity demand is higher.
3. **Distribution**: This is the final stage of delivery for electricity to reach the consumers.
4. **Customer**: The entire grid is created to support the customer domain. This is where the electricity is consumed. This is usually categorized into home, commercial/building and industrial with varying energy needs for each.
5. **Service Provider**: This provides services to business processes of different power system producers, customers and distributors.
6. **Operations**: This is responsible for continuous operations of the system and includes the network control centers for Energy Management Systems (EMS) and Distribution Management Systems (DMS).
7. **Markets**: This is place where power grid assets are traded. The supply/demand and prices are exchanged in this domain.

The publication according to Yu Cunjiang [16] further categorizes the characteristics of smart grids as following.

- It is robust and can deliver power without interruptions.
- It has self healing capabilities such that it can monitor its state in real time and, analyse any faults that may happen and also restore itself in-case of an incidents.
- It can be easily integrated with a unifying platform and share information with guarantees to managing the grid.
- Its interoperability feature enables logical grouping of standards among various components in the Smart Grid.
- It helps reduce the cost of operations and investments by efficiently managing loss of power and improving the utilization of power efficiency.
- It can also efficiently manage the users by monitoring their interactions and functions they use most.

### 2.1.2 Software Architecture of Smart Grid

In terms of software architecture, Figure 2.3 shows the embedded software view of a smart grid architecture as proposed in the bibliographical survey of software architectures for smart grids [17].

The software architecture shows that smart grid features are realized by software algorithms that interfaces with the grid sensors and actuators. The device drivers enabled with operating system ensures a real time control and operations of the smart grid system. The components are mainly placed in three layers namely physical (device) layer, communication (service) layer and application layer. The physical layer is the base layer having modules that have access directly to the micro controller and peripherals. Implementations of interfaces for sensors and device drivers are done in this layer. The communication layer has a pre-agreed protocol and information exchange with other layers and connects them to extract

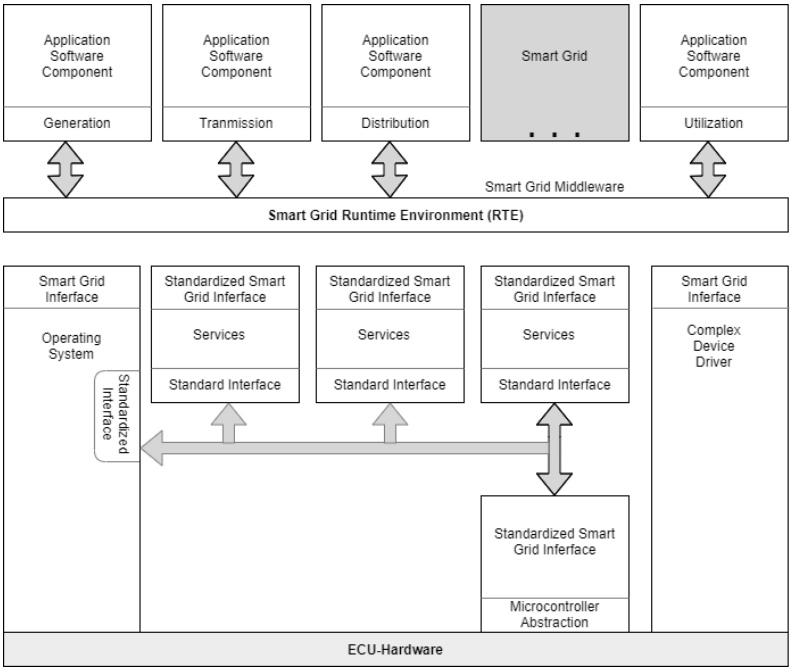



**Figure 2.3:** Embedded software view of Smart Grid architecture [17]

details of modules that are under them architectural wise with main modules like communication protocol interface, operating system interface and device independent service. The application layer is the layer that implements the software functionality. Several smart grid specific features like load forecasting or management of demand is done in this layer.

### 2.1.3 Logical Domains of Smart Grid

The NIST conceptual model of Smart grid is categorised into the following logical domains [15].

1. Bulk Generation
2. Transmission
3. Distribution
4. Operations
5. Market
6. Service Providers
7. Customers

From Figure 2.2, it can be seen that bulk generation, transmission, distribution and customers have bidirectional flow in terms of power generation, storage and delivery. On the other hand; service providers, operations and markets deal with consumer services, power flow and information exchanges. The conceptual model has the following types of customers.

1. **Home Area Network (HAN)**: This is a network confined inside home that connects devices within itself allowing to share resources and also connection to internet [18].
2. **Building Area Network**: This includes network communication within a building that encompasses of several homes [19].
3. **Industrial Area Network (IAN)**: This is a network that spans across a large industrial area. It will monitor and control the devices connected to it.

## 2.2 Overview of IoT

IoT in broad terms are all devices that are interconnected and communicate over internet. These are objects that are broadly scattered with low storage capability, processing capabilities that can improve performance, security and reliability. Some examples of it are smart devices including mobile phones and other objects like appliance, landmark, monument, work of art that can cooperate to provide a common target. Figure 2.4 shows a high level overview of IoT. It shows a communication dimension that can be maintained by anyone irrespective of the location that provides any services shown in the network.This section has been studied as a part of my term paper work in [12] and also extended by a paper where I am the lead author [13] that is under review.

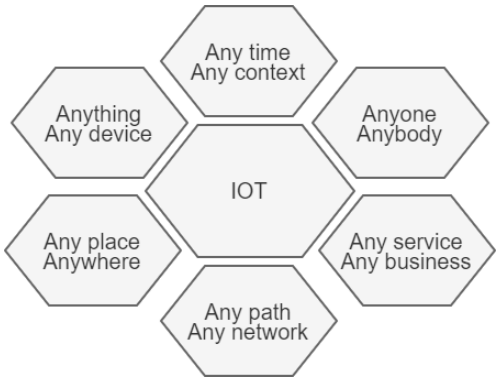



**Figure 2.4:** High level overview of IoT [20]

IoT platforms can be categorized as follows:

1. **Enterprise/Company based**: These platforms are used for management of companies and investment that are independent which are subsequently used by the society to help users.
2. **eGovernment related**: This promotes the economic development and management of a region. The welfare scheme of the government finances this and eventually helps development of eGovernment towards the IoT. Smart city is an example that can be considered as an eGovernment platform and has functions like controlling traffic, security, protection of environment, education and health.

3. **Business oriented**: This attracts investors in developing key sectors in the industry. Smart grids come under this category.

Figure 2.5 shows these categories and sub-categories of IoT platforms.

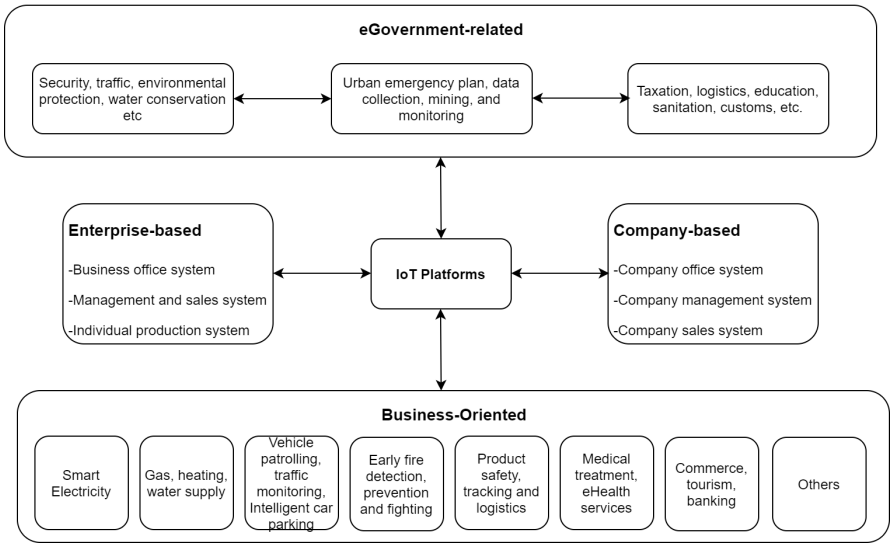



**Figure 2.5:** IoT platform categories [21]

### 2.2.1 IoT Security and privacy attacks and challenges

IoT faces many security and privacy attacks for the users and businesses. Security control depends on three security principles defined as Confidentiality, Integrity and Availability (CIA). Confidentiality is information protection from access without proper authorization, Integrity is the consistency of this information and Availability refers to the information availability whenever it is required [22] . More details about CIA are described in Section 3.4. Different types of attacks may violate the CIA principles such as passive and active attacks [23]. One the one hand, passive attacks do not change the function (does not jeopardize integrity) but leak information (compromise confidentiality) without impacting the expected business behaviour. On the other hand, active attacks aim in hindering the functionality directly (compromising integrity). In addition to potential attacks, the source of these attacks can be internal (insider attack) or external (outsider attack). Insider attack is attack that are originated from within the organization (e.g., attacks by an authorized employee). Outsider attacks are the ones that is originated from outside the organization (e.g., attacks by organized criminals) [24]. In comparison, internal attacks are more dangerous than external ones. Following are list of different attacks faced by IoT [25], to name a few.

1. **DoS Attacks**: DoS attacks aims at limiting the availability of IoT device for the users. Since IoT uses low resources and memory, it can be easily targeted

by DoS attacks. Mitigation against such attacks are difficult to meet due to the previous mentioned reasons. Targets of such attacks can lead to limiting bandwidth, memory, processor, disk space etc. [26].

2. **Physical attacks**: These are attacks on the physical IoT devices itself. Most IoT devices are located remotely with minimal physical security and can be easily manipulated with.
3. **Privacy attacks**: There are different types of privacy attacks that can happen in IoT devices namely:
    - Analysis of traffic: Traffic analysis can also lead to extracting useful information if the traffic is not encrypted. This can identify roles and activities in IoT devices.
    - Mining Data: Analysis of data can give much more information that it is protected for in some databases leading to several privacy issues.
    - Passive monitoring: This is method of obtaining data if transmissions are not encrypted.

Besides the described attacks, IoT also have different types of challenges as follows (among others):

1. **User privacy and data protection**: Privacy is a very important factor of IoT security. This is a delicate issue since devices communicate and data is shared across the internet which is receptive to leakage. There are many researches in the area but still some issues need to be resolved in the area of data collection, sharing, management and security [27].
2. **Authentication and identity management**: This is aimed at managing and protecting the information and also securing the identity. Authentication is essential in communication between devices. Since there are many devices that communicate between each other, management and protection is essential. There are many research areas in connection to this [27].
3. **Authorization and access control**: The method in which an object is able to access the resources after identification is called authorization while access control is the ability to grant or deny permissions based on criterion. Access control supplements authorization and is essential for a secure connection. This should be easily manageable to create and understand.

According to Vermesan and Friess [25] also described by Abomhara and Koien [28], the various requirements in addressing IoT security are as follows:

- IoT devices have less resources, so the solutions should be equally balanced and light-weighted to support it.
- Protect the data that is stored using encryption methodologies thereby preventing unauthorized access.
- Encryption key distribution that is light weight with limited communication and resources to support the resource limited IoT devices.
- Maintaining data privacy of individuals that can be extracted by observation of data exchange in IoT.

- Distributed computing and management of keys to keep information as local as possible.
- "Privacy by Design" concepts such as identification of data identification, authentication and anonymity is supported using different methods

Table 2.1 shows the different security components that influences the security functionality of IoT.

| **Component Name** | **Component Functionality** | **Security Goals** |
|---|---|---|
| Authorization | Access control on devices and services | Data Confidentiality, Data Integrity |
| Authentication | Authentication of service users and device users | Authentication, Accountability |
| Identity Management (IdM) | Management of identities, pseudonyms and related access policies | User privacy, Service privacy |
| Key exchange and management (KEM) | Exchange of cryptographic keys | Communication confidentiality, Communication integrity |
| Trust management and reputation | Service level trust and collecting user reputation scores | Service trust, Service reputation |

**Table 2.1:** Security components influencing IoT security functionality [25]

### 2.2.2 IoT Impact areas, technology and protocols

From the previous chapters it can be seen that IoT is a collection of different technologies and areas of implementation. For it to have maximum impact, an extensive acceptance is required and different policies and technologies need to be in place. Following are some of the enablers that can deliver maximum impact of IoT [29].

1. **Technology**: Inexpensive and low power devices availability is quite essential to the success of IoT. Coupled with it the security and integrity of the collection of IoT devices is required. Availability of enough bandwidth and fault free connection across the different IoT devices can provide continuous operations.
2. **Interoperability**: IoT consists of different devices working together and interoperability is essential to utilize its maximum potential.This involves creating technology and protocol standards that can be used to integrate different technology standards and protocols to communicate between different

devices. Access to different sources can also be standardized.

3. **Privacy and confidentiality**: The data collected by IoT devices are huge in number and this can cause privacy concerns by customers on how the data is treated and the integrity of the data. The providers of these IoT devices need to create a trust and transparency among customers on how the data is shared to create a value addition and ensure that data is protected.
4. **Security**: With the huge amounts of data gathered by IoT devices, they are vulnerable to different security threats. Care should be taken to prevent potential data breaches and security should be provided to the physical devices itself. Protecting the physical devices is essential as a breach to it can even cause physical dangers.
5. **Intellectual property**: IoT devices require different parties to collaborate together to achieve its goal. These include companies and industry verticals to work together and a common understanding of the ownership of the data is essential for smooth operation.
6. **Organization and talent**: IoT is a typical example of a cyber physical system. It combines both the physical and cyber worlds and with it comes various challenges unlike a traditional IT company that deals with only the software side of things. Competency need to be developed to focus on both the software and hardware side of things and investment need to made based on clear business cases.
7. **Public policy**: Regualtory policies need to be in place for implementation of certain IoT devices. A typical example is self driving in smart cars and healthcare industry. Government agreements need to be in place to ensure fair practise for data usage and entities need to be liable for actions taken.

Protocols in IoT are important part of the whole architecture. They are used in combination with the hardware to transfer useful data in a structured way. Protocols facilitate the communication between devices, sensors, servers, gateways and different user applications.Various protocols are used to obtain a standarad IoT environment. These include protocols like XMPP, MQTT, REST and CoAP [30]. XMPP consist of XSF, IETF, W3C, ISO, IEC, IEEE and uPnP; MQTT includes OASIS; REST includes W3C and CoAP includes IETF [30]. 6LoWPAN is the IPv6 over Low -Power Wireless Personal Area Networks is slowly taking over the more expensive IPv6 protocols into small link layer frames [31].

## 2.3 Machine Learning

Arthur Samuel described machine learning as "the field of study that gives computers the ability to learn without being explicitly programmed" [32]. Tom Mitchell introduces a later definition which states that "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E" [33]. An example of this is playing checkers where E is the experience of playing

many games of checkers, T is the task of playing checkers and P is the probability that the program will win the next game.

Machine learning is classified into supervised learning and unsupervised learning. Supervised learning is the one in which the data is labeled and categorized. The system is trained with these labeled datasets and any future data set is categorized according to this pre-trained labeled datasets. So in short it can compare the estimated output to the actual output and based on feedback do corrective actions . Unsupervised machine learning is the one in which the data set is not labeled. Instead the structure between the data points are derived from clusters or relationships after analysis of the data. There is no feedback associated with unsupervised learning. A general schema for machine learning methods is shown in Figure 2.6.

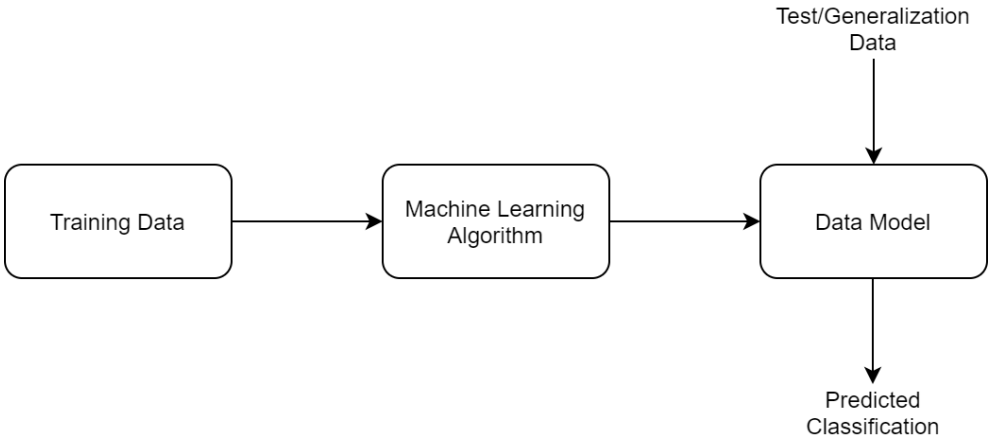



**Figure 2.6:** General Machine Learning Schema

Supervised learning is further divided into classification and regression. Classification problems maps inputs into discrete output categories while regression problems maps input into a continuous output function.

### 2.3.1 Classification

The output of a classification problem are typically classes or categories. These will be represented by categorical data points that are found relevant or simple integers that represent each individual class. Typical examples are predicting what is in a given image, prediction of a person going to default on credit card payments etc. So classification is essentially trying to predict a category. One of the most commonly used classification algorithm is the decision tree. A decision tree determines ways to segregate based on different conditions. A typical example is the titanic data set to predict whether a passenger survived or not. This is represented in a decision tree as shown in Figure 2.7.

This model uses three features from the data set namely sex, age and number of spouses or children along. The passenger survival here is shown as survived

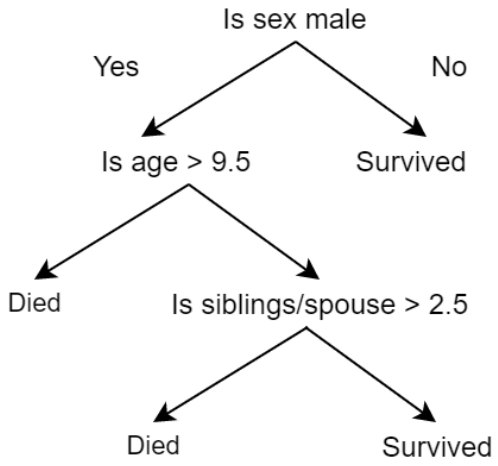



**Figure 2.7:** Decision Tree example

or died as the decision. As it can be seen from Figure 2.7, the algorithm is quite simple. Creating a tree involves the following:

- Features to choose.
- Conditions for splitting.
- Knowing when to stop.
- Pruning.

The different classification algorithms used in the experiment are as follows:

- Random Forest.
- Bayesian Network.
- SVM.
- C5.4 decision tree.

These are further detailed in Section 4.4.

### 2.3.2 Selection of features

The quality of features is quite important for the machine learning algorithms to give the correct results from analysing the datasets. This can be done manually or by using algorithms and should be done prior to executing the classification algorithms. This thesis uses a combination of both manual and algorithmic to select the feature set as detailed in Section 4.1.2 and Section 5.2 respectively. Feature selection enables getting a better insight into the outcome of executing the classification algorithm and hence is essential. The total number of features in the feature set also impacts the performance of the chosen algorithm and further the classification process and hence is better to have less but the right amount of features that will enable machine learning models to perform better. This is commonly referred to as the curse of dimensionality [34].

### 2.3.3 Evaluation metrics

The performance of various classifiers are also evaluated by the evaluation metrics like accuracy or recognition rate, confusion matrix, recall , false positive rate (FPR), sensitivity or true positive rate (TPR), specificity, learning time, precision and ROC curve. The four concepts behind these metrics are defined as follows:

- **True Positive (TP)**: Samples that are classified correctly as positive.
- **False Positive (FP)**: Samples that are classified as positive but not correct.
- **True Negative (TN)**: Samples that are classified correctly as negative.
- **False Negative (FN)**: Samples that are classified falsely as negative.

The confusion matrix generated using the above four concepts is given in Table 2.2 known as the confusion matrix.

| | 1 (Predicted) | 0 (Predicted) |
|---|---|---|
| 1 (Actual) | True Positive | False Negative |
| 0 (Actual) | False Positive | True Negative |

**Table 2.2:** Confusion Matrix

This leads to defining the below:

**Accuracy or recognition rate:** It is defined as the percentage of data set in the test set that is correctly classified and is represented with the formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \tag{2.1}$$

**Recall:** It is defined as the measure of the accuracy of detected positive instances.

$$Recall = \frac{TP}{TP + FN} \tag{2.2}$$

**Sensitivity or True positive rate (TPR):** This is the same as recall as mentioned before.

**False Positive Rate (FPR):** This is the percentage of falsely classified normal instances.

$$FPR = \frac{FP}{FP + TN} \tag{2.3}$$

**Specificity:** This is also known as the True Negative Rate (TNR) and calculates the ratio of negatively classified instances.

$$Specificity = \frac{TN}{TN + FP} \tag{2.4}$$

**Learning time:** The time taken to build a model based on the classifier and the training data set.

**Precision:** This is the percentage of relevant results represented with formula:

$$Precision = \frac{TP}{TP + FP} \tag{2.5}$$

**F-Measure:** It is the harmonic average of precision and recall represented by the following formula:

$$F - measure = \frac{2 * Precision * Recall}{Precision + Recall} \tag{2.6}$$

**ROC Curve:** The relation between sensitivity and specificity is shown by the ROC curve with the aim being to increase the area under curve (AUC).

# Chapter 3

# Related Work

This chapter focuses on the related research; IoT in Smart Grids and the security and privacy of IoT enabled Smart Grids. It also briefly gives an overview of anomaly detection and its algorithms which can be used to detect abnormalities in data and hence attacks.

## 3.1 IoT in Smart Grid

IoT technology have a important role in building a smart grid infrastructure. The processing ability and sensing, enable smart grid to achieve advanced characteristics like self healing, bi-directional communication, recovery from disaster etc. It can be also used for obtaining secure communication in different parts of the smart grid. This section has been studied as as a part of my term paper work in [12] and also extended by a paper where I am the lead author [13] that is under review. Some examples are as follows [35]:

- Monitor electricity generation in power plants like coal, wind, solar, biomass. It can also calculate the energy requirements of customers and its storage. Energy storage and different emissions can also be monitored.
- Monitor and protection of transmission lines used for power transfer, control the devices used for transmission and access the electricity consumption.
- From a consumer point, IoT have various uses. It can be used in smart meters to monitor power usage, control the electric car charging, manage the energy usage in a household and to ensure the continuous connectivity between different networks.

Some of the applications of IoT in smart grids are as follows: Monitoring of transmission lines, Smart homes, Electric vehicles charging systems and monitoring, Advanced Metering Infrastructure to gather various measurements like monitoring energy usage and pattern of usage.

| | | | | |
|---|---|---|---|---|
| Layer 4 | | Application | Social | Master station system |
| Layer 3 | Application | Cloud management | Application | Remote communication |
| Layer 2 | Network | Network | Network | Field network |
| Layer 1 | Perception | Perception | Perception | Perception |

**Table 3.1:** IoT architectures in smart grid proposals [35]

### 3.1.1 IoT architecture in Smart Grid

Table 3.1 shows some of the proposed IoT architectures in smart grids.

The proposed architectures are either three or four layered as shown in four different columns respectively. The first column shows three layers namely perception, network and application layer [35–37]. Perception layer collects data using various sensors, tags, readers. Network layer maps data gathered by perception layer to different communication protocols using wired or industry standard wireless networks. The industry standards include 3g, 4g, 5g broadband, Zigbee or Wi-Fi and further transmits the data to application layer that can monitor the IoT devices in real time. It contains an application structure that can compute and process information and enable interfacing and integration.

The third column in Table 3.1 shows four layers namely perception, network, cloud management and application layer [35, 38]. Here the perception layer consist of a thing layer that comprises of different sensors, tags, readers to sense, control, collect data and a gateway layer that comprises of microcontrollers and display which controls elements that connect to thing layer. As in previous case, the network layer transmits data from perception to application layer which inturn can provide services to consumers and managing energy pricing. Cloud management layer stores and analyses data and also manages users.

The fourth column in Table 3.1 have the same elements as before like perception, network, application layer and a social layer [35]. The social layer integrates and regulates various IoT applications.

The last column in Table 3.1 has a terminal, field network, remote communication and master station system layer [35]. This consist of remote units, smart devices, smart meters in the terminal layer; different communication channels like optic fiber, Wi-Fi, Zigbee etc. for field network layer; 3g, 4g, 5g or wired communication fro the remote communication layer; control systems for smart grids for master station layer.

### 3.1.2 IoT requirements in Smart Grid

IoT can be used in smart grids after following certain requirements. These are as follows [35]:

- Communication technologies: As mentioned in the previous subsection, com-

munication technologies processes state of the smart grids using the data it collects. It can either send or receive data and can be categorized in to short or long range of which optic fibre, mobile, satellite networks etc. are considered long range and Bluetooth, ZigBee etc. are considered short range communication.

- Data fusion techniques: This is used to acquire and merge data that is collected to broaden the administration of information that is gathered considering the fact that most IoT devices have less number of resources like storage, memory, battery capacity and bandwidth.
- Energy harvesting process: This can be used to oversee different parts of smart grids using different sensors for example as the IoT devices uses battery as their source of energy. Hence harvesting energy from different parts of the system is important.
- Operating in harsh environments: Since IoT devices are usually implemented in harsh environments, they should be resistant to different temperature scales, waterproof and resistant to electromagnetic waves. Extreme weather condition is one example of a harsh operating environment.
- Reliability: IoT is integrated into smart grids to compliment certain properties like self healing, reliability, organization on its own etc. Hence these should overcome any issues faced on its way and maintain reliable data at an acceptable level.
- Security: Data protection against leakage or losses, securing them while transmission, storage and management should be primary for IoT devices in terms of security.
- Sensors: Sensors are used for different purposes like measurement of energy, temperature, voltage, frequency etc. that further distribute the information collected for analysis and processing.

A summary of the different communication technologies and frequency ranges used by IoT devices are shown in Figure 3.1.

| Technology Name | Standard Name | Frequency Band | Coverage Range | Data Rate |
|---|---|---|---|---|
| ZigBee | ZigBee | 2.4 GHz | 100 m | 250 Kbps |
| WiFi | IEEE 802.11 | 2.4 GHz, 5 GHz | 150 m | 1 Gbps |
| WiMAX | IEEE 802.16 | 10–66 GHz | 50 km | 75 Mbps |
| Thread | IEEE 802.15.4 | 2.4 GHz | 30 m | 250 Kbps |
| Z-Wave | Z-Wave | 900 MHz | 30 m | 100 Kbps |
| Bluetooth | IEEE 801.15.1 | 2.4 GHz | 10 m | 1 Mbps |
| Cellular | 4G | 1.4–20 MHz | 50 km | 100 Mbps |
| LoRa | LoRa | 863, 915 MHz | +10 km | 100 Kbps |
| Sigfox | Sigfox | 863, 915 MHz | +10 km | 10, 100 Kbps |
| PLC | IEEE 1901 | 500 kHz | 3 km | 10–500 kbps |
| Ethernet | IEEE 802.3 | 100 MHz | 100 m | 100 Mbps–10 Gbps |
| Fiber optic | IEEE 802.3 | 500 MHz | 100 km | 40 Gbps |
| Satellite | IEEE 521 | 30–300 GHz | 6000 km | 1 Mbps |

**Figure 3.1:** Communication technologies in IoT [21]

## 3.2 Privacy and security implication in smart grid

Smart grids have several advantages over traditional power grids some of which are described in previous chapters. With the advancement in technology there are several security and privacy aspects that need to be taken in to consideration. These security and privacy aspects can be complicated and are essential factors in maintaining the confidentiality, integrity and availability of the system. Various measures need to be in place to keep the security and privacy intact. Several researches have been on going in this area and [30] have consolidated some of them as below.

Winter [39] has applied a framework that is relevant to the privacy established by Nissenbaum [40]. This is a tool to grasp subject response to implementation of sensible metering technology in home area networks. The research determined illegitimate use of data that is personal, interpretations made using data mining from the collected data and leakage of data. This can be quite dangerous as it can be used against specific individuals or selected groups.

Haddad et al. [41] have suggested a strategy to enact the basic security requirements like confidentiality, authentication, integrity of data and key agreements without confiding in LTE-A networks. The study is about protects AMI-UC communication through the LTE-A networks by implementing security and privacy. Conclusions of the study shows that this strategy is secure.

Suggestions from Eriksen [42] preserves the safety and privacy of the measurements that are susceptible in nature through encrypted protocols. This facilitates the supplier of the utility to gather the statistical information that is necessary. The suggested proficient protocol is reasonable for a group during a dynamic setting by joining the Chinese Remainder Theorem with altered homomorphic encryption.

Rahman et al. [43] suggests a completely unique protocol to share required information among users providing privacy, confidentiality, and integrity. They also suggest a replacement clustering-based, distributed multi-party computation (MPC) protocol. The author aims to implement a collaboration between legitimate and dishonest users in smart grid and prevent dishonest users from falsifying data of usage.

Saputro et al. [44] have organized experiments in a smart grid architecture that is hybrid in nature. It gathers data from smart meters using an LTE based wide area network and an 802.11s based smart meter without making changes to the components in the architecture. Results show that the IP address of the smart meter in the destination is found without having any separate overhead. This is used to gather data from smart meter depending on the identity of the user in the packets that is received.

Weiwei et al. [45] suggests two new protocols to attain the privacy of data in smart meters. These are namely the basic scheme and advanced scheme. It also analyzes and finds a new attack and establishes it by which an attacker can conclude readings from meter based on the information that a person is present or absent. The protocol is simple and quite practical with high efficiency as meas-

ured by performance and utility tools and analysis. Evaluation of the protocol is done using several methods, one of which is a java implementation with distinct parameters. The research guarantees privacy based on the security analysis of the protocol that is suggested.

There are also several literature reviews on smart grid authentication. One of them [46] enables the integrity, robustness and availability of the smart grid structure using less overhead security. This is done by collecting data from energy consumption of electrical devices in a smart grid architecture and analysing this data by suggesting an authentication system among the smart meters and the system utilities.

### 3.2.1 Security in IoT based smart grid

Safety, stability and reliability of smart grids depends on securing IoT devices integrated in them. Efforts on confidentiality, integrity, availability, authentication of smart grids and IoT have been gaining importance these days. [30] have again consolidated research areas in this field as stated in the following paragraphs.

Confidentiality is assured in wireless sensor networks (WSN) in the background of IoT as shown in the researches, surveys and literature's [47–51]. This is in relation to the security rules for ensuring confidentiality which provides an exclusive and precise clarification. The suggestions that are missing are proposed in [52].

Presentation of classifications of requirements for security of smart grids is described in [53]. Privacy, threats, security liabilities in IoT is composed through simple classifications in this paper and is expected to be investigated further. The role of wireless sensor networks for smart grids are proposed in [51]. This also suggests solutions for IoT in smart electricity consumption.

Another IoT application architecture in smart grid is proposed in [54] that provides a overall overview. It suggests a scheme with security instructions and present the secure access control system that is trusted for the safety of smart grids with IoT that is designed for the IoT terminals.

Discussions on the causes and according to what the manufacturing and usage of electric energy that will be indivisible or identical from IoT is shown in the paper [55]. The paper also lists other aspects that cause the integration of smart grids with IoT like outset of smart grids that has utilities, outset of smart grids that have consumers in one end and other troublesome technologies.

IoT in smart grids is presented from an architectural point in [56]. Here a three layered architectures is presented for the smart grid as per the system architecture of IoT and each layer describes in detail the utilization of IoT.

New risk assessment methodology in smart grids in presented in a straightforward and efficient manner in [57]. This paper determines the security, dependability and privacy of a smart grid. The procedure begins with assessing the components, then the subsystems and towards the end the assessment of the total system.

The adaptation of wireless sensor networks to IoT still have many compatibility issues considering the different applications that are suggested in the area. These are mainly due to the diversity of devices in the architecture of smart grids. Traditional security mechanisms like encryption may not be feasible for security such infrastructures and possibility of developing newer security mechanisms need to be investigated. Research areas on managing distinct keys, key sharing systems, authentication layer in the network, confirmation of integrity from beginning to end so that the entire system is prepared and immune to different kinds of attacks are required [58]. Some research papers like [59–62] answer the previous mentioned areas to an extent in terms of protocols for authentication in IoT.

Privacy protection is done using encryption mechanism established on XOR procedures as shown in the research in [59]. Remote access in facilitated to securely handle sessions keys using a sensor node as suggested in [60]. This research area is in the wireless sensor network background where authentication of users in a composite Wireless sensor network (WSN) and arrangement of a key plan is formulated. Establishment of session keys using elliptic curve cryptography (ECC) has been proposed in [61]. This has been enforced to institute access control policies that are based on attributes and also determining results that are resource constrained in IoT application level.

Protection of privacy protocols among smart meters and providers of energy are presented in the paper [62]. It can aptly conserve the series of data measurements that are gathered by smart meters. This is done by an adaptation of encryption strategy depending on elliptic curve cryptography (ECC).

## 3.3 Anomaly Detection Overview

This section has also been studied as a part of my recent project work in [63]. Anomaly analysis is done as part of experiments in Chapter 5.

### 3.3.1 Anomalies

Anomalies are identified as the abnormal instances within a data set which deviates from a normal pattern. Grubbs in 1969 [64] defined it as: “An outlying observation, or outlier, is one that appears to deviate markedly from other members of the sample in which it occurs”. Outlier is another term that is often used in lieu of anomaly. The present-day anomalies of today have the following characteristics:

- Anomalies are uncommon when related to instances that are normal in a data set.
- The features of anomalies are also distinct when compared to other instances.

Anomaly detection started with the removal of outliers from data used for training algorithms also known as data cleansing [65]. It is further developed in

to analyzing the outliers for more features.

### 3.3.2 Anomaly detection modes

Anomaly detection can be mainly classified into the following three modes of operation:

**Supervised Anomaly Detection:** The instances of the data sets are labelled for both the training and test data. The classes are strongly unbalanced classes in the sense that the anomalous instances are less in number than the normal instances. Support Vector Machines [66] and Artificial Neural Networks [67] can perform well with this type of detection whereas decision trees do not deal with this good.

**Unsupervised Anomaly Detection:** All the instances in the data set for unsupervised anomaly detection are unlabeled. Training and test data sets are not categorized separately in this process and mostly distances or densities of data clusters are often used to determine an anomaly as compared to normal instances. This category assumes that majority of data set instances are normal with few anomalies.

**Semi-supervised Anomaly Detection:** Here the instances of training data set contain are normal without outliers and is labelled. The training set is used to learn about normal instances any deviations from this is considered as an anomaly. This is called one-class classification [68]. One-class Support Vector Machine [69] and auto encoders [70] are most used algorithms in this category.

Figure 3.2 shows the pictorial representation of the above-mentioned modes of operation.

### 3.3.3 Anomaly Types

Anomalies can be local, global or micro clusters as show in Figure 3.3. Here $X_1$ and $X_2$ are global anomalies as it does not belong to any dense clusters. $X_3$ looks like a normal instance in the big picture but considered as an anomaly with respect to cluster $C_2$, hence called a local anomaly. Cluster $C_3$ here can be either seen as a group of anomalies and considered as micro cluster considering the context of the data set.

Anomalies can be further categorized into point anomaly, collective anomaly and contextual anomalies. When single data instance is identified in a bigger data set, it is defined as point anomaly [72]. Collective anomaly is defined as an outlier that is presented in a set of instance collection. An instance that is otherwise normal is categorized as anomaly in given context is defined as contextual anomaly.

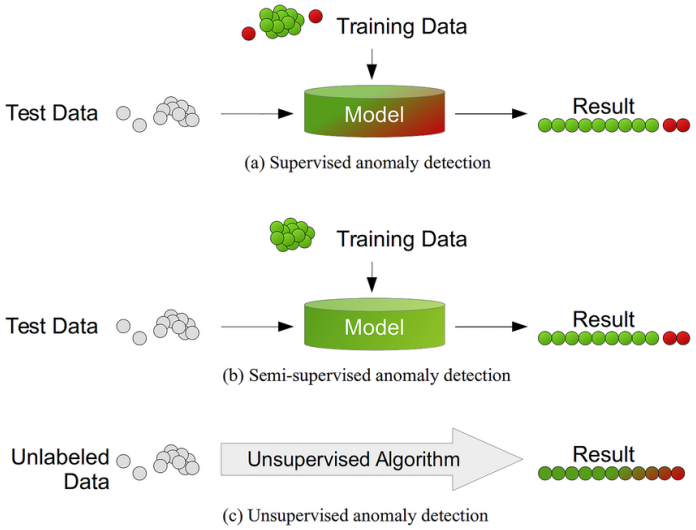



**Figure 3.2:** Anomaly detection modes [71]

Point anomalies can be used to detect contextual and collective anomalies. For collective anomalies aggregation and correlation is used for new data set generation and representing the features a different way. In the case of contextual anomaly this can be done by representing the context itself as a new feature.

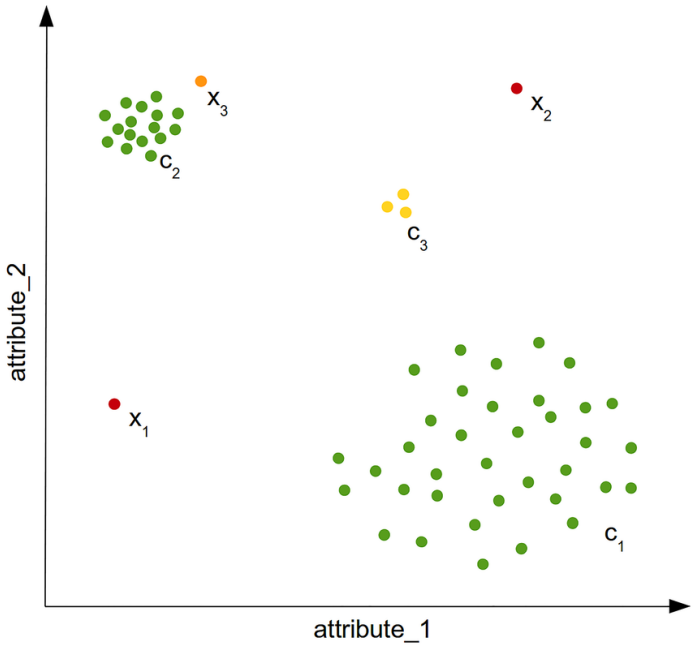



**Figure 3.3:** Local, Global and Micro clusters [71]

### 3.3.4 Output of Anomaly detection algorithms

Labels and Scores are the two possible outcomes of an anomaly detection algorithm. A label indicates whether a given instance is anomaly or normal whereas a score shows the abnormality degree. Scores are mostly used in unsupervised algorithms while labels are mostly used for supervised and semi supervised algorithms. Scores can be used to rank anomalies and inform the top outliers to the users of the system.

### 3.3.5 Applications of Anomaly Detection

There are different domains and application areas in which anomaly detection is used. Listed below are the majority and commonly used ones.

**Intrusion Detection:** This is one the most acknowledged areas of anomaly detection application. In this scenario, an anomaly in network traffic or applications running on the server is categorized as intrusion and will be detected using corresponding anomaly detection algorithms. This is called network-based intrusion detection. Intrusion detection systems that deal with system calls of operating systems are called host-based intrusion detection also known as behavioral analysis [73] in terms of anomaly detection. Since the data analyzed for such kind of systems should be in near real time, the algorithms should also be fast which is challenging. Semi-Supervised or unsupervised anomaly detection is preferred type of algorithms in the case of intrusion detection.

**Data Leakage prevention:** Data leakage is prevented in the initial stages thereby protecting confidential information loss by using this method [74]. Since the data loss needs to be prevented in an early stage, real-time protection is required. Frequent logging of all the access points helps to detect point of leaks in case of anomalies.

**Fraud detection:** Analysis of logs can detect fraudulent activities. This can be used to detect anomalous transactions in a bank, credit card misuse, frauds happening on online e-commerce and insurance claim frauds.

**Medical applications:** Anomaly detection in this area require high percentage of true positives as these are aimed at critical functions [75] monitoring health of patients like heart rate, detecting the outbreak of diseases like corona as in current pandemic situation world is facing. Other application of this include detecting anomalies from CT scan images and finding mutants.

**Other applications:** Anomaly detection can also be used in mobile networks, smart grids, smart buildings for detecting anomalies in consumption, tampering of meters and document forgery and surveillance cameras.

### 3.3.6 Anomaly detection algorithms

This section describes in short, the various anomaly algorithms detection algorithms that are commonly used.

**K-Means Clustering (K-Means):** K-Means is a type of clustering algorithm [76], that categorizes data using their features into predefined K clusters. The centroids of the initial clusters are initialized arbitrarily, and each record of the dataset is assigned uniquely to one of the centroids. This is based on similarity of features which is calculated by the distance to the centroid. The centroids are changed, and clustering of data is done repeatedly until its not possible to change the centroids. Anomalies are identified as the records that are farthest from centroids and scores are calculated based on the distance.

**Local Density Cluster-based Outlier Factor (LDCOF):** LDCOF [77] is used in the estimation of density of clusters based on the assumption that the members of the clusters have a distribution that is spherical. Clusters are categorized into small or large based after applying k-means. The distance from centroid of a cluster to each dataset instance divided by the average distance of all the dataset instances to the centroid gives the LDCOF score. This score can change according to the densities of the cluster and scores equivalent to 1 or less is considered as a normal dataset instance.

**Kth-Nearest Neighbor (kNN):** kNN [78] is used to identify global anomalies as opposed to local ones. The k-nearest neighbors are identified for each item in the dataset that contain similar features which are called the nearest neighbors (NN). If majority of the nearest neighbours are anomalies, then the identified item is the dataset is also considered as an anomaly.

**Local Outlier Factor (LOF):** LOF [79] is a local anomaly detection algorithm that detects anomalies based on local density of a record. After the calculation of NN for each record, Local Reachability Density (LRD) is computed based on this. The comparison of LRD of current and previous record gives the LOF score. Like LDCOF, scores above 1 are considered as anomalies.

**Connectivity-based Outlier Factor (COF):** COF [80] differs from LOF in density calculation. This is done by calculating the chaining distance that is the shortest path which is least possible summation of the distances that connects the dataset instance and all its K neighbors.

**Histogram-based Outlier Score (HBOS):** Anomaly detection can also be used in mobile networks, smart grids, smart buildings for detecting anomalies in consumption, tampering of meters and document forgery and surveillance cameras.

**Robust Principal Component Analysis (rPCA):** rPCA [81] is based on Principal Component Analysis (PCA) that detects sub-spaces in datasets. Anomalies

are reported based on the variation from the normal sub-spaces. The computation of PCA involves calculation of eigen vectors of the covariance matrix.

**Isolation Forest (IF):** IF [82] categorizes dataset instances as nodes of an isolation tree with the assumption that most instances are normal and anomalous ones are not large in number. Anomalies are isolated and near to the tree root and classification is done based on the instance distance from the root.

## 3.4 Attacks in Smart Grids

Smart Grid as mentioned in Section 2.1 is a cyber-physical system that consists of both the physical and cyber part. The addition of cyber part to the grid has made the grid more vulnerable to attacks and the importance of security is ever increasing. Since this a combination of two different systems;an attack on the cyber part will affect the physical part and vice- versa. The physical part is the infrastructure required for the smart grid. The attack in the physical part will affect the following areas:

- Components
- Connections
- Connectors

The attack on cyber part in general will be on the following:

- Accessing the system without authorization
- Data and information manipulation
- Connectors

This thesis focuses on the cyber part of the smart grid. To maintain the security of Smart Grids, NIST have defined certain criterion. These are Confidentiality, Integrity and Availability [83]. All these three should work together and any one is not working properly then the other two are not working well either and it is fundamental in terms of security. Following are short descriptions of these criterion:

- **Confidentiality**: Information that is contained in databases, systems etc. is confidential and other people who do not have access to this information are not allowed to view them. Examples of this in a smart grid scenario are billing information that is sent from power companies to customers, usage information by customers, access and control parameters of the meter should be confidential. If these information is compromised, it can be exploited and could lead to other consequences.

- **Integrity**: The information that is send back and forth within or between systems are things that cannot be modified without the knowledge of authorized people or systems. So the receiving end can trust the information that it receives. A typical example of attack on integrity is false data injection where the measurements of meters can be modified and be a serious

security threat.

- **Availability**: Like the name implies, the system will always be up and running and huge part of all the systems. Almost everyone in an organization will be involved in some aspect of availability. For example if the the system is not available the flow of information is also restricted which can lead to loss of control to the operators of the system.

In general, there are four steps in which an attacker tries to takes over a system. These are shown in Figure 3.4

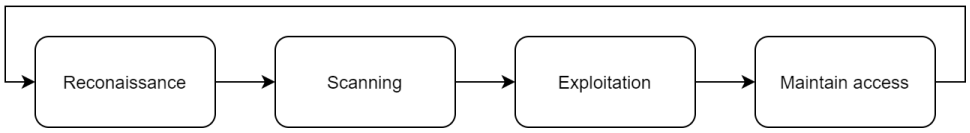



**Figure 3.4:** Cycle of attack [84]

Following are detailed explanation of the steps in Figure 3.4.

- **Reconnaissance**: This is the phase in which the intruder tries to assemble information about the target. Different methods are employed for this like analysing the traffic and social engineering. Social engineering is the process of getting a person to perform an action in the context of security like getting a person to disclose information, download malware or perform an unauthorized transaction. It is considered as a human attack as humans are considered commonly as the weakest link in the security chain. From an attackers point of view, it is easier to trick an employee perform an action like divulging a password than it is to circumvent and control like decryption of the password. Traffic analysis is the process of analysing traffic volumes, patterns and rates for gathering information about the devices they are connected to and its hosts. Combined with social engineering, they compromise the confidentiality in a system.

- **Scanning**: The second phase analyses the weaknesses in a system. Depending on the type of scan, different information can be gathered. The simplest information that can be gathered is to find the active hosts. This is called host scanning and to achieve this a scanner will progressively contact all the hosts in a target set which can either be a target network or list of IP addresses. Another type is port scanning to find the open ports in a host. This information can be considered as good approximation of which services are active on a host. There are TCP and UDP port scanning methods where TCP method uses TCP responses to find if port if open or closed. UDP scanning is simpler but unreliable than TCP as it needs more guessing to identify open ports. These type of attack also compromise the confidentiality of a smart grid system.

- **Exploitation**: The third phase exploitation manipulates the weaknesses of the units of a smart grid and gains access to it. This done by using attacks such as viruses, trojans, denial of service, man in the middle attack, violating privacy and integrity to name a few.

- **Maintain access**: This is the last phase where once the attacker had gained access to a system will try to maintain the access in the system. This is done by either creating backdoors, trojans or viruses. A backdoor is also a type pf trojan or virus and is designed such that it does not attack, destroy a system or consume bandwidth but allows to re-enter a system with little or no authentication. It can be a remote screen grab, screen control, capture inputs on a certain port and pipes. Using this access to the system it can affect confidentiality, integrity and availability.

# Chapter 4

# Methodology

The aim of this chapter is to examine and illustrate the methodologies used to justify the research questions in the thesis. The framework of this chapter is a top-down approach, where the different scenarios considered for experiment are first presented followed by the choice of datasets and the process involved in its collection. Data set pre-processing and extraction of features are described concluded by its classification methods are discussed. Quantitative methods [85] with a combination of literature review is used in this study. This have an impact on the methods described in the chapter and justifies it.

Below is a summary of different stages used in the methodology to execute the experiments:

1. Identify the different use/misuse cases
2. Gather datasets from different sources
3. Dataset pre-processing
4. Identify the features from the datasets based on scenarios
5. Application of feature selection on pre-processed dataset
6. Application of machine learning algorithms to detect attacks.
7. Results validation

## 4.1 Datasets

Different datasets have been considered for use in this research. The initial plan was to collect data from different Smart Grid providers across Norway. However due to data privacy and protection, this proved to be difficult. An alternate option considered was generating the required data using the National Smart Grid laboratory in NTNU, Trondheim [86]. The laboratory had a Smart home/Smart building automation domain that could have been used. Travel restrictions and access to the university as a result of COVID-19 made this as a non-viable option and remote access was not possible either.

Various other publicly available datasets were evaluated and there was two that were of interest. Following are the two used for the experiments:

**UMass data set [87]:** This is a publicly available dataset generated by the laboratory for advanced software systems as a part of optimizing consumption of energy in smart homes. This includes measurements from data like electrical usage and generation of different devices collected from 7 real smart homes over several years (2014-2016). The number of features varies according to the home and the meter from which data is measured.

**NTNU S-building dataset:** This is real data collected from NTNU, S-building in Gjøvik over a period from October 2019 - April 2020 measured every hour. This consists of mainly energy and power consumption of different rooms in the building. It also measures other values like consumption for ventilation, hot water etc. In total, there are about 100 features present out of which useful features needed for the machine learning algorithms are selected. This is not publicly available dataset. Following is short description of the building.

The NTNU S-building (Smaragd) is a special one that is categorized as a near-low energy and modern building with 5 floors. On the fifth floor there are many staff offices and are quite often occupied during weekdays. The rest are lecture rooms, classrooms and auditoriums to name some. On the ground floor is a manufacturing laboratory with energy intensive operations which also contribute to energy intensive consumption. The building is fully automated in terms of temperature, ventilation. It is heated using district heating which has been studied in [88] from Eidsiva Energy located in another place. The heat is transferred to NTNU S-building among other through underground pipes which in turn is used for space heating, office heating and domestic hot water in the pipes. There is also a substation in S-building that converts the incoming heat into suitable temperature that is usable. Here there are are IoT integrated smart meters that report back to Eidsiva. On top of the S-building there are 280 roof mounted photovoltaic (p.v) modules distributed across two roof levels. Tht total output capacity of this is 91.6 kWp and the estimated production is 74.7 MWh. If there is over production of energy, then it is fed back to the grid and back to Eidsiva. The p.v have an efficiency of around 20% which is quite good and there is a control system where operators view and control everything and measure several values from which the dataset is obtained.

### 4.1.1 DataSet description

As mentioned in Section 4.1, two different datasets were considered for the experiments. The first one, UMass data set consists of energy consumption from 7 different homes for a period of 3 years measured every 30 minutes. Following are the features from one of the homes (HomeA) of the UMass data set.

- Date & Time
- Use [kW]

- Gen [kW]
- FurnaceHRV [kW]
- CellarOutlets [kW]
- WashingMachine [kW]
- FridgeRange [kW]
- KitchenLights [kW]
- BedroomOutlets [kW]
- BedroomLights [kW]
- MasterOutlets [kW]
- MasterLights [kW]
- DuctHeaterHRV [kW]

As the name suggests these are different appliances and outlets in a smart house. Similar features can be found depending upon the meters from other homes. The experiments in Chapter 5 show that this data set cannot be further used to answer the research questions as the detailed semantics of the data is unknown.

The second data set from NTNU in Gjøvik is the result of collection of measurements from different points in the S-building. This is considered as a smart building. As mentioned in Section 4.1 this dataset have about 100 features out of which relevant features need to be selected. Some of the important ones considered for the scenarios are as follows.

- Date & Time: Date and time as to when the measurement in taken. In this case, data collection is done on an hourly basis from 00.00 hr to 23:00 hr from 16.10.2019 to 20.04.2020.
- Energy hot tap water-kWh: This is the energy used by the hot water pipes. This is also defined as social load and depends on activity.
- Energy ventilation-kWh: Energy consumed by the ventilation system
- Total Consumption-kWh: Total energy consumption for a particular section in the building.
- Power-kW: Power consumed by a particular section in the building.
- Energy measurement-kWh: Energy consumed by a particular point where the sensor is placed in the building.

### 4.1.2 Data pre-processing

The UMass dataset lesser features as compared to NTNU S-building dataset and the need for feature selection did not arise while the NTNU S-building dataset had several features which made the pre-processing necessary. Part of the feature selection is done manually and the rest using feature selection algorithms that is detailed in Section 5.2. The dataset for NTNU S-building that was received was four different csv files with different features for the same day. This four files were combined which created a single file for each day and resulting in 100 features. These features were manually analysed together with the domain expert and finally 50 relevant features where selected. Further steps of reducing the features as mentioned before are in Section 5.2.

There was some data pre-processing done for the UMass dataset as well. The dataset given was for several homes collected from three different years 2014, 2015 and 2016 and from different meters. A combination of the files for each meters from different years were done. Various machine learning algorithms can be used on this combined dataset and will be more efficient in performance and will provide better estimation in classification of the dataset.

### 4.1.3 Logical flow of the process

A data flow diagram as shown in Figure 4.1 supplements to further understanding of the logical flow of the process. The diagram shows the sources of raw data collection as explained in Section 4.1.

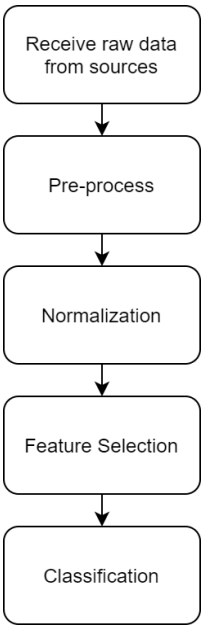



**Figure 4.1:** Data Flow

## 4.2 Scenarios

This sections illustrates different scenarios that can be experimented with the datasets in trying to answer the research questions.

### 4.2.1 Scenario 1

The initial plan was to collect data from NTNU S-building and use this data combined with machine learning algorithms to detect attacks. As the data collection started, COVID-19 was declared as a pandemic. This led to the closure of NTNU and the staff and students were asked to work/study from home. Courses were

conducted online and so as exams. Due to this, analysis is done to check if this has caused any effect and whether this can be considered as an indirect attack.

COVID-19 is considered as an event, and can be used for positive benefit for NTNU/organizations. As students, staff etc. were asked to work/study from home, monitoring of different parameters (like hot water consumption, ventilation energy) at offices and classrooms can tell if someone is 'violating the state' rule of home quarantine guidance. Some states have a restricted rule regarding quarantine and if someone violate these rules, it could lead to imprisonment of hefty fines. This monitoring can help the states/organizations (positive benefit) to execute home quarantine.

### 4.2.2 Scenario 2

False data injection on real data to see the effects of using machine learning algorithms. Analysis of how this affects the measurement of different parameters (like hot water consumption, ventilation energy, power and energy measured at different sensors in different parts of the building) can be done. For example, a system might be under attack but due to the false data injected one may believe that it is not under attack or vice versa can happen.

Data set used: NTNU S-building.

### 4.2.3 Scenario 3

Scenario in which ventilation is closed or opened. This opening and closing can change in the period before and during corona. Ventilation is closed to prevent the spreading of disease during corona time. If there are people present in the building when this happens, it might be also hazardous to health.

Data set used: NTNU S-building.

### 4.2.4 Scenario 4

This is a scenario in which energy and power consumption is monitored. The energy and power consumption can change in the period before and during corona. Ideally the power and energy consumption should go down during corona time. If there are people present in the building during this period it might indicate other wise.

Data set used: NTNU S-building.

### 4.2.5 Scenario 5

Analyze consumption data and check for change in consumption using machine learning algorithms. These are datasets obtained from usage in Homes and can be checked to see if there are any anomalies which could show that there is an attack.

Data set used: UMass.

## 4.3 Misuse Cases

Based on the scenarios mentioned in Section 4.2, a set of misuse cases can be derived. Use case represents the normal activities of the system while any potential attacks against normal behaviour is termed as misuse case [89].

### 4.3.1 Attack on data integrity

Integrity measures protect data from unauthorized alteration. Scenario 2 and 5 can be considered as misuse cases in this category. This leads to the loss of data and consumption data becomes corrupted or unintelligible.

### 4.3.2 Attack on service availability

Availability is the case in which uninterrupted service is not available to authorized users of the system. In other words, this results on disruption of access to or use of services. Scenarios 3 and 4 can be considered as misuse cases of this category.

## 4.4 Algorithms for Machine Learning Classification

The different classification algorithms used in machine learning models to classify the dataset are as follows:

**Random Forest:** This is one of the most powerful supervised machine learning algorithm that is capable of both regression and classification. The random forest is created a different decision trees with more trees giving more robust predictions and hence accuracy. It models decision trees as mentioned in Section 4.4 to create the forest. To classify a new object based on attributes, each tree provides a classification and the tree votes for that class. The forest then chooses the classification having the most votes over all the other trees in the forest. Bagging [90] and Boosting [91] are two different techniques used to combine the decision trees to random forest.

**Bayesian Network:** It is graphical representation of the different probabilistic relationships between various random variables which are there in a particular domain [92]. It has two key elements namely Directed Acyclic Graph (DAG) with all the nodes in the particular domain and probability table of a node which represents the dependencies or relationships between other entities or relationships which are there in the particular system. This is also has an important property known as conditional independence and it states that for a particular node in bayesian network, it is conditional independent of all of its non-descendants if its parent is known.

**Support Vector Machines (SVM):** SVM is a frontier which segregates two classes in a hyper-plane [93]. The selection of the hyper-plane is done by finding the

maximum distance between the data instances of all the classes combined thereby providing the ability to identify future classes with more accuracy. Data instances that are closer to this identified hyper-plane is called the support vectors that maximises the possibility of classifier margin.

**C4.5 decision tree:** This is based on the Iterative Dichotomiser 3 (ID3) algorithm that is used to generate decision trees from a given dataset [94]. This is also known as a statistical classifier as it is used for classification.

# Chapter 5

# Experiment

This chapter explains both the logical physical structure used to conduct the experiments. Data exploration methods are exercised on the datasets to get a better understanding of it. This involves both statistical analysis and representing data visually using various visualization patterns. Substantive results of the experiments are catalogued as an outcome of classification and selection of required features. Following is the different types of experiments performed:

1. Apply machine learning models to detect attacks in UMass dataset.
2. Anomaly analysis of UMass dataset.
3. Apply machine learning models to detect attacks in NTNU S-building dataset for both weekdays and weekend.
4. Data analysis to detect attacks in NTNU S-building dataset for both weekdays and weekend.
5. Anomaly analysis of NTNU S-building dataset for both weekdays and weekend.

## 5.1 Experimental environment

This section details the environment used in conducting the experiments along with a description the flow of data starting with collecting the raw samples and till the point where they are classified by different methods.

The main phases of the logical environment are data pre-processing and classification. Once the raw data is gathered, it is further pre-processed to remove any missing values and then balanced as mentioned in Section 4.1.2. Gathering the data is done in well defined regular time intervals and stored as comma separated flat files. Data cleansing is executed to increase the quality and reliability and further reduce any noise or inconsistencies if present. This is followed by a process of normalizing the data to avert any feature influence if they have large values. The normalization process transforms the raw data feature values between 0 and 1. Section 5.2 shows that feature selection is applied to the data set to remove redundant features and select required features for more analysis. Afterwards classi-

fication is done with algorithms as shown in Section 4.4 and models are evaluated on different metrics detailed in Section 5.3.

Relevant experiment results are shown and explained in this section and extended experiment results are displayed in Appendix B.

### 5.1.1 Physical environment

The resources used for computation and analysis in the thesis are catalogued in this section. Due to the large size of data and need to store everything on disk, the setup needs to have disc space corresponding to the size of the data. In order to process the dataset (raw datasets are usually large in size), enough memory and CPU is also quite essential. This enables to give results within acceptable time limits. A virtual machine hosted in Microsoft cloud services Azure with the following specifications was used for this purpose:

- Virtual Machine
- Windows 10 Operating System
- 500 GB disk
- 32 GB RAM
- 16 Core virtual CPU

The selection of features and some classification algorithms were conducted on personal laptop with the following specifications:

- Lenovo ThinkPad W540
- Windows 10 Operating System
- 500 GB SSD disk
- 32 GB RAM
- 4 Core CPU Intel i7-4900MQ

### 5.1.2 Logical environment

The following lists shows the major applications along with the version number used for executing the experiments.

- Weka 3.8.3
- Microsoft Excel 2016
- EmEditor 19.7.0

## 5.2 Feature selection

Selection of the features is very important in the training data set as it influences the performance of the classification of the dataset and also on how the information is depicted for dataset classification. Fewer features can often contribute to better performance in classification [34]. The accuracy of classification is also improved as a result of good feature selection strategies by choosing the required subset of features.

This thesis experiments on two datasets as listed in Section 4.1 and the multi-objective feature selection technique namely NSGA-II which was previously published in [95] is applied on NTNU S-building dataset. The aforementioned method is an improved version of the standard NSGA-II [96] which improves the diversity of solutions by removing the redundant solutions offered by NSGA-II. This helped to reduce the number of features from 50 to 26 this dataset.

## 5.3 Evaluation of Models

There here different machine learning methods mentioned in Section 4.4 and metrics are used in the evaluation of datasets. As described in Section 2.3.3 true positive, false positive rate, precision and F-measure are used to evaluate the different classifier performances. Table 5.1 shows the results obtained from the NTNU S-building data set using 26 attributes selected by, algorithm described in Section 5.2. All the classification algorithms are run with a split of 66% on training set and remainder with test set.

The table shows that all the learning methods have a true positive rate between 0.997 and 0.998. The false positive rate is equivalently quite low as 0.001 whereas the precision is in the range of 0.99. It also confirms this with ROC area close to 1. This is quite satisfactory and shows that the feature selection algorithm have selected the right features that is used by the classifiers.

| **Learning method** | **Attributes** | **TP Rate** | **FP Rate** | **Precision** | **ROC Area** | **F-Measure** |
|---|---|---|---|---|---|---|
| SVM | 26 | 0.997 | 0.001 | 0.997 | 0.998 | 0.997 |
| Bayesian | | 0.997 | 0.001 | 0.997 | 1 | 0.997 |
| Random Forest | | 0.998 | 0 | 0.998 | 1 | 0.998 |
| C4.5 | | 0.998 | 0 | 0.998 | 0.999 | 0.998 |

**Table 5.1:** Classifier Evaluation with 26 Features

## 5.4 Data exploration

In the dataset used for the experiments based on the NTNU S-building, there where 4464 data rows in total out of which 3528 data rows are before corona has occurred that led to the closing of the university. This is as shown in Table 5.2.

| **Dataset** | **Number of instances** | **4464** |
|---|---|---|
| NTNU S-building | Number of Normal Classes | 3528 |
| | Number of Attack Classes | 936 |

**Table 5.2:** Overview of NTNU S-building data set instances

However this dataset is highly unbalanced. Unbalanced datasets are a common problem in classification where the classes are not equally represented. This is especially demanding in case of enormous datasets as machine learning algorithms aim to determine anomalous situations from these datasets. This imbalance of classes leads the machine learning models to classifying the datasets incorrectly which in turn results in erroneous model accuracy. As this imbalance is a common problem identified in most cases this has been expected. As for the NTNU S-building data set, the imbalance is due to the fact that occurrence of corona is a recent event and data for is available only from March, 2020 onwards.

There are several methods to mitigate the unbalanced data set. Some of them are as follows :

- Gathering more data.
- Changing performance metrics
- Re-sampling datasets
- Generating synthetic samples
- Using penalized models

The method used in this thesis is to re-sample the data set. Re-sampling can be done in two different ways namely:

1. Removing instances from the class that have more items. This is known as under-sampling.
2. Adding copies of instances from the class that has less instances. This is known as over-sampling.

The experiment used the first method to remove some datasets from the majority instance classes that is before corona in the NTNU S-building data set. After the balancing the data set, the instances are selected from 25.01.2020 to 20.04.2020. This is further categorized into weekdays and weekends to see how the different selected feature values change.

## 5.5 Results of Experiments

Section 5.3 shows the comparison on different classification methods and it can be seen that the accuracy is in the region of 99% for all the learning methods. It be concluded from Table 5.1 that there is a change from before corona and during corona time. The algorithms classify both the classes as labelled correctly which can also be deduced from the false positive rate (in the range of 0.001% for most of the classifiers), true positive rate and ROC area.

As specified in Section 5.4, the datasets have been balanced and further categorized into weekday and weekend data. The important features as mentioned in Section 4.1.1, selected out of the 26 are further analysed. Following sub-sections display the results of these analysis as histograms. The x-axis shows the measured value for the selected feature and y-axis shows the frequency of occurrence. The blue color indicates the period before corona and red during corona time. These

analysis have been done to show the differences before and during corona.

### 5.5.1 Analysis of NTNU S-Building dataset for weekdays

Following are the results of data analysis from weekdays as shown by Figure 5.1, Figure 5.2, Figure 5.3 and Figure 5.4. This is done to see if there are attacks in NTNU S-building dataset for weekdays. It is further discussed in Chapter 6.

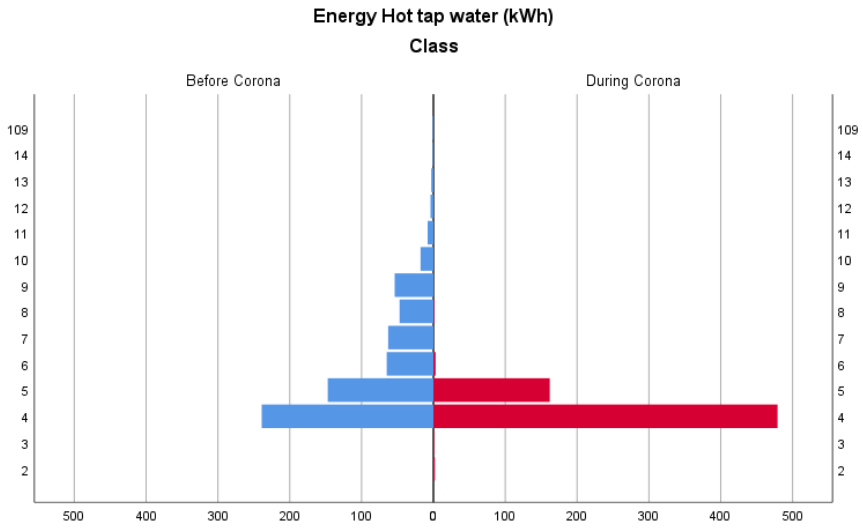



**Figure 5.1:** Energy Hot tap water

### 5.5.2 Analysis of NTNU S-Building dataset for weekend

Following are the results of data analysis from weekend as shown by Figure 5.5, Figure 5.6, Figure 5.7 and Figure 5.8. This is done to see if there are attacks in NTNU S-building dataset for weekend. It is further discussed in Chapter 6.

### 5.5.3 Anomaly analysis on NTNU S-building dataset

The features identified in Section 4.1.1 are further analysed for anomalies in the dataset to investigate for attacks and following are the results. From these, anomaly analysis have been conducted for ventilation, power and energy data that is specified in Section 4.2.3 and Section 4.2.4. This has been analysed again for weekdays and weekends and following are the results:

**Weekday:** Following graphs shows the results from weekdays analysis of anomalies as shown by Figure 5.9, Figure 5.10 and Figure 5.11.

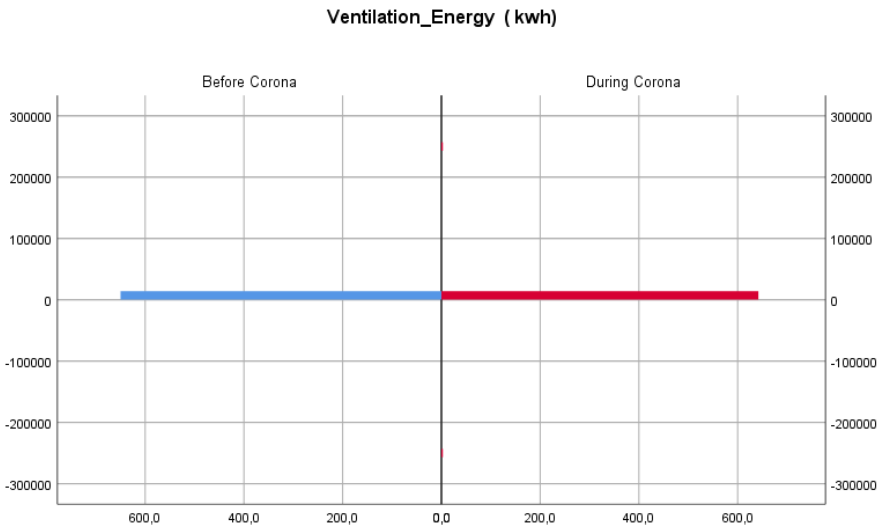



**Figure 5.2:** Ventilation Energy

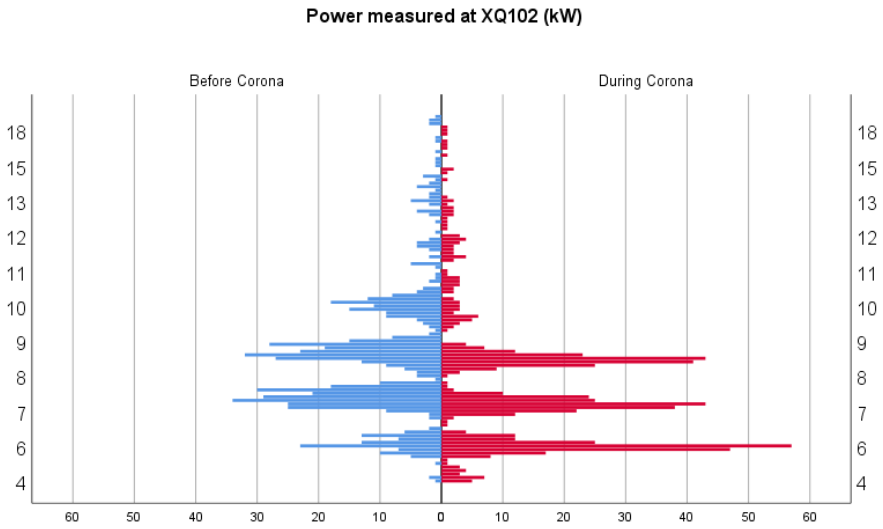



**Figure 5.3:** Power

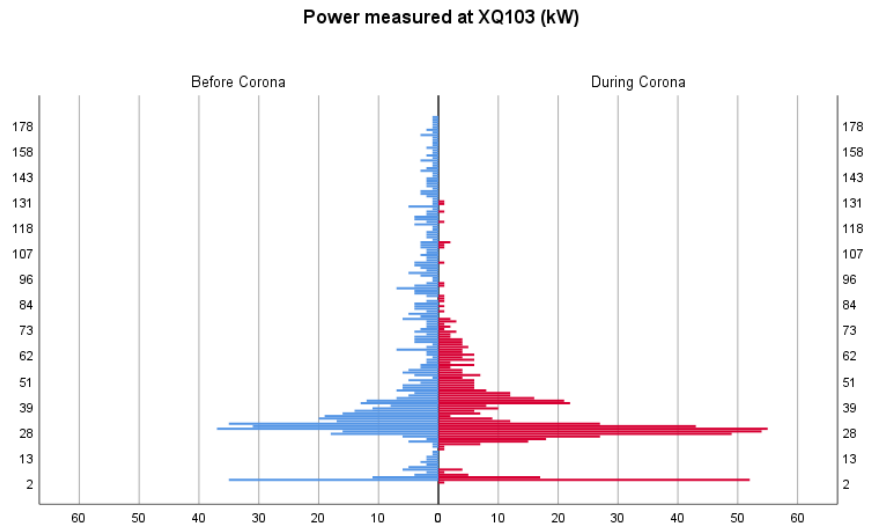



**Figure 5.4:** Energy

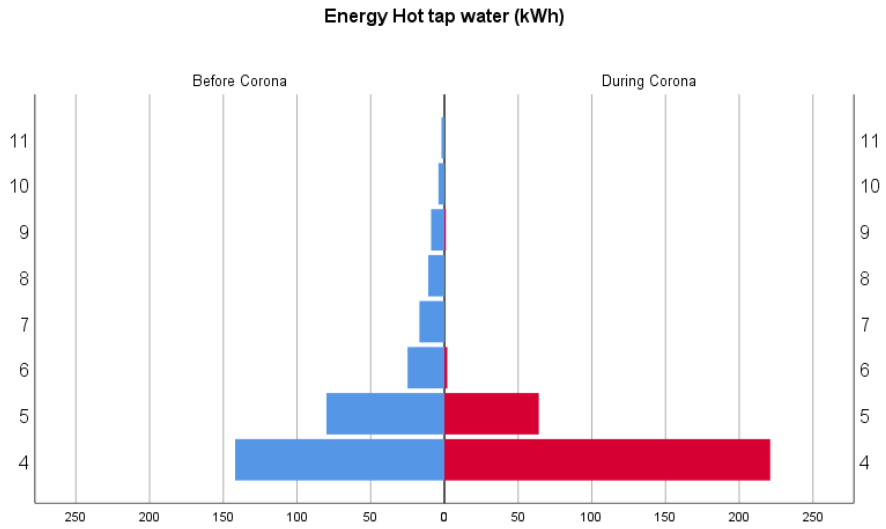



**Figure 5.5:** Energy Hot tap water

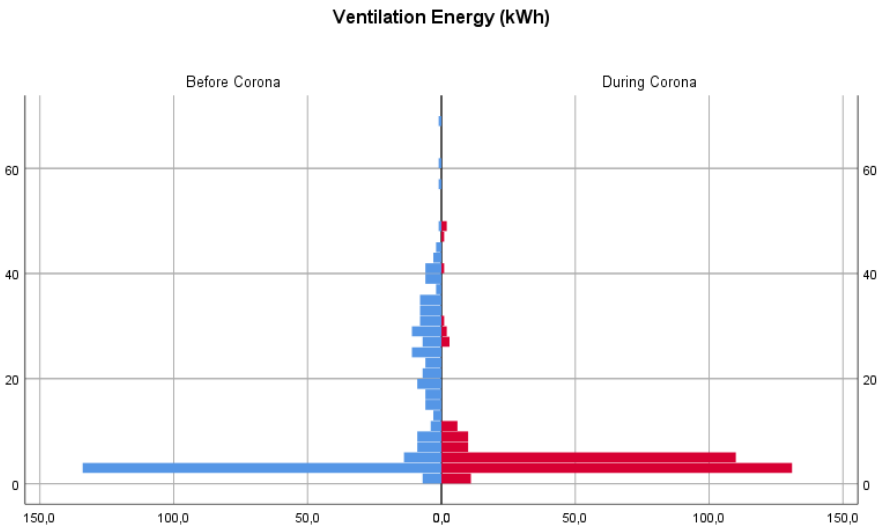



**Figure 5.6:** Ventilation Energy

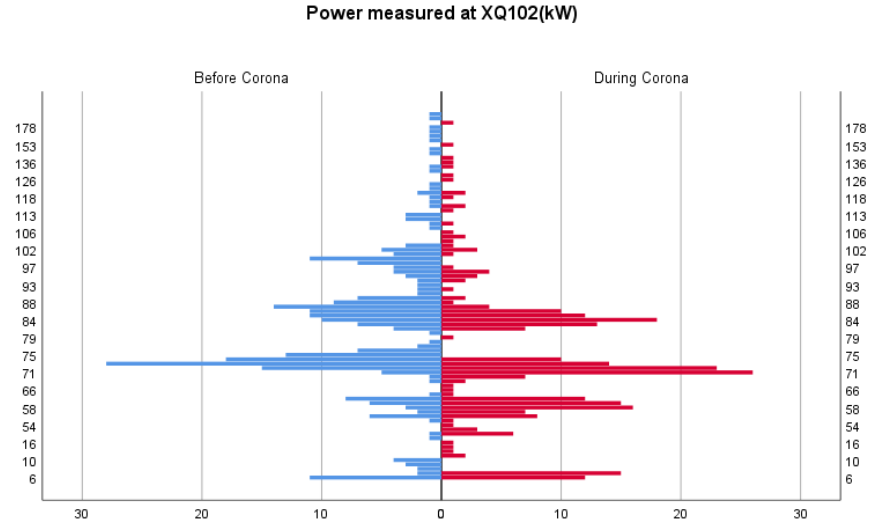



**Figure 5.7:** Power

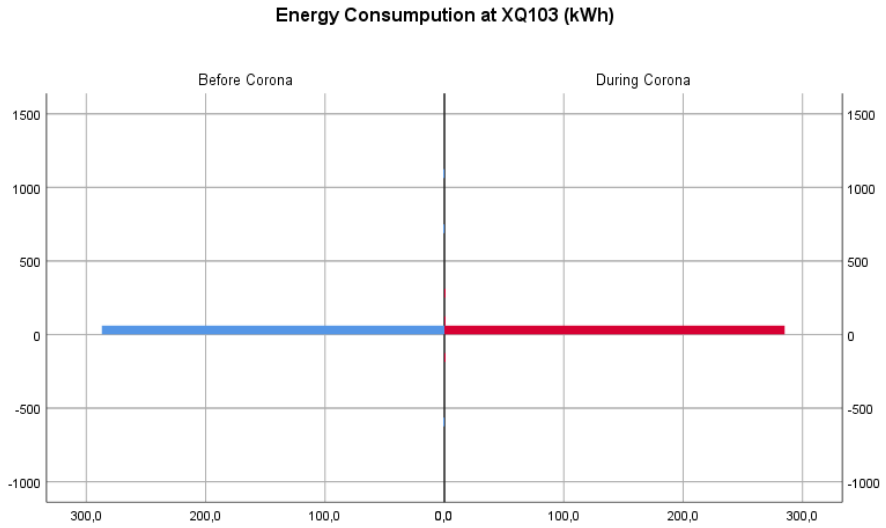



**Figure 5.8:** Energy

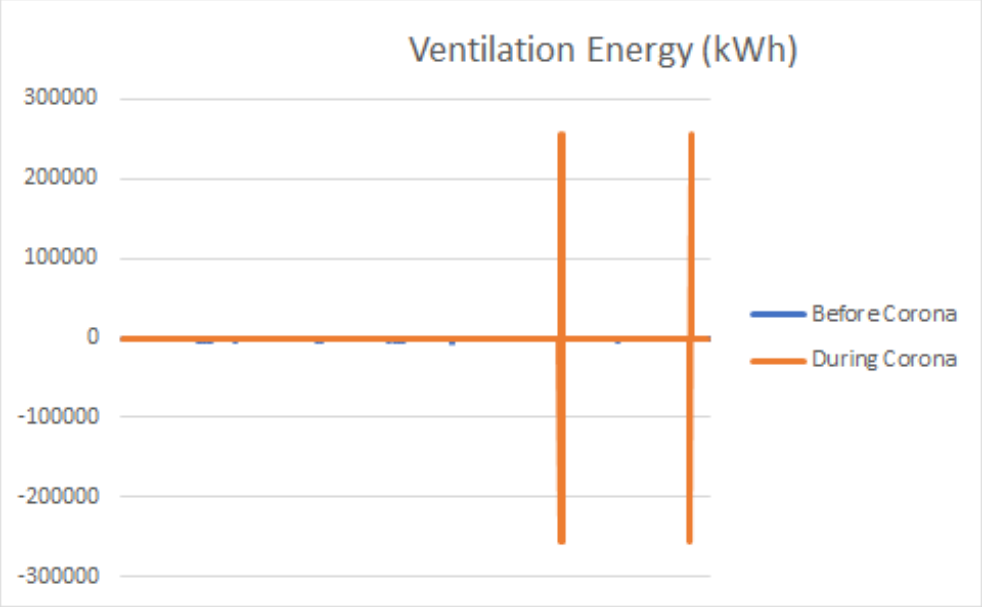



**Figure 5.9:** Ventilation

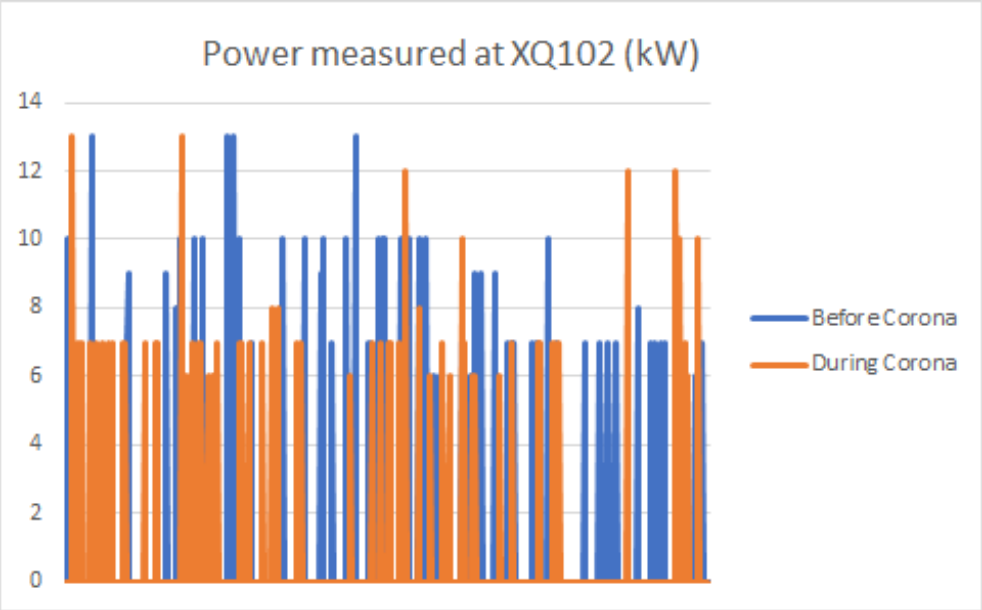



**Figure 5.10:** Power

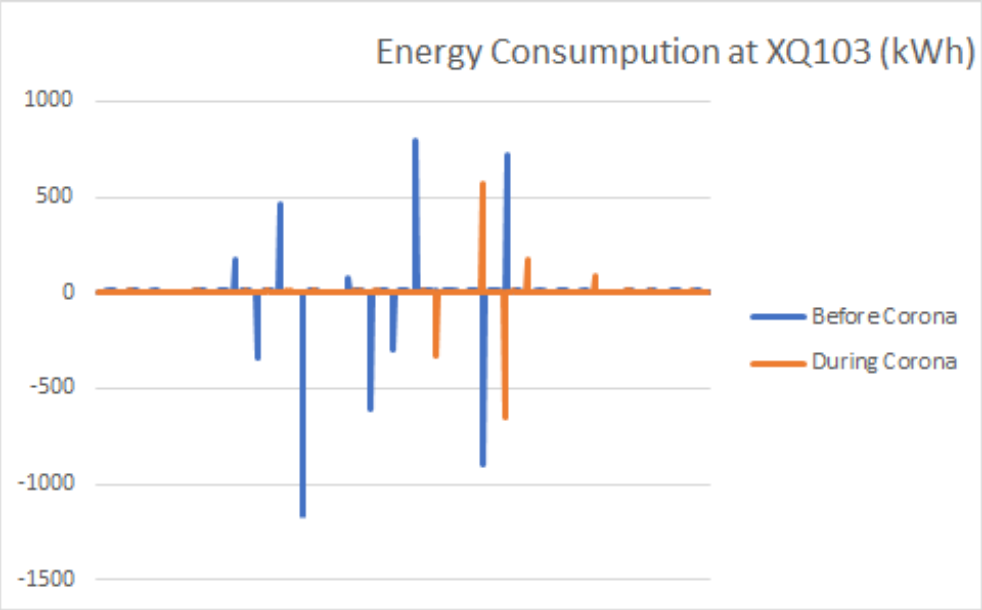



**Figure 5.11:** Energy

**Weekend:** Following graphs shows the results from weekend analysis of anomalies as shown by Figure 5.12, Figure 5.13 and Figure 5.14.

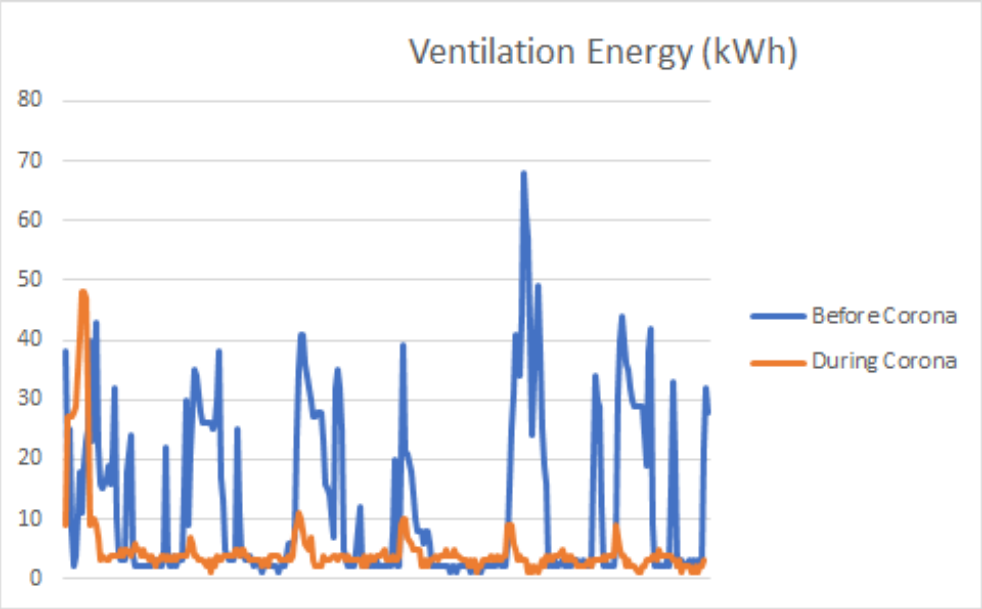



**Figure 5.12:** Ventilation

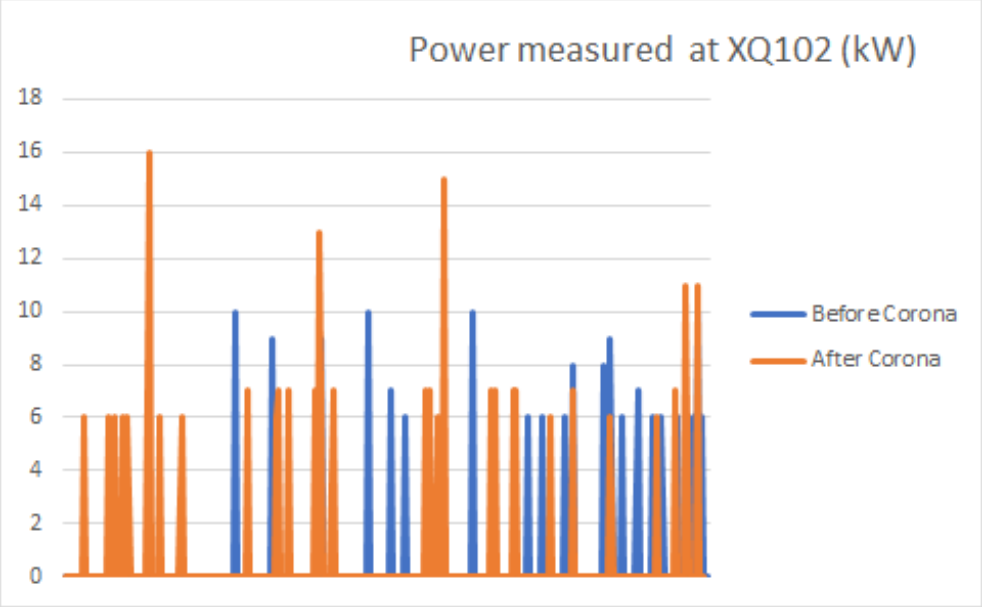



**Figure 5.13:** Power

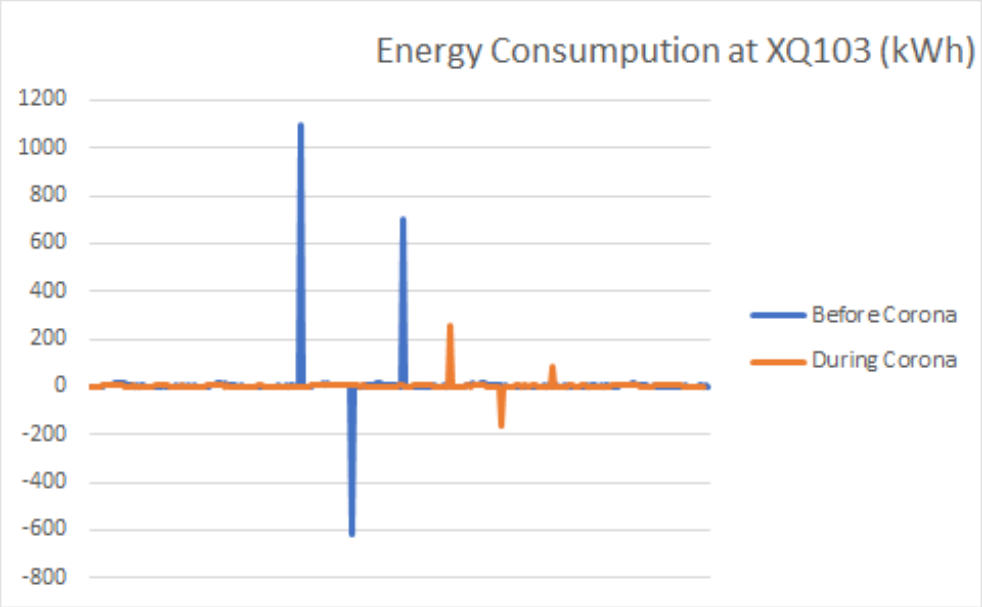



**Figure 5.14:** Energy

### 5.5.4 Unsupervised Machine Learning on UMass dataset

The UMass is a publicly available dataset and is not labelled. Hence it is not possible to run supervised machine learning algorithms on this dataset. This has led to the use of unsupervised machine learning methods on this dataset as mentioned in Section 2.3. Here the cluster algorithm simple K means was run on the Home A dataset for data from meters 2, 3 and 4. A sample of the relevant results is shown in Table 5.3 and rest in Appendix B. All the cluster analysis algorithms are run with k set to 2, a split of 66% on training set and remainder with test set.

| Attribute | Full Data | Cluster 0 | Cluster 1 |
|---|---|---|---|
| Furnace HRV | 0.1579 | 0.0631 | 0.5964 |
| Cellar Outlets | 0.0807 | 0.0798 | 0.0852 |
| WashingMachine | 0.0011 | 0.0009 | 0.0024 |
| FridgeRange | 0.0056 | 0.0052 | 0.0076 |
| DisposalDishWasher | 0.0018 | 0.0017 | 0.0023 |
| KitchenLights | 0.013 | 0.0043 | 0.053 |
| BedroomOutlets | 0.0111 | 0.0102 | 0.0155 |
| BedroomLights | 0.0129 | 0.0064 | 0.0426 |
| MasterOutlets | 0.0297 | 0.0113 | 0.115 |
| MasterLights | 0.0319 | 0.0148 | 0.1113 |
| DuctHeaterHRV | 0.0362 | 0.0109 | 0.1529 |

**Table 5.3:** Simple K means on HomeA-meter 2

It can be seen from Table 5.4 that 82% has been classified in cluster 0 and rest 18% in cluster 1.

| | Attributes | Percentage |
|---|---|---|
| Cluster 0 | 78808 | 82 |
| Cluster 1 | 16963 | 18 |

**Table 5.4:** Cluster Results

### 5.5.5 Anomaly analysis on UMass dataset

The Section 5.5.4 shows the dataset categorized into two different clusters. However to look into whether there has been some anomalies in the dataset, two features namely the Washing Machine and Fridge are selected from HomeA and meter 2 are analysed. Following are the results from these as shown by Figure 5.15 and Figure 5.16.

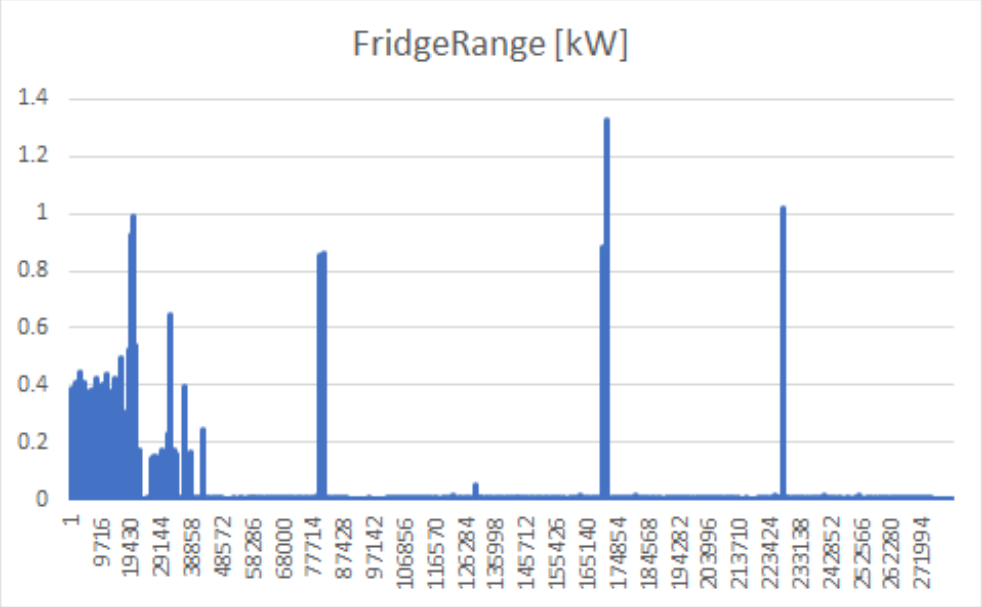



**Figure 5.15:** HomeA-Meter2-Fridge data

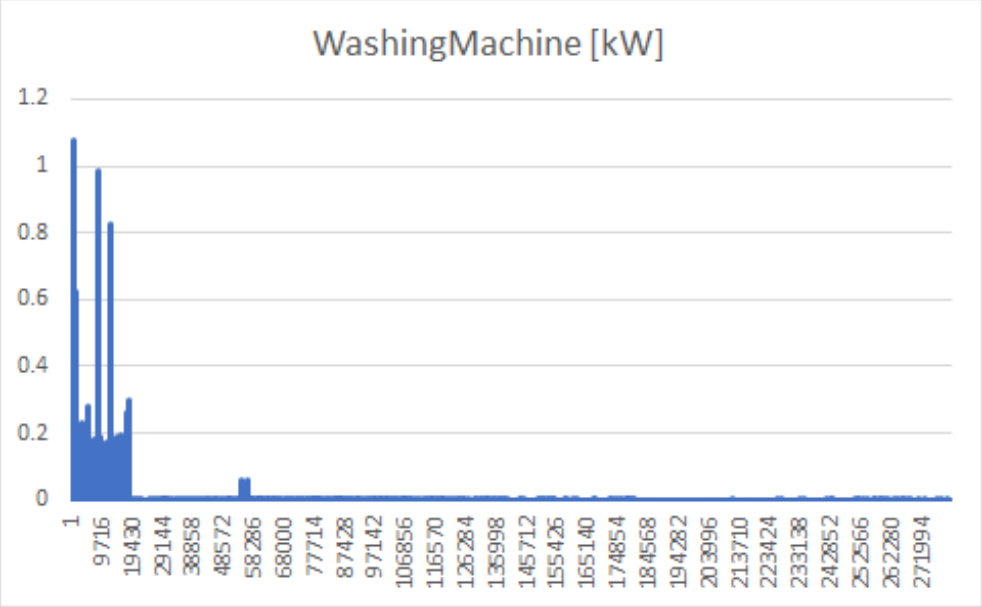



**Figure 5.16:** HomeA-Meter2-Washing Machine data

# Chapter 6

# Discussion

The research and experiments as detailed in Chapter 5 preformed in the thesis answer the research questions stated in Section 1.5.

*Research Question 1: How machine learning models can detect attacks in smart grids embedded with IoT?*

Section 5.3 shows the different machine learning classifiers that are used to identify the classes before and during corona for the NTNU S-building dataset. It can been seen that all of the models have true positive rate between 0.997 and 0.998 while having really low false positive rate and ROC area close to 1. The models classify the classes correctly as before corona or during corona and this shows that there has been a change and machine learning models are able to detect this change.

Section 5.5.4 shows the different unsupervised learning algorithms run on UMass dataset and that it categorize them into two different clusters. However we cannot conclude that there has been an attack on the system from both these. For the NTNU S-building dataset these, changes are expected during corona time. To detect anomaly/attack, the cumulative distribution must be fine grained as the attack data can be hidden in somewhere in the cumulative distribution. Each sampled consequent difference should be checked for spikes.

The results from Section 5.5.1 concludes that are visible differences before corona and during corona during weekdays. For example from Figure 5.1 or Figure 5.3, it can be seen that the consumption have changed from before corona as compared to during corona. The similar applies for the analysis from Section 5.5.2 where the different parameters are compared in weekends from before corona to during corona time. The energy usage for hot tap water (Figure 5.5) or ventilation energy (Figure 5.6) shows the consumption have been changed.

For both the above cases, it can be seen that the consumption is lower. The dataset that has been analyzed is the frequency distribution of the differences for all the features that is measured in KWh. This is due to the fact that for each hour when the value is measured, it is the cumulative sum from the previous value. Analysing on the actual value without taking the difference will not give

the correct data values required for analysis. We cannot conclude yet that there has been an attack as the lowering of consumption is expected during corona time. In order to do that, further analysis has been done to fine grain the results to detect anomalies. Anomaly analysis for NTNU S-building is conducted on ventilation, power and energy as shown in Section 5.5.3. The spikes in the figures show that there are anomalies present. This has been confirmed one of the professors in the Department of Manufacturing and Civil Engineering at NTNU (also one of my co-supervisor) who is an expert in the area. Similar anomaly analysis is also conducted on the UMass dataset for the features fridge and washing machine. As seen in the figures in Section 5.5.5, the spikes shows that there are anomalies present in the dataset.

From the above it can be concluded that the spikes present in the anomaly analysis may or may not be attacks. The ventilation, energy and power has been decreased as expected in Scenario 3 and Scenario 4 that is detailed in Section 4.2.3 and Section 4.2.4 respectively. An attack on these scenarios can be considered as attack on availability as mentioned in Section 4.3.2. These spikes can be examined further as part of future work to see there are real attacks or expected increase/-decrease in the consumption. According to experiment results, since there is a consumption reduction, it can also be seen that people are following the regulations as detailed in Section 4.2.1. Labelled data is needed in order to make machine learning model to learn different attack types and when an old/new/similar pattern of values or spikes are presented to the machine learning model, then it will be able to classify between different attack types. There is no such labelled data present and hence it was not possible to train machine learning models for detecting/classifying attacks.

*Research Question 1a: How does the behavior change in a Smart Grid as opposed to a traditional grid and if IoT is connected to a smart grid, how behavior of Smart grid changes? So, is anything connected to the system a threat?*

As explained in the literature review in Section 2.1 and Section 2.1.1, it can been seen that there are several differences in Smart Grid as opposed to a traditional grid. It is shown that a Smart Grid integrates modern sensor, measurement, communication, information, computing and control technologies. The information flow is bidirectional and it enables active participation by all the parties involved. It also provides quality power for the digital economy, optimizes asset utilization, anticipates and respond to attacks. Connection of IoT to the Smart Grid can cause various security threats as discussed in Section 2.2.1. Attacks can be targeted to affect the NIST security criterion as shown in Section 3.4. These can affect the confidentiality, integrity and availability of the system. All these three should be working together to maintain the security of the system. As shown in Figure 3.4, attack can happen in four different phases.

The discussions in Research Question 1 shows that for NTNU S-building dataset there has been a change in consumption and anomalies have been detected from before and during corona time. As with the UMass dataset there has been

detected anomalies as shown in Section 5.5.5.

It can be concluded that there are changes in behaviour between a Smart Grid as opposed to a traditional grid. These certain threats are associated with connecting an IoT device to a smart grid and attacks can happen as described in Chapter 3. Here we are able to detect the anomalies/outliers. On categorising whether it is an attack or not requires more research and considered as part of future work.

*Research Question 2: What are consequences if attack not detected?* The consequences of not detecting these attacks can vary according to the scenario. For example in the case of Scenario 3 mentioned here in which the ventilation in NTNU S-building is affected. It can be seen from previous research question discussions that there are detected some anomalies in the ventilation data. This data could mean that either these consumption changes are scheduled/expected or it can be an attack. In the case of an attack to the ventilation to the S-building either before corona time or during corona time, the consequences can vary. This is an attack on availability, i.e the ventilation service is interrupted.

Ventilation is basically used to circulate the air in a building or room by distributing it from outside to inside and vice-versa. It thereby provides clean air by dissolving the pollutants that is in the room or building and removing it [97]. Ventilation can be natural, mechanical or hybrid. Natural ventilation is using natural forces to control the air flow in the building by opening windows for example. Mechanical ventilation is the use of fans that are installed inside the building to control the air flow. If these are controlled by attacks it can cause serious health damages, in case a person is inside the building or enters the building when this attack happens. During the corona time, this can be even worse and may cause life threatening situations. As stated in [98], corona is highly transmissible and can be deadly if ventilation is manipulated.

There are other consequences of not detecting these attacks like causing power outages to the building or the country itself that cause a critical infrastructure damage that will have other repercussions.

*Research Question 2a: How is privacy breached by using IoT in smart grids?*

This thesis introduces Smart Grids and IoT in Section 2.1 and Section 2.2 respectively. The integration of IoT in Smart Grids can cause several privacy and security issues. Literature survey in Section 3.2 identifies these issues and hence answers this research question. This is also part of a paper I was the lead author with title "IoT in Smart Grids: A Perspective from Security and Privacy and the Road Ahead" under review for publication [13].

## 6.1 Limitations

Obtaining the dataset and enough items are considered as one of main challenges of the thesis. Compared to other researches which uses publicly available datasets that have millions of rows in dataset, there was comparatively lesser number of data in the dataset that was used. Machine learning algorithms could give bet-

ter analysis with more data present in datasets. The reason is most of the Smart Grid with integrated IoT contains sensitive and private data, obtaining this from energy companies was an issue. The publicly available datasets in this field does not contain enough schematics to provide conclusive evidence to support the research questions as seen from Chapter 5. However, it was possible to gather real dataset from NTNU S-building which was sufficient to yield dependable results as described in Section 5.5. Since these are also sensitive information in the aforementioned dataset, it would limit the possibility of making this public and restrict other researchers conducting experiments or do further analysis.

# Chapter 7

# Conclusion

This thesis has focused on analyzing the differences between a traditional grid and a smart grid. It then looked into the effects on integrating IoT in smart grids and the subsequent privacy related issues. A detailed analysis of different machine learning algorithms and how these can be used to detect any attacks has been looked into. This has shown the algorithms classify the anomalies correctly but also opened path for future work to look deeper and find the attacks in each scenario. It was also shown that data pre-processing and feature selection algorithms also play an important part in the different machine learning algorithms. Although there were difficulties in getting the required dataset as mentioned in the Section 6.1, this research managed to obtain real data. The importance of obtaining the right dataset understood in this research.

This work is a contribution to the field in security aspects IoT enabled smart grids and how machine learning can be used to identify it. It was started in relation to the CINELDI (Center for Intelligent Electricity Distribution) project and the field is still in advancement. Although many challenges are remaining, using machine learning to detect attacks such anomalies in IoT enabled Smart Grids is a promising solution. Therefore this research can be extended as proposed in the future work in Chapter 8.

# Chapter 8

# Future Work

The discussions and results have resulted in diverse areas that are engaging and have paved way for different investigation areas. Some of the topics can also contribute to future research directions. These are as follows:

**Obtaining data from energy companies:** Data collection have been a challenge throughout this thesis and future work can be improved by gathering data from energy companies. This can provide more detailed analysis on the different data types, the types of attack that can happen, train the machine learning algorithms even better to predict different types of attack that might occur and mitigate them when it happens.

**Analysis on live data:** This thesis conducted experiments on recorded dataset and a publicly available dataset. Although this provides training to the machine learning models, a live dataset can be of even more advantage. This will allow the models to predict attacks in real time and mitigate them immediately. A faster response time will be beneficial for businesses and society in terms of damage limitations to the systems and limit further issues.

**False data injection:** Scenario 2 as mentioned in Section 4.2.2 can be considered for analysing the experiments further analysis as mentioned the discussion of Research Questions 1 and 1a. This can be used to detect if analysed anomaly and consumption change is an attack or not. This could not be conducted as part of the existing thesis as the other areas of research where quite demanding and vast areas to research and the limited time.

**Using deep learning:** Use of deep learning was one of the aspects that was initially implored in the starting phase of the experiments. However the long execution time and limitation of existing hardware made this consider as a future work. It has the ability to process large number of features and process them efficiently.

**Simulations using Microgrid emulators:** Microgrids can be used to emulate the scenarios required and then generate the required data. There was a plan to

do this by travelling to the NTNU laboratory in Trondheim but was limited due the existing COVID-19 situation. The current issues faced while working on this thesis to gather data for the required scenarios can be mitigated by this.

**Examine Spikes and display attacks detected on screen:** Further examination of the analysed spikes can be conducted and can be displayed with whether it is an attack or not. This can lead to mitigation steps that can be done either after the attack has happened or prevent a situation where the same attack might happen in the future.

# Bibliography


[1] D. Abraham, 'Application of machine learning in iot enabled smart grids for attack detection', Department of Information Security, Communication Technology, NTNU – Norwegian University of Science and Technology, Project report in IMT4205, Dec. 2019.

[2] Y. Saleem, N. Crespi, M. H. Rehmani and R. Copeland, 'Internet of things-aided smart grid: Technologies, architectures, applications, prototypes, and future research directions', *IEEE Access*, vol. PP, Apr. 2019. DOI: 10.1109/ACCESS.2019.2913984.

[3] X. Yu and Y. Xue, 'Smart grids: A cyber–physical systems perspective', *Proceedings of the IEEE*, vol. 104, no. 5, pp. 1058–1070, May 2016, ISSN: 1558-2256. DOI: 10.1109/JPROC.2015.2503119.

[4] S. E. Collier, 'The emerging enernet: Convergence of the smart grid with the internet of things', *IEEE Industry Applications Magazine*, vol. 23, no. 2, pp. 12–16, Mar. 2017, ISSN: 1558-0598. DOI: 10.1109/MIAS.2016.2600737.

[5] F. Aloul, A. Al-Ali, R. Al-Dalky, M. Al-Mardini and W. El-Hajj, 'Smart grid security: Threats, vulnerabilities and solutions', *International Journal of Smart Grid and Clean Energy*, vol. 1, pp. 1–6, Sep. 2012. DOI: 10.12720/sgce.1.1.1-6.

[6] J. Bugeja, A. Jacobsson and P. Davidsson, 'On privacy and security challenges in smart connected homes', Aug. 2016, pp. 172–175. DOI: 10.1109/EISIC.2016.044.

[7] P. Chandel and T. Thakur, 'Smart meter data analysis for electricity theft detection using neural networks', *Advances in Science, Technology and Engineering Systems Journal*, vol. 4, Jan. 2019. DOI: 10.25046/aj040420.

[8] R. Anderson and S. Fuloria, 'Who controls the off switch?', Nov. 2010, pp. 96–101. DOI: 10.1109/SMARTGRID.2010.5622026.

[9] S. Sridhar, A. Hahn and M. Govindarasu, 'Cyber-physical system security for the electric power grid', *Proceedings of the IEEE*, vol. 100, pp. 210–224, Jan. 2012. DOI: 10.1109/JPROC.2011.2165269.

[10] P. McDaniel and S. Mclaughlin, 'Security and privacy challenges in the smart grid', *IEEE Security Privacy*, vol. 7, pp. 75–77, May 2009. DOI: 10.1109/MSP.2009.76.

[11] M. Ozay, I. Esnaola, F. T. Yarman Vural, S. R. Kulkarni and H. V. Poor, 'Machine learning methods for attack detection in the smart grid', *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 8, pp. 1773–1786, 2016.

[12] D. Abraham, 'Security and privacy of iot in smart grids', Department of Information Security, Communication Technology, NTNU – Norwegian University of Science and Technology, Project report in IMT4203, Jan. 2020.

[13] D. Abraham, S. Y. Yayilgan, F. Dalipi, M. Abomhara and A. Gebremedhin, 'Iot in smart grids: A perspective from security and privacy and the road ahead (under review)', in *The 10th International Conference on the Internet of Things, Malmo, Sweden*, 2020 October.

[14] A. Jain, 'Changes and challenges in smart grid towards smarter grid', Dec. 2016. DOI: 10.1109/ICEPES.2016.7915907.

[15] W. Meng, R. Ma and H. Chen, 'Smart grid neighborhood area networks: A survey', *IEEE Network*, vol. 28, no. 1, pp. 24–32, Jan. 2014, ISSN: 1558-156X. DOI: 10.1109/MNET.2014.6724103.

[16] Y. Cunjiang, Z. Huaxun and Z. Lei, 'Architecture design for smart grid', *Energy Procedia*, vol. 17, pp. 1524–1528, Dec. 2012. DOI: 10.1016/j.egypro.2012.02.276.

[17] A. Ramesh, P. Karthikeyan, S. Padmanaban, S. Balasubramanian and J. M. Guerrero, 'A bibliographical survey on software architectures for smart grid system', 2018.

[18] M. Sadiku, M. Tembely and S. Musa, 'Home area networks: A primer', *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 7, pp. 634–635, May 2017. DOI: 10.23956/ijarcsse/SV7I5/208.

[19] M. D. H. Abdullah, M. H. Zurina, Z. Zukarnain and M. A. Mohamed, 'Attacks, vulnerabilities and security requirements in smart metering networks', *KSII Transactions on Internet and Information Systems*, vol. 9, pp. 1493–1515, Apr. 2015. DOI: 10.3837/tiis.2015.04.013.

[20] F. Xia, L. T. Yang, L. Wang and A. Vinel, 'Internet of things', *International Journal of Communication Systems*, vol. 25, no. 9, pp. 1101–1102, 2012. DOI: 10.1002/dac.2417. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/dac.2417. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/dac.2417.

[21] Z. Ji, I. Ganchev and M. O'Droma, 'A generic iot architecture for smart cities', vol. 2014, Jan. 2014, pp. 196–199, ISBN: 978-1-84919-924-7. DOI: 10.1049/cp.2014.0684.

[22] J. T. F. T. I. I. W. Group, *Certification Accreditation of Federal Information Systems Volume I*. Scotts Valley, CA: CreateSpace, 2010, ISBN: 1453610022.

[23] M. Uma and G. Padmavathi, 'A survey on various cyber attacks and their classification', *I. J. Network Security*, vol. 15, pp. 390–396, 2013.

[24] R. Walton and W.-M. Limited, 'Balancing the insider and outsider threat', *Computer Fraud Security*, vol. 2006, no. 11, pp. 8–11, 2006, ISSN: 1361-3723. DOI: https://doi.org/10.1016/S1361-3723(06)70440-7. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1361372306704407.

[25] M. Abomhara and G. M. Køien, 'Security and privacy in the internet of things: Current status and open issues', in *2014 International Conference on Privacy and Security in Mobile Systems (PRISMS)*, May 2014, pp. 1–8. DOI: 10.1109/PRISMS.2014.6970594.

[26] P. Mahalle, B. Anggorojati, N. Prasad and R. Prasad, 'Identity authentication and capability based access control (iacac) for the internet of things', English, *Journal of Cyber Security and Mobility*, vol. 1, no. 4, pp. 309–348, Oct. 2012, ISSN: 2245-1439.

[27] A. Riahi Sfar, Y. Challal, E. Natalizio, Z. Chtourou and A. Bouabdallah, 'A systemic approach for iot security', May 2013, pp. 351–355, ISBN: 978-1-4799-0206-4. DOI: 10.1109/DCOSS.2013.78.

[28] N. Fantana, T. Riedel, J. Schlick, S. Ferber, J. Hupp, S. Miles, F. Michahelles and S. Svensson, 'Internet of things - converging technologies for smart environments and integrated ecosystems', in. Jan. 2013, pp. 153–204, ISBN: ISBN 978-87-92982-73-5 (print) ISBN 978-87-9282-96-4 (ebook).

[29] J. Manyika, M. Chui, J. W. P. Bisson, R. Dobbs, J. Bughin and D.Aharon, 'Unlocking the potential of the internet of things', *McKinsey Global Institute*, Jun. 2015.

[30] F. Dalipi and S. Y. Yayilgan, 'Security and privacy considerations for iot application on smart grids: Survey and research challenges', in *2016 IEEE 4th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)*, Aug. 2016, pp. 63–68. DOI: 10.1109/W-FiCloud.2016.28.

[31] C. Pautasso, E. Wilde and R. Alarcón, Eds., *REST: Advanced Research Topics and Practical Applications*. Springer, Dec. 2014, pp. 1–214, ISBN: 978-1-4614-9298-6. [Online]. Available: http://ws-rest.org/book/2/.

[32] A. L. Samuel, 'Some studies in machine learning using the game of checkers', *IBM J. Res. Dev.*, vol. 3, pp. 210–229, 1959.

[33] T. M. Mitchell, *Machine Learning*. McGraw-Hill, 1997, ISBN: 978-0-07-042807-2.

[34] H. Liu and H. Motoda, *Computational Methods of Feature Selection (Chapman Hall/Crc Data Mining and Knowledge Discovery Series)*. Chapman Hall/CRC, 2007, ISBN: 1584888784.

[35] A. Ghasempour, 'Internet of things in smart grid: Architecture, applications, services, key technologies, and challenges', *Inventions*, vol. 4, p. 22, Mar. 2019. DOI: 10.3390/inventions4010022.

[36] J. L. Mauri, J. Tomás, A. Canovas and L. Parra, 'An integrated iot architecture for smart metering', *IEEE Communications Magazine*, vol. 54, pp. 50–57, 2016.

[37] Q. Ou, Y. Zhen, X. Li, Y. Zhang and L. Zeng, 'Application of internet of things in smart grid power transmission', Jun. 2012, pp. 96–100, ISBN: 978-1-4673-1956-0. DOI: 10.1109/MUSIC.2012.24.

[38] S. K. Viswanath, C. Yuen, W. Tushar, W.-T. Li, C.-K. Wen, K. Hu, C. Chen and X. Liu, *System design of internet-of-things for residential smart grid*, 2016. arXiv: 1604.04009 [cs.NI].

[39] J. S. Winter, 'Citizen perspectives on the customization/privacy paradox related to smart meter implementation', *Int. J. Technoethics*, vol. 6, no. 1, pp. 45–59, Jan. 2015, ISSN: 1947-3451. DOI: 10.4018/ijt.2015010104. [Online]. Available: https://doi.org/10.4018/ijt.2015010104.

[40] H. Nissenbaum, 'Privacy in context: Technology, policy, and the integrity of social life', *Bibliovault OAI Repository, the University of Chicago Press*, Jan. 2010.

[41] M. A. Ferrag, L. Maglaras, A. Argyriou, D. Kosmanos and H. Janicke, 'Security for 4g and 5g cellular networks', *J. Netw. Comput. Appl.*, vol. 101, no. C, pp. 55–82, Jan. 2018, ISSN: 1084-8045. DOI: 10.1016/j.jnca.2017.10.017. [Online]. Available: https://doi.org/10.1016/j.jnca.2017.10.017.

[42] Z. Erkin, 'Private data aggregation with groups for smart grids in a dynamic setting using crt', in *2015 IEEE International Workshop on Information Forensics and Security (WIFS)*, Nov. 2015, pp. 1–6. DOI: 10.1109/WIFS.2015.7368584.

[43] M. A. Rahman, M. H. Manshaei, E. Al-Shaer and M. Shehab, 'Secure and private data aggregation for energy consumption scheduling in smart grids', *IEEE Transactions on Dependable and Secure Computing*, vol. 14, no. 2, pp. 221–234, Mar. 2017, ISSN: 2160-9209. DOI: 10.1109/TDSC.2015.2446492.

[44] N. Saputro, K. Akkaya and I. Guvenc, 'Privacy-aware communication protocol for hybrid ieee 802.11s/lte smart grid architectures', Oct. 2015. DOI: 10.1109/LCNW.2015.7365945.

[45] J. Chen, J. Shi and Y. Zhang, 'Eppdc: An efficient privacy-preserving scheme for data collection in smart grid', *Int. J. Distrib. Sen. Netw.*, vol. 2015, Jan. 2015, ISSN: 1550-1329. DOI: 10.1155/2015/656219. [Online]. Available: https://doi.org/10.1155/2015/656219.

[46] I. Doh, J. Lim and K. Chae, 'Secure authentication for structured smart grid system.', in *IMIS*, L. Barolli, F. Palmieri, H. D. S. Silva and H.-C. Chen, Eds., IEEE, 2015, pp. 200–204, ISBN: 978-1-4799-8873-0. [Online]. Available: http://dblp.uni-trier.de/db/conf/imis/imis2015.html#DohLC15.

[47] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam and E. Cayirci, 'A survey on sensor networks', *Comm. Mag.*, vol. 40, no. 8, pp. 102–114, Aug. 2002, ISSN: 0163-6804. DOI: 10.1109/MCOM.2002.1024422. [Online]. Available: https://doi.org/10.1109/MCOM.2002.1024422.

[48] Haowen Chan and A. Perrig, 'Security and privacy in sensor networks', *Computer*, vol. 36, no. 10, pp. 103–105, Oct. 2003, ISSN: 1558-0814. DOI: 10.1109/MC.2003.1236475.

[49] J. Yick, B. Mukherjee and D. Ghosal, 'Wireless sensor network survey', *Computer Networks*, vol. 52, pp. 2292–2330, Aug. 2008. DOI: 10.1016/j.comnet.2008.04.002.

[50] N. Li, N. Zhang, S. Das and B. Thuraisingham, 'Privacy preservation in wireless sensor networks: A state-of-the-art survey', English (US), *Ad Hoc Networks*, vol. 7, no. 8, pp. 1501–1514, Nov. 2009, ISSN: 1570-8705. DOI: 10.1016/j.adhoc.2009.04.009.

[51] L. Zheng, S. Chen, S. Xiang and Y. Hu, 'Research of architecture and application of internet of things for smart grid', Aug. 2012, pp. 938–941, ISBN: 978-1-4673-0721-5. DOI: 10.1109/CSSS.2012.238.

[52] G. Piro, G. Boggia and L. A. Grieco, 'A standard compliant security framework for ieee 802.15.4 networks', in *Proc. of IEEE World Forum on Internet of Things (WF-IoT)*, Seoul, South Korea, Mar. 2014. eprint: https://telematics.poliba.it/publications/2014/piro-wf-iot-2014.pdf.

[53] I. Alqassem and D. Svetinovic, 'A taxonomy of security and privacy requirements for the internet of things (iot)', *IEEE International Conference on Industrial Engineering and Engineering Management*, vol. 2015, pp. 1244–1248, Mar. 2015. DOI: 10.1109/IEEM.2014.7058837.

[54] Y. F. Wang, W. M. Lin, T. Zhang and Y. Y. Ma, 'Research on application and security protection of internet of things in smart grid', in *IET International Conference on Information Science and Control Engineering 2012 (ICISCE 2012)*, Dec. 2012, pp. 1–5. DOI: 10.1049/cp.2012.2311.

[55] S. E. Collier, 'The emerging enernet: Convergence of the smart grid with the internet of things', in *2015 IEEE Rural Electric Power Conference*, Apr. 2015, pp. 65–68. DOI: 10.1109/REPC.2015.24.

[56] Y. E. Song, Y. Liu, S. Fang and S. Zhang, 'Research on applications of the internet of things in the smart grid', in *2015 7th International Conference on Intelligent Human-Machine Systems and Cybernetics*, vol. 2, Aug. 2015, pp. 178–181. DOI: 10.1109/IHMSC.2015.131.

[57] J. Noll, I. Garitano, S. Fayyad, E. Åsberg and H. Abie, 'Measurable security, privacy and dependability in smart grids', *Journal of Cyber Security and Mobility*, vol. 3, pp. 371–398, Feb. 2015. DOI: 10.13052/jcsm2245-1439.342.

[58] S. Sicari, A. Rizzardi, L. Grieco and A. Coen-Porisini, 'Security, privacy and trust in internet of things: The road ahead', *Computer Networks*, vol. 76, Jan. 2015. DOI: 10.1016/j.comnet.2014.11.008.

[59] J. Lee, W. Lin and Y. Huang, 'A lightweight authentication protocol for internet of things', in *2014 International Symposium on Next-Generation Electronics (ISNE)*, May 2014, pp. 1–2. DOI: 10.1109/ISNE.2014.6839375.

[60] M. Turkanović, B. Brumen and M. Hölbl, 'A novel user authentication and key agreement scheme for heterogeneous ad hoc wireless sensor networks, based on the internet of things notion', *Ad Hoc Networks*, vol. 20, pp. 96–112, Apr. 2014. DOI: 10.1016/j.adhoc.2014.03.009.

[61] N. YE, Y. Zhu, R.-c. WANG, R. Malekian and L. Qiao-min, 'An efficient authentication and access control scheme for perception layer of internet of things', *Applied Mathematics  Information Sciences*, vol. 8, Jul. 2014. DOI: 10.12785/amis/080416.

[62] F. Borges, F. Volk and M. Mühlhäuser, 'Efficient, verifiable, secure, and privacy-friendly computations for the smart grid', in *2015 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, Feb. 2015, pp. 1–5. DOI: 10.1109/ISGT.2015.7131862.

[63] D. Abraham, 'Survey on anomaly detection for unknown presentation attack detection (adp)', Department of Information Security, Communication Technology, NTNU – Norwegian University of Science and Technology, Project report in IMT4126, May 2020.

[64] F. E.Grubbs, 'Procedures for detecting outlying observations in samples', *Technometrics*, vol. 11, no. 1, pp. 1–21, 1969. DOI: 10.1080/00401706.1969.10490657. eprint: https://www.tandfonline.com/doi/pdf/10.1080/00401706.1969.10490657. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/00401706.1969.10490657.

[65] H. S. Teng, K. Chen and S. C. Lu, 'Adaptive real-time anomaly detection using inductively generated sequential patterns', in *Proceedings. 1990 IEEE Computer Society Symposium on Research in Security and Privacy*, 1990, pp. 278–284.

[66] B. Scholkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2001, ISBN: 0262194759.

[67] K. Mehrotra, C. K. Mohan and S. Ranka, *Elements of Artificial Neural Networks*. Cambridge, MA, USA: MIT Press, 1996, ISBN: 0262133288.

[68] M. M. Moya and D. R. Hush, 'Network constraints and multi-objective optimization for one-class classification', *Neural Netw.*, vol. 9, no. 3, pp. 463–474, Apr. 1996, ISSN: 0893-6080. DOI: 10.1016/0893-6080(95)00120-4. [Online]. Available: https://doi.org/10.1016/0893-6080(95)00120-4.

[69] B. Schölkopf, J. C. Platt, J. C. Shawe-Taylor, A. J. Smola and R. C. Williamson, 'Estimating the support of a high-dimensional distribution', *Neural Comput.*, vol. 13, no. 7, pp. 1443–1471, Jul. 2001, ISSN: 0899-7667. DOI: 10.1162/089976601750264965. [Online]. Available: https://doi.org/10.1162/089976601750264965.

[70] S. Hawkins, H. He, G. J. Williams and R. A. Baxter, 'Outlier detection using replicator neural networks', in *Proceedings of the 4th International Conference on Data Warehousing and Knowledge Discovery*, ser. DaWaK 2000, Berlin, Heidelberg: Springer-Verlag, 2002, pp. 170–180, ISBN: 3540441239.

[71] M. Goldstein and S. Uchida, 'A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data', *PLOS ONE*, vol. 11, no. 4, pp. 1–31, Apr. 2016. DOI: 10.1371/journal.pone.0152173. [Online]. Available: https://doi.org/10.1371/journal.pone.0152173.

[72] V. Chandola, A. Banerjee and V. Kumar, 'Anomaly detection: A survey', *ACM Comput. Surv.*, vol. 41, no. 3, Jul. 2009, ISSN: 0360-0300. DOI: 10.1145/1541880.1541882. [Online]. Available: https://doi.org/10.1145/1541880.1541882.

[73] D.-Y. Yeung and Y. Ding, 'Host-based intrusion detection using dynamic and static behavioral models', *Pattern Recognition*, vol. 36, pp. 229–243, Jan. 2003. DOI: 10.1016/S0031-3203(02)00026-2.

[74] J. Sigholm and M. Raciti, 'Best-effort data leakage prevention in inter-organizational tactical manets', Oct. 2012, pp. 1–7, ISBN: 978-1-4673-1729-0. DOI: 10.1109/MILCOM.2012.6415755.

[75] J. Lin, E. Keogh, A. Fu and H. van herle, 'Approximations to magic: Finding unusual medical time series', Jul. 2005, pp. 329–334, ISBN: 0-7695-2355-2. DOI: 10.1109/CBMS.2005.34.

[76] E. Schubert, A. Koos, T. Emrich, A. Züfle, K. A. Schmid and A. Zimek, 'A framework for clustering uncertain data', *Proc. VLDB Endow.*, vol. 8, no. 12, pp. 1976–1979, Aug. 2015, ISSN: 2150-8097. DOI: 10.14778/2824032.2824115. [Online]. Available: https://doi.org/10.14778/2824032.2824115.

[77] M. Amer and M. Goldstein, 'Nearest-neighbor and clustering based anomaly detection algorithms for rapidminer', Aug. 2012. DOI: 10.5455/ijavms.141.

[78] S. Ramaswamy, R. Rastogi and K. Shim, 'Efficient algorithms for mining outliers from large data sets.', vol. 29, Jun. 2000, pp. 427–438. DOI: 10.1145/335191.335437.

[79] M. Breunig, H.-P. Kriegel, R. Ng and J. Sander, 'Lof: Identifying density-based local outliers.', vol. 29, Jun. 2000, pp. 93–104. DOI: 10.1145/342009.335388.

[80] J. Tang, Z. Chen, A. Fu and D. Cheung, 'Enhancing effectiveness of outlier detections for low density patterns', May 2002, pp. 535–548. DOI: 10.1007/3-540-47887-6_53.

[81] R. Kwitt and U. Hofmann, 'Unsupervised anomaly detection in network traffic by means of robust pca', *Computing in the Global Information Technology, International Multi-Conference on*, vol. 0, p. 37, Mar. 2007. DOI: 10.1109/ICCGI.2007.62.

[82] F. T. Liu, K. M. Ting and Z.-H. Zhou, 'Isolation forest', in *Proceedings of the 2008 Eighth IEEE International Conference on Data Mining*, ser. ICDM '08, USA: IEEE Computer Society, 2008, pp. 413–422, ISBN: 9780769535029. DOI: 10.1109/ICDM.2008.17. [Online]. Available: https://doi.org/10.1109/ICDM.2008.17.

[83] M. Harvey, D. Long and K. Reinhard, 'Visualizing nistir 7628, guidelines for smart grid cyber security', in *2014 Power and Energy Conference at Illinois (PECI)*, 2014, pp. 1–8.

[84] Z. Elmrabet, N. Kaabouch, H. el ghazi and H. Elghazi, 'Cyber-security in smart grid: Survey and challenges', *Computers Electrical Engineering*, vol. 67, May 2018. DOI: 10.1016/j.compeleceng.2018.01.015.

[85] J. E. O. Paul D. Leedy, *Practical Research Planning and Design*. Pearson, 2015.

[86] NTNU, *National smart grid laboratory*, 2020. [Online]. Available: https://www.ntnu.edu/smartgrid.

[87] T. Weibel, *Umasstracerepository*. [Online]. Available: http://traces.cs.umass.edu/index.php/Smart/Smart.

[88] F. Dalipi, S. Yildirim Yayilgan and A. Gebremedhin, 'Data-driven machine-learning model in district heating system for heat load prediction', *Appl. Comp. Intell. Soft Comput.*, vol. 2016, p. 1, Jun. 2016, ISSN: 1687-9724. DOI: 10.1155/2016/3403150. [Online]. Available: https://doi.org/10.1155/2016/3403150.

[89] G. Sindre and A. L. Opdahl, 'Capturing security requirements through misuse cases', 2001.

[90] L. Breiman, 'Bagging predictors', *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, Aug. 1996, ISSN: 0885-6125. DOI: 10.1023/A:1018054314350. [Online]. Available: https://doi.org/10.1023/A:1018054314350.

[91] D. Solomatine and D. Shrestha, 'Adaboost.rt: A boosting algorithm for regression problems', vol. 2, Aug. 2004, 1163–1168 vol.2, ISBN: 0-7803-8359-1. DOI: 10.1109/IJCNN.2004.1380102.

[92] N. Friedman and M. Goldszmidt, 'Building classifiers using bayesian networks', in *Proceedings of the Thirteenth National Conference on Artificial Intelligence - Volume 2*, ser. AAAI'96, Portland, Oregon: AAAI Press, 1996, pp. 1277–1284, ISBN: 026251091X.

[93] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, 2000. DOI: 10.1017/CBO9780511801389.

[94] S. L. Salzberg, 'C4.5: Programs for machine learning by j. ross quinlan. morgan kaufmann publishers, inc., 1993', *Machine Learning*, vol. 16, no. 3, pp. 235–240, Sep. 1994, ISSN: 1573-0565. DOI: 10.1007/BF00993309. [Online]. Available: https://doi.org/10.1007/BF00993309.

[95] K. Deb, A. Pratap, S. Agarwal and T. Meyarivan, 'A fast and elitist multiobjective genetic algorithm: Nsga-ii', *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.

[96] A. Golrang, A. Golrang, S. Yildirim Yayilgan and O. Elezaj, 'A novel hybrid ids based on modified nsgaii-ann and random forest', *Electronics*, vol. 9, Mar. 2020. DOI: 10.3390/electronics9040577.

[97] D. W. Etheridge and M. Sandberg, 'Building ventilation: Theory and measurement', 1996.

[98] S. Asadi, N. Bouvier, A. S. Wexler and W. D. Ristenpart, 'The coronavirus pandemic and aerosols: Does covid-19 transmit via expiratory particles?', *Aerosol Science and Technology*, vol. 54, no. 6, pp. 635–638, 2020. DOI: 10.1080/02786826.2020.1749229. eprint: https://doi.org/10.1080/02786826.2020.1749229. [Online]. Available: https://doi.org/10.1080/02786826.2020.1749229.

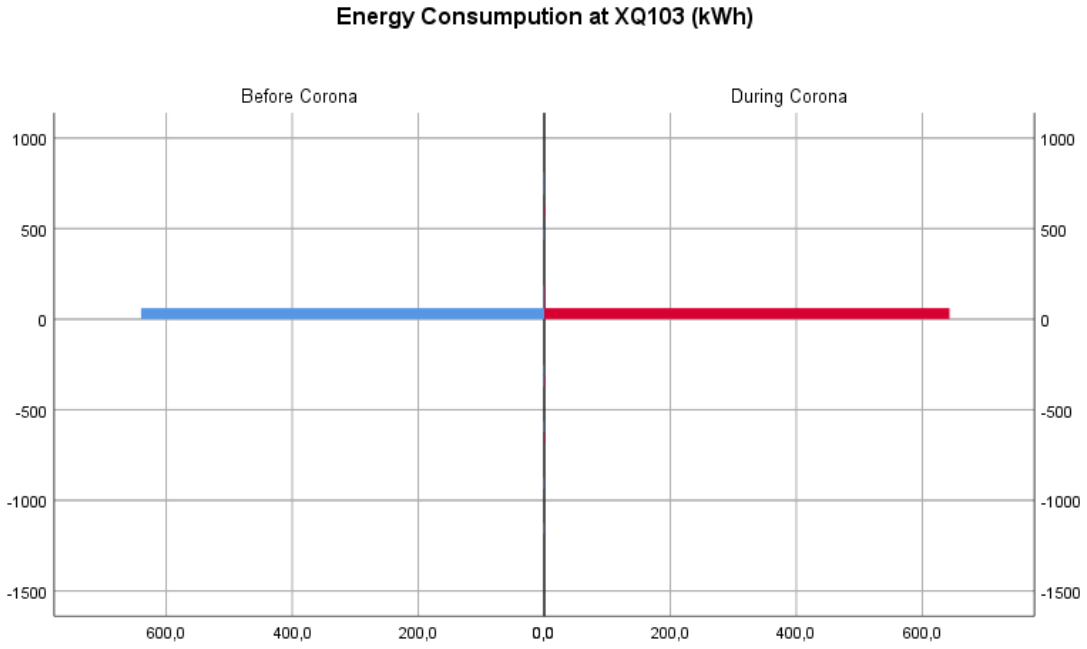
# Appendix A

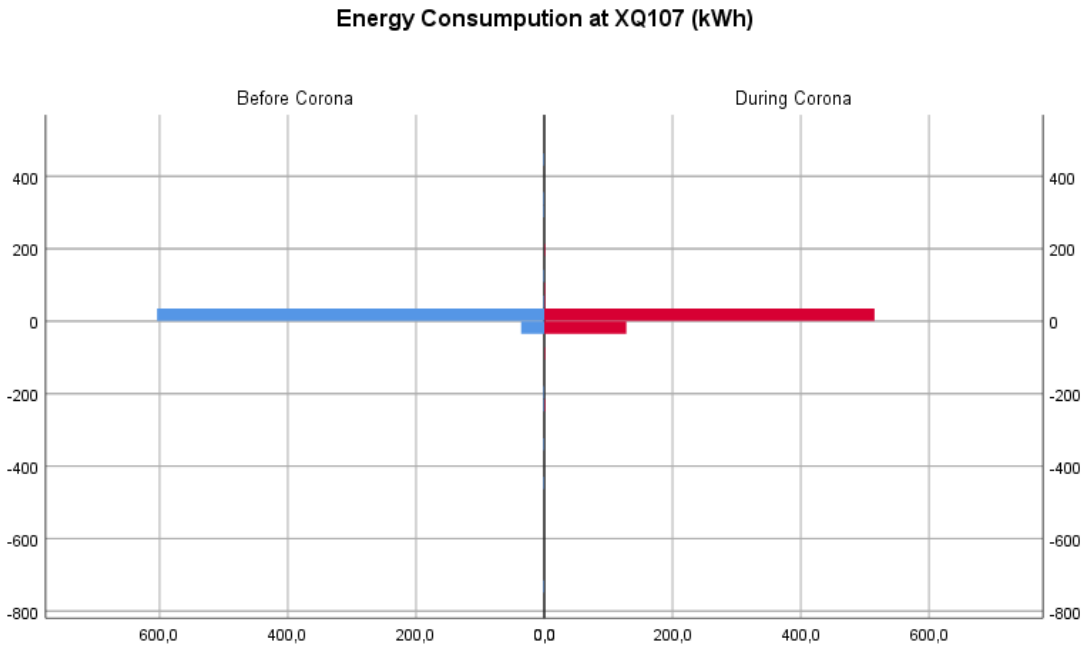
# NTNU S-building dataset analysis

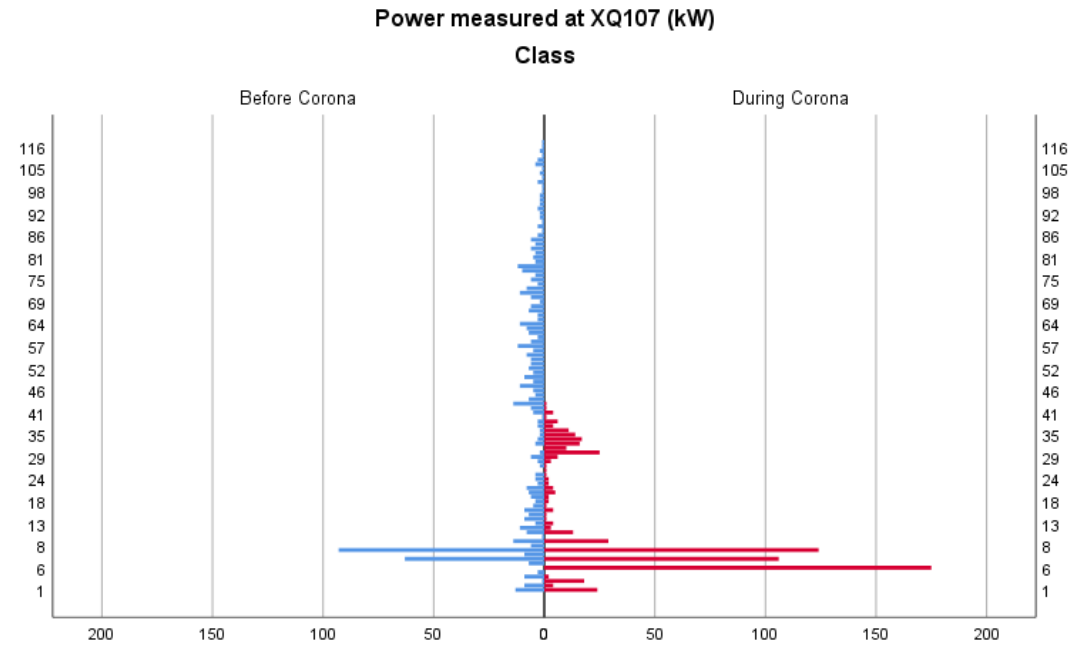
Following are rest of analysis results not included in Section 5.5.1, Section 5.5.2 and Section 5.5.3 but are still relevant. A sample dataset of some selected features is also shown here. Here it can be seen that the measurements are recorded hourly.

## Weekdays dataset analysis

Following are rest of analysis results not included in Section 5.5.1



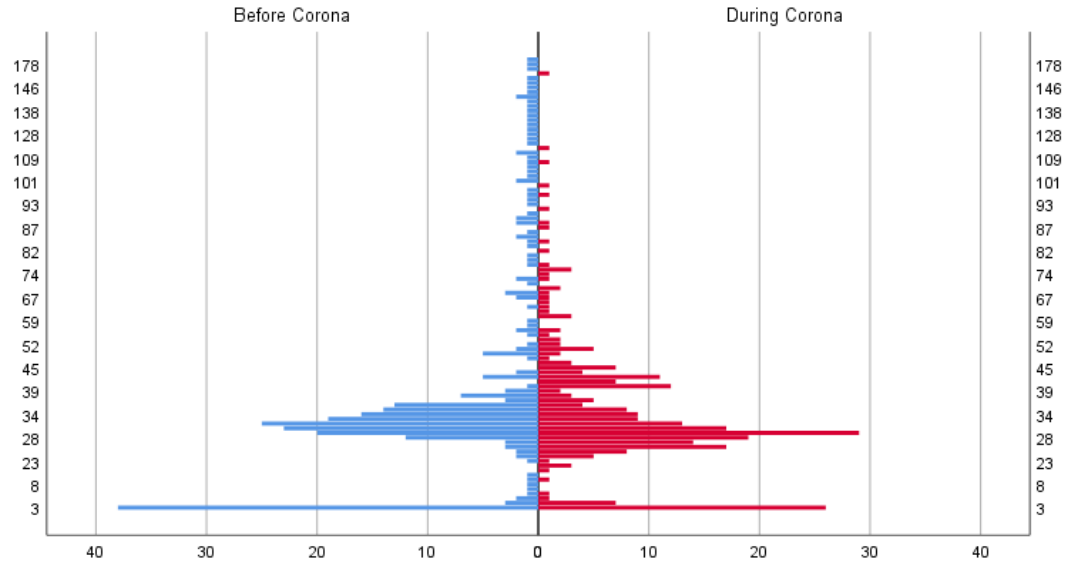


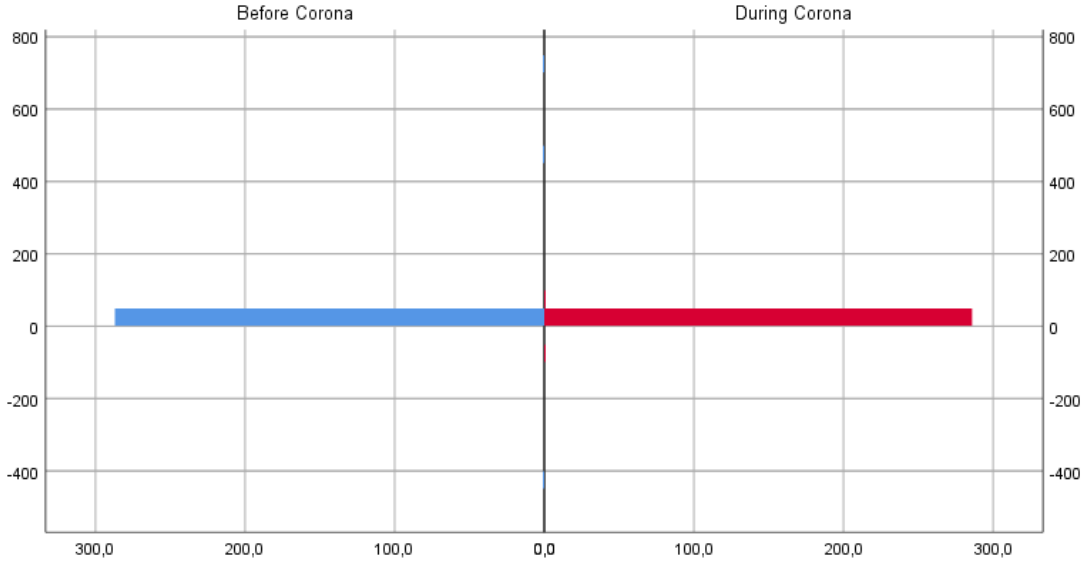

## Weekend dataset analysis

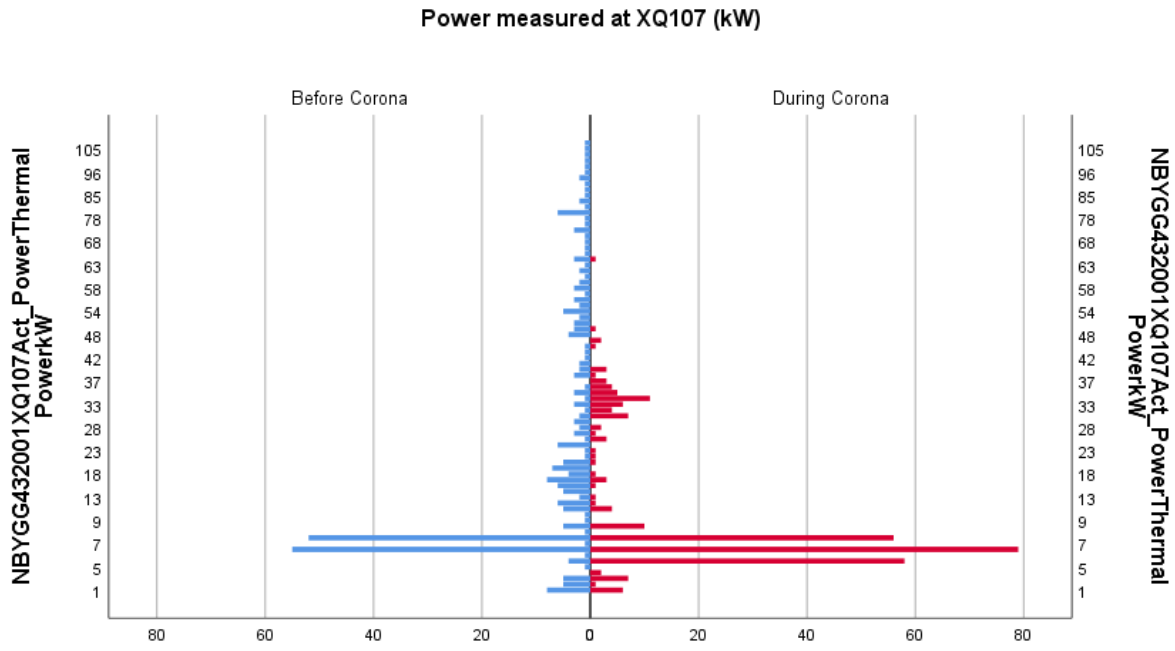
Following are rest of analysis results not included in Section 5.5.2

**Power measured at XQ103 (kW)**



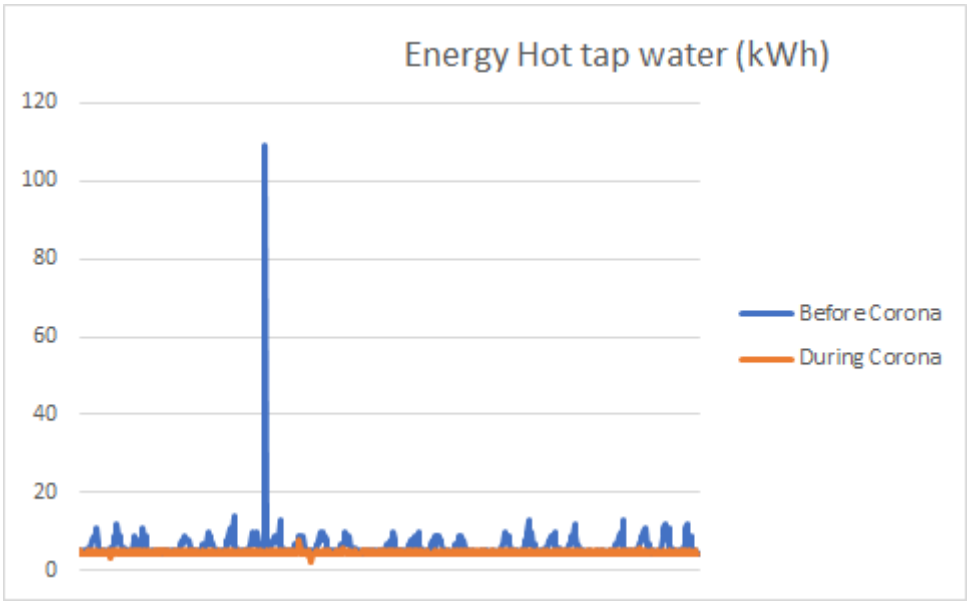

**Energy Consumption at XQ107 (kWh)**

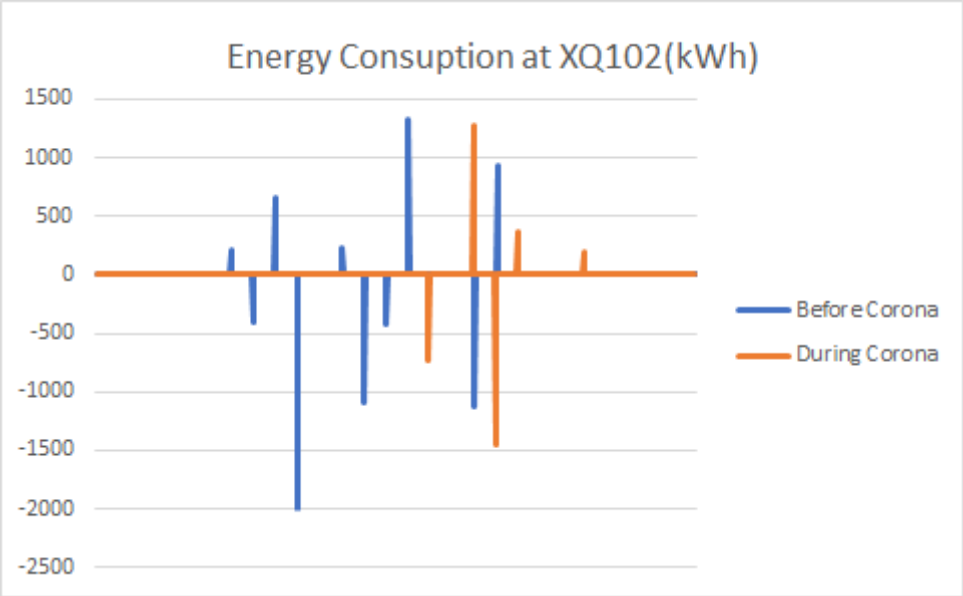


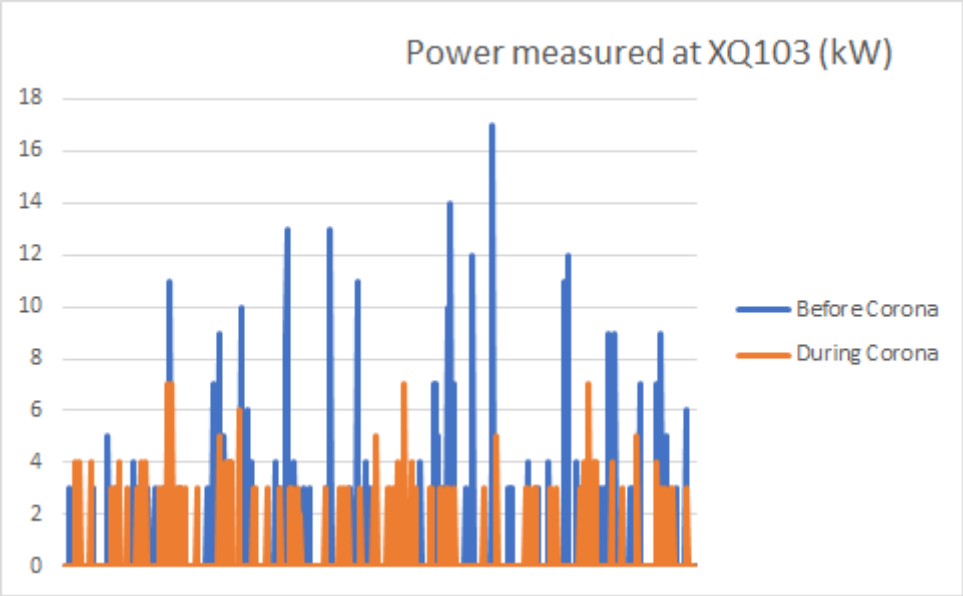


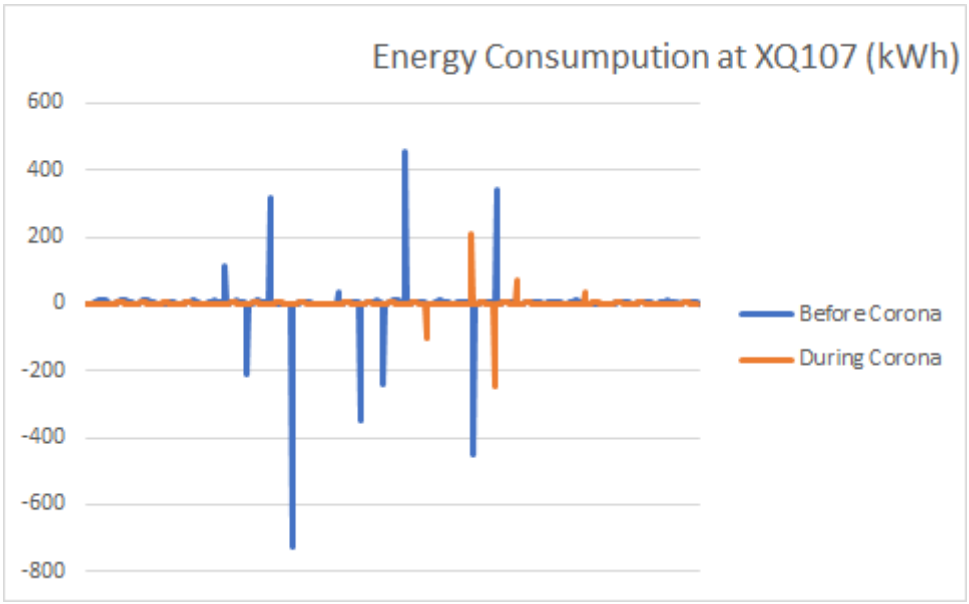

## Weekday dataset Anomaly analysis

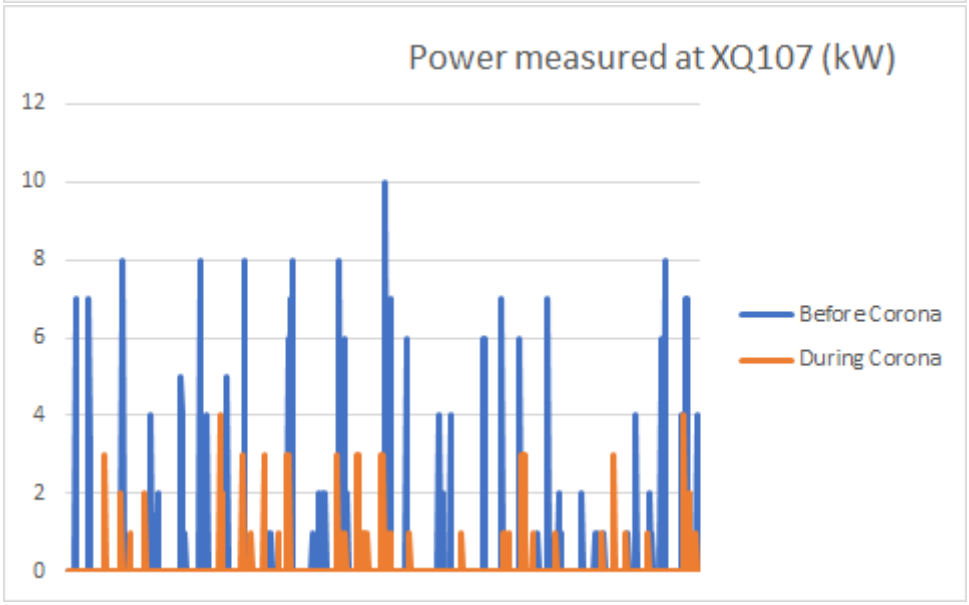
Following are rest of analysis results not included in Section 5.5.3




Energy Consuption at XQ102(kWh)
1500
1000
500
0
-500
-1000
-1500
-2000
-2500
Before Corona
During Corona

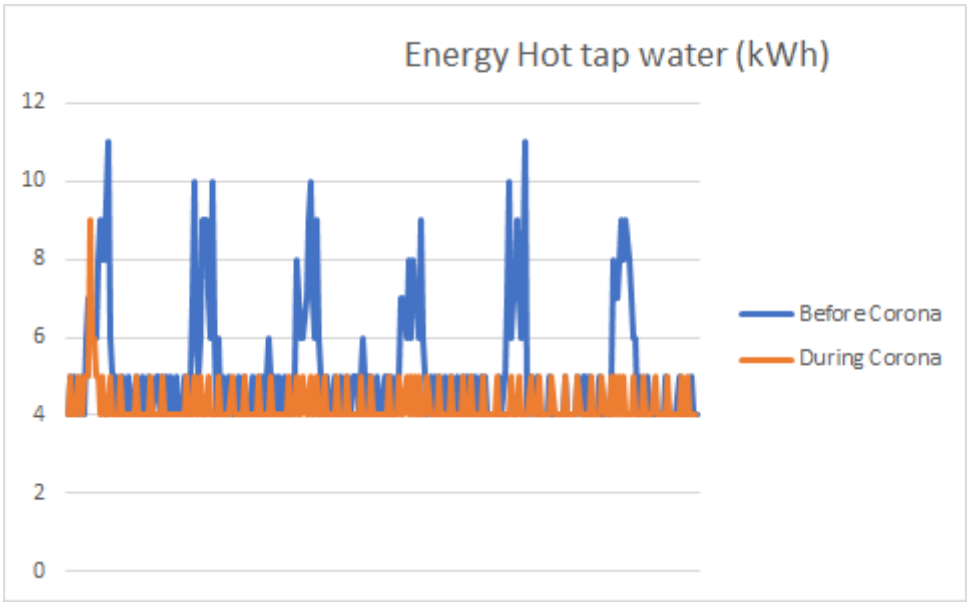

Power measured at XQ103 (kW)
18
16
14
12
10
8
6
4
2
0
Before Corona
During Corona

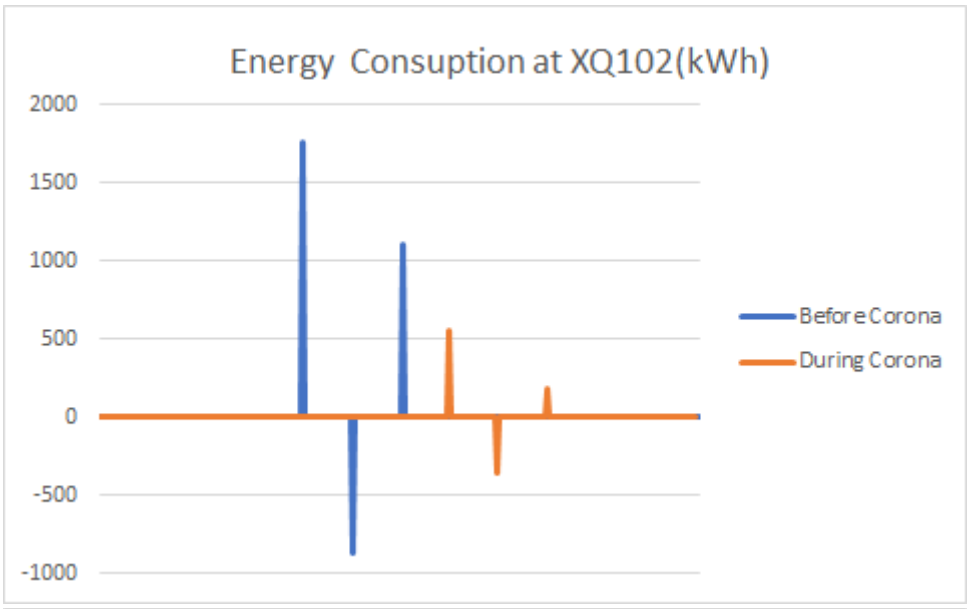


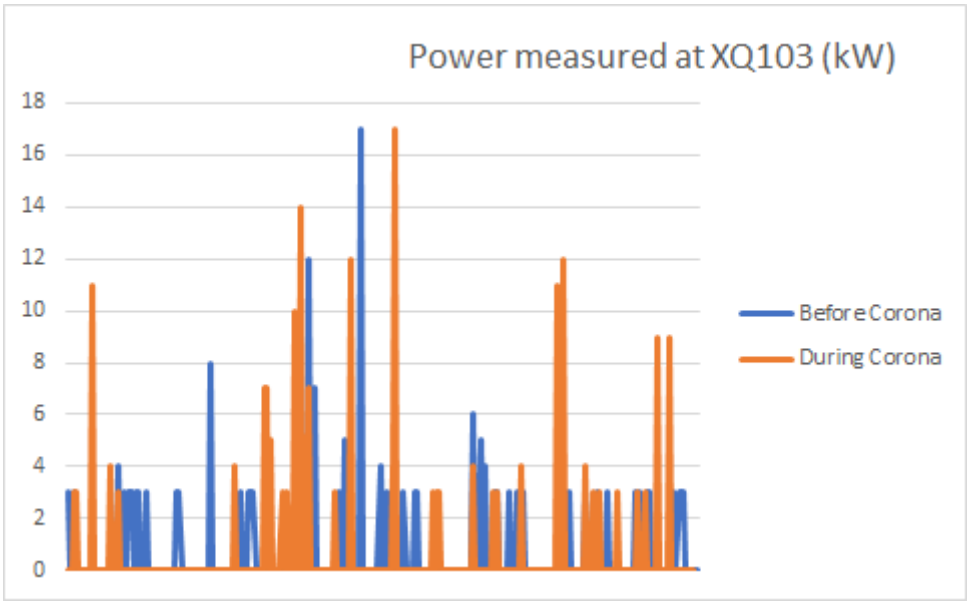



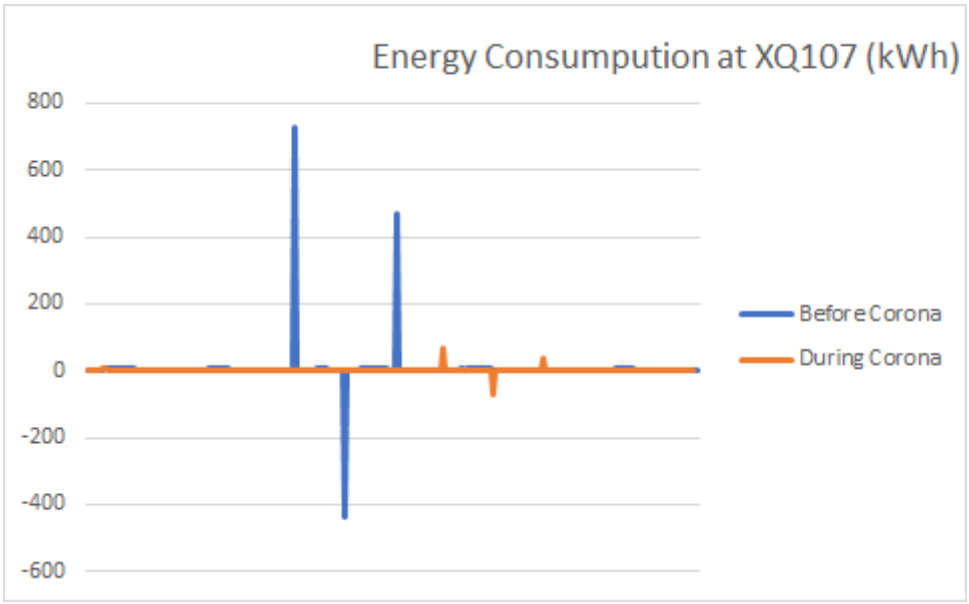

## Weekend dataset Anomaly analysis

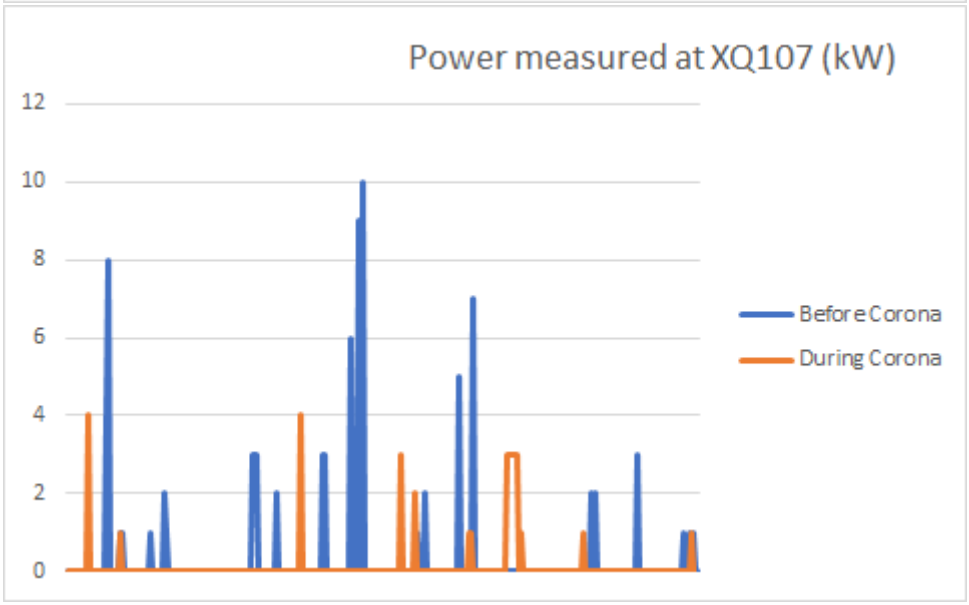
Following are rest of analysis results not included in Section 5.5.3


Energy Hot tap water (kWh)
12
10
8
6
4
2
0
Before Corona
During Corona


Energy Consuption at XQ102(kWh)
2000
1500
1000
500
0
-500
-1000
Before Corona
During Corona


Power measured at XQ103 (kW)
18
16
14
12
10
8
6
4
2
0
Before Corona
During Corona


Energy Consumption at XQ107 (kWh)
800
600
400
200
0
-200
-400
-600
Before Corona
During Corona


Power measured at XQ107 (kW)
12
10
8
6
4
2
0
Before Corona
During Corona

## Sample dataset of selected features

| DateTime | Hot tapwater Energy-kWh | Ventilation Energy-kWh | Power Consumption XQ102-kW | Power Consumption XQ103-kW |
|---|---|---|---|---|
| 16.10.2019 00:00 | 77 920,00 | 203 175,00 | 7,2 | 2,7 |
| 16.10.2019 01:00 | 77 923,00 | 203 178,00 | 9,7 | 2,8 |
| 16.10.2019 02:00 | 77 927,00 | 203 180,00 | 8,3 | 2,9 |
| 16.10.2019 03:00 | 77 931,00 | 203 181,00 | 7,1 | 2,5 |
| 16.10.2019 04:00 | 77 935,00 | 203 183,00 | 11,7 | 2,9 |
| 16.10.2019 05:00 | 77 938,00 | 203 185,00 | 7,1 | 3,1 |
| 16.10.2019 06:00 | 77 942,00 | 203 189,00 | 7,2 | 12,4 |
| 16.10.2019 07:00 | 77 948,00 | 203 200,00 | 9,8 | 9,2 |
| 16.10.2019 08:00 | 77 952,00 | 203 212,00 | 7,3 | 13,7 |
| 16.10.2019 09:00 | 77 958,00 | 203 223,00 | 7,2 | 7,6 |
| 16.10.2019 10:00 | 77 964,00 | 203 230,00 | 18,5 | 11,1 |
| 16.10.2019 11:00 | 77 972,00 | 203 236,00 | 11,6 | 14,9 |
| 16.10.2019 12:00 | 77 982,00 | 203 242,00 | 11,1 | 10,6 |
| 16.10.2019 13:00 | 77 991,00 | 203 249,00 | 12,1 | 13 |
| 16.10.2019 14:00 | 77 998,00 | 203 256,00 | 10,7 | 6,8 |
| 16.10.2019 15:00 | 78 005,00 | 203 264,00 | 15,3 | 7,1 |
| 16.10.2019 16:00 | 78 016,00 | 203 272,00 | 15,3 | 4,9 |
| 16.10.2019 17:00 | 78 025,00 | 203 282,00 | 10,4 | 3,7 |
| 16.10.2019 18:00 | 78 030,00 | 203 288,00 | 10,4 | 6,3 |
| 16.10.2019 19:00 | 78 034,00 | 203 294,00 | 10,4 | 6,4 |
| 16.10.2019 20:00 | 78 038,00 | 203 300,00 | 10,5 | 6,2 |
| 16.10.2019 21:00 | 78 042,00 | 203 303,00 | 16,2 | 5 |
| 16.10.2019 22:00 | 78 046,00 | 203 306,00 | 10,1 | 3,5 |
| 16.10.2019 23:00 | 78 050,00 | 203 308,00 | 9,6 | 2,7 |

# Appendix B

# U-Mass dataset machine learning analysis

Following are analysis results of running unsupervised algorithms on UMass dataset that is not included in Section 5.5.4. It can be seen that some of the experiments are with the full training dataset and some with percentage split of 66% between training and test dataset. A sample dataset of some selected features is also shown here.

# HomeA-meter2_Consolidated-2014-15-16

Instances: 281679

Attributes: 14

1. Date & Time
2. use [kW]
3. gen [kW]
4. FurnaceHRV [kW]
5. CellarOutlets [kW]
6. WashingMachine [kW]
7. FridgeRange [kW]
8. DisposalDishwasher [kW]
9. KitchenLights [kW]
10. BedroomOutlets [kW]
11. BedroomLights [kW]
12. MasterOutlets [kW]
13. MasterLights [kW]
14. DuctHeaterHRV [kW]

## 1. Simple K means (Cluster data using the k means algorithm.)

### a. Test mode: evaluate on training data

Time taken to build model (full training data) : 0.79 seconds

=== Model and evaluation on training set ===

Clustered Instances

0 231597 ( 82%)

**1 50082 ( 18%)**

Final Centroids

| **Attribute** | **Full Data** | **Cluster 0** | **Cluster 1** |
|---|---|---|---|
| use | 0 | 0 | 0 |
| gen | 0 | 0 | 0 |
| FurnaceHRV | 0.1579 | 0.0631 | 0.5964 |
| CellarOutlets | 0.0807 | 0.0798 | 0.0852 |
| WashingMachine | 0.0011 | 0.0009 | 0.0024 |
| FridgeRange | 0.0056 | 00.01.1900 | 00:10:57 |
| DisposalDishwasher | 0.0018 | 0.0017 | 0.0023 |
| KitchenLights | 0.013 | 0.0043 | 0.053 |
| BedroomOutlets | 0.0111 | 0.0102 | 0.0155 |
| BedroomLights | 0.0129 | 0.0064 | 0.0426 |
| MasterOutlets | 0.0297 | 0.0113 | 0.115 |
| MasterLights | 0.0319 | 0.0148 | 0.1113 |
| DuctHeaterHRV | 0.0362 | 0.0109 | 0.1529 |

# HomeA-meter3_Consolidated-2014-15-16

Instances: 281679

Attributes: 11

1. Date & Time
2. use [kW]
3. gen [kW]
4. ElectricRange [kW]
5. Dryer [kW]
6. GarageMudroomLights [kW]
7. DiningRoomOutlets [kW]
8. MudroomOutlets [kW]
9. MasterBathOutlets [kW]
10. GarageOutlets [kW]
11. BasementOutdoorOutlets [kW]

## 1. Simple K means

Test mode: evaluate on training data

Time taken to build model (full training data) : 1.47 seconds

=== Model and evaluation on training set ===

Clustered Instances

**0 14583 ( 3%)**

1 546556 ( 97%)

Final Centroids

| **Attribute** | **Full Data** | **Cluster 0** | **Cluster 1** |
|---|---|---|---|
| use [kW] | 0 | 0 | 0 |
| gen [kW] | 0 | 0 | 0 |
| ElectricRange [kW] | 0.0264 | 0.2528 | 0.0203 |
| Dryer [kW] | 0.0471 | 1.6393 | 0.0047 |
| GarageMudroomLights [kW] | 0.0121 | 0.0913 | 0.01 |
| DiningRoomOutlets [kW] | 0.0054 | 0.1382 | 0.0019 |
| MudroomOutlets [kW] | 0.0142 | 0.0378 | 0.0136 |
| MasterBathOutlets [kW] | 0.0108 | 0.0921 | 0.0087 |
| GarageOutlets [kW] | 0.0064 | 0.0153 | 0.0062 |
| BasementOutdoorOutlets [kW] | 0.0075 | 0.0252 | 0.007 |

# HomeA-meter4_Consolidated-2014-15-16

Instances: 572823

Attributes: 15

1. Date & Time
2. use [kW]
3. gen [kW]
4. KitchenDenLights [kW]
5. MasterBedBathLights [kW]
6. MasterOutlets [kW]

7. DenOutdoorLights [kW]
8. DenOutlets [kW]
9. RearBasementLights [kW]
10. KitchenOutletsEast [kW]
11. KitchenOutletsSouth [kW]
12. DishwasherDisposalSinkLight [kW]
13. Refrigerator [kW]
14. Microwave [kW]
15. OfficeLights [kW]

## 1. Simple K means

Test mode: evaluate on training data

Time taken to build model (full training data) : 7.13 seconds

=== Model and evaluation on training set ===

Clustered Instances

0 274747 ( 48%)

1 298076 ( 52%)

Final Cluster Centroids

| Attribute | Full Data | Cluster 0 | Cluster 1 |
|---|---|---|---|
| use [kW] | 0 | 0 | 0 |
| gen [kW] | 0 | 0 | 0 |
| KitchenDenLights [kW] | 0.04 | 0.0251 | 0.0537 |
| MasterBedBathLights [kW] | 0.0125 | 0.0095 | 0.0152 |
| MasterOutlets [kW] | 0.0102 | 0.0079 | 0.0123 |
| DenOutdoorLights [kW] | 0.0009 | 0.0007 | 0.0011 |
| DenOutlets [kW] | 0.0199 | 0.0024 | 0.0361 |
| RearBasementLights [kW] | 0.0032 | 0.0024 | 0.0039 |
| KitchenOutletsEast [kW] | 0.013 | 0.0095 | 0.0162 |
| KitchenOutletsSouth [kW] | 0.0145 | 0.0036 | 0.0246 |
| DishwasherDisposalSinkLight [kV | 0.0416 | 0.0051 | 0.0753 |
| Refrigerator [kW] | 0.0674 | 0.0102 | 0.1202 |
| Microwave [kW] | 0.0086 | 0.0042 | 0.0126 |
| OfficeLights [kW] | 0.0667 | 0.0569 | 0.0758 |

# HomeB-meter1_Consolidated-2014-15-16

Instances: 282640

Attributes: 18

1. Date & Time
2. use [kW]
3. gen [kW]
4. Grid [kW]
5. AC [kW]
6. Furnace [kW]
7. Cellar Lights [kW]
8. Washer [kW]
9. First Floor lights [kW]

10. Utility Rm + Basement Bath [kW]
11. Garage outlets [kW]
12. MBed + KBed outlets [kW]
13. Dryer + egauge [kW]
14. Panel GFI (central vac) [kW]
15. Home Office (R) [kW]
16. Dining room (R) [kW]
17. Microwave (R) [kW]
18. Fridge (R) [kW]

# 1. Simple K means

Test mode: evaluate on training data

Time taken to build model (full training data) : 0.96 seconds

=== Model and evaluation on training set ===

Clustered Instances

**0 72547 ( 26%)**

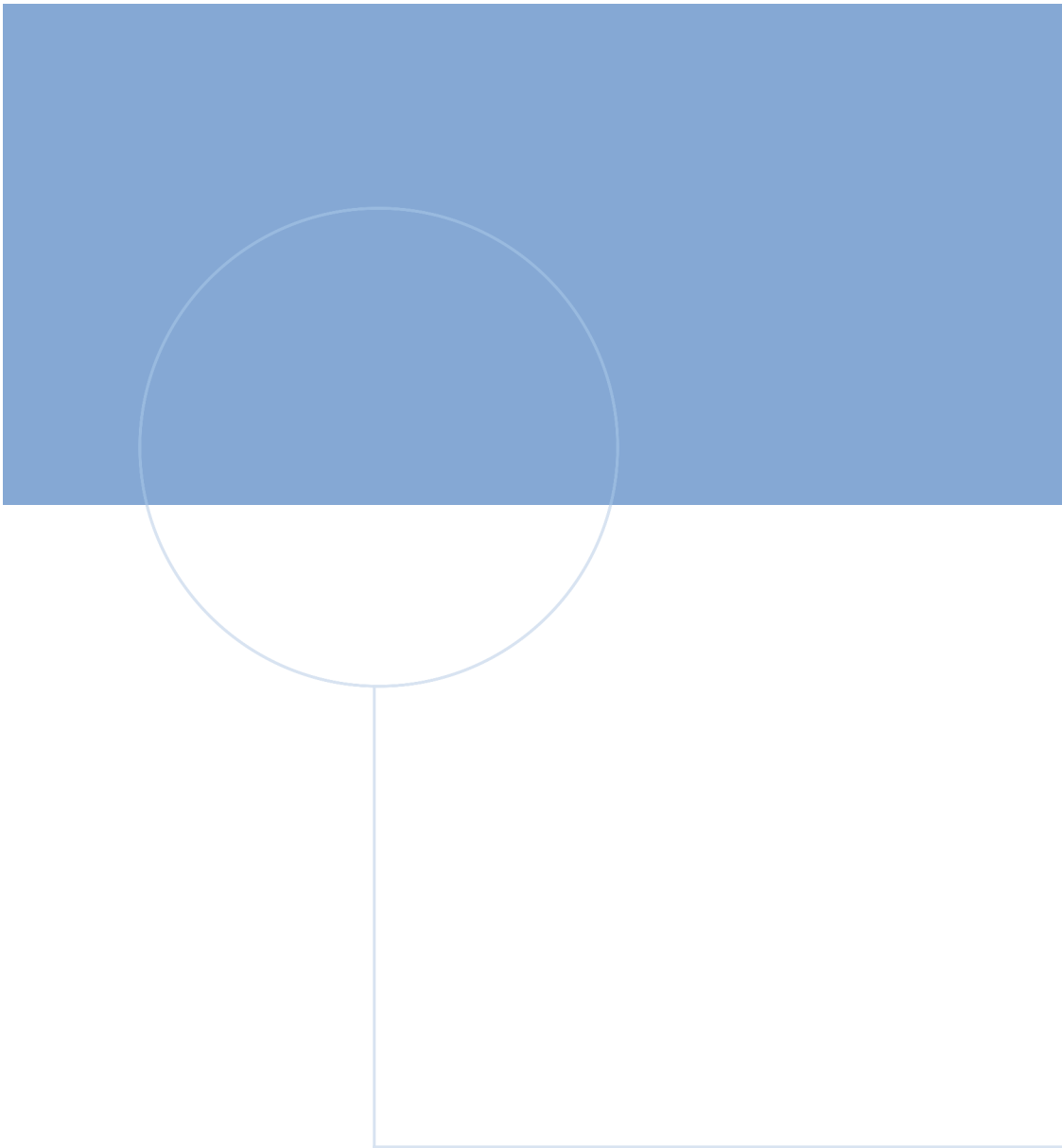
1 210093 ( 74%)

Final Cluster Centroids

| **Attribute** | **Full Data** | **Cluster 0** | **Cluster 1** |
|---|---|---|---|
| use [kW] | 0.9958 | 1.346 | 0.8749 |
| gen [kW] | 0 | 0 | 0 |
| Grid [kW] | 0.9958 | 1.346 | 0.8749 |
| AC [kW] | 0.3218 | 0.4464 | 0.2788 |
| Furnace [kW] | 0.1147 | 0.0924 | 0.1224 |
| Cellar Lights [kW] | 0.0091 | 0.0085 | 0.0093 |
| Washer [kW] | 0.0043 | 0.0048 | 0.0042 |
| First Floor lights [kW] | 0.0301 | 0.0252 | 0.0318 |
| Utility Rm + Basement Bath [kW] | 0.1434 | 0.2033 | 0.1227 |
| Garage outlets [kW] | 0.0054 | 0.0056 | 0.0054 |
| MBed + KBed outlets [kW] | 0.0804 | 0.0843 | 0.0791 |
| Dryer + egauge [kW] | 0.0261 | 0.0298 | 0.0248 |
| Panel GFI (central vac) [kW] | 0.0007 | 0.0003 | 0.0008 |
| Home Office (R) [kW] | 0.1119 | 0.3785 | 0.0199 |
| Dining room (R) [kW] | 0.0362 | 0.0414 | 0.0343 |
| Microwave (R) [kW] | 0.0122 | 0.0101 | 0.013 |
| Fridge (R) [kW] | 0.0722 | 0.0748 | 0.0713 |

# Sample dataset of selected features of HomeA

| Date & Time | FurnaceHRV [kW] | CellarOutlets [kW] | WashingMachine [kW] | FridgeRange [kW] | DisposalDishwasher [kW] | KitchenLights [kW] | BedroomOutlets [kW] |
|---|---|---|---|---|---|---|---|
| 01.01.2014 00:00 | 0.195337778 | 0.083204444 | 0.005686111 | 0.006891667 | 0.005568889 | 0.012153889 | 0.020451667 |
| 01.01.2014 00:30 | 0.182158333 | 0.036138889 | 0.005678889 | 0.094138333 | 0.005411667 | 0.0052 | 0.020570556 |
| 01.01.2014 01:00 | 0.134808333 | 0.047033889 | 0.005635 | 0.014786111 | 0.00551 | 0.003173333 | 0.020516111 |
| 01.01.2014 01:30 | 0.182125 | 0.071406667 | 0.005671667 | 0.082081111 | 0.005445 | 0.003071667 | 0.020506111 |
| 01.01.2014 02:00 | 0.092988333 | 0.014202778 | 0.00557 | 0.031901111 | 0.005401111 | 0.003153889 | 0.020411667 |
| 01.01.2014 02:30 | 0.130656111 | 0.070753333 | 0.005601667 | 0.054812778 | 0.005457222 | 0.003076111 | 0.020384444 |
| 01.01.2014 03:00 | 0.054559444 | 0.039048333 | 0.005543333 | 0.040308333 | 0.005352222 | 0.003192222 | 0.020511667 |
| 01.01.2014 03:30 | 0.082499444 | 0.043964444 | 0.005576111 | 0.049228889 | 0.005512778 | 0.003069444 | 0.020552778 |
| 01.01.2014 04:00 | 0.115085 | 0.069578889 | 0.005597222 | 0.037834444 | 0.005339444 | 0.003251111 | 0.020647222 |
| 01.01.2014 04:30 | 0.033995556 | 0.007424444 | 0.005532222 | 0.051721111 | 0.005597778 | 0.005637778 | 0.020636111 |
| 01.01.2014 05:00 | 0.08577 | 0.067398889 | 0.005601667 | 0.028158333 | 0.005308889 | 0.003345556 | 0.020737778 |
| 01.01.2014 05:30 | 0.088577222 | 0.051055556 | 0.005566111 | 0.060265 | 0.005663333 | 0.002962778 | 0.020489444 |
| 01.01.2014 06:00 | 0.236577222 | 0.020465556 | 0.005733333 | 0.012634444 | 0.005318889 | 0.0034 | 0.020813889 |
| 01.01.2014 06:30 | 0.690105556 | 0.078319444 | 0.00613 | 0.080910556 | 0.005888889 | 0.002851111 | 0.020662778 |
| 01.01.2014 07:00 | 0.645578889 | 0.028827778 | 0.006112222 | 0.018681111 | 0.005361111 | 0.00339 | 0.020769444 |
| 01.01.2014 07:30 | 0.573048333 | 0.044523889 | 0.006023333 | 0.075083333 | 0.005858333 | 0.002812222 | 0.020856667 |
| 01.01.2014 08:00 | 0.592766667 | 0.079576667 | 0.006055556 | 0.038977222 | 0.005079444 | 0.051691111 | 0.027404444 |
| 01.01.2014 08:30 | 0.521164444 | 0.010186667 | 0.005942222 | 0.055714444 | 0.005702222 | 0.25409 | 0.026731111 |
| 01.01.2014 09:00 | 0.563002222 | 0.078432222 | 0.005993889 | 0.074048333 | 0.005070556 | 0.238891111 | 0.027104444 |
| 01.01.2014 09:30 | 0.600661667 | 0.042062778 | 0.005998889 | 0.042467222 | 0.005863333 | 0.002761667 | 0.022361667 |
| 01.01.2014 10:00 | 0.499163333 | 0.044128333 | 0.005940556 | 0.062370556 | 0.004808333 | 0.003248889 | 0.012980556 |
| 01.01.2014 10:30 | 0.138334444 | 0.073493333 | 0.094268333 | 0.052166667 | 0.054229444 | 0.002788333 | 0.011717222 |
| 01.01.2014 11:00 | 0.051882778 | 0.011439444 | 0.144757222 | 0.050144444 | 0.746257778 | 0.003136111 | 0.012664444 |
| 01.01.2014 11:30 | 0.062891667 | 0.082459444 | 0.113875556 | 0.084619444 | 0.352852222 | 0.002701667 | 0.013569444 |
| 01.01.2014 12:00 | 0.102829444 | 0.031888333 | 0.005546667 | 0.059995 | 0.005025 | 0.00348 | 0.023377222 |
| 01.01.2014 12:30 | 0.054543333 | 0.053974444 | 0.005486667 | 0.021731667 | 0.005703889 | 0.002885556 | 0.022970556 |
| 01.01.2014 13:00 | 0.154086667 | 0.062461111 | 0.005612222 | 0.092938333 | 0.004983333 | 0.003293889 | 0.021939444 |
| 01.01.2014 13:30 | 0.071372222 | 0.022223889 | 0.016423333 | 0.036297222 | 0.005740556 | 0.00284 | 0.020083333 |
| 01.01.2014 14:00 | 0.184250556 | 0.082530556 | 0.164801111 | 0.059592778 | 0.004975 | 0.003297778 | 0.021072222 |
| 01.01.2014 14:30 | 0.709452778 | 0.016313889 | 0.14395 | 0.078263889 | 0.005476667 | 0.002570556 | 0.019667778 |
| 01.01.2014 15:00 | 0.558508333 | 0.060900556 | 0.011224444 | 0.010497222 | 0.00476 | 0.003323333 | 0.016923889 |
| 01.01.2014 15:30 | 0.614067778 | 0.102345556 | 0.127378333 | 0.116398333 | 0.005522222 | 0.002612222 | 0.034922778 |
| 01.01.2014 16:00 | 0.533529444 | 0.064127778 | 0.153873889 | 0.036248889 | 0.004738889 | 0.149671111 | 0.039352778 |
| 01.01.2014 16:30 | 0.562610556 | 0.115551111 | 0.009425556 | 0.101905556 | 0.005426667 | 0.249887222 | 0.039194444 |
| 01.01.2014 17:00 | 0.54488 | 0.075323889 | 0.128733889 | 0.068051667 | 0.004910556 | 0.249755 | 0.039561667 |
| 01.01.2014 17:30 | 0.622973889 | 0.051725 | 0.145342222 | 0.03125 | 0.005451111 | 0.252772222 | 0.039648889 |
| 01.01.2014 18:00 | 0.448922222 | 0.045747778 | 0.005894444 | 0.125758333 | 0.004642778 | 0.25443 | 0.03985 |
| 01.01.2014 18:30 | 0.132404444 | 0.078261111 | 0.005615556 | 0.0372 | 0.004981111 | 0.254537222 | 0.040141111 |
| 01.01.2014 19:00 | 0.108935556 | 0.013413889 | 0.005658333 | 0.120502222 | 0.004394444 | 0.256178333 | 0.031013333 |
| 01.01.2014 19:30 | 0.139362778 | 0.082696111 | 0.005632222 | 0.111287222 | 0.00517 | 0.256217222 | 0.022748333 |
| 01.01.2014 20:00 | 0.094052222 | 0.027533333 | 0.005608889 | 0.108448333 | 0.004882778 | 0.256898333 | 0.024024444 |
| 01.01.2014 20:30 | 0.122044444 | 0.059446111 | 0.005623889 | 0.138202778 | 0.008600556 | 0.252568333 | 0.023932222 |
| 01.01.2014 21:00 | 0.124163333 | 0.056362222 | 0.005658333 | 0.209431667 | 0.005358333 | 0.017547778 | 0.02415 |
| 01.01.2014 21:30 | 0.096300556 | 0.026252222 | 0.005623333 | 0.130559444 | 0.005555 | 0.051061667 | 0.023729444 |
| 01.01.2014 22:00 | 0.155921667 | 0.079721111 | 0.00567 | 0.081773889 | 0.005412222 | 0.003166111 | 0.023669444 |
| 01.01.2014 22:30 | 0.158684444 | 0.019642778 | 0.005666111 | 0.066070556 | 0.005592778 | 0.003053889 | 0.023643889 |
| 01.01.2014 23:00 | 0.196883889 | 0.067163333 | 0.005683333 | 0.033275556 | 0.005454444 | 0.003205 | 0.022086111 |
| 01.01.2014 23:30 | 0.12367 | 0.051918333 | 0.005638333 | 0.089881111 | 0.005535556 | 0.003057778 | 0.020751667 |



NTNU

Norwegian University of
Science and Technology