# DEVELOPMENT OF AN IMPROVED MACHINE LEARNING-BASED NON-TECHNICAL LOSS DETECTION MODEL FOR ADVANCED METERING INFRASTRUCTURE

**BY**

**ANIEDU AZUBUIKE NZUBECHUKWU**

**2009367006P**

**DEPARTMENT OF ELECTRONIC AND COMPUTER ENGINEERING,**

**NNAMDI AZIKIWE UNIVERSITY, AWKA**

**NOVEMBER 2017**

# DEVELOPMENT OF AN IMPROVED MACHINE LEARNING-BASED NON-TECHNICAL LOSS DETECTION MODEL FOR ADVANCED METERING INFRASTRUCTURE


**BY**

**ANIEDU AZUBUIKE NZUBECHUKWU**

**2009367006P**


**A DISSERTATION SUBMITTED TO THE DEPARTMENT OF ELECTRONIC AND COMPUTER ENGINEERING, NNAMDI AZIKIWE UNIVERSITY, AWKA**

**IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE AWARD OF DOCTOR OF PHILOSOPHY (Ph.D) IN COMPUTER AND CONTROL ENGINEERING**

**NOVEMBER 2017**

## CERTIFICATION PAGE

This research work “Development of an Improved Machine Learning-Based Non-Technical Loss Detection Model for Advanced Metering Infrastructure” was originally carried out by me under the supervision of Prof. H. C. Inyiama and has not been submitted in part or full to this university or other institutions for the award of a degree.

______________________ _______________

Aniedu, Azubuike Nzubechukwu Date

**APPROVAL PAGE**

This dissertation titled “DEVELOPMENT OF AN IMPROVED MACHINE LEARNING-BASED NON-TECHNICAL LOSS DETECTION MODEL FOR ADVANCED METERING INFRASTRUCTURE” has been approved by the Department of Electronic and Computer Engineering, Faculty of Engineering, Nnamdi Azikiwe University, Awka in partial fulfilment for the award of Doctor of Philosophy (Ph.D) in Computer and Control Engineering.

By:

| | |
|---|---|
| ______________________ | _____________ |
| Engr Prof. H. C. Inyiama<br>Supervisor | Date |
| ______________________ | _____________ |
| Engr. Prof C. C. Okezie<br>HOD, ECE Department | Date |
| ______________________ | _____________ |
| Engr. Prof. Gloria Chukwudebe<br>External Examiner | Date |
| ______________________ | _____________ |
| Engr Prof. V. E. Idigo<br>Dean, Faculty of Engineering | Date |
| ______________________ | _____________ |
| Prof. Ike Odimegwu<br>Dean, School of Postgraduate Studies | Date |

## DEDICATION

**This Ph.D dissertation is dedicated to God Almighty,**

**Possessor of the ends of the earth.**

**ACKNOWLEDGEMENT**

First, I want to thank God for seeing me through this struggle, He is indeed a merciful God.

Secondly, I wish to thank my supervisor, Prof. H. C. Inyiama, for putting up with me all these while supportively.

Then I wish to thank profusely Dr A.C.O Azubogu. Thank you for not giving up on me, and allowing God to use you to keep my faith and drive alive.

To my lecturers Prof. C.C. Okezie (my HOD and strong support), Prof. V.E. Idigo (my amiable Dean), Dr. C. O. Ohaneme, and the entire staff of Electronic and Computer Engineering, UNIZIK; thank you all for your unnumbered support.

And to my queen, my best half, Chioma Precious - Your support and encouragement is ever refreshing.

To my siblings (Chinyelu, Ifeanyi, and Chidiogo), my victory is your victory. We've come a long way and am grateful to have siblings like you.

To my parents-in-law, Mr and Mrs Ezeokpo, with your support, it felt like I never was orphaned afterall.

I must not also fail to acknowledge these who kept me on my toes; Prof and Prof Mrs J. E. Ahaneku (the Vice Chancellor of UNIZIK), Prof. C. O. Esimone (DVC (Acad), UNIZIK), Prof. Mrs P.K. Igbokwe, Prof and Dr Mrs Emma Okoye, Dr. Chinonso H. Achebe, Prof. Godson Okafor, Very. Rev. M.O Nwakanma (Chaplain, Chapel of Glory Unizik), Chikosolu Uzoka, Chinwe Ogbuefi, Sandra Nwokoye, Muna Onyendimele; your consistent urging kept me forging on.

Also to Charles Oranugo and Tony Morah who worked tirelessly to see to the actualization of this dream, thank you.

I love you all.

## ABSTRACT

Non-technical loss (NTL) defined as any consumed energy or service which is not billed because of measurement equipment failure or ill-intentioned and fraudulent manipulation of said equipment and which therefore results in inconsistencies in consumption profile of the consumers is a major problem for electricity utility companies. This dissertation presents an enhanced electricity consumption monitoring algorithm for non-technical loss detection in Advanced Metering Infrastructure (AMI) based on the analysis of consumers' consumption pattern leveraging Machine learning (ML) techniques. Support Vector Machines (SVM) was selected, modelled, trained and applied towards classifying consumer's electric energy usage readings (after filtering and formatting), and also for performing predictive analysis for the dataset after a careful survey of a number of machine learning classifiers and a methodical selection of the two main SVM parameters (ie Cost parameter, C, and kernel function, gamma). A novel pre-classifier was designed and developed which resulted in better prediction outcome with the SVM classifier. Classification accuracy (and subsequently, class prediction) of 99.2% was achieved with the developed pre-classifier as against 79.46% obtained without pre-classification (although there is a trade-off with processing time when the pre-classification time is taken into consideration). This implies that utility workers can predict the occurrence of NTL in electricity usage with about 99.2% accuracy using the developed model. This will enable them take intelligent decisions and promptly disconnect any fraudulent user remotely, using facilities embedded in AMI and smart meters. It was also observed from analysis of the energy usage dataset that there is normally a heightened use of electric energy during the winter period (especially among residential consumers) in contrast to other seasons of the year, which is a critical information for balancing the load in energy distribution. It has been shown, through this research, that fraud detection in electricity consumption, and hence a solution to non-technical losses can be achieved using the right combinations of Machine Learning techniques in conjunction with AMI technology.

**TABLE OF CONTENTS**

Table of Contents

**List of Figures**

**List of Tables**

## List of Abbreviations

| | |
|---|---|
| **AMI** | Automatic Metering Infrastructure |
| **AUC** | Area Under Curve |
| **CDU** | Credit Dispense Unit |
| **CV** | Cross Validation |
| **CVS** | Common Vending System |
| **DA** | Decryption Algorithm |
| **DDTK** | Dispenser Default STS Key |
| **DITK** | Dispenser Initialization STS key |
| **DK** | Meter Key |
| **DUCK** | Dispenser Common STS Key |
| **DUTK** | Dispenser Unique STS key |
| **EA** | Encryption Algorithm |
| **ECN** | Electricity Corporation of Nigeria |
| **ED** | Electricity Dispenser |
| **ELUPCI** | Electricity User Pre-Classifier Interface |
| **ESI** | Electricity Supply Industry |
| **FN** | False Negative |
| **FP** | False Positive |
| **GSM** | Global System of Mobil Telecommunication |
| **ISO** | International Organization for Standardization |
| **KEK** | Key Exchange Key |
| **KLF** | Key Load File |
| **KMC** | Key Management Center |
| **KRN** | Key Revision Number |
| **kW** | Kilo Watt |
| **LIBSVM** | Library for Support Vector Machines |
| **MIS** | Management Information System |
| **ML** | Machine Learning |

| | |
|---|---|
| **MSNO** | Meter Number |
| **MW** | Mega Watt |
| **NDA** | Niger Dam Authority |
| **NEPA** | National Electric Power Authority |
| **PHCN** | Power Holding Company of Nigeria |
| **POS** | Point of Sales |
| **QP** | Quadratic Problem |
| **RBF** | Radial Basis Function |
| **ROC** | Receiver Operating Characteristics |
| **SGC** | Supply Group Code |
| **SM** | Secure Module/Smart Meter |
| **SMS** | System Master Station |
| **STS** | Standard Transfer Specification |
| **STT** | Standard Token Translator |
| **SVM** | Support Vector Machines |
| **SVR** | Support Vector Regression |
| **TI** | Tariff Index |
| **TID** | Manufacturer Number Token Identification |
| **TM** | Transaction Manager |
| **TP** | True Positive |
| **UCI** | University of California, Irvine |
| **VK** | Vending Key |
| **WEKA** | Waikato Environment for Knowledge Analysis |

# CHAPTER ONE

# INTRODUCTION

## 1.1 Background of the Study

Energy distribution companies have, over the years, been faced with a barrage of problems in terms of equitable distribution of energy to users, proper metering and monitoring of usage, dealing with fraud and vandalism, to mention but a few.

It is a well-known fact that one other major problem in electric power systems management is the acquisition of electricity data. In countries like Nigeria, before the introduction of prepaid meters, staff of the electricity distribution company often have to go round, house by house to take energy reading of users' meters in order to be able to compute their monthly usage tariff. To this effect, the official gazette of the Federal Government of December 24, 2007 (Federal Republic of Nigeria Official Gazette, 2007), and August 1st 2012 (Federal Republic of Nigeria Official Gazette, 2012) among others stipulate the methodology for estimated billing and manual meter reading of consumer's energy meters. With the introduction of the prepaid meter, users have to travel to designated locations to buy tokens (a twenty digit number in which the amount of energy purchased is encrypted) and take this back to their meter for activation (whether the meter is the keypad type or the smartcard type). This method does not in any way encourage accumulation of usage data. Hence there is no comprehensive data bank that contains information on how much energy is used by who and when. This is neither ideal nor neat especially if one wants to have information for use in predicting or forecasting energy usage in order to aid future generation or distribution.

Speaking at the Nigeria Power Sector Investment Forum, the acting director of Electric Power Sector Team of the Bureau of Public Enterprises, Mr. Joe Anichebe listed inefficient usage of capacity, ineffective regulation, high technical losses and vandalism, inefficient use of electricity by consumers, inappropriate industry and market structure, among other issues, as part of the reasons necessitating the on-going reforms in that sector (Anichebe, 2013).

The issue of theft and vandalism (Non-Technical Losses, NTL) is another major problem to contend with. Energy distribution companies have had to part with billions of Naira to vandals and fraudsters due to theft, illegal generation and sale of tokens, fraudulent diversion of collected funds by staff and so on. But among all these, energy theft has been the most difficult to contend with, especially in emerging economies. A World Bank report discovers that up to 50% of electricity in developing countries is acquired through theft (Antmann, 2009). It is reported that each year over 6 billion dollars are lost due to energy theft in the United States alone (McDaniel & McLaughlin, 2009). In Canada, BC Hydro reports 100 million dollars in losses every year (CBC News, 2010). Utility companies in India and Brazil incur losses around 4.5 billion and 5 billion dollars respectively due to electricity theft (Ministry of Power, India, 2013), (Federal Court of Audit, Brazil, 2007). This creates a lot of avoidable losses in the system which affects the quality of supply, the electricity load on the generating station, and the tariff imposed on usage by genuine customers. Improper monitoring system also prevents the authorities from knowing exactly how much money had actually been realized from sales of energy.

The introduction of prepaid meter was intended to, among other things check the issue of theft, especially non-payment or underpayment of energy bills which was very rampant with manual meters. The prepaid energy meter is designed in such a way that the meters can only supply energy to the user for as long as there is energy credit unit remaining in the user's meter. Once the unit expires, the meter cuts off energy supply to user, thus enforcing a 'pay-as-you-use' policy.

The effective operation of the electricity metering system and that used for the vending of token in prepaid electricity market is essential to the success of the electricity distribution companies. The system currently in use is called the Common Vending System (CVS). According to Gareth (2006) this system has a number of shortfalls mostly due to the fact that it is an off-line system-There are devices present in the field that are independently capable of producing tokens that represent electricity credit. Thus the common vending system (CVS) currently used in many developing economies including Nigeria is grossly inefficient in tackling the problem of non-technical losses.

In order to surmount these problems, the architecture of the electricity metering system is being changed from off-line, semi-independent systems to a fully integrated online system described as advanced metering infrastructure (AMI).

Advanced Metering Infrastructure (AMI) is a hierarchical structure comprising the networking of varied digital or electronic hardware and software which combine interval data measurement with continuously available remote communication. It enables measurement of detailed, real-time information and frequent collection and transmission of such information to various parties for proper monitoring and recording. A key component of the AMI is the smart meter. Smart meters are energy meters which in addition to basic metering capabilities, are equipped with additional features enabling them to communicate to remote servers, monitor and control energy usage by home appliances with options of remote management (disconnection, reconnection, etc) and credit recharge.

Hence with AMI, devices that are capable of generating tokens are not required to be present in the field and direct communication of the meters and the servers will ensure that proper monitoring is done by the utility on general use and practice; this also provides the much needed data bank.

However, the impact of the adoption of AMI and smart meters may not be felt appreciably if there is no way of analysing the data these technologies provide and use them for preventive actions against non-technical losses. This research work zeros in on utilizing electricity usage data analysis (particularly with machine learning techniques) to solve this menacing problem of non-technical losses in the electricity distribution industry.

## 1.2 Statement of Problem

Although the AMI and smart grid proves to be significantly superior to the Common Vending System (CVS) in terms of energy monitoring and non-technical loss detection, it still has a number of shortfalls. Recent researches has shown that even AMI-based systems can be defrauded and some of its security features bypassed. Most of these AMI-based threats have been categorized into physical attacks, cyber hacks and data attacks and hijacks and these invariably undermines the efforts put in AMI in preventing NTL necessitating the need for further research towards curbing this issue.

The setbacks in monitoring electric energy usage by utilities in a bid to ensure proper and judicious utilization of this very important national resource and the provision of a platform for effective monitoring of deployed meters against tampering or other fraudulent activities motivates this research work.

## 1.3 Aim and Objectives

The aim of this research is to develop an enhanced electricity consumption monitoring for non-technical loss detection in AMI based on the analysis of consumers' consumption pattern leveraging Machine Learning (ML) techniques.

Specific objectives include:

1. To design and develop an effective algorithm that can be used in analysing and classifying electricity consumers based on their consumption patterns;
2. To develop this algorithm into a real-time pre-classifier which would improve the classification and prediction outcomes of classifier techniques;
3. To leverage Machine Learning (ML) technologies to find anomalies in customer energy consumption profile thereby identifying non-technical losses in electricity distribution and management;
4. To implement an improved model based on ML technologies for predicting usage patterns and hence easily detect abnormalities in usage trends
5. To evaluate the performance of the integrated infrastructure in terms of chosen parameters like user classification, pattern prediction, NTL detection and meter monitoring.

## 1.4 Justification for the Project

Electrical energy has become an integral part of our everyday life. Its importance cannot be overemphasised. In fact, one can convincingly argue that without electricity, most modern industries will fold up and the world will be threatened to return to the dark ages. But, indiscriminate transmission/distribution of this all important energy (or other forms of energy for that matter) will create enormous problems because while one part of the country/city is flooded with excess energy, another part is not having enough, hence resulting in non-uniform and irregular developmental growth.

This research therefore seeks to prevent this by enhancing the technology used to facilitate the monitoring of this energy sector hence the objectives. It is expected that the outcome of this research effort will enable utilities and government agencies responsible for the distribution of electricity energy monitor utilization better, detect and prevent fraudulent use and hence ensure adequate distribution of electric energy resources available.

## 1.5 Scope of Work

This research is focused on modelling an effective and enhanced system and then classifying consumption information of electricity consumers using a suitable Machine Learning algorithm for the purpose of monitoring usage and utilization of electric energy.

# CHAPTER TWO

# REVIEW OF RELATED LITERATURE

## 2.1 Historical Development of Electric Energy in Nigeria

Electricity development in Nigeria can be traced back to around 1898 when the first power plant was installed in the city of Lagos (Power Holding Company of Nigeria, 2014). Since then, the industry had seen several transformation notably the passing of the Electricity Corporation of Nigeria (ECN) ordinance No.15 in 1950 and the merging of the ECN and the Niger Dam Authority (NDA) to give birth to the National Electric Power Authority (NEPA) in 1973 (Power Holding Company of Nigeria, 2014).

Another major landmark in the history of this industry is the subsequent privatization of NEPA through the enactment of the National Electric Power Policy in 2001, and the Electric Power Sector Reform Act (EPSRA) in 2005 which gave rise to initial holding company named Power Holding Company of Nigeria (PHCN) (Nigeria Electricity Regulation , 2013). Subsequently in September 2013, PHCN was fully privatized and unbundled into 18 successor companies: six generation companied, 11 distribution companies covering all 36 states of Nigeria and Abuja, and a national power transmission company (Nigerian Electricity Privatization (PHCN). , 2013).

These were a direct fall out of the power sector reform embarked upon by the Federal Government and according to Anichebe, in (Anichebe, 2013), the government seek to achieve such objectives as ensuring that Nigeria has an Electricity Supply Industry (ESI) that can meet the needs of its citizens in the 21st Century; This entails that Nigeria (ESI) must be such that is able to:

- Meet all current and prospective economically justifiable demands for electricity throughout the country;
- Modernize and expand its coverage; and
- Support national economic and social development, including relations with neighbouring countries.

Current statistics by the Presidential Task Force on Power puts the total available capacity of generated electric power at 4842MW (Bureau of Public Enterprises (BPE), 2013) as against the promise of the former President, General Olusegun Obasanjo to increase it to 10,000MW by 2007. A comparative analysis of electric energy generation in other countries places Nigeria at a relatively bottom scale (Ifedi, 2005).

## 2.2 Electric Energy Metering in Nigeria: Current State Analysis

In 2006 PHCN introduced the prepaid metering system (Orukpe P., Agbontaen F., 2013). Since then, a number of the consumers have migrated to the prepaid meter. Part of the reasons given by PHCN for migrating to and introducing the prepaid meters include:

- Rampant case of meter tampering
- Difficult or very remote access to meters for meter reading
- Many customers had to be supported by very few number of personnel which often result in some areas continually being given estimated bills.
- Many customers do not understand nor had the budget to pay for the accumulated fixed charges or bills that arrive only after the electricity had been consumed.
- Difficulties experienced with customers withholding payment for electricity etc.

## 2.3 Basic Description of Prepayment Technology

The concept of prepaid electricity operates in virtually the same way as that of the GSM services. In other words, the customer pays upfront for energy before consumption. Much like that of the GSM technology, the energy consumption is controlled at the consumers premise through a prepaid meter on which is loaded an afore purchased electricity credit and continues to dispense electric service to the consumer until the purchased credit depletes, after which electric dispensing is cut off.

The typical prepayment system in use consists of the following components

The Credit Dispensing Unit (CDU)
The encrypted credit transfer token and
The Prepaid meter (or Electricity Dispenser, ED)
A management information system, MIS (Optional)

The Prepaid electricity setup operates in the following manner: the prepaid meter is installed in the premise of the customer to control the amount of electricity provided to the customer. The customer then purchases an encrypted credit transfer token from a credit dispensing unit. The token represents the amount of electricity that the customer has paid for. This token is then inputted into the ED to provide electricity to the customer. Where the MIS exists, it interfaces with the CDU periodically for the management of the token generator and for accounting purposes (see Figure 2.1).

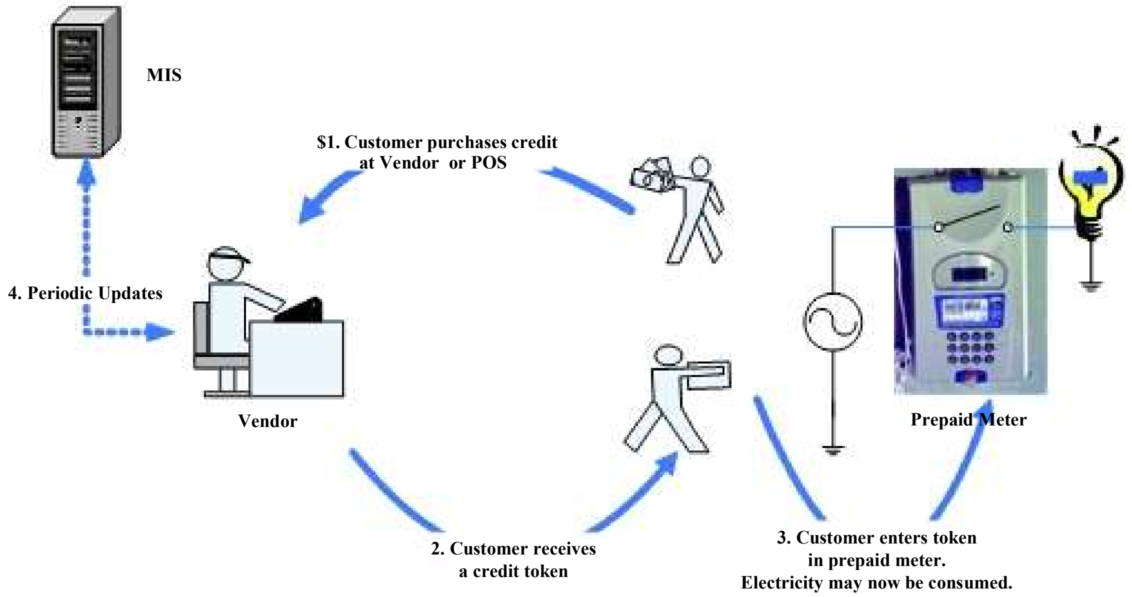



Figure 2.1: Generic Electricity Prepayment System.

The components of the prepayment system are discussed below:

### 2.3.1 The Credit Dispense Unit (CDU)

The typical first generation CDUs (figure 2.2), developed to the functional and performance requirements defined in NRS 009-2-2 (National Rationalised Specification, Functional and performance requirements: Credit dispensing units, NRS 009-2-2.) were built as specialized, self-contained devices for prepayment vending that could be securely mounted on a wall or counter. They have the capacity to vend magnetic and numeric tokens, using either a card reader (to read customer's meter cards) or printer. These devices were designed to assume a low operator skill and educational level. Also they were designed to withstand very harsh operating environment like dust, static electricity, power surges, vibration, insects and operator abuses (Kaplan, 1995).

Figure 2.2: Specialized CDU (Kaplan, 1995)

The newer generation CDUs (figure 2.3) were no longer built as specialised devices. They make use of a standard personal computer on which is attached peripherals like card reader etc.

Figure 2.3: Typical PC based CDU installation (Kaplan, 1995)

### 2.3.2 The Encrypted Credit Transfer Token and the Standard Transfer Specification (STS)

The Standard Transfer Specification (STS) was developed in order to facilitate the administration of prepaid electricity. It is a specification that defines how electricity is represented by a token. Its definition became imperative to forestall the confusion that arises as a result of incompatibility between different vending systems developed by different ED manufacturers (Kaplan, 1995). This incompatibility therefore results in the inability of the vending system of one manufacturer to vend for the ED from another manufacturer. This proves to be overtly expensive, inefficient and operationally inconvenient. Hence the STS was developed to provide an 'open system' standard for one-way prepayment meters. In other words, it allows information that is provided by a CDU to be transported to an ED and be meaningfully interpreted by the ED.

STS defines a secure message protocol that allows information to be carried between the CDU and meter. The information carried between the CDU and the meter is encrypted by the protocol into a fixed length value commonly referred to as a token. More than one token may be created depending on the amount of information being carried between the CDU and meter. STS caters for several message types such as credit, configuration, display and test instructions. It further specifies devices and codes of practice that allows for the secure management (generation, storage, retrieval and transportation) of cryptographic keys used within the system (Kaplan, 1995).

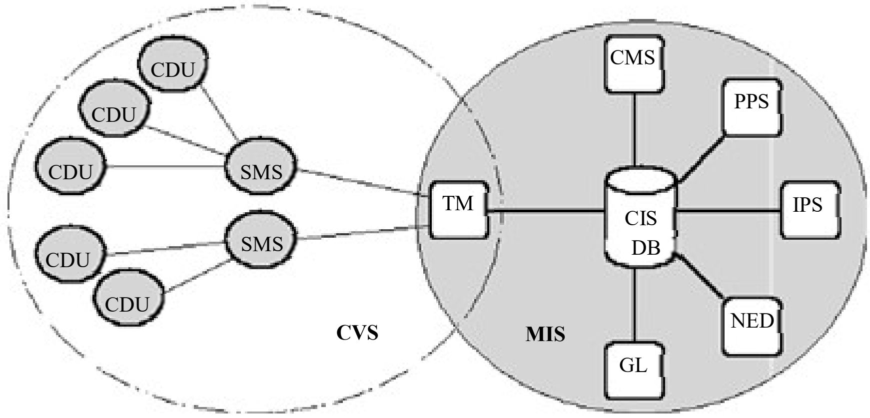
There are two types of token technologies: disposable magnetic card (STS Association, 2004) and numeric (STS Association, 2002). The disposable magnetic card technology adheres to the ISO (International Organisation for Standardisation) 7810 series of standards. It is made of paper and is intended to be used once only. A magnetic card reader is required at the ED. The numeric token technology, on the other hand, encodes the data as a string of 20 digits. Because of this, the physical transport medium can vary, and could even be communicated via audio or e-mail. The corresponding input mechanism at the ED is a numeric keypad. This is the more versatile of the two token technologies.

### 2.3.3 System Master Stations (SMSs)

System Master Stations (SMSs) are used to concentrate transaction data collected from CDUs. The collected transaction data is then uploaded to a higher-level MIS where it exists. Typical SMS functions include managing vending transactions, customer and meter management, tariff management, vendor credit and CDU configuration management. SMS functional and performance requirements are specified in (National Rationalised Specification, 2000). The SMS application is installed on a standard personal computer and operated at the utilities offices. Data transfer between the SMS and CDU can be accomplished by modem transfer or by disk transfer as specified in (National Rationalised Specification) and (National Rationalised Specification, 2000). An enhanced SMS can also function independent of an MIS and therefore operates in a standalone mode. In this mode, the SMS is responsible for the management of the customer and transaction databases.

## 2.4 The Common Vending System (CVS)

The Common Vending System (CVS) can be regarded as the collection of the entire infrastructure required for sale of token. The CVS comprise of the Credit Dispense Units (CDUs), System Master Stations (SMSs) and possibly a centralised database with its Mainframe Information System (MIS). A typical CVS consists of multiple groups of CDUs which are located at various Points of Sales (POS) locations. The data from each CDU group is then concentrated by an SMS. Data from the SMSs are in turn concentrated by a Transaction Manager (TM) on the Mainframe Information System, forming hierarchical network architecture as shown in Figure 2.4




CDU – Credit Dispensing Unit
CVS - Common Vending System
SMS – System Master Station
TM – Transaction Manager
________ Data Transfer
________ Data Storage

CIS DB – Customer Information System Data Base
CMS – Customer Management System
GL – General Ledger
IPS – Power Billing System
MIS – Mainframe Information System
NED – Nedisys – ED Tracking System

Figure 2.4: A Schematic Overview of CVS (Tewari D. D., 2003)

In the CVS, there are three different vending key types that are used to generate the STS keys that were described in Table 2.1. These types of keys are only used with the CVS; the STS is never aware of them. At any given moment, a unique VDDK exists for each default group defined in the CVS. Similarly, a unique VUDK exists for each unique group, and a unique VCDK exists for each common group.

Table 2.1: Vending Key Types

| Abbreviation | Name | Purpose |
|---|---|---|
| VDDK | Vending Default DES Key | Seed key for the generation of type 1 (DDTK) key values only |
| VUDK | Vending Unique DES Key | Seed key for the generation of type 2 (DUTK) key values only |
| VCDK | Vending Common DES Key | Seed key for the generation of type 3 (DCTK) key values only |

Figure 2.5 shows information flow in the CVS. Communication between the MIS and the SMS is bidirectional in order for transaction data and customer data to be exchanged and updated between the two. The CDU sends transaction information to the SMS via modem or disk; and also sends information to the ED via the token. There is no reverse communication between the CDU and SMS, and the CDU and ED.

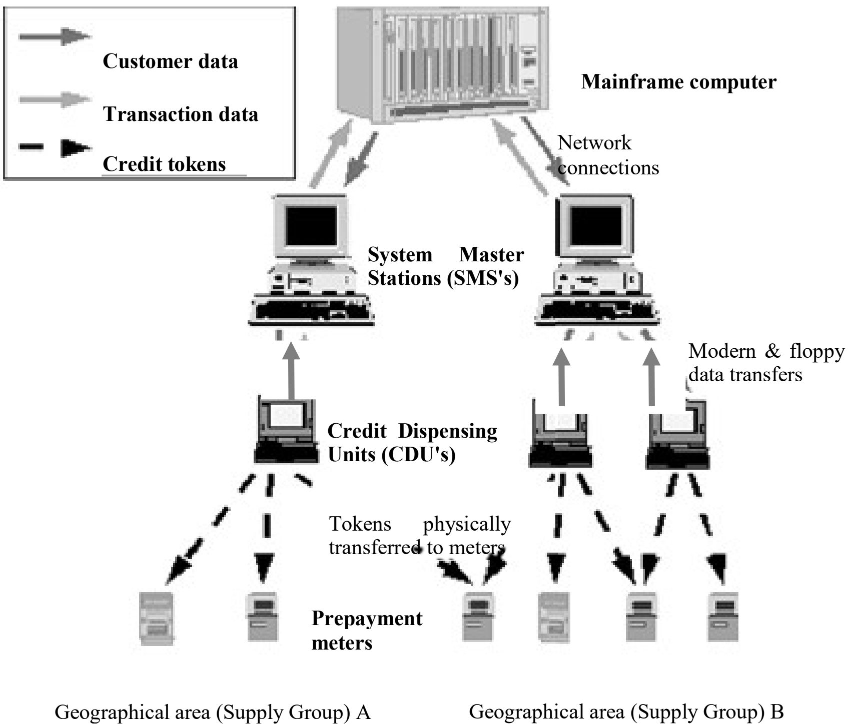



Figure 2.5: A Schematic Diagram Showing Information Movement within the CVS (Tewari D. D., 2003)

## 2.5 The Offline Vending System

The Common Vending System (CVS) was developed to provide total prepayment electricity system capable of supporting a widespread deployment of vending systems and meters. The system was standardised so as to ensure interoperability between CDUs and EDs from different manufacturers. The establishment of CDUs and POSs at various locations was geared towards providing convenience for customers to be able to purchase tokens easily. PHCN have even gone a step further (in partnership with some financial institutions) to introduce an online mode of payment known as PHCN easy payment solution (Shaibu, 2012) where customers can pay their electricity bills from ATM machines or on mobile apps.

However, the current setup of the CVS are still more or less offline, because the CDUs can only communicate with the EDs via the token. There is no reverse communication between the two. Also the CDU have a one way communication with the SMS. This makes the CVS more or less a stand-alone system or offline (See figure 2.5).

## 2.6 Advanced Metering Infrastructure and Smart Meters

The **Advanced Metering Infrastructure (AMI)** is a hierarchical structure and comprise of a number of different networks communicating with each other as shown in Figure 2.6 (Rong & etal, 2004). According to Electric Power Research Institute (EPRI) (Electric Power Research Institute, 2007), it comprise of state-of-the-art electronic/digital hardware and software, which combine interval data measurement with continuously available remote communications, and enable measurement of detailed, time-based information and frequent collection and transmittal of such information to various parties. Accordingly, AMI typically refers to the full measurement and collection system that includes meters at the customer site, communication networks between the customer and a service provider, and data reception and management system that make the information available to the service provider. AMI modernises the electricity metering system by replacing old mechanical meters with smart meters, which provide two-way communications between utility companies and energy customers (Rong & etal, 2004).

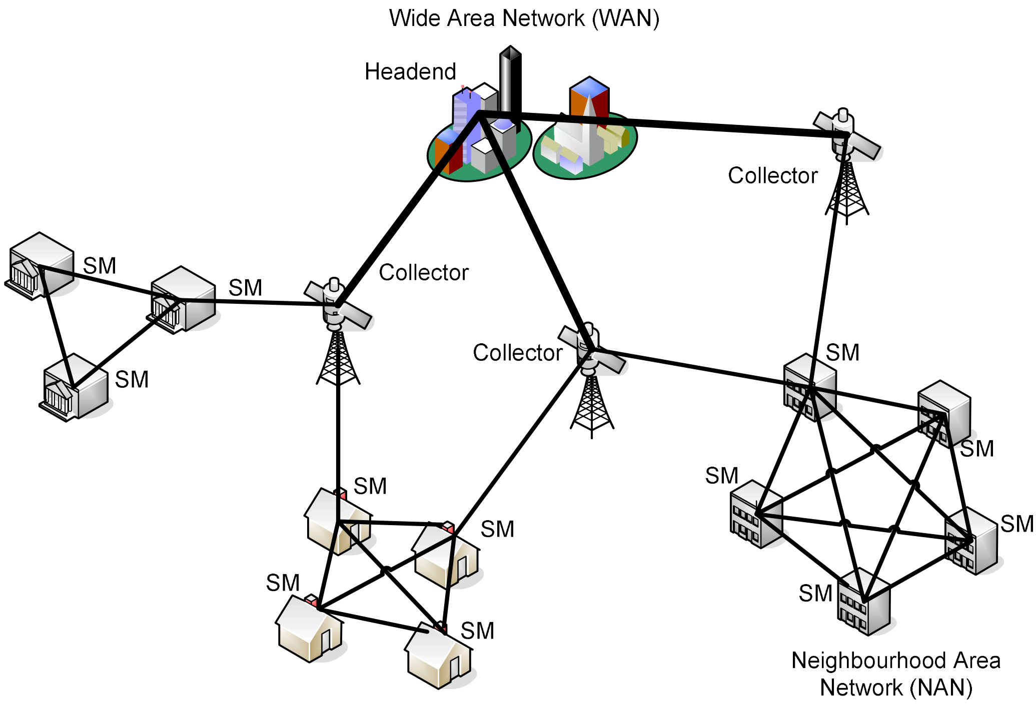



Figure 2.6: A Simple AMI Architecture (Rong & etal, 2004)

AMI technologies are rapidly overtaking the traditional meter reading technologies and millions of smart meters are equipped in the household all over the world (Rong & etal, 2004). For example, there are already more than 4.7 million smart meters used for billing and other purposes in Ontario, Canada (GSM Association, 2010). According to the American Institute for Electric Efficiency (IEE), approximately 36 million smart meters have been installed in the United State by May 2012, and additional 30 million smart meters will be deployed in the next three years (Wikipedia, the online encyclopedia, 2014).

Table 2.2 shows comparison between the basic features of a manual metering technology and an Advanced Metering Infrastructure.

Table 2.2: Comparison of Manual Metering Technology with AMI (King, 2004)

| System Element/Feature | Manual | Advanced Metering Infrastructure |
|---|---|---|
| **Meters** | Electromechanical | Hybrid or solid-state |
| **Data collection** | Manual, monthly | Remote via communication network, daily or more often |
| **Data recording** | Total consumption | Time-based (usage each hour or more often) |
| **Primary application** | Total consumption billing | Pricing options, customer options, utility operations, emergency demand response |
| **Key software interfaces** | Billing and customer information system | Billing and customer information system, customer data display, outage management, emergency demand response |
| **Additional devices enabled (but not included in base infrastructure)** | None | Smart thermostats, in-home displays, appliance controllers |

Some of the key benefits of AMI over conventional meters as identified by Chris King (King, 2004) include:

- Choice of billing date
- No estimated bills
- Projected month-end bill
- Choice of flat rates or dynamic pricing
- Automatic response and restoration verification by utilities
- Real-time meter readings

- First call problem resolution
- Web data access
- Monthly detailed usage reports
- Baseline threshold alarms
- Month-to-date usage
- Daily or hourly data for customer education

### 2.6.1 Smart Meters

Smart meters, a key component of AMI are meters equipped with additional features, apart from basic metering. These features include ability to communicate with other meters or concentrators using either wired or wireless technologies, ability to monitor and control energy usage by home appliances, capability for remote monitoring and management (disconnection, reconnection, credit recharge etc)

Several researchers and authors have attempted to propose a model for smart meters. Some of them are reviewed below.

Sudheer et al (Sudheer & etal, 2013) proposed the development of an equipment using a PIC16f887A microcontroller and a GSM module to enhance existing electromechanical meter's ability towards communication, exchange of usage data and tracking of energy theft by calculating stolen energy. This system's monitoring capability is limited and applied majorly to theft detection.

Sebola and Penzhorn (Sebola & Penzhorn, 2003), in researching against prevention of vending fraud, designed a secure mobile commerce system for vending of prepaid electricity token which they called m-commerce. This system however requires users to manually input generated tokens into the prepaid meter.

In their work ‘A prepaid meter using mobile communication’, Jain A. and Bagree M. (Jain & Bagree, 2011) modelled a prepaid card which is embedded into existing prepaid meter. This card incorporates a mobile wireless communication module to enhance communication between the meter and utility companies, hence making the meter smarter.

Raza A. et al (Raza & etal, 2013) in their work ‘online monitoring of electricity data through wireless transmission using radio frequency’ designed an energy meter possessing telemetry capabilities using ATmega32L and ATmega8L microcontrollers, PT2262 encoder and PT2272 decoder and interfacing it to the computer serial port through MAX232 and DB9. They were able to perform remote monitoring and collection of electricity usage data which is the primary purpose for this research.

Actaris Metering Systems (Actaris Metering Systems, 2005) in a publication titled ‘wireless data communication solution for prepayment meters’ described an award winning innovation of an RF device which is connected to back of an STS compliant meter. This device named ACE9000 InfoPOD is then interfaced to a handheld unit having an RF master transceiver unit to enhance communication between it and the InfoPOD. This enables utility workers to take energy reading and meter functionality check without requiring physical access to the meters. This setup however still requires that the worker move round block by block with the handheld device in order to take energy readings as the transmission distance of the InfoPOD is very limited.

In the work ‘Development of a smart power meter for AMI based on ZigBee Communication’, Shang-Wen (Shang-Wen & al., 2009) and his colleagues developed a smart power meter based on ZigBee network for energy meter reading and data communication.

Smart meters have the ability to measure readings of energy usage and send it to utility control centre through different techniques and protocols. Some of these techniques/protocols include Bluetooth for Home Area Networks (HAN) which employ the IEEE 802.15.1 protocol, Broadband Power Line communication (BPL) employing TCP/IP over radio frequency spectrum, Wi-Fi or WiMAX technology employing the 802.11a/b/g/n standard. Other techniques include protocols like ZigBee network employing the 802.15.4 standard, the Global system for Mobile Communication (GSM), the General Packet Radio Service (GPRS) etc.

## 2.7 Reported Threats in Electric Energy Management and Electricity Theft

As is it with every human invention, AMI-based energy management has its own issues and challenges.

Chief among them is the cost implications. Design, deployment and maintenance of the smart meter system involve large investment in funds. Initially, the process of replacing the existing energy meters with a smart meter system will be a challenge for utility companies. Lack of proper infrastructure for synchronizing this new technology with the existing ones might interrupt the introduction of smart meters. Though several devices are integrated with the smart meter system, they can be used to their fullest extent only when all the appliances and devices in the distribution and metering network are included in the communication network (Soma, Wang, Devabhaktuni, & Gudi, 2011)

Another major challenge is the issue of theft. Several authors agree that energy theft continues to be a major issue to utility companies, even after deployment of smart meters. Stephen McLaughlin et al in their papers titled 'Energy theft in the Advanced Metering Infrastructure' (McLaughlin, Podkuiko, & McDaniel,

2010) and ‘AMIDS: A Multi-Sensor Energy Theft Detection Framework for Advanced Metering Infrastructures’ (McLaughlin, Holbert, Zonouz, & Berthier, 2012) insist that the single requirement of energy theft is the manipulation of the demand data. In their work they identified three ways to tamper with the demand data, which includes a) while it is recorded (via electromechanical tampering), b) while it is at rest in the meter, and c) as it is in flight across the network.

Anas et al (Anas, et al., 2012) describes electricity theft as basically an illegal way of getting the energy for different uses, resulting in loss for utility company. In their work they categorized losses into two broad groups viz technical and non-technical losses.

The technical losses are due to energy dissipated in the conductors, equipment used for transmission Line, Transformer, sub- transmission Line and distribution Line and magnetic losses in transformers (Parmar, 2013), while non-technical losses are defined as any consumed energy or service which is not billed because of measurement equipment failure or Ill-intentioned and fraudulent manipulation of said equipment (Monedero, Biscarri, & Leon, 2006). Whereas technical losses are easy to determine and calculate, non-technical losses are not.

Sudheer K. et al (Sudheer & etal, 2013) categorized electricity theft under several categories including:

- By under voltage technology
- By under current technology
- Stealing electricity by phase-shifted technology
- Stealing electricity by difference expansion technology.

Other methods of theft as identified by Anas et al (Anas, et al., 2012) include:

- Taking connections directly from distribution lines
- Grounding the neutral wire
- Putting a magnet (like neodymium) on electromechanical meter
- Inserting some disc to stop rotating of the coil
- Hitting the meter to damage the rotating coil
- Interchanging input and output connections

Also, Depuru S. (Depuru, 2012) identifies many techniques for such meter tampering including:

- Exposing meters to strong magnetic fields to wipe out the memory.
- Inserting a film or depositing high viscous fluid to disturb the rotation of disc.
- Implementation of sophisticated technologies like remote sensing devices.
- Tampering the crystal frequency of integrated circuits.
- Creating a link between the breaking control wires in an energy meter which would divert the current reading in the meter thereby reflecting zero reading at all times.
- In the case of electronic meters, Radio Frequency (RF) devices are mounted to affect the accuracy of the meter.
- A shunt is installed between the incoming and outgoing meter terminals.
- Inter-changing the incoming and outgoing terminals of the meter.
- Damaging the pressure coil of the meter.
- Resetting the meter reading.
- Exposing the meter to mechanical shock.
- Voltage is regulated from the meter terminals, making it read lesser quantity then the original consumption.

## 2.8 Minimising Non-Technical Losses Using AMI and ML Techniques

Nabi Mohammad et al (Mohammad, Barua, & Arafat, 2013) postulated measures for controlling electricity theft in their work “A smart prepaid energy metering system to control electricity theft”. According to them, measures like protection against shorting the phase line and disconnecting the neutral line, protection against whole meter bypassing, control of electricity theft using observer meter and protection against tampering can be achieved using a smart prepaid energy metering system and AMI.

Regardless of how they occur, it is becoming increasingly obvious that detection and subsequent prevention of non-technical losses cannot be done easily without the assistance of smart grids and smart meters. Hence it is no wonder that several authors have suggested different ways of detecting these losses.

Some of the recent suggestions include a proposal to use genetic algorithm and support vector machines (SVM) to detect abnormalities by Nagi et al (Nagi, Yap, Tiong, Ahmed, & Mohammad, 2008). Monedero et al (Monedero, Biscarri, & Leon, 2006) proposed the application of data mining techniques including use of neural networks and statistical techniques for these detections. Thomas Hartmann and co (Hartmann, et al., 2015) proposed the use of live machine learning using multi-profiling. Nizar and co (Nizar, Dong, & Wang, 2008) in the same vain presented a novel approach for analysing NTL using modern computational technique called extreme learning machine (ELM). Also Depuru S. (Depuru, 2012) employed several data classification algorithms including Support Vector Machines (SVM), Genetic Algorithm (GA), Neural Network (NN) model, and Rule Engine Algorithm in order to classify illegal consumers based on their energy consumption patterns.

## 2.9 Machine Learning Technique

Discussed below is the ML techniques used in the classification of consumer profile for detection of non-technical losses and prediction of usage pattern.

### 2.9.1 Support Vector Machines (SVM)

SVM (Support Vector Machines) is a machine learning method that is widely used for data analysis and pattern recognition. The algorithm was invented by Vladimir Vapnik and the current standard was proposed by Corinna Cortes and Vladimir Vapnik (Vapnik, 1998) (Cristianini & Taylor, 2000). The idea of support vector machines is to create a hyper plane in between data sets to indicate which class it belongs to. The challenge is to train the machine to understand structure from data and mapping with the right class label (Xie, 2011). The goal of SVM is to produce a model (based on the training data) which predicts the target values of the test data given only the test data attributes. It is to find the hyperplane (classifier) that maximizes the gap between data points on the boundaries (so called support vectors), assuming an infinite number of such hyperplanes exists (Mokshyna, 2014). For example, given the hyperplane as shown in Figure 2.7.

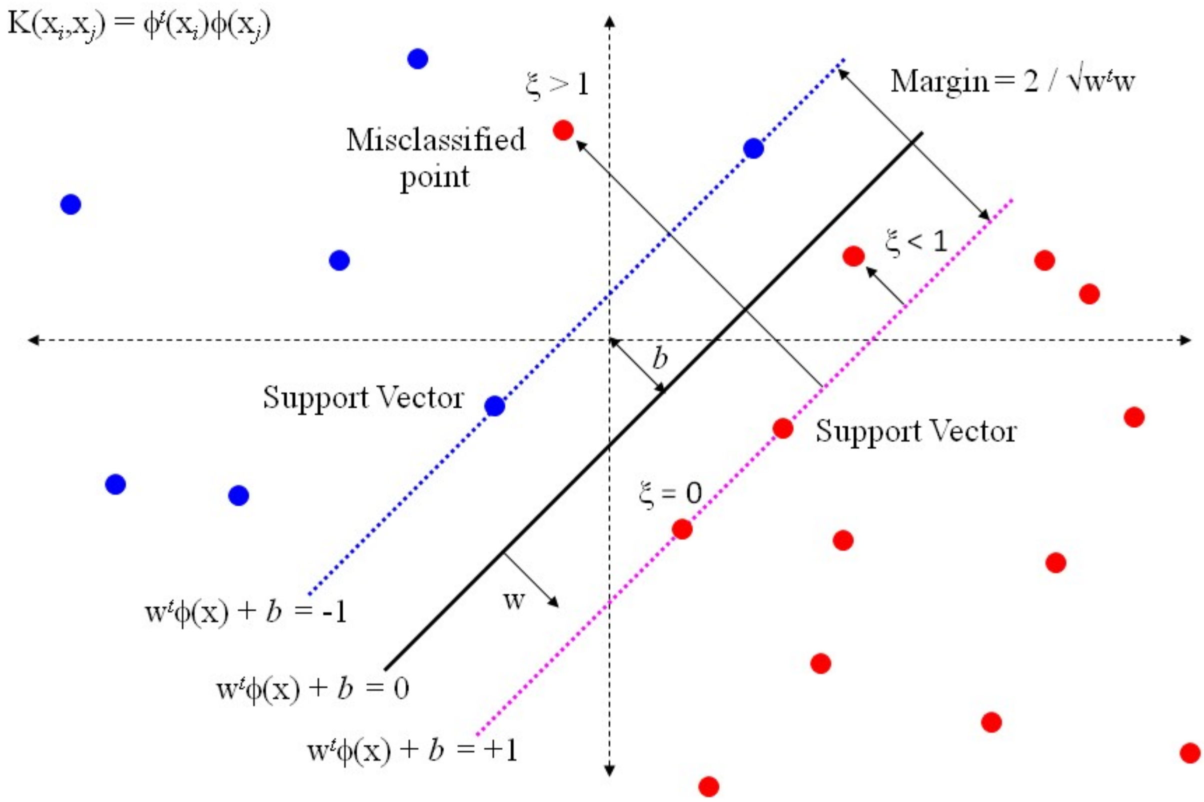



Figure 2.7: Support Vector Hyperplane (Mokshyna, 2014)

According to Nagi et al (Nagi, Yap, Tiong, Ahmed, & Mohammad, 2008), in SVM, training is performed in a way such to obtain a quadratic programming (QP) problem. The solution to this QP problem is global and unique. For empirical data $(x_1, y_1), \dots, (x_l, y_l) \in R^n * \{-1, +1\}$ that are mapped by $\varphi\colon R^n \to F$ into a "feature space", the linear hyperplanes that divide them into two labeled classes can be mathematically represented as:

$$w * \emptyset(x) + b = 0 \qquad w \in R^n, b \in R \tag{2.1}$$

To construct an optimal hyperplane with maximum-margin and bounded error in the training data (soft margin), the following QP problem is to be solved (Nagi, Yap, Tiong, Ahmed, & Mohammad, 2008):

$$\min_{w,b,\xi} \quad \frac{1}{2} w^T \mathrm{w} + C \sum_{i=1}^{l} \xi_i \tag{2.2}$$

Subject to $\quad y_i(w^T * \emptyset(x_i) + b) \geq 1 - \xi_i, \qquad (2.3)$

$$\xi \geq 0, i = 1, \dots, l$$

where $\emptyset(x_i)$ maps $x_i$ into a higher-dimensional space. The first term in cost function eqn. (2.2) makes maximum margin of separation between classes, and the second term provides an upper bound for the error in the training data. The constant $C \in [0,\infty)$, called the regularization parameter, creates a tradeoff between the number of misclassified samples in the training set and separation of the rest samples with maximum margin. Due to the possible high dimensionality of the vector variable $w$, usually the following dual problem is solve

$$\min_{\alpha} \ \frac{1}{2}\alpha^T Q\alpha - e^T\alpha$$

$$\text{Subject to} \quad y^T\alpha = 0 \tag{2.4}$$

$$0 \le \alpha_i \le C, \quad i = 1, \dots, l$$

Where $e = [1, \dots, 1]^T$ is the vector of all ones, $Q$ is an $l$ by $l$ positive semidefinite matrix, $Q_{ij} \equiv y_i y_j K(x_i, x_j)$, and $K(x_i, x_j) \equiv \emptyset(x_i)^T \emptyset(x_j)$ is the kernel function. After problem (eq. 2.4) is solved, using the primal-dual relationship, the optimal $w$ satisfies

$$w = \sum_{i=1}^{l} y_i \alpha_i \emptyset(x_i) \tag{2.5}$$

And the decision function is

$$sgn\,(w^T * \emptyset(x) + b) = sgn\left(\sum_{i=1}^{l} y_i \alpha_i K(x_i, x) + b\right) \tag{2.6}$$

$y_i\alpha_i\ \forall_i, b$, label names, support vectors and other information such as kernel parameters are stored for prediction

From eq. (2.1) it is seen that the optimal hyperplane in the feature space can be written as the linear combination of training samples with $\alpha_i \neq 0$. These informative samples, known as *support vectors*, construct the decision function of the classifier based on the kernel function:

$$f(x) = sgn\left(\sum_{i=1}^{m} y_i \alpha_i k(x, x_j) + b\right) \tag{2.7}$$

.LibSVM is a library for developing SVMs based on classification model developed by C.C. Chang and C.J. Lin (Chang & Lin, 2013). it could be adapted to most machine learning knowledge environments like Matlab, R, Pyton, WEKA (Waikato Environment for Knowledge Analysis) etc . LibSVM classifier is a wrapper class for the libsvm tools. It is used to build the SVM classifier and runs faster than other SVM tools.

SVM offers many advantages including, as listed by (Mokshyna, 2014), flexibility in the choice of the form of the threshold, robustness towards small number of data points, delivery of unique solution etc.

Despite its numerous advantages, SVM still suffers some setbacks including discrete data presentation problem and yet to be optimized design for multiclass SVM, it has slow test phase (Burges, 1998), and it is also computationally expensive (Mokshyna, 2014) etc.

## 2.10 Summary of Literature Review

From the review done in the previous sections, a few deductions could be adduced:

- The STS based electricity meter and the common vending system (CVS) are grossly inefficient in tackling the issue of non-technical losses in electricity distribution
- Although AMI and smart meters provide a significant advantage over the STS based meters, they are also susceptible to other forms of manipulations and bypass.
- The introduction of online transmission of electricity usage data through AMI also created a loophole especially via cyber-attack and data hijack

when these sensitive information are in transit from the meters to the control and data storage centers.

- The need to find a permanent solution for non-technical losses in electricity distribution and management has caught the attention of a number of researchers in recent times. Varied solutions have been postulated ranging from hardware additive approach to software redesign approach to machine learning based heuristics towards pattern recognition and usage prediction.
- Machine Learning approach to NTL detection, which seem to be the most plausible solution so far, still cannot guarantee hundred percent accuracy.
- There has not been a generally accepted, fool-proof methodology which would be able to holistically address this multi-faceted issue of non-technical losses in electricity distribution and management.

This research therefore aims at improving on Machine learning (ML) technique-based analysis of consumers' electricity consumption pattern in a bid to effectively minimise non-technical losses in AMI based systems.

# CHAPTER THREE

# METHODOLOGY

This chapter describes the steps taken to achieve the listed objectives of this work.

## 3.1 System Analysis

### 3.1.1 Data Acquisition

In order to monitor and correctly analyse electricity consumption of consumers, instantaneous electricity readings of each individual consumer has to be captured and transmitted to a central location for such analysis. This has been made possible by smart meters and AMI technology as elaborated in chapter two of this work.

However, as at the time of conducting this research, the electricity companies in Nigeria had not fully implemented smart grid network in the electricity sector. Survey showed that of all the households which had the prepaid meters installed before 2015, the meters installed had no smart capabilities (ie, they were simply STS prepaid meters with no communication capabilities to a central databank). Sometime in 2015, some of the electricity companies began rolling out the smart meters and installed to mostly new customers and a few of the old customers. However, the smart capabilities of these installed meters were turned off, hence there were no data readily available in Nigeria for analysis purposes.

In a view to surmount this problem, live electricity usage data recorded from 370 homes in California USA was collected and extracted through the assistance of the University of California Irvine (UCI) Machine Learning Repository (Lichman, 2015). The Dataset is a real instantaneous time-series electricity consumption of 370 clients taken at 15 minutes interval between January 2011 and December 2014. That is, each client has about 96 instances of their energy consumption taken daily for a period of 4 years. The energy readings are in kilowatts. A sample of the electricity usage data is in Appendix E.

### 3.1.2 Data Preparation and Feature Selection

Because using the whole 4 year data of 370 clients would be overbearing to the classifier, and also because some of the years (like 2011) have some missing data for some clients, the data for this analysis was streamlined to include only from 1st January 2012 through 31st December 2012 containing a total of 35131 attributes of the 370 clients (instances).

$$Total\ attrubutes, a_T = M_T \times N \times I \qquad 3.1$$

Where $M_T$ is the total number of Months under consideration, $N$ is the number of days in each Month and $I$ is the number of instantaneous electricity energy reading of each client taken at 15 minutes interval.

Out of the 35131 instances selected, a set containing 11808 instances were further selected as the training data.

$$\text{Hence, } a_{Tt} = \sum M_{i2} \times N \times I \qquad 3.2$$

Where $a_{Tt}$ represents the total training data and given that the months of the year are represented in a matrix S, in the form:

$$S = \begin{bmatrix} M_{11} & M_{12} & M_{13} & M_{14} \\ M_{21} & M_{22} & M_{23} & M_{24} \\ M_{31} & M_{32} & M_{33} & M_{34} \\ M_{41} & M_{42} & M_{43} & M_{44} \end{bmatrix} \qquad 3.3$$

The $M_{i2}$ formation hence selects the second month of the beginning of each season of the year. This was done so because electric energy consumption of users are affected by the season of the year, hence the dataset selected for training the classifier comprise of readings from January (Winter) , April (Spring), July (Summer) and October (Autumn) in order to capture variations in the different seasons of the year. The rest of the dataset (that is, $a_T - a_{Tt}$) was then used as the testing data.

Since the collected data is not in the format of the classifier input (see classifier selection in section 3.2.1), the data was further subjected to a transposition hence from equation 3.2,

$$a_{Tt} = (a_{Tt})^T \quad 3.4$$

Where $(a_{Tt})^T$ is the transpose values of $a_{Tt}$.

### 3.1.3 Data Analysis

To be able to understand the dataset properly and hence properly utilize the information embedded in it, a series of analysis had been conducted on the dataset.

First, the 370 consumers were grouped based on the understanding of the category they fall under according to their consumption capacity. According to the research done by Depuru (Depuru, 2012) and as correlated by other sources like the US Energy Information Administration (U.S Energy Information Administration (EIA), 2016) and OVO energy (Ovo Energy, 2016), consumers can be divided into three main groups; Agricultural, Commercial and Residential Customers. Table 3.2 shows a profile of the customer types and their daily average consumption in both kW and kWh.

Table 3.1: Customer Types and their Daily Average Consumptions

| Customer type | Daily consumption (kW) | Daily consumption (kWh) |
|---|---|---|

| | | |
|---|---|---|
| Large Commercial customers | >2000kW | >70kWh |
| Small Commercial customers | 700-2000kW | 50-70kWh |
| Large Residential Customers | 300-600kW | 10-50kWh |
| Small Residential | 100-300kW | 3.5-10kWh |
| Large Agricultural Customers | 100-300kW | 3.5-10kWh |
| Small Agricultural Customers | (<100kW | <3.5kWh |

Based on the information in Table 3.1, the 370 clients used for this research was then grouped into their various types and shown in Table 3.2 and Figure 4.1.

Table 3.2: Frequency of the different groups of consumers

| Types | Frequency |
|---|---|
| Large Commercial | 101 |
| Large Residential | 156 |
| Small Agricultural | 52 |
| Small Commercial | 27 |
| Small Residential/Large Agricultural | 34 |

Table 3.3 and Figure 3.2 shows the average energy consumption of the different consumer types represented in the dataset.

Table 3.3: Average daily consumption of different consumer types

| Group | Average consumption (kw/h) |
|---|---|
| Small Agriculture | 2.14 |
| Small Residence/ Large Agricultural | 7.19 |
| large Residence | 27.12 |
| Small Commercial | 59.11 |
| Large Commercial | 361.82 |

Further analysis were conducted to show samples of the represented customer types and their maximum and minimum consumption values. This information is shown in tables 3.4 – 3.8 and figures 3.3 to 3.7.

Table 3.4: Small Agricultural Consumers Showing Maximum and Minimum Consumption in kW

| Consumer | Min (kw) | Max (kw) | Type |
|---|---|---|---|
| **MT_003** | 0.87 | 151.17 | Small Agricultural |
| **MT_007** | 0.57 | 35.61 | Small Agricultural |
| **MT_019** | 3.02 | 22.61 | Small Agricultural |
| **MT_023** | 0.66 | 27.70 | Small Agricultural |
| **MT_080** | 3.80 | 46.84 | Small Agricultural |
| **MT_096** | 1.74 | 41.16 | Small Agricultural |
| **MT_141** | 1.03 | 57.49 | Small Agricultural |
| **MT_187** | 0.75 | 150.63 | Small Agricultural |

Table 3.5: Small Residential (and Large agricultural) Consumers Showing Maximum and Minimum Consumption in kW

| Consumer | Min (kw) | Max(kw) | Type |
|---|---|---|---|
| **MT_002** | 0.71 | 115.22 | Small Residential |
| **MT_016** | 12.53 | 80.30 | Small Residential |
| **MT_022** | 0.62 | 69.74 | Small Residential |
| **MT_026** | 7.63 | 66.95 | Small Residential |
| **MT_027** | 6.70 | 51.77 | Small Residential |
| **MT_036** | 6.88 | 123.92 | Small Residential |
| **MT_044** | 0.71 | 83.39 | Small Residential |
| **MT_047** | 8.47 | 70.44 | Small Residential |
| **MT_052** | 4.63 | 87.96 | Small Residential |
| **MT_055** | 8.39 | 97.83 | Small Residential |
| **MT_056** | 13.74 | 93.13 | Small Residential |
| **MT_060** | 11.58 | 58.41 | Small Residential |
| **MT_063** | 0.80 | 71.66 | Small Residential |
| **MT_079** | 7.80 | 129.70 | Small Residential |
| **MT_093** | 1.02 | 249.75 | Small Residential |

| | | | |
|---|---|---|---|
| **MT_094** | 1.52 | 245.07 | Small Residential |
| **MT_095** | 0.52 | 123.10 | Small Residential |
| **MT_125** | 3.69 | 96.75 | Small Residential |
| **MT_135** | 6.21 | 68.80 | Small Residential |
| **MT_149** | 0.65 | 61.28 | Small Residential |
| **MT_169** | 15.27 | 64.81 | Small Residential |
| **MT_188** | 2.03 | 51.82 | Small Residential |

Table 3.6: Large Residential Consumers Showing Maximum and Minimum Consumption in kW

| Consumer | Min (kw) | Max(kw) | Type |
|---|---|---|---|
| **MT_005** | 14.63 | 120.73 | Large Residential |
| **MT_006** | 2.98 | 491.07 | Large Residential |
| **MT_011** | 14.16 | 100.60 | Large Residential |
| **MT_013** | 6.63 | 169.98 | Large Residential |
| **MT_014** | 14.68 | 130.82 | Large Residential |
| **MT_017** | 4.49 | 110.51 | Large Residential |
| **MT_021** | 65.45 | 421.47 | Large Residential |
| **MT_029** | 0.63 | 152.16 | Large Residential |
| **MT_035** | 2.81 | 285.51 | Large Residential |
| **MT_037** | 48.36 | 292.07 | Large Residential |
| **MT_038** | 53.66 | 240.84 | Large Residential |
| **MT_040** | 43.96 | 337.91 | Large Residential |
| **MT_045** | 1.69 | 241.96 | Large Residential |
| **MT_046** | 37.91 | 345.97 | Large Residential |
| **MT_050** | 49.82 | 427.05 | Large Residential |
| **MT_051** | 15.99 | 140.32 | Large Residential |
| **MT_053** | 25.85 | 211.52 | Large Residential |
| **MT_054** | 32.37 | 146.88 | Large Residential |
| **MT_058** | 21.21 | 234.91 | Large Residential |

Table 3.7: Small Commercial Consumers Showing Maximum and Minimum Consumption in kW

| Consumer | Min (kw) | Max (kw) | Type |
|---|---|---|---|
| **MT_008** | 107.74 | 535.35 | Small Commercial |
| **MT_018** | 70.29 | 616.61 | Small Commercial |
| **MT_025** | 74.77 | 574.77 | Small Commercial |
| **MT_068** | 4.64 | 580.05 | Small Commercial |
| **MT_084** | 63.29 | 525.32 | Small Commercial |
| **MT_105** | 21.36 | 648.96 | Small Commercial |
| **MT_124** | 9.57 | 722.49 | Small Commercial |
| **MT_171** | 64.89 | 415.71 | Small Commercial |
| **MT_198** | 40.13 | 477.67 | Small Commercial |
| **MT_202** | 0.56 | 443.94 | Small Commercial |
| **MT_206** | 46.39 | 385.45 | Small Commercial |
| **MT_217** | 23.85 | 537.99 | Small Commercial |
| **MT_242** | 1.88 | 384.32 | Small Commercial |
| **MT_252** | 1.92 | 437.19 | Small Commercial |
| **MT_254** | 1.48 | 426.27 | Small Commercial |
| **MT_260** | 6.00 | 375.21 | Small Commercial |
| **MT_262** | 4.66 | 404.07 | Small Commercial |
| **MT_268** | 71.28 | 528.49 | Small Commercial |
| **MT_283** | 1.70 | 425.23 | Small Commercial |

Table 3.8: Large Commercial Consumers Showing Maximum and Minimum Consumption in kW

| Consumer | Min (kw) | Max (kw) | Type |
|---|---|---|---|
| **MT_043** | 22.73 | 2659.09 | Large Commercial |
| **MT_064** | 435.90 | 3025.64 | Large Commercial |
| **MT_086** | 147.37 | 1421.05 | Large Commercial |

| | | | |
|---|---|---|---|
| **MT_098** | 18.12 | 1792.42 | Large Commercial |
| **MT_099** | 167.97 | 3574.22 | Large Commercial |
| **MT_101** | 127.18 | 3080.14 | Large Commercial |
| **MT_102** | 108.32 | 2261.56 | Large Commercial |
| **MT_103** | 30.51 | 2500.00 | Large Commercial |
| **MT_104** | 5.94 | 1811.88 | Large Commercial |
| **MT_114** | 122.52 | 751.66 | Large Commercial |
| **MT_128** | 30.67 | 792.44 | Large Commercial |
| **MT_139** | 78.93 | 802.69 | Large Commercial |
| **MT_147** | 400.00 | 5229.27 | Large Commercial |
| **MT_148** | 48.01 | 688.97 | Large Commercial |
| **MT_157** | 51.87 | 3340.06 | Large Commercial |
| **MT_163** | 87.56 | 4923.96 | Large Commercial |
| **MT_166** | 13.74 | 2085.49 | Large Commercial |
| **MT_189** | 63.09 | 804.42 | Large Commercial |
| **MT_190** | 2.93 | 734.90 | Large Commercial |

Apart from the size and type of the customer, there are a number of other factors which affect and impact electricity consumption directly. Some of the major ones include geographic location, season of the year, weather conditions, and time of day. In order to appreciate the effect these factors have on energy consumption, the consumption profile of randomly selected clients from each of the customer types were plotted against the time of day through the four seasons of the year (ie Spring (April), Summer (July), Autumn (October) and Winter (January), both on a weekday (Monday) and a weekend (Sunday). The second month into each of the official start date of each season was selected to ensure accuracy of results gotten. The results of these plots are shown in Figure 4.8 – Figure 4.12.

## 3.2 System Modelling

### 3.2.1 Classifier Selection

A number of research techniques exists for adoption when carrying out a classification analysis as is the case of this work. As a matter of fact, the work of Manuel Fernandez-Delgado and co (Fernandez-Delgado, Cernadas, Barro, & Amorim, 2014) evaluated 179 different classifiers arising from 17 families and these are just the common ones as there are many more. Hence in choosing a suitable technique for this work, various criteria like availability of support resource, ease of manipulation, openness of source, as well as of course performance matrix of these classifiers were considered. From the work of Delgado et al, the Random forest (RF) version "are most likely to be the best classifiers" when "implemented in R and accessed via caret" (Fernandez-Delgado, Cernadas, Barro, & Amorim, 2014). The second best is the SVM with Gaussian kernel implemented in C using LibSVM followed by the neural networks.

In his article titled "A Tour of Machine Learning Algorithms", Jason Browniee (Browniee, 2017) identified SVM as among the top 5 classifiers (see Table 3.9).

Table 3.9: Five Top Machine Learning Classifiers (Browniee, 2017)

| S/N | Algorithm | Type |
|---|---|---|
| **1** | Logistic Regression | Regression |
| **2** | Support Vector Machines (SVM) | Vector-space based |
| **3** | Naive Bayes | Bayesian |
| **4** | k-Nearest Neighbours (KNN) | Instance Based |
| **5** | Classification and Regression Trees (CART) | Decision Tree |

In a bid to ascertain the most appropriate classifier for the dataset used for this work (see section 3.1), nine of the top listed classifiers were subjected to a paired corrected tester using Waikato Environment for Knowledge Analysis (WEKA) (The University of Waikato, 2013). The result is given in Table 3.10 and Table 3.11 while a column chart is shown in Figure 4.13.

Table 3.10: Paired Corrected Tester for Nine Data Classifiers

Tester: weka.experiment.PairedCorrectedTTester

Analysing: Percent_correct

Datasets: 1

Resultsets: 7

Confidence: 0.05 (two tailed)

Sorted by: -

Date: 3/15/17 8:32 PM

| Dataset | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| 4SeasonConsumption | 69.46 | 76.97 v | 70.49 | 76.11 v | 69.46 | 79.68 v | 19.62 | 50.85 | 69.45 |

Key:

(1) rules.ZeroR '' 48055541465867954

(2) functions.LibSVM '-S 0 -K 2 -D 3 -G 0.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -seed 1' 14172

(3) functions.SMO '-C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K \"functions.supportVector.PolyKernel -C 250007 -E 1.0\"' -6585883636378691736

(4) lazy.IBk '-K 1 -W 0 -A \"weka.core.neighboursearch.LinearNNSearch -A \\\"weka.core.EuclideanDistance -R first-last\\\"\"' -3080186098777067172

(5) trees.REPTree '-M 2 -V 0.001 -N 3 -S 1 -L -1' -9216785998198681299

(6) trees.RandomForest '-I 10 -K 0 -S 1' -2260823972777004705

(7) bayes.NaiveBayes '' 5995231201785697655

(8) functions.MultilayerPerceptron '-L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H 0' -5990607817048210779

(9) trees.SimpleCart '-S 1 -M 2.0 -N 5 -C 1.0' 4154189200352566053

Table 3.11: Result of Accuracy Test for Nine Data Classifier

| S/No | Classifier | Type | CV Accuracy |
|---|---|---|---|
| 1 | Zero Rules | Rule System | 69.46 |
| 2 | LibSVM | Support Vector Machines | 76.97 |
| 3 | Sequential Minimal Optimization (SMO) | Support Vector Machines | 70.49 |
| 4 | IBk (K-Nearest Neighbour) | Instance Based | 76.11 |
| 5 | REPTree | Decision Trees | 69.46 |

| | | | |
|---|---|---|---|
| 6 | Random Forest | Ensemble | 79.68 |
| 7 | Naïve Bayes | Bayesian | 19.62 |
| 8 | Multilayer Perceptron | Neural Networks | 50.85 |
| 9 | Simple CART | Decision Trees | 69.45 |

From Table 3.11, it could be observed that classifier RandomForest performed best (79.68) followed closely by LibSVM (76.97). However, RandomForest gives its best performance in the R environment. So against these backdrops, the LibSVM classifier was chosen for classifying the dataset, training the system and also for performing predictive analysis for the given dataset.

### 3.2.2 Support Vector Machines (SVM) Classifier

In classification of the electricity consumption profile of the 370 customers selected (see section 3.1), the optimization problems (given in section 2.9.1) was solved and decision functions were applied to predict labels (target values) of testing data. Let $x_1, \ldots, x_l$ be the testing data and $f(x_1), \ldots, f(x_l)$ be decision values predicted by LIBSVM. If the true labels (true target values) of testing data are known and denoted as $y_1, \ldots, y_l$, the prediction results can be evaluated by the following measures.

For classification, $$Accuracy = \frac{\#\, correctly\ predicted\ data}{\#\, total\ testing\ data} \times 100\% \qquad (3.5)$$

All LIBSVM's training and testing algorithms are implemented in the file libSVM.java. The two main classes are "buildClassifier" (which trains the model) and "distributionForInstance" (which is used to make predictions). The codes are shown in appendix C.

### 3.2.3 Solving Quadratic Problem in One Linear Constraint.

To solve the quadratic problem with one linear constraint, the following general form of C-SVC was considered:

$$\min_{\alpha} \quad f(\alpha)$$

Subject to $\quad y^T\alpha = \Delta,$ (3.6)

$$0 \leq \alpha_t \leq C, t = 1, \dots, l,$$

where

$$f(\alpha) \equiv \frac{1}{2}\alpha^T Q\alpha + p^T\alpha$$

and $y_t = \pm 1, t = 1, \dots, l$. The constraint $y^T\alpha = 0$ is called a linear constraint. It can be clearly seen that equation (2.2) is in the form of equation (3.6). Q is not assumed to be positive semi-definite (PSD) because sometimes non-PSD kernel matrices are used.

According to (Chang & Lin, 2013) the main difficulty in solving the problem of equation (3.6) is that Q is a dense matrix and may be too large to be stored. LIBSVM considers a decomposition method to conquer this difficulty. A decomposition method modifies only a subset of α per iteration, so only some columns of Q are needed. This subset of variables, denoted as the working set B, leads to a smaller optimization sub-problem. An extreme case of the decomposition methods is the Sequential Minimal Optimization (SMO) (Platt, 1998), which restricts B to have only two elements. Then, at each iteration, a simple two-variable problem is solved without needing any optimization software. LIBSVM considers an SMO-type decomposition method proposed in (Fan, Chen, & Lin, 2005) and is as follows:

1. Find $\alpha^1$ as the initial feasible solution. Set k = 1.
2. If $\alpha^1$ is a stationary point of equation (2.4), stop. Otherwise, find a two-element working set $B = \{i, j\}$ by Working Set Selection (WSS) described in section 3.4.2. Define $N \equiv \{1, \dots, l\}\backslash B$. Let $\alpha_B^k$ and $\alpha_N^k$ be sub-vectors of $\alpha^k$ corresponding to B and N, respectively.

3. If $a_{ij} \equiv K_{ii} + K_{jj} - 2K_{ij} > 0$, $\quad$ ($K(x_i, x_j)$ is abbreviated to $K_{ij}$)

   Solve the following sub-problem with the variable $\alpha_B = \left[\alpha_i \; \alpha_j\right]^T$.

$$\min_{\alpha_i \alpha_j} \quad \frac{1}{2}\left[\alpha_i \; \alpha_j\right]\begin{bmatrix} Q_{ii} & Q_{ij} \\ Q_{ij} & Q_{jj} \end{bmatrix}\begin{bmatrix} \alpha_i \\ \alpha_j \end{bmatrix} + (P_B + Q_{BN}\alpha_N^K)^T \begin{bmatrix} \alpha_i \\ \alpha_j \end{bmatrix}$$

$$\text{Subject to} \quad 0 \le \alpha_i, \alpha_j \le C, \qquad (3.7)$$

$$y_i\alpha_i + y_j\alpha_j = \Delta - y_N^T \alpha_N^k,$$

   else

   Let $\tau$ be a small positive constant and solve

$$\min_{\alpha_i \alpha_j} \quad \frac{1}{2}\left[\alpha_i \; \alpha_j\right]\begin{bmatrix} Q_{ii} & Q_{ij} \\ Q_{ij} & Q_{jj} \end{bmatrix}\begin{bmatrix} \alpha_i \\ \alpha_j \end{bmatrix} + (P_B + Q_{BN}\alpha_N^K)^T \begin{bmatrix} \alpha_i \\ \alpha_j \end{bmatrix} + \frac{\tau - a_{ij}}{4}\left(\left(\alpha_i - \alpha_i^k\right)^2 + \left(\alpha_j - \alpha_j^k\right)^2\right) \qquad (3.8)$$

   Subject to constraints of (eq. 3.7)

4. Set $\alpha_B^{k+1}$ to be the optimal solution of the sub-problems in eqs. (3.7) and (3.4), and $\alpha_N^{k+1} \equiv \alpha_N^k$ . Set $k \leftarrow k + 1$ and go to step 2.

Note that $B$ is updated at each iteration, but for simplicity, $B$ is used instead of $B^k$. If $Q$ is PSD, then $a_{ij} > 0$. Thus eq. (3.8) is used only to handle the situation where $Q$ is non-PSD

### 3.2.4 Working Set Selection (WSS)

For the selection of the working set B, the following procedure from Section II of (Fan, Chen, & Lin, 2005) is used.

1. For all t, s, define

$$a_{ts} \equiv K_{tt}K_{ss} - 2K_{ts}, \; b_{ts} \equiv -y_t\nabla_t f(\alpha^k) + y_s\nabla_s f(\alpha^k) > 0, \qquad (3.9)$$

   And $\quad \bar{a}_{ts} \equiv \begin{cases} a_{ts} & if \; a_{ts} > 0, \\ \tau & otherwise. \end{cases}$

Select $\quad i \in \arg\max_t \{-y_t \nabla_t f(\alpha^k) \mid t \in I_{up}(\alpha^k)\},$

$$j \in \arg\min_t \left\{-\frac{b_{it}^2}{\bar{a}_{it}} \mid t \in I_{low}(\alpha^k), -y_t \nabla_t f(\alpha^k) < -y_i \nabla_i f(\alpha^k)\right\}. \tag{3.10}$$

2. Return $B = \{i, j\}$.

The procedure selects a pair $\{i, j\}$ approximately minimizing the function value; see the term $-b_{it}^2/\bar{a}_{it}$ in eq. (3.10).

### 3.2.5 Multi-Class Classification

According to (Chang & Lin, 2013) LIBSVM implements the “one-against-one" approach as originally proposed by (Knerr, Personnaz, & Dreyfus, 1990) for multi-class classification. If $k$ is the number of classes, then $k(k-1)/2$ classifiers are constructed and each one trains data from two classes. For training data from the $ith$ and the $jth$ classes, the following two-class classification problems are solved

$$\min_{w^{ij}, b^{ij}, \xi^{ij}} \quad \frac{1}{2}\left(w^{ij}\right)^T w^{ij} + C \textstyle\sum_t \left(\xi^{ij}\right)_t \tag{3.11}$$

Subject to $\quad \left(w^{ij}\right)^T \emptyset(x_t) + b^{ij} \geq 1 - \xi_t^{ij},\ if\ x_t\ in\ the\ ith\ class,$

$$\left(w^{ij}\right)^T \emptyset(x_t) + b^{ij} \leq -1 + \xi_t^{ij},\ if\ x_t\ in\ the\ jth\ class,$$

$$\xi_t^{ij} \geq 0.$$

In classification a voting strategy is used: each binary classification is considered to be a voting where votes can be cast for all data points $x$ - in the end a point is designated to be in a class with the maximum number of votes.

In case that two classes have identical votes, the class appearing first in the array of storing class names is simply chosen.

### 3.2.6 Kernel Selection

Kernel functions in SVMs are selected based on the data structure and type of boundaries between the classes. The representative and widely applied kernel function based on Euclidean distance is the radial basis function (RBF) kernel, also known as the Gaussian kernel (Wang, Yeung, & Tsang, 2007):

$$K^{RBF}(x_{i,}x_j) = exp\left(-\gamma\|x_i - x_j\|^2\right) \tag{3.12}$$

Where $\gamma$ > 0 is the RBF kernel parameter. The RBF kernel induces an infinite-dimensional kernel space, in which all image vectors have the same norm, and the kernel width parameter "$\gamma$" controls the scaling of the mapping (Wang, Yeung, & Tsang, 2007)

However, to ascertain the exact kernel most suited for the dataset used for this research work, the four kernel types of SVM was subjected to a paired Corrected Testing. The result of the test is given in Table 3.12 and Figure 4.14. It could be seen from that table that for the dataset, the RBF and Sigmoid kernel had the best cross validation accuracy of 79.46%. Therefore, the Radial Basis Function kernel was selected for the kernel for use in the SVM classifier.

Table 3.12: Paired Corrected Tester for Four SVM Kernel Types

| | | | | |
|---|---|---|---|---|
| Tester: weka.experiment.PairedCorrectedTTester<br>Analysing: Percent_correct<br>Datasets: 1<br>Resultsets: 4<br>Confidence: 0.05 (two tailed) | | | | |
| Dataset | (1) | (2) | (3) | (4) |
| 4SeasonConsumption | 79.46 | 72.86 | 71.41 | 79.46 |
| Key:<br>(1) functions.LibSVM 'Radial Basis Function: $exp(-\gamma\|u - v\|^2)$<br>(2) functions.LibSVM ' Linear: $u' \times v$<br>(3) functions.LibSVM ' Polynomial: $(\gamma \times u' \times v + coef0)^{degree}$ | | | | |

| |
|---|
| (4) functions.LibSVM ' Sigmoid: $\tanh(\gamma \times u' \times v + coef0)$ |

### 3.2.7 Parameter Selection

To train SVM problems, some parameters must be specified. As explained by (Chang & Lin, 2013) LIBSVM provides a simple tool to check a grid of parameters. For each parameter setting, LIBSVM obtains cross-validation (CV) accuracy. Finally, the parameters with the highest CV accuracy are returned. The parameter selection tool assumes that the RBF (Gaussian) kernel is used. The RBF kernel takes the form of eq. (2.8) so $(C, \gamma\ )$ are parameters to be decided.

The cost parameter C is the parameter for the soft margin cost function, which controls the influence of each individual support vector; this process involves trading error penalty for stability. It "tunes" the algorithm between better fitting the available data or giving a larger margin.

Gamma, $\gamma$, is the free parameter of the Gaussian radial basis function. A small $\gamma$ means a Gaussian with a large variance so the influence of $x_j$ is more, i.e. if $x_j$ is a support vector, a small $\gamma$ implies the class of this support vector will have influence on deciding the class of the vector $x_i$ even if the distance between them is large. If $\gamma$ is large, then variance is small implying the support vector does not have wide-spread influence. Technically speaking, large $\gamma$ leads to high bias and low variance models, and vice-versa.

Possible intervals of $C\ (or\ \gamma)$ were provide with the grid space. Then, all grid points of $(C,\ \gamma)$ are tried to find the one giving the highest CV accuracy (see Table 3.13). Then the best parameters were used to train the whole training set (see section 3.1) and generate the final model. From Table 3.13, it could be observed that $(C, \gamma\ ) \equiv (1.0, \geq 1)$ gave the highest CV accuracy.

More advanced parameter selection methods were not considered because for only two parameters $(C\ and\ \gamma)$, the number of grid points was not too large. For multi-class classification, under a given $(C,\ \gamma)$, LIBSVM uses the one-against-one method to obtain the CV accuracy. Hence, the parameter selection tool suggests the same $(C,\ \gamma)$ for all $k(k\ -\ 1)/2$ decision functions.

Table 3.13: Determination of Best $(C, \gamma\ )$ for Highest CV Accuracy.

Tester: weka.experiment.PairedCorrectedTTester

Analysing: Percent_correct

Datasets: 1

Resultsets: 10

Confidence: 0.05 (two tailed)

Sorted by: -

Date: 4/11/17 10:54 AM

| Dataset | (1 | (2 | (3 | (4 | (5 | (6 | (7 | (8 | (9 | (10 |
|---|---|---|---|---|---|---|---|---|---|---|
| '4SeasonConsumption_abrid | 39 | 39 | 39 | 39 | 50 | 47 | 50 | 50 | 50 | 50 |

Key:

(1) functions.LibSVM '-S 0 -K 2 -D 3 -G 0.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -seed 1' 14172

(2) functions.LibSVM '-S 0 -K 2 -D 3 -G 0.0 -R 0.0 -N 0.5 -M 40.0 -C 3.0 -E 0.001 -P 0.1 -seed 1' 14172

(3) functions.LibSVM '-S 0 -K 2 -D 3 -G 0.0 -R 0.0 -N 0.5 -M 40.0 -C 30.0 -E 0.001 -P 0.1 -seed 1' 14172

(4) functions.LibSVM '-S 0 -K 2 -D 3 -G 0.0 -R 0.0 -N 0.5 -M 40.0 -C 300.0 -E 0.001 -P 0.1 -seed 1' 14172

(5) functions.LibSVM '-S 0 -K 2 -D 3 -G 0.01 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -seed 1' 14172

(6) functions.LibSVM '-S 0 -K 2 -D 3 -G 0.001 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -seed 1' 14172

(7) functions.LibSVM '-S 0 -K 2 -D 3 -G 1.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -seed 1' 14172

(8) functions.LibSVM '-S 0 -K 2 -D 3 -G 3.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -seed 1' 14172

(9) functions.LibSVM '-S 0 -K 2 -D 3 -G 30.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -seed 1' 14172

(10) functions.LibSVM '-S 0 -K 2 -D 3 -G 300.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -seed 1' 14172

## 3.3 Model Training

From the Primal SVM model given in eq. (2.2) and the Gaussian kernel given in eq. (3.12) ie

$$\min_{w,b} \frac{1}{2} |w|^2 + C \sum_{i=1}^{m} \xi_i \qquad (3.13)$$

$$K^{RBF}(x_i, x_j) = exp\left(-\gamma \|x_i - x_j\|^2\right) \qquad (3.14)$$

Utilizing the C-SVC SVM type, with the following parameters:

Cache size =40Mb,

Coefficient = 1,

Cost parameter C for C-SVC= 1.0,

Gamma $\gamma$ = 1.0,

$K^{RBF} = exp(-\gamma * |x_i - x_j|^2)$,

Weight for each class = (1 1 1 1)

Cross validation fold = 10

With standardized data, non-shrinking heuristic, the model was trained, subjecting it to a set of chosen attributes, cutting across the four seasons of the year (section 3.1.3 Data Analysis).

### 3.4 Data Classification Using Trained Model

Based on the model parameters given in section 3.5, the classifier haven been adequately trained, was used to classify the given data set. The summary of the classification of the 370 instances using the trained model is given in Table 3.14

Table 3.14: Summary of LibSVM Classifier Output on 370 Instances

```
=== Run information ===

Scheme:weka.classifiers.functions.LibSVM -S 0 -K 2 -D 3 -G 1.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0
-E 0.001 -P 0.1 -seed 1

Relation:     4SeasonConsumption

Instances:    370

Attributes:   11808

Test mode: evaluate on training data

=== Classifier model (full training set) ===

LibSVM wrapper, original code by Yasser EL-Manzalawy (= WLSVM)

Time taken to build model: 32.27 seconds



=== Summary ===
```

| | | |
|---|---|---|
| Correctly Classified Instances | 294 | 79.4595 % |
| Incorrectly Classified Instances | 76 | 20.5405 % |
| Kappa statistic | 0.442 | |
| Mean absolute error | 0.1027 | |
| Root mean squared error | 0.3205 | |
| Relative absolute error | 42.2306 % | |
| Root relative squared error | 92.1927 % | |
| Total Number of Instances | 370 | |

=== Detailed Accuracy By Class ===

| | TP Rate | FP Rate | Precision | Recall | F-Measure | ROC Area | Class |
|---|---|---|---|---|---|---|---|
| | 0 | 0 | 0 | 0 | 0 | 0.5 | RED |
| | 1 | 0.655 | 0.776 | 1 | 0.874 | 0.673 | GREEN |
| | 1 | 0.006 | 0.949 | 1 | 0.974 | 0.997 | BLACK |
| Weighted Avg. | 0.795 | 0.455 | 0.634 | 0.795 | 0.705 | 0.67 | |

=== Confusion Matrix ===

| a | b | c | d | | <-- classified as |
|---|---|---|---|---|---|
| 0 | 74 | 0 | 2 | \| | a = RED |
| 0 | 257 | 0 | 0 | \| | b = GREEN |
| 0 | 0 | 0 | 37 | \| | d = BLACK |

### 3.5 System Implementation: Class Prediction Using LIBSVM

Out of the total dataset selected for analysis (section 3.1.2), a further breakdown was done which filtered out four months representing the four seasons of the year (section 3.1.3). This was used as the training data whereas the remaining data is applied as testing data for prediction. Having successfully trained the model using the training data, the testing data was applied to it in order to ascertain the accuracy of the classifier's training and its ability to predict the class of each instance. Table 3.15 - Table 3.17 show the output of the prediction run. The preliminary information including classifier scheme with its attributes and the number of instances and attributes associated with the supplied test data is given in Table 3.15. A sample of the actual run information (20 instances) containing the predicted classes, error information and probability distribution of the classification can be seen in Table 3.16 (the entire predictions table is given in Appendix A), whereas the summary of the run showing that 294 instances representing 79.45% of the total instances were correctly classified with a mean absolute error of 0.1027 is shown in Table 3.17

Table 3.15: Prediction Run Preliminary Information

| |
|---|
| === Model information === |
| Filename: libsvm-final run model.model |
| Scheme:weka.classifiers.functions.LibSVM -S 0 -K 2 -D 3 -G 0.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -seed 1 |
| Relation:4SeasonConsumption |
| Attributes: 11808 |
| [list of attributes omitted] |
| === Classifier model === |
| LibSVM wrapper, original code by Yasser EL-Manzalawy (= WLSVM) |

=== Re-evaluation on test set ===

User supplied test set

Relation: prediction data

Instances: 370

Attributes: 11808

Table 3.16: Sample of Prediction Statistics

| === Predictions on test set === | | | | | | | |
|---|---|---|---|---|---|---|---|
| inst#, | actual, | predicted, | error, | probability distribution | | | |
| 1 | 1:RED | 2:GREEN | + | 0 | *1 | 0 | 0 |
| 2 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 3 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 4 | 3:RED | 2:GREEN | + | 0 | *1 | 0 | 0 |
| 5 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 6 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 7 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 8 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 9 | 1:RED | 2:GREEN | + | 0 | *1 | 0 | 0 |
| 10 | 3:RED | 2:GREEN | + | 0 | *1 | 0 | 0 |
| 11 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 12 | 4:BLACK | 4:BLACK | | 0 | 0 | 0 | *1 |
| 13 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 14 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 15 | 4:BLACK | 4:BLACK | | 0 | 0 | 0 | *1 |
| 16 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 17 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 18 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 19 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 20 | 3:RED | 2:GREEN | + | 0 | *1 | 0 | 0 |

Table 3.17: Summary Information for the Prediction Run

=== Summary ===

Correctly Classified Instances 294 79.4595 %

Incorrectly Classified Instances 76 20.5405 %

Kappa statistic 0.4608

| | |
|---|---|
| Mean absolute error | 0.1027 |
| Root mean squared error | 0.3205 |
| Total Number of Instances | 370 |

It could be seen that 79.46% accuracy can be achieved in terms of predicting the class of users in based on their energy usage and utilization patterns (Table 3.17). This is a significant step in the right direction as such will go a long way in early detection of inconsistent use (characterising fraud perhaps) of energy supplied. This could then be further inspected by utility personnel for confirmation purposes and hence prevent unnecessary wastage of generated power.

## 3.6 Pre-classifier Design and Development

In order to achieve an improvement with the SVM classifier output (ie 79.46% accuracy of classification achieved in section 3.4 and hence enhance its efficiency, it became necessary to pre-classify each user at the instance of input of each data into the system. Achieving this would require an automated pre-classifier which has real-time capabilities.

### 3.6.1 Pre-classifier Design – Flow Algorithm

To develop a pre-classifier with real-time capabilities, some criteria need to be applied to each input from the Smart meter:

i. If any of the instantaneous energy inputs is zero, flag the client as suspicious, else if there were no zero inputs, flag as valid customer and designate green.

ii. If there had not been any other zero inputs in the last 24 hours, designate orange and mark for close monitoring.
iii. If there had been other zero inputs in the last 24 hours but not in the last 72 hours, designate red and alert monitoring personnel for immediate investigation.
iv. If here had been consistent zero inputs in the last 72 hours, designate black and disconnect client.

The following process is shown using the flow algorithm given in Figure 3.1.

The algorithm of Figure 3.1 was also developed into an application named **Electricity Usage Pre-Classifier Interface (ELUPCI)**, a novel application interface for the monitoring and pre-processing of instantaneous electricity usage in real-time. The class for every client is updated as each stream of usage data is obtained from their smart meter and the resultant class is saved on the database against the client's record. This goes on indefinitely, hence the effective classes assigned a user is a function of their cumulative standing within the last 72 hours.

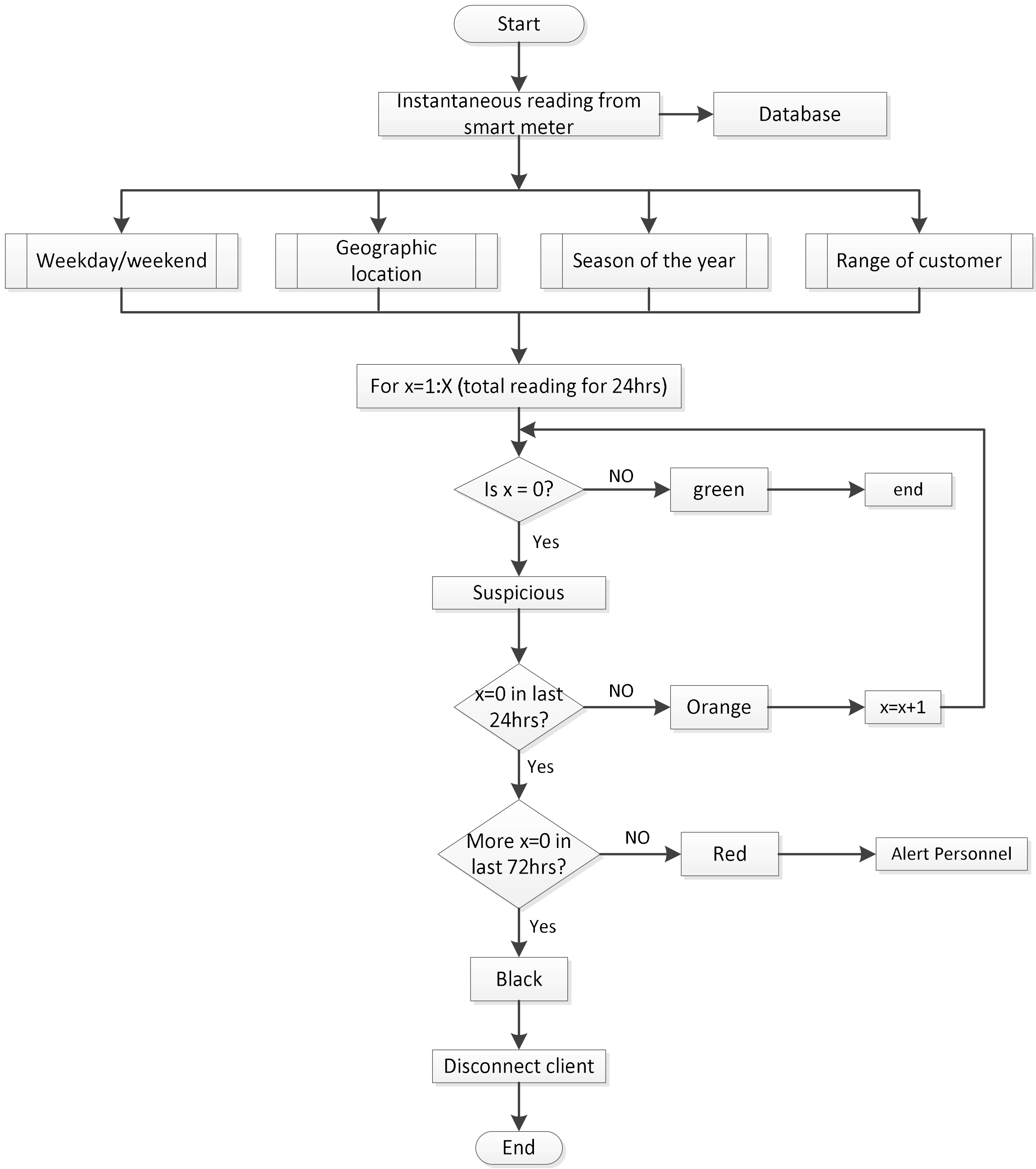



Figure 3.1: Flow Algorithm for Electricity Usage Pre-Classifier Interface (ELUPCI)

### 3.6.2 ELUPCI Design – pseudocode

From the flow algorithm, the following pseudocode was deduced for designing the system

- Initialize ELUPCI
- Accept input from smart meter

- Read/register user ID
- If user does not exist, reject input
- If user exists, send input to user table in Database
- Check: is input x= 0?
- If NO, do nothing, end.
- If YES, check of any other occurrence of zero against user in last 24hrs
- If NONE, label ORANGE, and mark for re-evaluation
- If YES, check for more zeros in last 72 hrs
- If NONE, label RED and alert fraud monitoring team
- If YES, label BLACK, initiate disconnect sequence, alert monitoring team
- End.

### 3.6.3 ELUPCI Design –Deployment Diagram and Use Case Diagram

Deployment diagram is an implementation state diagram showing static deployment view of the system, nodes/components and their relationships. The deployment diagram for ELUPCI is show in Figure 3.2.

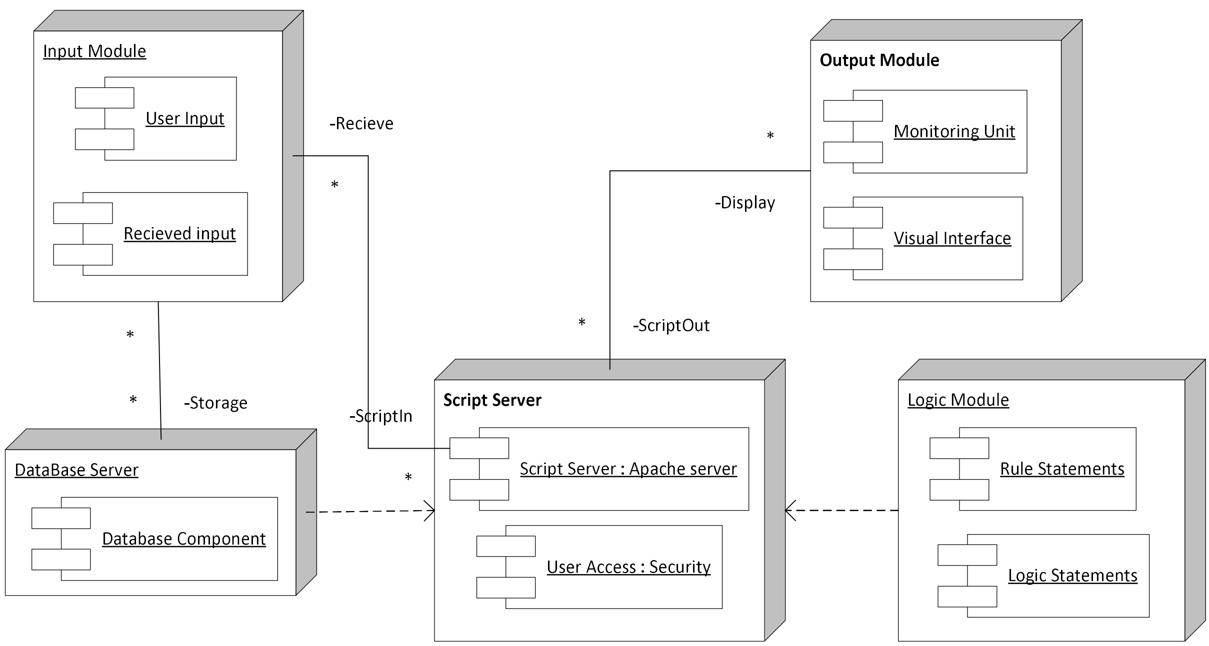



Figure 3.2: Deployment and Component Diagram for ELUPCI

Use case diagram for the implementation of ELUPCI is given in Figure 3.3 showing the interaction of the various components embedded in ELUPCI design.

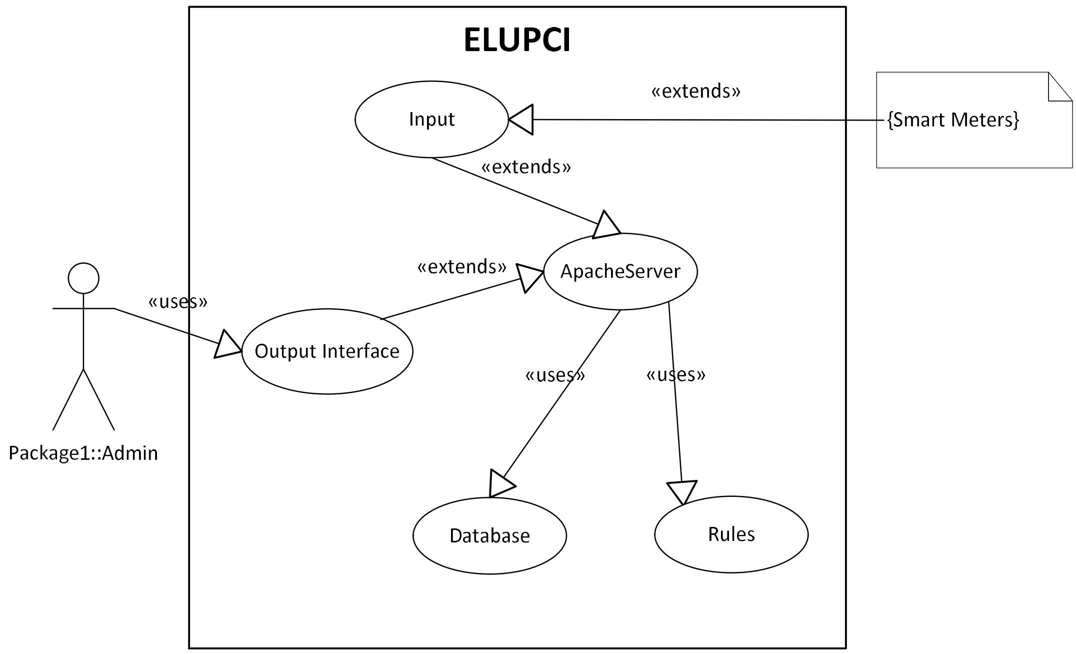



Figure 3.3: Use Case Diagram for ELUPCI

### 3.6.4 ELUPCI Database Design

**Conceptual design:** The conceptual design for ELUPCI database is as shown in Table 3.18.

Table 3.18: Conceptual Design of ELUPCI database

| |
|---|
| Admin (id:int(11), username:varchar(30), password:varchar(60)); |
| Customer (id:int(11), fullname:varchar(30), customerID:varchar(30), reg_date:date, flag:varchar(30)); |
| Mtc000 (id:int(6), dateTime:date, readings:double); |
| CustomerSearch (id:int(11), customerID:varchar(30), fullname:varchar(30), flag:varchar(30)); |

Figure 3.4 shows the Entity Relationship (ER) diagram of ELUPCI database design, showing the entities and their relationship with each other.

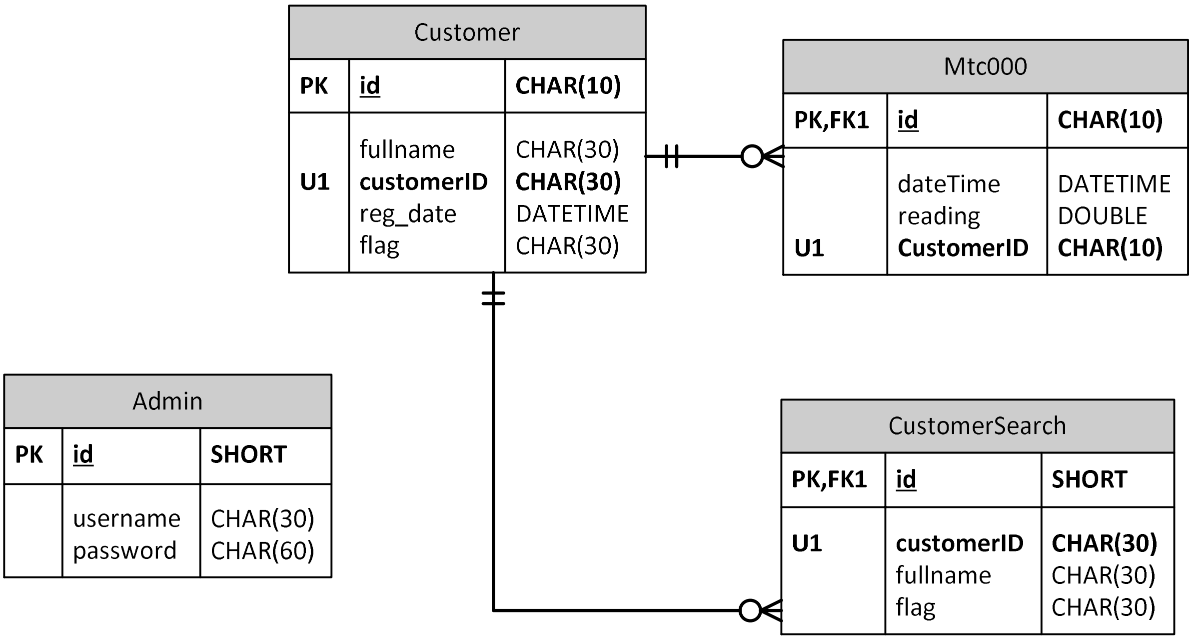



Figure 3.4: Entity Relationship Diagram of ELUPCI Database Design

### 3.6.6 LIBSVM Classification Using ELUPCI Pre-classified Data

Using the same parameters as stipulated in section 3.5 ie:

SVM type = C-SVC,

Cache size =40Mb,

Coefficient = 1,

Cost parameter C for C-SVC= 1.0,

Gamma $\gamma$ = 1.0,

$K^{RBF} = exp(-\gamma * |x_i - x_j|^2)$,

Weight for each class = (1 1 1 1)

Cross validation fold = 10

With standardized data, non-shrinking heuristic, the output of ELUPCI for 125 instances having 11089 attributes was used as input into the LIBSVM classifier, to further enhance the accuracy of the classifier. Table 3.19 shows the output of the enhanced classification.

Table 3.19: Summary of ELUPCI Enhanced LIBSVM Classification of 125 Clients

=== Run information ===

Scheme:weka.classifiers.functions.LibSVM -S 0 -K 2 -D 3 -G 1.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -seed 1

Relation: 4SeasonConsumptionELUPCI125

Instances: 125

Attributes: 11809

[list of attributes omitted]

Test mode:10-fold cross-validation

=== Classifier model (full training set) ===

LibSVM wrapper, original code by Yasser EL-Manzalawy (= WLSVM)

=== Stratified cross-validation ===

=== Summary ===

| | | |
|---|---|---|
| Correctly Classified Instances | 124 | 99.2 % |
| Incorrectly Classified Instances | 1 | 0.8 % |
| Kappa statistic | 0.9719 | |
| Mean absolute error | 0.008 | |
| Root mean squared error | 0.0894 | |
| Relative absolute error | 2.7223 % | |
| Root relative squared error | 23.4685 % | |
| Total Number of Instances | 125 | |

=== Detailed Accuracy By Class ===

| | TP Rate | FP Rate | Precision | Recall | F-Measure | ROC Area | Class |
|---|---|---|---|---|---|---|---|

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 0.045 | 0.99 | 1 | 0.995 | 0.977 | GREEN |
| | 0.955 | 0 | 1 | 0.955 | 0.977 | 0.977 | BLACK |
| Weighted Avg. | 0.992 | 0.037 | 0.992 | 0.992 | 0.992 | 0.977 | |
| === Confusion Matrix === | | | | | | | |
| a b <-- classified as<br>103 0 \| a = GREEN<br>1 21 \| b = BLACK | | | | | | | |

In Table 3.20, a sample (20 iterations) of the predicted class for the 125 instances is shown. The full prediction run for the ELUPCI enhanced LIBSVM prediction is shown in Appendix E – ELUPCI Enhanced LIBSVM Prediction Details

Table 3.20: Sample of ELUPCI Enhanced LIBSVM Prediction

| === Predictions on test data === | | | | | | |
|---|---|---|---|---|---|---|
| inst#, | actual, | predicted, | error, | probability distribution | | |
| 1 | 1:GREEN | 1:GREEN | | *1 | 0 | 0 |
| 2 | 1:GREEN | 1:GREEN | | *1 | 0 | 0 |
| 3 | 2:RED | 2:RED | | 0 | *1 | 0 |
| 4 | 3:BLACK | 3:BLACK | | 0 | 0 | *1 |
| 5 | 1:GREEN | 1:GREEN | | *1 | 0 | 0 |
| 6 | 2:RED | 2:RED | | 0 | *1 | 0 |
| 7 | 3:BLACK | 3:BLACK | | 0 | 0 | *1 |
| 8 | 1:GREEN | 1:GREEN | | *1 | 0 | 0 |
| 9 | 2:RED | 2:RED | | 0 | *1 | 0 |
| 10 | 3:BLACK | 3:BLACK | | 0 | 0 | *1 |
| 11 | 1:GREEN | 1:GREEN | | *1 | 0 | 0 |
| 12 | 1:GREEN | 1:GREEN | | *1 | 0 | 0 |
| 13 | 2:RED | 2:RED | | 0 | *1 | 0 |
| 14 | 3:BLACK | 3:BLACK | | 0 | 0 | *1 |
| 15 | 1:GREEN | 1:GREEN | | *1 | 0 | 0 |
| 16 | 2:RED | 2:RED | | 0 | *1 | 0 |
| 17 | 3:BLACK | 3:BLACK | | 0 | 0 | *1 |
| 18 | 1:GREEN | 1:GREEN | | *1 | 0 | 0 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 19 | 2:RED | 2:RED | | 0 | *1 | 0 |
| 20 | 3:BLACK | 3:BLACK | | 0 | 0 | *1 |

# CHAPTER FOUR

# RESULTS AND DISCUSSION

## 4.1 Results of Data Analysis of the 370 consumers

The results of the data analysis on the 370 consumers (represented by the 370 instances of the dataset) are shown in this section.

Figure 4.1 shows the distribution of the customer types represented in the 370 instances of the dataset. From the figure, it could be observed that large residential customer type have the highest frequency (156 instances), followed by large commercial (101 instances). Small Commercial/Large Agricultural consumers have the lowest frequency (27 instances). This implies that the dataset was the electricity usage from a fairly industrialized urban residential area.

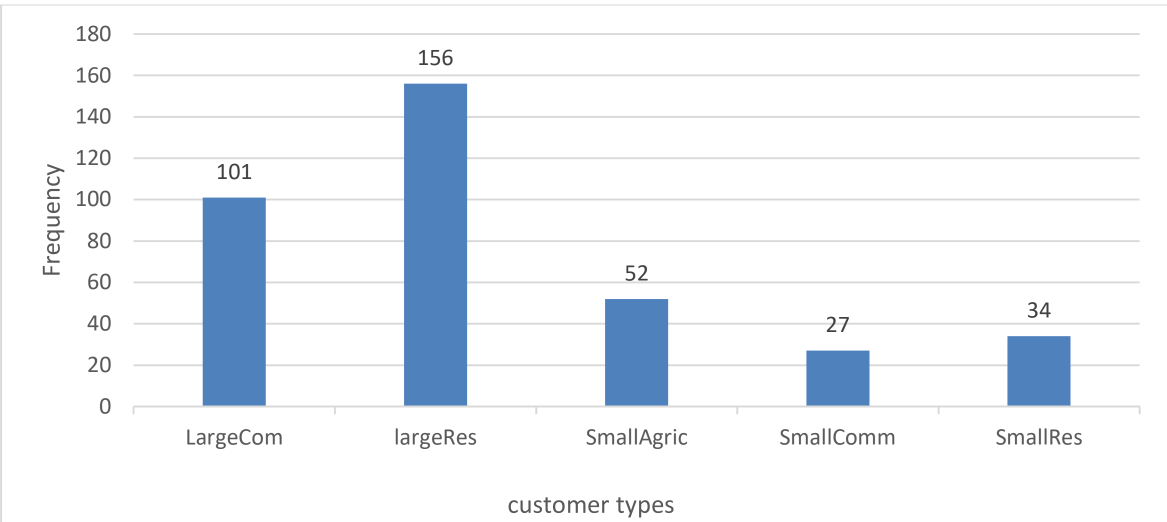



Figure 4.1: Distribution of Customer Types

Figure 4.2 shows the column chart of average energy consumption (in kilowatts/hour) of the different customer types. It could be observed that large commercial consumers utilizes the most energy (daily average of 361.82kW/h), followed by small commercial consumers (59.11kW/h). Small Agricultural consumers utilizes the least energy with an average daily consumption of 2.14kW/h

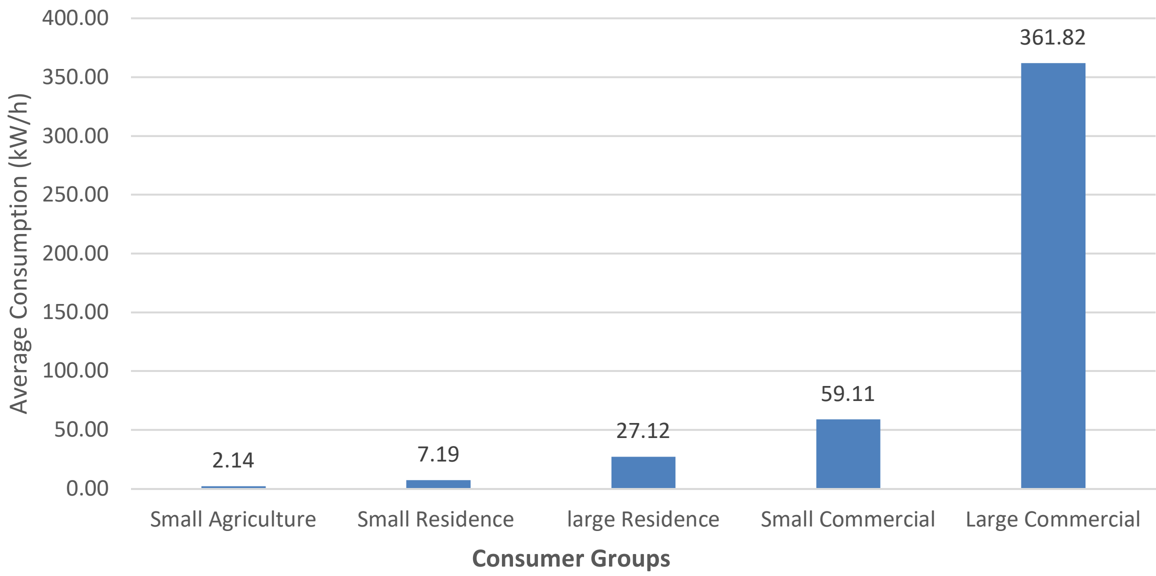



Figure 4.2: Average Energy Consumption (KWh) of Represented Consumer Types

Figure 4.3 - 4.7 shows the daily maximum and minimum energy consumption of randomly selected clients, each group according to their categories.

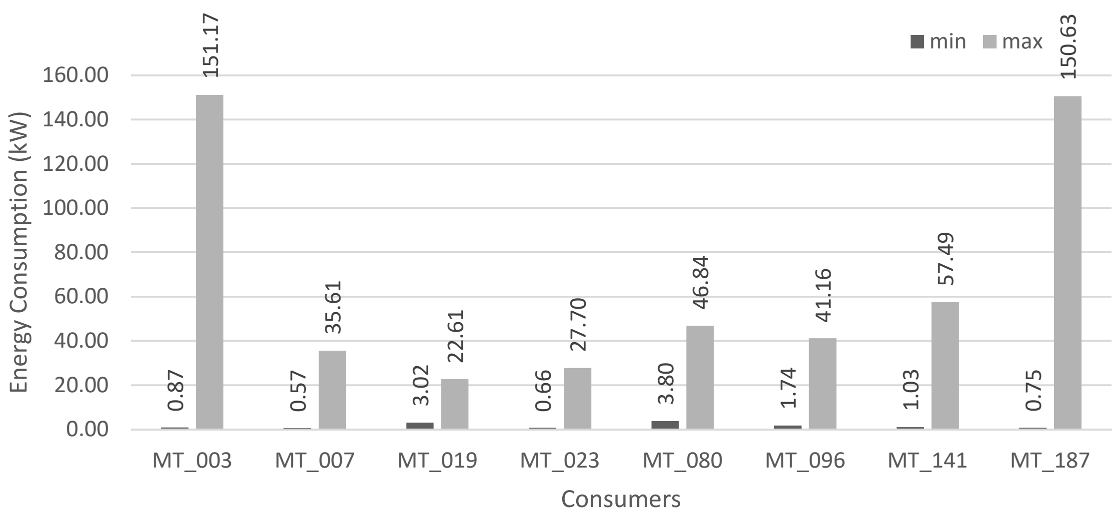



Figure 4.3: Small Agricultural Consumers Showing Maximum and Minimum Consumption in kW

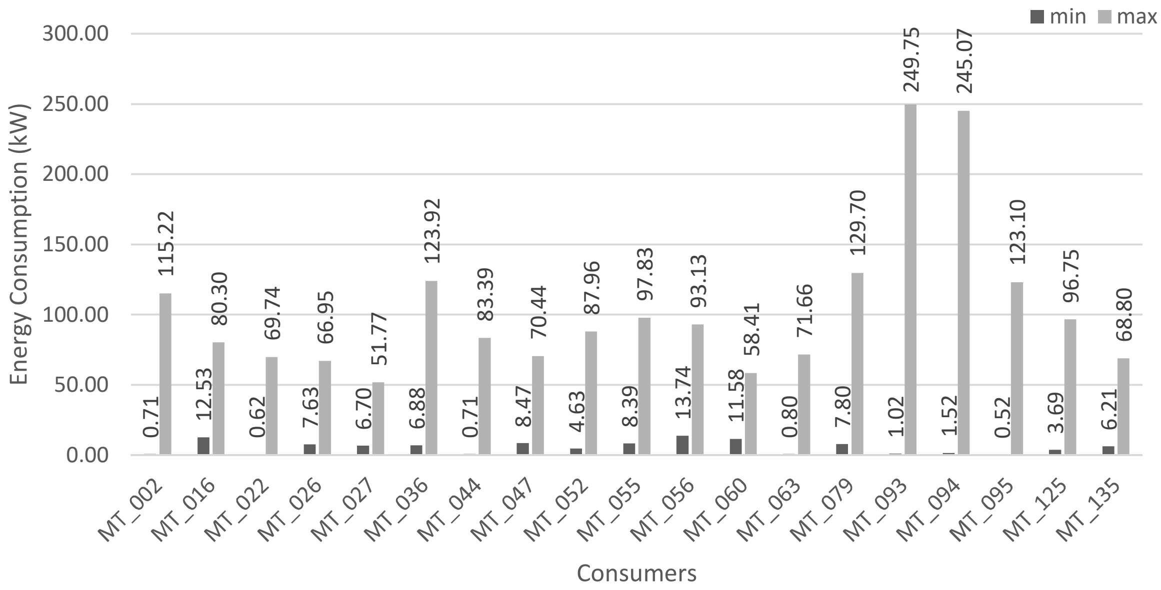



Figure 4.4: Small Residential (and Large agricultural) Consumers Showing Maximum and Minimum Consumption in kW

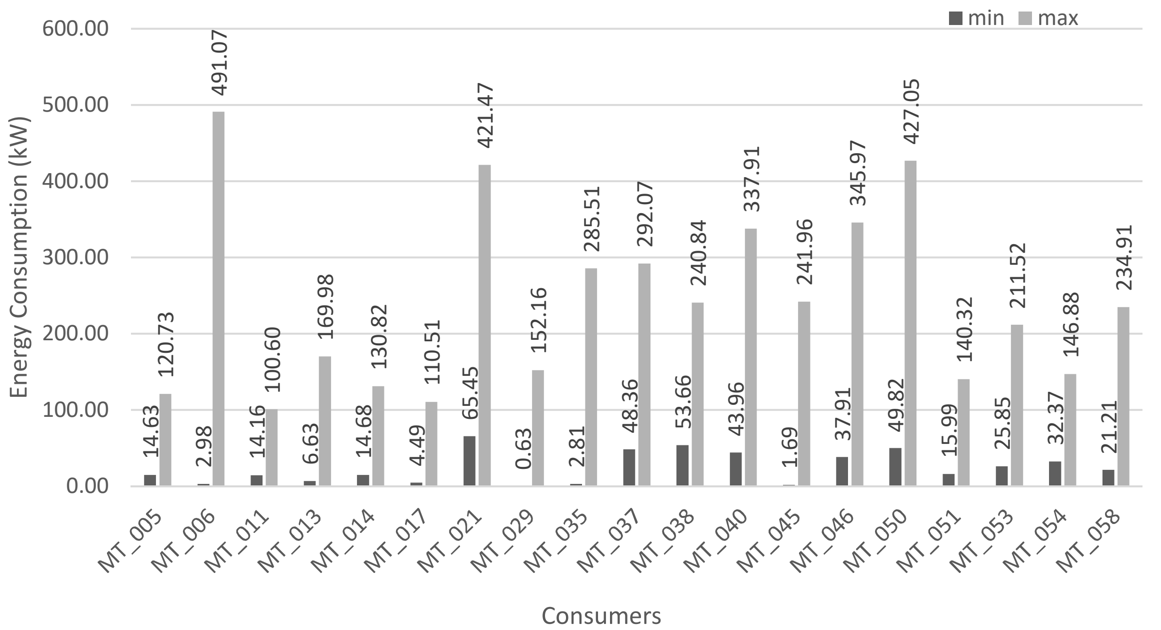



Figure 4.5: Large Residential Consumers Showing Maximum and Minimum Consumption in kW

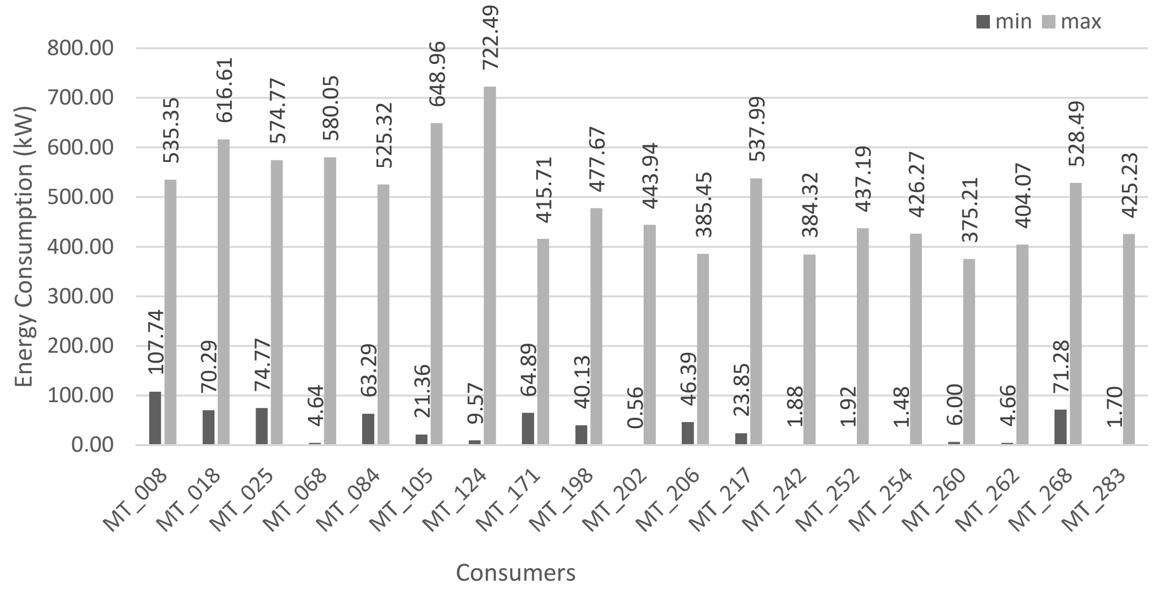



Figure 4.6: Small Commercial Consumers Showing Maximum and Minimum Consumption in kW

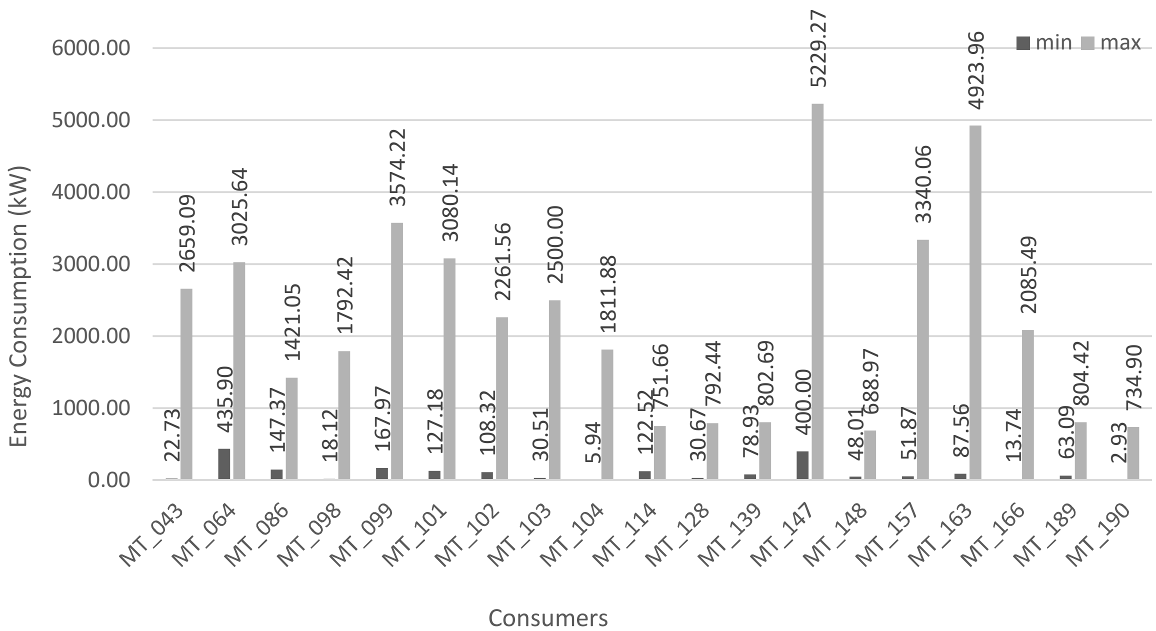



Figure 4.7: Large Commercial Consumers Showing Maximum and Minimum Consumption in kW

Figure 4.8 – Figure 4.12 presents the consumption profile of randomly selected clients from each of the customer types plotted against the time of day through the four seasons of the year (ie Spring (April), Summer (July), Autumn (October)

and Winter (January), both on a weekday (Monday) and a weekend (Sunday), in order to determine how these factors affect electricity consumption.

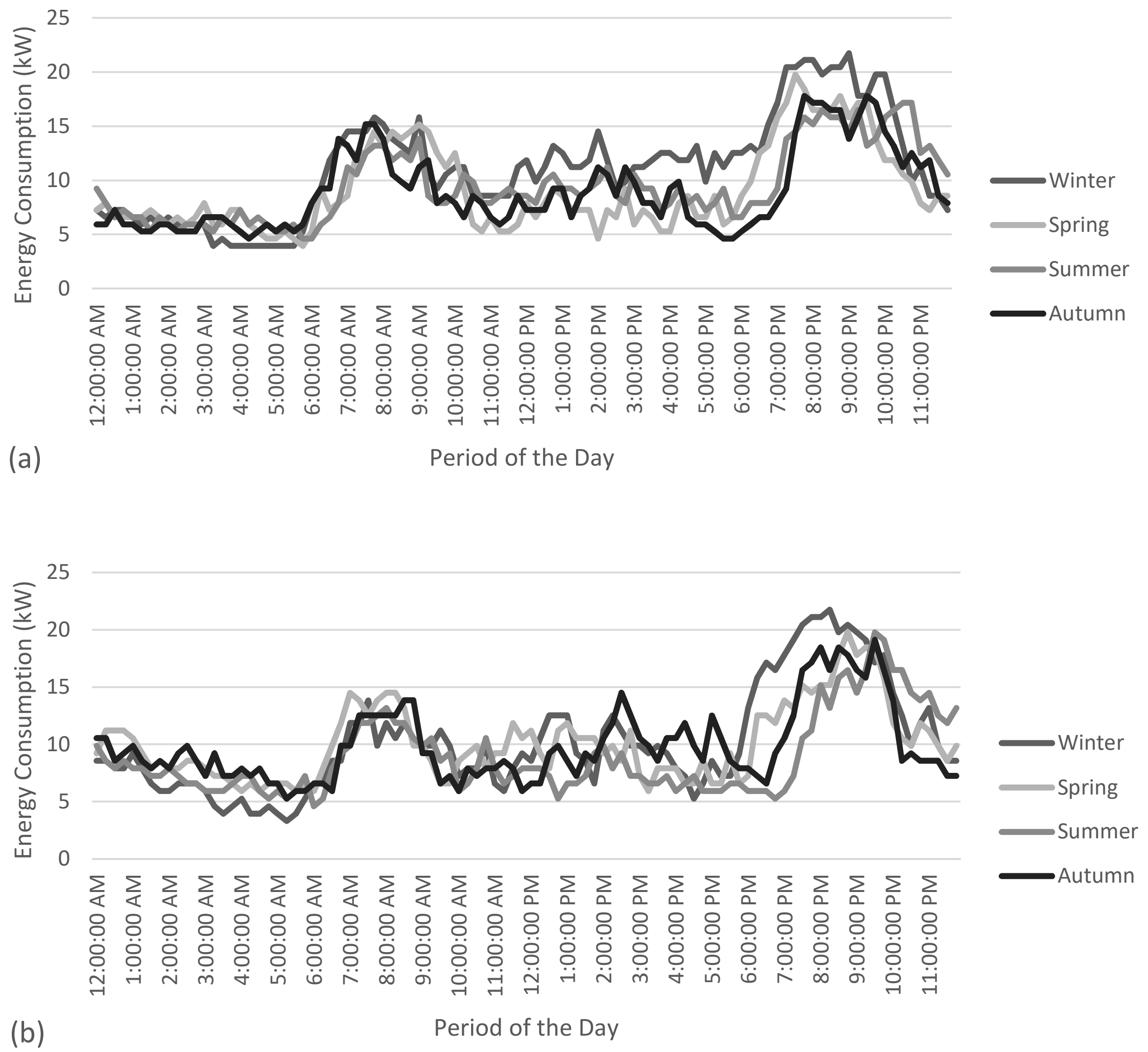
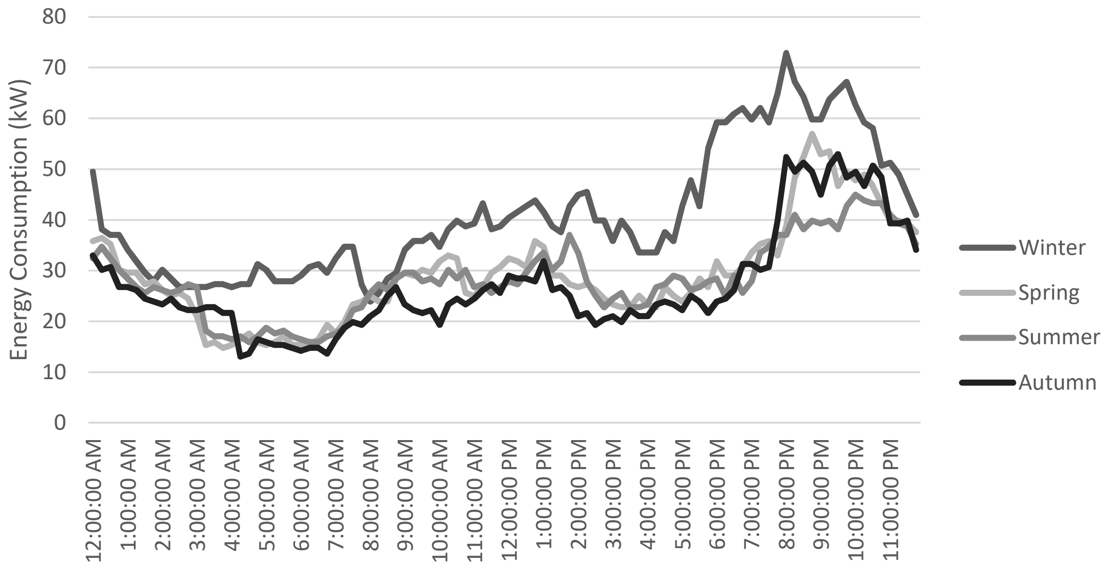
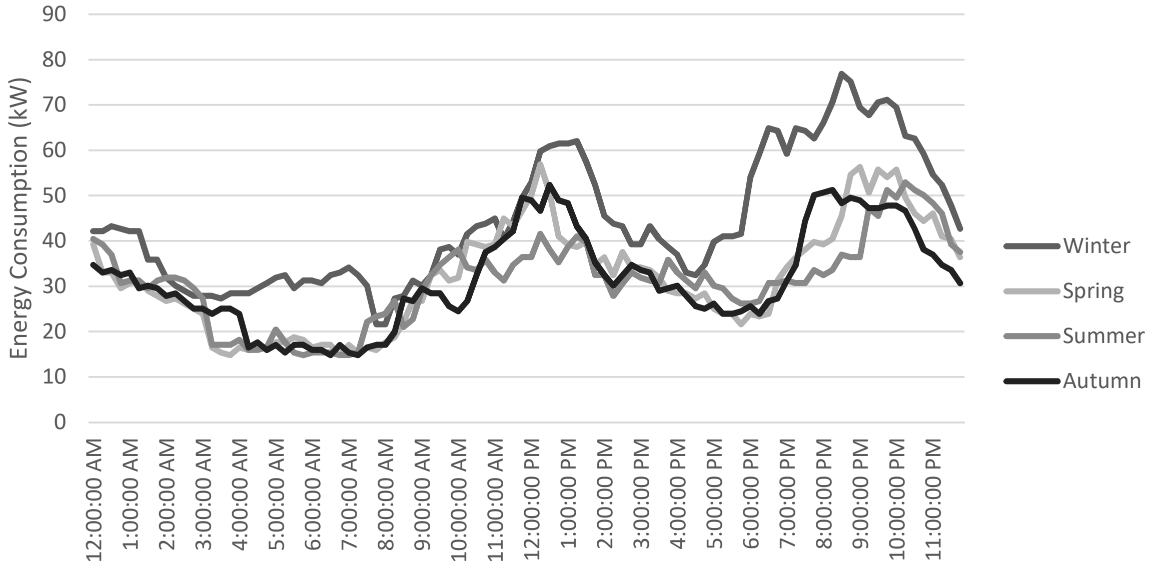



Figure 4.8: Typical Small Agricultural Consumer's Daily Consumption across Four Seasons on (a) Weekday. (b) Weekend

The plot in Figure 4.8 shows a significant increase in energy consumption from 6am through 9am and then again from 7pm through 10pm. Also it could be observed that consumption on weekdays (Figure 4.8a) during winter is slightly higher than at other seasons.



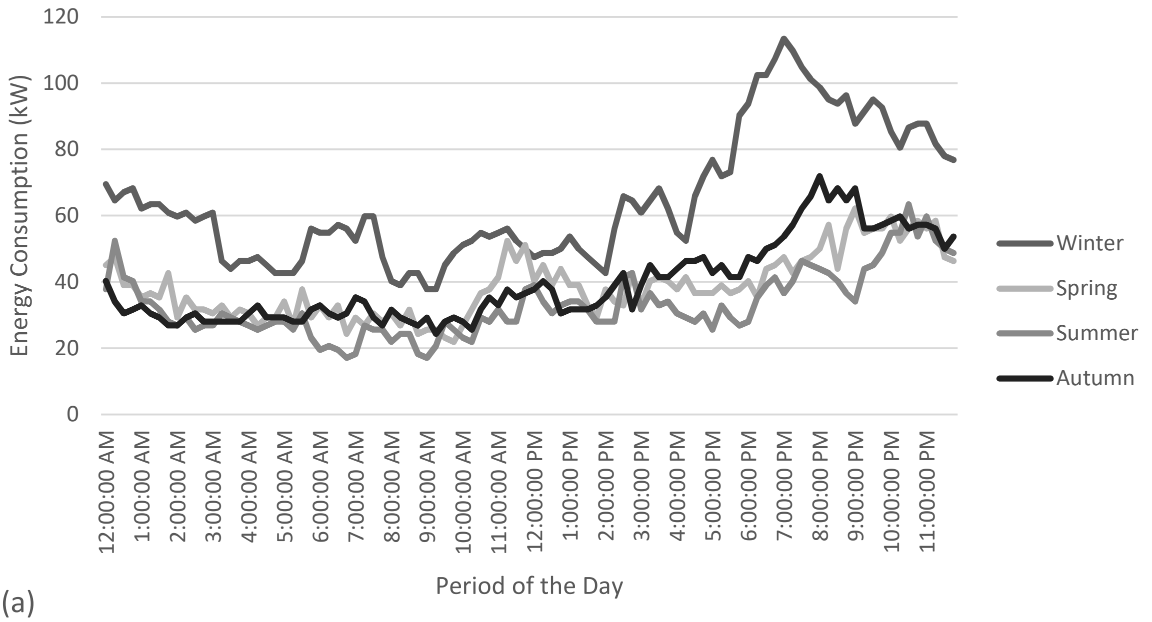

(a)



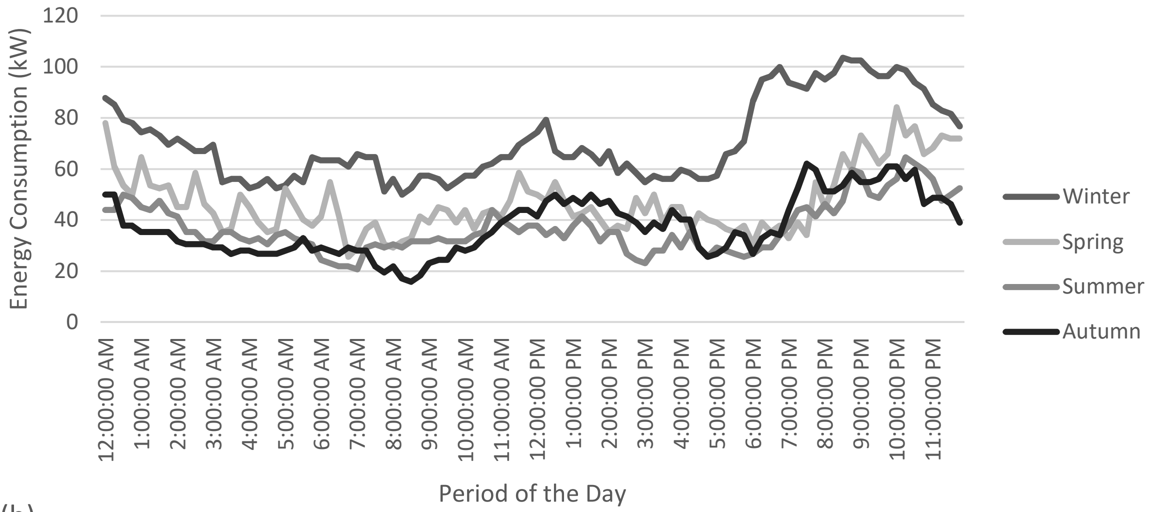

(b)

Figure 4.9: Typical Small Residential Consumer's Daily Consumption across Four Seasons on (a) Weekday. (b) Weekend

The plot for small residential consumers (Figure 4.9) shows marked increase in consumption during winter than at other seasons. This is more so as from 6pm through 10pm. Also a sharp spike in consumption could be observed as from 10am through 12pm on weekends (Figure 4.9b).



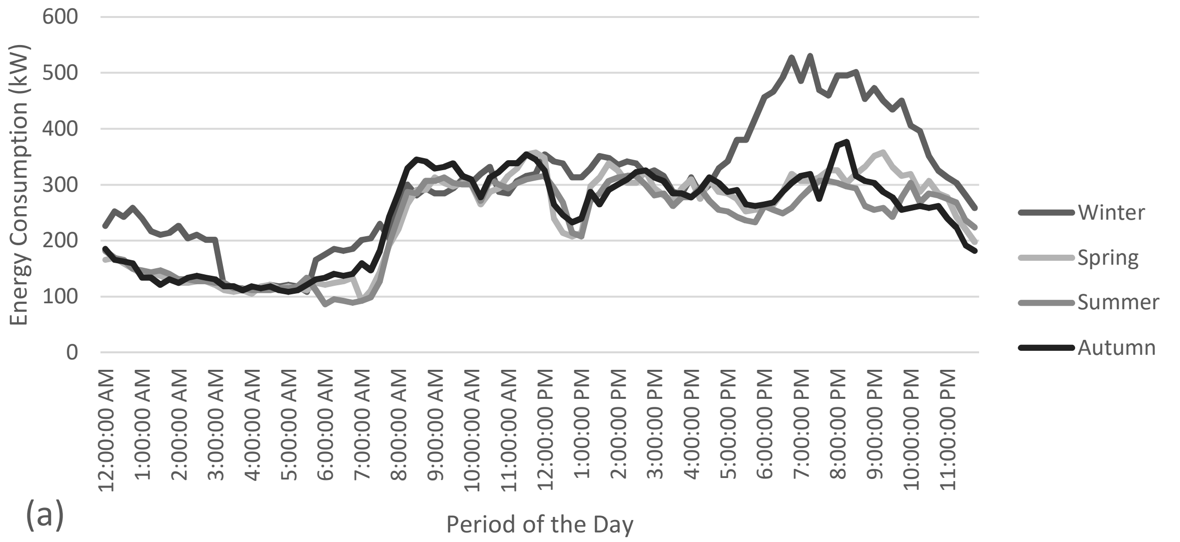

(a)



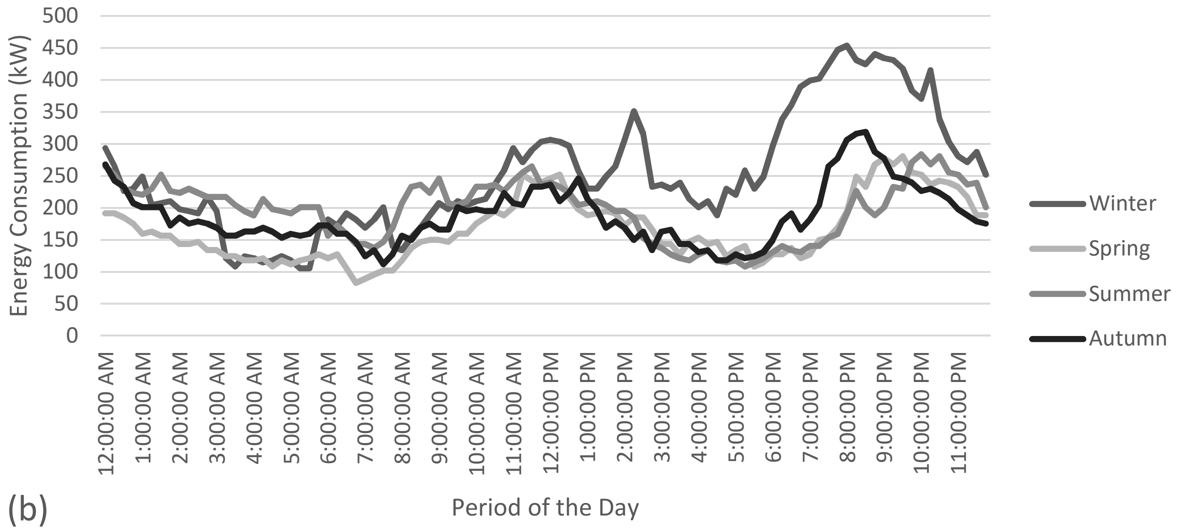

(b)

Figure 4.10: Typical Large Residential Consumer's Daily Consumption across Four Seasons on (a) Weekday. (b) Weekend

Just as with small residential consumers, large residential consumers also show heightened consumption during the winter season (Figure 4.10), the maximum point occurring at about 7pm on weekday (Figure 4.10a) and between 7pm and 10pm during weekends (Figure 4.10b).



(a)



(b)

Figure 4.11: Typical Small Commercial Consumer's Daily Consumption across Four Seasons on (a) Weekday. (b) Weekend

The plot for small commercial consumers (Figure 4.11) shows some distinction when compared with the residential consumers. Here there is a sharp upsurge in consumption as from 7am which is sustained till around 10pm, especially during weekdays (Figure 4.11a); an exception occurring during the winter period where the consumption also increases again significantly during the late evening.

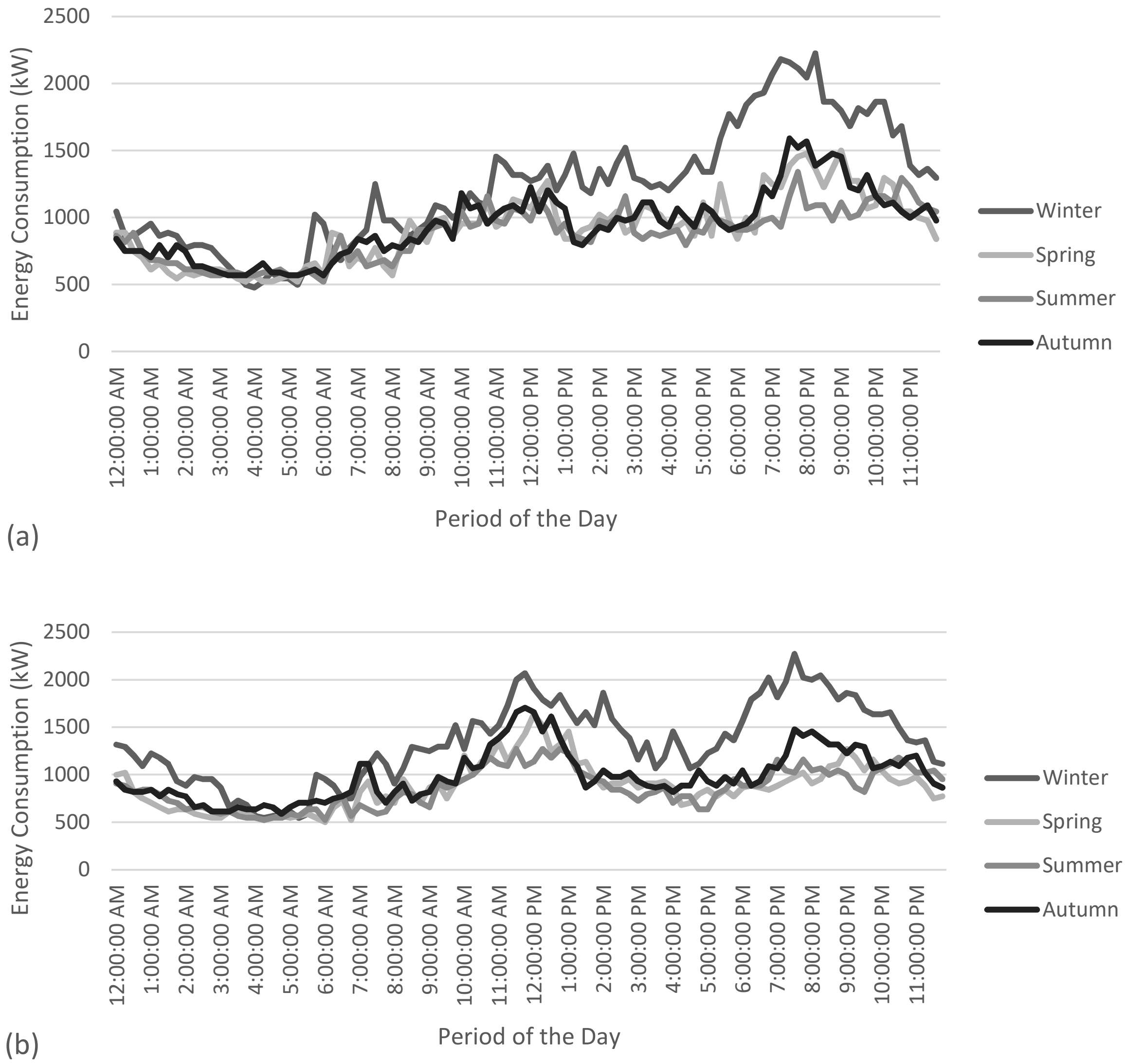



Figure 4.12: Typical Large Commercial Consumer's Daily Consumption across Four Seasons on (a) Weekday. (b) Weekend

The plot for the large commercial consumers (Figure 4.12) shows also a sustained increase from around 7am till about 9pm, except of the weekend which saw a dip around the afternoon period (Figure 4.12b).

## 4.2 Classifier and Parameter Selection Accuracy Tests

To be able to select the best classifier for the dataset used for this work, a Cross validation paired corrected test was conducted on nine popular classifiers. The result of the test was given in Table 3.10 and Table 3.11, while a column chart of the outcome is shown in Figure 4.13. From the figure it could be seen that Random Forest had the highest CV test accuracy followed closely by LibSVM and IBk. Due to the constraints experienced with Random Forest as explained in section 3.2.1, and given the fact that other researchers working on this field had posited that LIBSVM is one of the best classifiers for classification of electricity energy data, LIBSVM was chosen as the algorithm for this work.

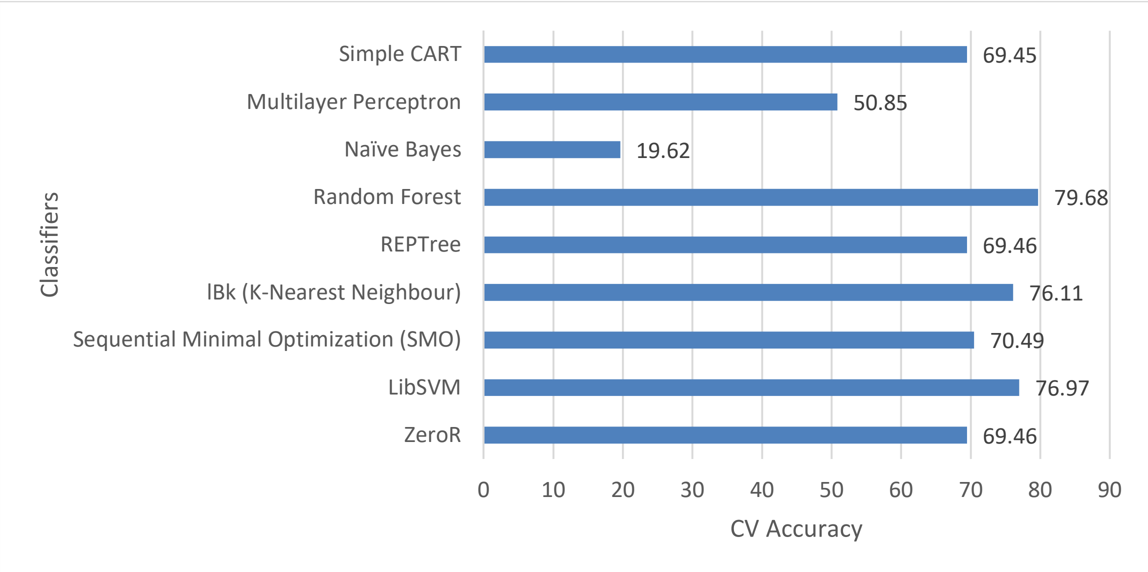



Figure 4.13: Column Chart for CV Accuracy Test on Nine Classifiers

In order to determine the best SVM kernel for our dataset (out of the four SVM kernels) a paired corrected tester was conducted (see Table 3.12) and a column chart also plotted from the results of that test. From the column chart (as shown in Figure 4.14), it could be seen that RBF and Sigmoid had a cross validation accuracy of 79.46% while Linear and Polynomial kernels had a CV accuracy of 72.86% and 71.41% respectively. Based on the obtained results and the fact (from studied literatures – Section 3.2.6) that RBF kernel is the widely applied kernel function based on Euclidean distance, RBF kernel was selected for this research work.

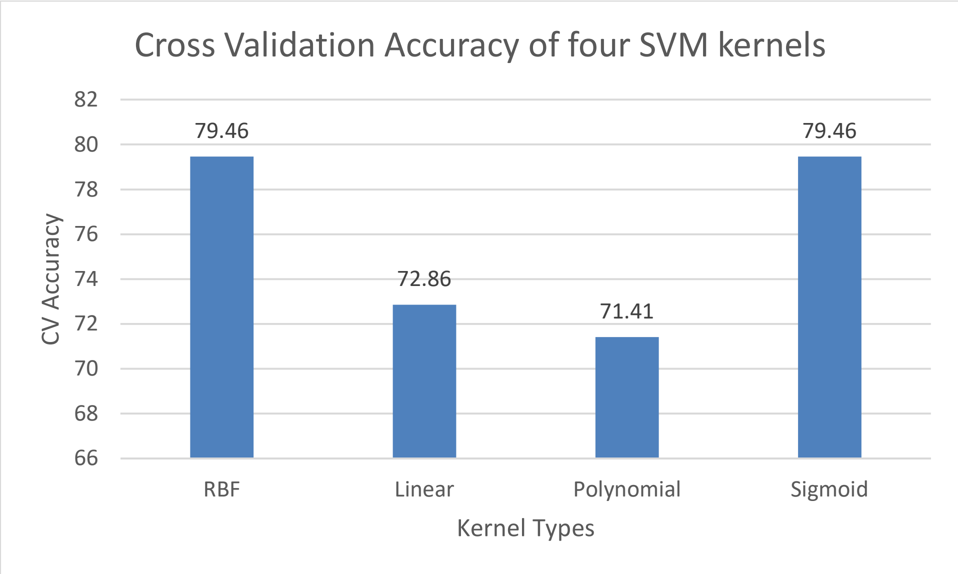



Figure 4.14: Column Chart Showing Cross Validation Accuracy Testing for Four SVM Kernel Types

Since the main parameters which determine the outcome of the SVM classifier are the $(C, \gamma)$ pair, a cross validation accuracy test was done on a range of $C, \gamma$ pair in order to determine which pair gives the best accuracy. From the result of the test (shown in Table 3.13) a stem plot (Figure 4.15) was drawn showing the different $C, \gamma$ combinations compared and the CV accuracy results obtained. From Figure 4.15, it could be seen that values of $(C, \gamma)$ corresponding to (1.0, ≥ 1.0) an accuracy of 50 was obtained. Hence a $(C, \gamma)$ value of (1.0, ≥ 1.0) was used in the SVM classifier for this work.

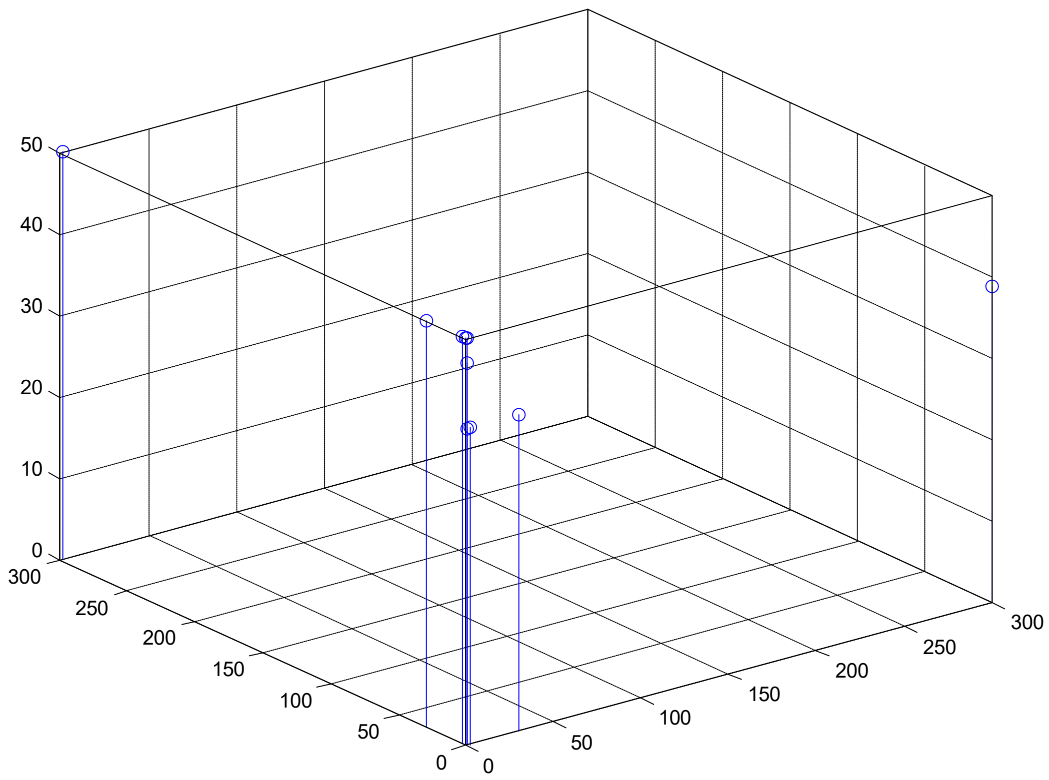



Figure 4.15: Stem plot Cross Validation Accuracy Tester to Determine Highest of $C, \gamma$ Values

### 4.3 ELUPCI Implementation

The novel application, Electricity Usage Pre-Classifier Interface (ELUPCI) was developed based on the developed algorithm (Figure 3.1) and design specifications highlighted in section 3.6. It automatically receives instantaneous energy readings from customer's smart meter, pre-classifies them and appends the result of the classification to the database against the client. This action is a real-time event and updates automatically to reflect current status of each user per time. The application was coded in Hypertext Pre-processor, an application development environment for developing web based application. Screen shots of ELUPCI are shown in Figure 4.16 - Figure 4.20 while the codes are given in Appendix D – Electricity Usage Pre-Classifier Interface (ELUPCI).

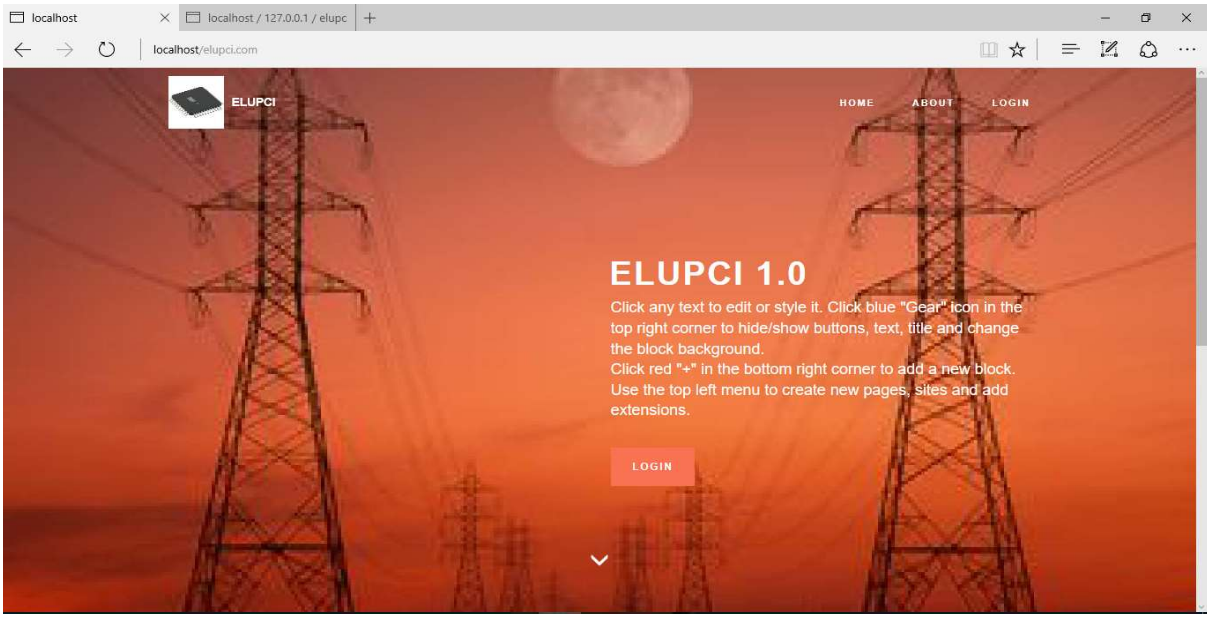



Figure 4.16: ELUPCI – Home Screen

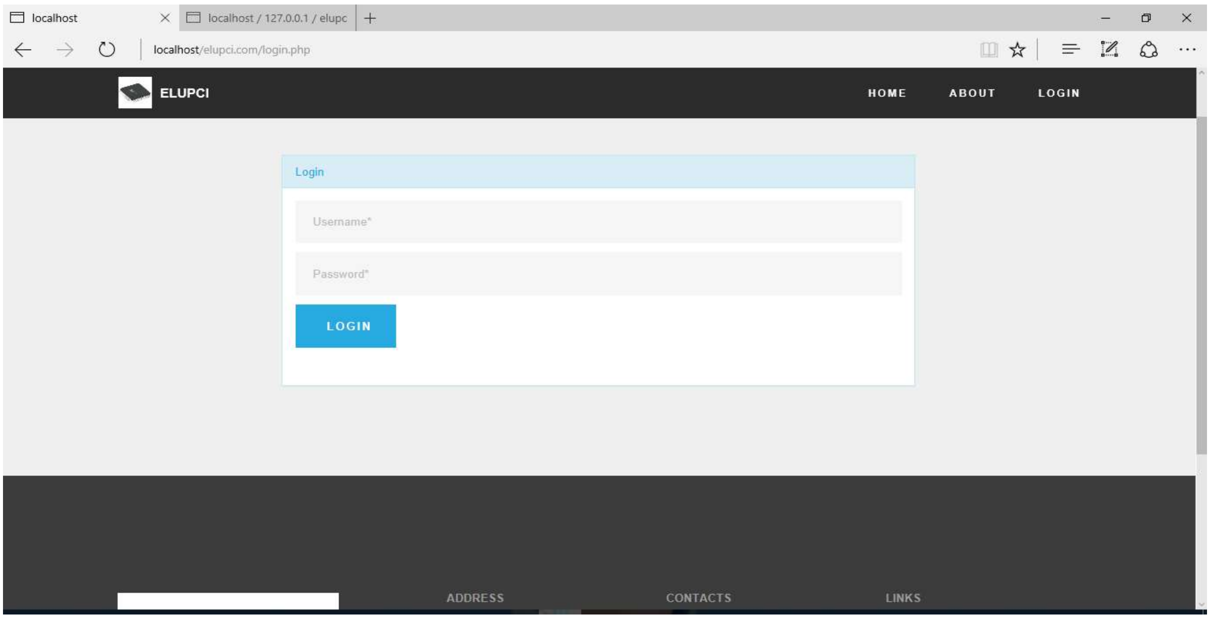



Figure 4.17: ELUPCI – Login Page

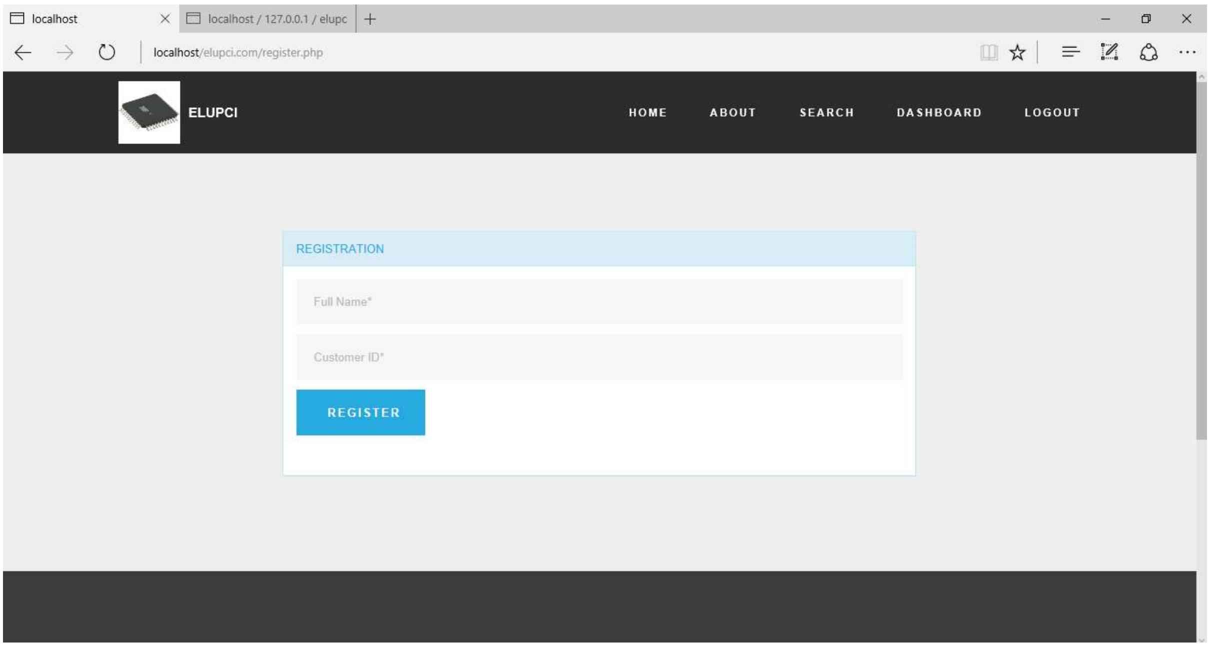



Figure 4.18: ELUPCI – Customer Registration Page

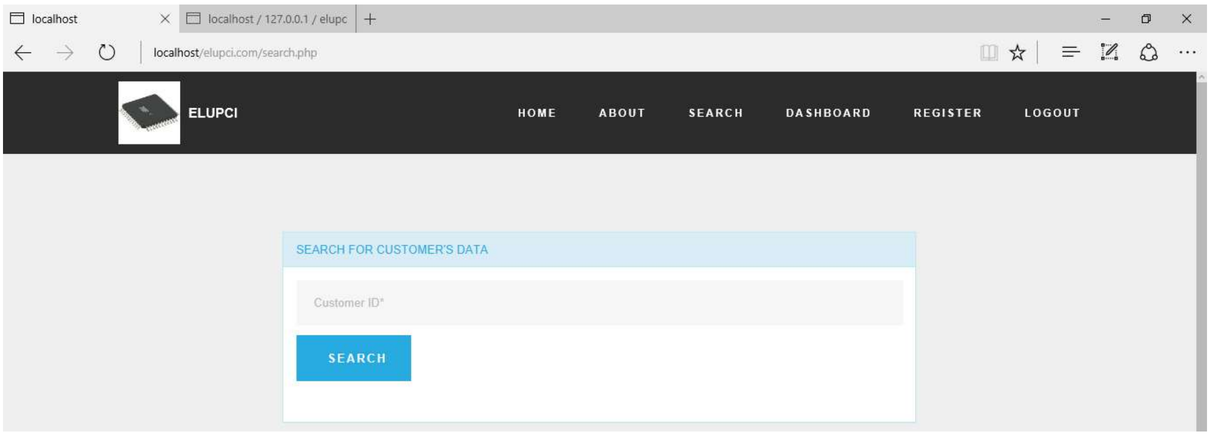



Figure 4.19: ELUPCI – Customer Data Search

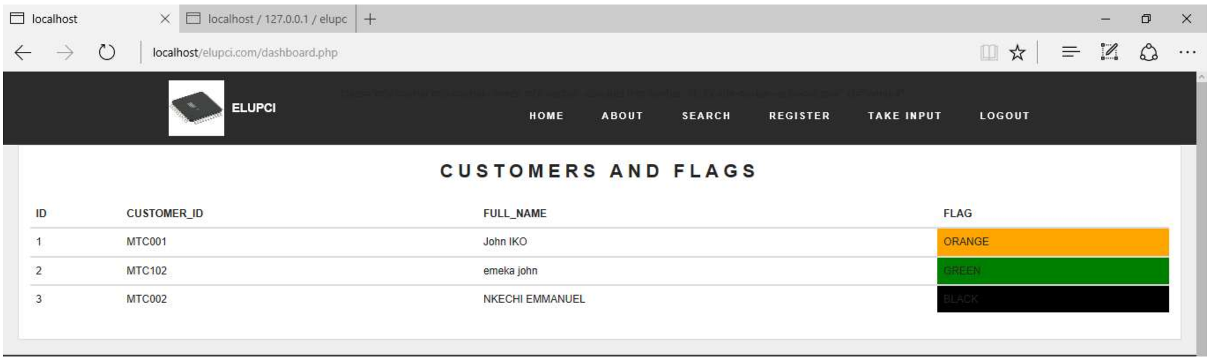



Figure 4.20: ELUPCI – Monitoring and Management Interface Showing Customers and Flags

## 4.4 Result of LIBSVM Classification

From Table 3.17, the LibSVM correctly classified 294 out of the 370 considering a total of 11808 attributes which corresponds to 79.46% accuracy. It could also be observed (from the detailed accuracy by class section) that the Receiver Operating Characteristics (ROC) area (also called Area Under Curve (AUC), which is the a plot of True Positive Rate (TP) against False Positive Rate (FP)) for RED class was 0.5% each while that of GREEN and BLACK classes were 0.673% and 0.997% respectively. AUC represents a probability that a positive will be ranked higher than a negative. The confusion matrix showed that the total 257 and 37 instances, belonging to the GREEN and BLACK class respectively were all correctly classified. Figure 4.21– Figure 4.23 shows a plot of the area under ROC for the four classes.

The f-measure (or f-score) is calculated using the formula

$$f - measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4.1)$$

Where $Precision = \frac{TP}{TP+F}$ and $Recall = \frac{TP}{TP+F}$

TP = True Positives, FP = False Positives, FN = False Negatives

Precision is defined as the fraction of elements correctly classified as positive out of all the elements the algorithm classified as positive, whereas recall is the fraction of elements correctly classified as positive out of all the positive elements.


Figure 4.21: Area Under ROC for Class Value RED

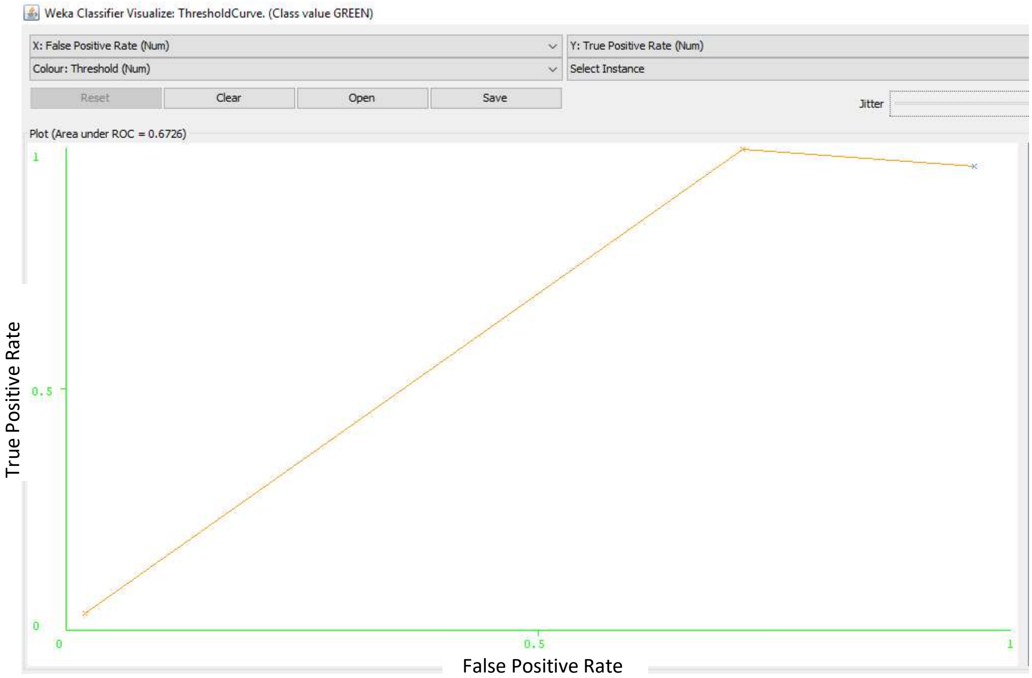



Figure 4.22: Area Under ROC for Class Value GREEN

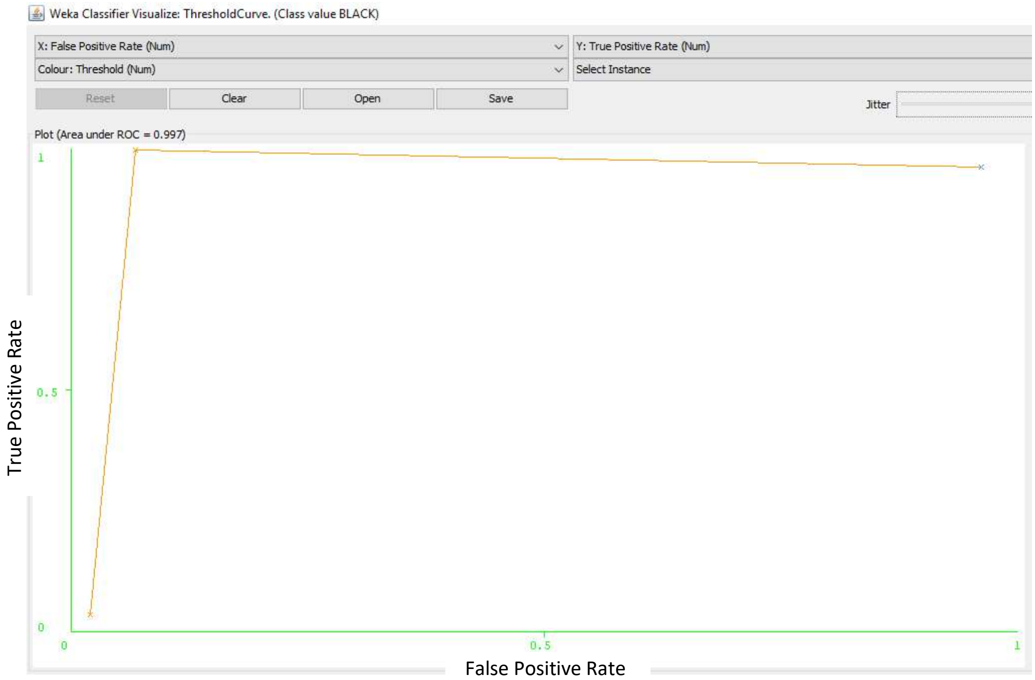



Figure 4.23: Area Under ROC for Class Value BLACK

In the multiclass case, each class $i$ have a respective precision and recall, in which a "true positive" is an element predicted to be in $i$ is really in it and a "true negative" is an element predicted to not be in $i$ that isn't in it.

The weighted f-measure is a weighted average of the classes' f-measure, weighted by the proportion of how many elements are in each class.

Figure 4.24 shows a scatter plot of the classifier error, showing the 4 classes and the correctness of their classification.

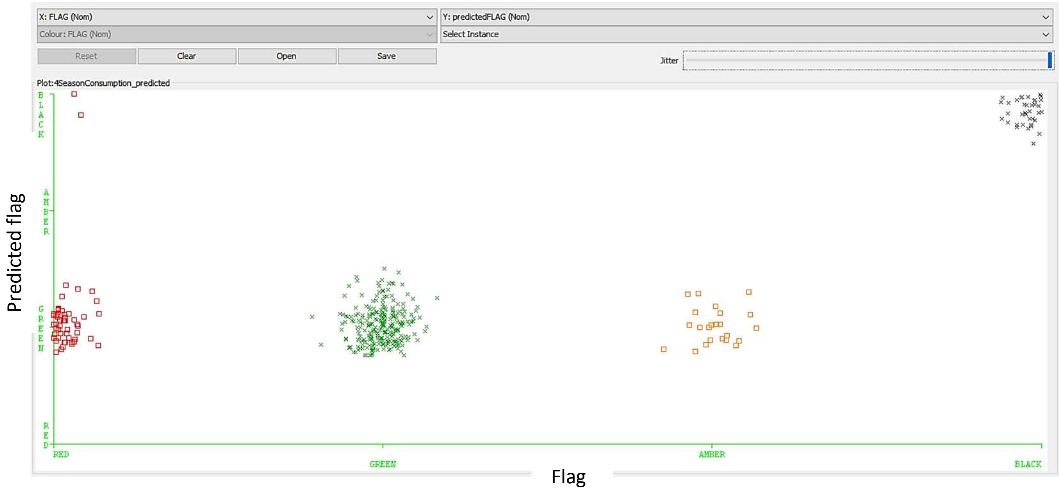



Figure 4.24: Classifier Error Showing Correctly and Incorrectly Classified Classes

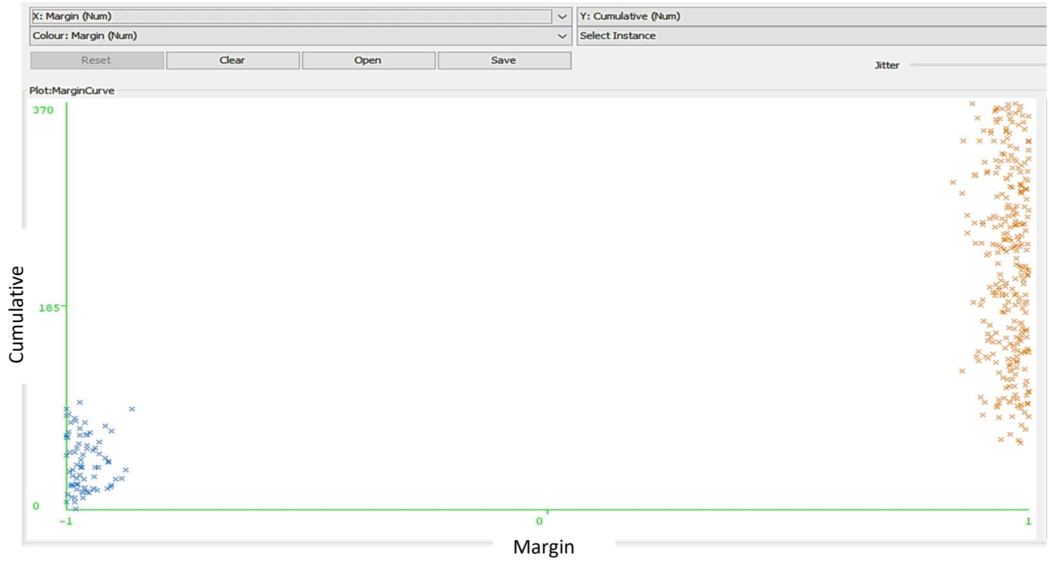



Figure 4.25: Margin Curve for 370 Classified Instances

The margin is defined as the difference between the probability predicted for the actual class and the highest probability predicted for the other classes. Figure 4.25 displays the margin curve of the 370 classified instances showing deviation from prediction of class by the instances.

Figure 4.26 shows a cross section of the output scatter plot of the classification process that is, the distribution of users over the time space.

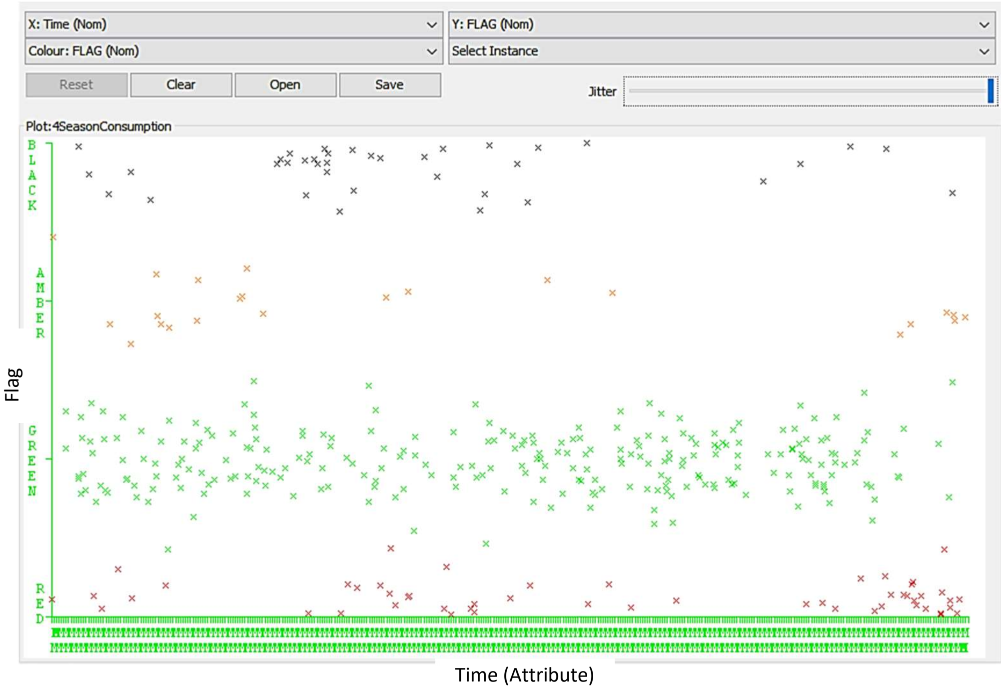



Figure 4.26: Scatter Plot of User Attribute (TIME) Against the Class Attribute (FLAG) Showing Distribution of Users

## 4.5 Results of ELUPCI-Enhanced LIBSVM Classification

Table 3.19, shows the output of classification of the client data that had already been pre-classified using ELUPCI (Electricity User Pre-Classifier Interface). From the table it could clearly be seen that LIBSVM was able to achieve a 99.2% accuracy with the 125 instances classified. However, since the initial classification was done with 370 clients, extrapolating the result from this 125 clients will give an approximate of 98.65% accuracy. The confusion matrix showed that it was able to classify all the three classes correctly into their respective class.

This is a marked improvement (19.19%) from the 79.46% gotten from the initial attempts at classifying without first pre-classifying the instantaneous readings with ELUPCI.

It could also be observed (from the detailed accuracy by class section) that the ROC area for all the classes are 0.977, meaning that most of the instances were classified correctly.

The scatter plot of the classifier error which shows that all 125 instances were correctly classified is shown in Figure 4.27.

The treshold curve for the three classes are shown in Figure 4.28 - Figure 4.29. The curve in each diagram shows that the classes – GREEN and BLACK were almost all correctly classified (ROC =0.977) which also corresponds to the result shown in Table 3.19 under “detailed accuracy by class” and “confusion matrix”.

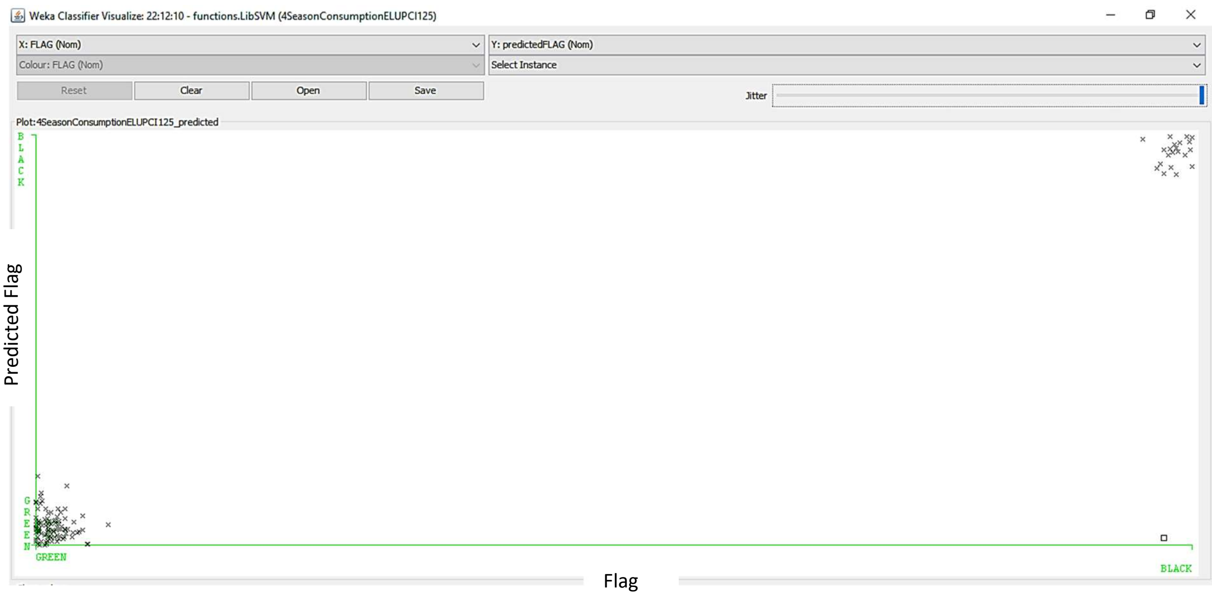



Figure 4.27: Classifier Error Showing Correctly Classified Classes

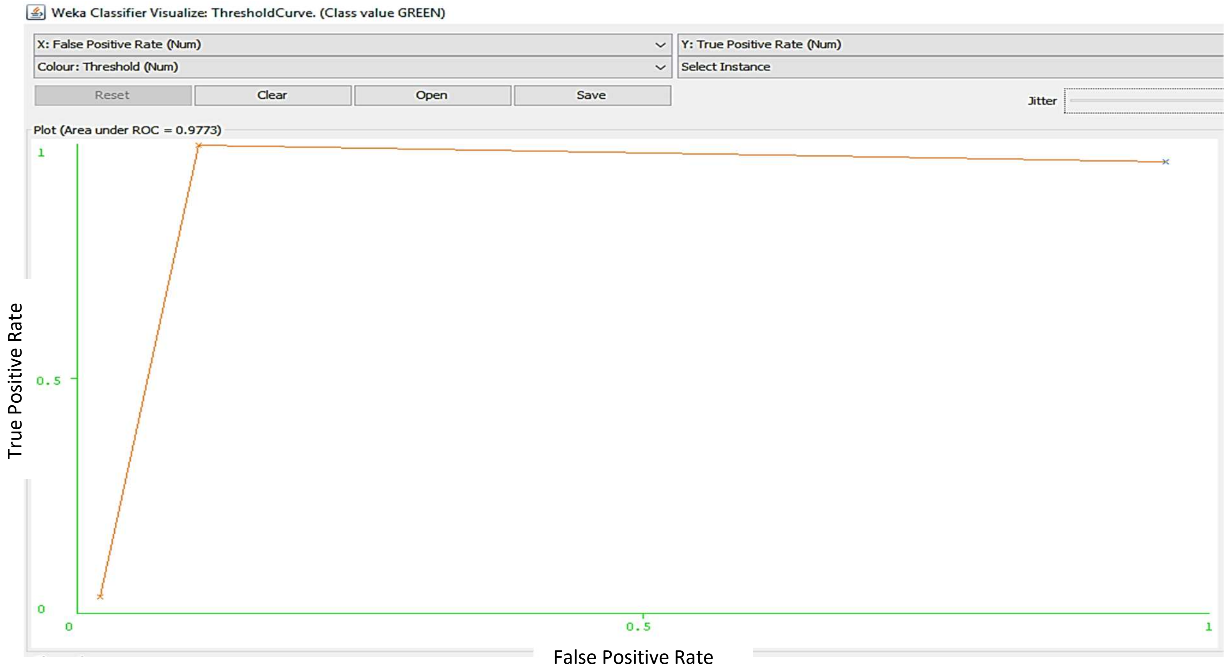



Figure 4.28: Area Under ROC for Class Value GREEN

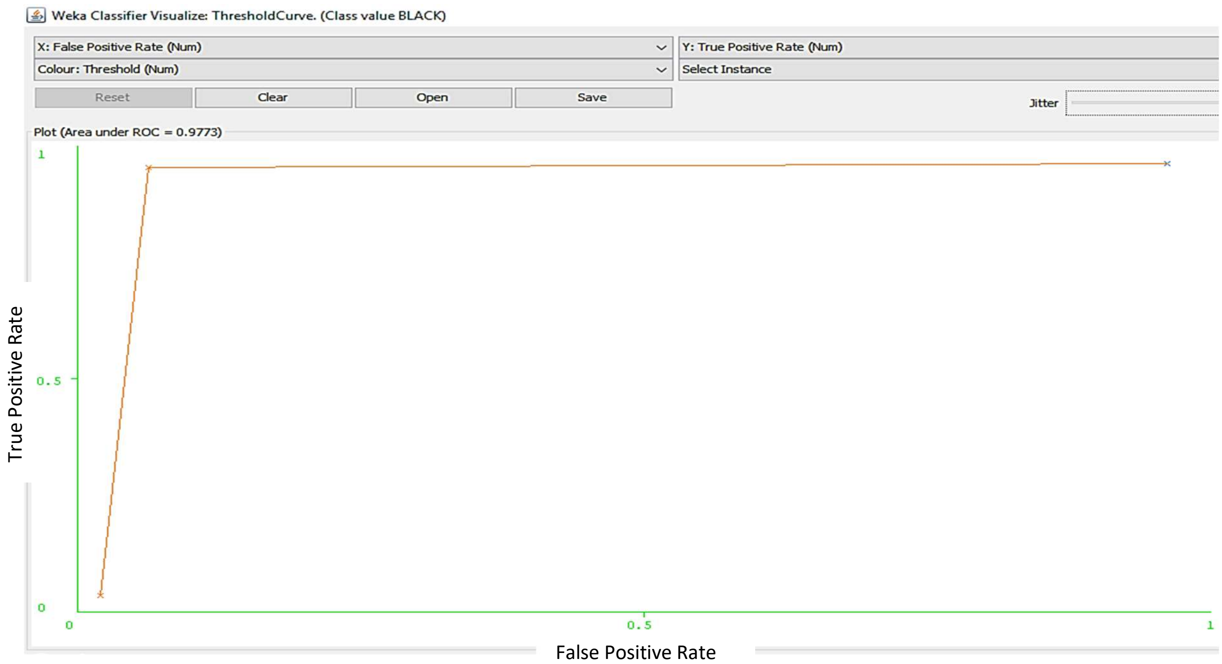



Figure 4.29: Area Under ROC for Class Value BLACK

The major implication of this result is that utility workers can predict the occurrence of NTL in electricity usage with about 98.65% accuracy using the developed model. This will enable them take intelligent decisions and promptly disconnect any fraudulent user remotely, using the remote disconnection facilities embedded in AMI and smart meters and hence save them the trouble (and its attendant cost implications) of making frequent home calls to determine when and where there is a tampering of the user’s energy metering facility or such fraudulent activities.

# CHAPTER FIVE

# CONCLUSION AND RECOMMENDATION

## 5.1 Conclusion

Monitoring and management of energy usage has become a trending issue among researchers and utilities chiefly due to the progressive search for the solution to the nagging issue of non-technical losses which continue to waste tons of funds for the utilities and the government.

Firstly research has shown that there is a close correlation between electricity usage data collected from consumers' energy meter (smart meters especially) and the energy usage habit or pattern of the consumers. This relationship was exploited in this work in a bid to discover a solution to NTL.

In this work, preliminary analysis of the consumers' energy usage data revealed that there is normally a heightened use of electric energy during the winter period (especially among residential consumers) in contrast to other seasons of the year. This increased use of energy was also observed between 7pm and 10pm daily with the residential consumers whereas the commercial consumers have their peak usage periods from around 10am through 6pm. This information could guide electricity generating and utility companies on proper distribution of electricity load during the days and seasons of the year.

In terms of classification of energy users into fraudulent or valid users, first this research showed that Support Vector Machines is one of the best machine learning classifiers for use in classifying dataset of the nature of electricity energy usage data. Also this research revealed that Radial Basis Function (RBF) kernel (or Gaussian kernel) is one of the best SVM kernels suitable for classifying this nature of data. Also it was determined that a cost parameter C of 1.0 and a free parameter of the Gaussian kernel $\gamma$, of ≥ 0.01 gives the highest cross-validation accuracy when applied to electricity usage dataset and hence is best suited for classifying it.

Hence in this work, machine learning, particularly support vector machine was employed in classifying users and also predicting user's activities in terms of energy usage, with the bid to detecting fraudulent users hence ensuring prompt disconnection of such from the grid. Using LibSVM, a library wrapper for SVM, a 79.46% accuracy in classification of users and subsequent prediction of future outcomes given similar parameters in terms of energy usage information was achieved.

However, when the instantaneous electric energy readings of consumers were taken and first pre-classified using ELUPCI, a novel pre-classifier which was modelled to serve as an input to the SVM classifier, a 98.65% classification accuracy was achieved, showing about 19% improvement in the classifier output due to the pre-classification.

This implies that using the developed model, electricity utility companies can determine when a consumer’s electric energy usage becomes abnormal with as much as 98.65% accuracy hence enhancing prompt detection of non-technical losses. Such prompt and accurate detection of NTL will enable them save lots in both human and capital resources.

## 5.2 Contributions to Knowledge

- This research has been able to design an algorithm that could be used to classify energy users into categories preceding an ML run for predicting fraudulent activities in the usage of electric energy.
- The algorithm designed was further developed into a novel smart program called ELUPCI (Electricity Usage Pre-Classifier Interface), which instantaneously classifies users based on their usage information in real-time
- LIBSVM has been demonstrated to be effective in classifying and predicting energy usage patterns given the correct parameters to approximately 80% accuracy. This goes to complement the works of some other researchers in this field.
- ELUPCI enhanced classification was demonstrated to improve the classifier accuracy of the SVM classifier by over 19%.
- Having such accuracy in classification and prediction outcome of energy consumers will arm utility companies with an invaluable tool for detecting and hence preventing NTL in that sector

## 5.2 Recommendations

This research work concentrated on scenarios where users bypass their energy meters or manipulate the AMI infrastructure in such a way that consumed energy is not accurately recorded and hence not reported nor billed.

It is recommended therefore that further research be conducted taking into accounts other forms of fraudulently consuming energy not paid for like illegal generation of energy credits, partial connection of total house load to the meter and even the use of non-smart meter as is still obtained in most developing countries.

ELUPCI could also be further developed to accommodate variations in consumptions due to factors like season of the year, day of the week, geographic location, etc.

Utilities and responsible government agencies should take the issue of collection of data, and hence proper monitoring of utilized energy seriously in order to forestall/prevent the colossal waste being recorded in that sector due to non-technical losses (including vandalism, theft and fraud).

It is also recommended that other aggregations of machine learning profiles be explored in Classification, regression etc. in other to improve the result achieved in this work.

## References

Actaris Metering Systems. (2005, June). *Wirelss data communication solution for prepayment meters.* Retrieved July 8, 2014, from http://www.ee.co.za/wp-content/uploads/legacy/Ditrib%201.pdf

Anas, M., Javaid, N., Mahmood, Raza, S. M., Qasim, U., & Khan, Z. A. (2012). Minimizing Electricity Theft Using Smart Meters in AMI. *3PGCIC '12 Proceedings of the 2012 Seventh International Conference on P2P, Parallel, Grid, Cloud and Internet Computing* (pp. 176-182). Victoria, Canada: IEEE Computer Society.

Anichebe, J. (2013). Overview of the Nigerian Electricity Industry (Roles, Responsibilities, Structure, Expectation). Abuja: Bureau of Public Enterprise. Retrieved from Nigerian Electricity Privatization (PHCN). .

Antmann, P. (2009). Reducing technical and non-technical losses in the power sector. *Background paper for the WBG Energy Strategy, Tech. Rep.,.* Washington, DC, USA: The World Bank.

Bluman, A. G. (2009). *Elementary Statistics: A step by step approach* (7th ed.). Mcgraw hill.

Browniee, J. (2017, January). *A Tour of Machine Learning Algorithms*. Retrieved 2 17, 2017, from Machine Learning Mastery: http://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/

Bureau of Public Enterprises (BPE). (2013). Power Generation (Status and Outlook). Abuja, Nigeria: Presidential Task Force on Power, Federal Republic of Nigeria.

Burges, C. J. (1998). Geometry and invariance in kernel based methods. In B. SchÖlkopf, C. J. Burges, & A. J. Smola, *Advances in Kernel Methods - Support Vector Learning.* Cambridge, MA.: MIT Press.

CBC News. (2010). *Electricity theft by B.C. grow-ops costs $100m a year.* Retrieved November 2014, from http://www.cbc.ca/news/canada/britishcolumbia/electricity-theft-by-b-c-grow-ops-cost-ayear-1.969837

Chang, C.-C., & Lin, C.-J. (2013, March 4). *LIBSVM - A Library for Support Vector Machines*. Retrieved from URL http://www.csie.ntu.edu.tw/~cjlin/libsvm

Cristianini, N., & Taylor, J. S. (2000). *An Introduction to Support Vector Machines.* Cambridge, MA: Cambridge University Press.

Depuru, S. S. (2012). *Modeling, Detection and Prevention of Electricity Theft for Enhanced Performance and Security of Power Grid.* Toledo: University of Toledo.

Electric Power Research Institute. (2007, February). *Advanced Metering Infrastructure (AMI).* Retrieved November 2014, from https://www.ferc.govt/EventCalender/Files/20070423091846-EPRI%20-%20Advanced%20Metering.pdf

Fan, R. E., Chen, P., & Lin, C. (2005). Working set selection using second order. *Journal of Machine Learning Research, 6*, 1889-1918. Retrieved from http://www.csie.ntu.edu.tw/~cjlin/papers/quadworkset.pdf

Federal Court of Audit, Brazil. (2007). *Operational audit report held in national agency of electrical energy.* Technical Report, Federal Court of Audit, Aneel, Brazil.

Federal Republic of Nigeria Official Gazette. (2007, December 24). *Nigerian Electricity Regulatory Commission's Meter Reading, Billing, Cash Collection and Credit Management for Electricity Supplies Regulations.* Retrieved April 23, 2014, from Nigerian Electricity Privatization: www.nigerianelectricityprivatization.com/?page_id=152

Federal Republic of Nigeria Official Gazette. (2012, August 1). *Nigerian Electricity Regulatory Commission(Methodology for Estimated Billing) Regulation.* Retrieved April 23, 2014, from Nigerian Electricity Privatization (PHCN): www.nigerianelectricityprivatization.com/?page_id=152

Fernandez-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014, 10). Do we Need Hundreds of Classifiers to Solve Real World Classification Problems? *Journal of Machine Learning Research 15*, 3133-3181.

GSM Association. (2010). *Market Data Summary.* Retrieved November 2014, from http://www.gsmworld.com/newsroom/marketdata/market_data_summary.htm

Hartmann, T., Moawad, A., Fouquet, F., Rechinger, Y., Mouelhi, T., Klein, J., & Le Traon, Y. (2015). Suspicous Electric Consumption Detection Based on Multi-Profiling Using Live Machine Learning. *Smart Grid Communications (SmartGridComm), 2015 IEEE International Conference on.* IEEE.

Hastie, T., Tibshirani, R., & Friedman, J. (2008). *The Elements of Statistical Learning.* Springer.

Hearn, G. B. (2006). *Design and Development of an On-line Vending System for Selling Prepaid Electricity Via the Internet.* Pretoria: University of Pretoria.

Ho, T. K. (1995). Random Decision Forests. *Proceedings of the 3rd International Conference on Document Analysis and Recognition*, (pp. 278-282). Montreal, QC.

Ho, T. K. (1998). The Random Subspace Method for Constructing Decision Forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*(8), 832-844. doi:10.1109/34.709601

Ifedi, V. (2005). Power Reform and Electricity Generation. *The Guardian*.

InetDaemon Enterprises. (2012). *Hayes AT Modem Commands*. Retrieved June 12, 2016, from InetDaemon.com - Free online IT Tutorial and Internet Training: http://www.inetdaemon.com/tutorials/computers/hardware/modems/at_commands.shtml

Jain, A., & Bagree, M. (2011). A Prepaid Meter Using Mobile Communication. *International Journal of Engineering Science and Technology, 3*(3), 160-166.

Kadir, H. A., Abd. Wahab, M. H., Tukiran, Z., & Mutalib, A. A. (2002). Principles of bridge rectifier. *9th WSEAS International Conference.*

Kaplan, R. (1995). *Standard Transfer Specification Guidelines, Measurement and Control Department, .* National PTM&C, Transmission Group, Eskom, .

King, C. (2004). *Advanced Metering Infrastructure (AMI): Overview of system features and capabilities.* eMeter Corporation, U.S.

Kleinberg, E. (1996). An Overtraining-Resistant Stochastic Modeling Method for Pattern Recognition. *Annals of Statistics, 24*(6), 2319-2349. doi:10.1214/aos/1032181157

Knerr, S., Personnaz, L., & Dreyfus, G. (1990). Single-layer learning revisited: a stepwise. In J. Fogelman, *Neurocomputing: Algorithms, Architectures and Applications.* Springer-Verlag.

Lichman, M. (2015, March 13). UCI Machine Learning Repository. *ElectricityLoadDiagrams20112014 Data Set*. University of California, Irvine, School of Information and Computer Sciences. Retrieved October 27, 2016, from UCI Machine Learning Repository: https://archive.ics.uci.edu/ml/datasets/ElectricityLoadDiagrams20112014

Maltoni, D., Maio, D., Jain, A. K., & Prabhaker, S. (2003). *Handbook on rectification.* New York: Springer.

McDaniel, P., & McLaughlin, S. (2009). Security and privacy challenges in the smart grid. *IEE Security and Privacy*, Vol. 7, no. 3, pp 75-77.

McLaughlin, S., Holbert, B., Zonouz, S., & Berthier, R. (2012). AMIDS: A Multi-Sensor Energy Theft Detection Framework for Advanced Metering Infrastructures. *Smart Grid Communications (SmartGridComm), 2012 IEEE Third International Conference on* (pp. 354 - 359). Tainan: Institute of Electrical Electronic Engineering (IEEE).

McLaughlin, S., Podkuiko, D., & McDaniel, P. (2010). Energy Theft in the Advanced Metering Infrastructure. In S.-V. B. Heidelberg, *Critical Information Infrastructures Security* (pp. pp 176-187). Berlin: Springer Berlin Heidelberg.

MIcrochip Technology Inc. (2016). Retrieved June 12, 2016, from Microchip - products: http://www.microchip.com/wwwproducts/Devices.aspx?dDocName=en024153

Microchip Technology Inc. (2016). *World's Most Reliable Memory*. Retrieved June 12, 2016, from Microchip - Design Centers: http://www.microchip.com/stellent/idcplg?IdcService=SS_GET_PAGE&nodeId=2128&param=en025151

Ministry of Power, India. (2013). *Overview of power distribution.* Retrieved November 2014, from http://www.powermin.nic.in

Mohammad, N., Barua, A., & Arafat, m. A. (2013). A Smart Prepaid Energy Metering System to Control Electricity Theft. *2013 International Conference on Power, Energy and Control (ICPEC)* (pp. 562-565). Institute of Electrical Electronic Engineering (IEEE).

Mokshyna, E. (2014). Support Vector Machines. Retrieved 4 10, 2017, from https://dl.acm.org/citation.cfm?id=1968168

Monedero, I., Biscarri, F., & Leon, C. (2006). MIDAS: Detection of Non-technical Losses in Electrical Consumption Using Neural Networks and Statistical Techniques. *International Conference on Computational Science and its Application (ICCSA)* (pp. 725-734). Berlin: Springer-Verilag.

Nagi, J., Yap, K. S., Tiong, S. K., Ahmed, S. K., & Mohammad, A. M. (2008). Detection of Abnormalities and Electricity Theft using Genetic Support Vector Machines. *2008 IEEE Region 10 Conference* (pp. 1-6). IEEE.

National Rationalised Specification. (2000). *Interface Standards: Data Transfer by physical media - System Master Station - Credit dispensing unit, NRS 009-6-4. .*

National Rationalised Specification. (2000). *Electricity meter cards and associated.*

National Rationalised Specification. (2000). *Functional and performance requirements:.*

National Rationalised Specification. (2000). *Functional and performance requirements:System master stations, NRS 009-2-1.*

National Rationalised Specification. (n.d.). *Interface Standards: System Master Station -Credit dispensing unit, NRS 009-6-3.*

(n.d.). *National Rationalised Specification, Functional and performance requirements: Credit dispensing units, NRS 009-2-2.*

Nigeria Electricity Regulation . (2013, February). *Nigeria Electricity Privatisation ((PHCN).* Retrieved April 23, 2014, from Nigerian Electricity Privatisation (PHCN): www.nigeriaelectricityprivatization.com/?page_id=2

Nigerian Electricity Privatization (PHCN). . (2013). *PHCN Successor Companies.* Retrieved April 23, 2014, from Nigerian Electricity Privatization (PHCN). : www.nercng.org/index.php/industry-operators/licensing-procedures/phcn-successor-companies.htm

Nizar, A. H., Dong, Z. Y., & Wang, Y. (2008). Power Utility Nontechnical Loss Analysis With Extreme Learning Machine Method. *IEEE Transactions on Power Systems* (pp. 946 - 955). IEEE.

Orukpe P., Agbontaen F. (2013). Prepaid Meter in Nigeria: The Story so Far and the Way Forward. *Vol 824*, 114-119.

Ovo Energy. (2016, 11). *How much electricity does a home use?* Retrieved 2 6, 2017, from Energy guides: https://www.ovoenergy.com/guides/energyguides/

Parmar, J. (2013, July). *Total Losses in Power Distribution & Transmission Lines-Part1*. Retrieved from Electrical Notes and Articles:

https://electricalnotes.wordpress.com/2013/07/01/total-losses-in-power-distribution-transmission-lines-part-1/

Platt, J. C. (1998). Fast training of support vector machines using sequential minimal optimization. In B. Sch  olkopf, C. J. Burges, & A. J. Smola (Eds.), *Advances in Kernel Methods - Support Vector Learning.* Cambridge, MA: MIT Press.

*Power Holding Company of Nigeria.* (2014, March 31). Retrieved April 20, 2014, from Wikipedia, the free encyclopedia: www.wikipedia.org/w/index.php?title=Power_Holding_Company_of_Nigeria&oldid=602122265

Raza, A., & etal. (2013). Online monitoring of electricity data through wireless transmission using radio frequency. *International Journal of Innovtion and Applied Studies, 3*(1), 5-67.

Rong, J., & etal. (2004, April). Energy-theft detection issues for advanced metering infrastructure in smart grid. *Tsinghua Science and Technology, 19*(2), 105-120.

Saravanavel, P. (1987). Research Methodology. Kitab Mahal, Allahabad.

Sebola, L., & Penzhorn, W. (2003). A Secure Mobile Commerce System for the Vending of Prepaid Electricity Tokens. Pretoria: Satnac.

Shaibu, M. (2012). *PHCN and Banks Introduce Online Electricity Payments.* Retrieved April 23, 2014, from Proshare (intelligent investing): www.proshareng.com/index.php

Shang-Wen, L., & al., e. (2009). Development of a smart power meter for AMI based on ZigBee communication. *International Conference on Power Electronics and Drive Systems* (pp. 661-665). Taipei: IEEE.

Silviya, M. E., Vinodhini, M. K., & Salai, T. T. (2008). *Active Energy Metering IC with Serial Interface.* Paiyanoor Chennai, India: Department of Electronics and Commmunication Engineering, Aarupadai Veedu Institute of Technology, Vinayaka Mission University.

Soma, S., Wang, L., Devabhaktuni, V., & Gudi, N. (2011). Smart Meters for Power Grid - Challenges, Issues, Advantages and Status. *Institute of Electrical Electronic Engineering*.

STS Association. (1999). Electricity Sales Systems - Part 7: Standard transfer specification / The management of cryptographic keys. *Rationalized User Specification, Edition 1.0, NRS 009-7* .

STS Association. (2002). *Electricity Sales Systems - Part 6: Interface standards - Section 7: Standard transfer specification / Credit dispensing unit - Electricity dispenser - Token encoding and data encryption and decryption.* Rationalized User Specification, Edition 2.2, NRS 009-6-7.

STS Association. (2002). Electricity Sales Systems - Part 6: Interface standards - Section 9: Standard transfer specification / Numeric token technology - Token encoding format and physical token definition. *Rationalized User Specification, Edition 1.0, NRS009-6-9*.

STS Association. (2004). Electricity Sales Systems - Part 6: Interface standards - Section 8: Standard transfer specification / Disposable magnetic token technology - Token encoding format and physical token definition,. *Rationalized User Specification, Edition 2.0, NRS 009-6-8*.

Sudheer, K., & etal. (2013). Equipment for Anti-Electricity Stealing with Remote Monitoring. *International Journal of Engineering Research and Applications, 1*(2), 241-245.

Taylor, D. (2002). *Standard Transfer Specification - Synopsis.* STS Association.

Tewari D. D., S. T. (2003). An Assessment of South African Prepaid Electricity experiment, lesson learned, and their policy implications for developing countriees. *Elsevier Energy Policy*, 911-927.

Texas Instruments Inc. (2016). *MAX232 (Active) Dual EIA-232 Driver/Receiver*. Retrieved June 12, 2016, from Texas Instruments: http://www.ti.com/product/MAX232?keyMatch=max232&tisearch=Search-EN-Everything

The University of Waikato. (2013). Waikato Environment for Knowledge Analysis (WEKA). Hamilton, New Zealand. Retrieved from http://www.cs.waikato.ac.nz/~ml/weka/

U.S Energy Information Administration (EIA). (2016, 10 18). *How much electricity does an American home Use*. Retrieved 2 6, 2017, from U.S Energy Information Administration FAQ: https://www.eia.gov/tools/faqs/faq.cfm?id=97&t=3

Vapnik, V. N. (1998). *Statistical Learning Theory.* New York: Wiley.

Verle, M., Mazidi, M. A., & Mazidi, J. G. (2010). *Microcontrollers and Embedded Systems.* Karachi, Pakistan: Pearson Education, NED University of Engineering and Technology.

Wang, D., Yeung, D. S., & Tsang, E. C. (2007, September). Weighted Mahalanobis Distance Kernels for Support Vector Machines. *IEEE Transactions on Neural Networks, 18*(5), 1453-1462.

Wikipedia, the free encyclopedia . ( 2014, September). *Agile software development.* Retrieved October 16, 2014, from http://www.en.wikipedia.org/wiki/Agile_software_development

Wikipedia, the free encyclopedia. (2014, September). *Spira model- [online].* Retrieved October 16, 2014, from http://en.wikipedia.org/wiki/Spiral_model

Wikipedia, the online encyclopedia. (2014, October). *Rapid application development.* Retrieved Occtober 16, 2014, from http://en.wikipedia.org/wiki/Rapid_application_development

Wikipedia, the online encyclopedia. (2014). *Subscriber Identity Module.* Retrieved November 2014, from http://en.wikipedia.org. http://en.wikipedia.org/wiki/Subscriber_identity_module.

Wikipedia, the online encyclopedia. (2014, October). *Waterfall model.* Retrieved October 16, 2014, from http://en.wikipedia.org/wiki/Waterfall_model

Xie, K. (2011). *Support Vector Machine - Concept and matlab build.* Retrieved 9 29, 201, from Michigan State University, College of Engineering: ww.egr.msu.edu/classes/ece480/capstone/spring11/group04/application_Kan.pdf

## Appendix A – Complete Prediction Run Statistics

=== Model information ===

Filename: libsvm-final run model.model

Scheme:weka.classifiers.functions.LibSVM -S 0 -K 2 -D 3 -G 0.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -seed 1

Relation:4SeasonConsumption

Attributes: 11808

[list of attributes omitted]

=== Classifier model ===

LibSVM wrapper, original code by Yasser EL-Manzalawy (= WLSVM)

=== Re-evaluation on test set ===

User supplied test set

Relation: prediction data

Instances: 370

Attributes: 11808

| === Predictions on test set === | | | | |
|---|---|---|---|---|
| inst#, | actual, | predicted, | error, | probability distribution |
| 1 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 2 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 3 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 4 | 3:AMBER | 2:GREEN | + | 0 *1 0 0 |
| 5 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 6 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 7 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 8 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 9 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 10 | 3:AMBER | 2:GREEN | + | 0 *1 0 0 |
| 11 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 12 | 4:BLACK | 4:BLACK | | 0 0 0 *1 |
| 13 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 14 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 15 | 4:BLACK | 4:BLACK | | 0 0 0 *1 |

| | | | | |
|---|---|---|---|---|
| 16 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 17 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 18 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 19 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 20 | 3:AMBER | 2:GREEN | + | 0 *1 0 0 |
| 21 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 22 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 23 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 24 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 25 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 26 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 27 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 28 | 3:AMBER | 2:GREEN | + | 0 *1 0 0 |
| 29 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 30 | 4:BLACK | 4:BLACK | | 0 0 0 *1 |
| 31 | 3:AMBER | 2:GREEN | + | 0 *1 0 0 |
| 32 | 1:RED | 4:BLACK | + | 0 0 0 *1 |
| 33 | 1:RED | 4:BLACK | + | 0 0 0 *1 |
| 34 | 3:AMBER | 2:GREEN | + | 0 *1 0 0 |
| 35 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 36 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 37 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 38 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 39 | 4:BLACK | 4:BLACK | | 0 0 0 *1 |
| 40 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 41 | 4:BLACK | 4:BLACK | | 0 0 0 *1 |
| 42 | 3:AMBER | 2:GREEN | + | 0 *1 0 0 |
| 43 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 44 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 45 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 46 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 47 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 48 | 3:AMBER | 2:GREEN | + | 0 *1 0 0 |
| 49 | 3:AMBER | 2:GREEN | + | 0 *1 0 0 |
| 50 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 51 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 52 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 53 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 54 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 55 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 56 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 57 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 58 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 59 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |

| | | | | |
|---|---|---|---|---|
| 60 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 61 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 62 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 63 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 64 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 65 | 3:AMBER | 2:GREEN | + | 0 *1 0 0 |
| 66 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 67 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 68 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 69 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 70 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 71 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 72 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 73 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 74 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 75 | 3:AMBER | 2:GREEN | + | 0 *1 0 0 |
| 76 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 77 | 3:AMBER | 2:GREEN | + | 0 *1 0 0 |
| 78 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 79 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 80 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 81 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 82 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 83 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 84 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 85 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 86 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 87 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 88 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 89 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 90 | 3:AMBER | 2:GREEN | + | 0 *1 0 0 |
| 91 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 92 | 4:BLACK | 4:BLACK | | 0 0 0 *1 |
| 93 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 94 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 95 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 96 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 97 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 98 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 99 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 100 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 101 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 102 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 103 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |

| | | | | |
|---|---|---|---|---|
| 104 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 105 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 106 | 4:BLACK | 4:BLACK | | 0 0 0 *1 |
| 107 | 4:BLACK | 4:BLACK | | 0 0 0 *1 |
| 108 | 4:BLACK | 4:BLACK | | 0 0 0 *1 |
| 109 | 4:BLACK | 4:BLACK | | 0 0 0 *1 |
| 110 | 4:BLACK | 4:BLACK | | 0 0 0 *1 |
| 111 | 4:BLACK | 4:BLACK | | 0 0 0 *1 |
| 112 | 4:BLACK | 4:BLACK | | 0 0 0 *1 |
| 113 | 4:BLACK | 4:BLACK | | 0 0 0 *1 |
| 114 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 115 | 4:BLACK | 4:BLACK | | 0 0 0 *1 |
| 116 | 4:BLACK | 4:BLACK | | 0 0 0 *1 |
| 117 | 4:BLACK | 4:BLACK | | 0 0 0 *1 |
| 118 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 119 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 120 | 4:BLACK | 4:BLACK | | 0 0 0 *1 |
| 121 | 4:BLACK | 4:BLACK | | 0 0 0 *1 |
| 122 | 4:BLACK | 4:BLACK | | 0 0 0 *1 |
| 123 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 124 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 125 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 126 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 127 | 1:RED | 4:BLACK | + | 0 0 0 *1 |
| 128 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 129 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 130 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 131 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 132 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 133 | 4:BLACK | 4:BLACK | | 0 0 0 *1 |
| 134 | 4:BLACK | 4:BLACK | | 0 0 0 *1 |
| 135 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 136 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 137 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 138 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 139 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 140 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 141 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 142 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 143 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 144 | 1:RED | 4:BLACK | + | 0 0 0 *1 |
| 145 | 3:AMBER | 2:GREEN | + | 0 *1 0 0 |
| 146 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 147 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |

| | | | | |
|---|---|---|---|---|
| 148 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 149 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 150 | 3:AMBER | 2:GREEN | + | 0 *1 0 0 |
| 151 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 152 | 4:BLACK | 4:BLACK | | 0 0 0 *1 |
| 153 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 154 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 155 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 156 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 157 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 158 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 159 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 160 | 4:BLACK | 4:BLACK | | 0 0 0 *1 |
| 161 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 162 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 163 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 164 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 165 | 4:BLACK | 4:BLACK | | 0 0 0 *1 |
| 166 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 167 | 1:RED | 4:BLACK | + | 0 0 0 *1 |
| 168 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 169 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 170 | 4:BLACK | 4:BLACK | | 0 0 0 *1 |
| 171 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 172 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 173 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 174 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 175 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 176 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 177 | 1:RED | 4:BLACK | + | 0 0 0 *1 |
| 178 | 4:BLACK | 4:BLACK | | 0 0 0 *1 |
| 179 | 4:BLACK | 4:BLACK | | 0 0 0 *1 |
| 180 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 181 | 4:BLACK | 4:BLACK | | 0 0 0 *1 |
| 182 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 183 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 184 | 1:RED | 4:BLACK | + | 0 0 0 *1 |
| 185 | 4:BLACK | 4:BLACK | | 0 0 0 *1 |
| 186 | 4:BLACK | 4:BLACK | | 0 0 0 *1 |
| 187 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 188 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 189 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 190 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 191 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 192 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 193 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 194 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 195 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 196 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 197 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 198 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 199 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 200 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 201 | 1:RED | 2:GREEN | + | 0 | *1 | 0 | 0 |
| 202 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 203 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 204 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 205 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 206 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 207 | 3:AMBER | 2:GREEN | + | 0 | *1 | 0 | 0 |
| 208 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 209 | 3:AMBER | 2:GREEN | + | 0 | *1 | 0 | 0 |
| 210 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 211 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 212 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 213 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 214 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 215 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 216 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 217 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 218 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 219 | 1:RED | 2:GREEN | + | 0 | *1 | 0 | 0 |
| 220 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 221 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 222 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 223 | 1:RED | 2:GREEN | + | 0 | *1 | 0 | 0 |
| 224 | 4:BLACK | 4:BLACK | | 0 | 0 | 0 | *1 |
| 225 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 226 | 1:RED | 2:GREEN | + | 0 | *1 | 0 | 0 |
| 227 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 228 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 229 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 230 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 231 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 232 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 233 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 234 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 235 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 236 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 237 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 238 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 239 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 240 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 241 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 242 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 243 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 244 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 245 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 246 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 247 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 248 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 249 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 250 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 251 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 252 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 253 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 254 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 255 | 1:RED | 4:BLACK | + | 0 | 0 | 0 | *1 |
| 256 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 257 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 258 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 259 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 260 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 261 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 262 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 263 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 264 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 265 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 266 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 267 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 268 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 269 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 270 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 271 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 272 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 273 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 274 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 275 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 276 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 277 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 278 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |
| 279 | 2:GREEN | 2:GREEN | | 0 | *1 | 0 | 0 |

| | | | | |
|---|---|---|---|---|
| 280 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 281 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 282 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 283 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 284 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 285 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 286 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 287 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 288 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 289 | 4:BLACK | 4:BLACK | | 0 0 0 *1 |
| 290 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 291 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 292 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 293 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 294 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 295 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 296 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 297 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 298 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 299 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 300 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 301 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 302 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 303 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 304 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 305 | 4:BLACK | 4:BLACK | | 0 0 0 *1 |
| 306 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 307 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 308 | 1:RED | 4:BLACK | + | 0 0 0 *1 |
| 309 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 310 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 311 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 312 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 313 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 314 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 315 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 316 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 317 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 318 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 319 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 320 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 321 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 322 | 4:BLACK | 4:BLACK | | 0 0 0 *1 |
| 323 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |

| | | | | |
|---|---|---|---|---|
| 324 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 325 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 326 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 327 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 328 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 329 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 330 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 331 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 332 | 1:RED | 4:BLACK | + | 0 0 0 *1 |
| 333 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 334 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 335 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 336 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 337 | 4:BLACK | 4:BLACK | | 0 0 0 *1 |
| 338 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 339 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 340 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 341 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 342 | 3:AMBER | 2:GREEN | + | 0 *1 0 0 |
| 343 | 3:AMBER | 2:GREEN | + | 0 *1 0 0 |
| 344 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 345 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 346 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 347 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 348 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 349 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 350 | 3:AMBER | 2:GREEN | + | 0 *1 0 0 |
| 351 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 352 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 353 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 354 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 355 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 356 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 357 | 3:AMBER | 2:GREEN | + | 0 *1 0 0 |
| 358 | 3:AMBER | 2:GREEN | + | 0 *1 0 0 |
| 359 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 360 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 361 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 362 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 363 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 364 | 1:RED | 2:GREEN | + | 0 *1 0 0 |
| 365 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 366 | 3:AMBER | 2:GREEN | + | 0 *1 0 0 |
| 367 | 1:RED | 2:GREEN | + | 0 *1 0 0 |

| | | | | |
|---|---|---|---|---|
| 368 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 369 | 2:GREEN | 2:GREEN | | 0 *1 0 0 |
| 370 | 4:BLACK | 4:BLACK | | 0 0 0 *1 |

=== Summary ===

Correctly Classified Instances 294 79.4595 %

Incorrectly Classified Instances 76 20.5405 %

Kappa statistic 0.4608

Mean absolute error 0.1027

Root mean squared error 0.3205

Total Number of Instances 370

## Appendix B – LIBSVM Classifier Trainer Codes (Chang & Lin, 2013)

```
 /**
  * builds the classifier.
   * @param insts      the training instances
  * @throws Exception  if libsvm encountered a problem   */
 public void buildClassifier(Instances insts) throws Exception {
   m_Filter = null;

   // remove instances with missing class
   insts = new Instances(insts);
   insts.deleteWithMissingClass();

   if (!getDoNotReplaceMissingValues()) {
     m_ReplaceMissingValues = new ReplaceMissingValues();
     m_ReplaceMissingValues.setInputFormat(insts);
     insts = Filter.useFilter(insts, m_ReplaceMissingValues);
   }
   // can classifier handle the data?
   // we check this here so that if the user turns off
   // replace missing values filtering, it will fail
   // if the data actually does have missing values
   getCapabilities().testWithFail(insts);

   double y0 = Double.NaN;
   double y1 = y0;
   int index = -1;
   if (!insts.classAttribute().isNominal()) {
     y0 = insts.instance(0).classValue();
     index = 1;
     while (index < insts.numInstances() && insts.instance(index).classValue() == y0) {
       index++;
     }
     if (index == insts.numInstances()) {
       // degenerate case, all class values are equal
       // we don't want to deal with this, too much hassle
       throw new Exception("All class values are the same. At least two class values should be
different");
     }
     y1 = insts.instance(index).classValue();
   }

   if (getNormalize()) {
     m_Filter = new Normalize();
     ((Normalize)m_Filter).setIgnoreClass(true); // Normalize class as well
     m_Filter.setInputFormat(insts);
     insts = Filter.useFilter(insts, m_Filter);
```

```
    }

    if (!insts.classAttribute().isNominal()) {
      if (m_Filter != null) {
        double z0 = insts.instance(0).classValue();
        double z1 = insts.instance(index).classValue();
        m_x1 = (y0 - y1) / (z0 - z1); // no division by zero, since y0 != y1 guaranteed => z0 != z1
???
        m_x0 = (y0 - m_x1 * z0); // = y1 - m_x1 * z1
      } else {
        m_x1 = 1.0;
        m_x0 = 0.0;
      }
    } else {
      m_x0 = Double.NaN;
      m_x1 = m_x0;
    }
    // nominal to binary
    m_NominalToBinary = new NominalToBinary();
    m_NominalToBinary.setInputFormat(insts);
    insts = Filter.useFilter(insts, m_NominalToBinary);

    double[] vy = new double[insts.numInstances()];
    svm_node[][] vx = new svm_node[insts.numInstances()][];
    int max_index = 0;
    for (int d = 0; d < insts.numInstances(); d++) {
      Instance inst = insts.instance(d);
      vx[d] = instanceToArray(inst);
      if (vx[d].length > 0) {
        max_index = Math.max(max_index, vx[d][vx[d].length - 1].index);
      }
      vy[d] = inst.classValue();
    }

    // calculate actual gamma
    if (getGamma() == 0)
      m_GammaActual = 1.0 / max_index;
    else
      m_GammaActual = m_Gamma;

    svm_problem p = getProblem(vx, vy);
    svm_parameter pars = getParameters();

    // check parameters
    String error_msg = svm.svm_check_parameter(p, pars);

    if (error_msg != null)
```

```
    throw new Exception("Error: " + error_msg);

  // make probability estimates deterministic from run to run
  svm.rand.setSeed(m_Seed);

  // Change printing function if no debugging output is required
  if (!getDebug()) {
    svm.svm_set_print_string_function(new svm_print_interface() {
      @Override
      public void print(String s) {
        // Do nothing
      }
    });
  }

  // train model
  m_Model = svm.svm_train(p, pars);

  // save internal model?
  if (!m_ModelFile.isDirectory()) {
    svm.svm_save_model(m_ModelFile.getAbsolutePath(), m_Model);
  }
}
```

## Appendix C – LIBSVM Predict Codes – (Chang & Lin, 2013)

```
/**
  * Computes the distribution for a given instance.
  * In case of 1-class classification, 1 is returned at index 0 if libsvm
  * returns 1 and NaN (= missing) if libsvm returns -1.
  *
  * @param instance          the instance for which distribution is computed
  * @return                  the distribution
  * @throws Exception                if the distribution can't be computed successfully
  */
 public double[] distributionForInstance (Instance instance) throws Exception {

  int[] labels = new int[instance.numClasses()];
  double[] prob_estimates = null;

  if (m_ProbabilityEstimates) {
   svm.svm_get_labels(m_Model, labels);

   prob_estimates = new double[instance.numClasses()];
  }

  if (!getDoNotReplaceMissingValues()) {
   m_ReplaceMissingValues.input(instance);
   m_ReplaceMissingValues.batchFinished();
   instance = m_ReplaceMissingValues.output();
  }

  if (m_Filter != null) {
   m_Filter.input(instance);
   m_Filter.batchFinished();
   instance = m_Filter.output();
  }

  m_NominalToBinary.input(instance);
  m_NominalToBinary.batchFinished();
  instance = m_NominalToBinary.output();

  svm_node[] x = instanceToArray(instance);
  double v;
  double[] result = new double[instance.numClasses()];
  if (m_ProbabilityEstimates
       && ((m_SVMType == SVMTYPE_C_SVC) || (m_SVMType == SVMTYPE_NU_SVC))) {
   v = svm.svm_predict_probability(m_Model, x, prob_estimates);

   // Return order of probabilities to canonical weka attribute order
   for (int k = 0; k < prob_estimates.length; k++) {
```

```
      result[labels[k]] = prob_estimates[k];
    }
  } else {
    v = svm.svm_predict(m_Model, x);

    if (instance.classAttribute().isNominal()) {
      if (m_SVMType == SVMTYPE_ONE_CLASS_SVM) {
        if (v > 0)
          result[0] = 1;
        else
          // outlier (interface for Classifier specifies that unclassified instances
          // should return a distribution of all zeros)
          result[0] = 0;
      } else {
        result[(int) v] = 1;
      }
    } else {
      result[0] = v * m_x1 + m_x0;
    }
  }

  return result;
}
```

## Appendix D – Electricity Usage Pre-Classifier Interface (ELUPCI)

```
<?php
	//error_reporting(0);
	require_once 'db.inc.php';
	require_once 'security.php';
	function loggedin($db, $table, $username, $password)
	{
		$password = md5($password);
		if (!empty($username) && !empty($password)) {
			# code...
			$result = $db->query("SELECT * FROM $table WHERE username = 
'$username' AND password = '$password'");
			if ($result->num_rows) {
				# code...
				return true;
			}else{
				return false;
			}
		}else
		{
			return false;
		}
	}
	function CreateTable($db, $tableName)
	{
		if (!empty($tableName)) {
			# code...
			$sql = "CREATE TABLE $tableName (
			id INT(6) UNSIGNED AUTO_INCREMENT PRIMARY KEY,
			dateTime DATETIME,
			readings FLOAT(30)
			)";
			if ($result = $db->query($sql)) {
				# code...
				return true;
			}else{
				return false;
			}
		}
	}
	function FetchFlag($db, $customerID)
	{
		$result = $db->query("SELECT * FROM $customerID ORDER BY id DESC");
		$total = $result->num_rows;
		$zeros = 0;
		$count = 0;
```

```
        if ($result->num_rows) {
                # code...
                while ($rows = $result->fetch_object())
                {
                        $value = $rows->readings;
                        if ($value == 0) {
                                # code...
                                $zeros++;
                        }
                        $count++;
                        if ($count == 289) {
                                # code...
                                break;
                        }
                }
                if ($zeros != 0) {
                        # code...
                $onethird = (1/3)*$total;
                $twothird = (2/3)*$total;
                $test = $zeros;
                /*if ($test <= $onethird) {
                        # code...
                        $flag = 'GREEN';
                        return $flag;
                }else*/
                        if ($test > $onethird && $test <= $twothird) {
                        # code...
                        $flag = 'RED';
                        return $flag;
                }elseif ($test > $twothird && $test == $total) {
                        # code...
                        $flag = 'BLACK';
                        return $flag;
                }
                }else{
                        # code...
                        $flag = 'GREEN';
                        return $flag;
                }

        }else{
                        # code...
                        $flag = 'NONE';
                        return $flag;
                }

}
```

```
	function AddCustomer($db, $table, $customerID, $fullname)
	{
		if ($result = $db->query("INSERT INTO $table(id,fullname,customerID,reg_date,flag) VALUES('','$fullname','$customerID',NOW(),'NONE')")) {
			# code...
			return true;
		}else{
			return false;
		}
	}
	function AddValues($db, $table, $reading)
	{
		if ($result = $db->query("INSERT INTO $table(id, dateTime, readings) VALUES('',NOW(),$reading)")) {
			# code...
			return true;
		}else{
			return false;
		}
	}
```

## Appendix E – ELUPCI Enhanced LIBSVM Prediction Details

=== Run information ===

Scheme:weka.classifiers.functions.LibSVM -S 0 -K 2 -D 3 -G 1.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -seed 1

Relation: 4SeasonConsumptionELUPCI125

Instances: 125

Attributes: 11809

[list of attributes omitted]

Test mode:10-fold cross-validation

=== Classifier model (full training set) ===

LibSVM wrapper, original code by Yasser EL-Manzalawy (= WLSVM)

Time taken to build model: 4.14 seconds

=== Predictions on test data ===

| inst#, | actual, | predicted, | error, | probability distribution | |
|---|---|---|---|---|---|
| 1 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 2 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 3 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 4 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 5 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 6 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 7 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 8 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 9 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 10 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 11 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 12 | 2:BLACK | 2:BLACK | | 0 | *1 |
| 13 | 2:BLACK | 2:BLACK | | 0 | *1 |
| 14 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 15 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 16 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 17 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 18 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 19 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 20 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 21 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 22 | 1:GREEN | 1:GREEN | | *1 | 0 |

| | | | | | |
|---|---|---|---|---|---|
| 23 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 24 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 25 | 2:BLACK | 2:BLACK | | 0 | *1 |
| 26 | 2:BLACK | 2:BLACK | | 0 | *1 |
| 27 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 28 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 29 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 30 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 31 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 32 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 33 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 34 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 35 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 36 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 37 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 38 | 2:BLACK | 2:BLACK | | 0 | *1 |
| 39 | 2:BLACK | 2:BLACK | | 0 | *1 |
| 40 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 41 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 42 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 43 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 44 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 45 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 46 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 47 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 48 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 49 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 50 | 2:BLACK | 2:BLACK | | 0 | *1 |
| 51 | 2:BLACK | 2:BLACK | | 0 | *1 |
| 52 | 2:BLACK | 2:BLACK | | 0 | *1 |
| 53 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 54 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 55 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 56 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 57 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 58 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 59 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 60 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 61 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 62 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 63 | 2:BLACK | 2:BLACK | | 0 | *1 |
| 64 | 2:BLACK | 2:BLACK | | 0 | *1 |
| 65 | 2:BLACK | 2:BLACK | | 0 | *1 |
| 66 | 1:GREEN | 1:GREEN | | *1 | 0 |

| | | | | | |
|---|---|---|---|---|---|
| 67 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 68 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 69 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 70 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 71 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 72 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 73 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 74 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 75 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 76 | 2:BLACK | 2:BLACK | | 0 | *1 |
| 77 | 2:BLACK | 2:BLACK | | 0 | *1 |
| 78 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 79 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 80 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 81 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 82 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 83 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 84 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 85 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 86 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 87 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 88 | 2:BLACK | 2:BLACK | | 0 | *1 |
| 89 | 2:BLACK | 2:BLACK | | 0 | *1 |
| 90 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 91 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 92 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 93 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 94 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 95 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 96 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 97 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 98 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 99 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 100 | 2:BLACK | 2:BLACK | | 0 | *1 |
| 101 | 2:BLACK | 2:BLACK | | 0 | *1 |
| 102 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 103 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 104 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 105 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 106 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 107 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 108 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 109 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 110 | 1:GREEN | 1:GREEN | | *1 | 0 |

| | | | | | |
|---|---|---|---|---|---|
| 111 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 112 | 2:BLACK | 2:BLACK | | 0 | *1 |
| 113 | 2:BLACK | 2:BLACK | | 0 | *1 |
| 114 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 115 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 116 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 117 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 118 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 119 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 120 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 121 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 122 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 123 | 1:GREEN | 1:GREEN | | *1 | 0 |
| 124 | 2:BLACK | 2:BLACK | | 0 | *1 |
| 125 | 2:BLACK | 1:GREEN | + | *1 | 0 |

=== Stratified cross-validation ===

=== Summary ===

| | | |
|---|---|---|
| Correctly Classified Instances | 124 | 99.2 % |
| Incorrectly Classified Instances | 1 | 0.8 % |
| Kappa statistic | 0.9719 | |
| Mean absolute error | 0.008 | |
| Root mean squared error | 0.0894 | |
| Relative absolute error | 2.7223 % | |
| Root relative squared error | 23.4685 % | |
| Total Number of Instances | 125 | |

=== Detailed Accuracy By Class ===

| | TP Rate | FP Rate | Precision | Recall | F-Measure | ROC Area | Class |
|---|---|---|---|---|---|---|---|
| | 1 | 0.045 | 0.99 | 1 | 0.995 | 0.977 | GREEN |
| | 0.955 | 0 | 1 | 0.955 | 0.977 | 0.977 | BLACK |
| Weighted Avg. | 0.992 | 0.037 | 0.992 | 0.992 | 0.992 | 0.977 | |

=== Confusion Matrix ===

```
a      b      <-- classified as
103    0   |  a = GREEN
1      21  |  b = BLACK
```

# Appendix F – A Sample the dataset (370 Instances/Electricity Consumers)

| Time | 1/1/2012 0:00 | 1/1/2012 0:15 | 1/1/2012 0:30 | 1/1/2012 0:45 | 1/1/2012 1:00 | 1/1/2012 1:15 | 1/1/2012 1:30 | 1/1/2012 1:45 | 1/1/2012 2:00 |
|---|---|---|---|---|---|---|---|---|---|
| MT_001 | 0 | 3.807106599 | 5.076142132 | 3.807106599 | 3.807106599 | 5.076142132 | 3.807106599 | 3.807106599 | 6.345177665 |
| MT_002 | 0 | 22.75960171 | 22.75960171 | 22.75960171 | 22.75960171 | 22.04836415 | 22.75960171 | 22.75960171 | 24.89331437 |
| MT_003 | 0 | 77.32406603 | 77.32406603 | 77.32406603 | 77.32406603 | 77.32406603 | 77.32406603 | 77.32406603 | 77.32406603 |
| MT_004 | 0 | 136.1788618 | 136.1788618 | 140.2439024 | 140.2439024 | 146.3414634 | 134.1463415 | 132.1138211 | 136.1788618 |
| MT_005 | 0 | 70.73170732 | 73.17073171 | 69.51219512 | 75.6097561 | 73.17073171 | 73.17073171 | 67.07317073 | 67.07317073 |
| MT_006 | 0 | 351.1904762 | 354.1666667 | 348.2142857 | 339.2857143 | 342.2619048 | 336.3095238 | 342.2619048 | 336.3095238 |
| MT_007 | 0 | 9.609949124 | 9.044657999 | 8.479366874 | 7.348784624 | 6.783493499 | 6.218202374 | 6.783493499 | 7.914075749 |
| MT_008 | 0 | 279.4612795 | 279.4612795 | 279.4612795 | 279.4612795 | 265.993266 | 272.7272727 | 282.8282828 | 282.8282828 |
| MT_009 | 0 | 75.17482517 | 73.42657343 | 75.17482517 | 68.18181818 | 69.93006993 | 66.43356643 | 64.68531469 | 68.18181818 |
| MT_010 | 0 | 87.09677419 | 84.94623656 | 91.39784946 | 88.17204301 | 86.02150538 | 83.87096774 | 79.56989247 | 79.56989247 |
| MT_011 | 0 | 60.35767511 | 57.37704918 | 59.61251863 | 61.84798808 | 59.61251863 | 57.37704918 | 54.39642325 | 52.1609538 |
| MT_012 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| MT_013 | 0 | 45.6053068 | 47.26368159 | 53.06799337 | 44.7761194 | 48.09286899 | 53.06799337 | 43.11774461 | 39.80099502 |
| MT_014 | 0 | 37.0134014 | 35.09891512 | 37.0134014 | 37.0134014 | 37.0134014 | 32.54626675 | 33.18442884 | 33.82259094 |
| MT_015 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| MT_016 | 0 | 56.9476082 | 57.51708428 | 55.23917995 | 50.6833713 | 51.82232346 | 58.08656036 | 50.11389522 | 48.97494305 |
| MT_017 | 0 | 76.37017071 | 74.57322552 | 76.37017071 | 77.26864331 | 77.26864331 | 73.67475292 | 70.08086253 | 71.87780773 |
| MT_018 | 0 | 309.9041534 | 341.8530351 | 348.2428115 | 313.0990415 | 319.4888179 | 300.3194888 | 287.5399361 | 316.2939297 |
| MT_019 | 0 | 16.58291457 | 15.57788945 | 15.57788945 | 14.57286432 | 16.08040201 | 15.57788945 | 15.07537688 | 15.07537688 |
| MT_020 | 0 | 55.50660793 | 51.10132159 | 53.74449339 | 59.91189427 | 51.10132159 | 51.10132159 | 51.98237885 | 50.22026432 |
| MT_021 | 0 | 204.1884817 | 196.3350785 | 191.0994764 | 198.9528796 | 188.4816754 | 188.4816754 | 180.6282723 | 178.0104712 |
| MT_022 | 0 | 29.8879203 | 28.01992528 | 28.64259029 | 31.13325031 | 29.26525529 | 30.51058531 | 32.37858032 | 32.37858032 |
| MT_023 | 0 | 11.21372032 | 9.234828496 | 11.21372032 | 8.575197889 | 9.234828496 | 8.575197889 | 10.55408971 | 9.234828496 |
| MT_024 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| MT_025 | 0 | 289.7196262 | 299.0654206 | 294.3925234 | 285.046729 | 271.0280374 | 257.0093458 | 257.0093458 | 242.9906542 |
| MT_026 | 0 | 40.6779661 | 39.83050847 | 40.6779661 | 38.13559322 | 36.44067797 | 35.59322034 | 37.28813559 | 36.44067797 |
| MT_027 | 0 | 35.3227771 | 32.88672351 | 32.88672351 | 31.05968331 | 31.66869671 | 32.27771011 | 29.23264312 | 28.62362972 |
| MT_028 | 0 | 136.6698749 | 139.5572666 | 147.2569779 | 142.4446583 | 151.1068335 | 147.2569779 | 141.4821944 | 145.33205 |
| MT_029 | 0 | 63.24358172 | 60.11271133 | 59.48653726 | 56.98184095 | 57.60801503 | 52.59862242 | 50.09392611 | 47.58922981 |
| MT_030 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| MT_031 | 0 | 129.9342105 | 118.4210526 | 110.1973684 | 110.1973684 | 110.1973684 | 111.8421053 | 116.7763158 | 106.9078947 |
| MT_032 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| MT_033 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| MT_034 | 0 | 27.23463687 | 28.63128492 | 25.83798883 | 27.23463687 | 25.1396648 | 25.1396648 | 27.23463687 | 26.53631285 |
| MT_035 | 0 | 150.4922644 | 147.6793249 | 149.0857947 | 147.6793249 | 149.0857947 | 150.4922644 | 156.1181435 | 150.4922644 |
| MT_036 | 0 | 10.32702238 | 12.04819277 | 12.04819277 | 13.76936317 | 15.49053356 | 15.49053356 | 13.76936317 | 12.04819277 |
| MT_037 | 0 | 141.1992263 | 133.4622824 | 131.5280464 | 139.2649903 | 145.0676983 | 137.3307544 | 123.7911025 | 123.7911025 |
| MT_038 | 0 | 117.8010471 | 113.8743455 | 113.8743455 | 116.4921466 | 115.1832461 | 123.0366492 | 112.565445 | 112.565445 |
| MT_039 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| MT_040 | 0 | 170.3296703 | 175.8241758 | 173.0769231 | 167.5824176 | 159.3406593 | 151.0989011 | 151.0989011 | 142.8571429 |
| MT_041 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| MT_042 | 0 | 64.80304956 | 53.36721728 | 57.17916137 | 54.63786531 | 52.09656925 | 57.17916137 | 57.17916137 | 52.09656925 |
| MT_043 | 0 | 1409.090909 | 1363.636364 | 1272.727273 | 1272.727273 | 1159.090909 | 1090.909091 | 1159.090909 | 1204.545455 |
| MT_044 | 0 | 51.59010601 | 49.46996466 | 48.76325088 | 50.88339223 | 47.34982332 | 47.34982332 | 42.40282686 | 40.28268551 |
| MT_045 | 0 | 72.75803723 | 69.37394247 | 74.4500846 | 65.98984772 | 65.98984772 | 64.29780034 | 62.60575296 | 64.29780034 |
| MT_046 | 0 | 255.9241706 | 279.6208531 | 274.8815166 | 251.1848341 | 251.1848341 | 236.9668246 | 208.5308057 | 218.0094787 |
| MT_047 | 0 | 38.66525424 | 36.01694915 | 35.48728814 | 36.54661017 | 34.95762712 | 32.30932203 | 32.30932203 | 32.30932203 |
| MT_048 | 0 | 22.50130822 | 22.50130822 | 21.45473574 | 25.64102564 | 19.36159079 | 22.50130822 | 19.88487703 | 21.45473574 |
| MT_049 | 0 | 1434.782609 | 1347.826087 | 1347.826087 | 1293.478261 | 1239.130435 | 1250 | 1239.130435 | 1228.26087 |
| MT_050 | 0 | 266.9039146 | 263.3451957 | 252.6690391 | 249.1103203 | 245.5516014 | 249.1103203 | 245.5516014 | 224.1992883 |
| MT_051 | 0 | 63.94316163 | 62.1669627 | 56.8383659 | 56.8383659 | 60.39076377 | 63.94316163 | 58.61456483 | 60.39076377 |
| MT_052 | 0 | 57.09876543 | 58.64197531 | 49.38271605 | 59.41358025 | 55.55555556 | 59.41358025 | 54.01234568 | 53.24074074 |
| MT_053 | 0 | 106.93302 | 108.1081081 | 108.1081081 | 115.1586369 | 111.6333725 | 103.4077556 | 105.7579318 | 108.1081081 |
| MT_054 | 0 | 93.52517986 | 92.92565947 | 90.52757794 | 89.92805755 | 89.92805755 | 93.52517986 | 85.73141487 | 83.93285372 |
| MT_055 | 0 | 80.36338225 | 73.37526205 | 68.48357792 | 67.08595388 | 62.89308176 | 63.59189378 | 61.49545772 | 61.49545772 |
| MT_056 | 0 | 53.4351145 | 52.4173028 | 50.38167939 | 49.36386768 | 49.87277354 | 52.4173028 | 48.34605598 | 47.32824427 |
| MT_057 | 0 | 40.88586031 | 41.73764906 | 43.44122658 | 43.44122658 | 38.33049404 | 37.47870528 | 38.33049404 | 33.2197615 |
| MT_058 | 0 | 171.2887439 | 172.9200653 | 166.3947798 | 163.132137 | 159.8694943 | 158.2381729 | 151.7128874 | 150.0815661 |
| MT_059 | 0 | 151.8876207 | 158.0333626 | 190.5179982 | 206.3213345 | 187.8841089 | 190.5179982 | 192.2739245 | 184.3722564 |
| MT_060 | 0 | 40.78549849 | 41.28902316 | 42.29607251 | 41.79254783 | 36.25377644 | 33.73615307 | 33.2326284 | 32.22557905 |
| MT_061 | 0 | 139.8201145 | 133.2788226 | 132.4611611 | 130.0081766 | 133.2788226 | 134.9141455 | 126.7375307 | 130.0081766 |
| MT_062 | 0 | 99.58506224 | 96.26556017 | 95.43568465 | 93.77593361 | 92.94605809 | 92.94605809 | 92.94605809 | 87.96680498 |
| MT_063 | 0 | 44.58598726 | 46.17834395 | 44.58598726 | 43.78980892 | 39.8089172 | 39.8089172 | 39.8089172 | 38.21656051 |
| MT_064 | 0 | 1333.333333 | 1384.615385 | 1256.410256 | 1307.692308 | 1230.769231 | 1230.769231 | 1282.051282 | 1179.487179 |
| MT_065 | 0 | 94.64285714 | 92.85714286 | 91.07142857 | 82.14285714 | 80.35714286 | 78.57142857 | 85.71428571 | 103.5714286 |
| MT_066 | 0 | 93.77462569 | 87.47044917 | 84.31836091 | 88.25847124 | 89.83451537 | 92.98660362 | 89.83451537 | 88.25847124 |
| MT_067 | 0 | 238.8663968 | 226.7206478 | 214.5748988 | 234.8178138 | 226.7206478 | 218.6234818 | 218.6234818 | 234.8178138 |
| MT_068 | 0 | 232.0185615 | 225.0580046 | 213.4570766 | 225.0580046 | 213.4570766 | 201.8561485 | 190.2552204 | 194.8955916 |
| MT_069 | 0 | 291.1392405 | 278.4810127 | 265.8227848 | 263.2911392 | 258.2278481 | 255.6962025 | 255.6962025 | 253.164557 |
| MT_070 | 0 | 74.5176314 | 69.86027944 | 69.86027944 | 71.19095143 | 71.19095143 | 71.19095143 | 65.20292748 | 65.20292748 |
| MT_071 | 0 | 65.85612969 | 61.29685917 | 63.32320162 | 61.80344478 | 57.75075988 | 57.75075988 | 57.75075988 | 54.7112462 |
| MT_072 | 0 | 128.9111389 | 131.4142678 | 135.1689612 | 127.6595745 | 133.9173967 | 128.9111389 | 120.1501877 | 122.6533166 |
| MT_073 | 0 | 39.73843058 | 38.73239437 | 39.23541247 | 40.7444668 | 36.21730382 | 35.71428571 | 36.21730382 | 35.71428571 |
| MT_074 | 0 | 154.1802389 | 149.8371336 | 147.6655809 | 150.9229099 | 141.1509229 | 133.5504886 | 132.4647123 | 130.2931596 |
| MT_075 | 0 | 106.6536204 | 90.99804305 | 90.01956947 | 88.06262231 | 85.12720157 | 84.14872798 | 90.01956947 | 82.19178082 |
| MT_076 | 0 | 100.9708738 | 93.2038835 | 94.17475728 | 96.11650485 | 96.11650485 | 99.02912621 | 99.02912621 | 99.02912621 |
| MT_077 | 0 | 39.78494624 | 41.93548387 | 41.93548387 | 41.39784946 | 40.32258065 | 40.86021505 | 40.86021505 | 39.78494624 |
| MT_078 | 0 | 77.9696494 | 81.10936682 | 82.1559393 | 80.58608059 | 72.73678702 | 73.26007326 | 67.50392465 | 65.93406593 |
| MT_079 | 0 | 46.06661942 | 43.23175053 | 40.39688164 | 41.10559887 | 41.81431609 | 43.23175053 | 41.10559887 | 40.39688164 |