Energy Analytics Assisted Model for Behaviour Monitoring and Abnormality Detection

Thesis submitted in partial fulfillment of the requirements for the award of degree of

Master of Engineering

in

Computer Science and Engineering

Submitted by

Anupam Kaur Grewal

(**Roll no: 801632001**)

Under the supervision of:
Dr. Maninder Kaur
Assistant Professor, CSED

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT
THAPAR INSTITUTE OF ENGINEERING AND TECHNOLOGY
PATIALA-147004

June 2018

Certificate

I hereby certify that the work which is being presented in the thesis entitled, *"Energy Analytics Assisted Model for Behaviour Monitoring and Abnormality Detection"*, in partial fulfillment of the requirements for the award of degree of Master of Engineering in *Computer Science and Engineering* submitted in Computer Science and Engineering Department of Thapar Institute of Engineering and Technology (Deemed to be University), Patiala, is an authentic record of my own work carried out under the supervision of *Dr. Maninder Kaur* and refers other researchers work which are duly listed in the reference section.

The matter presented in the thesis has not been submitted for award of any other degree of this or any other University.

Anupam Kaur Grewal

Signature:

(Anupam Kaur Grewal)

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.

Maninder Kaur

Dr. Maninder Kaur

Assistant Professor,

CSED

Abstract

With the rise in population of world, providing door to door healthcare services to citizens is turning into very expensive affair for the governments, thus there is dire need of digital healthcare model to provide services to citizens in efficient manner. Smart cities are becoming a largest infrastructure modernization process. Smart infrastructure is bringing revolutionary changes in different fields. Different smart techniques are employed in smart cities to make them better than traditional cities. One among the advancement is usage of smart data associated with human activities and encompassing environment for health care facilities. In this study, smart meter data is used for health care facilities. Smart meter data is utilized to figure out the relationship between energy utilization and daily life activities. Further, anomaly detect model for day to day life is proposed to detect the abnormal activity patterns of occupants of smart homes. Details of proposed work and implementation results are recorded in this study.

Acknowledgements

I would like to express my deep gratitude to my supervisor **Dr. Maninder Kaur** for their invaluable advice and encouragement at every step of my M.E. program. Without her unfailing support and belief in me, this thesis would not have been possible. Their contribution to this thesis goes well beyond their role as an academic supervisor and includes constant support on a personal level without which this journey may never have been completed. And for this, I am truly grateful.

I would like to express my gratitude to the Professor and Head, CSED **Dr.Maninder Singh** for his constant motivation and encouragement.He sets high principles for his students and motivates and guides them to meet those principles.

Before ending I would like to thank my parents and friends for their love, motivation, support and blessings. They have been a constant source of love, concern, support and strength for me all these years.

Finally, I would like to thank the management of Thapar Institute of Engineering and Technology for providing me a great opportunity for learning, not just in academics but also in many other creative things.

I sincerely regret any inadvertent omissions. With my heartiest thanks to all.

Anupam Kaur Grewal

Table of Contents

List of Figures

List of Tables

List of Notations

$G1$	Normal Region1
$G2$	Normal Region 2
$o1$	Anomaly Region 1
$o2$	Anomaly Region 2
n	Total no of values
i	Required values
$Modela$	Actual Model
$Model_p$	Predicted Model
P_o	Observed Probability
P_i	Probability of Success
k	Value for k means clustering
k_min	Minimum value of k for k_means
$x(k)$	Centroid for x-means
$x(a)$	new centroid1
$x(b)$	new centroid 2
log_p	no. of observations
log_q	log Probability
x	single Instance
X	group Of values

List of Abbreviations

AMI	Advanced Metering infrastructure
AAI	Ambient Assisted Living
ADL	Activities of Daily life
BIC	Bayesian Information Critarion
C 4.5	Algorithm based on decision tree for classification
CCTV	Close Circuit Television
ECLAT	Equivalence Class Transformation
F-Measure	Weighted Harmonic Mean of Precision and Recall
FP	Frequent Patterns
FP	False Positive
GA	Genetic Algorithm
ID	Identity
J48	Type of decision tree algorithm
K-Means	Clustering Algorithm
KNN	K-Nearest Neighbour
LOF	Local Outlier Factor
MAE	Mean Absolute Error
MAP	Maximum Aprosterior
ML	Machine Learning
MLP	Multi-Layer Preceptor
NN	Neural Network
PADL	Possible Executed Activity
PC	Personal Computers
PRC	Precision Recall Curve
RMSE	Root Mean Square Error
ROC	Receiver Operating Charaterstic
RRSE	Root Relative Squared Error
SM	Smart Meters
SMM	Semi Markov Model

SVM	Support Vector Machines
TID	Transaction ID
TP	True Positive
TV	Television
UK-Dale	United Kingdom Domestic Appliance Level Electricity
UKERC-EDC	United Kingdom Education Research Committee Energy Data Centre
UNIX	UNiplexed Information and Computing System
Wi-Fi	Wireless Fidelity
X-Means	Extension of K means algorithm

Chapter 1

Introduction

With the emergence of new technologies in last century world has been changed drastically. The technology brought us so far that if someone who is alive 100 years ago will see todays world he might think that he is teleported to some other planet. The technology has marked its presence in every field. The same can be said true about healthcare services also. The digitalization of different fields has impacted the healthcare facilities in many ways, as a result whole experience of patients as well as professions has been changed.

The smart cities are becoming a largest infrastructure modernization project of all the times and are using digital technologies to improve the performance and enhance the productivity and well being of cities. The main objective of smart cities is to use digital technologies to optimize the usage of various resources like: electricity, water, governance, healthcare etc [1]. The smart city concept relies on various constituting technologies and devices. The smart meters with associated devices are a main element of intelligent digital infrastructure on user side [2]. Real time monitoring of energy consumption is performed by smart meters,which provides short time and long term benefits to consumers as well as to utilities.

The data recorded by smart meter can be remodeled in numerous ways to gain insights in different fields; one of them is recognition of occupant's activity patterns. In this study, the information from smart meters is utilized to grasp its usage in health care services. The routine energy utilization patterns can provide insights on how the residents adhere to their normal behavior of energy consumption and induce more information regarding their normal day to day activities.

1.1 Ubiquitous Computing

Ubiquitous Computing is defined as revolutionary computing paradigm that has changed the way of interactions between computers, physical space and devices etc. It is considered as an emerging field of information and communication technology that can be integrated into everyday objects. Ubiquitous computing is

surrounding the users with comfortable and conventional information atmosphere that combines computational and physical infrastructure into a conjoined habitat [3]. In todays world, use of modern information and communication technologies (ICT) are considered as a significant condition for economic growth and viability of future. The new techniques and technologies have great impact on public administration, science, scholarships and private life. With mobilization of digital technologies and services, they can be called from anywhere. This is one of the main motives of ubiquitous computing. Everyday objects are becoming more and more smart, hence they can be called anywhere at any time. In recent years, smart and smaller devices such as notebooks, smart phones are replacing traditional computers. Computers are getting integrated into everyday objects. The home utilities like lights, driver assistance systems in cars can be controlled with the help of PCs or using smart phones. Ubiquitous computing is one of the complementary field of virtual reality. It focuses on converting all the objects in constituting environment into components of information and communication system. In ubiquitous computing variety of devices run in background and interact with other services and devices on the behalf of users instead of simulating the everyday objects with computers. User doesn'tt need to provide the explicit instructions, the environment provided by ubiquitous computing acts as cooperative partner of humans.

1.1.1 Application areas of Ubiquitous Computing

Ubiquitous Computing aims at correlating all components of living environment and thus to allow the uninterrupted flow of data and information. Some of the areas where ubiquitous computing is already giving the stellar performance are:

- **Communication:** Communication is considered as a cross application .It is one of the initial requirement for all the other fields to exist.Ubiquitous computing has significant role in field of communication to automate it and make it intelligent and smart.
- **Logistics:** Ubiquitous Computing plays an important role in tracking of goods along the transportation chain of raw materials , finished and semi

finished products. Ubiquitous Computing aims to reduce the gap between information flow and actual flow of finished products.

- **Motor Traffic:** In Motor traffic ubiquitous computing is an emerging technique. Now a days automobiles are being embedded with assistance systems which aids the drivers by provided assistance for driving.
- **Military:** Military aims at using Ubiquitous Computing for curbing internal and external threats which are correlated and internal meshes. Military is employing ubiquitous computing for development of new weapons.
- **Smart Homes:** Large number of devices for heating, ventilation, cooling, communication and lighting are becoming smart and automatically settle themselves to the user requirement.
- **Medical Technology:** In medical technology small framed, multifunctional, networked development of healthcare platforms and utilities become possible with the aid of ubiquitous computing. Ubiquitous computing offers the monitoring of health of old age and ill people in their homes using smart and intelligent implants.
- **Internal Security:** Identification systems like smart electronic passports, smart cards are examples of ubiquitous computing. In future, they will be integrated to secure the internal infrastructure like airports and power grids.

One of the main objectives of smart home research is to ease daily life by increasing user comfort. This is achieved in two ways. One is related human activity identification and event automation in local environments. The other is remote home management from distant locations. The following smart home projects aim to automate home appliances using knowledge of human activity and behavior. Figure 1.1 shows the hierarchy of the interconnection between ubiquitous computing and smart home infrastructure.

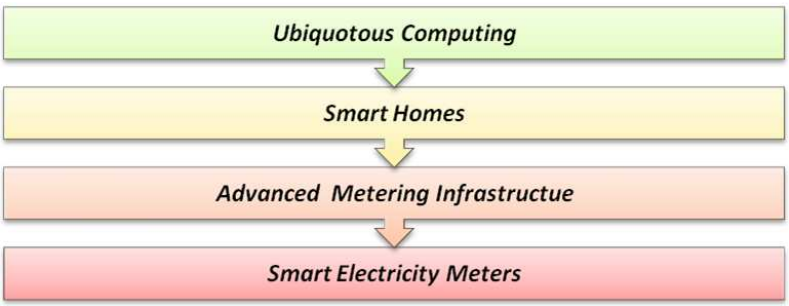

Figure 1.1: Relationship between Ubiquitous Computing and Smart Home Infrastructure

1.1.2 Smart Homes, Advanced Metering Infrastructure and smart meters

Large number of devices for heating, ventilation, cooling, communication and lighting are becoming smart and automatically settle themselves to the user requirement [2]. A smart house is associate application of ubiquitous computing in which the home atmosphere is monitored by ambient intelligence to provide context-aware services and facilitate remote home control. One of the main objectives of smart home research is to ease daily life by increasing user comfort. This is achieved in two ways. One is related human activity identification and event automation in local environments. The other is remote home management from distant locations. The following smart home projects aim to automate home appliances using knowledge of human activity and behavior. These assistive services sometimes optimize energy usage because the house is intelligent enough to reduce energy use by controlling unattended home appliances. Smart homes are also known as ubiquitous homes [3]. Ubiquitous home is defined as a house that incorporates main electrical devices and services .It allows them to be controlled accessed and monitored remotely. Smart homes take decisions based on their own context without the interference of human beings .All the computations performed in ubiquitous are invisible.

- **Advanced Metering Infrastructure:** Advanced Metering Infrastructure refers to the system of collecting, measuring, utilizing the energy consumption data and transferring the stored record to the utility companies, con-

sumers on demand or on regular intervals [3]. The system consists of hardware, management softwares and communication controls etc. The interconnection between measuring equipments and business structures permits collection and division of information to users ,suppliers, providers etc. These meters enable demand response services.

- **Smart Meters:** Smart Meters are one of the component of advanced metering infrastructure. Smart homes comprises of different types of microprocessors, embedded systems and utilization meters etc. Smart meters are significant component which are employed in smart homes to monitor and store records of electricity utilization in smart homes. It is an improved and enhanced version of traditional electricity and gas supply meters. They are being installed in homes so as to keep an eye on energy consumption data in association with time and amount of money spent on it. It measures the data in hourly and more frequent manner. It allows two way communications between consumers and suppliers.

1.2 Knowledge Discovery

Data Mining is also considered as synonym of Knowledge Discovery Process, but in reality it is a significant step of Knowledge Discovery Process. Data mining is the technique of pattern discovery in massive data sets .It involves strategies which constitutes methods at the intersection of machine learning, statistics, and information systems.

Data Mining is technique for recognition of interesting information such as associations among data, patterns, outliers , important structures from large amount of data stored in data warehouses or other sources of information. Knowledge Discovery Process consists of following different steps:

- **Data Cleaning:** Data Cleaning involves handling of missing, irrelevant, noisy data.
- **Data Integration:** Data integration involves integration of multiple heterogenous sources of data into a single data warehouse.

- **Data Selection:** In this step of knowledge discovery, Data which is relevant to the analysis is considered for further processing.
- **Data Mining:** This is a process in which intelligent techniques are applied to extract patterns from raw data.
- **Pattern Evaluation:** In Pattern Evaluation, extracted patterns are evaluated to trace out some useful information from them.
- **Knowledge representation:** In knowledge evaluation process, evaluated patterns are presented to the user in visualized forms.

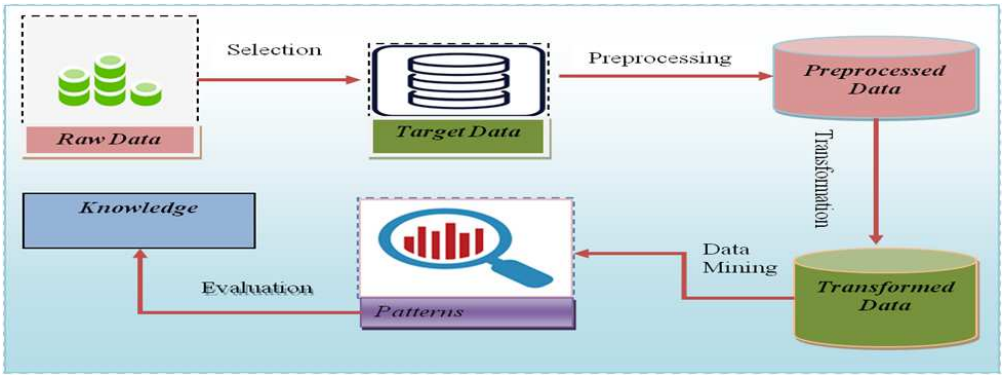

Figure 1.2: Knowledge Discovery

Figure 1.2 represent major processes of knowledge discovery. Further the analyses of the data mining techniques has been discussed.

1.2.1 Major Processes of Data Mining

Data Mining tasks are of two types descriptive and predictive .They are described concisely in following section:

- **Descriptive Techniques:** Descriptive Techniques describe the database in concise and summarized manner. Some of the examples of techniques are: sum, average, count etc.
- **Predictive Techniques:** In Predictive techniques models are constructed, inferences are derived from data, predictions of behavior from new dataset are attempted.

Some of the most significant data mining techniques are summarized as follows:

- **Class Descriptions:** Class Description provides concise summery of data collection which makes it distinguishable from other classes. Class Discovery should not cover only techniques like count, sum, average but also properties like data description properties such as variance, quartile etc.

- **Associations:** Association Mining deals with discovery of associations and correlations among distinct set of items .They are expressed in the rule based form which shows attribute value, condition that frequently appear together in given dataset.An association rule in the form of AB implies those data tuples who satisfy A are likely to satisfy B. Associations are widely used in directed Marketing, business Decision Process etc. In previous years, much research focused on Association Analysis with new efficient techniques consisting of Appriori Search, Mining Multiple levels, Multi Dimensional associations are performed.

- **Classification:** Classification is a technique which takes set of training data and constructs a learning model for each class based on features of data. Set of classification rules is generated with this technique .The rule set can be further utilized to for getting better understanding of classes in databases and for categorization of feature data. For example prediction of disease from set of symptoms is classification. There have been many classification techniques developed in the field of statistics, databases, machine learning ,rough sets ,neural nets. It has wide range of applications in the field of business management, credit analysis and customer segmentation etc.

- **Prediction:** The prediction function predicts possible outcomes of some classes or distribution of certain values of attributes in set of objects. It finds attributes relevant to attributes of interest and forecast the value using set of data similar to selected objects. Prediction of salary of employ based on his qualification and with respect to the his colleagues salary is example of predictions. Some of the prediction techniques are regression analysis, decision trees etc. Genetic algorithm and neural nets are also popularly used for predictions.

- **Clustering:** Clustering is a technique of identifying clusters of related data,

cluster stands for collection of data points which have similarities to each other. Similarities can be figured out with the aid of distance functions, expert specified methods etc. For example: clustering of houses on the basis of categories of floor area, geographical location etc.

- **Time Series Analysis:** This approach implies the analysis of large sets of time series data to trace out certain characteristics and regularities. Some of the interesting characteristics are searching of related sequences, trends and deviations in data, sequential pattern mining etc.Example of time series analysis is prediction of trends of data.

- **Frequent Pattern Mining:** Frequent Patterns are itemsets or substructures that appear in dataset with frequency more than threshold frequency specified by user [4]. Sequences whose frequency is more than or equal to minimum threshold value are coined as frequent patterns. For example buying camera is followed by purchase of PC and if it occurs repeatedly then it is known as frequent pattern. For pattern mining, different techniques are applied to find candidates after that frequent patterns are generated by utilizing Basic techniques of frequent pattern mining:Apriori, FP growth and Eclat are considered as basic pattern mining approaches. They are explained as follows:

  - **Apriori:** In apriori ,an itemset is known as frequent if all of its sub itemsets are also frequent.Frequent itemsets are extracted using hieratical methods of patterns mining. First of all, one itemset is mined ,next using one itemset itemsets of size two are mined [5]. Similarly k itemsets are mined using k-1 patterns.Approirialgorithm is not considered as the best algorithm as dataset is needed to be revisited repeatedly. To overcome the drawbacks of Appriori different improved techniques are proposed by different researchers like: Portioning Technique, Hashing Technique, Sampling Approach, Dynamic Itemset Counting, Integrity Mining etc. These techniques result in efficient ways of pattern mining which reduced the no. of database scans.

  - **FP- growth:** FP growth is another pattern identification algorithm

which surpasses the problems of apriori algorithm. Following three features make FP growth better then Apriori.

* It uses divide and conquer technique in which large dataset in divided into smaller problems which decrease the size of search space.
* It does not uses complex candidate generation technique for large no. of candidate itemsets.
* Databases are converted into smaller data structures known as FP trees, due to which repeated scans of database which are costly are avoided.

Appriori consists of two sub processes:

* FP-Tree construction.
* Generation of frequent patterns based on FP-Tree.

– **Equivalence Class Transformation (ECLAT):** Both the Apriori and FP-growth strategies mine frequent patterns from a group of transactions in horizontal data formatting (i.e.,), wherever TID may be a transaction-id and itemset is the set of things bought in transaction TID. as an alternative, mining can even be performed with knowledge conferred in vertical data format i.e. Equivalence class Transformation (Eclat) algorithmic [6] rule performs pattern mining by exploring the vertical formatting. the primary scan of the info builds the TID set of every single item. beginning with one item (k = 1), the frequent (k + 1) itemsets grown from a previous k itemset will be generated in line with the Apriori property, with a depth-first computation order like FP-growth. The computation is completed by an intersection of the TID sets of the frequent k itemsets to calculate the TID sets of the corresponding (k+1) itemsets. This method repeats till no frequent itemsets or no candidate itemsets will be found.Besides taking advantage of the Apriori property within the generation of candidate (k + 1) itemset from frequent k itemsets, another benefit of this methodology is that there's no ought to scan the information to seek out the support

of (k + 1) itemsets (for k 1).

1.3 Machine Learning

In recent years , there is much discussion about machine intelligence and what are its correlations with health, well being and productivity. Machine learning appeared to be as potential transformative technique , which tackle global issues like changes of global environment, saving lives, addition of trillions of dollars to world economy by increasing productivity. Machine leaning has brought both opportunities as well as challenges to world as it has changed the fundamental nature of work and the choices people make in day to day life. The risks and benefits of machine learning are needed to be navigated as it is becoming central to day to day activities. Machine learning is a technology which allows machines

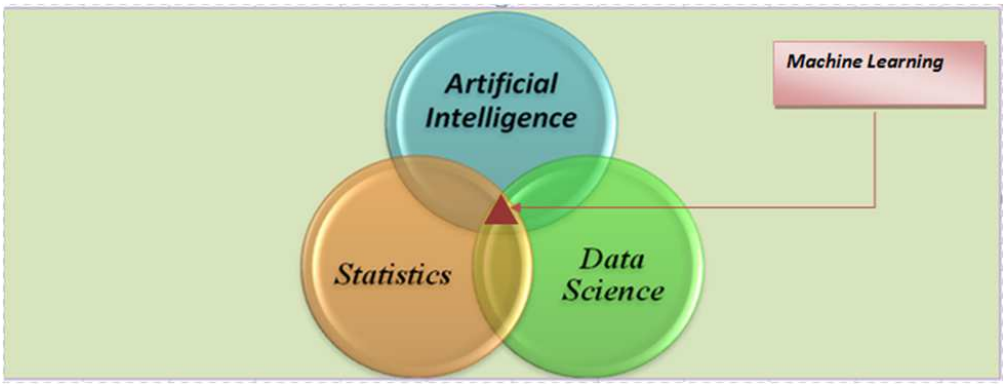

Figure 1.3: Machine Learning

to learn from data, experiences and examples. It permits the machines to perform tasks intelligently by learning from past examples. System can carry out complex tasks with the aid of machine learning rather then following pre programmed instructions. In day to day life people interact with machine learning system very frequently. Photo tagging feature of social media sites, voice recognition systems all are examples of machine learning driven systems. Machine learning holds significant potential in many other fields like education, healthcare etc. In healthcare system machine learning aid in providing better healthcare utilities to people.

Machine learning is constituting mixture of Artificial Intelligence, Statistics, Data Science Figure 1.3. It learns from these fields in such a way that can aid it in

learning from patterns, make predictions or make decisions.

1.3.1 Branches of Machine learning

There are three core branches of machine learning:

- **Supervised Machine Learning:** In supervised techniques system is trained on data which has pre existing labels. The labels classify the data into one or more categories. The system learns from training data about the structure of categories and uses this to categories the training data.
- **Unsupervised machine Learning:** In unsupervised learning labels are not present. It learns to detect the characteristics which makes the points more or less similar to each other .The technique for supervised learning is like construction of clusters and assigning data point to clusters.
- **Reinforcement Learning:** It focuses on learning from existing experiences. Reinforcement learning sits between supervised and unsupervised learning.In reinforcement learning an agent interacts with its environment,and provided with reward function which it tries to optimize.For example,the system might get rewarded for winning the game.The aim of the agent is to learn the consequences of decisions,like which moves are of much importance in game and further use those learning for maximizing its rewards.

On the basis of deployment machine learning systems are further classified as offline and online systems.

- **Offline Learning Systems:** These systems are trained and tested in offline mode. The trained models are then frozen before deploy them in online setting. Any other vocational training is also provided in offline setting, then tested and deployed in online setting using conventional software management systems. These techniques are more prevalent in current machine learning systems. In current scenarios these systems are preferred as they gave opportunity to validate the system performance before its deployment in real scenario.
- **Online Learning Systems:** These systems are also developed in offline

setting .But key different is related to the training model. The model keep on training itself in real life scenarios after it is being deployed. It means system performance keep on improving.

1.3.2 Canonical Problems in Machine learning

Machine learning allows the analysis of data to find out the patterns, and further make predictions on the basis of those detected patterns. The fundamental problems that machine learning seeks to resolve are summarized below:

- **Classification:** Classification deals with problem of categorization of data. On the basis of training data it assigns classes to the testing data. Some of the examples of classification data are: In medical diagnosis for detection of type of decease on the basis of pre existing classes, In banking sector to determine whether the transaction is fraudulent or not?, In computer vision to detect the type of object in picture whether it is object or human. Methods for such tasks include: Logistic Regression, Random Forests, Support Vector Machines ,Gaussian Process Classifiers etc.

- **Regression:** Regression analysis tries to predict continuous quantities from input data. Its applications include financial prognostication, and click rate prediction, which has a variety of applications in web advertising. Typical strategies to handle this task embody Linear Regression, Neural Networks, and Gaussian Processes.

- **Clustering:** Clustering decides which data points have similarity to each other. Examples of problem which are solved using clustering are: In e-commerce, customers showing same behavior are grouped together. In video streaming same genre videos are catalogued in one genre. Main techniques consists of k-means, Gaussian mixtures and Dirichlet process mixtures.

- **Dimensionality reduction:** Dimensionality reduction is utilized to decide which are most significant features of data and how they can be summarized. In E commerce what combination of feature allows to conclude the behavior of customer. Typical techniques include: Isomap, Gaussian Process Latent Variable Models, Principal Components Analysis, Multidimensional Scaling,

Factor Analysis.

1.4 Anomaly Detection

In data mining, anomaly detection refers to the problem of finding patterns in data which do not conform to an expected pattern or other items in a dataset[7]. Typically the anomalous items will translate to some kind of problem such as medical problems or errors in a text. Figure 1.4 shows anomalies in a simple two-dimensional data set. In data G1 and G2, are two normal regions, since most observations lie in these two regions. Points o1 and o2 are anomalies, because they are far away from the normal regions G1 and G2.Anomalies might influence the data for a variety of reasons, such as malicious activities, for example, credit card fraud, cyber-intrusion, terrorist activity or breakdown of a system. Due to critical effects of anomalies in data, anomaly detection is very important. Anomaly detection is a technique used to identify unusual patterns which do not resemble to the normal behavior.

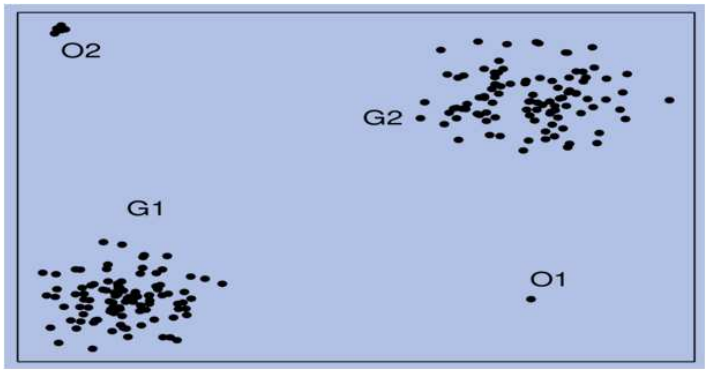

Figure 1.4: Anomalies in Normal Data

1.4.1 Categories of anomaly Detection

Anomaly detection is divided into basically three types:

- **Unsupervised anomaly detection** techniques detect anomalies in an unlabeled test data set under the assumption that the majority of the instances in the data set are normal by looking for instances that seem to fit least to the remainder of the data set.

- **Supervised anomaly detection** techniques require a data set that has been labeled as "normal" and "abnormal" and involves training a classifier (the key difference to many other problems is the inherent unbalanced nature of outlier detection).
- **Semi-supervised anomaly detection** techniques construct a model representing normal behavior from a given normal training data set, and then testing the likelihood of a test instance to be generated by the learnt model.

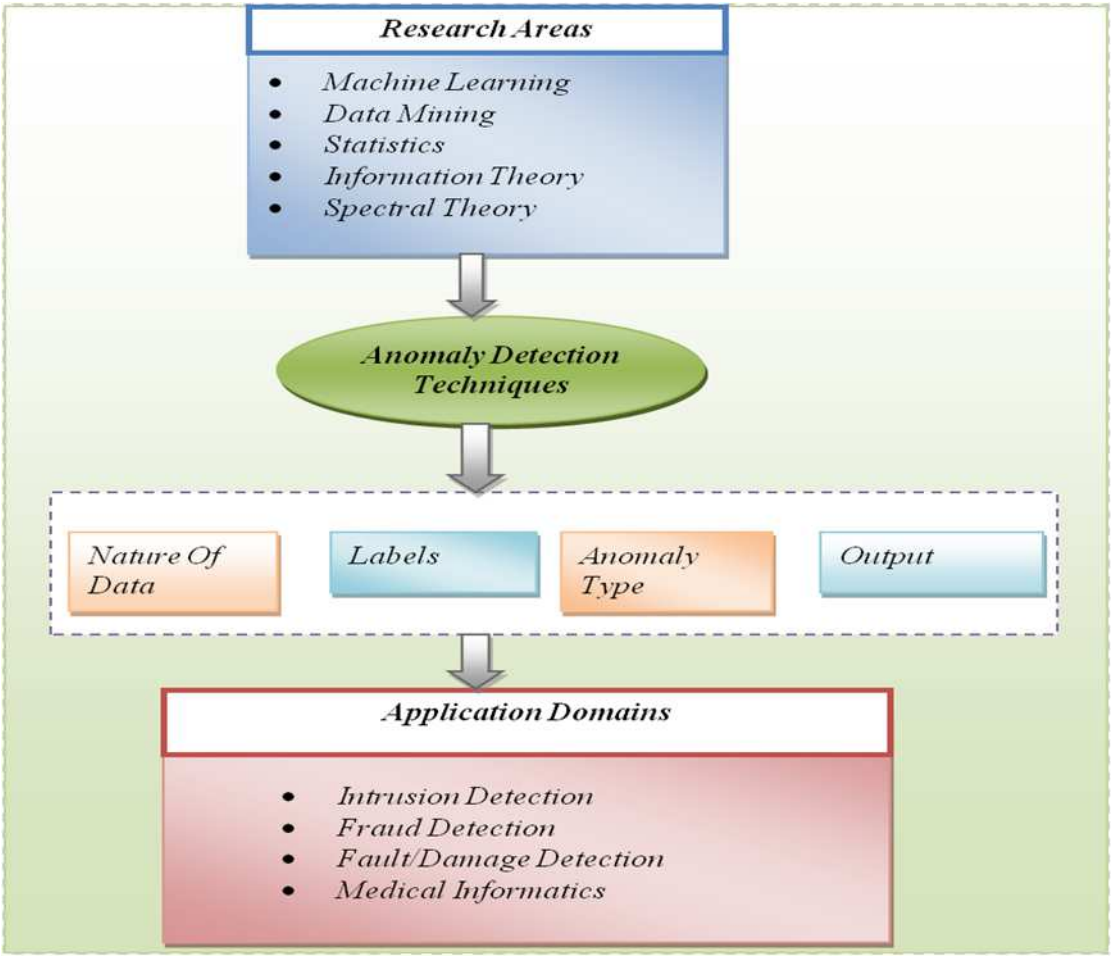

Figure 1.5: Key Components in Anomaly Detection Techniques

In Figure1.5 Key components in anomaly detection techniques are discussed. Anomaly detection techniques use techniques related to machine learning, data mining, statistics etc. Medical informatics is one of the application domains where anomaly detection techniques fits well. For applying anomaly detection techniques the problem characteristics like nature of data, Labels of data, what is the type of anomaly, what is the desired output is required to be clearly mentioned.

1.5 Thesis Organization

The thesis structure is divided into 6 chapters. A brief structure of thesis organization is presented in following section:

- **Chapter 1**: This chapter presents the brief overview of thesis structure and theoretical background of different techniques. It introduces the concept of ubiquitous computing which is the umbrella branch of smart cities and smart metering infrastructure. Next section provides the brief introduction of knowledge discovery, data mining and machine learning. Further, anomaly detection techniques are introduced in brief.

- **Chapter 2**: Chapter 2 covers the review of existing literature in the field of healthcare monitoring, smart metering infrastructure, utilization of techniques related to computer science in health monitoring and energy consumption.

- **Chapter 3**: Chapter 3 formulates the problem statement which is figured out from the literature review done in Chapter 2. Objectives for the proposed study are formulated by keeping research gaps in mind.

- **Chapter 4**: This chapter presents the proposed methodology to attain the objectives discussed in Chapter 3. Theoretical background and implementation sequence of proposed techniques is presented in this chapter. Algorithms and mathematical background of proposed techniques is briefly summarized in this chapter.

- **Chapter 5**: Chapter 5 presents the results. Results for one and all steps proposed to fulfill the objectives of proposed study are recorded in this chapter. Implementation results are filed in the same sequence as they are in proposed methodology.

- **Chapter 6**: Chapter 6 presents the summary of the overall results of proposed study. Results are evaluated on the basis of proposed objectives. Possibilities of extending the proposed methodology in future are also outlined in this chapter.

Chapter 2

Literature Review

Chapter 2 covers review of several existing studies in the field of smart meters, healthcare in association with ubiquitous computing techniques and anomaly detection in the field of computer science. With revolutionary changes in the field of computer science, techniques of interactions between humans and computers are changing for betterment and progress of society. Ubiquitous computing has provided a whole new definition to correlation between humans and computers. Involvement of computers in human life has totally changed its meanings with ubiquitous computing. Ubiquitous computing aims at involving computers in all the aspects of life in the ways which are invisible to humans. Main aim of involving ubiquitous computing in context of proposed stud is to correlate the energy consumption patterns with inhabitants health to assist the healthcare model.

2.1 Existing literature in different subfield with respect to proposed study

With emergence of smart technologies, every field is getting digitalized. As a result monitoring healthcare using Smart meter data conjointly become prevalent. The aim of the study is to analyze the inhabitants activity patterns to understand normal behavior and to detect the anomalous activities, which directly indicate health problems. In this section review of existing work is presented, which reflects the status of existing work done in this field till date.

Detection of human activities in smart homes is basically done using sensors and smart meters. In [8] smart meters are used to analyze the human activities. The paper proposed two techniques one technique is Semi-Markov-Model (SMM) which detects human actions and second technique uses impulse oriented technique to find out activities in daily living which focuses on simultaneously occurring activities. In same way [9] works on activity detection of elderly citizens by categorization of the sensors related to daily activities. Smart meter energy data is utilized in [10] for activity detection using Dumpster -Shafer theory of evidence and Non-

intrusive Appliances load monitoring technique. This study uses processed input data and then machine learning based algorithms are applied to separate the major activities inside the house with the aid of aggregated energy data. Other studies [11]-[17] utilized smart energy utilization data. They are utilizing sensor based data and Internet of things based infrastructure to develop applications which monitor and provide health care services to the residents of smart cities. In [12] demonstration for observing energy utilization of equipments is presented. [13] uses Bayesian network based approach to forecast inhabitants behavior. In [14] multi label classifiers based on time series are used to predict device utilization by using method of decision tree. In [15] clustering approach is used to find out the distribution of consumers consumption sequences .In this study no device level usage is considered. Work in [16] uses hierarchical and c means clustering to predict appliance usage patterns using ON and OFF status of devices. Study in [14] presented a graph based model algorithm to forecast interdependency of devices and human behavior. Further it uses these relations to predict different device usage with the aid of Bayesian prediction model. Previous models have not approached the problem of human behavior anomaly detection with the aid of smart meter energy consumption data.

Table 2.1: Existing Methods for Energy Management

Author	**Problem Statement**	**Research Method**	**Contribution**	**Key Findings**
Paula Carroll et al.[17]	Study conducted in response to a national rollout of smart electricity metering in Ireland planned by government.	Machine learning methods ,Neural Networks and Elastic Net Logistic regression	Usage of smart meter data as a potential new data source for better analysis of energy consumption is proposed .	• Models are useful in identifying energy usage patterns for houses of single occupancy. • Performance of the model worsens as the number of persons in a household increases.

to be cont'd on next page

Table 2.1: Existing Methods for Energy Management (Cont.)

Author	**Problem Statement**	**Research Method**	**Contribution**	**Key Findings**
Shailendra Singh et al.[21]	To understand the interdependencies among different device utilization within a house where multiple concurrent devices are operating.	Incremental frequent pattern mining, Statistical methods	Figured out relationship between association rules and device usage behavior of humans to forecast energy consumption.	• Occupants energy usage behavior is directly related to association of devices operating concurrently.
Mi Zhang et al.[28]	To detect most important features for human activities recognition using future importance techniques.	Statistical techniques, Classification technique of machine learning .	Figured out the Impact of the physical features on the performance of the recognition system, a single-layer feature selection framework is developed for classification.	Accuracy for recognition is improved to 90%, which is 8% better in comparison to previous analysis when only statistical features are used. performance is further improved by 3.8% by extending the single-layer framework to a multi-layer framework of classification.
Ngoc Cuong Truong et al.[20]	To forecast the usage of multiple electrical appliances by domestic users,with the aim of providing suggestions about the best time to run appliances in order to reduce carbon emissions and save money .	Graphical model based algorithm ,clustering techniques of machine learning	extensively evaluates the proposed methods on real-world data .The methods performs 47% more efficiently that existing techniques.	• All the algorithms suffer from uncertainty within the labeling process of home owners as the labels are not provided in efficient way.•limited training data is another drawback of this model.

to be cont'd on next page

Table 2.1: Existing Methods for Energy Management (Cont.)

Author	**Problem Statement**	**Research Method**	**Contribution**	**Key Findings**
Krzysztof Gajowniczek et al.[16]	Conducted for discovery of the sequence of home appliances usage patterns.	Unsupervised machine learning techniques.	Helped to examine the interdependencies between the usage patterns of home appliances and drive important associations between several related factors including time of the usage of devices and user activities.	• Determination of the household characteristics from smart meter data is feasible.• aids for quickly grasping general trends in data.
Kaustav Basu et al.[12]	To forecast whether the particular appliance will start working in a given hour or not.	Three machine learning techniques Decision Trees, C4.5, Bayes Net were employed. Oracle data storage is used.	• Concluded the best historical data in last 24 h is more relevant for predictions.• Prediction methodology gives the best results when considering large training data from oracle database .	The system require historical data from last 24 hours in oracle database for the efficient prediction of appliance operations in upcoming next hour
Abdulsalam Yassine et al.[18]	To address the need to analyze energy consumption patterns at the appliance level, which are directly correlated to human activities.	incremental frequent mining of device to device associations , prediction model based on Bayesian prediction	Presented a model to correctly infer multiple appliance usage concurrently and make short and long term prediction of device utilization with high accuracy.	Model is based on individual and simultaneous appliance usage in home environment.

to be cont'd on next page

Table 2.1: Existing Methods for Energy Management (Cont.)

Author	**Problem Statement**	**Research Method**	**Contribution**	**Key Findings**
Thomas et al. [13]	To propose a network model for health monitoring of elderly people in their homes , the focus of this study is on completely passive sensing.	Microprocessor board, wireless add on are used.	leveraging of smart grid technology based on network architecture for home-health monitoring.	• The model is completely passive which do not compromise any aspect of privacy of occupants.• Projected work is in its preliminary stage. 3. Testing stage of the work is successful.
Alam et al. [22]	To analyze the data to detect energy usage anomalies related to the behavioral abnormality of the residents.	Hierarchical probabilistic model-based group anomaly detection technique	Detects routined appliances usage patterns from smart meter and smart plugs in regular days and then learns the unique time segment group of each appliance's energy consumption.	• Activity logs are analyzed on daily basis using proposed model.• Proposed model detect abnormalities on the basis of static training data.
Shamim M Hossain et al. [43]	To propose a recognition system based on tasks performed by patients of some particular disease for the healthcare framework	Fourier Transformation ,Grey Level Conversion, log-likelihood score	The system analyzes the patient status taking two main types of input, video and audio which are captured in a multi-sensory environment.	• Approach is based on speech and video inputs.• Data used is real.100 people are recruited for collection of facial express data collection.

to be cont'd on next page

Table 2.1: Existing Methods for Energy Management (Cont.)

Author	**Problem Statement**	**Research Method**	**Contribution**	**Key Findings**
Charl Chelmers et al. [11]	To recognize the sudden changes in the behavior of patients suffering from Alzheimer's, Parkinson's disease and clinical depression	Neural Network based approach(Random Neural Net Classifier	Presents an approach for unobtrusively monitoring of people. Data classification techniques are employed to detect anomalies in behavior of people.	• System is capable of tracking sudden changes •Networks are better as compared to other classifiers at detecting changes in patient behavior
Jana Clement et al. [8]	To propose methods that can be used to monitor human behavior in single apartments.	Semi-Markov-Model (SMM), impulse based method	Most possible executed activity (PADL) will be calculated to allow an evaluation of the currently executed activity (ADL) of the inhabitant	Basic technique used is based on peak power consumption by device.
Alam et al.[24]	To present the overview of existing smart home research as well as of correlated technologies.	Algorithms from fields like Machine Learning , Self adaptive algorithms, access protocols are used.	Identifies the future direction of smart home research. As the centre of intelligent service consumption smart meters are going to be revolutionary field.	Distributed computing, middlewares should be used in the fields to increase the efficiency of existing work.
Charl Chalmer et al.[25]	To predict the activities based on energy consumption trends.	Perception classifiers ,data transfer over wide area network	Model figures out that reoccurring patterns of energy consumption indicates the same time of some particular activity.	Energy consumption patterns forecast many things like if low consumption of energy then no one is present at home.

to be cont'd on next page

Table 2.1: Existing Methods for Energy Management (Cont.)

Author	**Problem Statement**	**Research Method**	**Contribution**	**Key Findings**
Muhammad Fahim et al.[26]	To analyse the concurrent situation assessment and major dominating activities over minor activities.	Evolutionary ensemble approach based on GA .	Detects minor activities independent of major activities	Sensory data is prepared to extract feature vectors. These vectors are analyses using ensemble based genetic algorithm
Damminda Alahakoon et al.[27]	To extract various types of values from smart meter data for various sub processes like data acquisition, transmission, interpretation	Self organizing Maps, SVM, Fuzzy Logic Profiling etc.	Provides insights on opportunities as well as on challenges arising due to big data and cloud environment.	Shows possibilities of correlation between smart metering with big data, cloud computing ,internet of things.
Guanchen Zhang et al.[28]	To present the energy disaggregation algorithm based on hourly smart meter readings.	Clustering and Optimization Techniques	Outputs of proposed study presents the breakdown of energy into different load categories based on components having different power factors.	Approach can be applied to different houses based on random seasons to find out the load of each device
Henry Friday Nweke et al. [29]	To propose the combined functioning of different types of sensors to provide general framework for Ambient Assistant Living ,Activities of daily life.	Deep learning based classifiers	Data fusion Modelities have been implied Deep learning based human activity recognition model is developed.	Presents different classifier systems to aid AAl ,ADL.

This chapter presents literature review of the various fields to find out the current

status of related work .Several existing studies in fields of smart infrastructure, pattern recognition, machine learning in association with activity prediction,energy consumption and anomaly detection techniques are explored. From literature review the conclusion is drawn that although large amount of studies are conducted in different fields but very few authors have touched all the different aspects of ubiquitous computing to provide a single solution for home based healthcare system.Hense,there is a dire need of single model which will utilize the smart meter data from smart homes and provides healthcare assistance to inhabitants by automatically detecting anomalous behavior of people in dwellings.

Chapter 3

Problem Statement

Good Healthcare System is a fundamental necessity of todays social infrastructure. With growing population, providing good healthcare services to the people is one of the great challenge for governments. Hence some alternative approach for healthcare service is a big requirement in todays society. This chapter provides the concise description of different research issues present in existing literature. In this chapter current research gaps are identified that needs to be addressed. After thoroughly analyzing the existing issues, problem statement, research directions are formulated. Objectives for the proposed study are also framed which are presented in subsequent sections.

3.1 Problem Definition

With emergence of smart technologies, every field is getting digitalized. As a result monitoring healthcare with energy data analytics is becoming an emerging technique . The aim of the study is to analyze the inhabitants abnormal activity patterns , which directly indicates health issues in particular home.This study proposes a methodology that is capable of analyzing the readings of smart meter to recognize activity patterns and reveal the changes in electricity consumption patterns to indicate the diversion of normal behavior of individual to abnormal.

3.2 Research Gaps

This section presents the research gaps found in the existing literature .The research gaps which are identified during literature survey are given as follows:

- Presently, activities of daily living (ADL) are recognized using sensors, wearable bands and CCTV cameras etc [22]. Sensors have the high price associated with them, because of that, it's impracticable to extend the utilization of sensors for activity recognition on the massive scale.

- Wearable bands and CCTVs have an effect on users privacy, because of that many people deny to adopt them in their day to day life. In such a scenario, use of some passive approach like the smart meter data for activity recognition appears a decent plan.
- Techniques of anomaly detection are not applied for behaviour abnormality detection.
- Existing approaches did not consider any time frame that leads to efficient recognition of activities.

Although data analytics techniques used are not new ones but the combination of techniques to fulfill the objectives in proposed methodology is unique to the pre existing studies.

3.3 Objectives

Following are some research objectives which are formulated to fulfill the research gaps identified in literature survey of existing studies:

- To study and analyze existing literature and techniques that has been presented within the field of activity recognition, anomaly detection.
- To mine the patterns of simultaneous device operations using frequent mining techniques.
- To utilize energy consumption records of simultaneous appliance operation (frequent patterns of home appliances usage) for activity recognition.
- To discover the anomalies in daily life activities of inhabitants on the basis of past energy consumption data.

Chapter 4

Research Methodology

The Chapter provides the detailed summary of methodology employed in the study. Figure 4.1 illustrates the proposed model with its different phases like preparation of the data, pattern extraction, classification of patterns into manually annotated classes, Mood Profiling and Anomaly Recognition. This chapter provides discription of phases and details of respective mechanisms is provided along with related theoretical background.

The raw data that consists of numerous records of energy time series consump-

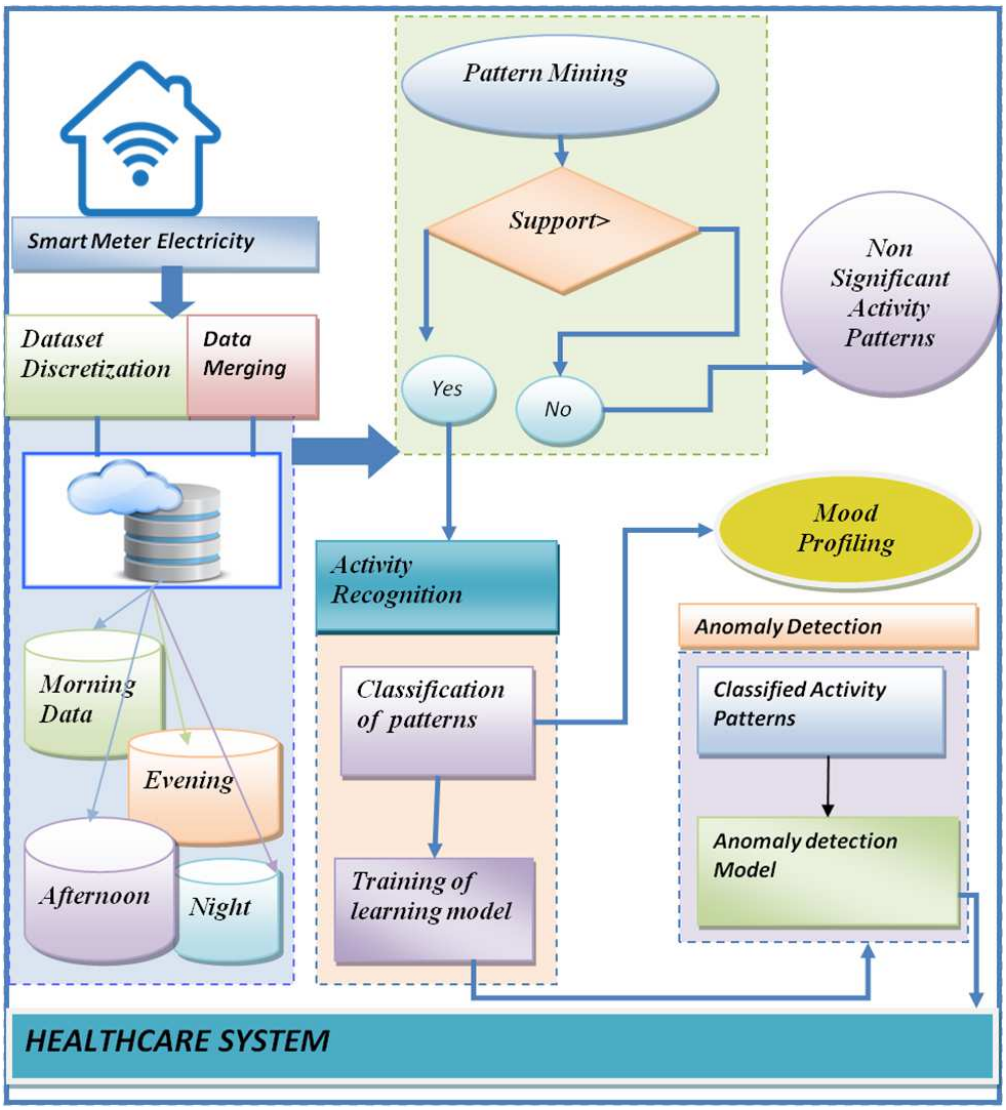

Figure 4.1: Conceptual Model of proposed methodology

tion data are processed and prepared for further analysis. Disaggregated energy consumption data from UK Dale [30] is used. Further, frequent pattern mining, multiclass classification is performed. In initial phase, Data preparation and preprocessing techniques are applied on each data file .Data discritization, data merging techniques are used for preprocessing. Preprocessed files are joined to make an intermediate database. The intermediate database is further divided into 4 distinct databases meant for morning, afternoon, evening and night energy utilization records. Frequent patterns points out repeated patterns of devices which frequently appear together in dataset. If sequence of devices for example bread maker and kettle often appears together, then they are considered as frequent patterns. The aim of this study is to unveil associations of various devices in conjunction with usage time(morning, evening, afternoon, night). For making the system more efficient progressive incremental data mining approaches are applied. As mentioned earlier energy utilization data is divided into four distinct databases according to time slots of morning, afternoon, evening, night. This is done to reduce the system overhead and for rapid detection of problems. System has to traverse lesser amount of data when it is stored in separate files.

The datasets are mined to extract the frequent patterns of energy utilization by different devices. If support value of the extracted patterns is more than threshold value then pattern is considered as normal. Similarly all the 4 distinct databases are mined to extract frequent patterns associated with time. Multiclass classification is performed to categories the frequent patterns into manually labeled multiple categories. Different machine learning prediction models are trained on data to predict classes.

Mood profiling is one of the objective of this proposed work. In Mood Profiling multiclass classified results are utilized to correlate persons mood with day to day activities. Final phase of proposed work is anomaly recognition. Aim of this section is to find out those abnormal patterns which are deviating from normal activity patterns. Customized model based on clustering and machine learning techniques is proposed. Anomaly Detection model is based on X-Means clustering which is extension of K-means clustering technique. It makes number of clusters automatically, which differentiates it from K-means clustering where number of

clusters are needed to be specified as input to algorithm. Local Outlier Factor (LOF) detects outliers by calculating the deviation of given data from neighbors. Anomaly Score is assigned to each transaction by local outlier factor method. Prediction model is applied to predict the anomalies for whole database. Next filtration method is applied to filter out those transactions whose anomaly score is more then 1.5. The patterns having anomaly score higher than 1.5 are considered as anomalies. They are filtered out and sent to healthcare system for further actions. Next section provides the brief outline of the methods, models, evaluation parameters used in the proposed work.

4.1 Data Preparation

The original dataset stores the readings from 53 different devices installed in a single house. Energy consumption records of each device are stored in separate files. The format of raw data files is as given below:

$$\langle \textit{Unix timestemp, Energy Consumption} \rangle$$

- **Conversion of Unix Timestamp into Date and time:** As shown in table 1 row data is in form of ⟨ UNIX timestamp, Energy consumption ⟩. UNIX timestamp is not easily understandable .Hence it is converted into date and time format.
- **Data Discretization:** Data discretization is data preprocessing technique. In data discretization large number of values are converted in smaller ones so that data evaluation and processing become easy. There are different techniques of data discretization .Equal width portioning is used in this study. Initial electricity consumption data is in redundant form. It is converted into 1 minute time gap with the aid of equal width partitioning.
- **Data Merging:** Data merging implies integration of different data files into single combined database. Many-to-one merge approach is used .This step is significant as single data files does not have any importance in proposed model. The intermediate database which is unified and has no redundancies

is created.

- **Division of data into different time slots:** Data is divided into 4 different time slots of morning, afternoon, evening, night. To make the pattern discovery process more efficient and making the proposed model rapid at tracking abnormalities, data divisions are done.

After conversion into date and time, data discretization and data merging an intermediate database is created the intermediate database format is as follows:

⟨ ***Date And time, energy consumption by device 1, device 2,...,..., device 53*** ⟩

Figure 4.2 is visualization of energy consumption in association with time. In Figure 2 X-axis represent the time of day when device is operating while y axis represent the amount of energy consumed by each device. Intermediate database

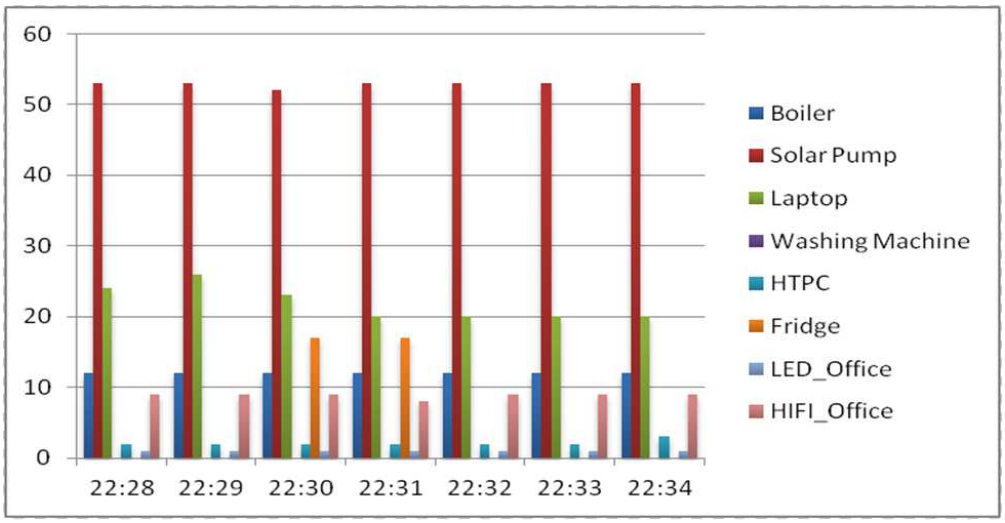

Figure 4.2: Association of energy consumption data with time in initial dataset

is further processed to convert it into format which records the data in table according to operating status of various appliances. The format ready to mine dataset is as follows:

⟨ ***Date and time, device1, device 2,...,...,..., device 53*** ⟩

4.2 Extracting Frequent Patterns of Activities of Daily Living

One of the significant step of the proposed work is discovery of hidden patterns of human activities from smart meter electricity utilization data.Preparing food, Studying, Using Computer, Watching TV, listening to radio are daily living activities. Our objective is to trace sequences of activities so that sudden change or imbalance in these activities can be monitored as earlier as possible and take appropriate actions. For frequent pattern mining all the devices which are active in source database are considered and utilized to mine the activity patterns. Pattern discovery is more related to discovery of association between multiple devices operating together like listening to radio while preparing food, washing clothes while watching television. Patterns mining is done in different time slots on distinct databases.[4] and [32] proposed FP growth using divide and conquer method. This approach works better than other pattern mining techniques.

The patterns are stored in different databases to accurately detect the patterns of activities performed in morning, afternoon, evening, night. Frequent patterns are sequences that are present in data more than the threshold value. Sequences whose frequency is more than or equal to minimum threshold value are coined as frequent patterns. For pattern mining, different techniques are applied to find candidates after that frequent patterns are generated by utilizing candidates. Apriori is popular algorithm used for association rule mining. Apriori algorithm uses candidate generation criteria. It makes largest candidate at level n to further form candidates at n+1 level. It scans database multiple times till largest frequent itemset is formed. The main drawback of apriori is traversing of database. As database is needed to be traversed many times, hence apriori is not an efficient technique to be implemented in those cases when size of datasets is large.

FP growth is another pattern identification algorithm which surpasses the problems of Apriori Algorithm. Following three features make FP growth better then Apriori:

- It uses divide and conquer technique in which large dataset in divided into

smaller problems which decrease the size of search space.

- It does not uses complex candidate generation technique for large no. of candidate itemsets.
- Databases are converted into smaller data structures known as FP trees, due to which repeated scans of database which are costly are avoided[6].

It consists of two sub processes:

- FP-Tree construction.
- Generation of frequent patterns based on FP-Tree.

Algorithm 4.1 Steps of construction of the FP-Tree are as follows

Method (divide-and-conquer)
•For one and all items, construct its conditional pattern-base, so its conditional FP-tree.
•Repeat the method on every recently created conditional FP-tree.
•till the resulting FP-tree is empty, or it contains just one path (single path can generate all the combos of its sub-paths, each of that may be a frequent pattern)
Step 1: Conditional Pattern Base construction
•Begin from the header table of frequent items within the FP-tree.
•By following the link of one and all frequent items traverse the complete Frequent pattern tree.
•All of remodeled prefix ways of that item are accumulated to create a conditional pattern base.
Step 2: Conditional FP-tree construction
•Begin from the tail of the header list
•For one and all pattern-base
Collect the count for each item in the base
For frequent items of the pattern base construct a FP tree

4.3 Multiclass Classification for activity representation

In this section, Frequent Patterns are classified into manually labeled categories with the aid of machine learning prediction models. As delineated in previous sections, information is distributed into different parts in line with time frames (morning, afternoon, evening, night). Each database is mined individually, therefore the patterns in every time slot are distinct. Multiclass classifications algorithms are applied on frequent patterns so as to map the device operative status with manu-

ally annotated classes. An example of manual classes is given as if kitchen lights, breadmaker, coffee maker are ON, it signifies that person is preparing food. Similarly, different patterns are manually labeled for classification. Different classes are allotted to patterns on basis of normal activities of daily living (ADL)[33] like bathing, showering, cleaning, maintenance of house, preparing meals etc.
Mostly binary category classification algorithms are tuned to perform the multiclass classification techniques. Existing strategies to handle multiclass classification problems are: Transformation to binary, Extension from binary, hierarchical classification.
In the projected model, binary classification techniques are extended to resolve the multiclass classification problems. Some algorithms are developed on the premise of support vector machines, Naive Bayes technique, k-nearest neighbor approach, neural network , decision tree etc to handle multiclass classification problems. These techniques are also known as algorithm adaptation techniques. Their brief description is given as follows:

- **Support Vector Machines:** SVM is based on principle of increasing the least distance from hyper plane which separates them to the nearest example set. Basically it is based on the idea of maximizing the margin. Basic SVM only deals with binary classification problems, but they are extended to handle the multiclass classification problems. The additional constraints and parameters are added to deal with multiple categories in multiclass.

- **Nave Bayes:** Nave Bayes is based on the technique of maximum a posteriori (MAP).It uses concept of conditional probability. It is naturally extensible to the multiclass problems; also they perform very well in those conditions in spite of their simplifying approach of solving the problem using conditional independence.

- **Reptree:** It belongs to decision tree category. Uses regression tree logic and creates multiple trees in incremental iterations. It is a powerful classification technique. Training data split infers a good general decision based on available features. It is automatically compatible to handle binary class as well as multiclass classification problems.

- **MLP:** Multilayer preceptor is class of artificial neural network. It consists of minimum of three layers of nodes each node is a neuron in MLP, except input nodes. It uses back propagation technique of supervised learning. Multiple layers and non linear activation differentiates MLP from other linear preceptors. It can distinguish the data which is difficult to separate linearly.

- **J48:**J48 is decision tree which is implementation of Iterative Dichotomiser 3 algorithm. For classifying new data it needs to make the decision tree of training data. It keep on comparing the testing data, when features are matched with any tree more than other then instances are classified to that class.

- **KNN:** It implements k nearest neighbor approach. It is based on similarity calculation between instances. It uses local average calculation for classification of data into different classes.

4.3.1 Evaluation Parameters

Main evaluation parameters used to determine the fair class division are Accuracy, Precision, Recall, RMSE and Kappa Statistics. Brief introduction about their theoretical background is given as follows:

- **Accuracy:** Accuracy is most important machine learning method which is used to evaluate the overall performance of any technique in predicting values as compared to actual.

$$Accuracy = \frac{TruePositive + TrueNegetive}{TruePositive + FalsePositive + FalseNegetive + TrueNegetive} \tag{4.1}$$

- **RMSE:** Root-Mean-Square Error (RMSE) is a frequently used evaluation parameter to measure the difference between $Model_a$ and the value $Model_p$ by the prediction model. It represents the sample standard deviation between the actual and predicted value of a prediction model [34].

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Model_p - Model_a)^2}{N}} \tag{4.2}$$

Here $Model_p$ Predicted Value by Prediction Model $Model_a$ Actual Value

- **Kappa Statistics:** Kappa measures the percentage of data values in the main diagonal of the table and then adjusts these values for the amount of agreement that could be expected due to chance alone [35]. The value of Kappa is defined as:

$$k = \frac{P_o - P_i}{1 - P_i} \tag{4.3}$$

The numerator represents the discrepancy between the observed probability of success and the probability of success under the assumption of an extremely bad case. Independence implies that pair of raters agree about as often as two pairs of people who effectively flip coins to make their ratings.

- **Precision:** Precision tells about how accurate model is in finding the predictive positive values, how many of them are actually positive.

$$Precision = \frac{TruePositive(TP)}{TruePositive + FalsePositive} \tag{4.4}$$

- **Recall:** It tells how many accurate positives model predicts by telling them positive.

$$Recall = \frac{Truepositive}{TruePositive + FalseNegetive} \tag{4.5}$$

4.4 Mood Profiling

Multiclass classification categorizes the sequence of devices operating conjointly into different categories. The nomenclature of multiple categories is done in such a simplest way that they're indicating the routined activities of a person based on device patterns. Persons mood is mirrored from his day to day activities. The proposed model make it possible to predict persons mood (sad, feeling hungry or tensed) based on his activities. If someone is preparing food in odd times like toaster is functioning at mid night then it indicates that the person is either tensed or hungry, if toaster isn't operating for whole day then the person is unwell. In mood profiling, results from multiclass classification are employed to further facilitate persons mood profiling supported by activity patterns. Figure 4.3 indicates

some of the sample mood profiles, which can be extracted from the combination patterns of simultaneously performed activities.

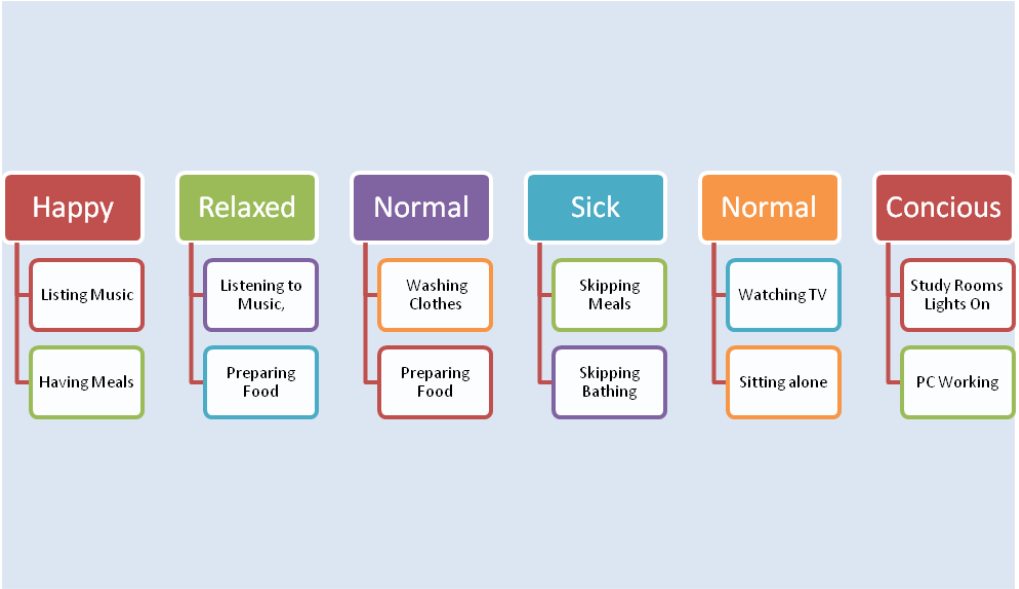

Figure 4.3: Mood profiling

4.5 Anomaly Recognition Model

Next step of proposed model is to discover those activities which don't seem to be expected or are missing out in line with time slot. This task is accomplished by using anomaly recognition techniques. Now question arises what are anomalies?

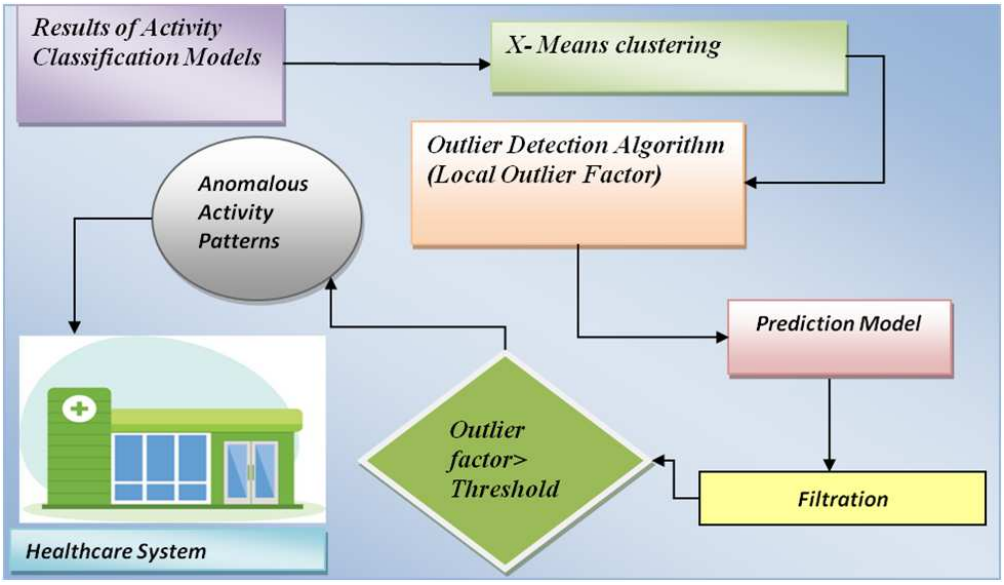

Figure 4.4: Anomaly Detection Model

Anomaly is outlined as uncommon pattern in data that don't tally to a normal behavior. Because of crucial impacts of anomalies, anomaly recognition is very important. Anomaly Recognition is technique, used to discover odd patterns, values, point that dont match with other items and patterns in dataset. When frequent patterns are mapped onto different activities of everyday life, then there may be some patterns that doesn't match to our day to day activities. Those activities need to be detected and evaluated whether those are outliers or our training model is incapable of recognizing them. Anomaly Recognition model works in several sub processes. First of all it performs x-means clustering on categorical data. Then it will discover outlier score using local outlier issue (LOF)[36] formula. Then prediction model is applied to filter out instances as normal or abnormal activity patterns. Then abnormal patterns are passed on to healthcare system. Concerted person may evaluate those patterns and take relevant actions according to problem. In following sections theoretical background of all the sub processes used in anomaly detection model are discussed. Section 3.5.1 illustrates the x-means clustering which is performed to make different clusters of multiple activity classes based on K nearest neighbor approach.

4.5.1 X-Means Clustering

X means clustering is performed on multiclass classification dataset. X means clustering algorithm determines exact no. of centroids on the basis of heuristic [38]. It starts with a minimum number of centroids and iteratively increases the centroids according to data.It uses concept of Euclidian distance.

Algorithm 4.2 X-means Algorithm

- Let the value of k = min_k.
- Implement K-means clustering method

loop

for do

- value of k = 1,2,3,....,k
- Change the value of centroid x(k) by two centroids values x(a) and x(b).

(Two new centroids for starting two different K-means clustering algorithm are obtained by transforming an initial value of centroid in two different directions along a randomly chosen value of vector by an amount equals to their cluster size.)

end for

end loop

• change the value of k=2 over the cluster k.

• Replace each centroid based on the criteria of model selection. (Model selection test BIC is performed to check if two new clusters better then original single cluster in each case. BIC (k)= 2* $\log P$ + K $\log q$ Where p no. of observations k clusters and $\log q$ log probability

if then

• condition of convergence condition is not fulfilled, go to Step 2. Otherwise Stop.

end if

4.5.2 Local Outlier Factor (LOF)

Local outlier factor (LOF) [36] is technique for outlier scoring.LOF is based on the concept of local density. Points are said local to each other on the basis of density. The identified common algorithmic scheme for local outlier detection consists of the following components:

Algorithm 4.3 Algorithm for Local Outlier Factor

Input: Database X

Output: (normalized) anomaly score for all and one x ϵ X

•Phase 1: Model Construction

loop

for all x ϵ X **do**

select condition (x)

Model(x):=build model(x, condition(x))

end for

end loop

•Phase 2: Model Comparison

loop

for all x ϵ X **do**

choose reference (x)

Score(x):= compare (model(x), model(n) n ϵ reference x

end for

end loop

•Phase 3:Normalization

if then

Normalization needed then

```
    loop
      for all x ϵ X do
        Normalized Score(X):=normalize (score(x))
      end for
    end loop
  end if
```

- Condition: a local condition of an object x for model building (condition(x))
- Model: the technique used for constructing the model
- Reference: a reference condition of object o for model comparison (reference(x))
- Comparison: the method used to compare the models
- Normalization: a (global) normalization procedure

When an object is compared to local densities of neighbors, the points which have lower density then neighbours are identified. These are termed as anomalies.

4.5.3 Prediction Model

Predictive modeling is a technique that employs data mining and probability to forecast outcomes. Every model is formed from variety of predictors, that are variables likely to influence future results. Predictive model can be applied to any problem regardless of the nature of occurrence of event. In context of anomaly detection, Model is first trained on example set using different learning algorithms like SVM, NN etc. Afterwards it can be applied on testing data, to transform the data.

4.5.4 Filtration

This technique is employed to selected only desirable output results. This method selects the favorable values and discards the unfavorable results. In context to this model, those Examples that match the given condition i.e. anomaly score $>$ 1.5 are filtered and coined as abnormal patterns using this method.

Chapter 5

Implementation and Results

The model analysis and experiments were run on a energy time series dataset collected from real house. The data is from United Kingdom Domestic Appliance Level Electricity dataset (UK-Dale). This energy time series dataset entails a total of fifty three appliances with time gap of six seconds . This dataset is published by UK Energy research Centre Energy data Centre (UKERC-EDC). [30]

5.1 Dataset Preparation Phase

To carry out the implementation of proposed model ,as discussed in section 4.1 data is converted into format which is suitable for proposed techniques.FP-growth algorithm for frequent pattern mining is applied on ready to mine dataset. The resulting patterns are of different sizes and shows different combinations. Minimum support value for FP-growth is 0.5. All the patterns having support more than 0.5 are considered as frequent patterns and stored in frequent item set database. Table 5.5 depicts some of sample frequent patterns .These Patterns provide an idea of conjoint operation of various devices. Data is converted from UNIX time stamp into 1 minute gap dataset as optimum for this proposed study. Different techniques of data preprocessing are implied to convert the raw data into ideal format.

Table 5.1: Initial Data

Unix Timestamp	**Energy Consumption**
1388534500	589
1388534506	565
1388534512	566
1388534519	600
1388534524	565
1388534531	558

Table 5.2: Intermediate Database

Date	**Boiler**	**Solar Pump**	**Laptop**	**Washing Machine**	**Dishwasher**	**TV**	**Kitchen Lights**
11/9/2012 22:28	12	53	24	0	1	1	0
11/9/2012 22:29	12	53	26	0	1	1	0
11/9/2012 22:30	12	52	23	0	1	1	0
11/9/2012 22:31	12	53	20	0	1	1	0
11/9/2012 22:32	12	53	20	0	1	1	0
11/9/2012 22:33	12	53	20	0	1	1	0
11/9/2012 22:34	12	53	20	0	1	1	0
11/9/2012 22:35	12	53	20	0	1	1	0
11/9/2012 22:36	12	53	20	0	1	1	0

Table 5.2 presents the unified view of intermediate database, which provides the information about the energy consumption by each device at particular period of time. Data is divided into 4 different time slots of morning, afternoon, evening, night. Table 5.3 gives better overview of how the data is portioned according to time the type of activities varies according to different phases of day. To make the pattern discovery process more efficient and making the proposed model rapid at tracking abnormalities, data divisions are done. Intermediate database shown in

Table 5.3: Divisions of data into different time slots

Time of day	**Database Name**
5:01am-12:00pm	Morning
12:1-5:0pm	Afternoon
5:1pm-8:00pm	Evening
8:00pm-5:00am	Night

Figure 5.2 is further processed to convert it into format which records the data in

table according to operating status of various appliances. Table 5.4 shows the final database which is ready to mine .The database is in binary format that represents the operating status of devices according to time. In Table 5.4, 1 implies device is on while 0 stands for no operation of device.

Table 5.4: Ready to mine dataset

Date	**Boiler**	**Solar Pump**	**Laptop**	**Washing Machine**	**Dishwasher**	**TV**
1/1/2014 0:10	1	0	0	0	1	1
1/1/2014 0:11	1	0	0	0	1	1
1/1/2014 0:12	1	0	0	0	1	1
1/1/2014 0:13	1	0	0	0	1	1
1/1/2014 0:14	1	0	0	0	1	1
1/1/2014 0:15	1	0	0	0	1	1
1/1/2014 0:16	1	0	0	0	1	1
1/1/2014 0:17	1	0	0	0	1	1
1/1/2014 0:18	1	0	0	0	1	1

5.2 Frequent Mining Results

Dataset preprocessed using different knowledge discovery techniques is further mined to extract the hidden patterns from data.Table 5.5 presents a glance of frequent patterns extracted from dataset by applying FP-growth algorithm.Patterns of different sizes are present in frequent pattern dataset.

Table 5.5: Frequent Activity Patterns

Support	**Device 1**	**Device 2**	**Device3**	**Device 4**	**Device5**	**Device 6**	**Device 7**
0.532	gas oven	Wifi	kitchen lights	Iron	kitchen stereo	Toaster	study room led
0.532	Router	gas oven	Wifi	Pc	Iron	kitchen stereo	Toaster
0.857	Router	Pc	kitchen stereo	Toaster	TV		
0.997	Router	Pc	Iron	kitchen stereo	study room led		
0.501	Router	gas oven	Iron	kitchen stereo	iPod charger		
0.693	kitchen radio	kitchen lights					
0.537	Toaster	living room sub-woofer					
0.532	gas oven	Wifi					

5.3 Results of Classification Models

After mining frequent patterns data is categorized into different manually annotated classes. The classes assigned to different patterns resemble to daily life activities like if person is working in kitchen then activity performed is preparing food. If wifi , PC, study room lights are on then person is studying. If iron is on and parallely TV is on then activity performed is pressing cloths while watching TV.

In this way classes are assigned to frequent patterns. For multiclass classification , five different classifiers are used which classify data into different classes. Classifiers employed are SVM, MLP, J48, RepTree, Nave Bayes. Different models have provided with mixed type of results. Some models performed very well while others have average performance. SVM, MLP have consistent performance

on different datasets. Different datasets signifies separate databases for morning, afternoon, evening, night as discussed in section 4.1. Different evaluation parameters like kappa statistics, RRSE,FP rate, TP rate are calculated for each prediction model .Distinct number of instances are used to testify the models .The results are recorded in following tables:

Table 5.6: Results of Classification Model

	Model name	**Accuracy**	**MAE**	**RAE**	**RRSE**	**RMSE**	**Kappa Statistics**
Morning	SVM	92.33%	0.1491	97.39%	95.52%	0.264	0.9082
Results	MLP	97.67%	0.0048	3.14 %	20.53 %	0.0567	0.9723
(5am-12pm)	KNN	92.81%	0.0395	25.91%	43.87 %	0.1212	0.9144
	J48	94.96%	0.0012	7.89 %	31.16 %	0.0861	0.9399
	RepTree	91.72%	0.053	34.82%	46.22%	0.1278	0.901
	Nave Bayes	79.91%	0.049	32.59%	38.43%	0.1615	0.761
Afternoon	SVM	95.32%	0.148	98.21%	95.421%	0.264	0.9438
Results	MLP	99.92%	0.0007	0.48%	3.7804 %	0.0104	0.9991
(12pm-5pm)	KNN	88.59%	0.148	96.33%	95.25%	0.2645	0.862
	J48	93.99%	0.0247	19.23%	40.17%	0.1017	0.9278
	RepTree	97.42%	0.035	27.39%	37.52%	0.095	0.969
	Nave Bayes	84.30%	0.0371	24.36%	52.15%	0.1439	0.8152
Evening	SVM	95.35 %	0.0114	7.49%	31.51 %	0.0871	0.944
Results	MLP	98.44%	0.0076	4.56%	17.71%	0.0511	0.981
(5pm-8pm)	KNN	96.51 %	0.0302	18.13%	34.43 %	0.0993	0.958
	J48	80.62%	0.042	32.79%	62.66%	0.159	0.766
	RepTree	82.62%	0.0312	24.27%	52.48%	0.133	0.792
	Nave Bayes	93.30%	34.70%	0.879	14.23%	0.018	0.919
(Night)	SVM	98.86%	0.0065	4.24%	4.24%	0.0402	0.9864
Results	MLP	99.97%	0.0007	0.45%	2.16%	0.006	0.9997
(8pm-5am)	KNN	96.55%	0.028	18.34%	34.90%	0.0964	0.958
	J48	96.89%	0.008	5.59%	24.64%	0.0681	0.0681
	RepTree	90.39%	0.022	14.56%	41.82%	0.1156	0.0885
	Nave Bayes	94.89%	0.213	13.91%	31.09%	0.0859	0.939

Further Table 5.7 gives the detailed accuracy statistics. Different evaluation patterns based on accuracy are calculated. From results ,performance of different

algorithms are evaluated.Support vector machines and MLP are the best performing classifiers.

Table 5.7: Detailed Accuracy Statistics

	Model name	TP Rate	FP Rate	Precision	Recall	F-Measure	PRC	ROC
Morning Results 5am-12pm	SVM	0.983	0.003	0.984	0.983	0.981	0.977	0.998
	MLP	0.977	0.003	0.979	0.977	0.977	0.986	0.990
	KNN	0.928	0.011	0.930	0.928	0.927	0.967	0.995
	J48	0.950	0.008	0.950	0.950	0.949	0.959	0.992
	RepTree	0.917	0.014	0.918	0.917	0.915	0.996	0.970
	Nave Bayes	0.799	0.029	0.811	0.799	0.791	0.981	0.888
Afternoon Results 12pm-5pm	SVM	0.954	0.007	0.951	0.954	0.952	0.945	0.987
	MLP	0.999	0.000	0.999	0.999	0.999	1.000	1.000
	KNN	0.965	0.006	0.965	0.965	0.965	0.989	0.998
	J48	0.940	0.011	0.941	0.940	0.946	0.994	0.968
	RepTree	0.974	0.02	0.974	0.974	0.973	0.999	0.987
	Nave Bayes	0.843	0.015	0.880	0.843	0.850	0.934	0.986
Evening Rsults 5pm-8pm	SVM	0.953	0.003	0.954	0.953	0.949	0.949	0.992
	MLP	0.999	0.000	0.998	0.999	0.999	1.000	1.000
	KNN	0.965	0.006	0.966	0.965	0.965	0.988	0.998
	J48	0.806	0.040	0.809	0.806	0.805	0.806	0.994
	RepTree	0.826	0.260	0.886	0.826	0.829	0.973	0.862
	Nave Bayes	0.931	0.007	0.912	0.993	0.921	0.999	0.978
Night Results 8pm-5pm	SVM	0.989	0.001	0.990	0.989	0.989	0.998	1.000
	MLP	0.999	0.000	1.000	0.999	1.000	1.000	1.000
	KNN	0.966	0.006	0.966	0.966	0.960	0.987	0.998
	J48	0.962	0.002	0.959	0.969	0.905	0.986	0.980
	RepTree	0.904	0.013	0.912	0.904	0.905	0.986	0.98
	Nave Bayes	0.949	0.006	0.935	0.949	0.941	0.999	0.988

Data is divided into four portions according to divisions of day (morning, afternoon, evening and night time slots). In morning data all the devices which are active in morning from 5 a.m. to 12 p.m. are considered. Figure 5.1,5.2, 5.3 shows the accuracy, precision, recall and kappa statistics graph for morning data. Model MLP performed best. Figure 5.4 depicts the results for afternoon data.

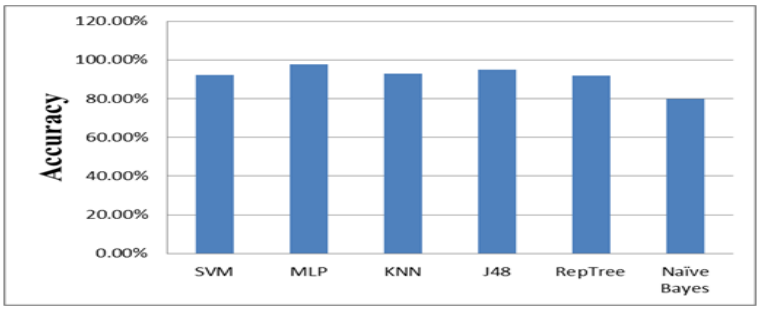

Figure 5.1: Accuracy Graph for Morning data

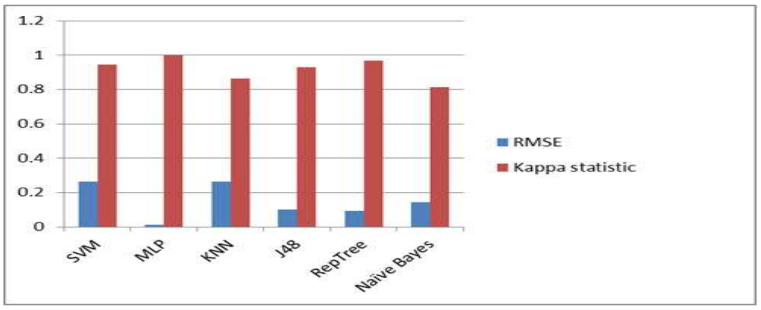

Figure 5.2: RMSE and Kappa Statistics Graph for Morning data

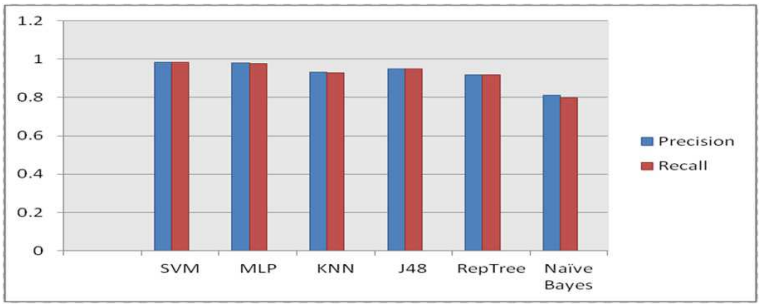

Figure 5.3: Precision, Recall Graph for morning data

Afternoon time slot lies between 12p.m. to 5 p.m. Different classifiers like SVM, MLP, J48 etc are applied to classify the patterns in different manually labeled categories. Figure 5.4, 5.5, 5.6 shows that results provided by MLP is best ,while J48 follows .Similarly graphs for Precision and recall are provided. Figure 5.7,

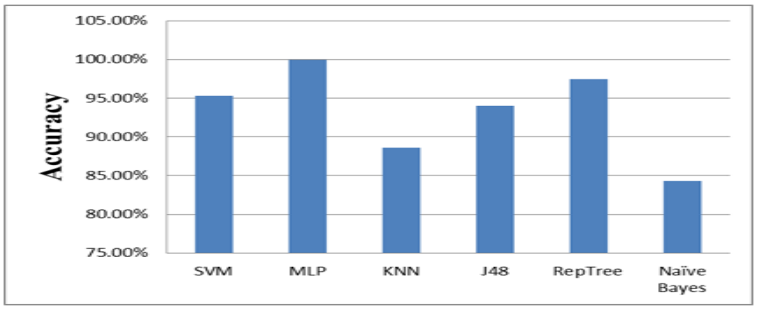

Figure 5.4: Accuracy Graph for afternoon data

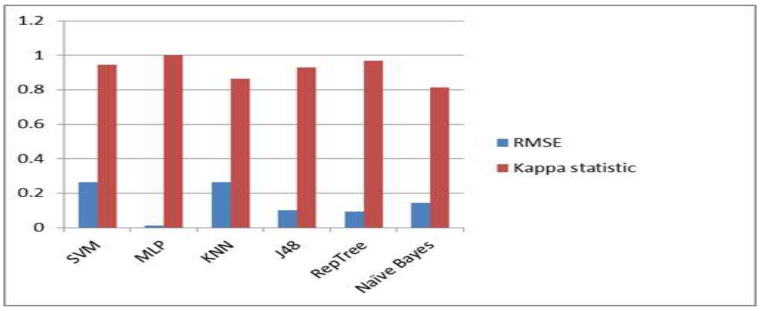

Figure 5.5: RMSE and Kappa Statistics Graph for afternoon data

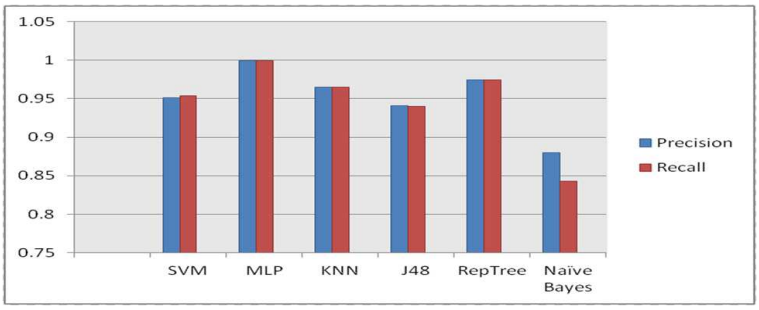

Figure 5.6: Precision, Recall Graph for afternoon Period

Figure 5.8, Figure 5.9 are associated with evening data. Evening time database consists of records of all the active devices between 5 p.m. to 8 p.m. Most of the activities in this time slot are related to preparing food and relaxing. Accuracy

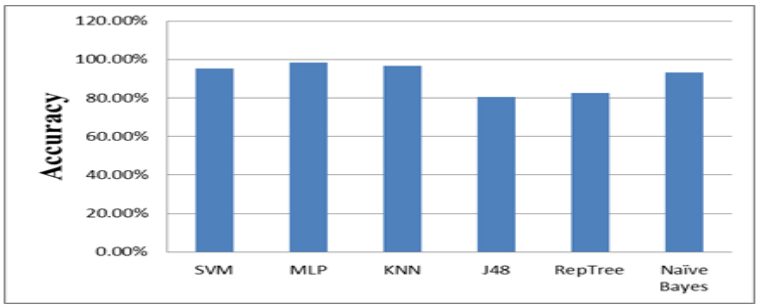

Figure 5.7: Accuracy Graph for Evening Data

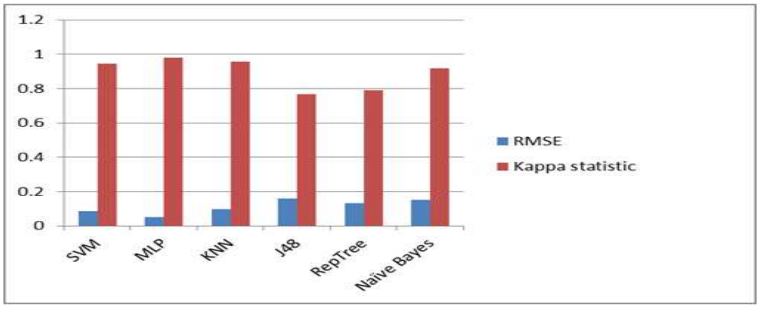

Figure 5.8: RMSE and Kappa Statistics Graph for Evening Data

provided by MLP is higher in this time slot. RMSE value of MLP is least followed by SVM. Figure 5.9 shows plot of precision and recall for evening time. Night

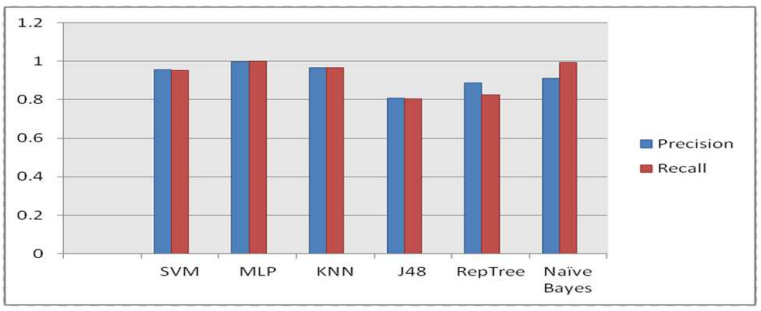

Figure 5.9: Precision and Recall graph for Evening Data.

time data files consider frequent patterns of devices which are active from 8 p.m. in night to 5 a.m. in morning. This is the largest time slot but activities in this time slot are limited as most of the devices are OF because occupants of house are resting. MLP performed best in this case also. Performance of Rep Tree is worst as compared to all the models. Next is the graph for precision and recall of night

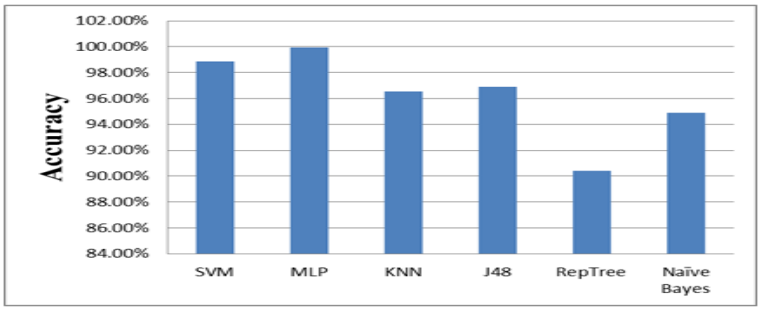

Figure 5.10: Accuracy Graph for Night Data

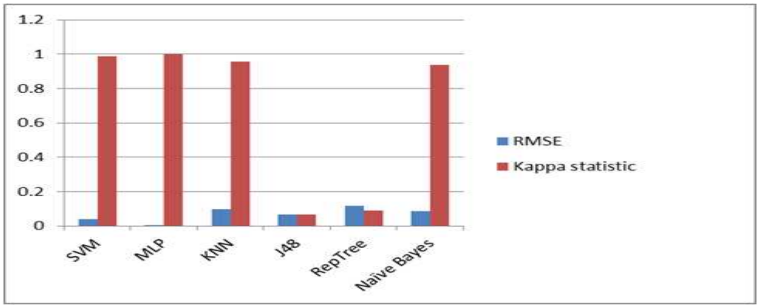

Figure 5.11: RMSE and Kappa Statistics for Night Data

time data as in accordance with in table 5.7. Figures 5.10,5.11,5.12 represents the accuracy, RMSE and Kappa statistics,Precision and Recall graph for night time period.MLP providesd the best results in this case also. Overall,from graph plots

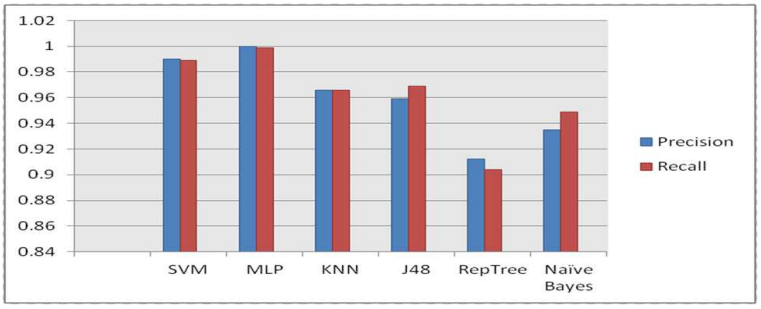

Figure 5.12: Precision and Recall Graph for Night Data

of accuracy, RMSE, Kappa Statistics, Precision and Recall performance of MLP model is best for all the four time slots.SVM is second best model which provided the good results.

5.4 Mood Profiling Results

Smart meter data can be analyzed in many different ways to trace out different hidden patterns related to persons physical activities as well as his mental status. Mood Profiling is one of the technique which is correlated to persons daily life activities. If person is preparing food and radio is on then it signifies that person is in happy mood. Similarly if person is skipping meals then maybe he is sad. The Mood Profiling is accomplished with the help of frequent patterns .Frequent patterns are directly or indirectly mapped to different moods domains. Like happy, sad, hungry, sick etc.

Table 5.8: Examples of Mood Profiling

Active Devices			**Activity**	**Mood**
Kitchen Stereo	boiler	kitchen Lamp	Preparing Food while listening to radio	Normal
Kitchen Stereo	toaster	Boiler	Daily life chores	Normal
Router	iron	TV	Daily life chores while watching TV	Relaxing

to be cont'd on next page

Table 5.8: Examples of Mood Profiling (Cont.)

Active Devices			**Activity**	**Mood**
Router	iron	Kitchen Radio	Daily life chores while watching TV	Relaxing
Gas Oven	kitchen radio	Toaster	Daily life chores while watching TV	Normal
Living Room lights	TV	Living Room subwoofer	Enjoying Music	Happy

5.5 Results for Anomaly Recognition

Anomaly Recognition is most significant and final step of proposed model. As already discussed in section 4.5 anomalies or outliers are patterns which deviates from normal behavior. In the context of this study, if imbalanced appliance operating status are detected then it means they are anomalies. Anomalies are those patterns to which classification model fails to assign a categorical label. The reason behind this is distinctive nature of these patterns from normal patterns. The proposed model for anomaly recognition is already discussed in section 3.5. Here results of anomaly recognition is summarized. Anomalies are detected using X-means clustering and anomaly score.

Table 5.9: Anomaly Score Table for Morning Data

ID	**Cluster Id**	**Anomaly Score**
581	cluster_0	1.73
582	cluster_0	1.51
587	cluster_0	1.75
589	cluster_0	1.52
601	cluster_1	1.50
602	cluster_1	1.56
604	cluster_1	1.52
605	cluster_1	1.59

to be cont'd on next page

Table 5.9: Anomaly Score Table for Morning Data (Cont.)

ID	**Cluster Id**	**Anomaly Score**
606	cluster_1	1.56
610	cluster_2	1.50
611	cluster_2	1.53
613	cluster_2	1.52
614	cluster_2	1.59
615	cluster_2	1.56

The Table 5.9,5.10 ,5.11,5.12 are the result of anomaly recognition where each table has an original id,cluster id,anomaly score.Id column depicts the original id of classification pattern which is anomalous to the normal behavior.Tables are adjoined by graphs which give the better overview of associated anomalies according to anomaly score. The table 5.9 depicts the anomalies recognized from morning time data.Here ID tells about the identity of anomalous pattern which is different from normal patterns. The table 5.10 depicts the anomaly score table

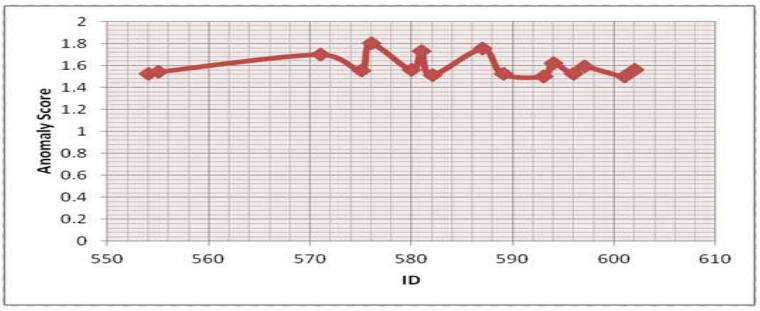

Figure 5.13: Anomaly graph for Morning

for afternoon database. From results we can see that anomaly score for most of the patterns is not much extreme only one or two values have shown extreme behavior of deviation from normal patterns. Most of the anomalies belong to cluster number three.

Table 5.10: Anomaly Score Table for Afternoon Data

ID	**Cluster Id**	**Anomaly Score**
47	cluster_0	1.86
208	cluster_3	1.57
213	cluster_0	1.52
1412	cluster_2	1.50
2357	cluster_2	1.57
2365	cluster_2	1.53
3211	cluster_2	1.59
3220	cluster_2	1.56
3761	cluster_2	1.70
3771	cluster_2	1.52
4009	cluster_2	1.56
4020	cluster_2	1.51

The Figure 5.14 is anomaly graph associated with afternoon time. In this case only one pattern has anomaly score more then 1.6 . Hence, there are very cases of anomalous behavior in afternoon time. Table 5.11 shows the anomaly values for

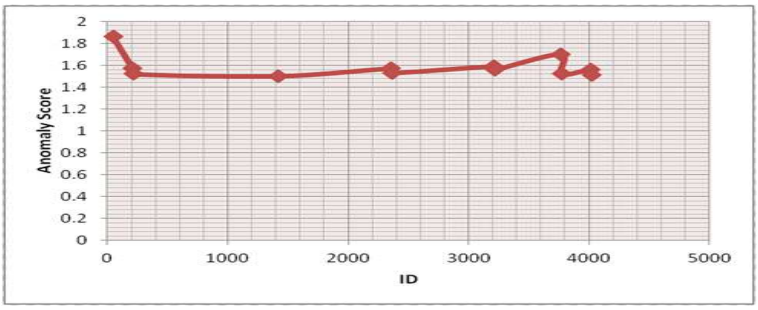

Figure 5.14: Anomaly graph for Afternoon

evening data which is recorded between 5 p m. in evening. to 8 p.m. Frequency of anomalous behavior in evening time is less as compared to morning and afternoon time. One thing which seeks attention is higher anomaly score in this time slot. Higher anomaly score signifies the extreme behavior of residents of house in particular time slot. It indicates that the patterns have much difference from normal pattern and it needs attention of concerned healthcare attendents as soon as possible.

Table 5.11: Anomaly Score Table for Evening Data

ID	Cluster Id	Anomaly Score
48	cluster_0	1.73
49	cluster_0	2.18
4043	cluster_3	1.73
4319	cluster_3	1.72

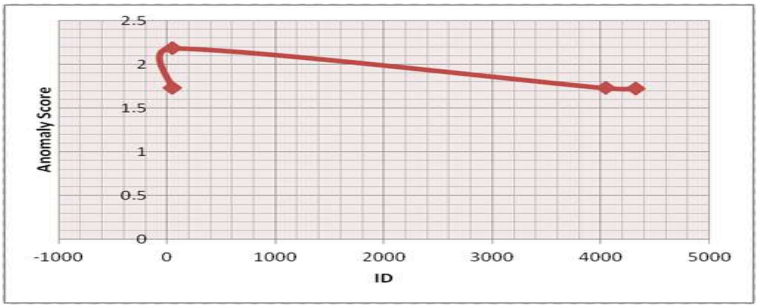

Figure 5.15: Anomaly graph for Evening

The Table 5.12 and Figure 5.16 shows the results for night time slot.

Table 5.12: Anomaly Score Table for Night Data

ID	Cluster Id	Anomaly Score
1	cluster_1	1.61
300	cluster_2	1.52
803	cluster_2	1.51
1624	cluster_2	1.60
2610	cluster_2	1.59
3487	cluster_2	1.63
4058	cluster_2	1.65
4321	cluster_2	1.61
4401	cluster_2	1.62

It is the largest time slot but most of the devices are inactive in this time slot as the occupants sleep at this time. The anomaly score in this time slot are not associated with much extreme values. From the anomaly values, probably of

anomalous activities is less.

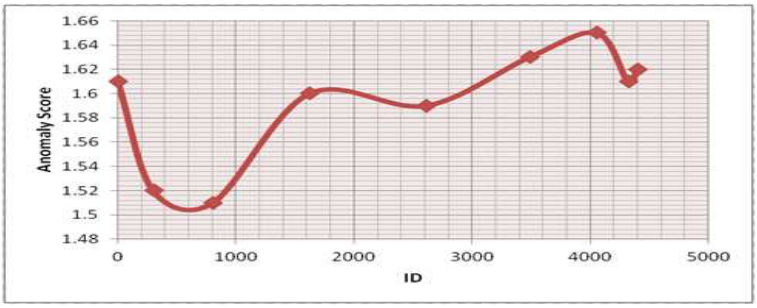

Figure 5.16: Anomaly graph for Night

This chapter was about the results and implementation details of behaviour monitoring and anomaly detection model of human activities. The MLP model provided the best results for behaviour monitoring. The anomaly detection model detected the variable number of anomalies for different time slots.

Chapter 6

Conclusion & Future Scope

This research demonstrates the activity recognition, mood profiling and anomaly recognition model to track individuals activities from low resolution smart meter energy consumption data on time series data.

6.1 Conclusion

Most of the times person perform similar habitual activities every day. But sometimes the activities are not performed in normal sequence by the person.

- Firstly time series dataset is preprocessed and changed into required 1 minute gap data from time series data.
- Frequent activity patterns based on simultaneous operation of different appliances are mined using FP growth algorithm.
- Multiclass classification approach is implemented to categories the activities into manually created classes. Different multiclass classification algorithms are used .SVM model and MLP performed better then all the other model with accuracy $> 95\%$ in all the four time slots.
- Further activity patterns are utilized to predict the mood of inhabitants of house.
- Anomaly detection Model is applied which detected anomalies in different time slots based on outlier detection algorithm. Different time slots are associated with different no. of anomalies.

When an anomalous pattern is confronted the anomaly detection model detects it and send regular reports to healthcare system about abnormal activities. Concerned authorities can take required actions according to severity of problem. If this system is implemented in healthcare sector then tracking the health status of people will become conventional. The cost of implementation of this model is also negligible as compared to deploying sensors in each house. In future this system

can be used for old age citizens for keeping an eye on their health status. It will also be efficient for keeping an eye on the health status of those patients who are suffering from serious health issues.

6.2 Future Scope

As a part of current research activity recognition, mood profiling, abnormal behavior of residents is detected based on daily life activities. However there is scope of future research to improve the techniques employed and to extend the objectives of study.

- Real time data from multiple smart homes can be used in future model.
- Other advanced metering infrastructures like gas, water consumption data can be merged with electricity data to predict the health status of occupant in much accurate way.
- Study can be extended to recognize emotional, cognitive and social behavior of resident of smart homes based in their day to day activities.
- Other techniques like parallel computing, cloud computing, Artificial Intelligence can be used to work on real time data.

References

[1] Using energy more efficiently and targeted control of consumption. [Online]. Available: http://www.ict-smart-cities-centre.com/en/expertise/smart-metering/

[2] Artificial intelligence:a smarter way to build smart cities. [Online]. Available: http://www.financialexpress.com/opinion//ArtificialIntelligence:Asmarterwaytobuildsmartcities/1202358

[3] J. Sen, "Ubiquitous computing: Potentials and challenges," 2010.

[4] C. Borgelt, "Keeping things simple: finding frequent item sets by recursive elimination," in *Proceedings of the 1st international workshop on open source data mining: frequent pattern mining implementations*. ACM, 2005, pp. 66–70.

[5] T. id Items, "Association analysis: Basic concepts and algorithms," 0.

[6] J. Han, J. Pei, Y. Yin, and R. Mao, "Mining frequent patterns without candidate generation: A Frequent-Pattern tree approach," *Data Min Knowl Disc*, vol. 8, no. 1, pp. 53–87, 2004.

[7] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *Acm Comput Surv Csur*, vol. 41, no. 3, p. 15, 2009.

[8] J. Clement, J. Ploennigs, and K. Kabitzsch, *Detecting Activities of Daily Living with Smart Meters*. springer, 2014.

[9] Y. Yang, H. Yang, Y. Hong, J. Kim, M. Park, H. Na, I. Han, and S. Kim, "Activities of daily living and dementia," *Dementia Neurocognitive Disord*, vol. 11, no. 2, pp. 29–37, 2012.

[10] X. Liu and P. Nielsen, "Scalable prediction-based online anomaly detection for smart meter data," *Inform Syst*, 2018.

[11] C. Chalmers, W. Hurst, M. Mackay, and P. Fergus, "Smart meter profiling for health applications," pp. 1–7, 2015.

[12] K. Basu, L. Hawarah, N. Arghira, H. Joumaa, and S. Ploix, "A prediction system for home appliance usage," vol. 67, 2013.

[13] T. Thomas, C. Cashen, and S. Russ, "Leveraging smart grid technology for home health care," 2013.

[14] C. Chelmis, J. Kolte, and V. K. Prasanna, "Big data analytics for demand response: Clustering over space and time," 2015.

[15] T. Cao, X. Wu, T. Hu, and S. Wang, *Active Learning of Model Parameters for Influence Maximization.* springer, 2011, vol. 6911.

[16] K. Gajowniczek and T. Zbkowski, "Data mining techniques for detecting household characteristics based on smart meter data," vol. 8, 2015.

[17] P. Pouladzadeh, P. Kuhad, S. Peddi, A. Yassine, and S. Shirmohammadi, "Mobile cloud based food calorie measurement," *2014 Ieee Int Conf Multimedia Expo Work Icmew*, pp. 1–6, 2014.

[18] A. Yassine, S. Singh, and A. Alamri, "Mining human activity patterns from smart home big data for health care applications," vol. 5, 2017.

[19] M. Zhang and A. Sawchuk, "A feature Selection-Based framework for human activity recognition using wearable multimodal sensors," 2011.

[20] N. Truong, M. James, T. Long, E. Costanza, and S. D. Ramchurn, "Forecasting multi-appliance usage for smart home energy management," 2013.

[21] S. Singh, A. Yassine, and S. Shirmohammadi, "Incremental mining of frequent power consumption patterns from smart meters big data," 2016.

[22] M. Alam, M. Reaz, and M. Ali, "A review of smart homesPast, present, and future," vol. 42, 2012.

[23] S. M. Hossain, "Patient state recognition system for healthcare using speech and facial expressions," *J Med Syst*, vol. 40, no. 12, p. 272, 2016.

[24] M. Alam, N. Roy, M. Petruska, and A. Zemp, "Smart-energy group anomaly based behavioral abnormality detection." 2016.

[25] P. Carroll, T. Murphy, M. Hanley, D. Dempsey, and J. Dunne, "Household classification using smart meter data," *J Off Stat*, vol. 34, no. 1, pp. 1–25, 2018.

[26] M. Fahim, I. Fatima, S. Lee, and Y. Lee, *Activity recognition.* acm, 2011.

[27] D. Alahakoon and X. Yu, "Smart electricity meter data intelligence for future energy systems: A survey," *Ieee T Ind Inform*, vol. 12, no. 1, pp. 425–436, 2016.

[28] G. Zhang, G. G. Wang, H. Farhangi, and A. Palizban, "Data mining of smart meters for load category based disaggregation of residential power consumption," *Sustain Energy Grids Networks*, vol. 10, pp. 92–103, 2017.

[29] H. F. Nweke, Y. W. Teh, M. A. Al-Garadi, and U. R. Alo, "Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges," *Expert Systems with Applications*, 2018.

[30] J. Kelly and W. Knottenbelt, "The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes," vol. 2, 2015.

[31] J. Han, H. Cheng, D. Xin, and X. Yan, "Frequent pattern mining: current status and future directions," *Data Min Knowl Disc*, vol. 15, no. 1, pp. 55–86, 2007.

[32] J. Han, J. Pei, and Y. Yin, *Frequent Patterns Tree Generation: A Frequent-Pattern Tree Approach.* acm, 2000, vol. 29.

[33] Smart homes:past,present and future. [Online]. Available: http://www.integratedio.com/smart-home-business-technology/smart-home-past-present-and-future/

[34] Statistics how to. [Online]. Available: http://www.statisticsHowTo.com/rmse/

[35] What is kappa coefficient (kohn's kappa). [Online]. Available: http://www.pmean.com/definitions/kappa.htm

[36] M. M. Breunig, H. Kriegel, R. T. Ng, and J. Sander. acm, 2000, vol. 29.

[37] H. Lin, "Efficient classifiers for multi-class classification problems," *Decis Support Syst*, vol. 53, no. 3, pp. 473–481, 2012.

[38] D. Pelleg and A. W. Moore, "X-means: Extending k-means with efficient estimation of the number of clusters." vol. 1, 2000.

[39] K. Basu, V. Debusschere, and S. Bacha, "Appliance usage prediction using a time series based classification approach," 2012.

[40] K. Vadim, "Overview of different approaches to solving problems of data mining," *Procedia Comput Sci*, vol. 123, pp. 234–239, 2018.

[41] W. Stephen, A. Yelundur, M. Charlie, and M. Landry, "Support vector machine implementations for classification & clustering," *Bmc Bioinformatics*, vol. 7, no. S2, pp. 1–18, 2006.

[42] K. Lin, I. Liao, and Z. Chen, "An improved frequent pattern growth method for mining association rules," *Expert Syst Appl*, vol. 38, no. 5, pp. 5154–5161, 2011.

[43] M. Hülsmann and C. M. Friedrich, *Comparison of a Novel Combined ECOC Strategy with Different Multiclass Algorithms Together with Parameter Optimization Methods.* springer, 2007, vol. 4571.

[44] Activities of daily living checklist & assessments. [Online]. Available: http://www.payingforseniorcare.com/longtermcare/activities-of-daily-living.html

[45] S. Peddi, P. Kuhad, A. Yassine, P. Pouladzadeh, S. Shirmohammadi, and A. Shirehjini, "An intelligent cloud-based data processing broker for mobile e-health multimedia applications," vol. 66, 2017.

[46] J. Alcalá, O. Parson, and A. Rogers, *Detecting Anomalies in Activities of Daily Living of Elderly Residents via Energy Disaggregation and Cox Processes*, 2015.

[47] E. Schubert, A. Zimek, and H. Kriegel, "Local outlier detection reconsidered: a generalized view on locality with applications to spatial, video, and network outlier detection," *Data Min Knowl Disc*, vol. 28, no. 1, pp. 190–237, 2012.

[48] M. Chan, D. Estève, C. Escriba, and E. Campo, "A review of smart homesPresent state and future challenges," *Comput Meth Prog Bio*, vol. 91, no. 1, pp. 55–81, 2008.

[49] S. Idowu, C. Åhlund, O. Schelén, and R. Brännström, "Machine learning in pervasive computing," 2013.

[50] H. Nweke, Y. The, G. Mujtaba, and M. Al-garadi, "Data fusion and multiple classifier systems for human activity detection and health monitoring: Review and open research directions," *Inform Fusion*, 2018.

List of Publications

1. Anupam Kaur Grewal, Maninder Kaur *A Survey on Activity Recognition and Health Monitoring Systems for Smart Homes.* [Communicated]
2. Anupam Kaur Grewal, Maninder Kaur *Human Scenario Abnormality Detection in Smart Metering Infrastructure.* [Communicated]