DETECTING CELL OUTAGE BY APPLYING DENSITY BASED ANOMALY DETECTION ALGORITHM USING MACHINE LEARNING TECHNIQUE: THE CASE OF ETHIO TELECOM UMTS NETWORK

BY: Aklilu Bisrat

ADVISER: Dr. –Ing. Dereje Hailemariam

A Thesis Submitted to

the School of Electrical and Computer Engineering,

Addis Ababa Institute of Technology

In Partial Fulfillment of the Requirements for the Degree of Masters of Science in Telecommunications Network Engineering

February 1, 2020

Addis Ababa University

Addis Ababa, Ethiopia

Declaration

I, the undersigned, declare that the thesis comprises my own work in compliance with internationally accepted practices; I have fully acknowledged and referred all materials used in this thesis work.

Aklilu Bisrat _____

Name *Signature*

ADDIS ABABA UNIVERSITY

ADDIS ABABA INSTITUTE OF TECHNOLOGY

SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

This is to certify that the thesis prepared by ***Aklilu Bisrat****, entitled "Detecting Cell Outage by Applying Density Based Anomaly Detection Algorithm using Machine Learning Technique: The case of ethio telecom* Universal Mobile Telecommunication System *(UMTS) network" and submitted in partial fulfillment of the requirements for the degree of Masters of Science in Telecommunication Engineering, complies with the regulation of the University and meets all the accepted standards with respect to originality and quality.*

Signed by the Examining Committee:

Name	Signature	Date
_____	_____	_____

(Examiner)

_____	_____	_____

(Examiner)

Dr. –Ing. Dereje Hailemariam	_____	_____

(Advisor)

_____

Dean, School of Electrical and Computer

Engineering

Abstract

This thesis develops a model to detect cell outage on real ethio telecom network data by using density based anomaly detection algorithms through the application of machine learning technique. Cell outage is the total/partial loss of radio network in a given area and the process of detecting cell outage is called cell outage detection. Cross-Industry Standard Process (CRISP-DM) machine learning methodology, which is a six staged open standard process model that describes common approaches for data mining or machine learning was used. The considered data was normal and problematic network environment obtained from ethio telecom UMTS network in Addis Ababa. The proposed detection framework used network performance data (incoming handover (inHO), which is the process of transferring an ongoing call from one cell to the other, and traffic data, originated from base station and terminated to mobile device) of the neighbor cells to capture the normal network state and to detect the outage of the target cell automatically in a pre-set time interval in the UMTS network environment.

To profile the normal network operation, the study used two density based anomaly detection algorithms; namely, the K- Nearest Neighbor (K-NN) and Local Outlier Factor (LOF) algorithms, of which one was selected based on its performance during the training. To validate the models, K-*fold cross validation technique was used and for the selection of the optimal model, parameter selection was done for different values of K (K=1,2, 3…30). To compare the two algorithms, Receiver Operating Characteristic (ROC) curves were used. Based on the results, the K-NNAD was found to be of a better performance than the LOFAD, thus was selected as a detector in the profiling stage. The proof of the system model was tested by using real problematic network state data and the results of classic data mining metrics were obtained. Based on the results obtained from the testing, the K-NNAD method was found to perform better in detecting outage cells in the proposed framework.*

Key words

Cell outage, Machine learning, Anomaly detection, Self-healing, UMTS network, CRISP-DM

ACKNOWLEDGMENTS

Frist and for most, I would like thank God for being a source of courage and confidence in my entire stay in the program. My very special gratitude also goes to my advisor, Dr.-Ing. Dereje Hailemariam for his enormous and genuine advises from the very beginning the end of the thesis. Dear advisor, your valuable comments and suggestions were all really vital for the successful completion of this thesis.

I would also like to thank the ethio telecom mobile network staff members, especially Selam G/Medihin, for their unreserved support while I was desperately in need of network data for my study. Without their support, this study could have not been a reality.

My deepest thanks also due to all my relatives and friends, especially Solomon Bekele, who have supported me in one way or the other throughout my study.

Last but not least, my heartfelt thanks go to my wife Aynalem Demisie and little angel Makiba Aklilu. You were very much cooperative, supporting and considerate throughout my study time. Thank you again and again…

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

3GPP	3rd Generation Partnership Project
AuC	Authentication Centre
BS	Base Station
CAPEX	Capital Expenditure
CNMS	Cellular Network Management System
COC	Cell Outage Compensation
COD	Cell Outage Detection
COM	Cell Outage Management
CRISP-DM	Cross-Industry Standard Process- Data Mining
CS	Circuit Switching
DM	Domain Managers
EIR	Equipment Identity Register
EMS	Element Management System
eNB	Evolved NodeB
GSM	Global Systems for Mobile Communications
GMSC	Gateway MSC
GGSN	Gateway GPRS Support Node
HO	Handover
IMSI	International Mobile Subscriber Identity
IP	Internet Protocol
KDD	Knowledge Discovery in Databases
K-NN	K-Nearest Neighbor
KPI	Key Performance Indicator
LOF	Local Outlier Factor

LTE	Long Term Evolution
MDS	Multi-Dimensional Scaling
MDT	Minimization of Drive Tests
MSC	Mobile Services Switching Center
NE	Network Element
NGMN	Next Generation Mobile Network
NMS	Network Management System
OAM	Operation Administration and Maintenance
OPEX	Operational Expenditure
PM	Performance Management
PS	Packet Switching
QoE	Quality of Experience
QoS	Quality of Service
RAN	Radio Access Network
RAT	Radio Access Technology
RLF	Radio Link Failure
RRM	Radio Resource Management
RNS	Radio Network Subsystems
SON	Self-Organizing Networks
SVM	Support Vector Machine
TTI	Transmission Time Interval
UE	User Equipment
UMTS	Universal Mobile Telecommunication System
UNMS	Unified Network Management System
USIM	UMTS Subscriber Identity Modules
UTRAN	UMTS Terrestrial Radio access network
VLR	Visitor Location Register

1 Introduction

1.1 Background of the Study

Global communication has been made easy since the start of telecommunication technologies, which contributed a lot for the creation of a globally interconnected world. Newly emerging technologies of telecommunication are making the lives of billions of peoples worldwide much simpler and easier than before. With over 5.1 billion mobile subscribers worldwide, telecom service providers are being challenged by the increasing demands for wider coverage and service quality by their customers[1]. This forces service operators to allocate ever increasing resources and work harder so as to increase their coverage and capacity with the purpose of reaching the ever expanding market both in terms of quantity and quality[2].

Operators deploy new infrastructures that are capable of handling larger volume of traffic. However, increasing infrastructure may cause challenges on operation and maintenance such as optimization problem and detecting network outages. It is, thus, a must for operators to use more efficient operational and maintenance strategies. Such strategies can improve system performance and fault management systems making the operator more effective and efficient. To improve the effectiveness and efficiency of network management systems, one possible solution is automation of the network management system.

The current technology in Mobile Cellular Network Management (MCNM) that can take over the roles of the traditional methods of network management is Self-Organizing Networks (SON)[3], which makes automation on network management system possible. SON has three major functionalities; namely, self-configuration, self-optimizations and self-healing. Self-healing is the one that is capable of overtaking the function of the traditional cellular network management

system and turns it into an automated system to detect and compensate cell outage. Thus, this study focusses on the self-healing function of SON on cell outage detection.

Traditionally, cell outages have been detected manually either by the analysis of fault alarms or by analyzing performance measurement at operation and maintenance centers. Sometimes, it might be a must to visit the sites or run a drive testing[4]. This consumes time and may result in reduction in service capacity and quality[5]. However, the weakness of the traditional outage detection method can be compensated if we employ the self-healing functionality of SON.

ethio telecom, the only telecom service provider in Ethiopia, uses traditional method of Cell Outage Management (COM), in which the network management is done manually without using an automated approach to detect and compensate cell outage. Such a traditional management method has made delays in cell outage detection a common occurrence in the operator. In some cases, it might take days and even weeks to detect cell outage [4]. This has effects on network coverage leading towards network coverage gap, decline in service quality as well as in service capacity. Thus, it seems a must for the operator to automate its COM system by deploying the self-healing function of SON.

SON is a self-organizing network that provides network intelligence, automation and network management features that can be used in automating the configuration and optimization of wireless networks so as to adapt them to varying radio channel conditions [6]. While using the self-healing function of SON to detect cell outage, we can employ supervised, semi-supervised or unsupervised machine learning techniques that use different types of algorithms such as Support Vector Machine (SVM), K-NN, LOF and Neural Network. The algorithms use various network measurements including HO data, traffic data, and Key Performance Indicator (KPI) data as well as other user level data like Reference Signal Received Power (RSRP) and Reference Signal Received Quality (RSRQ), which are extracted from Minimized Drive Test (MDT) data, to detect cell outage.

1.2 Statement of the Problem

ethio telecom is one of the largest telecom service providers in Africa and the only in Ethiopia with about 42 million mobile service subscribers as of July, 2019[7]. Its vision is to become a world-class telecom service provider. Recently, the Ethiopian government has decided to liberalize the sector and invited other operators to join the business. Thus, it is believed that other national and/or international telecom operators will join the service market in the near future. To become competent in the sector, ethio telecom is working hard towards tackling it challenges and there by improve its Quality of Service. One of the main challenges that ethio telecom faces is the issue related with its Cellular Network Management System (CNMS), one of the functions of which is cell outage detection and compensation, that needs proper attention and urgent solution. As the company is currently using traditional methods of Network Management System (NMS), automating its system, especially that of its cell outage detection and compensation, can be a viable solution to improve its service quality. It is, thus, with this intent that the researcher is motivated and decided to work on the issue in this thesis.

To improve service quality, cell outages must be managed properly. Cell Outage Management (COM) is a process by which cell outages are detected and compensation is made as required. It has two functions: Cell Outage Detection (COD) and Cell Outage Compensation (COC). Cell outage is a total/partial loss of radio service in a service area. While cell outage detection is a process of identifying outage cells, cell outage compensation is fixing the outage problem by using possible mechanisms to preserve services to the network's users.

COD is one of the major problems in many cellular networks and it is a major cause for reduction in network quality, coverage and efficiency, which will ultimately affect service quality. For example, the data generated from the monthly ethio telecom mobile network performance report issued in April, 2019, showed that 85.2% of the total numbers of sites in Huawei and ZTE circle were registered as outage. Given the fact that ethio telecom uses a traditional outage detection method, which may usually lead towards delayed detection and compensation [4], it is obvious

that the outage could have had greater impact on the Company's service quality and revenue. From this, we can understand that timely detection and recovery of cell outage is the best strategy to improve the network coverage and service quality of the operator. This can be achieved only if the outage detection of the company is automated so as to shorten the time needed for the detection and compensation of cell outage, which is a major problem in traditional detection. The automation can facilitate the quick start of the compensation process.

Most of the studies such as [8][9][10][11] conducted in the area were done by using MDT and KPI data generated from a simulated environment where the simulated data is different from the real one, which can only be obtained from telecom operators. Consequently, results may vary from what someone could have achieved if the studies were made based on the real data that is obtained from the operator. This can help in avoiding the bias in capturing the actual network characteristics thereby making outage detection easier. In another study, outage detection was made by using real ethio telecom UMTS inHO data and achieved 89% accuracy[4]. Nonetheless, the study had gaps on implementation as the detection was not made based on the assumption. The researcher started by assuming neighbor cells behavior/anomaly score value to detect outage cell but he, finally, detected the target cell status (outage/normal cell) based on the anomaly score of the cell itself. However, the author recommended the inclusion of neighbor's cell list together with incoming HO statistical data to achieve a higher level of accuracy than what was actually attained by the study. Thus, it is with this intention that this thesis is designed to detect cell outage by applying machine learning anomaly detection technique on ethio telecom UMTS network by considering neighbor cells performance data (traffic data, in addition to inHO) to detect the outage of the target cell automatically in a pre-set time interval.

1.3 Objectives of the Study

1.3.1 General Objective

The ultimate objective of this study is to detect cell outage through the application of density based anomaly detection algorithms (K-NNAD and LOFAD) and machine leaning technique by using real data of cell-level statistics of ethio telecom (i.e., inHO and Traffic data) report of neighbor cell on selected sample sites and propose its applicability over the entire networks of the operator.

1.3.2 Specific Objectives

In order to achieve its general objective, the study tried to address the following specific objectives. To:

- extract a data set that serves as an input in the modelling process by gathering and preprocessing the data set;
- select features through the application of exploratory data analysis technique;
- define the reference data to be used in the anomaly detection model;
- select algorithms to set an anomaly detection rule that can differentiate between normal and abnormal measurements by computing threshold based dissimilarity measure;
- profile the normal working behavior of the network by using fault free measurements;
- evaluate the performance of the generated model by using the test data extracted from problematic network state; and
- draw conclusions and recommendations based on the findings of the study.

1.4 Scope and Limitation

The study is conducted only on selected UMTS network sites of ethio telecom in Bole Sub-city in the city of Addis Ababa, Ethiopia, making its geographic coverage very much limited. In addition, the study focused only on cell outage detection by using density based anomaly detection and machine learning techniques on selected sample cells. Thus, it does not include the other part of

cell outage management techniques of COM, that is cell outage compensation. In addition, the technique is not capable of identifying the root causes of cell outage as well as the model may not apply on the other types of Radio Access Technology (RAT).

1.5 Significance of the Study

This study has a number of importance. First, it will give an experiment-based feedback and recommendation on the applicability of machine learning and anomaly detection techniques on cell outage detection in cellular network. Second, it will contribute to the stock of literatures on the area of cell outage management and detection. Last but not least, the findings of this study will serve other researchers as a spring board by indicating the possible gaps in the area.

1.6 Literature Review

A number of researches had been done in the area of COD and COC. The discussion below gives a brief review of what some of these studies are along with their respective methods, findings, strengths and weakness.

In [5] COM framework was developed for heterogeneous networks by using different detection methods for both macro and small cells. The study developed a framework for both COD and COC by applying machine learning and anomaly detection method, which could be taken as its strength. The COD of macro cell was made by using large amount of MDT to learn about the normal network scenario. MDT is a user level data that is collected by active user mobile with parameters like RSRP and RSRQ for both a serving and neighbor cell data [2]. The minimize complexity, the MDT report was reduced by using Multi-Dimensional Scaling (MDS) to a manageable size, thereby shortening computation time. MDS reduces the data by retaining the relevant information only. The authors used two algorithms to define anomaly detection rules to differentiate normal and abnormal MDT measurement by measuring dissimilarity and comparing with threshold value. They also compared the performance of the two anomaly detecting algorithms; namely, k-NN and

LOF based anomaly detector, to improve outage detection accuracy. They integrated COD algorithm to COC algorithm to detect and fix the outage area by optimizing the capacity and coverage by adjusting the antenna gain and transmission power of the surrounding BSs.

Paper [12], worked on COD framework that adopts a model driven approach that made use of mobile terminal-assisted data collection solution based on MDT functionality. They first collected User Equipment (UE) reported MDT sample measurements from a fault free and Radio Link Failure (RLF) working scenario to develop reference database and implemented MDS to transform it into a low-dimensional embedding space. The study used two kind of anomaly detection algorithms; namely, LOF-based detector and One-Class Support Vector Machine-based (OCSVM) detector together with the embedded measurements to learn the normal network behavior. The two learning algorithms were, then, compared and evaluated. Moreover, the geo-location associated with each measurement of COD framework was used to localize the position of the faulty cell.

In [13], integrated detection and diagnosis framework that can perform fault classification based on statistical analysis and find the most probable root cause of problems was addressed. For fault detection, monitored radio measurements and other KPIs were compared to their usual behavior captured automatically by profiles without threshold and manual setting. But, diagnosis was dependent on previous fault cases. The abnormality level was used to calculate the likelihood of a failure case and the target with largest likelihood value is considered to be the diagnosed failure.

In [4], considered a real ethio Telecom UMTS mobile network scenario and used inHO statistical data for the analysis of COD with LOF data mining algorithm. By considering the amount of the successful inHO from neighboring cells. the outage cell in a given time can be detect. The strength of this paper is that the algorithm detects a cell outage even when measurements from the considered cell are not available. Finally, he showed the experiment result and numerical performance analysis in the two algorithms and concluded that Modified LOF algorithm had a better performance than the original LOF detection algorithm.

All the studies in [5][8][9][11][12][14], a user–level data like RSRP and RSRQ were used while [15][4] used HO data to detect cell outages.. As can be seen from the discussion above, the studies employed different algorithms along with various dimension reduction methods and came up with results of varying accuracy. It was also learnt from these studies that they used different techniques of feature selection, parameter setting and model validation. However, using K-NN algorithm coupled with MDS can produce an even more accurate prediction than those used in the studies above. It is with this notion in mind that this study intends to implement a cell level data called cell inHO and traffic for detecting the outage cell using machine learning and anomaly detection method.

1.7 Methods of the Study

Of the existing methods of machine learning, this study used Cross-Industry Standard Process (CRISP-DM) which is an open standard process model that describes common approaches for data mining or machine learning. This method is chosen because it is most widely used in machine learning making it a well proven methodology in the area [16]. The method has a well-established framework that contains six phases, each with different tasks, that are interrelated to each other. Figure 1-1 shows the different phases of the model and in section IV the discussion gives a highlight on each of the phases.

To conduct this research, a four months i.e., from June 2019 – September 2019, inHO and traffic data in an hourly based, was collected from 80 sites in Addis Ababa UMTS network and analyzed by using Matlab, MS-office software and the results of which are then interpreted and presented in the form of tables, figures, charts and diagrams.

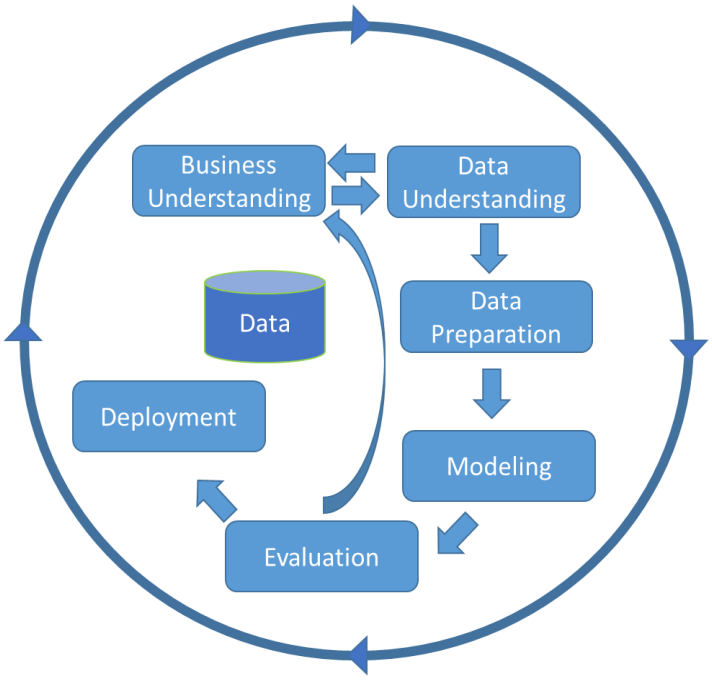

Figure 1-1 CRISP-DM Process Framework.

1.8 Thesis Organization

This thesis is organized in six chapters. The first chapter presents an introduction to the study including background of the study, statement of the problem, objectives of the study, scope and limitation of the study and significance of the study along with a review of related literatures. The second chapter gives an overview of UMTS network, application of SON with more emphasis on self-healing function of cell outage detection. The third chapter discusses the techniques of machine learning on SON and their applicability, anomaly detection methods on COD and CRISP-DM. In the fourth chapter, the activities of model building such as data preparation, setting detection rules, and model validation will be presented. Chapter five gives the discussion of results and the last chapter presents the conclusions and recommendations of the study.

2 Network Management and Self-Organizing Network in UMTS

This part attempts to present a highlight on existing technologies that are related with this thesis. It specifically addresses the issues related with the UMTS network architecture, Network Management System and the history, functions and implementation of SON.

2.1 Overview of UMTS Mobile Network

UMTS is the third generation (3G) of mobile cellular systems that was developed by 3rd Generation Partnership Project (3GPP). The rationale behind the development of UMTS is to give higher data rate and good voice services than its predecessor, the second generation (2G) Global System for Mobile Communications (GSM). As a network architecture, which is an infrastructure that delivers services between two end points, UMTS has three functional elements; namely, User Equipment (UE), Radio Access Network (RAN) and Core Network (CN). Figure 2-1(below) shows the network architecture of UMTS, each of these elements has different functions the outlines of which are presented below[17].

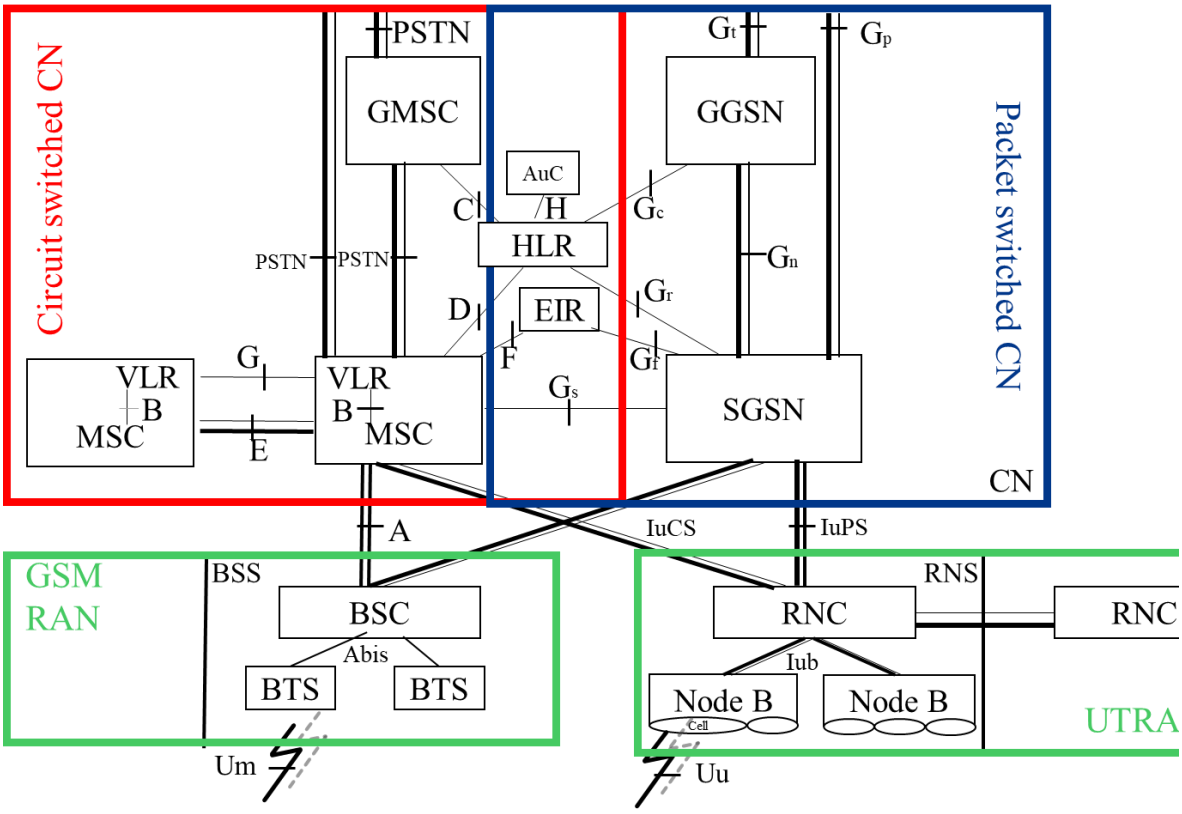

Figure 2-1 UMTS Network Architecture[17].

2.1.1 User Equipment

UE is a link between a user and radio air interface. It connects user to Node B and has two parts. The first is the Mobile Equipment (ME) or phone, which is an interface for radio services. The second, on the other hand, is UMTS Subscriber Identity Module (USIM), which is a smart card that stores essential subscriber data important to identify and authenticate the user who is given an access of the network[18].

2.1.2 Radio Access Network

RAN is part of a network infrastructure that helps users to access the network. There are different access technologies. Among others, GSM system uses Time Division Multiple Access (TDMA)

technology which lets a user to access network on the bases of time. On the other hand, the UMTS system uses Wideband Code Division Multiple Access (WCDMA) technology to let users access the network on the bases of code. WCDMA adds two access technologies; namely High Speed Downlink Packet Access (HSDPA) and High Speed Uplink Packet Access (HSUPA), to increase the data rate in order to improve Quality of Service (QoS)[19]. HSDPA supports high speed downlink data rate up to 10 Mbps (theoretically 14.4 Mbps) and HSUPA has a peak data rate increase up to 2 Mbps (theoretically 5.8 Mbps).

RAN is responsible for radio related functionalities and has two main components; namely, Node B and Radio Network Controller (RNC).

Node B

Node B (Base Station) is the transmitter and receiver unit between the UE and the rest of the network. It performs some basic radio resource management functions and its main functions are channel coding, interleaving, rate adaptation and spreading or dispreading[19].

Radio Network Controller

RNC is a network element that is responsible for controlling the radio access network or UMTS Terrestrial RAN (UTRAN). It controls, allocates and checks radio resources like time, frequency and codes. RAN also controls parameters like code allocation, admission control, load control, and controls handover processes.

2.1.3 Core Network

CN is the main backbone network of the entire network infrastructure. It connects the user with either Public Switch Telephone Network (PSTN) system or Internet system, as per the users' service request. The main elements of CN are Mobile Service Switching Center (MSC), Home Location Register (HLR), Visitor Location Register (VLR), Authentication Center (AuC), Equipment Identity Register (EIR), Getaway MSC (GMSC), Serving GPRS Support Node

(SGSN) and, Gateway GPRS Support Node (GGSN)[19]. There are also a number of interfaces that connect the different network elements.

Basically, CN contains the circuit switched domain and packet switched domain, which are used to access voice and packet data services, respectively.

Circuit Switched domain networks, such as Integrated Service Digital Network (ISDN) and Public Switched Telephone Network (PSTN), provide connection for voice call services in a mobile system. MSC is an element for voice call which is responsible for call establishment, call management and transmit the call to other network. It uses a number of data base system centers including HLR, VLR, AUC and EIR, each of which has different functions.

Packet Switching network provides connection for data services through SGSN, which controls data traffic and GGSN, which helps to access external packet data network services.

2.2 Mobile Network Management

A mobile network is consisted of base station networks, core network and a transmission networks that work together to provide the required service. To manage these networks, the management system uses Element Management Systems (EMS) for element level such as router and switch and Network Management System (NMS) to monitor the entire network in an integrated manner. The main functions of NMS are network configuration management, performance management, fault management and security management. Figure 2-2 shows the network management system architecture. The major function of each of these components is presented in the discussions below.

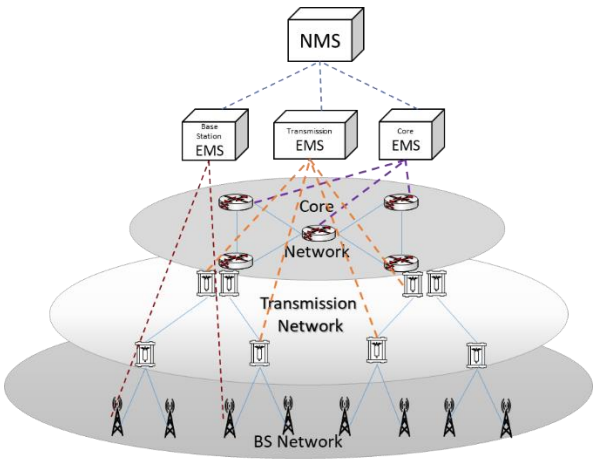

Figure 2-2 UMTS Network Management System Architecture [17]**.**

2.2.1 Configuration Management

Configuration management is concerned with monitoring system configuration data, such as neighbor cell list data, base station configuration data and element configuration data and any change that takes place on the network at any time. A proper configuration management strategy involves tracking all changes made to the network's hardware and software[17].

2.2.2 Performance Management

Performance management monitors the network performance like throughput, network response time, link utilization and alerts the network administrator when the performance parameters fall above or below the given performance threshold.

2.2.3 Fault Management

Fault management monitors the status of the network and reports the faults that occur in it such cell outage and equipment failure. This can be established by monitoring different alarms generated from the network elements like KPI, HO and MDT data or from the observed abnormal behavior of the network[17].

2.2.4 Security Management

Security management controls the access for network resources by using different security management functions, such as managing network authentication, authorization, and auditing to allow/deny access for legal/illegal users[17].

The above functions are served by both the traditional and SON network management systems. However, the traditional network management system has the following limitations as was learnt from[17].

It requires a lot of labor, hence, high cost, while network planning and to define neighbor relationship configuration;

BTs activation requires manual software loading and data configuration mistakes might be quite often while neighbor configuration is made manually; and

Cell outage detection is made manually causing delays in recovering outages.

However, these weaknesses of the traditional system can be minimized by deploying SON and automating it.

2.3 Self-Organizing Networks

The rapid growth of smart equipment users and the large number of applications that smart equipment has, have boosted the traffic volume of mobile data services. To meet the required service demand, multi RAT, Multi Band and Multi-Layer network are being used to increase network capacity[3].

Such networks are so complicated that they, for example, cause network management impact making the network operation difficult [20]. These days' mobile network elements need network parameters to be configured. Such a situation has significant associated operational cost. In addition, the existing traditional operation is time consuming and potentially error prone[3].

For each operator, a key challenge is guaranteeing high efficiency operation with low cost. To this end, recent technologies attempt to address the growing data capacity with minimum operational cost by deploying SON, which gives it a high demand and value network management. This permits reduction in CAPEX and OPEX during deployment and continuous operation, which is one of the most important assets of SON.

The concept of SON technology has been incorporated in 3GPP Release 8 with some features and then, were expanded in 3GPP Release 9 and 3GPP release 10 [6]. The first SON release included automatic inventory, software download, neighbor relation and physical cell ID assignment functionalities. The second 3GPP Release 9 added optimization functions such as mobility robustness optimization, random access channel optimization, load balancing optimization and inter-cell interference coordination functions to what existed in the earlier version. Likewise, 3GPP Release 10 includes coverage and capacity optimization, enhanced inter-cell interference coordination, cell outage detection and compensation, self-healing, minimization of drive testing and energy saving functionalities of the SON, giving it best performance.

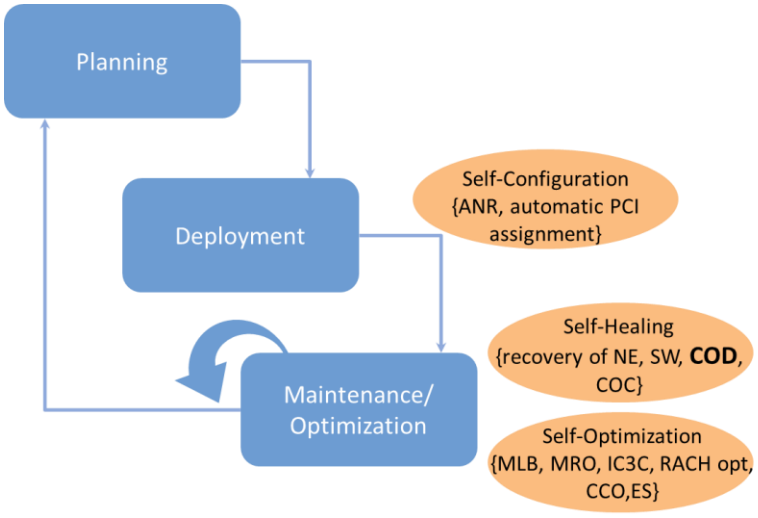

Figure 2-3 Self-Organizing Networks [3].

Generally, SON provides network intelligence, automation and network management features in order to automate the configuration and optimization of wireless networks to adapt them to varying radio channel conditions [6]. By doing so, it minimizes cost, improves network performance and flexibility.

SON can be deployed in three different ways[3]. These are:

- Deployment of Distributed SON at network element level;
- Deployment of Centralized SON at management center; and
- Deployment of Hybrid SON both at element level and management center.

Despite the presence of various functionalities of SON, this paper focuses only on cell outage detection, which is one of the components of self-healing and a feature of 3GPP Release 10.

2.3.1 Self-healing

Self-healing is the automated SON functionality for recovery of NE, cell outage detection and cell outage compensation. It triggers the appropriate recovery action to solve the fault/outage automatically [21]. It has two parts; the monitoring part and the healing process part. The first part monitors alarms in a real-time to activate the self-healing actions and the second part triggers healing process to solve the fault automatically.

In 3GPP Release 10 TS 32.541, the concept of Self-healing and its requirements are clearly presented. One of these is cell outage detection, which is identifying a sleeping, faulty or outage cell, indicating the presence of a partial/total loss of radio service in an area. The use case of Self-healing functions is defined for cell outage detection, recovery and compensation.

2.3.2 Cell Outage Detection

The most critical area in wireless cellular systems is RAN, and it is prone to fault and failure [3]. Cell outage is one of the problem in RAN, and it needs proper management techniques to detect and compensate the outage area. Cell outage detection is the process of detecting an outage cell

through the monitor performance indicators, which are compared against threshold and profile. Such a task needs information analysis to determine the cell status precisely.

There are different failure scenarios that happen in cellular network equipment. A case of complete Node B failure is discussed in [22], as a situation where by OAM will be unable to communicate with the Node B to decide the cell status. Lack of communication may be a symptom of failure on the OAM backhaul rather than an indication that the site is down. This type of symptom needs other evidence to learn the nature of the problem. Such a problem can be solved by analyzing the core network metrics to determine the specific status of the cell because if the cell is active, it will interact with the core network. Such detection needs evidence. This kind of cell outage detection scenario is the most challenging and needs anomaly statistics rather than an alarm.

Profiling network behavior is an acceptable mechanism for evidence based detection. Such a detection can be achieved by collecting data over a period of time and there by building a statistical picture of the expected performance. The profile may vary for a given time of day, weekday, or weekend. If the stats collected for a cell shows significant deviation from the normal profile of that cell, then, it is more likely that a latent fault has occurred.

3 Machine Learning and Anomaly Detection

3.1 Introduction

Anomaly detection is a technique that is used to identify something that has noticeable difference from the usual pattern of occurrence. It is one of the most commonly used techniques in machine learning environments[10]. Machine learning is an application of artificial intelligence that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. It focuses on the development of computer programs that are capable of accessing data and use it to learn for themselves. The primary aim of machine learning is to allow computers learn automatically without humans' intervention and take actions accordingly[10]. A number of reasons instigate the need for machine learning. Among others, one is the ever increasing demand to deal with complex tasks which are being performed by humans or even those that are beyond humans' capabilities. The main objective of machine learning is to improve the performance of a specific set of tasks by creating models that help to find patterns through learning algorithms [3]. Although a recently emerged technology, the utilization of machine learning in different disciplines has been expanding rapidly as the technology really simplifies the way things are done.

The telecom industry is one of the areas that can make use of machine learning for different purposes and in different scenarios. For example, it can be implemented in cellular network systems to handle network problems[23]. As such, the technology can be employed in anomaly

detection methods. There are many anomaly detection algorithms, specially designed to detect anomalous behavior, that use different techniques like spectral theory, Information theory, machine learning, statistics, and the like.

3.2 Machine Learning-based Anomaly Detection Techniques

By considering the type of data they use anomaly detection can be classified into three as Supervised, Unsupervised and Semi-supervised.

3.2.1 Supervised Learning

Supervised anomaly detection is a Machine learning technique in which training is made by using labeled data set (input vector and a desired output) which includes both normal and anomaly classes. The dataset has rows and columns, each row represents a single instance and the column represents features. The data set can be divided into two sets; namely training set and test set. While the training set is used to train the model, the test set is used to evaluate the prediction accuracy of the model. In this case, the model can be assumed as predictive model for both normal and anomalous classes. It functions by comparing the unseen data instance with the model to determine to which class that the unseen data instance belongs. This technique is useful when the network management function needs to address required estimation, prediction and classification of variable [24].

The main application of supervised learning technique is on classification or regression problem. There are a number of supervised learning algorithms including those that are popular in SON application such as K-NN, SVM, Naïve Bayesian, Artificial Neural Network, Decision Tree and Hidden Markov Model [3].

According to [16], two issues must be given due attention while applying supervised learning. The first issue is imbalanced data distribution, which comes from the existence of fewer number of anomalous instances from the normal instance. The second, on the other hand, is lack of accurate

and representative label, especially for the anomalous class. To solve such challenges, the study recommends injection of artificial anomalies in a normal data set.

3.2.2 Unsupervised Learning

Unsupervised anomaly detection is a machine learning technique that needs no training data set. As the majority of data sets are assumed normal, this technique appears being the most widely applicable one. The assumption is that the normal instance is far more frequent than anomalies in the test data. If the assumption is not true, such techniques may suffer from high false alarm rate [3]. In unsupervised anomaly detection, the algorithm profiles (model the most behavior) the input data that can be used to predict future input. Unsupervised learning is mainly implemented in network management to identify anomalous behavior, or reorganize pattern such as in self-healing, some examples are [25][26].

3.2.3 Semi-supervised Learning

Semi-supervised anomaly detection is a machine learning technique that uses training data that contain only normal instances to learn from interaction to achieve a certain goal[10]. To recognize the anomalous behavior, it measures the deviation of instances from normal training points. In semi-supervised techniques, normal data can be used as a training data to produce the model. The model will, then, be used to predict a test data, which contains both normal and abnormal instances. Semi-supervised technique can be used to address network management functions that require network parameter control. Some examples are[4] [9][27].

In anomaly detection technique, the anomalies are reported in two types as scores and labels.

Scores

Scoring techniques assign a value for each test instance based on the degree of anomaly is considered as anomaly detection score. Anomaly detection score is used to rank the instances based on the score value, which will, then, used to put a threshold to select anomalies based on the nature of the problem.

Labels

Labeling technique labels test instances as normal and abnormal based on the anomaly score. Putting threshold by considering some parameters that help to select the most appropriate class is important to classify test instances in the two category.

3.3 Anomaly Detection Methods

Anomaly detection uses different methods as outlined below, of which the first two are considered in this paper.

- Classification-based anomaly detection method;
- Nearest neighbor based anomaly detection method;
- Clustering based anomaly detection method;
- Statistical based anomaly detection method;
- Information theoretic method;
- Spectral anomaly detection method.

3.3.1 Classification-based Anomaly Detection Technique

Classification-based anomaly detection technique learns the classifier based on the given labeled (normal and anomalous) data at the training phase and classify the test instance by using the classifier in testing phase. The assumption on classification-based anomaly detection is that a classifier can be learnt in the given feature space [24].

Based on the available labels in the training phase, classification-based anomaly detection method is classified as one-class or multi-class anomaly detection techniques.

One-class Classification-based Anomaly Detection Technique

One-class classification-based anomaly detection technique assumes that all training instances have only one class label. It can be considered as a binary classification problem. The normal instance in the training stage creates distinguished boundary of normality. Thus, an instance that

lies out of this boundary will be considered as anomalous in the test phase. See Figure 3-1for illustration.

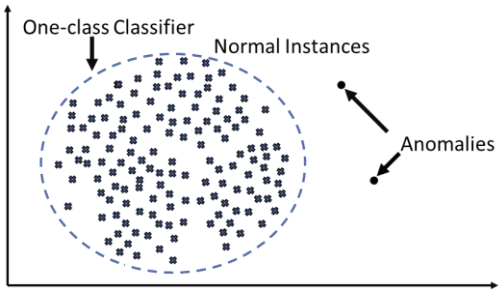

Figure 3-1 One-class Anomaly Detection[10]**.**

Multi-class Classification-based Anomaly Detection Technique

Multi-class classification-based anomaly detection technique works on the environment of two or more normal labeled classes [24], in which the classifier learns all normal classes against the rest of the class. To decide the normalness, it tests for each class inside a confidence level and if none of the classifier are confident in classifying the test instance as normal, the instance as declared to be anomalous, see Figure 3-2 for illustration.

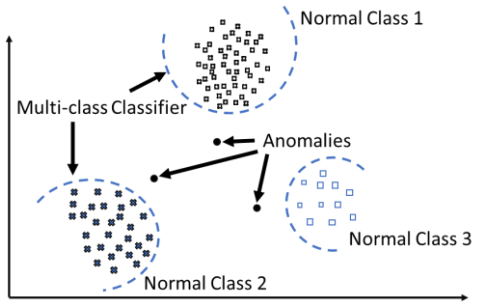

Figure 3-2 Multi-class anomaly Detection[10].

3.3.2 Nearest Neighbor-based anomaly Detection Method

The concept of Nearest neighbor analysis lies in the assumption that "normal data instances occur in dense neighborhoods, while anomalies occur far from their closest neighbors" [4]. The method needs a distance or similarity measure between two data instances. In this case, Euclidian distance is the most widely used distance measurement but other distance metrics like mahalanobis distance can also be used [27]. This method is further classified into two categories, as Distance to k^{th} Nearest Neighbor and Relative Density, based on the anomaly score assignment.

Distance to k^{th} Nearest Neighbor

Distance to k^{th} Nearest Neighbor measures distance of data instance to its K^{th} nearest neighbor to define the anomaly score. It also modified as sum of k nearest neighbor's distance of data instance to assign the anomaly score [9] .

Basically, there are different ways to define whether the given instance is anomalous or not in Distance to k^{th} Nearest Neighbor. In [5], the authors applied a threshold on anomaly score. Accordingly, those anomaly scores that are greater than the threshold value are taken as the anomalies. In [9] , on the other hand, takes with large anomaly score as anomalies.

Relative Density

Density-based anomaly detection technique estimates density of each data instance to determine anomalies. In this case, while an instance that lies in a dense neighborhood is declared as normal, those that lie in the low density are declared as anomalous. Density based technique performs poorly if the data has regions of varying density [24]. To address such an issue, one possible approach is weigh, in which the relative weight of neighboring dense neighborhoods has been developed.

Nowadays machine learning applications are widely implemented in cellular network system. Many researches have been conducted in the self-healing function of SON for detecting abnormal network behavior, fault classification, and cell outage management. Specifically, nearest neighbor based anomaly detection methods are implemented in [5] propose cell outage detection and compensation framework using machine learning method. In order to detect the sleeping cell automatically, a density based anomaly detection method is used in[8].

With the purpose of getting a better performance for cell outage detection, this thesis implements the two types of Nearest neighbor based anomaly detection methods discussed above. Two state of the art algorithms; namely, K-NNAD and LOFAD are compared and the one that has better prediction accuracy was selected for this purpose.

3.4 CRISP-DM Machine Learning Method

This study uses Cross-Industry Standard Process (CRISP-DM) method, which is an open standard process model that describes common approaches for data mining or machine learning. It consisted of six different but interrelated stages. The essence and major tasks of each stage is briefly discussed below.

3.4.1 Business Understanding

Business understanding is the first stage in CRISP-DM. It helps to see what someone wants to do in business perspective and clearly describes the relevance of the proposed project. With regard to this thesis, this stage helps to explain how important that the project might be to ethio telecom Cellular Network Management System. As such, this stage lays the base for a proper data mining which aligns with the problem statement in later stage [16].

3.4.2 Data Understanding

In data understanding stage, the available data sources will be identified out of which relevant data will be selected (collected). If data is obtained from different sources, it might be necessary to integrate these data for further exploration. This will help to select relevant attributes and decide the appropriate format for analysis[16].

To properly understand the data, it might be important to see the pattern and existing relationships between attributes, evaluate the information content of each attribute and test the stated assumptions. This stage will provide inputs to the next stage by producing substantial information about the hidden pattern.

3.4.3 Data Preparation

Data preparation involves the selection, cleaning, transformation and reduction of data from the row data.

- In the selection stage, data will be collected, selected and integrated from the row data to decide what should be used for analysis.
- In data cleaning, add new values missing values or remove their rows to improve the quality of the dataset and attain results with high accuracy.
- The transformation stage involves data normalization, which helps to maintain data uniformity as well as creating new attributes.

- In the reduction stage, dimension reduction techniques like Principal Component Analysis (PCA), MDS or other techniques will be selected to minimize the dimension of the predictors or attributes by maintaining the information and the structure with the purpose of reducing the computation complexity and there by processing time.

This process helps to generate the final dataset that will be used while fitting the dataset with required machine learning model.

3.4.4 Modelling

At this stage, the selection of the appropriate model from the available machine learning models that relates with the problem will be made. In this study, the selection can be made by using different techniques like comparing different models that work for the same problem but with different algorithms or try for different parameters in the training phase to achieve best efficiency.

3.4.5 Evaluation

In the evaluation stage, the model will be tested with unseen data to evaluate how well the model performs. The selection of different evaluation methods is based on the existing problem.

This study evaluates the performance of the model after post processing stage, which is a process of distributing the predicted values and measuring how the outage score deviates from the reference threshold for each site. This can be done by using the evaluation methods used in machine learning classification problems such as measuring the prediction accuracy, overall accuracy, precision, recall and area under ROC curve (AUC-ROC). Below is shown the evaluation metrics used in the study.

Confusion matrix

Table 3-1 Confusion matrix

		Predicted	
		Class 1	**Class 2**
Actual	Class 1	True Positive (TP)	**False Negative (FN)**
	Class 2	**False Positive (FP)**	**True Negative (TN)**

$$\textbf{\textit{Overall accuracy}}= \frac{TP+TN}{TP+FP+FN+TN} \quad (3\text{-}1)$$

$$\textbf{\textit{Precision}}= \frac{TP}{TP+FP} \quad (3\text{-}2)$$

$$\textbf{\textit{Recall}}= \frac{TP}{TP+FN} \quad (3\text{-}3)$$

4 Experimentation

As mentioned in section 1.7, this study followed CRISP-DM machine learning approach to develop a cell outage detection model for the selected area in UMTS mobile network using inHO and Traffic as input. It was also indicated that the study used two of the algorithms in Nearest Neighbor based Anomaly detection method; namely K-NN and LOF.

In the discussions below, the experimentation procedures are presented as per the six phases of the CRISP-DM approach.

4.1 Business Understanding

Business understanding, as the first stage of the CRISP-DM model, is used to explicitly describe the importance of the proposed project. As indicated in the objective of this thesis, the ultimate purpose of the study is to detect cell outage by applying machine learning and anomaly detection technique by using real data of cell-level statistics of ethio telecom (i.e., inHO and Traffic data) report of neighbor cell on selected sample sites and propose its applicability over the entire networks of the operator.

To this end, the automation of the NMS through the application of machine learning will serve as a pilot case to indicate that automation in the NMS can help in minimizing outage detection time and thereby compensation time leading to high network quality, which will in turn produce high customer experience and satisfaction. It will also contribute a lot to the company's effectiveness and efficiency in its resource management.

4.2 Data Understanding

Data understanding is the second stage in CRISP-DM. At this stage, the identification of the available data sources will be made from which those that are relevant to the study will be selected (collected). In this case study, the available data is the entire information obtained from OSS. From the existing information, selection was made based on relevance to the study. Accordingly, inHO and Traffic data, cell ID, site ID, location information and neighbor cell list for each cell were selected for the purpose of this study. These data were used to detect the abnormal behavior of a cell in the network of the selected area. The specific procedures of data understanding are experimented according to the pre-specified assumptions and presented as follows.

4.2.1 Data collection

In mobile cellular networks, huge amount of data is generated from different counters and analyzed for different purposes. KPI is one of the users of such data to describe the performance of the network. It is often used to detect failure of network equipment or cell outage. However, outage detection can also be made by using other data such as those related with traffic and user mobility, other than KPI[6]. In this study, mobility related data, i.e. inHO, and traffic are used, with the assumption that the analysis of such information obtained from neighbor cells can help us to detect the status of the target cell.

For services to continue uninterrupted, handover from one cell to another is the important feature of mobile cellular network. Handover is simply defined as changing the channel, i.e. the frequency, time slot, spreading code or combination of them, associated to the current connection while a call is in progress[10]. The most obvious cause for performing handover is user mobility, i.e. due to movement, a user can be served in another cell. It may, however, also be performed for other reasons such as system load balancing due to neighbor cell/site failure. Most of the time, when some failure or outage is occurred in one cell, there is a rapid increment in inHO to in the

neighboring cell. As this situation increases the number of users in a neighbor cell, the traffic may increase accordingly.

Generally, HOs have two categories; namely, inter-cell (inter-frequency) and intra-cell (inter-system) HOs[4]. In UMTS, HO has three categories. These are Hard Handover, Soft Handover and Softer Handover. The different types of air interface measurements in UMTS network are presented below.

- ***Hard Handover*:** it occurs when the handover is not noticeable to the user. In practice, hard handover that requires a change of the carrier frequency (inter-frequency handover) is always performed as hard handover.
- ***Softer Handover:*** such a handover occurs at Mobile Station (MS)/UE in the overlapping coverage of two adjacent sectors of a Node B. In this case, communication between BS and MS takes place concurrently via two air interface.
- ***Soft Handover:*** it occurs at a MS in the overlapping coverage of two different Node Bs. When such a handover happens, communication between MS and Node Bs is concurrently held via two interface channel from each base station separately.
- ***Inter-frequency hard handover:*** it occurs on high capacity Node B with several carriers to hand a mobile from one carrier frequency to the other.
- ***Inter-system handover:*** this kind of handover takes place between WCDMA FDD and different systems such as WCDMA TTD or GSM.
- ***Traffic Measurements:*** measurements on uplink traffic volume. It is a measurement of one cell.
- ***Adaptive Multi Ratio (AMR):*** it is adapting the AMR rates for coverage and capacity reasons. The AMR switching and rate control helps to optimize the usage of the air-interface.

To conduct this research, a inHO and traffic data was collected from the OSS of 80 sites in Addis Ababa UMTS network. These data are the most relevant data to predict the cell's/site's status. The

collection was made over a period of four months i.e., from June 2019 – September 2019, in an hourly base. Table 4-1 shows the selected predictors for the detection.

Table 4-1 List of Features

Features	Abbreviation (Defined in OSS center)
Inter-frequency Incoming Successful Hard Handover by circuit switched	HHO.SuccInterFreqIn.CS
Inter-frequency Incoming Attempts Hard Handover by circuit switched	HHO.AttInterFreqIn.CS
Inter-frequency Incoming Successful Hard Handover by packet switched	HHO.SuccInterFreqIn
Inter-frequency Incoming Attempts Hard Handover by packet switched	HHO.AttInterFreqIn
Factor of soft & softer HO for call (CS voice call + PS call)	Factor of soft&softer HO
Softer Handover Addition	Softer Handover Addition
Successful Soft Handover by adaptive modulation rate	SHO.AMR.Succ
Packet Switched Traffic in Kbit	PS Traffic (kbit)
Circuit Switched Traffic in Erl	CS Traffic (Erl)

4.2.2 Exploratory Data Analysis

Exploratory Data Analysis refers to the process of exploring for some of the main characteristics of the data to be used. It can be performed on raw original data, or post-processed data. This can

be done by employing a number of techniques including descriptive statistics, hypothesis testing, and visualization. In this case,

- descriptive statistics can be used to describe each feature (predictor) of the dataset.
- hypothesis testing can be employed in order to check whether the assumption required for model fitting is meet or not.
- visualization technique, such as bar plots and box plot, can be used to represent abstract ideas of the data better.

The results obtained from Exploratory Data Analysis usually indicate how the data should be prepared or pre-processed in the next stage.

To assess the statistical distribution of the different types of inHO, normal and outage time analysis was made and their description is presented below.

Normal Time Analysis

With the purpose of learning about the characteristics of cells, an attempt was made to explore the normal time scenario of the network in the selected cells. The following discussion presents the lessons learned from EDA in different cases.

Variation over 24 Hours

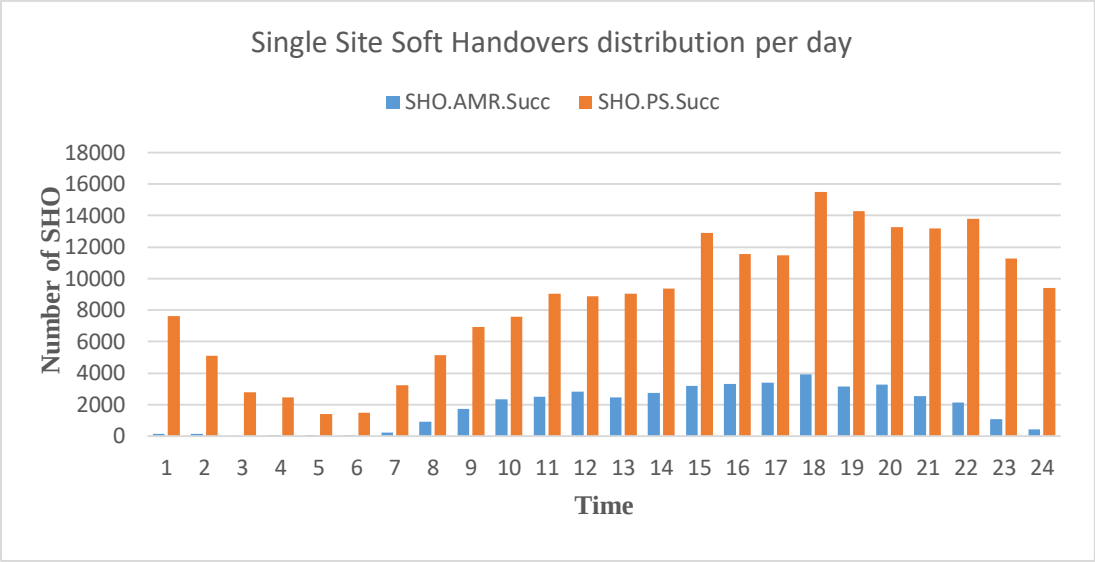

Figure 4-1 Different Soft Hos Distribution per day.

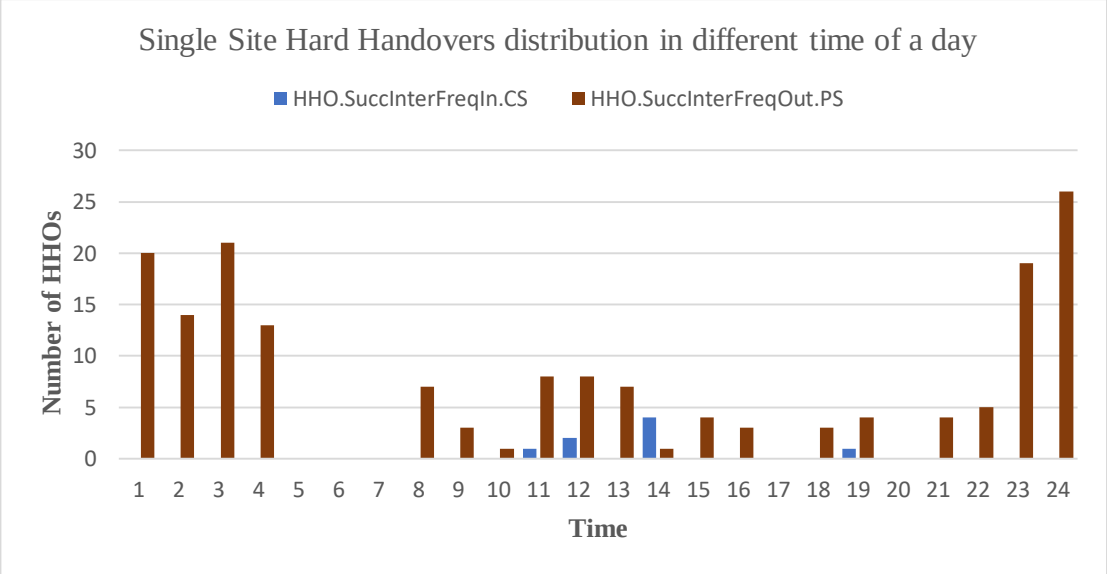

Figure 4-2 Different Hard Handovers distribution in different time of a day.

As Figure 4-1 and Figure 4-2 above depict, there exists variation in the types, number and hourly distribution of handover over the 24 hours of the observed date in a single cell. From this, we can learn that: handovers exist in all the 24 hours of a day; least handover is experienced from about 1:00 – 8:00, which are night hours; handovers gradually increase from about 8:00 to almost Midnight, with a little variation between hours, and the peak handover hour is 18:00.

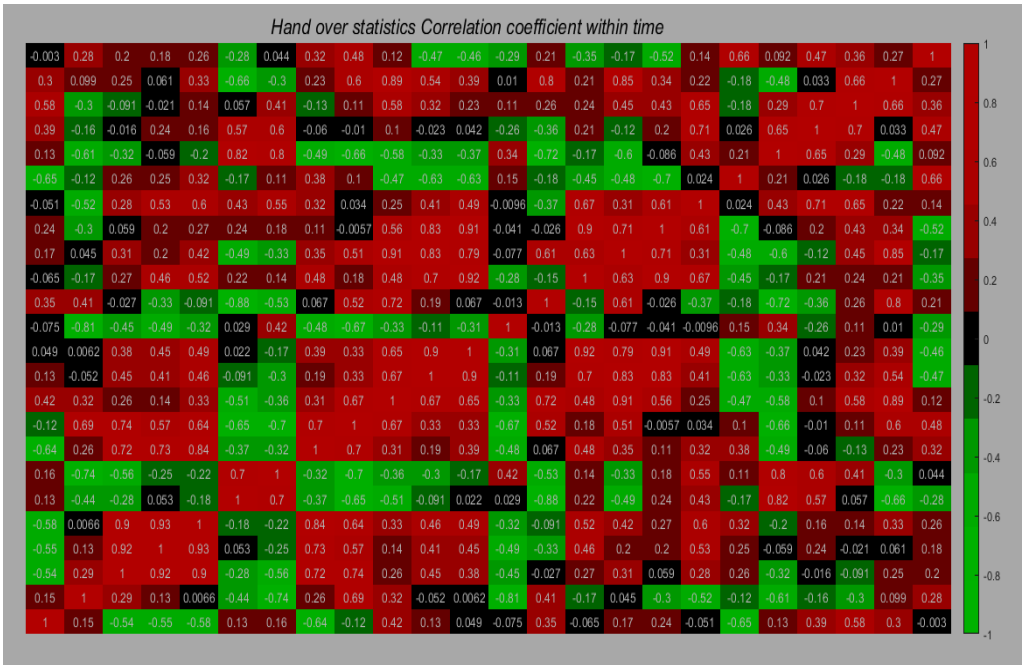

Figure 4-3 An Hour's Handover correlation coefficient within hours of a day.

Figure 4-3 shows the time correlation of a single hour hard handover in a day against the rest of the hours of that day. From this, an attempt was made to learn whether there exists any relationship between the observed Hard Handovers in an hour and the rest of the observed Hard Handovers. It was, thus, learnt that there is no uniform relationship between the two sets of observations as the correlation values vary significantly, with some having positive, others negative and still others zero values. This indicates that while some are positively/negatively correlated, others don't have any correlation at all.

Taking a lessons from the above two analyses, I have come to understand that due attention should be given to the time of the day while setting the parameters that will be used in the localization phase.

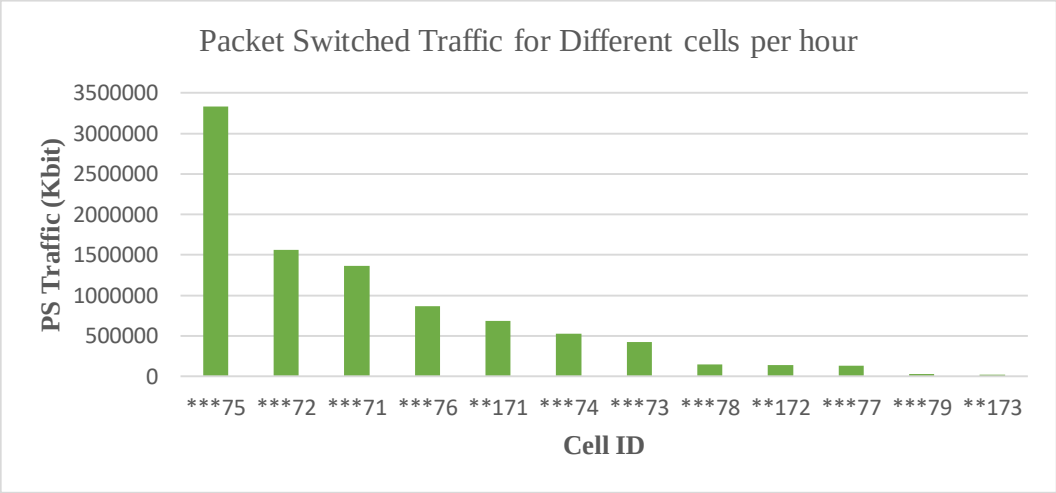

Figure 4-4 One hour Packet Switched Traffic of different cells in a single site.

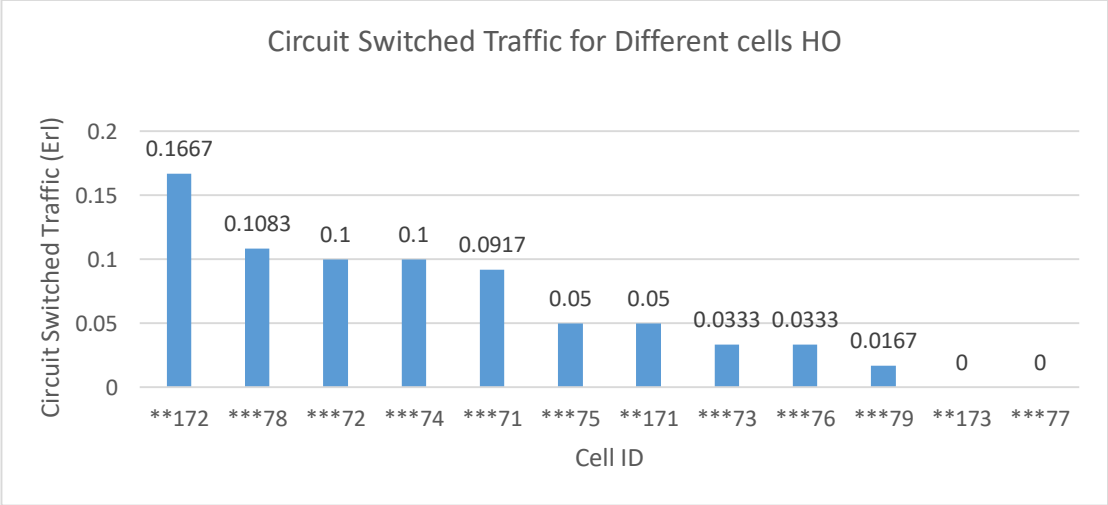

Figure 4-5 One hour Circuit Switched Traffic of different cells in a single site.

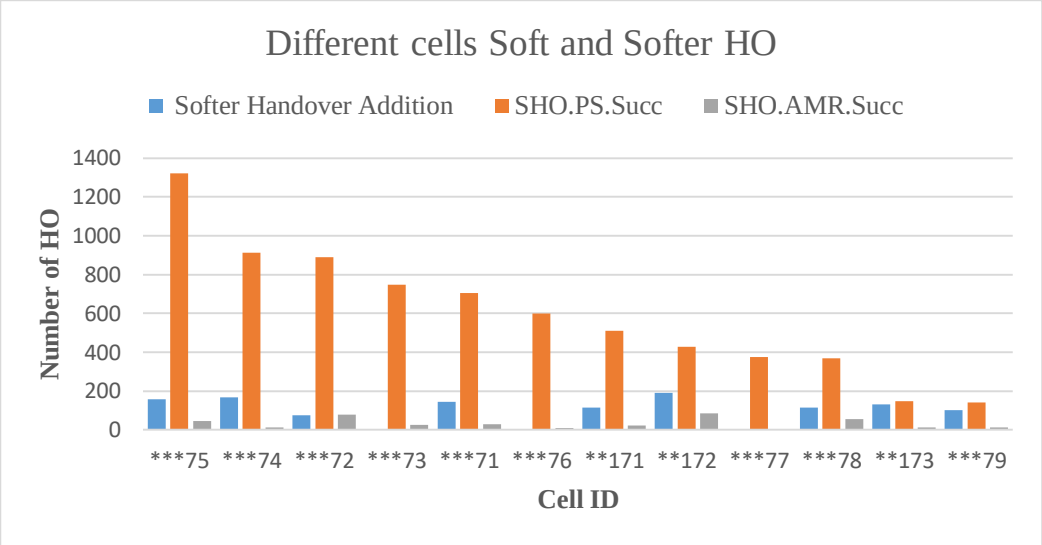

Figure 4-6 One hour Soft and Softer Hos of different cells in a single site.

The previous three figures show inHOs and traffics of an hour for different cells in a single site. As can be seen from the figures, different cells of a single site have different traffic density; while some are serving larger number of statistics, others serve only a few. Such a variation in the traffic density of the cells in a single site can be due to factors like number and types of users and sector/antenna direction. From this, we can learn that there can be variations in traffic density between different cells of the same site during the same hours of observation. Thus, it might seem feasible to categorize the cells in a single site into cells of high and low traffic density. Such an observation can help while labeling the training data.

Table 4-2 Single time Handover correlation coefficient within days.

	Mo	Tu	We	Th	Fr	Sa	Su
Mo	1	0.97994	0.95869	0.95003	0.979382	0.987522	0.951719
Tu	0.97994	1	0.962105	0.97075	0.985851	0.977069	0.954689
We	0.95869	0.962105	1	0.966626	0.959873	0.952856	0.94882
Th	0.95003	0.97075	0.966626	1	0.953923	0.932979	0.943422
Fr	0.979382	0.985851	0.959873	0.953923	1	0.985451	0.956578
Sa	0.987522	0.977069	0.952856	0.932979	0.985451	1	0.952329
Su	0.951719	0.954689	0.94882	0.943422	0.956578	0.952329	1

With the purpose of exploring whether there exists any similarity between the observed handovers of the same hours of each day over a week, an attempt was made to test the existing correlations. It was, thus, found, as presented in Table 4-2 above, that a single time specific handover has strong positive correlation with the observations of the same hour over all the rest of the days of a week. From this characteristics, we can conclude that we can use the same reference for single time within the days of a week.

Outage Time Analysis

This section tries to see the change in neighbor cell statistics when an outage occurs in another cell by considering the statistics from normal time as a reference. To illustrate the case, a site's normal and outage time scenarios of the same time series were considered (i.e. 12 AM). Accordingly, four neighboring sites (48 cells) of the outage site were selected and their inHOs and traffic data was used to check whether variations occur. The results from such a comparison can be useful in conducting the study confidently by taking the assumption that when a site or a cell outage occurs in the network, it creates statistical change in the neighboring cells or sites.

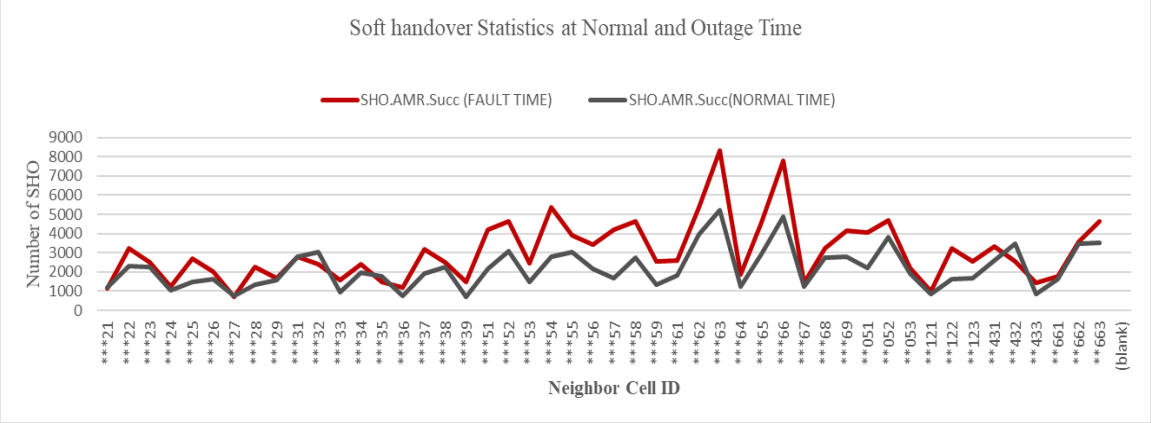

Figure 4-7 Soft handover statistics at normal and outage time.

Figure 4-7 shows successful soft handover statistics during the normal and outage time for each neighbor cells of the outage cell. From the above figure, we can observe that handover increases in almost all the neighboring cells/sites when outage occurs in a cell/site. The experiments on other types of handovers have also produced similar results.

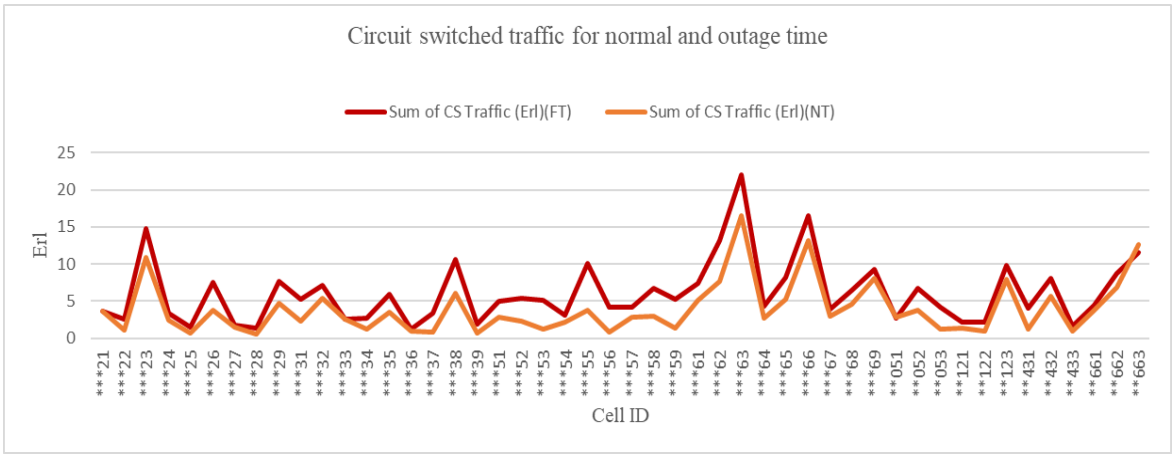

Figure 4-8 Circuit switched traffic for normal and outage time.

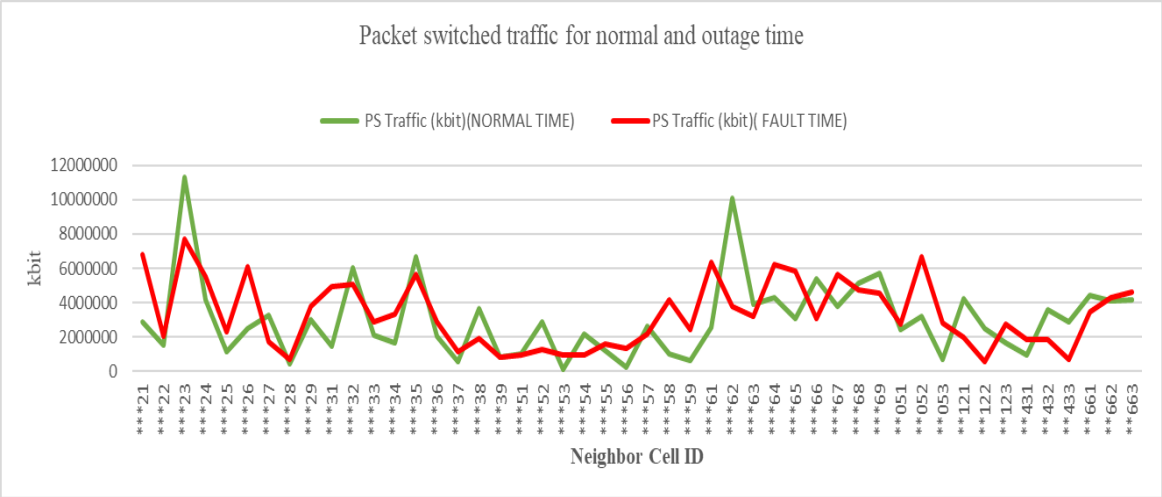

Figure 4-9 Packet switched traffic for normal and outage time.

The two line graphs, on Figure 4-8 and Figure 4-9 above, show the two types of service domains; namely, circuit switched and packet switched with an attempt to see the existing traffic variations between the normal and outage time for each neighbor cells of the outage cell.

From the first figure, we can observe that the CS traffic increases almost in all the neighbor cells at the time of outage. The increment of CS traffic is expected because there will be flow of users to the neighbor cells from the outage cell to maintain the service. Contrary to this, the traffic variation in the second graph appears being relatively low for many the cells. This can be due to the fact that the network gives priority for CS services than PS services. This characteristics of network is clearly seen in the second figure. From this, we can understand that PS traffic may even decrease in some cells while outage occurs in a neighbor cell. However, such a case may not happen when the outage occurs in off peak hours.

From the discussions made so far, we can conclude that at the time of outage, there will be a statistical change in the neighbor cells. Such statistical changes in the neighbor cells of the outage cell can easily be detected by using anomaly detection techniques. Such a detection method can

be used to improve the existing traditional cell outage detection method, which is based only on alarm observation, that is being used in the sector.

4.3 Data Preparation

One of the factors that affects model performance in machine learning is the input data. This stage is, therefore, useful in preparing the input data for future analysis. For the purpose of this study, both normal and outage time data were collected from SOS center and were pre-processed. The major tasks in the pre-processing stage of this study were filtering, feature extraction, normalization and dimensional scaling. All the stages were followed seriously to prepare the dataset for the training and test.

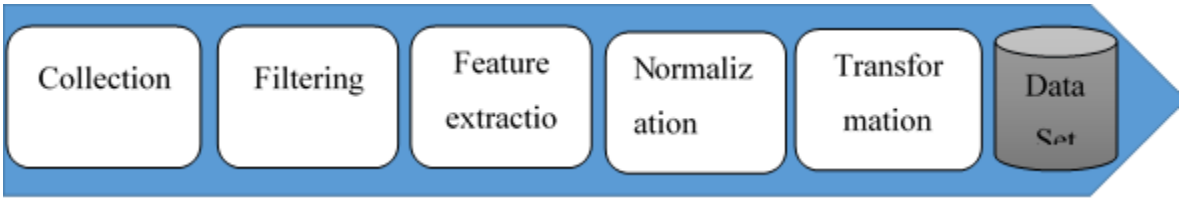

Figure 4-10 Data Preparation

In the discussion below, a brief description on how the last three stages were performed is presented. The first two stages were excluded from the discussion simply because they only involve the routine tasks of collecting and filtering the data from SOS.

4.3.1 Feature selection

After collecting the inHO and Traffic data from the selected sites and filtering it, the features were selected manually. The feature selection process was made based on the assumptions and using exploratory data analysis methods. The selected attributes are found to be highly representative of the behavior of the network and used for alarm generation in the network management system.

4.3.2 Normalization

While conducting feature selection, every feature may not necessarily have the same range. For example, in this paper, the various inHO data are measured in numbers having different ranges. Similarly, PS traffic data is measured in Mbts and CS traffic data is measured in Erlang with all having different ranges. So, to use such dataset in machine learning, normalization is needed as different features have different range of values, it affects the accuracy of the model[23]. Normalization is the technique for machine learning that is vital to transform the values of numeric columns in the dataset to a common scale often ranges between 0 and 1.

In this paper, Min-Max normalization method was used for the dataset X, which has N rows (entries) and M columns (feature). X [:, i] represents feature (single cell one hour measurements) and X [j,:] represents instance (single feature all cells measurements). Equation (4-1) below describes the Min-Max normalization formula.

$$x[:,i] = \frac{X[:,i] - \min(X[:,i])}{\max(X[:,i] - \min(X[:,i])} \quad (4\text{-}1)$$

Where:

x [:, i] = normalized value;

X [:, i] = single hour, all cell one time measurements;

max X [:, i] = maximum value in the column; and

min X [:, i] = minimum value in the column;

4.3.3 Dimensionality Reduction with Classical Multi-Dimensional Scaling

Dimensionality reduction is a process of converting high dimensional data to low dimensional data with the purpose of reducing the complexity of storing, processing and analyzing the dataset[5]. By applying dimensionality reduction, the dataset for this thesis was reduced from nine dimensional vector to three dimensions in the Euclidean space through the application of Classical Multi-Dimensional Scaling (CMDS) method[5]. The main aim of CMDS is to find a configuration in a small number of dimensions with a small distance deviation between the points, making the

plotting and visualization of hidden pattern in the data possible[8][14]. The input data is in the form of a distance matrix that represents the distance between a pair of objects. In this method, distance between the pair of points is treated as Euclidean distance. At the end of this process we can get the transformed training and testing data set.

After the preparation stage, each preprocessed recorded measurement must be attached with its own cell ID so that it can be used in the post processing stage. Accordingly, the obtained measurements were effectively correlated with their respective target cell ID.

4.4 Modeling

Different researches on machine learning define model in different ways. One of such definitions define a ML model as an artifact that is created by the training process[10]. The process of training a ML model involves providing a ML algorithm (that is the learning algorithm) with training data to learn from.

The central idea in this methodology, anomaly detection methodology, is using the cell level data (i.e. inHO and Traffic) that is acquired from fault free state to profile the behavior of the network. The learned profile will then be used to address the cell outage autonomously. The model will be used to predict a test data, which contains both normal and abnormal instances. In addition, we test the model with reference dataset. Therefore, this machine learning and anomaly detection method is semi supervised because it uses normal input data for training to produce the model.

In this section, the modeling process is executed. Figure 4-11 shows the system model that presents the activities, processes and flows in the framework (COD). The modeling process involves two stages; namely, profiling and detection & localization. The major tasks performed in each of these stages are presented below.

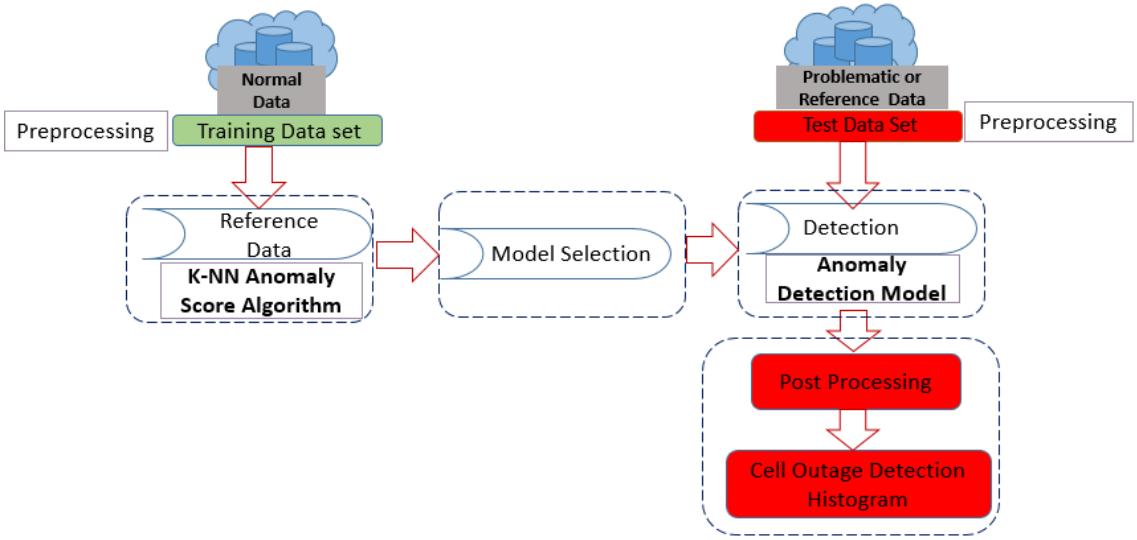

Figure 4-11 System Model

4.4.1 Profiling

After performing data preparation for the records collected from normal network operation, the prepared data is stored as reference data (D_R) that will be used to learn about the normal operation of the network. In order to learn about the normal operation, two state of the art density based anomaly detection algorithms; namely, K-NNAD and LOFAD, were applied on the reference data. The goal of the algorithms is to define anomaly detection rule that can differentiate between normal and abnormal reports by computing a threshold 'φ' based on a dissimilarity measure 'D', which is expressed as follows.

$$f(xi)\begin{cases} Normal, if\ D(xi, DR) \leq \varphi \\ Anomalous, if\ D(xi, DR) > \varphi \end{cases} \quad (4\text{-}2)$$

The two anomaly detection algorithms are measure the abnormality based on the position of the point in the MDS space classified as local and global anomalies. Local anomalies are localized to small spatial region or a neighborhood, whereas global anomalies are bounded to the entire dataset.

The two detection algorithms are discussed as follows.

4.4.1.1 K-NNAD

It computes the global dissimilarity $D_{k\text{-NNAD}}$, [5]which assigns a score to the test observation x_i based on the k^{th} nearest training point in the MDS space.

Let x_i be the test instance, and k be the k^{th} neighbor in the D_R. To label x_i as normal or abnormal, the k-NNAD computes a D_k-NNAD as:

$$D_{K-\text{NNAD}}(x_i, K, D_R) = \frac{1}{N_{tr}} \sum_{i=1}^{N_{tr}} I(d_t \leq d_i) \qquad (4\text{-}3)$$

Where:

$N_{tr} = | D_R |$,

d_t is the distance of x_i from its k^{th} nearest neighbor,

d_i is the distance between i and its k^{th} nearest training object in D_R, and

I(.) is an indicator function.

N.B. The indicator function is activated as soon as the condition $d_t < d_i$ is fulfilled.

The expression presented in (4.3) represents a global density based anomaly detection score as proposed in [5]. Accordingly, a test measurement is marked as anomalous if it receives a score greater than the 'φ' value.

4.4.1.2 LOF

The instead of the distance the LOFAD [5] compares the local density 'ρ' of the object to that of its k neighbors, which is obtained as per the following procedures. It constructs a local neighborhood of an instance x_i and defines its distance to the k^{th} nearest neighbor $NN(x_i; k)$:

$$d_b(x_i, k) = d\big(x_i, NN(x_i, k)\big) \tag{4-4}$$

Where:

In the above formula, the $d_b(x_i, k)$ is used to construct a neighborhood $N(x_i,k)$ by including all those points that fulfill the criteria:

$d(x_i; x_j) \leq d_b(x_i, k)$.

Reachability distance d_r is defined to estimate the ρ (x_i, k) as follows:

$$d_r(x_i, k) = \max\{d_b(x_j, k), d(x_j, x_i)\} \tag{4-5}$$

Similarly, ρ can be defined as:

$$\rho\,(x_i, k) = \frac{|N(x_i, k)|}{\sum_{x_j \in N(x_i,k)} d_r(x_i, x_j, k)} \tag{4-6}$$

The $d_r(xi, xj, k)$ ensures that instances that lie farther away from x_i have less impact on $\rho\ (x_i, k)$.

Finally, the D can be calculated by comparing the ρ of x_i to its N(x_i, k), which is formally defined as:

$$D_{\text{LOFAD}}(x_i, K; D_{\text{train}}) = \frac{\sum_{x_j \in N(x_i,k)} \frac{\rho(x_j, k)}{\rho(x_i, k)}}{|N(x_i, k)|} \tag{4-7}$$

D_{LOFAD} represents a local density-estimation score, with a value close to 1 indicating that x_i has the same density as its neighbors. On the other hand, a significantly high D_{LOFAD} score is an indication of anomaly.

Parameter selection

The parameter selection for both algorithms are performed using a cross-validation (CV) method and the algorithm is described below. It splits the D_R into training and validation dataset using K-fold approach. To select the optimal parameter, each detector (K-NNAD and LOFAD) is trained for different values of *K (*i.e. 1, 2, ……,30) and select the parameter with the highest average model detection score.

1: Split the target dataset D_R into K chunks.

2: for l = 1, 2, …., K: do

3: Set D_{val} to be the l^{th} chunk of data.

4: Set D_{train} to be the other K - 1 chunks.

5: Fit each model of D_{train} and evaluate its performance on D_{val}.

6: end for

7: Model Selection: Select the model with the highest average detection score.

Post processing

The other task performed in the profiling stage is post processing, which involves calculating the reference anomaly detection score for each BS and their respective Z-score. In the selected case sites, the high capacity Node B with several carriers has 3x4 configuration so it generates 12 cell level records per hour. By using the data labeled as abnormal by the anomaly detection algorithm, the anomaly detection score for each BS can be calculated by using the following procedures:

First, count the number of abnormal cells in the site based on the cell ID and location information in the record; and then

based on the neighbor list information, add the number of abnormal records for each site to get the total for each so that sum can be used as anomaly detection score for its respective site.

In this study, 4 neighboring sites were selected based on their dominance (closeness) on (to) the target site. Since each site has 12 cells, information has been collected from a total of 48 cells in the 4 neighboring sites. The anomaly detection score was, then, calculated by using the information from these cells.

Based on the anomaly detection scores obtained from the above procedures, the Z-score can be calculated as follows.

$$Zb = \frac{|Nb - \mu n|}{\sigma n} \tag{4-7}$$

Where:

Nb = Anomaly detection score;

μn = Mean of anomaly detection scores of the neighboring sites;

σn = The standard deviation of anomaly detection scores of the neighboring sites;

The calculated Z-score above can be used as a reference to identify outage cells/site in the detection phase.

4.4.2 Detection and Localization

In the detection and localization phase, the detection of outage sites along with identifying their location is done. To do so, the test record is preprocessed in a similar way used in the profiling stage. The embedded representation of test data that is obtained from the preprocessing stage is classified as normal or anomalous by using anomaly detection model and respective cell ID is attached to each.

The post processing is done by taking the anomalous ones to calculate Z-score in a similar was used in the profiling phase. By measuring degree of variation of the Z-scores from the normal, a histogram was plotted to clearly observe the changes that occurred in the network. Finally, outage

cell is detected and localized by observing the changes in Z-score for each site based on the histogram.

To increase detection accuracy, the anomaly counts were normalized by the total number of anomalies to amplify the output[9]. The site specific deviations are, then, used to create bar plot called site outage detection histogram.

4.5 Evaluation

Model selection is made by comparing the two algorithms based on the performance of the algorithms and used to profile the network behavior. performance of the system model was tested by using real problematic network state data and measure the results using classic data mining metrics such as ROC curve, Precision, recall. The description is presented below

ROC curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at different discrimination threshold settings. The formulae to compute TPR and FPR is shown in the figure. Demonstrating the performance of a binary classifier at different circumstances is the main purpose of the plot. In machine learning, the most common measure from ROC analysis is area under ROC curve (AUC). It indicates the ability of the classifier to make a good separation between positive and negative labels.

- ***Overall Accuracy*** *as* denoted by Equation (3-1) measures the ratio of correctly classified sites in both classes with respect to the whole sites.
- ***Precision*** for outage class denoted by Equation (3-2) is the number of outage sites, which are classified as outage, divided by the total outage sites classified by the model.
- ***Recall*** or TPR for outage class denoted by Equation (3-3) is the number of outage sites, which are classified as outage, divided by the total number of outage sites.
-

5 Result Analysis and Discussion

This section of the thesis presents the outcomes of the anomaly detection framework, that is the totality of phases used to develop the model and measure the models' performance for detecting cell outage in the real environment by testing the problematic data. For this thesis, data was collected from 80 sites of the case company in Addis Ababa. As was mentioned earlier, the selected time series for this experiment is 12 AM (noon). A total of 939 records per day were collected for four months from which normal network operation working hour records were filtered to profile the normal network behavior. The problematic data from the above dataset were, on the other hand, used for testing purpose. The average of the selected normal hour records data was used for profiling. The normal reference network operation data was also used to set the threshold for the Z-score deviation. Then, different testing experiments were applied on selected problematic records to measure the performance of the target model.

5.1 K-NN anomaly Detection

The first step in both anomaly score calculation is preprocessing the normal network operation hour data that will be used to train the two density based anomaly detection algorithms that are used to set a classification rule to be used in the testing stage so as to predict the unseen data as normal and abnormal. The test data were also preprocessed by following the steps indicated earlier. The figures below represent the output of the K-NN algorithm that was ran to assign the dissimilarity score for each instant. The interpretation of each figure is given in the discussions under.

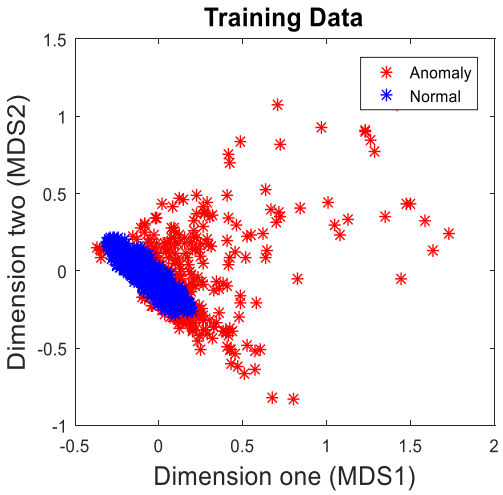

Figure 5-1Normal dataset used in the embedded space.

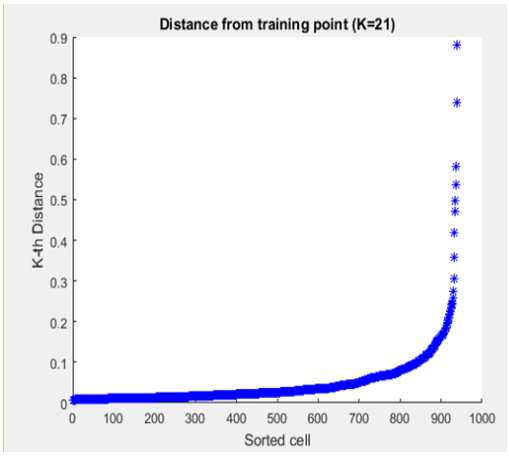

Figure 5-2 Sorted Distance from the training Data.

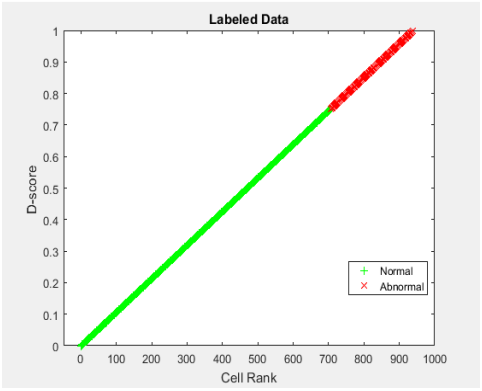

Figure 5-3 Anomaly detection score.

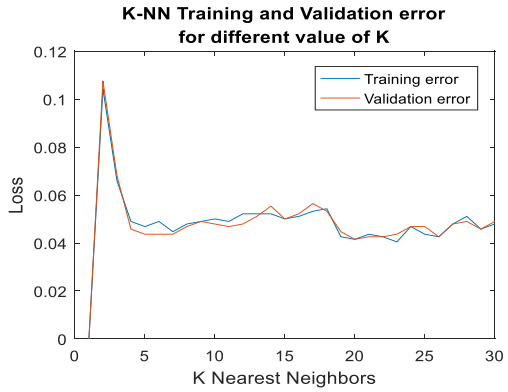

Figure 5-4 Training and validation error.

Figure 5-1 shows the output of the preprocessed training dataset in the embedded space, which was used for both algorithms. As can be seen from the figure, most of the data is highly

concentrated with a few outliers, which is an indicator of greater similarity among the cell records. Figure 5-2 shows sorted k^{th} distance of each record measured from the training data. It is clearly observable from the figure that the data collected from each site have similarities as most, around 80%, of the k^{th} distance lies below two.

Dissimilarity score $D_{k\text{-}NNAD}$ is presented in Figure 5-3 that was measured based on the equation presented in (4-3). Based on the results of $D_{k\text{-}NNAD,}$ classification of data points to normal and abnormal classes was done and the selection was made by considering the 75^{th} percentile of anomaly score in the training data.

In the stage of model training, training accuracy and validation of the model were performed by using cross validation. Figure 5-4 shows the training and validation accuracy of the model. To get optimal model, parameter selection was performed for different values of *k* (1, 2…30) based on K-fold cross validation. With an overall training error of the K-NNAD model that lies between 0.04 and 0.1, the optimal model was achieved at k=21 with a training error of 0.04.

5.2 LOF anomaly Detection

The same data and procedures used in the preprocessing the K-NN anomaly score calculation were employed in LOF anomaly score calculation. Accordingly, local reachability distance was measured, reachability density was calculated by using the results from reachability distance measurements, and, LOF anomaly score was calculated by following all the steps used in LOF calculation. Based on the scores obtained from the above calculations, data classification was made into normal and anomalous, by taking the 75^{th} percentile into consideration. Lastly, the model was trained by using the training data and validation was made by using K-fold cross validation method. The model's training and validation accuracy was also measured accordingly.

The results of the above procedures are presented by the figures below followed by a brief discussion on their outcomes.

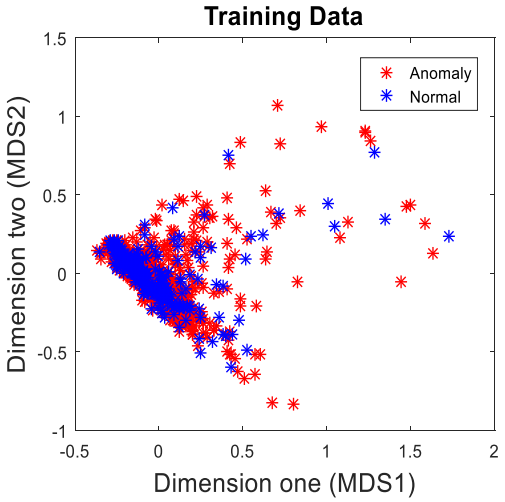

Figure 5-5 Normal dataset in the MDS embedded space.

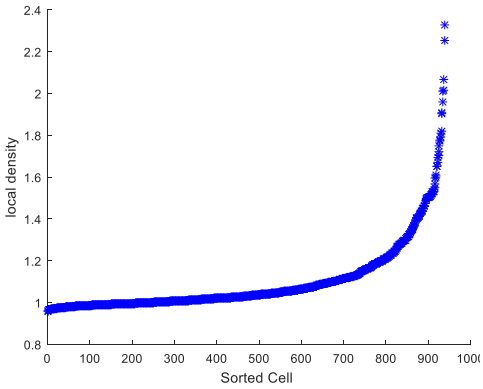

Figure 5-6 Local density estimation scores.

Figure 5-5 represents the training data, which is a normal dataset, in the MDS embedded space, as was classified by using the LOF algorithm. It can, thus, be learnt from the figure that there exists no clear boundary between the normal and anomalous data points, indicating that there exists mixing between the normal and anomalous data points among the outliers, which can be due to the fact that LOF algorithm measures local density of the dataset, unlike the K-NN, which measures global dissimilarity. This, in turn, would have negative impact in the classification model, as this data serves as an input in the model.

The LOF score is presented in Figure 5-6. As it can be seen from the figure, the data instance LOF score of most of the data points are found to be close to 1. This indicates that most of the data points' neighbors are similar.

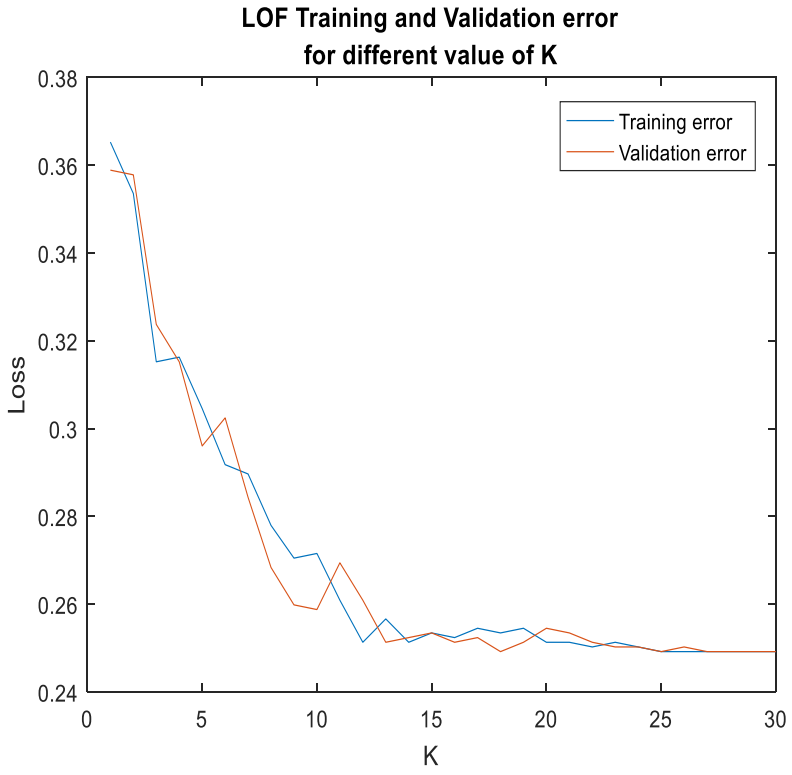

Figure 5-7 Training and Validation

Figure 5-7 shows LOF training and validation error for different values of k. As can be seen from the figure, the error values range between 0.24 and 0.38. This indicates that the LOF model may misclassify our dataset while prediction of the unseen data into normal and anomalous is performed. The lowest training error of 0.25 of the LOF model was obtained at K=25. However, this value of training error falls far away from the one obtained from the K-NNAD training error, which falls between 0.04 and 0.1.

The figure below gives the results of the comparative analysis of the K-NNAD and LOFAD models that was performed with the purpose of selecting the best performing model for our case. The analysis was made by using ROC curve that shows the coverage of the two models.

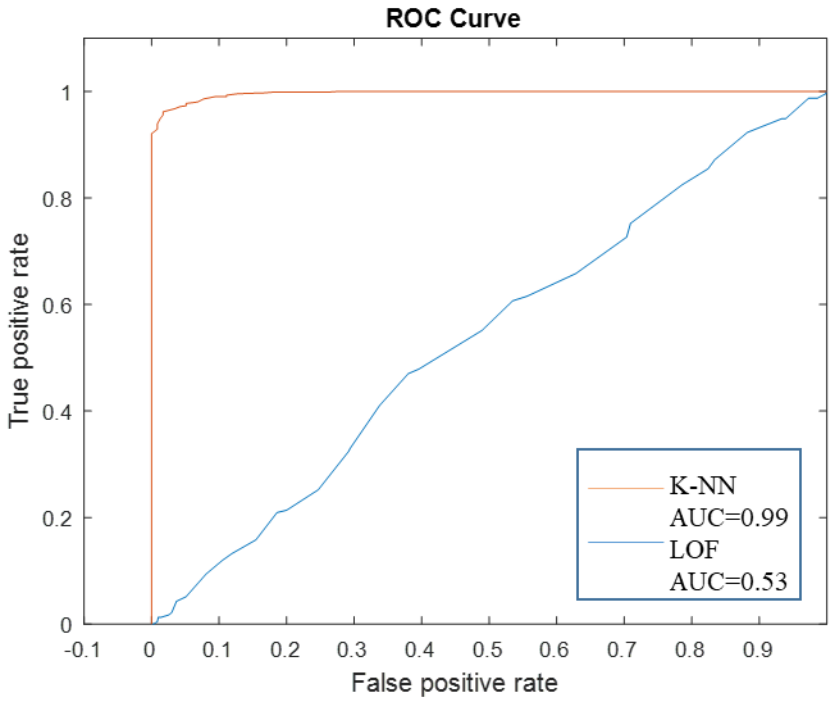

Figure 5-8 ROC Curve

In a ROC curve, an optimal model would find a TPR of 1 and a FPR of 0 and hence the ROC curve would have a point in the top-left corner of the chart. Accordingly, a visual comparison of the two anomaly detection methods on the above figure can help us to select the model that serves our interest best. As can be seen from Figure 5-8 curve, the Area Under Curve of the K-NNAD model is much larger (0.99) than that of the LOFAD model (0.53). From this and the results obtained from the procedures performed on the K-NNAD and LOFAD models earlier, we can learn that the K-NNAD model performs better than LOFAD in serving the interest of this thesis. Thus, profiling was done by using K-NN.

5.3 Detection

When high traffic measurement is obtained from normal network operation time, it shows similarity with the data sample obtained from the outage state of neighbor cell of the outage cell. Thus, in the embedded space outage cell neighbor cells, measurements are projected to be close to the measurements obtained from high traffic serving cell at normal time. From the classification perspective, the target model will classify such measurements as anomaly correctly.

Outages that occur in the last minutes of the collection time may not cause significant increment of measurement in the neighbor cell. Such outages can be magnified in the amplification stage to make them easily detectable. Similarly, partial outages are treated in the same manner.

At this stage, the performance of the selected method is tested. Based on the results obtained from the performed procedures and the decision made thereby, the K-NN anomaly detection algorithm was selected as the best model. In this section, the performance of the K-NNAD was tested by using data that was collected from problematic network states. The detection was made based on the spatial study of the different sites at a single temporal instance by using a number of experimental procedures. The results of the test are presented in the following discussion.

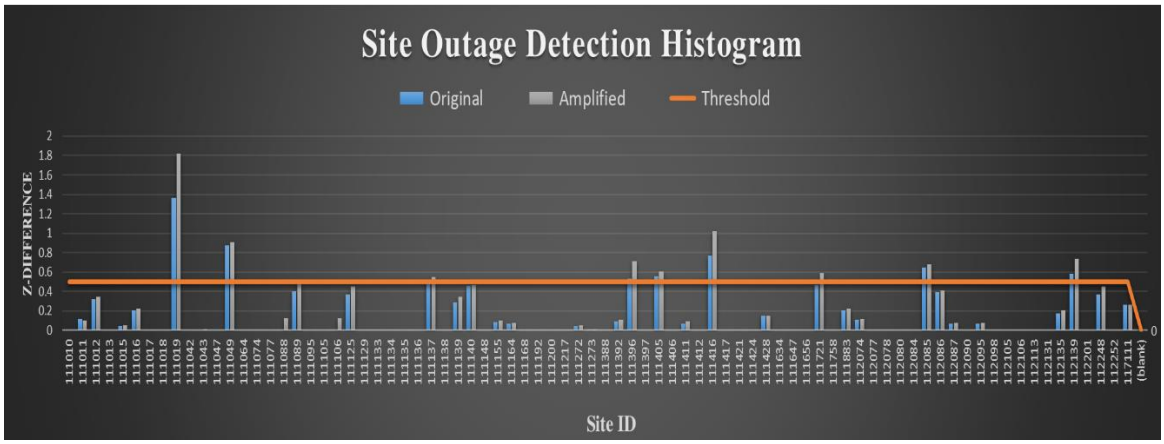

Figure 5-9 Spatial study of the different sites at a single temporal instance.

Several outage time data have been collected and were used to evaluate the performance of the method. Some of these data were outages that only affect one site, while others were outages that occur in multi-sites. The detection performance of the model was, then, analyzed by considering both conditions. Figure 5-9 represents only one instance of the many instances in the test, in which multi-site outages were observed. The test was made on the selected 80 sites, of which five were reported outage. The histogram was plotted based on the z-score difference between the test and the reference state.

As can be seen from the figure, the total number of outage sites is eight, three of which were correctly recognized as outage, as they were reported as outage earlier. The five outage sites are from the normal sites, hence, are incorrectly detected. In other words, from the 4 outage sites, 3 were classified as outage correctly while the rest are incorrectly identified as normal and from the 76 normal sites, 71 were classified as normal correctly while the rest are incorrectly identified as outages. The confusion matrix is presented in Table 5-1 below. The performance evaluation indicates that the method has an overall accuracy of 92.5% with recall and precision of 0.75 and 0.375, respectively.

Table 5-1 Confusion Matrix of the Method

		Predicted	
		Outage	**Normal**
Actual	Outage	**3**	**1**
	Normal	**5**	**71**

6 Conclusion, Recommendation and Future Work

6.1 Conclusion

Cell outage management has always been one of the major challenges that telecom service providers encounter in their network management system. Telecom companies need to detect outage cells as fast as possible and make compensations accordingly so as to provide a high quality service to their customers. ethio telecom is one of such companies providing its service in Ethiopia. As any of the other similar companies, ethio telecom faces similar challenges.

This thesis was prepared with the purpose of developing the cell outage detection. To this end, an attempt was made to develop an outage detection model by applying density based anomaly detection algorithm using machine learning technique. From the existing machine learning methodologies, CRISP-DM was employed. Accordingly, two algorithms; namely K-NNAD and LOFAD were selected for this purpose their performance was evaluated. Similarly, of the three machine learning techniques, i.e. semi-supervised, supervised and unsupervised, semi-supervised technique was implemented.

The performance of the two density based machine learning algorithms was tested. The algorithms were trained by using normal network operation hour performance data, i.e. inHO and traffic data. The validation of the two algorithms was, then, made by using k-fold cross validation technique. To select the optimal model, parameter search for different values of k was

done. By comparing the performance of the two algorithms based on the data obtained from the training accuracy and the ROC curve, the k-NNAD was found to be of a better performance. This anomaly detector was, then, used to profile the normal network operation behavior, which was used in the testing stage.

The overall performance of the method was evaluated by using real problematic network state data of a single instance time and an overall accuracy of 92.5% was obtained. The same performance evaluation also indicates that the method has 0.75 recall and 0.375 precision. From this, we can conclude that cell outage detection can be performed better by employing anomaly detection method.

6.2 Recommendations for Future Work

As indicated in the summary, better outage detection can be achieved by using anomaly detection methods. In this thesis, the K-NNAD was found to perform better than the LOFAD in detecting cell outages on the selected sites. Despite the methods 92.5% performance, the false negative alarm of the method was found to be high. To minimize this, it is recommended that other studies should consider other KPI data or user level data, like MDT, and an even better result can be achieved.

References

[1] GSM Association, “The Mobile Economy 2019,” p. 56, 2019.

[2] D. Baumann, “Minimization of Drive Tests (MDT) in Mobile Communication Networks,” no. March, pp. 9–16, 2014.

[3] J. Moysen and L. Giupponi, “From 4G to 5G : Self-organized Network Management meets Machine Learning,” pp. 1–23, 2017.

[4] S. Bekele, “Cell Outage Detection Through Density-based Local Outlier Data Mining Approach : In case of Ethio telecom UMTS Network,” no. November, 2018.

[5] O. Onireti, “A cell outage management framework for dense heterogeneous networks,” *IEEE Trans. Veh. Technol.*, vol. 65, no. 4, pp. 2097–2113, 2016.

[6] C. Sartori, H. Sanneck, K. Pedersen, J. Pekonen, and I. Viering, “SON for Heterogeneous Networks (HetNet).”

[7] Ethio telecom, *“Ethio telecom 2018/2019 EFY Bus. Perform. report”[July 23, 2019], Available https//www.ethiotelecom.et/2018-19-efy-p-reporte/, Accessed*.

[8] A. Zoha, A. Saeed, A. Imran, M. A. Imran, and A. Abu-dayya, “A SON Solution for Sleeping Cell Detection using Low-Dimensional Embedding of MDT Measurements,” no. 978, pp. 1626–1630, 2014.

[9] F. Chernogorov, S. Chernov, K. Brigatti, and T. Ristaniemi, “Sequence-based detection of sleeping cell failures in mobile networks,” *Wirel. Networks*, vol. 22, no. 6, pp. 2029–2048, 2016.

[10] P. V. Klaine, M. A. Imran, O. Onireti, and R. D. Souza, “A Survey of Machine Learning

Techniques Applied to Self-Organizing Cellular Networks," *IEEE Commun. Surv. Tutorials*, vol. 19, no. 4, pp. 2392–2431, 2017.

[11] H. Sanneck and C. Sartori, "LTE Self-Organising Networks."

[12] A. Zoha, A. Saeed, A. Imran, M. A. Imran, and A. Abu-Dayya, "Data-driven analytics for automated cell outage detection in Self-Organizing Networks," *2015 11th Int. Conf. Des. Reliab. Commun. Networks, DRCN 2015*, pp. 203–210, 2015.

[13] M. Amirijoo, "Cell outage management in LTE networks," *Proc. 2009 6th Int. Symp. Wirel. Commun. Syst. ISWCS'09*, pp. 600–604, 2009.

[14] A. Zoha, A. Imran, and A. Abu-dayya, "A Machine Learning Framework for Detection of Sleeping Cells in LTE Network," no. September 2015, 2014.

[15] T. Zhang, L. Feng, P. Yu, S. Guo, W. Li, and X. Qiu, "A handover statistics based approach for Cell Outage Detection in self-organized Heterogeneous Networks," *Proc. IM 2017 - 2017 IFIP/IEEE Int. Symp. Integr. Netw. Serv. Manag.*, no. March 2018, pp. 628–631, 2017.

[16] V. Nguyen, "Anomaly detection in self-organizing network A case study: Failure prediction in a real LTE network," 2019.

[17] K. Kawamura, M. Murata, S. Higuchi, and F. Kurokochi, "Management system for mobile networks," *Fujitsu Sci. Tech. J.*, vol. 48, no. 1, pp. 47–53, 2012.

[18] G. Stefan, D. T. Techniques, and F. Group, "Radio measurement collection for Minimization of Drive Tests (MDT); Overall description," vol. 0, pp. 31–52, 2018.

[19] N. • Kaaranen, Ahtiainen, Laitinen, Naghian, "UMTS System Architecture and Protocol Architecture UMTS / GSM Network Architecture," 2005.

[20] P. Yu and F. Zhou, "Self-Organized Cell Outage Detection Architecture and Approach for 5G H-CRAN," vol. 2018, 2018.

[21] 3rd Generation Partnership Project, "3GPP TS 32.541 V1.5.0 (2010-08)." .

[22] Amevcas, “Self-Optimizing Networks: The Benefits of Son in LTE,” pp. 1–69, 2011.

[23] Q. Liao, “Statistical Learning , Anomaly Detection , and Optimization in Self-Organizing Networks,” no. November 2016.

[24] B. V Chandola, A. Banerjee, and V. Kupin, “Anomaly Detection,” vol. 41, 2009.

[25] S. Aghabozorgi, A. Seyed Shirkhorshidi, and T. Ying Wah, “Time-series clustering - A decade review,” *Inf. Syst.*, vol. 53, pp. 16–38, 2015.

[26] T. Zhang, L. Feng, P. Yu, S. Guo, W. Li, and X. Qiu, “A Handover Statistics based Approach for Cell Outage Detection in Self-organized Heterogeneous Networks,” pp. 628–631, 2017.

[27] M. Nicolae, “Learning similarities for linear classification : theoretical foundations and algorithms To cite this version : HAL Id : tel-01975541 Learning Similarities for Linear Classification : Theoretical Foundations and Algorithms Apprentissage de similarités pou,” 2019.