MULTIPLE LINEAR REGRESSION BASED BASE TRANSCEIVER STATION POWER CONSUMPTION PREDICTION

Habtamu Mossie Mengistu

A Thesis Submitted to

The Department of Computer Science and Engineering

School of Electrical Engineering and Computing

Presented in Partial Fulfilment of the Requirement for the Degree of Masters in Computer Science and Engineering

Office of Graduate Studies

Adama Science and Technology University

July 2021

Adama, Ethiopia

Multiple Linear Regression Based Base Transceiver Station Power Consumption Prediction

Habtamu Mossie Mengistu

Advisor: Bahiru Shifaw (Ph.D.)

A Thesis Submitted to

The Department of Computer Science and Engineering

School of Electrical Engineering and Computing

Presented in Partial Fulfilment of the Requirement for the Degree of Masters in Computer Science and Engineering

Office of Graduate Studies

Adama Science and Technology University

July 2021

Adama, Ethiopia

APPROVAL SHEET

I, the advisors of the thesis entitled "Multiple Linear Regression based Base Transceiver Station Power Cconsumption Prediction" and developed by HabtamuMossie, hereby certify that the recommendation and suggestions made by the board of examiners are appropriately incorporated into the final version of the thesis.

____________________	____________________	________________
Advisor	Signature	Date

We, the undersigned, members of the Board of Examiners of the thesis by Habtamu Mossie Mengistu have read and evaluated the thesis entitled "Multiple Linear Regression based Base Transceiver Station Power Consumption Prediction" and examined the candidate during open defense. This is, therefore, to certify that the thesis is accepted for partial fulfillment of the requirement of the degree of Master of Science in Computer science and Engieering.

____________________	____________________	____________________
Chairperson	Signature	Date
____________________	____________________	____________________
Internal Examiner	Signature	Date
____________________	____________________	____________________
External Examiner	Signature	Date

Finally, approval and acceptance of the thesis is contingent upon submission of its final copy to the Office of Postgraduate Studies (OPGS) through the Department Graduate Council(DGC)and School Graduate Committee (SGC).

____________________	____________________	____________________
Departemnt Head	Signature	Date
____________________	____________________	____________________
School Dean	Signature	Date
____________________	____________________	____________________
Office of Postgraduate Studies, Dean	Signature	Date

DECLARATION

I, hereby declare that this Master Thesis entitled "Multiple Linear Regression Based Base Transceiver Station Power Consumption Prediction" is my original work. That is, it has not been submitted for the award of any academic degree, diploma or certificate in any other university. All sources of materials that are used for this thesis have been duly acknowledged through citation.

____________________ ____________________ ____________________

Name of the Student	Signature	Date

RECOMMENDATION

I, the advisor of this thesis, hereby certify that I have read the revised version of the thesis entitled “Multiple Linear Regression Based Base Transceiver Station Power Consumption Prediction” prepared under my guidance by Habtamu Mossie Mengistu submitted in partial fulfillment of the requirements for the degree of Mater’s of Science in computer scince and engineering.

Therefore, I recommend the submission of revised version of the thesis to the department following the applicable procedures.

______________________ ______________________ _______________

Advisor Signature Date

ACKNOWLEDGMENT

First and for most my sincere gratitude goes to almighty God for giving me the enablement and capability for this research.

I would like to express my truthful gratitude to my Advisor Bahiru Shifaw (Ph.D.) for this invaluable advice and guidance through the course of my study and research.

Last but not least, I would like to thank my Family for continuous support and motivation, Moreover, I am appreciative for those who support my work, giving me encouragement and valuable help.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACRONYMS

AC	Alternating Current
ANN	Artificial Neural Network Current
ANOVA	Analysis of Covariance
ATS	Automatic Transfer Switch
AUC	Authentication Center
BSC	Base Station Controller
BSS	Base Station Subsystem
BTS	Base Transceiver Station
CDMA	Code Division Multiplexing Access
CO2	Carbon Dioxide
DC	Direct Current
EIR	Equipment Identity Register
ET	Ethio Telecom
GRU	Gated Recurrent Unit
GSM	Global System for Mobile Communication
HLR	Home Location Register
ICT	International Mobile Subscriber Identity
IDE	Integrated Development Environment
IMSI	International Mobile Subscriber Identity
IQR	Interquartile Range
ISDN	Integrated Services Digital Network
LCD	Liquid Crystal Display
LCO	Lithium Cobalt Oxide
LED	Light Emitting Diode
LIB	Lithium Ion Batteries
LSTM	Long Short Term Memory
LTE	long-Term Evolution
MAE	Mean Absolute Error
ML	Machine Learning

MLP	Multilayer Perception
MLR	Multiple Linear Regression
MS	Microsoft
MSAG	Multi-Service Access Gateway
MSAN	Multi-Service Access Node
MSC	Mobile switching Services Center
MSE	Mean Square Error
MW	Micro Wave
NSS	Network and Switching Subsystem
OLS	Ordinary Least Square
OSS	Operation and Support Subsystem
PCB	Printed Circuit Board
PDF	Portable Document Format
PSTN	Public Switched Telephone Network
PV	Photovoltaic
QOS	Quality of Service
RAM	Random-Access Memory
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
RRU	Remote Radio Unit
RSSI	Received Signal Strength Indicator
SIM	Subscriber Identity Module
SLR	Simple Linear Regression
UMTS	Universal Mobile Telecommunications System
VLR	Visitor Location Register
WCDMA	Wideband Code Division Multiplexing Access

ABSTRACT

The development in wireless Technologies particularly cellular mobile networks is attracting people to ease their daily activities. In mobile networks, Base Transceiver Station (BTS) is the main infrastructure element that is performing the task of connecting the customer equipment with the cellular network. As there is a rapid growth of mobile subscribers and number of base stations demand, power consumption is also increasing. Thus studying the impact of factors related to power consumption at a base station has a significant advantage to save resources as well as providing sustainable service for telecom sectors. The fundamental aim of this thesis is, therefore, to analyze factors such as temperature, humidity and other factors to base transceiver station power consumption by implementing machine learning algorithm techniques to datasets of base station power system. Specifically, multiple linear regression technique is applied to deal with independent variables in respect of dependent variable. This technique was proposed as all the assumptions or prerequisites to use it in solving machine learning problems are satisfied by the behavior of the dataset on which this study is based. Furthermore, Python, Micro soft Access and Microsoft Excel are used as tools for data preprocessing. This research used datasets of 187484 rows having 4 variables which was split into 80 %, 10% and 10% training, validation and test sets respectively which has given best result of accuracy as compared to other split options. Hence, the experimental result confirms that temperature and DC load current have a positive relationship with power consumptions whereas voltage and humidity are negatively related with power consumption. The model which employed multiple regression techniques has a prediction performance accuracy of 99.5 %.

Keywords: Power consumption, Base Transceiver station, Machine learning, Prediction

CHAPTER 1 INTRODUCTION

1.1 Background of the Study

Currently telecommunication industry is evolving to high capacity networks and multi services delivery due to business requirement and technology innovations. The development in wireless technologies particularly cellular mobile networks is attracting customers to ease their life[1]. In environmental factors like temperature and humidity has effect on power consumption. A growing demand to reserve energy resource, primarily reserves of fossil fuel is increasing. There is also a change in climate that brought about global warming. This situations resulted in a number of studies on the aspects of energy consumption in relation to weather condition[2]. Operating temperature and computer system hardware is correlated. Therefore, in data center enviromntal factor like temperature control system used for reducing power consumption and also it protect telcom devices from failure[3]. According to[4], multiple regression model use one output variable and more than one input variables to make predictions as studying the relationships of variables through the application of regression model is simple and effective[5].

In mobile networks, Base Transceiver Station (BTS) is one key infrastructure element that is performing the task of connecting the customer equipment with the cellular network. BTS services may be interrupted due to many reasons like transmission failure, optical fiber cut, power system failure, natural disaster or many more. In the case of Ethio-Telecom power system failure takes the biggest share[6]. Because all the communication equipment is supplied from a common power source, their failure affects overall services provisioning. Power systems are critical elements in any communication. their failure may cause an interruption to the complete services or failure of subsystems. Therefore, identifying and eliminating the failure causes of the BTS power system minimizes the downtime of the overall system and it customer satisfaction there by maximizing revenue. At this time there are around 7093 BTSs over the whole country and above 742 of them are found in Addis Ababa[6]. Base stations involve global system for mobile communication (GSM), wideband code division multiplexing access (WCDMA), long-term evolution (LTE) and code division multiplexing access (CDMA) cellular networks. One BTS electric bill per month is 1552 birr so from all BTS monthly bill is around 11008336 birr. The average energy consumption of one base station (BTS) per day is 32.4kWh- around 2235.6kWh per month. birr. Hence, there is an average energy consumption of 13414MWh per month for all BTS. This is a huge amount of energy, as Ethiopia has a insufficiency of energy currently[7].

1.2 Motivation of the Study

Temperature and humidity are some of the factors that affects BTS power system and this makes the network infrastructure unreliable. So studying the relation between temperature, humidity and power system has significant advantage to facilitate communication between subscriber and telecom operator network. In the case of ethio telecom, power devices frequently fail and it requires high maintenance cost; and also impact on power consumption. As a result of this, it is interested on applying machine learning techniques. This research is initiated to study effect of temperature and other determinant factors there by identifying the cause of power system through the application of machine learning techniques.

1.3 Statement of the problem

Today machine learning has been used in many applications that have significantly changed human lives in many areas. Research's has not been done in temperature and humidity impact with power system. Base stations are the main targets of energy saving systems because of their large share in energy consumption [5].

Researchers [8][9] [10] has done on temperature effect with power but not done on for base station power system specifically in ethio-telecom [7]. Studies are available on energy efficiency to reduce power consumption; however, it has a gap in identifying temperature and other determinant effect on the system. Machine learning methods used to predict among variables with reason and result relations[4]. This paper attempts to propose an approach to use machine learning model to analysis the effect of base station power consumption prediction.

1.3.1 Research Questions

This study is aiming at answering the following research questions.

1. How strong is the relationship between BTS determinant factors have on power consumption?
2. How can the impact of determinant factors of power consumption are modeled using machine learning algorithms?
3. Is there any mechanism to notify the impact of environmental factors on BTS?

1.4 Objectives

1.4.1 General objective

The general objective of the research is predicting power consumption of Base transceiver station using multiple linear regressions.

1.4.2 Specific Objectives

To achieve the general objective of the research, the following specific objectives are to be carried out.

- Collecting BTS power related data and doing preprocesses so as to make the dataset appropriate for modeling.
- Analazing BTS data.
- Evalute the performance of developed model.
- Examining the effect of temperature and other determinant factors on BTS power consumption.
- Develop prototype to notifay determinant factors of BTS power consumption.

1.5 Scope and limitation of the study

1.5.1 Scope

The scope of this research is to study determinant factors such as humidity, temperature, DC load Current and voltage on BTS and its effect on power consumption through implementing multiple linear regression of machine learning technique based on the dataset collected about ethiotelcom at Addis Ababa.

1.5.2 Limitations

In this research there were limitation in getting data for regional sites. There is also limitation on separating the input power which is reached in the BTS.

1.6 Contributions of the study

The main contribution of the thesis is described below:

- Providing information about the impact of temperature and humidity on BTS power consumption.
- Generating information about BTS power consumption of ethio telecom
- Supporting the company to prepare a plan of power usage thereby reducing save its expenses.
- The result of this study can be used as a reference to studies related to BTS power consumption.

1.7 Thesis organization

This study contains the following seven chapters.

Chapter One is about introduction, background, statement of the problem, objective, scope, limitation and contributions of the study whereas Chapter Two reviews related literatures regarding BTS, power consumption, machine learning techniques, regression, temperature and humidity. Chapter Three presents methodology of the study. Chapter Four discusses the proposed system and its implementation. Chapter Five discusses the result of the analysis and develop prototype for this research work. In Chapter Six, conclusion, recommendations and future works are presented.

CHAPTER 2 LITERATURE REVIEW AND RELATED WORKS

2.1 Overview

Power consumption is the amount of energy consumed by equipment or machine measured in watt or kilowatt[8]. In the Present time communication equipment and network topologies works with high capacity of data in addition to it needs a high data rate of transferring information. Due to this broad development in wireless technology, consumption of power efficiency has become a major issue that has taken the attention of the world.

In reality this progress in communication network is proportional to the increased amount of CO_2 in the atmosphere. Thus, the world has been doing on reducing CO_2, European union has plan to decrease power consumption in 20% which they have named as Green Vision. Considering from operational viewpoint 3% of the world total electrical power which is equivalent to almost 600 Terawatt-Hour is absorbed by the Information and Communication Technology sector. This amount of power will be expected to rise to 1700TWh in end 2027[9].

One third (1/3) of the total ICT's power is spent by the telecommunication equipment. cell phone consumes half of the power consumed whereas 10% of the total mobile phone consumption is related with the terminal-user and 90% is absorbed by the network equipment's, around which two third(2/3) is consumed by base transmitter station [9]. The increasing responsiveness of global warming and environment consequences of Information and Communication Technology (ICT), scientists have been discovering techniques to decrease power consumption.

In Information and Communication Technology the cellular network become a significant issue which has greatly attracted many scholars to reduce the energy in cellular networks, since it increases the energy bills for telecommunication service providers and it increases the release of (CO_2) carbon foot print [10].

2.2 Mobile stations

In global mobil commincation system mobile stations comprises all user equipment and software needed for communication with a mobile network. Hardware and the subscriber identity module(SIM) are two portion of mobile station[7]. International Mobile Equipment Identity (IMEI) is a number that assign in mobile phone for identification. The SIM card holds (IMSI) and information of subscriber[11].

2.2.1 Base-Station Subsystem

Communicating with the cellular phone on the network is fundamentally associated with base station subsystem(BSS). There are two parts of BSS.

- Base transceiver station is an equipment used in a GSM network. BTS consists of the radio transmitter receivers, connected antennas that transmit and receive to directly communicate with the mobiles. Um interface used to connect base transceiver station with mobile station.
- Base station controller (BSC) used to monitor multiple base stations. Radio source also monitor using BSC.[12].

2.2.2 Network and Switching Subsystem

Network Switching Subsystem (NSS), is the main parts in GSM network. NSS used to transport the main control and interfacing for the whole mobile network. Therefor the following are part of core network.

- Mobile switching services center (MSC) is mostly associate with communication and it is the major part within the core network. In the other way MSC used as PSTN (public switched telephone network), ISDN (integrated services digital network), also MSC support the following activities like, inter-MSC handovers and call routing to a mobile subscriber[7].
- Home location register (HLR) is central database of all subscriber which associate services with subscriber. Every single SIM(subscriber identification module) on uses IMSI of the SIM as a primary key for searching its data[13].

- Visitor location register (VLR) has an advantage to provide information which enables by HLR to subscriber and implemented as single entity to provide service of the subscriber[7].
- Equipment identity register (EIR) consist the list of cell phone entity that decides whether given mobile device may be acceptable onto the network. Apiece of mobile equipment has a number termed as international mobile equipment identity[7].
- Authentication center (AUC) is an element a secure database that holds the secret key also contained in the user's SIM card[7].

2.2.3 Operation and Support Subsystem (OSS)

The Operation and Support Subsystem (OSS) is a part of GSM network architecture which connected to device of the NSS and the BSC and control both GSM network and traffic load of the BSS [7].

2.3 Climate change effect in power system

Climate change is expected to increase future temperatures. The potential impact of global warming and climate change on socioeconomic consequences has become an important and growing area of scientific study and evaluation. Separate lines of study include quantifying the likely impact of climatic change on measures of civil conflict[14]. Based on[2] temperature determine the impact of power consumption in residential and commercial sector. Climate change increase power demand for cooling.

2.3.1 Effect of temperature in Battery power system

Ambient temperature can significantly affect the battery cycle life. The higher the temperature, the less the battery cycle life[15]. Temperature is an important factor that affects the health and safe operation of lithium-ion batteries (LIBs). Working temperatures has impact on the performance of LIB therefore the battery life span will decrease. Controlling temperature maximize the life span of the battery and also keep it from deep discharge. High temperature for battery means fast chemical reaction this make the battery make shorten. Power system Examined based on the aging rate of the maximum charge storage capacity, on the aging rate of graphite electrode, The LCO (Lithium cobalt oxide) electrode, which is the cathode during discharging, is made from $LiCoO2$(LCO [16].

Temperature is one of determinate factors that affect battery performance significantly. Battery temperature prediction model if it is short it is appropriate put the battery as the surface temperature. Whereas long-term prediction the battery temperature may change due to the accumulation of the Joule heat. Thus, the battery's future temperature should be estimate before its long-term voltage and power[17].

2.3.2 Temperature effect on Photovoltaic power system

PV(Photovoltaic) system is sensitive to temperature, and it has direct relationship efficiency with ambient temperature due to this studying temperature has an advantage on performance analysis. Predicting photovoltaic panel performance is better compare to wind speed. Ambient temperature is direct proportionality between ambient temperature and solar photovoltaic module efficiency[18].

In solar system with rise of temperature, the short circuit current of the solar cell increases whereas open circuit voltage decreases[19]. Photovoltaic (PV) power output prediction is valuable to increase the effective operation of power system for their consumers though approximating the overall level of electricity that is able to be provided by their PV operators[20].

2.4 BTS power system

Power is important issue in wireless communication, due to this it needs alternative to get power access in case of ethio telecom commercial power (main power), diesel generator and battery is available. as shown below in the diagrams when main power off the generator start by ATS (automatic transfer switch) also if generator fail battery power will functional for BTSs. Rectifier convert AC to DC power to 48v and as some time it charges the battery from commercial power[6].

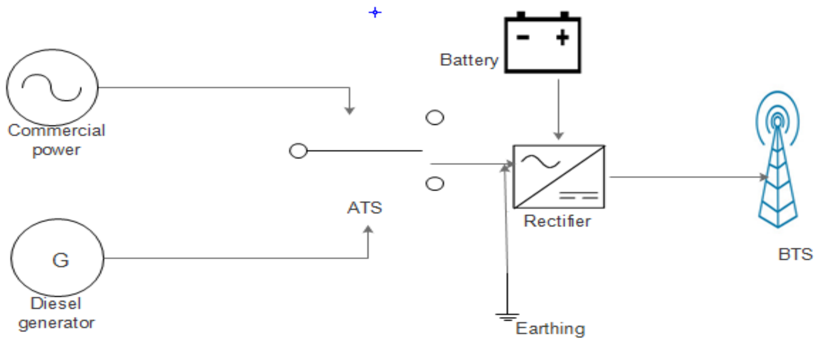

Figure 2-1BTS power system [8]

2.4.1 BTS power consumption determinates

2.4.1.1 Temperature

BTS increase demand of power consumption in mobile network with noticeable impact by environmental factors[21]. Signal and link quality has impact on communication due to temperature. Lower temperatures might require up to 16% less power to maintain a reliable communication. Power amplifier positively affects strength of the outgoing radio signal. Thus, higher temperature has weak signal. Therefore, it can be expected that with an increase in temperature, a higher transmission power is required to maintain the same signal strength and thus to ensure successful data communication[22]. High temperatures within a data center can cause a number of problems, such as increased cooling costs and increased hardware failure rates[3]. By collecting quantitative data, we can forecast the current Temperature based on current state of atmosphere.

The Neural Networks package supports different types of training or learning algorithms[23]. Indoor Environmental condition effectively predicate in RNN algorithm with considerable accuracy[24]. From [25] Cluster analysis used effectively in many climatic and atmospheric studies, especially for determining homogeneous climate regions based on meteorological variables. Researches show in [26] a multivariate linear regression model predicts indoor temperature, and how to define and add categorical variables also temperature prediction problem is studied.

The recommended operative temperature for communication equipment is from $+15^{\circ}$ C to $+25^{\circ}$ C [6].Environmental conditions can determine on mobile communication signals[27].

2.4.1.2 Humidity

Humidity is a concentration of water vapor exist in the air and stated as in percentage (%). High humidity indicate that accumulation of electric charge on the insulating and conductor material [36]. Thus corrosion, rust and short circuit happen on the device. According to [35] humidity range for communication equipment is 40% to 55% is recommended.

2.4.1.3 Traffic load

Researcher In [5] [27] show that power consumption of BTS influenced by traffic load. Thus with increase traffic load power consumption of BTS also increase.

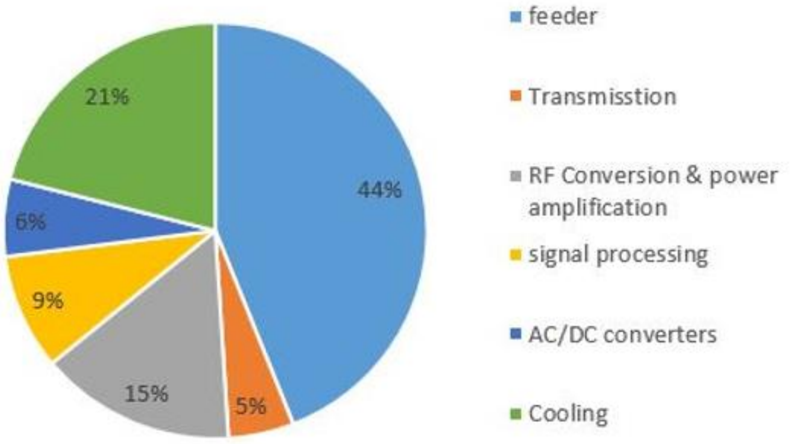

Figure 2-2 Power consumption inside base station [9]

2.5 Machine learning

The Development of ML created in 1950 by Alan Turning created “Turning Test” to check a machine’s intelligence. But a highly capable learning algorithm created in 1952 by Arthur Samuel which can play the game of Checkers with itself and get self-trained. Through learning ability, computers are gifted with the capability of acting as a human without being explicitly programmed for all tasks. Developing algorithms that can learn from data or experience and making data-based decisions or forecasts are other behavior of the machine-learning model[28].

Machine learning is algorithm to learn computer using data and understand the past and estimate the future. Machine learning uses theory of statistics in building mathematical models. The model predicts in the future, or descriptive to gain knowledge from data, or both. after a model is learned, its representation and algorithmic solution for inference needs to be efficient as well. From paste experience model assumption computer program optimizes using training data. In certain applications, the efficiency of the learning algorithm space and time complexity, is important as its predictive accuracy [29].

2.5.1 Unsupervised machine learning

The unsupervised machine learning approach is identifying hidden patterns in unlabeled input data. This technique is appropriate in a situation when the categories of data are unknown. Here, the training data is not labeled. Unsupervised learning is regarded as a statistic based approach for learning and thus refers to the problem of finding hidden structure in unlabeled data[28].

2.5.2 Clustering

Based on [30] Clustering is unsupervised a machine learning approach which groups are not labeled. Clustering technique applied in many sectors such as image analysis, pattern recognition, statistical data analysis, bioinformatics and others. Clustering is the task of grouping similar objects into different groups. Clustering work by computing data set into subsets to manipulate easily.

2.5.3 Supervised machine learning

Supervised machine learning uses label input and output data the algorithm learns to respond more accurately by comparing its output with those that are given as input. Supervised learning is also known as learning via examples and prediction application find based on past data [28].

2.5.4 Classification

Classification is supervised learning algorithm which process predicting a target value. The difference between regression and classification is that the variable forecasted in regression is continuous, whereas it’s discrete in classification[31] Classification, one the widespread and vital tasks of machine learning, is used to place an unknown piece of data into a known group. In order to build or train a classifier, you feed it training and finally it predicts or classify discrete value [31].

2.5.5 Regression

Regression is a technique for estimating the relationship among variables which have input and output. Regression used to predict a continuous value[4]. There are different types of regression. Here below, variants of regression namely simple linear and multiple linear regressions are presented. Linear regression is a model that captures the linear relationship between two (simple) or more (multiple) variables, one labeled as the dependent variable and the other(s) labeled as the independent variable(s). A linear relationship exists when increasing or decreasing the independent variable(s) results in a corresponding increase or decrease of the dependent variable.

2.5.5.1 Simple linear regression

Simple linear regression analysis focus on relationship between a dependent variable and one independent variable[4]. In Simple Linear Regression (SLR), the goal is to predict the value of the dependent variable y based on the independent variable x. In this method, only the relationship between these two variables are examined.

Though the relationship between x and y need not necessarily be linear, SLR models include errors in the data, otherwise known as residuals. A residual is the difference between the true value of y and the predicted value of [31].

2.5.5.2 Multiple linear regression

According to [32] multiple linear regression (MLR) method was used to analyze or predict the unknown effect of changing one variable over another. MLR function shows the linear relationship between dependent variable (Y) and several independent variables or functions of independent variables (X). In other words, MLR deals with more than one independent variable when creating a regression model. Technically; the linear regression estimates how much Y changes when X changes by one unit.

The MLR function is presented by

$$\mathbf{Y=\beta_0+\beta_1x_1+\beta_2x_{2+}\beta_3x+\ldots\ldots+\beta_nx_n+\epsilon} \tag{1}$$

Where y is the dependent variable,

β_0 is the intercept (the value of y when all the variables are 0),

β_1… n are the regression coefficients,

x_1…n represent the independent variables, and

ε is the random error component which measures how far above or below the True Regression Line (i.e. the line of means) the actual observation of Y lies. The mean of ε is zero. We used ordinary least squares (OLS) to estimate all the model regression coefficients. The coefficient of determination (R^2) measures the proportion of variance in the dependent variable that is predictable from the independent variables.

Based on [4] regression analysis is accomplished to determine the correlation between cause and effect relations and to make prediction by using the relation.

2.6 Measurements of Regression Model

2.6.1 Coefficient of determination

The strength of a relationship between the output variable and input variables or two input (independent variables) determined by correlation analysis.[33]. Correlation analysis basically linear association between variables in which the tendency for one variable to increase or decrease as the other increases or decreases. Correlation and regression analysis has similar tendency concerning with relationship among variables.

Correlation coefficient indicates how strong relationship have between variables. If the value is equal to zero it shows no correlation between variables. A correlation coefficient of positive one indicates variables are positively related or X and Y value is direct forward where as negative one is inverse relationship between variables. Thus as increase X value Y value will decrease[34].

2.6.2 Goodness of Fit of the Model

To check the goodness of fit of the model, it needs to evaluate the regression model, either using t-test or using ANOVA for regression the purpose of analysis of variance (ANOVA) is to test for significant differences between means. But in regression models it consists of calculations that deliver measurements about levels of predictability within a regression model and form a basis for tests of significance[34].

2.6.3 R-square

R-square is numerical processes for regression to delivers the proportion of variability. R^2 measures how regression model fit to the line[35].

2.6.4 F-test

The F-statistic is fundamentally a ratio of two variances. Variances are a measure of dispersion, or how far the data are scattered from the mean. Greater dispersion represented in larger value. F-test is often used when comparing statistical model that has been fitted to a data set. It uses the p-value and the degree of freedom of the two samples as input to calculate it using excel or F-table.

When F-statistics is greater than the F-test (F-table) value then the goodness of the fit is good[5].

2.6.5 P-value

P-value (probability value) specifies chance happened in results and the range is between 0 to 1. If significance value equals 0.05, then our confidence level is 0.95. If we increase significance value, the probability of incorrectly rejecting the null valued hypothesis and confidence level decrease. If the p-value is less than the significant value then the result is statistically significant[5].

2.6.6 T-test

A t-test is a form of inferential statistic which is used to decide if there is a significant difference between the means of two groups. A t-test can be found from t-table or calculated in excel by considering the degrees of freedom and p - value. Degrees of freedom are the number of values in a study that have the freedom to vary. The t-value is the ratio of variance between groups and variance with in groups. The larger the t score, the more difference between groups. The smaller the t score, the more similarity between groups.[5].

2.7 Related Works

Researchers in [5] [7] studied on BTS power consumption with traffic modeling and switching techniques of BTS in off- peak or in low traffic respectively on the case of ethio telecom. One third (1/3) of the total ICT's energy is consumed by the telecommunication equipment’s; cellular networks consume 90% of it. Base stations, in turn, consume 60 to 80% of the cellular network consumption.

The base stations switching off technique is done by turning off base stations remote radio unit (RRU) and applying antenna electrical down tilt and azimuth angles change on active base stations. [7]. In cellular network energy efficiency has an advantage in design new technology. Based on researches the base stations are the most power consumer part of cellular network. Due to this proper use of power in base stations is very important issue in green cellular networks.

The key objective of designing green base stations is for saving energy and reducing power consumption is for improving quality of service and energy efficient implementation. This can be succeeded by decreasing the base station energy consumption with different techniques like designing hardware, using sleep modes, energy aware cooperative base station power controlling with self-organizing cells, cell zooming according to traffic load, and applying renewable energy resource. So in green base stations power efficient solution analysis is concerned issue since base station is major power consumer in Mobil communication [36].

No	Year	Title	Objectives	Methods	Gaps
1	2015	monitoring and optimization of energy consumption of base transceiver stations	study and analysis of energy consumption BTS	simulation Monte Carlo method	Humidity not consider
2	2017	Green GSM Network Operation for Energy-Efficiency via Base stations Switching Technique	Reduce base stations power consumption and also study its performance	Neural network	Not consider determinate factors
3	2018	Traffic Modeling using Power Consumption of Base Station	study is to develop a model for BSs traffic load based on the power consumption	Linear regression	Power saving techniques not considered
4	2018	Energy assessment and optimization in second generation wireless access network, the case of ethio-telecom	Develop a model for the power consumption of BSs as per the carried traffic in a real time base station	Linear regression	Assume only voice calls
5	2018	Wireless sensor network based internet of things for Environmental impact analysis	develop a low cost sensor network	Real-time environmental monitoring application	Data size is not identified
6	2018	Impacts of Temperature and Humidity variations on RSSI in indoor Wireless	modification of temperature and humidity in indoor wireless network	Received signal Strength Indicator based applications	Data set not clearly set
7	2019	Building energy performance forecasting: A multiple linear regression approach	Provide an improved method that allows the evaluation of building energy performance	Multiple Linear Regression	Low data set used

Table 2-1 Summary of sample related works

2.8 Summary of literature review

The problem identified and the measure taken to solve the problem has been discussed in the above sections' literature review and related works. Many researchers have tried to study power consumption prediction using different machine learning and statistical approaches. As discussed in the first chapter, the contribution of this thesis is to study environmental impact on power consumption. To overcome this, the proposed work will address the solution using multiple linear regression machine learning algorithm and the gaps will be solved by doing this research work.

CHAPTER 3 METHODOLOGY

3.1 Over view

This chapter deals on selecting the appropriate research methods of data collection, data preprocessing, tools used, to carry the proposed work. In addition to this it covers approaches and methods, tools used, way of data collection, preparation techniques, data evaluation techniques will have done to conduct prediction for base transceiver station power consumption prediction.

3.2 Research procedures

The following Figure shows the procedures to achieve the general and specific objectives.

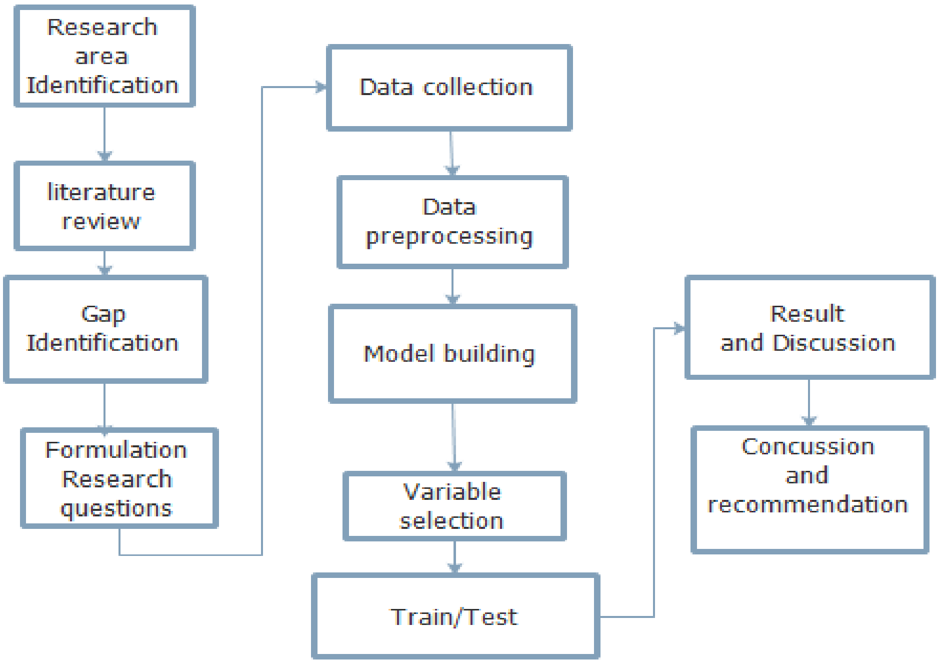

Figure 3-1 Methodology

3.3 Tools Used

In this research work hardware and software tools used in order to achieve the specific objective of the proposed work for implementing codes, data preparation and drawing diagrams easily and accurately.

3.3.1 Hardware tools

- Pc core i7
- RAM: 8GB
- Hard disk :1TB

3.3.2 Software tools

- Anaconda navigator python version 1.9.1.2
- Jupyter notebook version 6.0.1
- Arduino version1.6.12
- Protues 8.11
- Mendlay desktop version 1.19.4
- E-draw max: version 7.9

3.4 Data collection

BTS sites power system data are extract from ethiotelcom power management system which is known by ET net ECO (Network Ecosystem). The collected data has five-minute time interval.

3.5 Data set variables

Number of Variables collected from the system are 18. The variables are the following: site name, Subnet, Object Name, Collection Period, Outdoor Temperature(degC), start time, working temperature, working humidity, Battery current, Battery voltage, Battery temperature, Battery total discharge power, Battery cycle times, AC/DC system output current, Bus Bar voltage, DC load current, DC load power and week. From this features our target is studying the effect of environmental condition with load power (power consumption) thus independent and dependent variables are already known.

3.6 . Dataset Preprocessing

Data preprocessing is a fundamental techniques used to process complex data, it may consume large amounts of processing time[37]. The major purpose of data preparation is to manipulate and transform raw data so that the information content enfolded in the data set can made more easily accessible thus Data preprocessing techniques generate clear or quality data[38].

Data preprocessing is important for having the good prediction result unless it has impact on prediction quality. Each BTS site extracted from the monitoring system and prepared in time sequence. The collected data is numeric form and it is time based with five-minute interval.

3.7 Task of data preprocessing

3.7.1 Data Cleaning

Data cleaning is a process done by removing missing values or by filling empty space. smooth noisy data are data which is meaningless so it is not give information. Identify outlier used to remove unusual observation data from mass datasets. Data resolve inconsistencies and inappropriate data which decrease the accuracy of the data [39].

3.7.2 Data Integration

Data Integration applied to combine several data from different folder for having uniform information or to access in easily way. This is happened due to system limitation for example environmental features, battery features, load features are not in the same sheet so merging the columns was applied.

3.7.3 Data Reduction

Dimensionality reduction is method used to remove unimportant attributes. Reduction is the process of identifying and removing as much irrelevant and redundant information. Unimportant feature which is not to fit the objective of the study will identify and reduce in this research study[37].

3.7.4 Data set Outliers

outlier is an observation that is numerical value which is abnormal distance from the other of the data used ,and if we remove those outlier it shows different model result [40]. The interquartile range is one techniques of removing outlier. IQR ranking in 25% and 75% in a data set. The values are denoted as Q_1 and Q_3, respectively Quantile regression is an extension of the least squares method based on the classical conditional mean model. Thus only the data that satisfies the following conditions.

$$\mathbf{Q_1 - 1.5 * IQR \leq x \leq Q_3 + 1.5 * IQR} \quad (2)$$

where x is the residual between the estimated value and the true value, Q_1 and Q_3 are the quarter quantile and third-quarter quantile of the residual, respectively, and IQR is the interquartile range of the residual. Residuals outside this range are regarded as outliers based on quartile and interquartile range. It follows that the quartile is more resistant to disturbances since up to 25% of the data can be distributed as far as possible without greatly disturbing the quartile[41][42]. In the outlier result true indicates the outlier which far from other data or it is abnormal where as false means data set is normal so it used for implementation.

3.7.5. Normalization

Normalization basically used to change the values of numeric columns in the dataset to a common scale, without distorting differences in the ranges of values[37]. Min-max normalization is implemented in this research study.

3.8 Multiple linear regression algorithm

According to [32] multiple linear regression (MLR) method will used in this research study to analyze or predict the linear relationship between dependent variable (Y) and several independent variables or functions of independent variables (X).

3.8.1 Multiple linear regression assumptions

Data set assumption is conducted before multiple linear regressions. The following are criteria that associate with multiple regression those are Multicollinearity, Homoscedasticity, linearity, autocorrelation and normal distribution are assumptions to conducted in this research study.

The relationship between input variable and output variable is linear. According to this assumption if the linear dependency is violated then the model in hand is no more reliable. Multicollinearity (near-linear dependence), is a statistical phenomenon which shows the relationships between more than two predictors. But If there is no linear relationship between predictor variables, they are said to be orthogonal Multicollinearity[43] [44].

3.9 Data set split ratio

Data split is process of portioning data to develop model. Totally, ten (10) BTS sites data are collected for one hundred fifty days (from April to August 2019) by weekly data collection plan. The data set used for train and test contains four (4) features with a five (5) minute sampling period. The number of clear data is 184484 from this, the following split percentage is included.

- Train: 90%, validation :5% and Test: 5%
- Train: 80%, validation :10% and Test: 10%
- Train: 70%, validation :15% and Test: 15%
- Train: 60%, validation :20% and Test: 20%

Good training/testing ratio result applied in this research. So using the training set and then apply the model to the test set it gives best split ratio of the models. using this python code, we get the result.

```
import pandas as pd
df = pd.read_csv('/Users/HP/Desktop/best/five.csv')
print(df.head())
df1 = df[['Wtemp', 'Whumi', 'BusbarVol', 'DcloadCur', 'DcloadPow',]]
df2 = df1.dropna()
from sklearn. model selection import train_test_split
x_train, x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=0)
```

- Training Dataset is sample of data used to fit the model.
- Validation Dataset is the validation set is used to evaluate a given model.
- Test Dataset means sample of data used to provide an unbiased evaluation of a final model fit on the training dataset[45].

3.10 Evaluation Techniques

Evaluation techniques used to know how fit the actual data to predicated value the following are popular methods.

- Root means absolute error (RMSE): most appropriate techniques to measure the regression model's error rate.

$$\mathrm{RMSE} = \frac{\sqrt{\sum_{i}^{n} (yi - xi)^{2}}}{\mathrm{n}} \qquad (3)$$

x-actual value and y-predicted value

CHAPTER 4 IMPLEMENTATION

4.1 Overview

This chapter discusses on the proposed work to be done in multiple linear regression algorithm for base transceiver station power consumption prediction. To achieve this, model building, Dataset test, train and validation, assumptions of the algorithm, accuracy of the model will have implemented using python code. The table below shows sample preprocessed data.

Working Temp	Working Humidity	Battery Tempe	Bus Bar Volt	DC Load Current	Battery Volt	Battery Current	Battery Total Discharge	Battery Total Cycle	AC/DC System	DC Load Power(kW)
19.29	58.62	18.68	54.06	48.39	54.06	0	121362.7	421	49.2	2.62
19.19	55.97	18.37	54.06	50.68	54.12	0	55641.68	97	50.35	2.74
20.51	44.07	21.12	53.57	53.2	48.32	0	47438.77	115	55.2	2.85
26.31	30.95	25.6	53.5	62.03	53.55	0	17530.01	60	63.96	3.32
27.94	29.83	30.28	53.17	48.71	53.2	0	16630.59	68	50.39	2.59
18.99	55.97	17.76	54.1	59.58	54.19	0	64702.73	155	60.3	3.22
21.53	45.39	19.9	53.92	55.21	53.98	0	57435.76	183	56.78	2.98
20	32.37	19.6	53.99	66.34	54.05	0	44544.22	154	67.39	3.58
12.78	69.91	12.78	54.53	42.74	54.55	0	63060.2	169	44.15	2.33
19.29	52.01	18.48	54.09	49.26	54.09	0	121362.7	421	49.14	2.66
19.29	60.25	18.48	54.06	50.28	54.12	0	55641.68	97	50.47	2.72
20.71	43.56	21.43	53.56	52.75	48.32	0	47438.77	115	53.96	2.82
26.31	30.95	25.7	53.52	62.2	53.57	0	17530.01	60	63.45	3.33
27.84	29.83	30.28	53.13	47.95	53.19	0	16630.59	68	49.82	2.55
19.39	55.26	17.76	54.09	56.19	54.19	0	64702.73	155	57.77	3.04
21.12	29.63	19.19	53.97	54.04	54.03	0	57435.76	183	54.51	2.92
20	35.22	19.6	53.99	64.35	54.05	0	44544.22	154	65.8	3.47
13.7	73.78	14	54.43	42.26	54.45	0	63060.2	169	43.72	2.3
19.09	55.57	18.48	54.08	47.04	54.08	0	121362.7	421	49.31	2.54
19.39	59.53	18.58	54.03	50.86	54.1	0	55641.68	97	51.18	2.75
20.51	44.07	21.02	53.58	51.79	48.3	0	47438.77	115	53.3	2.77
26.31	30.95	25.7	53.51	61.56	53.55	0	17530.01	60	63.25	3.29
27.84	29.93	30.28	53.15	48.15	53.2	0	16630.59	68	49.92	2.56
19.29	42.14	17.66	54.1	57.04	54.19	0	64702.73	155	59.1	3.09

Table 4-1 Sample preprocessed data

As discuss in chapter 3 we implement outlier removal technique, Thus the data set summery is as follows. As shown below in Table 4-2 true outlier indicates is an observation that lies an abnormal distance from other values from the data sets. Whereas false shows non outlier which used for implementation. Variables total data true outlier false outlier aggregate after outlier.

Working temperature	339500	22379	317121	184484
Working humidity	339500	299	339201	184484
Battery temperature	339500	14413	325087	184484
Bus Bar Voltage	339500	60163	279337	184484
Dc load Current	339500	1113	338387	184484
Battery Voltage	339500	56481	283019	184484
Battery current	339500	58297	281203	184484
Battery discharge power	339500	102233	237267	184484
Battery total cycle	339500	33921	305579	184484
AC DC Current	339500	40008	299492	184484
DC load Power	339500	1041	338459	184484

Table 4-2 Data set with outlier

The maximum and minimum data record of the variables is shown in the the table blow. Before outlier data describe the range of each variables including outlier. Whereas after outlier means data excluding outlier.

Data set	Temperature (°C)		Humidity (%)		Voltage(v)		DC current(A)		Dcpower load(kw)	
Before outlier	Max	Min	Max	min	Max	Min	Max	min	max	Min
	42	10.64	82.52	0.13	56.51	43.91	83.66	1.12	4.17	0.05
After outlier	29.16	12.58	80.79	9.79	54.66	53.25	77.65	28.29	4.08	1.82

Table 4-2 Data range value of sample variables

4.2 Working with Criteria for Implementing Multiple Linear Regression

4.2.1 Correlation Result of Variables

The correlation coefficient or Pearson correlation coefficient is a numerical index of the strength of relationship between variables. Eleven variables were tested in python to check whether this assumption for implementing multiple linear regression. The following python code implemented for correlation result.

```
import pandas as pd
df = pd.read_csv('/Users/HP/Desktop/best/eleven.csv')
print(df.head())
df1=df[['Wtemp','Whumi','Battep','BusbarVol','DcloadCur','BatVol','Batcur','Battdi
schp','Batttcycl','AcdcoCur', 'DcloadPow',]]
df2 = df1.dropna()
corr=df2.corr()
print(corr)
```

Table 4-3 below show correlation output of the variables which is done in python code. The result shows the correlation strength and direction of the relationship between the variables. The correlation coefficient ranges between -1 to +1, thus -1 indicate negative correlation, and +1 indicating positive correlation.

	Wtemp	Whumi	Battep	BusbarVol	DcloadCur	BatVol
Wtemp	1.00	-0.70	0.90	-0.9	0.6	-0.8
Whumi	-0.70	1.00	-0.58	0.59	-0.58	0.5
Battep	0.90	-0.58	1.00	-0.99	0.54	-0.9
BusbarVol	-0.9	0.59	-0.99	1.00	-0.55	0.98
DcloadCur	0.61	-0.58	0.54	-0.55	1.00	-0.49
BatVol	-0.86	0.54	-0.97	0.98	-0.49	1.00
Batcur	NaN	NaN	NaN	NaN	NaN	NaN
Battdischp	-0.35	0.39	-0.28	0.27	-0.34	0.22
Batttcycl	-0.25	0.22	-0.18	0.18	0.16	0.11
AcdcoCur	0.61	-0.58	0.56	-0.57	0.99	-0.50
DcloadPow	0.59	-0.57	0.52	-0.53	0.99	-0.47

	Batcur	Battdisch	Batttcycl	AcdcoCur	DcloadPow
Wtemp	NaN	-0.35	-0.25	0.61	0.59
Whumi	NaN	0.39	0.22	-0.58	-0.57
Battep	NaN	-0.28	-0.18	0.56	0.52
BusbarVol	NaN	0.27	0.18	-0.57	-0.53
DcloadCur	NaN	-0.34	0.16	0.99	0.99

Table 4-3 Total Correlation result of variables

Result of correlation analysis is based on the following interpretation:

From 0.1 to 0.39 is weak,0.40 to 0.60 moderate, 0.70 to 0.89 strong and from 0.90 and above is very strong[46].

According to this guide line Dcloadcurrent and ac/dc current has very strong correlation where as temperature, humidity, battery temperature, basbar voltage battery voltage has moderate correlation. Battery current, battery discharge, battery cycle has weak correlation with dcload power. In the table 4.4 below shows that correlation result of the target variable.

	Wtemp	Whumi	Volt	DcloadCur	DcloadPow
Wtemp	**1.00**	-0.70	-0.91	0.61	0.59
Whumi	-0.70	**1.00**	0.59	-0.58	-0.57
Voltage	-0.91	0.59	**1.00**	-0.55	-0.53
DcloadCur	0.61	-0.58	-0.55	**1.00**	0.99
DcloadPow	0.59	-0.57	-0.53	0.99	**1.00**

Table 4-4 correlation result

4.2.2 Scatter Plot (Correlation analysis)

Scatter plot can analysis our model to understand the relation between variables. When the graph shows upward it is positive correlation. If it is downward pattern it indicated negative correlation. Where as if it is zero variables are uncorrelated. From this we can understand Working temperature and Dcloadcurrent has positive relation with Dc load power, working humidity negatively related to Dcloadpower.

The code below applied to check the scatter plot graph.

```
import seaborn as sns
sns.set(style='ticks', color_codes=True, font_scale=2)
g=sns.pairplot(df2, height=3, diag_kind='hist', kind='reg')
g.fig.suptitle('Scatter Plot', y=1.08)
```

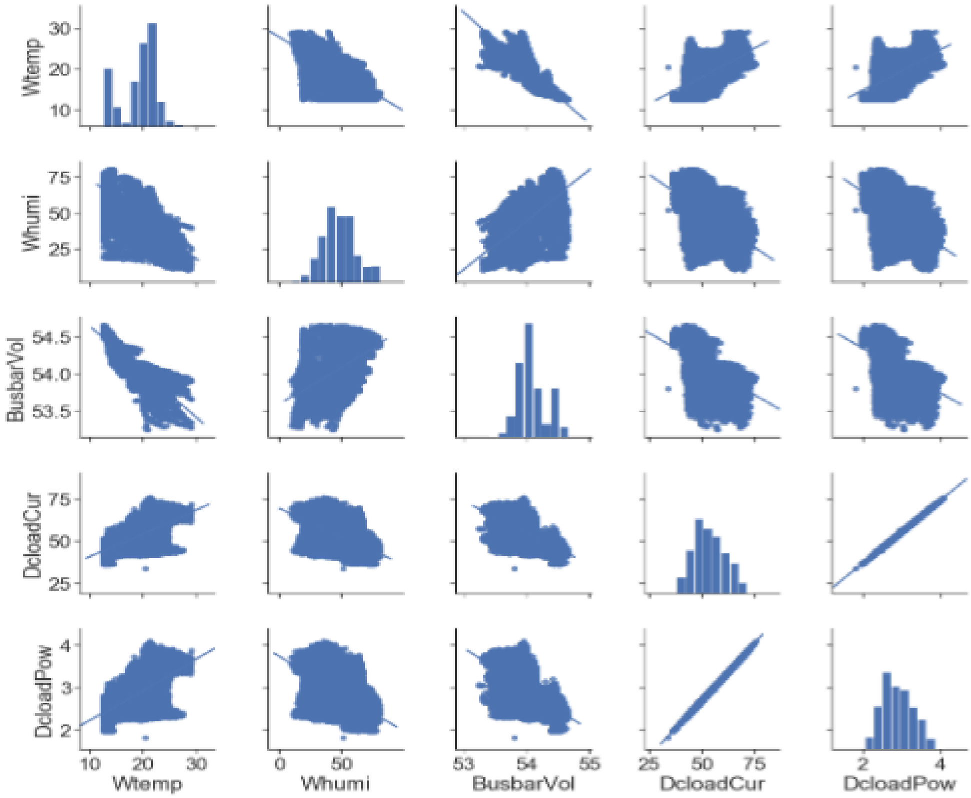

Figure 4-1Scatter plot result of target variables

4.2.3 Homoscedasticity

Homoscedasticity recommend equal level variability between dependent and independent variables by produce scatter plot. residuals plot is the same width for all predicted values.

```
import matplotlib.pyplot as plt
fig, ax=plt.subplots(figsize=(8,3.5))
pred_val=reg.fittedvalues.copy()
true_val=df2['DcloadPow'].values.copy()
resid=true_val-pred_val
sns.residplot(resid,pred_val)
plt.title('Homoscedasticity')
```

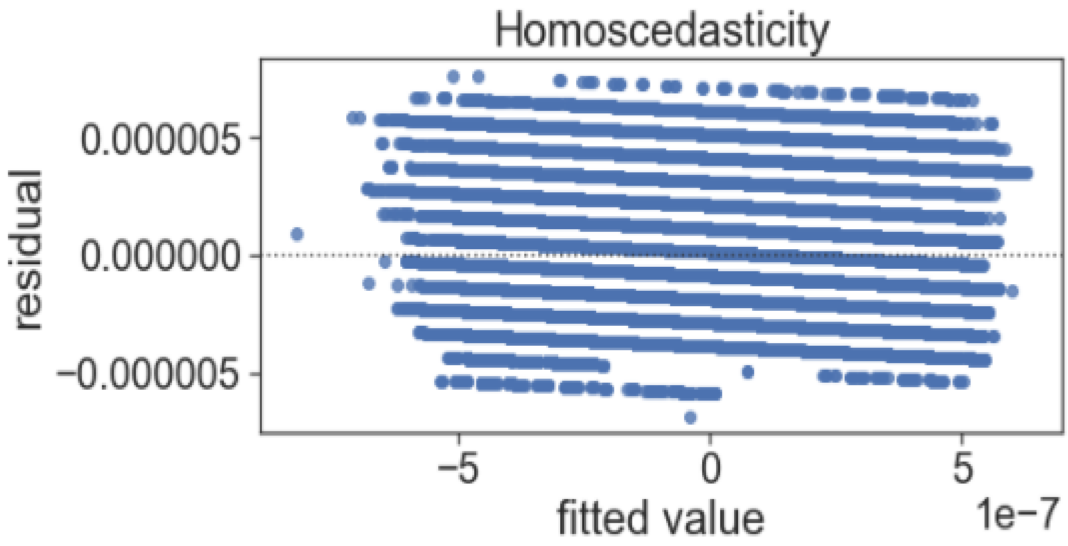

Figure 4-2 Homoscedasticity graph result target variables

Homoscedasticity graph show the set of data distance from the line, the figure 4-2 indicates how far is from the line or having the same scatter for it to exist in a set of data, the points must be about the same distance from the line the assumption of homoscedasticity describes same variance in the regression model. Data are homoscedastic if the residuals plot is the same width. Fitted value describes where a particular x-value fits the line of best fit.

4.2.4 Probability plot

Probability plot graph shows how model distributed normally with graph and identified outlier values. The following code is implemented for having probability result.

```
import scipy.stats as stats
fig, ax=plt.subplots(figsize=(10,6))
stats. probplot(resid, dist='norm', plot=plt)
plt. Show
```

The probability plot shows straight line; we can decide the observed data set come from normal distribution as shown Figure 4.3 it deviate line is the error but little from the line. Theoretical quantile plot is a graphical way of checking whether data are normally distributed or not. Normally distributed data shows a nearly straight line.

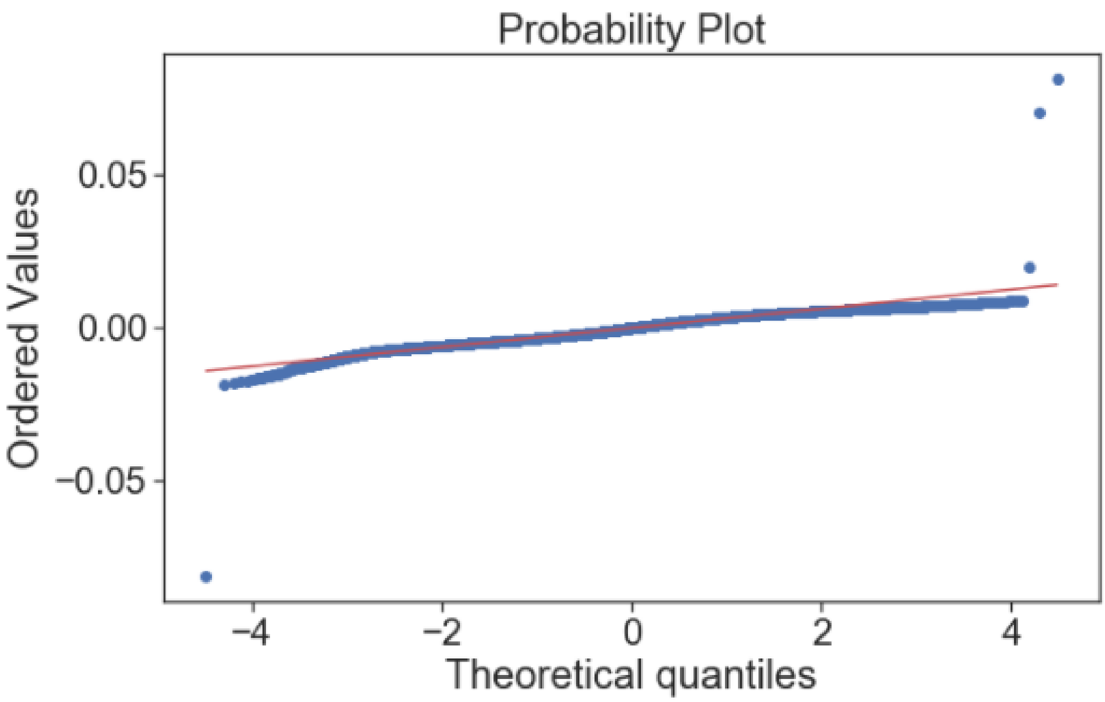

Figure 4-3 Probability plot result target variables

No of sites	No of Variables	Reduced Variables	Data size	Missed data	Total data
10	18	7	341265	1765	339500

Table 4-5 Summary of data sets

4.3 Model Building

In this paper, we propose multiple linear model based on machine learning by using power data. As per previous observations, there is a linear relationship between the dependent variable and independent variables in the data set. We therefore, use linear model technique to develop a linear BS power consumption model based on our measurement results. A power consumption model is developed based on the dependent and the independent features

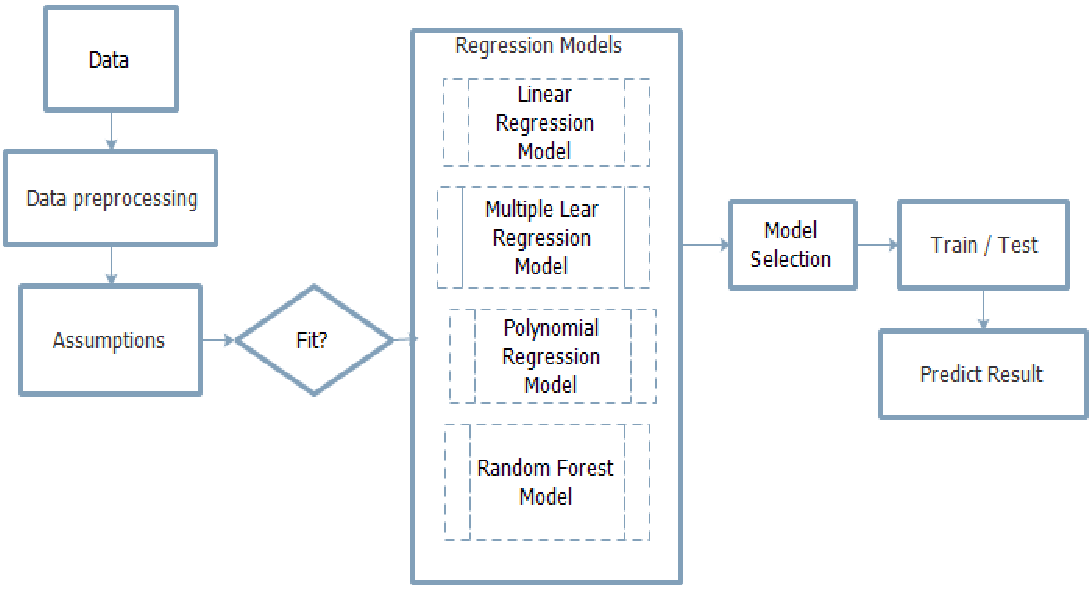

Figure 4-4. prediction model Architecture

We use the regression equation to predict the value of the dependent variable from the values of the independent variable.

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\ldots+\beta_nX_n+\epsilon$$

$$\mathbf{Dcloadpower}=\beta_0+\beta_1\times\mathbf{temp}+\beta_2\times\mathbf{humidity}+\beta_3\times\mathbf{voltage}+\beta_4\times\mathbf{dcloadcurrent}+\epsilon$$

- Dcloadpower: predictor or feature
- β0: intercept
- $\beta 1$, $\beta 2$, $\beta 3$, β_4: slope coefficient
- ϵ: A residual (or the error term) is the difference between the observed response y and the predicted response

- Dcloadpower is output variable and temperature, humidity, voltage and dc load current: are input variables.

4.3.1 Estimating Coefficients and intercept

Coefficients is change in the value of Y for each one unit change in X where Intercept is line where the regression cross the y-axis. The code below gives us the output of coefficient and intercept.

```
import pandas as pd
df = pd.read_csv('/Users/HP/Desktop/best/five.csv')
print(df.head())
df1 = df[['Wtemp', 'Whumi', 'BusbarVol', 'DcloadCur', 'DcloadPow',]]
df2 = df1.dropna()
from statsmodels.formula.api import ols
reg=ols (' DcloadPow~Wtemp+Whumi+BasbarVoltage+DcloadCur ‘, data=df2). fit ()
print (reg. summary ())
```

As it has been displayed through executing the above code, the multiple linear regression model (the equation) is the following

Dcloadpower = -2.741+ (0.0002) * temp+ (-0.000065) * hum+ (0.0507) * Basbar volt + (0.0540) *DcloadCur+0.0032

CHAPTER 5 RESULT, EVALUATION AND DISCUSSION

5.1 Overview

In this chapter, we show the result of the analysis and evaluate in detail which obtained from the implementation of the proposed solution for power consumption prediction using multiple linear regression for based transceiver station. Finally, we discuss the result obtained from each experiment in this study.

5.2 Data Split Ratio

Data splitting technique is used for evaluating the performance of a machine learning model. Data split ratio is checked in the following data split ratio.

- 60:40,
- 70:30,
- 80:20,
- 90:10

Data split ratios are checked using linear regerssion , multiple regerssion, polynomial regression and random forest regression models are applied and 80:20 split ratio in multiple linear regression model has better performance based on accuracy result.

From the four regression models all of regression models can said to be good but randomforest is relativley less comare to others models. linear regerssion and polynomial regression are good next to random forest.

R-square indicates how our data set fit to the regression model, higher r-square shows better fit to the model. Four type of split ratio applied as states in the table 5.1.

Model type	Split ratio %	R-square	RMSE
Linear regression	60/40	0.99944	0.0093
	70/30	0.9994	0.0093
	80/20	0.99944	0.00924
	90/10	0.99943	0.00928
Multiple regression	60/40	0.999934	0.00066
	70: 30	0.999933	0.00067
	80/20	0.99995	0.0006
	90/10	0.99993	0.00066
Polynomial regression	60/40	0.99994	0.00291
	70/30	0.99994	0.0029
	80/20	0.999948	0.0028
	90/10	0.999945	0.0029
Random forest	60/40	0.9999	0.0038
	70/30	0.9999	0.0034
	80/20	0.9999	0.0023
	90/10	0.9999	0.0032

Table 5-1 Different model split ratio

5.3 Prediction result

Actual value is the value which is achieved by measuring the available data. Predicted value shows a value estimated based on regression analysis and difference value is the difference between the predicted value and the observed value the actual value is the value that is obtained by observation or by measuring the available data. It is also called the observed value.

Actual value(kw)		Predicted value(kw)	Differnce
2.85		2.848303	0.001697
2.51		2.509680	0.000320
2.87		2.863938	0.006062
2.36		2.360939	-0.000939
2.72		2.719469	0.000531
2.33		2.333575	-0.003575
3.18		3.176354	0.003646
2.31		2.314762	-0.004762
2.68		2.680115	-0.000115
2.92		2.925437	-0.005437
2.68		2.679934	0.000066
2.95		2.953857	-0.003857
3.41		3.413844	-0.003844
3.52		3.524777	-0.004777
2.60		2.596483	0.003517
2.70		2.697834	0.002166
2.81		2.812776	-0.002776
3.19		3.186290	0.003710
2.33		2.329191	0.000809
2.81		2.806175	0.003825

Table 5-2 Actual and predicted value

The graph below shows blue color indicates the actual value and orange color is the predicted value. So we can understand predicted value is fitted to actual value it is good model.

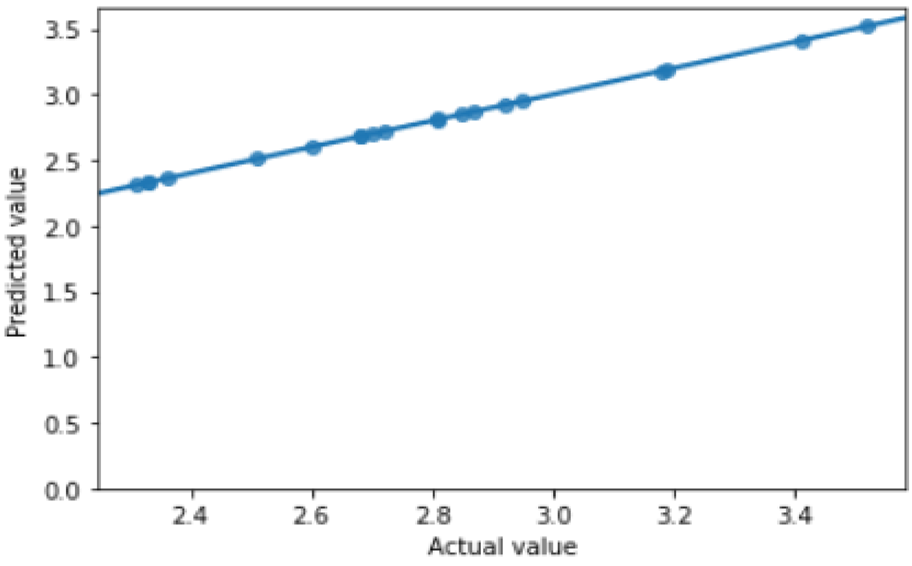

Figure 5-1 Actual and predicted value graph

5.4 Performance Evaluation Result

The performance of the proposed model is measured by root-mean-squared error (RMSE).

The Root Mean Squared Error result is 0.0032. The closer the values of RMSE to zero is more accurate the model results will be preferred as better model performance.

From the result we can understand the RMSE value prediction accuracy of the proposed model is very satisfactory.

5.5 Prototype development

This section presents the use of the Proteus environment for the study of Arduino-driven sensor circuits design. So environmental variables such as temperature, humidity and voltage has controlled on help of ICT and hardware and software tasks will combine and implement using Arduino Uno and the result displayed on LED and LCD display.

5.6 Hardware components

The hardware component has been substantiated and implemented, and the software of the system physical model has been developed by using modern Arduino Uno and sensor technologies. [47].

- DHT11 Sensor is most regularly electronics device used for environmental monitoring and it converts a change in physical occurrence into an electrical signal. DHT 11 sensor transfer humidity and temperature information to arduino.

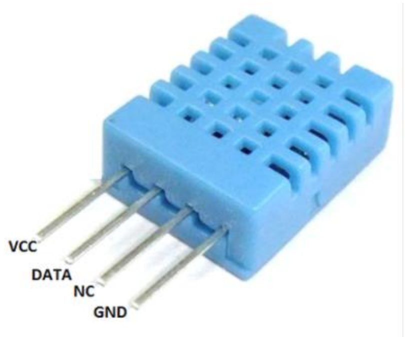

Figure 5-2 DHT 11 Sensor [48]

- LCD ((Liquid Crystal Display): Display information which send from sensor, the humidity level, and temperature.

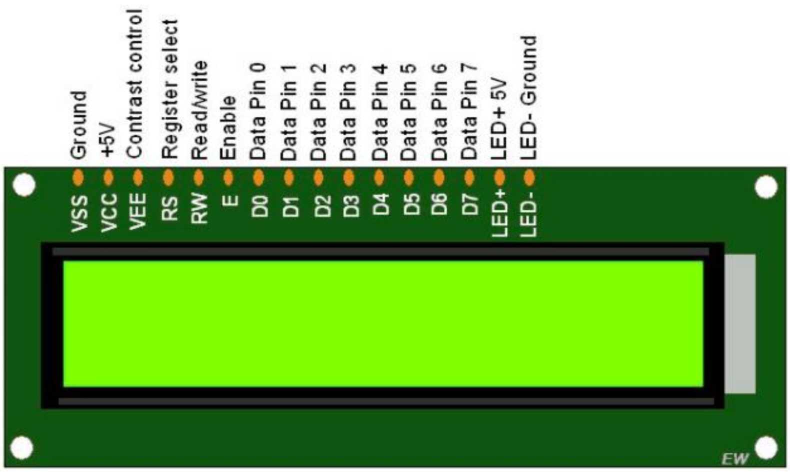

Figure 5-3 LCD [49]

- In the electrical circuit other components applied such as resister to drop voltage to protect the arduino from high voltage. DC fan is direct current fan used for cooling purpose. PCB Board- Stand for printed circuit board which supports by connecting electronic components electrically mechanically using conductive tracks from copper sheets laminated onto a non-conductive substrate[50].

5.7 Software Tools

Arduino is a newly developed open source hardware and software system. The Arduino software IDE (Integrated Development Environment) provides space to write codes in the language (programming languages C, C++)[48]

Figure 5-4 Arduino Uno   [48]

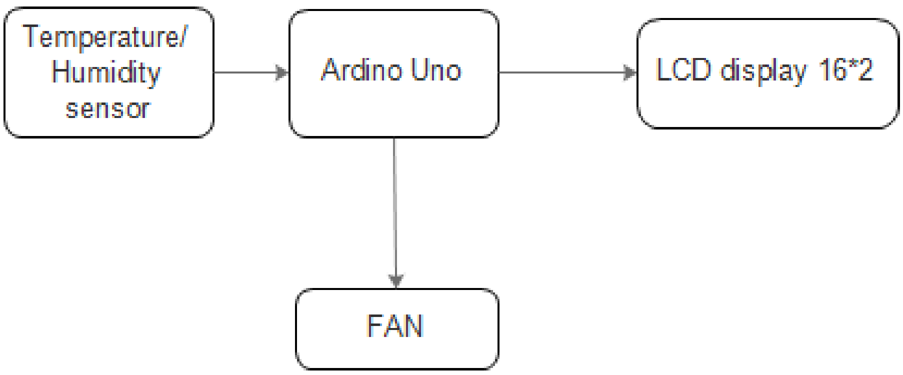

Figure 5-5 Temperature and humidity hardware control block diagram [51]

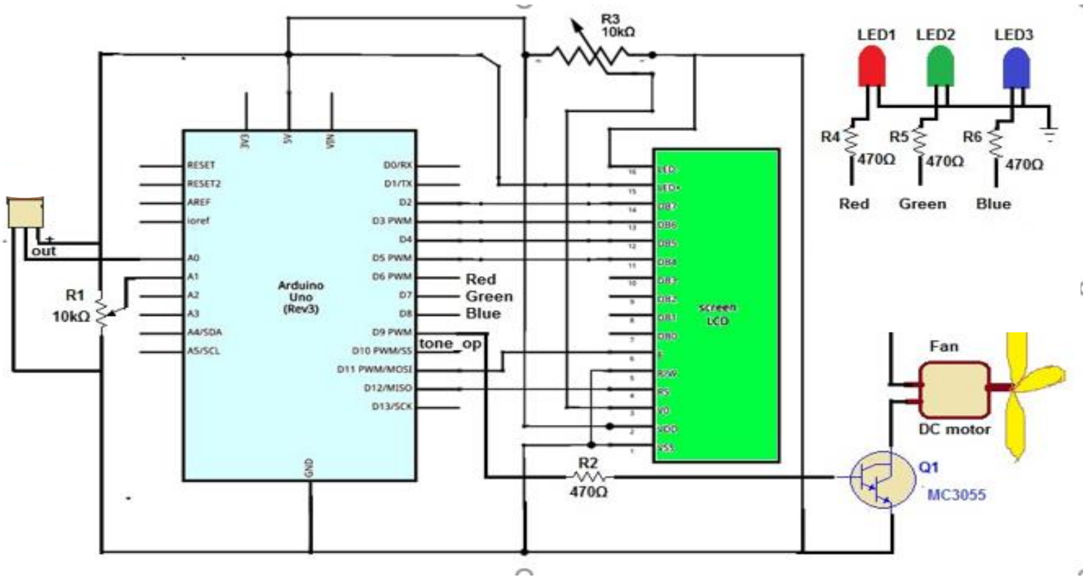

Figure 5-6 Temperature control system circuit diagram [51]

5.8 Temperature and humidity on PCB board

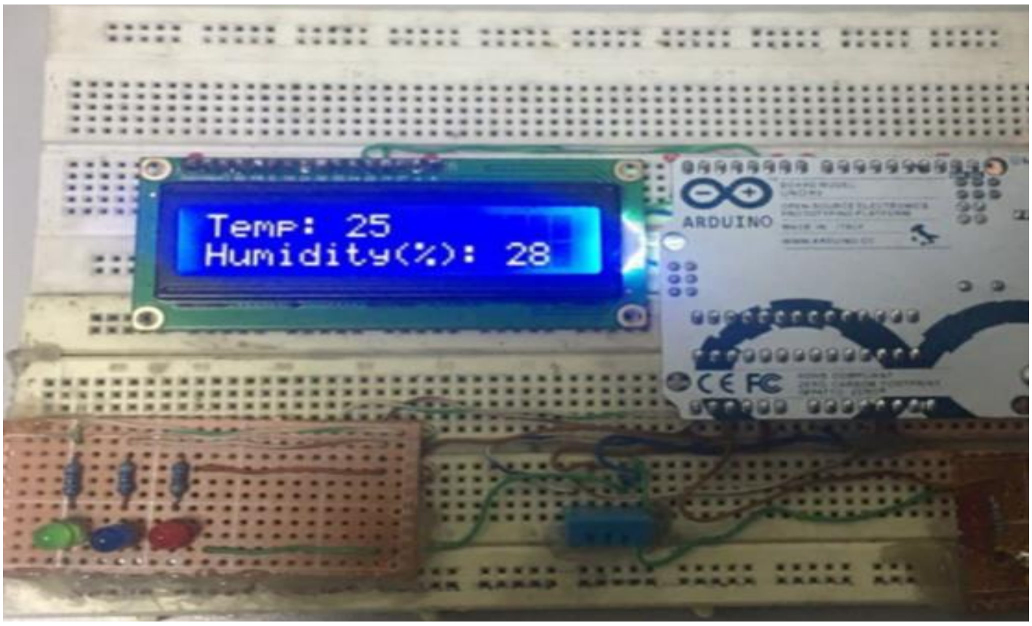

Figure 5-7 Temperature and humidity display

5.9 Voltage interface to lcd and led prototype

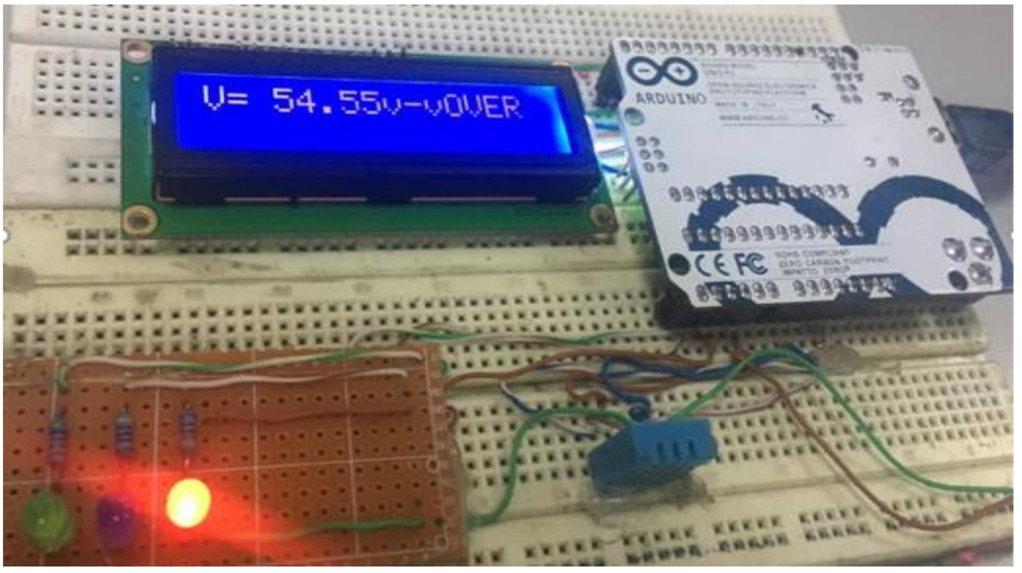

Figure 5-8 Voltage level indicator

The following figure shows interface of power consumption prediction models.

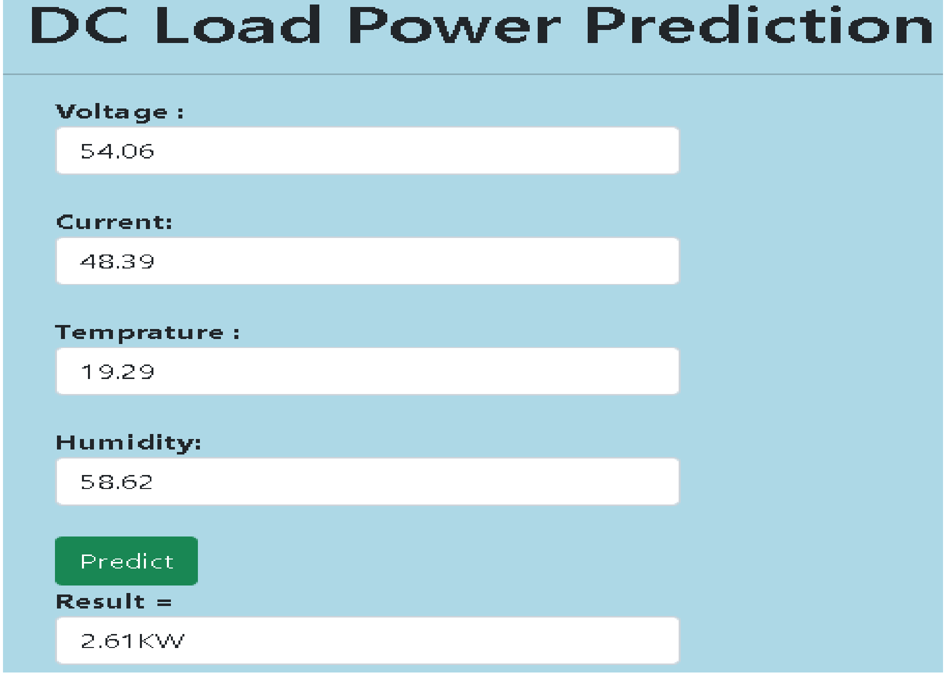

Figure 5-9 Model prediction

5.10 Discussion

The relationship between the environmental features and the power consumption of different base stations type is obtained. As shown on Table 5 3 the data of power consumption of 10 different base stations is collected for 150 days. There were total 187484 sample values for dependent and independent variables.

The maximum power consumption as shown on Table 5 4 is 4.17 kw and minimum power consumption is 0.05 kw. The range of power consumption is change in temperature and humidity varied. The upper bound of temperature is 42oc and lower bound is 10.64 oc where as upper bound of humidity is 82.52% and the lower bound is 0.13%. the temperature has positive correlation thus increase in temperature dc power load or power consumption also increase, whereas humidity has negative correlation so dc power consumption decrease when humidity increase. Temperature and humidity has moderate correlation with dc power load, and dc current has strong correlation with power consumption.

The effect of temperature change with Dc power consumption is 0.0084 and humidity effect is -0.0054 per one kw. Voltage has 0.288 effect with DC load power when it changes in one unit where as current 0.054 coefficient. Other determinant factor of power consumption such as Battery current, Battery temperature, Battery total discharge power, Battery cycle times, has positive correlation and Battery voltage has negative correlation and relatively weak correlation with DC power consumption. Controlling temperature and humidity to the recommended value support the system performance. So cooling mechanism should be operate based on sensing environmental features variation. This minimize the amount of power cost and it support for the network reliability. As increase voltage has high power consumption thus controlling voltage minimize the amount of power usage.

The model which employed using multiple regression techniques has a prediction performance accuracy of 99.5 % when evaluated using RMSE (Root Mean Square Error).

Conclusion

The exponential growth in the mobile communication technology increased the demand for base transceiver station power consumptions. So saving power in base stations is the major focus in green cellular network development.

This paper studies determinant factors like temperature, humidity voltage and dc current effect were included on the research. Data were collected for 150 days in 24 hours' basis in 5-minute interval. Due to goodness of model multiple linear regression prediction approach has been implemented for each base transceiver station.

The proposed model presents a significant result and can be acknowledged as a model for base transceiver station power consumption under real power consumption.

Dcloadcurrent current has very strong correlation where as temperature, humidity, basbar voltage has moderate correlation dcload power. The effect of temperature change with power consumption is 0.0084 and humidity effect is -0.0054 per one kw. Voltage has 0.288 effect with DC load power when it changes in one unit where as current 0.054 coefficient.

The model result can be further taken for prediction of the power usage of a base transceiver station from its annual DC load power forecast data of the planning strategy. Notifying the status of the variables used as information for new optimization. The results justify the use of multiple linear regressions as an alternative method, issuing a simple and immediate tool that can solve a complex datasets building energy efficiency prediction, thereby accelerating and helping some evaluation phases in energy planning presenting a valid criterion that could be indicated in standards and laws in the field of the building energy performance.

Finally Based on temperature, humidity, voltage and current coefficients we got significant effect on base transceiver station power consumption per kilowatt(kw).

Recommendation

Based on the study of this power consumption model it is an important to recommend those recommendations:

- Notifying base stations environmental features in during operation is one of the strategy for having information about BTS power system, so this model will help to identify the effect of those features with DC power load.
- This research used for scheduling and analyzing the network infrastructure based on power usage.

Future work

Future research will be focused on improving the following concepts.

- Study different regional site data set is important for comparison of environmental features impacts of the base transceiver station.
- Studying generator fuel consumption and Ac power input improve the research and also important for the company to have strategy in power efficiency.
- Implement other type of algorithm will give the study more advantages.

REFERENCE

[1] Asmelash Tesfay, "Neural Network based 3G Mobile Sites Fault Prediction : A Case Study in Addis Ababa , Ethiopia Neural Network based 3G Mobile Sites Fault Prediction : A Case Study in Addis Ababa , Ethiopia," 2018.

[2] S. Jovanović, S. Savić, M. Bojić, Z. Djordjević, and D. Nikolić, "The impact of the mean daily air temperature change on electricity consumption," *Energy*, vol. 88, no. June, pp. 604–609, 2015, doi: 10.1016/j.energy.2015.06.001.

[3] L. Wang, G. Von Laszewski, F. Huang, J. Dayal, T. Frulani, and G. Fox, "Task scheduling with ANN-based temperature prediction in a data center: A simulation-based study," *Eng. Comput.*, vol. 27, no. 4, pp. 381–391, 2011, doi: 10.1007/s00366-011-0211-4.

[4] G. K. Uyanık and N. Güler, "A Study on Multiple Linear Regression Analysis," *Procedia - Soc. Behav. Sci.*, vol. 106, pp. 234–240, 2013, doi: 10.1016/j.sbspro.2013.12.027.

[5] T. Hagos, D.-I. D. Hailemariam, and A. Ababa, "Addis Ababa Institute of Technology School of Electrical and Computer Engineering Traffic Modeling using Power Consumption of Base Station: The case of Ethio telecom," 2018.

[6] Y. Y. Tefera, T. Kibatu, B. S. Shawel, and D. H. Woldegebreal, "Recurrent Neural Network-based Base Transceiver Station Power Supply System Failure Prediction," *Proc. Int. Jt. Conf. Neural Networks*, no. December, 2020, doi: 10.1109/IJCNN48605.2020.9206978.

[7] M. Mesfin, "Green GSM Network Operation for Energy-Efficiency via Basestations Switching Technique: The Case of Addis Ababa, Ethiopia," *Green GSM Netw. Oper. Energy-Efficiency via Basestations Switch. Tech. Case Addis Ababa, Ethiop.*, vol. 13, no. 3, pp. 1576–1580, 2017.

[8] Z. Shang, D. Gao, Z. Jiang, and Y. Lu, "Towards less energy intensive heavy-duty machine tools: Power consumption characteristics and energy-saving strategies," *Energy*, vol. 178, no. April, pp. 263–276, 2019, doi: 10.1016/j.energy.2019.04.133.

[9] P. Giri and S. S. Singh, “An Analytical Survey of Power Consumption and Modeling in Different Areas of ICT Networks,” vol. 6, no. 1, pp. 866–870, 2015.

[10] D. Palani and M. Arulraj, “Energy Efficient Solutions for Green Cellular Networks – a Survey,” vol. 9, no. 5, pp. 1–16, 2018.

[11] B. Teshale, “Energy assessment and optimization in second generation wireless access network , the case of ethio-telecom,” 2018.

[12] N. Zarka, “BTS , BSC , MSC details Raghda Khadoor,” no. February, 2016, doi: 10.13140/RG.2.1.2075.6241.

[13] P. A. Ochang and P. J. Irving, “Evolutionary Analysis of GSM , UMTS and LTE Mobile Network Architectures,” *world Sci. New*, vol. 54, pp. 27–39, 2016, [Online]. Available: www.worldscient i ficnews.com.

[14] J. Tack, A. Barkley, and L. L. Nalley, “Effect of warming temperatures on US wheat yields,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 112, no. 22, pp. 6931–6936, 2015, doi: 10.1073/pnas.1415181112.

[15] K. Qian, C. Zhou, Y. Yuan, and M. Allan, “Temperature effect on electric vehicle battery cycle life in vehicle-to-grid applications,” in *2010 China International Conference on Electricity Distribution, CICED 2010*, 2010, pp. 1–6.

[16] F. Leng, C. M. Tan, and M. Pecht, “Effect of Temperature on the Aging rate of Li Ion Battery Operating above Room Temperature,” *Sci. Rep.*, vol. 5, pp. 1–12, 2015, doi: 10.1038/srep12967.

[17] X. Tang, K. Yao, B. Liu, W. Hu, and F. Gao, “Long-term battery voltage, power, and surface temperature prediction using a model-based extreme learning machine,” *Energies*, vol. 11, no. 1, 2018, doi: 10.3390/en11010086.

[18] T. Bhattacharya, A. K. Chakraborty, and K. Pal, “Effects of Ambient Temperature and Wind Speed on Performance of Monocrystalline Solar Photovoltaic Module in Tripura, India,” *J. Sol. Energy*, vol. 2014, pp. 1–5, 2014, doi: 10.1155/2014/817078.

[19] T. K. Routh *et al.*, “Artificial neural network based temperature prediction and its

impact on solar cell," *2012 Int. Conf. Informatics, Electron. Vision, ICIEV 2012*, no. May 2019, pp. 897–902, 2012, doi: 10.1109/ICIEV.2012.6317369.

[20] D. Lee and K. Kim, "Recurrent neural network-based hourly prediction of photovoltaic power output using meteorological information," *Energies*, vol. 12, no. 2, 2019, doi: 10.3390/en12020215.

[21] A. Spagnuolo, A. Petraglia, C. Vetromile, R. Formosi, and C. Lubritto, "Monitoring and optimization of energy consumption of base transceiver stations," *Energy*, vol. 81, no. March, pp. 286–293, 2015, doi: 10.1016/j.energy.2014.12.040.

[22] C. A. Boano, N. Tsiftes, T. Voigt, J. Brown, and U. Roedig, "The impact of temperature on outdoor industrial sensornet applications," *IEEE Trans. Ind. Informatics*, vol. 6, no. 3, pp. 451–459, 2010, doi: 10.1109/TII.2009.2035111.

[23] S. S. Baboo and I. K. Shereef, "ANN_AnEfficientWeather_ForecastingSystem," *Int. J. Environ. Sci. Dev.*, vol. 1, no. 4, pp. 321–326, 2010.

[24] W. Zaremba, I. Sutskever, and O. Vinyals, "Recurrent Neural Network Regularization," no. 2013, pp. 1–8, 2014, [Online]. Available: http://arxiv.org/abs/1409.2329.

[25] C. Iyigun *et al.*, "Clustering current climate regions of Turkey by using a multivariate statistical method," *Theor. Appl. Climatol.*, vol. 114, no. 1–2, pp. 95–106, 2013, doi: 10.1007/s00704-012-0823-7.

[26] A. Ghofrani, S. D. Nazemi, and M. A. Jafari, "Prediction of building indoor temperature response in variable air volume systems," *J. Build. Perform. Simul.*, vol. 13, no. 1, pp. 34–47, 2020, doi: 10.1080/19401493.2019.1688393.

[27] J. Lorincz, T. Garma, and G. Petrovic, "Measurements and modelling of base station power consumption under real traffic loads," *Sensors*, vol. 12, no. 4, pp. 4281–4310, 2012, doi: 10.3390/s120404281.

[28] J. Alzubi, A. Nayyar, and A. Kumar, "Machine Learning from Theory to Algorithms: An Overview," *J. Phys. Conf. Ser.*, vol. 1142, no. 1, 2018, doi: 10.1088/1742-6596/1142/1/012012.

[29] Y. Baştanlar and M. Ozuysal, *Introduction to Machine Learning Second Edition*, vol. 1107. 2014.

[30] M. G. H. Omran, A. P. Engelbrecht, and A. Salman, “An overview of clustering methods,” *Intell. Data Anal.*, vol. 11, no. 6, pp. 583–605, 2007, doi: 10.3233/ida-2007-11602.

[31] M. Awad and R. Khanna, *Machine Learning in Action: Examples*. 2015.

[32] M. Sakah, S. de la Rue du Can, F. A. Diawuo, M. D. Sedzro, and C. Kuhn, “A study of appliance ownership and electricity consumption determinants in urban Ghanaian households,” *Sustain. Cities Soc.*, vol. 44, no. October, pp. 559–581, 2019, doi: 10.1016/j.scs.2018.10.019.

[33] S. Kumar and I. Chong, “Correlation analysis to identify the effective data in machine learning: Prediction of depressive disorder and emotion states,” *Int. J. Environ. Res. Public Health*, vol. 15, no. 12, 2018, doi: 10.3390/ijerph15122907.

[34] J. N. Sa’eed Kewan, “Estimation of Multivariate Multiple Linear Regression Models and Applications,” p. 168, 2015.

[35] M. A. Bujang, N. Sa’at, and T. M. I. T. A. B. Sidik, “Determination of minimum sample size requirement for multiple linear regression and analysis of covariance based on experimental and non-experimental studies,” *Epidemiol. Biostat. Public Heal.*, vol. 14, no. 3, pp. e12117-1-e12117-9, 2017, doi: 10.2427/12117.

[36] C. Murthy and C. Kavitha, “A Survey of Green Base Stations in Cellular Networks,” *Ijcnwc.Org*, vol. 2, no. 2, pp. 232–236, 2012, [Online]. Available: http://www.ijcnwc.org/papers/vol2no22012/20vol2no2.pdf.

[37] S. García, S. Ramírez-Gallego, J. Luengo, J. M. Benítez, and F. Herrera, “Big data preprocessing: methods and prospects,” *Big Data Anal.*, vol. 1, no. 1, pp. 1–22, 2016, doi: 10.1186/s41044-016-0014-0.

[38] J. Han, M. Kamber, and J. Pei, “Data Preprocessing,” *Data Min.*, pp. 83–124, 2012, doi: 10.1016/b978-0-12-381479-1.00003-4.

[39] H. Yang, “Data Preprocessing-Chapter 3,” *Citeseerx*, 2013, [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.467.29&rep=rep1&type=pdf.

[40] S. M. A. K. Rahman, M. M. Sathik, and K. S. Kannan, “Multiple Linear Regression Models in Outlier Detection,” *Int. J. Res. Comput. Sci.*, vol. 2, no. 2, pp. 23–28, 2012, doi: 10.7815/ijorcs.22.2012.018.

[41] J. Yang, S. Rahardja, and P. Fränti, “Outlier detection: How to threshold outlier scores?,” *ACM Int. Conf. Proceeding Ser.*, pp. 1–6, 2019, doi: 10.1145/3371425.3371427.

[42] C. Zhang, Y. Wang, Y. Gao, F. Wang, B. Mu, and W. Zhang, “Accelerated fading recognition for lithium-ion batteries with Nickel-Cobalt-Manganese cathode using quantile regression method,” *Appl. Energy*, vol. 256, no. August, p. 113841, 2019, doi: 10.1016/j.apenergy.2019.113841.

[43] J. I. Daoud, “Multicollinearity and Regression Analysis,” *J. Phys. Conf. Ser.*, vol. 949, no. 1, 2018, doi: 10.1088/1742-6596/949/1/012009.

[44] L. Plonsky and F. L. Oswald, “MULTIPLE REGRESSION AS A FLEXIBLE ALTERNATIVE to ANOVA in L2 RESEARCH,” *Stud. Second Lang. Acquis.*, vol. 39, no. 3, pp. 579–592, 2017, doi: 10.1017/S0272263116000231.

[45] F. Emmert-Streib and M. Dehmer, “Evaluation of Regression Models: Model Assessment, Model Selection and Generalization Error,” *Mach. Learn. Knowl. Extr.*, vol. 1, no. 1, pp. 521–551, 2019, doi: 10.3390/make1010032.

[46] P. Schober and L. A. Schwarte, “Correlation coefficients: Appropriate use and interpretation,” *Anesth. Analg.*, vol. 126, no. 5, pp. 1763–1768, 2018, doi: 10.1213/ANE.0000000000002864.

[47] I. S. Laktionov, O. V. Vovna, A. A. Zori, and V. A. Lebedev, “Results of simulation and physical modeling of the computerized monitoring and control system for greenhouse microclimate parameters,” *Int. J. Smart Sens. Intell. Syst.*, vol. 11, no. 0, pp. 1–15, 2018, doi: 10.21307/IJSSIS-2018-017.

[48] N. Dangi, “Monitoring environmental parameters: humidity and temperature using Arduino based microcontroller and sensors,” pp. 3–8, 2017, [Online]. Available: https://www.theseus.fi/bitstream/handle/10024/142235/Dangi_Nagendra.pdf?sequence=1&isAllowed=y%0Ahttps://www.theseus.fi/bitstream/handle/10024/142235/Dangi_Nagendra.pdf?sequence=1.

[49] Y. Jin, J. Chen, and J. Zhang, “Final Report for Group # 30 REMOTE CONTROL LED WITH TIME / TEMPERATURE / DATE DISPLAYING LCD SCREEN AND TOUCH SENSOR,” 2019.

[50] ABADI HAIDER MOHAMMED AHMED, “Design and Implementation of Temperature and Humidity Monitoring System for High Performance Servers Room,” no. October, pp. 1–91, 2017.

[51] K. K. Khaing, K. Srujan Raju, G. R. Sinha, and W. Y. Swe, “Automatic temperature control system using arduino,” *Adv. Intell. Syst. Comput.*, vol. 1090, no. March, pp. 219–226, 2020, doi: 10.1007/978-981-15-1480-7_18.

APPENDICES

Appendices 1-Model Prediction code

```
<html>
<head>
<h1 align="center">DC Load Power Prediction</h1>
<link href="bootstrap/dist/css/bootstrap.css" rel="stylesheet">
<link href="bootstrap/dist/css/bootstrap.min.css" rel="stylesheet">
<style>
body {
 background-color: lightblue;
}
</style>
<script>
      //function that display value
      function dis()
      {
        let a = document.getElementById("volt").value
        let b= document.getElementById("current").value
        let c = document.getElementById("temprature").value
        let d= document.getElementById("humidity").value
        let y= -2.741+ (0.0002) * c + (-0.000066) * d + (0.0507) * a + b *(0.0540)
        document.getElementById("result").value = y
      }
    </script>
</head>
<hr>
```

```html
<body>
<div class="row">
<div class="col-sm-4"></div>
<div class="col-sm-3" >
<div class="panel panel-default">
  <div class="panel-body">
<form class="form-horizontal">
<div class="form-group">
<b>Voltage :</b> <input type="text" id="volt" class="form-control" /></div><br>
<b>Current:</b> <input type="text" id="current" class="form-control" /><br>
<b>Temprature :</b> <input type="text" id="temprature" class="form-control" /><br>
<b>Humidity:</b> <input type="text" id="humidity" class="form-control" /><br>
 <input type="button" id="btn" class="btn btn-success"  onclick= "dis()" value="Predict" />
<br>
<b>Result =</b> <input type="text" id="result" class="form-control" />
</form>
</div>
</div>
</div>
<div class="col-sm-5"></div>
</div>
</body>
<script src="bootstrap/dist/js/bootstrap.js"></script>
</html>
```

Appendices 2-Tempratuer and Humidity disply code

```c
#include <LiquidCrystal.h>
```

```
#include <SimpleDHT.h>

//Declaring digital pin no 6 as the dht11 data pin

int pinDHT11 = A0;

SimpleDHT11 dht11;

//Declaring the lcd pins

const int rs = 12, en = 11, d4 = 5, d5 = 4, d6 = 3, d7 = 2;

LiquidCrystal lcd(rs, en, d4, d5, d6, d7);

void setup() {

// Don't forget to choose 9600 at the port screen

  Serial.begin(9600);

//Telling our lcd to start up

  lcd.begin(16, 2);

  pinMode(A2,OUTPUT);

}

void loop() {

  //These serial codes are for getting readings on the port screen aswell as the LCD display, since they'll offer us a more detailed interface

  Serial.println("=================================");

  Serial.println("DHT11 readings...");

  byte temperature = 0;

  byte humidity = 0;

  int err = SimpleDHTErrSuccess;

  //This bit will tell our Arduino what to do if there is some sort of an error at getting readings from our sensor

  if ((err = dht11.read(pinDHT11, &temperature, &humidity, NULL)) != SimpleDHTErrSuccess) {

    Serial.print("No reading , err="); Serial.println(err);delay(1000);

    return;
```

```
}
Serial.print("Readings: ");
Serial.print((int)temperature); Serial.print(" Celcius, ");
Serial.print((int)humidity); Serial.println(" %");
//Telling our lcd to refresh itself every 0.75 seconds
lcd.clear();
//Choosing the first line and row
lcd.setCursor(0,0);
//Typing Temp(oc): to the first line starting from the first row
lcd.print("Temp(oc): ");
//Typing the temperature readings after "Temp: "
lcd.print((int)temperature);
//Choosing the second line and first row
lcd.setCursor(0,1);
//Typing Humidity(%): to the second line starting from the first row
lcd.print("Humidity(%): ");
//Typing the humidity readings after "Humidity(%): "
lcd.print((int)humidity);
if (temperature > 5) {
   digitalWrite(A2,HIGH);
   }
  else {
```

```
      digitalWrite(A2,LOW);
      }
    }
```

Appendices 3-Voltage display code

```
#include <LiquidCrystal.h>
// initialize the library with the numbers of the interface pins
LiquidCrystal lcd(12, 11, 5, 4, 3, 2);
int analogInput = A1;
float vout = 0.0;
float vin = 0.0;
float R1 = 100000.0; // resistance of R1 (100K) -see text!
float R2 = 10000.0; // resistance of R2 (10K) - see text!
int value = 0;
void setup() {
  pinMode(analogInput, INPUT);
   lcd.begin(16, 2);
   lcd.print("BTS Project");
   delay(3000);
 lcd.clear();
    pinMode(A5,OUTPUT);
  pinMode(A4,OUTPUT);
```

```
  pinMode(A3,OUTPUT);
}
void loop(){
   value = analogRead(analogInput);
   vout = (value) / (48.0*4.4); // see text
   vin = vout / (R2/(R1+R2));
   if (vin<0.09) {
   vin=0.0;//statement to quash undesired reading !
}

lcd.print("INPUT V= ");
lcd.print(vin);
delay(1000);
if (vin>50.0)
{
  lcd.setCursor(1,1);
      lcd.write("OVER Voltage");
          digitalWrite(A5, HIGH);
  digitalWrite(A4, LOW);
  digitalWrite(A3, LOW);
}
```

```
else if((vin>46.0) && (vin <=50.0))
{
  lcd.setCursor(1,1);
    lcd.write("NORMAL Voltage");
             digitalWrite(A5, LOW);
  digitalWrite(A4, HIGH);
  digitalWrite(A3, LOW);
}
else {
  lcd.setCursor(1,1);
    lcd.write("UNDER Voltage");
      digitalWrite(A5, LOW);
  digitalWrite(A4, LOW);
  digitalWrite(A3, HIGH);
}
   }
```