Tribhuvan University
Asian School of Management And Technology

Final Year Project Report
on
"Diabetes Prediction System"

Under the supervision of

Chakra Narayan Rawal

Submitted By:

Gaurab Mani Rimal

Aashish Pudasaini

Santosh Tamang

Submitted To:
Asian School of Management And Technology

Tribhuvan University

ACKNOWLEDGEMENT

We would like to whole heartedly express our deepest sense of gratitude and sincere thanks to our highly respected and esteemed guide Mr. Chakra Rawal(program coordinator CSIT, ASMT), for his valuable guidance, encouragement and help for completing this work. His useful suggestions for this whole work and co-operative behavior are sincerely acknowledged. We also thanks to Mr. Surya Bam and Anil Lal Amatya for their valuable guidance throughout the project.

We would also like to express our sincere thanks to all our colleagues and others who helped us directly or indirectly during this project work.

Thank you all once again!

Gaurab Mani Rimal (15929/074)

Aashish Pudasaini (15908/074)

Santosh Tamang (15945/074)

ABSTRACT

Nowadays, diabetes has become a common disease to the mankind from young to the old persons. The growth of the diabetic patients is increasing day-by-day due to various causes such as bacterial or viral infection, toxic or chemical contents mix with the food, auto immune reaction, obesity, bad diet, change in lifestyles, eating habit, environment pollution, etc. A lot of Nepalese people are suffering from diabetes but due to unavailability of diagnosis they are having trouble in day to day activity of the life. Hence, diagnosing the diabetes is very essential to save the human life from diabetes. The data analytics is a process of examining and identifying the hidden patterns from large amount of data to come to certain conclusions and provide a new input in solving the disease. In health care, this analytical process is carried out using machine learning algorithms for analyzing medical data to build the machine learning models to carry out medical diagnoses. This project tries to present a simple diabetes prediction system to diagnosis diabetes. Moreover, this paper explores the approaches to improve the accuracy in diabetes prediction using medical.

Diabetes Prediction is integral part of this system, according to given test data from user it will predict the possibility of having diabetes or not. To predict the diabetes from given test use the support vector machine algorithm which finds the probability of having diabetes.

TABLE OF CONTENT

ABBREVIATIONS

BIM	Body Mass Index
RAM	Random Access Memory
GB	Giga Byte
HTML	Hyper Text Markup Language
ERD	Entity Relational Diagram
CSV	Comma Separated Values
CSS	Cascading Style Sheet
ANN	Artificial Neural Network
DNN	Deep Neural Network
RBF	Radial Basis Function

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 Introduction

Diabetes is the fast-growing disease among the people even among the youngsters. Diabetes is caused by the increasing level of the sugar (glucose) in the blood. The prediction of diabetes system means that recognize diabetes on the particular person without the help of a doctor. This system helps us to predict diabetes if we have the test result of several parameters. In this system, we take a various parameter for a prediction like no of Pregnancy, Glucose level, Blood Pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function, Age. We can use the Naïve Bayes' Theorem to predict diabetes with different libraries. According to these parameters, the system gives result 0 or 1. Here 0 means have not diabetes and 1 means have diabetes. This project works on the principle that no one should be deprived of the health service because they do not have sufficient information about their illness. When the amount of glucose in the blood increases, e.g., after a meal, it triggers the release of the hormone insulin from the pancreas. Insulin stimulates muscle and fat cells to remove glucose from the blood and stimulates the liver to metabolize glucose, causing the blood sugar level to decrease to a normal level [1] Diabetes can be classified into two categories such as type 1 diabetes and type 2 diabetes. Type 1 diabetes is an autoimmune disease. In this case, the body destroys the cells that are essential to produce insulin to absorb the sugar to produce energy. Type 2 diabetes usually affects the adults who are obese. In this type, the body resists observing insulin or fails to produce insulin. Type 2 generally occurs in the middle or aged groups [2]

In the context of Nepal, a lot of people are deprived of the health services because of the limited doctors and hospital in portion to the population, low economic standard making it infeasible for people to visit the doctor every time. So, this project envisions to let of Nepal to receive diabetes prediction service without visiting a doctor without money.

1.2 Problem Statement

In developing countries like Nepal, the reach of health services and its awareness are poor. There is limited no of doctor causes the problem to the patient and they cannot visit a hospital. People cannot

visit health centers because they are limited in number and it is not always feasible to afford expenses for every health illness considering the weak economic condition of the people.

There is good hospital and equipment but not available of a doctor in time. In addition to this, there are also people who find it hard to take time out to visit health centers for all health problems. The project aims to be a common solution to all these problems and limits, this help to diagnose diabetes disease without the help of a doctor.

1.3 Objective

The fast-growing disease among all the people, so it is important to give all facility to the patient. So central aim of this project is to make health-related knowledge and information reach every individual in really and smartest way. The major objectives of the project are listed as follows.

- To reduce the dense queue in health centers by predicting diabetes according to a given test result.
- To measure the probability of a user for getting diabetes.
- To reduce the lack of a doctor in the rural area because one simple person can do test various feature and by using this system diagnose diabetes.
- To develop the system that function with the real problem.
- To implement machine learning algorithm as prediction technique into a system.

1.4 Scope and Limitation

With the aim of reaching every individuals who are in need test of diabetes this system predict the diabetes if given data set is correct. This system is useful where doctor can't reach. So this system help to those patient who do not have pay high cost to doctor and where do not doctor. Simply this system give result of given test data sets.

- This system is applied to all the health post or hospitals reduce the load of doctor
- This system is applied to all the individual who has the probability of diabetes and has the test result of different features.

- This system gives detailed information about diabetes, so this helps the patient to check different test result if symptoms do not match.

The limitation of the project are as follows.

- Not all individuals in need of this services are aware of the use of computer system or mobile and web application.
- The project requires computer machine with internet to operate on
- This system required first test results before check like Glucose, insulin, BIM, etc.

1.5 Report Organization

This chapter discusses about introduction of the project, problem statement, objectives, scope and limitations of the project. In chapter 2 the system and its literature review is investigated in detail. In chapter 3 System Analysis is investigated in details. This includes requirement collection, and feasibility study. In chapter 4 system design is studied in details. This includes data design and system design. The listing of algorithm used and its implementation are also includes in the very chapter. In chapter 5 implementation and testing is done and studied which includes tools used, implementation details of modules and testing is done. In chapter 6 conclusion and future recommendation is discussed.

CHAPTER 2: Background Study and Literature review

2.1 Background study

Diabetes is a chronic health condition that affects how your body converts food into energy. Most of the food you eat is broken down into sugar called glucose and released into your bloodstream. When your blood sugar goes up, it signals your pancreas to release insulin. Insulin acts like a key to let the blood sugar into your body's cell to be used as energy. For a normal person the sugar level cannot be low or more from the normal level which is 4.4 until 6.1 mmol/L [1].

Over the long period of time diabetes can cause serious health problems like heart disease, kidney disease, vision loss etcetera. There is no immediate cure for diabetes but it can be prevented by following balanced diet and living an active and healthy lifestyle.

Diabetes are known the one of the top disease in this world. Diabetes are not easily to be cured and need to depend on the medicine. If we know earlier that we had diabetes, we can control its impacts

become worse. The people only know that they have diabetes after the effects already become worse. So, the early prediction are required to aware all the people. There are three type of Diabetes that has been identified such as Type 1, Type 2 and Gestational Diabetes. All this type diabetes have its differences and characteristics. Also, the majority patient who has Diabetes is female than male [2].

Type 1, usually people who suffer this type diabetes, she/he cannot produce insulin anymore in their body because the pancreas totally damaged. Furthermore, the average people have this type of diabetes is below 20 years old [2]. Next, the patient with this type have weight loss. However, this disease is not easy to classify whether the patient can have this type diabetes or can become into Diabetes Type 2 [3].

As well as Type 1, Type2 Diabetes patient also have problem in producing insulin for their body, where their pancreas produce insulin ,however it still no enough because their body resistant toward insulin. The majority of the Diabetes patients had this type diabetes [3].

Gestational Diabetes, commonly the person who have this type diabetes are consist of pregnant women. During pregnancy moment, the pregnant women are advised to do a few test to check they have this kind of diabetes or not. If the person have this diabetes, the production of insulin cannot be produce as usual as before pregnant. The risk for the baby to suffer from diabetes also higher. For information, usually the high weight baby maybe delivered by the Gestational Diabetes mother. Next, for the next pregnancy, the patient have high risk to get the same problem. The bad effect to pregnant women who have this diabetes is bleeding during birth or miscarriage may occur [2].

2.2 Literature Review

2.2.1 Diabetes prediction system using Decision Tree

Diabetes prediction system is very useful in the healthcare field. An accurate system for diabetes prediction is proposed in this paper. The proposed system used K-nearest neighbor algorithm for eliminating the undesired data, thus reducing the processing time. However, a proposed classification approach based on Decision Tree(DT) to assign each data sample to its appropriate classes. By

experiment, the proposed system achieved high classification result i.e 98.7% comparing to the existing system using Pima Indian Diabetes Dataset [4].

2.2.2 Early diabetes prediction system using ANN

Diabetes is a common, chronic disease. Prediction of diabetes at an early stage can lead to improved treatment. Data mining algorithm are widely used for prediction of disease at an early stage. In this research paper, diabetes is predicted using significant attributes and the relationship of differing attributes are also characterized. Various tools are used to determine significant attribute selection, and for clustering, prediction, and association rule mining for diabetes. The ANN technique provided a best accuracy of 75.7% which may be useful to assist medical professional with treatment decision [5].

2.2.3 Diabetes prediction system using DNN

The deaths by diabetes are increasing each year so the need of developing a system that can effectively diagnose the diabetes patient becomes inevitable. In this work, an efficient medical decision system for diabetes prediction based on Deep Neural Network(DNN) is presented. Such algorithms are state-of-the-art in computer vision, language processing and image analysis and when applied in healthcare for prediction and diagnosis purpose these algorithm can produce highly accurate results. The obtained results showed that the proposed system based on DNN technique provides promising performances with an accuracy of 99.75% and FI score of 99.6%. This improvement can reduce time, efforts and labor in healthcare services as well as increasing the final decision accuracy [6].

CHAPTER 3: System Analysis

3.1 System Analysis

3.1.1 Requirement Analysis

The requirement analysis of a diabetes prediction system to be developed is laying out, functional and non-functional requirement and may include a prediction set of use cases that describe about how diabetes prediction works within the system. The requirement analysis enlists all the necessary and enough requirements for the project development. To derive requirements, clear and thorough understanding of the system is required.

3.1.1.1 Use Case Diagram

The use case diagram of our system shows user and admin as an actor accessing the system. The user can view information and check diabetes as admin can login and update information.

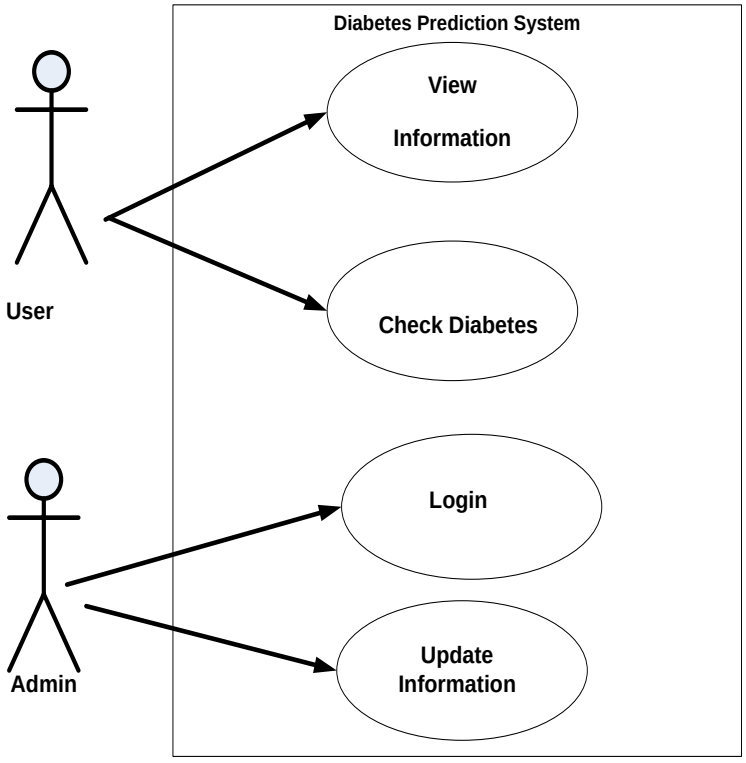

Figure 1: Use Case Diagram

i. Functional Requirement

Functional requirements specifies a function that a system or system component must be able to perform. The functional requirements of the system are to the implement the solution for inserting, updating, deleting and retrieving the information related to the different entities. This project has following functional requirement:

- **User Input:**

  The user gives the input to the system according to their test result of lab. After entering the tested values as parameters the system predict the disease like have diabetes or you haven't diabetes.

- **Processing:**

  The information regarding symptoms of disease, details of disease, treatment, self-care, diagnosis and other details are retrieved from the database.

- **Storage Requirements:**

  The information will be stored, accessed and retrieved from the database.

ii. Non-Functional Requirement

Non-functional requirements are any other requirement than functional requirements. Those are the requirements that specify criteria that can be used to judge the operation of a system, rather than specific behaviors**.** The new system provides different nonfunctional requirements, which can be categorized as follows.

- **Usability:**

The new system is very user friendly. Any person with basic computer knowledge

Can use it.

- **Efficiency:**

The new system is very much efficient. The information can be stored easily and retrieve as per required.

- **Portability:**

The system is portable across different platforms and internet services.

- **Security:**

The new system is very secure as it doesn‘t allow the unauthorized user to violate and modify the terms and conditions of the system.

- **Portability:**

The system is user friendly that meets the computer knowledge and skill of the general user. This is web based application that preforms independent so it can run easily in every device like mobile, computer, etc.

3.1.2 Feasibility Study

Feasibility study is the study of system about its ability to meet the user requirement. Feasibility is the study of impact, which happens in the organization by the development of a system. It is wise to think about the feasibility of any problem this project looks to solve or undertake. The impact can be either positive or negative. When the positives nominate the negatives, then the system is considered feasible. Here the feasibility study determine if the proposed project:

- Is it technically feasible?
- Is it feasible with the estimated cost?
- Is it profitable?

i. Technical Feasibility Study

Technical study considers the technical requirement of the proposed project. Technical requirement that are required during the project are system hardware and software. The technical resources needed for the system should be available and its must be reasonable to use. The technical requirement for this proposed system is considered as windows, Django, Bootstrap, Python, SQLite and Atom or sublime and machine learning concept. To develop this project we use the following hardware and software:-

Hardware Requirements:

The minimum hardware requirement for developed system and its operations include the following:

- RAM:-4GB
- Processor:-Intel Pentium 4
- Hard disk:- 500GB

Software Requirements:

The minimum software requirement for developed system and its operations include the following:

- Operating system: Microsoft Windows 8, 7, 8.1, 10 etc.
- Software: Atom, Django framework, Jupyter Notebook or Spyder, Command line etcetera
- Database: SQLite
- Language: HTML, Bootstrap, Python, Machine Learning etc.

These above hardware and software are easily available and can be used by anyone. Hardware used are very minimal and hence any computer can run our system. Also software are also easily available and easy to use. So analyzing these hardware and software used it is sure that this system is technically feasible.

ii. Economic Feasibility (Measure of the cost effectiveness of the system**)**

Economic feasibility determine if system is worth the cost and time we spent on it. Out system is economically feasible because it is web based and easy to use with minimum cost. The only cost related in this system will be for hosting space. It doesn't require further hardware and other resource. Science the project is developing using Window 10, Atom text editor, Jupyter that doesn't require the cost for installation and use. Hence the project is economically feasible.

iii. Operational Feasibility (measure of wellness and adaptability of the system)

Operational feasibility is dependent on human resources available for the project and involves projecting whether the system will be used if it is developed and implemented. Operational feasibility

is a measure of how will propose systems solves the problems. It reviews the willingness of the organization to support the proposed system.

Generally there is two aspects of operational feasibility I considered:

Is the problem worth solving, or will the solution to the problem work?

Yes this system solve the problem of user by taking user input and giving output.

How do the end user and management team feel about the solution?

As the system creates a graphically interactive system the end user can find it very easy for use and will help them to solve their problem. Hence the system is feasible operationally also.

iv. **Schedule Feasibility (**measure the reasonability of the project time table**)**

A project will fail if it takes too long to be completed before it is useful. Typically, this means estimating how long the system will take to develop, and if it can be completed in a given time period using some methods like payback period. Schedule Feasibility takes into consideration:

How reasonable the project timetable is?

Given our technical expertise are the project deadline achieve able?

3.1.3 Structured Analysis

3.1.3.1 Process Model

For this project we use spiral model as the model of the methodology, which has been widely applied in the other project. There are many advantages of using spiral model as any idea can be introduced in the later stages of the project, the budget of the system can also be known and the end user can always give idea or contribution towards the project.

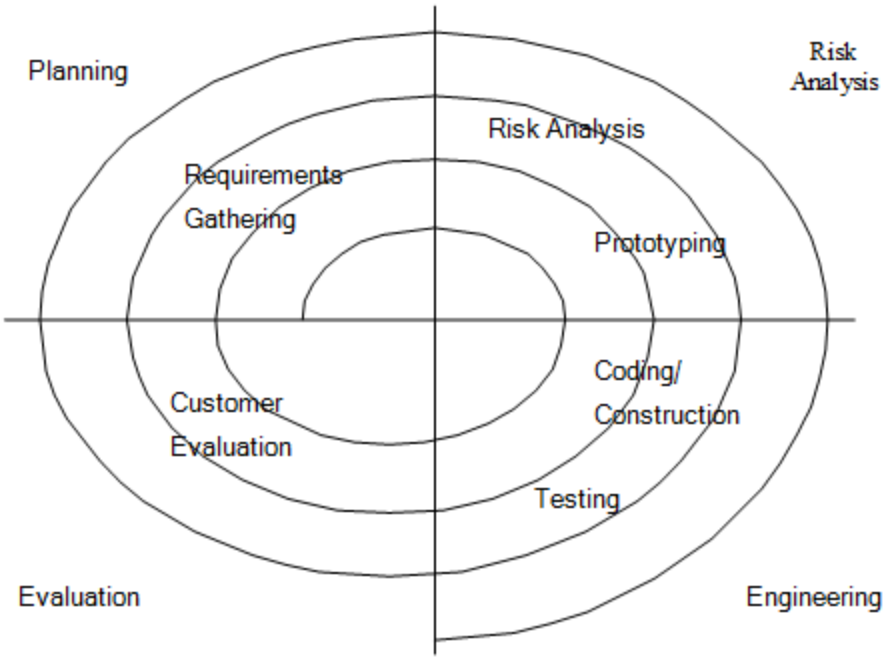

Figure 2: Spiral Model

There are four phases that involved in the spiral model that including: planning, risk analysis, engineering and evaluation. For each phase, there are activities are involved.

- **Planning phase:**

  Phase where the requirement is collected and risk is assessed. This phase where the title of the project has been discussed with project 13 supervisor. From that discussion, Diabetes Prediction System has been proposed

- **Risk analysis Phase:**

  Phase where the risk and alternative solution are identified. A prototype are created at the end this phase. If there is any risk during this phase, there will be suggestion about alternate solution.

- **Engineering phase**

  At this phase, a software are created and testing are done at the end this phase.

At this phase, the user do evaluation toward the software. It will done after the system are presented and the user do test whether the system meet with their expectation and requirement or not. If there is any error, user can tell the problem about system.

3.1.3.2 Context Diagram

Context Diagrams all represent all external entities that may interact with a system. Such a diagram pictures of the system at the center, with no details of its interior structure, surrounded by all its interacting systems, environments and activities. The objective of the system context diagram is to focus attention on external factors and events that should be considered in developing a complete set of systems requirements and constraints.

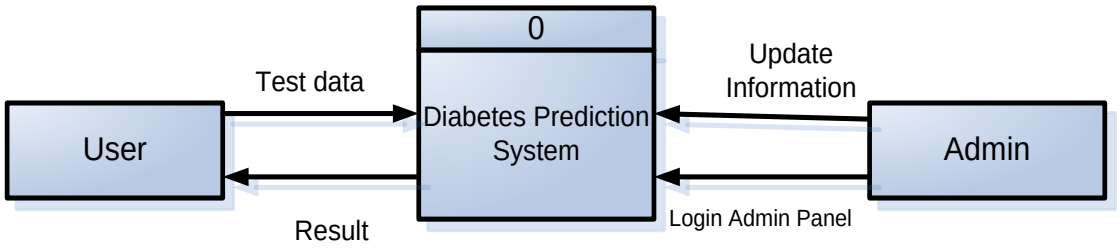

Figure 3: Context Diagram (DFD 0)

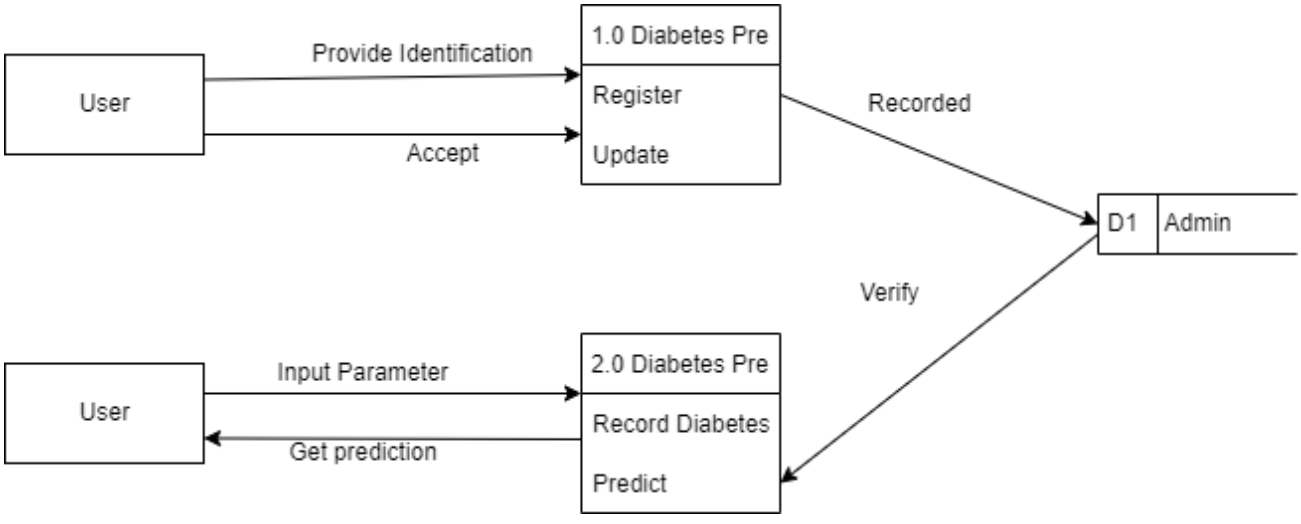

Figure 4: Context Diagram (DFD 1)

3.1.3.3 ER Diagram

An entity-relationship diagram (ERD) is a data modeling technique that graphically illustrates an information system's entities and the relationships between those entities. An ERD is a conceptual and representational model of data used to represent the entity framework infrastructure. The E-R diagram of whole system of this project is given below.

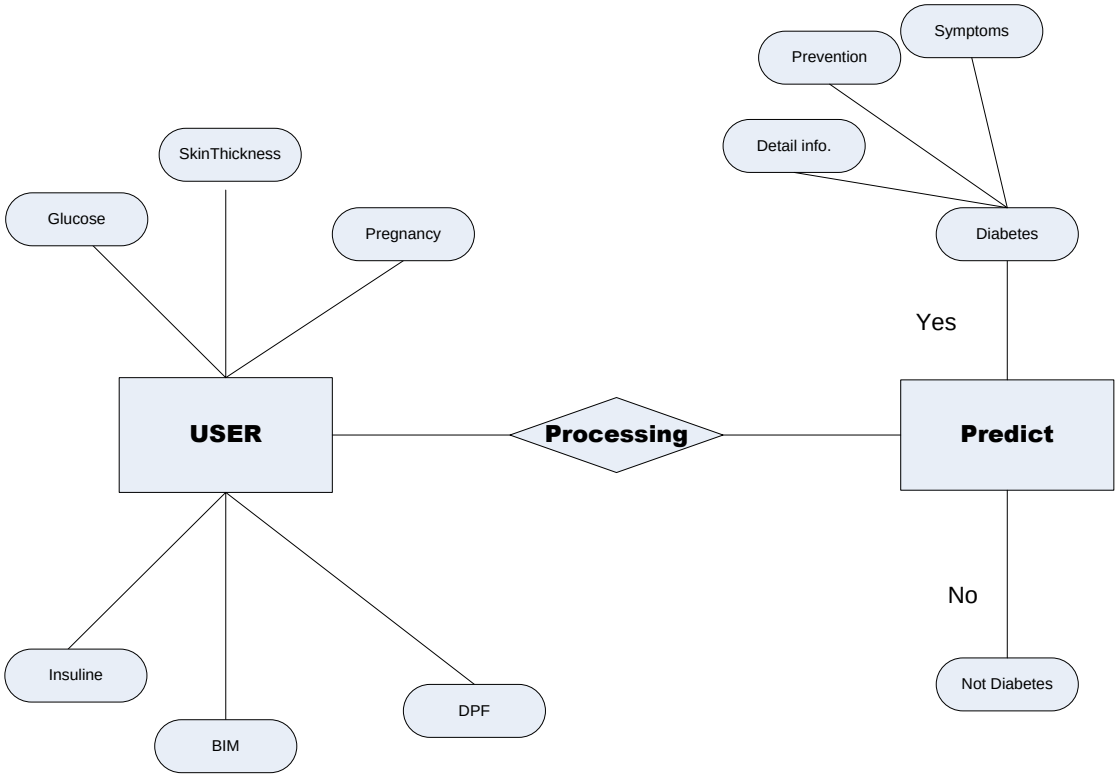

Figure 5: E-R diagram

CHAPTER 4: System Design

4.1 System Architecture

The system developed is a diabetes prediction system. It asks the user different values like no of pregnancies, glucose level in the body, blood pressure of the patient, skin thickness, Insulin, body mass Index, diabetes pedigree function and age. After getting these value the system can predict whether the user is diabetic or not.

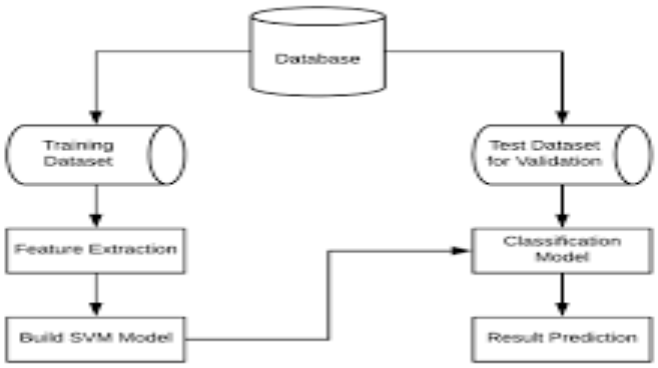

Figure 6: System Architecture

4.1.1 Dataset Design

Data set play important role in make sure the data and information in the system display properly. Data set contain the information of the past inserted data. These data sets are used for testing and training and finally these data used in prediction.

The dataset used in the system are saved as .CSV file. The data used in the system is originally from National Institute of Diabetes and Digestive and Kidney Disease.

Index	pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	etesPedigreeFunc	Age	Outcome
445	0	180	78	63	14	59.4	2.42	25	1
228	4	197	70	39	744	36.7	2.329	31	0
4	0	137	40	35	168	43.1	2.288	33	1
370	3	173	82	48	465	38.4	2.137	25	1
45	0	180	66	39	0	42	1.893	25	1
58	0	146	82	0	0	40.5	1.781	44	0
371	0	118	64	23	89	0	1.731	21	0
593	2	82	52	22	115	28.5	1.699	25	0
621	2	92	76	20	0	24.2	1.698	28	0
395	2	127	58	24	275	27.7	1.6	25	0
330	8	118	72	19	0	23.1	1.476	46	0
622	6	183	94	0	0	40.8	1.461	45	0
12	10	139	80	0	0	27.1	1.441	57	0
147	2	106	64	35	119	30.5	1.4	34	0
661	1	199	76	43	0	42.9	1.394	22	1
308	0	128	68	19	180	30.5	1.391	25	1

Table 1: Table for Diabetes Prediction System

4.1.2 Interface and Dialogue Design

The app interface is simple and easy to use. It has login page and a page where user can input the different parameters the system asks him to provide which will give the right accurate prediction.

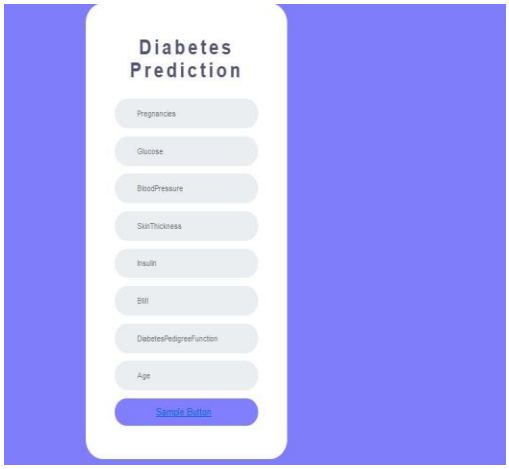

figure 7: Interface of Diabetes Prediction System

4.2 Algorithm

Support Vector Machine(SVM) is a supervised machine learning algorithm used for both classification and regression. Though we say regression problems as well its best suited for classification. The objective of SVM algorithm is to find a hyperplane in an N-dimensional space that distinctly classifies the data points. The dimension of the hyperplane depends upon the number of features. If the number of input features is two, then the hyperplane is just a line. If the number of input features is three, then the hyperplane becomes a 2-D plane. It becomes difficult to imagine when the number of features exceeds three.

A support vector machine takes these data points and outputs the hyperplane (which in two dimensions it's simply a line) that best separates the tags. This line is the decisionboundary: anything that falls to one side of it we will classify as *blue*, and anything that falls to the other as *red*.

The hyperplane (remember it's a line in this case) whose distance to the nearest element of each tag is the largest.

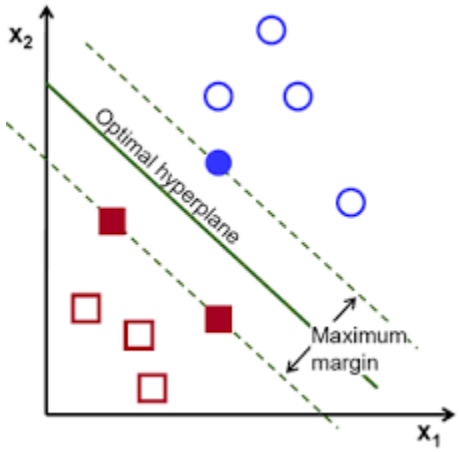

Figure 8: Support Vector Machine implementation for linear data

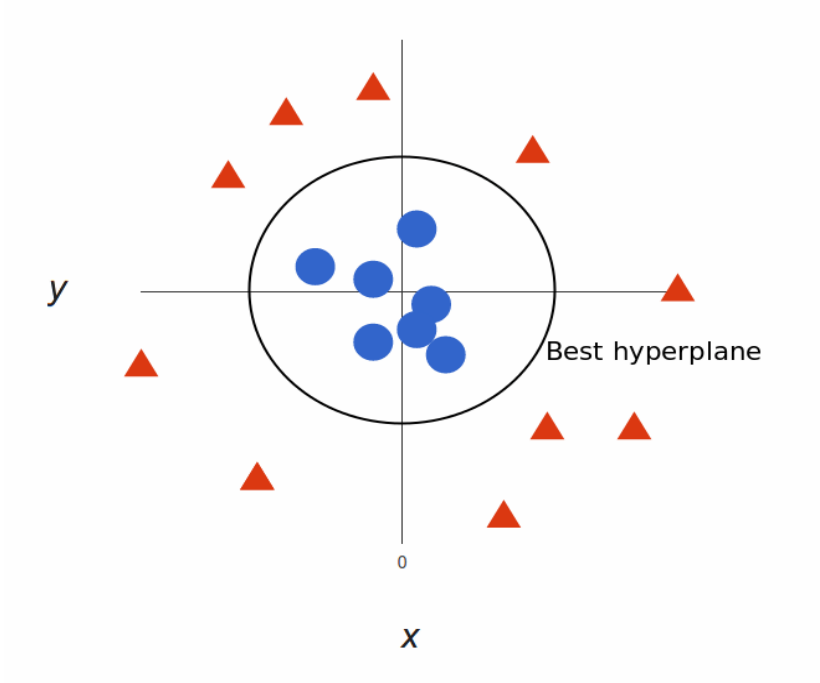

Figure 9: Support Vector Machine implementation for non-linear data

SVM is a supervised learning algorithm that can be used for both classifications as well as regression problems. However, mostly it is used for classification problems. It is a highly efficient and preferred algorithm due to significant accuracy with less computation power.

In this article, we are going to see the working of SVM and the different kernel functions used by the algorithm.

In the SVM algorithm, we plot each observation as a point in an n-dimensional space (where n is the number of features in the dataset). Our task is to find an optimal hyper plane that successfully classifies the data points into their respective classes.

Before diving into the working of SVM let's first understand the two basic terms used in the algorithm "The support vector" and" Hyper-Plane".

A hyper plane is a decision boundary that differentiates the two classes in SVM. A data point falling on either side of the hyper plane can be attributed to different classes. The dimension of the hyperplane depends on the number of input features in the dataset. If we have 2 input features the hyper-plane will be a line. Likewise, if the number of features is 3, it will become a two-dimensional plane.

Support vectors are the data points that are nearest to the hyper-plane and affect the position and orientation of the hyper-plane. We have to select a hyperplane, for which the margin, i.e the distance between support vectors and hyper-plane is maximum. Even a little interference in the position of these support vectors can change the hyper-plane.

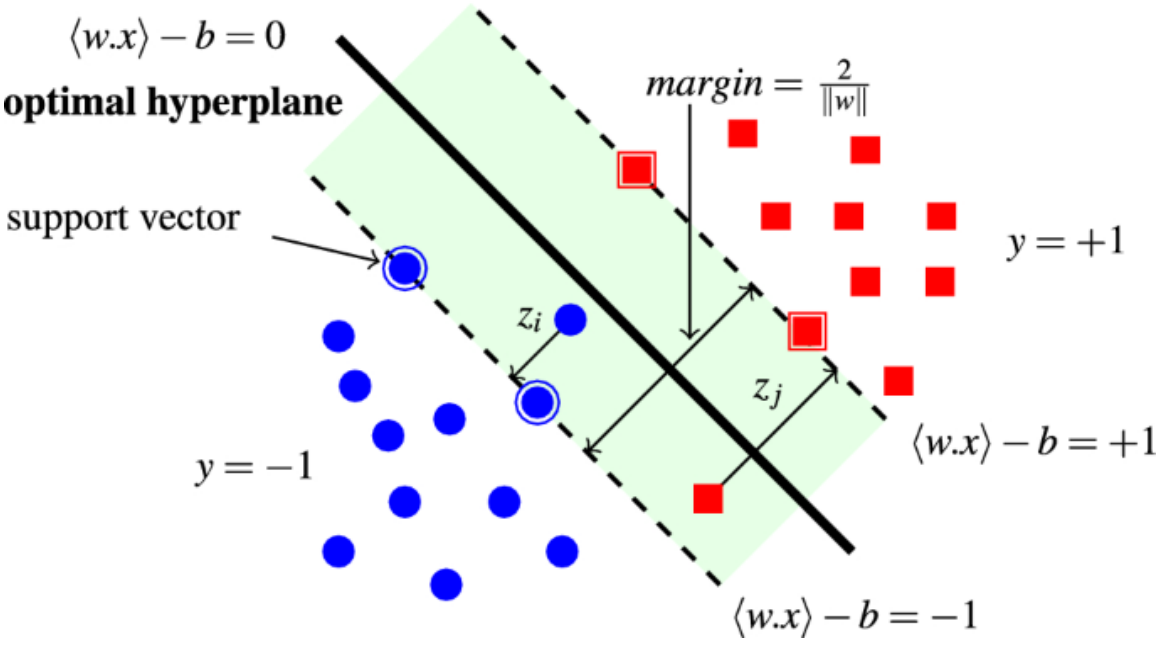

figure 10: representation of svm hyper line separating two support vectors

Kernel

Kernel function is a method that is used to take data as an input and transform it into the required form of processing data. "Kernel" is used due to set of mathematical functions used in support vector machine providing the window to manipulate data. The kernel function generally transforms the training set of data so that non-linear surface is able to transform the linear equation in a higher number of dimension spaces. It returns the inner product between two points in a standard feature dimension.

Kernel function	Mathematical formula
Linear kernel	$k(X_i, X) = X_i \cdot X$
Polynomial kernel	$k(X_i, X) = (X_i \cdot X)^e$
Normalized Polynomial kernel	$k(X_i, X) = [(X_i \cdot X)^e] \Big/ \sqrt{(X_i \cdot X_i)^e (X \cdot X)^e}$
Radial basis kernel (RBF)	$k(X_i, X) = e^{-\left(\lvert X_i - X \rvert^2 / 2\sigma^2\right)}$

CHAPTER 5: Implementation & Testing

5.1 Implementation

For the implementation of system different frontend, backend technologies and databases were used. The user is asked for input in through a form and behind the scenes the dataset from database is loaded, it is then preprocessed, then it is splinted into testing and training data and then the machine learning algorithm predicts whether the patient is diabetic or not.

5.1.1 Tools Used

5.1.1.1 Analysis and Design Tools

Several tools has been used to create the diagrams. Visual Paradigm was used to create dfd diagram, use case diagram etcetera.

5.1.1.2 Implementation Tools

Frontend

- HTML
- CSS
- Bootstrap

Backend

- SQLITE3
- Python
- Django

5.1.2 Implementation details of Modules

The final prediction is not generated in a single step. The system has to undergo many steps to give the desired output for the system.

5.1.2.1 Preparing Training Data

The PIMA Indian diabetes dataset donated by Vincent Sigilito, is a collection of medical diagnostic report of women from age 21 onwards living in Arizona, USA. In the dataset 0 means negative of diabetes and 1 means positive of diabetes. Out of 768 instances there is 500 cases of no diabetes and

268 cases of diabetes. The dataset is stored as CSV file so it can be used by using import function. There are no missing values in the data, but it needs to be cleaned for duplicate and missing data.

5.1.2.2 Training Data

In the training phase, the SVM algorithm first draws an N-dimensional hypercube by representing each feature as a separate dimension. It then uses the numerical values of those features to plot points on the N-dimensional hypercube. It then attempts to find a boundary that separates the two classes of data — points where outcome is 0 (no diabetes) and points where outcome is 1 (diabetes), for example. The boundary is a (N-1) dimension hyperplane.

5.1.2.3 Train/Test Split

To feed the machine we need to split the data into training and testing data. For training and Testing data we split it into 80-20 ratio respectively.

```
#Train Test split
X_train, X_test, Y_train, Y_test = train_test_split(X,Y,test_size=0.2, stratify=Y, random_state=2)
```

5.1.2.4 Testing Data

In the testing phase, we can start with real-time data about a patient such as age, number of pregnancies, insulin levels, and so on. The SVM algorithm determines a 1/0 outcome about diabetes based on which side of that boundary the data falls on.

5.1.2.5 Applying Support Vector Machine

The next step is to build our model using Support Vector Machines. The output of the SQL query above is available as a data frame (df). Skikit-learn package has an algorithm for SVM and we import it. The code for building our model is below. We select the features we want to include and pass that along with the outcomes to the fit method of SVC (Support Vector Classifier). This builds the model. Note that we are using the linear kernel function.

```
# SQL output is imported as a data frame variable called 'df'
import pandas as pd
```

```
from sklearn import svm
outcomes = df['OUTCOME']
features = df[['PREGNANCIES', 'GLUCOSE', 'BLOODPRESSURE', 'INSULIN', 'BMI', 'AGE']].as_matrix()
model = svm.SVC(kernel='linear')
model.fit(features, outcomes)
```

5.2 Testing

Testing is an integral part of the software development process. It is performed at each stage of the SDLC. It ensures that the developed parts conform to the user requirements. It helps to find out whether an input given to the system is well processed or not and output meets the specified objective of the system. It mainly ensures that the system performs as planned. The testing of system was carried out step by step. Testing is performed at key points that are crucial for the working of a system. In hardware system testing is done for individual components that are used to make up a system. In other word testing is a process of Validation and Verification. Validation is the process of checking if the system will meet the customer's actual needs, whereas Verification is concerned whether the system is well-engineered and error free.

Testing of system is carried out step by step. Test plan is to select a set of action that are carried out in in order to revels the defects of the system. There is various different phases of the system testing, these phases are defined in following diagram.

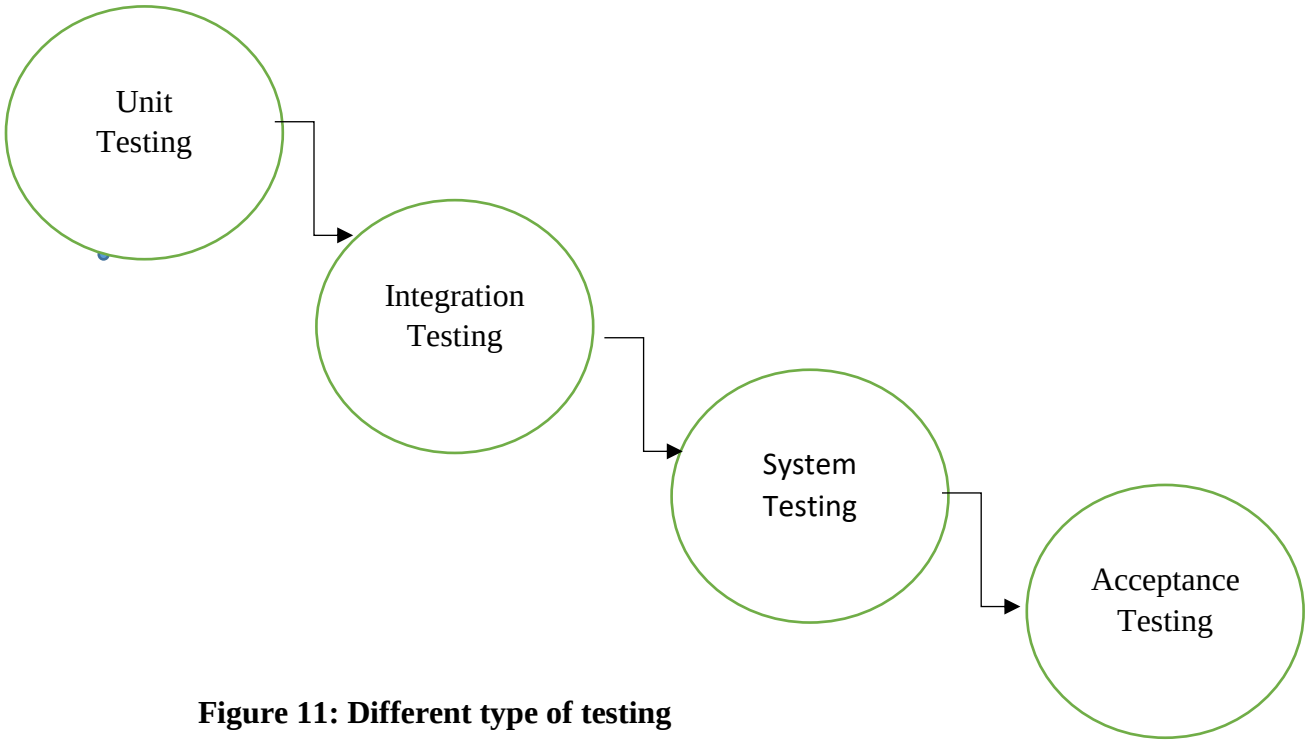

Figure 11: Different type of testing

5.2.1 Test Case for Unit Testing

Client and server side scripts has number of methods, each method is distinctly tested. In this phase each unit developed was tested just after the implementation. The errors were corrected and modification were made required. Each development unit is tested on its own. Each module is tested i.e. program, procedure, result, etc.

In unit testing we tested following:

The correctness of testing and training data set results.

The correctness of the result for given set of input by entering the data set which is already in the data set.

Correctness of the process and function that are designed in the UI.

5.2.2 Test Case for System Testing

The system testing is the combination of all the unit and integration. For the system to run smoothly on the unit and integration of different module comprises a system. Once all the required module were developed and integrated, the system as whole was to ensure that the system is functioning correctly

and effectively. This set is very helpful to determine the problems and bugs that have been escaped during the integration testing. This testing is ensure that the system meets the user requirements.

Index	Pregnanc	Glucose	Bp	Skin Thic	Insulin	Bmi	Diabetes	Age	E.output	O.output
0	6	148	66	35	0	33.6	0.627	50	1	1
1	1	85	72	29	0	26.5	0.354	31	0	0

Table 2: System Testing

5.3 Result Analysis

To analyze the performance of the classification, the accuracy and AUC measures are adopted. Four cases are considered as the result of the classifier.

TP(True Positive): The number of examples correctly classified to the class.

TN(True Negative): The number of examples correctly rejected from the class.

FP(False Negative): The number of example incorrectly rejected from the class.

FP(False Positive): The number of example incorrectly classified to that class.

The Accuracy, sensitivity and specificity of the system is given by:

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

Sensitivity=$\frac{TP}{TP+FN}$ Specificity= $\frac{FP}{FP+TN}$

Dataset	Accuracy	Sensitivity	Specificity
Diabetes	78%	80%	76.5%

Table 3: Performance of SVM Classifier

CHAPTER 6: Conclusion & Future

6.1 Conclusion

The developed system has successfully incorporated all the diabetes related problems that are of public concern. The system can be a huge facility to developing countries like Nepal and developed countries that don't have more time to go hospitals. And this project also probably covers the all project objectives.

The development of this project has helped us know the essence of the computer in solving daily problems of people who otherwise have to go through a difficult procedure. It also helped us understand the different phases in the life cycle of software development and working as a team to meet the development plan.

We look forward to take this system live to the public and help people sort out their health problems irrespective of its nature or find the best health center that can help them cure their deformity. We are also convinced that the system can extend the health knowledge and awareness among the people.

6.2 Future Recommendation

In this system we only make the prediction of the diabetes due to various reasons like time, data etc. But there is huge no of other factors we include in this project in future like recommend medicine according to types of diabetes, similarly predict diabetes in particular type of diabetes, similarly recommend the hospital according to user rating etc. so some of the feature that can enhance feature are below :

- Add feature to predict particular type of diabetes in particular user.
- Recommend the medicine according to diabetes type.
- Recommend best rated hospitals and near hospitals.
- Similar to this system we add the prediction for all disease called as disease prediction system and solution of this system.

REFERENCES

[1] "NBK1671", [Online]. Available :https ://www.ncbi.nlm.nih.gov/books/NBK1671/

[2] "Publication", [Online].Available:https"//www.researchgate.net/publication/316432650_Diabetes_Prediction_Using_Medicaldata/.

[3] "deepblue.lib.umich.edu,", [Online].Available:https://deepblue.lib.umich.edu/bitstream/handle/2021.42/75845/hagan_3.pdf?sequence=1/.

[4] Mustafa S.Kadham, Duaa Enteesha Mhawi"An Accurate diabetes prediction system based on k-means clustering" ,[Online]. Available:https://www.ripublication.com/ijaer18/ijaerv13n6_118.pdf

[5] Talha Mahboob "Amodelforearlypredictionofdiabetes", [Online].Available:https://doi.org/10.1016.j.imu.2019

[6] Yousef Berik, Tawfik Beghride "Efficient prediction system for diabetes disease based on deep neural network", [Online], Available:https://hindawi.com/journals/complexity

LOG

Date Of Visit	Work Discussed
2078/8/11	Discussed about proposal writing.
2078/9/20	Discussed about workflow and task management.
2078/10/10	Discussed about formal report writing.
2078/10/15	Discussed about algorithm implementation.