

Diabetes Health Care Routine Using Machine Learning

BY

Israt Jahan Eva
ID: 181-15-10723

Mariya Akter Mitu
ID: 181-15-11122

Sayra Jannat Anika
ID: 181-15-11252

This Report Presented in Partial Fulfillment of the Requirements for
The Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Abdus Sattar

Assistant Professor

Department of CSE

Daffodil International University



DAFFODIL INTERNATIONAL UNIVERSITY

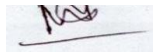
DHAKA, BANGLADESH

JANUARY 2022

APPROVAL

This Project titled “**Diabetes Health Care Routine Using Machine Learning**”, submitted by **Mariya Akter Mitu**, ID: 181-15-11122, **Sayra Jannat Anika**, ID: 181-15-11252, **Israt Jahan Eva**, ID: 181-15-10723 to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering (BSc) and approved as to its style and contents. The presentation has been held on 4, January 2022.

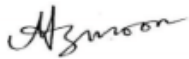
BOARD OF EXAMINERS



Dr. Md. Ismail Jabiullah
Professor

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Chairman



Nazmun Nessa Moon (NNM)
Assistant Professor

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



Aniruddha Rakshit (AR)
Senior Lecturer

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



Dr. Md Arshad Ali
Associate Professor

Department of Computer Science and Engineering
Hajee Mohammad Danesh Science and Technology University

External Examiner

DECLARATION

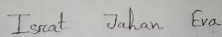
We hereby declare that, this thesis has been done by us under the supervision of **Abdus Sattar**, **Assistant Professor, Department of CSE** Daffodil International University. We also declare that neither this thesis nor any part of this thesis has been submitted elsewhere for award of any degree or diploma.

Supervised by:

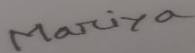


Abdus Sattar
Assistant Professor
Department of CSE
Daffodil International University

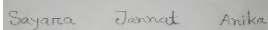
Submitted by:



Israt Jahan Eva
ID: 181-15-10723
Department of CSE
Daffodil International University



Mariya Akter Mitu
ID: 181-15-11122
Department of CSE
Daffodil International University



Sayra Jannat Anika
ID: 181-15-11252
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final thesis successfully.

We really grateful and wish our profound our indebtedness to **Abdus Sattar**, Assistant Professor, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of “Human Computer Interaction in Education” to carry out this thesis. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior draft and correcting them at all stage have made it possible to complete this thesis.

We would like to express our heartiest gratitude to **Professor Dr. Touhid Bhuiyan, Professor, and Head**, Department of CSE, for his kind help to finish our thesis and also to other faculty members and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and passion of our parents.

ABSTRACT

Day by day machine learning is getting its prominence. Recommending something using machine learning is a popular area for enthusiastic researchers. The recommendation system is getting more precise as days go by. In the medical industry, a recommendation system is being used vastly to recommend a good diet. A recommendation system would be a useful tool if we can make it for diabetes patients. As diabetes patients need to follow diet plans and work out strictly. Diabetes patients need to control blood sugar in his or her bloodstream to stay healthy. Using machine learning for recommend food and workout plan is an effort to make a precise recommendation system for Bangladeshi diabetes patients. Various classification algorithm is being used in this system to detect diabetes risk of the user as per his or her input. Such as, Decision tree, Naive Bayes and Support vector machine. After classify the diabetes risk the system will suggest some best diet plan and workout routine by query selection method.

TABLE OF CONTENTS

CONTENTS	PAGE
Board of examiners	i
Declaration	ii
Acknowledgements	iii
Abstract	iv
CHAPTER	
CHAPTER 1: INTRODUCTION	1-3
1.1 Introduction	01
1.2 Motivation	02
1.3 Problem Definition	02
1.4 Research Question	02
1.5 Research Methodology	03
1.6 Research Objective	03
1.7 Report Layout	03
CHAPTER 2: LITERATURE REVIEW	5-6
2.1 Introduction	05
2.2 Related work	05
2.3 Bangladesh Perspective	06
CHAPTER 3: RESEARCH METHODOLOGY	7-12
3.1 Introduction	07
3.2 Experiment Data Set	08
3.3 Process The Original Data	08
3.4 Level Selection	08
3.5 Data Scaling	09
3.6 Model Building	09
CHAPTER 4: RESULT AND ANALYSIS	12
4.1 Training, Testing and the Validation of the model	12
4.2 Model efficiency	12
4.3 Model Comparison	13
CHAPTER 5: CONCLUSION AND FUTURE WORK	14
REFERENCES	15-16

LIST OF FIGURES

FIGURES	PAGE NO.
Figure 3.1.1 Process Structure	07
Figure 3.6.1 Internal structure of SVM	10
Figure 4.3.1 confusion matrix of all algorithms	13

LIST OF TABLES

TABLE NO.	PAGE NO.
Table 3.4.1 Features of the model	08
Table 4.2.1 Accuracy table of the model	12

CHAPTER 1

INTRODUCTION

1.1 Introduction

It is essential for a human being to have food on a regular basis. This food later turns into energy in our inner system. Diabetes is a disease or condition that influences the system that turns food into energy [1]. Whenever we have food maximum of them change into glucose or sugar. After turning it down into sugar it spread into our bloodstream. When the blood sugar rises up it makes a signal to our pancreas to release the insulin. We can consider insulin as a key that decides when blood sugar goes to our body cells for use as energy. If someone is a diabetic patient, there must have two possible causes. First of all, either his or her body cannot make enough insulin that requires or the body cannot make the insulin in a proper way. When a patient suffers from diabetics that means there are too many sugars that stay in his or her bloodstream. Moreover, it is responsible for severe diseases. such as kidney disease, vision loss, and heart problems. Till now there is no proper treatment around the world that can cure diabetes forever. After doing lots of hard work researchers from all over the world are able to invent some medicine that can control blood sugar levels [2]. There are some other ways by which we can control our blood sugar levels. One is diet control another one is regular exercise. It is proved that if a patient can control his or her food and exercise on a regular basis, he or she can control his blood sugar level [3].

In the context of Bangladesh, lots of people are suffering from diabetes. The scenario is more complex because most of them don't know how to control this disease. Most people cannot know the proper diet that they should follow. A recommendation system is being developed by researchers to recommend a thing or product to its user. We can see the use of recommendation systems starting from digital media to online shopping everywhere. Although it's late, the medical industry has also started using recommendation systems to give patients the best treatment.

In this research, we have tried to build a recommendation system that will be able to detect diabetic risk based on user input and recommend a proper diet and exercise plan. As we discussed before most people are unaware of the proper food diet and exercise to control diabetes it proposed model will help to stay safe and healthy.

1.2 Motivation

As per WHO's report, Bangladesh is highly affected by diabetes. In the year 2016, 8% of the total population was affected by diabetes [4]. The number is almost 12.86 million. The problem is so concerning that 3 percent of total deaths in Bangladesh were directly linked with diabetes. As there is no redeem treatment of diabetes and it makes human life so harder, researchers from all over the world are researching and trying to make diabetes medicine. Till now, they are just able to produce some medicine and insulin that can control diabetes levels. As a student of computer science, we tried to build something that would help people and make their life easier.

1.3 Problem Definition

Machine learning has changed our modern technology greatly. However, using this advanced technology for the help of our society we need to know its deep implementation and how to use it. Before proceeding with the solution we need to have clarity about the problem. So we have to understand the problem thoroughly. After going through the problem we have to be specific about our requirements. If we cannot figure out what is our requirements we cannot proceed with our solution.

1.4 Research Questions

The main questions those are emphasize in this research are given below:

1. The current condition of diabetic patients in our country.
2. The awareness of diabetic patients regarding their diet and workout.
3. The possible way to overcome this situation.
4. Use of technology in Bangladesh's health system.
5. Limitation of technology in Bangladesh.

1.5 Research Methodology

In this part we briefly discuss about our working structure. Start from data collection to final product every single section has been discussed here. Apart from that, we also discuss about the model that we have built. We have evaluated the performance and compare our system with traditional model. We have used different machine learning algorithms and test the accuracy to use the suitable algorithm in our model.

1.6 Research Objectives

The objectives of our work are given below:

1. By automating routine processes, clinicians can devote less time to data entry and more time to patient interaction.
2. User's begin by selecting a diet plan such as vegan or gluten-free.
3. The app's Chabot assists using for planning weekly meals by recommending recipes or menu uses meal option fits a user's diets.
4. Patients may also gain an increased sense of control over their diabetes by machine learning.
5. A better continuity of care may be achieved between face-to-face visits with physician and patient.

1.7 Research Layout

Chapter 1: will shortly discuss about introduction, motivation, Problem Definition, Research Question, Research Methodology and the expected outcome of our system.

Chapter 2: will explains the problem about Bangladesh perspective. We will also discuss about related work regarding this topic.

Chapter 3: will present elaborately method of the proposed model in the field of agriculture in Bangladesh.

Chapter 4: will discuss about the models performance and compare the models.

Chapter 5: It elaborate the conclusion and future perspective.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Bangladesh's people age over 40 are highly affected by diabetes. Yet there is digital platform that help them to control their food and lifestyle. Therefore, the background of the traditional techniques to control diabetics and digital technique to control diabetes should have explored.

2.2 Related Works

Nowadays, Technology is used everywhere in human life. Especially in modern medical treatment technology is playing an important role. Researchers from all over the world are trying to use machine learning to control people's diabetes levels. There are lots of initiatives that have been taken by the researchers to make food recommendation systems for patients that recommend food for users based on different criteria. People need nutrition to live a healthy life. For this, they eat a variety of healthy foods. But not all foods are suitable for diabetics. So a team of researchers tried to make a food recommendation system that can suggest alternative and suitable food for diabetic patients which also contain a good amount of nutrition [5]. K-means clustering and the self-organizing map have been in this system to recommend food that deals with eight very important nutrition. Another paper has been published by the researcher [6], where they collected data from the internet and various hospital about patient's different features diseases and tried to make a food recommendation system for them with the help of a deep learning algorithm. On the other hand, another machine learning technique developed by the researchers in [7] to detect chronic kidney disease risk and suggest food diet based on condition. First of they collect the potassium level in the blood from the patients and apply various classification algorithms to detect the condition of the patient. After getting the report on the patients' condition they suggest a food diet by using query-matching from their data set. For classification, they used various types of machine learning algorithms such as support vector machine, Naive byes, random forest, and so on. In the paper [8] the authors provide a personalized food recommendation system to prevent coronary heart disease. The algorithm, first of all, considers some basic information about the user such as important symptoms, family history, and blood pressure. The user can also change and customize the recommendation as per his preference. The authors in the paper [9] made a recommendation system for the user that recommends a healthy lifestyle to its user. The system was developed

in a mobile system that recommends healthy food, necessary workout for the American Indian. Moreover, its performance is evaluated by human experts. Various types of recommendation systems have been developed by researchers from all over the world into only in the medical sector but also in different fields such as Movies, restaurants, products, and so on. Researchers in the paper [10] developed a restaurants recommendation system where the model analyzes user behavior by using collaborative filtering and content-based filtering and suggest restaurant based on user preferences. In the paper [11] the authors collect a movie rating dataset from Twitter and analyze users' preferences. Based on that analysis the authors have built a movie recommendation system. In paper [15] the authors provide us valuable information regarding data collection process. Till now there are lots of working on recommendation systems is going on to make life easy and safe.

From the above scenario, we can conclude that there are some similarities and dissimilarities with our research. There is some other works have been done by the various researchers to suggest food for various disease using machine learning. So, we can use machine learning to build our system.

2.3 Bangladesh Perspective

As a developing country Bangladesh is doing well in all sector from the last 20 years. It's tech industry is also booming and occupying a good position in the international market. People are not very sensitive about their health and life style. Smart phones are also increasing in Bangladesh's market. Hence, if we can build a system that can recommend food and workout, it will be also beneficial in terms of tech market as well.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

The methodology has 6 steps that concludes our research that is in Figure. The steps are the following

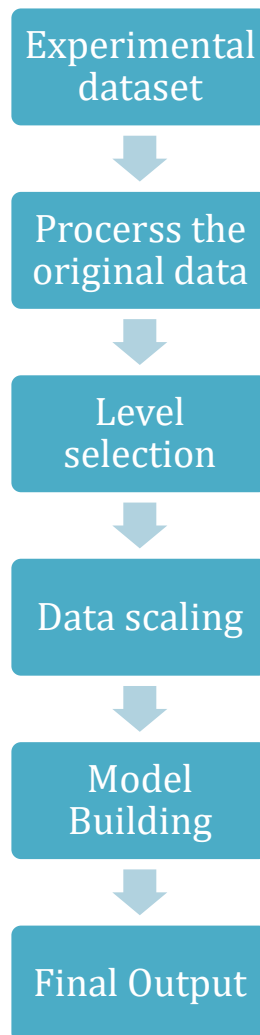


Figure 3.1.1: Process Structure.

3.2 Experiment Data Set

The first step to start a research project is collecting data. As there is no proper data set for Bangladeshi people about their lifestyle and diet plan, we had to invest a lot of time to collect data. We had created a google form and distributed it to our friends and family to collect information about their diet plan and lifestyle. In the end, we were able to collect 300 individuals' data.

3.3 Process the original data

The next step of our work is to process the raw data. As we collect the data in a google form. There were lots of noisy and irrelevant data that should have been tackled otherwise it would create problems for our next steps. Hence, we decide to dump messy and missing information. We have used various python libraries to process the dataset and give it a new shape.

3.4 Level Selection

One of the important parts of a machine learning algorithm is to select features wisely. There were almost 7-9 features in our initial dataset. However, some features are irrelevant in the context of our desired model. So, we have chosen 4 features that help us to make our project. They are as follows age, gender, sugar level, weight. We skipped the other's features. After all, for our desired system these four features are important because all the features that we mentioned are directly linked to our health condition.

Table 3.4.1: Features of the model

Feature Name	Feature type	Range
Age	Integer	(0- max)
Gender	String	<Male, Female>
Weight	Integer	(10-max)
Work out time	Integer	10,20,30.....
Blood Sugar level	String	<Normal, High, Low>

3.5 Data Scaling

Our next step is to convert our dataset in a format that will be understandable to our computers. Our dataset has some string columns but string columns are not understandable to the machine. So, we have used one-hot encoding from python's Sklearn and panda's dummies libraries to convert strings into numbers. Apart from that we also used min-max scaler module to make an equilibrium our whole dataset.

3.6 Model Building

When our dataset is prepared Now it's time to build the actual model that will help us to detect diabetic risk and recommend diet and workout as per our body need. We distribute our model into two different sections. In the first section, the system will classify users' condition and detect whether he or she is at risk of diabetes or not and, in the second part the system will recommend food and workout routine.

- Classification:

1. Support Vector Machine (SVM):

Support vector machine is a well-known and widely used classification algorithm to classify something into a group. It is a nonlinear classification algorithm so; it is quite convenient for us to use this algorithm in our dataset because our dataset has multiple features. This algorithm lies in how well we organize the decision plane. If there are multiple features and multiple clusters linear algorithms draw a random line to divide its cluster. However, SVM draws an optimal hyperplane through the dataset and give accurate decision [12].

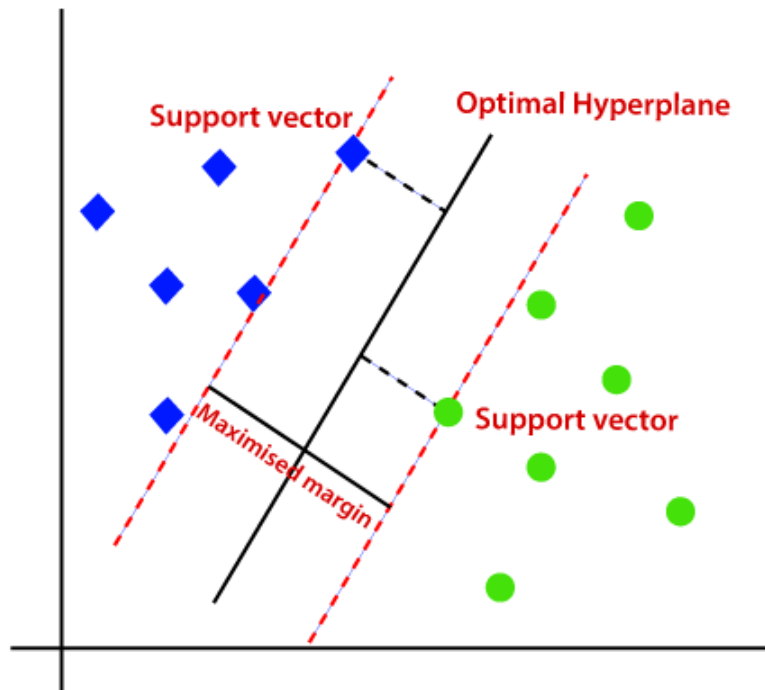


Figure 3.6.1: Internal structure of SVM

2. Naïve Bayes:

A simple yet powerful algorithm to classify something based on probability. Conditional probability is used in this algorithm to calculate the correct answer. This algorithm gives good results when the dataset is big as we know that probability performs well when a test case is big. The conditional probability works like this. Probability of event A knowing that event B has already occurred [13]. So, in our system, we give some information about the user such as age, weight, and so on of based on those events, the algorithm finds out the probability of diabetic risk.

$$p\left(\frac{A}{B}\right) = \frac{p\left(\frac{B}{A}\right) * p(A)}{p(B)} \dots\dots\dots(i)$$

3. Decision Tree Classifier:

The decision tree algorithm belongs to the supervised machine learning algorithm family. We can use it in classification problems. This algorithm determines the value of the target variable. To solve it the algorithm represents

the problem into a tree. First of all, we consider the whole dataset as a root [14]. There is a scientific term called entropy that helps to draw boundaries among the leaves of the decision tree. The impurity values lie between zero to one. The less the value the more accurate the decision. From the value of impurity, we can calculate the information gain. The information gain is used to determine which label to divide on at each step to build the tree.

- Recommendation:

After predicting whether the user is at risk of diabetes or not it's time to recommend the food diet and workout routine based on the results. To recommend the best food and living style we used the query selection method from the database to find out the best routine. For this, we set some parameters into our model. When a user gives information about him/her we first select the diabetic risk by using a classification algorithm. If the result is 0 that means the user is risk-free so, we can suggest his/her normal diet and workout. On the other hand, if the result gives us 1 that means the user has a high chance to be affected by diabetics. In this scenario, we use query onto our dataset to find out the best diet plan and workout routine similar to the user's age, weight and sugar level conditions. Then will suggest those top recommendations to the users who have diabetes risk zero. By this, we can recommend a healthy lifestyle to our users

CHAPTER 4

RESULT AND ANALYSIS

4.1 Training, Testing and the Validation of the model

We have divided our dataset into two parts. Train part to train our dataset. Test part to test the model that we have built. To do so we have used the Sklearn library of python. We have used 75% of our total dataset for training our model. Moreover, 25% of our dataset has been used for testing and evaluating the result.

4.2 Model efficiency

As we have used 3 algorithms in our dataset for classification Naive Bayes, Support vector machine, and Decision tree. All three models give us an impressive result of more than 50% accuracy.

Figure 4.2.1: Accuracy table of the model

Algorithm Name	Accuracy
Support Vector Machine	64%
Naïve Bayes	68%
Decision Tree	96%

From the above table, we can see that SVM and Naive Bayes gave us 64% and 68% accuracy respectively. Whereas, the decision tree classification model gives more than 90% accuracy.

4.3 Model comparison

To detect the diabetic risk, we need to use the best algorithm which can predict maximum right cases. Though we have used three different algorithms, we have to select only one out of this three. To select the best algorithm for our recommendation system we have used python's confusion matrix to find out which algorithm gives us true positive value most of the time.

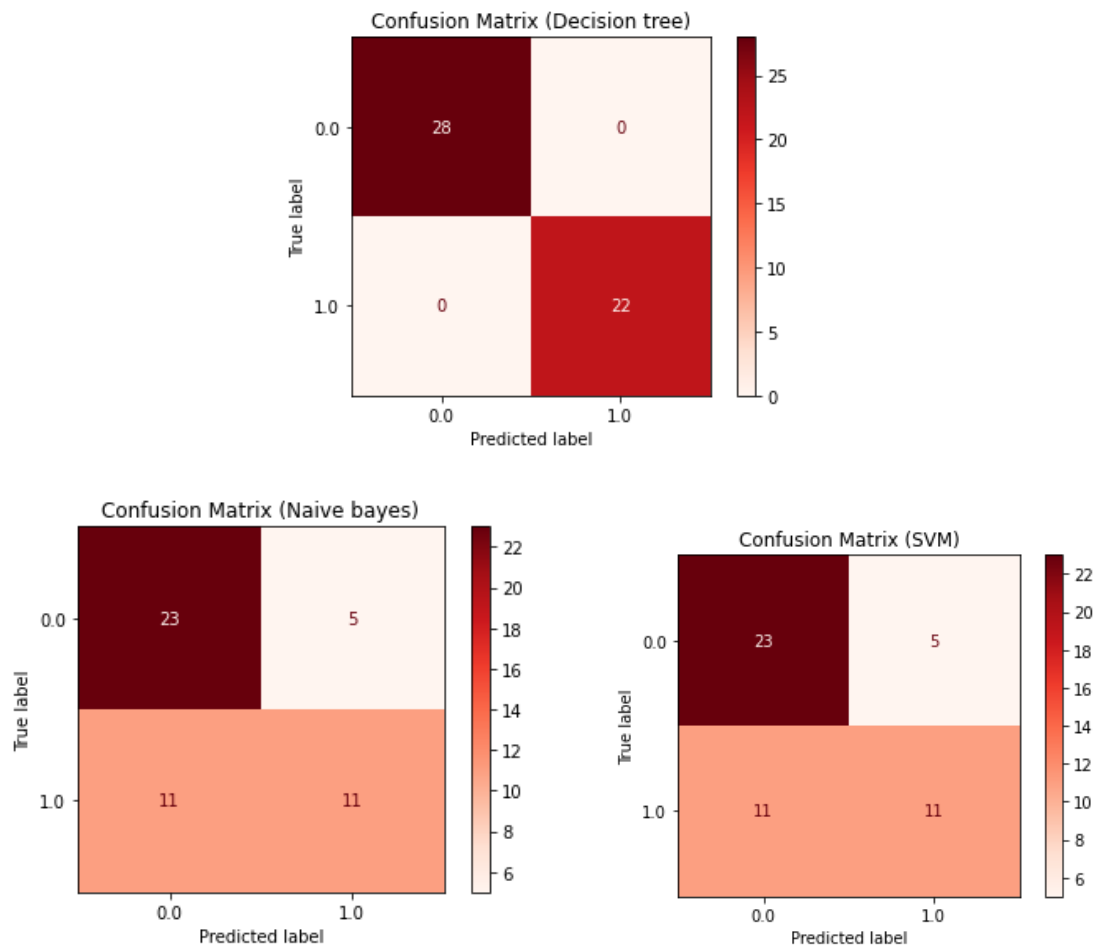


Figure 4.3.1: confusion matrix of all the model

From the above picture, we can see that the decision tree classification algorithm gives us the most accurate result as compared to Naive bayes and support vector machine algorithm. So we can use a decision tree as our best choice.

CHAPTER 5

CONCLUSION AND FUTURE WORK

The diabetic disease is a great concern in Bangladesh. Although our medical sector is doing good to fight against diabetes. However, we can use modern technology to help and secure our population from diabetes. In this paper, we have tried to utilize modern technology to fight against diabetes and give a healthy life to our people. We have tried to create a recommendation system that will help our users to control their diet and get a workout plan so that the user can stay fit and healthy. We have used various machine learning algorithms and user filter methods to present the recommendation. Though we have collected only 200 data so in near future we want to extend our project. In the future, we will try to collect more data about users and train it into our system. Hence, we will be able to make a more accurate recommendation system.

REFERENCE

- [1]. Introduction to diabetes <https://www.cdc.gov/diabetes/basics/diabetes.html>
- [2]. Medicine for diabetes patient <https://dtr.ucsf.edu/types-of-diabetes/type2/treatment-of-type-2-diabetes/medications-and-therapies/type-2-non-insulin-therapies/table-of-medications/>
- [3]. Foods for diabetes <https://www.niddk.nih.gov/health-information/diabetes/overview/diet-eating-physical-activity>
- [4]. Talukder, Ashis, and Md Zobayer Hossain. "Prevalence of diabetes mellitus and its associated factors in Bangladesh: application of two-level logistic regression model." *Scientific Reports* 10, no. 1 (2020): 1-7.
- [5]. Phanich, Maiyaporn, Phathrajarin Pholkul, and Suphakant Phimoltares. "Food recommendation system using clustering analysis for diabetic patients." In *2010 International Conference on Information Science and Applications*, pp. 1-8. IEEE, 2010.
- [6]. Manoharan, Samuel. "Patient diet recommendation system using K clique and deep learning classifiers." *Journal of Artificial Intelligence* 2, no. 02 (2020): 121-130.
- [7]. Banerjee, Anonnya, Alaa Noor, Nasrin Siddiqua, and Mohammed Nazim Uddin. "Food Recommendation using Machine Learning for Chronic Kidney Disease Patients." In *2019 International Conference on Computer Communication and Informatics (ICCCI)*, pp. 1-5. IEEE, 2019.
- [8]. Kim, Jong-Hun, Jung-Hyun Lee, Jee-Song Park, Young-Ho Lee, and Kee-Wook Rim. "Design of diet recommendation system for healthcare service based on user information." In *2009 Fourth International Conference on Computer Sciences and Convergence Information Technology*, pp. 516-518. IEEE, 2009.
- [9]. Alian, Shadi, Juan Li, and Vikram Pandey. "A personalized recommendation system to support diabetes self-management for American Indians." *IEEE Access* 6 (2018): 73041-73051.
- [10]. Varatharajan, Mr N., J. Guruprasad, and K. Mathumitha. "Restaurant Recommendation System Using Machine Learning." *International Educational Applied Research Journal* 4, no. 3 (2020): 1-4.
- [11]. Das, Debashis, Himadri Tanaya Chidananda, and Laxman Sahoo. "Personalized movie recommendation system using twitter data." In *Progress in Computing, Analytics and Networking*, pp. 339-347. Springer, Singapore, 2018.
- [12]. Noble, William S. "What is a support vector machine?" *Nature biotechnology* 24, no. 12 (2006): 1565-1567.
- [13]. Rish, Irina. "An empirical study of the naive Bayes classifier." In *IJCAI 2001 workshop on empirical methods in artificial intelligence*, vol. 3, no. 22, pp. 41-46. 2001.
- [14]. Safavian, S. Rasoul, and David Landgrebe. "A survey of decision tree classifier methodology." *IEEE transactions on systems, man, and cybernetics* 21, no. 3 (1991): 660-674.
- [15]. Das, Rozanee Kanta, Alaya Refat Tinni, Tanjina Zaman Rinvee, Sharun Akter Khushbu, and Abu Kaisar Mohammad Masum. "A Survey Analysis of Student Attachment on Computer Science Towards Universities Purposes." In *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1-8. IEEE, 2021.

PLAGIARISM REPORT

Plagiarism Checked by
Abdus Sattar, Assistant Professor, Department of CSE



01-12-2021

Diabetes Health Care Routine Using Machine Learning

ORIGINALITY REPORT

7%

SIMILARITY INDEX

6%

INTERNET SOURCES

0%

PUBLICATIONS

3%

STUDENT PAPERS

PRIMARY SOURCES

1

Submitted to Daffodil International University

Student Paper

3%

2

emerj.com

Internet Source

2%

3

www.mdpi.com

Internet Source

1%

4

dspace.nwu.ac.za

Internet Source

<1%

5

www.scribd.com

Internet Source

<1%

6

younglives-india.org

Internet Source

<1%

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off