Nutritionist for Diabetic Patients

Kavya Guruprasad
Indiana University
prasadk@umail.iu.edu

Nandini Goswami
Indiana University
goswamin@umail.iu.edu

Sarita Bhateja
Indiana University
sbhateja@umail.iu.edu

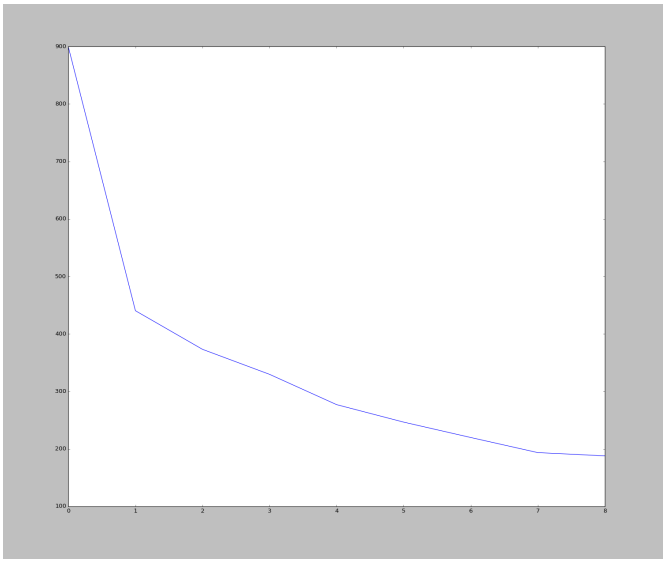
ABSTRACT
This paper presents our project Nutritionist for Diabetic Patients. This project helps diabetic patients find a better substitute of food, thereby helping them to reduce their overall blood sugar level. With the help of machine learning algorithms like linear regression and K-Nearest Neighbors, and AI techniques like Case Based Reasoning has helped us to achieve this. This project has the potential to positively impact the health of a diabetic patient.

Keywords
Linear Regression, K-Nearest Neighbor, Case Based Reasoning, dataset

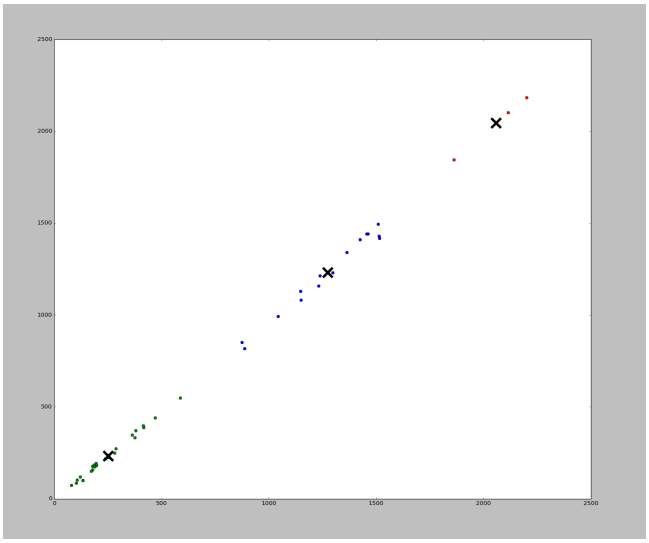
1. INTRODUCTION
According to statistics shared by CDC (Centers for Disease Control and Prevention), there are more than 29 million Americans who are suffering from diabetes and one in four don't know about it.Eighty Six million adults who constitute to more than one in three U.S. adults, have prediabetes. These numbers are alarming and needs increased focus on the dietary needs of these patients. With the help of Machine learning Algorithms like Linear regression and K-Nearest Neighbors and AI techniques like CBR, we are trying to find a better food substitute for the patient. Our dataset contains food names, their nutrition values and the Glycemic index values. A higher Glycemic index value indicates higher carbohydrate level and hence should be avoided by patients.

2. ALGORITHMS USED
2.1 Linear regression
The Linear Regression model helps to predict Glycemic index values depending on the food nutrients. The GI value ranges from 0-100. GI is classified as : Low GI(0-55),Medium GI(56-69) and High GI(70-100) We have considered applying linear regression model, as the GI has continuous values. Using GI value and net carbohydrate content we will calculate GL, that will in turn help us decide whether the food is appropriate for the patient or not. Using GI values we will classify foods into low, medium and high glycemic index food that will help us predict better food substitutes for the patient.

2.2 K-Nearest Neighbor
K-means is an unsupervised learning algorithm which groups similar data. Food samples from the dataset that have low range of GI value is given as input to the k-NN algorithm. When a user inputs a food item, clustering will be done in a way to find similar nutrients, but with a lower GI index. The best result will be returned as per the least distance. The number of clusters is given by k. Elbow method is used to find the optimal value of k. In this method, we plot a graph with number of clusters on the x-axis and average distance to the centroid and look for an elbow in the graph. The point when the line starts to flatten out is the k. We obtained the following graph



From the above graph, it is clear that we need to choose the value of k as 3. Giving the value of k as the input to our clustering algorithm, we can plot the graph of the data points and their assigned clusters.



2.3 Case Based Reasoning

Case Base involves four steps:

- Retrieve: We retrieve food samples from dataset and try to use its knowledge to solve new cases.
- Reuse: Once a substitute is found, we try to reuse this knowledge again and again.
- Revise: Once we find a solution we try to confirm from the user whether the solution is good enough.
- Retain: After the solution has been successfully adapted to the target problem, we store it into our database and use it for solving newer problems.

Thus, if a patient agrees to the food item suggested by the system, the combination will be saved in the knowledge base and if similar case occurs in future, the results will be fetched from the knowledge base.

2.4 How is our approach different?

Linear regression is a good approach to determine the GI index as we have all the related nutrient information.

The use of k-NN to provide a food substitute is a new approach followed in the project.

2.5 How to scale our project?

With the right data set, we can include an option of food preference factor(non-vegetarian, vegan, vegetarian, etc) to give more appropriate food substitute. Also, taking into consideration the blood group of the patient could give better results while suggesting substitutes.

2.6 Strengths and Weaknesses

Firstly, the project takes into consideration, the allergic restrictions that a person might have and gives a substitute if the user is not willing to take it.Secondly, it can be used by diabetic as well as non-diabetic people.For instance, patients suffering from obesity, thyroid, PCOD/PCOS can also use this application, as all these diseases are inter-linked. It can also be considered as a nutritionist, which helps to check the quantity of food intake and provides healthier food choices. Thirdly, It has an easy to use interface.

Some of the weaknesses are that we did not find a relevant dataset which can be used directly in the project. Two datasets have been merged to get the desired dataset which resulted into loss of data points. It also has a dependency on domain expert.

2.7 Sample program Output

- The first case is for when food exits in our data set and the GL is high for the food then a substitute is returned and saved in database. The substitute is found using K-means and saved to knowledgebase.

  Program output: Enter your age:27
  Enter your gender(F/M):F
  Enter your blood group(A,B,AB,O):B
  Enter your blood sugar level:143
  Enter your food : Bagel, from white flour, commercial
  Enter amount of food(gms): 1222
  GL=644.394816
  Food is inappropriate for consumption
  Do you want a healthier food substitute(Y/N):Y
  Healthier substitute will be: Apple, dried
  Is this a good substitute or are you allergic to the substitute(Y/N):Y

  Database:
  27, 'F', 'B', '143.0', 'Bagel, from white flour, commercial', 'Apple, dried'

- The second case is for when food exits in our data set and the GL is high for the food then a substitute is returned from the existing knowledge base that was saved in the knowledge base from the first case. No knowledge base update takes place here.

  Program output: Enter your age:30
  Enter your gender(F/M):F
  Enter your blood group(A,B,AB,O):B
  Enter your blood sugar level:144
  Enter your food:Bagel, from white flour, commercial
  Enter amount of food(gms):1233
  GL=650.195424
  Food is inappropriate for consumption
  Do you want a healthier food substitute(Y/N):Y
  Healthier substitute will be: Apple, dried
  Is this a good substitute or are you allergic to the substitute(Y/N):Y

- The third case is for an unknown food entered by the user and the the user is asked to enter all the nutrient value of food to be consumed. Using these values glycemic index is predicted by linear regression and further then that the process remains the same. Enter your age:33
  Enter your gender(F/M):M
  Enter your blood group(A,B,AB,O):O
  Enter your blood sugar level:122
  Enter your food:MANGO
  Enter amount of food(gms):13

Enter Energywithdietaryfibre(kJ):1
Enter Energywithoutdietaryfibre(kJ):1
Enter Moisture(g):1
Enter Protein(g):1
Enter Totalfat(g):1
Enter Available carbohydrates with sugar alcohols(g):1
Enter Availablecarbohydrateswithoutsugaralcohol(g):1
Enter Starch(g):1
Enter Totalsugars(g):1
Enter Addedsugars(g):1
Enter Freesugars(g):1
Enter Dietaryfibre(g):1
Enter Alcohol(g):1
Enter Ash(g):1
Enter Preformed vitamin A(retinol)(g):1
Enter Beta - carotene(g):1
Enter Provitamin A(b - carotene equivalents) (g):1
Enter Vitamin A retinol equivalents(g):1
Enter Thiamin(B1)(mg):1
Enter Riboflavin(B2)(mg):1
Enter Niacin(B3)(mg):1
Enter Niacin derived equivalents(mg):1
Enter Folate, natural(g):1
Enter Folic acid(g):1
Enter Total Folates(g):1
Enter Dietary folate equivalents(g):1
Enter Vitamin B6(mg):1
Enter Vitamin B12(g):1
Enter Vitamin C(mg):1
Enter Alpha - tocopherol(mg):1
Enter Vitamin E(mg):1
Enter Calcium(Ca)(mg):1
Enter Iodine(I)(g):1
Enter Iron(Fe)(mg):1
Enter Magnesium(Mg)(mg):1
Enter Phosphorus(P)(mg):1
Enter Potassium(K)(mg):1
Enter Selenium(Se)(g):1
Enter Sodium(Na)(mg):1
Enter Zinc(Zn)(mg):1
Enter Caffeine(mg):1
Enter Cholesterol(mg):1
Enter Tryptophan(mg):1
Enter Totalsaturatedfat(g):1
Enter Totalmonounsaturatedfat(g):1
Enter Totalpolyunsaturatedfat(g):1
Enter Linoleic acid(g):1
Enter Alpha - linolenic acid(g):1
Enter C20:5 w3 Eicosapentaenoic(mg):1
Enter C22:5 w3 Docosapentaenoic(mg):1
Enter C22:6 w3 Docosahexaenoic(mg):1
Enter Total long chain omega3 fatty acids(mg):1
Enter Totaltransfattyacids(mg):1
GI=37
GL=0.2405
Food is appropriate for consumption

- The fourth case is when user has entered food with Low GI but due to the amount of serving to be consumed the GL value is high. In this case the user is asked to consume food in lesser quantity or find a substitute. Also, if the returned substitute is not appropriate or the user is allergic substitutes the system keeps recommending substitutes till the user finds appropriate substitute.

Enter your age:55
Enter your gender(F/M):M
Enter your blood group(A,B,AB,O):B
Enter your blood sugar level:133
Enter your food:Apple, dried
Enter amount of food(gms):1222
462.11152
If food is consumed in lesser quatity, it would be fit for consumption
Do you want a healthier food substitute(Y/N):y
Healthier substitute will be: Apricot, dried
Is this a good substitute or are you allergic to the substitute(Y/N):n
Healthier substitute will be: Bread, tortilla, corn, commercial
Is this a good substitute or are you allergic to the substitute(Y/N):n
Healthier substitute will be: Bread, from white flour, sour dough, homemade from basic ingredients, toasted
Is this a good substitute or are you allergic to the substitute(Y/N):n
Healthier substitute will be: Bread, Naan, homemade
Is this a good substitute or are you allergic to the substitute(Y/N):n
Healthier substitute will be: Bread, mixed grain, commercial
Is this a good substitute or are you allergic to the substitute(Y/N):n
Healthier substitute will be: Bread, pumpkin
Is this a good substitute or are you allergic to the substitute(Y/N):n
Healthier substitute will be: Bread, Roti, commercial
Is this a good substitute or are you allergic to the substitute(Y/N):n
Healthier substitute will be: Bread, soy linseed, commercial
Is this a good substitute or are you allergic to the substitute(Y/N):n
Healthier substitute will be: Breakfast cereal, wheat bran, flakes, unfortified
Is this a good substitute or are you allergic to the substitute(Y/N):n
Healthier substitute will be: Breakfast cereal, oat whole wheat, biscuit, added salt, unfortified
Is this a good substitute or are you allergic to the substitute(Y/N):n
Healthier substitute will be: Cake mix, sponge, dry powder
Is this a good substitute or are you allergic to the substitute(Y/N):n
Healthier substitute will be: Cake or cupcake, banana, homemade from basic ingredients, oil, uniced
Is this a good substitute or are you allergic to the substitute(Y/N):n
Healthier substitute will be: Cake or cupcake, chocolate, commercial, uniced
Is this a good substitute or are you allergic to the substitute(Y/N):n
no more substitutes available

- The fifth case is when the user finds the substitute appropriate. Enter your age:12
  Enter your gender(F/M):F
  Enter your blood group(A,B,AB,O):A
  Enter your blood sugar level:140
  Enter your food:Biscuit, sweet, honey jumble, homemade from basic ingredients, fat not further defined, uniced
  Enter amount of food(gms):1200
  Glycemic Load: 5931.9036
  Food is inappropriate for consumption
  Do you want a healthier food substitute(Y/N):Y
  Healthier substitute will be: Chocolate, not further defined
  Is this a good substitute and ensures no allergies? (Y/N):Y

2.8 Evaluation

We tried other ML algorithms before coming down to Linear Regression and K-means such as Decision tree but results were not satisfactory. When the linear regression was run on complete dataset, the results were not good enough to proceed. We ran Random Forest algorithm to identify the importance of all the feature and features with least importance were removed. This process was repeated over and over to identify when to stop removing the features from the dataset. This process improved the performance of Linear Regression. The results are not at par, however, good enough to get started with the project. The next algorithm K-means is helping in the adaptation part and providing substitute to the user. In this algorithm, number of clusters to be chosen is an important factor for the algorithm to converge. To identify the best possible number of clusters required, we ran the K-means with number of clusters ranging from 1-10 and used Elbow method to choose the best k (number of clusters). The results were better with number of clusters = 3. The limited number of data points is a weakness because K-means performs good with large dataset. The K-means algorithm with CBR for adaptation is working better than expected and results are quite interesting. When we tested the results against test data points with various scenarios, the results were fair. Also, we shared the results with domain expert to get insight into it. To her knowledge, the results are average and there is scope of improvement. The project has a lot of potential to leverage if given enough data points to analyse in depth. Also, domain expert has a huge role to play in this project.

2.9 Clarity on the program

For this program we have merged two data sets where one contained all nutrient values and the other contained Glycemic index corresponding to food name. [2] [3] [1]

In our program we firstly take input from the user about their age, gender, blood group, sugar level, serving and the food they are going to consume.

If the food exists in our dataset we find the glycemic index for the food from dataset or asks the user to enter all nutritive values and using linear regression, predicts GI of the food. Further, we calculate the glycemic load for the food to find whether the food is appropriate for the user or not. Conditions for approving food:

- If food has Glycemic Load below 10 it is appropriate for people of all age, sugar level and gender.
- If food has Glycemic Load in range of 10-19 and user has sugar level below 100 which is upper value for normal blood sugar level, food is appropriate for user intake.
- If food has Glycemic Load above 20 but Glycemic Index is less than 56, then this food if taken is smaller quantities can be appropriate for the user for consumption.
- If Glycemic Load is above 20 and Glycemic Index is more than 56, then the food is inappropriate for the user.

If the food is inappropriate for consumption the system asks the user whether user wants a healthier substitute. If the user enters 'no' the code exists else the system finds a healthier substitute for the user. For finding healthier substitute the user performs two steps:

- If a substitute exits in our knowledge base for a user with similar medical background the substitute is returned from our case base.
- If food does not exists in our knowledge base then food substitute is found using kmeans.

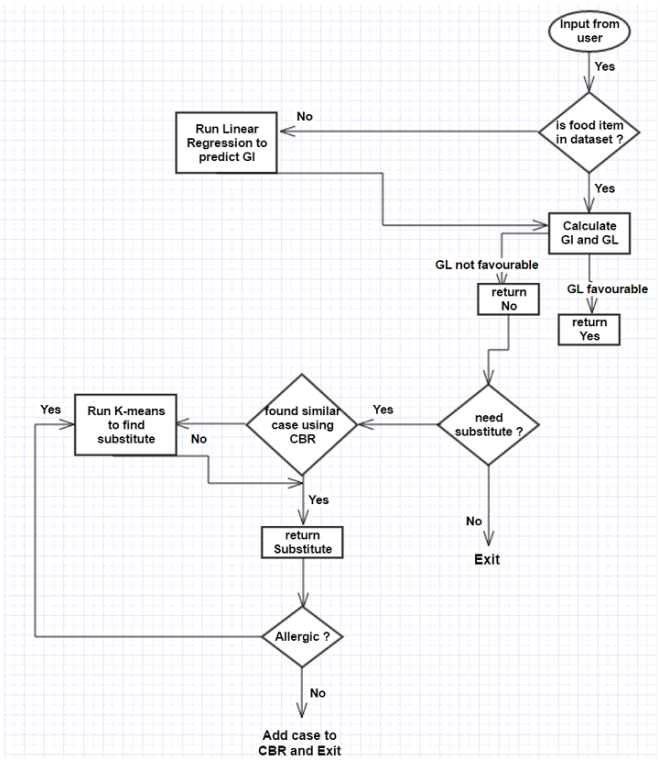
For running kmeans user entered food nutrients are added to our kmeans input data set. Kmeans input contains nutrients of all food containing Glycemic index below 55. After clustering all all the food that have same label as our food data point are taken as contending food substitutes. The one with lowest glycemic index and good nutritive value substitute food is returned.

The user is then asked whether the substitute is good and is the user allergic to the food substitute. If the user finds the substitute is good, the substitute is saved in the knowledge base and code exists. Else if the user enters no the system finds the second best food substitute and returns that. The process continues till the user does not find an appropriate substitute and enters yes or the substitutes available in the system exhausts.

2.10 Related work

2.10.1 The 4 Diabetes Support System, A Case Study in CBR Research and Development

This project helps patients with type 1 diabetes who are on insulin pump therapy achieve and maintain good blood glucose control. There has been a lot of changes in the techniques and machine learning techniques that have been used in this project. But the latest project uses Naive Bayes classification and support vector regression were extensively used. How Glycemic varies with food and also the blood glucose level prediction were taken into consideration. The AI technique Case Based Reasoning has been used. This new work has the potential to positively impact the health and well being of patients with diabetes.[5]



2.10.2 PEPPER

PEPPER Pepper is a personalised decision support system which empowers people who are taking in insulin injections to manage their conditions. It uses Case-Based reasoning to advise people their insulin doses, taking as input various aspects such as physiological, lifestyle, environmental, and social data. It also uses Model-Based reasoning to achieve users safety. It is integrated with the an insulin patch pump and aims at being patient-centric development approach that is it improves the patients self-efficacy to stick to a particular treatment.[7]

2.10.3 *Artificial Intelligence Techniques for Diabetes Management: the T-IDDM Project*

This application uses Artificial Intelligence(AI) methodologies in the telemedicine service domain,for diabetic patients management. The system architecture is distributed, and composed by a Patient Unit and by a Medical Unit, connected through a telecommunication link. The project uses many AI techniques to implement the functionality. The data base comprises of the domain knowledge that is given to the system. Case Based Reasoning (CBR) is applied to perform the Knowledge Management task. Finally, CBR is integrated with Rule Based Reasoning to provide which gives the physicians a model which is a good reasoning decision support system. This project could lead to the management of diabetes in type 1 diabetic patients, which leads to a more tight control of a patients' metabolic rate and it proves to be quiet cost effective.[9]

2.10.4 *Enhancing an Artificial Pancreas with an Adaptive Bolus Calculator based on Case-Based Reasoning*

In the previous systems, a pre-meal insulin bolus is required to avoid hyperglycemia. This pre-meal is computed by standard bolus calculation, as is found in insulin pumps. The performance of these calculations is less because the adapting nature is low. In this project, a new technique is introduced which will adapt a meal bolus on an artificial pancreas which is based on Case-Based Reasoning. Simulation results showed using adaptive meal bonus calculator improved the glycemic control of type 1 diabetic patients.[8]

2.10.5 *Temporal case-based reasoning for bolus decision support*

People suffering from Type 1 diabetes have to find the insulin consumption frequently before every meal, to keep their blood glucose level maintained. To find this, they calculate the bolus which gives the appropriate doses for consumption. However,these calculators do not automatically adapt to improve bolus suggestions. In order to automate this, the project has come up with artificial intelligence technique of case-based reasoning to personalise the bolus decision support. The CBR retrieves the appropriate cases for reuse.This approach is able to learn and improve the bolus prediction [6]

2.10.6 *Monitoring patients with diabetes using wearable sensors:Predicting glycaemias using ECG and respiration rate*

Sensors have proven to be a great monitoring device for patients with diabetes and can improve their quality of life. In the pilot study, patients with type I and II diabetes were equipped with a series of such sensors. The chest harness sensor records the ECG signal and the respiration rates. Predictions were made to recognise abnormal glucose blood levels in type I and II diabetes patients. An accuracy of 84 percent was obtained in predicting glycemia for patients with type I diabetes and 88 percent for patients with type II. For recognition of glycaemia, an accuracy of 78 percent for type I and 76 percent for type II was obtained.[4]

2.11 ACKNOWLEDGMENT

We would like to extend our respect and thank our Professor David Leake for the guidance and support he has offered. Learning from him was always a pleasure, as it made more sense to Artificial Intelligence.

3. REFERENCES

[1]
[2] M. Al Sears.
[3] M. Al Sears.
[4] A. G. a. M. L. BoÅ¿idara CvetkoviÄĞ, and UrÅąka Pangerc. Monitoring patients with diabetes using wearable sensors: Predicting glycaemias using ecg and respiration rate. Paper.
[5] T. C. R. B. J. S. F. S. Cindy Marling, Matthew Wiley. The 4 diabetes support system:a case study in cbr research and development.
[6] C. M. I. B. Daniel Brown, Rachel Harrison. Temporal case-based reasoning for bolus decision support. Paper.
[7] C. M. Pau Herrero, Beatriz Lopez. Pepper: Patient empowerment through predictive personalised decision support. Paper.

[8] P. P. N. O. P. G. Pau Herrero, Jorge Bondia. Enhancing an artificial pancreas with an adaptive bolus calculator based on case-based reasoning. Paper.

[9] A. R. C. L. L. P. M. S. Stefania Montani, Riccardo Bellazzi. Artificial intelligence techniques for diabetes management: the t-iddm project. Paper.