**DIET RECOMMENDATION SYSTEM FOR DIABETES USING**

**MACHINE LEARNING**

*Sumathy V [[1]](#footnote-0) , Sangeetha J [[2]](#footnote-1) , Prince Godwin I [[3]](#footnote-2)*

*1,2Loyola College, Chennai*

*3 St.Joseph College, Tiruchirapalli*

**Abstract:**

In recent years, diabetes has become a prevalent and rapidly increasing global health concern. The primary driver of this surge is attributed to cultural shifts in dietary habits and a decline in physical activity. Effectively diagnosing and managing diabetes through timely dietary interventions can mitigate the risk of severe health complications associated with the condition, including heart attacks, kidney failure, blindness, stroke, and lower limb amputation. The proposed diet recommendation system, employing machine learning, is envisioned as an efficient tool for managing diabetes in patients. Leveraging data from individuals with diabetes, encompassing factors such as age, gender, blood sugar levels, weight, triglycerides, BMI, and physician-prescribed remarks, a nutritionist has tailored appropriate dietary recommendations. The system further involves training classification models, such as the Random Forest classifier, Naïve Bayes classifier, and Decision Tree, using the patient data and prescriptions. Subsequent evaluation will determine the most effective model for personalized diet recommendations in diabetes management.

**Keywords:**

Diabetes, Diet Recommendation, Random Forest classifier, Naïve Bayes classifier, Decision Tree.

**1.Introduction:**

Recently, the fields of data science and information technology have become crucial in delivering healthcare services.(1). Diabetes is a persistent metabolic condition marked by heightened blood glucose levels, resulting in gradual harm to vital organs such as the heart, blood vessels, eyes, kidneys, and nerves. Approximately 422 million individuals worldwide suffer from diabetes, with the majority residing in low- and middle-income nations. Each year, 1.5 million deaths are directly linked to diabetes. The incidence and prevalence of diabetes have been consistently rising over recent decades. There is a unified global objective to cease the escalation of diabetes and obesity by 2025.(2).

Timely diagnosis and appropriate medication are crucial for effective management of diabetes.Body factors like diet, physical activity, obesity, and weight management significantly impact the initiation and advancement of conditions that contribute to the onset of type 2 diabetes and its subsequent complications.(3). Leveraging machine learning, the prediction and management of diabetes, along with the provision of suitable diet recommendations, can be significantly enhanced. This study focuses on utilizing machine learning methodologies, including the Gaussian Naive Bayes classifier, Decision Tree, and random forest, with the objective of developing an efficient diet recommendation system for the effective management of diabetes.

**2.Literature Survey:**

Shafi et al. used the machine learning techniques such as Naive Bayes classifier, Random forest Classifier and Decision Tree classifier for the diagnosis of diabetes and recommend proper diet for diabetic patients through diet recommendation system(DRS) (4).

Sultana et al. proposed an approach for developing diabetes prediction and recommendation system by implementing multiple linear regression (5). Hassan et al. by using ensembling of many machine learning classifiers and employing methods such as Multilayer Perception, XG Boost, Adaboost, k-nearest neighbor, Random Forest, Decision Tree and Naive Bayes to build the diabetes prediction system (6).

Sowah et al. built an Tensor flow neural network model for determining whether a meal should be recommended, implements k-nearest neighbor algorithm for the diet recommendation and uses cognitive sciences to build a chat bot for answering questions related to diabetes (7). Devi et al. employed clustering techniques to categorize patients based on features such as age, weight, blood glucose levels, etc. Additionally, they enhanced the Krill-Herd optimization method to identify the most suitable neighborhood for generating personalized nutritional recommendations.(8)(9).

In 2020, Manoharan et al. implemented k-clique deep learning classifier, to develop a diet recommendation system and compared it with the recommendation system developed by employing techniques such as logistic regression and Naive Bias and other deep learning classifier such as the MLP and RNN to demonstrate the proficiency of K-DLRS (10).

Norouzi et al. in 2017, analyzed various diabetes diet recommendation systems and revealed in the review that CFRS, KBRS and CARS methods are the most common systems used for the development of food recommender systems (11).

**3.Methodologies:**

Our primary goal is to examine the data-set containing details like age, gender, BMI (Body Mass Index), occupation, FBS (Fasting Blood Sugar), PPBS (Post-Prandial Blood Sugar), baseline HBA1C, cholesterol, triglycerides, HDL (High Density Lipo Protein), and LDL (Low Density Lipo Protein), with the aim of forecasting the suitable dietary recommendations for individuals with diabetes.

**3.1 Data Collection and Preprocessing:**

The information from 300 diabetes patient records was gathered from Apollo Hospital in Chennai. This dataset includes parameters such as height, weight, BMI, hip-waist ratio, job type, age, sex, FBS, PPBS, LDL, triglycerides, cholesterol, and diagnosis. Following data cleansing and exploratory analysis, the nutritionist's dietary recommendations were incorporated into the dataset, with a particular focus on features like BMI and the type of diagnosis. Based on the type of diabetes and BMI, the patients were divided into four groups: Diabetes Mellitus related to overweight, obese grade 1, obese grade 2, and severe obesity. The suggested diet from the dietitian was then included to the dataset as input.

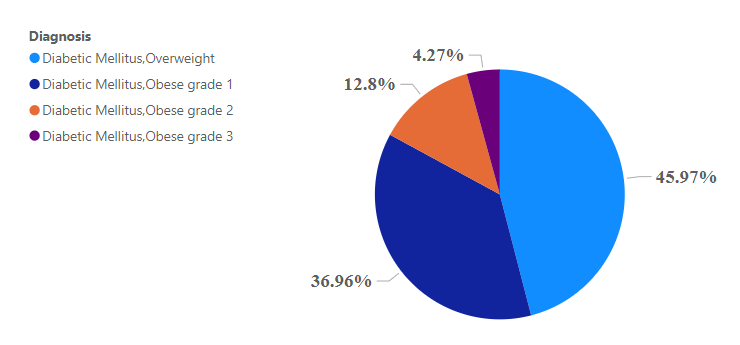


Figure 3.1.1 percentage of people vs diagnosis.

**3.2 Multi-target classification:**

The recommendation system needs to categorize each instance and suggest an appropriate diet for each category. Consequently, the target variables for this system are the type of diagnosis and the recommended diet, requiring the use of a multi-target classification approach. To build the model, various classifiers, including the random forest classifier, Gaussian Naive Bayes classifier, and Decision Tree classifier, have been employed.

**3.2.1 Random Forest Classification:**

A Random Forest Classifier is a type of ensemble learning approach utilized for both classification and regression tasks. The method involves creating numerous Decision Trees during training and generates predictions by taking the mode of the classes (for classification) or the average prediction (for regression) from the individual trees. Each tree within the forest is constructed using a distinct subset of the training data, and in the training process, random subsets of features are considered for each split in the Decision Trees.

**3.2.2 Gaussian Naive Bayes Classification:**

Gaussian Naive Bayes is a probabilistic machine learning algorithm used for classification tasks. Specifically, the algorithm assumes that the continuous-valued features follow a Gaussian (normal) distribution within each class. During training, Gaussian Naive Bayes calculates the mean and standard deviation of each feature for each class, forming a model of the Gaussian distribution for each feature in each class. When making predictions, the algorithm uses these distributions to calculate the likelihood of observed feature values given each class. The final prediction is made by selecting the class with the highest posterior probability, combining likelihoods with prior probabilities.

*P*(*A*∣*B*)=P(B\A)\*P(A)/P(B) ………..[1]

**3.2.3 Decision Tree Classification:**

The Decision Tree addresses the problem, by utilizing a tree representation in which the internal node of the tree represents attributes and each leaf node corresponds to a class label. Any Boolean function can be represented on discrete attributes using the Decision Tree. Data is split on the feature that yields the maximum Information Gain (IG), starting at the root.

**Entropy:**It quantifies the impurity or randomness of each outcome following its division at each level. The aim is to diminish it as additional levels are introduced.

𝐸(𝑆) = ∑𝑐 −𝑝 𝑙𝑜𝑔 𝑝 …………….[2]

The entropy value always falls within the range of 0 and 1.

**Gini Impurity:** Gini Impurity shares similarities with the concept of entropy, but the implemented formula differs slightly. In many instances, Gini impurity is favored over entropy because it is computationally more efficient, requiring less execution time due to the absence of a logarithmic function.

𝐺𝑖𝑛𝑖=1−∑𝑐 (𝑝)2  ….…….[3]

The Gini impurity value always falls within the range of 0 to 0.5.

**Information Gain:** Information gain represents the quantity of information that can be acquired at each level, aiding in the decision-making process to expand the Decision Tree model. The primary objective is to enhance the information gained at every stage. As entropy diminishes, information gain rises.

𝐺𝑎𝑖𝑛(𝑆,𝐴)=𝐸(𝑆)−∑ |𝑆𝑣|𝐸(𝑆 ) …………[4]

1. **Performance Analysis:**

In this segment, we are assessing the effectiveness of the multi-target classifiers incorporated during model training. We are evaluating the performance of the models by employing metrics like accuracy to gauge their effectiveness.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **Actual Class** | |
|  |  | Positive (P) | Negative (P) |
| **Predicted Class** | Positive (P) | True Positive (TP) | False Positive (FP) |
| Negative (N) | False Negative (FN) | True Negative (TN) |

Accuracy is a performance metric that measures the overall correctness of a classification model.

Accuracy= …………[5]

1. **Result Analysis:**

In the context of suggesting diets for diabetic patients using the dataset features, we employed multi-target classification models, including the Random Forest Classifier, Gaussian Naive Bayes Classifier, and Decision Tree. The respective accuracies achieved for these models were 0.96, **0.98**, and 0.97. Notably, the Gaussian **Naive Bayes Classifier** emerged as the most effective model among these, demonstrating the highest accuracy.

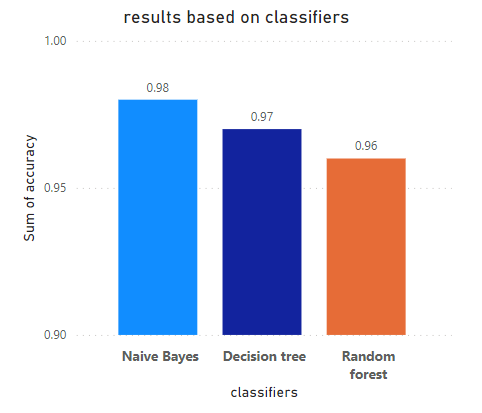


Figure 5.1 results based on classifiers.

1. **Conclusion:**

In conclusion, this study aimed to develop an efficient diet recommendation system for diabetes management within the healthcare system, utilizing machine learning methodologies. The objective was to enhance the process of diabetes management through timely and personalized dietary recommendations. Among the models employed, including the Random Forest Classifier and Decision Tree, the Gaussian Naive Bayes Classifier demonstrated the highest accuracy of 0.98. This achievement signifies the successful development of a diet recommendation system, showcasing the potential of machine learning in advancing healthcare practices for diabetes management.

# **7.References:**

1. N.Jothi, W.Husain,et al. *Data mining in healthcare-a review.* s.l. : Procedia Computer Science, 2015.

2. *www.who.int,WHO report of diabetes hybrid meeting .* s.l. : WHO, 2023.

3. *Lifestyle and outcome among patients with type 2 diabetes,Int.Congr.Series,vol.1303,pp.160-171.* G.Hu and J.Tuomiletho. 2007.

4. *Predictions of Diabetes and Diet Recommendation System for Diabetic patients using Machine Learning Techniques,https://www.researchgate.net/publication/352668574.* al., Salliah Shafi et. 2021. DOI: 10.1109/INCET51464.2021.9456365.

5. *An Approach for developing Diabetes Prediction and Recommendation System.* al., Saima Sultana et. Jan 2021, Dhaka,Bangladesh : International Journal of Computer Applications, 2021, Vol. 174.

6. *Diabetes Prediction using Ensembling of different Machine Learning Classifiers.* al., MD.Kamrul Hasan et.

7. *Design and development of diabetes management system using machine learning.* al., Sowah et. 2020. 8870141.

8. *Personalized nutrition recommendation for diabetic patients using improved k-means and krill-head optimization.* al., Devi et. 2020. 1076–1083..

9. *A systematic review on Food recommender systems for diabetic patients.* Raciel Yera, Ahmad A.Alzahrani et al.

10. *Patient Diet Recommendation system using K-clique and deep learning Classifiers.* al., Samuel Manahoran et. s.l. : https://www.researchgate.net/publication/348538735, 2020.

11. *Food recommender system for diabetic patients :a narrative review.* 2021.

​

1. [↑](#footnote-ref-0)
2. [↑](#footnote-ref-1)
3. [↑](#footnote-ref-2)