IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 8, June 2025
Impact Factor: 7.67

Diabetes Detection using K-Nearest Neighbour Classification Algorithm

M. Navyasri
Assistant Professor
School of Business, Galgotias University, Greater Noida, UP, India
mullapudi.navyasri@galgotiasuniversity.edu.in

Abstract: *Diabetes is a chronic disease with the potential to cause a worldwide health care crisis. According to International Diabetes Federation 382 million people are living with diabetes across the whole world. By 2035, this will be doubled as 592 million. Diabetes is a disease caused due to the increase level of blood glucose. This high blood glucose produces the symptoms of frequent urination, increased thirst, and increased hunger. Diabetes is a one of the leading causes of blindness, kidney failure, amputations, heart failure and stroke. When we eat, our body turns food into sugars, or glucose. At that point, our pancreas is supposed to release insulin. Insulin serves as a key to open our cells, to allow the glucose to enter and allow us to use the glucose for energy.The algorithms like K nearest neighbour, Logistic Regression, Random Forest, Support vector machine and Decision tree are used. The accuracy of the model using each of the algorithms is calculated. Then the one with a good accuracy is taken as the model for predicting the diabetes.*

Keywords: Machine Learning, Diabetes, Decision tree, K nearest neighbour, Logistic Regression, Support vector Machine, Accuracy

I. INTRODUCTION

Healthcare sectors have large volume databases. Such databases may contain structured, semi-structured or unstructured data. Big data analytics is the process which analyses huge data sets and reveals hidden information, hidden patterns to discover knowledge from the given data.

Considering the current scenario, in developing countries like India, Diabetic Mellitus (DM) has become a very severe disease. Diabetic Mellitus (DM) is classified as Non-Communicable Disease (NCB) and many people are suffering from it. Around 425 million people suffer from diabetes according to 2017 statistics. Approximately 2-5 million patients every year lose their lives due to diabetes. It is said that by 2045 this will rise to 629 million.[1]

Diabetes Mellitus (DM) is classified as-

Type-1known as Insulin-Dependent Diabetes Mellitus (IDDM). Inability of human's body to generate sufficient insulin is the reason behind this type of DM and hence it is required toinject insulin to a patient. Type-2also known as Non-Insulin-Dependent Diabetes Mellitus (NIDDM). This type of Diabetes is seen when body cells are not able to use insulin properly.Type-3Gestational Diabetes, increase in blood sugar level in pregnant woman where diabetes is not detected earlier results in this type of diabetes. DM has long term complications associated with it. Also, there are high risks of various health problems for a diabetic person. A technique called, Predictive Analysis, incorporates a variety of machine learning algorithms, data mining techniques and statistical methods that uses current and past data to find knowledge and predict future events. By applying predictive analysis on healthcare data, significant decisions can be taken and predictions can be made. Predictive analytics can be doneusing machine learning and regression technique. Predictive analytics aims at diagnosing the disease with best possible accuracy, enhancing patient care, optimizing resources along with improving clinical outcomes.

DOI: 10.48175/568

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 8, June 2025
Impact Factor: 7.67

II. LITERATURE SURVEY

The analysis of related work gives results on various healthcare datasets, where analysis and predictions were carried out using various methods and techniques.Various prediction models have been developed and implemented by various researchers using variants of data mining techniques, machine learning algorithms or also combination of these techniques.Dr Saravana Kumar N M, Eswari, Sampath P and LavanyaS (2015) implemented a system using Hadoop and Map Reduce technique for analysis of Diabetic data. This system predicts type of diabetes and also risks associated with it. The system is Hadoop based and is economical for any healthcare organization.[4] Aiswarya Iyer (2015) used classification technique to study hidden patterns in diabetes dataset. Naïve Bayes and Decision Trees were used in this model. Comparison was made for performance of both algorithms and effectiveness of both algorithms was shown asa result.[5] K. Rajesh and V. Sangeetha (2012) used classification technique. They used C4.5 decision tree algorithm to find hidden patterns from the datasetfor classifying efficiently.[8] Humar Kahramanli and Novruz Allahverdi (2008) used Artificial neural network (ANN) in combination with fuzzy logic to predict diabetes.[9] B.M. Patil, R.C. Joshi and Durga Toshniwal (2010) proposed Hybrid Prediction Model which includes Simple K-means clustering algorithm, followed by application of classification algorithm to the result obtained from clustering algorithm. In order to build classifiers C4.5 decision tree algorithm is used.[10] Mani Butwall and Shraddha Kumar (2015) proposed a model using Random Forest Classifier toforecast diabetes behaviour.[7] Nawaz Mohamudally1 and Dost Muhammad (2011) used C4.5 decision tree algorithm, Neural Network, K-means clustering algorithm and Visualization to predict diabetes.[11] Fig 1, represents taxonomy for Machine Learning Algorithms that can be used for diabetes prediction.The task of choosing a machine learning algorithm includes feature matching of the data to be learned based on existing approaches.Taxonomy of machine learning algorithms is discussed below- Machine learning has numerous algorithms which are classified into three categories: Supervised learning, Unsupervised learning, Semi-supervised learning.

III. METHODOLOGY

The used data in this research consists of two data types, the patient data obtained from an interface provided to the user to input personal details like age, sex, weight, height, and level of activity. The food nutrition data was obtained from the Department of Nutrition and Food Science, University of Ghana, and from the MyFitnessPal database [29]. The diet type of the patient is determined from the obtained data, and calorie needs calculated using the Harris-Benedict's equation. The steps followed using the Harris Benedict's equation are summarised below in the flowchart in below Figure. The flowchart diagram encapsulates the actions taken for the Calorie requirements computation and determination.

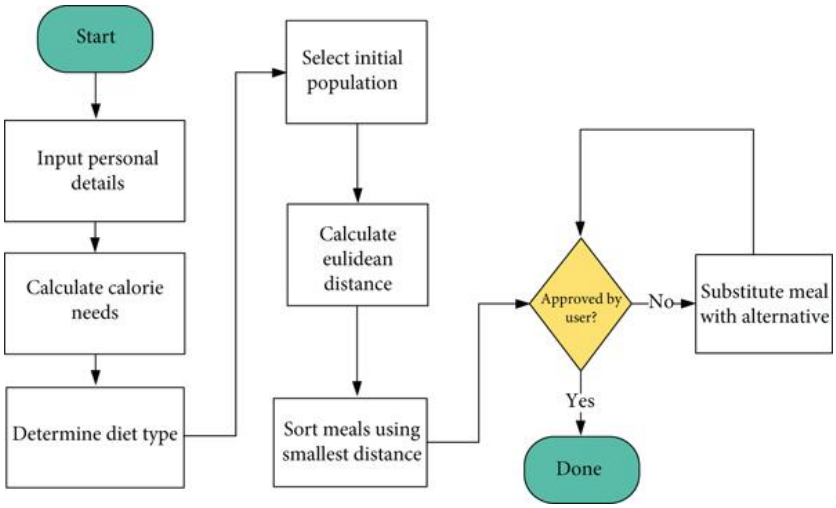

Figure 1: Proposed system model

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 8, June 2025
Impact Factor: 7.67

To save and update the user's information in real-time, Google's firebase was set up and implemented. A firebase project was created in the Firebase console, a setup user authentication and storage rules were generated using a configuration file for both Android application and web application. A JSON file containing meals and the nutrition content was then imported. To extend the reach of the application, there was the need to design a cross-platform application not limited to a particular Operating System. To achieve this, the Ionic framework, which uses AngularJS built on top of Cordova, was used to implement the system. Ionic was used because it has mobile-friendly controls that make cross-platform apps feel like native apps and has a lot of documentation. The navigation logic to enable users to move from one view to the other was implemented and firebase and Google services configuration file added to allow communication with the database. Finally, the various Cordova plugins needed were installed, and the interfaces implemented using HTML, JavaScript, and other Ionic components. To recognize food images, images of different local dishes were collected, labelled, and put into classes. Tensorflow for CPU support was installed, and the Inception model trained with the dataset. To test the classifier, a Python script was written to set up a Flask server that will provide an interface for retrieving classification results. Two graphs were generated after training the model, one containing the retrained model and the other containing the output labels. The trained graphs were imported into a Python function that is called to obtain predictions. Another function was implemented in the Flask server for the meal recommendation. This function loads user's information and meals from the firebase database, performs K-NN to find the best meals for the user, and sends the top best five meals to the database, which syncs data to the front end in real-time. For the question and answer bot, a Microsoft QnA service was created with a knowledge base which extracts question and answers pairs from pdfs with information of diabetes and URLs of diabetes forums.After successfully implementing the various modules, the system was integrated. This was achieved in 2 steps, namely,(1)Making HTTP requests from the front end to the Microsoft model and Flask web service(2)Adding the Firebase configuration file to the front end and the Flask app .

IV. RESULTS

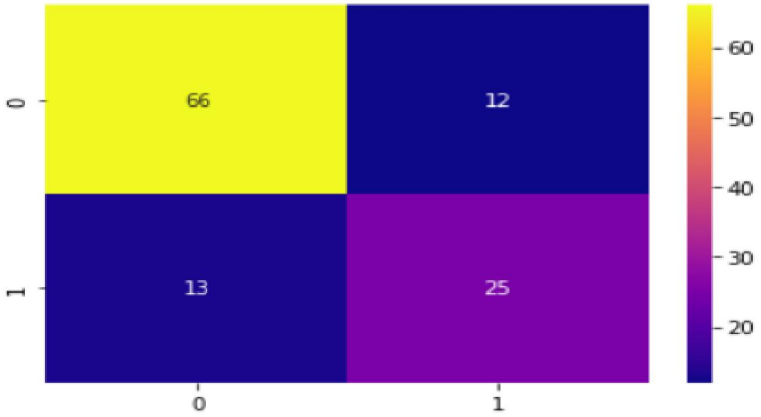

Figure 1: Confusion Matrix

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

Figure 2: Classification Model

DOI: 10.48175/568

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 8, June 2025 Impact Factor: 7.67

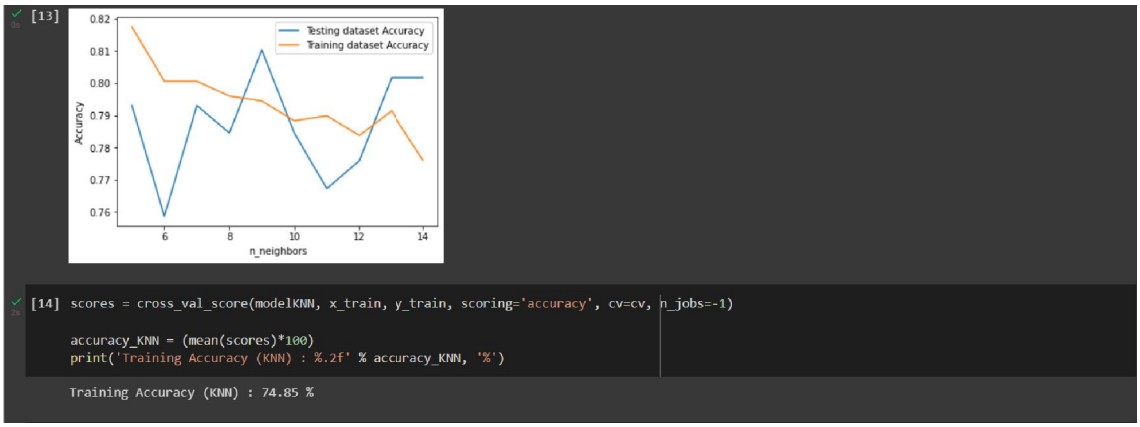

Figure 3: Accuracy Graph

V. CONCLUSION AND FUTURE SCOPE

This research paper has presented a meal recommendation system with food recognition capabilities which focused on generating daily personalized meal plans for the users, according to their nutritional necessities and previous meal preferences. The reviewed literature presented some gaps which informed the design and development of an integrated diabetes management platform for patients using K-Nearest Neighbour (KNN) algorithm, a supervised machine learning model for food recommendation system for diabetics, (2) scheduling and reminding diabetic patients to take their medication and blood glucose readings for doctor's intervention via mobile app, (3) encouraging and tracking the activity of diabetic patients, and (4) providing an interactive visual interface to help them make meaning of their readings and establishing a sufficient connection between the doctor and the diabetic patient using e-mail and chatbots. These integrated technologies present state-of-the-art solutions for the effective management of diabetes. This research paper required us to provide a framework with a user-friendly interface for people with diabetes to monitor their diet, medication, and activity levels. The task has been solved using state of the art algorithms in artificial intelligence.

REFERENCES

[1]. Aljumah, A.A., Ahamad, M.G., Siddiqui, M.K., 2013. Application of data mining: Diabetes health care in young and old patients. Journal of King Saud University - Computer and Information Sciences 25, 127–136. doi:10.1016/j.jksuci.2012.10.003.

[2]. Arora, R., Suman, 2012. Comparative Analysis of Classification Algorithms on Different Datasets using WEKA. International Journal of Computer Applications 54, 21–25. doi:10.5120/8626-2492.

[3]. Bamnote, M.P., G.R., 2014. Design of Classifier for Detection of Diabetes Mellitus Using GeneticProgramming. Advances in Intelligent Systems and Computing 1, 763–770. doi:10.1007/978-3- 319-11933-5.

[4]. Choubey, D.K., Paul, S., Kumar, S., Kumar, S., 2017. Classification of Pima indian diabetes dataset using naive bayes with genetic algorithm as an attribute selection, in: Communication and Computing Systems: Proceedings of the International Conference on Communication and Computing System (ICCCS 2016), pp. 451– 455.

[5]. Dhomse Kanchan B., M.K.M., 2016. Study of Machine Learning Algorithms for Special Disease Prediction using Principal of Component Analysis, in: 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication, IEEE. pp. 5–10.

[6]. C. B. Giorda, V. Manicardi, J. Diago Cabezudo, Theimpact of diabetes mellitus on healthcare costs initaly, Expert review of pharmacoeconomics& outcomes research 11 (6) (2011) 709–719.

[7]. Q. Zou, K. Qu, Y. Ju, H. Tang, Y. Luo, D. Yin, Predictingdiabetes mellitus with machine learning techniques,Frontiers in genetics 9 (2018) 515.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 8, June 2025
Impact Factor: 7.67

[8]. M. A. Sarwar, N. Kamal, W. Hamid, M. A. Shah, Pre-diction of diabetes using machine learning algorithmsin healthcare, in: 2018 24th International Conference on Automation and Computing (ICAC), IEEE, 2018, pp. 1–6.

[9]. D. W. Aha, D. Kibler, M. K. Albert, Instance-basedlearning algorithms, Machine learning 6 (1) (1991) 37–66.

[10]. W. H. Organization, et al., Global report on diabetes.2016 (2017).

[11]. I. H. Sarker, A. Kayes, P. Watters, Effectiveness analysisof machine learning classification models for predictingpersonalized context-aware smartphone usage, Journal of Big Data 6 (1) (2019) 57.

[12]. I. H. Sarker, A. Colman, J. Han, Recencyminer: miningrecency-based personalized behavior from contextual smartphone data, Journal of Big Data 6 (1) (2019) 49.

[13]. I. H. Sarker, Context-aware rule learning from smart-phone data: survey, challenges and future directions,Journal of Big Data 6 (1) (2019) 95.

[14]. I. H. Sarker, A machine learning based robust prediction model for real-life mobile phone data, Internet of Things 5 (2019) 180–193