International Conference on
Global Engineering & Management Trends

International Journal of Science, Engineering and Technology
ISSN: 2348-4098, P-ISSN: 2395-4752

Prediction of Diabetes with Web-Application using Machine Learning Algorithms

M.Tech. Student Abbareddi Anjali, Asst. Prof. R Althaf,Asst. Prof. V Subhasini
Dept of CSE, MJR College of Engineering & Technology,
Piler, AP, India.

Abstract-Manufacturing decisions inherently face uncertainties and imprecision. Fuzzy logic, and tools based on fuzzy logic, allow for the inclusion of uncertainties and imperfect information in decision making models, making them well suited for manufacturing decisions. In this study, we first review the progression in the use of fuzzy tools in tackling different manufacturing issues during the past two decades. We then apply fuzzy linear programming to a less emphasized, but important issue in manufacturing, namely that of product mix prioritization. The proposed algorithm, based on linear programming with fuzzy constraints and integer variables, provides several advantages to existing algorithm as it carries increased ease in understanding, in use, and provides flexibility in its application.

Keywords- Artificial intelligence, fuzzy logic, expert system, decision support system, tool selection; production; Industry 4.0

I.INTRODUCTION

Diabetes is becoming more common and widespread.Currently, there are around 382 million individuals livingwith the condition worldwide, and so by 2035, this numberwould have more than doubled to 592 million. The rise couldbe attributed to suburbanization, the progressive adoption ofpoor lifestyle choices, and the ageing of the populationwithout adequate preparation for prevention and treatment, posing various problems for diabetes care, which seems tohave become a major health problem in the majority ofnations around the world.Diabetes affects approximately 200 million peopleworldwide, with a yearly prevalence of 7%. For a lot longer,people suffered from numerous diseases that, in certain cases, might have been avoided, but owing to a lack of quickdetection of symptoms in patients, this can have disastrous repercussions. Several studies have lately been conducted onthe subject of illness prediction, to the point that some physicians now employ machine learning models for predicting certain diseases. As a result, it is critical to create adiabetes classifier that is both simple and accurate, as well as cost-effective. Research methodologies provide a variety ofideas that are useful to human-related areas of application,such as clinical diagnosis, which is a procedure in which aphysician must analyse a large number of factors in order todiagnose diabetes, making the physician's job difficult andtime-consuming. Deep learning and data gatheringapproaches have been shown to be highly useful inthedevelopment of automatic diagnostic techniques for a variety of medical diseases [1].

II. LITERATURE REVIEW

Recently, a number of studies have been conducted on the subject of illness prediction, to the point that some physicians now employ machine learning models to forecast various diseases. As a result, it's critical to create a diabetes classifier that's simple to use, accurate, and affordable. AIpowered

International Journal of Science, Engineering and Technology
ISSN: 2348-4098, P-ISSN: 2395-4752

techniques provide a variety of opinions that are useful in various fields of application, such as medical diagnosis, which is a process in which a physician must analyse a large number of factors before diagnosing diabetes, making the physician's job difficult and time-consuming. Machine learning and data mining approaches have been shown to be extremely useful in the development of automated diagnosis systems for a variety of medical disorders.Using some supervised learning algorithms and a case study of chosen hospitals within the Kaduna metropolitan, this research effort aims to construct a prediction model which is highly accurate for diabetes in individuals at an early stage, before it escalates to a point of morbidity or fatality.

This study will also benefit the health sector by providing patients with correct prior knowledge about their health state as it relates to diabetes, lowering the risk of complications, morbidity, and death associated with this type of disease. Gaining information and real-time insights from extensive, high-dimensional, and diverse biological data continues to be a major problem in health-care transformation. Conventional data mining and statistical learning methods are required to test on feature engineering to gain some effective and more extensive features from those data, and then create prediction or clustering data without having enough knowledge related to that field. However, the current advancement in deep learning technologies provides more effective dimensions to gain endto-end learning methods from a vast and complex design of data that can be analysed quite accurately by deep learning techniques and neural networks more frequently in coming days by experts in all types of fields.Related work analysis provides results in a variety of health care repositories, where analysis and forecasting are performed using a variety of methods and procedures. Various researchers have created and used speculative models based on data mining techniques, machine learning algorithms, or a mix of these methods. Dr. Saravana Kumar N M, Eswari, Sampath P, and Lavanya S (2015) analysed diabetes data using a Hadoop and MapReduce system. This algorithm can forecast the kind of diabetes and the dangers that come with it. This Hadoop-based application is beneficial to any health-care firm. Aiswarya Iyer (2015) investigated hidden patterns on a diabetic site using a different approach.

There have only been a few machine learning approaches that have been frequently used in healthcare and medicalrelated systems to provide results based on tested data. Many researchers and data scientists have tested as well as used them for the treatment of diabetes-related problems all over the globe. The Random Forest technique in ML and for this paper's result provided the maximum accuracy amongst the other ML algorithms. The paper's authors implemented the DT (Decision Tree) method to find people at a lower risk of type-2 diabetes for the Tehran Lipid and Glucose Study (TLGS) database.

III. ABOUT LIBRARIES

While developing this diabetes predicting web application made use of some fundamental libraries that aided by theproject and helped me make the predictive model, they wereas follows:

1. **Streamlit:** This is an open-source Python library that makes it convenient to create and deploy powerful custom-made web applications for machine learning and data science projects.
2. **Pandas:** It is a fast and efficient tool to carry out data analysis and data manipulation. In addition to that, it allows us to import data from varied file formats such as JSON, CSV's, SQL, and Microsoft EXCEL, enabling us to perform complex data cleaning and wrangling operations on these files to read, interpret and gather insights from raw data.
3. **Numpy:** It is an array processing package that allows us to perform several scientific computations on arrays, including several high-level mathematical functions.

International Journal of Science, Engineering and Technology
ISSN: 2348-4098, P-ISSN: 2395-4752

4. **Sklearn:** This is the fundamental machine learning library in Python, It contains a wide variety of modules to aid machine learning and statistical modeling processes such as classification, regression, clustering, and dimensionality reduction.
5. **PIL:** Python imaging library, is an open-source additional library that provides features that allow opening, manipulation, and saving many different image file formats.

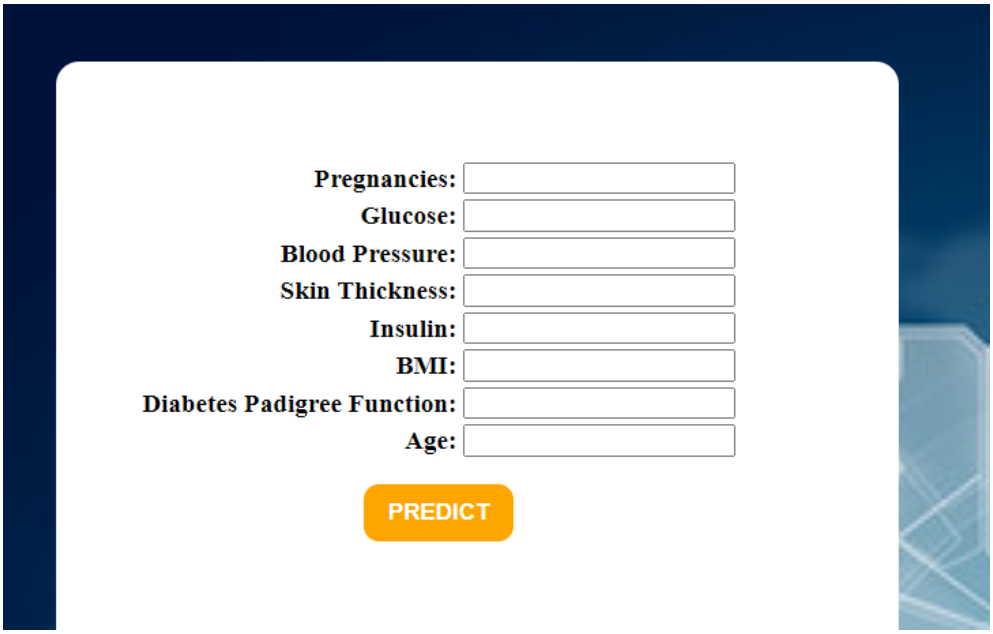

Fig. 1. Diabetes Prediction Classifier Web-Application Interface

ALGORITHMS

K-Nearest Neighbours (KNN)- KNN is a supervised machine learning algorithm, it assumes the similarity between the new data points and available data points and puts the new datapoint into the category that is most similar to the available set of categories. It is a non-parametric algorithm, in other words, it will not make any assumption on the underlying data.

Support Vector Machine(SVM): SVM is also a supervised machine learning algorithm, that can be used for classification and regression tasks. Here, we plot each data point in an n-dimensional space, and here 'n' is the number of features in our dataset. Basically, in SVM the algorithm will create a line or hyperplane due to which the data points get split into two distinct classes.

Random Forest: Random Forest is another supervised learning algorithm, where the forest is said to be an ensemble of decision trees. In simple terms, the random forest classifier will divide the dataset into different subsets and these subsets are fed into every tree of the random forest algorithm. Every decision tree will produce its specific output.

IV. METHODOLOGY

The chance of diabetes in a human can be predicted significantly more efficiently and correctly using cuttingedge technology. Diabetes will be predicted using machine learning. The dataset consists of around 1000 data entries (rows) and 9 attributes (columns). Our system will be able to predict people with and without diabetes using multivariate regression. In order to combine features so that our machine can predict quickly, we will use a total of four algorithms: SVM, RF, Nave Bayes, and Decision Tree. 75 percent of the data will be used to train our machine, while the remaining 25 percent will be used for testing purposes, to ensure that our computer accurately predicts the type of disease. All of

International Journal of Science, Engineering and Technology
ISSN: 2348-4098, P-ISSN: 2395-4752

our algorithms will be written in Python, and machine learning libraries such as Pandas, Numpy, Scikit-learn, Matplotlib.pyplot, TensorFlow, Seaborn, and others will be used.

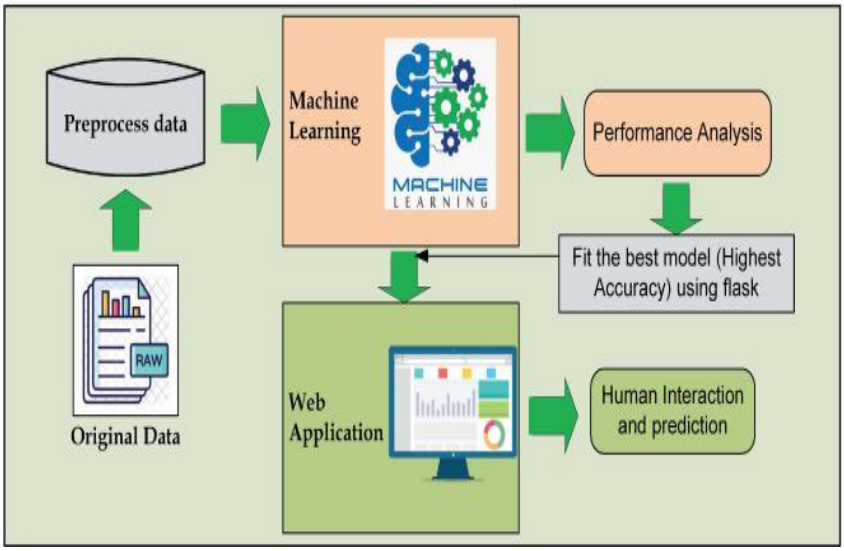

Fig. 2. Block Diagram for proposed model

We use the Flask web application to run and build the online web application. First, we load the machine learning model that we trained on a complete dataset, then we retrieve user input into an HTML image, make a prediction, and then return the result using a web application python (API) code. In addition, an advanced HTML template that allows the user to enter the patient's heart symptoms and determine if the patient has diabetes.

V. RESULTS AND INFERNECES

The suggested methodology will assist users in determining their risk of diabetes early and in acquiring future forecasts of their BG rise levels. The proposed methods are tested using the PIMA Indian Diabetes dataset. This system uses various models to solve the diabetes classification issue. This was proved by the SVM model's remarkable classification accuracy of 78.65 percent, the Decision Tree model's classification accuracy of 95 percent, the Random Forest model's classification accuracy of 96.6 percent, and the Nave Bias model's classification accuracy of 77.5 percent on the dataset. As a consequence, the Random Forest model produced the most accurate results for diabetes prediction.The proposed structure for the diagnosis and monitoring of diabetes in real time. It is important to consider processing sensor readings near the location where the data is received, such as on a smartphone, within a system design. The delay problem may be handled by inserting sensors near the point where the data is transmitted and received, such as on a smart phone. Maximum accuracy is achieved by the Random Forest algorithm for the Pima dataset.

International Journal of Science, Engineering and Technology
ISSN: 2348-4098, P-ISSN: 2395-4752

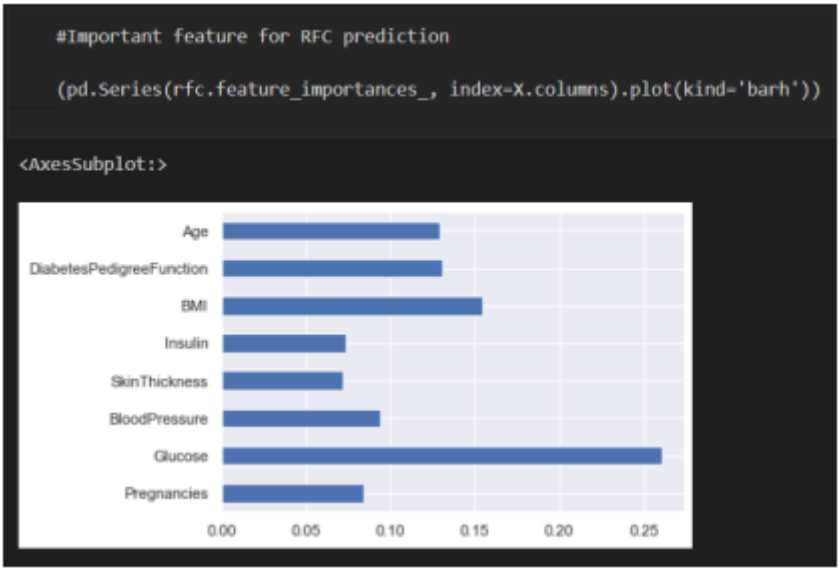

Fig. 3. RFC Prediction

VI. CONCLUSION

One of the major issues in the health-care business is the early identification of diabetes. In our research, we developed a technology that can accurately predict diabetes. Using the VS Code notebook program, we preprocessed the data.In this research, we applied several machine learning techniques to enhance diabetes prediction performance using available datasets. Data preparation techniques like correlation and polynomial regression were included in the framework to boost their performance. This approach works for the SVM Model, Decision Tree Model, Random Forest Model, and Naive Bias Model, and it does a good job of classifying diabetes. The SVM model had 77.25 percent classification accuracy on the dataset, the Decision Tree model had 95 percent classification accuracy, the Random Forest model had 96.6 percent classification accuracy, and the Nave Bias model had 76.75 percent classification accuracy. Therefore, it can be stated that the Random Forest model presented the most accurate results for diabetes prediction. The efficiency of the model is used to assess its functionality. What we mean by accuracy is the ability to determine which situations are positive and which are negative. Whenever a model has great accuracy, it is said to have low error.

REFERENCES

[1] A. E. Evwiekpaefe and S. M. Umar, "Predicting road traffic crash severity in Kaduna Metropolis using some selected Machine Learning Techniques," Nigerian Journal of Technology, vol. 40, no. 5, pp. 888– 900, 2022.

[2] G. Li, S. Peng, C. Wang, J. Niu, and Y. Yuan, "An energy-efficient data collection scheme using denoising autoencoder in wireless sensor networks," Tsinghua Science and Technology, vol. 24, no. 1, pp. 86– 96, 2019.

[3] D. Power, "Standards of medical care in diabetes," Diabetes Care, vol. 29, no. 2, pp. 476–476, 2006.

[4] H. Liu, H. Kou, C. Yan, and L. Qi, "Link prediction in Paper Citation Network to construct paper correlation graph," EURASIP Journal on Wireless Communications and Networking, vol. 2019, no. 1, 2019.

[5] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep Learning for Healthcare: Review, opportunities and challenges," Briefings in Bioinformatics, vol. 19, no. 6, pp. 1236–1246, 2017.

International Journal of Science, Engineering and Technology
ISSN: 2348-4098, P-ISSN: 2395-4752

[6] D. Fagella, 7, "Applications of Machine Learning in Pharma and Medicine",[online document 2017, Available: https://goo.gl/IS1R5k.
[7] G. D. Magoulas and A. Prentza, "Machine learning in medical applications," Machine Learning and Its Applications, pp. 300–307, 2001.
[8] M. Fatima and M. Pasha, "Survey of machine learning algorithms for disease diagnostic," Journal of Intelligent Learning Systems and Applications, vol. 09, no. 01, pp. 1–16, 2017.
[9] R. Zolfaghari, "Diagnosis of diabetes in female population of pimaindian heritage with ensemble of bp neural network and svm." International Journal of Computational Engineering & Management, vol. 15, no. 4, pp. 2230-7893, 2012
[10] N. Yuvaraj and K. R. SriPreethaa, "Diabetes prediction in healthcare systems using machine learning algorithms on Hadoop Cluster," Cluster Computing, vol. 22, no. S1, pp. 1–9, 2017.
[11] R. Ahuja, S. C. Sharma, and M. Ali, "A diabetic disease prediction model based on classification algorithms," Annals of Emerging Technologies in Computing, vol. 3, no. 3, pp. 44–52, 2019.