ISSN No:-2456-2165

Prediction of Diabetes using Machine Learning: A Modern User-Friendly Model

S. Umar Kalimulla[1], V. AlekyaSatyasri[2], K. Srunvitha[3], S. H. N. V. V. D. S. Sai Charan[4], A. V Satya Sai Ram[5,] DR. V. Venkateswara Rao[6]
[1,2,3,4,5]Department of Computer Science and Artificial Intelligence (CAI),
[6]Professor, Department of Computer Science and Engineering (CSE),
Sri Vasavi Engineering College, Tadepalligudem, Andhra Pradesh, India

Abstract:- Diabetes is a prevalent chronic disease affecting a significant portion of the global population. Early detection and accurate prediction of diabetes can play a crucial role in managing the condition and preventing complications. Machine learning (ML) techniques have shown promising results in diabetes prediction based on patient data. In this study, we propose a user-understandable approach utilizing the Random Forest classifier algorithm for accurate and interpretable diabetes prediction. To build our prediction model, we utilized a comprehensive dataset comprising various patient attributes, including age, body mass index (BMI), blood pressure, glucose levels, and medical history. Pre-processing techniques were applied to handle missing values and normalize the data, followed by feature selection to identify the most relevant attributes for diabetes prediction. The user-understandable representation of the model facilitated effective interpretation and communication of the prediction results. This allows healthcare professionals to explain the prediction rationale to patients, promoting shared decision-making and patient engagement.

I. INTRODUCTION

In an era where healthcare is increasingly intertwined with advanced technology, our project, which examines a crucial area of public health called THE PREDICTION OF DIABETES, focuses on a time when healthcare is becoming more and more entwined with cutting-edge technology. Early identification is essential for optimal care of diabetes, a chronic metabolic condition that affects millions of people worldwide. Our study aims to create a prediction model that can identify people at risk of diabetes before clinical signs appear by leveraging the power of machine learning. This project aims to contribute to proactive healthcare initiatives and enable people to make well-informed decisions about their future health by utilizing a wide range of health-related diseases.

II. LITERATURE SURVEY

They use CNN, K-NN, Choice Tree, and NB. They contrasted with a few commonplace expectations There are four pertinent papers that investigate diabetes prediction.

N.A. Farooqui, Ritika, and A. Tyagi proposed an examination article," Expectation Model for Diabetes Mellitus Utilizing AI Procedures". They distributed the Global Diary of Science and Medical Services. They researched the article in 2020. The writers utilized Choice Tree and k-Closest Neighbours to fabricate their model. They utilize different AI methods, viz., choice trees, k-closest neighbours, arbitrary woodland, and backing vector machines, and foresee the exhibition of various order procedures.

Dr. Mohammed Abdul Raheem, Shaik Ehetesham, Mohammad Faiz Ahmed Subhani, and Sayed Abdul Zakir proposed an examination article" Man-made Intelligence Calculation Framework for Expectations of Diabetes Utilizing Moderate Web Add IBM Cloud". They distributed the Global Diary of Science and Medical Services. They researched the article in 2020. They use AI, IBM Cloud, man-made brainpower calculations, and counterfeit brain organizations. Recognizing diabetes in its early stages is vital. Although the precision accomplished by these AI models is high, there are not many impediments to this task.

RinkalKeniya, Aman Khakharia, Vruddhi Shah, VrushabhGada, RachiManjalkar, Tirth Thaker, Mahesh Warang, and NinadMehendale proposed an examination article "Infection expectation from different side effects utilizing AI". They distributed SSRN in the year 2020. They utilized their closest neighbors. The weighted k-closest neighbour model gave the highest exactness of 93.5% for the expectation of illnesses utilizing the side effects.

Min Chen, Yixue Hao, Kai Hwang, Individual, IEEE, Lu Wang, and Lin Wang proposed an exploration article "Sickness Expectation by AI over Enormous Information from Medical Care Networks". They distributed IEEE in 2017. calculations. The expected precision of our proposed calculation.

III. PROBLEM STATEMENT IN EXISTING SYSTEM

In the day-to-day routine of an everyday person, the forecast of a persistent illness like diabetes at the beginning phase is exceptionally urgent. Expectations for these infections can be accomplished accurately by utilizing AI (ML) models with high precision. The current application consumes a greater amount of human time for determining and showing important information from the model. A client-reasonable model that is less tedious can be acquainted with getting client information anticipated in a client-friendly, efficient, and justifiable way.

ISSN No:-2456-2165

IV. DISADVANTAGES IN THE EXISTING SYSTEM

- **Data Privacy and Security:** Data privacy and security are major concerns in the healthcare industry when handling sensitive patient data. A significant challenge is maintaining patient data security and privacy while prediction models are being developed and used.
- **Limited Accuracy:** Many of the models now in use may not be very accurate at predicting diabetes, which could result in false positives and false negatives.
- Consumes a greater amount of human time.
- Not effective and user-friendly models.

V. PROPOSED SYSTEM

- The patient's clinical information is taken as a necessity for the model. An immense record of information connected with diabetes patients is utilized to prepare the model. The model Random Forest Classifier is prepared with the dataset connected to diabetic patients.
- The model is then made to anticipate the result as per the patient's information. Finally, a WhatsApp message is sent to the given portable number that says a client is diabetic or not. So, the proposed model has the following advantages in order to overcome the problem in the existing system:
- Early Recognition: AI models can break down understanding information, for example, clinical history, lab results, and way of life factors, to anticipate the probability of creating diabetes later on. Early identification permits medical care suppliers to immediately mediate and start preventive measures.
- Risk Definition: ML calculations can separate patients into various risk classifications in light of their probability of developing diabetes. This empowers medical services suppliers to assign assets and mediations all the more effectively, zeroing in on people at higher risk.
- Customized Medication: AI can assist with fitting treatment plans for diabetic patients. By breaking down information from a patient's wellbeing records, the model can suggest individualized treatment choices and lifestyle changes.
- Remote Checking: ML-controlled wearable gadgets and portable applications can persistently screen blood glucose levels and other significant information, giving ongoing bits of knowledge to patients and their medical service suppliers. This can prompt better illness for executives and convenient intercessions.
- Interpretability: While Random Forest models are not quite as interpretable as easier models like straight relapse, they can give some degree of interpretability through highlighted significance scores. This can assist clinicians and scientists with understanding which elements are driving the model's expectations.

VI. DATASET DESCRIPTION

Information collection from the Kaggle-acquired dataset is the underlying period of execution. For our review, a dataset associated with diabetes is required. This dataset incorporates names for the accompanying terms: pregnancy, glucose, blood pressure, skin thickness, insulin, BMI, and age. Given the large number of features provided in the dataset, classes are classified based on them.

Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	Age	Outcome
6	148	72	35	0	33.6	50	1
1	85	66	29	0	26.6	31	0
8	183	64	0	0	23.3	32	1
1	89	66	23	94	28.1	21	0
0	137	40	35	168	43.1	33	1
5	116	74	0	0	25.6	30	0
3	78	50	32	88	31	26	1
10	115	0	0	0	35.3	29	0
2	197	70	45	543	30.5	53	1
8	125	96	0	0	0	54	1
4	110	92	0	0	37.6	30	0
10	168	74	0	0	38	34	1
10	139	80	0	0	27.1	57	0
1	189	60	23	846	30.1	59	1

Fig. 1: Dataset Description

VII. METHODOLOGY

A decision tree is an ML algorithm and a visual representation of the decision-making process. Decision trees are built by recursively dividing a data set into subsets based on the values of input features. The main goal of a decision tree is to divide data into groups that are homogeneous in terms of the target variable, making it easier to make predictions or decisions.

ISSN No:-2456-2165

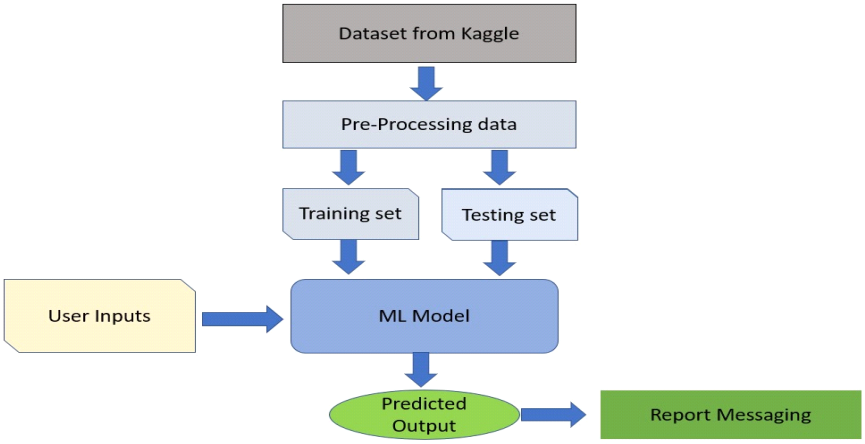

Fig. 2: Methodology of the proposed model

- **Data Collection:** Begin the process by choosing a reasonable diabetes dataset from a Google site that has significant features for prediction.
- **Pre-Processing Data:** Inspect the data to understand its structure, missing values, and data types.
- ✓ Handle missing data by imputing or removing incomplete rows.
- ✓ Encode categorical variables using techniques like one-hot encoding.
- ✓ Scale or normalize numerical features to have the same scale.
- **Feature Selection:** Determine which features are most relevant for predicting diabetes and remove the other irrelevant features.
- **Training and testing data:** Split the pre-processed data into testing sets and testing sets. In general, the training set is 80% and the testing set is 20%. Train the model and test the model to evaluate its performance using random forest.
- **Deployment:** The model can be deployed after being trained and evaluated successfully.

User inputs are given to the model that predicts outputs by constructing decision trees. At last, a WhatsApp reporting message is sent to the user that describes whether the user is diabetic or not.

VIII. SYSTEM ARCHITECTURE

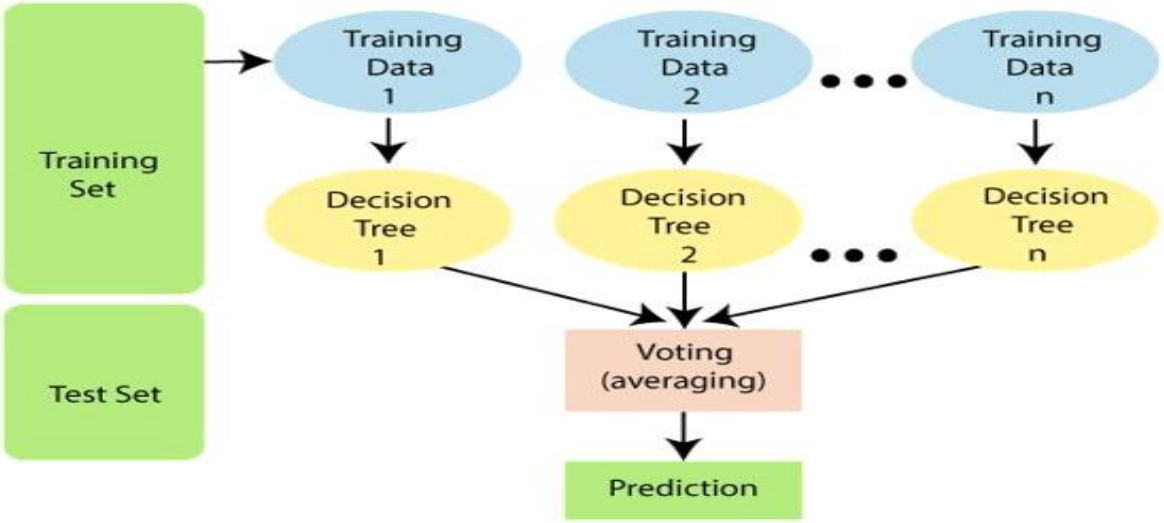

Fig. 3: System Architecture

- A decision tree is a straightforward and interpretable ML model utilized for both classification and regression.
- It is a variously levelled tree-like design that settles on choices by recursively dividing the information into subsets in view of the values of input features.
- The construction of the tree comprises nodes, branches, and leaves. Nodes address choices, branches address the potential results of decisions, and leaves address the last predictions.
- Decision trees make parts in view of component values to make nodes. The objective is to make splits that best separate the data into classes for classification. The splitting process continues until stopping criteria are met. For classification, a leaf node typically represents a class label. The final prediction in a random forest is obtained by aggregating the predictions of all individual decision trees; this can be done by majority voting.

ISSN No:-2456-2165

IX. EXPERIMENTAL RESULTS

The dataset is trained utilizing a random forest classifier, and that dataset should be liberated from commotion, which implies it shouldn't contain any irrelevant information or invalid qualities. Client data sources, for example, pregnancy, glucose, circulatory strain, skin thickness, insulin, BMI, and age, are given to the model. In the wake of giving the client inputs, they are compared with other information. The following figure shows the prepared information and the given client inputs.

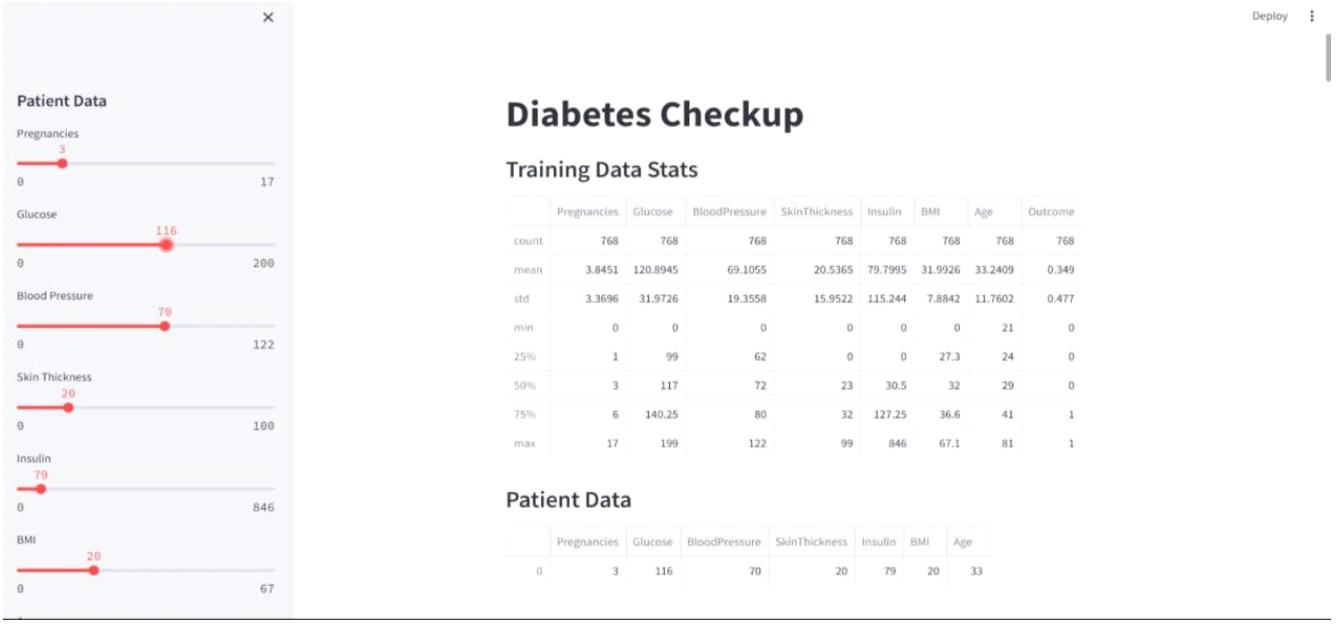

Fig. 4: Interface of the proposed model

A visualized patient report is made that addresses comparisons between the given client source information and other information. This perception report says a client is healthy or unhealthy. Here, '0' indicates a client is healthy, and '1' indicates a client is unhealthy. The accompanying figure shows a pregnancy count chart in which the x-axis addresses age and the y-axis addresses pregnancies.

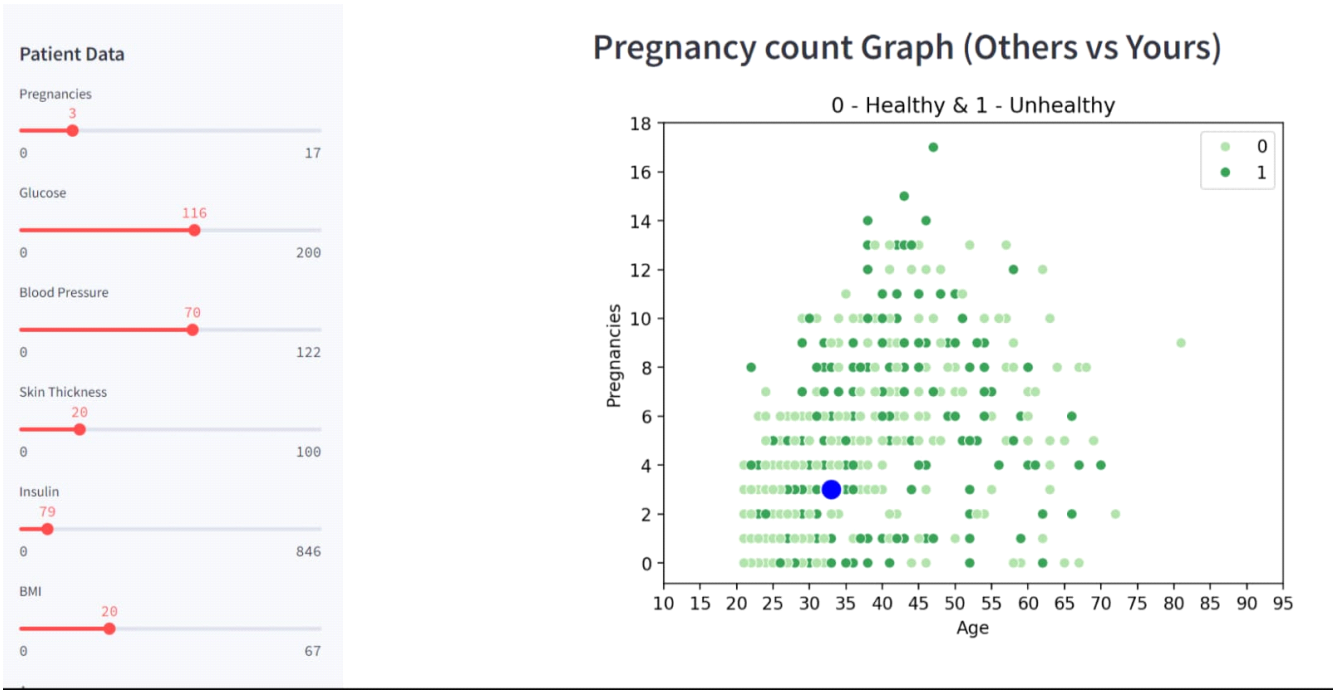

Fig. 5: Visualization of Pregnancies

ISSN No:-2456-2165

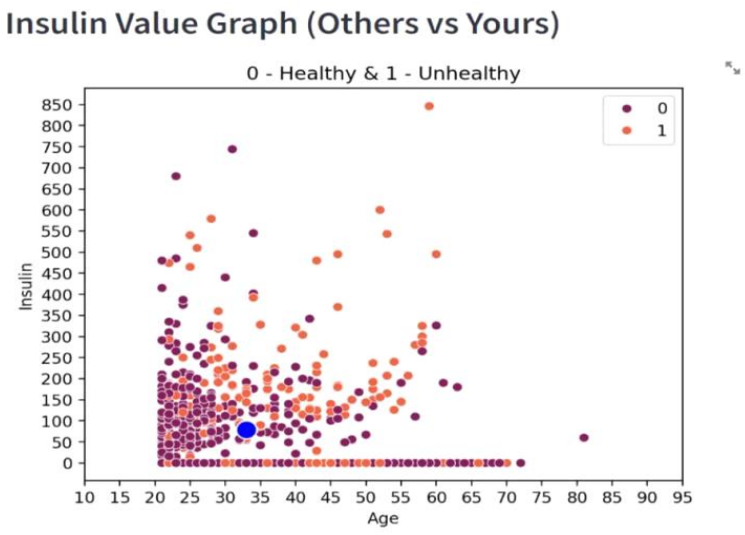

Fig. 6: Visualization of Insulin

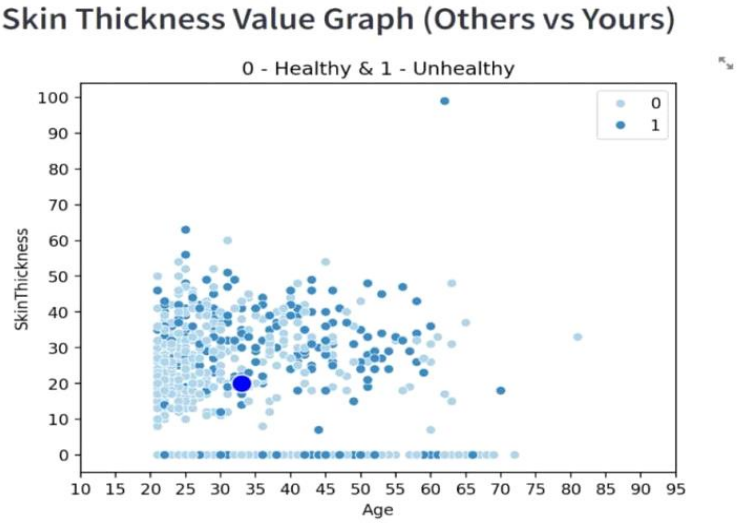

Fig. 7: Visualization of Skin Thickness

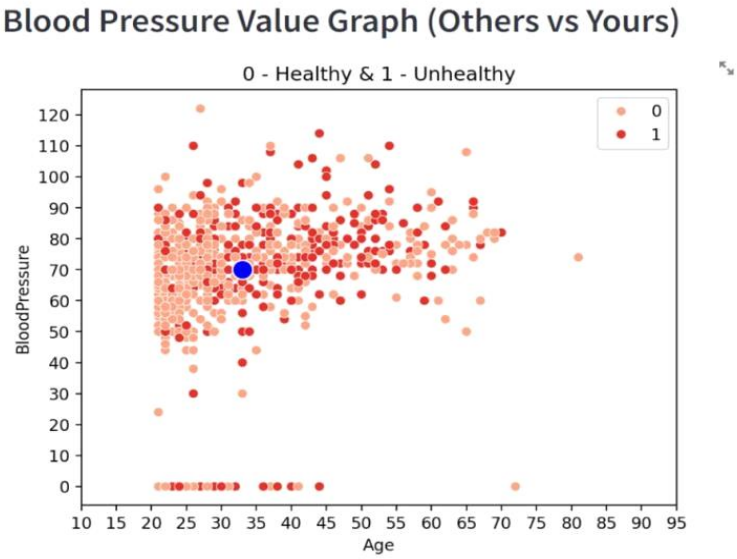

Fig. 8: Visualization of Blood Pressure

ISSN No:-2456-2165

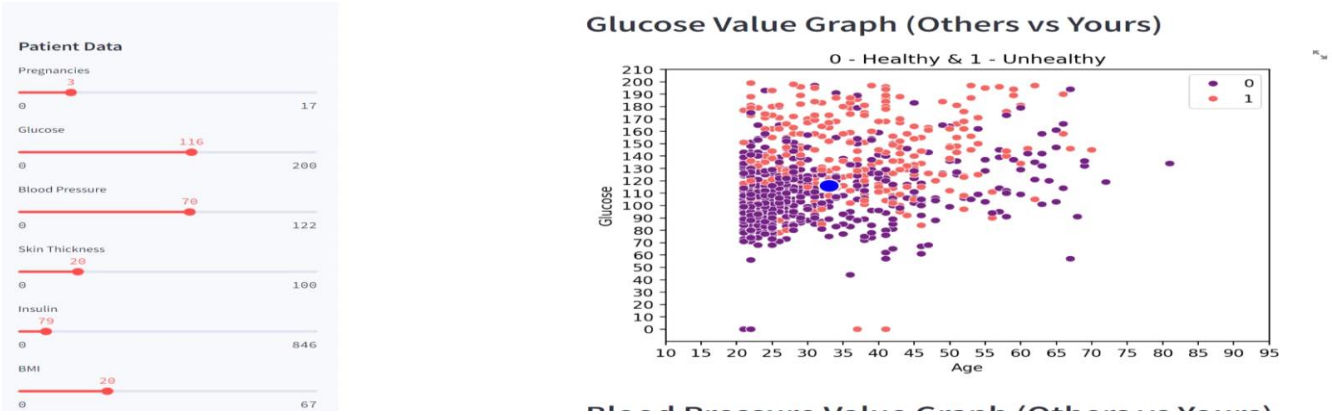

Fig. 9: Visualization of Glucose

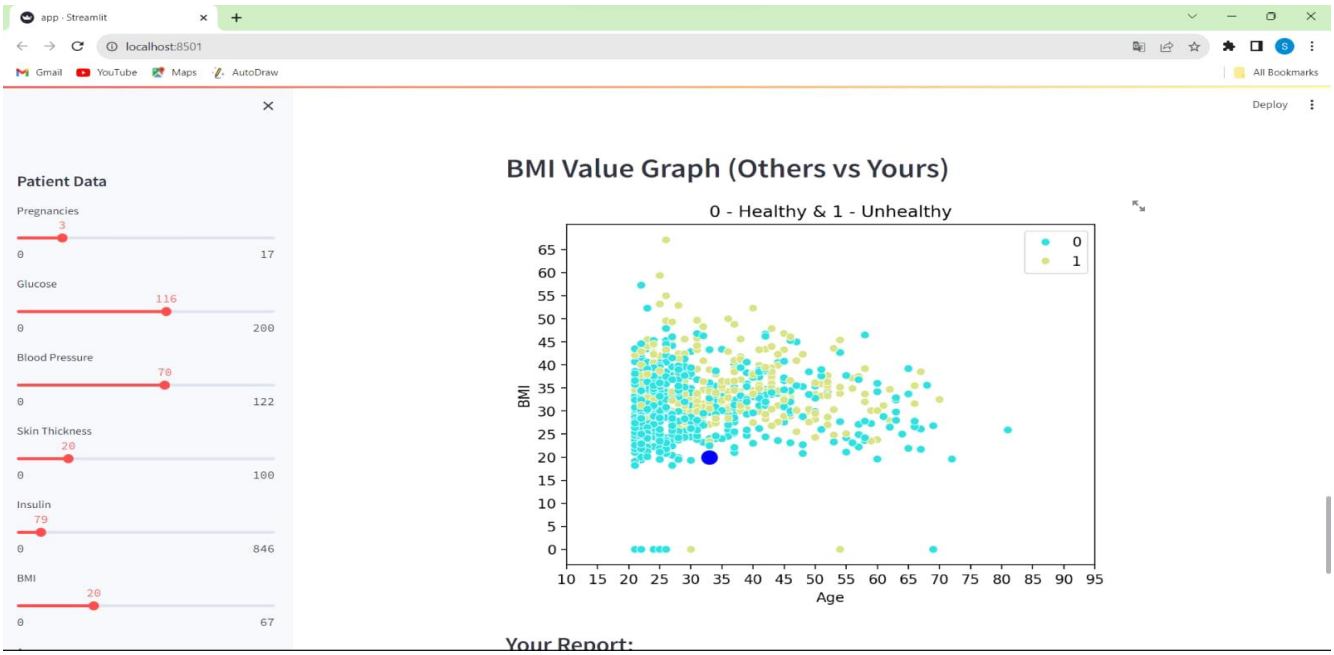

Fig. 10: Visualization of BMI

Lastly, based on the reports obtained, it is shown that a client is diabetic or not. To get the report, the client name and portable number are given. A WhatsApp message is sent to the given versatile number, which shows messages such as hi client, you are not diabetic if healthy and hi client, you are diabetic if unhealthy. The accompanying figure shows a WhatsApp message report.

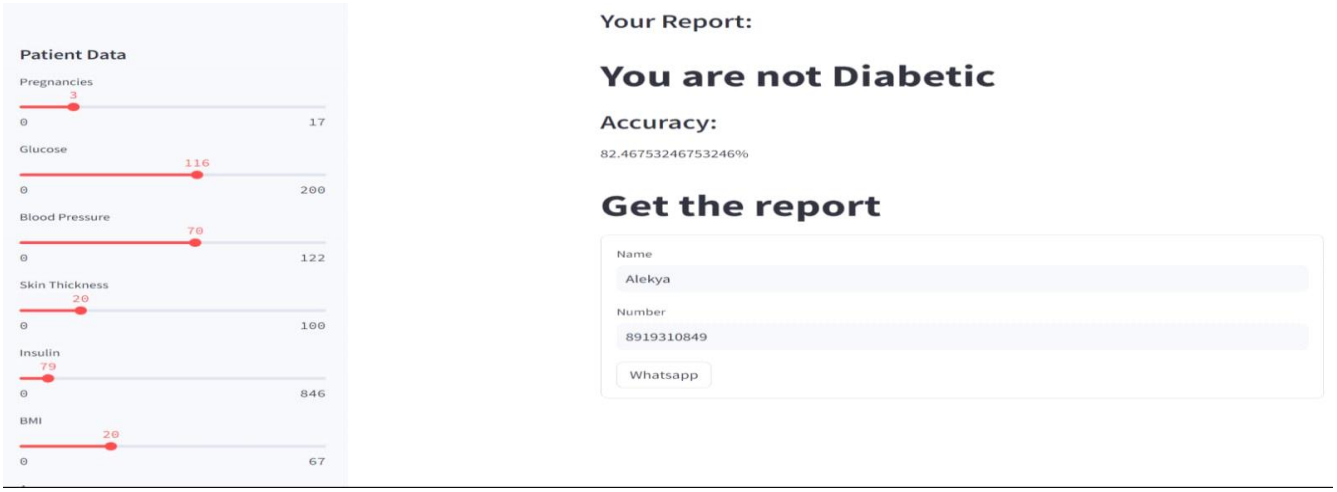

Fig. 11: Final output and Messaging feature

ISSN No:-2456-2165

X. CONCLUSION

All in all, the execution of the present-day diabetic forecast model shows a more prominent effect on humans' existence. In basic words, client inputs are contrasted with different information, which envisions a report with additional accuracy, and a WhatsApp message is sent saying a client is diabetic or not.In this quick world, foreseeing diabetes at the beginning is exceptionally pivotal. The current models are not less tedious, easy to use, and client-justifiable. So, our proposed model beats these impediments, i.e., it is less tedious, easy to understand, and gives highly precise outcomes.

REFERENCES

[1]. Parvin Soleimani, Prediction of Diabetes, Ryerson University Computer Science Department, Canada, 2020.
[2]. Al Juma, A.L., Ahmad, M.G., and 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.
[3]. Mukesh Kumari et al. (IJCSIT), International Journal of Computer Science and Information Technologies, Vol. 5(4), 2014, 5174-5178, “Prediction of Diabetes Using Bayesian Networks.".
[4]. G. D. Kalyankar, S. R. Poojara, and N. V. Dharwadkar, "Predictive analysis of diabetic patient data using machine learning and Hadoop," 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics, and Cloud) (I-SMAC), Palladam, 2017, pp. 619–624, doi: 10.1109/I-SMAC.2017.8058253.