

DAYANANDA SAGAR COLLEGE OF ENGINEERING

(An Autonomous Institute affiliated to VTU, Belagavi, Approved by AICTE & ISO 9001:2008 Certified)

Accredited by National Assessment & Accreditation Council (NAAC) with 'A' grade,

Shavige Malleshwara Hills, Kumaraswamy Layout, Bengaluru-111



Mini Project Report

on

“PREDICTIVE ANALYSIS OF DIABETES”

Submitted by

P R Sai Rahul (1DS19CS107)

Jaishree S (1DS19CS134)

Sanjeev F Annigeri (1DS19CS142)

Shreya Sri A N (1DS19CS153)

Sixth Semester B.E (CSE)

Mini Project

19CS6DCMIP

Under the guidance of

Chaitra S P

Professor

Dept. of CSE

DSCE, Bangalore

Department of Computer Science and Engineering

Dayananda Sagar College of Engineering

Bangalore-111

VISVESVARAYA TECHNOLOGICAL UNIVERSITY

DAYANANDA SAGAR COLLEGE OF ENGINEERING

Shavige Malleshwara Hills, Kumaraswamy Layout, Bangalore - 560078

Department of Computer Science & Engineering



CERTIFICATE

This is to certify that the project entitled **Predictive Analysis of Diabetes in India** is a bonafide work carried out by **P R Sai Rahul (1DS19CS107), Jaishree S (1DS19CS134), Sanjeev F Annigeri (1DS19CS142), Shreya Sri A N (1DS19CS153)** in partial fulfilment of 6th semester, Bachelor of Engineering in Computer Science and Engineering under Visvesvaraya Technological University, Belgaum during the year 2021-22.

Guide Name

Chaitra S P

Professor

Department of CSE,

DSCE

Signature.....

Dr.Ramesh Babu

Vice principal & Head

Department of CSE,

DSCE

Signature.....

Dr.CPS Prakash

Principal,

DSCE

Signature.....

Name of the Examiners:

1.

2.

Signature with date:

.....

.....

ACKNOWLEDGEMENT

We are pleased to have successfully completed the Mini project “**Predictive Analysis of Diabetes in India**”. We thoroughly enjoyed the process of working on this project and gained a lot of knowledge doing so.

We would like to take this opportunity to express our gratitude to **Dr. C P S Prakash**, Principal of DSCE, for permitting us to utilize all the necessary facilities of the institution. We also thank our respected Vice Principal, HOD of Computer Science & Engineering, DSCE, Bangalore, **Dr. Ramesh Babu D R**, for his support and encouragement throughout the process.

We are immensely grateful to our respected and learned guide, **Prof. Chaitra S P**, Designation CSE, DSCE for her valuable help and guidance. We are indebted to them for their invaluable guidance throughout the process and their useful inputs at all stages of the process.

We also thank all the faculty and support staff of Department of Computer Science, DSCE. Without their support over the years, this work would not have been possible.

Lastly, we would like to express our deep appreciation towards our classmates and our family for providing us with constant moral support and encouragement. They have stood by us in the most difficult of times.

P R Sai Rahul (1DS19CS107)

Jaishree S (1DS19CS134)

Sanjeev F Annigeri (1DS19CS142)

Shreya Sri A N (1DS19S153)

CONTENTS

SL. NO	CONTENT	PG. NO
1.	Abstract	4
2.	Introduction	5
3.	Literature Survey	7
4.	System design & Methodology	9
5.	Snapshots and Results	10
6.	Conclusion & Future Enhancements	13
7.	References	13
8.	Appendix	14

ABSTRACT

India is a fast-growing economy with a considerable number of diabetes patients. Age, obesity, lack of exercise, high blood pressure, bad diet, hereditary diabetes, etc can cause **Diabetes Mellitus**. This can lead to high risk of heart diseases, eye problem, nerve damage, kidney diseases, etc.

The practice which is followed currently in hospitals include collecting required information through various tests and providing appropriate treatment based on diagnosis.

Healthcare deals with a lot of data and big data analytics helps in finding the insights and hidden patterns in the data to make meaningful predictions of the outcomes.

A few external factors responsible for diabetes along with the regular factors include: Glucose, BMI, Age, Insulin, etc. The dataset we will be using for the project is “**Pima-Indians-diabetes**”

A **decision tree model** is used to train the dataset and make predictions. The diabetes prediction model for better classification of diabetes which includes few external factors responsible for diabetes along with regular factors like Glucose, BMI, Age, Insulin, etc.

A full stack platform is to be developed to collect, edit and visualize the patient’s data and provide awareness to them using facts and articles. The Machine Learning model is being integrated with the full stack model using *Django framework* to predict diabetes.

A *mobile application* is developed using *Flutter* and *Dart* which globalizes the impact of Diabetes. The ML model is integrated with the app using *flask* and *pythonanyhwere*.

Chapter 1

INTRODUCTION

Healthcare deals with a lot of data and big data analytics helps in finding the insights and hidden patterns in the data to make meaningful predictions of the outcomes. Diabetes is a lifelong condition wherein a person's glucose (sugar) levels in blood becomes very high. 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.

Various traditional methods, based on physical and chemical tests, are available for diagnosing diabetes. However, early prediction of diabetes is quite challenging task for medical practitioners due to complex interdependence on various factors as diabetes affects human organs such as kidney, eye, heart, nerves, foot etc.

The main types of diabetes include:

Type 1 Diabetes: This condition occurs when the body doesn't produce enough insulin. It is commonly seen in people at young age due to malfunctioning immune system.

Type 2 Diabetes: This condition occurs when the body is unable to use the insulin produced. It can occur at any age, even children yet commonly seen in middle-aged and older people due to obesity and sedentary lifestyle.

Gestational Diabetes: It is a temporary condition seen in pregnant women.

Healthcare industries have a massive number of databases consisting of different types of data such as structured, semi-structured, or unstructured. According to the healthcare sector, ER data sets are too huge. Big data analytics can process and analyse the large volume of data sets and discover hidden patterns and information and complex to be processed by traditional techniques. The application of predictive analytics in the healthcare sector has received a great amount of interest in the research community.

Prediction and detection of disease are made by adopting many predictive, quantitative, and statistical models. In recent times, diabetes becomes one of the leading causes of death in developing countries. Mainstream research has been funded to enhance analysis in the following fields and is driven by the emotive to find quick solutions. Diabetes is one of the most prevalent diseases that develops as a result of a high amount of blood glucose or blood sugar in the bloodstream. The glucose in the blood is the most important energy source for the human body, providing it with the energy it needs to complete the full task. This energy is derived through insulin, which is produced with the assistance of the pancreas, which

obtains energy from the consumption of food. As soon as a patient is diagnosed with diabetes, the glucose is unable to reach any cells in the body, which has an impact on the whole body's functioning. According to the findings of the study, 30.2 million individuals in the United States are suffering from diabetes. This diabetes contributes to the development of further diseases such as heart disease, stroke, and other health issues. In this section, we will cover diabetic difficulties that may affect someone as early as infancy and lead them to gain weight as a result of cells that are no longer functioning properly.

The diabetes prediction project reports discovery of knowledge from medical datasets such as Pima-Indians-Diabetes is important in order to make effective medical diagnosis. We apply various Machine Learning classification and ensemble Techniques to predict diabetes. Machine Learning Is a method that is used to train computers or machines explicitly. Various Machine Learning Techniques provide efficient result to collect Knowledge by building various classification and ensemble models from collected dataset.

Chapter 2**LITERATURE SURVEY**

- For the project, Predictive Analysis of Diabetes, the following research papers were referred before the implementation process. These research papers are studied and the summaries are as follows having the algorithm/technique used, as well as their performance and the dataset used in the implementation.

Title	Algorithm/Technique	Results	Dataset
<i>Random Forest Algorithm for the Prediction of Diabetes</i>	K.VijiyaKumar et al. proposed random Forest algorithm for the Prediction of diabetes develop a system which can perform early prediction of diabetes for a patient with a higher accuracy by using Random Forest algorithm in machine learning technique.	The proposed model gives the best results for diabetic prediction and the result showed that the prediction system is capable of predicting the diabetes disease effectively, efficiently and most importantly, instantly.	Self-prepared dataset
<i>Predicting Diabetes Onset: an Ensemble Supervised Learning Approach</i>	An ensemble supervised learning approach they used five widely used classifiers are employed for the ensembles and a meta-classifier is used to aggregate their outputs.	The results are presented and compared with similar studies that used the same dataset within the literature. It is shown that by using the proposed method, diabetes onset prediction can be done with higher accuracy.	Self-prepared dataset
<i>Diabetes Prediction Using Machine Learning Techniques</i>	Diabetes Prediction Using Machine Learning Techniques aims to predict	This project proposes an effective technique for earlier detection of the	Self-prepared dataset

	diabetes via three different supervised machine learning methods including: SVM, Logistic regression, ANN.	diabetes disease.	
<i>Diabetes Disease Prediction Using Data Mining</i>	In this system, they propose the use of algorithms like Bayesian and KNN.	Diabetes prediction using data mining assemble Intelligent Diabetes Disease Prediction System that gives analysis of diabetes malady utilizing diabetes patients' database.	Self-prepared dataset

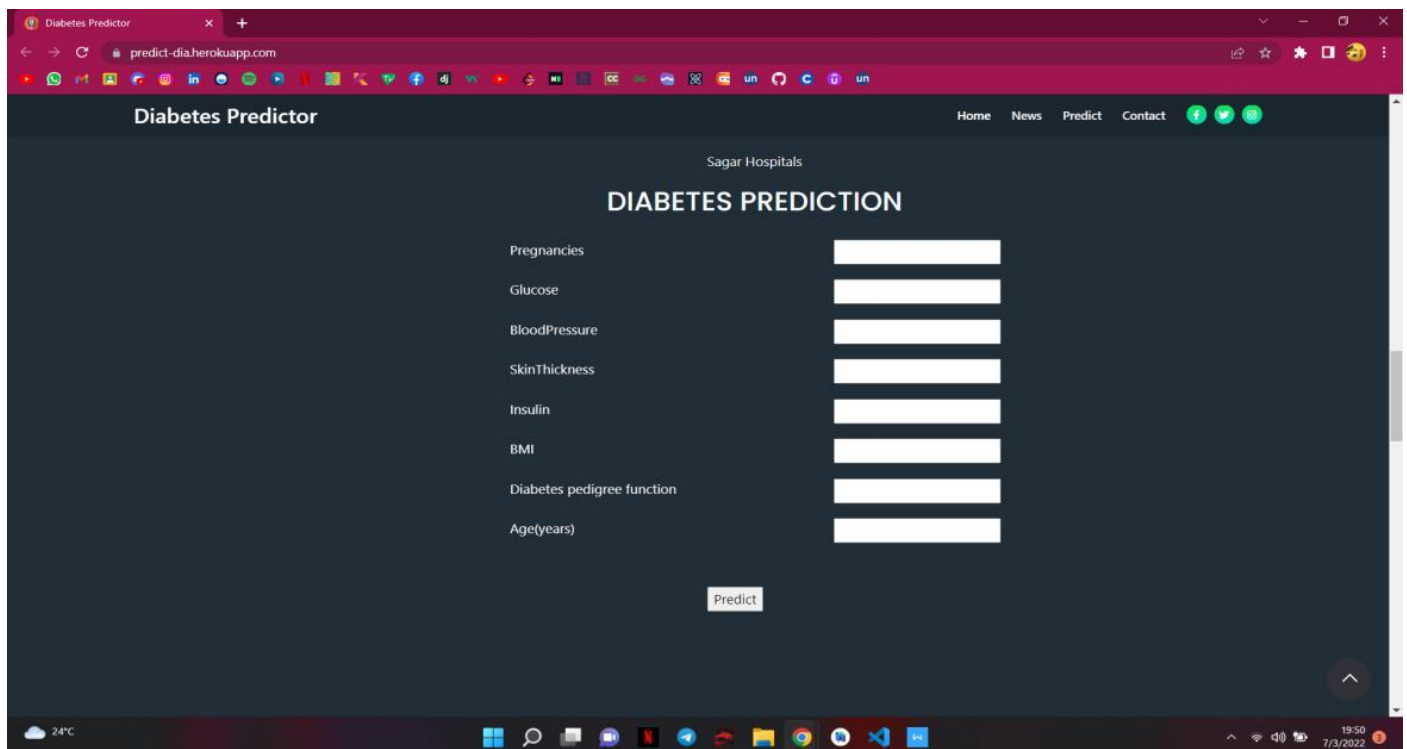
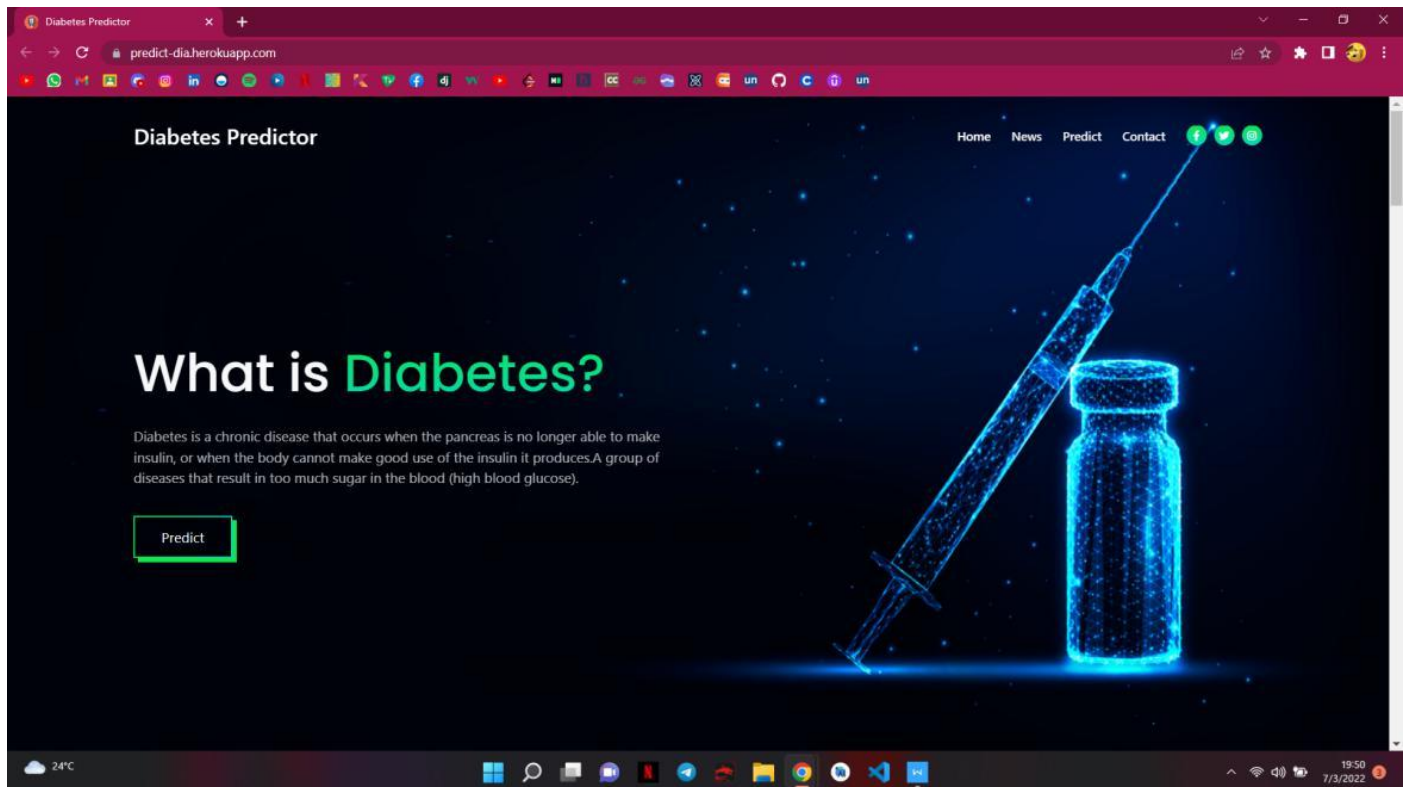
Chapter 3

SYSTEM DESIGN & METHODOLOGY

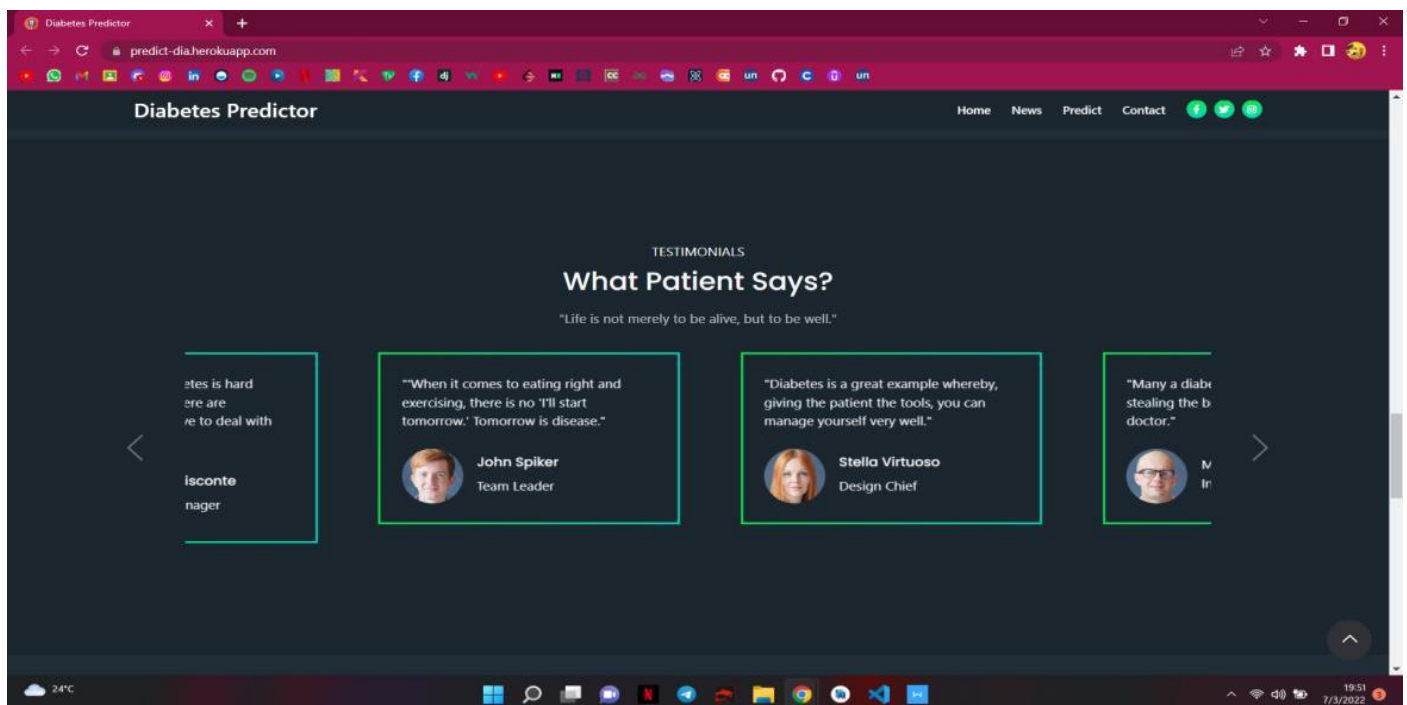
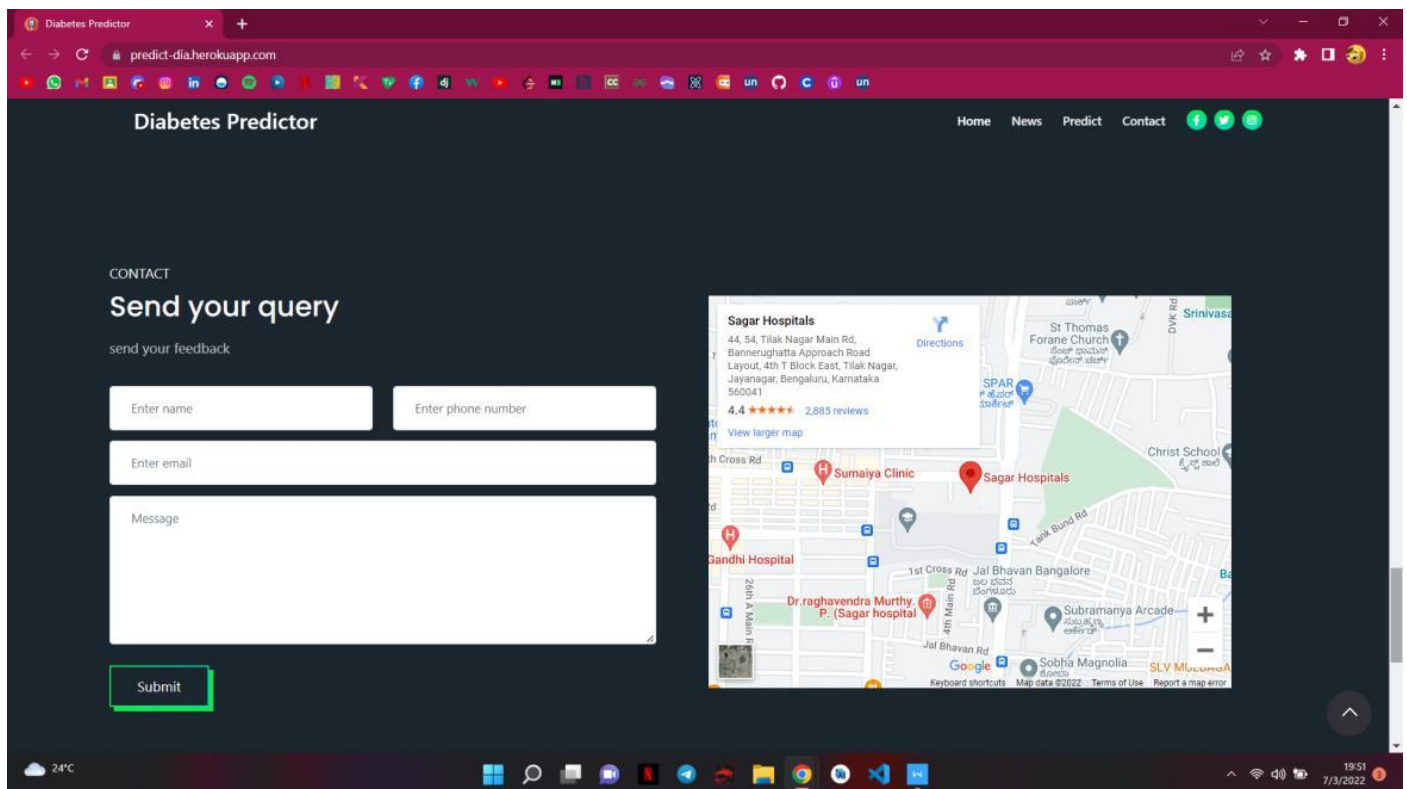


Chapter 4

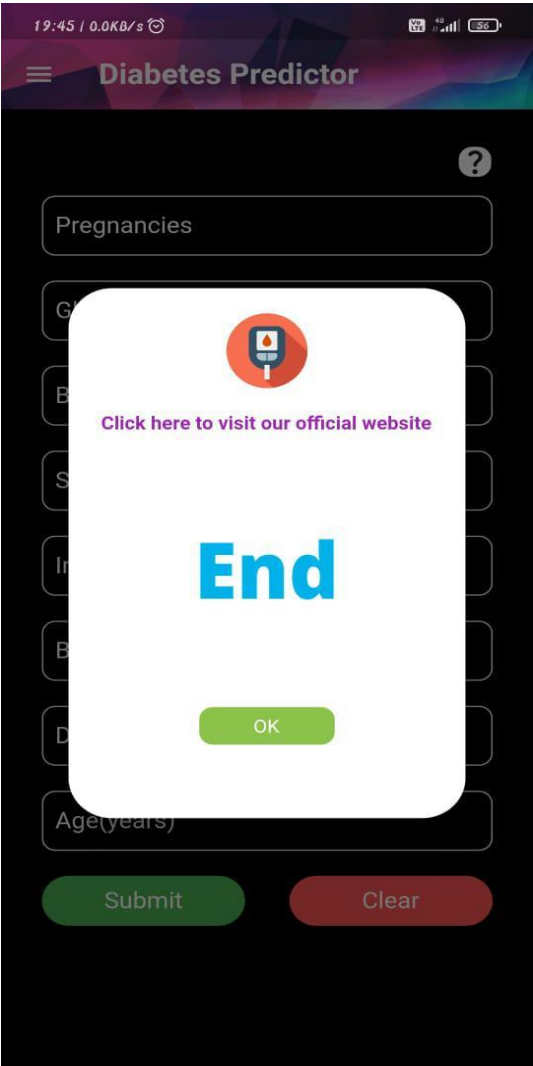
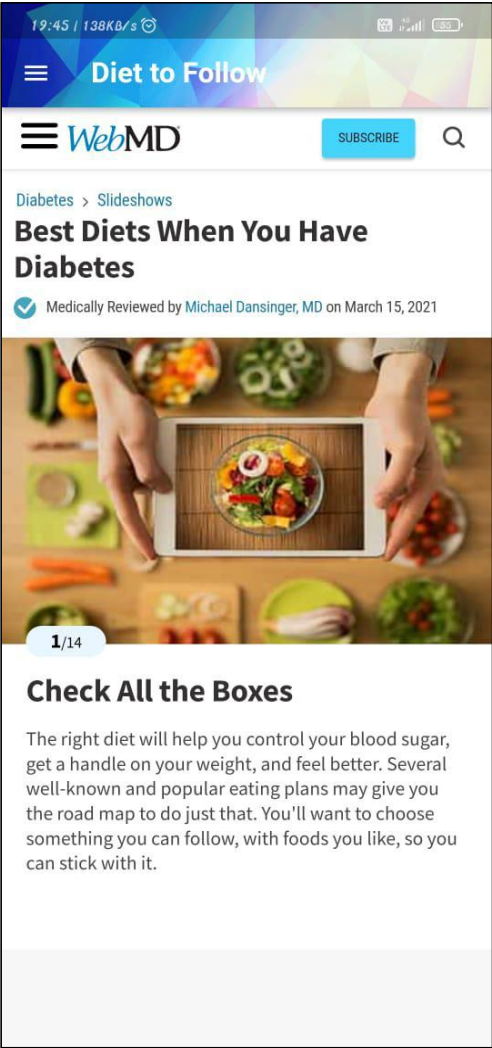
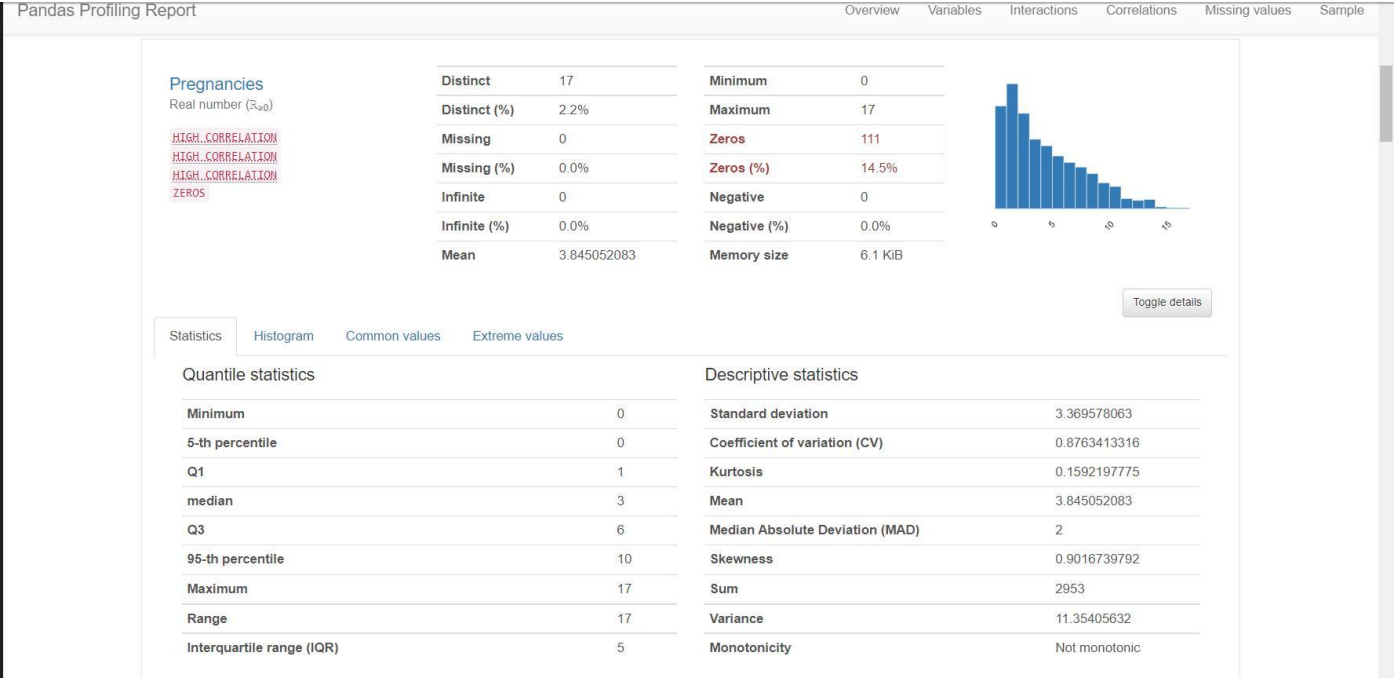
SNAPSHOTS AND RESULTS



Predictive Analysis of Diabetes



Predictive Analysis of Diabetes



Chapter 5

CONCLUSION

FUTURE ENHANCEMENTS

- ❖ To include disease prediction ML model(symptoms as input) in the site.
- ❖ To include IoT devices to calculate the input parameters for the model(eg : Insulin).
- ❖ To introduce BLOC architecturing in UI app for fast working.
- ❖ To apply ensemble methods to improve the accuracy of the prediction model.

REFERENCES

- 1) Gauri D. Kalyankar, Shivananda R. Poojara and Nagaraj V. Dharwadkar,” Predictive Analysis of Diabetic Patient Data Using Machine Learning and Hadoop”, International Conference On I-SMAC, 978-1-5090-3243-3, 2017.
- 2) Ayush Anand and Divya Shakti,” Prediction of Diabetes Based on Personal Lifestyle Indicators”, 1st International Conference on Next Generation Computing Technologies, 978-1-4673-6809-4, September 2015.
- 3) B. Nithya and Dr. V. Ilango,” Predictive Analytics in Health Care Using Machine Learning Tools and Techniques”, International Conference on Intelligent Computing and Control Systems, 978-1-5386-2745-7, 2017.
- 4) Dr Saravana kumar N M, Eswari T, Sampath P and Lavanya S,” Predictive Methodology for Diabetic Data Analysis in Big Data”, 2nd International Symposium on Big Data and Cloud Computing, 2015.
- 5) Diagnosis of Diabetes Using Classification Mining Techniques”, International Journal of Data Mining & Knowledge Management Process (IJDKP), 5 (1) (January 2015).

APPENDIX

Website:

<https://predict-dia.herokuapp.com/>

PROJECT SOURCE CODE LINK:

<https://github.com/sanjeevfa777/diabetes-predictor-using-Django>

https://github.com/rahulsaidileep/rest_flask_diabetes_predictor