

Project *On Diabetes* *Prediction*



DIABETES CHECKUP

TABLE OF CONTENTS::

<u>SR .NO</u>	<u>Title</u>	<u>Slide No</u>
1)..	<u>INTRODUCTION</u>	3
2)..	TITLE	4
3)..	ABSTRACT	5
4)..	RESEARCH INTRODUCTION	6
5)..	RESEARCH INTRODUCTION	7
6)..	INFORMATION	8
7)..	DATASET	9
8)..	DATA PREPROCESSING	10
9)..	PROPOSED METHOD	11
10..	INFORMATION	12
11)..	ARCHITECTURE	13
12)..	FACTS RELATED DIABETIC	14
13..	ACCURACY	15
14)..	CONCLUSION	16



INTRODUCTION

DATA SCIENCE PROJECT

Name : Gajanan Santosh Purud

Email :

gajananpurud36@gmail.com.



In this project, the objective is to predict whether the person has Diabetes or not based on various features like Number of Pregnancies, Insulin Level, Age, BMI.

TITLE: *#(Diabetes Checkup)*

******PREDICTION OF DIABETES USING
RANDOM FOREST CLASSIFIER*****



ABSTRACT

Abstract:- Diabetes is an illness caused because of high glucose level in a human body. Diabetes should not be ignored if it is untreated then Diabetes may cause some major issues in a person like: heart related problems, kidney problem, blood pressure, eye damage and it can also affects other organs of human body. Diabetes can be controlled if it is predicted earlier. To achieve this goal this project work we will do early prediction of Diabetes in a human body or a patient for a higher accuracy through applying, Various Machine Learning Techniques. Machine learning techniques Provide better result for prediction by constructing models from datasets collected from patients. In this

Research Introduction

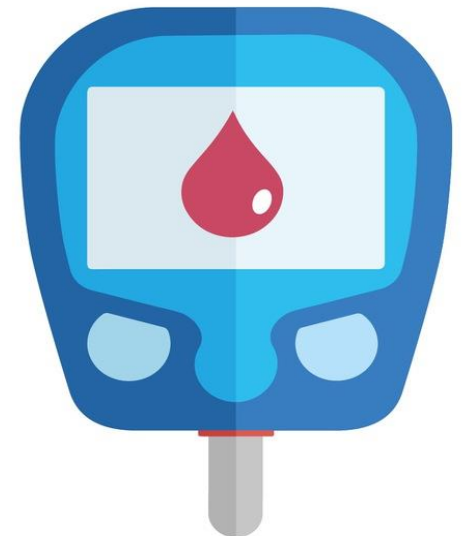
- Diabetes Mellitus (DM) is the most widely prevailing and rapidly spreading disease in all the countries all over the world.
- Retinal problem is the most dangerous and common complication which occurs due to diabetes. This problem is identified as Diabetic Retinopathy (DR).
- When the blood sugar of a patient goes high, it triggers injuries to the retinal blood vessels. In turn, these blood vessels may get swelled or it leakages blood. It may also stop blood circulation in that part.
- A survey says, approximately 80% of all patients who have had DM for at least 10 years, suffers from some degree of diabetic retinopathy.

Research Introduction

- According to National Diabetes and Diabetic Retinopathy Survey 2019, one in 46 diabetic patients in India becomes visually impaired and one in 7 diabetic patients have vision impairment.
- This survey also states that between 2015 and 2019, around 16.9% of the Indian population suffered from DR.
- Prevalence of blindness among diabetic patient was 2.1% and visual impairment was 13.7%.

Research Introduction

- Diabetic retinopathy (DR) is gradually becoming the major cause of blindness all over the globe.
- It keeps on progressing in silent mode and in maximum cases remains overlooked.
- Although this area is gaining interest of many researchers, major focus relies on prediction/diagnosis of DR by applying machine learning techniques on medical images.
- In our study, for the easy and early prediction of DR, we focus on prediction of DR using health records of the diabetic patients.



Dataset

- In this paper, records of 300 patients form our experimental dataset.
- We used the information of a group of regular outpatients.
- The data was collected and selected for the patients from different private hospitals in Central India.
- There are altogether 10 socio-demographic features identified for this study like age and gender of the patient, duration of the diabetes mellitus, Body Mass Index (BMI), systolic and diastolic blood pressure measurements, measurement of hemoglobin A1c (HbA1C), level of low-density lipoprotein (LDL), smoking status and level of Microalbuminuria (level of albumin).
- Both, male and female, patients with the age between the range 20 to 75 have been considered.

Table 1: Dataset Description

S No.	Attributes
1	Pregnancy
2	Glucose
3	Blood Pressure
4	Skin thickness
5	Insulin
6	BMI(Body Mass Index)
7	Diabetes Pedigree Function
8	Age

Data Pre-Processing

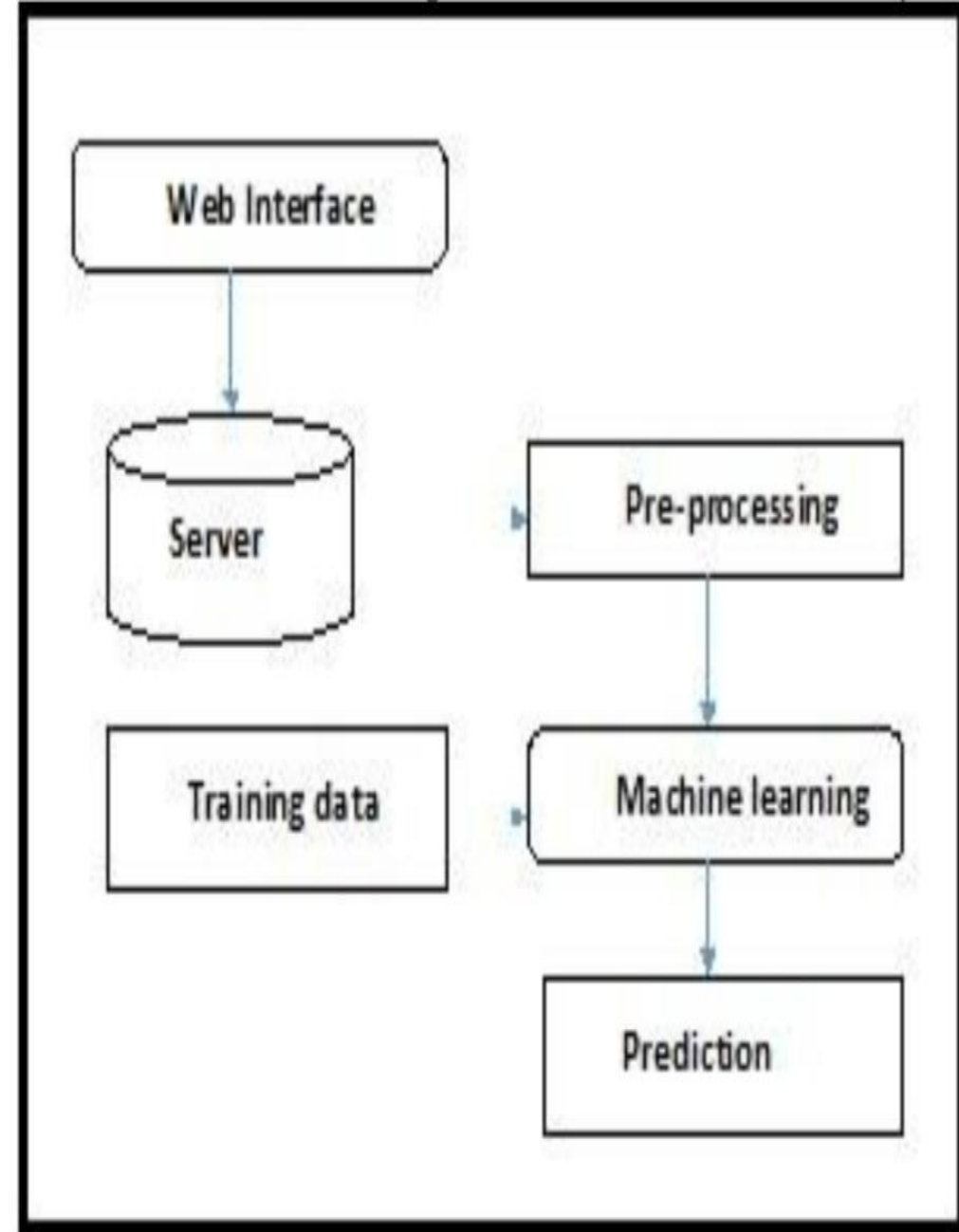
- Data pre-processing is crucial step.
- If the data collected contains any missing attributes or attribute values, contains noisy, outliers, duplicate or wrong data, then resultant accuracy will be depleted.
- Moreover, the inconsistencies in the collected data may also affect the subsequent work.
- That's why we have applied pre-processing on the gathered data.



#Proposed Method

We aim to build a model that can classify whether a patient has diabetes or not based on valid input. Raw data is pre-processing methods are applied (with the help of outlier analysis, missing values reduction, and feature scaling). Data is made ready for experimentation.

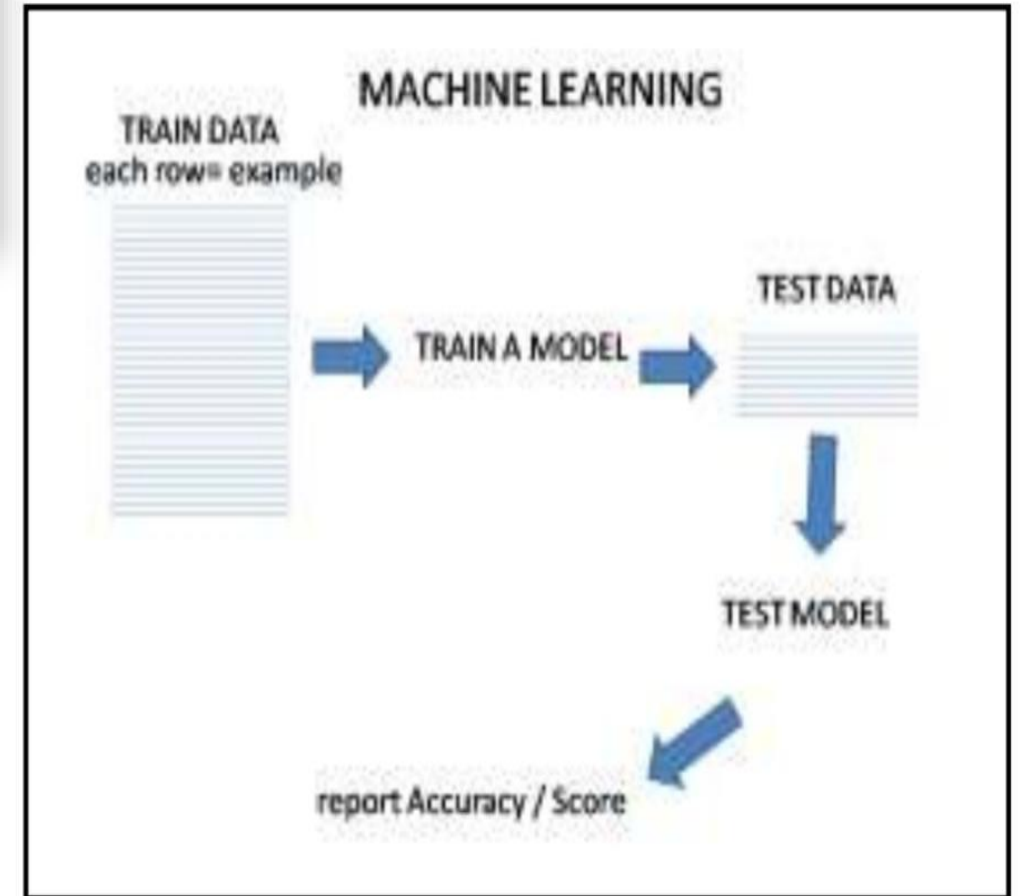
The set of available data is split for constructing and testing phases. Decision tree and random forest algorithms are applied on data to learn the class model. The optimal model is selected.



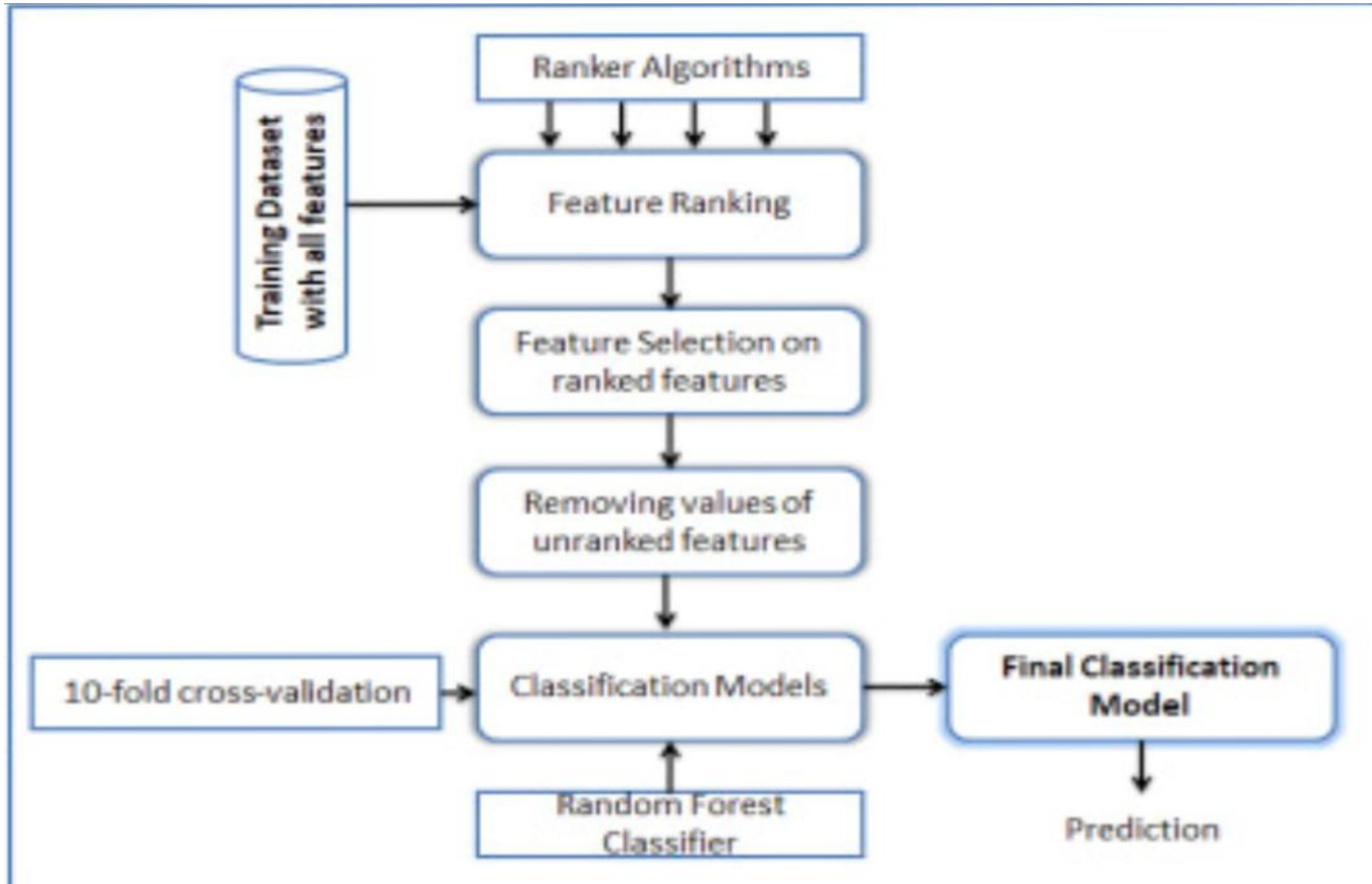


Machine Learning

Machine learning is an application of AI that has systems the capability to mechanically learn and improve from experience whereas not being expressly programmed.



ARCHITECTURE



Facts

- Diabetes recently affects around 346 million people.
- Also the mayor cause for:
 - Heart stroke
 - Kidney failure
 - Lower-limb amputation
 - Blindness
- One-third go undetected in early stage.
- Early detection and treatment - substantial health benefits, (avoiding or minimizing the mentioned complications).

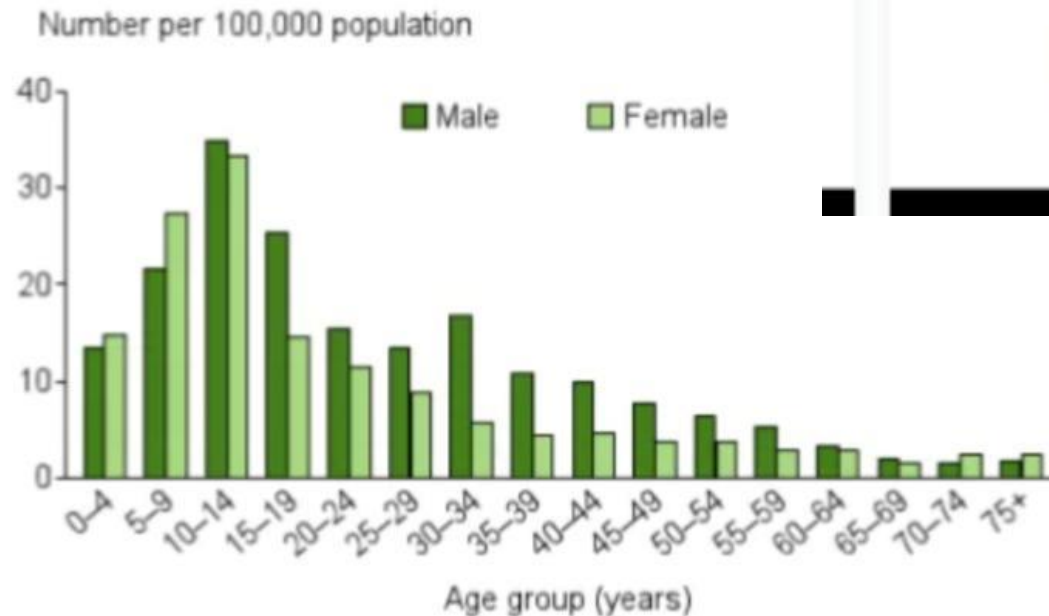
DIABETES CASES, MILLIONS

COUNTRY	2013	2035 (projected rise)
CHINA	98.4	142.7
INDIA	65.1	109
USA	24.4	29.7
BRAZIL	11.9	19.2
RUSSIA	10.9	11.2
INDONESIA	8.6	14.2
MEXICO	8.7	15.7
EGYPT	7.5	13.1
GERMANY	7.6	8.1
TURKEY	7	11.8
JAPAN	7.2	6.7
PAKISTAN	6.7	12.8

FACTS RELATED TO DIABETES

Amazing Facts

Diabetes compendium



Types of Diabetes

Type 1 diabetes

10%

The body does not produce insuline

No

Type 2 diabetes

90%

The body produces insuline but is not used efficiently

CAN BE PREVENTED

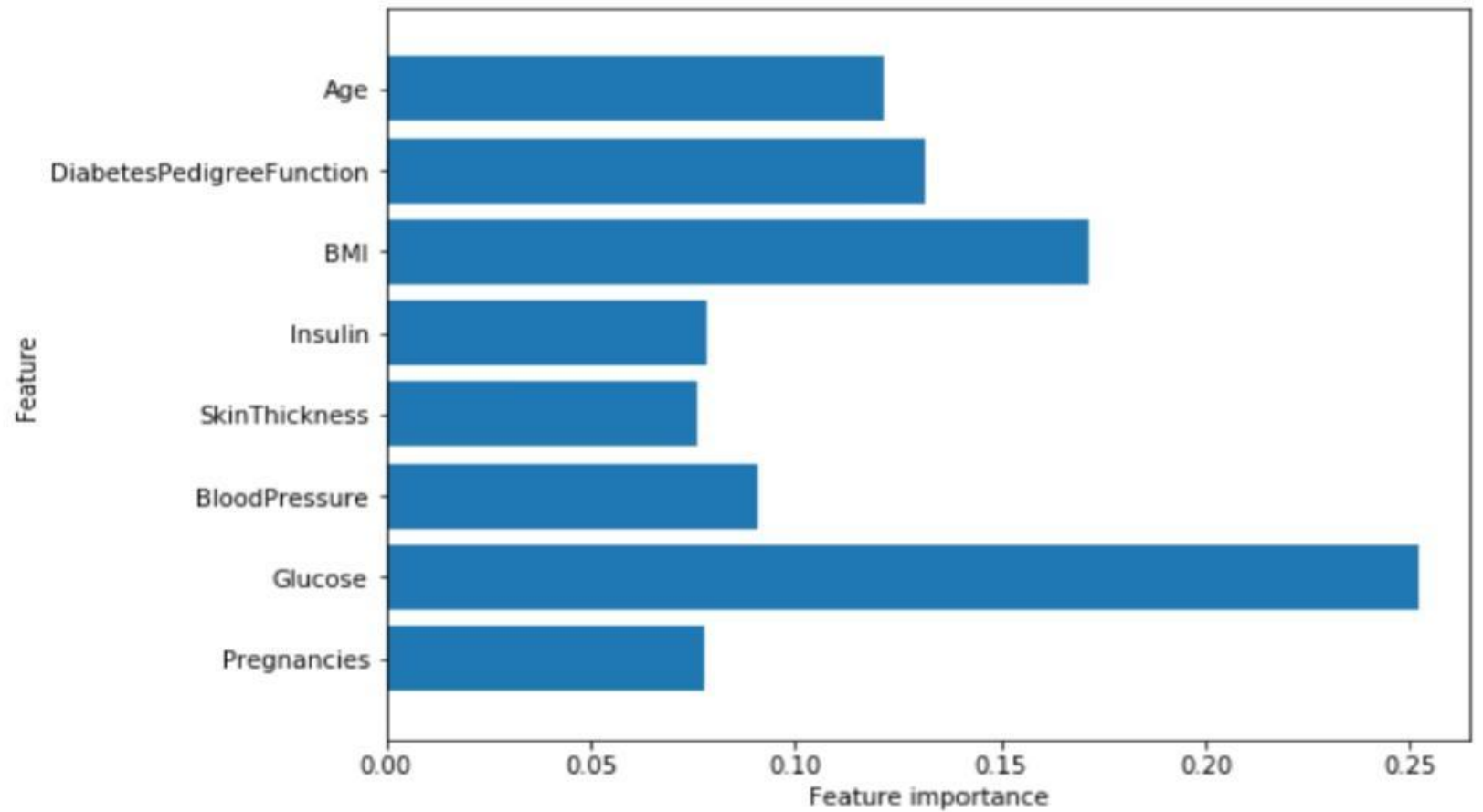


HOW TO PREVENT?



YES

RESULT



Results & Conclusion:

With Random Forest Classifier giving a reasonably well accuracy of close to **82.68%**, this approach and the ML model look promising in assisting healthcare professionals to provide a prediction. Moreover, before finalizing a health situation diagnosis based on ML models, it is essential to place a greater focus on interpreting the confusion matrix as False positives – False negatives can be risky.

Accuracy

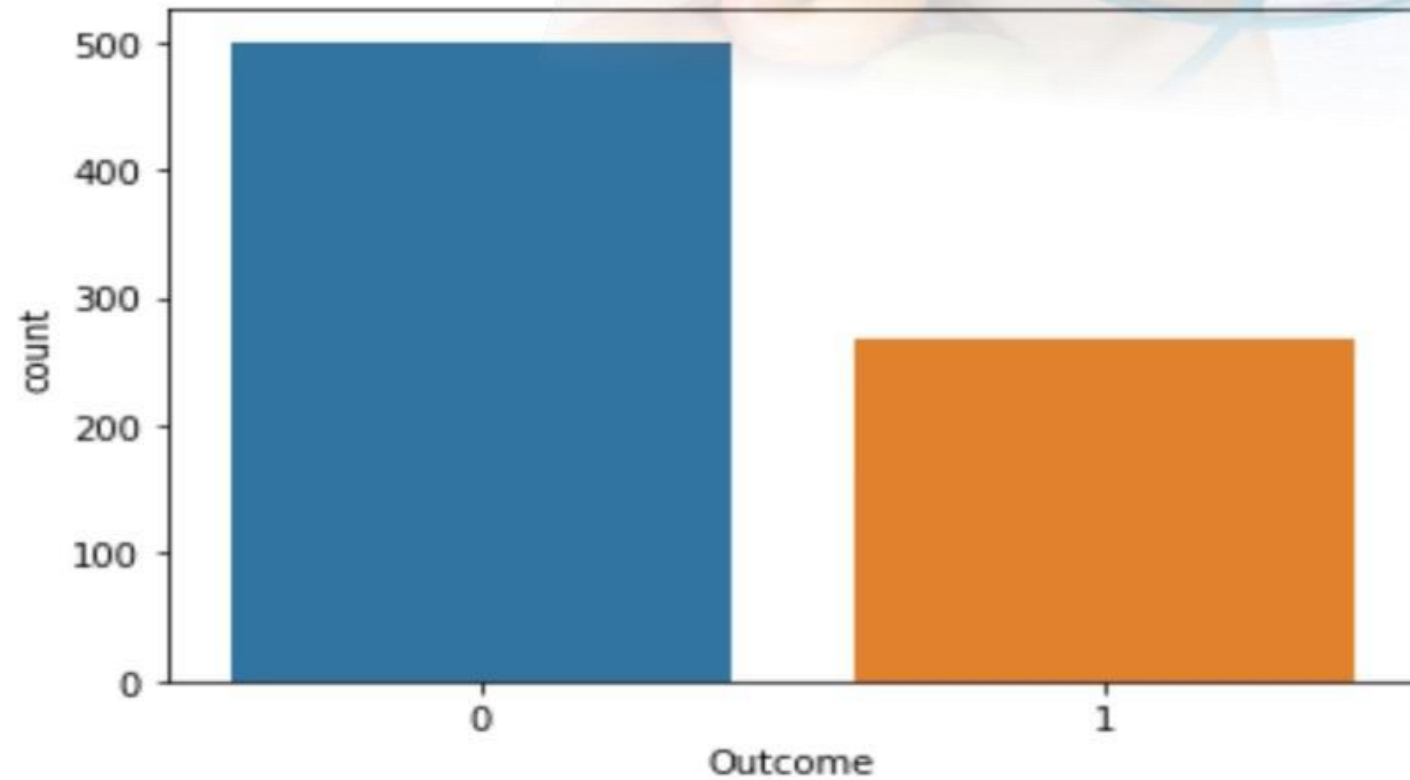


Figure 1: Ratio of Diabetic and Non Diabetic Patient

Thank You



Thank you.....

► *Exposys data labs*

