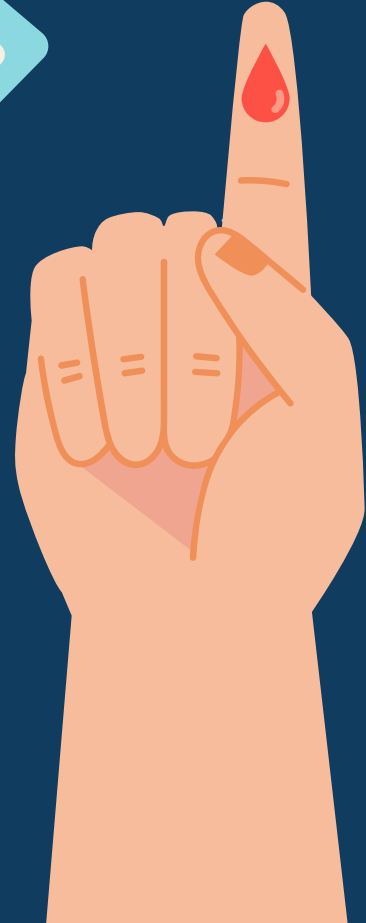


DIABETES CLASSIFICATION

Done by:
Nawaf Almutairi
Surayyi Alqahtani

Instructor :
Dr. Mejdal Alqahtani





OUTLINES

01

INTRO

02

**PROJECT
GOLE**



03

**EXPLORATORY
DATA
ANALYSIS**

04

**CLASSIFICATION
MODELS**

05

CONCLUSIONS



01

INTRODUCTION

What is the diabetes ?





INTRODUCTION



Diabetes is a chronic (long-lasting) health condition that affects how your body turns food into energy.

Most of the food you eat is broken down into sugar (also called glucose) and released into your bloodstream. When your blood sugar goes up, it signals your pancreas to release insulin.





SYMPTOMS OF DIABETES

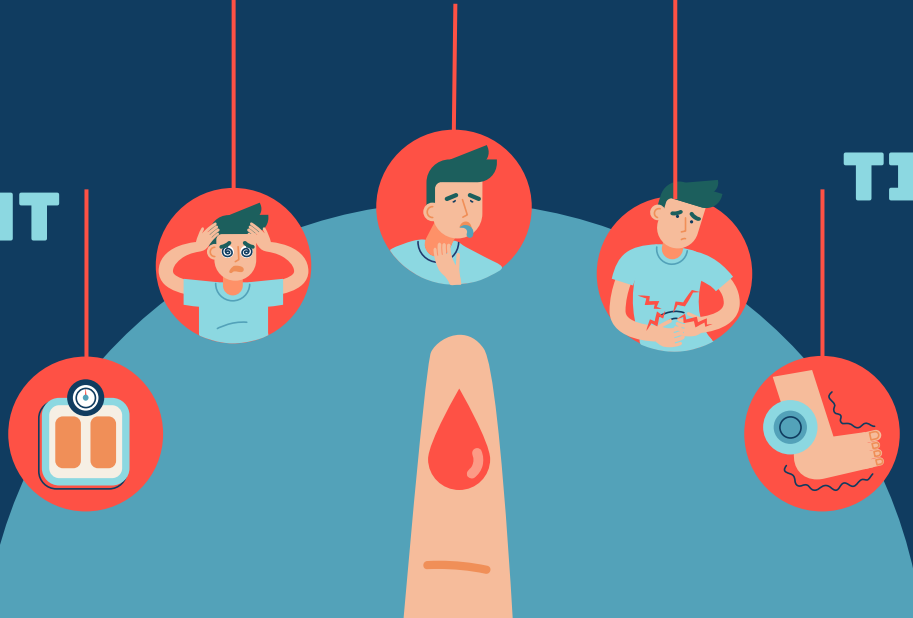
DIZZINESS

THIRSTY

HUNGRY

OVERWEIGHT

TINGLING





BLOOD SUGAR CHART

CATEGORY	Fasting Value (mg/dl)		Postprandial (mg/dl)
	Min. Value	Max. Value	Value 2h after eating glucose
NORMAL	70	100	Less than 140
EARLY DIABETES	101	126	140 to 200
ESTABLISHED DIABETES	More than 126	-	More than 200



THERE ARE 3 MAIN TYPES OF DIABETES

Type 1 Diabetes – where the body's immune system attacks and destroys the cells that produce insulin.

Type 2 Diabetes – where the body does not produce enough insulin, or the body's cells do not react to insulin.

Gestational Diabetes - Gestational diabetes develops in pregnant women who have never had diabetes.





+ PROJECT GOAL

“The goal of this project is to predict the diabetes and get prevention from it”



EXPLORATORY DATA ANALYSIS



DATA STRUCTURE:



It is consisted of 22 columns and 253K rows.

	Diabetes_binary	HighBP	HighChol	CholCheck	BMI	Smoker	Stroke	HeartDiseaseorAttack	PhysActivity	Fruits	...	AnyHealthcare	NoDocbcCost	Ge
0	0.0	1.0	1.0	1.0	40.0	1.0	0.0	0.0	0.0	0.0	...	1.0	0.0	
1	0.0	0.0	0.0	0.0	25.0	1.0	0.0	0.0	1.0	0.0	...	0.0	1.0	
2	0.0	1.0	1.0	1.0	28.0	0.0	0.0	0.0	0.0	1.0	...	1.0	1.0	
3	0.0	1.0	0.0	1.0	27.0	0.0	0.0	0.0	1.0	1.0	...	1.0	0.0	
4	0.0	1.0	1.0	1.0	24.0	0.0	0.0	0.0	1.0	1.0	...	1.0	0.0	
...	
253675	0.0	1.0	1.0	1.0	45.0	0.0	0.0	0.0	0.0	1.0	...	1.0	0.0	
253676	1.0	1.0	1.0	1.0	18.0	0.0	0.0	0.0	0.0	0.0	...	1.0	0.0	
253677	0.0	0.0	0.0	1.0	28.0	0.0	0.0	0.0	1.0	1.0	...	1.0	0.0	
253678	0.0	1.0	0.0	1.0	23.0	0.0	0.0	0.0	0.0	1.0	...	1.0	0.0	
253679	1.0	1.0	1.0	1.0	25.0	0.0	0.0	1.0	1.0	1.0	...	1.0	0.0	

253680 rows × 22 columns

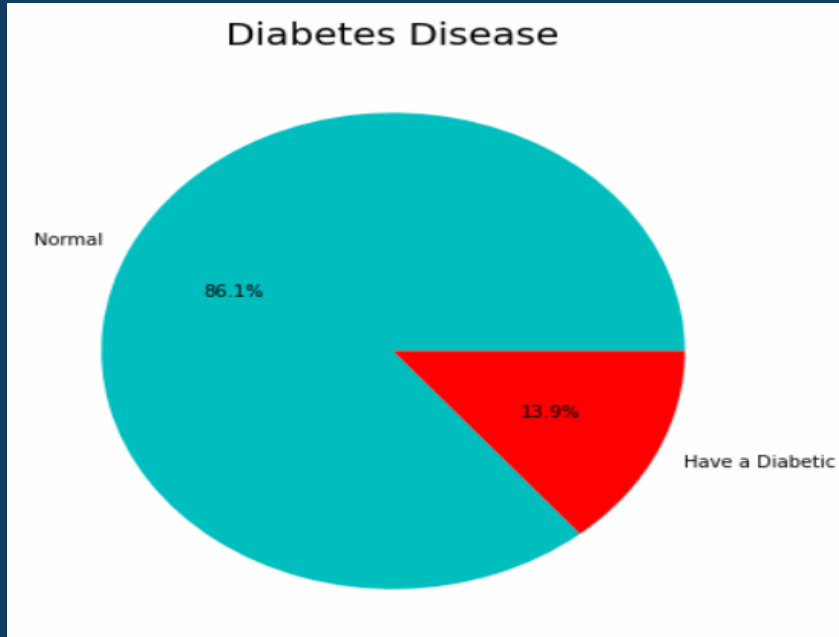




EXPLORATORY DATA ANALYSIS



DIABETES PIE CHART :



We have around 214k a Normal and 40k having a Diabetes



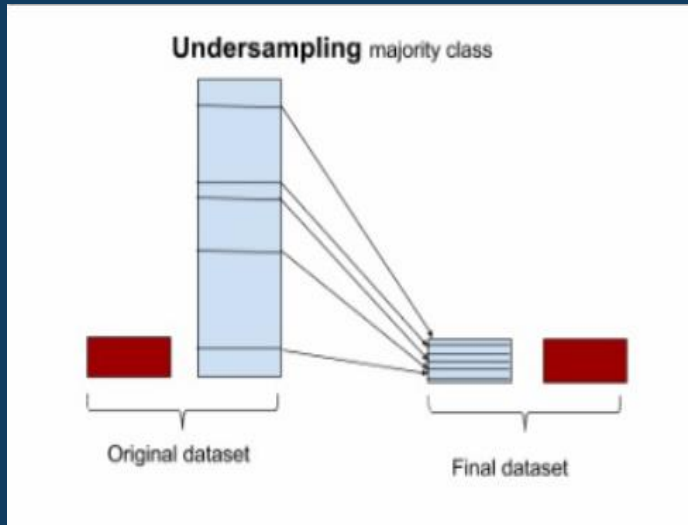


EXPLORATORY DATA ANALYSIS



HOW WE DEAL WITH IMBALANCED DATA :

By using resampling SMOTE (**Under Sampling**)



Now, we have balanced data for each classes.

40k for each.



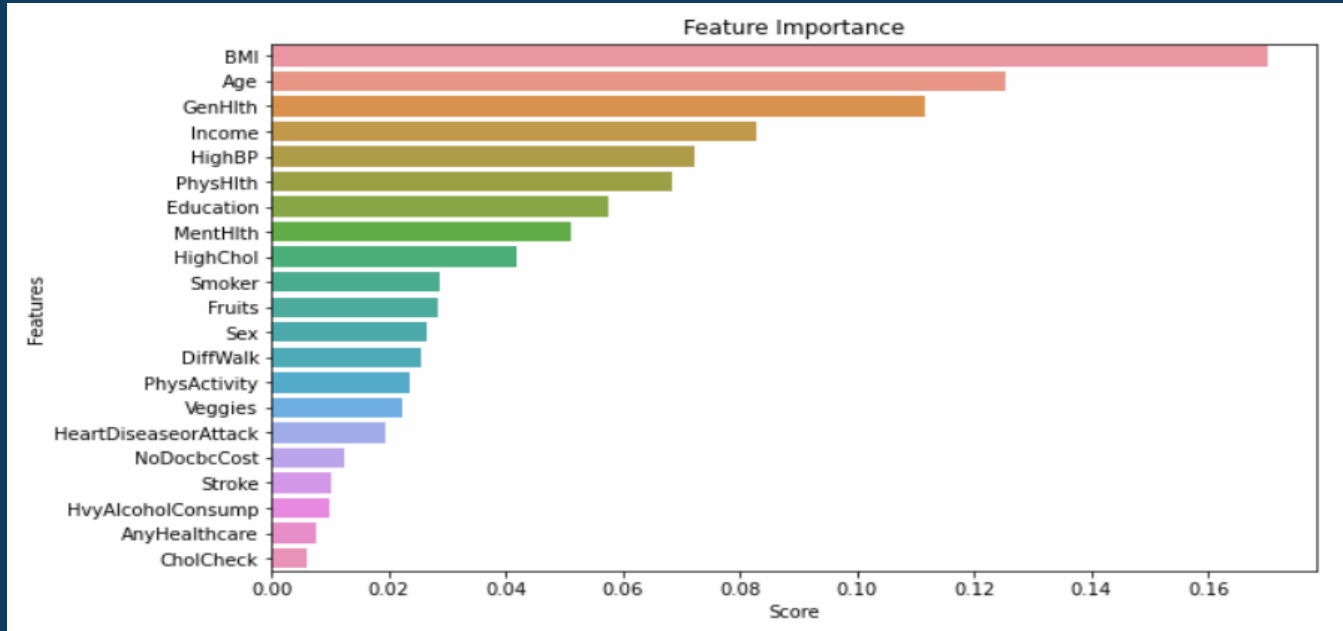


EXPLORATORY DATA ANALYSIS



THE IMPORTANT FEATURES :

This is a Bar chart shows the Feature Importance of Diabetes

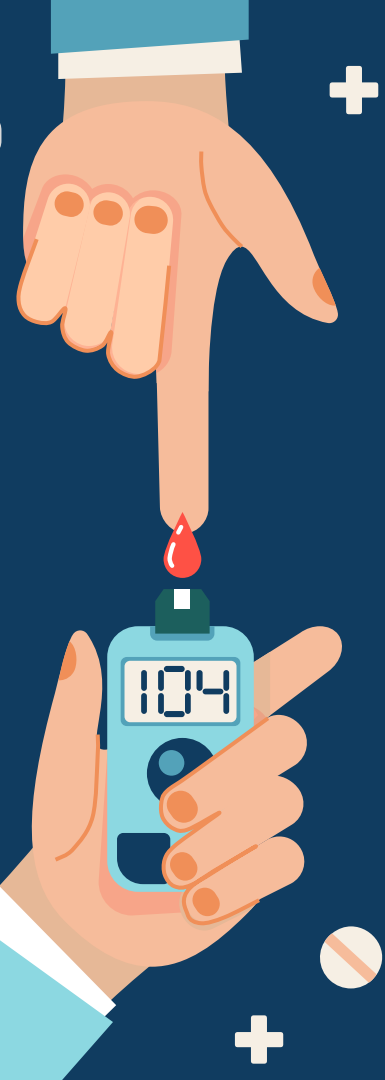




CLASSIFICATION MODELS

TRYING DIFFERENT MODELS FOR CLASSIFICATION :

- 1- Random Forest Classifier.
- 2- Decision Tree Classifier.
- 3- Logistic Regression.
- 4- Kneighbors Classifier.
- 5- Support Vector Machine SVM.





CLASSIFICATION MODELS



FIND OUT THE BEST MODELS:

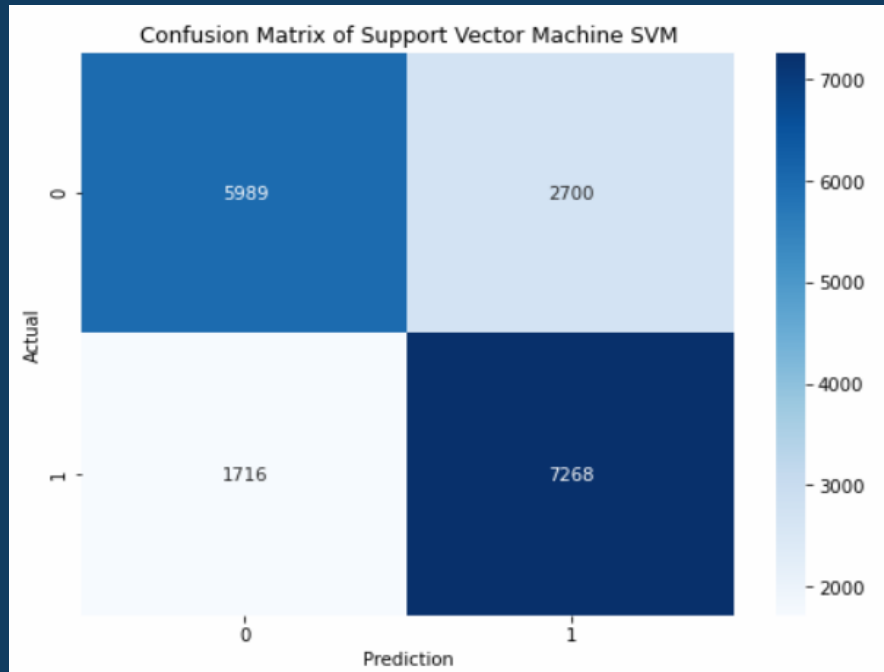
MODELS	ACCURACY	RECALL
RANDOM FOREST	0.74	0.77
DECISION TREE	0.61	0.50
LOGISTIC REGRESSION	0.74	0.76
KNN	0.69	0.70
SVM	0.75	0.80



CLASSIFICATION MODELS

THE SVM IS THE BEST MODELS:

This is a Confusion Matrix of Support Vector Machine SVM



TP = 7268

TN = 5989

FP = 2700

FN = 1716



CONCLUSIONS

- The diabetes is a chronic disease , and there is no main cause of being a diabetic.
- Body mass index is the highest factor can cause the diabetes then the age.
- We use some models to find the best accuracy and recall ,and the SVM turned out to be the best model.
- In the future we will try to distinction between diabetes types.





THANKS!

Do you have any questions?

