

CONTENT

01

INTRODUCTION

02

PROBLEM DEFINITION

03

DATA PROCESSING

04

VISUALIZATIONS

05

MODELLING

06

TRAINING & COMPARISON

07

CONCLUSION



GOALS AND OBJECTIVES

EDA n° 1

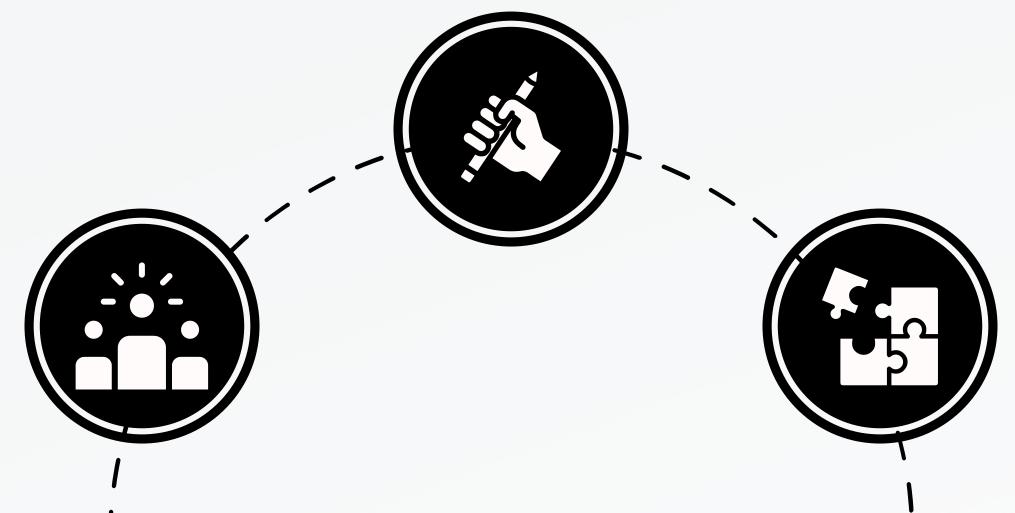
Exploring the data in order to gain insights into the structure and Modelling options

Modelling° 2

Selecting 2 models based on the findings from the EDA process

Training & Testingn° 3

Testing Models & adding Hyper Parameter in order to increase accuracy of Model



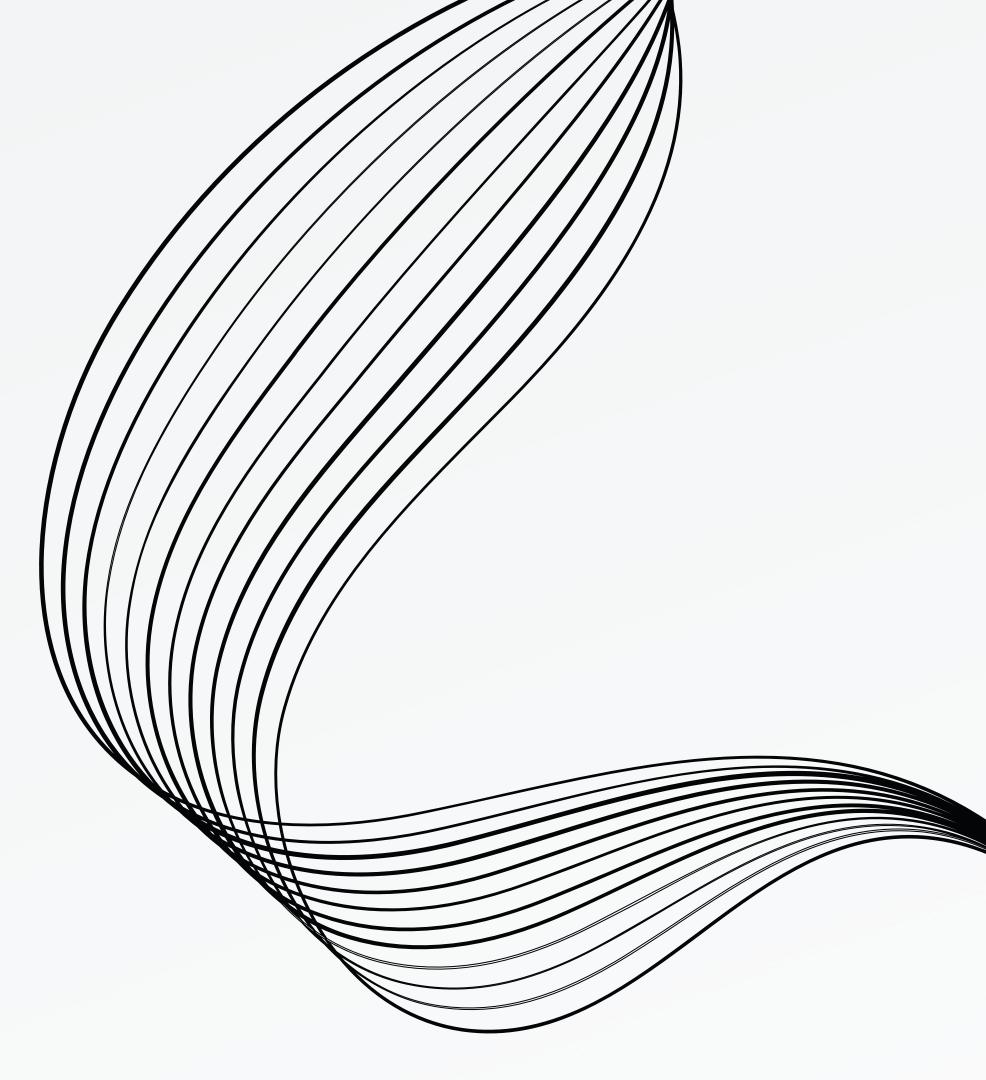


PROBLEM DEFINITION

Creating a Supervised Learning Model based on Diabetes Dataset

DATA PROCESSING

- Data Cleaning
- Feature Engineering
- Model Building



EXPLORATION & VISUALIZATION &

egnancies	Glucose	BloodPressure	SkinThickness	Insulin	вмі	DiabetesPedigreeFunction	Age	Outcome	C_BMI	C_InsulinScore	C_Gluco
6	148	72	35	0	33.6	0.627	50	1	Obesity 1	Abnormal	Hi
1	85	66	29	0	26.6	0.351	31	0	Overweight	Abnormal	Norn
8	183	64	0	0	23.3	0.672	32	1	Normal	Abnormal	Hi
1	89	66	23	94	28.1	0.167	21	0	Overweight	Normal	Norn
0	137	40	35	168	43.1	2.288	33	1	Obesity 3	Abnormal	Hi

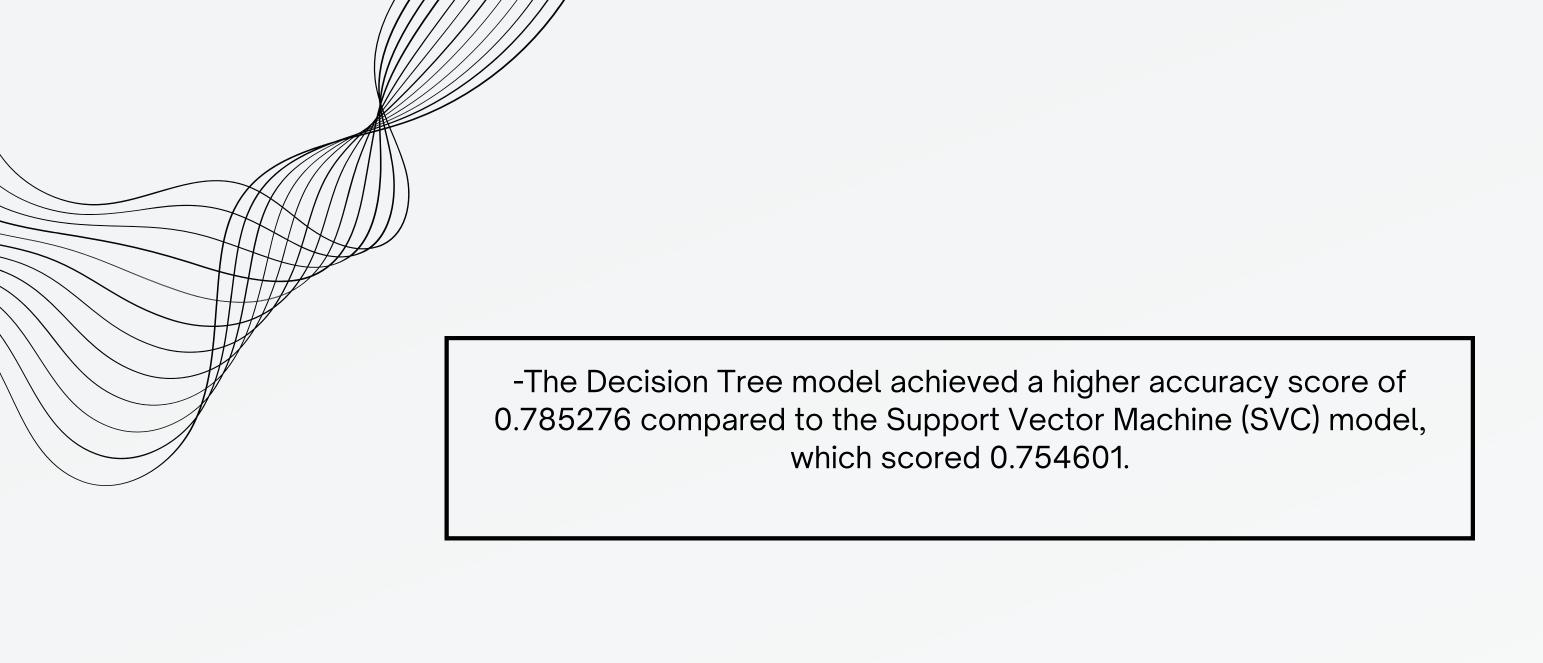
MODELLING

	precision	recall	f1-score	support
0	0.78	0.92	0.85	105
1	0.79	0.53	0.64	58
accuracy			0.79	163
macro avg	0.79	0.73	0.74	163
weighted avg	0.79	0.79	0.77	163

	precision	recall	f1-score	support
0	0.81	0.91	0.86	56
1	0.67	0.45	0.54	22
accuracy			0.78	78
macro avg	0.74	0.68	0.70	78
weighted avg	0.77	0.78	0.77	78

DECISION TREE

SVD MODEL



Further Analyzes required as the SVD model accuracy predicting closer to 1 needs More improvement

CONCLUSION

BOTH MODELS SHOW REASONABLE PERFORMANCE, WITH THE DECISION TREE MODEL HAVING A SLIGHTLY BETTER ACCURACY, MAKING IT A FAVORABLE CHOICE FOR THIS SPECIFIC TASK.