

Water

(وَجَعَلْنَا مِنَ الْمَاءِ كُلَّ شَيْءٍ حَيٍّ)

Design a suitable pattern recognition technique for classifying different water safety

Presented To : Safaa Alsafari

Presented By :

Layali alsolami

Bushra dajam

Jana osta



Agenda

PART 1

Introduction

PART 2

EDA

PART 3

Preprocessing

PART 4

Desicion Tree

PART 5

Support Vector Machine (SVM)

PART 6

Results

PART 7

Conclusion

Introduction

Unfortunately, there are communities throughout Africa that suffer from the lack of safe, clean water for drinking, cooking, and hygiene.



HYPOTHESIS



01

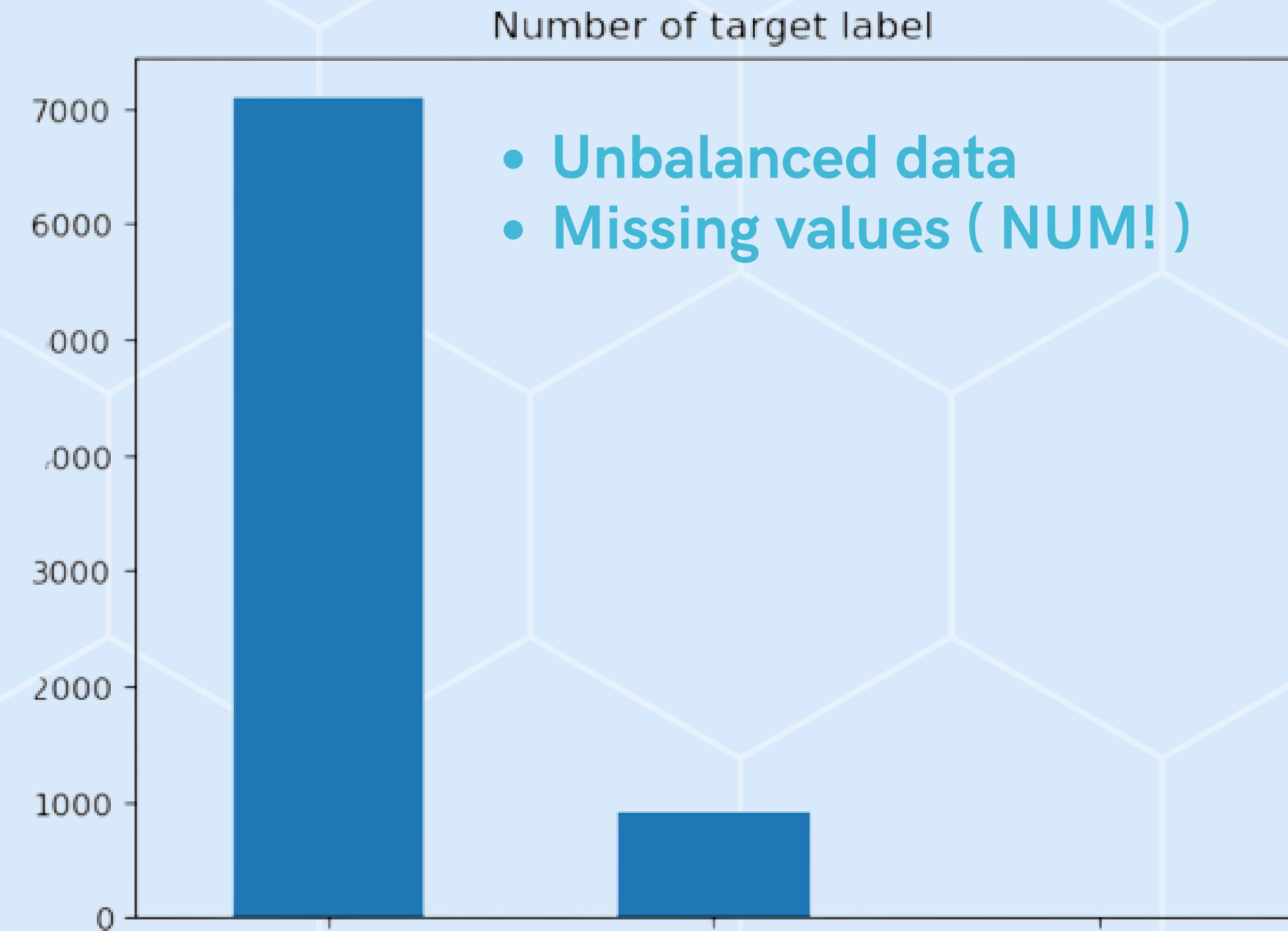
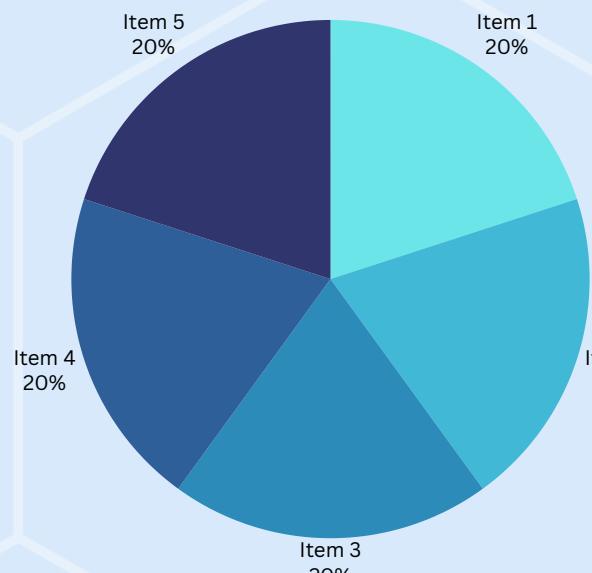
If one is water is safe ,the other it will be water is not safe.

02

Decision Tree is more effective than SVM in our dataset problem.

EDA

The data contain 7999 ‘Rows’ as known as samples and 21 ‘columns’ including the features and the class label.



- Unbalanced data
- Missing values (NUM!)

preprocessing



remove all records that contain NUM!



convert object columns to numeric data type

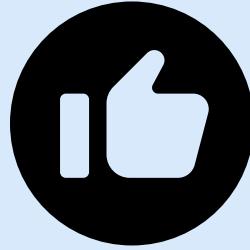


normalize dataset to make sure have same feature scaling



use random oversample technique to handle with unbalanced data

Decision Tree



STRENGTHS

- Don't requiring much computation
- handle both continuous and categorical variables
- handle large datasets
- handle missing value
- Requires little data preparation



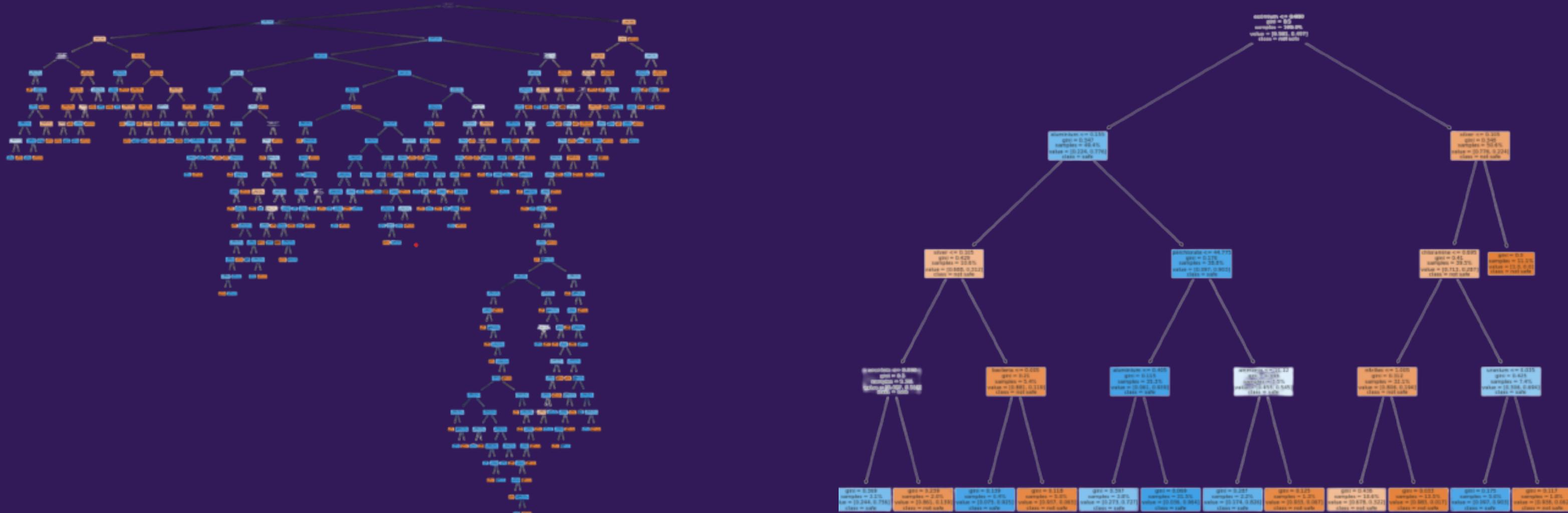
WEAKNESSES

- prone to overfitting
- Optimization
- unstable

Decision Tree

Hyperparameter Tuning

Best hyperparameters is: {'criterion': 'gini',
'max_depth': 4, 'max_features': None,
'splitter': 'best'}



SVM



01

What is SVM

02

How Does it Works ?

03

How to improve it ?

04

Advantages and
disadvantages (SVM)

WHAT IS IT?

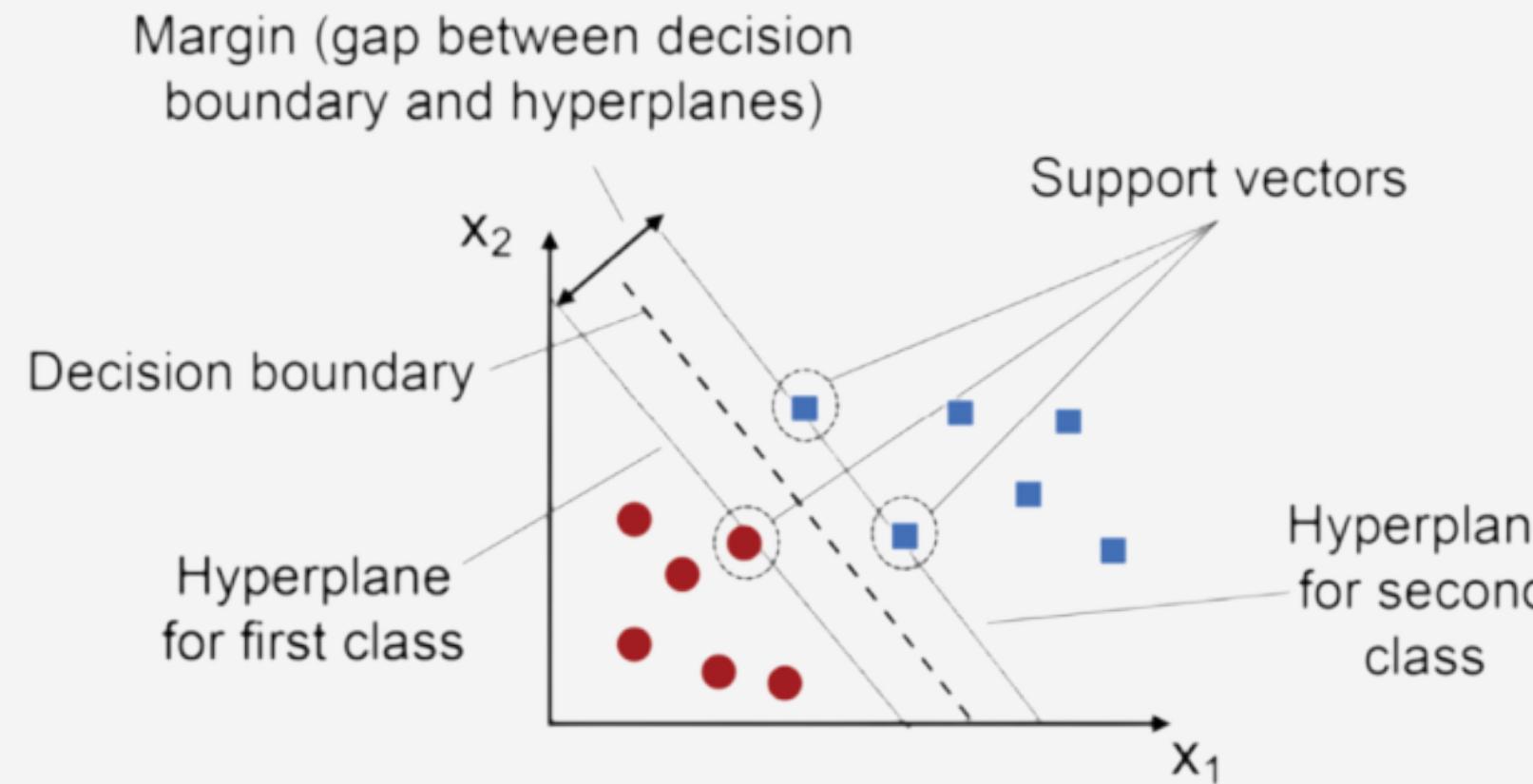
- is a supervised
- used for classification and regression
- used for both linear and non-linear

HOW TO IMPROVE IT?

- By choosing the best values for the hyperparameter :
1-Kernal
2-C
3-Gamma

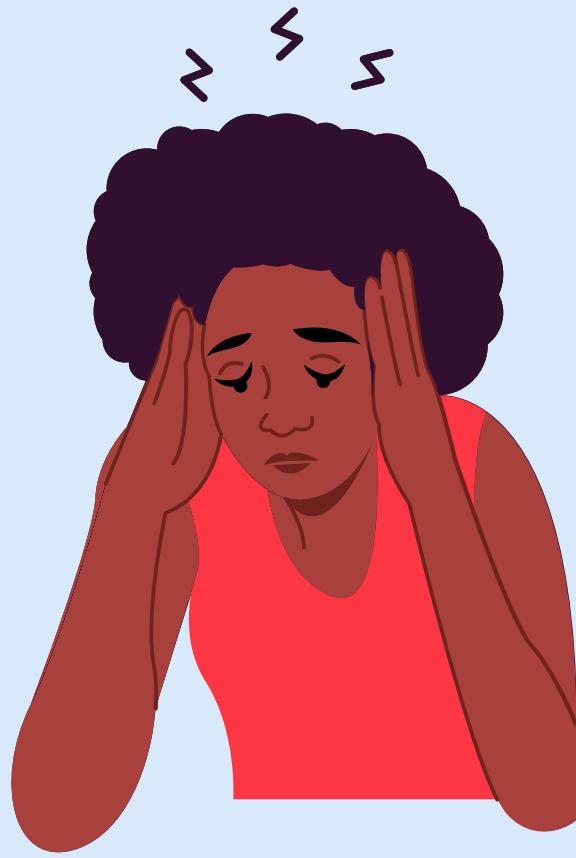
HOW DOES IT WORKS?

- by finding the best possible boundary or hyperplane that separates the data points into different classes
- The hyperplane is chosen in such a way that it maximizes the margin between the two classes





STRENGTH

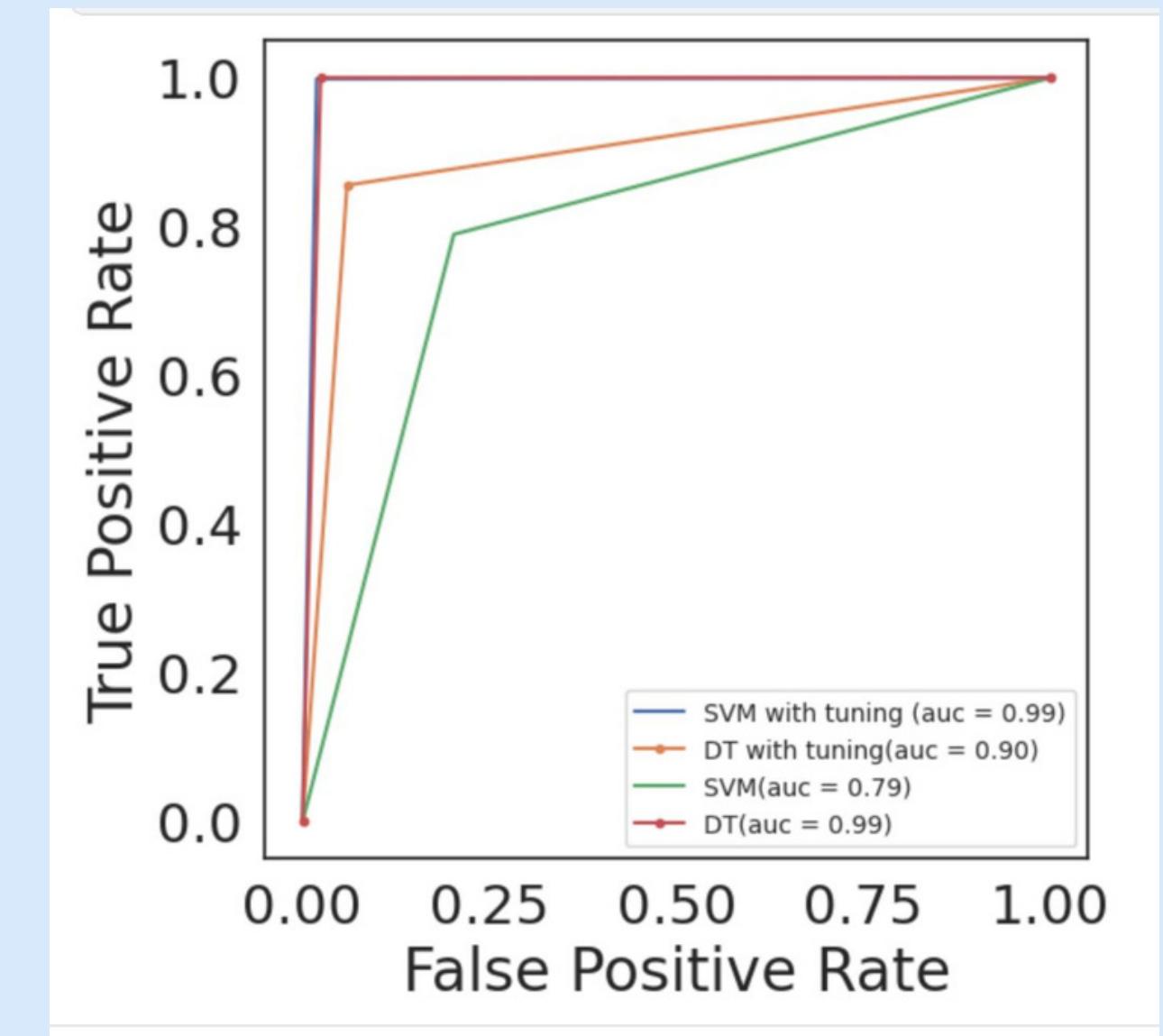


Highly Accurate , Handle many features

WEAKNESS

Very slow , Take time to process

RESULTS



	DT		DT_best		SVM		SVM_best	
Accuracy	0.986		0.898		0.7893		0.989	
precision	1.00	0.98	0.86	0.94	0.78	0.80	1.00	0.98
recall	0.98	1.00	0.94	0.86	0.80	0.78	0.98	1.00
F1-measure	0.99	0.99	0.90	0.90	0.97	0.97	0.99	0.99
misclassification	.0130		0.101		0.210		0.010	

At the end

NOW WE
HAVE SAVE
WATER!

*Thank
you!*

