Professional training

Classification with other model

Summary: In this Module, you will learn about other classification models.

Version: 1.00

Contents

Chapter I

Introduction

Greetings!

> If you haven't already done so, read en.toolkit.pdf.

What this Module will cover:

In this module, we're diving headfirst into the captivating world of classification using some of the most powerful algorithms. After laying down the groundwork with essential concepts like data preprocessing, regression logistique, and understanding evaluation metrics, we're all set to explore the capabilities of Decision Trees and XGBoost.

Wishing you success in your learning journey.

Chapter II

General instructions

Unless explicitly specified, the following rules will apply every day of this Professional training.

- This subject is the one and only trustable source. Don't trust any rumor.

- This subject can be updated up to one hour before the turn-in deadline.

- The assignments in a subject must be done in the given order. Later assignments won't be rated unless all the previous ones are perfectly executed.

- Be careful about the access rights of your files and folders.

- Your assignments will be evaluated by your peers.

- You <u>must not</u> leave in your turn-in your workspace any file other than the ones explicitly requested By the assignments.

- You have a question? Ask your left neighbor. Otherwise, try your luck with your right neighbor.

- Every technical answer you might need is available in the `man` or on the Internet.

- By Thor, by Odin! Use your brain!!!

Chapter III

Exercise 00

	Exercise 00
	Data exploration
Turn-in directory : *ex00/*	
Files to turn in : `Medium02.ipynb`	
Allowed functions : `All`	

For this first exercise, you'll need to load the data into colab and do some data exploration, i.e. you'll need to understand your data.

⚠ This time the data will resemble that which you may encounter in the professional world.
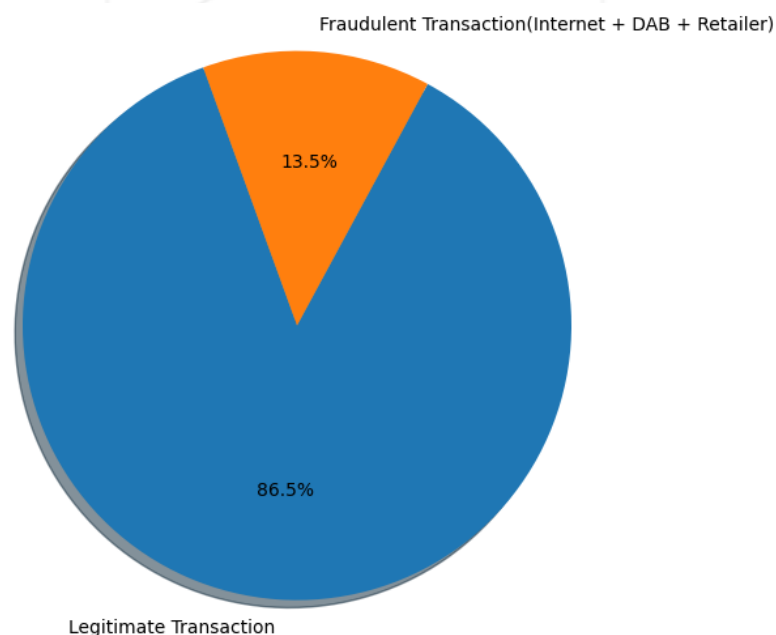
Chapter IV

Exercise 01

	Exercise 01
	Data Visualisation
Turn-in directory : *ex*01/	
Files to turn in : `Medium02.ipynb`	
Allowed functions : `All`	

In this exercise, you will delve into exploring the relationship between various numerical attributes within an unbalanced banking dataset and their potential influence on financial transaction outcomes. By employing visualization techniques, you will gain insights into how these numerical features contribute to the classification of financial transactions as fraudulent or legitimate.
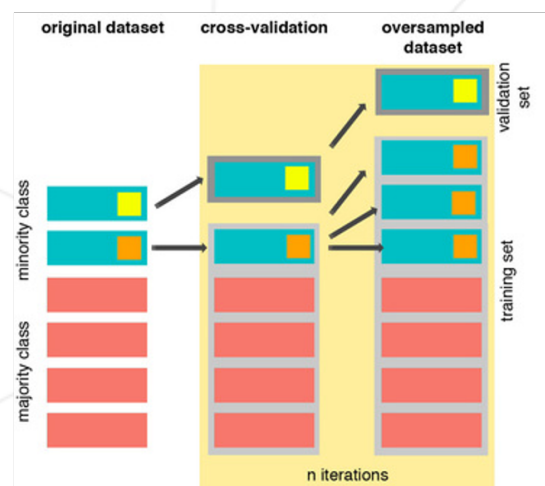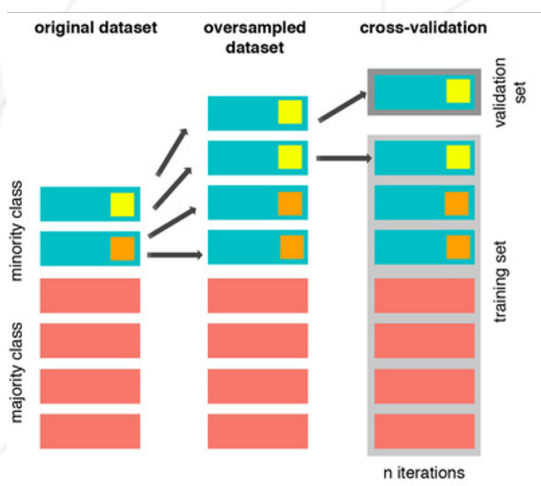
You should have something like this:

Chapter V

Exercise 02

![]	Exercise 02
	Data Augmentation
Turn-in directory : *ex02/*	
Files to turn in : `Medium02.ipynb`	
Allowed functions : `All`	

if you've done the first two parts well, you're sure to have understood that your data set is unbalanced. You'll have to rebalance it, but be careful how you go about it.

Chapter VI

Exercise 03

	Exercise 03
	Models
Turn-in directory : *ex03/*	
Files to turn in : `Medium02.ipynb`	
Allowed functions : `All`	

To avoid blocking transactions from legitimate customers, you will be required to have near-perfect accuracy.

Test with logistic regression, but personally the results are not precise enough.

```
Logistic Regression accuracy: 0.802
```

That's why we're going to have to use much more powerful models, so it's up to you to find out.

You must have an accuracy greater than 0.99 with your test set.
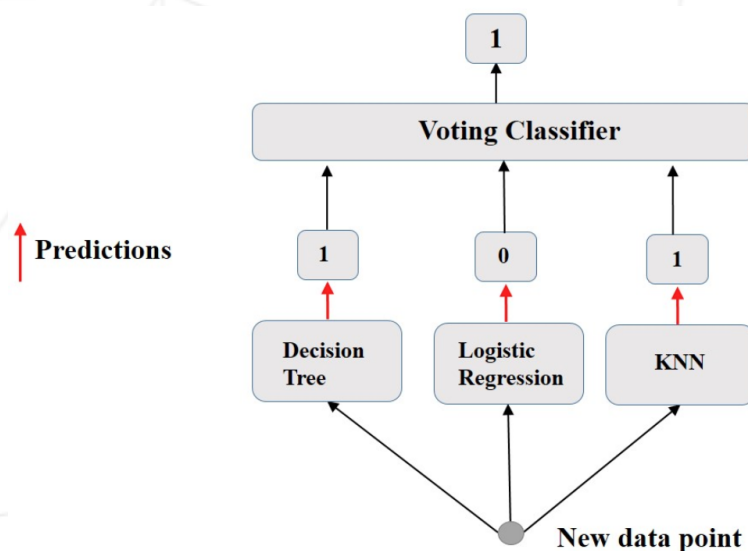
> You have to at least tested Decision Trees and XGBoost

Chapter VII

Bonus part

If you have some time, you can try to validate this little bonus. It is not mandatory, so don't waste too much time doing it.

Set up a vote classifier to improve your predictions.

Chapter VIII

Submission and peer-evaluation

- Create a `professional_training_medium` folder at the root of your home, and move around in it.

- Create a new `module02` folder and navigate to it.

> ⚠️ Please note, during your defense anything that is not present in the folder for the day will not be checked.