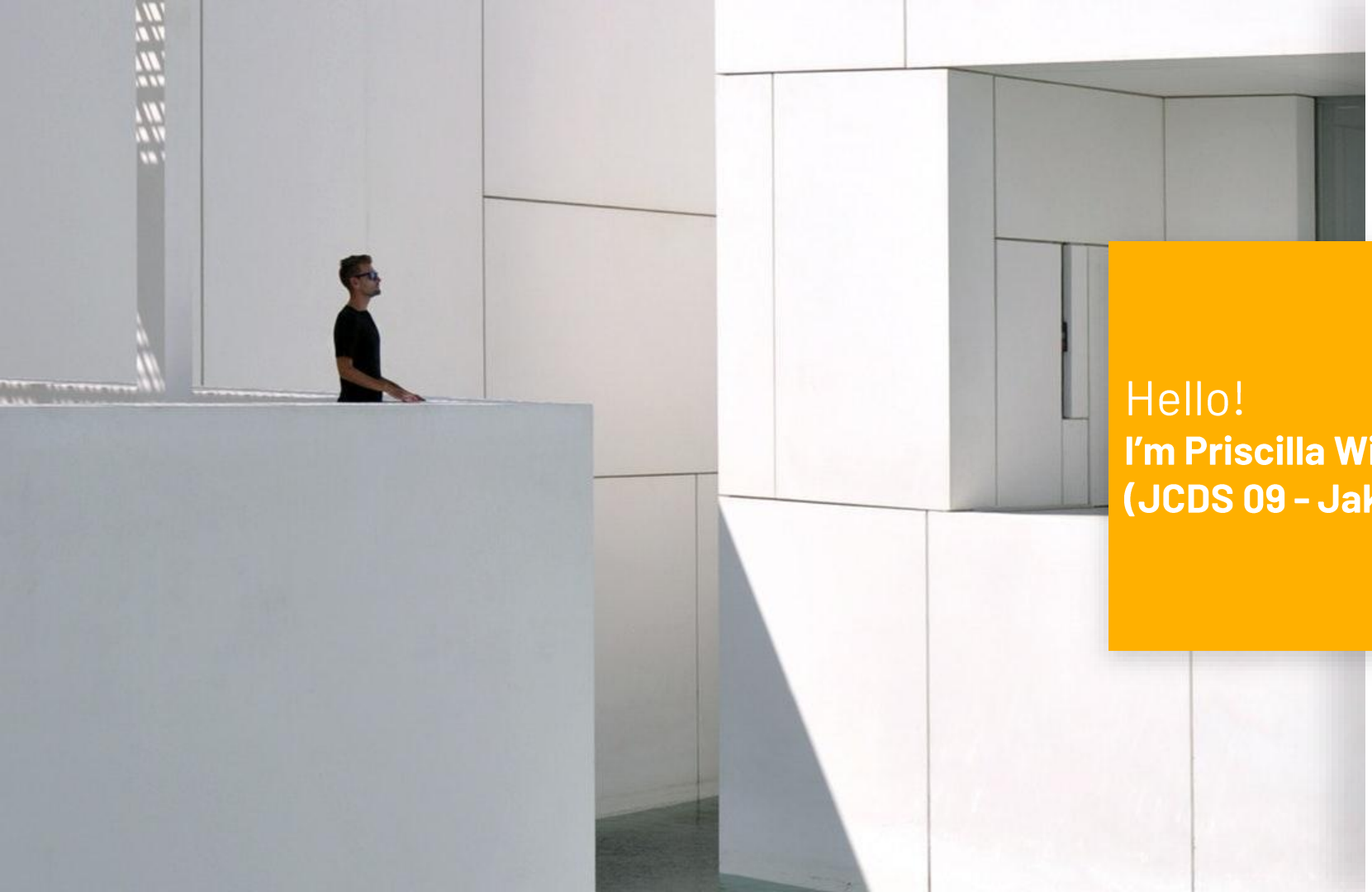
Term Deposit Subscription Prediction



Hello!

I’m Priscilla Widjaja
(JCDS 09 - Jakarta)

1. BUSINESS PROBLEM

Abstract

BANK MARKETING

Abstract

The data is related with direct marketing campaigns (phone calls) of a Portuguese banking institution. The classification goal is to predict if the client will subscribe a term deposit (variable y).

Problem

Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be ('yes') or not ('no') subscribed.

TERM DEPOSIT

- Created by deposits a certain amount of money in the bank for a preliminary fixed period of time.
- The bank is obliged to pay out certain interest at a definite time.
- With a fixed period for which the interest rate agreed between the bank and the client is valid.
- The interest rate is higher compared to that on a current account.
- The interest rate usually increases when the term of the deposit is enlarged.
- If the deposit is drawn earlier than the agreed period, the client usually loses the interest or a part of it.

TELEMARKETING

Contacting, qualifying, and canvassing prospective customers using telecommunications devices such as telephone, fax, and internet.

ADVANTAGES AND DISADVANTAGES OF TELEMARKETING

Advantages

- Allows to immediately gauge customer's level of interest.
- Provide a more interactive and personal sale service.
- Create an immediate rapport with your customers.
- Explain technical issues more clearly.
- Sell from a distance to increase your sales territory.
- Achieve results that are measurable.

Disadvantages

- Customer lists can be very costly.
- Telemarketing has a negative image that could damage your business' reputation - if carried out poorly.
- Training staff can be time-consuming and costly.
- **Often, more than one contact to the same client was required.**

2. ATTRIBUTE INFORMATION

Dataset : http://archive.ics.uci.edu/ml/datasets/Bank+Marketing#

BANK CLIENT DATA

Category	Descriptions	Data Type	Data Values
Age		Numeric	
Job	Type of Job	Categorical	'admin.', 'blue-collar', 'entrepreneur', 'housemaid', 'management', 'retired', 'self-employed', 'services', 'student', 'technician', 'unemployed'
Marital	Marital Status	Categorical	'divorced', 'married', 'single' Note: 'divorced' means divorced or widowed
Education		Categorical	'basic.4y', 'basic.6y', 'basic.9y', 'high.school', 'illiterate', 'professional.course', 'university.degree"

RELATED WITH THE LAST CONTACT OF THE CURRENT CAMPAIGN

Columns	Descriptions	Data Type	Data Values
Day of Week	Last Contact Day of The Week	Categorical	'mon','tue','wed','thu','fri'

OTHER ATTRIBUTES

Columns	Descriptions	Data Type	Data Values
Campaign	Number of contacts performed during this campaign and for this client	Numeric	Includes last contact
Pdays	Number of days that passed by after the client was last contacted from a previous campaign	Numeric	999 means client was not previously contacted
Poutcome	Outcome of the previous marketing campaign	Categorical	'failure','nonexistent', 'success'

SOCIAL AND ECONOMIC CONTEXT ATTRIBUTES

Columns	Descriptions	Data Type
Cons.conf.idx	Consumer Confidence Index - Monthly Indicator	Numeric
Nr.employed	Number of Employees - Quarterly Indicator	Numeric

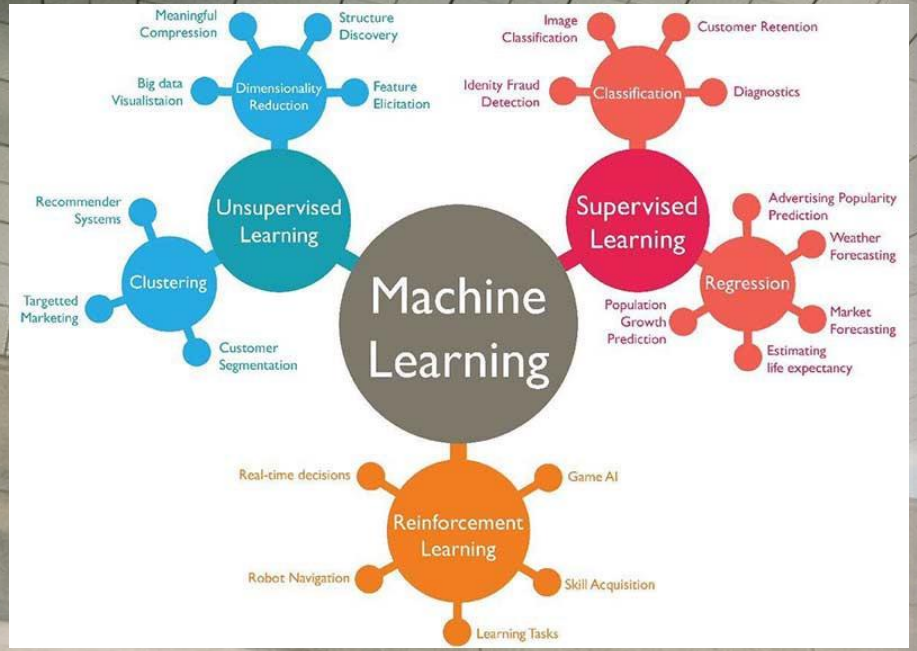
OUTPUT VARIABLE (DESIRED TARGET)

Column	Descriptions	Data Type	Data Values
y	Has the client subscribed a term deposit?	Binary	'yes', 'no'

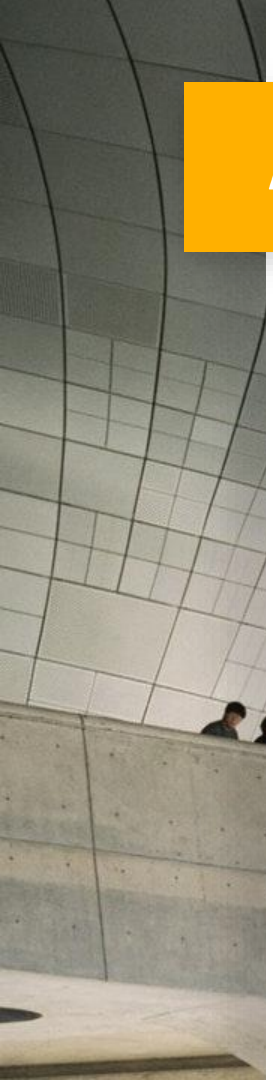
3. MACHINE LEARNING

Supervised : Classification

CLASSIFICATION



- The process of predicting class or category from observed values or given data points.
- Targets are also provided along with the input data set.
- Mathematically, it is the task of approximating a mapping function (f) from input variables (X) to output variables (Y).



ALGORITHMS

Logistic Regression

Used when the value of the target variable is categorical in nature. The data has binary output, so when it belongs to one class or another, or is either a 0 or 1.

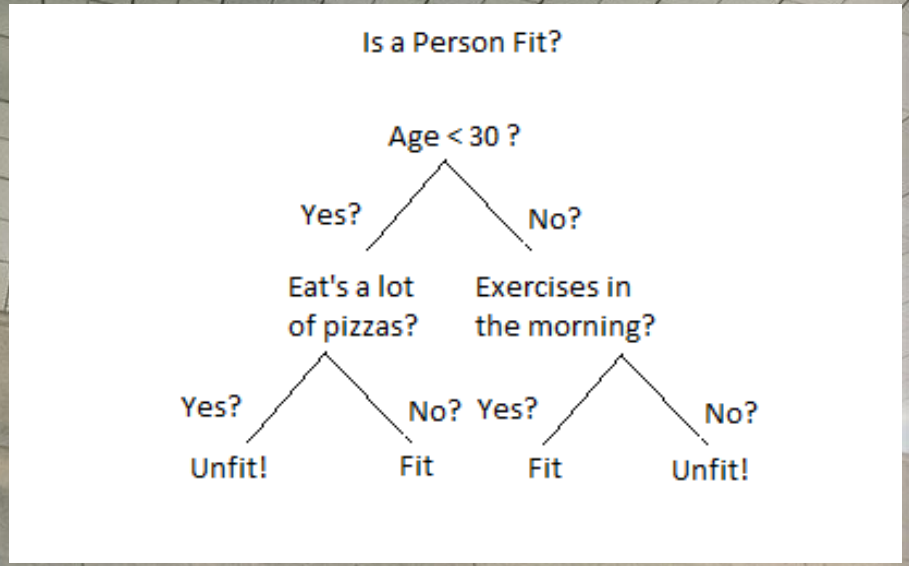
Random Forest

Creates decision trees on data samples. Gets the prediction from each of them. Selects the best solution by means of voting.

K Nearest Neighbors

An algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions).

DECISION TREE CLASSIFIER



Builds models in the form of a tree structure. It breaks down a data set into smaller and smaller subsets. The final result is a tree with decision nodes and leaf nodes.

A decision node has two or more branches. Leaf node represents a classification or decision.

The topmost decision node in a tree which corresponds to the best predictor called root node.

4. PROBLEM SOLVING

Building Model

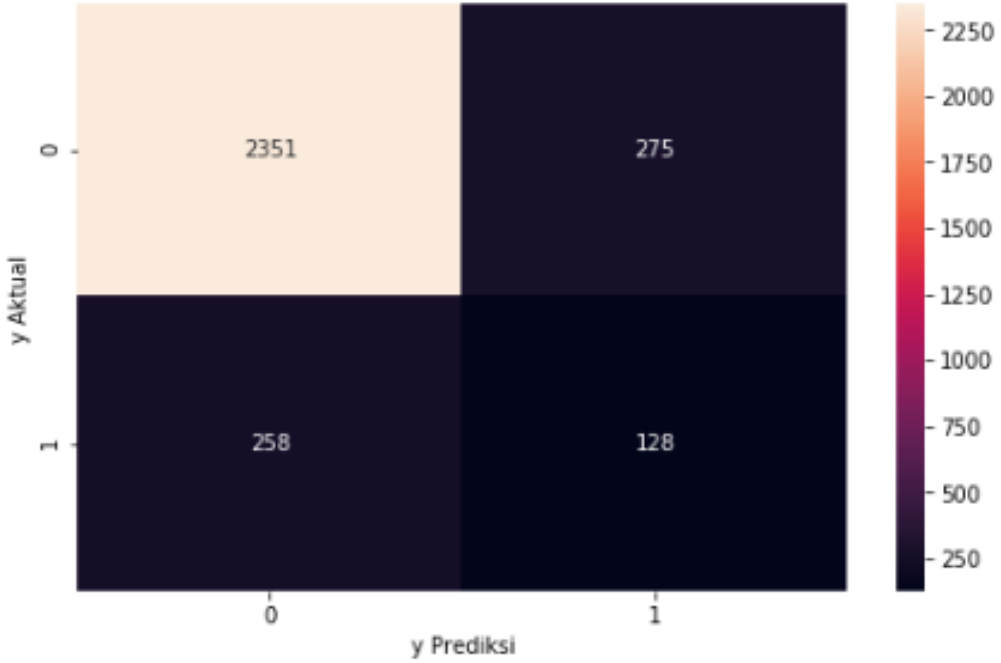
Step by Step

1. Load Dataset
2. Handling Missing Values
3. Handling Outliers
4. Encoding
5. Scaling
6. Features Selection
7. Splitting Dataset
8. Handling Imbalance Target
9. Building Machine Learning
10. Hyperparameter Tuning
11. Evaluating Model
12. Saving Model
13. Using Model

5. CONCLUSION

Evaluation : Decision Tree Classifier

	precision	recall	f1-score	support
0	0.90	0.90	0.90	2626
1	0.32	0.33	0.32	386
accuracy			0.82	3012
macro avg	0.61	0.61	0.61	3012
weighted avg	0.83	0.82	0.82	3012



DecisionTreeClassifier()

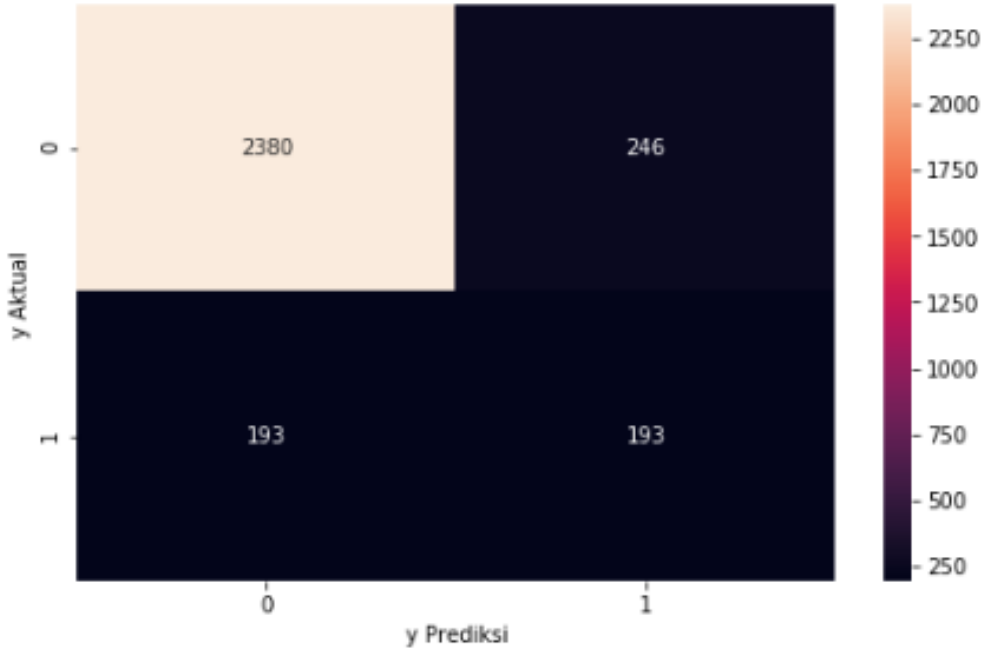
	Score (%)
accuracy	82.304117
recall	33.160622
precision	31.761787
roc_auc_score	61.344210
f1_score	32.446134

Before Hyperparameter Tuning

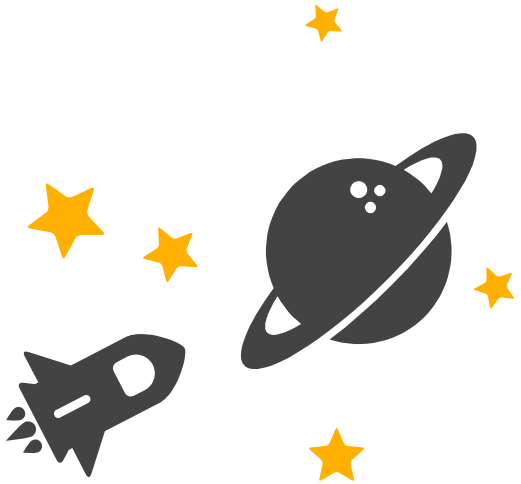
	Score (%)
accuracy	85.424967
recall	50.000000
precision	43.963554
roc_auc_score	70.316070
f1_score	46.787879

After Hyperparameter Tuning

	precision	recall	f1-score	support
0	0.92	0.91	0.92	2626
1	0.44	0.50	0.47	386
accuracy			0.85	3012
macro avg	0.68	0.70	0.69	3012
weighted avg	0.86	0.85	0.86	3012



Thank You



Any questions?

You can find me at roselinapriscilla.widjaja@gmail.com .