IJSE

Module	ITS 2130/ITS 2140-Assignment
Module leader	Suhash Rodrigo
Weighting:	80% for Part 1: Report on assignment 20% for Part 2: Presentation
Learning Outcomes Covered in this Assignment:	● Discuss, compare and contrast the advantages and disadvantages of applying a specific machine learning technique to a given learning task. ● Use a toolkit to develop a data mining application tailored to a given learning task. ● Effectively interpret the results of learning through an understanding of the strengths and limitations of data mining technology and the selection of an appropriate evaluation technique. ● Demonstrate knowledge of the state-of-the-art in data mining and an awareness of current areas of research. ● Apply and, where necessary, adapt an appropriate data mining technique to a given problem. ● Compare the operational concepts, implementation and performance issues of cloud computing systems, and the relative merits and suitability of each for complex data−intensive applications ● Critically appraise different cloud computing models, namely, infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS). ● Evaluate design choices when solving real−world cloud computing problems by analyzing and contrasting different cloud computing solutions. ● Integrate software components in novel ways to architect and develop cloud−based applications solutions for an enterprise. ● LSEPI - Security and ethics of cloud computing with public cloud solutions, Legal implications of off-premise hosted data, impact on society implications and benefits.
Handed Out:	7th March 2025
Due Date:	Part 1: Report on Assignment: 04th April 2025 Midnight. Part 2: Presentation: 05th and 06th April 2025. Each student will get 10 min time slots.
Expected deliverables	Part 1: One compressed electronic file containing Report and python notebook. Part 2: Working POC deployed in the cloud.
Method of Submission:	Google Classroom

Coursework Description

Use case: Bank Marketing Dataset

The dataset is related with direct marketing campaigns (phone calls) of a Portuguese banking institution. The goal is to predict if the client will subscribe a term deposit. The attributes information of the dataset as follows:

Input variables:
Bank Client Data:

1. - age (numeric)
2. - job : type of job (categorical)
3. - marital : marital status (categorical)
4. - education (categorical)
5. - default: has credit in default? (categorical: 'no','yes','unknown')
6. - housing: has housing loan? (categorical: 'no','yes','unknown')
7. - loan: has personal loan? (categorical: 'no','yes','unknown')

Related information with the last contact of the current campaign:

8. - contact: contact communication type (categorical: 'cellular','telephone')
9. - month: last contact month of year (categorical)
10. - day_of_week: last contact day of the week (categorical)
11. - duration: last contact duration, in seconds (numeric).

Important note: this attribute highly affects the output target (e.g., if duration=0 then y='no'). Yet, the duration is not known before a call is performed. Also, after the end of the call y is obviously known. Thus, this input should only be included for benchmark purposes and should be discarded if the intention is to have a realistic predictive model.

Other attributes:

12. - campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)
13. - 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)
14. - previous: number of contacts performed before this campaign and for this client (numeric)
15. - poutcome: outcome of the previous marketing campaign (categorical: 'failure','nonexistent','success')

social and economic context attributes:

16. - emp.var.rate: employment variation rate - quarterly indicator (numeric)
17. - cons.price.idx: consumer price index - monthly indicator (numeric)
18. - cons.conf.idx: consumer confidence index - monthly indicator (numeric)
19. - euribor3m: euribor 3 month rate - daily indicator (numeric)
20. - nr.employed: number of employees - quarterly indicator (numeric)

Output variable (desired target):

21. - y - has the client subscribed a term deposit? (binary: 'yes','no')

More details of the dataset can be found at https://archive.ics.uci.edu/ml/datasets/bank+marketing# and the dataset is provided with the assignment.

Produce Python Notebook, which performs the following task

A. Preprocess the dataset as specified in the data mining process.
    a. Handle Missing Values and Outliers if any
    b. If it is required, apply suitable feature coding techniques.
    c. Scale and/or standardized the features, produce relevant graphs to show the scaling/ standardizing effect.
    d.
B. Apply the following techniques to predict the value of Y for the test dataset (Training Dataset to Test Dataset is 0.8 to 0.2 ratio)
    a. Support Vector Machines (SVM)
    b. Logistic Regression (LR)
C. Discuss how significant your findings are.
D. Deploy the model in aws with a from when the user fill age, job,marital,education,default,housing,loan should display client subscribed a term deposit or not?
E. Describe the designed solution. Students should include the solution architecture diagram and the deployment architecture diagram into the report. In the solution diagram the student has to highlight the request flows and the data flows.
F. Select an appropriate hosting (EC2, Lambda , ECS)
G. Include the CI/CD pipeline designs diagrams and CI/CD process descriptions into the same report under CI/CD process section.

Coursework Marking scheme

The Coursework will be marked based on the following marking criteria:

Question	Marks	Marks	Comments
(A)	10		
(B)	10		
(C)	10		
(D)	20		
(E)	20		
(F)	20		
(G)	10		
Total	100		