THE UNIVERSITY OF HONG KONG

FACULTY OF ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE

FITE7410 Financial Fraud Analytics

Date: 9 December 2021 (WED) **Time: 6:30-8:30pm**

Answer ALL questions in SEPARATE ANSWER BOOK provided.

This is an open book examination. Candidates may bring to their examination any printed/written materials.

Only approved calculators as announced by the Examinations Secretary can be used in this examination. It is candidates' responsibility to ensure that their calculator operates satisfactorily, and candidates must record the name and type of the calculator used on the front page of the examination script.

QUESTION 1.

QUESTION 1.1.

Transaction ID	**Insurance Claimed Amount**	**Annual Income**	**Fraud**
1	350k	45k	YES
2	100k	40k	NO
3	130k	100k	NO
4	90k	140k	NO
5	100k	120k	NO
6	100k	200k	NO
7	120k	60k	YES
8	200k	120k	NO
9	90k	25k	YES
10	90k	40k	NO

Answer the following short questions:

a) Calculate the Gini Impurity of the (**Insurance Claimed Amount**)-split using (**Insurance Claimed Amount >= 100k**) as the spitting criteria.

b) Calculate the Gini Impurity of the (**Annual Income**)-split using (**Annual Income <= 50k**) as the spitting criteria.

c) Based on results of (a) and (b), which attribute should be used as the splitting attribute?

Show the steps of your calculation. NO marks will be given if only Gini Impurity values are written as the answers.

QUESTION 1.2.

What is Trade Based Money Laundering?

QUESTION 1.3.

Please explain what "Politically Exposed Person" is.

QUESTION 2.

			Actual	
			Fraud=Yes	**Fraud=No**
DT	**Decision Tree Prediction**	**Fraud=Yes**	38	9
		Fraud=No	12	241
LG	**Logistics Regression Prediction**	**Fraud=Yes**	41	1
		Fraud=No	9	249
RF	**Random Forest Prediction**	**Fraud=Yes**	45	4
		Fraud=No	5	246
NN	**Neural Network Prediction**	**Fraud=Yes**	43	5
		Fraud=No	7	245

You are the fraud detection expert of the credit card company and you have developed 4 models with the confusion matrix results listed in the above table. From the business perspective, the fraud detection model must satisfy the following criteria:

a) Have a recall rate of at least 85%

b) Have a false positive rate lower than 10%

Answer ALL questions in SEPARATE ANSWER BOOK provided.

(a) Write down the formula of calculating these 2 criteria

- Recall rate
- False Positive Rate

Write down the steps and results of calculating the classification metrics for each of the 4 models.

	DT	**LG**	**RF**	**NN**
Recall Rate				
False Positive Rate				

NO marks will be given if only calculated values are written as the answers.

(b) Base on the results of (a) above, which model is preferred for the company? Justify your answer.

(c) Sometimes stopping the use of a credit card whenever there is an alert of potential fraud case raised by the fraud detection model is not desirable by the credit card company. From the perspective of the credit card company, the business cost of false alarms is 5 times higher than a false negative result. In addition, management of the credit card company would like you to explain how the fraud detection model works. In this case,

1/ What additional criteria or changes that you might make for the model selection? List the new selection criteria and state your reasons for the changes made.

2/ With the new selection criteria (1) above, which fraud detection model you would choose? Justify your answer.

(d) From the confusion matrix, what are the possible problems with this dataset? Suggest solutions for handling these problems.

QUESTION 3.

Case study

Case Background

A whistle blower complaint was received by the Board of a US pharmaceutical company targeting at the senior management of their China operation.

There were allegations of bribery and conflict of interests among the senior management.

Please provide an Action Plan setting out how you would investigate the potential irregularities/ issues highlighted above based on what you have learned in fraud investigation.

Please also explain what technology you would use to assist with the investigation, e.g. data analytics and/or any other tools.

=== END OF PAPER ===