**Full Name (as shown in attendance sheet): KAR CHAUDHURI ANIRBAN**

**[Optional] LumiNUS Account:**

**[Optional] NRIC / Passport / NUS Matriculation No.: A0108517H**

(Select and provide the applicable)

**Institute of Systems Science**

**National University of Singapore**

**GRADUATE CERTIFICATE**

**INTELLIGENT REASONING SYSTEMS**

**Assessment**

**Subject: *Reasoning Systems***

SECTION B

|  |  |
| --- | --- |
| **Question** | **Marks** |
| **2** | **/1** |
| **3** | **/2** |
| **TOTAL** | **/3** |

**Instructions for Paper**

Duration: Sixty minutes exam

This is an *OPEN BOOK* examination. This examination paper consists of *one* Section and *two* Questions. You are to answer *ALL* questions. There are a total of *3 Marks* for this paper.

1. Read **ALL** instructions before answering any of the examination questions.
2. Write your Student ID number on the **front page** of this examination paper in the box provided.
3. This is an **Open Book** examination. If you wish, you may use reference materials to answer a question. Reference materials can be *books, manuals, handouts* or *notes*.
4. Answers are to be written **only** in this **examination paper** and any **attachments** provided and will be considered for credit. Answers written in any appendices will **NOT** be marked.
5. Use a pen for writing your answers. Pencil may only be used for drawing diagrams and writing program code.
6. Non-programmable calculators may be used if required. **However, computers of any form (laptops, tablets, smart watches etc.) are not permitted to be brought into the examination hall.**
7. State clearly any assumptions you make in answering any question where you feel the requirement is not sufficiently clear.
8. At the end of the examination:
9. Hand-in the examination paper for **each** section **separately**, any appendices and attachments.
10. You are **not** allowed to remove the examination paper, appendices or attachments from the examination hall.

***REMEMBER:***

***This is an OPEN BOOK exam.***

***There are a total of 3 Marks for this paper.***

***You are required to answer ALL questions.***

***State clearly any assumptions you make in answering any question where you feel the requirement is not sufficiently clear.***

**SECTION B**

**Question 2** *(Total: 1 Mark)*

Relate to your **own** workplace and profession, **Answer the following questions:**

* 1. **How/Where** would you apply the learnt in your workplace? (max 300 words)

*For example: which specific area to perform more efficiently using those learnt, etc.*

*(0.5 Mark)*

[Answer]

I am working with fraud detection involving e-commercial purchases on an online commercial purchasing platform. I use logistic regression and decision tree algorithms for machine learning who coefficients enable me to identify the important predictor variables and generate if-else rules. Some of the important factors influencing probability of fraud are age, income, hour of purchase, etc. One must look at deviations from normal underlying data patterns. The model is trained on training (familiar) data and tested on test data and this involves cross validation. Its performance is optimized by fine-tuning hyperparameters using Genetic algorithms or random search techniques.

The credit transactions data are collected real time on Azure cloud, ingested into Synaspses data warehouse for storage and analytics purposes involving dashboard visualization. The information obtained and analysed is real time.

* 1. What **business values** can be derived? (max 300 words)

*For example: prepare oneself and the organization to apply what new techniques, etc.*

*(0.5 Mark)*

[Answer]

I will be able to interpret the most important predictor variables influencing if a transaction is fraud or not like hour of purchase, amount, income of buyer involving decision trees and logistics regression algorithm coefficients.

The benefits are that I can get real-time rapid insights of fraudulent cases and their underlying factors and able to alert customers I am able to share insights to C-level management who are able to craft precautionary security and awareness based policies to ensure safe online commercial purchasing.

**Question 3** *(Total: 2 Marks)*

You have been employed by the ***NGL (NoGasLeak) Chemical Inc.*** to build a hybrid reasoning system to determine what emergency actions to take when a gas leak occurs in their chemical plant. You interview the world's top experts on gas leaks from chemical plants to acquire knowledge for the intelligent reasoning system. However, the experts can only provide a set of guidelines for a general gas leak situation and nothing specific to the ***NGL*** chemical plant. Examples of these guidelines are shown below:

***If*** *health risk category is A and gas dispersion radius is low*

***Then*** *no action is needed*

***If*** *health risk category is B and gas dispersion radius is medium*

***Then*** *alert Civil Defence*

***If*** *health risk category is C and gas dispersion radius is high*

***and***

*immediate danger zone is urban*

***Then*** *evacuate immediate danger zone population*

The variables used by these guidelines (e.g. health risk category, gas dispersion radius, immediate danger zone) cannot be measured directly but must be inferred from data obtained from the specific gas leak situation, e.g. the plant database, sensor data, and visual observations at the scene of the gas leak. When you ask the experts how they perform this inference, they are unable to give an explanation saying that it is an intuitive process that is hard to explain. Instead they offer to examine a database of 300 records containing data recorded during past ***NGL*** gas leaks (plant sensor data, wind speed, wind direction and other weather measurements and various visual observations) and then indicate what the associated values of health risk category, gas dispersion radius, danger zone etc. are for each record.



<http://www.trainingfordisastermanagement.com/wp-content/media/images/industrial/slider/industrial6.jpg>

**Answer the following questions:**

* 1. Draw a **sub-system level** block diagram (process flow/diagram) of the hybrid reasoning system. Annotate each sub-system/module with the appropriate technique to be used, e.g. various search, constraint solver module, simulation, data mining, genetic algorithms module, association rule module, knowledge/rule-based inference, decision and process automation, etc. Or design the system in the context of KIE suite.

*(1 Mark)*

[Answer]

I propose a hybrid cooperating system.

Analyst

Data mining

Data collection & preparation (Azure data synapses warehouse)

Business decision and solutions involving if-else rules

Deployment of finalised model

Model hyperparameter finetuning: genetic algorithms and heuristic search optimising

Model selection: classification report and confusion matrix of predictions

Decision Tree and logistic regression (python sklearn)

* 1. Illustrate and describe the system data flow (from start to end) in the scenario: ***evacuate immediate danger zone population***.

*(0.5 Mark)*

[Answer]

**Rules:**

***If*** *health risk category is A and gas dispersion radius is low*

***Then*** *no action is needed*

***If*** *health risk category is B and gas dispersion radius is medium*

***Then*** *alert Civil Defence*

***If*** *health risk category is C and gas dispersion radius is high*

***and***

*immediate danger zone is urban*

***Then*** *evacuate immediate danger zone population*

1. **Data Collection:** Data collected from sensors and manual surveys, ingested into Postgresql/mysql relational database
2. **Variables:** wind speed, wind direction, humidity, temperature, rainfall
3. Data mining, understanding of underlying distribution and feature selection done using statistical tests like ANOVA, Chi-Square, filter based criteria, stepwise regression using python libraries like statsmodels, scipy and sci-kit learn.
4. Reason for feature selection: Avoid overfitting due to curse of high dimensionality
5. Apply logistic regression and decision tree classification models that are interpretable
6. Choose between logistic regression and decision tree based on performance metrics like recall and precision
7. Data visualisation, model prediction interpretations helped by visualisation tools like Tableau connected to a postgresql database
8. Interpretations of model translated into above stated if-else rules to identify immediate danger zone population
   1. After your reasoning system is fielded, it is noticed that for a small proportion of gas leaks, the advice it offers is inappropriate. In these situations ***NGL*** falls back on human judgment to determine the correct action. These correct actions and all data regarding these situations are logged. Suggest how your reasoning system can be updated to take into account this new information.

*(0.5 Mark)*

[Answer]

**END OF ASSESSMENT PAPER**