(Knowledge for Development)

KIBABII UNIVERSITY

UNIVERSITY EXAMINATIONS
2020/2021 ACADEMIC YEAR

YEAR FOUR SEMESTER two EXAMINATIONS

FOR THE DEGREE OF
(COMPUTER SCIENCE)

COURSE CODE : CSC 423

COURSE TITLE : MACHINE LEARNING

DATE: 24/09/2021 TIME: 09:00 A.M – 11:00 A.M

INSTRUCTIONS TO CANDIDATES

ANSWER QUESTIONS ONE AND ANY OTHER TWO

QUESTION ONE (COMPULSORY) **[30 MARKS]**

a)

i. Describe Machine Learning and draw appropriate diagram **[4 Marks]**

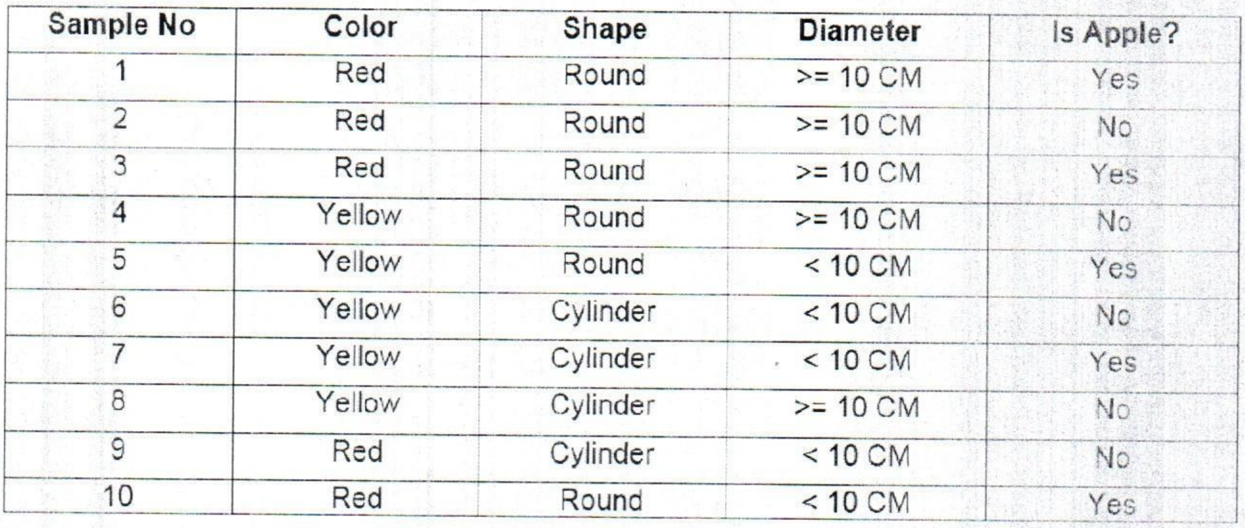
ii. Describe the **THREE** Main Problems Solved by Machine Learning **[6 Marks]**

iii. Distinguish between between Machine Learning Algorithms and Traditional Rule-Based Algorithms **[4 Marks]**

iv. Explain the Necessity of Feature Selection **[2 Marks]**

b)

i. Use the table to answer the question that Follow **[6Marks]**



Sample No	Color	Shape	Diameter	Is Apple?
1	Red	Round	>= 10 CM	Yes
2	Red	Round	>= 10 CM	No
3	Red	Round	>= 10 CM	Yes
4	Yellow	Round	>= 10 CM	No
5	Yellow	Round	< 10 CM	Yes
6	Yellow	Cylinder	< 10 CM	No
7	Yellow	Cylinder	< 10 CM	Yes
8	Yellow	Cylinder	>= 10 CM	No
9	Red	Cylinder	< 10 CM	No
10	Red	Round	< 10 CM	Yes

Calculate each conditional probability using naïve bayes:

***p*(Colour = Red | Apple = Yes) =**

***p*(Colour = Red | Apple = No) =**

***p*(Shape = Round | Apple = Yes) =**

p*(Shape = Round | *Apple = No) =

p(Diameter = > 10 cm | Apple = Yes) =

p(Diameter = > 10 cm | Apple = No) =

Multiply the conditional probabilities as follows:

p(Color = Red | Apple = Yes) X p(Shape = round | Apple = Yes) X p(Diameter = > 10 cm | Apple = Yes) =

p(Color = Red | Apple = No)p(Shape = round | Apple = No)p(Diameter => 10 cm | Apple = No) =

Compare p(CY | x) to p(CN | x):

$$\text{If } \frac{p(C_Y|\mathbf{x})}{p(C_N|\mathbf{x})} > 1 \therefore \mathbf{x} \in C_Y \text{ , else } \mathbf{x} \in C_N$$

ii. Describe Overall Procedure of Building a Model **[6 Marks]**

iii. Differentiate between underfitting and overfitting **[2 Marks]**

QUESTION TWO [20 MARKS]

a) Compare the three gradient descent methods Batch Gradient Descent (BGD), Stochastic Gradient Descent (SGD) and Mini-Batch Gradient Descent (MBGD) **[5 Marks]**

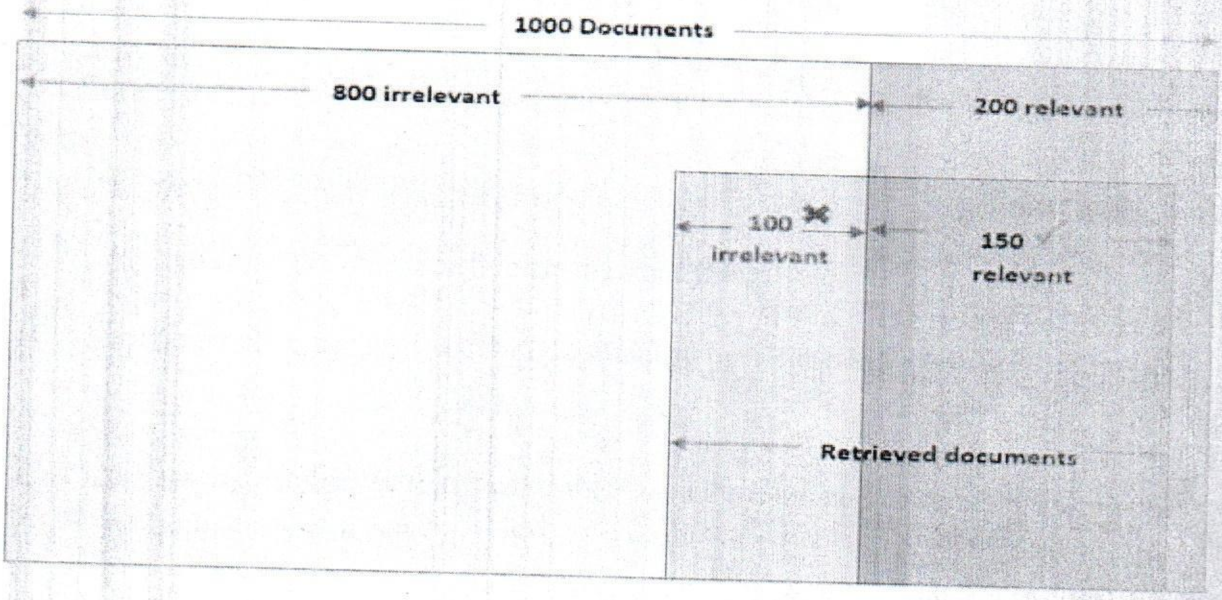
b) State and Explain the Common model hyperparameters **[5 Marks]**

c) Use the information to answer the Question that Follow **[10 Marks]**

Assume that you developed a search algorithm that helps you to retrieve related words from a corpus that contains 1000 documents. From these 1000 documents, assume 200 are relevant to the word cat, and the other 800 documents are irrelevant.

You ran a search test for the word "cat".

After the test ran, the search engine retrieved the documents that are shown here.



You test your solution by searching for the word "cat". Your algorithm returns 250 documents, where 150 documents are relevant (which means your algorithm missed 50 relevant documents) and 100 documents are irrelevant (which means your algorithm correctly eliminated 700 of the irrelevant documents). Use the information to identify:

- How many relevant documents were retrieved by the algorithm? True positive (Tp).
- How many irrelevant documents were retrieved by the algorithm? False positive (Fp)
- How many relevant documents did the algorithm not retrieve? False negative (Fn).
- How many irrelevant documents did the algorithm not retrieve? True negative (Tn).

i. Calculate the Accuracy of the algorithm

ii. Calculate the precision of the algorithm

iii. Calculate the Recall of the algorithm

iv. Calculate the F-Score measure of the algorithm

QUESTION THREE [20 MARKS]

Determine if the following statements are **TRUE/ FALSE** **[6 Marks]**

i. Artificial intelligence is a new technical science that studies and develops theories, methods and application systems for simulating, extending and extending human intelligence. It is one of the core research areas of machine learning.

ii. The word recognition in the speech recognition service refers to the synchronous recognition of short speech. Upload the entire audio at once, and the recognition result will be returned in the response

iii. Self-encoder is an unsupervised learning algorithm

iv. Loss function and model function are the same thing.

v. The commonly used functions for mathematical operations in Python are basically in the math module and the cmath module.

vi. The Python dictionary is identified by "{}", and the internal data consists of the key and its corresponding value.

a) Describe Ensemble learning **[4 Marks]**

b) Describe the classification of ensemble learning **[4 Marks]**

c) Describe reinforcement learning **[6 Marks]**

QUESTION FOUR [20 MARKS]

a) Describe the importance of Convergence in Gradient Descent **[4 Marks]**

b) Describe the following **THREE** concepts used machine learning

i. Dataset

ii. Training Set

iii. Test Set **[6 Marks]**

c) i. Describe K-means Clustering **[4 Marks]**

ii. Describe **THREE** Examples of application using K-means clustering **[6 Marks]**

QUESTION FIVE [20 MARKS]

a) Describe a support vector Machine. Use appropriate diagram where necessary. **[4 Marks]**

b) Describe the Importance of Pruning in Decision Tree. Give appropriate python Code to show Pruning. **[4 Marks]**

c) Describe the following THREE modules are important and show how they are imported in python

i. Pandas

ii. Math

iii. NumPy **[6 Marks]**

d) Describe the Importance of the Python Code below **[6 Marks]**

```
def calculate Entropy(df):

if algorithm == 'Regression':
return0
rows = df.shape[0]
decisions = df['Decision'].value_counts().keys().tolist()
entropy = 0
for i in range(0, len(decisions)):
num_of_decisions = df['Decision'].value_counts().tolist()[i]

class_probability = num_of_decisions / rows

entropy = entropy -class_probability * math.log(class_probability, 2)
    return entropy
```