B.Tech DEGREE EXAMINATION, DECEMBER 2023

Fifth & Sixth Semester

18CSE307T - MACHINE LEARNING IN DRUG DISCOVERY

(For the candidates admitted during the academic year 2020 - 2021 & 2021 - 2022)

Note:

i. **Part - A** should be answered in OMR sheet within first 40 minutes and OMR sheet should be handed over to hall invigilator at the end of 40th minute.

ii. **Part - B** and **Part - C** should be answered in answer booklet.

Time: 3 Hours

Max. Marks: 100

PART - A (20 × 1 = 20 Marks)
Answer **all** Questions

		Marks	BL	CO

1. Which of the following is not a carbohydrate?   (1, 1, 1)
   (A) Glucose   (B) Sucrose
   (C) Galactose   (D) Glycine

2. Amino acids are joined by   (1, 1, 1)
   (A) Phosphodiester bond   (B) Peptide bond
   (C) Glycosidic bond   (D) Hydrogen bond

3. Which of the following purine bases is present in RNA?   (1, 1, 1)
   (A) Adenine   (B) Uracil
   (C) Guanine   (D) Cytosine

4. How many essential amino acids are there: (1, 1, 1)
   (A) 25   (B) 20
   (C) 19   (D) 18

5. What is the first step in the X – ray crystallography of proteins?   (1, 1, 2)
   (A) Creating electron density map   (B) X-ray diffraction
   (C) Precipitation of protein   (D) Purification of protein

6. Which of the following statements best describes a lead compound?   (1, 1, 2)
   (A) A compound that contains the   (B) A compound from the research
       element lead   laboratory.
   (C) A molecule that shows some   (D) The first compound of a structural
       activity or property of interest and   class of compounds to reach the
       serves as the starting point for the   market.
       development of a drug

7. Which of the following terms is used to describe a drug that has the same effect on a receptor as the endogenous chemical messenger?   (1, 1, 2)
   (A) Agonist   (B) Antagonist
   (C) Partial agonist   (D) Inverse agonist

8. Alignment procedure that tries to align the entire sequence: (1, 1, 2)
   (A) Multiple sequence alignment (B) Pairwise alignment
   (C) Local alignment   (D) Global alignment

9. Write the correct biopython code   (1, 1, 3)

   ```
   >>> protein_seq = Seq("EVRNAK")
   >>> dna_seq = Seq("ACGT")
   >>> protein_seq + dna_seq
   ```
   (A) Evrnakacgt   (B) EVRNAKACGT
   (C) EvrnakaCGT   (D) EVRNnakacgt

10. What will be the output of the following Python function?
len(["hello".2. 4. 6])
(A) Error   (B) 6
(C) 4   (D) 3

        1   1   3

11. Write the biopython output for the code
my_seq = Seq("GATCGGGG")
len(my_seq)
(A) 7   (B) 8
(C) 9   (D) 10

        1   1   3

12. Write the biopython output for the code
Seq("AAAA").count("AA")
(A) 2   (B) 3
(C) 4   (D) 5

        1   1   3

13. NumPY stands for
(A) Numbering python   (B) Number in python
(C) Numerical python   (D) Number Problem

        1   1   4

14. What will be the number of arguments require to pass the pandas series
(A) 2   (B) 3
(C) 4   (D) 1

        1   1   4

15. What is Zero function in NumPy used to
(A) make a matrix with first column 0   (B) Make a matrix will all elements 0
(C) make a matrix with diagonal   (D) make a empty matrix
    elements

        1   1   4

16. Which of the following machine learning algorithm is based on idea of bagging
(A) Decision tree   (B) Random forest
(C) PCA   (D) Genetic Algorithm

        1   1   4

17. Machine learning is a subset of
(A) IOT   (B) Artificial intelligence
(C) Data science   (D) Data learning

        1   1   5

18. Most significant phase in genetic algorithm is
(A) mutation   (B) selection
(C) crossover   (D) fitness function

        1   1   5

19. Write the biopython output for the code
my_seq = Seq("GATCGGGG")
my_seq[:]
(A) Seq('G')   (B) Seq('GATCGGGG')
(C) Seq('GA')   (D) error

        1   1   5

20. Write the biopython output for the code
dna_seq = Seq("acgtACGT")
dna_seq.upper()
(A) Seq('ACGTACGT')   (B) Seq('ACCGT')
(C) Seq('ACGTAT')   (D) Seq('ACGGT')

        1   1   5

PART - B (5 × 4 = 20 Marks)
Answer any 5 Questions

Marks BL CO

No.	Question	Marks	BL	CO
21.	Explain amino acid classification based on R group?	4	1	1
22.	Explain the mechanism of enzymes and enzyme inhibitors?	4	2	2
23.	Explain the principle behind X ray crystallography technique?	4	1	2
24.	With a neat flow chart explain the lead identification strategy.	4	1	2

		Marks	BL	CO
25.	Explain the concept of arrays in Python?	4	1	3
26.	Mention the importance of boosting method in ML algorithm.	4	2	4
27.	How is a biological dataset is constructed for ML analysis?	4	1	5

PART - C (5 × 12 = 60 Marks)
Answer all Questions

		Marks	BL	CO
28.	(a) Explain with neat diagrams the structural classification of proteins?	12	1	1
	(OR)			
	(b) Explain the structural features of DNA and RNA with a neat diagram?			
29.	(a) Explain various scoring function used by docking algorithms?	12	1	2
	(OR)			
	(b) Explain the process of pairwise and multiple sequence alignment with a suitable example.			
30.	(a) Explain the the role of python in functional genomics and its application ML.	12	1	3
	(OR)			
	(b) Explain the importance of Pandas with suitable example?			
31.	(a) Explain the importance of Random forest algorithm in drug discovery?	12	1	4
	(OR)			
	(b) Explain the role of SVM in deterring new lead compounds discovery.			
32.	(a) Why we need to split the dataset to generate better ML model, explain in detail?	12	1	5
	(OR)			
	(b) Write the importance of biological databases in detail, and how it will be used for ML model construction.			

* * * * *

08DF5&6-18CSE307T