32. a. Explain data frames in PANDAS with example. 12 3 3 2

(OR)

b. Explain 'BBB' dataset preparation, pre-processing and feature generation using ML. 12 3 3 2

* * * * *

Reg. No.														

B.Tech. DEGREE EXAMINATION, JUNE 2023
Sixth Semester

18CSE307T – MACHINE LEARNING IN DRUG DISCOVERY
(For the candidates admitted from the academic year 2018-2019 to 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 & 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	PO
1. The basic building block of polysaccharides are (A) Glucose (B) Amino acid (C) Fatty acid (D) NADH	1	1	1	1
2. The process of making a exact copy of DNA in a cell is (A) Translation (B) Transcription (C) Replication (D) Polymerization	1	1	1	1
3. The Ramachandran plot for protein 2° structure prediction is based on (A) PSI and PHI angle (B) Torsion angle (C) Degree of rotation (D) Planar rotation	1	2	1	1
4. Which of the following is not true about enzymes? (A) Enzymes are proteins (B) Enzymes act as catalyst (C) Enzymes reduce activation energy (D) Enzymes are converted to product	1	2	1	1
5. What is termed as a lead compound? (A) A compound that contains element lead (B) A compound from the research laboratory that is chosen to go forward for preclinical and clinical trails (C) A molecule that shows some activity of property of interest and serves as the starting point of a drug (D) The first compound of a structural class of compound to reach the market	1	2	2	2
6. Which bond is primarily stabilizing the drug binding to target protein? (A) Vander Walls interaction (B) Hydrogen bond (C) Ionic bond (D) Hydrophobic interaction	1	2	2	2
7. Expand ADME (A) Affinity, dosage, marketing, efficacy (B) Absorption, distribution, metabolism, excretion (C) Agonism, dependence, mobility, efficiency (D) Antagonism, deficiency, mean, efflux	1	1	2	2

8.	Which of the following is a protein sequence database? (A) DD BJ (B) EMBL (C) Genbank (D) PIR	1	1	2	2
9.	Which of the following is a sequence alignment tool? (A) BLAST (B) Print (C) Prostite (D) PIR	1	1	2	2
10.	The term given for molecule that has no biological activity but resembles drug molecule is (A) Dummy drug (B) Lead (C) Placebo (D) Peptide	1	1	2	2
11.	What is the maximum length of a python identifier (A) 32 (B) 16 (C) 128 (D) No fixed length is specified	1	3	3	2
12.	Which of the following types of loops are not supported in python? (A) For (B) While (C) Do-while (D) If	1	2	3	2
13.	Which python library is similar to PANDAS? (A) NPY (B) RPY (C) Numpy (D) SPY	1	2	3	2
14.	ndim can be used for (A) Finding the dimension of the array (B) Size of the array (C) Weight of the array (D) Operational activities of array	1	2	3	2
15.	Which of the following statement for creating data frames is valid? (A) df = Pd.dataframe (dict 1) (B) df = Pd.Dataframe (dict 1) (C) df = Pd.dataFrame (dict 1) (D) df = Pd.DataFrame (dict 1)	1	3	3	3
16.	Machine learning is an application of (A) Block chain (B) Artificial intelligence (C) Python (D) Numpy	1	1	4	2
17.	Which among the following algorithms are not used in machine learning? (A) Naïve bayes (B) Support vector machines (C) K-nearest neighbors (D) Linear regression	1	1	4	2
18.	In random forest the memory requirement for storage process is _____. (A) High memory (B) Low memory (C) Random memory (D) Optimal memory	1	2	3	2
19.	In SVM, if the number of input features is 3, then the hyper plane is (A) Line (B) Circle (C) Plane (D) Square	1	2	3	2

20.	In SVM what is a hyper plane? (A) Decision boundaries (B) Data point (C) Features (D) Nodes	1	3	3	2

PART – B (5 × 4 = 20 Marks)
Answer **ANY FIVE** Questions

		Marks	BL	CO	PO
21.	With neat diagram explain secondary structure of protein.	4	2	1	1
22.	Explain the various structural features of a gene.	4	2	1	1
23.	Discuss various RNA molecule with its features.	4	2	1	1
24.	Discuss the features of connection table.	4	2	2	1
25.	Discuss string function in python for nuclic acid sequence analysis.	4	2	3	2
26.	Discuss the uses of Numpy.	4	2	2	1
27.	Discuss commonly used data structures in PANDAS.	4	2	3	1

PART – C (5 × 12 = 60 Marks)
Answer **ALL** Questions

		Marks	BL	CO	PO
28. a.	Discuss the classification of proteins with neat diagram.	12	2	1	1
	(OR)				
b.	Explain enzyme inhibition mechanism.	12	2	1	1
29. a.	With a flow chart explain the process involved in lead identification.	12	2	2	2
	(OR)				
b.	Explain various scoring function used in docking algorithm.	12	2	2	2
30. a.	Discuss the following (i) Rule of five representing smiles notation (ii) Role of python in functional genomics	12	2	3	1
	(OR)				
b.	Discuss the following (i) Needleman-Wunch algorithm (ii) Progressive alignment in MSA	12	2	3	1
31. a.	Distinguish supervised and unsupervised learning using SVM with an example.	12	2	3	1
	(OR)				
b.	With a flow chart, explain random forest algorithm for drug discovery.	12	2	3	1