Name: Enrolment No:	

UPES

End Semester Examination, December 2024

Set 1

Course: Artificial Intelligence and Bioinformatics — **Semester: 1st**

Program: MSc. (Microbiology) and (Nutrition and Dietetics) — **Duration: 3 Hours**

Course Code: CSAI7021 — **Max. Marks: 100**

Instructions: Read all questions carefully.

S. No.	Section A Short answer questions/ MCQ/T&F (20Qx1.5M= 30 Marks)	Marks	COs
Q 1	**When was Python first released?** A) 1991 B) 2000 C) 1989 D) 1995	1.5	CO2
Q 2	**Firmware is:** A) Temporary hardware B) Software programmed into read-only memory C) A type of web application D) None of the above	1.5	CO2
Q 3	**Which version of Python introduced the `print()` function?** A) Python 1 B) Python 2 C) Python 3 D) Python 4	1.5	CO1
Q 4	**Which of the following is NOT a Python IDE?** A) PyCharm B) Eclipse C) VS Code D) Notepad++	1.5	CO1
Q 5	**Which command is used to install Python packages?** A) install_package B) pip install C) Python load D) pkg add	1.5	CO1
Q 6	**What function is used to convert data from one type to another in Python?** A) `convert()` B) `cast()`	1.5	CO1

	C) `str()` D) `int()`		
Q 7	**Which of the following statements is correct for multiple assignments in Python?** A) `x, y = 5, 10` B) `x, y := 5, 10` C) `x := 5, y := 10` D) `assign(x, 5), assign(y, 10)`	**1.5**	**CO1**
Q 8	**Which of the following is an immutable data type in Python?** A) List B) Set C) Tuple D) Dictionary	**1.5**	**CO2**
Q 9	**How do you create a dictionary in Python?** A) `my_dict = {key1: value1, key2: value2}` B) `my_dict = {key1, key2}` C) `my_dict = dict([key1, value1], [key2, value2])` D) `my_dict = dict((key1, value1), (key2, value2))`	**1.5**	**CO2**
Q 10	**Which loop is generally used when the number of iterations is not known beforehand?** A) `for` loop B) `while` loop C) `do-while` loop D) `foreach` loop	**1.5**	**CO2**
Q 11	**What is homology modeling?** A. Predicting a protein structure from scratch B. Predicting a structure using known structures of similar sequences C. Predicting secondary structures only D. Predicting the DNA sequence of a gene	**1.5**	**CO5**
Q 12	**Unsupervised Learning is used for:** A. Predicting future values B. Clustering and pattern recognition C. Classification problems D. Reinforcement tasks	**1.5**	**CO3**
Q 13	**Which type of learning uses a reward-based system to improve performance?** A. Supervised Learning B. Unsupervised Learning C. Reinforcement Learning D. Self-supervised Learning	**1.5**	**CO4**
Q 14	**Self-supervised Learning mainly relies on:** A. Labeled data B. Partially labeled data	**1.5**	**CO3**

	C. Automatically generated labels D. No labels		
Q 15	**Which of the following is an example of Supervised Learning?** A. K-means clustering B. Decision trees for classification C. Principal Component Analysis D. Hidden Markov models	**1.5**	**CO3**
Q 16	**Which algorithm is used in supervised learning to predict continuous values?** A. K-means clustering B. Linear regression C. Random Forest D. Q-learning	**1.5**	**CO3**
Q 17	**Which evaluation metric would be most important in a scenario where false negatives are critical, such as cancer detection?** A. Accuracy B. Recall C. Precision D. F1 Score	**1.5**	**CO4**
Q 18	**F1 Score is the harmonic mean of:** A. Accuracy and Precision B. Precision and Recall C. Accuracy and Recall D. Recall and Specificity	**1.5**	**CO4**
Q 19	**What is the key difference between training and validation sets?** A. The validation set is used to adjust model hyperparameters B. The validation set is used for model training C. The training set is used for final model evaluation D. The validation set is always smaller than the training set	**1.5**	**CO3**
Q 20	**Which of the following roles would best suit a non-IT major interested in Machine Learning?** A. Data scientist B. Clinical data analyst C. Network engineer D. Software tester	**1.5**	**CO5**
Section B (4Qx5M=20 Marks)			
Q 1	Discuss the key differences between AI and traditional manual methods in problem-solving. What advantages does AI offer over manual methods?	**5**	**CO1**
Q 2	Describe the role of cross-validation techniques in model evaluation. What is k-fold cross-validation?	**5**	**CO4**

<table>
<tr><td>Q 3</td><td>Explain the application of AI-based protein structure prediction techniques such as AlphaFold in modern bioinformatics. How has AI improved the accuracy and speed of structure prediction?</td><td>5</td><td>CO5</td></tr>
<tr><td>Q 4</td><td>Explain the process of interacting with a Python program from taking user input to displaying the output. Illustrate the process with examples demonstrating the use of input statements, type conversion, and output formatting.
Hint:
<pre># Step 1: Taking User Input
Use the input() function to accept user input as a string

Step 2: Accepting an integer input
Use the int() function to convert string input to an integer

Step 3: Processing the input
Perform some calculations or operations

Step 4: Displaying the output
Use formatted string (f-string) for output with formatting</pre></td><td>5</td><td>CO3</td></tr>
<tr><td colspan="4">Section C
(2Qx15M=30 Marks)</td></tr>
<tr><td>Q 1</td><td>What are the different data structures available in Python? Explain lists, tuples, sets, and dictionaries with examples. Highlight the differences between them, especially with respect to mutability and usage.</td><td>15</td><td>CO2</td></tr>
<tr><td>Q 2</td><td>A healthcare provider has implemented a model to predict whether patients have a particular disease based on symptoms. You are provided with the following dataset of actual and predicted disease statuses for 15 patients
<table>
<tr><th>Patient</th><th>Actual Label (Disease Status)</th><th>Predicted Label</th></tr>
<tr><td>1</td><td>1</td><td>1</td></tr>
<tr><td>2</td><td>0</td><td>0</td></tr>
<tr><td>3</td><td>1</td><td>0</td></tr>
<tr><td>4</td><td>1</td><td>1</td></tr>
<tr><td>5</td><td>0</td><td>0</td></tr>
<tr><td>6</td><td>1</td><td>1</td></tr>
<tr><td>7</td><td>0</td><td>1</td></tr>
<tr><td>8</td><td>1</td><td>1</td></tr>
<tr><td>9</td><td>1</td><td>1</td></tr>
<tr><td>10</td><td>0</td><td>0</td></tr>
<tr><td>11</td><td>0</td><td>0</td></tr>
<tr><td>12</td><td>1</td><td>0</td></tr>
<tr><td>13</td><td>1</td><td>1</td></tr>
</table></td><td>15</td><td>CO5</td></tr>
</table>

	14 | 0 | 1 15 | 1 | 1 **Using this data:** Construct the confusion matrix. (6 marks) Calculate the following metrics for the model: Accuracy (2 mark) Precision (2 marks) Recall (2 marks) F1 Score (3 mark) Show your calculations for each metric.		
Section D **(2Qx10M=20 Marks)**			
Q 1	Trace the history of Python as a programming language. Discuss its unique features and explain how Python's evolution from Python 2 to Python 3 has impacted its use in modern programming.	**10**	**CO1**
Q 2	Discuss the significance of data quality in Machine Learning. How do data preprocessing techniques improve the quality of the data?	**10**	**CO4**