**STELLA MARIS COLLEGE (AUTONOMOUS) CHENNAI 600 086**
**(For candidates admitted from the academic year 2023– 2024)**

**M. Sc. DEGREE EXAMINATION, NOVEMBER 2024**
**BIOINFORMATICS**
**THIRD SEMESTER**

**COURSE : CORE**
**PAPER : MACHINE LEARNING, DEEP LEARNING AND ARTIFICIAL INTELLIGENCE**
**SUBJECT CODE : 23BI/PC/MA34**
**TIME : 3 HOURS** **MAX. MARKS: 100**

| Q. No. | SECTION A<br>ANSWER ALL QUESTIONS (10X1=10) | CO | KL |
|---|---|---|---|
| 1. | In dealing with missing data in a dataset, which method is most appropriate for handling missing values in a continuous variable?<br>a) Dropping the rows with missing values<br>b) Replacing missing values with the mean of the column<br>c) Replacing missing values with a random value<br>d) Ignoring the missing values and proceeding with the analysis | CO1 | K1 |
| 2 | Which statistical measure best describes the spread or variability of a dataset?<br>a) Mean b) Median<br>c) Standard deviation d) Mode | CO1 | K1 |
| 3 | Which of the following statements is true regarding the relationship between correlation and causation?<br>a) Correlation implies causation<br>b) Causation implies correlation<br>c) Correlation and causation are unrelated<br>d) Causation can be inferred from correlation alone | CO1 | K1 |
| 4 | What is the primary purpose of the K-means clustering algorithm?<br>a) To classify data into predefined categories<br>b) To group data into a specified number of clusters based on similarity<br>c) To predict the next value in a time series<br>d) To reduce the dimensionality of the dataset | CO1 | K1 |
| 5 | In text processing, what does tokenization specifically involve?<br>a) Stemming words to their root forms<br>b) Removing stop words from the text<br>c) Breaking text into smaller units, such as words or phrases<br>d) Converting all text to lowercase | CO1 | K1 |
| 6 | Which method of ensemble voting involves selecting the class label that has the majority of votes from multiple models?<br>a) Averaging b) Weighted voting<br>c) Majority voting d) Stacking | CO1 | K2 |
| 7 | In a perceptron, what is the primary function of the activation function?<br>a) To normalize the input data<br>b) To determine the output based on a linear combination of inputs<br>c) To adjust the weights during training<br>d) To compute the gradient for backpropagation | CO1 | K2 |

..2

| | | | |
|---|---|---|---|
| 8 | Which component of an autoencoder is responsible for reducing the dimensionality of the input data?<br>a) Decoder b) Encoder<br>c) Activation function d) Loss function | CO1 | K2 |
| 9 | In the context of AI and machine learning for disease prediction, which technique is typically used to handle large datasets with complex patterns?<br>a) Rule-based systems<br>b) Linear regression<br>c) Deep learning neural networks<br>d) Decision trees | CO1 | K2 |
| 10 | Which of the following describes a common application of deep learning (DL) in medical diagnosis?<br>a) Performing manual data entry and record-keeping<br>b) Predicting patient outcomes based on historical data using simple linear models<br>c) Analyzing medical images to detect abnormalities such as tumors or fractures<br>d) Creating and testing new pharmaceutical drugs through traditional chemistry | CO1 | K2 |
| **Q. No.** | **SECTION B**<br>**ANSWER IN ABOUT 50 WORDS. (10X2=20)** | **CO** | **KL** |
| 11. | Illustrate the different forms of statistics. | CO2 | K3 |
| 12. | Elucidate the importance of Feature generation. | CO2 | K3 |
| 13. | Draw ROC curve and mention its importance. | CO2 | K3 |
| 14. | List the distance metrics used in KNN. | CO2 | K3 |
| 15. | What does it mean PoS in NLP? | CO2 | K3 |
| 16. | Compare and contrast the Bagging vs Boosting. | CO3 | K4 |
| 17. | Depict the schematic diagram of MLP. | CO3 | K4 |
| 18. | Comment on the different gates of LSTM. | CO3 | K4 |
| 19. | Write any two applications of ML, DL and AI. | CO3 | K4 |
| 20. | List the types of RBM. | CO3 | K4 |
| **Q. No.** | **SECTION C**<br>**ANSWER IN ABOUT 600 WORDS. (4X10=40)** | **CO** | **KL** |
| 21. | a) Categorize the Machine learning concepts based on the types of data.<br>**(OR)**<br>b) Measure the importance of Data imputation and how to imply it. Discuss with an example. | CO3 | K4 |
| 22. | a) Examine the importance of Cross-fold K validation<br>**(OR)**<br>b) Discuss the working mechanisms of SVM algorithm. | CO3 | K4 |
| 23. | a) Comment on the different ensemble methods used in random forest algorithm.<br>**(OR)**<br>b) Elucidate the importance of N-grams in text mining process. | CO4 | K5 |

| | | | |
|---|---|---|---|
| 24. | a) Infer the structure and mechanism of ANN in detail.<br>**(OR)**<br>b) Discuss the role of RBM (Restricted Boltzmann Machines) in ANN. | CO4 | K5 |
| **Q. No.** | **SECTION D**<br>**ANSWER IN ABOUT 1200 WORDS. (2X15=30)** | **CO** | **KL** |
| 25. | a) Justify the use of machine learning (ML), deep learning (DL), and artificial intelligence (AI) in drug discovery and development, and discuss their impact on improving the process.<br>**(OR)**<br>b) Appraise the role of the ROC curve and evaluating metrics in measuring the performance of classification models and discuss its significance in model selection. | CO5 | K6 |
| 26. | a) Evaluate the importance of feedforward neural networks in machine learning, focusing on their structure, training processes, and applications.<br>**(OR)**<br>b) Assess the impact of text pre-processing techniques like tokenization, noise removal, and vectorization methods on text mining results. | CO5 | K6 |

*******