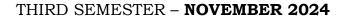
LOYOLA COLLEGE (AUTONOMOUS) CHENNAI – 600 034



M.Sc. DEGREE EXAMINATION – STATISTICS





PST3MC03 - DATA MINING AND MACHINE LEARNING

| | Date: 12-11-2024 Dept. No. Time: 01:00 pm-04:00 pm | Max. : 100 Marks | | | | | | |
|----------------------|---|---------------------------|--|--|--|--|--|--|
| CECTION A 1/1 (CO1) | | | | | | | | |
| SECTION A – K1 (CO1) | | | | | | | | |
| 4 | Answer ALL the questions | $(5 \times 1 = 5)$ | | | | | | |
| 1 | Fill in the blanks | | | | | | | |
| a) b) | The last phase in CRISP-DM is the | | | | | | | |
| c) | is a graphical method of identifying outliers. A decision tree is used for and tasks. | | | | | | | |
| d) | | | | | | | | |
| u) | A logistic regression tree is a machine learning method that partitions the data and fits an model in each partition. | | | | | | | |
| e) | • | sover, and mutation. | | | | | | |
| <u> </u> | SECTION A – K2 (CO1) | sover, and matarion. | | | | | | |
| | | | | | | | | |
| | Answer ALL the questions | $(5 \times 1 = 5)$ | | | | | | |
| 2 | Answer the following | | | | | | | |
| <u>a)</u> | Define Unsupervised Learning Algorithms. | | | | | | | |
| <u>b)</u> | What is information gain? Give the formula and the terminologies used. | | | | | | | |
| <u>c)</u> | Define Naïve Bayes algorithm. | | | | | | | |
| <u>d)</u> | What is Binary classifier in Machine Learning? | | | | | | | |
| e) | What is text mining? | | | | | | | |
| SECTION B – K3 (CO2) | | | | | | | | |
| | Answer any THREE of the following | $(3 \times 10 = 30)$ | | | | | | |
| 3 | Explain the three methods of data transformation using suitable examples. | | | | | | | |
| 4 | How does Random Forest work? List the key differences between Random Forest and other machine | | | | | | | |
| | learning algorithms? | | | | | | | |
| 5 | What are Support Vector Machines? Elaborate how they work. | | | | | | | |
| 6 | Explain the construction of Bayesian network with suitable example. | | | | | | | |
| 7 | Outline the foundations of genetic algorithms, search space, fitness score and operators. | | | | | | | |
| SECTION C – K4 (CO3) | | | | | | | | |
| | Answer any TWO of the following | $(2 \times 12.5 = 25)$ | | | | | | |
| 8 | Explain all the steps in data pre-processing in detail. | | | | | | | |
| 9 | Explain the k Nearest neighbor algorithm with special reference to k ar | nd the distance functions | | | | | | |
| 10 | What is back propagation in machine learning? Explain the working of | | | | | | | |
| 11 | What are the similarities and differences between Bagging and Boostin Learning? | g algorithms in Machine | | | | | | |
| | SECTION D – K5 (CO4) | | | | | | | |
| | Answer any ONE of the following $(1 \times 15 = 15)$ | | | | | | | |
| 12 | What are Neural Networks? List out their applications and explain the components in detail. | | | | | | | |
| 13 | Explain the concept of association rule mining from large datasets. Ela association mining with advantages and disadvantages. | | | | | | | |
| | | | | | | | | |

SECTION E – K6 (CO5)

Answer any ONE of the following

 $(1 \times 20 = 20)$

14 Grow a decision tree using the C 4.5 algorithm, for the following training set records.

| Customer | Savings | Assets | Income (\$1000s) | Credit Risk |
|----------|---------|--------|------------------|-------------|
| 1 | Medium | High | 75 | Good |
| 2 | Low | Low | 50 | Bad |
| 3 | High | Medium | 25 | Bad |
| 4 | Medium | Medium | 50 | Good |
| 5 | Low | Medium | 100 | Good |
| 6 | High | High | 25 | Good |
| 7 | Low | Low | 25 | Bad |
| 8 | Medium | Medium | 75 | Good |

Construct a Naïve Bayes Classifier to predict the probability of "Not playing cricket" given that the weather conditions are "Rainy outlook", "Temperature is hot", "high humidity" and "no wind".

| Outlook | Temperature | Humidity | Windy | Play Cricket |
|----------|-------------|----------|-------|--------------|
| Rainy | Hot | High | False | No |
| Rainy | Hot | High | True | No |
| Overcast | Hot | High | False | Yes |
| Sunny | Mild | High | False | Yes |
| Sunny | Cool | Normal | False | Yes |
| Sunny | Cool | Normal | True | No |
| Overcast | Cool | Normal | True | Yes |
| Rainy | Mild | High | False | No |