Seat No.	

T.Y. B.Tech. (CSE) (Data Science) (Semester - V)
(CBCS) Examination, December - 2023
DATA MINING & DATA WAREHOUSING
Sub. Code : 86858

Day and Date : Saturday, 02- 12 - 2023 **Total Marks : 70**
Time : 10.30 a.m. to 1.00 p.m.

Instructions :   1)   **All questions are compulsory.**
2)   **Assume suitable data wherever necessary.**
3)   **Figures to the right indicate full marks.**

Q1) Solve MCQs. (2 Marks Each)   **[14]**

a)   Identify the term used to define the task of inferring a model from labeled training data.
   i)   Supervised Learning   ii)   Unsupervised Learning
   iii)   Reinforcement Learning   iv)   None

b)   What is KDD in Data Mining?
   i)   Knowledge Discovery Database
   ii)   Knowledge Discovery Data
   iii)   Knowledge Data Definition
   iv)   Knowledge Data House

c)   Identify the correct answer which defines Datamart
   i)   A subgroup of Data warehouse
   ii)   Another type of data warehouse
   iii)   Not related to data warehouse
   iv)   None

d)   Why is the Snowflake Schema applied?
   i)   Transformation   ii)   Aggregation
   iii)   Normalization   iv)   Generalization

e)   On what is data warehouse based?
   i)   1D model   ii)   2D model
   iii)   3D model   iv)   Multidimensional Model

P.T.O.

f) Which of the following can be considered as the classification or mapping of a set or class with some predefined group or classes
   i)   Data set   ii) Data Characterization
   iii) Data Sub Structure   iv) Data Discrimination

g) The issues like efficiency, scalability of data mining algorithms comes under_____
   i)   Performance Issues
   ii) Diverse Data type issues
   iii) Mining Methodology and user interaction
   iv) All of the above

Q2) Solve any 2 of the following (7 Marks Each)   **[14]**
   a) What is Data Mining? Explain Data Mining as a step in the process of knowledge discovery.
   b) Write a short note on Data Integration.
   c) Differentiate between OLAP and OLTP.

Q3) Solve any 2 of the following (7 Marks Each)   **[14]**
   a) Write a short note on:
      i)   Data Characterization and Data Discrimination
      ii) Classification and Regression for predictive analysis
   b) How Missing values are handled in data cleaning?
   c) How indexing takes place on OLAP data? Explain Bitmap and Join Index with example.

Q4) Solve any 2 of the following (7 Marks Each)   **[14]**
   a) Explain Decision tree induction with example.
   b) What is cluster analysis? Explain Requirements for cluster analysis.
   c) Explain Categories of Sequence Data.

Q5) Solve any 2 of the following (7 Marks Each)   **[14]**
   a) Write a short note on: Precision and Recall
   b) Explain Agglomerative and Divisive Clustering.
   c) Explain Data mining trends.

Summer Examination March - 2023

Subject Name: B.Tech. CBCS_86877 - E-Commerce & Digital Marketing (CSDS)_05.07.2023_10.30 AM To 01.00 PM

Subject Code: 86877

Day and Date: - Wednesday, 05-07-2023
Time: - 10:30 am to 01:00 pm

Total Marks: 70

Instructions.:
   1) All questions are compulsory
   2) Figures to the right indicate full marks

Q.1. **Solve MCQs (All questions compulsory) (1 Mark each).** **[14]**
(i) Which of the following is a key characteristic of B2B transactions?
a) Small order volumes
b) Simple purchasing processes
c) Shorter sales cycles
d) Large order values

(ii) Which marketing channel is commonly used in B2B marketing?

a) Social media advertising
b) Television commercials
c) Outdoor billboards
d) Print newspapers

(iii) Which of the following is NOT a benefit of e-commerce for businesses?

a) Global reach and access to a larger customer base
b) Lower operational costs compared to physical stores
c) Limited opportunities for personalized customer interactions
d) Increased efficiency in inventory management

(iv) Which of the following describes E-Commerce?

a) Business activities
b) Sale of goods
c) Doing business online
d) All of the above

(v) What is the key advantage of using social media   **[14]**
marketing in digital marketing?

a) Targeting specific demographics and interests
b) Running advertisements on television and radio
c) Monitoring and analyzing website performance metrics
d) Enhancing the user experience on a website

(vi) What is the primary goal of search engine optimization
(SEO) in digital marketing?

a) Increasing website traffic and visibility through organic search
results
b) Running paid advertisements on search engines
c) Enhancing the user experience on a website
d) Monitoring and analyzing website performance metrics

(vii) Which of the following is a characteristic of a successful
online marketplace?

a) Limited product selection and variety
b) Exclusive partnerships with a single seller in each category
c) Secure payment processing and buyer protection
d) Offline, brick-and-mortar presence

(viii) What is the primary goal of relationship marketing?

a) To increase the number of one-time customers
b) To maximize short-term profits
c) To build and maintain long-term customer relationships
d) To focus solely on acquiring new customers

(ix) Which of the following is a component of the macro
environment?

a) Competitive analysis
b) Marketing strategy
c) Technological advancements
d) Supply chain management

(x) What is affiliate marketing?   **[14]**

a) Selling products directly to customers through an online marketplace
b) Earning commissions by promoting and driving sales for other companies' products or services
c) Running paid advertisements on social media platforms
d) Implementing email marketing campaigns to engage with customers

(xi) Which of the following is NOT a primary objective of CRM?

a) Acquiring new customers
b) Building long-term customer loyalty
c) Increasing operational efficiency
d) Minimizing product development costs

(xii) Which of the following is an example of a key performance indicator (KPI) in digital marketing?

a) Number of Facebook likes
b) Number of website visitors
c) Number of industry conferences attended
d) Number of office locations

(xiii) What is email marketing?

a) Sending unsolicited emails to random recipients
b) Creating and sending targeted promotional emails to a specific audience
c) Sharing personal emails with friends and family
d) Selling email software and tools

(xiv) What is the main benefit of paid search marketing?

a) Increased organic search rankings
b) Higher website traffic without any cost
c) Immediate visibility and targeted reach
d) Long-term brand awareness and loyalty

Q.2.   Solve any two of the following questions. (7 marks each)   **[14]**
  a) Explain advantages of E-Commerce to Organizations, Consumers and Society and explain technical and non technical disadvantages of e-commerce.
  b) Explain primary and secondary sales process in E-Commerce with neat diagram.
  c) Explain B2B business process model and its variants.

Q.3. **Solve any two of the following questions. (7 marks each)** **[14]**
a) Explain B2C Business and Customer Relationship Management(CRM) in detail.
b) What are the key features of digital marketing strategy.
c) Explain Multichannel marketing with Paid, owned, and earned media.

Q.4. **Solve any two of the following questions. (7 marks each)** **[14]**
a) What is online marketplace? Explain the key factors to consider in a situation analysis for digital marketing.
b) Explain the main elements of the online marketplace map with diagram.
c) Explain implementation of digital marketing strategy in detail.

Q.5. **Solve any two of the following questions. (7 marks each)** **[14]**
a) Explain implementation of digital marketing strategy in detail.
b) What is paid search marketing? Explain advantages and disadvantages of paid search marketing.
c) What is Search Engine Optimization (SEO) ? Explain Different Stages involved in producing a search engine listing for the natural listings.

Seat No.

T.Y. B.Tech. (CSE(DS)OE2) (Part - III) (Semester - VI) (CBCS) Examination, November - 2023
E-COMMERCE & DIGITAL MARKETING
Sub. Code : 86877

Day and Date : Saturday, 25 - 11 - 2023   **Total Marks : 70**
Time : 2.30 p.m. to 5.00 p.m.

Instructions :
1) **All questions are compulsory.**
2) **Assume suitable data wherever necessary.**
3) **Figures to the right indicate full marks.**

Q1) Select the correct option and rewrite complete sentence.   **[14]**

   i)   Which of the following is NOT a benefit of e-commerce for businesses?

      a)   Global reach and access to a larger customer base

      b)   Lower operational costs compared to physical stores

      c)   Limited opportunities for personalized customer interactions

      d)   Increased efficiency in inventory management

   ii)   Which of the following describes E-Commerce?

      a)   Business activities   b)   Sale of goods

      c)   Doing business online   d)   All of the above

   iii)   Which of the following is a key characteristic of B2B transactions?

      a)   Small order volumes   b)   Simple purchasing processes

      c)   Shorter sales cycles   d)   Large order values

   iv)   Which marketing channel is commonly used in B2B marketing?

      a)   Social media advertising   b)   Television commercials

      c)   Outdoor billboards   d)   Print newspapers

   v)   What is the primary goal of search engine optimization (SEO) in digital marketing?

      a)   Increasing website traffic and visibility through organic search results

      b)   Running paid advertisements on search engines

      c)   Enhancing the user experience on a website

      d)   Monitoring and analyzing website performance metrics

P.T.O.

vi) What is the key advantage of using social media marketing in digital marketing?

   a) Targeting specific demographics and interests

   b) Running advertisements on television and radio

   c) Monitoring and analyzing website performance metrics

   d) Enhancing the user experience on a website

vii) Which of the following is a characteristic of a successful online marketplace?

   a) Limited product selection and variety

   b) Exclusive partnerships with a single seller in each category

   c) Secure payment processing and buyer protection

   d) Offline, brick-and-mortar presence

viii) Which of the following is a component of the macro environment?

   a) Competitive analysis

   b) Marketing strategy

   c) Technological advancements

   d) Supply chain management

vix) What is the primary goal of relationship marketing?

   a) To increase the number of one-time customers

   b) To maximize short-term profits

   c) To build and maintain long-term customer relationships

   d) To focus solely on acquiring new customers

x) Which of the following is NOT a primary objective of CRM?

   a) Acquiring new customers

   b) Building long-term customer loyalty

   c) Increasing operational efficiency

   d) Minimizing product development costs

xi) Which of the following is an example of a key performance indicator (KPI) in digital marketing?

a) Number of Facebook likes

b) Number of website visitors

c) Number of industry conferences attended

d) Number of office locations

xii) What is the main benefit of paid search marketing?

a) Increased organic search rankings

b) Higher website traffic without any cost

c) Immediate visibility and targeted reach

d) Long-term brand awareness and loyalty

xiii) What is affiliate marketing?

a) Selling products directly to customers through an online marketplace

b) Earning commissions by promoting and driving sales for other companies' products or services

c) Running paid advertisements on social media platforms

d) Implementing email marketing campaigns to engage with customers

xiv) What is email marketing?

a) Sending unsolicited emails to random recipients

b) Creating and sending targeted promotional emails to a specific audience

c) Sharing personal emails with friends and family

d) Selling email software and tools

Q2) Solve any 2 of the following. (7 marks Each) **[14]**

a) Define E-Commerce? Explain different types of E-Commerce.

b) Explain primary and secondary sales process in E-Commerce with neat diagram.

c) Explain B2B business process model and its variants

Q3) Solve any 2 of the following. (7 marks Each)   **[14]**

   a)   What is B2B and B2C? Differentiate between B2B and B2C.

   b)   Explain Multichannel marketing with Paid, owned, and earned media.

   c)   What are the challenges in developing and managing digital marketing strategy

Q4) Solve any 2 of the following. (7 marks Each)   **[14]**

   a)   Explain the main elements of the online marketplace map with diagram.

   b)   What is online marketplace? Explain the key factors to consider in a situation analysis for digital marketing.

   c)   Explain implementation of digital marketing strategy in detail.

Q5) Solve any 2 of the following. (7 marks Each)   **[14]**

   a)   Explain how to structure digital marketing strategy with neat diagram.

   b)   What is email marketing? What are Advantages and disadvantages of email marketing.

   c)   What is Search Engine Optimization (SEO)? Explain Different Stages involved in producing a search engine listing for the natural listings.

▽ ▽ ▽ ▽

Seat No.

January - February (Winter) Examination - 2023

Subject Name: B.Tech. CBCS_86853_Fundamental of Data Science_13.01.2023_10.30 AM To 01.00 PM

Subject Code: 86853

Day and Date: Friday, 13-01-2023
Time: 10:30 am to 01:00 pm

Total Marks: 70

Instructions.:

1) All questions are compulsory
2) Figures to the right indicate full marks
3) Assume suitable data wherever necessary and mention it boldly

---

Q.1. **Solve the MCQ** **[14]**

a) Which of the following is not a part of data science process?
   i) Setting the research Goal
   ii) Retrieving Data
   iii) Communication Building
   iv) Data Preparation

b) Which of the following are the Data Sources in data science?
   i) Structured
   ii) UnStructured
   iii) Both A and B
   iv) None Of the above

c) Which of the following is not an application for data science?
   i) Recommendation Systems
   ii) image & Speech Recognition
   iii) Online Price Comparison
   iv) Privacy Checker

d) Which of the following is one of the key data science skills?
   i) Statistics
   ii) Machine Learning
   iii) Data Visualization
   iv) All of the above

e) Which of the following step is performed by data scientist after acquiring the data?
   i) Data Cleaning
   ii) Data Integration
   iii) Data Replication
   iv) All of the above

f) Unstructured data is not organized. **[14]**
    i) TRUE
    ii) FALSE
    iii) Can be true or false
    iv) Can not say

g) Which of the following is an example of Semi-Structured Data.
    i) RDBMS
    ii) Audio Files
    iii) JSON
    iv) Video Files

h) Amongst which of the following can be considered as the main source of unstructured data.
    i) Twitter
    ii) Facebook
    iii) Webpages
    iv) All of the mentioned above

i) Amongst which of the following is / are used to analyze the data in pandas.
    i) Dataframe
    ii) Series
    iii) Both A and B
    iv) None of the mentioned above

j) Which of the following is not the part of Data Preparation in Data Science Process.
    i) Data Cleansing
    ii) Data Transformation
    iii) Data Presentation
    iv) Combining Data

k) NumPy arrays can be ___.
    i) Indexed
    ii) Sliced
    iii) Iterated
    iv) All of the mentioned above

l) NumPy package is capable to do fast operations on arrays.
    i) TRUE
    ii) FALSE

m) What is the purpose of NumPy in Python?
    i) To do numerical calculations
    ii) To do scientific computing
    iii) Both A and B
    iv) None of the mentioned above

n) Amongst which of the following shows an example of unstructured data,
    i) Students roll number, age in Table
    ii) Videos
    iii) Audio files
    iv) Both B and C

Q.2. **Solve any 2 (Two) of the following questions (7 Marks Each)** **[14]**
a) What is Data Science ? Explain three types of data with examples.
b) Differentate between Data Scientist and Data Analyst.
c) Explain about various stages in Data Science process.

Q.3. **Solve any 2 (Two) of the following questions (7 Marks Each)** **[14]**
a) Describe various techniques used to handle the missing values in a Data Science Process
b) Expain different types of inconsistancies error within different dataset.
c) Explain any 3 Data Manipulation techniques using Excel.

Q.4. **Solve any 2 (Two) of the following questions (7 Marks Each)** **[14]**
a) What is NumPy Array ? Explain different types of built in function to create numpy array with examples.
b) Explain the basic NumPy Uuniversal functions('U' Functions) with examples.
c) Create an appropriate array to describe array concatenation & splitting in NumPy array.

Q.5. **Solve any 2 (Two) of the following questions (7 Marks Each)** **[14]**
a) What is Pandas Series? Explain different way of creating Pandas Series with examples.
b) Create a sample pandas series and with examples Explain different mathematical operations on Pandas Series.
c) With neat diagram, Explain basics of creating graphs in Matplotlib libraries.

Seat No.

T.Y. B.Tech. (CSE) (Semester - V) (CBCS) Examination, November - 2023
FUNDAMENTAL OF DATA SCIENCE (DS)
Sub. Code : 86853

Day and Date : Tuesday, 28 - 11 - 2023   **Total Marks : 70**
Time : 10.30 a.m. to 01.00 p.m.

Instructions :   1)   **All questions are compulsory.**
               2)   **Figures to the right indicate full marks.**
               3)   **Assume suitable data wherever necessary and mention it boldly.**

Q1) **Solve the MCQ.**   **[14]**

   i)   Point out the correct statement.

       A)   Raw data is original source of data

       B)   Preprocessed data is original source of data

       C)   Raw data is the data obtained after processing steps

       D)   None of the mentioned

   ii)   Which of the following is one of the key data science skills?

       A)   Statistics

       B)   Machine Learning

       C)   Data Visualization

       D)   All of the above

   iii)   What will be the output of the following Python code?

       from numpy import random

       x = random.randint(100)

       print(x)

   A)   120   B)   26

   C)   103   D)   None of the above

iv) The following code create a dataframe named 'D1' with _____ columns

import pandas as pd

LoD =[{'a':10, 'b':20}, {'a':5, 'b' :10, 'c':20}]

D1 = pd.DataFrame(LoD)

A)   0   B)   1

C)   2   D)   3

v) Which of the following function is used to create DataFrame?

A)   DataFrame()   B)   NewFrame()

C)   CreateDataFrame()   D)   None of the Above

vi) What will be output for the following code?

import pandas as pd

s = pd.Series ([1,2,3,4,5],index = ['a', 'b', 'c', 'd', 'e'])

print s['a']

A)   1   B)   2

C)   3   D)   4

vii) We can create DataFrame from_____

A)   Numpy arrays   B)   List of Dictionaries

C)   Dictionary of Lists   D)   All of the above

Q2) **Solve any 2 (Two) of the following questions (7 Marks Each)**   **[14]**

a)   What is Data Science? Explain three types of Data with examples.

b)   Explain benefits and uses of Data Science and Big Data.

c)   Differentiate between Data Scientist and Data Analyst

Q3) **Solve any 2 (Two) of the following questions (7 Marks Each)   [14]**

   a)   Explain different types of inconsistencies error within different dataset.

   b)   Describe various techniques used to handle the missing values in a Data Science Process

   c)   With neat diagram describe any four data manipulation techniques in Excel.

Q4) **Solve any 2 (Two) of the following questions (7 Marks Each)   [14]**

   a)   Explain the basic NumPy Universal functions ('U' Functions) with examples.

   b)   What is NumPy Array? Explain different types of built in function to create numpy array with examples.

   c)   What is Data Frame in Pandas? Explain any three ways of creating Dat Frame in Pandas with examples?

Q5) **Solve any 2 (Two) of the following questions (7 Marks Each)   [14]**

   a)   Explain different methods for detecting, removing, and replacing null values in Panda's data structures.

   b)   Create a sample pandas series and with examples Explain different mathematical operations on Pandas Series.

   c)   With neat diagram, explain various basics graphs in matplotlib libraries.

෴෴෴

Seat No.

Summer Examination March - 2023

Subject Name: B.Tech. CBCS_86856 - Feature Engineering_27.06.2023_02.30 PM To 05.00 PM
Subject Code: 86856

Day and Date: - Tuesday, 27-06-2023   Total Marks: 70
Time: - 02:30 pm to 05:00 pm

Instructions.:
      1) All questions are compulsory

Q.1. **Solve the following MCQ(1 mark each)** **[14]**

1_____ investigates how computers can learn (or improve their performance) based on data
  a. Machine learning
  b. Artificial intelligence
  c.Statistics
  d.Visualization

2. Choose the correct components of data science
  a.Domain expertise
  b.Data engineering
c.Advanced computing
d.All of the above

3. Mean is the _____ of a dataset
a.Average
b.Middle
c.Central
d.Ordered

4.The number that occurs most often within a set of data called as ____
a.Mean
b.Median
c.Mode
d.Range

5.Knowledge in AI can be represented as?
a.Predicate
b.Propositional logic
c.Both a) and b)
d.None of the above

6.Among the following identify the one in which dimensionality reduction reduces
a.Performance
b.Entropy
c.Stochastics
d.Collinearity

7.Multiple data sources may be combined is called
a.Data reduction
b.Data cleaning
c.Data integration
d.Data transformation

8.The different types of machine learning are?
a.Supervised
b.Unsupervised
c.Reinforcement
d.All of the above

9.Among the following options identify the one which is false
a.It is used for the prediction
b.It is used for interpretation
c.It relates inputs to outputs
d.It discovers causal relationship

10.Artificial Intelligence is associated with computers of which generation?
a.Second
b.First
c.Fifth
d.Third

11.Identify the key data science skills among the following
a.Data visualization
b.Machine learning
c.Statistics
d.All of the above

12.Which of the following is an example of raw data?
a.Original swath files generated from a sonar system
b.Initial time-series file of temperature values
c.A real-time GPS-encoded navigation file
d.All of the mentioned above

13.Data discretization is
a.To convert continuous attributes to discrete attributes
b.To scale up data
c.To convert discrete attributes to continuous
d.None

14.This is the cleaning/transforming the data set in the supervised learning model.
a.Problem Identification
b.Identification of Required Data
c.Data Pre-processing
d.Definition of Training Data Set

Q.2. **Answer Any two questions [2 X 7 = 14 Marks]**   **[14]**
1. What are supervised and unsupervised learning? Explain with example [7 marks]
2. What is quantization or binning? Explain types of Binning with example [7 marks]
3. How do you overcome challenges with missing data? Compare between feature extraction and feature selection? When should each one be used?   [7 marks]

Q.3. **Answer any two questions [2 X 7 = 14 Marks]** **[14]**
1. Define and explain the terms   i. Linear Projection, ii. Variance and empirical Variance, iii. Intuition, iv. Derivation [7 marks]
2. What is Feature Scaling or Normalization. Explain Min-Max Normalization with example [7 marks]
3. Explain the process of Tf-Idf in Natural Language Processing. Discuss advantages and disadvantages of Tf-Idf[7 marks]

Q.4. **Answer any two questions [2 X 7 = 14 Marks]** **[14]**
1. "Effect of irrelevant features" explain with case study [7 Marks]
2. what is feature transformation? Explain with diagram [7 Marks]
3. Describe the process i. Bag-of-words ii. N-grams iii. One-hot-encoding [7 marks]

Q.5. **Answer any two questions [2 X 7 = 14 Marks]** **[14]**
1. What are the three classes of feature selection methodologies? [7 marks]
2. what is dimensionality reduction? Describe dimensionality reduction with PCA [7 marks]
3. Explain reconstruction of data with Restricted Boltzmann's Machine (RBM) considering two layers.. [7 marks]

January - February (Winter) Examination - 2023

Subject Name: B.Tech. CBCS_86856_Feature Engineering_21.01.2023_10.30 AM To 01.00 PM

Subject Code: 86856

Day and Date: Saturday, 21-01-2023
Time: 10:30 am to 01:00 pm

Total Marks: 70

Instructions.:

   1) All questions are compulsory

Q.1.   **Solve the MCQ**   **[14]**

   a. Principal component is a technique for
   1. Feature selection
   2. Dimensionality reduction
   3. Exploration
   4. None of the above
   b. In feature extraction, some of the commonly used ___ are used for combining the original features.
   1. Operators
   2. Delimiters
   3. Words
   4. All of the above
   c. This is the cleaning/transforming data set in the supervised learning model.
   1. Problem Identification
   2. Identification of Required Data
   3. Data Pre-processing
   4. Definition of Training Data Set
   d. n-gram of size 1 is called
   1. Bigram
   2. Unigram
   3. Trigram
   4. None of the above
   e. "The Cat flys" after applying which n-gram gives the output as "The Cat","Cat flys"
   1. Unigram
   2. Bigram
   3. Trigram
   4. Quadrigrams
   f. What is the data cleaning method
   1. Binning
   2. Clustering
   3. Regression
   4. Classification

g. Select Normalization Method

1. Min-max normalization

2. Z-score Normalization

3. Decimal Scaling Normalization

4. Schema integration

h. Data discretization is

1. To convert continuous attributes to discrete attributes

2. To scale up data

3. To convert discrete attributes to continuous

4. None

i. To remove noise and inconsistent data _____ is needed

1. Data Cleaning

2. Data Transformation

3. Data Reduction

4. Data Integration

j. Bag of Words in text preprocessing is a-

1. Feature Scaling technique

2. Feature Extraction Technique

3. Feature selection technique

4. None

k. The primary use of data cleaning is

1. Removing the noisy data

2. Correction of the data inconsistencies

3. Transformations for correcting the wrong data

4. All of the above

l. ____ is a top-down spitting technique based on a specified number of bins

1. Normalization

2. Binning

3. Clustering

4. Classification

m. Machine learning is a subset of which of the following

1. Feature Engineering

2. Artificial Intelligence

3. Deep learning

4. None of the above

n. In ____, the attribute data are scaled so as to fall within a smaller range, such as -1.0 to 1.0, or 0.0 to 1.0.

1. Aggregation

2. Binning

3. Clustering

4. Normalization

Q.2.   Solve any two questions [2 X 7 = 14 Marks]   **[14]**
1. What is feature engineering? Why Feature Engineering is important in model building and lists out some of the techniques used for Feature Engineering.
2. What is quantization or binning? Explain types of Binning with example.
3. How do you turn the natural text to flat vector. Explain with Bag-of-words.

Q.3.   Solve any two questions [2 X 7 = 14 Marks]   **[14]**
1. Explain the process of cleaning and organizing the data with figure and example.
2. What is Feature Scaling or Normalization. Explain Min-Max Normalization with example.
3. Explain the process of Tf-Idf in Natural Language Processing. Discuss advantages and disadvantages of Tf-Idf.

Q.4.   Solve any two questions [2 X 7 = 14 Marks]   **[14]**
1. "Effect of irrelevant features" explain with case study.
2. what is feature transformation? Explain with diagram.
3. Explain reconstruction of data with Restricted Boltzmann's Machine (RBM) considering two layers.

Q.5.   Solve any two questions [2 X 7 = 14 Marks]   **[14]**
1. What are the three classes of feature selection methodologies?
2. what is dimensionality reduction? Describe dimensionality reduction with PCA.
3. What is data reconstruction? Define the term Bernoulli RBM with proper explanation.

Seat
No.

T.Y. B.Tech. (CSE) (Data Science) (Semester - V) (CBCS) Examination, December - 2023
FEATURE ENGINEERING
Sub. Code : 86856

Day and Date : Monday, 4 - 12 - 2023   **Total Marks : 70**
Time : 10.30 a.m. to 1.00 p.m.

Instructions :   1)   **All questions are compulsory.**
                     2)   **Assume suitable data wherever necessary.**
                     3)   **Figures to the right indicates full marks.**

Q1) Solve MCQs. (2 Marks Each)

a)   Principal component is a technique for
   i)   Feature selection   ii)   Dimensionality reduction
   iii) Exploration   iv)   None of the above

b)   In feature extraction, some of the commonly used _____ are used for combining the original features.
   i)   Operators   ii)   Delimiters
   iii) Words   iv)   All of the above

c)   This is the cleaning/transforming the data set in the supervised learning model.
   i)   Problem Identification
   ii)   Identification of Required Data
   iii) Data Pre-processing
   iv)   Definition of Training Data Set

d)   n-gram of size 1 is called
   i)   Bigram   ii)   Unigram
   iii) Tri gram   iv)   None of the above

e)   "The Cat flys" after applying which n-gram gives the output as "The Cat", "Cat flys"
   i)   Unigram   ii)   Bigram
   iii) Trigram   iv)   Quadrigrams

f)   What is data cleaning method
   i)   Binning   ii)   Clustering
   iii) Regression   iv)   Classification

P.T.O.

g) Select Normalization Method
   i) Min-max normalization
   ii) Z-score Normalization
   iii) Decimal Scaling Normalization
   iv) Schema integration

h) Data discretization is
   i) To convert continuous attributes to discrete attributes
   ii) To scale up data
   iii) To convert discrete attributes to continuous
   iv) None

i) To remove noise and inconsistent data _____ is needed.
   i) Data Cleaning   ii) Data Transformation
   iii) Data Reduction   iv) Data Integration

j) Bag of Words in text preprocessing is a-
   i) Feature Scaling technique
   ii) Feature Extraction Technique
   iii) Feature selection technique
   iv) None

k) The primary use of data cleaning is
   i) Removing the noisy data
   ii) Correction of the data inconsistencies
   iii) Transformations for correcting the wrong data
   iv) All of the above

l) _____ is a top-down spitting technique based on a specified number of bins
   i) Normalization   ii) Binning
   iii) Clustering   iv) Classification

m) Machine learning is a subset of which of the following
   i) Feature Engineering   ii) Artificial Intelligence
   iii) Deep learning   iv) None of the above

n) In _____, the attribute data are scaled so as to fall within a smaller range, such as - 1.0 to 1.0,or 0.0 to 1.0
   i) Aggregation   ii) Binning
   iii) Clustering   iv) Normalization

Q2) Answer Any two questions. **[2×7=14]**

a) What is feature engineering? Why Feature Engineering is important in model building and lists out some of the techniques used for Feature Engineering?

b) What are supervised and unsupervised learning? Explain with example.

c) How do you turn the natural text to flat vector. Explain with Bag-of-words.

Q3) Answer any two questions. **[2×7=14]**

a) Explain the process of cleaning and organizing the data with figure and example.

b) Define and explain the terms i. Linear Projection, ii. Variance and empirical Variance, iii. Intuition, iv. Derivation.

c) Explain the process of Tf-Idf in Natural Language Processing. Discuss advantages and disadvantages of Tf-Idf.

Q4) Answer any two questions. **[2×7=14]**

a) Describe the process
   i) Bag-of-words
   ii) N-grams
   iii) One-hot-encoding

b) What is feature transformation? Explain with diagram

c) Explain reconstruction of data with Restricted Boltzmann's Machine (RBM) considering two layers

Q5) Answer any two questions. **[2×7=14]**

a) What are the three classes of feature selection methodologies?

b) What is dimensionality reduction? Describe dimensionality reduction with PCA

c) What is data reconstruction? Define the term Bernoulli RBM with proper explanation.

♦ ♦ ♦

Summer Examination March - 2023

Subject Name: B.Tech. CBCS_86854_86854 - Internet of Things_24.06.2023_02.30 PM To 05.00 PM

Subject Code: 86854

Day and Date: - Saturday, 24-06-2023
Time: - 02:30 pm to 05:00 pm

Total Marks: 70

Instructions.:
      1) All questions are compulsory
      2) Figures to the right indicate full marks

Q.1. **What is IoT? Define different characteristics of IoT** **[7]**

Q.2. **Describe various technologies that are working as an IoT enablers.** **[7]**

Q.3. **Write a note on RFID. How RFID overcomes the limitations of Infrared?** **[7]**

Q.4. **Compare IPv4 with IPv6.** **[7]**

Q.5. **Design smart door for secure home using IoT.** **[7]**

Q.6. **Write a note on Raspberry-Pi. Explain device compatibility and steps involved in using raspberry-pi** **[7]**

Q.7. **Draw and Explain WPAN Architecture and it's Protocol Stack.** **[7]**

Q.8. **Describe functions of following in Bluetooth Protocol stack - i) RFCOMM ii) TCS iii) SDP iv)L2CAP v) Baseband** **[7]**

Q.9. **What is Internet of Things?** **[1]**
A. It is the inter-networking of physical devices embedded with electronics, software, sensors, actuators, and network connectivity that enable these objects to collect and exchange data.
B. It is protocol to access internet
C. It set of services used to access internet
D. None of above

Q.10. **What is EPC in IoT?** **[1]**
    A. Electrical Process Component
    B. Electronic Process Code
    C. Electrical Process Code
    D. Electronic Product Code

Q.11. _____ communication modules are usually integrated directly into target devices, such as automated meter readers (AMRs), vending machines, alarm systems, surveillance cameras, and automotive equipment, to list a few. **[1]**
A. H2H
B. M2M
C. MiH
D. M2H

Q.12. ITU –T views about IOT is **[1]**
A. IoT is just a concept
B. IoT is an infrastructure
C. Both a and b
D. None of above

Q.13. IoT was originally introduced by the Auto-ID research center at the **[1]**
A. Google
B. Bell Labs
C. MIT
D. Accenture Labs

Q.14. ITU view of ubiquitous networking contains **[1]**
A. anywhere connectivity
B. anytime connectivity
C. anything connectivity
D. all of the above

Q.15. Glucose sensor installed in body of human to get blood sugar level is the example of_____. **[1]**
A. H2H
B. M2M
C. MiH
D. M2H

Q.16. Temperature sensors installed in a server room transfer the temperature data to the server for controlling air conditioners automatically. This the example of _____. **[1]**
A. H2H
B. M2M
C. MiH
D. M2H

Q.17. e-Health applications include _____. **[1]**
A. health and fitness
B. road safety
C. Privacy and security
D. All of above

Q.18. What is MBANs? [1]
    A. Mobile Body Area Networks
    B. Medical Body Area Networks
    C. Mobile Body Access Networks
    D. Medical Body Access Networks.

Q.19. What is WBANs [1]
    A. Wireless Body Area Networks
    B. Wireless Body Access Networks.
    C. Wireless Body Access Networks
    D. Wireless Body Area Networks

Q.20. The bit length of the IPV6 address is _____ [1]
    A. 8 bit
    B. 16 bit
    C. 64 bit
    D. 128 bi

Q.21. Identification codes can be classified as : [1]
    A. object IDs (OIDs)
    B. communication IDs
    C. Both of above
    D. None of above

Q.22. Which of the following is not a structural aspect of IoT? [1]
    A. Traffic characteristic
    B. Scalability
    C. Interoperability
    D. Consistency

Seat No.	

January - February (Winter) Examination - 2023

Subject Name: B.Tech. CBCS_86854_Internet of Things_17.01.2023_10.30 AM To 01.00 PM

Subject Code: 86854

Day and Date: Tuesday, 17-01-2023
Time: 10:30 am to 01:00 pm

Total Marks: 70

Instructions.:

      1) All questions are compulsory
      2) Figures to the right indicate full marks

Q.1. Explain Bluetoth Network, How many maximum bluetooth devices can participate in single piconet ? **[7]**

Q.2. What is IoT? Describe the characteristics of the IoT system **[7]**

Q.3. Compare Piconet and Scatternet. **[7]**

Q.4. Explain Identification Technology in IoT. **[7]**

Q.5. Explain Sensor technology in detail? **[7]**

Q.6. What are the challenges faced by a modern WSN? Explain in detail **[7]**

Q.7. How IoT can be used to develop Smart City? Explain with suitable example **[7]**

Q.8. Explain Zigbee/IEEE 802.15.4 for IoT? **[7]**

Q.9. A _____ is an infrastructure comprising sensing (measuring), computing, and communication elements that gives the administrator the ability to the instrument, observe, and react to events and phenomena in a specified environment. **[1]**
A. device power
B. sensor network
C. device intelligence
D. internet of things

Q.10. Which of the following is the correct sentence **[1]**
A. Object Name Service (ONS) is a mechanism that leverages Domain Name System (DNS) to discover information about a product and related services from the Electronic Product Code (EPC).
B. object name service (ONS) will also be important in the IoT to map the "thing-friendly" names of objects which may belong to heterogeneous name spaces
C. Both a and b
D. None of above

Q.11. RFID is _____. **[1]**
A. Radio Frequency Identification
B. Random Frequency Identification
C. both of above
D. None of above

Q.12. Which of the following is not a component of RFID? **[1]**
A. Transmitter
B. Receiver
C. Controller
D. Pre-processor

Q.13. Due to its global reach and the ability to support mobility in all geographical environments, _____ can play a critical role in many broadly distributed M2M applications. **[1]**
A. satellite communications
B. sensor networks
C. RFIDs
D. Bluetooth networks

Q.14. WSNs stands for _____. **[1]**
A. Wired Sensor Networks
B. Wireless Sensitive Networks
C. Wireless Sensor Networks
D. None of above

Q.15. The sensor nodes are _____ **[1]**
A. Typically small in size
B. Consumes less power
C. Cost is low
D. All of the above

Q.16. What is the Ethernet/LAN cable used in RPi? **[1]**
a) Cat5
b) Cat5e
c) Cat6
d) RJ45

Q.17. **What is the default user in Debian on Raspberry Pi?** **[1]**
 a) Default
 b) User
 c) Pi
 d) Root

Q.18. **What is GPIO pins in Raspberry Pi ?** **[1]**
 a) General Purpose Input Output Pins
 b) Gnome Process Inode Output
 c) GNU Process Input Output
 d) General Purpose Interactive OS

Q.19. **Which Raspberry Pi port is used to plug into a monitor or modern** **[1]**
 television?
 a) Ethernet port
 b) HDMI port
 c) Micro USB power port
 d) None of the above

Q.20. **Following are examples of PANs (Personal Area Networks)** **[1]**
 A. ZigBee
 B. Bluetooth (BLE)
 C. NFC
 D. All of the above

Q.21. **A _____is a network used for communication among intelligent** **[1]**
 devices physically close to a person (including smartphones, tablets,
 body monitors, and so on)
 A. PAN / WPAN
 B. WAN
 C. MAN
 D. LAN

Q.22. **In WBAN, the key wireless standard includes -** **[1]**
 A. ZigBee (IEEE 802.15.4)
 B. Bluetooth (IEEE 802.15.1)
 C. IEEE 802.15.6 and IEEE 802.15.4
 D. All of the Above

Seat No.

B.Tech. (Data Science) (Part - III) (Semester - V) (CBCS) Examination, November - 2023
COMPUTER SCIENCE ENGINEERING
Internet of Things
Sub. Code : 86854

Day and Date : Thursday, 30 - 11 - 2023   **Total Marks : 70**

Time : 10.30 a.m. to 01.00 p.m.

Instructions :
1) **All questions are compulsory.**
2) **Figures to the right indicate full marks.**
3) **Assume suitable data wherever necessary and mention It boldly.**

Q1) **Solve the MCQs :**   **[14]**

1. Which of the following IEEE standard provides the lowest level of the ZigBee software stack?
   a) IEEE 802.11   b) IEEE 802.15.1
   c) IEEE 802.15.3   d) IEEE 802.15.4

2. A _____ is a network used for communication among intelligent devices physically close to a person (including Smartphone's, tablets, body monitors, And so on)
   a) PAN / WPAN   b) WAN
   c) MAN   d) LAN

3. Raspbian is _____
   a) Assembler   b) Language
   c) Compiler   d) OS

4. What is GPIO pins in Raspberry Pi?
   a) General Purpose Input Output Pins
   b) Gnome Process Inode Output
   c) GNU Process Input Output
   d) General Purpose Interactive OS

P.T.O.

5. A radio transceiver in WSN (node) contains which component internally?

   a) Antenna   b) Wire

   c) Electrode   d) None of the above

6. WSNs stand for _____

   a) Wired Sensor Networks   b) Wireless Sensitive Networks

   c) Wireless Sensor Networks   d) None of above

7. Which of the following is not a component of RFID reader?

   a) Microcontroller   b) Microprocessor

   c) RF signal generator   d) RF signal detector

8. Which of the following is a standard for RFID?

   a) IEEE 802.15.11   b) IEEE 802.15.4

   c) ISO 14443   d) None of above

9. Which of following is a structural aspect of IoT?

   a) Traffic characteristic   b) Scalability

   c) Interoperability   d) All of the above

10. A _____ is an infrastructure comprising sensing (measuring), computing and communication elements that gives the administrator the ability to instrument, observe and react to events and phenomena in a specified environment.

   a) device power   b) sensor network

   c) device intelligence   d) internet of things

11. The bit length of the IPV6 address is

   a) 8 bit   b) 16 bit

   c) 64 bit   d) 128 bit

12. e-Health applications include _____

   a) Health and fitness   b) Road safety

   c) Privacy and Security   d) All of above

13. ITU view of ubiquitous networking contains

   a)   anywhere connectivity   b)   anytime connectivity

   c)   anything connectivity   d)   all of the above

14. What is EPC in IoT?

   a)   Electrical Process Component

   b)   Electronic Process Code

   c)   Electrical Process Code

   d)   Electronic Product Code

Q2) **Solve any 2 (Two) of the following questions (7 Marks Each) :**   **[14]**

   a)   Short notes on : Machine to Machine (M2M) communication.

   b)   What is EPC? How EPC is used in RFID/sensor?

   c)   With the help of neat diagram explain RFID middleware.

Q3) **Solve any 2 (Two) of the following questions (7 Marks Each) :**   **[14]**

   a)   Explain Identification Technology in IoT.

   b)   Draw a neat diagram RFID reader and explain its operation?

   c)   Short notes on : Communication in Wireless sensor network.

Q4) **Solve any 2 (Two) of the following questions (7 Marks Each) :**   **[14]**

   a)   Explain major hardware and software components of Raspberry Pi.

   b)   Explain Zigbee/IEEE 802.15.4 for IoT?

   c)   Explain IoT automotive and transportation applications in brief.

Q5) **Solve any 2 (Two) of the following questions (7 Marks Each) :**   **[14]**

   a)   Explain the I/O Ports (input output ports) of Raspberry Pi.

   b)   Explain LTE in details.

   c)   How IoT is useful for development of Smart City?

□□□

Seat No.

Summer Examination March - 2023

Subject Name: B.Tech. CBCS_86857 - Information Security_30.06.2023_02.30 PM To 05.00 PM

Subject Code: 86857

Day and Date: - Friday, 30-06-2023
Time: - 02:30 pm to 05:00 pm

Total Marks: 70

Instructions.:

      1) All questions are compulsory
      2) Figures to the right indicate full marks
      3) Assume suitable data wherever necessary and mention it boldly

---

Q.1.   **Rewrite the sentence with correct answer (1 Mark each).**   **[14]**

1.In RSA, $\Phi(n)$ = _____ in terms of p and q.

A.(p)/(q)   **B.**(p)(q)

C.(p-1)(q-1)   **D.**(p+1)(q+1)

2.... is a collection facility can be implemented that generates audit records containing only that information required by the intrusion detection system.

A.Native audit records   **B.**Detection-specific audit records.

C.Both A & B   **D.**None of the above

3.Firewall examines each that are entering or leaving the internal network

A.emails users   **B.**updates

C.connections   **D.**data packets

4.DoS attack coming from a large number of IP addresses, making it hard to manually filter or crash the traffic from such sources is known as a ____

A.GoS attack   **B.**PDoS attack

C.DoS attack   **D.**DDoS attack

5.In Rail fence of depth 2, the ciphertext of Plain Text :- "MY COLLEGE IS ABCD".Cipher Text is:-____

A.MCLEESDBYOLGIAC   **B.**MCLEESBDOYLGIAC

C.MCLEESBDYOLGIAC   **D.**None of these

6.The DES algorithm has a key length of......

A.128 Bits   **B.**32 Bits

C.64 Bits   **D.**16 Bits

7.In cryptography, the order of letters in a message is rearranged by ____

A.Substitution ciphers   **B.**Transpositional ciphers

C.Both (A) and (B)   **D.**Quadratic ciphers

8. In Kerberos , TGS requests _____ from user.

A. Service granting ticket **B.** Ticket granting ticket

C. Ticket granting token **D.** Token gaining ticket

9. In RSA algorithm If two prime numbers are 11 and 19 then value of n will be___

A. 309 **B.** 209

C. 180 **D.** 187

10. When a hash function is used to provide message authentication, the hash function value is referred to as

A. Message Field **B.** Message Digest

C. Message Score **D.** Message Leap

11. PGP makes use of which cryptographic algorithm?

A. DES **B.** AES

C. RSA **D.** ROUND ROBIN

12. In Playfair cipher technique combining i&j, if Plain Text= INSTRUMENT and Keyword = MONARCHY, Cipher Text is:-_____

A. GATPMZCKRQ **B.** GATLMZCLRQ

C. GATRMZCLRQ **D.** GATLMZCNRQ

13. ....is a more advanced technique/s to get users' credentials by making effort to enter users into the website.

A. Phishing **B.** Pharming

C. Both A and B **D.** None the above

14. A firewall protects which of the following attacks?

A. Phishing **B.** dumpster diving

C. denial of service **D.** shoulder surfing

Q.2. **Solve any TWO of following questions (7 marks each)** **[14]**
a) What is a Transposition cipher? Explain Rail fence technique with an example.
b) Explain General Depiction of DES Encryption Algorithm.
c) Explain RSA algorithm with an example ?

Q.3. **Solve any TWO of following questions (7 marks each)** **[14]**
a) State the requirements of MAC. What are the basic uses of MAC ?
b) Explain RSA and DSS approaches to digital signature?
c) Explain X-509 certificate format in detail with diagram.

Q.4. **Solve any TWO of following questions (7 marks each)** **[14]**
a) Explain Tunnel and Transport mode of IP Security?
b) What is PGP? Explain various services provided by PGP.
c) What is Intruder? What are different classes of intruders and explain it with an example?

Q.5. **Solve any TWO of following questions (7 marks each)** **[14]**
   a) Explain SSL Architecture with suitable diagram.
   b) What is Phishing and Pharming? Differentiate between Phishing and Pharming.
   c) What is DoS and DDoS attack? Explain Dos and DDoS attack.

Seat No.

January - February (Winter) Examination - 2023

Subject Name: B.Tech. CBCS_86857_Information Security_27.01.2023_10.30 AM To 01.00 PM

Subject Code: 86857

Day and Date: Friday, 27-01-2023
Time: 10:30 am to 01:00 pm

Total Marks: 70

Instructions.:

      1) All questions are compulsory
      2) Figures to the right indicate full marks
      3) Assume suitable data wherever necessary and mention it boldly

Q.1. Rewrite the sentence with correct answer (1 Mark each). **[14]**

(i). In Rail fence of depth 2, the ciphertext of Plain Text :- "MY COLLEGE IS DYPSN".

Cipher Text is:-_____

A) MCLEESYYSOLGIDPN

B) MCLESEYSYOLGIDPN

C) MCLEESYSYOLGIDPN

D) None of these

(ii). The DES algorithm has a key length of......

A) 128 Bits

B) 32 Bits

C) 64 Bits

D) 16 Bits

(iii). In cryptography, the order of letters in a message is rearranged by _____

A) Substitution ciphers

B) Transpositional ciphers

C) Both (A) and (B)

D) Quadratic ciphers

(iv). In RSA algorithm If two prime numbers are 11 and 19 then value of n will be_____

A) 309

B) 187

C) 180

D) 209

(v). When a hash function is used to provide message authentication, the hash function value is referred to as

A) Message Field

B) Message Digest

C) Message Score

D) Message Leap

(vi). In Kerberos , TGS requests _____ from user.

A) Service granting ticket

B) Ticket granting ticket

C) Ticket granting token

D) Token gaining ticket

(vii). In X.509 certificate , _____ Consists of two dates: the first and last on which the certificate is valid.
A) Signature algorithm identifier
B) Period of validity
C) Issuer unique identifier
D) Subject unique identifier
(viii). PGP makes use of which cryptographic algorithm?
A) DES
B) AES
C) RSA
D) ROUND ROBIN
(ix). IPSec is designed to provide security at the
A) Transport layer
B) Network layer
C) Application layer
D) Session layer
(x). Fundamental tool for intrusion detection is ……………
A) Audit record.
B) Password management.
C) Both A & B
D) None of the above
(xi). SSL primarily focuses on
A) confidentiality and integrity
B) integrity and non-repudiation
C) authenticity and privacy
D) integrity and authenticity
(xii). A firewall protects which of the following attacks?
A) Phishing
B) dumpster diving
C) denial of service
D) shoulder surfing
(xiii)…………………..is a more advanced technique/s to get users' credentials by making effort to enter users into the website.
A) Phishing
B) Pharming
C) Both A and B
D) None the above
(xiv). In Playfair cipher technique combining i&j, if Plain Text= INSTRUMENT and Keyword = MONARCHY,
Cipher Text is:-_____
A) GATPMZCKRQ
B) GATLMZCNRQ
C) GATRMZCLRQ
D) GATLMZCLRQ

Q.2. **Solve any TWO of following questions (7 marks each)** **[14]**
a) Which parameters and design choices determine the actual algorithm of a Feistel cipher explain with diagram?
b) What is a Substitution cipher? Explain Playfair cipher with an example.
c) What are three broad categories of applications of public-key cryptosystems. Explain each category with neat diagram?

Q.3. **Solve any TWO of following questions (7 marks each)** **[14]**
a) What is message authentication? How to achieve message authentication using hash function.
b) What is Kerberos Realm ? Explain Kerberos Inter-realm Authentication in detail.
c) Explain X-509 certificate format in detail with diagram.

Q.4. **Solve any TWO of following questions (7 marks each)** **[14]**
a) What is PGP? Explain various services provided by PGP.
b) Explain ESP (Encapsulating Security Payload) format in IPSec in detail.
c) Explain SSL Architecture with suitable diagram.

Q.5. **Solve any TWO of following questions (7 marks each)** **[14]**
a) Explain SSL Handshake protocol.
b) What is DoS and DDoS attack? Differentiate between Dos and DDoS attack.
c) What is a Pharming attack? Explain Pharming attack in detail.

Summer Examination March - 2023

Subject Name: B.Tech. CBCS_86876 - Machine Learning_01.07.2023_10.30 AM To 01.00 PM

Subject Code: 86876

Day and Date: - Saturday, 01-07-2023
Time: - 10:30 am to 01:00 pm

Total Marks: 70

Instructions.:
      1) All questions are compulsory
      2) Figures to the right indicate full marks
      3) Assume suitable data wherever necessary and mention it boldly

Q.1.   **Solve the MCQ**   **[14]**
1) Which ML algorithm is suitable when we want to predict any continuous value?
A) Classification.
B) Regression.
C) Clustering.
D) None of the above
2) Cleaning of Data is done in _____
A) Data Collection
B) Data Preparation.
C) Data Splitting.
D) Data Testing.
3) Which of these are classification tasks?
A) Find the gender of a person by analyzing his writing style.
B) Predict whether there will be abnormally heavy rainfall next year.
C) Both a & b.
D) None of above.
4) Suitable evaluation metric for measuring the performance of a given regression model is ?
A) Mean absolute error
B) Root mean square error
C) Both a and b
D) None of above
5) What type of machine learning is suitable for predicting the dependent
   variables with two different values?
A) Logistic Regression
B) Linear Regression
C) Multiple linear Regression
D) Polynomial Regression
6) Which of following are categorical features?
A) Height of a person
B) Price of petroleum
C) Mother tongue of a person
D) Amount of rainfall in a day

7) Appropriate chart for visualizing the linear relationship between two variables is....
A) Scatter plot
B) Bar Chart
C) Histogram
D) None of the above.

8) _____gives the rate of speed where the gradient moves during gradient descent.
A) Learning rate
B) Cost Function
C) Hypothesis Function
D) None of above

9) What is formula to calculate error of single data point?
A) Actual value – Predicted value.
B) Actual value + Predicted value.
C) Predicted Value – Actual Value.
D) Predicted Value + Actual Value

10) --------is used to optimize the cost function or the error of the model.
A) Gradient Descent Algorithm
B) Hypothesis Function
C) Both a and b
D) None of above

11) Classification is what type of machine learning technique?
A) Supervised
B) Unsupervised
C) Both a and b
D) None of above

12) Pruning is
A) Removing of unwanted branches of the tree.
B) Formed by splitting of Tree
C) Dividing the root node in to different parts.
D) Roots divided into homogeneous sets

13) Neural networks can be used in different fields. such as –
A) Classification
B) Data processing
C) Compression.
D) All of the above

14) What are different Recommendation Engine techniques?
A) Content based filtering
B) Collaborative filtering
C) Knowledge based system
D) All of the above

Q.2. **Solve any 2 (Two) of the following questions (7 Marks Each)**   **[14]**
 a) What is Machine Learning? Explain any two types of Machine Learning.
 b) Describe Various Machine Learning Problem Categories?
 c) Illustrate the working of the Confusion Matrix with examples?

Q.3. **Solve any 2 (Two) of the following questions (7 Marks Each)** **[14]**
a) Describe Simple Linear Regression? Explain various assumptions of Linear Regression
b) Describe the hypothesis and Cost Function of Linear Regression with equations.
c) Differentiate between Linear Regression and Logistic Regression.

Q.4. **Solve any 2 (Two) of the following questions (7 Marks Each)** **[14]**
a) What is Decision Tree? Describe various terminologies related with Decision Tree.
b) What is Bayes Theorem? Explain the working of Naïve Bayes Classifier.
c) Construct Agglomerative Hierarchical clustering by mentioning all the steps with neat diagram.

Q.5. **Solve any 2 (Two) of the following questions (7 Marks Each)** **[14]**
a) Describe the working of K-Means Clustering with example?
b) What is Regular Expression (RegEx)? Analyze various built-in modules of Regular Expression.
c) What is the Recommendation Engine? Explain types of recommendation Engine?

Summer Examination March - 2023

Subject Name: B.Tech. CBCS_86874 - UNIX Shell Programming_27.06.2023_10.30 AM To 01.00 PM
Subject Code: 86874

Day and Date: - Tuesday, 27-06-2023   Total Marks: 70
Time: - 10:30 am to 01:00 pm

Instructions.:

      1) All questions are compulsory
      2) Figures to the right indicate full marks
      3) Assume suitable data wherever necessary and mention it boldly

---

Q.1. **Solve Following MCQs (All Questions are compulsory) (2**   **[14]**
marks each)
1.Which part of the Unix operating system interacts with the hardware?
a) Application program
b) Vi editor
c) Shell
d) Kernel

2. Which command is used to display the documentation of commands
in Unix?
a) man
b) whatis
c) help
d) search

3. Which Unix command is used for changing the current directory?
a) pwd
b) rm
c) cd
d) cp

4. Which command is used for printing the current working directory?
a) dir
b) HOME
c) cd
d) pwd

5. Which command is used for listing files in a directory?
a) list
b) ls
c) wc
d) ps

6. Which command is used by the user to change their login password in Unix?
a) cp
b) man
c) reset
d) passwd

7. Which command is used for displaying the contents of a file in Unix?
a) mkdir
b) cat
c) rm
d) cp

Q.2. **Solve any two (7 marks each)** **[14]**
A. Draw and Explain Block Architecture of UNIX.

B. Describe features of the Unix Operating System in detail?

C. Brief about the commands used in the vi Editor.

Q.3. **Solve any two (7 marks each)** **[14]**
A. Explain the security levels provided in the Unix environment. How to change permissions of a file.

B. What is the user and group in Unix? Explain the related commands for changing ownership and group.

C. Write about the operations that can be performed on both directories and file.

Q.4. **Solve any two (7 marks each)** **[14]**
A. What is inode and what it contains? Explain them in detail.

B. What is Pipe, explain in detail with types, usage and syntax.

C. In Unix, What are different data structures associated with the file?

Q.5. **Solve any two (7 marks each)** **[14]**
A. Write a shell program for counting characters, words and line

B. Explain Unix init process.

C. What information is presented when the following commands are entered?
 (a) cmp (b) diff (c) comm (d) cut (e) paste (f) pwd (g) stat

Seat No.

B.Tech. (Computer Science and Engineering) (Data Science) (Part - III) (Semester - VI) (CBCS) Examination, November - 2023 UNIX SHELL PROGRAMMING Sub. Code : 86874

Day and Date : Tuesday, 21 - 11 - 2023   Total Marks : 70

Time : 2.30 p.m. to 5.00 p.m.

Instructions : 1)   All Questions are compulsory.

2)   Assume suitable data wherever necessary.

3)   Figures to the right indicate full marks.

Q1) Solve MCQs.   **[14]**

i)   In UNIX, which command is used to change the permissions of a file?

a)   chown   b)   chgrp

c)   chmod   d)   ch

ii)   Which of the following files are known as special files in Unix?

a)   /dev/null   b)   sample.txt

c)   /dev/null and /dev/tty   d)   /dev/tty

iii)   In Unix, for what purpose be command is used?

a)   for copying files   b)   as a process table viewer

c)   as an editor   d)   as a calculator

iv)   Which command is used for renaming files in Unix?

a)   rename   b)   move

c)   mv   d)   cp

v)   What is the function of the cp command in Unix?

a)   delete a given file

b)   change the directory

c)   list all the available files in the current directory

d)   cp is a command used for copying files and directories

P.T.O.

vi) Which command is used to delete all files in a directory?

    a)   del *   b)   rm*

    c)   mv*   d)   rmdir*

vii) echo command is used for _____

    a)   displaying dialogue messages

    b)   displaying date and time

    c)   displaying errors

    d)   displaying operating system details

Q2) Solve any 2 of the following.   **[14]**

    a)   Draw and Explain Block Architecture of UNIX

    b)   Describe following Unix commands with syntax:

        i)   mkdir

        ii)   cat

        iii) pwd

        iv) cd

        v)   ls

        vi) chmod

    c)   What is mail? Explain it in detail

Q3) Solve any 2 of the following.   **[14]**

    a)   How Unix maintains a File on Hard Disk. How dynamic growth of file size is supported by the Unix File System?

    b)   Explain (i) Aliases (ii) Unix session

    c)   How does input, output redirection operators and pipe operators increase efficiency of Unix? Explain with suitable examples.

Q4) Solve any 2 of the following.   **[14]**

    a)   What is inode and what it contains? Explain them in detail.

    b)   What is Pipe, explain in detail with types, usage and syntax.

    c)   Describe exec system call and how it is different from fork system call?

Q5) Solve any 2 of the following.   **[14]**

    a)   Write a shell script that will do all arithmetic operations on two numbers and operations(+,–,*,/) given by the user.

    b)   Explain Signals in Unix. How signals are handled in Unix?

    c)   What information is presented when the following commands are entered?

        i)   cmp   ii)   diff

        iii)   comm   iv)   cut

        v)   paste   vi)   pwd

        vii)   stat

$$\bigcirc\bigcirc\bigcirc\bigcirc$$

Summer Examination March - 2023

Subject Name: B.Tech. CBCS_86873_86873 - Computer Algorithms_23.06.2023_10.30 AM To 01.00 PM

Subject Code: 86873

Day and Date: - Friday, 23-06-2023
Time: - 10:30 am to 01:00 pm

Total Marks: 70

Instructions.:

1) All questions are compulsory
2) Figures to the right indicate full marks
3) Use of calculator and statistical table is allowed
4) Assume suitable data wherever necessary and mention it boldly

---

Q.1. **Solve Following MCQs (All Questions are compulsory) (2 Marks Each)**   **[14]**

1. In Depth First Search _____ Data structure is used
a) Stack
b) Queue
c) Both
d) None

2. An algorithm is _____?
a)A problem
b) a procedure for solving a problem
c) A real-life mathematical problem
d) none

3. What should be considered when designing an algorithm?
a)If this software is used correctly
b)In the hardware is used correctly
c)If there is more than one way to solve the problem
d)All of the above are correct

4. Job Sequencing with deadline problem is a maximization problem.
a) True
b) False
c) None
d) Both

5. Which of the following algorithm is polynomial time algorithm?
a)Graph Coloring
b)Binary Search
c)Sum of subsets
d)0/1 Knapsack

6. For the Quick sort algorithm, what is the time complexity of the best/worst case?
a)best case: O(log(n)) worst case: O(n2)
b)best case: O(n2), worst case: O(n log(n))
c)best case: O(n log(n)) worst case: O(n2)
d)best case: O(n2), worst case: O(n2log(n))

7. The time factor when determining the efficiency of algorithm is measured by
a) Counting microseconds
b) Counting the number of key operations
c)Counting the number of statements
d)Counting the kilobytes of algorithm

Q.2. **Solve any Two questions (7 Marks Each)**   **[14]**
1.Explain General divide and conquer method with control abstraction and example.

2. Explain Performance analysis and performance measurement.

3.Define Spanning Tree. Explain Prim,s algorithm to find minimum cost spanning tree.

Q.3. **Solve any Two questions (7 Marks Each)**   **[14]**
1. Draw and explain permutation tree for 4-queen problem using backtracking.

2. Explain BFS and DFS with suitable example.

3. Apply dynamic programming method to find minimum cost of path from S-T in the multistage graph of following fig., Using Backward approach.
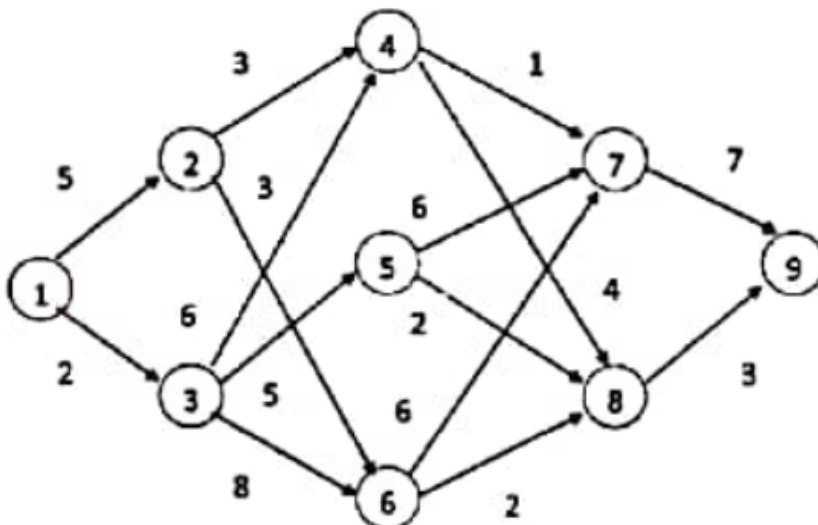
Q.4. **Solve any Two questions (7 Marks Each)** **[14]**

1. Find Optimal solution to given Knapsack problem using Greedy Method, where
   Knapsack capacity, m= 20 and N=6 , profit and wieght of objects is as follows,
   (P1,P2,P3,P4,P5,P6)= (12,5,15,7,6,18)
   (W1,W2,W3,W4,W5,W6)= (2,3,5,7,1,5)

2. Explain Data Concentration with mesh and hypercube.

3. Write Short note on:
A) Hamiltonian Cycle
B) Butterfly network
C) Optimal Merge Pattern

Q.5. **Solve any Two questions (7 Marks Each)** **[14]**

1. Explain Pre-order, In-order and Post-order Traversal Techniques for binary tree.

2. Explain Hypercube Computational Model.

3. Define the following terms:
   A) Deterministic and non-deterministic algorithm
   B) Decision and optimization problem
   C) P and NP problems

Seat
No.

T.Y. B.Tech. (Computer Science and Engineering) (Data Science) (Part - III) (Semester - VI) (CBCS) Examination, November - 2023
COMPUTER ALGORITHM
Sub. Code : 86873

Day and Date : Monday, 20 - 11 - 2023　　　　　　　**Total Marks : 70**
Time : 02.30 p.m. to 05.00 p.m.

Instructions :　　1)　**All questions are compulsory.**
　　　　　　　　　2)　**Assume suitable data wherever necessary.**
　　　　　　　　　3)　**Figures to the right indicate full marks.**

Q1) **Solve MCQs (2 marks each) :**　　　　　　　　　　　　　　**[14]**

　i)　When an algorithm is written in the form of a programming language, it becomes a _____?

　　　a)　Flowchart　　　　　　　　b)　Program

　　　c)　Pseudo code　　　　　　　d)　Syntax

　ii)　What is the best time complexity of all-pairs shortest paths in a weighted graph?

　　　a)　O(n^3)　　　　　　　　　　b)　O(n^2)

　　　c)　O(n)　　　　　　　　　　　d)　O(n^4)

　iii)　Which of the following algorithms are used to find the shortest path from a source node to all other nodes in a weighted graph?

　　　a)　BFS　　　　　　　　　　　b)　Djikstra's Algorithm

　　　c)　Prims algorithm　　　　　　d)　Kruskal's algorithm

　iv)　Hamiltonian path problem is _____?

　　　a)　NP problem　　　　　　　　b)　P class problem

　　　c)　NP-complete problem　　　　d)　N class problem

P.T.O.

v) Which of the following is a Divide and Conquer algorithm?

    a) Bubble sort   b) Selection Sort

    c) Heap Sort   d) Merge Sort

vi) Binary Search Method has Worst Case Time Complexity of

    a) $2n$ b) $n\log(n)$

    c) $\log(n)$ d) $n*n$

vii) Which of the following is used for solving the Job Sequencing with deadline Problem?

    a) Greedy Algorithm   b) Dynamic Programming

    c) Backtracking   d) Sorting

Q2) **Solve any 2 of the following (7 marks each) :**   **[14]**

    a) Explain Time Complexity of an algorithm. Explain step count and step table method for the example of matrix addition.

    b) Write a Short note on :

        i) Single Source Shortest Path

        ii) Optimal Merge Pattern

    c) Define Knapsack Problem. Consider the knapsack instance n = 3, (w1 ,w2, w3) = (2, 3, 4) and (p1, p2, p3) = (1, 2, 5).

    Knapsack capacity, m = 6. Solve this instance using dynamic programming method.

Q3) **Solve any 2 of the following (7 marks each) :**   **[14]**

    a) Define and explain asymptotic notations with example.

    b) Define job sequencing with deadline using Greedy approach. Solve following instance,

    N = 5, (P1, P2, P3, P4, P5) = (20, 15, 10, 5, 1)

    (d1, d2, d3, d4, d5) = (2, 2, 1, 3, 3)

    c) Write a short note on :

        i) Reliability Design

        ii) All pair shortest path

Q4) **Solve any 2 of the following (7 marks each) :** **[14]**

a) Explain Depth First Search Algorithm with example.

b) Write a short note on :

i)   Hamiltonian Cycle

ii) Graph Coloring

c) Explain Data concentration with Hypercube.

Q5) **Solve any 2 of the following (7 marks each) :** **[14]**

a) Explain solution to a knapsack problem using backtracking.

b) Explain Pre-order, In-order and Post-order Traversal Techniques for binary tree.

c) Write a short note on :

i)   Butterfly Network

ii) P, NP, NP-Complete and NP-hard Problems

□□□

Seat No.

Summer Examination March - 2023

Subject Name: B.Tech. CBCS_86875 - Database Engineering_02.07.2023_10.30 AM To 01.00 PM

Subject Code: 86875

Day and Date: - Sunday, 02-07-2023
Time: - 10:30 am to 01:00 pm

Total Marks: 70

Instructions.:
       1) All questions are compulsory

Q.1.   **[14]**

I. A relational database consists of a collection of
a) Tables
b) Fields
c) Records
d) Keys

II. Course(course_id,sec_id,semester) Here the course_id,sec_id and semester are _____ and course is a _____
a) Relations, Attribute
b) Attributes, Relation
c) Tuple, Relation
d) Tuple, Attributes a)

III. Which one of the following is a procedural language?
a) Domain relational calculus
b) Tuple relational calculus
c) Relational algebra
d) Query language

IV. The file organization which allows us to read records that would satisfy the join conditionby using one block read is
a) Heap file organization
b) Sequential file organization
c) Clustering file organization
d) Hash file organization

V. If a transaction has obtained a _____ lock, it can read but cannot write on the item
a) Shared mode
b) Exclusive mode
**c) Read only mode**
d) Write only mode Data and the database

VI. Which of the following are introduced to reduce the overheads caused by the log-based recovery?
a) Indices
b) Checkpoints
c) Locks
d) Stable Storage

VII. If a failure has occurred in the midst of a transfer, it is called as
_____
 a) Successful completion
 b) Partial failure
 c) Total failure
 d) None of the mentioned

Q.2. **1. Reduce the below ERD to relational schema.** **[14]**
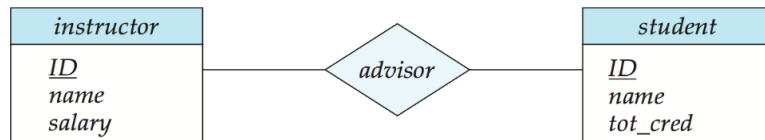
2.List and explain aggregate functions of SQL with appropriate examples.
3.Give difference between Trigger and stored procedure.

Q.3. **1. List and explain different types of attributes with appropriate** **[14]**
examples for each.
2. List and explain the desired properties of decomposition.
3. Assume the Relations given below.
Student(Enrno, name, courseId, emailId, cellno) Course(courseId, course_nm, duration)
Write SQL statements for following
1. Find out list of students who have enrolled in "computer" course. (2)
2. List name of all courses with their duration.   (1)
3. List name of all students start with "s".   (2)
4. List email Id and cell no of all Civil engineering students.   (2)

Q.4. **1. Define the terms Primary Index and Secondary Index. Differentiate** **[14]**
between them on basis of the Evaluation Criteria for indices
2. Draw & Explain state of transaction.
3. List and elaborate the Drawbacks of Shadow Paging.

Q.5. **1. What is transaction? Explain its ACID properties** **[14]**
2. State and explain various classes of failure in database system.
3. Explain log based recovery technique.

Seat No.

Summer Examination March - 2023

Subject Name: B.Tech. CBCS_86858 - Data Mining & Data Warehousing_26.06.2023_02.30 PM To 05.00 PM

Subject Code: 86858

Day and Date: - Monday, 26-06-2023   **Total Marks: 70**
Time: - 02:30 pm to 05:00 pm

Instructions.:
>   1) All questions are compulsory
>   2) Figures to the right indicate full marks

Q.1. **Rewrite the sentence with correct answer**   **[14]**
i) .....   is a comparison of the general features of the target class data objects against the general features of the target
   class data objects against the general features of objects from one or more contrasting classes.
A)Data Discrimination
B)Data Characterization
C)Data Integration
D)Classification
ii) Which of the following is not data preprocessing step?
A)Data Mining
B)Data Integration
C)Data Cleaning
D)Data selection
iii) In .... data encoding schemes are applied so as to obtain a reduced or compressed representation of original data..
A)Dimensionality Reduction
B)Numerosity Reduction
C)Data Reduction
D)Data Transformation
iv) For nominal attributes .... is used to detect redundancies.
A)Correlation coefficient
B)Covariance
C)Chi square test
D)Both A and B
v) ...... is a visualization operation that rotates the data axes in view to provide an alternative data presentation.
A) Drill Up
B)Drill Down
C)Pivot
D) Slice

vi)The cuboid that holds the lowest level of summarization is called the .....
A)Apex Cuboid
B)Base Cuboid
C)Monoid
D)Vertex Cuboid

vii) In starnet query model each abstraction level in a hierarchy is called........
A)Star schema
B)Generalization
C)Footprint
D)Specliazation

viii)The attribute with ..... Gini Index is selected as splitting attribute
A)Maximun
B)Minimum
C)Equal
D)None of the above

ix) Specificity is also referred as...
A)True Negative Rate
B)True Positive Rate
C)Misclassification rate
D)Resubstituting rate

x) Which of the following is top down hierarchical clustering method?
A)Agglomerative
B)Boosting
C)Divisive
D)Ada Boosting

xi) Classification is type of ...
A) Supervised Learning
B) Unsupervised Learning
C) Semi Supervised learning
D) Active Learning

xii)Which of the following is grid based clustering method?
A)CLIQUE
B)DBSCAN
C)OPTICS
D)DENCLUE

xiii) Which of the following is operational processing system?
A)OLAP
B)OLTP
C)ROLAP
D)MOLAP

xiv) A .... contains a subset of corporate-wide data that is of value to a specific group of users.
A)Datawarehouse
B)Data Mart
C)Enterprise warehouse
D)Virtual warehouse

Q.2. **Solve any TWO of the following questions.** **[14]**
a) What kinds of patterns can be mined in data mining?
b) What do you mean by data mining? Explain various knowledge discovery processes.
c) Write short note on: Data Integration

Q.3. **Solve any TWO of following questions (7 marks each)** **[14]**
a) What is noise? Which methods are used in data smoothing? Explain with example.
b) How indexing takes place in OLAP explain with example.
c) Explain different schemas used in multidimensional data models with the help of diagram.

Q.4. **Solve any TWO of following questions (7 marks each)** **[14]**
a) Explain various metrics used for evaluating classifier performance.
b) State Bayes Theorem. Explain working of naïve bayes classifier with the help of example.
c) State various density based clustering methods. Explain any one method in detail.

Q.5. **Solve any TWO of following questions (7 marks each)** **[14]**
1. Explain different operations used in online analytical processing with diagram.
2. Explain how text data and cyber physical system data is mined?
3. How similarity, regression and trend analysis takes place in time series data? Explain with example.

Seat No.

January - February (Winter) Examination - 2023

Subject Name: B.Tech. CBCS_86858_Data Mining & Data Warehousing_19.01.2023_10.30 AM To 01.00 PM

Subject Code: 86858

Day and Date: Thursday, 19-01-2023
Time: 10:30 am to 01:00 pm

Total Marks: 70

Instructions.:

     1) All questions are compulsory
     2) Figures to the right indicate full marks

Q.1. **Rewrite the sentence with correct answer** **[14]**

i) ..... is a summarization of the general characteristics or features of a target class of data.
A)Data Discrimination
B)Data Characterization
C)Data Integration
D)Classification

ii) .... is the science of searching for documents or information in documents.
 A)Data Mining
 B)Business Intelligence
 C)Information Retrieval
 D)Data Warehousing.

iii) In .... the data are replaced by alternative ,smaller representations using parametric models or nonparametric models.
  A)Dimensionality Reduction
  B)Numerosity Reduction
  C)Data Reduction
  D)Data Transformation

iv)For numeric attributes .... is used to detect redundancies.
A) Correlation coefficient
B) Covariance
C) Chi square test
D) Both A and B

v) ...... defines a sequence of mappings from a set of low level concepts to higher level, more general concepts.
A)Dimension Table
B)Fact table
C)Schema
D)Concept hierarchies

vi)The 0-D cuboid , which holds the highest level of summarization is called the .....
A)Apex Cuboid
B)Base Cuboid
C)Monoid
D)Vertex Cuboid

vii) In ..... Schema some dimension tables are normalized, thereby further splitting the data into additional tables.
  A) Star schema
  B) Snowflake schema
  C) Fact Constellation
  D) Galaxy schema

viii)Which of the following is NOT attribute selection method?
A)Information gain
B)Error Rate
C)Gini Index
D)Gain Ratio

ix) Sensitivity is also referred as...
 A) True Negative Rate
 B) True Positive Rate
 C) Misclassification rate
 D) Resubstituting rate

x) Which of the following is not ensemble method
A)Bagging
B)Boosting
C)Random Forests
D)Ada Boosting

xi)Clustering is also known as...
A)Supervised Learning
B) Unsupervised Learning
C)Semi Supervised learning
D)Active Learning

xii)Which of the following is NOT Density based clustering method?
A)BIRCH
B)DBSCAN
C)OPTICS
D)DENCLUE

xiii) .Which of the following is category of sequence data?
A)Spatial Data
B)Multimedia Data
C)Biological sequences
D)Web Data

xiv)........ is a set of views over operational databases.
A)Datawarehouse
B)Data Mart
C)Enterprise warehouse
D)Virtual warehouse

Q.2.   Solve any two of the following questions.   **[14]**
a)What kinds of data can be mined?
b)Explain major issues in data mining.
c)How correlation analysis is performed on nominal data and numeric data?

Q.3.   Solve any two of following questions.   **[14]**
a)How missing values are handled in data cleaning?
b)Explain architecture of data warehousing in detail.
c)Which operations are performed on OLAP ? Explain each operation with the help of diagram.

Q.4.   Solve any two of following questions.   **[14]**
a)How rule based classification takes place? Explain each method in detail.
b)Explain different techniques to improve classification accuracy
c)Explain k means algorithm with example.

Q.5.   Solve any two of following questions   **[14]**
a)State difference between agglomerative Clustering and Divisive Clustering
b)Explain how spatial data and multimedia data is mined?
c)Explain categories of sequence data.