c) Describe the Logical Database Design. Describe the purposes of key in Database Management System Explain with suitable example.

Q 5. (2X7=14)

a) Explain in detail Nested Queries with suitable example.

b) Describe the purpose of using SQL. Explain DDL, DML and DCL syntax with example.

c) Expain the turm Triggers in SQL. What are advantages and disadvantages of triggers?

CENTRAL UNIVERSITY OF HARYANA

Second Semester Term End Examinations June 2023

Programme: M.Sc Data Science **Session: 2022-23**

Semester: 2nd **Max. Time: 3 Hours**

Course Title: Scalable Database System **Max. Marks: 70**

Course Code: SBS CS 030 208 C 3014

Instructions:

1. Question no. 1 has seven parts and students are required to answer any four. Each part carries three and half Marks.

2. Question no. 2 to 5 have three parts and student are required to answer any two parts of each question. Each part carries seven marks.

Q 1. (4X3.5=14)

a) What are the disadvantages of file server architecture?

b) What do you mean by functional dependency? Explain with example.

c) Define the client server architecture in database management.

d) Differentiate between RDBMS and DBMS.

e) What do you mean by Query Evaluation in SQL?

f) Define primary key and super key with example.

g) What is data replication?

Q 2. (2X7=14)

a) What are the functions of DBA? Also Explain data dictionary.

b) Explain in detail the relational database management system with suitable example.

c) Why we are using DBMS? Difference between File system and DBMS.

Q3. (2X7=14)

a) Construct an ER diagram for Library Database. Identify entities, attributes for each entity, relationship among entities. Represent necessary constraints in this database design process in detail.

b) What is physical and logical data independence? Also Explain the difference between physical and logical data independence.

c) Explain the Normalization with example. How does 2NF differ from 3NF?

Q 4. (2X7=14)

a) What are integrity constraints? Explain various types of integrity constraints with suitable example.

b) Explain in detail Three schema architecture. Also describe its all types of views/level.

Q 4. (2X7=14)

a) What is decision tree? Explain its algorithms and construct decision tree for the following training set.

Day	Outlook	Temperature	Humidity	Wind	Play Tennis
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

b) What is Entropy and Information gain? How information gain is different from entropy?

c) *What is the difference between logistic regression and SVM? Why do we use kernels in SVM?*

Q 5. (2X7=14)

a) What is clustering? Draw a flowchart to show how the K-Mean Clustering algorithm works?

b) Explain the Methods of Hierarchical Clustering. Differentiate between Agglomerative and Divisive Hierarchical Clustering Algorithm?

c) Write Short note (any two):

(i) Stacking

(ii) Bagging

(iii) Boosting

.....

CENTRAL UNIVERSITY OF HARYANA

Term End Examinations July 2023

Programme:	**MSc Data Science**	**Session: 2022-23**
Semester:	**II**	**Max. Time: 3 Hours**
Course Title:	**Machine Learning**	**Max. Marks: 70**
Course Code:	SBS CS 030209 C 3014	

Instructions:

1. Question no. 1 has seven parts and students are required to answer any four. Each part carries three and half Marks.

2. Question no. 2 to 5 have three parts and students are required to answer any two parts of each question. Each part carries seven marks.

Q 1. (4X3.5=14)

a) What are the various types of machine learning problems?

b) Discuss the major drawbacks of K-nearest Neighbour learning Algorithm.

c) What is outlier analysis?

d) Define puring and its type.

e) Why SVM is an example of a large margin classifier?

f) What is the difference between classification and clustering?

g) State the categories of clustering methods?

Q 2. (2X7=14)

a) Explain the k-Means Algorithm with an example.

b) Explain Naïve Bayes Classifier with an Example.

c) How do you classify text using Bayes Theorem?

Q3. (2X7=14)

a) What is linear regression? *How does multicollinearity affect the linear regression?*

b) *How to interpret the results of a logistic regression model? Why can't linear regression be used in place of logistic regression for binary classification?*

c) *Explain gradient descent with respect to linear regression.*

Q4. **(2X7=14)**

a.) What is Real time Analytics Platform (RTAP)? Discuss its applications.

b.) Discuss the Apriori Algorithm through example in detail.

c.) Explain the Stream Data Model its architecture and frequent Item sets in a Stream.

Q 5. **(2X7=14)**

a.) What is Data Visualizations? Discuss Visual Data Analysis Techniques.

b.) Calculate the co-efficient of correlation (r) between age of cars and annual maintenance cost by using **karl Person's method** and comment.

Age of cares in years	2	4	6	7	8	10	12
Annual Maintenance cost in Rs.	1600	1500	1800	1900	1700	2100	2000

c.) Find the "**Support**" and "**Confidence**" between Bread and Cookies, Milk and Juice, Cookies and Coffee, Milk and Eggs with the help of following example.

Transaction Id	**Transaction Time**	**Item Bought**
709	6:10 am	Milk, Bread, Cookies, Juice
710	7:38am	Milk , Juice
711	8:00pm	Milk , Eggs
712	8:47pm	Bread, Cookies, Coffee

CENTRAL UNIVERSITY OF HARYANA

Term End Examinations June 2023

Programme: M.Sc (Data Science) | **Session: 2022-23**
Semester: IInd | **Max. Time: 3 Hours**
Course Title: Big Data Analytics | **Max. Marks: 70**
Course Code: SBS CS 030207 C 3014

Instructions:

1. Question no. 1 has seven parts and students are required to answer any four. Each part carries three and half Marks.

2. Question no. 2 to 5 have three parts and students are required to answer any two parts of each question. Each part carries seven marks.

Q 1. **(4X3.5=14)**

a.) What is ReSampling?

b.) Define Prediction Error.

c.) Discuss Fuzzy Decision Trees.

d.) Explain Stream Computing.

e.) What is Frequent Patterns?

f.) No SQL Databases?

g.) What is unstructured information?

Q 2. **(2X7=14)**

a.) What is Big data? Discuss its application in detail.

b.) Explain the Modern Data Analytic Tools with suitable examples.

c.) What is sampling? Explain its types through examples.

Q 3. **(2X7=14)**

a.) Discuss the Regression Modeling in detail.

b.) What is Neural Network? Discuss the terms Learning and Generalization.

c.) Discuss followings :

- Multivariate Analysis?
- Explain Time Series Analysis with suitable example.

CENTRAL UNIVERSITY OF HARYANA

Second Semester Term End Examination July 2023

Programme: MASTER OF SCIENCE (Data Science) **Session: 2022-23**

Semester: Second **Max. Time: 3 Hours**

Course Title: Data science for deep learning **Max. Marks: 70**

Course Code: SBS CS 030206 E 3014

Instructions:

1. Question no. 1 has seven parts and students are required to answer any four. Each part carries three and half Marks.

2. Question no. 2 to 5 have three parts and student are required to answer any two parts of each question. Each part carries seven marks.

Q 1. (4X3.5=14)

a) Which regularization method leads to weight sparsity?

b) Give two benefits of using convolutional layers instead of fully connected ones for visual tasks.

c) List different applications of artificial neural networks.

d) Before training your model, you want to decide the image resolution to be used. Why is the choice of image resolution important?

e) Why is it important to place non linearity between the layers of neural networks?

f) Define transfer learning.

g) Outline different difficulties of training deep neural networks.

Q 2. (2X7=14)

a) Discuss about different deep convolution neural network architectures.

b) Explain LSTM (Long Short Term Memory). Highlights its applications.

c) Differentiate between Deep and Shallow Network.

Q3. (2X7=14)

a) Elaborate deep learning. Describe its uses, applications, and history.

b) Explain back propagation with its algorithm.

c) Discuss about deep reinforcement learning and recurrent neural network.

Q 4. (2X7=14)

a) Differentiate between machine learning and deep learning.

b) Differentiate between Convolutional neural networks and Generative adversarial networks.

c) Explain about hyper-parameter optimization.

Q 5. (2X7=14)

a) Explain Auto encoders and regularization in details?

b) Illustrate Restrictive Boltzmann Machines (RBMs).

c) Write short notes on belief nets, learning sigmoid belief nets, and deep belief nets.