Questions from the topic - 1. Introduction (2 hours) - Data Mining

1. What is warehouse and data mart? Why do any organizations need warehouse separately though it has transactional database system **- 2066 Bhadra**
2. Explain the general architecture of Warehouse system with its three main phases of development - **2066 Bhadra**
3. What is data mining? Describe the application of data mining in banking in bio-medical. - **2065 Baisakh**
4. Describe the architecture of data warehouse along with different components of it. - **2065 Baisakh**
5. Explain the characteristics of Data Warehouse. How operational database is different from Data Warehouse **- 2065 Baisakh**
6. Write short notes on(any four)
    - Decision Support System
    - Likelihood and Distance Technique
    - Data Warehouse Schema
    - Partitioning and Hierarchical Clustering
    - Noise in Data Warehouse
    - Descision Tree

    - 2065 Baisakh

Questions from the topic - 2. Data Preprocessing (6 hours) - Data Mining

1. How do you relate KDD with data mining? Explain the phases of KDD with example. - **2066 Bhadra**
2. Explain the typical OLAP operations over warehouse multidimensional data. Define the following star schema using DMQL - **2066 Bhadra**
3. Explain the typical OLAP operations over warehouse multidimensional data. Define the following star schema using DMQL - **2066 Bhadra**
4. What is Online Analytical Processing (OLAP) in data mining? List the operations associated with OLAP. - **2065 Baisakh**
5. Describe the major task performed in each stage of Knowledge Discovery in Database (KDD) with example. - **2065 Baisakh**
6. Explain the characteristics of Data Warehouse. How operational database is different from Data Warehouse - **2065 Baisakh**

Questions from the topic - 3. Classification (12 hours) - Data Mining

1. "Do as your Neighbor Do". Explain this concept on behalf of Likelihood and Distance principle. Differentiate between hierarchical clustering and partitioning clustering technique **- 2066 Bhadra**
2. Write short notes on (any Three)
    1. Rules generation ID3 technique.
    2. Genetic Algorithm
    3. K-means clustering
    4. WWW mining and its advantages
    5. Rues of warehouse and query manager in DW

    - 2066 Bhadra

3. Write short notes on(any four)
    - Decision Support System
    - Likelihood and Distance Technique
    - Data Warehouse Schema
    - Partitioning and Hierarchical Clustering
    - Noise in Data Warehouse
    - Descision Tree

    - 2065 Baisakh

Questions from the topic - 4. Association Analysis (10 hours) - Data Mining

1. What is association analysis? Write A-priori Algorithm with its pruning principle. - 2066 Bhadra
2. Write the Apriori Algorithm. Use this algorithm to find the frequent item set and generate few rules associated with the transaction.

(Take minimum support=25%, minimum confidence%)

Transaction ID	Items
1	A, C, D
2	C, D, E
3	A, D, E
4	B, D, E
5	A, C, D, E
6	A, D, E
7	A, B, C
8	C, E

3. - 2065 Baisakh

Questions from the topic - 5. Cluster Analysis (9 hours) - Data Mining

1. Write short notes on (any Three)
    1. Rules generation ID3 technique.
    2. Genetic Algorithm
    3. K-means clustering
    4. WWW mining and its advantages
    5. Rues of warehouse and query manager in DW

    - 2066 Bhadra

2. Clusters the following instances of given data (2- Dimensional form) with the help of K-means algorithm (Take K = 2)

Instances	X	Y
1	1.0	2.5
2	1.0	4.5
3	2.5	3.0
4	2.0	1.5
5	4.5	1.5
6	4.0	5.0

    - 2065 Baisakh

3. Write short notes on(any four)
    - Decision Support System
    - Likelihood and Distance Technique
    - Data Warehouse Schema
    - Partitioning and Hierarchical Clustering
    - Noise in Data Warehouse
    - Descision Tree

    - 2065 Baisakh

Questions from the topic - 7. Advanced Applications (3 hours) - Data Mining

1. What is CRM? Describe the security and privacy aspects of Data Mining - 2066 Bhadra

Examination Control Division
2066 Bhadra

Programme	BCT	Pass Marks	32
Year / Part	IV / I	Time	3 hrs.

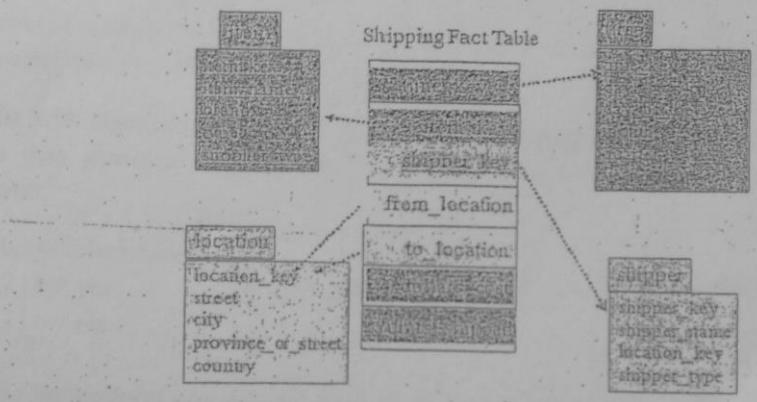
Subject: - Data Warehousing and Data Mining *(Elective I)*

- ✓ Candidates are required to give their answers in their own words as far as practicable.
- ✓ *Attempt All questions.*
- ✓ *The figures in the margin indicate Full Marks.*
- ✓ *Assume suitable data if necessary.*

[1] What is warehouse and data mart? Why do any organizations need warehouse separately though it has transactional database system? Explain with its application [2+2+6]

[2] How do you relate KDD with data mining? Explain the phases of KDD with example. [2+8]

[3] Explain the typical OLAP Operations over warehouse multidimensional data. Define the following star schema using DMQL. [4+6]



[4] Explain the general architecture of Warehouse System with its three main phases of development. [9]

[5] What kind of data preprocessing do we need before applying data mining algorithm to any dataset. Explain binning method to handle noisy data. [4+6]

[6] "Do as your Neighbor Do". Explain this concept on behalf of Likelihood and Distance principle. Differentiate between hierarchical clustering and partitioning clustering technique. [4+2]

[7] What is association analysis? Write A-Priori Algorithm with its pruning principle. [2+6]

[8] What is CRM? Describe the security and privacy aspects of data mining. [2+6]

[9] Write short notes on (any Three) [3+3+3]
- [a] Rules generation using ID3 technique.
- [b] Genetic Algorithm
- [c] K-Means Clustering
- [d] WWW mining and its advantages
- [e] Roles of warehouse and query manager in DW.

Attempt All the Questions

[1] What is warehouse and data mart? Why do any organizations need warehouse separately though it has transactional database system? Explain with its application [2+2+6]

[2] How do you relate KDD with data mining? Explain the phases of KDD with example. [2+8]

[3] Differentiate between OLAP & OLTP. Explain the warehouse schema, Fact table and Dimension table with example [3+6]

[4] Explain the Warehouse System Architecture [8]

[5] What are the major tasks to be followed in data preprocessing? Explain the techniques to handle noisy data [4+4]

[6] Define Likelyhood and Distance. Do we need likelyhood and distance in clustering? Explain the drawbacks of K-means clustering algorithm. [4+2+3]

[7] Write ID3 algorithm. Explain the significance of confusion matrix generated by WEKA [6+3]

[8] What is Web Log Mining? Explain WWW Mining Architecture [2+6]

[9] Write short notes on (any Three) [3+3+3]
- [a] Security and Privacy Aspects of Data Mining
- [b] Genetic Algorithm
- [c] DW Process Architecture
- [d] Phases of KDD
- [e] Apriori Algorithm and its drawbacks

Attempt All the Questions

[1] How do you define data warehousing? Why do any organization need data warehouse? Explain the process of creating a data warehouse. [3+3+4]

[2] Differentiate between OLAP and OLTP. Explain the conceptual modeling of Data Warehouse Schemas. How do you define Star Schema using DMQL? [3+4+3]

[3] Is KDD a data mining? Explain first two phases of KDD with example. [2+6]

[4] Briefly describe the Typical Process flow within the data warehouse system. [8]

[5] List out the techniques of Data Mining. "Data that are close to each other are very alike", verify this quotation with suitable supporting example. [2+6]

[6] explain with algorithm that how an ID3 works. Following is the confusion matrix of telephone user's data generated by ID3 algorithm using WEKA; calculate the TP rate, FP rate and precision. [6+6]

a b <-- classified as
7 2 | a = yes
3 2 | b = no

(Formula provided in Car question's back)

[7] Why do we perform cluster analysis in Data Mining? With K-means clustering algorithm, find out the clusters based on the following instances. (Choose K at least 2) [2+8]

Instances	Height	Width
1	0.5	2.2
2	1.5	3.5
3	0.7	1.5
4	2.0	1.0
5	3.2	2.3
6	1.0	1.8

[8] Explain the importance of World Wide Web Mining. Categorize and briefly describe the www mining techniques. [2+6]

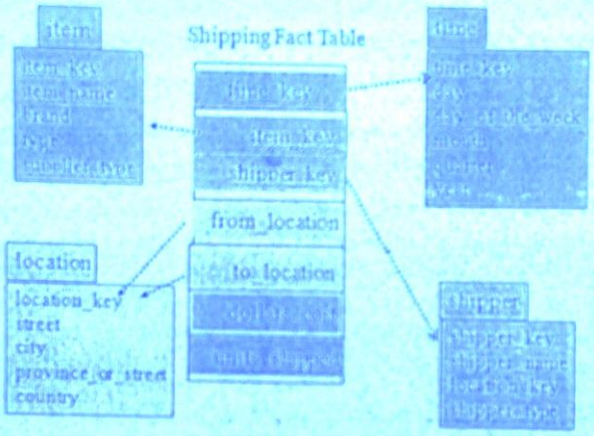
[9] Write short notes on (any two) [3+3]
- [a] A-priory Algorithm and its drawbacks
- [b] Data Preprocessing
- [c] Data Mining in CRM
- [d] Genetic Algorithm

Attempt All the Questions

[1] What is warehouse and data mart? Why do any organizations need warehouse separately though it has transactional database system? Explain with its application [2+2+6]

[2] How do you relate KDD with data mining? Explain the phases of KDD with example. [2+8]

[3] Explain the typical OLAP Operations over warehouse multidimensional data. Define the following star schema using DMQL. [4+6]



[4] Explain the general architecture of Warehouse System with its three main phases of development. [9]

[5] What kind of data preprocessing do we need before applying data mining algorithm to any dataset. Explain binning method to handle noisy data. [4+6]

[6] "Do as your Neighbor Do". Explain this concept on behalf of Likelihood and Distance principle. Differentiate between hierarchical clustering and partitioning clustering technique. [4+2]

[7] What is association analysis? Write A-Priori Algorithm with its pruning principle. [2+6]

[8] What is CRM? Describe the security and privacy aspects of data mining. [2+6]

[9] Write short notes on (any Three) [3+3+3]
- [a] Rules generation using ID3 technique
- [b] Genetic Algorithm
- [c] K-Means Clustering
- [d] WWW mining and its advantages
- [e] Roles of warehouse and query manager in DW

Data warehousing and data mining, question 2063

Attempt all questions:

Q. 1 what is Data Mining? Is it deferent from data warehouse? Explain the application of data mining in biomedical, CRM & Telecom Industries. [2+2+6]

Q. 2 is KDD a Data Mining? explain the phases of KDD with example. [2+8]

Q. 3 "**The Kathmandu Mall**" one of the largest departmental stores in the valley. Assume your project team got the opportunity to build the warehouse system of this departmental store. Develop the process architecture of the warehouse system and explain how you manage the responsibilities to build the warehouse. [10]

Q.4 why do we need cluster analysis? Differentiate between hierarchical and partitioning clustering technique. Identify the cluster of the following instances using K-Means clustering algorithm. [2+10]

Instance	X	Y
1	1.0	1.5
2	2.5	5.5
3	1.5	1.0
4	2.0	3.0
5	2.5	3.5
6	4.0	6.2

Q. 5 what are the differences between OLAP & OLTP? Define OLAP operations, fact table and dimension table. [3+6]

Q. 6 write Apriori algorithm. Identify the candidate and large item sets of the following transaction table. Use Apriori algorithm with minimum support 2. [2+10]

TID	Items
10	A, C, D
20	B, C, E
30	A, B, C, E
40	B, E

Q. 7 what is web mining? Explain with its importance. [2+6]

Q. 8 Write short note on (Any Three)[3+3+3]

[a] Neural Networks
[b] DSS
[c] Data Preprocessing
[d] Data Marts & Metadata in DW
[e] Architectural Aspects of Data Mining

Attempt All the Questions

[1] Explain the necessity of Data warehousing and data mining in government, medical and military. Characterize the warehouse data with example. [8]

[2] Define the terms Data Mart, Metadata, slicing, dicing, base cuboid and apex cuboid. [8]

[3] Assume that your college has perfect software package named College Management Information System (CMIS) which includes Teacher Database, Student Database, Staffs Database and Account Database and also manage tutorials, class notes through web. There is a facility of online assessment examination system that might be controlled by each subject teacher. Except this, now your College management board has passed to develop a project named **college warehouse system** and announced the tender notice through college web. Suppose that you are the project leader of this warehouse development system. As a project leader,

(a) What are your responsibilities? (b) Divide the project into sub-system and describe the responsibilities assigned to your other partners with the help of your warehouse system architecture. [9]

[4] What is entropy and information gain? Explain with algorithm that how can the ID3 tree be generated? [3+7]

[5] How could you solve the problem of over fitting in ID3 Implementation? Explain the rules and conditions to generate decision tree using ID3. [4+4]

[6] What is DMQL? Where is it applicable? Differentiate between Transactional databases with Data warehouse databases. [2+3+5]

[7] What is Association Analysis? Explain Apriori Algorithm with its pruning principle. [2+8]

[8] Explain Web Structure/Content/Usage mining. Explain the privacy Aspects of Data Mining. [5+3]

[9] Write short notes on (any Three) [3+3+3]

[a] Characteristics of Neural Network & Perceptron Learning rule

[b] OLAP & OLTP

[c] Likelyhood and Distance

[d] Genetic Algorithm

[e] KDD

Attempt All the Questions

[1] Explain the importance and uses of Data Warehouse with its characteristics. [2+2+6]

[2] What is KDD? How do you apply KDD in World Wide Web mining? [2+8]

[3] What are the uses of OLAP? Compare OLAP with OLTP. [4+6]

[4] Explain the role of Warehouse manager with respect to detail information, summary information, Meta data and data mart management. [10]

[5] Why do we need to preprocess data for mining? Explain the methods to handle noisy data. [4+6]

[6] What is K-nearest neighbor? Find the other two students with common behavior from anyone record of the following table (apply likelihood and distance principle) [2+10]

Name	Age	Weight
Std1	22	53
Std2	26	57
Std3	20	54
Std4	24	52

[7] How do you generate rules with ID3? Explain with example. [10]

[8] Write short notes on (any Two) [4+4]
- [a] Apriory algorithm and its drawbacks
- [b] Crossover and mutation in GA
- [c] Privacy Matters in CRM
- [d] Data mining visualizations.

Attempt All the Questions

[1] Compare Data Mining with RDBMS SQL in Knowledge Extraction. Explain the Cause of DM & DW Project Failure. [4+4]

[2] Define the terms Data Mart, Metadata, Fact table and Dimension table with example. [8]

[3] Explain ten golden rules to setup perfect KDD environment. [10]

[4] Explain the application areas of Cluster Analysis. Describe K-means clustering algorithm with example. [3+7]

[5] How could you solve the problem of over fitting in ID3 Implementation? Explain the rules and conditions to generate decision tree using ID3. [4+4]

[6] What is DMQL? Where is it applicable? Differentiate between Transactional databases with Data warehouse databases.

[7] What is Association Analysis? Explain Apriori Algorithm with its pruning principle. [2+6]

[8] Explain Web Structure/Content/Usage mining. Explain the privacy Aspects of Data Mining. [6+3]

[9] Write short notes on (any Three) [3+3+3]

[a] Characteristics of Neural Network & Perceptron Learning rule

[b] OLAP & OLTP

[c] Likelyhood and Distance

[d] Warehouse process architecture

[e] KDD

Attempt All the Questions

[1] Explain with reason the areas where data mining is applicable and not applicable. Is it compulsory to have data warehouse in an organization for data mining? Explain. [5+5]

[2] Write down the differences between RDBMS database with data warehouse database. Explain with example, the terms data mart, metadata, slicing, dicing and drilling. [3+5]

[3] Nepal Government is going to build a national data center (National Data Warehouse) for its running e-Governance project. Propose your own process architecture of the warehouse system and explain the responsibilities of each entity within the proposed warehouse architecture. [10]

[4] Explain the parameters and steps that ID3 algorithm follows to build a tree. How do you avoid the problem of over-fitting? Calculate TP rate, FP rate and Recall value from the following confusion matrix generated by WEKA from sales data. [3+2+6]

```
a b  <-- classified as
9 2 | a = yes
3 5 | b = no
```

[5] What are the different techniques used in data mining? Explain likelihood and distance principle with example. [3+5]

[6] List the steps of data preprocessing. Explain binning method to handle noisy data and normalization techniques for data transformation. [2+3+3]

[7] What is CRM? List the goals of CRM in relationship marketing and channels used for customer approach. [2+5]

[8] Write A-priori algorithm with its pruning principle. Find the candidate and large item sets from the following transaction in a restaurant using A-priori algorithm. (Min. support: 2) [4+8]

Transaction ID	Order Combination
1	Beer, chicken, peanut
2	Whisky, buff, coke, lemon
3	Beer, buff, coke, lemon
4	Buff, coke, peanut
5	Beer, peanut, lemon, fish
6	Peanut, coke

[9] Write short notes on (any two) [3+3]

[a] Characteristics of Neural Network & Perceptron Learning rule [b] WWW Mining

[c] OLAP & OLTP [d] DSS