Tribhuvan University
Institute of Science and Technology

SCHOOL OF MATHEMATICAL SCIENCES

First Assessment 2080

Subject: **Research Methodology** *Full marks*: **45**
Course No: **MDS 601** *Pass Marks:* **22.5**
Level: **MDS / II Year / III Semester** *Time:* **2hrs**

Candidates are required to give their answers in their own words as far as practicable.
Attempt All Questions.

Group A [5x3=15]

1. Explain the difference between fundamental research and applied research with suitable examples.
2. What is scientific research? Describe the major steps involved in the scientific research process?
3. Define literature survey. What are the benefits of a good literature survey?
4. What are the steps involved in the research design process?
5. What is ratio scale? What are its superiorities over other scale of measurement

Group B [5x6=30]

6. Define the term reliability. Describe the various types of reliability.
7. What is case study research? Describe the steps and characteristics of a case study research.
8. Describe the difference between statistical hypothesis and Research hypothesis.
9. Describe sampling. Why is sampling used in research? List the steps involved in the process of sampling.

OR

What is non-sampling error? How non-sampling error can be reduced in the process of data collection?

10. prepare a bibliographic entry in APA method and MLA method for the following :
    a) for a book with one author
    b) for a journal article

OR

Describe experimental research method. Point out the mandatory Steps in Experimental Methods.

***

Tribhuvan University
Institute of Science and Technology
SCHOOL OF MATHEMATICAL SCIENCES
First Assessment 2080

Subject: **Advanced Data Mining** | *Full marks:* **45**
Course No: **MDS 602** | *Pass Marks:* **22.5**
Level: **MDS / II Year / III Semester** | *Time:* **2hrs**

Candidates are required to give their answers in their own words as far as practicable.
Attempt All Questions.

Group A **[5x3=15]**

1. What is warehousing? What are the sources of data for building data warehouse.
2. List the differences between OLAP and OLTP.
3. List out the different techniques that can be applied for data mining with use case.
4. Why data pre-processing is important for data mining? List the possible task that can be done in data pre-processing phase.
5. Write different steps of K-NN algorithm.

Group B **[5x6=30]**

6. How can we build the Association based recommendation system. Explain with example.

OR

What is Apriori Algorithm. Discuss with its use case.

7. Explain data visualization techniques.

OR

Write ID3 Algorithm. With example discuss the pros and cons of the decision tree based classification technique.

8. Explain KDD with workflow.
9. What is Measurement of similarity? Why this is important in Data mining?
10. How ANN works? Explain the working of the perception for classification task.

***

Tribhuvan University
Institute of Science and Technology

SCHOOL OF MATHEMATICAL SCIENCES

First Assessment 2080

Subject: **Techniques for Big Data**
Course No: **MDS 603**
Level: **MDS / II Year / III Semester**

Full marks: **45**
Pass Marks: **22.5**
Time: **2hrs**

Candidates are required to give their answers in their own words as far as practicable.
Attempt All Questions.

Group A [5x3=15]

1. Explain about the different characteristics of big data.
2. How the functional programming language of MapReduce different from the object-oriented programming language?
3. What are the core components of Hadoop?
4. What do you understand by the term "running Hadoop"?
5. Explain about different map reduce input and output formats.

Group B [5x6=30]

6. What is map reduce? Explain its distributed execution overview.
7. Explain about the anatomy of read and write operation ofthe Hadoop Distributed File System (HDFS).

OR

Explain about the HDFS architecture with its running daemons.

8. Explain about the anatomy of map reduce job run, failures, shuffle and sort with reference to word frequency count example.

OR

What are the different configuration modes to setup the Hadoop? Which setup mode is preferred for the development and why? Also briefly explain about Hadoop streaming.

9. Write mongodb query for the following database:
   - db.order.find({})

```
[{
cust_id: "ID1",  "ord_date": ISODate("2018-05-04"),  "price": 400,  "status": "A",  "itemqty": 20
},
{
cust_id: "ID1",  "ord_date": ISODate("2020-05-04"),  "price": 400,  "status": "not A",  "itemqty": 20
},
{
cust_id: "ID2",  "ord_date": ISODate("2015-05-04"),  "price": 200,  "status": "not A",  "itemqty": 10
}]
```

   a) For each unique cust_id, sum the price field on your created order collection
   b) For each unique cust_id, ord_date grouping, sum the price field. Exclude the time portion of the date.
   c) For cust_id with multiple records, return the cust_id and the corresponding record count.
   d) For each unique cust_id, ord_date grouping, sum the price field and return only where the sum is greater than 250. Exclude the time portion of the date.
   e) For each unique cust_id with status A, sum the price field
   f) For each unique cust_id with status A, sum the price field and return only where the sum is greater than 250.

10. Give some examples of HDFS commands. What is the single point of failure in Hadoop version 1 and how Hadoop version 2 tries to solve it? Explain.

***

Tribhuvan University
Institute of Science and Technology

SCHOOL OF MATHEMATICAL SCIENCES

First Assessment 2080

Subject: **Decision Analysis**
Course No: **MDS 606**
Level: **MDS / II Year / III Semester**

Full marks: **45**
Pass Marks: **22.5**
Time: **2hrs**

Candidates are required to give their answers in their own words as far as practicable.
Attempt All Questions.

Group A **[5x3=15]**

1. Define Pure and Mixed strategies in Game Theory.
2. What is conditional probability? State Bayes theorem up to 3 events.
3. Mr. X quite often flies from town A to town B. He can use the airport bus which costs Rs.25 but if he takes it, there is a 0.08 chance that he will miss the flight. A Hotel Everest costs Rs.270 with a 0.96 chance of being on time for the flight. For Rs.350 he can use a taxi which will make 99 of 100 flights. If Mr. X catches the plane on time, he will conclude a business transaction which will produce a profit of Rs.10000, otherwise he will lose it. Which mode of transport should Mr. X use? Answer on the basis of the EMV criterion.
4. Mr. Shrestha is on a point to decide whether to stock paddy or rice. If he stocks paddy and if it is a success, he hopes that he can make profit Rs.10000 and if it is a flop, he will lose Rs.2000. If he stocks rice and if it is a success, he thinks that he can make Rs.20000 but if it is a flop, he would lose Rs.5000. From the following probability distribution, construct a decision tree indicating all events.

Probability Distribution

Act / Event	Paddy	Rice
Success	0.6	0.7
Flop	0.4	0.3
Total	1.00	1.00

   a) Find the total expected monetary value of the two decisions.
   b) Advice Mr. Shrestha as to which item he should stock and what would be his maximum possible outcome?
5. In a game of matching coins with two players A and B, suppose A wins 2 units of value when there are two heads, wins nothing when there are two tails and losses 1 unit of value when there is one head and one tail. Determine the payoff matrix, the best strategies for each player and the value of the game to A.

Group B **[5×6= 30]**

6. What are the five criteria of decision making under condition of uncertainty? Explain them with suitable examples.
7. The following matrix gives the payoff (in Rs. Lakh) of different strategies S_1, S_2 and S_3 against different conditions N_1, N_2 N_3 and N_4.

Strategy (Supply)	States of nature (demand)			
	High demand (N_1)	Moderate demand (N_2)	Low demand (N_3)	Failure (N_4)
Expand (A_1)	6500	3000	–2000	–5000
Build (A_2)	9000	5000	–3000	–10000
Subcontract (A_3)	3000	1500	–1000	–2000

   a) Indicate the decision taken under the following approaches: Optimistic crite
   b) Pessimistic criterion c) Minimax regret criterion d) Laplace criterion
   e) Hurwitz criterion, if the coefficient of optimism is 0.80.
8. An ice-cream retailer buys ice-cream at a cost of Rs.50 per cup and sells it for Rs.80 per cup. Any remaining unsold at the end of the day can be disposed of at a salvage price of Rs.20 per cup. Past sales have been ranged between 15 and 18 cups per day. There are no reasons to believe that sales volume will take on any other magnitude in future. If the sale history has the following probabilities, find how many cups the retailer should bought per day to maximize the profit.

Market size	15	16	17	18
Probability	0.10	0.20	0.40	0.30

a) What quantity should be bought to maximize expected profit?

b) What will be his maximum expected profit? If he has perfect information of the market, what would be the expected profit?

c) What is the cost of uncertainty?

OR

The following table shows three states of nature (events) and three actions. The amount that a person will gain in each combination of states of nature and action is shown in the table.

Events	Actions		
	A_1 (Rs.)	A_2(Rs.)	A_3(Rs.)
D_1	25000	-7000	-18000
D_2	30000	50000	-8000
D_3	40000	25000	60000

Probabilities for three states of nature (events) are also given below.

States of nature	D_1	D_2	D_3
Probability	0.3	0.5	0.2

a) Find out EMV for each decision alternative.

b) Find out EPPI and EVPI.

c) A researcher first has agreed to conduct a survey to provide the management with additional information regarding the state of nature. It will charge Rs.12000 for undertaking this survey. Do you think that the survey should be conducted?

9. The ABC Publication Pvt. Ltd. has to take a decision regarding publication of any of the three projects offered by an author: Statistical Methods (SM), Quantitative Techniques (QT) and Business Mathematics (BM).

The probability distribution of these three projects have been estimated as under:

	State of sales			
Project	Very good	Good	Fair	Poor
SM	0.40	0.30	0.20	0.10
QT	0.50	0.25	0.15	0.10
BM	0.25	0.40	0.20	0.15

The profits expected from the projects under the given states of sale are as follows:

	State of sales			
Project	Very good	Good	Fair	Poor
SM	40000	30000	10000	-5000
QT	25000	15000	5000	-3000
BM	50000	20000	10000	-5000

Construct a decision tree for the above data and advice the ABC as to which project it should take up for the publication.

10. Reduce the following two-person zero sum game to 2×2 order by dominance rule and obtain the optimal strategies for each player and the value of the game.

Player A	Player B			
	B_1	B_2	B_3	B_4
A_1	3	2	4	0
A_2	3	4	2	4
A_3	4	2	4	0
A_4	0	4	0	8

OR

Find the optimal strategies using the dominance rule for player A and B in the following game. Also obtain the value of the game.

Player A's Strategy	Player B's Strategy		
	B_1	B_2	B_3
A_1	–3	–2	2
A_2	–1	4	2

***

Tribhuvan University
Institute of Science and Technology

SCHOOL OF MATHEMATICAL SCIENCES

First Assessment 2080

Subject: **Monte Carlo Methods** | *Full marks:* **45**
Course No: **MDS 607** | *Pass Marks:* **22.5**
Level: **MDS / II Year / III Semester** | *Time:* **2hrs**

Candidates are required to give their answers in their own words as far as practicable.
Attempt All Questions.

Group A **[5x3=15]**

1. What is an error? How do you estimate it in Monte Carlo methods?

2. Explain the meaning of "Pseudo Random Numbers (PRN)".

3. Use idea of Neuman to generate five random numbers.

4. Write down the meaning of transition probabilities with examples.

5. Describe the meaning of "State Space".

Group B **[5x6=30]**

6. Discuss Monte Carlo method to estimate value of π.

OR

Illustrate the importance sampling to evaluate integral of a function.

7. What is Monte Carlo Method? Discuss its significance.

OR

How do you evaluate

$$I = \int_{-5}^{5} dx\, dy\, g(x, y)$$

with

$$g(x, y) = \exp(-x^2 - y^2 - (x - y)^2/2)$$

using Monte Carlo. Illustrate with algorithm.

8. Discuss Acceptance Rejection algorithm in Monte Carlo with example.

9. How do you distinguish "stationary distribution" from other distributions. Explain an algorithm to get "stationary distribution" using Python code.

10. Discuss the role of Master equation in Monte Carlo methods.

***