Shahjalal University of Science and Technology (SUST), Sylhet – 3114.

Department of Software Engineering

3rd Year 2nd Semester 2019 Final Examination (Session: **2016 – 2017**)

Course Code: **BUS – 301W**, Course Title: **Entrepreneurship Development**, Credit: **02**

1st Term – Test

Full Marks – 10 Time: **1 hour**

[Answer All from the followings Questions]

[All parts of a question must be answered sequentially, like 1 (a), 1 (b), 1 (c) .....]

1. a. Define Entrepreneur and Entrepreneurship on the basis of Evolution Concept of Entrepreneur. **02**

   c. How will you Distinguish among Entrepreneur, Manager and Intrapreneur? **03**

   c. Why Rural Entrepreneurship is needed in country like Bangladesh? **03**

   d. How Rural Entrepreneurship would be developed by overcoming the Problems? **02**

Scanned by CamScanner

CamScanner

Shahjalal University of Science and Technology

Institute of Information and Communication Technology

B.Sc. (Engg.) in Software Engineering

3rd Year 2nd Semester Final Examination, 2019

Session: 2016-17

Course No: BUS-301W Course Title: Entrepreneurship Development

Credits: 2 Full Marks: 50 Time: 2 Hours

Answer any two questions from the following of each part.

All part of question must be answered sequentially

PART-A

(Answer any two questions)

1.

a. Define Entrepreneur on the basis of Evolution Concept of Entrepreneur. 2

b. What would be the best possible way to characterize an Entrepreneur? 2.5

c. How will you distinguish among Entrepreneur, Manager and Intrapreneur? 3

d. Being an Entrepreneur, what type of business will you prefer for your own characteristics and why? 2.5

e. What your statement about Entrepreneur and Entrepreneurship? 2.5

2.

a. What would be the best possible way to perform for Women to develop an Organization as an Entrepreneur? 3

b. How will a Women Entrepreneur overcome the obstacles for setting up an Organization? 2.5

c. Why Rural Entrepreneurship is needed in country like Bangladesh? 2.5

d. How Rural Entrepreneurship would be developed by overcoming the problems? 2.5

e. Mention the basic difference between Maslow's Need Hierarchy Theory and McClelland's Acquired Needs Theory? 2

3.

a. What is the rural entrepreneurship? What are the factors will you consider to be a successful entrepreneur? Explain. 5.5

b. Describe the Kakinada Experiment. 4

c. As a student of Software Engineering, How can you explain the role of technopreneurs in economic development of developing countries like Bangladesh. 3

PART-B

(Answer any two questions)

4.

a. Define projects and describe its classification. 3.5

b. How and why will you select a project after identifying among various alternatives? 5

c. Why a project report is significant and mentions the basic contents of project report. 4

5.

a. Define quality and mention the Historical Development of Quality Concept and Total Quality Management (TQM). 2.5

b. What are the differences between ISO 9000 and TQM? 2.5

c. Why TQM Process is needed in Small-Scale Enterprise? 2.5

d. What is the concept of industrial sickness? 1

e. Explain the Signals and ~~Sickness~~ symptom of Industrial sickness. 4

6.

a. Do you believe that ethics and social responsibilities should be the part of entrepreneurs in Bangladesh? Justify your answer. 5

b. Why does and entrepreneur need to be aware of the dynamics associated with time? 3.5

c. Discuss how an employee can be held liable for trade secrets and knowledge attained while working is a sensitive position 4

Scanned by CamScanner

Shahjalal University of Science and Technology
Institute of Information and Communication Technology
Term Test 1
Course Code: SWE 331 Credits: 2 Course Title: Software Usability and Metrics
Time: 1 hr Total Marks: 20

Marks: 10

1. Read the following story carefully:

A user shall be able to sign up for this portal. He shall be able to edit his personal information. The system shall have two different types of users- i. registered users and ii. unregistered users. Registered users can be of two types- i. students and ii. teachers. There will be an admin. He shall be able to assign courses to teachers. The students shall be able to request the teachers to get enrolled into the course. The teacher shall be able to add students into his course. He shall be able to create groups for particular students. He shall be able to assign a single project to a particular group. The students shall be able to see which group do they belong to and they also shall be able to see other members of the group. The students shall also be able to upload their projects to the group. The teachers shall be able to see the projects; they shall be able to mark the submissions as complete or incomplete. The unregistered users shall be able to see the complete submissions and shall be able to download the projects. Also they shall be able to search through the entire site.

Within your organization average number of staff months per mandatory requirement is 1.05 with a standard deviation of 10%. What do you estimate the number of staff months to be to build the above mentioned system? As this report is for your boss you are advised to decorate it with some simple statistical analysis. Don't give your boss a single choice. Show him a few options so that he may reason.

2. The above mentioned system has simple input from four sectors: admin, student, teacher, unregistered user. It produces simple output for teacher, student and average output for unregistered user. It contains four average interfaces. The system also contains 3 database files of average complexity. It also has 3 simple and 1 complex inquiry. Assume all 14 GSC values are 2. Marks: 4 + 3 + 3

a. Now estimate the effort in staff months to build the system. Assume a productivity of 10 function points per staff month.

b. How will you tackle the scenario where your boss can only provide you with some fraction of your estimated staff months?

c. Predict the probable size of the project in terms of David Consulting Group Data. Assume the language that has been used to develop the system is JAVA.

Scanned by CamScanner

TABLE 4.3 Function Point Complexity Ratings

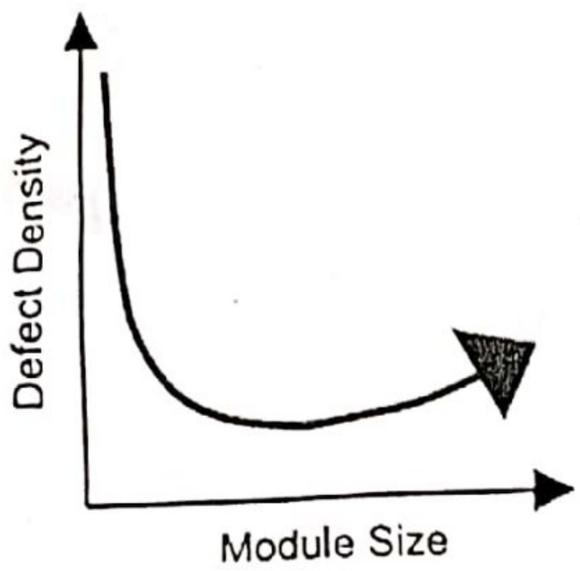
Component	Simple	Average	Complex
Inputs (I)	3	4	6
Outputs (O)	4	5	7
Data Files (F)	7	10	15
Interfaces (N)	5	7	10
Inquiries (Q)	3	4	6

	QSM SLOC/FP Data				David Consulting Group Data
Language	Average	Median	Low	High	
Access	35	38	15	47	—
Ada	154	—	104	205	—
Advantage	38	38	38	38	—
APS	86	83	20	184	—
ASP	69	62	32	127	—
Assembler	172	157	86	320	575 Basic/ 400 Macro
C	148	104	9	704	225
C++	60	53	29	178	80
C#	59	59	51	66	—
Clipper	38	39	27	70	60
Cobol	73	77	8	400	175
Cool:Gen/IEF	38	31	10	180	—
Culprit	51	—	—	—	—
DBase III	—	—	—	—	60
DBase IV	52	—	—	—	55
Easytrieve +	33	34	25	41	—
Excel	47	46	31	63	—
Focus	43	42	32	56	60
FORTRAN	—	—	—	—	210
FoxPro	32	35	25	35	—
HTML	43	42	35	53	
Ideal	66	52	34	203	—
IEF/Cool:Gen	38	31	10	180	—
Informix	42	31	24	57	—
J2EE	61	50	10	60	—
Java	60	59	14	97	80
JavaScript	56	54	44	65	50
JCL	60	48	21	115	400
JSP	59	—	—	—	—
Lotus Notes	21	22	15	25	—
Mantis	71	27	22	250	—
Mapper	118	81	16	245	—

Scanned by CamScanner

Shahjalal University of Science and Technology
Institute of Information and Communication technology
Software Engineering
Software Usability and Metrics (SWE 331)
Term Test 2
Full marks: 20 Time: 45 minutes

1. Describe what do you understand by the following graph. What are the reasons for this shape?
   Marks: 5



2. Find the LOC, CC, Weighted Information flow and Halsted metrics for the following pseudocode.
   Marks: 15

```
Eratosthenes (n) {
        e[1] := 0
        for i := 2 to n do e[i] := 1
        p := 2
        while p*2 < n do {
                j := p*2
                while (j<n) do {
                        e[j] := 0
                        j := j + p
                }
                repeat p := p + 1 until e[p] = 1
        }
        return(a)
}
```

n_1
n_2

= 9
*1
<1
+2
: 8
} 5 = n_1

n = 4
e = 4
i = 2
p = 7
j = 4
a = 1

Scanned by CamScanner

Shahjalal University of Science and Technology
Institute of Information and Communication Technology
Software Engineering
Final Examination, 3rd Year 2nd Semester, 2019

Course Code: SWE 331 **Credits: 2** **Course Title: Software Usability and Metrics**
Time: 2 hrs **Total Marks: 50**

Group A
[Answer all the questions]

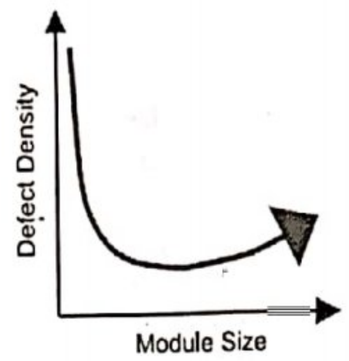
1. Answer any FIVE **5x1=5**

a) Most of the methods of Software Estimation is empirically defined – True or false. If false give correct answer.
b) What does Maintainability Index mean?
c) Define Index of Variation.
d) What is the most common measurement of software?
e) What is Coupling between objects?
f) What is KLOC?

2. Answer any FOUR **4x2.5=10**

a) What do you understand by the term "Manage by numbers"?
b) Assume you have three modules—A, B, and C—which have sizes of 10 KLOC, 24 KLOC, and 50 KLOC, respectively. Now find the Deviation and Variance from the data.

c) Describe what do you understand by the following graph. What are the reasons for this shape?



d) Briefly describe how Agresti–Card–Glass System Complexity Metric works.
e) What are the problems that may arise if you measure a system by counting the lines of codes that it has?

3. Answer any TWO **2x5=10**

a) Suppose ABC Software Ltd is a software firm. They have the following properties:
They don't allow ad-hoc software development.
They make plans and prepare estimation report based on real world scenario.
They can track progress of their product that is under development.
The employees of the firm have separate and properly defined responsibilities.
Now, answer the following questions.
   i. According to CMMI at which level does ABC Software Ltd belong to?
   ii. What are the things that they need to incorporate within their firm to make it to the highest standard?

b) Read the following story carefully:
"A user shall be able to sign up for this portal. He shall be able to edit his personal information. The system shall have two different types of users- i. registered users and ii. unregistered users. Registered users can be of two types- i. students and ii. teachers. There will be an admin. He shall be able to assign courses to teachers. The students shall be able to request the teachers to get enrolled into the course. The teacher shall be able to add students into his course. He shall be able to create groups for particular students. He shall be able to assign a single project to a particular group. The students shall be able to see which group do they belong to and they also shall be able to see other members of the group. The students shall also be able to upload their projects to the group. The teachers shall be able to see the projects; they shall be able to mark the submissions as complete or incomplete. The unregistered users shall be able to see the complete submissions and shall be able to download the projects. Also they shall be able to search through the entire site."

15

Scanned by CamScanner

Within your organization average number of staff months per mandatory requirement is 1.15 with a standard deviation of 10%. What do you estimate the number of staff months to be to build the above mentioned system? As this report is for your boss you are advised to decorate it with some simple statistical analysis. Don't give your boss a single choice. Show him a few options so that he may reason.

c) Find the LOC, CC and Weighted Information flow for the following code.

```
//this is a code to sort each row of a matrix
void matrixSort(int a[2][2]) {
  //bubble sorting each row
  for (int i = 0; i < 2; ++i){
    for (int j = 0; j < 1; ++j) {
      for (int k = j + 1; k < 2; ++k) {
        if (a[i][j] > a[i][k]){
          int temp = a[i][j];
          a[i][j] = a[i][k];
          a[i][k] = temp;
        }
      }
    }
  }
  //prints the sorted array
  for (int i = 0; i < 2; ++i) {
    for (int j = 0; j < 2; ++j) {
      printf("%d ", a[i][j]);
    }
    printf("\n");
  }
  return;
}
```

14 m
e
19
7

References:

CC = Binary Decisions + 1

Weighted IFC = length * (fanin * fanout)2

Scanned by CamScanner

Group B

[Answer all the questions]

4. Answer any FIVE **5x1=5**

a) What SEI stands for?
b) 3 SD means what percentage of the population?
c) What do you understand by Lack of Cohesion on Methods?
d) Give example of nominal scale.
e) What are Software Development Metaphors?
f) Having a deep inheritance tree is not a good software development practice. True or false?

5. Answer any FOUR **4x2.5=10**

a) What is GQM^2?
b) Why is Software Measurement a challenging task?

c) What does effort mean in Software Estimation?
d) What does it mean that a firm is "Quantitatively Managed"?

e) Extensive productivity research was done between 1927 and 1932 at the AT&T Hawthorne plant in Cicero, Illinois. They studied manufacturing and believed that if the environment was changed in certain ways (such as more light), it would improve productivity. And they were right. Productivity improved. Then they changed another factor and measured the result. Productivity improved. Then they changed it back to the original state. Productivity improved.
What do you think is the actual reason for this improvement?

6. Answer any TWO **2x5=10**

a) Suppose you have a system that has simple input from four sectors- admin, student, teacher, unregistered user. It produces simple output for teacher, student and average output for unregistered user. It contains four average interfaces. The system also contains 3 database files of average complexity. It also has 4 simple and 1 complex inquiry. Assume all 14 GSC values are 2.

TABLE 4.3 Function Point Complexity Ratings

Component	Simple	Average	Complex
Inputs (I)	3	4	6
Outputs (O)	4	5	7
Data Files (F)	7	10	15
Interfaces (N)	5	7	10
Inquiries (Q)	3	4	6

i. Now estimate the effort in staff months to build the system. Assume a productivity of 10 function points per staff month.
ii. Predict the probable size of the project in terms of David Consulting Group Data. Assume the language that has been used to develop the system is JAVA. According to David Consulting Group Data for JAVA it is 80 SLOC/FP.

b) Find Halstead's Metrics for the following code:

```
void DFS(int now)
{
    visited[now] = true;
    int i, to, k;
    k = graph[now].size();
    for(i=0;i<k;i++)
    {
        to = graph[now][i];
        if(visited[to] == false)
        {
            DFS(to);
        }
    }
    stk.push(now);
    return;
}
```

Scanned by CamScanner

Reference:

Halstead's metrics are defined as:

Length: $N = N_1 + N_2$
Vocabulary: $n = n_1 + n_2$
Volume: $V = N(\log_2 (n))$
Difficulty: $D = (n_1/2) * (N_2/n_2)$
Effort: $E = D * V$

where

n_1 = number of distinct operators
n_2 = number of distinct operands
N_1 = total number of operators
N_2 = total number of operands

c) Suppose you want to develop a result processing system for your university. Now identify the goal, at least 2 questions that helps to answer whether goal is reached or not and for each question develop at least 2 metrics that helps you to answer the question.

Scanned by CamScanner

Software Engineering

Course Code: SWE-333 **Term Test # 2**

Total Marks: 20 **Date:** 17th November, 2019

12

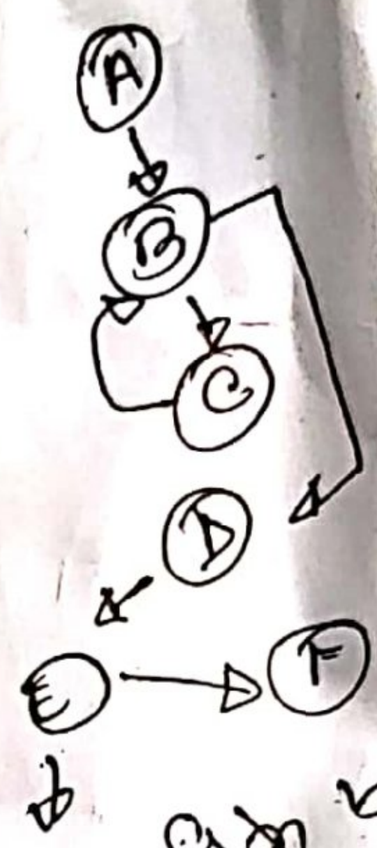
8

1. A program computes a b where a lies in the range [1,10] and b within [1,5]. Design test cases for this program using BVC, robust testing,

2.

```
main() {
    int num, small;
    int i, j, sizelist, list[10], pos, temp;
    clrscr();
    printf("\nEnter the size of list : \n ");
    scanf("% d",& sizelist);

    for (i = 0; i < sizelist; i++) {
        printf("\nEnter the number");
        scanf("% d",& list[i]);
    }

    for (i = 0; i < sizelist; i++) {
        small = list[i]; pos = i;
        for (j = i + 1; j < sizelist; j++) {

            if (small > list[j]) {
                small = list[j]; pos = j;
            }
        }
        temp = list[i];
        list[i] = list[pos];
        list[pos] = temp;
    }
    printf("\nList of the numbers in ascending order : ");
    for (i = 0; i < sizelist; i++)
        printf("\n % d", list[i]);

    getch();
}
```

A B C D E



6m + 1

(a) Draw the DD graph for the program.

(b) Calculate the cyclomatic complexity of the program using all the methods.

(c) List all independent paths.

Scanned by CamScanner

Shahjalal University of Science and Technology
Institute of Information and Communication Technology
3rd Year 2nd Semester Final Examination' 2019
Session: 2016-17

Course Code: SWE 333 **Credits: 2** **Course Title: Software Verification and Validation**
Time: 2 hrs. **Total Marks: 50**

Group A
[Answer all the questions]

1. Answer any FIVE **5x1=5**

a) What is Cyclomatic Complexity?
b) Write down the Software Development Life Cycle steps. proto
c) Why Must the Process of underlying software be written?
d) What are the basic activities of Spiral model?
e) What is software quality assurance?
f) Define Source node and Sink node for path based integration.
g) How can you verify a code?

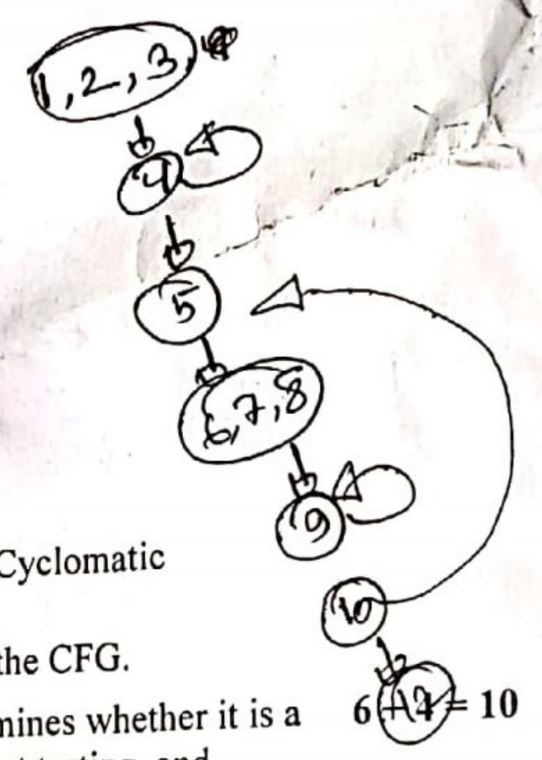
2. Answer any FOUR **4x2.5=10**

a) Write down the difference between White box testing and Black box testing. 2.5
b) Describe the waterfall model. 2.5
c) What is static testing? What are the different steps of inspections? 2.5
d) Describe the logic coverage criteria in white box testing. 2.5
e) What is the Equivalence class testing? Explain the identification of equivalent class. 2.5
f) Describe the states of the bugs. 2.5

3. Answer any One **10x1=10**

a) Consider the following code segment: 10

```
1 Void insertion_procedure (int a[], int p [], int N)
2 {
3   int i,j,k;
4   for (i=0; i<=N; i++) p[i] = i;
5   for (i=2; i<=N; i++)
6   {
7     k = p[i];
8     j = 1;
9     while (a[p[j-1]] > a[k]) {p[j] = p[j-1]; j--}
10    p[j] = k;
11  }
12 }
```



(a) Draw a Control Flow Graph (CFG) for the above code. Calculate the Cyclomatic Complexity.
(b) Prepare a test case for each of the independent paths require to cover the CFG. 6+4 = 10

b) i) A program reads an integer number within the range [1,100] and determines whether it is a prime number or not. Design test cases for this program using BVC, robust testing, and worst-case testing methods.
ii) Write down the formation of Decision table formation criteria.

Group B
[Answer all the questions]

4. Answer any FIVE **5x1=5**

a) How Driver class helps in unit testing?
b) Write down the types of Integration testing.
c) What is Defect in software?
d) Define the attributes of a Test case.
e) Explain the properties of Control Flow Graph.
f) Draw the Diagram of Inspection Process.
g) What is the Ramp up period?

5. Answer any FOUR **4x2.5=10**

a) Explain Rapid Prototyping model of Software development. 2.5
b) What is software crisis? State with an example. 2.5
c) Why incremental integration testing is preferred rather than using the non-incremental approach? 2.5
d) What is the benefit of TDD over the traditional unit testing approach? 2.5

Scanned by CamScanner

e) What is the difference between debugging and testing? 2.5

f) Show the difference between effective testing and exhaustive testing. 2.5

6. **Answer any ONE** 10x1=10

a) A wholesaler has three commodities to sell and has three types of customers. Discount is given as per the following procedure: 10

(i) For DGS & D orders, 10% discount is given irrespective of the value of the order.

(ii) For orders of more than Tk 50,000, agents get a discount of 15% and the retailer gets a discount of 10%.

(iii) For orders of Tk 20,000 or more and up to Tk 50,000, agents get 12% and the retailer gets 8% discount.

(iv) For orders of less than Tk 20,000, agents get 8% and the retailer gets 5% discount.

The above rules do not apply to the furniture items wherein a flat rate of 10% discount is admissible to all customers irrespective of the value of the order.

Design test cases for this system using decision table testing.

b) Consider the following registration page for using the tax web portal – 10

National Id

Email Adress

Username

Password

Submit

The fields in the form have the following properties -

i) The National Id is a 17-digit number where the first 4 digits indicate the birth year. The birth year should be such so that at the current year, the person's age falls within the range of 18 to 65. For this problem the current year is 2014.

ii) The pattern of the email address is as follows: it has two parts separated by the symbol @. The first part is alphanumeric, should not start with a number and it should be 6-12-character long. The second part should be 6-20 character long and must end with a top level domain (for this problem consider only *com* and *org*) preceded by the *dot* symbol.

iii) The user name should be alphanumeric, should not start with a number, 8-16 character long and must have any of these characters – *dot* and *underscore*.

iv) The password should be 6-24 characters long and must have a special character.

Now answer the following questions –

(a) List valid and invalid equivalence classes for each of the form fields.

(b) Propose the boundary values for this scenario.

Scanned by CamScanner

Shahjalal University of Science and Technology
Institute of Information and Communication Technology
Software Engineering
Term Test 2

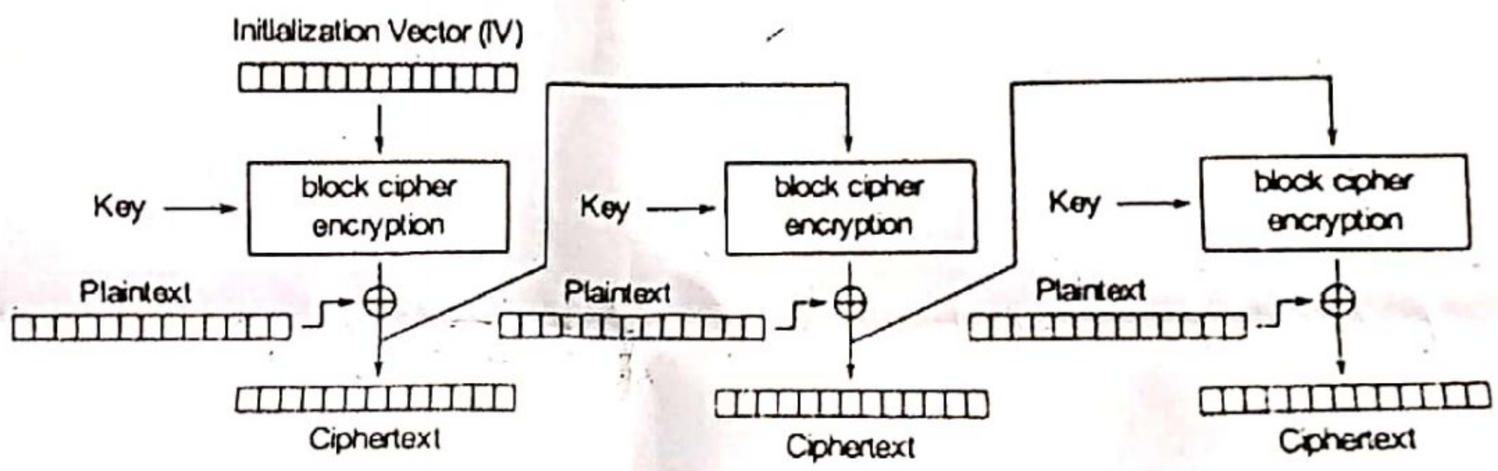
Course Code: SWE 337 **Credits: 2** **Course Title: Computer, Data and Network Security**
Time: 45 Minutes Total Marks: 20

1. Encrypt the text "CAB BCC ACC B" using Block Cipher with padding where
   m = 3;
   $$k = \begin{bmatrix} 1 & 2 & 3 \\ 5 & 6 & 7 \\ 9 & 10 & 11 \end{bmatrix};$$
   For the first block $\vec{x} = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$

   And for the last block $\vec{x} = \begin{bmatrix} 1 \\ 9 \\ 3 \end{bmatrix}$

   Hints: The last block will become B22 after padding. And you may assume that there are at most 4 characters in the entire character set. **Marks: 10**

2. Can you identify any problem with the above mechanism? How can you improve the above mechanism? (Not more than 3 lines/points) **Marks: 2**
3. Write the problems with ECB mode of Block Cipher. (Not more than 3 lines/points) **Marks: 2**
4. What would be the problem of the following Block Cipher architecture and why? Write within 6 lines/points. **Marks: 6**



Cipher Feedback (CFB) mode encryption

5 mod 25

25)5(

44
25
19

Scanned by CamScanner

Shahjalal University of Science and Technology
Institute of Information and Communication Technology
B.Sc. (Engg.) in Software Engineering
3rd Year 2nd Semester Final Examination, 2019
Session: 2016-17

Course No.: SWE 337 **Course Title: Computer, Data and Network Security**

Credits: 2 **Time: 2 hrs** **Total Marks: 50**

Group A

[Answer all the questions]

1. Answer any FIVE. **5x1=5**

a) What is the role of a Certificate Authority?
b) Write the full form of AES.
c) Why are public keys used?
d) Why is padding used in encryption?
e) Write the four modes of block cipher.
f) Vigenère cipher uses a key that is as long as the plain text. Is it True or False?

2. Answer any FOUR. **4x2.5=10**

a) What is Fail Safe Default? Give a practical example of a popular system that violates this principle.
b) A computer can test a password in every nanosecond. How much time will it take for an attacker to crack the password of length 10 (the password may consist of any of the 128 ASCII characters)?
c) Describe how a Brute Force attack happens.
d) What is the difference between an Encryption Algorithm and a Hashing Algorithm?
e) What is a Social Engineering Attack?

3. Answer any TWO. **2x5=10**

a) What do you understand by Data Confidentiality? Describe the steps to achieve Data Confidentiality.
b) Between CFB and OFB which one has a better encryption architecture? Why?
c) Suppose we have a network of 2300 users. They want to communicate in a secretive manner.
   i. How many keys are required if we use a Symmetric Key Cryptosystem?
   ii. How many key pairs are required if we use an Asymmetric Key Cryptosystem?

$n(n-1)$
$2n$

$10^9 = 1$

Scanned by CamScanner

[Answer all the questions]

4. Answer any FIVE. **5x1=5**

a) In CBC mode we can take advantage of pre computed IV. Is it True or False? If false give correct answer.

b) What is a "Quid pro quo" attack?

c) Write 3 components of a Crypto System.

d) "Brute-force decryption is likely to succeed for messages in natural language that are not too short". Is it True or false? If false give correct answer.

e) Why are RAIDs used?

f) What is cryptanalysis?

5. Answer any FOUR. **4x2.5=10**

a) How can you fulfill availability triad of a security system? Can you give an example?

b) Describe Man In The Middle attack.

c) The AA Corporation has a new refrigerator, the Fridge, which has a camera that takes a picture of the contents of the refrigerator and uploads it to the AA Corporation's web site. The Fridge's owner can then access this web site to see what is inside their refrigerator without opening the door. For security reasons, the AA Corporation encrypts this picture and gives a 4-digit PIN to decrypt this picture to the Fridge's owner, so he or she can get access to the pictures of their Fridge's interior. Suppose you have been hired as a security expert to find holes in the system. How will you test the system?

d) Suppose we are using ASCII encoding. Now each English alphabet is estimated to carry 1.25 bits of information. Calculate the probability that a randomly selected sequence of 32 bits is a meaningful text message.

e) What are the problems of security questions in password reset systems? Can you give a real life example of a hacking that was performed based on this?

6. Answer any TWO. **2x5=10**

a) How can Message Authentication Codes be used to ensure Data Integrity? Describe.

b) Encrypt the text "CBB BAC ABC B" using Block Cipher with padding where m = 3;

$$k = \begin{bmatrix} 1 & 2 & 3 \\ 5 & 6 & 7 \\ 9 & 10 & 11 \end{bmatrix};$$

For the first block $\vec{x} = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$

And for the last block $\vec{x} = \begin{bmatrix} 1 \\ 3 \\ 3 \end{bmatrix}$ B22

Hints: The last block will become ~~B22~~ 33 after padding and you may assume that there are at most 4 characters in the entire character set.

c) Suppose Rupa wants to send a secret message to Emma. There is possibility of eavesdropping. Now propose a mechanism that will ensure both Confidentiality and Integrity of the data. What is this technology called? There needs to be way so that Emma can be assuring of the fact that the message came from ~~Kamal~~ Rupa only and nobody else.

Scanned by CamScanner

Shahjalal University of Science and Technology
Institute of Information and Communication Technology
Software Engineering
3rd Year 2nd Semester Final Examination' Dec 2019 (Session: 2016-17)
Course Code: SWE 335 Credits: 2 Course Title: Data Science
Time: 2 hrs Total Marks: 50

Group A

[Answer all the questions]

5*1=5

1. Answer any FIVE

a) Define datum and data.
b) What is Residual Standard Error?
c) Define population and sample.
d) What is Messy data? What are the common causes of messiness?
e) What is selection bias?
f) State the complexities of computing mean and median of data.
g) What is 95% confidence interval for linear regression?

4*2.5=10

2. Answer any FOUR

a) Write the Data Science Process.
b) How is data represented and stored? *Give examples.*
c) Differentiate between regression and classification.
d) Assume there are 4 features (X_1, X_2, X_3, X_4) in a data set, where the interaction of X_2 and X_4 might have some impact on the response. Write the appropriate Linear Regression model for it.
e) Write down the imputation methods for missing data.
f) What is cross validation? Describe k-fold cross validation.

2*5=10

3. Answer any TWO

a) *Perform KNN regression for the following data and show the predictions for all the X values. Assume k = 2.*

X	1	2	3	4	5
Y	1	2	5	4	5

4.45

b) *Use K-means to cluster the given data: {20, 3, 9, 10, 9, 3, 1, 8, 5, 3, 24, 2, 14, 7, 8, 23, 6, 12, 18} into 3 groups (use 2 iterations).*

c) What is MSE? Calculate the MSE for the given model.

X	1	2	3	4	5
Y	6	7	4	3	2
$\hat{Y}$	6.8	5.6	4.4	3.2	2

3.625

Scanned by CamScanner

Group B

[Answer all the questions]

5*1=5

4. **Answer any FIVE**

a) Write down the type of missingness.

b) What are interacting predictors?

c) What is overfitting?

d) What is Variance and Standard Deviation?

e) What is agglomerative clustering? Give example.

f) Write a polynomial regression model of degree M.

g) What is Web Scrapping?

4*2.5=10

5. **Answer any FOUR**

a) What are the causes of over-fitting?

b) Show the difference between Prediction and Estimation.

c) What is LASSO Regression?

d) Define Entropy and Information Gain.

e) Use hierarchical cluster to cluster the given data {3, 7, 10, 16, 18, 20}. Show each steps.

f) Differentiate between parametric and non-parametric models with example.

2*5=10

6. **Answer any TWO**

a) For the given data, approximate the values of B_0 and B_1 and write the equation for simple liner regression.

X	1	2	3	4	5
Y	6	7	4	3	2

b) Construct the decision tree to decide what to do in the evening, find the root (use the given dataset).

Deadline?	Is there a Party?	Lazy?	Activity
Urgent	Yes	Yes	Party
Urgent	No	Yes	Study
Near	Yes	Yes	Party
None	Yes	No	Party
None	No	Yes	Pub
None	Yes	No	Party
Near	No	No	Study
Near	No	Yes	TV
Near	Yes	Yes	Party
Urgent	No	No	Study

Party
Study
Pub
TV

c) What is hypothesis testing? Write the steps of testing a hypothesis.

Scanned by CamScanner

Shahjalal University of Science and Technology
Department of Computer Science and Engineering
3rd Year 2nd Semester Final Examination' Dec 2019 (Session: 2016-17)
Course Code: CSE 345 Credits: 2 Course Title: Data Science
Time: 2 hrs Total Marks: 50

Group A

[Answer all the questions]

5x1=5

1. Answer any FIVE

a) What is Structured Data?
b) Define Quantitative Variable with examples.
c) What is a Scatter Plot?
d) Define Variance of a Sample.
e) What is a Statistical Model?
f) What is a Loss Function?
g) How to interpret a P-Value?
h) If B is a Linear Regression Coefficient and the Standard Error $(B) = 1.5$, then find the 95% Confidence Interval of B.

4x2.5=10

2. Answer any FOUR

a) Describe the process of Data Science.
b) Define Data, Distribution of Data, Population, Sample and Bias.
c) Name and define different Measures of Centrality; Mention the Computational Complexity of each.
d) What are the principles of Data Visualization?
e) What is a Histogram? Draw the Histogram of the following data:

Height(Feet)	Frequency
0-2	0
2-4	1
4-5	4
5-6	8
6-8	2

f) What are the Two main types of Prediction Problems we discussed in this course? Differentiate between them.

2x5=10

3. Answer any TWO

a) Assume there are 4 predictors (X_1, X_2, X_3, X_4) in a data set, where the interaction of X_2 and X_4 might have some impact on the response.

i. Write the appropriate Linear Regression Model for it.
ii. Write the Final Model if you get the following P-Values after performing hypotheses testing on the significance of the predictors.

p value < 0.05
$B_1 \rightarrow$ sign

Coefficients	P-Value
B_0	0.00
B_1	0.09
B_2	0.00
B_3	0.03
B_4	0.00
B_5	1.00

$\hat{y} = B_0 + B_1 X_1 + B_2 X_2 + B_3 X_3 + B_4 X_4 + B$

iii. What can you tell about the signification of the predictors?

b) What is F statistics? The true values and predicted values of a model are given below. Calculate the F statistics for the data. And interpret the result.

X	1	2	3	4	5
Y	1	2	.5	4	5
Y(Predicted)	3.5	3	3	5	4.5

c) Perform Linear Regression on the following data to find the equation of the Regression line.

Data	Age X	Glucose Level Y
1	43	99
2	21	65
3	25	79
4	42	75

Scanned by CamScanner

Group B

[Answer all the questions] **5x1=5**

4. Answer any FIVE

a) What is Logistic Regression?

b) What is Ensemble Learning?

c) What is a Random Forest?

d) Define Entropy and Information Gain.

e) Write the Bayes' Theorem.

f) What is a ROC Curve?

g) What is bootstrapping?

h) Write a polynomial regression model of degree M.

4x2.5=10

5. Answer any FOUR

a) What is Over-fitting? What are the causes of Over-fitting?

b) Define the steps to choose the subset of significant predictors using K-fold cross validation.

c) Why Regularization is used? Define L_1 and L_2 Regularization.

d) Differentiate between Parametric and Non-Parametric Models with example.

e) What is the most commonly used Loss Function? Define it and calculate its value from the following prediction.

X	1	2	3	4	5
Y	1	2	5	4	5
Y(Predicted)	3.5	3	3	5	4.5

f) How to deal with Missing Values in data?

2x5=10

6. Answer any TWO

a) What is Principal Component Analysis? Write the steps of computing principal components of N data with J features.

b) Suppose you have written a classifier to detect which images in your favorite social network are selfies. You have tested your classifier with some data and got the following predictions.

	Target	Prediction
1	Selfie	Not Selfie
2	Selfie	Selfie
3	Not Selfie	Not Selfie
4	Selfie	Selfie
5	Not Selfie	Selfie
6	Not Selfie	Not Selfie
7	Not Selfie	Not Selfie
8	Selfie	Selfie
9	Selfie	Not Selfie
10	Not Selfie	Not Selfie

Compute the Confusion Matrix for above and calculate Accuracy from it.

c) Briefly describe the process of building a Decision Tree.

Scanned by CamScanner