Computer & Data Science: second-round sample tasks

Module 1. Simple test questions with one correct answer (1 point for the correct answer)

Section "Applied Mathematics"

Task 1

Use the data in matrix $A = \begin{bmatrix} 12 & -1 \\ 2 & 11 \\ 1 & -1 \end{bmatrix}$, and select the set that will contain all the elements of the last row of the transposed matrix A.

a) {1, -1}
b) {1, 11, 1}
c) {12, 2, 1}
d) {-1, 11, -1}

Answer – d

Task 2

Given a set A = {1000_2, 110_2, 1100_2, 11000_2, 111_2}. Translate each element of set A from the binary number system to the decimal number system and find what will be the cardinality of the set if there will only be elements with GCD=2.

a) 4
b) 2
c) 1
d) 5

Answer – a

Task 3

There is a string: HelloWorld. Select a regular expression that matches this string (POSIX Extended).

a) /^[A-Z]$/
b) /^[A-Z]?[a-zA-Z]+$/
c) /^[A-Z]?[a-zA-Z]+[0-9]$/
d) /^[A-Z]?[a-z]+$ /

Answer – b

Task 4

Given a directed graph G = (V, E), where V – set of vertices, E – set of edges. If V = {1, 2, 3}, E = {(1; 2), (2; 1), (1;3), (3; 2)}, then …

a) graph G does not contain an Eulerian cycle
b) graph G contains an Eulerian cycle
c) the diameter of graph G is 1
d) the radius of graph G is 2

Answer – a

Section "Software Engineering"

Task 5

The Test Plan describes:
a) Testing procedure
b) Schedule for the implementation of test cases
c) Criteria for the end of testing

d) All of the above
Answer – d.

Task 6
What is the average estimate of the complexity of implementing tests in accordance with the plan compared to the total development costs:
a) 10%
b) 20%
c) 40%
d) 70%
Answer – c

Task 7
The criteria for system testing are:

a) Completeness of solving functional problems
b) Software stress resistance
c) Correctness of resource usage
d) All listed above
Answer – a

Task 8
Which of these models belong to the software lifecycle models:
a) Functional
b) V-shaped
c) Modular
d) Temporary
Answer – b

Section "Information systems and computing"
Task 9
Which of the options below can be an interface for a barcode label scanner?
a) COM
b) DRAM
c) DVI
d) FDD
Answer – a

Task 10
What type of RAID system does the following statement apply to: "The read speed doubles, but the write speed remains unchanged"?
a) RAID 0
b) RAID 1
c) RAID 2
d) RAID 3
Answer – b

Task 11
What is the maximum number of words L with m letters that can be composed from an alphabet with a power of N, if the words consist of 5 letters and the alphabet power is 8?

a) 2048
b) 128
c) 10340
d) 32768
Answer – d

Task 12
Determine the number of sound signal levels when using a 12-bit sound card in a computer
a) 1012
b) 4096
c) 3044
d) 6192
Answer – b

Section "Information security"
Task 13
What is the term for a channel that is used to transfer data in a manner not originally designed for that purpose?

a) an insecure channel
b) a secure channel
c) a covert channel
d) a side channel
Answer – c

Task 14
What abbreviation does not stand for an operation mode of block cypher?
a) CBC
b) OFB
c) ECB
d) BCF
Answer – d

Task 15
What virus encrypts itself and mutates a decryptor?
a) a polymorphic virus
b) a polygamous virus
c) a homomorphic virus
d) a pseudovirus
Answer – a

Task 16
What process can be described in short as a one-way data transformation using a certain function?
a) XORing
b) encryption
c) archiving
d) hashing
Answer – d

Section "Data preprocessing and analysis"

Task 17

Data Mining technology is used for (choose one answer):

a) Intelligent data analysis;
b) backing up data;
c) organizing and storing data;
d) information transmission.

Answer – a

Task 18

List the disadvantages of Gantt chart data analysis.

a) inflexibility, dependence, overflow;
b) inflexibility, independence, overflow;
c) flexibility, dependence, overflow;
d) flexibility, independence, overflow.

Answer – a

Task 19

Which classification metric does not depend on the threshold?

a) Precision
b) Recall
c) AUC-ROC
d) Accuracy

Answer – c

Task 20

Which of the knowledge representation models corresponds to the ideas about the organization of human long-term memory (choose one answer)?

a) production;
b) semantic networks;
c) frames;
d) integral.

Answer – b

Module 2. Complex test questions with several correct answers (3 or 4 points for the correct solution)

Section "Applied Mathematics"

Task 21

The function $f(x, y, z) = 0001\ 0101$ is given. Select **false** statements regarding the normalization result of the function f:

a) PKNF: $xz \oplus yz \oplus xyz$
b) PDNF: $\bar{x}yz \vee x\bar{y}z \vee xyz$
c) PDNF: $x\bar{y}\bar{z} \vee \bar{x}y\bar{z} \vee \bar{x}\bar{y}\bar{z}$
d) PKNF: $(x \vee y \vee z) \cdot (x \vee y \vee \bar{z}) \cdot (x \vee \bar{y} \vee z) \cdot (\bar{x} \vee y \vee z) \cdot (\bar{x} \vee \bar{y} \vee z)$

Answer – a, c

2 correct answers - 3 points

1 correct answer - 1 point
Maximum 3 points.

Task 22

The Boolean function is given as a formula $\bar{x}\bar{y} \vee y\bar{z} \vee xyz$. Choose equivalent ways of setting this function from the options presented below:

a) $\bar{x}\bar{y}z \vee \bar{x}\bar{y}\bar{z} \vee xy\bar{z} \vee \bar{x}y\bar{z} \vee xyz$
b) 1100 0011
c) 1110 0011
d) $\bar{x}\bar{y}\bar{z} \vee \bar{x}\bar{y}z \vee xy\bar{z} \vee xyz$

Answer – a, c
2 correct answers - 4 points
1 correct answer - 2 points
Maximum 4 points.

Section "Software engineering"

Task 23

What terms are used in software development:
a) Coding
b) Testing
c) Justifications
d) Destruction
e) Maintenance
f) Destruction

Answer – a, b, e
3 correct answers - 3 points
2 correct answers - 2 points
1 correct answer – 1 point
Maximum 3 points.

Task 24

What types of testing are used in software design?
a) Module testing
b) System testing
c) Periodic testing
d) Integration testing
e) Thematic testing
f) Game testing

Answer – a, b, d
3 correct answers - 4 points
2 correct answers - 2 points
1 correct answer – 1 point
Maximum 4 points.

Section "Information systems and computing"

Task 25

Name the protocols used in storage systems of the SAN category
a) jjSCSi

b) FCoE
c) SCSI
d) SRP
e) pSCSI
f) CPP

Answer – b, c, d
3 correct answers - 3 points
2 correct answers - 2 points
1 correct answer – 1 point
Maximum 3 points.

Task 26

By IPv4 address 123.221.125.12/25, set the four correct network configuration parameters:
a) IP address of the first host 123.221.125.0
b) IP address of the first host 123.221.125.1
c) IP address of the last host 123.221.125.126
d) The IP address of the last host is 123.221.125.254
e) Broadcast address 123.221.125.127
f) Broadcast address 123.221.125.255
g) Subnet mask 255.255.255.128
h) Subnet mask 255.255.255.254

Answer – b, c, e, g
4 correct answers - 4 points
3 correct answers - 3 points
2 correct answers - 2 points
1 correct answer – 1 point
Maximum 4 points.

Section "Information security"

Task 27

What steps are performed in AES (Rijndael) round function?
a) Shifting rows
b) Shifting columns
c) Mixing rows
d) Mixing columns
e) Table-based substitution
f) XORing with a round key

Answer – a, d, e
3 correct answers - 3 points
2 correct answers - 2 points
1 correct answer – 1 point
Maximum 3 points.

Task 28

The key used for the XOR cipher is generated using a modulo-17 linear congruential generator. It has been zero-initialized. The first two elements of the generated sequence are 3 and 7. What is the next element of the sequence?

Answer – 1
Maximum 4 points

Section "Data preprocessing and analysis"
Task 29
Imagine a bank customer as an object within a machine learning task.
Which of the options is the task of binary classification?
a) Predicting credit amounts;
b) Predicting when the customer will return the loan;
c) Predicting whether the customer will use the early repayment service;
d) Predicting whether the customer will connect the mobile bank.

Answer – c, d
2 correct answers - 3 points
1 correct answer - 1 point
Maximum 3 points.

Task 30
What is the objective of data storage (DS (DS)?
a) to ensure the collection, storage and quick access to the key information;
b) for create real-time systems;
c) to the operational analytical processing and data mining;
d) to view video files.
Answer – a, c
2 correct answers - 4 points
1 correct answer - 2 points
Maximum 4 points.

Module 3. Tasks with a detailed answer (15 points for the correct answer)

Section "Applied Mathematics"
Task 31

Consider the courses Algebra (A), Binary Trees (B), Calculus (C), and Differential Equations (D). Calculate the total number of students (N) residing in a dormitory using the provided data. Then, express the answer in the base-14 positional numeral system, which includes the digits 0 to 9 and the letters A, B, C, and D (0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D).
12 take A, 5 take A and B, 4 take B and D, 2 take B, C, D,
18 take B, 7 take A and C, 3 take C and D, 3 take A, C, D,
22 take C, 5 take A and D, 3 take A, B, C, 1 take all four,
8 take D, 16 take B and C, 3 take A, B, D, 64 take none.

Let T represent the total number of students who are enrolled in at least one course. We can determine T using the Inclusion-Exclusion Formula, where:
$s_1 = 12 + 18 + 22 + 8 = 60$,
$s_2 = 5 + 7 + 5 + 16 + 4 + 3 = 40$,
$s_3 = 3 + 3 + 2 + 3 = 11$,
$s4 = 1$.

Thus T = 30, and N = 64 + T = 94. 94_{10}= 6*14+10=$6A_{14}$

The correct use of the Inclusion-Exclusion formula - 7 points;
for the correct calculation - 3 points;
for the correct translation into the number system - 5 points.

Section “Information security”
Task 32:
Alice and Bob have chosen to create a secure channel for exchanging messages. Unfortunately, an intruder intercepted their conversation through an insecure channel:

Alice: "We should employ the XOR cipher to encrypt our messages. We'll encrypt each character using a key established by the Diffie-Hellman protocol."

Bob: "Agreed. Let's utilize the following numbers as algorithm parameters: 3D and FB. Let’s start"
Alice: 4D
Bob: E5
Alice: D9 D4 DD DD DE
What plaintext word was sent by Alice?
Only characters from Table 1 are allowed in plaintext.

Table 1. Characters and their codes in decimal and hexadecimal format

decimal	hex	character		decimal	hex	character
65	41	A		84	54	T
66	42	B		85	55	U
67	43	C		86	56	V
68	44	D		87	57	W
69	45	E		88	58	X
70	46	F		89	59	Y
71	47	G		90	5A	Z
72	48	H		48	30	0
73	49	I		49	31	1
74	4A	J		50	32	2
75	4B	K		51	33	3
76	4C	L		52	34	4
77	4D	M		53	35	5
78	4E	N		54	36	6
79	4F	O		55	37	7
80	50	P		56	38	8
81	51	Q		57	39	9
82	52	R		58	3A	:
83	53	S				

Solution: Upon analyzing the intercepted messages, it becomes evident that Alice and Bob are using 3D = 61 and FB = 251 as the algorithm parameters. Consequently, Alice transmitted the following message:

$$4D = 77 = 61^{x_A} \bmod 251,$$

Bob replied:

$$E5 = 229 = 61^{x_B} \bmod 251$$

We need to determine the values of x_A or x_B to compute the secret key using the equation:

61^(x_A x_B) mod 251.

In real-life scenarios, calculating these values can be challenging. However, in this particular case where non-secure algorithm parameters are used, we can employ a brute-force method to find them. Bob suggested using non-secure parameter values.
So, we get $x_A = 3$, $x_B = 7$.
(Correct x_A or x_B value – 5 points)
Then the shared secret key is

$$\text{Key} = 61^{x_A x_B} \bmod 251 = 77^7 \bmod 251 = 229^3 \bmod 251 = 145 = 91$$

(Correct shared secret key value – 5 points)

Finally, the message can be decrypted:

D9	D4	DD	DD	DE
11011001	11010100	11011101	11011101	11011110
		⊕		
10010001	10010001	10010001	10010001	10010001
		=		
01001000	01000101	01001100	01001100	01001111
48	45	4C	4C	4F
H	E	L	L	0

(Correct plaintext message – 5 points)
Answer: **HELLO**
(Total: 15 points)

Section "Data preprocessing and analysis"
Task 33
There is a sample containing 30 numerical values of some sign of a random variable *X*:

19	25	22	16	22	14	17	19	18	20
22	26	24	18	16	19	22	14	18	14
25	17	18	14	20	18	24	25	16	18

Create a sampling distribution and calculate the sample mean.
Solution:
The sampling distribution is a tabular representation where the first row consists of the possible values of a random variable arranged in ascending order, and the second row contains their corresponding frequencies, indicating how frequently each value occurred.

Assessing the provided definition of the sampling distribution - (3 points).

x_i	14	16	17	18	19	20	22	24	25	26
n_i	4	3	2	6	3	2	4	2	3	1

Creating a table - 4 points

Sample mean: $x_{\text{в}} = \frac{\sum n_i * x_i}{n}$

Writing the sample mean formula - 4 points

$$x_{\text{в}} = \frac{14 * 4 + 16 * 3 + 17 * 2 + 18 * 6 + 19 * 3 + 20 * 2 + 22 * 4 + 24 * 2 + 25 * 3 + 26 * 1}{30}$$

Sample mean = 19, 3.

Correct calculation of the sample mean – 4 points