

**Corso di Data Mining e Bioimmagini (A.A. 2023/2024)**  
**CdL Medicina e Chirurgia TD**  
**Simulazione di Prova scritta 15 Gennaio 2024 (Durata prova 90minuti) {BOZZA}**

Prima parte Modulo di DM

1) Definire l'importanza della statistica per caratterizzare i dati di un campione di pazienti in uno studio clinico.

2) Data la tabella seguente, definire alcuni esempi di analisi statistica per caratterizzare i dati.

AGE	GENDER (1=M)	PULSE	SYSTOLIC		DIASTOLIC		HDL	LDL	WHITE	RED	PLATE	WEIGHT	HEIGHT	
43	0	80	100	70	73	68	8.7	4.80	319	98.6	172.0	120.4	40.7	33.3
57	1	84	112	70	35	116	4.9	4.73	187	96.9	186.0	107.8	37.0	28.0
38	0	94	134	94	36	223	6.9	4.47	297	108.2	154.4	120.3	44.3	45.4
80	1	74	126	64	37	83	7.5	4.32	170	73.1	160.5	97.2	30.3	28.4
34	1	50	114	68	50	104	6.1	4.95	140	83.1	179.0	95.1	34.0	25.9
77	1	60	134	60	55	75	5.7	3.95	192	86.5	166.7	112.0	31.4	31.1
29	1	52	118	64	53	128	4.1	4.68	191	64.1	178.5	78.0	27.4	20.1
69	0	58	138	80	40	140	8.1	4.60	286	79.2	155.7	103.5	34.2	32.7
44	0	66	114	66	45	136	8.0	4.09	263	64.2	157.6	89.7	32.5	25.8
35	1	62	124	70	62	110	5.6	5.47	193	118.8	180.4	112.0	40.0	36.5
79	1	52	138	58	54	92	8.3	4.44	193	71.3	166.3	95.0	32.7	25.8
72	0	56	110	72	53	102	6.9	4.15	215	98.2	168.6	115.3	38.5	34.5
18	1	76	116	40	45	79	5.1	5.07	162	122.7	181.1	118.8	41.2	37.4
65	1	52	134	66	41	100	9.5	4.38	646	75.3	170.0	92.6	30.4	26.1
37	0	82	100	80	55	94	8.1	5.22	250	51.8	164.0	75.5	27.0	19.3
30	1	62	124	66	30	47	6.1	4.32	265	89.5	180.8	92.5	36.2	27.4
44	0	86	100	50	59	119	6.3	4.30	218	75.9	155.4	100.8	33.5	31.4
51	0	88	106	72	67	129	10.9	6.34	503	67.1	178.8	82.8	26.6	21.0

3) Costruire un esempio di tabella di popolazione per il dominio di infettivologia e descrivere metodi di utilizzo per predizione di patologie

4) Dato il seguente dataset identificare i valori anomali e proporre delle soluzioni di preprocessing prima di analizzare il dataset

AGE	GENDER (1=M)	PULSE	SYSTOLIC	DIASTOLIC	HDL	LDL	WHITE	RED	PLATE	WEIGHT	HEIGHT			
43	0	80	100	70	73	68	8.7	4.80	319	98.6	172.0	120.4	40.7	33.3
57	1	84	112	70	35	116	4.9	4.73	187	96.9	186.0	107.8	37.0	28.0
57	1	84	112	70	35	116	4.9	4.73	187	96.9	186.0	107.8	37.0	28.0
38	0	94	134	94	36	223	6.9	4.47	297	108.2	154.4	120.3	44.3	45.4
80	1	74	126	64	37	83	7.5	4.32	170	73.1	160.5	97.2	30.3	28.4
34	1	50	114	68	50	104	6.1	4.95	140	83.1	179.0	95.1	34.0	25.9
77	1	60	134	60	55	75	5.7	3.95	192	86.5	166.7	112.0	31.4	31.1
57	1	84	112	70	35	116	4.9	4.73	1000	96.9	186.0	107.8	37.0	28.0
--	1	-52	64	53	000	4.1	4.68	191	64.1	0000	78.0	27.4	20.1	
69	0	58	138	80	40	140	8.1	4.60	286	79.2	155.7	103.5	34.2	32.7
44	0	66	114	66	45	136	8.0	4.09	263	64.2	157.6	89.7	32.5	25.8
35	1	62	124	70	62	110	5.6	5.47	193	118.8	180.4	112.0	40.0	36.5
-35	1	00	124	70	62	110	5.6	5.47	193	118.8	180.4	xyv	40.0	36.5

5) Indicare quali sono le fasi di processo per una procedura di data mining su un dataset

(preprocessing, classificazione, clustering, supporto alle decisioni)

6) Indicare le fasi essenziali di un processo di data mining

Classificazione [Predittiva] si hanno le classi e voglio spiegarle e descriverli in termini di caratteristiche  
Clustering [Descrittiva] – scoprire gruppi di dati simili

Ricerca di regole associative [Descrittiva]  
Ricerca di pattern sequenziali [Descrittiva]  
Regressione [Predittiva] descrivere le forme (funzione) che  
approssima i dati  
Individuazione di deviazioni [Predittiva]

- 7) Descrivere le principali tecniche di classificazione e la regola di training e testing
- 8) Riportare degli esempi di algoritmi di clustering ed almeno un esempio su un dataset da te fornito
- 9) Cosa e a cosa servono gli alberi decisionali e riportare un esempio
- 10) Descrivere le metriche utilizzabili in un processo di data mining
- 11) Descrivere la tecnica KNN con un esempio
- 12) ....
- 13) (in dettaglio su classificazione) dato un insieme di regole di inferenze per la classificazione dire se rispettano le regole di copertura

### **Modulo di Bioimmagini**

- 1) descrivere le modalità di digitalizzazione di una immagine
- 2) descrivere la Tomografia assiale
- 3) su quale principio si basa la tecnica di RM
- 4) come si formano le immagini ecografiche
- 5) Quale è il principio di funzionamento dei raggi X
- 6) Quale è la principale differenza tra TC e RM
- 7) Descrivere l'uso di radiofarmaci nella diagnostica per immagini
- 8) Descrivere il funzionamento della PET tramite isotopi radioattivi