

Prima prova scritta – Traccia A

Determinazione INVALSI n. 185 del 31/08/2022 Concorso pubblico, per titoli ed esami, per l'assunzione a tempo pieno e indeterminato presso l'INVALSI di n. 1 (una) unità di personale nel profilo di Funzionario, V livello professionale di cui al D.P.R. n. 171/1991

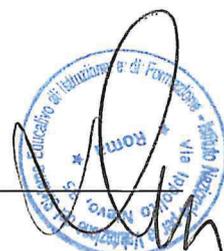
1. Cosa si intende per stemming nell'ambito del text mining?
 - Ridurre una parola alla sua radice
 - Identificare la categoria grammaticale di una parola
 - Convertire una frase in parole
 - Nessuna delle precedenti

2. Quale dei seguenti algoritmi NON può essere usato nella classificazione dei documenti?
 - Decision tree
 - Support vector machine
 - K-means clustering
 - Naive Bayes

3. Cosa misura la Term Frequency (TF)?
 - La frequenza di un termine all'interno di un documento
 - Il numero dei caratteri che compongono un documento
 - La frequenza di un termine all'interno dell'intero corpus (insieme dei documenti)
 - Il numero dei caratteri che compongono l'intero corpus (insieme dei documenti)

4. Cosa si può rappresentare attraverso una word cloud (nuvola di parole)?
 - L'associazione tra termini che si presentano più spesso insieme ad altri termini
 - Le frequenze delle parole più comuni nel corpus (insieme dei documenti) con un grafico a barre
 - I termini che ricorrono più frequentemente nel corpus (insieme dei documenti), differenziati per colore e per dimensioni
 - La frequenza di un termine utilizzato in due soli documenti a confronto

5. Se "X" è una variabile casuale, $E(X)$ è il valore atteso della variabile casuale e "a" e "b" sono due costanti allora:
 - $E(aX + b) = b^2E(X)$
 - $E(aX + b) = a^2E(X)$
 - $E(aX + b) = (a+b)^2E(X)$
 - $E(aX + b) = aE(X) + b$



Prima prova scritta – Traccia A

Determinazione INVALSI n. 185 del 31/08/2022 Concorso pubblico, per titoli ed esami, per l'assunzione a tempo pieno e indeterminato presso l'INVALSI di n. 1 (una) unità di personale nel profilo di Funzionario, V livello professionale di cui al D.P.R. n. 171/1991

6. Nell'ambito dell'inferenza statistica che cos'è l'Errore di Prima Specie (o di Tipo I)?
- Il quadrato dell'errore di Seconda Specie
 - L'errore che si commette quando un test statistico indica che va accettata un'ipotesi nulla che invece è falsa
 - Il complemento a 1 dell'errore di Seconda Specie
 - L'errore che si commette quando un test statistico indica che va rifiutata l'ipotesi nulla che invece è vera
7. Se l'MSE (Mean Squared Error) tende a zero quando la numerosità campionaria tende a infinito allora uno stimatore si definisce:
- Efficiente
 - Completo
 - Corretto
 - Consistente
8. Quale tra queste affermazioni relative agli intervalli di confidenza è falsa?
- L'ampiezza dell'intervallo aumenta quando diminuiamo il livello di fiducia alfa
 - L'ampiezza dell'intervallo aumenta quando aumenta il valore della varianza campionaria
 - L'ampiezza dell'intervallo non dipende dal livello di fiducia alfa
 - L'ampiezza dell'intervallo diminuisce quando aumenta la numerosità campionaria
9. Un intervallo di confidenza al 95% per la media della popolazione è stimato da 15.48 a 20.52. Se ora viene stimato un intervallo di confidenza al 90% per la media della popolazione sarà?
- Più ampio di quello al 95%
 - Lo stesso dell'intervallo al 95%
 - Meno ampio di quello al 95%
 - Non ci sono abbastanza informazioni per rispondere

Prima prova scritta – Traccia A

Determinazione INVALSI n. 185 del 31/08/2022 Concorso pubblico, per titoli ed esami, per l'assunzione a tempo pieno e indeterminato presso l'INVALSI di n. 1 (una) unità di personale nel profilo di Funzionario, V livello professionale di cui al D.P.R. n. 171/1991

10. Quale tra queste affermazioni sul Classificatore bayesiano è falsa?

- Assegna ogni osservazione alla classe più verosimile, dati i valori di un predittore
- Produce il tasso di errore di test più basso possibile
- Si può sempre calcolare su dati reali senza presenza di dati mancanti
- Si basa su un'applicazione del teorema di Bayes

11. Quale tra queste affermazioni sul K-means clustering è falsa?

- Garantisce sempre di raggiungere ottimi globali
- Il numero di cluster deve essere scelto a priori
- Non è sensibile alla presenza di outlier
- È sensibile alle strutture dei cluster non sferiche

12. Nell'ambito della cluster analysis in cosa consiste il metodo del legame singolo?

- Calcolare tutte le dissimilarità tra le osservazioni tra due cluster e registrare la più grande tra queste differenze
- Calcolare tutte le dissimilarità tra le osservazioni tra due cluster e registrare la più piccola tra queste differenze
- Calcolare tutte le dissimilarità tra le osservazioni tra due cluster e registrare la media tra queste differenze
- Calcolare la dissimilarità tra il centroide di due cluster

13. Che cos'è un dendogramma?

- Un metodo di classificazione
- Una rappresentazione delle aggregazioni tra unità
- Una misura di distanza tra unità
- Una rappresentazione delle componenti principali

Prima prova scritta – Traccia A

Determinazione INVALSI n. 185 del 31/08/2022 Concorso pubblico, per titoli ed esami, per l'assunzione a tempo pieno e indeterminato presso l'INVALSI di n. 1 (una) unità di personale nel profilo di Funzionario, V livello professionale di cui al D.P.R. n. 171/1991

14. Quale delle seguenti misure può essere usata come criterio di divisione binaria ricorsiva per gli alberi di regressione?
- VIF (Variance Inflation Factor)
 - Indice di entropia
 - Indice di dissimilarità
 - MSE (Mean Squared Error)
15. Quali dei seguenti algoritmi NON può essere considerato un modello di machine learning supervisionato?
- SVM (Support Vector Machine)
 - KNN (K-Nearest Neighbors)
 - PCA (Principal Component Analysis)
 - Regressione Logistica
16. Cosa misura la precisione (precision) per un modello di machine learning?
- Il rapporto tra Veri Positivi (TP) e la somma dei Veri positivi (TP) e Falsi Positivi (FP). Ossia: $TP / (TP + FP)$
 - Il rapporto tra Veri Positivi (TP) e la somma dei Veri positivi (TP) e Falsi Negativi (FN). Ossia: $TP / (TP + FN)$
 - Il rapporto tra la somma dei Veri Positivi (TP) e Veri Negativi (TN) e il totale delle osservazioni. Ossia: $(TP + TN) / (TP + TN + FP + FN)$
 - Il rapporto tra Veri Positivi (TP) e Falsi Positivi (FP). Ossia: TP / FP
17. Nei modelli di regressione lineare la statistica F per quale dei seguenti motivi si può utilizzare?
- Misurare l'accuratezza della previsione sulla base del modello
 - Fornire una misura della bontà di adattamento ai dati
 - Sottoporre a verifica l'ipotesi che almeno uno dei predittori sia utile per prevedere la variabile risposta
 - Sottoporre a verifica l'ipotesi di normalità della variabile risposta

Prima prova scritta – Traccia A

Determinazione INVALSI n. 185 del 31/08/2022 Concorso pubblico, per titoli ed esami, per l'assunzione a tempo pieno e indeterminato presso l'INVALSI di n. 1 (una) unità di personale nel profilo di Funzionario, V livello professionale di cui al D.P.R. n. 171/1991

18. Quale metodo in Python permette di rimuovere caratteri all'inizio e alla fine di una stringa?
- trim()
 - str_trim()
 - strip()
 - isspace()
19. Quale comando permette di ottenere la prima riga di un DataFrame Pandas?
- df[0]
 - df[1]
 - df.loc[0]
 - df.loc[1]
20. Il metodo head() della libreria Pandas di Python quante righe restituisce di default se non si specifica alcun parametro?
- 10
 - Tutte le righe del DataFrame
 - 1
 - 5
21. Quale parametro della funzione dropna() presente nella libreria Pandas di Python permette di modificare il DataFrame piuttosto che crearne uno nuovo?
- how
 - original
 - keep
 - inplace
22. Qual è un modo corretto per inserire un commento in un codice Python?
- //Questo è il mio commento
 - /*Questo è il mio commento*/
 - # Questo è il mio commento
 - """Questo è il mio commento

Prima prova scritta – Traccia A

Determinazione INVALSI n. 185 del 31/08/2022 Concorso pubblico, per titoli ed esami, per l'assunzione a tempo pieno e indeterminato presso l'INVALSI di n. 1 (una) unità di personale nel profilo di Funzionario, V livello professionale di cui al D.P.R. n. 171/1991

23. La variabile stringa `my_string` contiene il seguente testo: "a b c d". Quale dei seguenti risultati viene restituito dalla ricerca: `my_string.find("d")`?
- 4
 - 5
 - 6
 - 7
24. Quale di queste collezioni di dati definisce un oggetto SET di Python?
- ["Mario", "Rossi"]
 - {"name": "Mario", "surname": "Rossi"}
 - ("Mario ", "Rossi")
 - {" Mario ", " Rossi "}
25. La variabile stringa `my_string` contiene il seguente testo: "a b c d". Quale dei seguenti risultati viene restituito dal comando: `my_string.split()`?
- abcd
 - abc
 - ['a','b','c','d']
 - {'a','b','c','d'}
26. Quale comando di SQL si può utilizzare per cancellare dati da un database?
- COLLAPSE
 - UPDATE
 - DELETE
 - REMOVE
27. Quale comando di SQL puoi utilizzare per selezionare una colonna denominata "CodiceScuola" da una tabella denominata "Istituti"?
- SELECT CodiceScuola from Istituti
 - EXTRACT CodiceScuola from Istituti
 - SELECT CodiceScuola.Istituti
 - SELECT Istituti(CodiceScuola)

Prima prova scritta – Traccia A

Determinazione INVALSI n. 185 del 31/08/2022 Concorso pubblico, per titoli ed esami, per l'assunzione a tempo pieno e indeterminato presso l'INVALSI di n. 1 (una) unità di personale nel profilo di Funzionario, V livello professionale di cui al D.P.R. n. 171/1991

28. Quale comando di SQL puoi utilizzare per selezionare tutti i record da una tabella denominata "Istituti" che hanno valore della colonna "CodiceMeccanografico" che inizia con "RM"?
- SELECT * from Istituti WHERE CodiceMeccanografico = '%RM%'
 - SELECT * from Istituti WHERE CodiceMeccanografico LIKE 'RM%'
 - SELECT * from Istituti WHERE CodiceMeccanografico LIKE '%RM'
 - SELECT * from Istituti WHERE CodiceMeccanografico = 'RM'
29. Quale comando puoi utilizzare per inserire un nuovo record in una tabella denominata "Istituti" in SQL?
- INSERT ('RMIC000000','RM') INTO Istituti
 - INSERT INTO Istituti VALUES ('RMIC000000','RM')
 - INSERT VALUES ('RMIC000000','RM') INTO Istituti
 - INSERT VALUES Istituti ('RMIC000000','RM')
30. Quale comando puoi utilizzare per contare il numero dei record nella tabella Istituti in SQL?
- SELECT COLUMNS(*) FROM Istituti
 - SELECT TOP(*) FROM Istituti
 - SELECT LEN(*) FROM Istituti
 - SELECT COUNT(*) FROM Istituti

Prima prova scritta – Traccia B

Determinazione INVALSI n. 185 del 31/08/2022 Concorso pubblico, per titoli ed esami, per l'assunzione a tempo pieno e indeterminato presso l'INVALSI di n. 1 (una) unità di personale nel profilo di Funzionario, V livello professionale di cui al D.P.R. n. 171/1991

1. Nell'ambito del text mining quale tecnica di pre-processing del testo permette di convertire la parola 'computer' in 'comput'?
 - Stemming
 - Tokenizzazione
 - Tagging delle parole
 - Lemmatizzazione

2. Nella frase "Analisi del testo", quanti bigrammi (2-gram) ci sono?
 - 2
 - 3
 - 4
 - 5

3. Nell'ambito del text mining cosa si intende per stopwords?
 - Sono le parole non di interesse per l'analisi che possono essere eliminate
 - Sono parole che chiudono le frasi
 - Sono le parole che identificano le colonne della document-term matrix
 - Sono parole poco comuni all'interno del corpus (insieme dei documenti)

4. In che modo la metrica TF (Term Frequency) influisce sul peso assegnato a un termine all'interno di un documento?
 - Diminuisce il peso dei termini che compaiono frequentemente nel corpus (insieme dei documenti)
 - Aumenta il peso dei termini che compaiono raramente in un documento
 - Aumenta il peso dei termini che compaiono frequentemente in un documento
 - Non ha nessun effetto sul peso assegnato ai termini in un documento

5. Se "X" è una variabile casuale, $Var(X)$ è la varianza della variabile casuale e "a" e "b" sono due costanti allora:
 - $Var(aX + b) = b^2Var(X)$
 - $Var(aX + b) = a^2Var(X)$
 - $Var(aX + b) = (a+b)^2Var(X)$
 - $Var(aX + b) = aVar(X) + b$



Prima prova scritta – Traccia B

Determinazione INVALSI n. 185 del 31/08/2022 Concorso pubblico, per titoli ed esami, per l'assunzione a tempo pieno e indeterminato presso l'INVALSI di n. 1 (una) unità di personale nel profilo di Funzionario, V livello professionale di cui al D.P.R. n. 171/1991

6. Nell'ambito dell'inferenza statistica che cos'è l'Errore di Seconda Specie (o di Tipo II)?
- L'errore che si commette quando un test statistico indica che va accettata un'ipotesi nulla che invece è falsa
 - Il quadrato dell'errore di Prima Specie
 - L'errore che si commette quando un test statistico indica che va rifiutata l'ipotesi nulla che invece è vera
 - Il complemento a 1 dell'errore di Prima Specie
7. Se il valore atteso dello stimatore è uguale al "vero" valore del parametro incognito allora lo stimatore si dice:
- Efficiente
 - Completo
 - Corretto
 - Consistente
8. Quale tra queste affermazioni relative agli intervalli di confidenza è falsa?
- L'ampiezza dell'intervallo aumenta quando aumentiamo il livello di fiducia alfa
 - L'ampiezza dell'intervallo aumenta quando aumenta il valore della varianza campionaria
 - L'ampiezza dell'intervallo non dipende dal livello di fiducia alfa
 - L'ampiezza dell'intervallo diminuisce quando diminuisce la numerosità campionaria
9. Un intervallo di confidenza al 90% per la media della popolazione è stimato da 15.48 a 20.52. Se ora viene stimato un intervallo di confidenza al 95% per la media della popolazione sarà?
- Più ampio di quello al 90%
 - Lo stesso dell'intervallo al 90%
 - Più stretto di quello al 90%
 - Non ci sono abbastanza informazioni per rispondere
10. Per quale motivo generalmente si utilizzano gli alberi di regressione?
- Individuare costrutti latenti
 - Effettuare previsioni su una variabile numerica
 - Raggruppare le unità statistiche in insiemi di dati
 - Individuare la presenza di valori anomali

Prima prova scritta – Traccia B

Determinazione INVALSI n. 185 del 31/08/2022 Concorso pubblico, per titoli ed esami, per l'assunzione a tempo pieno e indeterminato presso l'INVALSI di n. 1 (una) unità di personale nel profilo di Funzionario, V livello professionale di cui al D.P.R. n. 171/1991

11. Quale tra queste affermazioni sul K-means clustering è falsa?

- Permette di raggiungere ottimi locali
- Il numero di cluster non deve essere scelto a priori
- È sensibile alle strutture dei cluster non sferiche
- Non è sensibile alla presenza di outlier

12. Nell'ambito della cluster analysis in cosa consiste il metodo del legame medio?

- Calcolare tutte le dissimilarità tra le osservazioni tra due cluster e registrare la più grande tra queste differenze
- Calcolare tutte le dissimilarità tra le osservazioni tra due cluster e registrare la più piccola tra queste differenze
- Calcolare tutte le dissimilarità tra le osservazioni tra due cluster e registrare la media tra queste differenze
- Calcolare la dissimilarità tra i centroidi di due cluster

13. Che cos'è il VIF (Variance Inflation Factor)?

- Una misura della bontà di adattamento del modello ai dati osservati
- Una misura di significatività statistica per un insieme di variabili di regressione
- Una misura della quantità di collinearità per un insieme di variabili di regressione
- Una misura utile a verificare l'ipotesi di normalità in un modello di regressione lineare

14. Qual è la principale differenza tra l'approccio bagging e l'approccio random forest?

- L'approccio bagging non si basa su campioni di addestramento bootstrap
- L'approccio random forest non si basa su campioni di addestramento bootstrap
- Nell'approccio bagging viene scelto un campione casuale diverso di m predittori per ogni albero decisionale
- Nell'approccio random forest viene scelto un campione casuale diverso di m predittori per ogni albero decisionale

Prima prova scritta – Traccia B

Determinazione INVALSI n. 185 del 31/08/2022 Concorso pubblico, per titoli ed esami, per l'assunzione a tempo pieno e indeterminato presso l'INVALSI di n. 1 (una) unità di personale nel profilo di Funzionario, V livello professionale di cui al D.P.R. n. 171/1991

15. Quali valori sono riportati sull'asse delle ordinate della Curva di ROC?

- La Sensibilità
- La Specificità
- Il complemento a 1 della Sensibilità
- Il complemento a 1 della Specificità

16. Cosa misura il recupero (recall) per un modello di machine learning?

- Il rapporto tra Veri Positivi (TP) e la somma dei Veri positivi (TP) e Falsi Positivi (FP). Ossia: $TP / (TP + FP)$
- Il rapporto tra Veri Positivi (TP) e la somma dei Veri positivi (TP) e Falsi Negativi (FN). Ossia: $TP / (TP + FN)$
- Il rapporto tra la somma dei Veri Positivi (TP) e Veri Negativi (TN) e il totale delle osservazioni. Ossia: $(TP + TN) / (TP + TN + FP + FN)$
- Il rapporto tra Veri Positivi (TP) e Falsi Positivi (FP). Ossia: TP / FP

17. Che cosa implica l'ipotesi di omoschedasticità nei modelli di regressione lineare?

- La media della variabile risposta è costante per tutti i valori della variabile esplicativa
- La varianza della variabile risposta è costante per tutti i valori della variabile esplicativa
- La covarianza tra la variabile risposta e la variabile esplicativa è nulla
- Nessuna delle risposte precedenti

18. Quale funzione presente nella libreria Pandas di Python permette di rimuovere i valori nulli in un DataFrame?

- `naomit()`
- `dropna()`
- `remove_null()`
- `delmiss()`

19. Quale sintassi permette di ottenere le prime 10 righe di un DataFrame?

- `df.head(10)`
- `df.top(10)`
- `df.row(10)`
- `df.start(10)`

Prima prova scritta – Traccia B

Determinazione INVALSI n. 185 del 31/08/2022 Concorso pubblico, per titoli ed esami, per l'assunzione a tempo pieno e indeterminato presso l'INVALSI di n. 1 (una) unità di personale nel profilo di Funzionario, V livello professionale di cui al D.P.R. n. 171/1991

20. Quale dei seguenti NON è un nome valido per una variabile in Python?

- my_var
- _myvar
- global
- myvar1

21. Quale metodo in Python permette di convertire tutti i caratteri di una stringa in maiuscolo?

- capitalize()
- upper()
- uppercase()
- toupper()

22. Quale funzione presente nella libreria Pandas di Python permette di rimuovere i valori duplicati da un DataFrame?

- remove_duplicates()
- delete_duplicates()
- drop_duplicates()
- duplicates()

23. La variabile stringa my_string contiene il seguente testo: "a b c d". Quale dei seguenti risultati viene restituito dalla ricerca: my_string.find("c")?

- 2
- 3
- 4
- 5

24. Quale di queste collezioni di dati definisce un oggetto TUPLE di Python?

- ["Mario", "Rossi"]
- {"name": "Mario", "surname": "Rossi"}
- ("Mario ", "Rossi")
- {" Mario ", " Rossi "}

Prima prova scritta – Traccia B

Determinazione INVALSI n. 185 del 31/08/2022 Concorso pubblico, per titoli ed esami, per l'assunzione a tempo pieno e indeterminato presso l'INVALSI di n. 1 (una) unità di personale nel profilo di Funzionario, V livello professionale di cui al D.P.R. n. 171/1991

25. Quale delle seguenti affermazioni è vera per una lista (LIST) di Python?
- Tutti gli elementi devono essere dello stesso tipo
 - Un dato elemento può essere presente nella lista più volte
 - Una lista può contenere qualsiasi tipo di oggetto eccetto un'altra lista
 - ["a","b","c"] e ["c","b","a"] rappresentano la stessa lista
26. Quale comando di SQL puoi utilizzare per eliminare tutte le righe in una tabella denominata "Istituti" in cui il valore della colonna "CodiceProvincia" è uguale a "RM"?
- DELETE FROM Istituti WHERE CodiceProvincia <> 'RM'
 - DELETE Istituti WHERE CodiceProvincia <> 'RM'
 - DELETE FROM Istituti WHERE CodiceProvincia = 'RM'
 - DELETE Istituti WHERE CodiceProvincia != 'RM'
27. Quale istruzione di SQL può essere utilizzata per aggregare i valori di una colonna e restituire un unico valore per ogni gruppo di righe di una tabella?
- GROUP
 - AGGREGATE
 - AGGREGATE BY
 - GROUP BY
28. Quale operatore si può utilizzare in SQL per selezionare i record che hanno valori per una determinata colonna compresi all'interno di un intervallo specificato?
- BETWEEN
 - LIKE
 - RANGE
 - WITHIN
29. Quale comando puoi utilizzare per selezionare tutti i record da una tabella denominata "Istituti" che hanno valore della colonna "CodiceProvincia" uguale a "RM" oppure uguale a "FR"?
- SELECT * from Istituti WHERE CodiceProvincia = 'RM' | CodiceProvincia = 'FR'
 - SELECT [all] from Istituti WHERE CodiceProvincia = 'RM' OR CodiceProvincia = 'FR'
 - SELECT * from Istituti WHERE CodiceProvincia = 'RM' OR CodiceProvincia = 'FR'
 - SELECT [all] from Istituti WHERE CodiceProvincia = 'RM' | CodiceProvincia = 'FR'

Prima prova scritta – Traccia B

Determinazione INVALSI n. 185 del 31/08/2022 Concorso pubblico, per titoli ed esami, per l'assunzione a tempo pieno e indeterminato presso l'INVALSI di n. 1 (una) unità di personale nel profilo di Funzionario, V livello professionale di cui al D.P.R. n. 171/1991

30. Quale comando puoi utilizzare per inserire "RMIC00000" come "CodiceMeccanografico" in una tabella denominata "Istituti" che non contiene chiavi primarie?

- INSERT ('RMIC000000') INTO Istituti (CodiceMeccanografico)
- INSERT INTO CodiceMeccanografico ('RMIC000000') INTO Istituti
- INSERT INTO Istituti (CodiceMeccanografico) VALUES ('RMIC000000')
- INSERT VALUES Istituti ('RMIC000000') INTO CodiceMeccanografico

Prima prova scritta – Traccia C

Determinazione INVALSI n. 185 del 31/08/2022 Concorso pubblico, per titoli ed esami, per l'assunzione a tempo pieno e indeterminato presso l'INVALSI di n. 1 (una) unità di personale nel profilo di Funzionario, V livello professionale di cui al D.P.R. n. 171/1991

1. Nell'ambito del text mining quale tecnica di pre-processing permette di convertire la parola 'andiamo' in 'andare'?
 - stemming
 - tokenizzazione
 - splitting
 - lemmatizzazione

2. Quale delle seguenti affermazioni sulla metrica IDF (Inverse Document Frequency) è vera?
 - Viene utilizzata per calcolare la frequenza relativa di un termine in un documento
 - Viene utilizzata per calcolare la frequenza relativa di un termine in un intero corpus (insieme di documenti)
 - Viene utilizzata per calcolare la frequenza assoluta di un termine in un documento
 - Premia le parole comuni che compaiono spesso in un corpus (insieme di documenti)

3. Nell'ambito del text mining in cosa consiste la tokenizzazione?
 - Rimuovere prefissi e suffissi dai termini dei documenti
 - Permette di arrivare alle radici dei termini nei documenti
 - Suddividere le parole in blocchi di testo atomico
 - Rimuovere dai documenti tutti i termini considerati irrilevanti ai fini dell'analisi

4. Dalla frase "Analisi del testo NLP", quanti bigrammi (2-gram) ci sono?
 - 2
 - 3
 - 4
 - 5

5. Se "X" è una variabile casuale, $\text{Var}(X)$ è la varianza della variabile casuale e "a" è una costante allora:
 - $\text{Var}(aX) = \text{Var}(X)$
 - $\text{Var}(aX) = a^2\text{Var}(X)$
 - $\text{Var}(aX) = 0$
 - $\text{Var}(aX) = a\text{Var}(X)$



Prima prova scritta – Traccia C

Determinazione INVALSI n. 185 del 31/08/2022 Concorso pubblico, per titoli ed esami, per l'assunzione a tempo pieno e indeterminato presso l'INVALSI di n. 1 (una) unità di personale nel profilo di Funzionario, V livello professionale di cui al D.P.R. n. 171/1991

6. Cosa rappresenta il p-value nei test d'ipotesi?
- la probabilità di ottenere una statistica test con un valore più estremo rispetto a quello della statistica test osservata
 - la probabilità di rifiutare l'ipotesi nulla quando è vera
 - la probabilità di non rifiutare l'ipotesi nulla quando è falsa
 - Il complemento a 1 dell'ipotesi di Prima Specie (o di Tipo I)
7. Quale di queste affermazioni sulla funzione di verosimiglianza è falsa?
- Il punto di massimo della funzione è detto stima di massima verosimiglianza
 - La funzione di verosimiglianza non può essere utilizzata per stimare il valore del parametro incognito
 - Può non esistere un unico punto di massimo della funzione
 - Può essere usata per confrontare il peso del supporto ricevuto da ipotesi diverse
8. Quale tra queste affermazioni relative agli intervalli di confidenza è falsa?
- L'ampiezza dell'intervallo aumenta quando aumenta il livello di fiducia alfa
 - L'ampiezza dell'intervallo aumenta quando diminuisce il valore della varianza campionaria
 - L'ampiezza dell'intervallo non dipende dal livello di fiducia alfa
 - L'ampiezza dell'intervallo diminuisce quando aumenta la numerosità campionaria
9. Un intervallo di confidenza al 99% per la media della popolazione è stimato da 15.48 a 20.52. Se ora viene stimato un intervallo di confidenza al 95% per la media della popolazione sarà?
- Più ampio di quello al 90%
 - Lo stesso dell'intervallo al 90%
 - Più stretto di quello al 90%
 - Non ci sono abbastanza informazioni per rispondere
10. Per quale motivo generalmente si utilizzano i metodi di clustering?
- Individuare costrutti latenti
 - Effettuare previsioni su una variabile numerica
 - Raggruppare le unità statistiche in insiemi di dati
 - Individuare la presenza di valori anomali

Prima prova scritta – Traccia C

Determinazione INVALSI n. 185 del 31/08/2022 Concorso pubblico, per titoli ed esami, per l'assunzione a tempo pieno e indeterminato presso l'INVALSI di n. 1 (una) unità di personale nel profilo di Funzionario, V livello professionale di cui al D.P.R. n. 171/1991

11. Nell'ambito della cluster analysis in cosa consiste il metodo del legame completo?
- Calcolare tutte le dissimilarità tra le osservazioni tra due cluster e registrare la più grande tra queste differenze
 - Calcolare tutte le dissimilarità tra le osservazioni tra due cluster e registrare la più piccola tra queste differenze
 - Calcolare tutte le dissimilarità tra le osservazioni tra due cluster e registrare la media tra queste differenze
 - Calcolare la dissimilarità tra il centroide di due cluster
12. In base a quale motivo si sceglie di ricorrere ad un albero di classificazione oppure ad un albero di regressione?
- Al numero di unità nel dataset
 - Alla dimensione dell'albero che si vuole ottenere
 - Alla natura dei predittori (quantitativa o qualitativa)
 - Alla natura della variabile risposta o outcome (quantitativa o qualitativa)
13. Quale tra le seguenti affermazioni sul Support Vector Machine è vera?
- Il SVM è un modello di regressione
 - Il SVM cerca di separare le classi utilizzando un iperpiano di separazione
 - Il SVM non prende in considerazione il margine di separazione tra le classi
 - Il SVM non può essere utilizzato per una classificazione lineare
14. Quale delle seguenti misure può essere usata come criterio di divisione binaria ricorsiva per gli alberi di classificazione?
- VIF (Variance Inflation Factor)
 - Indice di entropia
 - Indice di dissimilarità
 - MSE (Mean Squared Error)

Prima prova scritta – Traccia C

Determinazione INVALSI n. 185 del 31/08/2022 Concorso pubblico, per titoli ed esami, per l'assunzione a tempo pieno e indeterminato presso l'INVALSI di n. 1 (una) unità di personale nel profilo di Funzionario, V livello professionale di cui al D.P.R. n. 171/1991

15. Che cos'è il metodo bootstrap?

- Una tecnica statistica di ricampionamento con reimmissione
- Una tecnica statistica di ricampionamento senza reimmissione
- Una tecnica di convalida incrociata che consiste nella suddivisione dell'insieme di dati totale in k parti di uguale numerosità
- Una tecnica che permette di rimuovere valori anomali da un dataset

16. Quali valori sono riportati sull'asse delle ascisse della Curva di ROC?

- La Sensibilità
- La Specificità
- Il complemento a 1 della Sensibilità
- Il complemento a 1 della Specificità

17. Cosa misura l'accuratezza (accuracy) per un modello di machine learning?

- Il rapporto tra Veri Positivi (TP) e la somma dei Veri positivi (TP) e Falsi Positivi (FP). Ossia: $TP / (TP + FP)$
- Il rapporto tra Veri Positivi (TP) e la somma dei Veri positivi (TP) e Falsi Negativi (FN). Ossia: $TP / (TP + FN)$
- Il rapporto tra la somma dei Veri Positivi (TP) e Veri Negativi (TN) e il totale delle osservazioni. Ossia: $(TP + TN) / (TP + TN + FP + FN)$
- Il rapporto tra Veri Positivi (TP) e Falsi Positivi (FP). Ossia: TP / FP

18. Quale funzione presente nella libreria Pandas di Python permette di sostituire i valori nulli con un valore specificato dall'utente in un DataFrame?

- fillna
- notna
- nareplace
- insert

Prima prova scritta – Traccia C

Determinazione INVALSI n. 185 del 31/08/2022 Concorso pubblico, per titoli ed esami, per l'assunzione a tempo pieno e indeterminato presso l'INVALSI di n. 1 (una) unità di personale nel profilo di Funzionario, V livello professionale di cui al D.P.R. n. 171/1991

19. Quale funzione presente nella libreria Pandas di Python permette di individuare righe duplicate?

- duplicate()
- duplicated()
- dup()
- isduplicate()

20. Quale dei seguenti NON è un nome valido per una variabile in Python?

- my_var
- _myvar
- my-var
- myvar1

21. Quale metodo in Python permette di convertire tutti i caratteri di una stringa in minuscolo?

- capitalize()
- lower()
- lowercase()
- tolower()

22. Quale funzione presente nella libreria Pandas di Python permette di ottenere i valori delle ultime n righe di un DataFrame?

- last()
- foot()
- back()
- tail()

23. La variabile stringa my_string contiene il seguente testo: "a b c d". Quale dei seguenti risultati viene restituito dalla ricerca: my_string.find("b")?

- 1
- 2
- 3
- 4

Prima prova scritta – Traccia C

Determinazione INVALSI n. 185 del 31/08/2022 Concorso pubblico, per titoli ed esami, per l'assunzione a tempo pieno e indeterminato presso l'INVALSI di n. 1 (una) unità di personale nel profilo di Funzionario, V livello professionale di cui al D.P.R. n. 171/1991

24. Quale di queste collezioni di dati definisce un oggetto DICTIONARY di Python?
- ["Mario", "Rossi"]
 - {"name": "Mario", "surname": "Rossi"}
 - ("Mario ", "Rossi")
 - {" Mario ", " Rossi "}
25. Data la seguente lista di Python: a = [1,5,3]. Quale dei seguenti comandi rimuove l'elemento 5 dalla lista?
- a.del(5)
 - a.delete(5)
 - a.remove(5)
 - a.rm(5)
26. Quale comando di SQL si può utilizzare per ottenere solo i valori distinti?
- SELECT UNIQUE
 - SELECT DIFFERENT
 - SELECT DISTINCT
 - SELECT DUPLICATE
27. Quale comando puoi utilizzare per selezionare tutti i record da una tabella denominata "Istituti" che hanno valore della colonna "CodiceProvincia" uguale a "RM"?
- SELECT * FROM Istituti WHERE CodiceProvincia <> 'RM'
 - SELECT * FROM Istituti WHERE CodiceProvincia != 'RM'
 - SELECT * Istituti WHERE CodiceProvincia == 'RM'
 - SELECT * FROM Istituti WHERE CodiceProvincia = 'RM'
28. Quale comando puoi utilizzare per selezionare tutti i record da una tabella denominata "Istituti" ordinati in base alla colonna "CodiceProvincia"?
- SELECT * FROM Istituti ORDER CodiceProvincia
 - SELECT * FROM Istituti SORT CodiceProvincia
 - SELECT * FROM Istituti ORDER BY CodiceProvincia
 - SELECT * FROM Istituti SORT BY CodiceProvincia

Prima prova scritta – Traccia C

Determinazione INVALSI n. 185 del 31/08/2022 Concorso pubblico, per titoli ed esami, per l'assunzione a tempo pieno e indeterminato presso l'INVALSI di n. 1 (una) unità di personale nel profilo di Funzionario, V livello professionale di cui al D.P.R. n. 171/1991

29. Quale comando puoi utilizzare per inserire "RM" come "CodiceProvincia" in una tabella denominata "Istituti" che non contiene chiavi primarie?

- INSERT ('RM') INTO Istituti (CodiceProvincia)
- INSERT INTO CodiceProvincia ('RM') INTO Istituti
- INSERT INTO Istituti (CodiceProvincia) VALUES ('RM')
- INSERT VALUES Istituti ('RM') INTO CodiceProvincia

30. Quale comando puoi utilizzare per creare una tabella denominata "Istituti" in un database SQL?

- CREATE TABLE Istituti
- CREATE DATABASE TABLE Istituti
- CREATE DATABASE TAB Istituti
- CREATE DB Istituti