

Immagina che tu sia il manager di un'azienda che ha appena lanciato un nuovo prodotto. Il tuo obiettivo è aumentare le vendite di questo prodotto. Per farlo, hai deciso di lanciare una campagna di marketing. La campagna consiste in una serie di attività che conducono gli utenti a un risultato utile per rispondere all'obiettivo posto.

Periodo	Attività	Risultato
1. 15.06.2023	1. 15.06.2023	1. 15.06.2023
2. 15.06.2023	2. 15.06.2023	2. 15.06.2023
3. 15.06.2023	3. 15.06.2023	3. 15.06.2023
4. 15.06.2023	4. 15.06.2023	4. 15.06.2023
5. 15.06.2023	5. 15.06.2023	5. 15.06.2023
6. 15.06.2023	6. 15.06.2023	6. 15.06.2023
7. 15.06.2023	7. 15.06.2023	7. 15.06.2023
8. 15.06.2023	8. 15.06.2023	8. 15.06.2023
9. 15.06.2023	9. 15.06.2023	9. 15.06.2023
10. 15.06.2023	10. 15.06.2023	10. 15.06.2023
11. 15.06.2023	11. 15.06.2023	11. 15.06.2023
12. 15.06.2023	12. 15.06.2023	12. 15.06.2023
13. 15.06.2023	13. 15.06.2023	13. 15.06.2023
14. 15.06.2023	14. 15.06.2023	14. 15.06.2023
15. 15.06.2023	15. 15.06.2023	15. 15.06.2023
16. 15.06.2023	16. 15.06.2023	16. 15.06.2023
17. 15.06.2023	17. 15.06.2023	17. 15.06.2023
18. 15.06.2023	18. 15.06.2023	18. 15.06.2023
19. 15.06.2023	19. 15.06.2023	19. 15.06.2023
20. 15.06.2023	20. 15.06.2023	20. 15.06.2023
21. 15.06.2023	21. 15.06.2023	21. 15.06.2023
22. 15.06.2023	22. 15.06.2023	22. 15.06.2023
23. 15.06.2023	23. 15.06.2023	23. 15.06.2023
24. 15.06.2023	24. 15.06.2023	24. 15.06.2023
25. 15.06.2023	25. 15.06.2023	25. 15.06.2023
26. 15.06.2023	26. 15.06.2023	26. 15.06.2023
27. 15.06.2023	27. 15.06.2023	27. 15.06.2023
28. 15.06.2023	28. 15.06.2023	28. 15.06.2023
29. 15.06.2023	29. 15.06.2023	29. 15.06.2023
30. 15.06.2023	30. 15.06.2023	30. 15.06.2023
31. 15.06.2023	31. 15.06.2023	31. 15.06.2023
32. 15.06.2023	32. 15.06.2023	32. 15.06.2023
33. 15.06.2023	33. 15.06.2023	33. 15.06.2023
34. 15.06.2023	34. 15.06.2023	34. 15.06.2023
35. 15.06.2023	35. 15.06.2023	35. 15.06.2023
36. 15.06.2023	36. 15.06.2023	36. 15.06.2023
37. 15.06.2023	37. 15.06.2023	37. 15.06.2023
38. 15.06.2023	38. 15.06.2023	38. 15.06.2023
39. 15.06.2023	39. 15.06.2023	39. 15.06.2023
40. 15.06.2023	40. 15.06.2023	40. 15.06.2023
41. 15.06.2023	41. 15.06.2023	41. 15.06.2023
42. 15.06.2023	42. 15.06.2023	42. 15.06.2023
43. 15.06.2023	43. 15.06.2023	43. 15.06.2023
44. 15.06.2023	44. 15.06.2023	44. 15.06.2023
45. 15.06.2023	45. 15.06.2023	45. 15.06.2023
46. 15.06.2023	46. 15.06.2023	46. 15.06.2023
47. 15.06.2023	47. 15.06.2023	47. 15.06.2023
48. 15.06.2023	48. 15.06.2023	48. 15.06.2023
49. 15.06.2023	49. 15.06.2023	49. 15.06.2023
50. 15.06.2023	50. 15.06.2023	50. 15.06.2023
51. 15.06.2023	51. 15.06.2023	51. 15.06.2023
52. 15.06.2023	52. 15.06.2023	52. 15.06.2023
53. 15.06.2023	53. 15.06.2023	53. 15.06.2023
54. 15.06.2023	54. 15.06.2023	54. 15.06.2023
55. 15.06.2023	55. 15.06.2023	55. 15.06.2023
56. 15.06.2023	56. 15.06.2023	56. 15.06.2023
57. 15.06.2023	57. 15.06.2023	57. 15.06.2023
58. 15.06.2023	58. 15.06.2023	58. 15.06.2023
59. 15.06.2023	59. 15.06.2023	59. 15.06.2023
60. 15.06.2023	60. 15.06.2023	60. 15.06.2023
61. 15.06.2023	61. 15.06.2023	61. 15.06.2023
62. 15.06.2023	62. 15.06.2023	62. 15.06.2023
63. 15.06.2023	63. 15.06.2023	63. 15.06.2023
64. 15.06.2023	64. 15.06.2023	64. 15.06.2023
65. 15.06.2023	65. 15.06.2023	65. 15.06.2023
66. 15.06.2023	66. 15.06.2023	66. 15.06.2023
67. 15.06.2023	67. 15.06.2023	67. 15.06.2023
68. 15.06.2023	68. 15.06.2023	68. 15.06.2023
69. 15.06.2023	69. 15.06.2023	69. 15.06.2023
70. 15.06.2023	70. 15.06.2023	70. 15.06.2023
71. 15.06.2023	71. 15.06.2023	71. 15.06.2023
72. 15.06.2023	72. 15.06.2023	72. 15.06.2023
73. 15.06.2023	73. 15.06.2023	73. 15.06.2023
74. 15.06.2023	74. 15.06.2023	74. 15.06.2023
75. 15.06.2023	75. 15.06.2023	75. 15.06.2023
76. 15.06.2023	76. 15.06.2023	76. 15.06.2023
77. 15.06.2023	77. 15.06.2023	77. 15.06.2023
78. 15.06.2023	78. 15.06.2023	78. 15.06.2023
79. 15.06.2023	79. 15.06.2023	79. 15.06.2023
80. 15.06.2023	80. 15.06.2023	80. 15.06.2023
81. 15.06.2023	81. 15.06.2023	81. 15.06.2023
82. 15.06.2023	82. 15.06.2023	82. 15.06.2023
83. 15.06.2023	83. 15.06.2023	83. 15.06.2023
84. 15.06.2023	84. 15.06.2023	84. 15.06.2023
85. 15.06.2023	85. 15.06.2023	85. 15.06.2023
86. 15.06.2023	86. 15.06.2023	86. 15.06.2023
87. 15.06.2023	87. 15.06.2023	87. 15.06.2023
88. 15.06.2023	88. 15.06.2023	88. 15.06.2023
89. 15.06.2023	89. 15.06.2023	89. 15.06.2023
90. 15.06.2023	90. 15.06.2023	90. 15.06.2023
91. 15.06.2023	91. 15.06.2023	91. 15.06.2023
92. 15.06.2023	92. 15.06.2023	92. 15.06.2023
93. 15.06.2023	93. 15.06.2023	93. 15.06.2023
94. 15.06.2023	94. 15.06.2023	94. 15.06.2023
95. 15.06.2023	95. 15.06.2023	95. 15.06.2023
96. 15.06.2023	96. 15.06.2023	96. 15.06.2023
97. 15.06.2023	97. 15.06.2023	97. 15.06.2023
98. 15.06.2023	98. 15.06.2023	98. 15.06.2023
99. 15.06.2023	99. 15.06.2023	99. 15.06.2023
100. 15.06.2023	100. 15.06.2023	100. 15.06.2023

Immagina che ti venga richiesto di profilare gli utenti che effettuano una spesa alta per accedere (superiore ai 10 euro). Descrivi tutte le attività che conducessi per arrivare ad un risultato utile per rispondere all'obiettivo posto.

Immagina che ti venga richiesto di individuare quali sono i due prodotti da inserire nella prossima promozione in home page. Descrivi tutte le attività che conducessi per arrivare ad un risultato utile per rispondere all'obiettivo posto.

Appello di MODELLI DI E-BUSINESS E BUSINESS INTELLIGENCE

Client ID	Product ID	Session ID	Date	Number of products	Amount	Payment type
1	1	1	15.06.2023	7	10	Credit card
1	2	1	15.06.2023	2	5	
1	3	2	20.06.2023	6	2	Cash
1	4	2	20.06.2023	7	1	
2	2	1	15.06.2023	1	5	Cash
2	3	1	15.06.2023	9	2	
2	4	2	20.06.2023	8	1	Credit card
2	5	2	20.06.2023	8	0.5	
3	1	1	15.06.2023	7		Cash
3	4	1	15.06.2023	2	1	
3	5	1	15.06.2023	9	0.5	
3	6	2	20.06.2023	10		Credit card
3	3	2	20.06.2023	9	2	
4	1	1	15.06.2023	10		Credit card
4	2	1	15.06.2023	4	5	
4	3	1	15.06.2023	9	2	
4	4	2	20.06.2023	2	1	Cash
4	6	2	20.06.2023	4		
5	1	1	15.06.2023	1		Credit card

La attività di pre-processing è in difficoltà
 1) cleaning, 2) transformation, 3) reduction
 Nell'attività di cleaning si rimuovono i dati non necessari o fuori dall'analisi,
 nell' attività

13. Si considerino i risultati di un modello di classificazione addestrato sui risultati di un questionario in cui, oltre alcune informazioni sull'intervistato e sul servizio utilizzato, viene richiesto di fornire un giudizio sul servizio distinguendo tra un servizio non raccomandabile, raccomandabile, molto raccomandabile, prioritario oppure altamente prioritario. Se l'obiettivo è classificare e distinguere tra un servizio non accettabile e uno accettabile (senza ulteriori distinzioni di merito), il risultato ottenuto si può considerare come buono? Se sì, perché. Se no, perché e cosa faresti per migliorare.

```

Number of Leaves : 301
Size of the tree : 429

Time taken to build model: 0.07 seconds

--- Stratified cross-validation ---
--- Summary ---

Correctly Classified Instances      9490      96.89 %
Incorrectly Classified Instances    311        3.11 %
Kappa statistic                    0.9533
Mean absolute error                 0.0153
Root mean squared error             0.0875
Relative absolute error              1.0982 %
Root relative squared error         26.4503 %
Total Number of Instances          10000

--- Detailed Accuracy By Class ---

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
-----
1.000  0.000  1.000  1.000  1.000  1.000  1.000  1.000  not_recom
0.000  0.000  ?      0.000  ?      ?      0.499  0.000  recomend
0.729  0.007  0.776  0.729  0.756  0.746  0.965  0.717  very_recom
0.488  0.024  0.964  0.488  0.961  0.835  0.965  0.960  priority
0.977  0.014  0.982  0.977  0.987  0.987  0.995  0.992  spec_prior
Weighted Avg.  0.949  0.012  ?      0.949  ?      ?      0.994  0.995

--- Confusion Matrix ---

      a      b      c      d      e  -- classified as
-----
1333  0      0      0      0      0  a = not_recom
0      0      2      0      0      0  b = recomend
0      0  229  39      0      0  c = very_recom
0      0  43 1755 103      0  d = priority
0      0      0  54 1332      0  e = spec_prior

```

14. Si consideri il risultato della seguente regressione addestrata su dati riferiti agli acquisti effettuati dagli utenti su di un sito web. L'obiettivo è prevedere l'importo speso dagli utenti. Dopo una serie di tentativi si ottiene il risultato in basso dove le variabili inserite sono "quantità" (numero di prodotti acquistati), "sconto" (indica se c'è stata l'applicazione di uno sconto). Il risultato ottenuto si può considerare buono? Se sì, perché. Se no, perché e cosa faresti per migliorare.

--- Linear Classification model (with training set) ---			
Linear Regression Model			
Equation =			
$y = 0.0001x + 0.0001x + 0.0001x + 0.0001x$			
Regression statistics:			
	Multiple R	Adjusted R Square	F-Stat
Intercept	0.0001	0.0001	0.0001
Quantità	0.0001	0.0001	0.0001
Sconto	0.0001	0.0001	0.0001
Degrees of Freedom = 100			
R-Square = 0.0001			
Adjusted R-Square = 0.0001			
F-Statistic = 0.0001			
Time taken to build model = 0.0001			
--- Prediction on training set ---			
Time taken to build model on training set = 0.0001			
--- Summary ---			
Correctly Classified Instances	0.0001		
Incorrectly Classified Instances	0.0001		
Mean absolute error	0.0001		
Root mean squared error	0.0001		
Relative absolute error	0.0001 %		
Root relative squared error	0.0001 %		
Total Number of Instances	0.0001		

Name: _____

Surname: _____

Matriculation number: _____

Appello di MODELLI DI E-BUSINESS E BUSINESS INTELLIGENCE

1. Se l'indice di Gini misurato per la variabile "categoria acquistata" è pari a 1, vuol dire che l'utente:
 - a. acquista categorie di prodotto sempre diverse.
 - b. acquista categorie di prodotto sempre uguali.
 - c. acquista una sola categoria di prodotto.
2. Per Immobiliare.it, l'aumento dei tassi di interesse è:
 - a. Forza di mercato.
 - b. Forza industriale.
 - c. Forza macro-economica.
3. Per un campo rappresentativo della zona di residenza in Italia, che assume i valori "nord" e "sud", un valore mancante può essere sostituito con:
 - a. Il valore più frequente.
 - b. Un valore scelto arbitrariamente.
 - c. Il valore medio.
4. Un metodo di apprendimento supervisionato presuppone la presenza di una:
 - a. variabile categorica.
 - b. variabile continua.
 - c. variabile target.
5. Nell'ipotesi di avere un database contenente le rilevazioni metereologiche dell'ultimo anno, per prevedere la temperatura posso utilizzare:
 - a. Regressione.
 - b. Classificazione.
 - c. Regole associative.
6. Se l'indice di Skewness del campo "soddisfazione" (1=non soddisfatto, 10=soddisfatto) è pari a 0,8, vuol dire che:
 - a. È minore la frequenza di utenti non soddisfatti.
 - b. Molti utenti hanno soddisfazione pari a 5.
 - c. È maggiore la frequenza di utenti non soddisfatti.
7. Se prendo in gestione un punto vendita di Starbucks, nel modello di business di Starbucks sono:
 - a. Cliente.
 - b. Revenue stream.
 - c. Partnership.
8. Posso trasformare una variabile continua in categorica?
 - a. Sì, solo se non ci sono valori mancanti.
 - b. No.
 - c. Sì, discretizzando la variabile.
9. Se voglio migliorare l'accuratezza di un albero decisionale posso provare ad:
 - a. Aumentare la dimensione dell'albero.
 - b. Ridurre la dimensione dell'albero.
 - c. Utilizzare il cross-validation.
10. Per verificare se l'età di una popolazione è molto concentrata attorno la media, posso utilizzare:
 - a. Indice di Kurtosis.
 - b. Indice di Skewness.
 - c. Media.
11. Quali sono le value propositions, i customer segments e le revenue streams del sito Immobiliare.it? Quale pattern di modello di business segue? Spiegarne il perché.

1. Se lo skewness index del campo "soddisfazione" (1=non soddisfatto, 10=soddisfatto) è pari a 0,8, vuol dire che:
 - a. È maggiore la frequenza di utenti soddisfatti.
 - ☒ b. La distribuzione è iponormale.
 - c. È maggiore la frequenza di utenti non soddisfatti.
2. Se la coppia di variabili a-b presenta un indice di Pearson pari a 0,8 mentre la coppia c-d presenta un indice di Pearson pari a -0,9, la coppia con la relazione più intensa è:
 - a. a-b
 - b. a-d
 - ☒ c. c-d
3. Per un campo rappresentativo del grado di istruzione che può assumere valori come "scuola superiore", "università" e "dottorato", un valore mancante può essere sostituito con:
 - ☒ a. Il valore più frequente.
 - b. Eliminato.
 - c. Un valore scelto arbitrariamente.
4. Nel modello di business di Spotify, concentrandoci sul ruolo che hanno le case produttrici di musica, si segue un pattern di tipo:
 - a. Free.
 - b. Open.
 - ☒ c. Multi-sided.
5. Una variabile rappresentativa dello stadio di produzione di un componente meccanico, che assume valori da 1 a 10, è di tipo:
 - ☒ a. Discreto.
 - b. Ordinale.
 - c. Continuo.
6. La value proposition dipende da:
 - ☒ a. I segmenti di clienti che voglio raggiungere.
 - b. Le partnership che voglio instaurare.
 - c. Le revenue streams che voglio ottenere.
7. Per individuare il miglior bundle (combinazione) di prodotti da vendere, posso utilizzare:
 - a. Regressione.
 - b. Regole associative.
 - ☒ c. Classificazione.
8. Il modello di Airbnb segue il pattern long-tail:
 - ☒ a. No.
 - b. Dipende dal canale.
 - c. Sì.
9. Per una testata giornalistica, il proliferare di piattaforme di intelligenza artificiale come ChatGPT è:
 - ☒ a. Forza di mercato.
 - b. Trend chiave.
 - c. Forza industriale.
10. Per un produttore di cellulari, i sistemi operativi su cui si poggia sono:
 - a. Clienti.
 - b. Partnership chiave.
 - ☒ c. Risorse chiave.
11. Quali sono le value propositions, i customer segments e le revenue streams di YouTube? Quale pattern di modello di business segue? Spiegare il perché.

VP: possibilità di vedere film in streaming; app scaricabile; easy to use; ottimo rapporto qualità/prezzo; veloce; possibilità di inserire pubblicità; grande quantità di dati

CS: teenager; youtuber; famiglia; inserzionisti (pubblicità)

RS: ricavi derivanti dalla pubblicità; visitatori (tramite app premium scaricate)

YouTube segue il modello di business del "multi-sided platform", in quanto si r
12. Supponiamo di avere il seguente dataset contenente informazioni circa le transazioni svolte in un punto vendita. Conosciamo l'utente che ha svolto l'acquisto, il prodotto acquistato, in quale sessione ha svolto l'acquisto, in che giorno e in quale quantità è stato acquistato il prodotto, quanto è stato speso e come si è scelto di pagare (carta di credito o contante). Descrivere tutte le attività di pre-process necessarie, e il modello di data mining da utilizzare, per poter individuare quali sono i prodotti da mettere in promozione assieme.

segmenti di clienti che sono interdipendenti tra di loro; inoltre solotta un modello "freemium", in quanto offre anche la possibilità di servizi a pagamento (nominati) per la versione di film in streaming.

Nome: _____

Cognome: _____

12. Immagina di avere
agli

Cognome: _____

Matricola: _____

12. Immagina di avere il seguente database contenente una serie di informazioni relative alla navigazione e agli acquisti effettuati su di un sito web. In particolare, viene registrato l'identificativo dell'utente, l'identificativo della pagina visualizzata, l'identificativo della sessione durante la quale l'utente ha visualizzato la pagina (durante ogni sessione possono essere state visualizzate più pagine), la data, il prodotto su cui si è cliccato nella pagina, se il prodotto su cui si è cliccato è stato acquistato, la spesa effettuata dall'utente durante la sessione, l'età e il genere dell'utente.

User ID	Page ID	Session ID	Date	Product ID	Purchased the product	Amount of money spent during the session	Age	Gender
1	1	1	15.06.2023	5	YES			
1	2	1	15.06.2023	6	NO		10	18 M
1	3	2	20.06.2023	10	YES		10	18
2	4	2	20.06.2023	2	NO		2	18 F
2	2	3	15.06.2023	6	NO		2	18
2	3	4	15.06.2023	10	YES		4	20 F
2	4	5	20.06.2023	2	NO		4	20
3	1	1	20.06.2023	3	NO			20
3	4	1	15.06.2023	5	NO			20
3	5	2	15.06.2023	2	YES		15	33 F
3	6	1	15.06.2023	3	NO		15	33
3	3	2	20.06.2023	1	NO		15	33
4	1	2	20.06.2023	10	NO		0	33
4	2	1	15.06.2023	5	YES			33
4	3	1	15.06.2023	6	NO		20	25 M
4	4	2	15.06.2023	10	NO		20	25
4	6	3	20.06.2023	2	YES			25
4	1	4	20.06.2023	1	NO		15	25
5	1	1	15.06.2023	5	YES		8	25 M

Immagina che ti venga richiesto di profilare gli utenti che effettuano una spesa alta per sessione (superiore ai 10 euro). Descrivi tutte le attività che condurresti per arrivare ad un risultato utile per rispondere all'obiettivo posto.

Immagina che ti venga richiesto di individuare quali sono i due prodotti da inserire nella prossima promozione in home page. Descrivi tutte le attività che condurresti per arrivare ad un risultato utile per rispondere all'obiettivo posto.

Nome: _____

Surname: _____

Matriculation number: _____

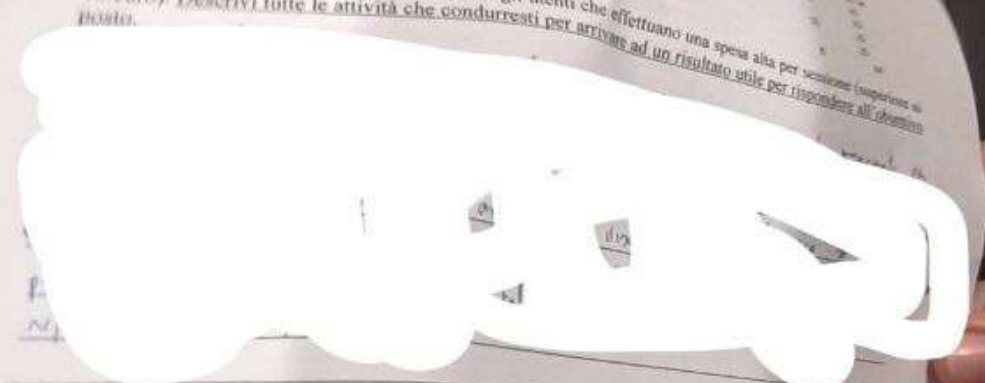
Appello di MODELLI DI E-BUSINESS E BUSINESS INTELLIGENCE

1. Se la coppia di variabili a-b ha una covarianza di 0,8 mentre la coppia c-d ha una covarianza pari a -0,9 vuol dire che:
☐ la coppia c-d ha la relazione più intensa.
☐ la coppia c-d presenta una discordanza.
☐ la coppia a-b ha la relazione più intensa.
2. Per Sony Playstation, l'ingresso di Apple nel mercato del videogaming è:
☐ Forza di mercato.
☐ Forza industriale.
☐ Trend chiave.
3. Per un campo rappresentativo del momento della giornata, che può assumere i valori "giorno" e "notte", un valore mancante può essere sostituito con:
☐ Il valore più frequente.
☐ Un valore scelto arbitrariamente.
☐ Il valore medio.
4. La variabile rappresentativa della risposta alla domanda di un questionario (su scala da 1 a 6) è di tipo:
☐ Continuo.
☐ Nominale.
☐ Ordinale.
5. Nell'ipotesi di avere un database contenente gli esiti delle campagne promozionali precedenti, per identificare gli utenti a cui inviare uno sconto del 10%, 15% oppure del 20%, posso utilizzare:
☐ Regressione.
☐ Classificazione.
☐ Regole associative.
6. Se l'indice di Kurtosis del campo "soddisfazione" (1=non soddisfatto, 10=soddisfatto) è pari a 0,8, vuol dire che:
☐ È minore la frequenza di utenti non soddisfatti.
☐ Molti utenti hanno soddisfazione pari a 5.
☐ È maggiore la frequenza di utenti non soddisfatti.
7. La value proposition principale di Netflix è:
☐ Guadagnare dagli abbonamenti.
☐ Ampio catalogo.
☐ Permettere di visualizzare contenuti in streaming.
8. Se ho una variabile target categorica, posso utilizzare una regressione?
☐ Sì, normalizzando la variabile.
☐ No.
☐ Sì, discretizzando la variabile.
9. Se riduco il numero di regole di un albero decisionale posso aspettarmi:
☐ Un aumento dell'accuratezza dell'albero.
☐ Una riduzione dell'accuratezza dell'albero.
☐ Un peggioramento della leggibilità dell'albero.
10. Per verificare se l'età di una popolazione è prevalentemente bassa o alta posso utilizzare:
☐ Indice di Kurtosis.
☐ Indice di Skewness.
☐ Media.
11. Quali sono le value propositions, i customer segments e le revenue streams del sito del Corriere della Sera? Quale pattern di modello di business segue? Spiegarne il perché.

Immagina di essere il seguente responsabile di un sito e-commerce. Il tuo sito è un sito di vendita di prodotti di lusso. Il tuo obiettivo è aumentare la conversione della pagina di checkout. Il tuo compito è di cliccare sulla pagina di checkout e di cliccare sulla pagina di checkout.

Session ID	Date	Conversion Rate
1	15.06.2023	5 YES
2	16.06.2023	5 YES
3	20.06.2023	5 YES
4	20.06.2023	5 YES
5	15.06.2023	5 YES
6	15.06.2023	5 YES
7	15.06.2023	5 YES
8	15.06.2023	5 YES
9	15.06.2023	5 YES
10	15.06.2023	5 YES
11	15.06.2023	5 YES
12	15.06.2023	5 YES
13	15.06.2023	5 YES
14	15.06.2023	5 YES
15	15.06.2023	5 YES
16	15.06.2023	5 YES
17	15.06.2023	5 YES
18	15.06.2023	5 YES
19	15.06.2023	5 YES
20	15.06.2023	5 YES
21	15.06.2023	5 YES
22	15.06.2023	5 YES
23	15.06.2023	5 YES
24	15.06.2023	5 YES
25	15.06.2023	5 YES
26	15.06.2023	5 YES
27	15.06.2023	5 YES
28	15.06.2023	5 YES
29	15.06.2023	5 YES
30	15.06.2023	5 YES
31	15.06.2023	5 YES
32	15.06.2023	5 YES
33	15.06.2023	5 YES
34	15.06.2023	5 YES
35	15.06.2023	5 YES
36	15.06.2023	5 YES
37	15.06.2023	5 YES
38	15.06.2023	5 YES
39	15.06.2023	5 YES
40	15.06.2023	5 YES
41	15.06.2023	5 YES
42	15.06.2023	5 YES
43	15.06.2023	5 YES
44	15.06.2023	5 YES
45	15.06.2023	5 YES
46	15.06.2023	5 YES
47	15.06.2023	5 YES
48	15.06.2023	5 YES
49	15.06.2023	5 YES
50	15.06.2023	5 YES

Immagina che ti venga richiesto di profilare gli utenti che effettuano una spesa alta per sessione (superiore ai 10 euro). Descrivi tutte le attività che condurresti per arrivare ad un risultato utile per rispondere all'obiettivo posto.



Immagina che ti venga richiesto di individuare quali sono i due prodotti da inserire nella prossima promozione in home page. Descrivi tutte le attività che condurresti per arrivare ad un risultato utile per rispondere all'obiettivo posto.

Handwritten text in the second section of the form, including the words "prodotti" and "promozione".

Nome: _____

Surname: _____

Matriculation number: _____

Appello di MODELLI DI E-BUSINESS E BUSINESS INTELLIGENCE

1. Se la coppia di variabili a-b ha una covarianza di 0,8 mentre la coppia c-d ha una covarianza pari a -0,9 vuol dire che:
☐ la coppia c-d ha la relazione più intensa.
☐ la coppia c-d presenta una discordanza.
☐ la coppia a-b ha la relazione più intensa.
2. Per Sony Playstation, l'ingresso di Apple nel mercato del videogaming è:
☐ Forza di mercato.
☐ Forza industriale.
☐ Trend chiave.
3. Per un campo rappresentativo del momento della giornata, che può assumere i valori "giorno" e "notte", un valore mancante può essere sostituito con:
☐ Il valore più frequente.
☐ Un valore scelto arbitrariamente.
☐ Il valore medio.
4. La variabile rappresentativa della risposta alla domanda di un questionario (su scala da 1 a 6) è di tipo:
☐ Continuo.
☐ Nominale.
☐ Ordinale.
5. Nell'ipotesi di avere un database contenente gli esiti delle campagne promozionali precedenti, per identificare gli utenti a cui inviare uno sconto del 10%, 15% oppure del 20%, posso utilizzare:
☐ Regressione.
☐ Classificazione.
☐ Regole associative.
6. Se l'indice di Kurtosis del campo "soddisfazione" (1=non soddisfatto, 10=soddisfatto) è pari a 0,8, vuol dire che:
☐ È minore la frequenza di utenti non soddisfatti.
☐ Molti utenti hanno soddisfazione pari a 5.
☐ È maggiore la frequenza di utenti non soddisfatti.
7. La value proposition principale di Netflix è:
☐ Guadagnare dagli abbonamenti.
☐ Ampio catalogo.
☐ Permettere di visualizzare contenuti in streaming.
8. Se ho una variabile target categorica, posso utilizzare una regressione?
☐ Sì, normalizzando la variabile.
☐ No.
☐ Sì, discretizzando la variabile.
9. Se riduco il numero di regole di un albero decisionale posso aspettarmi:
☐ Un aumento dell'accuratezza dell'albero.
☐ Una riduzione dell'accuratezza dell'albero.
☐ Un peggioramento della leggibilità dell'albero.
10. Per verificare se l'età di una popolazione è prevalentemente bassa o alta posso utilizzare:
☐ Indice di Kurtosis.
☐ Indice di Skewness.
☐ Media.
11. Quali sono le value propositions, i customer segments e le revenue streams del sito del Corriere della Sera? Quale pattern di modello di business segue? Spiegarne il perché.

Nome: _____

Cognome:

Matricola:

12. Immagina di avere il seguente database contenente una serie di informazioni relative alla navigazione e agli acquisti effettuati su di un sito web. In particolare, viene registrato l'identificativo dell'utente, l'identificativo della pagina visualizzata, l'identificativo della sessione durante la quale l'utente ha visualizzato la pagina (durante ogni sessione possono essere state visualizzate più pagine), la data, il prodotto su cui si è cliccato nella pagina, se il prodotto su cui si è cliccato è stato acquistato, la spesa effettuata dall'utente durante la sessione, l'età e il genere dell'utente.

User ID	Page ID	Session ID	Date	Product ID	Purchased the product	Amount of money spent during the session	Age	Gender
1	1	1	15.06.2023	5	YES			
1	2	2	15.06.2023	6	NO			
1	3	3	20.06.2023	10	YES		10	18 M
2	4	2	20.06.2023	2	NO		10	18
2	2	3	15.06.2023	6	NO		2	18 F
2	3	4	15.06.2023	10	YES		2	18
2	4	5	20.06.2023	2	NO		4	20 F
3	1	1	20.06.2023	3	NO		4	20
3	4	1	15.06.2023	5	NO			20
3	5	4	15.06.2023	2	YES		15	33 F
3	6	5	15.06.2023	3	NO		15	33
3	3	6	20.06.2023	1	NO		15	33
4	1	1	20.06.2023	10	NO		0	33
4	2	2	15.06.2023	5	YES			33
4	3	3	15.06.2023	6	NO		20	25 M
4	4	4	15.06.2023	10	NO		20	25
4	6	6	20.06.2023	2	YES		15	25
5	1	1	20.06.2023	1	NO			25
		1	15.06.2023	5	YES		8	M

Immagina che ti venga richiesto di profilare gli utenti che effettuano una spesa alta per sessione (superiore ai 10 euro). Descrivi tutte le attività che condurresti per arrivare ad un risultato utile per rispondere all'obiettivo posto.

Immagina che ti venga richiesto di individuare quali sono i due prodotti da inserire nella prossima promozione in home page. Descrivi tutte le attività che condurresti per arrivare ad un risultato utile per rispondere all'obiettivo posto.

Nome: _____

Cognome: _____

Matricola: _____

12. Ipotizza di avere un franchising composto da 5 punti vendita e di voler profilare il comportamento dei clienti di ciascuno di questi per poi impostare una strategia commerciale opportuna per far crescere tutti i punti vendita. Ipotizza di possedere una serie di informazioni per ciascun cliente (ad esempio, età, genere, spesa, modalità di pagamento, etc.) e in più una variabile target che indica in quale punto vendita acquista (ad esempio, punto vendita 1, punto vendita 2, etc.). Lanciando un classificatore avente variabile target il punto vendita, quale sarebbe il primo risultato (non tutti i risultati) che controlleresti? Perché?

13. Immagina di avere un data base storico delle transazioni bancarie e che, per ogni bonifico, conosci se è stato fraudolento o meno. Quale sarebbe la matrice di confusione nel caso in cui volessi prevedere gli utenti che effettuano un bonifico fraudolento? Fornisci anche indicazione in merito ai costi associati a ciascun errore della matrice così definita.

14. Considera di avere il seguente risultato per un modello di regressione.

```

FATT_CLIENTE =
-32.8032 * TRANSAZIONI_PER_CLIENTE +
5.1552 * FATT_PER_TRANS +
31.1331 * TOT_COSTI +
-47.8107 * PRGO_MEDIO_PER_TRANSAZIONE +
-1258.0722 * CATEGORIA_FUT_ACOUITATA=ABBONAMENTO, SCARPE +
3404.0641 * CATEGORIA_FUT_ACOUITATA=SCARPE +
344.133

Regression Analysis:
Variable          Coefficient    SE of Coef    t-Stat
TRANSAZIONI_PER_CLIENTE    -32.8032      3.6816      -8.91
FATT_PER_TRANS              5.1552      0.8996       5.73
TOT_COSTI                   31.1331      0.5402      57.6279
PRGO_MEDIO_PER_TRANSAZIONE -47.8107      9.7451      -4.9061
CATEGORIA_FUT_ACOUITATA=ABBONAMENTO, SCARPE -1258.0722    772.1749     -1.6283
CATEGORIA_FUT_ACOUITATA=SCARPE  3404.0641    423.0477      8.0459
CONST                     344.133      695.0747      0.4951

Degrees of Freedom = 614
R^2 value = 0.9497
Adjusted R^2 = 0.94917
F-statistic = 1850.7311

Time taken to build model: 0 seconds

--- Cross-validation ---
--- Summary ---
Correlation coefficient      0.9471
Mean absolute error         3014.6211
Mean squared error          7832.2204
RM absolute error           21.0281
Relative squared error      28.4907
Number of instances         621

```

Considereresti un buon risultato o un cattivo risultato? Perché?

Name: _____

Surname: _____

Matriculation number: _____

Appello di MODELLI DI E-BUSINESS E BUSINESS INTELLIGENCE

1. Se l'indice di Gini misurato per la variabile "categoria acquistata" è pari a 1, vuol dire che l'utente:
 - a. acquista categorie di prodotto sempre diverse.
 - b. acquista categorie di prodotto sempre uguali.
 - c. acquista una sola categoria di prodotto.
2. Per Immobiliare.it, l'aumento dei tassi di interesse è:
 - a. Forza di mercato.
 - b. Forza industriale.
 - c. Forza macro-economica.
3. Per un campo rappresentativo della zona di residenza in Italia, che assume i valori "nord" e "sud", un valore mancante può essere sostituito con:
 - a. Il valore più frequente.
 - b. Un valore scelto arbitrariamente.
 - c. Il valore medio.
4. Un metodo di apprendimento supervisionato presuppone la presenza di una:
 - a. variabile categorica.
 - b. variabile continua.
 - c. variabile target.
5. Nell'ipotesi di avere un database contenente le rilevazioni meteorologiche dell'ultimo anno, per prevedere la temperatura posso utilizzare:
 - a. Regressione.
 - b. Classificazione.
 - c. Regole associative.
6. Se l'indice di Skewness del campo "soddisfazione" (1=non soddisfatto, 10=soddisfatto) è pari a 0,8, vuol dire che:
 - a. È minore la frequenza di utenti non soddisfatti.
 - b. Molti utenti hanno soddisfazione pari a 5.
 - c. È maggiore la frequenza di utenti non soddisfatti.
7. Se prendo in gestione un punto vendita di Starbucks, nel modello di business di Starbucks sono:
 - a. Cliente.
 - b. Revenue stream.
 - c. Partnership.
8. Posso trasformare una variabile continua in categorica?
 - a. Sì, solo se non ci sono valori mancanti.
 - b. No.
 - c. Sì, discretizzando la variabile.
9. Se voglio migliorare l'accuratezza di un albero decisionale posso provare ad:
 - a. Aumentare la dimensione dell'albero.
 - b. Ridurre la dimensione dell'albero.
 - c. Utilizzare il cross-validation.
10. Per verificare se l'età di una popolazione è molto concentrata attorno la media, posso utilizzare:
 - a. Indice di Kurtosis.
 - b. Indice di Skewness.
 - c. Media.
11. Quali sono le value propositions, i customer segments e le revenue streams del sito Immobiliare.it? Quale pattern di modello di business segue? Spiegarne il perché.