

Name: [REDACTED]

Surname: [REDACTED]

Matriculation number: _____

Appello di MODELLI DI E-BUSINESS E BUSINESS INTELLIGENCE

- 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.
- Per Sony Playstation, l'ingresso di Apple nel mercato del videogaming è:
- Forza di mercato.
 - Forza industriale.
 - Trend chiave.
- 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.
- La variabile rappresentativa della risposta alla domanda di un questionario (su scala da 1 a 6) è di tipo:
- Continuo.
 - Nominale.
 - Ordinale.
- Nell'ipotesi di avere un database contenente gli esisti 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.
- 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.
- 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.
- La value proposition principale di Netflix è:
- Guadagnare dagli abbonamenti.
 - Ampio catalogo.
 - Permettere di visualizzare contenuti in streaming.
- Se ho una variabile target categorica, posso utilizzare una regressione?
- Si, normalizzando la variabile.
 - No.
 - Si, discretizzando la variabile.
- 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.
- Per verificare se l'età di una popolazione è prevalentemente bassa o alta posso utilizzare:
- Indice di Kurtosis.
 - Indice di Skewness.
 - Media.

1.

- a) La coppia c-d ha le relazioni più intense

Questa affermazione è corretta perché la covarianza da sola non è utile per misurare l'intensità delle relazioni tra le variabili

LA COVARIANZA PUÒ INDICARE LA DIREZIONE DELLA RELAZIONE (POS O NEG) MA NON LA FORZA O INTENSITÀ DELLA RELAZIONE

- b) La coppia c-d presenta una discordanza

Una covarianza negativa (-0,9) significa che c e d tendono a variare in direzioni opposte indicando una discordanza tra due variabili

- c) La coppia a-b ha le relazioni più intense

Come già detto la covarianza non è utile a misurare l'intensità delle relazioni

2.

- a) FORZA MERCATO

Le forze mercato si riferiscono ai fattori che influenzano la domanda e l'offerta nel mercato, come le preferenze dei consumatori, le tendenze dei prezzi e la concorrenza generale.

in questo contesto specifico, stiamo guardando più all'IMPATTO COMPETITIVO DIRETTO piuttosto che alle dinamiche di mercato più ampio

- b) FORZA INDUSTRIALE

Le forze industriali riguardano le strutture competitive di un settore ed includono fattori come le rivalità tra i concorrenti, la minaccia di nuovi entranti e il potere di fornitori e acquirenti

L'ingresso di Apple nel mercato dei videogiochi rappresenta una nuova minaccia competitiva per Sony PlayStation, poiché introduce un nuovo e potente concorrente con significative risorse finanziarie

- c) TREND CHIAVE

I trend chiave si riferiscono ai cambiamenti macroeconomici, socioculturali, tecnologici e regolamentari che possono influenzare un settore a lungo termine

Sebbene l'ingresso di Apple possa essere influenzato da trend tecnologici più ampi

3.

a) IL VALORE PIÙ FREQUENTE

In caso di DATI CATEGORIALI come 'giorno' e 'mese', sostituire i valori mancanti con il valore più frequente (MODA) è una tecnica comune.

Questa tecnica assume che il valore più comune sia una buona rappresentazione dei dati mancanti.

b) UN VALORE SCELTO ARBITRARIAMENTE

NON È Raccomandato in quanto introduce un elemento soggettivo e potenzialmente distorsivo nei dati.

c) IL VALORE MEDIO

Il valore medio NON si applica direttamente ai dati categociali come 'giorno' e 'mese'.

La media è una misura statistica utilizzata per valori numerici, quindi non è pertinente per categorie discrete.

4.

a) CONTINUO

Le variabili continue possono assumere un numero infinito di valori entro un intervallo.

Ad esempio l'altezza, il peso e il tempo.

Una scala da 1 a 6 NON è continua in quanto i valori sono discreti e limitati.

b) NOMINALE

Le variabili nominali sono categociali e non hanno un ordine intrinseco.

Ad esempio il colore degli occhi o il genere.

La scala da 1 a 6 ha un ordine intrinseco, quindi NON è nominale.

c) ORDINALE

Le variabili ordinali sono categociali con un ordine intrinseco.

Ad esempio includono i livelli di istruzione o le risposte di un questionario su una scala di valutazione.

La scala da 1 a 6 riflette un ordine di preferenza o intensità, rendendo la variabile di tipo ordinale.

5.

a) REGRESSIONE

La regressione viene utilizzata per prevedere un valore continuo

Meglio per prevedere una probabilità di risposta o il valore atteso di una matrice continua come ad esempio le spese future di un utente

b) CLASSIFICAZIONE

Utilizzando un modello di classificazione si possono prevedere le categorie di sconto per ciascun utente basandosi sui dati storici

Un algoritmo di classificazione potrebbe comprendere le caratteristiche degli utenti che hanno risposto bene a ciascuna di sconti e utilizzare queste informazioni per assegnare il giusto sconto a nuovi utenti

c) REGOLE ASSOCIATIVE

Sono utili per trarre pattern e correlazioni all'interno dei dati. Possono essere utilizzate per identificare combinazioni di caratteristiche degli utenti che sono associate a risposte positive a determinati sconti

ad esempio si potrebbe scoprire che utenti di una certa età e con un certo comportamento di acquisto hanno una maggiore possibilità di rispondere positivamente ad uno sconto del 20%

6

a) È MINORE LA FREQUENZA DI UTENTI NON SODDISFATTI

Questa affermazione non è supportata direttamente da kurtosis. La kurtosis positiva indica minore dispersione ma non dice nulla sulle frequenze specifiche dei valori bassi o alti

b) MOLTI UTENTI HANNO SODDISFAZIONE PARI A 5

Una kurtosis positiva implica una maggiore concentrazione attorno alla media ma non possiamo affermare con certezza che il valore mode sia 5 solo sulla base della kurtosis

c) È MAGGIORE LA FREQUENZA DI UTENTI NON SODDISFATTI

Dato che la kurtosis positiva suggerisce minore dispersione e quindi maggiore concentrazione attorno alla media, non un aumento nelle frequenze dei valori estremamente bassi

7.

a) GUADAGNARE DAGLI ABBONAMENTI

Questo descrive il modello di business di Netflix, cioè come l'azienda genera entrate

Tuttavia non rappresenta la value proposition dal punto di vista del cliente

b) AMPIO CATALOGO

Un ampio catalogo di contenuti è sicuramente una componente importante della value proposition di Netflix

Tuttavia non è un elemento distintivo rispetto ad altre piattaforme streaming

c) PERMETTE DI VISUALIZZARE CONTENUTI IN STREAMING

Questa è la principale value proposition di Netflix.

La capacità di accedere facilmente ad una vasta gamma di contenuti in qualsiasi momento e su qualsiasi dispositivo è il motivo principale per cui gli utenti scelgono Netflix.

Il servizio di streaming on-demand con le possibilità di visualizzare contenuti senza interruzioni pubblicitarie, rappresenta il valore centrale offerto da Netflix

8.

Se si ha una variabile target categorica, l'uso di una regressione tradizionale (come la regressione lineare) non è appropriato

a) SI, NORMALIZZANDO UNA VARIABILE

Normalizzare una variabile si riferisce alle scale delle variabili indipendenti.

La normalizzazione non rende appropriato l'uso della regressione lineare per una variabile target

b) NO

Questo è in gran parte corretto se si sta considerando la regressione lineare classica.

Tuttavia, ci sono altre forme di regressione come quelle logistiche che sono adatte per variabili categoriali.

c) SI, DISCRETIZZANDO LA VARIABILE

Discretizzare si riferisce alle conversioni di variabili continue in variabili categroiche ma non è rilevante per l'uso della regressione lineare con una variabile target categorica

a) UN AUMENTO DELL'ACCURATEZZA DELL'ALBERO

Ridurre il numero di regole tramite potatura può EFFETTIVAMENTE aumentare l'accuratezza sui dati nuovi perché il modello sarà meno complesso e quindi meno incline all'OVERFITTING TUTTAVIA non garantisce un aumento dell'accuratezza sul Training set dove il modello potrebbe apparire meno accurato

b) UNA RIDUZIONE DELL'ACCURATEZZA DELL'ALBERO

Se considereremo l'accuratezza sul Training set questa è l'opzione più corretta.

Ridurre il numero di regole comporta un modello meno complesso che potrebbe non catturare tutte le variazioni presenti nei dati di training, riducendo così l'accuratezza con questi dati

TUTTAVIA il modello potrebbe generalizzare meglio sui dati di test

c) UN PEGGIORAMENTO DELLA LEGGIBILITÀ DELL'ALBERO

Ridurre il numero di regole generalmente migliora la leggibilità dell'albero decisionale

Un albero meno complesso è più facile da interpretare e capire

10.

a) INDICE DI KURTOSIS

Misura la pesantezza delle code di una distribuzione ovvero quanto frequentemente si verificano valori estremamente alti o bassi rispetto alla media.

NON è utile per determinare se una distribuzione è spostata verso altri valori bassi o alti ma piuttosto per capire la forma della distribuzione in termini di code

b) INDICE DI SKEWNESS - INDICE DI ASIMMETRIA

- misure la simmetria della distribuzione dei dati.

POSITIVA coda più lunga a destra

NEGATIVA coda più lunga a sinistra

c) MEDIA

La media fornisce il valore medio della distribuzione dei dati ma non informa sulla della distribuzione stessa

TRACCIA INDEFINITA

	Surname	Matriculation number:
1.	Se l'indice di Gini misurato per la variabile "categoria acquistata" è pari a 1, vuol dire che l'utente:	
a.	acquista diverse categorie di prodotto	
b.	acquista una sola categoria di prodotto sempre uguali	
c.	acquista categorie di prodotto sempre differenti	
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 "addoddiamento" è pari a 0,8, vuol dire che:	
a.	È minore la frequenza di utenti 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.	Possiamo trasformare una variabile continua in categorie:	
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 <i>value propositions</i> . Quale pattern di modello di business segue? Spiegarne il perché.	

12. Immagine di avere il seguente database contenente una serie di informazioni relative alla navigazione e all'identificazione degli utenti su di un sito web. In particolare, viene registrato l'identificativo dell'utente, ha visualizzato la pagina visualizzata, l'identificativo della sessione durante la quale l'utente ha prodotto su cui si è clonato ogni sezione, se questa sia stata visualizzata più pagine, la data effettuata dall'utente durante la sessione, l'età e il genere dell'utente.						
User ID	Page ID	Session ID	Date	Product ID	Purchase the product	Amount of money spent during the session
1	1	1	11.06.2023	1	NO	18 M
1	2	1	11.06.2023	2	NO	18 F
1	4	2	20.06.2023	10 YES	10	18 M
2	2	2	20.06.2023	2 NO	2	18 F
2	3	2	20.06.2023	6 NO	2	18 F
2	4	3	15.06.2023	20 YES	4	18 F
3	5	3	20.06.2023	2 NO	4	20 M
3	1	3	15.06.2023	2 NO	20	20 M
3	4	3	15.06.2023	5 NO	15	20 F
3	5	3	15.06.2023	2 YES	15	20 F
3	6	3	20.06.2023	3 NO	15	20 M
4	3	4	20.06.2023	1 NO	0	20 M
4	1	4	15.06.2023	12 NO	20	20 M
4	2	4	15.06.2023	5 YES	20	20 M
4	3	4	15.06.2023	8 NO	20	20 M
4	4	4	20.06.2023	12 NO	20	20 M
4	5	4	20.06.2023	2 YES	15	20 M
4	6	4	20.06.2023	2 NO	15	20 M
5	1	5	15.06.2023	5 YES	8	20 M

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

1. L'indice di Gini misura le disegualanze di una distribuzione. Il valore 1 indica MASSIMA DISUGUALIANZA mentre 0 indica PERFETTA OMOGENEITÀ

In questo caso 0 acquista in modo omogeneo tutte le categorie di prodotto / Totele concentrazione su una sola categoria

a) ACQUISTA CATEGORIE DI PRODOTTO SEMPRE DIVERSE

NO perchè implicherebbe un indice di GINI più vicino a 0 perché ci sarebbe una distribuzione uniforme delle categorie acquistate



b) ACQUISTA CATEGORIE DI PRODOTTO SEMPRE UGUALI

c) ACQUISTA UNA SOLA CATEGORIA DI PRODOTTI

Questo' opzione implica un indice di GINI pari a 1 perchè c'è massima disegualanza nella distribuzione delle categorie acquistate

2.

a) FORZA MERCATO

Si riferisce a fattori che influenzano il comportamento dei clienti e dei fornitori in un mercato specifico.

b) FORZA INDUSTRIALE

Riguarda le dinamiche interne di un settore o di un'industria specifica come competit., borsiere dell'entrate

c) FORZA MACRO-ECONOMICA

Si riferisce a fattori economici che influenzano l'economia in generale come i tassi di interesse, inflazione, tasso di dissoluzione

3.

a) IL VALORE PIÙ FREQUENTE

In caso di DATI CATEGORIALI come 'giorno' e 'notte', sostituire i valori mancanti con il valore più frequente (MODA) è una tecnica comune.

Questa tecnica assume che il valore più comune sia una buona rappresentazione dei dati mancanti.

b) UN VALORE SCELTO ARBITRARIAMENTE

NON È RACCOMANDATO in quanto introduce un elemento soggettivo e potenzialmente distorsivo nei dati.

c) IL VALORE MEDIO

Il valore medio NON si applica direttamente ai dati categoriali come 'giorno' e 'notte'.

La media è una misura statistica utilizzata per valori numerici, quindi non è pertinente per categorie discrete.

4

a) VARIABILE CATEGORICA

Non necessariamente poiché la variabile target può essere sia categorica (PER LA CLASSIFICAZIONE) che continua (PER LA REGRESSIONE).

↓
UGUALE

b) VARIABILE CONTINUA

c) VARIABILE TARGET

Un metodo di apprendimento supervisionato richiede la presenza di una variabile target che può essere sia categorica che continua a seconda del tipo di problema.

5.

a) REGRESSIONE

È una tecnica di apprendimento supervisionato utilizzata per prevedere valori continui.

In questo contesto può essere utilizzata per modellare le relazioni tra le Temperature ed altre variabili meteorologiche come ad es l'UMIDITÀ.

b) CLASSIFICAZIONE

È un'altra tecnica di apprendimento supervisionato ma viene utilizzata quando la variabile target è categorica. La CLASSIFICAZIONE prevede l'assegnazione di etichette ad un insieme di dati in base alle caratteristiche di ingresso.

In questo caso per classificare i giorni come 'caldi', 'freddi', 'piacevoli'.

c) REGOLE ASSOCIATIVE

Le regole associative sono utilizzate per trovare associazioni o pattern fra variabili in grandi insiemi di dati.

Questa tecnica è più comunemente utilizzata nell'analisi di mercato MA NON APPLICA per le previsioni di valori comuni come le temperature.

6. L'INDICE DI SKEWNESS misura la SIMMETRIA delle distribuzioni dei dati



a) È MINORE LA FREQUENZA DI UTENTI NON SODDISFAZIONE

Indice positivo Indice che ci sono meno utenti con valori alti di soddisfazione non meno utenti con valori bassi

b) MOLTI UTENTI HANNO SODDISFAZIONE PARI A 5

Questo valore è difficile verificare solo con il valore di SKEWNESS, perché non ci dice dove si concentra esattamente il picco delle distribuzioni.

c) È MAGGIORE LA FREQUENZA DI UTENTI NON SODDISFAZIONE

Indice che ci sono più valori bassi rispetto ai valori alti

7.

a) CLIENTE I clienti sono individui o gruppi che acquistano prodotti o servizi
Nel caso del punto vendita, i clienti sono le persone che entrano nei negozi per acquistare

b) REVENUE STREAM Queste opzioni si riferisce alle fonti di entrate principali.

Nel caso dei punti vendita il flusso di ricevi proviene dalla vendita diretta ai clienti dei prodotti

Queste entrate sono essenziali per la redditività complessiva dell'azienda

c) PARTNERSHIP

Starbucks può essere coinvolto in partnership con altre aziende o entità per varie iniziative, come collaborazioni per prodotti speciali, accordi di distribuzione e alleanze strategiche con altre aziende.

Tuttavia, la gestione di un punto vendita di Starbucks non viene direttamente sotto le categorie partnership

8

a) SI, SOLO SE NON CI SONO VALORI MANCANTI

Non è necessariamente vero che la presenza di valori mancanti impedisce la trasformazione di una variabile continua in una categoreica

b) NO Palesemente errata

c) SI, DISCRETIZZANDO LA VARIABILE

Discretizzare una variabile continua significa suddividerla in intervalli o categorie distinte

Questo processo consente di trasformare una variabile continua in una variabile categoreica

9

a) AUMENTARE LA DIMENSIONE DELL'ALBERO

Aumentare la complessità dell'albero decisionale può portare ad uno sovradattamento ai dati di addestramento, riducendo le capacità di generalizzazione del modello e quindi non migliorando necessariamente l'accuratezza.

b) RIDURRE LA DIMENSIONE DI UN ALBERO

Ridurre la complessità dell'albero può essere una strategia per prevenire l'OVERFITTING ma non una tecnica per migliorare l'accuratezza di un albero decisionale esistente.

c) UTILIZZARE IL CROSS-VALIDATION

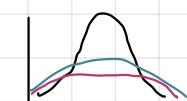
È una tecnica utilizzata per valutare le prestazioni di un modello e per stimare quanto bene si generalizzerà ai dati non visti.

QUESTA TECNICA AIUTA AD EVITARE L'OVERFITTING ED A MIGLIORARE L'ACCURATEZZA DEL MODELLO, ottimizzando i parametri o la struttura dell'albero decisionale.

10

a) INDICE DI KURTOSIS

Indice quanto una distribuzione di dati è 'appuntito' o piatto



b) INDICE DI SKEWNESS

Misura l'assimetria rispetto alla media

c) MEDIA

È semplicemente il valore medio dei dati non fornisce informazioni sulle distribuzione o sulla concentrazione