

Informazioni Preliminari

Il documento presenta, per ogni argomento trattato nel corso di *Fondamenti di Intelligenza Artificiale*, una serie di domande di preparazione alla prova scritta dell'esame. Il documento sarà soggetto a continui aggiornamenti (nuove domande, riformulazione di quelle già presenti, correzione di eventuali errori); la data in oggetto è perciò indicativa dell'ultimo aggiornamento compiuto.

Le domande sono raggruppate per argomento, sulla base del programma del corso disponibile al link: <https://corsi.unisa.it/informatica/didattica/insegnamenti?anno=2022&id=511550>.

Le slide e il materiale aggiuntivo fornito dal docente sono disponibili sulla piattaforma e-learning di ateneo al seguente link: <http://elearning.informatica.unisa.it/el-platform/course/view.php?id=817>. L'accesso è ristretto agli studenti registrati al corso.

Prima di continuare la lettura del documento, è necessario sottolineare che l'insieme di domande riportate **è da considerarsi indicativo per la prova scritta, ma NON esaustivo**. Pertanto, è fortemente consigliato di utilizzare questo documento come **una guida per la preparazione dell'esame e non come l'unica sorgente da considerare**.

La partecipazione alle lezioni nonché lo studio individuale, svolto sulla base delle slide del corso e dei libri di testo consigliati, sono necessari per poter superare con successo la prova d'esame. All'atto della prova scritta, infatti, gli studenti potrebbero trovare domande diverse rispetto a quelle presentate in questo documento; saranno preservate le sole tipologie di domande (a risposta multipla e aperte).

Per eventuali chiarimenti sugli argomenti di esame e/o sulle domande di preparazione riportate in questo documento, è possibile contattare il docente inviando una e-mail al seguente indirizzo: fpalomba@unisa.it.

Part I

Fondamenti: Cos'è l'Intelligenza Artificiale

**In grassetto sono riportate le risposte corrette.*

1. Quali dei seguenti aspetti **non** è compreso nella definizione di Intelligenza Artificiale?
 - A. Agire umanamente;
 - B. **Agire lucidamente;**
 - C. Pensare umanamente;
 - D. Pensare razionalmente.
2. Si completi la seguente frase. “L'introspezione fa parte [...]”?
 - A. [...] **dell'approccio della modellazione cognitiva.**
 - B. [...] dell'approccio del test di Turing.
 - C. [...] dell'approccio logicista.
 - D. [...] dell'approccio degli agenti razionali.
3. Si completi la seguente frase. “Nel gioco delle imitazioni, l'obiettivo del giocatore umano B è quello di [...]”
 - A. [...] fingere di essere una macchina.
 - B. [...] indurre in errore il secondo giocatore A.
 - C. [...] **aiutare l'interrogante nella corretta identificazione della macchina.**
 - D. [...] aiutare la macchina nella corretta identificazione dell'interrogante.
4. Si completi la seguente frase. “In origine, il pensiero corretto veniva codificato tramite [...]”
 - A. [...] la neuroscienza.
 - B. [...] l'imaging cerebrale.
 - C. [...] **i sillogismi aristotelici.**
 - D. [...] il test di Turing.
5. Fornire una definizione completa di Intelligenza Artificiale.

6. Descrivere il ruolo della psicologia cognitiva all'interno dell'Intelligenza Artificiale.

7. Descrivere il test di Turing.

8. Descrivere i modelli di rappresentazione del cervello umano.

Part II

Formulazione di Problemi

**In grassetto sono riportate le risposte corrette.*

1. Un agente basato su obiettivi possiede le seguenti caratteristiche:
 - A. **Si basa su una rappresentazione atomica degli stati del mondo;**
 - B. **Non conosce la rappresentazione interna di uno stato;**
 - C. Si basano su regole condizione-azione;
 - D. Ha un unico obiettivo e mira a raggiungerlo.
2. In un ambiente deterministico e osservabile, quali di queste caratteristiche NON sono necessariamente valide?
 - A. **A partire da uno stato, esiste un numero finito di azioni tra l'agente può scegliere.**
 - B. L'agente conosce sempre lo stato corrente del mondo.
 - C. Ogni azione avrà uno ed unico risultato.
 - D. **L'agente saprà sempre quali stati saranno raggiunti da ciascuna azione.**
3. Si completi la seguente frase. "Un modello di transizione descrive [...]"
 - A. [...] l'insieme di azioni possibili da uno stato.
 - B. [...] l'obiettivo che l'agente dovrà verificare.
 - C. [...] **il risultato di ogni azione attuabile dall'agente.**
 - D. [...] il costo numerico che porta da uno stato di partenza ad un obiettivo.
4. Si completi la seguente frase. "Un'astrazione è valida se [...]"
 - A. [...] **possiamo espandere ogni soluzione astratta in una soluzione del mondo più dettagliata.**
 - B. [...] eseguire ogni azione nella soluzione è più facile che nel problema originale.
 - C. [...] mantiene quanti più dettagli possibile dello stato del mondo.
 - D. [...] specifica una sequenza di azioni che porta al raggiungimento di un obiettivo.
5. Si completi la seguente frase. "Nella formulazione di un problema, un cammino ciclico è [...]"
 - A. [...] anche detto cammino additivo.
 - B. [...] esprime il concetto di frontiera.
 - C. [...] **un cammino che contiene più copie di uno stesso stato.**
 - D. [...] **un caso particolare di cammino ridondante.**
6. Descrivere i componenti di un problema.

7. Descrivere il processo di ricerca.

8. Riportare i principali passi di un semplice agente risolutore di problemi.

Part III

Algoritmi di Ricerca Non Informata

**In grassetto sono riportate le risposte corrette.*

1. Quale delle seguenti definizioni meglio esprime il concetto di ricerca in profondità?
 - A. Strategia di ricerca in cui un nodo viene espanso solo se la lunghezza del cammino corrispondente è minore rispetto al massimo stabilito.
 - B. Strategia di ricerca in cui tutti i nodi di profondità d sono espansi prima di quelli di profondità $d+1$.
 - C. Strategia di ricerca che espande prima i nodi n con il minimo costo di cammino $g(n)$.
 - D. **Strategia di ricerca in cui viene sempre espanso prima il nodo più profondo nella frontiera corrente dell'albero di ricerca.**
2. Si completi la seguente frase. "Il principale vantaggio di una strategia di ricerca bidirezionale è [...]"
 - A. [...] la maggiore facilità di implementazione rispetto ad altri algoritmi di ricerca non informata.
 - B. [...] **la minore complessità spaziale rispetto ad altri algoritmi di ricerca non informata.**
 - C. [...] la maggiore facilità di definizione di un obiettivo rispetto ad altri algoritmi di ricerca non informata.
 - D. [...] la maggiore flessibilità di definizione rispetto ad altri algoritmi di ricerca non informata.
3. Definire il concetto di ricerca ad approfondimento iterativo.

4. Descrivere le principali strutture dati necessarie per la formulazione di un algoritmo di ricerca.

5. Discutere delle prestazioni (in termini di completezza, ottimalità, complessità temporale e spaziale) della ricerca in ampiezza.

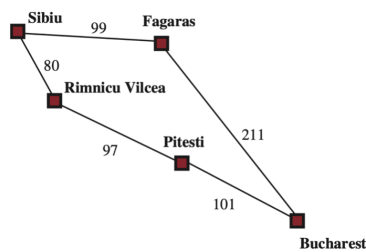


Figure 1: Percorso che collega il punto di partenza Sibiu all'obiettivo Bucarest.

6. Si consideri il percorso tra Sibie e Bucarest raffigurato in Figura 1. Riportare le espansioni dei nodi effettuate da un algoritmo di ricerca in profondità.

Part IV

Algoritmi di Ricerca Informata

**In grassetto sono riportate le risposte corrette.*

1. La ricerca best-first è, da un punto di vista implementativo molto simile ad una ricerca a costo uniforme. Perché?
 - A. **Un nodo viene espanso sulla base del valore costo stimato più alto rispetto al valore di costo effettivo più alto.**
 - B. La frontiera è memorizzata in una coda ordinata per costo di cammino rispetto al costo stimato.
 - C. Un nodo viene espanso sulla base del valore costo effettivo più alto rispetto al valore di costo stimato più alto.
 - D. **La frontiera è memorizzata in una coda ordinata per costo costo stimato rispetto a quello di cammino.**
2. In qualche condizione l'algoritmo A* risulta essere ottimale?
 - A. L'euristica $h(n)$ deve essere consistente.
 - B. L'euristica $h(n)$ deve rispettare la disuguaglianza triangolare tra il costo effettivo e quello stimato.
 - C. L'euristica $h(n)$ deve essere ammissibile.
 - D. **L'euristica $h(n)$ deve essere sia ammissibile che consistente.**
3. In quali condizioni l'algoritmo Simplified Memory Bounded A* risulta essere sia completo che ottimale?
 - A. La soluzione non si trova nella frontiera finora esplorata.
 - B. **La soluzione è raggiungibile.**
 - C. La soluzione è raggiungibile in tempo lineare.
 - D. **La soluzione raggiungibile è ottima.**
4. Si completi la seguente frase. "L'algoritmo best-first greedy [...]"
 - A. [...] **espande sempre il nodo più vicino all'obiettivo.**
 - B. [...] espande sempre il nodo meno promettente in termini di $h(n)$.
 - C. [...] espande sempre il nodo più profondo.
 - D. [...] compie una ricerca in ampiezza dell'albero di ricerca.
5. Descrivere le caratteristiche dell'algoritmo di ricerca A*, specialmente in termini di funzione euristica.

6. Descrivere le caratteristiche della Beam Search, indicando le criticità che non la rendono né completa né ottimale.

7. Quali sono le differenze tra l'algoritmo di ricerca ad approfondimento iterativo standard e la Iterative Deepening A* (IDA*)?

Part V

Algoritmi di Ricerca Locale

**In grassetto sono riportate le risposte corrette.*

1. Cos'è un algoritmo di ricerca locale?
 - A. Un algoritmo di miglioramento parallelo.
 - B. Un algoritmo deterministico.
 - C. Un algoritmo sostitutivo.
 - D. **Un algoritmo di miglioramento iterativo.**
2. Si completi la seguente frase: “Una struttura dei vicini è definita come [...]”
 - A. [...] una funzione F che assegna a ogni soluzione s il suo valore.
 - B. [...] **una funzione F che assegna a ogni soluzione s dell'insieme di soluzioni S un insieme di soluzioni $N(s)$ sottoinsieme di S .**
 - C. [...] una funzione F che assegna il una stima di quanto è vicino la stato s alla soluzione.
 - D. [...] una funzione F che assegna a una singola soluzione s dell'insieme di soluzioni S un insieme di soluzioni $N(s)$ sottoinsieme di S .
3. Si completi la seguente frase. “Uno dei tipici problemi dell'algoritmo Hill-Climbing è rappresentato [...]”
 - A. [...] **dalla presenza di plateau, ovvero un regione dello spazio dove gli stati vicini hanno tutti lo stesso valore.**
 - B. [...] dalla generazione casuale dei successori di un nodo, che porta alla mancata osservazione del panorama dello spazio degli spazi.
 - C. [...] dalla presenza di massimi globali, che bloccano la ricerca non consentendo di identificare la soluzione migliore.
 - D. Nessuna delle opzioni precedenti.
4. Qual è la limitazione principale dell'algoritmo Hill-Climbing con random-walk che porta all'utilizzo della tecnica di memorizzazione di una funzione euristica di stima di costo?
 - A. La mancanza di completezza del primo algoritmo.
 - B. **La lentezza con cui ambienti complessi vengono esplorati.**
 - C. La mancanza di ottimalità del primo algoritmo.
 - D. La probabilità di finire in un vicolo cieco.
5. Si completi la seguente frase. “Un *ridge* è [...]”
 - A. [...] un punto piatto della regione dello spazio.
 - B. [...] un punto dello spazio che presenta uno spigolo in salita.
 - C. [...] **una porzione dello spazio di ricerca che presenta una brusca variazione.**
 - D. [...] un massimo locale nello spazio di ricerca.
6. Quale delle seguenti affermazioni sugli algoritmi di ricerca locale è *falsa*?
 - A. Nella ricerca locale, lo stato obiettivo è esso stesso la soluzione al problema.
 - B. Gli algoritmi di ricerca locale sono algoritmi di miglioramento iterativo.
 - C. **Gli algoritmi di ricerca locale non possono essere generalmente applicati in problemi con spazio degli stati grandi/infiniti.**
 - D. Gli algoritmi di ricerca locale usano poca memoria, molto spesso avendo una complessità costante.
7. Si completi la seguente frase. “Un algoritmo genetico è [...]”
 - A. [...] un algoritmo di ricerca locale ispirato alla teoria evoluzionistica di Darwin.
 - B. [...] **una meta-euristica con la quale poter definire algoritmi di ricerca.**
 - C. [...] un algoritmo di ricerca locale ispirato alla teoria della Swarm Optimization.
 - D. [...] un algoritmo di ricerca simile a quello della Ant Colony.

-
8. Quale delle seguenti meglio riflette la definizione di Roulette Wheel selection?
- A. Gli individui della prossima popolazione vengono casualmente selezionati.
 - B. Gli individui della popolazione vengono ordinati rispetto al loro valore di fitness e vengono selezionati i primi k elementi.
 - C. Gli individui della popolazione vengono ordinati rispetto al loro valore di fitness e a ciascuno di essi viene assegnato un rango in base alla posizione.
 - D. **Gli individui della popolazione ricevono una probabilità di selezione pari al valore della loro fitness relativa all'intera popolazione.**
9. Quale delle seguenti meglio riflette la definizione di mutazione nel contesto degli algoritmi genetici?
- A. La mutazione è un metodo di ordinamento delle soluzioni di un algoritmo genetico.
 - B. La mutazione è una variazione progettata per sostituire un gene di un individuo con un gene di un secondo individuo.
 - C. La mutazione è una operazione che consente a due individui della popolazione di combinare i propri geni.
 - D. **La mutazione è una variazione arbitraria di uno o più geni di un individuo.**
10. Quale delle seguenti meglio riflette la definizione di Elitism?
- A. Tecnica che consente, ad ogni iterazione di un algoritmo genetico, di restringere l'insieme delle funzioni di fitness da valutare a seconda del risultato dell'iterazione precedente.
 - B. **Una tecnica che consente ai migliori individui di una popolazione di sopravvivere ed essere portati nella generazione successiva di un algoritmo genetico.**
 - C. Una tecnica che consente di fornire un ordinamento totale tra gli individui nel fronte di Pareto attraverso la definizione di una funzione di preferenza.
 - D. Una tecnica che consente di mantenere una popolazione aggiuntiva che non evolve ma che contiene gli individui che hanno soddisfatto obiettivi non soddisfatti in iterazioni precedenti.
11. Si completi la seguente frase. "L'elitism è definito come [...]"
- A. una tecnica che consente di fornire un ordinamento totale tra gli individui nel fronte di Pareto attraverso la definizione di una funzione di preferenza.
 - B. una tecnica che consente di mantenere una popolazione aggiuntiva che non evolve ma che contiene gli individui che hanno soddisfatto obiettivi non soddisfatti in iterazioni precedenti.
 - C. **una tecnica che consente ai migliori individui di una popolazione di sopravvivere ed essere portati nella generazione successiva di un algoritmo genetico.**
 - D. una tecnica che consente, ad ogni iterazione di un algoritmo genetico, di restringere l'insieme delle funzioni di fitness da valutare a seconda del risultato dell'iterazione precedente.
12. Quali delle seguenti affermazioni sul rapporto di competitività nella ricerca online è falsa?
- A. Questo rappresenta il rapporto tra il costo di una soluzione ottenuta e il costo della soluzione che l'agente potrebbe ottenere se conoscesse in anticipo lo spazio di ricerca.
 - B. **Il rapporto di competitività non può essere infinito.**
 - C. Nessun algoritmo può evitare vicoli ciechi in tutti gli spazi degli stati.
 - D. Questo viene utilizzato per studiare il costo di un algoritmo di ricerca.
13. Quali dei seguenti algoritmi possono essere utilizzati per risolvere problemi online?
- Nota: una o più risposte potrebbero essere corrette.*
- A. Un algoritmo di ricerca in ampiezza
 - B. Un algoritmo basato su un gioco non cooperativo.
 - C. **Un algoritmo di ricerca in profondità.**
 - D. **Un algoritmo basato su Hill-Climbing.**
14. Si completi la seguente frase: "La ricerca local beam [...]"
- A. [...] **sceglie k successori in maniera casuale rispetto a scegliere i k successori migliori.**

-
- B. [...] ha l'obiettivo di selezionare un singolo stato e migliorarlo in maniera iterativa.
 - C. [...] sceglie i k successori migliori rispetto a sceglierli in maniera casuale.
 - D. [...] è equivalente all'algoritmo Hill-Climbing con riavvio casuale.

15. Descrivere le principali differenze tra algoritmi di ricerca tradizionali e algoritmi di ricerca locali.

16. Definire il concetto di struttura dei vicini.

17. Elaborare sui principali vantaggi e svantaggi degli algoritmi genetici.

18. Definire il processo generale di definizione di un algoritmo genetico.

19. Descrivere due algoritmi (a scelta) di mutazione applicabili nel contesto di un algoritmo genetico in cui gli individui hanno una codifica in forma di stringa di interi.

20. Descrivere il concetto di fronte di Pareto nel contesto di algoritmi genetici multi-obiettivo.

21. Fornire una formulazione di un problema di ricerca online.

22. Descrivere le variazioni all'utilizzo di Hill-Climbing per la risoluzione di problemi di ricerca online.

Part VI

Algoritmi di Ricerca con Avversari

**In grassetto sono riportate le risposte corrette.*

1. Si completi la seguente frase. “In un gioco a somma zero [...]”
 - A. [...] l'ambiente è necessariamente di tipo singolo agente.
 - B. [...] i giocatori conoscono sempre l'insieme di azioni che l'avversario può effettuare.
 - C. [...] due giocatori si sfidano alternando azioni fino al termine della partita.
 - D. [...] **i valori di utilità sono sempre uguali ma di segno opposto.**
2. Quale delle seguenti affermazioni relative agli alberi di gioco è *vera*?
 - A. Sono costruiti complessi da rappresentare ma semplici da esplorare.
 - B. Sono costruiti semplici da rappresentare ed esplorare.
 - C. **Sono costruiti semplici da rappresentare ma complessi da esplorare.**
 - D. Sono costruiti che possono essere rappresentati tramite l'utilizzo di un array di stringhe.
3. Quale delle seguenti meglio riflette la definizione di equilibrio di Nash?
 - A. **L'equilibrio riflette una combinazione di strategie in cui ciascun giocatore effettua la migliore scelta possibile sulla base delle aspettative di scelta dell'altro giocatore.**
 - B. L'equilibrio riflette una combinazione di strategie in cui i giocatori effettuano la migliore scelta possibile sulla base di un ipotetico accordo di cooperazione.
 - C. L'equilibrio riflette una combinazione di strategie in cui i giocatori effettuano la migliore scelta possibile sulla base di un ipotetico accordo di non cooperazione.
 - D. L'equilibrio riflette una combinazione di strategie in cui i giocatori ignorano le dinamiche dominanti che governano l'ambiente di gioco.
4. Si completi la seguente frase. “Nella teoria dei giochi, un ottimo paretiano [...]”
 - A. [...] è raggiunto in una situazione in cui è possibile migliorare la condizione di un giocatore senza peggiorare la condizione di un altro.
 - B. [...] è raggiunto in una situazione in cui è consentito ai giocatori di poter comunicare e operare secondo una strategia comune.
 - C. [...] **è raggiunto in una situazione in cui non è possibile migliorare la condizione di un giocatore senza peggiorare la condizione di un altro.**
 - D. [...] è raggiunto in una situazione in cui non esiste solo un modo per migliorare la condizione di un giocatore senza peggiorare la condizione di un altro.
5. Quale delle seguenti è una limitazione dell'algoritmo Minimax?
 - A. L'eccessiva complessità spaziale.
 - B. **L'eccessiva complessità temporale.**
 - C. La non completezza.
 - D. La non ottimalità.
6. Quale delle seguenti affermazioni relative alla potatura alfa-beta è *falsa*?
 - A. **La potatura può potare rami che influenzano la decisione finale.**
 - B. La potatura può essere applicata ad alberi di qualunque profondità.
 - C. La potatura sarebbe più efficace se le mosse fossero ordinate.
 - D. La potatura restituisce lo stesso risultato della tecnica minimax standard.
7. Si completi la seguente frase. “L'applicazione di un algoritmo di ricerca ad approfondimento iterativo alla potatura alfa-beta [...]”
 - A. [...] consente di disporre di una lista di stati già esplorati in precedenti momenti della partita.

-
- B. [...] **consente un ordinamento dinamico delle mosse che la ricerca dovrà eseguire.**
- C. [...] consente di limitare il numero di mosse da considerare.
- D. [...] consente di disporre di una funzione di stima della complessità spaziale della ricerca.
8. Quale delle seguenti affermazioni relative all'uso delle trasposizioni è *vera*?
- A. Una trasposizione è equivalente ad un algoritmo di selezione di tipo Roulette Wheel.
- B. Una trasposizione memorizza un insieme di stati obiettivo da raggiungere.
- C. Una trasposizione necessita della specifica di una funzione euristica.
- D. **Una trasposizione memorizza la valutazione di una configurazione.**
9. Descrivere il concetto di Equilibrio di Nash e contestualizzarlo nel contesto del Dilemma del Prigioniero.
- _____
- _____
- _____
10. Analizzare le prestazioni dell'algoritmo Minimax in termini di completezza, ottimalità, complessità spaziale e temporale.
- _____
- _____
- _____
11. Descrivere il funzionamento dell'algoritmo Minimax con potatura alfa-beta.
- _____
- _____
- _____
12. Descrivere i principali metodi di ordinamento delle mosse, spiegando inoltre il vantaggio che questi portano alle prestazioni della potatura alfa-beta.
- _____
- _____
- _____
13. Descrivere la variazione, in termini di formulazione del problema e cambiamenti implementativi, richiesta per poter consentire all'algoritmo Minimax di prendere decisioni imperfette in tempo reale.
- _____
- _____
- _____

Part VII

Teoria dell'Apprendimento

**In grassetto sono riportate le risposte corrette.*

1. Quale delle seguenti affermazioni sull'apprendimento non supervisionato è *vera*?
 - A. E' un tipo di apprendimento che richiede la presenza di dati etichettati sulla variabile dipendente.
 - B. Per poter utilizzare in maniera corretta un algoritmo non supervisionato, il progettista dovrà fornire indicazioni di dettaglio sulle caratteristiche delle classi da predire.
 - C. **Un algoritmo non supervisionato sarà in grado di stimare il valore della variabile dipendente senza conoscerne il valore reale.**
 - D. Tipicamente, gli algoritmi non supervisionati vengono utilizzati per problemi meno complessi.
2. Si completi la seguente frase. "L'errore irriducibile [...]"
 - A. [...] dipende esclusivamente dai dati che abbiamo a disposizione.
 - B. [...] **indica una misura di variabilità intrinseca del fenomeno in esame.**
 - C. [...] rappresenta l'unico errore che può verificarsi in un modello di machine learning.
 - D. [...] è proporzionale alla dimensione dei dati di input ed è del tutto indipendente dalle caratteristiche del problema.
3. Quale delle seguenti operazioni NON è utilizzata per mitigare il rischio di underfitting e overfitting?
 - A. La selezione delle caratteristiche rilevanti.
 - B. **La diminuzione della dimensione del dataset.**
 - C. La convalida incrociata delle prestazioni del modello.
 - D. La configurazione dei parametri di un algoritmo di machine learning.
4. Quale delle seguenti meglio riflette la definizione di machine learning?
 - A. **Il machine learning consente la definizione di algoritmi che possano imparare dai dati e sulla base di questi fare previsioni.**
 - B. Il machine learning consente la definizione di algoritmi multi-obiettivo che risolvono situazioni complesse tramite ottimizzazione dei diversi obiettivi.
 - C. Il machine learning consente la definizione di strumenti che possano stimare i dati di input e fornire indicazioni su eventuali dati mancanti che il progettista dovrà provvedere ad integrare.
 - D. Il machine learning consente la definizione di algoritmi tramite i quali poter configurare appropriatamente i parametri di algoritmi di ricerca.
5. Descrivere i concetti di errore, bias e varianza, con particolare riferimento al compromesso bias-varianza.

6. Descrivere le componenti di un agente capace di apprendere.

7. Fornire una definizione di machine learning, oltre che una categorizzazione delle varie tipologie di apprendimento.

Part VIII

Ingegneria del Machine Learning

**In grassetto sono riportate le risposte corrette.*

1. Si completi la seguente frase. “Considerando il modello CRISP-DM, la fase di data preparation [...]”
 - A. [...] consente di esplorare i dati, così da identificare eventuali dati mancanti da integrare.
 - B. [...] è riservata alla validazione di un modello di machine learning.
 - C. [...] è quella in cui il progettista dovrà sperimentare diversi classificatori, così da identificare il migliore da usare per predire nuovi dati.
 - D. [...] **ha come attività principale quella del feature engineering.**
2. Quale delle seguenti meglio riflette la definizione di data modeling?
 - A. **E’ una fase del modello CRISP-DM nella quale poter selezionare la tecnica da usare e procedere al suo addestramento.**
 - B. E’ una fase del modello CRISP-DM nella quale poter pulire i dati e selezionare le caratteristiche rilevanti da utilizzare nel problema in esame.
 - C. E’ una fase del modello TDSP eseguita dal product owner.
 - D. E’ una fase del modello TDSP nella quale è possibile comunicare con il cliente per identificare i requisiti del modello che dovrà essere sviluppato.
3. Quale delle seguenti affermazioni, riferite allo SCRUM Master, è *falsa*?
 - A. Le responsabilità dello SCRUM Master includono il controllo e la gestione dei conflitti tra gli sviluppatori.
 - B. **Le responsabilità dello SCRUM Master includono la gestione del personale che svilupperà il progetto.**
 - C. Le responsabilità dello SCRUM Master includono il controllo dei processi di sviluppo.
 - D. Le responsabilità dello SCRUM Master includono il coordinamento delle attività di sviluppo.
4. Fornire una descrizione del modello CRISP-DM.

5. Descrivere i cambiamenti richiesti al modello CRISP-DM per poter consentire uno sviluppo agile.

6. Descrivere la fase di data understanding, con particolare riferimento alle attività che un progettista dovrà svolgere.

7. Descrivere le principali differenze tra i modelli CRISP-DM e TDSP.

Part IX

Qualità dei Dati e Feature Engineering

**In grassetto sono riportate le risposte corrette.*

1. Quale delle seguenti affermazioni, riferite ai dati strutturali, è *vera*?
 - A. **Sono dati aventi forma tabulare ed in cui righe e colonne sono ben definite.**
 - B. Sono dati non aventi una struttura ben precisa.
 - C. Sono dati rappresentati da qualsiasi tipo di file che non ricade nella categoria dei dati non strutturati.
 - D. Nessuna delle precedenti.
2. Si completi la seguente frase. “La data imputation [...]”
 - A. è l'insieme delle tecniche che possono normalizzare i dati non strutturati da utilizzare nel contesto di un modello di machine learning.
 - B. **è l'insieme delle tecniche che possono stimare il valore di dati mancanti sulla base dei dati disponibili.**
 - C. è l'insieme delle tecniche che possono determinare le caratteristiche (feature) dai dati grezzi estraibili tramite metodi di data mining.
 - D. è l'insieme delle tecniche che possono selezionare le caratteristiche più correlate al problema in esame, a partire da un insieme di caratteristiche esistenti.
3. Quale delle seguenti meglio riflette la definizione di feature engineering?
 - A. Il feature engineering è un processo nel quale il progettista normalizza e/o scalarl'insieme di valori di una caratteristica.
 - B. Il feature engineering è un processo nel quale il progettista seleziona le caratteristiche più correlate al problema in esame, a partire da un insieme di caratteristiche esistenti.
 - C. Il feature engineering è un processo nel quale il progettista converte un dataset sbilanciato in un dataset bilanciato.
 - D. **Il feature engineering è un processo nel quale il progettista utilizza la propria conoscenza del dominio per determinare le caratteristiche (feature) dai dati grezzi estraibili tramite metodi di data mining.**
4. Descrivere un tipico processo di data preparation, con particolare riferimento alle operazioni di pulizia dei dati necessari per disporre di dati utili per un modello di machine learning.

5. Descrivere il processo di pulizia dei dati testuali, facendo riferimento alle varie operazioni necessarie in questa fase.

6. Descrivere il motivo per cui le tecniche di data balancing sono talvolta necessarie. Fornire inoltre una panoramica delle tecniche di data balancing, dettagliando le differenze tra metodi di oversampling e undersampling.

Part X

Problemi di Classificazione

**In grassetto sono riportate le risposte corrette.*

1. Si completi la seguente frase. “L’information gain [...]”
 - A. misura il grado di decadimento delle prestazioni di un modello di machine learning.
 - B. misura il grado di purezza di un attributo, ovvero quanto un certo attributo sarà in grado di dividere adeguatamente il dataset.**
 - C. misura il grado di data leakage nel processo di validazione di un modello di machine learning.
 - D. misura la probabilità che una caratteristica del machine learning sia utilizzata da un algoritmo bayesiano per processare le sue decisioni.
2. Quale delle seguenti meglio riflette la definizione di classificazione?
 - A. Task in cui l’obiettivo è predire il valore di una variabile numerica, chiamata variabile dipendente o di risposta, tramite l’utilizzo di un training set, ovvero un insieme di osservazioni per cui la variabile dipendente è nota.
 - B. Task in cui l’obiettivo è raggruppare oggetti in gruppi che abbiano un certo grado di omogeneità ma che, al tempo stesso, abbiano un certo grado di eterogeneità rispetto agli altri gruppi.
 - C. Task in cui l’obiettivo è predire il valore di una variabile categorica, chiamata variabile dipendente, target, o classe, tramite l’utilizzo di un training set, ovvero un insieme di osservazioni per cui la variabile target è nota.**
 - D. Nessuna delle precedenti.
3. Quale delle seguenti affermazioni, riferite agli alberi decisionali, è vera?
 - A. Sono algoritmi di apprendimento che appartengono alla categoria degli algoritmi probabilistici.
 - B. Sono algoritmi di apprendimento che mirano a creare un albero i cui nodi rappresentano delle decisioni e i cui archi rappresentano un sotto-insieme di caratteristiche del problema.
 - C. Sono algoritmi di apprendimento che mirano a raggruppare dati in maniera tale da formare gruppi coesi e distanti da altri gruppi.
 - D. Sono algoritmi di apprendimento che mirano a creare un albero i cui nodi rappresentano un sotto-insieme di caratteristiche del problema e i cui archi rappresentano delle decisioni.**
4. Descrivere le principali differenze tra algoritmi di apprendimento probabilistici e basati su entropia.

5. Fornire una descrizione dei passi necessari alla costruzione di un albero di decisione.

6. Descrivere il processo di validazione di un modello di machine learning, con particolare riferimento a come evitare il fenomeno del data leakage.

Giocare	Meteo	Temperatura	Umidità
NO	Soleggiato	Caldo	Elevata
NO	Soleggiato	Caldo	Elevata
SI	Nuvoloso	Caldo	Elevata
SI	Piovoso	Mite	Elevata
SI	Piovoso	Freddo	Normale
NO	Piovoso	Freddo	Normale
SI	Nuvoloso	Freddo	Normale
NO	Soleggiato	Mite	Elevata
SI	Soleggiato	Freddo	Normale
SI	Piovoso	Mite	Normale
SI	Soleggiato	Mite	Normale
SI	Nuvoloso	Mite	Elevata
SI	Nuvoloso	Caldo	Normale
NO	Piovoso	Mite	Elevata

Figure 2: Dataset di partenza.

7. Si consideri l'insieme di dati mostrato in Figura 2. Simulando il comportamento di un classificatore bayesiano, fornire una predizione della variabile dipendente "Giocare" per il seguente dato: "Meteo = Soleggiato; Temperatura = Caldo; Umidità = Elevata".
NB: E' necessario mostrare i passi che portano alla predizione.

8. Descrivere le principali metriche di validazione di modelli di machine learning, facendo particolare riferimento ai problemi potenziali che la metrica di accuratezza può causare per l'interpretazione dei risultati.

Part XI

Problemi di Regressione

**In grassetto sono riportate le risposte corrette.*

1. Quale delle seguenti meglio riflette la definizione di regressione?
 - A. **Task in cui l'obiettivo è predire il valore di una variabile numerica, chiamata variabile dipendente o di risposta, tramite l'utilizzo di un training set, ovvero un insieme di osservazioni per cui la variabile dipendente è nota.**
 - B. Task in cui l'obiettivo è raggruppare oggetti in gruppi che abbiano un certo grado di omogeneità ma che, al tempo stesso, abbiano un certo grado di eterogeneità rispetto agli altri gruppi.
 - C. Task in cui l'obiettivo è predire il valore di una variabile categorica, chiamata variabile dipendente, target, o classe, tramite l'utilizzo di un training set, ovvero un insieme di osservazioni per cui la variabile target è nota.
 - D. Nessuna delle precedenti.
2. Quale delle seguenti NON è un'assunzione fatta dalla regressione lineare?
 - A. Linearità dei dati.
 - B. Normalità dei residui.
 - C. Omoschedasticità
 - D. **Nessuna delle precedenti.**
3. Si completi la seguente frase. "La regressione multipla [...]"
 - A. [...] serve a predire il valore della variabile dipendente sulla base di un'unica variabile indipendente.
 - B. [...] assume la non linearità dei dati.
 - C. [...] mira a stimare le classi di appartenenza sulla base delle etichette a disposizione.
 - D. [...] **serve a predire il valore della variabile dipendente sulla base di un più variabili indipendenti.**
4. Descrivere i passi necessari per l'esecuzione dell'approccio di predizione basato sui minimi quadrati.

5. Fornire una descrizione delle assunzioni di base richieste per l'applicazione della regressione lineare.

6. Fornire una descrizione delle metriche di valutazione per gli algoritmi di regressione lineare.

Part XII

Problemi di Clustering

**In grassetto sono riportate le risposte corrette.*

1. Quale delle seguenti meglio riflette la definizione di clustering?
 - A. Task in cui l'obiettivo è predire il valore di una variabile numerica, chiamata variabile dipendente o di risposta, tramite l'utilizzo di un training set, ovvero un insieme di osservazioni per cui la variabile dipendente è nota.
 - B. **Task in cui l'obiettivo è raggruppare oggetti in gruppi che abbiano un certo grado di omogeneità ma che, al tempo stesso, abbiamo un certo grado di eterogeneità rispetto agli altri gruppi.**
 - C. Task in cui l'obiettivo è predire il valore di una variabile categorica, chiamata variabile dipendente, target, o classe, tramite l'utilizzo di un training set, ovvero un insieme di osservazioni per cui la variabile target è nota.
 - D. Nessuna delle precedenti.
2. Si completi la seguente frase. "Il clustering gerarchico agglomerativo [...]"
 - A. [...] procede nel raggruppamento degli elementi con l'obiettivo di minimizzare l'errore quadratico rispetto al valore medio dei cluster.
 - B. [...] procede nel raggruppamento degli elementi partendo da un unico cluster fino ad arrivare al punto in cui esistono N cluster, uno per ogni elemento del problema.
 - C. **[...] procede nel raggruppamento degli elementi partendo da cluster atomici fino ad arrivare al punto in cui esiste un unico cluster.**
 - D. [...] procede nel raggruppamento degli elementi combinando più misure di distanza e criteri di ottimizzazione, così da bilanciare il numero di cluster generati.
3. Quale delle seguenti affermazioni, riferite alle metriche di valutazione del clustering, è *vera*?
 - A. Il punto di gomito rappresenta una misura di coesione e separazione tra i dati nei cluster generati da un algoritmo.
 - B. Il coefficiente di forma rappresenta i valori candidati di un parametro k rispetto alla somma degli errori quadratici ottenuti dall'algoritmo configurato per generare k cluster.
 - C. La MoJo distanza misura il numero minimo di operazioni di spostamento ed eliminazione di cluster necessari per passare da una partizione ad un'altra.
 - D. **Nessuna delle precedenti.**
4. Discutere le proprietà di una misura metrica.

5. Discutere le differenze principali tra algoritmi di clustering partizionali ed algoritmi di clustering gerarchici.

6. Fornire una descrizione dell'algoritmo k-means, con particolare riferimento alle potenziali problematiche che possono portare ad una riduzione delle sue prestazioni .
