Machine Learning - Prof. G. Polese, Prof.ssa L. Caruccio **Anno Accademico 2023/2024**

Prova in Itinere - Aula F4 (ore 11:00) **del 20/12/2023**

Nome e Cognome: **Matricola/Alias:**

(Scrivere solo nello spazio bianco. Se necessario, usare il retro del foglio. Non sono ammessi elaborati su fogli diversi.)

Si consideri uno scenario di classificazione binaria in cui in caso di attesa al ristorante si vuole predire se il cliente aspetterà il tavolo o meno, e il seguente training set:

Alternate	**Bar**	**Patrons**	**Price**	**Raining**	**Reservation**	**Estimation**	***WillWait***
True	False	Some	$$$	False	True	0-10	*Yes*
True	False		$	False	False	30-60	*No*
False	True	Some	$	False		0-10	*Yes*
True	False	Full		False	False	10-30	*Yes*
True	False	Full	$$$		True	>60	*No*
False	True	Some	$$	True	True		*Yes*
False	True	None	$	True	False	0-10	*No*
False	False	Some	$$	True	True	0-10	*Yes*
False	True	Full	$	True	False	>60	*No*
True	True	Full	$$$	False	True	10-30	*No*
False	False	None	$	False	False	0-10	*No*
True	True	Full	$	False	False	30-60	*Yes*

1. **Alternate**: se è disponibile un ristorante alternativo nelle vicinanze
2. **Bar**: se il ristorante ha a disposizione un'area bar confortevole per aspettare
3. **Patrons**: quante persone sono presenti nel ristorante (possibili valori: *None, Some, Full*)
4. **Price**: il range di prezzo del ristorante (possibili valori: *$, $$, $$$*)
5. **Raining**: se sta piovendo
6. **Reservation**: se il tavolo è stato prenotato
7. **Estimation**: il tempo di attesa previsto (possibili valori: *0-10 minuti, 10-30, 30-60, >60*)

Esercizio 1 (punti 5 su 30)

Determinare i valori mancanti nei vari attributi applicando l'imputazione per *moda*.

Machine Learning - Prof. G. Polese, Prof.ssa L. Caruccio **Anno Accademico 2023/2024**

Prova in Itinere - Aula F4 (ore 11:00) **del 20/12/2023**

Esercizio 2 (punti 8 su 30)

Considerando il dataset imputato dell'Esercizio 1, definire il miglior criterio di split sull'attributo **Patrons** che generi due raggruppamenti, utilizzando la metrica di *Gini Impurity* come metrica di valutazione. (N.B.: Mostrare le valutazioni anche mediante le tabelle con i gruppi)

$$Gini\ Impurity = 1 - \sum_{k=1}^{K} {f_k}^2$$

Machine Learning - Prof. G. Polese, Prof.ssa L. Caruccio **Anno Accademico 2023/2024**

Prova in Itinere - Aula F4 (ore 11:00) **del 20/12/2023**

Esercizio 3 (punti 6 su 30)

Per lo stesso scenario di classificazione assumiamo che siano stati riportati i seguenti risultati su un'analisi effettuata secondo il tipo di pasto servito al momento della raccolta dei dati.

	LUNCH	DINNER	TOTAL
Yes	**98**	**510**	**608**
No	**182**	**275**	**457**
TOTAL	**280**	**785**	**1065**

Per cui, supponendo di sapere che un cliente NON sia disposto ad attendere per il tavolo, qual è la probabilità che questo sia avvenuto a cena (DINNER)? (N.B.: Applicare il Teorema di Bayes per ottenere la soluzione)

Machine Learning - Prof. G. Polese, Prof.ssa L. Caruccio **Anno Accademico 2023/2024**

Prova in Itinere - Aula F4 (ore 11:00) **del 20/12/2023**

Esercizio 4 (punti 6 su 30)

Supponiamo di aver utilizzato due differenti classificatori binari per effettuare le predizioni, valutare quale dei due ha ottenuto migliori performance sulla base delle metriche di Recall, Precision, Accuracy ed F1-Score e delle seguenti confusion matrix:

Classificatore 1

	ACTUAL *Yes*	**ACTUAL *No***
PREDICTED *Yes*	**90**	**38**
PREDICTED *No*	**27**	**45**

Classificatore 2

	ACTUAL *Yes*	**ACTUAL *No***
PREDICTED *Yes*	**66**	**55**
PREDICTED *No*	**15**	**64**

Esercizio 5 (punti 5 su 30)

Descrivere in generale il task di classificazione. Specificare le differenti categorie di questo task e indicare qualche esempio della loro applicazione nei problemi del mondo reale.