

LavorAI

Francesco Scarati MAT. 693181	Isabell Paradiso MAT. 685685
--	---

Indice

Sommario

1. Introduzione	2
2. Intelligenze Artificiali	2
2.1 Classificatore KNN	2
2.2 Gaussian Naive Bayes	2
2.3 Decision Tree.....	2
3. Valutazione	3
3.1 K-Nearest Neighbors	3
3.2 Gaussian NB	4
3.3 Decision Tree (nessun limite d'altezza)	5
3.4 Decision Tree (limite altezza = 3).....	6
3.5 Decision Tree (limite altezza = 8).....	7
4. Conclusioni.....	8

1. Introduzione

LavorAI è un progetto sviluppato per l'esame di Ingegneria della Conoscenza all'Università degli Studi Aldo Moro di Bari. Il suo scopo è quello di profilare degli individui che cercano lavoro e indirizzarli verso la giusta direzione.

Tre diversi modelli vengono utilizzati (di cui uno ha 3 varianti), ognuno di loro si allena su un dataset di questionari ed è successivamente in grado di dare risposte ben ottimizzate riguardo ogni questionario.

2. Intelligenze Artificiali

LavorAI dispone di 3 diversi modelli di IA in grado di profilare l'utente

2.1 Classificatore KNN

Il k-nearest neighbors (KNN) è un algoritmo utilizzato nel riconoscimento di pattern per la classificazione di oggetti basandosi sulle caratteristiche di ciascun oggetto vicino a quello considerato. L'input è costituito dai k esempi di addestramento che rappresentano lo stato del mondo.

Nella classificazione k-NN, l'output è un'appartenenza a una classe. Un oggetto è classificato da un voto di pluralità dei suoi vicini, con l'oggetto assegnato alla classe più comune tra i suoi k vicini (k è un numero intero positivo, tipicamente piccolo). Se $k = 1$, l'oggetto viene semplicemente assegnato alla classe di quel singolo vicino più prossimo.

Nella regressione k-NN, l'output è il valore della proprietà per l'oggetto. Questo valore è la media dei valori di k vicini più prossimi.

2.2 Gaussian Naive Bayes

I classificatori Naive Bayes sono una famiglia di semplici "classificatori probabilistici" basati sull'applicazione del teorema di Bayes con ipotesi di indipendenza forti (ingenue) tra le caratteristiche. Sono tra i più semplici modelli di rete bayesiana, ma, insieme alla stima della densità del kernel, possono raggiungere livelli di precisione più elevati.

I classificatori NB sono altamente scalabili e richiedono un numero di parametri lineari nel numero di variabili (caratteristiche/predittori) in un problema di apprendimento. L'addestramento alla massima verosimiglianza può essere effettuato valutando un'espressione in forma chiusa, che richiede tempo lineare, piuttosto che mediante costose approssimazioni iterative come quelle usate per molti altri tipi di classificatori.

Nella letteratura statistica e informatica, i modelli NB sono conosciuti con una varietà di nomi, inclusi Bayes semplice e Bayes indipendente. Tutti questi nomi fanno riferimento all'uso del teorema di Bayes nella regola decisionale del classificatore, ma l'NB non è (necessariamente) un metodo bayesiano.

2.3 Decision Tree

Un albero decisionale è uno strumento di supporto decisionale che utilizza un modello di decisioni simile ad un albero e le loro possibili conseguenze, inclusi i risultati di eventi casuali, i costi delle risorse e l'utilità. È un modo per visualizzare un algoritmo che contiene solo istruzioni di controllo condizionali.

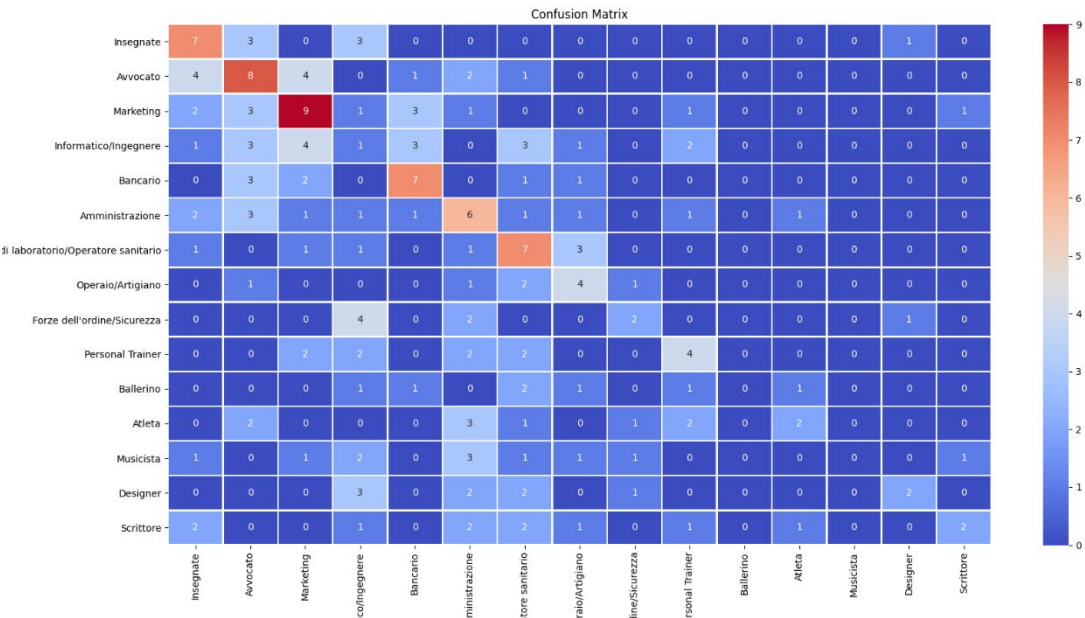
Gli alberi decisionali sono comunemente usati nella ricerca operativa, in particolare nell'analisi decisionale, per aiutare a identificare una strategia che ha maggiori probabilità di raggiungere un obiettivo, ma sono anche uno strumento popolare nell'apprendimento automatico.

3. Valutazione

Valutiamo ora le prestazioni di ogni modello utilizzato.

3.1 K-Nearest Neighbors

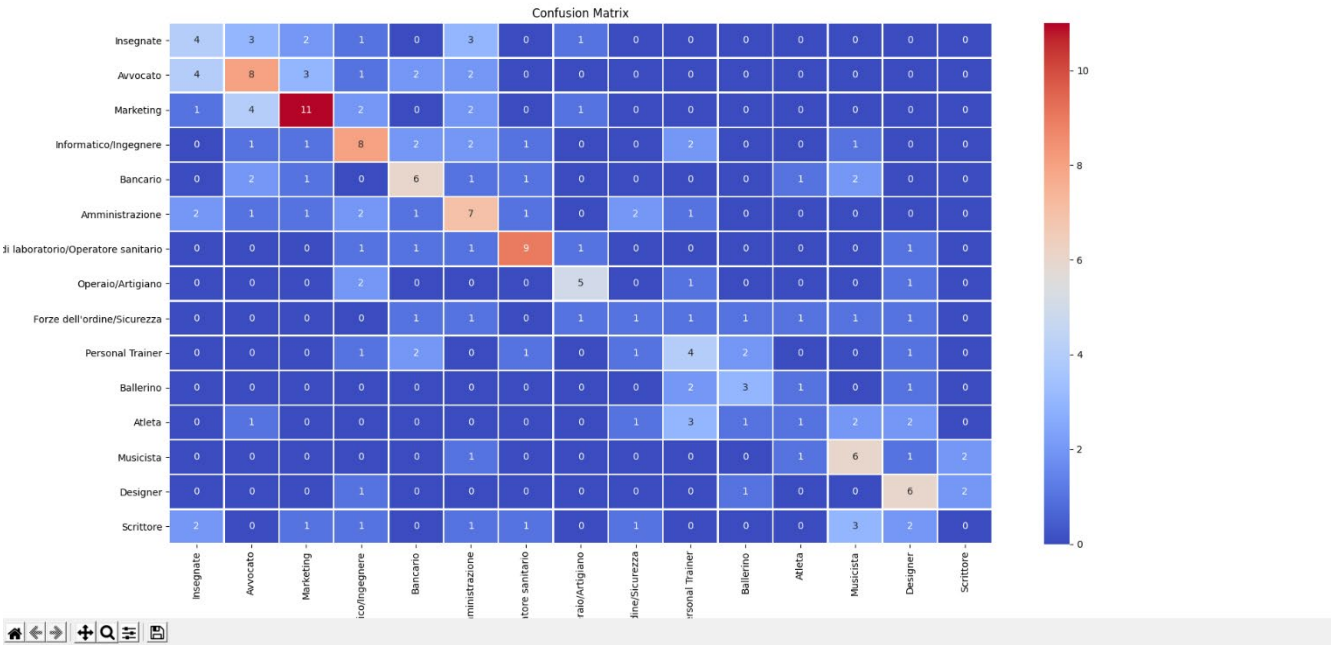
		precision	recall	f1-score	support
	Insegnate	0.35	0.50	0.41	14
	Avvocato	0.31	0.40	0.35	20
	Marketing	0.38	0.43	0.40	21
	Informatico/Ingegnere	0.05	0.06	0.05	18
	Bancario	0.44	0.50	0.47	14
	Amministrazione	0.24	0.33	0.28	18
Scienziato di laboratorio/Operatore sanitario		0.28	0.50	0.36	14
	Operaio/Artigiano	0.31	0.44	0.36	9
	Forze dell'ordine/Sicurezza	0.33	0.22	0.27	9
	Personal Trainer	0.33	0.33	0.33	12
	Ballerino	0.00	0.00	0.00	7
	Atleta	0.40	0.18	0.25	11
	Musicista	0.00	0.00	0.00	11
	Designer	0.50	0.20	0.29	10
	Scrittore	0.50	0.17	0.25	12
	accuracy			0.30	200
	macro avg	0.29	0.28	0.27	200
	weighted avg	0.30	0.30	0.28	200



3.2 Gaussian NB

		precision	recall	f1-score	support
	Insegnate	0.31	0.29	0.30	14
	Avvocato	0.40	0.40	0.40	20
	Marketing	0.55	0.52	0.54	21
	Informatico/Ingegnere	0.40	0.44	0.42	18
	Bancario	0.40	0.43	0.41	14
	Amministrazione	0.33	0.39	0.36	18
Scienziato di laboratorio/Operatore sanitario		0.64	0.64	0.64	14
	Operaio/Artigiano	0.56	0.56	0.56	9
	Forze dell'ordine/Sicurezza	0.17	0.11	0.13	9
	Personal Trainer	0.29	0.33	0.31	12
	Ballerino	0.38	0.43	0.40	7
	Atleta	0.20	0.09	0.13	11
	Musicista	0.40	0.55	0.46	11
	Designer	0.38	0.60	0.46	10
	Scrittore	0.00	0.00	0.00	12
	accuracy			0.40	200
	macro avg	0.36	0.39	0.37	200
	weighted avg	0.37	0.40	0.38	200

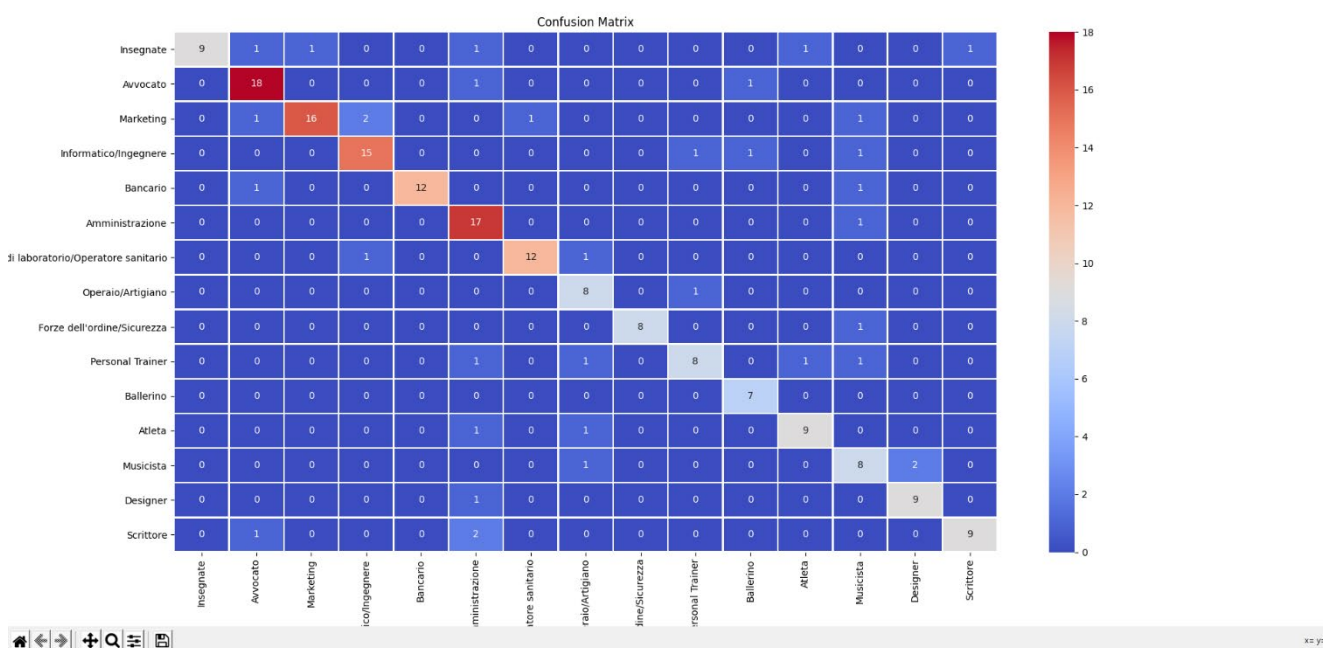
Figure 1



3.3 Decision Tree (nessun limite d'altezza)

		precision	recall	f1-score	support
	Insegnante	1.00	0.64	0.78	14
	Avvocato	0.82	0.90	0.86	20
	Marketing	0.94	0.76	0.84	21
	Informatico/Ingegnere	0.83	0.83	0.83	18
	Bancario	1.00	0.86	0.92	14
	Amministrazione	0.71	0.94	0.81	18
Scienziato di laboratorio/Operatore sanitario		0.92	0.86	0.89	14
	Operaio/Artigiano	0.67	0.89	0.76	9
	Forze dell'ordine/Sicurezza	1.00	0.89	0.94	9
	Personal Trainer	0.80	0.67	0.73	12
	Ballerino	0.78	1.00	0.88	7
	Atleta	0.82	0.82	0.82	11
	Musicista	0.57	0.73	0.64	11
	Designer	0.82	0.90	0.86	10
	Scrittore	0.90	0.75	0.82	12
	accuracy			0.82	200
	macro avg	0.84	0.83	0.83	200
	weighted avg	0.85	0.82	0.83	200

Figure 1

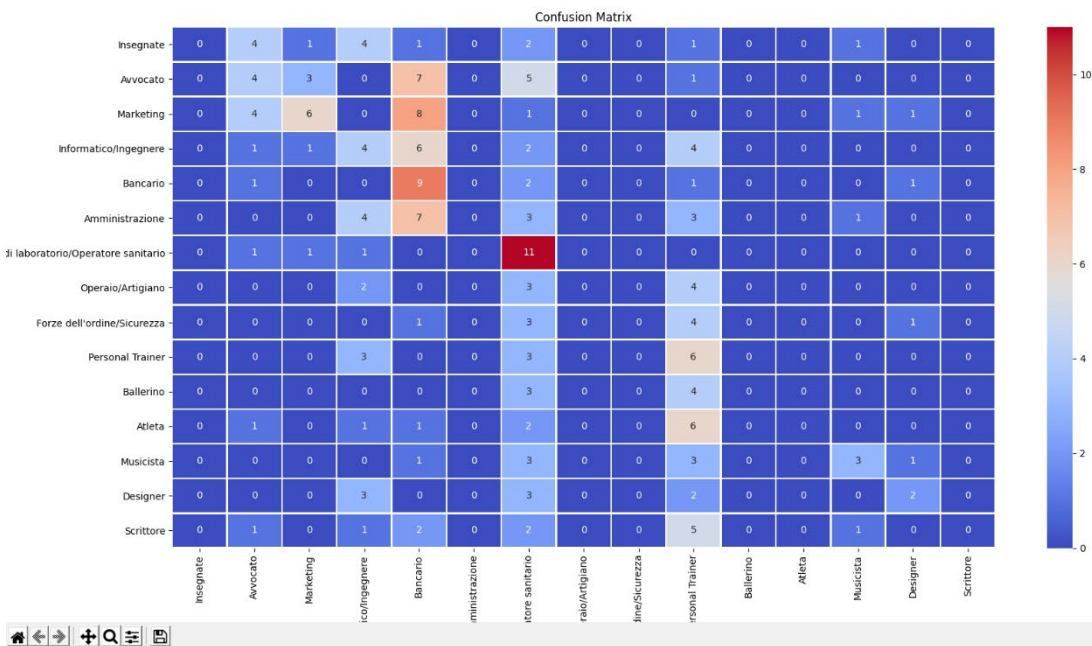


3.4 Decision Tree (limite altezza = 3)

	precision	recall	f1-score	support
Insegnate	0.00	0.00	0.00	14
Avvocato	0.24	0.20	0.22	20
Marketing	0.50	0.29	0.36	21
Informatico/Ingegnere	0.17	0.22	0.20	18
Bancario	0.21	0.64	0.32	14
Amministrazione	0.00	0.00	0.00	18
Scienziato di laboratorio/Operatore sanitario	0.23	0.79	0.35	14
Operaio/Artigiano	0.00	0.00	0.00	9
Forze dell'ordine/Sicurezza	0.00	0.00	0.00	9
Personal Trainer	0.14	0.50	0.21	12
Ballerino	0.00	0.00	0.00	7
Atleta	0.00	0.00	0.00	11
Musicista	0.43	0.27	0.33	11
Designer	0.33	0.20	0.25	10
Scrittore	0.00	0.00	0.00	12
accuracy			0.23	200
macro avg	0.15	0.21	0.15	200
weighted avg	0.17	0.23	0.17	200

Figure 1

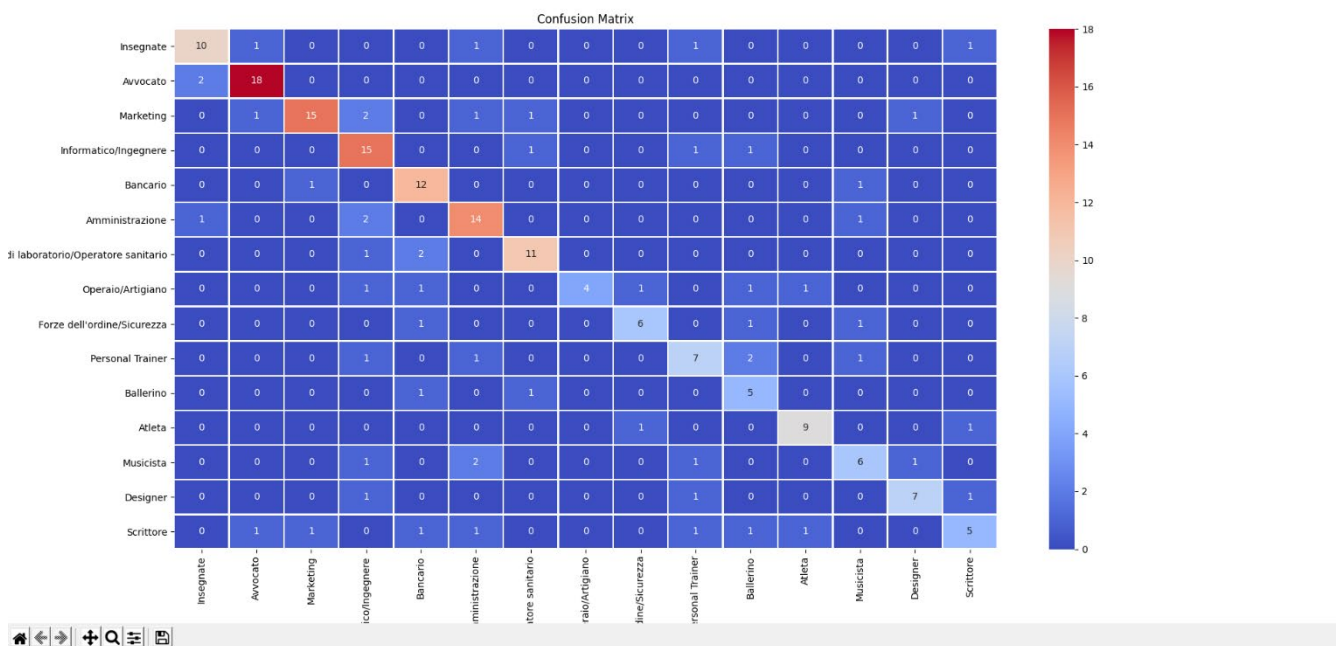
— 0 x



3.5 Decision Tree (limite altezza = 8)

	precision	recall	f1-score	support
Insegnante	0.77	0.71	0.74	14
Avvocato	0.86	0.90	0.88	20
Marketing	0.88	0.71	0.79	21
Informatico/Ingegnere	0.62	0.83	0.71	18
Bancario	0.67	0.86	0.75	14
Amministrazione	0.70	0.78	0.74	18
Scienziato di laboratorio/Operatore sanitario	0.79	0.79	0.79	14
Operaio/Artigiano	1.00	0.44	0.62	9
Forze dell'ordine/Sicurezza	0.75	0.67	0.71	9
Personal Trainer	0.58	0.58	0.58	12
Ballerino	0.45	0.71	0.56	7
Atleta	0.82	0.82	0.82	11
Musicista	0.60	0.55	0.57	11
Designer	0.78	0.70	0.74	10
Scrittore	0.62	0.42	0.50	12
accuracy			0.72	200
macro avg	0.73	0.70	0.70	200
weighted avg	0.74	0.72	0.72	200

Figure 1



4. Conclusioni

Possiamo infine concludere che il modello di albero decisionale senza limite d'altezza è il più efficace, anche in presenza di dataset sbilanciato.