

APPROFONDIMENTO

ANALISI DELLA QUALITÀ DEL VINO

PRINCIPI E MODELLI DELLA PERCEZIONE

CARTANESE ALESSANDRO E ROTA GABRIELE

DESCRIZIONE DATASET

DESCRIZIONE DATASET

UNIONE DI DUE DATASET

- VINI BIANCHI E VINI ROSSI

DIMENSIONE DEL DATASET

- 6497 CAMPIONI TOTALI

TARGET

- COLONNA QUALITY

STATO

- DATASET PULITO, NESSUN VALORE NULLO

VARIABILI INDIPENDENTI (INPUT)

- ACIDITÀ FISSA
- ACIDITÀ VOLATILE
- ACIDO CITRICO
- ZUCCHERI RESIDUI
- CLORURI
- ANIDRIDE SOLFOROSA LIBERA
- ANIDRIDE SOLFOROSA TOTALE
- DENSITÀ
- PH
- SOLFATI
- ALCOOL

DESCRIZIONE DATASET

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
mean	7.2	0.3	0.3	5.4	0.1	30.5	115.7	1.0	3.2	0.5	10.5	5.8
std	1.3	0.2	0.1	4.8	0.0	17.7	56.5	0.0	0.2	0.1	1.2	0.9
min	3.8	0.1	0	0.6	0.0	1.0	6.0	1.0	2.7	0.2	8.0	3.0
25%	6.4	0.2	0.2	1.8	0.0	17.0	77.0	1.0	3.1	0.4	9.5	5.0
50%	7.0	0.3	0.3	3.0	0.0	29.0	118.0	1.0	3.2	0.6	10.3	6.0
75%	7.7	0.4	0.4	8.1	0.1	41.0	156.0	1.0	3.3	0.6	11.3	6.0
max	15.9	1.6	1.6	65.8	0.6	289.0	440.0	1.0	4.0	2.0	14.9	9.0

DESCRIZIONE DATASET

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
mean	7.2	0.3	0.3	5.4	0.1	30.5	115.7	1.0	3.2	0.5	10.5	5.8
std	1.3	0.2	0.1	4.8	0.0	17.7	56.5	0.0	0.2	0.1	1.2	0.9
min	3.8	0.1	0	0.6	0.0	1.0	6.0	1.0	2.7	0.2	8.0	3.0
25%	6.4	0.2	0.2	1.8	0.0	17.0	77.0	1.0	3.1	0.4	9.5	5.0
50%	7.0	0.3	0.3	3.0	0.0	29.0	118.0	1.0	3.2	0.6	10.3	6.0
75%	7.7	0.4	0.4	8.1	0.1	41.0	156.0	1.0	3.3	0.6	11.3	6.0
max	15.9	1.6	1.6	65.8	0.6	289.0	440.0	1.0	4.0	2.0	14.9	9.0

DESCRIZIONE DATASET

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
mean	7.2	0.3	0.3	5.4	0.1	30.5	115.7	1.0	3.2	0.5	10.5	5.8
std	1.3	0.2	0.1	4.8	0.0	17.7	56.5	0.0	0.2	0.1	1.2	0.9
min	3.8	0.1	0	0.6	0.0	1.0	6.0	1.0	2.7	0.2	8.0	3.0
25%	6.4	0.2	0.2	1.8	0.0	17.0	77.0	1.0	3.1	0.4	9.5	5.0
50%	7.0	0.3	0.3	3.0	0.0	29.0	118.0	1.0	3.2	0.6	10.3	6.0
75%	7.7	0.4	0.4	8.1	0.1	41.0	156.0	1.0	3.3	0.6	11.3	6.0
max	15.9	1.6	1.6	65.8	0.6	289.0	440.0	1.0	4.0	2.0	14.9	9.0

DESCRIZIONE DATASET

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
mean	7.2	0.3	0.3	5.4	0.1	30.5	115.7	1.0	3.2	0.5	10.5	5.8
std	1.3	0.2	0.1	4.8	0.0	17.7	56.5	0.0	0.2	0.1	1.2	0.9
min	3.8	0.1	0	0.6	0.0	1.0	6.0	1.0	2.7	0.2	8.0	3.0
25%	6.4	0.2	0.2	1.8	0.0	17.0	77.0	1.0	3.1	0.4	9.5	5.0
50%	7.0	0.3	0.3	3.0	0.0	29.0	118.0	1.0	3.2	0.6	10.3	6.0
75%	7.7	0.4	0.4	8.1	0.1	41.0	156.0	1.0	3.3	0.6	11.3	6.0
max	15.9	1.6	1.6	65.8	0.6	289.0	440.0	1.0	4.0	2.0	14.9	9.0

The background consists of several wine bottles standing upright in a dark room. The bottles are partially visible, with their labels and necks catching some light, creating a warm, golden glow against the dark background.

**ANALISI
ESPLORATIVA**

BOX PLOT

SPIEGAZIONE GRAFICO

Sull'asse orizzontale (X) abbiamo i voti di qualità, dal peggiore (3) al migliore (9).

Sull'asse verticale (Y) c'è la percentuale di alcool. Ogni 'scatola' colorata rappresenta la distribuzione dei vini che hanno preso quel determinato voto.

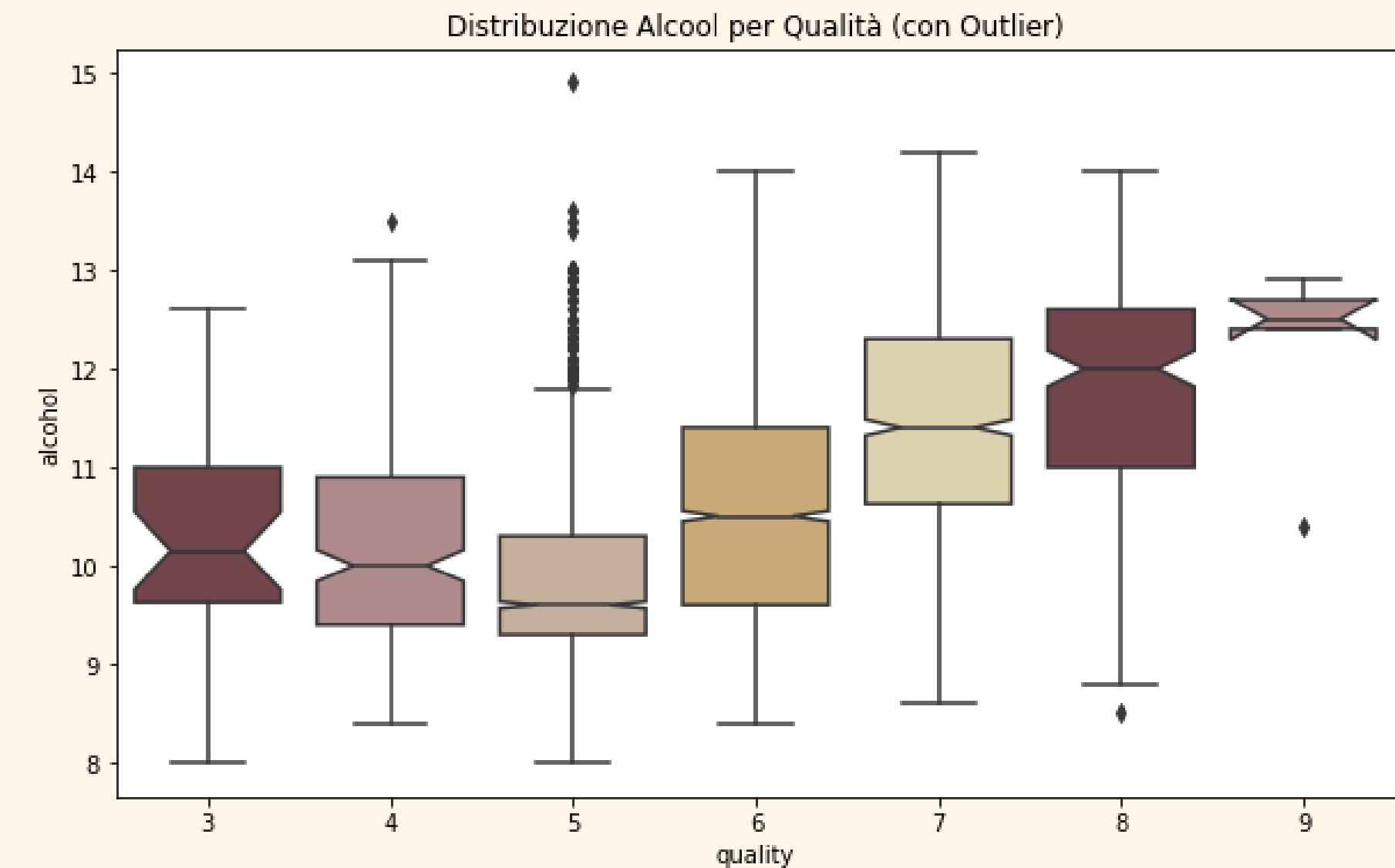
La linea nera al centro della scatola è la mediana.

COSA NOTIAMO IMMEDIATAMENTE?

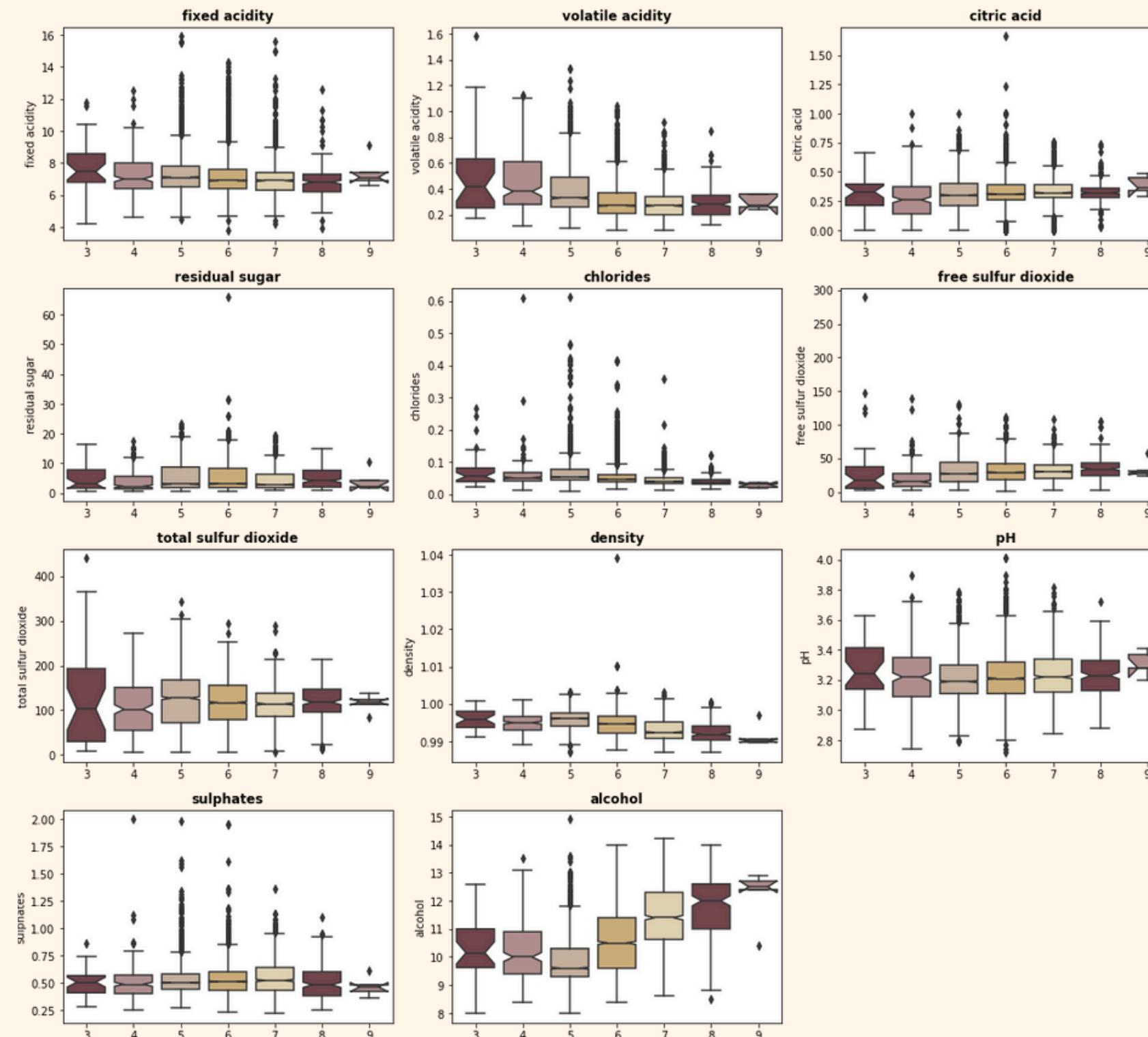
C'è un trend chiaro e crescente.

Se guardate i vini mediocri (voti 3, 4, 5), la linea dell'alcool è bassa.

Man mano che ci spostiamo verso l'eccellenza (voti 7, 8, 9), le scatole salgono.

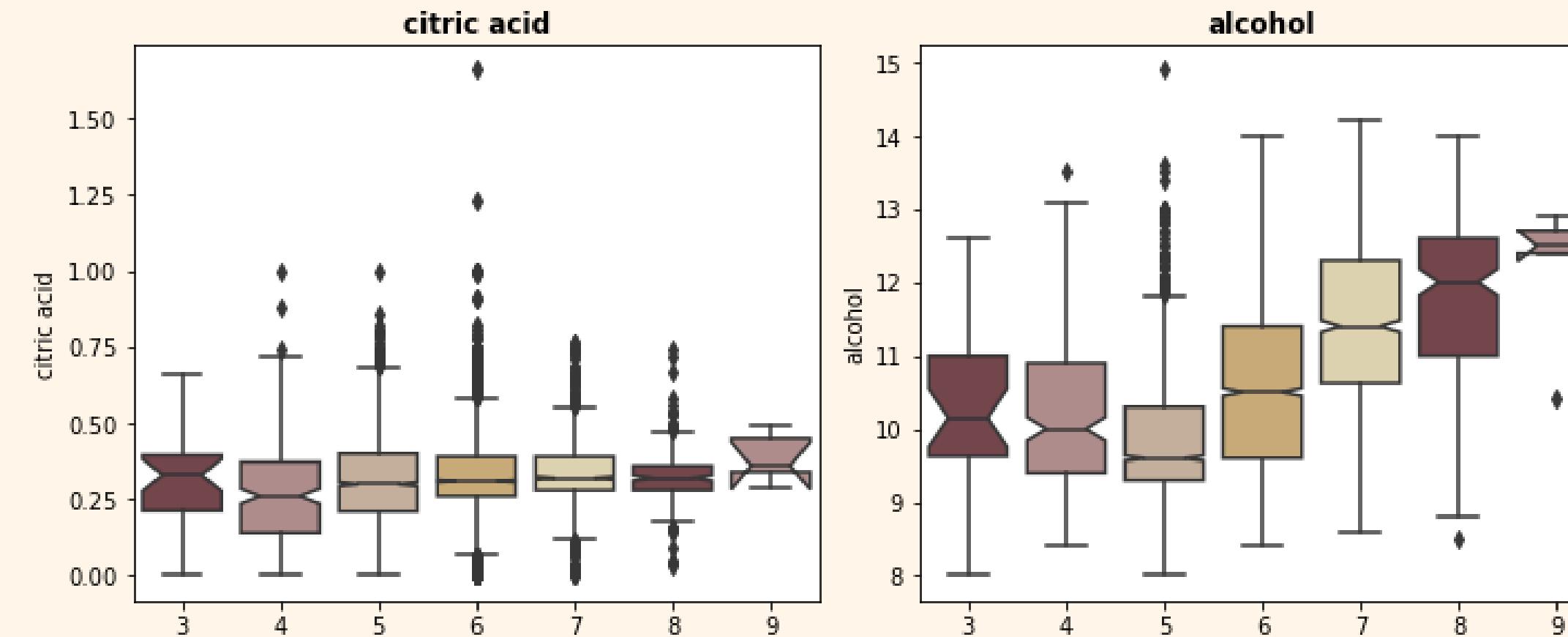


SMALL MULTIPLES



SMALL MULTIPLES

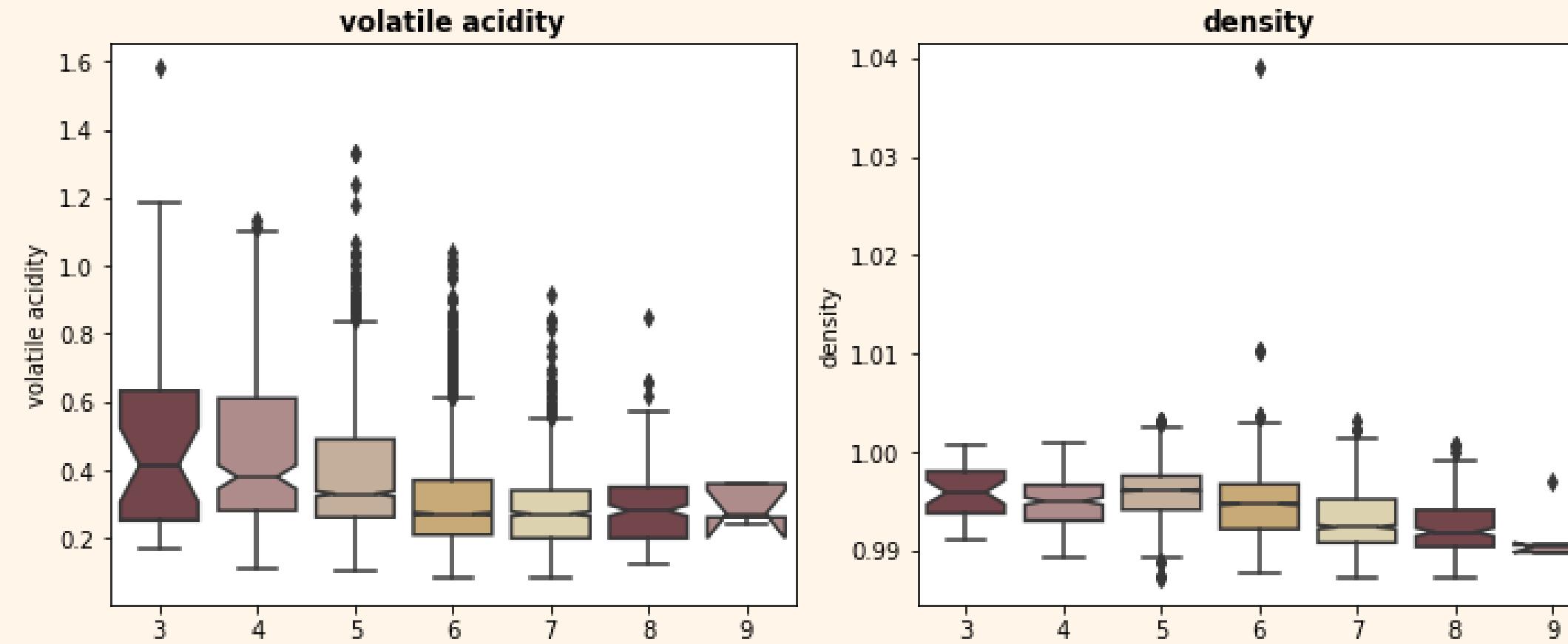
CORRELAZIONI POSITIVE



ACIDO CITRICO E ALCOOL TENDONO A SALIRE CON LA QUALITÀ

SMALL MULTIPLES

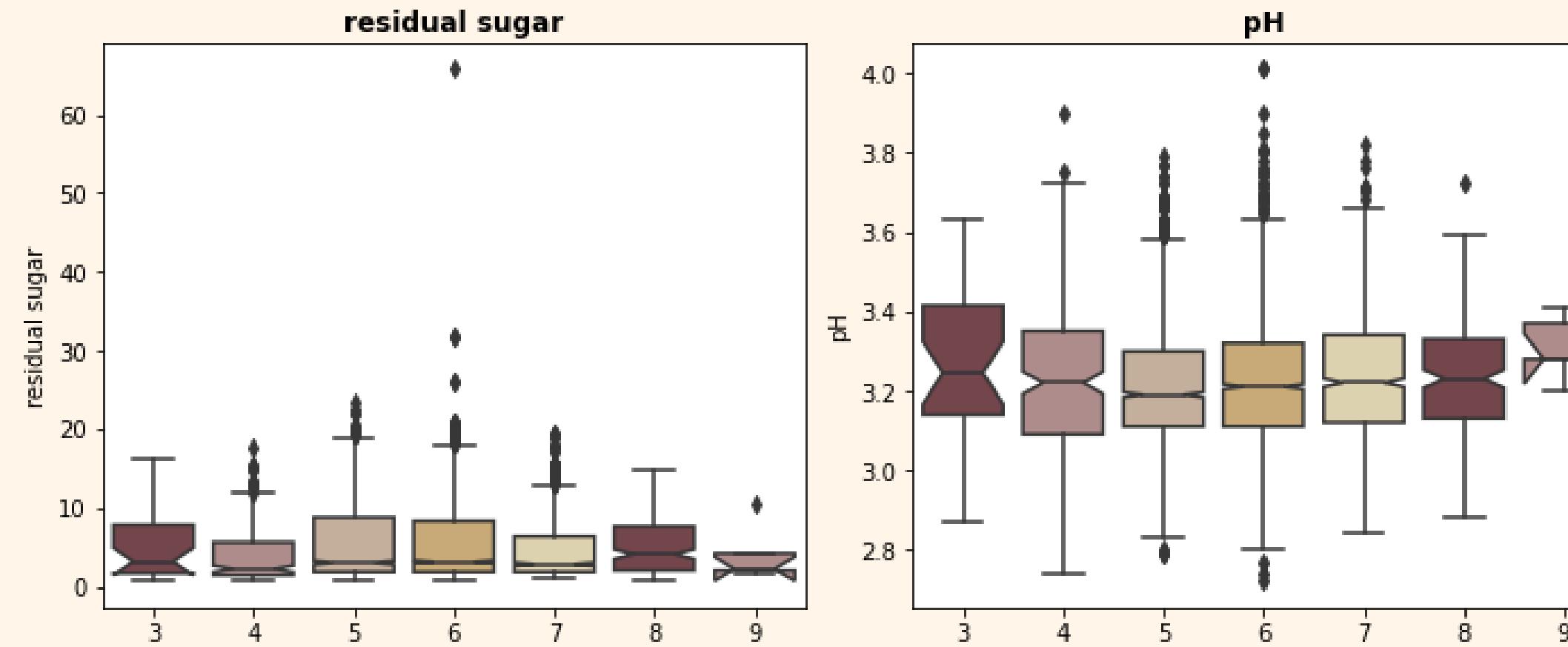
CORRELAZIONI NEGATIVE



ACIDO VOLATILE E DENSITÀ TENDONO A SCENDERE CON LA QUALITÀ

SMALL MULTIPLES

VARIABILI IRRILEVANTI



PH E ZUCCHERI RESIDUI NON MOSTRANO TREND CHIARI

The background of the image shows a row of wine bottles standing upright in a dark room. The bottles are arranged in a perspective that recedes towards the right. The lighting is dramatic, highlighting the necks and shoulders of the bottles against the dark background.

PRE-PROCESSING

PRE-PROCESSING

NORMALIZZAZIONE

Abbiamo usato StandardScaler per portare tutte le variabili sulla stessa scala (fondamentale perché variabili come "Zolfo" hanno valori alti, mentre "Cloruri" valori bassissimi).

STRATIFICAZIONE

Abbiamo diviso i dati in Training (75%) e Test (25%) usando stratify=y per mantenere le proporzioni delle classi di qualità (evitando che i vini rari spariscano dal test set)

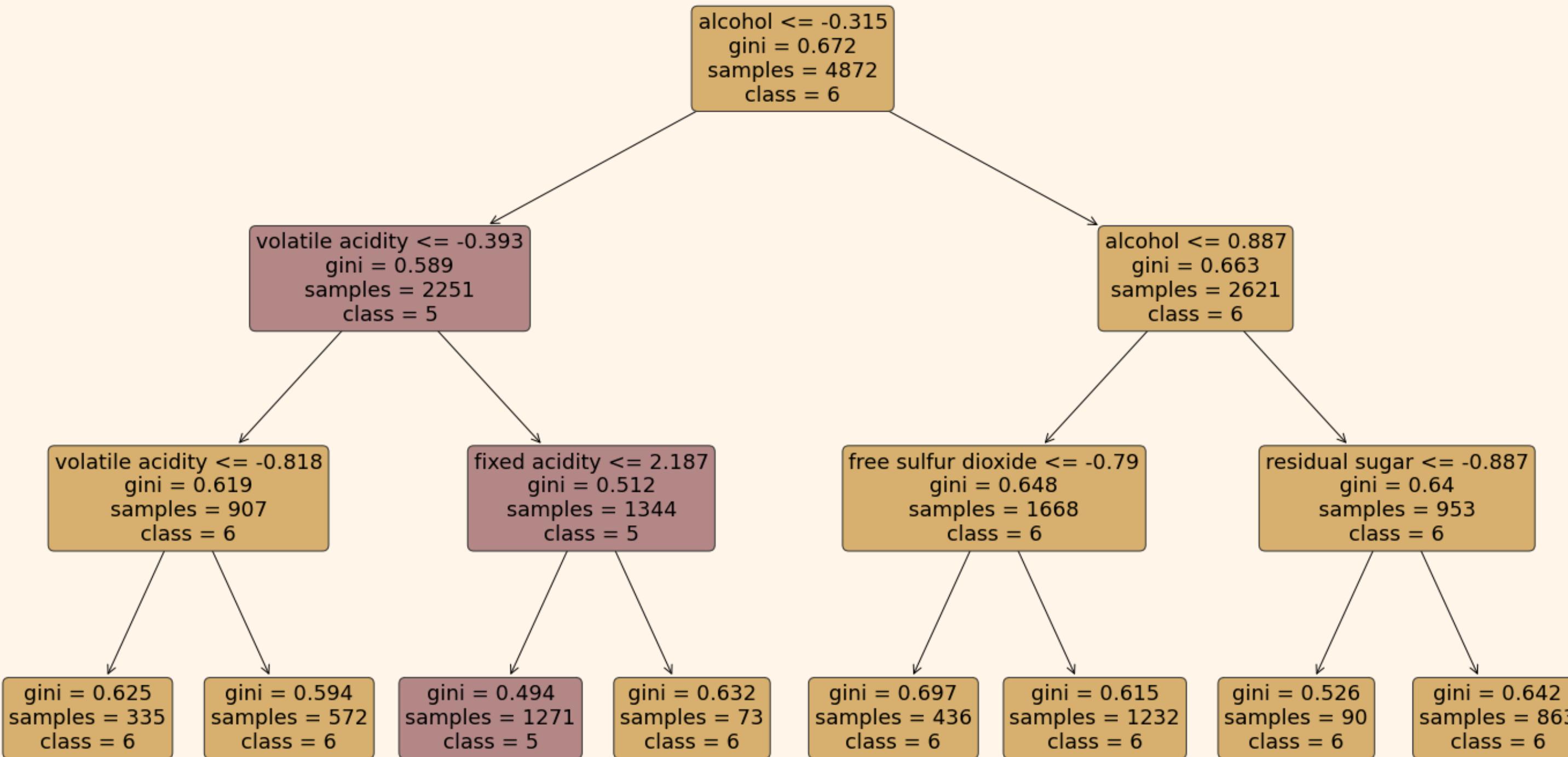
```
# Normalizzazione
# Separiamo l'input (X) dall'output (y)
X = df.drop('quality', axis=1)
y = df['quality']
scaler = StandardScaler()
X_scaled = pd.DataFrame(scaler.fit_transform(X), columns=X.columns)

# Stratificazione
X_train, X_test, y_train, y_test = train_test_split(
    X_scaled, y,
    test_size=0.25,
    stratify=y, # Questo parametro attiva la stratificazione
    random_state=42
)
print("Training set shape:", X_train.shape)
print("Test set shape:", X_test.shape)
```

The background consists of several wine bottles standing upright in a dark room. The bottles are partially illuminated from the side, creating a dramatic play of light and shadow. The labels on the bottles are mostly illegible, appearing as dark rectangles. The overall atmosphere is moody and sophisticated.

DECISION TREE

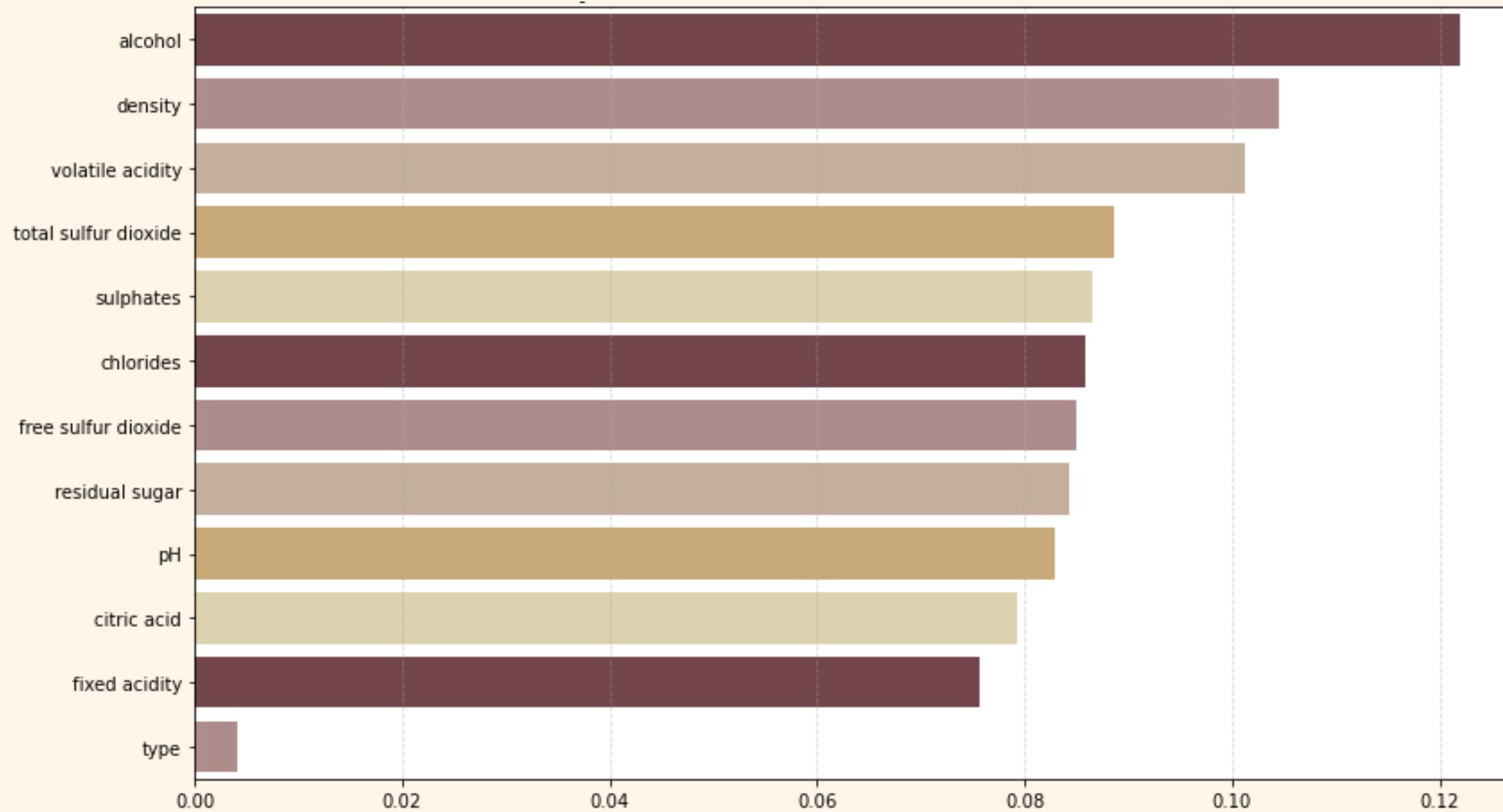
DECISION TREE



The background of the image shows a row of wine bottles standing upright in a dark, low-light environment. The bottles are arranged in a perspective that recedes towards the right side of the frame. The lighting highlights the necks and shoulders of the bottles, creating a dramatic effect against the dark background.

RANDOM FOREST

RANDOM FOREST



ENSEMBLE LEARNING

- Unione di 100 alberi decisionali.

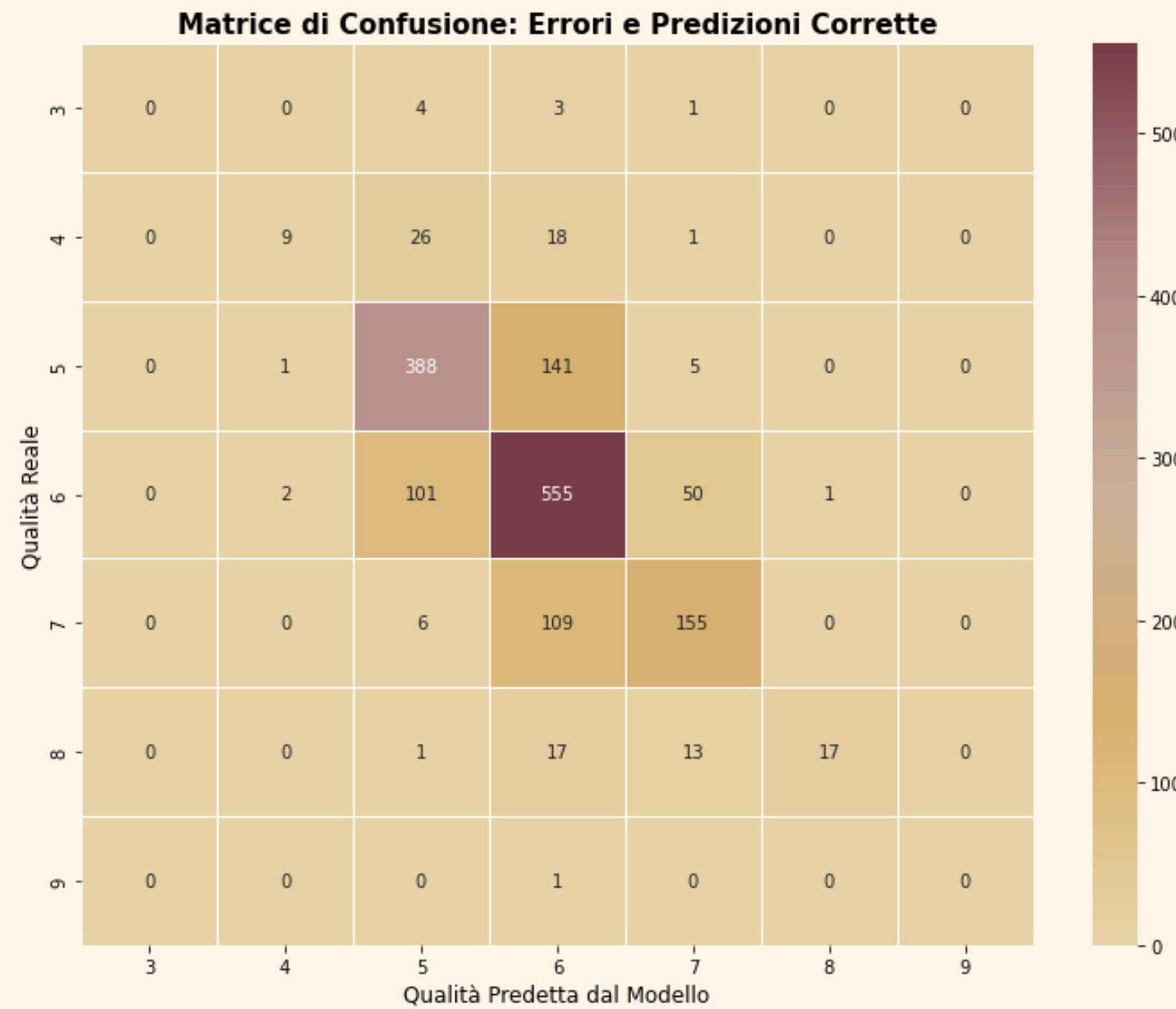
STABILITÀ

- Riduce il rischio di errori su casi specifici

PERFORMANCE

- Accuratezza sul Test Set superiore rispetto all'albero singolo

MATRICE DI CONFUSIONE



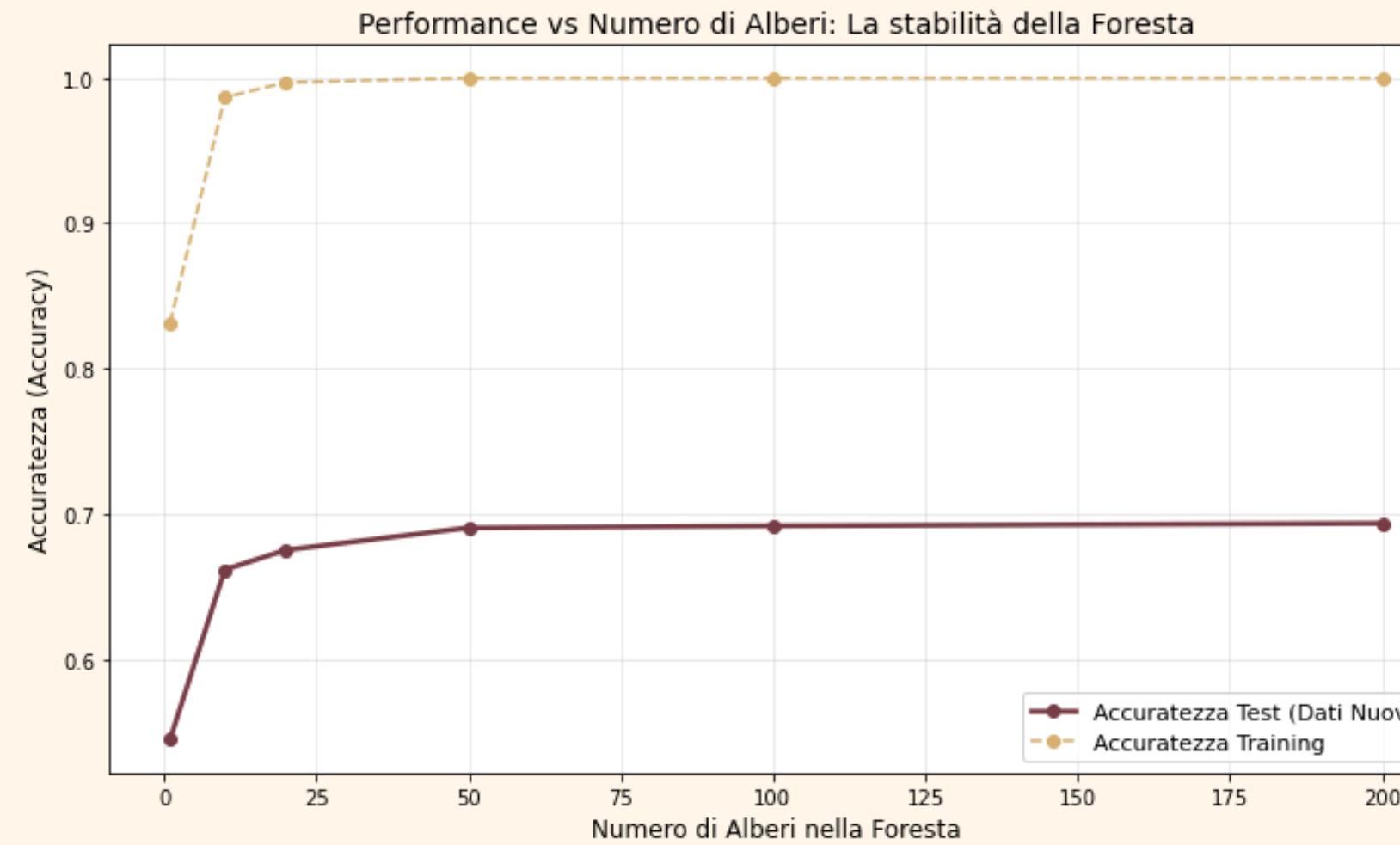
ERRORI VICINI

- Il modello confonde spesso classi adiacenti (es. 5 vs 6)

ERRORI GRAVI

- Sono quasi assenti (il modello non scambia mai un vino pessimo per uno eccellente).

CURVA DI APPRENDIMENTO



LINEA ORO

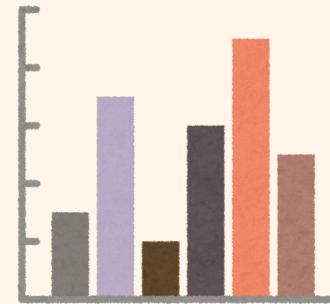
- Rappresenta quanto bene il modello impara dai dati che ha già visto.
“Schizza” subito al 100%, significa che il modello ha una memoria perfetta, ma questo non basta.

LINEA COLOR VINO

- Ci dice quanto è bravo il modello a giudicare vini mai visti prima. È il nostro indice di affidabilità nel mondo reale.

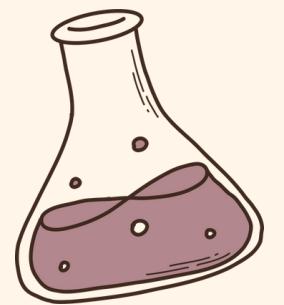
CONCLUSIONI

LA “FIRMA” CHIMICA DELLA QUALITÀ



MODELLO SOLIDO

Abbiamo sviluppato una Random Forest da 6.500 vini per distinguere con precisione i vini mediocri da quelli eccellenti.



LA CHIMICA

Non serve magia per capire la qualità, ma basta sapere quali parametri guardare



IL FUTURO

Questo strumento non serve a licenziare i sommelier, ma a potenziare il loro lavoro.

A dark, atmospheric photograph showing a row of wine bottles standing upright. The bottles are mostly black with some gold-colored foil at the top. They are arranged in a perspective that recedes towards the right side of the frame. The lighting is low, creating strong shadows and highlights on the bottles.

**GRAZIE PER
L'ATTENZIONE**

CARTANESE ALESSANDRO E ROTA GABRIELE