

CONFUSION MATRIX

STRUMENTO UTILIZZATO PER VALUTARE LE PRESTAZIONI DI UN CLASSIFICATORE IN UN PROBLEMA DI APPRENDIMENTO SUPERVISIONATO. SERVE A:

- VISUALIZZARE COME UN MODELLO CLASSIFICA DEI DATI IN DIVERSE CATEGORIE
- AIUTA AD IDENTIFICARE DOVE IL MODELLO FALLISCE

STRUTTURA DELLA MATRICE

		PREDICTED VALUE		
		Class A	Class B	Class C
ACTUAL VALUE	Class A	38	12	0
	Class B	5	43	2
	Class C	6	0	44

VALORI CORRETTI ASSEGNAZI AI DATI → VALORI PREDETTI DAL MODELLO

VALORI PREDETTI PER LA CLASSE A:

- 38 CAMPIONI DI CLASSE A SONO STATI CORRETTAMENTE PREDETTI COME CLASSE A.
- 12 CAMPIONI DI CLASSE A SONO STATI ERRONEAMENTE PREDETTI COME B.
- 0 NESSUN CAMPIONE CLASSIFICATO COME CLASSE C.

METRICHE

- **TRUE POSITIVE (TP)**: QUANDO IL MODELLO PREDICE CORRETTAMENTE UN'ISTANZA POSITIVA, $\text{e.g. SPAM} \rightarrow \text{SPAM}$
- **TRUE NEGATIVE (TN)**: QUANDO IL MODELLO PREDICE CORRETTAMENTE UN'ISTANZA NEGATIVA, $\text{e.g. NON SPAM} \rightarrow \text{NON SPAM}$
- **FALSE POSITIVE (FP)**: QUANDO IL MODELLO PREDICE POSITIVAMENTE MA LA REALTÀ È NEGATIVA (FALSO ALLARME) $\text{e.g. NON SPAM} \rightarrow \text{SPAM}$
- **FALSE NEGATIVE (FN)**: QUANDO IL MODELLO PREDICE NEGATIVAMENTE MA LA REALTÀ È POSITIVA (MANCATO RILEVAMENTO) $\text{e.g. SPAM} \rightarrow \text{NON SPAM}$

ESEMPIO

Valore Reale	Predizione del Modello	Risultato
Malato	Malato	True Positive (TP)
Non Malato	Non Malato	True Negative (TN)
Non Malato	Malato	False Positive (FP)
Malato	Non Malato	False Negative (FN)

METRICHE CALCOLABILI

ACCURATEZZA: PROPORZIONE DI PREVISIONI CORrette SUL TOTALE:

$$\text{ACCURACY} = \frac{\text{TP} + \text{TN}}{\text{TOTALE}}$$

PRECISIONE: QUANTE TRA LE ISTANZE PREDette POSITIVE SONO EFFETTIVAMENTE POSITIVE:

$$\text{PRECISIONE} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

RECALL (SENSIBILITÀ): QUANTE ISTANZE POSITIVE REALI SONO STATE IDENTIFICATE COME POSITIVE

$$\text{RECALL} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

F1: MEDIA ARMONICA DI PRECISIONE E RECALL, UTILE CON CLASSI SBILANCiate. COMBINA PRECISIONE E RECALL FORNENDO UNA MISURA PIÙ COMPLETA DELLE PERFORMANCE DEL MODELLO RISPETTO AD OGNI METRICA PRESA SINGOLARMENTE.

$$F1 = \frac{2 \cdot \text{PRECISIONE} \cdot \text{RECALL}}{\text{PRECISIONE} + \text{RECALL}}$$

ERRORE:

$$\frac{\text{FP} + \text{FN}}{\text{CLASSIFICAZIONI}}$$

ESEMPIO

									
Dog	Dog	Dog	No Dog	Dog	No Dog	Dog	No Dog	Dog	Dog

COSA VOGLIAMO FARE: CAPIRE SE UN IMMAGINE È UN CANE

POSITIVE: IMMAGINI CONTRASSEGNAME DALL'ETICHETTA **DOG** (CONSISTONO NELLE IMM. CHE IL MODELLO PENSA SIA UN CANE, POTREBBE SBAGLIARSI). TOTALE: 7

TRUE POSITIVE: IMMAGINI CONTRASSEGNAME COME CANI, QUINDI POSITIVE, CHE

SONO EFFETTIVAMENTE CANI. TOTALE: 4

FALSE POSITIVE: IMMAGINI CONTRASSEGNAME CONE CANI, QUINDI POSITIVE MA CHE NON SONO DAVVERO CANI. TOTALE: 3

NEGATIVE: IMMAGINI CONTRASSEGNAME DALL'ETICHETTA **NO DOG**, IL MODELLO CREA CHE NON SONO CANI. TOTALE: 3

TRUE NEGATIVE: IMMAGINI CONTRASS. COME NEGATIVE, CHE EFFETTIVAMENTE NON SONO CANI. TOTALE: 1

FALSE NEGATIVE: IMMAGINI CONTRASS. COME NEGATIVE, MA CHE IN REALITÀ SONO CANI. TOTALE: 2

ACCURACY

$$\text{QUANTE PREDETTE CORRETTAMENTE?} = \frac{\text{TP+TN}}{\text{TOT}} = \frac{5}{10} = 0.5$$

PRECISION & ACCURACY PER LA CLASSE DOG

PRECISION (LA BASE SONO GLI ELEM. PREDETTI)

$$\text{QUANTI CANI PREDETTI SONO EFFETTIVAMENTE CANI?} = \frac{\text{TP}}{\text{TP+FP}} = \frac{4}{7} = 0.57$$

RECALL (LA BASE SONO I VALORI REALMENTE VERI)

$$\text{QUANTI CANI REALI SONO STATI PREDETTI COME CANI?} = \frac{\text{TP}}{\text{TP+FN}} = \frac{4}{6} = 0.67$$

PRECISION & ACCURACY PER LA CLASSE NO DOG

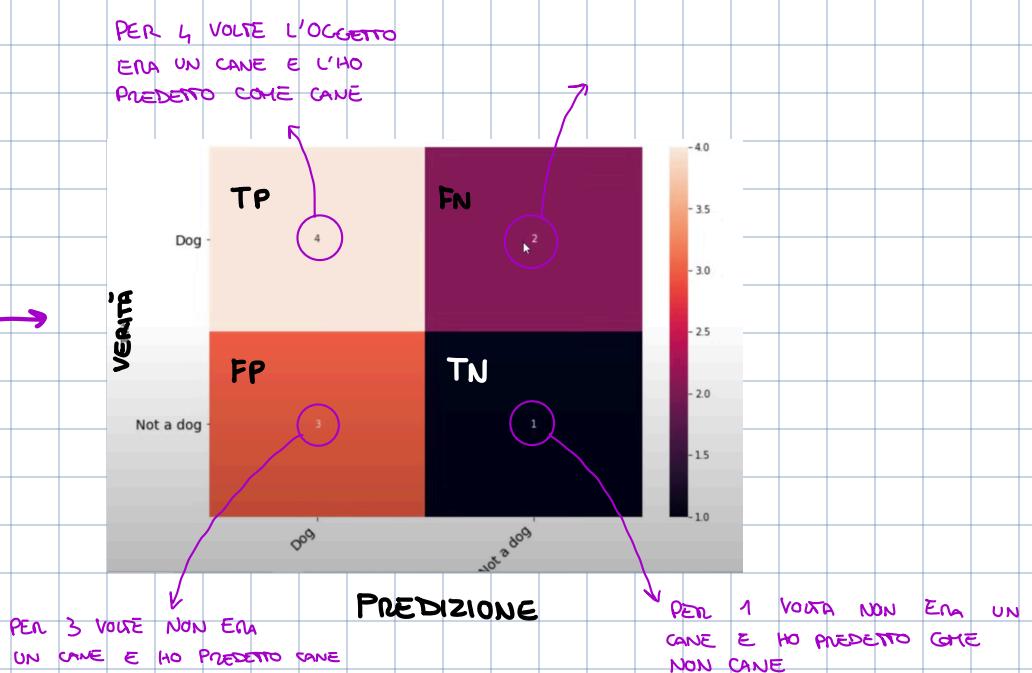
Precision

$$\frac{\text{TN}}{\text{N}} = \frac{1}{3} = 0.333$$

Recall

$$\frac{\text{TN}}{\text{TN+FP}} = \frac{1}{4} = 0.25$$

CONFUSION →
MATRIX



MACRO AVERAGE & MICRO AVERAGE

MACRO: CALCOLA LE MEDIE DI PRECISIONE TRATTANDO LE CLASSI IN MODO UNIFORME.

$$M\text{PRECISION} = \frac{1}{m} \cdot \sum_{i=1}^m \text{PRECISION}_i;$$

MEDIA SEMPLICE DELLE PRECISIONI DELLE VARIE CLASSI

$$M\text{RECALL} = \frac{1}{m} \cdot \sum_{i=1}^m \text{RECALL}_i;$$

$$MF_1 = \frac{2 \cdot M\text{PRECISION} \cdot M\text{RECALL}}{M\text{PRECISION} + M\text{RECALL}}$$

MICRO: AGGREGA I RISULTATI CONSIDERANDO IL PESO DELLE DIVERSE CLASSI IN BASE ALLA LORO FREQUENZA NEI DATI.

$$\mu\text{Precision} = \frac{\sum_{i=1}^m \alpha_i}{\sum_{i=1}^m (\alpha_i + b_i)}$$

$\alpha_i = TP$ PER OGNI CLASSE
 $b_i = FP$ PER OGNI CLASSE

$$\mu\text{Recall} = \frac{\sum_{i=1}^m \alpha_i}{\sum_{i=1}^m (\alpha_i + c_i)}$$

$c_i = FN$

$$\mu\text{BEP} = \frac{\mu\text{Precision} + \mu\text{Recall}}{2}$$

$$\mu F_1 = \frac{2 \cdot \mu\text{Precision} \cdot \mu\text{Recall}}{\mu\text{Precision} + \mu\text{Recall}}$$

TEST DEI DATI

PER OTTENERE UNA BUONA STIMA I DATI DI TEST NON DEVONO ESSERE ISTANZE DEI DATI DI TRAINING.

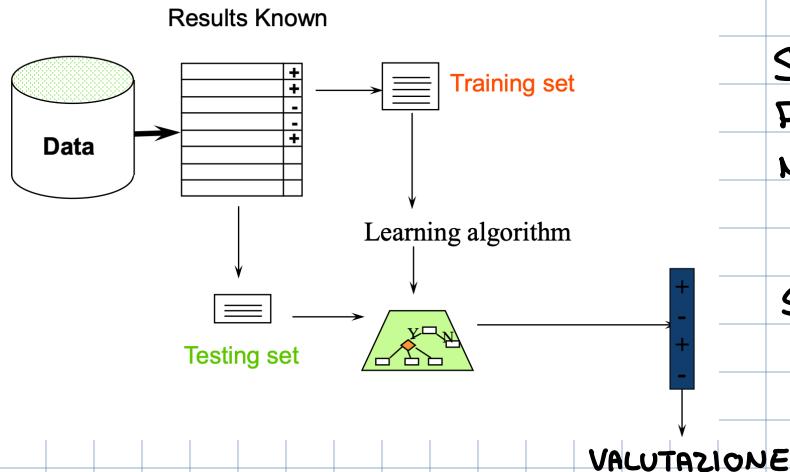
OVERTFITTING: IL MODELLO SI ADATTA TROPPO FEDELEMENTE / ESATTAMENTE AI DATI DI TRAINING, APPENDENDO "RUMORE" O "DETTAGLI"

IRRILEVANTI". PORTANDO AD UNA SCARSA GENERALIZZAZIONE SUI NUOVI DATI.

L'OBBIETTIVO PRINCIPALE NELLA VALUTAZIONE DI UN MODELLO DI MACHINE LEARNING È VERIFICARE QUANTO SIA ACCURATO NEL FARE PREVISIONI, PIUTTOSTO CHE MISURARE ALTRE CAPACITÀ COME AD ESEMPIO LA MEMORIZZAZIONE.

↳ L'OBBIETTIVO È MISURARE LA QUALITÀ DELLE PREVISIONI DEL MODELLO SU DATI NUOVI E MAI VISTI (GENERALIZZAZIONE)

- **Memorizzazione:** Il modello ricorda i dati di addestramento in modo esplicito, senza comprendere i modelli sottostanti. Questo porta a **overfitting**, cioè il modello funziona bene sul training set ma male sui dati nuovi.
- **Apprendimento:** Il modello identifica regole o schemi generali nei dati che possono essere applicati anche a nuovi esempi.



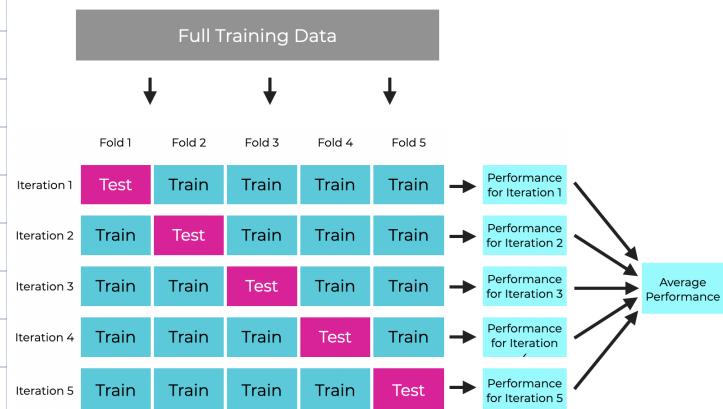
SE I DATI SONO SCARSI UN SINGOLO PROCESSO DI VALUTAZIONE POTREBBE NON ESSERE ABBASTANZA RAPPRESENTATIVO.



SOLUZIONE: **N-FOLD SPLITTING**

NEGLI SPAGNOLI DI SPLITTING I DATI SONO DIVISI IN **M** SUBSETS E OGNI SUBSET A TURNO VIENE USATO PER IL TESTING E GLI ALTRI **M - 1** PER IL TRAINING.

VENGONO POI CALCOLATE LE METRICHE PER CIASUNA ITERAZIONE, LA PRESTAZIONE FINALE È DATA DALLA MEDIA DELLE METRICHE CALCOLATA SUGLI **M** TEST SET.



TUNING

MOLTI ALGORITMI DI MACHINE LEARNING SI BASANO SU SPECIFICI PARAMETRI CHE POSSONO ESSERE REGOLATI. L' OBIETTIVO DEL TUNING È CERCARE DI REGOLARE QUESTI PARAMETRI PER MIGLIORARE LE PRESTAZIONI DEL MODELLO.

AD ESEMPIO VIENE STABILITO UN SET DI CONFIGURAZIONI, OGUNA DI QUESTE VIENE VALUTATA SU UN **VALIDATION SET**.



SOTTOINSIEME DI DATI NON USATO NEL **TRAINING SET**.

- **TRAINING SET**: SET DI DATI UTILIZZATO PER L'ADDESTRAMENTO, IL MODELLO IMPARA DA QUESTI DATI TROVANDO RELAZIONI o PATTERN.
VIENE UTILIZZATO PER REGOLARE I PESI E I PARAMETRI DEL MODELLO.
- **VALIDATION SET**: SET USATO PER OTTIMIZZARE IL MODELLO DURANTE L'ADDESTRAMENTO.
SERVE A VALUTARE LE PERFORMANCE DEL MODELLO SU DATI MAI VISTI PRIMA E VEDERE SE IL MODELLO STA GENERALIZZANDO BENE, USATO ANCHE X IL TUNING DEI PARAMETRI (es. PROFONDITÀ DELL'ALBERO), VENGONO CALCOLATE LE METRICHE PER LA VALUTAZIONE.
UN ERRORE **MOLTO BASSO** SUL TRAINING SET MA **ALTO** SUL VALIDATION SET INDICA OVERFITTING.
- **TEST SET**: SET UTILIZZATO PER VALUTARE IL MODELLO DOPO CHE È STATO ADDESTRATO E OTTIMIZZATO, SERVE AD OTTENERE UNA **MISURA FINALE** DEL MODELLO SU DATI CHE NON HA MAI VISTO PRIMA, DANDOCI UN'IDEA SU COME SI COMPONERÀ SUI DATI FUTURI.
IL TEST È USATO PER VALUTARE LE PERFORMANCE CALCOLANDO LE METRICHE COME: **F1, RECALL, PRECISIONE**.

ESEMPIO

- Dividi il dataset in tre parti:
 - **Training set (70%)**: 700 righe.
 - **Validation set (15%)**: 150 righe.
 - **Test set (15%)**: 150 righe.
- Usi il **validation set** (150 righe) per ottimizzare il modello.
- Supponiamo che l'albero di decisione abbia un parametro chiamato **profondità massima** (`max_depth`), che controlla quanto il modello può essere complesso.
- Provi diversi valori di `max_depth` (ad esempio, da 3 a 10) e osservi le performance sul validation set:
 - Con `max_depth = 3`, l'accuratezza è 85%.
 - Con `max_depth = 5`, l'accuratezza è 90%.
 - Con `max_depth = 10`, l'accuratezza scende a 80% (sintomo di overfitting, perché il modello si adatta troppo ai dati di training).
- Decidi di scegliere `max_depth = 5`, perché ottimizza le performance sul validation set.

