CORRIERE DELLA SERA

UNIVERSITÀ DI SIENA 1240

DIISM

CINECA

DISCLAIMER

ULTIME AVVERTENZE PRIMA DELLA RIVOLUZIONE

BIOMIND

INTELLIGENZA ARTIFICIALE PER L'INNOVAZIONE IN FARMACIA E BIOINFORMATICA

KNOWLEDGE PARTNER

FONDAZIONE Mondo digitale

Diritto alla conoscenza

Typography

Aa

Brera Condensed

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed

lorem ipsum

~~do eiusmod tempor incididunt labore et dolore~~

14 Regular

14 *Italic*

16 Bold

16 ***Bold Italic***

ABCDEFHIJKLM NOPQURSTUVWXYZ

abcdefghijklmnopqurstuvwxyz

1234567890!@#$%&

BRAND
Color

BLUE	R30 - G42 - B58
LIGHT BLUE	R51 - G60 - B115
VIOLET	R116 - G37 - B85
DARK RED	R160 - G38 - B49
RED	R255 - G0 - B0

INTRODUZIONE

- Cos'è l'algoritmo - Il Random Forest è un algoritmo di apprendimento supervisionato basato su un insieme di alberi decisionali.

- A cosa serve - È ampiamente utilizzato per compiti di classificazione (e anche di regressione), grazie alla sua robustezza e capacità di ridurre il rischio di overfitting.

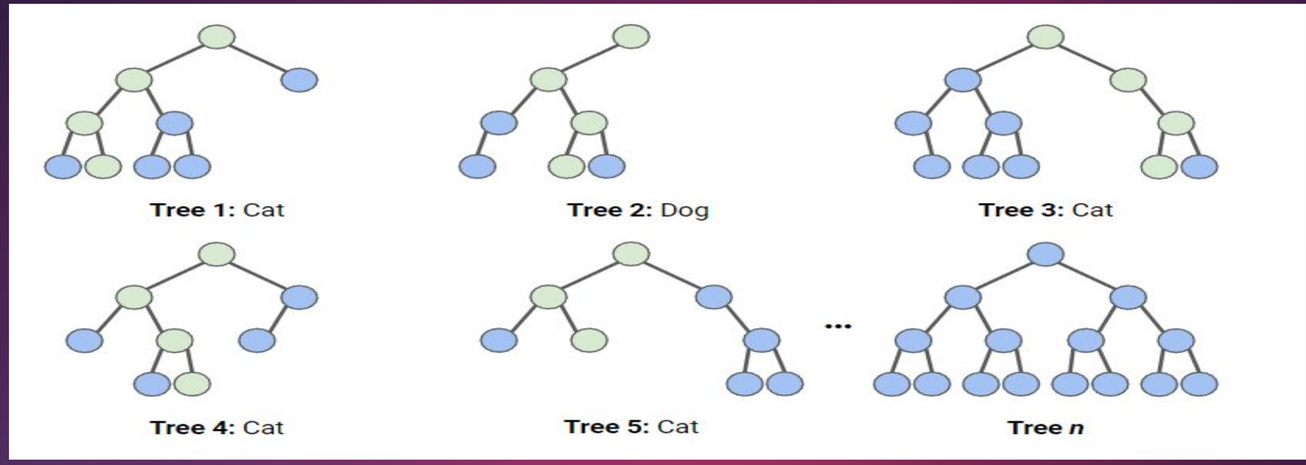
- Contesto applicativo - In economia, finanza, bioinformatica, farmacologia, biologia e in qualsiasi contesto dove il machine learning con pochi dati a disposizione e dove abbiamo bisogno di un modello robusto.

CONCETTI CHIAVE

- Struttura/architettura - Il modello funziona creando una "foresta" di alberi decisionali, ognuno addestrato su un campione casuale del dataset (con tecnica di bagging).

- Elementi principali - Ogni albero produce una predizione per la classe di appartenenza. - Le predizioni vengono combinate tramite voto di maggioranza, fornendo l’output finale.

- Schema/diagramma -



COME FUNZIONA

- Processo di addestramento - Funziona con un algoritmo di ottimizzazione (“gini” quello utilizzato nel nostro caso) che raffina iterattivamente gli alberi decisionali del random forest utilizzando i dati di addestramento che abbiamo inserito all’inizio.

- Tecniche usate -

1. Abbiamo cambiato e manipolato il numero di alberi decisionali (da 100 a 1000) e abbiamo provato diversi algoritmi di ottimizzazione e provato diversi random seed per l’inizializzazione della foresta.
2. L’obbiettivo e’ quello di migliorare la precisione di ridurre gli errori.

VANTAGGI E LIMITI

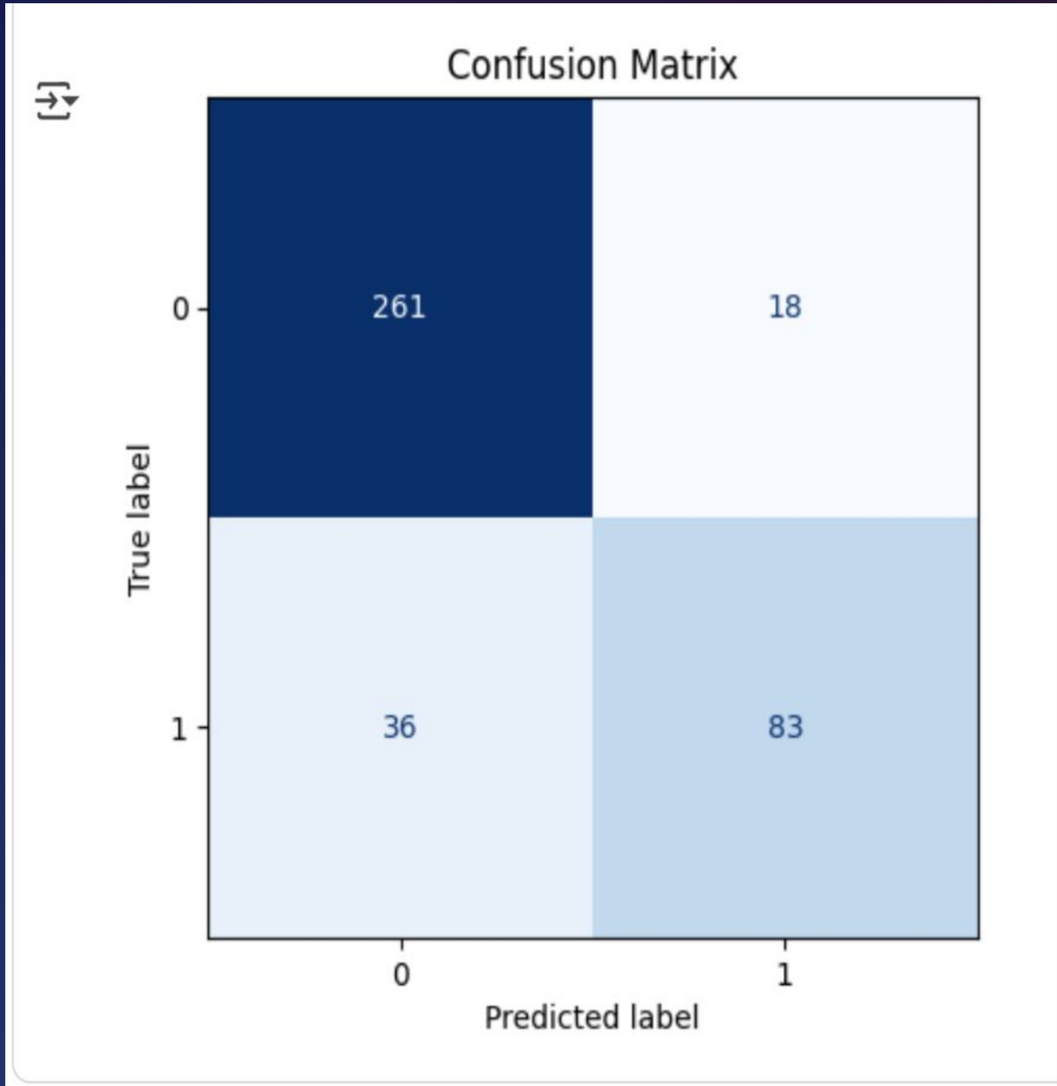
- ✅ Vantaggi:

1. Robusto al rumore
2. Riduce l’overfitting rispetto a un singolo albero
3. Gestisce bene dataset ad alta dimensionalità.

- ❌ Limiti:

1. Può essere più lento nell’inferenza rispetto a modelli più semplici,
2. Meno interpretabile rispetto a un singolo albero.

RISULTATI INIZIALI

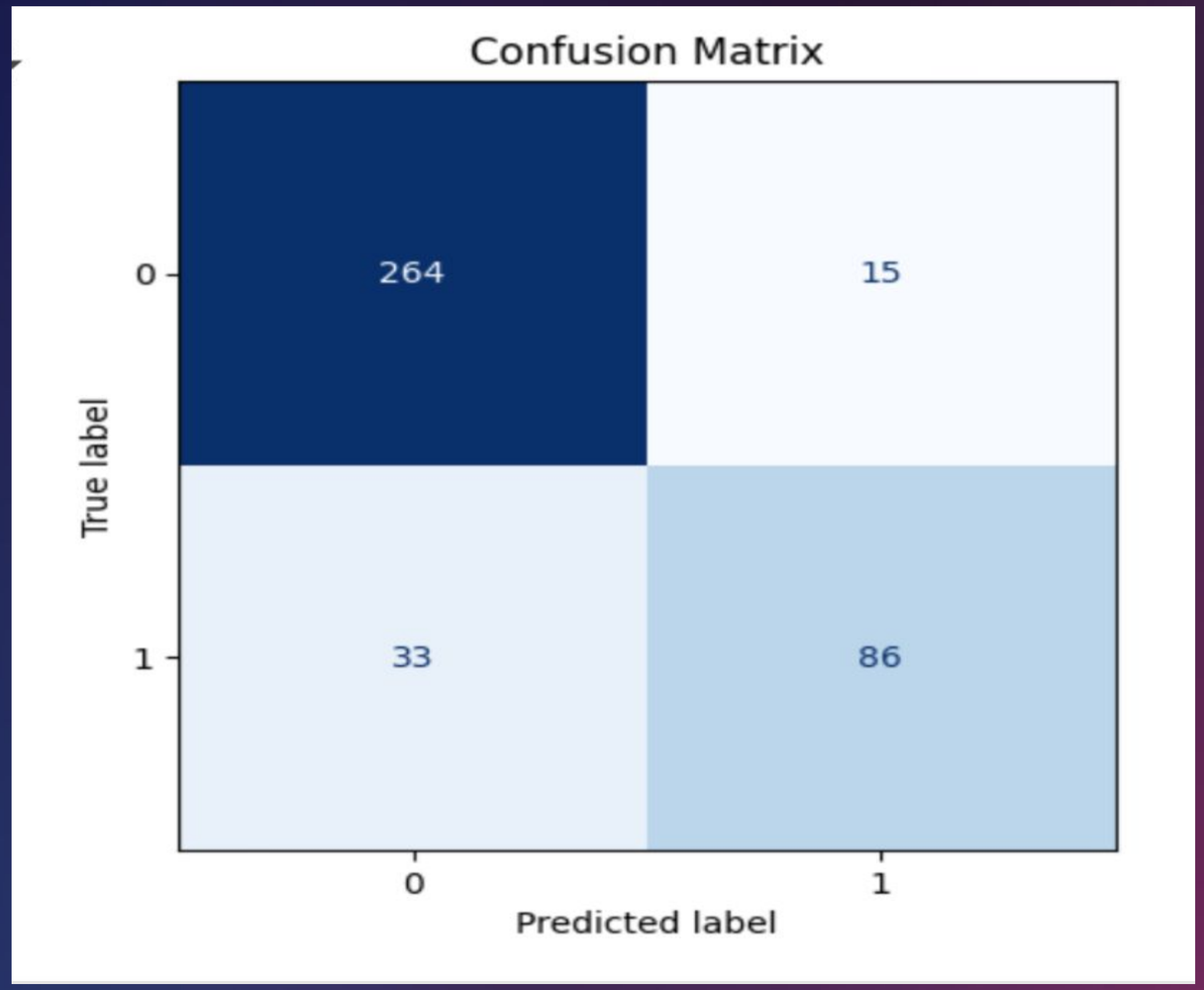



```
Accuracy: 0.86432
F1 Score: 0.75455
ROC AUC Score: 0.81648
Classification Report:
              precision    recall  f1-score   support

           0       0.88      0.94      0.91       279
           1       0.82      0.70      0.75       119

    accuracy                           0.86       398
   macro avg       0.85      0.82      0.83       398
weighted avg       0.86      0.86      0.86       398
```

RISULTATI FINALI

- 


```
Accuracy: 0.8794
F1 Score: 0.78182
ROC AUC Score: 0.83446
Classification Report:
              precision    recall  f1-score   support

           0       0.89      0.95      0.92       279
           1       0.85      0.72      0.78       119

    accuracy                           0.88       398
   macro avg       0.87      0.83      0.85       398
weighted avg       0.88      0.88      0.88       398
```

1 to 11 of 11 entries Filter

index	compound_id	protein_id	Actual	Predicted	Predicted Label
0	CID_00083	PID_170	unknown	0	Inactive
1	CID_00232	PID_170	unknown	0	Inactive
2	CID_00251	PID_170	unknown	1	Active
3	CID_00335	PID_170	unknown	1	Active
4	CID_00493	PID_170	unknown	1	Active
5	CID_00672	PID_170	unknown	0	Inactive
6	CID_00762	PID_170	unknown	1	Active
7	CID_01044	PID_170	unknown	1	Active
8	CID_01306	PID_170	unknown	1	Active
9	CID_01450	PID_170	unknown	0	Inactive
10	CID_01679	PID_170	unknown	0	Inactive

Show 25 per page

Like what you see? Visit the data table notebook to learn more about interactive tables.

Next steps: Generate code with evaluation | New interactive sheet

CORRIERE DELLA SERA

UNIVERSITÀ DI SIENA 1240

DIISM

CINECA

GRAZIE DELL'ATTENZIONE!

DISCLAIMER

ULTIME AVVERTENZE PRIMA DELLA RIVOLUZIONE

BIOMIND

INTELLIGENZA ARTIFICIALE PER L'INNOVAZIONE IN FARMACIA E BIOINFORMATICA

KNOWLEDGE PARTNER

FONDAZIONE Mondo digitale

Diritto alla conoscenza