

Machine Learning: Algoritmi e Modelli

Decision Tree e Random Forest

presentato da
Gianluca Malato

PROFESSION 

L'Albero come Modello



L'Albero come Modello



Radice (root)

base di partenza dell'albero

L'Albero come Modello



Archi (archs)

base di partenza dell'albero

L'Albero come Modello



Nodi (nodes)

contengono le informazioni

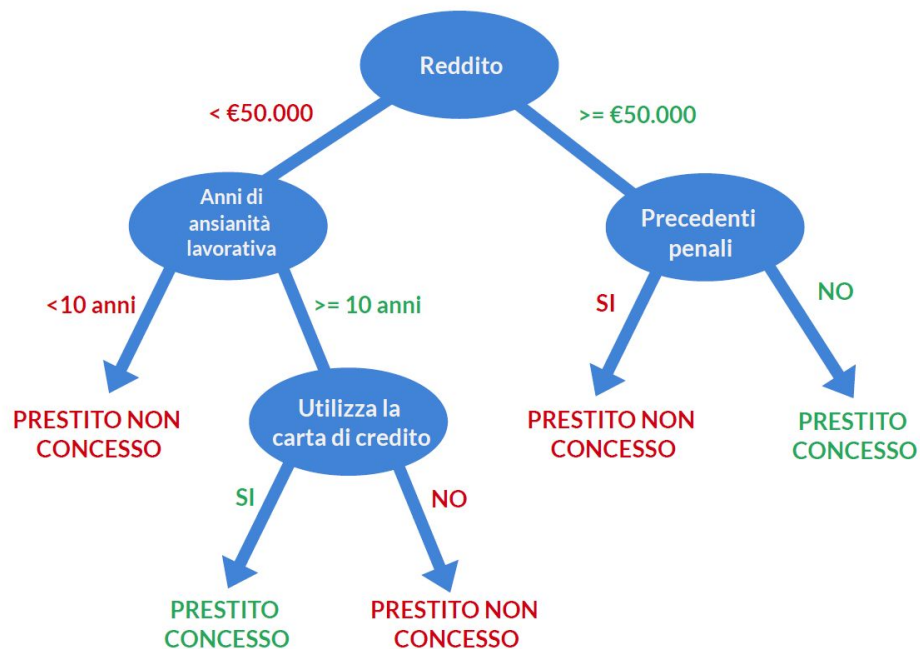
L'Albero come Modello



Foglie (leaves)

Punti di arrivo dell'albero

Otterrò il prestito dalla banca?



Reddito	Anzianità lavorativa	Carta di credito	Precedenti penali
35.000	15	SI	NO

Otterrò il prestito dalla banca?



Reddito	Anzianità lavorativa	Carta di credito	Precedenti penali
35.000	15	SI	NO

Otterrò il prestito dalla banca?



Reddito	Anzianità lavorativa	Carta di credito	Precedenti penali
35.000	15	SI	NO

Otterrò il prestito dalla banca?

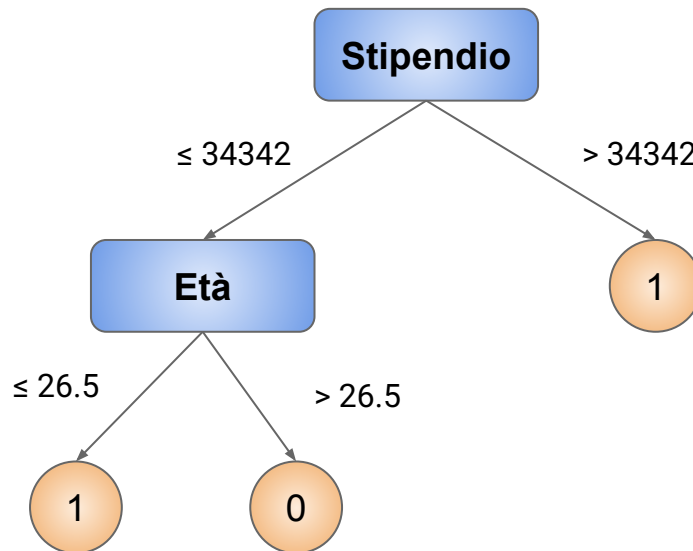


Reddito	Anzianità lavorativa	Carta di credito	Precedenti penali
35.000	15	SI	NO

Alberi decisionali

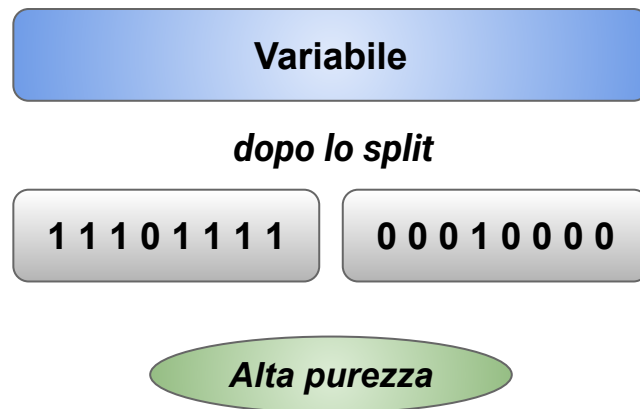
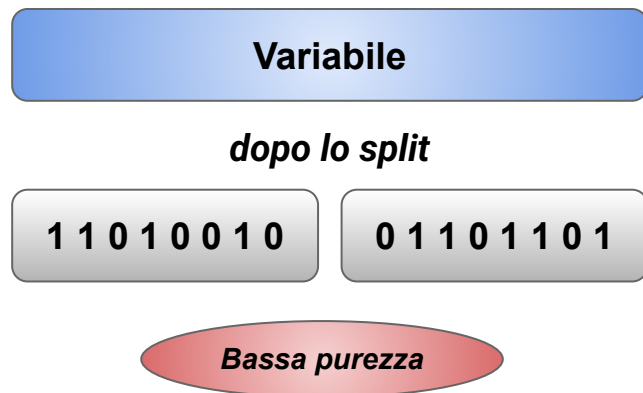
Gli alberi decisionali sono modelli che rappresentano la relazione tra le feature e un target in termini di albero decisionale binario, creando dei **punti di split** nelle variabili.

<i>Età</i>	<i>Stipendio</i>	<i>Target</i>
30	12343	0
40	23221	0
35	23342	0
45	45342	1
23	13111	1
54	21234	0



Alberi decisionali

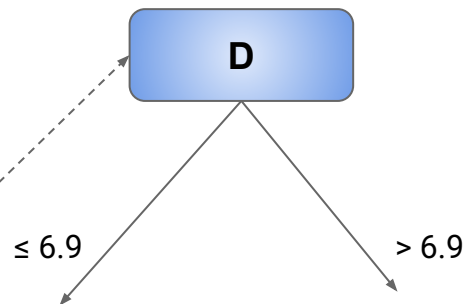
Un albero decisionale è costruito sul concetto di **purezza**. La purezza è alta se lo split è in grado di separare efficacemente i valori del target ed è bassa se la divisione crea **miscele casuali** dei valori del target nei nodi figli. In un albero decisionale cerchiamo la **massima purezza** possibile.



Alberi decisionali

Per ogni variabile numerica, lo split migliore viene calcolato come quella soglia che crea il **miglior incremento della purezza** dei nodi figli. Si seleziona quindi la variabile con il miglior incremento.

<i>Variabile</i>	<i>Incremento nella purezza</i>	<i>Soglia</i>
A	0.3	3.4
B	0.6	8.9
C	0.8	5.7
D	0.9	6.9

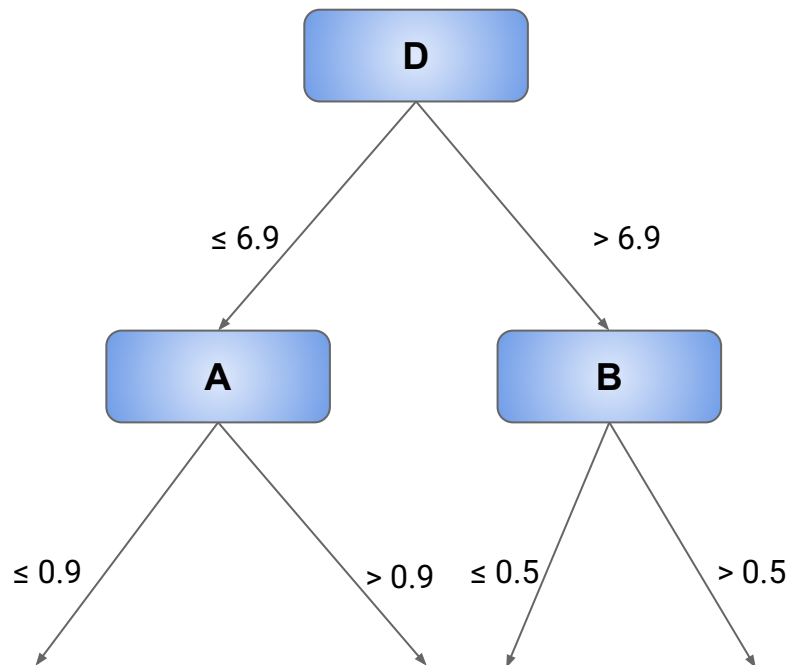


Alberi decisionali

La procedura si ripete per **ogni ramo dell'albero** secondo la regola decisionale che caratterizza di ogni nodo.

Il miglior punto di split potrebbe cambiare e la stessa variabile potrebbe apparire più di una volta.

Gli ultimi nodi di un albero sono chiamati **foglie**.



Alberi decisionali

Criteri di arresto

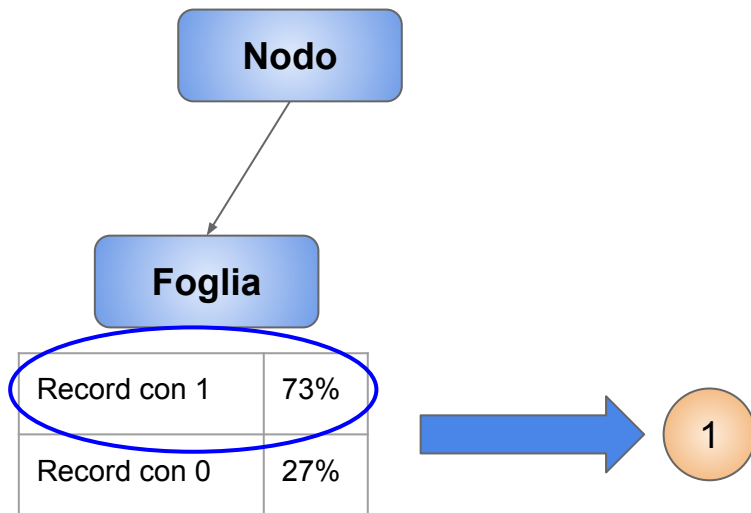
Esistono alcuni criteri possibili per interrompere l'addestramento di un albero decisionale.

- L'incremento di purezza è inferiore a una soglia fissa
- Il numero di record di un nodo è inferiore al numero minimo consentito per uno split
- L'albero raggiunge un numero massimo predefinito di livelli
- Raggiungiamo un nodo puro (ovvero otteniamo un singolo valore del target per ogni record del nodo)

Alberi decisionali

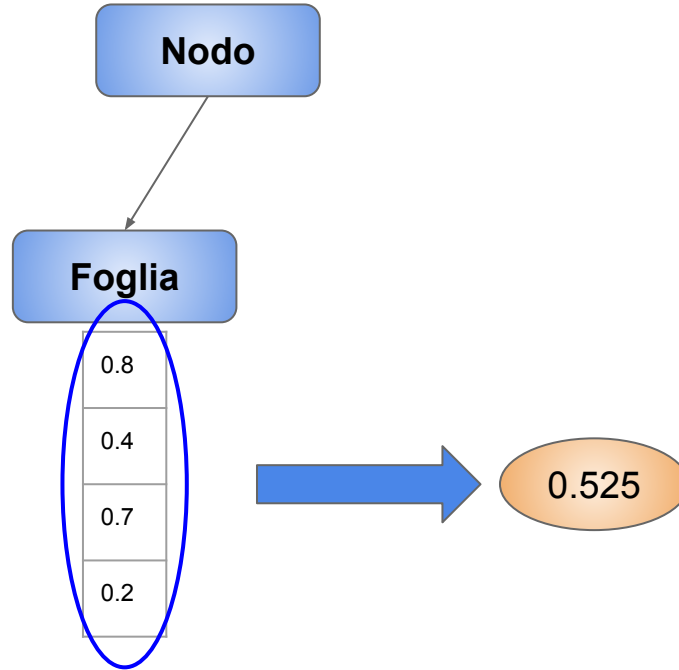
Alcune foglie potrebbero **non essere pure** (cioè potrebbero essere legate a record con più di un valore del target). Quando ciò accade, dobbiamo fare alcune scelte sulle previsioni.

Per la classificazione, la classe di **maggioranza** viene restituita



Alberi decisionali

Per la regressione, viene restituito il **valore medio**



Alberi decisionali

Come calcolare lo split

Si possono usare diversi approcci



Classificazione

Minimizzare l'indice di **Gini**

Massimizzare l'**Information Gain**



Regressione

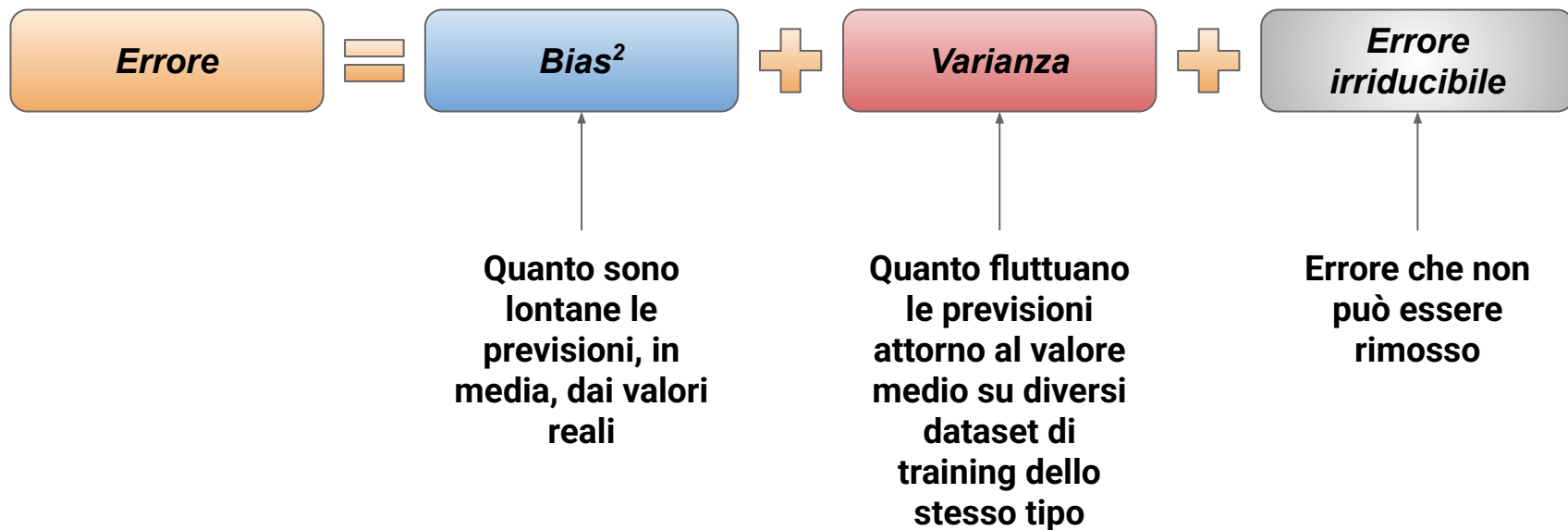
Minimizzare la **Varianza**

Alberi decisionali

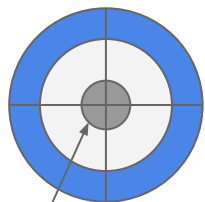
Gli iperparametri più importanti di un albero decisionale sono:

1. Il criterio di split (Gini o Information Gain)
2. La profondità dell'albero
3. Il numero minimo di record necessari per dividere un nodo
4. Il numero minimo di record necessari per creare una foglia
5. Il numero massimo di variabili da considerare in ogni divisione

Bias-variance tradeoff

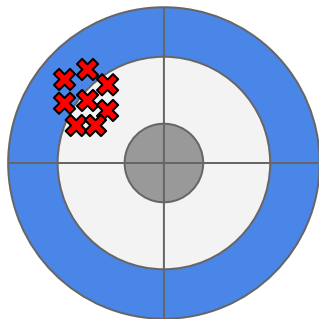


Bias-variance tradeoff

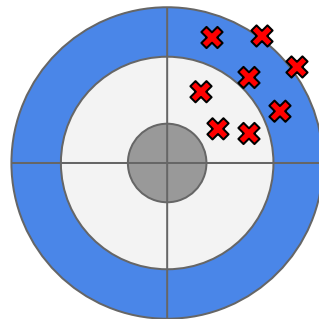


Valore reale

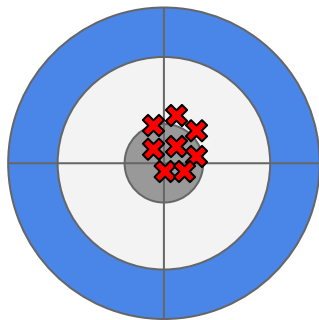
*Alto bias
Bassa varianza*



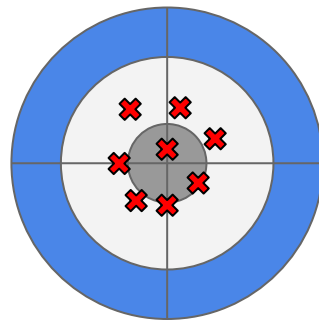
*Alto bias
Alta varianza*



*Basso bias
Bassa varianza*



*Basso bias
Alta varianza*



Bias-variance tradeoff

I modelli ensemble mescolano tra loro le previsioni di un **insieme di modelli diversi** per ridurre il bias o la varianza

Riduzione del bias



Boosting

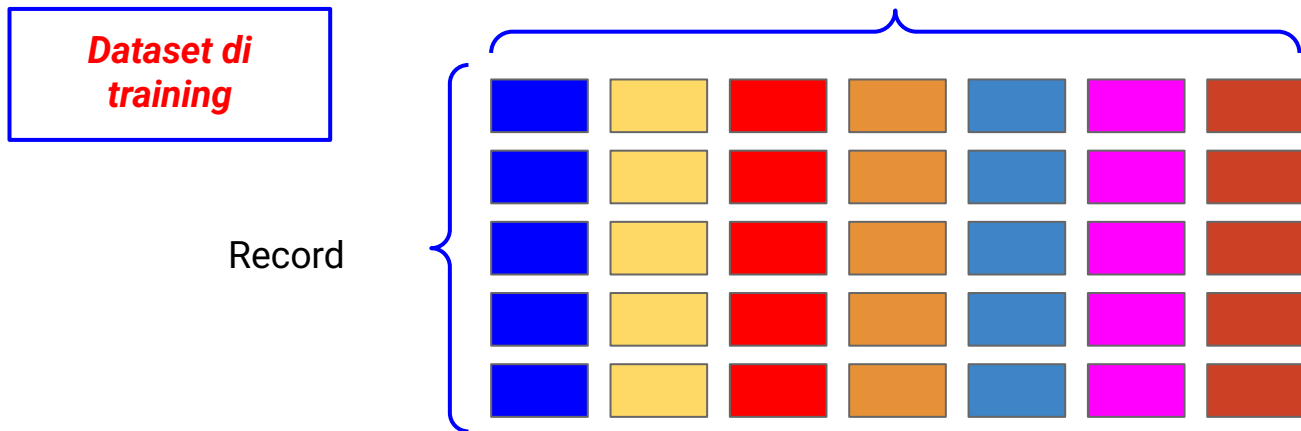
Riduzione della varianza



Bagging

Bagging

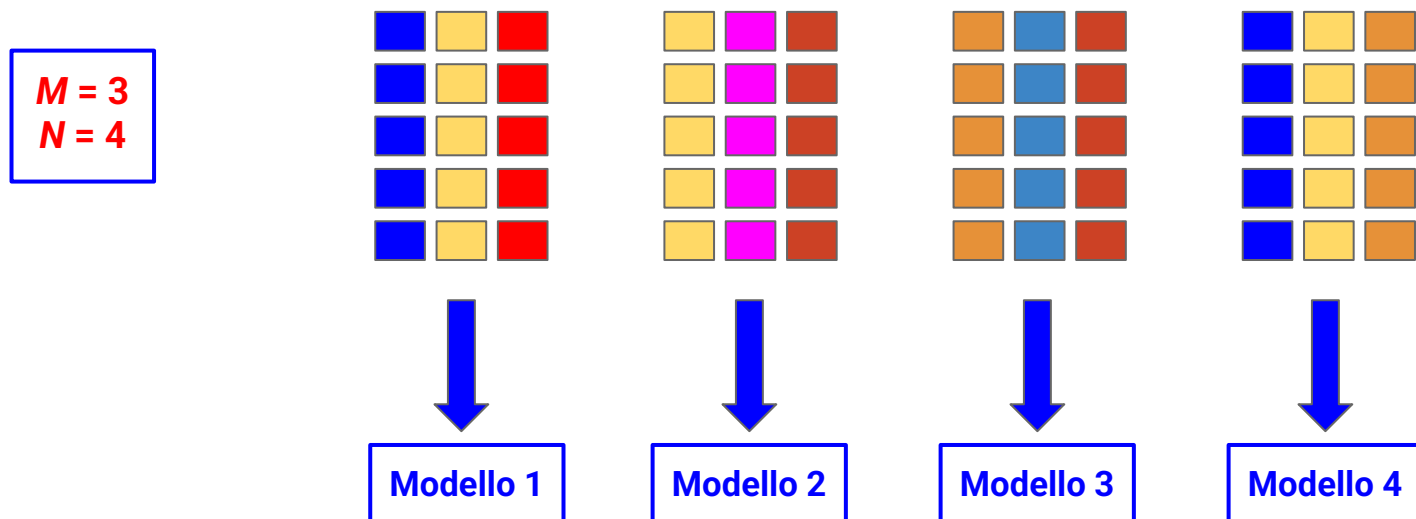
Il bagging è una tecnica che addestra diversi modelli dello stesso tipo (*weak learner*) su **diversi ricampionamenti** del dataset di training. Infine mescola le previsioni dei *weak learner* per produrre l'output finale.



Bagging

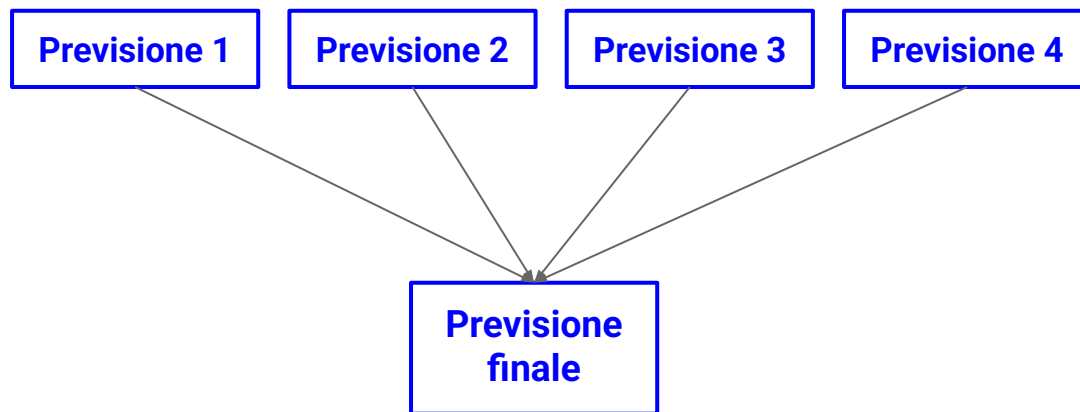
Passaggio 1: creare N campioni scelti **casualmente** dal dataset di training con M variabili scelte **casualmente**.

Addestrare un modello su ogni nuovo set di dati



Bagging

Passaggio 2: eseguire una previsione per ciascun modello. Per la regressione, calcolare la **media** delle previsioni. Per la classificazione, utilizzare il soft o hard **voting**.



Bagging

Il bagging introduce alcuni **iperparametri**:

- Il numero di weak learner
- Il numero di feature da considerare in ciascun campione
- Il numero di record di ciascun campione
- Se applicare o meno il ricampionamento con sostituzione (bootstrap) a colonne o righe

Random Forest

Random Forest è un modello ensemble basato su **alberi binari** e **bagging**. Quindi, gli iperparametri sono gli stessi di entrambi i modelli.

