Svolgimento indagine predittiva

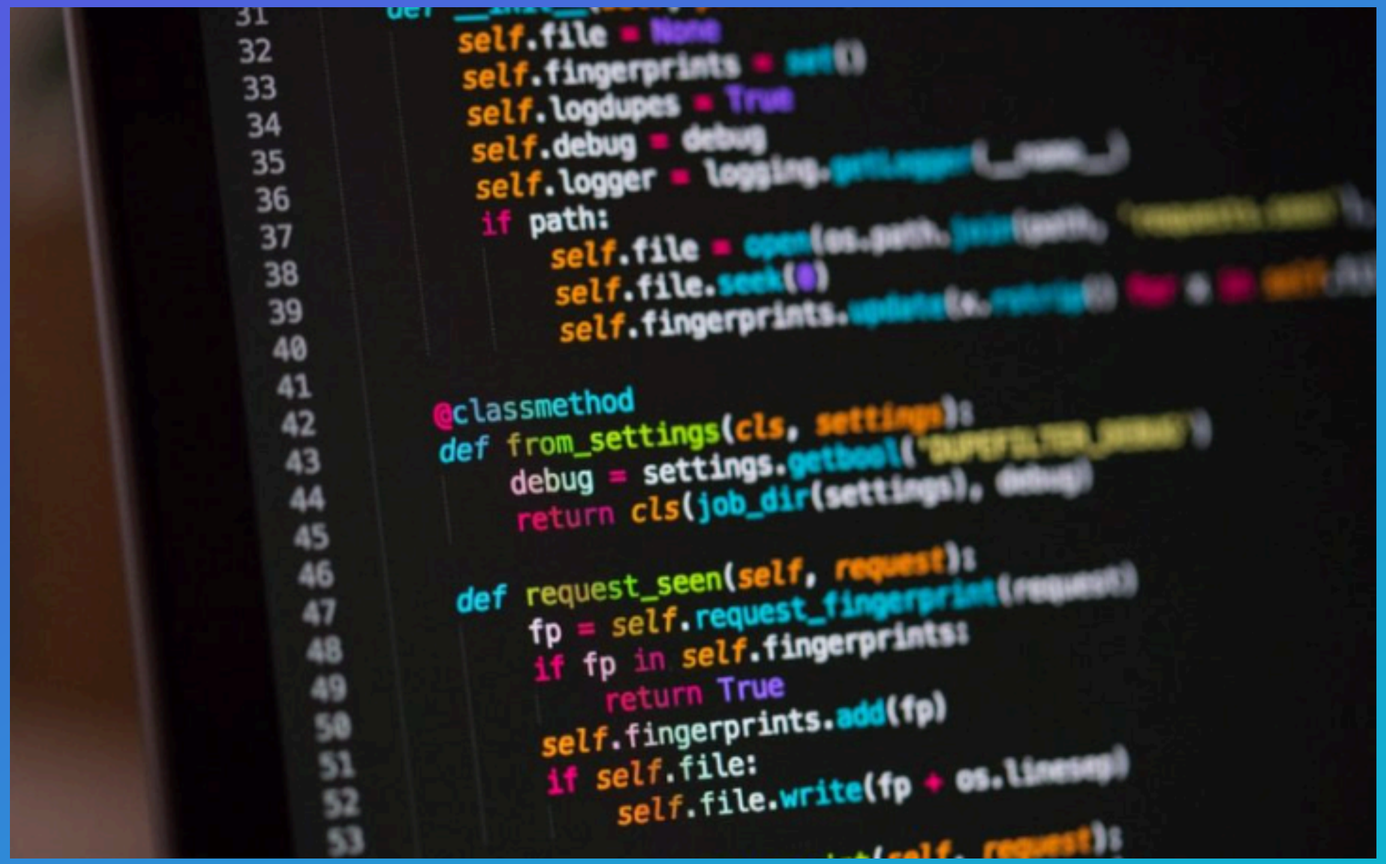
Spiegazione processo di Machine Learning su un dataset avente come oggetto i passeggeri del Titanic

Link per il notebook:

https://drive.google.com/file/d/1oIA6rWKtpNxRc9ZpWzCIXMd2IXTFpiVv/view?usp=drive_link

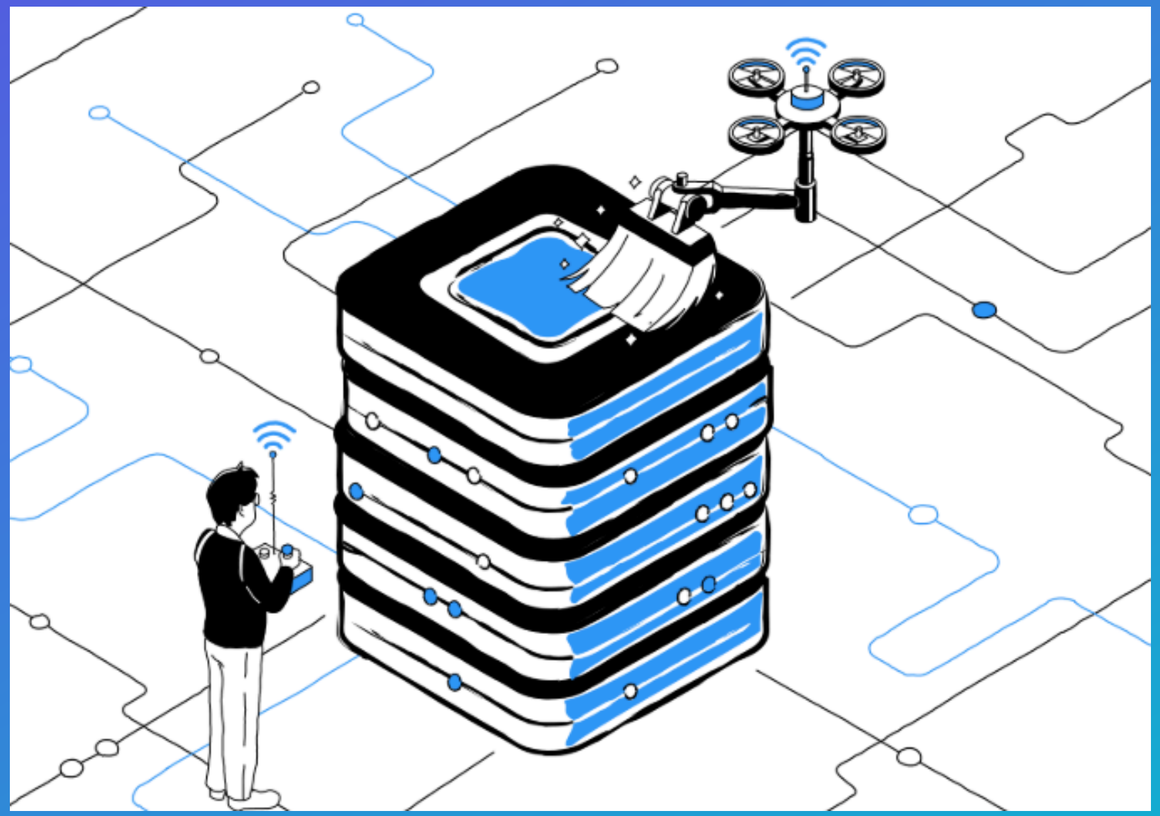
Fase 1

Il primo passo è stato quello di importare il dataset e convertirlo sottoforma di DataFrame per le ulteriori lavorazioni.



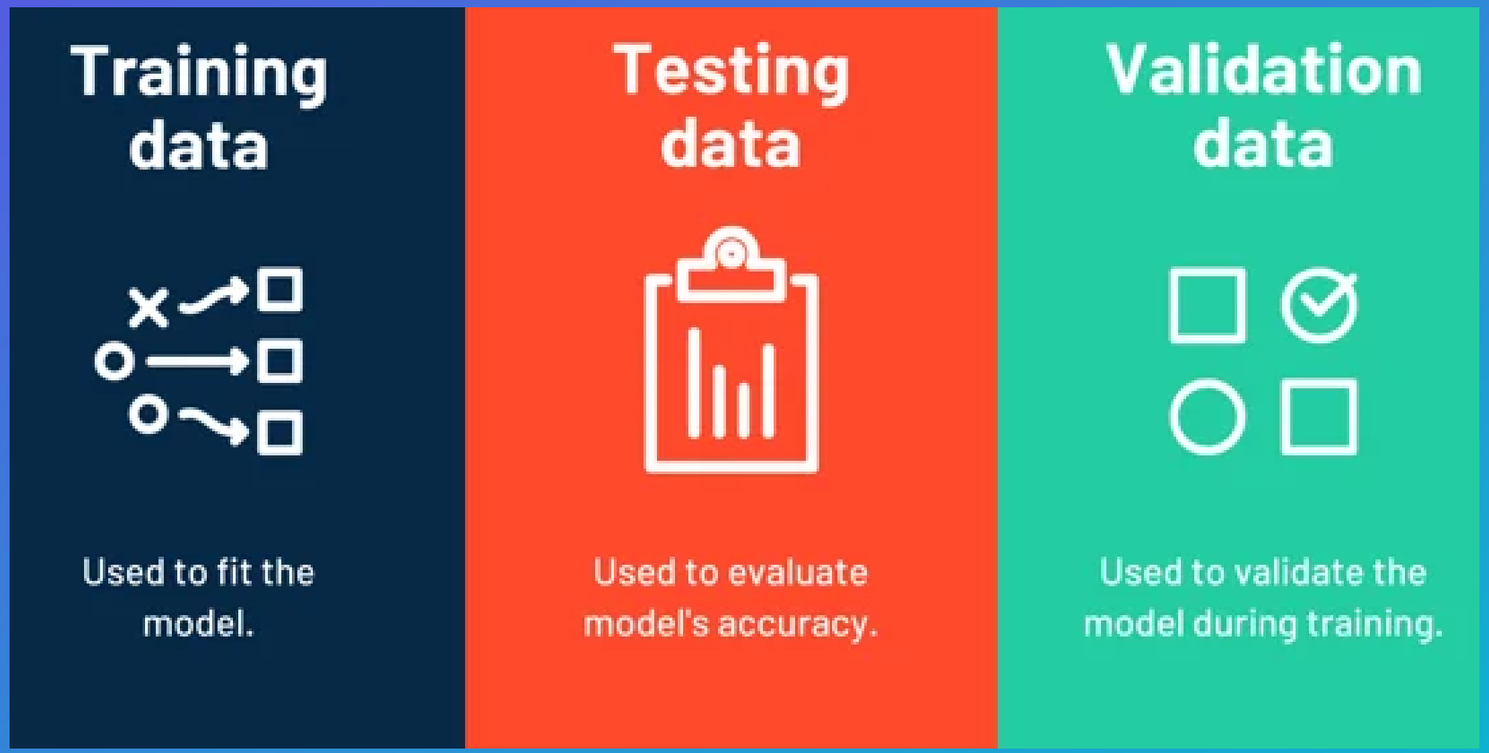
Fase 2

Una volta importato il Dataframe la prima idea è stata quella di visualizzarne il contenuto e valutarne le features. Si intravede subito che le labels (variabile y) sono contenute nella colonna “Survived”; inoltre è anche evidente che sono presenti variabili numeriche e variabili categoriche, queste ultime per essere “digerite” dal modello necessitano una conversione numerica, è quindi necessaria un’azione di pre-processing.

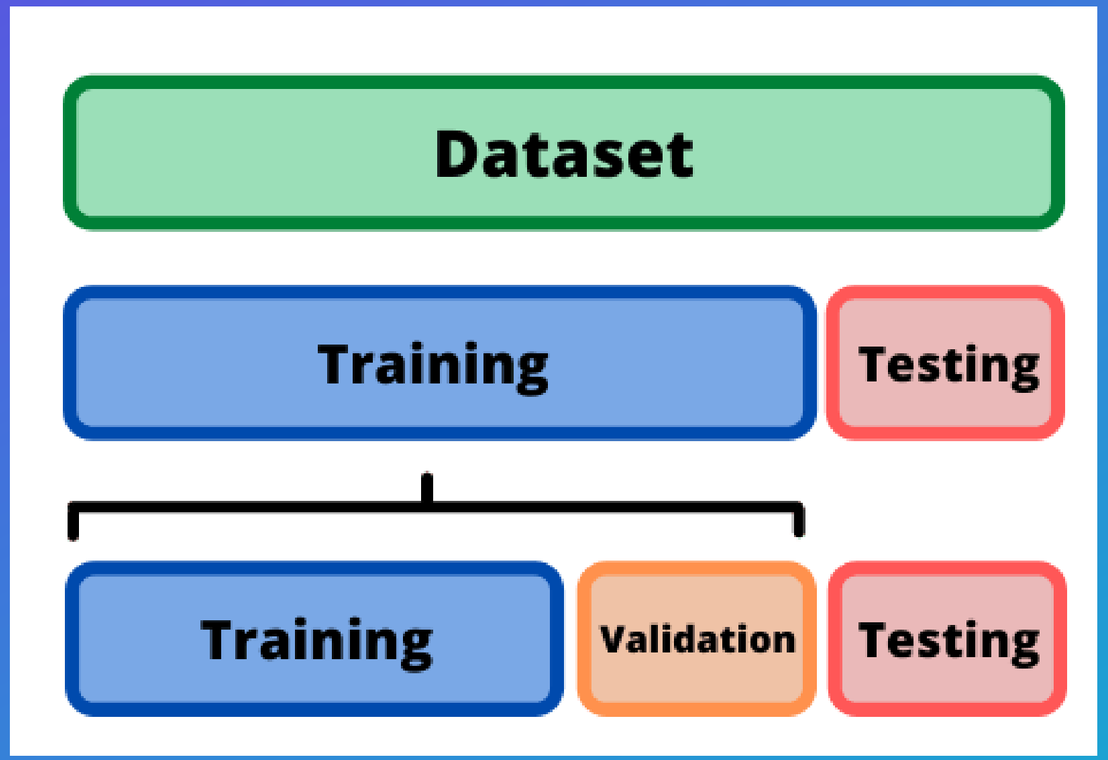


Fase 3

Una volta eseguito il pre-processing posso finalmente dividere il dataset in features “X” e in labels “y” (Survived).
Ora che le variabili sono pronte, divido il dataset in due parti, una riservata all’ addestramento del modello predittivo (Decisonal Tree) e un'altra riservata al suo effettivo utilizzo su dati mai visti. Le percentuali delle porzioni sul totale del dataset saranno rispettivamente il 75% e il 25%.

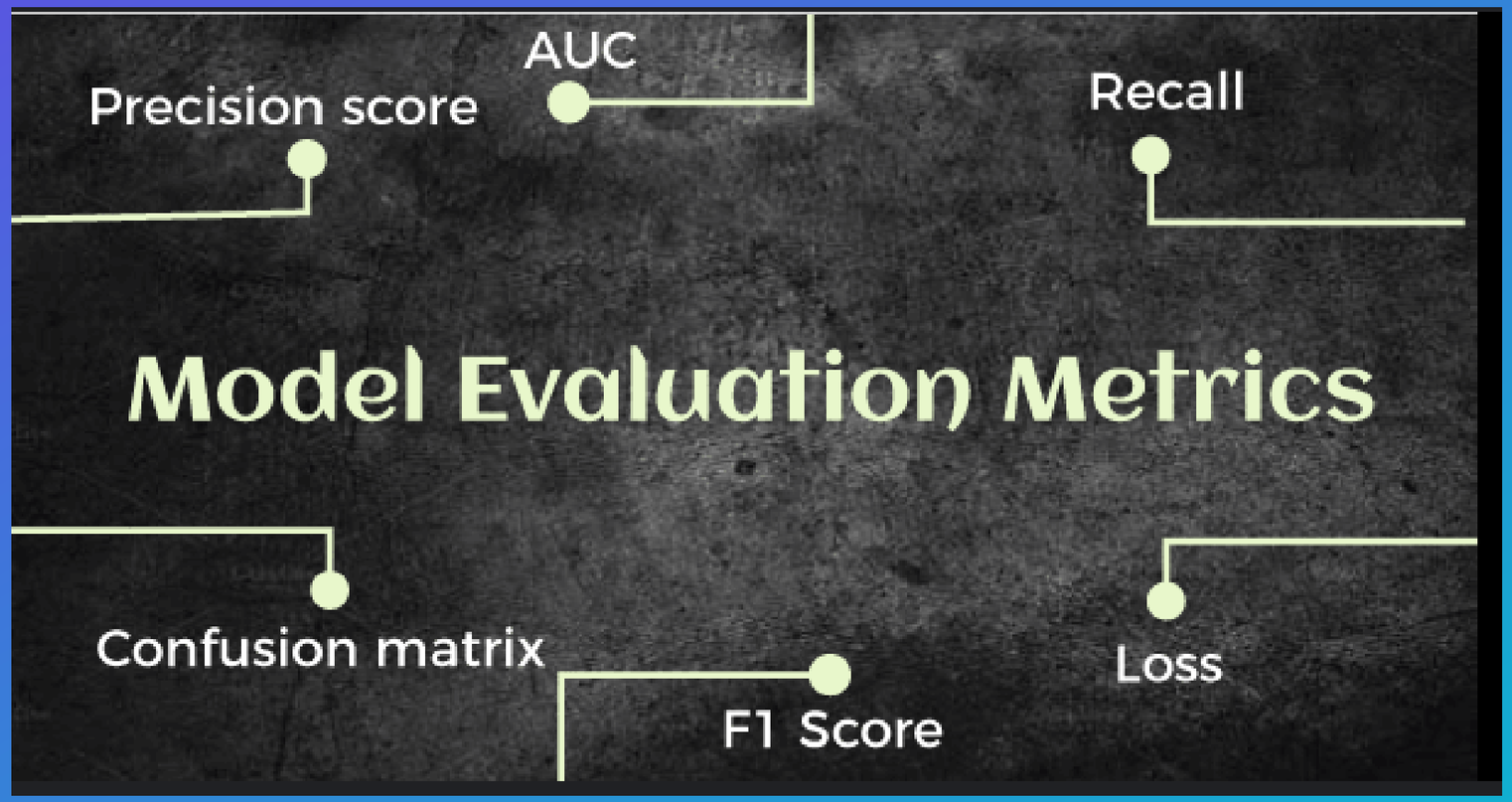


Prima di applicare il modello (dopo l’addestramento) direttamente sulla porzione di test, scelgo un ulteriore suddivisione nel training set, ovvero il validation set (25% del training set) dove posso testare il modello con gli iperparametri più ideali, evitando fenomeni come l’overfitting in fase di test.



Fase 4

In seguito, valuto l’accuratezza dei livelli di profondità (iperparametri) risultanti nel validation set, scelgo così quello che ha dimostrato delle performance migliori (profondità= 5) . Una volta scelto il modello con la profondità più idonea, lo applico finalmente alla porzione di test e ne quantifico allo stesso modo l’accuretezza. In questo caso ho raggiunto un valore di accuratezza di 0.8072.



Grazie per l’attenzione!