



CARVAL

L'AI al servizio del mercato delle auto





Chi sono

Raffaele Coppola

0512116487

Obiettivo

- Sviluppare un modello in grado di predire il **prezzo** di un'automobile in pochi passaggi semplici ...



sia in formato categorico che numerico



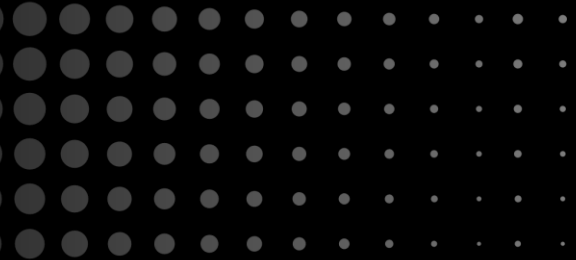


■ Cosa si vuole offrire

- Fornire supporto a venditori e acquirenti.
- Ridurre il tempo per stimare il valore delle vetture.

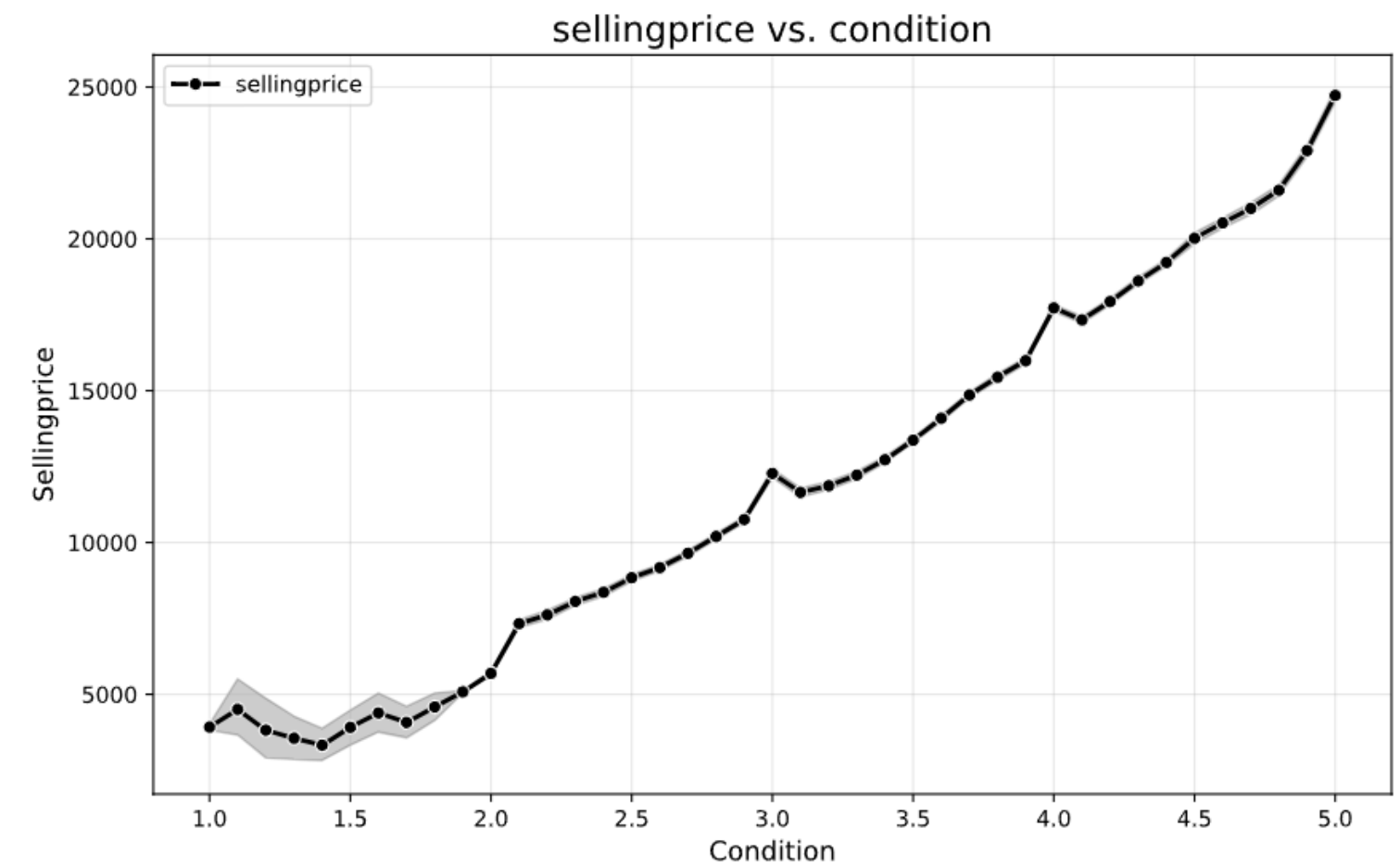
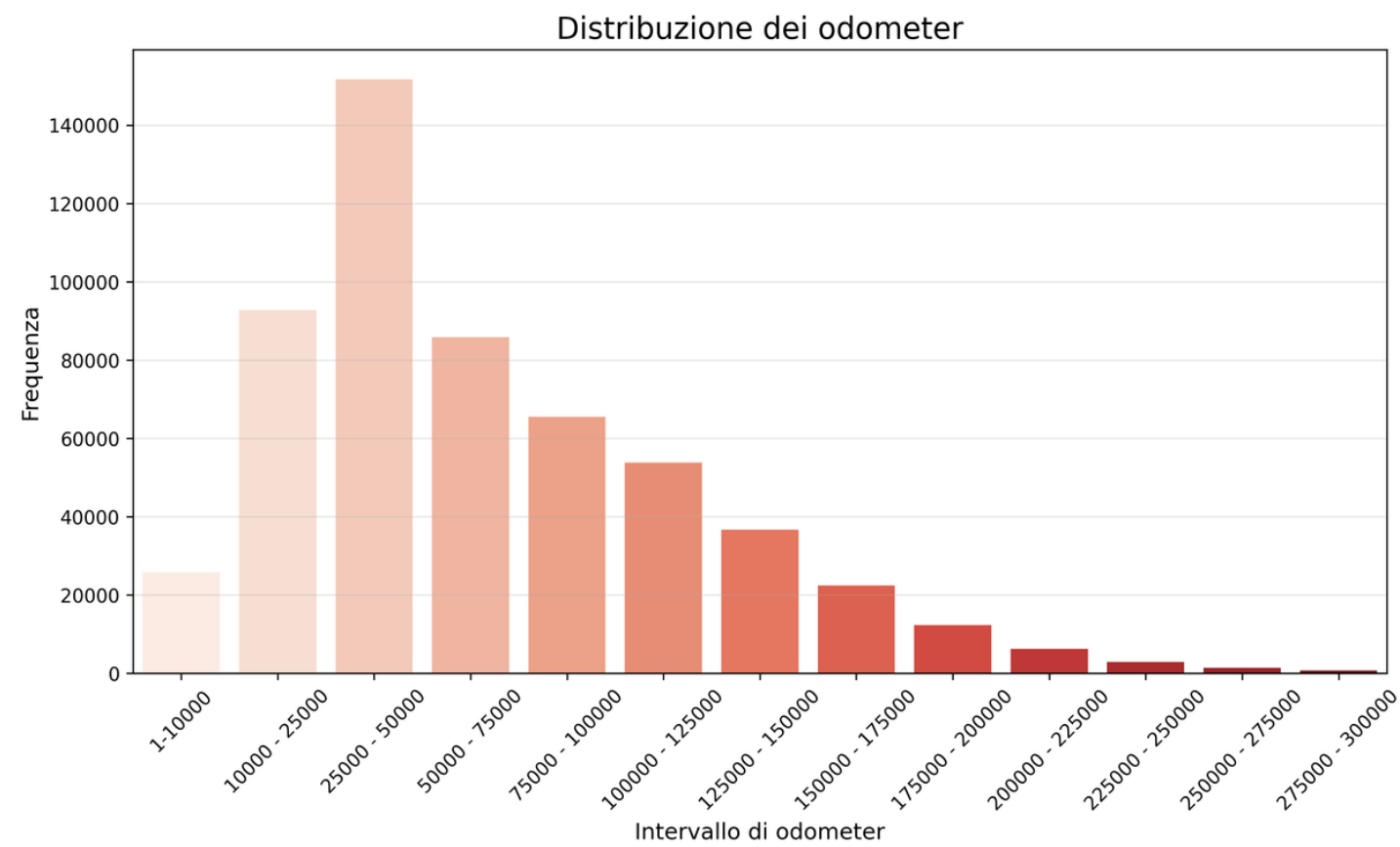
■ Valutazione delle prestazioni finali

- MAE e MAPE per la *regressione*
- Accuracy e precision per la *classificazione*.
- Il tempo necessario per generare una previsione deve essere inferiore a 3 secondi.

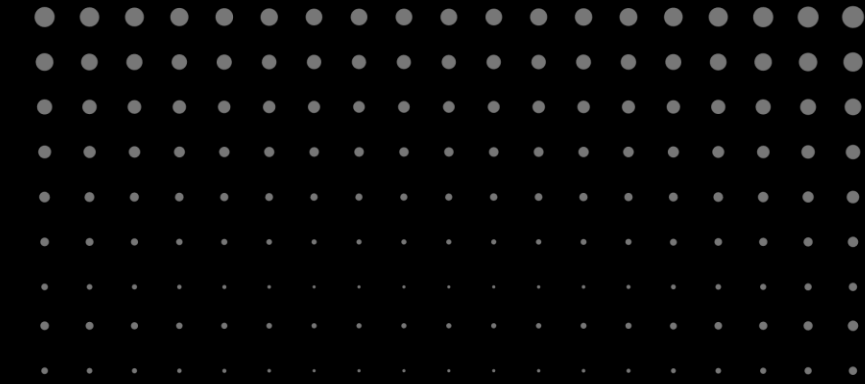


Data understanding

Il dataset di partenza è costituito da **558811** osservazioni organizzate in **16** caratteristiche.



Le caratteristiche



Anno produzione

Colore interni

Carrozzeria

Colorazione

Modello

Allestimento

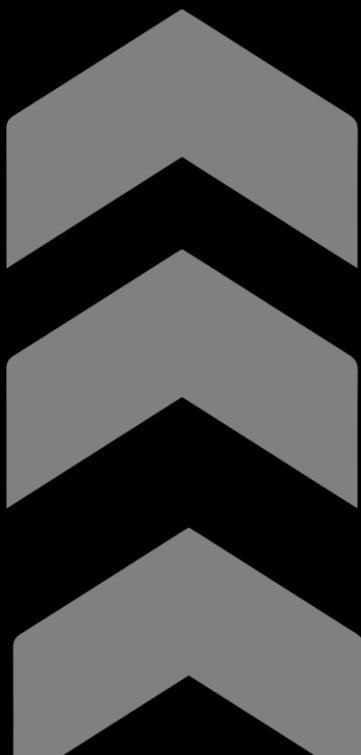
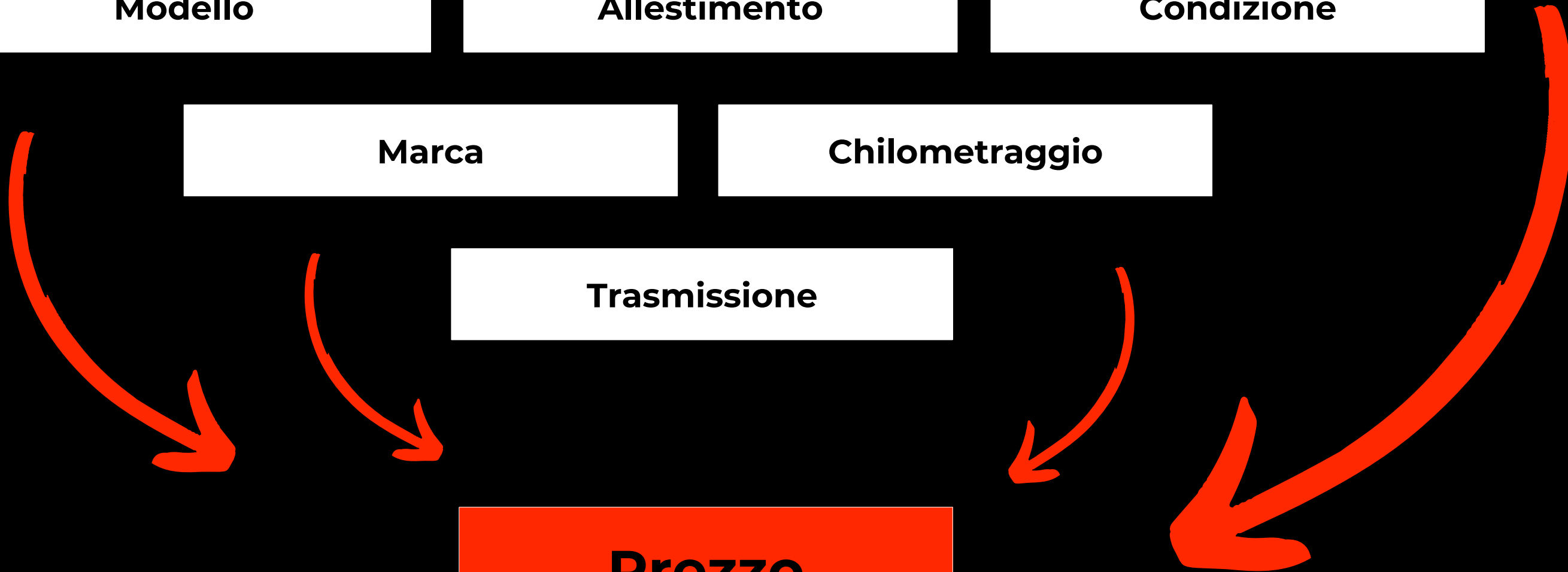
Condizione

Marca

Chilometraggio

Trasmissione

Prezzo

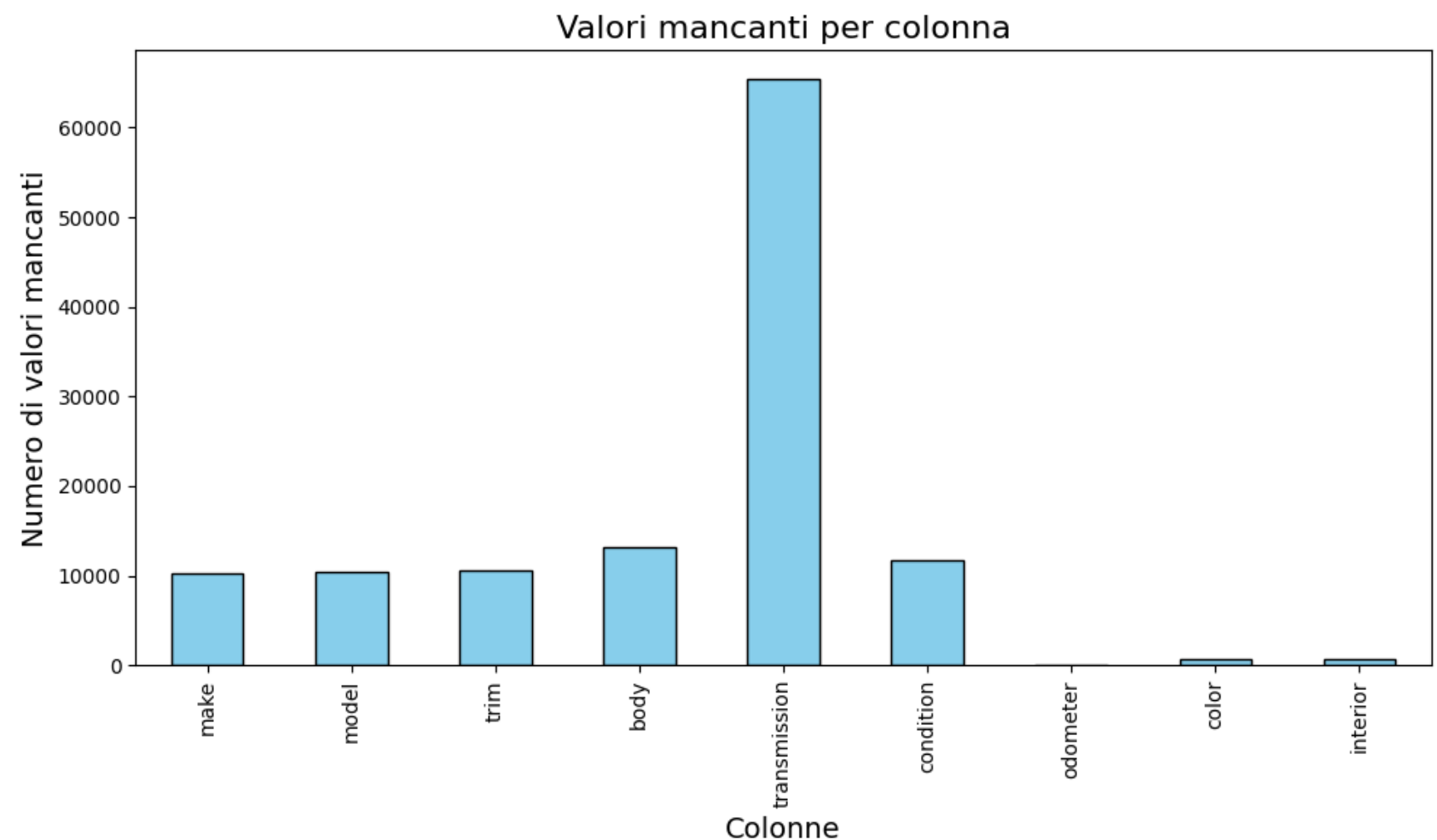


Data preparation

Data cleaning

Capire come sopravvivere alla **mancanza** o al **rumore** dei dati di partenza.

Scartare basi di esperienza oppure stimare i valori mancanti (**imputazione**).



Normalizzazione dei dati

Encoding

Le variabili categoriali non possono essere direttamente utilizzate in modelli di Machine Learning senza un adeguato processo di codifica. (**smoothed target encoding**)

Scaling

Modificare i valori delle caratteristiche allo scopo di evitare che le scale più ampie possano monopolizzare la fase di apprendimento. (**z-score normalization**)

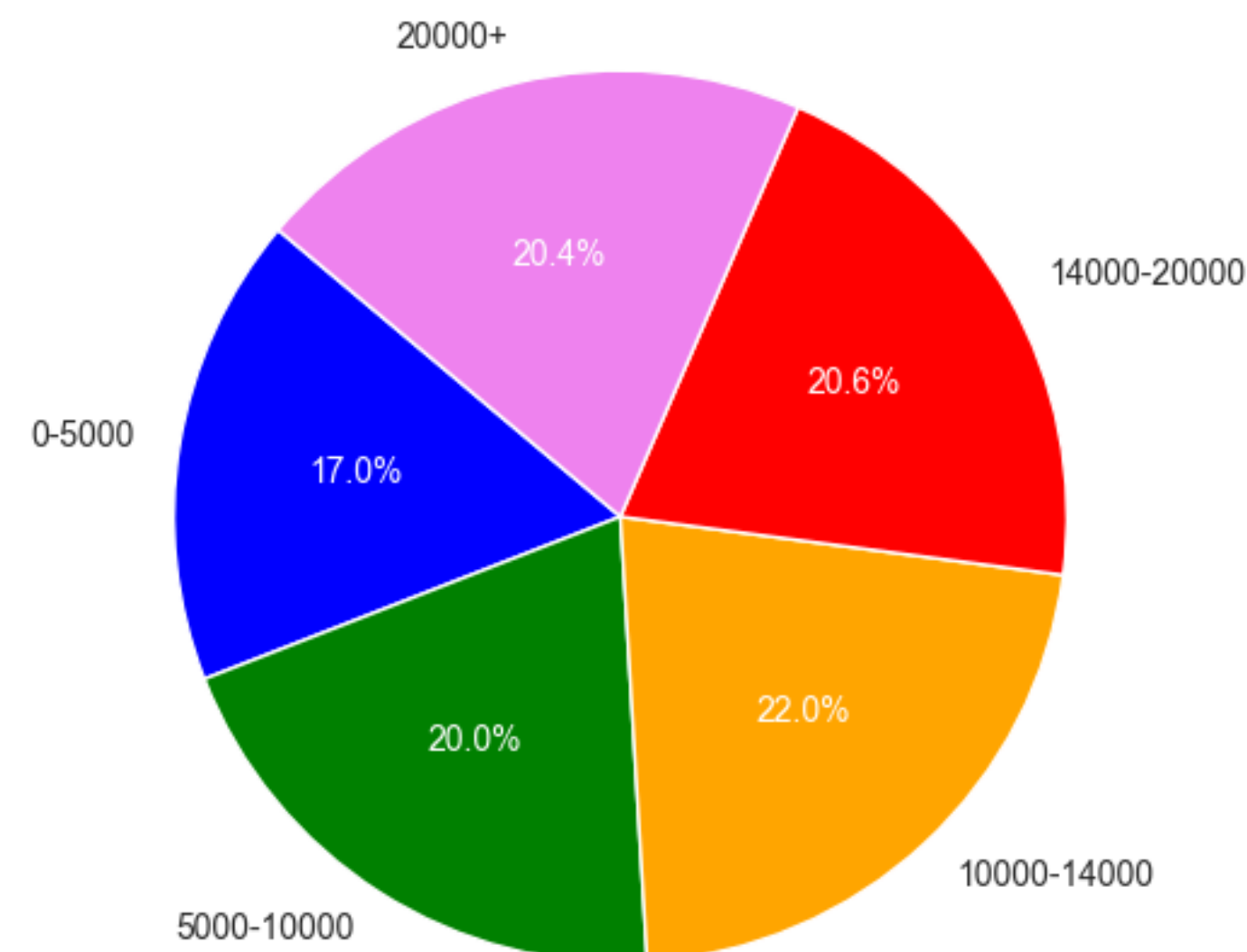
Bilanciamento delle classi

Per la classificazione

Discretizzazione del target e suddivisione delle classi in **fasce** di prezzo predefinite.

Tecniche di **oversampling** per aumentare i campioni delle classi minoritarie con *dati sintetici*.

Distribuzione delle classi di prezzo (Pie Chart)



Modeling

Random Forest

Metodo di **ensemble learning**, in cui molteplici alberi decisionali vengono addestrati su sottoinsiemi del dataset.



- | **Campionamento**

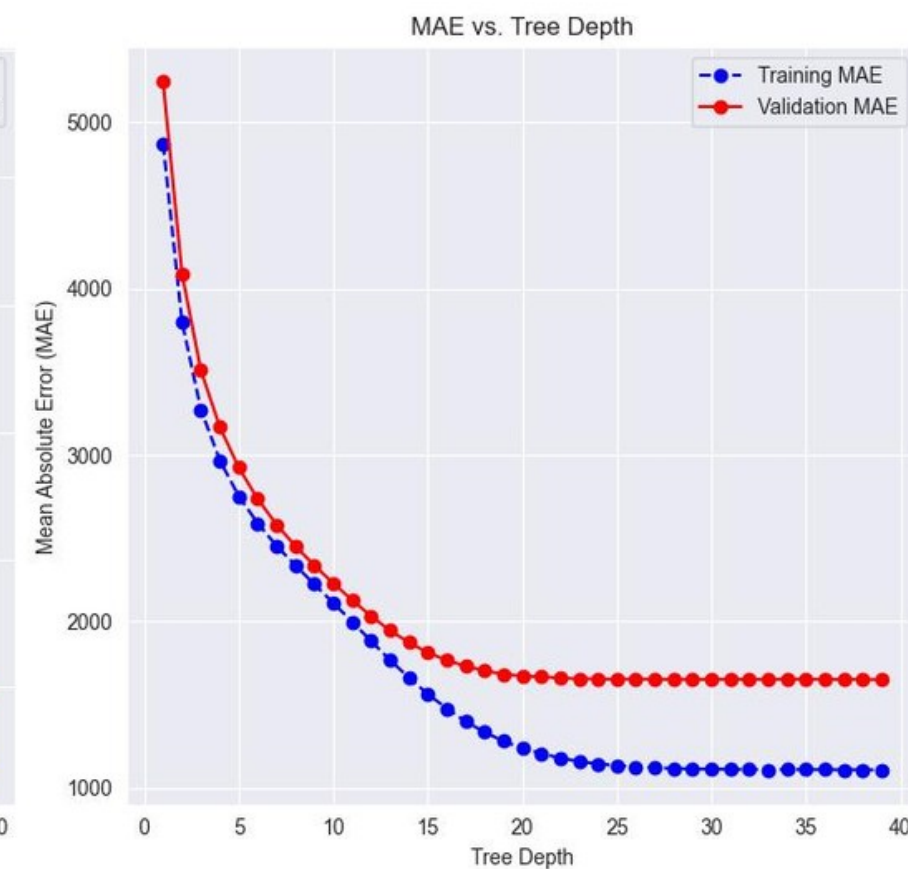
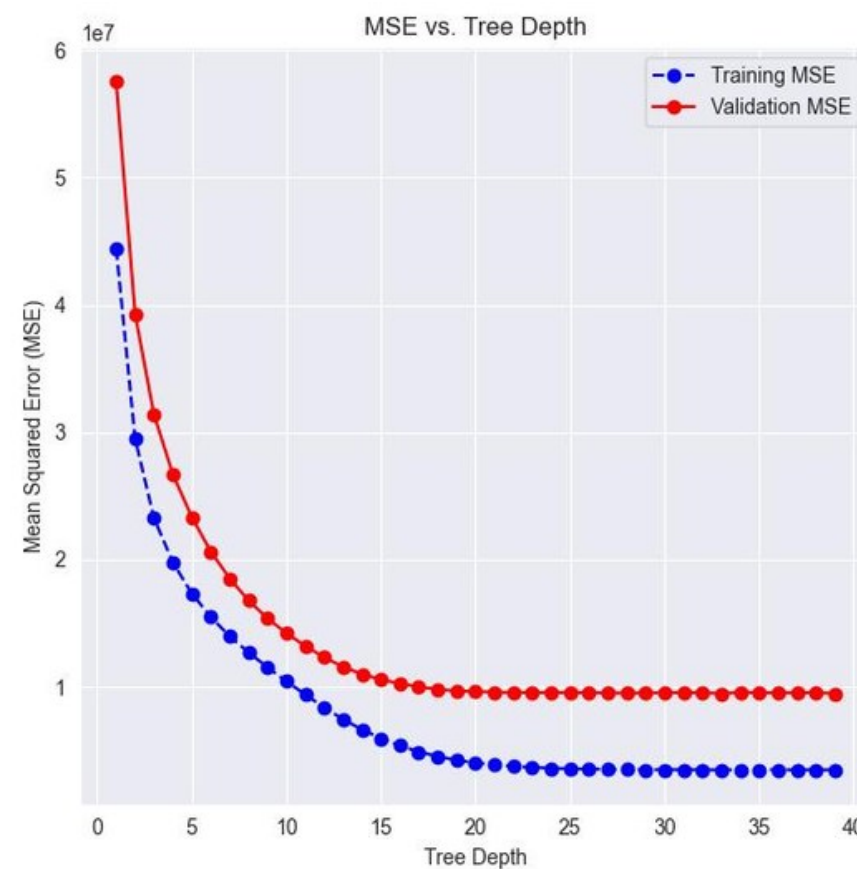
- | **Addestramento**

- | **Aggregazione**



CAPIRE LA PROFONDITÀ

Sono stati plottati dei grafici per comprendere l'andamento della **validation loss** all'aumentare del valore dell'iperparametro *max_depth*.



Evaluation

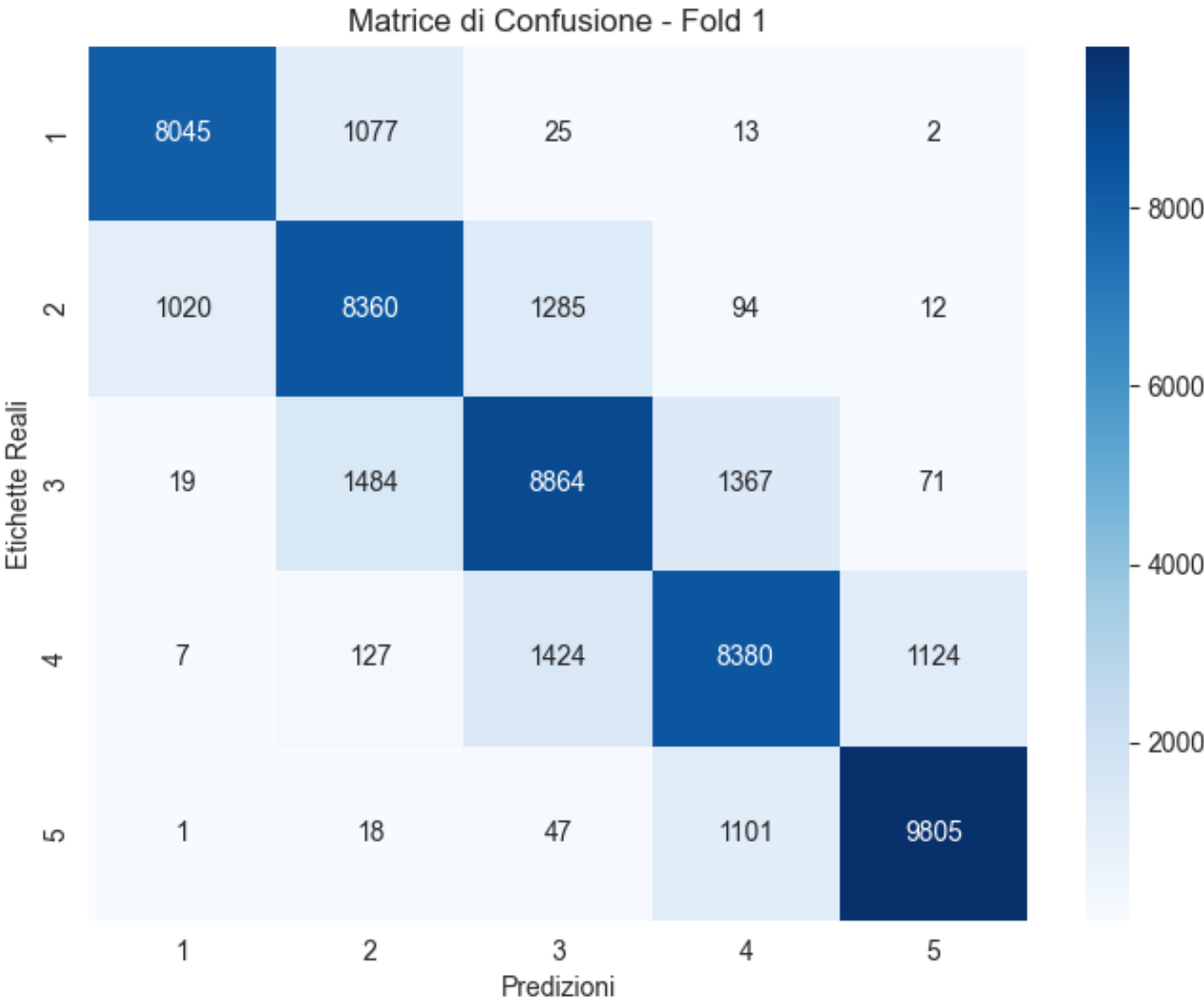
Validazione incrociata

Per evitare le problematiche derivanti dal semplice train/test split, si è deciso di effettuare la **k-fold-cross validation**.

Metodo statistico che consiste nella ripetuta partizione e valutazione dell'insieme dei dati di partenza.



Evaluation



Fold	MAE	MSE	RMSE	MAPE
1	1395.638	4920057.676	2218.120	0.157
2	1416.425	5258219.227	2293.080	0.157
...
10	1404.051	5240750.286	2289.268	0.156

Deployment

Per rendere **usabile** il modello è stata sviluppata una semplice **interfaccia** grafica.

Stima del Prezzo di un'Auto

Inserisci la marca

ford

Inserisci il modello:

Inserisci l'anno di produzione:

2010

Inserisci l'allestimento:

Inserisci la carrozzeria:

Inserisci la condizione dell'auto:

★ ★ ★ ☆ ☆

base

Inserisci la carrozzeria:

suv

Inserisci la condizione dell'auto:

★ ★ ★ ★ ☆

Inserisci il chilometraggio:

240000

Inserisci il colore esterno:

red

Inserisci il colore degli interni:

black

Stima il prezzo

Il prezzo stimato è: 12107.57€

Considerazioni finali

Possibili miglioramenti

- L'uso di strategie avanzate di configurazione degli **iperparametri**, magari tramite **algoritmi di ricerca**, potrebbe migliorare le prestazioni generali del modello.
- **Espandere, aggiornare** e rendere maggiormente **granulare** il dataset da cui l'agente matura la propria esperienza.
- Sperimentazione di diversi modelli più avanzati a discapito dell'**explainability**.



Grazie per l'attenzione!

