

Relazione

Home sales prices

Aldo Solari, matricola 2575

La valutazione si baserà sulla correttezza, chiarezza e precisione della relazione. Non è previsto un limite di pagine, ma verrà premiata la capacità di sintesi, ovvero una struttura argomentativa ben articolata, elegante e leggibile. E' possibile utilizzare fonti (libri, Internet, persone e così via) ma è richiesto di citarle nel testo. L'uso di fonti senza citarle si traduce in un voto nullo.

Tutti gli studenti sono tenuti ad aderire ad un codice di condotta, che vieta il plagio, la falsificazione, l'assistenza non autorizzata, imbrogli e altri atti gravi di disonestà accademica.

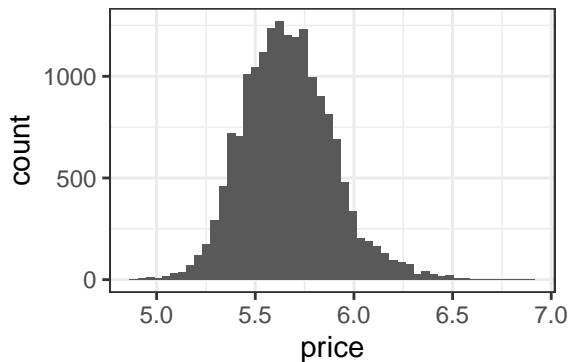
Relazione

L'obiettivo dell'analisi è la previsione del prezzo di vendita (*price*, in scala log10) di $m = 4320$ abitazioni del test set.

Sintesi del processo di modellizzazione

1. Analisi esplorativa dei dati

E' stata considerata la distribuzione della variabile risposta, rappresentata con l'istogramma che segue:



2. Pre-elaborazione dei dati

Non è stata effettuata alcuna pre-elaborazione dei dati.

3. Costruzione di nuove variabili (*feature engineering*)

Non è stata costruita nessuna nuova variabile.

5. Selezione delle variabili (*feature selection*)

Non è stata effettuata alcuna selezione delle variabili.

6. Confronto tra modelli

Non è stata effettuato alcun confronto, è stato considerato solo il modello lineare. Non è stata effettuata la regolazione del modello (*model tuning*) in quanto non prevista per il modello lineare. Non è stata effettuata la stima dell'errore di previsione (e.g. attraverso il metodo della convalida incrociata).

7. *Scelta del modello finale*

Non è stata effettuata alcuna scelta, è stato considerato solo il modello lineare.