

# Un modello di previsione per il mercato immobiliare

RealEstateAI Solutions si propone di ottimizzare la valutazione dei prezzi immobiliari attraverso l'uso di tecniche avanzate di regolarizzazione in modelli di regressione lineare. L'obiettivo è fornire previsioni di prezzo più accurate e affidabili, riducendo il rischio di overfitting e migliorando la capacità di generalizzazione del modello.

Nel settore immobiliare, ottenere stime precise dei prezzi delle proprietà è cruciale per prendere decisioni informate. Tuttavia, i modelli di regressione lineare tradizionali possono soffrire di overfitting, compromettendo l'accuratezza delle previsioni. È necessario esplorare metodi di regolarizzazione efficaci per migliorare le performance predittive e gestire la complessità del modello.

Implementando e confrontando metodi di regolarizzazione come Lasso, Ridge e Elastic Net, RealEstateAI Solutions offrirà un sistema in grado di fornire previsioni di prezzo immobiliari più accurate e stabili. Questo permetterà agli agenti immobiliari e agli investitori di prendere decisioni basate su dati più affidabili, aumentando la loro competitività nel mercato.

## Requisiti del Progetto:

### 1. Preparazione del Dataset:

- Caricamento e preprocessing dei dati sui prezzi immobiliari.
- Gestione dei valori mancanti, codifica delle variabili categoriche e normalizzazione/scalatura dei dati.

### 2. Implementazione dei Modelli di Regressione:

- **Ridge Regression:** Implementazione e addestramento del modello con regolarizzazione Ridge.
- **Lasso Regression:** Implementazione e addestramento del modello con regolarizzazione Lasso.
- **Elastic Net Regression:** Implementazione e addestramento del modello con regolarizzazione Elastic Net.

### 3. Valutazione delle Performance:

- Utilizzo di tecniche di validazione incrociata.
- Calcolo del Mean Squared Error (MSE) per ciascun modello.
- Confronto della complessità dei modelli valutando il numero di coefficienti non nulli.
- Analisi e confronto dei risultati dei vari metodi di regolarizzazione.

### 4. Visualizzazione dei Risultati:

- Creazione di grafici per visualizzare e confrontare le performance dei modelli.
- Visualizzazione della distribuzione dei residui per valutare l'adeguatezza del modello.

Il progetto deve includere il codice sorgente completo, con commenti dettagliati che spiegano i vari passaggi, le scelte fatte e i risultati ottenuti, per garantire trasparenza e replicabilità del lavoro.

## Il dataset

Il dataset è disponibile qui: <https://proai-datasets.s3.eu-west-3.amazonaws.com/housing.csv> (liberamente tratto dal seguente dataset: <https://www.kaggle.com/datasets/yasserh/housing-prices-dataset>)

- **Price:** il prezzo, il target da prevedere
- **Area:** superficie dell'immobile
- **Bedrooms:** numero di camere da letto
- **Bathrooms:** numero di bagni
- **Stories:** numero di piani
- **Mainroad:** vale 1 se l'immobile affaccia su una strada principale, 0 altrimenti
- **guestroom:** vale 1 se l'immobile ha una stanza degli ospiti, 0 altrimenti
- **basement:** vale 1 se l'immobile ha un seminterrato, 0 altrimenti
- **hotwaterheating:** vale 1 se l'immobile ha una caldaia, 0 altrimenti
- **airconditioning:** vale 1 se l'immobile ha l'aria condizionata, 0 altrimenti
- **parking:** numero di parcheggi
- **prefarea:** vale 1 se l'immobile è in una zona prestigiosa, 0 altrimenti
- **Furnishingstatus:** vale 0 se l'immobile non è arredato, 1 se è parzialmente arredato, 2 se è completamente arredato

**Modalità di consegna:** Link pubblico a notebook di Google Colab

