



Comune di Parma

BUSINESS INTELLIGENCE PER LE POLITICHE URBANE LOCALI

-

VERSO PARMA SMART AND WISE

A cura di
Avv. Pasquale Criscuolo
Dott.ssa Lara Berzieri
Dott. Massimiliano Ferrari

Direzione Generale
Comune di Parma

GRUPPO DI LAVORO



□ COMUNE DI PARMA

- Criscuolo Pasquale - Direttore Generale del Comune di Parma;
- Lara Berzieri - Responsabile Ufficio Statistica e Docente a contratto di «Introduction to Statistics» a Parma;
- Massimiliano Ferrari – Tirocinante Ufficio Statistica.

□ CENTRO DI RICERCA Robust Statistics Academy di Parma:

- Fabrizio Laurini - Responsabile scientifico del progetto e Professore Ordinario di Statistica Economica;
- Marco Riani - Direttore Ro.S.A. e Professore Ordinario di Statistica;
- Aldo Corbellini - Professore Associato di Statistica Economica;
- Gianluca Morelli - Ricercatore di Statistica Economica.

OBIETTIVI DEL PROGETTO

- **Migliorare la qualità e la disponibilità dell'informazione statistica sul territorio ai fini dello svolgimento delle funzioni affidate alle autonomie locali;**
- **Rafforzare la funzione statistica come strumento di «data science» per trasformare i dati in decisioni e orientare gli amministratori;**
- **Promuovere la cultura statistica nell'organizzazione e nell'intera collettività;**
- **Attivare leve di *governance* mirate al benessere collettivo**

FASI DEL PROGETTO	INIZIO	FINE
Identificazione variabili esistenti disponibili e setting dei "case study" prioritari	set-23	dic-23
Ricerca banche dati esterne	ott-23	apr-24
Integrazione base dati - armonizzazione indicatori, ricerca e identificazione di anomalie	nov-23	mag-24
Analisi dati e taratura ottimale di modelli di "statistical learning" (performance e usabilità)	nov-23	mar-25
Discussione risultati ottenuti - ricalibrazione modelli, scelta indicatori significativi e tuning dei parametri	dic-23	mar-25
Identificazione e progettazione soluzione hardware/cloud	mar-24	mar-25
Produzione e integrazione software in house	apr-24	dic-25
Realizzazione web application e dashboard	lug-24	dic-25

PARMA CITY SMART AND WISE



Comune di Parma

POPOLAZIONE

TERRITORIO

ECONOMIA

AMBIENTE

MOBILITA'

GOVERNANCE

GIUSTIZIA



HARDWARE & SOFTWARE

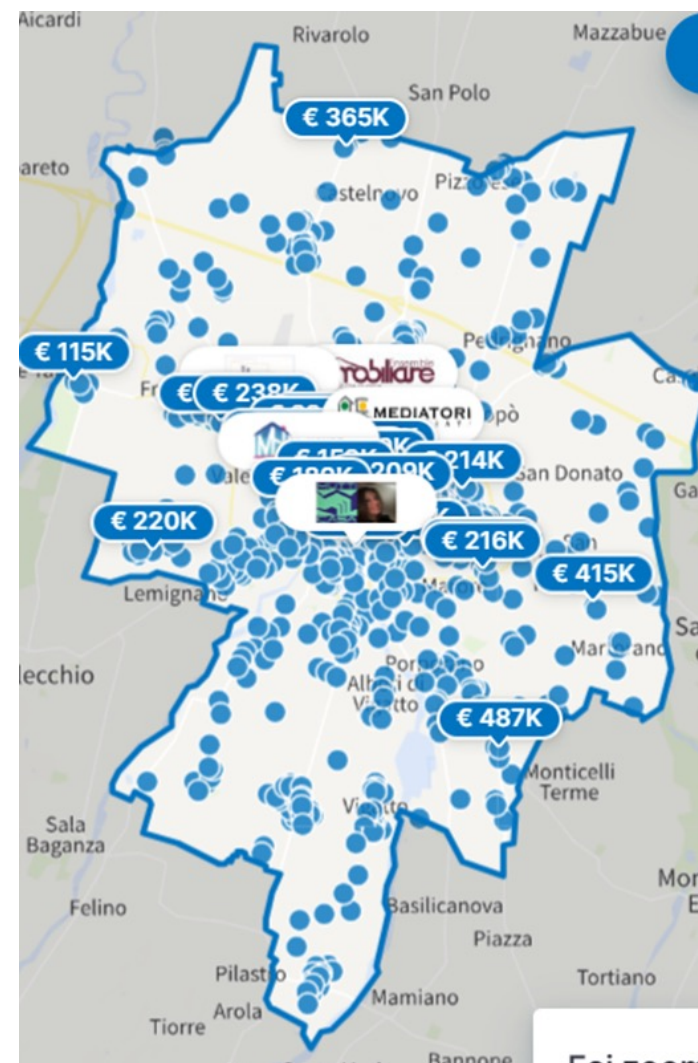
➤ **DATALAKE – SQL**

➤ **POWER B. I.**

➤ **MATLAB - FSDA**

MATLAB CASE STUDIES: IL MERCATO IMMOBILIARE ABITATIVO LOCALE

- 2000 unità statistiche: annunci nel territorio comunale al primo agosto 2024
- 90+ variabili: ex. anno di costruzione, locali, quartiere, tipologia ...
- Obiettivo a breve termine (predittivo): pricing in funzione delle caratteristiche dell'abitazione
- Obiettivo a medio-lungo termine: costruzione di serie storiche sul mercato immobiliare del comune





Fasi

1. Webscraping dal sito di annunci
2. Creazione del dataset e scelta delle variabili
3. Imputazione di alcuni missing ed eliminazione di alcune unità statistiche: da 2000 a circa 1500 immobili
4. EDA e rilevazione di potenziali unità anomale
5. Fase modellistica
6. Indagine sui residui



Modelli impiegati

Nell'ottica del pricing dell'immobile si sono adottate 4 metodologie di statistical learning con le quali si riesce a spiegare fra il 70 e l'80 della varianza della variabile dipendente <prezzo>:

- Gradient boosting con alberi di decisione come learner implementata con le funzioni *templateTree*, *fitrensemble*
- Random Forest implementata con le funzioni *templateTree*, *fitrensemble*
- Lasso (Least Absolute Shrinkage and Selection Operator) implementato con la funzione *lasso*
- Group Lasso (senza una funzione predefinita)

Nel proseguo ci si concentrerà tuttavia su Lasso e GroupLasso

Lasso

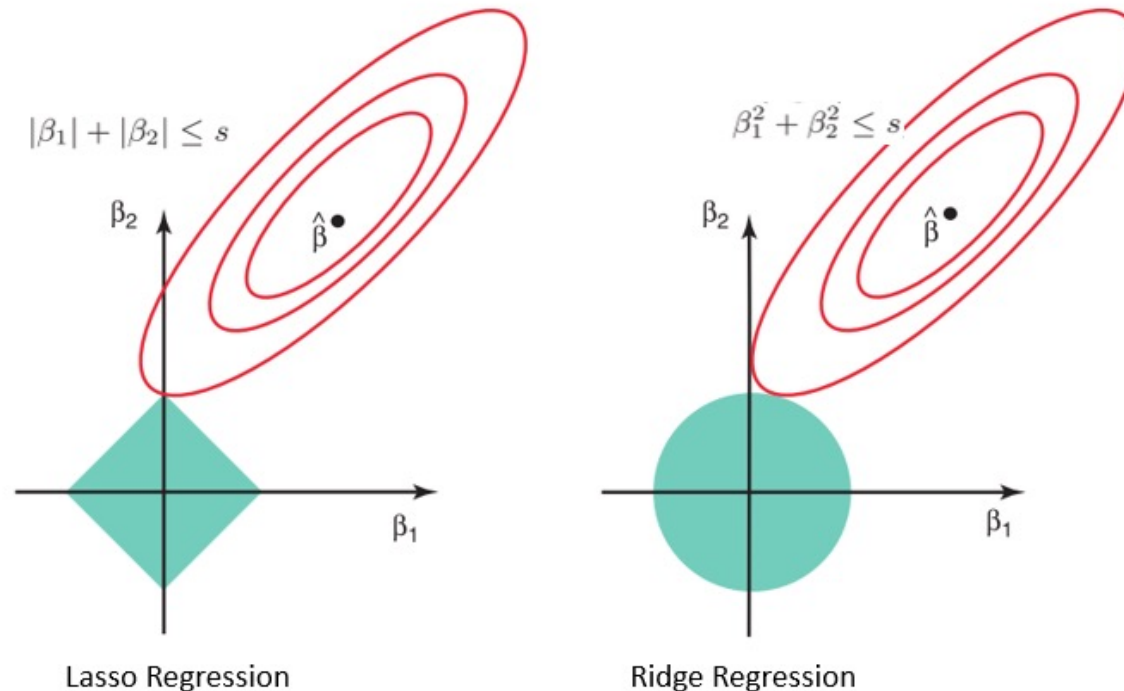
Problema. Affrontare il tema della sparsità e del sovradattamento per migliorare la performance su dati non ancora visti (possibilità di generalizzazione a nuovi dati)

- Solo una decina delle 91 variabili sono quantitative
- Le variabili categoriche (ad esempio quartiere, tipologia, stato di conservazione) sono codificate come variabili dummy $\{0,1\}$

Tecniche di Regolarizzazione

Soluzione. Inserire nel problema di ottimizzazione una penalità per la complessità del modello: l'iperparametro lambda (λ); maggiore è lambda, maggiore è il numero di parametri posti uguali a 0

$$\frac{1}{N} \sum_{i=1}^N (y_i - x_i \beta)^2 + \lambda \|\beta\|_1$$



Problema. Desideriamo che le variabili siano incluse o escluse a gruppi e non singolarmente, in altre parole non deve accadere ad esempio che la variabile CentroStorico sia inclusa e quella Oltretorrente no essendo entrambe variabili relative a <quartiere>

Group Lasso

$$\min_{\beta \in \mathbb{R}^d} \left\{ \frac{1}{2} \sum_{i=1}^N (y_i - x_i \beta)^2 + \lambda a \sum_{g=1}^G \|\beta_g\|_2 \right\}$$

Metodo del gradiente prossimale

- Inizializzazione dei parametri

Ciclo:

1. Calcolo del gradiente della funzione di errore quadratico
2. Aggiornamento dei parametri: $\text{new_params} = \text{old_params} - \text{step_size} * \text{gradient}$
3. Applicazione dell'operatore prossimale al vettore di parametri di ogni gruppo
4. ...
5. Fino a convergenza alla soluzione

$$\text{proximal_operator}(\beta_g) = \underset{\beta_{g\text{new}} \in \mathbb{R}^d}{\text{argmin}} \left\{ \frac{1}{2} \|\beta_{g\text{new}} - \beta_g\|_2^2 + \lambda a \|\beta_{g\text{new}}\|_2 \right\}$$

$$\text{proximal_operator}(\beta_g) = \max \left\{ 0, 1 - \frac{\lambda a}{\|\beta_g\|_2} \right\} \beta_g$$

Abbiamo quindi due casi:

- Se $\lambda a > \|\beta_g\|_2$, i parametri del gruppo sono posti uguale a 0
- Se $\lambda a < \|\beta_g\|_2$, i parametri sono aggiornati sottraendo $\frac{\lambda a}{\|\beta_g\|_2} \beta_g$ mantenendo la direzione



Confronto sul test set

