

Valutazione del danno alluvionale mediante strumenti di Statistica Multivariata e Machine Learning

Simone Montanari^(a), Christian Natale Gencarelli^(b), Simone Sterlacchini^(b), Maurizio Vichi^(c)

a) Tesista triennale Dipartimento Scienze Statistiche, La Sapienza, Roma – Volontario Information Management Team, Croce Rossa Italiana


b) CNR – Istituto Geologia Ambientale e Geoingegneria, Milano


c) Dipartimento Scienze Statistiche, La Sapienza, Roma

Presentazione

Chi sono

Simone Montanari

➤ laureando triennale in Scienze
Statistiche presso l'Università
di Roma La Sapienza  SAPIENZA
UNIVERSITÀ DI ROMA

➤ volontario per Information
Management Team di Croce
Rossa Italiana 

Croce Rossa Italiana

Perché questo lavoro

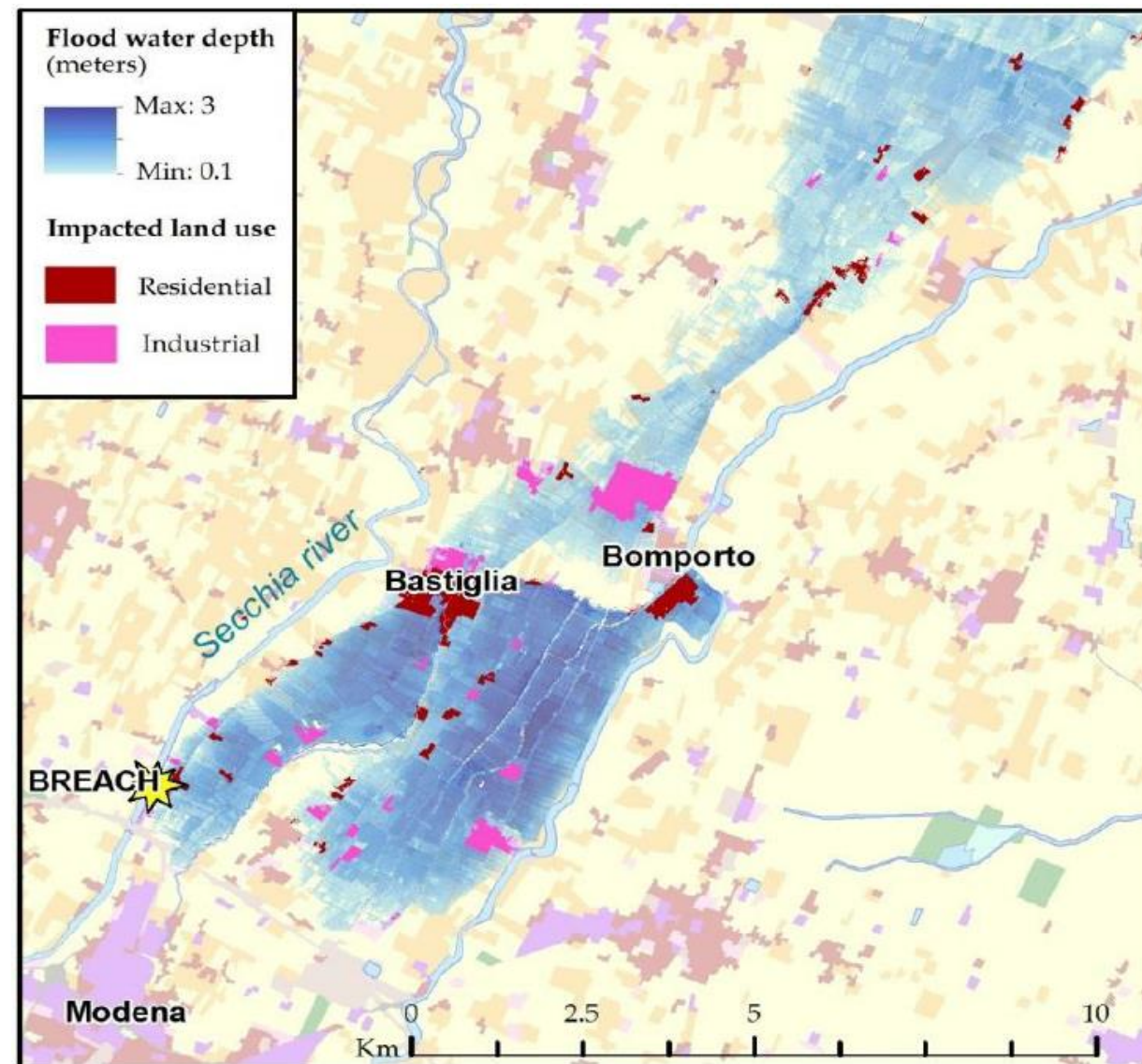
Il progetto nasce grazie a
un'intesa tra il laboratorio
LARGE del CNR-IGAG e
Croce Rossa Italiana



Di cosa stiamo parlando?

Alluvione del 19 gennaio 2014 nei comuni di Bastiglia e Bomporto (Modena)

- 52 km² invasi da 37 milioni di m³ d'acqua in meno di 30 ore
- Circa 1500 sfollati
- Circa 500 milioni di euro di danni



Fonte: Amadio et al., Improving flood damage assessment models in Italy, 2016

I dati

COMUNE DI BASTIGLIA
n. progressivo scheda B: 76

SCHEDA B

Ricognizione del fabbisogno per il
ripristino del patrimonio edilizio privato, dei beni mobili e dei
beni mobili registrati

REGIONE EMILIA ROMAGNA

EVENTI ALLUVIONE DEL ____19/____01____/____2014____

SEGNALAZIONE E QUANTIFICAZIONE DEL DANNO
(Autocertificazione ai sensi del D.P.R. 445/2000)

COMUNE DI ____BASTIGLIA____ PROVINCIA ____MO____

Il sottoscritto _____

nato/a a _____ il _____

residente a BASTIGLIA _____ CAP ____41030 Indirizzo _____

Tel. _____; Cell. _____; Fax. _____

consapevole delle conseguenze penali previste dall'art.76 del D.P.R. 445/2000 per le falsità in atti e le dichiarazioni mendaci

**DICHIARA
SOTTO LA PROPRIA RESPONSABILITA'**

1) che l'immobile è ubicato in

via / viale / piazza _____ n. civico: _____

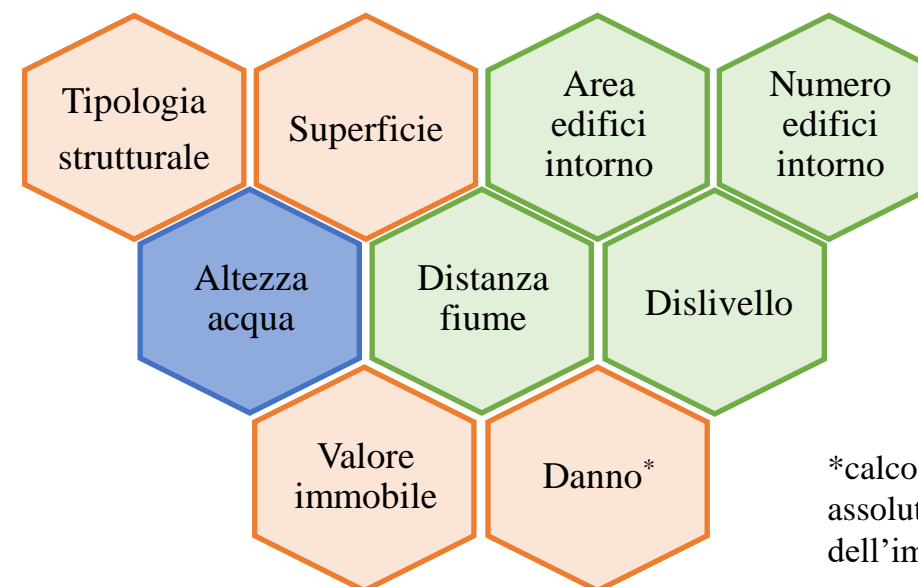
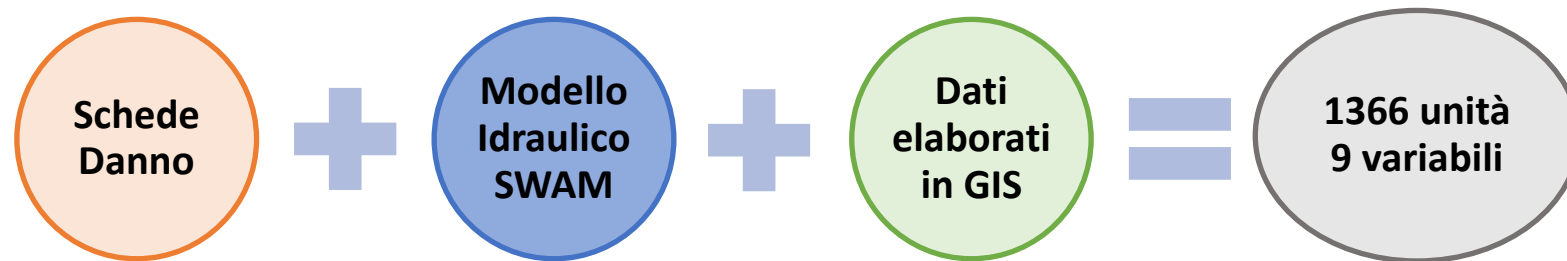
località: ____BASTIGLIA____ CAP ____41030____

L'immobile è

☒ di proprietà ☐ in comproprietà
(nome del comproprietario: _____)

☐ altro diritto reale di godimento (specificare: _____)

COMUNE DI BASTIGLIA

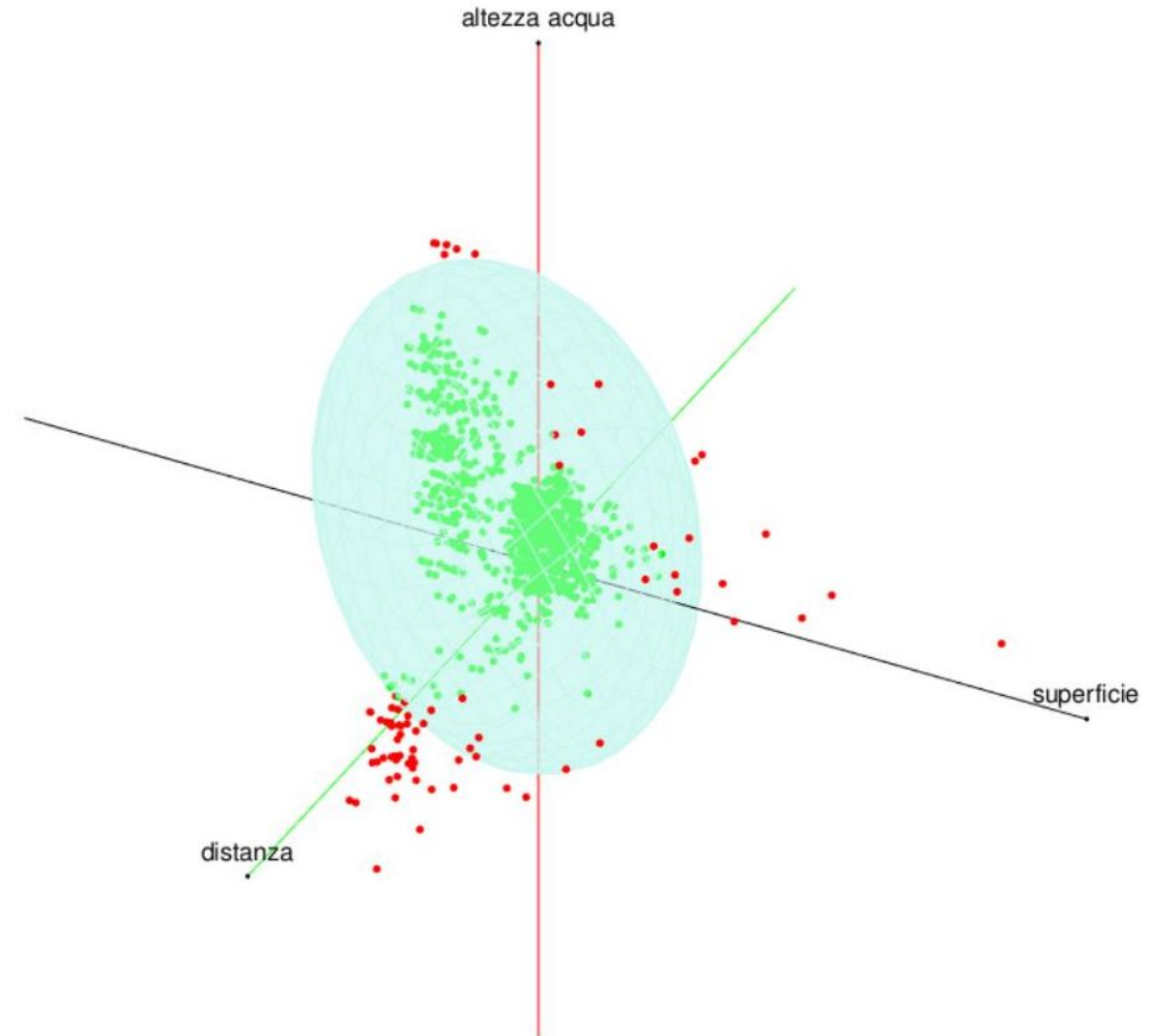
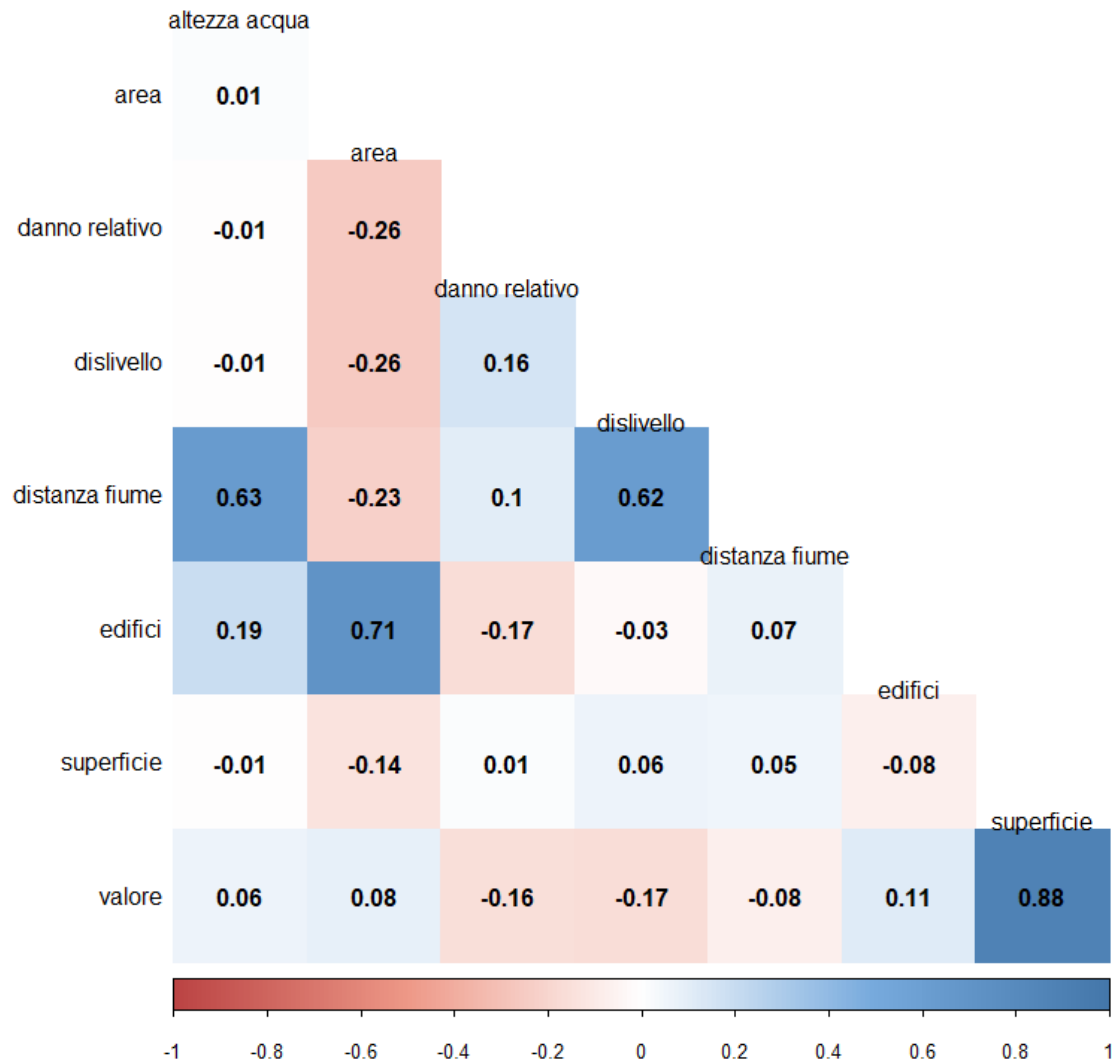


*calcolato come danno
assoluto/valore
dell'immobile

I dati

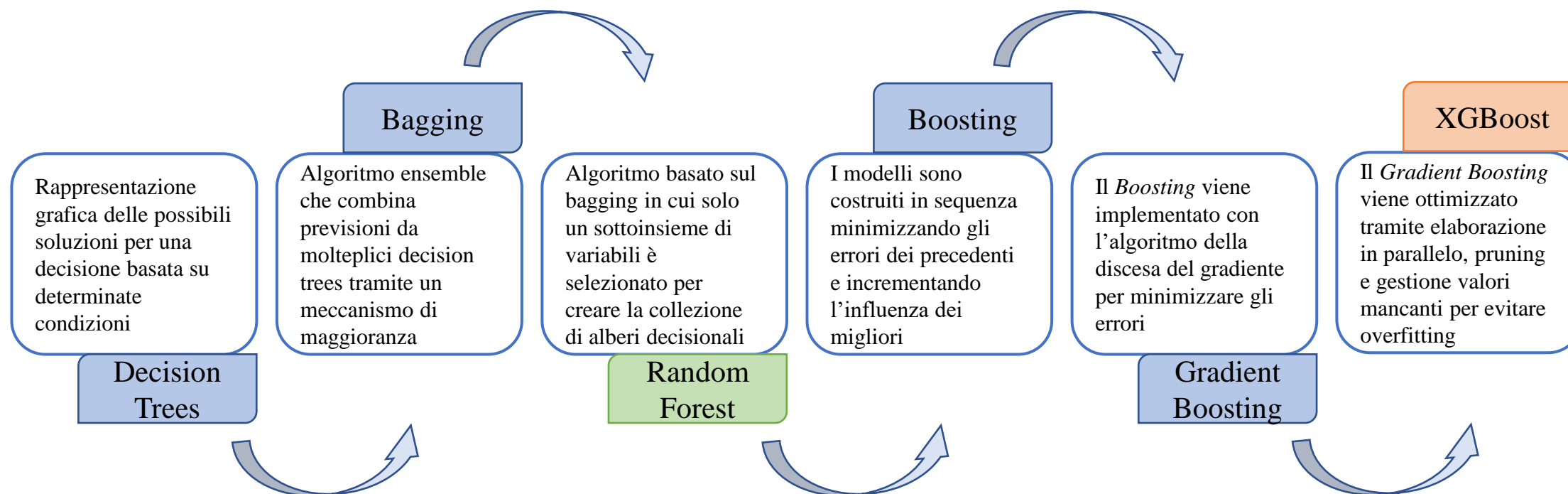
	A	B	C	D	E	F	G	H	I
1	altezza dell'acqua	distanza dal fiume	dislivello rispetto gli argini	tipologia strutturale	superficie	area edifici	# edifici	valore dell'immobile	danno relativo
2	0.39	1106.13	10.58	muratura	58.00	2222.00	7.00	68150.00	0.00
3	0.39	1106.13	10.58	muratura	70.00	2222.00	7.00	82250.00	0.00
4	0.55	1103.58	10.99	muratura	40.00	1155.00	6.00	47000.00	0.23
5	0.44	1113.50	10.76	muratura	30.00	2409.00	5.00	35250.00	0.18
6	0.55	1109.58	10.61	cemento	200.00	1155.00	6.00	235000.00	0.03
7	0.77	900.32	10.80	muratura	110.00	2135.00	9.00	129250.00	0.26
8	0.67	979.88	11.00	cemento	120.00	1041.00	5.00	141000.00	0.11
9	0.73	901.15	10.81	cemento	150.00	1695.00	8.00	176250.00	0.02
10	0.81	992.80	11.14	cemento e muratura	60.00	2771.00	7.00	70500.00	0.00
11	0.59	951.41	10.80	muratura	84.00	2098.00	10.00	98700.00	0.00
12	0.59	951.41	10.80	cemento	50.00	2098.00	10.00	58750.00	0.01
13	0.59	951.41	10.80	muratura	80.00	2098.00	10.00	94000.00	0.01
14	0.59	951.41	10.80	cemento e muratura	80.00	2098.00	10.00	94000.00	0.01

Analisi descrittiva

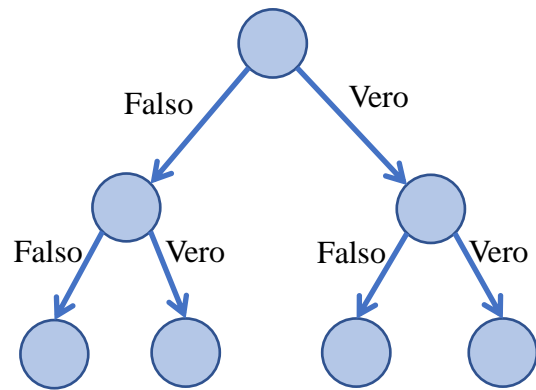


Valori anomali (288 su 1366)

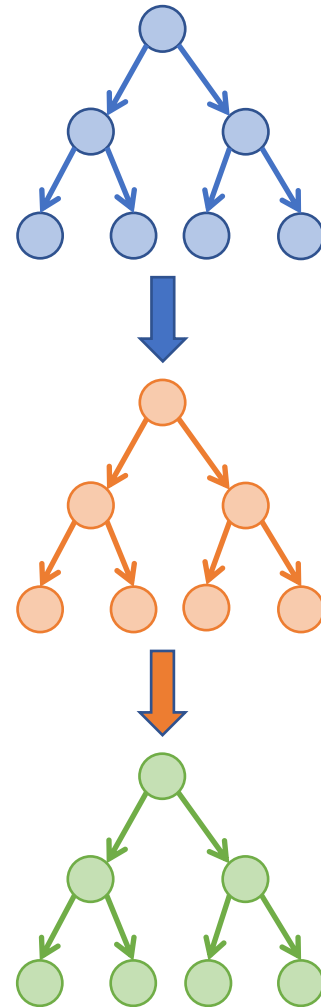
Regressione: Alberi decisionali



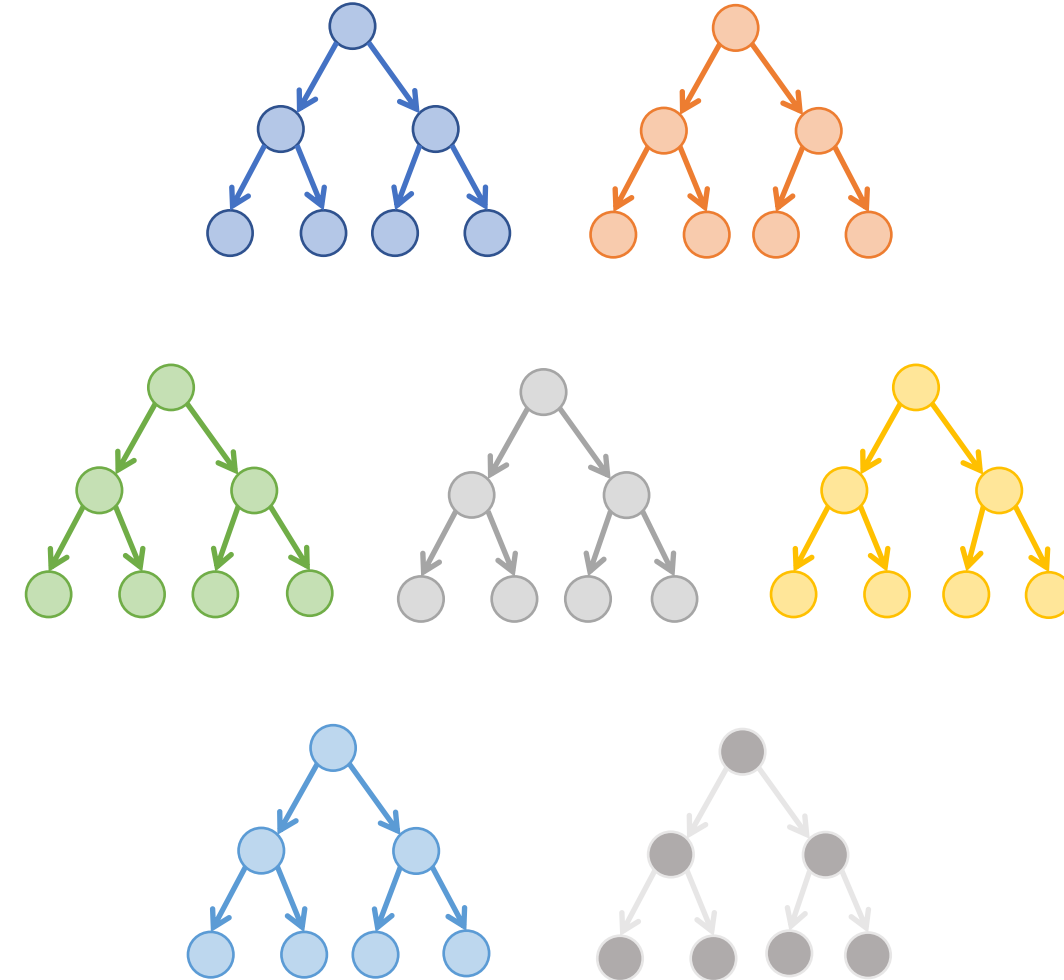
Regressione: RF vs XGBoost



Decision Tree



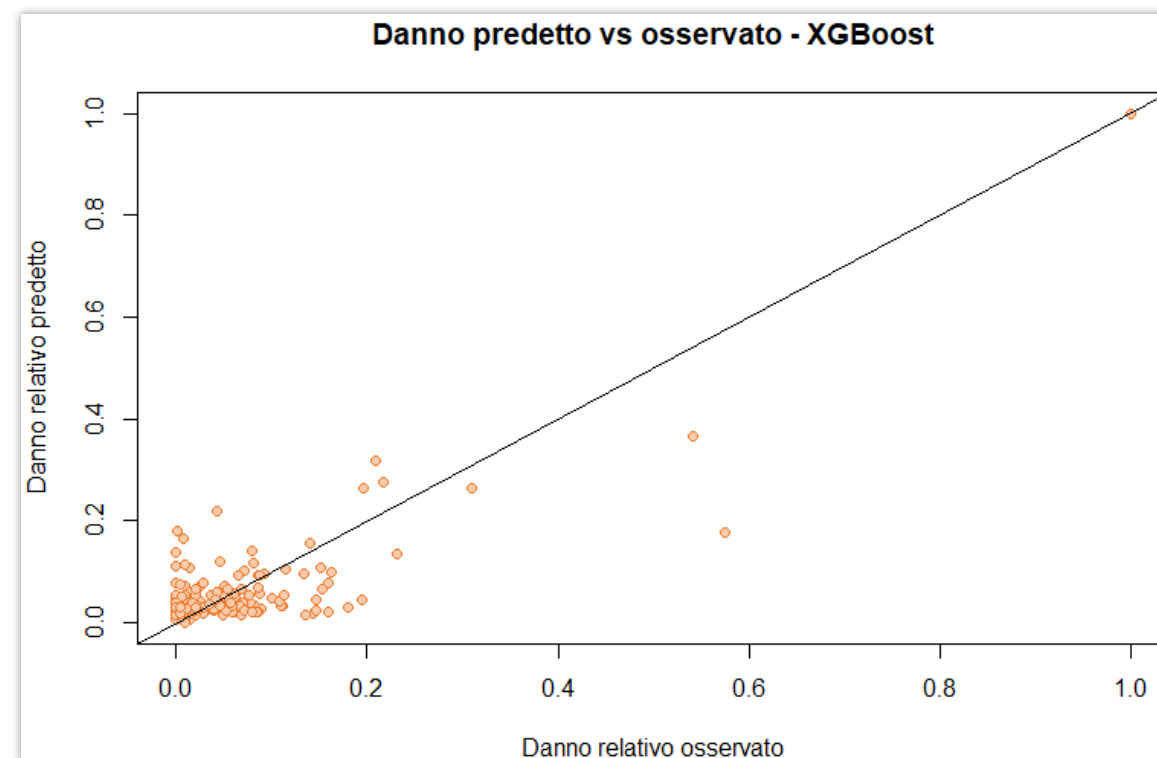
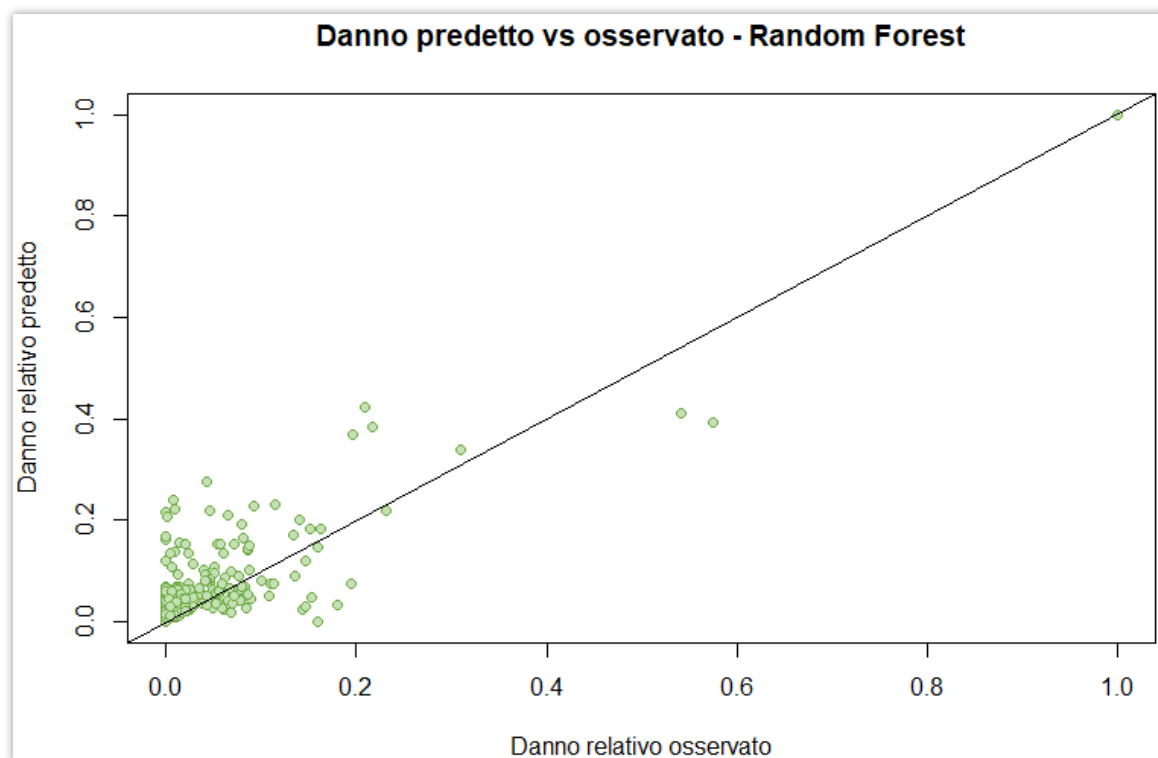
Gradient Boosted Tree



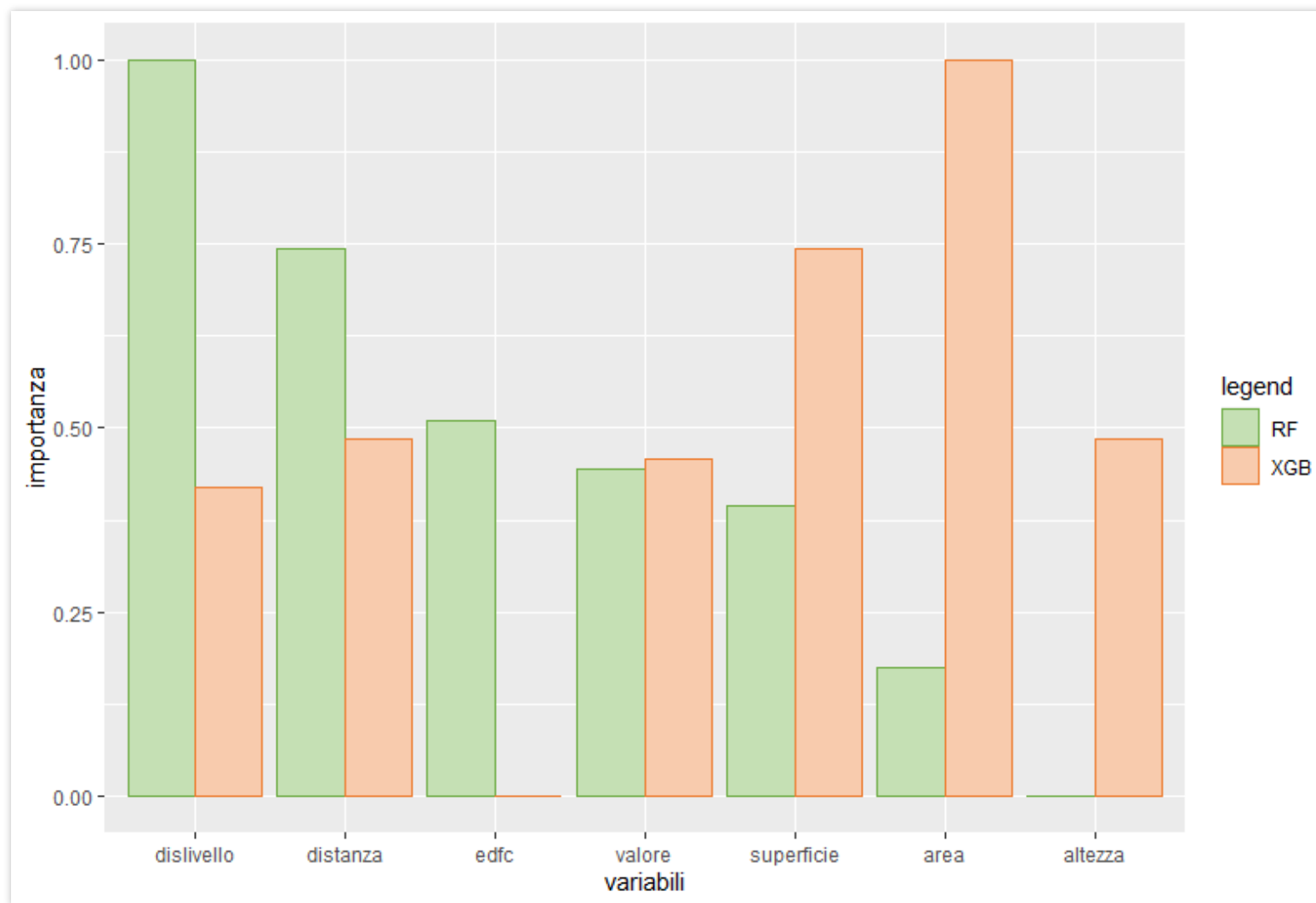
Random Forest

Regressione: Risultati

	R^2	Random Forest	XGBoost
85%	Train set	0,78	0,75
15%	Test set	0,64	0,67



Regressione: Variabili



Importanza delle variabili nei due modelli

Regressione: Regole decisionali

Random
Forest

Distanza fiume ≤ 4.360 mt & dislivello ≤ 11 mt & area > 495 mq & edifici $> 1,50$ & valore immobile > 53.400 €

Previsione: 0,08

$1,05$ mt $<$ altezza acqua $\leq 1,30$ mt & dislivello > 11 mt

Previsione: 0,88

XGBoost

Superficie ≤ 220 mq & 30.500 € $<$ valore immobile ≤ 370.000 €

Previsione: 0,10

Superficie $\leq 68,5$ mq & altezza acqua $> 0,13$ mt & area ≤ 85 mq

Previsione: 0,86

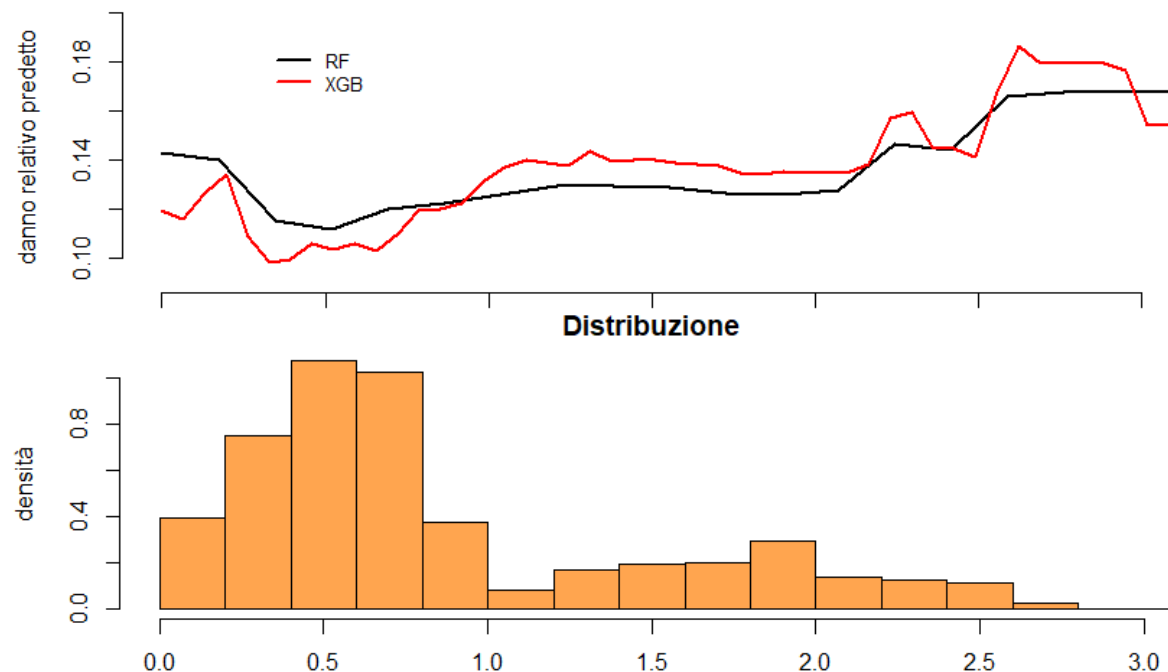
Esempio di interpretazione:

Gli alberi del modello Random Forest che rispettano la prima condizione restituiscono come output una previsione in media pari a 0,08, ossia un danno pari all' 8% del valore dell'immobile

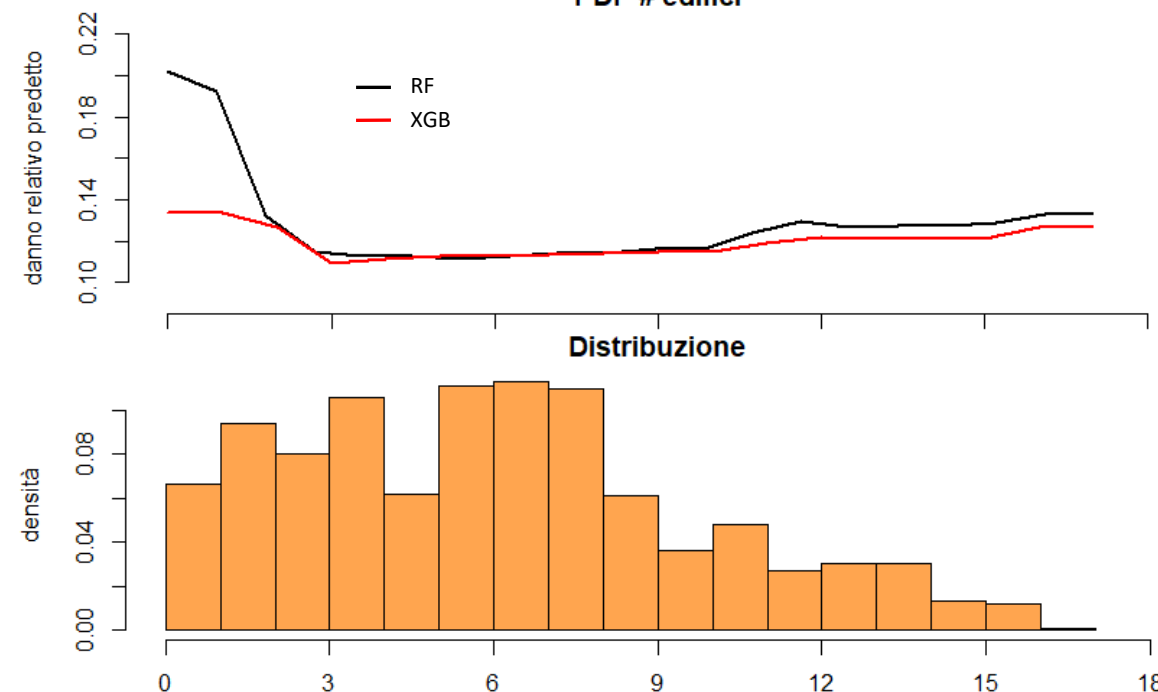
Regressione: PDP

Il *Partial Dependence Plot (PDP)* mostra l'effetto marginale di una variabile sul risultato previsionale medio di un modello di machine learning

PDP altezza acqua



PDP # edifici



Clustering

Approccio

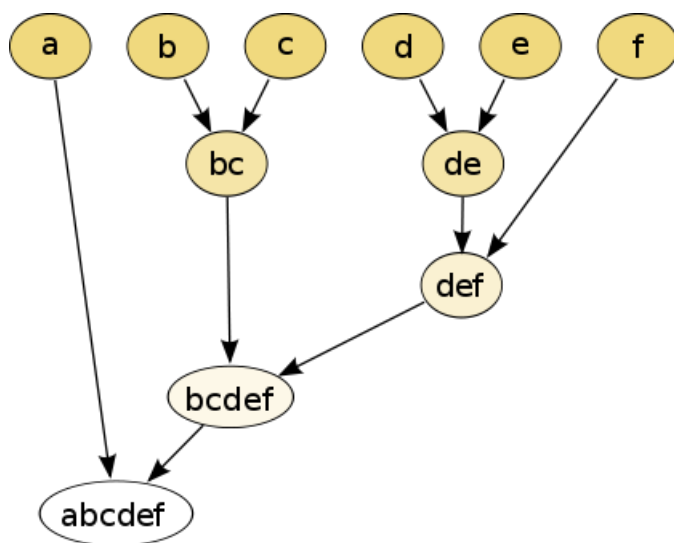
Gerarchico
Agglomerativo

Metrica

Metodo di Ward
(minimizzazione
varianza intra-cluster)

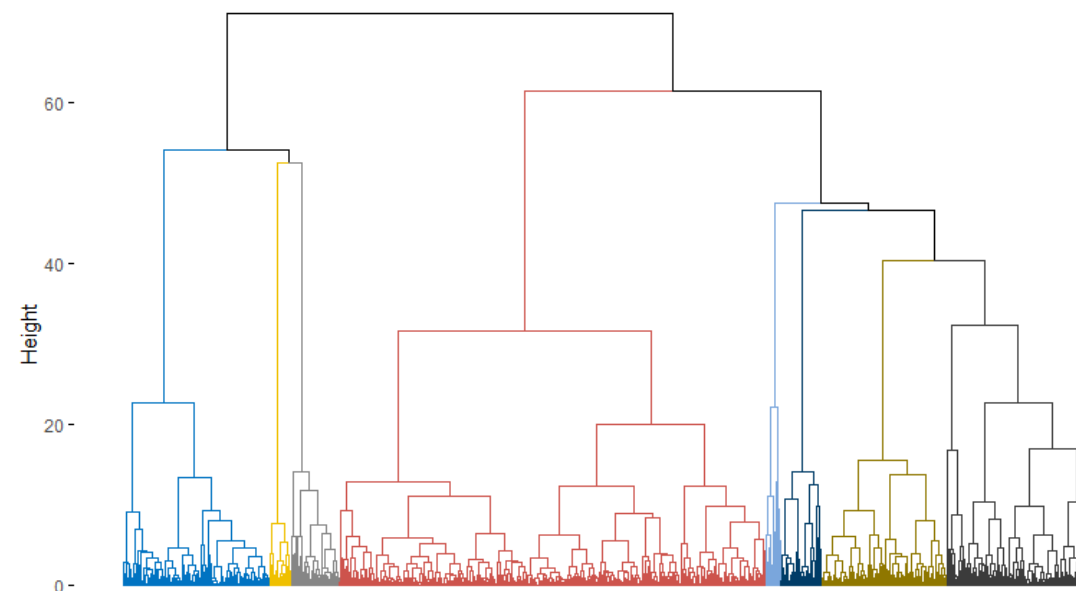
Risultato

8 Cluster



Fonte: Wikipedia

Cluster Dendrogram



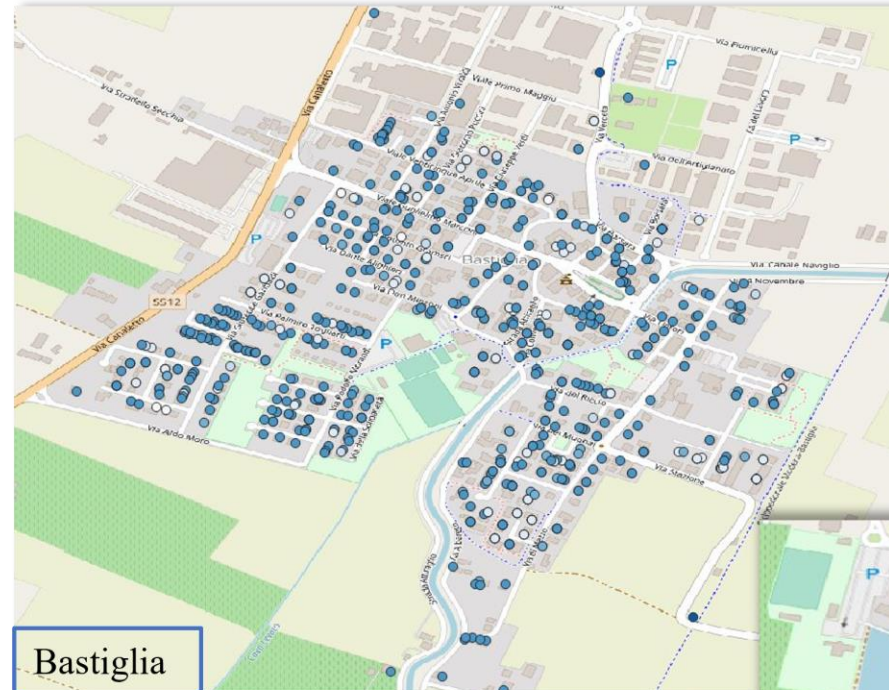
Clustering: Mappa



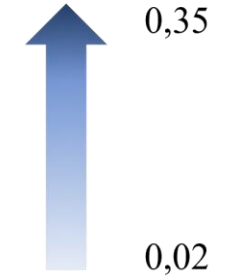
Consiglio Nazionale
delle Ricerche



Fiume Secchia



Danno medio



Clustering: Impatto variabili

	Varianza nei cluster (WSS)	Varianza tra cluster (GSS)	Pseudo F*
Distanza fiume	286	1079	731
Dislivello	293	1072	711
Altezza acqua	548	817	289
Valore immobile	726	639	171
Superficie	822	543	128
Area edifici intorno	964	401	81
Numero edifici intorno	1034	331	62
Danno	1131	234	40

$$*Pseudo F = \frac{gss/(k-1)}{wss/(n-k)} = \frac{gss}{wss} \frac{n-k}{k-1},$$

$k = \#cluster, n = \#osservazioni$

Variabili legate alla
posizione dell'edificio

Variabili legate alle
caratteristiche urbane

Conclusioni: Risultati



Importante ampliare il numero di variabili analizzate

Modelli ad albero restituiscono risultati soddisfacenti

Impatto delle variabili (importanza, regole, pdp)

Utilità del clustering

Valori Pseudo F

Conclusioni: E ora?



Schede di danno compilate da esperti



Creazione database unico a livello nazionale



Integrare nuove variabili



Approfondimento valori anomali



Miglioramento modelli ad albero



Modelli più avanzati



Analisi PDP avanzata



Approfondimento Pseudo-F

Grazie dell'attenzione!

Simone Montanari

simonemontanar@gmail.com