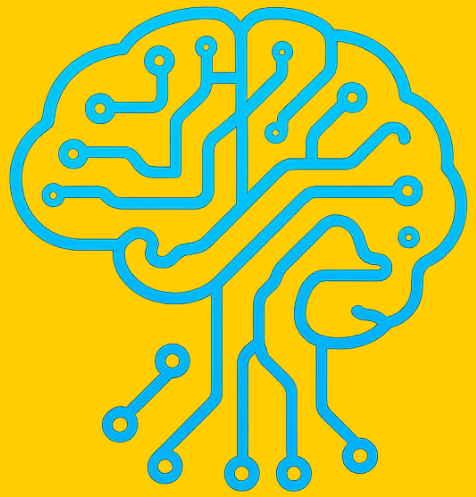
IZS TERAMO

ISTITUTO ZOOPROFILATTICO SPERIMENTALE DELL'ABRUZZO E DEL MOLISE "G. CAPORALE"

Approcci Machine Learning nell'identificazione dei fattori di rischio per talune malattie infettive in sanità animale.

Progetto Ricerca Corrente 05/22



Cerella Angelo
Mazzagufo Sabrina

IZS TERAMO
ISTITUTO ZOOPROFILATTICO SPERIMENTALE DELL'ABRUZZO E DEL MOLISE "G. CAPORALE"

Obiettivo

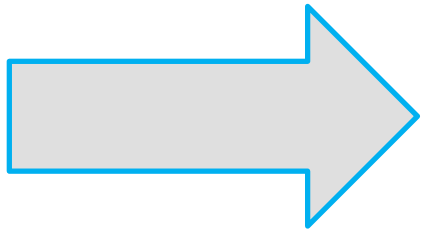
- Identificare il livello di rischio alle aziende per alcune malattie valutando una serie di dati presenti in VETINFO:
  - Brucellosi
  - Tubercolosi
- Sistemi coinvolti:
  - BDN
  - SANAN
  - SIMAN
  - FARMACO

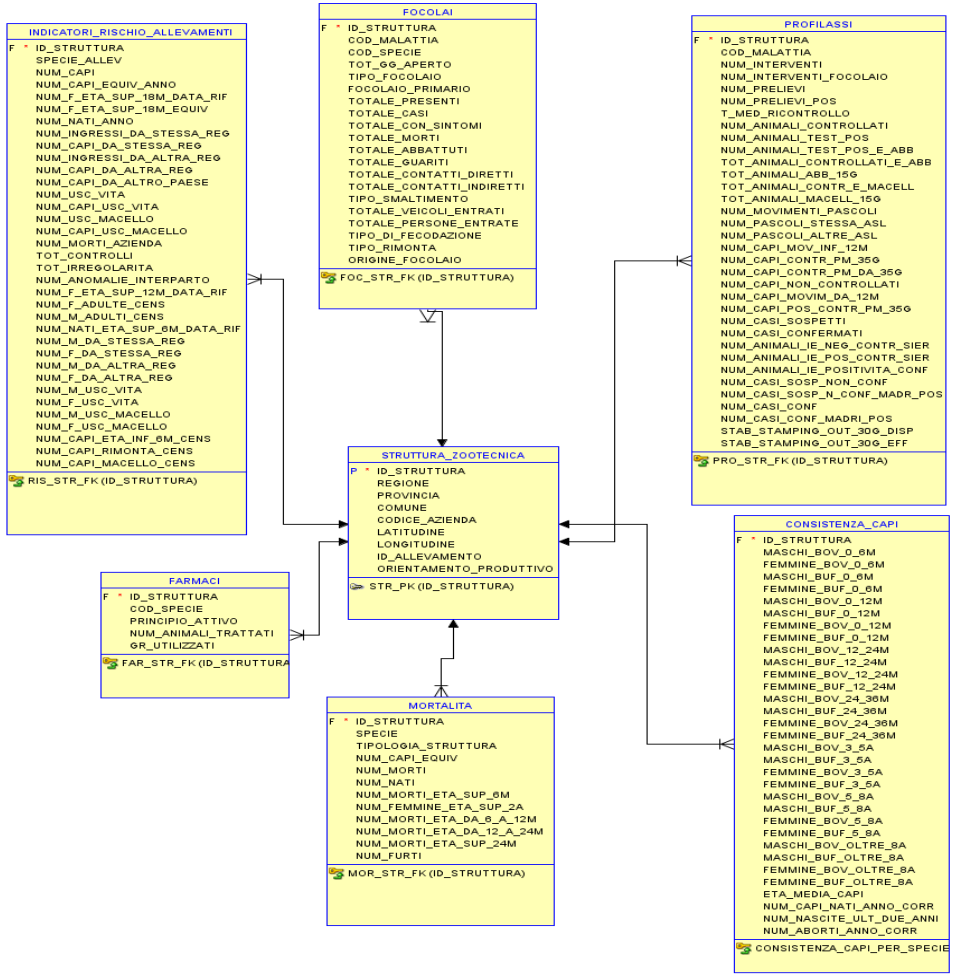
IZS

TERAMO

ISTITUTO ZOOPROFILATTICO SPERIMENTALE DELL'ABRUZZO E DEL MOLISE "G. CAPORALE"

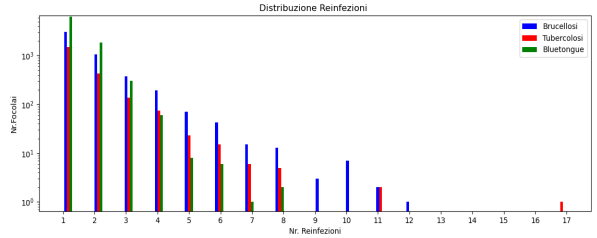
Data pipeline



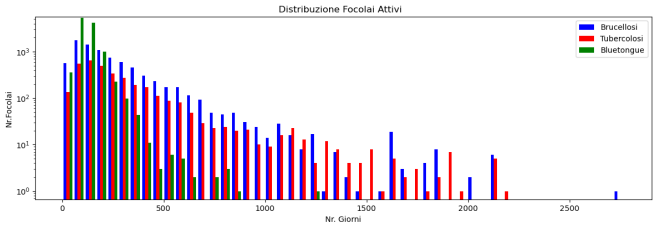


Data cleaning (outlier)

- Distribuzione del **numero di reinfezioni** per ciascuna malattia al fine di escludere le aziende con un valore di reinfezioni superiore al 97,5° percentile.
- Sono state quindi rimosse tutte le aziende il cui focolaio presentava una **durata** inferiore o uguale al 5° percentile oppure superiore o uguale al 95° percentile, poiché considerate valori anomali potenzialmente influenti sull'analisi.
- stati rimossi tutti i capi per i quali l'intervallo **interparto** risultava superiore al 97,5° percentile, in quanto considerati valori anomali potenzialmente distorsivi.



	Brucellosi	Tubercolosi	Bluetongue
Media	1.66	1.55	1.32
Mediana	1.0	1.0	1.0
Dev. Standard	1.18	1.1	0.61
Min	1.0	1.0	1.0
Max	12.0	17.0	8.0
Totale casi	4814.0	2190.0	8527.0
Percentile 2.5	1.0	1.0	1.0
Percentile 5	1.0	1.0	1.0
Percentile 10	1.0	1.0	1.0
Percentile 90	3.0	3.0	2.0
Percentile 95	4.0	4.0	2.0
Percentile 97.5	5.0	4.0	3.0



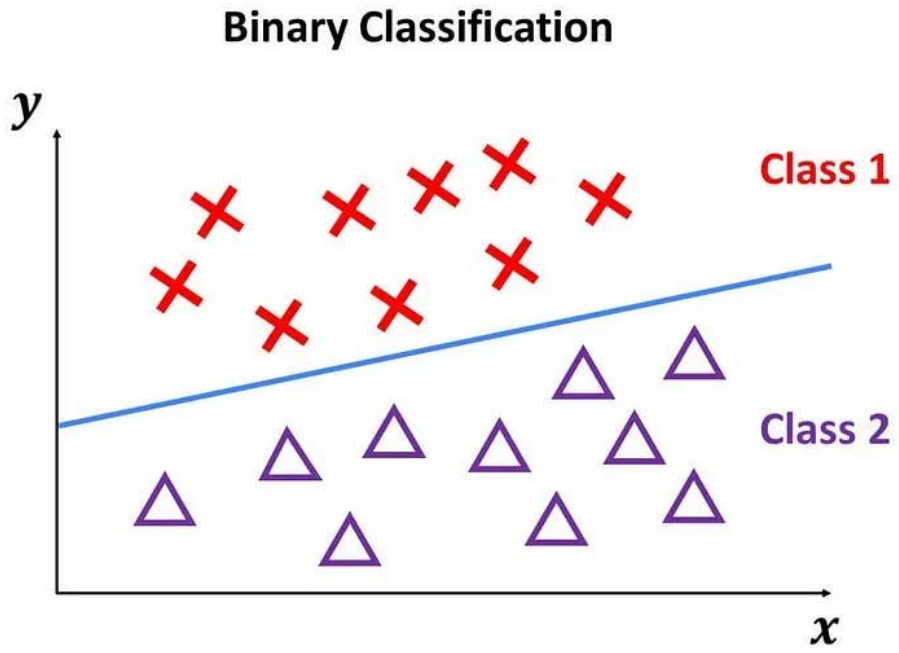
	Brucellosi	Tubercolosi	Bluetongue
Media	251.28	294.56	116.32
Mediana	177.0	204.0	110.0
Dev. Standard	236.85	283.38	55.96
Min	0.0	0.0	0.0
Max	2772.0	2205.0	1256.0
Totale casi	7989.0	3387.0	11267.0
Percentile 2.5	37.0	34.65	51.0
Percentile 5	49.0	61.0	60.0
Percentile 10	63.0	88.0	61.0
Percentile 90	526.0	575.4	172.0
Percentile 95	690.0	877.0	198.0
Percentile 97.5	871.3	1145.0	259.0

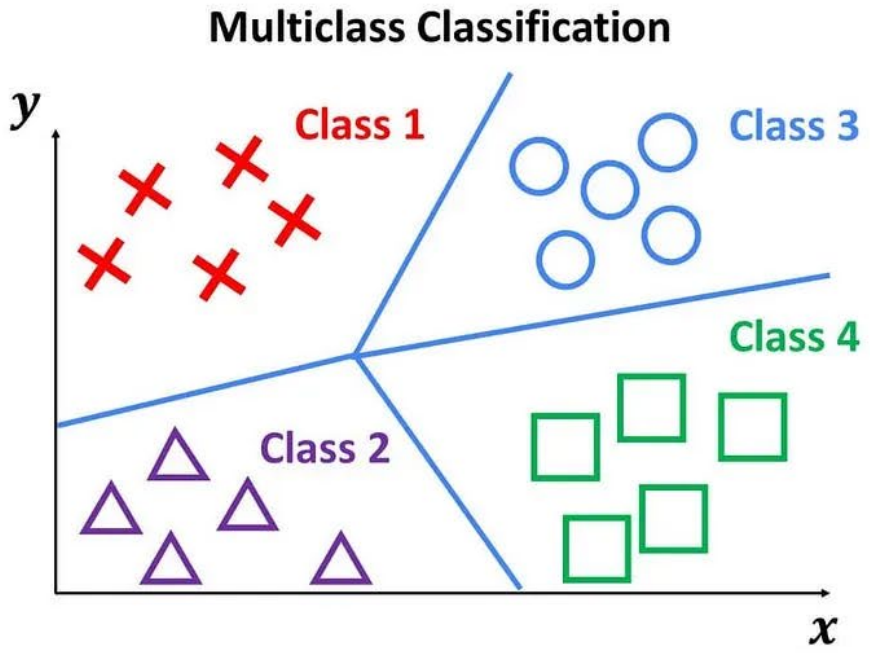
IZS TERAMO

ISTITUTO ZOOPROFILATTICO SPERIMENTALE DELL'ABRUZZO E DEL MOLISE "G. CAPORALE"

Problema di Classificazione

- A Rischio (con valore 1)
- Non a Rischio (con valore 0)



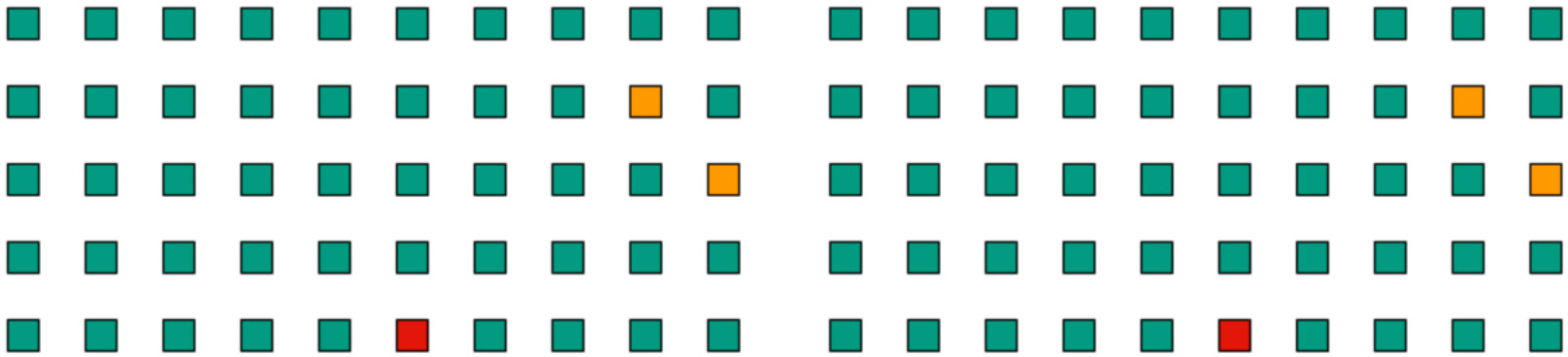


IZS TERAMO

ISTITUTO ZOOPROFILATTICO SPERIMENTALE DELL'ABRUZZO E DEL MOLISE "G. CAPORALE"

Definizione classi

- Quali sono le aziende a Rischio?
  - Hanno avuto almeno 1 focolaio;
  - Il numero di animali positivi è superiore al 95° percentile rispetto al totale degli animali presenti in azienda;
  - Il numero totale di morti è superiore al 95° percentile rispetto al totale degli animali presenti in azienda;
  - Il numero di animali trattati è superiore al 95° percentile rispetto al totale degli animali presenti in azienda;



IZS
TERAMO
ISTITUTO ZOOPROFILATTICO SPERIMENTALE DELL'ABRUZZO E DEL MOLISE "G. CAPORALE"

Modelli analizzati

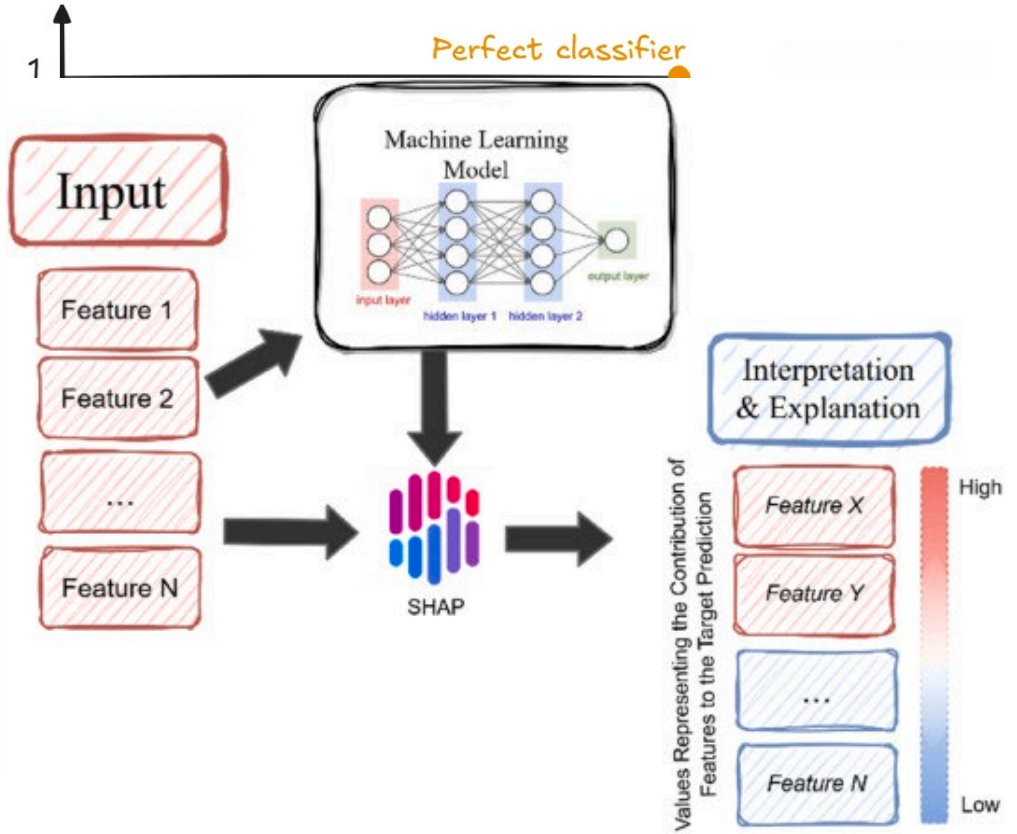
- **Logistic Regression:** stima la probabilità che un dato appartenga a una determinata classe.
- **K-Nearest Neighbors** (**KNN**): classifica la nuova previsione sulla base dei K punti più vicini nel dataset, secondo una misura di similarità (generalmente la distanza euclidea)
- **Support Vector Machine** (**SVM**): cerca di individuare un confine (linea o iperpiano) che separi al meglio le due classi.
- **Random Forest:** è un algoritmo che basa la sua previsione su più alberi decisionali.
- **CATBoost** e **XGBoost:** algoritmi che costruiscono una serie di alberi decisionali, uno dopo l'altro dove ogni nuovo albero cerca di correggere gli errori fatti dagli alberi precedenti

IZS TERAMO

ISTITUTO ZOOPROFILATTICO SPERIMENTALE DELL'ABRUZZO E DEL MOLISE "G. CAPORALE"

Valutazione dei Risultati

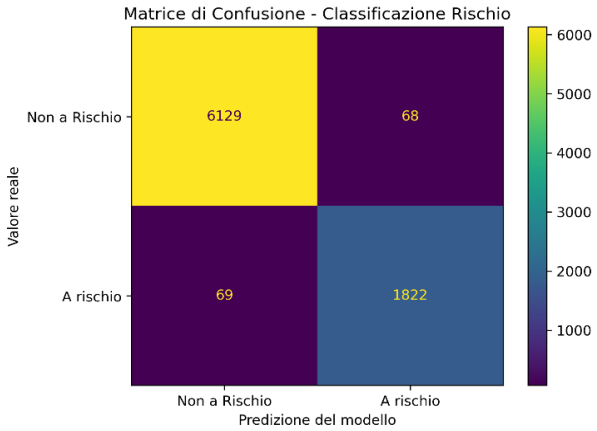
- **Matrice di confusione**
- **Curva ROC**
- **Precision-Recall Curve**
- **SHAP (SHapley Additive exPlanat**

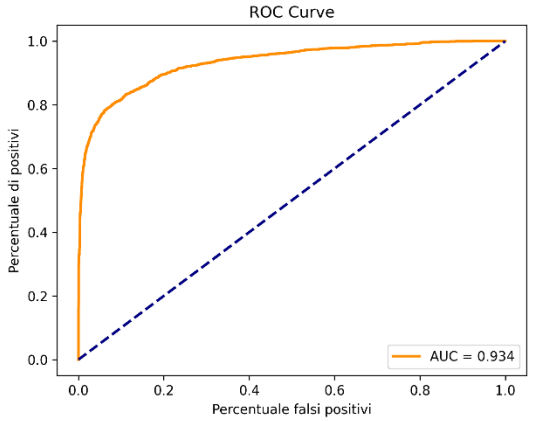


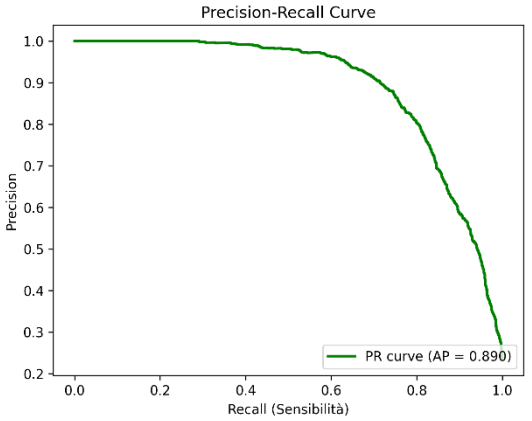
IZS TERAMO

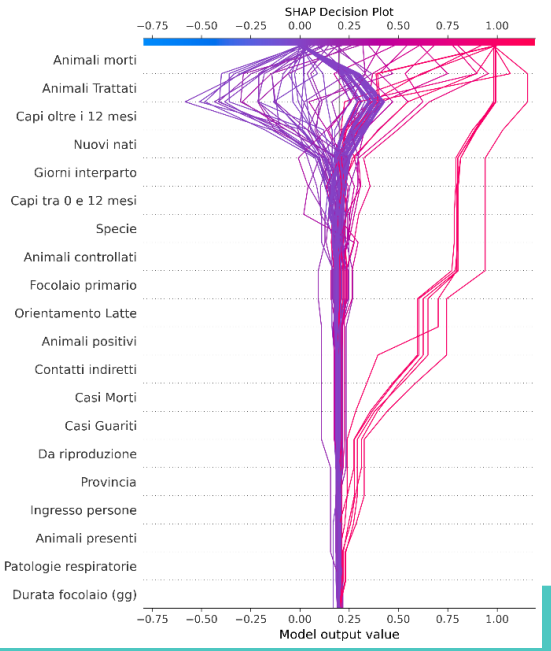
ISTITUTO ZOOPROFILATTICO SPERIMENTALE DELL'ABRUZZO E DEL MOLISE "G. CAPORALE"

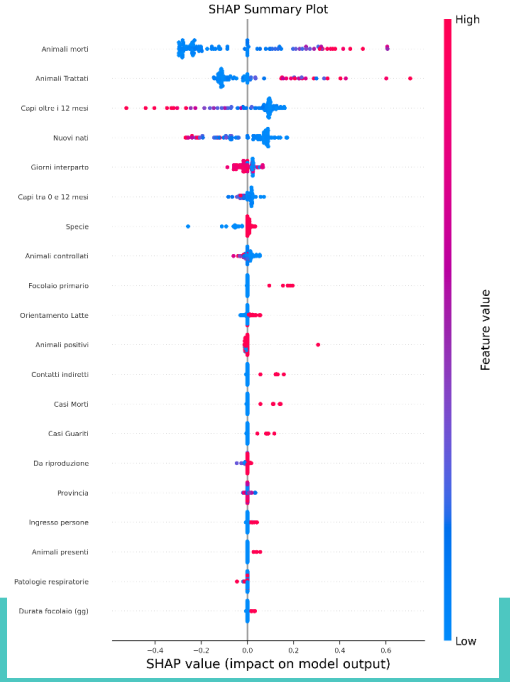
Valutazione dei Risultati

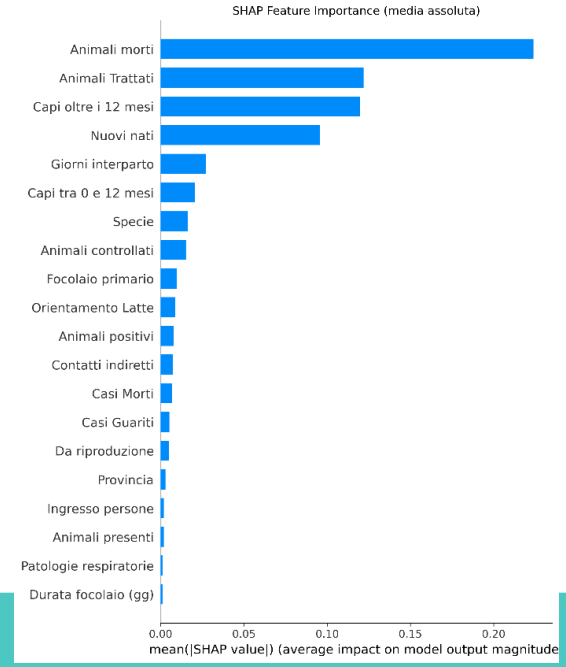












IZS TERAMO

ISTITUTO ZOOPROFILATTICO SPERIMENTALE DELL'ABRUZZO E DEL MOLISE "G. CAPORALE"

Matrice di Correlazione

- Rappresentazione della dipendenza tra features.
- Eliminazione di una delle variabili con coefficiente di correlazione superiore all'85%:
  - esclusione la variabile di minore importanza;
  - le variabili con un numero elevato di valori ricostruiti durante il pre-processing dei dati.

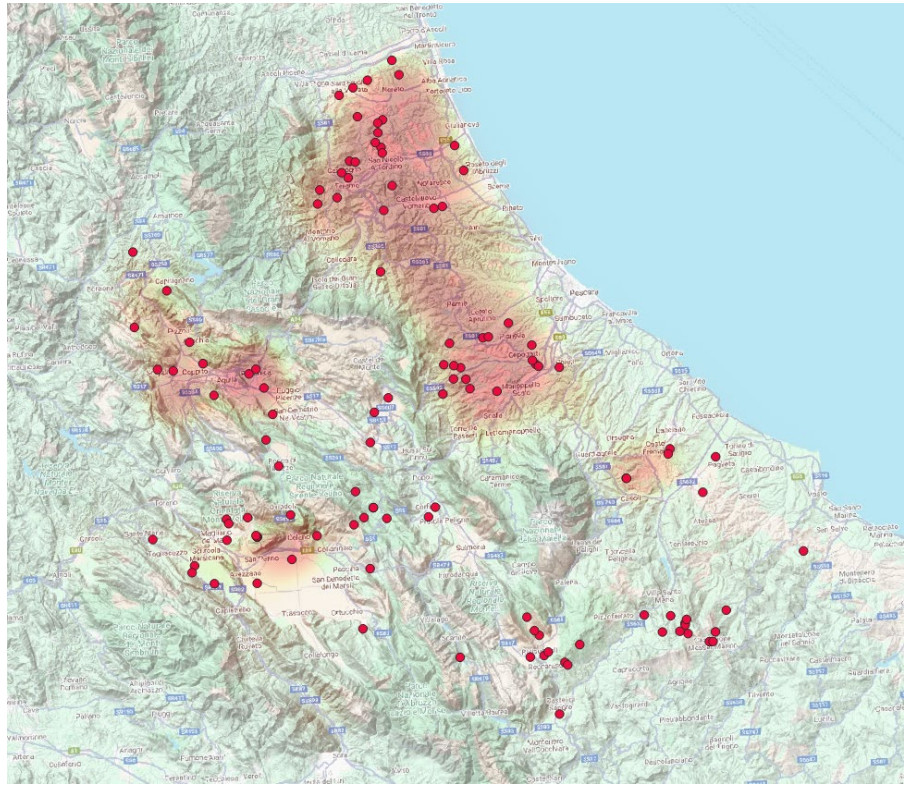
	Capi oltre i 12 mesi	Animali postivi abbattuti	Animali abbattuti	Nuovi nati	Animali morti	Focolaio primario	Numero Casi	Casi con sintomi	Casi Morti	Casi Abbattuti	Casi Guariti	Contatti diretti	Contatti indiretti	Ingresso Veicoli	Ingresso persone	Fecondazione Naturale	Rimonta Enterna
Capi tra 0 e 12 mesi	0,70																
Capi oltre i 12 mesi				0,95	0,76												
Animali positivi		0,96	0,95														
Animali postivi abbattuti			0,99														
Nuovi nati					0,87												
Durata focolaio (gg)						0,85		0,80	0,85				0,84		0,84		
Focolaio primario								0,94	0,99		0,65	0,68	0,98	0,67	0,96	0,63	0,69
Animali presenti							0,61										
Numero Casi										0,79							
Casi con sintomi									0,94		0,61	0,64	0,93	0,63	0,91		0,63
Casi Morti											0,65	0,69	0,98	0,67	0,97	0,64	0,68
Casi Guariti													0,64		0,63		
Contatti diretti													0,71		0,67		
Contatti indiretti														0,68	0,96	0,64	0,68
Ingresso Veicoli															0,68		
Ingresso persone																0,62	0,69
Fecondazione Naturale																	0,67

IZS TERAMO

ISTITUTO ZOOPROFILATTICO SPERIMENTALE DELL'ABRUZZO E DEL MOLISE "G. CAPORALE"

Validazione dei Risultati

- Esecuzione del modello sulle aziende Abruzzesi:
  - Punti rossi: Aziende classificate a rischio
  - Sfondo rosso: Densità maggiore di allevamenti
- Maggior rischio:
  - Zone ad alta densità di allevamenti
  - Aree alto Sangro interessate dalla transumanza dal Foggiano



IZS
TERAMO

ISTITUTO
ZOOPROFILATTICO
SPERIMENTALE
DELL'ABRUZZO
E DEL MOLISE
"G. CAPORALE"

Grazie per l'attenzione