

Progetto Machine Learning

Davide Grandesso 852078

Fabio Marini 851977



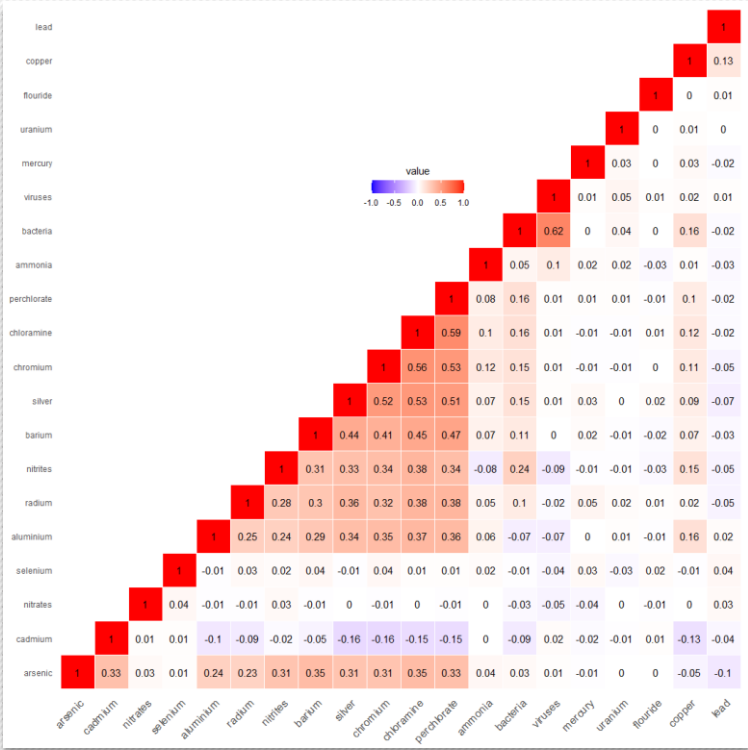
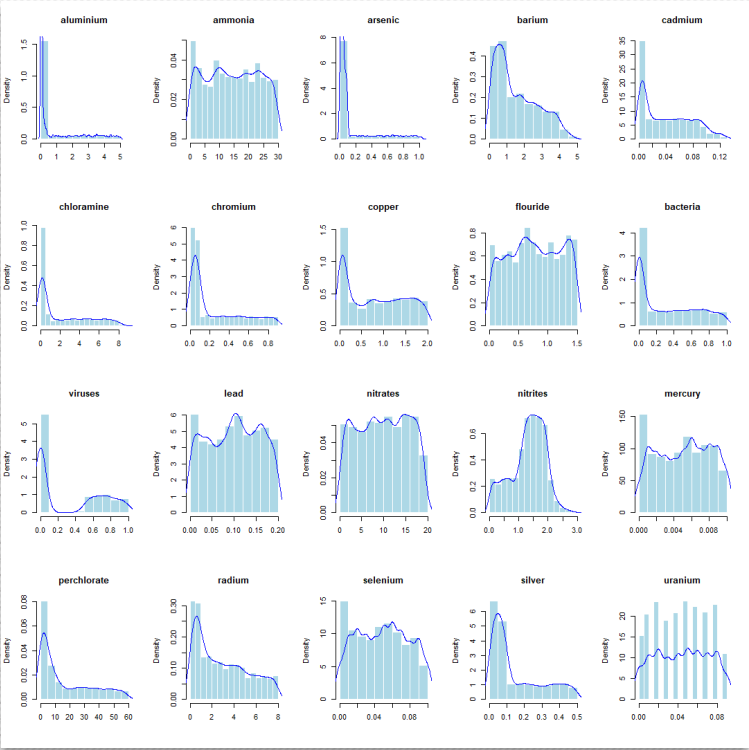
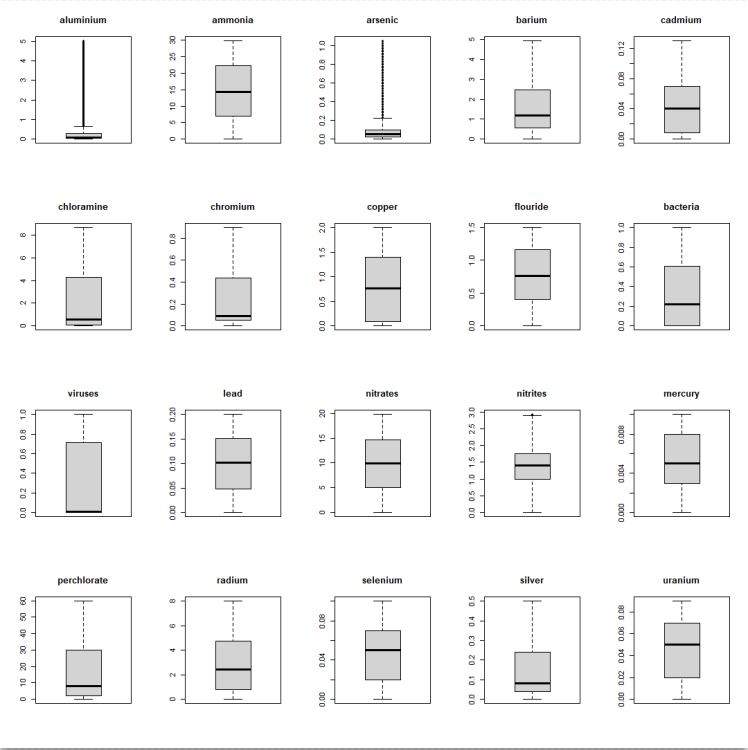
L'obiettivo del progetto è classificare se i campioni di acqua osservati sono sicuri o meno.

La classificazione è fatta prendendo in considerazione gli elementi presenti in essa.

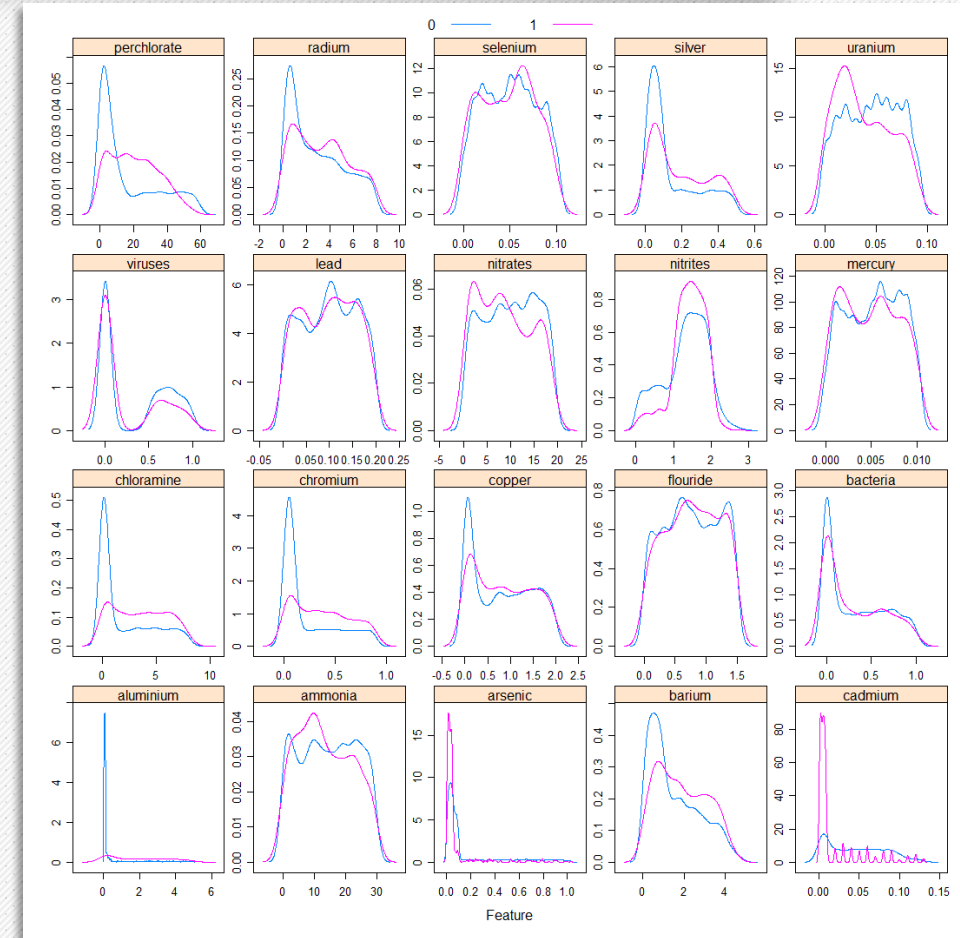
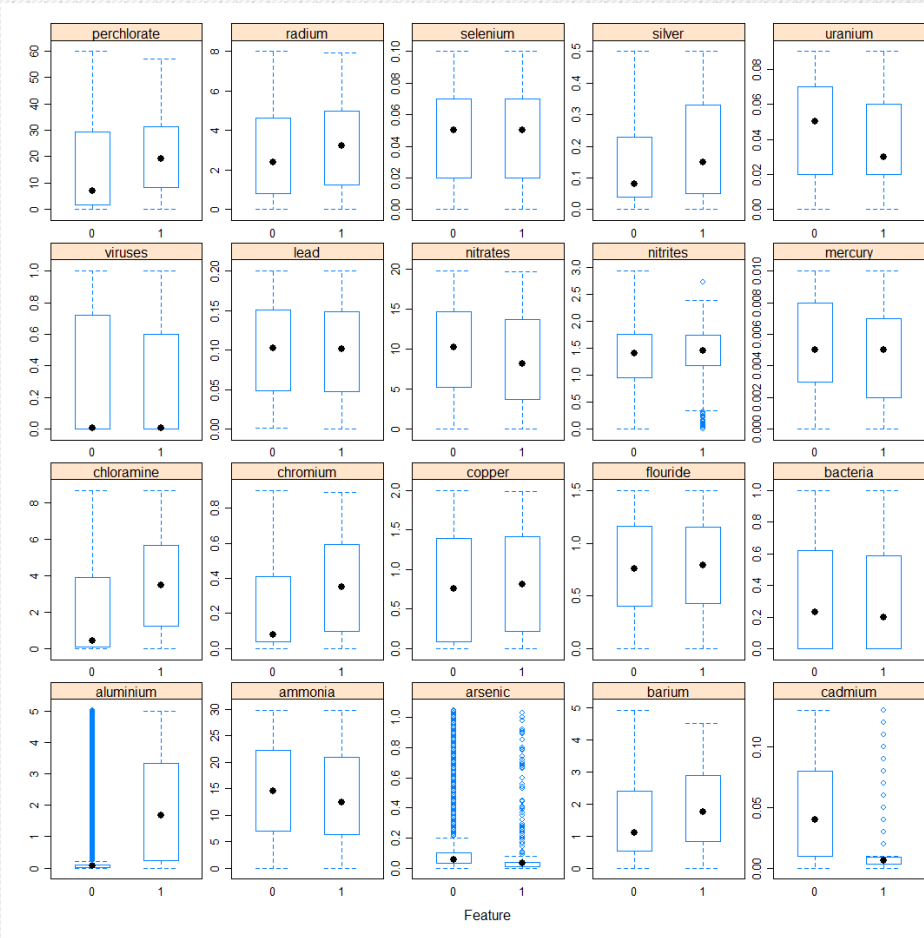
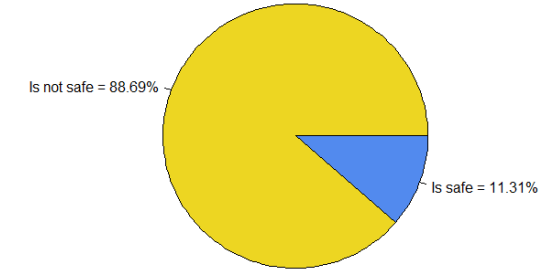


Dataset ~ Water quality

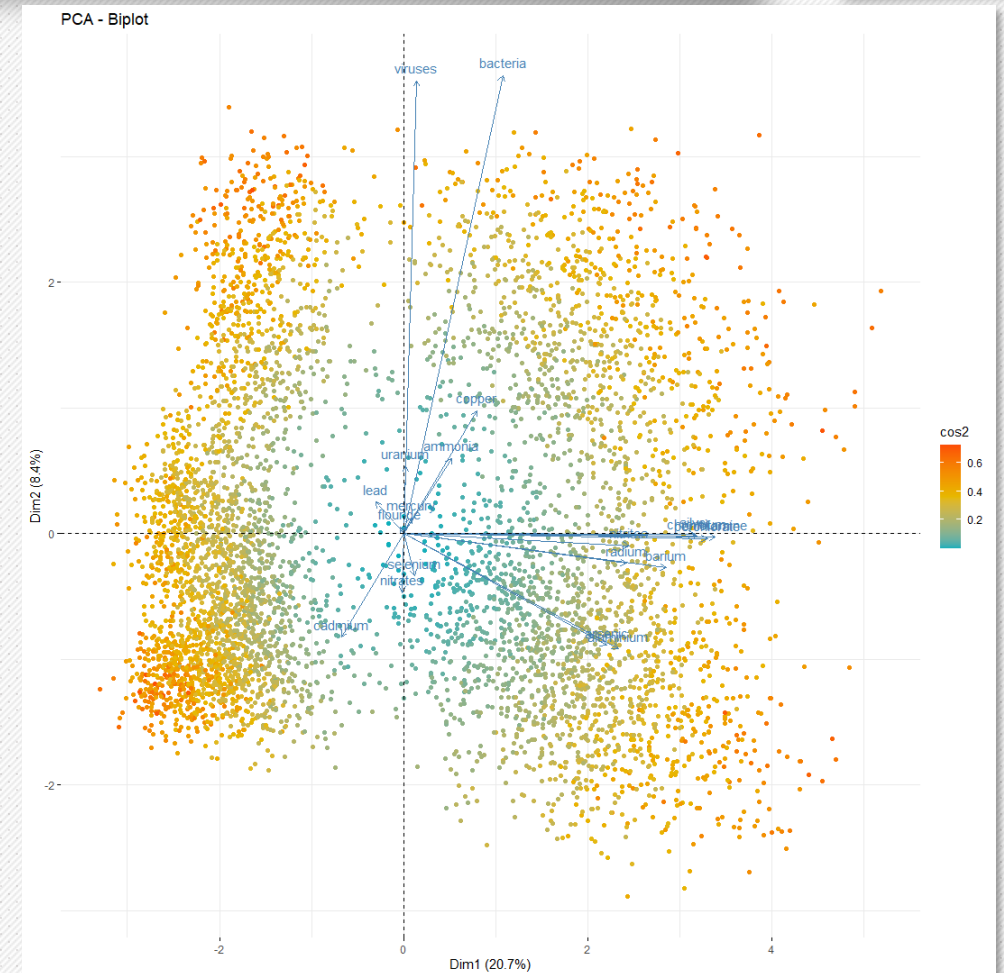
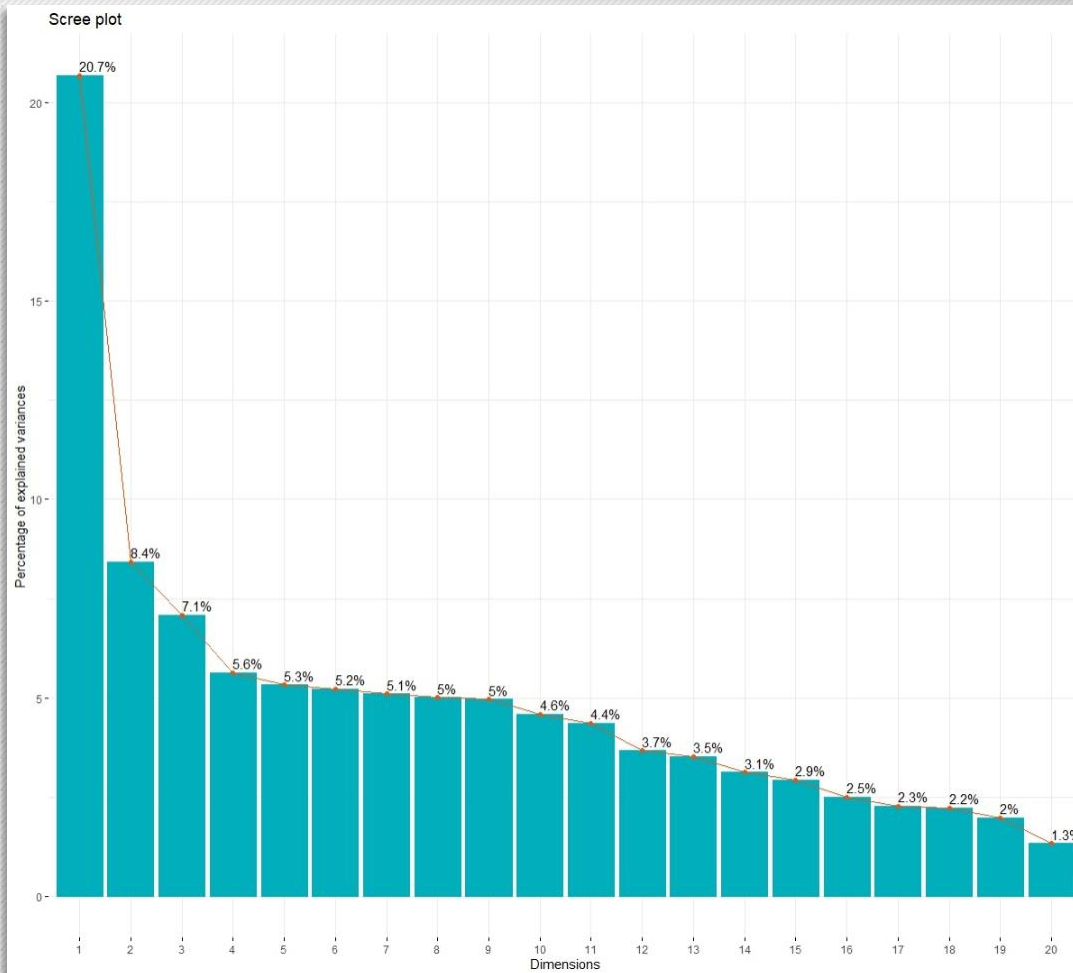
Analisi Esplorativa



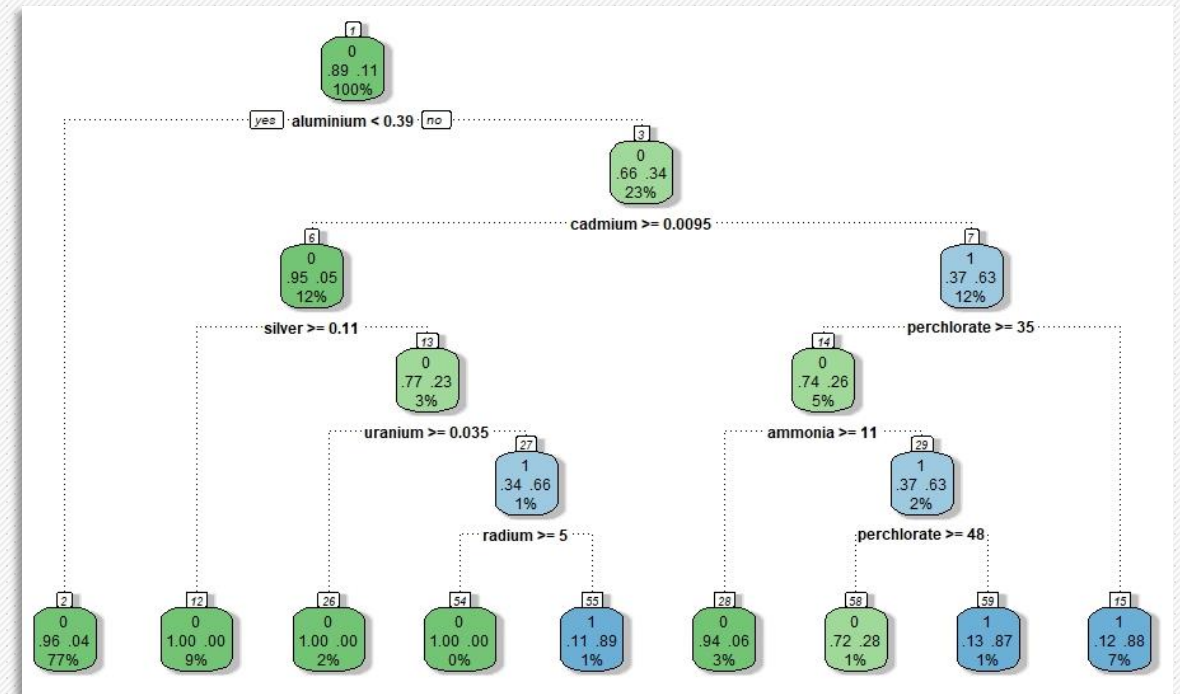
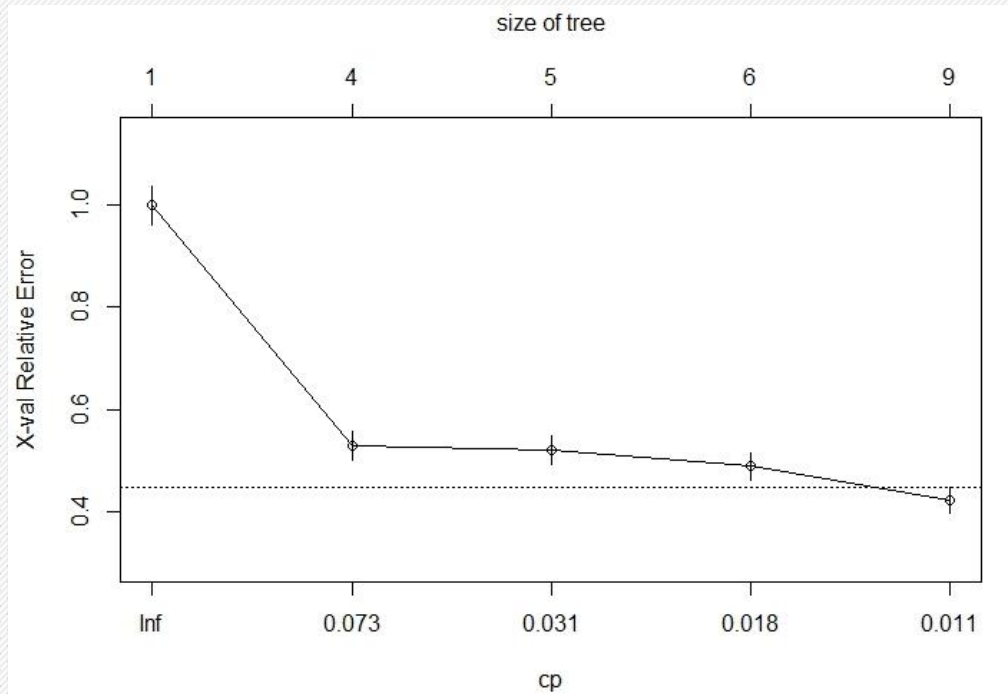
Confronto degli attributi



Analisi delle Componenti Principali ~ PCA

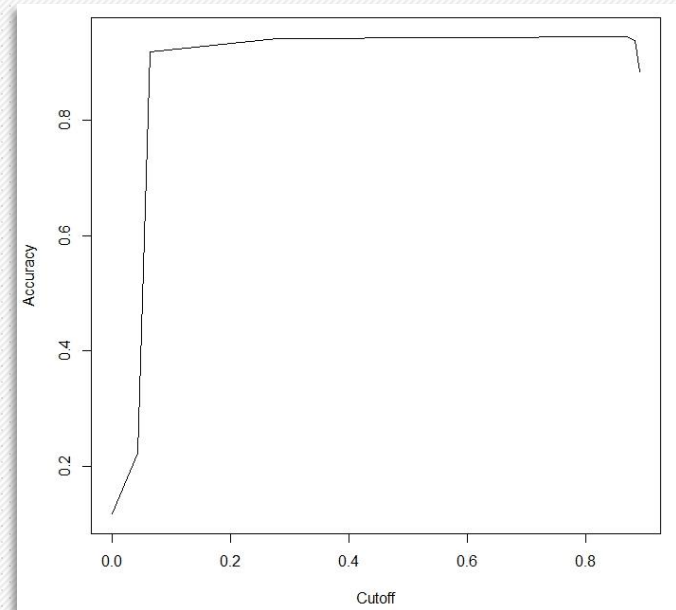
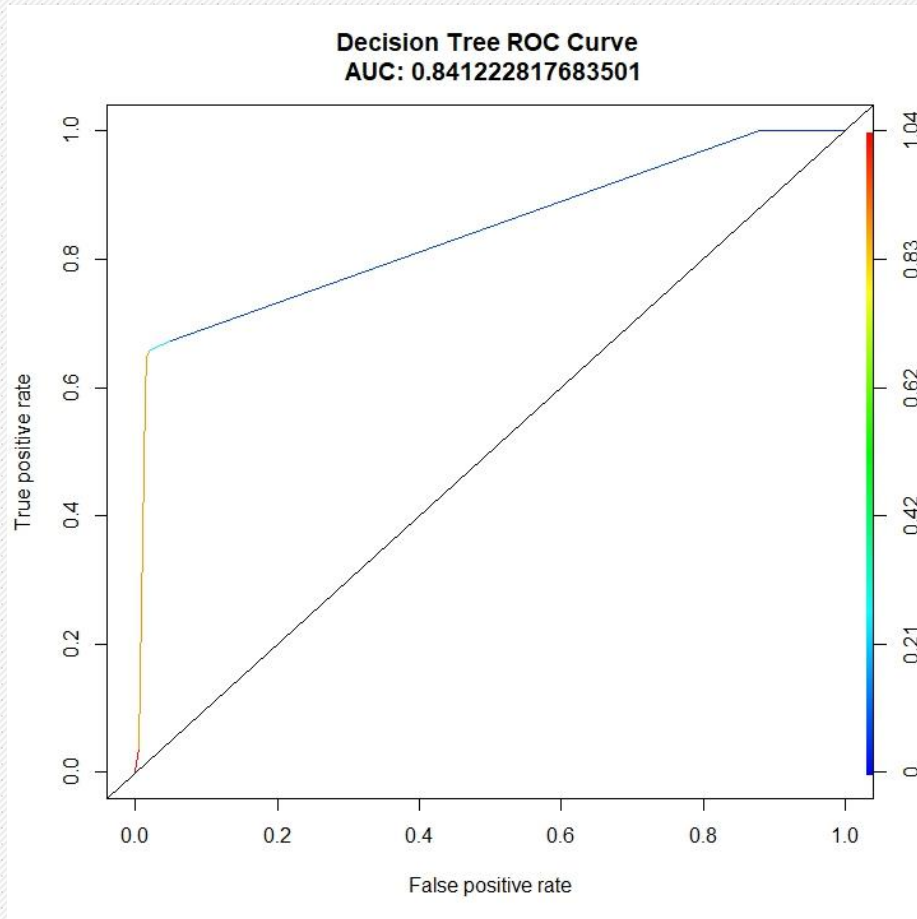


Modelli di Machine Learning utilizzati



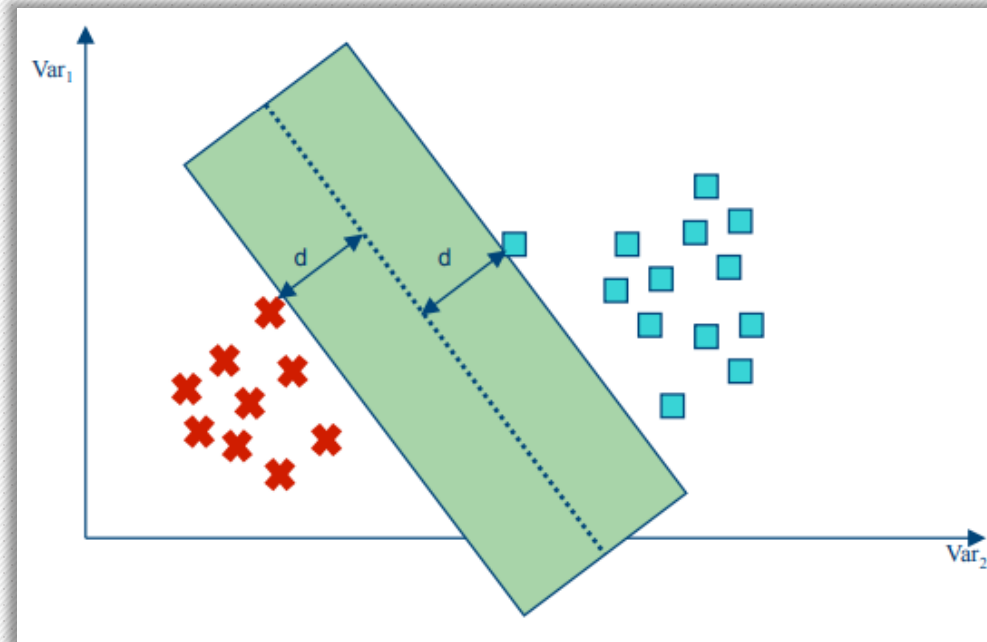
Alberi di Decisione ~ DT

Misure di Performance ~ DT



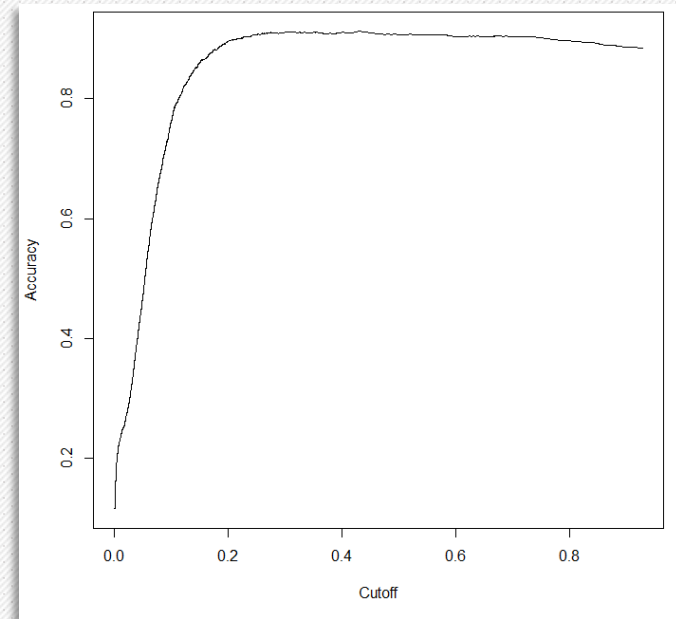
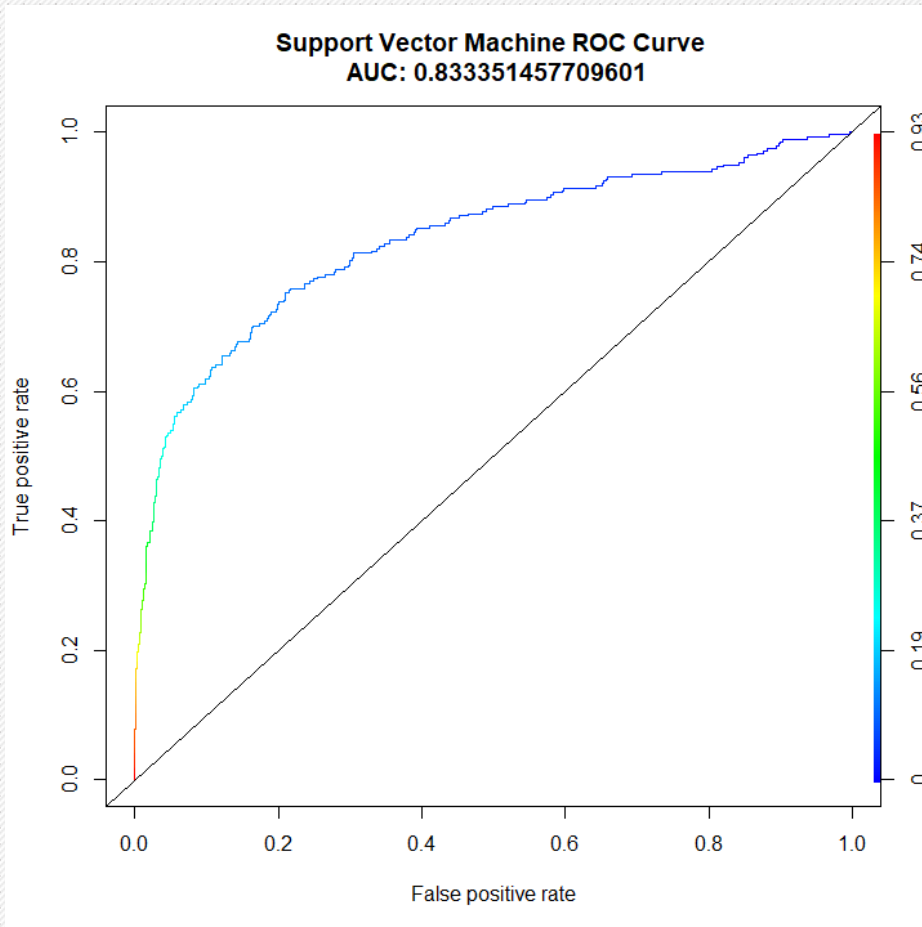
Accuratezza	0.9449
Precisione	0.8411
Recall	0.6475
F-measure	0.7317
Sensibilità	0.6475
Specificità	0.9839

		Predicted class	
		Campioni sicuri	Campioni non sicuri
Actual class	Campioni sicuri	180	98
	Campioni non sicuri	34	2083



Support Vector Machine ~ SVM

Misure di Performance ~ SVM



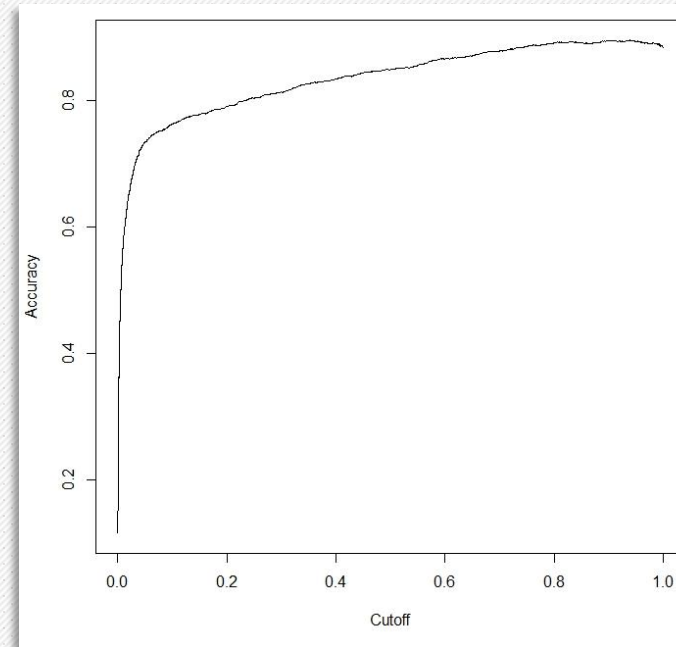
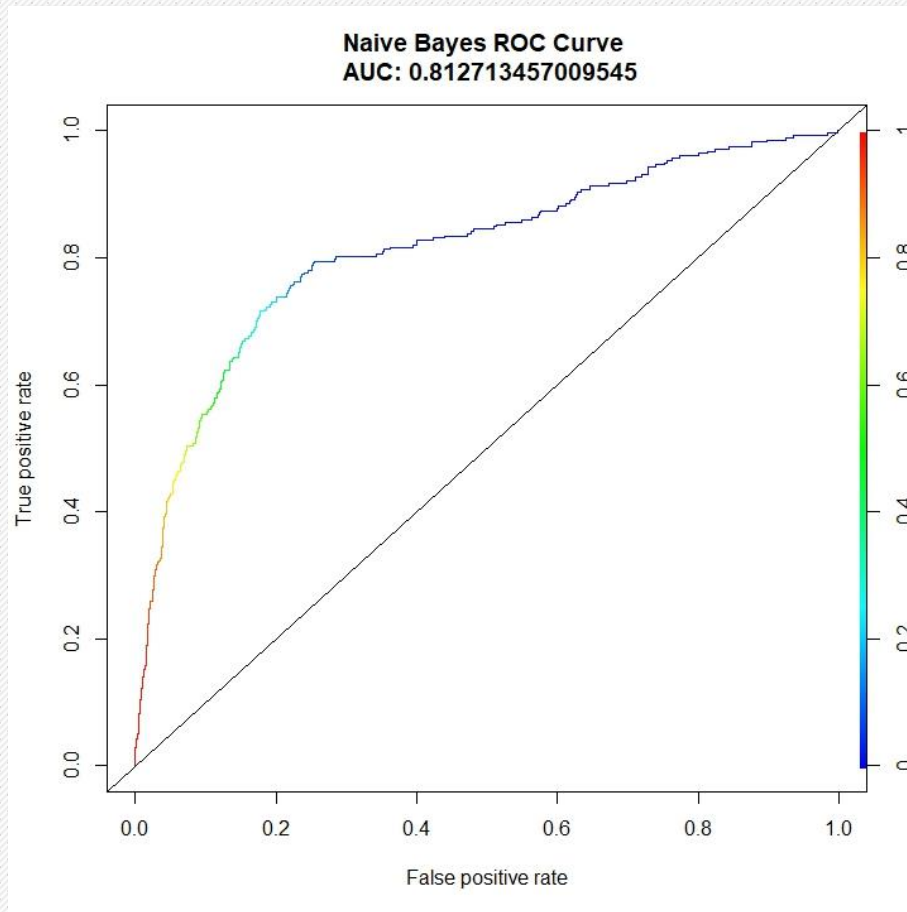
Accuratezza	0.9073
Precisione	0.7593
Recall	0.2950
F-measure	0.4249
Sensibilità	0.2950
Specificità	0.9877

		Predicted class	
		Campioni sicuri	Campioni non sicuri
Actual class	Campioni sicuri	82	196
	Campioni non sicuri	26	2091

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

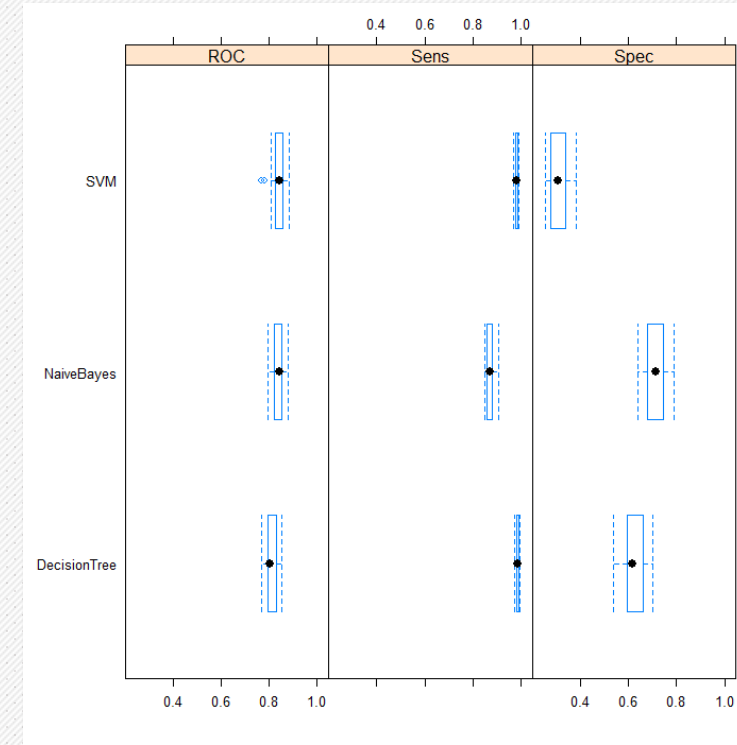
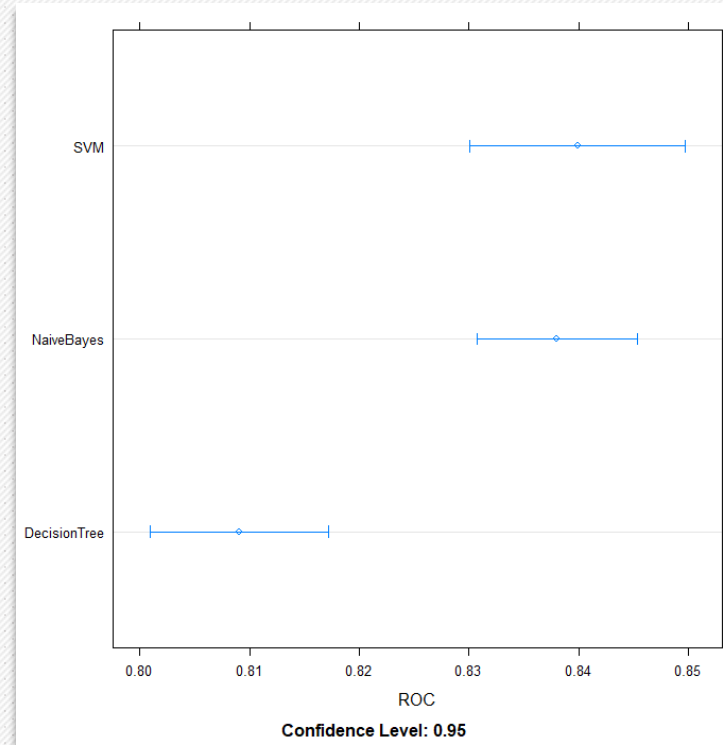
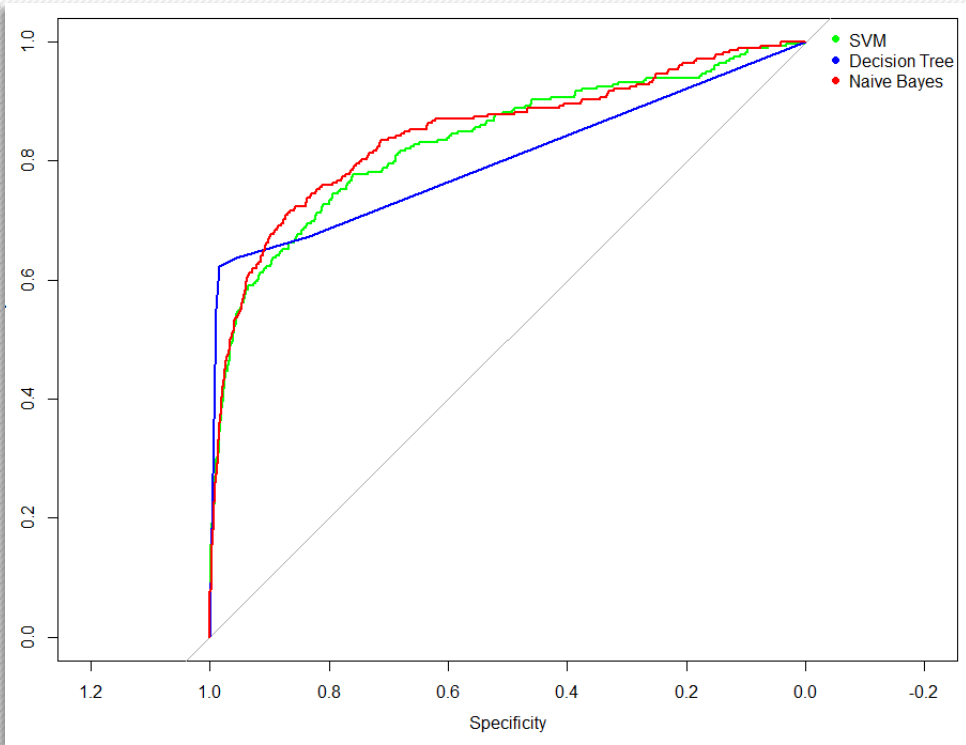
Naive Bayes ~ NB

Misure di Performance ~ NB



Accuratezza	0.8493
Precisione	0.39753
Recall	0.57914
F-measure	0.47145
Sensibilità	0.57914
Specificità	0.88474

		Predicted class	
		Campioni sicuri	Campioni non sicuri
Actual class	Campioni sicuri	161	117
	Campioni non sicuri	244	1873



10-Fold Cross-Validation

Grazie per l'attenzione

Davide Grandesso 852078
Fabio Marini 851977

