

Covid-19 prediction based on blood test result

A machine Learning project

Mattia Marchi - 817587
Stoppa Miguel - 820506
Tomasoni Lorenzo - 829906



Obiettivi e Dominio di Riferimento

Creare un tool di Machine Learning per poter classificare la positività al COVID19 di un paziente sulla base dei suoi parametri sanguinei.

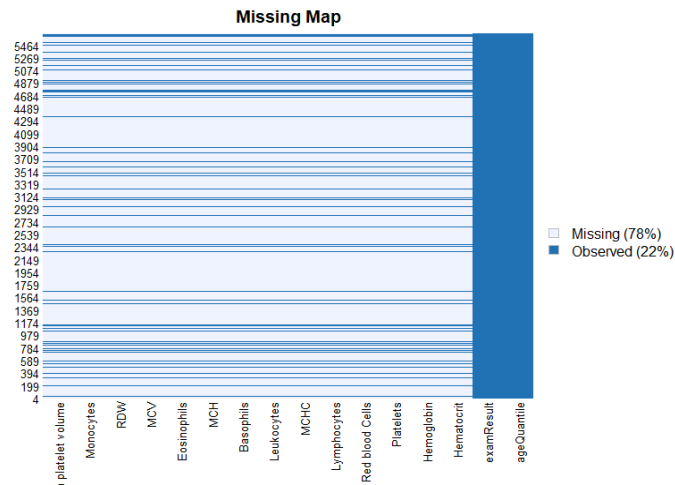
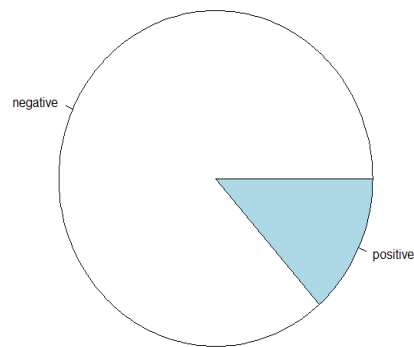


DATASET

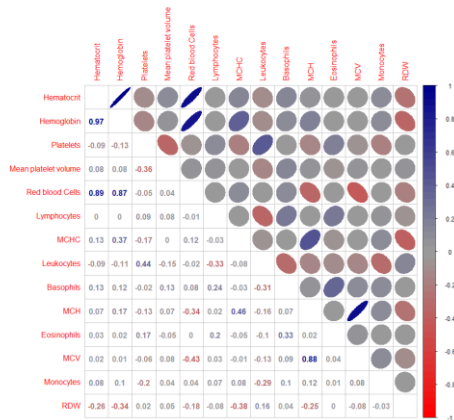
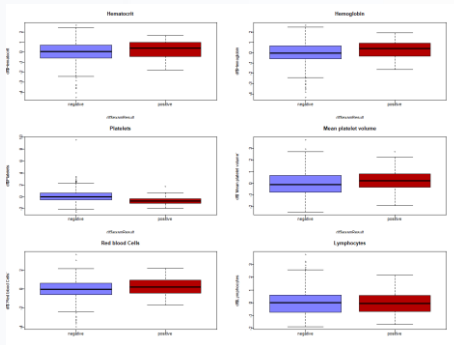
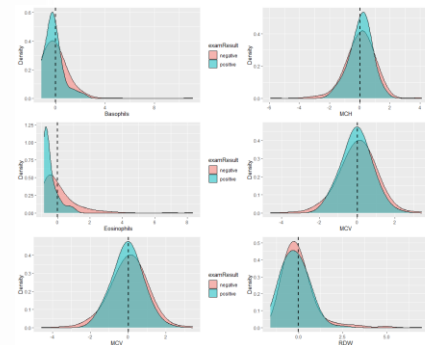
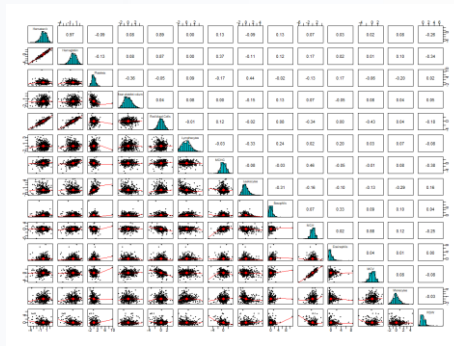
Il dataset è stato ottenuto tramite il sito [kaggle.com](https://www.kaggle.com) e contiene dati anonimi di pazienti provenienti dall'ospedale «Israelita Albert Einstein» di San Paolo, Brasile.

Preprocessing

- Selezione delle feature relative ai parametri ematici
- Rimozione dei valori mancanti

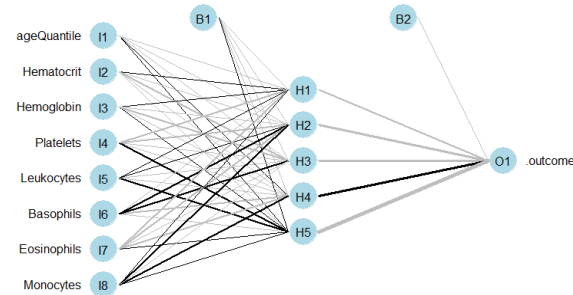
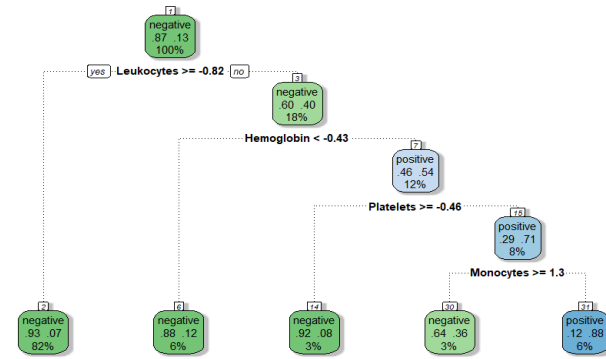


Analisi Esplorativa



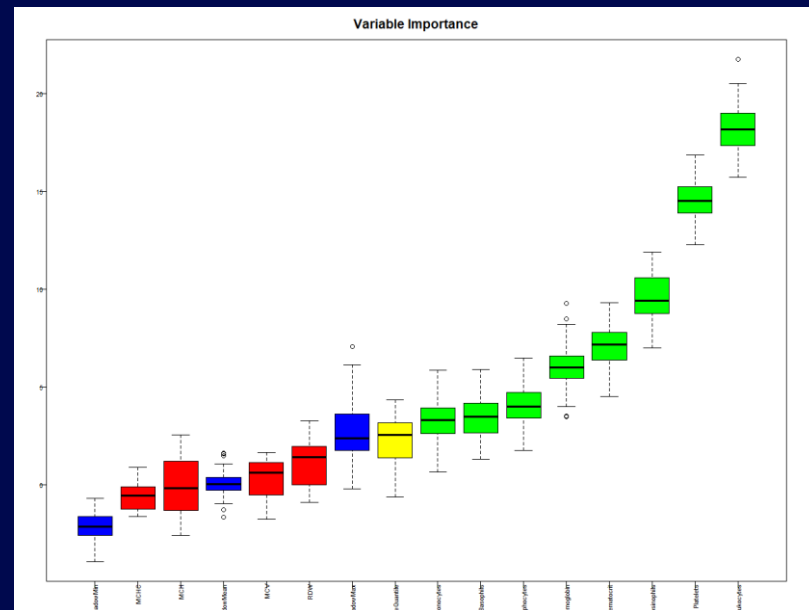
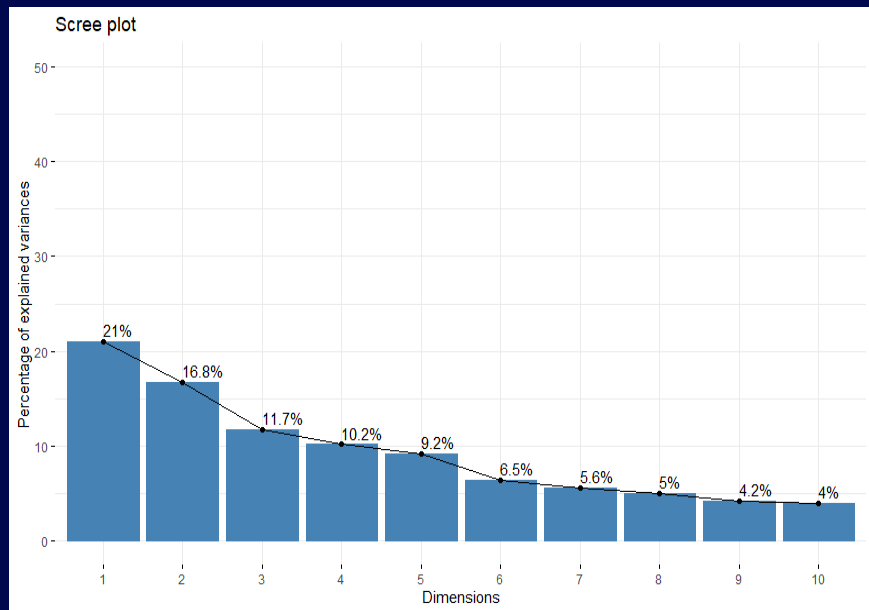
Modelli e Tuning

- Decision Tree
- Neural Network
- Naive Bayes





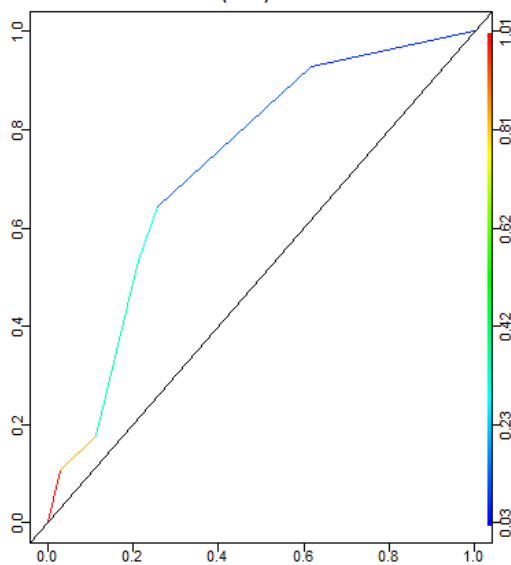
PCA VS FEATURE SELECTION



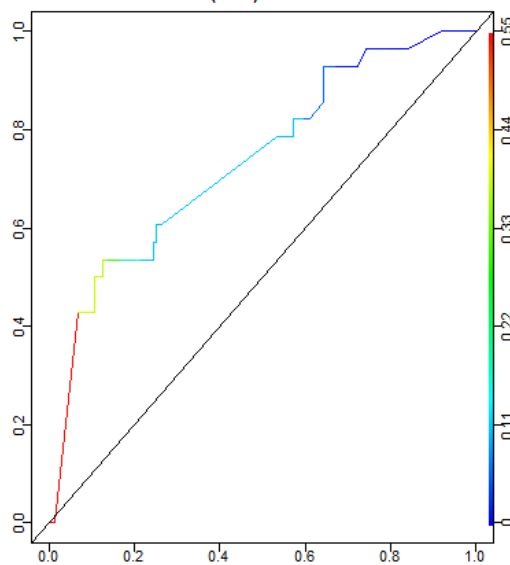


Principal Component Analysis

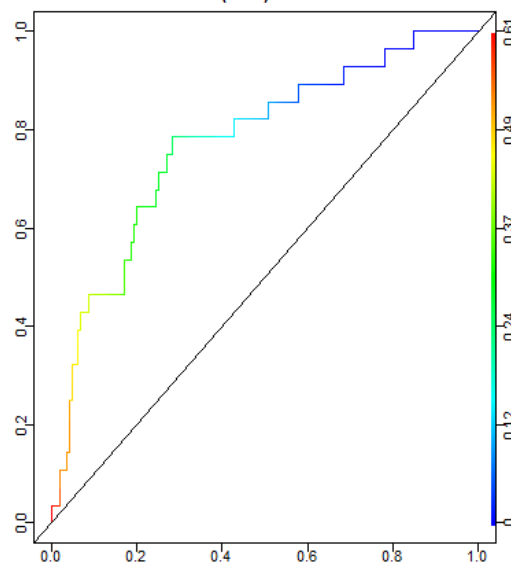
DECISION TREE (PCA): 0.726864330637916



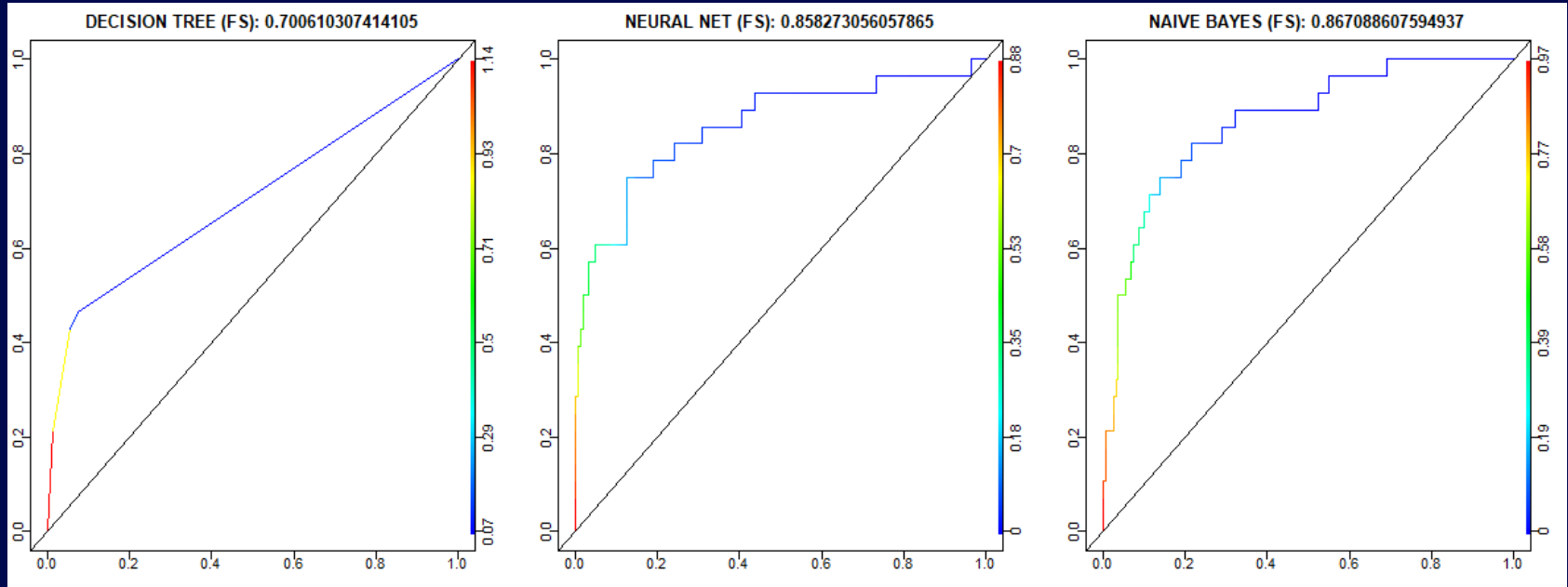
NEURAL NET (PCA): 0.735062893081761



NAIVE BAYES (PCA): 0.77156334231806

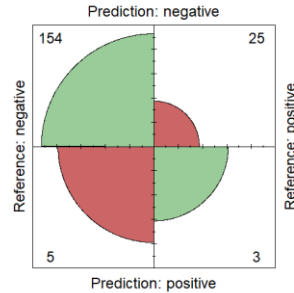


Feature Selection

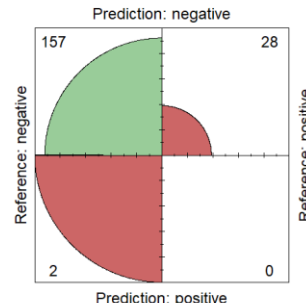


Matrice di Confusione Complessiva

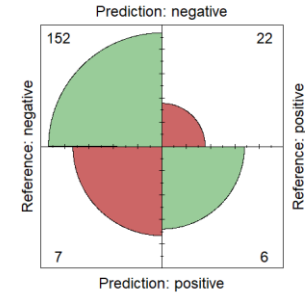
TREE (PCA) ACCURACY: 0.839572192513369



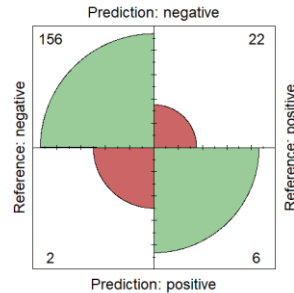
NEURAL NETWORK (PCA) ACCURACY: 0.839572192513369



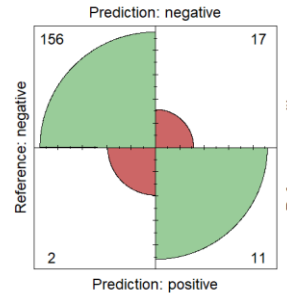
NAIVE BAYES (PCA) ACCURACY: 0.844919786096257



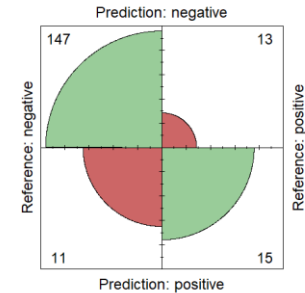
TREE (FS) ACCURACY: 0.870967741935484



NEURAL NETWORK (FS) ACCURACY: 0.897849462365591

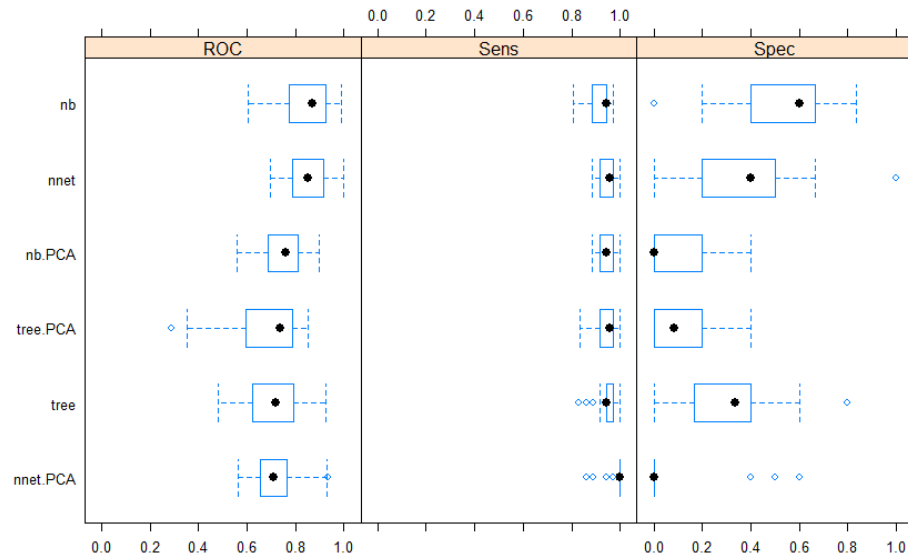
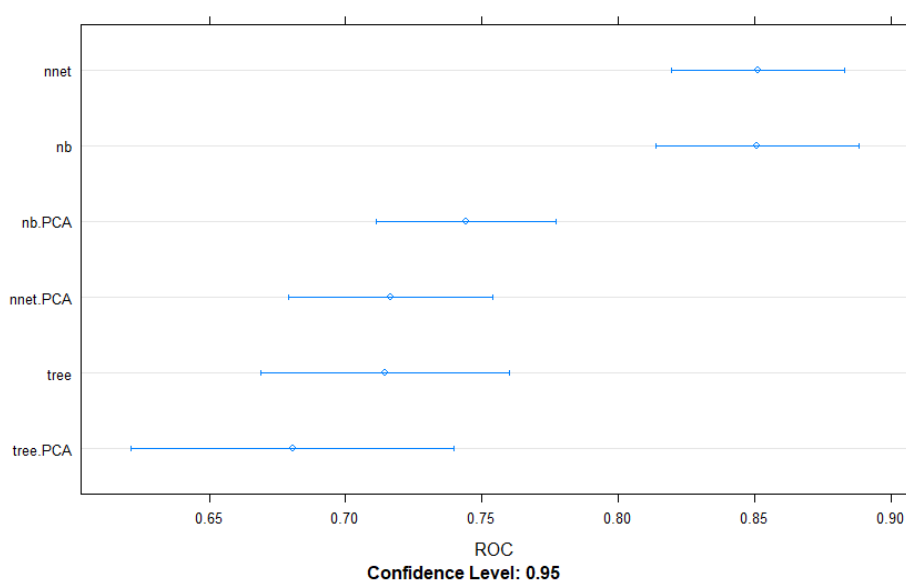


NAIVE BAYES (FS) ACCURACY: 0.870967741935484





Misure di Performance



Conclusioni

È possibile stimare la positività di un paziente al COVID-19 sulla base dei suoi parametri sanguinei?

Con questo dataset **NO**, ma...

Dataset futuri più ricchi
e bilanciati

SI

