

VALENTINA VETRUGNO, MATRICOLA: 704356

LUIGI MATERA, MATRICOLA: 697882

*PROGETTO
INGEGNERIA
DELLA
CONOSCENZA*

INTRODUZIONE

- Il progetto consiste in un sistema che si occupa di *classificare* e fornire le *probabilità* con cui una persona, in 10 anni, potrà essere affetta da problemi all'arteria coronaria.
- Funzioni principali:
 1. Classificare i soggetti che potrebbero ammalarsi nei prossimi 10 anni
 2. Calcolare la probabilità con cui un soggetto possa risultare malato o meno nei prossimi 10 anni

STRUTTURA DEL SISTEMA

Pre-processing Dataset
(file .csv)

Calcolo probabilità con
Classificatore Naive Bayes

Valutazione
Classificatori

Calcolo probabilità con
Rete Bayesiana



PREPROCESSING DATASET

```
#controllo dei valori nulli  
dataCoronary.isnull().sum()
```

✓ 0.4s

male	0
age	0
education	105
currentSmoker	0
cigsPerDay	29
BPMeds	53
prevalentStroke	0
prevalentHyp	0
diabetes	0
totChol	50
sysBP	0
diaBP	0
BMI	19
heartRate	1
glucose	388
TenYearCHD	0



Fill dei
valori
nulli

male	0
age	0
education	0
currentSmoker	0
cigsPerDay	0
BPMeds	0
prevalentStroke	0
prevalentHyp	0
diabetes	0
totChol	0
sysBP	0
diaBP	0
BMI	0
heartRate	0
glucose	0
TenYearCHD	0

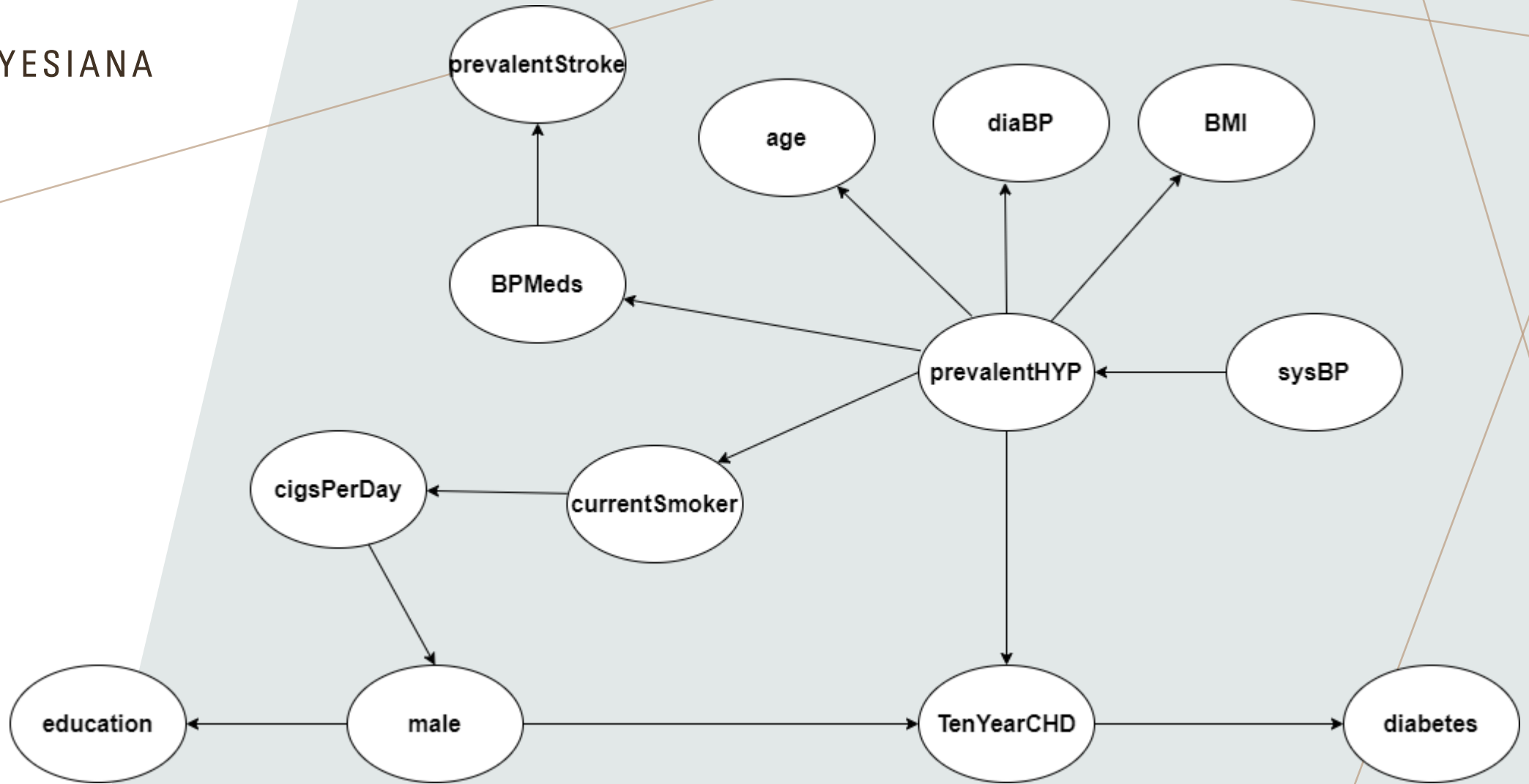
Media delle metriche del GaussianNB:
Media Accuracy: 0.809917
Media Precision: 0.308511
Media Recall: 0.232000
Media f1: 0.264840

Media delle metriche del RandomForest
Media Accuracy: 0.858323
Media Precision: 0.692308
Media Recall: 0.072000
Media f1: 0.130435

CONFRONTO TRA I DUE CLASSIFICATORI
UTILIZZANDO IL K – FOLD CROSS VALIDATION

*VALUTAZIONE DEI
CLASSIFICATORI*

RETE BAYESIANA



RISULTATI:

Classificatore Gaussian Naive Bayes

```
NonMalato      Malato
[[3.64416896e-06 9.99996356e-01]]
```

Rete Bayesiana

Probabilità per un soggetto potenzialmente malato:

TenYearCHD	phi(TenYearCHD)
TenYearCHD(0)	0.4485
TenYearCHD(1)	0.5515

Probabilità per un soggetto potenzialmente sano:

TenYearCHD	phi(TenYearCHD)
TenYearCHD(0)	0.8755
TenYearCHD(1)	0.1245