Myocardial infarction complications

Business Understanding

Definizione

Con infarto miocardico (MI) si intende la necrosi di una parte del muscolo cardiaco a seguito dell'ostruzione di una delle coronarie, arterie deputate alla sua irrorazione.

Come si manifesta?

L'infarto miocardico si può manifestare a riposo, dopo un'emozione intensa, durante uno sforzo fisico rilevante o quando lo sforzo è già terminato.

Negli Stati Uniti più di un milione di persone soffrono di MI ogni anno, e 200-300 mila di loro muoiono prima di arrivare in ospedale.

Fonti:

https://www.humanitas.it/malattie/infarto-del-miocardio/ https://archive.ics.uci.edu/ml/datasets/Myocardial+infarction+complications

Business Understanding

L'infarto miocardico può manifestarsi con o senza complicazioni, le quali possono portare anche al peggioramento della malattia e persino alla morte.

Anche uno specialista esperto non può sempre prevedere lo sviluppo di queste complicazioni. A questo proposito, un supporto alla previsione delle complicanze dell'infarto miocardico, al fine di attuare tempestivamente le necessarie misure preventive, è un compito importante.

Business Understanding

TASK DI DATA MINING

- TASK 1: Clustering dei dati per complicazioni.
- TASK 2: Estrazione di regole di associazione.

DATASET

Myocardial infarction complications disponibile nel seguente link https://archive.ics.uci.edu/ml/datasets/Myocardial+infarction+complications

Nel dataset sono raccolti dati relativi a 1700 pazienti in cui si è presentato l'infarto miocardico.

124 Attributi di cui 12 si riferiscono alla presenza o assenza di complicazioni.

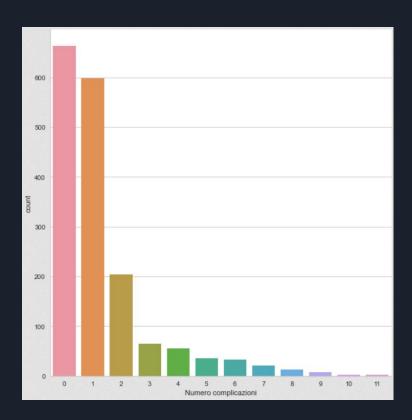
In particolare, sono presenti 13 attributi quantitativi e 111 attributi categorici.

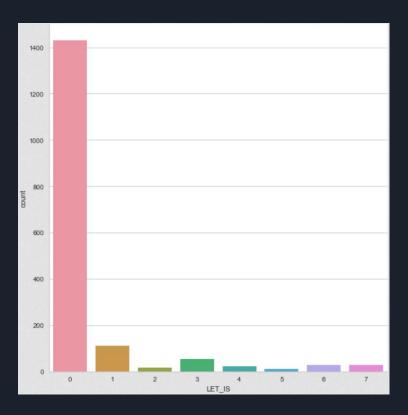
Complicazioni:

- FIBR_PREDS: assenza/presenza di fibrillazione atriale;
- PREDS_TAH: assenza/presenza di tachicardia sopraventricolare;
- JELUD_TAH: assenza/presenza di tachicardia ventricolare;
- FIBR_JELUD: assenza/presenza fibrillazione ventricolare;
- A_V_BLOCK: assenza/presenza di blocco atrioventricolare di III grado;
- OTEK_LANC: assenza/presenza di edema polmonare;
- RAZRIV: assenza/presenza di sindrome di Dressler;
- ZSN: assenza/presenza d'insufficienza cardiaca cronica;
- REC_IM: assenza/presenza di ricaduta dell'infarto miocardico.
- P_IM_STEN: assenza/presenza angina post-infartuale
- LET_IS: il paziente è vivo o è morto per shock cardiogeno, edema polmonare, rottura del miocardio, progressione dell'insufficienza cardiaca congestizia, tromboembolismo, asistolia o fibrillazione ventricolare.

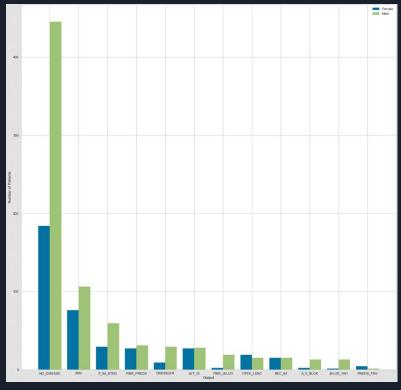
df_original[df_original[columns_output].sum(axis='columns') > 1].shape[0]
438

Ci sono 438 pazienti con più di una complicazione.

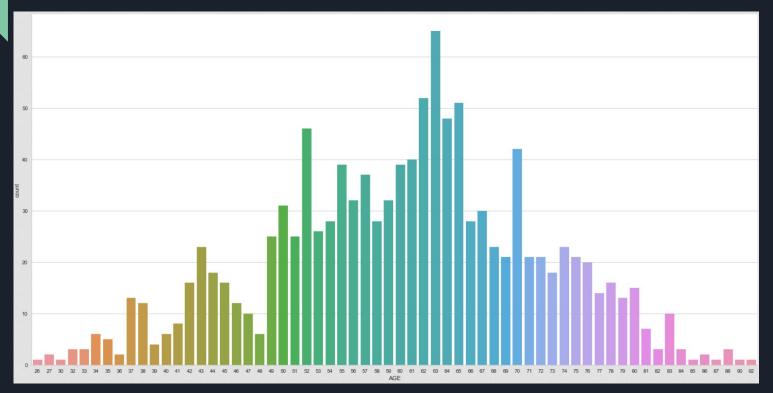




Valori di LET_IS



Numero di pazienti con una determinata complicazione divisi per sesso (post data preparation). 395 femmine e 774 maschi.



Istogramma per età (post data preparation)

Qualità dei dati

- Precisione: conformità del valore memorizzato al valore reale.
- Completezza: quantità dei valori mancanti.
- Coerenza: rappresentazione uniforme.
- Aggiornamento: i dati memorizzati non sono obsoleti.

Qualità dei dati

Completezza

Nel dataset sono presenti 15974 valori mancanti.

	<u> </u>				
KFK_BLOOD	1696	n_p_ecg_p_08	115	SVT_POST	12
IBS_NASL	1628	n_p_ecg_p_10	115	endocr_01	11
S_AD_KBRIG	1076	n_p_ecg_p_11	115	B_BLOK_S_n	11
D_AD_KBRIG	1076	n_p_ecg_p_12	115	LID_S_n	10
NOT_NA_KB	686	n_p_ecg_p_09	115	NOT_NA_1_n	10
LID_KB	677	n_r_ecg_p_04	115	fibr_ter_02	10
NA_KB	657	NOT_NA_2_n	110	fibr_ter_01	10
GIPER_NA	375	R_AB_2_n 108 fibr_ter_		fibr_ter_05	10
NA_BLOOD	375	NA_R_2_n	108	endocr_03	10
K_BLOOD	371	STENOK_AN	106	fibr_ter_03	10
GIPO_K	369	ant_im	83	fibr_ter_06	10
AST_BLOOD	285	inf_im	80	fibr_ter_07	10
ALT_BLOOD	284	lat_im	80	fibr_ter_08	10
S_AD_ORIT	267	FK_STENOK	73	endocr_02	10
D_AD_ORIT	267	post_im	72	GB	9
DLIT_AG	248	ZSN_A	54	NITR_S	9
ROE	203	IBS_POST	51	SIM_GIPERT	8
ritm_ecg_p_06	152	nr_03	21	AGE	8
ritm_ecg_p_08	152	nr_08	21	zab_leg_06	7
ritm_ecg_p_01	152	nr_07	21	zab_leg_02	7
ritm_ecg_p_07	152	nr_04	21	zab_leg_03	7
ritm_ecg_p_02	152	nr_01	21	zab_leg_04	7
ritm_ecg_p_04	152	nr_02	21	zab_leg_01	7
NA_R_3_n	131	nr_11	21	NA_R_1_n	5
NOT_NA_3_n	131	np_07	18	INF_ANAM	4
R_AB_3_n	128	np_09	18	IM_PG_P	1
TIME_B_S	126	np_08	18	P_IM_STEN	0
L_BLOOD	125	np_01	18	OTEK_LANC	0
n_r_ecg_p_06	115	np_10	18	REC_IM	0
n_r_ecg_p_01	115	np_04	18	ZSN	0
n_r_ecg_p_02	115	np_05	18	DRESSLER	0
n_r_ecg_p_03	115	ASP_S_n	17	RAZRIV	0
n_r_ecg_p_05	115	GEPAR_S_n	17	ID	0
n_r_ecg_p_09	115	R_AB_1_n	16	A_V_BLOK	0
n_r_ecg_p_08	115	TIKL_S_n	16	FIBR_JELUD	0
n_r_ecg_p_10	115	TRENT_S_n	16	JELUD_TAH	0
n_p_ecg_p_01	115	K_SH_POST	15	PREDS_TAH	0
n_p_ecg_p_03	115	MP_TP_POST	14	FIBR_PREDS	0
n_p_ecg_p_04	115	ANT_CA_S_n	13	SEX	0
n_p_ecg_p_05	115	O_L_POST	12	LET_IS	0
n_p_ecg_p_06	115	FIB_G_POST	12	3-6-1	
n p ecg p 07	115	GT POST	12		

Numero di valori mancanti di ogni attributo

• Eliminate le righe riguardanti pazienti con più di una complicazione.

```
#Rimuovo i pazienti con più di una complicazione
num_righe_pre_rim = df_original.shape[0]
index_more_compl = df_original[df_original[columns_output].sum(axis='columns') > 1].index
print(f'Righe pre-rimozione: {df_original.shape[0]}')
df = df_original.drop(index_more_compl)
df.reset_index(drop=True, inplace=True)
print(f'Righe post-rimozione: {df.shape[0]}')
print(f'Sono state rimosse {num_righe_pre_rim-df.shape[0]} righe')

Righe pre-rimozione: 1700
Righe post-rimozione: 1262
Sono state rimosse 438 righe
```

- Trasformazione della colonna LET_IS.
- Aggiunta la colonna 'class', la quale indica la complicazione di cui soffre il paziente + valore NO_DISEASE.
- Eliminazione delle colonne riguardanti le complicazioni.

Valori mancanti:

- Eliminate le colonne con più valori mancanti: KFK_BLOOD, IBS_NASL, S_AD_KBRIG, D_AD_KBRIG, NOT_NA_KB, LIB_KB, NA_KB. Inoltre, è stata eliminata la colonna ID, la quale risulta inutile in analisi.
- 2. Eliminate le righe con più del 20% di valori mancanti.

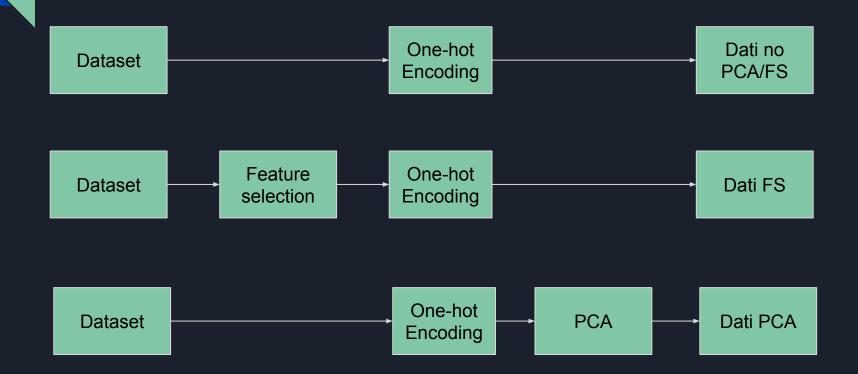
```
#Rimuovo le righe che hanno più del 20% di valori nulli
num_righe_pre_rim = df.shape[0]
print(f'Righe pre-rimozione: {df.shape[0]}')
df.dropna(thresh=df.shape[1] - 20*df.shape[1]/100, inplace=True)
df.reset_index(inplace=True, drop=True)
print(f'Righe post-rimozione: {df.shape[0]}')
print(f'Sono state rimosse {num_righe_pre_rim-df.shape[0]} righe')

Righe pre-rimozione: 1262
Righe post-rimozione: 1169
Sono state rimosse 93 righe
```

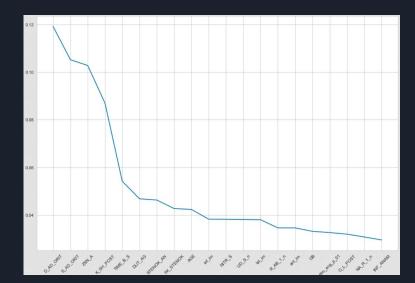
Valori mancanti:

Imputazione

- Variabili categoriche: i valori nulli sono stati imputati con la moda del relativo attributo tenendo in considerazione l'attributo 'class'.
- Variabili quantitative: i valori nulli sono stati imputati con la media del relativo attributo tenendo in considerazione l'attributo 'class'.



La Mutual information tra due variabili è un valore non negativo che misura la dipendenza tra le variabili. È uguale a zero se e solo se due variabili sono indipendenti, mentre valori più alti rappresentano una maggiore dipendenza.
Il gomito è sul sesto attributo.

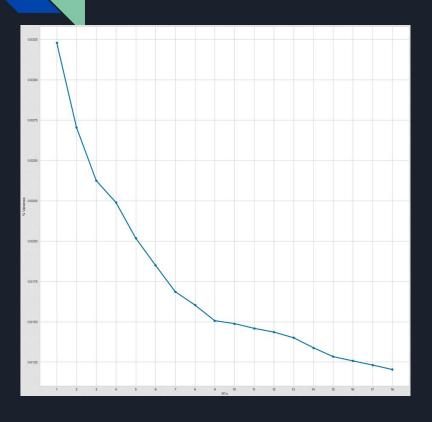


D AD ORIT	0.120834931	n p ecg p 12	0.009701	np 01	
S AD ORIT	0.105050556	fibr ter 03	0.009619	ALT BLOOD	
ZSN A	0.102824273	n_r_ecg_p_04	0.009005	ROE	
K_SH_POST	0.087009435	nr_04	0.008659	AST_BLOOD	
TIME B S	0.049883056	FIB G POST	0.008549		
DLIT_AG	0.046816076	n_r_ecg_p_01	0.0083		
STENOK AN	0.046294922	n_p_ecg_p_07	0.008211		
FK STENOK	0.042714746	n_r_ecg_p_03	0.008066		
AGE	0.042310113	B BLOK S n	0.007776		
ant_im	0.041438362	endocr_02	0.007644		
inf_im	0.038257933	nr_03	0.007436		
NITR_S	0.038225913	ritm_ecg_p_08	0.006901		
LID S n	0.038126954	TIKL S n	0.006589		
lat_im	0.038029625	fibr_ter_07	0.005779		
R AB 1 n	0.034603852	SVT_POST	0.005296		
GB	0.032985528	TRENT S n	0.005168		
ritm_ecg_p_01	0.032622358	zab_leg_04	0.00488		
O L POST	0.031970724	zab leg 03	0.004833		
NA_R_1_n	0.030773575	n_p_ecg_p_11	0.004815		
INF ANAM	0.029533535	GT_POST	0.004563		
ANT_CA_S_n	0.02735731	GIPER_NA	0.004401		
ASP_S_n	0.027344855	n_p_ecg_p_10	0.004263		
IBS_POST	0.026328536	SIM_GIPERT	0.004023		
R_AB_2_n	0.025944883	np_08	0.003939		
NOT_NA_1_n	0.024390087	n_r_ecg_p_09	0.003816		
NA_BLOOD	0.023547455	n_r_ecg_p_02	0.003815		
ritm_ecg_p_07	0.022695154	n_r_ecg_p_10	0.003459		
SEX	0.022589331	n_p_ecg_p_09	0.003449		
ritm_ecg_p_02	0.020102285	fibr_ter_01	0.003271		
zab_leg_02	0.01918922	endocr_03	0.003238		
R_AB_3_n	0.018248655	np_09	0.00319		
L_BLOOD	0.01754613	zab_leg_06	0.002936		
n_r_ecg_p_05	0.016874471	fibr_ter_06	0.002848		
n_p_ecg_p_06	0.016583799	nr_02	0.002667		
post_im	0.016284206	n_p_ecg_p_04	0.002624		
endocr_01	0.015903647	n_p_ecg_p_05	0.002622		
GIPO_K	0.015637562	np_10	0.002622		
zab_leg_01	0.015380028	np_05	0.002476		
n_p_ecg_p_03	0.014209869	fibr_ter_05	0.002465		
MP_TP_POST	0.013946084	n_r_ecg_p_08	0.002382		
IM_PG_P	0.013913784	nr_08	0.00234		
NOT_NA_2_n	0.013824664	fibr_ter_08	0.002217		
NA_R_2_n	0.013207967	ritm_ecg_p_06	0.001593		
GEPAR_S_n	0.01250405	nr_01	0.001593		
n_r_ecg_p_06	0.011703765	K_BLOOD	0.001481		
ritm_ecg_p_04	0.011561439	n_p_ecg_p_08	0.00131		
NOT_NA_3_n	0.01088005	np_04	0.001062		
NA_R_3_n	0.01075916	n_p_ecg_p_01	0.001062		
nr_11	0.009831952	np_07	0.00053		
fibr_ter_02	0.009726184	nr_07	0		

Mutual Information tra le variabili di input e la variabile class

Variabili categoriche

• Trasformazione delle variabili FK_STENOK, IBS_POST, GB, ZSN_A, ant_im, lat_im, inf_im, post_im con la codifica One-Hot.



PCA

```
X = df[df.columns.difference(['class'])]
standardScaler = StandardScaler()
X scaled = standardScaler.fit transform(X)
pca = PCA(n_components=18, random_state=17)
X PCA = pca.fit transform(X scaled)
features = range(1, pca.n components +1)
serie PCs = pd.Series(np.cumsum(pca.explained variance ratio ), features)
for k, v in zip(serie_PCs[:18].keys(), serie_PCs[:18].values):
    print(f'{k}
      0.03229529011093845
      0.05933791104732604
      0.08308038272206524
      0.10546444396267166
      0.1256451305386452
      0.14414138890429684
      0.1610008557420771
      0.17702946137151732
      0.192085325681049
10
       0.2069617701471705
11
       0.22154231636402685
12
       0.23589327068762492
13
       0.24989256433966675
14
       0.26327238024763305
15
       0.2761007844561868
16
       0.28866537997316355
17
       0.3009719203495881
18
       0.31300203977023217
```

Per il primo task sono stati utilizzati vari algoritmi di clustering cambiando i dati di input. In particolare, sono stati utilizzati:

- 1. tutti gli attributi di input;
- 2. i primi sei attributi con il mutual information più alto;
- 3. le prime nove componenti principali calcolate con la PCA.

K-Means con 12 cluster

```
Class attribute: class
Classes to Clusters:
                         6 7 8 9 10 11 <-- assigned to cluster
Cluster 0 <-- No class
Cluster 1 <-- NO DISEASE
Cluster 2 <-- REC IM
Cluster 3 <-- DRESSLER
Cluster 4 <-- FIBR PREDS
Cluster 6 <-- JELUD_TAH
Cluster 7 <-- No class
Cluster 8 <-- LET IS
Cluster 9 <-- P IM STEN
Cluster 10 <-- OTER LANC
Cluster 11 <-- FIBR_JELUD
Incorrectly clustered instances :
                                      935.0 79.9829 %
```

```
Class attribute: class
Classes to Clusters:
                           7 8 9 10 11 <-- assigned to cluster
        0 2 3 1 8 1 30 6 0 2 | LET IS
Cluster 0 <-- P IM STEN
Cluster 1 <-- REC IM
Cluster 2 <-- DRESSLER
Cluster 3 <-- PREDS TAH
Cluster 4 <-- FIBR PREDS
Cluster 5 <-- NO DISEASE
Cluster 6 <-- A V BLOK
Cluster 7 <-- ZSN
Cluster 8 <-- LET IS
Cluster 9 <-- OTEK LANC
Cluster 10 <-- JELUD TAH
Cluster 11 <-- FIBR_JELUD
Incorrectly clustered instances :
                                     933.0 79.8118 %
```

```
Class attribute: class
Classes to Clusters:

0 1 2 3 4 5 6 7 8 9 10 11 <-- assigned to cluster
78 73 69 18 12 43 27 87 6 51 128 37 | No_DISEASE
13 22 15 12 17 7 7 14 7 25 21 12 | 25N
13 3 4 4 8 1 3 4 0 7 5 6 | PIBR_PREDS
3 5 3 2 0 3 1 4 1 3 2 3 | REC_IM
2 2 0 0 0 1 0 0 0 0 0 0 0 | PREDS_TAH
2 1 4 0 1 3 2 4 0 9 7 5 | DRESSLER
0 1 1 0 0 2 2 0 0 0 0 8 1 | A_V_BLOK
13 16 14 4 2 7 1 14 1 7 9 0 | P_IM_STEN
6 2 2 3 2 5 3 0 5 2 2 2 | OTEK_LANC
0 0 0 1 0 1 0 1 2 1 1 1 0 0 3 4 | JEUD_TAH
0 1 3 2 3 5 6 3 0 26 1 1 4 | LET_IS

Cluster 0 <-- FIBR_PREDS
Cluster 1 <-- ZeN
Cluster 2 <-- P_IM_STEN
Cluster 3 <-- No class
Cluster 4 <-- JEUD_TAH
Cluster 5 <-- OTEK_LANC
Cluster 6 <-- A_V_BLOK
Cluster 6 <-- A_V_BLOK
Cluster 9 <-- DERSSLER
Cluster 9 <-- DERSSLER
Cluster 1 <-- ZeN
Cluster 9 <-- DERSSLER
Cluster 9 <-- DERSSLER
Cluster 9 <-- DERSSLER
Cluster 10 <-- OTEK_LANC
Cluster 6 <-- A_V_BLOK
Cluster 9 <-- DERSSLER
Cluster 10 <-- NO CLESSEE
Cluster 11 <-- DERSSLER
Cluster 11 <-- DERSSLER
Cluster 11 <-- DERSSLER
Cluster 11 <-- DISEASE
Cluster 11 <-- FIBR_JELUD

Incorrectly clustered instances : 937.0 80.154 $
```

E-M non specificando il numero di cluster

```
Class attribute: class
Classes to Clusters:
  0 1 2 <-- assigned to cluster
 252 204 173 | NO DISEASE
 82 43 57 | ZSN
 29 14 15 | FIBR PREDS
         12 | REC IM
         2 | PREDS TAH
  14 14 10 | DRESSLER
         6 | A V BLOK
     25 19 | P IM STEN
         10 | OTEK LANC
          7 | JELUD TAH
          9 | FIBR JELUD
 17 3 35 | LET IS
Cluster 0 <-- NO DISEASE
Cluster 1 <-- P IM STEN
Cluster 2 <-- ZSN
Incorrectly clustered instances :
                                              71.4286 %
```

E-M con 12 cluster

```
Class attribute: class
Classes to Clusters:
                        6 7 8 9 10 11 <-- assigned to cluster
            0 0 1 0 1 5 1 1 4 | FIBR JELUD
Cluster 0 <-- JELUD TAH
Cluster 1 <-- FIBR JELUD
Cluster 3 <-- NO DISEASE
Cluster 4 <-- LET IS
Cluster 5 <-- REC IM
Cluster 6 <-- No class
Cluster 7 <-- P IM STEN
Cluster 9 <-- PREDS TAH
Cluster 10 <-- OTEK LANC
Cluster 11 <-- FIBR PREDS
Incorrectly clustered instances :
                                     950.0 81.266 %
```

```
Class attribute: class
Classes to Clusters:
                          6 7 8 9 10 11 <-- assigned to cluster
Cluster 0 <-- DRESSLER
Cluster 1 <-- FIBR PREDS
Cluster 2 <-- ZSN
Cluster 3 <-- REC IM
Cluster 4 <-- A_V_BLOK
Cluster 5 <-- OTEK LANC
Cluster 6 <-- P IM STEN
Cluster 7 <-- FIBR JELUE
Cluster 8 <-- NO DISEASE
Cluster 9 <-- No class
Cluster 10 <-- JELUD TAH
Cluster 11 <-- LET IS
Incorrectly clustered instances :
                                      928.0 79.3841 %
```

```
Class attribute: class
Classes to Clusters:
                            7 8 9 10 11 <-- assigned to cluster
Cluster 0 <-- FIBR PREDS
Cluster 1 <-- FIBR JELUD
Cluster 2 <-- LET IS
Cluster 3 <-- OTEK LANC
Cluster 4 <-- REC IM
Cluster 5 <-- P IM STEN
Cluster 6 <-- ZSN
Cluster 8 <-- NO DISEASE
Cluster 9 <-- PREDS TAH
Cluster 10 <-- JELUD TAH
Cluster 11 <-- A_V_BLOK
Incorrectly clustered instances :
                                      850.0 72.7117 %
```

Gerarchico Single-Linkage con 12 cluster

```
Class attribute: class
Classes to Clusters:
Cluster 0 <-- NO DISEASE
Cluster 1 <-- No class
Cluster 2 <-- OTEK LANC
Cluster 3 <-- No class
Cluster 4 <-- No class
Cluster 5 <-- ZSN
Cluster 6 <-- LET IS
Cluster 7 <-- No class
Cluster 8 <-- No class
Cluster 9 <-- No class
Cluster 10 <-- No class
Cluster 11 <-- No class
Incorrectly clustered instances :
                                       525.0 44.9102 %
```

```
Class attribute: class
Classes to Clusters:
                                     9 10 11 <-- assigned to cluster
Cluster 0 <-- NO DISEASE
Cluster 1 <-- PREDS TAH
Cluster 2 <-- No class
Cluster 3 <-- No class
Cluster 5 <-- ZSN
Cluster 6 <-- No class
Cluster 7 <-- No class
Cluster 9 <-- LET IS
Cluster 10 <-- No class
Cluster 11 <-- No class
Incorrectly clustered instances :
                                      537.0 45.9367 %
```

```
Class attribute: class
Classes to Clusters:
Cluster 0 <-- NO DISEASE
Cluster 1 <-- No class
Cluster 2 <-- No class
Cluster 3 <-- No class
Cluster 5 <-- No class
Cluster 6 <-- ZSN
Cluster 7 <-- OTEK LANC
Cluster 8 <-- P IM STEN
Cluster 9 <-- No class
Cluster 10 <-- No class
Cluster 11 <-- LET IS
Incorrectly clustered instances :
                                       538.0 46.0222 %
```

Gerarchico Complete-Linkage con 12 cluster

```
Class attribute: class
Classes to Clusters:
                     5 6 7 8 9 10 11 <-- assigned to cluster
Cluster 0 <-- FIBR PREDS
Cluster 1 <-- ZSN
Cluster 2 <-- P IM STEN
Cluster 3 <-- NO DISEASE
Cluster 4 <-- OTEK LANC
Cluster 5 <-- No class
Cluster 6 <-- No class
Cluster 7 <-- No class
Cluster 8 <-- No class
Cluster 9 <-- A_V_BLOK
Cluster 10 <-- LET IS
Cluster 11 <-- No class
Incorrectly clustered instances :
                                     824.0 70.4876 %
```

```
Class attribute: class
Classes to Clusters:
                            7 8 9 10 11 <-- assigned to cluster
Cluster 0 <-- FIBR PREDS
Cluster 1 <-- FIBR JELUD
Cluster 2 <-- DRESSLER
Cluster 3 <-- ZSN
Cluster 4 <-- NO DISEASE
Cluster 5 <-- P IM STEN
Cluster 6 <-- No class
Cluster 7 <-- No class
Cluster 8 <-- No class
Cluster 9 <-- LET IS
Cluster 10 <-- OTEK LANC
Cluster 11 <-- A V BLOK
Incorrectly clustered instances :
                                              63.0453 %
```

```
Class attribute: class
Classes to Clusters:
                         6 7 8 9 10 11 <-- assigned to cluster
                        4 16 21 3 73 0 | NO DISEASE
Cluster 0 <-- REC IM
Cluster 2 <-- NO DISEASE
Cluster 3 <-- ZSN
Cluster 4 <-- A V BLOK
Cluster 5 <-- OTEK LANC
Cluster 6 <-- No class
Cluster 7 <-- FIBR PREDS
Cluster 9 <-- JELUD TAH
Cluster 10 <-- P IM STEN
Cluster 11 <-- LET IS
Incorrectly clustered instances :
                                     868.0 74.2515 %
```

Gerarchico Average-Linkage con 12 cluster

```
Class attribute: class
Classes to Clusters:
                                     9 10 11 <-- assigned to cluster
Cluster 0 <-- P IM STEN
Cluster 1 <-- ZSN
Cluster 2 <-- NO DISEASE
Cluster 3 <-- OTEK_LANC
Cluster 4 <-- No class
Cluster 5 <-- No class
Cluster 6 <-- No class
Cluster 7 <-- No class
Cluster 8 <-- LET IS
Cluster 9 <-- No class
Cluster 10 <-- No class
Cluster 11 <-- No class
Incorrectly clustered instances :
                                       836.0 71.5141 %
```

```
Class attribute: class
Classes to Clusters:
                             7 8 9 10 11 <-- assigned to cluster
Cluster 0 <-- NO DISEASE
Cluster 2 <-- PREDS TAH
Cluster 3 <-- No class
Cluster 4 <-- No class
Cluster 5 <-- LET IS
Cluster 6 <-- No class
Cluster 7 <-- No class
Cluster 8 <-- OTEK LANC
Cluster 9 <-- A V BLOK
Cluster 10 <-- No class
Cluster 11 <-- No class
Incorrectly clustered instances :
                                       507.0 43.3704 %
```

```
Class attribute: class
Classes to Clusters:
                                     9 10 11 <-- assigned to cluster
Cluster 0 <-- NO DISEASE
Cluster 1 <-- ZSN
Cluster 2 <-- No class
Cluster 3 <-- No class
Cluster 4 <-- No class
Cluster 5 <-- No class
Cluster 6 <-- No class
Cluster 7 <-- DRESSLER
Cluster 8 <-- P IM STEN
Cluster 9 <-- No class
Cluster 10 <-- OTEK LANC
Cluster 11 <-- LET IS
Incorrectly clustered instances :
                                      539.0 46.1078 %
```

Gerarchico Ward-Linkage con 12 cluster

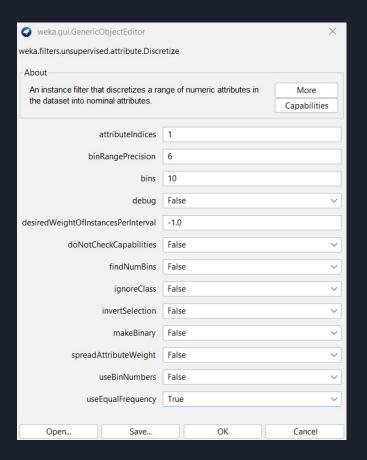
```
Class attribute: class
Classes to Clusters:
        2 3 4 5 6 7 8 9 10 11 <-- assigned to cluster
Cluster 0 <-- FIBR PREDS
Cluster 1 <-- JELUD TAH
Cluster 2 <-- ZSN
Cluster 3 <-- No class
Cluster 4 <-- P IM STEN
Cluster 5 <-- REC IM
Cluster 6 <-- DRESSLER
Cluster 7 <-- FIBR JELUI
Cluster 8 <-- NO DISEASE
Cluster 9 <-- A V BLOK
Cluster 10 <-- OTEK_LANC
Cluster 11 <-- LET IS
Incorrectly clustered instances :
                                             80.4106 %
```

```
Class attribute: class
Classes to Clusters:
                            7 8 9 10 11 <-- assigned to cluster
Cluster 0 <-- ZSN
Cluster 1 <-- OTER LANC
Cluster 2 <-- FIBR JELUD
Cluster 3 <-- NO DISEASE
Cluster 4 <-- P IM STEN
Cluster 5 <-- A V BLOK
Cluster 6 <-- REC IM
Cluster 7 <-- PREDS TAH
Cluster 8 <-- FIBR PREDS
Cluster 9 <-- JELUD TAH
Cluster 10 <-- LET IS
Cluster 11 <-- DRESSLER
Incorrectly clustered instances :
                                      950.0 81.266 %
```

```
Class attribute: class
Classes to Clusters:
 0 1 2 3 4 5 6 7 8 9 10 11 <-- assigned to cluster
 34 79 23 93 77 43 68 10 78 26 54 44 | NO DISEASE
 1 4 0 3 1 2 0 1 1 0 1 0 | JELUD TAH
24 4 4 0 0 6 5 4 2 1 3 2 | LET IS
Cluster 0 <-- LET IS
Cluster 1 <-- ZSN
Cluster 2 <-- No class
Cluster 3 <-- NO DISEASE
Cluster 4 <-- DRESSLER
Cluster 5 <-- OTEK_LANC
Cluster 6 <-- PREDS TAH
Cluster 7 <-- JELUD TAH
Cluster 8 <-- P IM STEN
Cluster 9 <-- REC IM
Cluster 10 <-- FIBR PREDS
Cluster 11 <-- A V BLOK
Incorrectly clustered instances :
                                     975.0 83.4046 %
```

Algoritmo	Metodo	Dati	Istanze	Commento
Aigoritino	Wietodo	Dati	clusterizzate	Commento
			incorrettamente	
K-Means		Dati FS	79.98%	
		Dati PCA	79.81%	Weka riesce ad associare i
				cluster alle classi
L		Dati no PCA/FS	80.15%	
E-M	Non specificando i	Dati FS	71.42%	
	cluster	Dati PCA	74.59%	
		Dati no PCA/FS	69.89%	
	12 cluster	Dati FS	81.26%	
		Dati PCA	79.38%	
		Dati no PCA/FS	72.71%	Weka riesce ad associare i
				cluster alle classi
Gerarchico	Single-Linkage	Dati FS	44.91%	Quasi tutte le istanze fanno
	242245 342465	Dati PCA	45.94%	parte di un cluster
		Dati no PCA/FS	46.02%	
	Complete-Linkage	Dati FS	70.49%	Quasi tutte le istanze fanno
		Dati PCA	63.05%	parte di alcuni cluster
		Dati no PCA/FS	74.25%	
	Average-Linkage	Dati FS	71.51%	Quasi tutte le istanze fanno
			100000000000000000000000000000000000000	parte di alcuni cluster
		Dati PCA	43.37%	Quasi tutte le istanze fanno
				parte di due cluster
		Dati no PCA/FS	46.11%	Quasi tutte le istanze fanno
				parte di un cluster
	Ward	Dati FS	80.41%	
	3	Dati PCA	81.26%	Weka riesce ad associare i
			to was autopart SPS A SPaties	cluster alle classi
		Dati no PCA/FS	83.40%	

- Preparazione del dataset (eliminazione valori nulli, imputazione, encoding).
- Discretizzazione delle variabili quantitative con approccio EqualFrequency.



Discretizzazione

Il numero di bin è stato scelto in modo da avere circa lo stesso numero di istanze in ogni bin.

Attributo	#bins	
AGE	10	
S_AD_ORIT	8	
D_AD_ORIT	3	
K_BLOOD	4	
NA_BLOOD	5	
L_BLOOD	6	
ROE	11	

Algoritmo testati in Weka:

Apriori:

 Pro: permette di scegliere quali attributi devono essere presenti nel conseguente della regola.

Contro: molto lento e richiede molta

RAM.

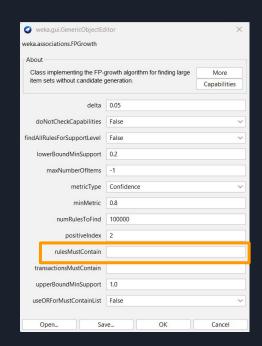


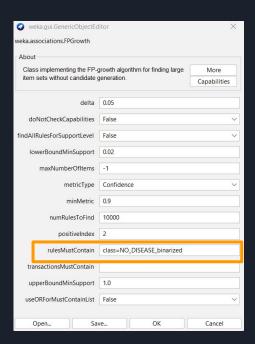
FPGrowth:

- Pro: veloce e richiede meno RAM di Apriori.
- Contro: permette di scegliere solo quali attributi devono essere presenti nelle regole, senza specificare se devono essere presenti nell'antecedente o nel conseguente.

Sono state cercate le regole di associazione:

- senza vincoli sugli attributi presenti nelle regole;
- con il vincolo che le regole devono contenere una complicazione. Questo è stato fatto variando il parametro rulesMustContain e considerando solo le regole che hanno la relativa complicazione nel conseguente.





Senza il vincolo

Con il vincolo

Senza il vincolo supporto 0.2

```
1. [NA BLOOD='(132.368421-132.868421]' binarized=1]: 246 ==> [K BLOOD='(3.702632-3.85]' binarized=1]: 239 <conf:(0.97)> lift:(3.65) lev:(0.15) conv:(22.57)
 2. [ritm_ecg_p_01_binarized=1, IBS_POST_1_binarized=1]: 267 ==> [FK_STENOK_2_binarized=1]: 243 <conf:(0.91)> lift:(1.9) lev:(0.1) conv:(5.56)
 3. [IBS POST 1 binarized=1]: 378 ==> [FK STENOK 2 binarized=1]: 343 <conf: (0.91) > lift: (1.89) lev: (0.14) conv: (5.47)
 4. [GEPAR S n binarized=1, IBS POST 1 binarized=1]: 266 ==> [FK STENOK 2 binarized=1]: 240 <conf: (0.9) > lift: (1.88) lev: (0.1) conv: (5.13)
 5. [ANT CA S n binarized=1, IBS FOST 1 binarized=1]: 264 ==> [FK STENOK 2 binarized=1]: 238 <conf: (0.9)> lift: (1.88) lev: (0.1) conv: (5.09)

    [ASP S n binarized=1, IBS POST 1 binarized=1]: 281 ==> [FK STENOK 2 binarized=1]: 253 <conf: (0.9) > lift: (1.88) lev: (0.1) conv: (5.05)

 7. [ritm ecg p 01 binarized=1, GEPAR S n binarized=1, ANT CA S n binarized=1, SEX 1 binarized=1]: 288 ==> [ASP S n binarized=1]: 250 <conf: (0.87)> lift: (1.13) lev: (0.02) conv: (1.72)
 8. [GEPAR S n binarized=1, SEX 1 binarized=1, class=NO DISEASE binarized=1]: 328 ==> [ASP S n binarized=1]: 281 <conf: (0.86)> lift: (1.12) lev: (0.03) conv: (1.59)
 9. [ritm ecg p 01 binarized=1, GEPAR S n binarized=1, ANT CA S n binarized=1]: 411 ==> [ASP S n binarized=1]: 352 <conf: (0.86) > lift: (1.12) lev: (0.03) conv: (1.59)
10. [ritm ecq p 01 binarized=1, GEPAR S n binarized=1, SEX 1 binarized=1]: 430 ==> [ASP S n binarized=1]: 368 <conf: (0.86)> lift: (1.12) lev: (0.03) conv: (1.59)
11. [ritm ecg p 01 binarized=1, GEPAR S n binarized=1, class=NO DISEASE binarized=1]: 345 ==> [ASP S n binarized=1]: 295 <conf:(0.86)> lift:(1.11) lev:(0.03) conv:(1.57)
12. [GEPAR S n binarized=1, ANT CA S n binarized=1, SEX 1 binarized=1]: 374 ==> [ASP S n binarized=1]: 317 <conf: (0.85) > lift: (1.1) lev: (0.03) conv: (1.5)
13. [ritm ecq p 01 binarized=1, GEPAR S n binarized=1, GB 2 binarized=1]: 311 ==> [ASP S n binarized=1]: 263 <conf: (0.85) > lift: (1.1) lev: (0.02) conv: (1.48)
14. [ASP S n binarized=1, D AD ORIT='(78-87.5]' binarized=1]: 278 ==> [GEPAR S n binarized=1]: 235 <conf:(0.85)> lift:(1.19) lev:(0.03) conv:(1.85)
15. [ritm ecg p 01 binarized=1, GEPAR S n binarized=1]: 601 ==> [ASP S n binarized=1]: 508 <conf:(0.85)> lift:(1.1) lev:(0.04) conv:(1.49)
16. [GEPAR S n binarized=1, SEX 1 binarized=1]: 562 ==> [ASP S n binarized=1]: 473 <conf: (0.84) > lift: (1.1) lev: (0.04) conv: (1.45)
17. [GEPAR S n binarized=1, class=NO DISEASE binarized=1]: 451 ==> [ASP S n binarized=1]: 379 <conf: (0.84) > lift: (1.1) lev: (0.03) conv: (1.44)
19. [GEPAR S n binarized=1, ANT CA S n binarized=1, lat im 1 binarized=1]: 313 ==> [ASP S n binarized=1]: 262 <conf:(0.84)> lift:(1.09) lev:(0.02) conv:(1.4)
21. [GEPAR S n binarized=1. GB 2 binarized=1]: 431 ==> [ASP S n binarized=1]: 360 <conf:(0.84)> lift:(1.09) lev:(0.03) conv:(1.39)
22. [GEPAR S n binarized=1, ANT CA S n binarized=1, class=NO DISEASE binarized=1]: 316 ==> [ASP S n binarized=1]: 263 <conf:(0.83)> lift:(1.08) lev:(0.02) conv:(1.36)
24. [ritm ecg p 01 binarized=1, GEPAR S n binarized=1, lat im 1 binarized=1]: 330 ==> [ASP S n binarized=1]: 274 <conf: (0.03) > lift: (1.08) lev: (0.02) conv: (1.35)
25. [GEPAR S n binarized=1, D AD ORIT='(87.5-inf)' binarized=1]: 289 ==> [ASP S n binarized=1]: 239 <conf:(0.83)> lift:(1.08) lev:(0.01) conv:(1.32)
26. [GEPAR S n binarized=1, SEX 1 binarized=1, lat im 1 binarized=1]: 287 ==> [ASP S n binarized=1]: 236 <conf: (0.82) > lift: (1.07) lev: (0.01) conv: (1.28)
28. [GEPAR S n binarized=1, ANT CA S n binarized=1, GB 2 binarized=1]: 307 ==> [ASP S n binarized=1]: 252 <conf:(0.82)> lift:(1.07) lev:(0.01) conv:(1.28)
29. [ritm ecq p 01 binarized=1, ANT CA S n binarized=1, lat im 1 binarized=1]: 327 ==> [ASP S n binarized=1]: 268 <conf:(0.82)> lift:(1.07) lev:(0.01) conv:(1.27)
30. [ritm ecq p 01 binarized=1, ANT CA S n binarized=1, GB 2 binarized=1]: 310 ==> [ASP S n binarized=1]: 253 <conf: (0.82)> lift: (1.06) lev: (0.01) conv: (1.24)
31. [GEPAR S n binarized=1, lat im 1 binarized=1]: 449 ==> [ASF S n binarized=1]: 366 <conf: (0.82) > lift: (1.06) lev: (0.02) conv: (1.24)
32. [GEPAR S n binarized=1, IBS POST 2 binarized=1]: 322 ==> [ASP S n binarized=1]: 262 <conf:(0.81)> lift:(1.06) lev:(0.01) conv:(1.23)
33. [ASP S n binarized=1, ritm ecg p 01 binarized=1, SEX 1 binarized=1]: 454 ==> [GEPAR S n binarized=1]: 368 <conf: (0.81)> lift: (1.15) lev: (0.04) conv: (1.53)
34. [ANT CA S n binarized=1, GB 2 binarized=1]: 436 ==> [ASP S n binarized=1]: 352 <conf:(0.81)> lift:(1.05) lev:(0.01) conv:(1.19)
35. [ritm ecq p 01 binarized=1, GB 2 binarized=1]: 426 ==> [ASP S n binarized=1]: 343 <conf:(0.81)> lift:(1.05) lev:(0.01) conv:(1.18)
36. [ANT CA S n binarized=1, lat im 1 binarized=1]: 450 ==> [ASP S n binarized=1]: 362 <conf:(0.8)> lift:(1.05) lev:(0.01) conv:(1.18)
37. [ANT CA S n binarized=1, D AD ORIT='(87.5-inf)' binarized=1]: 305 ==> [ASP S n binarized=1]: 245 <conf:(0.8)> lift:(1.05) lev:(0.01) conv:(1.16)
38. [class=NO DISEASE binarized=1, GB 2 binarized=1]: 345 ==> [ASP S n binarized=1]: 277 <conf:(0.8)> lift:(1.05) lev:(0.01) conv:(1.16)
39. [lat im 1 binarized=1, GB 2 binarized=1]: 350 ==> [ASP S n binarized=1]: 281 <conf:(0.8)> lift:(1.05) lev:(0.01) conv:(1.16)
40. (ANT CA S n binarized=1, IBS POST 2 binarized=1): 338 ==> [ASP S n binarized=1]: 271 <conf:(0.8)> lift:(1.04) lev:(0.01) conv:(1.16)
41. [ritm ecg p 01 binarized=1, ANT CA S n binarized=1]: 583 ==> [ASP S n binarized=1]: 467 <conf:(0.8)> lift:(1.04) lev:(0.02) conv:(1.16)
42. [ritm ecq p 01 binarized=1, ANT CA S n binarized=1, class=No DISEASE binarized=1]: 356 ==> [ASF S n binarized=1]: 285 <conf:(0.8)> lift:(1.04) lev:(0.01) conv:(1.15)
43. [ASP S n binarized=1, SEX 1 binarized=1]: 591 ==> [GEPAR S n binarized=1]: 473 <conf:(0.8)> lift:(1.13) lev:(0.05) conv:(1.45)
44. [SEX 1 binarized=1, class=NO DISEASE binarized=1]: 445 ==> [ritm ecq p 01 binarized=1]: 356 <conf: (0.8) > lift: (1.12) lev: (0.03) conv: (1.42)
45. [ANT CA 8 n binarized=1. SEX 1 binarized=1. class=NO DISEASE binarized=11: 315 ==> [ritm ecg n 01 binarized=11: 252 | <conf: (0.8) > lift: (1.12) lev: (0.02) conv: (1.41)
```

Vincolo NO_DISEASE supporto 0.01 confidenza 0.6

```
[IBS_POST_2_binarized=1, ant im_1_binarized=1, K_BLOOD='(3.85-4.45]' binarized=1]: 31 ==> [class=NO_DISEASE_binarized=1]: 29 <conf:(0.94)> lift :(1.74)
[ritm_ecg_p_01_binarized=1, ANT_CA_S_n_binarized=1, SEX_1_binarized=1]: 31 ==> [class=NO_DISEASE_binarized=1]: 30 ==> [class=NO_DISEASE_binarized=1]: 28 <conf:(0.93)> lift :(1.73)
[ASF_S_n_binarized=1, ANT_CA_S_n_binarized=1, ANT_CA_S_n_binarized=1, SEX_1_binarized=1, lat_im_1_binarized=1, AGE='(44.5-51.5]'_binarized=1]: 28 ==> [class=NO_DISEASE_binarized=1]: 26 <conf:(0.93)> lift :(1.73)
[GEPRA_S_n_binarized=1, ANT_CA_S_n_binarized=1, ant_im_1_binarized=1, TIME_B_S_9_binarized=1]: 27 ==> [class=NO_DISEASE_binarized=1]: 25 <conf:(0.93)> lift :(1.72)
[ASF_S_n_binarized=1, ritm_ecg_p_01_binarized=1, ANT_CA_S_n_binarized=1, SEX_1_binarized=1, lat_im_1_binarized=1, AGE='(44.5-51.5]'_binarized=1]: 27 ==> [class=NO_DISEASE_binarized=1]: 27 ==> [class=NO_DISEASE_binarized=1]: 25 <conf:(0.93)> lift :(1.72)
[ASF_S_n_binarized=1, IBS_POST_2_binarized=1, ant_im_1_binarized=1, K_BLOOD='(3.85-4.45]'_binarized=1]: 26 ==> [class=NO_DISEASE_binarized=1]: 24 <conf:(0.92)> lift :(1.72)
[ASF_S_n_binarized=1, IBS_POST_2_binarized=1, ant_im_1_binarized=1, ant_im_1_binarized=1, ant_im_1_binarized=1, ANB_LOOD='(13.65-140.5]'_binarized=1]: 26 ==> [class=NO_DISEASE_binarized=1]: 24 <conf:(0.92)> lift :(1.72)
[ASF_S_n_binarized=1, ANT_CA_S_n_binarized=1, SEX_1_binarized=1, AGE='(44.5-51.5]'_binarized=1]: 35 ==> [class=NO_DISEASE_binarized=1]: 24 <conf:(0.92)> lift :(1.72)
[ASF_S_n_binarized=1, ANT_CA_S_n_binarized=1, lat_im_1_binarized=1, AGE='(44.5-51.5]'_binarized=1]: 35 ==> [class=NO_DISEASE_binarized=1]: 32 <conf:(0.91)> lift :(1.72)
[ASF_S_n_binarized=1, ANT_CA_S_n_binarized=1, lat_im_1_binarized=1, AGE='(44.5-51.5]'_binarized=1]: 34 ==> [class=NO_DISEASE_binarized=1]: 31 <conf:(0.91)> lift :(1.72)
[ASF_S_n_binarized=1, ANT_CA_S_n_binarized=1, lat_im_1_binarized=1, AGE='(44.5-51.5]'_binarized=1]: 34 ==> [class=NO_DISEASE_binarized=1]: 31 <conf:(0.91)> lift :(1.75)
[ASF_S_n_binarize
```

Le 10 regole di associazione con confidenza più alta. Con supporto 0.2 si ricavano regole di associazione con Confidenza ≈0.6

Vincolo FIBR_PREDS

supporto 0.01 confidenza 0.6

```
Scheme: weka.associations.FPGrowth -P 2 -I -1 -N 100000 -T 0 -C 0.6 -D 0.05 -U 1.0 -M 0.01 -rules class=FIBR_PREDS_binarized

Relation: df_final-weka.filters.unsupervised.attribute.Discretize-F-B10-M-1.0-R1-precision6-weka.filters.unsupervised.attribute.Discretize-F-B8-M-1.0-R25-precision6-weka.filters.unsupervised.attribute.Discretize-F-B8-M-1.0-R25-precision6-weka.filters.unsupervised.attribute.Discretize-F-B8-M-1.0-R25-precision6-weka.filters.unsupervised.attribute.Discretize-F-B8-M-1.0-R25-precision6-weka.filters.unsupervised.attribute.Discretize-F-B8-M-1.0-R25-precision6-weka.filters.unsupervised.attribute.Discretize-F-B8-M-1.0-R25-precision6-weka.filters.unsupervised.attribute.Discretize-F-B8-M-1.0-R25-precision6-weka.filters.unsupervised.attribute.Discretize-F-B8-M-1.0-R25-precision6-weka.filters.unsupervised.attribute.Discretize-F-B8-M-1.0-R25-precision6-weka.filters.unsupervised.attribute.Discretize-F-B8-M-1.0-R25-precision6-weka.filters.unsupervised.attribute.Discretize-F-B8-M-1.0-R25-precision6-weka.filters.unsupervised.attribute.Discretize-F-B8-M-1.0-R25-precision6-weka.filters.unsupervised.attribute.Discretize-F-B8-M-1.0-R25-precision6-weka.filters.unsupervised.attribute.Discretize-F-B8-M-1.0-R25-precision6-weka.filters.unsupervised.attribute.Discretize-F-B8-M-1.0-R25-precision6-weka.filters.unsupervised.attribute.Discretize-F-B8-M-1.0-R25-precision6-weka.filters.unsupervised.attribute.Discretize-F-B8-M-1.0-R25-precision6-weka.filters.unsupervised.attribute.Discretize-F-B8-M-1.0-R25-precision6-weka.filters.unsupervised.attribute.Discretize-F-B8-M-1.0-R25-precision6-weka.filters.unsupervised.attribute.Discretize-F-B8-M-1.0-R25-precision6-weka.filters.unsupervised.attribute.Discretize-F-B8-M-1.0-R25-precision6-weka.filters.unsupervised.attribute.Discretize-F-B8-M-1.0-R25-precision6-weka.filters.unsupervised.attribute.Discretize-F-B8-M-1.0-R25-precision6-weka.filters.unsupervised.attribute.Discretize-F-B8-M-1.0-R25-precision6-weka.filters.unsupervised.attribute.Discretize-F-B8-M-1.0-R25-precision6-wek
```

Nessuna regola di associazione con FIBR_PREDS nel conseguente. Non è possibile diminuire il supporto.

Vincolo REC_IM

supporto 0.01 confidenza 0.6

Nessuna regola di associazione con REC_IM nel conseguente. Non è possibile diminuire il supporto.

Vincolo PREDS_TAH

supporto 0.01 confidenza 0.6

Nessuna regola di associazione. Non è possibile diminuire il supporto.

Vincolo DRESSLER

supporto 0.01 confidenza 0.6

```
=== Run information ===
              weka.associations.FPGrowth -P 2 -I -1 -N 100000 -T 0 -C 0.6 -D 0.05 -U 1.0 -M 0.01 -rules class=DRESSLER binarized
Scheme:
              df final-weka.filters.unsupervised.attribute.Discretize-F-B10-M-1.0-R1-precision6-weka.filters.unsupervised.attribute.Discretize-F
Relation:
Instances:
              1169
Attributes:
              217
              [list of attributes omitted]
=== Associator model (full training set) ===
FPGrowth found 125 rules (displaying top 125)
Showing only rules that contain: class=DRESSLER binarized
 1. [LID S n binarized=1, class=DRESSLER binarized=1]: 14 ==> [GEPAR S n binarized=1]: 13 <conf:(0.93)> lift:(1.31) lev:(0) conv:(2.05)
  2. [ant im 4 binarized=1, class=DRESSLER binarized=1]: 13 ==> [GEPAR S n binarized=1]: 12 <conf: (0.92) > lift: (1.3) lev: (0) conv: (1.9)
  3. [ant im 4 binarized=1, class=DRESSIER binarized=11: 13 ==> [SEX 1 binarized=11: 12 <conf:(0.92)> lift:(1.39) lev:(0) conv:(2.2)
```

Nessuna regola di associazione con DRESSLER nel conseguente. Non è possibile diminuire il supporto.

Vincolo A_V_BLOK

supporto 0.01 confidenza 0.6

Nessuna regola di associazione con A_V_BLOK nel conseguente. Non è possibile diminuire il supporto.

Vincolo P_IM_STEN

supporto 0.01 confidenza 0.6

Nessuna regola di associazione con P_IM_STEN nel conseguente. Non è possibile diminuire il supporto.

Vincolo OTEK_LANC

supporto 0.01 confidenza 0.6

```
Scheme: weka.associations.FPGrowth -P 2 -I -1 -N 100000 -T 0 -C 0.6 -D 0.05 -U 1.0 -M 0.01 -rules class=OTEK_LANC_binarized

Relation: df_final-weka.filters.unsupervised.attribute.Discretize-F-B10-M-1.0-R1-precision6-weka.filters.unsupervised.attribute.Discretize-F-B8-
Instances: 1169
Attributes: 217

[list of attributes omitted]
=== Associator model (full training set) ===

FPGrowth found 153 rules (displaying top 153)

Showing only rules that contain: class=OTEK_LANC_binarized

1. [ritm_ecg_p_01_binarized=1, class=OTEK_LANC_binarized=1]: 15 ==> [ASP_S_n_binarized=1]: 15 <conf:(1)> lift:(1.3) lev:(0) conv:(3.49)

2. [D_AD_ORIT='(78-87.5]'_binarized=1, class=OTEK_LANC_binarized=1]: 12 ==> [ASP_S_n_binarized=1]: 12 <conf:(1)> lift:(1.3) lev:(0) conv:(2.79)

3. [LID_S_n_binarized=1, class=OTEK_LANC_binarized=1]: 13 ==> [ASP_S_n_binarized=1]: 12 <conf:(1)> lift:(1.3) lev:(0) conv:(2.79)
```

Nessuna regola di associazione con OTEK_LANC nel conseguente. Non è possibile diminuire il supporto.

Vincolo JELUD_TAH

supporto 0.01 confidenza 0.6

Nessuna regola di associazione con JELUD_TAH nel conseguente. Non è possibile diminuire il supporto.

Vincolo FIBR_JELUD

supporto 0.01 confidenza 0.6

```
=== Run information ===

Scheme: weka.associations.FFGrowth -P 2 -I -1 -N 100000 -T 0 -C 0.6 -D 0.05 -U 1.0 -M 0.01 -rules class=FIBR_JELUD_binarized

Relation: df_final-weka.filters.unsupervised.attribute.Discretize-F-B10-M-1.0-R1-precision6-weka.filters.unsupervised.attribute.Discretize-F-

Instances: 1169

Attributes: 217

        [list of attributes omitted]
=== Associator model (full training set) ===

FFGrowth found 200 rules (displaying top 200)

Showing only rules that contain: class=FIBR_JELUD_binarized

1. [class=FIBR_JELUD_binarized=1]: 21 ==> [LID_S_n_binarized=1]: 21 <conf:(1)> lift:(3.55) lev:(0.01) conv:(15.09)

2. [ASP_S_n_binarized=1, class=FIBR_JELUD_binarized=1]: 15 ==> [LID_S_n_binarized=1]: 15 <conf:(1)> lift:(3.55) lev:(0.01) conv:(10.78)

3. [ritm ecg p 01 binarized=1, class=FIBR_JELUD_binarized=1]: 18 ==> [LID_S_n_binarized=1]: 18 <conf:(1)> lift:(3.55) lev:(0.01) conv:(12.93)
```

Nessuna regola di associazione con FIBR_JELUD nel conseguente. Non è possibile diminuire il supporto.

Vincolo LET_IS supporto 0.01 confidenza 0.9

Le 10 regole di associazione con confidenza più alta.

Non inserendo vincoli tra i parametri, FPGrowth ricava regole di associazione con un'alta confidenza, utilizzando un supporto minimo 0.2.

Aggiungendo i vincoli invece, solo con NO_DISEASE si ricavano regole di associazione utilizzando un supporto minimo 0.2, ma con confidenza intorno a 0.6.

Problema:

Non sono state ricavate regole di associazione relative ad alcune complicazioni.

Soluzione:

Aumentare il numero di istanze di quelle classi o diminuire il supporto (potrebbero esserci regole poco significative).

	Supporto	Confidenza	Commento
Senza vincoli	0.2	> 0.6	Molte regole
NO_DISEASE	0.2	1	Regole con confidenza
			≈ 0.6
NO_DISEASE	0.01	> 0.6	Molte regole
FIBR_PREDS	0.01	> 0.6	Nessuna regola con
			FIBR_PREDS nel
			conseguente
REC_IM	0.01	> 0.6	Nessuna regola con
			REC_IM nel conseguente
PREDS_TAH	0.01	/	Nessuna regola
DRESSLER	0.01	> 0.6	Nessuna regola con
			DRESSLER nel conseguente
A_V_BLOK	0.01	> 0.6	Nessuna regola con
			A_V_BLOK nel
			conseguente
P_IM_STEN	0.01	> 0.6	Nessuna regola con
2000 43290			P_IM_STEN nel
			conseguente
OTEK_LANC	0.01	> 0.6	Nessuna regola con
			OTEK_LANC nel
			conseguente
JELUD_TAH	0.01	> 0.6	Nessuna regola con
6005			JELUD_TAH nel
			conseguente
FIBR_JELUD	0.01	> 0.6	Nessuna regola con
			FIBR_JELUD nel
			conseguente
LET_IS	0.01	> 0.9	Molte regole