

First & Last name: Ourmassi Alaa Eddine

Specialty: RSSI Groupe 02

TP 02 : Weka

Introduction :

EN: Ulcerative colitis is a long-term condition where the colon and rectum become inflamed. The colon is the large intestine (bowel) and the rectum is the end of the bowel where stools are stored.







FR: La recto-colite hémorragique (RCH) est une maladie inflammatoire du rectum et du gros intestin (côlon). Elle se manifeste par une diarrhée accompagnée de sang et, le plus souvent, de douleurs abdominales intenses.

Partie explorer :

Dataset Name : Ulcerative_colitis

Number of instances = 52

Number of attributes = 06

-  age
 - ✓ 10-19
 - ✓ 20-29
 - ✓ 30-39
 - ✓ 40-49
 - ✓ 50-59
 - ✓ 60-69
 - ✓ 70-79
-  Gender
 - ✓ Female
 - ✓ Male
-  Type
 - ✓ Ulcerative_proctitis
 - ✓ Proctosigmoiditis
 - ✓ left_sided_colitis
 - ✓ Pancolitis
 - ✓ Fulminant_Colitis
-  Pain
 - ✓ hight_crampy
 - ✓ medium_crampy
 - ✓ low_crampy
-  Death
 - ✓ yes
 - ✓ no
-  Class
 - ✓ no_recover
 - ✓ recover

Algorithme 1:

OneR :

Options	Correctly Classified Instances	Precision	Recall	F-Measure
Cross validation fold=50	82.6923 %	0,837	0,827	0,829
Cross validation fold=25	82.6923 %	0,837	0,827	0,829
Cross validation fold=05	82.6923 %	0,837	0,827	0,829
Cross validation fold=12	82.6923 %	0,837	0,827	0,829
Pourcentage split = 66%	27.7778 %	0,214	0,278	0,242
Pourcentage split = 33%	57.1429 %	0,544	0,571	0,554
Pourcentage split = 11%	58.6957 %	0,537	0,587	0,540
Pourcentage split = 90%	80 %	0,850	0,800	0,781

Algorithme 2 :

Naïve Bayes :

Options	Correctly Classified Instances	Precision	Recall	F-Measure
Cross validation fold=50	76.9231 %	0,798	0,769	0,773
Cross validation fold=25	76.9231 %	0,798	0,769	0,773
Cross validation fold=05	76.9231 %	0,798	0,769	0,773
Cross validation fold=12	76.9231 %	0,798	0,769	0,773

Pourcentage split = 66%	77.7778 %	0,787	0,778	0,772
Pourcentage split = 33%	80 %	0,797	0,800	0,792
Pourcentage split = 11%	69.5652 %	0,696	0,696	0,696
Pourcentage split = 90%	80 %	0,850	0,800	0,781

Algorithme 3 :

Id3:

Options	Correctly Classified Instances	Precision	Recall	F-Measure
Cross validation fold=50	55.7692 %	0,653	0,659	0,656
Cross validation fold=25	55.7692 %	0,653	0,659	0,656
Cross validation fold=05	59.6154 %	0,689	0,689	0,689
Cross validation fold=12	55.7692 %	0,653	0,659	0,656
Pourcentage split = 66%	44.4444 %	0,389	0,444	0,398
Pourcentage split = 33%	28.5714 %	0,540	0,556	0,519
Pourcentage split = 11%	60.8696 %	0,563	0,609	0,556
Pourcentage split = 05%	61.2245 %	?	0,612	?

Algorithme 4 :

C45:

Options	Correctly Classified Instances	Precision	Recall	F-Measure
Cross validation fold=50	82.6923 %	0,837	0,827	0,829
Cross validation fold=25	82.6923 %	0,837	0,827	0,829
Cross validation fold=05	82.6923 %	0,837	0,827	0,829
Cross validation fold=12	82.6923 %	0,837	0,827	0,829
Pourcentage split = 66%	94.4444 %	0,951	0,944	0,945
Pourcentage split = 33%	65.7143 %	?	0,657	?
Pourcentage split = 11%	63.0435 %	?	0,630	?
Pourcentage split = 90%	80 %	0,850	0,800	0,781

Algorithme 5 :

KNN

K=3, avec Euclidean distance

Options	Correctly Classified Instances	Precision	Recall	F-Measure
Cross validation fold=50	75 %	0,772	0,750	0,754
Cross validation fold=25	75 %	0,772	0,750	0,754
Cross validation fold=05	73.0769 %	0,748	0,731	0,735

Cross validation fold=12	75 %	0,772	0,750	0,754
Pourcentage split = 66%	50 %	0,294	0,500	0,370
Pourcentage split = 33%	65.7143 %	0,617	0,657	0,597
Pourcentage split = 11%	65.2174 %	0,776	0,652	0,535
Pourcentage split = 90%	80 %	0,850	0,800	0,781

The decision tree:

