**TD 2**

**EX1 : weka**

**EX2 :**

classifieur naif de Bayse

Probabilités conditionnelles

Le naive Bayes classifier se base sur le théorème de Bayes. Ce dernier est un classique de la théorie des probabilités. Ce théorème est fondé sur les probabilités conditionnelles.

Probabilités conditionnelles : Quelle est la probabilité qu’un événement se produise sachant qu’un autre événement s’est déjà produit.

Pour bien comprendre le classifier naif de Bayse on va s'entrainer sur les données fournit par le texte ou l'ennoncé du ex2 de td2 .

Etape 1 :

d'après une lecture attentive de notre texte on a eu une aidé sur les caractéristiques de chaque spécim .

les attributs de chaque specime est :

@attribute odeur @attribute chapeau @attribute spores

@attribute lieu @attribute lisseté @attribute class

L’attribut classe est l'attribut principale (comestible/atoxique) sur lequel on vas se baser pour classier nos attributs on verra ca par la suite.

Étape 2 :

Dans la 2eme étape on va extraire les valeurs de chaque attribut cité là-dessus.

@attribute odeur {anis,amande,piquant}

@attribute chapeau {jaune,blanc}

@attribute spores {violet,marron,noir}

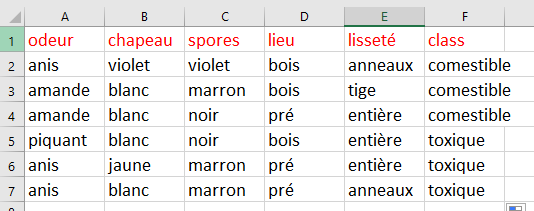
@attribute lieu {bois,pré}

@attribute lisseté {anneaux,tige,entière}

@attribute class {comestible,toxique}

Et puis construire le train data pour appliquer là-dessus notre algorithme de classification.

Train data sous forma <arff> et <csv>



@DATA

anis,jaune,violet,bois,anneaux,comestible

amande,blanc,marron,bois,tige,comestible

amande,blanc,noir,pré,entière,comestible

piquant,blanc,noir,bois,entière,toxique

anis,jaune,marron,pré,entière,toxique

anis,blanc,marron,pré,anneaux,toxique

Le format <csv> pour faire des calcules sur excel en utilisant l’algo naif de Baise on verra par la suite comment appliquer naife de baise pour classifier et obtenir l’ensemble des probabilités qui vas nous permettre de trouver la connaissance et avoir la prédiction.

Le format <arff> pour faire une classification de naife de Bayse a l’aide de Weka sans perdre du temps sur les calculs manuelle .

Etapes 3 :

Avant de commencer les calculs de **P(Xi/Cj)**

Xi l’ensemble des valeurs {i…n} d’un attribut X

**Cj** c’est notre classe principale qui indique si un champignon est toxique ou non.

Le principe d’avoir tout ces probabilités est pour de construire une connaissance qui va nous permettre de faire une prédiction sur d’autre champignon si ils sont toxique ou non a la base de son caractéristiques.

Dans notre exemple on doit répondre à la problématique suivante :

Un champignon {anis,jaune,marron,bois,tige} est il toxique ou non …. ?

1 : il faut calculer les probabilités de P(Xi/Cj) de chaque valeur d’attribut xi pour chaque class Cj comme il est indiqué là-dessous :

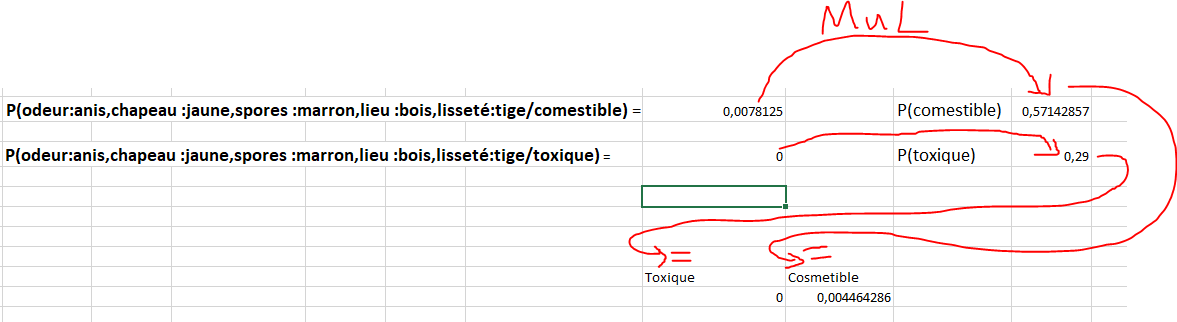
|  |  |
| --- | --- |
| P(comestible) | 0,571428571 |
| P(toxique) | 0,29 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  | **P(ODEUR/CLASS)** | |  |  |
| odeur | comestible | toxique |  |  |  |  |
| anis | 2 | 1 | P(anis/comes) = | 0,5 | P(anis/tox) = | 0,5 |
| piquant | 0 | 1 | P(piquant/comes)= | 0 | P(piquant/tox)= | 0,5 |
| amande | 2 | 0 | P(amande/comes)= | 0,5 | P(amande/tox)= | 0 |
| total | 4 | 2 |  |  |  |  |
|  |  |  |  |  |  |  |
| chapeau | comestible | toxique | **P(CHAPEAU/CLASS)** |  |  |  |
| violet | 1 | 0 | P(violet/comes) = | 0,25 | P(anis/tox) = | 0 |
| blanc | 2 | 2 | P(blanc/comes) = | 0,5 | P(piquant/tox)= | 1 |
| jaune | 1 | 0 | P(jaune/comes) = | 0,25 | P(amande/tox)= | 0 |
| total | 4 | 2 |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| spores | comestible | toxique | **P(SPORES/CLASS)** |  |  |  |
| violet | 1 | 0 | P(violet/comes) = | 0,25 | P(anis/tox) = | 0 |
| marron | 2 | 1 | P(blanc/comes) = | 0,5 | P(piquant/tox)= | 0,5 |
| noir | 1 | 1 | P(jaune/comes) = | 0,25 | P(amande/tox)= | 0,5 |
| total | 4 | 2 |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| lieu | comestible | toxique | **P(LIEU/CLASS)** |  |  |  |
| bois | 2 | 1 | P(violet/comes) = | 0,5 | P(anis/tox) = | 0,5 |
| pré | 2 | 1 | P(blanc/comes) = | 0,5 | P(piquant/tox)= | 0,5 |
| total | 4 | 2 |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| lissté | comestible | toxique | **P(SPORES/CLASS)** |  |  |  |
| anneaux | 1 | 1 | P(violet/comes) = | 0,25 | P(anis/tox) = | 0,5 |
| tige | 1 | 0 | P(blanc/comes) = | 0,25 | P(piquant/tox)= | 0 |
| entière | 2 | 1 | P(jaune/comes) = | 0,5 | P(amande/tox)= | 0,5 |
| total | 4 | 2 |  |  |  |  |

2 : la 2eme étape c’est de calculer la probabilité d’avoir un champignon toxique ou comestible selon les caractéristique suivant : {odeur :anis,chapeau :jaune,spores :marron,lieu :bois,lisseté :tige} ->[ testing data]

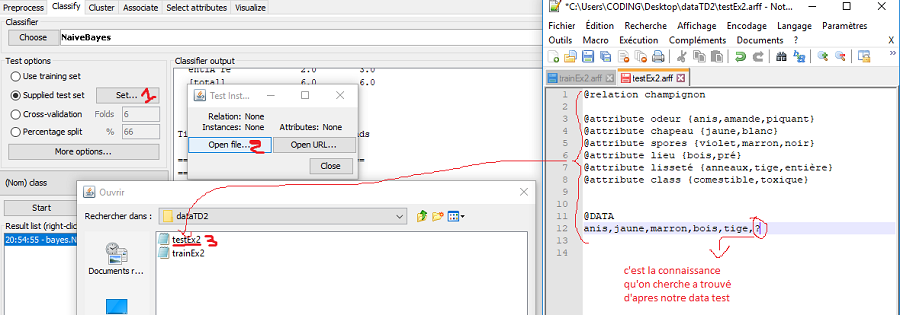
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
|  | | | | | |  |
|  |  |  |  |  |  |  |
|  | | | | | |  |

3 : la dernière étape va nous servir a savoir si le champignon est toxique ou pas avec une % de certitude de notre prédiction :

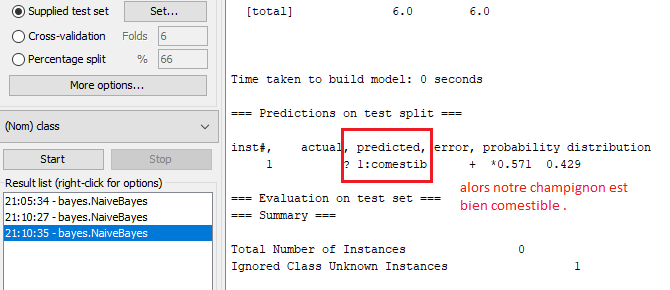


D’après cette résultat notre champignon est comestible parce que P comestible est > que toxique .

**Solution avec weka**



D’après effectué notre test sur le data train et avec l’utilisation d’une classification naif de Bayes on peut prédire que notre champignon est comestible.



EX3 :

1.

Naif de Bayse

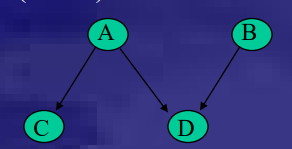
-est très facile à programmer, sa mise en œuvre est aisée ; d’autre part, parce que l’estimation de ses paramètres, la construction du modèle, est très rapide sur de très grandes bases de données, que ce soit en nombre de variables ou en nombre d’observations.

-s. Il y a beaucoup de méfiance vis-à-vis du classifieur bayesien naïf parce qu’il est mal compris, parce que le modèle n’est pas interprétable.

- On ne voit pas très bien de quelle manière chaque variable pèse sur la décision. L’interprétation des résultats n’est pas aisée

Reseau Bayésien

Modèles de représentation des connaissances, fondés sur une description graphique des variables aléatoires : Directed Acyclic Graph (DAG)



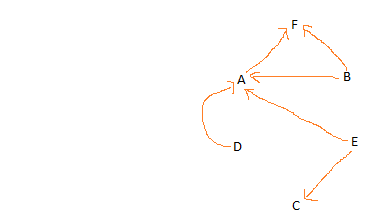
Apport des réseaux bayésiens

• Ils conjuguent les avantages de diverses approches : – la compréhensibilité des modèles symboliques – les fondements probabilistes rigoureux des méthodes statistiques – la structure en réseau de composants simples des approches connexionistes

• Les réseaux bayésiens représentent toutes les relations entre les attributs (décrivant les exemples)

• Et permettent une utilisation multidirectionnelle

2.



3.

4.