

Compte-rendu du TP n°2

Préparation des Données

1) Le jeu de données sont manquants pour les intervalles de températures]72.4, 74.5] et]76.6, 78.7].

2) Liste des intervalles générés :

Intervalle	Nombre de valeurs
] -inf, 69.25]	4
]69.25, 74.5]	4
]74.5, 79.75]	2
]79.75, inf[4

3) Liste des intervalles générés :

Intervalle	Nombre de valeurs
] -inf, 64.5]	1
]64.5, 66.5]	1
]66.5, 69.5]	2
]69.5, 70.5]	1
]70.5, 71.5]	1
]71.5, 73.5]	2
]73.5, 77.5]	2
]77.5, 80.5]	1
]80.5, 84]	2
]84, inf[1

4) Liste des intervalles générés :

Intervalle	Nombre de valeurs
] -inf, 70.5]	5
]70.5, 77.5]	5
]77.5, inf[4

5) Description des 4 premières barres de l'histogramme :

N°	Label	Nombre de valeurs
1	65	1
2	70	3
3	75	1
4	80	2

6) Description de l'histogramme obtenu :

N°	Label	Nombre de valeurs
1	neg_5-10	7
2	pos_5-10	7

7) Après une rapide observation des histogrammes, on remarque que plus les valeurs de température et d'humidité sont élevées, plus la possibilité de jouer est restreinte. Des valeurs faibles en températures et en humidité permettent alors de jouer de façon optimale.

8) Instances de (temperature, humidity) vérifiant play=yes :

temperature	humidity	Nombre d'occurrences
cool	low-medium	3
	high	1
medium	low-medium	2
	high	1
hot	low-medium	1
	high	1

9) Listing des valeurs représentatives des attributs, pour en déterminer la jouabilité :

Attribut	Représentant play=yes	Représentant play=no
outlook	overcast	sunny
temperature	cool	hot
humidity	low-medium	high
windy	FALSE	TRUE

Règles d'association

10) L'algorithme ne peut être lancé avec ce fichier, cela provient du fait que l'algorithme ne peut être utilisé qu'avec des valeurs nominales

11) Règles les plus sûres pour prédire la possibilité de jouer (play = yes) :

- outlook = overcast
- outlook = rainy && windy = FALSE
- temperature = cool && windy = FALSE

12) Règles les plus sûres pour prédire l'impossibilité de jouer (play = no) :

- outlook = sunny && temperature = hot
- outlook = sunny && humidity = high
- outlook = rainy && windy = TRUE

13 → 16) L'algorithme Prism n'est pas disponible sur l'installation Weka du poste de TP.

Comparaison de méthodes

17) La première règle de temperature impliquant que l'on peut jouer est : temperature = cool.

18) La première règle de temperature impliquant que l'on ne peut pas jouer est : temperature = hot.

19) Cela confirme les différentes estimations effectuées dans les parties précédentes.

20) Les premières règles de l'attribut humidity ...

- Impliquant que l'on peut jouer : humidity = low.
- Impliquant que l'on ne peut pas jouer : humidity = high.

Cela confirme, les hypothèses sur l'influence de l'humidité, établies précédemment. (low-medium représentant de play=yes, et high représentant de play=no).