

### projet 6 NSI :

Dans un premier temps nous importons les modules dont nous avons besoin :

- pandas pour manipuler et analyser les données de type CSV (comma separated values)
- sklearn pour l'apprentissage des données et leur exploitation pour l'algorithme knn et les prédictions

Le programme fonctionne sur trois fonctions principales, une pour lire le fichier CSV (dans notre cas nommé *Titanic.csv*) à l'aide du module pandas, une pour entraîner notre algorithme knn grâce au module sklearn et enfin une dernière qui permet à l'utilisateur de créer un naufragé fictionnel pour tester l'algorithme knn et voir si selon lui le naufragé aurait survécu :

-la première fonction nommée *read\_titanic\_data* permet de lire le fichier *titanic.csv* et de le modifier tout d'abord en supprimant les colonnes contenant des données incomplètes et inutiles ensuite en modifiant les données 'male' et 'female' par 1 et 0 pour que le module sklearn puisse les utiliser et en modifiant aussi la colonne 'cabine' ensuite on crée deux tables différentes, une avec le Label "Survived" nommée *titanic\_label* et une autre avec les données 'Sex', 'Cabin', 'Pclass' et 'Age' nommée *titanic\_data*.

-La fonction suivante s'appelle *modeltraining* et se sert des données traitées par la première fonction pour entraîner l'IA de knn avec le module sklearn, pour cela on prend comme label la table *titanic\_label* et comme données la table *titanic\_data* créées précédemment. Pour entraîner l'IA il faut consacrer une partie des données à son entraînement et une autre partie pour la tester et voir à quel point elle répond juste, la commande *train\_test\_split* nous permet de choisir ce partage de données, nous avons choisi d'utiliser 80% des données pour l'entraînement et 20% pour les tests, on calcule ensuite le pourcentage de réponses justes, pour ce programmes nous avons une moyenne de réponses juste d'environ 70%, le maximum que nous avons atteint a été de 81,82% et le minimum d'environ 54,6%.

```
longeur total de la table titanic_csv : 162
Longeur de la table d'entrainement : 129
Longeur de la table de test : 33
Nombre d'exemple utilisé : 3
Taux de prediction juste : 75.76 %
```

-Dans la troisième et dernière fonction *survivor\_prediction* on demande d'abord à l'utilisateur l'id, la classe, l'age et le sexe d'un naufragé inventé de toute pièce ou qui fait partie de la liste du fichier csv, puis on remplace ensuite le sexe par 1 ou 0 puis on demande ensuite la cabine dans laquelle se trouve le passager puis on transforme pour que ce soit utilisable. Enfin on cherche à savoir si le passager a survécu ou non et on affiche le résultat. L'IA a donc environ 70% de chance de donner une prédiction