

CREATION D'UNE APPLICATION CONSOLE POUR LE DIAGNOSTIC MÉDICAL

Ludovic Laflamme (étudiant) & Yacine Yaddaden

Département de mathématiques, informatique et génie, Université du Québec à Rimouski



Problématique

Dans un contexte médical, les victimes des maladies du cœur se retrouvent dans plusieurs catégories en fonction des attributs spécifiques de leur condition. Toutefois, il est difficile de déterminer à quelle catégorie une nouvelle victime appartient.[1] ...

Objectifs

Développer un programme qui permet de récupérer les attributs d'une personne souffrant de maladies du cœur et ensuite déterminer à quelle catégorie cette personne appartient. Spécifiquement,

- Implémenter un algorithme robuste de comparaison de données.
- Récupérer l'ensemble des données historiques et les traduire en objets informatiques utilisables.
- Utilisation d'algorithmes de tri pour organiser les données.

Méthodologie suivie

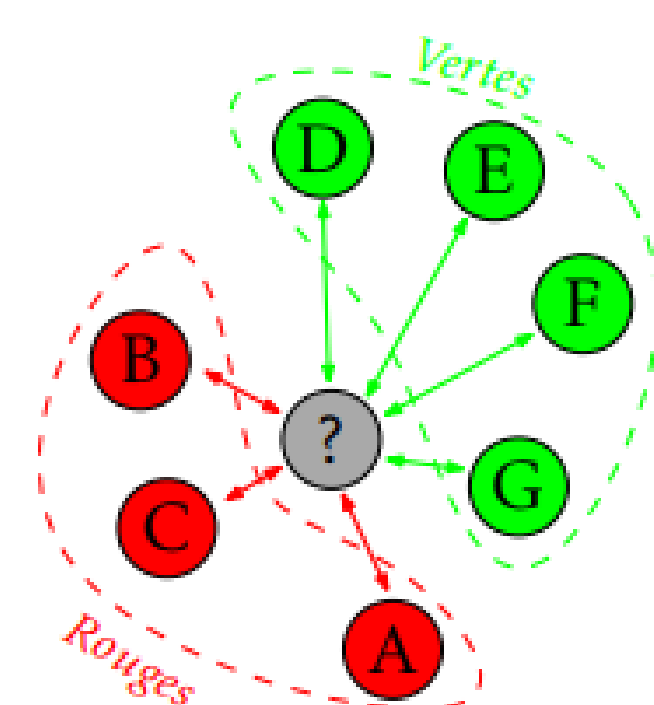
Les données se retrouvent dans des fichiers .csv elles sont récupérées à l'aide de la librairie CsvHelper qui facilite leur manipulation. Il s'agit d'une base de données composée de plusieurs colonnes représentant les caractéristiques numériques de chaque victime. Il en est de même pour l'élément inconnu à analyser. Une fois les données en banque, chaque élément représentant une victime est intégré dans le programme et trié dans une liste.

La distance euclidienne est calculée à avec les caractéristiques sur une rangée dans la base de données. La distance de l'élément dont la catégorie est inconnue est ensuite comparé à tous ceux dans la liste à l'aide de l'algorithme "k-Nearest Neighbors", qui récupère un nombre "k" d'éléments de la liste dont les attributs sont semblables et établit ainsi la catégorie de notre inconnu.

$$D_{\text{Euclidienne}} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2} \quad (1)$$

avec $A = [x_1, y_1, z_1]$ et $B = [x_2, y_2, z_2]$ deux échantillons avec les attributs : x , y et z .

Pour l'exécution du k-NN, supposons que $k = 3$:



Les distances :	Le tri :
$D_A = 3$	$D_C = 1$
$D_B = 2$	$D_B = 2$
$D_C = 1$	$D_G = 2$
$D_D = 5$	$D_A = 3$
$D_E = 5$	$D_F = 4$
$D_F = 4$	$D_D = 5$
$D_G = 2$	$D_E = 5$

FIGURE 1 – Exemple de k-NN

Évaluation

L'évaluation des résultats est démontrée à travers la matrice de confusion qui compare la similitude de l'élément inconnu à ses plus proches voisins. Établissant ainsi la précision de l'algorithme.

Résultats

Dans le but de tester l'exactitude du programme, nous possédons les diagnostics d'experts de nos données "inconnues". La catégorie de ce diagnostic est comparée à notre résultat, affichant ainsi si le modèle a réussi à calculer la catégorie appropriée.

```
Confusion Matrix :
      True  False
True  165   4
False  23  138
Classification Accuracy → 91,81818 %
```

FIGURE 2 – Mesure de performance

```
[ sample 01 ] Prediction by model → False | by expert → False
```

FIGURE 3 – Prédiction d'un échantillon

Conclusion

Une application réelle de ce projet permettrait de déterminer précisément la catégorie à laquelle une personne souffrant de maladies du cœur appartient sans avoir besoin d'un expert. Cette personne pourrait ensuite suivre un traitement approprié.

Références

[1] Yacine Yaddaden. TP01 – Creation d'une application console pour le diagnostic médical. UQAR, 2023.