Application Windows pour le diagnostic médical automatique - Mediag

Fabien Airaud (Étudiant) & Yacine Yaddaden (Professeur)

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

Objectifs

L'objectif de ce projet est de développer une **application Windows en C#** dans le but d'effectuer un **diagnostic médical automatique**, que ce soit de la maladie du cœur [1] ou du cancer du sein [3].

- ✓ Récupération des données du dataset CSV
- ✓ Entraînement et test de l'arbre de décision
- ✓ Prédictions des données médicales des patients
- ✓ Enregistrement des données et résultats

Technologies utilisées

Microsoft Visual Studio Community ⟶ programmation langage C#

CSVHelper ⟶ gestion des fichiers CSV

WPF de .Net ⟶ interface application Windows

ORM **Entity Framework** ⟶ gestion des données vers la base de données SQL

Interface utilisateur

L'interface utilisateur est accédée par les médecins à travers un compte utilisateur. Elle leur permet de gérer les entités de l'hôpital ainsi que l'arbre de décision et le dataset CSV.
L'interface respecte le **motif d'architecture MVVM** qui vise à séparer la logique en 3 couches :

- Les modèles (***Model***) pour la gestion des données de la logique métier
- Les vues (***View***) sont l'interface pour l'utilisateur
- Les modèles de vue (***ViewModel***) permettent le lien entre les données et les vues

Figure 1 – Accueil application Mediag

Arbre de décision

Un arbre de décision est un algorithme d'**apprentissage supervisé** très simple [2].
Chaque **noeud** de l'arbre représente une **caractéristique des données** et les liens sont choisis en fonction de leurs valeurs. Le **dernier noeud** de chaque branche est appelé une feuille et représente le **résultat** de la décision.

1. Déterminer la **meilleure caractéristique** de l'ensemble par le **ratio de gain** (amélioration du gain d'information)
2. **Diviser** les données en **sous-ensembles** contenant les valeurs possibles de la meilleure caractéristique
3. **Générer** récursivement les **sous-arbres**
4. On s'arrête lorsqu'on arrive aux **feuilles** (résultats)

Méthodologie suivie

La figure 2 résume la structure logique de l'application.

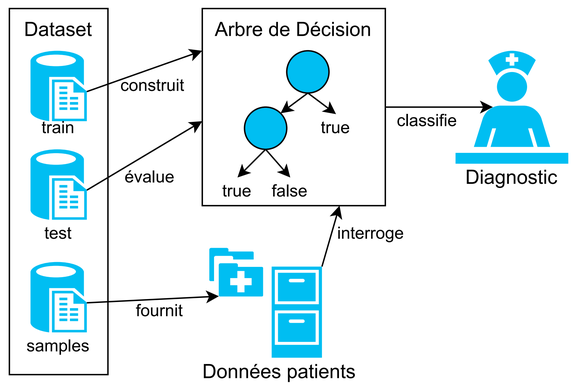

Figure 2 – Structure logique de l'application.

1. **Extraction des données** du dataset comprenant entraînement (*train*), tests (*test*) et échantillons (*samples*)
2. **Entraînement** de l'arbre de décision (données train)
3. **Evaluation** de l'arbre de décision (données test)
4. Enregistrement des échantillons en tant que données médicales des premiers patients
5. Les médecins peuvent demander à l'arbre de décision d'effectuer une **prédiction** sur la maladie des patients, qui recevront un **diagnostic**.

Évaluation

- Les datasets sont séparés pour environ **67%** des données pour l'**entraînement** et **33%** des données pour les **tests**.
- L'arbre de décision prédit toutes des données de test et renvoie une **matrice de confusion**.
- On extrait aussi la **précision** (*accuracy*), représentée par la proportion de données correctement prédites sur l'ensemble des données.

Résultats

act // pred	malade	bénin
malade	58	10
bénin	5	110

Table 1 – Cancer du sein

act // pred	malade	bénin
malade	140	21
bénin	21	147

Table 2 – Maladie du coeur

- La table 1 représente la matrice de confusion pour le cancer du sein avec **183 données testées**, pour une **précision** de presque **92%**.
- La table 2 représente la matrice de confusion pour la maladie du coeur avec **330 données testées**, pour une **précision** de presque **87%**.

Conclusion

- ✓ Approfondissement des connaissances en C#
- ✓ Découverte de WPF pour les UI Windows
- ✓ Fonctionnement détaillé des arbres de décision
- ✓ Bons résultats malgré la simplicité des arbres.

Références

[1] Pfisterer Matthias Janosi Andras, Steinbrunn William and Detrano Robert. Heart Disease. UCI Machine Learning Repository, 1988.

[2] Ian H. Witten, Eibe Frank, and Mark A. Hall. *Data Mining : Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann, Amsterdam, 3 edition, 2011.

[3] Street Nick Wolberg William, Mangasarian Olvi and Street W. Breast Cancer Wisconsin (Diagnostic). UCI Machine Learning Repository, 1995.