

SAE R4.04 Méthodes d'Optimisation

SAE R4.04 Méthodes d'Optimisation



Genetet Maud
Kadri Yanis
Leveque Loïc
Petit Juliette
Tomera Clément

Sommaire

Sommaire	1
Introduction	2
Présentation du contexte et du sujet de recherche	2
Objectifs	2
Questions	3
Glossaire	3
Principe des arbres de décisions	3
Définition	3
Comment ça marche	4
Premier essai : Arrêts Cardiaques	5
Choix et description des données	5
Description du jeu de données	5
Méthodologie	6
Résultats	7
Arbre de décision après entraînement (gini et entropy)	7
Résultat des test	8
Interprétation premier essai	8
Conclusion et Hypothèses	9
Deuxième essai : Espérance de vie	11
Choix et description des données	11
Méthodologie	12
Résultats	13
Extrait de l'arbre de décision après entraînement (gini)	14
Interprétation	15
Conclusion deuxième essai	15
Conclusion Finale	16
Récapitulation des résultats de la recherche	16
Réponse aux questions de recherche et hypothèses établies	16
Suggestions pour des recherches futures	16
Bibliographie	17
Annexes	17

Introduction

Présentation du contexte et du sujet de recherche

Ce projet est dans le cadre du cours de Méthode d'optimisation du BUT2 Informatique à l'IUT de Bordeaux, enseigné par Mr Casteigts.

Nous avons 1 mois pour réaliser ce projet en groupe de 5 étudiants, pour faire son rendu le 14 mars 2023.

La consigne de ce mini-projet, en équipe, consistait à reproduire un exemple d'utilisation de l'algorithme KNN ou d'un arbre de décision, avec les données de notre choix.

Le résultat à fournir est un notebook Python restituant nos algorithmes et nos recherches et ce rapport décrivant le cheminement de notre projet pour à l'issue produire une soutenance de 10 minutes devant nos professeurs.

Objectifs

Pour ce projet nous avons décidé de nous tourner vers le monde de la santé car on voulait faire un projet qui nous permet de nous rapprocher de la réalité et qui nous permet de voir comment les nouvelles technologies peuvent nous aider réellement dans notre vie quotidienne.

Notre but est donc d'analyser l'espérance de vie de personnes en fonction de plusieurs paramètres comme le pays, le niveau de vie, la consommation de tabac, l'alcool, les maladies, etc... et de voir si on peut trouver des corrélations entre ces paramètres et l'espérance de vie pour pouvoir ainsi essayer de prédire l'espérance de vie d'une personne en fonction de ces paramètres et ainsi essayer de trouver des moyens pour améliorer l'espérance de vie des personnes.

Nous avons aussi choisi ce thème car beaucoup de données sur la santé sont disponibles sur internet, cela était donc beaucoup plus pratique pour effectuer ce sujet.

Notre objectif à l'issue de ce projet est aussi de produire un notebook de nos recherches et de vous décrire notre cheminement à travers ce rapport.

Questions

Après avoir choisit ce sujet nous avons choisit d'utiliser un algorithme de machine Learning basé sur les arbres de décisions car on a pensé que cela serait plus visuel et compréhensible pour les gens qui ne sont pas familiers avec les algorithmes de machine learning, vu que notre but est de faire un projet qui vise à aider les gens.

Les arbre de décision sont donc plus adapter à la vulgarisation d'après nous, or part la suite de cette reflexions des questions se sont posées :

- Est ce que les arbres de décisions sont les plus appropriés techniquement pour ce projet ?
- Aurait-il un algorithme plus approprié techniquement pour ce projet ?

Nous essayerons donc de répondre à ces questions dans la suite de ce document en parallèle de notre réflexion sur le sujet de base car nous avons supposé que les arbres de décisions ne sont peut-être pas les plus appropriés techniquement pour ce projet.

Glossaire

Machine Learning : Technique qui permet aux ordinateurs d'apprendre à partir d'exemples, sans avoir besoin d'être explicitement programmés pour résoudre un problème spécifique.

KNN : L'algorithme KNN (k-plus proches voisins) est une méthode d'apprentissage automatique. L'idée de base de l'algorithme KNN est de placer nos données dans plusieurs plans pour ensuite avec des données que l'on veut traiter les placer à leur tour dans les différents plans et par la suite de trouver les points voisin afin d'utiliser la majorité de leurs valeurs (pour la classification) ou la moyenne de leurs valeurs (pour la régression) pour prédire la valeur du point que l'on veut.

Kaggle : Site en ligne de base de données gratuites

Principe des arbres de décisions

Définition

Arbre de décision : Méthode d'analyse de données qui permet de prendre des décisions en se basant sur des règles simples et hiérarchisées. Ces règles sont représentées sous forme d'un arbre, où chaque sommet représente une décision et chaque branche représente une conséquence de cette décision, jusqu'à mener à une feuille qui est donc le résultat de l'analyse.

Comment ça marche

Un arbre de décision est un apprentissage supervisé, c'est-à-dire qu'au départ nous avons un jeu de données et pour chaque données il est relié à son label, autrement dit sa classe associée. Ensuite ce jeu de données il faut le séparer en deux, en données d'apprentissage (70% des données) et en jeu de test (30% des données). Comme expliqué précédemment, un arbre il faut l'entraîner pour qu'il se forme, l'arbre d'entraîne avec les données d'apprentissage. Une fois son entraînement ou son apprentissage fait, avec des nouvelles données, on peut les classer.

Dans les grandes lignes, l'arbre de décision pendant son entraînement fait un algorithme glouton et va essayer de trouver l'arbre le plus court possible, c'est-à-dire faire la prédiction d'une nouvelle donnée le plus rapide donc une profondeur de l'arbre le plus petit possible. Pour cela il construit l'arbre par la sélection de la variable ou attribut la plus discriminante en utilisant des critères. Ensuite, il sépare les données en deux ou plusieurs sous ensemble en fonction de la variable choisie.

Enfin, une fois l'algorithme fini, nous avons l'arbre, on peut maintenant faire une prédiction d'une classe sur des nouvelles données en parcourant l'arbre.

Tout d'abord, on part de la racine de l'arbre et on parcourt les branches selon la valeur des caractéristiques de la nouvelle donnée (on regarde la première information si la caractéristique est plus grande ou plus petite que la valeur calculé) jusqu'à arriver à une feuille qui prédit la classe de la nouvelle donnée.

Pour finir, à chaque étape c'est-à-dire à chaque nœud interne, on peut voir des informations. Il peut y avoir une formule utilisée qui permet de prendre des mesures et de choisir la meilleure variable pour séparer les données en sous-ensemble (gini pour mesurer l'erreur de classification ou encore entropy qui mesure de l'incertitude d'un ensemble de données). Il apparaît aussi l'information "samples" qui permet de dire combien y a de donnée est à cette étape. Enfin comme dernière information c'est la valeur "value" qui est un tableau où chaque indice représente une classe, et la valeur à ces indices représente le nombre de personnes qui est dans la classe concernée.

Premier essai : Arrêts Cardiaques

Choix et description des données

Pour commencer nous avons cherché sur le site Kaggle un jeu de données qui nous permettrait de réaliser notre projet. Nous avons donc trouvé un jeu de données qui contient des données sur le potentiel risque de mourir d'un problème cardiaque durant le suivi d'un patient en fonction de plusieurs paramètres comme le sexe, l'âge, le nombre de plaquette dans le sang...

Nous avons choisi ce jeu de données car il contient des données sur le risque de mourir d'un problème cardiaque et nous avons pensé que cela pourrait être utile pour les personnes qui souffrent de ce problème et qui voudraient savoir si leur état va s'aggraver ou non.

Aussi les données sont assez simples et nous avons pensé que cela pourrait être plus facile pour nous de les analyser et de les traiter.

Description du jeu de données

Ce jeu de données a été créé par Svetlana Ulianova il y a 4 ans et contient environ 300 enregistrements de données de patients avec 13 colonnes.

Il y a 3 types d'entrées :

Objectif : informations factuelles;

Examiné : résultats d'examen médical;

Subjectif : informations fournies par le patient.

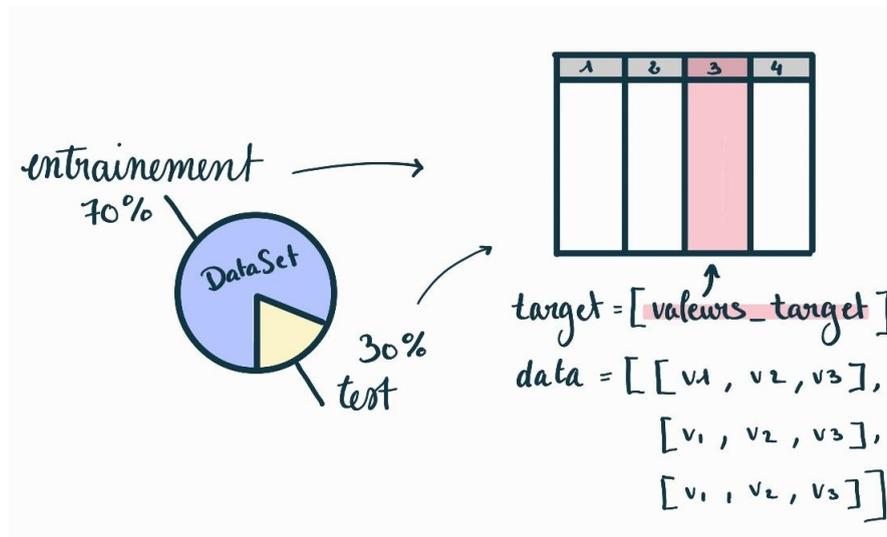
Voici donc les différentes colonnes du jeu de données :

- Age : âge du patient (en année) (Objectif)
- Anémie : diminution du nombre de globules rouges ou de l'hémoglobine (booléen) (examen)
- Créatinine Phosphokinase : niveau de l'enzyme CPK dans le sang (mcg/L) (examen)
- Diabète : si le patient a le diabète (boolean) (examen)
- Fraction d'éjection : pourcentage de sang qui sort du cœur à chaque contraction (pourcentage) (examen)
- Hypertension : si le patient a l'hypertension (boolean) (examen)
- Plaquettes : plaquettes dans le sang (kilo plaquettes/mL) (examen)
- Créatinine sanguine : niveau de la créatinine sanguine dans le sang (mg/dL) (examen)
- Sodium sanguin : niveau du sodium sanguin dans le sang (mEq/L) (examen)
- Sexe : femme ou homme (boolean) (Objectif)
- Fumeur : si le patient fume ou non (boolean) (Subjectif)
- Temps : période de suivi (jours) (Objectif)

Et la colonne cible : Mort : si le patient est décédé pendant la période de suivi (boolean)

Méthodologie

La méthodologie utilisée pour la construction de l'arbre pour ce premier essai consiste à transformer notre dataset en tableau à deux dimensions et de prendre 30% de notre dataset pour le test et 70% pour l'entraînement. Par la suite, nous prenons nos données cibles et les plaçons dans un tableau, puis nous les supprimons de nos deux tableaux. Nous aurons donc un tableau contenant nos données cibles et un autre tableau à deux dimensions contenant nos données, pour le test et l'entraînement.



Ensuite, nous entraînons notre arbre de décision avec notre tableau de données d'entraînement et notre tableau de données cibles d'entraînement, grâce à `tree.DecisionTreeClassifier()`, une méthode de la librairie `sklearn.tree`, qui nous permet de construire notre arbre de décision. Nous utiliserons aussi l'indice gini et l'entropie pour la construction de nos arbres de décisions.

Nous utiliserons `matplotlib` pour afficher notre arbre de décision. Nous utiliserons aussi la méthode `score()` de la librairie `sklearn.tree`, qui nous permet de calculer la précision de notre arbre de décision.

Après notre entraînement, nous avons obtenu deux arbres de décision identique selon gini et l'entropie, car on a un jeu de donnée simple (209 données d'entraînement)

Cet arbre nous permet de déterminer si une personne ayant subi un arrêt cardiaque a une chance de survie ou non. Nous pouvons observer que le premier nœud de l'arbre est relatif au temps que le patient a passé à l'hôpital, plus précisément s'il a passé plus ou moins de 73,5 heures à l'hôpital. Nous constatons ainsi que si le patient a passé plus de 73,5 heures à l'hôpital, il a moins de chances de survie, car il se situe sur un nœud orange (orange = décès, tandis que bleu = survie). À l'inverse, s'il a passé moins de 73,5 heures à l'hôpital, il a plus de chances de survie, car il se situe sur un nœud bleu.

Nous pouvons ensuite observer que pour le nœud bleu, la prochaine caractéristique est le taux de sodium dans le sang, tandis que pour le nœud orange, la prochaine caractéristique est le taux de sang qui sort du cœur. Nous pouvons également voir que ce nœud a d'autres nœuds ou feuilles, et ainsi de suite. L'arbre a donc déduit que ces deux caractéristiques étaient plus discriminantes statistiquement pour ces deux populations.

Une autre observation que nous pouvons faire est que certaines colonnes de notre dataset n'ont pas été utilisées pour la construction de cet arbre, telles que le fait que le patient est fumeur ou non, s' il a du diabète ou encore une bonne pression artérielle.

Cependant, nous pouvons constater que notre arbre de décision est relativement simple et qu'il peut donc être facilement expliqué à quiconque.

Résultat des test

Après avoir construit notre arbre de décision, nous avons utilisé notre jeu de test pour évaluer sa précision. Nous avons obtenu un taux de précision de 0,80, ce qui est relativement élevé. Cela signifie que notre arbre de décision est assez précis et qu'il peut être utilisé pour prédire si un patient a une chance de survie ou non.

Ce résultat indique que nous avons une bonne réponse pour 80% des cas, ce qui est bien meilleur que le hasard qui aurait donné 50% de bonnes réponses.

Interprétation premier essai

En d'autres termes, l'arbre de décision créé à partir du dataset ne prend en compte que des facteurs statistiques et ne tient pas compte des facteurs humains tels que le tabagisme qui peut avoir un impact important sur la survie d'un patient après un arrêt cardiaque, comme dit le site CoeurEtAVC, *"les fumeurs sont 2 fois plus susceptibles de subir une crise cardiaque ou un AVC, et même d'en mourir."*

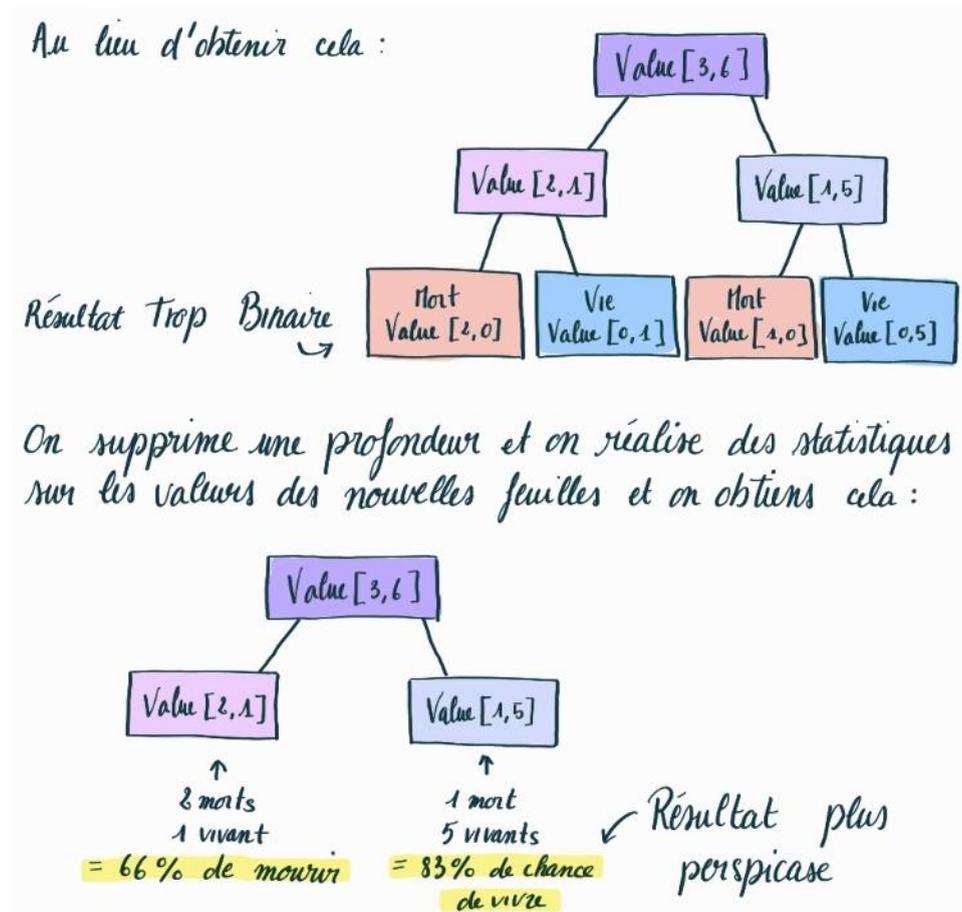
Cela peut remettre en question la fiabilité de notre arbre de décision, car son objectif est simplement de classer les populations de la manière la plus efficace possible, sans tenir compte de la réalité des situations.

Conclusion et Hypothèses

En conclusion, bien que notre arbre de décision donne des résultats satisfaisants sur le plan statistique avec un taux de précision de 80%, son manque de prise en compte des facteurs humains importants tels que le tabagisme, le diabète ou la pression artérielle pose un problème de fiabilité. Par conséquent, il est difficile de conclure que notre arbre de décision est fiable et peut être utilisé pour prédire avec précision les chances de survie d'un patient après un arrêt cardiaque.

Par la suite de l'analyse de nos données, nous n'étions pas aussi totalement satisfaits, car nous avons constaté que le résultat était trop binaire et peu réaliste. Nous avons jugé qu'il était insensé d'utiliser un arbre décisionnel comme celui là pour déterminer le sort d'un patient à l'hôpital, car le résultat était soit "vous vivrez à 100%" ou "c'est fini à 100%", ce qui est peu réaliste. Nous aurions préféré un résultat plus statistique, tel que "46% de probabilité de mourir" ou "91% de chance de vivre", mais notre jeu de données et notre arbre décisionnel ne permettaient que deux résultats possibles, car l'arbre à l'entraînement ne peut pas s'inventer des feuilles.

Afin de résoudre ce problème, nous avons envisagé de raccourcir notre arbre, c'est-à-dire de le rendre moins profond, pour obtenir des feuilles avec des résultats plus variés. Pour illustrer notre idée, voici un exemple graphique :



Cependant, après avoir envisagé cette solution, nous n'étions toujours pas satisfaits, car nous nous sommes demandé comment déterminer la profondeur optimale de notre arbre. En effet, supprimer des nœuds de l'arbre reviendrait en plus à retirer des données de notre analyse, et ces données pourraient être importantes.

Étant donné que nous ne sommes pas des médecins, et ne souhaitant pas jouer aux apprentis sorciers en faisant n'importe quoi avec des données sensibles telles que la santé de personnes et après des recherches infructueuses sur Internet à votre question.

Nous nous sommes dit qu'il est possible que le problème vienne de la nature binaire de la colonne cible, qui ne permet pas une classification plus fine des patients. Pour résoudre ce problème, nous avons décidé de réaliser une seconde expérience en utilisant une colonne cible constituée de plusieurs valeurs.

Deuxième essai : Espérance de vie

Choix et description des données

Nous désirions obtenir un jeu de données similaire, donc toujours sur le thème de la santé. Nous nous sommes arrêtés sur un jeu de données répertoriant les espérances de vie de différents pays du monde. Les chiffres proviennent de l'organisation mondiale de la santé (OMS), qui est une source réputée. Les durées d'espérance de vie sont fournies pour les années allant de 2000 à 2015. Présentons donc les colonnes de ce jeu de données :

- Pays: (String) Le pays que concernent les données
- Année : (int) L'année que concernent les données
- Statut : (Développé / en voie de développement) statut économique du pays
- Espérance de vie : (Float) Moyenne d'âge de la population lors du décès
- Mortalité Adulte : (int) pour mille des habitants morts entre 15 et 60 ans
- Mortalité Infantile : (int) pour mille des habitants de moins d'un an, morts
- Consommation d'alcool : (float) en litres par habitant chez les + de 15 ans
- Pourcentage des dépenses en santé: (float) en pourcentage du PIB/habitant
- Hépatite B : (int) pourcentage de la couverture vaccinale contre l'hépatite B
- Rougeole : (int) pour mille des cas confirmés de rougeole dans la population
- IMC : (float) Indice de masse corporelle de l'ensemble de la population
- Morts sous 5 ans : (int) pour mille des habitants
- Polio : (float) Pourcentage de couverture vaccinale contre la polio des enfants de 1 an
- Dépenses de santé: (float) Pourcentage des dépenses de santé du gouvernement sur le total des dépenses;
- Diphtérie : (float) Pourcentage de couverture vaccinale par l'anatoxine diphtérique, le tétanos et la coqueluche chez les enfants de 1 an
- VIH : (float) Morts du sida pour 1000 naissances (entre 0 et 4 ans)
- PIB : (float) Produit intérieur brut par habitant (en USD)
- Population : (int) Population d'un pays
- Sous-poids 10-19: (float) Prévalence de la sous-nutrition chez les enfants et adolescents de 10 à 19 ans (%)
- Sous-poids 5-9 : (float) Prévalence de la sous-nutrition chez les enfants et adolescents de 5 à 9 ans (%)
- Indice de développement humain : (float) Indice de développement humain en tant que composante du revenu tiré des ressources (indice de 0 à 1) ;
- Education : (int) Nombre d'années de scolarité

Si ce jeu de données contient plus d'attributs que le précédent, et comporte aussi un nombre de données de loin supérieur (près de 3000 entrées pour 193 pays différents). Il ne contient qu'une colonne de données dites "subjectives" : la colonne Statut ; le reste de ces données est soit constatable, calculé ou mesuré. L'idée avec ce jeu de données était de voir si nous pouvions prédire l'espérance de vie d'une personne dans un pays donné avec de nombreux facteurs en entrée.

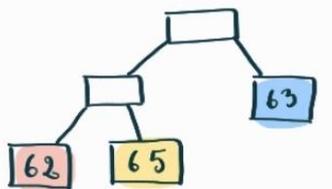
Méthodologie

Notre objectif était de prédire l'espérance de vie, qui est un nombre décimal. Cependant il est impossible de faire deviner une valeur précise d'une variable continue sur un arbre binaire de recherche.

Nous avons considéré que l'utilisation d'intervalles était préférable à celle de valeurs continues pour les arbres de décision, car ces derniers sont plus adaptés à la classification de valeurs avec un nombre fini de classes à déterminer. Si nous n'avions pas utilisé d'intervalles, il aurait été difficile de présenter toutes les valeurs possibles dans l'arbre, ce qui aurait pu conduire à des prévisions incomplètes. Même avec un grand ensemble de données, il est possible que des informations importantes manquent, par exemple si le jeu d'entraînement ne contient pas toutes les tranches d'âge, l'arbre de décision ne sera pas en mesure de prédire pour les personnes qui ne sont pas incluses dans le jeu d'entraînement.

Exemple :

...	62 ans
...	63 ans
...	65 ans
...	..



si la valeur 64 ans n'existe pas dans notre database

⇒

notre arbre ne vas pas non plus contenir 64

or par la suite dans notre jeux de test peut se trouver 64, l'arbre sera donc incapable de trouver la bonne réponse, car il ne peut pas s'inventer des feuilles.

C'est donc pourquoi on a choisi d'utiliser les intervalles car cela permet d'avoir un nombre de target fixes.

Nous avons donc testé différentes granularités de tranches d'âges et nous sommes arrivés aux meilleurs résultats avec des découpes par tranches d'âges de 15 ans (de 0 à 15, de 15 à 30, etc...).

Une prédiction par notre arbre est donc considérée comme juste si elle tombe dans la bonne tranche d'âge. Ceci autorise donc une certaine marge d'erreur à notre problème de prédiction, et élimine le problème de classification des variables continues.

Pour notre banc de tests nous avons isolé des lignes de notre jeu de données, ces valeurs ne faisant bien évidemment pas partie de la base d'entraînement. Nous avons réservé une part de 30% du jeu de données pour les tests, les 70% étant laissées pour l'entraînement de ce second arbre.

Résultats

Notre arbre été entraîné à différencier les espérances de vie dans les catégories de tranche d'âge que nous avons imposées. Après des tests, nous avons déterminé un taux de résultats positifs de 90%

Interprétation

Selon une première interprétation, ce résultat de 90% de positifs est élevé. Il faut dire que les variations de l'espérance de vie d'un pays à l'autre ne sont pas très grandes. Notre arbre de décision résultant ne doit classer que dans trois ou quatre tranches d'âge car on a de grandes intervalles, ce qui n'est pas une tâche "complexe" à proprement parler.

Cela dit, un résultat supérieur à 14% (7 classes) nous confirme que la sélection est basée sur autre chose que le hasard, et donc que l'attribut de la longévité dans un pays est bien un facteur qui a un lien avec les autres attributs de notre jeu de données. L'avantage d'une représentation algorithmique comme celle des arbres est que cette dernière est relativement transparente : on peut savoir quels sont les attributs qui ont la plus grande conséquence sur le problème. Nous nous sommes donc penchés sur la question. Cependant, la classification n'étant pas binaire l'arbre résultant est beaucoup moins lisible que le précédent produit.

Conclusion deuxième essai

A l'issue de notre deuxième essai, le résultat était plus convaincant que le premier car notre taux de résultats positifs à notre phase de tests était plus élevé. Nous avons un jeu de données plus grand que le premier ; et nous étions en conséquence conscients qu'un plus grand nombre d'attributs nous donnerait une plus grande précision dans les résultats. En effet, les calculs d'exactitude sont plus constants dans notre deuxième essai.

Par la suite, il est possible que les excellents résultats de ce second arbre ne soient dus qu'à nos tranches d'âges qui sont relativement larges.

Et que si elle était plus petite, le taux de réussite de notre arbre aurait donc baissé en conséquence.

On peut donc aussi se poser comme question: quelle tranche d'âge optimale devront nous fournir pour que l'utilisation d'un arbre devienne vraiment utile ?

Effectivement la longévité de vie mondiale se trouve autour des 70-80 ans, on peut donc se dire que l'un programme ne retournant que la target ou ce trouve 70-80 ans est possiblement tout aussi efficace qu'un arbre de décision.

Pour répondre à cela nous avons donc regardé qu'elle est le pourcentage de notre population du jeu d'entraînement qui se trouve dans chaque target et nous avons trouvé cela :

Target (ans)	0-15	15-30	30-45	45-60	60-75	75-90	90+	Total
Nombre de ligne	0	0	11	187	678	220	0	1154
en (%)	0	0	0,95	16,20	58,75	19,06	0	100

Cela nous permet donc de conclure que notre arbre de décision est quand même plus efficace qu'un programme qui retourne uniquement la target ou l'on retrouve la majorité de

nos valeurs d'entraînement, ici : "60-75". Ici un programme qui ferait ça aurait donc un pourcentage de réussite de 58,75% contre environ 90% pour notre arbre on peut donc prouver statistiquement que notre arbre est meilleur.

Conclusion Finale

Récapitulation des résultats de la recherche

En somme, nous avons obtenu de très bons résultats sur le premier arbre, avec environ 85% de réussite sur la prédiction de mort ou de décès lors d'un arrêt cardiaque, cependant ce résultat, n'est pas totalement satisfaisant. En effet il s'agit d'une prédiction binaire, et nous aurions préféré obtenir un pourcentage qui justifierait la probabilité de survivre ou de mourir d'un arrêt cardiaque. Cet apprentissage est envisageable, mais pas avec un arbre de décision, plutôt un réseau de neurones.

Le deuxième arbre que nous avons produit permet lui, de prédire l'espérance de vie dans un pays, selon une série de critères démographiques et de santé publique. Il nous permet d'obtenir des résultats assez fiables (de l'ordre de 90%). Cependant les critères de sélection sont précis et nombreux.

Réponse aux questions de recherche et hypothèses établies

Nous nous demandions si, malgré leur indisposition à évaluer des variables continues, les arbres de décision nous permettent tout de même de classer ces valeurs. Nous avons obtenu dans l'ensemble des résultats satisfaisants du point de vue statistique. En revanche, pour ce qui est de l'utilité pratique de l'information fournie par nos arbres, elle est *de facto* quasi nulle. En effet, il serait malvenu en milieu médical, de présenter un pronostic vital binaire. Ou encore avec notre second arbre, il est fort peu pratique de mesurer autant de valeurs telles que l'indice de masse corporel d'un habitant moyen, sans avoir mesuré l'espérance de vie d'un habitant dans le pays.

En somme, nous en sommes venus à la conclusion que présenter des paramètres continus à un arbre de décision est une mauvaise idée. Cela ne devrait pas être l'usage primaire qu'il faut en faire.

Suggestions pour des recherches futures

La prédiction et la classification ne sont pas les seuls usages que l'on peut faire des arbres de décision. La sélection d'attributs est notamment un domaine à explorer. En effet l'approche "à cœur ouvert" d'un arbre de recherche nous permet de voir quels sont les attributs qui sont les plus importants dans la détermination du résultat, notamment avec les notions d'entropie et de gini. Une analyse en ce sens, nous permettrait, dans le cas de notre premier arbre, de déterminer quels sont les facteurs les plus importants à privilégier pour la survie d'un patient lors d'un arrêt cardiaque.

Se pencher sur cette approche pourrait être intéressant dans des recherches futures.

Bibliographie

Cours Méthode d'optimisation de Mr Ramet :

- ramet.gitlab.io/r4.04-optim/

Site (consulté le 14/03/23) :

- [Tabagisme | Fondation des maladies du cœur et de l'AVC](#)

Annexes

Lien du premier jeu de données sur les arrêts cardiaques :

- [Heart Failure Prediction | Kaggle](#)

Lien du second jeu de données sur l'espérance de vie dans le monde :

- [Life Expectancy \(WHO\) | Kaggle](#)