NEXIALOG CONSULTING

NEXIASEARCH

Benchmark sur la corrélation et l'importance des variables : guide des modèles de ML supervisé

Ernesto LOPEZ FUNE
Valentin MESSINA
Amande EDO

THINK SMART ACT DIFFERENT

TABLE DES MATIÈRES

INTRODUCTION

Le développement rapide de l'intelligence artificielle (IA) a conduit à des avancées remarquables dans divers domaines, allant de la reconnaissance d'image à la prédiction financière. Cependant, ces modèles sont souvent perçus comme des "boîtes noires" en raison de la complexité de leurs algorithmes et de la difficulté à comprendre leurs mécanismes de prise de décision. Cette opacité pose un défi majeur pour les utilisateurs et les décideurs qui exigent des modèles plus transparents pour pouvoir se fier aux prédictions générées [1].

Les deux principaux piliers de la modélisation par ML sont l'interprétabilité et l'explicabilité des modèles. L'interprétabilité [2, 3, 4] se réfère à la capacité à comprendre les mécanismes internes d'un modèle permettant à un humain de suivre et de comprendre ses processus internes sans nécessiter de connaissances techniques approfondies. En revanche, l'explicabilité concerne la capacité à fournir des explications sur le comportement du modèle, notamment pourquoi et comment il génère certaines prédictions. Nous nous concentrerons sur les méthodes existantes pour quantifier l'importance des variables prédictives dans la prise de décision, soit par une analyse de corrélation (ad-hoc) soit par des méthodes de mesure de l'importance des variables des modèles de ML supervisé après l'entraînement (post-hoc), en soulignant ainsi leurs avantages et inconvénients. Cette approche par importance de variables est une des méthodes d'explicabilité des modèles de ML qui consiste à attribuer à chaque variable prédictive d'un modèle un score qui quantifie son importance sur les prédictions.

Pour ce faire, nous commencerons par présenter les différents types de corrélation, essentiels pour la sélection de variables et la validation des résultats obtenus via les méthodes de mesure d'importance des variables. Ensuite, nous aborderons la mesure de l'importance des variables par permutation, et examinerons les méthodes de mesure de l'importance des variables telles que les valeurs de Shapley, et LIME. Ces méthodes sont implémentées dans trois packages Python : Eli5, SHAP et LIME, qui permettent une explication globale et locale des modèles de ML supervisé.

I. Importance de la corrélation dans la sélection de variables

1. Différents types de corrélation

Dans cette section, nous présentons les différents types de corrélation et les tests statistiques utiles pour la sélection de variables avant l'optimisation d'un modèle de ML supervisé. En effet, il est important de sélectionner les variables les plus corrélées à la variable cible pour obtenir un modèle prédictif, tout en ne conservant pas de variables trop corrélées entre elles pour éviter la multi-colinéarité. D'autre part, selon le format des variables dont on calcule l'association (catégorielle, numérique, etc.), les calculs de corrélation sont différents. De plus, ces outils sont aussi utiles pour confronter théoriquement les résultats des méthodes de mesure de l'importance des variables vues après.

1.1 Association entre variables catégorielles/nominales

Le test du χ^2 (Khi-deux) [5] est une procédure statistique largement utilisée, dont les résultats sont évalués par référence à la distribution du χ^2. Il permet d'évaluer l'indépendance entre deux variables catégorielles en calculant la différence entre les fréquences observées et attendues sous l'hypothèse nulle d'indépendance, comme détaillé sur *l'Algorithme 1*.

Une statistique du χ^2 élevée indique qu'il existe une différence significative entre les fréquences observées et attendues, suggérant une association entre les deux variables.

La signification de la statistique du χ^2 est déterminée en la comparant à une valeur critique de la distribution du χ^2 avec des degrés de liberté appropriés ou en calculant la p-value. Les limitations du test résident dans le fait qu'il nous indique uniquement si une association existe, et non la force ou la direction de ladite association.

Entrée: Deux variables catégorielles avec des fréquences observées dans un tableau de contingence
Sortie: χ^2−statistique et p-value

Initialiser les variables :
$O_{ij} \leftarrow$ nombre de fréquences observées pour la cellule (i, j).
$R_i \leftarrow$ somme des fréquences observées pour la ligne i.
$C_j \leftarrow$ somme des fréquences observées pour la colonne j.
$N \leftarrow$ nombre total d'observations
$E_{ij} \leftarrow 0$ (fréquence attendue pour la cellule (i, j)).
$\chi^2 \leftarrow 0$ (la statistique du test).
Calculer les fréquences attendues :
foreach *cell* (i, j) *dans le tableau de contingence* **do**
  $E_{ij} \leftarrow \frac{R_i \times C_j}{N}$
  Calculer la statistique du test χ^2 *:*
  foreach *cell* (i, j) *dans le tableau de contingence* **do**
    if $E_{ij} \neq 0$ **then**
      $\chi^2 \leftarrow \chi^2 + \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$
  Déterminer les degrés de liberté :
  $df \leftarrow$ (le nombre de lignes $- 1$) $\times$ (le nombre de colonnes $- 1$).
  Calculer la p-value :
  $p \leftarrow \text{PValue}(\chi^2, df)$
  return χ^2, p.
  Fonction `PValue`(χ^2, df) **:**
    return
    1 - $\text{CDF}_{\chi^2}(df)$.

Algorithme 1 : Test du χ^2 pour l'indépendance.

En complément du test précédent, le coefficient de Cramér V [6] nous fournit une mesure de l'association, allant de 0 à 1, où 0 indique l'indépendance et 1 indique une association parfaite. L'algorithme de calcul de ce test est disponible sur *l'Algorithme 2*. Contrairement au test du χ^2, le coefficient de Cramér V fournit une mesure de la force de l'association. En revanche, une des limites de ce coefficient est qu'il ne fournit pas la direction de ladite association.

Entrée: Deux variables catégorielles avec des fréquences observées dans un tableau de contingence $\mathbf{T}$ de taille $r \times c$ où r est le nombre de lignes et c est le nombre de colonnes.
Sortie: Le coefficient de Cramér V.

Calculer le χ^2 observé à partir du tableau de contingence $\mathbf{T}$.
Calculer la taille de l'échantillon total n : $n = \sum_{i=1}^{r} \sum_{j=1}^{c} T_{ij}$.
Trouver la valeur minimale entre le nombre de lignes r et de colonnes c : $k = \min(r-1, c-1)$.
Utiliser la formule suivante pour calculer Cramér V : $V = \sqrt{\frac{\chi^2}{n \cdot k}}$.
return V.

Algorithme 2 : Coefficient de Cramér V.

En résumé, les deux tests sont utilisés dans le cas où l'on souhaite mesurer une association entre une variable prédictive de format catégoriel et une variable cible discrète. Alors que le test du χ^2 est utilisé pour tester la présence d'une potentielle association, le coefficient de Cramér V est utilisé pour mesurer sa force. Ces deux techniques sont très utiles pour la sélection des variables renforçant ainsi la capacité explicative et l'interprétabilité des modèles de ML supervisé.

1.2 Association entre variables ordinales

D'autre part, la corrélation de rang de Spearman [7] est un test non paramétrique qui évalue la relation monotone entre deux variables ordinales en calculant le coefficient de corrélation basé sur leurs rangs, comme on peut le voir sur *l'Algorithme 3* ci-dessous.

Entrée: Deux ensembles de données de même taille, $X = \{x_1, x_2, \ldots, x_n\}$ et $Y = \{y_1, y_2, \ldots, y_n\}$.
Sortie: Coefficient de corrélation de rang de Spearman ρ.
Trier les ensembles X et Y et attribuer des rangs aux valeurs :
$R_X \leftarrow$ rangs de X, $R_Y \leftarrow$ rangs de Y.
Calculer les différences des rangs pour chaque paire d'éléments :
$d_i \leftarrow R_{X_i} - R_{Y_i}$ pour $i = 1, 2, \ldots, n$.
Élever chaque différence au carré :
$d_i^2 \leftarrow (R_{X_i} - R_{Y_i})^2$ pour $i = 1, 2, \ldots, n$.
Calculer la somme des carrés des différences :
$\sum d_i^2 \leftarrow \sum_{i=1}^{n} d_i^2$.
Calculer le coefficient de corrélation de rang de Spearman :
$\rho \leftarrow 1 - \frac{6 \sum d_i^2}{n(n^2-1)}$.
return ρ.

Algorithme 3 : Calcul de la corrélation de rang de Spearman.

S'il n'y a pas de valeurs de données répétées, une corrélation de Spearman parfaite de +1 ou −1 se produit lorsque chacune des variables est une fonction monotone parfaite de l'autre.

Le signe du coefficient de corrélation de Spearman ρ indique le sens de l'association entre X (la variable prédictive) et Y (la variable cible). Si Y tend à augmenter/diminuer lorsque X augmente, le coefficient de corrélation de Spearman est positif/négatif. Une corrélation de Spearman égale à zéro indique qu'il n'y a pas de tendance claire pour Y à augmenter/diminuer lorsque X augmente.

1.3 Association entre variables numériques

Le test de corrélation le plus connu pour l'association de variables numériques est probablement le test de corrélation de Pearson [8]. Ce test mesure la relation linéaire entre deux variables numériques en évaluant le degré de dépendance linéaire entre elles. Il se calcule comme le rapport entre la covariance des deux variables et le produit de leurs écarts types, ce qui en fait essentiellement une mesure normalisée de la covariance, comme on peut le voir sur *l'Algorithme 4*. Le coefficient de corrélation de Pearson prend toujours une valeur comprise entre −1 et 1. Toutefois, à l'instar de la covariance elle-même, cette mesure ne reflète qu'une corrélation linéaire entre les variables et ignore de nombreux autres types de relations.

Entrée: Deux ensembles de données de même taille, $X = \{x_1, x_2, \ldots, x_n\}$ et $Y = \{y_1, y_2, \ldots, y_n\}$.
Sortie: Coefficient de corrélation de Pearson r.
Calculer les moyenne de X et Y : $\overline{X}$ et $\overline{Y}$.
Calculer la covariance entre X et Y : $\mathrm{Cov}(X, Y)$.
Calculer les écart-types de X et Y : σ_X et σ_Y.
Calculer le coefficient de corrélation de Pearson :
$r = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}$.
return r.

Algorithme 4 : Calcul de la corrélation de Pearson.

Une valeur absolue de 1 du coefficient de corrélation de Pearson r, indique qu'une relation linéaire parfaite existe entre les variables X et Y, où tous les points de données se situent exactement sur une ligne droite. Le signe de la corrélation est déterminé par la pente de la régression : une valeur de +1 signifie que Y augmente linéairement avec X, tandis qu'une valeur de -1 signifie que Y diminue linéairement avec X. En revanche, une valeur de 0 indique l'absence de dépendance linéaire entre les variables. En cas de suspicion d'une corrélation non linéaire entre les variables, par exemple dans le cas d'une loi de puissance du type $Y \sim X^{\alpha}$, ou exponentielle $Y \sim e^{\alpha X}$, la fonction logarithme peut être appliquée aux deux variables et étudier la corrélation linéaire entre les nouvelles variables transformées $\log X, X$ et $\log Y$.

1.4 Association entre variables catégorielles/nominales et numériques

La corrélation point-bisériale [9] est une mesure statistique utilisée pour évaluer la relation entre une variable catégorielle binaire et une variable numérique continue.

Il s'agit d'un cas particulier du coefficient de corrélation de Pearson et est utilisé lorsqu'une variable est dichotomique et l'autre continue. Ce test peut être décrit par *l'Algorithme 5* ci-dessous :

Entrée: Variable numérique continue X, variable binaire Y
Sortie: Coefficient de corrélation point-bisérial : r_{pb}
Diviser les données numériques X en deux groupes :
X_1 et X_2, en fonction de la variable catégorielle binaire Y.
Calculer la moyenne de la variable continue pour les deux groupes :
$\bar{X}_1$ et $\bar{X}_2$.
Calculer l'écart type de la variable continue $\sigma(X)$.
Calculer la proportion de cas dans chaque groupe : $p = n_1/n$ et $q = n_2/n$
Calculer le coefficient de corrélation point-bisérial à l'aide de la formule :
$r_{pb} = \frac{\bar{X}_1 - \bar{X}_2}{\sigma(X)} \sqrt{\frac{pq}{n}}$.
return r_{pb}.

Algorithme 5 : Calcul de la corrélation point-bisériale.

Le coefficient de corrélation point-bisérial mesure la force et la direction de l'association entre une variable binaire et une variable continue. Un coefficient positif indique que des valeurs plus élevées de la variable continue sont associées à la catégorie binaire codée 1, tandis qu'un coefficient négatif indique une association avec la catégorie codée 0. La signification statistique de cette corrélation peut être évaluée par un T-test, où une p-value inférieure à 0,05 suggère une relation significative entre les variables.

1.5 Conclusions sur les tests statistiques

Ainsi, ces différents tests statistiques, notamment utiles en sélection de variables, présentent des avantages et des désavantages que nous rappelons ci-après.

Avantages : Ces types de tests présentent premièrement l'avantage d'être agnostiques aux modèles, s'appuyant uniquement sur les propriétés statistiques des données.

II. Mesures de l'importance des variables des modèles de ML supervisé

Cette indépendance permet une évaluation objective des relations entre les variables sans être influencée par les spécificités d'un modèle particulier. De plus, les tests statistiques fournissent un cadre robuste pour identifier les variables fortement associées à la variable cible. Cette capacité à repérer les associations importantes en fait une étape préliminaire essentielle dans le processus de modélisation, aidant à orienter les analyses ultérieures et à garantir que les variables pertinentes sont prises en compte.

Désavantages : Leur principal inconvénient est leur nature univariée, ce qui signifie qu'ils analysent les relations entre couple de variables. Cette approche univariée limite leur capacité à saisir les motifs complexes et les interactions entre plusieurs variables prédictives. En conséquence, ils peuvent ne pas capter toute la richesse des structures de données multivariées. Malgré cela, les tests statistiques restent précieux pour optimiser la sélection des variables. En intégrant ces tests dans le processus de modélisation, les analystes peuvent améliorer la performance globale des modèles en s'assurant que les variables sélectionnées apportent une contribution significative aux prédictions finales.

2. Mesures de l'importance des variables des modèles de ML supervisé

Désormais, nous présentons des méthodes de quantification de l'importance des variables des modèles de ML supervisé à utiliser après leur entraînement (post-hoc).

2.1 Importance des variables par permutation

L'importance des variables par permutation, également connue sous le nom de *Permutation Feature Importance* (PFI), représente une mesure cruciale de l'importance des variables prédictives pour un modèle supervisé [10]. Une variable est considérée comme importante si la permutation aléatoire de ses valeurs entraîne une diminution significative des performances du modèle. Plus précisément, les PFI mesurent la variation de l'erreur de prédiction du modèle après avoir permuté les valeurs de la variable prédictive concernée, comme on peut le voir sur *l'Algorithme 6*. Ce calcul permet de déterminer s'il existe une relation entre cette variable et la sortie du modèle [11, 12].

Entrée: Modèle entraîné $\hat{f}$, matrice des variables X, vecteur cible y, métrique ou mesure d'erreur $\mathcal{L}(y, \hat{f})$
Sortie: Vecteur de features importances FI

Estimer l'erreur originale du modèle $e_{orig} = \mathcal{L}(y, \hat{f}(X))$
for *chaque variable* $j \in [\![1, p]\!]$ **do**
  Générer la matrice de features X_{perm} en permutant aléatoirement les valeurs de la colonne de features j dans les données X.
  Estimer l'erreur $e_{perm} = \mathcal{L}(y, \hat{f}(X_{perm}))$
  Calculer la PFI $FI_j = \frac{e_{perm}}{e_{orig}}$
return PFI triés par ordre croissant.

Algorithme 6 : Calcul des PFI.

Les PFI dans les algorithmes d'ensemble comme Random Forest sont basées sur l'évaluation de l'erreur en mesurant la diminution de l'impureté causée par chaque variable lors de la division des arbres, utilisant souvent la mesure d'impureté de Gini [1] [13]. Une visualisation typique des PFI est donnée sur la *Figure 1*, montrant un cas d'usage en classification sur la base de données du Boston Dataset [14]. Ce jeu de données contient des caractéristiques des propriétés à Boston qui sont utilisées pour prédire les prix des biens immobiliers.

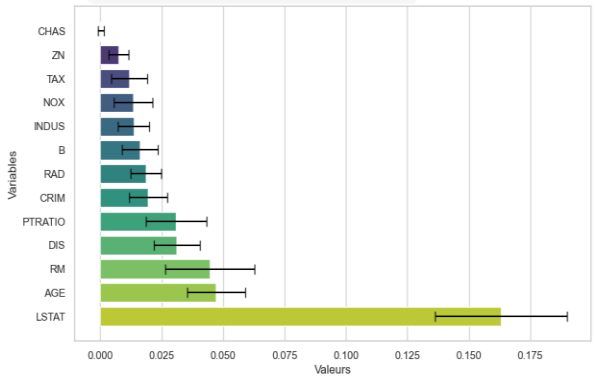

Figure 1 : PFI d'un modèle Random Forest Classifier.

Les barres horizontales noires représentent des barres d'erreur.

L'explication des PFI montre que la variable LSTAT (pourcentage de la population à faible revenu) impacte significativement les performances du modèle, qui prédit si le prix d'un immeuble surpasse un seuil fixe, tandis que la variable CHAS (proximité de la rivière Charles) a peu d'importance, car ses permutations aléatoires n'affectent pas notablement les performances du modèle.

Avantages : Les PFI fournissent une mesure de l'impact des variables sur la performance du modèle lorsqu'elles sont permutées aléatoirement, offrant une interprétation simple de l'importance des variables : une variable est importante si, lorsque l'on modifie aléatoirement ses valeurs, on observe un grand impact sur l'erreur du modèle.

Désavantages : Cette méthode peut être coûteuse en calcul, particulièrement pour les modèles complexes avec de nombreuses variables. Une autre limite de cette méthode est que l'interprétation des résultats peut dépendre de l'architecture du modèle.

La multi-colinéarité des variables prédictives est un autre problème : lorsque les variables sont corrélées, permuter l'une d'elles peut perturber leur relation avec la cible, augmentant artificiellement leur importance. Pour résoudre cela, des techniques comme le regroupement hiérarchique ou l'élimination récursive des variables peuvent être utilisées.

2.2 LIME

La méthode Local Interpretable Model-agnostic Explanations ou LIME [14], est devenue très prisée pour accroître la compréhension des modèles d'IA. Cette approche offre une explication locale des prédictions individuelles générées par les modèles de ML supervisé. Globalement, LIME fonctionne en approximant localement un modèle complexe global f par un modèle interprétable local g en fournissant une observation spécifique X_0, comme illustré dans *l'Algorithme 7* ci-dessous.

Par modèle interprétable, on entend un modèle facilement compréhensible par l'utilisateur à l'aide de représentations visuelles, comme par exemple, une régression linéaire/logistique dans le cas d'un problème de régression/classification. L'explication est obtenue en minimisant les divergences entre la prédiction du modèle interprétable g et celle du modèle complexe f, tout en favorisant la simplicité de g par régularisation.

Le fonctionnement simple de LIME peut s'observer visuellement sur la *Figure 2* : la forme locale de la fonction de décision bleue autour de l'instance verte est utilisée pour expliquer cette instance. L'explication est alors obtenue en générant des échantillons aléatoires autour de cette instance, puis en apprenant un modèle interprétable sur ces données générées.

Entrée: Modèle complexe f, instance à expliquer X_0,
nombre d'échantillons N, distance métrique D, régularisation λ.
Sortie: Explication locale de f pour l'instance X_0.

Générer un ensemble d'échantillons Z autour de l'instance X_0 :
$Z = \{z_1, z_2, \ldots, z_N\}$ où $z_i \sim$ distribution $D(X_0)$.
Calculer les prédictions du modèle complexe f pour chaque échantillon :
$F = \{(z_i, f(z_i)) \mid z_i \in Z\}$.
Calculer les poids des échantillons en fonction de leur proximité à X_0 :
$W = \{w_i \mid w_i = \text{kernel}(D(X_0, z_i)), z_i \in Z\}$ où kernel est une fonction
de pondération (par exemple, une fonction exponentielle).
Apprendre un modèle interprétable g (par exemple, une
régression linéaire/logistique) en utilisant les échantillons pondérés :
$g = \arg\min_g \sum_{i=1}^{N} w_i(f(z_i) - g(z_i))^2 + \lambda\Omega(g)$
où $\Omega(g)$ est un terme de régularisation pour encourager la simplicité de g.
return Paramètres du modèle g comme explication locale de f pour l'instance X_0

Algorithme 7 : Fonctionnement de LIME.

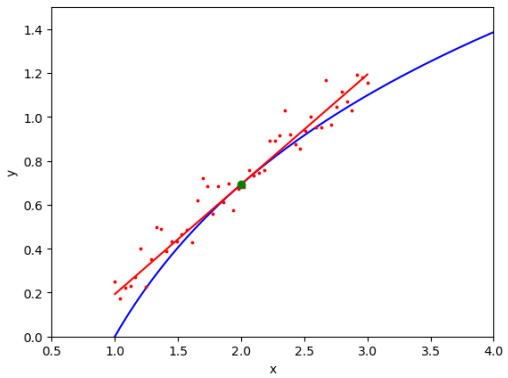

Figure 2 : Principe de fonctionnement de LIME.

Une sortie typique du package LIME est donnée sur la *Figure 3*, montrant un cas d'usage en régression par le modèle Random Forest sur la base de données citée précédemment, pour prédire, dans ce cas, le prix des immeubles.

Selon LIME, les variables LSTAT et RM (nombre moyen de pièces par logement) sont les plus importantes et augmentent la valeur de la prédiction obtenue par ce modèle. En revanche, la variable CRIM (taux de criminalité par habitant) est moins importante et diminue la valeur des prédictions.

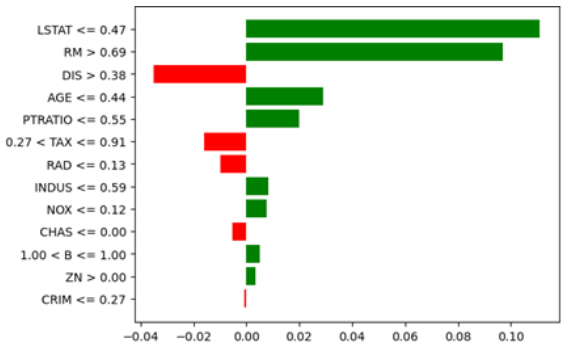

Figure 3 : Explicabilité locale d'une instance d'un modèle de régression Random Forest par LIME.

Avantages : LIME possède l'avantage de reposer sur un principe simple, compréhensible par tous les utilisateurs. De plus, le package LIME est facile à utiliser et demande peu d'efforts pour être exécuté.

Désavantages : Le principal désavantage de la méthode proposée par LIME est lié au fait que, lors de la génération d'un échantillon autour du point d'intérêt, le voisinage utilisé n'est pas correctement défini et doit être choisi de manière empirique à chaque utilisation. De plus, l'échantillonnage autour du point d'intérêt se fait avec une distribution gaussienne qui ignore les corrélations entre les variables prédictives. Il est crucial de résoudre ces problèmes pour assurer la précision de l'explicabilité fournie par LIME.
c

LIME est une méthode prisée pour maîtriser l'explicabilité des modèles de ML mais souffre du manque de résultats théoriques. En revanche, les valeurs de Shapley implémentées dans le package SHAP, bénéficient d'une théorie sous-jacente solide.

2.3 Valeurs de Shapley

Les valeurs de Shapley attribuent une importance à chaque variable prédictive dans un modèle supervisé, en s'appuyant sur la théorie des jeux coopératifs [16]. Elles répartissent équitablement l'importance de chaque variable selon leur rôle, comme illustré à la *Figure 4*.

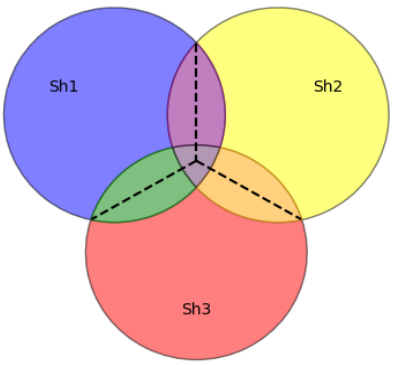

Figure 4 : Répartition égalitaire des valeurs de Shapley entre 3 variables.

Le calcul des valeurs de Shapley est visible sur *l'Algorithme 8* ci-dessous :

Entrée: Ensemble des variables X, Fonction de valeur f
Sortie: Valeurs Shapley ϕ_i pour chaque variable $X_i \in X$

for *chaque variable* $X_i \in X$ **do**
  $\phi_i \leftarrow 0$
  for *chaque sous-ensemble* $S \subseteq X \setminus \{X_i\}$ **do**
    $|S| \leftarrow$ taille de S ;
    $|X| \leftarrow$ taille de X ;
    Poids $\leftarrow \frac{|S|!(|X|-|S|-1)!}{|X|!}$;
    $\phi_i \leftarrow \phi_i +$ Poids $\cdot [f(S \cup \{X_i\}) - f(S)]$.
return ϕ_i pour chaque variable X_i.

Algorithme 8 : Calcul des valeurs de Shapley.

Une valeur de Shapley positive indique que la variable contribue positivement à la prédiction du modèle, tandis qu'une valeur négative indique une influence négative. L'amplitude de ces valeurs quantifie la force de l'effet de chaque variable sur la prédiction, facilitant ainsi la compréhension et la validation des modèles complexes. Les valeurs de Shapley peuvent également être utilisées pour calculer l'importance globale de chaque variable en faisant la moyenne des valeurs absolues des contributions de cette variable sur toutes les prédictions du modèle [10].

Le package SHAP (SHapley Additive exPlanations) [17] applique ce schéma d'explicabilité globale aux variables prédictives des modèles de ML supervisé. Une sortie typique du package SHAP est donné sur la *Figure 5*, sur le même cas d'usage de régression mentionné précédemment.

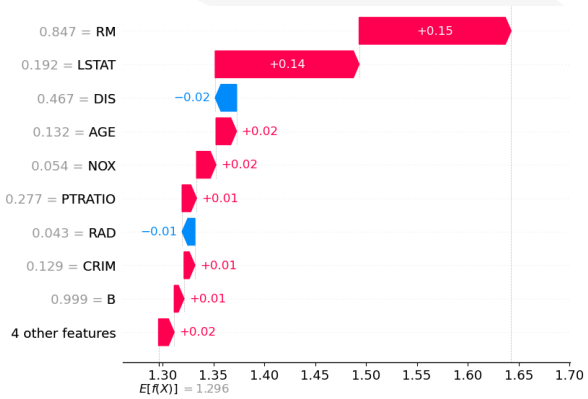

Figure 5 : Explicabilité locale d'une instance du modèle de régression Random Forest par SHAP.

L'explication fournie par SHAP montre que les variables RM et LSTAT améliorent la performance du modèle de régression, tandis que la variable RAD (indice d'accessibilité aux autoroutes) a peu d'importance dans la prise de décision du modèle.

On remarque que les résultats sont cohérents avec ceux obtenus par LIME précédemment sur les quatre variables les plus pertinentes, mais l'ordre d'importance est différent.

Avantages : Le principal avantage de SHAP par rapport aux autres outils d'explicabilité par importance des variables est qu'il s'appuie sur une théorie solide, la théorie des jeux, ce qui le rend plus convaincant et fiable. Cette base théorique assure que les valeurs de Shapley fournissent des explications cohérentes et équitables des contributions des variables prédictives.

Désavantages : Malgré sa base théorique robuste, les valeurs de Shapley présentent certaines limitations. Une variable non présente dans le modèle mais corrélée aux autres peut se voir attribuer de l'importance, un phénomène connu sous le nom de « blague de Shapley » [1]. De plus, le package SHAP fait l'hypothèse d'indépendance des variables prédictives, ce qui peut induire en erreur en présence de variables fortement corrélées. Enfin, les valeurs de Shapley supposent que la contribution d'une variable est indépendante de l'ordre d'ajout des variables au modèle, ce qui n'est pas toujours vrai, surtout pour les modèles avec des interactions non linéaires.

2.4 Package Eli5

Eli5 est un package puissant spécialement conçu pour améliorer la transparence et la quantification de l'importance des variables prédictives des modèles de ML supervisé [18]. Son nom, qui signifie "Explain Like I'm 5" (explique comme si j'avais 5 ans), reflète sa mission de simplifier la compréhension des modèles complexes et de leurs prédictions.

Eli5 excelle dans l'analyse de l'importance des variables prédictives, fournissant aux utilisateurs diverses techniques pour évaluer quelles variables influencent le plus les résultats de leurs modèles. Pour l'analyse de l'importance des variables prédictives, Eli5 propose plusieurs méthodes, comme LIME et les PFI. De plus, Eli5 s'intègre parfaitement aux bibliothèques populaires telles que scikit-learn [18], XGBoost [19] et LightGBM [20], offrant un support natif pour leurs modèles et améliorant la polyvalence de ses explications. Eli5 fournit également des visualisations détaillées, telles que la mise en évidence des contributions des variables prédictives individuelles pour des prédictions spécifiques et l'affichage des vecteurs de poids pour les modèles linéaires.

En conclusion, Eli5 est un package puissant qui permet de mettre en œuvre facilement des méthodes comme PFI ou LIME. Son utilisation peut grandement améliorer la confiance des utilisateurs et des parties prenantes dans les décisions des modèles, tout en facilitant le débogage et la réduction des biais.

Avantages : En comparant Eli5 à d'autres outils d'interprétabilité comme SHAP et LIME, Eli5 se distingue par sa facilité d'utilisation et son intégration robuste. Alors que SHAP offre des insights théoriques plus profonds avec les valeurs de Shapley et LIME fournit des explications de fidélité locale, l'équilibre d'Eli5 entre simplicité, efficacité et large applicabilité en fait un choix solide pour les praticiens cherchant à démystifier leurs modèles de ML.

Désavantages : En revanche, lorsque Eli5 fait des calculs de PFI et LIME, cette méthode souffre des mêmes limitations. En particulier, elle peut induire en erreur l'utilisateur en présence de variables corrélées.

CONCLUSION

Cette étude a exploré les principales méthodes de mesure de l'importance des variables des modèles de ML supervisé. L'analyse de la corrélation joue un rôle crucial dans la sélection des variables prédictives, permettant de déterminer les attributs les plus informatifs et de renforcer l'interprétabilité des modèles.

Nous avons comparé différentes techniques de quantification de l'importance des variables, y compris les méthodes indépendantes du modèle, comme les coefficients de corrélation, et les méthodes dépendantes du modèle, telles que LIME et les valeurs de Shapley. Bien que les méthodes dépendantes du modèle offrent une compréhension plus approfondie des mécanismes internes des algorithmes, elles peuvent être influencées par la multi-colinéarité des variables prédictives, compromettant ainsi la fiabilité des explications. Dans ce contexte, la sélection de variables par l'étude des corrélations reste essentielle pour construire des algorithmes de ML supervisé interprétables.

L'application des packages LIME, SHAP et Eli5 a démontré leur utilité dans l'explicabilité des modèles complexes, fournissant des insights précieux sur la contribution de chaque variable prédictive. Cependant, des défis persistent, notamment en ce qui concerne la gestion des corrélations entre les variables lors de la génération d'échantillons synthétiques [21].

En pratique, ce qui est recommandé est une approche combinée utilisant à la fois les méthodes de mesure de l'importance des variables pour la compréhension des modèles de ML supervisé, mais aussi une analyse des corrélations pour valider ces résultats et réaliser des sélections de variables lorsque cela est nécessaire. Cette stratégie permet de garantir des prédictions fiables et transparentes, répondant ainsi aux besoins des utilisateurs et des décideurs.

RÉFÉRENCES

1. S. Ali, T. Abuhmed, *et al.* Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. *Information Fusion*, 99, 2023.
2. A. Cinar. Overview of existing approaches for the interpretation of machine learning models. *Hochschule Esslingen, 1-19, 2019*.
3. G. Schwalbe, B. Finzel. A comprehensive taxonomy for explainable artificial intelligence : a systematic survey of surveys on methods and concepts. *Data Mining and Knowledge Discovery*, 2023.
4. E. Lopez Fune, V. Messina, A. Edo. Benchmark sur les algorithmes de ML supervisé interprétables. *Production interne R&D*, Nexialog Consulting, 2024.
5. K. Pearson. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, 1958. https://doi.org/10.1080/14786440009463897
6. H. Cramér. Mathematical Methods of Statistics. *Princeton University Press,* 1946. https://books.google.fr/books?id=_db1jwEACAAJ
7. C. Spearman. The Proof and Measurement of Association between Two Things. *The American Journal of Psychology*, 15.1 : 72-101, 1904.
8. J. Cohen. Statistical power analysis for the behavioral sciences. *Hillsdale, NJ : Lawrence Erlbaum Associates*, 1998.
8. G. V Glass, K. D Hopkins. Statistical Methods in Education and Psychology. *Boston : Allyn et Bacon*, 1996.
9. J. Kim. Explainable AI (XAI) Methods Part 4 – Permutation Feature Importance, 2022. https://medium.com/geekculture/explainable-ai-xai-methods-part-4-permutation-featureimportance-72b8a5d9be05
10. C. Molnar. Interpretable Machine Learning A Guide for Making Black Box Model Explainable, 2023. https://christophm.github.io/interpretable-ml-book/
11. F. Dominici, A.Fisher, C. Rudin. All Models are Wrong, but Many are Useful : Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously. *Journal of Machine Learning Research*, 20 : 1–81, 20138.
12. L. Breiman. Random Forests. *Machine Learning*, 45.1 : 5-32, 2001.
13. M.T. Ribeiro, S. Singh, C. Guestrin. "Why Should I Trust You?": Explaining the Predictions of Any Classifier, . e-print arXiv : 1602.04938, 2016.
14. D. Harrison, D.L. Rubinfeld. Hedonic prices and the demand for clean air. *J. Environ. Economics & Management*, 5 : 81-102, 1978.

15. L.S. Shapley. A Value for n-Person Games. *Contributions to the Theory of Games Princeton University Press*, 307-317, 1953.

16. S.L. Scott Lundberg. A Unified Approach to Interpreting Model Predictions, e-print arXiv : 1705.07874 , 2017.

17. Eli5 Documentation. https://eli5.readthedocs.io/en/0.11.0/overview.html

18. F. Pedregosa, *et al*. Scikit-learn : Machine Learning in Python. *Journal of Machine Learning Research,* 12 : 2825-2830, 2011.

19. T. Chen, C. Guestrin. XGBoost : A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '16. ACM*, 2016.

20. G. KE , *et al*. "LightGBM : A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30 : 3146-3154, 2017.

21. Scikit-Learn documentation. https://scikit-learn.org/stable/auto_examples/inspection/plot_permutation_importance_multicollinear.html

NEXIALOG CONSULTING

ACTUARIAT | **GESTION DES RISQUES** | **DATA** | **FINANCE DURABLE**

Nexialog Consulting est un cabinet de conseil spécialisé en Stratégie, Actuariat, Gestion des risques et Data qui dessert aujourd'hui les plus grands acteurs de la banque et de l'assurance. Nous aidons nos clients à améliorer de manière significative et durable leurs performances et à atteindre leurs objectifs les plus importants.

Les besoins de nos clients et les réglementations européennes et mondiales étant en perpétuelle évolution, nous recherchons continuellement de nouvelles et meilleures façons de les servir. Pour ce faire, nous recrutons nos consultants dans les meilleures écoles d'ingénieur et de commerce et nous investissons des ressources de notre entreprise chaque année dans la recherche, l'apprentissage et le renforcement des compétences.

Quel que soit le défi à relever, nous nous attachons à fournir des résultats pratiques et durables et à donner à nos clients les moyens de se développer.

CONTACTS

Retrouvez toutes nos publications sur Nexialog R&D

www.nexialog.com

ALI BEHBAHANI

Associé, Fondateur

+33 (0) 1 44 73 86 78

abehbahani@nexialog.com

ARESKI COUSIN

Directeur Scientifique

+33 (0) 7 88 03 51 87

acousin@nexialog.com

CHRISTELLE BONDOUX

Associée, Directrice Commerciale, Recrutement & Marketing

+33 (0) 1 44 73 75 67

cbondoux@nexialog.com

BAPTISTE BOBEL

Account Manager

+33 (0)6 64 59 12 48

bbobel@nexialog.com