Comparaison des valeurs de Shapley et des valeurs du poids de l'évidence dans le cas du classifieur naïf de Bayes

Vincent Lemaire*, Fabrice Clérot*, Marc Boullé*

* Orange, Lannion, France

Résumé. La sélection de variables et/ou la mesure d'importance des variables en entrée d'un modèle de machine learning est (re)devenue le centre de nombreuses recherches du fait de la réglementation européenne sur la protection de la vie privée. Avoir un bon modèle ne suffit plus il faut aussi expliquer ses décisions. Il existe de ce fait aujourd'hui de nombreux algorithmes d'intelligibilité. Parmi ces derniers on trouve beaucoup , ces derniers temps, d'algorithmes d'estimation des valeurs de Shapley, une méthode d'intelligibilité reposant sur la théorie des jeux coopératifs. Cet article propose une comparaison des valeurs de Shapley dans le cas particulier du classifieur naïf de Bayes avec un autre indicateur fréquemment utilisé "le poids de l'évidence".

1 Introduction

Il existe de nombreux algorithmes d'intelligibilité, souvent empiriques et parfois sans justifications théoriques. C'est là l'une des raisons principales pour lesquelles la bibliothèque Python SHAP a été créée en 2017 par Scott Lundberg à la suite de sa publication (Lundberg et Lee, 2017), pour proposer des algorithmes d'estimation des valeurs de Shapley, une méthode d'intelligibilité reposant sur la théorie des jeux coopératifs. Depuis son lancement, cette bibliothèque connaît un succès grandissant, notamment grâce à de meilleures justifications théoriques et à des visualisations qualitatives.

On propose dans ce document une expression analytique des valeurs de Shapley dans le cas particulier du classifieur naïf de Bayes et une comparaison avec "le poids de l'évidence". Expliquer la prédiction à l'aide des valeurs de l'un de ces indicateurs consiste à attribuer à chaque variable d'entrée, plus précisément à chaque valeur décrivant l'individu considéré un coefficient réel. Chacun de ces coefficients indique comment cette valorisation a contribué à impacter la prévision.

2 Shapley pour le classifieur naïf de Bayes

A notre connaissance il n'existe pas dans la littérature de calcul "analytique" des valeurs de Shapley pour le classifieur naïf de Bayes. Cette première section est donc dédiée à une proposition de calcul de ces valeurs, en exploitant l'hypothèse d'indépendance conditionnelle des variables qui caractérise ce classifieur.

2.1 Rappels sur le classifieur naïf de Bayes

Le classifieur naïf de Bayes (NB) est un outil largement utilisé dans les problèmes de classification supervisée. Il a pour avantage de se montrer efficace pour de nombreux jeux de données réels (Hand et Yu, 2001). Cependant, l'hypothèse naïve d'indépendance des variables peut, dans certains cas, dégrader les performances du classifieur. Aussi, des méthodes proposant de réaliser de la sélection de variables ont vu le jour (Langley et Sage, 1994). Elles consistent en la mise en place d'heuristiques d'ajout et de suppression de variables afin de sélectionner le meilleur sous-ensemble de variables maximisant un critère de performance du classifieur, selon une approche wrapper (Guyon et Elisseeff, 2003). Il a été montré par Boullé (Boullé, 2007) que moyenner un grand nombre de classifieurs Bayésiens naïfs sélectifs, réalisés avec différents sous-ensembles de variables, revenait à ne considérer qu'un seul modèle avec une pondération sur les variables. La formule de Bayes sous l'hypothèse d'indépendance des variables conditionnellement aux classes devient :

$$P(C_k|X) = \frac{P(C_k)\prod_i P(X_i|C_k)^{W_i}}{\sum_{j=1}^{K}(P(C_j)\prod_i P(X_i|C_j)^{W_i})} \quad (1)$$

où W_i représente le poids de la variable i. La classe prédite est celle qui maximise la probabilité conditionnelle $P(C_k|X)$. Les probabilités $P(X_i|C_i)$ peuvent être estimées par intervalle à l'aide d'une discrétisation pour les variables numériques. Pour les variables catégorielles, cette estimation peut se faire directement si la variable prend peu de modalités différentes ou après un groupage dans le cas contraire.

2.2 Définition et notations

On pose les notations suivantes :

* le classifieur utilise d variables : $[d] = \{1, 2, ..., d\}$
* pour un sous ensemble, u, de $[d]$ on note $|u|$ la cardinalité de u
* pour deux ensembles u et r disjoints de $[d]$ on pose $u + r$ comme étant $u \cup r$
* pour un sous-ensemble u de $[d]$, on désigne par $-u = [d]\backslash u$, le complément de u dans d

On définit une 'value function' $v(.)$ indiquant pour chaque sous ensemble de variables, u, la "contribution" maximale qu'elles peuvent obtenir ensemble, c.à.d $v(u)$, à la sortie du classifieur. La valeur maximale (ou gain total) de la 'value function' est quand a elle atteinte lorsqu'on considère toutes les variables, $v([d])$. La valeur de Shapley pour la variable j est notée ϕ_j. Le théorème de Shapley nous dit qu'il existe une unique répartition des valeurs de Shapley satisfaisant les quatre propriétés suivantes :

— Efficacité
  — $v([d]) = \sum_j \phi_j$
  — le gain total est distribué sur l'ensemble des variables
— Symétrie
  — si $\forall u \subset [d] - \{i, j\}$, $v(u + j) = v(u + i)$, alors $\phi_j = \phi_i$
  — si les variables i et j apportent le même gain à tout sous-ensemble de variables, alors elles ont la même valeur de Shapley
— Joueur nul
  — si $\forall u \subset [d] - \{i\}$, $v(u + j) = v(u)$, alors $\phi_j = 0$

— si la variable i n'apporte rien à n'importe quel sous-ensemble de variables, alors sa valeur de Shapley est nulle

— Additivité
  — si les d variables sont utilisées pour deux problèmes de classification indépendants A et B associés à v_A, v_B, alors les valeurs de Shapley pour l'ensemble des deux problèmes sont la somme des valeurs de Shapley pour chaque problème

2.3 Valeurs de Shapley pour le classifieur naïf de Bayes

2.3.1 'Value Function'

Dans le cas du NB on propose de prendre comme 'Value Function' (cas d'un problème de classification à deux classes) le log ratio des probabilités :

$$LR = log\left(\frac{P(C_1|X)}{P(C_0|X)}\right) \tag{2}$$

$$= log\left(\frac{P(C_1)\prod_{i=1}^{d} P(X_i|C_1)^{W_i}}{\sum_{j=1}^{K}(P(C_j)\prod_{i=1}^{d} P(X_i|C_j)^{W_i})}\frac{\sum_{j=1}^{K}(P(C_j)\prod_{i=1}^{d} P(X_i|C_j)^{W_i})}{P(C_0)\prod_{i=1}^{d} P(X_i|C_1)^{W_i}}\right) \tag{3}$$

$$= log\left(\frac{P(C_1)\prod_{i=1}^{d} P(X_i|C_1)^{W_i}}{P(C_0)\prod_{i=1}^{d} P(X_i|C_1)^{W_i}}\right) \tag{4}$$

$$= log\left(\frac{P(C_1)}{P(C_0)}\right) + \sum_{i=1}^{d} W_i log\left(\frac{P(X_i|C_1)}{P(X_i|C_0)}\right) \tag{5}$$

Ce choix du log odd ratio comme 'value function' est motivé par deux raisons (i) le log odd ratio est en bijection avec le score produit par le classifieur selon une transformation monotone (ii) le log odd ratio a une forme linéaire qui simplifie les calculs.

Pour un sous ensemble, u, de variables [1] :

$$v(u) = \mathbb{E}_{X_{-u}|X_u=x_u}\left[LR(X_u = x_u^*, X_{-u})\right] \tag{6}$$

qu'on écrira de manière "simplifiée" par la suite

$$v(u) = \mathbb{E}\left[(LR(X)|X_u = x_u^*)\right] \tag{7}$$

Il s'agit d'un proxy de l'information sur la cible apportée par u au point $X = x^*$. Il va de soi que cette formulation ne vaut que dans le cas de variables indépendantes conditionnellement à la cible ce qui, heureusement, est le cas supposé dans la formule du classifieur naïf de Bayes.

On a donc pour un point (exemple) d'intérêt $x*$

— $v([d]) = LR(X = x^*)$, tout est conditionné à x^* donc on a le log odd ratio pour $X = x^*$

— $v(\emptyset) = \mathbb{E}_X\left[LR(X)\right] = \mathbb{E}_X\left[log(\frac{P(C_1|X)}{P(C_0|X)})\right]$, rien n'est conditionné donc a l'espérance du log odd ratio

1. sur les covariables en u, nous faisons une moyenne sur la distribution conditionnelle de X_{-u} étant donné $X_u = x_u$

2.3.2 Valeurs de Shapley

Par définition des valeurs de Shapley (Shapley et Shubik, 1954) on a pour une variable m :

$$\phi_m = \frac{1}{d} \sum_{u \in -m} \frac{v(u+m) - v(u)}{\binom{d-1}{|u|}}, \tag{8}$$

Pour obtenir ϕ_m il faut donc calculer, pour un sous ensemble de variables dans lequel la variable m n'apparaît pas, la différence de gain $v(u+m) - v(u)$. Cela permet de comparer le gain obtenu par le sous ensemble de variables avec et sans la variable m, afin de mesurer son impact lorsqu'elle "collabore" avec les autres. On doit donc calculer $v(u+m) - v(u)$ dans le cas du classifieur naïf de Bayes. Si cette différence est positive, cela signifie que la variable contribue positivement. A l'inverse, si la différence est négative, cela signifie que la variable pénalise le gain. Enfin, si la différence est nulle, cela indique que la variable n'apporte rien.

En suivant l'exemple de Lundberg et Lee (2017) et le Corollary1 avec un modèle linéaire dont les covariables sont le log odd ratio comme 'value function' on peut décomposer les sous ensembles de variables en 3 groupes $\{u\}, \{m\}, -\{u+m\}$.

Calcul de $v(u)$: Sur $\{u\}$, nous conditionnons sur $X_u = x_u$ tandis que sur $\{m\}$, $\{u+m\}$, on fait un moyennage Par conséquent on a

$$\begin{aligned} v(u) &= \mathbb{E}\left[LR(X)|X_u = x_u^*)\right] \\ &= log(P(Y_1)/P(Y_0)) \\ &+ \sum_{k(k\in u)} w_k log\left(\frac{P(X_k = x_k{}^*|Y_1)}{P(X_k = x_k{}^*|Y_0)}\right) \\ &+ w_m \mathbb{E}_{X_m}\left[P(X_m = x_m) log\left(\frac{P(X_m = x_m|Y_1)}{P(X_m = x_m|Y_0)}\right)\right] \\ &+ \sum_{k(k\in -\{u+m\})} w_k \mathbb{E}_{X_k}\left[P(X_k = x_k) log\left(\frac{P(X_k = x_k|Y_1)}{P(X_k = x_k|Y_0)})\right)\right] \end{aligned} \tag{9}$$

(10)

Calcul de $v(u+m)$: La seule différence c'est que l'on conditionne aussi sur X_m

$$\begin{aligned} v(u+m) &= \mathbb{E}\left[LR(X)|X_{u+m} = x_{u+m}^*)\right] \\ &= log(P(Y_1)/P(Y_0)) \\ &+ \sum_{k(k\in u)} w_k log\left(\frac{P(X_k = x_k{}^*|Y_1)}{P(X_k = x_k{}^*|Y_0)}\right) \\ &+ w_m \left[log\left(\frac{P(X_m = x_m^*|Y_1)}{P(X_m = x_m^*|Y_0)}\right)\right] \\ &+ \sum_{k(k\in -\{u+m\})} w_k \mathbb{E}_{X_k}\left[P(X_k = x_k) log\left(\frac{P(X_k = x_k{}^*|Y_1)}{P(X_k = x_k{}^*|Y_0)})\right)\right] \end{aligned} \tag{11}$$

(12)

On a donc $v(u+m)-v(u)$:

$$v(u+m)-v(u) = w_m \left(log\left(\frac{P(X_m = x_m^*|Y_1)}{P(X_m = x_m^*|Y_0)}\right) - \mathbb{E}_{X_m}\left[P(X_m = x_m) log\left(\frac{P(X_m = x_m|Y_1)}{P(X_m = x_m|Y_0)}\right)\right]\right) \quad (13)$$

ce qui correspond à la différence entre le contenu de l'information de X_m conditionnée à $X_m = x_m^*$ et l'entropie conditionnelle de la variable X_m. Autrement dit, comme l'entropie est une espérance, il s'agit de l'apport en information de la variable X_m pour la valeur $X_m = x_m^*$ de l'instance considérée, contrasté par l'apport moyen sur l'ensemble de la base. On voit aussi que $v(u+m)-(u)$ (dans le cas du NB) ne dépend que de m ; la sommation (voir Equation 8) est donc égale à l'équation 13. Equation qui peut être réécrit (on omet juste le produit par w_m) sous la forme [2] :

$$\begin{aligned} &- \left[log\left(\frac{1}{P(X_m = x_m^*|Y_1)}\right) - \mathbb{E}_{X_m}\left[P(X_m = x_m) log\left(\frac{1}{P(X_m = x_m|Y_1)}\right)\right]\right] \\ &+ \left[log\left(\frac{1}{P(X_m = x_m^*|Y_0)}\right) - \mathbb{E}_{X_m}\left[P(X_m = x_m) log\left(\frac{1}{P(X_m = x_m|Y_0)}\right)\right]\right] \end{aligned} \quad (14)$$

Résultat : L'équation 13 fournit donc l'expression analytique de la valeur de Shapley. On s'aperçoit que dans le cas du classifieur naïf de Bayes on obtient une formulation exacte (non approchée) qui est, de plus, peu coûteuse en temps de calcul dans le cas où les variables numériques (réciproquement catégorielles) ont été préalablement discrétisées (groupage de modalités).

3 Comparaison avec le "poids de l'évidence"

Dans le cas du classifieur naïf de Bayes ainsi que celui de la régression logistique (Hosmer et Lemeshow, 2000) il existe un certain nombre de méthodes "usuelles" de calcul d'importance des variables. Le lecteur pourra trouver dans (Robnik-Sikonja et Kononenko, 2008) un large panel de ces indicateurs. Cette section présente le "weight of evidence" et une comparaison avec les valeurs de Shapley proposées dans la section précédente.

Le "Weight of evidence" (WoE) (Good, 1950) est parmi les indicateurs les plus utilisés. On utilisera cet indicateur aussi à titre de comparaison car comme nous le verrons plus tard, cet indicateur est proche de l'équation présentée ci-dessus (équation 13). La différence principale

2. En effet : L'équation 13 est de la forme

$$\begin{aligned} log\left(\frac{A}{B}\right) - \mathbb{E}\left[P(X) log\left(\frac{A}{B}\right)\right] = \\ log(A) + log\left(\frac{1}{B}\right) - \mathbb{E}\left[P(X)\left(log(A) + log\left(\frac{1}{B}\right)\right)\right] = \\ log(A) + log\left(\frac{1}{B}\right) - \mathbb{E}\left[P(X) log(A)\right] - \mathbb{E}\left[P(X) log\left(\frac{1}{B}\right)\right] = \\ -log\left(\frac{1}{A}\right) + log\left(\frac{1}{B}\right) + \mathbb{E}\left[P(X) log\left(\frac{1}{A}\right)\right] - \mathbb{E}\left[P(X) log\left(\frac{1}{B}\right)\right] = \\ - \left[log\left(\frac{1}{A}\right) - \mathbb{E}\left[P(X) log\left(\frac{1}{A}\right)\right]\right] + \left[log\left(\frac{1}{B}\right) - \mathbb{E}\left[P(X) log\left(\frac{1}{B}\right)\right]\right] \end{aligned}$$

réside dans la valeur de référence. Dans l'équation 13 on peut s'apercevoir que le deuxième terme utilise une référence vis-à-vis de l'ensemble de la population alors que dans le WoE utilise une référence en zéro. On donne ci-dessous la définition du WoE (dans le cas à deux classes) qui est un log odds ratio calculé entre la probabilité de la sortie du modèle et cette dernière privée de la variable X_m :

$$(WoE)_m = log\left(\frac{\frac{p}{1-p}}{\frac{q}{1-q}}\right) = log\left(\frac{\frac{P(Y_1|X)}{P(Y_0|X)}}{\frac{P(Y_1|X\setminus X_m)}{P(Y_O|X\setminus X_m)}}\right) = log\left(\frac{P(Y_1|X)}{P(Y_0|X)}\right) - log\left(\frac{P(Y_1|X\setminus X_m)}{P(Y_0|X\setminus X_m)}\right) \quad (15)$$

$$(WoE)_m = w_m\left(log\left(\frac{\frac{P(Y_1|X)}{P(Y_0|X)}}{\frac{P(Y_1|X\setminus X_m)}{P(Y_0|X\setminus X_m)}}\right)\right) = w_m\left(log\left(\frac{P(Y_1|X)P(Y_0|X\setminus X_m)}{P(Y_0|X)(Y_1|X\setminus X_m)}\right)\right) \quad (16)$$

$$(WoE)_m = w_m\left(log\left(\frac{P(Y_1)\left[\prod_{i=1}^{d} P(X_i|Y_1)\right] P(Y_0)\left[\prod_{i=1,i\neq m}^{d} P(X_i|Y_0)\right]}{P(Y_0)\left[\prod_{i=1}^{d} P(X_i|Y_0)\right] P(Y_1)\left[\prod_{i=1,i\neq m}^{d} P(X_i|Y_1)\right]}\right)\right) \quad (17)$$

par simplification du numérateur et du dénominateur :

$$(WoE)_m = w_m\left(log\left(\frac{P(X_m = x_m^*|Y_1)}{P(X_m = x_m^*|Y_0)}\right)\right) \quad (18)$$

On notera que "priver" le classifieur naïf de Bayes d'une variable est équivalent à mener un calcul de "saliency" tel que proposé dans (Lemaire et Clérot, 2006) et qui prend en compte la distribution de probabilité de la variable X_m. En effet il suffit pour priver le classifieur (dans le cas du naïf de Bayes) de la variable X_m de recalculer la moyenne des prédictions du classifieur pour toutes les valeurs possible de la variables X_m tel que démontré dans (Robnik-Sikonja et Kononenko, 2008).

Lien en $(WoE)_m$ et ϕ_m : on peut donc voir que c'est la référence qui change. Dans l'équation 13 on peut s'apercevoir que le deuxième terme pose une référence vis-à-vis de l'ensemble de la population alors que le WoE pose une référence en zéro.

En effet si on pose que la variable X_m possède k valeurs distinctes Robnik et al. (Robnik-Sikonja et Kononenko, 2008) on montré que le calcul de saliency de (Lemaire et Clérot, 2006) est exact dans le cas du naïf Bayes et revient bien à "effacer" la variable X_m :

$$P(Y_.|X\setminus X_m) = \sum_{q=1}^{k} P(X_m = X_q)\frac{P(Y_.|X, X_m = X_q)}{P(X, X_m = X_q)} \quad (19)$$

$$P(Y_.|X\setminus X_m) = \sum_{q=1}^{k} P(X_m = X_q)\left(P(Y_.)\left(\prod_{i=1,i\neq m}^{d}\frac{P(X_i|Y_.)}{P(X_i)}\right)\frac{P(X_m = X_q|Y_.)}{P(X_m = X_q)}\right)$$

$$P(Y_.|X\setminus X_m) = P(Y_.)\prod_{i=1,i\neq m}^{d} P(X_i|Y_.)\left(\sum_{q=1}^{k}\frac{P(X_m = X_q)P(X_m = X_q|Y_.)}{P(X_m = X_q)}\right) \quad (20)$$

$$P(Y_.|X\setminus X_m) = P(Y_.)\prod_{i=1,i\neq m}^{d} P(X_i|Y_.) \quad (21)$$

avec $P(Y|X, X_m = X_q)$ étant $P(Y|X)$ mais où la valeur de la variable X_m a été remplacée par une autre valeur de sa distribution X_q.

Ce dernier résultat est intéressant car à l'aide de l'équation 25 on peut réécrire l'équation 21 en :

$$(WoE)_m = w_m \left(log \left(\frac{P(Y_1|X)P(Y_0|X\backslash X_m)}{P(Y_0|X)(Y_1|X\backslash X_m)} \right) \right)$$

$$(WoE)_m = w_m log \frac{\left(P(Y_1) \prod_{i=1}^{d} P(X_i|Y_1) \right) \left(P(Y_0) \prod_{i=1, i\neq m}^{d} P(X_i|Y_0) \sum_{q=1}^{k} P(X_m = X_q|Y_0) \right)}{\left(P(Y_0) \prod_{i=1}^{d} P(X_i|Y_0) \right) \left(P(Y_1) \prod_{i=1, i\neq m}^{d} P(X_i|Y_1) \sum_{q=1}^{k} P(X_m = X_q|Y_1) \right)}$$

$$(WoE)_m = w_m \left(log \left(\frac{P(X_m = x_m^*|Y_1)}{P(X_m = x_m^*|Y_0)} \frac{\sum_{q=1}^{k} P(X_m = X_q|Y_0)}{\sum_{q=1}^{k} P(X_m = X_q|Y_1)} \right) \right)$$

$$(WoE)_m = w_m \left(log \left(\frac{P(X_m = x_m^*|Y_1)}{P(X_m = x_m^*|Y_0)} \right) + log \left(\frac{\sum_{q=1}^{k} P(X_m = X_q|Y_0)}{\sum_{q=1}^{k} P(X_m = X_q|Y_1)} \right) \right) \quad (22)$$

$$(WoE)_m = w_m \left(log \left(\frac{P(X_m = x_m^*|Y_1)}{P(X_m = x_m^*|Y_0)} \right) + log \left(\frac{1}{1} \right) \right) \quad (23)$$

Ce résultat permet peut être de mieux comprendre pourquoi le WoE est référencé en zéro. De plus en comparant l'équation 28 et l'équation 13 on voit bien l'effet de prendre comme 'value function' comme étant la sortie du classifieur dans le premier cas (WoE) alors que c'est l'odds dans le deuxième cas (Shapley).

On voit ci-dessus que la différence entre ϕ_m (equation 13) et WoE provient du fait que dans le premier cas on calcule une espérance sur la variation du log ratio $(P(Y_1|X)/P(Y_0|X)$ alors que dans le deuxième cas cette espérance n'est calculée que sur les variations de $f(X) = P(Y_1|X)$ (ou réciproquement $P(Y_0|X)$).

4 Conclusion

Nous avons proposé dans cet article une méthode de calcul analytique des valeurs de Shapley dans le cas du classifier naïf de Bayes. Cette méthode exploite l'hypothèse d'indépendance des variables conditionnellement à la cible pour obtenir la valeur exacte des valeurs de Shapley avec une complexité algorithmique linéaire avec le nombre des variables. Contrairement aux méthodes d'évaluation/approximation alternatives, on exploite des hypothèses parfaitement conformes au classifieur sous-jacent et on évite les méthodes d'approximation particulièrement coûteuses en temps de calcul.

Références

Boullé, M. (2007). Compression-based averaging of selective naive Bayes classifiers. *Journal of Machine Learning Research 8*, 1659–1685.

Good, I. J. (1950). *Probability and the weighing of evidence*. C. Griffin & Company Limited.

Guyon, I. et A. Elisseeff (2003). An introduction to variable and feature selection. *J. Mach. Learn. Res. 3*, 1157–1182.

Hand, D. J. et K. Yu (2001). Idiot's bayes-not so stupid after all ? *International Statistical Review 69*(3), 385–398.

Hosmer, D. W. et S. Lemeshow (2000). *Applied logistic regression*. John Wiley and Sons.

Langley, P. et S. Sage (1994). Induction of selective bayesian classifiers. In *Proceedings of the Tenth International Conference on Uncertainty in Artificial Intelligence*, San Francisco, CA, USA, pp. 399–406. Morgan Kaufmann Publishers Inc.

Lemaire, V. et F. Clérot (2006). An input variable importance definition based on empirical data probability distribution. In *Feature extraction, foundations and applications*, pp. 509–516. Springer Berlin Heidelberg.

Lundberg, S. M. et S.-I. Lee (2017). A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, et R. Garnett (Eds.), *Advances in Neural Information Processing Systems*, Volume 30. Curran Associates, Inc.

Robnik-Sikonja, M. et I. Kononenko (2008). Explaining classifications for individual instances. *Knowledge and Data Engineering, IEEE Transactions on Knowledge and Data Engineering 20*, 589 – 600.

Shapley, L. S. et M. Shubik (1954). A method for evaluating the distribution of power in a committee system. *American Political Science Review 48*(3), 787–792.

Summary

Variable selection and/or importance measurement of input variables to a machine learning model has become the focus of much research due to the European regulation on privacy protection. Having a good model is no longer enough, one must also explain its decisions. Therefore, there are many intelligibility algorithms available today. Among these, there are many algorithms for estimating Shapley values, a method of intelligibility based on the theory of cooperative games. This article proposes a comparison of Shapley values (in the particular case of the Naive Bayes Classifier) with a frequently used indicator: the weight of evidence.