Projet XAI

Machine learning interprétable

L'année 2018 a été marquée par la mise en place du RGPD en Europe qui souligne qu'à l'heure actuelle trouver un **compromis** entre **précision** et **explicabilité** des algorithmes de machine learning devient de plus en plus nécessaire.

Le développement et l'engouement rapides autour de l'IA ont conduit à prioriser la performance des algorithmes souvent qualifiés de "**boîtes noires**", alors qu'à présent le respect de **normes**, de l'**éthique** et de **transparence** notamment, dessinent une autre dynamique dans laquelle l'explicabilité pourrait devenir le nouveau critère d'évaluation des modèles.

Dans cette perspective, nous avons développé une tool box d'interprétabilité Python générant un report complet.

Résultat de machine learning

Développement
TOOL BOX

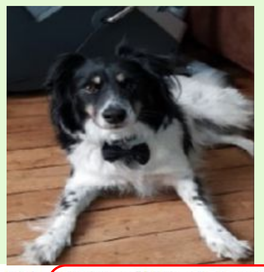
BLACK BOX

Analyse

Pourquoi

Machine learning **interprétable**

Une image qui a du chien



74% Border collie

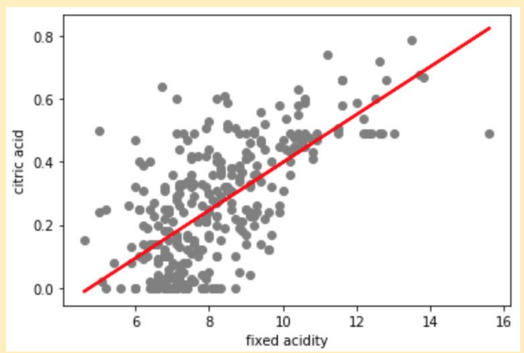
10% Collie

0.6% English setter

```
[('n02106166', 'Border_collie', 0.7452205),
 ('n02106030', 'collie', 0.10160939),
 ('n02100735', 'English_setter', 0.0065846588),
 ('n02086910', 'papillon', 0.0036008318),
 ('n02101388', 'Brittany_spaniel', 0.0034225571)]
```

Classification avec inception V3

Données saveur Vin Rouge



Régression linéaire simple

Pourquoi ces prédictions ?

TOOLBOX

PDP

SHAP

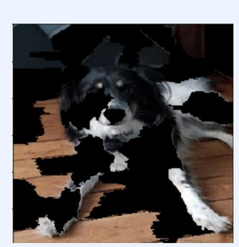
ICE

Shapley values

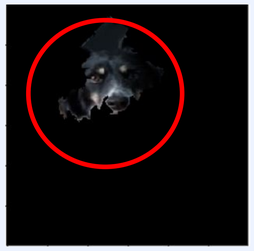
Lime

Permutation features

Résultats



Exemple de perturbations apportées à l'image



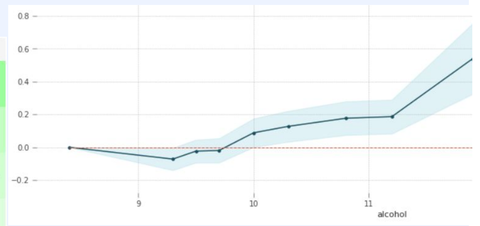
Zone importantes pour la prédiction

Conclusion: désigné comme un border collie à cause de sa tête

+	Weight	Feature
	0.1659 ± 0.0168	alcohol
	0.0811 ± 0.0166	sulphates
	0.0533 ± 0.0088	total sulfur dioxide
	0.0343 ± 0.0087	chlorides
	0.0313 ± 0.0082	volatile acidity
	0.0216 ± 0.0036	free sulfur dioxide
-	0.0181 ± 0.0041	density

Feature permutation

Importance pour la prédiction de "quality"



PDP pour le feature "alcohol"

Le niveau d'alcool a une influence positive sur la prédiction de la "quality" pour des valeurs entre 10 et 12

Morgane Besnier, Marjolaine Claret, Yovana Dentika, Uthamy Thanabalasingam - E3T - Giovanni Chierchia