

Projet Machine Learning

Python – Streamlit

Guide utilisateur

Fatima-Zahra ZABAKA

Darcy NGUYEN

Victor LEBRETON

26/11/2024

Présentation de l'application

Vous retrouverez dans cette documentation les principaux écrans et les principales fonctionnalités de l'application Streamlit Machine Learning.

Elle vous permettra d'uploader un fichier de données. De transformer ces données et de les visualiser. Dans un second temps vous pourrez entraîner vos données dans notre modèle de machine learning afin de « prévoir » la valeur de certains champs.

Nous allons vous expliquer comment réaliser ceci à travers un guide utilisateur complet.

#

#

#

#

Page d'accueil

[Traitement des données](#) [Visualisations](#) [Modélisation](#) [Evaluation](#)

Traitement de données

Données sources

	alcohol	malic_acid	ash	alkalinity_of_ash	magnesium	total_phenols	flavanoids	nonflavanoid_phenols	proanthocyanins	color_intensity	hue	od280/od315_of_diluted_wines	proline	target
0	14.23	1.71	2.43	15.6	127	2.8	3.06	0.28	2.29	5.64	1.04	3.92	1,065	Vin amer
1	13.2	1.78	2.14	11.2	100	2.65	2.76	0.26	1.28	4.38	1.05	3.4	1,050	Vin amer
2	13.16	2.36	2.67	18.6	101	2.8	3.24	0.3	2.81	5.68	1.03	3.17	1,185	Vin amer
3	14.37	1.95	2.5	16.8	113	3.85	3.49	0.24	2.18	7.8	0.86	3.45	1,480	Vin amer
4	13.24	2.59	2.87	21	118	2.8	2.69	0.39	1.82	4.32	1.04	2.93	735	Vin amer
5	14.2	1.76	2.45	15.2	112	3.27	3.39	0.34	1.97	6.75	1.05	2.85	1,450	Vin amer
6	14.39	1.87	2.45	14.6	96	2.5	2.52	0.3	1.98	5.25	1.02	3.58	1,290	Vin amer
7	14.06	2.15	2.61	17.6	121	2.6	2.51	0.31	1.25	5.05	1.06	3.58	1,295	Vin amer
8	14.83	1.64	2.17	14	97	2.8	2.98	0.29	1.98	5.2	1.08	2.85	1,045	Vin amer
9	13.86	1.35	2.27	16	98	2.98	3.15	0.22	1.85	7.22	1.01	3.55	1,045	Vin amer

Modification des données

	↗ alcohol	↗ malic_acid	↗ ash	↗ alkalinity_of_ash	↗ magnesium	↗ total_phenols	↗ flavanoids	↗ nonflavanoid_phenols	↗ proanthocyanins	↗ color_intensity	↗ hue	↗ od280/od315_of_diluted_wines	↗ proline	↗ target
0	14.23	1.71	2.43	15.6	127	2.8	3.06	0.28	2.29	5.64	1.04	3.92	1,065	Vin amer
1	13.2	1.78	2.14	11.2	100	2.65	2.76	0.26	1.28	4.38	1.05	3.4	1,050	Vin amer

Voici l'écran d'accueil sur lequel vous retrouverez vos données sources.

Vous avez la possibilité de modifier vos données si besoin dans le tableau « **Modification des données** ». Même principe de modification que sur Excel. Vous pourrez y remplir des données manquantes ou modifier des données erronées.

La modification des données entraine une mise à jour des dataframes, et les modifications sont conservées tout au long de la navigation sur l'application.

Vous pourrez également supprimer une ou plusieurs colonnes selon vos besoins d'analyse.

Supprimer une ou plusieurs colonnes

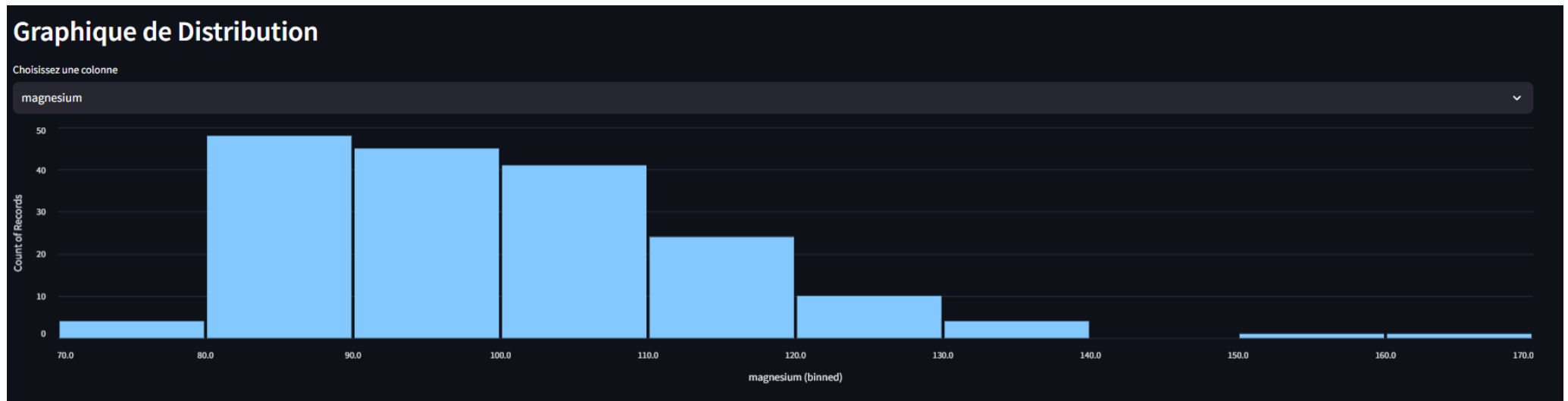
Sélectionner une colonne à supprimer

Choose an option

Supprimer

Supprimer une colonne entraîne une mise à jour des données. Les Dataframes présents sur la page seront mis à jour excepté celui qui contient les sources originales.

Vous y retrouverez également un graphique de distribution de chaque variable, groupée sur un intervalle différent en fonction de chaque champ.



Ici la distribution de la variable « Magnesium » sur l'échantillon de données.

Analyse descriptive du Dataframe sur vos données sources. Permet de connaître différentes métriques.

Analyse descriptive du dataframe													
	alcohol	malic_acid	ash	alcalinity_of_ash	magnesium	total_phenols	flavanoids	nonflavanoid_phenols	proanthocyanins	color_intensity	hue	od280/od315_of_diluted_wines	proline
count	178	178	178	178	178	178	178	178	178	178	178	178	178
mean	13.0006	2.3363	2.3665	19.4949	99.7416	2.2951	2.0293	0.3619	1.5909	5.0581	0.9574	2.6117	746.8933
std	0.8118	1.1171	0.2743	3.3396	14.2825	0.6259	0.9989	0.1245	0.5724	2.3183	0.2286	0.71	314.9075
min	11.03	0.74	1.36	10.6	70	0.98	0.34	0.13	0.41	1.28	0.48	1.27	278
25%	12.3625	1.6025	2.21	17.2	88	1.7425	1.205	0.27	1.25	3.22	0.7825	1.9375	500.5
50%	13.05	1.865	2.36	19.5	98	2.355	2.135	0.34	1.555	4.69	0.965	2.78	673.5
75%	13.6775	3.0825	2.5575	21.5	107	2.8	2.875	0.4375	1.95	6.2	1.12	3.17	985
max	14.83	5.8	3.23	30	162	3.88	5.08	0.66	3.58	13	1.71	4	1,680

Page Visualisation

Dans cette page vous retrouverez les colonnes de vos données sous forme de SELECT avec deux types de graphiques différents. En fonction de vos besoins en termes d'analyse de données vous pourrez sélectionner des colonnes différentes.

Visualisation

Graphique de distribution

Choisissez un model de graphe

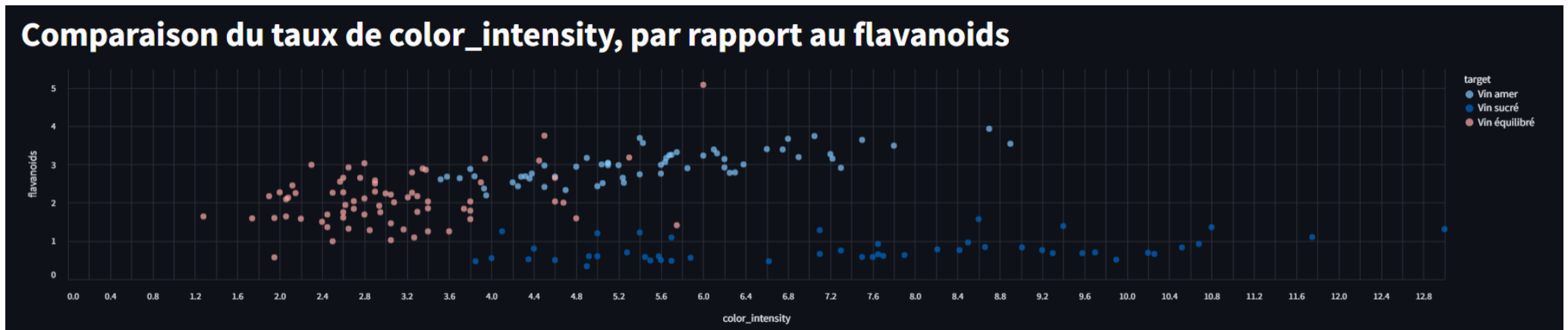
Area Chart with Gradient

Choisissez la première colonne

color_intensity

Choisissez la deuxième colonne

flavanoids

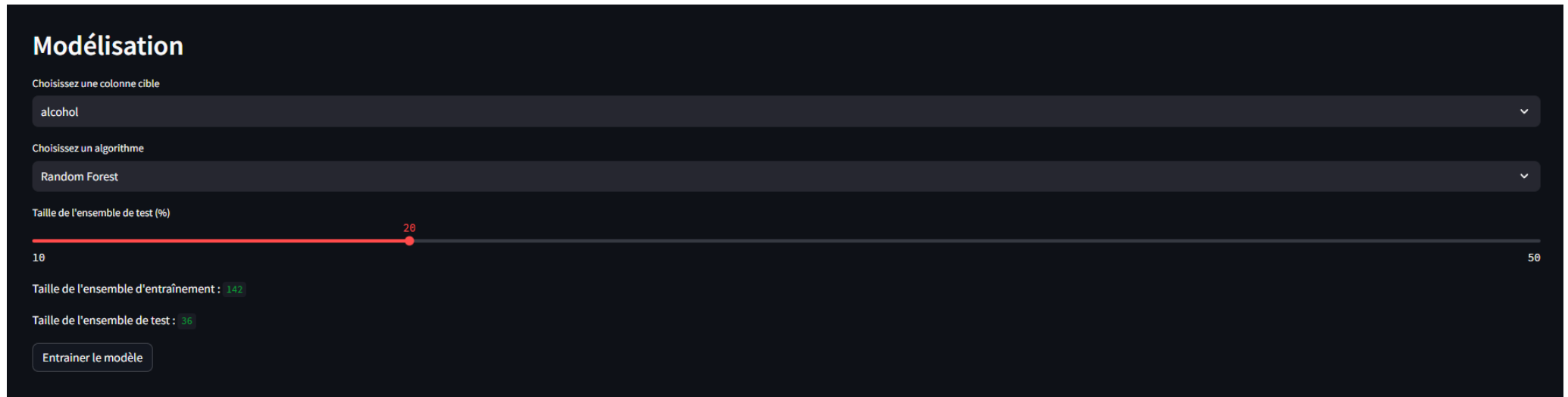


Ici, un graph « Area Chart With Gradient » qui représente le taux de l'intensité de couleur par rapport au flavanoids. Les colonnes sont modifiables et permutable.

Attention cependant, dans ce graphique vous ne pouvez pas sélectionner des colonnes de types Object ou String car elles ne sont pas numériques et ne permettent pas la visualisation des valeurs.

Page Modélisation

Partie importante de notre application, elle vous permettra d'entraîner vos données dans un modèle de machine learning. Vous aurez la possibilité de choisir un algorithme différent en fonction de vos attentes.



Modélisation

Choisissez une colonne cible

alcohol

Choisissez un algorithme

Random Forest

Taille de l'ensemble de test (%)

10 20 50

Taille de l'ensemble d'entraînement : 142

Taille de l'ensemble de test : 36

Entraîner le modèle

Dans un premier temps vous allez sélectionner une colonne dont vous souhaitez prédire les données. En fonction du type de champs contenus dans cette colonne vous allez adapter l'algorithme qui entraîne le modèle.

Random Forest ou Linear Regression.

Vous aurez la possibilité de choisir la taille de vos échantillons d'entraînement et de test. (Plus vous avez de données d'entraînement, plus les résultats seront « précis »).

Une fois que vous avez paramétré le modèle vous pouvez l'entraîner et ainsi voir les « prédictions ».

S'affiche alors un nouveau dataframe contenant les résultats.

Vous pouvez alors vous rediriger vers l'onglet Evaluation qui vous permettra de connaitre les métriques de réussite de notre modèle.

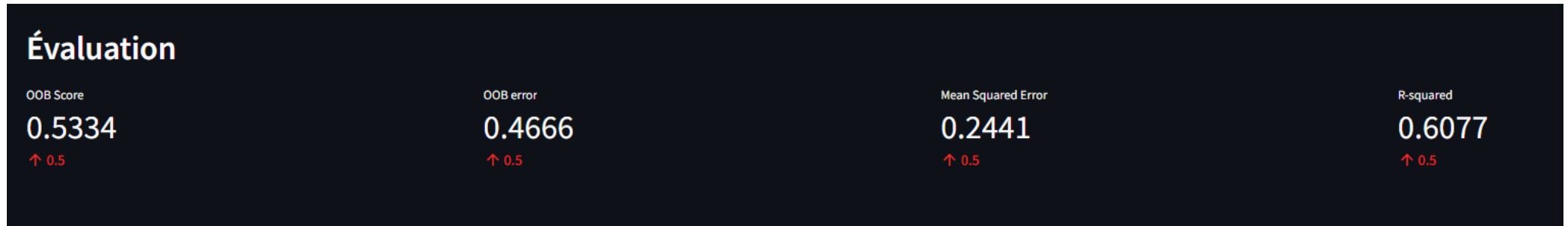
Modèle entraîné avec succès

	malic_acid	ash	alcalinity_of_ash	magnesium	total_phenols	flavanoids	nonflavanoid_phenols	proanthocyanins	color_intensity	hue	od280/od315_of_diluted_wines	proline	taux	target_Vin amer	target_Vin sucré	target_Vin équilibré	Prévision_alcohol	target
59	0.94	1.36	10.6	88	1.98	0.57	0.28	0.42	1.95	1.05	1.82	520	342.1053	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	12.3519	12.37
130	1.35	2.32	18	122	1.51	1.25	0.21	0.94	4.1	0.76	1.29	630	328	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	12.8952	12.86
17	1.57	2.62	20	115	2.95	3.4	0.4	1.72	6.6	1.13	2.57	1,130	194.1176	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	13.773	13.83
115	1.51	2.2	21.5	85	2.46	2.17	0.52	2.01	1.9	1.71	2.87	407	87.5576	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	12.2997	11.03
97	1.41	1.98	16	85	2.55	2.5	0.29	1.77	2.9	1.23	2.74	428	116	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	12.2685	12.29
24	1.81	2.61	20	96	2.53	2.61	0.28	1.66	3.52	1.12	3.82	845	134.8659	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	13.3129	13.5
67	1.17	1.92	19.6	78	2.11	2	0.27	1.04	4.68	1.12	3.48	510	234	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	12.953	12.37
86	1.61	2.31	22.8	90	1.78	1.69	0.43	1.56	2.45	1.33	2.26	495	144.9704	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	12.2144	12.16
74	1.09	2.3	21	101	3.38	2.14	0.13	1.65	3.21	0.99	3.13	886	150	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	12.1388	11.96
137	5.51	2.64	25	96	1.79	0.6	0.63	1.1	5	0.82	1.69	515	833.3333	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	13.2365	12.53

Rendez-vous dans l'onglet Evaluation pour monitorer les résultats

Page Evaluation

En fonction de l'algorithme sélectionné nous affichons des métriques différentes. Voici ici celles du Random Forest.



Les flèches vers le haut permettront à terme de voir l'évolution de la précision du modèle de machine learning.

Vous pouvez constater grâce à ces données, la précision et la pertinence des données qui sont générées par le modèle.

Nous arrivons au bout de cette présentation, nous avons pour projet de la faire évoluer et nous publieront d'ici 15 ans une ROAD MAP. Qui contiendra tous les changements que nous souhaitons apporter. En plus de prendre en compte les retour utilisateurs.