

# Analyse de données et Classification avec SVM

## Aperçu des données

	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	reason	guardian	traveltime	studytime	fai
0	GP	F	18	U	GT3	A	4	4	at_home	teacher	course	mother	2	2	
1	GP	F	17	U	GT3	T	1	1	at_home	other	course	father	1	2	
2	GP	F	15	U	LE3	T	1	1	at_home	other	other	mother	1	2	
3	GP	F	15	U	GT3	T	4	2	health	services	home	mother	1	3	
4	GP	F	16	U	GT3	T	3	3	other	other	home	father	1	2	

## Informations sur les données

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 649 entries, 0 to 648

Data columns (total 33 columns):

# Column Non-Null Count Dtype

```
-----
0 school    649 non-null object
1 sex       649 non-null object
2 age       649 non-null int64
3 address   649 non-null object
4 famsize   649 non-null object
5 Pstatus   649 non-null object
6 Medu      649 non-null int64
7 Fedu      649 non-null int64
8 Mjob      649 non-null object
9 Fjob      649 non-null object
10 reason   649 non-null object
11 guardian 649 non-null object
12 traveltime 649 non-null int64
```

### Paramètres

Choisissez un fichier CSV

Drag and drop file here  
Limit 200MB per file • CSV

Browse files



student-por-174584554...  
91.0KB

### Paramètres du modèle SVM

Kernel

linear

Paramètre de régularisation (C)

1.00

0.01

10.00

Gamma

scale

```
13 studytime 649 non-null int64
14 failures 649 non-null int64
15 schoolsup 649 non-null object
16 famsup 649 non-null object
17 paid 649 non-null object
18 activities 649 non-null object
19 nursery 649 non-null object
20 higher 649 non-null object
21 internet 649 non-null object
22 romantic 649 non-null object
23 famrel 649 non-null int64
24 freetime 649 non-null int64
25 goout 649 non-null int64
26 Dalc 649 non-null int64
27 Walc 649 non-null int64
28 health 649 non-null int64
29 absences 649 non-null int64
30 G1 649 non-null int64
31 G2 649 non-null int64
32 G3 649 non-null int64
dtypes: int64(16), object(17)
memory usage: 167.4+ KB
```

## Statistiques descriptives

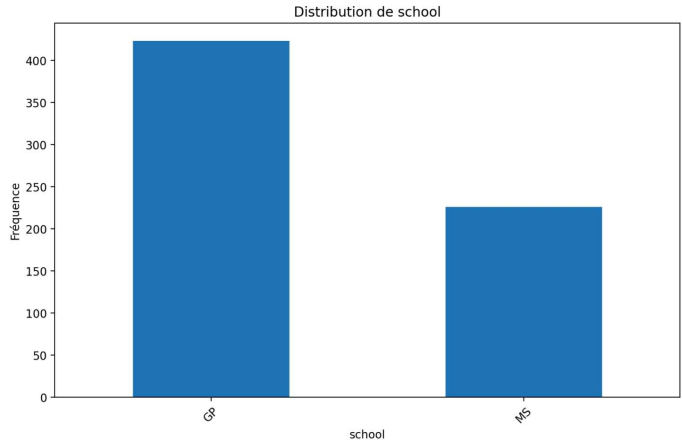
	age	Medu	Fedu	traveltime	studytime	failures	famrel	freetime	goout	Dalc	Walc	health	absences	G1	G2
count	649	649	649	649	649	649	649	649	649	649	649	649	649	649	649
mean	16.7442	2.5146	2.3066	1.5686	1.9307	0.2219	3.9307	3.1803	3.1849	1.5023	2.2804	3.5362	3.6595	11.3991	11.4736
std	1.2181	1.1346	1.0999	0.7487	0.8295	0.5932	0.9557	1.0511	1.1758	0.9248	1.2844	1.4463	4.6408	2.7453	2.7279
min	15	0	0	1	1	0	1	1	1	1	1	1	0	0	0
25%	16	2	1	1	1	0	4	3	2	1	1	2	0	10	10
50%	17	2	2	1	2	0	4	3	3	1	2	4	2	11	11
75%	18	4	3	2	2	0	5	4	4	2	3	5	6	13	13
max	22	4	4	4	4	3	5	5	5	5	5	5	32	19	19

# Visualisation des données

## Diagramme en barres - Variable catégorielle

Sélectionnez une variable catégorielle

school



## Nuage de points - Variables numériques

Sélectionnez la variable X

age

Sélectionnez la variable Y

Medu

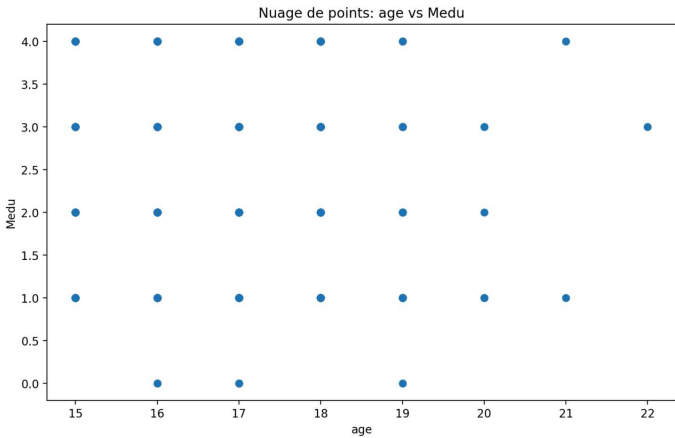
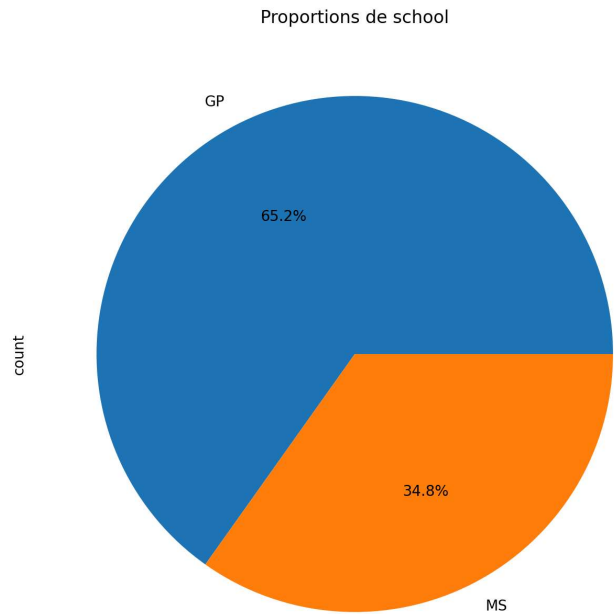
Colorer par (optionnel)

Aucun

# Diagramme circulaire - Proportions

Sélectionnez une variable pour le diagramme circulaire

school



# Modèle de Machine Learning - SVM

Sélectionnez la variable cible (classe)

school

Sélectionnez les variables prédictives

age x Medu x

Taille de l'ensemble de test (%)



Taille de l'ensemble d'entraînement: 519 échantillons

Taille de l'ensemble de test: 130 échantillons

Encodage des classes

```
▼ {  
  "0" : "GP"  
  "1" : "MS"  
}
```

Entraîner le modèle SVM

Modèle entraîné avec succès!

Performances du modèle

Accuracy

0.6692

Recall

0.6692

F1-Score

0.5366

Rapport de classification

	precision	recall	f1-score	support
GP	0.6692	1	0.8018	87
MS	0	0	0	43
accuracy	0.6692	0.6692	0.6692	0.6692
macro avg	0.3346	0.5	0.4009	130
weighted avg	0.4479	0.6692	0.5366	130

Matrice de confusion



