

UNIVERSITE SIDI MOHAMED BEN ABDELLAH
Faculté des Sciences Dhar El Mahraz – Fès
Master BDSAS
Année universitaire 2023-2024



جامعة سيدي محمد بن عبد الله
كلية العلوم دهر المصراخ
- فاس -



Réaliser par :

HALIMA ELHAGOUCHI

- nettoyer le data set (remplacer les valeurs manquants)

#importer les biblio

```
import pandas as pd
import numpy as np
```

```
df=pd.read_csv('tp1.csv')
```

premier chose on va afficher notre origine dataset

```
df
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
...

=>on voit que on a des valeurs nul ou manquant

Alors on doit les remplacer

Pour ça on a pas mal de méthode :

-regression linear (prediction)

-KNN (prediction)

-remplacer par la moyenne la médian .. Si il est une valeur quantitative (numérique)

-remplacer par l'avant ou après

=>on a 177 nul dans âge ,et 2 and embarked et 687 dans cabin

1. utilisant la moyenne pour l'attribue âge :
2. remplacer Cabin par 0 si il est nul sinon par 1
3. remplacer embarked par l'avant ou précédant
4. assurez que que df contient pas des valeurs nul après le prétraitement
5. maintenant on va traiter les règles d'association
c'est pour ça on doit avoir un data set composer par juste des 1 et 0

les règles d'associations

- rendre Notre data set Matrice cruse

=>df avant le transformation :

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	
0	1	0	3	Braund, Mr. Owen Harris	male	22.000000	1	0	A/5 21171	7.2500	0	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.000000	1	0	PC 17599	71.2833	1	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.000000	0	0	STON/O2. 3101282	7.9250	0	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.000000	1	0	113803	53.1000	1	S
4	5	0	3	Allen, Mr. William Henry	male	35.000000	0	0	373450	8.0500	0	S
...

1. On va retirer l'attribue Ticket et Fare
2. Mapper les valeurs de 'Embarked' à des numéros
3. traite l'attribue Name et diviser le par 6 class :master dr mr mrs miss name
4. Traite l'Age et la diviser par quatre class: adulte adolescents enfant aines
5. Faire la même pour sex ,embarked et Pclass ,Cabin ,Parch,Survived,Sib

⇒ **Df après la transformation :**

PassengerId	dr	master	miss	mr	mrs	name	adolescents	adult	ainés	...	s8	sub0	sub1	par0	pr1	par2	par3	pr4
0	1	0	0	0	1	0	0	1	0	0	...	0	1	0	1	0	0	0
1	2	0	0	0	0	1	0	0	1	0	...	0	0	1	1	0	0	0
2	3	0	0	1	0	0	0	0	1	0	...	0	0	1	1	0	0	0
3	4	0	0	0	0	1	0	0	1	0	...	0	0	1	1	0	0	0
4	5	0	0	0	1	0	0	0	1	0	...	0	1	0	1	0	0	0
...
886	887	0	0	0	0	0	1	0	1	0	...	0	1	0	1	0	0	0
887	888	0	0	1	0	0	0	1	0	0	...	0	0	1	1	0	0	0
888	889	0	0	1	0	0	0	0	1	0	...	0	1	0	0	0	1	0
889	890	0	0	0	1	0	0	0	1	0	...	0	0	1	1	0	0	0
890	891	0	0	0	1	0	0	0	1	0	...	0	1	0	1	0	0	0

- **Algorithme apriori**
=> Pour cette étape on a besoin d'autre bibliothèque appelle mlxtend

```
from mlxtend.preprocessing import
TransactionEncoder
```

```
from mlxtend.frequent_patterns import apriori,
association_rules
```

1. Convertir les transactions en une matrice binaire
2. Appliquer l'algorithme Apriori pour trouver les itemsets fréquents
3. Générer les règles d'association

- **Rules :**

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
0	(miss)	(female)	0.204265	0.352413	0.204265	1.000000	2.837580	0.132279	inf	0.813822
1	(mr)	(adult)	0.580247	0.710438	0.446689	0.769826	1.083594	0.034460	1.258014	0.183787
2	(mr)	(male)	0.580247	0.647587	0.580247	1.000000	1.544194	0.204487	inf	0.839572
3	(male)	(mr)	0.647587	0.580247	0.580247	0.896014	1.544194	0.204487	4.036626	1.000000
4	(mr)	(cab0)	0.580247	0.771044	0.475870	0.820116	1.063644	0.028474	1.272800	0.142550
...
3492	(class3, cab0, mr)	(0, par0, male, sub0)	0.352413	0.389450	0.255892	0.726115	1.864462	0.118645	2.229217	0.715969
3493	(class3, mr, sub0)	(0, cab0, male, par0)	0.317621	0.395062	0.255892	0.805654	2.039311	0.130412	3.112682	0.746855
3494	(class3, mr, par0)	(0, cab0, male, sub0)	0.331089	0.353535	0.255892	0.772881	2.186150	0.138841	2.846374	0.811131
3495	(class3, sub0, par0)	(0, cab0, mr, male)	0.331089	0.388328	0.255892	0.772881	1.990281	0.127321	2.693184	0.743833

Interprétation :

- La première règle entre miss item et female item indique que il y a une probabilité de **100%** que "miss" et 'female' ont une forte association en vue de **confidence** qui égale **1**, Le **lift** de **2.83** suggère une association positive.

- la deuxième règle entre mr item et male item indique que il y a une probabilité de **75%** que "mr" et 'male' ont une forte association en vue de **confidence** qui égale **0.75**, Le **lift** de **1.083** suggère une association positive

Alors 75% des adultes sont des mr

Et comme ça on lit toutes les autres lignes.....