Centre d'Excellence Interdisciplinaire en Intelligence Artificielle pour le Développement

université virtuelle Burkina Faso

# CITADEL AI Summer School 2022

## Machine Learning


19-25.09.22

Rodrique Kafando, PhD
rodrique.kafando@citadel.bf
kafando.rodrique@gmail.com

# Objectifs

- Donnez des exemples d'apprentissage automatique dans diverses industries;

- Décrire les étapes utilisées par l'apprentissage automatique pour résoudre les problèmes ;

- Donner des exemples de diverses techniques utilisées dans l'apprentissage automatique ;

- Décrire les bibliothèques Python généralement utilisées pour l'apprentissage automatique ;

- Expliquer les différences entre les algorithmes supervisés et non supervisés ;

- Décrire les capacités des différents algorithmes.

# Overwiew

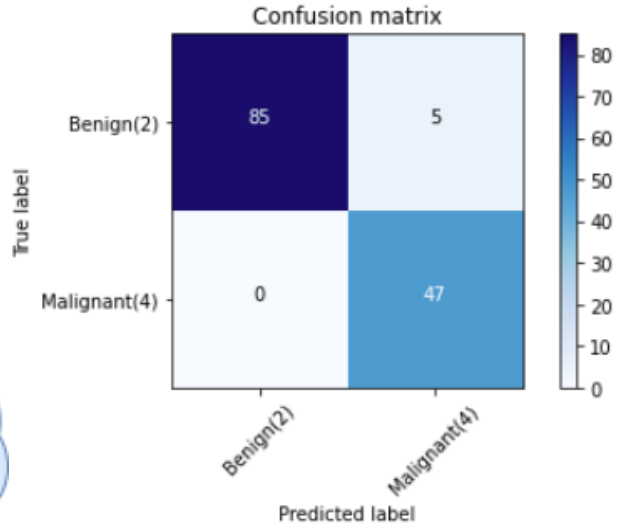

Confiance
Responsabilité
Sécurité
Inclusion
CONFIDENTIALITÉ & EXPLICABILITÉ
FINTECH & SÉCURITÉ
SANTÉ & BIEN-ÊTRE
AGRICULTURE & PASTORALISME
LANGUES & PATRIMOINE
ÉNERGIE & CHANGEMENT CLIMATIQUE

# Machine Learning pour faire de la prédiction..

| | ID | Clump | UnifSize | UnifShape | MargAdh | SingEpiSize | BareNuc | BlandChrom | NormNucl | Mit | Class |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 689 | 654546 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 8 | benign |
| 690 | 654546 | 1 | 1 | 1 | 3 | 2 | 1 | 1 | 1 | 1 | benign |
| 691 | 695091 | 5 | 10 | 10 | 5 | 4 | 5 | 4 | 4 | 1 | malignan |
| 692 | 714039 | 3 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | benign |
| 693 | 763235 | 3 | 1 | 1 | 1 | 2 | 1 | 2 | 1 | 2 | benign |
| 694 | 776715 | 3 | 1 | 1 | 1 | 3 | 2 | 1 | 1 | 1 | benign |
| 695 | 841769 | 2 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | benign |
| 696 | 888820 | 5 | 10 | 10 | 3 | 7 | 3 | 8 | 10 | 2 | malignan |
| 697 | 897471 | 4 | 8 | 6 | 4 | 3 | 4 | 10 | 6 | 1 | malignan |
| 698 | 897471 | 4 | 8 | 8 | 5 | 4 | 5 | 10 | 4 | 1 | malignan |

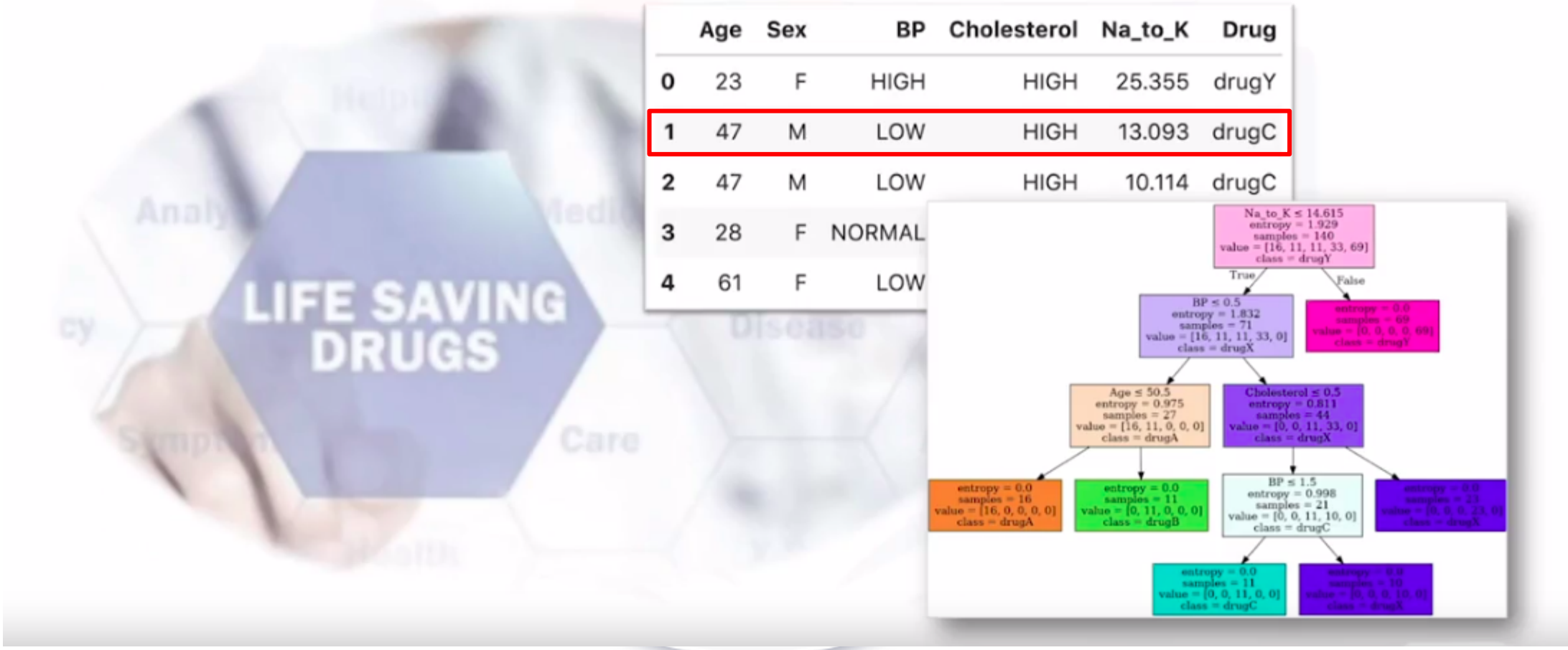
Modeling

Prediction




Centre d'Excellence Interdisciplinaire en Intelligence Artificielle pour le Développement

# Machine Learning pour faire de la prédiction...




| | Age | Sex | BP | Cholesterol | Na_to_K | Drug |
|---|---|---|---|---|---|---|
| 0 | 23 | F | HIGH | HIGH | 25.355 | drugY |
| 1 | 47 | M | LOW | HIGH | 13.093 | drugC |
| 2 | 47 | M | LOW | HIGH | 10.114 | drugC |
| 3 | 28 | F | NORMAL | | | |
| 4 | 61 | F | LOW | | | |

# Machine Learning pour la prise de decision..

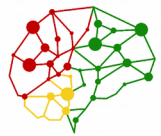
| age | ed | employ | address | income | debtinc | creddebt | othdebt | default |
|---|---|---|---|---|---|---|---|---|
| 41 | 3 | 17 | 12 | 176 | 9.3 | 11.359 | 5.009 | 1 |
| 27 | 1 | 10 | 6 | 31 | 17.3 | 1.362 | 4.001 | 0 |
| 40 | 1 | 15 | 14 | 55 | 5.5 | 0.856 | 2.169 | 0 |
| 41 | 1 | 15 | 14 | 120 | 2.9 | 2.659 | 0.821 | 0 |
| 24 | 2 | 2 | 0 | 28 | 17.3 | 1.787 | 3.057 | 1 |
| 41 | 2 | 5 | 5 | 25 | 10.2 | 0.393 | 2.157 | 0 |
| 39 | | | 9 | 67 | 30.6 | 3.834 | 16.668 | 0 |
| 43 | | | | 38 | 3.6 | 0.129 | 1.239 | 0 |
| | | | | 19 | 24.4 | 1.358 | 3.278 | 1 |
| | | | | 25 | 19.7 | 2.778 | 2.147 | 0 |

| ome | debtinc | creddebt | othdebt | default |
|---|---|---|---|---|
| 30 | 9.3 | 10.23 | 3.21 | No |


Categorical Variable

Modeling

Prediction

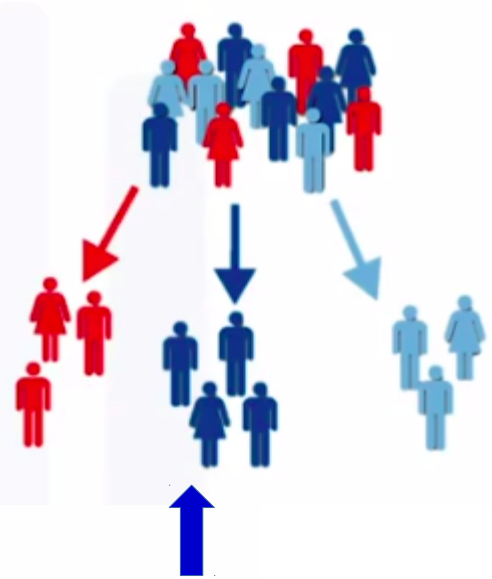


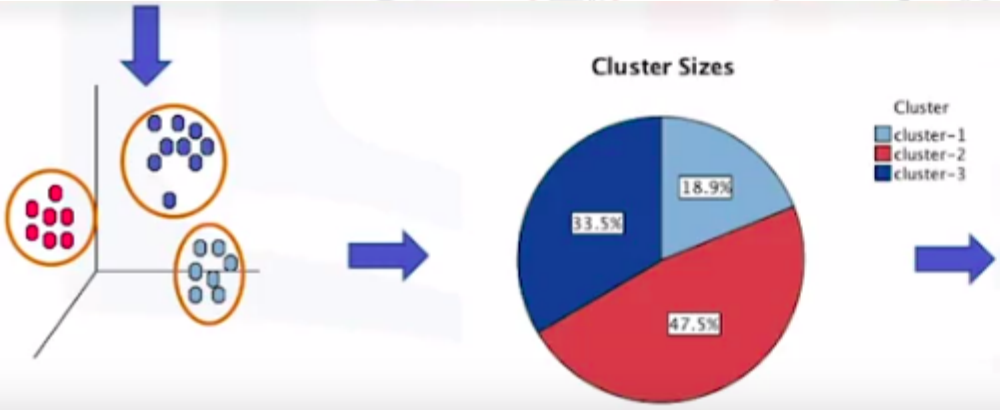
CITADEL

Centre d'Excellence Interdisciplinaire en Intelligence Artificielle pour le Développement

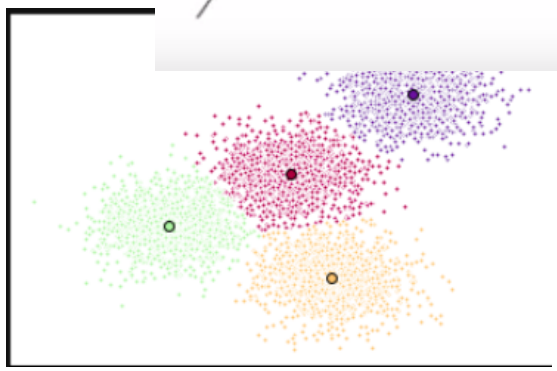
# M Learning pour la segmentation des clients...

| | Customer Id | Age | Edu | Years Employed | Income | Card Debt | Other Debt | Defaulted | Address | DebtIncomeRatio |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 41 | 2 | 6 | 19 | 0.124 | 1.073 | 0.0 | NBA001 | 6.3 |
| 1 | 2 | 47 | 1 | 26 | 100 | 4.582 | 8.218 | 0.0 | NBA021 | 12.8 |
| 2 | 3 | 33 | 2 | 10 | 57 | 6.111 | 5.802 | 1.0 | NBA013 | 20.9 |
| 3 | 4 | 29 | 2 | 4 | 19 | 0.681 | 0.516 | 0.0 | NBA009 | 6.3 |
| 4 | 5 | 47 | 1 | 31 | 253 | 9.308 | 8.908 | 0.0 | NBA008 | 7.2 |

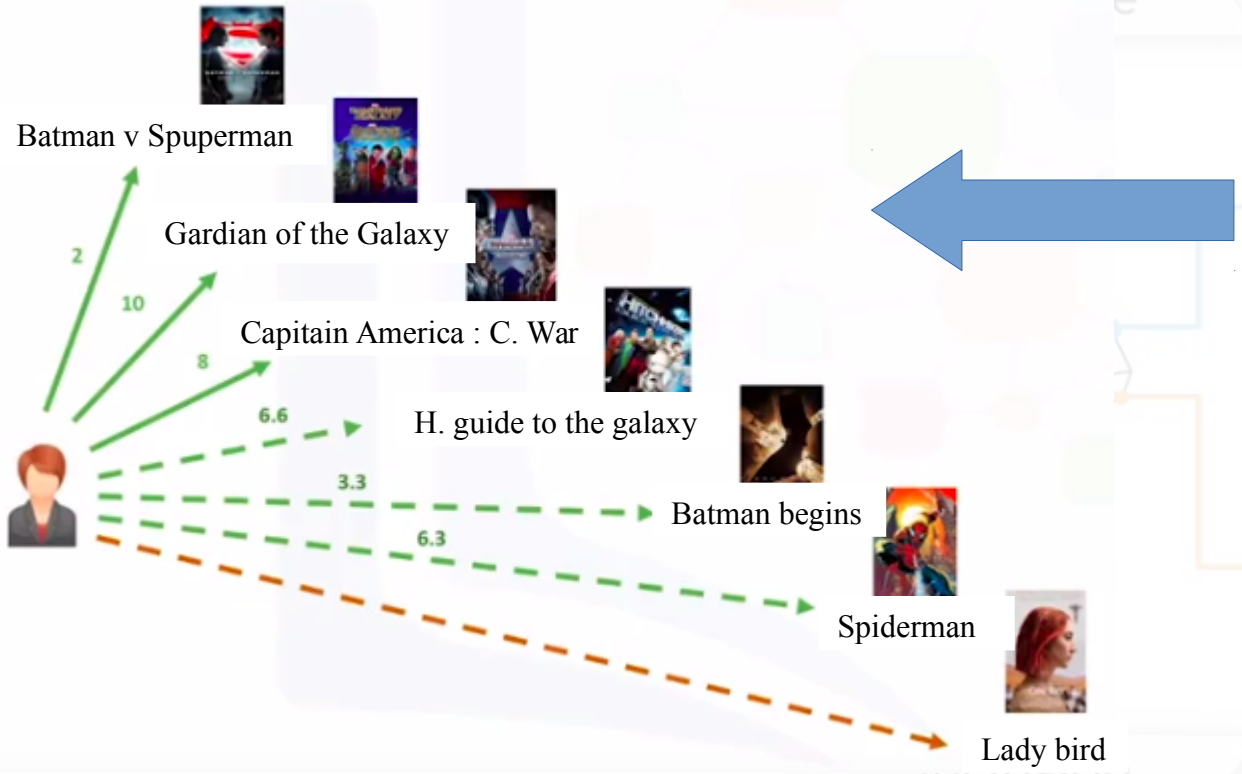


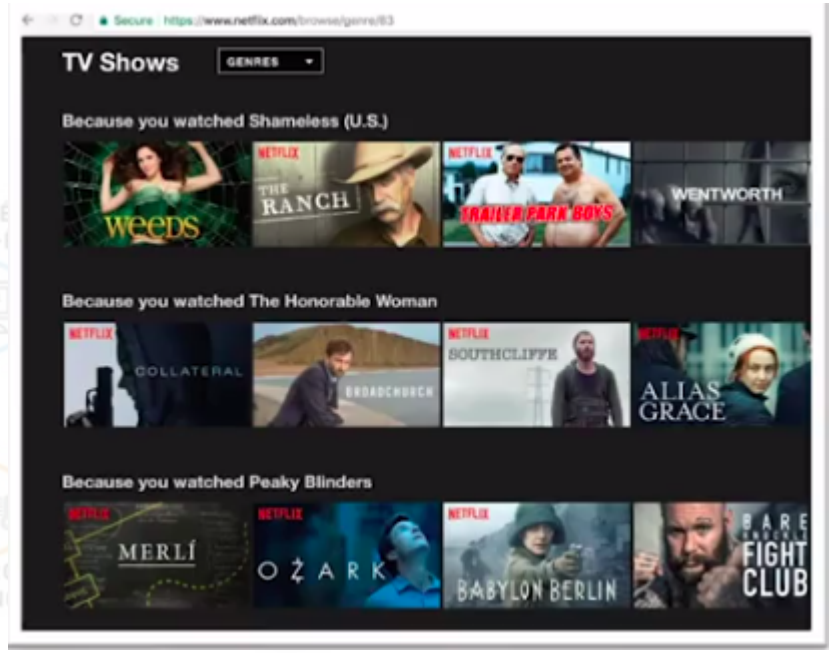



| Cluster | Segment Name |
|---|---|
| cluster-1 | AFFULUENT AND MIDDLE AGED |
| cluster-2 | YOUNG EDUCATED AND MIDDLE INCOME |
| cluster-3 | YOUNG AND LOW INCOME |



# M Learning pour systèmes de recommandations...

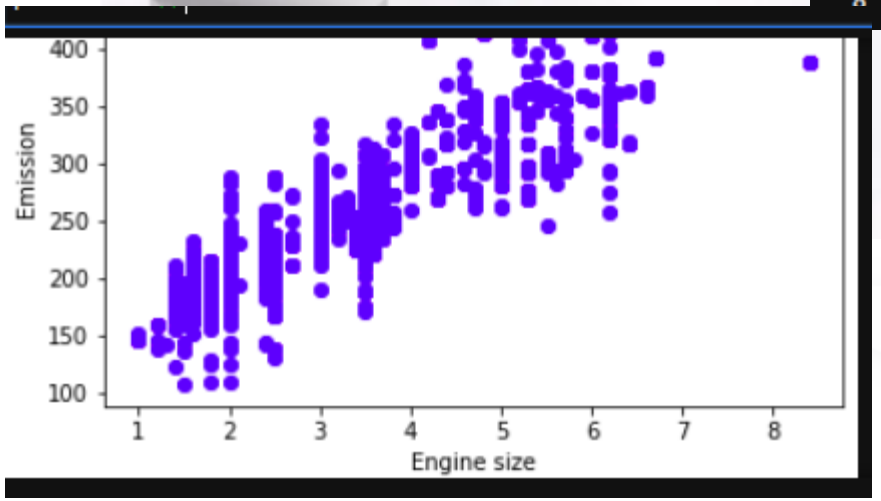



CITADEL
Centre d'Excellence Interdisciplinaire en Intelligence Artificielle pour le Développement

# Machine Learning sous python en quelques lignes...

| | ENGINESIZE | CYLINDERS | FUELCONSUMPTION_COMB | CO2EMISSIONS |
|---|---|---|---|---|
| 0 | 2.0 | 4 | 8.5 | 196 |
| 1 | 2.4 | 4 | 9.6 | 221 |
| 2 | 1.5 | 4 | 5.9 | 136 |
| 3 | 3.5 | 6 | 11.1 | 255 |
| 4 | 3.5 | 6 | 10.6 | 244 |
| 5 | 3.5 | 6 | 10.0 | 230 |
| 6 | 3.5 | 6 | 10.1 | 232 |
| 7 | 3.7 | 6 | 11.1 | 255 |
| 8 | 3.7 | 6 | 11.6 | 267 |



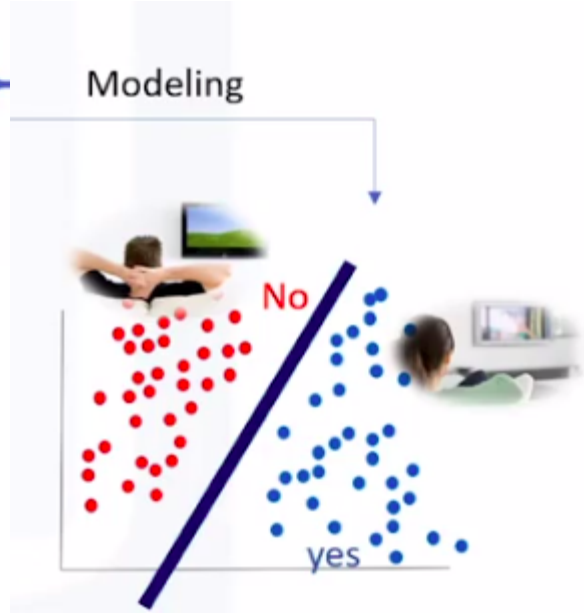

```
clf = linear_model.LinearRegression()
train_y_ = clf.fit(train_x_poly, train_y)
# Coefficients and Intercept
print ('Coefficients: ', clf.coef_)
print ('Intercept: ',clf.intercept_)
```

```
Coefficients:  [[ 0.         47.33821833 -1.03899124]]
Intercept:  [112.18904389]
```

# Machine Learning sous python en quelques lignes...

| | tenure | age | address | income | ed | employ | equip | callcard | wireless | churn |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 11.0 | 33.0 | 7.0 | 136.0 | 5.0 | 5.0 | 0.0 | 1.0 | 1.0 | Yes |
| 1 | 33.0 | 33.0 | 12.0 | 33.0 | 2.0 | 0.0 | 0.0 | 0.0 | 0.0 | Yes |
| 2 | 23.0 | 30.0 | 9.0 | 30.0 | 1.0 | 2.0 | 0.0 | 0.0 | 0.0 | No |
| 3 | 38.0 | 35.0 | 5.0 | 76.0 | 2.0 | 10.0 | 1.0 | 1.0 | 1.0 | No |
| 4 | 7.0 | 35.0 | 14.0 | 80.0 | 2.0 | 15.0 | 0.0 | 1.0 | 0.0 | ? |

No



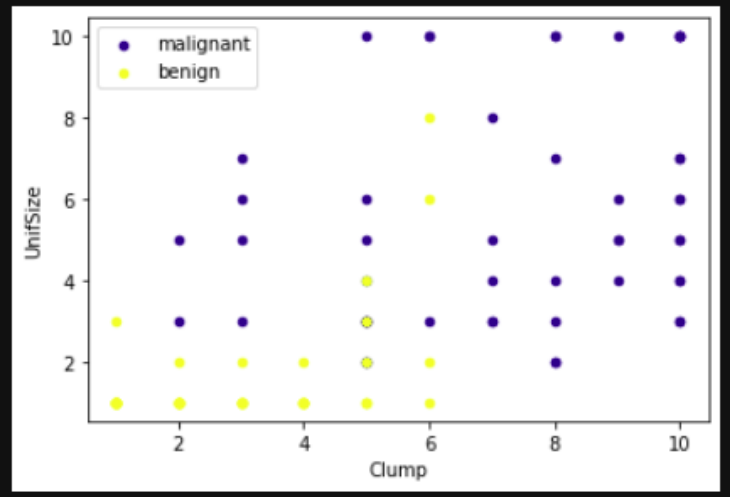
# Environnement de travail...

1 to 10, with 1 being the closest to benign.

The Class field contains the diagnosis, as confirmed by separate medical procedures, as to whether the samples are benign (value = 2) or malignant (value = 4).

Let's look at the distribution of the classes based on Clump thickness and Uniformity of cell size:

[40]:
```
ax = cell_df[cell_df['Class'] == 4][0:50].plot(kind='scatter', x='Clump', y='UnifSize', color='DarkBlue', label='malignant');
cell_df[cell_df['Class'] == 2][0:50].plot(kind='scatter', x='Clump', y='UnifSize', color='Yellow', label='benign', ax=ax);
plt.show()
```




## Data pre-processing and selection

Let's first look at columns data types:

[41]:
```
cell_df.dtypes
```

jupyterlab TNG

# Ce que vous retiendrez de ce cours....

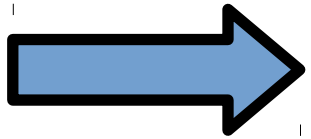
**Des compétences :**

Régression

Classification

Clustering

Sklearn & Scipy



Appliquées sur des
cas d'études concrets