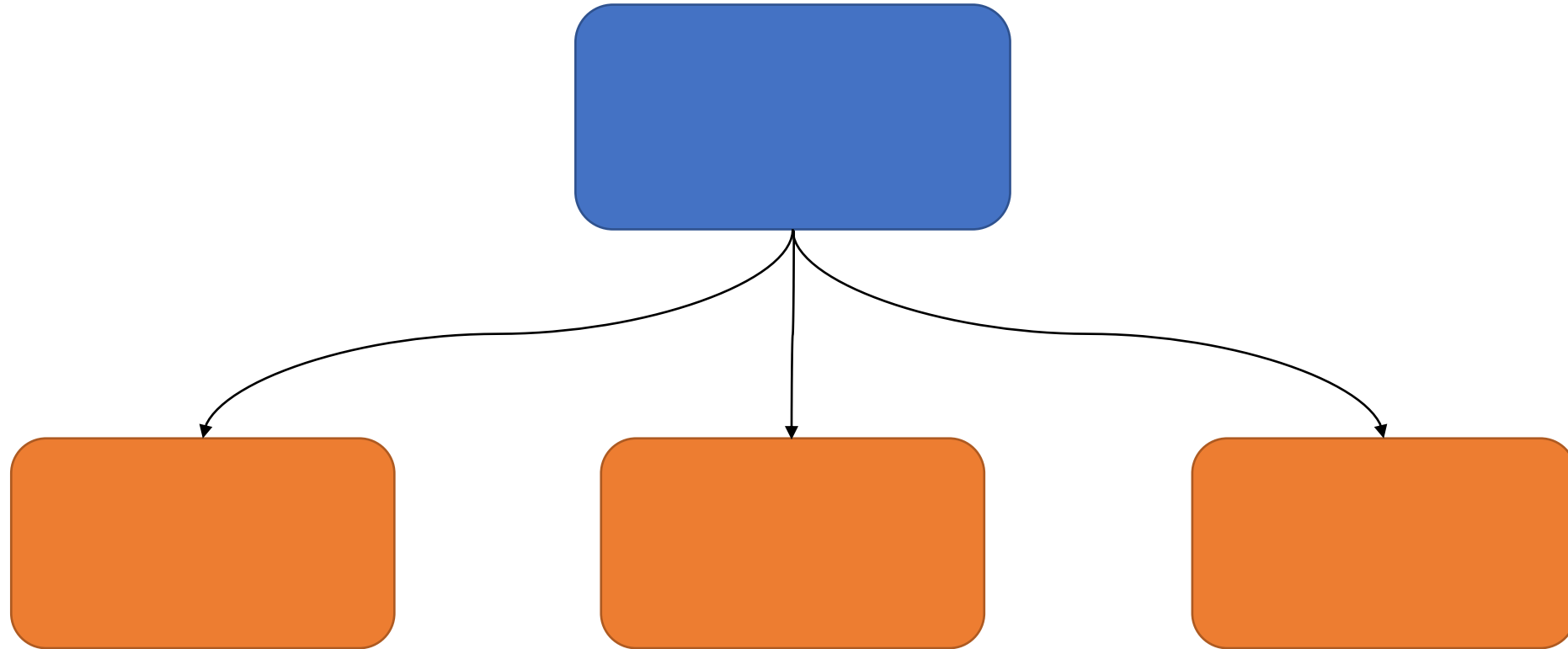


# Arbre de décision – ID3



<https://github.com/ForrayGabriel/PROJ631-Arbre-de-decision>

# L'algorithme ID3

Dataset(attributs, valeurs) + situation(valeur) → résultat

## ❑ Phase d'apprentissage:

- Choisir le meilleur attribut : Calcule du gain d'information de chaque attribut grâce à son entropie
- Créer un nouveau nœud pour le meilleur attributs
- Partitionner en sous ensemble ayant la même valeur pour l'attribut sélectionné
- Créer un arc pour chaque valeur et recommencer jusqu'à avoir fait tous les attributs

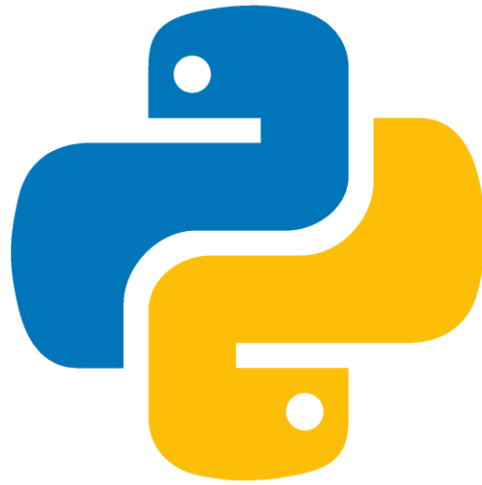
## ❑ Phase de prédiction:

- Récupérer l'arbre
- Passer de nœud en nœud jusqu'au résultat

$$\text{Entropy}(S) = - \sum p_i * \log_2(p_i)$$

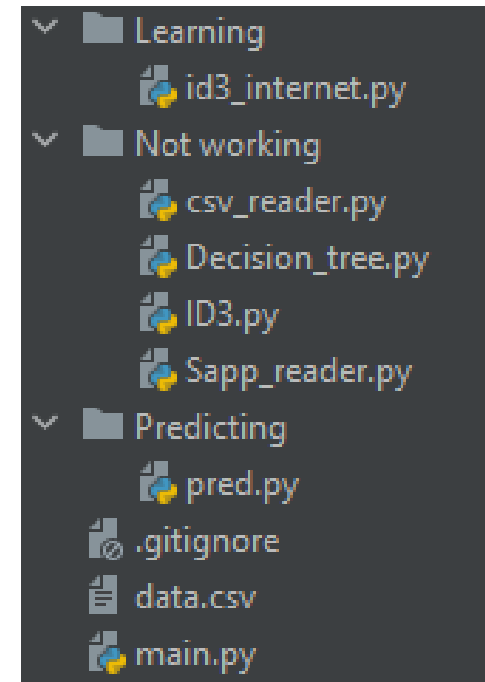
$$\text{IG}(S, A) = \text{Entropy}(S) - \sum ((|S_v| / |S|) * \text{Entropy}(S_v))$$

# Aspect technique

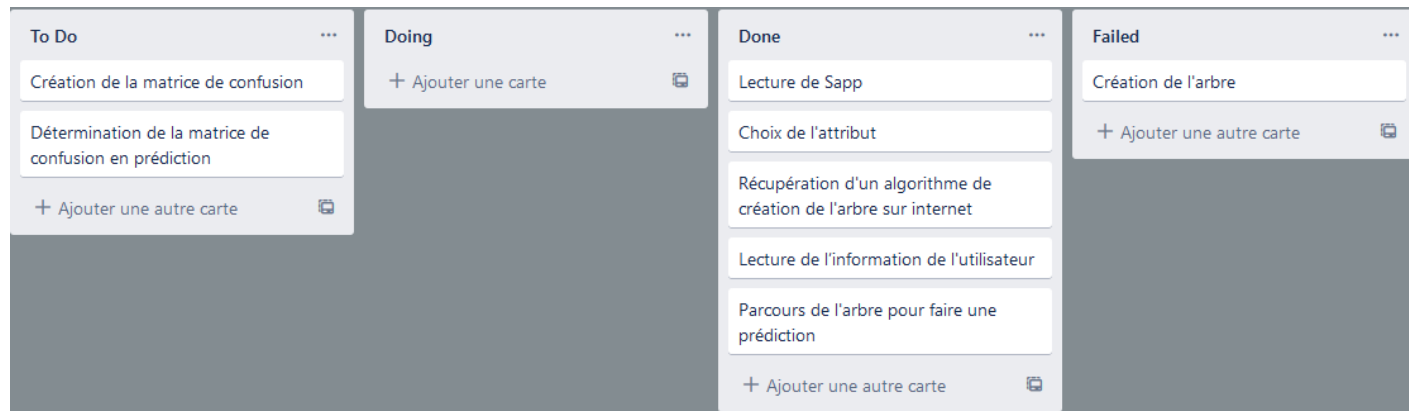


Python

With PEP8 formatting

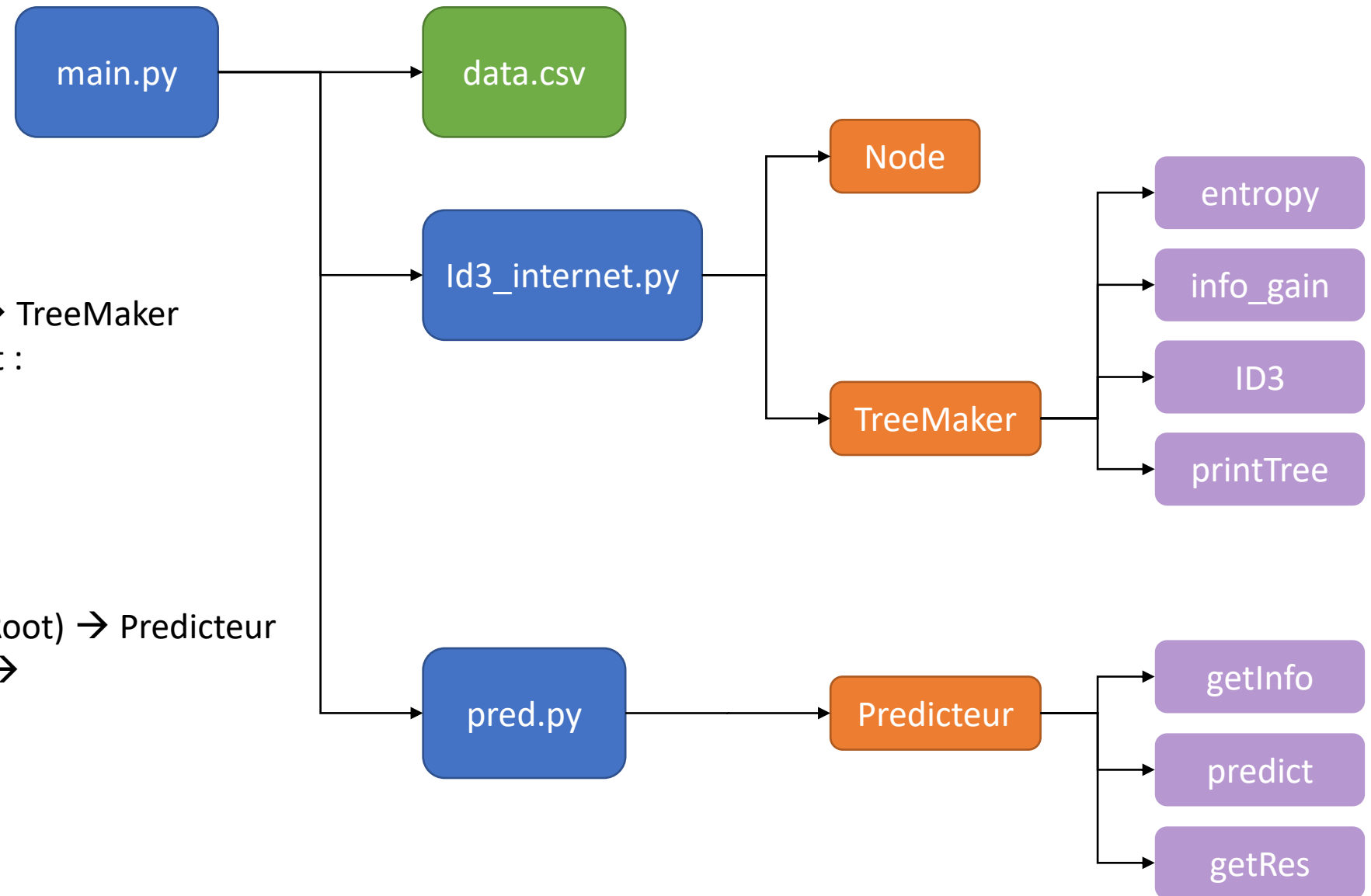


Organisation des fichiers



Trello avec les tâches à faire, entrain d'être faites, faites et échouées

# Décomposition fonctionnelle du programme



Main.py →

- data.csv → data
- Id3\_internet.py.TreeMaker → TreeMaker
- TreeMaker.ID3(data) → Root :
  - Info\_gain →
    - Entropy
- User choice :
  - 1 – print(data)
  - 2 – TreeMaker.printTree
  - 3 – pred.py.Predicteur(Root) → Predicteur
    - Predicteur.getRes →
      - getInfo
      - Predict(Root)

# Exemple

	A
1	outlook,temperature,humidity,wind,answer
2	sunny,hot,high,weak,no
3	sunny,hot,high,strong,no
4	overcast,hot,high,weak,yes
5	rain,mild,high,weak,yes
6	rain,cool,normal,weak,yes
7	rain,cool,normal,strong,no
8	overcast,cool,normal,strong,yes
9	sunny,mild,high,weak,no
10	sunny,cool,normal,weak,yes
11	rain,mild,normal,weak,yes
12	sunny,mild,normal,strong,yes
13	overcast,mild,high,strong,yes
14	overcast,hot,normal,weak,yes
15	rain,mild,high,strong,no



```
The data are :
outlook temperature humidity wind answer
0 sunny hot high weak no
1 sunny hot high strong no
2 overcast hot high weak yes
3 rain mild high weak yes
4 rain cool normal weak yes
5 rain cool normal strong no
6 overcast cool normal strong yes
7 sunny mild high weak no
8 sunny cool normal weak yes
9 rain mild normal weak yes
10 sunny mild normal strong yes
11 overcast mild high strong yes
12 overcast hot normal weak yes
13 rain mild high strong no
```

Donnée brute du .csv

Donnée extraient par python

```
outlook
  overcast -> ['yes']
  rain
    wind
      strong -> ['no']
      weak -> ['yes']
  sunny
    humidity
      high -> ['no']
      normal -> ['yes']
```

Arbre de décision obtenu

```
What is the outlook :
overcast
What is the temperature :
cool
What is the humidity :
normal
What is the wind :
strong
The prediction is : yes
```

Prédiction pour une certaine météo

```
What is the outlook :
sunny
What is the temperature :
mild
What is the humidity :
high
What is the wind :
weak
The prediction is : no
```

Prédiction pour une autre météo

```
What is the outlook :
wonderful
What is the temperature :
horrible
What is the humidity :
wet
What is the wind :
absent
Error : user entry doesn't correspond to any known data
```

Gestion des erreurs de saisie