Prédiction conformelle et base de données

Pl@ntNet-CrowdSWE

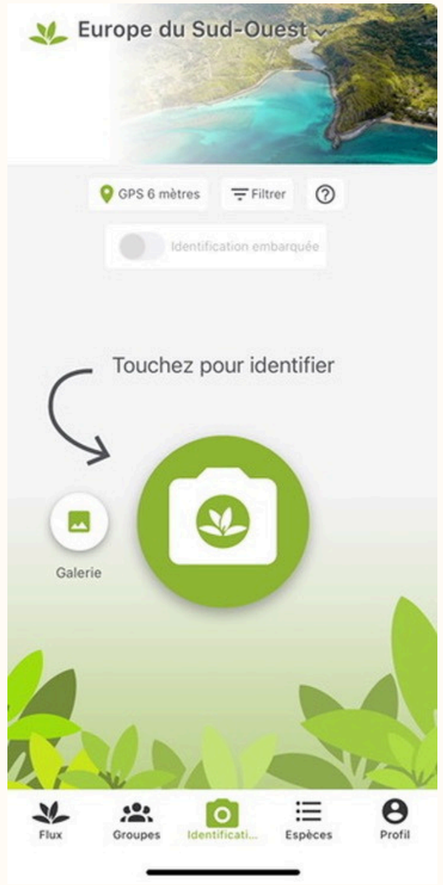
Présenté par :

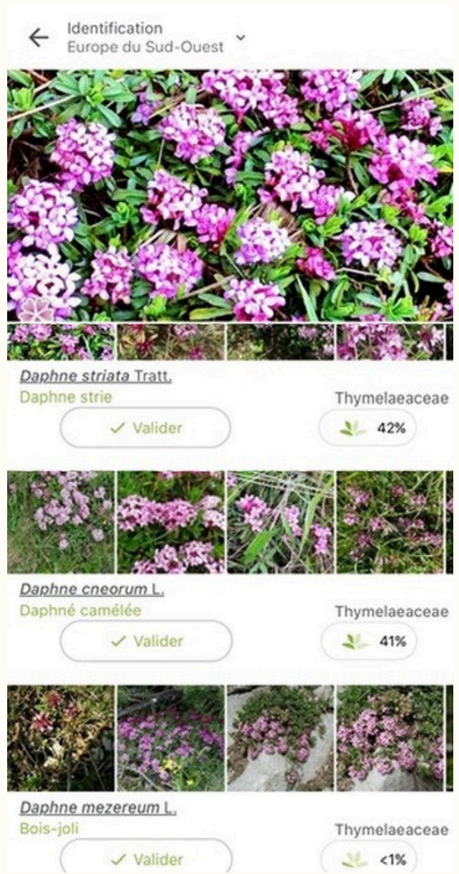
AIGOIN Emilie
CLETZ Laura
THOMAS Anne-Laure

Encadré par :

BOTELLA Christophe
SALMON Joseph

Application Pl@ntNet





Objectifs du projet :

Nombre faible de propositions d'espèces

Prédictions les plus correctes possibles

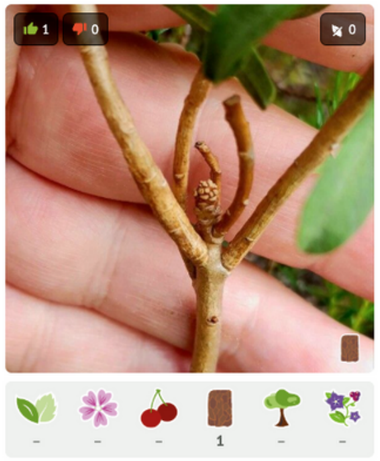
Application Pl@ntNet

Application Pl@ntNet

Feuille

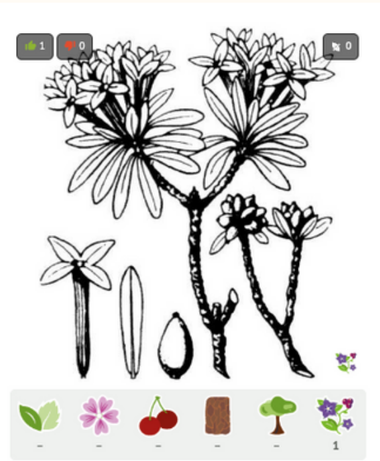
Fleur

Fruit



Ecorce

Plante entière



Autre

Application Pl@ntNet

Valider | Valider cette observation (qualité de l'image, organe, détermination)

Détermination

Déterminations proposées

NOM LE PLUS PROBABLE

Daphne striata Tratt. Daphne strie — H

73 %

NOM SOUMIS

Rosmarin Seidelbast — M

27 %

Suggérer une autre détermination

Nom de l'espèce | Suggérer | Flore mondiale

6 699 593 observations

6 673 092 non-jugées par un expert

26 501 jugées par un expert

Un exemple concret !

1014342716

DAPHNE CNEORUM L. (2443)
0,9424
A REÇU 2 VOTES !

dont l'expert 40062

DAPHNE STRIATA TRATT. (4553)
0,1638
A REÇU 0 VOTE !

DAPHNE CNEORUM L. (2443)
0,9424

DAPHNE STRIATA TRATT. (4553)
0,1638

DAPHNE ALPINA L. (1681)
0,0458

Aeonium arboreum Webb & Berthel.

Pavonia spinifex (L.) Cav.

Aloe distans Haw.

Corylus avellana L.

Fagus sylvatica L.

Ranunculus ficaria L.

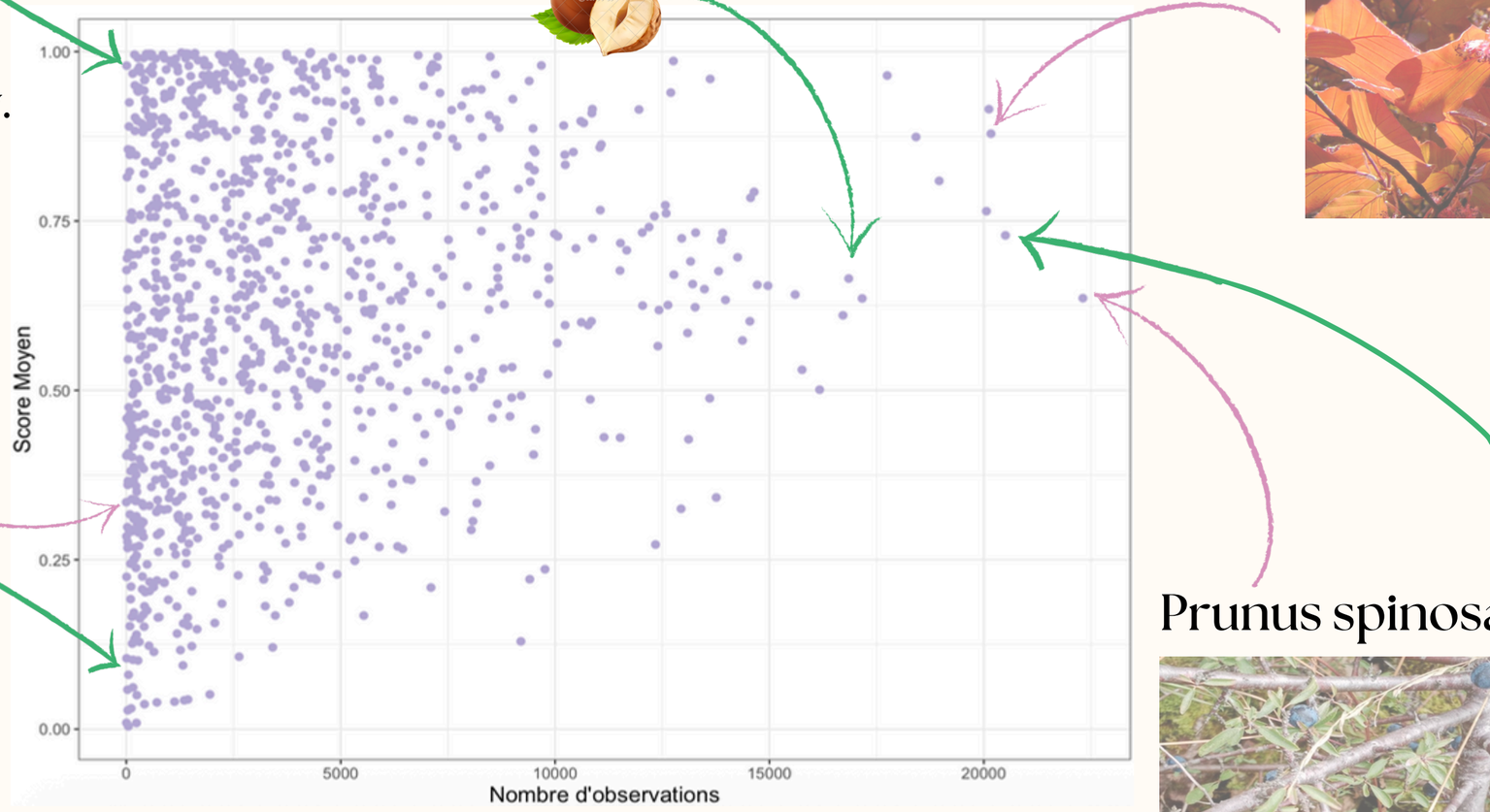
Prunus spinosa L.

Aeonium arboreum Webb & Berthel.

Corylus avellana L.

Fagus sylvatica L.

Pavonia spinifex (L.) Cav.



Ranunculus ficaria L.

Aloe distans Haw.

Prunus spinosa L.

Étude autour de la moyenne

1011121287

par Walter Reider le 23 Juin 2021

DAPHNE STRIATA TRATT.
(4553)
66,45%

DAPHNE CNEORUM L.
(2443)
21,49%

DAPHNE ALPINA L.
(1681)
1,99%

1011121287

par Walter Reider le 23 Juin 2021

DAPHNE STRIATA TRATT.
(4553)
66,45%

DAPHNE CNEORUM L.
(2443)
21,49%

DAPHNE ALPINA L.
(1[illegible])
1,[illegible]9%

Une solution :

la prédiction conformelle

Le cadre mathématique

Soient $(X_i, Y_i)_{\{1 \leqslant i \leqslant n\}}$, $X_i \in \mathbb{R}^d$, $Y_i \in \{1, \ldots, K\}$ et $\alpha \in]0, 1[$.
On considère

$$\mathbb{P}(Y_{n+1} \in \hat{C}_n(X_{n+1})) \geq 1 - \alpha$$

A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification

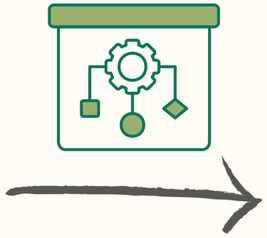
Anastasios N. Angelopoulos and Stephen Bates
December 8, 2022
https://arxiv.org/pdf/2107.07511

✘

- ○ _____
- ○ _____
- ○ _____
- ○ _____
- ○ _____
- ○ _____

✘

- ○ _____
- ○ _____



✓

- ○ _____
- ○ _____
- ○ _____
- ○ _____

95%

ÉTAPE 2 : Calculs des scores

Le score softmax :

$$s_1(X_i, Y_i) = 1 - p_{Y_i}(X_i)$$

Le score cumulatif APS (Adaptive Prediction Sets) :

$$s_2(X_i, Y_i) = \sum_{j=1}^{r(X_i,Y_i)-1} p_{(j)}(X_i)$$

→ Données expertes (~26 000)

→ Données non expertes (~6 millions)

ÉTAPE 1

Attribution des bons labels

→ Vote experts = justes
→ Vote non-experts = vote majoritaire pondéré de 3

ÉTAPE 3

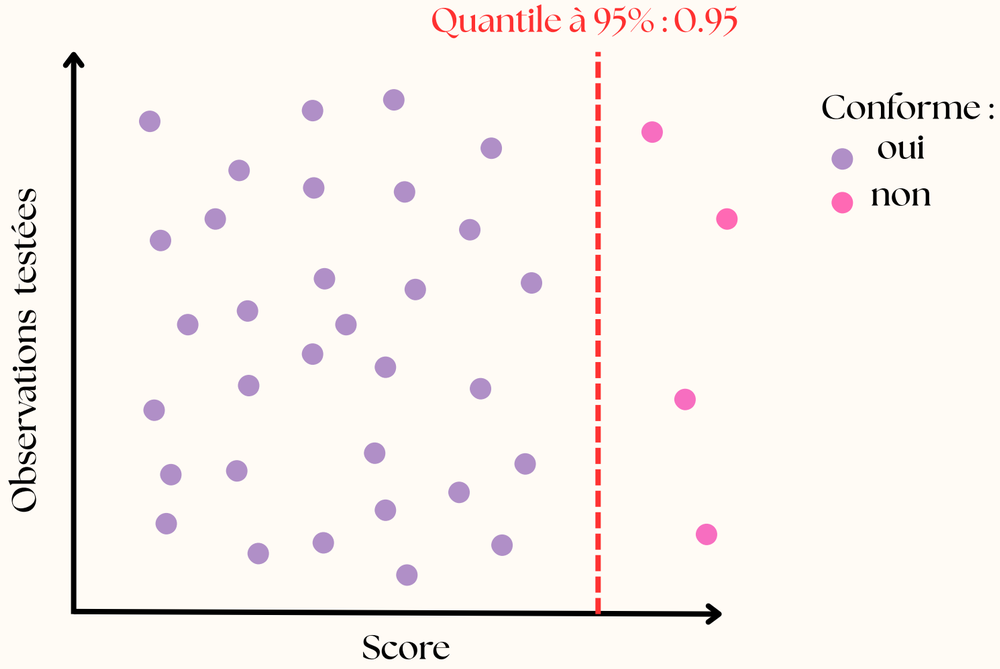
Construction des ensembles de calibration et de test

- Ensemble de calibration → calculer un seuil (le quantile).
- Ensemble de tests, pour vérifier si le modèle atteint bien les 95 % de couverture.

→ Observations imputées

Le quantile

- $\frac{\lceil (n+1)(1-\alpha) \rceil}{n}$, avec $\alpha = 0,05$
- Softmax
- Cumulatif APS



Exemple de graphe de conformité

Tableau des configurations

Score	Calibration	Test
Softmax	Non-expert	Seconde moitié des experts
	Première moitié des experts	
Cumulatif APS	Non-expert	
	Première moitié des experts	

Mesures :

Taille moyenne

$$\frac{1}{n}\sum_{i=1}^{n} \text{taille}(C_i)$$

- n : nombre total d'observations testées
- C_i : ensemble de prédiction pour l'observation i
- $\text{taille}(C_i)$: nombre d'espèces dans l'ensemble C_i

Taille médiane = valeur centrale des tailles (Ci)

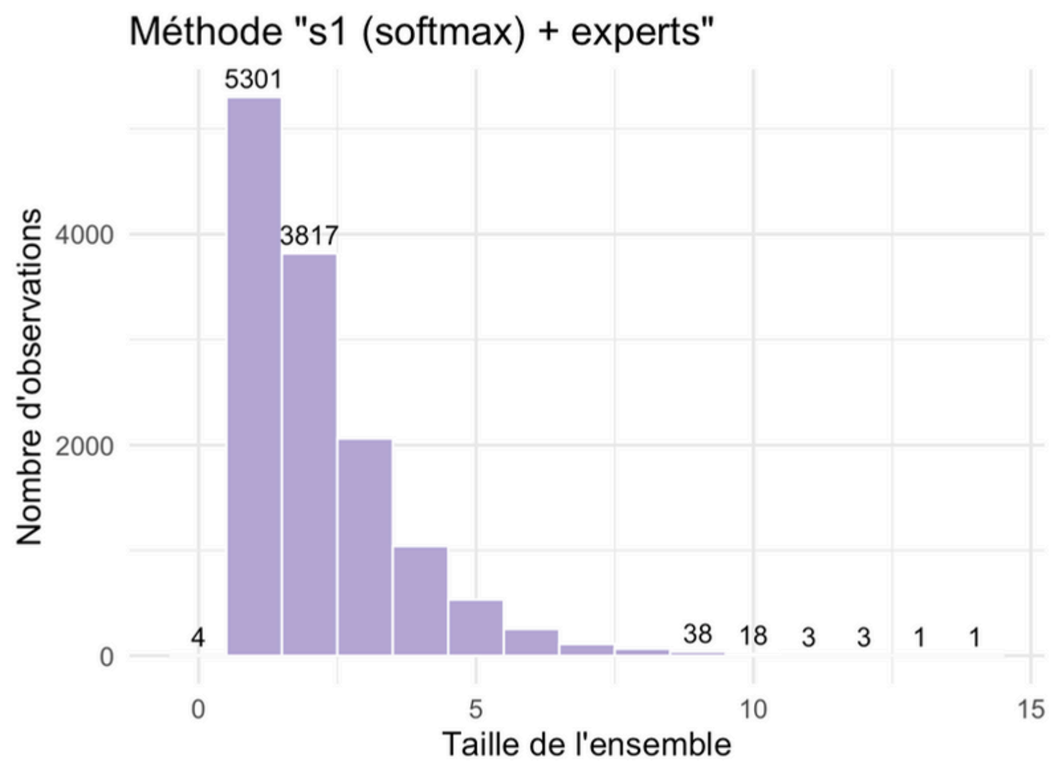
Test statistique du Chi²
→ Seuil de 3,84

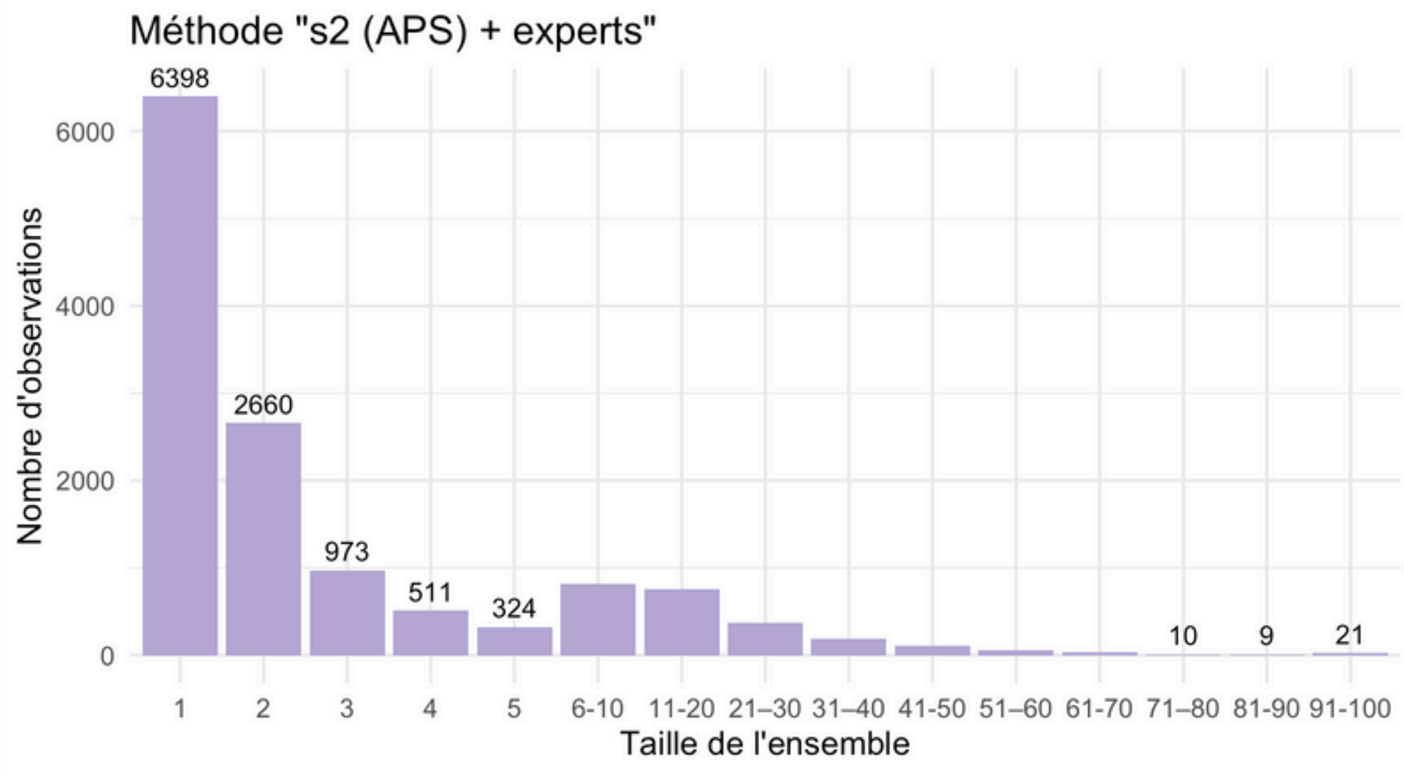
Graines

Taux de couverture

Tableau des résultats des configurations

Score	Données de calibration	Quantiles	Taux de couverture	Test du Chi-deux
~~Softmax~~	~~Non expertes~~	~~0,9990~~	~~96,45%~~	~~58,9~~
Softmax	Expertes	0,9649	94,82%	0,87
~~Cumulatif APS~~	~~Non expertes~~	~~0,9990~~	~~96,45%~~	~~58,9~~
Cumulatif APS	Expertes	0,7932	94,76%	1,57
			Objectif de 95%	Seuil de 3,84





Médiane : 2
Moyenne : 2,13

Médiane : 2
Moyenne : 4,51

Etape 1

Etape 2

Daphne striata Tratt.	0,7547
Daphne cneorum L.	0,2100
Daphne mezereum L.	0,0045
Daphne alpina L.	0,0020
Daphne sericea Vahl	0,0014

Etape 3

Score softmax :

$$s_1(X_i, Y_i) = 1 - p_{Y_i}(X_i)$$

Score cumulatif APS :

$$s_2(X_i, Y_i) = \sum_{j=1}^{r(X_i,Y_i)-1} p_{(j)}(X_i)$$

Etape 4

Score < quantile

→ conservation

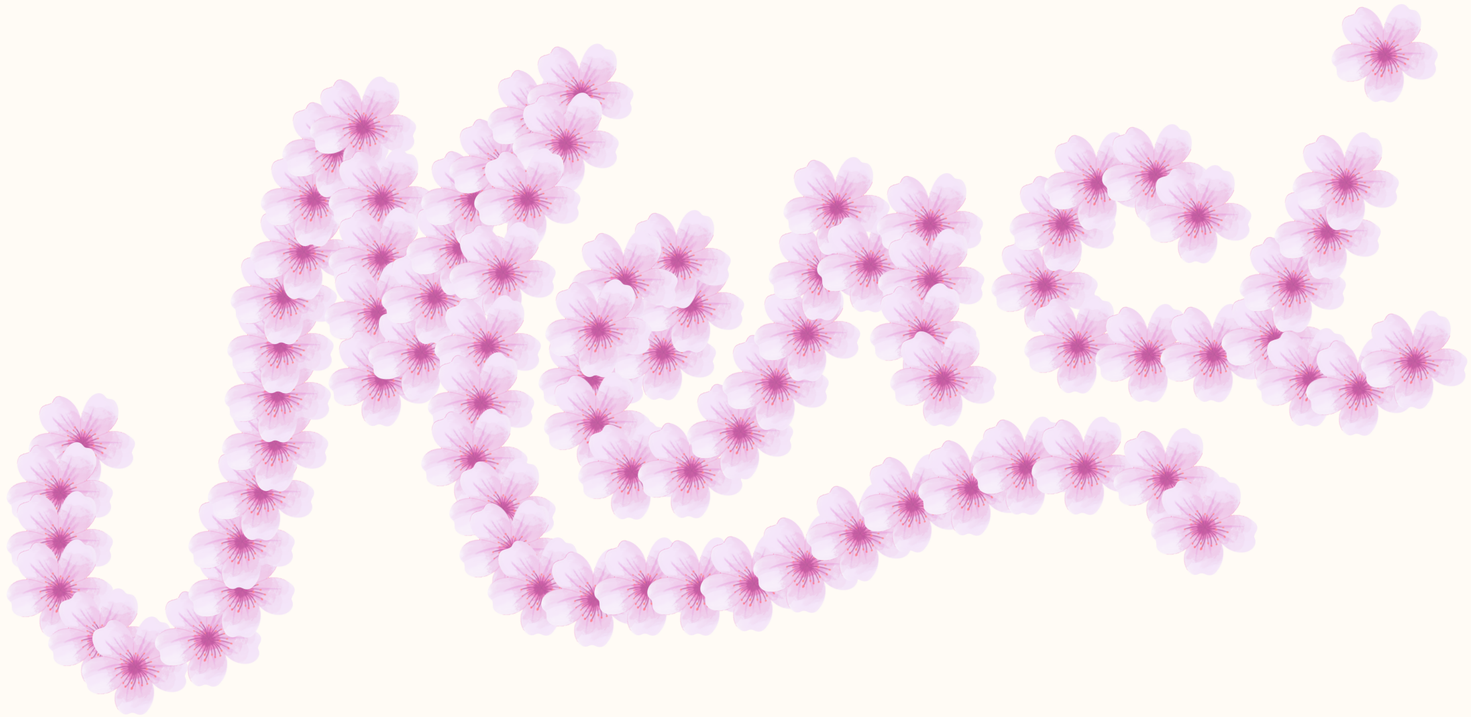
Score > quantile

→ exclusion

Etape 5

Daphne striata Tratt.	0,7547
Daphne cneorum L.	0,2100

Etape 6 → Pl@ntNet


Heal

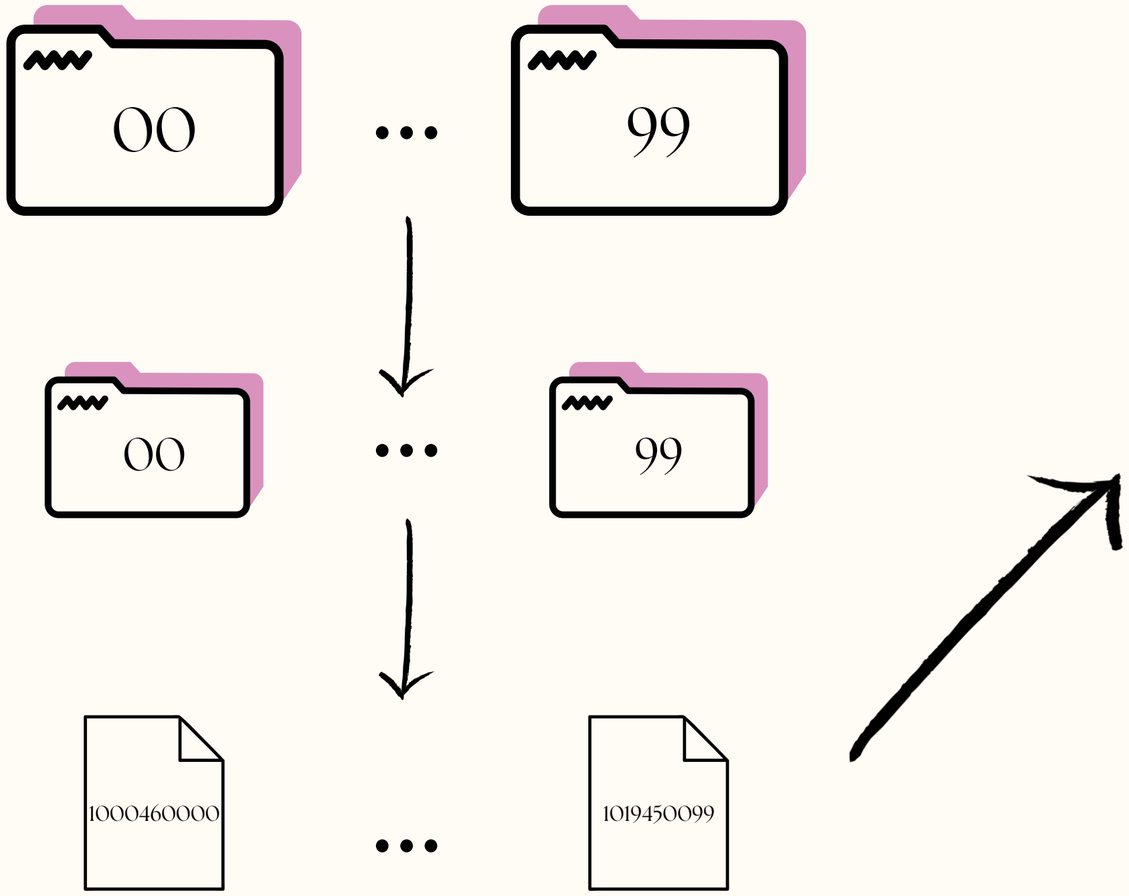
Questions

Tableau des couvertures avec les différentes graines

Score	Calibration	Graine 42	Graine 123	Graine 545	Ecart-max
Softmax	Non expertes	96,45%	96,52%	96,66%	0,21%
	Expertes	94,82%	94,95%	95,15%	0,33%
Cumulatif APS	Non expertes	96,45%	96,42%	96,66%	0,21%
	Expertes	94,76%	95,01%	95,24%	0,48%

Questions

Organisation des dossiers

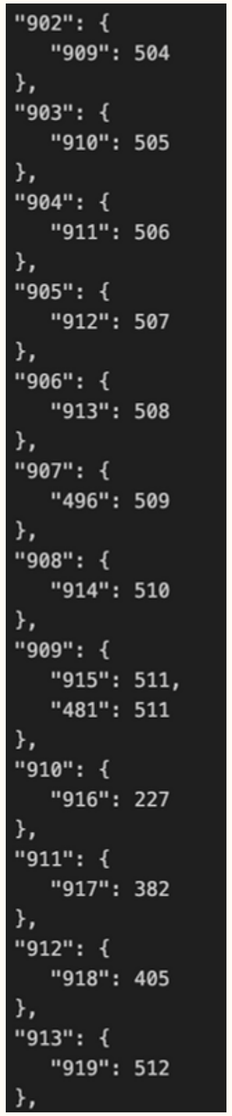


Contenu d'un fichier brut


```
{
    "status": "OK",
    "results": [
        {
            "name": "Cotoneaster franchetii Bois",
            "id": "1391408",
            "score": 0.88048
        },
        {
            "name": "Cotoneaster pannosus Franch.",
            "id": "1391394",
            "score": 0.03945
        },
        {
            "name": "Cotoneaster integerrimus Medik.",
            "id": "1391411",
            "score": 0.00761
        },
        {
            "name": "Cotoneaster dielsianus E.Pritz.",
            "id": "1666972",
            "score": 0.00481
        },
        {
            "name": "Cotoneaster simonsii Baker",
            "id": "1391419",
            "score": 0.00413
        },
        {
            "name": "Cotoneaster coriaceus Franch.",
            "id": "1390095",
            "score": 0.00224
        }
    ]
}
```

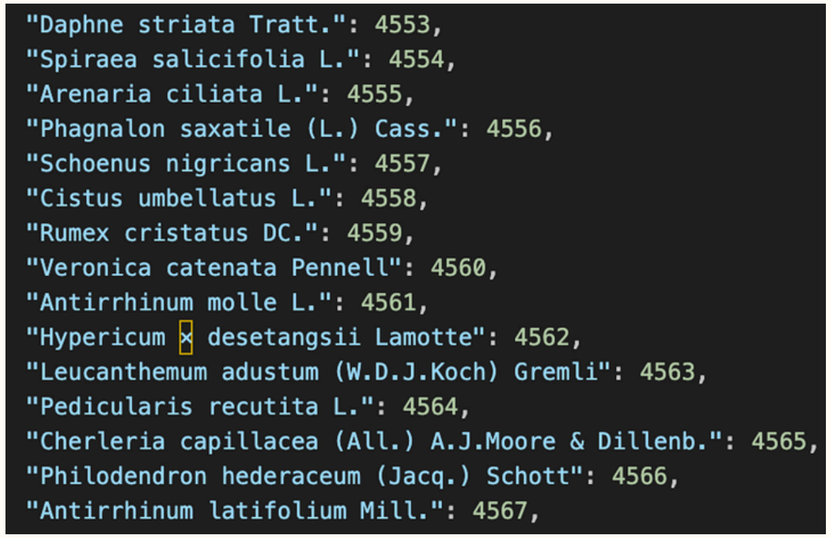
Questions

Votes des utilisateurs



```
"902": {
   "909": 504
},
"903": {
   "910": 505
},
"904": {
   "911": 506
},
"905": {
   "912": 507
},
"906": {
   "913": 508
},
"907": {
   "496": 509
},
"908": {
   "914": 510
},
"909": {
   "915": 511,
   "481": 511
},
"910": {
   "916": 227
},
"911": {
   "917": 382
},
"912": {
   "918": 405
},
"913": {
   "919": 512
},
```

Correspondances noms scientifiques - identifiants



```
"Daphne striata Tratt.": 4553,
"Spiraea salicifolia L.": 4554,
"Arenaria ciliata L.": 4555,
"Phagnalon saxatile (L.) Cass.": 4556,
"Schoenus nigricans L.": 4557,
"Cistus umbellatus L.": 4558,
"Rumex cristatus DC.": 4559,
"Veronica catenata Pennell": 4560,
"Antirrhinum molle L.": 4561,
"Hypericum × desetangsii Lamotte": 4562,
"Leucanthemum adustum (W.D.J.Koch) Gremli": 4563,
"Pedicularis recutita L.": 4564,
"Cherleria capillacea (All.) A.J.Moore & Dillenb.": 4565,
"Philodendron hederaceum (Jacq.) Schott": 4566,
"Antirrhinum latifolium Mill.": 4567,
```

Fichiers finaux - après croisements

```
  {
    "name": "Chondrilla juncea L.",
    "obs": "1357492",
    "proba": 0.001,
    "id_SWE": 381,
    "correct": 0
  }
],
"1017784395": [
  {
    "name": "Daphne cneorum L.",
    "obs": "1391649",
    "proba": 0.9656,
    "id_SWE": 2443,
    "correct": 1
  },
  {
    "name": "Daphne striata Tratt.",
    "obs": "1391654",
    "proba": 0.0046,
    "id_SWE": 4553,
    "correct": 0
  },
  {
    "name": "Daphne petraea Leyb.",
    "obs": "1420793",
    "proba": 0.0011,
    "id_SWE": 18748,
    "correct": 0
  }
],
"1017794395": [
  {
    "name": "Myosotis sylvatica Hoffm.",
    "obs": "1397441",
    "proba": 0.2239,
    "id_SWE": 571,
    "correct": 0
  },
```