

# Analyse des modèles de typologie

Approche avec des classes "weighted"

# A. Construction du poids des classes

1. Approche par seuil
2. Approche  $n\_samples / (n\_classes * np.bincount(y))$
3. Approche  $F(n\_samples / (n\_classes * np.bincount(y)))$ 
  1.  $F(x) = \log(x+300)$
  2.  $F(x) = x^{**}(1/2)$
  3.  $F(x) = x^{**}(1/1.5)$
  4.  $F(x) = x^{**}(1/1.2)$

## B . Critère d'évaluation

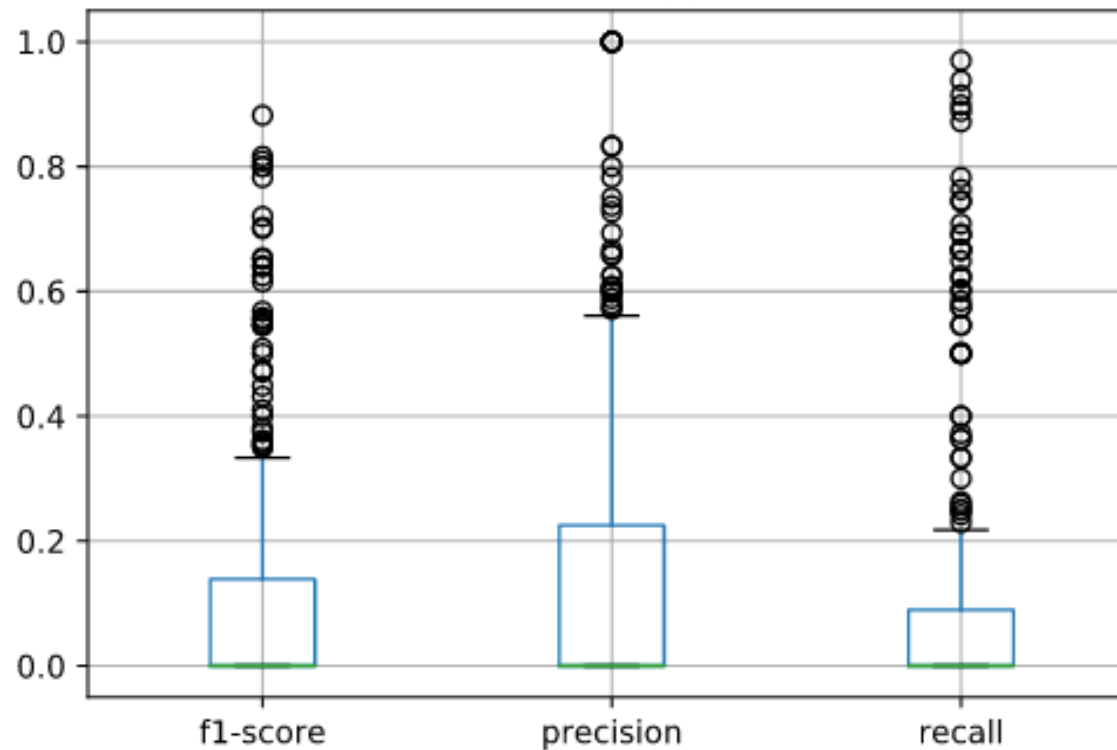
- % de classes avec un F1-score non nul et un support non nul
- Distribution f1, recall et précision sur les classes avec un support non nulles

### Contexte:

- Modèle entraîné sur 32 000 lignes
- Modèle testé sur 9000 lignes

## C. Résultat - Sans poids

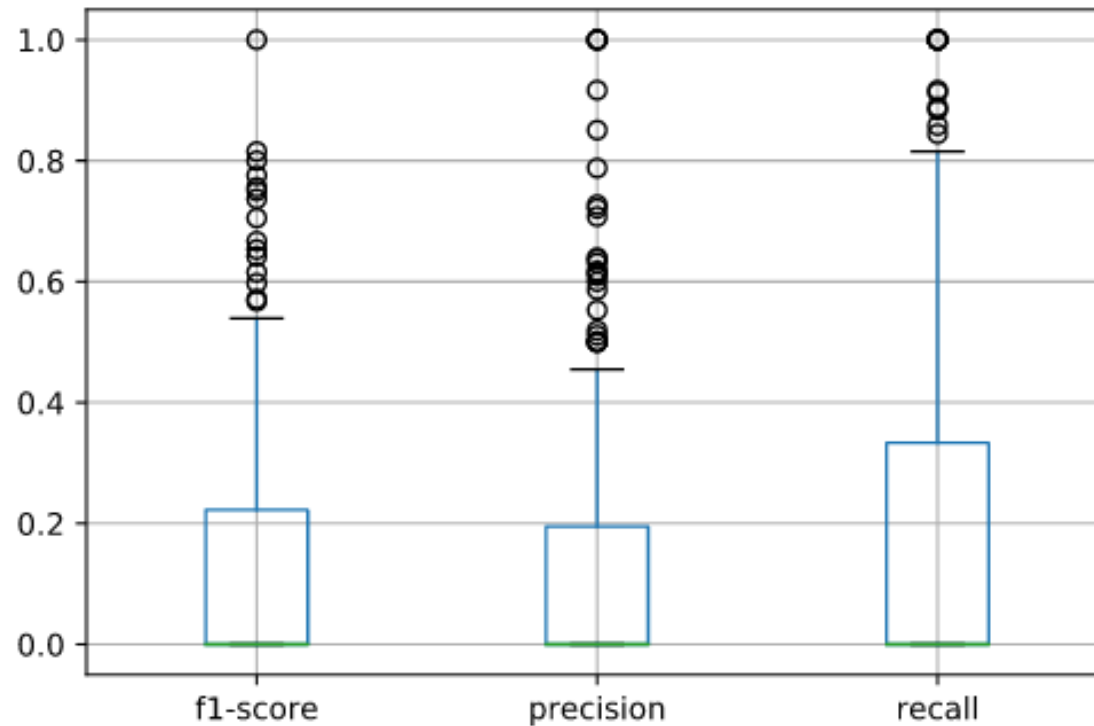
70 % Classes avec un f1 nul et un support non nul



Distribution f1, précision, recall, pour les classes avec un support non nul

## C. Résultat - Seuil Naïf

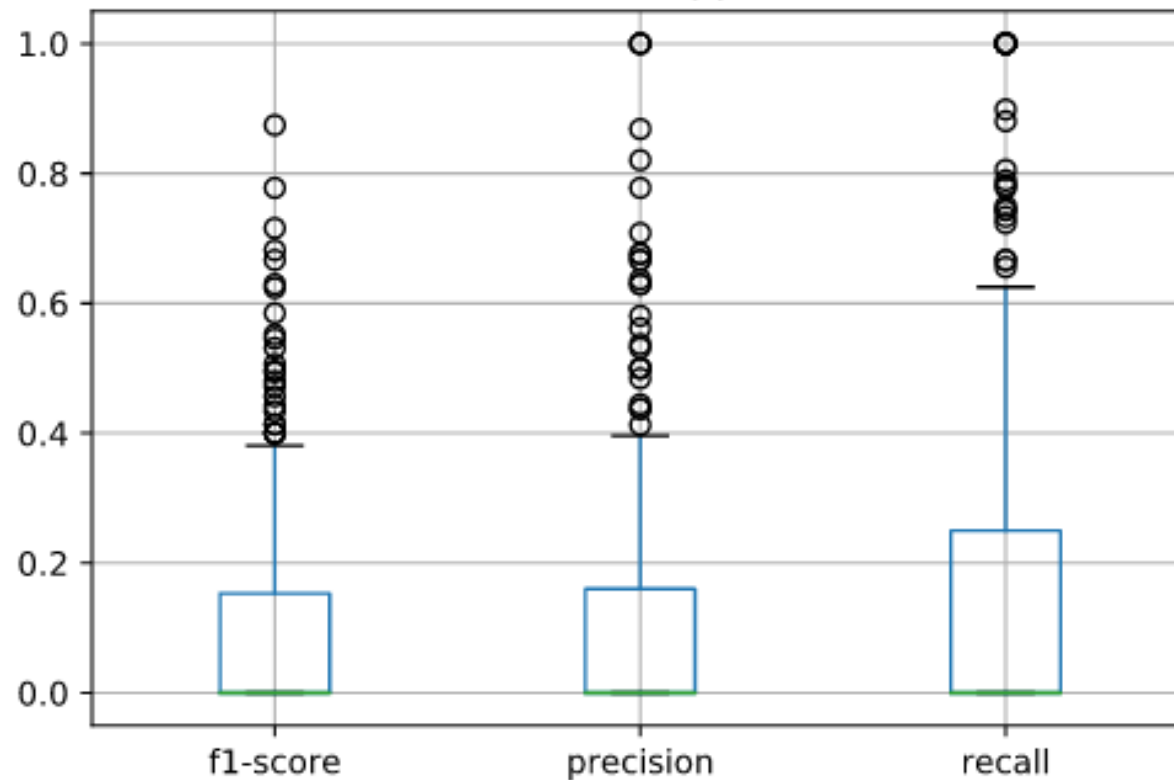
56 % Classes avec un f1 nul et un support non nul



Distribution f1, précision, recall, pour les classes avec un support non nul

## C. Résultat - approche $n\_samples / (n\_classes * np.bincount(y))$

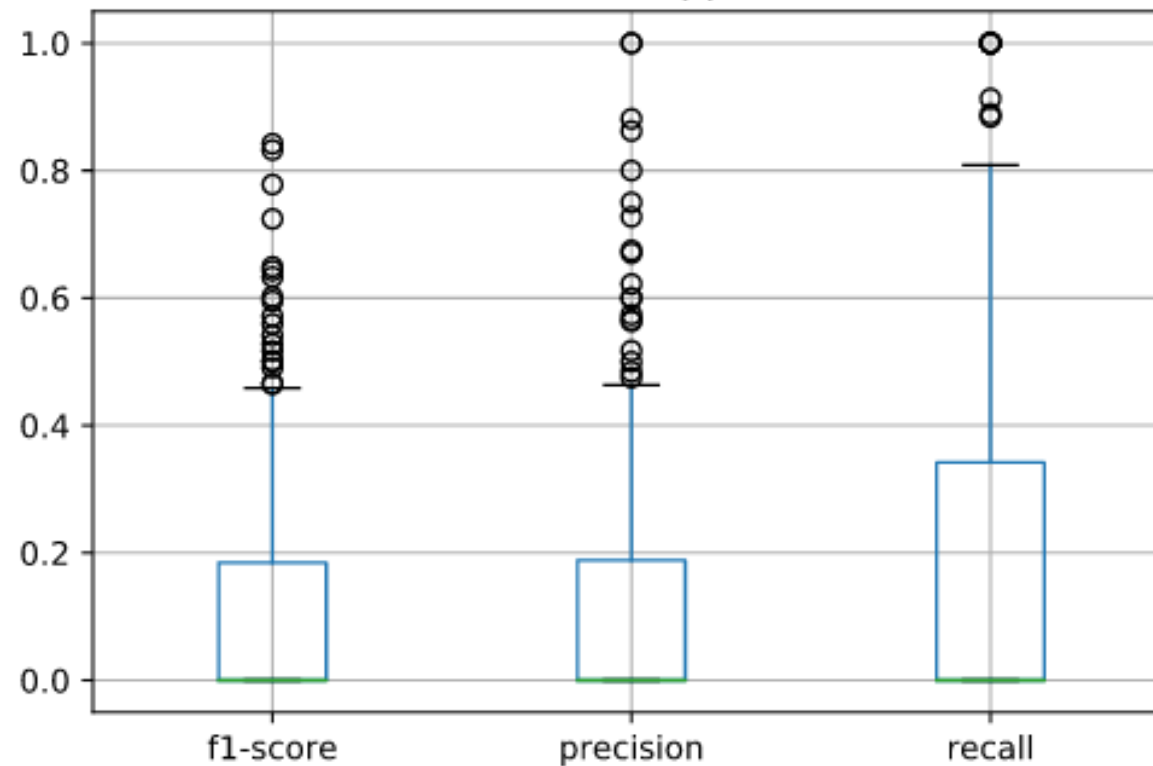
62 % Classes avec un f1 nul et un support non nul



Distribution f1, précision, recall, pour les classes avec un support non nul

## C. Résultat - approche $f(n\_samples / (n\_classes * np.bincount(y)))$

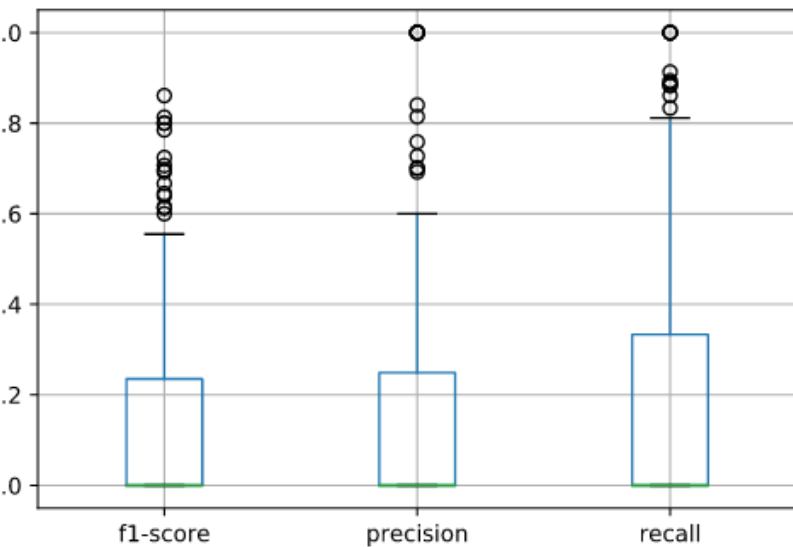
70 % Classes avec un f1 nul et un support non nul,  $F(x) = \log(x+300)$



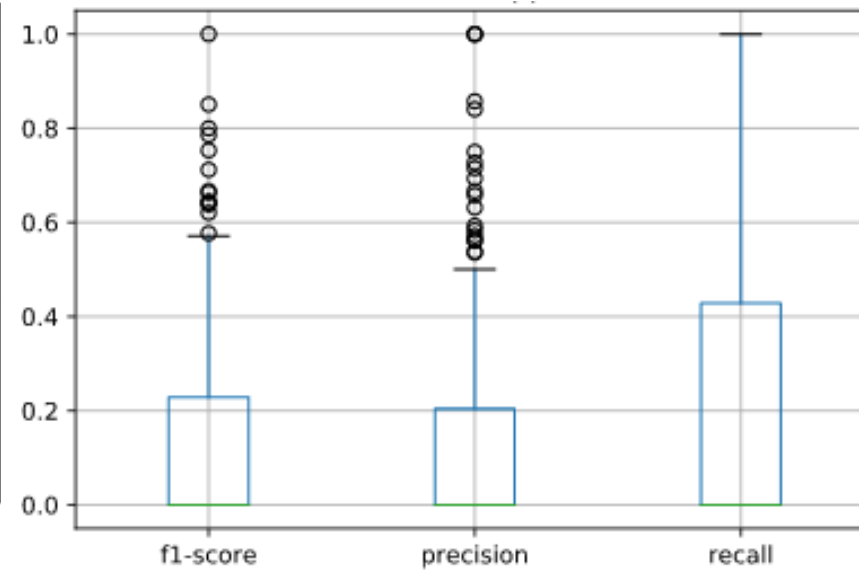
Distribution f1, précision, recall, pour les classes avec un support non nul

# C. Résultat - approche $F(n\_samples / (n\_classes * np.bincount(y)))$

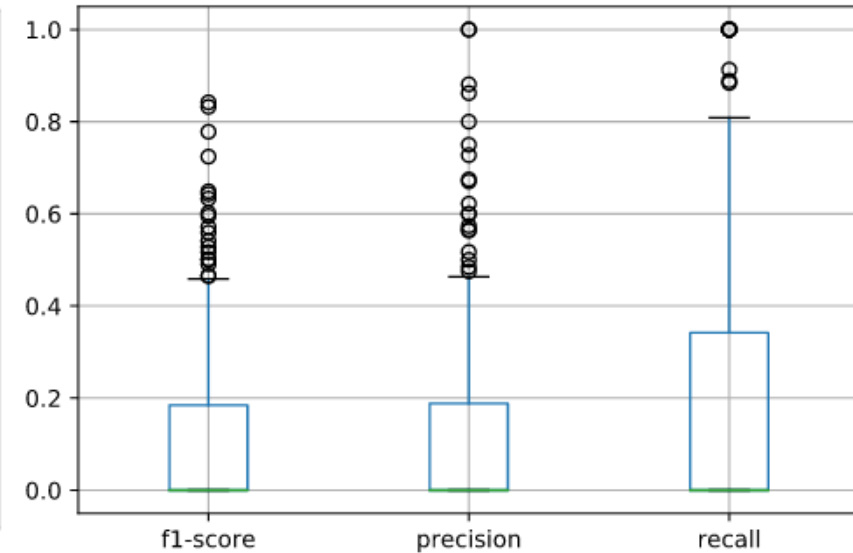
57 %,  $F(x) = x^{**}(1/2)$



54 %,  $F(x) = x^{**}(1/1.5)$



58 %,  $F(x) = x^{**}(1/1.2)$



Distribution f1, précision, recall, pour les classes avec un support non nul



# Conclusion

Nous voulons minimiser le % de classe avec un f1 nul et maximiser le f1 des classes avec un support non nul.

L'approche retenue est donc  $F(x) = x^{1/2}$  sur les poids construit par `n_samples / (n_classes * np.bincount())`