



Pénélope GITTOS



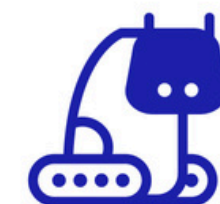
Business Developer

Training & Certification
:proabl.

Introduction au machine learning avec



:proabl.



: probabl.

Probabl est la spinoff d'Inria, l'Institut national de recherche en sciences et technologies du numérique, qui détient à présent la marque scikit-learn.

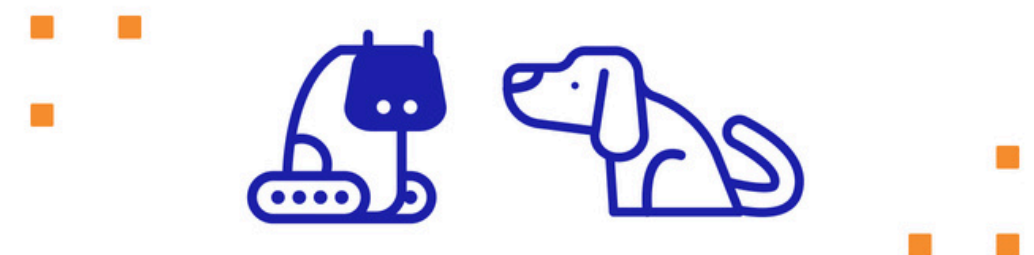
produit (WIP) pour les
data scientists

skore

certifications scikit-learn



: probabl.



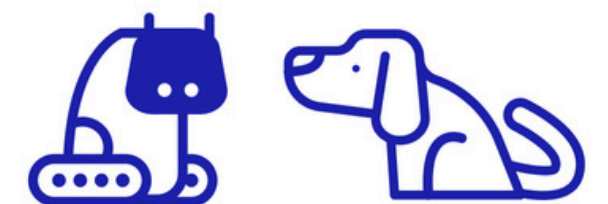
Quelques statistiques intéressantes...

- Plus de 80 millions de téléchargements par mois.
- Son rôle de middleware : fondamental pour près d'un million de projets et plus de 15 000 packages structurés.
- Élément central de l'écosystème qui stimule l'innovation en machine learning et en data science.



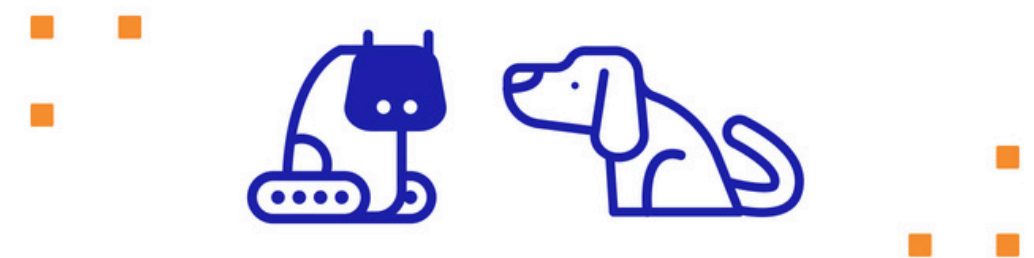
C'est la boîte à outils open source de ML la plus utilisée au monde par les Data Scientist.

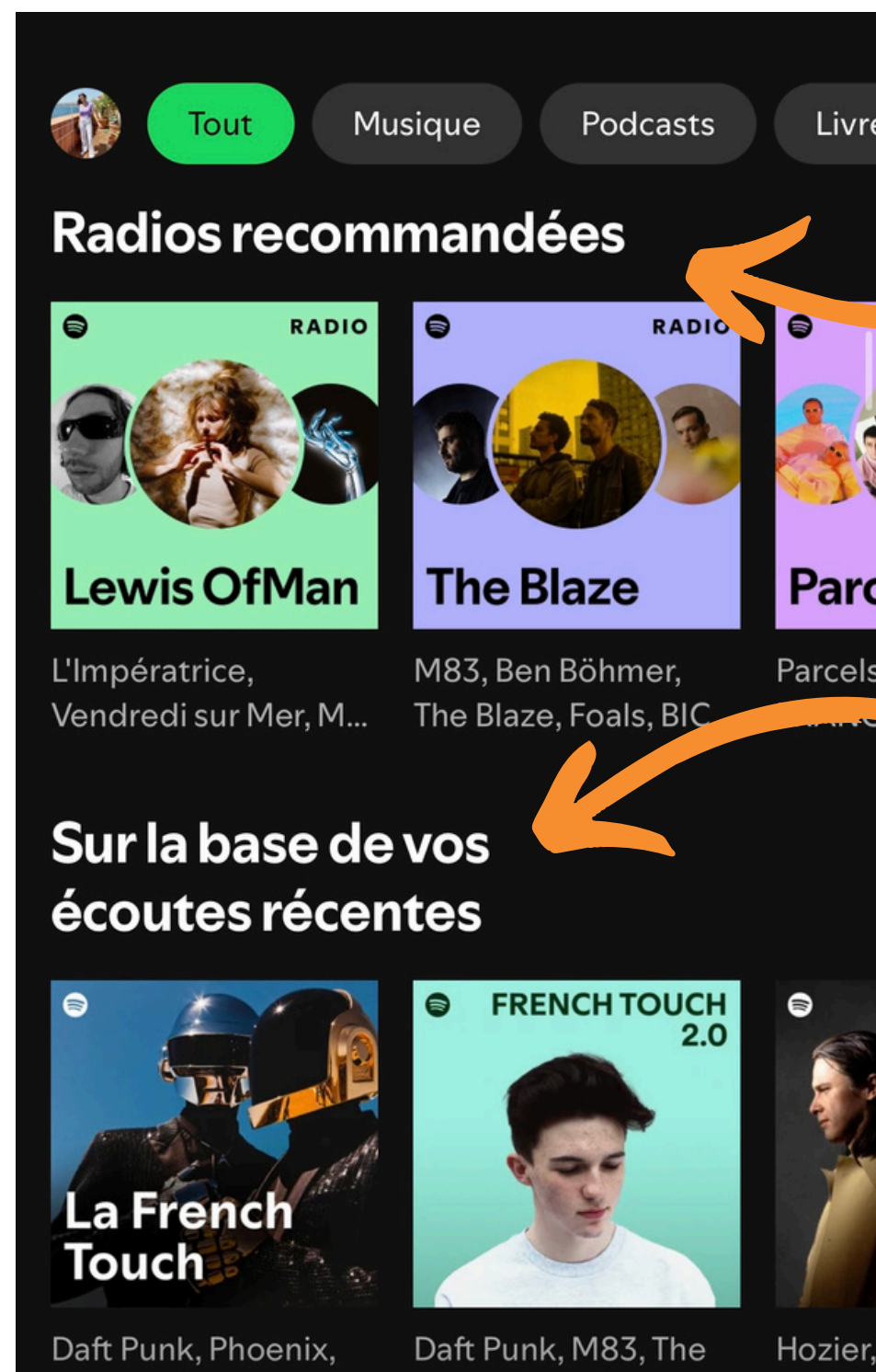
:probabl.



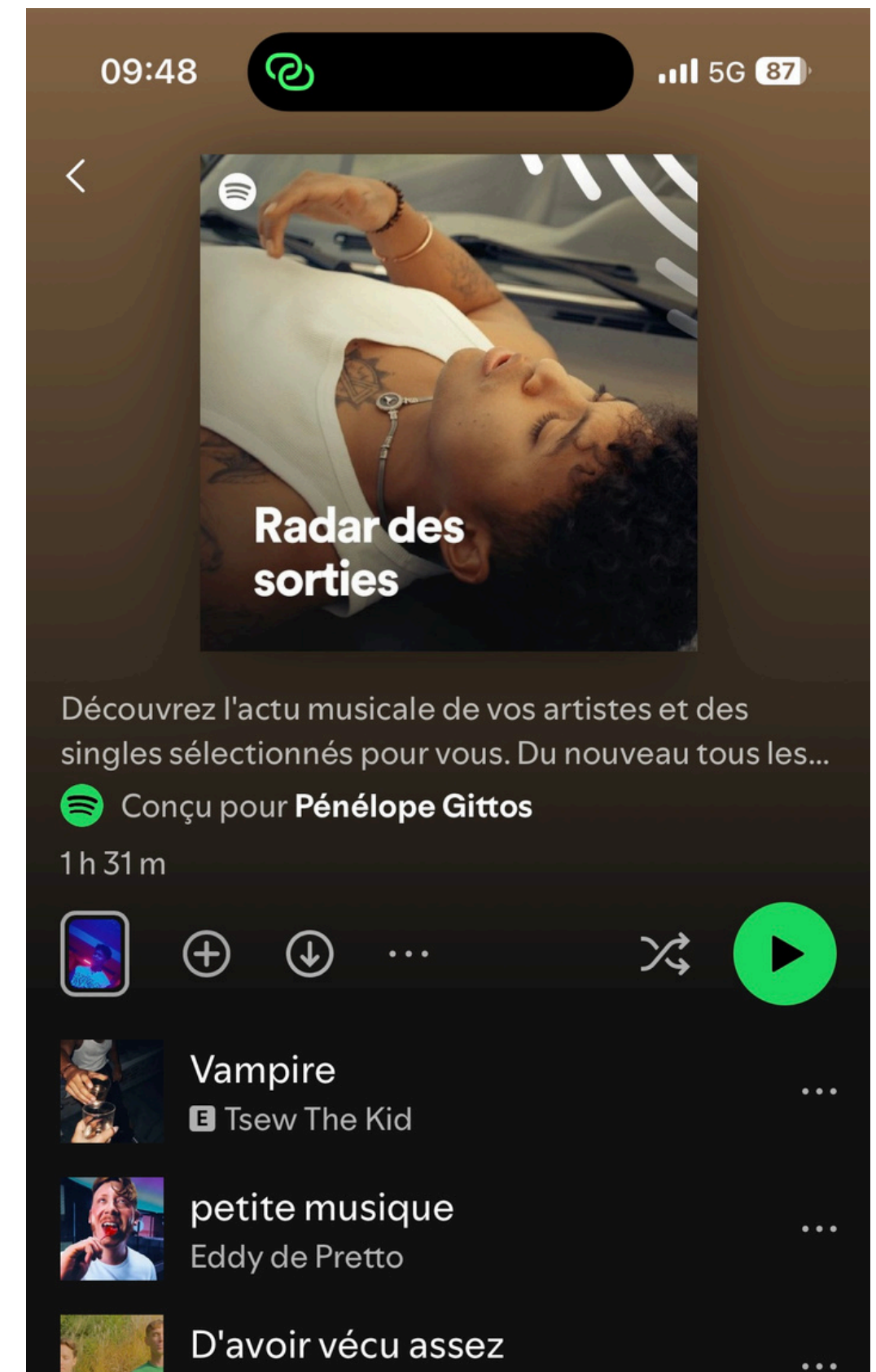
Scikit-learn permet aux ordinateurs d'effectuer des tâches telles que:

- la segmentation client (par tranche d'âge, préférences de consommation)
- la détection de fraude (bancaires)
- les prédictions de consommation
- créer un système de recommandation

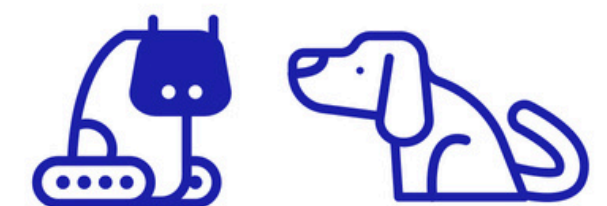




Recommandations
personnalisées



:probabl.



Mais comment ça marche ?

1

Collecte des données sur mes choix passés (artistes, genres...)

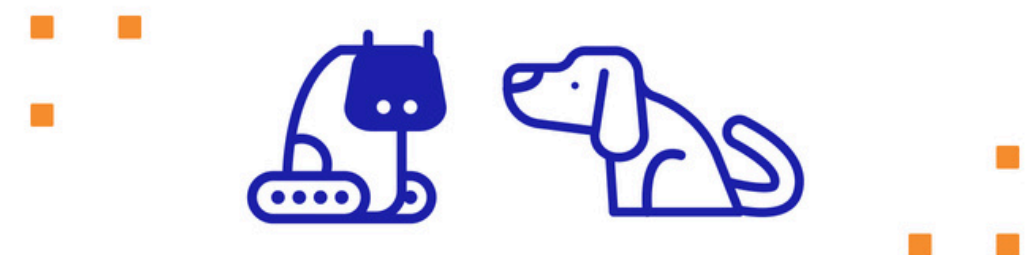
2

La machine s'entraîne à comprendre mes goûts et va établir des modèles qui représentent mes préférences

3

Une fois le modèle entraîné, la machine peut prédire ce qui pourrait me plaire.

:probabl.





Apprentissage supervisé :

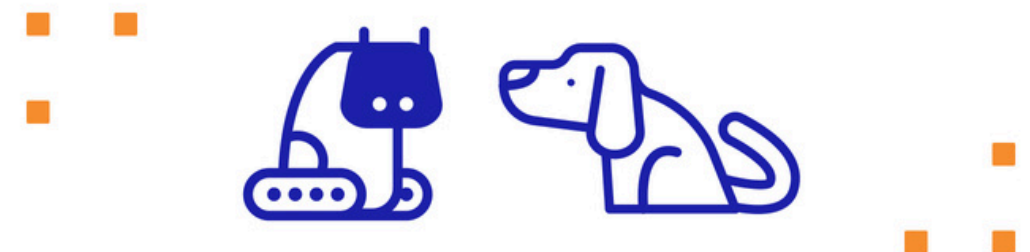
Le modèle apprend à partir de données étiquetées

Apprentissage non supervisé :

Le modèle explore des données non étiquetées et trouve des structures cachées



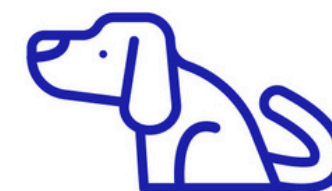
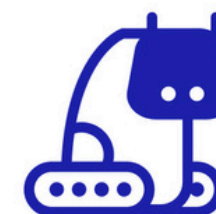
:probabl.

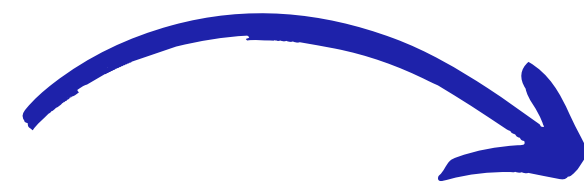




Démo

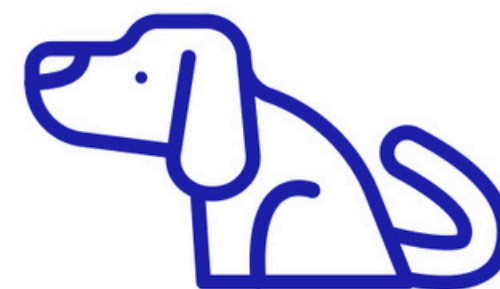
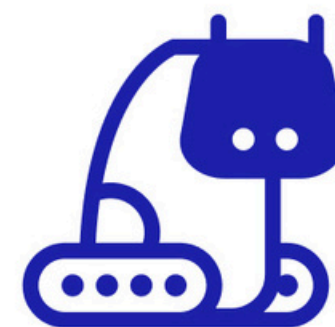
:probabl.





Pourquoi ?

- qualité des données
- variété des formats
- volume plus important
- dépendance et variabilité



:probabl.