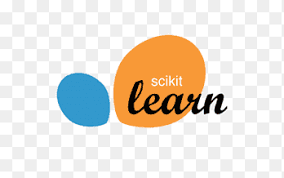
** Résumé Machine-Learning**

Introduction

Le machine Learning est une discipline de l’informatique (intégrée dans l’intelligence artificielle) destinée à modéliser les relations entre les données. Dans un autre domaine, on parlerait de modélisation statistique, ou de méthodes de data mining, ou encore d’analyse de données.. Des lors de nombreux librairies ont vu le jour pour faciliter son utilisation tel, que tensor flow ,python ,scikit-learn .Durant ce résumé nous allons prendre l’exemple du Scikit-learn qui est une librairie pour Python spécialisée dans le machine learning (apprentissage automatique).Ce travail s’articulera d’abord sur l’historique ,ensuite nous verrons son application sur le machine learning à savoir :l’apprentissage supervisé et l’apprentissage non supervisé.

# Historique

Scikit-learn est un point d’entrée pour essayer de comprendre le machine learning et est un projet né en 2006 ,décollé particulièrement en 2010 .Ce projet à été participé par beaucoup de personnes et massivement drivé par des laboratoires de recherches en terme de réalisation. Environ 400.0000 utilisateurs ont visionnées le projet dans un mois .

Cependant il ya deux (2) façons de faire du machine learning ,maintenant nous pouvons avoir un algorithme qui va correspondre à ces 2 apprentissage qui va être utilisé le plus adéquate selon les cas:

L’apprentissage supervisé et l’apprentissage non supervisé

## L’Apprentissage Supervisé

L’apprentissage supervisé est le fait de faire de la**classification,** de la **régression** et de la **ranking.**

L’apprentissage supervisé est un problème de prédicat :Il regroupe

* la classification

la classification: est principalement de définir des règles permettant de classer des objets dans des classes partir de variables qualitatives ou quantitatives caractérisant ces objets. Cependant nous avons quelques algorithmes qu’il utilise à savoir le SVM,

* la regression:La régression linéaire est l’un des algorithmes d’apprentissage supervisé les plus populaires.

Il est aussi simple et parmi les mieux compris en statistique et en apprentissage automatique.

* la ranking :il vise a donné d'abord un model ou les données seront entrées par l'utilisateur en prenant compte les paramètres sur les modèles .En terme d'API il existe un paramètre appelée param2="auto" qui va permettre de sélectionner les données entrées parmi les données prévus pour le param=le-8 .Apres nous avons le model.fit ,le model.predict donne le paramètre de prédiction de test model.score permet de donner le x\_test ,y\_test et fait la classification ,pour permettre de garder la régression.

## Apprentissage non supervisé

Quand on fait la non supervisé c'est la même chose ,mais juste nous avons pas la variable y .C'est à dire nous avons pas à prédire .Autrement dit l'apprentissage non supervisé permet de transformé des données ,de réduire la dimension ,la détection d'anomalie (sa centre dans ce domaine ),par ensemble nous avons des données ,il nous permet de dire ceux qui est anormale.

La non supervisé nous avons plein d'algorithmes pour faire du clustering parmi ces algorithmes on peut citez:

K-mean ou où l'on divise des observations en k partitions.

DBSCAN /BIRCH,PCA / ICA / NMF(Non negative factoriel)

LDA qui est très utilisé pour faire du topic mapping c'est à dire du texte

En terme du reduction de dimension qui sert à comprendre la structure des données ,pour cela nous avons l’algorithme :

* PCA permet de projeter sur des axes en utilisant l’algèbre linéaire
* TSNE

En terme de licence scikit c'est une licence BSD ou on peut l'utiliser non string attache, un produit commercial .En terme de demo il utilise python /numpy l'algèbre linéaire pour manipuler les tableaux de donner ,SciPy/Cython qui permet de typer du code python.

Enfin nous allons parler de clustering qui est la classification non supervisé; il consiste a donné la table sql et de trouvé tous les lignes qui sont a peu près similaires ou des paquets d'échantillon qui sont proches ,et ou de mettre les un des autres dans des groupes qui sont s'y proches .

Conclusion

Inutile de dire que SkLearn a certainement laissé sa marque sur l'apprentissage automatique, Python, l'informatique, ainsi que sur l'automatisation, et il est très peu probable qu'il aille de sitôt. SkLearn est le seul package d'apprentissage automatique parmi les cent meilleurs packages Python, et se classe de manière assez significative lorsque les chiffres sont mis en perspective.