



---

UNIVERSITÉ PARIS 8 - VINCENNES À SAINT-DENIS

Master Informatique des Systèmes Embarqués

Big-Data

Predict-Price-hotel-Machine-Learning

Aimen GANA

Hammou SABBAR

Université Paris VIII : Rakia JAZIRI

Date de rendu : le 06/01/2019

# Projet : Predict-Price-hotel-Machine-Learning

Notre projet consiste à prédire la catégorie des prix des hôtels (cher, moyen, pas cher) en Île-de-France, à partir d'apprentissage de Machine Learning avec Sklearn, Scikit-learn est une bibliothèque libre Python destinée à l'apprentissage automatique. Elle comprend notamment des fonctions pour estimer des forêts aléatoires, des régressions logistiques, des algorithmes de classification, et les machines à vecteurs de support. Nous ce qui va nous intéressé c'est les algorithmes de classification dans ce projet.

Elle est développée par de nombreux contributeurs notamment dans le monde académique par des instituts français d'enseignement supérieur et de recherche comme Inria et Télécom ParisTech.

Nous sommes basés sur ces Statistiques suivantes :

- Les prix
- Les avis
- Nombre des étoiles d'hôtel

Nous avons récupéré les données sur le site Tripadvisor les noms des hôtels, localisation, les prix, les avis, services proposés, dans un fichier csv, puis nous avons remarqué que les noms et les localisations pouvaient pas être traités par machine learning en format String, nous pouvons trouver une autre présentation sauf nous ne disposons pas d'un temps suffisant. Pour l'apprentissage nous basons sur 9 critères les prix, nombre d'étoile et les avis (Excellent, Très bon, Moyen, Médiocre, Horrible et le nombre des commentateurs)

Nous avons utilisé l'algorithme KNN que nous avons vu durant le cours pour faire la prédiction sur la classe des hôtels, nous avons utilisé 80