Azure Machine Learning

Présentation de l'outil de data mining / statistique

Julie Chalon - Chloé Devin - Maïté Garcia - Jean-louis Phimmasane

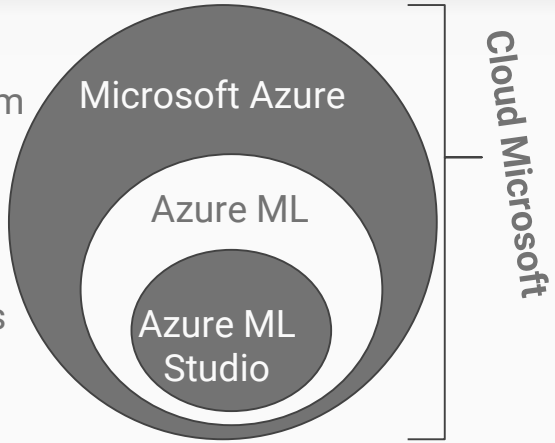
M2 SISE - 25 octobre 2016

- Qu’est ce c’est ?
- Lancement
- Fonctionnalités
- Méthodes de data mining
- Gestion des données
- Avantages vs inconvénients
- Positionnement
- Evaluation critique

- Qu’est ce c’est ?
- Lancement
- Fonctionnalités
- Méthodes de data mining
- Gestion des données
- Avantages vs inconvénients
- Positionnement
- Evaluation critique

Qu’est ce c’est ?

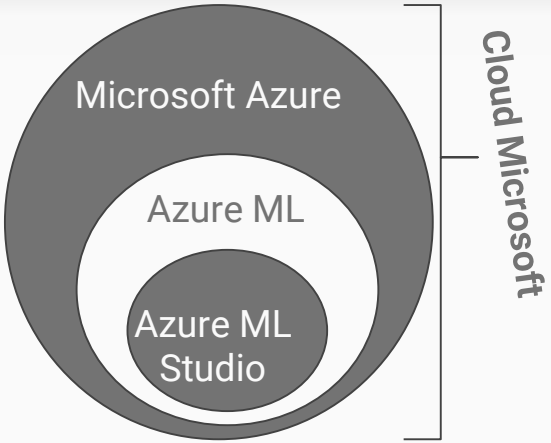
- Pour la petite histoire …
  - 2008 : service could annoncé sous le nom de Windows Azure Platform
  - 2014 :
    - Projet renommé **Microsoft Azure**
    - Panne affectant certains clients
    - Gestion des domaines défaillante à cause d’un gros volume des transactions
  - 2016 : accusé par la Cnil d’une collecte de données excessive sans satisfaire une confidentialité suffisante



Concept de « cloud computing » ou informatique en nuage :
- offre d'hébergement (applications et données)
- services (stockage et synchronisation des données)

Qu’est ce c’est ?

- Azure Machine Learning :
  - Service dédié à l’analyse prédictive
  - Outil collaboratif pour créer, déployer et partager des solutions dans le cloud
- Azure Machine Learning Studio :
  - Environnement d’essai
  - Analyse de données sous forme d’expérience
  - Palette de modules et interface visuelle
  - Front end
  - Service web



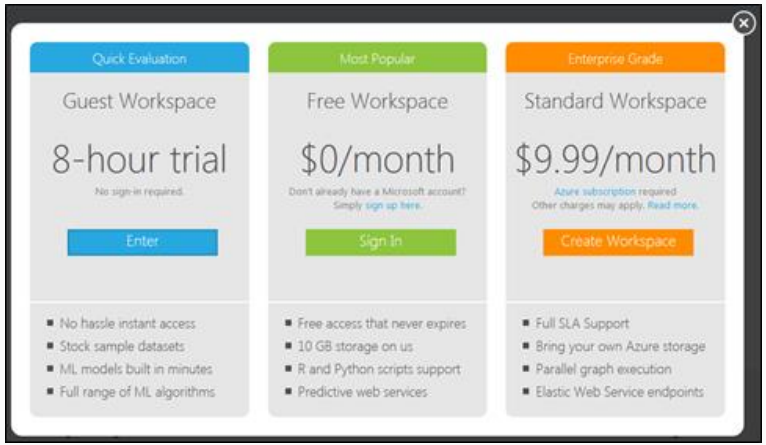
- Qu’est ce c’est ?
- Lancement
- Fonctionnalités
- Méthodes de data mining
- Gestion des données
- Avantages vs inconvénients
- Positionnement
- Evaluation critique

Lancement du logiciel

1. Se rendre sur le site de **Microsoft Azure Machine Learning Studio** :

➡ https://studio.azureml.net/

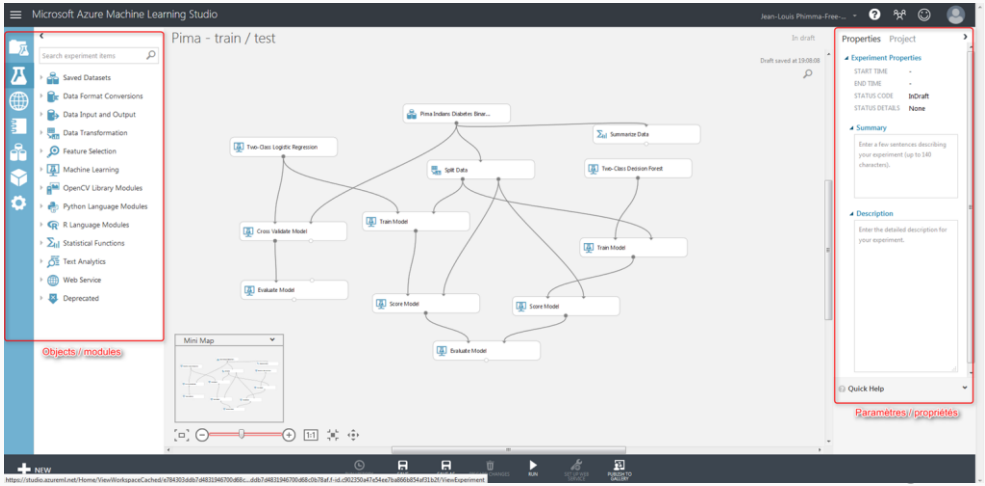
2. Sélection du mode d'utilisation
3. S'identifier avec un compte Hotmail ou Live



- Qu’est ce c’est ?
- Lancement
- Fonctionnalités
- Méthodes de data mining
- Gestion des données
- Avantages vs inconvénients
- Positionnement
- Evaluation critique

Fonctionnalités

- Création de projets et d'expériences
- Manipulation des données
- Différentes études de statistique / data mining
  - Exploration des données
  - Analyse descriptive
  - Analyse prédictive
- Méthodes apprentissage
  - Supervisé
  - Non supervisé
- Evaluation des modèles
- Mettre en service (web) / gérer



- Qu’est ce c’est ?
- Lancement
- Fonctionnalités
- Méthodes de data mining
- Gestion des données
- Avantages vs inconvénients
- Positionnement
- Evaluation critique

Méthodes de data mining

- De nombreuses méthodes de Data Mining sont proposées et sont souvent complémentaires.

Statistique descriptive	✓
Analyse factorielle ACP	✓
Autres analyses factorielles (AFC, ACM)	✗
Classification : méthode des K-means	✓
Classification : méthode de la CAH	✗
Régression	✓
Analyse prédictive et Scoring (Régression logistique, Arbres de décision, réseaux de neurones,…)	✓
Text mining	✓

- Qu’est ce c’est ?
- Lancement
- Fonctionnalités
- Méthodes de data mining
- Gestion des données
- Avantages vs inconvénients
- Positionnement
- Evaluation critique

Gestion des données

- Gestion de données volumineuses
  - Exploration des données
  - 10 Go de stockage alloué sur le cloud
  - Bibliothèque d’algorithmes
- Lenteur d'exécution
  - Calcul - déduction au travers des serveurs du cloud
- Stockage en ligne sur le cloud
  - Projets, expériences, données

- Qu’est ce c’est ?
- Lancement
- Fonctionnalités
- Méthodes de data mining
- Gestion des données
- Avantages vs inconvénients
- Positionnement
- Evaluation critique

Avantages vs inconvénients

+

- Outil collaboratif
- Cloud - Stockage en ligne
- Intègre R et Python
- Mise à jour automatique
- Ergonomique
- Intuitif

-

- Lent
- Visuel des arbres de décision
- Importation de fichiers Excel
- Dépendance à Internet
- Sécurité

- Qu’est ce c’est ?
- Lancement
- Fonctionnalités
- Méthodes de data mining
- Gestion des données
- Avantages vs inconvénients
- Positionnement
- Evaluation critique

Positionnement par rapport à d'autres outils

	Azure ML	Amazon ML	Google Prediction	Big ML
Export de modèle	-	-	-	X
Import de modèle	-	-	X	-
Classification	X	X	X	X
Régression logistique	X	X	X	X
Clustering	X	-	-	-
Arbres de décision	X	-	-	X
Réseau de neurones	X	-	-	X
Séparation app./test	X	X	-	X
Transformations maths	X	-	X	X
Normalisation	X	X	-	X

- Qu’est ce c’est ?
- Lancement
- Fonctionnalités
- Méthodes de data mining
- Gestion des données
- Avantages vs inconvénients
- Positionnement
- Evaluation critique

Evaluation critique

- Innovation :
  - Service could
- Utilisation qui convient plus au milieu scolaire ou encore aux petites structures :
  - Prix abordable
  - Problématique de sécurité et confidentialité des données
- Reste à tester la version payante du logiciel :
  - Y’a t-il plus de fonctionnalités ?

Merci pour votre attention :)

Rejoindre le groupe sur le BV pour le TD :
azure_machine_learning