

## Liste des Sujets

### Sujet1 : La théorie de machine Learning et ces applications.

1. Introduction <https://www.toptal.com/machine-learning/machine-learning-theory-an-introductory-primer>
2. Types: Supervised, semi supervised, Unsupervised.
3. Différentes techniques (regression, Classification ...)
4. Exemples algorithms.
5. Application de machine Learning.
6. Conclusion.

### Sujet 2: Data Préparation for machine Learning.

1. Introduction (définition et pourquoi c'est important de préparer les données pour Les algorithmes de machine Learning).
2. Les étapes de préparation des données. (<https://developers.google.com/machine-learning/data-prep> )
  - 2.1.1. Collection des données. (Taille, intégration des données, Labels...)
  - 2.1.2. Data Cleansing. (Les différentes techniques : missing values ...)
  - 2.1.3. Transformation des données (Transformation numérique et catégoriale, normalisation ...).
3. Exemples.
4. Conclusion.

### Sujet3 : Hadoop/ MapReduce.

1. Introduction (présentation de Hadoop et ces Modules (Solutions pour Big Data) en se basant sur MapReduce).
2. Fonctionnement de MapReduce.
3. Algorithme + exemple ( <https://hadoop.apache.org/docs/current/hadoop-mapreduce-client/hadoop-mapreduce-client-core/MapReduceTutorial.html> ).
4. Avantages / Inconvénients.
5. Conclusion.

### Sujet4 : Intégration des données (Cloud Data intégration).

1. Introduction.
2. Avantages / Inconvénients.
3. Type de Cloud intégration
4. Plateformes et outils
5. Open Source outils (présentation + comparaison)

### Sujet5 : Réduction de la dimensionnalité en utilisant Machine Learning.

1. Introduction (Quoi ? et Pourquoi utiliser réduction de dimensionnalité)
2. Différentes Techniques de réduction de dimensionnalité.  
<https://thenewstack.io/3-new-techniques-for-data-dimensionality-reduction-in-machine-learning/>,  
<https://www.ritchieng.com/machine-learning-dimensionality-reduction/>  
<https://elitedatascience.com/dimensionality-reduction-algorithms>
3. Avantages/ Inconvénients.
4. Conclusion

### Sujet6 : Apach Spark

1. Introduction

2. Pourquoi utiliser Spark?
3. Le stack e spark:
  - 3.1.1. Spark Core.
  - 3.1.2. Spark Streaming.
  - 3.1.3. Spark SQL.
  - 3.1.4. MLIB (Machine learning Library).
  - 3.1.5. GraphX.
4. Exemples d'utilisation.
5. Conclusion.

## Sujet 7 : Data Mapping

1. Introduction
2. Techniques de Data Mapping. (Manual Data Mapping, Semi-Automated Data Mapping, Automated Data Mapping).
3. Outils de Data Mapping.
4. Etapes de Data Mapping (Définir, Map the Data, Transformation, Test, Déploiement, Maintenir et Mise à jour) <https://www.talend.com/resources/data-mapping/>
5. Avantages / Inconvénients.
6. Conclusion.

## Sujet 8 : Data Warehouse

1. Introduction (Définition + Architecture). <https://www.tutorialspoint.com/dwh/index.htm>
2. Processus de Data Warehouse (DWH).
3. Caractéristiques de Data Warehouse, et son utilité et avantages.
4. Types et application de data Warehouse.
5. Les Schema de Data Warehouse
6. OLAP ...
7. Conclusion

## Sujet 9 : Natural Langage Processing (NLP) et Big data

1. Introduction (NLP ?). <https://www.xenonstack.com/blog/artificial-intelligence/>
2. Difference entre NLP et Text Mining.
3. Le process de NLP (Compréhension et Génération).
4. Les Technique et etapes de NLP processus (Sentence Segmentation, Tokenization, Stemming/Lemmatization, Part-of-Speech tagging, Parsing, Named Entity Recognition, Coreference resolution)
5. Deep Learning pour NLP.
6. Exemple d'application: ChatBot(Agents intelligents automatisés)
  - 6.1.1. Definition.
  - 6.1.2. Fonctionnement.
7. Conclusion

## Sujet 10 : OpenStack

1. Introduction (Quoi ? Historique + distributions récentes, Utilisateurs d'OpenStack.)
2. Architecture
3. Différents Modules et composants (il existe 21 modules dans OpenStack : Nova, Trove, Glance, Neutron, Mistral, Sahara, Keystone, Cinder, Horizon ...) avec définition de chaque module.
4. Models de déploiement.
5. Conclusion.

## **Sujet 11 :Business Intelligence (BI)**

1. Introduction
2. Fonctions du BI.
3. Types de bases de données pour BI.
4. Outils de BI.
5. Indicateurs de Mesure selon la dimension des bases de données.
6. Application de BI.
7. Conclusion.

## **Sujet 12 : Système multi-agents**

1. Introduction (Définition et domaines d'interaction)
2. Avantages/ Inconvénients.
3. Catégories et organisation des agents.
4. Architecture des systèmes Multi-agents.
5. Domaines d'application + exemple.
6. Conclusion.

## **Sujet 13 :Apache Cassandra**

1. Introduction (définition de Cassandra + No SQL).
2. Caractéristiques.
3. Langage de requête + exemple.
4. Avantages / Inconvénients.
5. Modele de données.
6. Applications.
7. Conclusion.

## **Sujet 14 : Table de hachage distribuée**

1. Introduction
2. Structure et différents designs.
3. Caractéristiques.
4. Critères d'implémentation.
5. Exemples d'applications.
6. Conclusion

## **Sujet 15 : Large Objects (LOBs)**

1. Introduction [https://docs.oracle.com/cd/B28359\\_01/appdev.111/b28393/adlob\\_design.htm#1012133](https://docs.oracle.com/cd/B28359_01/appdev.111/b28393/adlob_design.htm#1012133)
2. Avantages/ Inconvénients.
3. Types de LOBs.
4. Types de Données pour LOBs.
5. Composants et fonctionnement des LOB (avec exemples de requêtes)
6. Exemple d'application.
7. Conclusion.

## **Sujet 16 : Clustering**

1. Introduction (définition du clustering en relation avec data mining et machine Learning).
2. Types de Clustering.
3. Validation du Clustering (définition + types de validation).

4. Algorithmes de machine Learning pour le Clustering.
5. Domaine d'applications + exemple.
6. Conclusion.

## **Sujet 17 : System expert**

1. Introduction
2. Avantages/ Inconvénients.
3. Etapes de construction de system expert
4. Architecture de system expert.
5. Machine learning pour les system expert.
6. Domains application.
7. Conclusion.

## **Sujet 18: Predictive Analytics**

1. Introduction <https://blog.infodiagram.com/2019/01/big-data-presentation-appealing-diagrams-ppt.html>
2. Machine Learning et Big data pour prediction.
3. Techniques et défis.
4. Algorithmes de machine Learning utilisés.
5. Exemple d'application.
6. Conclusion.

## **Sujet 19 : Sécurité du cloud**

1. Introduction (Définition du cloud et différents composants)
2. Attaques sur le cloud.
3. Sécurité du cloud
  - 3.1.1. Sécurité des données.
  - 3.1.2. Infrastructure.
4. Problèmes de sécurité (Défis).
5. Conclusion.

## **Sujet 20 : Social data Analytics**

1. Introduction
2. Types des données, et sources.
3. Facteurs d'analyse.
4. Méthodes d'analyse
5. Machine learning pour Social Data.
6. Applications