



DK
Consultant Big Data

SYNTHESE DE COMPETENCES

❖ **Compétences techniques :**

- **Programmation :**

Python (pandas, scikit-learn, keras, matplotlib), Spark (PySpark, MLlib), Java, MATLAB, GNU Octave, R, Prolog.

- **Big Data:**

Hadoop ecosystem (Spark, Hive, Pig, Oozie, Flume), Elastic Stack (Logstash / Elasticsearch / Kibana), Splunk (connaissance de base).

- **Bases de données :**

MySQL, Cassandra (connaissance de base).

- **IDEs :**

Jupyter, PyCharm, Spyder, Eclipse, Rstudio.

- **Collaboration :**

Git, GitLab.

- **Systèmes d'exploitation :** Linux, Windows.

- **Software (autres):** Talend (intégration de données), Protégé (ontologies), JADE (Java Agent Development Framework).

❖ **Connaissances fonctionnelles :**

- **Intelligence Artificielle :** Aide à la Décision, Machine Learning, Théorie des Jeux, Systèmes Multi-Agents.

- **Gestion de projet**

- **Compétences rédactionnelles**

- **Enseignement**

❖ **Langues :**

- **Français**

- **Anglais**

- **Espagnol**

- **Grec**





EXPERIENCES PROFESSIONNELLES

Février 2018 – Février 2019 : Data Scientist / Engineer - BNP

Descriptif mission : conception & développement d'applications Data Science / Big Data. Interface entre l'équipe Big Data BP2I (fournisseur d'infrastructure Hadoop HDP) et d'autres équipes BNP (ses clients) qui font de la Data Science.

❖ Tâches :

- Conception & développement d'application Big Data :
Ingestion de données incidents du client (équipe Incident Management) dans un Datalake mutualisé (**cluster Hadoop HDP**). Traitement de données (nettoyage, calculs de KPI sur les incidents, enrichissements) avec la librairie **pandas de Python**. Outputs prêts pour faire du reporting.
- Conception & développement d'application Big Data Engineering :
Reprise de l'application précédente et développement de **code Spark** pour remplacer les traitements les plus coûteux et diminuer le temps de calcul.
- Ajout des fonctionnalités dans le Datalake :
 - Mise en place des scripts pour l'ingestion de données dans le **HDFS** à partir de serveurs de collecte.
 - Ingestion et archivage de données (index) **Splunk** dans le **HDFS**. Lecture des index avec **Hive**, à l'aide de l'application **Bucket Reader**.
- Conception & développement des **POC Data Science**:
 - Analyse de chemins fréquents d'assignation de tickets incidents (**librairie MLlib de Spark, librairie NetworkX de Python**).
 - Prédiction de groupes de résolution des tickets incidents (**Spunk Machine Learning Toolkit**).
 - Prédiction de commandes (à taper sur un serveur) qui emmèneront à la résolution d'un incident, basée sur un modèle de **Machine Learning**.
- Support aux clients de l'équipe Big Data BP2I, ayant des problèmes liés à la Data Science (**algorithmes, Python, Spark, SQL, Hive etc**).





Août 2017 – Janvier 2018 : Analyste Big Data - JEMS

Descriptif mission : Conception et développement d'un outil d'analyse du marché de l'emploi.

- ❖ **Tâche :** Aider les entreprises à mieux rechercher des profils d'employés, et aider les chercheurs d'emploi à se former d'une façon efficace pour répondre aux besoins du marché.
- Collecte de données :
Scraping du Web (en *Python*, avec la librairie *Scrapy*), données Open Data, construction d'un dictionnaire de termes.
- Intégration de données :
Avec l'utilisation de l'outil *Talend* (et du langage *Java*).
- Stockage, indexation, visualisation de données :
Utilisation de *Logstash / Elasticsearch / Kibana*.
- Visualisation (additionnelle) de données:
Utilisation de *Tableau*.
- Data Science / Machine Learning :
 - ✓ But : Identification de différents types d'offres d'emploi en Big Data. Identification de différences entre les métiers de Data Scientist – Data Engineer – Data Architect (par exemple). Identification des technologies (prérequis) les plus importantes par métier.
 - ✓ Construction automatisée d'un dictionnaire de mots-clés décrivant les profils demandés dans les offres d'emploi (e.g. Hadoop, Python, R etc.) et translation de chaque offre d'emploi en forme de vecteur de mots-clés.
 - ✓ Utilisation d'algorithmes d'apprentissage non-supervisé (*k-means clustering*), tournant sur un cluster Hadoop de 3 machines, avec l'utilisation de Spark, Python (*PySpark*) et de la librairie d'algorithmes Machine Learning *MLlib*.

Janvier 2015 – Août 2015 : Associé de Recherche – Imperial College London

Descriptif mission : Travailler sur une composante du projet DesMOLD, financé par la Commission Européenne.

❖ **Tâche :**

Ajouter de la fonctionnalité sur l'outil de travail collaboratif du projet DesMOLD.

- Analyse du comportement des utilisateurs pendant un débat argumentatif.
- Identification des utilisateurs dont le comportement est "suspect" (menteurs, trolls).





Environnement technique :

- Développement d'un prototype de simulation et d'analyse de débats argumentatifs, en Java (avec Eclipse). Développement d'algorithmes d'apprentissage (régression logistique) pour l'identification d'utilisateurs "suspects".

Environnement fonctionnel :

- Travailler sur une composante du projet Européen DesMOLD, visant à optimiser la production industrielle de modèles très détaillés en plastique.

Octobre 2010 – Août 2014 : Thèse en Informatique (Intelligence Artificielle) - Université Paris Descartes

Descriptif mission : Analyser et améliorer les méthodes de raisonnement argumentatif multi-agents.

❖ **Tâche :**

Analyser et améliorer les méthodes de raisonnement argumentatif dans un système multi-agents, où les agents ont des conflits de croyances (sur des faits qu'ils pensent être vrais) et des conflits de préférences (sur la meilleure décision à prendre). Les agents, en utilisant le raisonnement argumentatif (basé sur la théorie des graphes, et sur la Logique) arrivent à trouver la meilleure option / décision (pour la collectivité).

Environnement technique :

- Développement d'algorithmes de raisonnement multi-agents, basés sur des graphes d'argumentation (en Java).

- Simulations de débats argumentatifs multi-agents (en Java).

- Développement d'algorithmes (en Prolog) pour l'analyse du changement de systèmes d'argumentation.

Environnement fonctionnel:

- Travail dans le cadre du laboratoire *LIPADE* de l'Université Paris Descartes.

Septembre 2013 – Août 2014 : ATER – Université Paris Descartes

Descriptif mission : Enseignement

❖ **Tâche :**





- Cours TD / TP : Introduction à la Programmation (langage C), Numération et Logique, Algorithmique et Structures de Données.

Octobre 2010 – Août 2013 : Moniteur – Université Paris Descartes

Descriptif mission : Enseignement

❖ **Tâche :**

- Cours TD / TP : Introduction à la Programmation (langage C), Numération et Logique, Algorithmique et Structures de Données.

Février 2010 – Août 2010 : Stagiaire M2 – Thales R.T.

Descriptif mission : Définition et implémentation d'un protocole de négociation entre agents par argumentation.

❖ **Tâche :**

- Conception et développement d'un système multi-agents d'aide à la décision.

Environnement technique :

- Développement d'un prototype en Java, avec la plateforme multi-agents JADE.

Environnement fonctionnel :

- Travail au sein d'une équipe de recherche et développement de Thales R.T.

Novembre 2006 – Mai 2007 : Analyste / Programmeur - Nossis S.A. (Datamation)

Descriptif mission : Développement d'applications interbancaires.

❖ **Tâche :**

- Travailler sur les modules d'analyse et de vérification de messages interbancaires SWIFT.

Environnement technique :

- Eclipse, Java, XML, Swift.





Environnement fonctionnel:

- Finance

Février 2006 – Juin 2006 : Stagiaire - Dynamic Business Solutions Services d'Informatique (DBS)

Descriptif mission : Développement d'applications web.

❖ **Tâche :**

- Développement d'applications web.

Environnement technique :

- Oracle JDeveloper, Java Server Faces (JSF), Java Server Pages (JSP).

FORMATION / CERTIFICATIONS

- ❖ **Janvier 2019:**
Formation SSI / Cybersecurité (You Are Digital)
- ❖ **Novembre 2016 – Février 2017:**
Machine Learning (Stanford University – Coursera MOOC)
Data Scientist's Toolbox (Johns Hopkins University – Coursera MOOC)
- ❖ **Octobre 2010 – Novembre 2014 :**
Thèse en Informatique (Intelligence Artificielle) – Université Paris Descartes
- ❖ **Septembre 2009 – Juin 2010 :**
Master 2 (Recherche) en Informatique (Systèmes Multi-Agents) – Université Paris Descartes
- ❖ **Septembre 2008 – Juin 2009 :**
Master 1 en Informatique – Université Paris Descartes
- ❖ **Septembre 2001 – Septembre 2006 :**
Licence en Informatique – Université Economique d'Athènes (Grèce)
- ❖ **Juin 2001:**
Baccalauréat grec "Apolytirion" (Lycée Léonin à Néa Smyrni)





JEMS - 6 bis rue des Gravieres 92200 Neuilly-sur-Seine - www.jems-group.com

Votre contact JEMS : Etienne DOLEON – 07 83 87 10 39