



ASA DATA SCIENTIST

SYNTHESE DE COMPETENCES

Domaines de compétences

- ◆ Machine Learning
- ◆ Datamining
- ◆ Intelligence Artificielle
- ◆ Scoring
- ◆ Classification
- ◆ Clustering

Environnements d'interventions

- ◆ Banque
- ◆ Industrie pharmaceutique
- ◆ Innovation Digitale
- ◆ Marketing

Environnements techniques

- ◆ Python
- ◆ R
- ◆ SAS
- ◆ Langages : SQL, Mongodb, Mysql
- ◆ Système d'exploitation : windows, Unix
- ◆ Logiciel : SAS, R, Spyder, Microsoft Azure





EXPERIENCES PROFESSIONNELLES

<IN-TEAM/ startup>

< 6 mois / Mai 2018, Dec 2018>

Domaine de compétences : Python, R

Intitulé de l'intervention : Développement d'algorithmes de Machine Learning

Objectif(s) :

- ◆ Détection des pannes des portes SNCF.
- ◆ Traitement et nettoyage des données, construction des modèles d'apprentissage automatique pour prédire la panne.

Réalisation(s) :

- ◆ Traitement et nettoyage des données, Détection des valeurs aberrantes(valeur extrême) en les éliminant de données
- ◆ Mise en place d'algorithmes d'apprentissage machine pour prédire la panne.
- ◆ Mise en place des modèles de séries temporelles journalières et mensuelles(GARCH, ARMA,IGARCH..) pour prédire la panne.

Environnement(s) technique(s) :

- ◆ R
- ◆ Python
- ◆ Microsoft ML Azure

<NOKIA/ Data >

< 6 mois / Oct 2017, Avril 2018>

Domaine de compétences : Python, R

Intitulé de l'intervention : Comparaison de différents API Speech to Speech

Objectif(s) :

- ◆ Speech to Speech : Processus de traduction vocale par lequel les phrases parlées sont instantanément traduites et parlées à voix haute dans une deuxième langue.
- ◆ Traitement et nettoyage des données, construction des modèles d'apprentissage automatique pour prédire la panne.

Réalisation(s) :





- ◆ Benchmark théorique de comparaison des API pour différents fournisseurs (IBM, google et microsoft), benchmark opérationnel: 3 API pour chaque fournisseur: Speech to text, Text to text et text to Speech.
- ◆ Tester les différents API au niveau temps de réponse et exactitude de la traduction à travers le calcul du WER Word Error Rate)

Environnement(s) technique(s) :

- ◆ R
- ◆ Python

<Pharmasimple/ Marketing >

< 1 an / Oct 2016, Septembre 2017>

Domaine de compétences : R

Intitulé de l'intervention : Classifications des clients et produits

Objectif(s) :

- ◆ Classifications des clients et produits
- ◆ Développement de Moteur de recommandation.

Réalisation(s) :

- ◆ Classifications des clients à travers des Scores RFM
- ◆ Moteur de recommandatio: Calcul de similarités entre clients et produits
- ◆ Définition de règles d'association,
- ◆ Application du filtrage collaboratif

Environnement(s) technique(s) :

- ◆ R
- ◆ Python





<Jacare-Technologies/ startup>

< 1 an / Oct 2015, Sept 2016>

Domaine de compétences : Mongodb, microsoft Azure, Python, R

Intitulé de l'intervention : Développement d'algorithmes de Machine Learning

Objectif(s) :

- ◆ Amélioration du Dashboard de la solution;
- ◆ Construction des modèles d'analyse et de prédiction des audiences et du trafic;
- ◆ Création de nouvelles variables ;
- ◆ Prédiction des enchères (Prix des publicités);
- ◆ Application des modèles à d'autres sources de données.

Réalisation(s) :

- ◆ Amélioration du Dashboard avec les outils Qlikview et ClicData
- ◆ Création de nouvelles variables : temps d'attention, distance à chaque personne par rapport à l'écran, nombre de vues, trafic et audience
- ◆ Prédiction d'audience et de trafic avec des modèles ARIMA
- ◆ Construction de modèles intelligents pour prédire le trafic et l'audience en utilisant les modèles Random forest et Boosting
- ◆ Tester l'outil Microsoft ML AZURE : prédiction de l'audience et prix de publicité par Deep learning.
- ◆ Tester les modèles construits sur d'autres projets comme : mesure du trafic exact dans les magasins des stations essences, le but est de cibler les clients et augmenter les revenus en très peu de temps.

Environnement(s) technique(s) :

- ◆ R
- ◆ Python
- ◆ Mongodb
- ◆ Microsoft ML Azure

<Creativ-Ceutical / conseil en industrie pharmaceutique>

< 7 mois / Février 2015- Juillet 2015>

Domaine de compétences : Analyse des données

Intitulé de l'intervention Analyse des données PMSI et Chainage Anonyme

Objectif(s) :

- ◆ Traitement et Analyse des données
- ◆ Sélection des hospitalisations liées au cancer du sein



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

Votre contact JEMS : Etienne D'OLEON – 07 83 87 10 39



- ◆ Développement d'un algorithme de chaînage Anonyme

Réalisation(s) :

- ◆ Traitement des valeurs manquantes et valeurs aberrantes
- ◆ Comparaison de la gestion du cancer du sein entre établissement public et privé
- ◆ Expliquer la transition entre les deux établissements par une chaîne de Markov

Environnement(s) technique(s) :

- ◆ SAS
- ◆ R
- ◆ SQL

<Amen Bank/ Banque et assurance>

<6 mois / Février 2013, Juillet 2013>

Domaine de compétences : Scoring, segmentation, classification, Analyse des données

Intitulé de l'intervention : Développement de nouvelles méthodes de prédictions du risque du crédit bancaire.

Objectif(s) :

- ◆ Reporting des données avec des packages sql.
- ◆ Traitement et nettoyage des données
- ◆ Prédire la rentabilité des clients
- ◆ Développer de nouvelles méthodes de classification

Réalisation(s) :

- ◆ Développement des packages de reporting : création de tables, chargement des tables avec un loader et exécution des scripts schell pour automatiser le reporting.
- ◆ Traitement statistique des valeurs manquantes, valeurs aberrantes et nettoyage des données.
- ◆ Application de méthodes d'Analyses factorielles (ACP et ACM) pour réduire le nombre de variables
- ◆ Construction de nouveaux modèles de prédiction en utilisant: Random Forest, Neural Network, SVM, Bagging, Boosting.
- ◆ Comparaison avec les modèles classiques(Analyse discriminante, régression logistique..)

Environnement(s) technique(s) :

- ◆ R, Sql





Projets Universitaires

<Projet 1/ Détection d'inversion de l'électrode en ECG : projet sur la plateforme Challenge Data >

< 2 mois / Novembre 2015, Décembre 2015 >

Objectif(s) :

- ◆ Prédire un signal électrique indiquant si oui ou non, un ECG a été enregistré avec des électrodes inversées (0 : pas d'inversion, 1 : inversion).

Réalisation(s) :

- ◆ Traitement statistique des signaux
- ◆ Transformation des données périodiques en des données fréquentielles en utilisant le transformé de fourrier par la fonction fftpack de la bibliothèque scipy
- ◆ Appliquer des filtres pour éliminer le bruit dans les données en utilisant le transformé des ondelettes.
- ◆ Construction de modèles de prédiction de signaux pour différents scénarios(avec et sans changement en fréquences, avec et sans différents critères de sélections)

Environnement(s) technique(s) :

- ◆ R, Python

<Projet 2/ Value at Risk et Backtesting de la VaR>

< 2 mois / Septembre 2015, Novembre 2015 >

Objectif(s) :

- ◆ Modéliser la volatilité des rendements de l'action Hisense grâce à des modèles de marché.
- ◆ Prévion de la VaR avec le modèle choisi comme le plus pertinent.
- ◆ Appliquer un backtesting pour comparer entre ces modèles et contrôler leurs performances





Réalisation(s) :

- ◆ Détermination du modèle ARIMA général de série temporelle : élimination de la tendance linéaire et saisonnalité des données.
- ◆ Détermination des modèles GARCH(m,s), GARCH-M , modèles IGARCH, les modèles asymétriques EGARCH, GJR et APARCH
- ◆ Comparaison de ces modèles par Backtesting et méthode de dépassement de VAR et le Expected Sh (ES)

Environnement(s) technique(s) :

- ◆ R

FORMATION

- ◆ **2011-2013 :** Diplôme national d'ingénieur en Statistiques et Analyse d'information à l'école d'ingénieur ESSAI- Tunisie
- ◆ **2014-2015 :** Diplôme de master 2 en Statistiques spécialité Big Data à l'université Pierre et Marie Curie paris 6.

