

Samuel Kaci

- Data Scientist -

DOMAINES DE COMPETENCES

- Data sciences :
 - Utilisation de l'analyse discriminante
 - Utilisation des arbres et forêts
 - Utilisation des SVM (Support Vector Machine)
 - Utilisation des Réseaux de neurones
 - Scorer une population
 - Partitionnement des données (Segmentation)
- Data analyse :
 - Trier des données
 - Stockage de données
 - Création de dashboard
 - Création de reporting
 - Statistique descriptive
- Statistique :
 - Création et analyse de sondage
 - Utilisation de tests statistiques

ENVIRONNEMENT TECHNIQUE

- R ★★★★★
- Python ★★★★
- SQL ★★★★
- Tableau ★★★★
- Power BI ★★★★★
- SAS ★★★
- Talend ★★★
- Linux ★★★
- AWS ★★

★ : Notion (présentation produit)

★★ : Pratique occasionnelle (fonctions de base)

★★★ : Pratique régulière (fonctions de base)

★★★★ : Confirmé (fonctions avancées)

★★★★★ : Expert (design, tuning de la solution)

FORMATION

Master SIAD (Statistiques pour l'Information et l'Aide à la Décision)

IAE Montpellier 2020

- La gestion et l'analyse de données statistiques
- L'extraction d'informations et de connaissances
- La modélisation en environnement big data
- L'algorithmique et l'intelligence artificielle
- La maîtrise d'outils et de techniques d'aide à la décision
- La gestion de projet en environnement risqué
- La réalisation d'études de marché et de prévision...

Master MIND (Mathématiques de l'Information et de la Décision)

Faculté des sciences de Montpellier 2020

- Être capable d'extraire les données pertinentes
- Réaliser le pré-traitement des données (nettoyage et mise en forme éventuels)
- Mener l'analyse exploratoire des données en utilisant des outils de visualisation et de réduction de dimension.
- Modéliser un problème : maîtriser les méthodes usuelles en science des données moderne et savoir proposer la/les méthode(s) adéquate(s) pour la résolution du problème posé, écrire un / plusieurs modèle(s) mathématique(s) adapté(s) au service de la problématique, et le(s) mettre sous forme idoine au traitement par les méthodes usuelles de la science des données moderne.
- Mettre en œuvre la méthode d'un point de vue computationnel et être capable de proposer des stratégies de choix de modèles
- Programmer efficacement dans au moins un langage (python, R)
- Savoir analyser et interpréter les résultats, c'est-à-dire produire de la connaissance à partir de l'information extraite.
- Relier la connaissance produite à la décision de sorte à éclairer et optimiser cette dernière.
- Être capable de communiquer les résultats à l'écrit et à l'oral

LANGUES

Français : Natif

Anglais : Professionnel

DISPONIBILITE

Disponible sous asap.

DETAIL DES MISSIONS REALISEES

Freelance, mission à Piwwop – Data Scientist
Bordeaux

Contexte

Les Echos Judiciaires Girondins reçoivent quotidiennement par email des demandes de publication d'annonces légales. Le traitement manuel de ces demandes (extraction des informations, création des annonces et facturation) est chronophage et sujet à erreurs. Une automatisation de ce processus permet d'accélérer le traitement tout en garantissant sa fiabilité.

Mon objectif

Développer une solution automatisée de traitement des demandes d'annonces légales.

Tâches effectuées

- Conception d'une application RShiny pour le traitement des demandes
- Intégration avec l'API Outlook pour récupération automatique des emails
- Développement d'un module IA pour l'analyse des emails et pièces jointes
- Extraction automatique des informations clés (contenu de l'annonce, coordonnées de facturation)
- Génération de JSON standardisé pour le système de publication
- Mise en place du flux automatique de facturation

Environnement technique

- R
- Traitement NLP (IA)
- Json
- CI/CD
- Gitlab

Fonctions effectuées

- Structurer et coordonner les travaux d'un projet.
- Être garant du développement logiciel et de la méthodologie utilisée
- Développeur full-stack (front Shiny + back R)
- Concepteur de la solution d'automatisation
- Travailler en tant que Data Scientist et data Analyst

Contexte

Face à l'évolution rapide des technologies Data, une ESN a besoin d'adapter en temps réel ses formations aux besoins du marché. L'analyse manuelle des offres d'emploi étant impossible à grande échelle, un outil automatisé permet d'identifier les technologies les plus demandées par secteur et métier.

Mon objectif

Créer un outil d'analyse des tendances du marché de l'emploi Data.

Tâches effectuées

- Développement d'un scraper d'offres d'emploi (Data Science, Data Analyst, etc.)
- Analyse des technologies mentionnées dans les annonces
- Classification par secteur (banque, assurance, retail, luxe...)
- Synthèse par famille de métiers Data
- Création de visualisations pour l'analyse des tendances
- Intégration des résultats dans le processus de formation

Environnement technique

- Python
- Flask
- Pandas/Numpy

Fonctions effectuées

- Structurer et coordonner les travaux d'un projet.
- Être garant du développement logiciel et de la méthodologie utilisée
- Participer et arbitrer si nécessaire les débats techniques
- Travailler en tant que Data Scientist et data Analyst

Contexte

Bnp paribas a dû s'adapter aux nouvelles habitudes de travail imposées par la pandémie de Covid-19 en instaurant le télétravail. Maintenant, l'entreprise a pour objectif de réduire les espaces de bureaux pour réaliser des économies, tout en veillant à ce que les collaborateurs disposent d'un espace de travail optimal et adapté à leurs besoins. Il s'agit donc de trouver un équilibre entre ces deux aspects pour que les employés puissent continuer à travailler efficacement tout en réalisant des économies pour l'entreprise.

Mon objectif

Aider à optimiser les espaces de Bnp paribas

Tâches effectuées

- Amélioration du reporting existant (Tableau Server)
- Création d'un reporting (Power Bi)

Environnement technique

- Python
- R
- Tableau
- Power Bi

Fonctions effectuées

- Structurer et coordonner les travaux d'un projet.
- Être garant du développement logiciel et de la méthodologie utilisée
- Participer et arbitrer si nécessaire les débats techniques
- Travailler en tant que Data Scientist et data Analyst

Contexte

SAORIA est une start up créée en mars 2022 qui développe des outils d'aide à la décision permettant aux organisations d'anticiper et de mesurer les opportunités et les menaces des décisions politico-juridiques. Plus précisément, 4 volets sont à observer :

JADE, la plateforme permettant l'affichage personnalisée des nouveautés politico-juridiques

AKANE, l'analyse descriptive : elle consiste en l'analyse d'impact des nouveautés politico-juridiques par secteur, nature et niveau sur l'activité de l'entreprise

OPALE, l'analyse prédictive, se sert techniquement de l'analyse descriptive avec un processus et modèles mathématiques et informatiques plus poussés, permettant l'analyse d'impact des signaux faibles.

Enfin, ÉON la réalisation de recommandations par le biais de proposition de scénario.

Mon objectif

Améliorer la pertinence de l'algorithme.

Tâches effectuées

- Comprendre et apprendre le fonctionnement des IA de text mining
- Analyser les travaux déjà mis en place par la start up
- Apport de solution pour les différents algorithmes.

Environnement technique

- Python
- R

Fonctions effectuées

- Structurer et coordonner les travaux d'un projet.
- Être garant du développement logiciel et de la méthodologie utilisée
- Participer et arbitrer si nécessaire les débats techniques
- Travailler en tant que Data Scientist.

Contexte

BNP Paribas utilise un système de pré-acceptation sur le crédit à la consommation (cela permet de déterminer l'acceptation d'un crédit et sa mensualité maximale). Les résultats générés les ont amenés à la recherche d'une optimisation de ce système.

Mon objectif

Augmenter le nombre de personnes près acceptées en limitant le risque pris par BNP Paribas

Tâches effectuées

Répertorier les règles existantes

- Lecture des différentes documentations internes
- Création d'une nouvelle documentation plus synthétique

Reproduction de l'algorithme de pré-acceptation sur SAS afin d'y tester des modifications

- Coder l'algorithme sous SAS à partir de la documentation créée

Challenger chaque les règles d'éligibilité une à une, pour chaque règle :

- Évaluer le gain potentiel si la règle est supprimée ou modifiée
- Évaluer le risque potentiel si la règle est supprimée ou modifiée
- Validation de la suppression ou modification des règles par un comité

Challenger la formule de calcul de la mensualité

- Chercher comment augmenter la mensualité moyenne
- Évaluer le gain et le risque pris avec cette nouvelle mensualité

Regarder le taux d'éligible parmi les profils les plus appétents

Environnement technique

- SAS
- SQL
- Excel
- R

Fonctions effectuées

- Structurer, planifier et coordonner les travaux d'un projet.
- Être garant des méthodologies de science des données et de développement logiciel
- Participer et arbitrer si nécessaire les débats techniques
- Travailler en tant que Data Scientist.

Contexte

Chez l'Olivier Assurance les Data Analysts interviennent en soutien pour les différents corps de métier dans l'extraction de données et dans la mise en place de reportings récurrents.

Mon objectif

Aider les employés métier dans la réalisation de différentes d'études

Tâches effectuées

Création de reporting

Récupération des données en SQL

- Calcul avec SAS
- Mise en page sur Excel

Environnement technique

- SAS
- SQL
- Excel

Fonctions effectuées

- Être garant des méthodologies de science des données.
- Participer aux débats techniques

Contexte

La caisse régionale du Languedoc utilisait des outils de profiling développés par le Centre de Services Partagés (c'est un service transverse aux 39 caisses régionales, qui développe des outils et contenus communautaires). Leur utilisation a relevé des divergences entre leurs résultats et ceux obtenus par l'entreprise. Ces écarts étaient dus à des différences dans les bases de données et dans les définitions d'indicateurs.

Le Crédit Agricole m'a alors demandé de développer un outil de substitution et j'ai en outre proposé d'y intégrer mon savoir en data science, pour en créer un, plus complet, et approprié.

Mon objectif

Création d'un outil de profiling en R, composé de 2 parties :

- Première partie : Dashboard (statistique descriptive)
- Seconde partie : Scoring à l'aide de forêt aléatoire (data sciences)

Tâches effectuées

Création de la base de données utilisée par l'outil

- Choisir les variables
- Créer la table
- Historiser la table

Création d'un reporting numérique

- Découper les variables quantitatives par tranches
- Calculer la répartition des différentes populations sur ces tranches (en nombre et en pourcentage)
- Appliquer des tests statistiques pour comparer les écarts de répartition

Création d'un reporting graphique

- Afficher les répartitions précédentes sous forme de graphiques

Trouver les variables les plus discriminantes

- Création d'une forêt aléatoire pour trouver les variables les plus discriminantes dans le reporting

Scorer des individus

- Création d'une forêt aléatoire pour scorer des individus
- Trouver le score optimal pour une fonction gain / perte donnée

Environnement technique

- SAS
- SQL
- R

Fonctions effectuées

- Structurer, planifier et coordonner les travaux d'un projet.
- Être garant des méthodologies de science des données et de développement logiciel
- Participer aux débats techniques

Contexte

La loi de probabilité de Newcomb-Benford est de plus en plus utilisée dans les applications de la statistique, notamment en détection de fraude. Dans ces contextes, il importe de déterminer si un jeu de données est issu de cette loi de probabilité en contrôlant les risques d'erreur de Type I, soit de faussement identifier une fraude, et de Type II, soit de ne pas la détecter. L'outil statistique qui permet d'exécuter ce genre de tâche est le test d'adéquation. Pour la loi de Newcomb - Benford, le test d'adéquation le plus populaire est le test du khi-deux de Pearson dont la probabilité d'erreur de Type II est reconnue comme étant assez grande. En conséquence, d'autres tests ont été récemment introduits. Le but de ce travail est de proposer de nouveaux tests d'adéquation pour cette loi.

Mon objectif

Création d'un nouveau test plus puissant que les tests déjà existants.

Tâches effectuées

Répertorier les tests existants

- Lecture des différents articles concernant la loi de Benford et ces tests d'adéquation

Etude des tests lisses

- Lecture des articles sur les tests lisses

Création des tests lisses appliqués à la loi de Benford

- Calcul des différentes fonctions à utiliser

Création de la table des quantiles pour chaque test

- Calculs des différents quantiles

Comparaison des tests (évaluations des performances)

- Création de lois alternatives
- Utilisations des tests sur ces nouvelles lois

Environnement technique

- R

Fonctions effectuées

- Structurer, planifier et coordonner les travaux d'un projet.
- Être garant des méthodologies de science des données et de développement logiciel
- Travailler en tant que statisticien.

Contexte

Intégration de l'équipe de statisticien en charge de l'amélioration de l'algorithme qui associe un greffon à une personne en attente de greffe.

Tâches effectuées

- Fournir des informations statistiques fiables (dashboard) sur la situation des greffés du foie en France.
- Conduite et organisation d'une étude statistique en formulant une question scientifique.
-

Environnement technique

- R

Fonctions effectuées

- Structurer, planifier et coordonner les travaux d'un projet.
- Être garant des méthodologies de science des données.
- Travailler en tant que statisticien.