



Passerelle public-privé en “data science” ...

... chez un éditeur logiciel, d'ERP, agroalimentaire

Mehdi Kaytoue

Rencontre annuelle
MaDICS 2018

Strasbourg
21 novembre 2018



Déroulement

- I. Passerelle public-privé**
- II. Brève présentation de l'entreprise**
- III. Des défis pour assurer la compétitivité**
- IV. Des défis organisationnels pour la conquête**
- V. Conclusion**

Qui-suis je ?

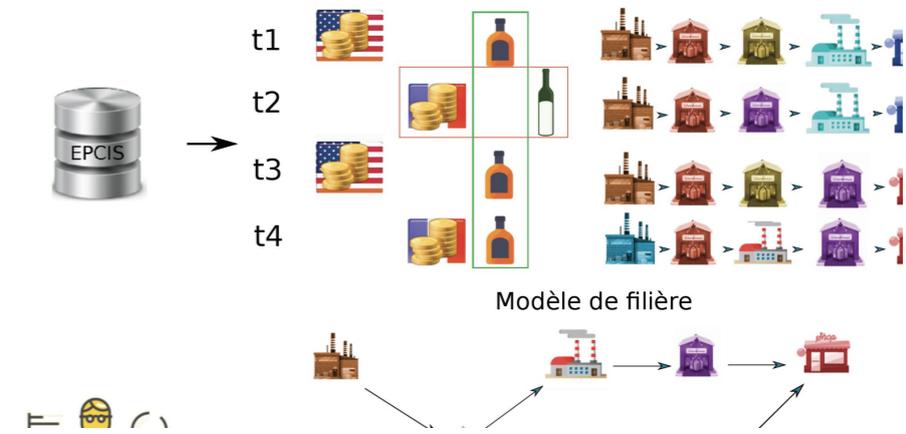
- **Parcours académique**

- PhD Informatique, Université de Lorraine (INRIA/LORIA), *Mining numerical data with pattern structure and formal concept analysis* avec A. Napoli (2011)
- Post-doc, Universidade Federal de Minas Gerais avec W. Meira Jr., Brasil (2011)
- Post-doc, INSA (LIRIS) avec J.-M. Petit, Lyon (2012)
- Maître de conférences à l'INSA/LIRIS (2012) avec l'équipe Data Mining & Machine Learning (DM2L)

- **Depuis février 2018 : “en disponibilité”**

- Chef de projet R&D chez Infologic
- Chercheur associé au laboratoire LIRIS (UMR CNRS 5205)

Pourtant !

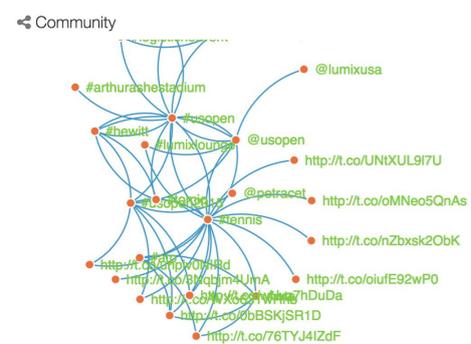


- J'aime le métier de MCF

- Véritable passion pour l'enseignement et le contact étudiant
- Environnement de travail de laboratoire, multifacette, riche
- Collègues au département, de l'équipe de recherche

- Activité scientifique

- Bon niveau de publication [Machine Learning, 2018; Data Mining and Knowledge Discovery 2017; IJCAI 2017]
- Encadrement de doctorants gratifiant [best student paper award ECML/PKDD 2018]
- Domaines d'application en fouilles de données variés
- Préparation de l'HDR en cours

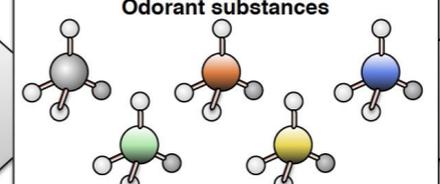




Physicochemical properties

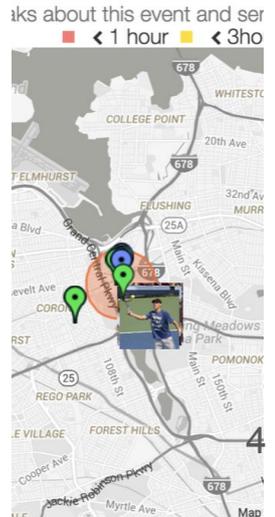
- Molecular weight
- Volume
- #atoms C, ...

Odorant substances



Olfactive qualities

- fruity
- woody
- vanillin
- citrus
- ...



Pourquoi alors ?

- Curiosité, connaître le monde industriel
- Les entreprises sont plus ouvertes avec la “hype” ML/DM/IA
- Une PME, un excellent choix
 - comprendre le fonctionnement global d’une entreprise
 - en fort besoin pour l’innovation et la R&D
 - comprendre ce qu’est transformation digitale par la pratique
- C’est le cas d’Infologic
 - Une reconnaissance du monde académique
 - Volonté d’intégrer des chercheurs pour “mettre de l’intelligence” dans le logiciel
 - 2 thèses CIFRE et 1 en évaluation,
 - Données massives, liées, comportementales... toute donnée transitant par l’ERP ou augmentant ces dernières

Brève présentation d'INFOLOGIC

- Éditeur d'un progiciel de gestion d'entreprise
- Spécialisé dans l'industrie agroalimentaire

Infologic : éditeur d'un progiciel de gestion d'entreprise

- **Entreprise Ressource Planning (ERP)**

- Capter, croiser et restituer l'information au bon endroit au bon moment pour la surveillance des activités et la prise de décision
- Intervient dans toute l'activité et administration : comptabilité, réception marchandise, gestion commerciale, gestion de la production, du stock...
- Centre névralgique, épine dorsale de l'entreprise

- **Quelques chiffres**

- Créé en 1982 par André Chabert, aujourd'hui 200 employés répartis sur 5 sites
- #Clients : 24 clients (1980's), 200 clients (1990's), 450 (2018)
- CA doublé en 10 ans : 8M (2008), 12M (2013), 18M (2018)

- **Industrie agro-alimentaire**

- lourde et spécifique : DLC (frais/ultra-frais), volumes importants (2M de colis/jour , 200K messages EDI/jour, 20K postes installés avec la solution Copilote)
- Clients de tailles différentes (CA de 100k à 4 milliards)
- Différents métiers (négociants/producteurs/consommateurs - viande/alcool/plats...) ⁷

L'ERP Copilote

- Infologic édite, vend et intègre, réalise la maintenance et assure
 - Un vaste périmètre fonctionnel (“tout sauf la fiche de paie”)
 - Un logiciel robuste et mature (>30 ans d’expérience, pas le droit à la panne, couche outil interne pour réactivité accrue)
 - Un Infocentre (business intelligence) : accès données liées et temps réel
 - Une solution multimétiers ultra-personnalisée (>10K paramètres)



Des défis pour assurer la compétitivité

- Une entreprise en bonne santé
- ... qui pourrait vivre sur ses acquis et nouveaux clients
- ... un temps seulement sans innovation et R&D



R&D : suivre les avancées, s'adapter et innover

- **Monde de la technologie en constante évolution**
 - ERP 2.0 : mobilité, tactile, centré utilisateur, ergonomie au coeur du métier
 - Industrie 4.0 : hyper-connectivité, capter et exploiter les données, piloter (jumeaux numériques)
 - Veille technologique, virages lourds (Pick (1982) → Java (2000) → Web (2014))
- **Monde scientifique plus accessible**
 - Relations académiques/laboratoires de recherche est devenu réalité
 - IA, Machine Learning, Data Mining, Recherche opérationnelle, Big data
 - Décloisonner des avancées à fort potentiel, les comprendre, les adapter, les intégrer, les industrialiser
- **Copilote : “pilote dont le rôle est d’assister le premier pilote”**
 - Optimisation, recommandation, clusters, prédictions informées/transparentes
 - Semi-automatisation : pas d’idée d’“IA qui fait tout”, le décideur reste maître à bord

Recherche opérationnelle & Optimisation de processus

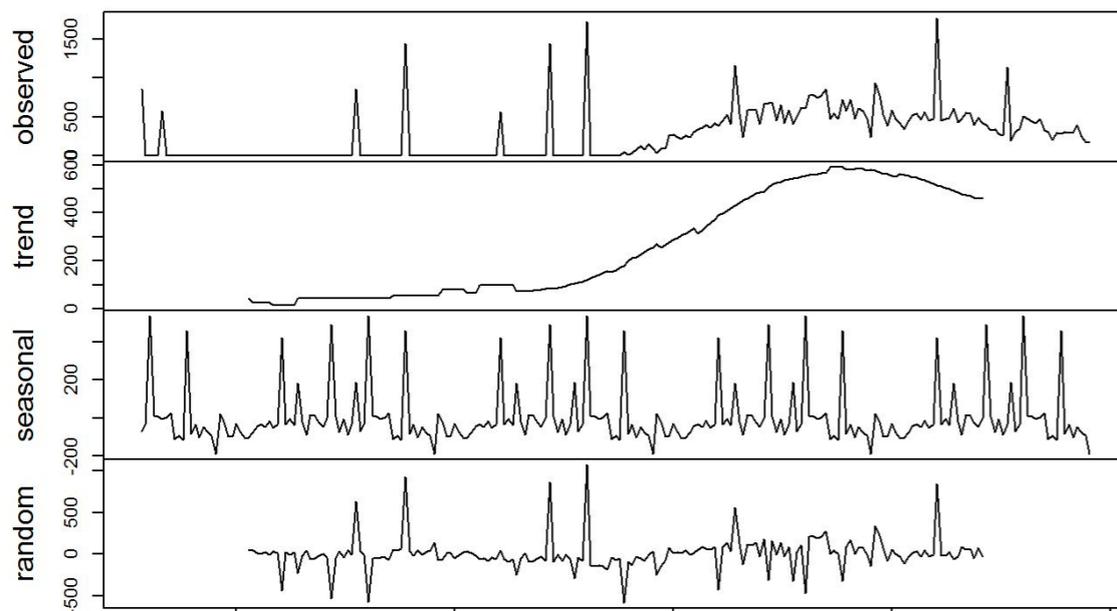


- De nombreux cas d'application (par définition de l'industrie)
 - Optimiser les ordres de fabrication : ensemble d'opérations (temps, dépendances), disponibilité machine, maintenance, opération de nettoyage...
 - Optimiser la préparation de commande : picking, pesée, colisage, étiquetage... sous contraintes (de place au sol, départ de commandes, postes...)
 - Optimiser l'emplacement des stocks
 - Ordonnancement des tournées logistiques
 - Choix des clients/fournisseurs en cas de pénurie (CRM/SRM)
- Anticiper dans l'incertain, intégrer pannes (maintenance prédictive)
 - Prendre en compte la probabilité de disponibilité d'une ressource
 - S'adapter aux changements rapidement
- Utilisation de techniques classiques
 - Programmation linéaire
 - Programmation par contraintes (p.e. solveur Choco)
 - ...

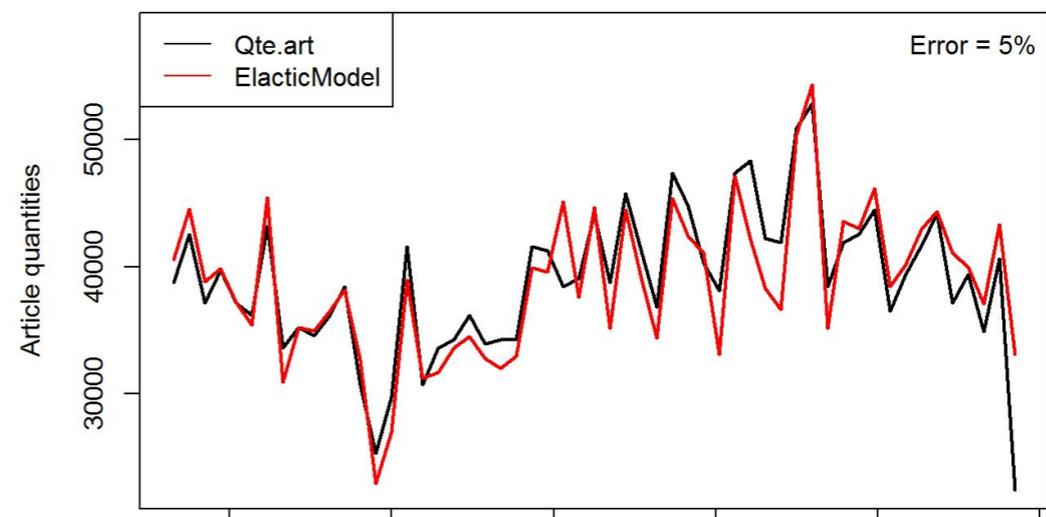
Machine Learning & Data Mining : Prévision de ventes

- Modèles de prévisions
 - N-1 : parfois trop simple, mais déjà efficace !
 - Decomposable time series model [Harvey & Peters, 1990]
 - Generalized additive model [Hastie & Tibshirani, 1987], ARIMA ...
 - Intégration de données externes, saisonnalité, ...
- Utilisée pour les planning d'approvisionnement, de production et budget comptable

Decomposition of additive time series



FRESHVALE PANACHE VIENNOIS CTA



Data Mining & Machine Learning

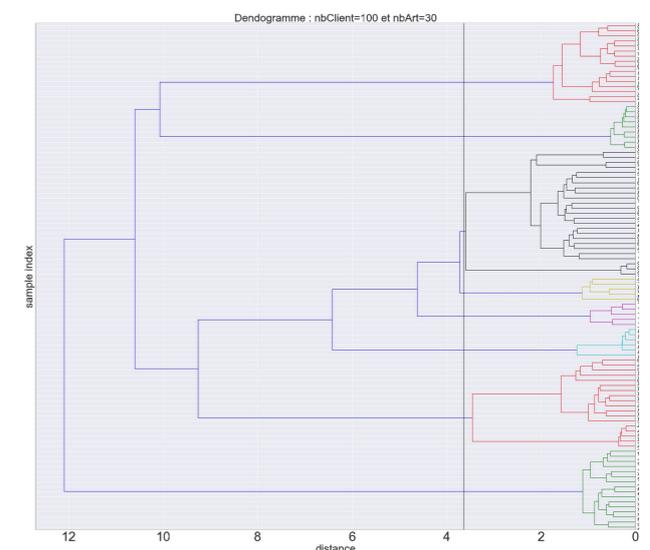
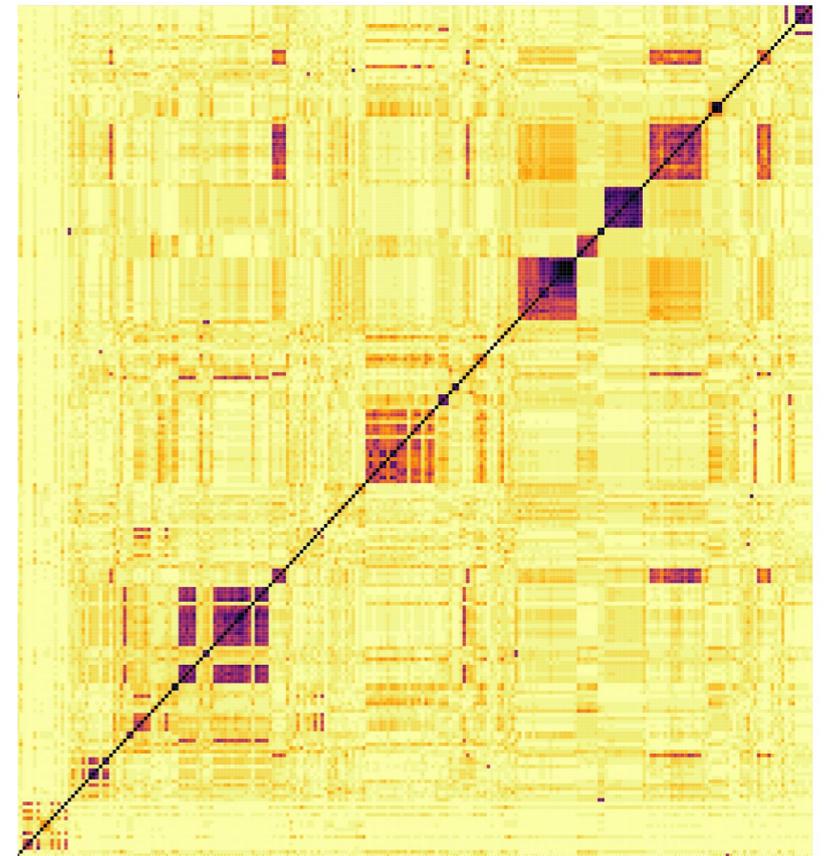
Pousser des articles à la vente

- **Cadencier en gestion commerciale**

- liste des produits pour un client assez statique (produits achetés < 60 jours)
- Commercial en visite chez un client
- Opérateur.trice en prise de commande

- **Recommandation**

- Profilage des clients (variables, modèles FRM et ses dérivés), segmentation du marché, attrition
- Calcul de similarités et présentation de hiérarchie de clusters
- Proposition de listes de produits
- Là encore : modèles simples dans un premier temps



Data Mining & Analyse de données comportementales



- Données
 - Données d'interaction avec le logiciel
 - Données organisationnelles des utilisateurs (planning, rôle, fonctionnalités métiers...)
 - Données externes
- Profilage utilisateur
 - Identifier des séquences d'actions caractéristiques d'une fonctionnalité
 - Comprendre le métier du client via son utilisation des fonctionnalités
 - Identification de planning types
 - Anticiper son activité pour le préchargement de données et proposition d'actions
 - Compréhension du travail, saisie du planning
 - Calcul d'efficacité de réalisations d'actions pour
 - la gestion de compétences et la gestion d'équipe
 - l'aide à la validation de congés...
- Détection d'anomalies ergonomiques (p.e., allers-retours entre onglets)

Des défis organisationnels pour la conquête

- Les possibilités et défis de valorisation de données s'appliquent aussi à nous, éditeur !
- Nous utilisons notre ERP en interne et sommes donc aussi clients
- Passer à l'échelle n'est pas simplement multiplier les RH

Copilote : Forces et faiblesses

- **Forces**

- Solution unique, multimétier, ultra-paramétrable et modulaire
- Vient épouser l'activité de l'entreprise
- Demande le moins d'adaptation possible du client lors de l'installation
- Développements spécifiques au besoin

- **Faiblesses**

- Beaucoup de paramètres (>10K)
- Formation des intégrateurs difficiles
- Paramétrage complexe

- **Besoin d'outils d'aide au paramétrage reposant sur les données internes, ma mission principale**

Recueil des données, consolidation et mise en consultation

- **Copilote**

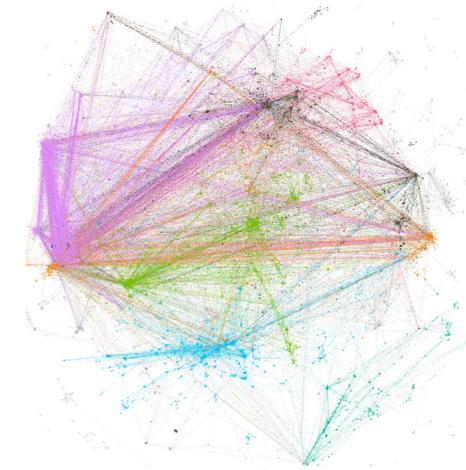
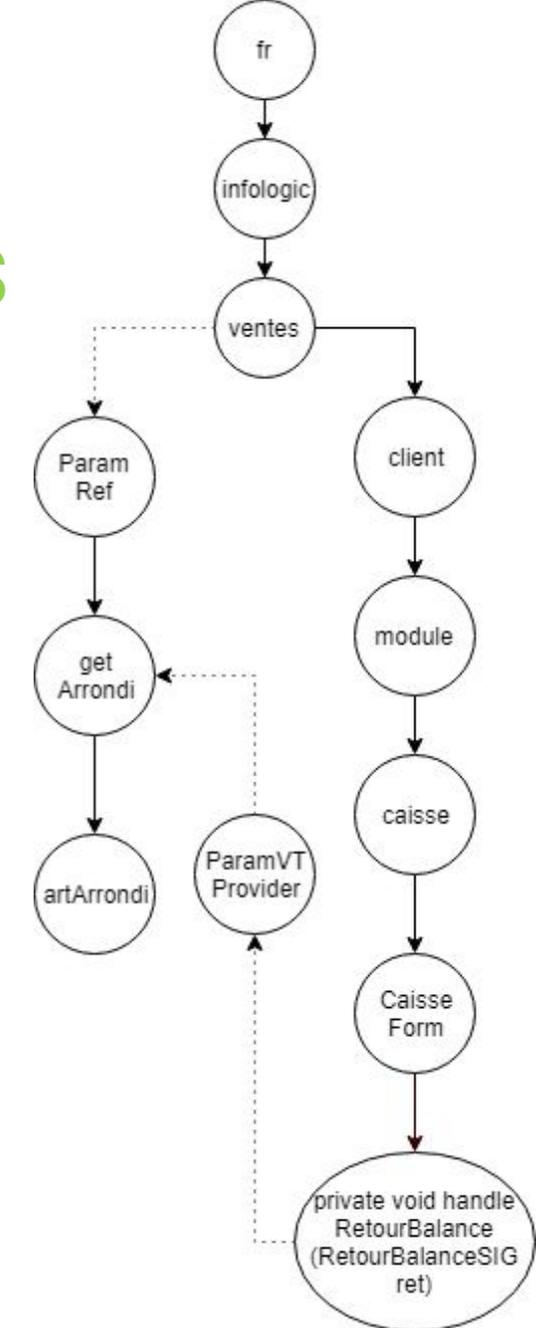
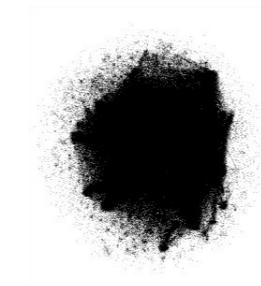
- Versionné (code source avec versions majeures et unitaires = commit git)
- Instancié : valeurs des paramètres chez tous les clients
- Utilisé : audit par les installateurs / clients (audit complet)
- Modularisé
 - Fonctionnalités (saisir une commande, réceptionner une matière)
 - Écrans de saisie, modules haut niveau vendus, ...
- Maîtrisé
 - Compétences du personnel sur chaque module
 - Procédés d'installation

- **Autant de sources de données liées et valorisables**

- Mais non disponibles de manière centralisées
- Liens manquants, entités non clairement identifiées, respect RGPD, contraintes réseaux...

Base de connaissances liées à travers des entités identifiées par analyse statique du code

- Construire l'AST de Copilote (Abstract syntax tree)
 - Permet de régénérer un programme sémantiquement équivalent
 - SPOON : librairie facile à manipuler (SPIRALS@INRIA Lille)
- Identifier des entités et leurs liens
 - Entités : paramètres, écrans, documentation, ...
 - Liens dans le code : parcours d'un grand graphe attribué
- Lier et instancier les entités trouvées
 - Leur instances chez les clients
 - Leur utilisation chez les clients, pendant le paramétrage historique
- Fouille de données
 - Trouver des familles d'entités, d'instanciations, de domaines d'attribut
 - Les intégrer dans des outils pertinents



Conclusion

- Des défis conditionnés par la disponibilité des données
- Retour d'expérience passerelle public-privé

Conclusion

- La transformation digitale d'une PME éditeur logiciel
 - à forte croissance ou qui peut (doit ?) se le permettre
 - a un besoin d'innovation qui passe par la R&D
 - qui nécessite des ingénieurs et chercheurs
- La R&D data chez Infologic
 - 1 chercheur (mcf dispo)
 - 3 docteurs
 - 3 doctorants
 - un ingénieur spécialiste en recherche opérationnelle
- Mettre en avant des relations avec laboratoire autre que CIFRE ?
- Comment interagir avec MADICS ?