

La méthode scientifique en génie et en apprentissage automatique

Camille Besse

Département d'Informatique et de Génie Logiciel
Université Laval, Québec, Canada

camille.besse@ift.ulaval.ca

16 mai 2019



Sources :

Marzuki B. Khalid : Research Methodology

Christine Dufour : Recherche et méthode scientifique

1 La recherche, c'est quoi ?

2 Classifier la recherche

- Selon l'utilisation
- Selon les objectifs
 - Recherche exploratoire
 - Recherche descriptive
 - Recherche explicative
- Selon les méthodes
 - Méthode qualitative

3 Le processus de la recherche

- Problème
 - État de l'art
- Méthodologie
- Données
- Analyse
- Résultats

La recherche, c'est quoi ?

- Toto a écrit un rapport sur l'usage d'internet au Québec, après avoir bien étudié la littérature sur le sujet
Est ce de la recherche ?

- Toto a écrit un rapport sur l'usage d'internet au Québec, après avoir bien étudié la littérature sur le sujet
Est ce de la recherche ? **non**

- Toto a écrit un rapport sur l'usage d'internet au Québec, après avoir bien étudié la littérature sur le sujet
Est ce de la recherche? **non**
- Mario a complété une recherche sur le personnel de la FSG et complété un document qui donne des infos sur l'âge, le salaire, les liens sociaux, etc.
Est ce de la recherche?

- Toto a écrit un rapport sur l'usage d'internet au Québec, après avoir bien étudié la littérature sur le sujet
Est ce de la recherche? **non**
- Mario a complété une recherche sur le personnel de la FSG et complété un document qui donne des infos sur l'âge, le salaire, les liens sociaux, etc.
Est ce de la recherche? **non**

- Toto a écrit un rapport sur l'usage d'internet au Québec, après avoir bien étudié la littérature sur le sujet
Est ce de la recherche? **non**
- Mario a complété une recherche sur le personnel de la FSG et complété un document qui donne des infos sur l'âge, le salaire, les liens sociaux, etc.
Est ce de la recherche? **non**
- Jules a participé à un atelier pour le développement et préparé un rapport technique donnant quelques "recettes". Il a fait pour cela une revue de la littérature et a questionné les intervenants du workshop.
Est ce de la recherche?

- Toto a écrit un rapport sur l'usage d'internet au Québec, après avoir bien étudié la littérature sur le sujet
Est ce de la recherche? **non**
- Mario a complété une recherche sur le personnel de la FSG et complété un document qui donne des infos sur l'âge, le salaire, les liens sociaux, etc.
Est ce de la recherche? **non**
- Jules a participé à un atelier pour le développement et préparé un rapport technique donnant quelques "recettes". Il a fait pour cela une revue de la littérature et a questionné les intervenants du workshop.
Est ce de la recherche? **non**

Seb gestionnaire d'une compagnie assemblant des PCs, reçoit des plaintes d'utilisateurs : après quelques mois, la carte mère est HS.

- Seb demande aux techniciens les informations nécessaires en vue d'identifier les facteurs qui ont influencé le problème ;
- Seb identifie plusieurs problèmes potentiels et émet des hypothèses ;
- Seb construit une grille de tests et demande à des utilisateurs les informations nécessaires pour la remplir ;
- Seb analyse les données obtenues de la part des utilisateurs, interprète les résultats à la lumière des hypothèses et tire ses conclusions.

Est ce de la recherche ?

Seb gestionnaire d'une compagnie assemblant des PCs, reçoit des plaintes d'utilisateurs : après quelques mois, la carte mère est HS.

- Seb demande aux techniciens les informations nécessaires en vue d'identifier les facteurs qui ont influencé le problème ;
- Seb identifie plusieurs problèmes potentiels et émet des hypothèses ;
- Seb construit une grille de tests et demande à des utilisateurs les informations nécessaires pour la remplir ;
- Seb analyse les données obtenues de la part des utilisateurs, interprète les résultats à la lumière des hypothèses et tire ses conclusions.

Est ce de la recherche ? **oui !**

Dans l'exemple précédent :

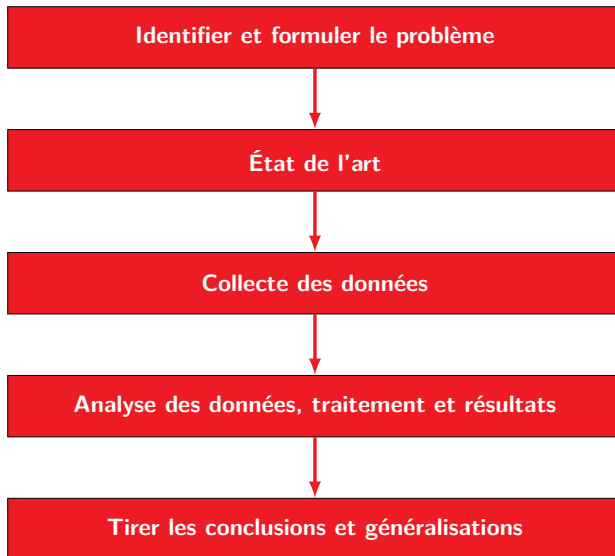
- Seb, y va à travers une séquence d'étapes ordonnées et donc systématiques ;
- Seb, ne saute pas directement aux conclusions, mais utilise une **méthode** scientifique qui consiste à investiguer en vue d'atteindre des conclusions ;
- Deux caractéristiques importantes de la recherche : Généralement une **procédure systématique** qui suit une **méthode scientifique** d'investigation.

La recherche est :

- la poursuite de faits ou de vérités sur un sujet
- une investigation organisée méthodologiquement
 - ▶ voulant résoudre des problèmes,
 - ▶ en testant des hypothèses,
 - ▶ pour finalement inventer de nouveaux produits.

La recherche est systématique dans la mesure où elle suit des étapes ordonnées de manière logique :

- Comprendre la nature du problème étudié et identifier les champs de connaissances qui y sont liés ;
- Établir l'état de l'art, i.e. collecter/étudier la littérature pour comprendre le problème a déjà été approché ;
- Collecter les données de manière organisée et contrôlée en vue d'arriver à des décisions valides ;
- Analyser les données appropriées au problème étudié ;
- Tirer les conclusions qui s'imposent et faire les généralisations qu'il faut.

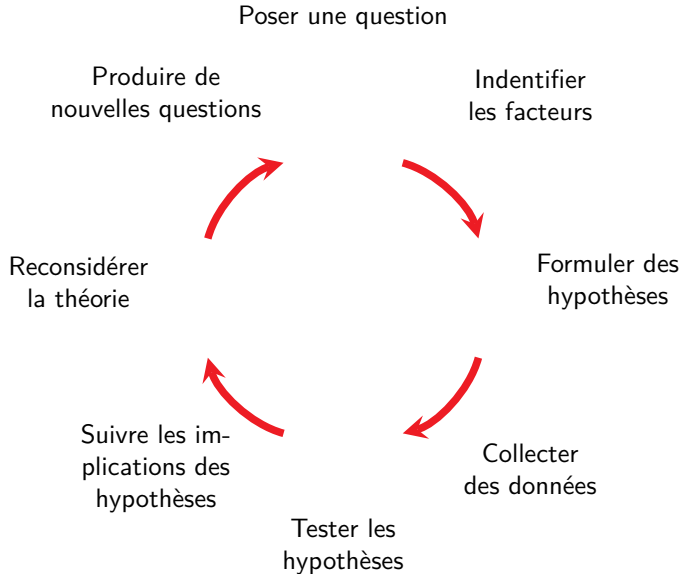


- Elle est reproductible ;
- Elle est incrémentale car souvent basée sur le travail d'autrui ;
- Elle est généralisable et transférable à d'autres contextes ;
- Elle est basée sur des arguments logiques et attachée à une théorie ;
- Elle soulève de nouvelles questions ;
- Elle est apolitique, éthique, objective,...

- Tout l'opposé
- S'acharner à vouloir montrer quelque chose qui n'existe pas ;
- Copier le travail d'autrui ;
- Falsifier des données pour montrer quelque chose ;
- Masquer ou déformer de l'information disponible dans le but de tromper.

- Faire de la recherche, c'est suivre une méthode scientifique
- I.e. faire un usage intégré de raisonnements **déductifs** et **inductifs**
- Cela permet d'expliquer et/ou prédire certains phénomènes
- Un postulat de base de la méthode scientifique est que **tout phénomène a une cause**.

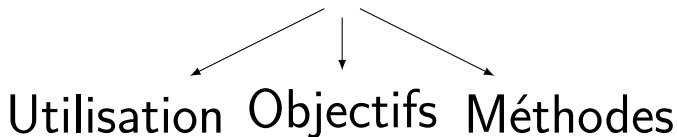
- Cela commence par la construction d'hypothèses depuis des observations ponctuelles et une connaissance du domaine (induction)
 - vers la construction d'hypothèses sur les conséquences ou implications possibles (déduction)
 - suivi par le tests de ces implications pour la confirmation ou la rejection des hypothèses.
- ⇒ L'usage intégré de l'induction et de la déduction est l'essence de la méthode scientifique.



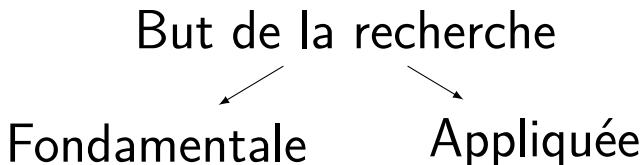
Classifier la recherche

- Passer en revue les recherches passées est une étape importante de la recherche
- Cela aide à la formulation du problème, à la construction d'hypothèse, et à la sélection de schémas comparatifs
- Il vous faut donc positionner votre recherche dans une certaine catégorie
- Chaque catégorie utilise un ensemble spécifique de procédures

On a trois grands types :



Exprime le degré avec lequel les résultats de recherche sont applicables et généralisables à l'enseignement et à la vie pratique.

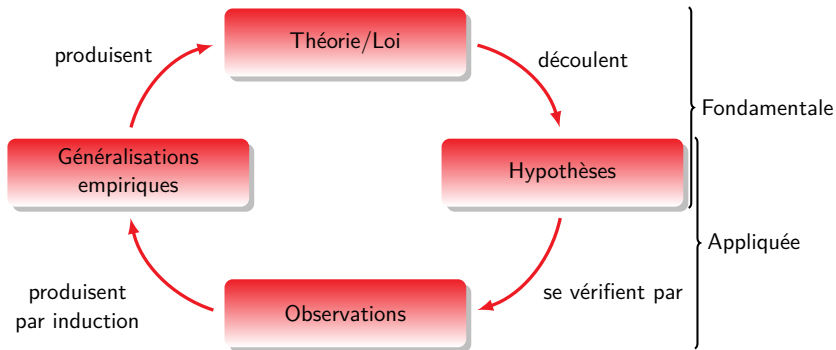


- Fondamentale : qui vise à éprouver des théories, des lois, des principes. afin d'accroître les connaissances dans un domaine et ce, sans préoccupation pratique immédiate
- Appliquée : qui vise à trouver des solutions pratiques à des problèmes pratiques

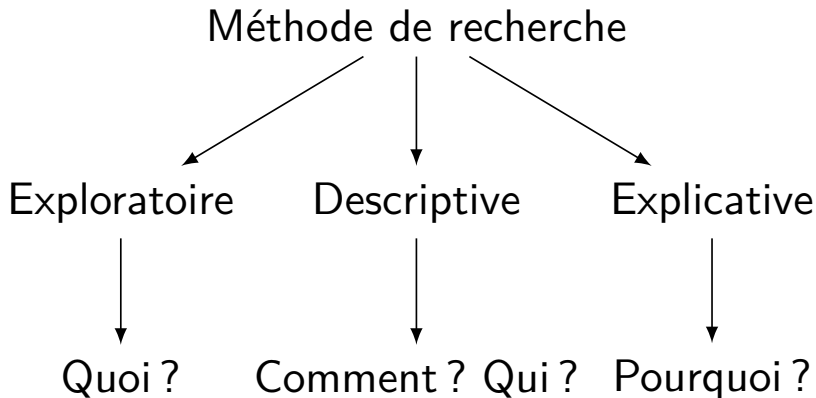
Mais à quoi ça sert ? [critiques usuelles] :

- Fondamentale : Perte d'argent et de temps puisqu'il n'y a pas d'application "réelle" ... *pelletage de nuage* ...
- Appliquée : Pas sérieux, non généralisable (rapide et petite-échelle), utile à court terme seulement ...

Oui ... mais les deux types de recherche se nourrissent l'une de l'autre.



En fonction de la nature des questions posées



Pour des phénomènes nouveaux, peu ou pas documentés.

- Se familiariser avec des faits, des données, des situations et des questions de base
- Formuler des questions pour des recherches futures
- Générer de nouvelles idées, conjectures ou hypothèses

Pour des phénomènes que l'on connaît un peu et que l'on veut décrire en profondeur

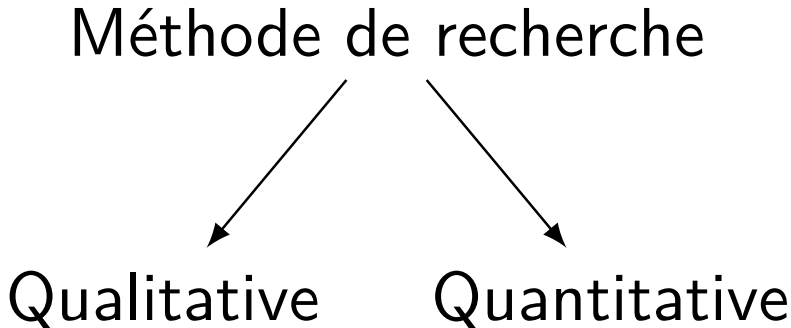
- Fournir une image détaillée et très précise
- Trouver de nouvelles données qui appuient ou contredisent d'anciennes données
- Clarifier une série d'étapes (processus)
- Documenter un processus ou mécanisme causal

La frontière entre recherche exploratoire et descriptive est parfois floue. Dans la pratique, on parle de recherche exploratoire-descriptive.

Pour des phénomènes connus, déjà décrits, pour lesquels on veut comprendre pourquoi les choses sont comme elles sont

- Tester un théorie
- Élaborer et enrichir l'explication d'une théorie
- Supporter ou réfuter un explication
- Déterminer quelle explication parmi plusieurs est la meilleure

En fonction des fondements philosophiques qui diffèrent selon les perceptions individuelles de la réalité, de la science, et de la nature humaine.



Pour comprendre un phénomène selon la perspective des sujets, en contexte.

- La réalité est multiple et se découvre dynamiquement en interagissant avec l'environnement : Connaissance relative ou contextuelle
- Les phénomènes humains sont uniques et non prévisibles

Critiques venant des "quantitatifs" :

- Manque de fiabilité, mensonges et biais humains dans les réponses au entrevues
- Biais inconscients dans l'analyse des résultats par les croyances des chercheurs

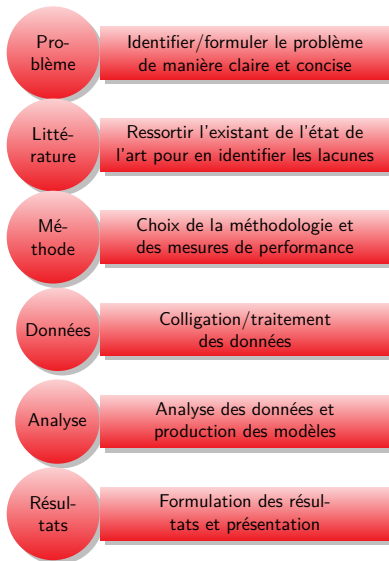
Pour décrire, vérifier des relations entre des variables, et examiner les changements observés chez la variable dépendante, à la suite de la manipulation de la variable indépendante.

- La réalité est perçue comme unique et statique mais probabiliste.
- Les phénomènes humains sont prévisibles et contrôlables
- Les faits objectifs existent en dehors du chercheur et peuvent être découverts : connaissance absolue

Critiques venant des "qualitatifs" :

- Manque de pertinence : on cherche à tout mesurer avec parfois des variables inadéquates
- Il est possible de faire dire n'importe quoi à des chiffres si l'on ne fait pas attention

Le processus de la recherche



Ceci est le point de départ de toute recherche : c'est l'aspect le plus ardu de la recherche. Un problème **sous ou sur spécifié** risque d'engendrer pas mal de difficultés par la suite.

- Spécifier le problème précisément ;
- Identifier les variables et les définir adéquatement ;
- Générer des hypothèses ou donner explicitement les questions de recherche ;
- Évaluer le problème quant à son importance d'un point de vue recherche ;
- Lier le problème posé à **l'état de l'art** et voir comment il a été approché et quelles méthodes ont été utilisées.

- Définir le problème à résoudre selon le domaine d'application, les contraintes et le contexte d'utilisation de la solution
- Identifier les variables d'entrées et de sorties, et les définir adéquatement, elles seront révisées au besoin si des variables exogènes ont à être ajoutées
- Générer des hypothèses sur les structures possibles de la résolution à valider
- Évaluer le problème quant à son importance du point de vue du résultat de l'application de la solution
- Lier le problème posé à **l'état de l'art** et voir comment il a été approché et quelles méthodes ont été utilisées dans des applications similaires

C'est quoi ?

- Lire les précédents travaux sur le sujet
- Assurer le caractère incrémental de la recherche
- Assurer la pertinence de la recherche

Objectifs :

- Circonscrire et préciser le problème
- Trouver de approches existantes pour comparaison
- Éviter des répétitions ou des approches inefficaces

Plusieurs techniques possibles :

- Définir les termes de la recherche
- Effectuer plusieurs recherches à l'aide de synonymes si possible
- Trouver un article effectuant la revue de littérature sur le problème si possible
- Sinon, trouver au moins un article décrivant et résolvant le problème considéré
- Une fois une référence pertinente trouvée, chercher parmi les articles le citant, et parmi ceux cités par lui

Mon algorithme personnel pour accélérer la lecture, si un point semble pertinent passer au suivant, sinon, changer d'article

- 1 Résumé : pour cerner le contenu
- 2 Conclusion : pour valider rapidement les résultats obtenus
- 3 Introduction : pour assurer que le problème est bien le même
- 4 Revue de littérature : pour ajouter les articles pertinents dans le pipeline de lecture
- 5 Résultats : pour comprendre les résultats obtenus
- 6 Méthodologie : pour valider qu'elle est adéquate et pertinente

Si vous vous êtes rendus à la méthodologie, une relecture complète et en profondeur de l'article est de mise dans le futur.

Une fois le problème bien posé ; il convient

- d'identifier le type de méthode de recherche et les méthodes elles-même ;
- de spécifier les sujets à étudier (les objets d'étude) ;
- de sélectionner adéquatement les échantillons, les données, etc.
- de sélectionner/construire des méthodes fiables pour faire les mesures des variables ;
- d'établir et décrire le protocole de recherche en étapes claires.

- d'identifier le type de méthode à employer, généralement quantitative, mais pourrait être exploratoire/descriptive ;
- d'identifier les techniques qui vont être employées pour faire les modèles ;
- de spécifier les objets à étudier qui ont été décomposés en caractéristiques ;
- de construire une méthode fiable pour faire la colligation et le prétraitement des données ;
- d'établir et décrire le protocole de recherche en étapes claires et reproductibles.

- Manipuler adéquatement les variables expérimentales ;
- Utiliser l'instrumentation pour mesurer les variables ;
- Observer et collecter les informations nécessaires ;
- Préparer les données en vue de l'analyse et de la modélisation ;
- Documenter la préparation.

- Manipuler adéquatement les données pour en assurer la fiabilité et la cohérence ;
- Collecter et colliger les données exogènes qui pourraient être utiles ;
- Observer et explorer les données pour en extraire les connaissances nécessaires ;
- Utiliser les techniques de préparation de données garantissant une bonne mise en oeuvre des modèles ;
- Préparer les données en vue de l'analyse et de la modélisation ;
- Documenter la préparation pour en assurer la reproductibilité.

- Les résultats de la recherche sont générées à cette étape ;
- Les données sont analysées en vue de fournir l'information nécessaire pour tester les hypothèses ;
- Les méthodes statistiques appropriées d'analyse sont utilisées pour tester les hypothèses ;
- L'analyse peut être faite à la main, par machine, voire par un cluster de machines.

- Création et entraînement des différents modèles testant les hypothèses de recherche ;
- Analyse des mesures des modèles permettant de valider ou de réfuter les hypothèses testées ;
- Interprétation des résultats à la lumière des mesures effectuées et des modèles utilisés.

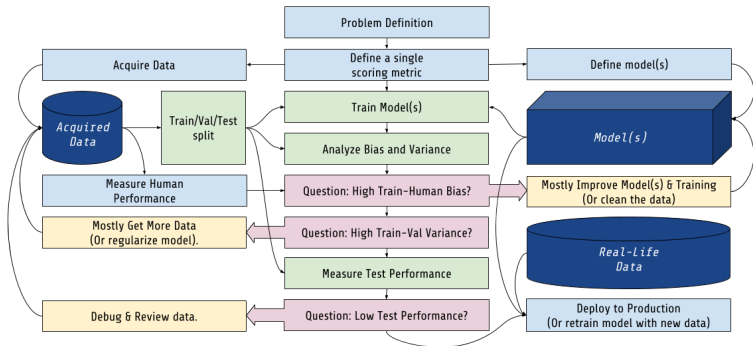
Vous devriez savoir cela ... mais juste pour être sûr ... 3 questions :

- 1 Paramètres vs hyper-paramètres ?
- 2 grid-search vs random-search ?
- 3 train / val / test ?
- 4 cross-validation vs hold-out ?

- Une fois l'analyse complétée, les résultats sont regroupés ou mis sous forme condensée ;
- Les résultats sont alors interprétés à la lumière des hypothèses et du problème de recherche étudié ;
- S'ensuit alors une discussion sur la consistance ou l'inconsistance avec des résultats existants, la place relativement à la science ;
- Les conclusions finales sont alors tirées et le tout doit finir en un écrit scientifique, si les résultats sont probants.

That's all folks !

Questions ?



Source : <https://github.com/Neuraxio/Machine-Learning-Figures>