

Faisabilité et interprétabilité du Machine Learning

Dominique Boullier

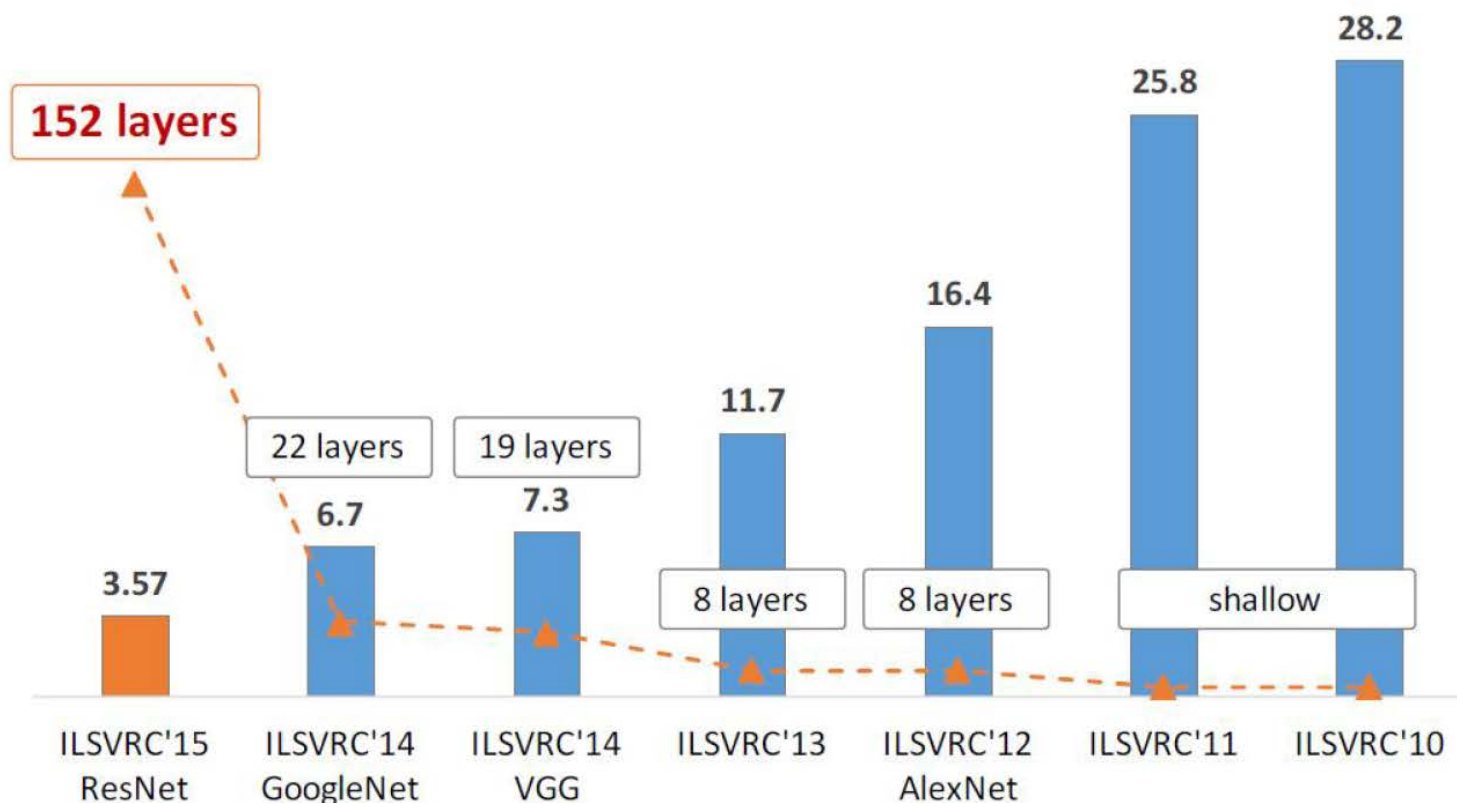
EPFL –Digital Humanities Institute

Novembre 2018



Les progrès du Deep Learning (en reconnaissance d'image) de 28 % à 3,5% d'erreurs en 5 ans

Visual classification with Supervised Deep Learning



ImageNet Classification top-5 error (%)

Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. "Deep Residual Learning for Image Recognition". CVPR 2016.

Source: Kaiming He, ICML 2016 tutorial:

http://icml.cc/2016/tutorials/icml2016_tutorial_deep_residual_networks_kaiminghe.pdf

La captation de la traduction par le machine learning: une mutation culturelle profonde

DeepL

Traducteur Linguee DeepL Pro Blog Infos

Traduire **français** (langue identifiée)

Traduire en **anglais**

pour qui sont ces serpents qui sifflent sur vos têtes?

>

who are those snakes that whistle on your heads for?

DeepL Pro

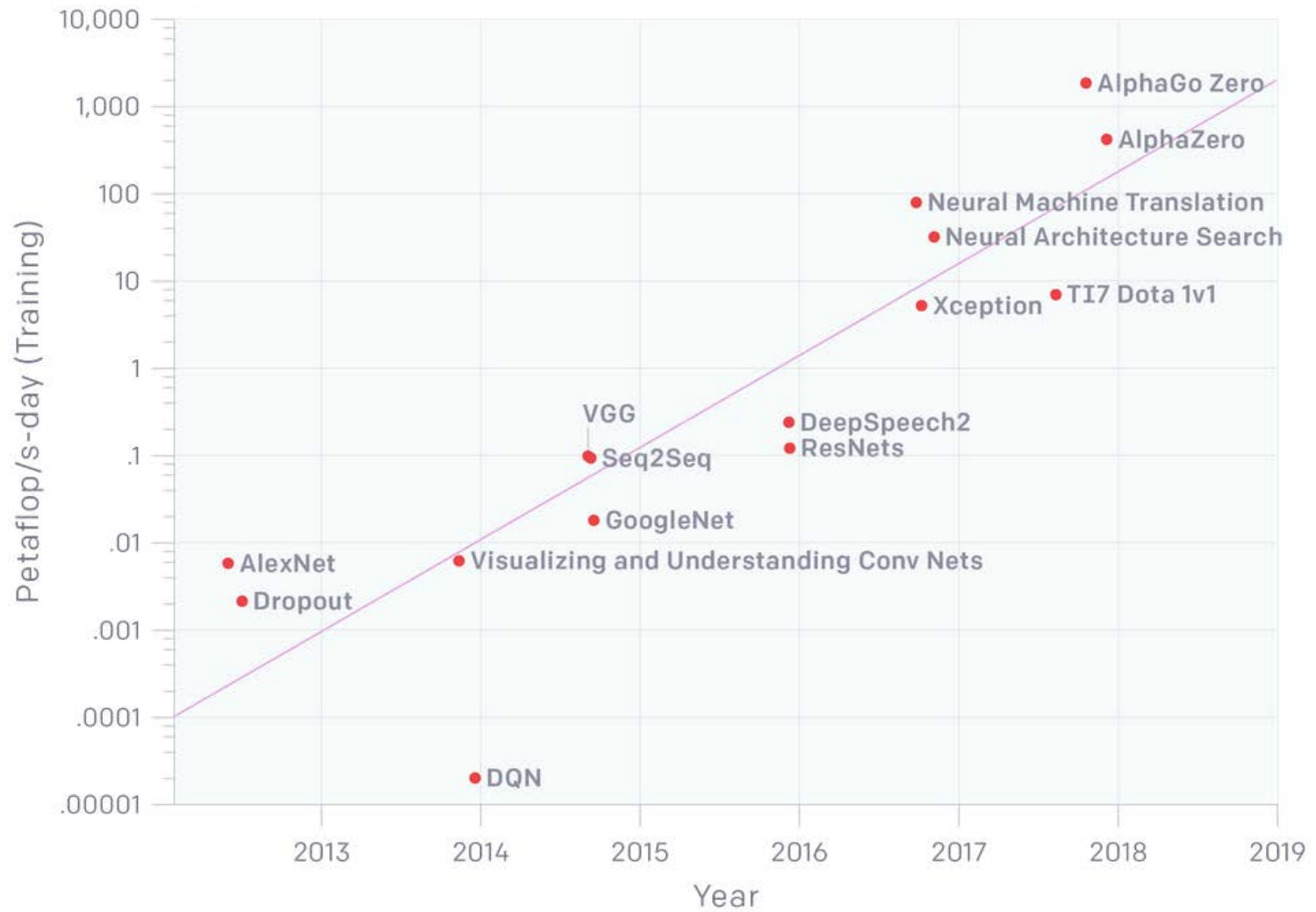
Exploitez tout le potentiel de DeepL

[En savoir plus](#)

- Accès à l'API de DeepL
[En savoir plus](#)
- Assurez la confidentialité de vos textes
[En savoir plus](#)
- Intégration à SDL Trados Studio 2017
[En savoir plus](#)

DeepL utilise des cookies. Pour plus de détails, consultez notre [politique de confidentialité](#).

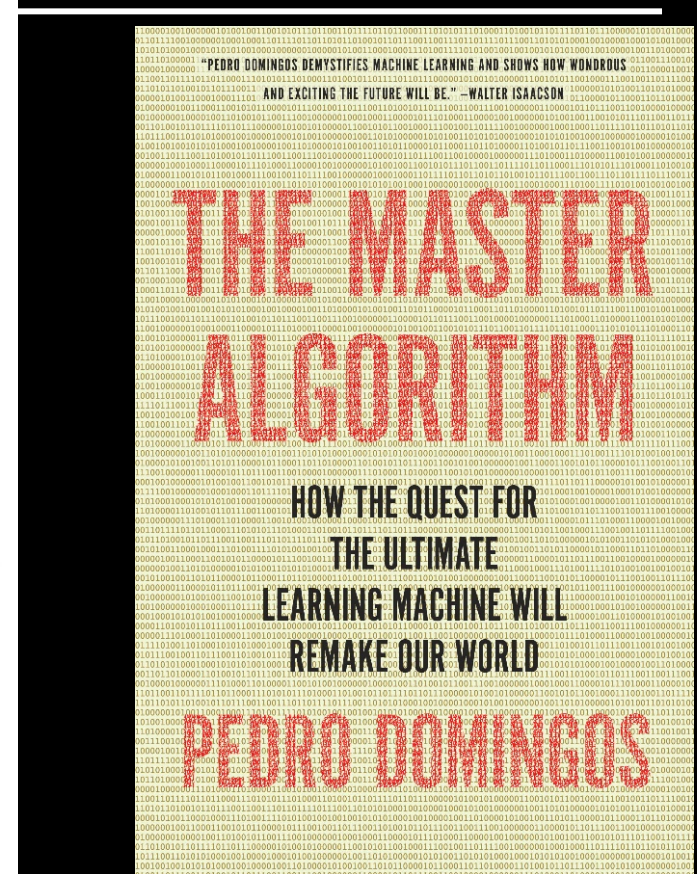
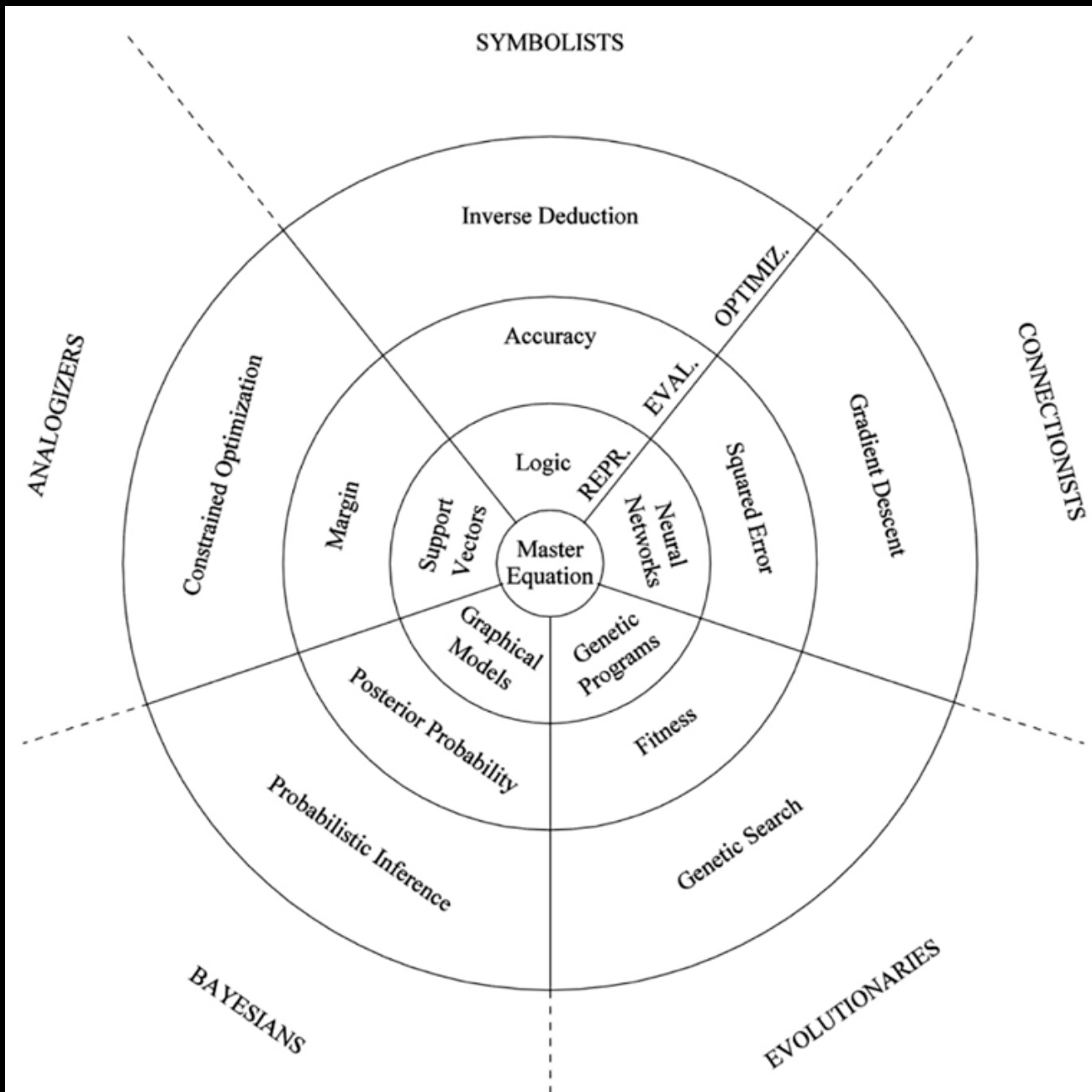
AlexNet to AlphaGo Zero: A 300,000x Increase in Compute



LOG SCALE

LINEAR SCALE

Une vision pluraliste du Machine Learning



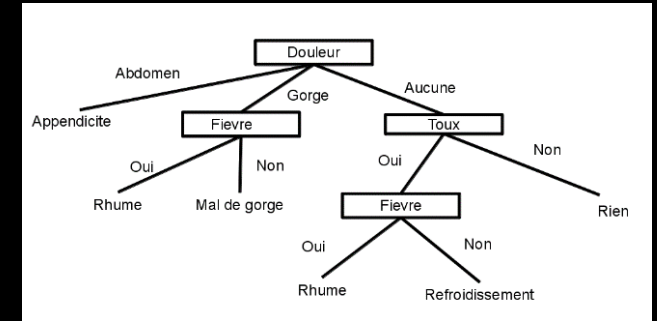
2 approches d'un statut différent

- **Evolutionnistes** (programmation génétique)
= une méta-analyse pour définir une stratégie de sélection des algorithmes
- **Bayesiens** (inférence probabiliste)
= présents dans tout le ML pour établir des probabilités dans des environnements incertains

3 modèles conceptuels

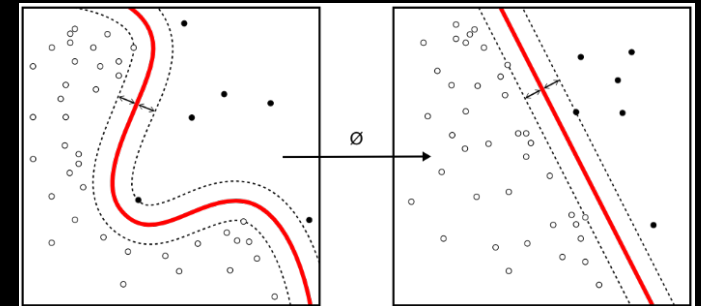
- **Symbolistes** : déduction inverse

- Trouver l'information qui manque,
- Faire remonter une déduction
- La généraliser autant que possible



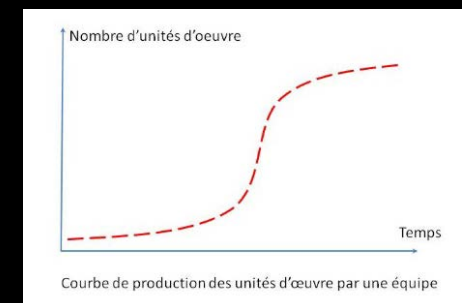
Analogistes: Optimisation contrainte pour la détection de similarité

- Quelle expérience conserver
- Comment les combiner
- Faire de nouvelles prédictions pour des situations nouvelles



- **Connexionnistes**: rétropropagation

- Comparer la sortie du système avec celle attendue
- Changer au fur et à mesure la pondération des connexions dans chaque couche de neurones,
- S'approcher le plus possible du résultat espéré (sur un dataset d'apprentissage)



La réalité du travail des machine learners les challenges Kaggle

The screenshot shows a web page from Kaggle.com. The header includes the title "No Free Hunch" and the URL "KAGGLE.COM". Below the header, there is a breadcrumb trail: "PREDICTING HOUSE PRICES PLAYGROUND COMPETITION: WINNING KERNELS". The main content area features a large title "Predicting House Prices Playground Competition: Winning Kernels" by Megan Risdal, dated 03.29.2017. A large, abstract image with a grid pattern and a starry background is positioned below the title. The text below the image describes the "House Prices playground competition" which ran on Kaggle from August 2016 to February 2017, involving over 2,000 competitors using advanced regression techniques like XGBoost to predict home sale prices based on 79 features. The text concludes by mentioning that the blog post features authors of kernels recognized for their excellence in data exploration, feature engineering, and more. At the bottom of the text, it says "In these sets of mini-interviews, you'll learn:". On the left side of the page, there is a sidebar with the text "The Official Blog of Kaggle.com", a search bar, and a "Categories" section listing various topics with their respective counts: DATA SCIENCE NEWS (61), KAGGLE NEWS (136), KERNELS (41), OPEN DATASETS (8), TUTORIALS (49), UNCATEGORIZED (2), and WINNERS' INTERVIEWS (219). Below the categories is a "Want to subscribe?" section with an "Email Address*" field and a "First Name" field.

> The Official Blog of Kaggle.com

Search

Categories

- DATA SCIENCE NEWS (61)
- KAGGLE NEWS (136)
- KERNELS (41)
- OPEN DATASETS (8)
- TUTORIALS (49)
- UNCATEGORIZED (2)
- WINNERS' INTERVIEWS (219)

Want to subscribe?

Email Address*

First Name

No Free Hunch

KAGGLE.COM

> PREDICTING HOUSE PRICES PLAYGROUND COMPETITION: WINNING KERNELS

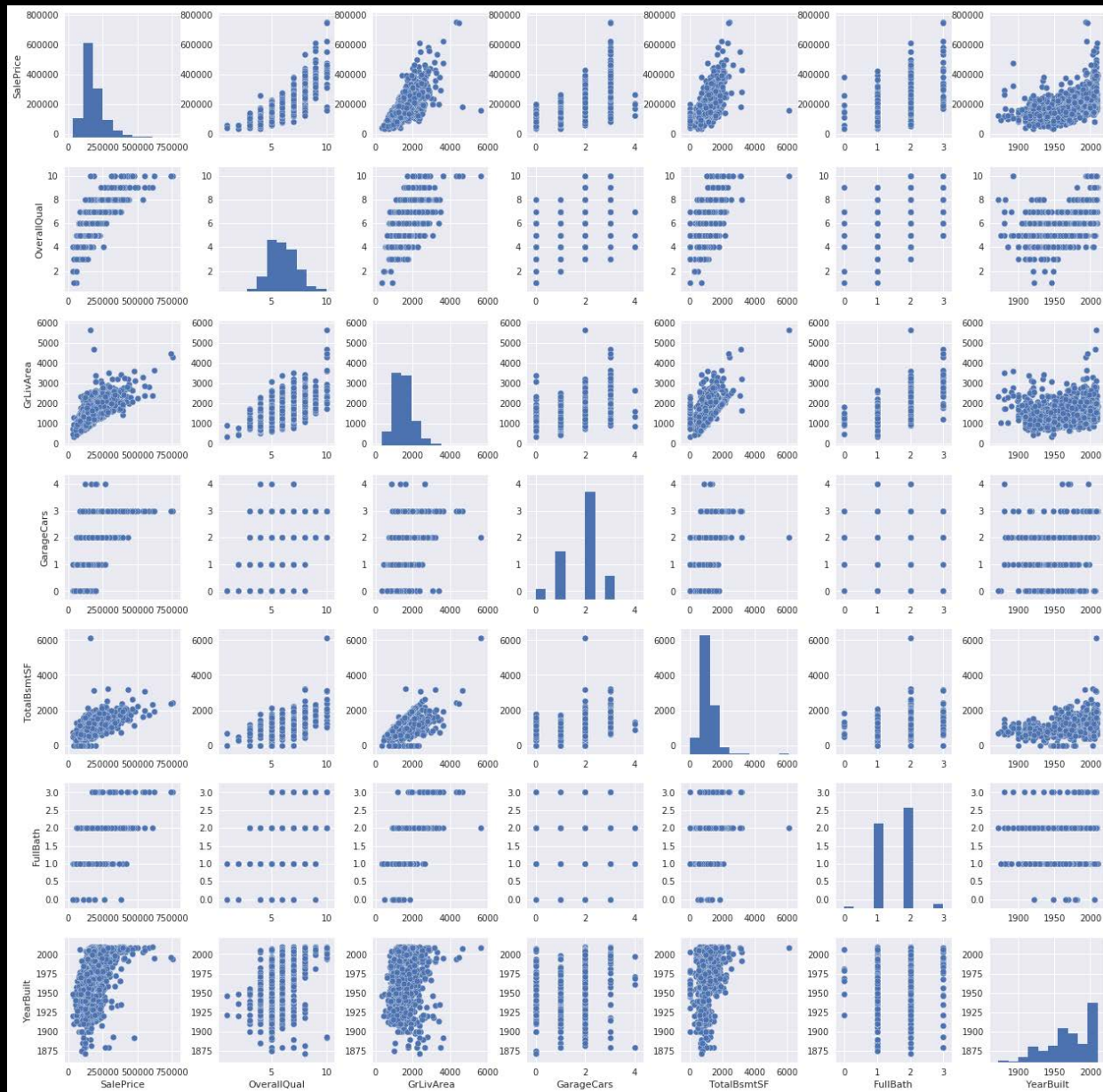
Predicting House Prices Playground Competition: Winning Kernels

Megan Risdal | 03.29.2017

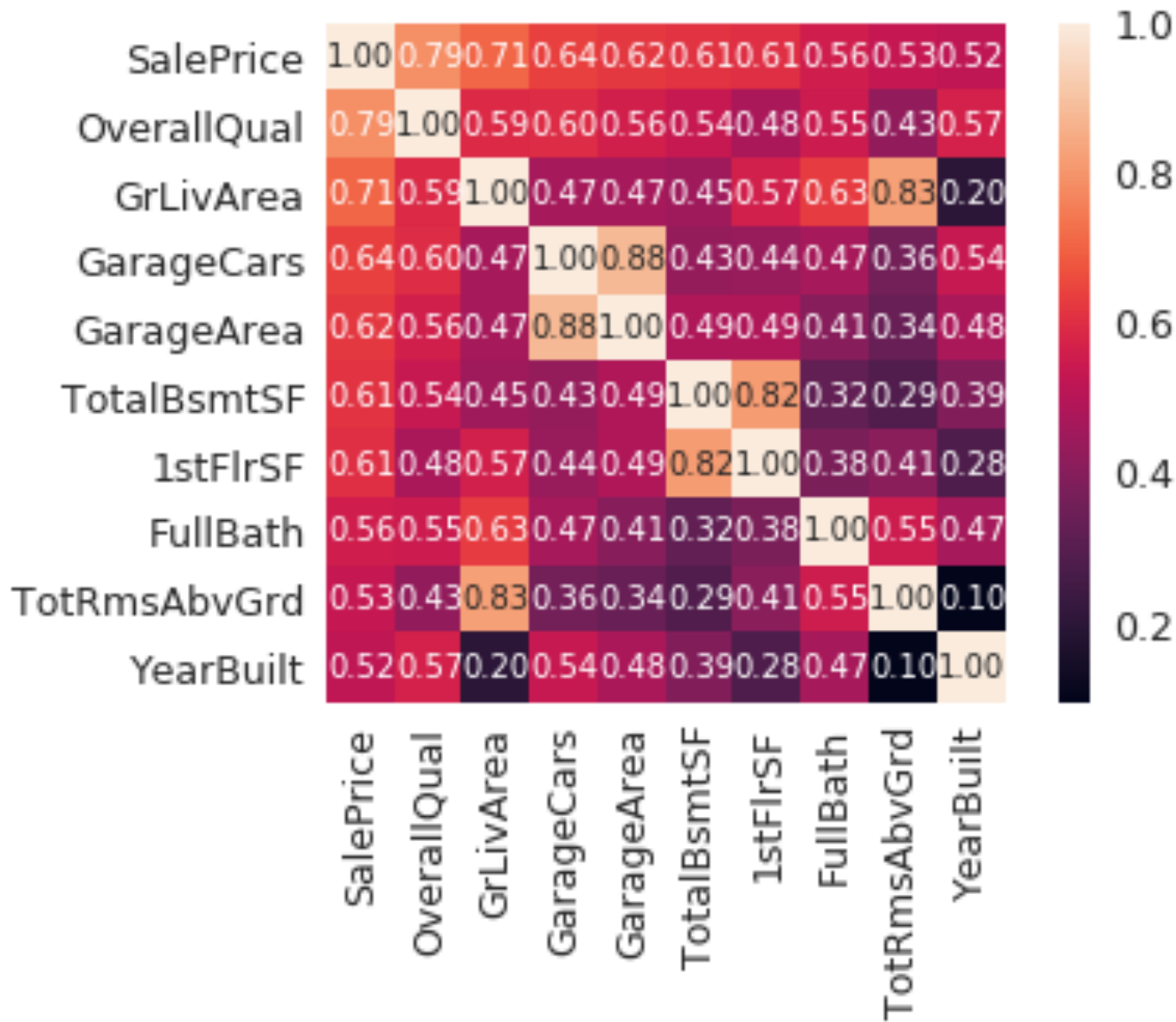
The [House Prices playground competition](#) originally ran on Kaggle from August 2016 to February 2017. During this time, over 2,000 competitors experimented with advanced regression techniques like XGBoost to accurately predict a home's sale price based on 79 features. In this blog post, we feature authors of kernels recognized for their excellence in data exploration, feature engineering, and more.

In these sets of mini-interviews, you'll learn:

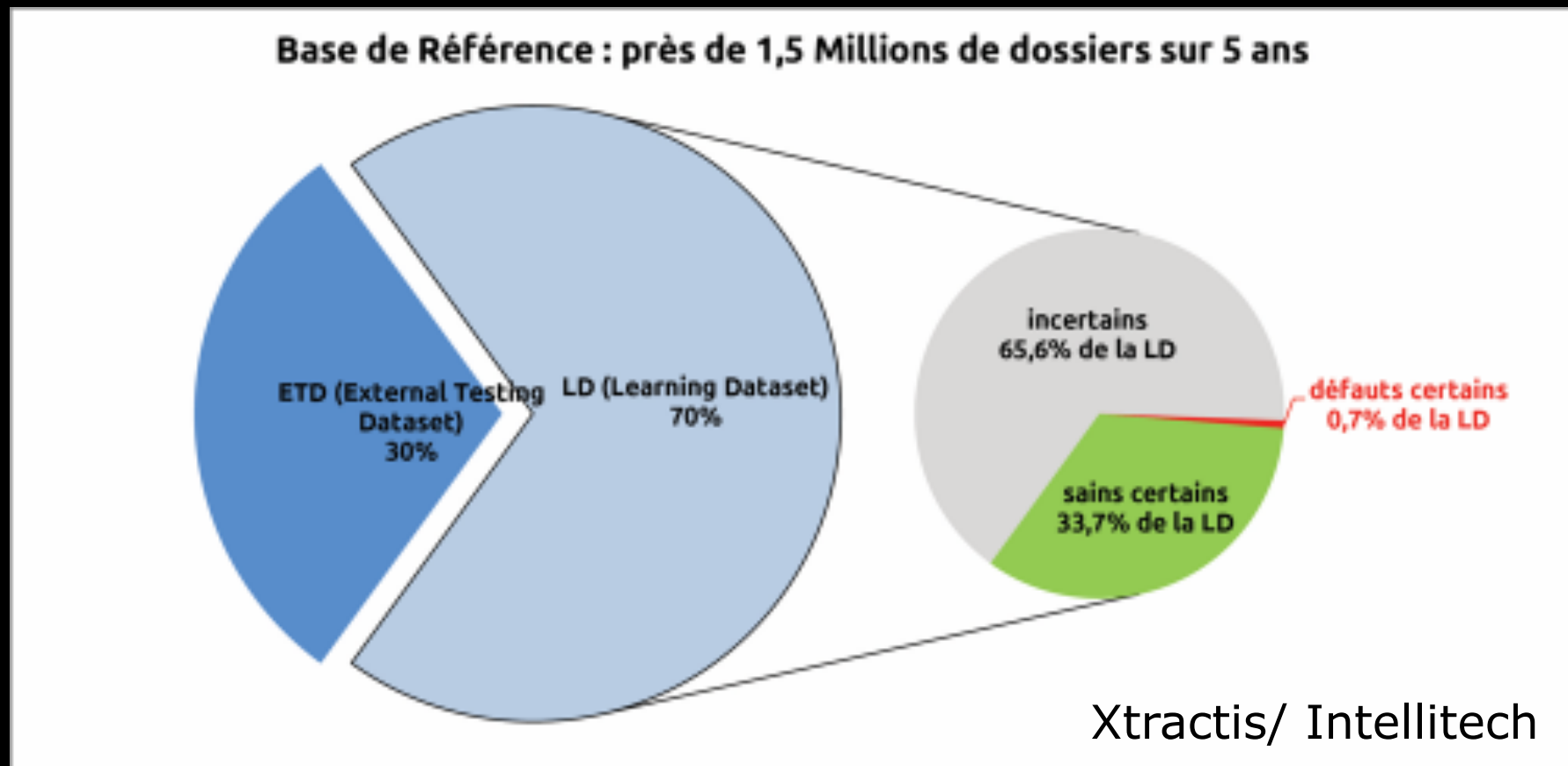
Une pratique beaucoup plus bricolée , agnostique et centrée sur le data engineering



Prédire un prix en fonction de 82 features des logements (arbre de décision)



Dataset d'apprentissage/ dataset de validation



L'extension de la supervision

Les automates ne sont pas ceux qu'on croit

- L'extension des robots humains qui classent, notent, corrigent, cliquent, pour faire apprendre les algorithmes
- Digital labor: Des tâcherons travaillant en micro-tâches sur des data (dataïfication) (Casilli)

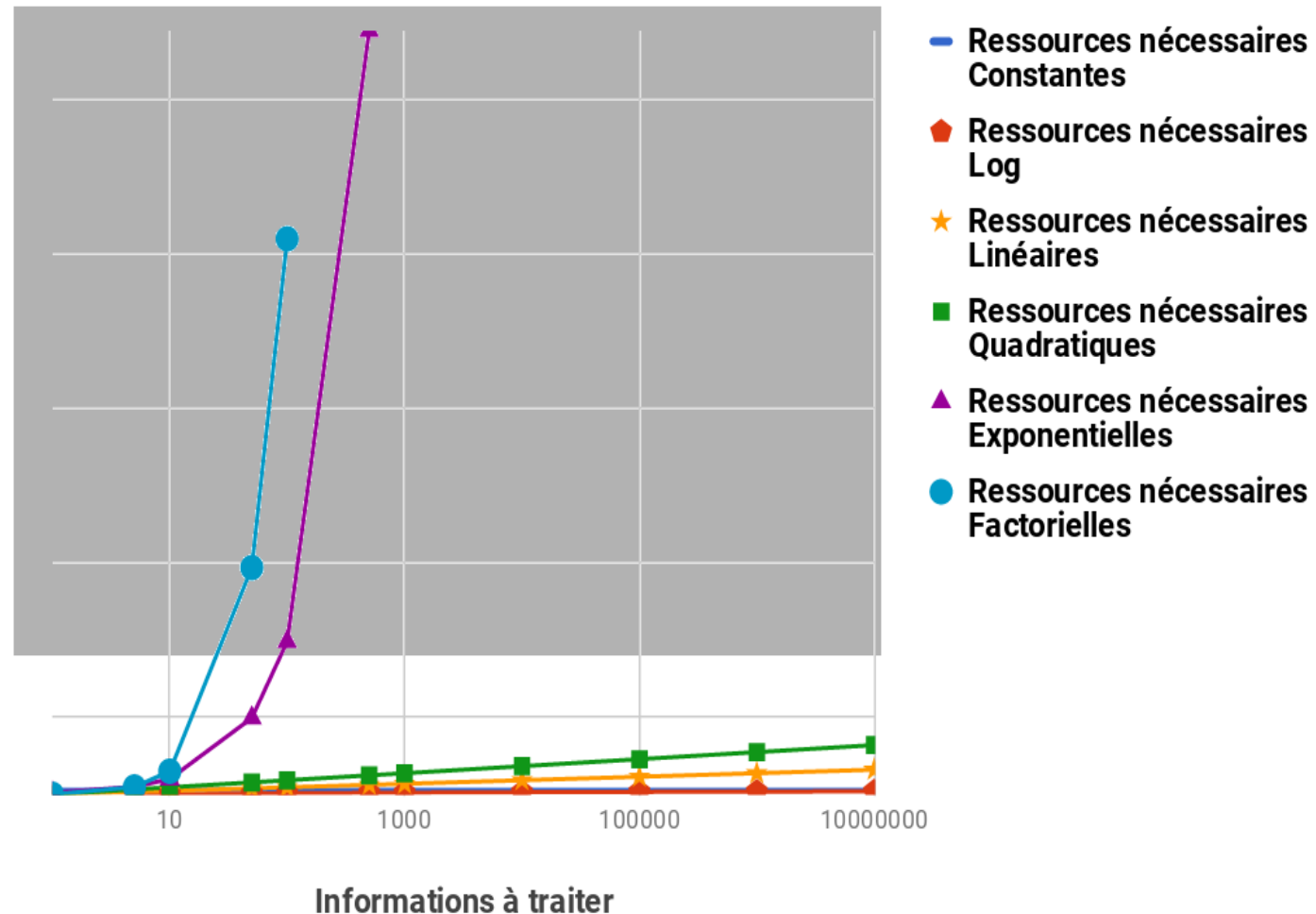
In 2016, Bloomberg highlighted the plight of the humans spending 12 hours a day pretending to be chatbots for calendar scheduling services such as X.ai and Clara. The job was so mind-numbing that human employees said they were looking forward to being replaced by bots.

- La tacheronnisation = l'extension de l'automation à tous les emplois = division en sub tâches

LES CONDITIONS DE FAISABILITÉ DU MACHINE LEARNING

Evaluer l'échelle de complexité des problèmes (Turing, Kolmogorov, VC dimension)

Ressources nécessaires en fonction des données à traiter



Plus automate que moi tu meurs = IA faibles au pouvoir

L'IA devient un automatisme « éclairé »:

- qui calcule en permanence
- qui vise la complétude de l'information
- qui est capable de corriger les défauts des non-décisions réflexes biaisées des humains automatisés

Des IA faibles comme aide à la décision

là où il y a peu de décision et beaucoup d'automatismes

Le cas de l'IA dans le conseil d'administration

The image shows a screenshot of a BBC News article. At the top, the BBC logo is on the left, and navigation links for 'Sign in', 'News', 'Sport', 'Weather', 'Shop', 'Earth', 'Travel', and 'More' are on the right. Below this is a red banner with the word 'NEWS' in white. Underneath the banner are more navigation links: 'Home', 'Video', 'World', 'UK', 'Business', 'Tech', 'Science', 'Stories', and 'Entertainment & Arts'. The article is in the 'Technology' section, indicated by a blue underline. The main headline is 'Algorithm appointed board director' in a large, bold, black font. Below the headline is the date '16 May 2014' and a row of social media sharing icons for Facebook, Twitter, Messenger, Email, and a 'Share' button. The article's main image shows a woman in a grey suit standing at the head of a long wooden conference table, pointing at a large screen displaying a line graph. Several people are seated around the table, some with laptops open. The image is watermarked 'THINKSTOCK' in the bottom right corner. Below the image is a caption: 'Would board meetings be improved by sitting next to an algorithm?'. At the very bottom of the screenshot, the beginning of the article text is visible: 'A venture capital firm has appointed a computer algorithm to its board of'.

Les placements financiers

Les recrutements

= Le court terme

**= se rassurer
(augures et haruspices
du XXIe ?)**

Quand l'IA ne sert plus à rien

Les décisions de long terme

+ de long terme = + d'incertitude



Le passé, seule source de l'IA, obstrue les devenirs en en faisant des futurs probabilisables

Les décisions procédurales

Le cas des diagnostics

- 80% sont des examens et des diagnostics de routine = aisément automatisables
- Le contrôle humain lorsque doute il y a
- Le staff pour discuter lorsque le problème est complexe
- = **design organisationnel**

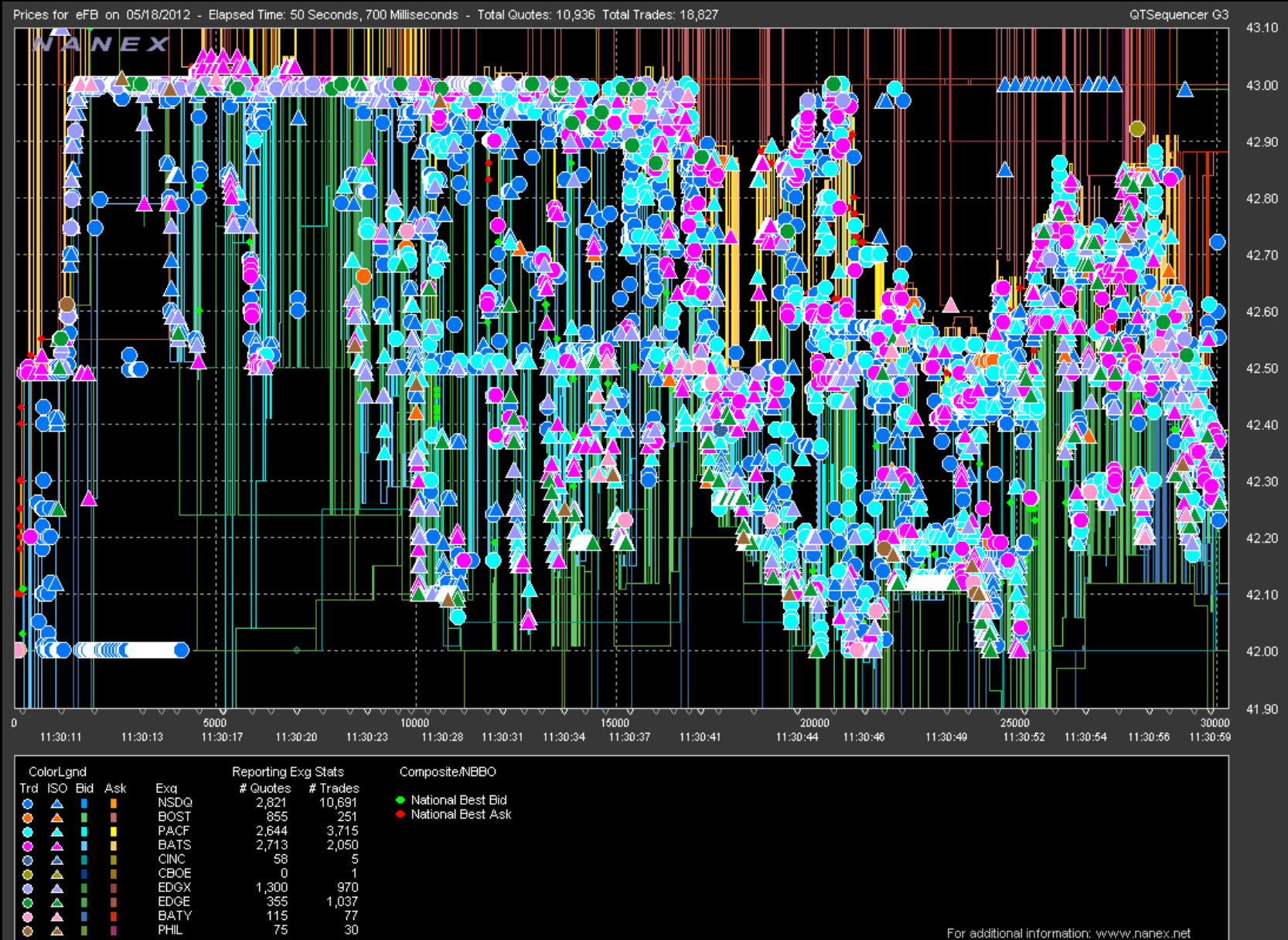


Les décisions en urgence

le coût cognitif des retours d'expérience



Les décisions spéculatives: la probabilisation du monde



Les décisions négociées: ni dilemmes ni théorie des jeux



Deux nouvelles rassurantes

- **1/ La pratique du ML** consiste en un assemblage de méthodes très empirique et non dogmatique
- **2/ Réduire les dimensions** est nécessaire pour rendre les problèmes tractables (feature engineering)
 - Les experts des domaines devraient être au pilotage de cela ou des pondérations

Deux problèmes sérieux

3/ Le Deep Learning (connexionnistes) génère une opacité pour les ML eux-mêmes.

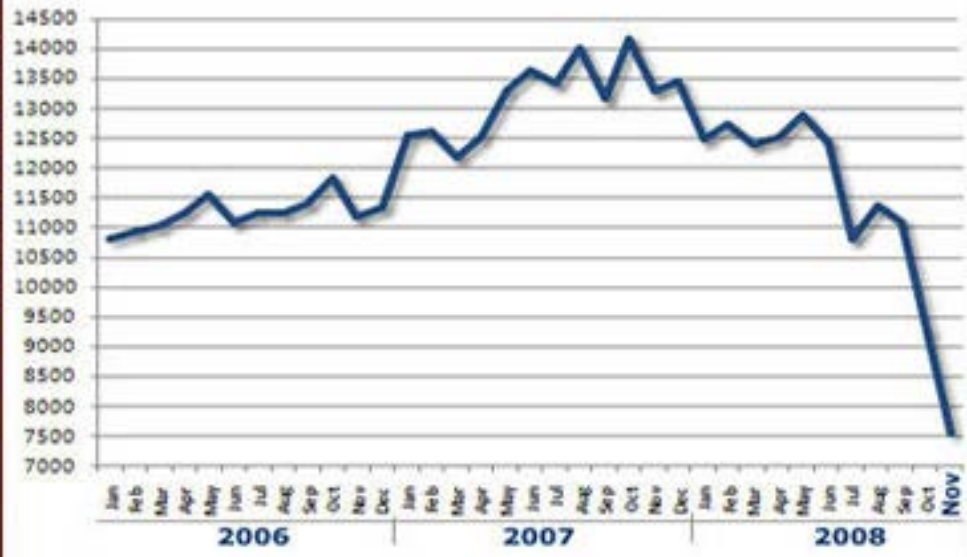
Comment le rendre **interprétable** ?

4/ Probabilités: Les datasets d'apprentissage sont essentiels pour apprendre

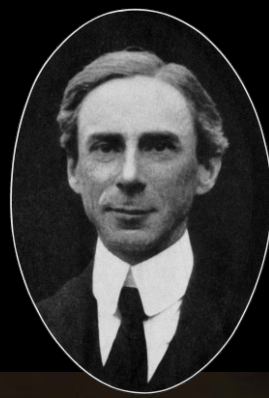
Dès lors, le passé gouverne le présent et peut conduire à **l'ignorance de l'ignorance**

L'INTERPRÉTABILITÉ DU MACHINE LEARNING

Dow Jones Industrial Average Jan 2006 - Nov 2008



World-Crisis.net



Le risque de désencastrement de l'IA

La nouvelle époque:
Big Data et IA se
désencastrent comme
l'a fait l'économie
(Polanyi)
= hors des organisations
et des interprétations
sociales
= autoréférentielles et
non interprétables

Medium Sign

LeCun vs Rahimi: Has Machine Learning Become Alchemy?



The image shows a screenshot of a Medium article. At the top, the word 'Medium' is visible in a large, bold font. To its right, there is a small 'Sign' button. Below the header, the article title 'LeCun vs Rahimi: Has Machine Learning Become Alchemy?' is displayed in a bold, black font. Underneath the title is a painting of an alchemist. The alchemist is an elderly man with a long white beard, wearing a brown robe and a cap. He is holding a small, glowing yellow flask in his right hand, looking at it intently. The background is a dimly lit workshop filled with various alchemical apparatuses, including a large round flask on a stand, a mortar and pestle, and several jars on shelves. The overall atmosphere is one of mystery and ancient knowledge.

Mais l'interprétabilité est possible!

The Annals of Applied Statistics

2015, Vol. 9, No. 3, 1350–1371

DOI: 10.1214/15-AOAS848

© Institute of Mathematical Statistics, 2015

INTERPRETABLE CLASSIFIERS USING RULES AND BAYESIAN ANALYSIS: BUILDING A BETTER STROKE PREDICTION MODEL

BY BENJAMIN LETHAM^{*,1}, CYNTHIA RUDIN^{*,1}, TYLER H. MCCORMICK^{†,2}
AND DAVID MADIGAN^{‡,3}

*Massachusetts Institute of Technology**, *University of Washington*[†]
and Columbia University[‡]

We aim to produce predictive models that are not only accurate, but are also interpretable to human experts. Our models are decision lists, which consist of a series of *if...then...* statements (e.g., *if high blood pressure, then stroke*) that discretize a high-dimensional, multivariate feature space into a series of simple, readily interpretable decision statements. We introduce a generative model called Bayesian Rule Lists that yields a posterior distribution over possible decision lists. It employs a novel prior structure to encourage sparsity. Our experiments show that Bayesian Rule Lists has predictive accuracy on par with the current top algorithms for prediction in machine learning. Our method is motivated by recent developments in personalized medicine, and can be used to produce highly accurate and interpretable medical scoring systems. We demonstrate this by producing an alternative to the CHADS₂ score, actively used in clinical practice for estimating the risk of stroke in patients that have atrial fibrillation. Our model is as interpretable as CHADS₂, but more accurate.

Interpretable Machine Learning



@ChristophMolnar