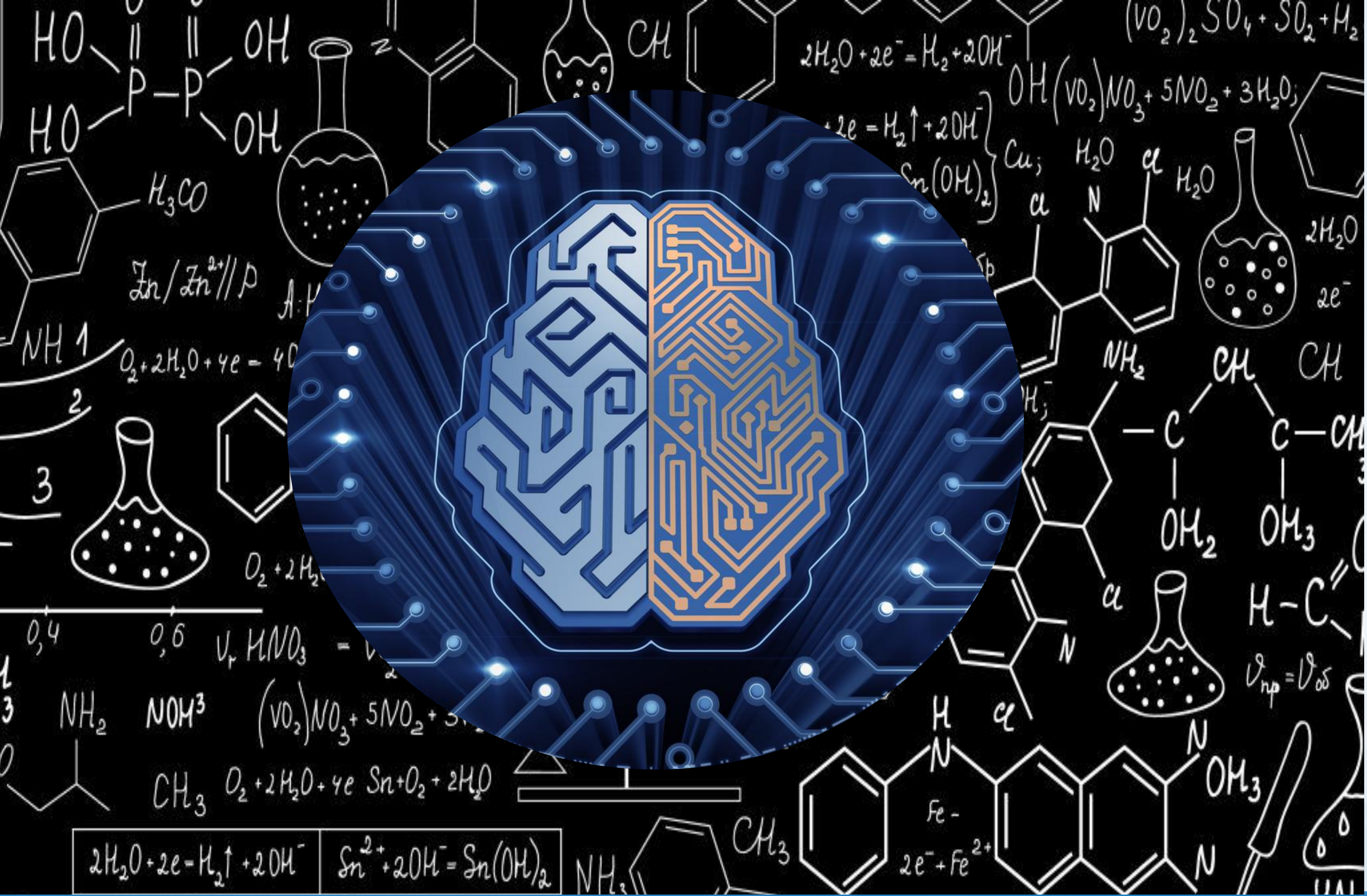


Materials Discovery at Solvay
Applying AI tools to 150 years of historical data

Jean Yves Delannoy - AI/ML/Simulation Manager - Research

07/12/2018

SOLVAY

asking more from chemistry®

CORPORATE RESEARCH & INNOVATION

WE ARE BUILDING A MODEL OF SUSTAINABLE CHEMISTRY TO MEET THE CHALLENGES OF SOCIETY

SOLVAY

asking more from chemistry®

WE ARE A MULTI-SPECIALTY CHEMICAL COMPANY

24,500 employees[1]

61 countries[1]

124 industrial sites[1]

21 major R&I centers[1]

0.65 occupational accidents at Group sites per million hours worked[2]

€ 10.1 billion of net sales[1]

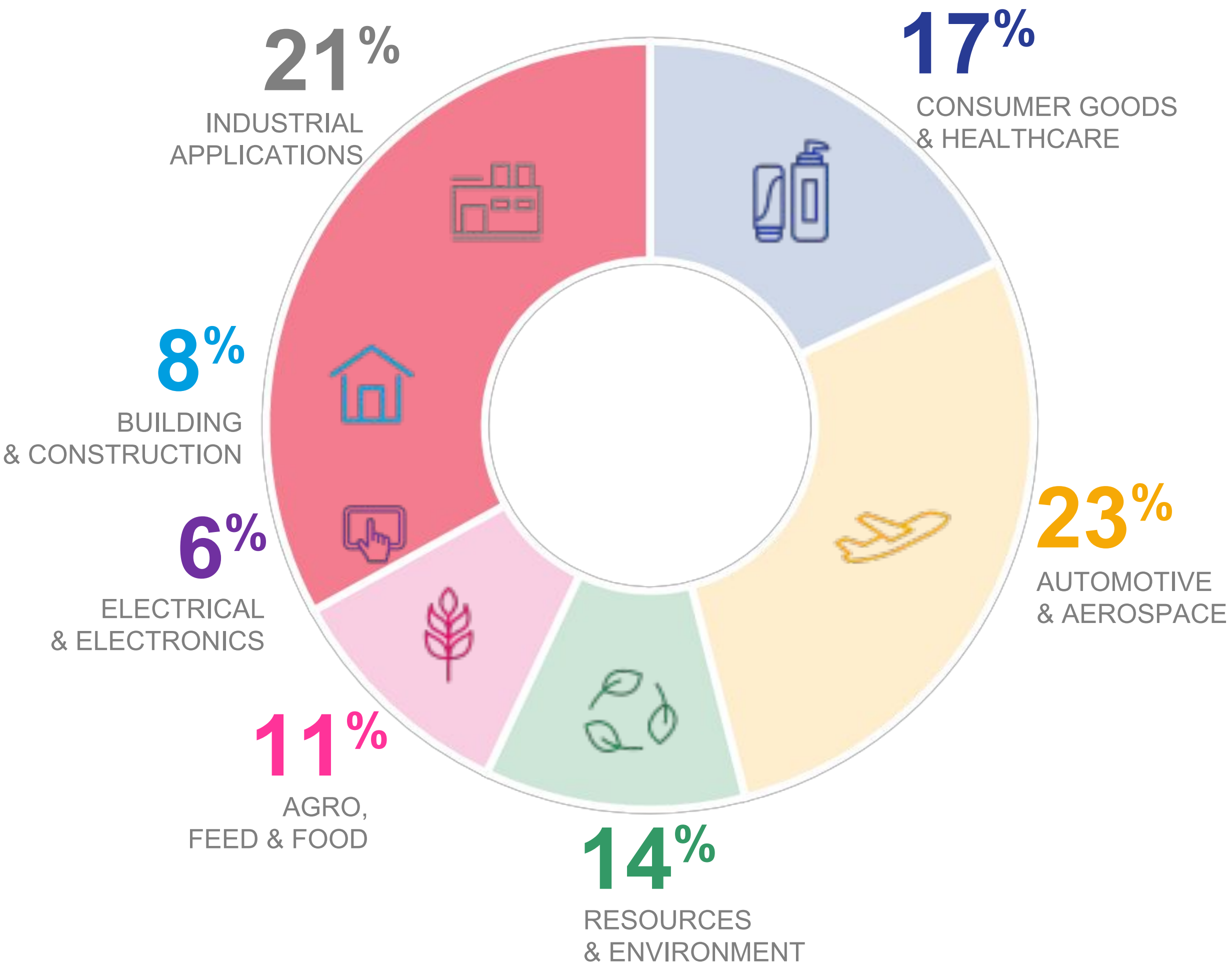
€ 2.2 billion of EBITDA[1]

5.53 greenhouse gas intensity
kg CO_2 eq. per € EBITDA

49% sustainable solutions
Group net sales

1. 2017 underlying results
2. MTAR: Medical Treatment Accident Rate

WE ADAPT OUR PRODUCT OFFERING TO FAST-EVOLVING MARKETS



Distribution of net sales

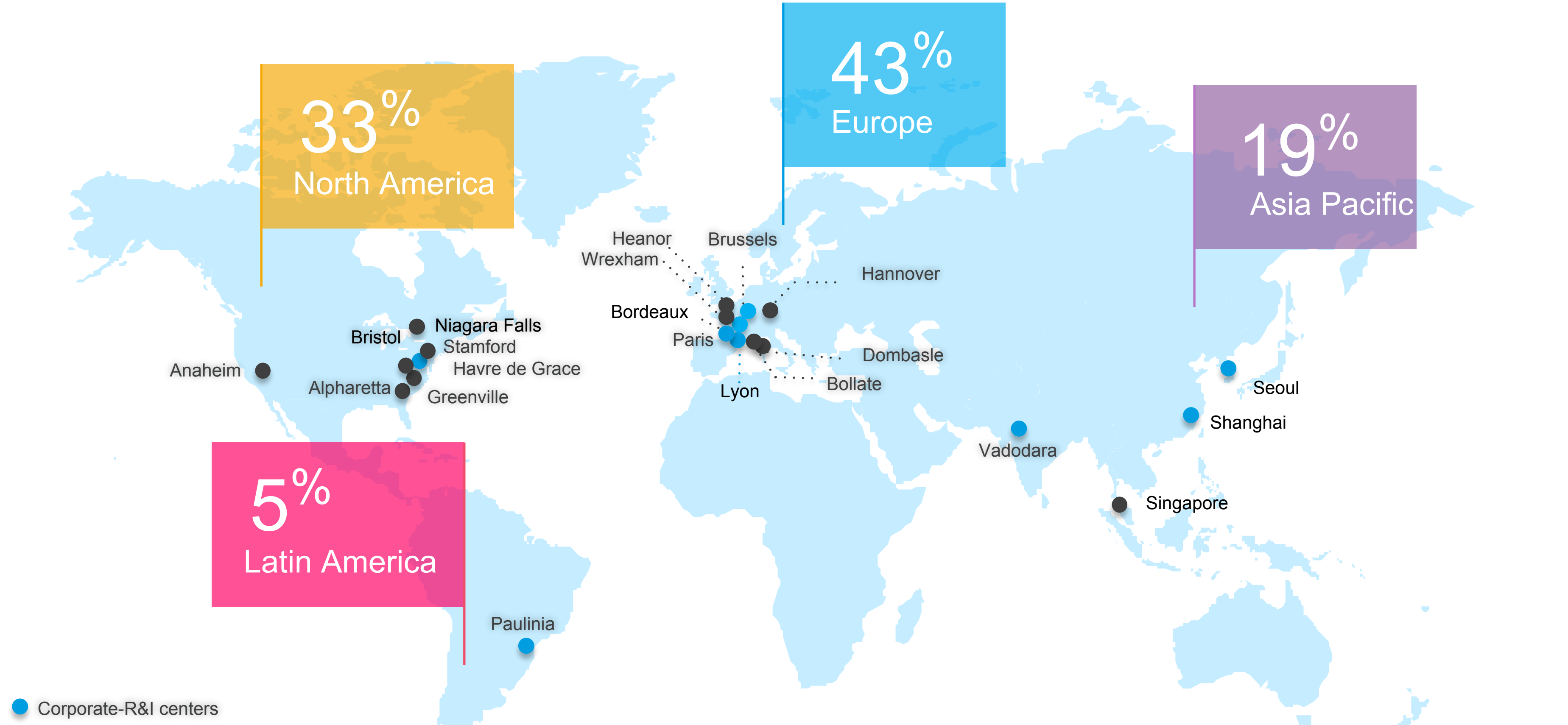
SPIRIT OF INNOVATION

Professor Ben Feringa
Laureate of the Chemistry for the Future in 2015, was awarded the 2016 Nobel Prize in Chemistry for his groundbreaking work on molecular motors

1911 | **Ernest Solvay established first prestigious meetings of top scientist as the Council of Physics**

First round-the-world solar flight
Solar Impulse 2 accomplished the first trip around the world without a single drop of fuel

OUR GLOBAL R&I FOOTPRINT & EXPERTISES



● Corporate-R&I centers

Fields of expertise

Chemistry
Organic & inorganic chemistry, catalysis, nano-material synthesis, polymer synthesis

Process
Chemicals engineering, environment, science, process safety, pilots

Enabling technologies
Analysis, characterization simulation, digital, high throughput technologies & microfluidics

Material science

Material processing

Soft matter & formulation

Biotechnology

AUTOMOTIVE AND AEROSPACE

Cleaner mobility

LIGHTWEIGHTING

Lightweight materials (high-performance polymers, advanced composite materials, etc.) for ligther vehicles (SolvaLite^{TM} , Tegracore^{TM}).

POWERTRAIN EFFICIENCY

Products (fluorinated elastomers, polymers, etc.) improve the motor longevity (Nocolok® Flux, Tecnoflon®).

ELECTRIFICATION

Flame-retardant materials and heat-resistant engineering plastics improve the lifespans of hybrid and electric vehicles (Solef® PVDF, LiTFSI salts, Amodel® PPA).

GREEN TECHNOLOGIES

Catalytic materials and highly dispersible silica, limit polluting emissions and fuel consumption (Premium SW, Optalys®).

We help manufacturers meet the challenges of sustainable mobility.

RESOURCES AND ENVIRONMENT

Affordable resources and environment protection

Our proven expertise in the oil & gas, mining and energy sectors enables us to develop eco-friendly solutions.

OIL AND GAS

- Solutions based on guar and on surfactants increase yields and limit the environmental impact of drilling.
- High-performance polymers (Solef® PVDF) for improved operating efficency.

MINING

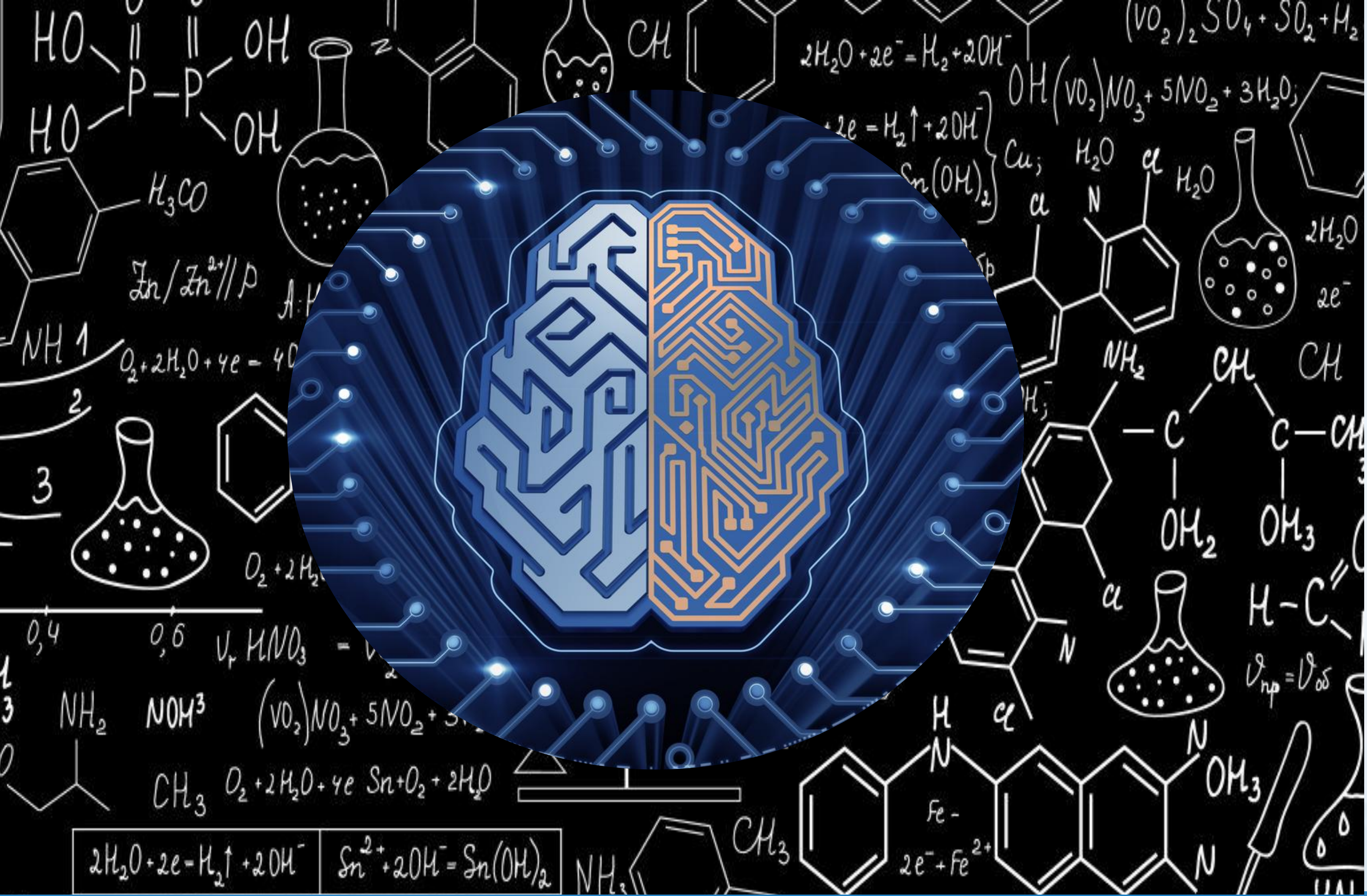
Chemical reagents improve customers' productivity and operating costs of the recovery of metals and minerals (Interox®).

ENERGY SOLUTIONS

Products and technologies for producing and storing renewable energies, and improving energy efficiency (Halar® ECTFE, LiTFSI salts).

ENVIRONMENTAL PROTECTION

- Solutions for air and water treatment using filtration, gas separation, absorption, and chemical reactions (Udel® PSU, Interox®).
- Range of products and systems for controlling air emission and managing associated waste (SOLVAIR Solutions®).



Materials Discovery at Solvay

Applying AI tools to 150 years of historical data

Jean Yves Delannoy - AI/ML/Simulation Manager - Ressearch

07/12/2018

Why is digital key for Solvay’s innovation and what is our ambition?

Why Digital?

Digital can affect or change drastically:

- Market analysis
- Competitive intelligence
- Data acquisition
- Data analysis
- Creation of knowledge
- Creation of new business models

What is our ambition for R&I

We want to:

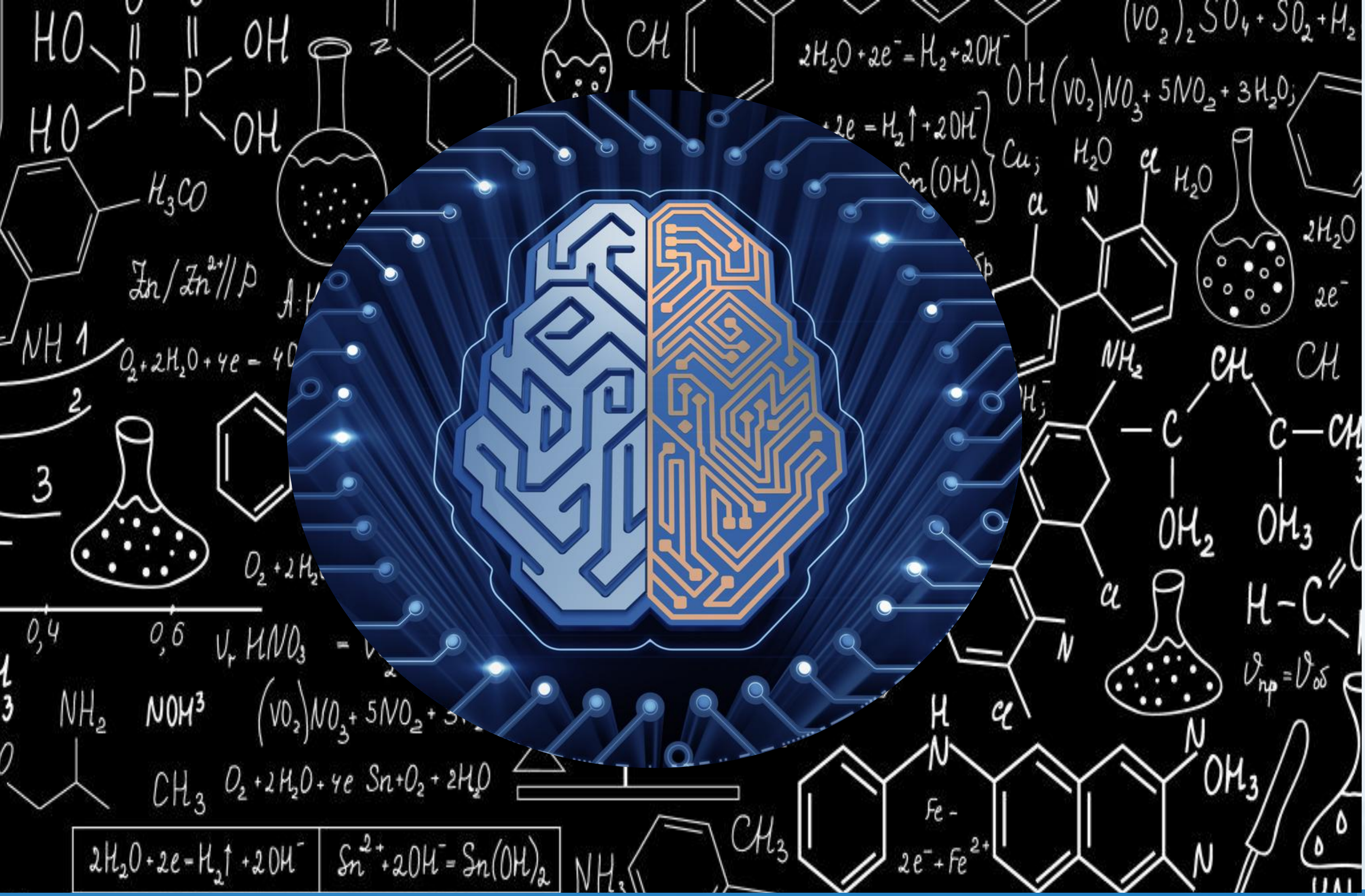
- Create a step-change in our labs efficiency (more focused experiments, more simulations upfront, more productivity...)
- Transform the way we get information on literature and competitive intelligence
- Transform the way we interact with customers and create new business models
- Create knowledge by pulling more value through data analysis and AI

Digital is a game changer in the way chemical companies can conduct their innovation.

→ We want to create a step change in the efficiency of our lab operations

→ We want to transform the way we access and create information.

→ We want to transform the way we generate value from data



Avenues for Collaboration and developments

What do we need to do ?

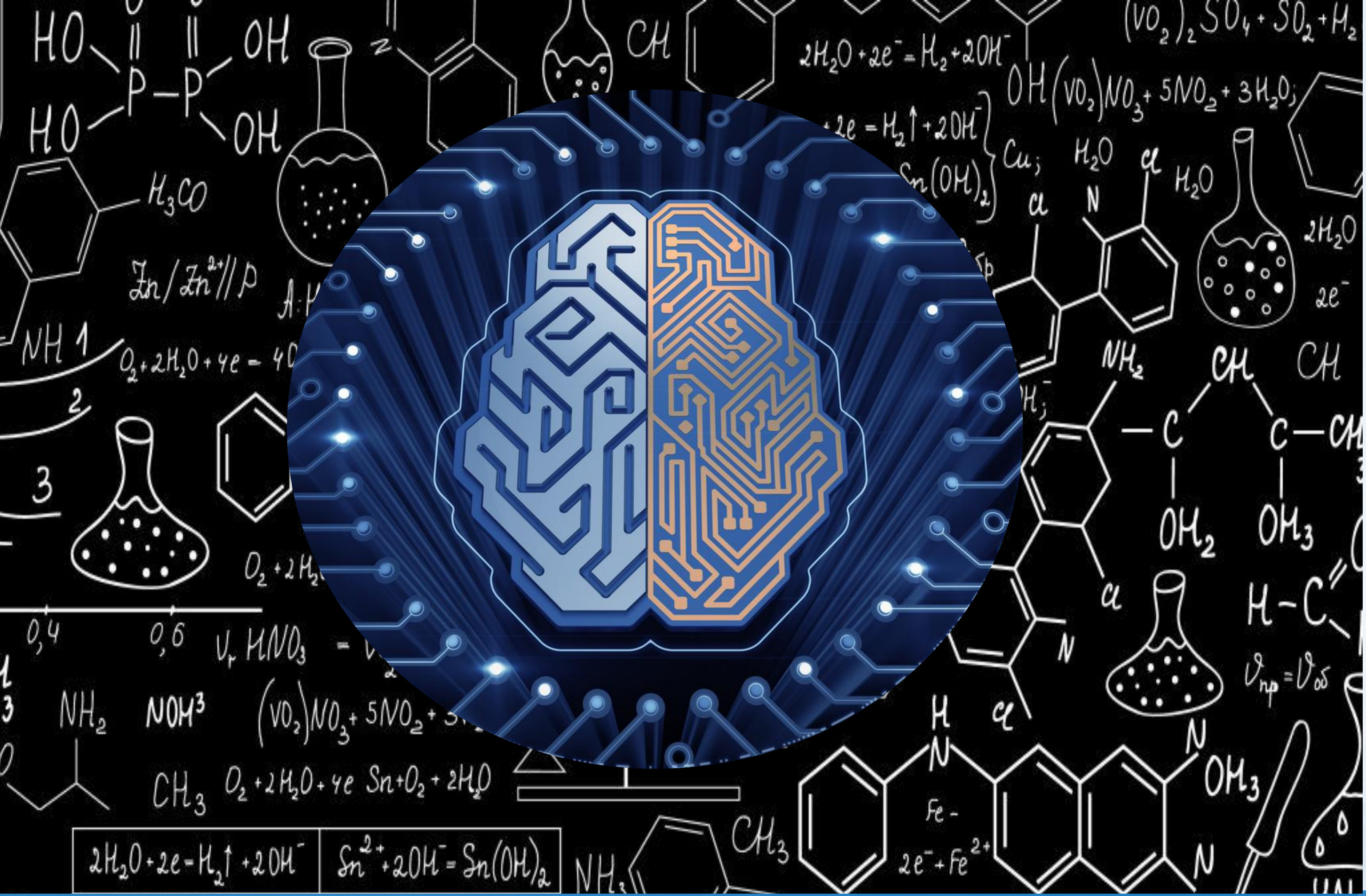
In 5 years, every chemists will have on their bench a chatbot to suggest the next experiment

Materials and Energy using data analytics

- **Saving energy by reducing experiments.**
  - Leveraging from the Past
  - Preparing the future
  - Facilitating the use of data.
- **Saving energy by improving manufacturing efficiency**
- **Saving energy by improving materials**
  - Leightweighting
  - Improving Efficiency
  - Electrification & Batteries
  - Green Technologies

Data Analytics and Data Management needs

- **Data Platform for democratization of data analytics**
- **Semantic Approach : Speaking Chemistry and Materials**
- **Development of new materials with more efficient properties.**
- **Development of Green/biodegradable materials**
- **Market Analysis (not discussed today)**
- **Manufacturing excellence (not discussed today)**



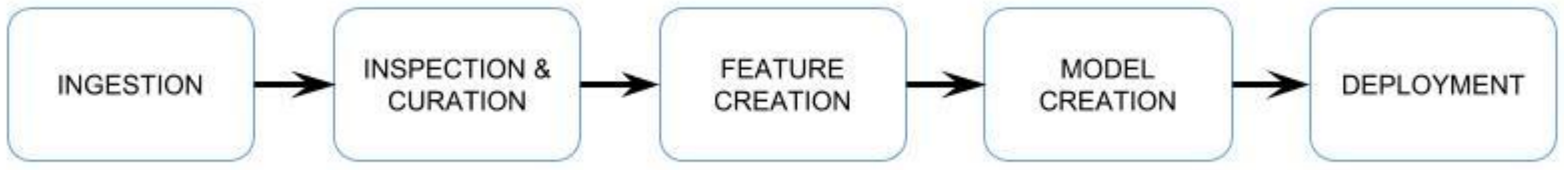
Some realizations & some needs

07/12/2018

Machine Learning ?

Data Platform for democratization of data analytics

Data Platform for democratization of data analytics



INGESTION → INSPECTION & CURATION → FEATURE CREATION → MODEL CREATION → DEPLOYMENT

The technological bricks necessary to get a democratic platform are :

- Data Ingestion : different data formats and recognition of chemical entities in different contexts

⇒ Uniformization of experimental machine data format ?

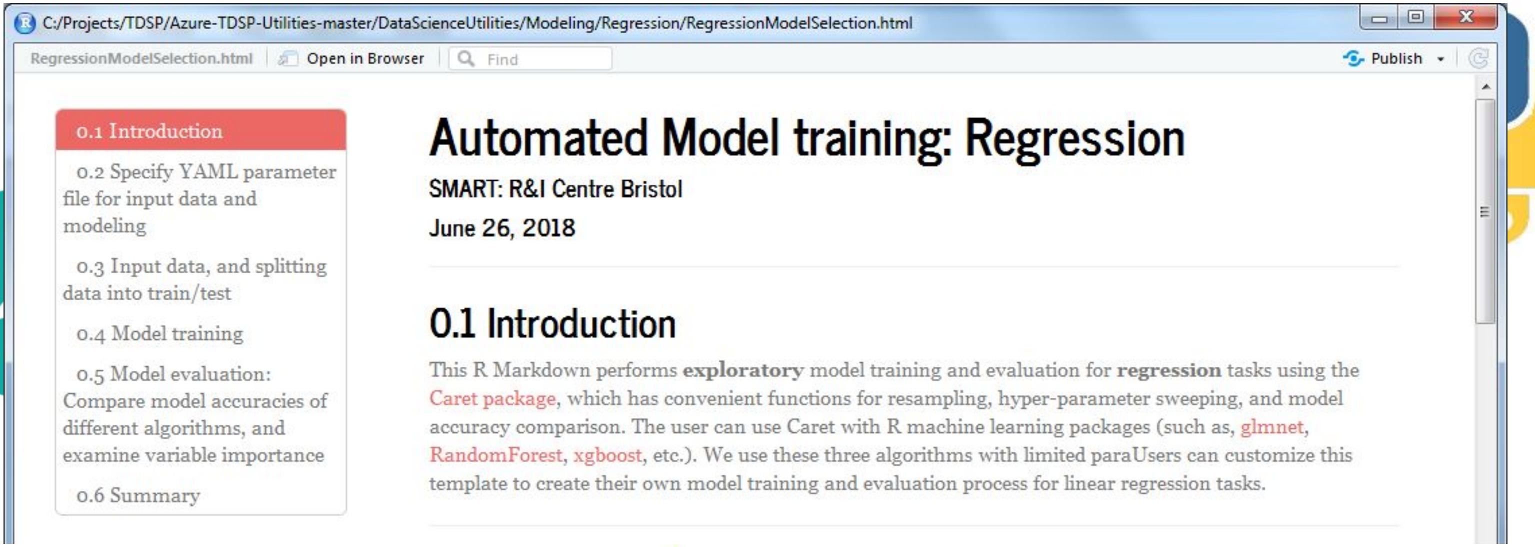
- Data Inspection & Cleaning (curation)
- Feature creation : Automatic feature extraction and processing. This is where chemistry, material science or formulation expertise are involved. The platform should “understand” Chemistry and Materials.
- Build & Evaluate Models : Auto machine learning will be an option to speed up the work of non-data scientists when the previous steps are already known
- Model Deployment : it should be easy, when a model has be created, to transfer it onto an interface

⇒ This is a non competitive issue. help is welcomed

Data Platform for democratization of data analytics

R

CITRINE
INFORMATICS


C:/Projects/TDSP/Azure-TDSP-Utilities-master/DataScienceUtilities/Modeling/Regression/RegressionModelSelection.html
RegressionModelSelection.html
Open in Browser
Find
Publish

- 0.1 Introduction
- 0.2 Specify YAML parameter file for input data and modeling
- 0.3 Input data, and splitting data into train/test
- 0.4 Model training
- 0.5 Model evaluation: Compare model accuracies of different algorithms, and examine variable importance
- 0.6 Summary

Automated Model training: Regression

SMART: R&I Centre Bristol

June 26, 2018

0.1 Introduction

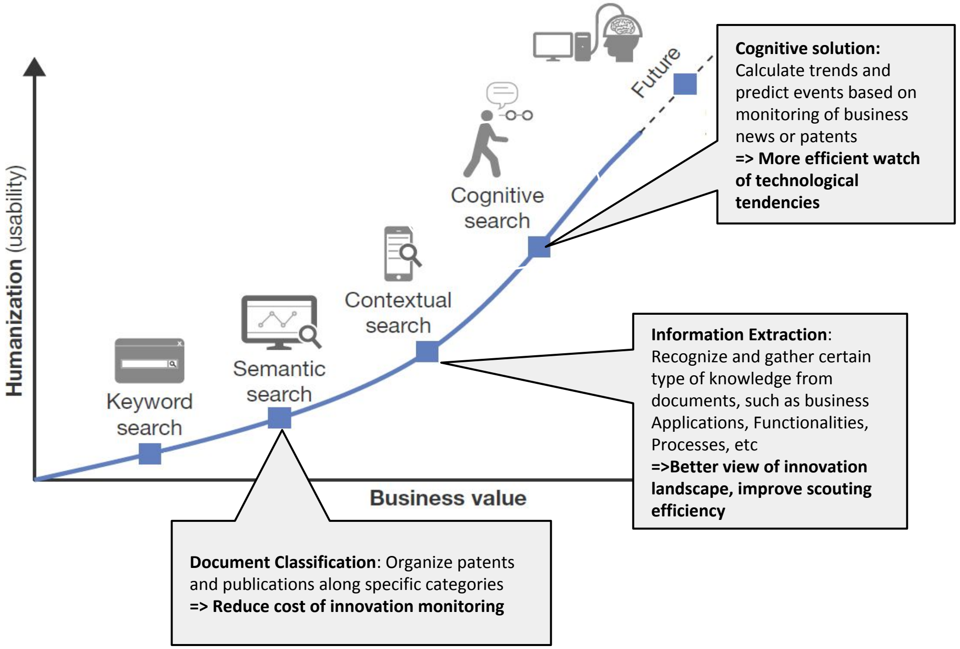
This R Markdown performs **exploratory** model training and evaluation for **regression** tasks using the Caret package, which has convenient functions for resampling, hyper-parameter sweeping, and model accuracy comparison. The user can use Caret with R machine learning packages (such as, glmnet, RandomForest, xgboost, etc.). We use these three algorithms with limited paraUsers can customize this template to create their own model training and evaluation process for linear regression tasks.

KNIME

Semantic Approach : Speaking Chemistry and Materials

Semantic Approach : Speaking Chemistry and Materials

- **The basics is not there :**
  - How do we identify in a document
    - What is an Table
    - What is a graph.
    - What is a chemical formula ?
    - What is a picture ?
  - How do we pre organize and structure unstructured data ?



Cognitive solution: Calculate trends and predict events based on monitoring of business news or patents
=> More efficient watch of technological tendencies

Information Extraction: Recognize and gather certain type of knowledge from documents, such as business Applications, Functionalities, Processes, etc
=>Better view of innovation landscape, improve scouting efficiency

Document Classification: Organize patents and publications along specific categories
=> Reduce cost of innovation monitoring

- **Technological Bricks :** Building a cognitive tool requires
  - choosing a Search Engine,
  - adding concept Extractors
  - Recognizing specific knowledge inside documents (locations, companies, people, compounds...).
  - ….

=> We need to leverage on our history

- Internal documents
- Patents
- Publications
- ….

UTILITIES
DATA SOURCING
INFORMATION RETRIEVAL
INFORMATION EXTRACTION
QUESTION ANSWERING
VISUALIZATION

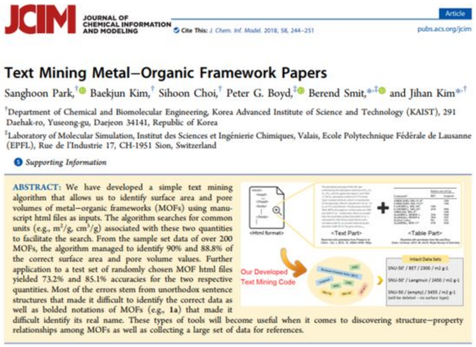
Help is welcomed :

★ **Systematically structure unstructured data**
★ **Digest large corpus and extract the key elements.**
★ **……**

Development of new materials with more efficient properties.

Approach

literature,
patents,
websites, …



JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

Cite This: *J. Chem. Inf. Model.* 2018, 58, 244–251

pubs.acs.org/jcim

Text Mining Metal–Organic Framework Papers

Sanghoon Park,[†] Baekjun Kim,[†] Sihoon Choi,[†] Peter G. Boyd,[‡] Berend Smit,[*,‡] and Jihan Kim[*,†]

[†] Department of Chemical and Biomolecular Engineering, Korea Advanced Institute of Science and Technology (KAIST), 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea

[‡] Laboratory of Molecular Simulation, Institut des Sciences et Ingénierie Chimiques, Valais, Ecole Polytechnique Fédérale de Lausanne (EPFL), Rue de l'Industrie 17, CH-1951 Sion, Switzerland

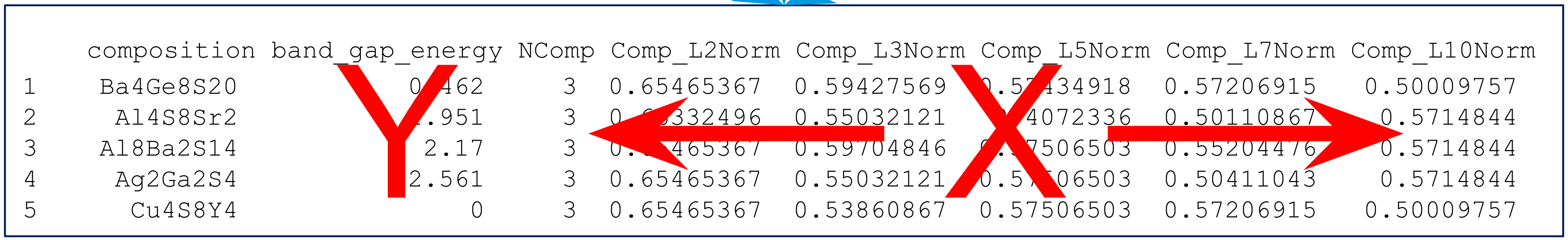
Supporting Information

ABSTRACT: We have developed a simple text mining algorithm that allows us to identify surface area and pore volumes of metal–organic frameworks (MOFs) using manuscript html files as inputs. The algorithm searches for common units (e.g., m^2/g, cm^3/g) associated with these two quantities to facilitate the search. From the sample set data of over 200 MOFs, the algorithm managed to identify 90% and 88.8% of the correct surface area and pore volume values. Further application to a test set of randomly chosen MOF html files yielded 73.2% and 85.1% accuracies for the two respective quantities. Most of the errors stem from unorthodox sentence structures that made it difficult to identify the correct data as well as bolded notations of MOFs (e.g., **1a**) that made it difficult identify its real name. These types of tools will become useful when it comes to discovering structure–property relationships among MOFs as well as collecting a large set of data for references.

<Html format>
<Text Part>
<Table Part>
Our Developed
Text Mining Code
Intact Data Sets
SNU-50' / BET / 2300 / m2 g-1
SNU-50' / Langmuir / 2450 / m2 g-1
SNU-50 / (empty) / 3455 / m2 g-1

local
data
repositories

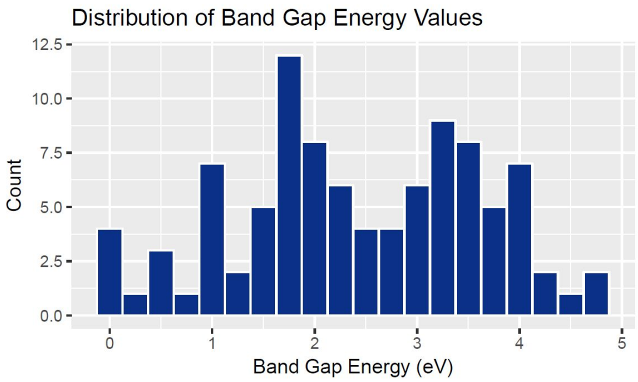
global
(distributed)
databases



```
     composition band_gap_energy NComp Comp_L2Norm Comp_L3Norm Comp_L5Norm Comp_L7Norm Comp_L10Norm
1     Ba4Ge8S20            0.[illegible]462     3  0.65465367  0.59427569  [illegible].57[illegible]34918  0.57206915   0.50009757
2      Al4S8Sr2            [illegible].951     3  0.[illegible]332496  0.55032121  [illegible]4072336  0.50110867    0.5714844
3     Al8Ba2S14             2.17     3  0.[illegible]465367  0.59704846  [illegible]7506503  0.55204476    0.5714844
4      Ag2Ga2S4            2.561     3  0.65465367  0.55032121  0.57[illegible]06503  0.50411043    0.5714844
5       Cu4S8Y4                0     3  0.65465367  0.53860867  0.57506503  0.57206915   0.50009757
```

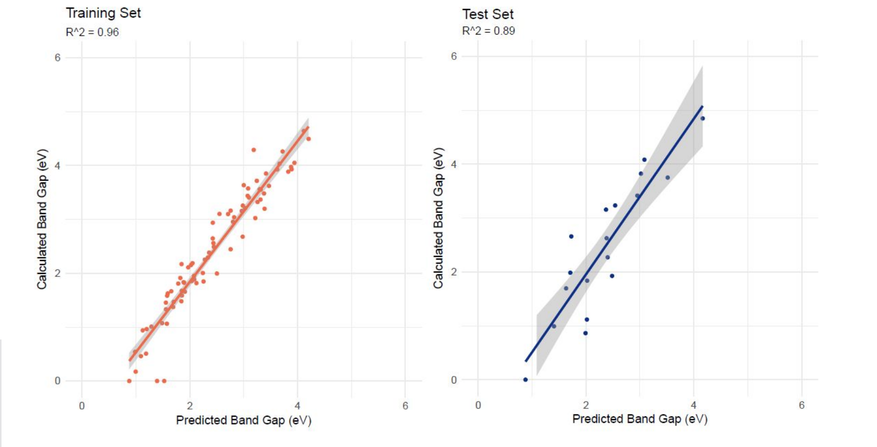
QDots: Band Gap Energy

- **Problem** : Modeling DFT Calculations using Random Forests



- **Approach**
  - Data
    - Wein2k dataset
    - DFT calculation results for 97 compounds
    - endpoint of interest = band gap energy
  - Features
    - 145 features calculated / molecule using *Magpie* (Materials
    - Agnostic Platform for Informatics and Exploration)
    - 37 features / molecule retained following feature curation
  - Model = Random Forest

- **Results**



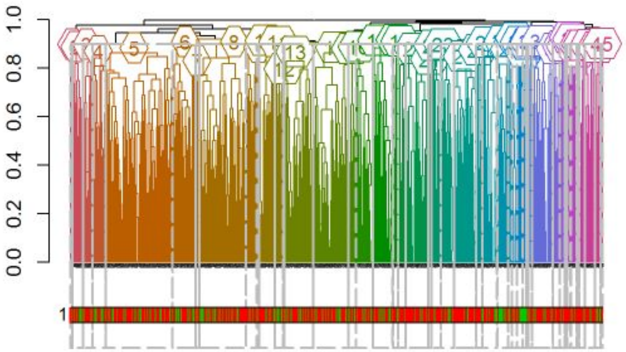
Very Predictive Models that can be combined with DFT to understand the microscopic origin
=> Approach to generalize

Development of Green/biodegradable materials

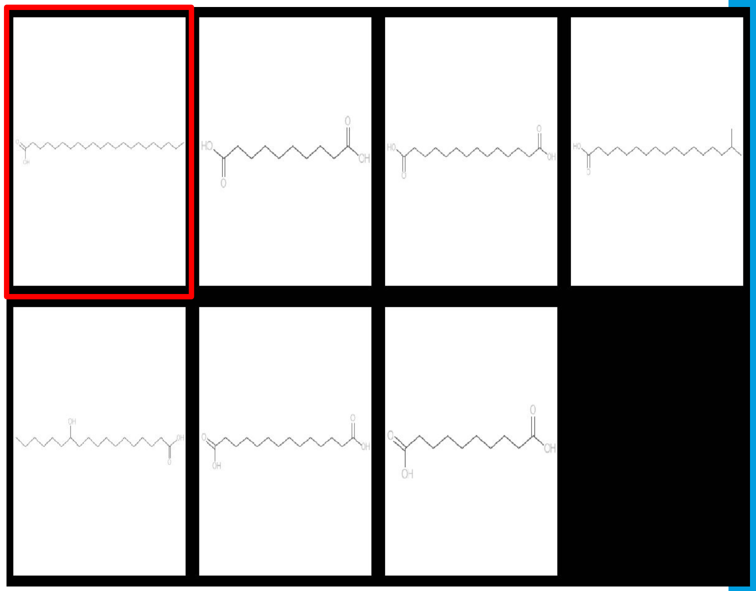
Development of Green/biodegradable materials

- **Question** : How can predict if a new molecule is gonna be biodegradable ?

- **Approach**
  - Internal developments.
  - Data collection combined public and internal databases
  - Use of Similarities and clustering



Non-biodeg

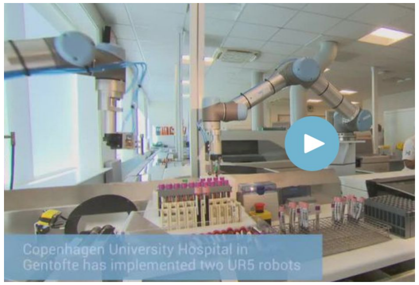


- **Results**
  - Simple connection/correlation established
  - Recommendations to researchers when dealing with new, or untested molecules.

Our Needs

Needs

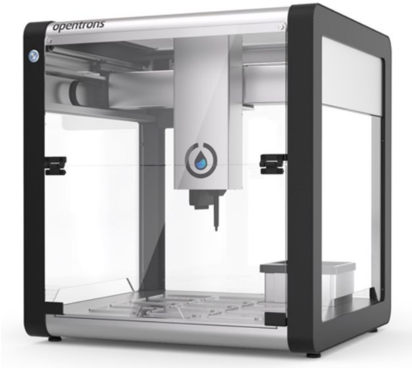
- **Needs of a Digital Platform that speaks Materials and Chemistry**

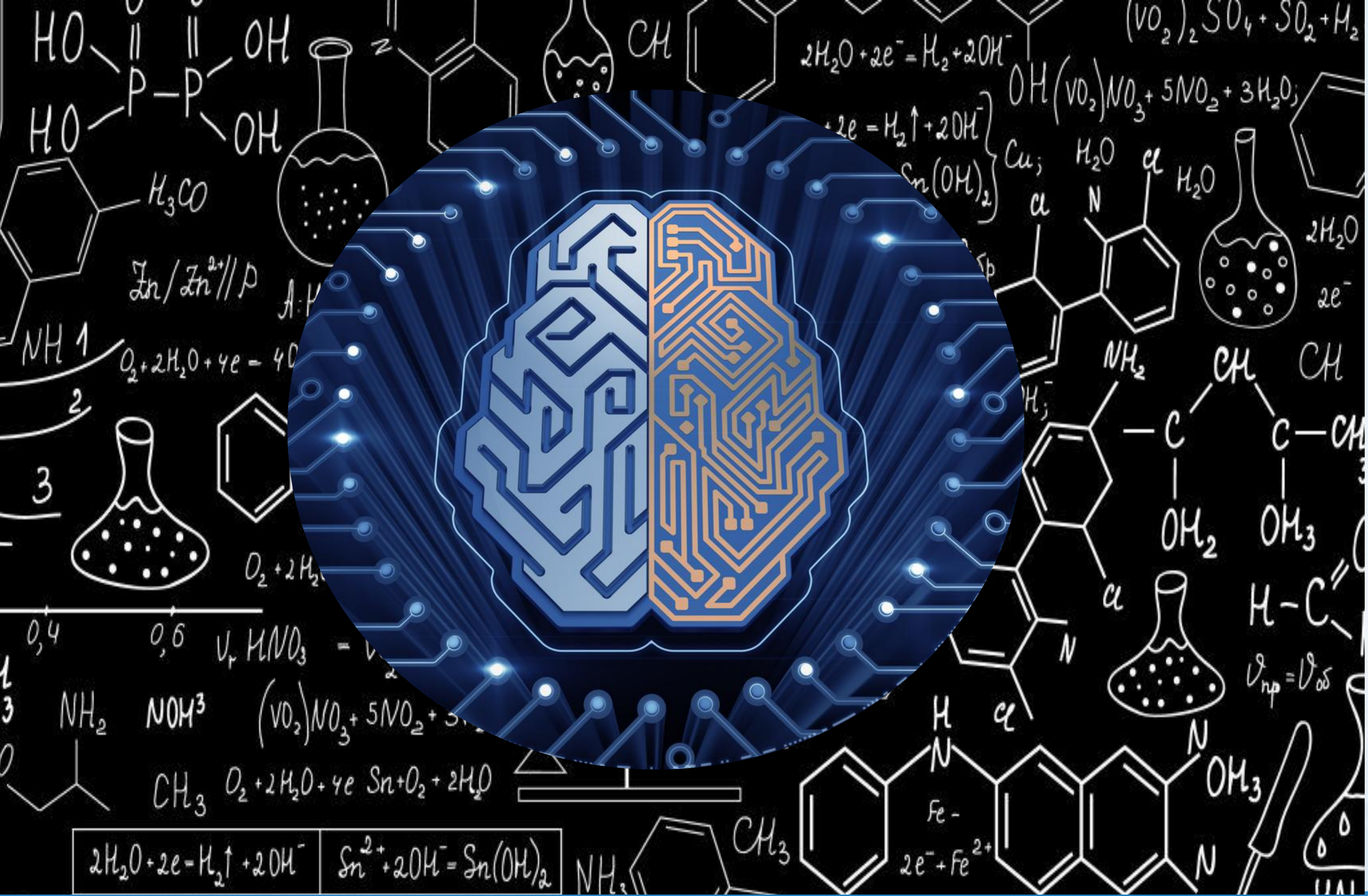

Copenhagen University Hospital in Gentofte has implemented two UR5 robots

- **Needs in Data Sources**
  - Access to more litterature data
  - Uniformization of Data sources (from experiments)
  - Trustable repository of public data.
  - Semantic developments

- **Needs to develop Iterative learning**
  - Connect experiments to Data analytics tools
  - Improve robotization and automatization.
  - Connect more molecular modeling with Data analytics to improve understanding.

- **Work on non competitive issues like sustainability**





Conclusions

Conclusions

- Data transformation is a key enabler for the future of Solvay

- Some projects are confidential and deal with IP protected Data but there is room for improvement in many aspects of competences developments :
  - Data Platform
  - Semantic Approach : Speaking Chemistry and Materials
  - Uniformization of Data format or Development of universal translators

- Our ambition is to use data to develop better and more sustainable materials.