FIRMENICH D-LAB

AI FORMULA GENERATOR

DR GUILLAUME GODIN

SCIENTIFIC DIRECTOR AI

IDSIA LUGANO, 2ND SCHOOL AIDD 2022

FIRMENICH TODAY
FAST FACTS

4.3
Billion CHF
Revenues

+16.8%
Including acquisitions at constant currency

+4.7%
On organic basis at constant currency

#1
Privately-owned Fragrance & Taste company

#1
Perfumery & Ingredients*

14
Acquisitions Last 5 Years

10%
Invested on average in R&D (FY17-FY21)

4,000+
Patents

Nobel
Prize
1939

10,000
Colleagues

100+
Markets

Started in 1895 at Geneva

*ranked #1 as co-leader

PERFORMANCE
Industry Leader in Fragrance & Taste

FIRMENICH
A LEGACY OF:

RESPONSIBILITY
Role Model Company in ESG Credentials

WORLD-CLASS SCIENCE:
Leveraging Science Leadership for the Best of Society

PERFUMERY

Fine Fragrance

Consumer Fragrances

WHAT WE DO

POWERED BY WORLD-CLASS SCIENCE & INNOVATION

INGREDIENTS

Naturals

Molecules

TASTE & BEYOND

Beverages

Sweet Goods

Savory

THE RIGHT MOMENT

Revolutionizing the Industry, Together

50+ YEARS — GENERATING DATA

11 YEARS — DIGITAL DEVELOPMENT

7 YEARS — ENHANCED CREATION TOOLS

4 YEARS — d-lab EPFL

DIGITAL ECONOMY AWARD

24 YEARS
Chem. Eng.
Norvatis
Vienna

22 YEARS
PhD
Iminosugars synthesis
Chaperone therapy
Orléans

19 YEARS
EPFL
In vivo Fluorescent biomarkers
Kai Johnsson
Lausanne

16 YEARS
Firmenich
J-M Lehn
A Herrmann
Pro-perfumes
Geneva

14 YEARS
ELN by CAMBRIDGESOFT (ChemDraw…)
Geneva

11 YEARS
PM Optimization
Reformulation Computer Assisted (ORCA)
Geneva

Few Open-source contributions

- **All 3D Dragon like descriptors RDKit C++** **https://github.com/rdkit/rdkit**
- Smiles randomization RDKit C++
- Coulumb Matrix RDKit C++
- HOSE code invariant type in RDKit C++
- * RDKit javascript (rebranded last year minimalLib) https://github.com/rdkit/RDKitjs
- YEeHMOP + EEM atomic "Partial Charges" RDKit Fortran / C++
- GetAtomFeatures for graph (GIN version) RDKit C++
- Pytorch: Attentive FP, DimenetPP in PyG https://github.com/pyg-team/pytorch_geometric
- **Keras: GIN, AttentiveFP, DimenetPP & PAiNNs Graphs** **https://github.com/aimat-lab/gcnn_keras**

***From Static representation of vectors to Dynamic autoML vector generation**

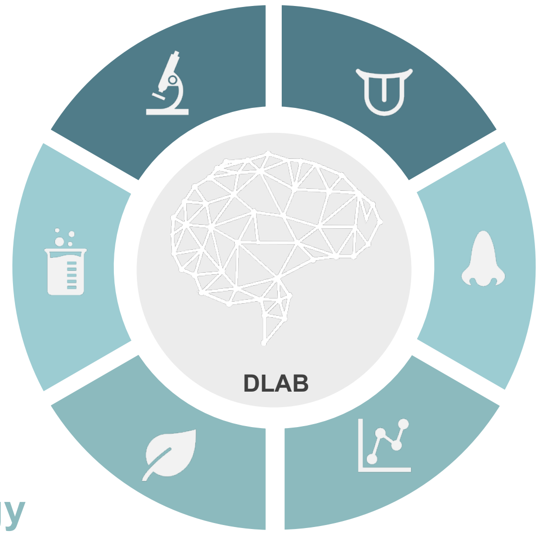
Main digitalization research activities

Ingredient - Chemistry

Properties Predictions

Assisted Chemistry

Assisted Biotechnology



Taste Receptor

Olfactory Receptor

Analytical Chemistry

Formulation

How to find missing/bad data means ?

Ruddigkeit *et al. Journal of Cheminformatics* 2014, **6**:27
http://www.jcheminf.com/content/6/1/27

Journal of Cheminformatics

RESEARCH ARTICLE **Open Access**

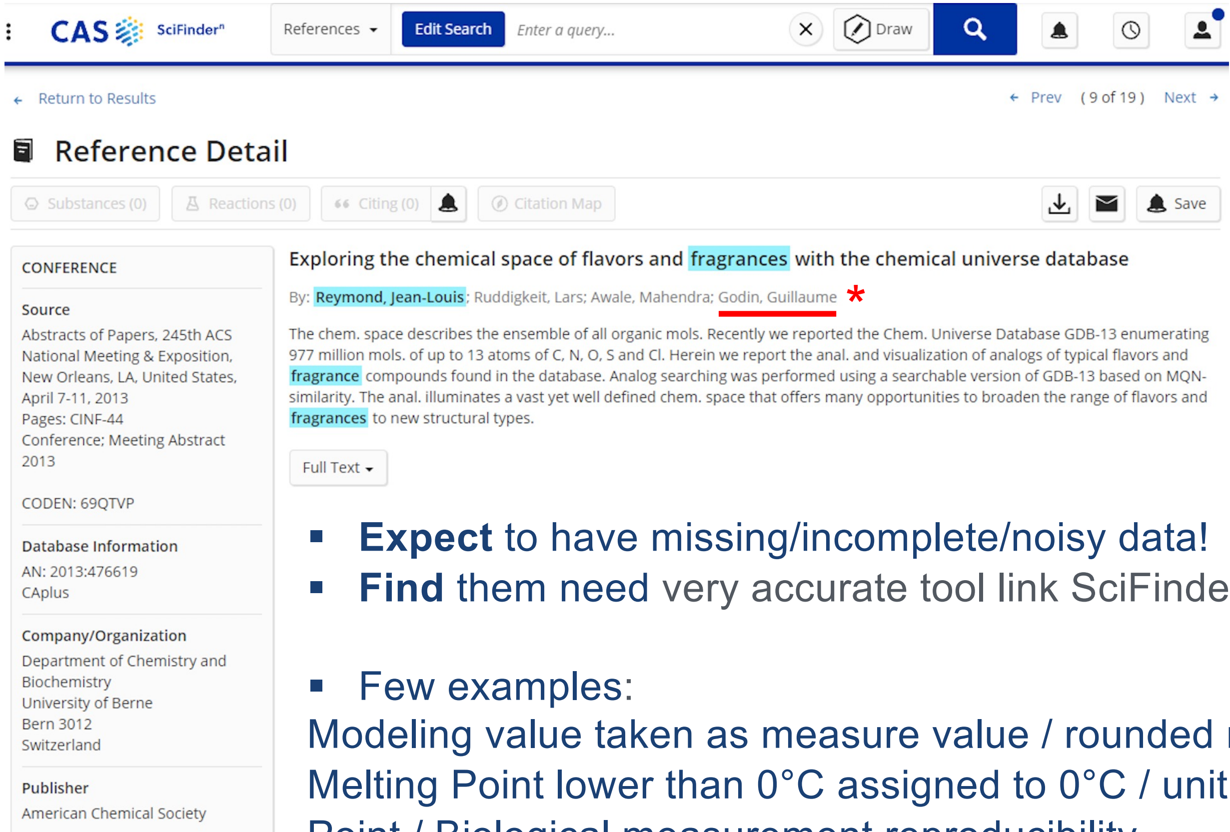
Expanding the fragrance chemical space for virtual screening

Lars Ruddigkeit, Mahendra Awale and Jean-Louis Reymond*

Q: Is list of authors complete ?

In fact, do we know how good is our data ?

How to find missing/bad data means ?



Short Answer: Yes & No

- **Expect** to have missing/incomplete/noisy data!
- **Find** them need very accurate tool link SciFinder

- Few examples:

Modeling value taken as measure value / rounded reaction yield (40% instead of 43% / Melting Point lower than 0°C assigned to 0°C / unit issues / Boiling point bellow Melting Point / Biological measurement reproducibility

Firmenich

**Guillaume GODIN* contribution found in abstract ACS New Orleans conference (spring 2013) before the paper (2014)

Short review of AI / ML fundamentals : Don’t read it, just do it !

Recent Fast Model Architectures revolution

From Mathematical concepts:
1967: K-Nearest Neighbour (**KNN**) averaging similar points
1963/1986: Multilayer perceptron (**MLP**)
1995: Support Vector Machine (**SVM**)
2001: Random forest (**RF**) based on random tree (1996)
2016: Extreme gradient boost (**XGBoost**)

To GAFA + Universities + Companies needs:
Mainly due to hardware accessibility (**GPU/TPU**):
2016: Graph Neural Network (**GCN**) (>100 architectures in 6 years)
2019: Transformers & Natural Language Processing (**NLP**)

Strong methodology == Systematic validation

A good model process:
- Prepare data (inputs, targets) & remove duplicates to avoid data “leaks”
- Always think data can be **noisy**
- ***Split Data***: Train / validation / test
- Compare multiple models (hyperparameters, variation, architecture, etc…)

Pick the correct metric*

- *Classification Metrics (accuracy, precision, recall, F1-score, ROC, AUC, …)*
- *Regression Metrics (MSE, MAE)*
- *Ranking Metrics (MRR, DCG, NDCG)*
- *Statistical Metrics (Correlation)*
- *Computer Vision Metrics (PSNR, SSIM, IoU)*
- *NLP Metrics (Perplexity, BLEU score)*
- *Deep Learning Related Metrics (Inception score, Frechet Inception distance)*

Main issues of AI / ML

Applicability Domain limits:
- Difficult to control, model & database dependent
- Overfitting & non-generalization are strongly correlated
- RF can slightly overfit, while it is a general case of Neural Network (high # of parameters)

User Acceptante:
- RF “explained”
- while NN are generally consider to be black box (recent trend Explainable AI)

Single model stability:
- Ensemble models improve generalization

**https://towardsdatascience.com/20-popular-machine-learning-metrics-part-1-classification-regression-evaluation-metrics-1ca3e282a2ce

OCHEM: Our dedicated solution for easy & robust AI / ML models

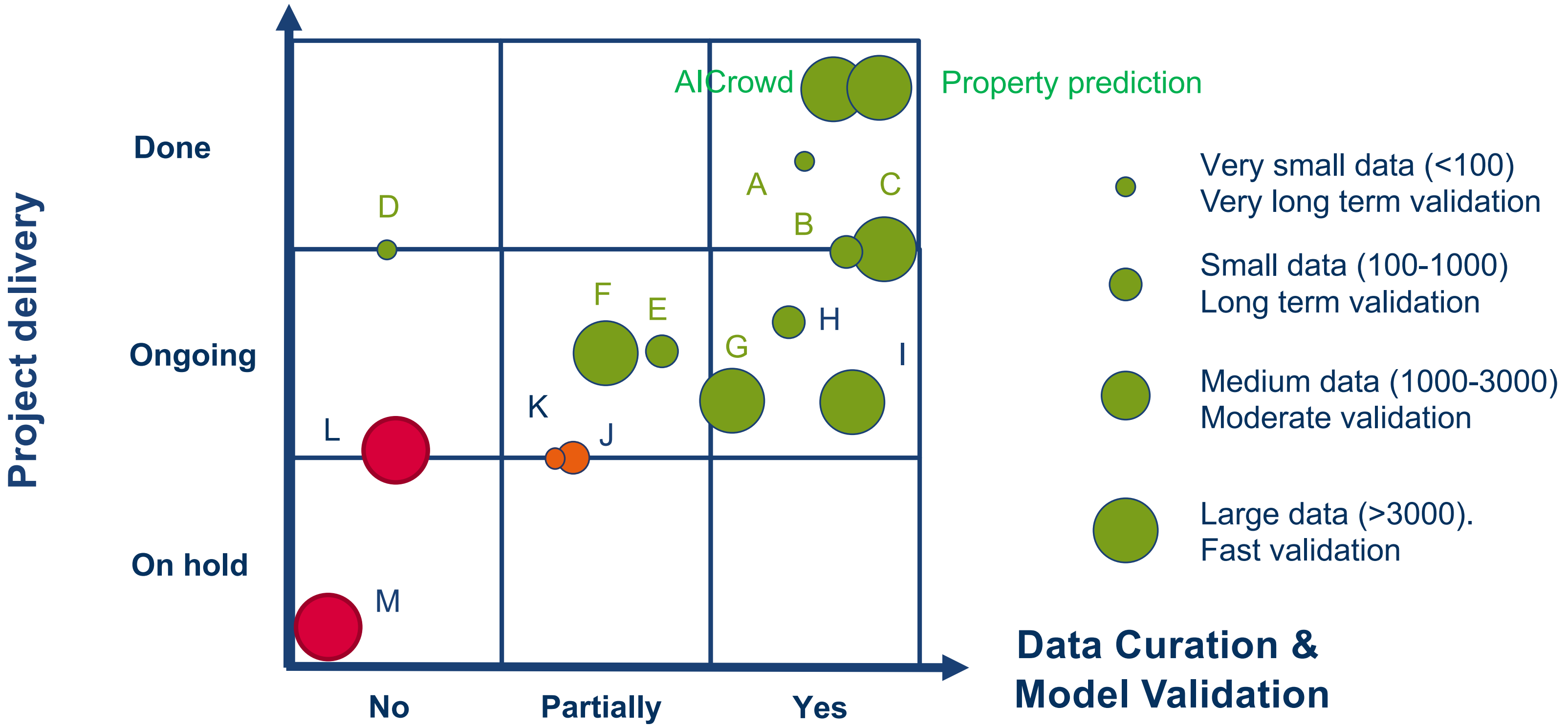
We use open OCHEM* to benchmark models & features

It gives access to

- **lot of architecture**
- **lot of metric**
- **a robust validation protocol**

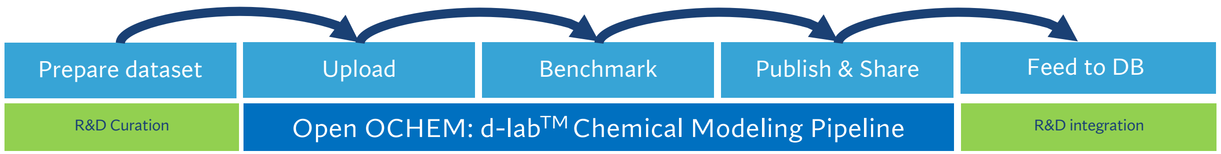
*https://ochem.eu

Relation between quality/curation vs project delivery



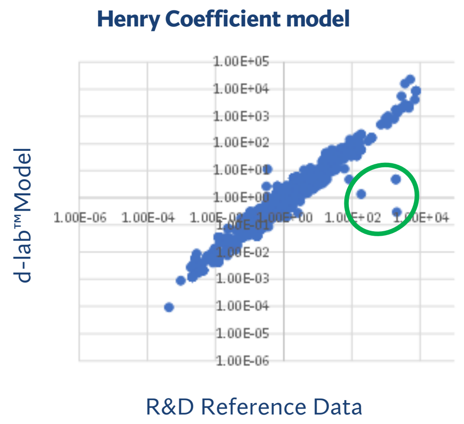
Firmenich

Open OCHEM integration at Firmenich



OCHEM advantages:

- **All in one:**
  - State of the art Firmenich AI architectures
  - Open-source AI architectures
  - Classical ML methods for comparison
- **Trust:**
  - **Fair** performance evaluation using standardized cross validation
- **Robust:**
  - Version control for models & datasets
  - Future data collection can be targeted to improve model performance where it is weak, improving experimental efficiency.



Outliers really?
Human check 3 of 4 are database molecule error

*https://ochem.eu

Firmenich

OCHEM model publication & comparison example

Predicted property: logHenry
Training set: studentmodelcosmo_csv

Metrics RMSE - Root Mean Square Error for Training set Validation: Cross-Validation (11 models)

	ChemProp	GNN GIN	TRANSNNI (F) (3D)	CHEMNLP (3D)	CHEMNLP (3D) (2)	ATTFP	GINX	PAINNKERAS	CHEMPROPKERAS	ATTFPKERAS	SMILESX (3D)
SMILES	0.58 0.042 (0.311)	0.55 0.029 (0.29)	+	0.55 0.44 (0.495)	0.58 0.44 (0.51)	0.51 0.028 (0.269)	0.5 0.43 (0.465)	0.48 0.025 (0.253)	0.45 0.023 (0.237)	0.52 0.027 (0.274)	running
SMILES 10/10	+	+	0.57 0.03 (0.3)	+	+	+	+	+	+	+	+

Pytorch

Keras

Data source → Model benchmark → Model selection → Publish

Model training tasks | All tasks statuses | logHenry+deltaHenry_A [Refresh] [Delete all matching tasks] See tasks of other users Refresh every minute

1 - 1 of 1

User	Task type / Time started	Model / Task name	Property / Set	Method	Status	Priority	Details	
tgg!	Model training 2021-11-18 19:05:26	logHenry+deltaHenry_ATTFPKERAS - 1000007038	logHenry studentmodelcosmo_csv	ATTFPKERAS	ready	normal	-	recalculate

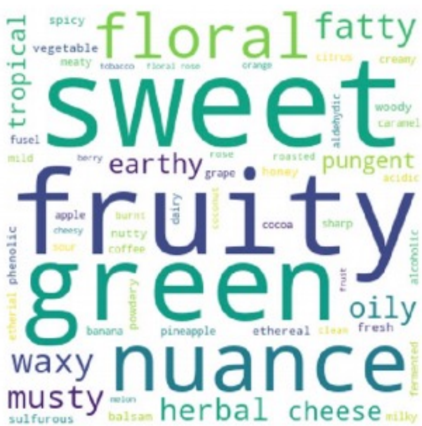
*https://ochem.eu

01 MOLECULE OLFACTION

d-lab - Let start by molecule olfaction model project

Generally, olfactive a molecule has multiple descriptors:

Sweet molecules are also



Musky molecules are also

Task is a “multi label classification” you want to predict several weighted “descriptor”:

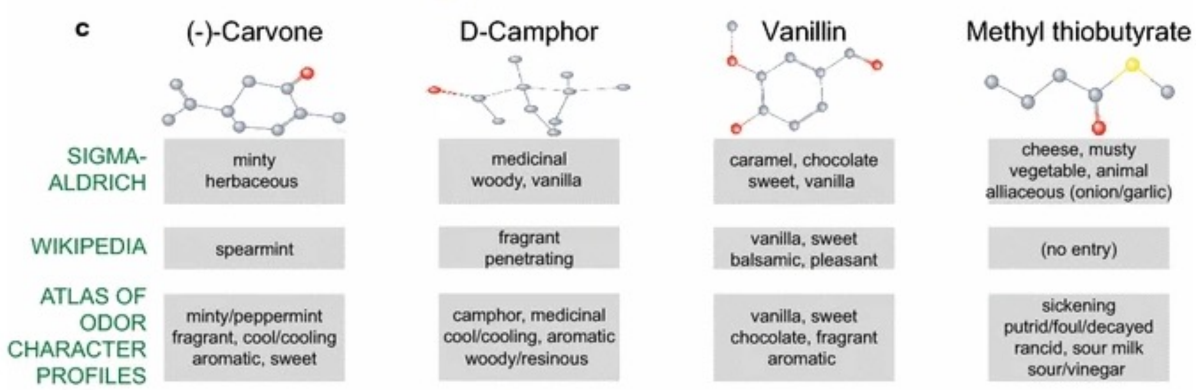
- **Labels are linked** (some are **orthogonal**, but **it’s rarely the case**)
- Labels are **not always reproductible** (Human’s bias)

Firmenich

*https://www.nature.com/articles/s41598-020-73978-1.pdf

Data unstability / interpretability

Human Panelist "deviation" ⟷ **External databases description "deviation"**



Not identical + mixing olfactive descriptors & oral sensation & liking:
- Cooling, Pleasant are not olfactive terms

How to qualify models coming from 2016 "Dream Challenge"* ?

- Small dataset: 480 molecules described/ not very structurally diverse
  - Applicability domain is very limited or sparse (weak)
  - Data quality & high variability observed

Google* GCN approach analysis (2019)

What change in between ?

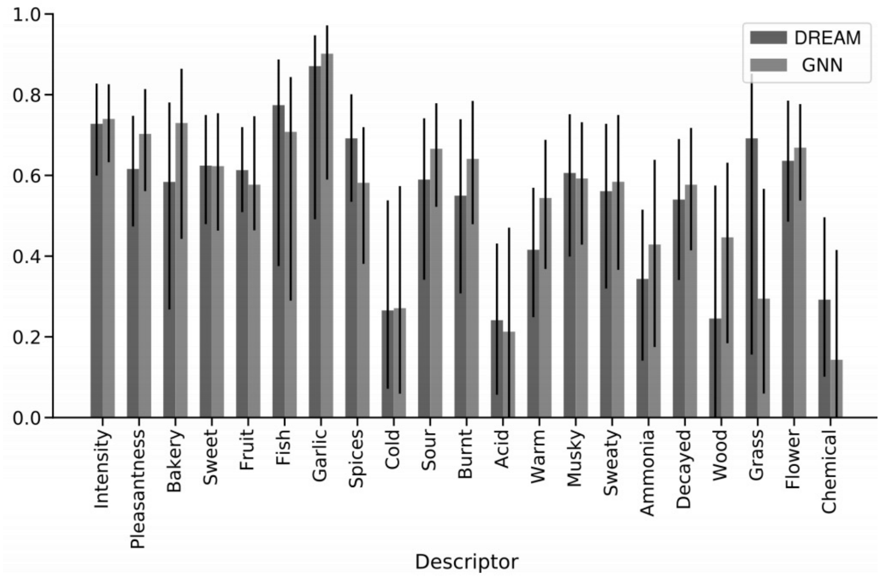
- Dataset increase:
    - 480 to 5000 molecules
- RF to Graph Neural Network

What did not change in between ?

- Data quality
  SD highly instable from splits

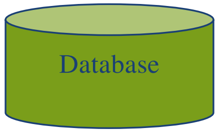
Conclusion:

- Very little improvement with 10x more data points



*https://ai.googleblog.com/2019/10/learning-to-smell-using-deep-learning.html

Our simple KNN Firmenich model (2018)



To build it:
- Select a fingerprint aka “FP”
- Determine Olfaction terms based on closest neighbours using “FP” projection space

Known Issues of such method:
- Applicability domain is strictly Firmenich centered and maybe not compatible with external descriptions
- KNN is not adapted for olfaction of isomers & mixtures*

*https://arxiv.org/pdf/2010.01027.pdf : not enough chirality in dataset (racemic mixture, unknow deviation), by removing chirality of our input molecule, we do not see prediction variation.

Ask help for external open source community

Status after round 2 on external dataset

Δ	#	Participants	top5_tss(adjusted)	top2_tss(adjusted)	
•	01	Sleeper	0.503	0.389	Japan
•	02	rejulien	0.492	0.382	France
•	03	alarih	0.416	0.305	
•	04	zuklei	0.404	0.270	

Learning to Smell*

- **Dataset**: (5000 training + 1000 testing) of external molecules + internal curated process
- **Metric**: Top5 best score other the first 3 words describing a molecule

*https://www.aicrowd.com/challenges/learning-to-smell

What happends if we change the population that describe ?

Final Best solution proposed => Random Forest

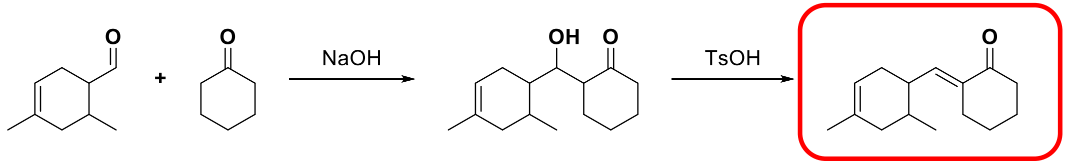
Δ	#	Participants	top5_tss(adjusted)	top2_tss(adjusted)
▲	01	rejulien	0.416	0.290
▼	02	Sleeper	0.400	0.277
●	03	alarih	0.317	0.214
▲	04	hjuinj	0.305	0.219

Score decreasing by 20% between round 2 to 3*

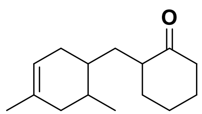
Probable explainations:
- Overfitting
- AD sharing
- SD split variability
- data expertise

Final challenge
- **Tweak** dataset by adding **anonymous** internal molecules in Test set only:
- +10% in training set
- + 200 data point in testing set

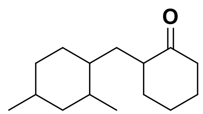
Validation of the olfactive models: synthesis is not fast



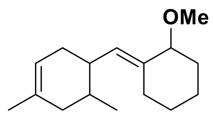
Prediction: woody (31%)
Evaluation: green, compost



Prediction: woody (33%)
Evaluation: dusty, metallic, fatty



Prediction: woody (100%)
Evaluation: fatty, metallic



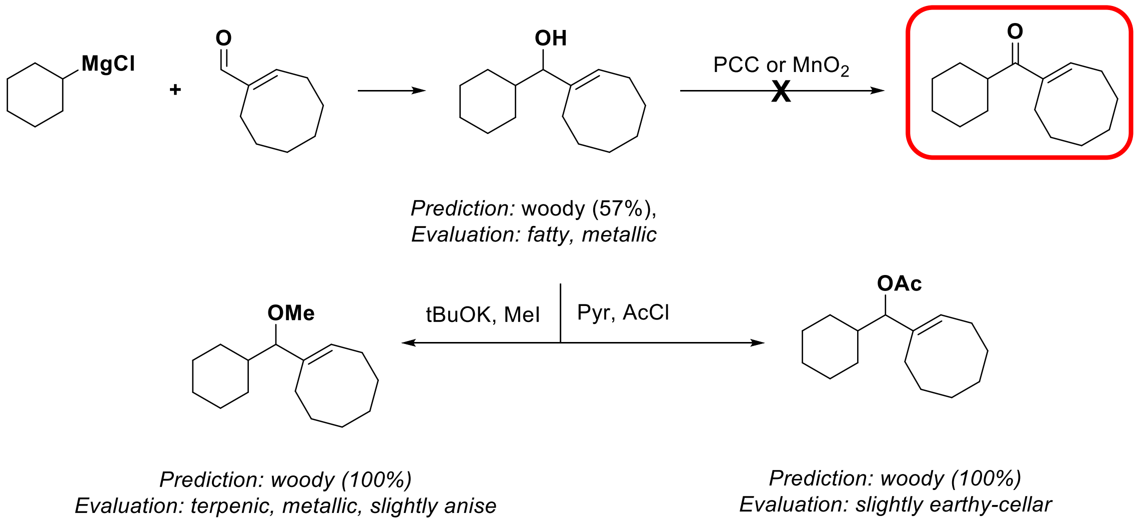
Prediction: woody (100%)
Evaluation: compost, earthy, green

Main molecule & Derivatives

Expected to be woody but none of them are!

Firmenich

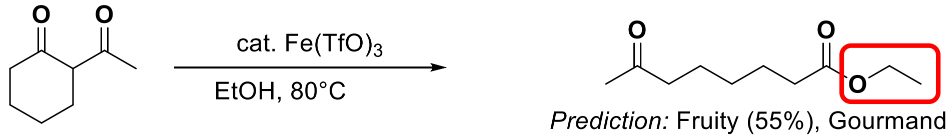
Validation of the olfactive models: synthesis is may failed



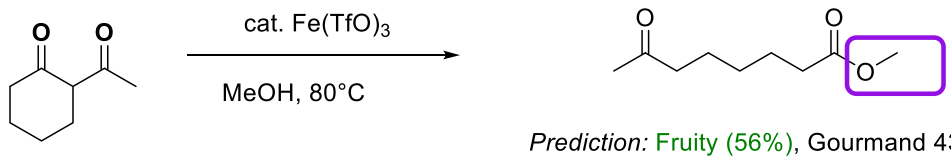
Expected to be Woody but got Terpenic, or Earthy-Moussy instead

Firmenich

Validation of the olfactive models: esters are easy to do !



Prediction: Fruity (55%), Gourmand (44%)
Evaluation: fatty, oily, green,vaguely fruity

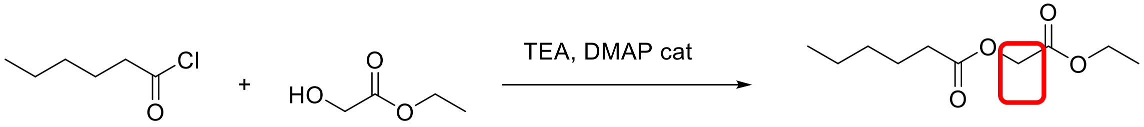


Prediction: Fruity (56%), Gourmand 43%)
Evaluation: fruity, dusty, butyric,

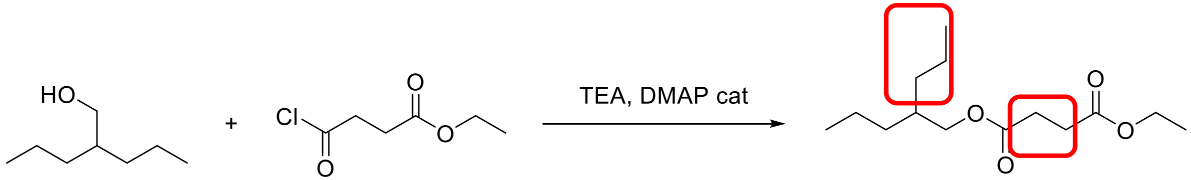
Fruity is simpler to predict (link to esters functions) than Woody

- Small modification have huge impact
  - **ethyl** 7-oxooctanoate
  - **methyl** 7-oxooctanoate

Validation of the olfactive models: esters are easy to do ?



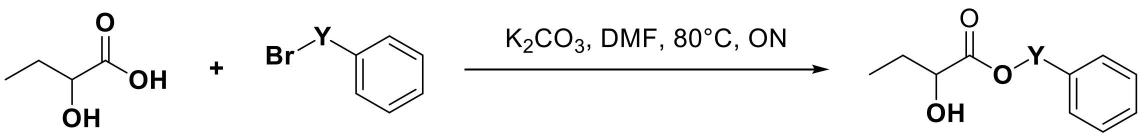
Prediction: Fruity (38%), Animalic (31%), gourmand (29%)
Evaluation: fruity, tutti frutt

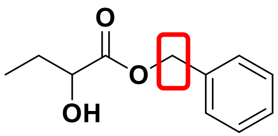


Prediction: Fruity (55%), Gourmand (44%)
Evaluation: odorless

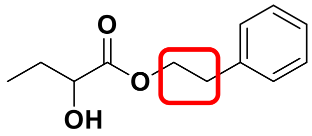
Model cannot deal with odour cut-off of high MW (C10→C14):
"Applicability Domain limit reached"
"Fruity to Odorless"

Our solution for robust AI / ML experience

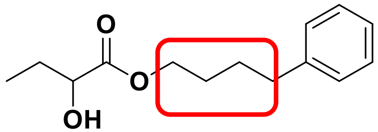




Prediction: Floral (28%), Gourmand (26%), Leathery (24%), Animalic (20%)
Evaluation: animal-fecal, metallic, dusty, almond-benzoic



Prediction: Gourmand (45%), Floral (54%)
Evaluation: rose, phenylacetic, honey



Prediction: Floral (57%), Gourmand (42%)
Evaluation: almost odorless, slightly metallic

Model cannot deal with odour cut-off of high MW (C10→C14):
"Applicability Domain limit reached"
Evolution of the response with MW is very high even if the "fragments are similar"

Firmenich

*https://ochem.eu/home/show.do

Final result using real validation

Expert vs Model Matching	No	One	Several	Perfect
Evaluated ratio	48%	37%	13%	1%

internal validation results was carried done by experts (Perfumery division)

Alignment with :
- Google*: F1 score = 36%
- AICrowd**: 40% Top5 best score

*https://arxiv.org/pdf/1910.10685.pdf
** https://www.aicrowd.com/challenges/learning-to-smell

02 FORMULA GENERATION

AI in art and sensory

MACKMYR
A WHISKY

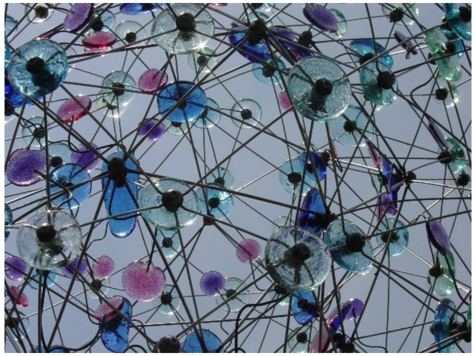
AI-DA THE
ROBOT

SWISS
BEER

THE

UNFINIESHED

What was the goals ?



- Assisting Perfumer and Flavorist

- Increasing Firmenich's signature

- Enhance creativity and authenticity

d-lab – Building the AI Advantage

Our vision:

“To augment human creativity using leading edge technologies, to accelerate innovation and create winning solutions for our customers, ethically.”

3 characteristics

- Safe and agile
- Ethical
- High tech

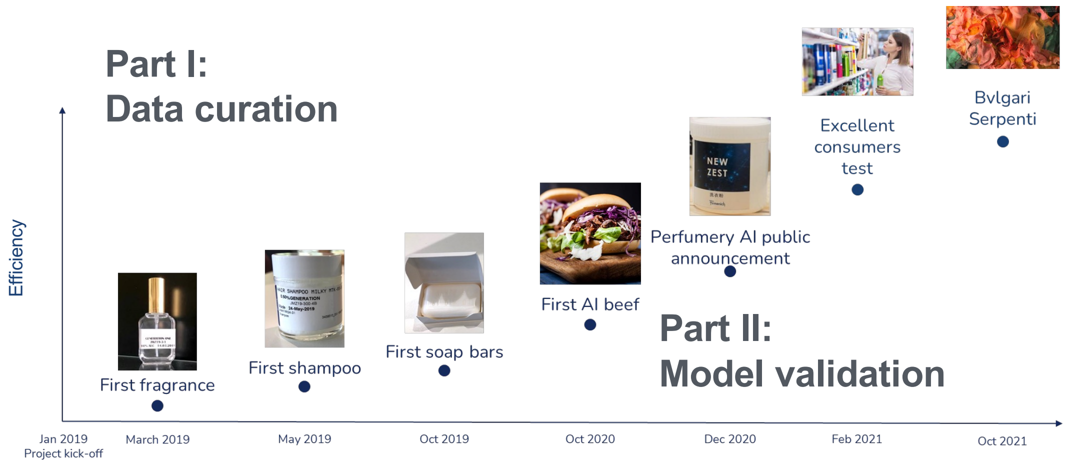
DIGITAL ECONOMY AWARD

Winner of the Digital Innovation of the year 2021

Final result using real validation

AI formula generator

- A machine learning model able to generate formula integrating multiple constraints



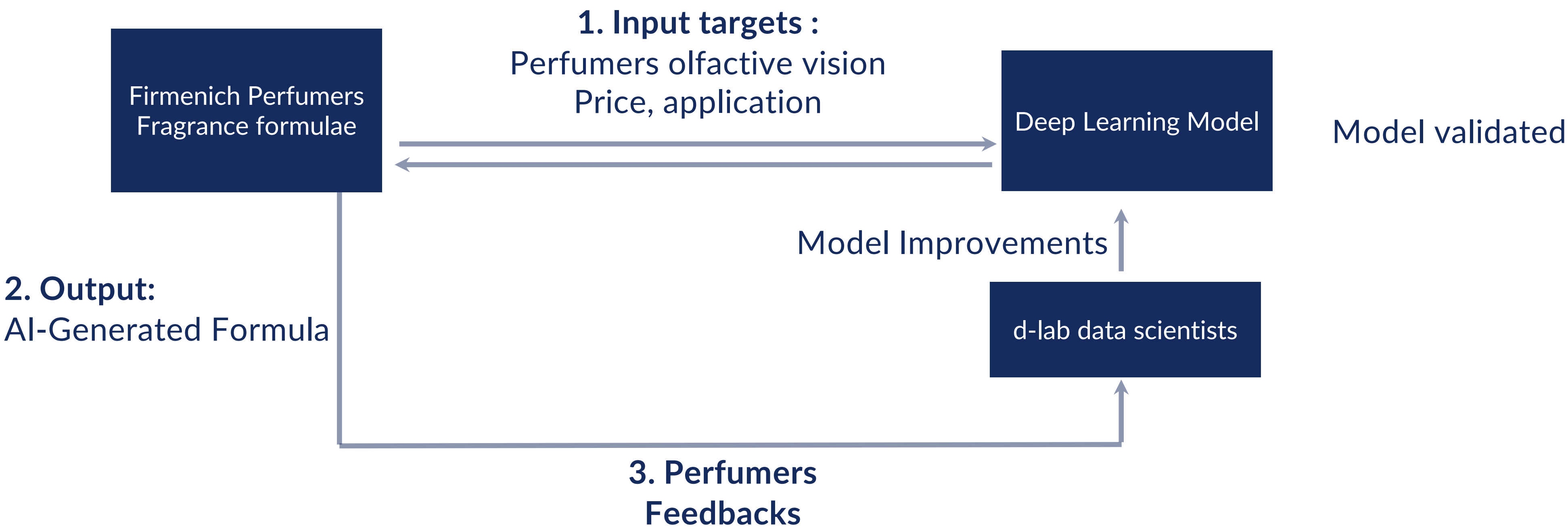
Database continuous integration

Generator for Flavors

Today up to 19 Tonalitites in Production

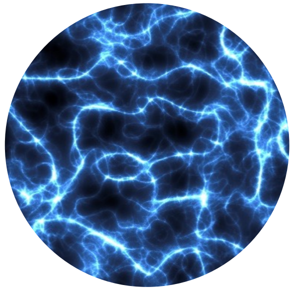
AI BEEF: lightly grilled

How are we working ?



How is it working ?

Leveraging data from Firmenich’s databases and use advanced techniques of statistics, mathematical modelling, data mining and machine learning (AI)



Formula similarity

Ingredient Knownledge

Optimization

Evaluation and Continuous improvement

Firmenich

Generator Interaction by Keywords

- Application: Shampoo, Soaps, Fine Fragrance, etc…
- Olfactive Descriptors: «Floral: 3», «Fruity: 4», «Vanilla: 2», etc
    - The weights in front of the descriptors correspond to the strength of a particular olfactive description.
- Target Price (CC)
- Sustainability: biodeg, carbon renewable
- …

03 FINDINGS

Pillars of Success

People & culture
Open-mindedness, OK to fail, agile, owning

Interaction with business
Building cross-disciplinary agile teams, involving business from day one

Skills
Understanding fundamentals – it's not just about doing a train/test split

Tech
This is at the core – but not an end in itself

Final thoughts

You need both low-hanging fruit and moonshots

Firmenich
for good, naturally

Innovative Craftsmanship in Fragrances, Taste and Beyond

Family Owned, Founded in Geneva, 1895