Practical machine learning based on cloud computing resources

Kyriakos N. Agavanakis[1,a)], George. E. Karpetas[2,b)], Christos M. Michail[3,c)] ,Michael Taylor[4,d)], Lamprini Kontopoulou[7,h)], Varvara Trachana[5,e)], Evangelia Pappa [6,f)], John Filos[6,g)]

1.*Atrinno*, Attica Research and Innovation PC

2.Department of Medical Physics, Faculty of Medicine, *University of Thessaly*

3.*University of West Attica*, Department of Biomedical Engineering Radiation Physics, Materials Technology and Biomedical Imaging Laboratory

4. Department of Meteorology, *University of Reading*, Reading RG6 6BB, UK

5. Laboratory of Biology, Faculty of Medicine, *University of Thessaly*

6.Department of Public Administration School of Economy and Public Administration, *Panteion University of Social and Political Sciences*

7. General Department, *University of Thessaly*, 41110, Larissa, Greece

Purpose

Investigate the *practical* applications of machine learning (ML) algorithms in several scientific areas,

and

Utilize cloud resources to provide *usable* services not only within the scientific community, but to everybody!

Case studies

- quality evaluation metrics for the tomographic image reconstruction of positron emission tomography (PET) images
- health implications of the vitamin D absorption function. Results showed that commercially available cloud resources are over sufficient to consolidate results from a variety of teams and applications and contribute to the built up of a valuable shared knowledge repository
- the investigations of the demographic determinants influencing the perception of corruption incidents within different industry sectors

Achievements

Using the suggested approach in the context of a widely available cloud service for feeding the training algorithms, will contribute to more accurate automation and successful operation of related activities in the application domains, breaking thus the knowledge silos and contributing to a more sustainable environment.

CASE STUDY

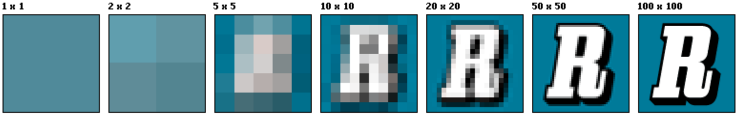
Modulation Transfer Function calculation using cloud-based Machine Learning Services

Evangelia Pappa
John Filos

Definitions

Spatial resolution – the amount of geometric detail

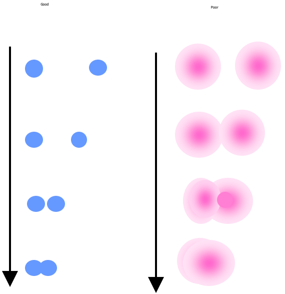
- How close can two points be before you can't distinguish them



Imaging

As spatial separation decreases, the "good" system maintains clear separation of point source images, while the "poor" system eventually can no longer distinguish them.

MTF quantifies this phenomenon in terms of contrast between the center peak intensities versus intensity at their midpoint across a scale of separation distances.



At large separations, even a poor system can completely resolve the two images. As separation decreases, only the good systems can still recognize separate sources.

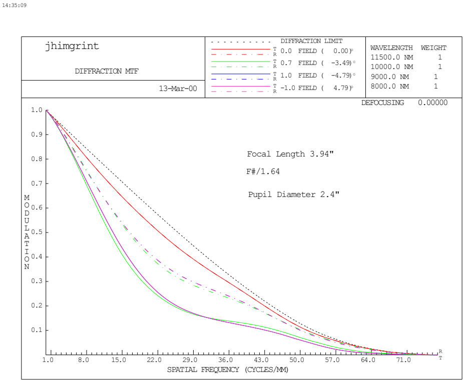
Image Quality in Nuclear Imaging

The response of the system to the incident signal amplitudes can be described by the :

Modulation **T**ransfer **F**unction (**MTF**),

which expresses the system's response in the spatial frequency domain by taking the Fourier transform of the corresponding PSF from a reconstructed cross-sectional image.

- MTF is a measure of intensity contrast transfer per unit resolution of an image or signal.
- It is used in optics, electronics, and related signal processing applications.



MTF curves obtained from iterative STIR reconstructed LSF images (the number of subsets was kept fixed and the number of iterations was increased with a step of 2)

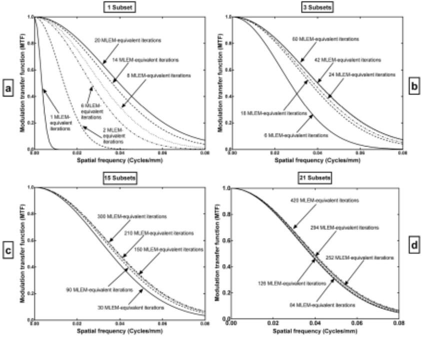
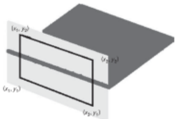


Fig. 2. a-d) MTFs obtained with the LSF method, from the plane source reconstructed image with various MLEM-equivalent iterations.

Simulation of the plane source for the MTF measurement



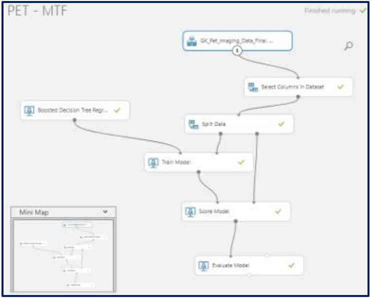
Schematic representation of the line profile selection

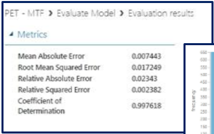
Data format…

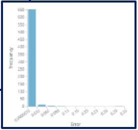
Spatial Frequency											
	Subsets		1	1	1	1	1	1		3	3
	Iterations		1	2	6	8	14	20		2	6
0,000000	X		1,00000 0	1,00000 0	1,00000 0	0,99533 4	1,00000 0	1,0000 0		1,00000 0	1,0000 0
0,001151888			0,92441 6	0,99517 2	0,9987	1	0,99930 5	0,99941 7		0,99875 1	0,99931 5
0,002303776			0,73024 6	0,98082 7	0,99480 9	0,9981					
0,003455664			0,49295 4	0,95737 7	0,98835 9	0,9870					
0,004607552			0,28436 6	0,92548 5	0,97939 8	0,9848					
0,00575944			0,14017 9	0,88603 8	0,96799 6	0,9793					
0,006911328			0,05905 1	0,8401	0,95424 1	0,9627					

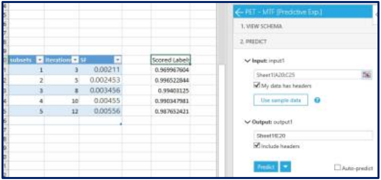
Data conversion

	Subsets	Iterations	Spatial Frequency	MTF
1	1	1	0,000000	1,000000
2	1	2	0,000000	1,000000
3	1	6	0,000000	1,000000
. . . .		8	0,000000	0,995334
3240	21	20	0,000000	1,000000


PET - MTF
Finished running
Select Columns in Dataset
Boosted Decision Tree Regr...
Split Data
Train Model
Score Model
Evaluate Model
Mini Map


PET - MTF › Evaluate Model › Evaluation results
Metrics
Mean Absolute Error 0.007443
Root Mean Squared Error 0.017249
Relative Absolute Error 0.02343
Relative Squared Error 0.002382
Coefficient of Determination 0.997618




subsets
iterations
SF
Scored Labels
1 3 0.00211 0.969967604
2 5 0.002453 0.996522844
3 8 0.003456 0.99403125
4 10 0.00455 0.990347981
5 12 0.00556 0.987652421
PET - MTF [Predictive Exp.]
1. VIEW SCHEMA
2. PREDICT
Input: input1
My data has headers
Use sample data
Output: output1
Include headers
Predict
Auto-predict

➢ CASE STUDY

Bio-uv products

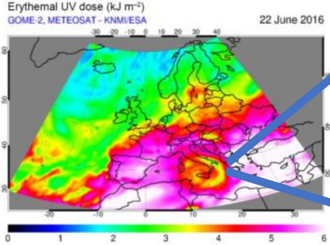
Neural network calculation of Vitamin-D and DNA-damage doses from spectral UV irradiance using cloud-based services

Michael Taylor,
Lamprini Kontopoulou,
Varvara Trachana[5,]

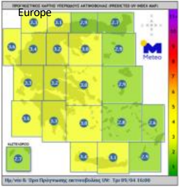
Surftemp Satellite Remote Sensing Group

SATELLITE UV DOSE

- Satellites like SCIAMACHY and GOME-2 have operational processing algorithms that retrieve erythemal UV dose (kJ m^{-2}) from space:



Used to calculate the UV index in Greece and across Europe



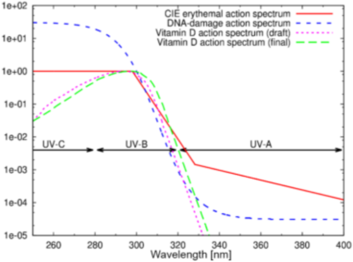
http://meteo.gr/uv.cfm

Van Geffen, J., Van Weele, M., Allaart, M. and Van der A, R.: 2017, TEMIS UV index and UV dose operational data products: http://www.temis.nl/uvradiation/UVarchive.html

BIOLOGICAL UV PRODUCTS

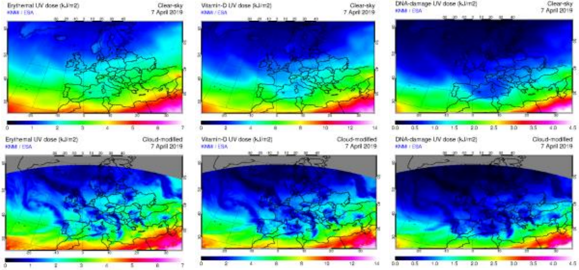
- Interestingly, you can use the satellite UV together with window functions ("action spectra") to calculate important biological UV products across the Earth's surface:

1) Vitami n-D dose
2) DNA-damage dose



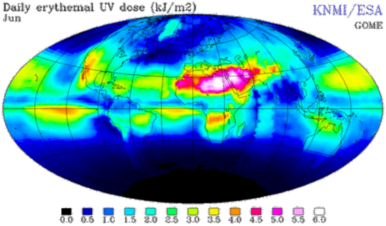
Zempila, M. M., van Geffen, J. H., Taylor, M., Fountoulakis, I., Koukouli, M. E., van Weele, M., Bais, A., Meleti, C., Balis, D. (2017). TEMIS UV product validation using NILU-UV ground-based measurements in Thessaloniki, Greece. *Atmospheric Chemistry and Physics*, *17*(11), 7157-7174.

- Using the viewing potential of satellites, this means we can generate maps of these UV products for most of the Earth surface - **but only once a day**:



Van Geffen, J., Van Weele, M., Allaart, M. and Van der A, R.: 2017, TEMIS UV index and UV dose operational data products: http://www.temis.nl/uvradiation/UVarchive.html

- As well as being sensitive to cloud, the UV reaching ground is also sensitive to absorbing aerosol (e.g. desert dust) – the combination of these 2 factors is a challenge for neural network models:

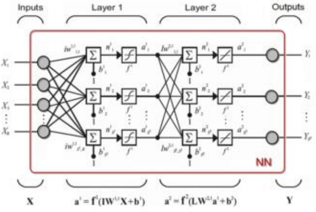


Van Geffen, J., Van Weele, M., Allaart, M. and Van der A, R.: 2017, TEMIS UV index and UV dose operational data products: http://www.temis.nl/uvradiation/UVarchive.html

BPNN MODEL

- A high frequency (1 minute interval) back-propagation neural network (BPNN) model has recently been developed to calculate these biological products from UV irradiances at 5 wavelengths plus the solar zenith angle (SZA) as inputs:

Ir (305 nm)
Ir (312 nm)
Ir (320 nm)
Ir (340 nm)
Ir (380 nm)
SZA



Erythemal UV dose
Vitamin-D dose
DNA-damage dose

Zempila, M. M., van Geffen, J. H., Taylor, M., Fountoulakis, I., Koukouli, M. E., van Weele, M., Bais, A., Meleti, C., Balis, D. (2017). TEMIS UV product validation using NILU-UV ground-based measurements in Thessaloniki, Greece. *Atmospheric Chemistry and Physics*, *17*(11), 7157-7174.

CNN MODEL

- Initial simulations using convolutional neural network (CNN) model trained on the same data are demonstrating similar levels of precision:

Ir (305 nm)

Ir (312 nm)

Ir (320 nm)

Ir (340 nm)

Ir (380 nm)

SZA

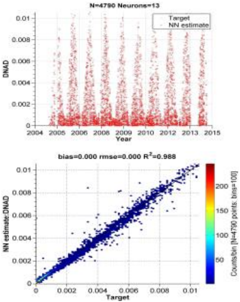
Erythemal UV dose

Vitamin-D dose

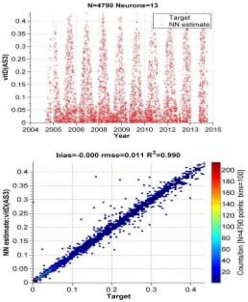
DNA-damage dose

Zempila, M. M., van Geffen, J. H., Taylor, M., Fountoulakis, I., Koukouli, M. E., van Weele, M., Bais, A., Meleti, C., Balis, D. (2017). TEMIS UV product validation using NILU-UV ground-based measurements in Thessaloniki, Greece. *Atmospheric Chemistry and Physics*, *17*(11), 7157-7174.

BPNN: DNA damage dose



BPNN Vitamin-D dose



Zempila, M. M., van Geffen, J. H., Taylor, M., Fountoulakis, I., Koukouli, M. E., van Weele, M., Bais, A., Meleti, C., Balis, D. (2017). TEMIS UV product validation using NILU-UV ground-based measurements in Thessaloniki, Greece. *Atmospheric Chemistry and Physics*, *17*(11), 7157-7174.

➢ CASE STUDY

Health and Care

Working environments and Business ethics

Kyriakos N. Agavanakis
George. E. Karpetas
Evangelia Pappa
John Filos

investigate the demographic determinants influencing the perception of corruption incidents within different industry sectors.

The major research instrument is a self-administered questionnaire which distributed to a random sample of individuals working in Greece.

Machine Learning as a Detecting Tool

Fraud incidents	Detect fraudulent scandals Identify integrity violations
Corruption cases	• Detect corruption • Detect low incidence events • Identify red flags

Our dataset

Persons	1000 in different industry sectors

Inputs	Nepotism
	Using expense claims unethically
	Long private telephone calls
	Surfing the internet for private purposes during working hours
	Taking company resources home from private use
	Arriving late at work
	Insufficient effort from staff members.
	Taking the credit of other people's work. → . *estimated*

Confusion Matrix

Predicted Class

Actual Class	1	2	3	4	5
1	93.5%	6.5%			
2	46.7%	46.7%	3.3%	3.3%	
3	20.0%	52.0%	20.0%	8.0%	
4		14.3%		85.7%	
5					

Overall accuracy	**0.77**
Average accuracy	**0.908**

”It is all interconnected

Platforms, Big Data, analytics, algorithms, machine learning, and artificial intelligence”

AI is the area of engineering intelligent machines

capable of perceiving the environment through activities such as

perception, learning & reasoning,

and take actions that maximize their chance of success at some goal

Machine learning – evolved from data analytics and pattern recognition – infers models from data streams, by combining their historical relations (often including hidden patterns) and their current trends.

An important role to this evolution has been played by the maturity of the associated enabling technological fields such as

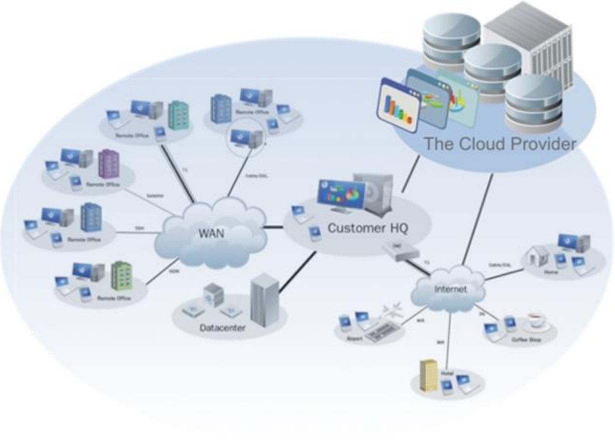
- Cloud computing
- Big Data
- Accessibility/reachability
- Telecommunications, smart devices

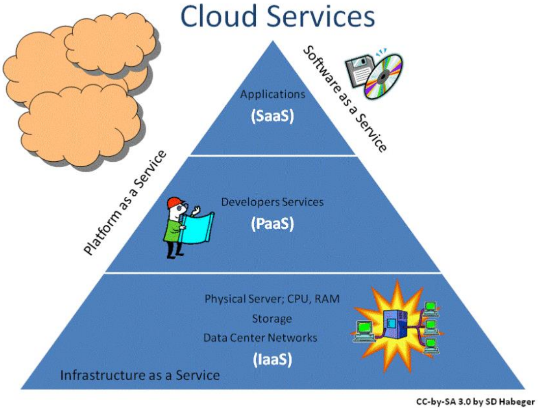
Deep learning is the application of artificial neural networks (neural networks for short) to learning tasks using networks of multiple layers. Essentially a statistical technique for classifying patterns, based on sample data, using neural networks with multiple layersDeep learning is the application of artificial neural networks (neural networks for short) to learning tasks using networks of multiple layers. Essentially a statistical technique for classifying patterns, based on sample data, using neural networks with multiple layers has shown numerous impressive results and became one of the most efficient areas of AI, with results such as

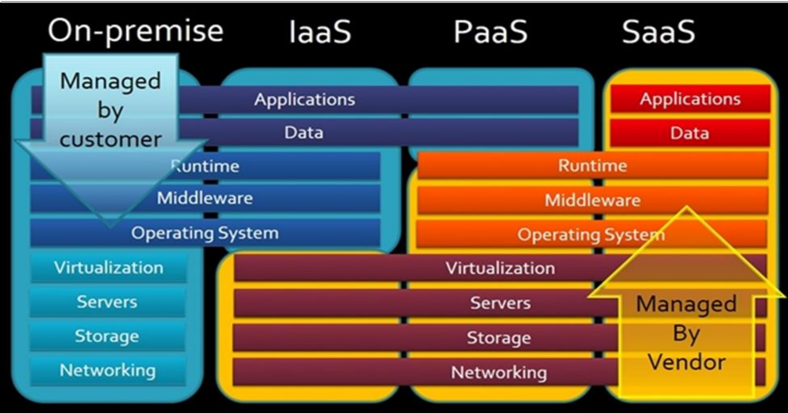
- speech recognition,
- image recognition,
- image deconvolution,
- language translation,
- game playing,
- bioinformatics,
- information retrieval,
- content recognition,
- security (e.g. intrusion detection, malware detection)

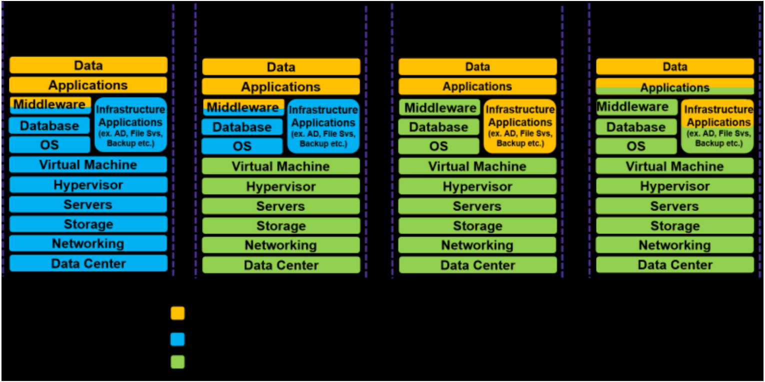
Several industry domains are already making use of the positive results of the ML application in their area, such as

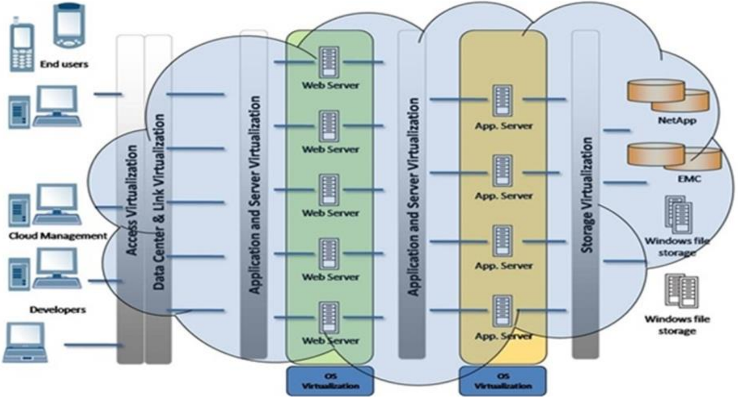
- retail shopping (personalized advertising, suggestions, campaigns),
- b2b (supply planning and customer insights),
- financial services (identification of important data insights, fraud detection),
- government (utilities),
- health care (wearable sensors, medical exams)

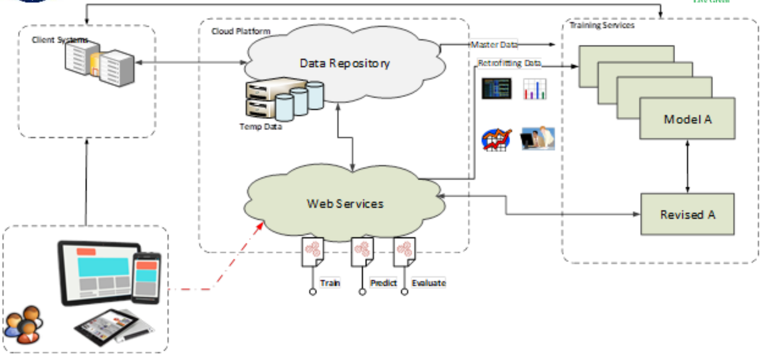

The Cloud Provider
WAN
Customer HQ
Datacenter
Internet
Remote Office

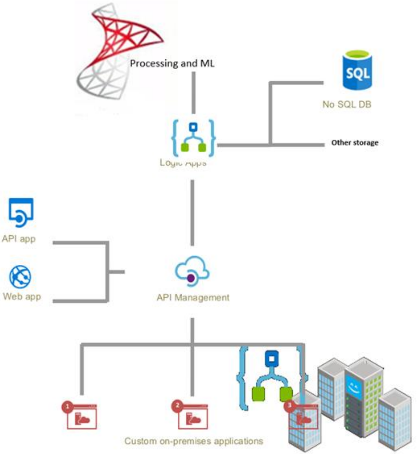

Cloud Services
Software as a Service
Applications
(SaaS)
Platform as a Service
Developers Services
(PaaS)
Physical Server; CPU, RAM
Storage
Data Center Networks
(IaaS)
Infrastructure as a Service
CC-by-SA 3.0 by SD Habeger


On-premise
IaaS
PaaS
SaaS
Managed by customer
Applications
Data
Runtime
Middleware
Operating System
Virtualization
Servers
Storage
Networking
Managed By Vendor

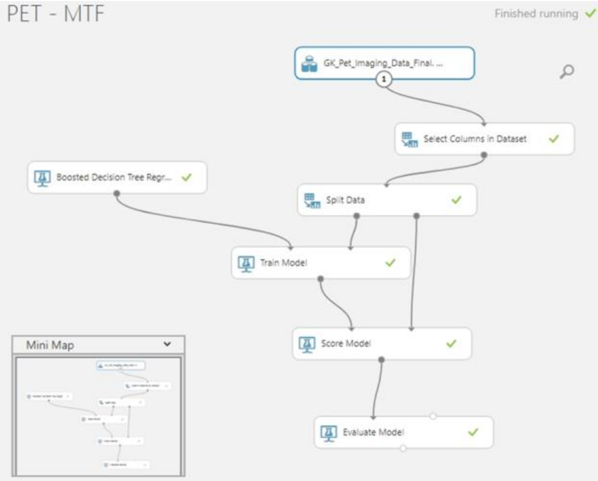

Data
Applications
Middleware
Database
OS
Infrastructure Applications (ex. AD, File Svs, Backup etc.)
Virtual Machine
Hypervisor
Servers
Storage
Networking
Data Center
Data
Applications
Middleware
Database
OS
Infrastructure Applications (ex. AD, File Svs, Backup etc.)
Virtual Machine
Hypervisor
Servers
Storage
Networking
Data Center
Data
Applications
Middleware
Database
OS
Infrastructure Applications (ex. AD, File Svs, Backup etc.)
Virtual Machine
Hypervisor
Servers
Storage
Networking
Data Center
Data
Applications
Middleware
Database
OS
Infrastructure Applications (ex. AD, File Svs, Backup etc.)
Virtual Machine
Hypervisor
Servers
Storage
Networking
Data Center


End users
Cloud Management
Developers
Access Virtualization
Data Center & Link Virtualization
Application and Server Virtualization
Web Server
Web Server
Web Server
Web Server
Web Server
OS Virtualization
Application and Server Virtualization
App. Server
App. Server
App. Server
App. Server
OS Virtualization
Storage Virtualization
NetApp
EMC
Windows file storage
Windows file storage


Client Systems
Cloud Platform
Data Repository
Temp Data
Master Data
Retrofitting Data
Training Services
Model A
Revised A
Web Services
Train
Predict
Evaluate


Processing and ML
SQL
No SQL DB
Other storage
Logic Apps
API app
Web app
API Management
Custom on-premises applications

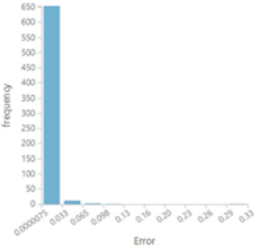
subsets	iterations	SF	MTF
1	8	0.001151888	1
1	14	0.001151888	0.9993054
1	20	0.001151888	0.9994166
3	2	0.001151888	0.9987505
3	6	0.001151888	0.9993148
3	8	0.001151888	0.9993687
3	14	0.001151888	0.9994235
3	20	0.001151888	0.9994295
15	2	0.001151888	0.9992808
15	6	0.001151888	0.9993671
15	10	0.001151888	0.9993962
15	14	0.001151888	0.9994069

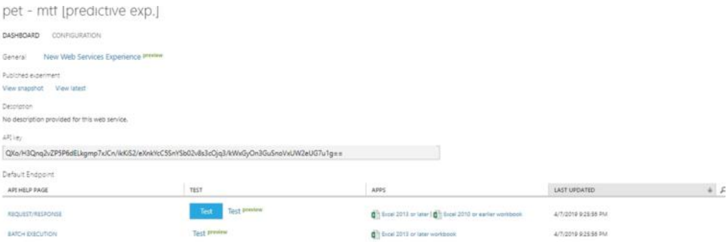

PET - MTF
Finished running
GK_Pet_Imaging_Data_Final. ...
Select Columns in Dataset
Boosted Decision Tree Regr...
Split Data
Train Model
Score Model
Evaluate Model
Mini Map

PET - MTF › Evaluate Model › Evaluation results

Metrics

Mean Absolute Error	0.007443
Root Mean Squared Error	0.017249
Relative Absolute Error	0.02343
Relative Squared Error	0.002382
Coefficient of Determination	0.997618




pet - mtt [predictive exp.]
DASHBOARD
CONFIGURATION
General
New Web Services Experience preview
Published experiment
View snapshot
View latest
Description
No description provided for this web service.
API key
Default Endpoint
API HELP PAGE
TEST
APPS
LAST UPDATED
REQUEST/RESPONSE
Test
Test preview
Excel 2013 or later
Excel 2010 or earlier workbook
4/7/2019 9:25:55 PM
BATCH EXECUTION
Test preview
Excel 2013 or later workbook
4/7/2019 9:25:56 PM

Request
POST

Request Headers

Request Header	Description
Authorization:Bearer abc123	Required. Pass the API Key here. Obtain this key from the publisher of the API.
Content-Length	Required. The length of the content body.
Content-Type:application/json	Required if the request body is sent in JSON format.
Accept: application/json	Optional. Use the header to receive the response in JSON format.

```
{
  "Inputs": {
    "input1": {
      "ColumnNames": [
        "subsets",
        "iterations",
        "SF"
      ],
      "Values": [
        [
          "0",
          "0",
          "0"
        ],
        [
          "0",
          "0",
          "0"
        ]
      ]
    }
  },
  "GlobalParameters": {}
}
```

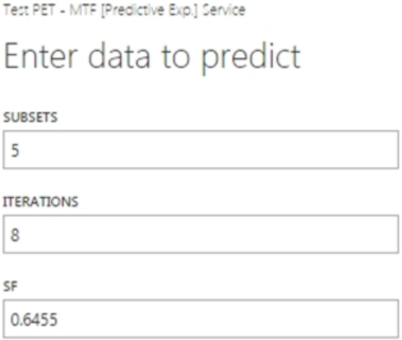
Response

Status Code

A successful operation returns status code 200 (OK)

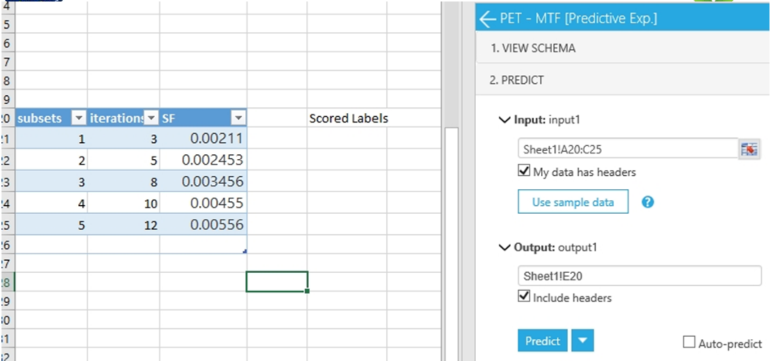
Response Header	Description
Content-Type:application/json	Indicates that the content body is in json format.

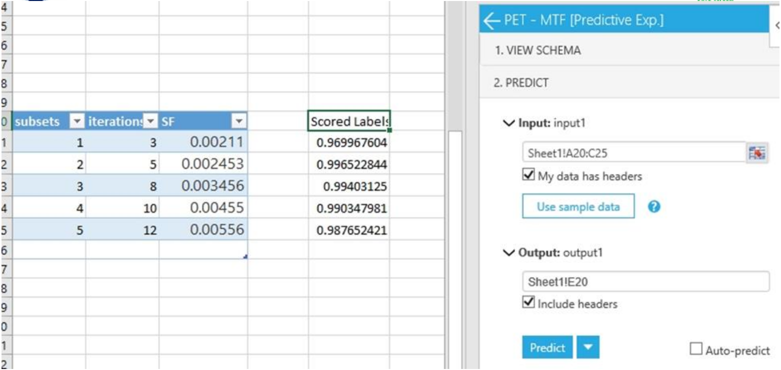

```
{
  "Results": {
    "output1": {
      "type": "DataTable",
      "value": {
        "ColumnNames": [
          "Scored Labels"
        ],
        "ColumnTypes": [
          "Numeric"
        ],
        "Values": [
          [
            "0"
```


Test PET - MTF [Predictive Exp.] Service
Enter data to predict
SUBSETS
5
ITERATIONS
8
SF
0.6455

```
PET - MTF [Predictive Exp.]' test returned ["5","3","0.2245","1.04300057888031"]...

Result: {
    "Results":{
        "output1":{
            "type":"table",
            "value":{
                "ColumnNames":["subsets","iterations","SF","Scored Labels"],
                "ColumnTypes":["Nullable`1","Nullable`1","Nullable`1","Double"],
                "Values":[["5","3","0.2245","1.04300057888031"]]
                }
            }
        }
    }
```


subsets
iterations
SF
Scored Labels
1 3 0.00211
2 5 0.002453
3 8 0.003456
4 10 0.00455
5 12 0.00556
PET - MTF [Predictive Exp.]
1. VIEW SCHEMA
2. PREDICT
Input: input1
Sheet1!A20:C25
My data has headers
Use sample data
Output: output1
Sheet1!E20
Include headers
Predict
Auto-predict


subsets
iteration
SF
Scored Labels
1 3 0.00211 0.969967604
2 5 0.002453 0.996522844
3 8 0.003456 0.99403125
4 10 0.00455 0.990347981
5 12 0.00556 0.987652421
PET - MTF [Predictive Exp.]
1. VIEW SCHEMA
2. PREDICT
Input: input1
Sheet1!A20:C25
My data has headers
Use sample data
Output: output1
Sheet1!E20
Include headers
Predict
Auto-predict

In the first CASE study, the usability of results may be dependent on other factors as well such as

Quantitative factors

- NNPS, Normalized Noise Power Spectrum
- DQE, Detective Quantum Efficiency
- SNR, Signal to Noise Ratio
- CNR, Contrast to Noise Ratio
- IC, Information Content

..as well as **Qualitative** ones

- Patient's movement (typical exam: 30')
- Body type and fat (thin, normal, obese)
- PET scanner operation mode (2D, 3D)
- PET machine structure, type and operation configuration

The same is true for different experiment parameters for the other case studies as well

..therefore, in order for the results to be broadly useful

→ Similar experiments have to be repeated for several influencing combinations

Whether that is related to the measurements conditions, geo-location or whatever else is applicable.

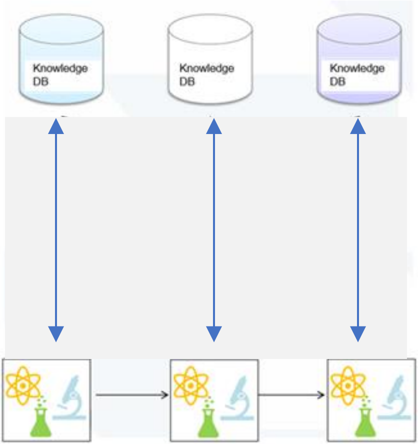
- teams working in the similar or even the same problem cannot easily combine their research results

- Even when they are not reluctant to share their results, and they publish them, the outcome is not always directly (re)usable

- Even when they do provide detailed results, and they can be used, there is a huge delay incorporated in order to be included in the product life cycle of some product and be practically useful to other scientists or end users
  e.g. in the PET case study, we need similar datasets for a wide variety of the influencing factors (PET configurations, energy, model, type) in order to have universally useful data set to be incorporated by industry manufacturers in a product and server the needs of real end users, all over the world

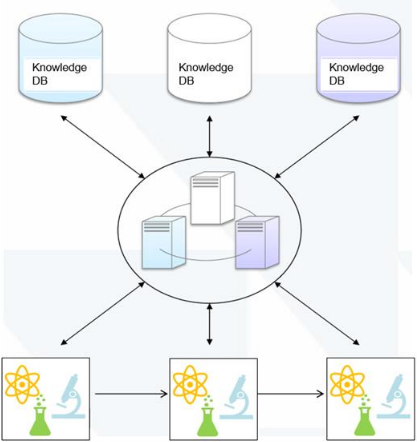
So instead of scientific teams to work in Knowledge silos,

It is better to form dynamic ecosystems



Vision: Knowledge silos

Products, services or technologies
developed by one,
serve as foundations
upon which others can build
complementary
products, services or technologies



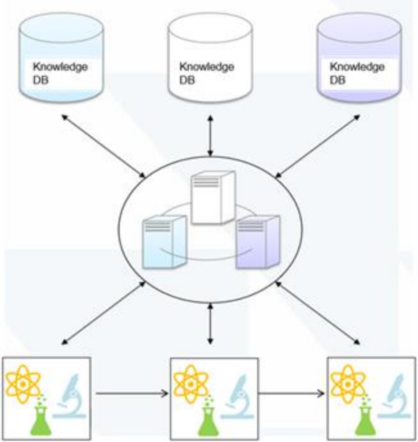
Software, cloud services, IoT, CPS
…dynamic ecosystems,
Where actors interact across boundaries

Highest added value when

platforms are made accessible to complementary

third-party technologies, products and services

that create value for everybody



The proposed infrastructure can

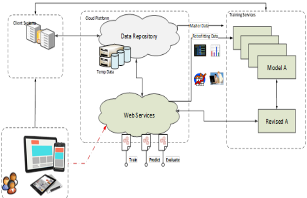
- Guarantee the results ownership
  (data used to train the models can be digitally signed and secured)

- Make them useful to a world wide variety of users, without exposing them
  (the trained models are needed for the applications, not the input data themselves)

- Services can be easily integrated to end users applications and be useful through
  - web sites
  - mobile applications
  - desktop applications
  - social apps
  - other 3rd party applications

- Besides of providing useful predictions to end users,
  end user's data may be further used to retrofit the models

  and contribute to their continue improvement



Let me find, understand and use my data

M. Taylor, "Should research data be publicly available?" 22-May-2013. https://www.elsevier.com/connect/should-research-data-be-publicly-available.

In an open, peaceful society,
knowledge shared is power multiplied

European Parliamentary Research Service Blog

Thank you!