Demonstration on OpenRiskNet approach on modelling for prediction or read across with the Jaqpot platform
ModelRX case study

Philip Doganis
National Technical University of Athens

OpenRiskNet: Open e-Infrastructure to Support Data Sharing, Knowledge Integration and *in silico* Analysis and Modelling in Risk Assessment
Project Number 731075
# OpenRiskNet webinars series

<table>
<thead>
<tr>
<th>Topic</th>
<th>Date &amp; Time</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Past events</strong></td>
<td></td>
</tr>
<tr>
<td>Introduction sessions to the OpenRiskNet e-infrastructure</td>
<td>Webinar recordings:</td>
</tr>
<tr>
<td></td>
<td>● [Session 1](24 Sep 2018)</td>
</tr>
<tr>
<td></td>
<td>● [Session 2](27 Sep 2018)</td>
</tr>
<tr>
<td></td>
<td>● [Session 3](4 Oct 2018)</td>
</tr>
<tr>
<td></td>
<td>● [Session 4](30 Oct 2018)</td>
</tr>
<tr>
<td>Learn how to deploy the OpenRiskNet virtual research environment</td>
<td>Webinar recordings (25 Feb 2019)</td>
</tr>
<tr>
<td>Demonstration on data curation and creation of pre-reasoned datasets in the OpenRiskNet framework</td>
<td>Webinar recordings (18 Mar 2019)</td>
</tr>
<tr>
<td>Identification and linking of data related to AOPWiki (an OpenRiskNet case study)</td>
<td>Webinar recordings (26 March 2019)</td>
</tr>
<tr>
<td>The Adverse Outcome Pathway Database (AOP-DB)</td>
<td>Webinar recordings (8 April 2019)</td>
</tr>
<tr>
<td>How to describe OpenRiskNet services and their functionality by semantic annotation</td>
<td>Webinar recordings (13 May 2019)</td>
</tr>
<tr>
<td>Use of Nextflow tool for toxicogenomics-based prediction and mechanism identification in OpenRiskNet e-infrastructure</td>
<td>Webinar recordings (27 May 2019)</td>
</tr>
<tr>
<td><strong>Current event</strong></td>
<td>Tuesday, 11 June 2019, 16:00 CEST Registration: <a href="https://openrisknet.org/events/67/">https://openrisknet.org/events/67/</a></td>
</tr>
<tr>
<td>Demonstration on OpenRiskNet approach on modelling for prediction or read across (ModelRX case study)</td>
<td><a href="https://openrisknet.org/events/67/">https://openrisknet.org/events/67/</a></td>
</tr>
<tr>
<td><strong>Future events</strong></td>
<td>Wednesday, 12 June 2019, 20:00 CEST Registration: <a href="https://openrisknet.org/events/73/">https://openrisknet.org/events/73/</a></td>
</tr>
<tr>
<td>Combining neXtProt and WikiPathways strengths using SPARQL federated queries</td>
<td><a href="https://openrisknet.org/events/73/">https://openrisknet.org/events/73/</a></td>
</tr>
<tr>
<td>Deploying Applications to an OpenRiskNet Virtual Environment</td>
<td>Monday, 24 June 2019, 16:00 CEST Registration: <a href="https://openrisknet.org/events/66/">https://openrisknet.org/events/66/</a></td>
</tr>
<tr>
<td>AOPlink workflow</td>
<td>Monday, 15 July 2019, 16:00 CEST Registration: <a href="https://openrisknet.org/events/70/">https://openrisknet.org/events/70/</a></td>
</tr>
</tbody>
</table>
1. Introduction to ModelRX
2. Jaqpot 4
   a. Modelling over UI
   b. QPRF (QSAR prediction reporting format) report.
3. Jaqpot 5
   a. Modelling over UI
   b. QMRF reports example
   c. Modelling over Jupyter Notebook
Question 1

1. Where is most of your work?
   a. Experimental
   b. Computational
   c. Information Technology
Case studies based on risk assessment framework
ModelRX Case Study

Modelling for Prediction or Read Across [ModelRX]

The ModelRX case study contributes to OpenRiskNet by providing:

- **computational methods** to support suitability assessment of existing data and identification of analogues
- **predictive modelling** functionalities, which are essential in the field of final risk assessment.

<table>
<thead>
<tr>
<th>Case Study leader</th>
<th>Team</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harry Sarimveis (NTUA)</td>
<td>NTUA, JGU Mainz, U Uppsala</td>
</tr>
</tbody>
</table>

https://openrisknet.org/e-infrastructure/development/case-studies/case-study-modelrx/
ModelRX Tools

- JaqPot (NTUA)
  - V4
    - GUI: https://ui-jaqpot.prod.openrisknet.org/
    - API: https://api-jaqpot.prod.openrisknet.org/jaqpot/swagger/
  - V5
    - GUI: https://app.jaqpot.org
    - API: http://api.jaqpot.org/jaqpot/swagger/
- Jupyter Notebooks
  - https://jupyterhub-jupyter.prod.openrisknet.org/hub/login
- Squonk Computational Notebook
  - https://squonk-notebook.prod.openrisknet.org/portal
- Lazar (JGU/IST)
  - https://lazar.prod.openrisknet.org/ (API + GUI)
- JGU WEKA Rest service (JGU)
  - https://jguweka.prod.openrisknet.org/ (API)
- CPSign/MetPred/cpLogD (UU)
Question 2

Rate your experience level in modelling of chemicals

a. Experienced
b. Frequent
c. Occasional
d. Beginner
e. No Experience
Jaqpot

- Jaqpot is a web application for modelling
  - data pre-processing
  - algorithms: statistical analysis, data mining and machine learning
  - methods for defining the Domain of Applicability of a predictive (nano-)QSAR model.

- Developed following the extended OpenTox APIs

- Jaqpot 4 offers modelling and *in silico* capabilities

- Jaqpot 5
  - expanded the modelling functionality
  - added QMRF support
  - remaining QSAR-focused features from Jaqpot 4 are integrated gradually
## Jaqpot - Features by version (as of June 2019)

<table>
<thead>
<tr>
<th></th>
<th>Modelling</th>
<th>Validation</th>
<th>QMRF report</th>
<th>QPRF report</th>
<th>Integration with Python</th>
<th>Integration with R jaqpotR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jaqpot 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>over JPDI</td>
<td>in progress</td>
</tr>
<tr>
<td></td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
<tr>
<td>Jaqpot 5</td>
<td>all Scikit-learn algorithms</td>
<td>all Scikit-learn algorithms</td>
<td>all Scikit-learn algorithms</td>
<td>all Scikit-learn algorithms</td>
<td>over jaqpotpy</td>
<td>in progress</td>
</tr>
<tr>
<td></td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td><img src="image9.png" alt="Image" /></td>
<td><img src="image10.png" alt="Image" /></td>
<td><img src="image11.png" alt="Image" /></td>
<td><img src="image12.png" alt="Image" /></td>
</tr>
</tbody>
</table>
Infrastructure

Docker is used to run software packages called "containers". Containers are isolated from each other and bundle their own application, tools, libraries and configuration files;

OpenShift is a multifaceted, open source container application platform from Red Hat Inc. for the development, deployment and management of applications.

An open-source system for automating deployment, scaling, and management of containerized applications

An open source software product to allow OpenID Connect single sign-on with Identity Management and Access Management aimed at modern applications and services.

Under the stewardship of:
Create new account
My resources
Algorithms

Regression
- MLR - Weka (multi-response linear regression implemented in Java-WEKA)
- SVM - Weka (LibSVM, Support vector machines implemented in Java-WEKA)
- PLS - Weka (Partial Least Squares implemented in Java-WEKA)
- Linear Regression (Implemented in Python-Scikit Learn)
- Lasso Regression (Implemented in Python-Scikit Learn)
- PLS - with VIP scores (Implemented in Python)
- Random Forest (Implemented in Python-Scikit Learn)
- Multi-layer Perceptron (Implemented in Python-Scikit Learn)

Classification
- SVM - Weka (LibSVM) Implementation
- iSVM - with MCI (Implemented in Python-Scikit-Learn)
- iOD Decision Tree (Implemented in Python-Scikit-Learn)
- CMI Decision Tree (Implemented in Python-Scikit-Learn)
- Generalised Naive Bayes (Implemented in Python-Scikit-Learn)
- Multinomial Naive Bayes (Implemented in Python-Scikit-Learn)
- Bernoulli Naive Bayes (Implemented in Python-Scikit Learn)
- Random Forest (Implemented in Python-Scikit Learn)
- Multi-layer Perceptron (Implemented in Python-Scikit Learn)
- Gradient Boosting (Implemented in Python-Scikit Learn)
# Ontological annotation of algorithms

<table>
<thead>
<tr>
<th>Category / URI</th>
<th>Description</th>
<th>OpenTox Ontological Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preparation-additional</td>
<td>Scaling</td>
<td>&quot;ot:Algorithm&quot;, &quot;ot:Scaling&quot;, &quot;ot:Transformation&quot;</td>
</tr>
<tr>
<td>WEKA (Java)</td>
<td>MLR - Weka (multi-response linear regression implemented in Java-WEKA)</td>
<td>&quot;ot:Algorithm&quot;, &quot;ot:Regression&quot;, &quot;ot:SupervisedLearning&quot;</td>
</tr>
<tr>
<td>Python</td>
<td>Id3 - with MCI (Implemented in Python-Scikit-Learn)</td>
<td>&quot;ot:Algorithm&quot;, &quot;ot:Classification&quot;, &quot;ot:SupervisedLearning&quot;</td>
</tr>
</tbody>
</table>

OpenTox Algorithm Ontology [http://old.opentox.org/data/documents/development/RDF%20files/AlgorithmTypes](http://old.opentox.org/data/documents/development/RDF%20files/AlgorithmTypes)
Model training

Example

Modelling the Solubility of C60 fullerene in various solvents (Gharageizi et.al., 2008).

The full dataset of 124 solvents has been uploaded to Jaqpot and can be viewed in the URI
http://www.jaqpot.org/data_detail?name=nGF3G5SBo4wk5h

Training dataset of 93 solvents
http://www.jaqpot.org/data_detail?name=XmCQVC7o5jKKRv

Test dataset of 31 solvents
(http://www.jaqpot.org/data_detail?name=3UbgEJPldT2Ovs)
## Model training - Dataset selection

### Select dataset:

#### Example datasets:

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Contributors</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gażewicz_10_29</td>
<td>10 Metal Oxide NPs, used for predicting HaCaT toxicity.</td>
<td>Gażewicz et al.</td>
<td>2018-07-05</td>
</tr>
<tr>
<td>Gażewicz_10_29_class</td>
<td>10 Metal Oxide NPs, used for predicting HaCaT toxicity.</td>
<td>Gażewicz et al.</td>
<td>2018-07-05</td>
</tr>
<tr>
<td>Gażewicz_18_29</td>
<td>18 Metal Oxide NPs, used for predicting HaCaT toxicity.</td>
<td>Gażewicz et al.</td>
<td>2018-07-05</td>
</tr>
<tr>
<td>Gażewicz_18_29_class</td>
<td>18 Metal Oxide NPs, used for predicting HaCaT toxicity.</td>
<td>Gażewicz et al.</td>
<td>2018-07-05</td>
</tr>
<tr>
<td>Gażewicz_18_29</td>
<td>18 Metal Oxide NPs, used for predicting HaCaT toxicity.</td>
<td>Gażewicz et al.</td>
<td>2018-07-05</td>
</tr>
<tr>
<td>Gażewicz_18_29_class</td>
<td>18 Metal Oxide NPs, used for predicting HaCaT toxicity.</td>
<td>Gażewicz et al.</td>
<td>2018-07-05</td>
</tr>
</tbody>
</table>

#### All Datasets:

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Contributors</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>lrnC2VC7u59K9X8</td>
<td>Solubility of C60 fullerene in various solvents (training)</td>
<td></td>
<td>2019-05-14</td>
</tr>
<tr>
<td>nQg3GSSb4x55</td>
<td>Solubility of C60 fullerene in various solvents</td>
<td></td>
<td>2019-05-14</td>
</tr>
<tr>
<td>Sm5744uiB8REE6e5p</td>
<td>Numerical values of the calculated descriptors along with solubility of C60</td>
<td>Gharehgoz et al.</td>
<td>2019-05-14</td>
</tr>
<tr>
<td>new dataset</td>
<td>New dataset</td>
<td></td>
<td>2019-05-14</td>
</tr>
<tr>
<td>xRnZ023I1u43G</td>
<td>Numerical values of the calculated descriptors along with cytotoxicity values</td>
<td></td>
<td>2019-05-14</td>
</tr>
<tr>
<td>iterations</td>
<td>New dataset</td>
<td></td>
<td>2019-05-14</td>
</tr>
<tr>
<td>nQbpbvht1778i</td>
<td>Properties of metal oxides</td>
<td></td>
<td>2019-05-07</td>
</tr>
</tbody>
</table>

---

OpenRiskNet  | RISK ASSESSMENT E-INFRASTRUCTURE  | www.openrisknet.org
# Model training - Choice of algorithm

## Train model

### Choose Algorithm

#### Regression
- MLR - Weka (multi-response linear regression implemented in Java-WEKA)
- SVM - Weka (LibSVM, Support vector machines implemented in Java-WEKA)
- PLS - Weka (Partial Least Squares implemented in Java-WEKA)
- Linear Regression (Implemented in Python-Scikit Learn)
- Lasso Regression (Implemented in Python-Scikit Learn)
- PLS - with VIP scores (Implemented in Python)
- Random Forest (Implemented in R-base library)
- Gradient Boosting (Implemented in Python-Scikit Learn)
- Random Forest (Implemented in Python-Scikit Learn)
- Multi-Layer Perceptron (Implemented in Python-Scikit Learn)

#### Classification
- SVM - Weka (LibSVM) Implementation
- ID3 - with MCI (Implemented in Python-Scikit-Learn)
- ID3 Decision Tree (Implemented in Python-Scikit Learn)
- C4.5 Decision Tree (Implemented in Python-Scikit Learn)
- Generalised Naive Bayes (Implemented in Python-Scikit Learn)
- Multinomial Naive Bayes (Implemented in Python-Scikit Learn)
- Bernouilli Naive Bayes (Implemented in Python-Scikit Learn)
- Random Forest (Implemented in Python-Scikit Learn)
- Multi-layer Perceptron (Implemented in Python-Scikit Learn)
- Gradient Boosting (Implemented in Python-Scikit Learn)
Model training - Algorithm parameters

Algorithm
Title: python-lm
Title: Linear Regression implemented in Python Scikit-learn
Model name: Linear Model for C60 Fullerene Stoich
Model description: The model is provided in the following publication: Fahad Gharabeghi & Reza Farzighi Alamian (2008) A Molecular-Based Model for Prediction of Solubility of C60 Fullerene in Various Solvents, Fullerenes, Nanotubes, and Carbon Nanostructures. 16:1, 45-57, DOI: 10.1063/1.3623930701.77915

Select variables:
- Select input variable(s) and endpoint
- Select PMML file
- Select endpoint only (all other variables will be used as input variables)

Select variable(s) and endpoint:
Input variable(s): 
- Seipp
- Solvents
- log5 Exp
- H1m
- More23e
- ATS5m
- pPCO2

Endpoint:
- Seipp
- Solvents
- log5 Exp
- H1m
- More23e
- ATS5m
- pPCO2

Select scaling method:
- Scaling between zero and one

Select domain of applicability method:
- Leverage method

Train
Model training - Select variables

Select variables:
- Select Input variable(s) and endpoint
- Select PMML
- Upload PMML file
- Select endpoint only (all other variables will be used as input variables)

Select variable(s) and endpoint:

Input variable(s) | Endpoint
---|---
Select All | Select All
Seigp | Seigp
Solvents | Solvents
log5 Exp | log5 Exp
H1m | H1m
More23e | More23e
ATS1m | ATS1m
piPC03 | piPC03
Model training - Select variables

Select variables:
- Select input variable(s) and endpoint
- Select PMML
- Upload PMML file
- Select endpoint only (all other variables will be used as input variables)

Select endpoint:
- Seigp
- Solvents
- logS Exp...
- H1m
- More23e
- ATS1m
- pIGC03
- None

Select of applicability method:

Train
Model training - Scaling / DoA

Scaling options

Select scaling method:
- Scaling between zero and one
- None
- Normalization
- Leverage method

Applicability method:

Select domain of applicability method:
- Leverage method
- None
- Leverage method
- Train

Domain of Applicability (DoA) calculations
Task management

Task: Training on algorithm: python-Im

Status: COMPLETED
Type: TRAINING
Date: 05/14/19

Description:
Training task using algorithm python-Im
Model webpage

URI of model: http://www.jaqpot.org/m_detail?name=72GEEmGhhavY00n7O209

Model: #72GEEmGhhavY00n7O209
Title: Linear Model for C60 Fullerene Solubility
Description:
The model is provided in the following publication: Forhad Gharagheizi & Reza Faragehi Farmand (2008) A Molecular-Based Model for Prediction of Solubility of C60 Fullerene in Various Solvents, Fullerences, Nanotubes, and Carbon Nanostructures, 16:1, 40-57, DOI: 10.1080/15363830701779315
Transformations:
http://jaqpot.org/8090/jaqpot/services/model/4eqE44m2ACA1m0qgPeU6
Dos:
http://jaqpot.org/8090/jaqpot/services/model/A6DgjyW4rFe5aJ7cHy
Algorithm:
Python-im
Features:
Required Features
Dependent Features
Independent Features
Predicted Features
Representation:
Validation of models - Dataset choice/entry
Validation report

http://www.jaqpot.org/report?name=cnJlt1LZ3fnvIcl
## Validation report

<table>
<thead>
<tr>
<th>Compound</th>
<th>Real</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>compound1</td>
<td>-4</td>
<td>-3.91355789487</td>
</tr>
<tr>
<td>compound2</td>
<td>-2.6</td>
<td>-3.72266687304</td>
</tr>
<tr>
<td>compound3</td>
<td>-4.9</td>
<td>-4.1994532239</td>
</tr>
<tr>
<td>compound4</td>
<td>2.5</td>
<td>-4.26442038285</td>
</tr>
<tr>
<td>compound5</td>
<td>3.1</td>
<td>-2.23685026688</td>
</tr>
<tr>
<td>compound6</td>
<td>4.3</td>
<td>3.3976177416</td>
</tr>
<tr>
<td>compound7</td>
<td>5.2</td>
<td>-3.7723420825</td>
</tr>
<tr>
<td>compound8</td>
<td>6.9</td>
<td>-3.8210765451</td>
</tr>
<tr>
<td>compound9</td>
<td>7.2</td>
<td>-4.88724242914</td>
</tr>
<tr>
<td>compound10</td>
<td>8.2</td>
<td>-8.3504783112</td>
</tr>
<tr>
<td>compound11</td>
<td>5.7</td>
<td>-6.0156585465</td>
</tr>
<tr>
<td>compound12</td>
<td>6.7</td>
<td>-5.6106328885</td>
</tr>
<tr>
<td>compound13</td>
<td>6.7</td>
<td>-4.33212595712</td>
</tr>
<tr>
<td>compound14</td>
<td>2.5</td>
<td>-3.56827270109</td>
</tr>
<tr>
<td>compound15</td>
<td>3.3</td>
<td>-3.13184967997</td>
</tr>
<tr>
<td>compound16</td>
<td>3.4</td>
<td>-2.2390345204</td>
</tr>
<tr>
<td>compound17</td>
<td>3.8</td>
<td>-5.0782245678</td>
</tr>
<tr>
<td>compound18</td>
<td>3.2</td>
<td>-2.812257864</td>
</tr>
<tr>
<td>compound19</td>
<td>4.4</td>
<td>-4.6884361486</td>
</tr>
<tr>
<td>compound20</td>
<td>4.7</td>
<td>-4.9162777548</td>
</tr>
<tr>
<td>compound21</td>
<td>3.5</td>
<td>-2.87194312599</td>
</tr>
<tr>
<td>compound22</td>
<td>3.9</td>
<td>-3.59602881815</td>
</tr>
<tr>
<td>compound23</td>
<td>3.4</td>
<td>-3.2803840934</td>
</tr>
<tr>
<td>compound24</td>
<td>2.2</td>
<td>-2.4346738494</td>
</tr>
<tr>
<td>compound25</td>
<td>3.7</td>
<td>-4.1836838527</td>
</tr>
<tr>
<td>compound26</td>
<td>5.9</td>
<td>-3.87655881681</td>
</tr>
<tr>
<td>compound27</td>
<td>4.2</td>
<td>-4.1408958734</td>
</tr>
<tr>
<td>compound28</td>
<td>6.1</td>
<td>-3.650065068</td>
</tr>
<tr>
<td>compound29</td>
<td>2.9</td>
<td>-2.7826828953</td>
</tr>
<tr>
<td>compound30</td>
<td>3.4</td>
<td>-3.5830861919</td>
</tr>
<tr>
<td>compound31</td>
<td>3.3</td>
<td>-2.41555506252</td>
</tr>
</tbody>
</table>
Validation report
Validation report

The report is downloadable as PDF

External validation report

Description: [External validation with model:
http://jaqpot.org:8080/jaqpot/services/model/72GEEEmGhhavY00n7O209 and dataset:
http://jaqpot.org:8080/jaqpot/services/dataset/XmCQVC7o5jKKrV]

Model: 72GEEEmGhhavY00n7O209

Dataset: XmCQVC7o5jKKrV

Algorithm Type: REGRESSION

F-Value: 247.42

Number of predictor variables: 5

RMSE: 0.34

R^2 (OECD): 0.9

R^2 Adjusted (if applicable): 0.9

StdError: 0.35
Predictions
QPRF report

The QPRF (QSAR prediction reporting format) report generated by Jaqpot contains all the fields required by the OECD guidelines, namely:

- **Substance**
  - Contains information such as CAS and EC numbers, SMILES, InChi, etc.

- **General**
  - Information such as date and creator name and email.

- **Prediction**
  - Biological endpoint, variables, model used, DoA, etc.

- **Adequacy**
  - Optional field, containing regulatory purpose, conclusion etc
QPRF report

PCA of Query Instance vs. Training Dataset

3D Projection of Datapoints
QPRF report

The report is downloadable as PDF
Model Validation Schemes

Options:

- **Cross**: performs Cross Validation
- **Data Split**: splits data into training and test datasets according to a split ratio
- **External**: validation with an external dataset
Read Across

• Users select **distance metric** and a **threshold** between 0 and 1 (the closest to 0 the strictest) used to assess nanoparticle similarity.

• Optionally select a method to calculate the **prediction confidence level** (Currently “Nearest Neighbour Confidence”)

![Predicted values of dataset](image)
Jaqpot 5 - Login
Jaqpot 5 - Menu
Jaqpot 5 - Shared space - Models
Jaqpot 5 - QMRF report

1. QSAR identifier

1.1. QSAR identifier (title):
Linear nanoQSAR model predicting Solubility of C60 Fullerene in Various Solvents. The model has been presented in the publication "A Molecular Based Model for Prediction of Solubility of C60 Fullerene in Various Solvents" Farhad Gharagheizi & Reza Farzighi Almadari, Fullerenes, Nanotubes and Carbon Nanostructures, Volume 16, 2008 - Issue 1

1.2. Other related models:
- Neural Network nanoQSAR model predicting Solubility of C60 Fullerene in Various Solvents.

1.3. Software coding the model:
Jaqpot
Jaqpot is a web platform that supports development, validation and sharing of QSAR models
Haralambos Sarimveis
apps.jaqpot.org

2. General information

2.1. Date of QMRF:
21 April 2019
Jaqpot 5 - Model page

Share model

Read
- Control and Information
- NanoCommons

Write
- Control and Information
- NanoCommons

Execute
- Control and Information
- NanoCommons

Notes about sharing

Deleting is only available for the creator

When something is shared with Jaqpot it becomes available for all the users

The privileges are given to all the users of an organization shared with

View

Affiliated Orgs

1.2 Other related models:
Jaqpot 5 - Model page - Data tab

MODEL
Title: Neural Network Model for Predicting Solubility of C60 Fullerenes in Various Solvents
Owner: hsarmv

Dependent feature / Predicted feature

logS Exp.
Description: Feature created to link to independent feature of model Neural Network Model for Predicting Solubility of C60 Fullerenes in Various Solvents

Independent features

ATS1m
Description: Feature created to link to independent feature of model Neural Network Model for Predicting Solubility of C60 Fullerenes in Various Solvents

piPC03
Description: Feature created to link to independent feature of model Neural Network Model for Predicting Solubility of C60 Fullerenes in Various Solvents

Seigp
Description: Feature created to link to independent feature of model Neural Network Model for Predicting Solubility of C60 Fullerenes in Various Solvents
Jaqpot 5 - Model page - Predict tab

MODEL
Title: Linear Model for Predicting Solubility of C60 Fullerenes in Various Solvents
Owner: tsarinym
Description:
Linear Model for Predicting Solubility of C60 Fullerenes in Various Solvents

Choose method
Predict
Validate

Upload dataset with the required independent features and values

Input values for the independent features
piPC03
ATS1m
Selgp
More23e
H1m
Jaqpot 5 - Predict tab - Entering values

MODEL
Title: Linear Model for Predicting Solubility of C60 Fullerenes in Various Solvents
Owner: hassimv
Description:
Linear Model for Predicting Solubility of C60 Fullerenes in Various Solvents

Choose method
Predict

Upload dataset with the required independent features and values

Input values for the independent features

<table>
<thead>
<tr>
<th>Feature</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>pIC03</td>
<td>2.485</td>
</tr>
<tr>
<td>AT81m</td>
<td>2.197</td>
</tr>
<tr>
<td>H1m</td>
<td>0.358</td>
</tr>
<tr>
<td>More20e</td>
<td>-1.227</td>
</tr>
</tbody>
</table>
Jaqpot 5 - Predict tab - Downloading template

Upload dataset with the required independent features and values

Download template dataset (csv)

Input values for the independent features

plPC03
ATS1m
Ssigp
H1m
More23e
Upload dataset with the required independent features and values

Input values for the independent features

- pIPC03
- ATS1m
- Ssigp
- H1m

More23e
<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Seigp</td>
<td>ATS1m</td>
<td>piPC03</td>
<td>More23e</td>
<td>H1m</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>2.485</td>
<td>2.303</td>
<td>-2.598</td>
<td>0.249</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>2.639</td>
<td>2.485</td>
<td>-3.047</td>
<td>0.263</td>
</tr>
</tbody>
</table>
Upload dataset with the required independent features and values

Upload dataset (should at least contain independent features)

Input values for the independent features

- piPC03
- ATS1m
- Seigp
- H1m

- More23e
Jaqpot 5 - Predict tab - Uploading data

Upload dataset with the required independent features and values

<table>
<thead>
<tr>
<th>id</th>
<th>ATS1m</th>
<th>Selgp</th>
<th>More23e</th>
<th>H1m</th>
<th>piPC03</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.792</td>
<td>0</td>
<td>-0.521</td>
<td>0.197</td>
<td>1.792</td>
</tr>
<tr>
<td>1</td>
<td>2.639</td>
<td>0</td>
<td>-3.047</td>
<td>0.263</td>
<td>2.485</td>
</tr>
</tbody>
</table>

Erase dataset

Start procedure

Start prediction with input dataset
Jaqpot 5 - Predict tab - Entering values

Input values for the independent features

<table>
<thead>
<tr>
<th>Feature</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>piPC03</td>
<td>2.485</td>
</tr>
<tr>
<td>ATS1m</td>
<td>2.197</td>
</tr>
<tr>
<td>H1m</td>
<td>0.358</td>
</tr>
<tr>
<td>More23e</td>
<td>-1.227</td>
</tr>
</tbody>
</table>

Start prediction with input values
**Jaqpot 5 - Predictions**

Starting Prediction...
Prediction completed successfully.
Dataset was built successfully.
Now saving to database...
Task Completed Successfully.

View predicted value only

<table>
<thead>
<tr>
<th>Id</th>
<th>Saip</th>
<th>logS Exp.</th>
<th>AT'S1m</th>
<th>Mora22e</th>
<th>H1 m</th>
<th>pIPG03</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>-4.186807750539802</td>
<td>2.197</td>
<td>-1.227</td>
<td>0.358</td>
<td>2.485</td>
</tr>
</tbody>
</table>

Download
**Jaqpot 5 - Predictions as CSV**

**Image Description:**
- A screenshot of the Jaqpot platform showing the predictions as CSV.
- The platform interface displays a message indicating that the dataset has been retrieved successfully.
- A table with predicted values is visible, including columns for ID, Sigg, log10 Exp., ATSE, MnxSE, ITrm, and pPIC22.
- The table shows a single row with predicted values for Sigg, log10 Exp., ATSE, MnxSE, ITrm, and pPIC22.

**Table Content:**
<table>
<thead>
<tr>
<th>ID</th>
<th>Sigg</th>
<th>log10 Exp.</th>
<th>ATSE</th>
<th>MnxSE</th>
<th>ITrm</th>
<th>pPIC22</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>-4.186807760539832</td>
<td>2.197</td>
<td>-1.227</td>
<td>0.398</td>
<td>2.485</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
</tr>
<tr>
<td>---</td>
<td>----</td>
<td>------</td>
<td>---------------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>1</td>
<td>My Report</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Id</td>
<td>Seigp</td>
<td>logS Exp.</td>
<td>ATS1m</td>
<td>More23e</td>
<td>H1m</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>-4,97982</td>
<td>1,792</td>
<td>-0,521</td>
<td>0,197</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>-3,96677</td>
<td>2,639</td>
<td>-3,047</td>
<td>0,263</td>
</tr>
</tbody>
</table>
Jaqpot 5 - Social network of models
Jaqpot 5 - User profile

Philip Doganis
filipposd

Organizations
On the Internet
Quota

Organizations
Organizations I am a member

NanoCommons

Jaqpot
Athens, Greece

Lab of Procass Control and Informa...
Athens, Greece

OpenRiskNet

CREATE
Jaqpot 5 - User profile - Quota

Philip Doganis
filipposd

Organizations
On the internet
Quota

Quota
My quota

Capabilities

Models
20

Algorithms

Reports
20

Tasks in parallel
5

Datasets
20

Organizations
3

About

Occupation

Occupation at

Lives at city

Country

Api key

eyJhbGciOiJSUzI1NiIsInR5cCI6IkpXVCJ9.eyJzdWIiOiJib3g2NzgyMCIsInJvbGUiOiJib3g2NzgyMCIsImlzaWQiOiJib3g2NzgyMCIsImNvbXQiOiJib3g2NzgyMCIsInN1YiI6IiIsInZ1bGUiOjEwMCwiYXV0aG9yIjoiaHR0cHM6Ly9yZXlut3dpMy5jb20iLCJlbnNpdGlvbiI6ImF1dG9uc2VfZWNobyIsImJlYWRlcnNpb24iOiJyZXlut3dpMy5jb20ifQ..
Jaqpot 5 - Organizations

OpenRiskNet

Contact: hsrinv@central.ntua.gr

About
Jaqpot 5 - OpenRiskNet Organization page
Jaqpot 5 - Sharing of resources
Jaqpot 5 - Sharing notification

MODEL SHARED

model shared through organization OpenRiskNet

<table>
<thead>
<tr>
<th>From user</th>
<th>hsarimv</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shared with</td>
<td>OpenRiskNet</td>
</tr>
<tr>
<td>model</td>
<td>Shin et al. Cytotoxicity classification Model</td>
</tr>
</tbody>
</table>

View  remove
Jaqpot 5 - Shared model

Logistic Regression Model based on the following publication:


SAR and QSAR in Environmental Research, 2017 VOL. 28, NO. 11, 875-888 https://doi.org/10.1080/1062936X.2017.1460098

Model in PMML Format

```xml
<xml version="1.0" encoding="UTF-8" standalone="yes"?>
<PMML version="4.3" xmlns="http://www.dmg.org/PMML-4.3" xmlns:data="http://pml露天.org/pml-mml-model/TslineTable" version="4.3">
<Header>
  <Application name="PMML-5A2Learn" version="1.5.11"/>
  <TimeStamp>2018-06-10T10:43:30Z</TimeStamp>
</Header>
<MiningBuildTask>
  <Extension Name="PMMLPipelineStep">
    <sage scaler="MinMaxScaler(copy=True, feature_range=(0, 1))",
      LOCAINES, logisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
      intercept_scaling=1, max_iter=100, multi_class='warn',
      n_jobs=1, penalty='l2', random_state=None, solver='warn',
      tol=0.0001, verbose=0, warm_start=False))"></Extension>
  </Extension>
  <Extension Name="PMMLBuildTable"/>
</MiningBuildTask>
<Dictionary>
  <DataField name="Hosu" optype="categorical" dataType="integer">
    <Value value="0"/>
    <Value value="1"/>
  </DataField>
  <DataField name="Hosur_hoho" optype="continuous" dataType="double"/>
  <DataField name="pol_dil_hyd" optype="continuous" dataType="double"/>
</Dictionary>
<TransformationDictionary>
  <DerivedField name="mix_max_scaler(Hosur_hoho)" optype="continuous" dataType="double"/>
  <DerivedField name="mix_max_scaler(pol_dil_hyd)" optype="continuous" dataType="double"/>
</TransformationDictionary>
</PMML>
```
**MODEL**

**Title:** ORN consensus RFE 5

**Description**

Logistic Regression+RFE Binary Blood-Brain-Barrier Penetration Data after applying Recursive Feature Elimination (Penetrating/Non-Penetrating).

---

**Overview**

"Model for Binary Blood-Brain-Barrier Penetration used on LogisticRegression."

Model was trained with 10-fold CV after applying Recursive Feature Elimination (Penetrating/Non-Penetrating) on the initial dataset in order to reduce the number of variables to 20.


Jaqpot 5 - Datasets
Jaqpot 5 - Datasets - Uploading new dataset
Jaqpot 5 - Datasets - Uploading new dataset
Jaqpot 5 - New dataset - Dataset id

Dataset title: Nanoreg Substances
Apr 1, 2019

Dataset

Filename: XRFESYall.csv

- EState_VSA2
- RotRatio
- JGI5
- JGI8
- Blood-Brain-Barrier Penetration
Jaqpot 5 - Dataset creation prompt
Title: Blood-Brain-Barrier Penetration

Description:
Jaqpot 5 - Dataset page - Discussion

DATASET
Title: Blood-Brain-Barrier Penetration

Owner: filipposd

Description:
Binary Blood-Brain-Barrier Penetration Data after applying Recursive Feature Elimination (Penetrating/Non-Penetrating).

Leave a comment

filipposd, Jun 10, 2019
I have a suggestion for the ontological annotation of the ATSC8I term. Please read my article at Medium: http:// ...

Reply

filipposd, Jun 10, 2019
What parameter values did you use?

Reply

filipposd, Jun 10, 2019
I have used method X and achieved 0.yz accuracy. More details at https://app.jaqpot.org/model/Hmc3S3VhCADYzK07Vet

Reply
Jaqpot in EOSC Catalogue

Jaqpot
Generate, store and share predictive statistical and machine learning models

www.openrisknet.org

https://catalogue.eosc-portal.eu/service/openrisknet.jaqpot

OpenRiskNet
RISK ASSESSMENT E-INFRASTRUCTURE

www.openrisknet.org
Question 3

Are you interested in deploying a predictive model through Jaqpot?
   a. Yes
   b. No
Conclusion

● Jaqpot is an **open source web infrastructure** with a flexible structure:
  ○ generalization across disciplines
  ○ integration to third-party tools in order to fulfil modelling needs on the Cloud
  ○ easy integration to diverse architectures/ontologies/knowledge domains.

● Jaqpot offers a seamless way to:
  ○ **take models out of the desktop or paper** and automatically make them available as web services via their URI
  ○ functionality is provided both through an **API and GUIs**.

Jaqpot constitutes a **universal platform** that **exchanges and produces** semantically annotated datasets and employs machine learning techniques to generate **predictive models**.
Acknowledgements

**OpenRiskNet** (Grant Agreement 731075) is a project funded by the European Commission within Horizon 2020 Programme

**Project partners:**

- P1 Edelweiss Connect GmbH, Switzerland (EwC)
- P2 Johannes Gutenberg-Universität Mainz, Germany (JGU)
- P3 Fundacio Centre De Regulacio Genomica, Spain (CRG)
- P4 Universiteit Maastricht, Netherlands (UM)
- P5 The University Of Birmingham, United Kingdom (UoB)
- P6 National Technical University Of Athens, Greece (NTUA)
- P7 Fraunhofer Gesellschaft Zur Foerderung Der Angewandten Forschung E.V., Germany (Fraunhofer)
- P8 Uppsala Universitet, Sweden (UU)
- P9 Medizinische Universität Innsbruck, Austria (MUI)
- P10 Informatics Matters Limited, United Kingdom (IM)
- P11 Institut National De L'environnement Et Des Risques INERIS, France (INERIS)
- P12 Vrije Universiteit Amsterdam, Netherlands (VU)