

Online QSAR Modelling Hackathon by Easy Access to Jaqpot: Deploy your model as a web service in a few minutes

Tue, Apr 13, 2021 3:00-4:30 PM CEST

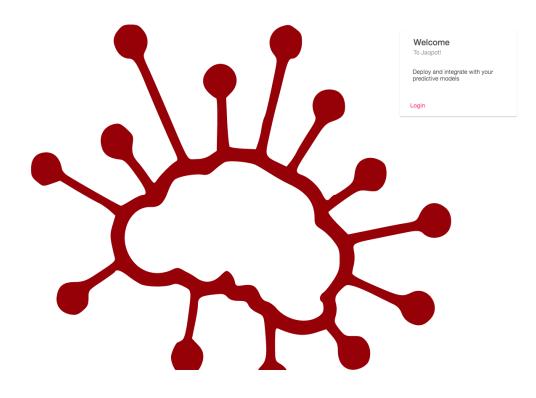
Briefly

- You will learn how a local QSAR model can be transformed to an online QSAR model web service, no installations required
- Prerequisites:
 - For access to Jaqpot, you will need a Jaqpot account
 - For access to Google collab you will need a Google account

Introduction	3
Problem studied	3
Accessing Jaqpot	4
Retrieving the dataset from NanoPharos	6
Accessing the Google colab notebook	6
Running the colab notebook to create the model web service on Jaqpot	8
1. Setting up packages	8
2. Uploading the data to the colab notebook	8
3. Viewing and Preprocessing the dataset	9
4. Training the model	9
5. Getting authentication to access Jaqpot	9
6. Deploying the model	10
ADDITIONAL STEPS	12

Introduction

Jaqpot 5 (<u>https://app.jaqpot.org</u>) is a powerful and versatile platform for toxicological in silico predictions, allowing you to deploy machine learning models incredibly easy and make them available to the community as web services.



Problem studied

In this tutorial we will demonstrate how to reproduce and publish in Jaqpot 5, a Linear nanoQSAR model predicting Solubility of C60 Fullerene in Various Solvents, with just a few lines of code in a Colab notebook. The model has been originally presented in the following publication: Farhad Gharagheizi & Reza Fareghi Alamdari (2008) *A Molecular-Based Model for Prediction of Solubility of C60 Fullerene in Various Solvents*. Fullerenes, Nanotubes and Carbon Nanostructures, 16:1, 40-57, DOI: 10.1080/15363830701779315. https://www.tandfonline.com/doi/full/10.1080/15363830701779315



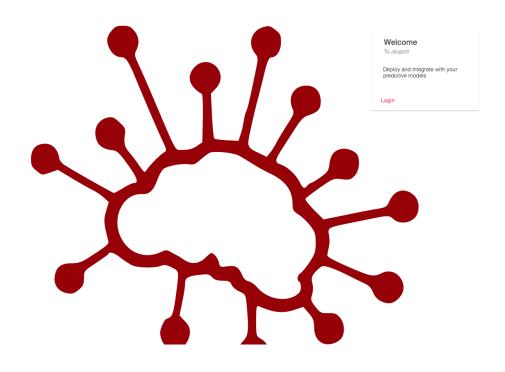
In this presented work, a quantitative structure-property relationship study (QSPR) was done for prediction of solubility of C_{60} fullerene in various solvents. In this study, genetic algorithm-based multivariate linear regression (GA-MLR) was applied to obtain most statistically effective molecular descriptors on solubility of C_{60} in various solvents. All of these molecular descriptors are only calculated from the chemical structure of solvents. For considering nonlinear behavior of appearing molecular descriptors in GA-MLR section, a feed forward neural network (FFNN) was constructed and optimized for prediction of solubility of C_{60} fullerene in solvents. Obtained models considerably showed better accuracy in comparison with the previous models.



1. Accessing Jaqpot

In order to access Jaqpot, you need a Jaqpot account. For the purposes of this workshop, please avoid using the Google/Github option, to enable a more uniform approach among participants. If you are a new user, you can create a Jaqpot user account following the instructions provided in a separate document.

First click on Login on the welcome page:



You are then redirected to your Jaqpot homepage, which gives you access to your models and datasets.

≡	Jaqpot 🖄		۹	۵	٥	θ
A	Home	Quick view				≣
e	Datasets Shared / Private			No it	em selec	ted
	Models Shared / Private					
Î	Trash					

Currently you don't have any models or datasets. A populated homepage looks like this:

≡	Jaqpot	diarran o		٩	128	١	8
A	Home		Quick view				≡
0	Datasets Shared / P	*	Model title: :Linear Model for Predicting Solubility of C60 Fullerenes in Various Solvents Mar 11, 2021		fo	inear 1 r Pred olubili	icting ty of
	Models Shared / P	8	Dataset title: D5.7 Example Jan 11, 2021		_	C60 ullerer Vario Solve	nes in Sus
Î	Trash	8	Dataset title: C60 Fullerene solubility Jun 21, 2019			Mode	
		8	Dataset title: NanoCommons Hackathon dataset Jun 21, 2019		Cr	Owne filippo eated: N	osd
		8	Dataset title: Blood-Brain-Barrier Penetration Jun 10, 2019]	202 Descrip Line	tion: ar
						Mod	el

Each time you access Jaqpot on Google Collab (as we will see in a subsequent step) you need to identify yourself, in order to be granted access to your private resources by providing your username and password.

2. Retrieving the dataset from NanoPharos

The dataset can be retrieved in ready-to-model format from the NanoPharos database, which is accessible

at:	https://	db.nanopharos.eu/Queries/Datasets.zul	

S NanoPharos Datasets' Query					
Select a dataset by name/paper		Description			
Solubility of C60 Fullerene in Various Solvents	•	Farhad Gharagheizi & Reza Fareghi Alamdari (2008) A Molecular-Based Model for Prediction of Solubility of C60 Fullerene in Various Solvents. <i>Fullerenes, Nanotubes, and Carbon Nanostructures</i> , 16:1, 40-57, DOI: 10.1080/15363830701779315. https://www.tandfonline.com/doi/full/10.1080/15363830701779315			
Download dataset					

Please select the Solubility of C60 Fullerene in Various Solvents entry from the drop-down menu on the

left and click the button. An Excel file named '**70_model_reduced.xlsx**', will be downloaded and stored in your local computer.

3. Accessing the Google colab notebook

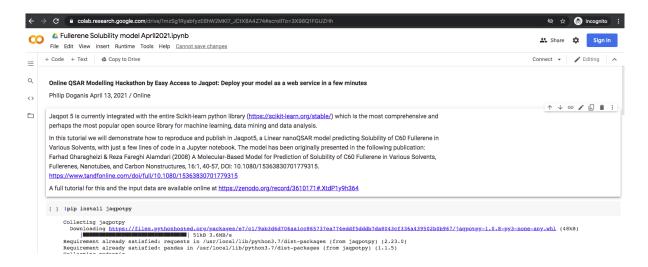
The Google colab platform allows users to run code in the Python programming language without needing to perform any installation in their local machines. A Google colab notebook that performs all the calculations and contains all the commands for model creation and deployment has been prepared for you and is available at:

https://drive.google.com/file/d/1mzSg1RyabfyzE6hW2MKI7_JCtX8A4Z74/view?usp=sharing

If a screen like this appears, please click on "Open with Google Colaboratory"

e Solubility model April2021.ipynb	CO Open with Google Colaboratory	■	Ð	•	:
<pre>("name":"supthon", "version":3), "file_extens python", "name":"python", "hocovert exporter April2021.ipynb", "provenance":[), "collapsed ["****online GSAB Modelling Hackstono by En 13, 2021 / Online"]), ("cell type", "markdonn linrary (htps://collt-learn.org/stable/) data analysis. \n", "\n", "In this tutorial w Fullerene in Various Solvents, with just a Gharagheii G Rera Farephi Alamdari (2008) Carbon Nonstructures, 16:1, 40-57, DOI: 10. and the input data are available online at packages\n", "\n", "\n"]). ("cell type": "code" ("status": "oko", "timestamp": 16178201531," Doganis", "photoUrl": "", "userId", "0515790350 jagotpy"), "execution_count": "1, "outputs" [[</pre>	<pre>("kernelspec":("display_name":"Fython 3", "isngage": "python", "name": "python3"), "language_info tom": "python", "mimetype":"text/x- ''python", "pygments_lexer": "spython3", "wersion": A.T. 6"), "colab":("name": "Fullerene Solubilit sections":(], "tocc visible":true)), "colls":("coll_chr": "markdown", "metadta"("id":"CORTWIGX") y Access to Jagobi Deploy your model as a web servici nn a few minutes****'n", "("a", "in", "bh' ", "metadta":("id":"SS\$901F0UIXh"), "source":("Jagobi 5 % oursently integrated with the entire which is the most comprehensive and perhaps the most populs open source library for machine 1 e will demonstrate how to reproduce and publish in Jagobi5. Jinear nancOSAR model predicting Bew lines of code in a Jupyter notebook. The most populs open source library for machine 1 e will demonstrate how to reproduce and publish in Jagobi5. Jinear nancOSAR model predicting 1000/1536380701779315. https://www.tandfoilne.com/doi/ful/10.10/15363830701779315.https://www.tandfoilne.com/doi/ful/10.10/15363830701779315.https://bymla/ ttps://senodo.cog/record/361011"]), ("cell_type":"markdown", "metaata": ("di":"BoxYafbaf", colab: "[baget uri" "https://senodo.cog/record/3610171]), ("cell_type":"markdown", "metaata": ("dis"); "display(11*3953394*dd)=doi-de=47cd35064607), "source": ("ipip install Gatput_type":"stream", "text": ("Collecting jagotpyln", "ud01b/1231 Downloading /92dbabe50134505/fb5040celabd0e50256fe00c2526fe00c2526fe10d5074360507-1.0.11-w3-none-axv.v</pre>	y model J2HW"),"source": iip Doganish","\n","April Soikit-learn python earning, data mining and Solubility of C60 ing publication: Farhad lerenes, Manotubes, and "% full tutorial for this "source":["#1. Setting up ":			

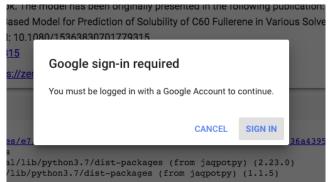
The Google colab notebook starting page looks like that:



In order to be able to use the notebook and make any changes, please save a copy of the notebook in the Drive by clicking the menu as shown below:

$\leftarrow \ \rightarrow$	C C colab.research.google.com	n/drive/1mzSg1RyabfyzE6hW2MKI7_JCtX8A4Z74#scrollTo=3X98Q1FGUZHh
co	Fullerene Solubility model A File Edit View Insert Runtime T	
😑 Ta	Locate in Drive	Code + Text 🔥 Copy to Drive
Q [New notebook Open notebook	$_{\ensuremath{\texttt{H}/Ctrl+O}}$)nline QSAR Modelling Hackathon by Easy Access to Jaqpot: D
\diamond	Upload notebook	hilip Doganis April 13, 2021 / Online
		aqpot 5 is currently integrated with the entire Scikit-learn pythor erhaps the most popular open source library for machine learning
	Save a copy in Drive	this tutorial we will demonstrate how to reproduce and publish
	Save a copy as a GitHub Gist	'arious Solvents, with just a few lines of code in a Jupyter noteb

This can be done seamlessly when you are logged in as a Google user, if however you receive the message below you need to log in.



Now you have a local copy of the notebook that you modify and run. This is in your private space managed by Google.

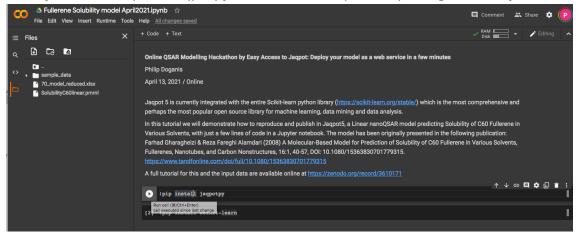
4. Running the colab notebook to create the model web service on Jaqpot

The notebook contains both text with comments, and cells containing python commands that will handle the dataset, make the necessary calculations and deploy the model as a web service. Green text in code cells are comments and are not executable.

1. Setting up packages

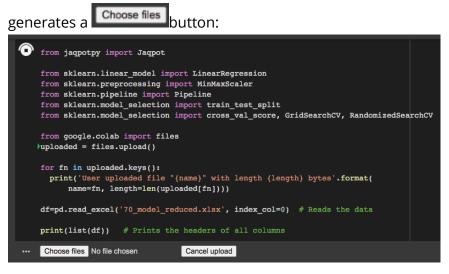
In order to execute code on the collab notebook you can go to the left of each cell and hit the Line icon. You see the outcome of the cell expanding right below the cell.

First, you will set up the Jaqpotpy, scikit-learn and pandas packages, so they can be used within the notebook:



2. Uploading the data to the colab notebook

In this step, you will upload the dataset retrieved from NanoPharos. Executing the code in this cell



Clicking the button leads to a file browser window, where you can locate the file to be uploaded, which here is the 70_model_reduced.xlsx file.

	Vorkshop	٢		earch
Name				Date
🐴 70_m	odel_reduced.xlsx			5 Ap
_				
_				
			Cancel	Open

3. Viewing and Preprocessing the dataset

In this step, you will:

- take a first look at the dataset
- generate descriptive statistics
- split the dataset randomly into training and test sets consisting of 75% and 25% of the data respectively
- generate the pipeline that will do:
 - preprocessing: transform features by scaling each feature to a given range using the MinMaxScaler function: <u>https://scikit-</u> <u>learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html</u>
 - modelling: The linear regression algorithm will be applied to the training dataset to generate the model

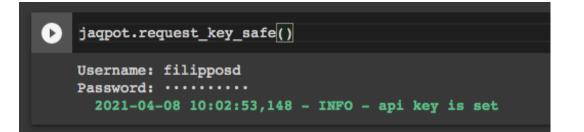
4. Training the model

Here you will:

- train the model
- print model performance metrics on the training, test dataset and on the total dataset
- perform a 5-fold cross validation test (<u>https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.cross_val_score.html</u>).

5. Getting authentication to access Jaqpot

Please provide your username, press enter and then provide your password in the prompt that emerges.



6. Deploying the model

You will now deploy the model using the *jaqpot.deploy_sklearn* function with the following arguments:

Pipelinelinear	the pipeline to be used for preprocessing the data and the algorithm to be applied.
Xall	Dataset with the input (known) features.
Yall	Dataset with the output feature (the feature we want to predict with the model).
title="Solubility of C60 Fullerenes in Various Solvents- USERNAME"	Title of the model, as it will appear in the Jaqpot user interface. Please change this by adding your username or any other character string that can distinguish your model and avoid multiple entries with the same model name.
description="Description"	A short description intended to inform users about the model.
model_meta=True	By choosing True we enable listing of automatically generated model metadata (version of sklearn used, details on pipeline, preprocessing transformations and algorithm used for model). False disables this.
doa=X_train	Providing the dataset here enables Domain of Applicability calculations based on the leverage method. False disables this.

The command with the arguments listed above is:

Please change USERNAME by any string of characters that can distinguish your model.

The web service has been created in Jaqpot and has a unique **modelURI** of the following form: <u>https://app.jaqpot.org/model/99vBdRvvD8NoiFA6IN17</u>.



Now your QSAR model is a web service! You can access your model through the Jaqpot interface

https://app.jaqpot.org

ADDITIONAL STEPS

After completing the steps above, you can take a look at the following:

- 1. Accessing your model through the colab notebook
- 2. Share the model to Organisation
- 3. Adding meta information to your model: QMRF report
- 4. Test and use your Jaqpot model
- 5. More information on the Jaqpot platform

1. Access the model through the colab notebook

You can use the **jaqpot.predict** function in your colab notebook providing the dataset name (for example **Xall**) and the **modelID** as arguments in order to get predictions from the model, which are stored in the predictions variable:

```
[ ] predictions = jaqpot.predict(Xall, modelId)
      2021-04-07 21:44:14,731 - INFO - completed 10.0
      2021-04-07 21:44:16,281 - INFO - completed 100.0
[] predictions
                            piPC03 ATS1m More23e
                                                       H1m logS Exp. Seigp
                             1.099 1.609 -0.922 0.148 -5.818155
                                                                          0.0
     pentane
                             1.386 1.792
                                            -1.156 0.173 -5.419509
                                                                          0.0
     hexane
                                            -1.656 0.208 -4.862671
-1.529 0.242 -4.839394
                             1.792 2.079
                                                                          0.0
     octane
                             1.792 2.079
                                                                          0.0
     iso-octane
     decane
                             2.079 2.303
                                            -2.151 0.233 -4.479601
                                                                          0.0
                                            -0.446 0.394 -4.148183
      pyridine
                             3.056 1.992
                                                                         -0.6
                                            -0.750 0.591 -2.817290
      quinoline
                            4.123 2.512
                                                                        -0.6
                                            -0.504 0.351 -3.839706
-0.462 0.407 -3.556255
-0.435 0.399 -3.278345
                             3.248 2.100
      aniline
                                                                         -0.6
      N-methylaniline
                             3.359 2.234
                                                                         -0.6
      N.N-dimethylaniline
                            3.458 2.351
                                                                         -0.6
     [124 rows x 6 columns], 'logS Exp.')
```

2. Share your model with the WorkshopApril2021 group

Access your model through the Jaqpot application: <u>https://app.jaqpot.org</u>:

← -	→ C ■ appjaqpotorg/home				
≡	🗉 Jaqpot 🌺				
A	Home I	Models > Mine >	items per page: 20 1 - 1 of 1 < > 📰		
e	Datasets Shared / Private	Model title: Solubility of C60 Fullerenes in Various Solvents-hsarimv Apr 11, 2021	Solubility of C60 Fullerenes in Various Solvents-hsarimy		
	Models Shared / Private		Model		
Î	Trash		Owner: Harry Sarimveis Created: Apr 11, 2021		
			Description: description		

Select your model and click on the "View" icon

	C 🗎 appjagpotorg/home	☆ 🙂 🗄
=	Jaqpot 🌺	९ 🌲 🕚 😫
ŧ	Home Quick view	
e	Datasets Shared / Private Model title Solubility of C60 Fullerenes in Various Solvents-hsarimv Apr 11, 2021	Solubility of C60 Fullerenes in Various Solvents-hsarimv
	Models Shared / Private	Model
Î	Trash	Owner: Harry Sarimveis
		Created: Apr 11, 2021
		Description: description
	Edit	
At t	he bottom right of the model page, click on the edit button	
	Share	

Clicking the Share button gives you the option to share the model with organisations of choice. Please select the WorkshopApril2021 group and give Execute rights only:

Share model		
	Execute	
Write	WorkshopApril2021 😣	
	These Organizations will be able to use the dataset	
		•
Notes about sharing		
Deleting is only availabl	le for the creator	
When something is shar	ed with Jaqpot it becomes available for all the users	
The priviledges are give	n to all the users of an organization whared with	
View		•
4		
>		

Adding an organisation to share with, activates the Share button, turning it from grey

to blue

. Clicking this button causes the "Successfully shared" message to appear:

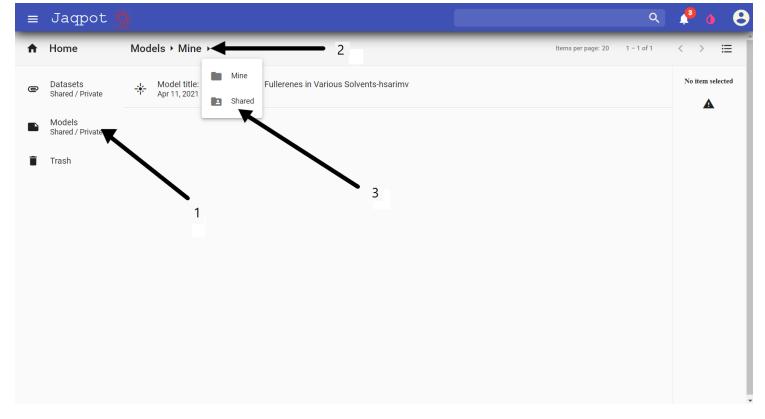
>

Share model			
Write	Execute WorkshopApril2021	Succesfully shared	
Notes about sharing Deleting is only available for the	creator		Â
When something is shared with J	aqpot it becomes available for	all the users	
The priviledges are given to all t	ne users of an organization wha	ared with	
View 4			+
>			

In order to view the shared models, please click on the menu button and select Models:

≡	Jaqpot	۹ مر	1 0 8
A	Home	Quick view	i i i
API's		Model title: Solubility of C60 Fullerenes in Various Solvents-hsarimv Apr 11, 2021	No item selected
BIOKINE			
	Httk		
Info	0		

Select Models [1], click the triangle icon next to Mine [2] and select Shared [3]



Select the **WorkshopApril2021** organisation. You are directed to a page containing all models shared through the **WorkshopApril2021** organisation.

≡	Jaqpot 🛓	२	1 0 8
A	Home	Models → Shared → With WorkshopApril2021 ▼ Items per page: 20 1 - 2 of 2	< > ≡
e	Datasets Shared / Private	Model title: Solubility of C60 Fullerenes in Various Solvents-hsarimv Apr 11, 2021	No item selected
	Models Shared / Private	Model title: Linear Model for Predicting Solubility of C60 Fullerenes in Various Solvents Mar 31, 2021	
Î	Trash		

3. Adding meta information to your model

You can add any other information about the model over the user interface at <u>https://app.jaqpot.org/</u> (for example detailed description, standard reports like Quantitative Model Reporting Format (QMRF), PMML representations, ontological annotations etc., descriptions of variables, ontological classes etc.). To demonstrate this, a fully documented model titled **"Linear Model for Predicting Solubility of C60 Fullerenes in Various Solvents"** is available at <u>https://app.jaqpot.org/model/RqCRtRpY85kpbgGtsiXp</u> and has been shared with the **WorkshopApril2021** organisation:

= Jaqpot 🌺		۹	J 🖗	• •
	Overview Features Predict / Validate Discussion Model meta			_
MODEL Title: Linear Model for Predicting Solubility of C60 Fullerenes in Various Solvents Commer Janon Solvents Commer Janon Solvents	Linear nanoQSAR model predicting Solubility of C60 Fullerene in Various Solvents. The model is provided in the following publication: Farhad Gharaghezi & Raza Fareghi Alamdari (2008) A Molecular-Based Model for Prediction of Solubility of C60 Fullerene in Various Solvents, Hanotubes, and Carbon Konstructures, 16:1, 40-57, DOI: 10.1080/15363830701773315 Ful dataset is available in this link Training dataset is available in this link Test dataset is available in this link Adownoodable OMBF Report is available in this link			
Linear Model for Predicting Solubility of C60 Fullerenes In Various Solvents	A Commandation of the second			
	2. General Information 2.1.Date of QMRF:			
	21 April 2019 2.2.QMRF author(s) and contact details:			•

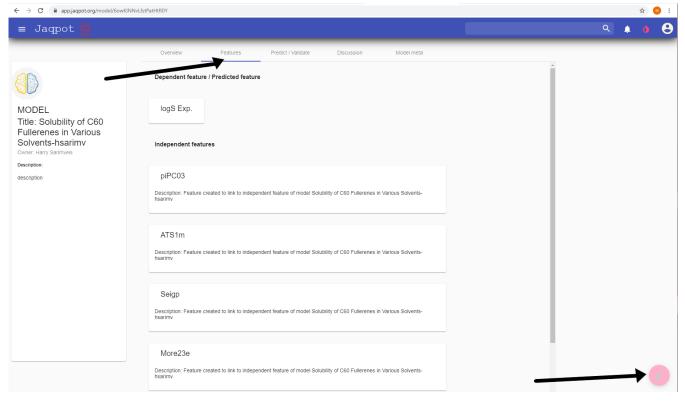
You can add information to your model by selecting the overview tab [1] and then clicking on the edit button.

← → C (app.jaqpot.org/model/6owK3)	iNvLfstPatHtR0Y						야 ☆ 😶 :
= Jaqpot 🖄							९ 📌 । 😌
MODEL Title: Solubility of C60 Fullerenes in Various Solvents-hasrimv	Overview	Features	Predict / Validate	Discussion	Model meta		
Dencer tary sames Description: description	•						
						2	

You can write any text (Markdown language is supported) and save it by clicking on the save button on the bottom right part of the page. You can return back and re-edit or extend the model description at any time.

\leftarrow \rightarrow C \cong app.jaqpot.org/model/6owKINNvl	.fstPatHtR0Y							07 1	¢ 🕖	:
= Jaqpot 🖄							٩	1		0
	Overview	Features	Predict / Validate	Discussion	Model meta					
<pre>verify and the second se</pre>	Edt overview *	t model								
								•	C	

Information about the variables used in the model (descriptions, units, ontological classes) can be added by selecting the "Features" tab and clicking on the Edit Button



After entering the information, please click on the Save Button to store this information to the system. Descriptions can be updated and extended at any time.

\leftrightarrow \rightarrow C $($ app.jaqpot.org/model/6owKINNvLt	fstPatHtR0Y	☆ 🙂 :
≡ Jaqpot 🌺		۹ 🖡 ۵ 😝
	Overview Features Predict / Validate Discussion Model meta	
	Dependent feature / Predicted feature	
MODEL	logS Exp. Description Add a description here	
Title: Solubility of C60 Fullerenes in Various Solvents-hsarimv	Vinits	
Owner: Harry Sarimveis Description description 4	// Ontological Classes	
Tags:	<u>Å</u>	
•	Independent features	
	piPC03 Description Feature created to link to independent feature of model Solubility of C60 Fullerenes in Various Solvents-hsarimv	
	Units #	
	Ontological Classes	
	ATS1m Description Feature created to link to independent feature of model Solubility of C60 Fullerenes in Various Solvents-hsarimv #	

To assist the users in adding editable versions of QMRF reports, we have created a QMRF markdown template, which can be downloaded at:

https://github.com/ntua-unit-of-control-and-informatics/QSAR-Models/blob/master/QMRF%20template.md .

by right clicking on the **Raw** button, as shown:

P master - QSAR-Models / QMRF template.md	Go to file
A dphilip QMRF TEMPLATE	Latest commit d89781d on 21 Jun 2019 🕥 History
Rx 1 contributor	
266 lines (202 sloc) 3.96 KB	Raw Open Link in New Tab Dean, Link in New Window Open Link in New Window
QMRF identifier (JRC Inventory): To be entered by JRC QMRF Title: ABCDEFGHIJ Printing Date:XX-XX-20XX	Send link to Xisomi Phone Save Link As
Palt text	Copy Link Address Z Save to Zotero
1.QSAR identifier	Inspect

The user only needs to provide the necessary information under each section and the QMRF report is generated in an easy-to-read format.

There are online tools to edit MarkDown documents, such as: <u>https://dillinger.io/</u>

CCUMENT NAME .QSAR identifier.md			READING TIME: 2 MIN READ WORDS: 3 CHARACTERS: 32
ARKDOWN	w ^R	PREVIEW	<
1 QMRF identifier (JRC Inventory): To be entered by JRC			
2 QMRF Title: ABCDEFGHIJ		QMRF identifier (JRC Inventory): To be entered by JRC	
3 Printing Date:XX-XX-20XX		QMRF Title: ABCDEFGHIJ	
<pre>4 ![alt text](<u>https://qsardb.jrc.ec.europa.eu/qmrf/images/logo_menu.png</u> "QMRF")</pre>		Printing Date:XX-XX-20XX	
5 *		Ralt text	
6 - # 1.QSAR identifier		-	
7 * ## 1.1.QSAR identifier (title):			
8 * ***		1.QSAR identifier	
9			
10 - ```		1.1.QSAR identifier (title):	
11 - ## 1.2.0ther related models:			
12 - ```			
13			

Alternatively, users can use a specialised QMRF editor for Windows at http://qmrf.sourceforge.net/.

4. Test and use your Jaqpot model

You can validate or use your model or any other Jaqpot model, by following the steps in the following tutorial:

Jaqpot 5: How to access and use an existing predictive model

The data should be provided in a csv file. A sample csv file can be downloaded by clicking on the following link:

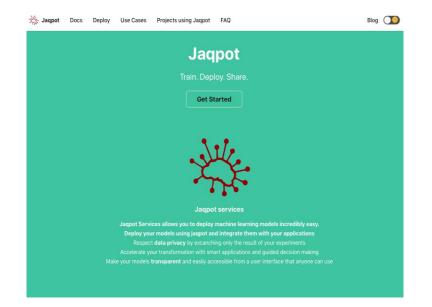
https://zenodo.org/record/4671069/files/70_model_reduced.csv?download=1

5. More information on the Jaqpot platform

Jaqpot offers a suite of tools for QSAR, PBPK, image analysis, read-across methods:



A comprehensive technical documentation is provided at <u>https://www.jaqpot.org</u>.



A collection of tutorials describing other Jaqpot functionalities is available on Zenodo:

- 1. Jaqpot 5 User accounts
- 2. Jaqpot 5: How to manage and use organisations
- 3. Jaqpot 5: How to deploy a predictive model using the jaqpotpy library
- 4. Jaqpot 5: How to simulate biodistribution scenarios using custom PBPK models