



# Jaqpot5 tutorials

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

USE:	How to use an existing predictive model
VERSION:	V.2.0 April 2021
CONTACT DETAILS:	Haralambos Sarimveis: <a href="mailto:hsarimv@central.ntua.gr">hsarimv@central.ntua.gr</a> Pantelis Karatzas: <a href="mailto:pantelispanka@gmail.com">pantelispanka@gmail.com</a> Philip Doganis: <a href="mailto:filipposd@gmail.com">filipposd@gmail.com</a> Jason Sotiropoulos: <a href="mailto:jasonsoti1@gmail.com">jasonsoti1@gmail.com</a>

# INTRODUCTION

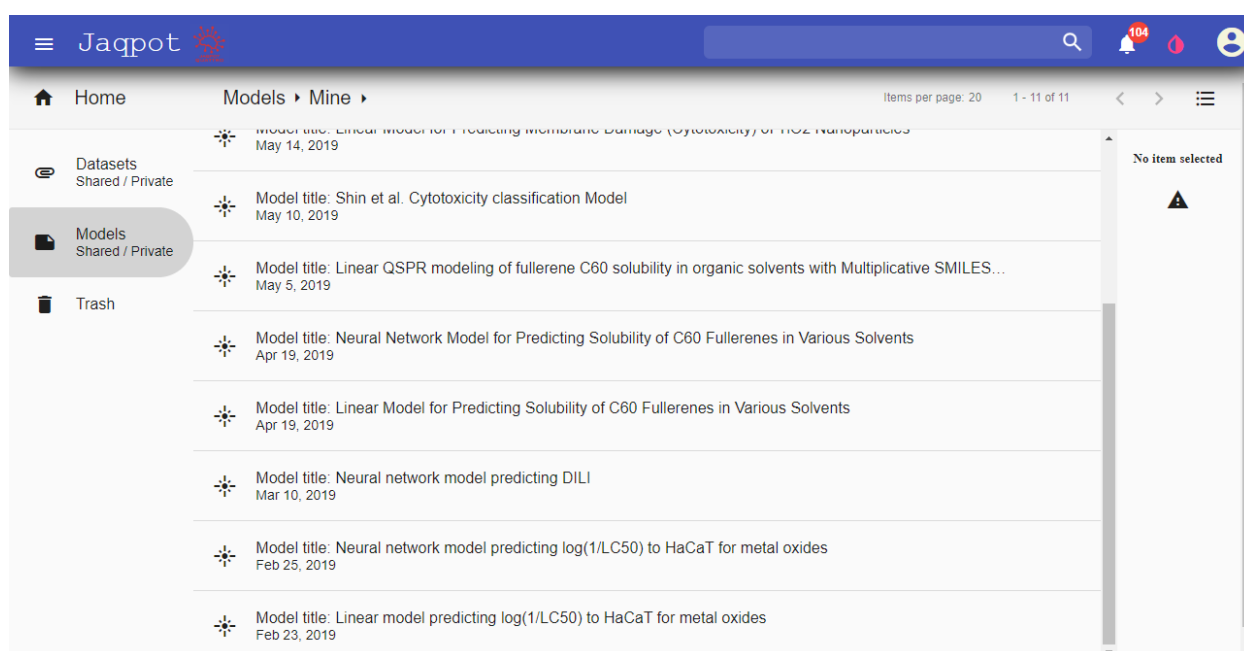
Jaqpote 5 is a user-friendly web-based e-infrastructure that allows model developers to deploy their predictive models and share them through the web. The Jaqpote 5 GUI directs the model developers to further document their models in a way that can be easily understood and used by end-users with little or no experience on machine learning and statistical analysis. The GUI also allows the end-users to apply the models on their own data for validation and/or prediction purposes and the results are collected and visualised in automatically generated tables, graphs and reports. All major machine learning and statistical data-driven algorithms are supported in Jaqpote 5, by integrating popular libraries such as the Python Scikit-learn and the R Caret libraries. Jaqpote 5 has been designed as a generic modelling and machine learning web platform, but particular emphasis is given on serving the needs of the chemo/bio/nano/pharma communities by integrating QSAR, biokinetics, dose-response and read-across models. Jaqpote 5 has been developed by the [Unit of Process Control and Informatics](#) in the School of Chemical Engineering at the National Technical University of Athens.

This document provides a tutorial on accessing and using an existing predictive model in Jaqpote 5. The resource has been made available at <https://app.jaqpote.org/>. Detailed documentation is available at: <https://www.jaqpote.org/>

# ACCESSING AND USING A PREDICTIVE MODEL

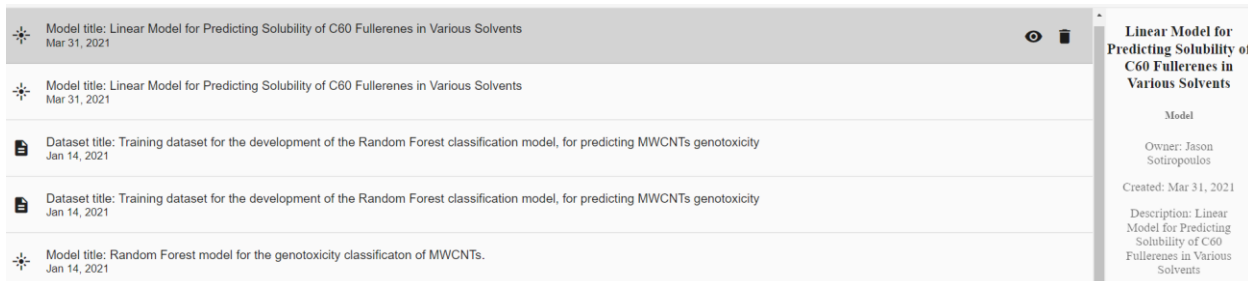
The user has access to a Jaqpot 5 predictive model, if he is the creator of the model or if the model is shared with him through an organisation. For more information, please visit tutorials on creating a model and sharing models through organisations.

When the user enters Jaqpot 5, a list of all resources (Models, Datasets) created by the user is displayed. By left-clicking on the Models tab, only the models are displayed (please see Figure 1).



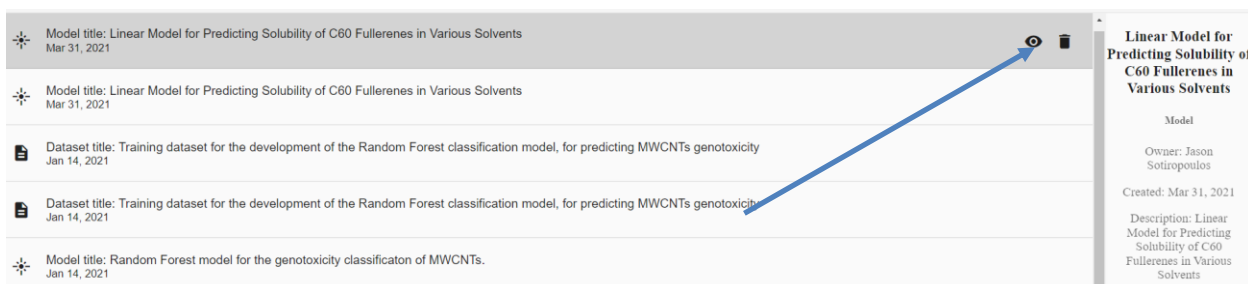
**Figure 1.** List of resources created by the user.

The user can select a model by placing the cursor on the model. (Figure 2). Automatically the colour of the model is becoming darker and more information about the model are shown on the right part of the screen (model creator, date title and description).



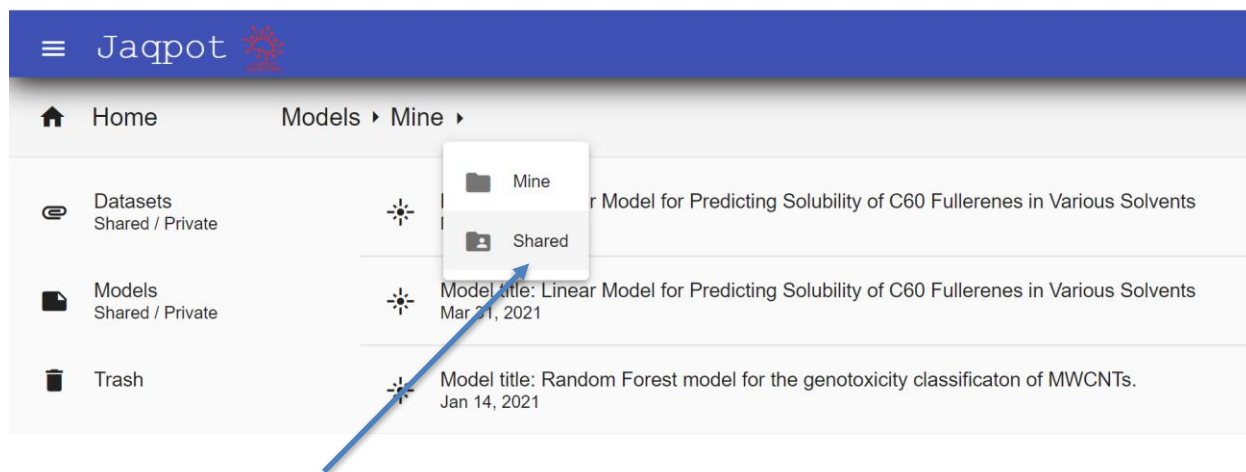
**Figure 2.** Model selection.

The user can access the model by clicking on the “eye” icon (Figure 3)



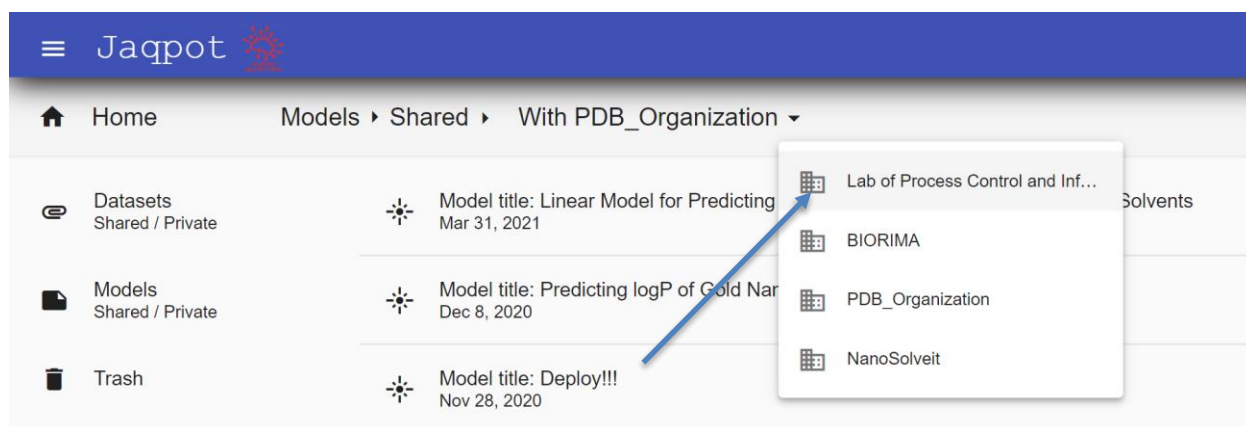
**Figure 3.** Opening a model

To access a model created by other users through organisations, the user should click on the right arrow after "Mine" on the top of the screen and select "Shared" from the dialogue box that appears (Figure 4).



**Figure 4.** Accessing models shared by other users through organisations

A new dialogue box appears listing all the organisations where the user is a member. To become a member of the organisation, the user should send an e-mail to the administrator of the organisation. The user clicks on the organisation of his interest, in the example here the users selects the Lab of Process Control and Informatics organisation (Figure 5)



**Figure 5.** Selecting an organisation

A list of all models shared through the organisation appears and the user proceeds exactly as shown in Figure 3 to open a model (Figure 6)

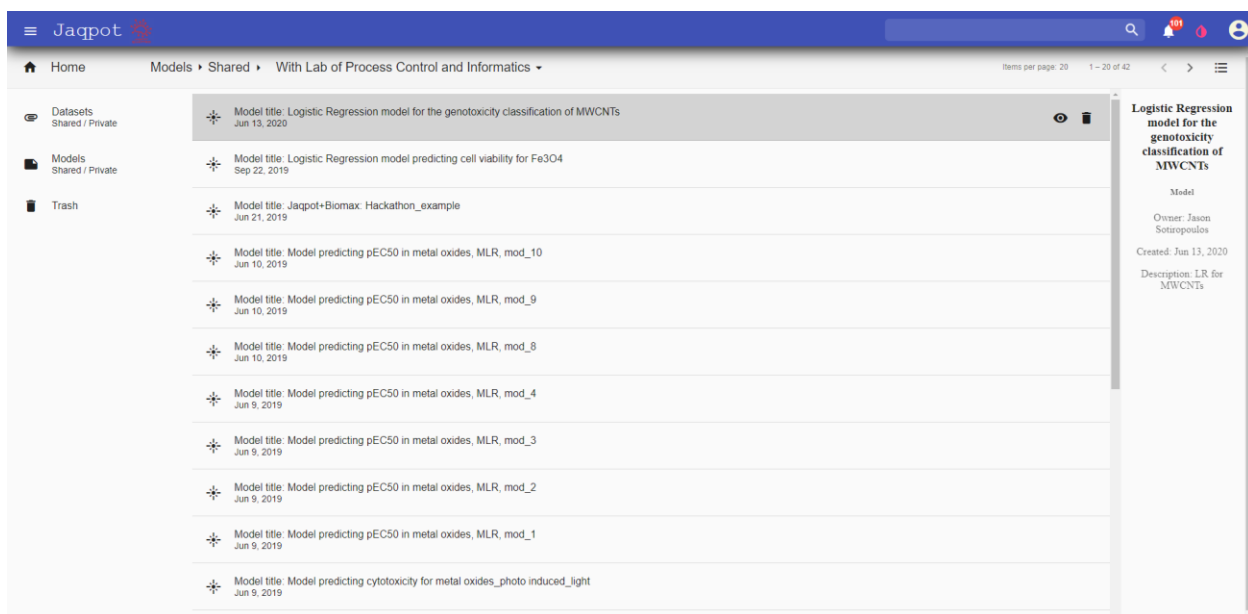
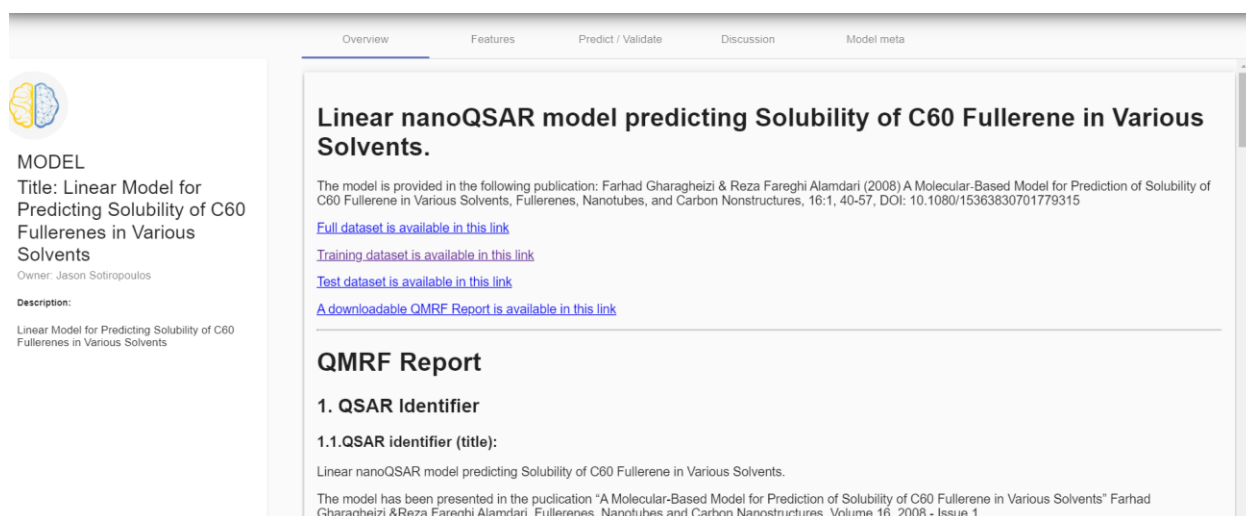


Figure 6. Models shared through an organisation

When the user opens a model, five tabs become available. As an example we will use the implementation of the linear model predicting Solubility of C60 Fullerene in Various Solvents, which is hosted in Jaqpot and is shared with the members of the NanoCommons organisation under the following URL: <https://app.jaqpot.org/model/RqCRtRpY85kpbGtsiXp>. The model has been published in the following paper: Farhad Gharagheizi & Reza Fareghi Alamdari (2008) A Molecular-Based Model for Prediction of Solubility of C60 Fullerene in Various Solvents, 16:1, 40-57, DOI: 10.1080/15363830701779315. The first tab available to the user is the “Overview” tab. Here, various information about the model can be shared by the model creator, for example information about the publication, Jaqpot 5 links to the datasets that were used to train and test the model, a full QSAR Model Reporting Format (QMRF) report and the PMML representation of the model (Figure 7).



The screenshot shows the 'Overview' tab of a model page. At the top, there is a navigation bar with five tabs: 'Overview' (selected), 'Features', 'Predict / Validate', 'Discussion', and 'Model meta'. On the left side, there is a sidebar with a logo of a brain and the text 'MODEL'. Below the logo, the title is 'Linear Model for Predicting Solubility of C60 Fullerenes in Various Solvents', the owner is 'Jason Sotiropoulos', and the description is 'Linear Model for Predicting Solubility of C60 Fullerenes in Various Solvents'. The main content area has the title 'Linear nanoQSAR model predicting Solubility of C60 Fullerene in Various Solvents.' followed by a paragraph describing the model's origin and publication. Below this, there are four blue links: 'Full dataset is available in this link', 'Training dataset is available in this link', 'Test dataset is available in this link', and 'A downloadable QMRF Report is available in this link'. The 'QMRF Report' section is titled '1. QSAR Identifier' and contains a sub-section '1.1. QSAR identifier (title):' with the text 'Linear nanoQSAR model predicting Solubility of C60 Fullerene in Various Solvents.' and a reference to the publication.

**Figure 7.** The “Overview” tab

In the “Features” tab the user can see information about the independent and the dependent variables of the model like descriptions, units and ontological classes where they belong (Figure 8).

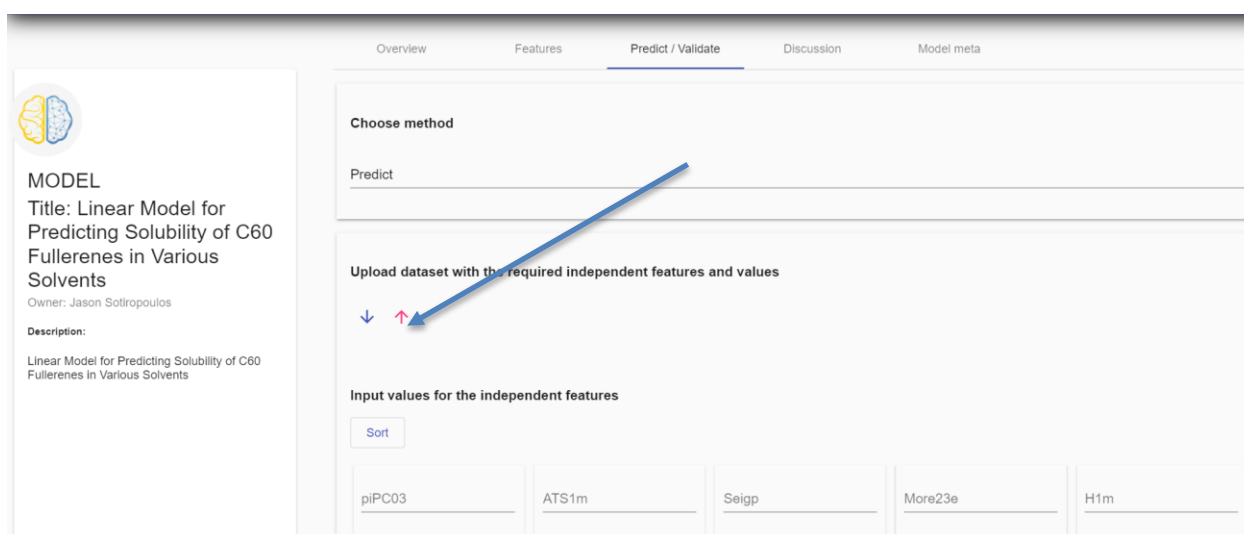
The screenshot shows the Jaqpot web application interface. At the top, there is a navigation bar with the Jaqpot logo and a search icon. Below the navigation bar, there are tabs for 'Overview', 'Features', 'Predict / Validate', 'Discussion', and 'Model meta'. The 'Features' tab is currently selected. On the left side, there is a sidebar with a brain icon and the following text: 'MODEL Title: Linear Model for Predicting Solubility of C60 Fullerenes in Various Solvents', 'Owner: Jason Sotiropoulos', and 'Description: Linear Model for Predicting Solubility of C60 Fullerenes in Various Solvents'. The main content area is divided into two sections: 'Dependent feature / Predicted feature' and 'Independent features'. The 'Dependent feature / Predicted feature' section contains a box with the text 'logS Exp.'. The 'Independent features' section contains several boxes, each with a feature name and a description: 'piPC03' (Description: Molecular multiple path count of order 03), 'ATS1m' (Description: Broto-Mireau autocorrelation of a topological structure-lag 1/weighted by atomic masses), 'More23e' (Description: Eigenvalue sum from polarizability weighted distance matrix), 'Seigp' (Description: 3D-MORSE-signal 23/weighted by atomic sanderson electronegativities), and 'H1m...' (Description: ...).

Figure 8. The “Features” tab



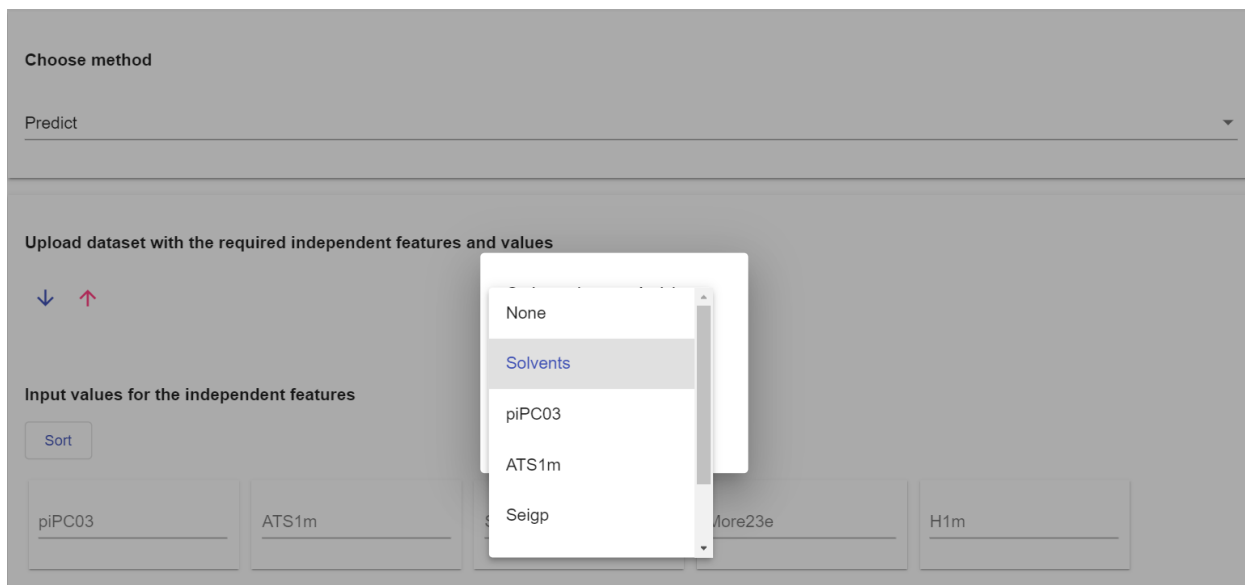
The “Predict/Validate” tab contains the main functionalities of the model (Figure 9). The user who has access to the model can either generate predictions for NMs with unknown end-point values or test the model with a data set containing end-point values.

The “Predict” option is used when the user wants to obtain predictions of new instances, where the end-point (dependent) value is unknown. Values for the independent variables can be entered by hand (for relatively small datasets). Alternatively, the user can upload data through a csv template which is automatically generated (by clicking on the blue arrow). The template contains all input variable names, so the user just enters the values in each column and uploads the data by clicking the red arrow. The procedure of uploading data through a csv is exactly the same with the validation option and is explained in more details later in the tutorial.



**Figure 9.** The “Predict/Validate” tab – Predict Option

After uploading the dataset a list of the names of all variables appears. We can select one column to serve as the substance ID (Figure 10).



**Figure 10.** List of variable names contained in the csv file

A preview of the dataset appears and the user is prompted to start the prediction procedure (Figure 11). If the model has been deployed having the DOA option enabled, he can select whether the leverage values will be calculated or not.

Upload dataset with the required independent features and values

↓ ↑

Dataset formed

Id	piPC03	ATS1m	Seigp	More23e	H1m
pentane	1.099	1.609	0	-0.922	0.148
hexane	1.386	1.792	0	-1.156	0.173
octane	1.792	2.079	0	-1.656	0.208
iso-octane	1.792	2.079	0	-1.529	0.242
decane	2.079	2.303	0	-2.151	0.233
dodecane	2.303	2.485	0	-2.598	0.249
tetradecane	2.485	2.639	0	-3.047	0.263

Erase dataset

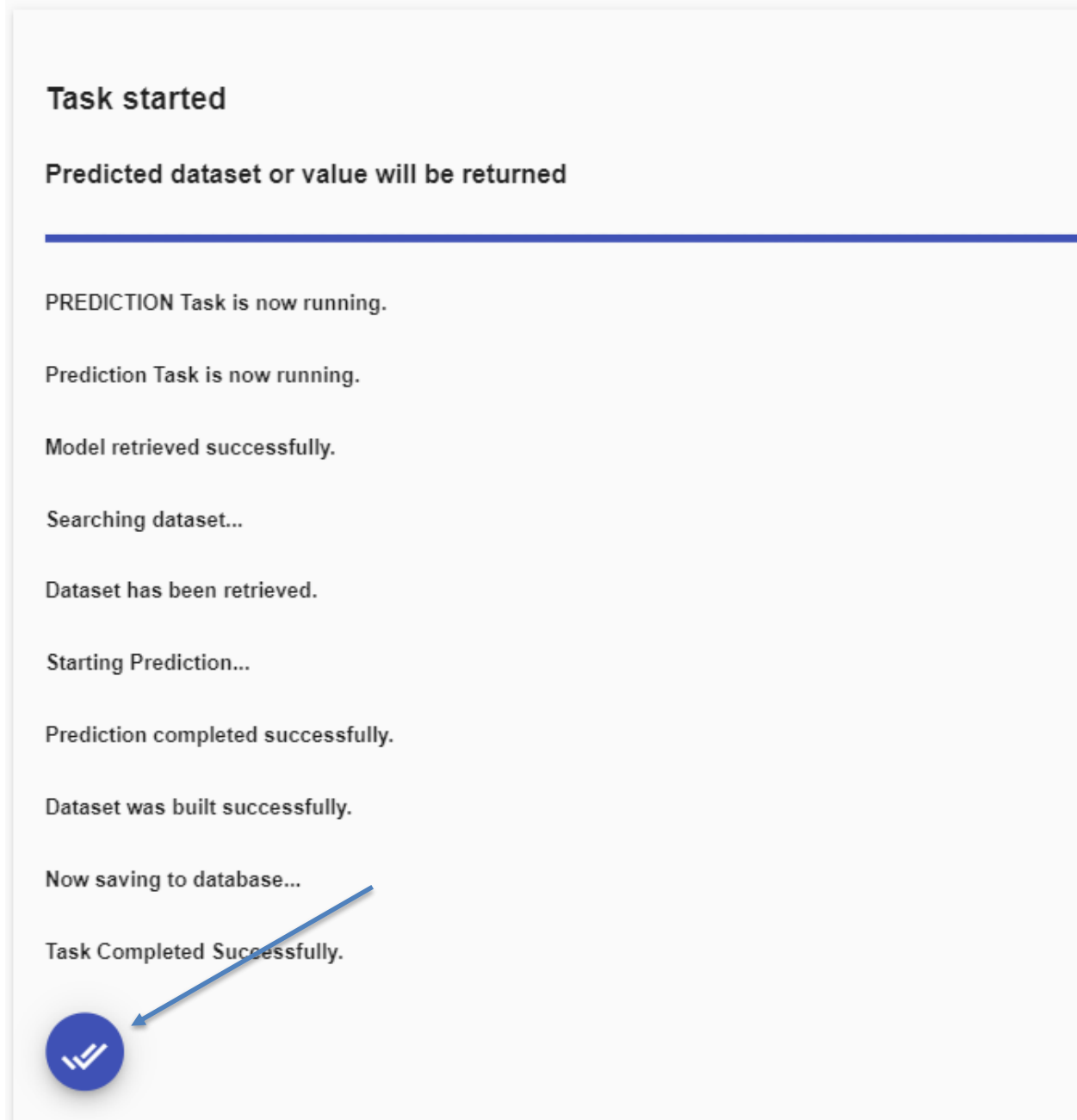
Start procedure

DOA

Start Procedure

Figure 11. Starting the prediction procedure

The user is informed about the progress of the task. When the task is completed, a double-check icon appears in the bottom of the screen. By clicking on this icon the user can view the prediction results (Figure 12).



**Figure 12.** Starting the prediction procedure

The end-point predictions and the leverage (DOA) values are shown along with all the independent variables. The definition of the DOA threshold value appears above the table. A leverage value below the DOA threshold value indicates that the query substance lies within the DOA of the model, otherwise the prediction is not reliable. The user can download these results by clicking on the “Download” button. (Figure 13), By clicking on the “View predicted values only”, only the end-point predictions are displayed (Figure 14).

The predicted DOA value should be less than 0.14516129

View predicted value only

Id	ATS1m	H1m	More23e	DOA	piPC03	Seigp	logS Exp.
pentane	1.609	0.148	-0.922	0.018870271764833198	1.099	0	-5.818155433224528
hexane	1.792	0.173	-1.156	0.027521590734008394	1.386	0	-5.419508786237124
octane	2.079	0.208	-1.656	0.0563441835378012	1.792	0	-4.8626710175980605
iso-octane	2.079	0.242	-1.529	0.04644666032482139	1.792	0	-4.839393874175334
decane	2.303	0.233	-2.151	0.09920129317714098	2.079	0	-4.479601112133886
dodecane	2.485	0.249	-2.598	0.15029120495907358	2.303	0	-4.186383981558108
tetradecane	2.639	0.263	-3.047	0.214362838913921	2.485	0	-3.966770281802667

Items per page: 30 1 – 15 of 15

Download

**Figure 13.** Viewing the complete dataset (independent variables and end-point predictions)

The predicted DOA value should be less than 0.14516129

View all values

Id	DOA	logS Exp.
pentane	0.018870271764833198	-5.818155433224528
hexane	0.027521590734008394	-5.419508786237124
octane	0.0563441835378012	-4.8626710175980605
iso-octane	0.04644666032482139	-4.839393874175334
decane	0.09920129317714098	-4.479601112133886
dodecane	0.15029120495907358	-4.186383981558108
tetradecane	0.214362838913921	-3.966770281802667

Items per page: 30 1 - 15 of 15 < >

Download

Figure 14. Viewing end-point predictions only

The “Validate” option is used when the user wants to test the models with observation containing the end-point (dependent) values. Here only the option of uploaded data through the csv template is available (Figure 15).

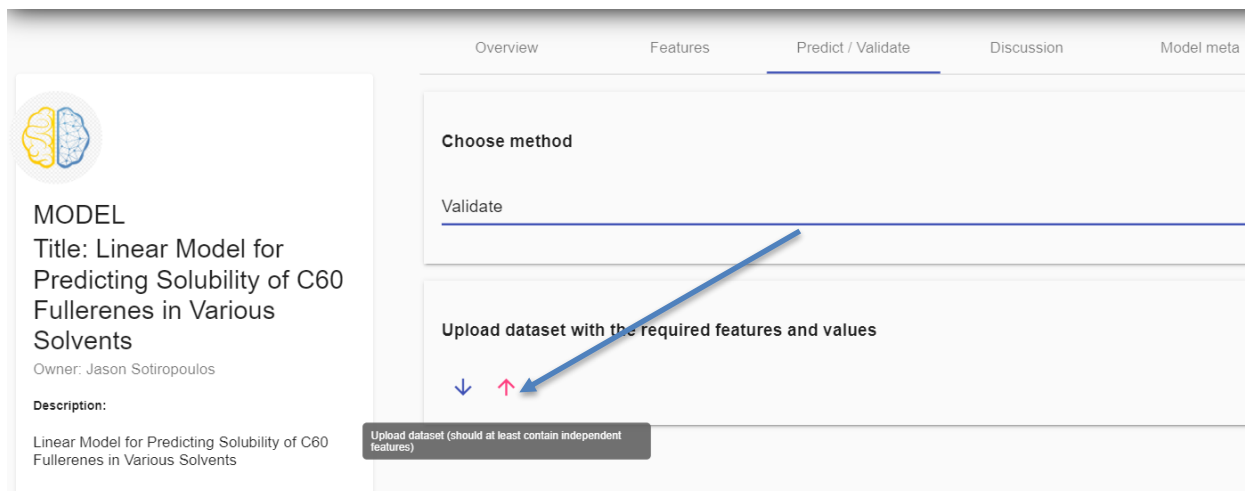


Figure 15. The “Predict/Validate” tab – Validate Option

After uploading the data a list of the names of all variables appears. We can select one column to serve as the substance ID (Figure 16).

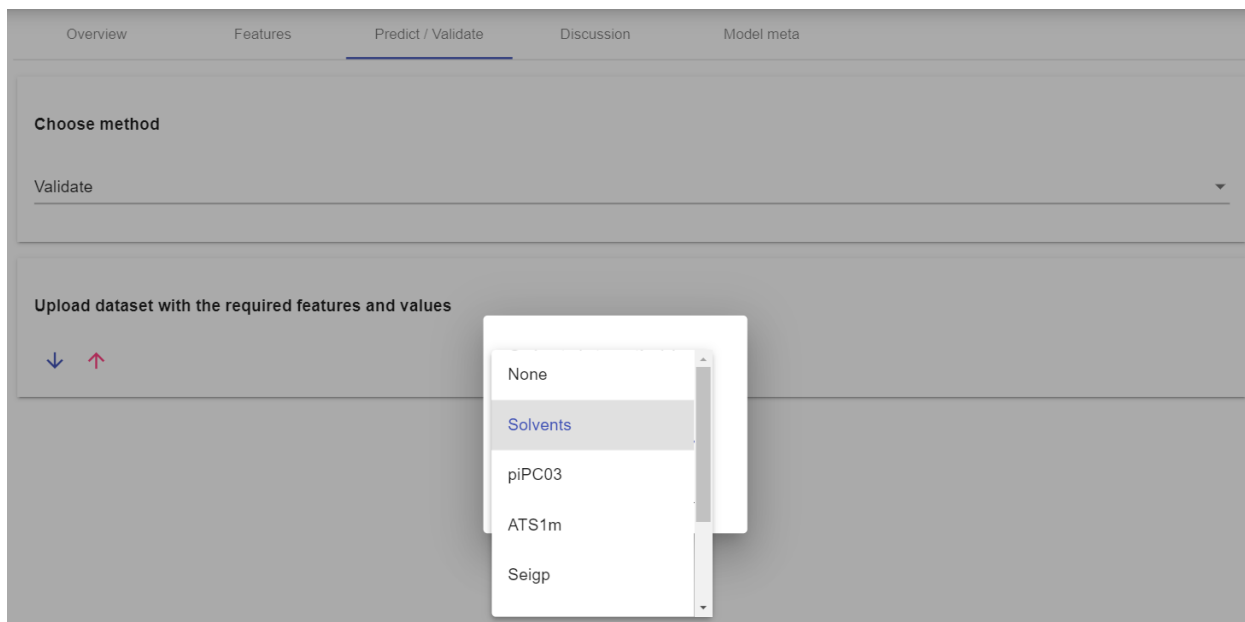
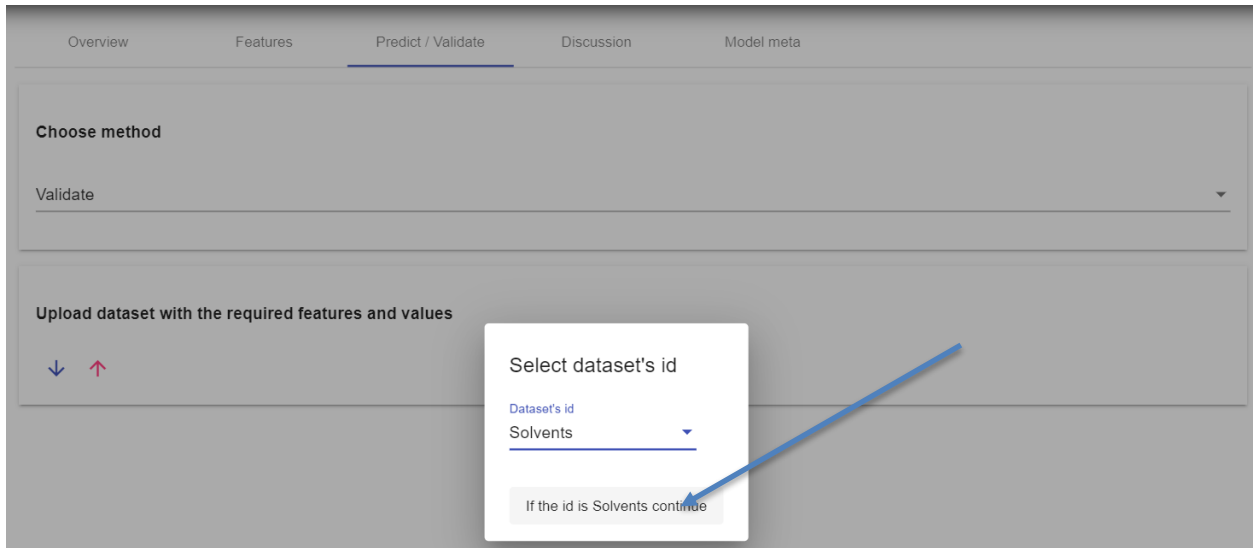


Figure 16. List of variable names contained in the csv file

Here the user selects the “Solvents” column and confirms his selection by clicking on the button that prompts him to continue (Figure 17):



**Figure 17.** Selecting the “id” column



A preview of the dataset appears and the user is prompted to start the validation procedure by selecting the type of validation method (Regression or Classification) and clicking on the “Start Procedure” button (Figure 18).

Upload dataset with the required features and values

↓ ↑

Dataset formed

Id	piPC03	ATS1m	Seigp	More23e	H1m	logS Exp.
pentane	1.099	1.609	0	-0.922	0.148	-6.1
hexane	1.386	1.792	0	-1.156	0.173	-5.1
octane	1.792	2.079	0	-1.656	0.208	-5.2
iso-octane	1.792	2.079	0	-1.529	0.242	-5.2
decane	2.079	2.303	0	-2.151	0.233	-4.7
dodecane	2.303	2.485	0	-2.598	0.249	-3.5
tetradecane	2.485	2.639	0	-3.047	0.263	-4.3

Choose validation method

Regression / Classification

Start procedure Erase dataset

Start Procedure

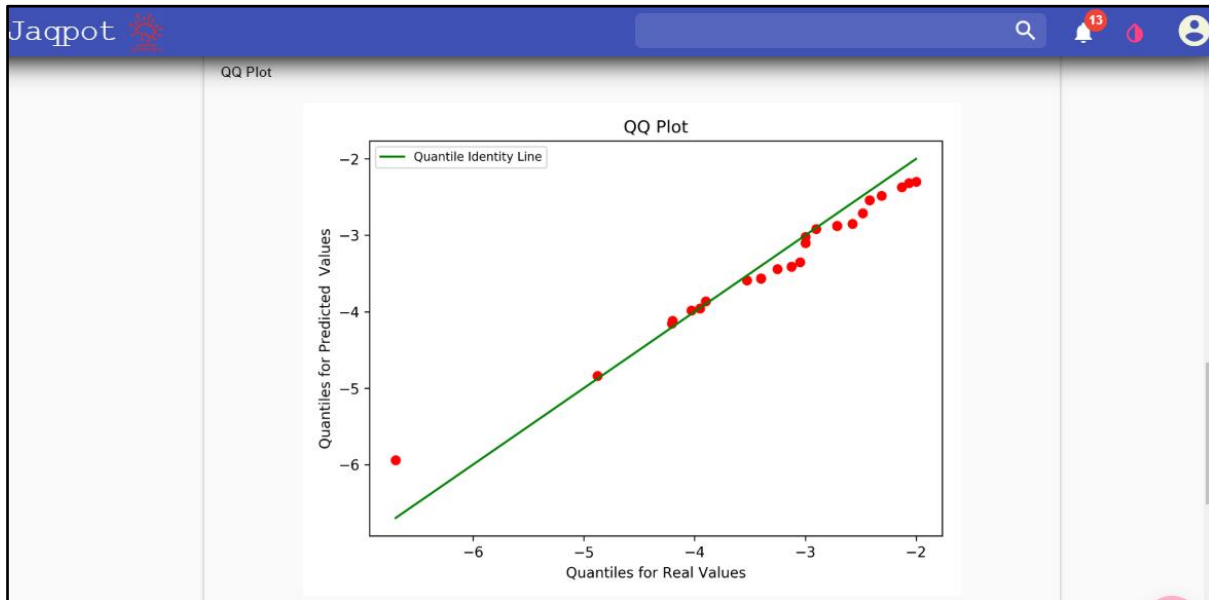
Figure 18. Starting the validation procedure.

The user is informed about the process of the task and waits until the task has been completed successfully (Figure 19). Then he can click on the “Check button” button to view the results.

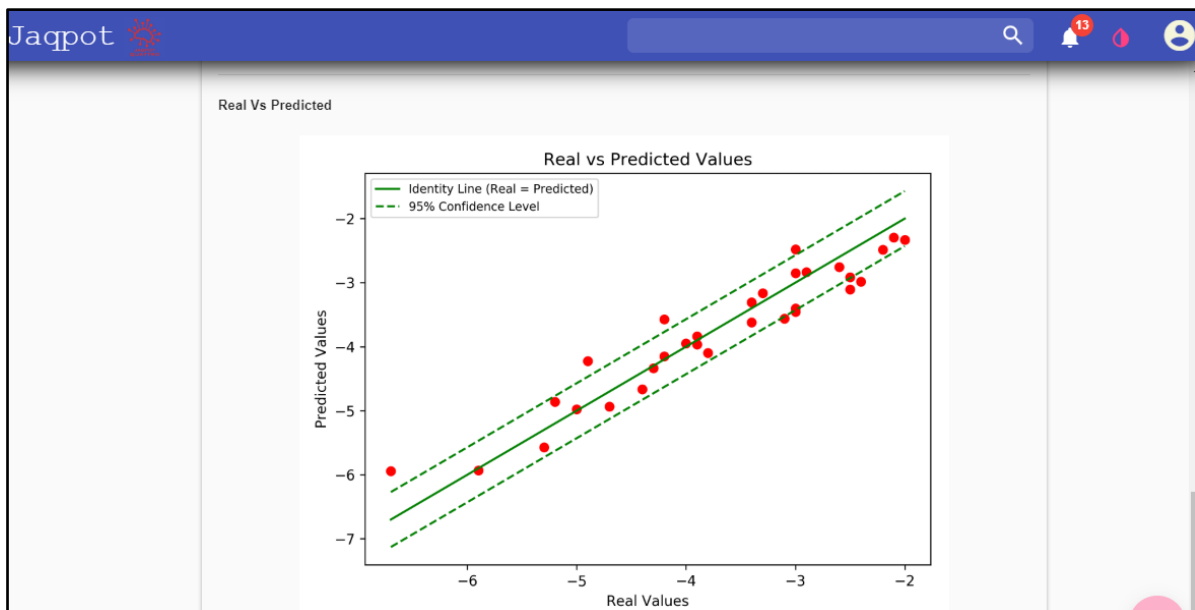


**Figure 19.** Task progress.

A report is automatically generated with validation statistics, accompanied with a QQ plot and *Real vs Predicted* values plot, giving insight to the effectiveness of the model (Figures 20-21)

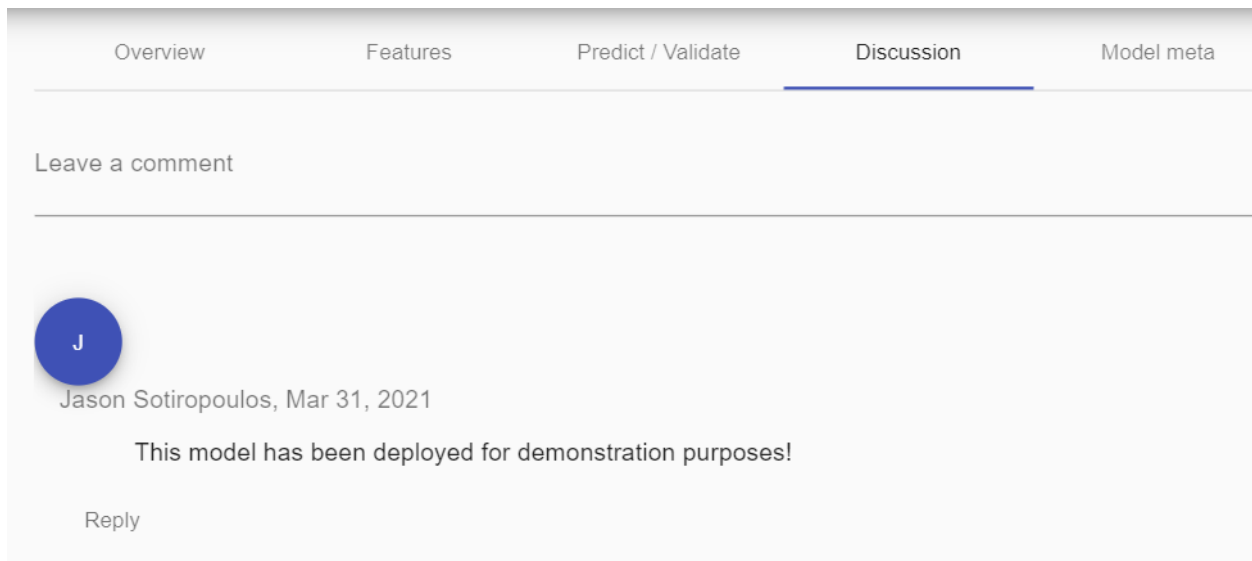


**Figure 20.** Validation results - QQ plot .



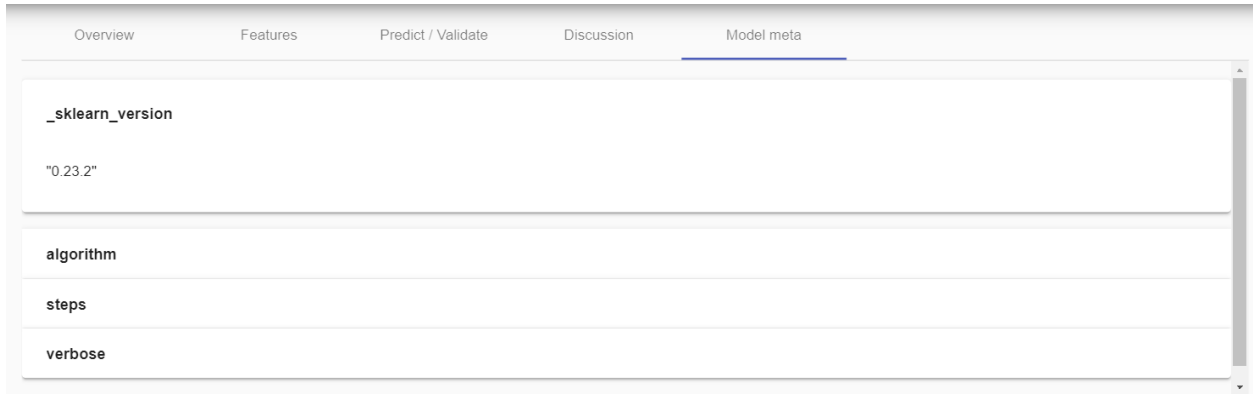
**Figure 21.** Validation results - *Real vs Predicted* plot .

In the discussion tab any user of the model can leave a comment (Figure 22):



**Figure 22.** The “Discussion” tab.

The “Model meta” tab contains meta information about the model, for instance the Sci-Kit Learn’s version, the model’s algorithm the model’s hyperparameter values and in case of a pipeline an analytical representation of each step (Figure 23).



**Figure 23.** The “Model meta” tab.

**Support**

**OpenRiskNet**  
RISK ASSESSMENT E-INFRASTRUCTURE



**NanoCommons**

Nano-Knowledge Community