Jaqpot5 tutorials

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

USE:	How to use an existing predictive model
VERSION:	V.1.0
CONTACT DETAILS:	Haralambos Sarimveis: hsarimv@central.ntua.gr Pantelis Karatzas: pantelispanka@gmail.com Philip Doganis: filipposd@gmail.com

INTRODUCTION

Jaqpot 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 Jaqpot 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 Jaqpot 5, by integrating popular libraries such as the Python Scikit-learn and the R Caret libraries. Jaqpot 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. Jaqpot 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 Jaqpot5. The resource has been made available at https://app.jaqpot.org/.

ACESSING 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 right-clicking on the models tab, the tab changes its colour to darker grey and only the models are displayed (please see Figure 1).

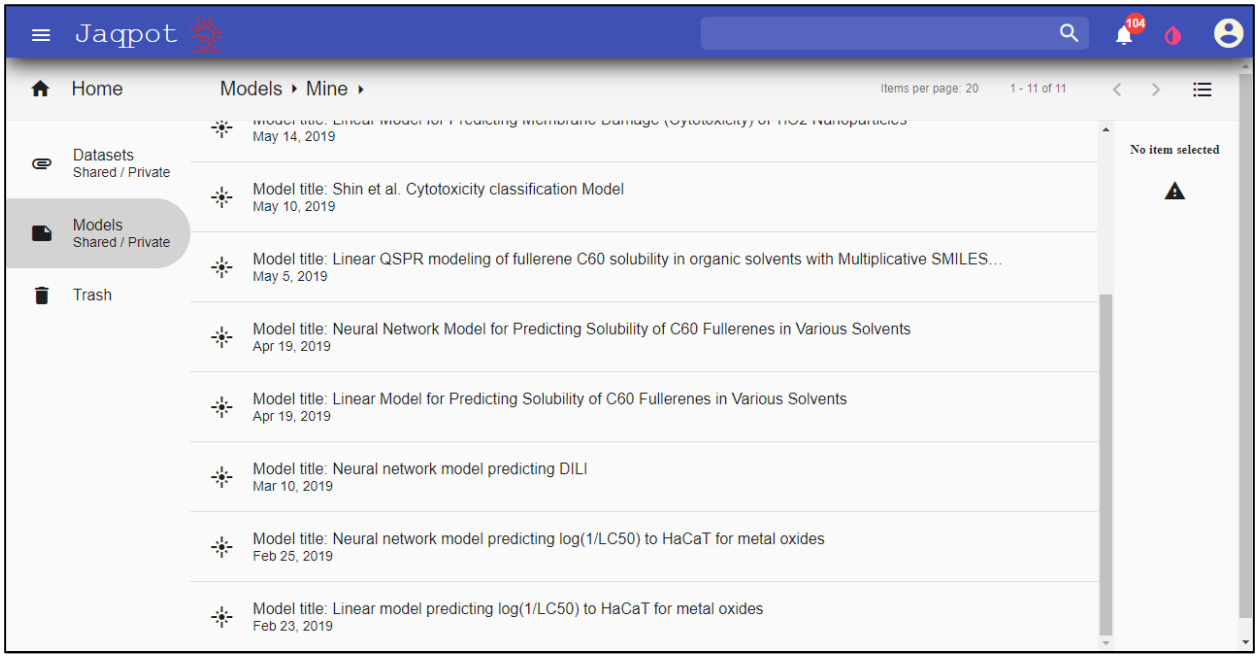

Figure 1. List of resources created by the user.

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, description and organisations with the model is shared).

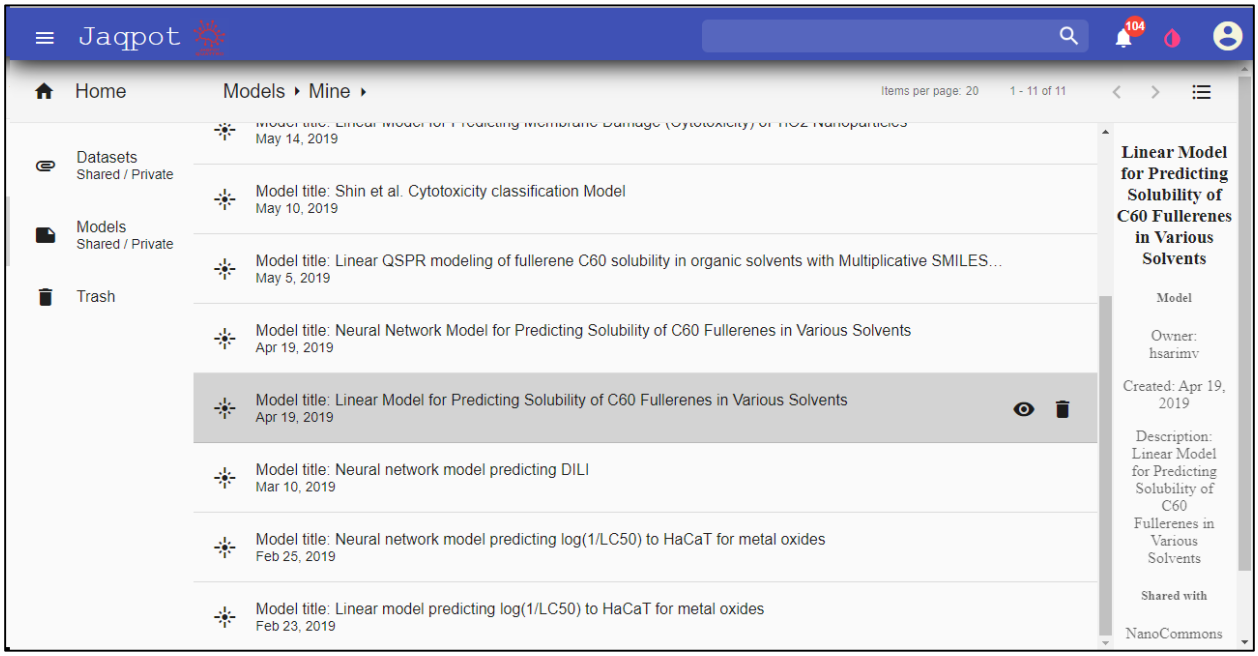

Figure 2. Model selection.

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

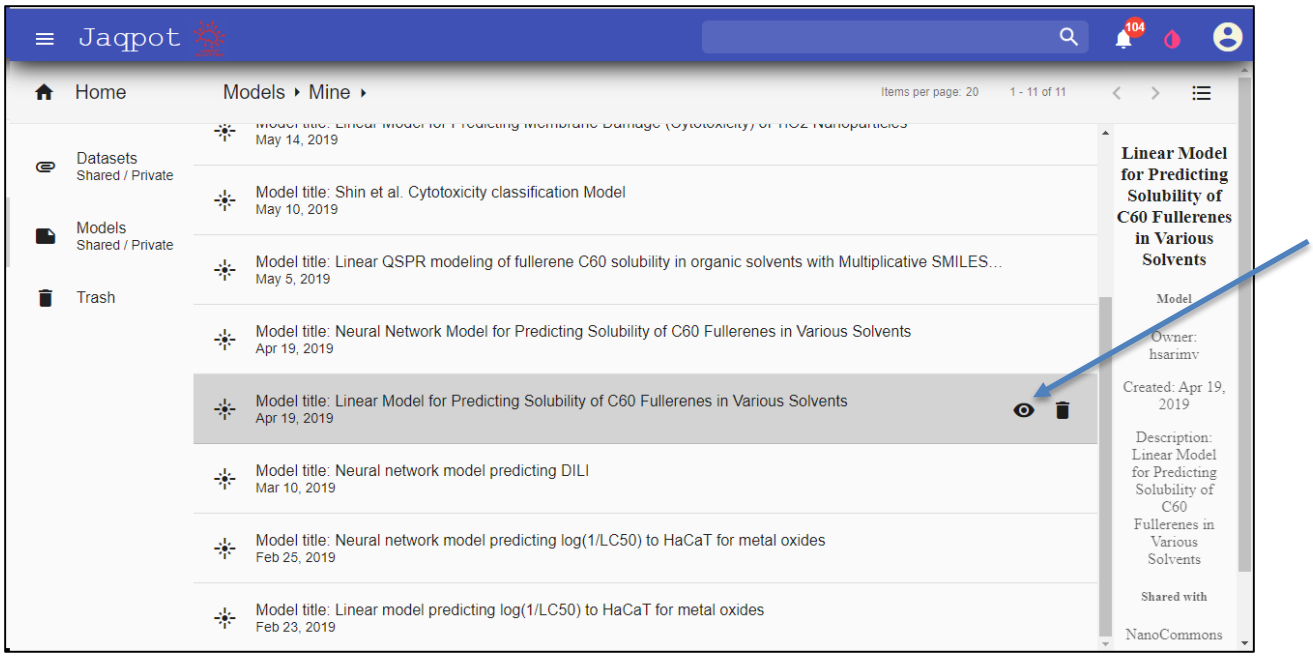

Figure 3. Opening a model

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).

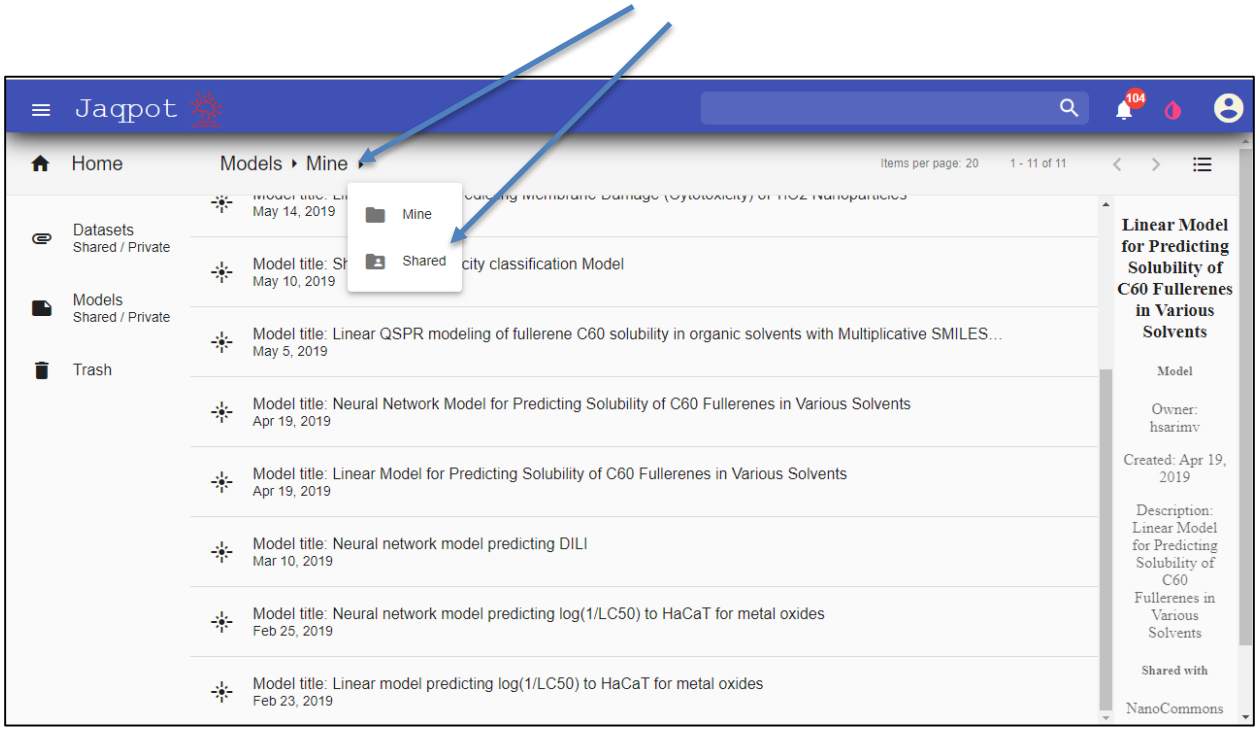

Figure 4. Accessing models shared by other users through organisations

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. Other options will be available shortly. The user clicks on the organisation of his interest, in the example here the users selects the NanoCommons organisation (Figure 5)

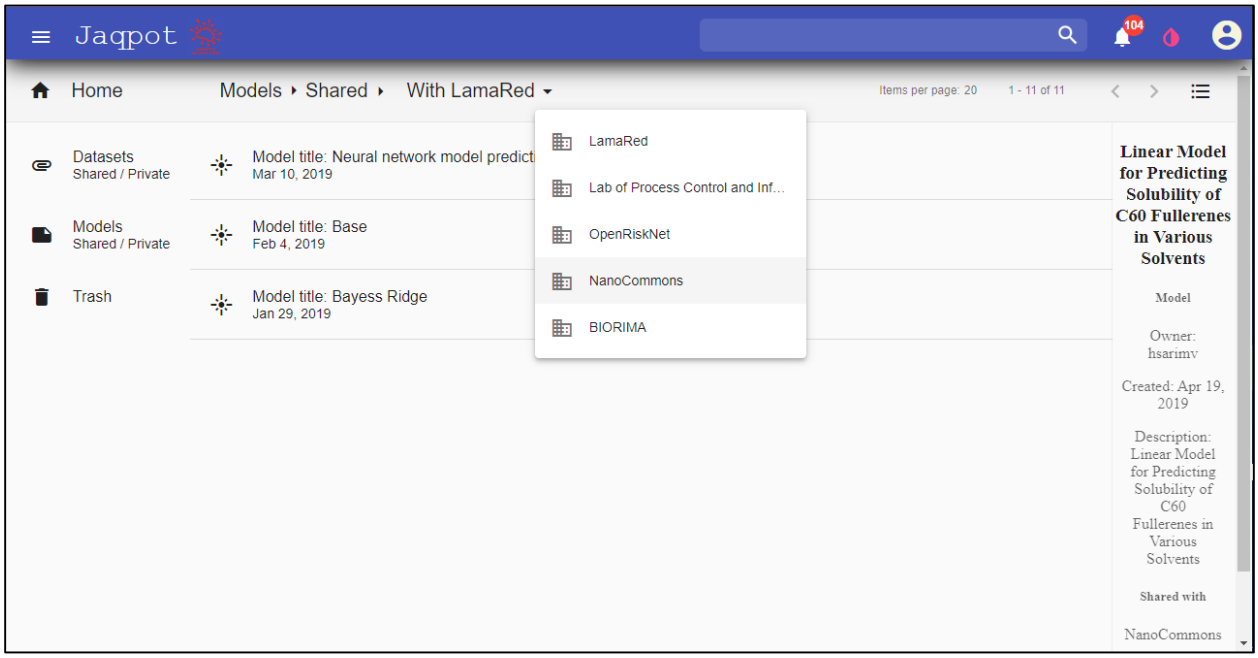

Figure 5. Selecting an organisation

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

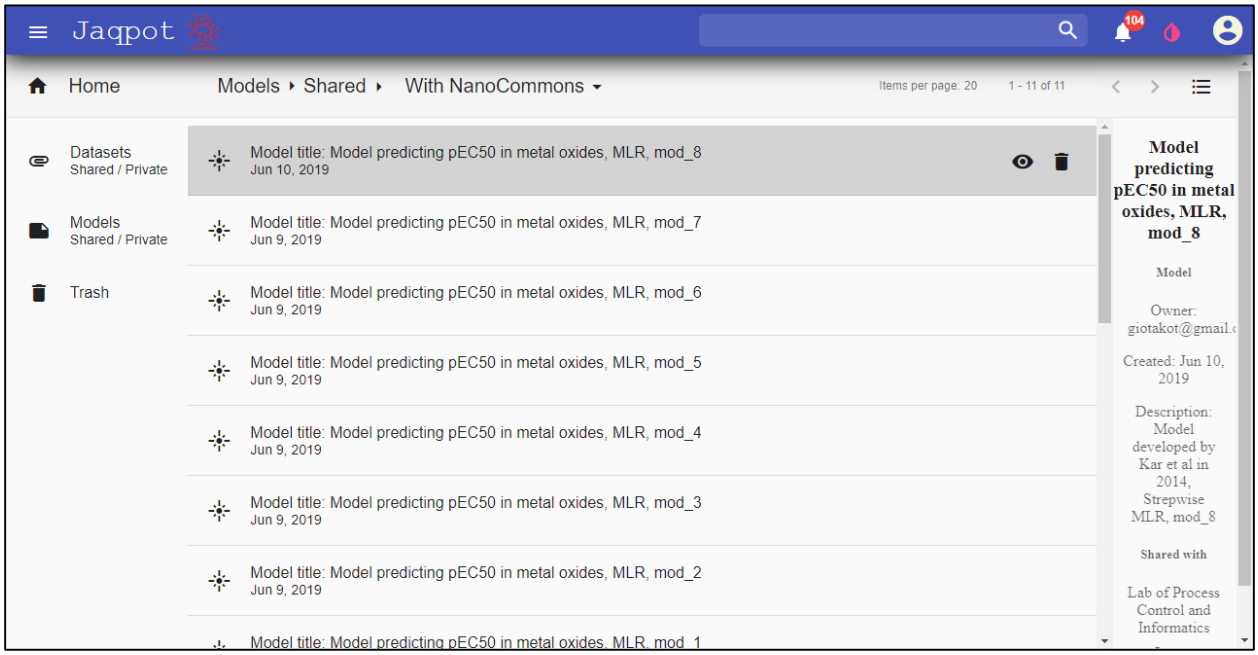

Figure 6. Models shared through an organisation

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 shared through the NanoCommons organisation. 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, Fullerenes, Nanotubes, and Carbon Nonstructures, 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).

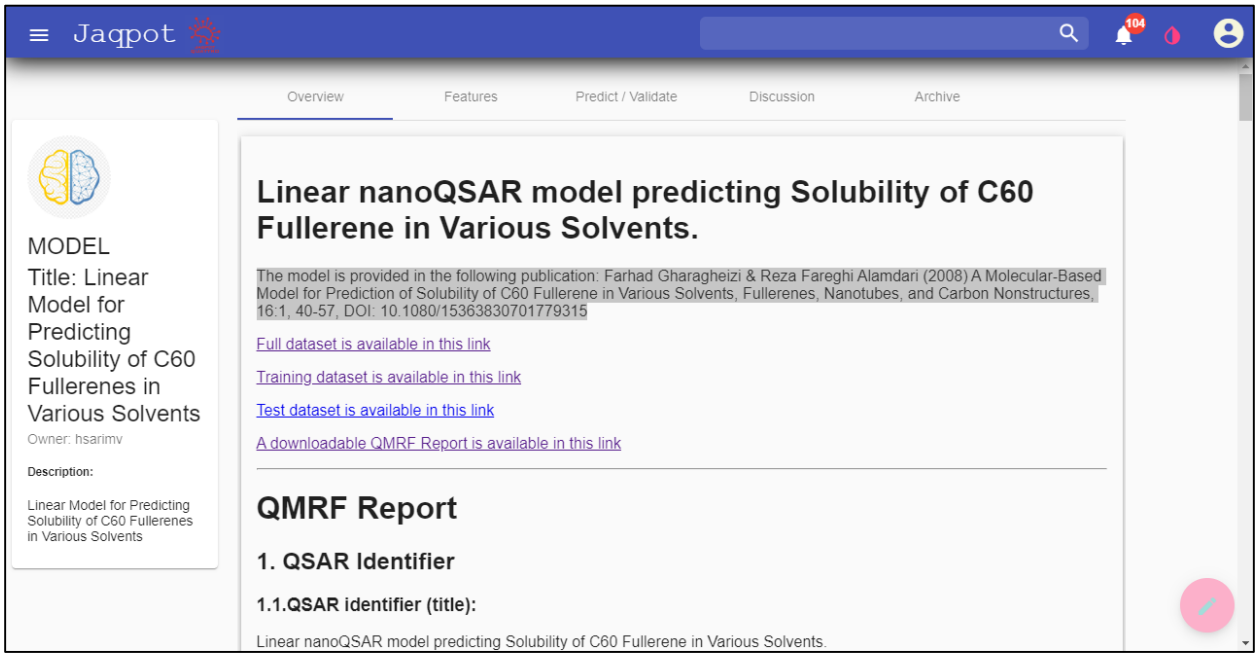

Figure 7. The "Overview" tab

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).

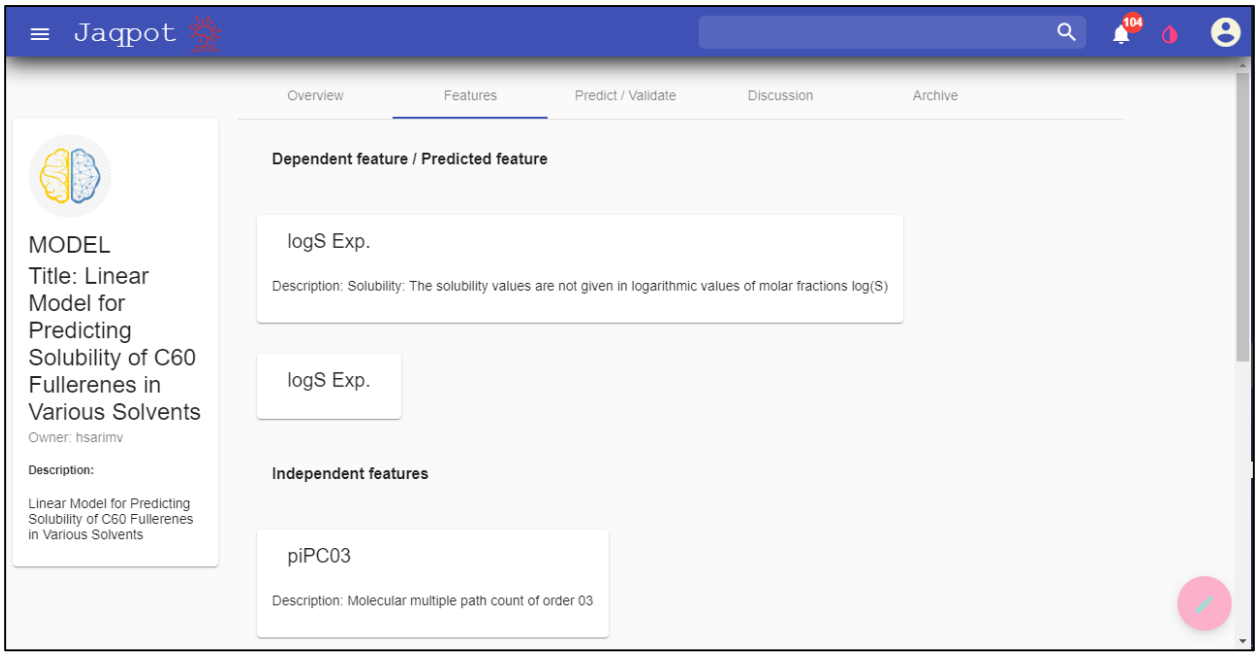

Figure 8. The “Features” tab

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 down-pointing 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 upwards-pointing 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.

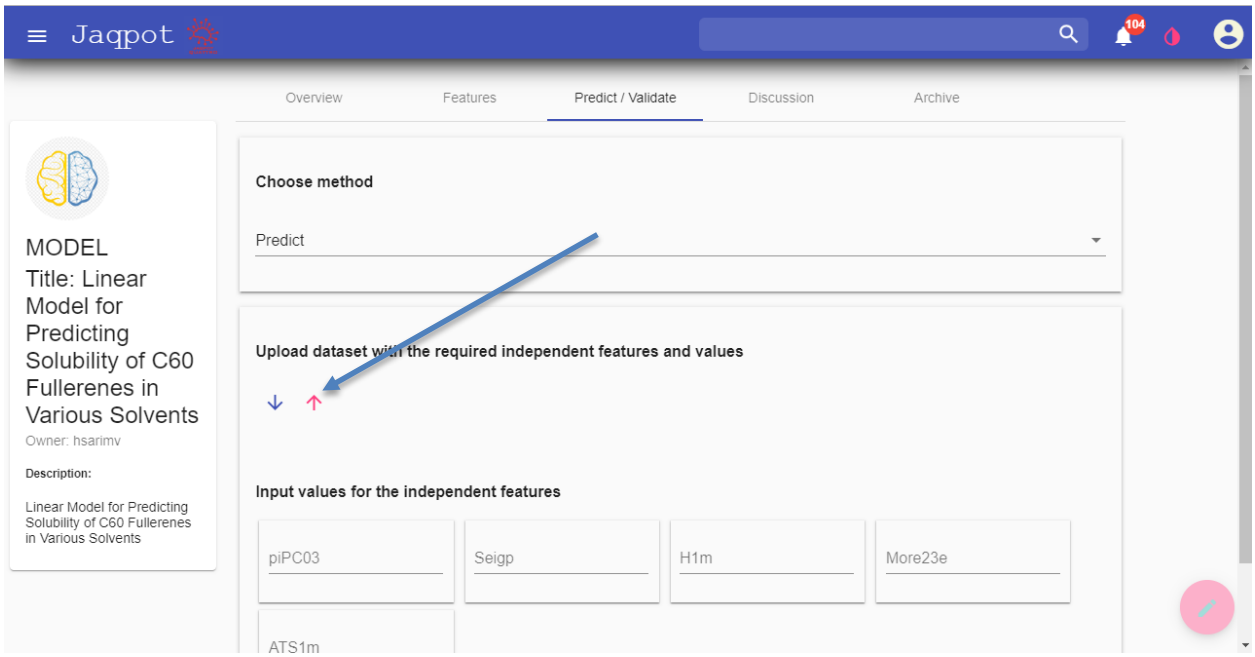

Figure 9. The “Predict/Validate” tab – Predict Option

A preview of the dataset appears and the user is prompted to start the validation procedure (Figure 10)

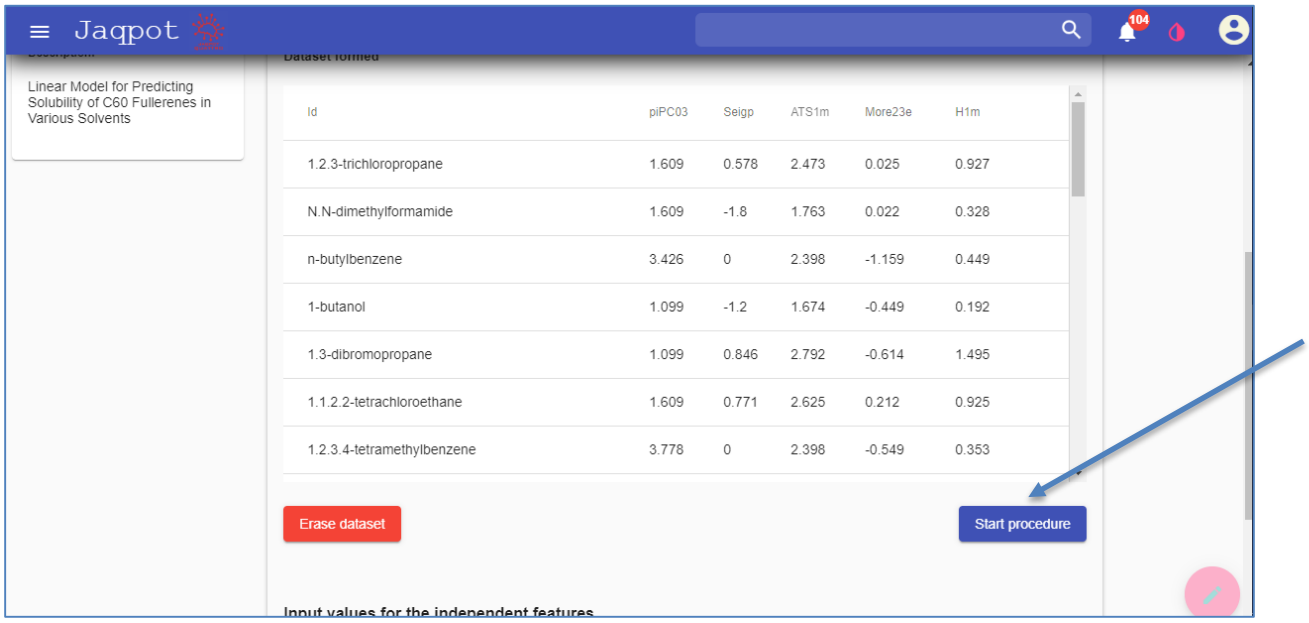

Figure 10. Starting the prediction procedure

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 11).

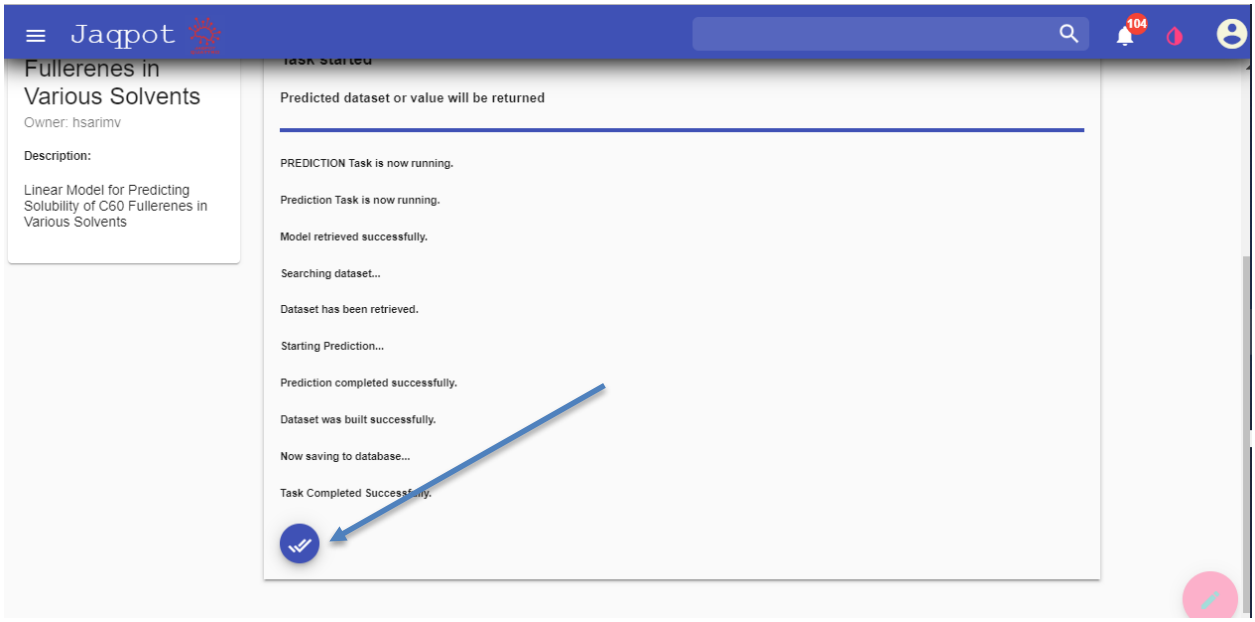

Figure 11. Starting the prediction procedure

The end-point predictions are shown along with all the independent variable values. The user can download these results by clicking on the “Download” button. (Figure 12), By clicking on the “View predicted values only”, only the end-point predictions are displayed (Figure 13).

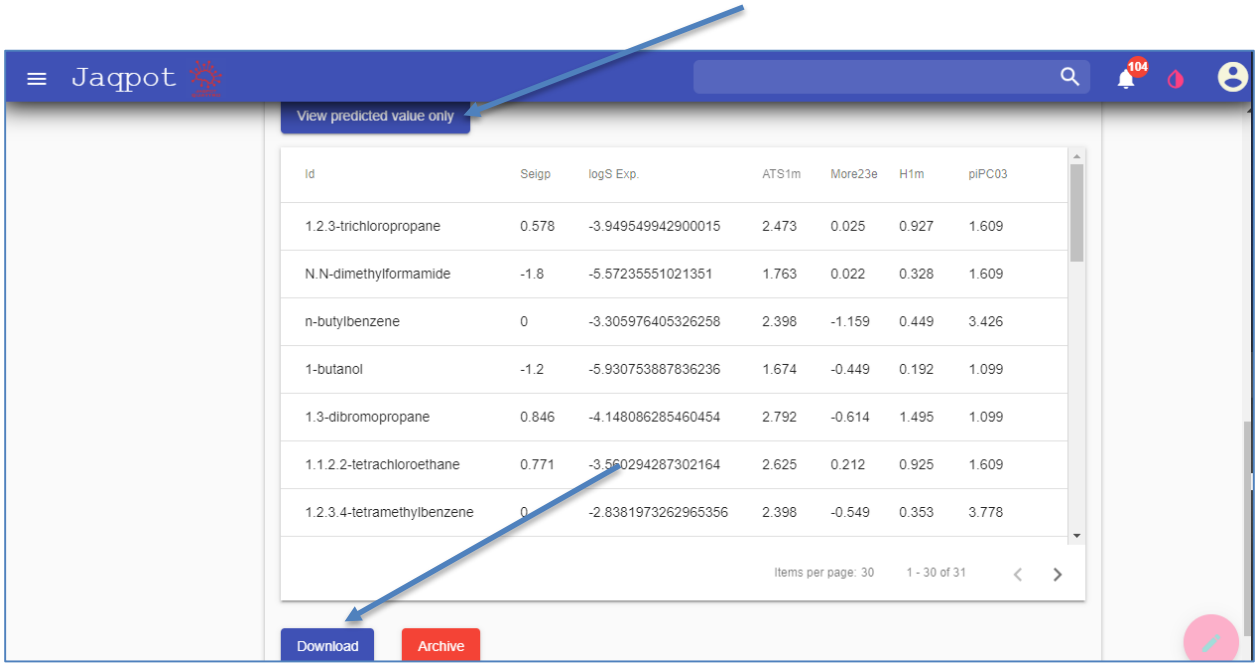

Figure 12. Viewing the complete dataset (independent variables and end-point predictions)

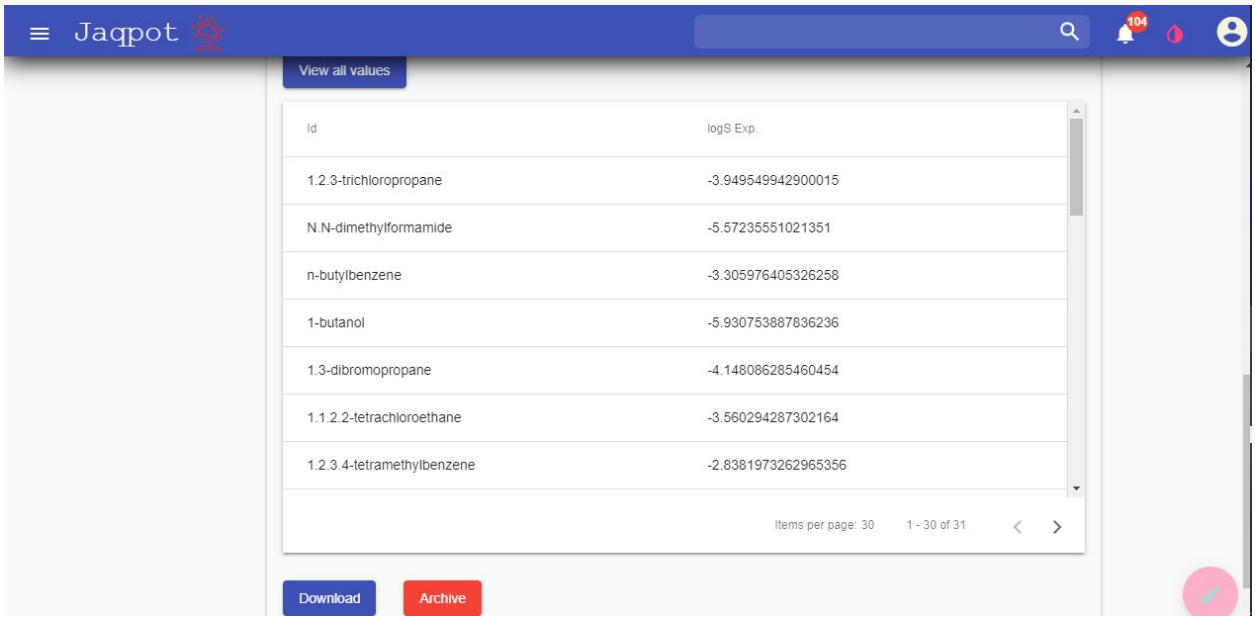

Figure 13. Viewing end-point predictions only

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 14).

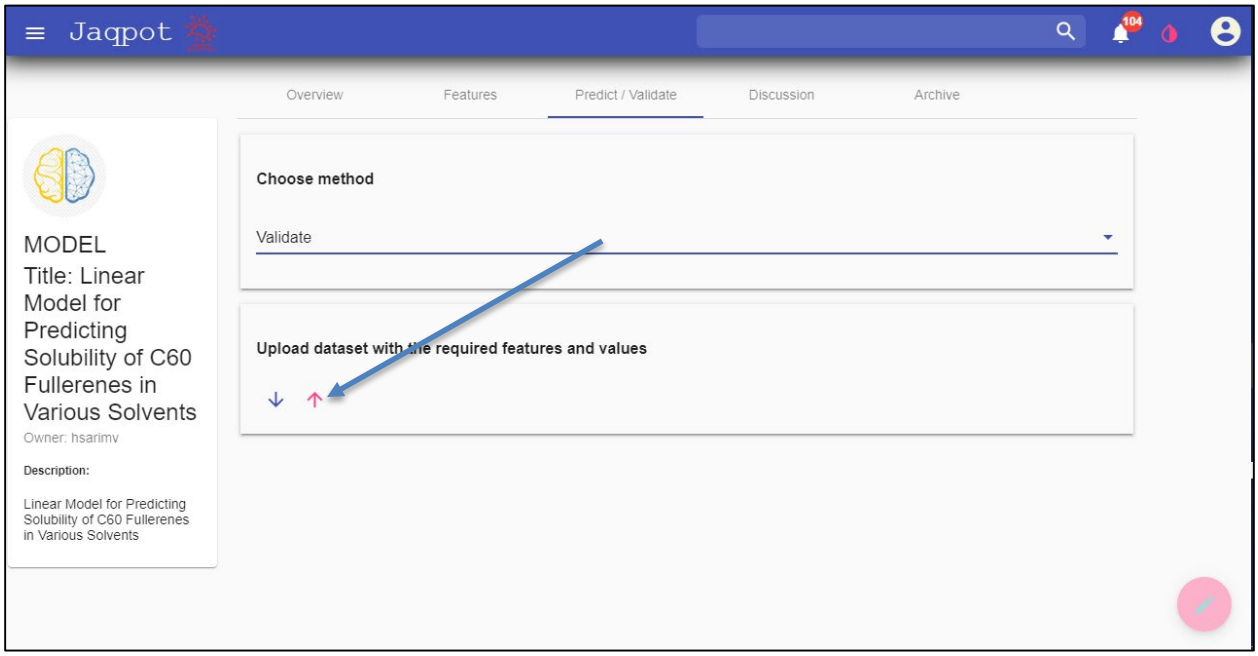

Figure 14. The “Predict/Validate” tab – Validate Option

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 15).

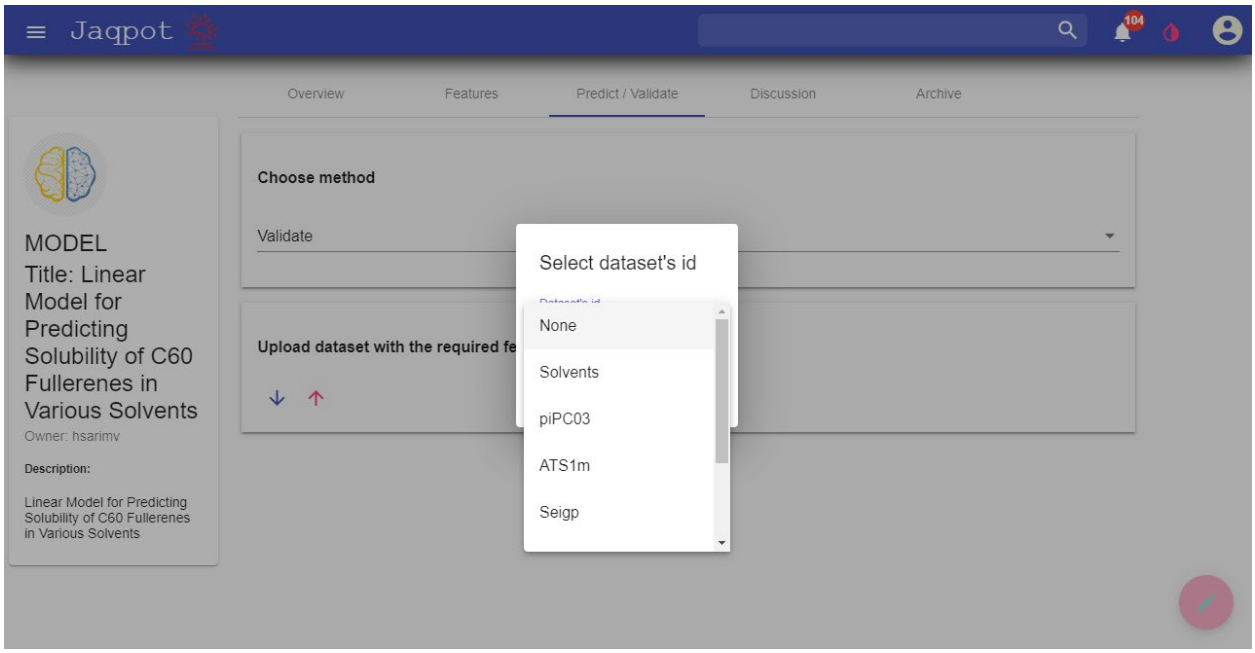

Figure 15. List of variable names contained in the csv file

Here the user selects the “Solvents” column and confirms his selection by clicking on the button that prompts him to continue (Figure 16):

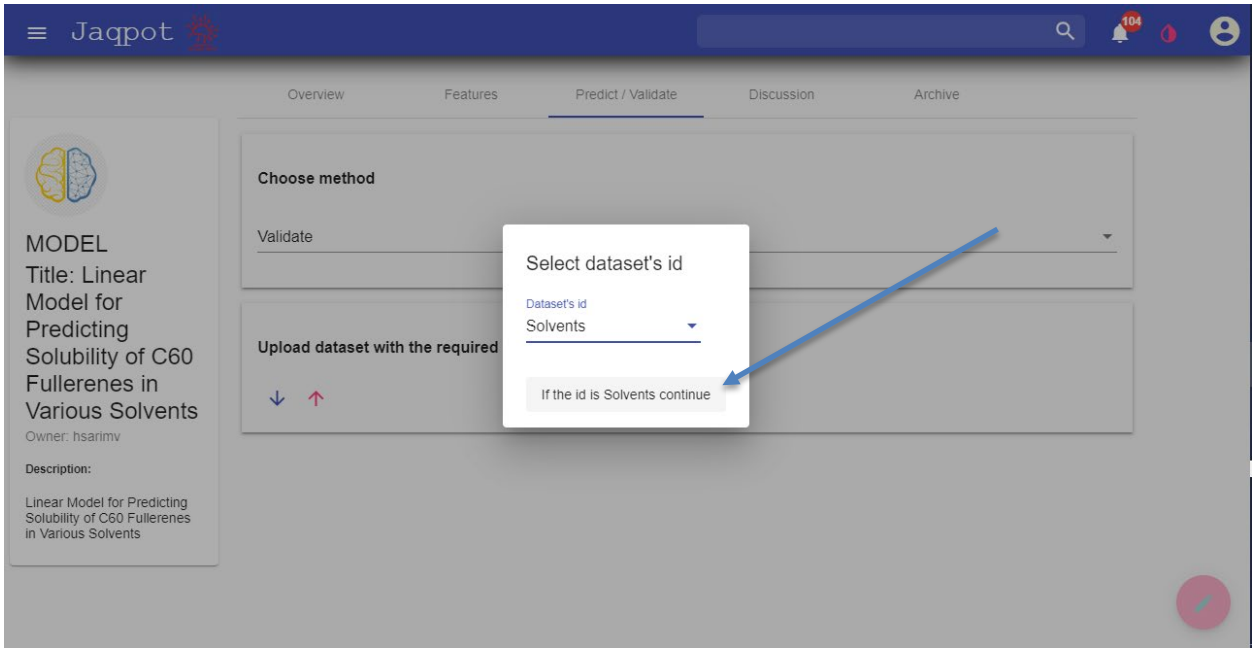

Figure 16. Selecting the “id” column

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 17).

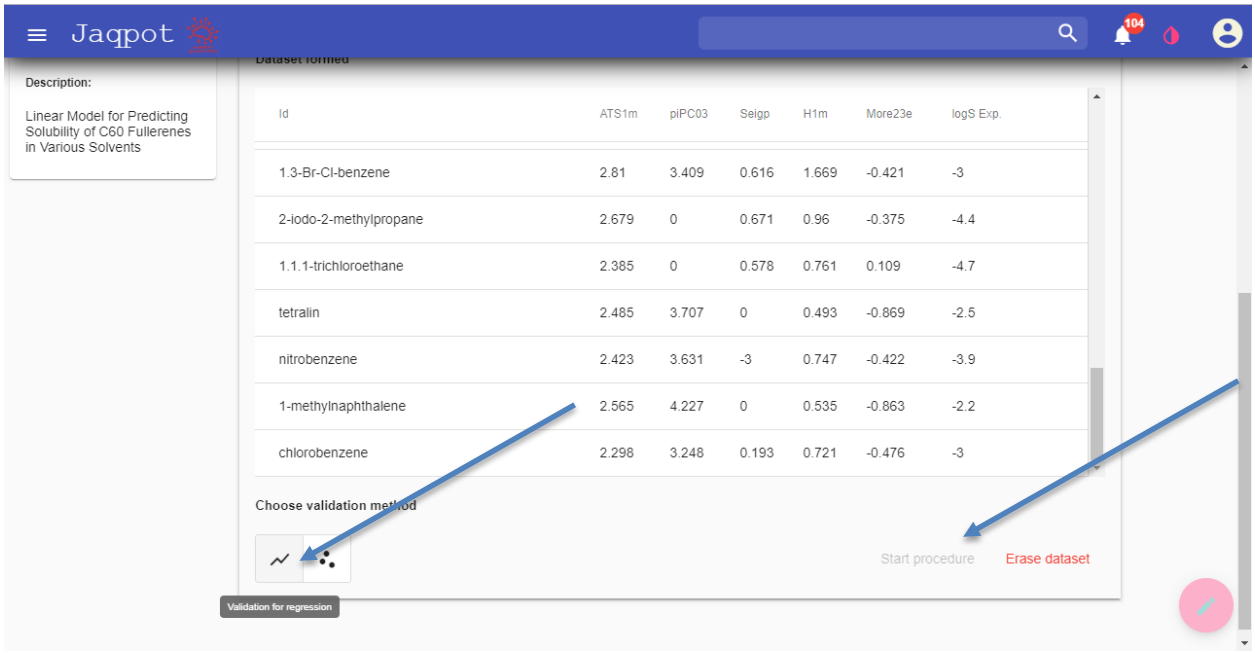

Figure 17. Starting the validation procedure.

The user is informed about the process of the task and waits until the task has been completed successfully (Figure 18). Then he can click on the “Check button” button to view the results.

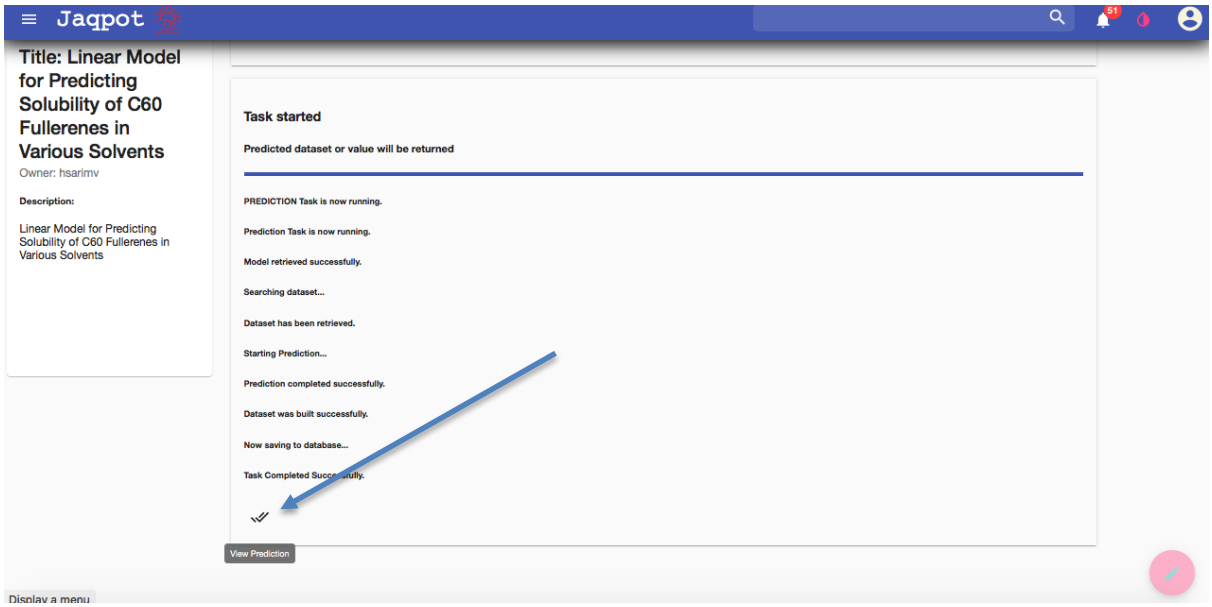

Figure 18. Task progress.

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 19-20)

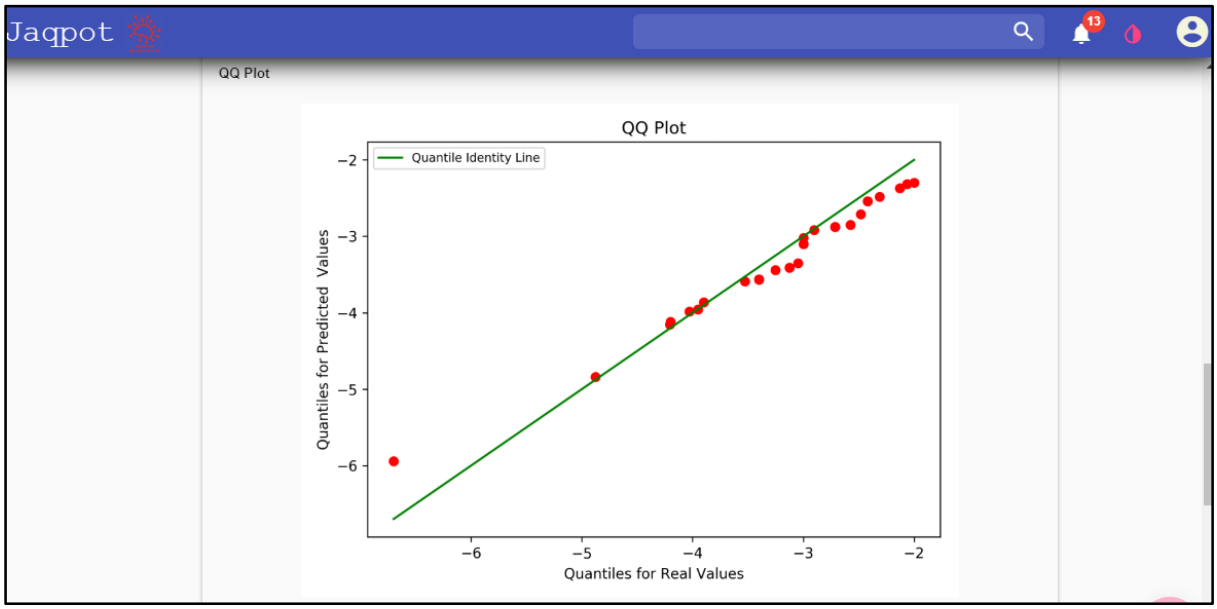

Figure 19. Validation results - QQ plot .

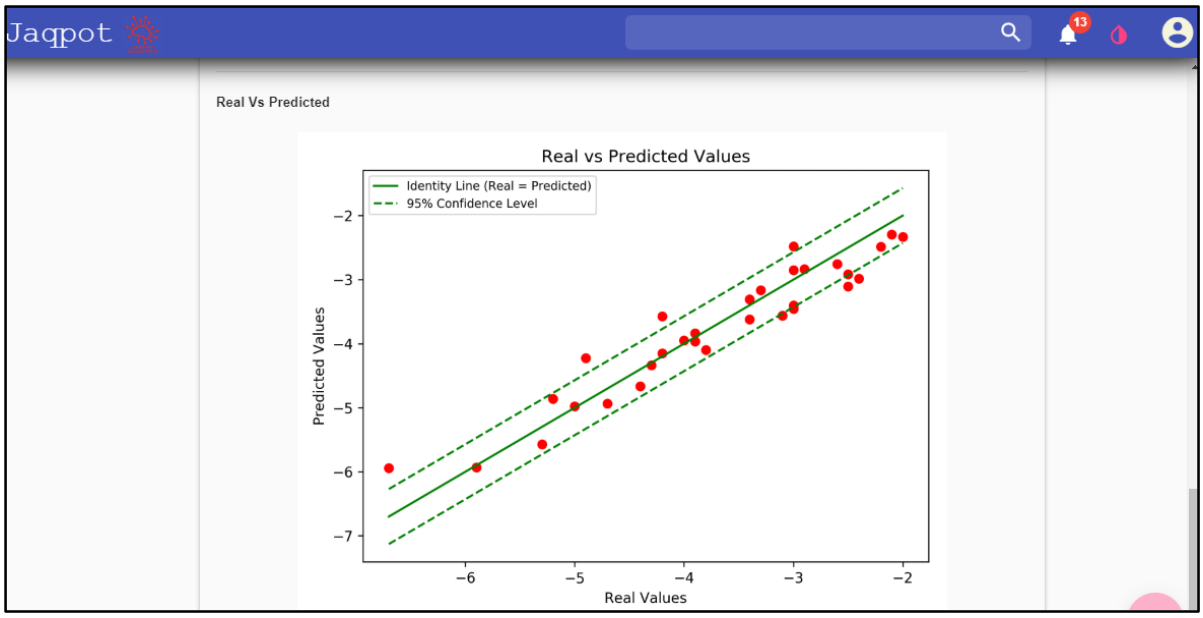

Figure 20. Validation results - *Real vs Predicted* plot .

In the discussion tab any user of the model can leave a comment (Figure 21):

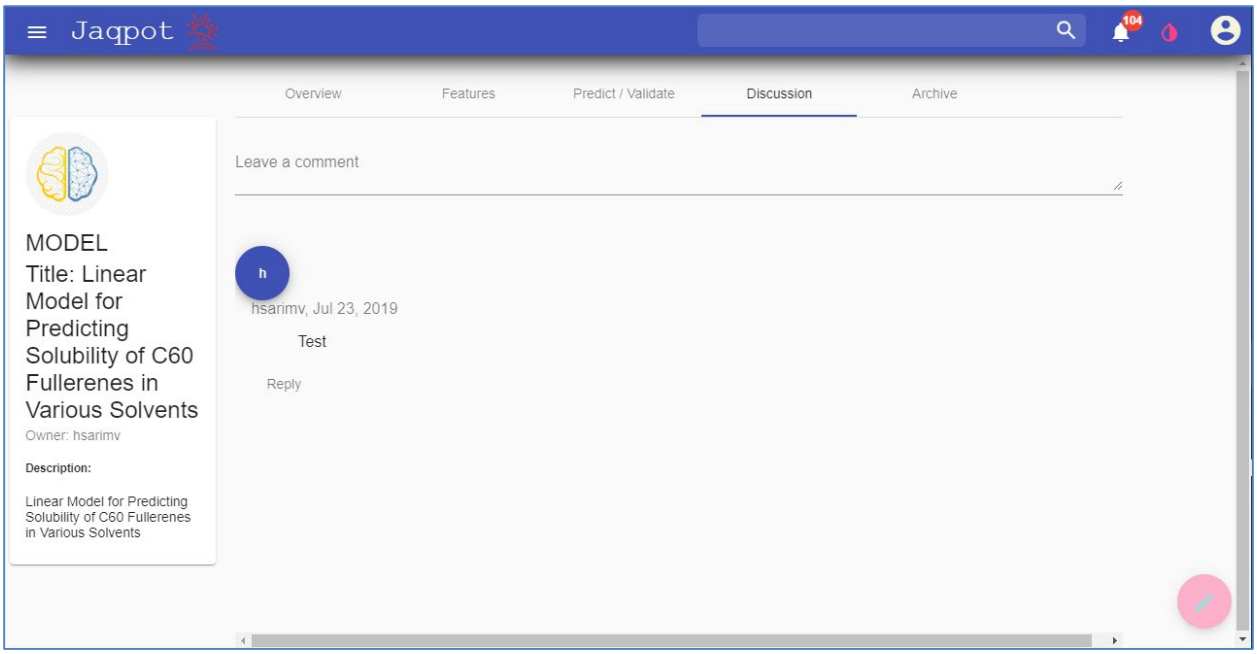

Figure 21. The "Discussion" tab.

The “Archive” tab is not activated yet. When activated the user will be able to save results of model runs so that he will not need to repeat the calculations each time he enters Jaqpot 5.

Support


NanoCommons
Nano-Knowledge Community


OpenRiskNet
RISK ASSESSMENT E-INFRASTRUCTURE