

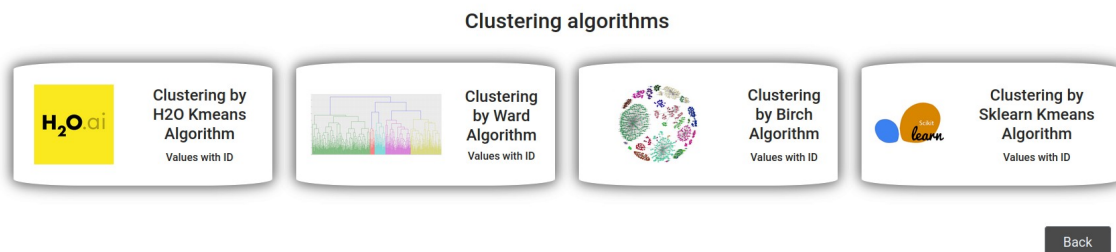
Manual iHelp Model Trainer

Once inside iHelp model trainer (<https://modeltrainer.ihelp-project.eu/>), we log in, user: pepe45 and pass: pepe45.

The first thing to do is to generate a model as we don't have any in the My Models section.

- First we click or select 'New Model'.

We have four algorithms, but I am going to use only one because the way to generate the models is similar in all of them, the only thing that could change are the properties of each algorithm.



- Once the algorithm is selected, we will immediately see the name inputs, we will fill in the model name that identifies the model and data file, we can add files with extension **.csv**, after adding the dataset file, we press next.

Clustering algorithm Sklearn

Model name *Required*

e.g.: myModel, anomalies, tabberMachine_Jun2018, etc.

Data file *Required*

No se ha seleccionado ningún archivo.

To train the model with a sample dataset

- Then a table will be generated where the data from our data file will be visible, in this table we can select the columns we want to use to generate our models, we can also delete columns we don't want to use and we can also filter the data we want to use from our rows of each column. In this case, we are deleting (for whatever reason, we don't need or don't like) the **sample_id**, **pation_cohort** and **sample_origin** columns and we are selecting the **LYVE1**, **REG1B** and **TFF1** columns as prediction columns of our dataset.

show entries

Showing 1 to 20 of 590 entries

First Previous **1** 2 3 4 5 ... 30 Next Last Search:

☐ All Columns For Train Model ☐ Delete All Columns

<input type="checkbox"/> sample_id	<input type="checkbox"/> patient_cohort	<input type="checkbox"/> sample_origin	<input type="checkbox"/> age	<input type="checkbox"/> sex	<input type="checkbox"/> stage	<input type="checkbox"/> benign_sample_diagnosis	<input type="checkbox"/> plasma_CA19_9	<input type="checkbox"/> creatinine	<input checked="" type="checkbox"/> LYVE1	<input checked="" type="checkbox"/> REG1B	<input checked="" type="checkbox"/> TFF1	<input type="checkbox"/> REG1A	<input type="checkbox"/> diagnosis
S1	Cohort1	BPTB	33	F	NA	NA	11.7	1.83222	0.8932192	52.94884	654.282174	1262	no pancreatic disease
S10	Cohort1	BPTB	81	F	NA	NA	NA	0.97266	2.037585	94.46703	209.48825	228.407	no pancreatic disease
S100	Cohort2	BPTB	51	M	NA	NA	7	0.78039	0.1455889	102.366	461.141	NA	no pancreatic disease
S101	Cohort2	BPTB	61	M	NA	NA	8	0.70122	0.00280488	60.579	142.95	NA	no pancreatic disease
S102	Cohort2	BPTB	62	M	NA	NA	9	0.21489	0.00089596	65.54	41.088	NA	no pancreatic disease
S103	Cohort2	BPTB	53	M	NA	NA	NA	0.84825	0.003393	62.126	59.793	NA	no pancreatic disease
S104	Cohort2	BPTB	70	M	NA	NA	NA	0.62205	0.1743808	152.277	117.516	NA	no pancreatic disease
S105	Cohort2	BPTB	58	F	NA	NA	11	0.89349	0.00357396	3.73	40.294	NA	no pancreatic disease
S106	Cohort2	BPTB	59	F	NA	NA	NA	0.48633	0.00194532	7.021	26.782	NA	no pancreatic disease
S107	Cohort2	BPTB	56	F	NA	NA	24	0.61074	0.2787785	83.928	19.185	NA	no pancreatic disease
S108	Cohort2	BPTB	77	F	NA	NA	NA	0.29406	0.00117624	6.218	28.297	NA	no pancreatic disease
S109	Cohort2	BPTB	71	M	NA	NA	23	1.05183	0.8603368	243.082	608.284	NA	no pancreatic disease
S11	Cohort1	BPTB	49	F	NA	NA	NA	0.85956	1.416114	161.83077	74.099025	505.571	no pancreatic disease
S110	Cohort2	BPTB	53	M	NA	NA	7	1.91139	1.516773	150.89	590.686	NA	no pancreatic disease
S111	Cohort2	BPTB	56	F	NA	NA	12	0.91611	0.5996449	93.811	93.576	NA	no pancreatic disease
S112	Cohort2	BPTB	60	F	NA	NA	28	0.50895	0.00203058	24.366	19.698	NA	no pancreatic disease
S113	Cohort2	BPTB	69	F	NA	NA	9	0.41847	0.00167388	17.102	0.03264066	NA	no pancreatic disease
S114	Cohort2	BPTB	60	F	NA	NA	47	0.80301	0.00321204	3.588	30.071	NA	no pancreatic disease
S115	Cohort2	BPTB	55	M	NA	NA	17	1.28934	2.285351	67.468	269.805	NA	no pancreatic disease
S116	Cohort1	BPTB	28	F	NA	NA	8.7	0.50895	0.5830097	13.61906	267.193539	381	no pancreatic disease

☒ delete ☒ delete ☒ delete ☐ delete ☐ delete ☐ delete ☐ delete ☐ delete ☐ delete ☐ delete ☐ delete ☐ delete ☐ delete

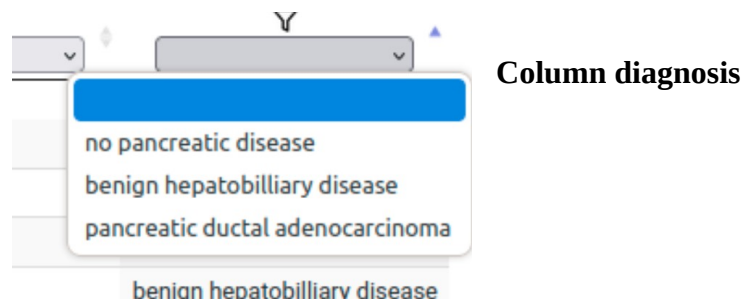
Generating new dataset CSV

- Then to finish we must click on the '**Generating new dataset CSV**' button, which is at the base of the data table, as here we generate the new dataset that the algorithm will use, as you may have filtered for example by sex and only generate models for women or men...

Generating new dataset CSV

- Then we have the parameters of the algorithm, if we pass the mouse pointer over them, we will see a small explanation of the property or parameters of the algorithm, if we double click on the input of each parameter the recommended parameters will be displayed, we can add the one we want but possibly the algorithm will produce an error and we will not generate appropriate and functional models or with a low percentage of predictions.

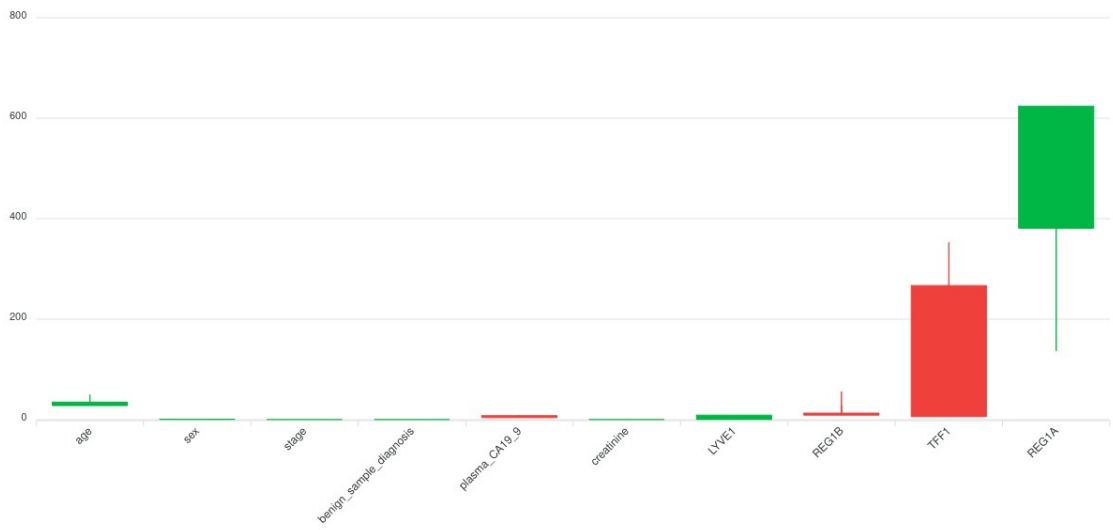
One parameter that we must choose is the **class column**, which is the column that the algorithm will predict, which in our dataset is the '**diagnosis**' column which is where we have the diagnosis for each patient (each row and as we see in the image below) and the remaining columns are a component of a urinalysis.



Random State:	Number init:	Init:	Advanced models
<input type="text" value="47"/>	<input type="text" value="100"/>	<input type="text" value="k-means++"/>	<input type="button" value="False"/>
Max Iterations:	Clusters:	Column class	NULL & NAN values
<input type="text" value="500"/>	<input type="text" value="3"/>	<input type="text" value="diagnosis"/>	<input type="button" value="Drop rows"/>

My advice is to use these parameters that we have in the image above, and then try several iterations, not moving away from these base parameters, which we show in the image above, finally we press next and our algorithm will run, we wait for it to finish (a long time in the version of the iHelp platform), and then the graphs and images will be generated, which will tell us the quality of the model generated by the algorithm.

133. Regression Chart

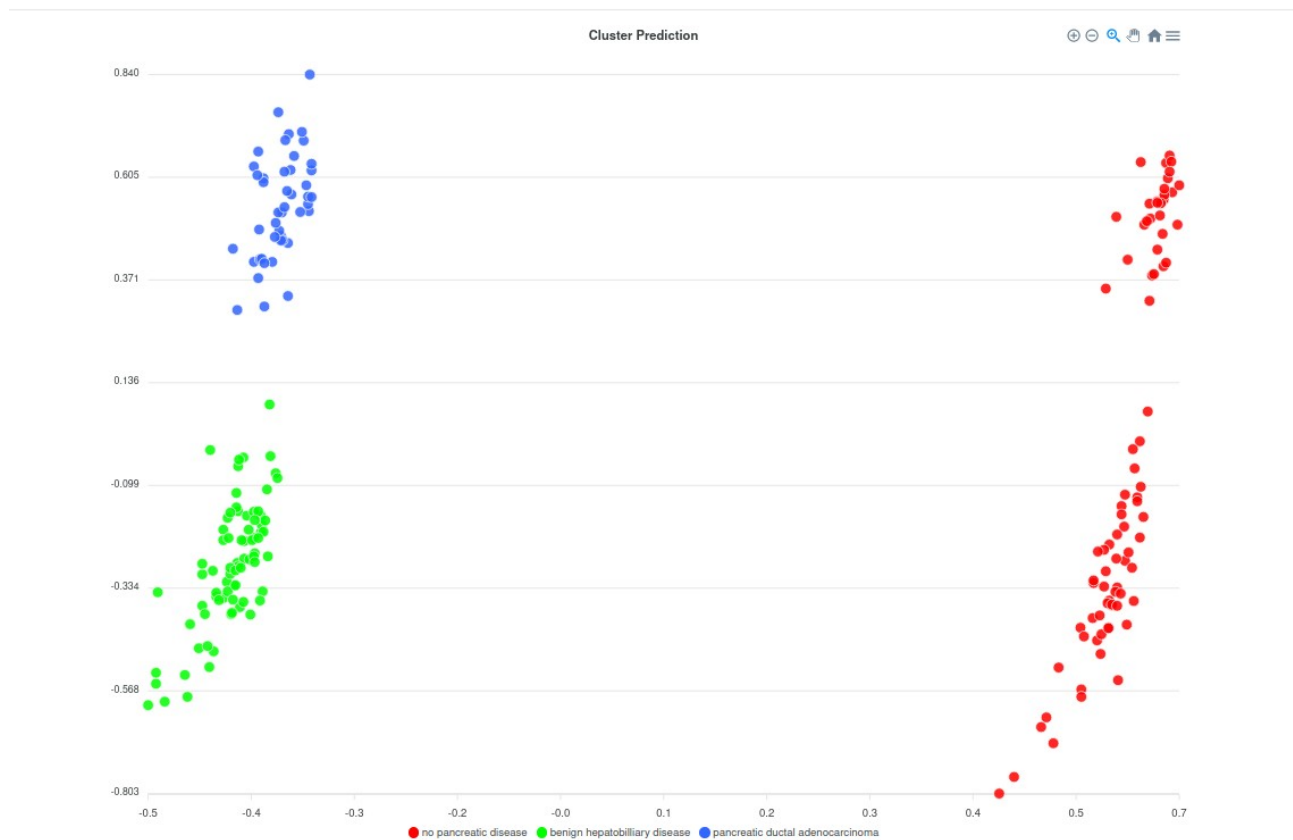


Show 10 entries

Search:

age	sex	stage	benign_sample_diagnosis	plasma_CA19_9	creatinine	LYVE1	REG1B	TFF1	REG1A	di
28	F	none	none	8.7	0.50895	0.5830097	13.61906	267.193539	381.0	pa d
30	M	none	none	10.8	1.18755	1.600313	22.471281	299.11836	570.0	pa d
35	M	none	none	4.0	0.24882	9.9528E-4	9.245166	6.03070104	624.0	pa d
40	M	none	none	12.84	0.42978	1.056174	1.8277425	33.106275	64.821	pa d
44	F	none	none	8.47	0.18096	0.01192871	2.1460546	0.01411488	13.616	pa d
45	M	none	none	9.6	1.3572	2.392864	28.509285	353.656656	563.0	pa d
46	M	none	none	6.58	0.35061	0.00140244	2.191019	33.19384	0.0	pa d

In our opinion, this is the most important image, as we can see how the algorithm has separated the clusters well by the classes of the column class that we have commented above (**diagnosis: pancreatic ductal adenocarcinoma, no pancreatic disease and benign hepatobiliary disease**). but we can see that the red cluster (no pancreatic disease) is not completely united, but it would not matter either (according to criteria) as it is the class that tells us that there is no pancreatic cancer, but the clusters should be the points together for each cluster or colour, to adjust the model we should change the parameters until the red points are all more or less together in a cluster.



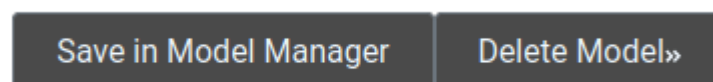
- The last part of the graphs is the SHAP explanation or **Explanation of the model built by the SHAP method**.

Where in our opinion it looks better, it informs about the weights of each parameter to generate the model, for example in the **Waterfall** tab:



As we see in the image above left the **value $f(x) = 1.939$** this is almost a 2 which would be the prediction of the model, but this data should be reviewed by the doctors as it informs which component of the analysis is most important to generate a model.

- After reviewing the data and graphs we can either save the model in the model manager or delete it.



- The **My Models** section is internal to **iHelp Model Trainer**, only the generated models are there, but it is possible that they are not in the model manager, in order to be able to use the models in the model trainer we must press **Save in model manager** when we finish generating the model.

