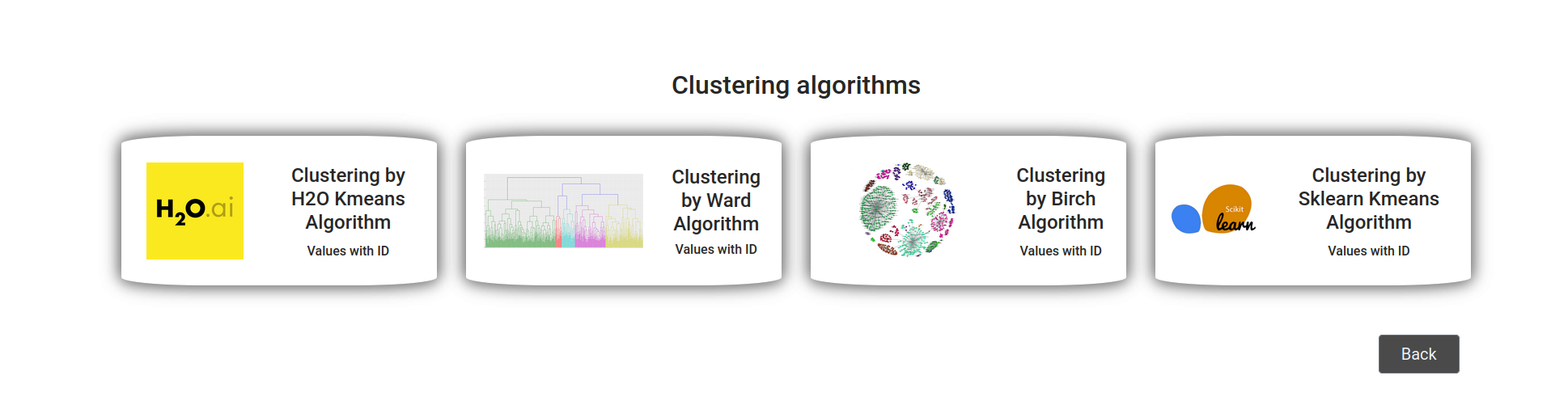
**Manual iHelp Model Trainer**

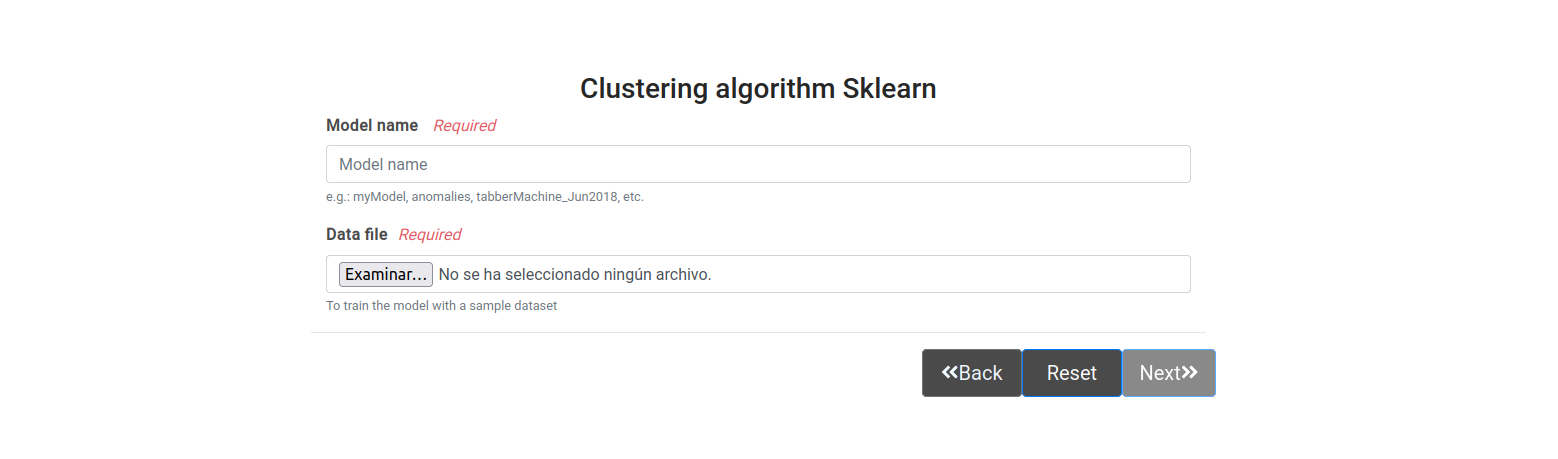
Once inside iHelp model trainer (<https://modeltrainer.ihelp-project.eu/> ), we log in, user: pepe45 and pass: pepe45.

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

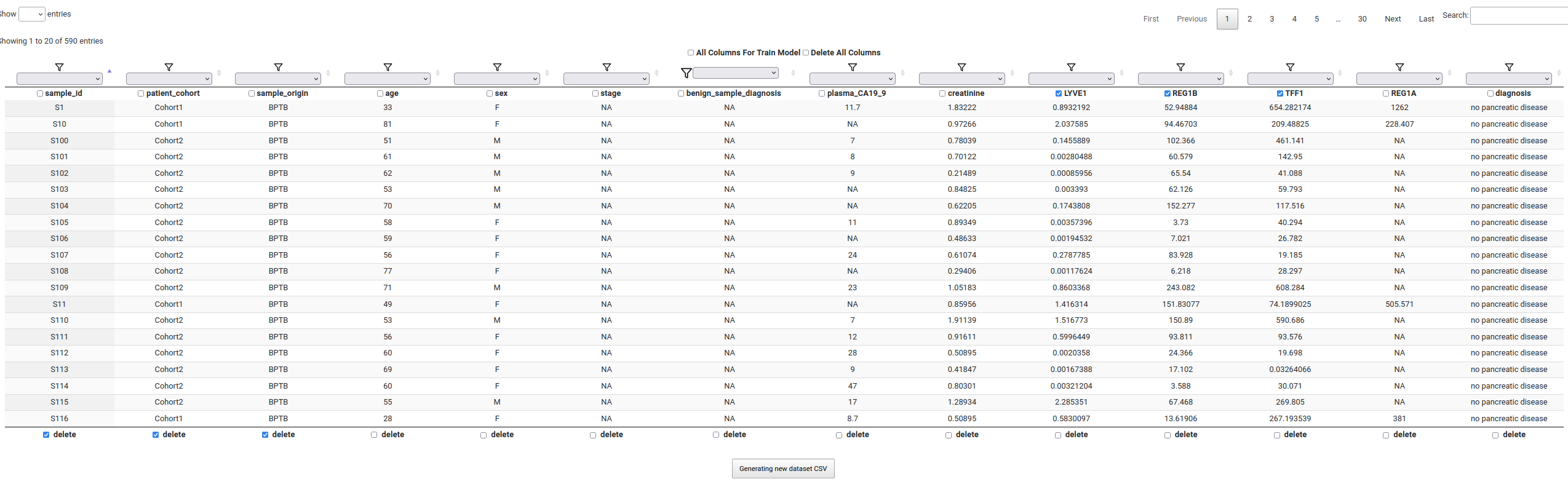
- First we click or select 'New Model'.

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. We will use the scikit-learn.

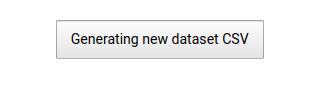


- Once the algorithm is selected, we will immediately see the name inputs, we will fill in the model name that identifies the model and data file, we can add files with extension .**\*csv**, after adding the dataset file, we press next.

- 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\_sample\_origin** columns and we are selecting the **LYVE1, REG1B and TFF1** columns as prediction columns of our dataset.

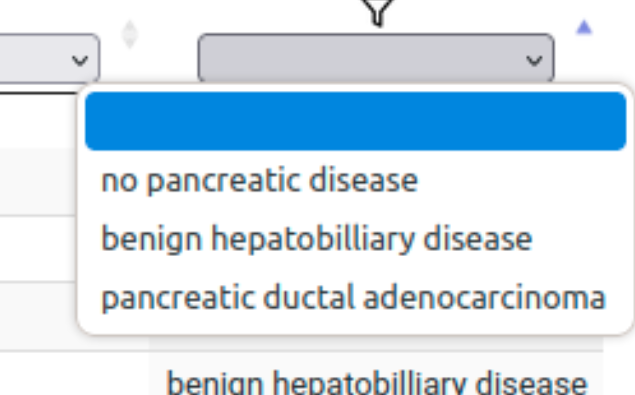


- Then to finish we must click on the '**Generating new dataset CSV**' button, which is at the base of the data table, as here we generate the new dataset that the algorithm will use, as you may have filtered for example by sex and only generate models for women or men...

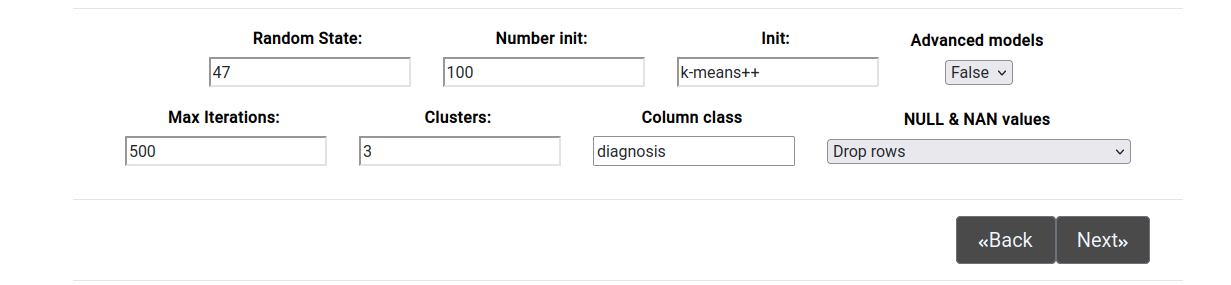


- Then we have the parameters of the algorithm, if we pass the mouse pointer over them, we will see a small explanation of the property or parameters of the algorithm, if we double click on the input of each parameter the recommended parameters will be displayed, we can add the one we want but possibly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

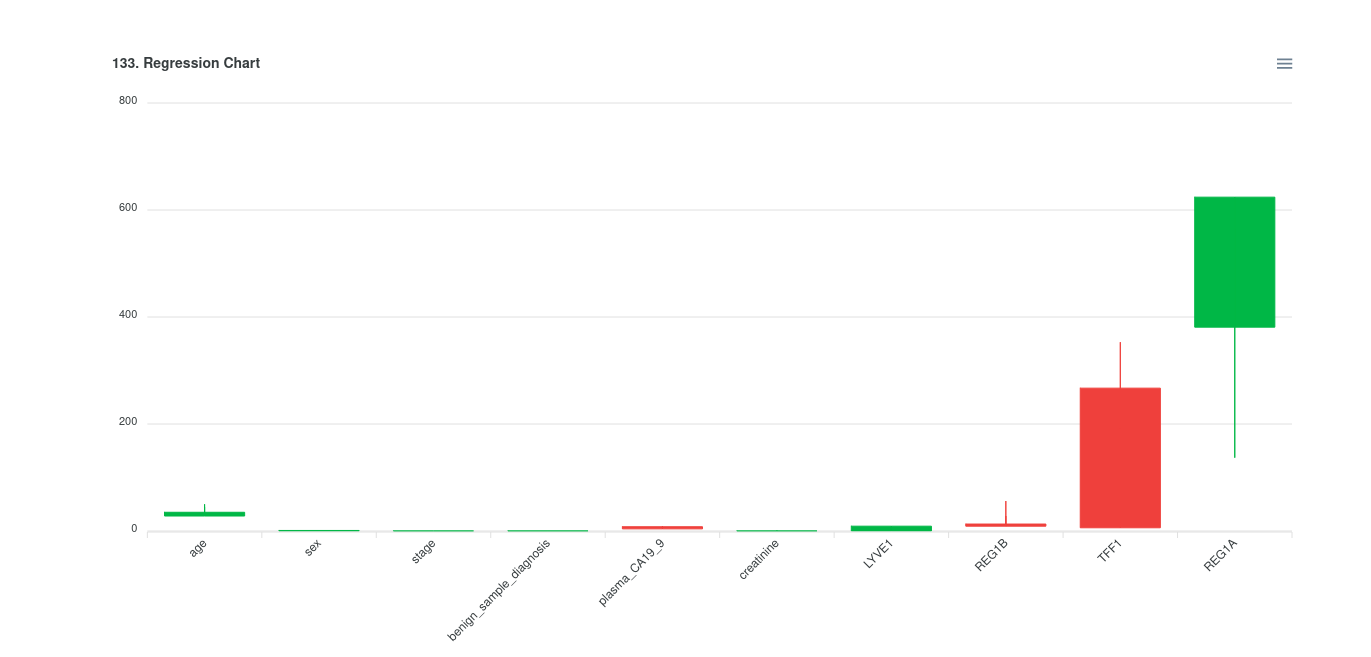
One parameter that we must choose is the **class column**, which is the column that the algorithm will predict, which in our dataset is the '**diagnosis**' column which is where we have the diagnosis for each patient (each row and as we see in the image below) and the remaining columns are a component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

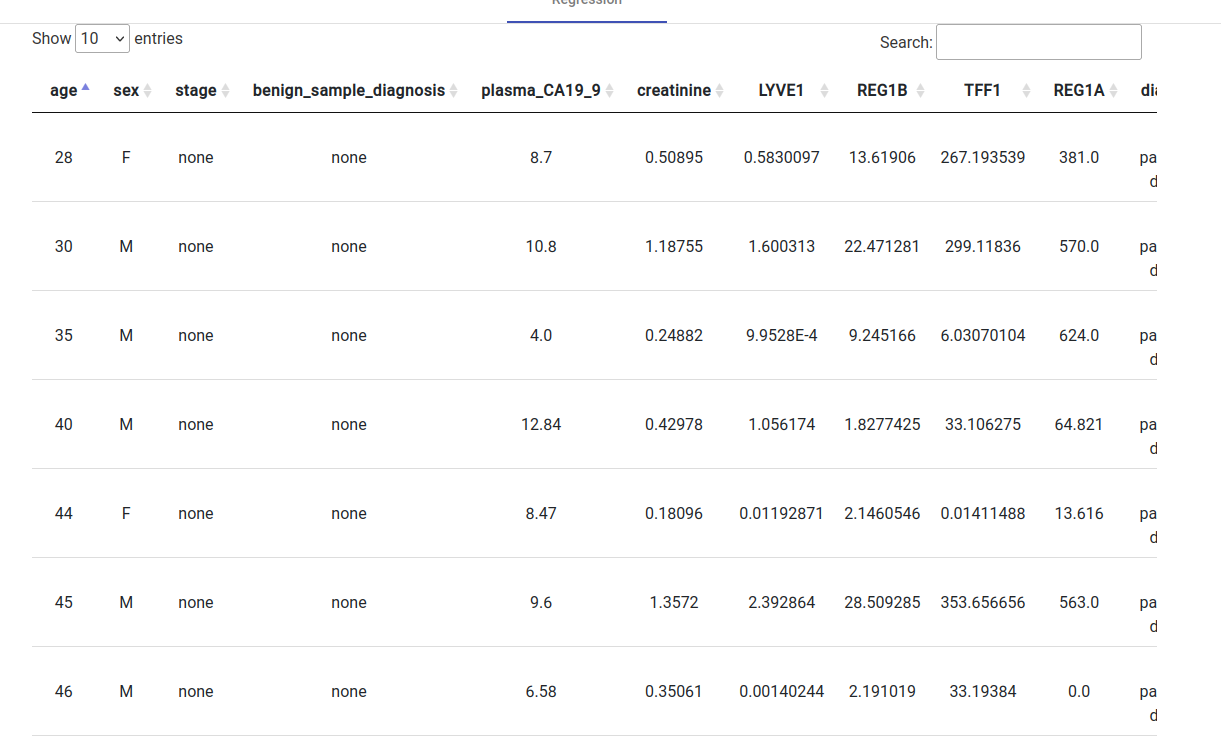


**Column diagnosis**

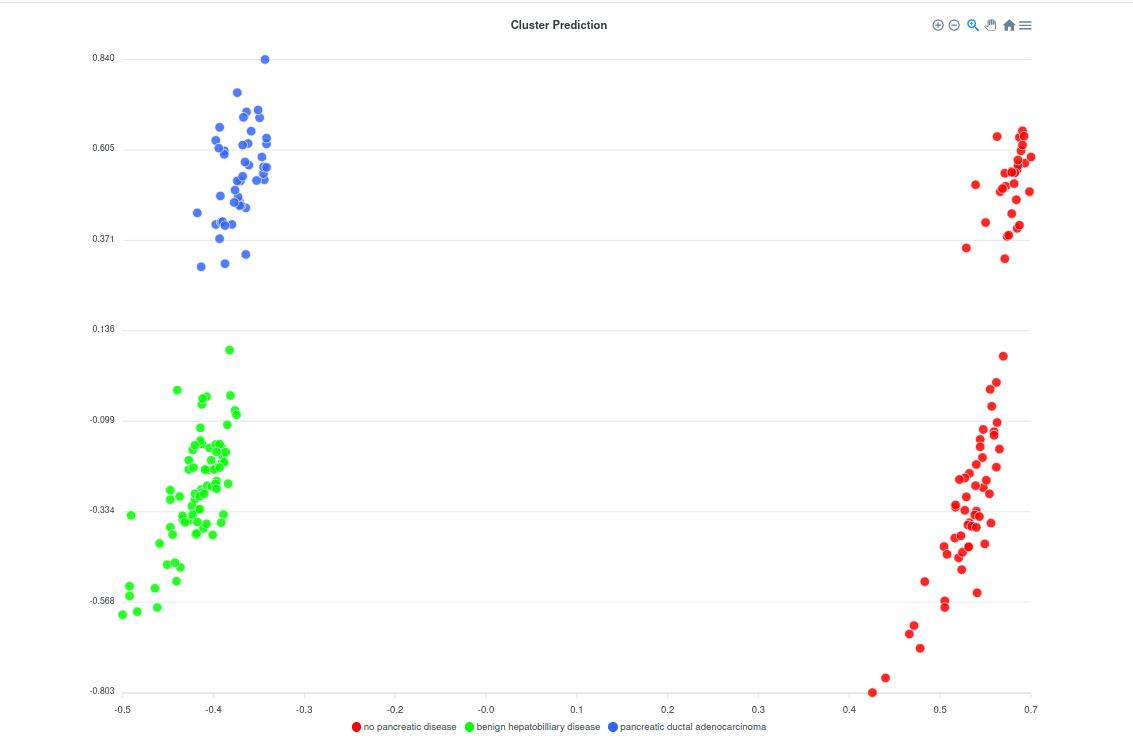


My advice is to use these parameters that we have in the image above, and then try several iterations, not moving away from these base parameters, which we show in the image above, finally we press next and our algorithm will run, we wait for it to finish (a long time in the version of the iHelp platform),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





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.

