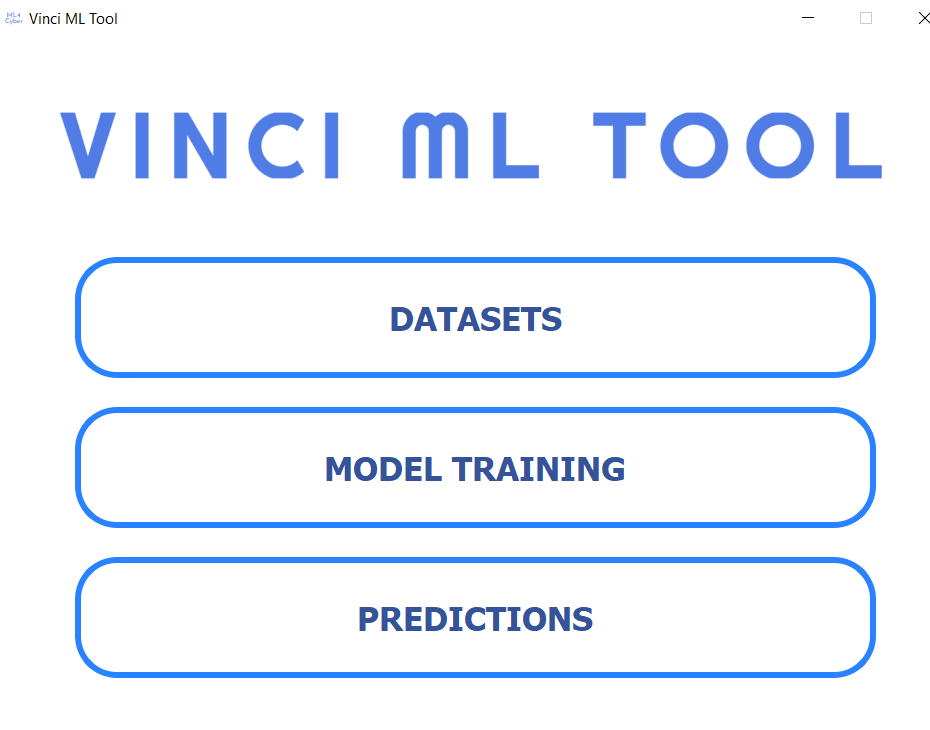
# VINCI User Guide

1. To run the tool double click the executable
2. Below is the main VINCI interface:

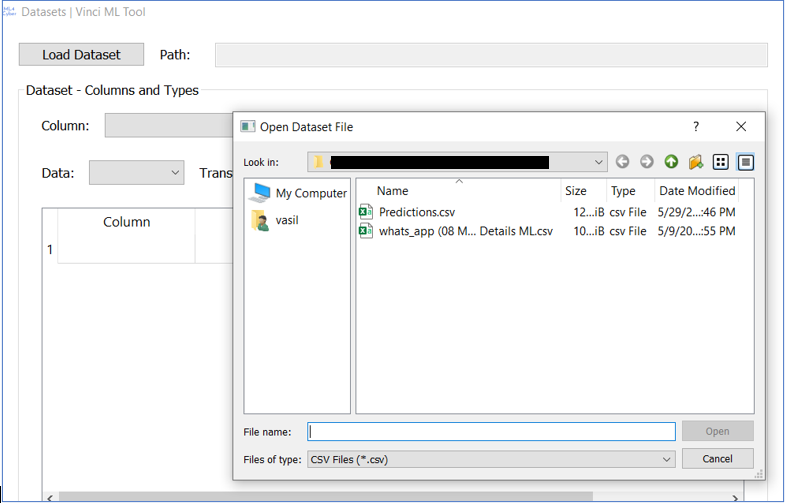


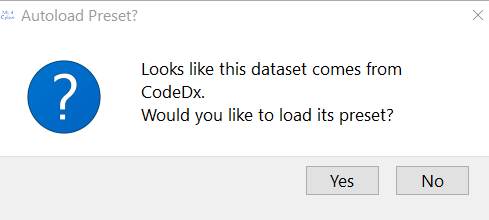
# Load Dataset and Save Schema

1. Click the **Datasets** button, press the **Load Dataset** on the top left corner and select the csv or xml file that contains the dataset. (It seems to work mostly with the XMLs with the CVS you have to manually pick the “preset” on the right)

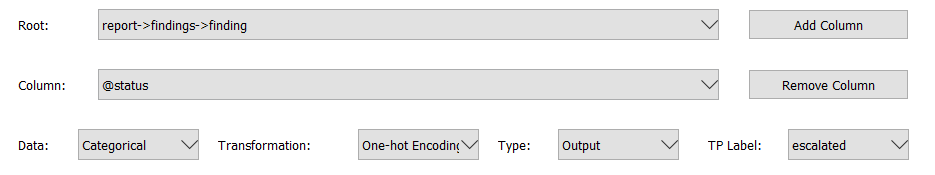
For testing proposes start by using the whats\_app(08 May 2021).xml that contains code Dx results.

1. The tool will recognize the type of report and preload the recommended features for best prediction accuracy.





1. You can add or remove columns.
2. Make sure you select the “TP Label” as it appears in the dataset for example for. Select the **Column**-> **Status** make sure the **Type** is “Output” and from the drop down option at the **TP Label** select “True Positive” or “Escalated” (for Code Dx results) and add column (if the column exists you do not need to add it, this is only for datasets that do not have one), then the dataset.



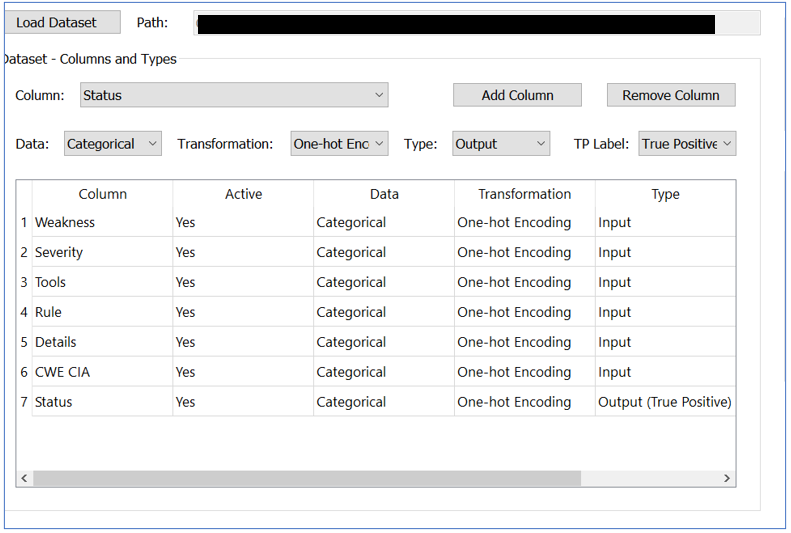
IMPORTANT: you need to save the dataset as CSV to be able to import it at the Labeler.

I noticed that for the files that do not have a status field like the code dx xml even if you assign a new column as output and then save the schema that schema does not work to make predictions.

If you open the Code Dx xml and then save the schema the schema does not make predictions you have to import it and select the output again even if it is already displayed from the preset you have to remove the original from the preset and add it again for it to work.

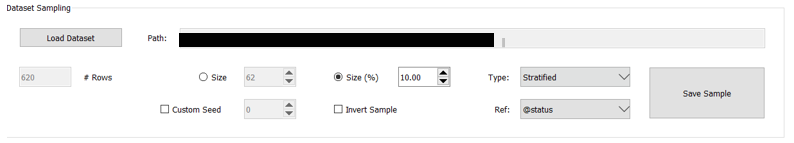
When something doesn' t work the GUI disappears, like when you save the spotbugs xml as csv then separate the training set the GUI disappears

1. Save new Dataset as a .csv file do not forget to add the extension. This dataset will be used later for predictions it contains all findings.

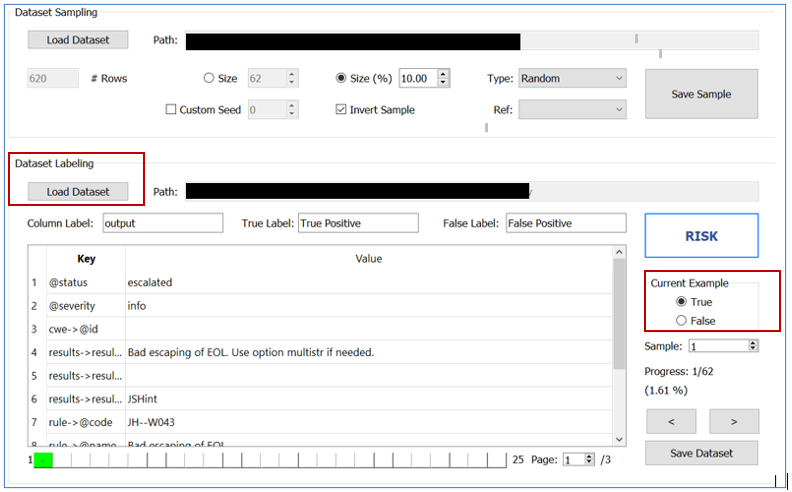


## LABELER

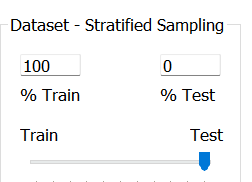
1. You can separate here the training set.
2. Load the dataset you saved as a .csv file before, select the **Stratified** type and save the training set.

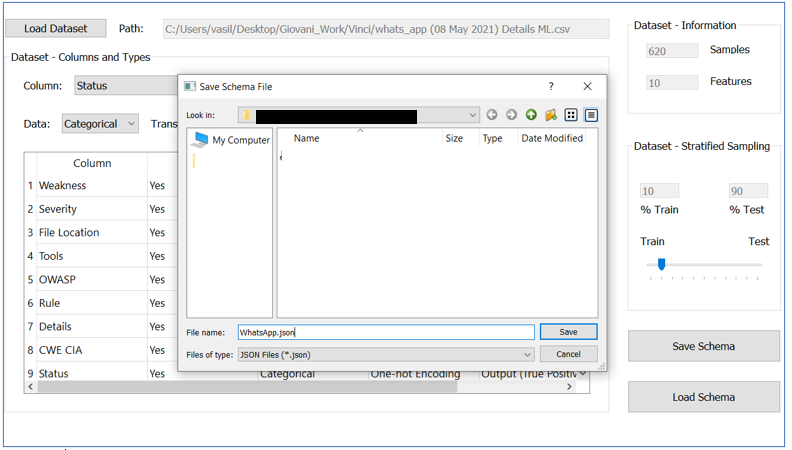


1. Load the training dataset you just created at the Dataset Labeling section. Each finding will be displayed separately and you will be able to assign the **True** or **False** label using the check boxes on the right.
2. Save the dataset and close the window.



1. You then have to create the schema as indicated in section “Load Dataset and Save Schema” steps 1 through 8 above.
2. IMPORTANT: you need to make sure before you save the training schema to save the whole dataset so slide up to 100%

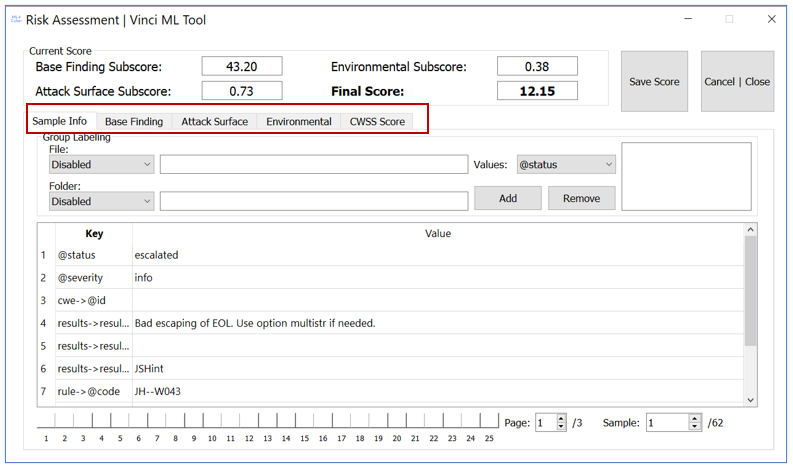




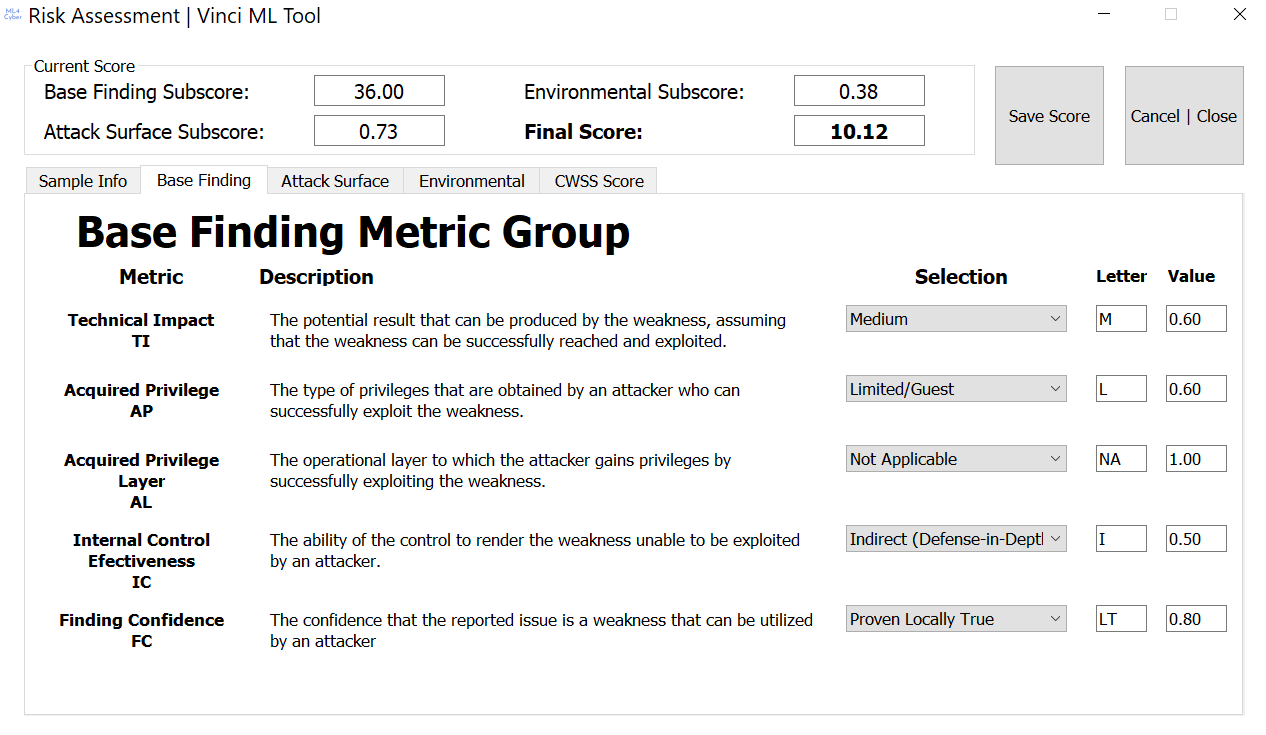
## Risk

We need the path displayed here (for risk and to answer the questions)

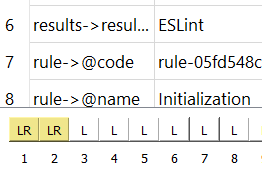
1. We can assign risk per finding or to a group of findings.
2. Select the Base Finding, Attack Surface, Environmental tabs to assign risk based the CWSS Score.



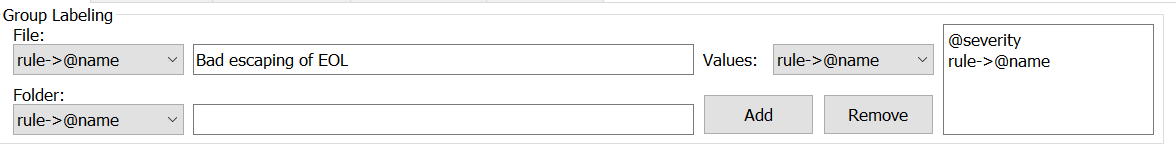
1. Answer as much questions as possible to get a more accurate risk score.



1. After you answered the questions select **Save Score.** Close the risk assessment interface to view and move to the next findings.
2. Your process will be tracked under the finding details where the “**L“** will indicate that the specific numbered finding is labeled as true or false and the “**R”** that a risk score is assigned.



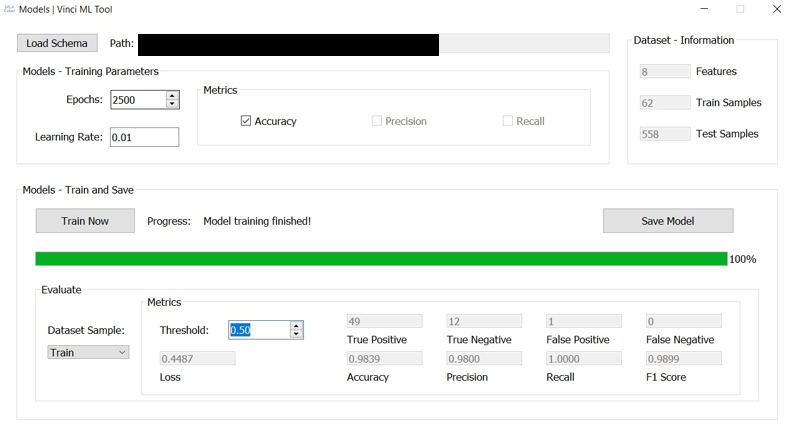
1. For group findings select **File** and **Folder.**



1. After you finish labeling and/or assigning risk save the dataset as a .csv file.
2. You then have to create the schema as indicated in section “Load Dataset and Save Schema” steps 1 through 8 above.

# Model Training

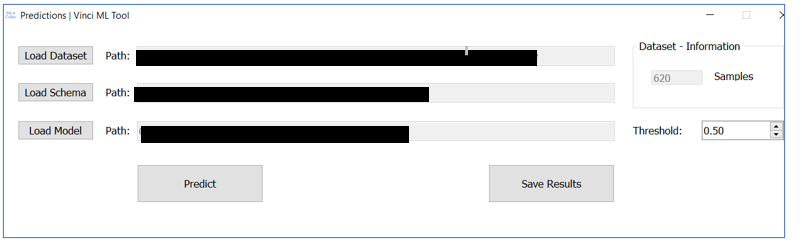
1. Load the schema you created and press “**Train Now**”, select a high **Epochs** number for higher accuracy.



1. At the lower half of this page you can select to view the Test set (**Dataset Sample** drop box) accuracy and metrics also if the whole dataset is labeled. See how changing the threshold the values are affected also.
2. Save the model by pressing **Save Model** on the middle right (picture above). Save it by adding the .h5 extension.

# Predictions

1. Load the original dataset, the schema you created and the training model. Press Predict and Save Results when finished to save a .csv file with the predictions.



IMPORTANT:

* If you use an xml first save it as csv
* For the predictions when you load the dataset you should import the csv not the xml
* The schema should be labeled and it should be created from the training set (also make sure before you save the schema that you moved the slider to 100% training)
* The csv you import in the predictions should contain all findings , the schema only the training