Help and Supplementary Documentation

Table of Contents

Table of Figures

Welcome Message

Hello and thank you for coming to the help documentation! This documentation covers some basic functionality and some of the concepts the application was built on. While this document covers the basic concepts for the application a more in-depth review can be found in Chapter 2 of the thesis Titled “Development of Software Tools for Efficient and Sustainable Process Development and Improvement” by Jake Stengel from the Rowan University Department of Chemical Engineering. During his time at Rowan University Jake Stengel was the main developer of the application and the backend algorithms. If you have any questions that the document does not answer, feel free to contact the developer or his advisor Dr. Yenkie:

Stengel.jake2@gmail.com

yenkie@rowan.edu

Model Explanation/ Random Forest Information:

Random Forest Classifier (RFC) was the Machine Learning (ML) algorithm which was used for the backend of the application. RFC is a collection of decision trees, each of which generates a predictive answer. A decision tree is a collection of random inequalities generated by the computer. As Figure 1 depicts, each tree is fed the information from the dataset and passes through each of the inequalities. If the inequality is true it goes to the left branch while it is false it goes down the right branch. For example, in Figure 1 the first decision in all trees is false since it goes down the right branch of each tree.

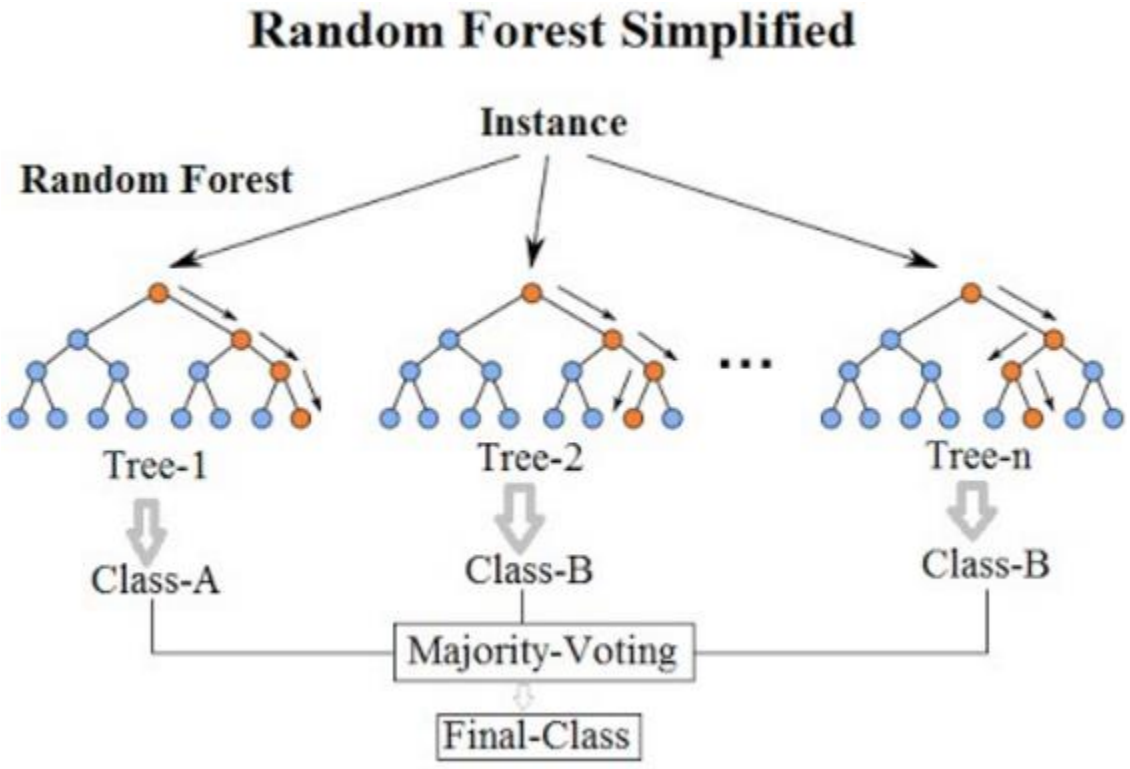

Figure 1: simplified version of random forest classification, showing Class-B as the correct answer for the RFC algorithm.

As Figure 1 shows, RFC creates, manages, and manipulates multiple decision trees to mathematically compute the impact of different independent variables on the dependent variable. RFC allows for the user to modify how many trees appear in the model, which can have a positive or negative influence on the model accuracy. Each tree generates a predictive classification based on the random decisions from the inequalities. The predictive answer which occurs the most will be the answer generated by the random forest. As Figure 1 shows, Class-B will be the correct answer predicted by the RFC because it occurs 2 times while Class-A only occurs once.

Independent variables (factors) and Dependent Variables

To correctly generate a predictive computational model, the RFC needs a set of independent and dependent variables. The independent variable used were determined to be important pipeline factors which would correctly define the dependent variables. The dependent variables for the two models are the Risk Probability (RP) and the Failure Impact (FI). The independent variables used to estimate the RP and the FI are given in Table 1.

Table 1: Factors used in base model given to the ACUA

Factor	Description
Years Since Last Inspection	The number of years since an asset was last inspected or rehabilitated
Pipe Diameter	The diameter of a pipe in inches
Material Type	Material the pipeline is made of
Segment Length	Length of a pipe segment in feet
Original Installation	Year of original asset installation
Flow Type (G/F)	Flow designation of pipeline. These flows can either be a gravity flow or force flow
Up/down stream orientation	Relative placement of asset. The closer the asset is the higher the value of this factor.
Average Flowrate	Estimated flowrate of pipeline from pumpstation data. The data was taken over a three-year period.
Population Density	Average human population per square mile. Taken from US Census Data
Remaining life	Estimated remaining lifespan of pipeline

When the RFC is developed each factor is given a rating on "how important it is to the system". This rating is called the factor importance and is calculated for both the RP and FI by the ML model. After the ML models for both the RP and FI are completed, each asset is given a score from 1-5 and the factor importance is computed for each model. These scores are then multiplied together into a new variable called the Overall Preventative Measurement Number (OPMN), where the factor importance for both models are combined. This combination allows for the user to see how important each factor from this table is in generating the score of the OPMN.

Asset Predictor Tool

This tool can be accessed at any time in the application by selecting the button “Asset Predictor Tool”, in the upper right-hand corner. This tool allows the user to input real or theoretical asset data and see how it is rated using the ML algorithm.

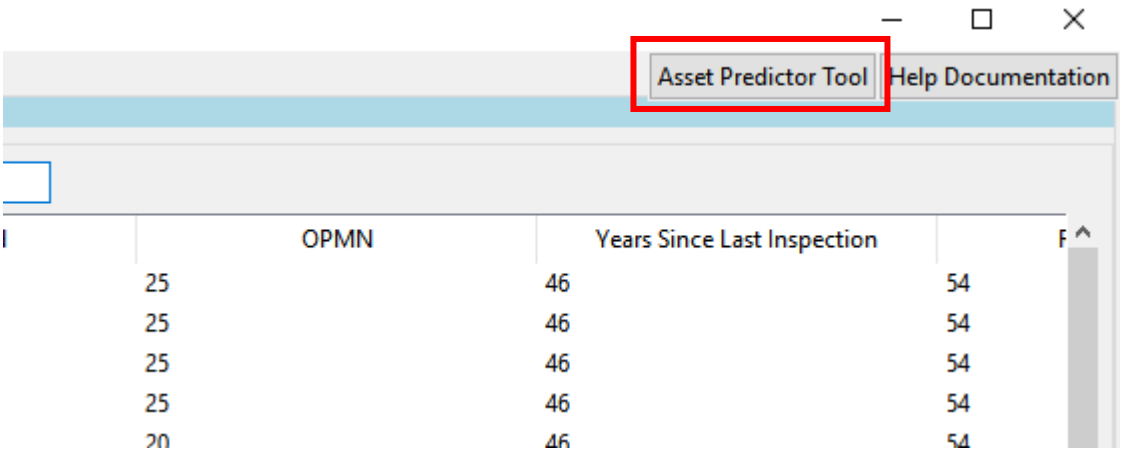

Figure 2: Button on Interface to use the Asset Predictor Tool

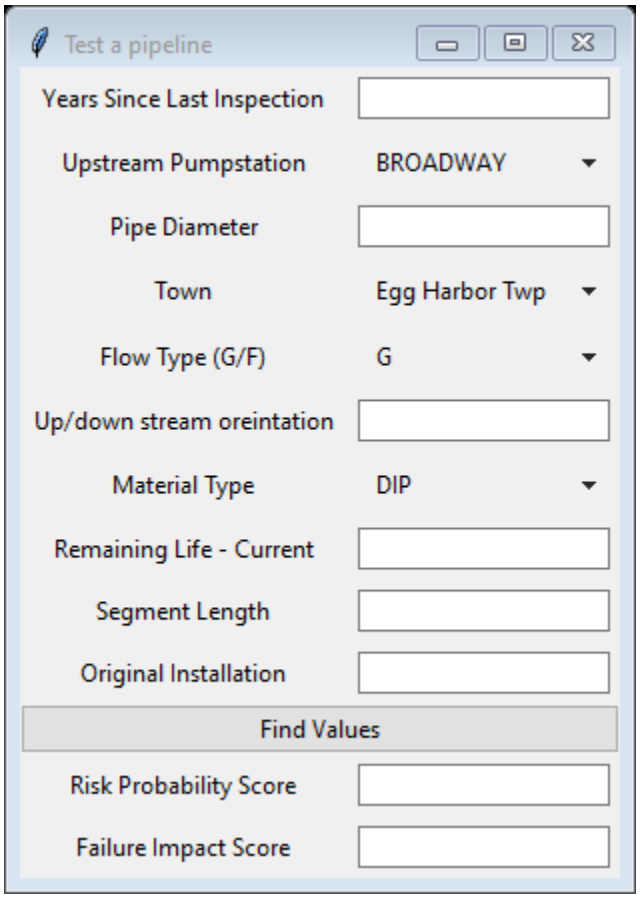

Figure 3: Asset predictor tool interface

Some errors which may arise & warnings to user:

1. Each textbox is not filled in
2. There are letters in one of the textboxes

To fix error 1:

1. Make sure each textbox is filled in except for the boxes under the “Find Values” button
2. If you do not know what number to put into Up/down stream orientation
    a. Find a similar pipeline in the “Pipeline Database” tab
    b. Each pipeline in the database is assigned a value in the table under the column “Up/ down stream” and use the value that is there

To fix error 2:

1. Check each textbox carefully and make sure there are no letters or words in the textbox

Pipelines at Most Risk Tab

The purpose of this tab is to display the pipelines scoring the highest on the risk assessment chart. The pipelines are automatically sorted in descending order. The corresponding sections will go over some of the functions in this tab:

Enter How Many Assets you would like to see:

This function allows the user to modify how many high-risk assets they would like to see in the table on the first tab "Pipelines at Most Risk".

Follow these general guidelines to use this function:

1. Locate the boxed area found in Figure 4
2. Inside the box you will find a textbox which would be automatically populated with the value '10'
3. Click on this value and put in desired number
4. This function is bound to your enter key, therefore, to use it simply press enter
5. When you press enter the corresponding table below the boxed area found in Figure 4 should change to the desired number

Some errors which may arise & warnings to user:

1. Make sure you are putting in a number value, no letters or words will work here
2. If you are pressing enter and the table is not changing, try:
    a. Scrolling down the table to verify the amount is not changing
    b. Clicking back on the application and pressing enter. If the application is not selected, then the function will not work
3. Make sure your value does not exceed the number of assets that is currently being used

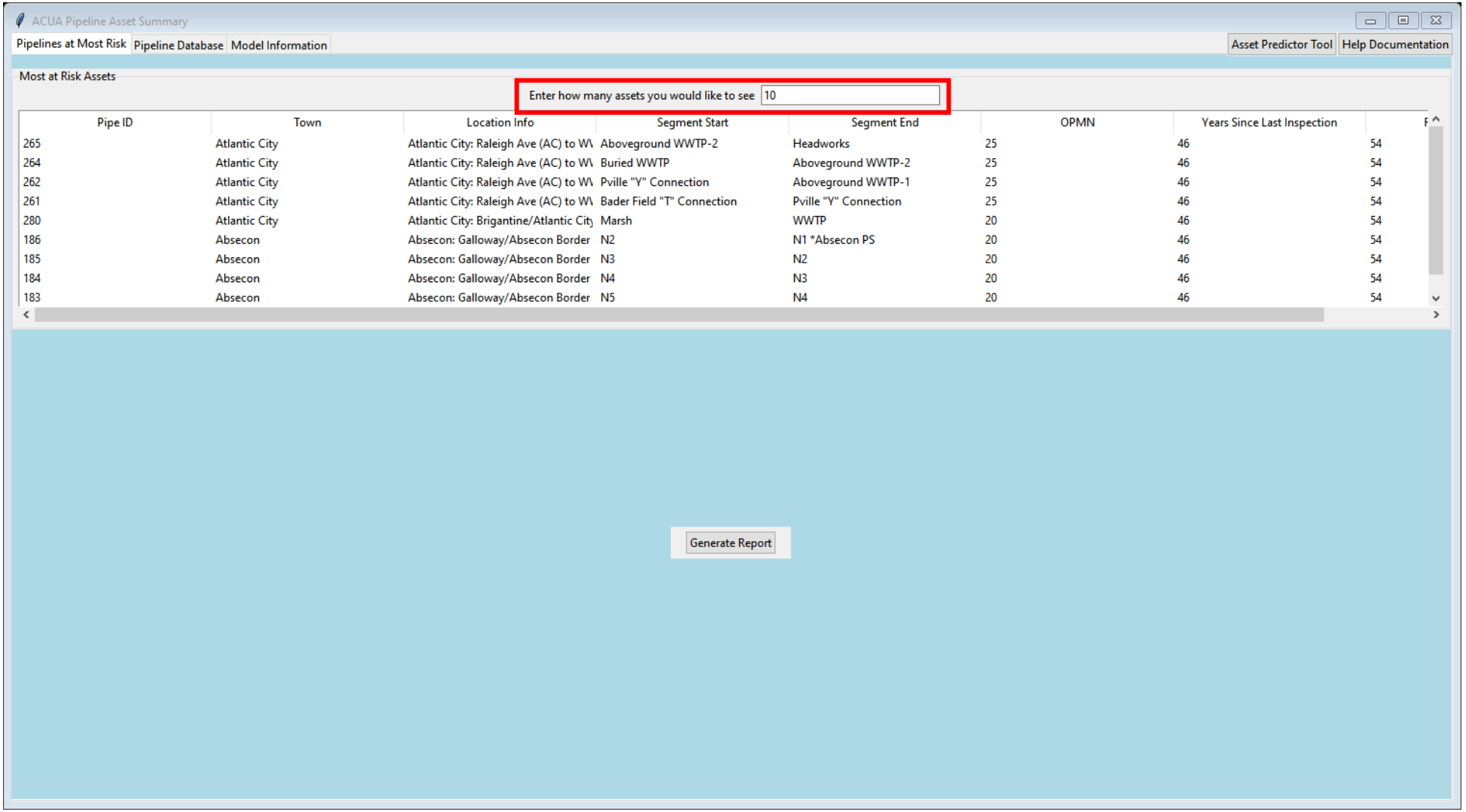

Figure 4: "Enter How many assets you would like to see" function

Generate Report Function

This function allows the user to generate an excel report and access the data through excel. This allows the user to generate data in such a way that is easy to format, change, and send if needed.

Follow these general guidelines to use this function:

1. Locate the boxed area found in Figure 5
2. Click on the button “Generate Report”
3. This should make an excel file which the user can use with some useful information
4. The excel file should open with the name “Report_ACUA.xlsx”

Some errors which may arise & warnings to user:

1. Cannot find file:
    a. The button should automatically open the file for the user
    b. If you cannot find the report it sometimes saves where the application is running. Go to the folder where you saved the application file and the report may be there.
    c. If you are trying to find old files, the report button will overwrite old files so try to save old reports under different names
2. The report won’t open
    a. If you are getting an error message, see if there are any other excel files with the same name “Report_ACUA.xlsx”. The report will not open if you have a previous version open.

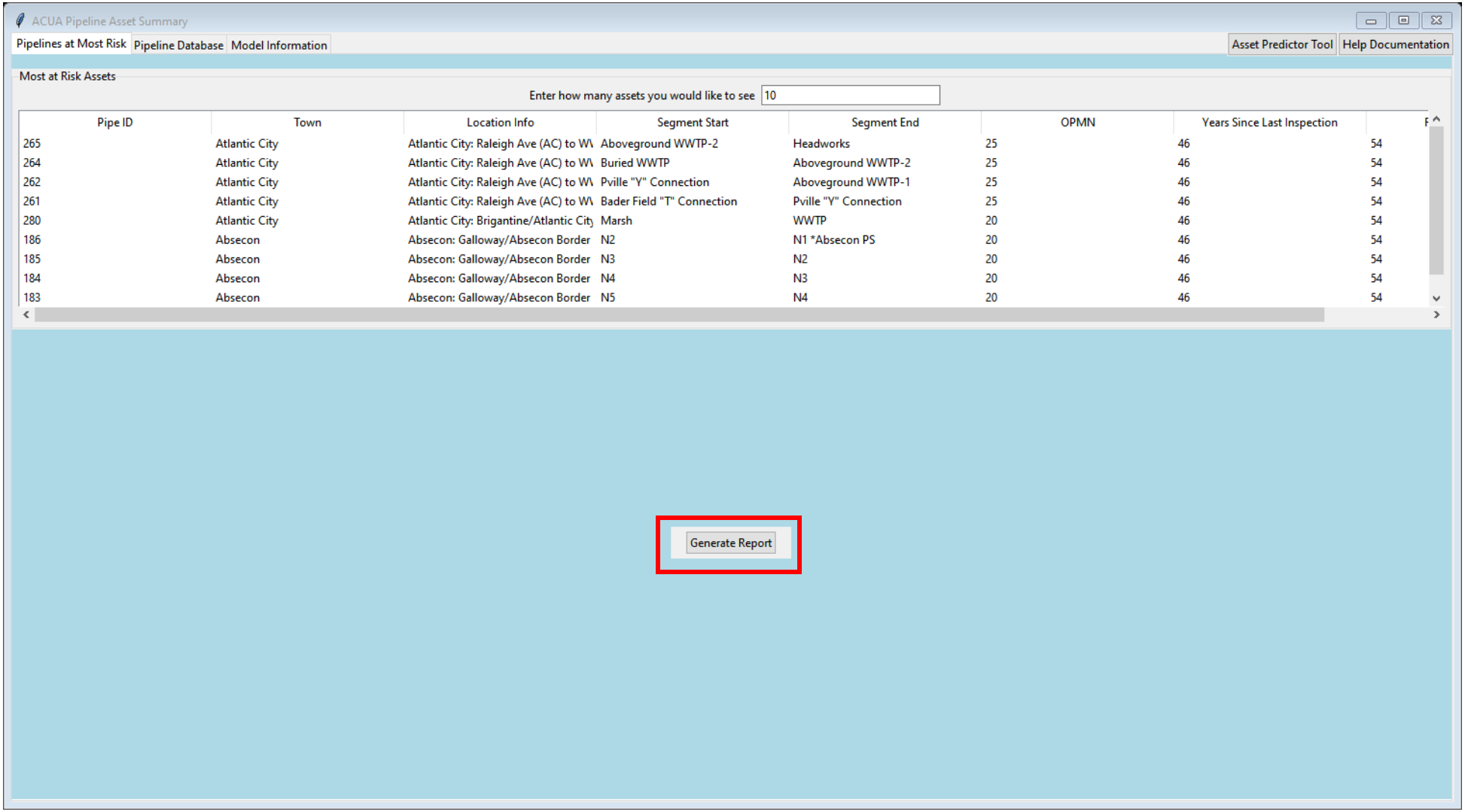

Figure 5: "Generate Report" function

Pipeline Database Tab

This tab allows you to view and search through the entire pipeline database in the system, as well as manipulating the database information from within the application. The bottom half of the tab has additional features when viewing asset data such as asset scores and factor assessments. The corresponding sections will go over some of the functions in this tab:

Database Search Functionality

This function allows the user to search through the database for all the information they are concerned about. At the time of writing this document the search functionality can only handle one search term.

Follow these general guidelines to use this function:

1. Select a property to search from the drop-down menu in the top left of the tab indicated by the red arrow in Figure 6.
2. Type a search term into the textbox adjacent to the drop-down menu indicated by the blue arrow in Figure 6.
3. Click the "Search" button indicated by the green arrow in Figure 6.
4. Use scrollbars on the central view to navigate all the results.
5. The view can be reset to normal but clicking the "Reset" button indicated by the black arrow in Figure 6.
    a. The view does not need to be reset to filter another search.
6. Searching nothing will also reset the central view

Some errors which may arise & warnings to user:

1. The search functionality does not support searching multiple terms. If you are trying to search for multiple terms, please only search one
2. The search functionality does not support inequalities. You cannot search for pipelines that have a diameter greater than 20 you can only search for one number at a time.
3. Make sure the search term in the textbox is applicable to the property you are searching over. Try not to enter in a number when searching for a town or location.
4. If you get an error message:
    a. Make sure your search word is applicable to the column you are searching in
    b. Make sure you have selected the right term in the drop-down menu

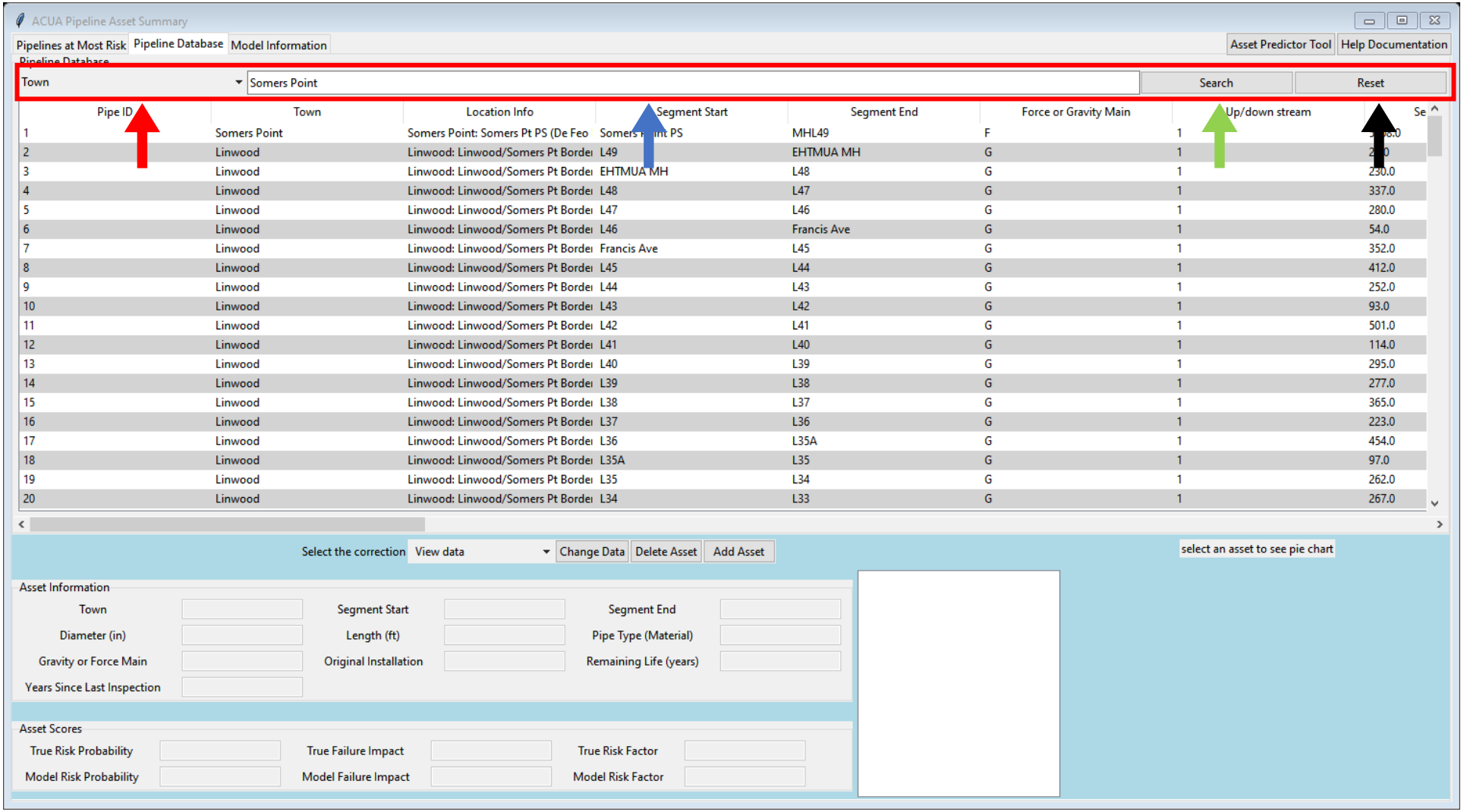

Figure 6: "Database Search" function

Database Selection Functionality

This function allows for the user to look at all the relevant information for individual pipeline segments. This function also gives a breakdown of the factor importance in relation to the OPMN as a pie chart.

Follow these general guidelines to use this function:

1. Make sure you are in the “Pipeline Database” Tab
2. Select on any row of the central view to see the corresponding information
3. The selection will populate the bottom view with all the asset information and the pie chart on the factor importance
4. Next to the pie chart there is a table titled “Factors in Other Category”
    a. This table shows all the factors that are group in the “Other” Category in the Pipeline Factor Assessment Pie chart
    b. Any factor that is less than 5% is put into this “Other” category
    c. The corresponding value of the category is the summation of the percentages of all factors in the table
    d. Figure 7 has a value of 7.7% in the other category. Therefore Up/down stream orientation% + Material Type% + Original Installation% = 7.7%
5. This selection can be done when the database is filtered also

Some errors which may arise & warnings to user:

1. Make sure you select the asset you wish to view
    a. If you correctly select the asset it should light up blue like in Figure 7

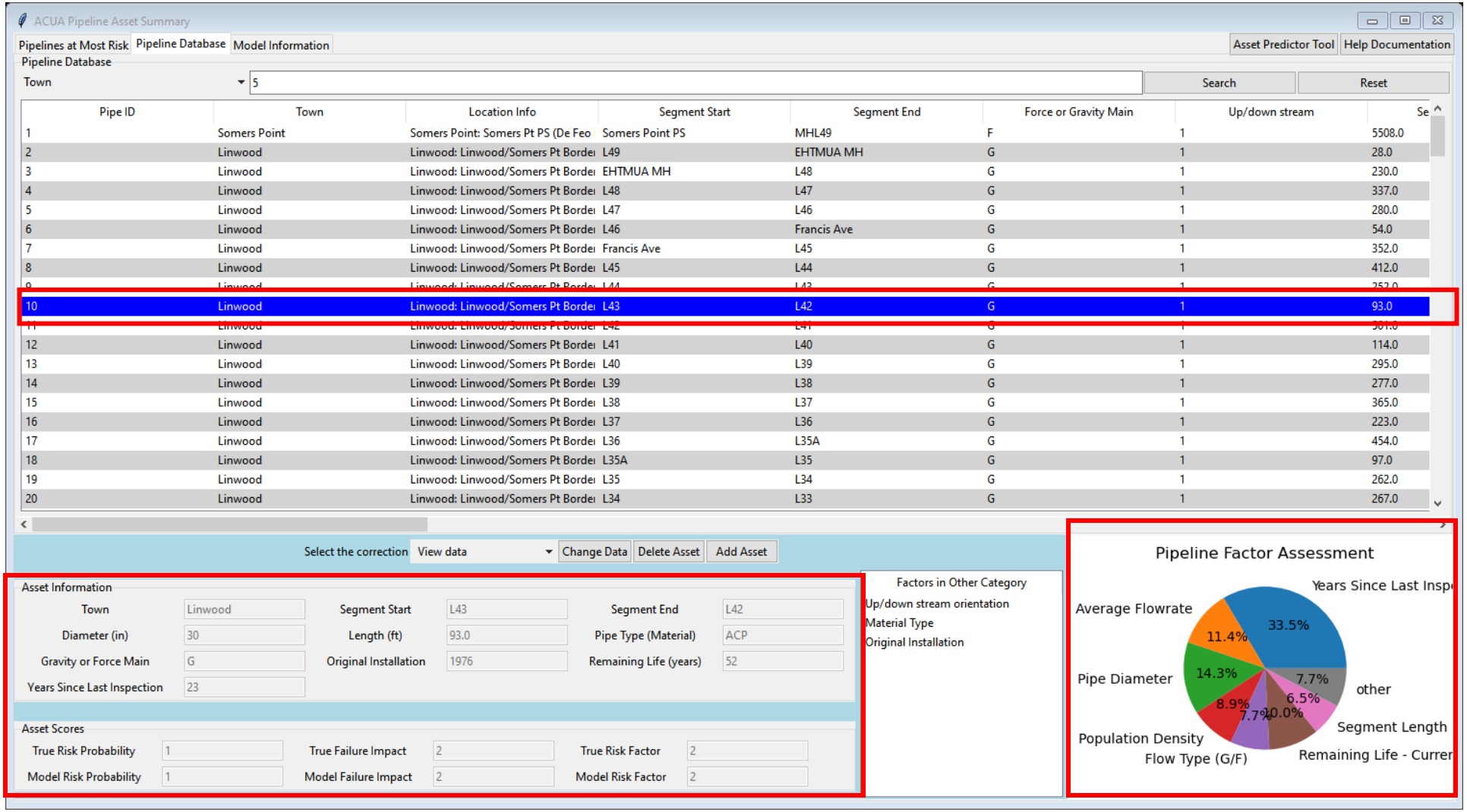

Figure 7: "Database Selection" function

Changing Asset Data

This function allows for the user to change the asset data and update it. This communities between the application and the database so it is always up-to-date with the correct information.

Follow these general guidelines to use this function:

1. Make sure you are in the “Pipeline Database” Tab
2. Select on any row of the central view to see the corresponding information
3. The selection will populate the bottom view with all the asset information and the pie chart on the factor importance.
4. Once an asset is selected use the dropdown menu to select the change you would like to implement
    a. The drop-down is in the red box in Figure 8
5. Once you have selected the drop-down menu option, the corresponding textbox should be editable
    a. The drop-down in Figure 8 is set to a diameter change and the corresponding text box for the diameter is white. (Shown in a blue outline)
6. After entering in the correct information select “Change Data” to modify the data in the database

Some errors which may arise & warnings to user:

1. If you do not select an asset, you will not be able to change the data. Make sure you have selected an asset in the central view. Selected assets should light up blue.
2. At the time of writing this document, this functionality only allows for the modification of one asset at a time. Modifying multiple assets at once is nots supported in this application.
3. The last selection in the drop-down “Alter Pipeline Data” allows the user to alter any of the databoxes in the bottom of the screen. Altering any other information outside of the textboxes is not supported at the time of writing this document.

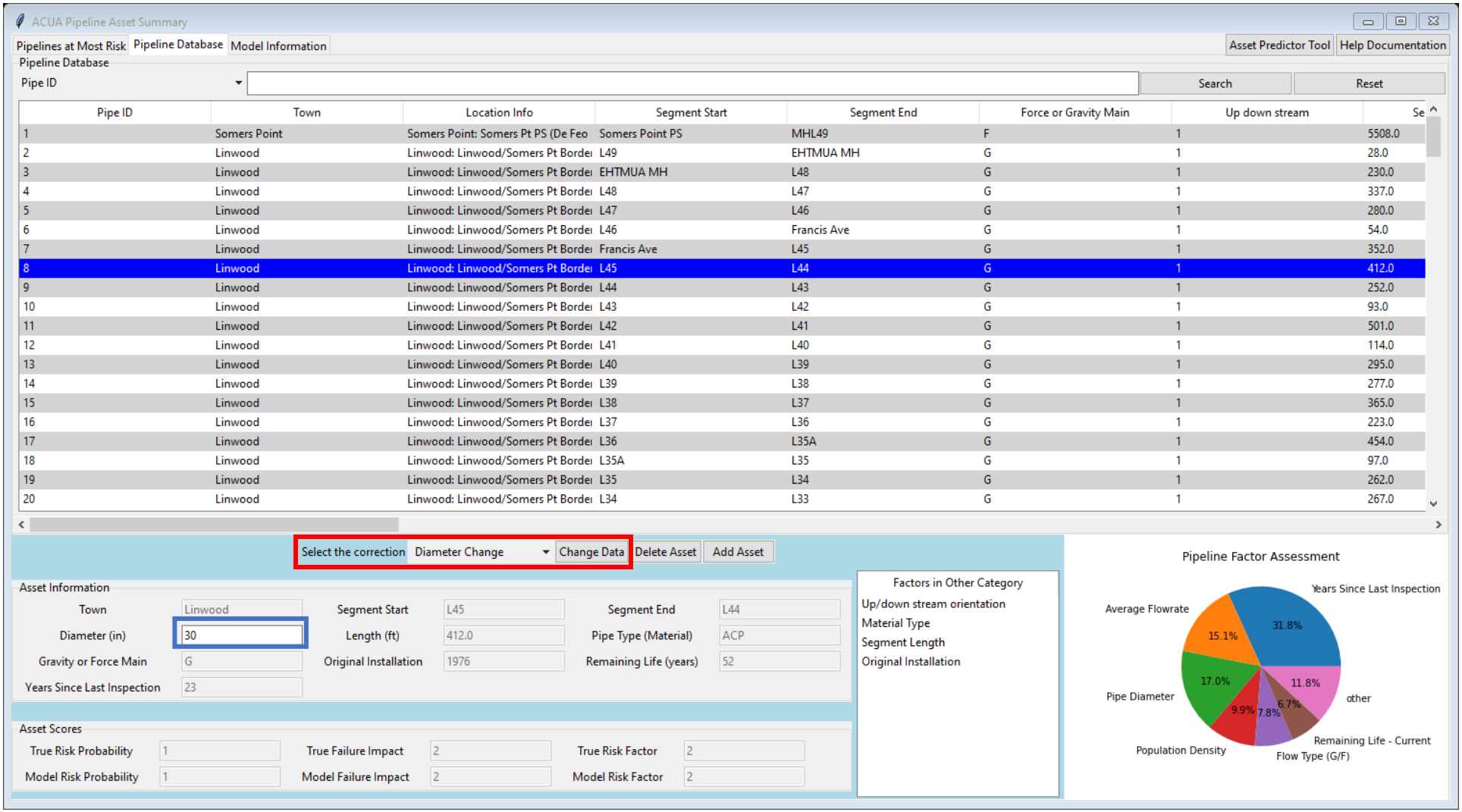

Figure 8: Changing Asset information function

Deleting Asset Data

This function allows for the user to delete any asset which is not in use or is no longer being monitored.

Follow these general guidelines to use this function:

1. Make sure you are in the “Pipeline Database” Tab
2. Select on any row of the central view to see the corresponding information
3. The selection will populate the bottom view with all the asset information and the pie chart on the factor importance.
4. Once you see the selection you would like to delete, press the “Delete Asset” button.
    a. This button is outlined in red in Figure 9
5. Once pressed another window will come up verifying you would like to delete the selected asset
6. When you press yes, the asset will be purged from the database and the application
    a. If you press no, the window will close, and the asset will remain in the database

Some errors which may arise & warnings to user:

1. When deleting an asset, a selection needs to be made in the application
    a. Make sure that the asset you wish to delete is highlighted in blue if you are having difficulty deleting the asset

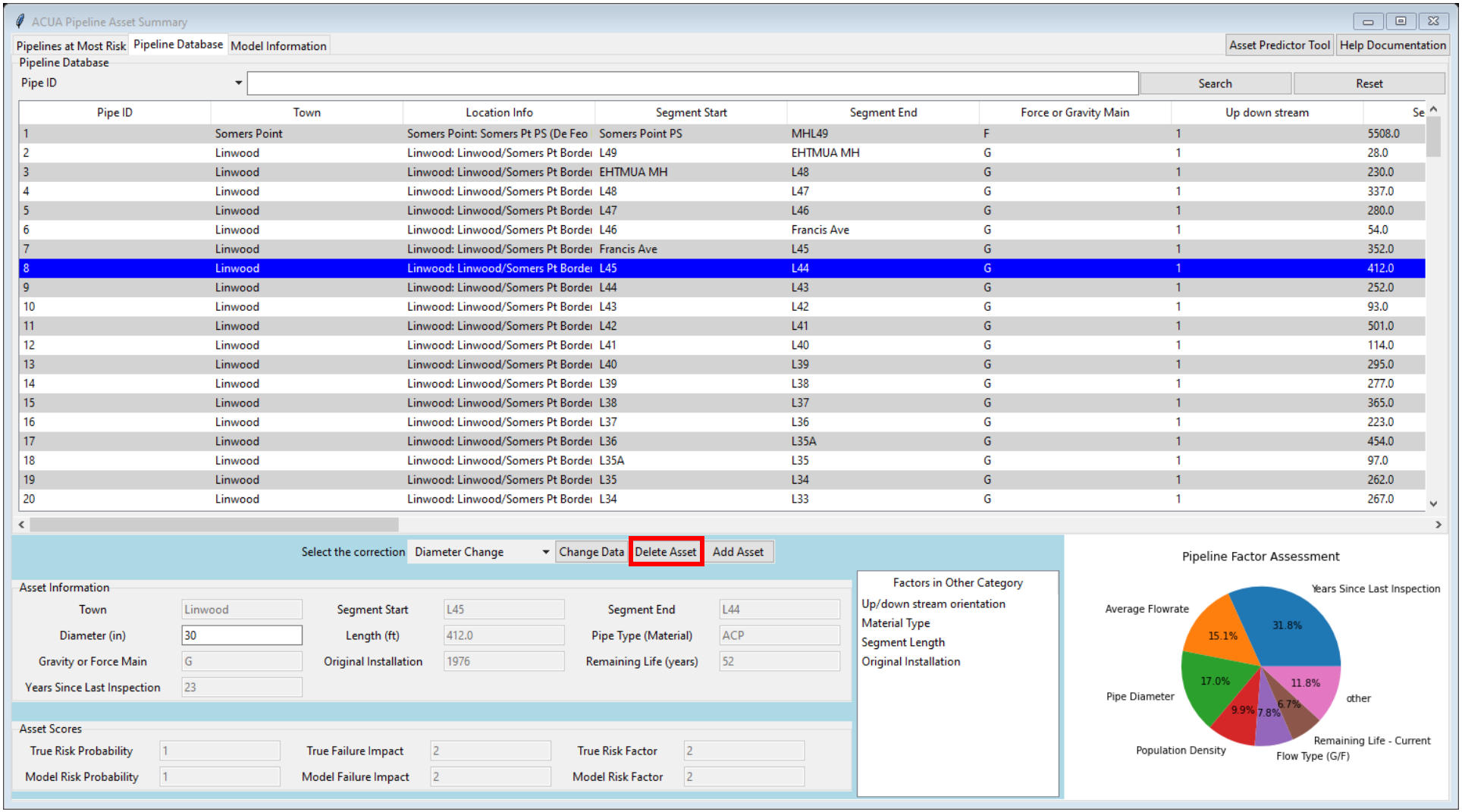

Figure 9: Deleting Asset function

Adding New Asset Data

If the ACUA was to install pipelines in the future, this feature allows for the company to correctly add in the necessary data. This will add it to the database in the correct way so that it can be used in the machine learning model.

Follow these general guidelines to use this function:

1. Make sure you are in the "Pipeline Database" Tab
2. Go to the button "Add Asset"
    a. This button is outlined in red in Figure 10
3. Once pressed Figure 11 will appear on your screen
4. Populate every textbox with the relevant information
    a. Make sure to select the correct selection from the drop-down menus
5. Once everything is filled out press the "Add Pipeline to Database" button and the pipeline will be put into the database and start to be used in the model.

Some errors which may arise & warnings to user:

1. Make sure you are putting in the correct type of information in each textbox
    a. Textboxes that should have words:
        i. Segment Start
        ii. Segment End
        iii. Location information of Pipeline
        iv. Rehabilitation Type
        v. Pipeline notes
    b. If the textbox is not in the above list, then it should have a numerical value assigned to it.
2. The "Pipeline Notes" section allows for the user to put important information about the pipeline which does not go in any other section
3. If you do not see the new pipeline in the database, then try to follow the guidance from step 1. The pipeline may not populate the database if the wrong information is put in.

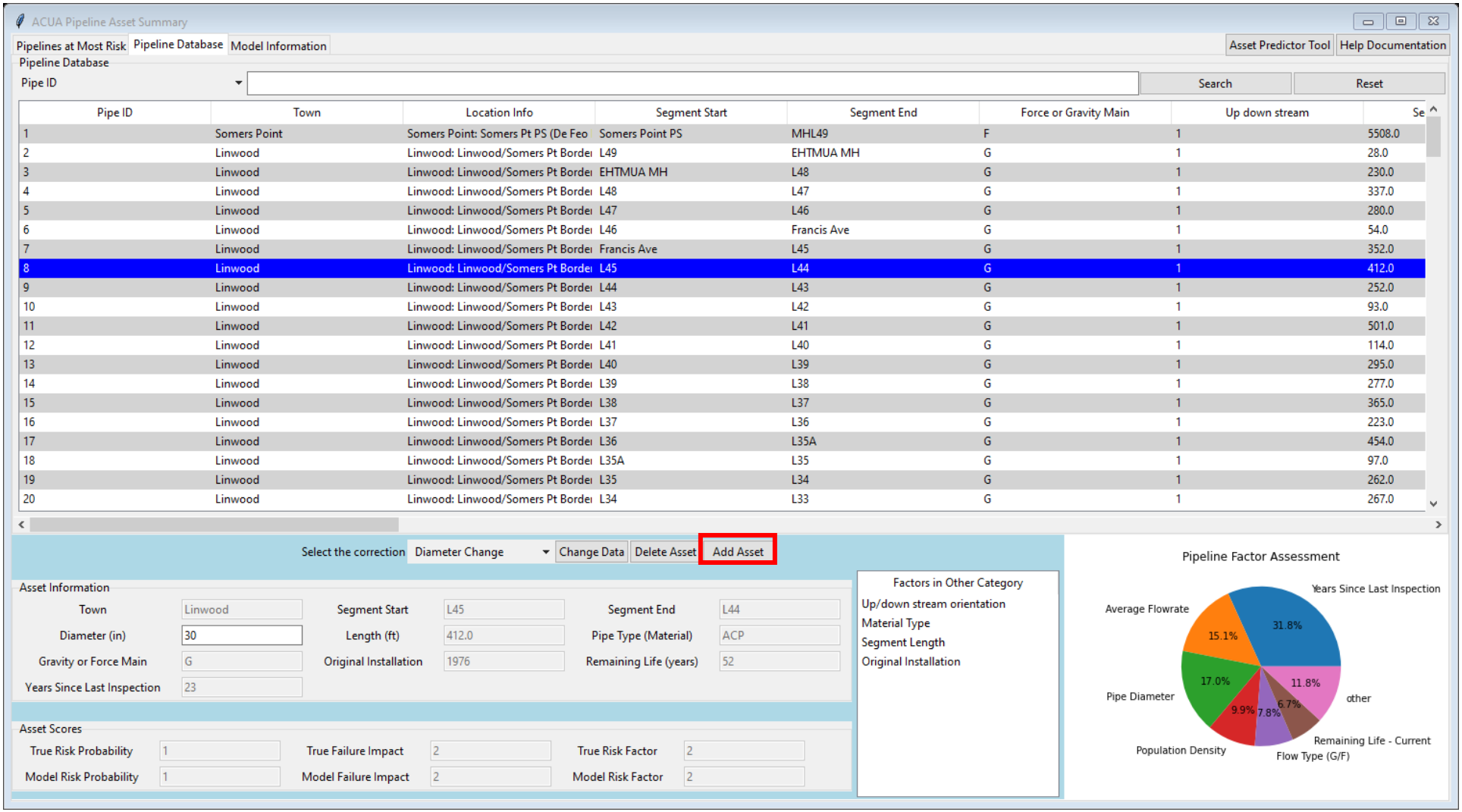

Figure 10: Adding Asset function

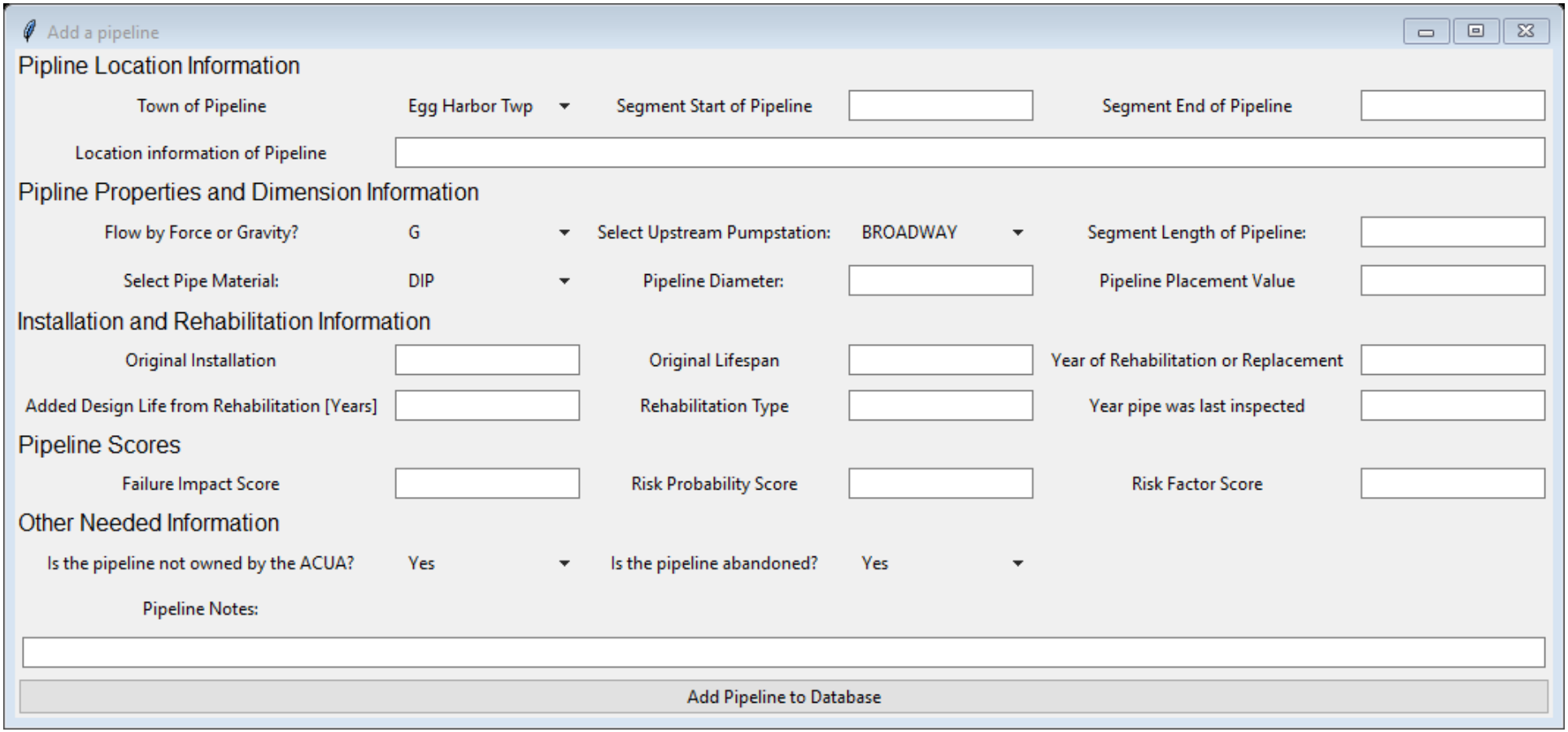

Figure 11: Add a pipeline window

Model Information Tab

This tab allows for the user to view all the information corresponding to the ML models for both RP and FI. This tab allows the user to view what factors are used and the accuracy of the model. This tab allows for the model to be changed if needed. The corresponding sections will go over some of the functions in this tab:

Viewing Base Model Data

This functionality allows for the user to see the base model which comes with the application and rates the RP and FI in the database.

Follow these general guidelines to use this function:

1. Navigate to the third tab "Model information"
2. The red area in Figure 12 shows where the user can see all the base model information
3. The default figure will be the feature importance's

Some errors which may arise & warnings to user:

1. This model cannot be modified without using other functions.
2. The purpose is to inform the user what is currently being used in the application
3. If the user wants to change the information in the model, test the new model first using the other functions of this tab.
4. The Changing the Machine Learning Random Forest Classification Model section goes over how to change the information in the model

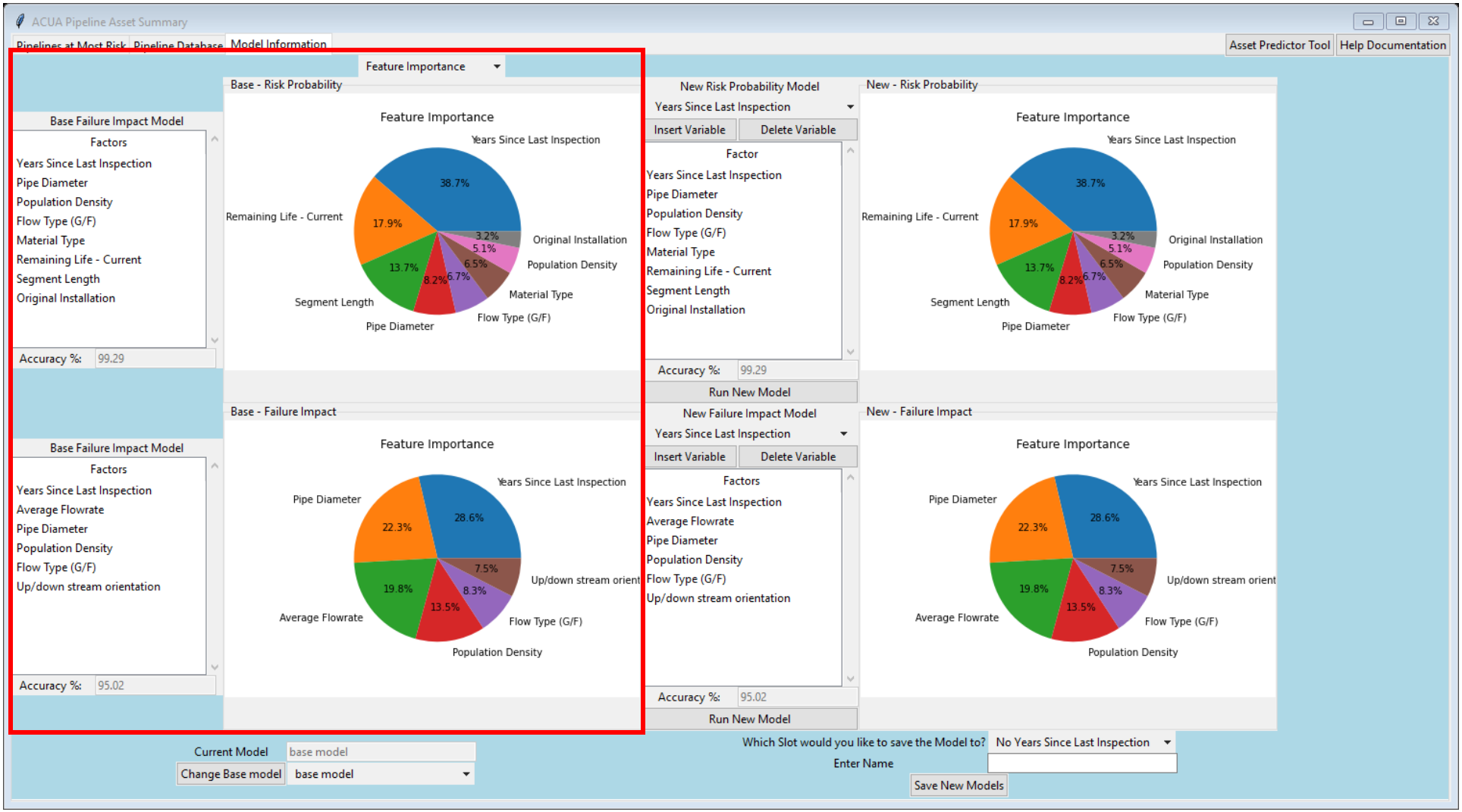

Figure 12: "Base Model" functionality

Changing the Machine Learning Random Forest Classification Model

This allows for the user to modify the machine learning model at any time they see fit. There is a side-by-side comparison with the current model being used so the user can see how much changes when modifications are made to the model.

Follow these general guidelines to use this function:

1. Navigate to the "Model Information" tab
2. Here you will see two sections
    a. The base model section on the left (shown in a red outline in Figure 13)
    b. The new model section on the right (shown in green and blue outlines in Figure 13)
3. The new model section will be populated with the base model by default; therefore, the images should always start out the same
4. By using the buttons in the green outline in Figure 13 the model can be changed
5. The "Insert Variable" button allows you to insert the selected factor from the drop-down menu and include it in the model
6. The "Delete Variable" button allows you to select a variable from the list below it and delete the factor
7. Once the list of factors are to your liking, press the run new model button to run either the new risk probability model or the new failure impact model.
    a. Risk probability is found on top
    b. Failure Impact is found on bottom
8. The corresponding figure in the blue outline in Figure 13 will change to reflect the new model as well as the accuracy textbox below.

Some errors which may arise & warnings to user:

1. When trying to delete a factor make sure you select a value from the table then press delete.
2. If you have two of the same factors, it will only show up once in the pie chart.

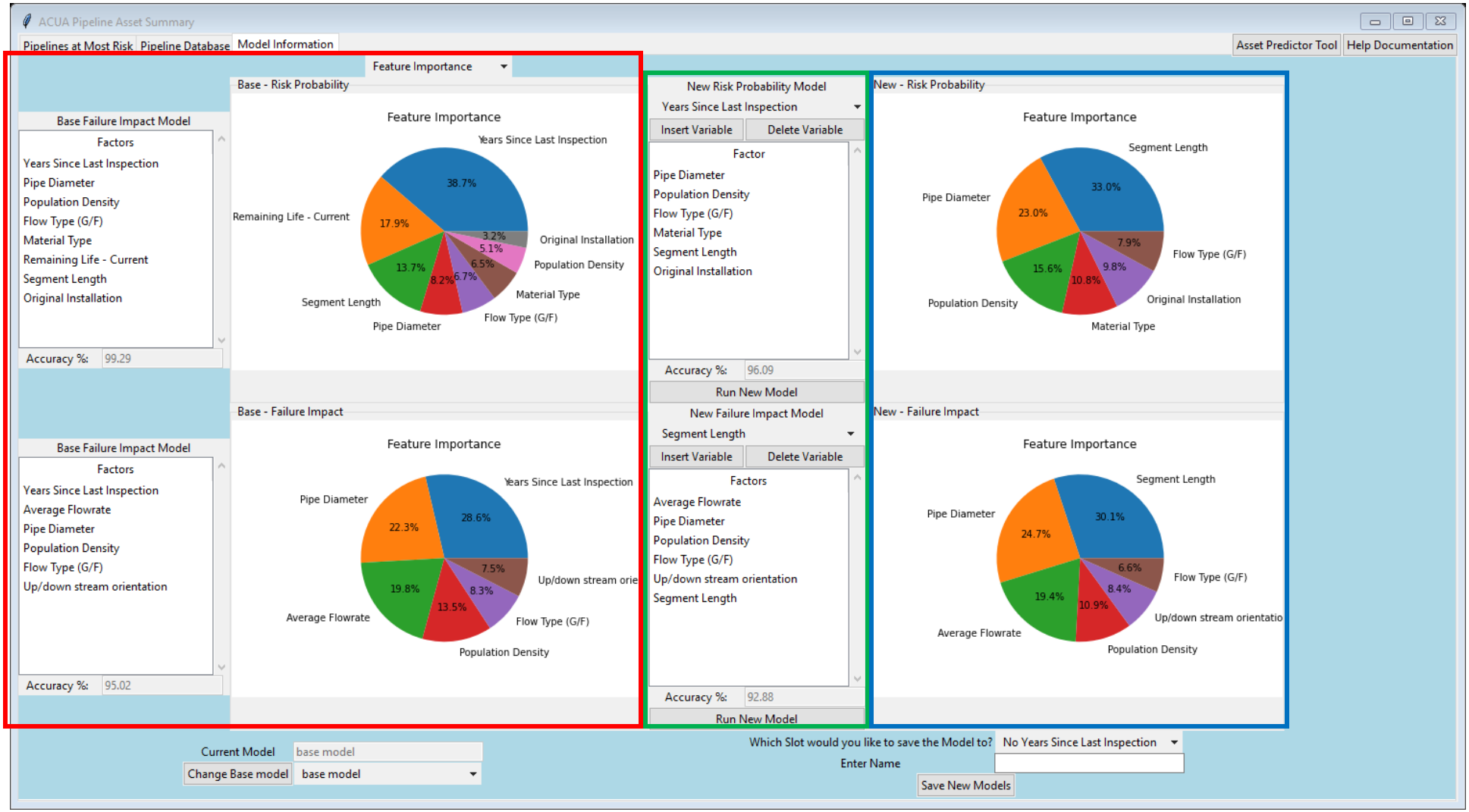

Figure 13: Changing the Machine Learning model

Viewing the Feature Importance and Confusion Matrices

This feature allows the user to assess the accuracy and the factors of each model. This is done through graphing the feature importance and the confusion matrix for each model that is computed.

Follow these general guidelines to use this function:

1. Navigate to the "Model Information" tab
2. Click the drop down outlined in red in Figure 14
3. Here you will have two selections:
    a. Feature Importance
    b. Confusion Matrix
4. If "Feature Importance" is selected, then four pie charts will populate the application
5. If "Confusion Matrix" is selected, then four confusion matrices will populate the application

Some errors which may arise & warnings to user:

1. A confusion matrix is an analysis of the model accuracy
    a. The x-axis represents the **true** values given by the ACUA
    b. The y-axis represents the **predicted** values from the model
    c. By looking at Figure 15 there are two examples outlined in blue and green
        i. The blue box shows that the model predicted 7 assets to have a score of 4 while the true value shows it had a score of 1
        ii. The green box shows that the model predicted 0 assets to have a score of 4 while the true value shows it had a score of 1
2. The percentages shown by the pie chart in Figure 14 represent how important the factor is to the model, but does not represent an equation to determine the score

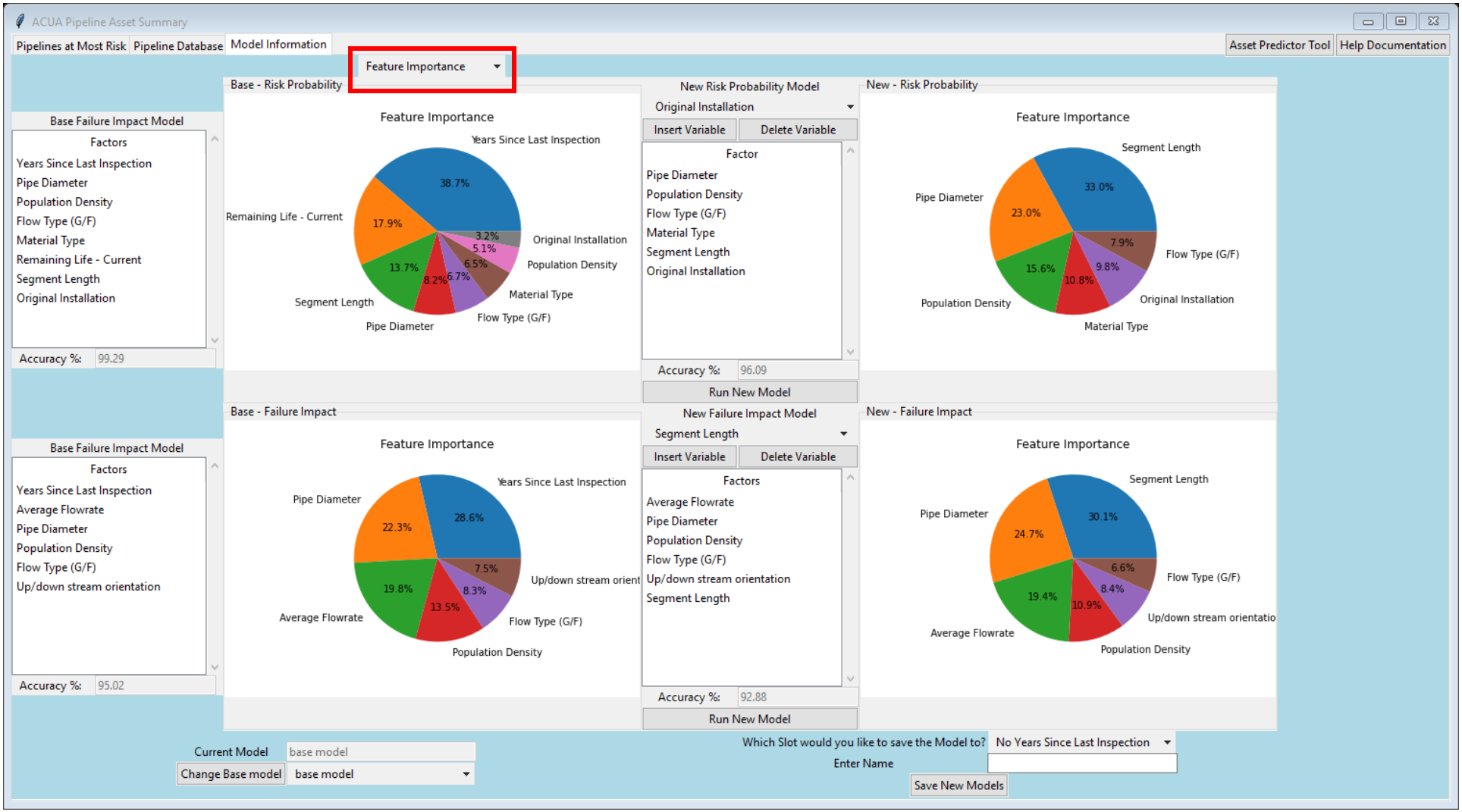

Figure 14: Showing the Feature importance function from the drop-down menu

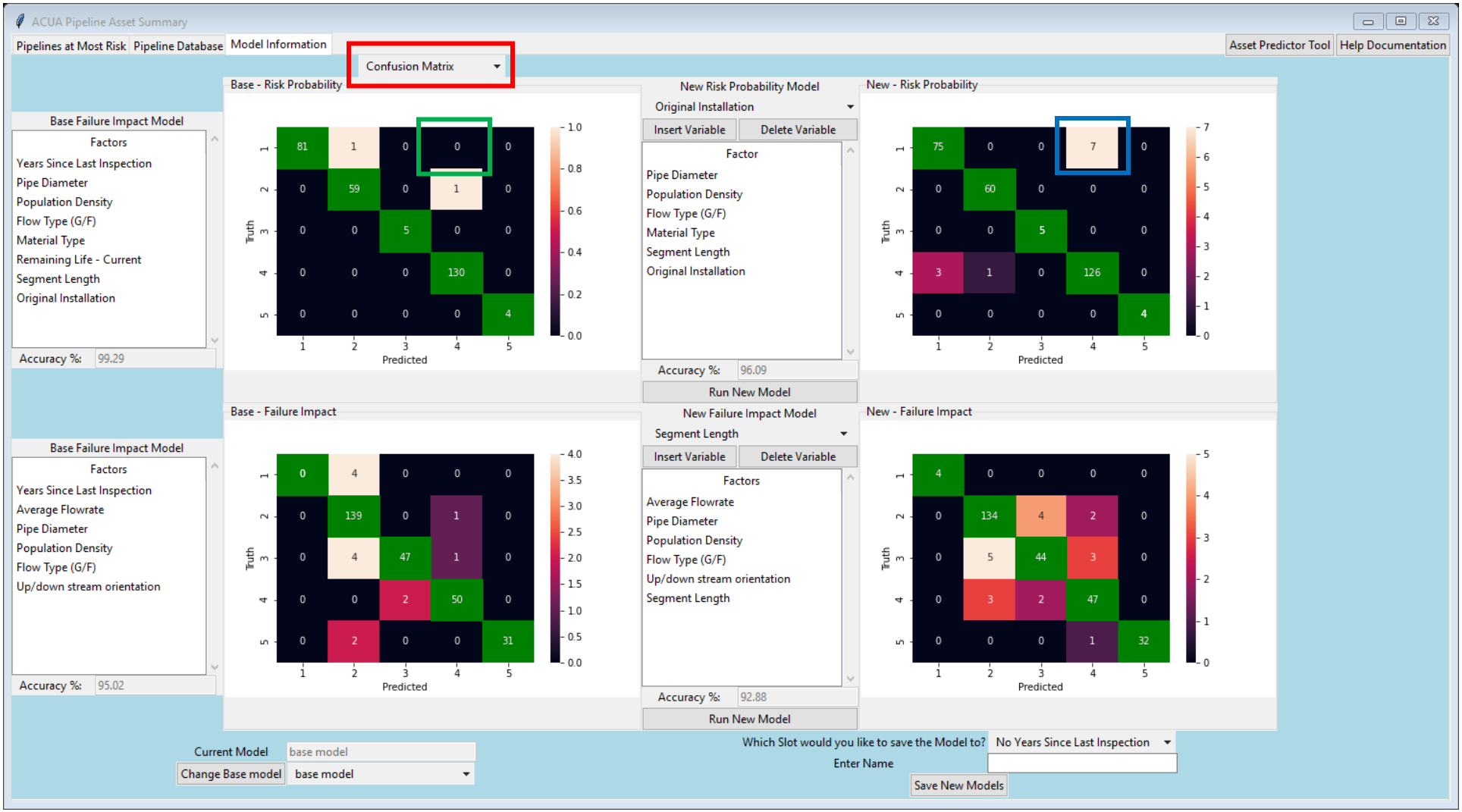

Figure 15: Showing the confusion matrix function from the drop-down menu

Changing the saved model

This function allows the user to filter through the saved models and choose which one they would like to run the application. This allows for flexibility while also keeping old models if they are needed at a later date.

Follow these general guidelines to use this function:

1. Navigate to the "Model Information" tab
2. The red outline in Figure 16 shows where the user can change the model to any of the saved models.
3. The textbox shows the current model which is being used by the whole application
    a. In the example, the model being used is the "base model"
4. In the drop-down menu below the textbox the user can select which model is to be used for the application
5. When the desired model is selected from the dropdown click the "Change Base model" button.
6. Once clicked this will change the model the application is using
7. Once changed the application will rerun and recalculate all the machine learning models and generate new answers.
    a. These answers may not change the scores of assets if the accuracies are very similar

Some errors which may arise & warnings to user:

1. There are five slots in the first version of the application. If the user selects a slot which is empty, then the application will give an error message
2. Make sure the model you are choosing has a loaded or saved model in it before trying to load it in as the base model.
3. The base model that was given will always be an option for the ACUA as the next feature does not allow for the base model to be overwritten.
    a. This allows for the ACUA to see the starting model whenever they need to

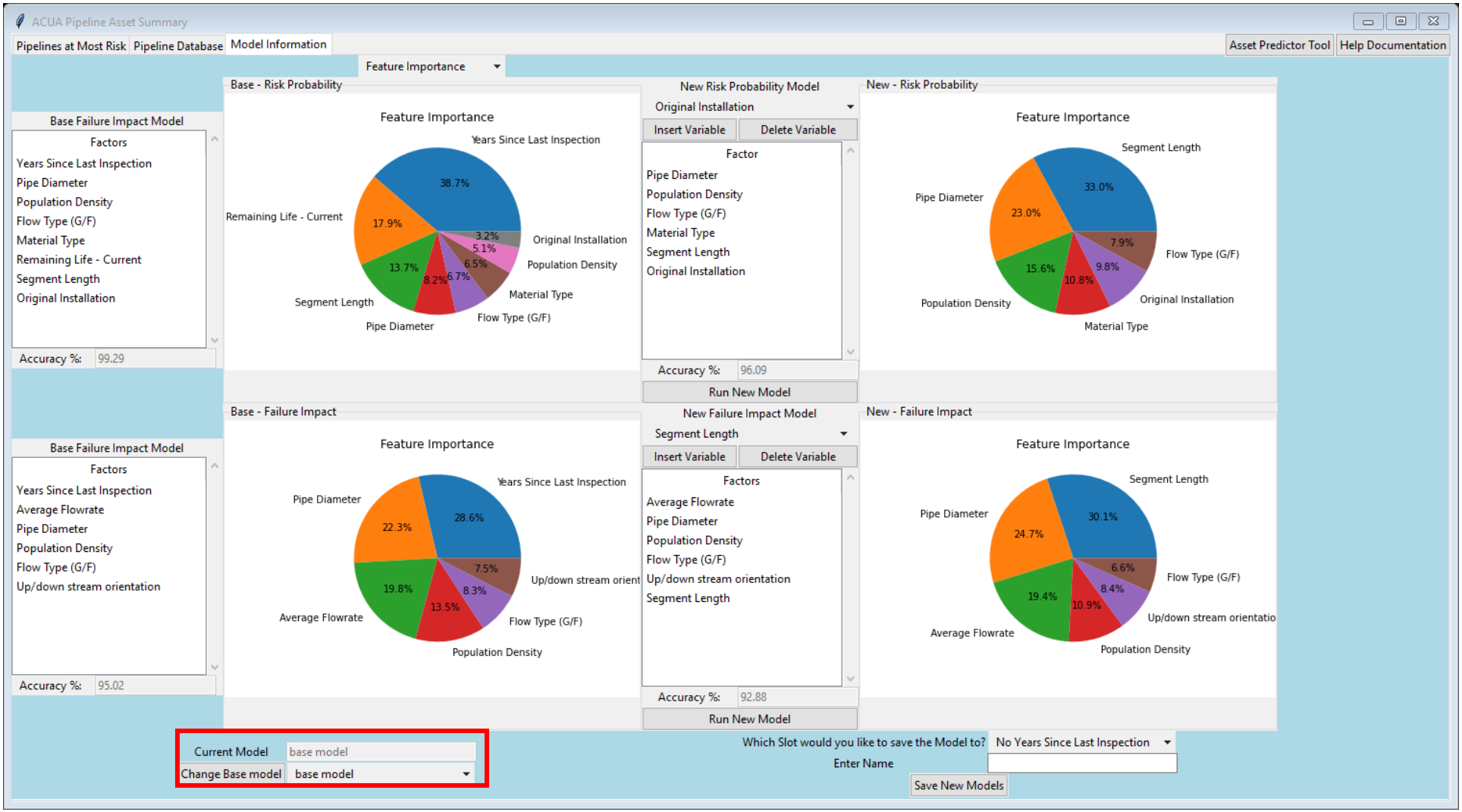

Figure 16: How to change the model, and displaying what model is currently being used

Saving a new model

This function allows for the ACUA to save up to 4 machine learning models within the application. These models can be changed and manipulated at any time.

Follow these general guidelines to use this function:

1. Navigate to the “Model Information” tab
2. The red outline in Figure 17 allows for the user to save the new models to any unused slot or overwrite any model besides the base model
3. First go to the drop-down menu in the red outline
4. In the drop-down select a slot which you would like to save a new model to
    a. If you do not run new models, you will not be able to save the models to a slot.
5. Once the desired slot is selected, enter the name you wish to call the model in the white textbox under the drop-down.
6. When all of this information is in, click the “Save New Models” button
7. This will make a pop-up asking if you are sure you would like to save this model
8. When you click yes, the model will be saved and can be used at any time for the application
    a. If you click no, the window will close, and you can continue modifying the model

Some errors which may arise & warnings to user:

1. A new model will not save if the user does not click the “Run New Model” Button for both the risk probability and failure impact
    a. If you still like the model for the risk probability or failure impact, simply keep the same factors and click the “run new model” button

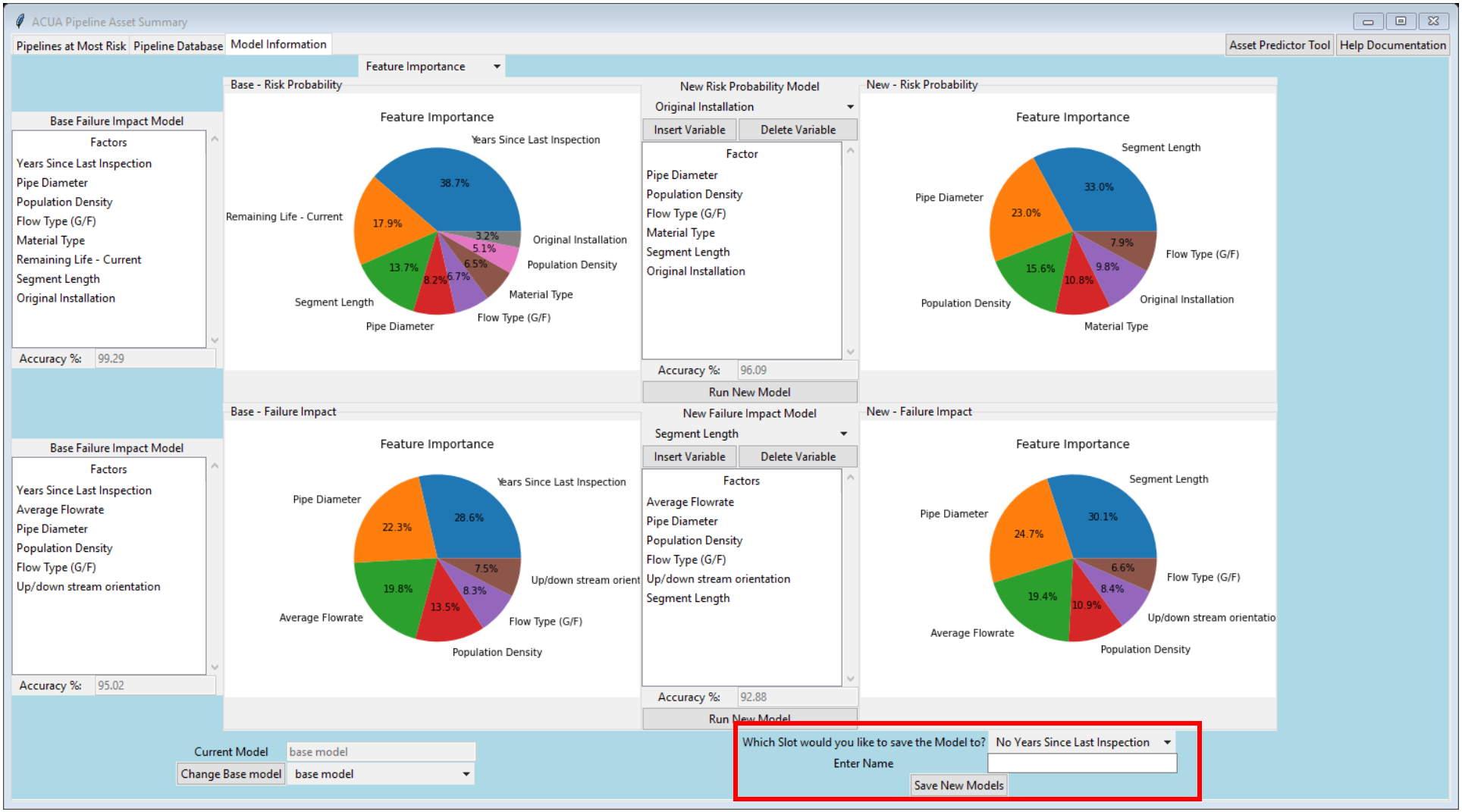

Figure 17: Showing the Feature importance function from the drop-down menu