



Predictive Toolkit

CA400 Final Year Project: User Manual

Predictive Analytics Toolkit

Sean Quinn 13330146

Supervisor: Mark Roantree

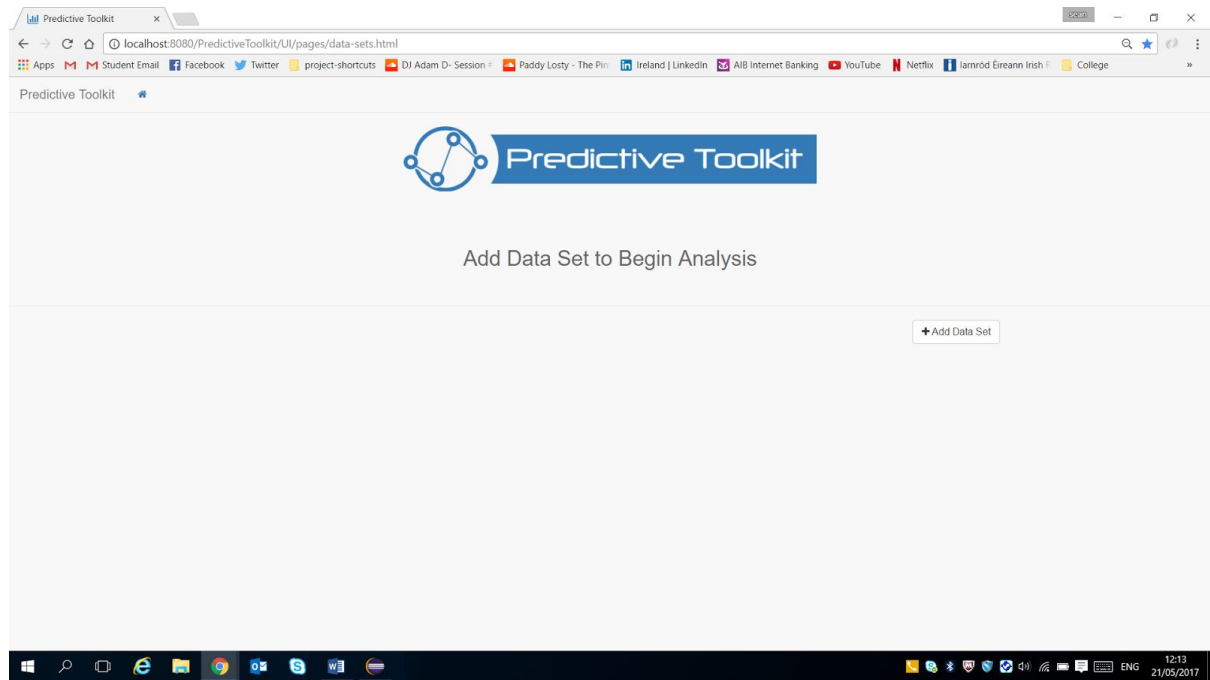
21/05/17

Application Summary

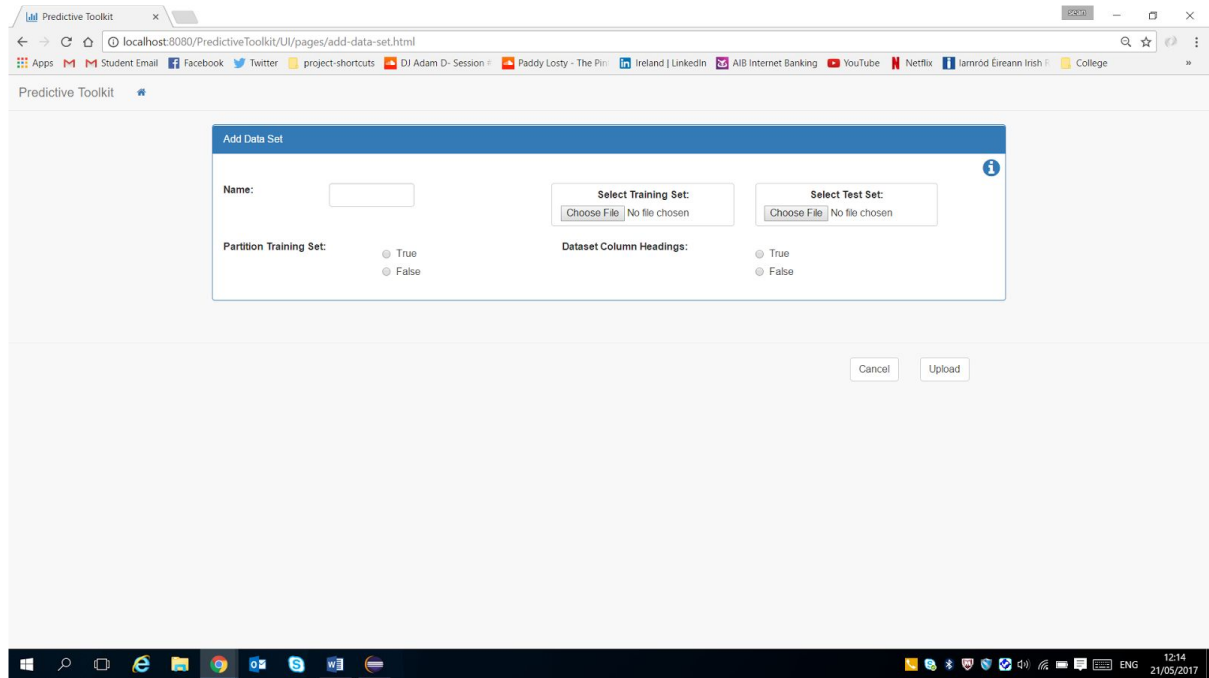
This application allows the user to upload csv format data sets and then configure and run several different types of predictive algorithms simultaneously on the data through a browser-based graphical user interface. This allows contrasting of the performance of the different algorithms and various configurations of their run parameters for any given dataset, principle component analysis and investigation of the nature of the data. The application features the K-Nearest Neighbours, Decision Tree, Naïve Bayes and Rocchio classification algorithms.

Uploading Datasets

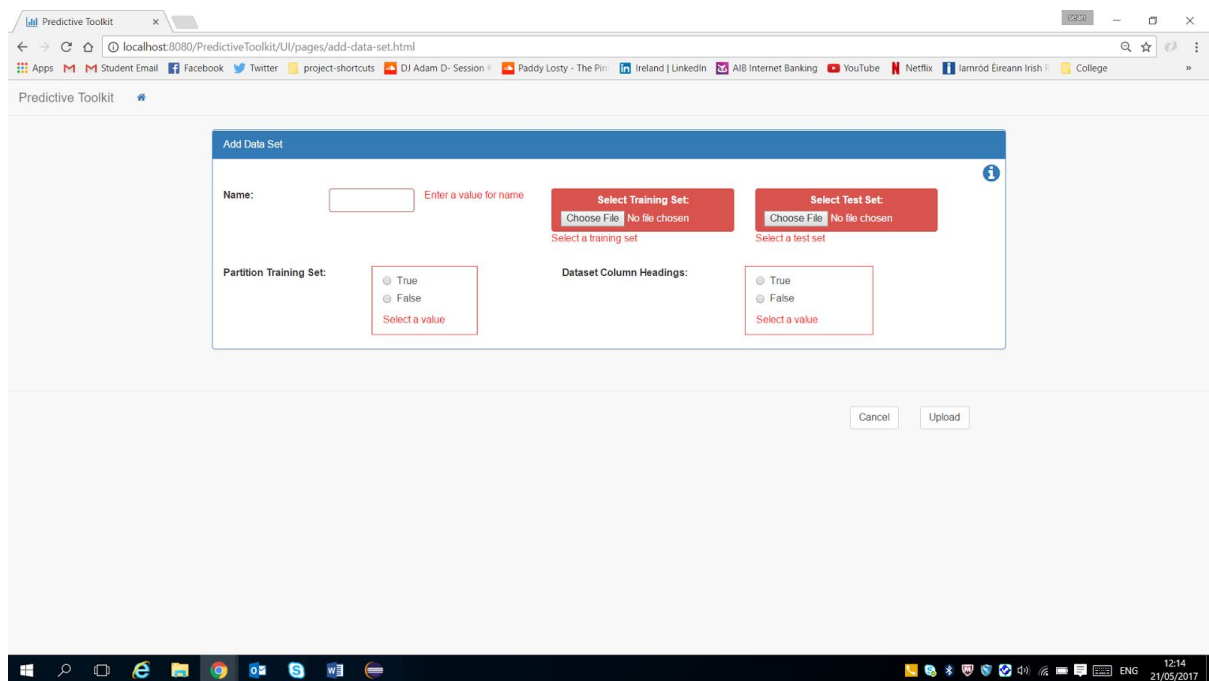
- To get started, start the application in eclipse and go to <http://localhost:8080/PredictiveToolkit/UI/pages/data-sets.html>
- You will be greeted with the following page. This is the home page of the application which will display all your datasets in a dashboard style.
- We have not uploaded anything yet, so for now the dashboard is empty.



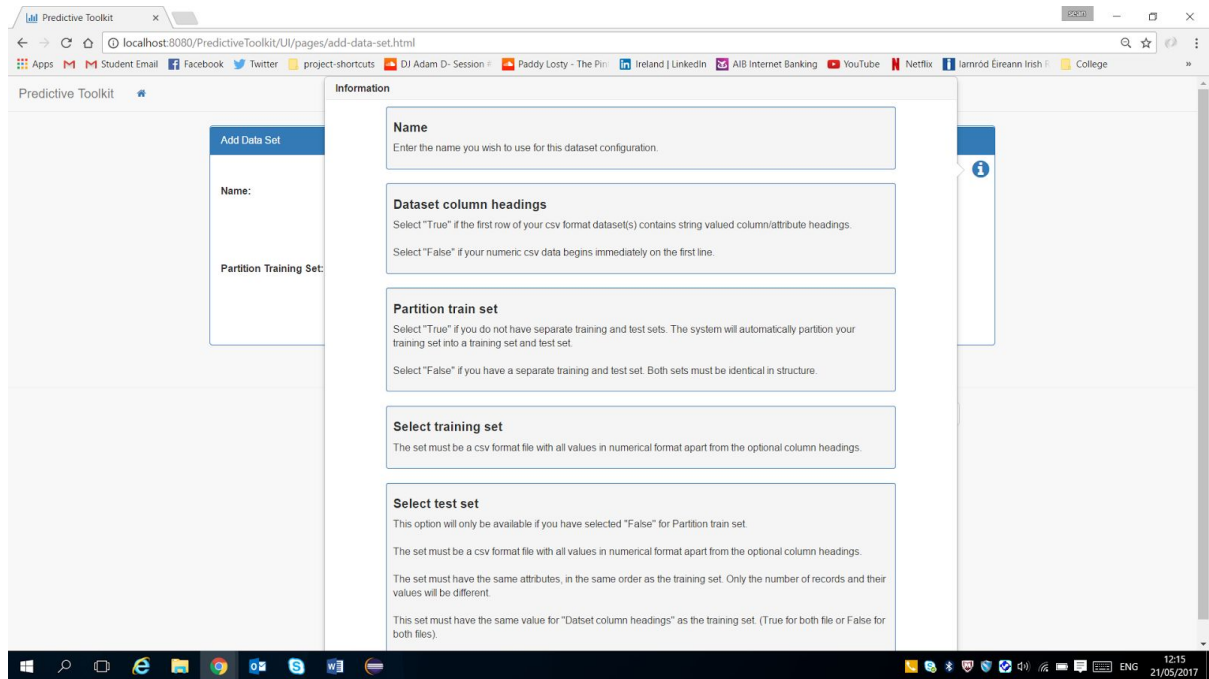
- When you select “Add Data Set” you will be brought to the following page.
- This page allows you to configure the parameters for the set you're uploading
 - Name - The alias you wish to refer to this dataset with.
 - Select Training Set - The CSV file you wish to use to build the classifiers on.
 - Select Test Set - The CSV file you wish to use to test the classifier with.
 - Partition Training Set - If You only have one CSV, selecting true for this option will automatically partition your training set into two.
 - Dataset Column Headings - If the first row of the csv a row of column headings select true.



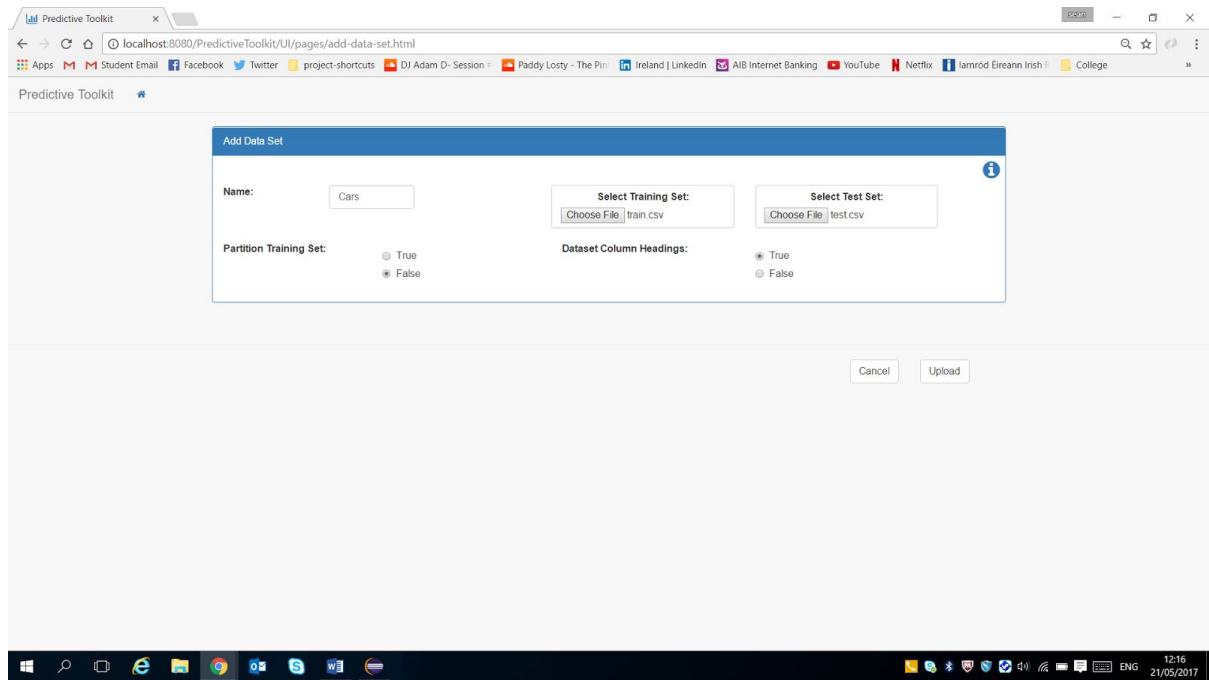
- If you forget to fill in one of the fields or fill them in incorrectly then the application will notify you by highlighting the problem field in red and displaying a text description of the problem when you click upload.
- You will not lose any of the information you have already entered and will be able to fix the error.



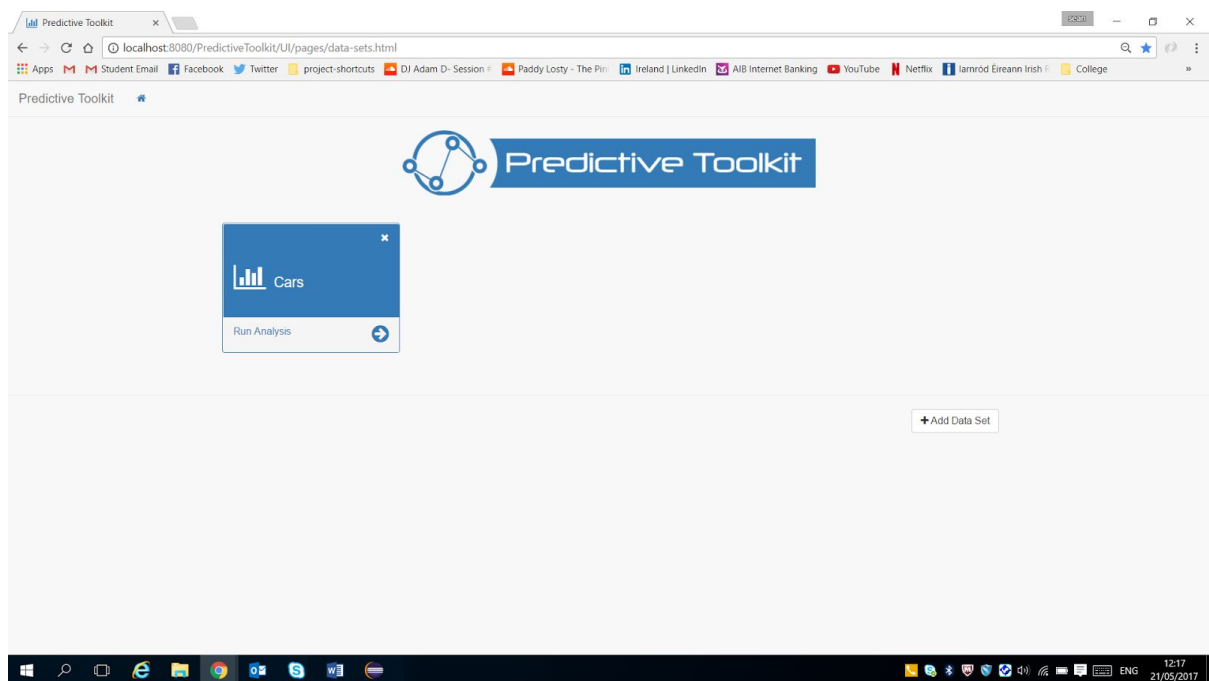
- Clicking on the large blue “i” help icon in the top right hand corner will display descriptions of what each of the fields are and about the format they must each be entered in.



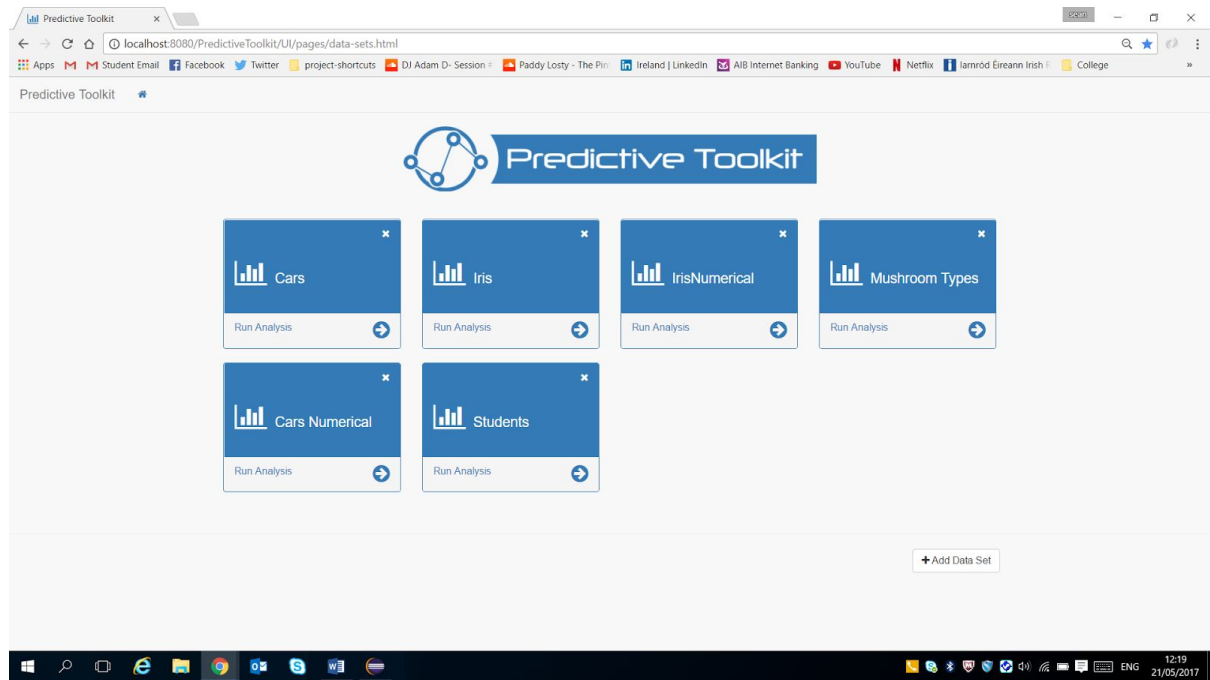
- The following shows an example of correctly entered data.



- When you have completed the fields, select upload. You will be brought to the home page where you will see the dataset has been uploaded.

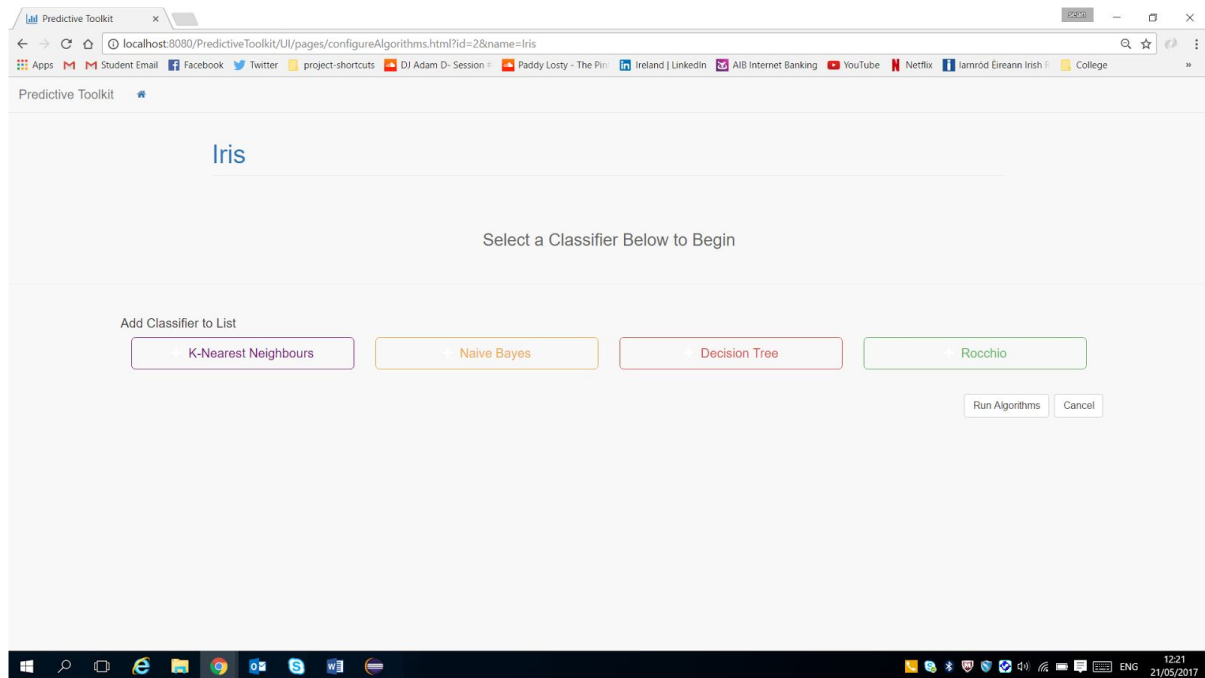


- You can upload as many datasets as you want.

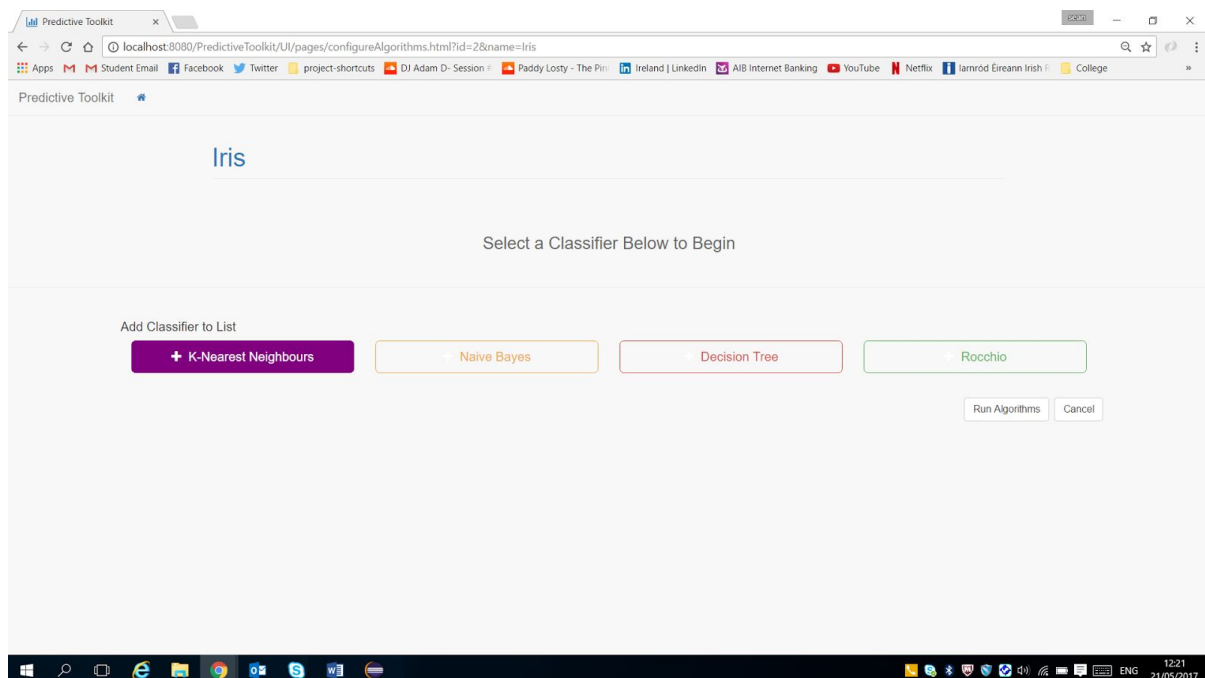


Configuring Algorithms

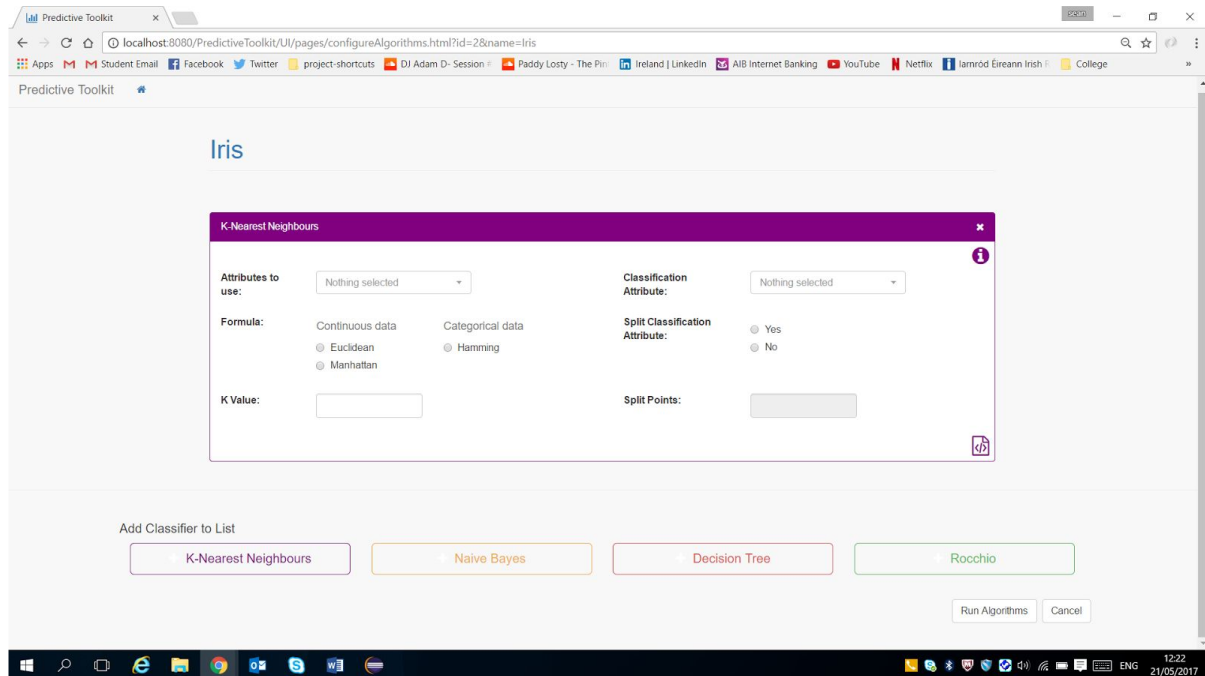
- Once you have selected “Run Analysis” on any dataset you will be then brought to the configuration page shown below. It will display the name of the dataset you are running on at the top of the page.



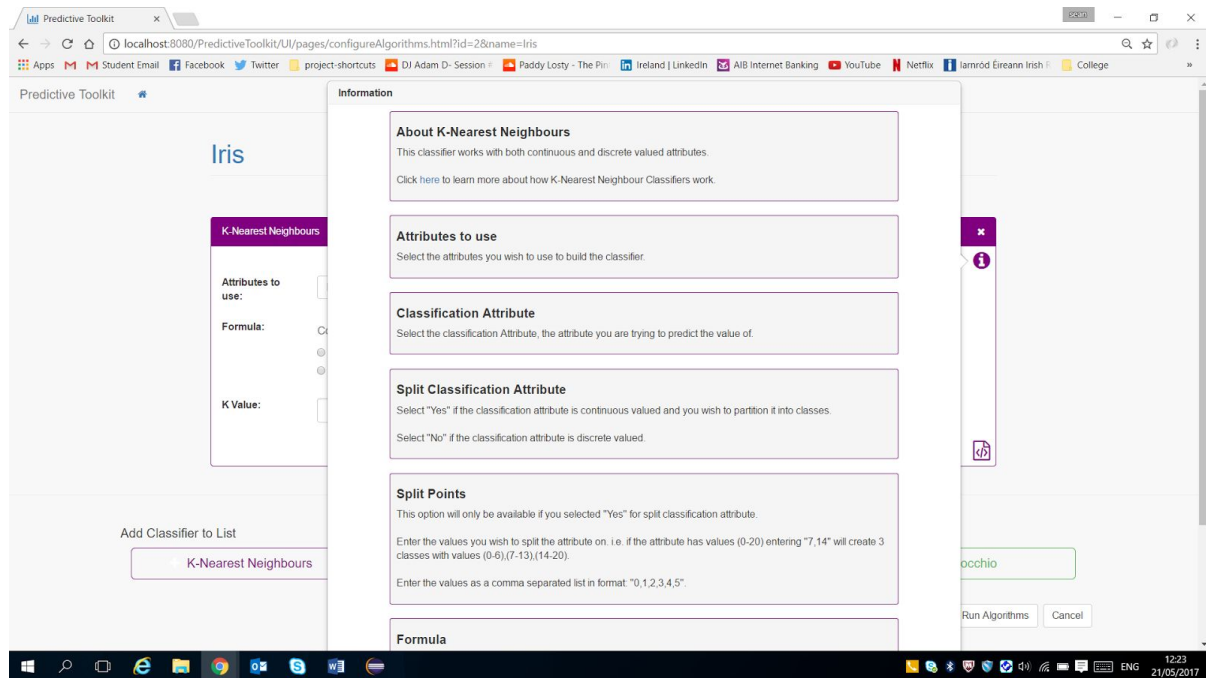
- Select one of the four classifier algorithm buttons along the bottom of the page to add a configuration to the run.



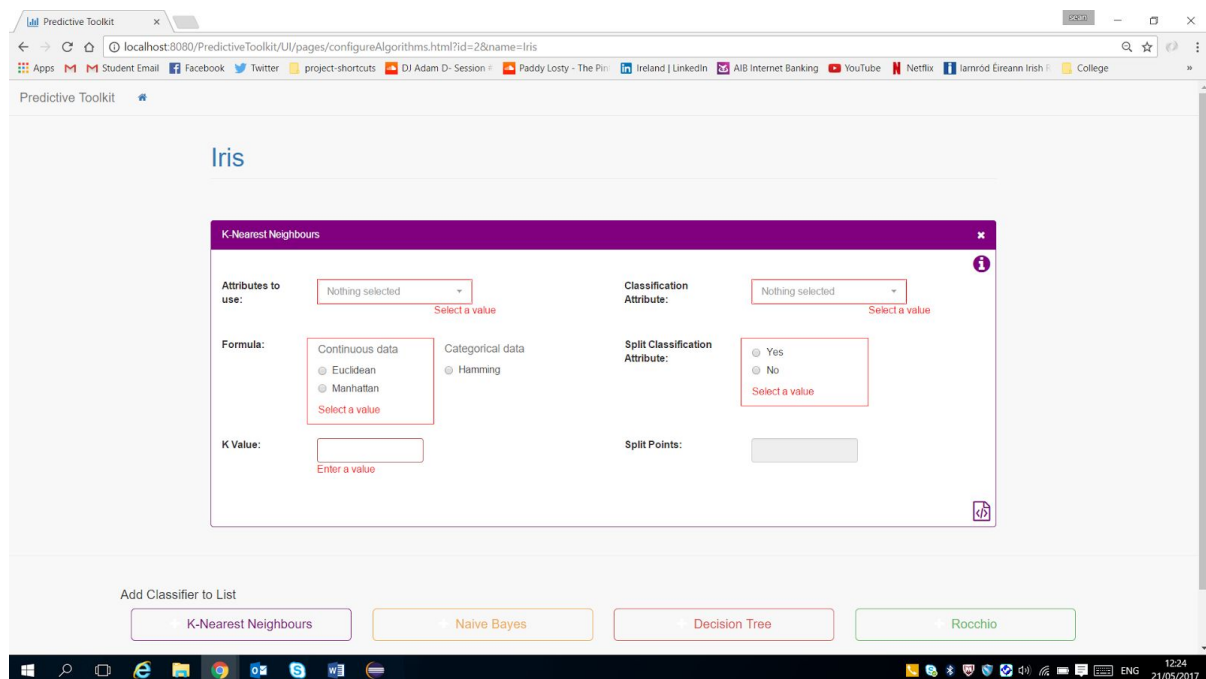
- When you have selected a button a tile will be created for that algorithm.



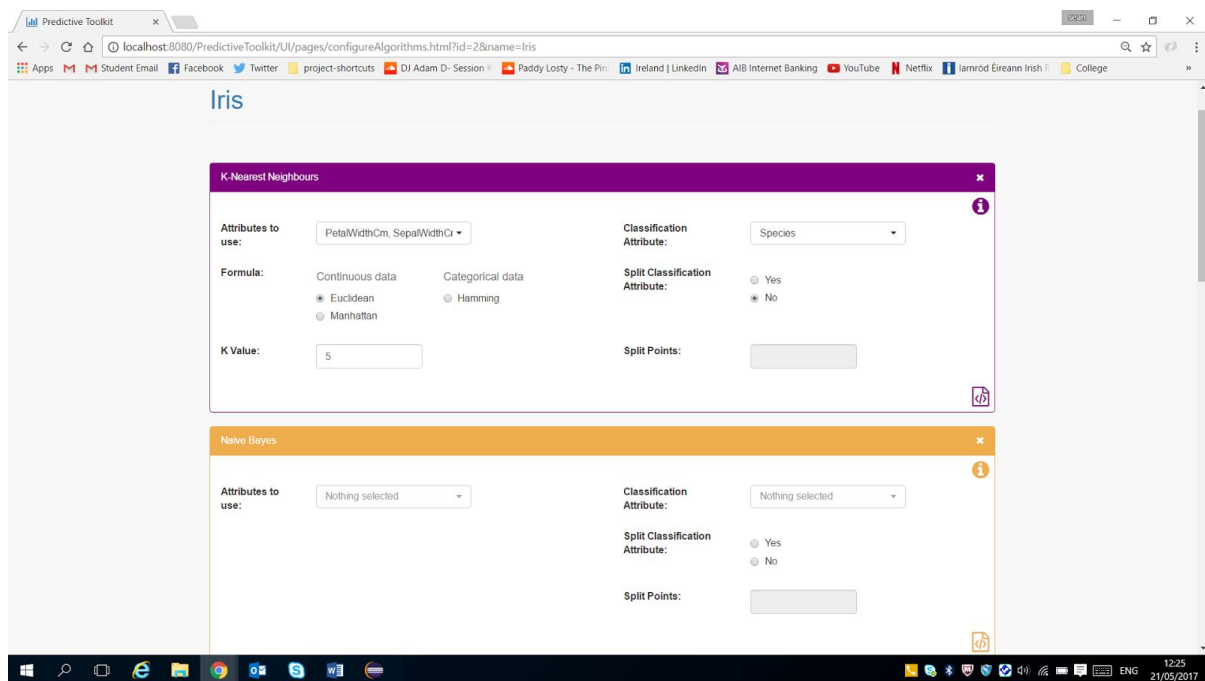
- Clicking on the large “i” help icon in the top right hand corner will display descriptions of what each of the fields are and about the format they must each be entered in.



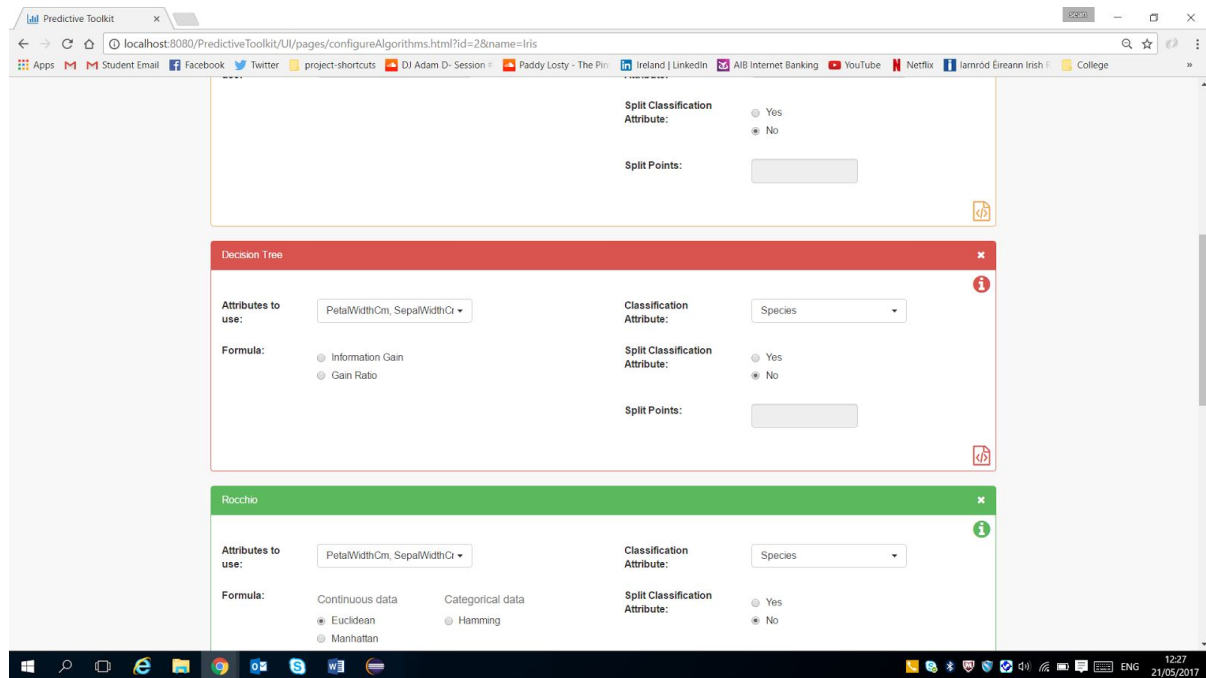
- If you forget to fill in one of the fields or fill them in incorrectly then the application will notify you by highlighting the problem field in red and displaying a text description of the problem when you click “Run Algorithms”.
- You will not lose any of the information you have already entered and will be able to fix the error.



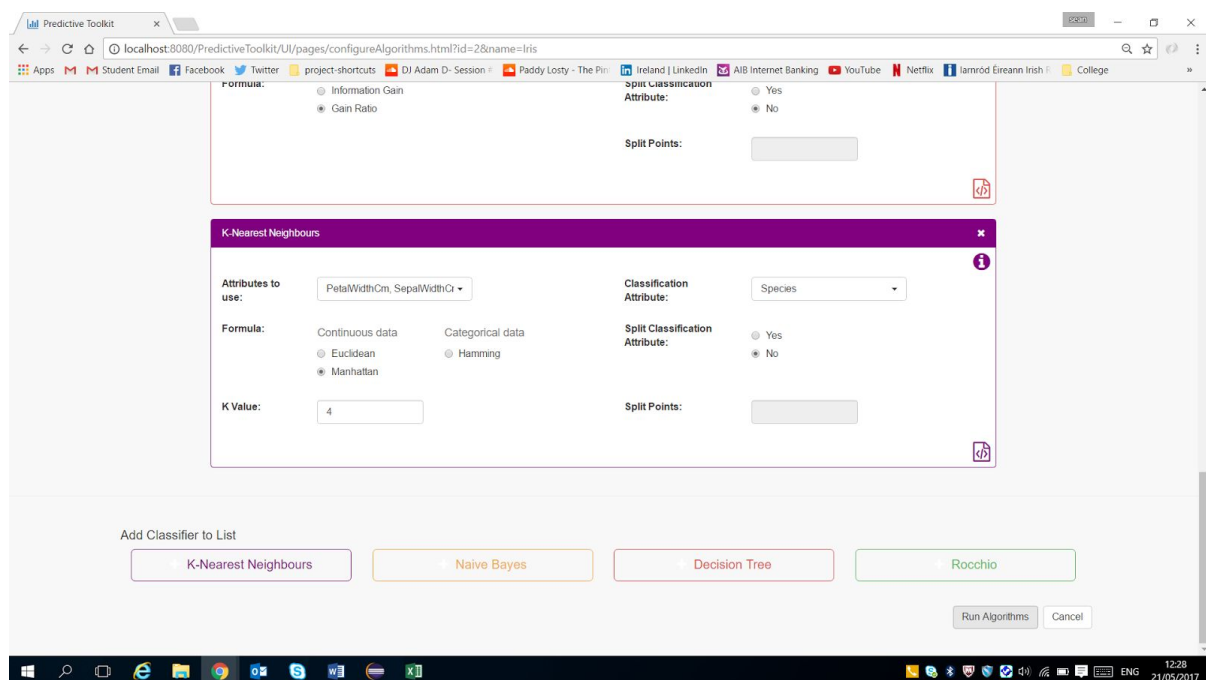
- You can add as many algorithms as you want to the run configuration.
- Below shows multiple tiles, I have filled in the values in the first configuration.



- clicking the copy option, in the bottom right hand corner of a tile, will populate the values of all other tiles with the values of that tile. It speeds up the configuration process so you're not entering the same thing multiple times.
- You can then go along and make the required changes and add the algorithm specific selections.

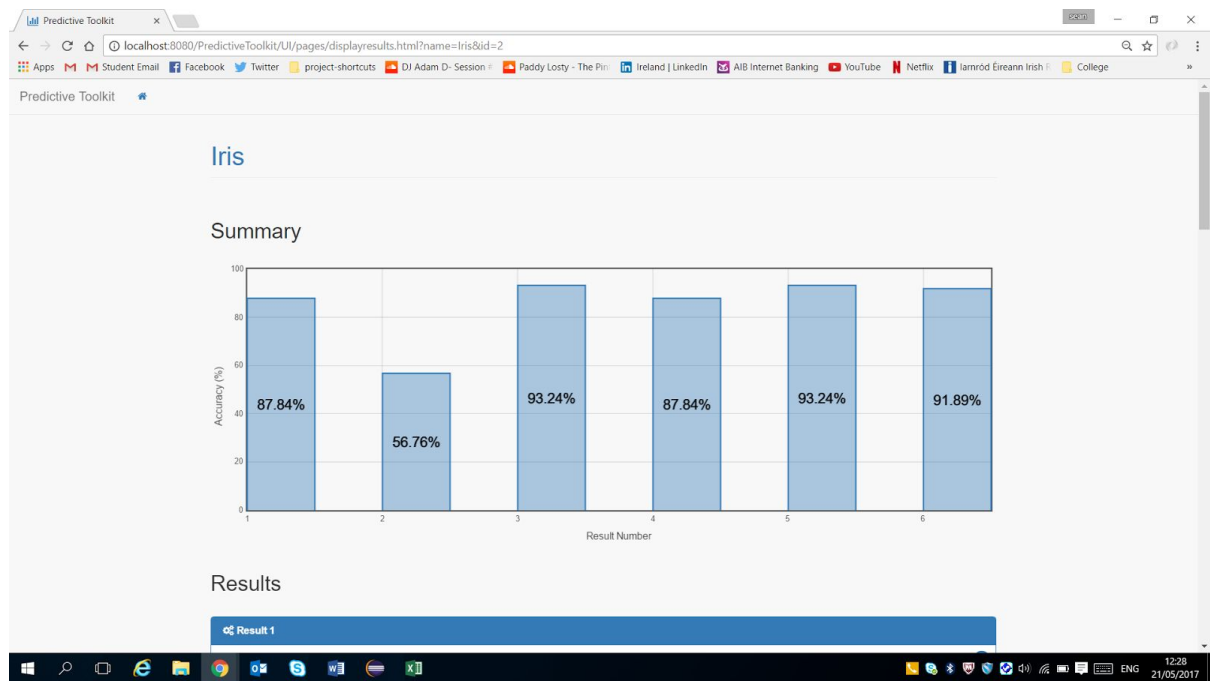


- When you have finished filling in all the required configurations select “Run Algorithms” to run these algorithms and be brought to the results page or cancel to abandon the configuration and go to the home page.

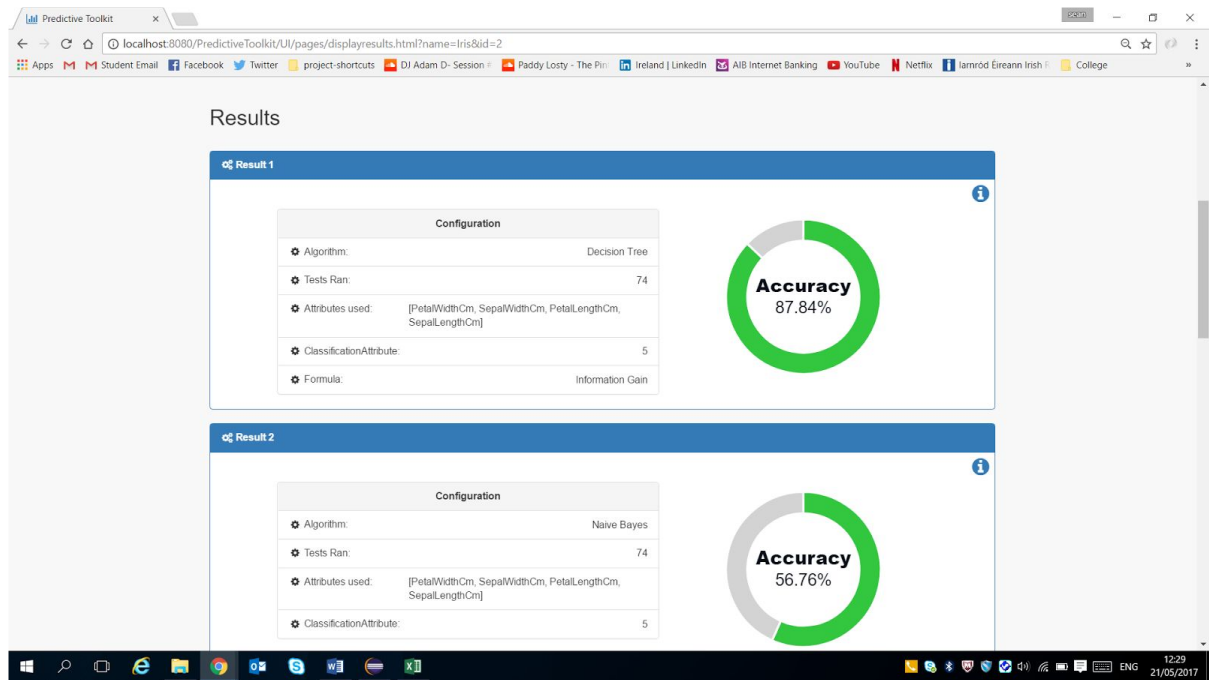


Reviewing Results

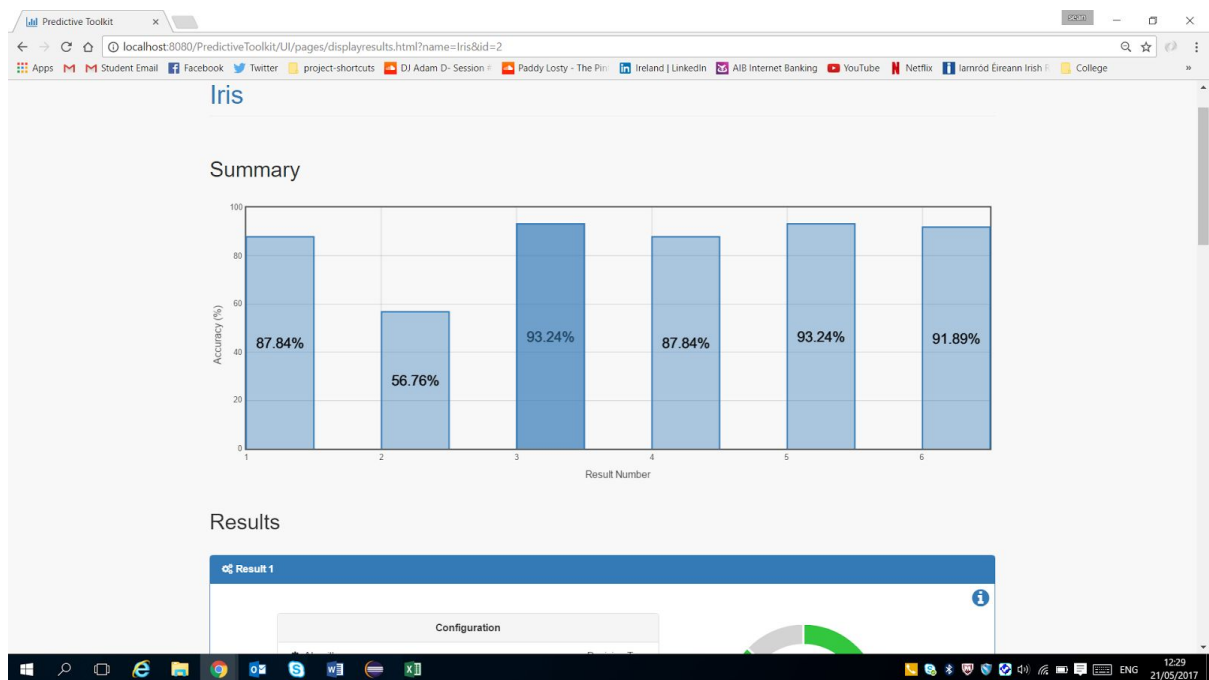
- You will be presented with a graph summary of the performance of all the configurations. You can easily identify if any particular runs have performed very well or very poorly for your data.

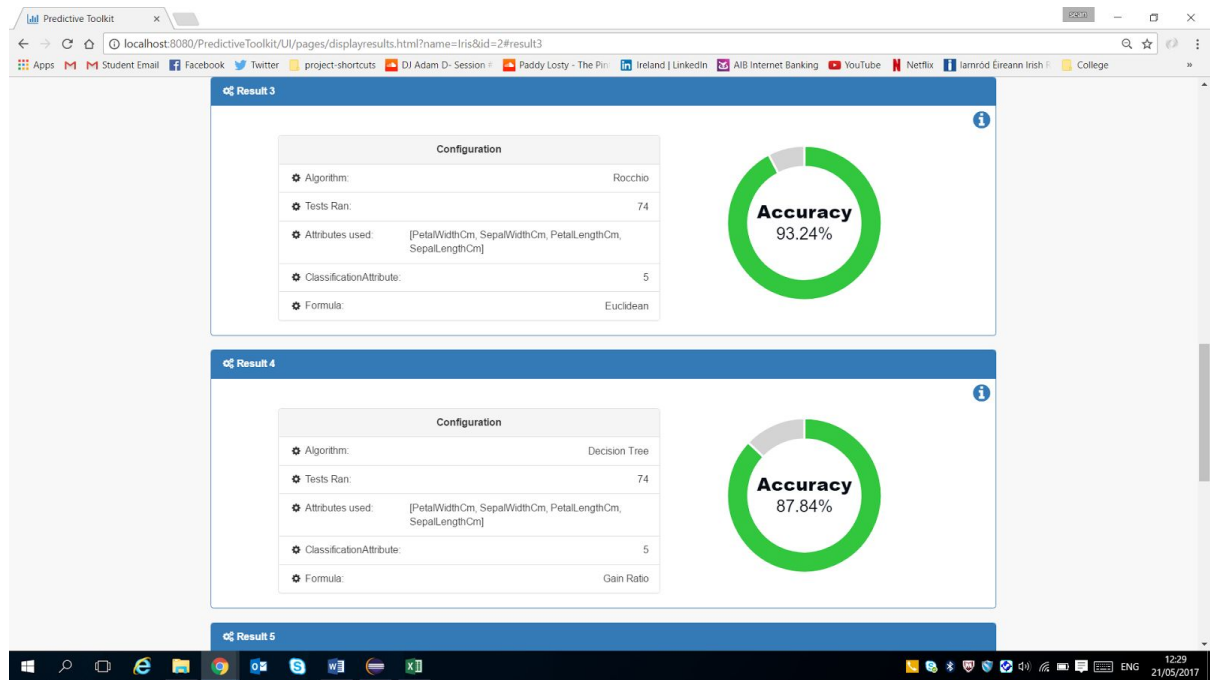


- You can scroll down to see more detailed information on each result.

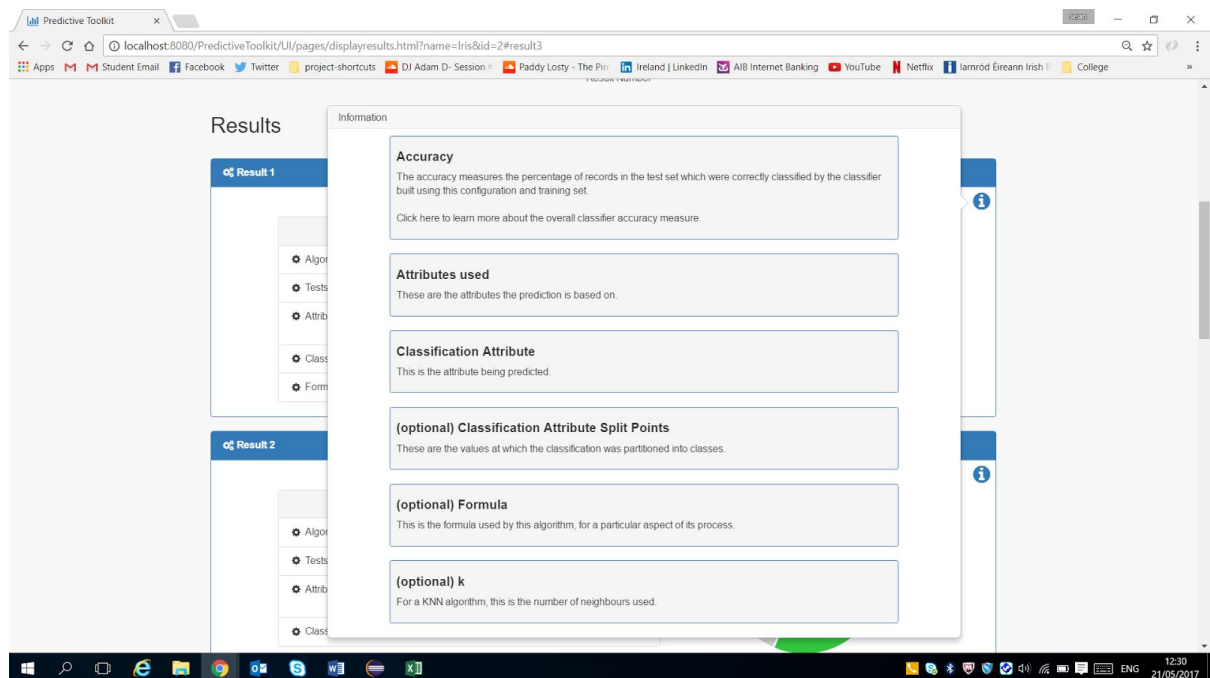


- You can also click on a bar in the results to jump straight to that result.

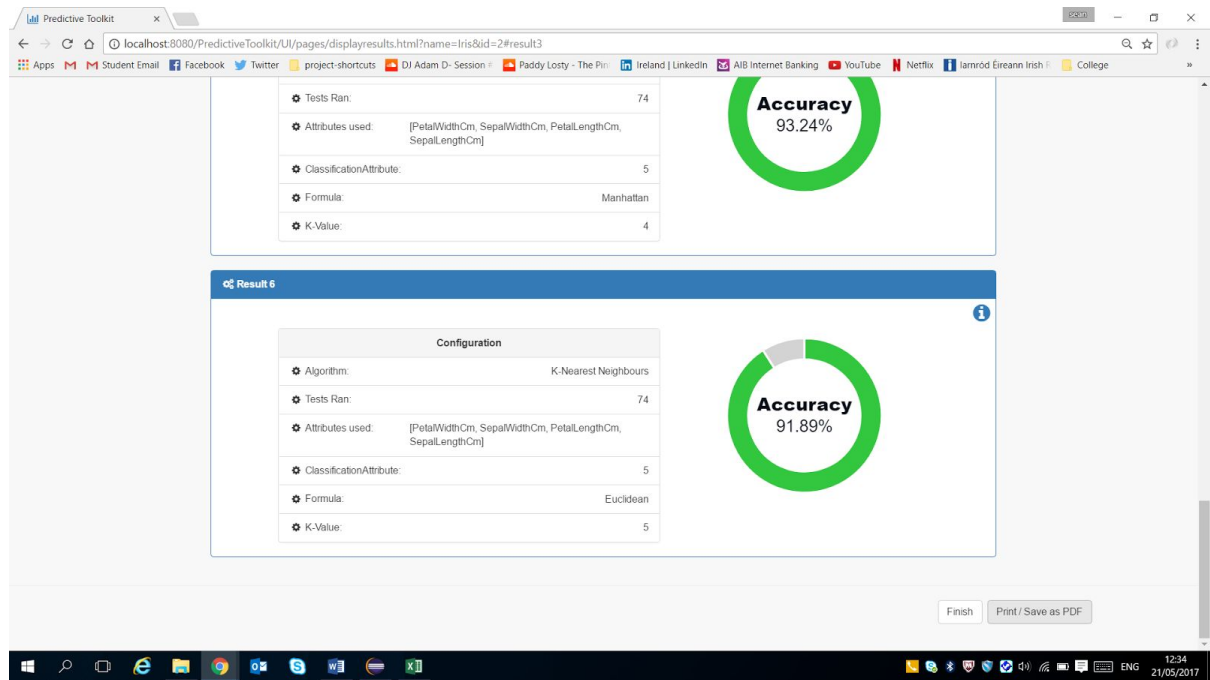




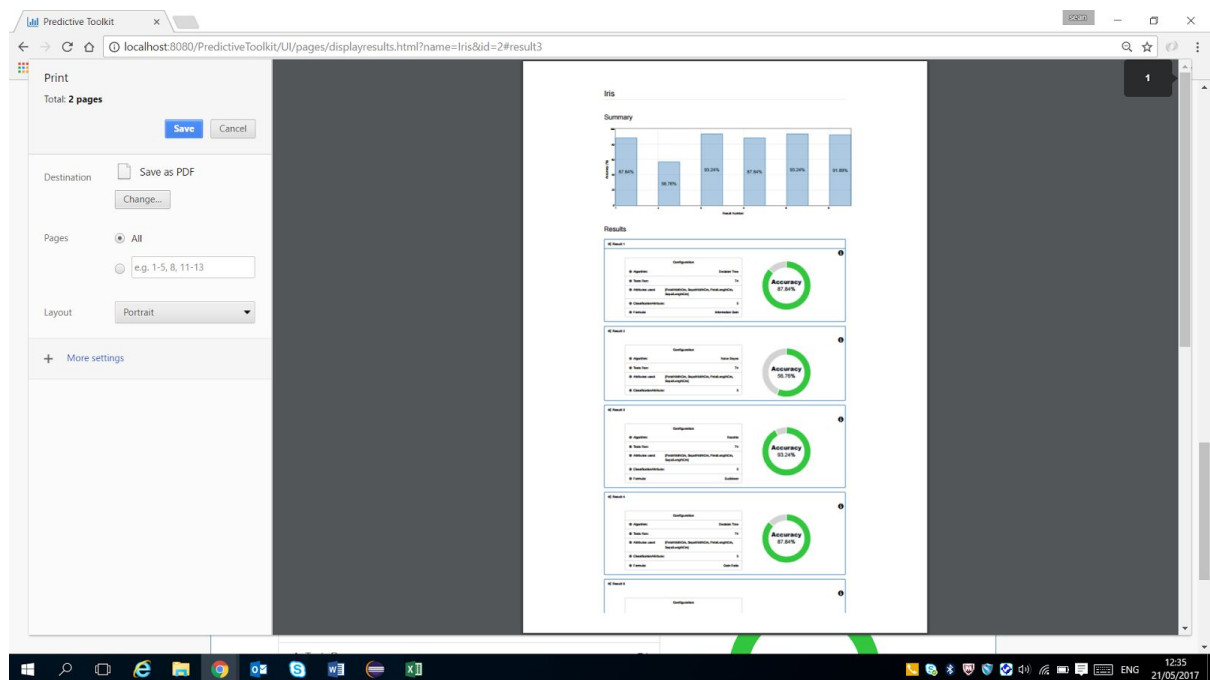
- Clicking on the large “i” help icon in the top right hand corner will display descriptions of what each of the result parameters are.



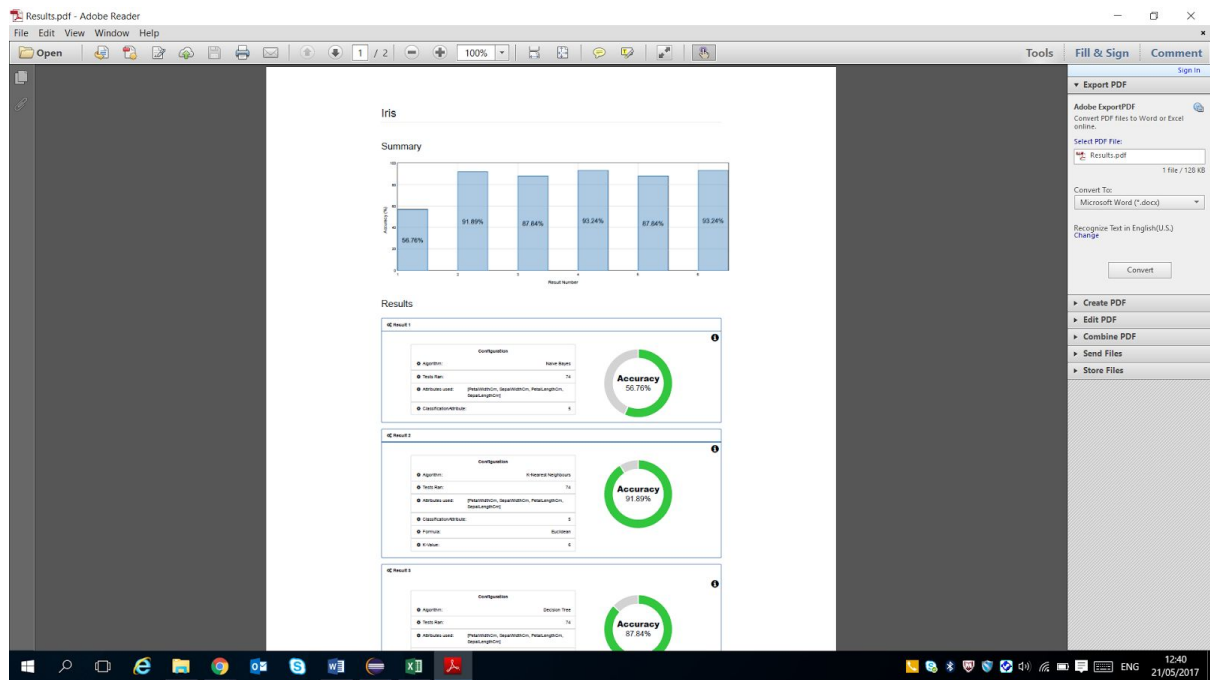
- If you want to save or print the results you can select the “Print / Save as PDF” button at the bottom of the screen.



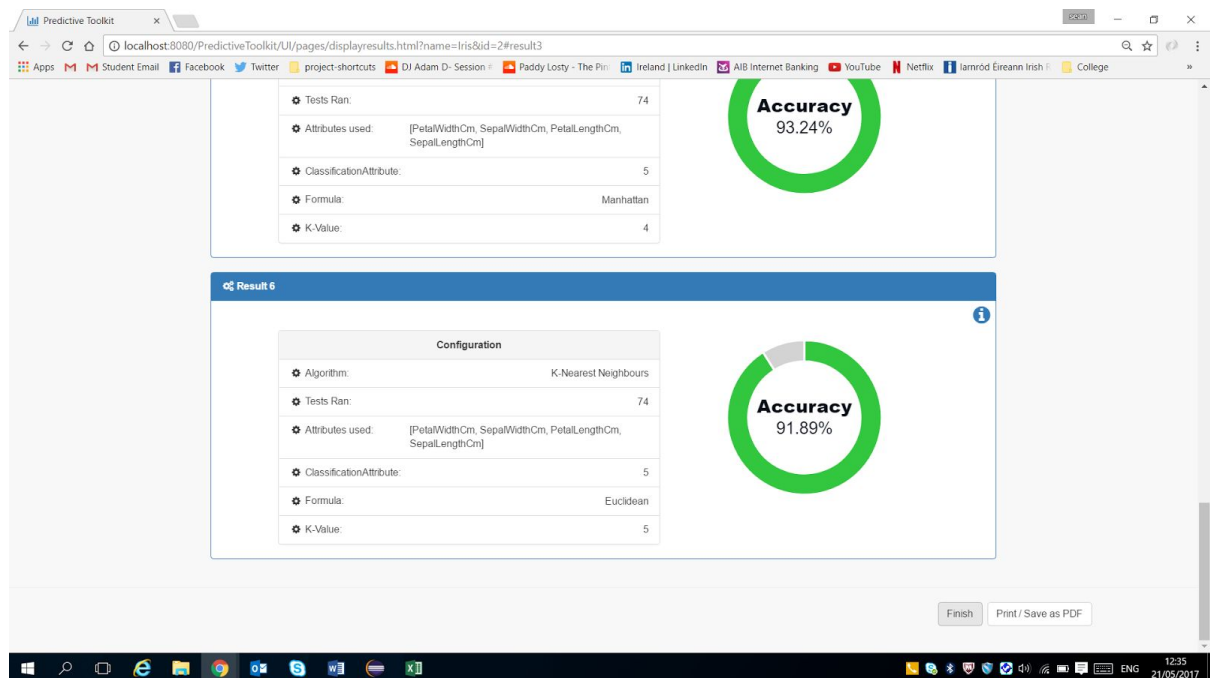
- This brings up Google's Print & save dialogue.



- This is an example of what the results look like once they have been saved to PDF.



- When you have finished analysing your results you can select “finish” at the bottom of the page.



- Once you select finish you will then be brought back to the home page.

