**COMP309 Assignment 2 Part One:**

Chosen Dataset:
https://data.mfe.govt.nz/table/52500-estimated-fish-and-invertebrate-bycatch-in-deep-water-fisheries-by-year-19912012/data/

This dataset displays the amount of fish and invertebrate bycatch from deep water fisheries, this information is tracked by year from 1991-2012, I have removed 2 data-points from the dataset, the 'field-5' and 'estimate type' as both have little to no relevance to the relationship being looked at.
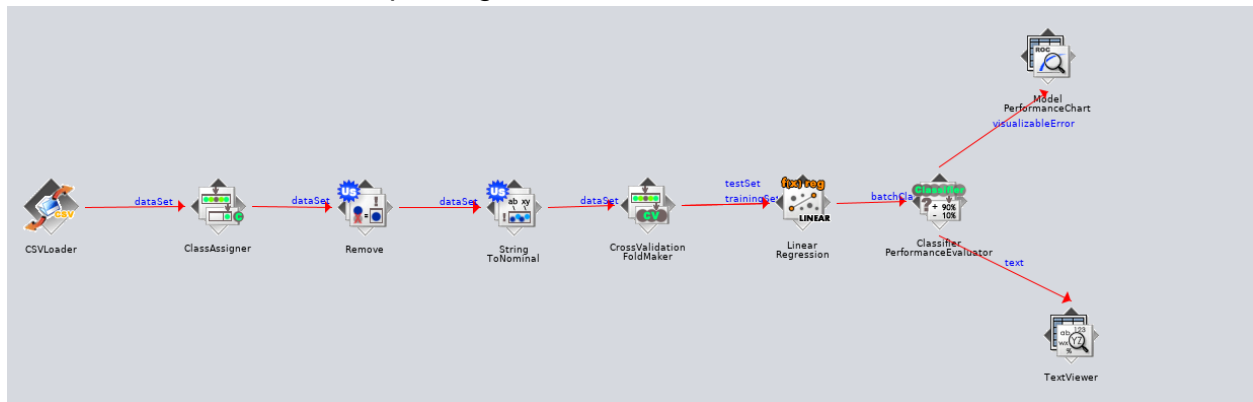

Linear Regression:
Using linear regression as data analysis allows for the benefit of linear predictability and an easily interpretable model able to show clear relationships, this is shown in the data with
Having chosen to use linear regression for analysis for its benefit of linear predictability onto an easily interpretable model allows for a straightforward and quick analysis of the data.
This is shown in the following graphical model, plotting each fishery type (type of targeted fish) against the volume of bycatch, with the relationship over time shown by the change in colour from blue (1991) to orange (2012), this relationship is showing a downward trend over time.
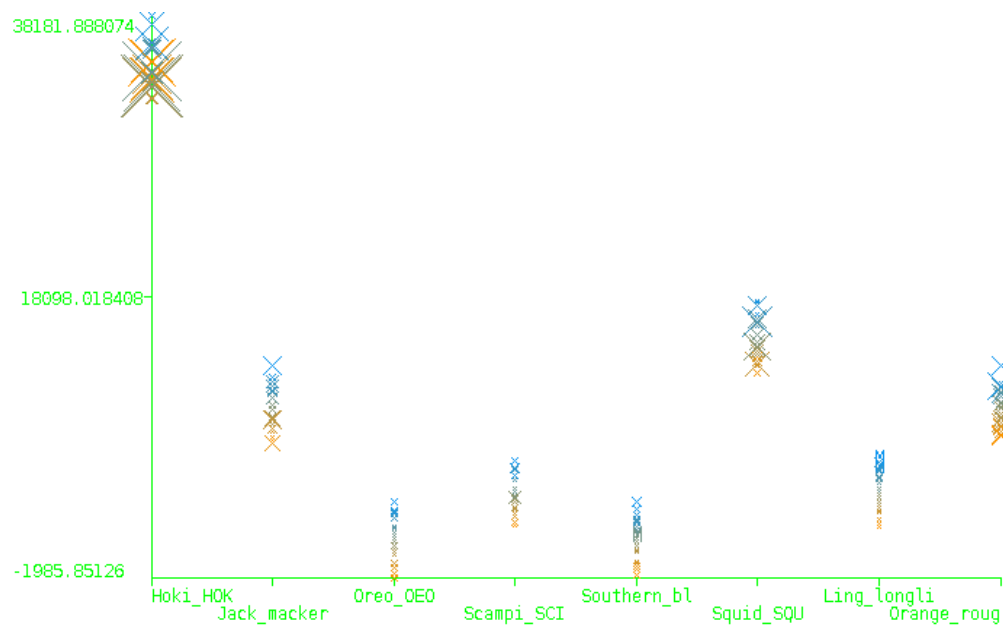
Other techniques were looked at for data analysis, however due to the nature of the dataset chosen they weren't appropriate. Clustering is unnecessary because this dataset is already split into defined groupings, this similarly applies to an attempt to use a classifier, as the dataset is fairly simplistic in its provided information, this is all beneficial to the use of linear regression to look at the singular relationship in the data.

"Is there any evidence of fish stocks collapsing in NZ?"
Taking a look at this question through the lens of this dataset, there could be a correlation between the lowering levels of bycatch and overall fish stocks collapsing, as a collapse would cause less fish overall to be caught. Alternatively the reduction in bycatch across the industry could be attributed to other factors such as better tools or other such improvements to increase the accuracy of the fishers catching their targeted types.

"Is there any evidence shown of over-fishing in New Zealand waters?"
The trends displayed by the linear regression model given by my analysis has shown evidence that could be correlated to over-fishing, as with the above question there is a case to be had for and against overfishing.

"Is there any evidence of fish levels being reduced due to environmental changes?"
Unfortunately this dataset does not contain information relevant enough to be immediately applied to environmental changes, and as with the above questions possible environmental changes could have an effect on the bycatch levels but it is impossible to state causation.

PART TWO:

Taking a further look at the question from part one:

"*"Is there any evidence of fish levels being reduced due to environmental changes?"*

This is an interesting question due to the abundance of talk about global warming and climate change, as well as general environmental change caused by humans. This is easily applicable to the concept of overfishing and the possibilities of how we could be doing damage to our oceans.

From this I found this dataset: https://data.mfe.govt.nz/table/53487-number-of-trawl-tows-19902014/data/

This is directly relatable to the previous data, as most of the fisheries would be using or directly affected by the use of trawls, and the subsequent damage of the ecosystem caused.

This data would be of use for tracking the environmental impact of trawling and the scale of the damage being caused, being able to combine this dataset with others to get a better understanding of any environmental impact caused by the fishing industries. From a business understanding, it can be used to regulate and keep an eye on fishing trends, and the predictions can provide an overview of any potential business competitors, and trends in methods being used.

Merging the dataset was done within excel using a tried-and-true methodology of copy and paste, due to the small size of the dataset this was a suitable method to get done in a quick and efficient time frame. Given more time a more comprehensive method of merging would be far more appropriate, such as WEKA command line.

I have removed redundant and/or irrelevant data-points from the data set, however not using any dimensionality reduction techniques, the data set was concise and clear enough that I had identified the irrelevant columns before even downloading the dataset. The remainder of the dataset is integral in the analysis of any trends and relationships between the variables.
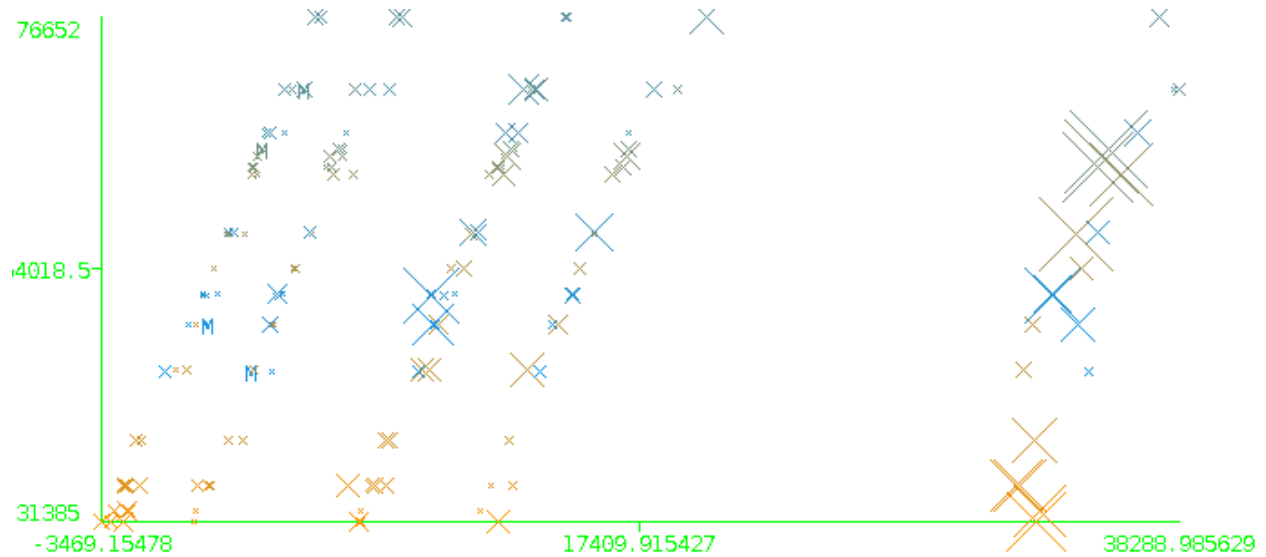
The output from my pipeline on the new dataset, plotting predicted bycatch levels against number of tows, with the year colouring being the same as before, blue (1991) to orange (2012).

Showing a downward trend again, this time in overall fishing tows and overall bycatch, the implication here is not that overfishing is possibly leading to the collapse of the fish stocks, but more so that the overall fishing rates are reducing over time.