Simplifying The Queer Divide

1 Introduction.

1.1 Dataset and Problem

A 2019 analysis [1] suggested that country-level changes to the legal standing of LGBT people over the last 25 years were significantly impacted by the form that each country's international relationships took. Velasco's two way fixed effect model seems to say that the more a country relates to the international community via LGBT focused INGOs (International Non Governmental Organisation) the more likely that LGBT people's legal standing improves. Conversely, the model suggests that aid relationships between countries (particularly when said aid makes up a significant proportion of recipient GDP) are sometimes tied to reactive responses against "new" LGBT norms – that is to say, the passing of anti-LGBT legislation.

Velasco's dataset [2] records 23 independent variables with annual observations for 110 non-OECD (The Organisation for Economic Co-Operation and Development) countries covering the dates 1990 – 2016. The model created from it is complex, and perhaps opaque to an audience unfamiliar with statistical methods. My goal was to create transparent predictive models for LGBT legislative change at the country level using only those two independent variables that were easiest to explain: Number of LGBT INGOs and Number of OECD donors.

This problem was modelled both as regression ("how much change was there?") and as categorisation ("was there any change at all?")

1.2 Data Pre-processing

1.2.1 Choice of Platform

All work took place in the interactive Python environment of a Jupyter Notebook. This method was chosen as the platform is open source and easy to share, and this fitted well with my goal of improving accessibility. Similarly, the libraries chosen (sklearn, seaborn, matplotlib) are well documented and readily accessible.

1.2.2 Dependent variables

Velasco's measure of LGBT legal standing is a yearly point score based on the presence or absence of certain LGBT policies in a country (see overleaf). Originally, scores on this index ranged from -2 to 9. This range was scaled between 0 and 11 to eliminate divide-by-zero errors. I have condensed these observations into the following averaged figures for each country.

Table 1. LGBT policies used with coding score

Policy	*Score*
Same-sex acts legal	1
Equal age of consent	1
Employment discrimination	1
Hate crime protections	1
Incitement to hatred	1
Civil unions	1
Marriage equality	1
Joint adoptions	1
LGB military	1
Constitutional anti-discrimination protections	1
Death penalty	−1
Propaganda laws	−1
Same-sex acts legal	−1
Unequal age of consent	−1
Ban on marriage equality	−1
LGB military bans	−1

Velasco's scoring system for a country's LGBT legislative situation. [1]

Average Score: This is the mean score across all observations for a country.

Average Yearly Change: The average relative increase/decrease of points every year expressed as a percentage. This measure communicates the general direction of legislative change.

Average Absolute Change: In the above measure, positive and negative change values cancel each other out. This measure ignores the direction of change and focuses only how much change of any sort happens.

Presence of Change: This Boolean registers whether change of any sort has happened.

1.2.3 Independent Values

For each country and year two variables were taken from Velasco's dataset: "**Number of LGBT INGOS**" and "**Number of OECD Donors**". A mean figure for each variable across the period was then calculated.

1.2.4 Scale of transformation

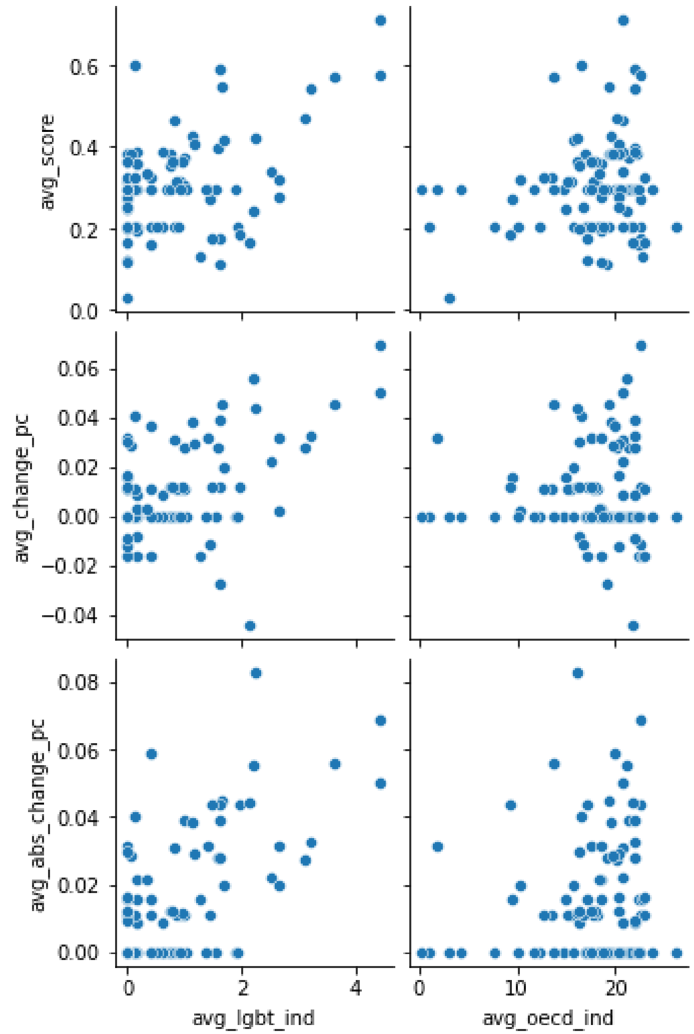
Velasco's original dataset contained 2,826 country/year observations. This study derived from that dataset a condensed dataset containing 110 country records with 660 averaged observations total. NaNs, where encountered, were discarded.

1.3 Methodology & Rationale

1.3.1 Exploratory Data Analysis

All dependent values were pair-plotted against all independent values as a preliminary step before choosing methods and target variables. These plots seemed to back up the idea that increases of either kind of relationship (INGOs or OECD donors) were correlated with LGBT legislative change.

Also of note was that though many countries experienced no change to LGBT legislature during the period, those countries with the most extreme positive change for LGBT people were those with more LGBT INGOs in place. Conversely, more OECD donor relationships seemed to correlate to more absolute change, but sometimes change leading to worse outcomes for LGBT people.



1.3.2 Choice of Methods

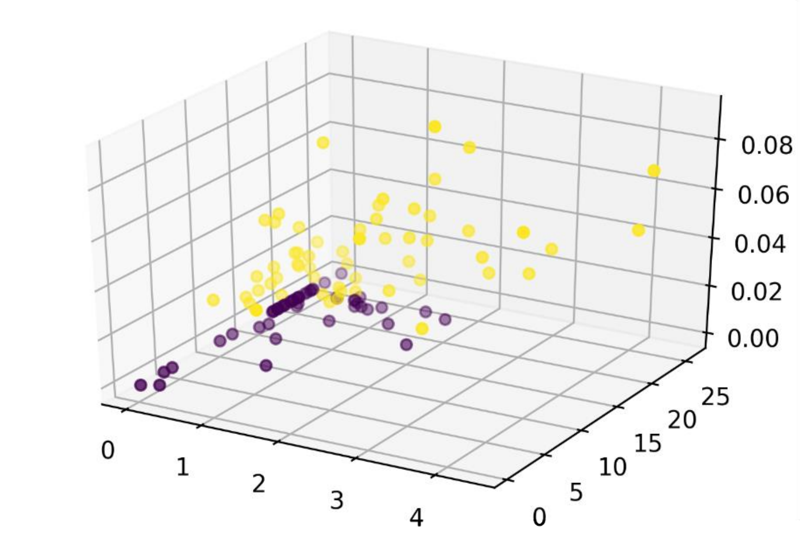
I decided to model average yearly change via a regression model as this would allow me to predict rate and direction of change as continuous variable. Linear Regression was chosen as its models are easy to communicate visually.

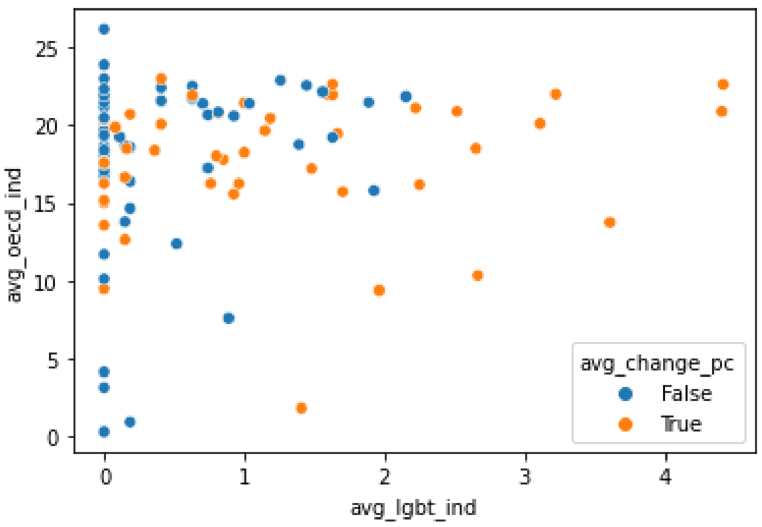
The presence or absence of legislative change was modelled as two categorisation models. The first categorisation method chosen was logistic regression. This was chosen, again, because of the visual simplicity of its models. Support Vector Machines were considered, but discarded as their reliance on higher-dimensionality would make them harder to explain to a lay audience.

As the dataset lacks linear separability (see figs) a decision tree was chosen for the second categorisation. Though a decision tree's decision boundary can make for a confusing plot, the ability to plot a tree in its entirety arguably makes it one of the most transparent categorisation models.

1.3.3 Test/Training Splits

All models were built on the same training data (70% of total observations) to make comparison between them easier.



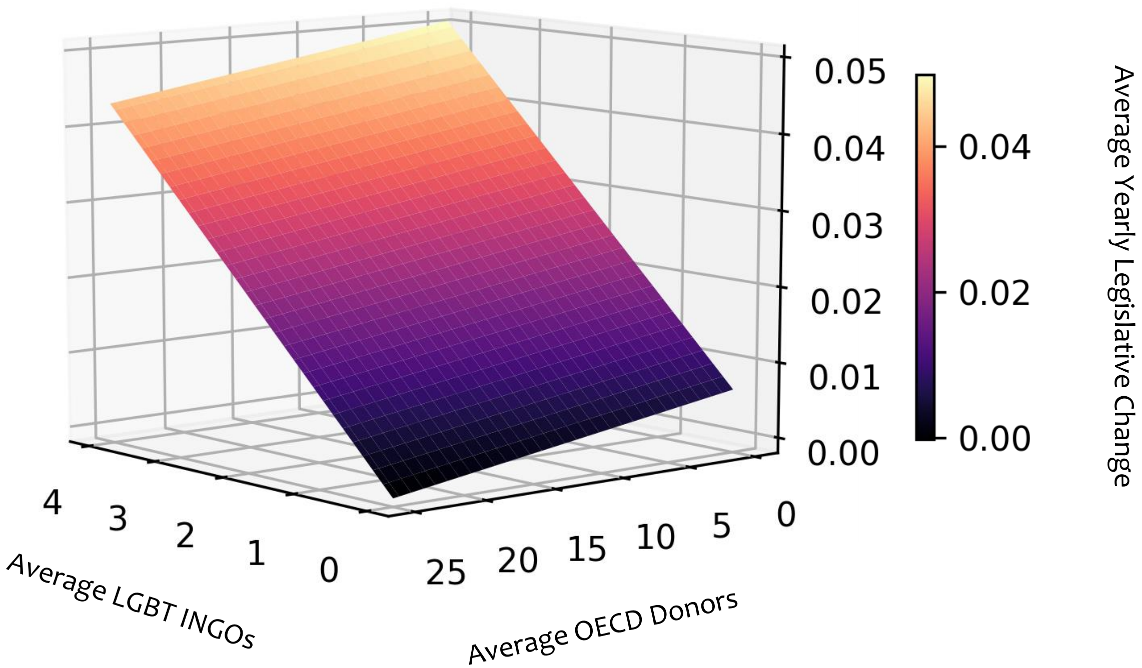


2 Machine Learning Models

2.1 Multivariate Linear Regression

This model was formed using Scikitlearn's Linear Regression functions. The model was plotted (see below) using Matplotlib. The average yearly change had a range of 10.6% between its minimum and maximum values. This model's Root Mean Square Error was 1.6 percent, which equalled around 15% of the total range.

As this model is so reductive compared to Velasco's data, I was impressed to see the relationships described in their study effectively recreated here.



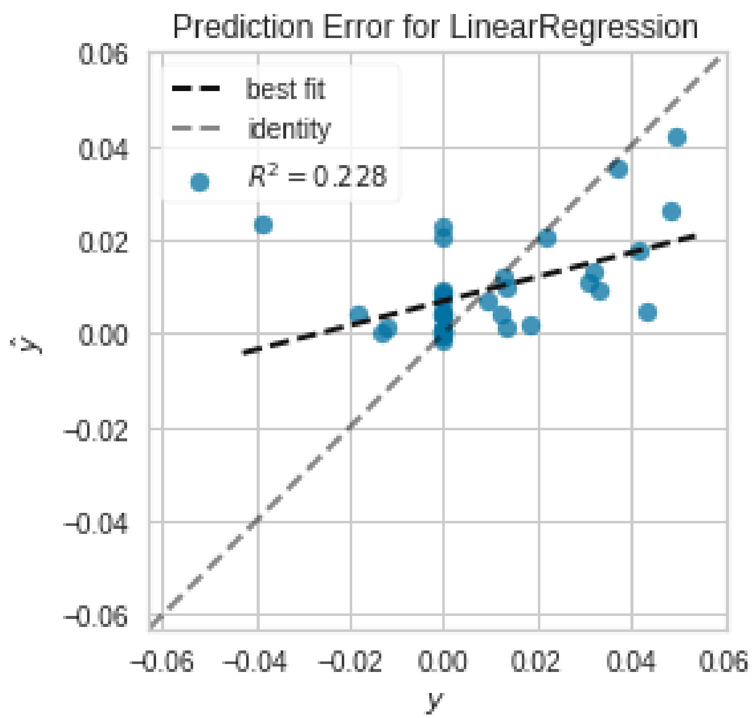
The Yellowbrick addition to Scikitlearn was used to create a prediction error plot as well as a plot of residuals.

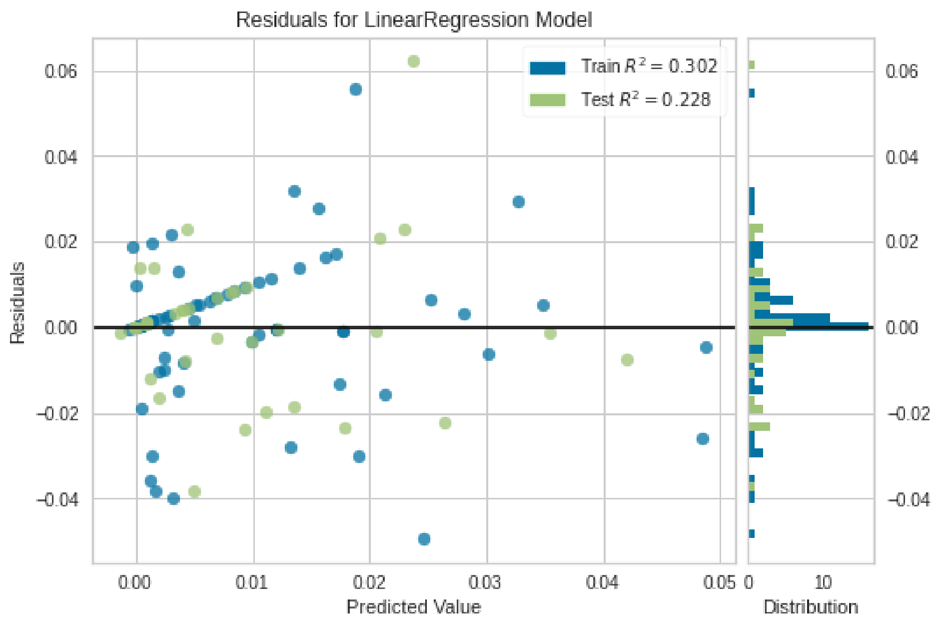
The prediction error plot suggests the model moves most results closer to 0 than they are, boosting negative figures and decreasing positive ones. This demonstrates the gravitational effect of the many sampled countries that did not change at all. Those countries (as the central line on this plot shows) were consistently over-estimated by the model.

Similarly, the plotted residuals show a strong diagonal line between 0 and 2%. The exact relationship between predicted value and residuals here shows the model struggling to accommodate the presence of the countries with 0% average yearly change.

Parameter tuning could be attempted, but the presence of these unchanging countries effectively places a limit on how accurate a simple linear model can be.

The R^2 figure of 0.228 is indicative of how simple the three variable model is compared to the complexity of the global situation, but also suggests a significant relationship requiring further exploration.



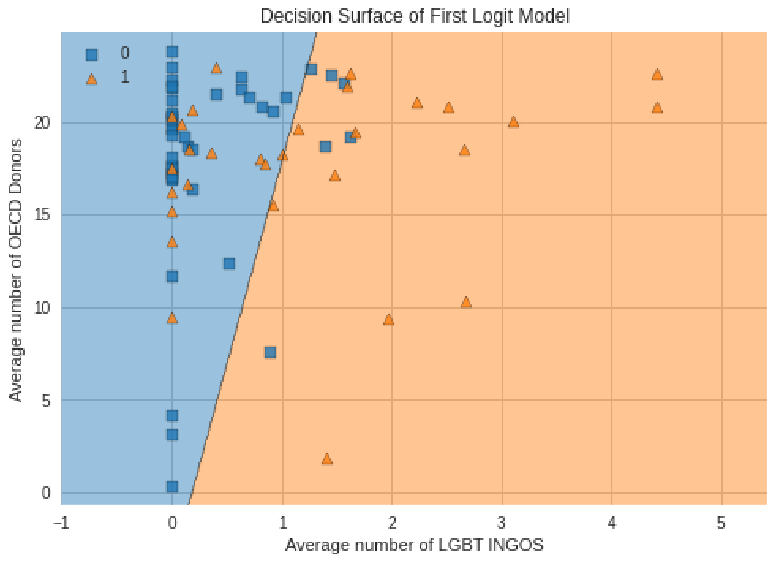


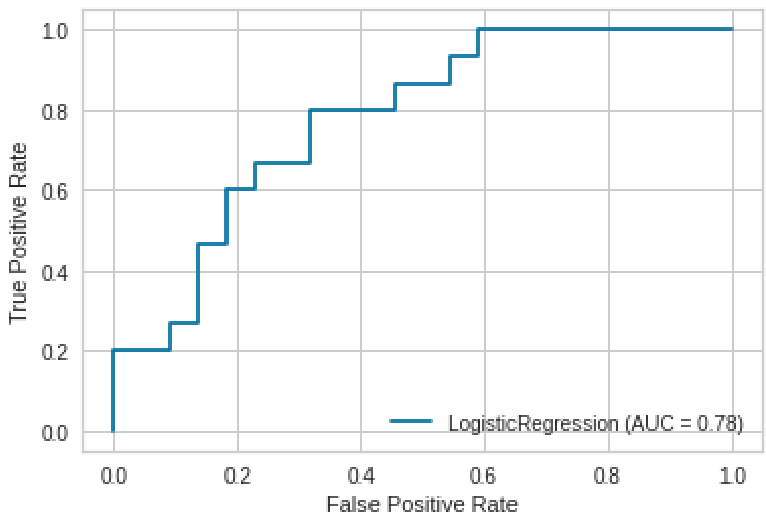
2.2 Logistic Regression

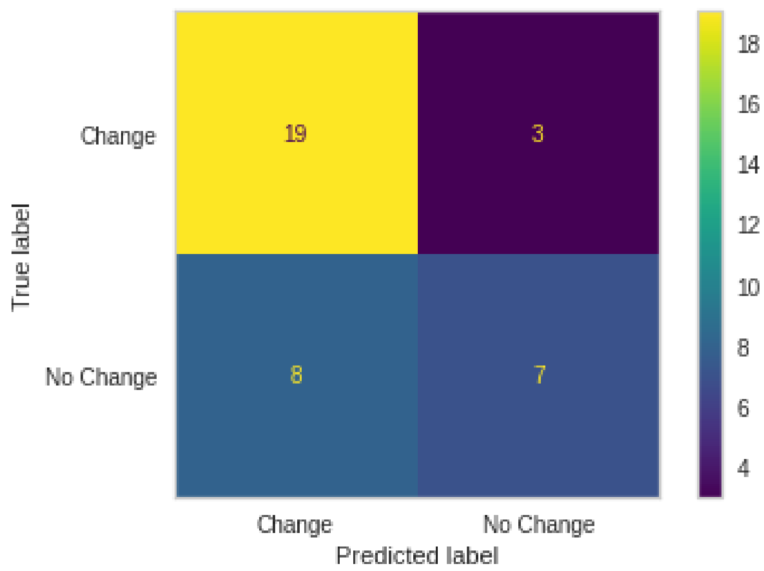
The second model made was a logistic regression to predict whether each country made any legislative change at all over the 25 year period. The logistic regression function of Scikitlearn was used.

The first iteration of the model was ran with the function's default parameters. It's decision boundary shows clearly the difficulties the model had with the test data's lack of linear separability. However, the ROC graph (and associated AUC of 0.78) suggests a reasonably high level of accuracy for such a simplified model.

The confusion matrix shows that the model found it much easier to identify countries that did change than those that did not. This might be explained by the overlap of the category ranges, which is to say, the area of the plotted samples containing only changing countries is much larger than the area containing only non-changing countries.

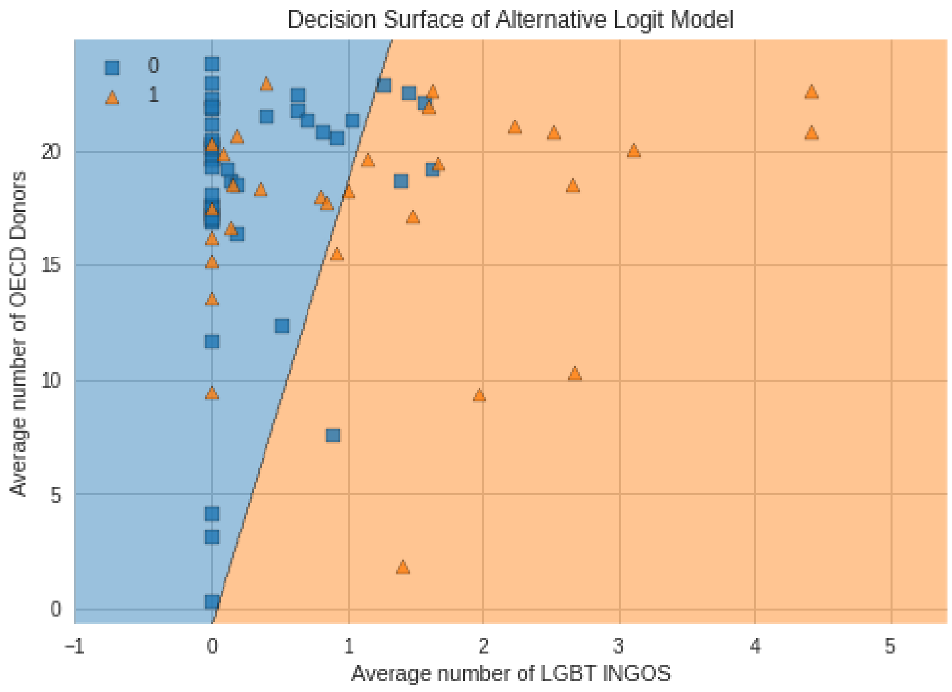


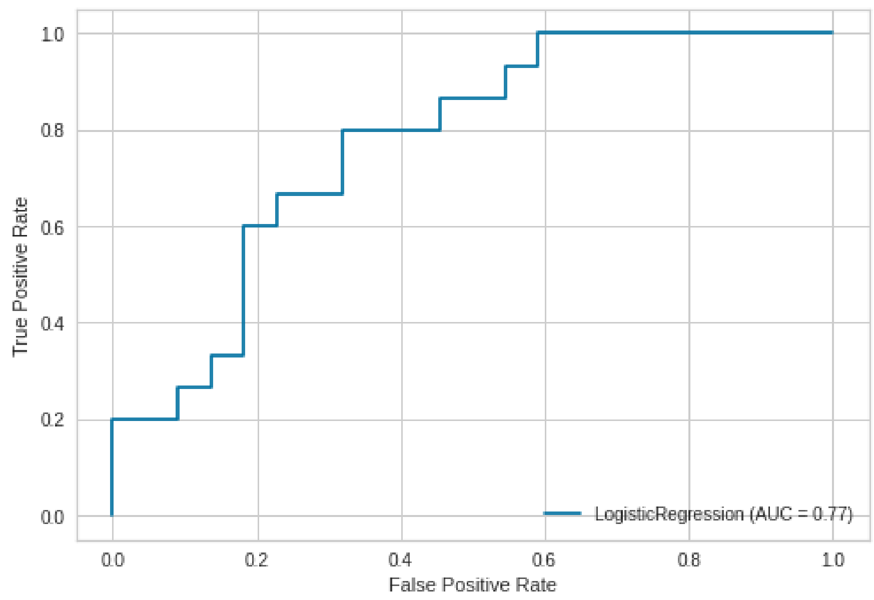




As there was some evidence of collinearity between the two independent variables, several variations on the hyper-parameters were tried. Use of Lasso and Elastic Net regularisation resulted in a small decrease in accuracy, but identical results when categorizing countries that had not changed.

The models' ability to accurately categorize unchanging samples performs at around the level of random guessing. That suggests that while likelihood of legislative change is positively correlated with both independent variables, one or more absent variables is needed to explain why some countries do not change despite scoring similar values on these metrics to those that do.



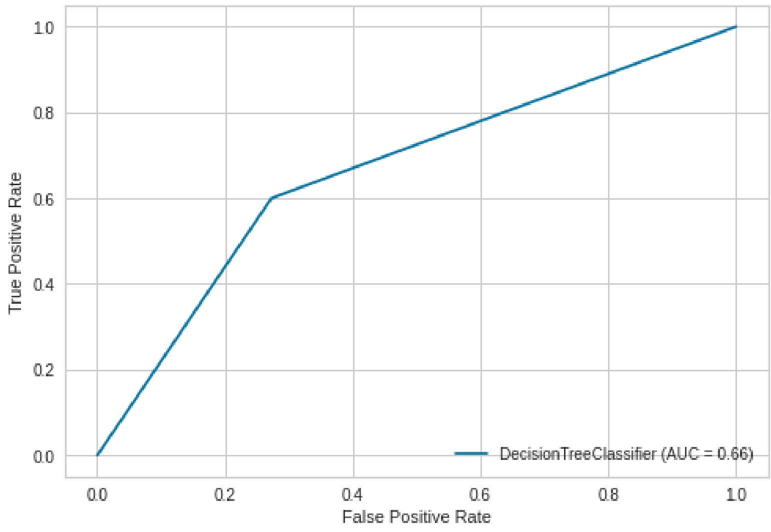


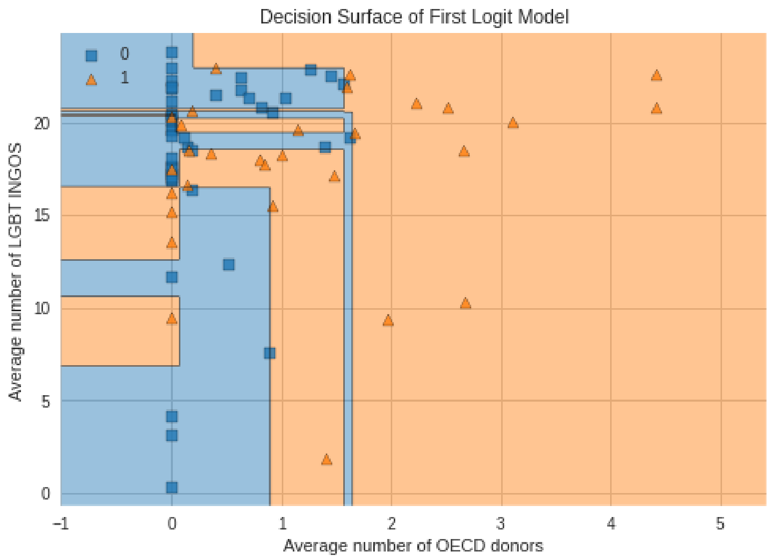
2.3 Decision Tree Classification

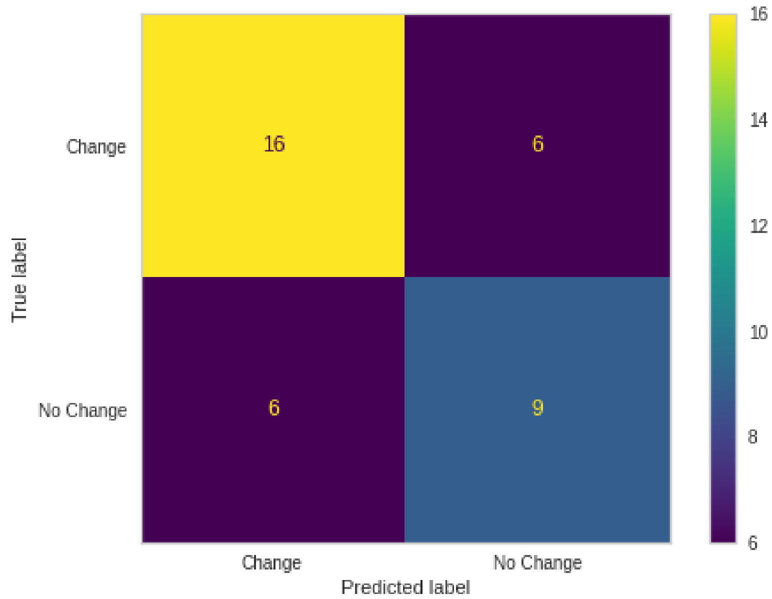
The second categorisation model made using the same variables was a decision tree. Again, the first instance was made using Scikitlearns' defaults for their decision tree classifier function.

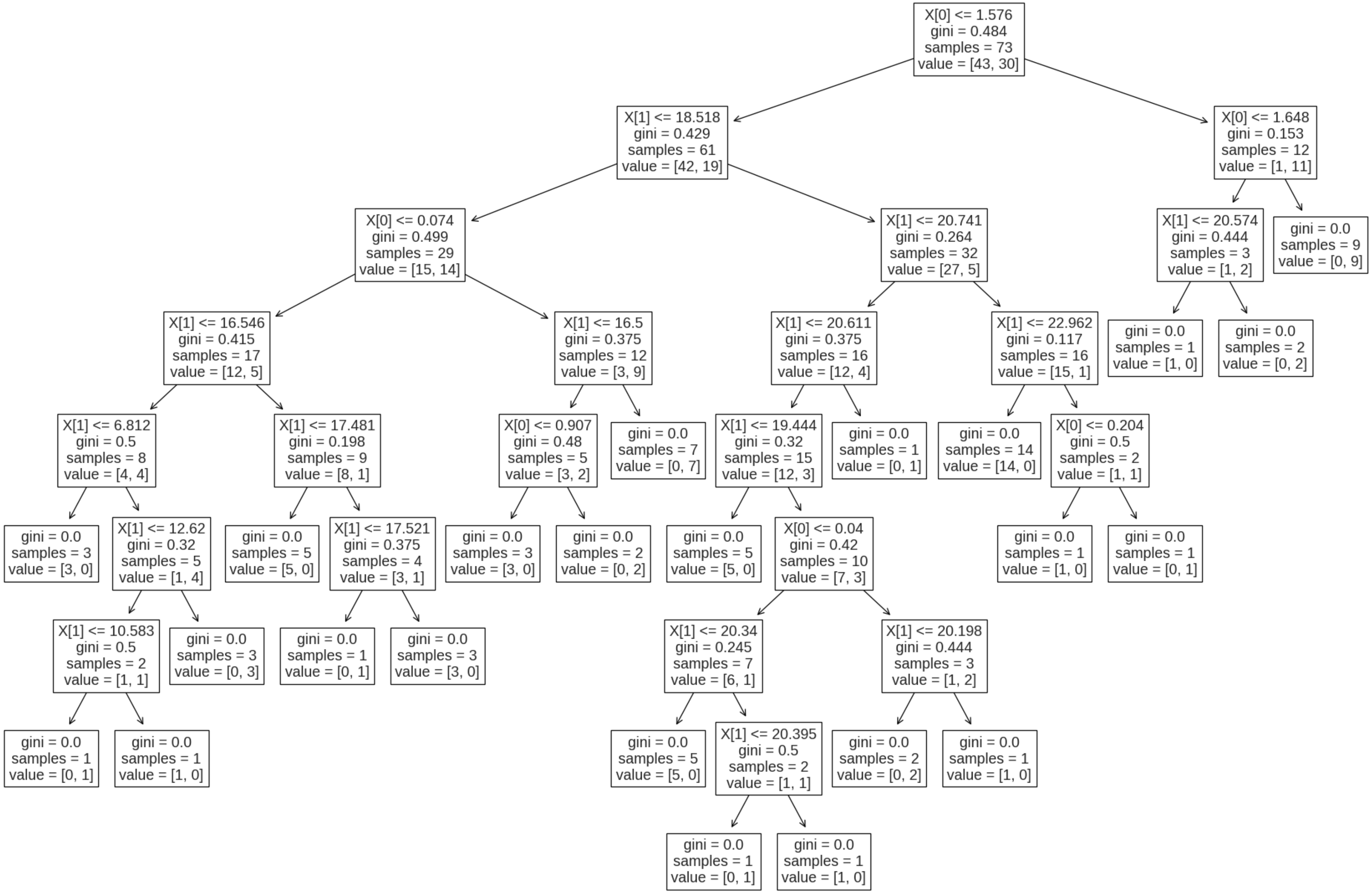
Initial results (AUC = 0.66) were significantly less accurate than the original logistic regression model (AUC = 0.78). Though the distribution of misclassifications was similar (with most difficulties being found with classifying countries without legislative change) the decision tree also had 100% more false negative categorisations.

Analysis of the model's decision surface (below) and tree plot (over leaf) suggested that the model was overfit, containing many "leaf" nodes with only one observation.





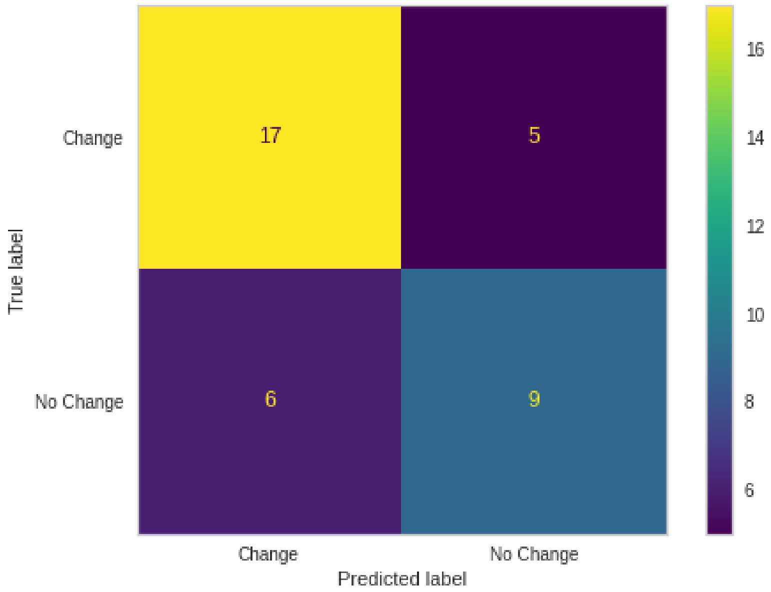


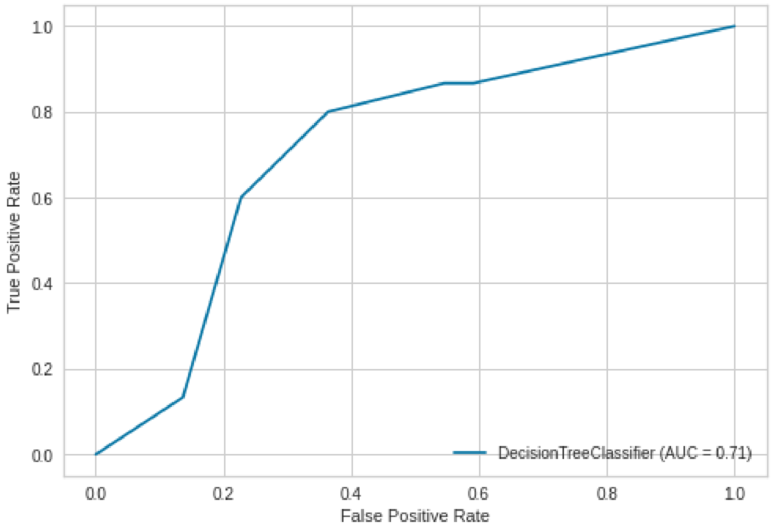

X[0] <= 1.576 gini = 0.484 samples = 73 value = [43, 30]
X[1] <= 18.518 gini = 0.429 samples = 61 value = [42, 19]
X[0] <= 1.648 gini = 0.153 samples = 12 value = [1, 11]
X[0] <= 0.074 gini = 0.499 samples = 29 value = [15, 14]
X[1] <= 20.741 gini = 0.264 samples = 32 value = [27, 5]
X[1] <= 20.574 gini = 0.444 samples = 3 value = [1, 2]
gini = 0.0 samples = 9 value = [0, 9]
X[1] <= 16.546 gini = 0.415 samples = 17 value = [12, 5]
X[1] <= 16.5 gini = 0.375 samples = 12 value = [3, 9]
X[1] <= 20.611 gini = 0.375 samples = 16 value = [12, 4]
X[1] <= 22.962 gini = 0.117 samples = 16 value = [15, 1]
gini = 0.0 samples = 1 value = [1, 0]
gini = 0.0 samples = 2 value = [0, 2]
X[1] <= 6.812 gini = 0.5 samples = 8 value = [4, 4]
X[1] <= 17.481 gini = 0.198 samples = 9 value = [8, 1]
X[0] <= 0.907 gini = 0.48 samples = 5 value = [3, 2]
gini = 0.0 samples = 7 value = [0, 7]
X[1] <= 19.444 gini = 0.32 samples = 15 value = [12, 3]
gini = 0.0 samples = 1 value = [0, 1]
gini = 0.0 samples = 14 value = [14, 0]
X[0] <= 0.204 gini = 0.5 samples = 2 value = [1, 1]
gini = 0.0 samples = 3 value = [3, 0]
X[1] <= 12.62 gini = 0.32 samples = 5 value = [1, 4]
gini = 0.0 samples = 5 value = [5, 0]
X[1] <= 17.521 gini = 0.375 samples = 4 value = [3, 1]
gini = 0.0 samples = 3 value = [3, 0]
gini = 0.0 samples = 2 value = [0, 2]
gini = 0.0 samples = 5 value = [5, 0]
X[0] <= 0.04 gini = 0.42 samples = 10 value = [7, 3]
gini = 0.0 samples = 1 value = [1, 0]
gini = 0.0 samples = 1 value = [0, 1]
X[1] <= 10.583 gini = 0.5 samples = 2 value = [1, 1]
gini = 0.0 samples = 3 value = [0, 3]
gini = 0.0 samples = 1 value = [0, 1]
gini = 0.0 samples = 3 value = [3, 0]
X[1] <= 20.34 gini = 0.245 samples = 7 value = [6, 1]
X[1] <= 20.198 gini = 0.444 samples = 3 value = [1, 2]
gini = 0.0 samples = 1 value = [0, 1]
gini = 0.0 samples = 1 value = [1, 0]
gini = 0.0 samples = 5 value = [5, 0]
X[1] <= 20.395 gini = 0.5 samples = 2 value = [1, 1]
gini = 0.0 samples = 2 value = [0, 2]
gini = 0.0 samples = 1 value = [1, 0]
gini = 0.0 samples = 1 value = [0, 1]
gini = 0.0 samples = 1 value = [1, 0]

To correct the overfit, variations on several hyperparameters were explored. The default criterion for the function is Gini Index reduction. Changing this criterion to focus on Entropy instead made little difference.

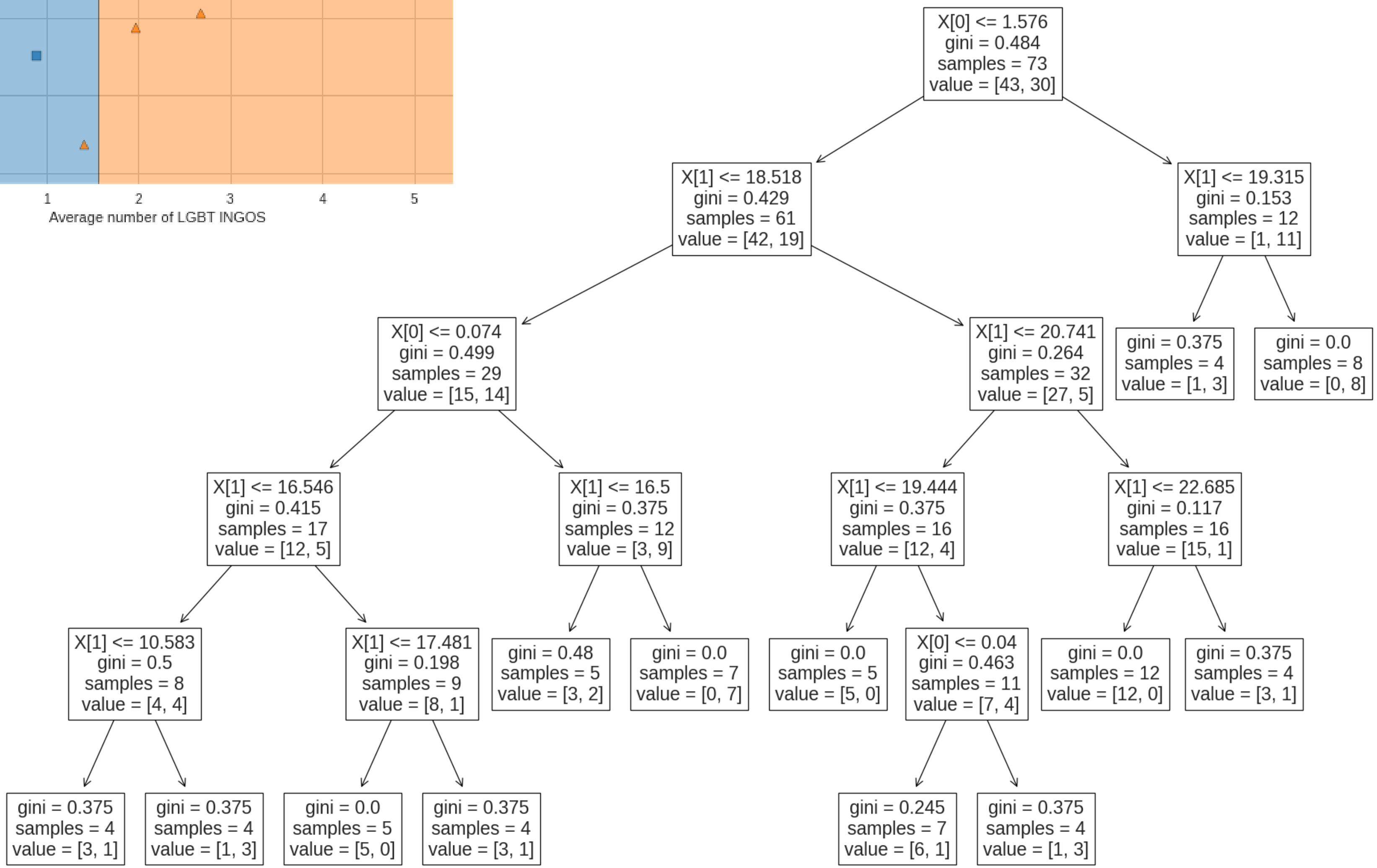
Limitations on the minimum number of samples needed to split or to form a leaf were tried, as were limitations on the maximum number of leaves and minimal purity increase at each branch. While there were subtle differences between these models, the AUC remained between 0.6 and 0.66%. Returning the function to default values, but requiring at least 4 samples per leaf, yielded an AUC of 0.71.

This also made the tree shorter and simpler to describe (see overleaf). Variations on the model with limitations on the maximum number of branches, to further simplify the tree, dramatically worsened the accuracy of the model.






Decision Surface of Second Decision Tree Model
0
1
Average number of OECD Donors
20
15
10
5
0
-1
0
1
2
3
4
5
Average number of LGBT INGOS


X[0] <= 1.576
gini = 0.484
samples = 73
value = [43, 30]
X[1] <= 18.518
gini = 0.429
samples = 61
value = [42, 19]
X[1] <= 19.315
gini = 0.153
samples = 12
value = [1, 11]
X[0] <= 0.074
gini = 0.499
samples = 29
value = [15, 14]
X[1] <= 20.741
gini = 0.264
samples = 32
value = [27, 5]
gini = 0.375
samples = 4
value = [1, 3]
gini = 0.0
samples = 8
value = [0, 8]
X[1] <= 16.546
gini = 0.415
samples = 17
value = [12, 5]
X[1] <= 16.5
gini = 0.375
samples = 12
value = [3, 9]
X[1] <= 19.444
gini = 0.375
samples = 16
value = [12, 4]
X[1] <= 22.685
gini = 0.117
samples = 16
value = [15, 1]
X[1] <= 10.583
gini = 0.5
samples = 8
value = [4, 4]
X[1] <= 17.481
gini = 0.198
samples = 9
value = [8, 1]
gini = 0.48
samples = 5
value = [3, 2]
gini = 0.0
samples = 7
value = [0, 7]
gini = 0.0
samples = 5
value = [5, 0]
X[0] <= 0.04
gini = 0.463
samples = 11
value = [7, 4]
gini = 0.0
samples = 12
value = [12, 0]
gini = 0.375
samples = 4
value = [3, 1]
gini = 0.375
samples = 4
value = [3, 1]
gini = 0.375
samples = 4
value = [1, 3]
gini = 0.0
samples = 5
value = [5, 0]
gini = 0.375
samples = 4
value = [3, 1]
gini = 0.245
samples = 7
value = [6, 1]
gini = 0.375
samples = 4
value = [1, 3]

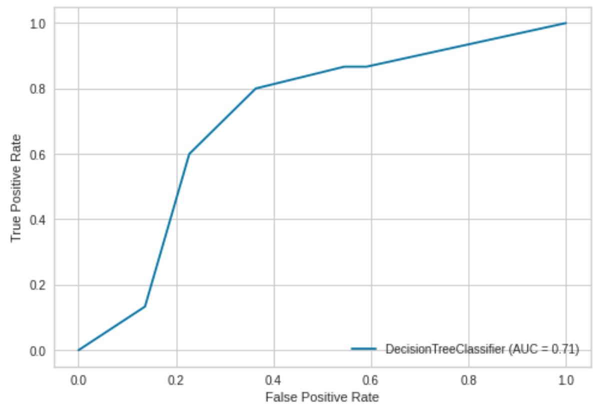
3 Comparison of Models

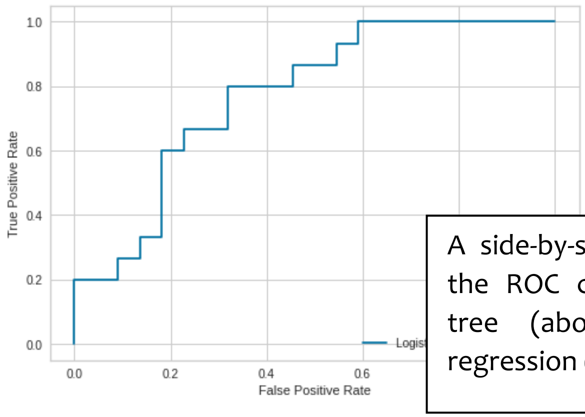
As the models were trained on a relatively small dataset containing only 77 samples, execution speed was not a significant criterion when selecting them. Similarly, as this is not a problem that is likely to require significant upscaling, speed is one of the less important metrics for comparison. That said, Linear Regression was the faster model for both fitting and prediction.

Considering the extreme reduction in dimensionality from Velasco's dataset, and considering also the intensely more complex political and social factors at work in the global landscape during the period in question, the predictive power of these simple models points towards a significant set of correlations.

The linear regression model was the most accurate of the categorisation models. Whilst all the models were fully mappable and transparent, the decision tree's output is less intuitive to read. All models struggled to predict the values of countries whose LGBT legislative score did not change in the period, though the decision tree fared a little better with these. However, the two other models' description of a general trend seems the more useful communicative device. It also seems likely that the "clusters" identified by the tree might be broken up if other independent values were introduced to the models.

	Linear Regression	Logistic Regression	Decision Tree
Training Speed	0.211	0.26	0.25
Prediction Speed	0.208s	0.267	0.214
Accuracy	MSE ≈ 0.00028% RMSE ≈ 0.0167%	AUC = 0.77	AUC = 0.71
Transparency	Fully mappable	Fully mappable	Fully mappable





A side-by-side comparison of the ROC charts for decision tree (above) and logistic regression (below) models.

4 Conclusions

Each of these three models makes sacrifices of predictive capability to improve comprehensibility and transparency. Despite this, yearly averages for LGBT INGOS and OECD donors have proven good predictors of a country's year on year legislative change.

The introduction of other variables might help models deal with the many countries that have not enacted legislative change but would make models it is harder to visualise or explain in two dimensions. Animation, however, might suggest ways of communicating these higher dimensional models.

Though Velasco's paper ends with a caution about the differing outcomes that state aid and LGBT INGOs suggest, it is worth remembering that correlation is not causation. Though a higher presence of LGBT INGOs is tied to increased LGBT legislative progress, we have not shown which of these variables (if either) precedes the other. The models do, however, concur with Velasco's finding that higher levels of OECD donor relationships are likely to inspire worse legislative movement in countries that lack LGBT INGOs.

5 References

[1] K. Velasco, 'A Growing Queer Divide: The Divergence between Transnational Advocacy Networks and Foreign Aid in Diffusing LGBT Policies', *Int. Stud. Q.*, p. sqz075, Aug. 2019, doi: 10.1093/isq/sqz075.

[2] K. Velasco, 'Replication Data for: A Growing Queer Divide: The Divergence between Transnational Advocacy Networks and Foreign Aid in Diffusing LGBT Policies'. Mar. 13, 2021. doi: 10.7910/DVN/XINDPB.

[3] Andriy Burkov, 'The Hundred-Page Machine Learning Book'. http://themlbook.com/ (accessed May 20, 2021).

[4] J. Brownlee, 'A Gentle Introduction to Imbalanced Classification', *Machine Learning Mastery*, Dec. 22, 2019. https://machinelearningmastery.com/what-is-imbalanced-classification/ (accessed May 20, 2021).

[5] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*, vol. 103. New York, NY: Springer New York, 2013. doi: 10.1007/978-1-4614-7138-7.

[6] jamesdmccaffrey, 'Logistic Regression and Linear Separability', *James D. McCaffrey*, Dec. 26, 2019. https://jamesmccaffrey.wordpress.com/2019/12/26/logistic-regression-and-linear-separability/ (accessed May 27, 2021).

[7] F. Pedregosa *et al.*, 'Scikit-learn: Machine Learning in Python', *J. Mach. Learn. Res.*, vol. 12, no. 85, pp. 2825–2830, 2011.

[8] B. Bengfort *et al.*, *Yellowbrick v1.3*. Zenodo, 2021. doi: 10.5281/zenodo.4525724.

[9] 'seaborn: statistical data visualization — seaborn 0.11.1 documentation'. https://seaborn.pydata.org/ (accessed Jul. 04, 2021).

[10] J. D. Hunter, 'Matplotlib: A 2D Graphics Environment', *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 90–95, May 2007, doi: 10.1109/MCSE.2007.55.