Jupiter – PBA Report

Dataset – Investment Treaty Cases

- UN Website – UNCTAD
- Data on 640 Decided and 341 Pending Investor State Cases
- Investor State Case – what is involved?
- Second UN Website – Human Development Index
- Geomteric mean – measures Hdi between 0 and 1 (most develoepd country)

Problem Definition

- Is it possible to build models on identified key features from the chosen UN datasets which has some predictive power as to case outcome?
- Or is factual review of each case critical?
- Possible features to consider:
  - HDI Score
  - Identity of Judges (Arbitrators)
  - Claim for Direct Expropriation?
  - Industrial Sector, e.g. Electricity, Mining
  - Size of Claim

Data Preparation (1)

- Data Extraction – web scraping

- Data Cleaning:
  - Amount Fields
  - Arbitrators
  - Home / Claimant State Field
  - Update HDI

Data Preparation (2)

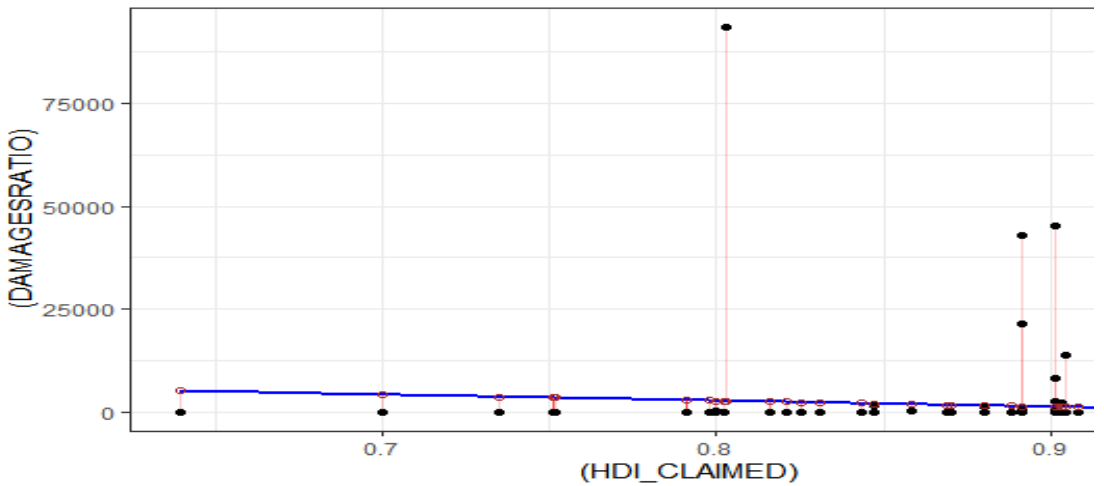
- Model Features – Final Variable Selection in Categorical Models
  - Arbitral Rules – 1 hot
  - Breaches Alleged – Direct Expropriation – binary
  - Economic Sector – 1 hot
  - Arbitrators (all 1 hot):
    - President
    - Claimant
    - Respondent
  - Claimant /Respondent HDI – continuous
  - Success – binary – Settled or Award in favour of investor

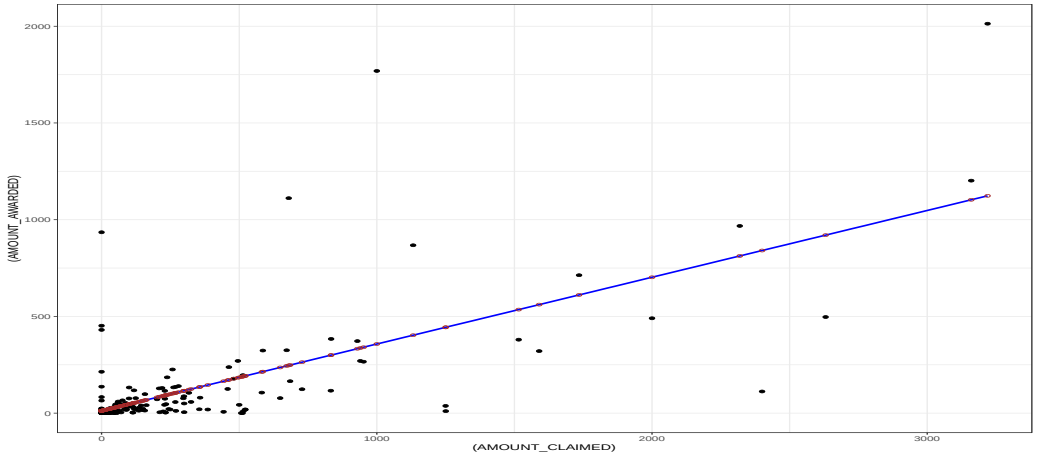
Modelling Approaches

- Mixed data set – continuous and discrete categorical
- Regression on continuous data:
  - Correlation between Claimant/Respondent HDI and the amount of damages?
  - Renormalise Awarded Amount by dividing by Claimed Amount to give Damages Ratio, then rerun.
  - Correlation between damages claimed and damages awarded?
- Classification methods on categorical data, to model probability of successful case outcome, using:
  - Decision Tree; Boosted DT; Random Forest; Deep Neural Network; Multi-layer perceptron.
  - 6-fold validation (450/90)

Model Output (1)

- Regression on continuous data of 181 cases where damages were awarded:
  - Essentially zero correlation between Claimant and Respondent HDI and amount of damages or damages normalised by claim amount (shown opposite).
  - Moderate Correlation between damages claimed and damages awarded (slope 0.34, R-squared 0.48). (Shown opposite)

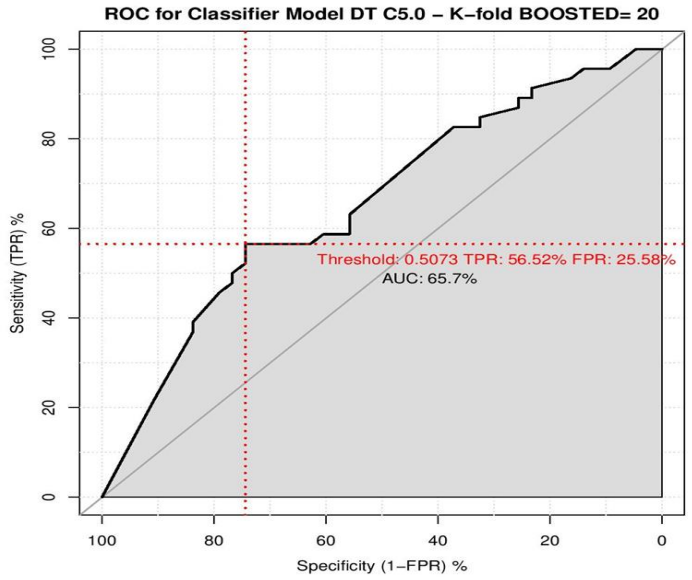


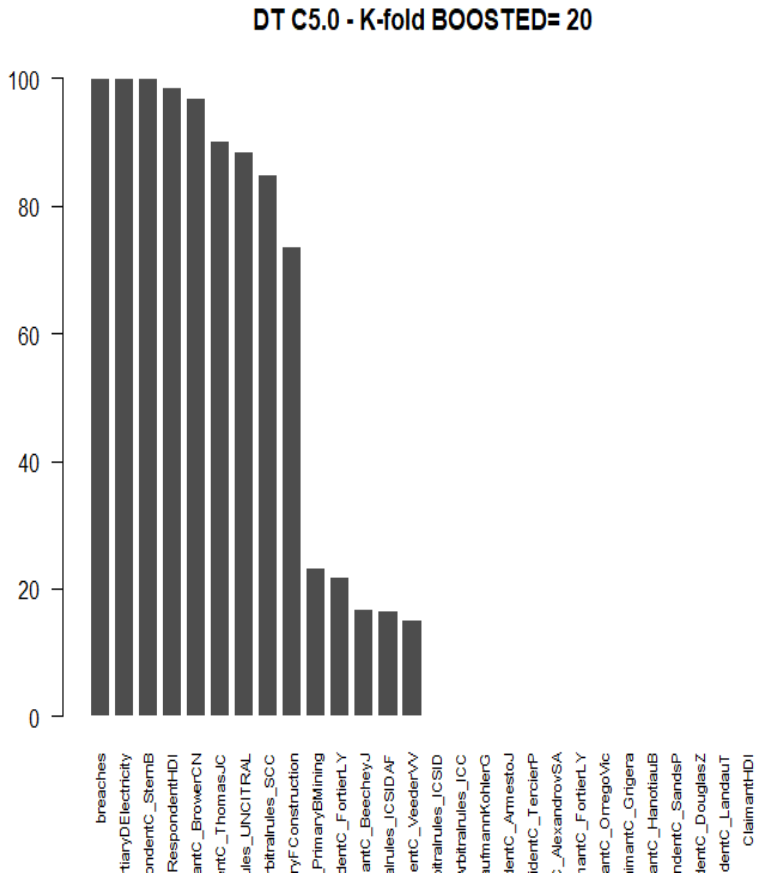


Model Output (2)

- Categorical and continuous data – results for DT Boosted in predicting case outcome using 28 variables:

Measure	Averaged Value
TP	25
FN	20
TN	30
FP	13
Accuracy	62.06
pgood	66.39
pbad	59.10
FPR	30.73
TPR	55.38
TNR	69.27
MCC	0.251
Threshold	0.54





Model Assessment

	TP	FN	TN	FP	accuracy	pgood	pbad	FPR	TPR	TNR	MCC	threshold	varGood	varAccuracy	varMCC	folds
RandomForest	27	19	30	12	64.27	67.94	61.58	29.22	58.26	70.78	0.29	0.54	15.59	27.55	0.01	6
DT_boost	25	20	30	13	62.06	66.39	59.10	30.73	55.38	69.27	0.25	0.54	27.16	17.55	0.01	6
DT_Kfold	26	19	28	14	61.47	64.55	58.95	34.28	57.53	65.72	0.23	0.33	14.35	9.19	0.00	6
MLP	25	21	29	13	61.12	65.08	58.17	31.54	54.29	68.46	0.23	0.51	18.61	13.31	0.01	6
DT_Holdout	36	21	30	21	61.11	63.16	58.82	41.18	63.16	58.82	0.22	0.31	0.00	0.00	0.00	6
Deep_Neural	27	19	24	19	57.78	59.49	56.85	43.83	59.28	56.17	0.16	0.52	8.36	9.79	0.00	6

- Random Forest / DT (Boosted) performed the best on most measures.
- Each had an MCC in excess of 0.25, and the highest accuracy rate
- They also had the greatest AUC in the ROC curves indicating superior performance in TPR vs FPR.
- Only in variance did Random Forest and DT Boosted (slightly) underperform relative to Deep Neural Network and MLP.
- Importantly, Random Forest and DT Boosted had the highest win-rate (the p-good measure) i.e. each successfully predicted winning cases approx. 66% of the time in the test samples of 90 cases.

Project Evaluation and Future Deployment

- Key Results:
  - Hypothesis - States with a lower HDI ranking are unfairly treated (and vice versa for claimants in higher HDI states) – no evidence;
  - Hypothesis - Damages claimed is positively correlated to damages awarded – some evidence;
  - Hypothesis - individual factual case analysis should be determinative when deciding on the likely outcome of a case – some evidence to reject partly.

    DT Boosted/RF focused on a limited number of exogenous variables appeared able to predict winning cases with approximately 66% probability versus (88 case) sample mean of 51.8%. Appeared significant at 0.01 level given probability of getting 66% success cases out of random sample of 88 cases.
- Future:
  - deploy the various models against the pending cases (341 cases)
  - isolate the impact specific features have on case outcome, e.g. does a specific arbitrator or combination of arbitrators make a difference?
  - combine existing model with a factual enquiry into case sample, reflecting additional factual features proposed by domain experts, then build improved model.