Misconduct in World Bank Contracts

Carlos Grandet | Héctor Salvador | Santiago Matallana

06.06.2016

https://github.com/smatallana/MLearning_WB_project

Introduction

- Fraud, corruption, and collusion represent $900 billion/year
- WB awards 20K-30K contracts/year with a value of ~$20 bn
- WB relies on the complaints of whistleblowers
- In 2015 alone, WB substantiated investigations on 93 contracts worth more than $500 million, of which $140 were not awarded because the wrongdoing was detected prior to contract award
- We want to help better allocate scarce investigative resources

Related work

- Summer of 2015: DSSG team
- Major DSSG team challenge: clean and merge contracts, projects, and investigations datasets (Bulk of the team's efforts)
- We built on the work of the DSSG team, from their final dataset, with a focus on:
  - Feature generation (≈20 vs. ≈160)
  - Perfecting evaluation: expected loss of wrongdoing + precision

General description of approach

- Explored original data to inform feature generation process
- Built a loop of machine learning classifiers and used it initially on “raw” DSSG 2015 clean/merged dataset
- Generated near 160 features, and reduced to 23 with linear SVM with lasso (in at least 20 of 100 iterations)
- Ran ML loop with new dataset, with train/test subsets after chronological ordering to prevent “predicting from the future”
- Calculated “expected value” (or expected loss/damage) per contract and prioritized accordingly
- Evaluated models using precision on top *n*% of the contracts, in line with resource availability

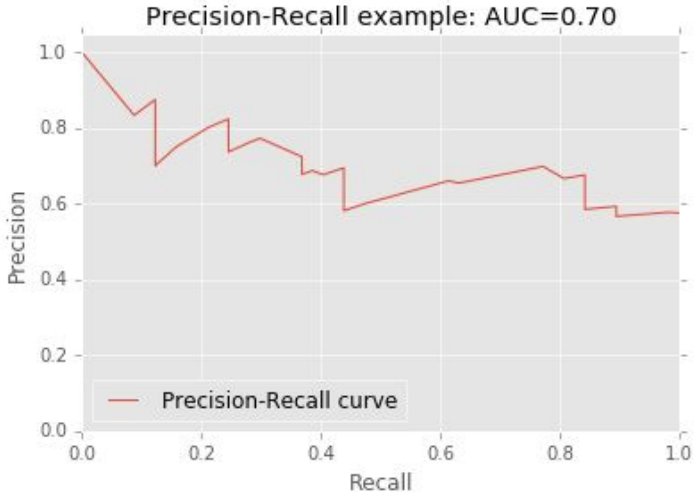
Data description

- DSSG team worked originally with three databases from the World Bank:
  - Contracts (+200K)
  - Projects
  - Investigations
- Cleaned and merged dataset:
  - 1,324 contracts with complaints
    - 567 labelled investigations: “substantiated”, “unsubstantiated”

Model proposal

- ML looped through LR, DT, LSVM, NB, GB, and RF models
- Preferred model:

RF	0.24	0.10	0.50	1	{'max_depth': 10, 'min_samples_split': 2, 'max_features': 'log2', 'n_estimators': 10}



Model evaluation

- Without constraints, detect all contracts with misconduct
- Given limited resources, we maximized precision (intuitively, the ability of the model not to label as fraudulent a contract that is OK)
- We evaluated the precision of our models assuming that the top *n%* of a ranked version of the contracts can be investigated
- We ranked contracts by their “expected value” :
  EV = (predicted probability of misconduct) * (contract value)

Results and policy recommendation

ID	yscore	expected_value ↓
226	1.00	$ 259,484,347
45	0.96	$ 3,418,803
47	0.96	$ 2,445,325
44	0.96	$ 1,857,527
46	0.96	$ 938,200
134	1.00	$ 812,488
199	0.97	$ 411,984
198	0.97	$ 411,984
324	0.80	$ 389,514
83	1.00	$ 340,864

- Prioritize investigation of reported contracts according to our ranked list: “prefer to investigate a millionaire contract over a low value one, even if the latter has a higher predicted probability of misconduct”

- Run A/B tests on future investigations

Limitations and future work

- Limitations:
  - Analysis is restricted to reported contracts. There are potentially other contracts with misconduct that are never reported.
  - Train/test splits are done on the basis of contract award dates. We do not know if these are consistent with investigation dates. (We could, in fact, be training the past with data from the future.)
- Future work:
  - KNN classifiers to incorporate more features
  - Experimentation with a randomized control trial in the field
  - Loop over probability threshold and top *n*% of contracts to maximize precision and AUC
  - Refine model to separately predict fraud and collusion (vs. “misconduct”)

Misconduct in World Bank Contracts

Carlos Grandet | Héctor Salvador | Santiago Matallana

06.06.2016

https://github.com/smatallana/MLearning_WB_project