



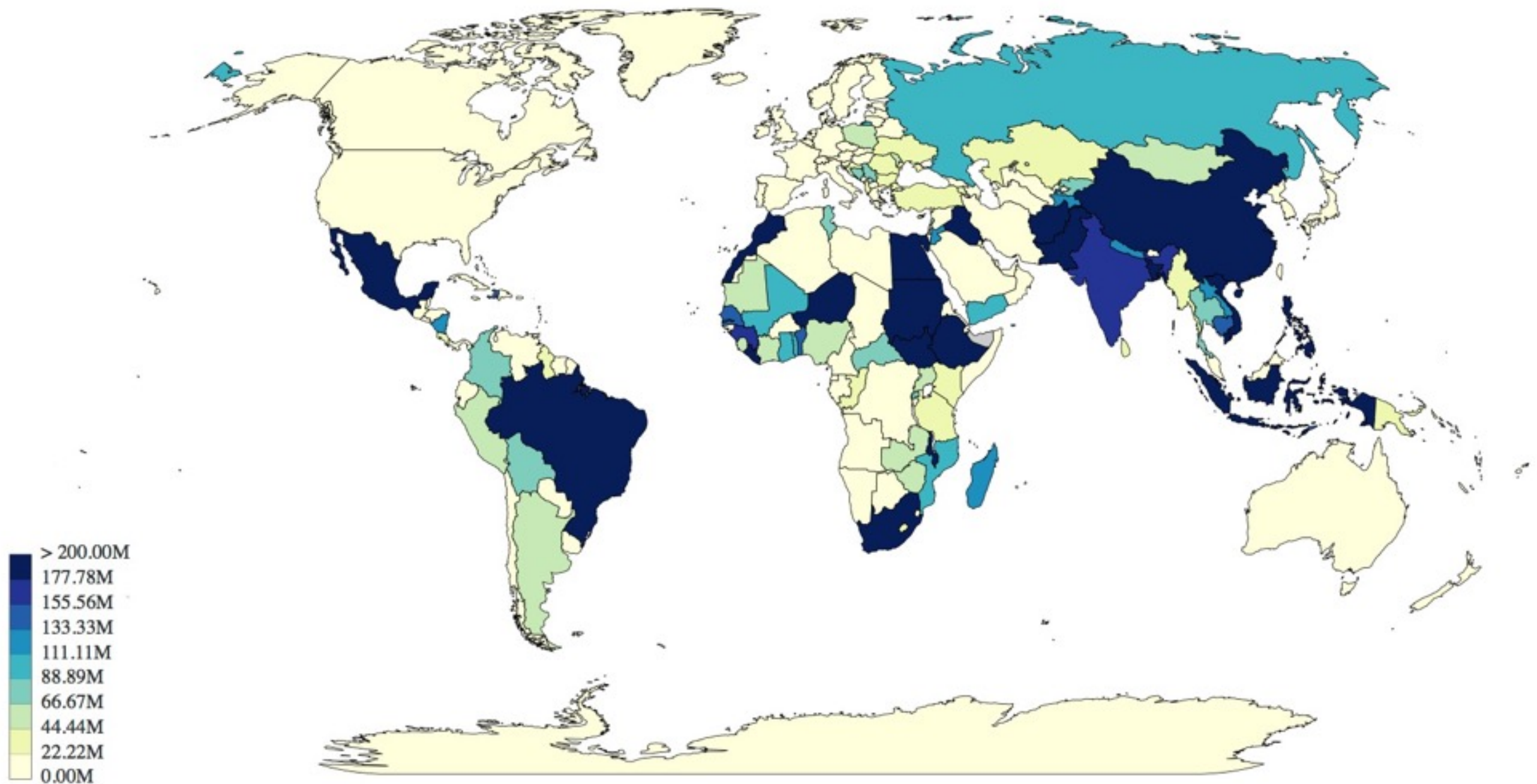
# Fraud Prediction in World Bank Projects

*by Yulia Piterbarg*



# Global effort to help weak economies

**Total Financing by World Bank since 1998**





# World Bank Projects

## Haiti:

**Cholera Emergency Response,  
US\$15 Million Grant, 2010**





# World Bank Projects

## Haiti:

Cholera Emergency Response,  
US\$15 Million Grant, 2010



## Indonesia,

Coral Reef Rehabilitation,  
US\$50 Million Loan, 2014

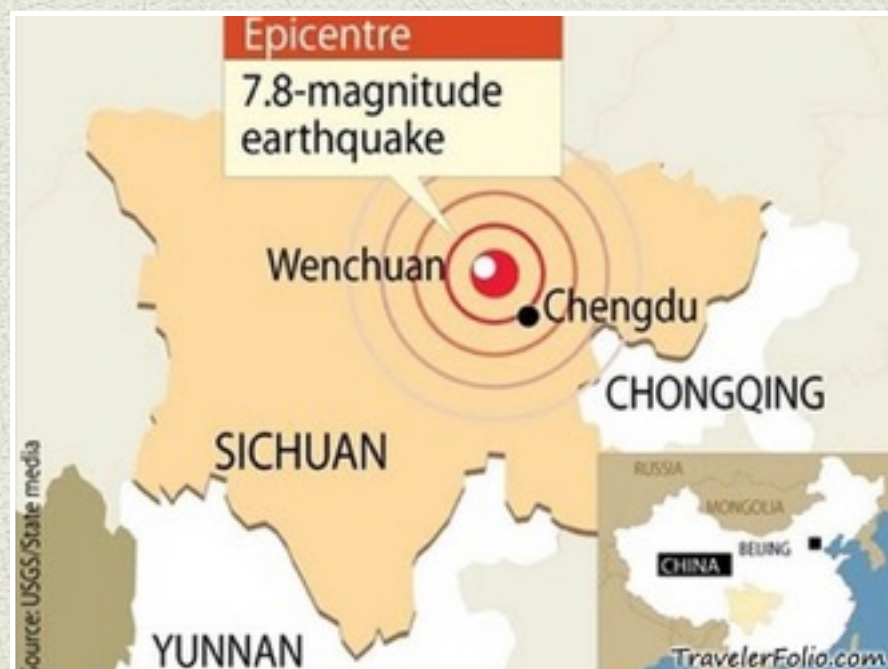




# World Bank Projects

## Haiti:

Cholera Emergency Response,  
US\$15 Million Grant, 2010



## Indonesia,

Coral Reef Rehabilitation,  
US\$50 Million Loan, 2014



## China:

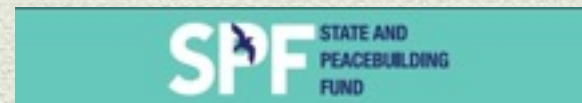
Earthquake Recovery Project,  
\$710 Million Loan, 2009



# Data:

## Obtain Fraud Information

- Identify a list of debarred firms and individuals.
  - Obtain pdf docs, Notices of Sanctioned Proceedings.
- Challenge 1: connect companies to projects
- Challenge 2: tesseract (ocr), regex





# Feature Engineering

*Project Information Document:  
health, transportation, education,  
energy, etc*

*International vs national bidding,  
competitive vs non-competitive*

*Construction, materials,  
consulting services*

*GNI per capita, poverty index,  
corruption index*

*IBRD, IRA, Trust Funds*

**LDA Topics**

**Bidding Format**

**Service Category**

**Economic Factors**

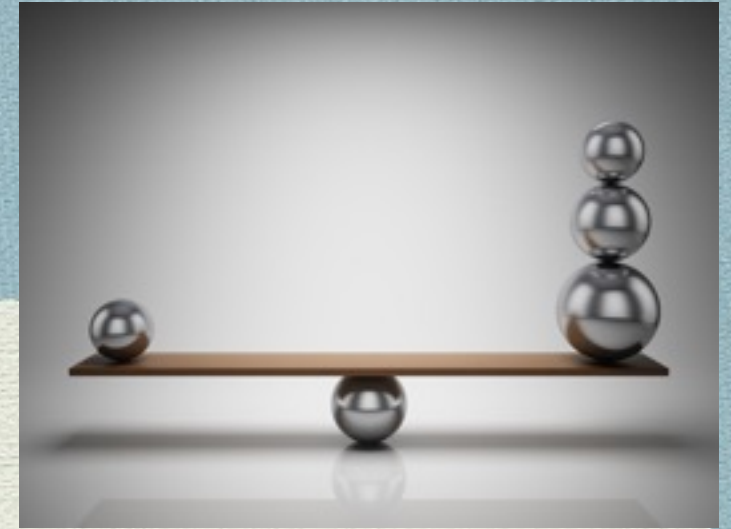
**Lending Source**

***Model***

```
graph LR; LDA[LDA Topics] --- V[ ]; Bidding[Bidding Format] --- V; Service[Service Category] --- V; Economic[Economic Factors] --- V; Lending[Lending Source] --- V; V --- Model((Model));
```



# Modeling



## ◆ Challenge:

- 89 projects with fraud vs 2100 successful projects.

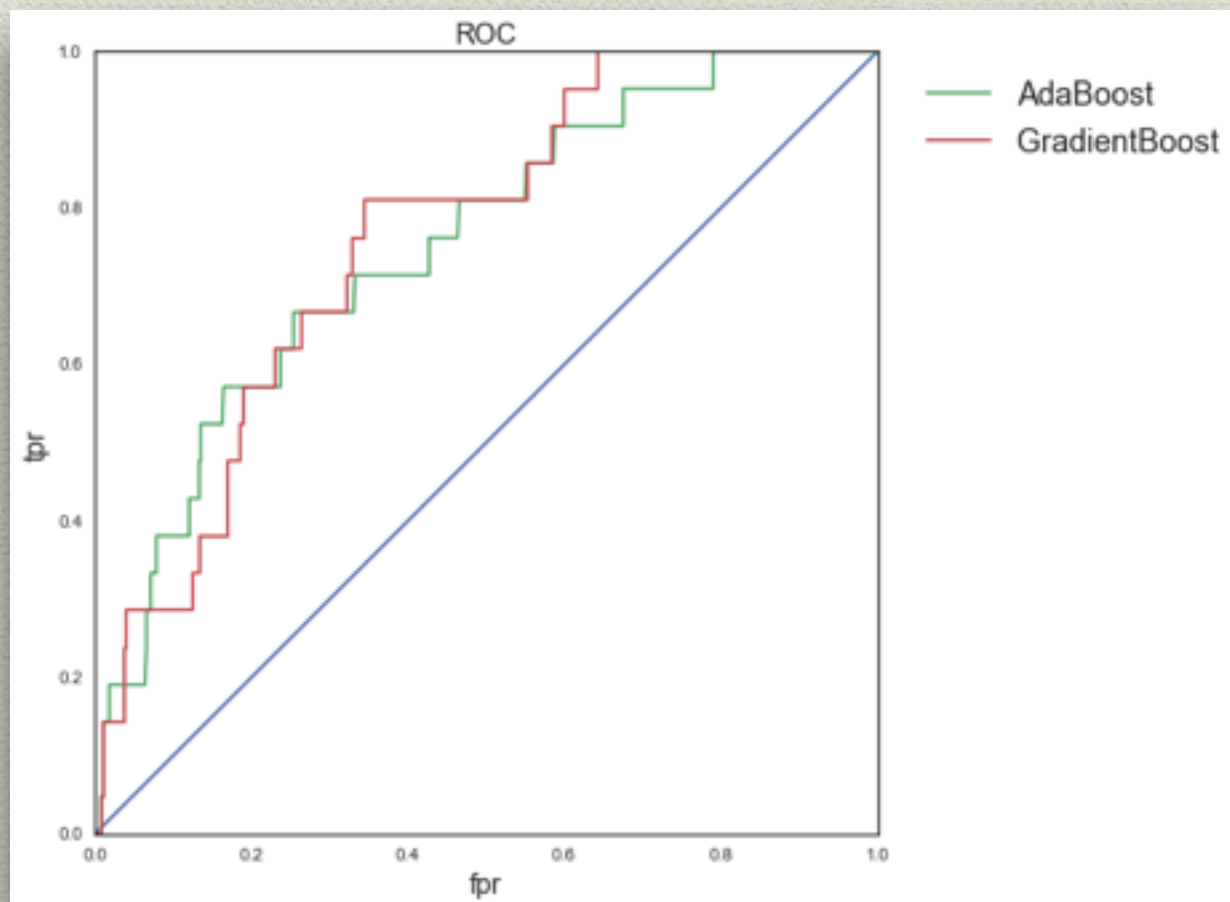
## ◆ Solution:

- Oversample minority class and / or undersample majority class.
- Use class weights, boosting methods - AdaBoost and BalanceCascade
- Recall Scoring - percentage of correctly identified “trouble” projects



# Analysis and Conclusion

	RandomOverSampler + AdaBoost	EasyEnsemble + Adaboost	EasyEnsemble + GradientBoost
Recall	0.10	0.62	0.81



## Important Features:

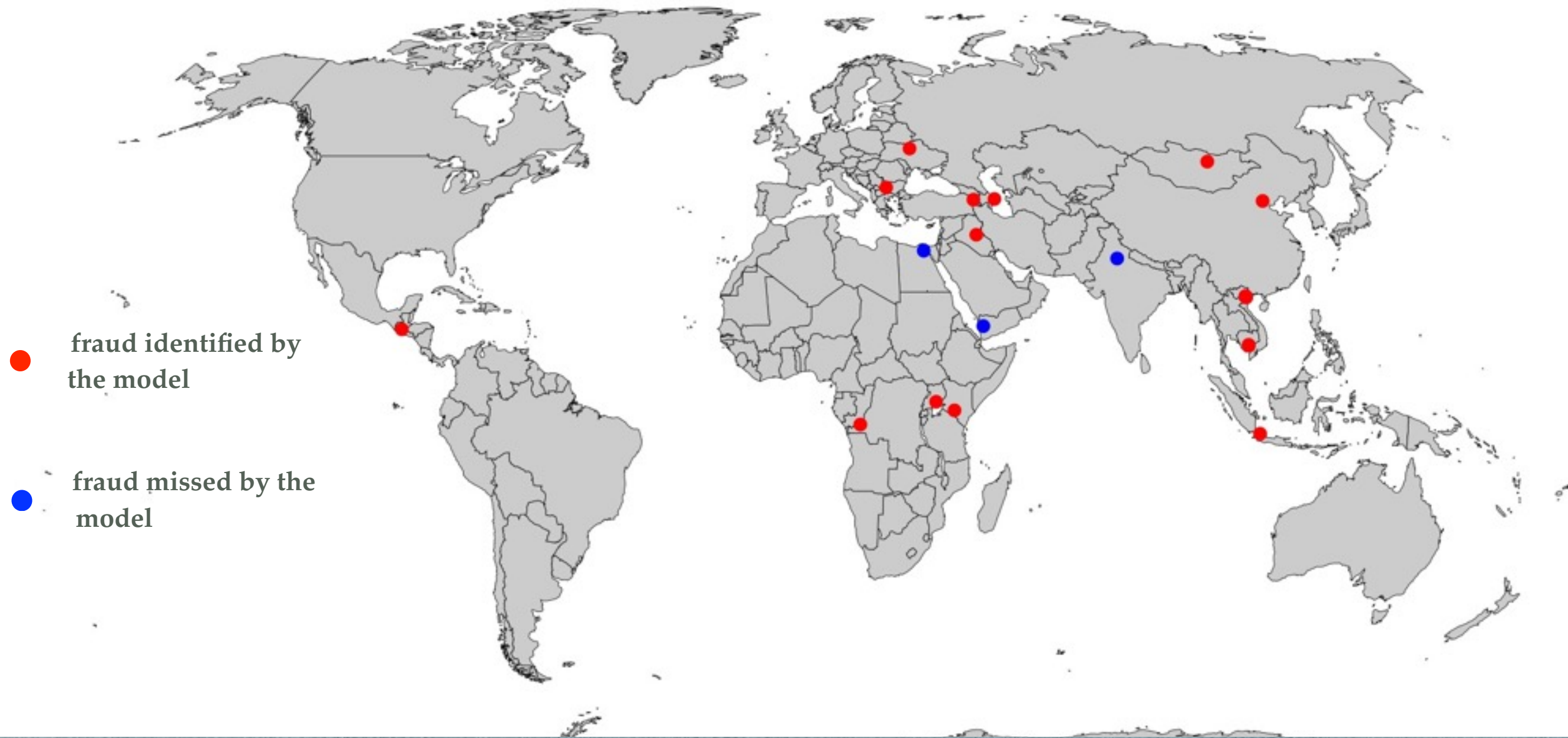
- topics
  - water&sanitation,
  - education,
  - climate change
- corruption rank
- projects with high number of consultants



# Analysis and Conclusion

## Results. Fraud Prediction.

//





“Perseverance is not a long race; it is many short races one after the other”

– *Walter Elliot*

[github.com/jpiter](https://github.com/jpiter)  
[linkedin.com/in/yulia-piterbarg](https://www.linkedin.com/in/yulia-piterbarg)

