

# Public Spending OBSERVATORY

**Risk of Public Contracts:  
Machine Learning + Multi-criteria Decision Analysis**

# Environment

## CORREIO BRAZILIENSE

### Company breaks the record on number of Governmental Contracts

Campeã de contratos de terceirização com o governo federal, a PH Serviços atrasa o pagamento a terceirizados

## DIARIO de PERNAMBUCO

Default epidemic in the Federal Government: 4 companies went bankrupt



MENU

G1

PIRACICABA E REGIÃO



Construction company abandons 3 projects

# Environment

Could we predict such situations?

Are there any features making possible to distinguish good from bad companies?



*Example:* Different registered activities per company

In Brazil: average of **1,99** registered activities

↳ Companies hired by the Government: **6,10**

↳ Defaulters: **11,61**

# Goals

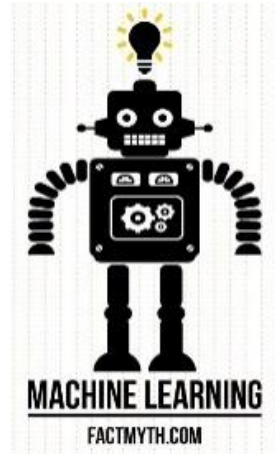
## 1. Classify companies & contracts (supervised learning)



**Supplier Risk Score**  
(Punishment risk)



**Contract Risk Score**  
(Termination/default risk)



## 2. Create decision model for auditing, including expert opinion.



### Logistical Issues

- Public agency has already been audited?
- Is it located at a Capital City?
- Does it require an “expert” to audit?

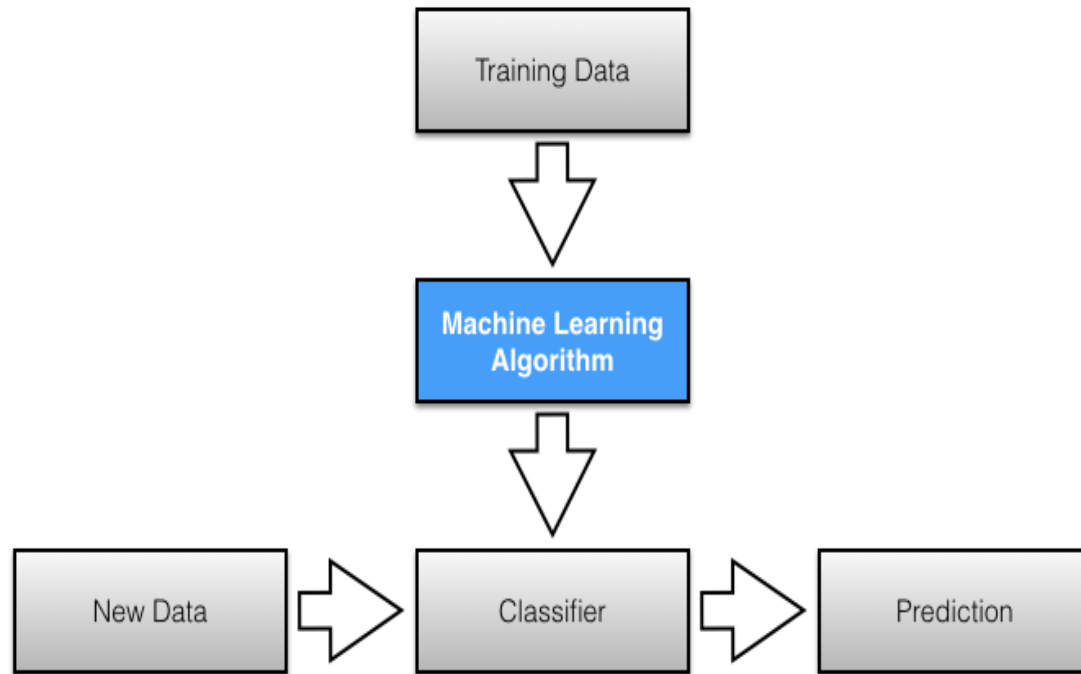
**MCDA**

**Multiple  
Criteria  
Decision  
Analysis**

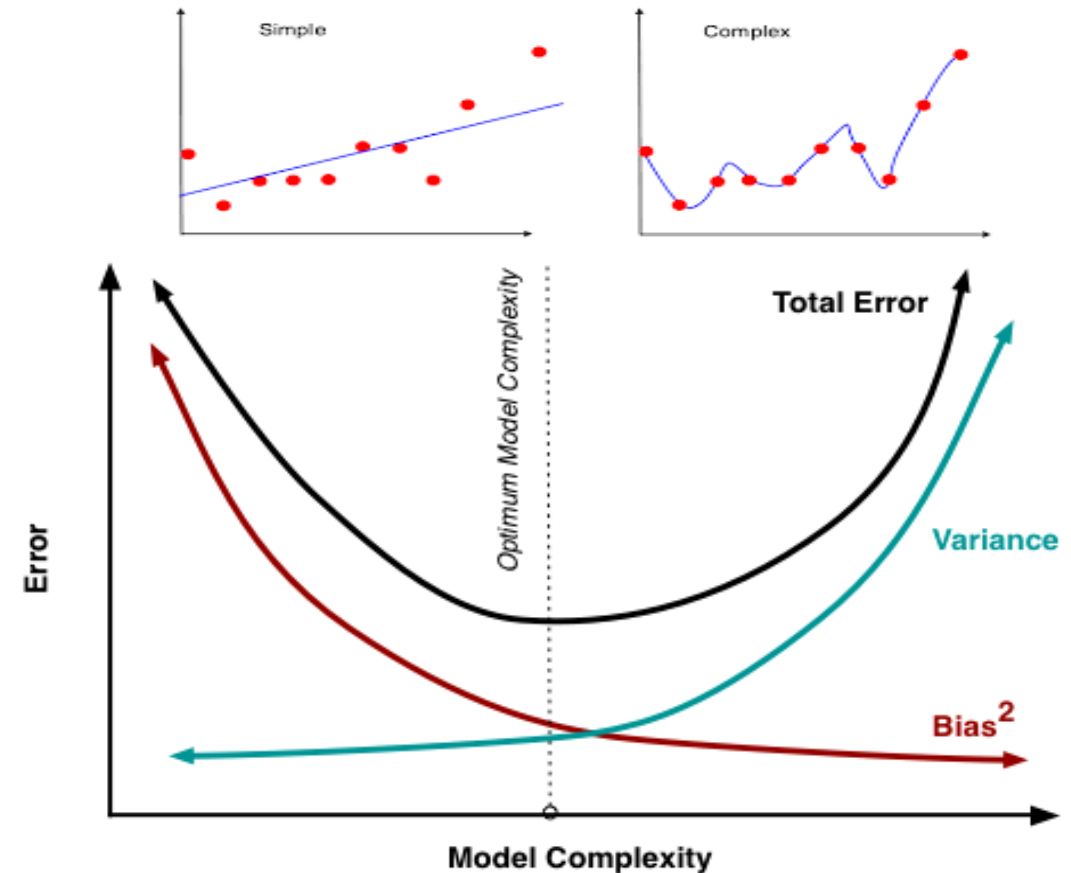
# Methods

## Supervised learning models

### Basic workflow of supervised learning



### Trade off - Bias x Variance





# Methods

How to choose the predictors? Economics!

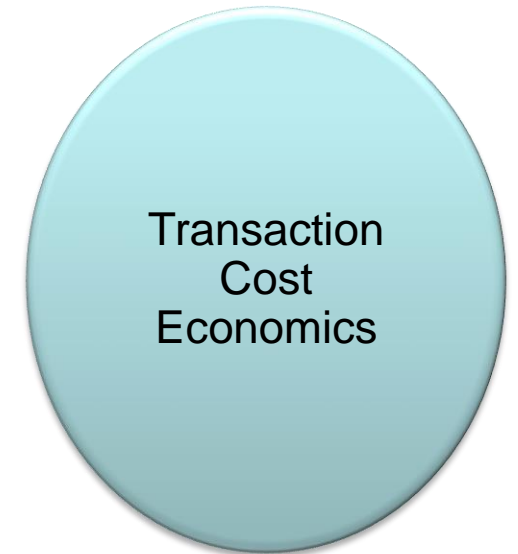


Qty. of bids  
Qty. of bidders  
Companies age



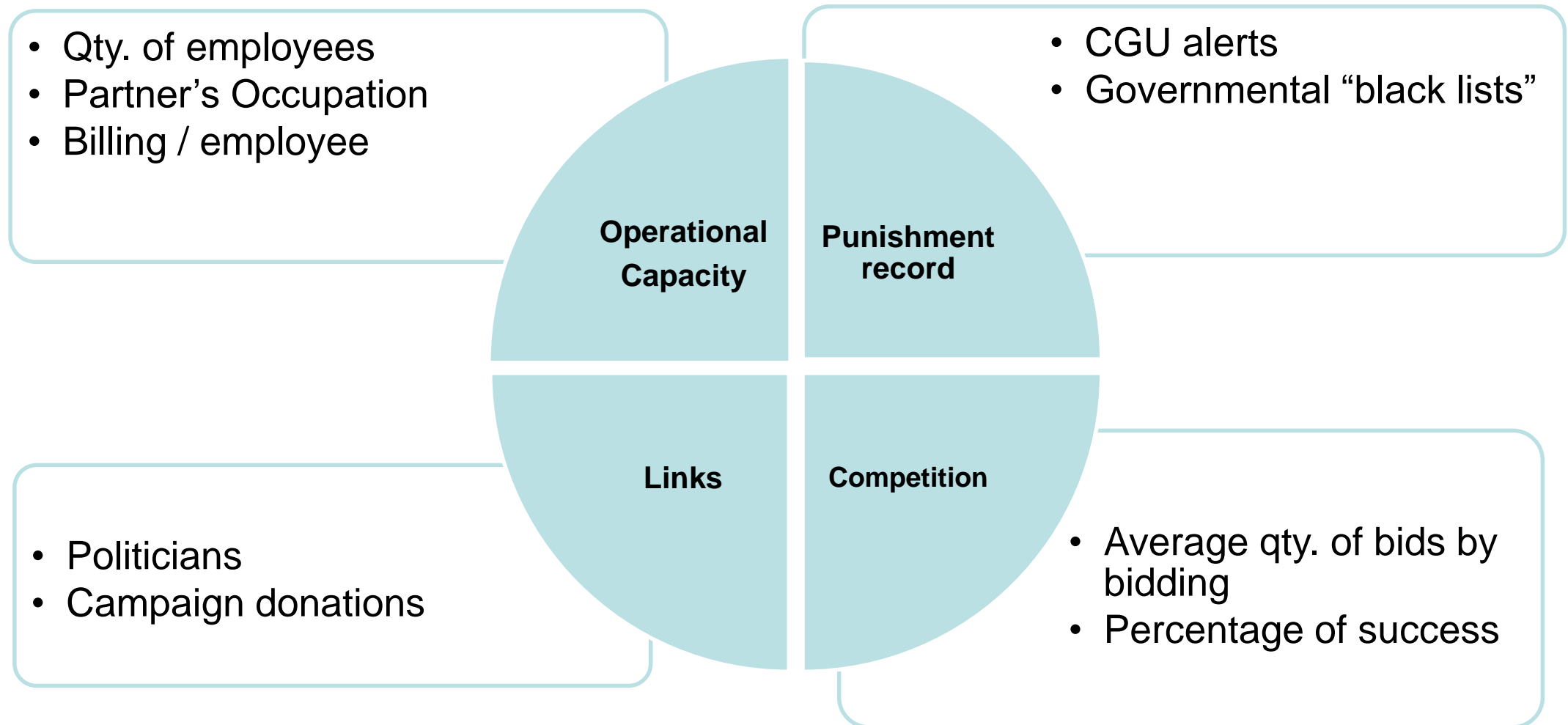
Frequency of the  
purchased item  
Frequency of bidders'  
participation

Specification quality  
Company size  
Product type  
Previous contract  
defaults  
Type of bidding  
Complexity of  
goods purchased



# Model 1: Supplier Risk

## 1<sup>st</sup> Phase: Identification of risk dimensions



# Model 1: Supplier Risk

## 2<sup>nd</sup> Phase: Creating Database

- A) 1446 companies:
  - a. 723 in the “High Risk” group
  - b. 723 in the “Low Risk” (**Under sampling**)
- B) 46 predictor variables
- C) 1 dependent variable (LABEL)

### Criteria for qualification as "High Risk":

- Having had an active contract in 2015 or 2016.
- Has been punished over that years with one of the following penalties:
  - Temporary suspension to bid (foreseen in Law #8666/93);
  - Impediment to bid and hire (foreseen in Law #10520/02);
  - Disreputable declaration (foreseen in Law #8666/93).

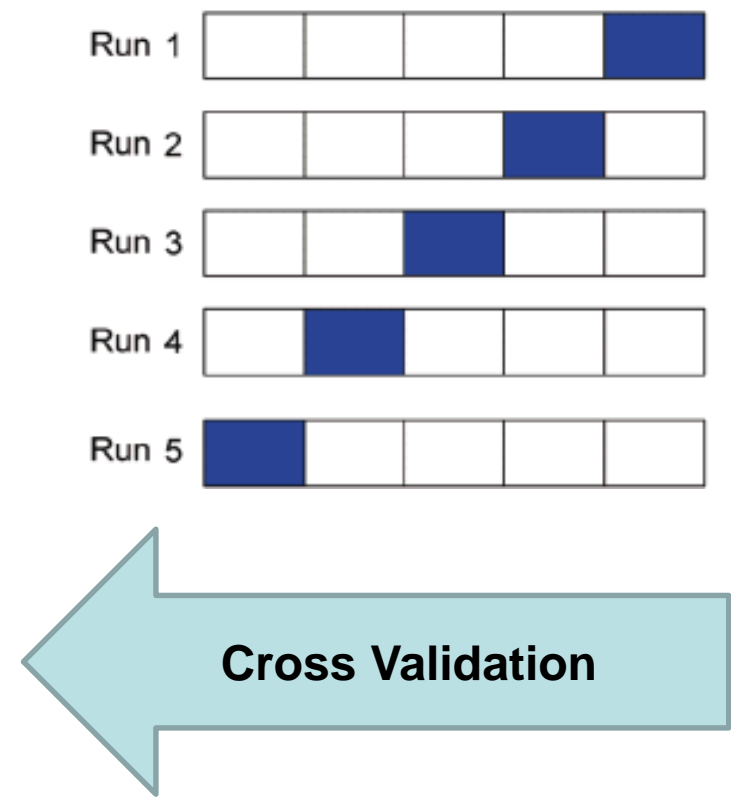
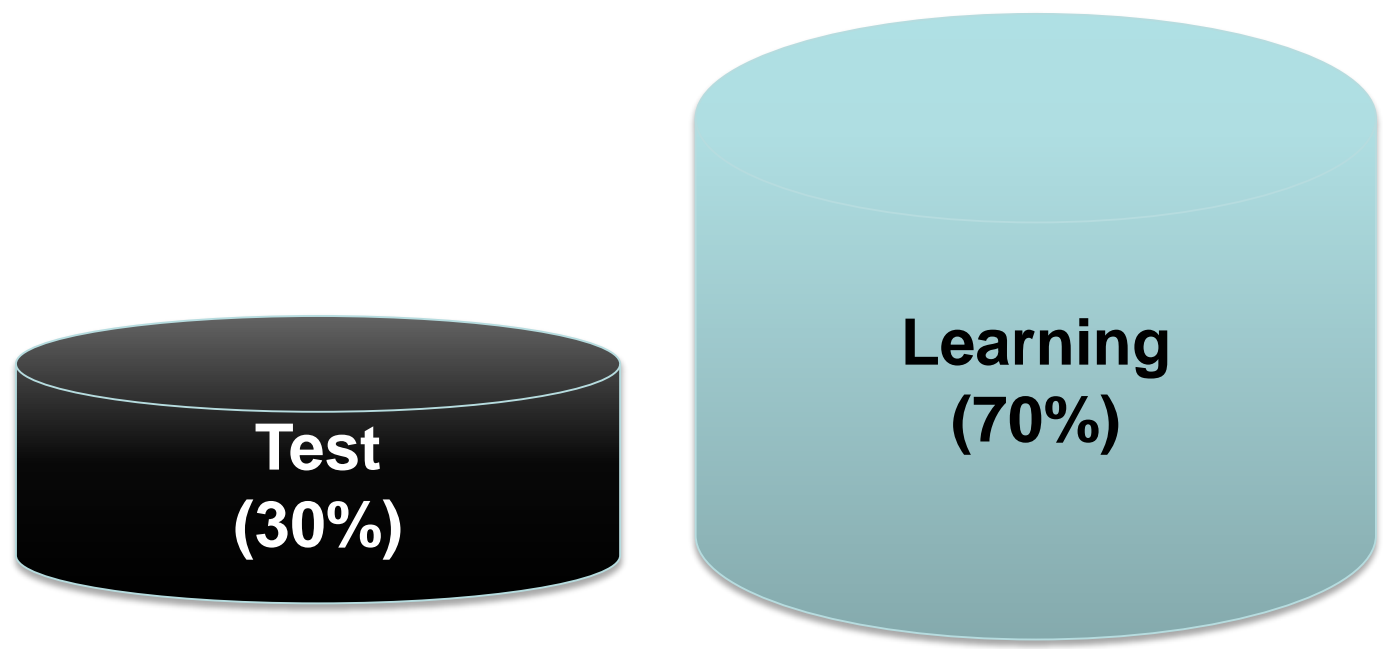




# Model 1: Supplier Risk

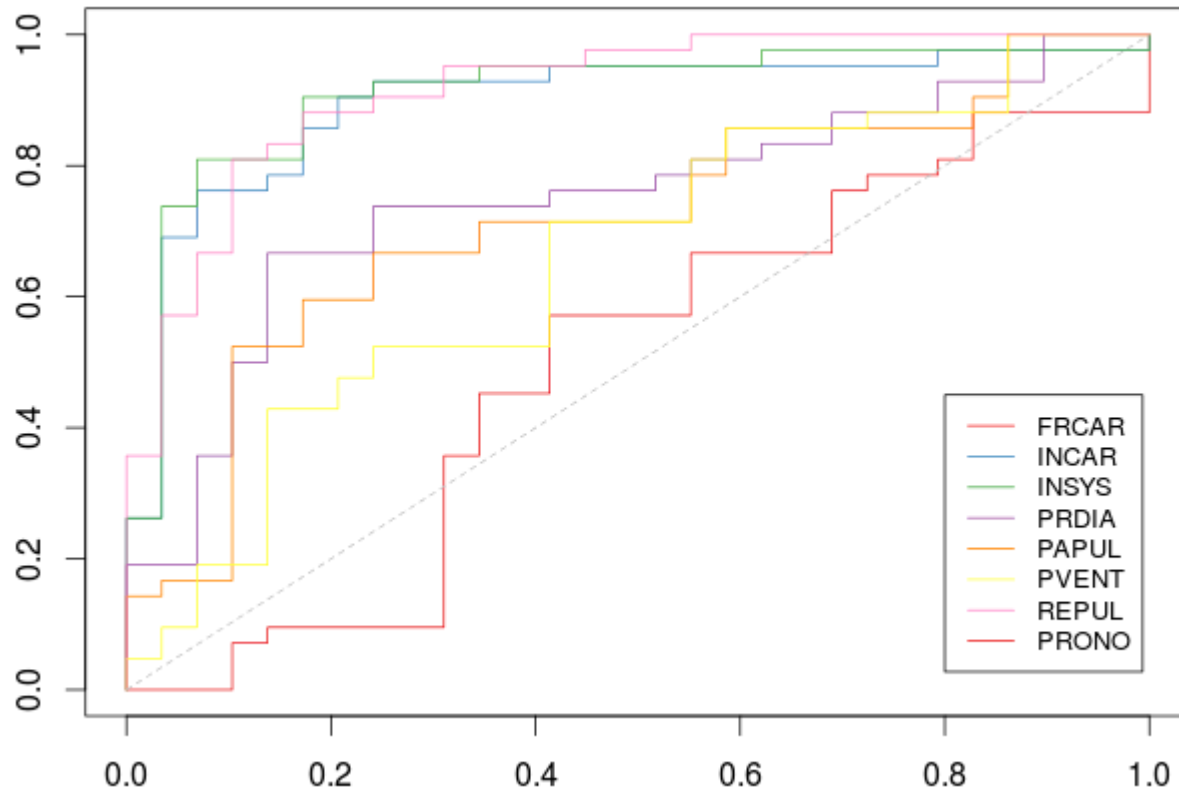
2<sup>nd</sup> Phase: Creating Database

D) Splitting in test and learning datasets



# Model 1: Supplier Risk

3<sup>rd</sup> Phase: Identifying the most important variables (Stepwise Algorithm)



Forward Stepwise

## 29 variables selected

Donated value on elections

Amount of activities carried out

Number of employees

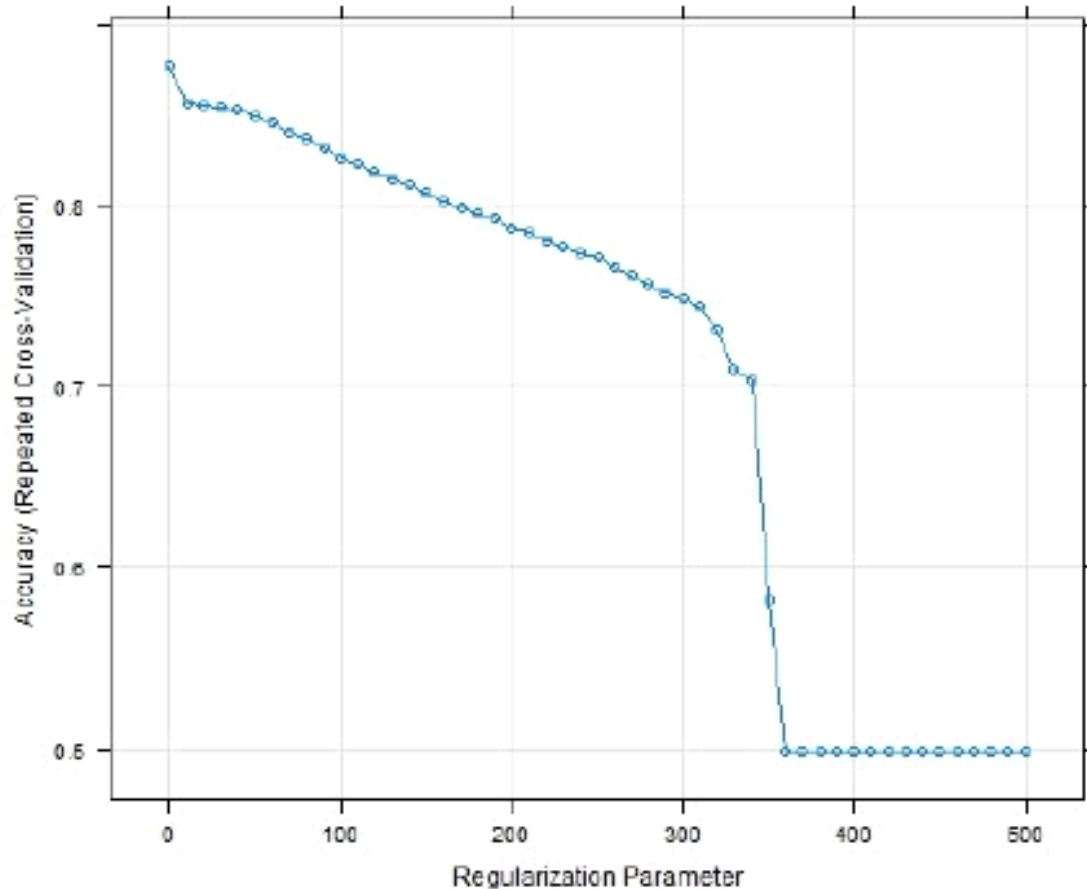
Partners' salary

Company age

CGU alerts

# Model 1: Supplier Risk

4<sup>th</sup> Phase: *Tuning*



Algorithm: GLMnet (Logistic Regression)  
Final penalty parameters:

$$\alpha = 0$$

$$\lambda = 1$$

**Cost Function =**

$$\underbrace{\sum_{i=1}^n \{y_i - \sum w_j x_{ij}\}^2}_{\text{RSS}} + \underbrace{\lambda (\alpha \sum |w_j| + (1 - \alpha) \sum w_j^2)}_{\text{Penalty parameters}}$$

# Model 1: Supplier Risk

5<sup>th</sup> Phase: Creating final model using the training database

6<sup>th</sup> Phase: Applying model in test database

## Results: Confusion Matrix

	Predict 1	Predict 0
1	208	9
0	54	163

**Accuracy:**  $(208+163)/(208+163+9+54) = 85.5\%$

**Sensitivity:**  $208/(208+9) = 95.9\%$

**Specificity:**  $163/(163+54) = 75.1\%$

**Precision:**  $208/(208+54) = 79.4\%$

## Results: Interpretation

Amount donated in elections



Registered Activities qty.



Employees qty.



Partners salary



Company's age

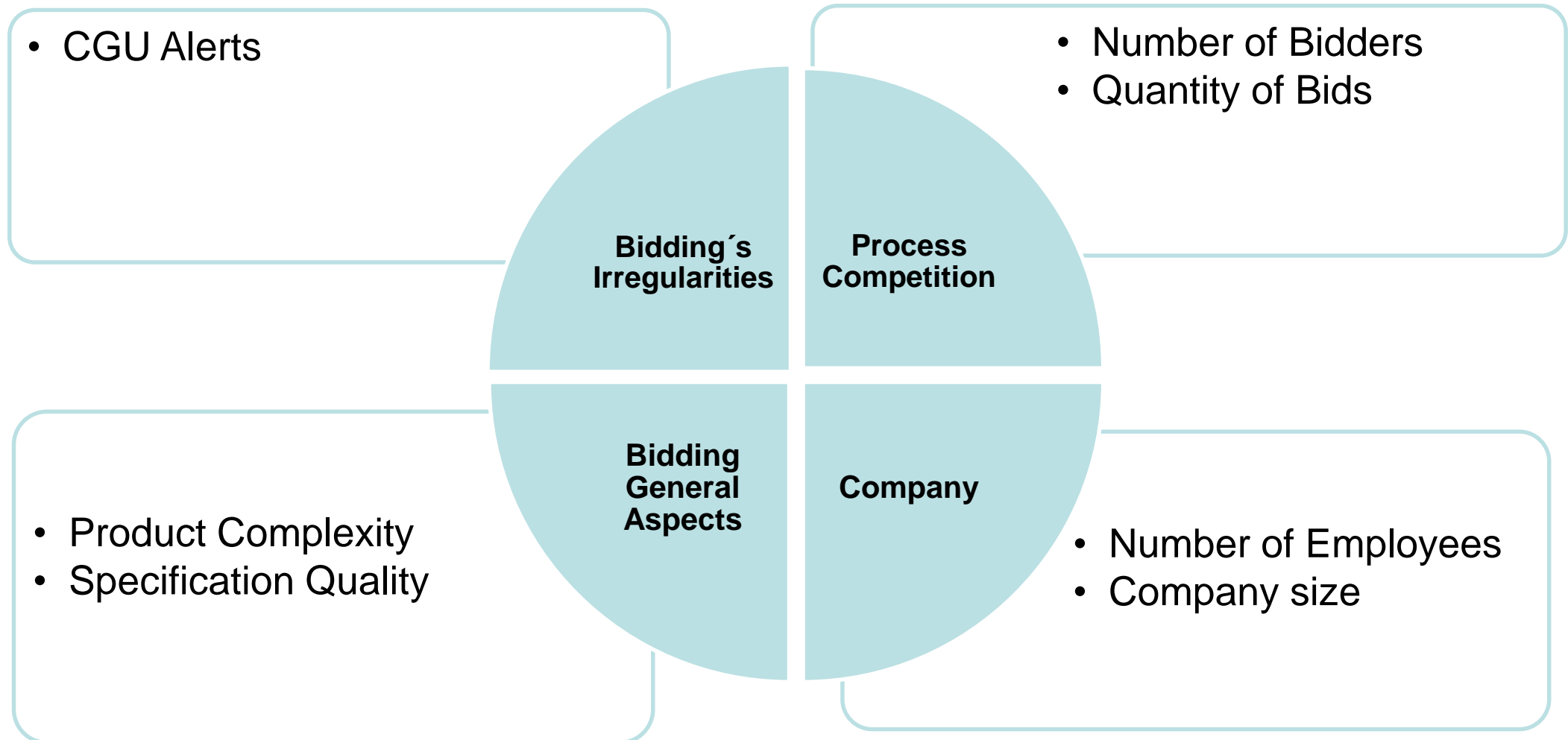


CGU alerts



# Model 2: Contract Risk

1<sup>st</sup> Phase: Identifying risk dimensions





# Model 2: Contract Risk

2<sup>nd</sup>, 3<sup>rd</sup> and 4<sup>th</sup> Phases

Same methodology as shown on Supplier Risk:

- Creating Database
- Listing variables
- Splitting data test/learning
- Forward Stepwise
- Tuning

# Model 2: Contract Risk

5th Phase: Creating a final model using the full training database

6th Phase: Applying model in test database

## Results: Confusion Matrix

	Predict 1	Predict 0
1	133	18
0	24	127

Accuracy	86.1%
Sensitivity	88.1%
Specificity	84.1%
Precision	84.7%

## Results: Interpretation

Contract of service



Discount obtained



Bids per participant



Complexity



Qty. of registered activities



Qty. of companies' partners



# Model 3: Contract Selection



**Contract Risk**

+



**Supplier Risk**

+

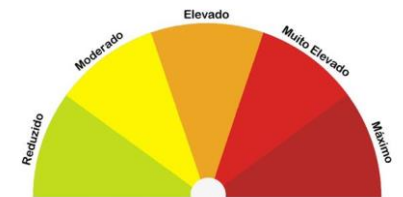


**Logistic Issues**

- Is the Agency already in the Audit Plan?
- Is the agency located in a Capital City?
- Does it require an “expert” to audit?
- What is the contract value?

=

**Auditing Score**



# Model 3: Contract Selection

## Multi-criteria Decision Analytic Hierarchy Process (AHP)

### Comparison Matrix of the criteria pairs

Criteria	C1	C2	C3	C4
C1	<b>C1/C1</b>	C1/C2	C1/C3	C1/C4
C2	C2/C1	<b>C2/C2</b>	C2/C3	C2/C4
C3	C3/C1	C3/C2	<b>C3/C3</b>	C3/C4
C4	C4/C1	C4/C2	C4/C3	<b>C4/C4</b>

# Model 3: Contract Selection

Final criteria for contract evaluation:

- Supplier risk
- Contract risk
- Is the Agency near a Capital City?
- Is the Agency already in the Annual Audit Plan?
- Is there any requirement/availability of specialized work team?
- Total contract value



# Model 3: Contract Selection

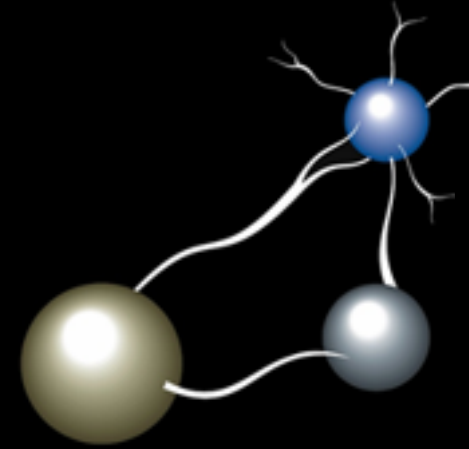
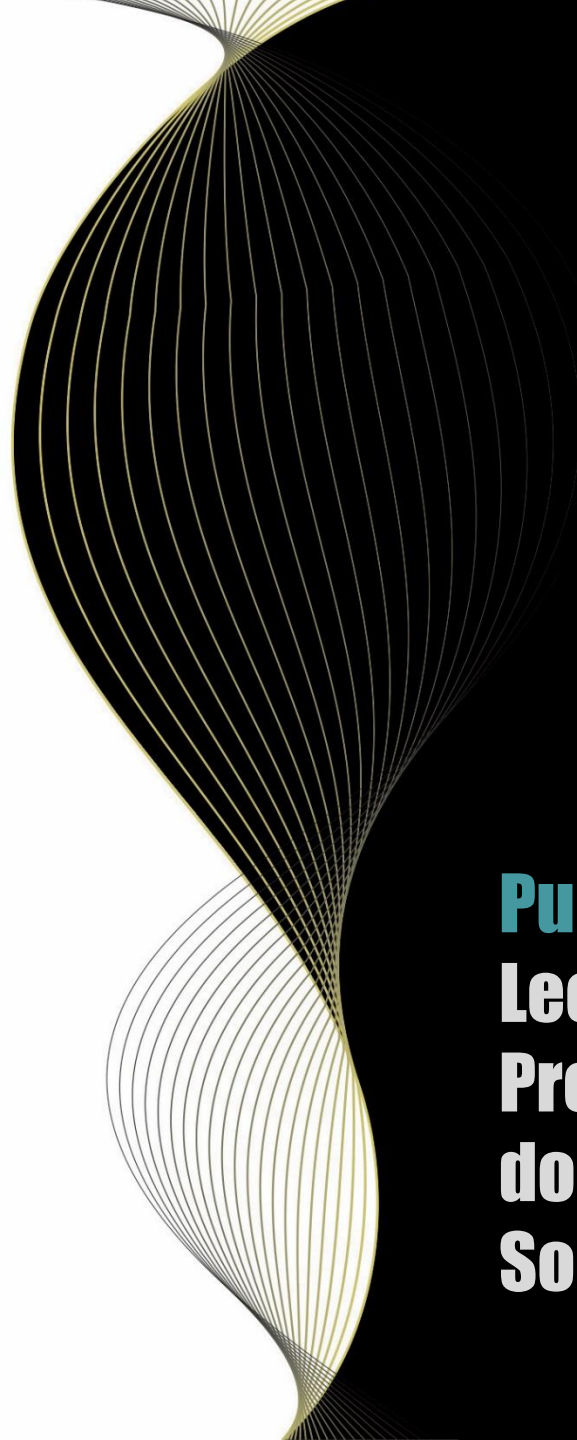
## Simulation: applying AHP to High Risk Contracts

- The contract evaluated with the highest risk dropped to the 45th position. Why?
  - Low value
  - Agency was out of the Audit Plan
- The contract evaluated at the 20th in risk ranked to the first position. Why?
  - Company located in state capital area.
  - Agency was already in the Audit Plan
  - High contract value

# Forward thoughts

**Already implemented: IT biddings - Federal Government**

**Creation of a shared indicators and code repository**  
(GITHUB – R Code)



# Public Spending OBSERVATORY

---

## **Publication:**

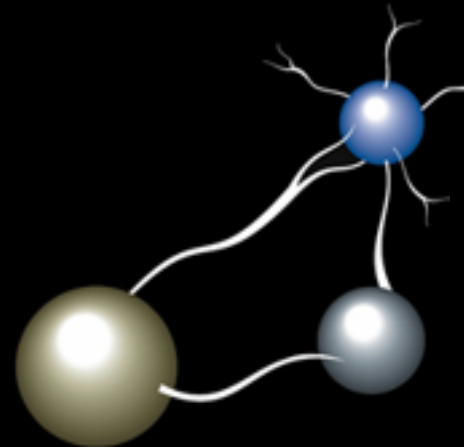
**Leonardo Jorge Sales, M. Sc.**

**Proposta de Modelo de Classificação  
do Risco de Contratos Públicos.**

Soon in **<http://mesp.unb.br/ano-2016>**



**Thank you !**



**Public Spending  
OBSERVATORY**

---

**Claudio Grunewald**  
**claudio.soares@cgu.gov.br**  
**David Cosac**  
**david.cosac-junior@cgu.gov.br**