



IV SEMANA DE
CONTROLE INTERNO, TRANSPARÊNCIA,
OUVIDORIA E CORREIÇÃO DA CGDF



Soluções de Inteligência Artificial no Combate à Corrupção





Por que Ciência de Dados num
órgão de controle governamental?



Estrutura



Especialistas

Auditores, TI, Analistas, Advogados, ...



Tecnologia

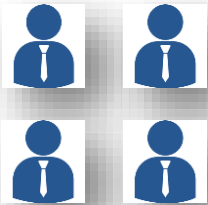
Técnicas e ferramentas avançadas



Metodologia

Objetivo comum

Equipe dedicada na CGU



Análise de
negócio



Gestão de
dados

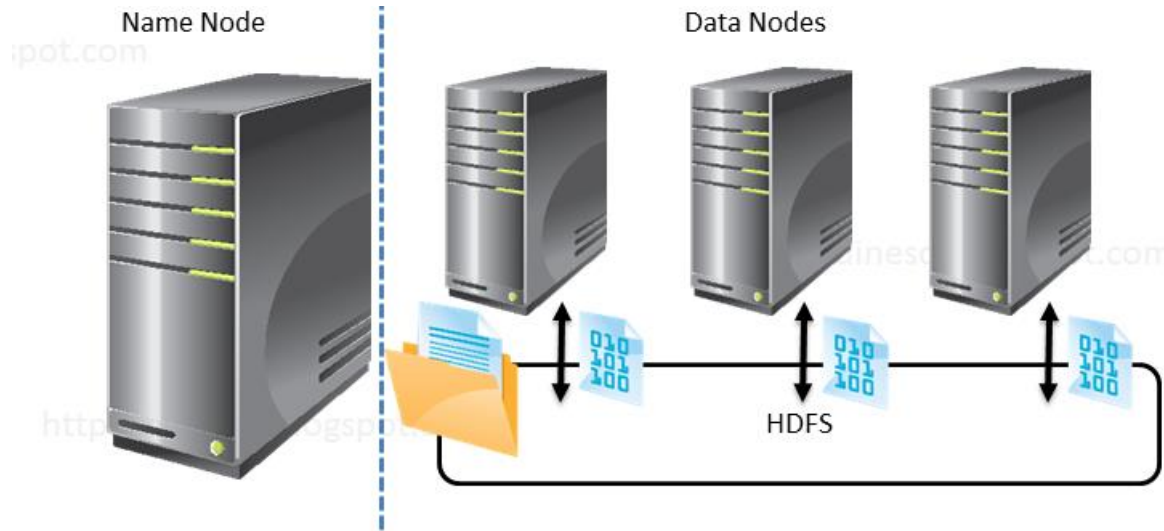


Mineração e
análise de dados

Ferramentas Utilizadas

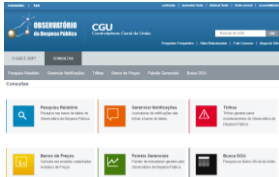


Ambiente de Big Data



- 5 máquinas
- 1 TB de RAM e 56 núcleos (x 2), cada uma

Técnicas



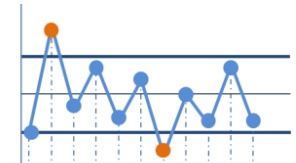
**Auditoria
Contínua**



Classificação



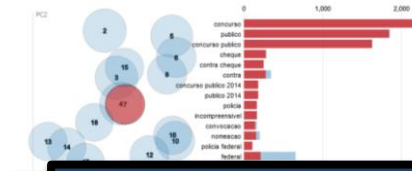
Regressão



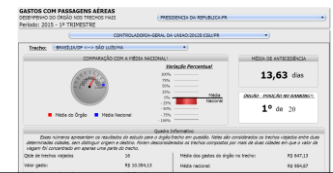
**Detecção de
Anomalia**



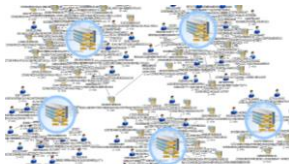
**Mineração de
Textos**



**Modelagem de
Tópicos**



Inteligência de Negócio



Análise de Redes Sociais



Clusterização



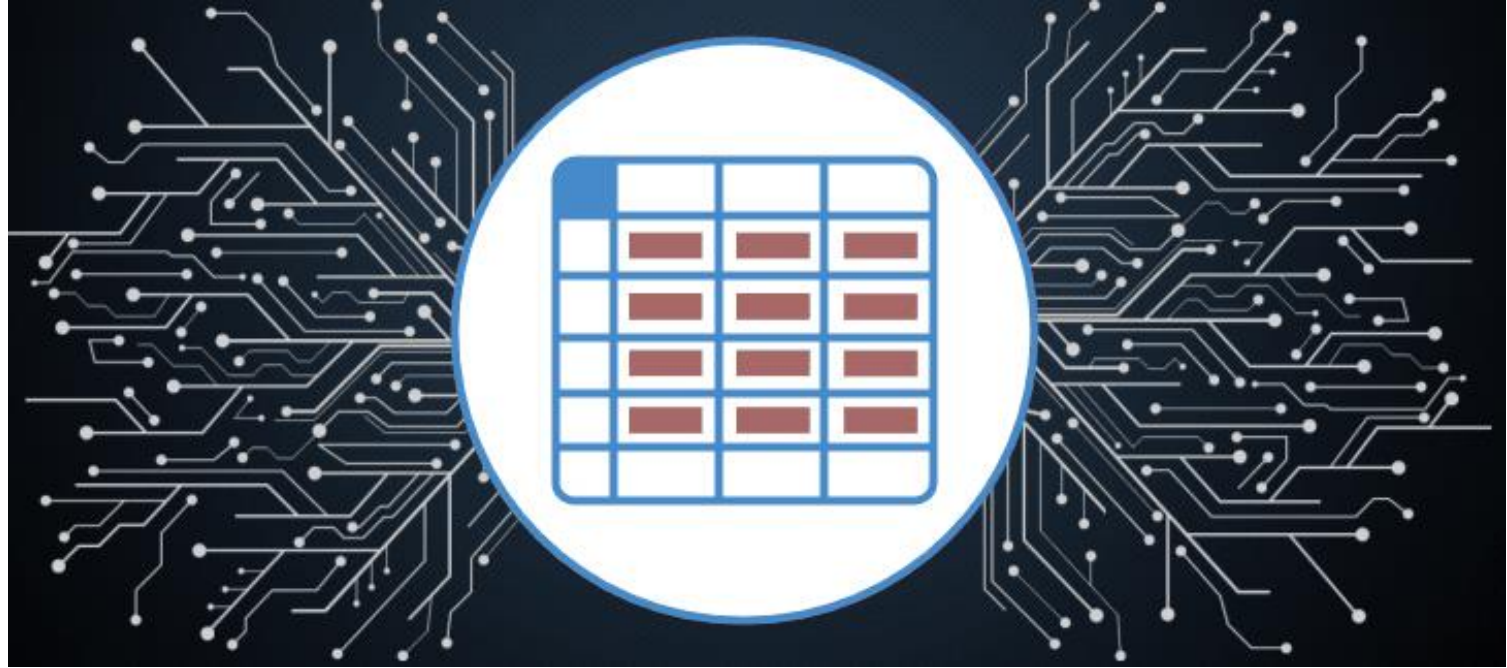
Webscrapping

Portfólio

- Monitoramento computacional de gastos
- Estudos técnicos
- Aplicações de IA

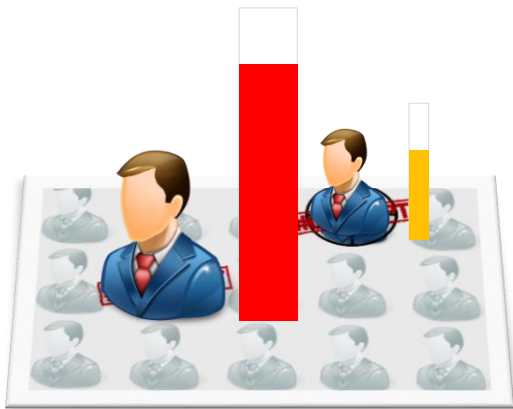
Portfólio

- Monitoramento computacional de gastos
- Estudos técnicos
- **Aplicações de IA**



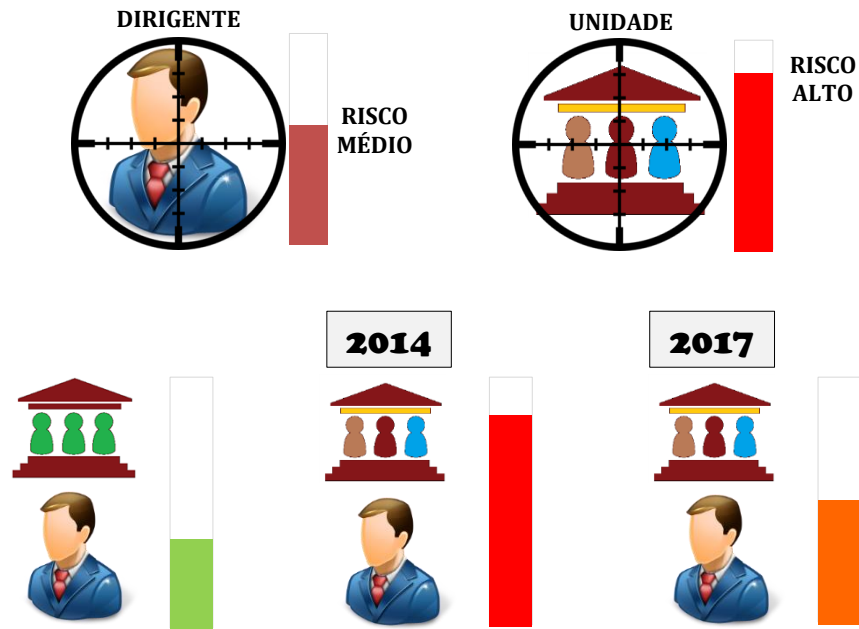
Modelos preditivos com
dados estruturados

Modelo de Risco de Corrupção



Identificação de risco de corrupção através da análise cadastral, histórica e comportamental do servidor público

Modelo de Risco de Corrupção



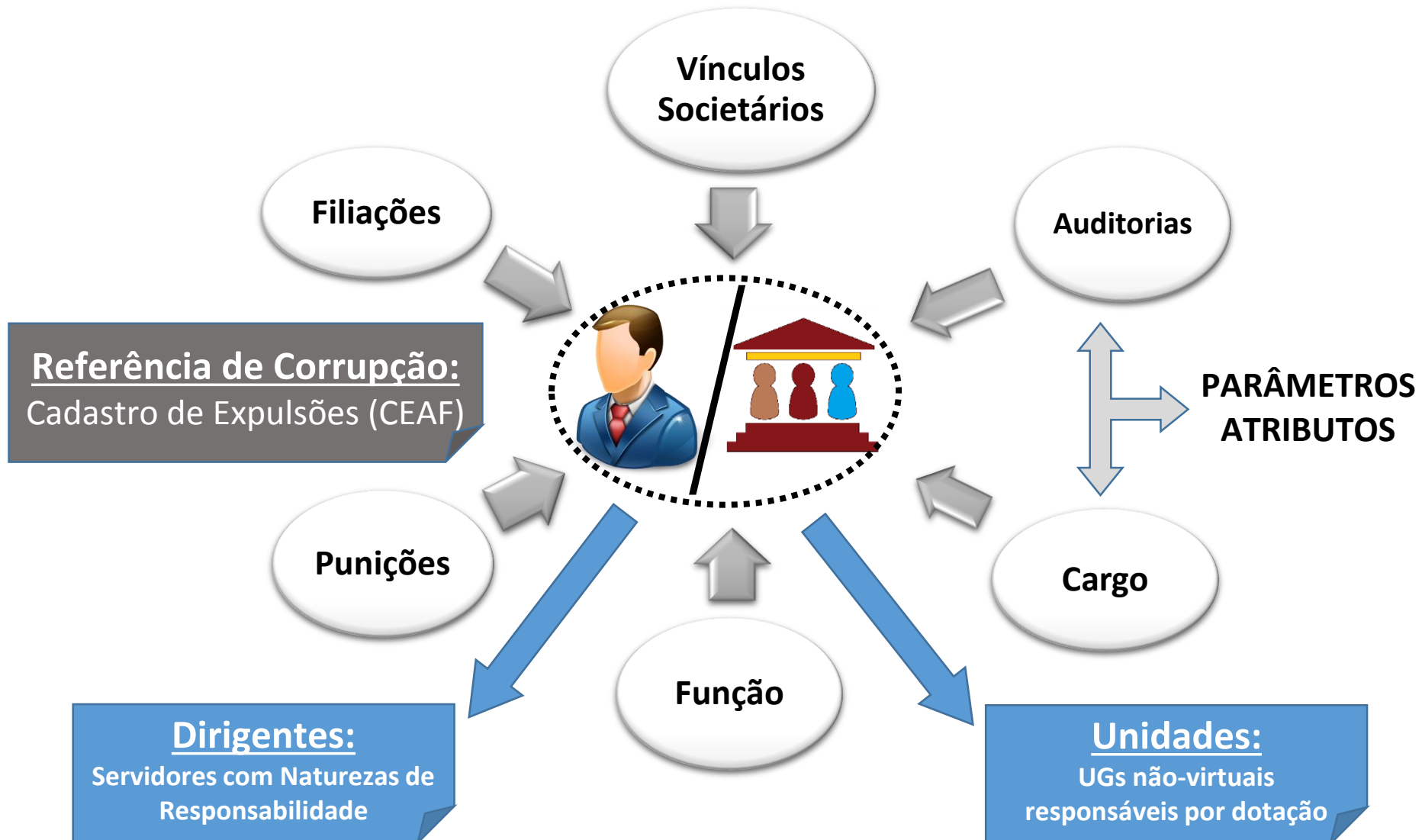
Desafio: Transformar dados em indicadores confiáveis



Como?

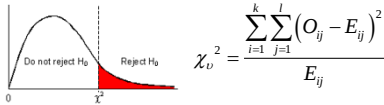


Modelo de Risco de Corrupção



Modelo de Risco de Corrupção

Suporte Estatístico Inicial



Embasamento Teórico



Tratamento Básico de Dados

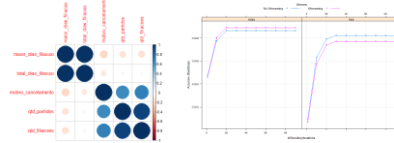


Agregação
Pivoteamento



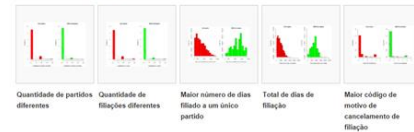
Padronização
Atualização

Seleção de Atributos Relevantes



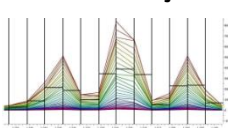
Histogramas de dados balanceados [judits]

Nas figuras a seguir são apresentados os histogramas dos dados balanceados, ou seja, os dados compreendidos são uma amostra dos. Para essa amostra foram pegos todos os Corruptos e, em seguida, de forma aleatória, a mesma quantidade de Não Corruptos. Isto foi feito devido ao desbalanceamento dos dados.

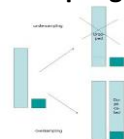


Pré-Processamento de Dados

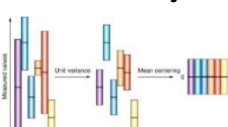
Discretização



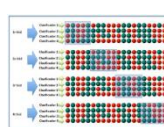
Sampling



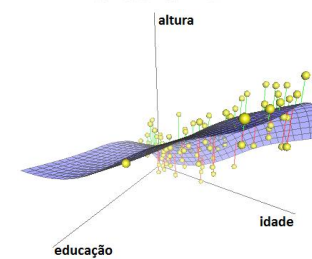
Normalização



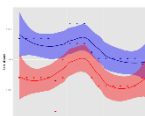
Datasets



Aprendizagem de Máquina



Validação de Modelos



```
'data.frame': 360 obs. of 14 variables:
 $ cost : num 0.0188 0.0288 0.009 0.001 0.
 $ cp : num 0.001 0.006 0.006 0.003 0.
 $ auc : num 0.47 0.486 0.452 0.463 0.4
 $ sens : num 0.033 0.0099 0.0157 0.0309
 $ spec : num 0.999 0.4 0.999 0.999
 $ ppv : num 0.955 0.95 0.95 0.95 0.94
 $ nrv : num 10 520 0 0 0 0 0 0 0 0 0 0 0
 $ pressure : num 0.0889 0.048 0.0752 0.072
 $ aucsd : num 0.0248 0.0182 0.0248 0.02
 $ senssd : num 0.0297 0.044 0.024 0.029
 $ specsd : num 0.95229 0.000579 0.000571
 $ ppvssd : num 0.0718 0.1 0.1 0.0821 0.01
 $ nrvssd : num 0.00443 0.00204 0.00343 0.
 $ pressuressd : num 0.0423 0.0223 0.0261 0.04
```

Ajuste de Modelos



Implantação
de Modelo no
Sistema

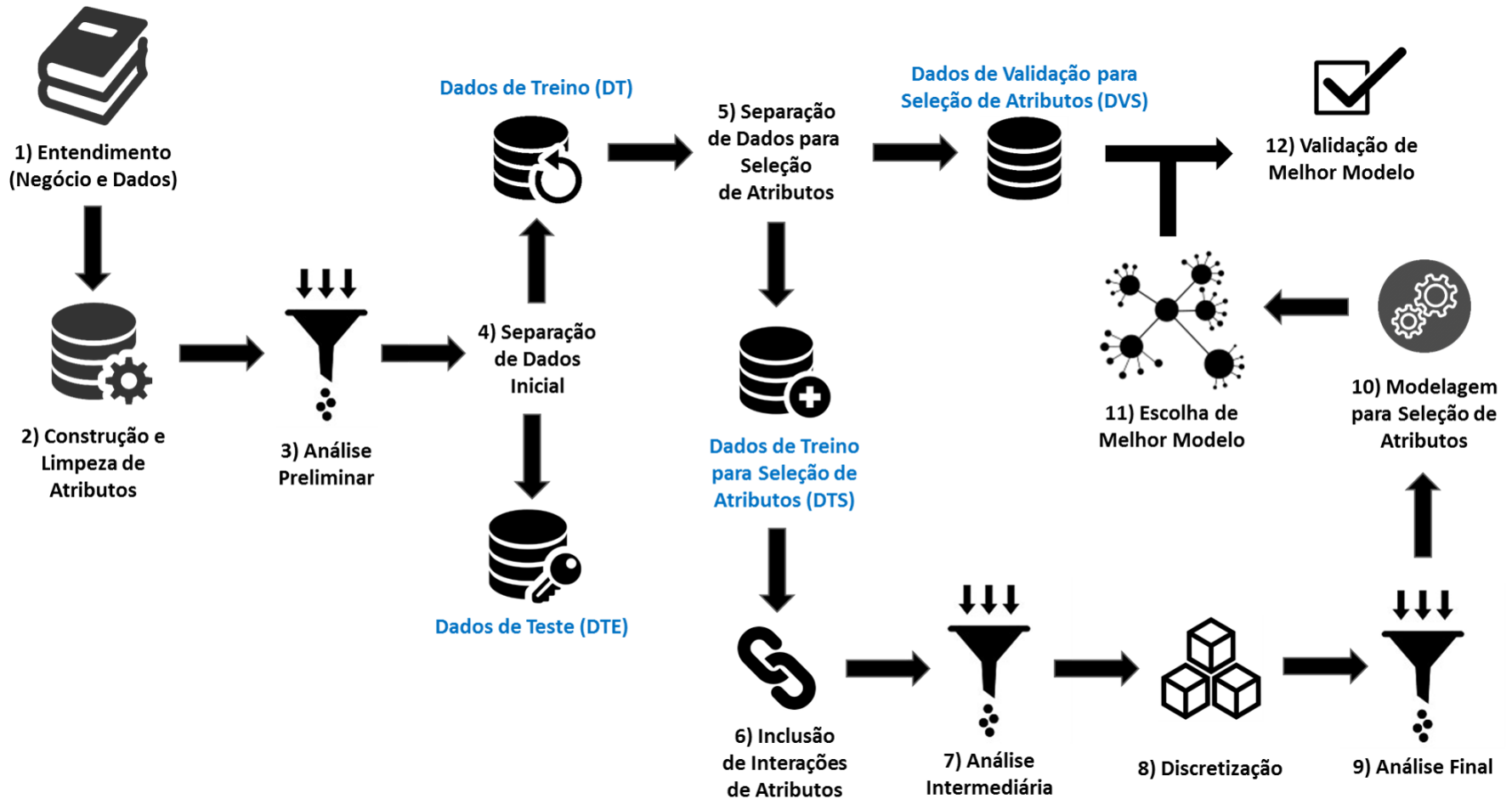
SIM

Modelo
Final
Confiável?

Teste de
Modelo Final



Modelo de Risco de Corrupção



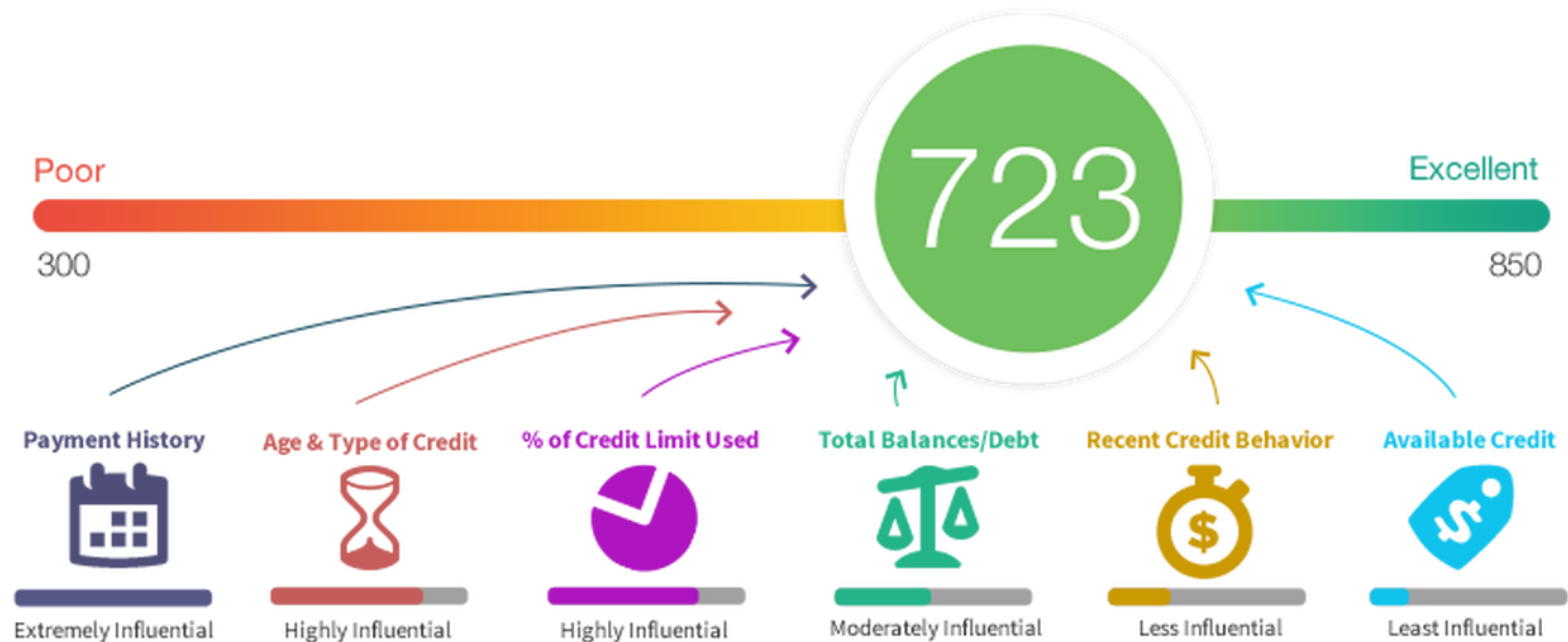
Modelo de Risco de Corrupção



Alguns dados da construção da versão Beta:

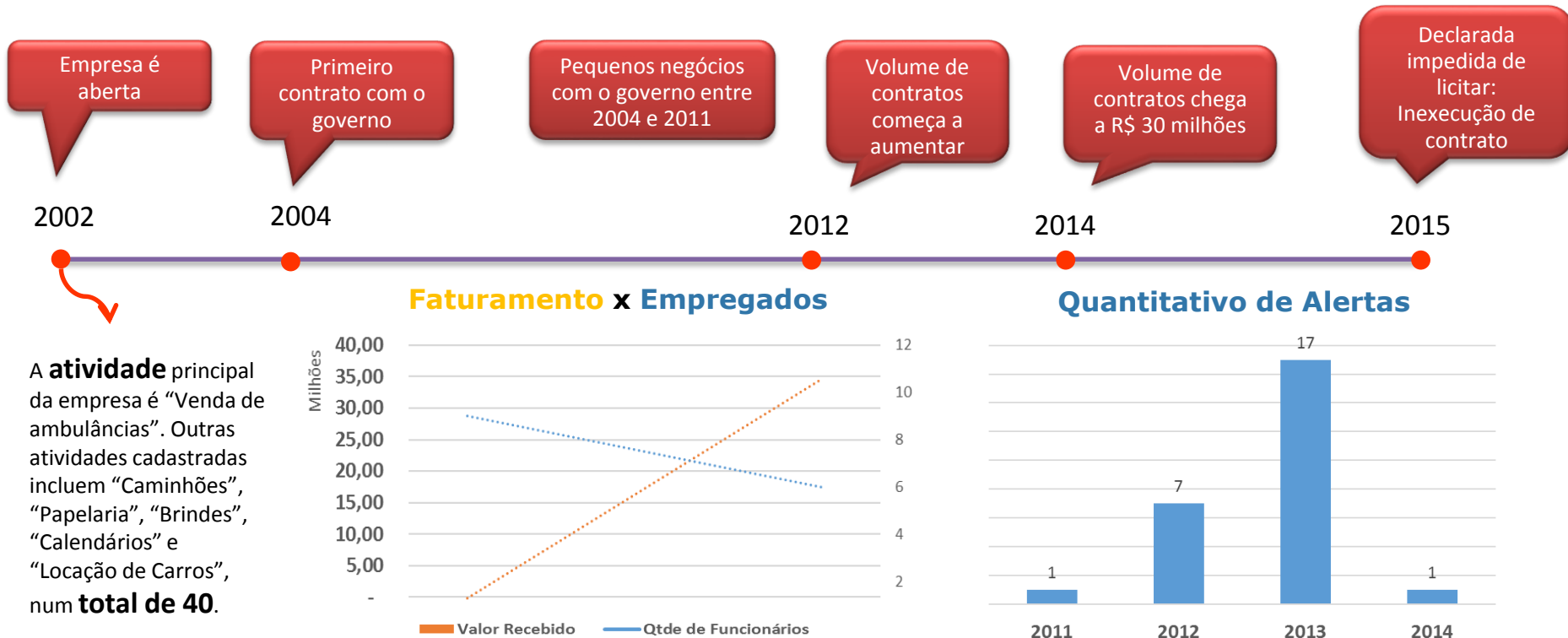
- ✓ 18 bases de dados utilizadas (+500 GB)
- ✓ 5.000 horas de trabalho
- ✓ 4.376 atributos elaborados
- ✓ 62.204 atributos de interação
- ✓ 76 atributos selecionados > 10^{22} possibilidades
- ✓ Godzilhões de cálculos estatísticos
- ✓ 10 produtos descritos, sendo 3 implementados
- ✓ 296 páginas de Wiki
- ✓ 150 arquivos produzidos
- ✓ 23.000 linhas de código
- ✓ 31.476 dirigentes e 5.186 UGs abordados

Modelo de Risco de Fornecedores



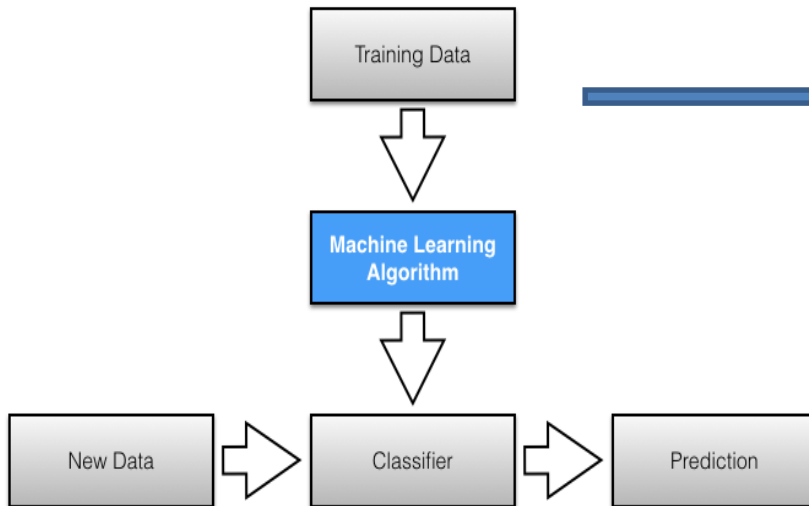
Modelo de Risco de Fornecedores

Quem conhece a Construtora Emporium?



Modelo de Risco de Fornecedores

Processo básico de aprendizagem de máquina supervisionada.



Base de treino

Histórico de contratações, separando:

- Fornecedores adimplentes
- Fornecedores inadimplentes

1446 empresas

180 variáveis peditoras

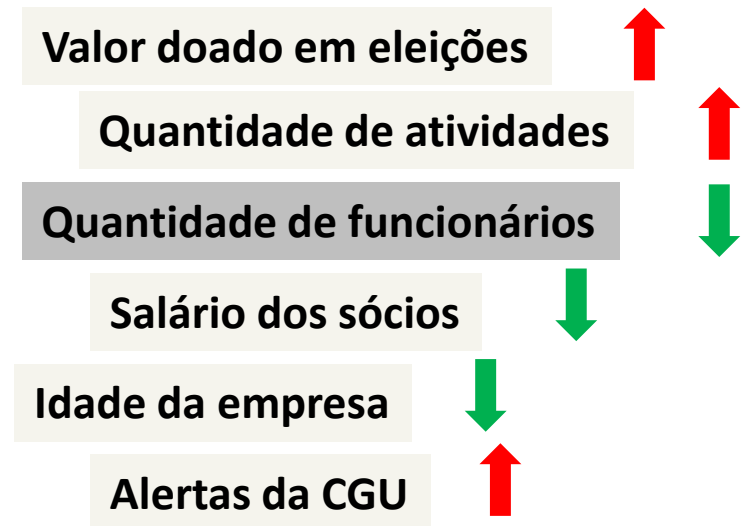
Modelo de Risco de Fornecedores

Algoritmos

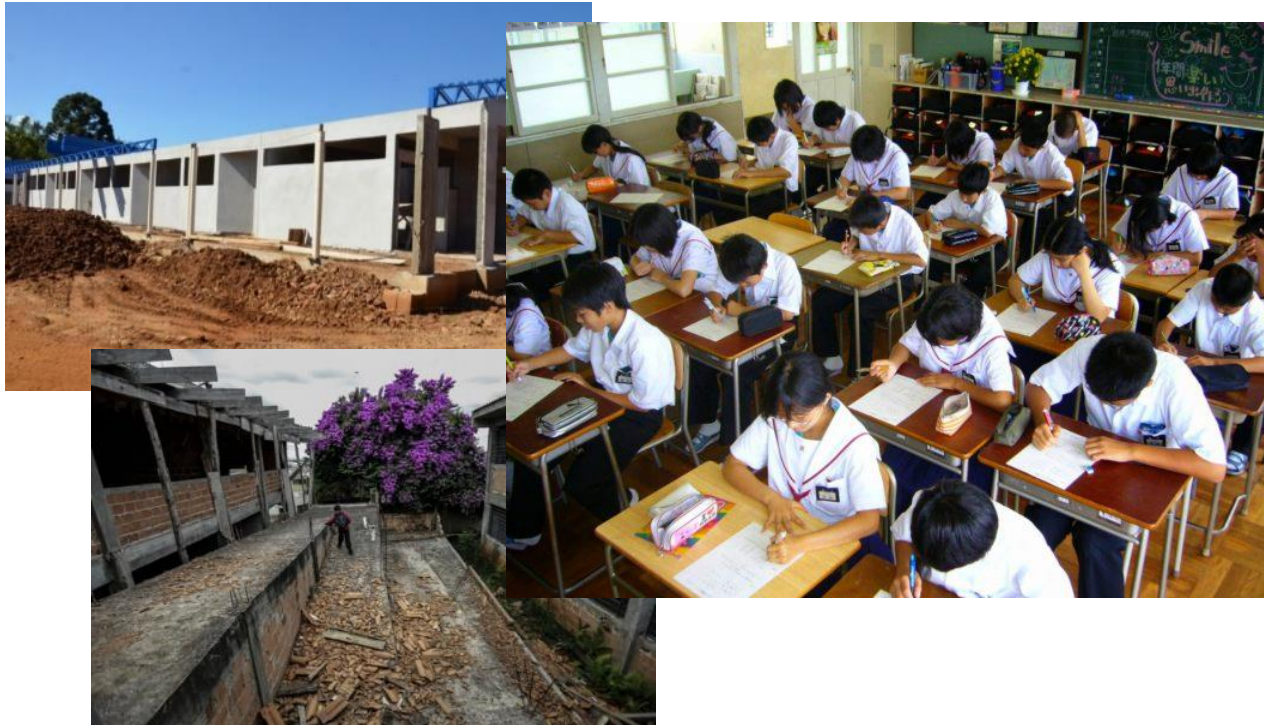
- 1ª versão: Regressão Logística
- 2ª versão: Redes Neurais
- 3ª versão: PU Learning (em implementação)

Precisão atual:
60%

Resultados: Interpretação



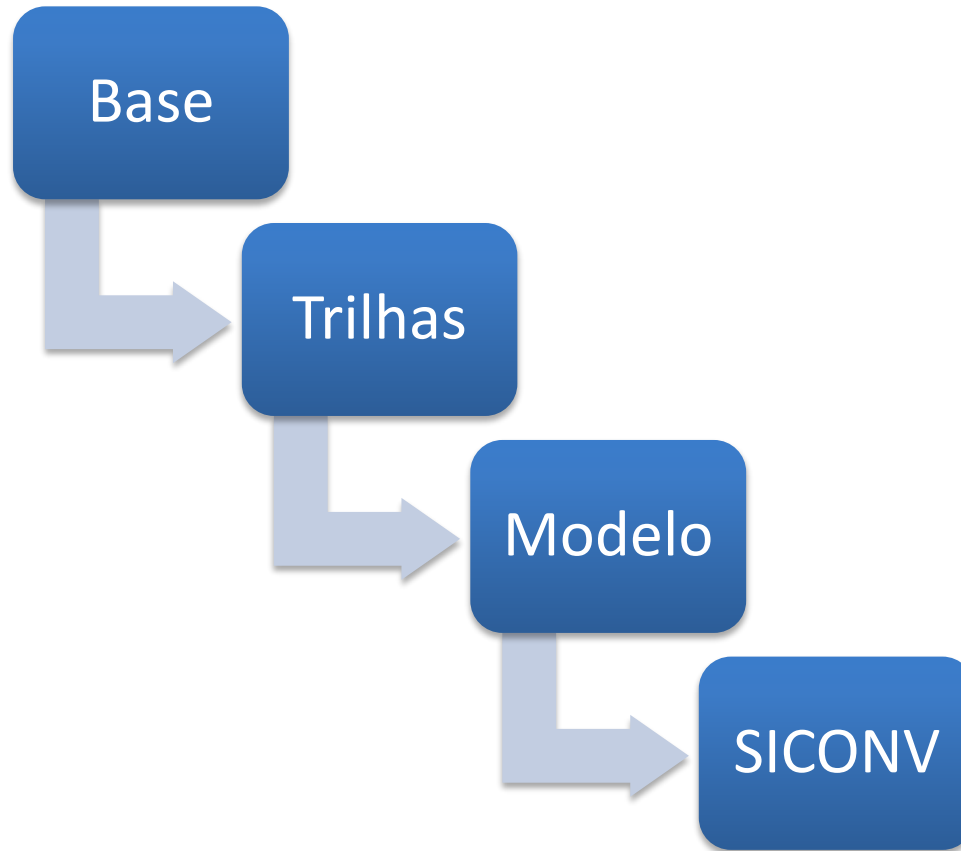
Modelo de Risco de Convênios

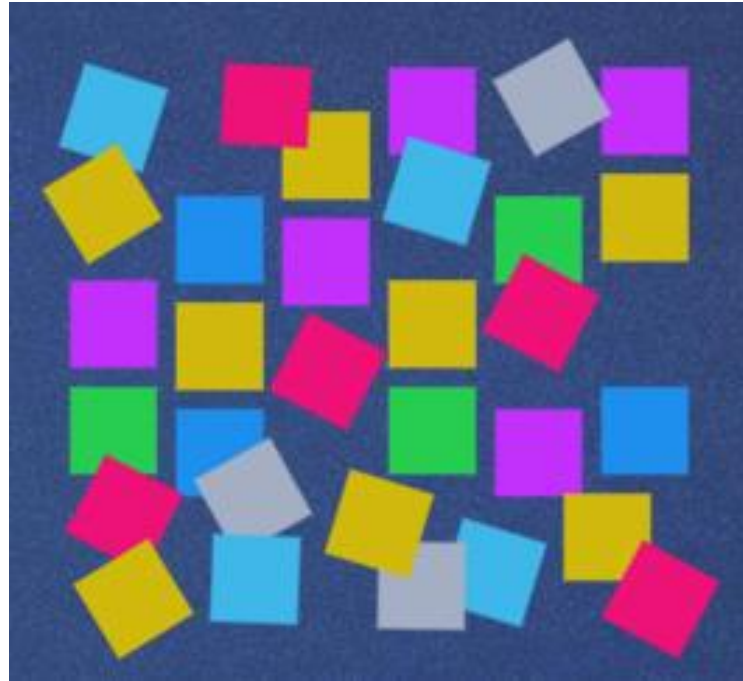


Modelo de Risco de Convênios

15 trilhas em produção +
Modelo de previsão +
Integração com o SICONV

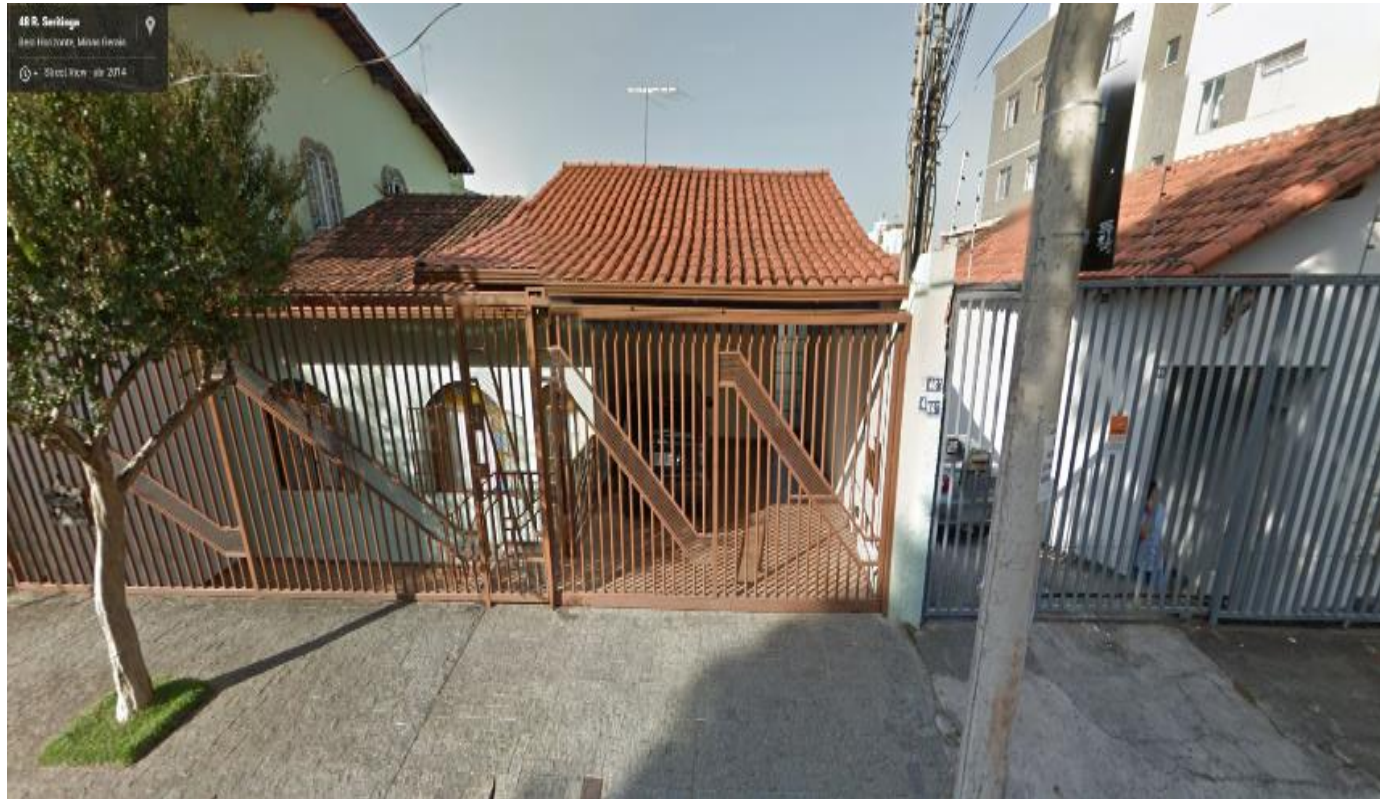
Modelo de Risco de Convênios





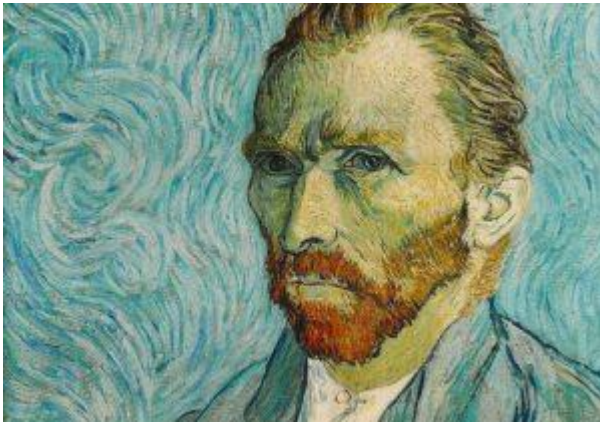
Modelos preditivos com
dados NÃO estruturados

Análise de Fornecedores de Fachada



Análise de Fornecedores de Fachada

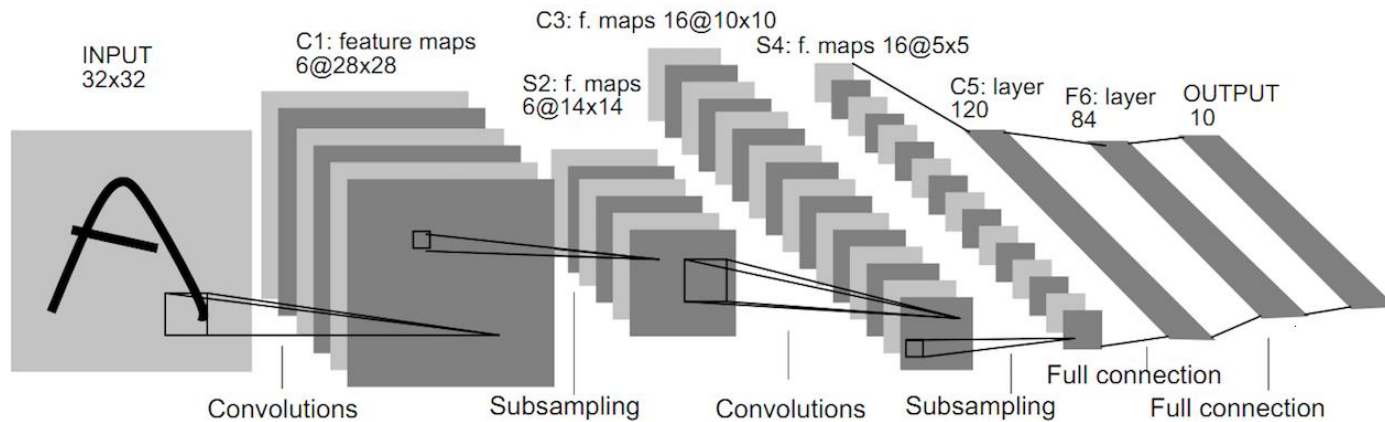
Uma imagem, mais de perto.



			24	25	26
		12	13	14	29
0	1	2	17	32	
3	4	5	20	35	
6	7	8	23		
9	10	11			

Análise de Fornecedores de Fachada

Redes Neurais Convolucionais.



Análise de Fornecedores de Fachada



- 1º passo: Loja x Não loja
- 2º passo: Cenários

Análise de Fornecedores de Fachada

R\$ 1.399.894,26



R\$ 19.964.550,22



R\$ 456.673,83



R\$ 661.044,73



R\$ 1.066.003,86



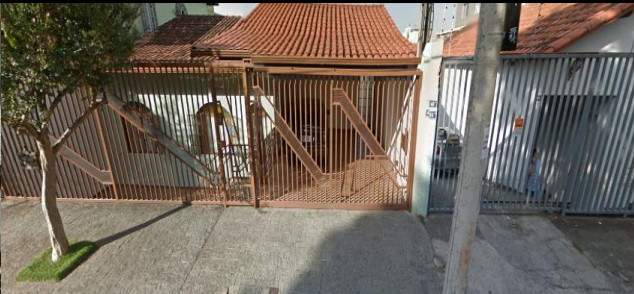
R\$ 1.914.453,57



R\$ 935.472,81



R\$ 6.736.946,00



R\$ 1.990.191,10



Análise de Imagens de Creches



Análise de Imagens de Creches



Análise de Imagens de Creches



Análise do Texto de Denúncias

“Eu, João Silva Torres Carvalho, denuncio o servidor Paulo Soares Neiva de Assis lotado no Ministério da Saúde (no DF) por desvio de recurso público de convênio envolvendo a empresa PAPA TUDO (CNPJ 11.888.555/0001-71), que além disso emprega outro servidor público Felipe Luiz (SIAPE 1999100) e beneficiários do Bolsa Família, como Maria Lopes (NIS 171.33333.00-1) e Mário Filho (CPF 992.991.990-44), com salários absurdos em torno de 10 milhões por ano para cada um. Favor verificar convênio 985283 de janeiro de 2018 de R\$ 230.000.000,00 e o convênio 652763”

**Texto de
Denúncia**

Trâmite?

Já investigados?

Importantes?

Materialidade?

Relacionados?

Análise do Texto de Denúncias

“Eu, João Silva Torres Carvalho, denuncio o servidor Paulo Soares Neiva de Assis lotado no Ministério da Saúde (no DF) por desvio de recurso público de convênio envolvendo a empresa PAPA TUDO (CNPJ 11.888.555/0001-71), que além disso emprega outro servidor público Felipe Luiz (SIAPE 1999100) e beneficiários do Bolsa Família como Maria Lopes (NIS 171.333333.00-1) e Mário Filho (CPF 992.991.990-44), com salários absurdos em torno de 10 milhões por ano para cada um. Favor verificar convênio 985283 de janeiro de 2018 de R\$ 230.000.000,00 e o convênio 652763”

1ª Tarefa: Identificar

NOME

CPF

CNPJ NIS

Contrato

SIAPE

Convênio

Materialidade

Palavras

“fortes”

Análise do Texto de Denúncias

“Eu, João Silva Torres Carvalho, denuncio o servidor Paulo Soares Neiva de Assis lotado no Ministério da Saúde (no DF) por desvio de recurso público de convênio envolvendo a empresa PAPA TUDO (CNPJ 11.888.555/0001-71), que além disso emprega outro servidor público Felipe Luiz (SIAPE 1999100) e beneficiários do Bolsa Família, como Maria Lopes (NIS 171.33333.00-1) e Mário Filho (CPF 992.991.990-44), com salários absurdos em torno de 10 milhões por ano para cada um. Favor verificar convênio 985283 de janeiro de 2018 de R\$ 230.000.000,00 e o convênio 652763”

2ª Tarefa: Expandir



Análise do Texto de Denúncias

“Eu, João Silva Torres Carvalho, denuncio o servidor Paulo Soares Neiva de Assis lotado no Ministério da Saúde (no DF) por desvio de recurso público de convênio envolvendo a empresa PAPA TUDO (CNPJ 11.888.555/0001-71, que além disso emprega outro servidor público Felipe Luiz (SIAPE 1999100) e beneficiários do Bolsa Família, como Maria Lopes (NIS 171.333333.00-1) e Mário Filho (CPF 992.991.990-44), com salários absurdos em torno de 10 milhões por ano para cada um. Favor verificar convênio 985283 de janeiro de 2018 de R\$ 230.000.000,00 e o convênio 652763”

2ª Tarefa: Expandir

CPF 123.321.111-01

CPF 222.333.444-11

Dependente: 747.747.747-11

Sócios: 222.333.444-11

992.991.990-44

Empregado: 213.213.213-22

Familiares: 766.766.766-55

812.666.482-21

Valor global: R\$ 100 milhões

Subcontratada:

+1 nível 22.153.444/0001-88

Análise do Texto de Denúncias

3ª Tarefa: Qualificar

“Eu, João Silva Torres Carvalho denuncio o servidor Paulo Soares Neiva de Assis lotado no Ministério da Saúde (no DF) por desvio de recurso público de convênio envolvendo a empresa PAPA TUDO (CNPJ 11.888.555/0001-71), que além disso

Servidor Federal e Estadual, Análise DIE, Cad. Único CEIS, CEPIM, PEPs, Doação TSE, Falhas CGU, PAD Inabilitados, Contas Julgadas Irregulares, Contratos Contratos Vigentes, Convênios, Convênios Vigentes

emprega Luiz (SIA Bolsa Família 171.3333 992.991. em torno de cada um. Favor verificar convênio 985283 de janeiro de 2018 de R\$ 230.000.000,00 e o convênio 652763’



Auditores Virtuais

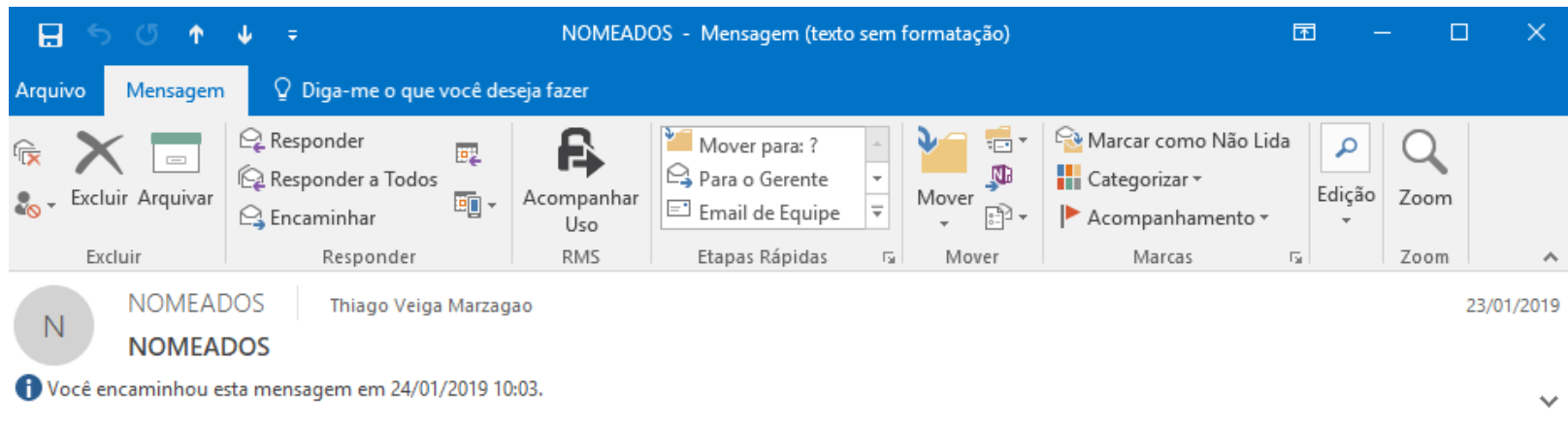
Robô Barnabette



Leitura e qualificação de
atos do DOU

Robô Barnabette

- **Barnabette** varre o D.O.U. todo dia útil e envia os achados por e-mail.



O(A) servidor(a) JOTAVIO BORGES GOMES, CPF 29160677104, foi nomeado(a) hoje para um cargo público pelo órgão Ministério de Minas e Energia/Agência Nacional de Mineração: http://www.in.gov.br/materia/-/asset_publisher/Kujrw0TZC2Mb/content/id/60172104 Mas já foi expulso(a) da Administração Federal e, em virtude da(s) conduta(s) penalizada(s) (8112-117-IX , 8112-132-IV), não está apto(a) a retornar (vide CEAF).

Robô Alice

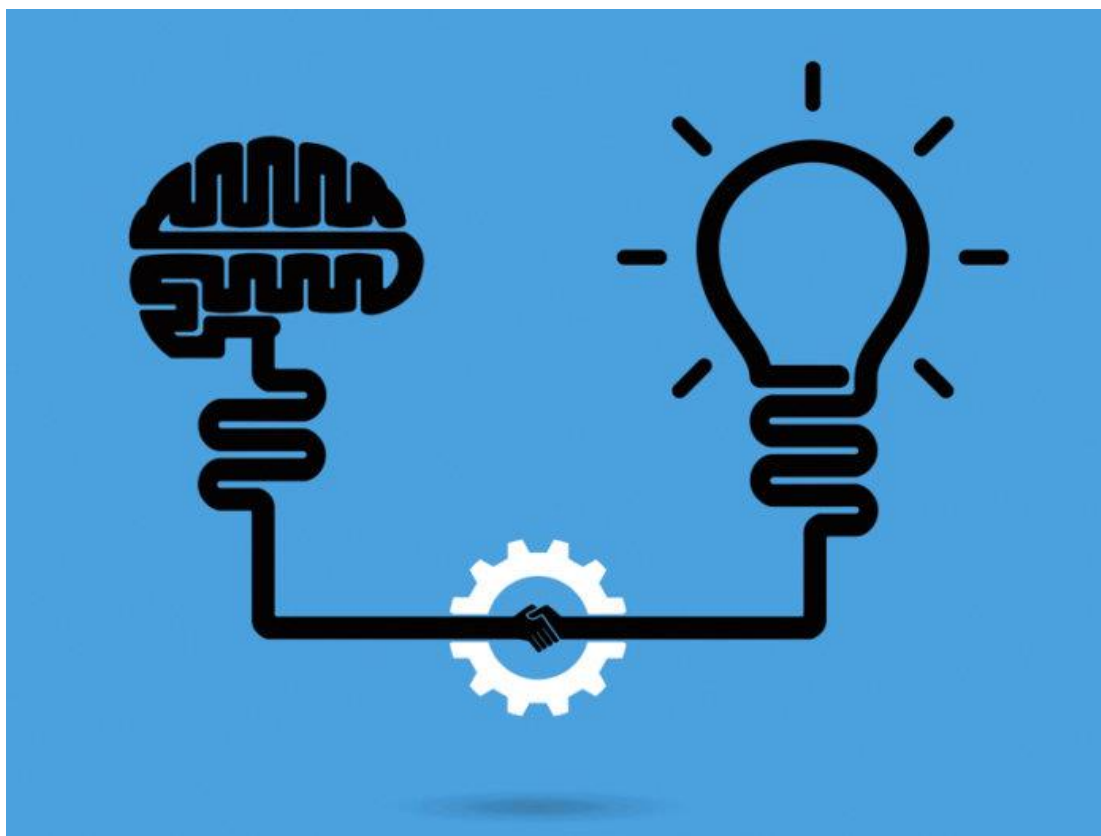


Leitura e identificação de
indícios de direcionamento
em editais

CONTROLADORIA-GERAL
DA UNIÃO



TRIBUNAL DE CONTAS DA UNIÃO
Fiscalização a serviço da sociedade



Outras iniciativas

Banco de Vínculos



Análise de Redes Sociais

Criação de Banco de Vínculos Modelagem em Grafo



- ✓ Apoio a pesquisas de relacionamentos na DIE

Análise de nepotismo



- ✓ Apoio ao Comitê de Risco de Integridade
- ✓ Mapeamento na CGU
- ✓ Mapeamento no Governo



“ANYONE CAN COOK!”

Capacitação

Agenda 2019

- Consultas e exploração de dados **usando SQL**
 - Introdução à Análise de dados com **Python**
 - Análise de Dados Avançada com **Python**
 - Análise **Massiva de Dados** x 4
-

Publicações e Prêmios

Artigos Nacionais e Internacionais



Eficiência na Administração Pública



Conhecimento Governamental



Governo para cidadãos



Gestão da Informação



Informação de Custos e Qualidade do Gasto no Setor Público





IV SEMANA DE
CONTROLE INTERNO, TRANSPARÊNCIA,
OUVIDORIA E CORREIÇÃO DA CGDF

Leonardo Jorge Sales
Controladoria-Geral da União
leonardo.sales@cgu.gov.br
(61) 2020-7277

**Escola de Governo
do Distrito Federal**

**Secretaria de
Fazenda, Planejamento,
Orçamento e Gestão**



**Escola Nacional de
Administração Pública**



**Controladoria-Geral
do Distrito Federal**

