

# Aprendizado de Máquina em Ações de Controle no Tribunal de Contas do Estado do Maranhão

Paulo N. S. Carmo<sup>1</sup>, Bruno F. Souza<sup>1</sup>, Marcelo V. G. Reis<sup>2</sup>, Jairo C. Vieira<sup>3</sup>

<sup>1</sup> Engenharia da Computação – Universidade Federal do Maranhão (UFMA)  
Av. dos Portugueses, 1966 – 65080-805 – São Luís – MA – Brasil

<sup>2</sup>Ministério Público do Estado do Maranhão (MPE-MA)  
Av. Prof. Carlos Cunha, 3347 – 65076-820 – São Luís – MA – Brasil

<sup>3</sup>Ministério Público de Contas – Tribunal de Contas do Estado do Maranhão (TCE-MA)  
Av. Prof. Carlos Cunha, s/n – 65076-820 – São Luís – MA – Brasil

paulonayron.s@hotmail.com, {bferes, marcelovidigal}@gmail.com

jcvieira@tce.ma.gov.br

**Abstract.** *The growing relationship between information technology and public administration management has generated new needs for government agencies in information storage and data processing. This work aims to study the use of Machine Learning in control practices in the Courts of Account, with direct application in the Court of Accounts of the state of Maranhão. Based on the experiences of other States and the Union, local opportunities for the development of automatic and intelligent solutions capable of subsidizing audits were envisaged. This work presents itself as an initial guiding document for such enterprise.*

**Resumo.** *A crescente relação entre a tecnologia da informação e a gestão da administração pública tem gerado novas necessidades aos órgãos governamentais no que se refere ao armazenamento e processamento de informações. Este trabalho visa estudar o emprego de Aprendizado de Máquina em práticas de controle nos órgãos oficiais, com aplicação direta no Tribunal de Contas do Estado do Maranhão. Com base nas experiências de outros Estados e da União, vislumbrou-se oportunidades locais para o desenvolvimento de soluções automáticas e inteligentes capazes de subsidiar auditorias. Este trabalho se apresenta como um documento norteador inicial de tal empreendimento.*

## 1. Introdução

A soma das despesas da União, dos Estados e dos Municípios, entes que compõem a Administração Pública no Brasil, totalizou, aproximadamente, R\$ 3.966.890 milhões no ano de 2017 [Brasil 2018]. De acordo com a Constituição Federal, em especial no artigo 71, cabe aos Tribunais de Contas a verificação deste grande volume de despesas, julgando contas dos administradores e demais responsáveis por bens públicos, realizando auditorias, fiscalizando aplicação de recursos, aplicando sanções, determinando correção de atos ou sustá-los e solicitando providências por parte de Poderes competentes.

Ao tratar dos Tribunais de Contas, a literatura especializada detém-se nos temas de jurisdição, processos, procedimentos e técnicas de auditoria. Segundo [Jacoby 2016],

no modelo atual brasileiro, “ficou indelevelmente definido o exercício da função jurisdicional pelos Tribunais de Contas”. Em [Speck 2000], o autor explica que “o controle exercido via julgamento das contas dos administradores recebe grande atenção por parte do Tribunal de Contas. Muitos integrantes veem nele o mecanismo principal de fiscalização”.

Durante a formulação de seu Plano Estratégico para 2015-2021, o Tribunal de Contas da União observou a tendência de uma utilização mais intensa de recursos computacionais para monitoramento e análise de dados [Brasil 2015]. Neste contexto, destaca-se o uso de inteligência de controle para identificar, em bases de dados grandes e diversas, riscos de inexecução ou execução inadequada de produtos e serviços.

Assim, o presente artigo pretende contribuir com o tema de tecnologia da informação em ações de controle de contas públicas. Especificamente, o trabalho aqui conduzido foca em duas vertentes. Na primeira, é realizado um estudo de prospecção de como técnicas de Aprendizado de Máquina (AM) [Bishop 1996] tem sido aplicadas nos Tribunais de Controle. Na segunda, estuda-se formas de aplicação de AM no Tribunal de Contas do Estado do Maranhão (TCE-MA), de forma a agregar valor às análises realizadas por esse órgão. Ressalta-se que este é um trabalho em andamento, sendo este documento norteador da pesquisa vindoura.

Este trabalho está organizado como segue. Na Seção 2, são apresentados conceitos fundamentais de Aprendizado de Máquina, pertinentes ao entendimento do trabalho. Na Seção 3, estudos relevantes envolvendo a aplicação da AM em ações de controle são discutidos. Na Seção 4, a concepção preliminar do sistema de controle proposto é exposta. Por fim, na Seção 5, apresentam-se as conclusões do trabalho.

## 2. Aprendizado de Máquina

Atualmente, para desempenhar a contento suas funções, os Tribunais de Controle necessitam manipular uma quantidade maciça de dados, oriundos de diversas fontes [Brasil 2015]. Para que informações qualitativamente interessantes possam então ser obtidas desses dados, faz-se uso geralmente de ferramentas computacionais. Como grande parte das análises realizadas envolve extrair padrões dos dados e entender seus relacionamentos, pode ser apropriado considerar a extração automática de conhecimento por meio de Aprendizado de Máquina como uma solução factível.

O Aprendizado de Máquina corresponde a uma área que estuda e modela como sistemas computacionais podem aprender [Bishop 1996]. Ela baseia-se fortemente na Inteligência Artificial e na Estatística para o desenvolvimento de artefatos que melhorem seu desempenho em determinada tarefa utilizando um conjunto de experiências disponíveis na mesma. Com isso, é possível extrair conceitos a partir de amostras de dados.

Dentre as estratégias para o desenvolvimento de abordagens de AM, destacam-se aquelas baseadas no conceito de indução. O aprendizado indutivo acontece quando conclusões genéricas são obtidas a partir de observações particulares. Assim, pode-se extrapolar as informações contidas nos dados e modelar conceitos de nível mais alto. Neste contexto, embora não sejam os únicos, dois paradigmas de AM tem sido predominantemente estudados, a saber: o supervisionado e o não-supervisionado, conforme ilustrado na Figura 1, extraída de [Facelli et al. 2011].

No paradigma supervisionado, os algoritmos preditivos são capazes de induzir



**Figura 1. Hierarquia de AM**

conceitos a partir de exemplos específicos dos mesmos. Para tanto, eles operam sobre descrições de exemplos rotulados e mapeiam características de entrada e de saída de maneira consistente com os dados disponíveis. Quando um modelo baseado neste mapeamento é gerado, torna-se possível prever os rótulos de exemplos não vistos anteriormente.

Os algoritmos de AM supervisionado se apresentam em dois tipos, de acordo com a natureza do rótulo dos exemplos. Nos problemas de classificação, os rótulos são entidades discretas, chamadas de classes. Já nos problemas de regressão, os valores dos rótulos se apresentam de forma contínua. Segundo [Facelli et al. 2011], algoritmos desse paradigma se dividem em baseados em distância, probabilísticos, baseados em procura e baseados em otimização. Algoritmos representativos destas abordagens incluem *k-Nearest Neighbors* (*k*-NN), *Naive Bayes*, Árvores de decisão e Redes Neurais Artificiais, respectivamente.

No paradigma não-supervisionado, os algoritmos descritivos têm como objetivo descobrir padrões e regras gerais capazes de explicar os dados. Para tanto, eles não possuem a facilidade de rótulos associados aos exemplos, fazendo uso tão somente das características de entrada para determinar a estrutura intrínseca dos dados.

Dentre as tarefas que podem ser realizadas por meio de algoritmos de AM não-supervisionado, destacam-se três. Na de agrupamento, o objetivo é formar conjuntos de dados agrupados de acordo com a similaridade dos exemplos. Na de sumarização, pretende-se produzir descrições compactas dos conjuntos de dados, geralmente textos, na forma de resumos com sentenças relevantes. Na de associação, busca-se encontrar padrões frequentes nos dados, de tal forma que itens que ocorrem juntos regularmente sejam descobertos. Algoritmos comumente utilizados dessas abordagens incluem algoritmos hierárquicos, *Latent Semantic Analysis* e Apriori, respectivamente

### 3. AM em Ações de Controle

Em [Balaniuk 2010], é proposta uma arquitetura de um sistema baseado na metodologia CRISP-DM (do inglês, *Cross-Industry Standard Process for Data Mining*) que consiste na subdivisão de processos de trabalho iterativo, constituída por: **Compreensão do negócio**, onde há participação dos auditores para propor tipologias que classificam atos como ilícitos; **Compreensão dos dados**, são escolhidos fontes de dados estratégicos que evidenciem as tipologias; **Preparação dos dados**, busca disponibilizar as bases a serem analisadas; **Modelagem**, escolha de estratégias de cruzamento de dados; **Avaliação e Aplicação**, gera relatórios de análise e métricas de riscos.

Em [Sousa 2015], é realizado uma análise de aplicação de redes neurais para classificação de padrões no contexto de auditoria. O autor identificou inúmeros cenários onde essa abordagem tem sido utilizada, a saber: na identificação de fraudes e erros, na emissão de opinião do tipo *going concern*; identificação de situações de demasiada exposição a riscos financeiros e previsão de falência, por meio de exames de coeficiente e informações contábeis de empresas; E identificação possíveis falhas nas opiniões de auditoria sobre demonstrações financeiras.

[Dutra et al. 2016] descreve o uso de técnicas de Inteligência Artificial (IA), Aprendizado de Máquina, OCR (*Optical Character Recognition*), similaridade léxica e semântica, NER (*Named Entity Recognition*), dentre outras, para subsidiar ações de controle implementadas no Tribunal de Contas da União. A ferramenta desenvolvida automatiza a extração de informações em dados do tipo texto não estruturado, presentes em documentos. O trabalho apresenta os resultados para as técnicas utilizadas, bem como, as que pretende-se adotar em trabalhos futuros.

Em [Costa and Perrota 2018] são mencionados três sistemas que são utilizados no TCU desde o final de 2016, Alice, Sofia e Mônica que agem na identificação de fraudes em licitações públicas. As três ferramentas fazem parte do chamado Laboratório de Informações de Controle (Labcontas), que utiliza de algoritmos amparados em modelos de Aprendizado de Máquina.

#### **4. AM no Tribunal de Contas do Estado do Maranhão**

Alicerçado no entendimento do papel dos Tribunais de Contas e na necessidade dos mesmos em trabalhar com grande volume de dados, fica evidente a conveniência do uso de abordagens baseadas em Aprendizado de Máquina para a realização das análises devidas. Nessas circunstâncias, este trabalho visa nortear os passos iniciais da colaboração científica entre o curso de Engenharia da Computação da Universidade Federal do Maranhão e o Tribunal de Contas do Estado do Maranhão (TCE-MA).

Em comunicação preliminar com a equipe técnica do TCE-MA, verificou-se que uma fonte de dados muito importante corresponde aos editais de licitação contidos no Sistema de Acompanhamento Eletrônico de Contratação Pública (SACOP), disponibilizado no site do órgão. No Mural de Contratos do sistema, são fornecidos arquivos do tipo PDF em texto não estruturado. Verificou-se então que a maior parte dos editais é digitalizada (não editável) e de baixa qualidade, semelhantes aos que foram utilizados em [Dutra et al. 2016]. Assim, oportunidade se apresenta para a utilização de técnicas de AM tanto para realizar o reconhecimento de caracteres em um sistema de OCR quanto para analisar os textos resultantes por meio de Mineração de Texto [Nahm and Mooney 2002]. Com isso, objetiva-se encontrar padrões, tendências ou regras úteis a partir dos conjuntos de dados.

Outra fonte de dados que demonstrou ter importância, foi o Portal da Transparência dos Municípios. A Lei da Transparência (Lei Complementar 131/2009) determina a disponibilização dos dados em tempo real e pormenorizado das despesas orçamentárias por parte da União, dos Estados, do Distrito Federal e dos Municípios. Em [Cruz et al. 2012] é feito um levantamento que torna notória a falta de padronização entre os portais. No Estado do Maranhão, segundo os técnicos do TCE, não há sistema semelhantes ao Comprasnet (utilizado pela Alice em âmbito Federal). Assim, para realizar

a aquisição de dados dos 217 municípios, propõe-se desenvolver um sistema automatizado para a captura das informações dos portais empregando técnicas de *web scraping* e *web crawling*. Nesta etapa, AM pode ser utilizado em seu paradigma não-supervisionado para identificar padrões indicativos de fraudes em operações de empenho.

Uma vez obtidos os dados, análises envolvendo AM podem ser vislumbradas. Baseado nos trabalhos da revisão bibliográfica da Seção 3, pretende-se atuar junto com os auditores para eleger as melhores trilhas de auditoria para auxiliar na busca por irregularidades em contas públicas. Dentre as tipologias já exploradas, citam-se: empresas que ganham licitações em áreas onde não possui experiência ou que não possui patrimônio equivalente ao valor empenhado, empresas que possuem como sócios servidores públicos ou pessoas impedidas, empresas que ganham muitas licitações, etc. Neste, sentido, sistemas de AM poderiam ter úteis para encontrar associações delituosas entre empresas e prefeituras ou para predizer, com base em características das empresas, quais estariam mais propensas a cometer faltas. É possível também agrupar entes públicos e privados a fim de detectar padrões de comportamento suspeitos. Por fim, mas sem exaurir o tema, pode-se usar AM na previsão de falência de empresas que ganham licitações.

## 5. Conclusão

Como resultado dos trabalhos reunidos neste estudo, foi possível observar as principais abordagens de utilização de Aprendizado de Máquina para subsidiar ações de controle. Uma vez realizado tal levantamento, buscou-se viabilizar parceria com o TCE-MA a fim de implementar, futuramente, soluções inteligentes alinhadas com o estado da arte no tema. Destaca-se o entendimento de que a presença de um corpo técnico qualificado continua sendo imprescindível para as ações de controle, fazendo com que os sistemas baseados em AM atuem como instrumentos eficientes para tornar o trabalho desses profissionais mais eficaz.

## Referências

- Balaniuk, R. (2010). A mineração de dados como apoio ao controle externo. *Revista do Tribunal de Contas da União*, 117:77–84.
- Bishop, C. M. (1996). *Neural networks for pattern recognition*. Oxford University Press.
- Brasil (2015). *Plano Estratégico do Tribunal de Contas da União: 2015-2021*. TCU, Secretaria do Planejamento, Governança e Gestão (Seplan), Brasília.
- Costa, B. F. and Perrota, R. P. C. (2018). Inteligência artificial no direito—uma realidade a ser desbravada. *Revista de Direito, Governança e Novas Tecnologias*, 4(1):1–16.
- Cruz, C. F., de Souza Ferreira, A. C., da Silva, L. M., and da Silva Macedo, M. Á. (2012). Transparência da gestão pública municipal: um estudo a partir dos portais eletrônicos dos maiores municípios brasileiros. *Revista de Administração Pública*, 46(1):153–176.
- Dutra, L. A. et al. (2016). Uso de técnicas de inteligência artificial para subsidiar ações de controle. *Revista do TCU*, (137):124–129.
- Facelli, K., Lorena, A. C., ao Gama, J., and de Carvalho, A. (2011). *Inteligência Artificial. Uma Abordagem de Aprendizado de Máquina*. LTC.
- Jacoby, J. U. F. (2016). *Tribunais de Contas do Brasil, jurisdição e competência*. Fórum, Belo Horizonte, 4 edition.

- Nahm, U. Y. and Mooney, R. J. (2002). Text mining with information extraction. In *Proceedings of the AAAI 2002 Spring Symposium on Mining Answers from Texts and Knowledge Bases*, pages 60–67. Stanford CA.
- Sousa, R. M. (2015). Inteligência computacional aplicada ao controle externo da administração pública: aplicações da classificação de padrões utilizando redes neurais artificiais. *Revista Controle: Doutrinas e artigos*, 13(2):201–218.
- Speck, B. W. (2000). *Inovação e rotina no Tribunal de Contas da União: o papel da instituição superior de controle financeiro no sistema político-administrativo do Brasil*. Fundação Konrad Adenauer, São Paulo.