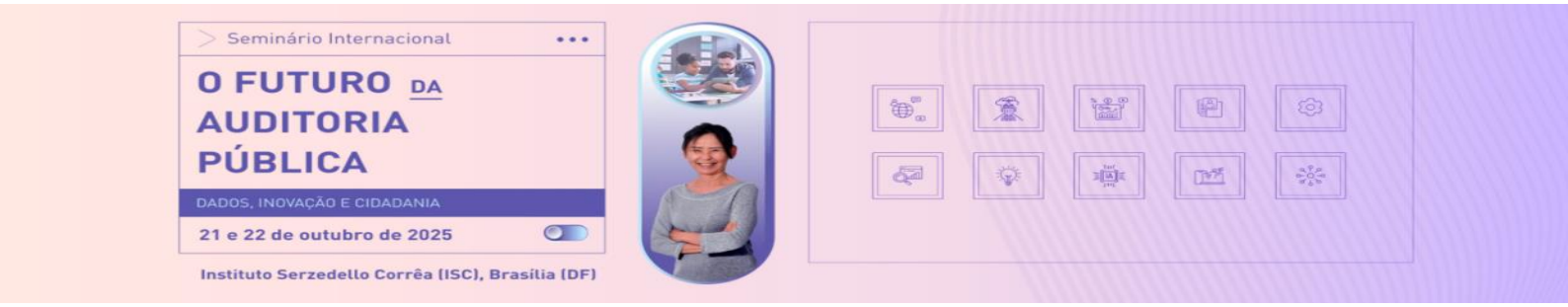

EIXO 4 – INTELIGÊNCIA ARTIFICIAL E TRANSFORMAÇÃO DIGITAL NA AUDITORIA PÚBLICA

COMPARAÇÃO ENTRE RANDOM FOREST E LLMS VIA PROMPT PARA CLASSIFICAÇÃO DE ATOS ADMINISTRATIVOS EM DIÁRIO OFICIAL

Wellington Souza Amaral
Auditor de Controle Externo do Tribunal de Contas do Estado do Rio de Janeiro (TCE-RJ). Mestre em Ciência da Computação pelo CEFET/RJ, com dissertação em mineração de dados em grafos aplicada à fiscalização de recursos públicos. Especialista em Gestão de Projetos (UCAM, 2012) e graduado em Sistemas para Internet (CEFET/RJ, 2008). Atua desde 2012 no TCE-RJ como auditor especializado em **tecnologia da informação e análise de dados aplicada à auditoria governamental**, desenvolvendo métodos e ferramentas para produção de informações estratégicas voltadas ao controle externo. Tem experiência em projetos de **inteligência artificial, aprendizado de máquina e processamento de linguagem natural**, aplicados a compras públicas, diários oficiais e riscos em contratações governamentais. Publicou artigos e participou de projetos de pesquisa sobre **classificação automatizada de notas fiscais eletrônicas** e sobre **governança de TI em municípios jurisdicionados**. Também é professor em cursos de capacitação da Escola de Contas e Gestão do TCE-RJ, ministrando disciplinas de **análise de dados aplicada à auditoria governamental**.

Gustavo Alexandre
Funcionário do TCE-RJ e doutorando em computação na UFF, atuando como Cientista de Dados nas áreas de BI, Analytics e Inteligência Artificial da Subsecretaria de Tecnologia da Informação. Possui experiência em Engenharia de Software, Administração de Banco de Dados, Ciência de Dados e Inteligência Artificial. Já participou de vários projetos de Ciência de Dados nos segmentos de Engenharia, Óleo & Gás, Educação e Auditoria e Controle. É também professor do curso de Pós-graduação Big Data e Inteligência de Marketing (ESPM) e do MBA em Ciência de Dados do Programa de Pós-graduação em Administração (UFF). É mestre em Ciência da Computação (CEFET/RJ), MBA em Data Warehouse e Business Intelligence (PUC-Rio) e Graduado em Ciência da Computação (UESC).

RESUMO

Este artigo compara o método Random Forest, previamente aplicado a 3.501 parágrafos do Diário Oficial do Estado do Rio de Janeiro (acurácia de 98,96%), com

um experimento utilizando *Large Language Models* (LLMs) via Microsoft Copilot. O LLM alcançou 99,37% de acurácia, com baixo esforço técnico, já que a execução ocorreu em nuvem e sem necessidade de programação. Os resultados mostram que LLMs oferecem alternativa acessível e eficaz para auditoria pública, alinhada à agenda de inovação e transformação digital no controle externo.

Palavras-chave: Auditoria Pública; Diário Oficial; Inovação em Controle Externo; Inteligência Artificial; Random Forest; Large Language Models (LLMs); Transformação Digital.

1 INTRODUÇÃO

A identificação de atos administrativos publicados em Diários Oficiais (DO) representa um desafio recorrente para órgãos de controle, devido ao grande volume de dados desestruturados e à necessidade de classificar corretamente categorias como nomeações e exonerações. Em estudo anterior, Amaral, Santos, Silva, Lima e Almeida (2025) aplicou o classificador Random Forest (RF) sobre 3.501 parágrafos do DO do Estado do Rio de Janeiro, alcançando 98,96% de acurácia. O presente trabalho propõe a comparação desse resultado com uma abordagem baseada em *Large Language Models* (LLMs), utilizando a ferramenta Microsoft Copilot, que dispensa codificação e torna o processo acessível a usuários não especialistas.

Diversos estudos têm explorado o uso de técnicas de processamento de linguagem natural e aprendizado de máquina para estruturar informações publicadas em DO, incluindo iniciativas como o *Querido Diário* da Open Knowledge Brasil (OKBR, 2024) e pesquisas acadêmicas que aplicaram métodos de classificação em contextos jurídicos e administrativos (Constantino et al., 2022; Pinto et al., 2021). Nesse cenário, Amaral et al. (2025) desenvolveu um experimento utilizando o classificador RF, associado à tokenização por meio do modelo DistilBERT, sobre uma amostra. A base foi dividida em 80% para treino e 20% para teste, alcançando 99% de acurácia, o que evidenciou o potencial desse método supervisionado para a tarefa de classificação de atos de nomeação e exoneração.

O método RF, proposto por Breiman (2001) a partir das árvores de decisão (BREIMAN et al., 1984), foi implementado com o apoio da biblioteca scikit-learn (PEDREGOSA et al., 2011) e obteve alto desempenho na classificação. Para lidar com a linguagem dos DO, foi utilizada a tokenização pelo modelo DistilBERT (SANH et al., 2019), derivado do BERT (DEVLIN et al., 2019), que permite tratar variações e termos raros. Esses elementos metodológicos sustentam a pesquisa anterior e servem de base para a comparação com os resultados obtidos neste estudo com LLM via Copilot.

Para garantir uma comparação direta, foi utilizada a mesma base de dados do estudo de Amaral et al. (2025). Dessa forma, é possível avaliar de forma objetiva o desempenho relativo das duas abordagens — RF e LLM —, destacando não apenas a acurácia obtida, mas também aspectos como acessibilidade, esforço de implementação e aplicabilidade prática em contextos de auditoria pública.

A proposta de comparar métodos tradicionais de aprendizado de máquina, como o RF, com abordagens baseadas em LLMs via Copilot da Microsoft, dialoga diretamente com a transformação digital, uso de inteligência artificial e inovação em auditoria pública. Ao apresentar evidências empíricas sobre a viabilidade do uso de LLMs em tarefas de classificação de atos administrativos, este trabalho contribui para a reflexão sobre como novas tecnologias podem tornar a auditoria mais acessível, eficiente e centrada no cidadão, fortalecendo a transparência e a capacidade de controle do setor público.

2 METODOLOGIA

A metodologia consistiu em aplicar LLMs via Microsoft Copilot para classificar os mesmos 3.501 parágrafos já rotulados no estudo de Amaral et al. (2025). O processo foi realizado em planilha Excel, com prompts estruturados, orientando o modelo a identificar atos de nomeação, exoneração ou outros. O Copilot retornou à categoria e, quando possível, metadados adicionais. A avaliação considerou acurácia global, matriz de confusão e esforço de implementação.

O procedimento consistiu em estruturar um prompt detalhado, definindo: critérios formais para identificação de nomeações e exonerações; exemplos positivos e

negativos; regras de exclusão e desambiguação (e.g., “designar”, “dispensar” ou “tornar sem efeito” não devem ser confundidos). Cada parágrafo do Diário Oficial, disposto nas colunas Id e Descrição do Parágrafo, foi submetido ao Copilot, que retornou a categoria atribuída (Nomeação, Exoneração ou Outro) e, quando possível, metadados adicionais (nome, cargo, unidade gestora).

Os resultados foram avaliados com base em métricas de acurácia global, matriz de confusão e taxa de erro residual. Além da precisão, foram registradas observações sobre esforço de implementação, acessibilidade e usabilidade da ferramenta, de modo a discutir os ganhos práticos da abordagem LLMs em comparação ao modelo RF já reportado em Amaral et al. (2025).

3 RESULTADOS E DISCUSSÃO

A Figura 1 ilustra o prompt utilizado no Microsoft Copilot para classificar os parágrafos do Diário Oficial como Nomeação, Exoneração ou Outro. O texto do prompt inclui definições, exemplos positivos e regras de exclusão, além de instruções para extrair metadados como nome, cargo, matrícula e unidade gestora. Essa estrutura orienta o modelo a lidar com variações linguísticas comuns em atos oficiais e a produzir uma saída organizada em formato Excel, preservando as informações originais e acrescentando colunas com os dados relevantes.

Figura 1 – *Prompt* usado para identificar atos de nomeação e exoneração e estruturar as informações relevantes em formato Excel.

Você receberá um arquivo Excel contendo textos do Diário Oficial (uma linha por item ou parágrafo).
Seu objetivo é classificar cada linha como "Nomeação", "Exoneração" e extrair metadados quando houver um ato válido.
Como ler o arquivo: Coluna Id, identifica de maneira única cada registro. Coluna DescricaoParagrafo, registro/parágrafo do diário oficial.
Definições e critérios - Ato de NOMEAÇÃO: Ocorre quando alguém é nomeado para um cargo ou função pública (ex.: cargo em comissão – CC; função gratificada – FG).
Sinais fortes (variações aceitas, inclusive com pronomes, artigos e flexões):
"Nomear [nome] … para [cargo|função] …";"Fica nomeado(a) [nome] … para/exercer [cargo|função] …";"Resolve nomear [nome] …"; "Nomeia-se [nome] …", "Nomeio [nome] …";"Nomeia [nome]..."
Ignore atos que sejam exoneração, dispensa, designação, substituição, lotação, remoção, cessão, revogação, tornar sem efeito, apostilamento, licença, posse (sem nomeação), ou outros que não criem provimento em cargo/função. Exemplos corretos de Nomeação: "Nomear Anderson Guedes Ribeiro, para ocupar o cargo de provimento em comissão de Assessor de Secretaria, símbolo CC1, da(o) Instituto de Educação."; "Fica nomeada Norma Suely Cristina Cataldo Izoldi para exercer a função gratificada de Assistente Pedagógico de Empreendedorismo, símbolo FGPE."; "Resolve nomear Jackeline de Fátima Pena para o cargo de Chefe de Setor, símbolo CC4, na Secretaria Municipal de Saúde."
Ato de EXONERAÇÃO: Ocorre quando alguém é desligado de um cargo ou função (ex.: CC, FG, FGS).
Sinais fortes (variações aceitas): "Exonerar [nome] … do cargo/da função …"; "Exonera [nome] ..."; "Fica exonerado(a) [nome] …";"Exonera-se [nome] …", "Exonero [nome] …"
Ignore atos que sejam nomeação, designação, dispensa (não confundir com exoneração), substituição, lotação, remoção, cessão, revogação, tornar sem efeito, licença etc.
Exemplos corretos de Exoneração: Exonerar Fábio Castilho de Souza, da função gratificada de Assistente Operacional, símbolo FGS, da(o) Secretaria Municipal de Educação." ; "Fica exonerado Plinio Marcus Dutra Pinheiro, do cargo de Gerente de Atendimento, símbolo CC3, do Instituto de Educação do Município de Resende."; "EXONERAR, com validade a contar de 25 de abril de 2020, ANDRÉ HENRIQUE DE OLIVEIRA SILVA, id funcional Nº 2393945-1, CEL PM, do cargo em comissão de Comandante Intermediário, símbolo DG, do Comando de Policiamento Ambiental, da Subsecretaria de Gestão Operacional da Polícia Militar, da Subsecretaria Geral de Polícia Militar, da Secretaria de Estado de Polícia Militar. Processo nº SEI-350009/010314/2024."; " Id: 2565319 Id: 2565283 Secretaria de Estado de Administração Penitenciária SECRETARIA DE ESTADO DE ADMINISTRAÇÃO PENITENCIÁRIA ATOS DA SECRETÁRIA DE 08.05.2024 EXONERA JOSE ROBERTO LEITÃO MAUES, Inspetor de Polícia Penal, ID Funcional nº 43547613, com validade a contar de 03 de março de 2024, do cargo em comissão de Chefe, símbolo DAI-5, da Seção II de Turma de Inspetor, do Serviço de Segurança e Disciplina, da Cadeia Pública Inspetor Luis Cesar Fernandes Bandeira Duarte, da Coordenação de Unidades Prisionais do Grande Rio, da Superintendência de Gestão Operacional, da Subsecretaria de Gestão Operacional, da Secretaria de Estado de Administração Penitenciária. Processo nº SEi-210001/031886/2024."
Observações de desambiguação:
"Dispensar" (ex.: de comissão de licitação, conselho) não é exoneração nem nomeação.
"Designar" (atribuição temporária/encargo) não é nomeação.
"Tornar sem efeito"/"Revogar": tratar como Outro (a menos que o próprio texto descreva claramente a nomeação/exoneração vigente).
Textos podem vir em caixa alta, com pontuação irregular, quebras de linha, símbolos de cargo (CC1, FG-3, FGS), ou com a forma passiva ("fica nomeado/exonerado"): todas as variações valem.

Para cada linha classificada como Nomeação ou Exoneração, retorne: nome: pessoa nomeada/exonerada (texto completo do nome próprio). matrícula: se aparecer (ex.: "matrícula 12345"); senão, vazio. cargo: cargo ou função (ex.: "Assessor de Secretaria", "Chefe de Setor", "FGPE", "CC3"). ato: "Nomeação" ou "Exoneração". parágrafo: o parágrafo completo (ou a linha/registro) onde o ato foi identificado. unidade: órgão/unidade gestora (ex.: secretaria, instituto) se identificável no texto.
Como devolver o resultado (no próprio Excel)
Gere um novo arquivo com o nome do arquivo original seguido de "_classificado "+a data e hora atual para download.
Copie todas as colunas originais e adicione as seguintes colunas novas, nesta ordem: nome; matrícula; cargo; ato; parágrafo; unidade
Validação mínima: Se detectar apenas parte dos campos (ex.: sem matrícula ou sem data), preencha o que for possível e deixe os demais vazios. Garanta que o ato esteja sempre preenchido com um dos três valores.

Fonte: Elaboração pelos autores.

A tabela a seguir mostra uma matriz de confusão, métrica usada para avaliar o desempenho de um modelo de classificação. Nela, as linhas representam os valores corretos (o que realmente ocorreu) e as colunas representam os valores preditos pelo modelo (o que o modelo classificou).

No exemplo da tabela, vemos duas classes principais — Nomeação e Exoneração — além da categoria “Outro”. Quando o valor correto é Nomeação, o modelo acertou em 47,47% dos casos, não confundiu com Exoneração, mas classificou 0,11% incorretamente como “Outro”. Já quando o valor correto é Exoneração, o modelo acertou em 51,90%, errou em 0,37% classificando como Nomeação e em 0,14% como “Outro”.

Tabela 1: Resultado da matriz de confusão com os resultados com o método LLM:

		Valor Predito			
		Nomeação	**Exoneração**	**Outro**	**Total por classe**
Valor Correto	**Nomeação**	47,47%	0%	0,11%	**47,58%**
	Exoneração	0,37%	51,90%	0,14%	**52,41%**

Classificado corretamente acurácia geral do modelo LLM: 99,37%

Dos resultados alcançados por Amaral et al. (2025), o desempenho foi igualmente elevado, mas com algumas diferenças. Entre os registros de nomeação, 55,47% foram corretamente identificados e 1% confundidos como exoneração. Já entre os registros de exoneração, 43,50% foram corretamente classificados e 0,46% confundidos como nomeação. O modelo obteve taxa global de acerto de 98,96% e erro de 1,04%. Embora o índice de acerto seja ligeiramente inferior ao do LLM, a árvore binária mostrou maior precisão para identificar nomeações, mas menor para exonerações, o que sugere uma tendência do modelo em privilegiar a classe majoritária.

Tabela 2: Resultado da matriz de confusão com os resultados com o método (RF):

		Valor Predito		
		Nomeação	**Exoneração**	**Total por classe**
Valor Correto	**Nomeação**	55,47%	1%	**56,04%**
	Exoneração	0,46%	43,50%	**43,96%**

Classificado corretamente acurácia geral do modelo RF: 98,96%

O experimento com LLM obteve desempenho elevado na classificação dos atos administrativos. A análise da matriz de confusão demonstrou que o modelo alcançou 99,37% de acurácia global, com erros residuais distribuídos de forma equilibrada entre as classes. No caso das nomeações, 47,47% dos registros foram corretamente identificados, enquanto apenas 0,11% foram classificados incorretamente como “Outro”. Para as exonerações, 51,90% foram reconhecidas corretamente, com pequenas taxas de erro: 0,37% como nomeação e 0,14% como “Outro”.

Do ponto de vista prático, o LLM via Copilot mostrou-se mais acessível para usuários sem experiência em programação, permitindo a classificação direta com a elaboração de prompts adequados. A execução ocorreu de forma simplificada, sem necessidade de programação. O principal esforço foi o aperfeiçoamento do prompt, que passou a ser o elemento central para orientar o modelo a identificar corretamente os atos administrativos. Já o RF, relatado por Amaral et al. (2025), exige conhecimento técnico em programação, bibliotecas de NLP e ajustes de hiperparâmetros, o que limita seu uso por equipes de auditoria sem suporte especializado.

Em síntese, ambos os métodos apresentaram resultados robustos e confiáveis, mas a escolha entre eles depende do contexto: enquanto o RF é mais indicado em cenários com infraestrutura técnica consolidada e necessidade de replicabilidade controlada, o LLM via Copilot destaca-se pela rapidez de implementação, acessibilidade e flexibilidade, alinhando-se aos debates sobre inovação e transformação digital em auditoria pública.

4 CONCLUSÃO

Os resultados obtidos demonstram que o uso de LLMs via Microsoft Copilot constitui uma alternativa eficaz para a classificação de atos administrativos em DO. A abordagem apresentou desempenho comparável ao método tradicional de RF descrito por Amaral et al. (2025), alcançando acurácia superior a 99% e reduzindo significativamente o esforço técnico necessário para a execução do experimento.

Como desdobramento desta pesquisa, sugere-se ampliar os experimentos para outros atos administrativos (contratos, licitações, convênios), bem como explorar o uso de LLMs em multiclassificação e extração de entidades. Também se destacam discussões sobre riscos de alucinação, custos de operação em larga escala e alternativas de código aberto, reforçando o potencial da inteligência artificial para a inovação e a transformação digital no controle externo.

5 REFERÊNCIAS

AMARAL, Wellington Souza; SANTOS, Gustavo Alexandre Sousa; SILVA, Eduardo Bezerra da; LIMA, Leonardo Silva de; ALMEIDA, Augusto César Benvenuto de. Desenvolvimento de uma metodologia para a coleta e identificação de atos administrativos de interesse nos Diários Oficiais dos jurisdicionados do Tribunal de Contas do Estado do Rio de Janeiro (TCE-RJ). Revista Síntese, Rio de Janeiro, aceito para publicação em 2025, no prelo.

BREIMAN, L., Friedman, J., Olshen, R.A., & Stone, C.J. (1984). Classification and Regression Trees (1st ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781315139470.

BREIMAN, L. Random Forests. Machine Learning 45, 5–32. 2001. https://doi.org/10.1023/A:1010933404324.

K. Constantino, V. A. L. Cruz, O. M. Zucheratto, C. França, M. Carvalho, T. H. Silva, A. H. Laender, and M. A. Gonçalves. Segmentação e classificação semântica de

trechos de diários oficiais usando aprendizado ativo. In Anais do XXXVII Simpósio Brasileiro de Bancos de Dados, pages 304–316. SBC, 2022.

DEVLIN J., CHANG M., LEE K., and TOUTANOVA K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186, 2019. Minneapolis, Minnesota. Association for Computational Linguistics.

OKBR (Open Knowledge Brasil). Sobre o Querido Diário. [S.l.: s.n.], 2024. Disponível em: https://queridodiario.ok.org.br/sobre.

PEDREGOSA, F. et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, v. 12, p. 2825-2830, 2011.

F. A. S. PINTO, E. H. Haeusler, and S. Lifschitz. Transparência pública automatizada a partir da gramática do diário oficial. Anais do Workshop de Computação Aplicada em Governo Eletrônico (WCGE 2021), 2021. URL https://api.semanticscholar.org/CorpusID:237674517.

SANH, V., DEBUT, L., CHAUMOND, J., & WOLF, T. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. 2019. arXiv preprint arXiv:1910.01108.