

INTELIGÊNCIA ARTIFICIAL APLICADA A DOCUMENTOS EM SUPORTE

PAPEL: o caso do sistema SEP no ES¹

ARTIFICIAL INTELLIGENCE APPLIED TO PAPER SUPPORT DOCUMENTS:

the case of the SEP system in ES

Jussara Teixeira*
Juliana Oliveira de Almeida**
Tania Barbosa Salles Gava***
Tatiana Canelhas****
Jose Márcio Moraes Dorigueto*****

RESUMO

Com a transformação digital ocorrida nos últimos tempos, principalmente após o cenário pandêmico, a automatização de processos tornou-se parte importante da gestão administrativa, inclusive na gestão arquivística de documentos. Atualmente, também tem crescido o interesse pelo uso de ferramentas de Inteligência Artificial no contexto da transformação digital. Sendo assim, este artigo tem como objetivo geral apresentar a funcionalidade do sistema SEP que utiliza a Inteligência Artificial (IA) para a classificação automática de processos administrativos em suporte papel. Trata-se de uma pesquisa do tipo aplicada, de caráter exploratório, cujo método é a pesquisa bibliográfica e documental. Os principais resultados da pesquisa são uma breve contextualização da gestão de documentos arquivísticos e do aprendizado de máquina (Machine Learning), e a apresentação de uma funcionalidade do sistema SEP, que classifica processos administrativos de forma automática utilizando aprendizado de máquina.

Palavras-chave: Gestão de Documentos; Classificação automática de documentos; Aprendizado de Máquina.

ABSTRACT

With the recent digital transformation, particularly following the pandemic scenario, process automation has become an important part of administrative management, including archival document management. Currently, there is also a growing interest in the use of Artificial Intelligence tools within the context of digital transformation. Therefore, the general objective of this article is to present the functionality of the SEP

¹Artigo atualizado, oriundo da comunicação oral realizada no X Congresso Nacional de Arquivologia (CNA) em Salvador, Bahia, entre 14 e 18 de outubro de 2024.

* Analista de TI | Mestranda em Informática | jussarateixeirafalcao@gmail.com

** Arquivista | Graduada | julianaalmeida.apees@gmail.com

*** Professor | Pós-Doutor em Ciência da Informação | tania.gava@ufes.br

**** Cientista da Computação e Arquivista | Mestranda em Ciência da Informação | tatyacs@gmail.com

***** Analista de TI | Mestre em Engenharia Ambiental | jose.dorigueto@prodest.es.gov.br



system that uses Artificial Intelligence (AI) for the automatic classification of administrative processes in paper support. This is applied, exploratory research, using bibliographic and documentary methods. The main results of the research include a brief contextualization of archival document management and machine learning, and the presentation of a SEP system functionality that automatically classifies administrative processes using machine learning.

Keywords: Document Management; Automatic Document Classification; Machine Learning.

1 INTRODUÇÃO

O Governo do Estado do Espírito Santo durante quase três décadas utilizou o SEP (Sistema Eletrônico de Protocolo) como sistema de gerenciamento e controle das operações referentes às funções de recebimento, registro e tramitação de processos na Administração Pública Estadual. O sistema começou a ser utilizado no final da década de 1980 e até o ano de 2005 ainda não tinham sido utilizados os instrumentos básicos da Gestão Documental, que são o Plano de Classificação de Documentos (PCD) e a Tabela de Temporalidade de Destinação de Documentos (TTD). Assim, por mais de 20 anos foram atuados processos administrativos sem critérios arquivísticos, o que acarretou o acúmulo de cerca de 09 milhões de processos até o período de maio de 2024.

Com o crescente acúmulo de documentos nos órgãos e entidades da Administração Estadual ao longo dos anos, surgiu a necessidade da busca de outros espaços, fora dos órgãos, para guarda intermediária dos acervos. Como resultado, houve um aumento significativo na contratação de serviços de guarda externa, gerando um alto custo para a Administração Pública, chegando a 18 milhões de reais por ano em 2018, como também a locação de espaços para instalação dos arquivos intermediários. É importante ressaltar que a guarda externa e a locação de imóveis não resolvem os problemas de massa documental acumulada, sendo necessário o tratamento arquivístico, que inclui classificação e avaliação para a correta destinação dos documentos, ou seja, sua eliminação ou recolhimento para guarda permanente, conforme a legislação vigente (Lei nº 8.159, de 8 de janeiro de 1991).

Em 2017 a Secretaria de Estado de Controle e Transparência do Estado do Espírito Santo (SECONT), realizou uma auditoria em Gestão Documental em que um dos indicadores de desempenho analisados foi a contratação de sistemas, serviços,



consultorias e equipamentos relacionados à Gestão Documental. Dentre os 10 órgãos auditados, 04 (quatro) apresentaram cópias dos contratos vigentes, e o valor pago entre os anos de 2016 e 2017 chegou ao total de R\$ 11.925.994,18 (onze milhões, novecentos e vinte e cinco mil, novecentos e noventa e quatro reais e dezoito centavos).

Para sanar os problemas apresentados faz-se necessário o tratamento arquivístico e a consequente eliminação dos documentos que já cumpriram seus prazos de guarda. A eliminação trará para a Administração Pública redução nos gastos com contratos de guarda externa de documentos, economia com insumos e manutenção de espaços que poderiam ser melhor aproveitados. Além das economias citadas, a eliminação tem o papel social, tendo em vista que todo material proveniente da fragmentação é revertido em benefício de associações de catadores de material reciclável.

O atual desafio da Administração Estadual é tratar o legado a fim de eliminar a documentação que já cumpriu os prazos de guarda, bem como fazer o recolhimento da documentação de caráter permanente. A dificuldade em realizar uma gestão documental efetiva se dá por diversos fatores, tais como falta de arquivistas nos órgãos, membros das comissões de avaliação de documentos setoriais (CADS) com muitas atribuições além das atividades voltadas ao arquivo, órgãos da Administração que ainda não possuem Plano de Classificação de Documentos e Tabela de Temporalidade e Destinação para apoiar as atividades de gestão, e em muitos casos falta de apoio da alta gestão para priorizarem o trabalho nos acervos.

Para o adequado tratamento arquivístico, a primeira etapa é a correta classificação dos documentos. No cenário ideal os processos recebem a classificação no momento de sua produção, o que não aconteceu no passado. De acordo com o Dicionário Brasileiro de Terminologia Arquivística (Arquivo Nacional, 2020, p. 49), entende-se por classificação:

- “1) Organização dos documentos de arquivo ou coleção, de acordo com um plano de classificação, código de classificação ou quadro de arranjo.
- 2) Análise e identificação do conteúdo de documentos, documentos seleção da categoria de assunto sob a qual sejam recuperados, podendo-se-lhes atribuir códigos.

Neste contexto, este artigo tem como objetivo geral apresentar uma funcionalidade do Sistema Eletrônico de Protocolo (SEP) que utiliza a Inteligência Artificial para a classificação automática de processos administrativos em suporte papel.



A pesquisa desenvolvida neste trabalho é do tipo aplicada, de caráter exploratório, cujo método é a pesquisa bibliográfica e documental (Gil, 2002), e encontra-se dividida da seguinte maneira: a seção 2 apresenta a gestão de documentos arquivísticos. Na seção 3 faz-se uma introdução ao aprendizado de máquina (Machine Learning). A seção 4 descreve uma visão geral da aplicação de aprendizado de máquina para a classificação de processos administrativos em suporte papel. A seção 5 apresenta as considerações finais, seguida das referências bibliográficas.

2 GESTÃO DE DOCUMENTOS ARQUIVÍSTICOS

Com a transformação digital ocorrida nos últimos tempos, principalmente após o cenário pandêmico, a automatização de processos tornou-se parte importante da gestão administrativa. No entanto, no contexto arquivístico, essa gestão precisa primordialmente abarcar uma gestão arquivística comprometida em adotar normas, padrões, modelos e a legislação arquivística vigente. Assim, para a aplicação da ferramenta de Inteligência Artificial na classificação de processos no contexto do estado do Espírito Santo, foi necessário que o governo avançasse na gestão de documentos, de modo que a grande maioria dos órgãos e entidades da administração estadual tivessem CADS atualizadas e ativas, bem como seus instrumentos de Gestão Documental, o Plano de Classificação de Documentos (PCD) e a Tabela de Temporalidade e Destinação de Documentos (TTD) publicados e atualizados.

O processo de classificação e avaliação consistem, respectivamente, em analisar o conteúdo do documento avulso ou processo a fim de recuperar o seu contexto de produção, atribuindo assim um código conforme o PCD, e a posterior consulta a TTD para identificar em que fase do ciclo de vida o documento se encontra: se na corrente, intermediária ou se está apto para a destinação final (eliminação ou recolhimento para guarda permanente). Tais etapas são a base para uma gestão documental efetiva tendo em vista que é ela que permite identificar a destinação dos documentos. Classificar os documentos é uma tarefa relativamente simples, e o ideal é que os documentos recebam a classificação no momento de sua produção, mas o cenário atual é a existência de uma massa documental acumulada sem critérios arquivísticos.



3 INTRODUÇÃO AO APRENDIZADO DE MÁQUINA

O aprendizado de máquina (*machine learning* - ML) é um subcampo da inteligência artificial focado no desenvolvimento de algoritmos capazes de aprender a partir de dados e fazer previsões ou tomar decisões sem a necessidade de programação explícita. Esses algoritmos de aprendizado de máquina permitem que os computadores analisem grandes quantidades de dados, identifiquem padrões e façam previsões ou tomem decisões em tempo real, conforme analisado por Li, Lin e Guan (2023). A evolução desse campo remonta ao Teste de Turing, que buscava definir operacionalmente a inteligência de um computador. Segundo Russell e Norvig (2010):

O teste de Turing, proposto por Alan Turing em 1950, foi feito para prover uma definição operacional satisfatória de inteligência. Um computador passa no teste se um interrogador humano, após fazer algumas perguntas por escrito, não conseguir distinguir se as respostas escritas vêm de uma pessoa ou de um computador (tradução nossa; Russell, Norvig; 2016; Capítulo 1, p. 2).

A partir da década de 1960, algoritmos de aprendizado de máquina começaram a ser desenvolvidos para resolver problemas voltados ao reconhecimento facial e de objetos. Os princípios básicos do aprendizado de máquina envolvem o uso de dados para treinar modelos matemáticos que identificam padrões e fazem inferências. Esses modelos são divididos em três categorias principais: aprendizado supervisionado, não supervisionado e por reforço. No aprendizado supervisionado, como explica Murphy (2012), o modelo é treinado com um conjunto de dados rotulados onde as respostas corretas são conhecidas, sendo ideal para tarefas de classificação e regressão. O aprendizado não supervisionado, como descrito por Hastie, Tibshirani e Friedman (2009), envolve o treinamento do modelo em dados sem rótulos, buscando descobrir estruturas ou padrões ocultos. Técnicas comuns incluem clustering e redução de dimensionalidade. Já o aprendizado por reforço (*reinforcement learning* - RL), segundo Sutton e Barto (2018), envolve a aprendizagem de tomar decisões sequenciais com feedback na forma de recompensas ou penalidades, sendo amplamente utilizado em jogos e controle de robôs.



Uma das abordagens dentro do ML é o uso de redes neurais artificiais. Goodfellow, Bengio e Courville (2016) descrevem essas redes como sistemas que imitam o funcionamento do cérebro humano, compostos por camadas de neurônios artificiais que processam e transmitem informações. Elas são eficazes em tarefas complexas como o reconhecimento de padrões em grandes volumes de dados. As redes neurais profundas, que utilizam múltiplas camadas de neurônios, são capazes de extrair características de alto nível dos dados. Um exemplo é a rede neural profunda *feedforward* ou *perceptron* multicamada (MLP). Segundo Goodfellow, Bengio e Courville (2016),

O exemplo clássico de um modelo de aprendizado profundo é a rede neural profunda *feedforward* ou perceptron multicamada (MLP). Um perceptron multicamada é apenas uma função matemática que mapeia um conjunto de valores de entrada para valores de saída. A função é formada pela composição de muitas funções mais simples. Podemos pensar em cada aplicação de uma função matemática diferente como fornecendo uma nova representação da entrada. (tradução nossa; Goodfellow, Bengio, Courville; 2016; Capítulo 1, p. 5).

Essas técnicas são usadas no Processamento de Linguagem Natural (PLN), onde as redes neurais podem ser aplicadas para a classificação automática de documentos, analisando grandes volumes de dados textuais, identificando padrões linguísticos e extraindo informações relevantes, ou seja, automatizando tarefas que seriam altamente complexas e demoradas se realizadas manualmente.

O treinamento de modelos para qualquer projeto de aprendizado de máquina exige um dataset de qualidade. Vancauwenbergh (2019), em seu texto *Data Quality Management*, argumenta que dados de alta qualidade permitem que os modelos aprendam corretamente e façam previsões precisas. A ausência de um *dataset* bem estruturado pode levar a resultados inadequados e modelos com desempenho inferior. A autora também destaca que *datasets* de qualidade devem ser completos, precisos e representativos do problema que se deseja resolver. Além disso, é importante que os dados sejam bem rotulados para que os algoritmos supervisionados possam aprender corretamente as associações entre as entradas e as saídas, conforme argumentado por Shaveta:

Para antecipar eventos futuros, algoritmos de aprendizado supervisionado usam exemplos rotulados para aplicar o que aprenderam no passado a novos dados. O método de aprendizado cria uma função inferida para prever valores de saída examinando um conjunto de dados de treinamento conhecido. Após um treinamento adequado, o sistema



pode fornecer alvos para qualquer nova entrada. Para identificar falhas e corrigir o modelo conforme necessário, ele também pode comparar sua saída com a que é correta e pretendida. (tradução nossa; Shaveta; 2023; p. 282).

A capacidade do aprendizado de máquina de processar grandes volumes de dados e identificar padrões automaticamente o torna uma ferramenta eficaz para o gerenciamento de documentos. Shaveta (2023) observa que os avanços no ML, baseados na criação de modelos matemáticos, aprimoram a precisão e os processos de tomada de decisão por meio de informações extraídas dos dados. Com o aprendizado de máquina, as organizações podem automatizar a classificação de documentos, consolidando essa tecnologia como um componente importante na modernização e transformação digital de suas operações.

4 CLASSIFICAÇÃO DE DOCUMENTOS UTILIZANDO ML

O desafio inicial do projeto de uso de IA no sistema SEP foi implementar uma solução de aprendizado de máquina, apesar da falta de experiência prática da equipe envolvida. Para suprir essa lacuna, houve um esforço direcionado para a realização de uma Prova de Conceito (PoC) e a criação de Mínimos Produtos Viáveis (MVPs). A PoC permite testar se uma solução proposta funciona conforme esperado e atende aos requisitos iniciais, ajudando a identificar possíveis problemas e desafios antes de um desenvolvimento completo. Com os resultados da PoC, a equipe pôde então desenvolver um MVP, que é uma versão inicial do produto com o conjunto mínimo de funcionalidades necessário para ser lançado e usado por clientes. Além disso, a equipe se empenhou em participar de cursos e workshops para adquirir os conhecimentos técnicos necessários para a entrega final.

O segundo grande desafio foi a identificação e o tratamento dos dados necessários para treinar o modelo algorítmico. Para mitigar o problema da qualidade dos dados foi realizada uma análise inicial para identificar padrões e inconsistências. Em seguida, foram implementadas ações automatizadas de limpeza e preparação de dados, utilizando técnicas de pré-processamento para garantir a qualidade dos dados.



O desafio subsequente envolveu a determinação da abordagem mais adequada para o processamento dos dados. Foram testados quatro algoritmos distintos de processamento de linguagem natural. Após uma análise criteriosa baseada em tempo de processamento versus volume de dados, foi identificado que o uso de redes neurais constituía a abordagem mais eficaz.

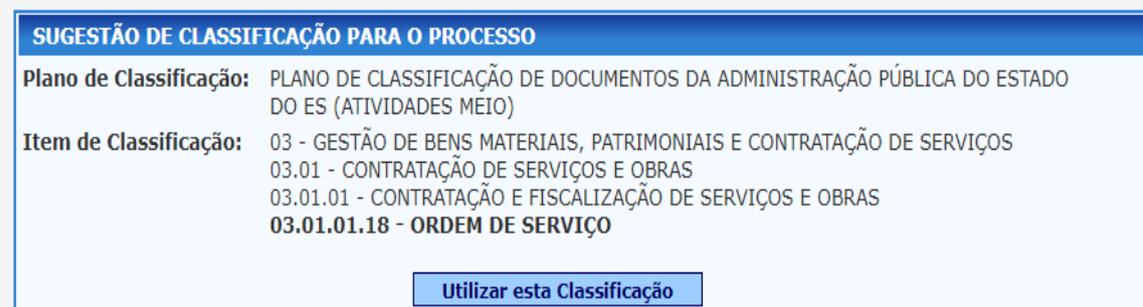
Após a escolha do algoritmo, buscou-se garantir que a solução fosse integrada ao sistema SEP sem comprometer seu funcionamento. Para isso, adotou-se a abordagem de integração por meio de APIs (*Application Programming Interfaces*). As APIs, que são conjuntos de definições e protocolos que permitem a comunicação entre diferentes softwares, foram fundamentais para assegurar uma integração eficiente. Essa abordagem permitiu a adição de uma nova funcionalidade ao sistema SEP sem interferir nas operações existentes, garantindo assim a continuidade e a estabilidade do sistema. Foram realizados testes com esta abordagem e os resultados mostraram que a integração foi bem-sucedida, sem interrupções no sistema SEP. Isso permitiu a implementação gradual da solução garantindo uma transição suave e minimizando qualquer impacto negativo sobre os usuários e as funcionalidades já existentes.

O próximo e último desafio foi garantir que a solução fosse escalável e que pudesse lidar com volumes crescentes de classificação sem perda de desempenho. Para isso, a solução utiliza a plataforma *OpenShift* da *Red Hat*, que propicia uma infraestrutura robusta e flexível, mantendo um desempenho consistente e confiável, mesmo com o aumento do volume de classificação decorrente de sua utilização por todos os órgãos.

A funcionalidade implementada é de fácil uso para o usuário. Ao consultar um número de processo na funcionalidade de “Alterar classificação” no SEP, o usuário recebe a informação exibida na Figura 1. Se a classificação apresentada estiver correta, basta confirmar clicando na opção “utilizar esta classificação”.



Figura 1: Excerto da funcionalidade de classificar processos usando IA no SEP



Fonte: Sistema SEP (<https://www.protocolo.es.gov.br/Default.aspx>)

Consciente da existência de milhões de processos a serem classificados, o SEP também possui recursos para apoiar o usuário na realização de classificações em larga escala. Assim, o usuário pode lidar de maneira eficiente com grandes volumes de dados, otimizando o processo de classificação.

A partir da entrega desta primeira solução, as perspectivas futuras para a continuidade do uso da IA em gestão documental desdobram-se em três vertentes: a primeira é promover sua adoção por todos os órgãos que utilizam o SEP, para que possamos ter toda a massa documental de processos em suporte papel classificada; a segunda é o desenvolvimento tecnológico contínuo da solução; e a terceira é ampliar a utilização da IA para outros problemas relacionados à gestão documental.

5 CONSIDERAÇÕES FINAIS

O projeto representa um avanço significativo ao incorporar o aprendizado de máquina na classificação de documentos arquivísticos. Essa inovação tecnológica substitui um procedimento manual por um sistema automatizado, elevando a eficiência e a precisão no tratamento de grandes quantidades de dados. Com sua adoção, os usuários podem redirecionar seu foco para tarefas mais estratégicas, como o planejamento de descarte de documentos, a atualização contínua do plano de classificação e da tabela de temporalidade, além de aprimorar a gestão de custos relacionados à guarda terceirizada. Isso prevê a diminuição de despesas com aluguel para a guarda de documentos e contribui para uma gestão documental mais eficaz e sustentável. Em conclusão, destaca-se a importância de uma busca constante para uma gestão documental eficiente e moderna,



utilizando tecnologias avançadas, como o uso de ferramentas de IA, para a busca na solução de problemas complexos que necessitam manipular um grande volume de dados.

REFERÊNCIAS

- ARQUIVO NACIONAL. **Dicionário brasileiro de terminologia arquivística**. Rio de Janeiro, 2005. 232p. Disponível em: https://www.gov.br/conarq/pt-br/centrais-de-conteudo/publicacoes/dicionario_de_terminologia_arquivistica.pdf. Acesso em: 21 mai. 2024.
- GIL, A. C. **Como Elaborar Projetos de Pesquisa**. 4. ed. São Paulo: Atlas, 2002. p. 1-175.
- GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning**. Cambridge, MA: MIT Press, 2016. 800 p.
- HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. 2nd ed. New York: Springer, 2009. 764 p.
- LI, Xiaolong; LIN, Wenliang; GUAN, Bei. *The impact of computing and machine learning on complex problem-solving*. **Engineering Reports**, v. 5, n. e12702, p. 1-2, 2023. DOI: 10.1002/eng2.12702.
- MURPHY, Kevin P. **Machine Learning: A Probabilistic Perspective**. Cambridge, MA: MIT Press, 2012. 1067 p.
- RUSSELL, Stuart; NORVIG, Peter. **Artificial Intelligence: A Modern Approach**. 3rd ed. Upper Saddle River, NJ: Pearson, 2010. 1152 p.
- SHAVETA. *A review on machine learning*. **International Journal of Science and Research Archive**, v. 9, n. 1, p. 281-285, 2023. DOI: 10.30574/ijrsra.2023.9.1.0410.
- SUTTON, Richard S.; BARTO, Andrew G. **Reinforcement Learning: An Introduction**. 2nd ed. Cambridge, MA: MIT Press, 2018. 552 p.
- VANCAUWENBERGH, Sadia. **Data Quality Management**. London: IntechOpen, 2019.
- WENLING, L. *The impact of computing and machine learning on complex problem-solving*. **Engineering Reports**, 2023. DOI: 10.1002/eng2.12702.
- YOUNG, Tom; HWANG, Karl; TIAN, Zhuang; CAO, Siyu. *Recent trends in deep learning based natural language processing*. **IEEE Computational Intelligence Magazine**, v. 13, n. 3, p. 55-75, 2018.



Agradecimentos

Agradecemos aos usuários do sistema SEP, cuja participação e feedback são fundamentais para o sucesso da funcionalidade de inteligência artificial, contribuindo para a modernização e eficiência na gestão documental no Poder Executivo do Estado do Espírito Santo.

Também agradecemos à equipe de Machine Learning do Prodest, cujo trabalho e dedicação viabilizaram o desenvolvimento dessa solução pioneira para a classificação automática de processos administrativos. Reconhecemos e valorizamos suas contribuições técnicas e compromisso com o projeto.

A equipe Prodest foi composta por:

- GEDAD - Gerência de Dados
 - Messias Gomes da Silva

- GESIT - Gerência de Sistemas
 - José Márcio Moraes Dorigueto
 - Juan Xabier Esteban de Aquino Calles
 - Jussara Teixeira

- GEPTIC - Gerência de Infraestrutura de TIC
 - Daniel Carvalho

