



APLICAÇÃO DE APRENDIZADO DE MÁQUINA EM METADADOS NA REPRESENTAÇÃO DOS DADOS DE PESQUISA

Edvaldo Wellington da Cunha Monteiro

Universidade Federal Rural da Amazônia (UFAM), Brasil | edvaldo.monteiro@ufra.edu

 <https://orcid.org/0009-0006-2373-8376>

DOI: 10.22477/xiv.biredial.413

EJE TEMÁICO: Infraestructura tecnológica

RESUMEN

Esta pesquisa aborda, de maneira teórica, o aprendizado de máquina e investiga sua relação com metadados e se justifica apresentando a relevância dessa conexão tanto em técnicas quanto em aplicações nos dados de pesquisa. Por meio de uma abordagem exploratória e bibliográfica, o estudo apresenta alguns conceitos fundamentais dos temas e estabelece um entrelaçamento teórico, reforçando a necessidade de considerar o aprendizado de máquina como uma ferramenta capaz de melhorar a qualidade de dados e metadados. Conclui-se com a sugestão de técnicas automatizadas de análise de dados para geração e produção de metadados na representação e administração dos dados de pesquisa.

Palabras-clave: Metadatos; Aprendizado de Máquina, Dados de pesquisa.

ABSTRACT

This research addresses, in a theoretical manner, machine learning and investigates its relationship with metadata, and is justified by presenting the relevance of this connection in both techniques and applications within research data. Through an exploratory and bibliographic approach, the study presents some fundamental concepts of the themes and establishes a theoretical interweaving, reinforcing the need to consider machine learning as a tool capable of improving the quality of data and metadata. It concludes with the suggestion of automated data analysis techniques for the generation and production of metadata in the representation and administration of research data.)

Keywords: Metadata; Machine learning; Research data.

1 INTRODUÇÃO

Ao se considerar a dinamização do acesso e recuperação de informações e a produção e uso massivo de dados em escala global, leva-se em conta os impactos advindos da e-Science, a qual demanda a especialização de tarefas e serviços por meio compartilhamento de recursos



computacionais, acesso a grandes conjuntos de dados e uso de plataformas digitais (Sales & Sayão, 2023).

Esses impactos, fundamentam um novo enfoque metodológico baseado no uso intenso de dados para a ciência, no qual estes dados são examinados de novas maneiras e consequentemente isso ocorre através do uso de tecnologias digitais de informação, como a simulação computacional e algoritmos especializados (Gray, 2007), além disso, tais processos requerem novos métodos de análises de grandes conjuntos de dados (Sales e Sayão, 2023).

É essencial, pois, que os conjuntos de dados sejam acessíveis e interpretáveis por humanos e máquinas através de metadados. Neste sentido, considera-se a relevância do aprendizado de máquina, subárea interdisciplinar da Ciência da computação, que pode transformar a representação de dados de pesquisa na Ciência da Informação, através de seus algoritmos que podem atuar na análise de grandes volumes de dados.

Deste modo, torna-se essencial analisar, de modo breve mas informativo, os impactos do aprendizado de máquina quando aplicados a metadados no escopo da Ciência da Informação, buscando compreender as transformações que estes podem trazer aos metadados ao otimizá-los em processos como descrição, acesso, uso e recuperação de informação em dados de pesquisa, além de avaliar as implicações decorrentes de sua utilização.

Portanto o objetivo deste trabalho é apresentar os conceitos de aprendizado de máquina e descrever sua funcionalidade e aplicabilidade através de alguns modelos de algoritmos de aprendizado de máquina, bem como aprofundar conceito de metadados e analisar sua relação com dados de pesquisa.

1.1 METODOLOGIA

Como metodologia esta pesquisa combina um estudo exploratório, cuja abordagem metodológica utilizada para esta pesquisa é uma revisão de literatura combinada com uma pesquisa bibliográfica e para o procedimento de coleta de dados utilizou-se uma análise de conteúdo da literatura especializada disponível em bases científicas, pois busca aprofundar a compreensão sobre um tema ainda em desenvolvimento e identificar lacunas de pesquisa, neste caso, investiga a aplicação do aprendizado de máquina em metadados no escopo da Ciência da Informação. A escolha de tal método se dá pelo fato da pesquisa exploratória, como informa Gil (2019, p. 26), ter como propósito trazer maior familiaridade para um problema ao torná-lo mais explícito.

Para reforçar a coleta bibliográfica, foram selecionadas fontes relevantes a partir de bases de dados acadêmicas reconhecidas nacional e internacionalmente, incluindo: Web of Science: Para identificar artigos de alto impacto relacionados à Ciência da informação, Aprendizado de máquina, dados de pesquisa e metadados. Elsevier (ScienceDirect): Com foco em



artigos técnicos e revisões sistemáticas na área de Aprendizado de máquina e metadados e BRAPCI (Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação): Para identificar estudos específicos do contexto brasileiro na Ciência da Informação a respeito de metadados e dados de pesquisa.

Em sequência, foram aplicados certos critérios de seleção dos artigos que incluíram:

1. Publicação de trabalhos nos últimos 10 anos (2013-2023), salvo referências clássicas da área.
2. Presença dos termos “machine learning,” “metadata,” “research data” e seus equivalentes em português (aprendizado de máquina, metadados e dados de pesquisa).
3. A relevância comprovada pela quantidade de citações e relação direta com o tema.

Foram recuperados 148 arquivos (Web of science: 49 artigos; Elsevier: 37 artigos e BRAPCI: 62 arquivos), os quais após análise de critérios de exclusão, como textos repetidos (42), Cobertura do assunto de modo muito breve (29), Textos técnicos e que necessitavam de maior expertise (18), totalizando (89) textos retirados da pesquisa. Quanto aos artigos restantes (59), foram objetos de análise para estruturar critérios para identificar, analisar e correlacionar conceitos e aplicações relativas ao aprendizado de máquina, metadados e dados de pesquisa.

Tais critérios foram necessários para delimitar as etapas da pesquisa, na definição do Corpus teórico, Onde Foi estabelecida uma lista inicial de termos-chave, como “aprendizado de máquina,” “metadados,” e “dados de pesquisa” e seus equivalentes em inglês para nortear as buscas. Em seguida foi realizada a triagem de fontes subdivididas em: leitura de títulos, resumos e, posteriormente, do texto completo para identificar a pertinência ao tema. A partir disso, pode-se categorizar os artigos conforme suas contribuições práticas, teóricas ou metodológicas e portanto, ao final foram selecionados 26 artigos para o desenvolvimento deste artigo.

Ao finalizar a triagem das fontes conseguiu-se realizar a análise de conteúdo, a qual é uma técnica de análise proposta por Bardin (2011) para identificar padrões, categorias e relações entre os conceitos analisados. Isto tornou possível realizar a análise dos dados coletados, os quais quando sistematizados permitiram destacar: algumas possíveis aplicações práticas do aprendizado de máquina em metadados e benefícios e limitações das abordagens atuais e desafios do aprendizado de máquina no escopo da Ciência da Informação.

2 APRENDIZADO DE MÁQUINA

O Aprendizado de Máquina (*Machine learning*), termo cunhado por Arthur Samuel em 1958, é um campo vasto da Inteligência Artificial, sendo uma das suas principais subáreas. O Aprendizado de máquina é utilizado em diferentes ramos de pesquisa e se baseia em modelos computacionais e estatísticos na interpretação de dados.



A origem do aprendizado de máquina está relacionada com a pesquisa de Arthur Samuel sobre o desenvolvimento de máquinas capazes de resolverem problemas sem serem explicitamente programadas.

Segundo Hackeling (2014) o aprendizado de máquina é o estudo e design de artefatos de software que usam a experiência passada para tomar decisões futuras, ou seja, estuda como programas aprendem com os dados.

Tal definição complementa o conceito proposto por Mitchell (1997), no qual disse que o aprendizado de máquina tem por finalidade o desenvolvimento de programas de computador capazes de se aprimorar automaticamente com base na experiência adquirida através de dados.

Essa assertiva é próxima do que propôs Samuel em 1959, quando disse que o Aprendizado de máquina é um: “campo de estudo que oferece aos computadores a capacidade de aprender sem serem explicitamente programados”.

Por fim, autores como Facelli *et al.* (2011), Goodfellow, Bengio e Courville (2016) e Perichó e Bianchi (2011) enfatizam que o aprendizado de máquina combina heurísticas (técnicas que influenciam a eficiência na solução de problemas, acelerando o aprendizado) e estatísticas aplicadas (métodos para coletar, analisar, interpretar e apresentar dados). Ambas as abordagens são focadas na solução de problemas e na aceleração do aprendizado por meio de algoritmos especializados, destacando a análise de dados como um componente essencial para que isso ocorra.

2.1 ALGORITMOS DE APRENDIZADO DE MÁQUINA

Os algoritmos são usados em diferentes contextos, dependendo dos objetivos (classificação, previsão, agrupamento) e da natureza dos dados (rotulados, não rotulados, ambientes interativos). E a utilização destes nos tipos de aprendizado são determinantes para o aprendizado de máquina, pois na resolução de problemas, os algoritmos devem se aprimorar automaticamente e tal aprimoramento vem através da escolha do treinamento de experiência, o que por sua vez impacta nas tarefas de aprendizado da máquina.

Para realizar essas tarefas, os algoritmos passam por uma fase de treinamento, na qual analisam um conjunto de dados para identificar padrões e relações. Durante esse processo, utilizam métodos estatísticos para ajustar seus parâmetros internos, com o objetivo de minimizar a diferença entre os resultados previstos e os reais.

Logo, os algoritmos de aprendizado de máquina funcionam analisando dados para identificar padrões e, com base nisso, podem generalizar o conhecimento adquirido para fazer previsões sobre novos dados. Dentre as características do aprendizado de máquina está a



predição de eventos no futuro baseado no passado, sendo esta para autores como Daumé et al (2015) o núcleo da maioria das aprendizagens automáticas.

Este aspecto permite que os algoritmos possam ser aplicados em diversos domínios, incluindo saúde, finanças e marketing, fornecendo soluções personalizadas que atendam a necessidades específicas, visto que no aprendizado de máquina, grandes quantidades de dados são ditadas por algoritmos e suas soluções (SIMEONE, 2018).

Finalmente, o aprendizado de máquina pode melhorar a personalização em serviços e produtos, levando a melhores experiências de usuário e maior satisfação do cliente, como no caso de metadados, através da melhoria e otimização de serviços como descrição e representação dos dados, visto que existem vários tipos de algoritmos de aprendizado, cada um projetado de acordo com os objetivos estabelecidos e problemas específicos, dentre eles pode-se citar: as redes neurais, análise de regressão, clustering (agrupamento), deep learning, mineração de dados e árvores de decisão. Para este estudo, alguns serão abordados mais a frente na discussão.

3 METADADOS

Os metadados têm uma ampla gama de aplicações, aqui serão apresentados alguns aspectos e conceitos que os relacionam com os dados de pesquisa e aprendizado de máquina. Uma vez que os aspectos dos metadados, como no caso de sua criação, seja relativa a um documento, arquivo ou outro ativo de informação, registros em formato físico ou objetos digitais, além de itens informacionais modificados e dados excluídos, tornam muito mais fácil para alguém localizar um documento específico ao terem a capacidade de procurar um elemento (ou elementos) específico através de sua descrição ou representação.

Outros aspectos dos metadados apresentam-se na sua utilidade para prolongar a vida útil dos dados existentes, contribuição para a precisão dos dados e ajuda aos usuários na encontrabilidade de informação, além de possibilitarem novas maneiras de acessar a informação. Organizam um objeto de dados usando termos associados a esse objeto em particular. Também permite que objetos diferentes sejam identificados e emparelhados com objetos semelhantes para ajudar a otimizar o uso de ativos de dados.

Quanto as suas ações e aplicações, os metadados podem ocorrer das seguintes maneiras, nas mais diversas estruturas, conforme Alves e Santos (2013):

- a) Exercem a mesma função dos elementos de descrição nos catálogos e nos sistemas de descrição para a busca, localização, acesso e recuperação da informação;
- b) Em relação aos sistemas, sua função é facilitar a interoperabilidade e intercâmbio de dados, facilitando, também, a busca e o gerenciamento de informação;



- c) Em relação aos usuários, indicam a informação acessível, sua localização e as condições para se chegar até ela.

Em relação à linguagem, os metadados relacionam-se com a escrita na intenção de tornar o dado/informação comprehensível tanto para sistemas de computadores quanto para seres humanos, em um nível de padronização que contribui para uma melhor interoperabilidade e integração entre aplicações e sistemas de informação.

Sendo que, para este estudo a pesquisa de Igere (2023) é crucial, pois considera algumas características

dos metadados, devido serem importantes na criação de associação automática de elementos relevantes de objeto digital, isto permite a organização dos dados, facilita a descoberta de dados e promove a gestão de dados que não estão estruturados, além de fornecer formas de identificar várias classes de dados; além do fato dos metadados incluírem uma linguagem legível por máquina, o que permitem a máquina entender informações da web, possibilitando seu uso na descrição do conteúdo de um recurso informacional e com isto favorecem o gerenciamento adequado de dados e informação em processo para criar acessibilidade e recuperação de dados.

4 DADOS DE PESQUISA

Os dados de pesquisa podem ser entendidos como qualquer tipo de informação que pode ser coletada, acessada, armazenada, observada, criada ou gerada em variados campos de pesquisa. Seu principal conceito é que significam dados na forma de fatos, observações, imagens, resultados de programas de computador, gravações, medições ou experiências nas quais um argumento, teoria, teste ou hipótese, ou outra saída de pesquisa é baseada e que possuem como resultado de qualquer investigação sistemática que envolva processos de observação, experimentação ou simulação de procedimentos de pesquisa científica, cuja função é validar e certificar pesquisas e facilitar a interpretação, comunicação e processamento de dados, sendo que podem ser dados físicos ou digitais. (SALES, 2014; SAYÃO; SALES 2020; SEMELER; PINTO, 2019;).

Informa-se que os dados de pesquisa apresentam um gama de finalidades e necessitam de maior escopo para sua conceitualização e definição, sendo a definição de Silva (2019, p. 14) apropriada aqui, pois para o referido autor os:

Dados de pesquisa se convertem em um fator essencial para a cadeia de inovação, a cooperação internacional, a promoção de novas investigações, a formação de novos pesquisadores e, sobretudo, a promoção de atividades científicas mais abertas e transparentes que tenham como princípio a produção do conhecimento à disposição do público.



Esta disponibilidade ao público, juntamente com o acesso para várias atividades científicas, bem como a representação de seu conteúdo necessita de descrição, representação, acesso e recuperação, o que torna viável a aplicação de metadados, logo, esta aplicação estabelece um ponto, quase indissociável entre dados e metadados, visto que para que se possa ter a existência de um é necessário a descrição, acesso, uso e reuso através do outro.

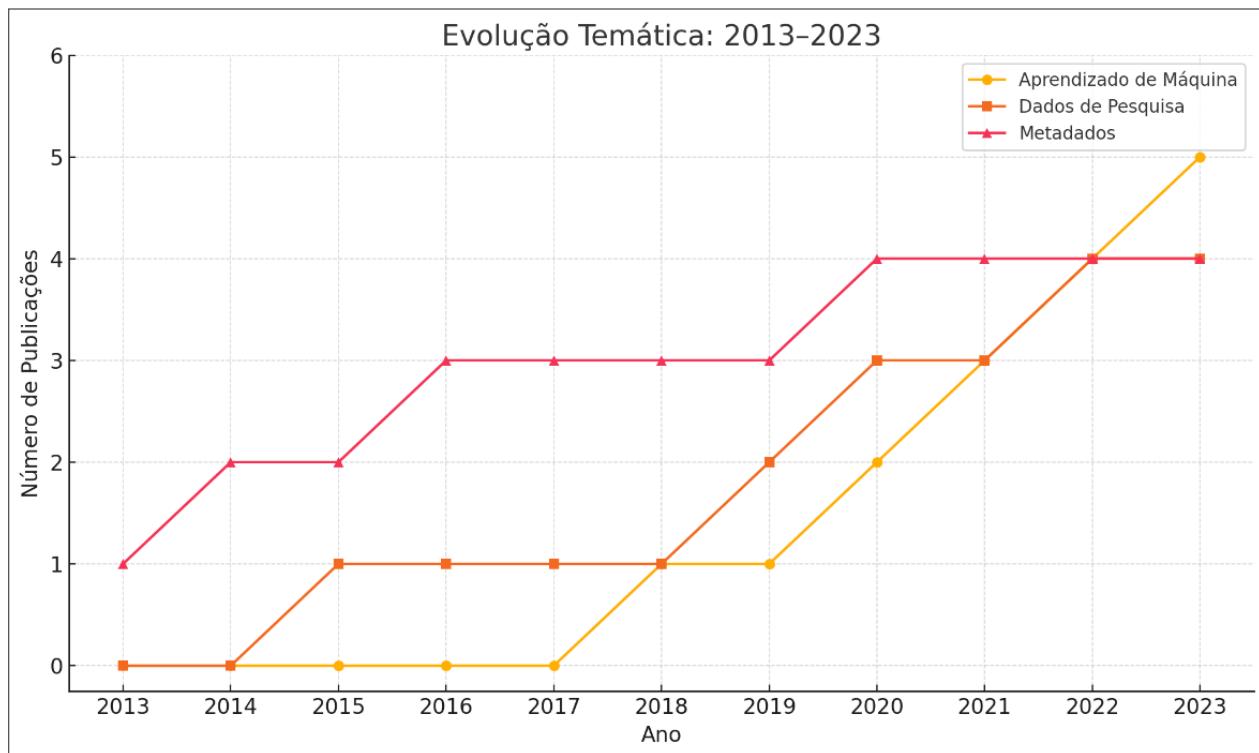
Neste sentido, os metadados são essenciais para o gerenciamento e ciclo de vida dos dados, visto que conforme o ENAP (2019) os metadados auxiliam uma organização a entender os dados de seus sistemas e fluxos de trabalho (com informações sobre processos técnicos e de negócio) e permitem a avaliação da qualidade dos dados (com estruturas de dados lógicas e físicas) estes devem fazer parte do gerenciamento de bancos de dados (regras e restrições de dados), contribuindo, assim, para a capacidade de processar, manter, integrar, proteger, auditar e governar outros dados.

Uma vez que dados brutos de pesquisa necessitam de diferentes tipos metadados, que não se relacionem com diferentes descritores de dados, ou seja, cada tipo de pesquisa irá demandar metadados cuja finalidade seja descrever, administrar, preservar, usar e reutilizar o potencial desses dados em específico e não de maneira geral.

5 RESULTADOS

Os resultados aqui apresentados estão de acordo com o que foi proposto na apresentação dos objetivos desta pesquisa. Com isto foi possível chegar ao gráfico abaixo que exemplifica a evolução da abordagem dos assuntos, dentro do recorte temporal de 2013 a 2023:

Figura 1 - Evolução temática dos assuntos de 2013 a 2023



Fonte: Os autores (2035).

A partir do levantamento nas bases, Elsevier, BRAPCI e Web of Science foi possível perceber a evolução do número de publicações nas temáticas sobre Aprendizado de Máquina: surgindo com força a partir de 2018 e crescendo de forma acentuada até 2023, quanto a Dados de Pesquisa: com crescimento mais constante desde 2014, especialmente após 2019 e para Metadados, percebe-se que são consolidados desde o início do período, mantendo um nível estável de publicações.

E baseado nisso, após análise e reunião de textos que seriam utilizados e abordados na pesquisa, pode-se chegar no quadro abaixo com os autores citados no texto, cujas obras foram de inestimável valor para o enriquecimento da pesquisa. Sendo que a construção da estrutura do quadro segue o período de tempo e a abordagem de assunto e a relação de cada autor com cada assunto em específico:

Quadro 1 - Relação dos autores com assuntos abordados na pesquisa

Ano	Aprendizado de Máquina	Dados de Pesquisa	Metadados
2013	—	—	Alves & Santos
2014	—	Sales	Santos, Simionato & Arakaki
2017	Simeone	—	—



Ano	Aprendizado de Máquina	Dados de Pesquisa	Metadados
2019	—	ENAP	—
2020	Almaghrabi & Chetty	—	—
2021	Chen <i>et al.</i>	—	—
2022	França, Cordel	Chen, Chiu & Cline	IGERE
2023	Mou <i>et al.</i> , Caliskan <i>et al.</i>	Sales (2023)	—

Fonte: Os autores (2025).

Acrescenta-se que, mesmo com a escolha destes textos, optou-se pela utilização de certos livros da área de aprendizado de máquina para maior alcance do escopo teórico do assunto, visto que é um assunto vasto e advindo de outra área de domínio que exige maior interpretação e explicação para enriquecer o contexto desta pesquisa, como Mitchell (1997), Alves (2010) e Simeone (2017).

5.1 DISCUSSÃO

Em projetos de aprendizado de máquina, no campo da Ciência da computação, a relação com metadados apresenta estudos entre as áreas bem estabelecidas. Visto que a documentação dos conjuntos de dados e dos experimentos com metadados é crucial para garantir a reprodutibilidade de dados.

Caliskan et al (2024), corroboram isso ao dizerem que os metadados contém detalhes sobre o pipeline (etapas de processamento de dados), devido documentarem como os dados foram coletados, processados e utilizados para treinar um modelo, isso facilita o entendimento e a replicação do processo de aprendizado.

Quanto ao aspecto da importância de estudos sobre aprendizado de máquina no escopo da Ciência da informação, uma das proposições é a que neste tipo de ambiente, a implementação de algoritmos de aprendizado de máquina ajudaria na descoberta de padrões dentro dos dados e metadados, e com isso auxiliaria na escolha do tipo e padrão de metadados para a descrição, administração e representação dos dados de modo otimizado.

Outro aspecto dos algoritmos de aprendizado aplicados em metadados no escopo da Ciência da Informação, pode ser na sua ação para possibilitar a descoberta fácil dos dados (discoverability). O trabalho de CHIU, CHEN, CLINE (2022) propõe que a descoberta fácil dos dados é particularmente importante em áreas científicas (dados de pesquisa) e industriais, onde modelos precisam ser testados e validados em diferentes contextos. Esta descoberta fácil se assemelha ao que Sayão (2010, p.3) nomeou reprodutibilidade, que é “quando submergimos no mundo dos documentos digitais, constatamos outras dimensões dos metadados, que ultra-



passam os limites de ferramenta para a descrição e descoberta de recursos que precisam ser reveladas e exploradas". Devido ao fato dos objetos digitais precisarem de gerenciamento e para serem usados necessitam de processos de maior amplitude, o que implica em identificar informações precisas através dos metadados.

Acrescenta-se o estudo de Allalouf, Mendelsson e Mishustin (2014) sobre o funcionamento do Visfacet, que utiliza visualização facetada interativa. Seu funcionamento ocorre quando os metadados são inseridos como índices na base de dados interna VuFind, juntamente com os seus metadados associados. Ao inserir um termo para pesquisar, o motor de busca do catálogo recupera os registros associados a partir do motor de busca, isso ocorre devido a interface do catálogo permitir aos usuários restringir ou ampliar seus resultados de busca filtrados por tópicos ou facetas. Ao mesmo tempo, ele recupera uma lista de facetas pré-definidas (às vezes também chamadas de clusters), cada um dos quais representa uma conjunção lógica do conjunto atual dos resultados filtrados por vários campos pré-definidos, cita-se este trabalho por apresentar a visualização da informação como componente integrante de mineração de dados, um dos variados tipos de algoritmos de aprendizado de máquina.

Ainda em relação a esta temática do Visfacet, o estudo de França (2022), no cenário nacional, investiga a relação entre deep learning e organização da informação, demonstrando como algoritmos podem melhorar a eficiência dos sistemas de classificação e consequentemente organizar os dados e metadados.

Destaca-se questões envolvidas com a localização, interoperabilidade, acesso e reuso de dados, ou seja os Princípios FAIR (Achável, Acessível, Interoperável e Reutilizável), tais princípios se envolvem com metadados e aprendizado de máquina e dados, tanto para questões envolvendo gestão de dados quanto para soluções envolvendo privacidade.

Argumenta-se isso devido a multiplicidade de funções dos metadados necessários à gestão FAIR dos objetos digitais de pesquisa, com princípios necessários para subsidiar ações que os tornam encontráveis, acessíveis, interoperáveis, de forma que eles possam ser compreendidos ao longo do tempo, com a sua "reutilização em diversos contextos diferentes do projeto original", o que resulta em um volume considerável e diversificado de metadados adicionado para cada objeto digital (Sayão e Sales, 2024). Desta maneira, os elementos de metadados irão adicionar valor aos dados durante seu ciclo de vida através de agentes, incluindo softwares e instrumentos científicos.

Entretanto, Sales e Sayão (2024) dizem que muitas vezes esses elementos estão desordenados, são ambíguos, complexos e desestruturados, e precisam de um grau de esquematização para apoiar a construção de esquemas de metadados disciplinares mais representativos.

Neste ponto, cita-se o trabalho de Mou et al (2023), onde os autores exploram as possíveis soluções de uma arquitetura de sistema que incorpore os princípios de dados FAIR, através de uma colaboração eficaz e segura ao utilizar um conjunto mínimo de esquemas de



metadados adaptados ao aprendizado de máquina, utilizando para isto um mecanismo descentralizado de autenticação e autorização baseado em identidade autônoma, com a utilização de identificadores descentralizados (DIDs), criptograficamente verificáveis, globalmente únicos e resolvíveis, como identificadores persistentes (PIDs) para garantir a capacidade de localização de objetos digitais e o desenvolvimento de um conjunto mínimo de esquemas de metadados adaptados para aprendizado de máquina distribuído, o que melhora a descoberta e interoperabilidade dos dados.

6 CONSIDERAÇÕES FINAIS

Esta pesquisa buscou explorar a relação entre aprendizado de máquina e metadados na Ciência da Informação, quando aplicados em dados de pesquisa. Pelo fato de que os metadados, ao estruturar e descrever dados, são elementos centrais no processo de aprendizado de máquina. Além disso, a análise dos textos utilizados nesta pesquisa demonstra como a integração dessas tecnologias pode melhorar a recuperação de informações, identificar padrões e otimizar serviços de informação.

Muito em virtude dos aspectos dos metadados, visto que dentre suas funcionalidades apresentam a utilidade para prolongar a vida útil dos dados existentes, atuarem na precisão dos dados e possibilitem novas maneiras de acessar a informação, isto em relação a dados de Pesquisa e Princípios FAIR é essencial. Quanto ao organização de um objeto de dados usando termos associados a esse objeto em particular, isto se aproxima do que Cordell (2020) que discute a utilização de aprendizado de máquina para enriquecer coleções, abordando aspectos técnicos e éticos dos algoritmos, destacando questões como a identificação de conjuntos de dados e a análise de padrões desses dados, com o uso de algoritmos de clustering para otimizar a pesquisa e a classificação de informações. Isto vai muito próximo do que os metadados permitem ao identificarem diferentes objetos e emparelhados com objetos semelhantes para ajudar a otimizar o uso de dados (Sendo esta uma das tarefas realizadas pelos algoritmos de clustering).

Desta forma, este estudo destaca a necessidade de novas pesquisas sobre os impactos do aprendizado de máquina nos serviços de informação brasileiros. Onde as aplicações promissoras como a automatização da geração de metadados, a identificação de informações ocultas e a redução de vieses nos algoritmos seriam essenciais para melhoria de serviços tanto para acesso e recuperação de dados. Portanto, tais avanços podem resultar em ferramentas mais eficazes e inclusivas, alinhadas às demandas contemporâneas da Ciência da Informação.

REFERENCIAS

Allalouf, A., Mendelsson, S., & Mishustin, A. (2014, September). *VisFacet: Facet visualization module for modern library catalogues* [Conference paper]. International Conference on Theory



and Practice of Digital Libraries (TPDL), London, United Kingdom. In *Lecture Notes in Computer Science* (Vol. 8674, pp. 82–93). Springer. https://doi.org/10.1007/978-3-319-08425-1_8

Almaghrabi, M., & Chetty, G. (2020, December). *Deep machine learning digital library recommendation system based on metadata for Arabic and English languages* [Conference paper]. 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE) (pp. 1–6). IEEE. <https://doi.org/10.1109/CSDE50874.2020.9411525>

Alves, R. C. V. (2010). *Metadados como elementos do processo de catalogação* [Doctoral dissertation, Universidade Estadual Paulista “Júlio Mesquita Filho”]. Repositório Institucional UNESP. <http://repositorio.unesp.br/handle/11449/103361>

Alves, R. C. V., & Santos, P. L. V. A. C. (2013). *Metadados no domínio bibliográfico*. Intertexto.

Caliskan, A., Mou, C., Saha, S., Nair, V., Srivastava, A., & Zhao, J. (2023). Metadata integrity in bioinformatics: Bridging the gap between data and knowledge. *Computational and Structural Biotechnology Journal*, 21, 4895–4913. <https://doi.org/10.1016/j.csbj.2023.08.020>

Chen, C., Ramos, G., Terry, M., & Yao, Y. (2021, July 27–29). *Toward a unified metadata schema for ecological momentary assessment with voice-first virtual assistants* [Conference paper]. ACM International Conference on Conversational User Interfaces (CUI '21), Bilbao, Spain. <https://doi.org/10.1145/3469595.3469612>

Chen, H. L. (O.), Chiu, T. H., & Cline, E. (2022). Academic libraries and research data management: A case study of Dataverse global adoption. *PCOM Scholarly Papers*. https://digitalcommons.pcom.edu/scholarly_papers/3301

Cordell, R. (2020). *Machine learning + libraries: A report on the state of the field*. Library of Congress Labs. <https://labs.loc.gov/static/labs/work/reports/Cordell-LOC-ML-report.pdf>

Daumé III, H., Kumar, A., & Sontag, D. (2015, July). *Deep unordered composition rivals syntactic methods for text classification* [Conference paper]. International Joint Conference on Natural Language Processing (IJCNLP), Beijing, China. <https://aclanthology.org/P15-1162>

Escola Nacional de Administração Pública. (2019). *Gerenciamento de metadados e da qualidade de dados*. Repositório ENAP. <https://repositorio.enap.gov.br/handle/1/5008>

Fran  a, I. O. (2022). *Deep learning: A rela  o entre intelig  ncia artificial e organiz  o da informa  o* [Undergraduate thesis, Universidade Federal do Rio de Janeiro]. Reposit  rio Pantheon. <https://pantheon.ufrj.br/handle/11422/16903>

Gil, A. C. (2019). *Métodos e técnicas de pesquisa social* (7^a ed.). Atlas.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. <https://www.deeplearningbook.org/>



Gray, J. (2007). *Jim Gray on eScience: A transformed scientific method*. In T. Hey, S. Tansley, & K. Tolle (Eds.), *The fourth paradigm: Data-intensive scientific discovery* (pp. xvii–xxxi). Microsoft Research.

Hackeling, G. (2014). *Mastering machine learning with scikit-learn*. Packt Publishing.

Igere, M. A. (2023). Metadata for effective research data management in university libraries. *Library Philosophy and Practice*, 2023(1), 1–12. <https://digitalcommons.unl.edu/libphilprac/7654>

Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill.

Mou, C., Caliskan, A., Saha, S., Srivastava, A., & Zhao, J. (2023). FAIR data principles for machine learning in bioinformatics. *Patterns*, 4(9), 100762. <https://doi.org/10.1016/j.pat.2023.100762>

Pericho, J. S., & Bianchi, C. (2011). *Aprendizado de máquina: Conceitos e algoritmos*. LTC.

Sales, L. F. (2014). *Dados de pesquisa: Perspectivas para a ciência e a comunicação científica*. Editora UFMG.

Sales, L. F., & Sayão, L. F. (2020). *Gestão e curadoria de dados de pesquisa*. Editora Fiocruz. <https://doi.org/10.7476/9786557080602>

Sales, L. F., & Sayão, L. F. (2023). Reproducibility and metadata challenges in open science. *Perspectivas em Ciência da Informação*, 28(2), 25–43. <https://doi.org/10.1590/1981-5344-2023v28n2a003>

Sales, L. F., & Sayão, L. F. (2024). Metadata schemas and FAIR principles: Challenges for information science. *Revista Ibero-Americana de Ciência da Informação*, 17(1), 45–61. <https://doi.org/10.26512/rici.v17i1.2024>

Sayão, L. F. (2010). Reproduibilidade e metadados na ciência contemporânea. *Ciência da Informação*, 39(1), 3–12. <https://doi.org/10.1590/S0100-19652010000100001>

Semeler, A. R., & Pinto, R. A. (2019). Dados de pesquisa e metadados na comunicação científica. *Revista Brasileira de Biblioteconomia e Documentação*, 15(1), 56–72. <https://rbbd.febab.org.br/rbbd/article/view/1221>

Simeone, A. (2018). *Fundamentals of machine learning for predictive data analytics*. MIT Press.

ANEXO 1

RESUMEN BIOGRÁFICO DEL AUTOR

Edvaldo Wellington da Cunha Monteiro

Possui graduação em Biblioteconomia pela Universidade Federal do Pará (2010). Pós Graduado em Gestão Tecnológica da Informação pela Escola Superior da Amazônia (ESAMAZ) (2011), atualmente mestrando em Ciência da Informação pela UFPA, no Programa de Pós-Graduação em Ciência da Informação PPGCI-2023. Tem experiência na área de Biblioteconomia e boa prática em Arquivos e Ciência da Informação. Possui interesse em assuntos correlatos com: Ciência da Informação, Web Semântica, Metadados, Padrões de metadados, Classificação, Catalogação, Indexação, Organização da Informação, Mediação e Uso da Informação, Ambientes Informacionais Digitais, Repositórios Institucionais, Ciência Aberta, Bases de Dados, Ontologias, Arquitetura da Informação, Recuperação da Informação, Encontrabilidade da Informação, Taxonomias, Inteligência Artificial, Sistemas de Informação, Informação Eletrônica, Sistemas de Organização do Conhecimento (SOC), Tecnologia da Informação e Comunicação (TICs), Mineração de Textos, Big Data, Dados de pesquisa e Gestão Eletrônica de Documentos.

ANEXO 2

REQUERIMIENTOS DE EQUIPO TÉCNICO PARA LA PRESENTACIÓN DE LA PONENCIA

Necessitarei dos seguintes equipamentos para a minha apresentação: Computador, projetor, microfones, conexão com a Internet, tradução simultânea, mesas, etc.