Technol. innov. j. (2026). Vol. 5 Núm. 2 págs. 7-24

TECHNOLOGICAL INNOVATIONS JOURNAL
ISSN: 2955-8239
https://tijournal.science
Editada por: Instituto Universitario de Innovación Ciencia y Tecnología Inudi Perú

ARTIGO ORIGINAL

Transduções informacionais, Machine Learning e a inciência dos atores envolvidos

Informational transductions, Machine Learning, and the unawareness of the actors involved

Transducciones informacionales, Machine Learning y la inciencia de los actores involucrados

Fábio Moreira[1]
Universidade Estadual Paulista, Marília – São Paulo, Brasil
https://orcid.org/0000-0002-9582-4218
fabio.moreira@unesp.br (correspondencia)

Marcos Siqueira
Universidade Estadual Paulista, Marília – São Paulo, Brasil
https://orcid.org/0009-0002-8412-8820
marcos.siqueira@unesp.br

Rachel Alves
Universidade Estadual Paulista, Marília – São Paulo, Brasil
https://orcid.org/0000-0002-1649-3722
rachel.vesu@unesp.br

Ricardo Sant'Ana
Universidade Estadual Paulista, Marília – São Paulo, Brasil
https://orcid.org/0000-0003-1387-4519
ricardo.santana@unesp.br

DOI: https://doi.org/10.35622/j.ti.2026.02.001

Recebido: 03/01/2026 Aceito: 26/03/2026 Publicado: 14/04/2026

PALAVRAS-CHAVE

aprendizado de máquina, inconsciência, inteligência artificial, transdução informacional.

RESUMO. O estudo realizado tem como objetivo discorrer sobre os aspectos da inconsciência dos atores envolvidos no desenvolvimento e uso de modelos de Machine Learning, com foco nas transduções informacionais que ocorrem ao longo de suas etapas, isto é, nas transformações que a informação experimenta ao longo das diferentes fases do ciclo de vida dos dados e do modelo, desde a coleta até a aplicação. O procedimento metodológico adotado baseia-se em uma pesquisa descritiva, exploratória e qualitativa, e coletou dados nas plataformas Hugging Face e GitHub, com a finalidade de analisar as possíveis transformações e transduções informacionais que ocorrem na fase de treinamento dos modelos disponíveis. Os resultados indicam que, embora as plataformas ofereçam um grande volume de modelos e ferramentas, existe opacidade em relação às transformações aplicadas aos dados originais, o que limita a compreensão dos usuários sobre a integridade e a confiabilidade dos modelos disponíveis. Conclui-se, portanto, que há necessidade de que as plataformas adotem protocolos de transparência e estruturas padronizadas de descrição para representar as transduções e as decisões técnicas envolvidas.

[1] Doutor em Ciência da Informação pela Universidade Estadual Paulista, Brasil.

Esta obra está bajo una licencia internacional Creative Commons Atribución 4.0.

KEYWORDS

artificial intelligence, informational transduction, machine learning, unawareness.

ABSTRACT. The study aims to discuss aspects of the unawareness of the actors involved in the development and use of Machine Learning models, focusing on the informational transductions that occur throughout their stages, that is, the transformations that information undergoes across the different phases of the data and model life cycle, from collection to application. The methodological procedure adopted is based on descriptive, exploratory, and qualitative research, and data were collected from the Hugging Face and GitHub platforms in order to analyze the possible transformations and informational transductions that occur during the training phase of the available models. The results indicate that, although these platforms offer a large volume of models and tools, there is opacity regarding the transformations applied to the original data, which limits users' understanding of the integrity and reliability of the available models. It is therefore concluded that there is a need for platforms to adopt transparency protocols and standardized descriptive structures to represent the transductions and technical decisions involved.

PALAVRAS-CHAVE

aprendizaje automático, insciencia, inteligencia artificial, transducción informacional.

RESUMEN. El estudio realizado tiene como objetivo discurrir sobre los aspectos de la inciencia de los actores involucrados en el desarrollo y uso de modelos de Machine Learning, con foco en las transducciones informacionales que ocurren a lo largo de sus etapas, es decir, las transformaciones que la información experimenta a lo largo de las diferentes fases del ciclo de vida de los datos y del modelo, desde la recolección hasta la aplicación. El procedimiento metodológico adoptado se basa en una investigación descriptiva, exploratoria y cualitativa, y recolectó datos en los portales Hugging Face y GitHub, con el fin de analizar las posibles transformaciones y transducciones informacionales que ocurren en la fase de entrenamiento de los modelos disponibles. Los resultados indican que, aunque los portales ofrecen un gran volumen de modelos y herramientas, existe opacidad respecto a las transformaciones aplicadas a los datos originales, lo que limita la comprensión de los usuarios sobre la integridad y confiabilidad de los modelos disponibles. Se concluye, por lo tanto, que existe la necesidad de que las plataformas adopten protocolos de transparencia y estructuras de descripción estandarizadas para representar las transducciones y las decisiones técnicas involucradas.

1. INTRODUÇÃO

A sociedade contemporân<ea tem testemunhado um avanço significativo no desenvolvimento e na aplicação da Inteligência Artificial (IA). Esse fenômeno impacta diretamente diversos setores da sociedade, como, por exemplo: na área da saúde já é possível observar a utilização de sistemas de IA para a realização de diagnósticos médicos, identificando padrões em exames e auxiliando na detecção precoce de doenças (Ghaffar Nia et al., 2023); no setor financeiro, modelos de Machine Learning (ML) são empregados para detectar fraudes e otimizar investimentos (Ali et al., 2022); enquanto na educação, algoritmos personalizados adaptam conteúdos didáticos às necessidades dos alunos (De Souza Souza et al., 2024).

Não obstante a esses avanços, pesquisadores têm levantado questionamentos sobre a confiabilidade dos modelos de ML (Aldieri et al., 2025), uma subárea da IA que baseia-se no desenvolvimento de modelos de identificação de padrões matemáticos em grandes volumes de dados para realizar previsões e classificações. Esses modelos utilizam conceitos de estatística, álgebra linear, cálculo e teoria da probabilidade para ajustar os parâmetros a partir dos dados de entrada e fornecer uma saída que corresponda a determinado comportamento ou padrão (Géron, 2017).

Diante desse contexto, este artigo investiga os fatores que influenciam a confiança em modelos de ML, especialmente no que se refere à documentação e à transparência dos processos de desenvolvimento e treinamento desses modelos. Atualmente, a credibilidade de um modelo de ML muitas vezes está associada à reputação da empresa desenvolvedora, o que levanta reflexões sobre a aceitação de modelos desenvolvidos por

Esta obra está bajo una licencia internacional Creative Commons Atribución 4.0.

instituições menos conhecidas. Sendo assim, o problema central deste estudo é, portanto, compreender quais aspectos podem aumentar a confiança nos modelos de ML de forma a conectar a reputação das empresas de tecnologia da informação com a necessidade de transparência documental.

Para responder a essa questão, o artigo tem como objetivo analisar o processo de desenvolvimento e treinamento de modelos de ML sob a perspectiva da transdução informacional (Sant'Ana, 2021), explorando suas etapas fundamentais a partir da documentação detalhada desses processos. Desta forma, pretende-se contribuir para o debate sobre a confiança em modelos de ML, enfatizando que a credibilidade desses sistemas não deve depender exclusivamente da reputação das empresas desenvolvedoras, mas sim de práticas transparentes, documentadas e acessíveis a pesquisadores, reguladores e ao público em geral.

Considerando que a credibilidade dos sistemas de IA deve derivar de práticas transparentes e acessíveis, e não apenas da autoridade institucional do desenvolvedor, a construção da confiança nos modelos de ML – e consequentemente nos agentes – passa por um compromisso coletivo com a explicabilidade, a governança e a ética no desenvolvimento e na aplicação dessas tecnologias. Portanto, a pesquisa se concentrou em refletir especificamente sobre possíveis insciências dos atores envolvidos no desenvolvimento e uso de modelos de ML, principalmente sobre as transduções informacionais (Sant'Ana, 2021) que ocorrem durante as etapas de desenvolvimento dos modelos.

2. METODOLOGIA

A metodologia adotada para tal investigação baseia-se em uma pesquisa descritiva, de caráter exploratório e de abordagem qualitativa. O procedimento metodológico partiu de consultas na literatura sobre 'Machine Learning e desenvolvimento de modelos', somados à análise do conceito de transdução informacional que, Segundo Sant'Ana (2021), refere-se ao impacto do controle sobre os dados na geração de conhecimento, evidenciando como a estruturação, a modelagem e a curadoria dos dados influenciam os resultados produzidos por sistemas de IA. Os procedimentos de busca e registro dos dados para a formulação deste estudo consistiram no acesso aos websites Hugging Face <https://huggingface.co/> e GitHub <https://github.com/onnx/models>, que disponibilizam para download modelos de ML desenvolvidos pela comunidade de programadores e demais entusiastas no tema.

Os critérios para a avaliação desses sites focaram na identificação das informações sobre os modelos disponíveis em cada website para analisar se eram suficientes para compreender as possíveis transformações e transduções informacionais que ocorreram nos dados originalmente coletados para a fase de treinamento dos modelos de ML. Neste ponto, ressalta-se que para um modelo de ML ser treinado e, consequentemente, aprender o que se deseja, um conjunto de dados deve ser coletado, organizado e apresentado. Também é necessário esclarecer que nessa fase de organização dos dados podem acontecer transformações nos dados originais como, por exemplo, ocasionada pela aplicação da técnica denominada normalização de dados. Sendo assim, a análise dos websites apoiou-se na observação de quais tipos de transformações os dados originais coletados e usados para a fase de treinamento do modelo de ML foram submetidos. A análise dos resultados inclui a identificação dos principais desafios e limitações na construção da confiabilidade em modelos de ML, bem como a proposição de diretrizes para aumentar a transparência e a aceitabilidade desses modelos no contexto social e científico. Nas seções seguintes, fundamentam-se os conceitos que apoiarão esta análise: transdução informacional e desenvolvimento de modelos de Machine Learning.

Esta obra está bajo una licencia internacional Creative Commons Atribución 4.0.

3. RESULTADOS E DISCUSSÃO

Transdução informacional e insciência do usuário

A necessidade de identificar com clareza os tipos de transformações que os dados originais coletados e usados para o treinamento de um modelo de ML conduziu esta pesquisa para o conceito de transduções informacionais (Sant'Ana, 2021), o que permitiu refletir sobre possíveis insciências dos atores envolvidos no processo de desenvolvimento e no uso de modelos de ML.

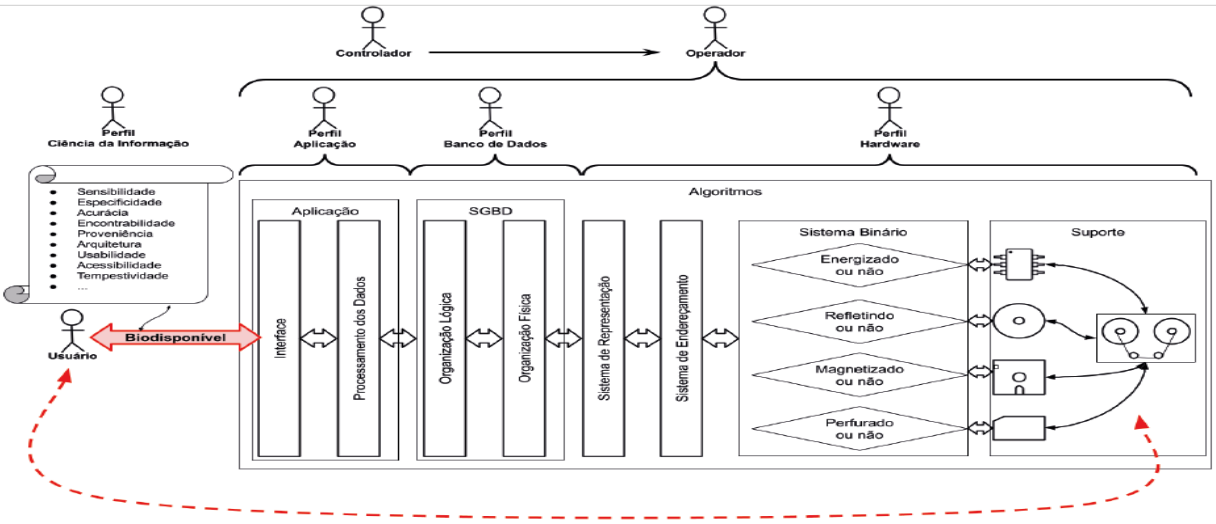
O conceito de transdução informacional, para Sant'Ana (2021), procura dar conta das transformações múltiplas e muitas vezes invisíveis ou "opacas" pelas quais a informação passa quando manipulada em meios digitais. Essas transformações ocorrem em níveis que os usuários típicos (e mesmo atores do ecossistema informacional) não percebem ou não têm controle. Por exemplo, o que é exibido em uma interface de usuário pode ser uma representação "suavizada" dos dados, enquanto nos bastidores houve compressões, indexações, junções, anonimizações, recálculos, geração de metadados e filtros algorítmicos. Essas camadas estruturais tornam "opacas" (isto é, pouco visíveis ou transparentes) as transformações efetivamente feitas sobre os conteúdos.

O termo "transdução" já é usado em outras áreas (biologia, engenharia) para indicar a conversão de um tipo de sinal ou energia. Sant'Ana (2021) adota essa ideia para o domínio da informação digital: ele entende que os conteúdos informacionais (dados, documentos, registros) não fluem inalterados nos sistemas digitais, mas são submetidos a sucessivas camadas de abstração, tradução, reinterpretação, filtragem e codificação/decodificação.

A transdução informacional, em síntese, trata-se de um processo de transformação de dados dirigido para algo que possa ser compreendido e visualizado por seres humanos. A Figura 1 ilustra este conceito destacando uma dinâmica na qual o usuário realiza uma entrada com dados biodisponíveis e o processo de transformações desses dados ocorre com a mobilização de inúmeros atores e profissionais, principalmente daqueles ligados às áreas da Ciência da Computação (desenvolvimento do aparato tecnológico) e da Ciência da Informação (representação e tratamento da informação).

Figura 1

Processo de transformação dos dados



Fonte. Sant'Ana (2021).

Esta obra está bajo una licencia internacional Creative Commons Atribución 4.0.

No bojo das transduções informacionais, a relação entre a insciência dos atores humanos e a carência de informações revela uma complexidade inerente aos processos de manipulação e transformação de dados na era digital. Essa insciência dos atores, isto é, o desconhecimento sobre os processos internos que ocorrem durante a manipulação dos dados, limita a compreensão completa do impacto dessas operações sobre a informação original (Sant'Ana, 2021).

Portanto, o conceito de transdução informacional pode incrementar um estado de consciência (ou reduzir a "insciência") dos diversos atores envolvidos nos fluxos informacionais (ciclos de coleta, armazenamento e recuperação), até quem consome a informação (usuário e/ou elemento alvo do fluxo). Reconhecer que está ocorrendo transdução informacional ajuda a perceber os "gaps" e os "pontos de controle" que antes não eram invisíveis, assim, a subseção seguinte põe em voga a definição do conceito de insiência, que é parte determinante para compreensão dos processos de transdução informacional.

A insciência dos atores sobre as transduções informacionais também está diretamente relacionada à ausência de transparência nos processos tecnológicos. À medida que os dados são transformados, eles passam por diferentes estados de codificação, alterando-se em forma, estrutura e energia. Cada uma dessas transformações pode introduzir alterações ou perda de informações que são imperceptíveis para o usuário final (ex: quando uma fotografia é digitalizada, ela passa por diversas etapas de transformação que convertem a imagem física em uma sequência binária de zeros e uns). Essas transformações, por mais imperceptíveis que sejam ao olhar humano, envolvem processos complexos que não são de conhecimento geral, o que gera uma lacuna significativa entre o uso cotidiano das tecnologias e a compreensão real de como elas operam (Sant'Ana, 2021).

Além disso, a falta de informações sobre as transduções informacionais também limita a capacidade dos indivíduos de questionar, avaliar e entender os sistemas tecnológicos que utilizam. De acordo com Sant'ana (2021), o uso de algoritmos e sistemas de processamento automatizados facilita a manipulação de grandes volumes de dados, mas também oculta as operações que ocorrem nos bastidores, aumentando a insciência dos usuários sobre o modo como suas informações são tratadas. Assim, a dependência crescente de tecnologias de processamento de dados reforça essa lacuna de conhecimento, ao mesmo tempo que aumenta o impacto das decisões automatizadas e processos computacionais sobre a vida cotidiana.

A insciência, portanto, não se limita apenas à falta de conhecimento técnico, mas também à ausência de informações acessíveis e compreensíveis para o público em geral. A complexidade técnica dos sistemas que realizam transduções informacionais muitas vezes torna impraticável para os não especialistas entenderem as nuances desses processos. Destacado por Sant'ana (2021), essa falta de compreensão pode levar a uma aceitação passiva de como os dados são gerenciados e manipulados, comprometendo a capacidade dos atores de exercer um controle consciente sobre as informações que compartilham ou consomem.

Em última análise, compreender a relação entre a insciência dos atores e a falta de informações sobre as transduções informacionais é essencial para promover um uso mais crítico e consciente das tecnologias digitais. O aumento da transparência e a educação sobre como os dados são transformados e utilizados podem ajudar a mitigar os riscos associados ao desconhecimento desses processos. Dessa forma, é possível fortalecer a capacidade dos indivíduos de tomar decisões informadas e de se engajar de maneira mais ativa na governança de seus próprios dados, promovendo um equilíbrio mais saudável entre a dependência tecnológica e a autonomia informacional (Sant'Ana, 2021).

Esta obra está bajo una licencia internacional Creative Commons Atribución 4.0.

No âmbito da fundamentação do conceito de insiência é importante destacar suas implicações epistemológicas. Por exemplo, apesar de não mencionar explicitamente a insciência, Aleixo (2020) tangencia o conceito ao discutir a falta de padronização e clareza na legislação e documentos que orientam a abertura dos dados. A falta de estrutura e a consequente dificuldade na compreensão dos processos internos de manipulação dos dados governamentais é explorada pela autora. Sem uma definição clara e precisa dos termos, sem a garantia de que os dados disponibilizados são realmente abertos e sem um entendimento dos processos que permeiam sua coleta e tratamento, o usuário fica à mercê de informações incompletas e, em alguns casos, até incorretas, sem ter a capacidade de avaliar completamente o impacto das operações realizadas sobre os dados originais.

De acordo com De Souza de Aguiar Monteiro e Gonçalves Sant'Ana (2022), os conceitos de anomia e anomia informacional se referem a ausência, falta ou desconhecimento de normas, regras, diretrizes ou valores que deveriam orientar o comportamento ou as ações em determinado contexto. No caso da anomia, a falta de normas claras e coercitivas na sociedade leva ao desconhecimento, por parte dos indivíduos, sobre as regras do chamado jogo social. Já a anomia informacional, refere-se à ausência de regras e padrões nos sistemas de informação, resultando em descompasso entre as necessidades dos usuários e a forma como a informação é disponibilizada. Portanto, o texto aborda diretamente a ignorância, entendida como a falta de conhecimento sobre regras, e a falta de conhecimento como elementos centrais na compreensão dos conceitos de anomia e anomia informacional.

A distinção entre insciência, ignorância, falta de conhecimento e anomia informacional é fundamental para compreender os limites epistemológicos da transparência nos modelos de ML. Embora os quatro conceitos tratem de lacunas cognitivas ou informacionais, apenas a insciência está diretamente ligada à opacidade estrutural das transformações que ocorrem no interior dos sistemas — especialmente durante as transduções informacionais. O Quadro 1 sistematiza essas diferenças e indica como cada uma delas impacta o desenvolvimento e a confiabilidade dos modelos de ML sob a perspectiva da Ciência da Informação.

Quadro 1

Definindo 'Insiência informacional', 'Ignorância', 'Falta de conhecimento' e 'Anomia informacional'

Conceito	Característica-chave (delimitação conceitual)	Implicações para Machine Learning (ML)	Implicações para Ciência da Informação (CI)
Insciência informacional	Desconhecimento sobre os processos internos de manipulação e transformação dos dados. Não é ignorância geral, mas ausência de consciência crítica sobre 'como' os dados foram transformados durante as etapas do ML.	Usuários (inclusive técnicos) podem utilizar modelos sem saber quais transformações os dados sofreram — especialmente nas fases de pré-processamento e normalização. Risco de aderir ao modelo sem avaliar sua integridade.	Indica necessidade de registrar e documentar as etapas do ciclo de vida dos dados, especialmente as transduções. Aponta para a importância da rastreabilidade informacional como critério de confiabilidade.
Ignorância	Falta de conhecimento geral, sem relação direta com processos técnicos específicos. Pode ser	Afeta a compreensão conceitual dos modelos — mas é superável com	Relaciona-se à educação informacional e formação em data literacy. É uma lacuna

Esta obra está bajo una licencia internacional Creative Commons Atribución 4.0.

	resolvida por treinamento e formação.	capacitação técnica. Não implica opacidade estrutural.	cognitiva superável por meio de ensino sistemático.
Falta de conhecimento	Ausência de domínio sobre determinado conteúdo, mas sem obstrução estrutural. É contingencial, pontual, e pode ser sanada com acesso à informação.	Pode dificultar o uso adequado de modelos, mas não impede, necessariamente, a compreensão do processo.	Relaciona-se ao déficit de acesso à informação — que pode ser sanado por documentação e práticas de acessibilidade da informação.
Anomia informacional	Ausência de normas, padrões ou diretrizes claras que orientem a organização e disponibilização da informação.	As plataformas de ML não seguem padrão para documentar transformações dos dados: cada modelo adota critérios distintos, tornando difícil a análise de confiabilidade.	A CI é chamada a propor padrões de estruturação e interoperabilidade da informação. Reforça a demanda por governança e padronização informacional.

Fonte. Elaboração própria.

Com a fundamentação do conceito de transduções informacionais e seus conceitos adjacentes é possível refletir sobre as transformações de dados que ocorrem nas etapas do processo de desenvolvimento de um modelo de ML, que por sua vez são tratadas na seção seguinte deste artigo.

Machine Learning

Machine Learning (ML) é uma subárea da Inteligência Artificial (IA) que permite que as máquinas identifiquem padrões matemáticos nos dados disponíveis, aprendendo com esses padrões para fazer previsões ou tomar decisões sem a necessidade de programação explícita. Através de algoritmos especializados, os sistemas de ML podem melhorar seu desempenho à medida que são expostos a mais dados, tornando-se mais precisos ao longo do tempo (Géron, 2017).

Esses recursos possibilitam a aplicação de ML em uma variedade de áreas, como, por exemplo, no reconhecimento de áudio e voz (Georgescu et al. 2021), visão computacional para análise de imagens e vídeos (Khan & Al-Habsi, 2020), análise preditiva e automação de tomada de decisão (Islam et al., 2024) e em sistemas de recomendação (Roy & Dutta, 2022). Tais funcionalidades permitem que os agentes de IA compreendam e interpretem dados complexos do ambiente, viabilizando uma percepção semelhante à humana, ampliando significativamente as capacidades de percepção, decisão e interação de agentes de IA.

A possibilidades geradas com o desenvolvimento dos modelos de ML sustentam não apenas a automação de tarefas, mas também a autonomia operacional, uma vez que os agentes podem aprender continuamente a partir de dados e interações. Complementarmente, arquiteturas avançadas, como redes neurais profundas e modelos multimodais, possibilitam a generalização de padrões, o raciocínio em diferentes domínios e a personalização de respostas em interações com humanos, consolidando os agentes de IA como sistemas cognitivos capazes de agir de forma proativa e adaptativa (Ludermir, 2021).

Munir et al., (2019) destacam a capacidade dos modelos de ML e suas possibilidades utilizando como exemplo o uso de agentes de IA aplicados no diagnóstico de um paciente para determinar a probabilidade de existir um

Esta obra está bajo una licencia internacional Creative Commons Atribución 4.0.

câncer, sem a intervenção de um médico. Esse uso já é uma realidade e comprovadamente, para alguns tipos de cânceres, os modelos de ML têm potencialmente fornecido melhores resultados em comparação aos médicos.

Essa realidade levanta uma questão social significativa baseada na identificação de quais são os fatores que podem contribuir para aumentar a confiança em modelos ML. Atualmente, a informação de que um determinado software ou modelo de ML foi desenvolvido por uma grande empresa multinacional de tecnologia da informação naturalmente conduz a algum grau de credibilidade. No entanto, caso o modelo de ML seja desenvolvido por uma empresa menos conhecida, a reputação não se mostra suficiente para garantir a confiança necessária.

Neste sentido, é importante compreender o processo de desenvolvimento de modelos de ML, que pode ser organizado nas seguintes etapas (De Mattos Paixão et al., 2022): pré-processamento, treinamento e avaliação do modelo.

Na primeira etapa realiza-se o tratamento dos dados brutos para torná-los adequados ao uso em algoritmos de ML. Isso envolve atividades como limpeza de dados (remoção de valores ausentes ou inconsistentes), normalização ou padronização de escalas, codificação de variáveis categóricas, seleção ou extração de características relevantes (engenharia de atributos) e divisão dos dados em subconjuntos de treino, validação e teste. O objetivo é garantir que os dados estejam organizados, representem bem o fenômeno estudado e não introduzam vieses ou ruídos excessivos no modelo (De Mattos Paixão et al., 2022).

Parte crucial do desenvolvimento dos modelos de ML, a etapa de treinamento consiste no processo sistemático de ajustar os parâmetros internos de um algoritmo a partir de um conjunto de dados de entrada, de modo a minimizar erros e maximizar sua capacidade de generalização para novos exemplos. Para Talaei Khoei e Kaabouch (2023), esse procedimento envolve a definição de uma função de custo ou perda, que quantifica a discrepância entre as previsões geradas pelo modelo e os valores reais, e a aplicação de métodos de otimização, como o gradiente descendente, para atualizar iterativamente os parâmetros. Durante o treinamento, técnicas como validação cruzada, regularização e ajuste de hiperparâmetros são utilizadas para evitar problemas de sobreajuste e garantir que o modelo aprenda padrões relevantes em vez de ruídos específicos dos dados de treino. Assim, o treinamento configura-se como a etapa central do ciclo de desenvolvimento pois é nela que o modelo adquire a capacidade de realizar inferências preditivas ou classificatórias de forma robusta em contextos práticos.

Na terceira e última etapa, o modelo treinado é testado em dados que não foram usados no treinamento (ou no ajuste final) para verificar seu desempenho e sua capacidade de generalização. Essa etapa emprega métricas adequadas ao problema (precisão, sensibilidade, especificidade, AUC, erro quadrático médio etc.) para quantificar o quão bem o modelo funciona em cenários não vistos. Também é nessa fase que se verifica se o modelo está sofrendo de overfitting (quando ele ajustou demais ao conjunto de treino e perdeu generalidade) ou underfitting, e decide-se se ajustes (por exemplo, novos hiperparâmetros, mudança de modelo ou mais dados) são necessários (De Mattos Paixão et al., 2022).

Com o entendimento das etapas de desenvolvimento de um modelo de ML, a seção seguinte apresenta os resultados obtidos ao analisar este processo sob a perspectiva do conceito de transdução informacional, a fim de levantar fatores que conduzam a credibilidade destes recursos (Sant'Ana, 2021).

Esta obra está bajo una licencia internacional Creative Commons Atribución 4.0.

Desenvolvimento

As reflexões estabelecidas e amparadas nos resultados a seguir buscam contribuir com a orientação da Federação Internacional de Associações e Instituições Bibliotecárias (IFLA em inglês), em sua declaração sobre bibliotecas e inteligência artificial, sobre a importância da Ciência da Informação obter uma compreensão de como modelos de ML funcionam, o que podem e não podem fazer (IFLA FAIFE, 2020).

Quanto às suas delimitações, esta pesquisa não realizou uma análise abrangente de diversos portais ou plataformas, entretanto, considerando o volume e a disponibilidade de modelos de ML acessíveis nos dois websites analisados, foi possível estabelecer análises endereçadas ao atendimento do objetivo da pesquisa, cujas quais encontram-se organizadas em duas subseções: a primeira abordando a síntese das evidências encontradas no que se refere a transformação de dados no processo de desenvolvimento de modelos de ML, dando ênfase às técnicas de normalização de dados; e a segunda versando sobre tais aspectos sob a perspectiva das transduções informacionais a partir do conteúdo analisado nos modelos de ML identificados nos dois portais observados; e a terceira destacando as limitações analíticas da pesquisa e próximos passos.

Síntese das evidências encontradas

Goldschmidt e Passos (2005, p. 37) afirmam que em um projeto de ML *"[...] é comum que os dados estejam incompletos, ruidosos ou inconsistentes"*. Os autores ainda destacam a necessidade dos dados serem submetidos a *"[...] uma verificação da consistência das informações, a correção de possíveis erros e o preenchimento ou a eliminação de valores desconhecidos e redundantes, além da eliminação de valores não pertencentes ao domínio"* (Goldschmidt & Passos, 2005, p. 37).

Wickham (2014) e Torgo (2017) destacam diversas possibilidades de transformações de dados no contexto do desenvolvimento de um modelo de ML, por exemplo, binarização, desordenação, agrupamento, ordenação, thresholding, normalização, discretização e calibração.

Outro aspecto importante a ser considerado são os procedimentos que podem ser adotados quando é necessário o preenchimento de valores ausentes. Para Goldschmidt e Passos (2005, p. 37), *"medidas estatísticas podem ser empregadas como alternativa à utilização de constantes padrões no processo de preenchimento de valores ausentes"*, por exemplo, a média, mediana ou moda. Ou seja, para preencher dados que estavam faltando nos dados originalmente coletados, a transformação aqui inclui usar essas medidas estatísticas para injetar dados sintéticos no conjunto de dados originais.

Diante das diversas possibilidades de transformações de dados existentes é possível verificar o quanto os atores envolvidos com o desenvolvimento e uso de modelos de ML estão cientes ou não de tais fatos, sendo assim, é necessário considerar alguns detalhes relacionados ao procedimento denominado normalização dos dados.

A normalização é uma técnica que pode ser aplicada durante a preparação dos dados que serão usados no desenvolvimento de um modelo de ML. Normalizar os dados significa alterar os valores das colunas numéricas no conjunto de dados para uma escala comum, evitando distorções nos intervalos de valores e perdas de informações (Pereira Gonçalves & Lima, 2013). Por exemplo, na possibilidade de existir um conjunto de dados que contenha uma coluna com valores variando de 0 a 1 e outra coluna com valores variando de 10.000 a 100.000, essa grande diferença na escala dos números poderia causar problemas no desenvolvimento do modelo de ML. A normalização dos dados cria novos valores que mantêm a distribuição geral e as proporções

Esta obra está bajo una licencia internacional Creative Commons Atribución 4.0.

dos dados originais. A normalização coloca os dados no intervalo entre 0 e 1 ou -1 e 1, caso haja valores negativos, sem distorcer as diferenças nas faixas de valores.

Pereira Gonçalves e Lima (2013), ao apresentarem maneiras de normalizar os dados, ressaltam que a similaridade em comum entre as técnicas é o uso de fórmulas matemáticas que quando aplicadas atingem o objetivo de alterar os valores das colunas numéricas no conjunto de dados para uma escala comum. Seguem alguns exemplos de técnicas de normalização de dados apresentados pelos autores.

A) O Z-score ou Desvio Padrão utiliza o desvio padrão e a média dos valores para normalizar (Pereira Gonçalves & Lima, 2013). A fórmula necessária para a aplicação da técnica é a seguinte: $x' = (x - \bar{A}) / \sigma A$. Onde: x': valor normalizado; x: valor do atributo a ser normalizado; $\bar{A}$: valor da média do atributo A; σA: valor do desvio padrão atributo A; Cálculo da Média ($\bar{A}$): $\bar{A} = \Sigma A / n$. E, ainda, onde: ΣA: Soma de todos os valores do atributo A; n: Número total de valores no atributo A; Cálculo do Desvio Padrão (σA): $\sigma A = \sqrt{(\Sigma(A - \bar{A})^2 / (n - 1))}$. Onde: $(A - \bar{A})^2$: Quadrado da diferença entre cada valor do atributo A e a média do atributo A; e $\Sigma(A - \bar{A})^2$: Soma dos quadrados das diferenças entre cada valor do atributo A e a média do atributo A.
B) A técnica denominada Linear (Min-Max) realiza uma interpolação linear que leva em consideração os valores mínimo e máximo dos atributos (Gonçalves e Lima, 2013). A fórmula necessária para a aplicação da técnica é a seguinte: x' = ((x - Min_A) / (Max_A - Min_A)) * (novo_Max_A - novo_Min_A) + novo_Min_A. Onde: x': valor normalizado; x: valor do atributo a ser normalizado; Min_A: valor mínimo do atributo A; Max_A: valor máximo do atributo A; novo_Min_A: novo valor mínimo do atributo A; e novo_Max_A: novo valor máximo do atributo A
C) Por fim, a técnica da Escala Decimal desloca o ponto decimal dos valores do atributo que se quer normalizar, baseando-se no número de casas decimais do atributo que tiver o maior valor (Pereira Gonçalves & Lima, 2013). A fórmula necessária para a aplicação da técnica é a seguinte: x' = x / 10^j. Onde: x': valor normalizado; x: valor do atributo a ser normalizado e j: menor valor inteiro tal que Max(|x'|) < 1

Além desses três exemplos de técnicas de normalização de dados existem outras técnicas disponíveis que transformam os valores originais em outros, e que embora matematicamente possam ser considerados os mesmos, são resultados de dados transformados em outros com base em técnicas que o responsável pelo desenvolvimento do modelo de ML escolheu. Outro ponto importante é esclarecer que a técnica de normalização de dados não é necessária em todos os projetos de desenvolvimento de um modelo de ML, por isso, apresentá-la aqui, tem o objetivo de ilustrar possíveis transformações nos dados que os atores envolvidos, principalmente, no uso do modelo de ML desenvolvido, podem estar cientes ou não.

Embora a técnica de normalização de dados não seja necessária em todos os projetos, a transformação dos dados sempre vai ocorrer, usando outras possibilidades já citadas no início desta seção, confirmando que os modelos de ML desenvolvidos sempre serão resultados da aplicação desta ou de outras possibilidades de transformações de dados. Para alcançar os resultados, nesta pesquisa foram analisados dois websites que oferecem o download gratuito de modelos de ML (Huggin Face e GitHub), cuja discussão encontra-se apresentada a seguir.

Esta obra está bajo una licencia internacional Creative Commons Atribución 4.0.

Análise das transformações e transduções em modelos de Machine Learning

A pesquisa buscou-se identificar, nos dois websites selecionados, informações que apresentassem com clareza as transformações e transduções informacionais que ocorreram nos dados usados para treinar os modelos de ML disponíveis. A seguir, são apresentadas as reflexões estabelecidas a partir de cada portal.

a) Huggin Face

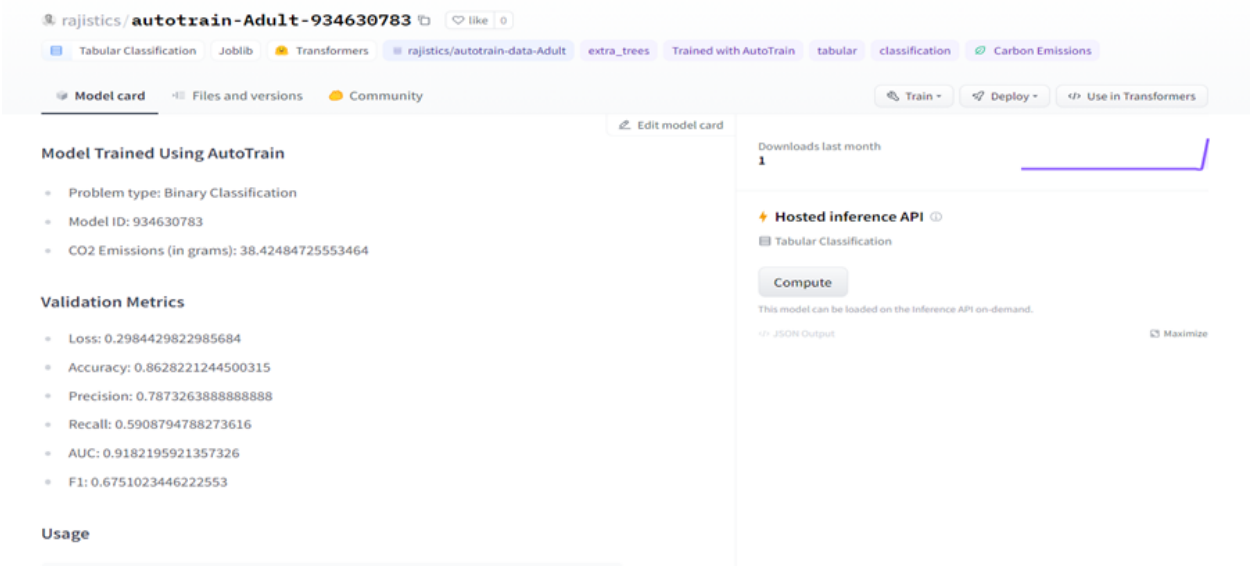
Ao realizar a análise (setembro de 2024) o Huggin Face informava ter disponível 935.539 modelos de ML para download. O portal organiza seus modelos de ML em diferentes categorias conforme a tarefa para a qual foram treinados. Em NLP, há modelos para classificação de texto, sumarização, tradução, geração de texto, question answering e extração de características. Em visão computacional, destacam-se modelos de classificação de imagens, detecção de objetos, segmentação, geração de imagem a partir de texto e análise de vídeo. Na área de áudio, estão disponíveis ferramentas para reconhecimento automático de fala, síntese de voz, classificação de áudio e detecção de atividade vocal. Há ainda modelos multimodais, que combinam texto, imagem, áudio e vídeo em tarefas como descrição de imagens, perguntas visuais e análise de documentos, além de suporte a dados tabulares, cobrindo assim um amplo espectro de aplicações de IA.

A Figura 2 ilustra a apresentação de informações no Huggin Face. Nela, é possível observar determinado modelo de ML (ex: AutoTrain) e suas métricas de validação, entretanto, não foram localizados detalhes sobre as transformações, ou transduções realizadas nos dados usados para a fase de treinamento do modelo de ML.

Na Figura 2 também é possível notar que os dados originais usados na fase de treinamento do modelo não estão facilmente disponíveis para uma consulta. Ainda que estivessem, não fica claro pelas informações apresentadas no site a questão em discussão nesta pesquisa sobre transformações realizadas nos dados. Portanto, não é possível identificar a quais transformações os dados originais foram submetidos no desenvolvimento dos modelos de ML disponíveis no site.

Figura 2

Modelo 1 de ML disponível no portal Huggin Face



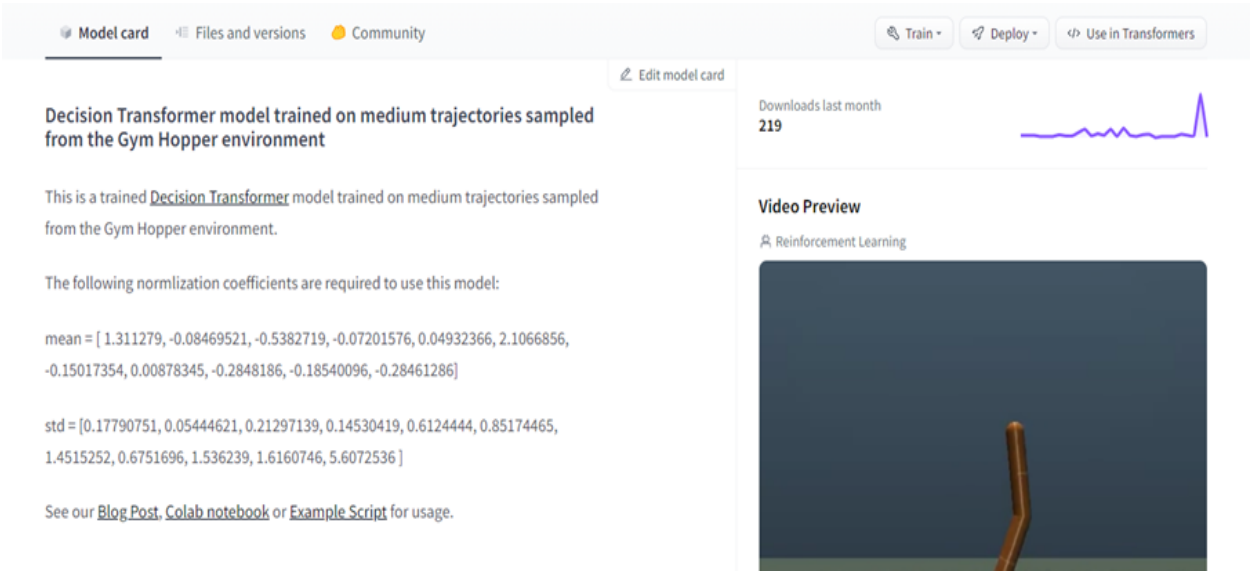
Nota. Rajistics. (2023). autotrain-Adult-934630783. Hugging Face.

Esta obra está bajo una licencia internacional Creative Commons Atribución 4.0.

Na figura 3 é possível notar em outro exemplo de modelo de ML disponível no website, que os coeficientes de normalização necessários para usar o modelo são informados, contudo, não fica claro quais transformações, ou transduções, nos dados originais, foram realizadas e que técnica de normalização foi utilizada.

Figura 3

Modelo 2 de ML disponível no portal Huggin Face



Nota. Hugging Face. (2023). Decision transformer gym hopper medium.

Neste exemplo (Figura 3) também não estão disponíveis facilmente os dados originais usados no treinamento do modelo de ML.

2) GitHub

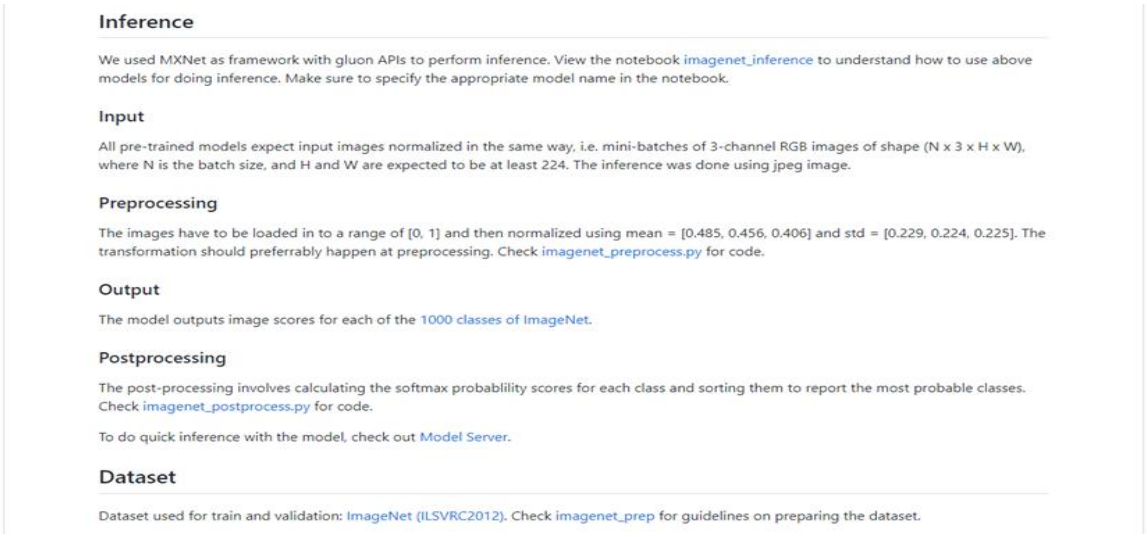
O website GitHub se apresenta como uma coleção de modelos pré-treinados e de última geração no formato Open Neural Network Exchange, ONNX, um formato padrão aberto para representar modelos de ML. Durante a análise (setembro de 2024), o website informava que para acessar os dados originais usados no treinamento do modelo, bem como referências à arquitetura de cada modelo de ML disponível no site, era necessário fazer o download do código fonte desenvolvido pelos programadores para encontrar esses detalhes. Neste caso, não se exige dos colaboradores que disponibilizam modelos de ML no site a apresentação dos detalhes sobre transformações que os dados originais foram submetidos.

Na Figura 4, é exibido um modelo de ML disponível neste website, na qual é possível observar que o padrão se repete no que diz respeito a pouca informação sobre as transformações, ou transduções que os dados originais foram submetidos para o treinamento. Consultando a documentação completa desse exemplo, notou-se algumas orientações de como o modelo deve ser usado fazendo referência a algumas configurações de normalização de dados que devem ser respeitadas, porém, detalhes de como os dados originais foram tratados ou transformados não são apresentados.

Esta obra está bajo una licencia internacional Creative Commons Atribución 4.0.

Figura 4

Modelo 1 de ML disponível no portal GitHub



Inference

We used MXNet as framework with gluon APIs to perform inference. View the notebook imagenet_inference to understand how to use above models for doing inference. Make sure to specify the appropriate model name in the notebook.

Input

All pre-trained models expect input images normalized in the same way, i.e. mini-batches of 3-channel RGB images of shape (N x 3 x H x W), where N is the batch size, and H and W are expected to be at least 224. The inference was done using jpeg image.

Preprocessing

The images have to be loaded in to a range of [0, 1] and then normalized using mean = [0.485, 0.456, 0.406] and std = [0.229, 0.224, 0.225]. The transformation should preferrably happen at preprocessing. Check imagenet_preprocess.py for code.

Output

The model outputs image scores for each of the 1000 classes of ImageNet.

Postprocessing

The post-processing involves calculating the softmax probablility scores for each class and sorting them to report the most probable classes. Check imagenet_postprocess.py for code.

To do quick inference with the model, check out Model Server.

Dataset

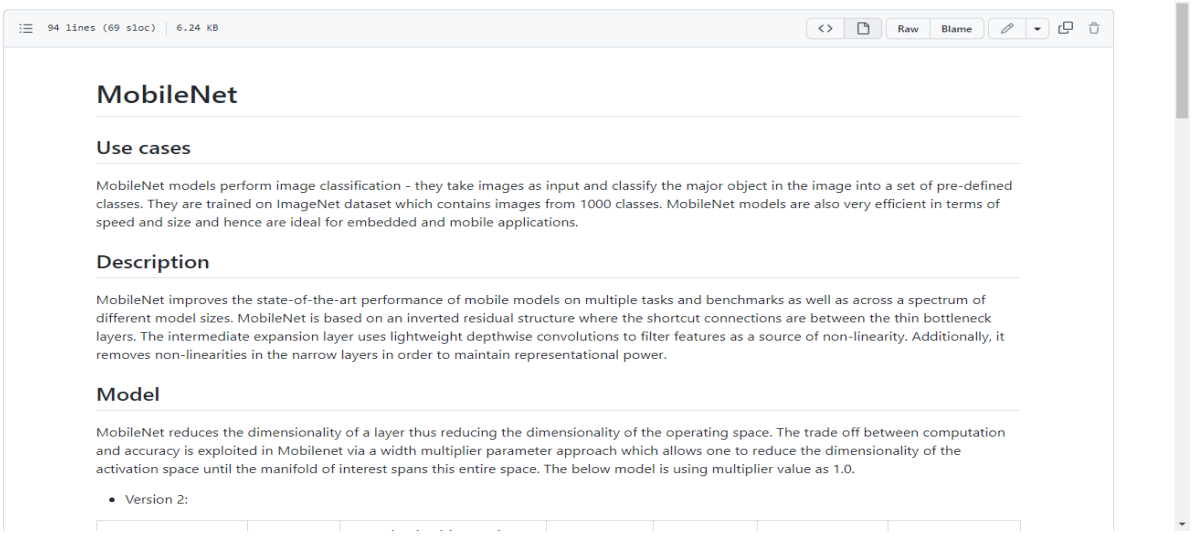
Dataset used for train and validation: ImageNet (ILSVRC2012). Check imagenet_prep for guidelines on preparing the dataset.

Nota. ONNX. (2023). MobileNet model README. GitHub.

No exemplo ilustrado na Figura 5, o modelo foi treinado para classificar a imagem que é apresentada como entrada em uma das 1000 classes que o modelo pode reconhecer. Como pode ser visto na Figura 5, algumas informações sobre o método de normalização de dados são apresentadas para as imagens de entrada, por exemplo, definindo que a imagem de entrada sejam mini-lotes de imagens RGB de 3 canais de forma (N x 3 x H x W), onde N é o tamanho do lote e espera-se que H e W ser pelo menos 224. Quanto ao pré-processamento das imagens, estas devem ser carregadas em um intervalo de [0, 1] e depois normalizadas usando média = [0,485, 0,456, 0,406] e std = [0,229, 0,224, 0,225]. Um ponto importante aqui, é que essas informações se referem a imagem que será classificada pelo modelo, mas nada é informado sobre as imagens originais que foram usadas para o treinamento do modelo e sobre as transformações, ou transduções, que teriam sido realizadas.

Figura 5

Modelo 2 de ML disponível no portal GitHub



94 lines (69 sloc) | 6.24 KB

Raw Blame

MobileNet

Use cases

MobileNet models perform image classification - they take images as input and classify the major object in the image into a set of pre-defined classes. They are trained on ImageNet dataset which contains images from 1000 classes. MobileNet models are also very efficient in terms of speed and size and hence are ideal for embedded and mobile applications.

Description

MobileNet improves the state-of-the-art performance of mobile models on multiple tasks and benchmarks as well as across a spectrum of different model sizes. MobileNet is based on an inverted residual structure where the shortcut connections are between the thin bottleneck layers. The intermediate expansion layer uses lightweight depthwise convolutions to filter features as a source of non-linearity. Additionally, it removes non-linearities in the narrow layers in order to maintain representational power.

Model

MobileNet reduces the dimensionality of a layer thus reducing the dimensionality of the operating space. The trade off between computation and accuracy is exploited in Mobilenet via a width multiplier parameter approach which allows one to reduce the dimensionality of the activation space until the manifold of interest spans this entire space. The below model is using multiplier value as 1.0.

- Version 2:

Nota. GitHub. (2023). MobileNet model.

Esta obra está bajo una licencia internacional Creative Commons Atribución 4.0.

As análises estabelecidas permitiram clarear como se dá os processos de transformações e transduções informacionais que ocorrem nos dados usados para treinar os modelos de ML disponíveis a partir de uma amostra de modelos concretos. Entretanto, algumas limitações relevantes precisam ser elencadas, como pode-se observar na última subseção destes resultados.

c) Limitações analíticas da pesquisa e próximos passos

Ao analisar os dados apresentados, observa-se um grau de dificuldade em localizar os detalhes necessários para a reflexão acerca das transformações que ocorrem no processo de desenvolvimento dos modelos de ML, considerado fator preponderante para a garantia da credibilidade. Ainda assim, foi possível refletir sobre possíveis insciências dos atores envolvidos.

É necessário destacar que os websites analisados foram desenvolvidos para um público específico, ou seja, desenvolvedores de modelos de ML, portanto, pode-se inferir que parte majoritária de seus usuários compreendem as nuances técnicas envolvidas. Entretanto, para as questões principais discutidas neste estudo, identificar as transduções informacionais no desenvolvimento de um modelo de ML e refletir sobre possíveis insciências dos atores envolvidos, percebe-se que faltam informações nesses sites para uma completa ciência daqueles que o acessam.

Especificamente sobre o conceito de transduções informacionais no contexto do desenvolvimento de modelos de ML, quando a técnica de normalização de dados é aplicada pode ocorrer uma transdução em duas fases porque inicialmente os dados originais são coletados e inseridos em um banco de dados, quando ocorre a primeira fase de uma transdução informacional. A segunda fase desta transdução parece ocorrer quando logo em seguida, com a utilização de algoritmos de ML e a técnica de normalização dos dados, esses mesmos dados são transformados em outros e mais uma vez registrados ou persistidos no banco de dados usado para ser de fato usado no treinamento do modelo de ML. Contudo, os websites analisados não apresentam os detalhes de como esses dados usados na fase de treinamento do modelo de ML teriam sido transformados.

A transdução informacional no processo de desenvolvimento de modelos de ML pode, portanto, levantar preocupações legítimas. Para a aplicação da técnica de normalização dos dados existem algumas possibilidades, questionar se a técnica escolhida foi a mais apropriada para o projeto em desenvolvimento parece ser justificável. Por outro lado, o cenário descrito pode ilustrar uma assimetria, ou distância entre o que os atores envolvidos pensam que sabem sobre os dados usados e o que realmente ocorre durante o processo. A insciência informacional que pode existir é causada pelas transduções informacionais ocorridas no desenvolvimento dos modelos de ML e pela carência de informações sobre o tema nos websites pesquisados.

Como foi possível verificar nos sites analisados, parece não ser um ponto importante destacar informações sobre as transformações que os dados originais foram submetidos para a realização do treinamento do modelo de ML. O fato de terem um público bem específico (programadores que desenvolvem modelos de ML) pode justificar a dificuldade de identificar quais transformações ocorreram nos dados originais.

Por outro lado, aos profissionais de outras áreas que desejam pesquisar como os dados e as informações são tratados e manipulados durante o processo de desenvolvimento de um modelo de ML, saber que possíveis transduções informacionais podem ter ocorrido embasaria uma cautela sobre a qualidade do modelo de ML desenvolvido e usado, não ficando restrito apenas a sua qualificação baseada na autoridade da empresa que a criou. Se os modelos estão sendo usados por outras áreas, por que não há garantias de transparência para não especialistas? Existe ou deveria existir responsabilidade epistemológica? Isso é apenas falha técnica ou problema

Esta obra está bajo una licencia internacional Creative Commons Atribución 4.0.

ético? Esse cenário abre uma discussão sobre a necessidade de padrões de governança e accountability algorítmica, conforme discutido por Floridi e Cowls (2022), se os portais não consideram usuários não especialistas, há um risco real de assimetria epistêmica na sociedade da informação.

Outro aspecto relevante abordado neste estudo é a explicabilidade dos modelos de ML. Modelos de caixa-preta, que não oferecem clareza sobre seu funcionamento interno, podem gerar desconfiança e resistência em sua adoção. A interpretabilidade dos algoritmos é, portanto, um fator essencial para garantir a aceitação e o uso responsável dessas tecnologias.

Portanto, acredita-se que a pouca informação sobre como ocorre o desenvolvimento de modelos de ML disponibilizadas nos sites analisados estão criando assimetrias no controle ou na percepção do processo como um todo (Sant'Ana, 2021), por isso, é necessário trabalhar no sentido de buscar caminhos que possam minimizar as assimetrias (Sant'Ana, 2021). Por fim, com base na orientação do IFLA FAIFE (2020) sobre a necessidade de uma compreensão de como modelos de Machine Learning funcionam, o que podem e não podem fazer, este estudo contribui para a Ciência da Informação indicando a importância desta ciência para a sociedade atual influenciada por modelos de ML.

4. CONCLUSÕES

O estudo realizado permitiu levantar reflexões críticas acerca da insciência dos atores envolvidos no desenvolvimento e uso de modelos de ML, com foco nas transduções informacionais que ocorrem ao longo de suas etapas. A análise dos portais Hugging Face e GitHub evidenciou que, embora disponibilizem um grande volume de modelos e ferramentas, não oferecem informações detalhadas sobre as transformações aplicadas aos dados originais, limitando a compreensão dos usuários — tanto especialistas quanto não especialistas — sobre a integridade e confiabilidade dos modelos disponibilizados.

Constatou-se que as transduções informacionais permanecem opacas e pouco documentadas, o que gera uma assimetria entre o que os desenvolvedores acreditam dominar e o que de fato ocorre nos processos técnicos. Essa constatação corrobora os resultados de Gallagher et al. (2021) que analisaram problemas similares no contexto do ensino superior, argumentando que a expansão da datificação e das infraestruturas algorítmicas torna muitos processos técnicos invisíveis para professores e estudantes, limitando a capacidade de ambos de compreenderem criticamente como essas tecnologias moldam práticas pedagógicas e relações de poder. A falta de transparência ontológica na automação gera dificuldade de distinguir interações humanas de interações mediadas por sistemas automatizados e compromete a capacidade dos estudantes de compreenderem plenamente quem está atuando no processo educativo e quais lógicas o estruturam (Gallagher et al., 2021).

No contexto do desenvolvimento de ML, essas lacunas comprometem a explicabilidade e a confiabilidade dos modelos, uma vez que nem sempre é possível rastrear as técnicas de pré-processamento aplicadas, como a normalização de dados ou outras transformações estatísticas. Por isso, a reputação de grandes empresas não deve ser o único critério de confiança: a transparência metodológica é essencial para a aceitação social e científica desses sistemas.

Assim, o presente trabalho não busca apenas levantar a existência de insciência informacional, mas defender a urgência de um protocolo mínimo de rastreabilidade das transformações algorítmicas. Propõe-se que futuras plataformas de disponibilização de modelos de ML adotem: (a) descrição padronizada das transduções aplicadas; (b) uso de metadados sobre decisões técnicas; (c) e mecanismos de auditoria humana mais

Esta obra está bajo una licencia internacional Creative Commons Atribución 4.0.

transparentes e acessíveis para comunicar as transduções informacionais que ocorrem durante o desenvolvimento de um modelo.

Portanto, abre-se caminho para novas investigações que poderiam desenvolver mecanismos padronizados de registro e disponibilização das transduções informacionais, aplicáveis em plataformas de compartilhamento de modelos; explorar a criação de ferramentas de auditoria capazes de rastrear e expor as transformações de dados em diferentes fases do ciclo de vida do modelo; investigar os impactos sociais e éticos da insciência informacional, sobretudo em setores críticos como saúde, finanças e educação; e até mesmo ampliar o diálogo interdisciplinar entre Ciência da Informação, Ciência da Computação e áreas regulatórias, visando à formulação de diretrizes que fortaleçam a governança dos sistemas de IA.

No campo da Ciência da Informação, a pesquisa reforça a importância de compreender os fluxos e mediações que estruturam a produção e o uso de conhecimento em ambientes digitais, destacando como a opacidade das transduções impacta o acesso, a confiabilidade e a ética informacional. Já na Ciência da Computação, contribui ao evidenciar a necessidade de boas práticas de documentação, transparência e explicabilidade nos modelos de ML, superando a lógica do "caixa-preta" e aproximando o desenvolvimento tecnológico das exigências sociais de confiança e responsabilidade. Como os modelos de ML moldam percepções do real, qualquer opacidade sobre suas transduções informacionais não é apenas uma limitação técnica, mas uma questão epistemológica e social — e deve ser tratada como tal.

Conflicto de intereses / Conflito de interesses:

Os autores declaram que não incorreram em conflitos de interesses.

Rol de los autores / Papel dos autores:

Fábio Moreira: Redação – revisão e edição, visualização.

Marcos Siqueira: Conceituação, metodologia, curadoria de dados, análise formal, investigação, redação – rascunho original.

Rachel Alves: Conceituação, metodologia, validação, supervisão, administração do projeto.

Ricardo Sant'Ana: Conceituação, metodologia, validação, supervisão

Fuentes de financiamiento / Fontes de financiamento:

Os autores declaram que não receberam financiamento específico para esta pesquisa.

Aspectos éticos / Legais; Ethics / legals:

Os autores declaram não ter incorrido em práticas antiéticas, nem ter omitido aspectos legais na realização da pesquisa.

REFERÊNCIAS

Aldieri, A., Gamage, T. P. B., Amedeo La Mattina, A., Loewe, A., Pappalardo, F., & Viceconti, M. (2025). Consensus statement on the credibility assessment of machine learning predictors. *Briefings in Bioinformatics*, *26*(2), bbaf100. https://doi.org/10.1093/bib/bbaf100

Aleixo, D. V. B. S. (2020). *O estado de anomia dos dados no acesso aos dados governamentais abertos no Brasil* [Tese de Doutorado, Universidade Estadual Paulista "Júlio de Mesquita Filho"]. Repositório Institucional do Universidade Estadual Paulista. http://hdl.handle.net/11449/191686

Ali, A., Abd Razak, S., Othman, S. H., Eisa, T. A. E., Al-Dhaqm, A., Nasser, M., Elhassan, T., Elshafie, H., & Saif, A. (2022). Financial fraud detection based on machine learning: a systematic literature review. *Applied Sciences*, *12*(19), 9637. https://doi.org/10.3390/app12199637

Esta obra está bajo una licencia internacional Creative Commons Atribución 4.0.

De Mattos Paixão, G. M., Campos Santos, B., Martins de Araujo, R., Horta Ribeiro, M., Lopes de Moraes, J., & Ribeiro, A. L. (2022). Machine learning na medicina: Revisão e aplicabilidade. *Arquivos Brasileiros de Cardiologia, 118*(1), 95–102. https://doi.org/10.36660/abc.20200596

De Souza de Aguiar Monteiro, E. C., & Gonçalves Sant'Ana, R. C. (2022). Anomia de pesquisadores no compartilhamento de dados. *Em Questão*, *29*, 122627. https://doi.org/10.19132/1808-5245.29.122627

De Souza Souza, A. P., De Jesus Conceição, C., Pancoto, M. A., Cecote, N. Q. B., Pedra, R. R., da Silva Oliveira, R. M., ... & Gomes, W. T. (2024). Personalização da aprendizagem com inteligência artificial: Como a IA está transformando o ensino eo currículo. *Aracê*, *6*(3), 5816-5831. https://doi.org/10.56238/arev6n3-092

Floridi, L., & Cowls, J. (2022). A unified framework of five principles for AI in society. In S. Carta (Ed.), *Machine learning and the city: Applications in architecture and urban design* (pp. 535–545). John Wiley & Sons. https://doi.org/10.1002/9781119815075.ch45

Gallagher, M., Breines, M., & Blaney, M. (2021). Ontological transparency,(in) visibility, and hidden curricula: Critical pedagogy amidst contentious edtech. *Postdigital Science and Education*, *3*(2), 425-443. https://doi.org/10.1007/s42438-020-00198-1

Georgescu, A. L., Pappalardo, A., Cucu, H., & Blott, M. (2021). Performance vs. hardware requirements in state-of-the-art automatic speech recognition. *EURASIP Journal on Audio, Speech, and Music Processing*, *2021*(1), 28. https://doi.org/10.1186/s13636-021-00217-4

Géron, A. (2017). *Mãos à obra: Aprendizado de máquina com Scikit-Learn, Keras e TensorFlow: Conceitos, ferramentas e técnicas para a construção de sistemas inteligentes* (3ª ed.). Alta books.

Ghaffar Nia, N., Kaplanoglu, E., & Nasab, A. (2023). Evaluation of artificial intelligence techniques in disease diagnosis and prediction. *Discover Artificial Intelligence*, *3*(1), 5. https://doi.org/10.1007/s44163-023-00049-5

Goldschmidt, R., & Passos, E. (2005). *Data mining: Um guia prático.* Elsevier.

IFLA FAIFE. (2020). *IFLA Statement on libraries and artificial intelligence.* International Federation of Library Associations and Institutions. https://repository.ifla.org/handle/20.500.14598/1646

Islam, M. M., Hassan, S., Akter, S., Jibon, F. A., & Sahidullah, M. (2024). A comprehensive review of predictive analytics models for mental illness using machine learning algorithms. *Healthcare Analytics*, *6*, 100350. https://doi.org/10.1016/j.health.2024.100350

Khan, A. I., & Al-Habsi, S. (2020). Machine learning in computer vision. *Procedia Computer Science, 167*(1), 1444–1451. https://doi.org/10.1016/j.procs.2020.03.355

Ludermir, T. B. (2021). Inteligência artificial e aprendizado de máquina: Estado atual e tendências. *Estudos Avançados, 35*(101), 85–94. https://doi.org/10.1590/s0103-4014.2021.35101.007

Munir, K., Elahi, H., Ayub, A., Frezza, F., & Rizzi, A. (2019). Cancer diagnosis using deep learning: a bibliographic review. *Cancers*, *11*(9), 1235. https://doi.org/10.3390/cancers11091235

Esta obra está bajo una licencia internacional Creative Commons Atribución 4.0.

Pereira Gonçalves, R., & Lima, E. C. P. (2013). Multidimensional interpolation filters development for risers' database analysis. In *Proceedings of the ASME 2013 32nd International Conference on Ocean, Offshore and Arctic Engineering: Volume 1: Offshore Technology*. American Society of Mechanical Engineers. https://doi.org/10.1115/OMAE2013-10372

Roy, D., & Dutta, M. (2022). A systematic review and research perspective on recommender systems. *Journal of Big Data*, *9*(1), 59. https://doi.org/10.1186/s40537-022-00592-5

Sant'Ana, R. C. G. (2019). Transdução informacional: Impactos do controle sobre os dados. En D. Martínez-Ávila; E. A. Souza; M. E. Q. Gonzalez (Ed.), *Informação, conhecimento, ação autônoma e big data: Continuidade ou revolução?* (pp. 117–128). Cultura Acadêmica/FiloCzar. https://doi.org/10.36311/2019.978-85-7249-055-9.p117-128

Talaei Khoei, T., & Kaabouch, N. (2023). Machine learning: Models, challenges, and research directions. *Future Internet*, *15*(10), 332. https://doi.org/10.3390/fi15100332

Torgo, L. (2017). *Data mining with R: Learning with case studies.* Chapman and Hall/CRC.

Wickham, H. (2014). Tidy data. *Journal of Statistical Software, 59*(10), 1–23. https://doi.org/10.18637/jss.v059.i10

Esta obra está bajo una licencia internacional Creative Commons Atribución 4.0.