

EDGARD JOSEPH KIRIYAMA

POSTECH

MACHINE LEARNING ENGINEERING

FUNDAMENTOS DE IA E MACHINE LEARNING

AULA 01

SUMÁRIO

O QUE VEM POR AÍ?	3
HANDS ON	4
SAIBA MAIS	5
O QUE VOCÊ VIU NESTA AULA?	10
REFERÊNCIAS.....	11

EMSE

O QUE VEM POR AÍ?

Bem-vindos e bem-vindas ao mundo fascinante da Inteligência Artificial (IA) e do Machine Learning (ML)! Se você já se perguntou como os computadores podem aprender com a experiência ou imitar a inteligência humana, você está prestes a mergulhar em um universo incrível de algoritmos e dados inteligentes. Em termos simples, a IA busca tornar as máquinas inteligentes, enquanto o ML é uma das principais estrelas do show, permitindo que essas máquinas aprendam e melhorem por conta própria.

Imagine seu smartphone entendendo suas preferências, carros que dirigem sozinhos e assistentes virtuais que parecem antecipar suas necessidades. Essas maravilhas são produzidas por IA e ML. No centro desse universo, temos algoritmos que não apenas processam informações como também aprendem com elas, ajustando-se e evoluindo ao longo do tempo.

Então, prepare-se para explorar o que está por trás da cortina da inteligência artificial e descobrir como os computadores estão se tornando não apenas ferramentas poderosas, mas também aprendizes ágeis em um mundo cada vez mais digital.

Vamos embarcar em uma jornada interativa cheia de conceitos importantes e entendimentos valiosos, antes de colocarmos nossos modelos em produção.

Bora lá?

HANDS ON

Vamos entrar em uma fascinante jornada pelo universo da Inteligência Artificial (IA) e desvendar as complexidades que os dados apresentam nesse cenário dinâmico. Preparem-se para uma experiência empolgante e repleta de insights transformadores.

Na primeira parte de nossa aventura, mergulharemos nas definições essenciais da Inteligência Artificial. Vamos desmistificar o real significado de IA, desde sua estrutura baseada em redes neurais até as nuances do aprendizado de máquina e aprendizado profundo. Vamos explorar como esses elementos se entrelaçam para criar sistemas inteligentes que podem aprender, adaptar-se e surpreender.

Em seguida, vamos decifrar o intrigante mundo dos dados. Descubra por que, na maioria das vezes, trabalhar com dados estruturados é mais simples e como a verdadeira complexidade reside nos dados não estruturados.

Vamos discutir sobre todos esses temas?

SAIBA MAIS

Inteligência Artificial

A Inteligência Artificial (IA) emerge como um dos campos mais recentes nas áreas de ciências e engenharia. Seu desenvolvimento teve início logo após a Segunda Guerra Mundial, sendo que o termo "Inteligência Artificial" foi cunhado em 1956. Em conjunto com a biologia molecular, a IA é frequentemente mencionada como "o campo no qual eu mais gostaria de estar" por cientistas de diversas disciplinas.

Enquanto um estudante de física possa argumentar, com justificativa, que todas as grandes ideias já foram concebidas por mentes como Galileu, Newton e Einstein, a IA, ao contrário, oferece um vasto terreno em que várias mentes brilhantes, equivalentes a Einsteins e Edisons em tempo integral, ainda podem contribuir significativamente (RUSSEL; NORVIG, 2013).

A IA representa um campo fascinante e dinâmico da ciência da computação, dedicado a desenvolver sistemas capazes de realizar tarefas que normalmente exigiriam inteligência humana. Essa disciplina procura replicar, de certa forma, a capacidade cognitiva, aprendizado e adaptação que caracterizam a inteligência humana. Ao longo dos anos, a IA evoluiu de conceitos teóricos para aplicações práticas, transformando indústrias e redefinindo a maneira como interagimos com a tecnologia.

Em uma perspectiva abrangente, delineia-se a interconexão dos principais elementos da Inteligência Artificial (IA). No ponto mais elevado, encontra-se a IA englobando uma diversidade de teorias e tecnologias. Em seguida, essa amplitude pode ser subdividida em duas categorias fundamentais: aprendizado de máquina (machine learning) e aprendizado profundo (deep learning).

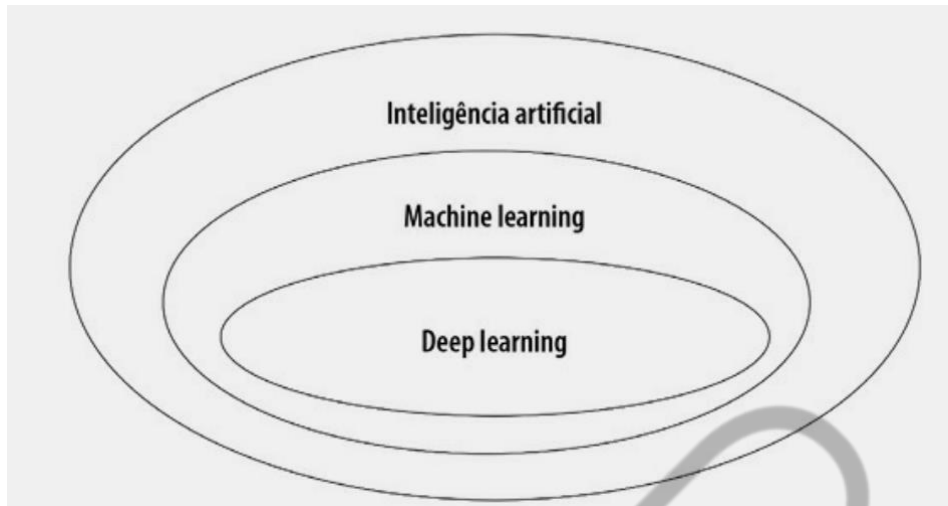


Figura 1 – Estrutura da Inteligência Artificial
Fonte: Taulli (2020)

Dentro desta estrutura, algumas definições são importantes:

- **Inteligência Artificial:** a IA refere-se à capacidade de um sistema computacional em realizar tarefas que, quando executadas por humanos, requerem inteligência. Isso inclui aprendizado, raciocínio, resolução de problemas, compreensão de linguagem natural e interação social.
- **Aprendizado de Máquina (Machine Learning):** parte central da IA, o aprendizado de máquina permite que sistemas aprendam padrões a partir de dados, sem programação explícita. Algoritmos de aprendizado de máquina são cruciais para a capacidade de adaptação e melhoria contínua dos sistemas.
- **Redes Neurais Profundas (Deep Learning):** inspiradas no funcionamento do cérebro humano, as redes neurais são estruturas fundamentais em muitas implementações de IA. Elas consistem em camadas interconectadas de unidades de processamento (neurônios) que são treinadas para reconhecer padrões complexos.

Dados: o Combustível Essencial que Impulsiona a Inteligência Artificial

A era da Inteligência Artificial (IA) está intrinsecamente ligada à abundância e qualidade dos dados que alimentam seus algoritmos inovadores. Em um mundo cada

vez mais digital, a coleta massiva de informações tornou-se uma fonte vital para o desenvolvimento e o aprimoramento contínuo de sistemas inteligentes.

A analogia de "dados como combustível" reflete a dependência crítica que a IA tem desses conjuntos de informações. Assim como um motor necessita de combustível para operar de maneira eficiente, os algoritmos de IA requerem dados diversos e relevantes para aprender, adaptar-se e fornecer resultados significativos.

Os dados, nesse contexto, não são simplesmente números e informações brutos, mas sim o reflexo digital da complexidade do mundo. Desde padrões de comportamento humano até as nuances de linguagens naturais, os dados alimentam as máquinas com insights que desencadeiam a capacidade de tomar decisões, aprender com experiências passadas e até antecipar eventos futuros.

Empresas, instituições de pesquisa e organizações têm investido consideravelmente na coleta ética e curadoria eficaz de dados, reconhecendo que sua qualidade é diretamente proporcional à eficácia da IA resultante. O processo de coleta, armazenamento e processamento de dados, muitas vezes denominado "big data", tornou-se uma disciplina por si só, moldando a paisagem da IA.

De acordo com Taulli (2020), os dados podem vir de muitas fontes diferentes, entre elas:

- Web/rede social (Facebook, Twitter, Instagram, YouTube).
- Dados biométricos (rastreadores de atividades físicas, testes genéticos).
- Sistemas de ponto de venda (de lojas físicas e sites de comércio eletrônico).
- Internet das coisas ou IoT (Etiquetas de identificação e dispositivos inteligentes).
- Sistemas de nuvem (aplicativos de negócios, como salesforce.com).
- Bancos de dados corporativos e planilhas.

Em muitas situações, lidar com dados estruturados é geralmente mais simples. Esses dados costumam originar-se de sistemas como CRM (Gestão do Relacionamento com o Cliente) e ERP (Planejamento de Recursos Empresariais) e,

frequentemente, apresentam volumes menores. Além disso, tendem a ser mais diretos em termos de análise.

Diversos softwares de Business Intelligence (Inteligência de Negócios) estão disponíveis para extrair insights desses dados. No entanto, é importante notar que esses dados estruturados compõem aproximadamente 20% de um projeto de Inteligência Artificial (TAULLI, 2020).

A maior parcela, por outro lado, consistirá em dados não estruturados, ou seja, informações sem um formato predefinido. Trabalhar com esses dados exigirá formatação manual, uma tarefa que pode ser tanto tediosa quanto demorada. Contudo, ferramentas avançadas, como os bancos de dados de próxima geração, incluindo aqueles baseados em NoSQL, podem facilitar esse processo.

Além disso, os sistemas de IA mostram-se eficazes na administração e estruturação de dados, visto que seus algoritmos têm a capacidade de identificar padrões de maneira automatizada (TAULLI, 2020).

São alguns exemplos de dados não estruturados:

- Imagens.
- Vídeos.
- Arquivos de áudio.
- Arquivos de texto.
- Informações de redes sociais.
- Imagens de satélites.

Boa parte destes dados estão armazenados em duas grandes estruturas: Big Data e Data Lakes.

Big Data refere-se a conjuntos massivos de dados que excedem a capacidade de processamento dos métodos de análise de dados tradicionais. Esse paradigma implica não apenas grandes volumes de dados, mas também variedade e velocidade.

Os dados do Big Data podem ser estruturados, semiestruturados ou não estruturados, originando-se de diversas fontes, como mídias sociais, dispositivos IoT e transações financeiras, entre outros (MARQUESONE, 2016).

A abordagem do Big Data envolve o processamento e a análise desses dados para extrair insights significativos, identificar padrões e suportar a tomada de decisões informadas. Tecnologias como Hadoop e Spark são frequentemente utilizadas para processar e armazenar grandes conjuntos de dados de forma distribuída.

Já os Data Lakes são estruturas de armazenamento que permitem a retenção de grandes volumes de dados brutos, sejam eles estruturados ou não estruturados. Diferentemente dos bancos de dados tradicionais, os Data Lakes oferecem flexibilidade no armazenamento de dados sem a necessidade imediata de definir sua estrutura ou finalidade. Eles atuam como repositórios unificados, agregando dados de diversas fontes e permitindo análises mais abrangentes.

A vantagem dos Data Lakes está na capacidade de preservar a integridade dos dados originais, possibilitando análises retrospectivas e a descoberta de novos insights à medida que as necessidades evoluem. Essa flexibilidade também permite que organizações estejam mais bem preparadas para enfrentar os desafios em constante mudança do ambiente de negócios (INMON, 2016).

Contudo, a discussão sobre dados na IA não está isenta de desafios éticos e preocupações com a privacidade. À medida que a importância dos dados cresce, a responsabilidade em gerenciá-los de maneira ética e transparente também se intensifica.

O QUE VOCÊ VIU NESTA AULA?

Começamos desvendando as definições fundamentais da Inteligência Artificial. A IA, em sua essência, refere-se à capacidade de sistemas computacionais executarem tarefas que normalmente exigiriam inteligência humana. Delineamos sua estrutura destacando a presença central das redes neurais, inspiradas na complexidade do cérebro humano.

Além disso, abordamos o Aprendizado de Máquina (Machine Learning) e o Aprendizado Profundo (Deep Learning) como componentes vitais, capacitando os sistemas a aprenderem padrões e a se adaptarem de maneira contínua.

Avançamos para explorar o papel crucial dos dados no ecossistema da IA. Identificamos que, em muitos casos, trabalhar com dados estruturados é mais direto, originando-se de sistemas como CRM e ERP. No entanto, aproximadamente 80% de um projeto de IA é composto por dados não estruturados, informações sem formatação predefinida que demandam esforço manual para sua organização. Adicionalmente, discutimos a importância ética e prática de gerenciar dados de maneira transparente.

Em nossa jornada sobre dados, destacamos as estruturas avançadas, como Big Data e Data Lakes. O Big Data, com sua capacidade de processar grandes volumes e variedades de dados, revelou-se vital para análises sofisticadas. Por outro lado, os Data Lakes oferecem uma abordagem flexível para o armazenamento de dados brutos, proporcionando um ambiente unificado para análises retrospectivas e insights em constante evolução.

IMPORTANTE: não esqueça de praticar com o desafio da disciplina para que você possa aprimorar seus conhecimentos!

Você não está sozinho(a) nesta jornada! Te esperamos no Discord e nas lives com professores(as) especialistas, em que você poderá tirar dúvidas, compartilhar conhecimentos e estabelecer conexões!

Bons estudos!

REFERÊNCIAS

INMON, B. **Data Lake Architecture: designing the data lake and avoiding the garbage dump**. Basking Ridge: R A Peters, 2016.

MARQUESONE, R. **Big Data: técnicas e tecnologias para extração de valor dos dados**. São Paulo: Casa do Código, 2016.

RUSSEL, S.; NORVIG, P. **Inteligência Artificial**. São Paulo: Dtphoenix Editorial, 2013.

TAULLI, T. **Introdução à Inteligência Artificial: uma abordagem não técnica**. São Paulo: Novatec, 2020.

PALAVRAS-CHAVE

Palavras-chave: Inteligência Artificial. Machine Learning. NoSQL. Banco de Dados. Dados.

EMEND



POSTECH