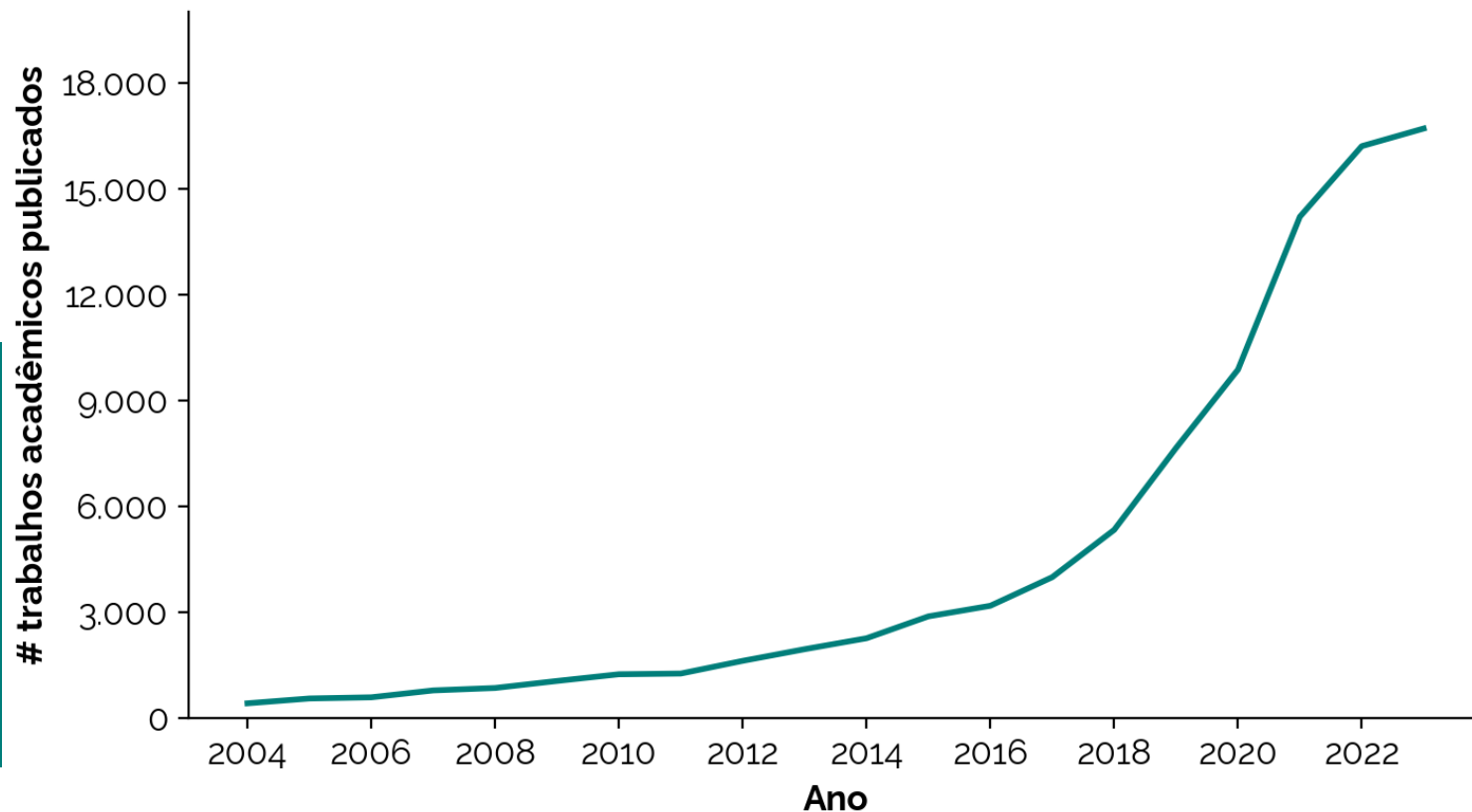


INTELIGÊNCIA ARTIFICIAL NA GEOTECNIA

DESAFIOS E OPORTUNIDADES

EXISTE UM USO CRESCENTE DE INTELIGÊNCIA ARTIFICIAL (IA) E APRENDIZADO DE MÁQUINA NAS GEOCIÊNCIAS

Publicações sobre **Inteligência Artificial** e **Aprendizado de Máquina** em **Geociências** à cada ano



O número de trabalhos no tema saltou de 417 para 16.700 em cerca de 20 anos - com uma grande aceleração no volume de publicações a partir de 2016

CERCA DE 25% DOS TRABALHOS PUBLICADOS PARECEM ESTAR RELACIONADOS À ENGENHARIA GEOTÉCNICA

Fonte: resultados baseados em uma busca realizada no Google Acadêmico no dia 06 de junho de 2024, utilizando o conjunto de termos ("artificial intelligence" OR "machine learning") AND "geosciences" na busca e examinando os resultados encontrados ano a ano.

SLOPE STABILITY 2024

BELO HORIZONTE - BRASIL

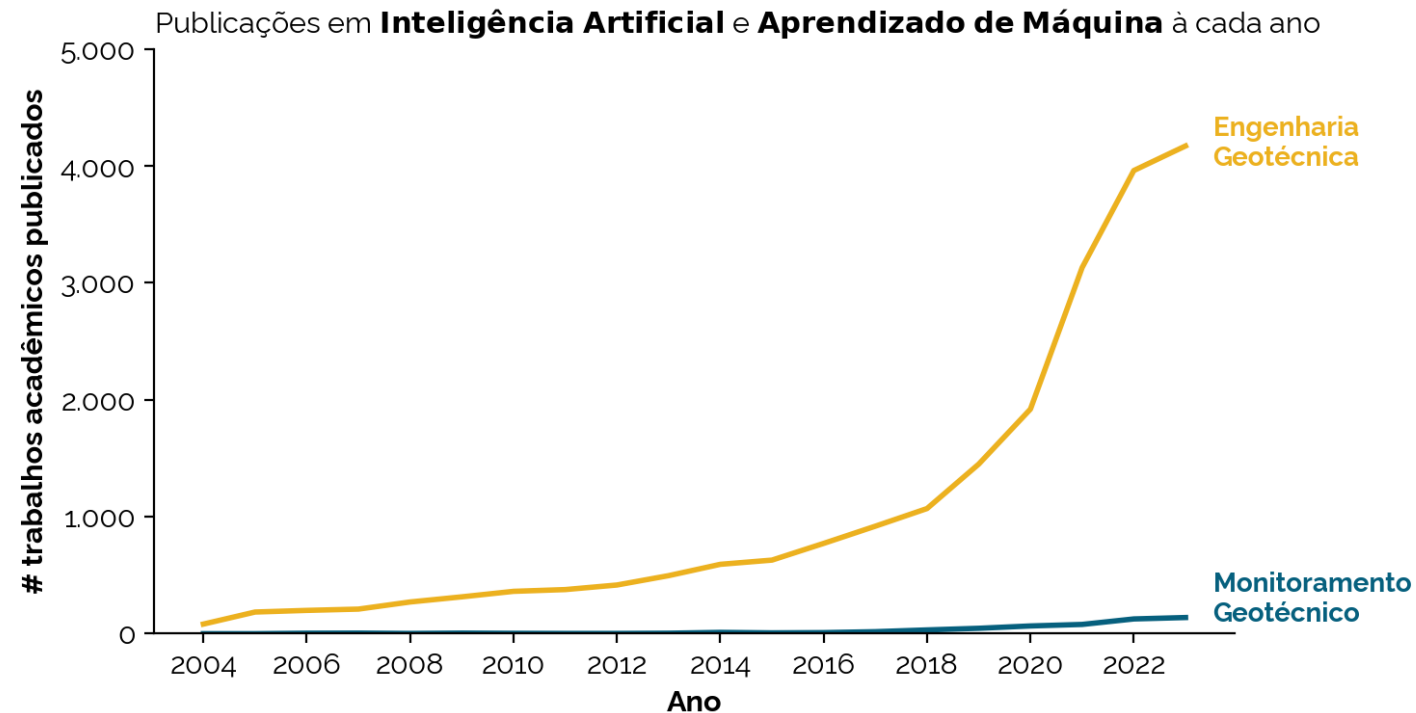
O TEMA DE INTELIGÊNCIA ARTIFICIAL
JÁ APARECE EM EVENTOS DA ÁREA,
COMO O SLOPE STABILITY – MAS
APLICAÇÕES NA INDÚSTRIA AINDA
SÃO POUCO REPRESENTATIVAS.

~10%

dos trabalhos
apresentados no último
simpósio estavam
relacionados à IA



APESAR DA TENDÊNCIA DE USO DE INTELIGÊNCIA ARTIFICIAL, AINDA EXISTE UM GAP NA ADOÇÃO DESTA FERRAMENTA EM ÁREAS QUE PODERIAM SE BENEFICIAR MUITO DE SEU POTENCIAL



O que tem sido feito com IA dentro da área?

O que pode estar por traz deste atraso na adoção de IA nestas áreas?

Como superar os obstáculos que podem estar por traz disto?

Agenda

1. O QUE É, COMO FUNCIONA, QUANDO E POR QUÊ USAR IA?
2. DESAFIOS DA UTILIZAÇÃO DE IA NA GEOTECNIA
3. COMO SUPERAR ESSES DESAFIOS?

A **Inteligência Artificial (IA)** é um campo da ciência da computação que se concentra no desenvolvimento de **software e máquinas** capazes de **reproduzir o comportamento e ações** que normalmente exigiriam a **inteligência humana**.

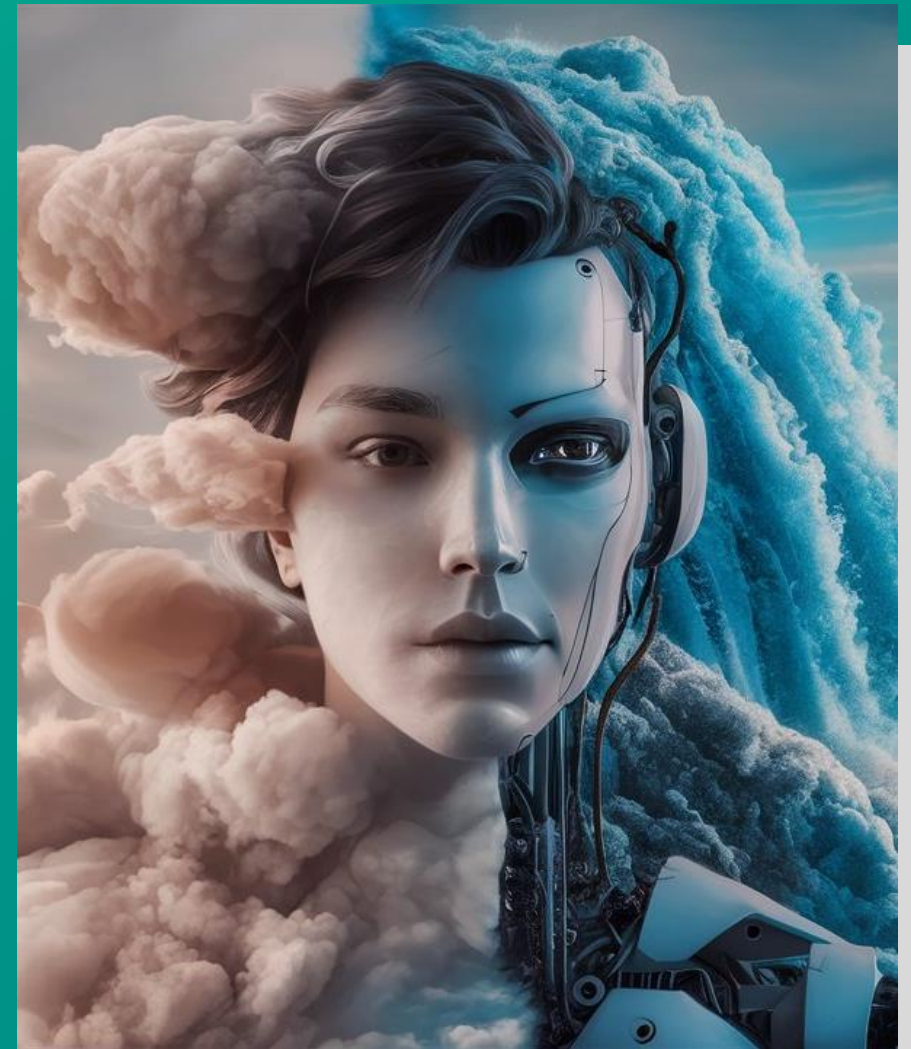
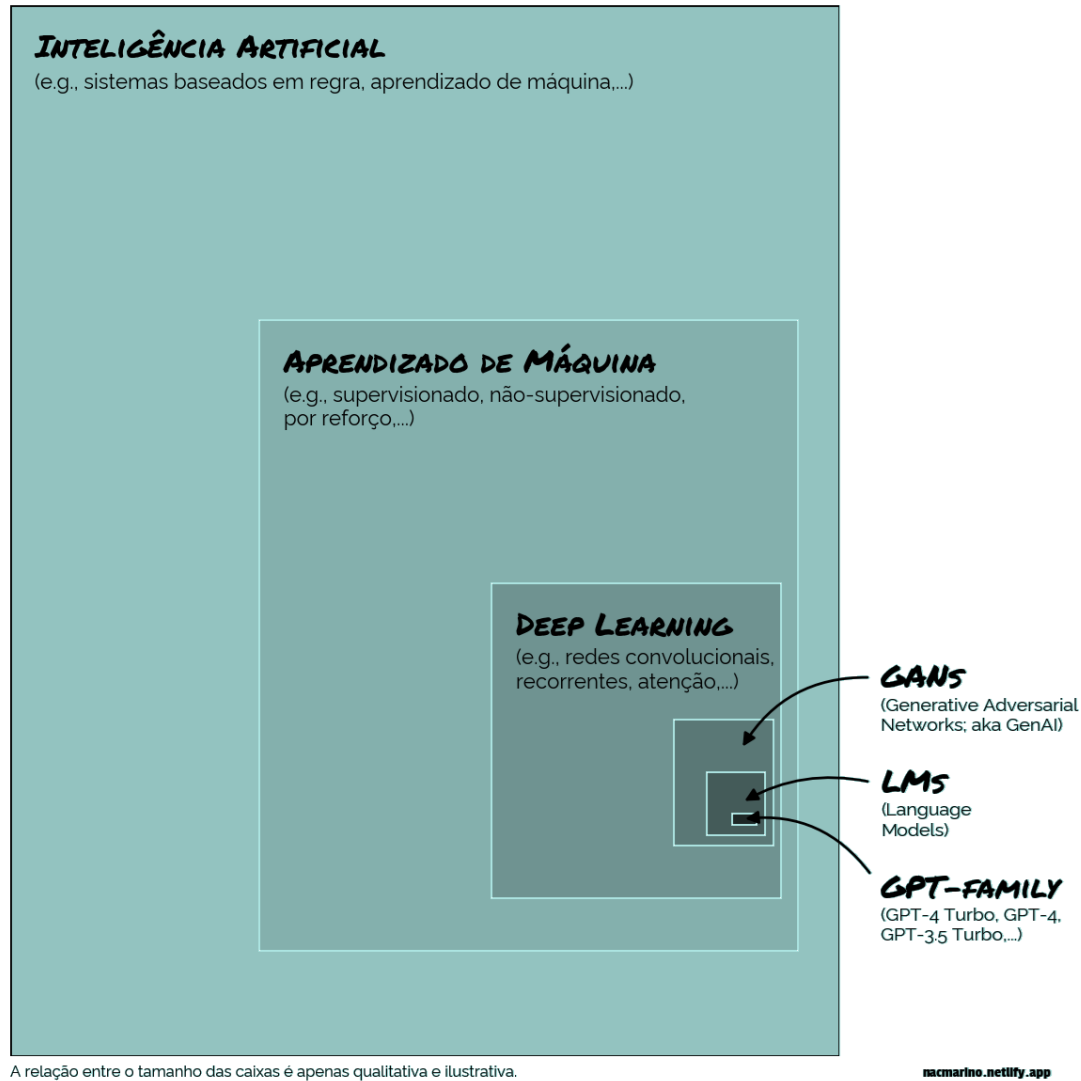


Imagem gerada por IA



DIFERENTES ABORDAGENS CARACTERIZAM ESTE CAMPO, ONDE TÉCNICAS DE **APRENDIZADO DE MÁQUINA** TÊM SE DESTACADO NA FORMA COMO **DADOS E RESULTADOS SE RELACIONAM PARA A IDENTIFICAÇÃO DE PADRÕES**

O **aprendizado de máquina** permite que um sistema aprenda a gerar previsões a partir dos padrões encontrados em um conjunto de dados históricos, possibilitando que o determinismo de **regras pré-definidas** (e muitas vezes não embasadas pela realidade) **sejam substituídas por uma abordagem de tomada de decisão mais direcionada por dados.**



O APRENDIZADO DOS ALGORITMOS SE DÁ PELA **BUSCA ITERATIVA DO CONJUNTO DE PADRÕES QUE MELHOR REPRODUZEM UM RESULTADO A PARTIR DOS DADOS**

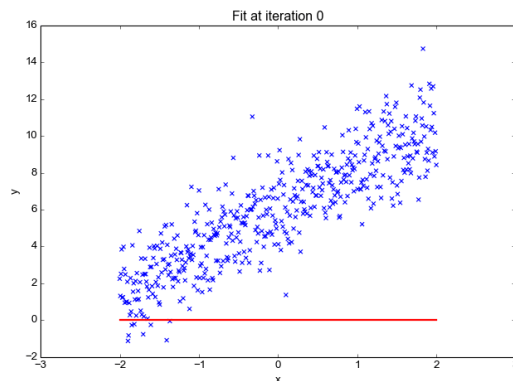
Esta busca é **guiada por algum tipo de métrica** que descreve quão distante o resultado gerado pelo algoritmo está daquele que seria considerado o 'ideal' – o qual depende do **tipo de tarefa de aprendizado do algoritmo e seu propósito final**.

SUPERVISIONADO

O aprendizado do algoritmo é fortemente guiado por **exemplos rotulados**, onde ele aprende a **reproduzir a relação** explícita **entre entradas e saídas** (i.e., dados e resultados).

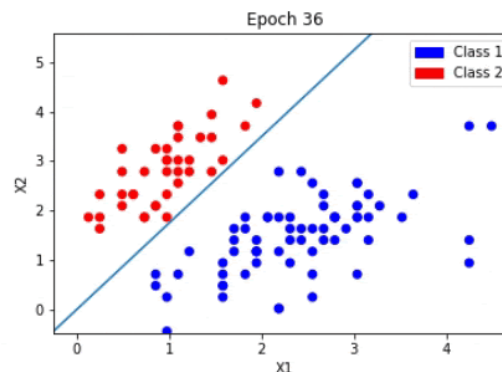
REGRESSÃO

A **saída** que o algoritmo tenta aprender **é um valor numérico**, onde o seu objetivo é encontrar uma 'reta' cuja **distância entre cada valor predito e observado seja a menor possível**.



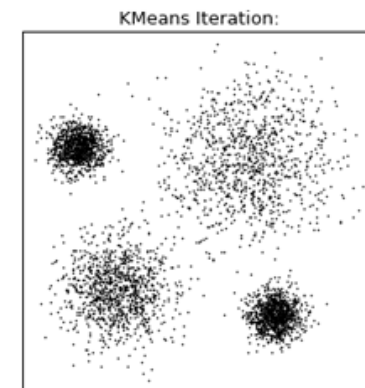
CLASSIFICAÇÃO

A **saída** que o algoritmo tenta aprender **é uma categoria ou probabilidade**, através da identificação de uma **função que melhor separe as categorias** na saída.



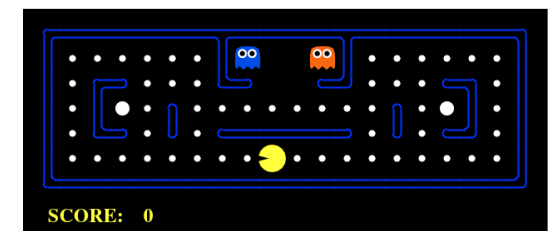
NÃO-SUPERVISIONADO

Não há exemplos rotulados, sendo o aprendizado do algoritmo guiado por **reproduzir a estrutura latente nos dados de entrada** – geralmente, **agrupando observações mais similares entre si** e/ou separando àquelas mais diferentes da norma.



REFORÇO

O objetivo do aprendizado é **identificar uma política que maximize a recompensa acumulada de um agente que interage com um ambiente dinâmico**. Esta política é um conjunto de ações, que são aprendidas através de ajustes baseados em um feedback contínuo de interação entre agente e ambiente.



MAS COMO É QUE ISSO FUNCIONA, AFINAL?

Uma **base de dados** contém o **histórico de informações sobre a entidade** que desejamos analisar, onde deve existir um conjunto de informações sobre as **características (i.e., features)** de cada **observação** bem como, **no caso do aprendizado supervisionado, um ou mais rótulos** que definem aquilo que o algoritmo deverá aprender a prever. Apresentamos aqui um **exemplo de classificação**.

Ponto	Evento	Deformação	Slope	Terreno
1	Sim	200 mm	60°	A
2	Não	30 mm	10°	A
3	Sim	100 mm	40°	B
4	Não	100 mm	40°	A
...

INSTÂNCIA
Um identificador do pontos no espaço, indivíduos, tempo, imagens, livro, texto,..

RESPOSTA (Y)
Rótulo associado à observação (categoria ou número)

CARACTERÍSTICAS OU FEATURES (X)
Conjunto de informações que descreve cada um dos aspectos de interesse da unidade observacional

Alguns algoritmos, como as regressões e redes neurais, utilizam uma soma ponderada dos valores de cada características para fazer uma previsão sobre o rótulo. Outros algoritmos, como árvores de decisão, utilizam critérios diferentes para estabelecer esses 'pesos'.

X

PESOS

Deformação	Slope	Terreno
200 mm	60°	A
30 mm	10°	A
100 mm	40°	B
100 mm	40°	A
...

*

$\beta_{\text{Deformação}}$
β_{Slope}
β_{Terreno}

=

Previsão	Evento	Δ
0.84	1	0.16
0.2	0	-0.2
0.6	1	0.4
0.1	0	-0.1
...

O objetivo do algoritmo é encontrar um conjunto de pesos que façam com que o erro de previsão seja (quase) zero. Esta busca é feita de forma iterativa, por exemplo, usando o gradiente da derivada do erro para ajustar os valores dos pesos.

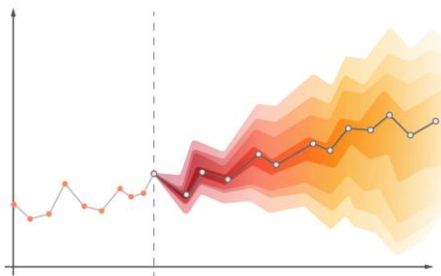
...que são geralmente transformados em variáveis binárias quando são categóricos: Evento = Sim \rightarrow 1; Evento = Não \rightarrow 0.

O erro de previsão é a diferença entre o que o algoritmo deveria ter previsto e o que ele previu, e é uma medida de sua performance.

O output do algoritmo de classificação são probabilidades, que definem o quanto cada instância está relacionada à cada rótulo.

Estas previsões são comparadas com os rótulos...

ÀQUELAS TÉCNICAS DE APRENDIZADO SÃO A BASE DE DIFERENTES TIPOS DE APLICAÇÃO DE IA



APRENDIZADO DE MÁQUINA (GERAL)

Tarefas de previsão associados a dados diferentes de imagens e textos, bem como simulação e otimização matemática.

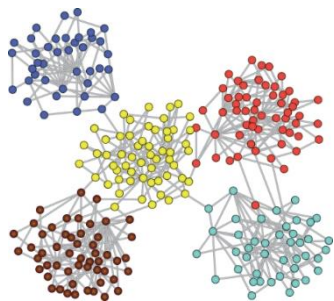
- Previsão de produção de minério no próximo dia;
- Estimativa da emissão de particulados;
- Otimização de rotas;
- Simulação de cenários.



VISÃO COMPUTACIONAL

Aplicações associadas à aquisição de informação através da análise de imagens ou vídeos.

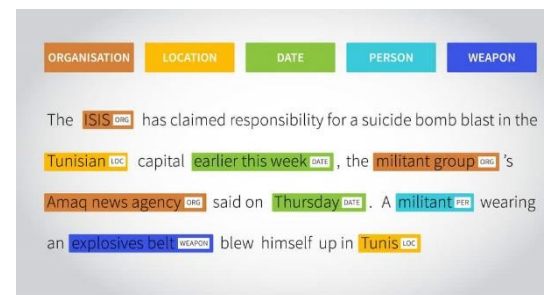
- Classificação de imagens;
- Detecção de objetos, pessoas, itens,...;
- Identificação ou reconhecimento facial;
- Reconhecimento ótico de caracteres.



AGRUPAMENTO DE ENTIDADES

Aplicações onde precisamos identificar grupos de entidades mais similares entre si do que as demais.

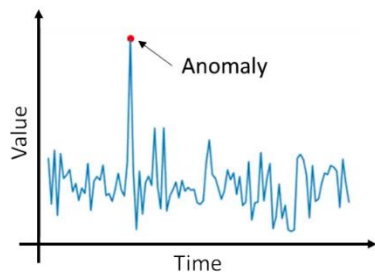
- Mapeamento de grupos de usuários;
- Itens com especificação similares;
- Grupos de municípios com condição socio-econômica similar;
- Agrupamento de séries temporais com dinâmicas similares.



PROCESSAMENTO DE LINGUAGEM NATURAL

Aplicações associadas à análise, interpretação, compreensão e/ou tradução de texto.

- Identificação e extração de entidades de um texto;
- Classificação de texto;
- Tradução de texto;
- Reconhecimento de fala.



DETECÇÃO DE ANOMALIAS

Aplicações onde buscamos identificar comportamentos anômalos em um fenômeno.

- Identificação de valores aberrantes;
- Mudanças inesperadas no comportamento de um instrumento;
- Identificação de novos comportamentos de um instrumento.



IA CONVERSACIONAL

Aplicações orientadas ao desenvolvimento de sistemas capazes de manter uma conversa com seres humanos. Atualmente, sinônimo de IA Generativa.

- Chatbots;
- Assistentes virtuais.

ESTAS APLICAÇÕES PODEM TER PROPÓSITOS DISTINTOS

Todavia, o que todas elas têm em comum é que oferecem uma forma de realizar **análises quantitativas** dos dados de uma forma **sistemática** e **objetiva**



DESCRITIVO

Fornecer insights sobre aquilo que ocorreu no passado.

"Quantas anomalias ocorreram na estrutura?"



DIAGNÓSTICO

Busca entender o porquê de algo ter ocorrido.

"Por que o instrumento X varia tanto quando chove?"



PREDITIVO

Tentar prever o que pode acontecer no futuro de acordo com os padrões do passado.

"Qual será a leitura do instrumento X amanhã, quando chover?"



PRESCRITIVO

Recomendar o melhor conjunto de ações para se atingir um objetivo.

"Como otimizar a malha de monitoramento?"

QUANDO E POR QUÊ UTILIZAR INTELIGÊNCIA ARTIFICIAL?

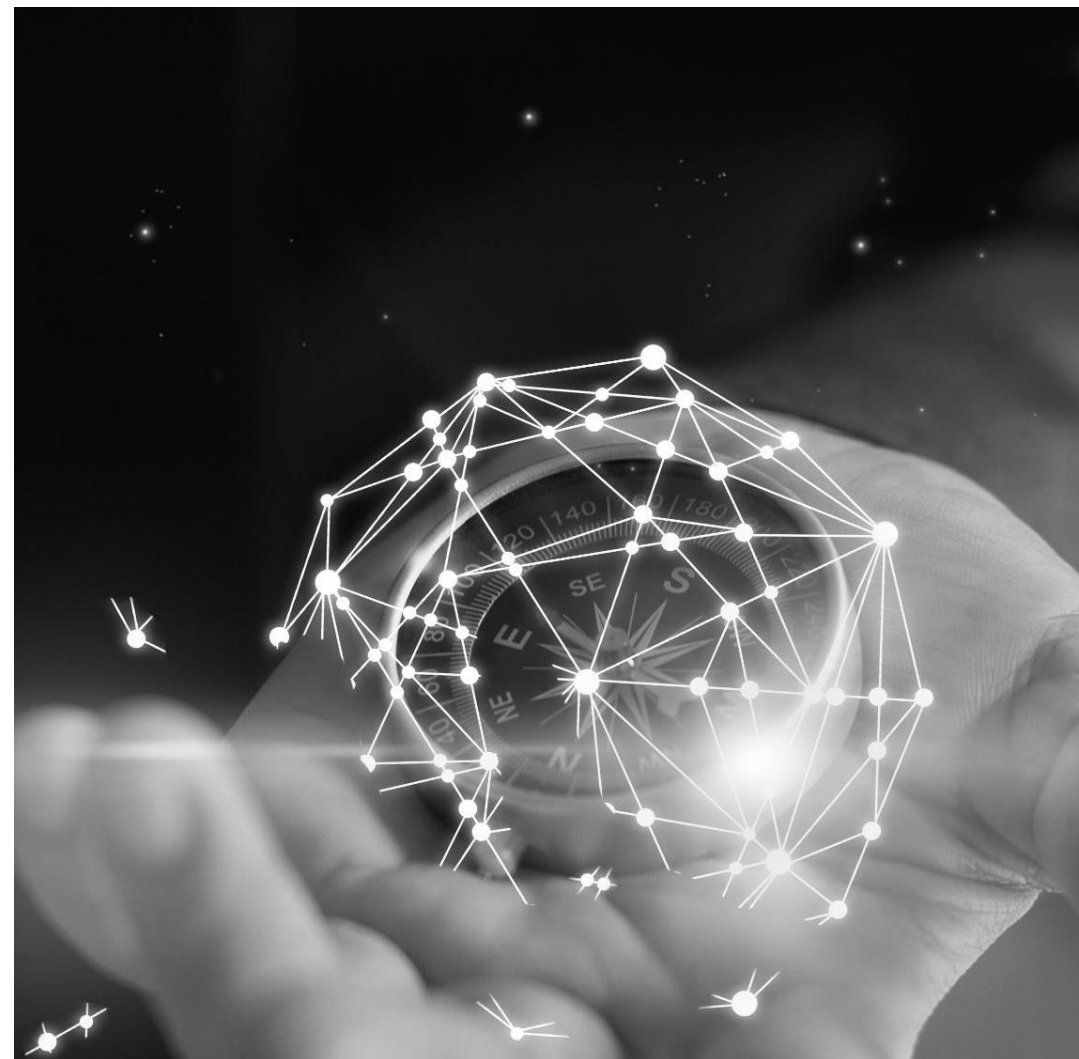
No geral, **qualquer tarefa que seja fácil de ser executada por um humano** pode ser reproduzida através de um algoritmo de inteligência artificial. No entanto, **a implementação desta tecnologia** faz mais sentido quando ela pode trazer pelo menos um de dois **benefícios para o negócio**.

MELHORIA DE PROCESSO

- Tempo de execução (e.g., tempo de análise)
- Execução inteligente (e.g., manutenção)
- Otimização do processo (e.g., automação)

AUMENTO DE MARGEM

- Aumento de produção (e.g., fazer mais com menos)
- Redução de desperdício (e.g., gestão de inventário)
- Produção inteligente (e.g., produção ~ demanda)



A INTELIGÊNCIA ARTIFICIAL REPRODUZ O COMPORTAMENTO HUMANO – TANTO O QUE HÁ DE BOM, QUANTO O QUE HÁ DE RUIM

Os **dados** que alimentam estes algoritmos **contém viéses e erros históricos de nossa sociedade e decisões**. Portanto, **devemos tomar ações intencionais para projetar, implementar e utilizar a IA para trazer seus benefícios de forma responsável**, criando mecanismos de proteção, correção e mitigação dos potenciais riscos associados à ela. Alguns princípios são fundamentais aqui.



CENTRADO NO SER HUMANO

Entender o impacto humano sempre que uma solução de IA for implementada.



JUSTO

Tratamento equitável para todos os grupos, com ações para mitigar ou evitar impactos de viéses.



TRANSPARENTE, EXPLICÁVEL E ACURADO

Divulgar o uso de IA quando apropriado, todos devem ser capazes de entender e avaliar adequadamente as decisões da IA e os processos de tomada de decisão.



SEGURO

Avaliar potenciais preocupações de segurança e tomar medidas para mitigar danos.



RESPONSABILIDADE

Documentar as estruturas de governança, com papéis, políticas e responsabilidades claras.



CONFORMIDADE, PRIVACIDADE DE DADOS E CIBERSEGURANÇA

Todo uso de IA deve estar em conformidade com as leis relevantes; proteção dos dados com medidas apropriadas de privacidade e segurança contra ataques cibernéticos.



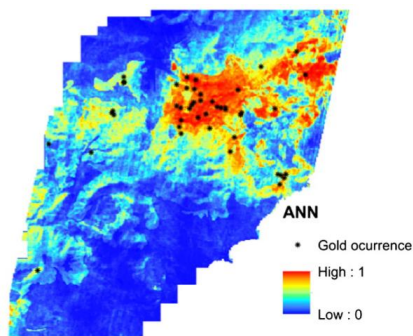
SUSTENTABILIDADE

Implementar IA levando em consideração o impacto no planeta, com medidas apropriadas tomadas para mitigar impactos negativos.

ALGUNS GRANDES TEMAS DOMINAM A LITERATURA ACADÊMICA E APLICAÇÕES NA INDÚSTRIA

PROSPECÇÃO MINERAL

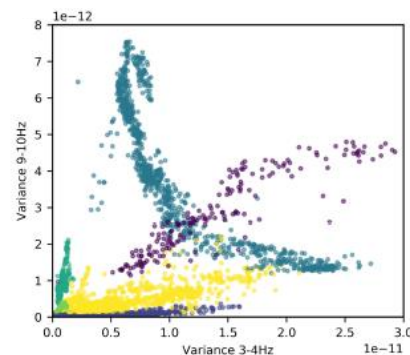
Utilização de dados de levamentos geofísicos e mapeamentos geológicos para a prospecção, amostragem e geração de modelos.



Rodriguez-Galiano et al, 2015, Ore Geol Rev

COMPORTAMENTO DE VULCÕES

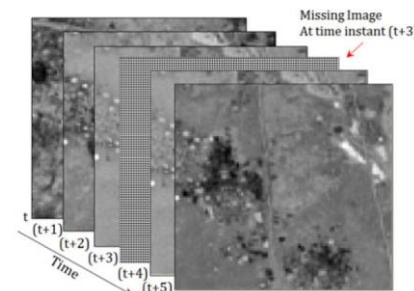
Utilização de informações sísmicas e/ou padrões de deformação do solo para a previsão de erupções.



Ren et al, 2020, Geo Res Lett

SENSORIAMENTO REMOTO

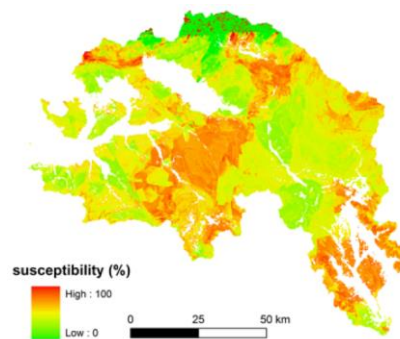
Segmentação, detecção de mudanças, classificação e previsão de eventos baseado principalmente em imagens de satélite.



Das et al, 2020, Appl Earth Obs Rem Sens

MAPAS DE SUSCEPTIBILIDADE

Uso de características do terreno, paisagem e eventos georreferenciados para a previsão de novas áreas susceptíveis.



Ruiz et al, 2019, Nat Haz; Rahmati et al, 2019, Sci Tot Env; Catani et al, 2013, Nat Haz Earth Sys Sci

**Quais os algoritmos
mais comumente
utilizados nestes
trabalhos?**

REDES NEURAIS ARTIFICIAIS E SUAS VARIANTES

Algoritmos com alta flexibilidade e poder preditivo, adaptados para utilizar dados tabulares, mas também imagens (CNNs) e dados sequenciais (RNNs e LSTMs).

RANDOM FOREST E ALGORITMOS DE ÁRVORE

Aprendem uma estrutura de árvore de decisão com alto poder preditivo; além disso, sua arquitetura os torna mais facilmente explicáveis.

ALGORITMOS DE AGRUPAMENTO (KMEANS)

Colocam coisas similares dentro de um mesmo grupo, sendo muito utilizados para a mineração de dados e/ou segmentação 'automática' de áreas de estudo.

SE EXISTE TANTO DESENVOLVIMENTO TÉCNICO NO CAMPO DE IA E ALGUMAS APLICAÇÕES JÁ VALIDADAS NA LITERATURA, ENTÃO POR QUÊ...

...ainda existem alguns *gaps* na aplicação desta tecnologia em algumas áreas da Geotecnia?

...encontramos dificuldade em implementar e escalar algumas das aplicações já existentes?



An aerial photograph of a large dam and reservoir. The reservoir is a calm body of water, surrounded by dense green forest. The dam is a long, straight structure with a concrete spillway on the left side. The surrounding landscape is hilly and covered in thick vegetation.

MELHORAR NOSSA PERSPECTIVA SOBRE

O que de fato é uma análise de dados?

FRASES PARA PENSAR SOBRE

O piezômetro subiu...

Os ensaios não são representativos...

Não acredito nos dados...

A ETR está com tendência...

O instrumento não é confiável...

O dado é um outlier, pode deletar...

EXISTEM PELO MENOS 5 OBSTÁCULOS QUE PRECISAMOS SUPERAR PARA MUDAR ÀQUELAS PERSPECTIVAS, ALAVANCAR IA NA GEOTECNIA E CAPTURAR SEU VALOR...

DISPONIBILIDADE E QUALIDADE DOS DADOS

Ter um conjunto de dados não quer dizer que você têm os dados necessários para uma solução de IA/ML.

SEMÂNTICA DOS FENÔMENOS

Ter clareza na definição do fenômeno estudado é fundamental para que a solução tenha aderência ao problema.

AValiação DAS SOLUÇÕES

Saber estabelecer e implementar critérios objetivos para definir como avaliar a qualidade de uma solução.

CONHECIMENTO SOBRE IA E ML NA ÁREA

Em um momento de *hype*, é necessário saber separar moinhos de vento de gigantes.

VISÃO SISTÊMICA DE UMA SOLUÇÃO DE IA

Entender que a IA não faz nada sozinha: ela é um componente de um sistema muito maior.

DISPONIBILIDADE E QUALIDADE DOS DADOS

É necessário que haja uma **fundação de dados robusta** para sustentar aplicações de Inteligência Artificial e Aprendizado de Máquina. Esta fundação **deve garantir a disponibilidade, consistência e veracidade dos dados. A ausência dela**, no melhor dos cenários, faz com que as **soluções não passem da etapa de PoC**.

PROBLEMAS

Dados em planilhas Excel ou PDFs

Geralmente esporadicamente, *offline*, estáticos. Podem ser difíceis de ser consumidos diretamente por aplicações de IA/ML.

Dados com fornecedores

Dificuldade de acesso e risco de perda de dados históricos por políticas dos próprios fornecedores.

Liberdade de edição de registros

Registros são livremente modificados com pouca ou nenhuma governança do porquê ou quem os modificou.

Histórico sobre os dados com as pessoas

Entendimento do histórico, porquê de alguns comportamentos nos dados e etc com quem conhece a estrutura.

Múltiplas fontes 'da verdade'

Risco de inconsistências, principalmente se houverem diferenças na frequência de atualização e tratamentos aplicados aos dados.

Repositórios pensados para armazenar dados

Silos de dados com acesso pensado apenas para seres humanos, que não sustentam requisições massivas de dados.

SOLUÇÕES POSSÍVEIS

Banco de Dados ou Data Hub

Dados atualizados em tempo real, com uma estrutura previsível e consistente ao longo do tempo.

Dados dentro de casa

Plena governança dos dados: onde estão, quem pode acessar, quando movimentar para arquivo e etc.

Governança de edição de registros

Múltiplas camadas de tratamento e governança, garantindo que toda a linhagem do dado esteja disponível e auditável.

Histórico sobre os dados documentado

Explicações e documentações sobre o comportamento dos dados disponível em um repositório central e acessível.

Repositório de dados centralizado

Uma única fonte de dados alimenta todas as aplicações (analíticas ou não), com tratamentos customizados para cada uma delas.

Repositórios também para tarefas analíticas

Estruturas de armazenamento que sustentam *workloads* massivos de dados, com acesso disponibilizado através de APIs.

SEMÂNTICA DOS FENÔMENOS

Uma solução de IA geralmente foca em resolver um problema específico, e precisa dos dados que **descrevem este problema** para ser desenvolvida e implementada. Por exemplo, o que seria uma solução de detecção de anomalias? O que é uma anomalia? Que tipo de anomalias ela deveria detectar? A **má definição** desse tipo de entidade **e/ou a indisponibilidade dos dados que a descrevem** podem **torná-la inviável**.

“ANOMALIA” GEOTÉCNICA



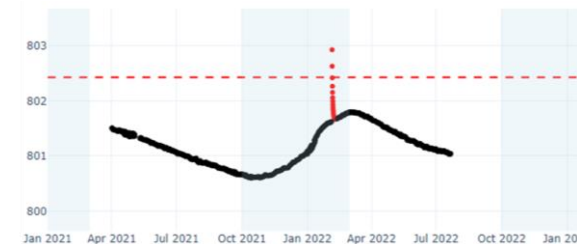
Uma fissura, deformação ou qualquer fenômeno que ocorra diretamente em uma estrutura e que deve ser refletida no monitoramento

“ANOMALIA” NOS DADOS



ERRO NA CAPTURA DOS DADOS

Valores discrepantes e isolados, associado à problemas no registro e/ou erros de processamento de dados.



VARIAÇÃO ‘NATURAL’

Comportamentos incomuns ou fora do padrão, mas explicáveis – normalmente associados à intervenções na estrutura e/ou instrumentação



NOVIDADES OU MUDANÇAS

Mudanças na distribuição dos dados, para padrões nunca antes observados – com pouca clareza da origem deles.

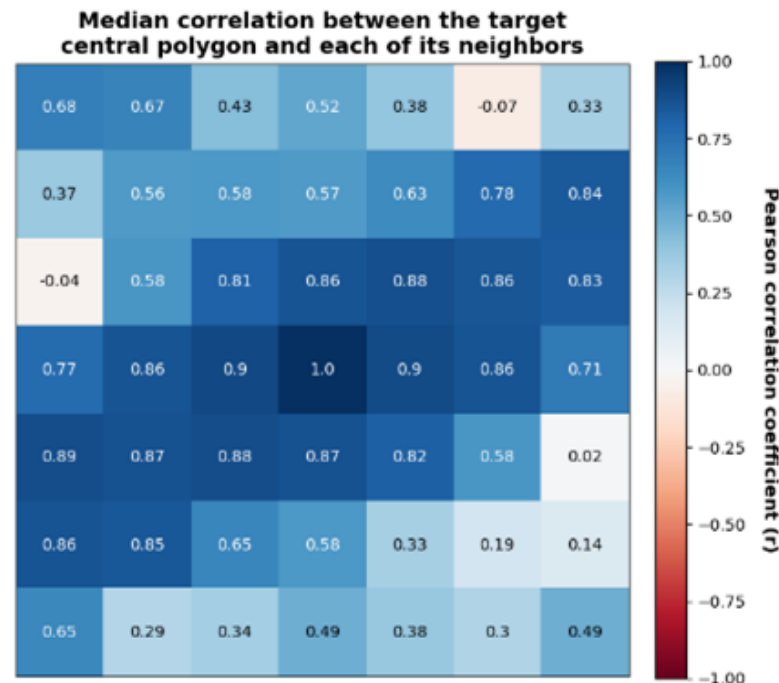
AVALIAÇÃO DAS SOLUÇÕES

Soluções de IA/ML podem memorizar padrões, sem aprendê-los

Se este processo de **memorização e vazamento de dados não forem ativamente combatidos**, observaremos excelentes resultados da solução no conjunto de dados que a gerou, mas **um péssimo desempenho em novos conjuntos de dados**. Infelizmente, é muito fácil se deixar levar por bons resultados e falhar na **elaboração de procedimentos robustos de avaliação dessas soluções** – o que exige conhecimento funcional sobre o problema e técnico sobre IA/ML.

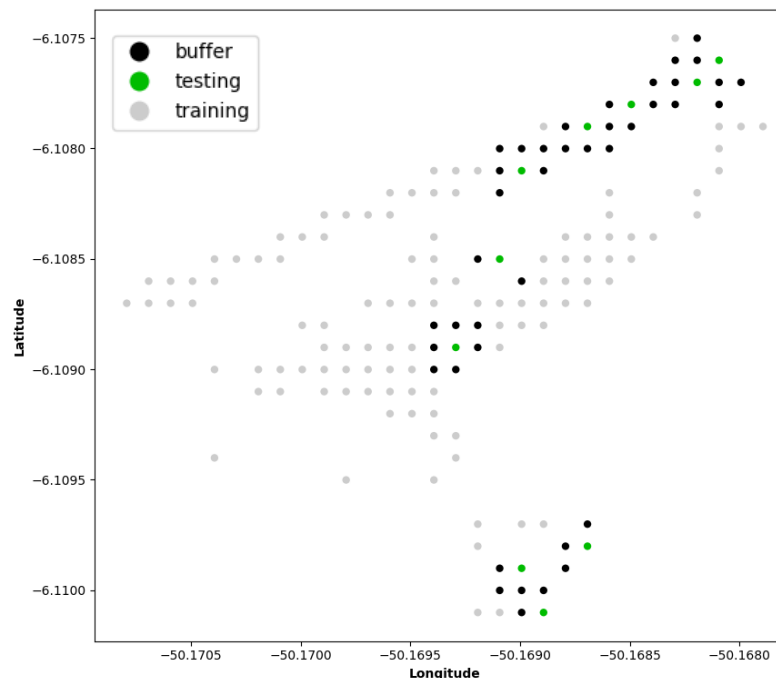
VAZAMENTO DE DADOS

Processo sutil, onde a solução tem acesso, direta ou indiretamente aquilo que ela precisa prever. A auto-correlação espacial e temporal podem levar ao vazamento de dados, bem como erros na preparação dos dados e falta de entendimento sobre o processo de geração dos mesmos.



AUTO-CORRELAÇÃO ESPACIAL EM DADOS DE INSAR

Áreas mais próximas tendem a ser mais similares: basta o algoritmo copiar um padrão de um quadradinho para o do lado, sem aprender nada, que ele acertará



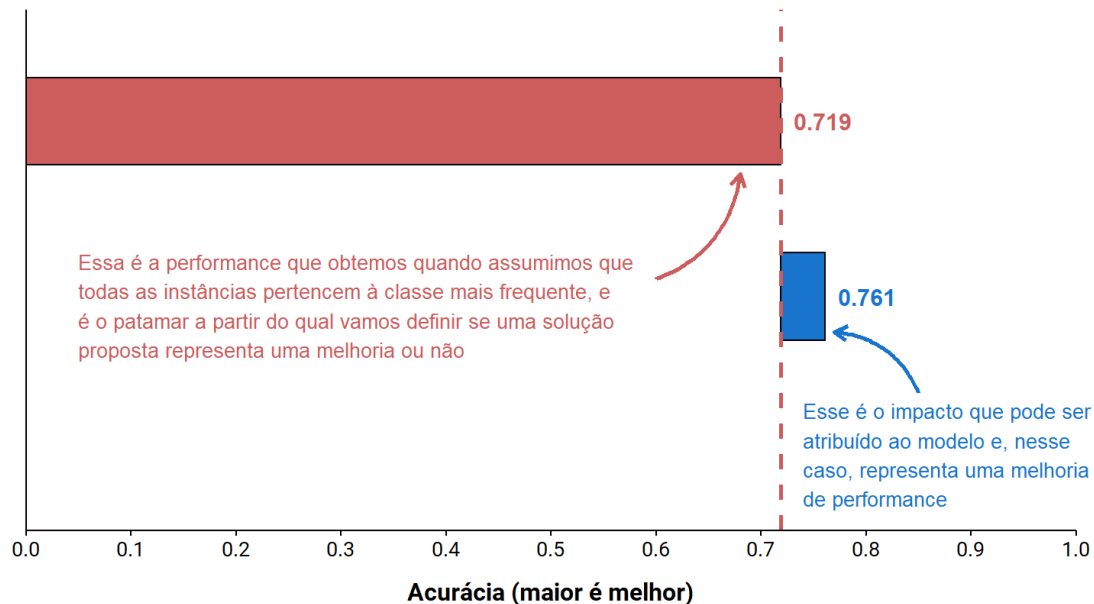
Pontos de monitoramento na Barragem Jacaré

SOLUÇÃO IMPLEMENTADA PARA AVALIAR O ALGORITMO

Isolamos alguns pontos para o teste do algoritmo, evitando que ele tenha visto a vizinhança destes durante o seu treinamento.

Avaliação da solução preditiva

Medimos o impacto do **modelo** comparando a sua performance com a do **baseline**



A SOLUÇÃO DEVE SUPERAR UM BASELINE PARA SER JUSTIFICADA

O exemplo abaixo ilustra um caso onde 71.9% das observações pertencem à uma única categoria, e mostra o poder que o 'chute' pode ter. Já observamos casos em que a performance do algoritmo foi inferior à uma regra de negócio ou solução mais simples, mas isso não foi notado por quem desenvolveu.

PREVISÕES EM RETROSPECTO NÃO SÃO REALMENTE PREDITIVAS

Já observamos muitas análises em retrospecto serem vendidas como previsões, bem como soluções cujas informações necessárias não estariam disponíveis no momento em que a previsão precisava ser feita.

CONHECIMENTO SOBRE IA E ML NA ÁREA

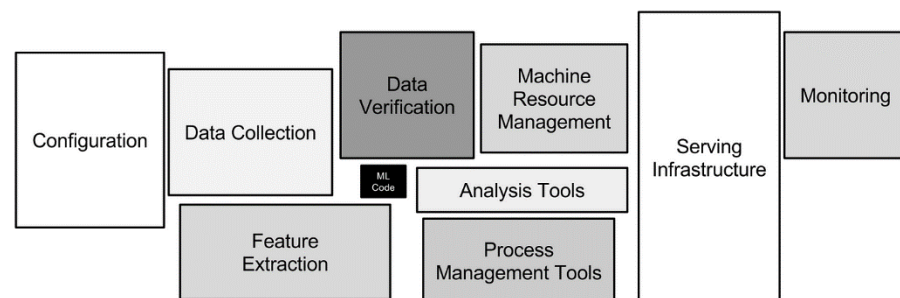
A implementação de uma solução de IA/ML deve seguir um processo científico rigoroso

É pouco produtivo investir em tais soluções caso haja evidência de que elas não serão capazes de trazer as melhorias desejadas e/ou que prometem. No entanto, **a falta de conhecimento** sobre estas tecnologias pode **dificultar a crítica sobre a solução apresentada**. Além disso, existem casos em que uma boa análise de dados já pode ser o suficiente para resolver um problema.

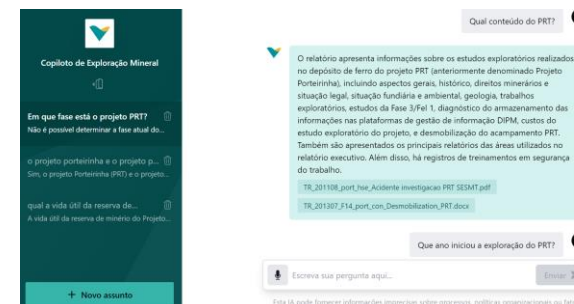
Ter uma solução de IA/ML por ter não é justificativa para a sua implementação.

VISÃO SISTÊMICA DE UMA SOLUÇÃO DE IA

Uma solução de IA, como um modelo de aprendizado de máquina, **é só um componente de um sistema muito maior**. É preciso considerar a totalidade deste sistema para gerar engajamento e trazer o valor que ela foi projetada para ter: **pessoas, processos e tecnologia**.



Diversos outros recursos de tecnologia precisam estar no lugar para que uma solução de IA possa ser usada fora do computador da pessoa que a desenvolveu de forma escalável, sustentável, segura e auditável.



Uma solução de IA também **precisa entrar nos processos onde elas foram desenhadas para atuar**, o que depende da forma como ela vai ser disponibilizada e de sua **adoção das pelas pessoas e lideranças**.

O QUE PODE SER FEITO PARA SUPERAR AQUELES OBSTÁCULOS?

TIMES MULTIDISCIPLINARES

A área de IA é multidisciplinar por natureza, e combinar a diversidade de conhecimentos em um time é uma solução que se mostra efetiva em nossa vivência.

CAPACITAÇÃO TÉCNICA

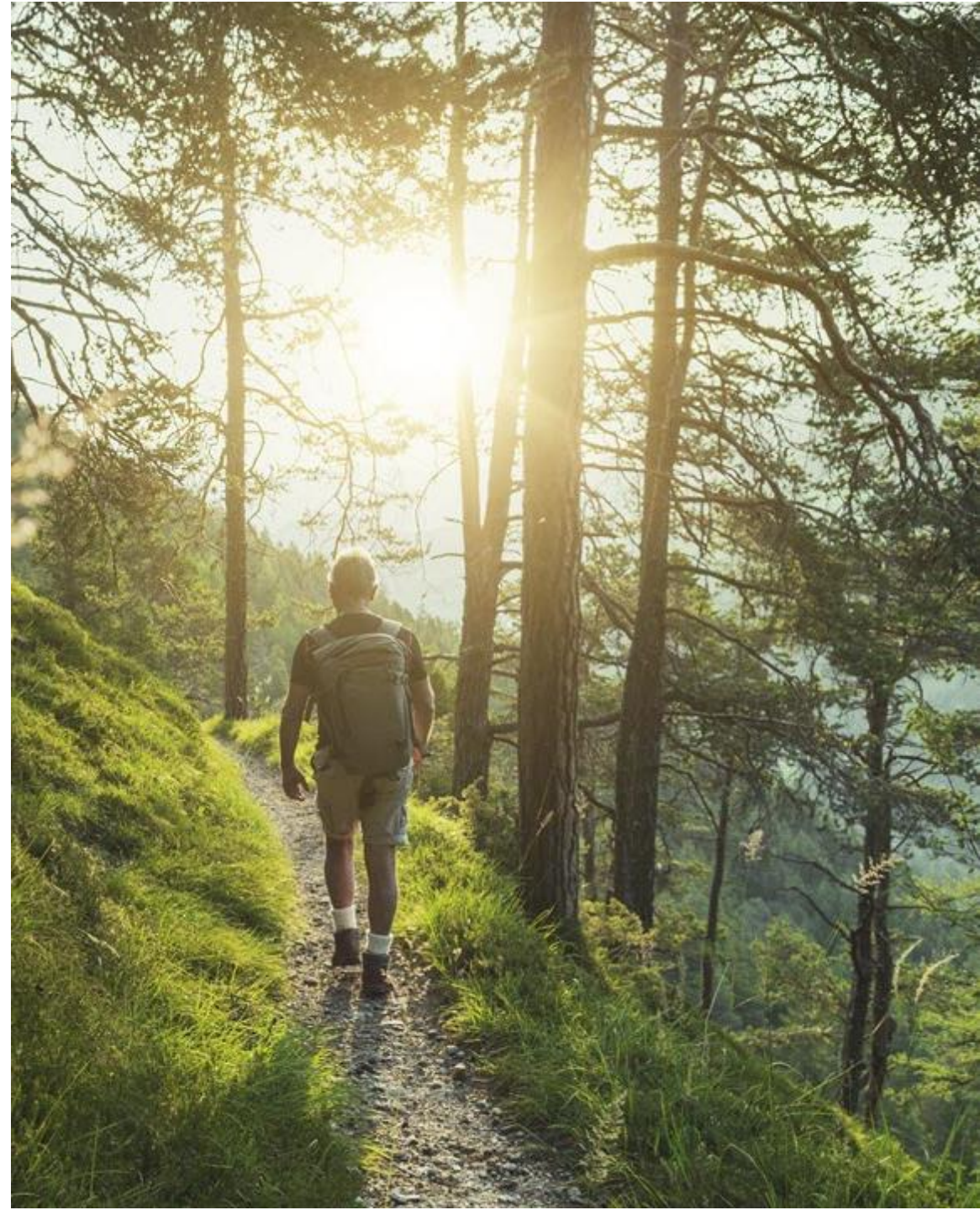
Trazer literacia em dados e IA é importante para ampliar perspectivas, dar mais autonomia e transformar a cultura para àquela orientada à dados.

FERRAMENTAS OPEN SOURCE

A dependência de soluções caixa preta ou específicas de fornecedores externos dificulta e limita a criatividade para a solução de problemas. Além disso, uma diversidade de soluções abertas podem ser implementadas com baixo custo.

BUSCAR O CONHECIMENTO DENTRO DE CASA

A empresa oferece opções que podem acelerar essa jornada de transformação e adoção de IA na área da Geotecnia.



“Embora a aplicação de Inteligência Artificial na Geotecnia apresente desafios significativos, como a necessidade de dados de alta qualidade, a complexidade dos modelos e a interpretação dos resultados, as oportunidades para inovação e avanço no campo são imensas. Superar essas barreiras não só impulsionará a precisão e a eficiência dos projetos geotécnicos, mas também contribuirá para a construção, monitoramento e gestão de estruturas de forma mais segura e sustentável.”

Resposta do ChatGPT ao prompt: “Preciso terminar uma apresentação que fala sobre os desafios do uso de Inteligência Artificial na Geotecnia. Sugira alguma frase ou algo do gênero para o slide de conclusão”. Realizamos apenas pequenos ajustes e adições, mas 95% do texto é dele.