**Nicolas Gustavo Magrini Pereira**

**Ciência de Dados**

A ciência de dados combina matemática, estatística, programação especializada, análise avançada, inteligência artificial (IA) e machine learning com conhecimento em assuntos específicos ao projeto para descobrir insights práticos, que são descobertas obtidas através da análise de dados e que podem ser aplicadas para resolver determinados problemas.

O volume crescente de fontes de dados tornou a ciência de dados um dos campos/áreas que mais crescem dentre todos os setores em Tecnologia da Informação (TI). Por conta disso, as organizações dependem cada vez mais dos cientistas de dados para interpretar o que foi coletado e fornecer recomendações para melhorar os resultados de negócios da empresa. Conclui-se isto levando em conta a proliferação de dispositivos que podem coletar e armazenar informações automaticamente.

**Ciclo de vida da ciência de dados**

Envolve várias funções, ferramentas e processos. Normalmente, um projeto de ciência de dados passa pelas seguintes etapas:

* **Ingestão de dados:** coleta de dados não processos, estruturados ou não, de todas as fontes relevantes por diversos métodos. Esses métodos incluem entrada manual, dados de streaming em tempo real etc. As fontes de dados podem incluir dados estruturados como dados do cliente, junto com dados não estruturados como vídeos, fotos, redes sociais, entre outros;
* **Armazenamento e processamento de dados:** as empresas consideram diferentes sistemas de armazenamento baseando-se nos tipos de dados a serem capturados. As equipes de gerenciamento de dados ajudam a definir os fluxos de trabalho para modelos de análise, machine learning e deep learning. Esta fase inclui a limpeza, desduplicação, transformação e combinação de dados por meio de tarefas **ETL** (extração, transformação, carregamento);
* **Análise de dados:** são examinados padrões, intervalos e distribuições de valores nos dados. Esta descoberta por análise guiará a geração de hipóteses para testes a/b. Ela também permite que os analistas determinem a relevância dos dados para uso dentro dos esforços de modelagem. Dependendo da precisão de um modelo, as organizações podem se tornar dependentes desses insights para a tomada de decisões de negócios
* **Comunicação:** os insights são apresentados como relatórios e outras visualizações de dados que simplificam a sua compreensão, além de seus impactos nos negócios, para analistas de negócios e outros tomadores de decisões. Uma linguagem de programação de ciência de dados, como R ou Python, inclui componentes para gerar visualizações; como alternativa, os cientistas de dados podem usar ferramentas de visualização dedicadas.

**Benefícios da ciência de dados**

* **Descobrir padrões transformadores desconhecidos:** a descoberta de padrões tem potencial para transformar uma organização. Como exemplo, podem ser feitas alterações de baixo custo no gerenciamento de recursos, o que consegue maximizar o lucro de uma empresa;
* **Inovar produtos e soluções:** falhas e problemas podem ser revelados graças a ciência de dados. Mais insights sobre decisões de compra, feedback de clientes e processos de negócios podem impulsionar a inovação em operações internas e soluções externas.
* **Otimização em tempo real:** É muito desafiadora para as empresas, especialmente as de grande porte, responder às mudanças nas condições em tempo real. Isso pode causar perdas significativas ou interrupções na atividade empresariais. A ciência de dados pode ajudar as empresas a preverem mudanças e reagir de maneira ideal a diferentes circunstâncias.

**Relação entre Data Science, Machine Learning e Inteligência Artificial**

Como dito anteriormente, o Machine Learning é uma ferramenta para a análise de dados em ciência de dados, pois esta possui conhecimentos específicos ao assunto sobre o qual o projeto é voltado.

Já, a Inteligência Artificial é “mãe” do Machine Learning e uma área complementar à Ciência de Dados, pois uma empresta conceitos e ferramentas da outra, principalmente as empregadas em Machine Learning. Ao mesmo tempo que fornecem informações para o aprendizado de máquina, os cientistas de dados recorrem a recursos de processamento e padronização de IA para desenvolverem seu trabalho.

Para exemplificar a relação entre ciência de dados e inteligência artificial, imagine a descrição de uma vaga para cientista de dados e outra para um analista de IA. Na primeira, será exigido do profissional de TI que conheça ferramentas de IA para classificar dados e realizar previsões, como algoritmos de deep learning.

Na segunda, provavelmente o título do cargo revelará a diferença entre o trabalho que será realizado pelo candidato e por um cientista de dados: Deep Learning Scientist, Machine Learning Engineer, PNL Scientist e por aí vai. Mesmo assim, quem preencher a vaga provavelmente usará ferramentas de ciência de dados no dia a dia, como R e Python.

**Referências**

IBM. **O que é ciência de dados?** Disponível em: https://www.ibm.com/br-pt/topics/data-science. Acesso em: 24 set. 2023.

AMAZON. **O que é ciência de dados?** Disponível em: https://aws.amazon.com/pt/what-is/data-science/. Acesso em: 24 set. 2023.

SILVEIRA, Paulo. **Qual a diferença entre Data Science, Machine Learning e Inteligência Artificial?** Disponível em: https://www.alura.com.br/artigos/qual-e-a-diferenca-de-data-science-machine-learning-e-inteligencia-artificial. Acesso em: 24 set. 2023.

BALDISSERA, Olívia. **Ciência de dados e inteligência artificial: combinação em alta no mercado de tecnologia.** Disponível em: https://posdigital.pucpr.br/blog/ciencia-de-dados-e-inteligencia-artificial. Acesso em: 24 set. 2023.

FERNANDES, Leandro Carlos. **Ciência de dados, inteligência artificial e aprendizado de máquina são a mesma coisa?** Disponível em: https://www.linkedin.com/pulse/ci%C3%AAncia-de-dados-intelig%C3%AAncia-artificial-e-m%C3%A1quina-s%C3%A3o-fernandes/?originalSubdomain=pt. Acesso em: 24 set. 2023.