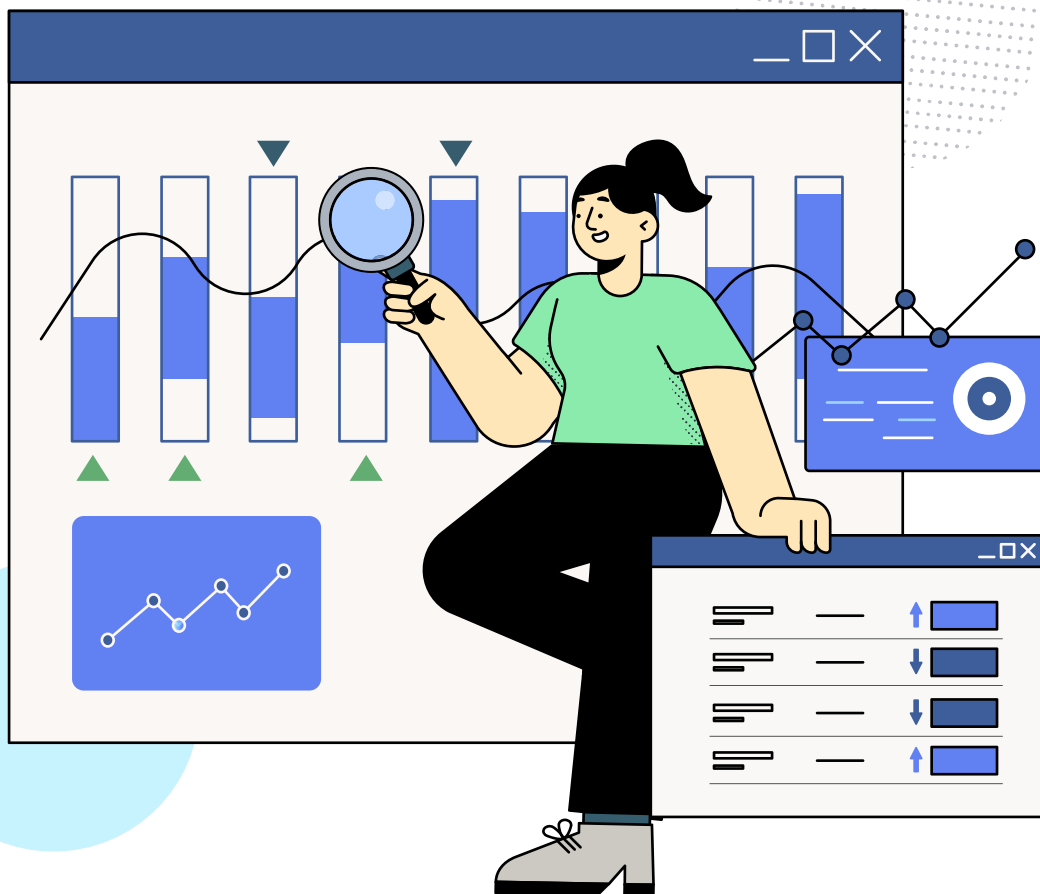


CARREIRA EM DADOS

Para quem busca uma transição ou está ingressando no mercado, as posições associadas ao Big Data e Ciência dos Dados oferecem grande projeção profissional. **Saiba mais**



SUMÁRIO INTERATIVO

| | |
|--|-----------|
| Introdução | 3 |
| Capítulo 1 - Histórico | 4 |
| Box 1 - Principais termos relacionados a carreiras de dados | 4 |
| Capítulo 2 - Mercado Aquecido | 7 |
| Box 2 - Conheça as principais ferramentas de análise de dados | 8 |
| Box 3 - Quem é quem no mundo dos dados? | 8 |
| Box 4 - Saúde: uma oportunidade crescente | 10 |
| Capítulo 3 - Transição e Evolução | 11 |
| Box 5 - Já pensou em trilhar uma carreira em dados? | 12 |
| Box 6 - Quanto ganha um profissional de dados? | 13 |
| Box 7 - De olho nas oportunidades | 13 |
| Capítulo 4 - O futuro das carreiras de dados | 16 |

Introdução

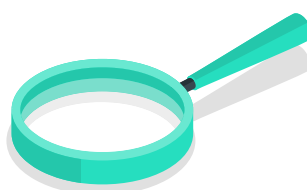
A evolução das inteligências artificiais, mais recentemente a inteligência generativa, tem demonstrado como os dados abertos podem ajudar as empresas. Essas IAs generativas, como o ChatGPT, são capazes de criar textos e imagens, por meio de machine learning (aprendizado de máquina, em inglês), um método de análise de dados que automatiza a construção de modelos analíticos.

Todo esse potencial criativo promovido pelos algoritmos da IA generativa promete revolucionar ainda mais a interação entre indivíduos e tecnologia, a partir do aprendizado com dados e identificação de padrões, resultando na tomada de decisões com o mínimo de intervenção humana.

Ao passo que essas evoluções serão responsáveis pela eliminação de milhões de empregos em todo o mundo, as posições de **especialistas em IA e aprendizado de máquina e analistas de inteligência de negócios** se destacam entre os empregos de crescimento mais rápido, segundo o Relatório sobre o Futuro dos Empregos 2023 do World Economic Forum.

Isso porque devido à escassez de mão de obra qualificada em dados, as oportunidades seguirão aumentando não apenas para os cientistas de dados e para quem já possui um perfil mais de “exatas”, como também para quem vem de outras áreas. Em comum, as diferentes carreiras em dados exigem pensamento analítico e habilidade de fazer as perguntas certas. Para quem busca uma transição ou está ingressando no mercado, as posições associadas ao universo do Big Data e Ciência dos Dados oferecem grande projeção de carreira!

Boa leitura!



Histórico

Antes da era Big Data, as organizações contavam com sistemas nos quais armazenavam grandes quantidades de informações, os chamados data warehouse, bem como ferramentas de análise descritiva e diagnóstica que relatavam os fatos da organização, por exemplo, a quantidade de vendas. “Elas já tinham pessoas especializadas em dados, mas eram carreiras mais técnicas, como profissionais de ETL (sigla para extract, transform, load - de extrair, transformar e carregar, na tradução) que usavam plataformas proprietárias, como Teradata, Oracle Exadata, SQL Server”, conta Regina Cantele, coordenadora acadêmica do MBA em Engenharia de Dados do Centro Universitário FIAP.

Esses perfis técnicos, como administradores de dados, administradores de banco de dados (database administrators - DBAs) e especialistas em visualização de dados, trabalhavam separados dos times de negócios, que geralmente demandavam dos primeiros relatórios baseados em dados internos.

A era Big Data teve início há aproximadamente 15 anos, juntamente com uma avalanche de dados. O conceito é baseado nos 3 V's: volume (grandes quantidades de dados), variedade (dados estruturados e não estruturados) e velocidade (agilidade de processamento). Nesse período, também surgem as chamadas big techs e a necessidade dessas plataformas de armazenar esses volumes de dados e processá-los em tempo real.

As big techs investiram no desenvolvimento de novas tecnologias de armazenamento de dados,

PRINCIPAIS TERMOS RELACIONADOS A CARREIRAS DE DADOS

Analytics: uso aplicado de dados, análises e raciocínio sistemático para tomada de decisão.

Big Data: conjunto de técnicas capazes de analisar grandes quantidades de dados para a geração de resultados.

Business intelligence (BI): combina análise empresarial, mineração de dados, visualização de dados, ferramentas/infraestrutura de dados e outras práticas para tomada de decisão orientadas por dados.

Data lake: repositório centralizado projetado para armazenar, processar e proteger grandes quantidades de dados (estruturados, semiestruturados e não estruturados).

substituindo os tradicionais bancos de dados relacionais por novos bancos de dados não relacionais, chamados de NoSQL (“não apenas SQL”). Essas soluções chegaram ao mercado no formato open source, permitindo que outras empresas pudessem processar e armazenar esse volume de dados para ganharem mais inteligência.

“O dado é, então, democratizado. Antes, o dado era proprietário, a infraestrutura era on-premise, as ferramentas eram caras e ficavam restritas à TI. Agora, as organizações têm acesso a dados internos e externos, as ferramentas são open source e disponíveis na nuvem e podem ser contratadas de forma flexível. As equipes de negócio se tornaram mais independentes da TI, pois têm acesso a uma base de dados e podem conduzir suas próprias análises via ferramentas de visualização”, diz a professora.

Esse conceito de “self-service analytics”, no qual os departamentos acessam os dados que precisam, exige que eles sejam preparados. Daí nascem



Ciência de Dados:

une princípios e práticas das áreas de matemática, estatística, inteligência artificial e engenharia da computação para analisar grandes quantidades de dados a fim de extrair insights para os negócios.

Data-driven: gestão orientada por dados.

Deep Learning (aprendizado profundo):

subárea de machine learning que utiliza as redes neurais artificiais para o aprendizado de máquina.

Inteligência Artificial (IA):

campo da Ciência da Computação que se concentra no desenvolvimento de sistemas e algoritmos capazes de realizar tarefas que exigem inteligência humana, como aprendizado e reconhecimento de voz e imagem.

as carreiras de dados e seus diferentes perfis necessários para fazer com que os dados sejam capturados, tratados, analisados e disponibilizados.

“Surgem figuras técnicas como engenheiro de dados, que desenvolve o ambiente que o pessoal de negócios acessa para construir as visualizações”. Em paralelo, as áreas de negócio também adquirem uma competência técnica a partir de profissionais de estatística, que fazem a correlação entre os dados e aplicam técnicas simples de machine learning, por exemplo, para fazer uma previsão de churn de clientes. “E juntamente com o engenheiro de dados, surge o cientista de dados para criar modelos analíticos integrando informações de dentro e fora da empresa, como dados da concorrência”, compartilha Regina.

Posteriormente, novas posições são demandadas para melhorar a qualidade dos dados e definir regras e políticas de uso desse ambiente analítico. Assim, temos as figuras responsáveis pela governança dos dados, como arquiteto de dados e DPO (Data Protection Officer).

“Uma empresa grande precisa de todos esses profissionais trabalhando juntos, enquanto uma menor prioriza um analista de dados e business intelligence que coloque a mão na massa tanto na hora de pensar em quais ferramentas utilizar, como na hora de armazenar e tratar esses dados para criar relatórios”, analisa Viviane Sampaio, division manager da Robert Half.

IA generativa: IAs com capacidade de criar informações a partir de conjuntos de dados pré-existentes.

Machine learning (aprendizado de máquina): ramo da IA no qual sistemas podem aprender com dados, identificar padrões e tomar decisões com o mínimo de intervenção humana. Esse método de análise de dados automatiza a construção de modelos analíticos.

Mercado Aquecido

A busca das empresas por informações valiosas para melhorar a produção, os processos financeiros, o gerenciamento de fornecedores e da logística e até mesmo a contratação de novos colaboradores têm levado à alta demanda por profissionais especializados na extração, análise e utilização desses dados.

Mario Maffei, gerente da divisão de tecnologia do PageGroup, analisa que quatro figuras despontam entre as mais procuradas: arquiteto de dados, engenheiro de dados, cientista de dados e analista de dados/ business intelligence (BI). Segundo o executivo, um terço das vagas da divisão é relacionado a dados, o que confirma uma imensa oportunidade para quem quer mergulhar na área.

“Essas posições demandam raciocínio lógico para lidar com algoritmos e modelagem de dados. Por serem cargos técnicos, exigem que o profissional se capacite na área, seja por meio de uma graduação, pós ou cursos especializados, e esteja sempre preparado para o novo, pois é uma área onde as mudanças se dão de forma rápida”, afirma Maffei.

Viviane Sampaio, da Robert Half, analisa que é muito comum nas carreiras de tecnologia que as pessoas queiram ser um desenvolvedor de software, enquanto poucas se interessam pelos dados. Como reflexo, o mercado está sofrendo com a dificuldade de contratar profissionais qualificados em dados.

“Muitos jovens têm contato com games e aplicativos, mas desconhecem como as informações circulam nesses sistemas e nas empresas. A área de dados não formou muita gente qualificada ao longo dos anos, por isso é importante que as pessoas conheçam essas carreiras e como elas transitam entre tecnologia e negócios”.

A grande maioria das vagas abertas no mercado, segundo ela, é para engenheiro de dados e cientista de dados. “São empresas que já possuem uma maturidade analítica e estão estruturadas para utilizar dados em larga escala. Nossa maior dificuldade atualmente na Robert Half é encontrar profissionais qualificados em Python e cloud”, revela Viviane.



Conheça as principais ferramentas de análise de dados

- **Visualização:** Microsoft Power BI, Tableau, Qlik e Looker
- **Ingestão, tratamento e pipeline de dados:** Fivetran, Airbyte, dbt, Databricks (Delta Lake), Snowflake, Amazon Redshift, BigQuery, Airflow, NiFi, Kafka
- **Construção de modelos de dados:** Glue Data Catalog, Informatica (gestão de metadados), Collibra, Talend, Trifacta
- **Modelos analíticos:** Python (linguagem de programação bastante utilizada na ciência de dados e machine learning), PySpark, R, TensorFlow, Jupyter, RapidMiner, Orange, H2O e DataRobot

Quem é quem no mundo dos dados?

Entenda as diferentes posições com as dicas da professora Regina Cantele, da FIAP.

Os três principais perfis

Engenheiro de Dados

É quem planeja e executa o pipeline de dados, permitindo que eles estejam disponíveis para uso com segurança no data lake (sejam os dados estruturados ou não). Também é responsável por construir o modelo analítico e pelo seu resultado. (por exemplo, quando você entra em um site de aluguel de imóveis e recebe sugestões de casas).

Entende de diferentes tecnologias, como armazenamento, nuvem, ferramentas de integração e da biblioteca Python, Big Data, machine learning etc. É alguém com raciocínio lógico e um perfil mais técnico e ligado à programação, que gosta de lidar com ingestão, tratamento e enriquecimento dos dados e fazer com que eles cheguem em tempo real até os cientistas de dados.

Cientista de dados

Profissional com raciocínio lógico apurado e utiliza técnicas mais sofisticadas de transformação de dados, como machine learning, deep learning, processamento de linguagem natural, para analisar e interpretar dados complexos. Como em toda Ciência, o cientista é aquele que testa e comprova uma hipótese para chegar a uma conclusão. Na Ciência dos Dados, ele é quem cruza dados de dentro e de fora da empresa e desenvolve os melhores modelos analíticos para fazer descobertas a partir de conclusões estatísticas. Os cientistas de dados também ajudam a criar aplicações de IA e machine learning, por exemplo, um chatbot que agiliza o atendimento de clientes de uma companhia.

Analista de dados e Business Intelligence (BI)

São pessoas de diferentes áreas (vendas, comercial, logística, financeira, marketing, RH etc.) que analisam dados para tomar decisões. Têm acesso aos dados do data lake e são capazes de construir relatórios, gráficos e dashboards por meio de planilhas ou ferramentas de visualização de dados, para inteligência de negócio e análise competitiva.

Arquiteto de dados: planeja o ambiente analítico que será desenvolvido pelo engenheiro de dados na nuvem. Ele define como será a arquitetura e a estratégia de dados organizacionais, quais ferramentas serão utilizadas, os padrões de qualidade dos dados, o fluxo dos dados e os requisitos de segurança.

Analista de suporte de ferramentas de analytics: trabalha na TI e dá suporte ao negócio no uso de ferramentas de análise. Porta de entrada muito comum para quem já entende de programação e quer se tornar um cientista ou engenheiro de dados.

Gerente de projetos analíticos: é quem gerencia os projetos que envolvem os dados de uma organização para que os times trabalhem de forma integrada, possibilitando uma entrega ágil.

Prompt engineer: carreira em plena ascensão por conta da IA generativa. É quem cria as instruções para treinar uma IA. Esse profissional tem que saber de análise de dados, Ciência de Dados e Engenharia de Dados para fazer as melhores perguntas.

DPO (Data Protection Officer): encarregado de cuidar de questões referentes à proteção de dados da organização e de seus clientes.



Saúde: uma oportunidade crescente

A transformação digital na saúde se deu muito mais tarde do que em outras áreas. De acordo com o professor Murilo Gazzola, professor da Faculdade de Computação e Informática da Universidade Presbiteriana Mackenzie, isso não diz respeito somente ao aspecto técnico, mas também ao impacto dessas soluções na sociedade, uma vez que vidas estão em jogo.

Apesar disso, Gazzola avalia que a integração entre inteligência artificial e saúde oferece inúmeros benefícios aos pacientes. É o caso das aplicações de medicina de precisão, que alia dados para diagnóstico e tratamento fornecidos por dispositivos conectados (como índice metabólico, batimentos cardíacos etc.) ao perfil genético do indivíduo. Elas possibilitam unir dados históricos, informações de exames e hábitos para trazer diagnósticos e tratamentos mais assertivos.

“Há um gap gigante na saúde para essas aplicações e este é um campo aberto. A ideia não é substituir um médico. No fim das contas, a computação traz escalabilidade e disponibilidade e a inteligência artificial pode ajudar a promover a saúde para mais pessoas e trazer mais personalização”, explicita Gazzola, que é especialista na área de Big Data com ênfase na saúde.

As possibilidades de unir Big Data à saúde são tanto para cientistas de dados quanto para engenheiros de dados. Além disso, profissionais de Radiologia também têm buscado esse campo de atuação.

O professor aconselha que é importante que esses profissionais tenham um conhecimento profundo sobre fundamentos de análise de dados, IA e Engenharia de Software, o que geralmente é propiciado por cursos de graduação em Ciências da Computação e Sistemas da Informação. Outro caminho é partir para cursos tecnólogos ou de pós-graduação com foco em Ciências dos Dados, que abrangem desde os fundamentos dos dados até a entrega de projetos. São cursos de curta duração ideais para quem busca uma virada de carreira.

Transição e Evolução

A posição de analista de dados/business intelligence (BI) é uma porta de entrada para quem não é do universo da tecnologia e deseja trilhar uma carreira em dados. É bastante demandada especialmente por empresas que ainda não possuem muita maturidade analítica e trata-se do caminho menos doloroso para quem busca uma transição de carreira, na opinião de Maffei.

A dica é aprofundar nas principais ferramentas de análise de dados, considerando que muitas são gratuitas, e conhecer o negócio para entender como a informação pode ajudar. E, conforme for se especializando, poderá migrar para os perfis mais técnicos, como engenheiro de dados e cientista de dados.

Basicamente, o analista de dados e BI vai construir relatórios e dashboards em cima dos dados que um engenheiro já estruturou. Não há necessidade desse profissional saber Python, mas deve dominar SQL, uma linguagem padrão para manipulação de registros em bancos de dados relacionais, e uma ferramenta de visualização de dados, como Power Bi e Tableau.

No entanto, Viviane, da Robert Half, observa que alguns setores que estão se tornando mais data-driven demandarão novas capacidades desses analistas. “Fintechs, bancos, indústrias, empresas de delivery e e-commerce estão precisando desses profissionais mais completos que além de fazer o dashboard, sabem trabalhar autonomamente para usar outros tipos de dados, com uso de Python, machine learning e deep learning”.



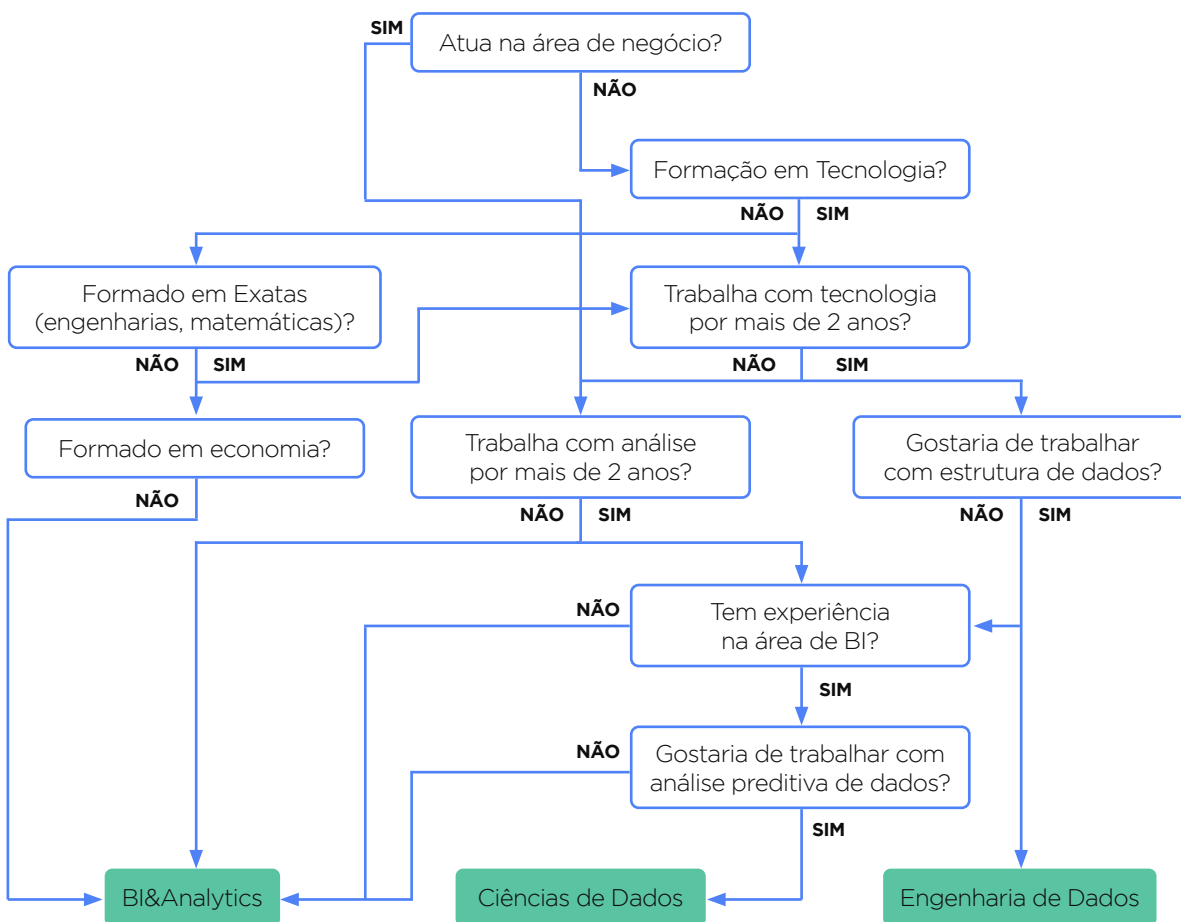
■ **“Cursos de pós-graduação com foco em dados permitem que profissionais que já atuam no mercado consigam se especializar e se destacar”** ■

Regina, da Fiap

Como lembra Regina Cantele, a carreira de dados é uma evolução. Uma pessoa de negócios pode começar se especializando em ferramentas de visualização, já alguém de TI pode conhecer as principais ferramentas de armazenamento e de pipeline de dados. Se quer trabalhar com governança, o caminho é tirar certificações do framework DAMA. “A graduação em Ciências da Computação ou Sistemas de Informação não é obrigatória, mas ajuda a desenvolver o pensamento analítico, por meio das disciplinas de cálculo, por exemplo, e da expansão do conhecimento. Já os cursos de pós-graduação com foco em dados permitem que profissionais que já atuam no mercado consigam se especializar e se destacar”.

Já pensou em trilhar uma carreira em dados?

Veja nesse diagrama qual seria o melhor caminho, de acordo com sua carreira hoje. O modelo é usado nas entrevistas para os cursos de pós-graduação da Fiap:



Quanto ganha um profissional de dados?



Fontes: *Page Group, **Guia Robert Half



De olho nas oportunidades

Soft skills cobiçadas: pensamento analítico, conhecimento do negócio, habilidade comunicacional e de relacionamento interpessoal (constante interação com outras áreas), poder de síntese, capacidade de trabalhar sob pressão, foco nas entregas e resiliência.

Os maiores empregadores: educação (análise de dados para captar novos alunos), e-commerce (dados para compreensão de características de consumo) e indústria (dados de people analytics, logística, consumo etc.). Vale ficar de olho também nos provedores de tecnologia que têm como projeto social a formação de especialistas em dados para reduzir a falta de mão de obra.

Áreas promissoras: saúde (integração entre dados para recomendações de tratamento) e empresas com iniciativas de sustentabilidade ambiental, social e de governança corporativa - ESG (usam dados para reduzir desperdício de recursos, minimizar impactos no meio ambiente e aumentar a eficiência produtiva).

Uma jornada pelo ecossistema de dados

Por Flávio Nakasato, sócio e cientista de dados da Novelo

Quando partiu em busca de um trabalho ao final da graduação em Física, Flávio percebeu que muitas das habilidades analíticas das vagas nas áreas de dados eram comuns ao seu curso. Iniciou como analista de BI em uma startup de app de táxi, onde aprendeu a trabalhar com bancos de dados e Python. Em menos de um ano, decidiu migrar para a Ciência de Dados e passou por startups de educação, consultoria e finanças, até chegar na Novelo, estúdio de data analytics baseado em São Paulo e fundado em 2018.

“Era comum que uma empresa contratasse um cientista de dados sem entender direito que precisava na verdade de um engenheiro de dados, um analista de BI, um devops ou uma equipe com tudo isso, antes de fato precisar de um cientista ou estatístico”. Por isso, ele acabou se desenvolvendo indiretamente em todo o ecossistema de dados nesses oito anos de carreira. Recentemente, começou um mestrado focado em Aprendizado por Reforço e Processamento de Linguagem Natural (NLP, na sigla em inglês), e, nas horas vagas, também dá aulas de programação, cloud e bancos de dados. Saiba mais sobre a carreira de Flávio.

Dia a dia do cientista de dados

Estamos construindo um produto inteligente que endereça problemas comuns de clientes, então muito do meu trabalho como cientista se confunde com o trabalho de engenheiro e de desenvolvimento de produtos. Um dia comum passa por reuniões com clientes, técnicos ou não, para entender como seus problemas podem ser mapeados com soluções de engenharia e de modelos de machine learning; con-



siderações sobre a infraestrutura que será necessária para dar suporte às soluções; construção e testes de modelos para atacar o problema; e tudo de novo, sempre interativamente com o cliente.

Desafio de quem atua na área

Resolver o problema correto - ou, traduzir o problema de negócio em termos técnicos claros e convencer quem toma as decisões a rever planos eventualmente mal construídos. Em oito anos, nunca peguei um problema para resolver que já estivesse bem definido. Com frequência, gestores ou PMs (product managers) subestimam ou superestimam pontos críticos de um projeto, fazendo com que cientistas e engenheiros trabalhem por semanas ou meses em não-problemas: ou já existe uma solução de prateleira disponível, ou as métricas-alvo não são as mais adequadas para o modelo, ou o caminho pensado não faz sentido mesmo.

Dicas para quem quer começar

A definição sobre o que faz um cientista de dados varia, então vale pensar no nicho em que se pretende trabalhar. Se for em chatbots, é importante focar em redes neurais e Transformers, que são a base dos LLMs (Large Language Models) atuais; em finanças, é melhor dedicar a séries temporais ou modelos tradicionais de previsão para dados tabulares, como Random Forest, LGBM, XGBoost etc. É importante ser objetivo e aprender o mínimo necessário de programação (comece com Python), matemática (um pouco de cálculo e álgebra de vetores e matrizes) e dos modelos para conseguir rapidamente uma posição de estagiário ou júnior na área, porque é na prática que você sente o que vai precisar para sua carreira. Não é fundamental, mas quem sabe um pouco de nuvem (AWS principalmente) e de bancos de dados (SQL) tem já um excelente diferencial na área de ciência de dados.

■ **Era comum que uma empresa contratasse um cientista de dados sem entender direito que precisava na verdade de um engenheiro de dados.** ■

Flávio Nakasato, da Novelo

O futuro das carreiras de dados

Como enfatiza o cientista de dados Flávio Nakasato é extremamente difícil prever como será o futuro da profissão - e de qualquer profissão - principalmente depois do boom dos modelos de IA generativa, como o ChatGPT.

Mário Maffei, da PageGroup, reconhece que a inteligência artificial e o machine learning proporcionarão cada vez mais agilidade nas entregas de dados. “Não acredito que essas inteligências substituirão um cientista de dados e, sim, permitirão que ele seja mais assertivo. Pode ser que ele se torne menos técnico, mas sempre haverá a necessidade de fazer novas descobertas a partir dos dados”.

Mesmo com esses avanços, Viviane Sampaio, da Robert Half, acredita que os profissionais que se dedicarem agora, terão amplas perspectivas. “As empresas são desafiadas a trabalhar orientadas a dados, fazendo previsões sobre sua competitividade no futuro”.

Para Nakasato, se o problema de alucinação desses modelos de IA generativa for reduzido para níveis mais aceitáveis, inúmeros postos de trabalho em todo o mercado se tornarão desnecessários. “Mesmo em ciência de dados, softwares de Automated Machine Learning (AML), isto é, que criam e testam inúmeras combinações de modelos inteligentes, como os da H2O, por exemplo, já não são simples de superar. Em NLP, há um ano ainda era necessário fazer refinamentos trabalhosos em modelos de prateleira, como o BERT, para obter algo minimamente funcional. Ou seja, era necessário um especialista para isso”.

Por outro lado, ele examina que hoje há vários modelos, como GPT-4, que só dependem de um prompt (uma instrução básica) e são “terrivelmente” superiores a vários desses modelos do passado. “A barreira de entrada baixou bastante. Resumo da ópera: a área de dados é uma das mais diretamente afetadas pela própria evolução gerada por ela mesma. Assim, vale aprender os fundamentos por trás das tecnologias que se usa, mas também aquilo que os robôs não farão bem tão cedo, como entender problemas de negócio e bolar soluções flexíveis e realistas para eles”, finaliza o sócio da Novelo.

EQUIPE EDITORIAL



Sobre o Eu Capacito

O Eu Capacito é um projeto social que tem o objetivo de formar uma legião de profissionais para a economia digital. Apoiado por diversas empresas da iniciativa privada, a plataforma Eu Capacito promove a capacitação profissional gratuita, focadas em habilidades de tecnologia, seja do ponto de vista conceitual, técnico (desenvolvimento) ou ferramental (manuseio para áreas de negócio), além de conhecimento em outras áreas consideradas importantes para a vida corporativa ou empreendedorismo (soft skills).

➤ [CONHEÇA O EU CAPACITO](#) ◀

Sigam nos, nas redes sociais

