

Garimpeiro moderno

Preparado para interpretar grande volume de informações, cientista de dados é profissional em alta no mercado, com múltiplas possibilidades de atuação

A pontada como campo de pesquisa em ascensão, que conjuga conhecimentos relacionados à computação, inteligência artificial, matemática e estatística, a ciência de dados consiste na análise de complexos volumes de informações geradas em diferentes plataformas. No Brasil, para atender a demanda crescente por cientistas de dados, instituições de ensino superior têm investido na criação de bacharelados e cursos de pós-graduação. O objetivo é formar profissionais aptos a interpretar, estruturar e analisar conteúdos que podem chegar à ordem dos petabytes, unidade de armazenamento que corresponde a 1.024 terabytes.

“Além de conhecimentos de computação, matemática e estatística,

o cientista de dados precisa ser curioso e gostar de resolver problemas”, observa André Ponce de Leon Ferreira Carvalho, vice-diretor do Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (ICMC-USP), em São Carlos, e do Centro de Ciências Matemáticas Aplicadas à Indústria (Cemeai), um dos Centros de Pesquisa, Inovação e Difusão (Cepid) apoiados pela FAPESP. A formação profissional envolve também a capacidade de identificar informações valiosas em bases gigantescas de dados, área que se tornou conhecida como big data. A instituição, que já forma cientistas de dados em nível de pós-graduação, inaugura agora curso de bacharelado específico na área.

“Os profissionais que já trabalham com ciência de dados vêm, em geral, de áreas correlatas, como ciência da computação, física, matemática, engenharia e estatística”, informa Carvalho. “O aumento na demanda justifica essa formação mais direcionada.”

Desde 2018, a ênfase em ciência de dados integra o currículo de todos os cursos do ICMC-USP, evidenciando sua relevância como campo de estudos multidisciplinar. Em 2020, o bacharelado em estatística, por exemplo, passou a ser denominado estatística e ciência de dados. “Essa é uma tendência mundial tanto nos cursos de estatística quanto nos de computação, que passaram a dar mais importância para a ciência de dados”,

completa Carvalho. Dentre as disciplinas que integram a matriz curricular do novo curso estão engenharia de software, inteligência artificial, computação de alto desempenho, redes de computadores e mineração em grandes bases de dados.

Além do quase imensurável volume de informações produzidas todos os dias, inclusive pelo uso de smartphones, assistentes virtuais, fechaduras eletrônicas, relógios, geladeiras, aspiradores de pó, aparelhos de ar-condicionado e televisores também já geram dados que podem ser utilizados para melhorar produtos e serviços ou simplesmente facilitar o dia a dia. “As informações produzidas pelos diversos sensores presentes nos automóveis, como câmera de marcha à ré e medidores de temperatura e velocidade, são armazenadas em bancos de dados e podem ser usadas para a realização de manutenção preventiva do veículo, além de avisar o condutor que ele está em uma área sujeita a alagamento ou com grande registro de acidentes”, exemplifica Carvalho.

A ampla base de dados que possibilita a análise sistematizada por parte dos cientistas divide-se, em geral, entre dados estruturados e dados não estruturados. Os primeiros são aqueles que surgem de forma já organizada, tais como a quantidade de acessos a determinado site ou aplicativo, número de usuários, produtos mais consumidos e locais que apresentam maior movimentação de pessoas. Os dados não estruturados reúnem textos e imagens publicados em redes sociais, sons captados a partir de microfones que podem estar instalados no oceano, em florestas ou em ambientes urbanos, por exemplo. Uma vez sistematizados, podem ajudar a prever fenômenos climáticos como furacões, identificar queimadas ou a ocorrência de assaltos em determinada região.

DEMANDA CRESCENTE

Se no início o computador chegou a ser visto como uma ameaça a postos de trabalho, sendo a automação associada ao aumento do desemprego, a partir de seu uso novas profissões

surgiram, dentre elas as relacionadas ao campo das tecnologias de informação e comunicação, as chamadas TICs. Dados divulgados pela empresa de consultoria Bain & Company, com sede em Boston, nos Estados Unidos, estimam que em 2020 aproximadamente 1 milhão de novos cientistas de dados tenha se formado em níveis de bacharelado e pós-graduação em todo o mundo. Dados da Associação Brasileira das Empresas de Tecnologia da Informação e Comunicação (Brasscom) indicam que, no país, a busca por profissionais de tecnologia não apenas segue em ascensão como pode sofrer com a falta de mão de obra qualificada. Até 2024, a demanda anual prevista é de 70 mil novos profissionais.

“Além das empresas que estão passando por transformações digitais ou que já nasceram com essa vocação, há oportunidades em diversas áreas da ciência, que cada vez mais dependem de cientistas de dados para interpretar informações produzidas em grande quantidade” afirma Bianca Zadrozny, pesquisadora e gerente sênior da área de Modelagem Espaço-temporal do Laboratório de Pesquisa da IBM Brasil. A empresa estima que, nos Estados Unidos, o número de vagas abertas continuará crescendo 5% ao ano, com 60 mil postos de trabalho abertos só em 2020. “O cientista de dados tende a ser uma das ocupações de maior destaque dentro das empresas, por sua capacidade de levantar hipóteses, projetar experimentos, avaliar resultados e apresentá-los de forma compreensível”, completa Zadrozny.

Ao avaliar o panorama nacional, Luis Gustavo Nonato, coordenador do novo curso do ICMC-USP, segue na mesma direção. “No Brasil existe um déficit enorme. As empresas tentam contratar cientistas de dados o tempo todo e, mesmo com a diversidade de cursos, ainda faltam profissionais”, diz. Segundo ele, o crescimento da demanda ocorre em todas as regiões do Brasil, mas é no Sul e Sudeste que está concentrado o maior número de vagas. “Além das empresas, governos e órgãos públicos dão cada vez mais importância para

o gerenciamento de dados regionais, essenciais para a formulação de políticas públicas.”

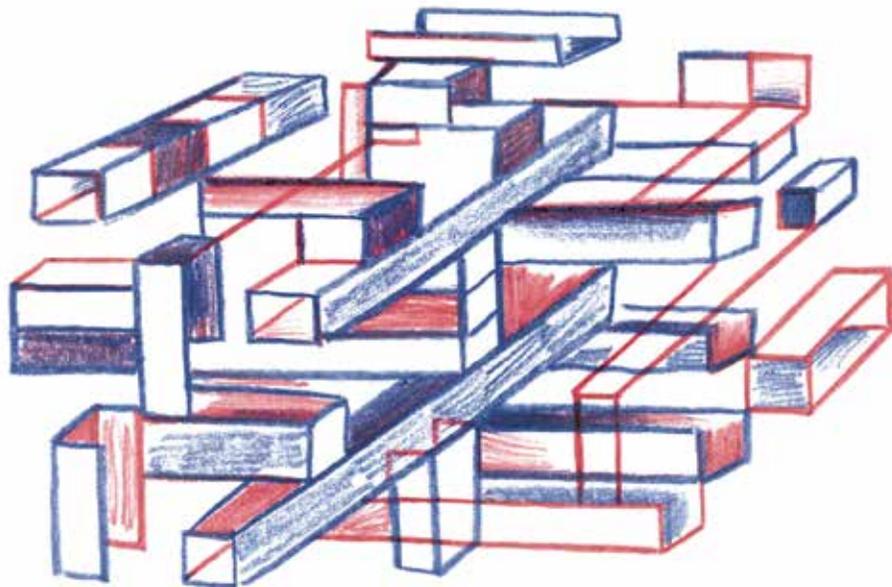
Um exemplo disso são as metodologias e os modelos que vêm sendo utilizados para aumentar a eficiência de processos jurídicos e sistematizar dados gerados no âmbito legislativo. Desde setembro de 2020 o ICMC-USP mantém com o Tribunal de Justiça do Estado de São Paulo (TJSP) convênio para criação de ferramentas de inteligência artificial que possibilitem a formação de banco de dados analítico com base nas informações constantes em documentos processuais. “O objetivo é analisar o conteúdo dos textos e levantar os assuntos mais recorrentes, além de apontar semelhanças entre processos”, explica Carvalho, do ICMC-USP.

O instituto também acaba de firmar contrato com a Câmara dos Deputados, em Brasília, para desenvolver métodos de aprendizado de máquina e processamento de linguagem natural que visam analisar o conteúdo dos canais de participação popular mantidos pelo órgão. “Com isso, será possível identificar os argumentos da população em relação às propostas legislativas, compilando posicionamentos contrários ou favoráveis a um determinado projeto de lei.”

Atenta ao aumento da demanda, a Pontifícia Universidade Católica de Campinas (PUC-Campinas) dará início, também neste ano, a um curso de bacharelado na área. Intitulado ciência de dados e inteligência artificial, o curso terá duração de quatro anos, sendo os três primeiros ofertados no período matutino e o último no período noturno. “Essa organização permite que ao final do curso os alunos possam realizar estágios nas diversas empresas de tecnologia presentes na região”, explica Daniele Maia Rodrigues, diretora da Faculdade de Engenharia de Computação da PUC-Campinas. Dentre os objetivos propostos pelo programa, Rodrigues destaca a formação de habilidades técnicas e interpessoais dos cientistas de dados, que devem conjugar conhecimentos sobre algoritmos e sistemas

Requisitos de um cientista de dados

- 1 Ter afinidade com a área de exatas, principalmente matemática e estatística
- 2 Familiaridade com linguagens de programação
- 3 Capacidade de solucionar problemas e identificar oportunidades de inovação
- 4 Disposição para trabalhar em equipes multidisciplinares
- 5 Boa capacidade de comunicação



computacionais, além da capacidade para trabalhar em equipes multidisciplinares. Para que tais habilidades sejam desenvolvidas, os alunos precisam entender de desenvolvimento de sistemas computacionais, programação, redes de computadores, computação em nuvem, infraestrutura de sistemas, inteligência artificial, aprendizado de máquinas e processamento de linguagem natural, ou seja, a capacidade que um computador tem de interpretar a linguagem humana, por meio da escrita ou da fala. “É importante observar que os cientistas de dados provavelmente integrarão equipes relacionadas a outros campos do conhecimento, de acordo com áreas de negócio específicas”, completa Rodrigues.

PÓS-GRADUAÇÃO

Para além da atualização de conhecimento, profissionais sobretudo da área de exatas têm visto na ciência de dados uma oportunidade de atuar em um campo de trabalho diferente. Sendo o bacharelado em ciência de dados uma alternativa de formação mais recente, empresas já vinham apostando na contratação de engenheiros, matemáticos, administradores, economistas e físicos que agora buscam complementar sua formação nessa área.

“Ter noções de linguagens de programação como Python e R

é indispensável para que o profissional obtenha êxito”, explica Eduardo Barbosa, coordenador do curso de pós-graduação *lato sensu* em data science e decisão do Instituto de Ensino e Pesquisa (Insper). Voltadas para a análise de bibliotecas de dados robustas, as linguagens Python e R são amplamente usadas na programação de aplicativos. Com o propósito de formar cientistas de dados capazes de dar suporte à tomada de decisões empresariais, o curso baseia-se na aprendizagem de modelos estatísticos e aprendizagem de máquina, programação e *design thinking* para ciência de dados, que envolve desenvolvimento da capacidade para compreensão de problemas e a proposição de soluções. “No exame de seleção buscamos identificar se o candidato atende aos requisitos básicos para admissão no programa. O objetivo é evitar a frustração de profissionais sem afinidade com o campo das ciências exatas”, informa Barbosa. Com duração de 20 horas divididas em três dias, a instituição também oferece uma modalidade expressa, voltada à formação de executivos que queiram utilizar os conceitos da área para melhorar os resultados de seus negócios.

Com primeira turma iniciada em 2018, o curso de especialização em ciência de dados e big data da

Universidade Federal da Bahia (UFBA) divide-se em três módulos, cada um com quatro meses de duração, compreendendo disciplinas como estatística aplicada, álgebra matricial e métodos numéricos, programação em R e Python, aprendizado de máquina, fundamentos de big data, inteligência artificial e reconhecimento de padrões de imagem, som e vídeo. “Além de estatísticos, matemáticos, cientistas da computação e engenheiros, entre os matriculados estão também profissionais de comunicação, direito, administração, dentre outros”, afirma Jalmar Manuel Farfan Carrasco, coordenador do curso oferecido pelo Departamento de Estatística do Instituto de Matemática e Estatística da UFBA. Para aproximar alunos de áreas tão distintas do conhecimento e facilitar a compreensão de conceitos mais específicos, o programa de pós-graduação tem investido em atividades em grupo, reunindo, sempre que possível, alunos formados em estatística e computação e a profissionais de outras áreas, como direito, comunicação, psicologia, dentre outros. “Ao observar um mesmo problema de formas distintas, essa interação permite que cada integrante contribua de forma singular na busca por soluções”, completa Carrasco. ■ Sidnei Santos de Oliveira