

# Nome do Curso: Ciência de Dados Aplicada a Saúde Online

- Justificativa: As dimensões territoriais e populacionais, o envelhecimento da população - que faz com que as pessoas utilizem mais os serviços de saúde - e o direito constitucional à saúde mostram a dimensão, a complexidade e a importância do mercado de saúde no Brasil. Segundo o Ministério da Saúde, atualmente o Brasil é o 8º maior mercado global de saúde e o volume de transações financeiras do segmento é responsável por quase 9% do PIB do País. Projeções apresentadas pela Confederação Nacional da Saúde estimam uma taxa de crescimento de vagas de empregos nos serviços de saúde para 2020 de, no mínimo, 4,3% e, no máximo, de 5,4% em relação a 2019. Uma simples análise do comportamento desse mercado permite concluir que ele tem sido positivamente afetado pela Tecnologias da Informação (TI). O avanço tecnológico é cada vez mais rápido e inclui, além das áreas próprias da medicina, toda uma estrutura social e de comunicação. Esse cenário propício para desenvolvimento pode ser confirmado pelo surgimento de diversas healthtechs focadas em soluções nos diversos segmentos que permeiam os serviços de saúde. Dados da Associação Brasileira de Startups (ABStartups) mostram que a quantidade healthtechs associadas quase dobrou entre 2015 e meados de 2019, passando de 235 para 406. Esse surgimento nasceu a partir da percepção dos empreendedores de que a área da saúde é uma das mais promissoras no país, que possui grande potencial para soluções inovadoras e também uma das que mais demandam melhorias. Estudos e pesquisas mostram que uma das questões críticas para ações no processo de decisão em saúde é a ausência de informação. Existem dados em abundância e em diversos tipos de mídias. A estratégia para a transformação significativa desse mercado está no uso dos dados como um recurso. Cada vez mais as empresas descobrem a potencialidade de se conhecer bem suas estruturas de dados e buscar através deste conhecimento ganhos tanto de qualidade quanto financeiro, seja através de previsibilidade, predições ou análises de dados históricos. A possibilidade de utilizar estas informações dentro dos processos tradicionais e transacionais das empresas pode proporcionar vantagens competitivas e ganhos nos fluxos e processos, podendo levar as empresas a um novo patamar de qualidade e produtividade. Ser capaz de processar e analisar grandes volumes de dados para obter uma compreensão mais profunda dos processos de gestão é o um dos principais objetivos das empresas. Mais especificamente, a própria gestão da saúde pública constitui mercado em que o uso eficiente de dados e informações pode gerar soluções inovadoras. Com seus mais de 5.500 municípios e tendo a gestão do Sistema Único de Saúde (SUS) realizada nas esferas municipal, estadual e federal, a saúde pública carece tanto de profissionais com visão estratégica de TI quanto de gestores habilitados para trabalharem com dados e informações de saúde. Embora, existam sistemas de informação com dados estruturados, a gestão pública também é feita por meio de dados não estruturados. A própria qualidade da informação para a população constitui em desafio seja pela forma de apresentar, pela profusão de fake News ou falta de exatidão. Ao propor uma ação ou política de saúde, é fundamental demonstrar o ganho que o cidadão/eleitor/financiador terá. Para tanto, dados são fundamentais tanto para a elaboração da ação quanto para o convencimento da

população. A pesquisa em dados não estruturados, por exemplo nas mídias sociais, pode ser útil tanto para instituições públicas quanto privadas em medir a eficácia e a repercussão das ações. Nota-se que fatores humanos – preparo para gerenciar as informações e favorecer o processo decisório – quanto fatores não humanos – sistemas interoperáveis, intuitivos e que favoreçam a leitura e interpretação dos dados – são fundamentais para a ciência de dados e sua aplicação na gestão em saúde. Nesse contexto, evidencia-se que aplicações de Ciência de Dados ganham cada vez mais importância. Por consequência, torna-se indispensável o profissional denominado Analista de Dados. Eles devem ser capacitados e ter competências e habilidades para entender bem as estratégias e necessidades do negócio e gerar soluções baseadas em dados para ajudar na melhoria dos serviços ofertados. No entanto, esse mercado apresenta carência desses profissionais. Previsões indicam que por volta de 2021 a área de TI terá que contar com 150% mais cientistas de dados para serem capazes de gerenciar a explosão de dados de saúde. É relevante, então, o engajamento das entidades de ensino na formação de profissionais para esse cenário. Este é o propósito da PUC Minas: prover formação sólida para profissionais analistas de dados para atuarem em um mercado com alta demanda de pessoas com capacidade de gerar valor para as empresas do setor de saúde. Matriz curricular atualizada para atender as necessidades do mercado, oferecendo uma formação focada no uso de Ciência de Dados Aplicada à Saúde em tecnologias atuais; Conteúdos apresentados por meio de casos reais, que colocam o aluno próximo de situações comuns no dia a dia de um usuário de dispositivo móvel; Tradição de ensino PUC Minas; Professores com muita experiência de mercado e com uma sólida formação acadêmica; Mentores experientes focados em orientar e motivar para otimizar o aprendizado; Abordagens inovadoras de ensino-aprendizagem em que as aulas e atividades pedagógicas são centradas nas necessidades dos alunos. Elas seguem dinâmicas orientadas por princípios de metodologias ativas; Experiência de aprendizado é suportada por ferramentas interativas - acessível via Web ou dispositivos móveis - incluindo salas virtuais, bate-papos e fóruns de discussão para estimular o aluno a um maior engajamento com o seu curso; Aprendizagem flexível em que o aluno planeja o próprio ritmo para alcançar seus objetivos pessoais;

- Objetivos: Identificar e compreender as diversas perspectivas da Ciência de Dados na área da saúde e verificar seus impactos em novos projetos; Compreender questões conceituais, padrões e práticas relacionadas ao tratamento de dados de saúde, bem como elaborar e executar os processos de governança e gestão de dados para garantir a qualidade dos mesmos; Conhecer as principais ferramentas e tecnologias relacionadas às soluções de Ciência de Dados e suas aplicações processo decisório na gestão em saúde; Projetar, desenvolver e gerenciar soluções inovadoras de ciência de dados na área de saúde; Atuar forma empreendedora na criação de novos negócios, utilizando os conceitos e as técnicas aprendidos no curso; Prospectar tendências em Ciência de Dados na área de saúde. O especialista em Ciência de Dados Aplicada à Saúde poderá atuar na área de saúde como Analista de Dados (Data Analyst) e, também, nas mais diversas áreas que demandam conhecimentos de ciência de dados e em diversos papéis como: cientista de dados, engenheiro de IA, Gerente de Projetos de Análise de Dados entre outros. Além disso, ele poderá atuar nos mais diversos tipos de projetos inovadores em TI e em

Healthcare.

- Público Alvo: Profissionais com formação superior: Atuantes na área de gestão da saúde ou áreas afins; Que atuam no mercado em projetos relacionados a inovação transformação digital; Que já atuam em projetos relacionados a análise e gestão de dados e que queiram complementar e aperfeiçoar seus conhecimentos na área de gestão de dados aplicados na área de saúde.

## **Disciplinas:**

### **Disciplina 1: PREPARAÇÃO E INTEGRAÇÃO DE DADOS**

Ementa: Melhoramento, enriquecimento e preparação de dados. Montagem do conjunto de dados. Feature Engineering ETL, ELT e Data Lake. Processo de integração de dados. Ferramentas. Projeto e desenvolvimento de aplicação de preparação e integração de dados. Operação. Conceitos e técnicas de ingestão de dados.

### **Disciplina 2: MACHINE LEARNING**

Ementa: Processo de aprendizagem de máquina. Feature Engineering. Técnicas e algoritmos de aprendizado supervisionado e não-supervisionado. Combinação de modelos. Métricas e avaliação de modelos.

### **Disciplina 3: PROCESSAMENTO DE LINGUAGEM NATURAL**

Ementa: Algoritmos e técnicas de processamento em linguagem natural. Expressões regulares. Medidas de similaridade textual. Parsing, tokenização, lematização, stemming. Marcação textual. Reconhecimento de entidades nomeadas. Extração de informação. Arquitetura de aplicação para processamento de Linguagem Natural. Análise de sentimento.

### **Disciplina 4: TÉCNICAS DE AMOSTRAGEM E MODELOS DE REGRESSÃO**

Ementa: Noções básicas de distribuição e amostragem. Tipos de amostragem. Dimensionamento da amostra. Técnicas de amostragem. Regressão linear simples e múltipla. Modelo de regressão linear múltipla. Estimação dos parâmetros do modelo. Inferências sobre a regressão linear múltipla. Estudo da adequação do modelo. Problemas da regressão linear múltipla. Variáveis Dummy. Modelos Lineares Generalizados: Família Exponencial de Distribuições. Componente sistemática. Função de ligação. Modelos Lineares Generalizados Especiais. Função desvio. Função Escore. Testes de Hipóteses. Matriz de Informação de Fisher. Análise de diagnóstico.

### **Disciplina 5: SÉRIES TEMPORAIS**

Ementa: Conceitos básicos e modelos de séries temporais. Estacionariedade. Função de autocorrelação. Modelos no domínio do tempo e da frequência. Método de decomposição. Modelos de tendência: determinística e estocástica. Método de medias moveis. Alisamento exponencial. Modelagem de séries temporais estacionárias: Modelos Autoregressivos e de Médias Móveis (ARMA). Modelagem de séries temporais não estacionárias: transformações e/ou diferenciação. Modelos Autoregressivos Integrados e de Médias Móveis (ARIMA). Modelos Sazonais Autoregressivos Integrados e de Médias Móveis (SARIMA). Análise de intervenção. Regressão em séries temporais. Regressão Dinâmica.

### **Disciplina 6: PYTHON PARA CIÊNCIA DE DADOS**

Ementa: Tipos de dados. Estruturas de controle: condicional e repetição. Estruturas de dados: listas, tuplas, conjuntos, dicionários, séries e dataframes. Funções. Vetorização e matrizes numéricas. Bibliotecas de manipulação de dados, de visualização de dados e vetorização de matrizes.

### **Disciplina 7: ESTATÍSTICA GERAL - TEORIA E APLICAÇÕES**

Ementa: Estatística descritiva. Introdução a probabilidade. Distribuições de probabilidade. Inferência estatística: estimação pontual e intervalar de parâmetros, testes de hipóteses e regressão linear simples. Utilização de software para análises estatísticas e análise de casos aplicados à gestão.

### **Disciplina 8: GESTÃO E COMPLIANCE EM SAÚDE**

Ementa: Sistema de saúde brasileiro. Sistema Único de Saúde. Principais players em saúde: consultórios e clínicas particulares, hospitais, cooperativas do setor. Serviços de Saúde. Abordagem do Quadruple Aim. Saúde digital: Telemedicina, telessaúde, wearables, IoMT. Processo de tomada de decisão em saúde. Ciência de Dados e Analytics em saúde. Papel das equipes multidisciplinares na análise de dados de saúde. Liderança em Análise de Saúde. Lei Geral de Proteção de Dados (LGPD) e especificidades da saúde. Aspectos ético-legais da gestão da informação em saúde. Padrões e protocolos de informação na saúde. Requisitos do Registro Eletrônico em Saúde. Tendências emergentes na área da saúde. Cases.

### **Disciplina 9: GESTÃO DE INDICADORES DE SAÚDE**

Ementa: Necessidades estratégicas de dados. Empresas Orientadas a Dados. Cultura Data Driven. Data literacy. Métodos para criação de estratégia de dados. Estratégia de dados para tomada de decisão. Conceituação e Classificação de Indicadores. Construção de Indicadores: Metodologia e Sistematização. Planejamento e implementação de medidas de desempenho. Alinhamento de Indicadores a processos e estratégia. Processo de monitoramento e análise Indicadores de saúde. Cases. Processo de monitoramento de qualidade de dados. Tomada de decisão e construção de planos de ação.

### **Disciplina 10: REDES NEURAIS E DEEP LEARNING**

Ementa: Introdução a redes neurais artificiais. Deep Learning. Técnicas, métodos e arquiteturas para redes neurais. Redes convolucionais modernas. Aplicações de redes convolucionais e deep learning.

### **Disciplina 11: GENERATIVE AI & ADVANCED ANALYTICS**

Ementa: Fundamentos de IAs Generativas (GenAI). Plataformas de GenAI. Engenharia de Prompt. Exploração de modelos preditivos. Utilização de APIs de modelos externos. AutoML. Cloud Analytics. Implementação de pipeline da solução. Conceitos de AI-as-a-Service, AI-as-a-commodity, AI Gateways. Tendências.

### **Disciplina 12: ANÁLISE DE IMAGEM E VISÃO COMPUTACIONAL**

Ementa: Introdução à visão computacional. Tipos de Visão computacional. Modelos de representação de imagem. Manipulação e processamento de imagens digitais com OpenCV. Descriptores de imagens. Recuperação de imagens com base no conteúdo visual utilizando BoW. Redes Neurais Convolucionais (CNN), arquitetura de redes neurais convolucionais. Classificação de imagens usando CNNs. Estratégias de data augmentation de imagens. Modelos pré-treinados e estratégias de fine-tuning em redes neurais para classificação de imagens. Técnicas de detecção de objetos e

reconhecimento de faces.

### **Disciplina 13: CIÊNCIA DE DADOS EM EPIDEMIOLOGIA**

Ementa: Fundamentos do Método Epidemiológico. Bases de dados públicas em saúde. Variáveis, proporções, indicadores e taxas em Epidemiologia. Indicadores populacionais de saúde. Medidas de ocorrência de doenças e óbitos. Validade e confiabilidade de testes diagnósticos e rastreamento. Estimativas de sobrevivência e prognóstico. Delineamento de estudos epidemiológicos. Estimativas de risco e medidas de associação.

### **Disciplina 14: HUMANIDADES**

Ementa: O ser humano, o processo de humanização e o conceito de pessoa. Desafios contemporâneos e o lugar da religião e da espiritualidade. Autonomia e heteronomia na sociedade atual. Princípios éticos e ética profissional.

### **Disciplina 15: DATA DISCOVERY E ANALYTICS**

Ementa: Fundamentos da descoberta de dados. Fundamentos e requisitos de aplicações de suporte a decisão. Princípios de projeto, arquitetura e construção de aplicações OLAP. Análise, visualização e comunicação de dados. Ferramentas de Data Discovery e Self-Service Analytics.

### **Disciplina 16: CULTURA E PRÁTICAS DATAOPS E MLOPS**

Ementa: A cultura DevOps. Integração contínua e entrega contínua. Estratégias de deploy. Projeto de pipeline para build e deployment. Automação de testes. Infrastructure as Code (IaC). Ferramentas e infraestrutura do ambiente integrado DevOps: Containers, Docker, Kubernetes e OpenShift.