

# Nome do Curso: Ciência de Dados e Big Data Online

- Justificativa: O mundo dos negócios tem-se transformado com soluções inovadoras e, segundo consultoria Gartner Group, o Big Data é uma das que integram esse cenário. Percebe-se que o dia-a-dia de uma economia altamente competitiva e globalizada tem posto como fator crítico de sucesso e gerador de diferenciais competitivos significantes para o negócio a capacidade das organizações em ter acesso e poder analisar, no tempo certo e de forma segura, seus dados, informações e conhecimentos necessários para a sua gestão. Nesse contexto, é cada vez mais significativa a contribuição e o papel que os sistemas de informação, alinhado aos conceitos de Big Data e análise de dados, têm no desempenho das organizações. A utilização do grande volume de dados, denominados de Big Data, para geração de valor nas organizações deve ser um objetivo constante. Assim, é quase imperativo que as empresas promovam o entendimento desses dados em tempo real e avaliem como utilizá-los para obtenção de informação útil para se diferenciarem no mercado. Para potencializar isso elas demandam por profissionais com facilidade para absorver rapidamente as regras de negócio da empresa e propor soluções que auxiliem os gestores a tomarem decisões de forma rápida e segura. Esse profissional está sendo tratado como Cientista de Dados. Ressaltam-se também as previsões para o mercado de Big Data. O Gartner Group destaca Big Data como uma das principais tendências de crescimento nos próximos anos. Ele mostra em um estudo que 73% das empresas almejam fazer uso de ferramentas de análises de dados; outros 49% apostam no Big Data como a porta para novos negócios. O Gartner afirma ainda que a capacidade de gerar e transmitir dados crescerá de forma acelerada e que o avanço das técnicas de analytics torna a tarefa de estruturar esses dados cada vez mais fácil, deixando o Big Data mais útil. O International Data Corporation – IDC afirma que investimentos em infraestrutura, software e serviços de Big Data vão crescer 23,1% ao ano a partir de 2016. Ainda de acordo o IDC, em 2020, cerca de 50% de todos softwares de business analytics vão incluir análises prescritivas construídas com base em funcionalidades da computação cognitiva e serviços cognitivos serão integrados a novos aplicativos. A grande maioria das empresas será capaz de analisar todos os dados relevantes para obter insights úteis, criando uma vantagem em produtividade em relação aos concorrentes menos habilitados em analytics. De acordo com o Forrester, em 2016 um número maior de empresas tentará tirar valor e receita de suas informações por meio da comercialização de seus próprios dados. Evidencia-se, então, que conceitos como Big Data e Ciência de Dados ganham cada vez mais importância dentro das empresas. Com isso, induz-se também que o mercado precisará então de profissionais capacitados e com competências e habilidades para entender bem as estratégias e necessidades do negócio. Isso para gerenciar, projetar e desenvolver com qualidade projetos relacionados à gestão e à análise de dados.

- Objetivos: Formar profissionais capazes de analisar o estado da arte de Big Data e Business Analytics com forte embasamento conceitual e prático; Capacitar os participantes a projetar, desenvolver e gerenciar projetos que demandam técnicas atuais para análise de grandes volumes de dados. Mostrar a importância dos dados no âmbito da organização, bem como elaborar e executar o processo de garantia de

qualidade dos mesmos e desenvolver os conceitos relacionados a sua governança. Capacitar profissionais para prospectar tendências na área de Ciência de Dados.

- Público Alvo: Profissionais com formação superior: Em Engenharia de Software, Sistemas de Informação, Ciência da Computação, Engenharia de Computação e tecnólogos da área de Tecnologia da Informação e outros cursos correlatos; Que atuam em Inteligência de Negócios e TI, gerentes de TI, coordenadores de projetos, analistas de negócios, analistas de redes sociais e mídias e outros profissionais com interesse em tecnologias para análise de bases de dados para apoio a tomada de decisão.

## **Disciplinas:**

### **Disciplina 1: 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 2: 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 3: 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 4: 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 5: FRAMEWORKS PARA DEEP LEARNING**

Ementa: Introdução a Keras e Tensorflow. Anatomia de redes neurais com keras e tensorflow. Classificação com deep learning: classificação binária e classificação multiclasse. Regressão linear com deep learning. Monitoramento do modelo e Inferências em modelos treinados. Introdução ao Pytorch e outros frameworks para deep learning. AutoML. Projeto e utilização de infraestrutura e serviços de computação em nuvem. Implementação de pipeline da solução.

### **Disciplina 6: ARMAZENAMENTO E PROCESSAMENTO MASSIVO E DISTRIBUÍDO DE DADOS**

Ementa: Introdução aos sistemas distribuídos. Princípios de processamento e de volumes de dados massivos. Sistemas de arquivos distribuídos. Modelo MapReduce. Balanceamento de carga. Replicação e redundância de dados e metadados. Gestão de metadados. Sincronização e disponibilidade. Operações de leitura, escrita e listagem de arquivos distribuídos. Soluções em Processamento Distribuído. Técnicas de

compressão, otimização e particionamento de dados.

### **Disciplina 7: BANCOS DE DADOS RELACIONAIS E NÃO RELACIONAIS**

Ementa: Modelo Relacional. SQL. Bancos de Dados NoSQL: definição; motivação; modelo de Transações. Modelos NoSQL. Propriedades Modelo Relacional x Propriedades Modelos NoSQL. Principais SGBD's.

### **Disciplina 8: 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 9: 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 10: 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. Tendencias.

### **Disciplina 11: 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 12: 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 13: GOVERNANÇA EM CIÊNCIA DE DADOS**

Ementa: Alinhamento com o planejamento estratégico e com a estratégia de dados da organização. Engajamento dos stakeholders. Liderança em Ciência de dados (CD). Papéis e responsabilidades em equipes de CD. Gestão de equipes e de recursos. Definição dos processos de CD. Definição de métricas e indicadores de desempenho.

Definição e atualização da Toolchain para Ciência de Dados e IA. Planejamento e gestão dos projetos. Apoio às diversas equipes. Gestão do ciclo de vida das soluções de CD. Gerencia de modelos. Impactos da LGPD em CD. Definição de políticas e diretrizes de uso de dados. Regulamentos, riscos e segurança em CD. Estudo de casos.

#### **Disciplina 14: 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.

#### **Disciplina 15: 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 16: 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.