

Nome do Curso: Estatística para Ciência de Dados

- Justificativa: Os dados desempenham um papel fundamental no mundo dos negócios. Soluções inovadoras têm sido criadas a partir de insights sobre eles. Como se sabe, Ciência de Dados é o estudo de dados em diferentes formas para fazer suposições adequadas sobre comportamentos e tendências. Porém, quando o volume de dados é grande e as estruturas e tipos de dados são variados, os métodos estatísticos são essenciais. Para que os projetos tenham sucesso, é importante que as informações sejam organizadas de acordo com os conceitos da estatística para que o estudo se torne fácil e conseqüentemente as descobertas se tornem mais precisas e confiáveis. A análise de dados e os modelos estatísticos suportam muitos aspectos do mundo moderno, desde ciência e tecnologia até finanças e negócios. Eles nos permitem superar problemas científicos, industriais e sociais e uma compreensão avançada deles é benéfica em muitas carreiras. O cientista de dados é um profissional capaz de analisar fontes de dados, limpar e processar os dados, entender por que e como esses dados foram gerados, obter insights e fazer alterações que beneficiem a organização. Mas para potencializar o resultado dos projetos de análise de dados, o mercado precisará cada vez mais de profissionais com forte formação em estatística, isso para gerenciar, projetar e desenvolver com qualidade projetos relacionados à gestão e à análise de dados. Nesse sentido, o curso de Estatística para Ciência de Dados deve prover habilidades e competências de análise de dados e desenvolver a capacidade de gerar informações e articular claramente como elas podem ser usadas. O curso abrange tanto a teoria quanto a aplicação de técnicas modernas de modelagem estatística e matemática para problemas aplicados na indústria, serviços públicos, comércio e pesquisa. Ele deve oferecer uma forte base teórica, mas deve ser orientado na prática para desenvolver a capacidade de enfrentar problemas novos e não padronizados com confiança. A dependência mútua da prática e da teoria é enfatizada ao longo do curso. Matriz curricular atualizada para atender as necessidades do mercado, oferecendo uma formação focada na definição de arquiteturas de software 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; Professores 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: O Curso de Especialização em Estatística para Ciência de Dados visa capacitar profissionais de diversas áreas que atuam com estratégias e análise de dados. O objetivo principal é capacitar profissionais para aplicar métodos estatísticos e modelos de Machine Learning para realizar diagnósticos e favorecer a tomada de decisão orientada a dados. Ao final do curso, o aluno será capaz de: Projetar, desenvolver e gerenciar projetos que demandam técnicas atuais para análise de

dados; Aplicar métodos e programas estatísticos para análise de dados em projetos empresariais; Realizar planejamento estatístico e desenvolver relatórios gerenciais que permitam apresentar resultados embasados estatisticamente de forma clara e concisa e em linguagem adequada; Conhecer e aplicar as principais funções e modelos, compreendendo suas especificidades e diferenças. Entender e promover a cultura de dados para garantir a geração de valor para o negócio. Projetar e desenvolver soluções complexas que demandam técnicas atuais de aprendizado de máquina.

- Público Alvo: Profissionais com formação superior: Que atuam em projetos ou atividades relacionados à análise e gestão de dados e que queiram complementar e aperfeiçoar seus conhecimentos em metodologia estatística; 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 dados para apoio a tomada de decisão; Que atuam ou pretendem atuar no desenvolvimento de soluções inovadoras e inteligentes para favorecer a transformação digital.

Disciplinas:

Disciplina 1: TÉCNICAS DE AMOSTRAGEM E REGRESSÃO LINEAR

Ementa: Noções básicas de distribuição e amostragem. Tipos de amostragem. Dimensionamento da amostra. Estimador Razão e Regressão. Técnicas de amostragem. Amostragem aleatória e viés de Amostra. Viés de seleção. Viés e erros não amostrais de não resposta. Estimação da variância. Distribuição de amostragem. Bootstrap. Intervalos de Confiança. Tipos de distribuição. Distribuição Poisson e Distribuições Relacionadas. Regressão linear simples e múltipla. Regressão com Variáveis Qualitativas. Análise de Resíduos. Diagnóstico em Regressão. Seleção de variáveis e construção de modelos. Validação de Modelos. Transformação de Variáveis.

Disciplina 2: 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 3: 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 4: 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 5: ANÁLISE DESCRITIVA E PROBABILIDADE

Ementa: Tipos de Variáveis. Representação tabular e gráfica. Medidas de posição e dispersão. Análise Bivariada para variáveis qualitativas e quantitativas. Análise bidimensional. Análise de Correlação. Técnicas de visualização para os diversos tipos de variáveis. Experimento. Espaço Amostral e Eventos. Probabilidade e propriedades. Variáveis aleatórias discretas e principais modelos de distribuição discretas. Variáveis aleatórias contínuas e principais modelos de distribuições contínuas. Principais distribuições de Probabilidade. Teorema do Limite Central. Modelos de Convergência. Desigualdade de Tchebyshev. Lei dos grandes números.

Disciplina 6: INFERÊNCIA ESTATÍSTICA

Ementa: Noções gerais de inferência estatística. Métodos de estimação. Propriedades dos estimadores. Estatísticas suficientes. Teste de hipótese. Métodos estatísticos Não-Paramétricos. Testes para uma amostra. Testes para duas amostras relacionadas e duas independentes. Testes para k amostras relacionadas e k amostras independentes. Medidas de correlação e seus testes de significância. Métodos de Reamostragem.

Disciplina 7: ANÁLISE DE SOBREVIVÊNCIA

Ementa: Conceitos básicos e dados de sobrevivência. Métodos Não-Paramétricos para estimação da função de sobrevivência. Modelos probabilísticos em análise de sobrevivência. Modelos de Regressão Paramétricos. Análise de resíduos. Modelagem via função de risco: modelo de Cox e modelo de taxa acelerada. Testes Acelerados: conceitos básicos, relação estresse-resposta e aplicações.

Disciplina 8: ANÁLISE MULTIVARIADA

Ementa: Introdução a estatística multivariada. Álgebra vetorial e matricial. Dados multivariados: vetores de médias, matriz de variância e covariância. Distribuições multivariadas: Normal, T2. Análise: de variância, de agrupamento, fatorial, discriminante, de correlação canônica, multidimensional, de correspondência. Redução de dimensão.

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: 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 11: 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 12: MODELOS DE REGRESSÃO

Ementa: Regressão Linear 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: 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 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: GOVERNANÇA DE DADOS

Ementa: Contexto organizacional de dados. Conceitos de Governança de Dados (GD). Framework DMBOK. Políticas, padrões e procedimentos aplicados aos dados: Data Stewardship, Data Owners, Dados Mestres, Dados Referência, Metadados, Data Catalog. Processo de implantação de GD. Modelos de maturidade de dados. GD aplicada em leis de Proteção (LGPD-GDPR). Compliance e Risk Assessment. GD 2.0: Ética nos dados, Agilidade em GD, Gerência de Mudanças.

Disciplina 16: 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.