

DADOS DETALHADOS DO CURSO

Especialização em:

(MBA em Ciência, Governança e Análise de Dados)

Carga Horária:
(375 horas)

Justificativa:

O MBA em Ciência, Governança e Análise de Dados (MB-CGA), é uma ação de capacitação de recursos humanos, associada a uma especialização da Faculdade Mackenzie, com o objetivo de formar profissionais com habilidades nas técnicas modernas de inteligência de negócio com competências para utilizar frameworks disponíveis e resolver problemas práticos em empresas de tecnologia. Este objeto é de fundamental importância para que se possa formar recursos humanos qualificados e com experiência na resolução de problemas reais com foco no processo de transformação digital atualmente em curso.

Objetivos:

Objetivo Geral:

O curso é baseado em conceitos e métodos sobre como coletar, compilar, analisar, interpretar, modelar, melhorar, e monitorar a análise de dados. Conceitos de modelagem de dados, diagnóstico preditivo, processos de negócios, técnicas de solução de problemas do negócio, e inteligência do negócio através dos dados serão apresentados.

Objetivos Específicos:

Dentro deste contexto, o objetivo deste curso é fornecer o conhecimento sobre gestão e análise de grande quantidade de dados de modo a auxiliar na tomada de decisão durante todas as fases de processo, seja ele de negócios, industrial ou tecnológico visando formar profissionais com um perfil inovador e tecnicamente preparado para aplicar os conceitos, ferramentas e métodos de gerenciamento no seu campo de atuação.

Público-Alvo:

O Curso foi concebido para capacitar profissionais especializados de todas as áreas, aptos a idealizar e conceber soluções inovadoras para todos os campos profissionais.

Programa:

Módulo	Disciplinas	Carga Horária
Básico	Metodologia e pesquisa científica	20h
	Introdução a banco de dados (relacional e NoSql)	25h
	Governança de Dados	25h
	Fundamentos de Big Data	20h
	Inteligência de Negócios	20h
	Análise Estatística de Dados com Linguagem R	30h
	Intermediário	Virtualização e Computação em Nuvem
Preparação de Dados (ETL)		25h
Inteligência e Modelagem na WEB		25h
Data Mining		25h
Introdução a Inteligência Artificial com Python		25h
Avançado	Modelagem MAP Reduce	20h
	Machine Learning	25h
	Deep Learning	25h
	Análise de dados com dashboards de negócio	25h
Projeto final	Seminários	15h
	Projeto final	
	Defesa de monografia	
Carga Horária Total		375 horas

Coordenação:

Nome: Antônio José Dias da Silva

Titulação: Mestre

Currículo Resumido: Engenheiro Elétrico com ênfase em Eletrônica pela Sociedade Educacional Professor Nuno Lisboa (1984), Mestre em Informática pela Universidade Federal do Rio de Janeiro (2012), Especialista em Métodos e Técnica de Ensino pela UCB (1993), Conselheiro do CREA por 2 anos.

Atuou como Engenheiro de Desenvolvimento de Sistemas Digitais na Indústria de Material Bélico do Brasil – IMBEL, por 12 anos. Possui mais de 36 anos de experiência acadêmica no ensino superior, tendo atuado como Diretor de Centro Acadêmico, Coordenador de Curso e Professor na UCB, onde participou na elaboração de vários projetos pedagógicos e de extensão. Atuou também como professor nas Universidades, UGF, UNESA e Universidade da Cidade. Participou da criação do Curso de Sistemas de Informação da Faculdade Cesgranrio, onde atualmente é Coordenador e professor do Curso, atuando também como professor no Centro Universitário Augusto Motta (UNISUAM), nas áreas de Ciência da Computação, Sistemas de Informação, Engenharias, onde também faz parte do NDE das Engenharias. Link para Currículo Lattes: <http://lattes.cnpq.br/8922587549718409>

Período e Periodicidade:

Período estimado de duração do Curso: 20 meses

Dias da Semana: Sábados

Horário: 08:00h às 18:00h

Trabalho Final:

Elaboração de Monografia, Seminários, apresentação da Monografia, apresentação de relatórios técnicos e projetos na área do curso.

Conteúdo Programático:**MÓDULO BÁSICO**

Disciplinas	Ementa
Metodologia de Pesquisa Científica	Introdução. O Conhecimento. A Ciência e suas Características. A Pesquisa Científica e o Desenvolvimento Tecnológico. O Método Científico e Suas Aplicações. Uso do Método Científico no Cotidiano Acadêmico e Profissional. Planejamento de uma Pesquisa. O Projeto de Pesquisa e sua Elaboração. A Comunicação Científica: Monografia, Tese, Relatórios, “Papers”.
Introdução a Banco de Dados (Relacional e NoSql)	Banco de Dados Relacional; Banco de Dados Distribuídos; Bancos no S QL; Arquitetura de Dados; Abstração de Dados; Modelagem Conceitual de Dados; Fundamentos de Projeto de Banco de Dados; Projeto Conceitual de Banco de Dados; Projeto Lógico de Banco de Dados; Projeto Físico de Banco de Dados; Projeto de Banco de Dados Orientado a Objetos; Administração de Dados; Banco de Dados Não convencionais etc.
Governança de Dados	Arquitetura de Soluções Big Data; Pipeline cobrindo desde a coleta à visualização de dados; Exemplos de utilização de tecnologias ditas Big Data em diversas áreas, como por exemplo, médica, vendas, educação, financeira etc.
Fundamentos de Big Data	Arquitetura de Soluções Big Data; Pipeline cobrindo desde a coleta à visualização de dados; Exemplos de utilização de tecnologias ditas Big Data em diversas áreas, como por exemplo, médica, vendas, educação, financeira etc.
Inteligência de Negócios	Business Intelligence. Arquitetura e soluções de Business Intelligence. Data Warehouse. Aplicações ETL. A importância das Ferramentas de BI. As aplicações de BI front-end. BI: primeiros passos: Por que implementar um projeto de Business Intelligence? É caro implementar uma solução de Business Intelligence? O que fazer para reduzir os riscos nesse início?
Análise Estatística de Dados com Linguagem R	Aplicações da análise estatística de dados no mundo corporativo. Assegurando a qualidade dos dados. Analisando o comportamento estatístico de indicadores corporativos. Melhorando a qualidade de processos e produtos. Analisando a probabilidade de eventos de interesse; Análise descritiva; Análise Conclusiva.

MÓDULO INTERMEDIÁRIO

Disciplinas	Ementa
Virtualização e Computação em Nuvem	<p>Conceitos gerais de virtualização e computação em nuvem, problemas de escalonamento de máquinas virtuais, isolamento de tráfego entre aplicações, interfaces de programação para computação em nuvem. Princípios gerais e tipos de virtualização; Definições para computação em nuvem; Relação entre virtualização e nuvens; Desafios em virtualização; Modelo econômico de computação em nuvem; Controle de tráfego em ambientes virtualizados e datacenters; Computação Verde; Computação em nuvem: software, plataforma ou infraestrutura como serviço; Avaliar o relacionamento entre o Amazon Redshift e outros sistemas de big data; Avaliar casos de uso de cargas de trabalho de data warehousing e examinar implementações reais de serviços de dados e análises da AWS como parte de uma solução de data warehousing; Avaliar as abordagens e metodologias de projeto de data warehouses; Identificar fontes de dados e avaliar requisitos que afetam o projeto do data warehouse; Projetar o data warehouse para uso eficiente dos métodos de compactação, distribuição de dados e classificação; Carregar e descarregar dados e executar tarefas de manutenção de dados; Escrever consultas e avaliar planos de consulta para otimizar sua performance; Configurar o banco de dados para alocar recursos, como memória para filas de consulta, e definir critérios para rotear determinados tipos de consulta para suas filas de consulta configuradas para aprimorar o processamento; Usar um aplicativo de inteligência de negócios (BI) para executar análise de dados e tarefas de visualização nos seus dados.</p>
Preparação de Dados (ETL)	<p>Análise onde os dados são verificados em busca de erros e anomalias, Workflow onde após a verificação dos dados é constituído uma sequência de operações para tratar os erros dos dados, Validação onde a correção do workflow é avaliada em um comparativo com uma amostra representativa do conjunto de dados, Transformação onde uma vez que a eficácia do workflow é validada na amostra, é executada a transformação de todo o conjunto de dados (e o “verdadeiro” processo de data preparativo ocorre) e Substituição onde a etapa derradeira do processo de data preparativo consiste em substituir os dados falhos pelos dados “limpos”; O que é o processo ETL; Requisitos para ETL; Conceitos de Dimensões; Conceitos de Fatos; Modelagem Dimensional; Staging Área; Processo de carga; principais ferramentas de ETL.</p>

Inteligência e Modelagem na WEB	Inteligência Coletiva. Web Semântica. Ontologia. Buscas avançadas: pessoas, negócios, redes sociais. Ferramentas para coleta de dados. Coleta de dados na Web – Web Crawler (redes sociais, blogs e notícias). API's de redes sociais.
Data Mining	Fornece ao aluno o aprendizado sobre Data Mining que consiste em um processo analítico projetado para explorar grandes quantidades de dados (tipicamente relacionados a negócios, mercado ou pesquisas científicas), na busca de padrões consistentes e/ou relacionamentos sistemáticos entre variáveis e, então, validá-los aplicando os padrões detectados a novos subconjuntos de dados. O processo consiste basicamente em 3 etapas: exploração, construção de modelo ou definição do padrão e validação/verificação. A premissa do Data Mining é uma argumentação ativa, isto é, em vez do usuário definir o problema, selecionar os dados e as ferramentas para analisar tais dados, as ferramentas do Data Mining pesquisam automaticamente os mesmos a procura de anomalias e possíveis relacionamentos, identificando assim problemas que não tinham sido identificados pelo usuário.
Introdução a Inteligência Artificial com Python	Programação com Python; NumPy; DataFrames de pandas para resolver tarefas complexas; pandas para lidar com arquivos Excel; Obtenção de dados na Web; Matplotlib e Seaborn para visualizações de dados; Plotly para visualizações interativas; Regressão linear; K nearest neighbors; K Means Clustering; Árvores de decisão; Florestas aleatórias; Processamento de linguagem natural; Suporte Máquinas Vector; Algoritmos de sistemas de recomendação (NetFlix, Youtube, Amazon, Spotify), etc.

MÓDULO AVANÇADO

Disciplinas	Ementa
Modelagem MAP Reduce	Padrões de design para MapReduce. Tuning. Combine. Aplicação do Spark em problemas modelados por MapReduce.
Machine Learning	Técnicas de Machine Learn, aprendizagem de algoritmos supervisionada e a não supervisionada; métodos de regressão, classificação, clustering e afins;

Deep Learning	Fundamentos de Deep Learning; Análise de Viés em Inteligência Artificial; Aplicações com Deep Learning; Classificação, Predição e Simulação com Deep Learning; Frameworks para Redes Neurais; Limitações de Redes Neurais; Modelagem de Problemas com Deep Learning; Detecção de Anomalias com Deep Learning; Processamento de Linguagem Natural; Visão Computacional com Deep Learning
Análise de Dados com Dashboards de Negócio	Introdução a Dashboards, introdução a projetos de BI e Data Warehouse – Conceitos. Categorias dos Dashboards, estrutura e utilizando métricas e indicadores. Estudos de Gráficos e objetos Visuais: Gráficos/Textos/Alertas. Os erros mais comuns na construção de um Dashboard. Boas Práticas de Construção de Dashboards. Principais Metas para um Dashboard Visual. Desenvolvendo Dashboards Empresariais.

MÓDULO BÁSICO

Metodologia de Pesquisa Científica

Formalização de técnicas e metodologias aplicadas à pesquisa científica, com ênfase sobre a pesquisa WEB e técnicas para a escrita de Monografia, Tese, Relatórios, "Papers".

Introdução. O Conhecimento. A Ciência e suas Características. A Pesquisa Científica e o Desenvolvimento Tecnológico. O Método Científico e Suas Aplicações. Uso do Método Científico no Cotidiano Acadêmico e Profissional. Planejamento de uma Pesquisa. O Projeto de Pesquisa e sua Elaboração. A Comunicação Científica: Monografia, Tese, Relatórios, "Papers".

Introdução à Banco de Dados (Relacional e NoSql)

Os conjuntos de dados estão se tornando cada vez maiores porque os dados estão sendo recolhidos por fontes cada vez mais comuns, como celulares, drones, logs da Internet, câmeras, microfones, leitores de RFID etc. Aprender a trabalhar e estruturar base de dados abre as portas profissionais para o apoio à inteligência das organizações.

Banco de Dados Relacional; Banco de Dados Distribuídos; Bancos noSQL; Arquitetura de Dados; Abstração de Dados; Modelagem Conceitual de Dados; Fundamentos de Projeto de Banco de Dados; Projeto Conceitual de Banco de Dados; Projeto Lógico de Banco de Dados; Projeto Físico de Banco de Dados; Projeto de Banco de Dados Orientado a Objetos; Administração de Dados; Banco de Dados Não Convencionais etc.

Governança de Dados Objetivo

Estudo de métodos para gerenciar a incalculável quantidade de dados continuamente gerados, remetendo à análise e busca de respostas em relação às áreas onde o Big Data está sendo utilizado, na análise de dados históricos e padrões.

Arquitetura de Soluções Big Data; Pipeline cobrindo desde a coleta à visualização de dados; Exemplos de utilização de tecnologias ditas Big Data em diversas áreas, como por exemplo, médica, vendas, educação, financeira etc.

Fundamentos de Big Data

Estudo de métodos para gerenciar a incalculável quantidade de dados continuamente gerados remetendo à análise e busca de respostas em relação às áreas onde o Big Data está sendo utilizado, na análise de dados históricos e padrões.

Arquitetura de Soluções Big Data; Pipeline cobrindo desde a coleta à visualização de dados; Exemplos de utilização de tecnologias ditas Big Data em diversas áreas, como por exemplo, médica, vendas, educação, financeira etc.

Inteligência de Negócios Objetivo

Fornece ao aluno o aprendizado sobre o processo de coleta, organização, análise, compartilhamento e monitoramento de informações que oferecem suporte a gestão de negócios. Mostrar o conjunto de teorias, metodologias, processos, estruturas e tecnologias que transformam uma grande quantidade de dados brutos em informação útil para tomadas de decisões estratégicas.

Business Intelligence. Arquitetura e soluções de Business Intelligence. Data Warehouse. Aplicações ETL. A importância das Ferramentas de BI. As aplicações de BI front-end. BI: primeiros passos: Por que implementar um projeto de Business Intelligence? É caro implementar uma solução de Business Intelligence? O que fazer para reduzir os riscos nesse início?

Análise Estatística de Dados com Linguagem R Objetivo

Estudo da aplicação de análises estatísticas para examinar o panorama geral de um negócio, bem como diferentes mercados individualmente. Os resultados das análises estatísticas são frutos de serviços de inteligência de mercado que envolvem a coleta e o exame detalhado dos dados de um negócio.

Aplicações da análise estatística de dados no mundo corporativo. Assegurando a qualidade dos dados. Analisando o comportamento estatístico de indicadores corporativos. Melhorando a qualidade de processos e produtos. Analisando a probabilidade de eventos de interesse; Análise descritiva; Análise Conclusiva.

MÓDULO INTERMEDIÁRIO

Virtualização e computação em Nuvem

Discutir os aspectos principais das tecnologias de virtualização e computação em nuvem. Conceitos de virtualização, focando-se aspectos da computação em nuvem aplicados na distribuição de serviços, redes para datacenters e no desenvolvimento de aplicações distribuídas.

Conceitos gerais de virtualização e computação em nuvem, problemas de escalonamento de máquinas virtuais, isolamento de tráfego entre aplicações, interfaces de programação para computação em nuvem. Princípios gerais e tipos de virtualização; Definições para computação em nuvem; Relação entre virtualização e nuvens; Desafios em virtualização; Modelo econômico de computação em nuvem; Controle de tráfego em ambientes virtualizados e datacenters; Computação Verde; Computação em nuvem: software, plataforma ou infraestrutura como serviço; Avaliar o relacionamento entre o Amazon Redshift e outros sistemas de big data; Avaliar casos de uso de cargas de trabalho de data warehousing e examinar implementações reais de serviços de dados e análises da AWS como parte de uma solução de data warehousing; Avaliar as abordagens e metodologias de projeto de data warehouses; Identificar fontes de dados e avaliar requisitos que afetam o projeto do data warehouse; Projetar o data warehouse para uso eficiente dos métodos de compactação, distribuição de dados e classificação; Carregar e descarregar dados e executar tarefas de manutenção de dados; Escrever consultas e avaliar planos de consulta para otimizar sua performance; Configurar o banco de dados para alocar recursos, como memória para filas de consulta, e definir critérios para rotear determinados tipos de consulta para suas filas de consulta configuradas para aprimorar o processamento; Usar um aplicativo de inteligência de negócios (BI) para executar análise de dados e tarefas de visualização nos seus dados.

Preparação de Dados (ETL)

É o processo de coletar, limpar, normalizar, combinar, estruturar e organizar os dados para análise. É o primeiro passo, e fundamental, para que o trabalho com Big Data seja bem-sucedido, uma vez que aumenta a qualidade dos dados e conseqüentemente, dos resultados com Data Mining. Dados "pobres", de qualidade ruim, geram resultados incorretos e não confiáveis ao fim do processo de uso das tecnologias.

Análise onde os dados são verificados em busca de erros e anomalias, Workflow onde após a verificação dos dados é constituído uma seqüência de operações para tratar os erros dos dados, Validação onde a correção do workflow é avaliada em um comparativo com uma amostra representativa do conjunto de dados, Transformação onde uma vez que a eficácia do workflow é validada na amostra, é executada a transformação de todo o conjunto de dados (e o "verdadeiro" processo de data preparation ocorre) e Substituição onde a etapa derradeira do processo de data preparation consiste em substituir os dados falhos pelos dados "limpos"; O que é o processo ETL; Requisitos para ETL; Conceitos de Dimensões; Conceitos de Fatos; Modelagem Dimensional; Staging Área; Processo de carga; principais ferramentas de ETL.

Inteligência e Modelagem na WEB

O Objetivo é discutir e avaliar o uso da WEB para obtenção de dados na geração de informação para o negócio.

Inteligência Coletiva. Web Semântica. Ontologia. Buscas avançadas: pessoas, negócios, redes sociais. Ferramentas para coleta de dados. Coleta de dados na Web - Web Crawler (redes sociais, blogs e notícias). API's de redes sociais.

Data Mining Objetivo

Fornece ao aluno o aprendizado sobre Data Mining que consiste em um processo analítico projetado para explorar grandes quantidades de dados (tipicamente relacionados a negócios, mercado ou pesquisas científicas), na busca de padrões consistentes e/ou relacionamentos sistemáticos entre variáveis e, então, validá-los aplicando os padrões detectados a novos subconjuntos de dados. O processo consiste basicamente em 3 etapas: exploração, construção de modelo ou definição do padrão e validação/verificação. A premissa do Data Mining é uma argumentação ativa, isto é, em vez do usuário definir o problema, selecionar os dados e as ferramentas para analisar tais dados, as ferramentas do Data Mining pesquisam automaticamente os mesmos a procura de anomalias e possíveis relacionamentos, identificando assim problemas que não tinham sido identificados pelo usuário.

Análise Estatística de Dados, Tratamento de dados para os processos de Data Mining, O Processo de Descoberta de Conhecimento em Bancos de Dados, Data Mining, suas principais funcionalidades, técnicas e algoritmos, Análise de Associações, Classificação de Dados, Árvores de Decisão, Regressão Logística, Redes Neurais, Segmentação e Análise de Cluster, Estudos de casos com Aplicações de Data Mining a CRM Analítico, Redução de Inadimplência, Detecção de Fraude etc.

Introdução a Inteligência Artificial com Python Objetivo

Será demonstrado como usar todo o poder do Python para analisar dados, criar excelentes visualizações e usar algoritmos de Machine Learning poderosíssimos.

Programação com Python; NumPy; DataFrames de pandas para resolver tarefas complexas; pandas para lidar com arquivos Excel; Obtenção de dados na Web; Matplotlib e Seaborn para visualizações de dados; Plotly para visualizações interativas; Regressão linear; K nearest neighbors; K Means Clustering; Árvores de decisão; Florestas aleatórias; Processamento de linguagem natural; Suporte Máquinas Vector; Algoritmos de sistemas de recomendação (NetFlix, Youtube, Amazon, Spotify), etc.

MÓDULO AVANÇADO

Modelagem MAP Reduce

Fornecer a compreensão da técnica MapReduce, mostrando como ajuda no desenvolvimento de aplicações críticas que trabalham com BIG Data, que são bancos de dados difíceis de serem manipulados por tecnologias convencionais. Entretanto, a adoção de uma implementação MapReduce, como o Hadoop, por exemplo, requer habilidades específicas sobre o modelo de programação MapReduce, que introduz novas exigências na forma de modelar a abstração sobre os dados e processar BigData.

Padrões de design para MapReduce. Tuning. Combine. Aplicação do Spark em problemas modelados por MapReduce.

Machine Learning Objetivo

Fornece ao aluno o aprendizado para projetar o desenvolvimento de soluções baseadas em Machine learning. Efetuando uma análise em grandes quantidades de dados por meio de métodos estatísticos específicos, utilizando-se de técnicas variadas de algoritmos para encontrar padrões no banco de dados. Com base nesses padrões, fazer determinações ou predições.

Técnicas de Machine Learn, aprendizagem de algoritmos supervisionada e a não supervisionada; métodos de regressão, classificação, clustering e afins;

Deep Learning

Fornecer iniciativas que envolvam a modelagem, implementação e validação de sistemas inteligentes, para a solução dos mais diversos tipos de problemas atuais. Conheça as melhores alternativas para aplicação de algoritmos e modelos computacionais inteligentes.

Fundamentos de Deep Learning; Análise de Viés em Inteligência Artificial; Aplicações com Deep Learning; Classificação, Predição e Simulação com Deep Learning; Frameworks para Redes Neurais; Limitações de Redes Neurais; Modelagem de Problemas com Deep Learning; Detecção de Anomalias com Deep Learning; Processamento de Linguagem Natural; Visão Computacional com Deep Learning.

Análise de Dados com Dashboards de Negócio Objetivo

Capacitar o aluno no desenvolvimento e implementação de Dashboard com as principais informações financeiras, comerciais, recursos humanos, logísticas e dos processos de manutenção da empresa.

Introdução a Dashboards, introdução a projetos de BI e Data Warehouse - Conceitos. Categorias dos Dashboards, estrutura e utilizando métricas e indicadores. Estudos de Gráficos e objetos Visuais: Gráficos/Textos/Alertas. Os erros mais comuns na construção de um Dashboard. Boas Práticas de Construção de Dashboards. Principais Metas para um Dashboard Visual. Desenvolvendo Dashboards Empresariais.

Corpo Docente:

Professores	Titulação	Link Lattes
Alexandre Thomaz Azevedo Pereira	Mestre	http://lattes.cnpq.br/1463106992405747
Ana Caroline Torres de Carvalho	Mestre	http://lattes.cnpq.br/8923808957175080
Ana Gabriella Amorim Abreu Pereira	Doutora	http://lattes.cnpq.br/251349184131437
André Przewodowski Filho	Doutor	http://lattes.cnpq.br/3906668574849249
Carlos Henrique Faria Alves	Mestre	http://lattes.cnpq.br/5898172089878701
Claudio Marcio do Nascimento Abreu Pereira	Doutor	http://lattes.cnpq.br/2341184189645578
Geraldo Motta Azevedo Junior	Doutor	http://lattes.cnpq.br/0567535260948255
José Ricardo Costa Basílio	Mestre	http://lattes.cnpq.br/6515713884837264
Manoel Villas Boas Junior	Mestre	http://lattes.cnpq.br/2694980072582390
Rogério Pinheiro de Souza	Mestre	http://lattes.cnpq.br/0114994563959832
Coordenador	Titulação	Link Lattes
Antônio José Dias da Silva	Mestre	http://lattes.cnpq.br/8922587549718409