



Ao Tribunal de Justiça do Piauí

A/c. Sr. Gleydson Vilanova Viana

gleydson.vilanova@tjpi.jus.br

(86) 99434-5011

Ref.: Proposta Comercial

Prezado(a) Senhor(a),

É com imensa satisfação que a **Data Science Academy**, produto do CNPJ No. 13.183.890/0001-66, situada em Brasília-DF, na SHN Qd. 01, Bloco A, Entr. A, Sala 1414, telefone/fax (61) 3255-1326, e-mail: pagamento@datascienceacademy.com.br apresenta proposta comercial cujo objeto é o fornecimento de Conteúdo dos Cursos Online – Formação Cientista de Dados, Formação Inteligência Artificial e Formação Engenheiro de Dados, que consistem em cursos com aulas on-line, vídeos, arquivos de texto, pdf, e-books, imagens e sons, disponibilizadas através de arquivo eletrônico por meio do site www.datascienceacademy.com.br.

Apresentação da Empresa

A **Sucesso Tecnologia e Informação** é a detentora exclusiva no Brasil do produto Data Science Academy (DSA). Idealizada desde 2003 e fundada em 2012 a Sucesso Tecnologia, tem por premissa básica a prestação de serviços com esmero e qualidade com foco no cliente. A Sucesso Tecnologia por meio da Plataforma DSA dissemina conhecimento tecnológico nas áreas de Big Data, Ciência de Dados, Inteligência Artificial, RPA, Blockchain dentre outros para um público Brasileiro, localizado no Brasil e no Mundo, por meio de cursos e treinamentos totalmente on-line.

A Plataforma DSA de aprendizado possui ferramentas que permitem ampliar o leque de conhecimentos e habilidades pessoais e profissionais, para os alunos por meio de uso do computador, tablet ou smartphone, em qualquer lugar, a qualquer hora, no momento mais adequado para o aluno.

Nossos profissionais contam com mais de 20 anos de experiência nas áreas de Tecnologia, Informação, Big Data, Ciência de Dados, Inteligência Artificial e Realidade Virtual.

Somos credenciados ABED – Associação Brasileira de Ensino a Distância.

Com sede em Brasília- DF, a Sucesso possui profissionais alocados no Brasil e no mundo para atender uma gama de mais de 340.000 alunos atualmente cadastrados.



Nossa Equipe de Instrutores Líderes

Nós contamos com um grande número de instrutores que realizaram trabalhos sob demanda. Destacamos abaixo os instrutores líderes dos principais cursos de Ciência de Dados da DSA:

David M. – Cientista de Dados – 22 anos de experiência - graduado em Ciência da Computação com MBA em Negócios e Finanças. David teve um forte background em banco de dados, Business Intelligence e Soluções de Enterprise Performance Management, antes de migrar para Data Science. Responsável por um dos principais blogs em Data Science e Big Data do Brasil, o Ciência e Dados, David está conduzindo um dos maiores projetos de Big Data do Canadá, onde vive atualmente.

Eduardo M. – Cientista de Dados – 20 anos de experiência na área de Banco de Dados, professor da UFRJ, Coordenador Do curso do Big Data do Instituto Inffinet no Rio de Janeiro. Consultor nas áreas de Banco de Dados, Data Base Architect & DBA.

Regis E. – Prof. Doutor em Economia pela Universidade do RS. Formação em R Programming – JHU – Estados Unidos. Cientista de Modelos Macroeconômicos. Desenvolveu a árvore do Impeachment, previsão do resultado do Impeachment brasileiro.

Julio Z.- Cientista de dados: Consultor e professor de Estatística. Bacharel em Estatística (UFRGS) com Mestrado em Engenharia de Produção (UFRGS)

Suemar C. - Graduado em Ciência da Computação e Pós-graduado em Segurança da Informação em Redes de Computadores e Sistemas. Professor universitário em disciplinas de programação como C# e Android. Desenvolvedor com foco em mobile, principalmente Android. Especialista em Machine Learning e Aplicações Analytics

Marconi V. – PMP, MVP in Project. Experiência em planejamento e consultoria de gerenciamento de projetos nas indústrias de TI, petroquímica, mineração, energia, siderúrgica, automobilística e construção civil, incluindo gestão de custos, análise de riscos, implantação e administração de Sistemas de Big Data. Autor do livro: Gerenciamento de Projetos de Tecnologia da Informação.

Alguns de Nossos Clientes

- Banco do Brasil
- Serasa Experian
- Embrapa
- IBM
- ATM Digital
- PGFN
- CEF
- Procergs
- Volvo
- TOTVs
- Santander
- UNESP
- Claro
- Deloitte
- Itaú
- TCE-RS
- Serpro
- TJ-RO



Valor da Proposta para 36 Matrículas

| Descrição | Quantidade | Valor Unitário (R\$) | Valor Total (R\$) |
|---|------------|----------------------|-------------------|
| Formação Cientista de Dados - 454h/a: - Big Data Analytics com R e Microsoft Azure Machine Learning (108 h/a) - Big Data Real-Time Analytics com Python e Spark. 644 h/a) - Engenharia de Dados com Hadoop e Spark (44 h/a) - Machine Learning (120 h/a) - Business Analytics (52 h/a) - Visualização de Dados e Design de Dashboards (54 h/a) - Curso Bônus – Carreiras para Cientista de Dados (32h/a) | 16 | 2.500,00 | 40.000,00 |
| Formação Engenheiro de Dados - 304 h/a: - Design e Implementação de Data Warehouses (64 h/a) - Data Lake – Design, Projeto e Integração (72 h/a) - Segurança e Alta Disponibilidade de Dados (64 h/a) - Machine Learning e IA em Ambientes Distribuídos (64 h/a) - Analytics, Visualização, Relatórios e Tomada de Decisões com Big Data (40 h/a) - Curso Bônus – Administração Linux | 10 | 2.500,00 | 25.000,00 |
| Formação Inteligência Artificial - 448 h/a: - Introdução à Inteligência Artificial (64 h/a) - Análise em Grafos Para Big Data (40 h/a) - Deep Learning Frameworks (40 h/a) - Deep Learning I (40 h/a) - Deep Learning II (40 h/a) - Processamento de Linguagem Natural (54 h/a) - Programação paralela em GPU (40 h/a) - Sistemas Cognitivos (40 h/a) - Visão computacional (54 h/a) - Projeto Final (36 h/a) - Curso Bônus : Módulos Extras de Capacitação Profissional | 10 | 3.170,00 | 31.700,00 |
| Valor Global | 36 | | 96.700,00 |
| Noventa e Seis Mil e setecentos reais | | | |

- Declaramos expressamente que na composição de preços estamos prevendo todos os custos com mão-de-obra, tributos, despesas, encargos trabalhistas ou de qualquer natureza, resultantes da prestação dos serviços.



Metodologia do Curso

Todos nossos cursos são online. As aulas são gravadas e os vídeos são em alta definição. O aluno pode assistir os vídeos de acordo com sua disponibilidade, por meio de nossas APP's gratuitas para IOS e Android ou qualquer computador conectado à internet. Ao final de cada módulo há um arquivo PDF com os slides de todas as aulas e todos os scripts são disponibilizados passo-a-passo. Haverá um fórum exclusivo para os alunos dos cursos, com a participação de no mínimo 4 instrutores. As dúvidas serão respondidas em até 24 horas e o acesso a todo o conteúdo ficará disponível por 24 meses a partir da data de início do contrato e/ou liberação do curso. Os cursos possuem explanação teórica dentre outras ferramentas, serão mostrados na prática passo a passo. Com exemplos, quizzes, exercícios e estudo de casos, sendo possível você testar seus conhecimentos e aplicá-los.

Avaliação e Certificado do Curso

As avaliações dos cursos são independentes. Haverá uma avaliação para cada um dos cursos e uma avaliação final, apenas para os alunos que se inscreverem na Formação Completa. Cada avaliação terá 50 questões e o aluno terá 3 chances para fazer a avaliação. As questões farão referência a cada um dos capítulos estudados. A última avaliação terá cunho prático, com questões que vão requerer execução de procedimentos para que se possa obter as respostas. Ao longo dos cursos e ao fim de cada capítulo, haverá quizzes que irão permitir ao aluno treinar os modelos de questões das avaliações finais. Nosso objetivo é que o aluno aprenda de verdade. Cada curso terá seu próprio certificado, incluindo os certificados da Formação Final.

Validade da Proposta

Até 60 (sessenta) dias, contados da data de recebimento da mesma.

Entrega de Conteúdo e Dinâmica do Curso On-line

Nossos cursos e disponibilização de conteúdos são 100% on-line. O curso on-line possui uma organização e disponibilização de conteúdos de forma totalmente diferente do curso presencial. A partir do momento que o aluno está cadastrado e matriculado na Plataforma de treinamento, todo o conteúdo já estará à disposição do aluno. Portanto, ele pode concluir o curso no período que desejar. Sendo assim, nossa condição de prestação de serviço também já estará concretizada e cumprida. Pois, nós vendemos e entregamos conteúdo. O tempo de acesso aos conteúdos, a forma de estudar, a forma de acesso, o prazo de retenção de conteúdo, todas essas questões, são alheias a nossa vontade e ao nosso controle. Essa é uma nova dinâmica da sociedade moderna, que nosso país e nossas organizações terão de se adaptar. O advento da Internet trouxe essa extraordinária opção. Mas, que devemos nós também nos adaptarmos aos novos modelos de contratação e prestação de serviços.



Condições de Pagamento

Por se tratar de uma nova modalidade de curso, onde todo o conteúdo é 100% on-line, assim que o aluno estiver matriculado na plataforma de treinamento, todo o conteúdo do curso já estará à disposição do aluno. Portanto, nossa condição de prestação de serviços estará cumprida. Nós vendemos e entregamos conteúdo. As avaliações podem ser feitas a qualquer momento e cabe ao aluno realizá-las ao seu tempo, e enviá-las diretamente ao chefe da coordenação pública. A partir desse entendimento. O nosso pagamento deve ser realizado em até 30 **(trinta) dias**, contados do aceite da presente proposta e envio da nota de empenho, por meio de depósito em conta e após apresentação de Nota Fiscal.

Dados Bancários

Banco do Brasil – 001
Agência: 1887-2
Conta corrente: 37.485-7

Dados Cadastrais

Razão social: **Sucesso Tecnologia e Informação Eireli-ME**
CNPJ: **13.183.890/0001-66**
Inscrição Estadual – DF – CF/DF **07.565.417/001-20**
Inscrição Municipal - isento
OPTANTE PELO SIMPLES NACIONAL (**SIM**)
Endereço: **SHN Qd. 01, Bloco A, Entr. A, Sala 1414**
Bairro: **Asa Norte**
CEP: **70.701.010**
Telefone: **(61) 3255-1326**

Desde já agradecemos a confiança em nós depositada e nos colocamos a sua disposição para dirimir quaisquer dúvidas.

Atenciosamente,

Leticia Oliveira
Departamento Financeiro



Especificação detalhada do Objeto

Formação Cientista de Dados

Big Data deixou de ser uma palavra da moda e está se transformando em realidade. O mercado de trabalho está em busca de profissionais de Big Data com capacidades analíticas.

Fazer análise de Big Data é utilizar técnicas avançadas para tratamento de grande volume de dados. São milhares de dados sendo alimentados e armazenados em tempo real 24 horas por dia. Trata-se de registros pessoais, informações sobre saúde, trabalho, comércio, cada ação realizada por usuários através de milhões de sites na internet, celulares, sensores, transações eletrônicas de negócios e assim por diante. Nós precisamos agora de tecnologias que nos ajudem a preparar esses dados, sejam eles estruturados ou não estruturados e em diferentes tamanhos e formatos, para dar sentido a eles de forma a torná-los mais inteligentes e auxiliar as nossas decisões.

Analisar grandes volumes de dados é uma tarefa para analistas, pesquisadores e usuários de tecnologia e negócios, os quais irão tomar decisões mais acertadas e objetivas a partir desses dados tratados. Por meio de técnicas de análise avançadas, tais como análise de conteúdo textual, análise preditiva de dados, mineração, estatísticas e aplicações de Machine Learning. Esses profissionais terão acesso a dados singulares, nunca acessados anteriormente o que os permitirão sair a frente de seus concorrentes.

A Formação Cientista de Dados da Data Science Academy é um treinamento intensivo, do básico ao avançado, para sua aprendizagem nas tecnologias mais modernas em Big Data, Data Science e Analytics. A Formação é composta de 6 cursos mais um curso bônus preparação para carreira de cientista de dados, detalhados abaixo.

Os Pré-requisitos Desejáveis são: Conhecimentos básicos de sistemas operacionais. Noções de programação de computadores. Habilidade com Matemática e pensamento lógico.



Conteúdo Programático: Big Data Analytics com R e Microsoft Azure Machine Learning, Versão 2.0

01- Introdução

Bem-vindo(a)

Apresentação dos Instrutores

Formação Cientista de Dados

O Que Há de Novo na FCD 2.0?

Perguntas e Respostas

Navegando pela Data Science Academy

Suporte e Canais de Comunicação

Termos e Condições de Uso

Introdução ao Curso

O Que Esperar e o Que Não Esperar Deste Treinamento?

E Como Deve Ser Sua Abordagem Neste Curso?

Como Participar do Fórum e Comunidade DSA?

Big Data Fundamentos e Introdução à Ciência de Dados

A Carreira de Cientista de Dados

Erros Que Devem Ser Evitados Na Sua Caminhada Para se Tornar Cientista de Dados

Por que Cientistas de Dados usam R?

Vantagens e Desvantagens da Linguagem R

Por que R e Microsoft Azure Machine Learning?

Azure Machine Learning Workflow

Dados x Informação x Conhecimento x Inteligência

A Importância do Big Data Analytics - Parte 1/2

A Importância do Big Data Analytics - Parte 2/2

Data Science e a Evolução dos Sistemas Analíticos

Data Science e Inteligência Artificial - Há Diferença?

A Importância do Portfólio Para o Cientista de Dados

Instalando R, RStudio e Rtools no Windows

Instalando R e RStudio no Mac

Instalando R e RStudio no Linux

Bônus - Como Preparar um Ambiente Virtual Para Desenvolvimento em Data Science



A História da Linguagem R
Competição DSA de Machine Learning
Bibliografia, Referências e Links Úteis
Slides Módulo 01

02- Fundamentos da Linguagem R

Introdução
Conhecendo a Linguagem R
Explorando o RStudio
Definindo o Diretório de Trabalho
Primeiros Passos na Linguagem R
Instalando e Carregando Pacotes
Como Obter Help na Linguagem R
Operadores Básicos, Relacionais e Lógicos em R
Definindo Variáveis - Parte 1/2
Definindo Variáveis - Parte 2/2
Tipos Básicos de Dados - Parte 1/2
Tipos Básicos de Dados - Parte 2/2
Tipos Avançados de Dados - Parte 1/2
Tipos Avançados de Dados - Parte 2/2
Vetores, Operações com Vetores e Vetores Nomeados - Parte 1/2
Vetores, Operações com Vetores e Vetores Nomeados - Parte 2/2
Matrizes, Operações com Matrizes e Matrizes Nomeados - Parte 1/2
Matrizes, Operações com Matrizes e Matrizes Nomeados - Parte 2/2
Listas, Operações com Listas e Listas Nomeadas - Parte 1/2
Listas, Operações com Listas e Listas Nomeadas - Parte 2/2
Operações com Strings - Parte 1/2
Operações com Strings - Parte 2/2
DataFrames e Operações com DataFrame - Parte 1/2
DataFrames e Operações com DataFrame - Parte 2/2
Quiz- 5 questões
Scripts e Datasets do Capítulo
Lista de Exercícios
Bibliografia, Referências e Links Úteis
Slides Módulo 02

03- Linguagem R - Fatores, Estruturas de Controle e Funções

Introdução
Solução - Lista de Exercícios - Parte 1/2
Solução - Lista de Exercícios - Parte 2/2
Big Data na Prática 1 - Carregando 8.5 Milhões de Registros em Menos de 5 Segundos - Parte 1/2
Big Data na Prática 1 - Carregando 8.5 Milhões de Registros em Menos de 5 Segundos - Parte 2/2
Variáveis Qualitativas e Quantitativas
Fatores e Fatores Ordenados - Parte 1/3
Fatores e Fatores Ordenados - Parte 2/3
Fatores e Fatores Ordenados - Parte 3/3
Fatores e Dataframes - Compreendendo a Ordem dos Fatores
Estruturas de Controle



Trabalhando com Estruturas de Controle em R
O Que São Funções?
Trabalhando com Funções em R - Parte 1/2
Trabalhando com Funções em R - Parte 2/2
Família Apply - Uma Forma Elegante de Fazer Loops - Parte 1/3
Família Apply - Uma Forma Elegante de Fazer Loops - Parte 2/3
Família Apply - Uma Forma Elegante de Fazer Loops - Parte 3/3
Funções Especiais unlist(), do.call() e strsplit() - Parte 1/3
Funções Especiais unlist(), do.call() e strsplit() - Parte 2/3
Funções Especiais unlist(), do.call() e strsplit() - Parte 3/3
Pacotes e Instalação de Pacotes - Parte 1/2
Pacotes e Instalação de Pacotes - Parte 2/2
Criação de Pacotes em R
Expressões Regulares em R
Tabela - Expressões Regulares em R
Trabalhando com Datas - Parte 1/2
Trabalhando com Datas - Parte 2/2
Diferença Entre Operadores de Atribuição
Quiz- 5 questões
Scripts e Datasets do Capítulo
Lista de Exercícios
Bibliografia, Referências e Links Úteis
Slides Módulo 03

04- Linguagem R - Gráficos

Introdução
Solução - Lista de Exercícios - Parte 1/5
Solução - Lista de Exercícios - Parte 2/5
Solução - Lista de Exercícios - Parte 3/5
Solução - Lista de Exercícios - Parte 4/5
Solução - Lista de Exercícios - Parte 5/5
O Que é Visualização de Dados?
O que são Gráficos?
Como a Linguagem R Trata as Visualizações de Dados?
Gramática dos Gráficos
Gráficos em R - Base Plotting System - Parte 1/4
Gráficos em R - Base Plotting System - Parte 2/4
Gráficos em R - Base Plotting System - Parte 3/4
Gráficos em R - Base Plotting System - Parte 4/4
Construindo ScatterPlots
Construindo BoxPlots
Construindo Histogramas
Construindo BarPlots
Construindo Pie Charts
Explorando ggplot2 - Parte 1/3
Explorando ggplot2 - Parte 2/3
Explorando ggplot2 - Parte 3/3
Folha de Referência do ggplot2



Visualizações com Lattice
Mapas em R
Quiz- 5 questões
Big Data na Prática 2 - Visualizações Interativas
Scripts e Datasets do Capítulo
Lista de Exercícios
Bibliografia, Referências e Links Úteis
Slides Módulo 04

05- Manipulação de Arquivos TXT, CSV e Planilhas Excel em R

Introdução
Solução - Lista de Exercícios - Parte 1/3
Solução - Lista de Exercícios - Parte 2/3
Solução - Lista de Exercícios - Parte 3/30
Importação e Manipulação de Arquivos em R - Definição
Importação e Manipulação de Arquivos em R - Principais Funções
Trabalhando com Arquivos TXT - Parte 1/2
Trabalhando com Arquivos TXT - Parte 2/2
Trabalhando com Arquivos CSV - Parte 1/2
Trabalhando com Arquivos CSV - Parte 2/2
Importação e Manipulação de Planilhas Excel em R
Por Que Precisamos Configurar o Java JDK?
Instalando e Configurando o Java JDK - Parte 1/2
Instalando e Configurando o Java JDK - Parte 2/2
Mais Configurações do Sistema Operacional
Importação e Manipulação de Planilhas Excel - Parte 1/2
Importação e Manipulação de Planilhas Excel - Parte 2/2
Dicas Para Importação e Manipulação de Arquivos em Linguagem R
Big Data na Prática 3 - Mineração de Regras de Associação (Market Basket Analysis) - Parte 1/5
Big Data na Prática 3 - Mineração de Regras de Associação (Market Basket Analysis) - Parte 2/5
Alerta Para Usuários Windows
Big Data na Prática 3 - Mineração de Regras de Associação (Market Basket Analysis) - Parte 3/5
Big Data na Prática 3 - Mineração de Regras de Associação (Market Basket Analysis) - Parte 4/5
Big Data na Prática 3 - Mineração de Regras de Associação (Market Basket Analysis) - Parte 5/5
Quiz- 5 questões
Scripts e Datasets do Capítulo
Lista de Exercícios
Bibliografia, Referências e Links Úteis
Slides Módulo 05

06- Trabalhando com Bancos de Dados Relacionais e NoSQL em R

Introdução
Solução - Lista de Exercícios - Parte 1/4
Solução - Lista de Exercícios - Parte 2/4
Solução - Lista de Exercícios - Parte 3/4
Solução - Lista de Exercícios - Parte 4/4
Introdução aos Bancos de Dados Relacionais
Principais Bancos de Dados Relacionais



Conhecendo o Banco de Dados MySQL
Instalando e Configurando o Banco de Dados MySQL no Windows
Instalando e Configurando o Banco de Dados MySQL no MacOS
Instalando e Configurando o Banco de Dados MySQL no Linux
Introdução a Linguagem SQL
Quais os Benefícios da Linguagem SQL?
Como a Linguagem SQL é Processada pelo SGBD?
Tipos de Instruções SQL
Manipulação de Dados com Linguagem SQL - Parte 1/5
Manipulação de Dados com Linguagem SQL - Parte 2/5
Manipulação de Dados com Linguagem SQL - Parte 3/5
Manipulação de Dados com Linguagem SQL - Parte 4/5
Manipulação de Dados com Linguagem SQL - Parte 5/5
Carregando e Analisando Dados do MySQL com Linguagem R - Parte 1/2
Carregando e Analisando Dados do MySQL com Linguagem R - Parte 2/2
Trabalhando com R e SQLite - Parte 1/3
Trabalhando com R e SQLite - Parte 2/3
Trabalhando com R e SQLite - Parte 3/3
Introdução aos Bancos de Dados NoSQL
Principais Bancos de Dados NoSQL
Conhecendo o Banco de Dados MongoDB
Instalando e Configurando o MongoDB no Windows
Instalando e Configurando o MongoDB no MacOS
Instalando e Configurando o MongoDB no Linux
Carregando Dados no MongoDB - Parte 1/2
Carregando Dados no MongoDB - Parte 2/2
Trabalhando com R e MongoDB - Parte 1/2
Trabalhando com R e MongoDB - Parte 2/2
Big Data na Prática 4 - Customer Churn Analytics - Definindo o Problema de Negócio
Big Data na Prática 4 - Customer Churn Analytics - Compreendendo o Conjunto de Dados
Big Data na Prática 4 - Customer Churn Analytics - Carregando e Limpando os Dados
Big Data na Prática 4 - Customer Churn Analytics - Análise Exploratória e Feature Selection
Big Data na Prática 4 - Customer Churn Analytics - Modelo Preditivo com Regressão Logística
Big Data na Prática 4 - Customer Churn Analytics - Análise de Variância - ANOVA
Big Data na Prática 4 - Customer Churn Analytics - Modelo Preditivo com Árvore de Decisão
Big Data na Prática 4 - Customer Churn Analytics - Modelo Preditivo com Random Forest
Big Data na Prática 4 - Customer Churn Analytics - Relatório Final com RMarkdown - Parte 1/2
Big Data na Prática 4 - Customer Churn Analytics - Relatório Final com RMarkdown - Parte 2/2
Quiz- 5 questões
Scripts e Datasets do Capítulo
Lista de Exercícios
Bibliografia, Referências e Links Úteis
Slides Módulo 06

07- Manipulação de Dados com R

Introdução
Solução - Lista de Exercícios - Parte 1/3
Solução - Lista de Exercícios - Parte 2/3



Solução - Lista de Exercícios - Parte 3/3
O Que é Data Wrangling?
Pacotes Para Manipulação de Dados em R - Parte 1/3
Pacotes Para Manipulação de Dados em R - Parte 2/3
Pacotes Para Manipulação de Dados em R - Parte 3/3
Operador de Concatenação
Limpeza, Transformação e Manipulação de Dados em R - Parte 1/5
Limpeza, Transformação e Manipulação de Dados em R - Parte 2/5
Limpeza, Transformação e Manipulação de Dados em R - Parte 3/5
Limpeza, Transformação e Manipulação de Dados em R - Parte 4/5
Limpeza, Transformação e Manipulação de Dados em R - Parte 5/5
Estudo de Caso - Limpando, Transformando e Manipulando Dados de Voos - Parte 1/2
Estudo de Caso - Limpando, Transformando e Manipulando Dados de Voos - Parte 2/2
Remodelagem de Dados - Parte 1/3
Remodelagem de Dados - Parte 2/3
Remodelagem de Dados - Parte 3/3
Remodelando os Dados com Reshape - Parte 1/2
Remodelando os Dados com Reshape - Parte 2/2
Aplicando a Técnica Split-Apply-Combine com plyr
Utilizando o Pacote data.table - Parte 1/2
Utilizando o Pacote data.table - Parte 2/2
Subconjuntos de Dados com Subsetting - Parte 1/4
Subconjuntos de Dados com Subsetting - Parte 2/4
Subconjuntos de Dados com Subsetting - Parte 3/4
Subconjuntos de Dados com Subsetting - Parte 4/4
Estudo de Caso - Extraindo Dados da Web com Web Scraping em R - Parte 1/3
Estudo de Caso - Extraindo Dados da Web com Web Scraping em R - Parte 2/3
Estudo de Caso - Extraindo Dados da Web com Web Scraping em R - Parte 3/3
Importando Dados de Softwares Estatísticos (SAS, STATA, SPSS) - Parte 1/2
Importando Dados de Softwares Estatísticos (SAS, STATA, SPSS) - Parte 2/2
Big Data na Prática 5 - Análise de Séries Temporais no Mercado Financeiro - Parte 1/3
Big Data na Prática 5 - Análise de Séries Temporais no Mercado Financeiro - Parte 2/3
Big Data na Prática 5 - Análise de Séries Temporais no Mercado Financeiro - Parte 3/3
Quiz- 5 questões
Scripts e Datasets do Capítulo
Lista de Exercícios
Bibliografia, Referências e Links Úteis
Slides Módulo 07

08- Introdução à Análise Estatística de Dados - Parte 1

Introdução
Solução - Lista de Exercícios - Parte 1/3
Solução - Lista de Exercícios - Parte 2/3
Solução - Lista de Exercícios - Parte 3/3
O Papel da Estatística em Ciência de Dados
As 3 Grandes Áreas da Estatística
População e Amostra
Como Garantir que a Amostra Representa Fielmente a População?



Parâmetros x Estatísticas
Dados x Informação
Principais Fontes de Dados
Observações x Variáveis
Tipos de Variáveis
Exercício - Colocando os Dados em Contexto
Solução do Exercício
E-book - A Importância da Estatística para o Cientista de Dados
Quiz- 10 questões
Scripts e Datasets do Capítulo
Bibliografia, Referências e Links Úteis
Slides Módulo 08

09- Introdução à Análise Estatística de Dados - Parte 2

Introdução
Estatística Descritiva
Medidas de Tendência Central - Média e Mediana
Medidas de Tendência Central - Moda, Valores Máximo e Mínimo e Amplitude
Medidas de Posição Relativa - Quartil e Percentil
Medidas de Dispersão - Desvio Padrão e Variância - Parte 1/2
Medidas de Dispersão - Desvio Padrão e Variância - Parte 2/2
Medidas de Dispersão - Coeficiente de Variação05:07
Quiz 10 questões
Representação Gráfica dos Dados
Interpretando um Histograma - Parte 1/2
Interpretando um Histograma - Parte 2/2
Coeficiente de Assimetria (Skewness)
Coeficiente de Curtose (Kurtosis)
Interpretando um BoxPlot
Covariância e Coeficiente de Correlação - Parte 1/2
Covariância e Coeficiente de Correlação - Parte 2/2
Introdução à Probabilidade
Evento, Experimento e Espaço da Amostra
Probabilidade e Possibilidade São a Mesma Coisa?
Probabilidade Clássica
Probabilidade Empírica
Probabilidade Subjetiva
Regras Básicas da Probabilidade
Regras Básicas da Probabilidade Para Mais de Um Evento - Parte 1/3
Regras Básicas da Probabilidade Para Mais de Um Evento - Parte 2/3
Regras Básicas da Probabilidade Para Mais de Um Evento - Parte 3/30
Variáveis Aleatórias Discretas e Contínuas
Tipos de Distribuição de Probabilidade
Como Construir Uma Distribuição de Probabilidade
Distribuição Binomial
Distribuição Poisson
Distribuições Contínuas
Distribuição Normal



A Regra Empírica
Distribuição Normal e Função Densidade
Estatística na Prática - Valor Monetário Esperado (VME)
Quiz- 10 questões
Scripts e Datasets do Capítulo
Lista de Exercícios
Bibliografia, Referências e Links Úteis
Slides Módulo 09

10- Introdução à Análise Estatística de Dados - Parte 3

Introdução
Solução - Lista de Exercícios - Parte 1/3
Solução - Lista de Exercícios - Parte 2/3
Solução - Lista de Exercícios - Parte 3/3
Amostragem
Tipos de Amostragem
Descrição dos Tipos de Amostragem Probabilística
Amostragem Probabilística (Bootstrapping)
Erros de Amostragem
Teorema do Limite Central - Definição
Teorema do Limite Central - Exemplo
A Importância do Tamanho da Amostra no Teorema do Limite Central
Escore z
Intervalo de Confiança
Nível de Confiança
Valor Crítico
Teste de Hipótese - Definição
Teste de Hipótese - Aplicação
A Lógica do Teste de Hipótese
Teste de Hipótese - Exercício
Análise de Regressão - Parte 1/6
Análise de Regressão - Parte 2/6
Análise de Regressão - Parte 3/6
Análise de Regressão - Parte 4/6
Análise de Regressão - Parte 5/6
Análise de Regressão - Parte 6/6
Premissas da Análise de Regressão
Quiz- 10 questões
Scripts e Datasets do Capítulo
Lista de Exercícios
Bibliografia, Referências e Links Úteis
Slides Módulo 10

11- Machine Learning em Linguagem R

Introdução
Solução - Lista de Exercícios - Parte 1/5
Solução - Lista de Exercícios - Parte 2/5
Solução - Lista de Exercícios - Parte 3/5



Solução - Lista de Exercícios - Parte 4/5
Solução - Lista de Exercícios - Parte 5/5
Introdução ao Aprendizado de Máquina (Machine Learning)
Inteligência Artificial x Machine Learning x Deep Learning
Machine Learning x Estatística
O Que São Algoritmos?
Machine Learning Frameworks
Principais Machine Learning Frameworks
Tipos de Aprendizagem em Machine Learning
Regressão x Classificação
O Processo de Aprendizagem em Machine Learning
O Processo de Aprendizagem em Detalhes - Parte 1/3
O Processo de Aprendizagem em Detalhes - Parte 2/3
O Processo de Aprendizagem em Detalhes - Parte 3/3
Treinamento, Validação e Teste
O Que é um Modelo de Machine Learning?
Algoritmos de Machine Learning - Parte 1/2
Algoritmos de Machine Learning - Parte 2/2
A Importância da Análise Exploratória de Dados
Análise Exploratória de Dados Para Variáveis Numéricas
Análise Exploratória de Dados Para Variáveis Categóricas - Parte 1/2
Análise Exploratória de Dados Para Variáveis Categóricas - Parte 2/2
Machine Learning - Regressão - Prevendo Despesas Hospitalares - Definindo o Problema
Machine Learning - Regressão - Prevendo Despesas Hospitalares - Coleta de Dados e Análise Exploratória
Machine Learning - Regressão - Prevendo Despesas Hospitalares - Construção e Treinamento do Modelo
Machine Learning - Regressão - Prevendo Despesas Hospitalares - Fazendo Previsões com o Modelo
Machine Learning - Regressão - Prevendo Despesas Hospitalares - Avaliação e Interpretação do Modelo - Parte 1/2
Machine Learning - Regressão - Prevendo Despesas Hospitalares - Avaliação e Interpretação do Modelo - Parte 2/2
Machine Learning - Regressão - Prevendo Despesas Hospitalares - Otimização do Modelo
Machine Learning - Regressão - Prevendo Despesas Hospitalares - Documentando o Resultado Final
Machine Learning - Classificação - Prevendo a Ocorrência de Câncer - Definindo o Problema
Machine Learning - Classificação - Prevendo a Ocorrência de Câncer - Coleta de Dados e Análise Exploratória
Machine Learning - Classificação - Prevendo a Ocorrência de Câncer - Pré-Processamento de Dados
Machine Learning - Classificação - Prevendo a Ocorrência de Câncer - Normalização
Machine Learning - Classificação - Prevendo a Ocorrência de Câncer - Construindo e Treinando o Modelo com KNN
Machine Learning - Classificação - Prevendo a Ocorrência de Câncer - Avaliando e Interpretando o Modelo
Machine Learning - Classificação - Prevendo a Ocorrência de Câncer - Otimizando o Modelo
Machine Learning - Classificação - Prevendo a Ocorrência de Câncer - Construindo e Treinando o Modelo com SVM
Machine Learning - Classificação - Prevendo a Ocorrência de Câncer - Construindo e Treinando o Modelo com RandomForest
Quiz- 10 questões
Scripts e Datasets do Capítulo
Lista de Exercícios



Bibliografia, Referências e Links Úteis
Slides Módulo 11

12- Microsoft Azure Machine Learning

Introdução02:51

Solução - Lista de Exercícios - Parte 1/1

Solução - Lista de Exercícios - Parte 2/1

Solução - Lista de Exercícios - Parte 3/1

Solução - Lista de Exercícios - Parte 4/1

Solução - Lista de Exercícios - Parte 5/1

Solução - Lista de Exercícios - Parte 6/1

Solução - Lista de Exercícios - Parte 7/1

Solução - Lista de Exercícios - Parte 8/1

Solução - Lista de Exercícios - Parte 9/1

Solução - Lista de Exercícios - Parte 10/1

O que é o Microsoft Azure Machine Learning?

Azure Machine Learning Studio

Azure Machine Learning Workflow

Azure Machine Learning Toolkit

Como Criar sua Conta no Azure Machine Learning

Conhecendo o Azure Machine Learning

Explorando a Interface do Azure Machine Learning Studio

Criando Experimentos no Azure ML

Explorando os Módulos de Análise Estatística e Modelagem Preditiva

Trabalhando com Módulos no Azure ML

Importando Dados no Azure ML

Módulos de Entrada e Saída de Dados

Criando e Avaliando um Modelo de Machine Learning

Comparando a Performance de Modelos de Machine Learning

Documentação e Exemplos

Usando o Pacote Caret Para Criar Modelos de Machine Learning em R - Parte 1/2

Usando o Pacote Caret Para Criar Modelos de Machine Learning em R - Parte 2/2

Quiz- 5 questões

Scripts e Datasets do Capítulo

Lista de Exercícios

Bibliografia, Referências e Links Úteis

Slides Módulo 12

13- Data Munging no Azure Machine Learning

Introdução

Solução - Lista de Exercícios - Parte 1/2

Solução - Lista de Exercícios - Parte 2/2

Executando Script R no Azure ML

Desenvolvendo Script R Para Manipulação de Dados e Biblioteca de Utilitários

Importando Scripts R no Azure ML

Executando Experimento de Análise de Dados no Azure ML

Editando os Metadados

Tratamento de Dados Duplicados e Missing - Parte 1/3



- Tratamento de Dados Duplicados e Missing - Parte 2/3
- Tratamento de Dados Duplicados e Missing - Parte 3/3
- Transformação e Padronização
- Tratamento de Erros e Outliers
- Manipulação de Dados com Linguagem SQL no Azure Machine Learning - Parte 1/2
- Manipulação de Dados com Linguagem SQL no Azure Machine Learning - Parte 2/2
- Usando Joins para Combinar Datasets
- Usando o Pacote dplyr no Azure ML - Parte 1/2
- Usando o Pacote dplyr no Azure ML - Parte 2/2
- Instalando Pacotes R no Azure ML
- Criando Gráficos com ggplot2 no Azure ML
- Quiz- 5 questões
- Scripts e Datasets do Capítulo
- Lista de Exercícios
- Bibliografia, Referências e Links Úteis
- Slides Módulo 13

14- Análise de Regressão com Linguagem R e Azure Machine Learning

- Introdução02:42
- Solução - Lista de Exercícios - Parte 1/5
- Solução - Lista de Exercícios - Parte 2/5
- Solução - Lista de Exercícios - Parte 3/5
- Solução - Lista de Exercícios - Parte 4/5
- Solução - Lista de Exercícios - Parte 5/5
- Definindo o Problema de Negócio
- Coletando os Dados
- Data Munging - Carregando os Dados
- Data Munging - Limpeza e Transformação em R - Parte 1/2
- Data Munging - Limpeza e Transformação em R - Parte 2/2
- Data Munging - Executando o Experimento no Azure ML
- Análise de Correlação - Parte 1/2
- Análise de Correlação - Parte 2/2
- Análise Exploratória de Dados - Séries Temporais
- Análise Exploratória de Dados - BoxPlots
- Análise Exploratória de Dados - Density Plots
- O Que é Feature Selection?
- Usando Modelo RandomForest Para Seleção de Variáveis
- Filter Based Feature Selection
- Construindo e Treinando o Modelo Preditivo
- Melhorando a Performance do Modelo Preditivo
- Criando um Módulo R no Azure ML
- Treinando o Modelo
- Computação dos Resíduos do Modelo
- Interpretando os Resíduos
- Otimização do Modelo
- Storytelling
- Quiz- 5 questões
- Scripts e Datasets do Capítulo



Bibliografia, Referências e Links Úteis
Slides Módulo 14

15- Classificação com Linguagem R e Azure Machine Learning

Introdução

O Que é Classificação?

Principais Características de Classificação

Performance dos Modelos de Classificação - Confusion Matrix

Performance dos Modelos de Classificação - Medidas de Performance

Definindo o Problema de Negócio

Coletando os Dados

Convertendo Variáveis Categóricas

Nomeando as Variáveis

Aplicando Engenharia de Atributos em Variáveis Numéricas

Aplicando Quantization em Variáveis Numéricas

Análise Exploratória de Dados - Parte 1/2

Análise Exploratória de Dados - Parte 2/2

Balanceamento do Dataset - Synthetic Minority Oversampling Technique (SMOTE)

Feature Selection (Seleção de Variáveis)

Filter Based Feature Selection

Permutation Feature Importance

Construindo o Experimento

Construindo o Modelo - Classificador Bayesiano x Redes Neurais x SVM

Treinamento dos Modelos de Machine Learning

Manual de Interpretação de Modelos de Classificação

Como Interpretar Modelos de Classificação

Avaliando os Modelos

Comprando os Modelos

Criando o Modelo de Machine Learning em R

Avaliando o Modelo - Valores Observados x Valores Previstos

Gerando Curva ROC em Linguagem R

Otimizando o Modelo

Recomendações Sobre Otimização

Quiz- 5 questões

Scripts e Datasets do Capítulo

Bibliografia, Referências e Links Úteis

Slides Módulo 15

16- Publicação Online do Modelo Preditivo

Introdução

O Que é Deploy de Modelos de Machine Learning?

Publicando seu Modelo Azure Machine Learning como um Serviço Web

Deploy do Modelo Preditivo

Slides Módulo 16

17- Mini-Projeto - Análise de Sentimentos em Redes Sociais

Como Usar os Mini-Projetos?

O Que é Análise de Sentimentos?



Por Que Analisar Dados de Redes Sociais?
Especificação do Mini-Projeto
Configurando a API na Rede Social
Proposta de Solução

18- Mini-Projeto - Análise de Risco de Crédito

Especificação do Mini-Projeto
Proposta de Solução

19- Mini-Projeto - Mapeando a Ocorrência do Vírus Zika em um Gráfico Interativo

Especificação do Mini-Projeto
Proposta de Solução

20- Projetos Com Feedback

Projeto 1 - Detecção de Fraudes no Tráfego de Cliques em Propagandas de Aplicações Mobile
Projeto 2 - Prevendo Demanda de Estoque com Base em Vendas

21- Bônus

Instalando o Oracle R Enterprise
Instalando o Microsoft R

22- Avaliação Final e Certificado de Conclusão

Avaliação Final -50 questões



Conteúdo Programático: Big Data Real-Time Analytics com Python e Spark, Versão 2.0

1- Introdução

Bem-vindo(a)

Apresentação dos Instrutores

Formação Cientista de Dados

O Que Há de Novo na FCD 2.0?

Perguntas e Respostas

Navegando pela Data Science Academy

Suporte e Canais de Comunicação

Termos e Condições de Uso

Introdução ao Curso

O que é o Apache Spark?

Principais Benefícios do Apache Spark

Por que Aprender Apache Spark?

Por que Python e Spark?

Apache Spark Workflow

Real-Time Analytics e Computação Distribuída

O Que é Streaming de Dados?

Processamento de Lotes (Batch) x Processamento de Streaming de Dados

Apache Spark x Apache Hadoop - Parte 1/2

Apache Spark x Apache Hadoop - Parte 2/2

Preparando o Ambiente Python e Spark

Instalando Anaconda Python e Apache Spark no Windows - Parte 1/3

Instalando Anaconda Python e Apache Spark no Windows - Parte 2/3

Instalando Anaconda Python e Apache Spark no Windows - Parte 3/3

Instalando Anaconda Python e Apache Spark no MacOS

Instalando Anaconda Python e Apache Spark no Linux

Competição DSA de Machine Learning

Bibliografia, Referências e Links Úteis

Slides Módulo 01



02- Manipulação de Dados com Python

Introdução

Por Que Linguagem Python Para Análise de Dados?

Principais Módulos Python Para Análise de Dados - Parte 1/2

Principais Módulos Python Para Análise de Dados - Parte 2/2

Preparando o Ambiente de Análise de Dados

Computação Numérica com NumPy - Parte 1/3

Computação Numérica com NumPy - Parte 2/3

Computação Numérica com NumPy - Parte 3/3

Indexação e Slicing com NumPy - Parte 1/3

Indexação e Slicing com NumPy - Parte 2/3

Indexação e Slicing com NumPy - Parte 3/3

Transposição e Reshaping de Arrays

Manipulação de Dados com Pandas - DataFrames

Manipulação de Dados com Pandas - Operações com DataFrames - Parte 1/2

Manipulação de Dados com Pandas - Operações com DataFrames - Parte 2/2

Manipulação de Dados com Pandas - Objetos Multidimensionais - Parte 1/2

Manipulação de Dados com Pandas - Objetos Multidimensionais - Parte 2/2

Exercício - Analisando Dados de Cidades Digitais - Group By e Reshape

Quiz- 5 questões

O que é Pré-Processamento?

Scripts e Datasets do Capítulo

Bibliografia, Referências e Links Úteis

Slides Módulo 02

03- Organização e Visualização de Dados

Introdução

Solução Exercício - Analisando Dados de Cidades Digitais - Group By e Reshape - Parte 1/3

Solução Exercício - Analisando Dados de Cidades Digitais - Group By e Reshape - Parte 2/3

Solução Exercício - Analisando Dados de Cidades Digitais - Group By e Reshape - Parte 3/3

Funções Map, Filter e Reduce - Parte 1/2

Funções Map, Filter e Reduce - Parte 2/2

List Comprehension e Função Lambda

Operações com Arrays NumPy

Concatenação, Join e Split de Arrays NumPy

Operações Matemáticas com Arrays e Matrizes - Parte 1/2

Operações Matemáticas com Arrays e Matrizes - Parte 2/2

Vetorização de Funções e Operações com Arrays NumPy

Operações com Pandas - Series - Indexação

Operações com Pandas - Series - Valores Missing

Operações com Pandas - Series - Concatenação de Series e Tratamento de Valores Missing

Operações com Pandas - Series - Operações e Cópia

Operações com Pandas - DataFrames - Indexação - Parte 1/2

Operações com Pandas - DataFrames - Indexação - Parte 2/2

SQL Join e Junção de Tabelas com Pandas - Parte 1/3

SQL Join e Junção de Tabelas com Pandas - Parte 2/3

SQL Join e Junção de Tabelas com Pandas - Parte 3/3

Estudo de Caso - Análise Exploratória de Dados em Arquivos CSV - Parte ¼



Estudo de Caso - Análise Exploratória de Dados em Arquivos CSV - Parte 2/4
Estudo de Caso - Análise Exploratória de Dados em Arquivos CSV - Parte 3/4
Estudo de Caso - Análise Exploratória de Dados em Arquivos CSV - Parte 4/4
Configuração e Customização Avançada do Matplotlib - Parte 1/4
Configuração e Customização Avançada do Matplotlib - Parte 2/4
Configuração e Customização Avançada do Matplotlib - Parte 3/4
Configuração e Customização Avançada do Matplotlib - Parte 4/4
Construindo um Dashboard com Matplotlib
Gráficos Estatísticos em Python com Seaborn - Parte 1/2
Gráficos Estatísticos em Python com Seaborn - Parte 2/2
Construindo Gráficos com ggplot em Python
Quiz- 5 questões
Scripts e Datasets do Capítulo
Lista de Exercícios
Bibliografia, Referências e Links Úteis
Slides Módulo 03

04- Análise Estatística de Dados - Parte 1

Introdução
Solução Lista de Exercícios - Parte 1/2
Solução Lista de Exercícios - Parte 2/2
Definindo Estatística
Tipos de Dados
Observação x Experimentação
Principais Áreas da Estatística
Estatística Descritiva
Tabela de Frequência
Distribuição de Frequência
Estudo Dirigido - Construindo Distribuição de Frequência em Python - Parte 1/2
Estudo Dirigido - Construindo Distribuição de Frequência em Python - Parte 2/2
Ferramentas Oferecidas Pela Estatística Descritiva
Ferramentas Para Análise Univariada - Parte 1/2
Ferramentas Para Análise Univariada - Parte 2/2
Ferramentas Para Análise Bivariada
Medidas de Tendência Central
Medidas de Dispersão
Medidas de Forma - Skewness e kurtosis
Coeficiente de Correlação
Estudo Dirigido - Estatística Descritiva em Python - Parte 1/2
Estudo Dirigido - Estatística Descritiva em Python - Parte 2/2
Scripts e Datasets do Capítulo
Quiz- 5 questões
Bibliografia, Referências e Links Úteis
Slides Módulo 04

05- Análise Estatística de Dados - Parte 2

Introdução
Introdução à Probabilidade



Experimento Aleatório
Tipos de Probabilidade
Regras da Probabilidade
Eventos e Espaço Amostral
O Fenômeno Aleatório
Probabilidade Empírica e a Lei dos Grandes Números
Um Pouco Mais Sobre os Tipos de Probabilidade
Probabilidade Conjunta
Probabilidade Condicional e Independência - Parte 1/2
Probabilidade Condicional e Independência - Parte 2/2
Independente ou Mutuamente Exclusivo?
Tabela de Contingência
Distribuições de Probabilidade - Variável Aleatória
Distribuições de Probabilidade Discreta
Distribuições de Probabilidade Contínua
A Distribuição Normal
Como Determinar Se a Distribuição é Normal?
Teorema do Limite Central
Árvores de Probabilidade
O Que é Estatística Inferencial?
População e Amostra
Princípios da Amostragem
Tipos de Amostragem
Testes de Hipótese
Teste de Hipótese Unilateral
Teste de Hipótese Bilateral
Erros Tipo I e Tipo II
Intervalo de Confiança e Significância Estatística
Intervalos de Confiança para Desvio Padrão da População Conhecido
Caso de Uso
Quiz- 10 questões
Bibliografia, Referências e Links Úteis
Slides Módulo 05

06- Machine Learning em Python

Introdução
Algoritmos de Machine Learning
Aprendizagem Supervisionada
Aprendizagem Não Supervisionada
Aprendizagem Por Reforço
Principais Algoritmos de Machine Learning - Parte 1/2
Principais Algoritmos de Machine Learning - Parte 2/2
O Processo de Construção de Modelos de Machine Learning
Soluções de Machine Learning
Deep Learning x Machine Learning x Inteligência Artificial
Template Para Construção de Modelos de Machine Learning
Definição do Problema de Negócio
Extraindo e Carregando os Dados



Análise Exploratória de Dados - Estatística Descritiva
Análise Exploratória de Dados - Correlação
Análise Exploratória de Dados - Matplotlib
Análise Exploratória de Dados - Seaborn
O Que é Normalização e Quando Aplicar?
O Que é Padronização e Quando Aplicar?
Normalização x Padronização
A Importância do Pré-Processamento dos Dados
Pré-Processamento - Normalização
Pré-Processamento - Padronização e Binarização
O Que é Feature Selection?
Feature Selection - Seleção Univariada
Feature Selection - Eliminação Recursiva (Recursive Feature Elimination - RFE)
Feature Selection - Método Ensemble para Seleção de Variáveis
Redução de Dimensionalidade com Principal Component Analysis (PCA)
Resampling e Divisão em Dados de Treino/Teste
Cross Validation
Métricas Para Modelos de Classificação - Acurácia, ROC, Confusion Matrix
Algoritmos de Classificação - Regressão Logística, LDA, Naive Bayes, KNN, CART, SVM
Seleção do Modelo Preditivo
Otimização do Modelo - Ajuste de Hyperparâmetros
Grid Search Parameter Tuning e Random Search Parameter Tuning
Salvando e Carregando o Modelo Treinado
O Que São Métodos Ensemble - Bagging, Boosting e Voting?
Métodos Ensemble - Bagged Decision Trees, RandomForest, AdaBoost, Gradient Boosting
Classificação com Algoritmo XGBoost
Definição do Problema de Negócio
Métricas Para Modelos Modelos de Regressão
Principais Métricas Para Modelos Modelos de Regressão - MSE, MAE, R2
Algoritmos de Regressão - Regressão Linear, Ridge, Lasso, ElasticNET, KNN, CART, SVM
Seleção, Avaliação e Otimização do Modelo
Quiz- 5 questões
Scripts e Datasets do Capítulo
Desafio - Pipeline com PCA e Regressão Logística
Bibliografia, Referências e Links Uteis
Slides Módulo 06

07- Processando Big Data com Apache Spark

Introdução
Solução - Desafio - Pipeline com PCA e Regressão Logística - Parte 1/2
Solução - Desafio - Pipeline com PCA e Regressão Logística - Parte 2/2
Apache Spark e Big Data
Ecossistema e Componentes do Apache Spark - Parte 1/3
Ecossistema e Componentes do Apache Spark - Parte 2/3
Ecossistema e Componentes do Apache Spark - Parte 3/3
Quando Devemos Usar o Spark?
Principais Características do Apache Spark
Medidas de Tendência Central



- Introdução ao PySpark
- Executando Aplicação PySpark
- Operação de MapReduce com PySpark
- Hadoop MapReduce x Apache Spark
- O Processo de MapReduce no Apache Spark
- Profissionais Que Trabalham com Apache Spark
- Anatomia de Uma Aplicação Spark
- Arquitetura Spark
- Spark Modes
- Deploy Mode e Fontes de Dados
- RDD's - Resilient Distributed Datasets - Parte 1/2
- RDD's - Resilient Distributed Datasets - Parte 2/2
- O que são transformações?
- Principais Operações de Transformação
- Transformações - PySpark10:35
- Operações Set, Outer Join e Distinct
- Transformação e Limpeza
- O Que São Ações?
- Ações - PySpark
- Mini-Projeto 1 - Analisando Dados do Uber com Spark - Especificação
- Mini-Projeto 1 - Analisando Dados do Uber com Spark - Proposta de Solução
- Scripts e Datasets do Capítulo
- Bibliografia, Referências e Links Úteis
- Slides Módulo 07

08- Apache Spark SQL

- Introdução
- Computação em Nuvem e em Cluster - Parte 1/2
- Computação em Nuvem e em Cluster - Parte 2/2
- Linguagem SQL - Introdução
- Linguagem SQL - Instruções SQL
- Linguagem SQL - Terminologia
- O Que é o Apache Spark SQL?
- Componentes do Apache Spark SQL
- Ajustando a Versão do Java JDK
- Spark SQL - Spark Session e SQL Context
- Spark SQL - RDD's e a Função Row
- Spark SQL - Criando Dataframes
- Spark SQL - Aplicando SQL a Tabelas Temporárias
- Spark SQL e Arquivos CSV
- Spark SQL e Machine Learning - Parte 1/2
- Spark SQL e Machine Learning - Parte 2/2
- Spark SQL com Arquivos JSON
- Spark SQL e Tabelas Temporárias
- Instalando e Configurando o Banco de Dados MySQL no Windows
- Instalando e Configurando o Banco de Dados MySQL no MacOS
- Instalando e Configurando o Banco de Dados MySQL no Linux
- Spark SQL com Banco de Dados Relacional - Instalando o Driver JDBC



- Spark SQL com Banco de Dados Relacional - Criando o Banco de Dados e Importando os Dados
- Spark SQL com Banco de Dados Relacional - Manipulando os Dados com Spark SQL
- Instalando e Configurando o MongoDB no Windows
- Instalando e Configurando o MongoDB no MacOS
- Instalando e Configurando o MongoDB no Linux
- Criando Banco de Dados no MongoDB e Conectando com PySpark
- Ajustando a Versão do Conector
- Leitura dos Dados do MongoDB com Spark
- Gravação dos Dados no MongoDB com Spark
- Agrupamento com Pares RDD's
- Acumuladores e Broadcast
- Particionamento de RDD's
- Quiz- 5 questões
- Scripts e Datasets do Capítulo
- Bibliografia, Referências e Links Úteis
- Slides Módulo 08

09- Real-Time Analytics com Spark Streaming

- Introdução
- Batch x Streaming
- Apache Spark Streaming
- Streaming de Dados - A Velocidade com que Você passa o Cartão de Crédito
- Arquitetura Apache Spark Streaming
- O que são DStreams (Discretized Streams)?
- Windowing - Agregando Stream de Dados ao Longo do Tempo
- Tolerância a Falhas
- Integração com Outros Sistemas
- Integração com Outros Sistemas - Apache Kafka
- Integração com Outros Sistemas - Apache Flume
- Integração com Outros Sistemas - AWS Kinesis
- Introdução ao Processamento de Linguagem Natural
- NLTK - Processamento de Linguagem Natural em Python
- Quiz- 5 questões
- Como o Uber Utiliza Spark e Hadoop
- Scripts e Datasets do Capítulo
- Bibliografia, Referências e Links Úteis
- Slides Módulo 09

10- Análise de Sentimentos do Twitter em Tempo Real com Spark Streaming e NLTK

- Introdução
- Criando Aplicação no Twitter - Parte 1/4
- Criando Aplicação no Twitter - Parte 2/4
- Criando Aplicação no Twitter - Parte 3/4
- Criando Aplicação no Twitter - Parte 4/4
- Gerando um Streaming de Dados em Tempo Real e Coletando com Spark Streaming - Parte 1/2
- Gerando um Streaming de Dados em Tempo Real e Coletando com Spark Streaming - Parte 2/2
- Mini-Projeto 2 - Análise de Dados do Twitter em Tempo Real
- Análise de Sentimentos de Streaming de Dados do Twitter em Tempo Real - Parte 1/6



Análise de Sentimentos de Streaming de Dados do Twitter em Tempo Real - Parte 2/6
Análise de Sentimentos de Streaming de Dados do Twitter em Tempo Real - Parte 3/6
Análise de Sentimentos de Streaming de Dados do Twitter em Tempo Real - Parte 4/6
Análise de Sentimentos de Streaming de Dados do Twitter em Tempo Real - Parte 5/6
Análise de Sentimentos de Streaming de Dados do Twitter em Tempo Real - Parte 6/6
Visualização de Dados e Apresentação do Resultado com D3.js
Scripts e Datasets do Capítulo
Bibliografia, Referências e Links Úteis

11- Apache Spark Machine Learning

Introdução

Machine Learning com Apache Spark

Analytics e Dataficação

Tipos de Analytics

Análise Exploratória de Dados

Análise Preditiva

Trade-off Entre Viés e Variância - Parte 1/2

Trade-off Entre Viés e Variância - Parte 2/2

Spark Machine Learning Library

Spark MLLib - Regressão Linear - Compreendendo o Problema de Negócio e Carregando os Dados

Spark MLLib - Regressão Linear - Limpeza do Dataset e Remoção de Valores Missing

Spark MLLib - Regressão Linear - Análise Exploratória e Análise de Correlação

Spark MLLib - Regressão Linear - Pré-Processamento de Dados, Vetores Densos e Esparsos

Spark MLLib - Regressão Linear - Criação, Treinamento e Avaliação do Modelo de Machine Learning

Spark MLLib - Classificação - Decision Tree - Problema de Negócio, Limpeza e Análise de Correlação

Spark MLLib - Classificação - Decision Tree - Pré-Processamento, Criação, Treino e Avaliação do Modelo

Spark MLLib - Classificação - Random Forest - Problema de Negócio, Limpeza e Análise de Correlação

Spark MLLib - Classificação - Random Forest - Pré-Processamento, Redução de Dimensionalidade e String Indexer

Spark MLLib - Classificação - Random Forest - Construção, Treino e Teste do Classificador

Spark MLLib - Classificação - Naive Bayes - Definição do Problema de Classificação de Spam

Spark MLLib - Classificação - Naive Bayes - Processamento de Linguagem Natural

Spark MLLib - Classificação - Naive Bayes - Pipeline, Criação, Treinamento e Avaliação do Modelo

Spark MLLib - Clustering - K-Means - Aprendizagem Não Supervisionada

Spark MLLib - Clustering - K-Means - Criação, Treinamento e Avaliação do Modelo

Mini-Projeto 3 - Sistema de Recomendação

Sistema de Recomendação - Parte 1/2

Sistema de Recomendação - Parte 2/2

Deploy dos Modelos de Machine Learning

Quiz- 5 questões

Scripts e Datasets do Capítulo

Bibliografia, Referências e Links Úteis

12- Projetos Com Feedback

Projeto 3 - Prevendo o Nível de Satisfação dos Clientes do Santander

Projeto 4 - Prevendo Customer Churn em Operadoras de Telecom

13- Bônus



Usando R e Spark com SparkR
Streaming de Dados do Twitter com Scala e Spark Streaming

14- Avaliação Final e Certificado de Conclusão

Avaliação Final 50 questões



Conteúdo Programático: Engenheiro de Dados com Hadoop

01- Introdução

- Bem-vindo(a)
- Apresentação dos Instrutores
- Formação Cientista de Dados
- O Que Há de Novo na FCD 2.0?
- Perguntas e Respostas
- Navegando pela Data Science Academy
- Suporte e Canais de Comunicação
- Termos e Condições de Uso
- Introdução ao Curso
- O que é o Apache Hadoop?
- Uma Breve História do Apache Hadoop
- Quais os benefícios para as Empresas ao utilizar o Hadoop?
- Ecosistema Hadoop
- Projetos Principais do Ecosistema Hadoop
- Apache HDFS - Conceito e Importância
- Apache HDFS - Arquitetura
- Definindo MapReduce
- Hadoop x Bancos de Dados Relacionais
- Competição DSA de Machine Learning
- Bibliografia, Referências e Links Úteis



02- Instalando o Ecossistema Hadoop - Parte 1

Introdução

Por que Cientistas de Dados Precisam Conhecer o Hadoop?

Modos de Execução do Hadoop

Softwares Usados para o Ambiente Virtual

Instalando o Ambiente Virtual - Parte 1/5

Instalando o Ambiente Virtual - Parte 2/5

Instalando o Ambiente Virtual - Parte 3/5

Instalando o Ambiente Virtual - Parte 4/5

Instalando o Ambiente Virtual - Parte 5/5

Instalando Utilitários do Sistema Operacional

Instalando MySQL Server (Será usado nos testes de importação para o Hadoop)

Instalando e Configurando o SSH

Instalando e Configurando a Máquina Virtual Java - Parte 1/2

Instalando e Configurando a Máquina Virtual Java - Parte 2/2

Bibliografia, Referências e Links Úteis

03- Instalando o Ecossistema Hadoop - Parte 2

Introdução

Não Esqueça do Backup

Instalação do Apache Hadoop - Parte 1/6

Instalação do Apache Hadoop - Parte 2/6

Instalação do Apache Hadoop - Parte 3/6

Instalação do Apache Hadoop - Parte 4/6

Instalação do Apache Hadoop - Parte 5/6

Instalação do Apache Hadoop - Parte 6/6

Troubleshooting - Alerta de Aprendizado

Inicializando o HDFS

Processando Big Data com Hadoop - Parte 1/3

Processando Big Data com Hadoop - Parte 2/3

Processando Big Data com Hadoop - Parte 3/3

Configurando e Inicializando o YARN

Executando Job MapReduce com Gerenciamento Pelo YARN - Parte 1/2

Executando Job MapReduce com Gerenciamento Pelo YARN - Parte 2/2

Ecossistema Hadoop

Instalando e Configurando o Zookeeper

Instalando e Configurando o HBase

Instalando e Configurando o Hive

Instalando e Configurando o Pig

Instalando e Configurando o Spark

Instalando e Configurando o Sqoop

Instalando e Configurando o Flume

Hadoop - Completo

Instalando e Configurando Máquina Virtual Cloudera

Instalando e Configurando Máquina Virtual Hortonworks

Apache Hadoop com Containers Docker

Quiz- 5 questões

Link Para Download do Material Deste Capítulo



Bibliografia, Referências e Links Úteis

04- Planejando e Configurando um Cluster Hadoop

Introdução

O que é um Cluster?

Arquitetura do Cluster Hadoop

Customizando o Ambiente Virtual Data Science Academy

Topologia de Rede para o Cluster Hadoop

Workflow no Cluster Hadoop

Planejamento do Cluster Hadoop

Hardware e Configuração de Rede do Cluster Hadoop

Arquivos de Configuração

Parâmetros de Configuração

HDFS - Hadoop Distributed File System

HDFS Writes e HDFS Reads

Trabalhando com o HDFS - Parte 1/3

Trabalhando com o HDFS - Parte 2/3

Trabalhando com o HDFS - Parte 3/3

Importando Dados do MySQL para o HDFS - Parte 1/2

Importando Dados do MySQL para o HDFS - Parte 2/2

Bibliografia, Referências e Links Úteis

Quiz- 5 questões

Slides Módulo 3

05- Usando MapReduce em Grandes Volumes de Dados

Introdução

Configurando o Ambiente Virtual DSA para Executar Jobs MapReduce em Python - Parte 1/3

Configurando o Ambiente Virtual DSA para Executar Jobs MapReduce em Python - Parte 2/3

Configurando o Ambiente Virtual DSA para Executar Jobs MapReduce em Python - Parte 3/3

Computação Distribuída e Computação Paralela

Computação Distribuída - Cloud Computing

O Modelo de Programação MapReduce

Map, Shuffle e Reduce

Como o MapReduce Utiliza a Computação Distribuída

YARN

Arquivos do Capítulo

Job MapReduce para Gerar Média de Amigos em Rede Social por Idade - Parte 1/2

Job MapReduce para Gerar Média de Amigos em Rede Social por Idade - Parte 2/2

Data Mining com MapReduce em Dados Não Estruturados - Parte 1/3

Data Mining com MapReduce em Dados Não Estruturados - Parte 2/3

Data Mining com MapReduce em Dados Não Estruturados - Parte 3/3

Analisando Logs de Servidores Web no Cloudera - Parte 1/2

Analisando Logs de Servidores Web no Cloudera - Parte 2/2

Criando Multi Node Cluster com Amazon Elastic MapReduce (EMR) - Parte 1/5

Criando Multi Node Cluster com Amazon Elastic MapReduce (EMR) - Parte 2/5

Criando Multi Node Cluster com Amazon Elastic MapReduce (EMR) - Parte 3/5

Criando Multi Node Cluster com Amazon Elastic MapReduce (EMR) - Parte 4/5

Criando Multi Node Cluster com Amazon Elastic MapReduce (EMR) - Parte 5/5



Bibliografia, Referências e Links Úteis

Quiz- 5 questões

Slides Módulo 4

06- Armazenamento de Dados com Hbase e Hive

Checkpoint

Introdução

Comandos Hbase - Parte 1

Trabalhando com Hbase no Cloudera - Parte 1

Trabalhando com Hive no Hortonworks - Parte 1

Conhecendo o Apache Hbase

Bancos de Dados NoSQL

Modelo de Dados Hbase

Comandos Hbase - Parte 2

Trabalhando com Hbase no Cloudera - Parte 2

Arquitetura Hbase

Importando Dados para o Hbase com Pig - Parte 1/3

Importando Dados para o Hbase com Pig - Parte 2/3

Importando Dados para o Hbase com Pig - Parte 3/3

Comandos Hbase - Parte 3

Trabalhando com Hbase no Cloudera - Parte 3

Operações CRUD no Hbase com API Java - Parte 1/2

Operações CRUD no Hbase com API Java - Parte 2/2

Introdução ao Hive

Trabalhando com Hive no Hortonworks - Parte 2

Trabalhando com Hive no Hortonworks - Parte 3

Bibliografia, Referências e Links Úteis

Quiz- 5 questões

Slides Módulo 5

07- Conectividade ETL com o Sistema Hadoop

O que é ETL (Extract, Transform and Load)?

Qual o Papel do ETL no Big Data?

Preparando o Ambiente Oracle 12c - Parte 1/6

Preparando o Ambiente Oracle 12c - Parte 2/6

Preparando o Ambiente Oracle 12c - Parte 3/6

Preparando o Ambiente Oracle 12c - Parte 4/6

Preparando o Ambiente Oracle 12c - Parte 5/6

Preparando o Ambiente Oracle 12c - Parte 6/6

Manual de Instalação Oracle 12c

Carregando 15 Milhões de Registros no Banco de Dados Oracle 12c

Apache Sqoop

Movendo Dados do Oracle para o Hadoop HDFS - Parte 1/2

Movendo Dados do Oracle para o Hadoop HDFS - Parte 2/2

Mini-Projeto1

ETL Tools Benchmarking

Hortonworks Tutorial - Analisando Dados de Sensores

Hortonworks Tutorial - Analisando Clickstream Data



Bibliografia, Referências e Links Úteis

Quiz- 5 questões

Slides Módulo 6

08- Administração e Manutenção do Hadoop

Introdução

NameNode e Estrutura de Diretórios

A Importância do Secondary NameNode

DataNodes e Estrutura de Diretórios

Metadados do FileSystem

Procedimento de Checkpoint04:31

Procedimento de Recuperação à falha do Namenode

Modo de Segurança

Backup - Parte 1/2

Backup - Parte 2/2

Solução de Problemas no Cluster Hadoop

Autenticação e Segurança Hadoop

Melhores Práticas de Monitoramento do Cluster Hadoop

Usando o Apache Ambari para Monitorar o Cluster Hadoop

Hortonworks Tutorial - Analisando Arquivos de Log

Hortonworks Tutorial - Analisando Dados de Geolocalização

Hortonworks Tutorial - Análise de Sentimento com Apache Hive

Bibliografia, Referências e Links Úteis

Quiz- 5 questões

Slides Módulo 7

09- Hadoop Machine Learning com Mahout

Introdução ao Mahou

Apache Mahout x Outros Frameworks de Machine Learning

Machine Learning - Algoritmos de Classificação

Instalação e Configuração do Apache Mahout

Cloudera - Naive Bayes - Parte 1/2

Cloudera - Naive Bayes - Parte 2/2

Cloudera - Random Forest

Modelos Ocultos de Markov (Hidden Markov Model)

Perceptrons Multi-Camadas (MLP - Multi Layer Perceptrons)

Cloudera - K Means Clustering - Parte 1/2

Cloudera - K Means Clustering - Parte 2/2

Bibliografia, Referências e Links Úteis

Quiz- 5 questões

Slides Módulo 8

10- Hadoop e Spark

Introdução

Hadoop e Spark Juntos

Anatomia de Uma Aplicação Spark

Deploy do Spark em um Cluster Hadoop com YARN - Parte ¼

Deploy do Spark em um Cluster Hadoop com YARN - Parte 2/4



Deploy do Spark em um Cluster Hadoop com YARN - Parte 3/4

Deploy do Spark em um Cluster Hadoop com YARN - Parte 4/4

Deploy do Cluster Spark no Amazon EC2 - Parte 1/3

Deploy do Cluster Spark no Amazon EC2 - Parte 2/3

Deploy do Cluster Spark no Amazon EC2 - Parte 3/3

RDD's, cache() e persist()

Spark SQL

Spark MLLib

Spark MLLib x Apache Mahout

Spark Streaming

Spark GraphX

Bibliografia, Referências e Links Úteis

Quiz- 5 questões

Slides Módulo 9

11- Mini-Projeto 1 - Importando Dados do Banco de Dados Oracle para o HDFS

Especificação do Mini-Projeto 1

Proposta de Solução

12- Mini-Projeto 2 - Prevendo Casos de Doenças Cardíacas

13- Mini-Projeto 3 - Design de um Job MapReduce com Spark para os Gastos Totais por Cliente

Especificação do Mini-Projeto 3

Proposta de Solução

14- Projetos Com Feedback

Projeto 5 - Implementando Slow Changing Dimensions em um DW Usando Hive e Spark

Projeto 6 - Processamento de Logs em Tempo Real Com Flume, Spark Streaming e Hbase

15- Bônus

Microsoft Azure HDInsight - Parte 1

Microsoft Azure HDInsight - Parte 2

Configuração e Tuning do Cloudera Hadoop

Fazendo o Deploy do Cluster Hadoop no AWS

16- Avaliação

Base de Conhecimento do Curso

Avaliação Final- 50 questões



Conteúdo Programático: Machine Learning

01- Introdução

Bem-vindo(a)
Formação Cientista de Dados
Perguntas e Respostas
Navegando Pela Data Science Academy
Suporte e Canais de Comunicação
Termos e Condições de Uso
Introdução
Lista de Mini-Projetos
Lista de Projetos Especiais
Lista de Projetos com Feedback
O que é Aprendizado de Máquina? - Parte 1/3
O que é Aprendizado de Máquina? - Parte 2/3
O que é Aprendizado de Máquina? - Parte 3/3
Por que Machine Learning Está Transformando o Mundo?
Que ferramentas usaremos neste curso?
Processo de Aprendizagem
Tipos de Aprendizagem
Aprendizagem Supervisionada
Aprendizagem Não Supervisionada
Aprendizado Indutivo
Reinforcement Learning
Treinamento, Validação e Teste
O que é um Modelo?
Classificação
The Dark Side of Big Data
Regressão
Clustering
Machine Learning é Dividir para Conquistar
Como Selecionar o Algoritmo Ideal para Cada Problema
E-book - Inteligência Artificial como Interface
Instalando R, RStudio e Rtools no Windows



- Instalando R e RStudio no Mac
- Instalando R e RStudio no Linux
- Instalando Anaconda Python no Windows
- Instalando Anaconda Python no Mac
- Instalando Anaconda Python no Linux
- Conhecendo o Seu Sistema Operacional - Windows 10 - Parte 1/2
- Conhecendo o Seu Sistema Operacional - Windows 10 - Parte 2/2
- Conhecendo o Seu Sistema Operacional - MAC - Parte 1/2
- Conhecendo o Seu Sistema Operacional - MAC - Parte 2/2
- Conhecendo o Seu Sistema Operacional - Linux - Parte 1/2
- Conhecendo o Seu Sistema Operacional - Linux - Parte 2/2
- Slides do Módulo 1
- Competição DSA de Machine Learning
- Bibliografia, Referências e Links Úteis

02- Algoritmos de Machine Learning e Modelos Preditivos

- Introdução
- Exemplo de Utilização de Modelos Preditivos
- O que é um Modelo Preditivo
- Aprendizagem Baseada em Instância
- Aprendizagem Baseada em Métodos Probabilísticos
- Aprendizagem Baseada em Procura
- Aprendizagem Baseada em Otimização
- Redes Neurais Artificiais - Parte 1/3
- Redes Neurais Artificiais - Parte 2/3
- Redes Neurais Artificiais - Parte 3/3
- Máquinas de Vetores de Suporte (SVM)
- Clustering
- Métodos Ensemble
- Bagging, Gradient Boosting e Adaboost - Parte 1/2
- Bagging, Gradient Boosting e Adaboost - Parte 2/2
- Redução de Dimensionalidade
- Modelos Lógicos, Geométricos e Probabilísticos
- Deep Learning
- Simulação
- Modelos Determinísticos x Modelos Estocásticos
- Otimização e Gradient Descent
- Slides do Módulo 2
- Quiz- 5 questões
- Bibliografia, Referências e Links Úteis

03- Como Funciona a Aprendizagem de Máquina

- Introdução
- Introdução ao Processo de Aprendizagem
- Teste de Hipóteses - Definindo H0 e H1
- Teste de Hipóteses - Testes Paramétricos
- Teste de Hipóteses - Erro Tipo I e Erro Tipo II
- Teste de Hipóteses - Valor-p



Processo de Aprendizagem - Visualizando um Algoritmo de Machine Learning
Processo de Aprendizagem - Elementos Essenciais do Processo de Aprendizagem
Processo de Aprendizagem - Componentes do Processo de Aprendizagem - Parte 1/2
Processo de Aprendizagem - Componentes do Processo de Aprendizagem - Parte 2/2
Processo de Aprendizagem - Modelo de Aprendizagem - Parte 1/2
Processo de Aprendizagem - Modelo de Aprendizagem - Parte 2/2
Machine Learning é Aprendizado a partir de Dados
Cost Function
Gradiente Descendente (Com Exemplo em Python)
Underfitting x Overfitting
Bias e Variance
E-book - O Teste de Turing
Slides do Módulo 3
Quiz- 5 questões
Bibliografia, Referências e Links Úteis

04- Regressão

Introdução
O que é Regressão? - Parte 1/2
O que é Regressão? - Parte 2/2
Avaliando o Modelo de Regressão: SST, SSE, SSR, MSE, MSA e RMSE - Parte 1/2
Avaliando o Modelo de Regressão: SST, SSE, SSR, MSE, MSA e RMSE - Parte 2/2
Modelo de Regressão Linear Simples - Explorando os Dados
Modelo de Regressão Linear Simples - Estatística com StatsModels - Parte 1/2
Modelo de Regressão Linear Simples - Estatística com StatsModels - Parte 2/2
Modelo de Regressão Linear Simples - Modelo com Scikit-Learn
Modelo de Regressão Linear Simples - Minimizando a Cost Function - Parte 1/2
Modelo de Regressão Linear Simples - Minimizando a Cost Function - Parte 2/2
Modelo de Regressão Linear Simples - Aplicando Gradiente Descendente
Regressão Linear Simples x Regressão Linear Múltipla, Estatística F e Valor-p
Modelo de Regressão Linear Múltipla - Matriz de Correlação
Modelo de Regressão Linear Múltipla - Multicolinearidade (Autovalores e Autovetores)
Modelo de Regressão Linear Múltipla - Gradiente Descendente
Modelo de Regressão Linear Múltipla - Importância de Atributos e R Square
Regularização LASSO, Ridge
Regressão Logística
Implementando Regressão Logística em Linguagem R - Parte 1/3
Implementando Regressão Logística em Linguagem R - Parte 2/3
Implementando Regressão Logística em Linguagem R - Parte 3/3
Vantagens e Desvantagens da Regressão
Tratamento de Colinearidade em Modelos de Regressão
Regressão Linear Múltipla em Linguagem Julia
Desafio DSA
Slides do Módulo 4
Quiz- 5 questões
Bibliografia, Referências e Links Úteis

05- K-Nearest Neighbours (KNN)



Introdução

KNN e Estrutura de Células de Voronoi

Como funciona o Algoritmo KNN?

Classificação KNN em Python - Carregando o Dataset de Dígitos

Classificação KNN em Python - Pré-Processamento e Normalização

Classificação KNN em Python - Construção do Modelo e Redução de Dimensionalidade

Classificação KNN em Python - Alterando Parâmetros do Modelo

Classificação KNN em R - Pacote Caret - Parte 1/2

Classificação KNN em R - Pacote Caret - Parte 2/2

Vantagens e Desvantagens KNN

E se der empate na votação do KNN?

Slides do Módulo 5

Quiz- 5 questões

Bibliografia, Referências e Links Úteis

06- Decision Tree, Random Forest e Métodos Ensemble

Introdução

O que são Árvores de Decisão?

Ganho de Informação, Entropia, Índice Gini e Pruning - Parte 1/4

Ganho de Informação, Entropia, Índice Gini e Pruning - Parte 2/4

Ganho de Informação, Entropia, Índice Gini e Pruning - Parte 3/4

Ganho de Informação, Entropia, Índice Gini e Pruning - Parte 4/4

Compreendendo a heurística ID3

Construindo uma Árvore de Decisão em R e Interpretando o Resultado - Parte 1/2

Construindo uma Árvore de Decisão em R e Interpretando o Resultado - Parte 2/2

Pruning da Árvore de Decisão em R

Calculando a Entropia e Construindo Modelo com C5.0 em R - Parte 1/2

Calculando a Entropia e Construindo Modelo com C5.0 em R - Parte 2/2

Construindo uma Árvore de Decisão em Python baseada em Entropia

Random Forest Classifier em Python - Parte 1/3

Random Forest Classifier em Python - Parte 2/3

Random Forest Classifier em Python - Parte 3/3

Random Forest Regressor em Python - Parte 1/2

Random Forest Regressor em Python - Parte 2/2

Métodos Ensemble - Parte 1/2

Métodos Ensemble - Parte 2/2

Bagging, ExtraTrees e AdaBoost em Python - Parte 1/2

Bagging, ExtraTrees e AdaBoost em Python - Parte 2/2

Otimização dos Parâmetros com Randomized Search em Python - Parte 1/2

Otimização dos Parâmetros com Randomized Search em Python - Parte 2/2

Grid Search x Randomized Search para Estimação dos Hyperparâmetros

Gradient Boosting Model

Gradient Boosting Classifier em Python

Gradient Boosting Regressor em Python

Stochastic Gradient Boosting em Python - Parte 1/2

Stochastic Gradient Boosting em Python - Parte 2/2

eXtreme Gradient Boosting (XGBoost) em R - Parte 1/2

eXtreme Gradient Boosting (XGBoost) em R - Parte 2/2



Slides do Módulo 6
Quiz- 5 questões
Bibliografia, Referências e Links Úteis

07- Naive Bayes

Introdução
Teorema de Bayes
Construindo um Classificador de Spam em R - Parte 1/2
Construindo um Classificador de Spam em R - Parte 2/2
Gaussian Naive Bayes em Python - Introdução
Gaussian Naive Bayes em Python - Machine Learning em Astronomia com AstroML
Gaussian Naive Bayes em Python - MeshGrid
Multinomial Naive Bayes em Python - Classificador de Notícias - Parte 1/2
Multinomial Naive Bayes em Python - Classificador de Notícias - Parte 2/2
Bernoulli Naive Bayes em Python
Construindo um Classificador Naive Bayes
Melhorando o Poder do Modelo Naive Bayes
Vantagens e Desvantagens do Naive Bayes
Bibliografia - Referências - Links Uteis
Slides do Módulo 7
Quiz- 5 questões

08- Clustering

Introdução
O que é Clustering? - Parte 1/2
O que é Clustering? - Parte 2/2
Tipos de Clustering - Hierárquico e Particional
Algoritmo K-Means
Algoritmo K-Means++
Escolhendo o Número Adequado de Clusters e Método Elbow
K-Means em Linguagem R - Parte 1/4
K-Means em Linguagem R - Parte 2/4
K-Means em Linguagem R - Parte 3/4
K-Means em Linguagem R - Parte 4/40
K-Means em Python - Visualização dos Clusters
K-Means em Python - Agrupando Clientes por Consumo de Energia - Parte 1/2
K-Means em Python - Agrupando Clientes por Consumo de Energia - Parte 2/2
Density-Based Spatial Clustering of Applications with Noise (DBSCAN) em Python
Análise de Cluster com Mean Shift em Python
Vantagens e Desvantagens do K-Means
Extra - Pipelines Para Automatização de Workflows de Machine Learning
Extra - Algoritmos Genéticos
Slides do Módulo 8
Quiz- 5 questões
Bibliografia - Referências - Links Uteis

09- Support Vector Machines

Introdução



- O que são SVM's?
- Dados Linearmente Separáveis
- Dados Não Linearmente Separáveis
- Modelo SVM em Python - Kernel Linear
- Modelo SVM em Python - Kernel RBF
- Teoria do Aprendizado Estatístico - Parte 1/2
- Teoria do Aprendizado Estatístico - Parte 2/2
- SVM's Lineares x SVM's Não Lineares
- Métodos Kernel
- Funções de Kernel
- Kernel Trick
- SVM's com Margens Rígidas x SVM's com Margens Flexíveis
- Modelo SVM em R - Construindo o Modelo e Definindo Parâmetros do Kernel
- Modelo SVM em R - Ajuste e Otimização do Modelo
- Modelo SVM em R - Caso de Uso - Parte 1 - Pré-Processamento
- Modelo SVM em R - Caso de Uso - Parte 2 - Criação do Modelo e Feature Selection
- Modelo SVM em R - Caso de Uso - Parte 3 - Grid Search e Otimização do Modelo
- SVC e SVR em Python
- Parâmetros C, Gamma e Kernel
- Otimizador RBF x Otimizador Linear
- Reconhecimento de Imagens com SVM em Python
- Vantagens e Desvantagens das SVM's
- Slides do Módulo 9
- Quiz- 5 questões
- Bibliografia, Referências e Links Úteis

10- Processamento de Linguagem Natural

- Introdução
- Análise Morfológica, Sintática, Semântica e Pragmática - Parte 1/2
- Análise Morfológica, Sintática, Semântica e Pragmática - Parte 2/2
- Python - Arquivos do Capítulo
- Boas Práticas em Processamento de Linguagem Natural - Parte 1/2
- Boas Práticas em Processamento de Linguagem Natural - Parte 2/2
- Expressões Regulares
- NLTK Python - Tokenization, Stopwords, Collocations, Stemming e Corpus - Parte 1/3
- NLTK Python - Tokenization, Stopwords, Collocations, Stemming e Corpus - Parte 2/3
- NLTK Python - Tokenization, Stopwords, Collocations, Stemming e Corpus - Parte 3/3
- NLTK Python - Bag of Words
- Construindo um Classificador SVM com Bag of Words Para um Site de E-commerce
- NLTK Python - Similaridade de Texto
- Usando Word2vec Para Computar Similaridades entre Vetores
- NLTK Python - Wordnets
- NLTK Python - Chunking e Parsing
- NLTK Python - Reconhecimento de Padrões
- OpenNLP em R - Parte 1/2
- OpenNLP em R - Parte 2/2
- Text Analytics em R e C++ - Parte 1/2
- Text Analytics em R e C++ - Parte 2/2



Bibliografia, Referências e Links Úteis

Slides do Módulo 10

Quiz- 5 questões

11- Redes Neurais Artificiais

Introdução às Redes Neurais Artificiais

O que são Redes Neurais Artificiais?

Redes Neurais Artificiais - Origem e Evolução

O Neurônio Biológico

O Neurônio Matemático

A Arquitetura de Redes Neurais Artificiais

Processo de Aprendizagem de Uma Rede Neural

Perceptron

Todos os Arquivos Python

Perceptron e Funções de Ativação

Adaline e Regra Delta

Perceptron em Python - Parte 1/2

Perceptron em Python - Parte 2/2

Aprendizado pelo Gradient Descent e Stochastic Gradient Descent

Perceptron com Scikit-Learn06:29

Perceptrons de Múltiplas Camadas (Multilayer Perceptron - MLP) - Parte 1/2

Perceptrons de Múltiplas Camadas (Multilayer Perceptron - MLP) - Parte 2/2

Multilayer Perceptron com Scikit-Learn - Parte 1/3

Multilayer Perceptron com Scikit-Learn - Parte 2/3

Multilayer Perceptron com Scikit-Learn - Parte 3/3

O Algoritmo Backpropagation

Backpropagation e Função Sigmóide em Python

Multilayer Perceptron com Neurolab em Python

Redes Neurais Artificiais em R - Parte 1/2

Redes Neurais Artificiais em R - Parte 2/2

Bibliografia, Referências e Links Úteis

Slides do Módulo 11

Quiz- 5 questões

12- Introdução ao Deep Learning

Introdução

O que é Deep Learning?

Como Funciona Deep Learning?

Onde Deep Learning Está Sendo Utilizado?

Deep Neural Networks

Redes Neurais Profundas e o Córtex Visual

Convolutional Neural Networks

Recurrent Neural Networks

LSTM - Long Short-Term Memory

Função Softmax

Hot Encoding

Stochastic Gradient Descent

Momentum e Learning Rate



Regularização e Dropout

Deep Learning Frameworks - Theano, TensorFlow e Keras

Deep Learning em Python com Keras - Parte 1/3

Deep Learning em Python com Keras - Parte 2/3

Deep Learning em Python com Keras - Parte 3/3

Deep Learning em Python - Redes Convolucionais com Tensorflow - Parte 1/2

Deep Learning em Python - Redes Convolucionais com Tensorflow - Parte 2/2

Deep Learning em R - Parte 1/2

Deep Learning em R - Parte 2/2

Otimização com Gradiente Descendente e Stochastic Gradiente Descendente em Linguagem R

Bibliografia, Referências e Links Úteis

Slides do Módulo 12

Quiz- 5 questões

13- Sistemas de Recomendação

Introdução

O que são Sistemas de Recomendação?

Tipos de Sistemas de Recomendação

Sistema de Recomendação Baseado no Item Mais Popular

Associação e Modelos Market Basket

Filtros Colaborativos

Filtragem de Conteúdo

Modelos Híbridos

Sistema de Recomendação em Linguagem R - Parte 1/3

Sistema de Recomendação em Linguagem R - Parte 2/3

Sistema de Recomendação em Linguagem R - Parte 3/3

Evolução dos Sistemas de Recomendação

Sistemas de Recomendação - Neighborhood-based

Sistemas de Recomendação - Mathematical-based

Sistemas de Recomendação - Machine Learning-based

Sistema de Recomendação em Python - Parte 1/3

Sistema de Recomendação em Python - Parte 2/3

Sistema de Recomendação em Python - Parte 3/3

Sistema de Recomendação com Recommender Lab

Bibliografia, Referências e Links Úteis

Slides do Módulo 13

Quiz- 5 questões

14- Mini-Projeto 1 - Implementando um Classificador de Spam com Naive Bayes

Especificação - Implementando um Classificador de Spam com Naive Bayes

Proposta de Solução

15- Mini-Projeto 2 - Construindo um Sistema de Recomendação de Filmes

Proposta de Solução

16- Mini-Projeto 3 - Criando um Modelo de Machine Learning para Retorno Sobre Investimentos

Especificação - Criando um Modelo de Machine Learning para Retorno Sobre Investimentos

Proposta de Solução



17- Mini-Projeto 4 - Otimizando o Sistema de Voos de uma Companhia Aérea

Especificação - Otimizando o Sistema de Voos de uma Companhia Aérea

Proposta de Solução

18- Mini-Projeto 5 - Análise para Prever a Força do Real em Relação a Outras Moedas

Especificação - Análise para Prever a Força do Real em Relação a Outras Moedas

Proposta de Solução

19- Bonus - Deploy do Modelo de Machine Learning em Produção com App Web

Projeto

Introdução

Definindo o Projeto

Construindo e Treinando o Modelo de Deep Learning

Módulo para Carregar o Modelo em Formato Json

Design da Interface WEB

Construindo a Aplicação WEB

Usando o Modelo de Deep Learning para Previsões em Tempo Real Via Aplicação Web

Fazendo o Deploy em Produção em Nuvem na AWS

Publicando a Aplicação Web e Fazendo Previsões em Tempo Real

Configurando o DNS Para a Aplicação Web

20- Projetos Com Feedback

Projeto 7 - Medicina Personalizada - Redefinindo o Tratamento de Câncer

Projeto 8 - Modelagem Preditiva em IoT - Previsão de Uso de Energia

21- Revisão e Avaliação Final

Revisão - Parte 1

Revisão - Parte 2

Revisão - Parte 3

Mini-Projetos

Avaliação Final- 50 questões



Conteúdo Programático: Business Analytics

01- Introdução

Bem-vindo(a)

Formação Cientista de Dados

Perguntas e Respostas

Navegando Pela Data Science Academy

Suporte e Canais de Comunicação

Termos e Condições de Uso

Introdução

Lista de Mini-Projetos

Lista de Projetos Especiais

Lista de Projetos com Feedback

O que é Business Analytics - Parte 1/2

O que é Business Analytics - Parte 2/2

O que é Data Mining?

Big Data, Métricas e Classificação de Dados

Metodologia Analítica - Parte 1/2

Metodologia Analítica - Parte 2/2

Definindo um Problema de Dados

Business Analytics - Processos e Práticas

Business Analytics x Business Intelligence

Instalando R, RStudio e Rtools no Windows

Instalando R e RStudio no Mac

Instalando R e RStudio no Linux

Instalando Anaconda Python no Windows

Instalando Anaconda Python no Mac

Instalando Anaconda Python no Linux

Conhecendo o Seu Sistema Operacional - Windows 10 - Parte 1/2

Conhecendo o Seu Sistema Operacional - Windows 10 - Parte 2/2

Conhecendo o Seu Sistema Operacional - MAC - Parte 1/2

Conhecendo o Seu Sistema Operacional - MAC - Parte 2/2

Conhecendo o Seu Sistema Operacional - Linux - Parte 1/2



Conhecendo o Seu Sistema Operacional - Linux - Parte 2/2

Slides Módulo 1

Competição DSA de Machine Learning

Bibliografia, Referências e Links úteis

02- Analytics

Introdução

Análise Descritiva

Análise Descritiva em R - Parte 1/4

Análise Descritiva em R - Parte 2/4

Análise Descritiva em R - Parte 3/4

Análise Descritiva em R - Parte 4/4

Definindo as Estatísticas

Análise Descritiva x Análise Diagnóstica

Análise Preditiva x Análise Prescritiva

Análise Preditiva - Modelando Relacionamento e Tendência nos Dados

Distribuições de Probabilidade - Parte 1/3

Distribuições de Probabilidade - Parte 2/3

Distribuições de Probabilidade - Parte 3/3

Análise de Regressão - Pré-Processamento e Análise Exploratória

Análise de Regressão - Construção do Modelo e Análise de Variância (ANOVA)

Análise de Regressão - Inferência Estatística e Análise Diagnóstica

Análise Preditiva em R - Pré-Processamento e Feature Selection

Análise Preditiva em R - Árvores de Decisão

Análise Preditiva em R - Regressão Logística

Análise Preditiva em R - Redes Neurais

Teste de Hipóteses - Parte 1/2

Teste de Hipóteses - Parte 2/2

Teste de Hipótese em R

Simulação de Monte Carlo

Simulação de Monte Carlo em Python - Projetando o Preço de Ativos

E-book - Todo Business é Analytics

Slides Módulo 02

Quiz5 questões

Bibliografia, Referências e Links úteis

03- Marketing Analytics

Introdução

Definindo Marketing Analytics

O que são Sistemas CRM?

Caso de Uso - Problema de Negócio

Marketing Analytics em R - Todos os Scripts

Marketing Analytics em R Parte 1 - Coleta de Dados de Marketing e Análise Exploratória

Marketing Analytics em R Parte 2 - Segmentação Estatística de Clientes

Marketing Analytics em R Parte 3 - Segmentação Hierárquica de Clientes

Marketing Analytics em R Parte 4 - Segmentação Gerencial

Marketing Analytics em R Parte 5 - Calculando o Faturamento por Segmento

Marketing Analytics em R Parte 6 - Construindo o Modelo Preditivo



Marketing Analytics em R Parte 7 - Aplicando o Modelo e Prevendo Faturamento
Marketing Analytics em R Parte 8 - Previsão de Clientes por Segmento
Marketing Analytics em R Parte 9 - Life Time Value
O que é um Teste A/B?
Quanto Tempo Deve Durar um Teste A/B?
Avaliando um Teste A/B em R Parte 1 - Carregando os Dados para Avaliar o Teste A/B
Avaliando um Teste A/B em R Parte 2 - Selecionando o Teste Estatístico (t-test ou z-test)
Avaliando um Teste A/B em R Parte 3 - Definindo as Hipóteses Nula e Alternativa
Avaliando um Teste A/B em R Parte 4 - Executando o z-test
Avaliando um Teste A/B em R Parte 5 - Determinando e Avaliando o Resultado
Avaliando um Teste A/B em R Parte 6 - Intervalos de Confiança
Gerando uma Calculadora para Testes A/B em R com Shiny
Avaliando Teste A/B em Python
Softwares para Testes A/B
Extra 1 - Linguagem R - Identificando Clientes Alvo para Ações de Vendas
Extra 2 - Linguagem R - Analisando Tráfego de Web Sites
Mini-Projeto 1 - Analisando o Resultado de uma Campanha de Marketing com R
E-book - O Futuro do Marketing no Business Analytics
Slides Módulo 03
Quiz- 5 questões
Bibliografia, Referências e Links úteis

04- RH Analytics

Introdução
Definindo o People Analytics - Como o Google Revolucionou a Gestão de Pessoas
Como o People Analytics Está Sendo Usado?
People Analytics e o Big Data
Riscos do People Analytics
Benefícios e Limites do People Analytics
Como Implementar People Analytics?
Por que o People Analytics é Importante para a Gestão de Pessoas?
Construindo Modelo Preditivo - Quais Funcionários Vão Deixar a Empresa? - Parte 1/6
Construindo Modelo Preditivo - Quais Funcionários Vão Deixar a Empresa? - Parte 2/6
Construindo Modelo Preditivo - Quais Funcionários Vão Deixar a Empresa? - Parte 3/6
Construindo Modelo Preditivo - Quais Funcionários Vão Deixar a Empresa? - Parte 4/6
Construindo Modelo Preditivo - Quais Funcionários Vão Deixar a Empresa? - Parte 5/6
Construindo Modelo Preditivo - Quais Funcionários Vão Deixar a Empresa? - Parte 6/6
Slides Módulo 04
Quiz- 5 questões
Bibliografia, Referências e Links úteis

05- Financial Analytics

Introdução
Financial Analytics - Finanças no Século XXI
Financial Analytics e os Processos de Negócios
Financial Analytics e Tecnologia Analítica
Financial Analytics e Machine Learning - Parte 1/2
Financial Analytics e Machine Learning - Parte 2/2



Gestão Baseada em Financial Analytics
A Revolução das Fintechs
Principais Relatórios Financeiros
Definindo Métricas Financeiras
Conceitos Básicos de Balanced Scorecard
Análise de Séries Temporais - Tratamento de Datas em R - Parte 1/4
Análise de Séries Temporais - Tratamento de Datas em R - Parte 2/4
Análise de Séries Temporais - Tratamento de Datas em R - Parte 3/4
Análise de Séries Temporais - Tratamento de Datas em R - Parte 4/4
Análise de Séries Temporais - Modelagem Financeira com quantmode em R - Parte 1/4
Análise de Séries Temporais - Modelagem Financeira com quantmode em R - Parte 2/4
Análise de Séries Temporais - Modelagem Financeira com quantmode em R - Parte 3/4
Análise de Séries Temporais - Modelagem Financeira com quantmode em R - Parte 4/4
Análise de Séries Temporais - Usando o Pacote zoo em R
Análise de Séries Temporais - Usando o Pacote xts em R
Modelando e Prevendo Séries Temporais Lineares com Modelos ARIMA em R - Parte 1/2
Modelando e Prevendo Séries Temporais Lineares com Modelos ARIMA em R - Parte 2/2
Análise Financeira, Liquidez e Análise DuPont
Otimização de Portfólio
Capital Asset Pricing Model
Mini-Projeto 2 - Implementando um Modelo Analítico para Análise de Portfólio
Slides Módulo 05
Quiz- 5 questões
Bibliografia, Referências e Links úteis

06- Fraud Analytics (Análise para Detecção de Fraudes)

Introdução
Conceito de Fraude - Parte 1/2
Conceito de Fraude - Parte 2/2
Principais Tipos de Fraudes
Detecção de Fraude e Prevenção
Big Data e Detecção de Fraudes
Ciclo de Vida do Modelo de Análise de Fraude - Parte 1/3
Ciclo de Vida do Modelo de Análise de Fraude - Parte 2/3
Ciclo de Vida do Modelo de Análise de Fraude - Parte 3/3
Inteligência Artificial para Detecção de Fraudes
Sumário Executivo - Fraudes em Transações de Cartões de Crédito
Detecção de Fraudes em R - Compreendendo o Problema e os Dados
Detecção de Fraudes em R - Análise Exploratória - Parte 1/2
Detecção de Fraudes em R - Análise Exploratória - Parte 2/2
Detecção de Fraudes em R - Definindo as Regras de Negócio
Detecção de Fraudes em R - Análise Preditiva para Detecção de Fraudes - Parte 1/2
Detecção de Fraudes em R - Análise Preditiva para Detecção de Fraudes - Parte 2/2
Detecção de Fraudes em R - Definindo o Fraud Detection Rate
Detecção de Fraudes em R - Random Forest
Detecção de Fraudes em R - Redes Neurais
Detecção de Fraudes em R - Avaliando o Modelo Preditivo
Detecção de Fraudes em R - Métricas



Apresentação Executiva
Detecção de Fraude Online com R e Microsoft Azure
E-book - Cientistas de Dados Especializados em Detecção de Fraude
Mini-Projeto 3 - Criando um Sistema de Detecção de Fraudes Para um Site de E-Commerce
Slides Módulo 06
Quiz- 5 questões
Bibliografia, Referências e Links úteis

07- Text Analytics

Introdução
O Processo de Descoberta de Conhecimento 04:33
Text Analytics em Dados Estruturados - Parte 1/2
Text Analytics em Dados Estruturados - Parte 2/2
Text Analytics em Dados Não-Estruturados
O Processo de Text Analytics
O Processo de Text Analytics - Análise Semântica
O Processo de Text Analytics - Análise Estatística
O Processo de Text Analytics - Preparação dos Dados - Parte 1/3
O Processo de Text Analytics - Preparação dos Dados - Parte 2/3
O Processo de Text Analytics - Preparação dos Dados - Parte 3/3
O Processo de Text Analytics - Indexação e Normalização - Parte 1/3
O Processo de Text Analytics - Indexação e Normalização - Parte 2/3
O Processo de Text Analytics - Indexação e Normalização - Parte 3/3
O Processo de Text Analytics - Cálculo da Relevância
O Processo de Text Analytics - Seleção de Termos
Modelos de Linguagem e Modelo Oculto de Markov
Métodos Quantitativos em Linguística
Pacotes R Para Text Mining
Text Analytics em R - Carregando os Arquivos
Text Analytics em R - Convertendo Arquivos PDF em TXT no R
Text Analytics em R - Criação do Corpus e Pré-processamento
Text Analytics em R - Processamento Estatístico
Text Analytics em R - Relacionamentos e Análise Quantitativa
Text Analytics em R - Visualização de Dados e Sumário dos Resultados
Bonus - Topic Modeling em R
Slides Módulo 07
Quiz- 5 questões
Bibliografia, Referências e Links úteis

08- Social Network Analytics

Introdução
Mídias Sociais e Sua Importância
Desafios em Social Network Analytics
Técnicas de Social Media Analytics
O Processo de Social Media Analytics
O Processo de Social Media Analytics - Autenticação
O Processo de Social Media Analytics - Visualização e Exploração
O Processo de Social Media Analytics - Limpeza e Pré-Processamento - Parte 1/2



O Processo de Social Media Analytics - Limpeza e Pré-Processamento - Parte 2/2
O Processo de Social Media Analytics - Pré-Processamento de Texto
O Processo de Social Media Analytics - Modelagem de Dados e Análise de Sentimentos
O Processo de Social Media Analytics - Visualizando o Resultado
Criando App no Twitter
Análise de Redes Sociais - Twitter - Coleta, Limpeza e Transformação
Análise de Redes Sociais - Twitter - Merge e Apresentação
Análise de Redes Sociais - Twitter - TweetBot
Análise de Redes Sociais - Twitter Analytics - Coletando Dados Offline
Análise de Redes Sociais - Twitter Analytics - Análise Exploratória07:28
Análise de Redes Sociais - Twitter Analytics - Testes Estatísticos06:19
Análise de Redes Sociais - Twitter Analytics - Visualizando os Resultados08:43
Análise de Redes Sociais - Facebook - Facebook Developer05:00
Análise de Redes Sociais - Facebook - Criando a Aplicação no Facebook
Análise de Redes Sociais - Facebook - Instalando o Pacote Rfacebook
Análise de Redes Sociais - Facebook - Gerando Autenticação de 2 Horas
Análise de Redes Sociais - Facebook - Gerando Autenticação de 2 Meses
Análise de Redes Sociais - Facebook - Coletando Dados do Facebook
Análise de Redes Sociais - Facebook - Social Network Analysis e pacote igraph
Análise de Redes Sociais - Facebook - Degree, Betweenness, Closeness e Cluster
Mini-Projeto 4 - Analisando Dados do Facebook
Análise de Redes Sociais - Facebook - Coletando Páginas, Posts, Likes e Comentários
Análise de Redes Sociais - Facebook - Usando Linguagem SQL para Minerar os Dados do Facebook
Análise de Redes Sociais - Facebook - Avaliando a Performance de uma Página
Análise de Redes Sociais - Instagram - Criando e Registrando a App
Análise de Redes Sociais - Instagram - Instalando Pacotes e Testando a Conexão
Análise de Redes Sociais - Instagram - Usando End-points e Coletando Dados
Análise de Redes Sociais - Instagram - Coleta de Dados
Análise de Redes Sociais - Instagram - Analisando Seguidores, Likes e Comentários
Análise de Redes Sociais - Instagram - Analisando Dados Por Localidade
Análise de Redes Sociais - Instagram - Wordcloud com Dados do Instagram
Análise de Redes Sociais - Instagram - Identificando Clusters de Usuários - Parte 1/2
Análise de Redes Sociais - Instagram - Identificando Clusters de Usuários - Parte 2/2
Mini-Projeto 5 - Construindo um Sistema de Recomendação
Análise de Redes Sociais - Instagram - Sistema de Recomendação - Parte 1/2
Análise de Redes Sociais - Instagram - Sistema de Recomendação - Parte 2/2
Slides Módulo 08
Quiz5 questões
Bibliografia, Referências e Links úteis

09- Projetos Com Feedback

Projeto 9 - Prevendo Tendências Macroeconômicas

Projeto 10 - Sistema de Recomendação Para Rede de Varejo Usando Market Basket Analysis

10- Avaliação Final e Certificado de Conclusão

Avaliação- 50 questões



Conteúdo Programático: Visualização de Dados e Design de Dashboards

01- Introdução

Bem-vindo(a)

Formação Cientista de Dados

Perguntas e Respostas

Navegando pela Data Science Academy

Suporte e Canais de Comunicação

Termos e Condições de Uso

Introdução

Lista de Mini-Projetos

Lista de Projetos Especiais

Lista de Projetos com Feedback

O que é Visualização de Dados - Definição

O que é Visualização de Dados - Compreendendo sua Audiência

O que é Storytelling?

Definindo o Propósito da Visualização de Dados

Escolhendo o Melhor Design para a Visualização de Dados

Design Thinking para Cientistas de Dados

Ferramentas de Visualização de Dados

Infográficos

Técnicas de Apresentação

E-book - Cientista de Dados, O Contador de Histórias

Slides Módulo 01

Conhecendo o Seu Sistema Operacional - Windows 10 - Parte 1/2

Conhecendo o Seu Sistema Operacional - Windows 10 - Parte 2/2

Conhecendo o Seu Sistema Operacional - MAC - Parte 1/2

Conhecendo o Seu Sistema Operacional - MAC - Parte 2/2

Conhecendo o Seu Sistema Operacional - Linux - Parte 1/2

Conhecendo o Seu Sistema Operacional - Linux - Parte 2/2

Competição DSA de Machine Learning

Bibliografia, Referências e Links úteis

02- Métodos de Visualização



Introdução

Por que Seu Cérebro Precisa de Visualização de Dados?

Que Tipo de Gráfico Escolher Para Sumarizar Meus Dados?

Tipos de Dados

Visual Encodings

O Processo de Criação de Charts e Plots

Visualização de Dados Numéricos e Não-Numéricos

Gráficos Padrões (Barra, Coluna, Linha, Pizza, Rosca, Area, Stacked Area)

Gráficos Comparativos (Bubble Plots, Radar Plot, Waterfall Plot, Packed Circle Diagrams)

Gráficos Espaciais e Mapas (Tree Maps, Choropleths, Point Maps, Raster Surface Maps)

Plots (Histogramas, Scatter Plots, Scatter Plots Matrix)

Word Clouds (Nuvem de Palavras)

Sankey Diagrams

Stream Graphs

Heatmaps

Gantt Charts

Sunburst Diagram

Box and Whisker Plot

Density Plot

Network Diagram

Melhorando a Leitura e Entendimento de Gráficos

Galeria de Exemplos com Outros Tipos de Gráficos

Slides Módulo 02

Quiz- 5 questões

Bibliografia, Referências e Links úteis

03- Organização Visual

Introdução

Identidade Visual x Design Gráfico

UI Design x UX Design

Visualizando Dados Multidimensionais

Técnicas de Visualização de Dados Multidimensionais - Projeções Geométricas

Técnicas de Visualização de Dados Multidimensionais - Técnicas Baseadas em Pixels

Técnicas de Visualização de Dados Multidimensionais - Iconografia

Técnicas de Visualização de Dados Multidimensionais - Hierarquia de Visualização

Técnicas de Visualização de Dados Multidimensionais - Grafos

Visualização de Dados e Big Data Analytics

Escolha de Técnicas de Visualização de Dados - Parte 1/2

Escolha de Técnicas de Visualização de Dados - Parte 2/2

Web Semântica

Visualização de Dados com Realidade Aumentada

Slides Módulo 03

Quiz- 5 questões

Bibliografia, Referências e Links úteis

04- Dashboard Design

Introdução

O que são Dashboards?



Dashboards Operacionais x Dashboards Analíticos x Dashboards Executivos

Tipos de Dashboards

O Processo de Design de Dashboards

Desenvolvimento do Layout

Características de um Dashboard Eficiente

Dashboards - Desenvolver ou Usar Soluções Proprietárias

Storyboard

Narrativa Visual

Erros Comuns no Design de Dashboards

Analytics Dashboard

Real-Time Dashboards

Alinhando a Terminologia: Dataviz, BI, Reporting, Analytics e KPI's

Mini-Projeto 1 - Construção de Real-Time Dashboard

Construindo um Real-Time Dashboard - Parte 1/3

Construindo um Real-Time Dashboard - Parte 2/3

Construindo um Real-Time Dashboard - Parte 3/3

Slides Módulo 04

Quiz- 5 questões

Bibliografia, Referências e Links úteis

05- Visualização de Dados com D3.js

Introdução

Introduzindo a Biblioteca D3.js para Visualização de Dados

Criando Elementos com D3 - Parte 1/3

Criando Elementos com D3 - Parte 2/3

Criando Elementos com D3 - Parte 3/3

O que é SVG?

Arrays e Objetos - Parte 1/2

Arrays e Objetos - Parte 2/2

Função Filter

Função Map

Funções Especiais do D3.js

Inicializando um Chart com SVG

Construindo um Bar Chart

Ajustando a Escala e Alterando o CSS

Adicionando Texto ao Chart

Criando uma Função para Construção de Charts

Definindo as Margens do Chart

Criando Gráficos Bidimensionais

Adicionando Cores e Eixos x e y

Invertendo os Eixos

Rotacionando os Eixos

Adicionando Labels

Adicionando Interatividade ao Gráfico

Criando um Plot de Series Temporais

Criando um Scatter Plot Interativo - Parte 1/2

Criando um Scatter Plot Interativo - Parte 2/2

Criando Aplicação de Visualização



Todos os Arquivos da Aplicação de Visualização
Criando Charts Reutilizáveis - Layout Padrão
Criando Charts Reutilizáveis - Adicionando Código de Barras
Criando Charts Reutilizáveis - Carregando Código de Barras com Arquivo CSV
Construindo um Radar Chart Interativo
Construindo um Bubble Chart com Legenda a partir de Arquivo JSON
Criando Gráfico com Slider
Adicionando um Color Picker
Adicionando Tooltips ao Gráfico
Criando Área de Seleção Interativa
Criando Mapas com Dados Geoespaciais, Integração D3 e Mapbox
Criando Choropleth Map
Criando Heatmaps
Construindo Mapas Avançados
Criando Seções no Dashboard - Parte 1/2
Criando Seções no Dashboard - Parte 2/2
Construindo um Dashboard
Mini-Projeto 2 - Criando Visualizações de Dados do Twitter em Tempo Real
Slides Módulo 05
Quiz- questões
Bibliografia, Referências e Links úteis

06- Visualização de Dados com R

Introdução
Base Plotting System
Compreendendo a Gramática dos Gráficos
Biblioteca ggplot2
Função Plot e Paleta de Cores - Parte 1/3
Função Plot e Paleta de Cores - Parte 2/3
Função Plot e Paleta de Cores - Parte 3/3
Gerando Plots Padrões (Histogramas, Boxplots e Scatterplots) - Parte 1/2
Gerando Plots Padrões (Histogramas, Boxplots e Scatterplots) - Parte 2/2
ggplot2 - Definindo as Camadas
ggplot2 - Colocando Gráficos Lado a Lado na Área de Desenho
ggplot2 - Plot Multivariado
ggplot2 - Scatter Plots e Scatter Plots 3D
ggplot2 - Barplots, Histogramas e Polígonos de Frequência
ggplot2 - Personalizando os Gráficos0
ggplot2 - Facets, Extra Grid e Plotly
Heat Maps
Correlation Heat Maps
Sumarizando Dados Multivariados em um único Heat Map
Countour Plot e Filled Countour Plot
Surface Plots 3D
Visualizando Séries Temporais com Calendar Heat Maps
Heat Maps com Plotly
Criando Gráficos com Lattice - Parte 1/2
Criando Gráficos com Lattice - Parte 2/2



Boxplots e Violin Plots
Gráficos com ggvis - Parte 1/2
Gráficos com ggvis - Parte 2/2
Construindo Mapas Interativos com ggvis
Construindo Mapas Interativos com ggvis e Shiny
Criando Mapas com RGoogleMaps - Parte 1/2
Criando Mapas com RGoogleMaps - Parte 2/2
Mapeando o Trânsito com Heat Maps e RGoogleMaps
Gráficos Interativos de Séries Temporais com Dygraphs
Criando Aplicações Web Interativas com Shiny Dashboard - Google Analytics - Parte 1/3
Criando Aplicações Web Interativas com Shiny Dashboard - Google Analytics - Parte 2/3
Criando Aplicações Web Interativas com Shiny Dashboard - Google Analytics - Parte 3/3
Slides Módulo 06
Quiz- 5 questões
Bibliografia, Referências e Links úteis

07- Visualização de Dados com Python

Introdução
Preparando o Ambiente
Plots com Matplotlib - Parte 1/3
Plots com Matplotlib - Parte 2/3
Plots com Matplotlib - Parte 3/3
Linha de Regressão com Matplotlib
Scatter Plot 3D e Plotly
Configuração e Customização Avançada do Matplotlib
Criando Módulos Personalizados
Carregando e Normalizando os Dados
Plots e GridSpec
Radar Plot
Plots Combinados
Construindo um Dashboard com Matplotlib
Usando o Pacote ggplot no Python
Criando Gráficos Estatísticos com Seaborn - Parte 1/3
Criando Gráficos Estatísticos com Seaborn - Parte 2/3
Criando Gráficos Estatísticos com Seaborn - Parte 3/3
Criando SVG Charts com Svgwrite
Criando SVG Charts com PyGal
Mapas Interativos com Geoplotlib
Mapa Interativo com Uso do Teclado
Construindo um Mapa com Animação Automática
Quadtree
Mapeamento de Rotas - Carga e Manipulação de Dados
Mapeamento de Rotas - Analisando os Dados
Mapeamento de Rotas - Construindo os Mapas
Mini-Projeto 3 - Web Analytics Dashboard
Slides Módulo 07
Quiz- 5 questões
Bibliografia, Referências e Links úteis



08- Visualização de Dados com Tableau

Introdução

Tableau Desktop e Tableau Public

Tableau Desktop x Tableau Public

Criando um Gráfico de Vendas em Apenas 3 Passos

Fontes de Dados e Navegando pelo Tableau

Criando Vários Gráficos Simultaneamente

Criando Campos Calculados

Customizando o Gráfico - Paleta de Cores

Customizando o Gráfico - Labels e Formatação

Customizando o Gráfico - Caption, Título e Summary

Exportando o Gráfico Para o Word

Inner, Left, Right e Full Outer Join

Definindo Hierarquia

Criando um Mapa do Brasil

Criando um Mapa do Brasil com Total de Vendas Por Cidade

Construindo um Dashboard - Gráfico Individual - Margem de Lucro por Estado

Construindo um Dashboard - Gráfico Individual - Lucratividade Por Cliente

Construindo um Dashboard - Montagem do Painel

Adicionando Interatividade ao seu Dashboard

Preparando um Dashboard para Storyline - Corrigindo Erros no Dataset

Preparando um Dashboard para Storyline - Trabalhando com Porcentagem

Preparando um Dashboard para Storyline - Convertendo os Bins

Preparando um Dashboard para Storyline - Adicionando Parâmetros aos Gráficos

Preparando um Dashboard para Storyline - Tree Map

Preparando um Dashboard para Storyline - Montagem do Painel

Como Apresentar uma Storyline - Parte 1/2

Como Apresentar uma Storyline - Parte 2/2

Criando um Portfólio com Public Tableau

Bonus - Conectando em Um Banco de Dados Oracle e Gerando Visualizações

Mini-Projeto 4 - Construindo uma Visualização de Dados para Apresentação Executiva

Slides Módulo 08

Quiz- 5 questões

Bibliografia, Referências e Links úteis

09- Visualização de Dados com Qlik Sense

Introdução

Conhecendo o Qlik Sense

Qlik View x Tableau x Power BI

Instalando Qlik Sense Desktop

Navegando Pelo Qlik Sense Hub

Dimensões e Medidas

Criando uma App

Importando e Carregando Dados

Associando Tabelas

Gráficos Para Avaliar o Orçamento de uma Startup

Customizando os Gráficos

Slides Módulo 09



Bibliografia, Referências e Links úteis

10- Outras Ferramentas de Visualização

Como Escolher uma Ferramenta de Visualização?

Chart.js

Leaflet

Datawrapper

Dygraphs

Highcharts

Google Charts

Polymaps

Weka

Bibliografia, Referências e Links úteis

11- Projetos com Feedback

Projeto 11 - Desenvolvendo um Business Plan Com Base em Modelos Preditivos

Projeto 12 - Design de Arquitetura Para Projetos IoT Analytics

12- Avaliação

Certificado da Formação Cientista de Dados

Avaliação- 50 questões



Formação Inteligência Artificial

Desde que os computadores foram criados, sua capacidade de realizar várias tarefas vem crescendo vertiginosamente. Bem como, sua habilidade em realizar grandes processamentos, num período de tempo extraordinário e com um nível de assertividade sem precedentes.

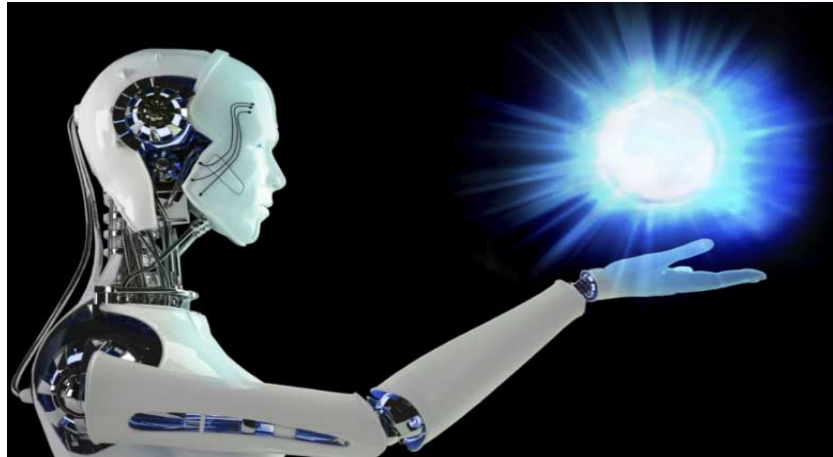
E com esse crescimento uma área que vem se destacando na Ciência da Computação é a Inteligência Artificial (I.A.), que busca criar computadores ou máquinas tão inteligentes quanto os seres humanos.

A Inteligência Artificial é realizada estudando como o cérebro humano pensa e como os seres humanos aprendem, decidem e trabalham enquanto tentam resolver um problema e, em seguida, utilizam os resultados deste estudo como base para o desenvolvimento de software e sistemas inteligentes.

Ao explorar o poder dos sistemas de computador, a maior curiosidade do ser humano é: "Pode uma máquina pensar e se comportar como os humanos?" Assim, o desenvolvimento da I.A. começou com a intenção de criar inteligência semelhante a dos seres humanos. Implementar a Inteligência Humana em Máquinas nada mais é do que criar sistemas que compreendam, pensam, aprendem e se comportam como seres humanos.

Nesse curso o aluno irá aprender os fundamentos da Inteligência Artificial e verá quais são os métodos e algoritmos utilizados para construir um sistema inteligente. Serão disponibilizados 9 cursos e um projeto final que será a construção de um sistema de inteligência artificial para atendimento, bem como, o curso bônus de Machine Learning, conforme detalhados a seguir.

Pré-requisitos Desejáveis: Conhecimentos de Sistemas Operacionais. Curso Programação Python Fundamentos. Curso Machine Learning.



Desde que os computadores foram criados, sua capacidade de realizar várias tarefas vem crescendo vertiginosamente. Bem como, sua habilidade em realizar grandes processamentos, num período de tempo extraordinário e com um nível de assertividade sem precedentes. E com esse crescimento uma área que vem se destacando na Ciência da Computação é a Inteligência Artificial (I.A.), que busca criar computadores ou máquinas tão inteligentes quanto os seres humanos. Nesse curso você aprenderá os fundamentos da Inteligência Artificial e verá quais são os métodos e algoritmos utilizados para construir um sistema inteligente. A Inteligência Artificial é realizada estudando como o cérebro humano pensa e como os seres humanos aprendem, decidem e trabalham enquanto tentam resolver um problema e, em seguida, utilizam os resultados deste estudo como base para o desenvolvimento de software e sistemas inteligentes. Ao explorar o poder dos sistemas de computador, a maior curiosidade do ser humano é: "Pode uma máquina pensar e se comportar como os humanos?" Assim, o desenvolvimento da I.A. começou com a intenção de criar inteligência semelhante a dos seres humanos. Implementar a Inteligência Humana em Máquinas nada mais é do que criar sistemas que compreendam, pensam, aprendem e se comportam como seres humanos.

Conteúdo Programático: Introdução a Ciência Artificial

1. Introdução

Bem-vindo

O que é Inteligência Artificial

História da Inteligência Artificial

Áreas de Aplicação da Inteligência Artificial

A Ciência dos Sistemas Inteligentes

Consciência e Auto-consciência

Limites da Computação Digital

A Sala Chinesa

A Máquina e o Teste de Turing

Introdução à Teoria dos Grafos

O Futuro da Inteligência Artificial

2. Agentes Inteligentes e Sistemas de Busca



- Introdução
- Agentes e Ambientes
- A Natureza dos Ambientes
- A Estrutura dos Agentes
- Conceito de Racionalidade
- Agentes de Resolução de Problemas
- Estratégias de busca sem Informação
- Estratégia de Busca com Informação (Heurística)
- Funções Heurísticas
- Algoritmos de Busca
- Problemas de Otimização
- Busca Local
- Busca com Ações Não Determinísticas
- Agentes e Busca Online em Ambientes Desconhecidos
- Decisões Imperfeitas em Tempo Real
- Jogos Estocásticos

3. Agentes Lógicos

- Introdução
- Agentes Baseados em Conhecimento
- Lógica e Lógica Proposicional
- Modelos Proposicionais Eficientes
- Agentes Baseados em Lógica Proposicional
- Lógica de Primeira Ordem
- Sintaxe e Semântica da Lógica de Primeira Ordem
- Inferência Proposicional
- Unificação e Elevação
- Encadeamento
- Planejamento Clássico
- Grafos de Planejamento
- Tempo Escalonamento e Recursos
- Planejamento Multiagente

4. Representação do Conhecimento

- Introdução
- Engenharia Ontológica
- Categorias e Objetos
- Eventos
- Sistemas e Objetos Mentais
- Sistemas de Raciocínio
- Quantificando a Incerteza
- Inferência com o Uso de Distribuições Conjuntas
- Independência
- A Regra de Bayes
- Raciocínio Probabilístico
- Representação do Conhecimento em um Domínio Incerto
- Semântica das Redes Bayesianas
- Inferência Exata em Redes Bayesianas



Modelo de Probabilidade Relacional
Raciocínio Probabilístico Temporal
Modelos Ocultos de Markov
Redes Bayesianas Dinâmicas

5. Tomada de Decisão

Introdução
A Base da Teoria da Utilidade
Redes de Decisão
Sistemas de Teoria da Decisão
Problemas de Decisão Sequencial
Iteração de Valor
Iteração de Política
Decisões com Vários Agentes - Teoria dos Jogos

6. Aprendizagem

Introdução
Teoria da Aprendizagem
Avaliação e Escolha da Melhor Hipótese
Regressão e Classificação com Modelos Lineares
Redes Neurais Artificiais
Modelos Não Paramétricos
Máquinas de Vetor de Suporte
Aprendizagem Por Agrupamento
Formulação Lógica da Aprendizagem
Aprendizagem Baseada na Explicação
Programação em Lógica Indutiva
Aprendizagem de Modelos Probabilísticos
Aprendizagem Estatística
Aprendizagem com Dados Complexos
Aprendizagem com Variáveis Ocultas

7. Busca Heurística e Métodos Estocásticos

Introdução
Programação Dinâmica
Algoritmo de Busca Pela Melhor Escolha
Admissibilidade e Monotonicidade
Usando Heurísticas em Jogos
Aspectos da Complexidade
Métodos Estocásticos
Elementos da Teoria da Probabilidade
Aplicação de Metodologia Estocástica

8. Solução de Problemas por Métodos Fortes

Introdução
Sistemas Especialistas Baseados em Regras
Sistemas Baseados em Modelos, Casos e Híbridos
Solução de Problemas Distribuída e Baseada em Agentes



Grafos Conceituais
Linguagem de Rede
Representações e Ontologias

9. Avaliação

Avaliação Final
Projeto - Parte 1

10. Apêndice - Fundamentos Matemáticos

Vetores, Matrizes e Álgebra Linear
Análise de Complexidade
Distribuições de Probabilidade



Deep Learning requer a utilização de um framework que permita facilitar o processamento de redes neurais com bibliotecas pré-construídas e habilitadas para processamento paralelo. Neste curso estudaremos os principais frameworks para Deep Learning: Theano, TensorFlow, Keras, Caffe e Microsoft CNTK. Vamos construir nossos modelos com linguagem Python e daremos ênfase ao framework que mais cresce atualmente, o Google TensorFlow.

Conteúdo Programático: Deep Learning Frameworks

1. Introdução

Bem-vindo
Navegando pela Data Science Academy
Introdução
O que é um Framework?
Principais Frameworks para Inteligência Artificial
NVIDIA Deep Learning SDK
NVIDIA CUDA Deep Neural Network library (cuDNN)

2. Theano

Introdução



- O que é o Framework Theano?
- Instalação e Configuração do Theano
- Linguagens Suportadas (Python)
- Variáveis
- Funções
- Expressões
- Otimização
- Extensores e Wrappers
- Construindo uma Rede Neural com Theano
- Processando a Rede Neural em GPU
- Técnicas de Regularização
- Redes Neurais Artificiais Baseadas em Classes

3. TensorFlow - Parte 1

- Introdução
- O que é o Framework Tensorflow?
- Instalação e Configuração do Tensorflow
- Instalação a Partir do Código Fonte
- Linguagens Suportadas (C++, Python)
- Modelo de Programação do TensorFlow
- Variáveis, Funções, Strings e Classes
- Tratamento de Exceções
- Expressões
- Como Utilizar o Tensor Board
- Otimização
- Construindo uma Rede Neural com Tensorflow
- Processando a Rede Neural em GPU
- Técnicas de Regularização
- Redes Neurais Artificiais Baseadas em Classes

4. TensorFlow - Parte 2

- Introdução
- Tensores Unidimensionais e Bidimensionais
- Tensores Tridimensionais
- Grafos de Fluxos de Dados
- Gradient Computing
- Usando o Método de MonteCarlo
- Construção do Modelo
- Machine Learning com TensorFlow
- Regressão Linear
- Clustering
- Cost Function e Gradient Descent
- Cost Function e Otimização

5. TensorFlow - Parte 3

- Introdução
- Redes Neurais Artificiais com TensorFlow
- Perceptron de Camada Única



Perceptron de Múltiplas Camadas
Função de Aproximação
Construção de Modelo

6. TensorFlow - Parte 4

Introdução
Técnicas de Deep Learning
Arquitetura das Redes Neurais Convolutivas
Arquitetura de Redes Neurais Recorrentes
Processamento de Linguagem Natural
Coletando Dados
Construindo os Modelos
Treinando e Avaliando os Modelos

7. TensorFlow - Parte 5

Introdução
Programação Paralela em GPU
TensorFlow Serving
Bazel
gRPC
Treinando e Exportando o Modelo TensorFlow

8. Keras

Introdução
O que é o Framework Keras?
Theano e Tensorflow como Backends para o Keras
Construindo Modelos de Deep Learning com o Keras

9. Microsoft CNTK

Introdução
O que é o Microsoft Cognitive Toolkit (CNTK)?
Instalando o CNTK
Regressão Logística no CNTK
Parâmetros de Aprendizagem
Dados Sintéticos
Processamento Paralelo em GPU
Classificação Multiclasse
Função Softmax
Descrição da Rede
Deep Learning com Microsoft CNTK

10. Caffe

Introdução
O que é o Framework Caffe?
Instalando o Caffe
Linguagens Suportadas (C, C++, Python e Matlab)
Definindo o Modelo de Deep Learning
Classificação de Imagens



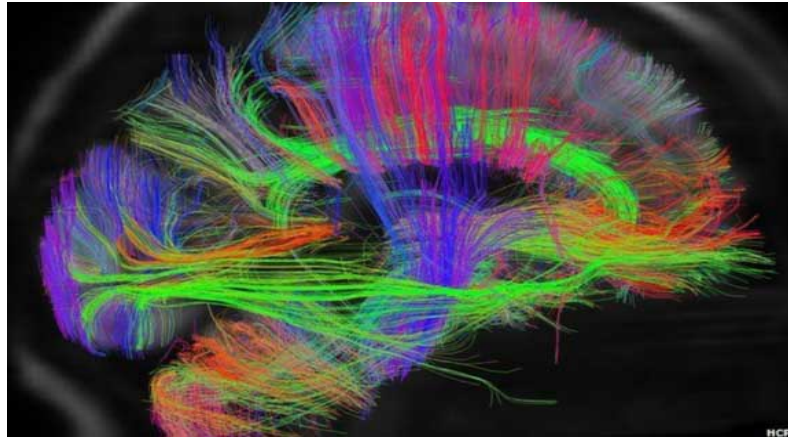
CUDA e OpenCV
BLAS (Basic Linear Algebra Subprograms)
Preparando os Dados
Definindo a Rede
Ajustando o Modelo

11. Outros Frameworks

Introdução
Torch
Deeplearning4j
Lasagne (Keras)
Leaf
maxDNN
Mxnet
Brainstorm
Chainer

12. Avaliação e Projeto

Avaliação Final
Projeto - Parte 2



Construir e treinar modelos de Deep Learning para Inteligência Artificial, requer hardware com alta capacidade computacional para processamento paralelo em Unidades de Processamento Gráfico (GPU's). Neste curso o aluno aprenderá os métodos e técnicas para processamento paralelo, extraindo o melhor que as GPU's tem a oferecer em capacidade de processamento. Ao longo do curso vamos construir programas de processamento paralelo com CUDA (plataforma de computação paralela da Nvidia) e linguagem Python.

Conteúdo Programático: Programação Paralela em GPU

1. Introdução

Bem-vindo

Navegando pela Data Science Academy

Introdução

O que é Programação Paralela?

HPPC - High Performance Parallel Computing

Por que Programação Paralela é importante em Inteligência Artificial

Programação Distribuída, Concorrente e Paralela

CPU x GPU

Processamento Paralelo em GPU Nvidia

Placas Nvidia e TitanX

Linguagens de Programação Paralela (CUDA, C++, Python e R)

Tipos de Paralelismo

Latência x Largura de Banda

Bibliotecas e SDK's

2. Hardware para Construção de Modelos em GPU

Introdução

Ambiente Virtual Data Science Academy - Nvidia Titan X

Acessando o Ambiente DSA e Configurando a Conta de Acesso

Configurando um Ambiente GPU no Amazon Web Service (AWS)

Criando e Configurando uma conta com EC2

Utilizando o Keras no AWS



Construindo Modelos de Deep Learning em Nuvem

3. Computação em GPU

Introdução

Computação Paralela em GPU

Arquitetura de Unidades de Processamento Gráfico (GPU)

General Processing GPU (GPGPU)

Modelo de Execução Paralela de Dados

Organização de Threads

Soluções Multi-CPU e Multi-GPU

Identificando Problemas com Programação Paralela (Deadlock, Starvation)

Design de Algoritmos Paralelos

Performance e Utilização de Energia

4. Programação Paralela CUDA - Parte 1

Introdução

CUDA - Instalação e Configuração (Windows, Mac e Linux)

CUDA Computing SDK

CUDA Runtime e Driver

PTX, nvcc e Microcode

Paralelismo de Dados

Estrutura de Programação CUDA

Funcionamento do Kernel

Memória Global e Transferência de Dados

Tratamento de Memória

Registradores e Memória Compartilhada

Tipos de Dispositivos de Memória CUDA

Erros com Diretivas CUDA

5. Programação Paralela CUDA - Parte 2

Introdução

Streams e Eventos

Grids de Computação 2D

Stencils e Memória Compartilhada

Interação com Threads

Interdependência de Threads

Grids de Computação 3D

Slicing, Renderização e Raycasting

Bibliotecas CUDA (cuDNN, cuSOLVER, cuRAND, cuBLAS)

Endereçamento de Blocos

6. Padrões de Programação em Paralelo

Introdução

Processamento de Dados em Paralelo

Concorrência, Streaming e Eventos

Padrões de Comunicação

Mapeamento e Stencil



Transposição
Blocos de Threads
Divergência de Threads
Intensidade Algorítmica

7. Programação Paralela em Python - Parte 1

Introdução
Programação Paralela em Python
Comunicação Inter-processos
Compreendendo o Estado Compartilhado
Usando Named Pipes
Arquitetura Celery
Transporte de Mensagens (Broker)
Operações Assíncronas

8. Programação Paralela em Python - Parte 2

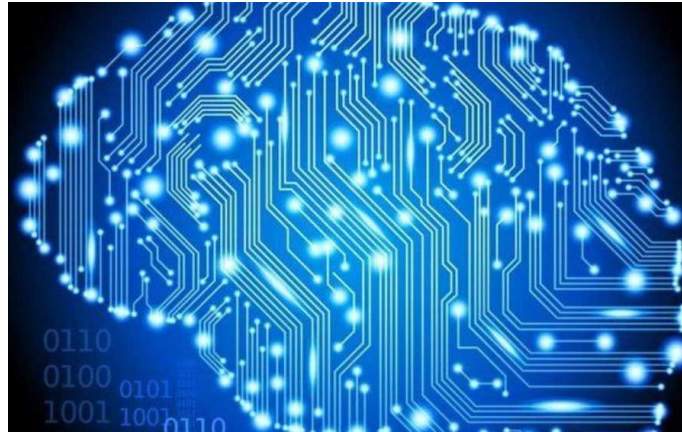
Introdução
Usando PyCUDA
Construindo Aplicações com PyCUDA
Manipulação de Matrizes
Invocação de Kernels com GPU Array
Avaliação de Expressões Element-wise
Operações de MapReduce
Usando o Módulo PyOpenCL
Testando Aplicações em GPU com PyOpenCL

9. Paralelismo Dinâmico

Introdução
Paralelismo em Bloco
Paralelismo Aninhado
Paralelismo Recursivo
Organização Paralela Dinâmica
Otimização de Aplicações para Processamento Paralelo
Trabalhando com múltiplas GPU's
Sincronização Inter-GPU

10. Avaliação e Projeto

Avaliação Final
Projeto - Parte3



Deep Learning é a técnica de aprendizado de máquina a partir de Redes Neurais Artificiais. A tecnologia de Deep Learning vai ter grandes evoluções nos próximos anos e se destacará muito no seu paradigma de “habilitar o computador a aprender a partir da observação dos dados”.

A vantagem dos algoritmos de Deep Learning é sua capacidade de aprendizagem em grandes quantidades de dados de uma forma não-supervisionada, sendo assim uma ferramenta valiosa para Big Data Analytics onde a maioria dos dados são desta natureza, também designados por dados não-estruturados. Deep Learning vem sendo largamente utilizado para construção de aplicações de Inteligência Artificial. O assunto é tão extenso, que dividimos em 2 cursos. Neste primeiro vamos estudar o que são Redes Neurais Artificiais, Perceptrons de Camada Única e Perceptrons de Múltiplas Camadas.

Conteúdo Programático: Deep Learning I

1. Introdução

- Bem-vindo
- Navegando pela Data Science Academy
- Introdução
- O que é uma Rede Neural?
- O que é Deep Learning?
- História da Inteligência Artificial
- Áreas de Aplicação da Inteligência Artificial

2. Álgebra Linear

- Introdução
- Escalares, Vetores, Matrizes e Tensores
- Multiplicação de Vetores e Matrizes
- Matriz Inversa
- Dependência Linear
- Tipos Especiais de Vetores e Matrizes
- Singular Value Decomposition
- Determinante

3. Processo de Aprendizagem



Introdução
Aprendizagem Por Correção e Erro
Aprendizagem Baseada em Memória
Aprendizagem Hebbiana
Aprendizagem Competitiva
Aprendizagem de Boltzmann
Tarefas de Aprendizagem
Natureza Estatística do Processo de Aprendizagem

4. Redes Neurais Artificiais

Introdução
Modelos de Redes Neurais Artificiais
Topologias
Treinamento e Visualização
Codificação de Dados
Normalização
Normalização Min-Max
Normalização Gaussiana
Redes Clássicas
Treinamento Não Supervisionado
Redes de Kohonen
Redes RBF
Outros Tipos de Redes Neurais Artificiais

5. Perceptrons de Camada Única

Introdução
Filtragem Adaptativa
Otimização Irrestrita
Filtro Linear de Mínimos Quadrados
Curvas de Aprendizagem
Variação da Taxa de Aprendizagem
Perceptron
Teorema de Convergência do Perceptron
Classificador Bayesiano
Ambiente Gaussiano

6. Perceptrons de Múltiplas Camadas - Parte 1

O que são MLP's?
Neurônios e Perceptrons de Múltiplas Camadas
Mecanismo de Input e Output
A Definição de Classe Perceptron
Perceptron e Adaline
O Problema do XOR
Redes de Neurônios
Usando a Classe Perceptron
Limitações
Treinando a Rede
Definindo o Problema e Coletando Dados



Construindo o Modelo com Tensorflow e Keras
Fazendo Previsões
Avaliando o Modelo
Cross Validation

7. Perceptrons de Múltiplas Camadas - Parte 2

Introdução
Salvando a Rede Neural em JSON e YAML
Checkpointing
Compreendendo o Comportamento do Modelo
Acessando e Visualizando o Histórico de Treinamento
Regularização e Redução do Overfitting
Dropout
Camadas Ocultas e Visíveis
Taxas de Aprendizagem
Taxa de Aprendizagem Time-Based e Drop-Based
Melhores Práticas para MLP's

8. Perceptrons de Múltiplas Camadas - Parte 3

Introdução
Algoritmo de Retropropagação
Heurísticas de Desempenho
Representação da Saída e Regras de Decisão
Experimento Computacional
Detecção de Características
Retropropagação e Diferenciação
Matriz Hessiana
Generalização
Aproximação de Funções
Cross Validation
Vantagens e Limitações da Aprendizagem Por Retropropagação

9. Avaliação e Projeto

Avaliação Final



As máquinas inspiradas em Deep Learning conseguem aprender abstrações complexas dos dados através de um processo de aprendizagem hierárquica muito similar ao que ocorre no cérebro humano – sobretudo no córtex visual. Estes métodos conseguem obter uma representação complexa e abstrata dos dados de forma hierárquica usando várias camadas de processamento não linear. Os dados são alimentados a uma primeira camada sendo a saída de cada camada uma entrada da camada seguinte. O empilhamento de várias camadas destes neurónios é a ideia básica dos algoritmos de Deep Learning. Neste curso estudaremos mais avançados em Deep Learning, como as Redes Neurais Convolutivas, Redes Neurais Recursivas, Mapas Organizacionais e Redes RBF.

Conteúdo Programático: Deep Learning II

1. Convolutional Neural Networks - Parte 1

Introdução

O que são as CNN's?

Camadas Convolucionais

Grupos de Pooling

Grupos de Normalização de Resposta

Função Softmax

Reconhecimento de Escrita Manual

Carregando o Dataset MNIST no Keras

CNN Simples para o MNIST

CNN Complexa para o MNIST

2. Convolutional Neural Networks - Parte 2

Introdução

API de Imagem Aumentada para o Keras

Pontos de Comparação

Padronização de Atributos

Rotações, Flips e Shifts Randômicos

Reconhecimento de Objetos em Imagens

Criação e Avaliação do Modelo

Melhores Práticas para CNN's



3. Recurrent Neural Networks - Parte 1

Introdução

O que são as RNNs?

Arquitetura de Redes Recorrentes

O Poder Computacional das Redes Recorrentes

Retropropagação Através do Tempo

Aprendizagem Recorrente em Tempo Real

Filtros de Kalman

Filtros de Kalman Estendido

4. Recurrent Neural Networks - Parte 2

Introdução

Redes com Memórias de Longa Duração

Previsões de Séries Temporais

Usando o Método Window

Redes com Memórias de Longa Duração

Previsões de Séries Temporais

Usando o Método Window

Extinção de Gradientes em Redes Recorrentes

Controle Adaptativo Por Referência a Modelo

5. Métodos Teóricos da Informação

Introdução

Entropia

O Princípio da Máxima Entropia

Informação Mutua

Infomax e Redução da Redundância

Análise de Componentes independentes

Método da Máxima Entropia

Máquinas Estocásticas

Mecânica Estatística

Máquina de Boltzmann

Redes de Crença Sigmóide

6. Mapas Auto-Organizáveis

Introdução

O Mapa Auto-Organizável

Simulações Computacionais

Quantização Vetorial Por Aprendizagem

Classificação Adaptativa de Padrões

Quantização Vetorial Hierárquica

Mapas Contextuais

7. Redes de Função de Base Radial

Introdução

Interpolação

Reconstrução de Hipersuperfície Malformulada

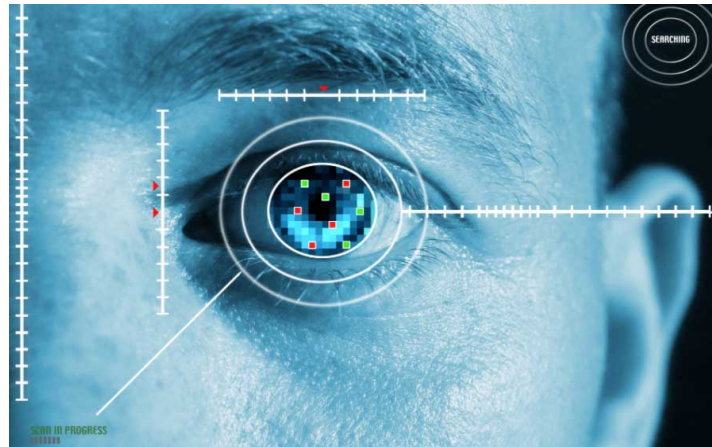
Teoria da Regularização



Redes de Função de Base Radial
Estimação do Parâmetro de Regularização
Propriedades Aproximadas de Redes RBF
Comparação Entre Redes RBF e Perceptrons de Múltiplas Camadas
Regressão de Núcleo e sua Relação com as Redes RBF
Estratégias de Aprendizagem
Classificação de Padrões

8. Avaliação e Projeto

Avaliação Final
Projeto - Parte 5



Em um mundo onde tiramos fotos o tempo todo, o reconhecimento de imagens tem mostrado avanços consideráveis. Podemos agora buscar e organizar coleções de fotos, mesmo que elas não tenham uma tag de identificação. E isso é feito em apenas alguns segundos. Vamos pensar sobre isso. Para ser capaz de identificar imagens de cães, uma app precisa ser capaz de identificar qualquer tipo de cão, desde um Chihuahua até um Pastor Alemão e não pode se equivocar com imagens invertidas ou parcialmente obscuras, sob neblina ou neve, no sol ou na sombra. A app deve ainda ser capaz de excluir lobos e gatos. A app precisa identificar um cachorro usando apenas os pixels de uma imagem. Como isso pode ser possível? É o que vamos ajudá-lo a descobrir neste incrível e curso. Estudaremos reconhecimento facial, reconhecimento de objetos em imagens, usaremos um dos mais populares frameworks para visão computacional, o OpenCV e ainda estudaremos reconhecimento de imagens em smartphones Android e iOS.

Conteúdo Programático: Visão Computacional e Reconhecimento de Imagens

1. Introdução

Bem-vindo

Navegando pela Data Science Academy

Introdução

Computer Vision x Machine Vision x Processamento de Imagem

Visão Humana x visão Computacional

O que é uma Imagem?

Formação da Imagem

Projeção e Perspectiva

A Imagem Digital

Histograma de Imagem Digital

Problemas da Visão Computacional

Detecção de Bordas

Reconhecimento de Objetos

Visão Estereoscópica

Aplicações da Visão Computacional (Medicina, Realidade Aumentada e OCR)



2. Sistema de Visão Computacional

- Introdução
- Aquisição de Imagens
- Restauração e Realce
- Segmentação
- Extração de Atributos ou Características
- Classificação e Reconhecimento
- Sistemas de Visão Binária

3. Bibliotecas de Computer Vision

- Introdução
- OpenCV
- Simplecv
- Libccv
- Open eVision
- Adaptative Vision
- Common Vision
- Instalando e Configurando o OpenCV

4. Tratamento de Arquivos de Imagens

- Introdução
- Operações Básicas de Leitura e Escrita com Imagens
- Convertendo uma Imagem em Bytes
- Extraindo os Atributos de Imagens
- Filtrando Imagens
- Criando Módulos
- Simulando Espaço de Cores RC, RGV e CMV
- Cartonizando uma Imagem
- Detecção de Arestas, Círculos e Linhas
- Operações Lógicas em Imagens
- Criando Imagens Panoramicas
- Segmentação de Imagens
- Transformação
- Escala

5. Reconhecimento Facial

- Introdução
- Tracking
- Definindo uma Face como uma Hierarquia de Retângulos
- Detectando Arestas e Aplicando Filtro às Imagens
- Blurring
- Motion Blur
- O que são Imagens Integrais?
- Detectando Atributos Faciais
- Detectando Formatos
- Reconhecimento de Gestos
- Treinamento e Visualização



6. Reconhecimento de Objetos

Introdução
Detecção de Movimentos e Tracking
Formação de Imagens
Reconhecimento de Objetos Por Aparência
Reconstrução do Mundo em 3D
Reconhecimento de Objetos a Partir de Informação Estrutural
Utilização da Visão
Histograma de Equalização

7. Visão Computacional para Smartphones

Introdução
Android NDK
Acessando a Câmera no Android e iOS
Algoritmos de Detecção
Criando Projetos com OpenCV para Android
Processamento de Imagens no Android
Portanto Aplicação Desktop para o Android
Ajustando a Resolução
Customizando a App
Criando Projetos iOS com OpenCV
Arquitetura da Aplicação
Binarização e Renderização da Imagem
Reconhecimento do Marcador
Refinamento do Marcador
Marcador 3D
Renderização de Objetos 3D
Marcadores Baseados em Realidade Aumentada
Remoção de Outliers

8. Captura de Movimentos

Introdução
Estruturas de Movimento com OpenCV
Estimando o Movimento da Câmera
Fluxo Óptico
Reconstrução de Cenas
Descritores de Atributos
Visualização de Pontos 3D

9. Reconhecimento de Padrões

Introdução
Algoritmo ANPR
Segmentação
Classificação
Segmentação OCR
Reconhecimento de Padrões com Support Vector Machines
Reconhecimento de Padrões com Deep Learning



10. Avaliação e Projeto

Avaliação Final

Projeto - Parte



Você percebeu ao longo dos últimos anos o incrível salto de qualidade em diversas tecnologias que usamos em nosso dia a dia? Se você possui um smartphone com reconhecimento de voz, provavelmente percebeu. Essa funcionalidade em nossos smartphones está melhor do que nunca. Com um simples comando de voz, fazemos ligações, acessamos a internet, fazemos buscas e abrimos aplicativos. Nunca foi tão fácil. Estamos conversando com nosso computador mais do que nunca. Assistiu o filme Her (Clique aqui para ver o Trailer do filme)? Não estamos muito longe disso! Softwares como o Cortana da Microsoft, Siri da Apple, Alexa da Amazon e recursos de reconhecimento de voz do Google, nos permitem interagir com o computador ou smartphone de forma simples. Um simples comando de voz e o computador faz o que pedimos. Neste curso aprenderemos o que está por trás desta tecnologia e como utilizar processamento de linguagem natural e reconhecimento de voz. E usaremos Deep Learning para isso.

Conteúdo Programático: Processamento de Linguagem Natural e Reconhecimento de Voz

1. Introdução

Bem-vindo

Navegando pela Data Science Academy

Introdução

O que é Processamento de Linguagem Natural

Processamento de Linguagem Natural e Inteligência Artificial

Módulos Python para Processamento de Linguagem Natural

Modelos de Linguagem

Classificação de Texto



Recuperação de Informação

2. Linguagem Natural para Comunicação

Introdução

Gramática com Estrutura Frasal

Análise Sintática

Interpretação Semântica

Tradução Automática

Reconhecimento de Voz

3. Manipulando Strings

Introdução

Tokenização

Normalização

Eliminação de Pontuação

Removendo Pontuação

Tratamento de Caracteres Repetidos

Substituição de Palavras por Sinônimos

Tratamento de Stopwords

Stemming

4. Modelagem Estatística da Linguagem

Introdução

Frequência de Palavras

Desenvolvendo MLE para um Texto

Estimativas Usando Modelo Oculto de Markov

Aplicando Smooth em um Modelo MLE

Interpolação

Amostragem Gibbs em Processamento de Linguagem

Avaliando o Modelo de Linguagem

Seq2Seq e Large Scale DL

5. Morfologia

Introdução

Introdução à Morfologia

Stemmer

Stemmer para Inglês e Português-BR

Gerador Morfológico

Analisador Morfológico

Engine de Busca

Identificando Palavras

GLoVe (Global Vectors for Word Representation)

6. API's de Reconhecimento de Voz

Introdução

Google Speech Recognition

Microsoft Bing Voice Recognition

IBM Speech



Wit.ai

Api.ai

Módulo SpeechRecognition do Python

Módulo Pygsr

PyAudio

FLAC Encoder

Script para Reconhecimento de Voz

Reconhecimento de Voz a Partir do Microfone

Transcrevendo o Arquivo de Áudio

Calibrando o Reconhecimento de Voz para Ambientes com Barulho

Corpora Pos-tagged

7. Usando Redes Neurais para Problemas de NLP

Introdução

NLP e Deep Learning

Simple Word Vector: word2vec, GloVe

Representações Avançadas de word vector: Modelos de Linguagem, Softmax, Redes de Camada Única

Neural Networks e Backpropagation - Reconhecimento de Entidades Nomeadas

Gradient Check, Overfitting, Regularização

Usando o TensorFlow

Machine Learning em Larga Escala

8. Redes Neurais Recursivas

Introdução

Descrição de Dados para Redes Neurais Recursivas

RNN's para Modelagem da Linguagem

RNN'S Parsing

CNN's para Classificação de Sentenças

Tree Neural Networks (TNN)

Construindo uma TNN com Recursividade

Redes Neurais Recursivas com Theano

Redes Neurais Recursivas com Tensorflow

9. Avaliação e Projeto

Avaliação Final

Projeto - Parte 7



Um dos grandes desafios da atualidade é produzir tecnologias que consigam armazenar e analisar de forma eficiente a massa de dados gerada atualmente. O modelo relacional tem se mostrado ineficiente, pois o volume de dados e seus relacionamentos afetam diretamente o seu desempenho. Portanto, utilizar um modelo flexível, como o orientado a grafos, pode ser uma boa alternativa para a criação de modelos complexos de dados. A teoria dos grafos é um ramo da matemática que estuda as relações entre os objetos de um determinado conjunto. Para tal são empregadas estruturas chamadas de grafos. Estruturas que podem ser representadas por grafos estão em toda parte e muitos problemas de interesse prático podem ser formulados como questões sobre certos grafos. O desenvolvimento de algoritmos para manipular grafos é um tema importante da ciência da computação e desenvolvimento de aplicações de Inteligência Artificial. O objetivo deste curso é dar a você condições de utilizar análise em grafos em seu projeto, analisando grandes conjuntos de dados.

Conteúdo Programático: Análise em Grafos para Big Data

1. Introdução

- Bem-vindo
- Navegando pela Data Science Academy
- Introdução
- O que são Grafos?
- Introdução à Teoria dos Grafos
- Análise de Grafos
- Um Grafo para Cada Problema
- Relacionamentos
- Hierarquias
- Comunidades
- Fluxos
- Redes Espaciais
- Graph Neural Networks

2. Ferramentas e Graph Databases

- Introdução



Ferramentas de Grafos "Point-and-Click" (Gephi, NodeXL)

Neo4j

OrientDB

SAP Hana

Oracle Spatial e Graph

DataStax

ArangoDB

3. Processamento

Introdução

Coletando, Limpando e Conectando Dados

Coleta - Identificação dos Dados

Limpeza: Organização e Transformação

Conexão: Organização dos Dados em Grafos

Processo de Construção de Grafos

4. Layout e Visualização

Introdução

Estatística Básica de Grafos

Layouts

Atributos Visuais

Atributos Key Node

Atributos Key Edge

Combinação de Atributos

5. Análise Visual de Grafos

Introdução

Relacionamentos

Links de Relacionamentos

Links e Nodes

Hierarquias

Árvores de Decisão

Árvores e Grafos

Comunidades

O que Define uma Comunidade?

Graph Clustering

Fluxos

Fluxos de Layout de Comunidades

Diagramas Chord

Árvore de Fator Comportamental

Redes Espaciais

Layout e Esquemas

Roteamento de Padrões

6. Análise em Grafos para Big Data

Introdução

Graph Query Languages

Analisando Vizinhanças



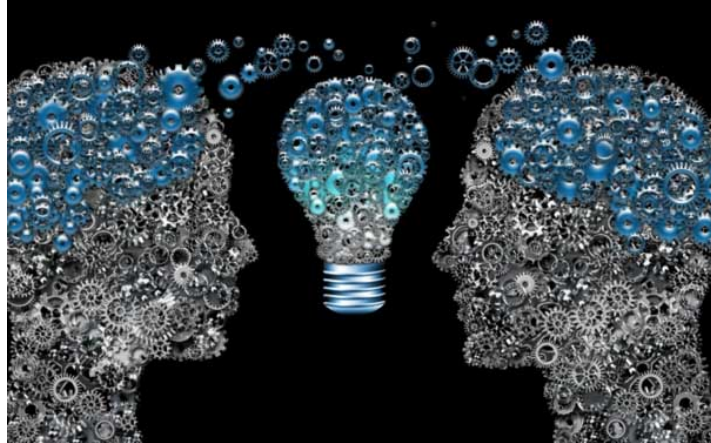
Plotando a Atividades da Rede
Visualização de Comunidades
Grafos Dinâmicos
Grafos Transacionais
Design

7. Visualização de Dados com D3.JS

Introdução
Configurando o Ambiente
Manipulação de DOM, SVG e CSS
Funções Built-in
Geodata
Interpoladores
Visualização Dinâmica de Dados

8. Avaliação e Projeto

Avaliação Final
Projeto - Parte 8



Os sistemas cognitivos têm o potencial de criar rupturas nas empresas e na sociedade, mudando inclusive a natureza do trabalho. Não apenas as tarefas explícitas podem ser automatizadas, mas tarefas tácitas (um veículo autônomo, por exemplo, pode dispensar motorista). Atividades realizadas hoje por indivíduos, como o atendimento em call center e o suporte administrativo, podem ser inteiramente substituídos por estes sistemas. Também não seria inimaginável pensar que diversas tarefas ligadas a setores como educação, direito e saúde também poderiam ser efetuadas por sistemas “inteligentes”. Neste curso vamos estudar os conceitos de sistemas cognitivos e estudaremos 2 importantes sistemas que estão ajudando a transformar o mundo como o conhecemos, o IBM Watson e o Amazon Alexa. Aprenderemos ainda como construir sistemas cognitivos com o Microsoft Cognitive Toolkit (CNTK).

Conteúdo Programático: Sistemas Cognitivos

1. Introdução

- Bem-vindo
- Navegando pela Data Science Academy
- Introdução
- O que são Sistemas Cognitivos?
- Arquitetura de Sistemas Cognitivos
- Design e Compreensão de Sistemas Cognitivos
- O que é Inteligência?
- Inteligência Artificial Baseada em Conhecimento
- O que a Inteligência Artificial pode Fazer?
- Conexão Cognitiva
- Inteligência Artificial, Sistemas Cognitivos e Ética

2. Redes Semânticas

- Introdução
- Estrutura de Redes Semânticas
- Representação
- Partes Fundamentais da Representação



Identificação de Objetos Basedada em Atributos
Analogia Geométrica
Análise Meios-Fins
Sistemas Geradores e Testes

3. Sistemas Baseados em Regras

Introdução
Aprendizado Incremental
Classificação
Árvores de Discriminação

4. Modelagem Cognitiva

Introdução
Engenharia do Conhecimento
Frames
Entendimento
Raciocínio de Senso Comum
Raciocínio Analógico
Aprendizado Baseado em Explicação
Raciocínio Baseado em Explicação
Raciocínio Visual e Espacial
Metacognição

5. IBM Watson - Parte 1

Introdução
IBM Watson em Perspectiva Empresarial
Versão Free do Watson
Enriquecimento de Dados
Interpretação de Dados
Modelo de Conhecimento do Watson
Normalização e Modelagem
Gerenciamento de Objetos
Enterprise Watson Roadmap
Watson Professional
Integrações

6. IBM Watson - Parte 2

Introdução
O que é o IBM Watson?
Desenvolvimento de Aplicações para o IBM Watson
Construindo um Módulo Cognitivo - Parte 1
Construindo um Módulo Cognitivo - Parte 2
Construindo um Módulo Cognitivo - Parte 3
Construindo um Módulo Cognitivo - Parte 4

7. Amazon Alexa

Introdução
Alexa SDK para Desenvolvedores



Interface de Reconhecimento de Voz
Framework Alexa e Skills Kit
SSML (Speech Synthesis Markup Language)
Estado de Sessão
Desenvolvendo Sistema de Conversação com Alexa

8. Sistemas Cognitivos com Microsoft CNTK

Introdução
Configuração do CNTK
LSTM Sequence Classification
Modificando uma Rede CNTK
Operações Pré-Definidas (Dropout)
Criando Funções
Implementando Estruturas de Rede Customizadas
Usando Loops Recursivos
Treinamento Paralelo
Reconhecimento e Classificação de Imagens

9. Avaliação e Projeto

Avaliação Final



Formação Engenheiro de Dados



Nesse curso o aluno irá aprender a preparar os dados para uso analítico ou operacional, por meio de uma trilha de aprendizagem de cinco cursos, únicos, e com conceitos avançados sobre como ele, o aluno, irá lidar com conjuntos de dados estruturados e não estruturados, aplicados a uma variedade de ferramentas tecnológicas de última geração, explicadas passo a passo.

Projete, Crie, Mantenha e Solucione Problemas de Processamento de Dados com Ênfase em Segurança, Confiabilidade, Tolerância a Falhas, Escalabilidade, Eficiência e Profissionalismo, acompanhe a seguir o conteúdo programático completo de cada curso da Formação Engenheiro de Dados.



Conteúdo Programático: Design e Implantação de Data Warehouses

1. Introdução

Introdução

O Que é Data Warehouse?

Data Warehouses Ainda São Relevantes na Era do Big Data?

Quando Usar um Data Warehouse?

Principais Características do Data Warehouse

Vantagens e Limitações do Data Warehouse Tradicional

Data Warehouses, Business Intelligence, Data Marts e OLTP

Componentes Típicos de um Data Warehouse

Extraction, Transformation e Loading (ETL)

Oracle OLAP e o Modelo Dimensional

Oracle Data Mining

Data Warehouse x OLTP

Componentes de um Data Warehouse

Arquitetura Multi-Camada

Arquitetura Corporativa - Enterprise Data Hub

O Que é e Quando Usar Uma Staging Area?

Como Criar um Business Case - Template

Como Criar uma Especificação Funcional - Template

Como Criar uma Especificação Técnica - Template

2. Business Intelligence e Modelagem de Dados

Introdução

O Que é Business Intelligence?

Business Intelligence x Data Science

Ferramentas de BI x Ferramentas Para Data Science

Definindo o Modelo de Negócio

Definindo o Modelo Lógico



- Definindo o Modelo Dimensional
- Definindo o Modelo Físico
- Star Schema, Snowflake Schema e Terceira Forma Normal
- Tabelas Fato e Dimensão
- Convertendo Dimensões de Negócio em Tabelas Dimensão
- Convertendo o Modelo Dimensional Para o Modelo Físico
- Granularidade
- Dimensão de Tempo
- Slow Changing Dimension

3. Modelo Entidade-Relacionamento

- Fases do Projeto de Base de Dados
- Definição e Objetivo do Modelo Entidade-Relacionamento
- Entidades e Conjuntos-Entidade
- Atributos e Domínio de Atributos
- Relacionamentos
- Grau de Relacionamento
- Projeto de Chaves
- Restrições de Cardinalidade
- Integridade Referencial
- Agregação
- Caso de Uso
- Usando Ferramenta Case e Criando o Modelo

4. Modelo Dimensional

- Modelagem Dimensional
- Benefícios da Modelagem Dimensional
- O Que é uma Dimensão?
- O Que é um Fato?
- Fatos Aditivos, Semi-Aditivos e Não-Aditivos
- Surrogate Key
- Star Schema
- Snowflake Schema
- Star Schema x Snowflake Schema
- Estudo de Caso
- Bitmap Index
- B-Tree Index
- Bitmap Index x B-Tree Index

5. Banco de Dados Oracle

- Instalando o Banco de Dados Oracle
- Configurando o Banco de Dados Oracle
- Database Sizing
- Estimando o Tamanho do Banco de Dados
- Projeto e Desenvolvimento da Data Staging Area
- Implementando o Modelo
- Considerações de Tuning e Performance
- Particionamento e Indexação



Estratégia Inicial de Agregação
Estratégia Inicial de Indexação
Paralelismo
Considerações Sobre Segurança

6. Extração, Transformação e Carga de Dados (ETL)

O Que é ETL?
O Processo de ETL
Ferramentas ETL Para Data Warehouses
Métodos de Extração de Dados de Data Warehouses
Métodos de Extração Físico e Lógico
Mapeamento de Dados
Oracle ETL Tools: Warehouse Builder, SQL*Loader e Data Pump
Usando Pentaho Data Integration
Técnicas de Extração e Manutenção de Metadados
Modelos de Staging Onsite e Remoto
Transformação e Checagem de Anomalia nos Dados
Técnicas de Transformação de Dados
Técnicas de Transformação de Metadados
Flat Files, Distributed Systems e Transportable Tablespaces
Construindo o Processo de Carga de Dados

7. Projeto - Implementando Data Warehouse - Parte 1

Por Que Usaremos Banco de Dados Oracle?
Definindo o Projeto
Levantamento de Requisitos
Design e Arquitetura do Projeto
Modelo de Dados
Fase ETL
Acessando os Dados na Fonte
Carregando os Dados
Verificando a Integridade dos Dados
10 Erros Comuns a Se Evitar em Projetos de Data Warehouse

8. Projeto - Implementando Data Warehouse - Parte 2

Automatizando o Processo de Refresh
Capturando Dados Modificados
Cálculos, Exceções e Drill Down
Opções de Entrada de Dados
Customização
Usando Sumários Para Aumentar a Performance
Usando Materialized Views
Query Rewrite
Cost-Based Query Rewrite Process
Trabalhando com Dimensões e Hierarquias
Deploy e Roll-out
Considerações e Lições Aprendidas
Documentação



Gestão de Metadados
Reporting e Segurança

9. Implementando Data Warehouse em Nuvem - Parte 1

- A Importância do Cloud Computing
- Benefícios de um Data Warehouse em Nuvem
- Introdução ao Amazon Web Services (AWS)
- O Que é o Amazon Redshift?
- Como as Empresas Estão Usando o Redshift?
- Vantagens, Desvantagens e Custo do Amazon Redshift
- Definindo o Projeto
- Arquitetura do Data Warehouse com AWS Redshift
- Criando um Cluster Redshift
- Criando um Cluster em uma Nuvem Virtual Privada (VPC)
- Configurando o SQL Workbench
- Carregando Dados no Amazon S3
- Carregando Dados no Cluster Redshift
- Tuning e Performance
- Conectando no Cluster
- Redshift CLI e API
- Efetutando Queries no Amazon Redshift

10. Implementando Data Warehouse em Nuvem - Parte 2

- Gerenciando Subnets no Cluster Redshift
- Gerenciando Grupos de Parâmetros
- Criando Snapshots
- Configurando Encriptação
- Migrando um Cluster Não-Encriptado para um Cluster Encriptado
- Usando Instâncias Reservadas com Redshift
- Autenticação e Controle de Acesso ao Data Warehouse
- Amazon Redshift Security Group
- Usando Políticas IAM para Customizar a Segurança
- Configurando Conexão JDBC e ODBC com Credenciais IAM
- Monitoramento do Cluster
- Auditoria e Logging
- Restaurando um Snapshot

11. Avaliação e Certificado de Conclusão

Avaliação Final

Bônus - Curso de Sistema Operacional Linux



Conteúdo Programático: Data Lake – Design, Projeto e Integração

1. Introdução

Navegando pela Data Science Academy
O Que São Dados Corporativos?
Relevância e Qualidade dos Dados
As Fontes do Big Data
Dados Estruturados x Dados Não Estruturados
Infraestrutura em Nuvem x On-premise
Bancos de Dados Relacionais x Bancos de Dados NoSQL
Data Lake x Data Warehouse Tradicional
Apache Hadoop e Big Data Stack
Sistemas de Arquivos Distribuídos
Qual o Custo de Implementação de um Data Lake?
Definindo um Business Case
Especificação Funcional e Técnica
Arquitetura Corporativa - Enterprise Data Hub

2. O Que São Data Lakes?

Conceito de Data Lake
Data Lakes - Benefícios, Vantagens e Desvantagens
Arquiteturas de Data Lakes
Data Lake Storage
Processamento e Gerenciamento de Dados
Gestão de Metadados e Segurança
Qualidade e Ciclo de Vida dos Dados
Quais Plataformas São Usadas Para Construir Data Lakes?
Como Construir um Data Lake?
Data Lake - Camada de Aquisição de Dados
Data Lake - Camada de Ingestão de Dados



Data Lake - Camada de Processamento de Dados
Data Lake - Camada de Armazenamento de Dados
Data Lake - Camada de Mensagens
Roadmap Para um Data Lake de Sucesso
Dividindo um Data Lake em Zonas
Data Lakes - Caso de Uso - Análise Estatística de Dados em Tempo Real

3. Data Lake - Design

Definindo as Fontes de Dados
Clickstream Data
Logs de Servidores
Dados de Mídias Sociais
Coordenadas de Geolocalização
Dados de Máquinas e Sensores
Conhecendo as Distribuições Hadoop
Apache Spark Framework
Fatores de Seleção de um Big Data Stack
Definindo a Arquitetura
NameNode x DataNode
Camada de Aquisição de Dados com Apache Sqoop
Camada de Aquisição de Dados com Apache Flume
Fontes e Eventos
Canais de Fluxos de Eventos
Spark Streaming
Dstreams, DataFrames e GraphFrames
Processo de Exportação/Importação de Dados
Data Lake - Casos de Uso - Capturando e Analisando Dados de Sensores

4. Data Lake - Aquisição de Dados em Batch

Contexto no Data Lake
O Que São Dados em Batch?
Real-Time Data x Batch Data
Importação e Exportação de Dados
Conectores Para Bancos Relacionais
Implementação do Data Lake - Parte 1 - Instalação Java
Implementação do Data Lake - Parte 2 - Instalação Hadoop
Implementação do Data Lake - Parte 3 - Instalação HUE
Implementação do Data Lake - Parte 4 - Instalação Sqoop
Configurando a Ingestão de Dados
Configurando o HDFS como Destino
Criando e Automatizado Jobs de Importação
Agendamento de Jobs
Importando Dados do Data Warehouse
Usando Pentaho Para Importação de Dados
Usando Pentaho Data Integration Para Integração Data Lake x Data Warehouse
Data Lake - Casos de Uso - Análise de Logs de Servidores

5. Data Lake - Aquisição de Dados em Streaming



- Contexto no Data Lake
- O Que é Stream de Dados?
- O Que é o Apache Flume?
- Arquitetura Flume
- Pipeline Distribuído
- Arquitetura Avançada
- Agentes, Fonte, Canal, Eventos e Sink
- Seletor de Canais
- Processador de Canais
- Grupos de Sink
- Instalação e Configuração
- Implementação do Data Lake - Parte 5 - Instalação do Apache flume
- Implementação do Data Lake - Parte 6 - Configuração do Apache flume
- Coleta de Dados em Tempo Real
- Avaliando Outras Opções - Apache Flink

6. Data Lake - Camada de Mensagens

- Contexto no Data Lake
- Camada de Mensagens
- Arquitetura e Topologia
- O Que é o Apache Kafka?
- Vantagens e Desvantagens
- Arquitetura do Kafka
- Ciclo de Vida de Streams de Dados
- Kafka Message
- Kafka Producer
- Persistência de Dados Usando Tópicos
- Partições
- Kafka Message Broker
- Kafka Consumer
- Kafka Core API's
- Kafka REST interface
- Segurança no Kafka
- Implementação do Data Lake - Parte 7 - Instalação do Apache Kafka
- Implementação do Data Lake - Parte 8 - Configurando um Multi-Broker Cluster
- Implementação do Data Lake - Parte 9 - Do RDBMS Para o Kafka
- Implementação do Data Lake - Parte 10 - Spool Messages no Kafka
- Avaliando Outras Opções - RabbitMQ, ZeroMQ e Apache Active MQ

7. Data Lake - Armazenamento de Dados

- Contexto no Data Lake
- O Que é o Apache Hadoop?
- Por que Usar o Apache Hadoop?
- Vantagens e Desvantagens do Apache Hadoop
- Arquitetura Hadoop
- Armazenamento Distribuído - Hadoop HDFS
- Computação Distribuída - Hadoop MapReduce
- Computação Distribuída - Apache Spark



Gerenciamento de Dados

Big Data Stack

Modos de Deployment

Implementação do Data Lake - Parte 11 - Instalação do Apache Hadoop

Implementação do Data Lake - Parte 12 - Configuração do Apache Hadoop

Implementação do Data Lake - Parte 13 - Instalação e Configuração do Hive e Hbase

Implementação do Data Lake - Parte 14 - ETL

8. Data Lake - Deploy, Roll-out e Boas Práticas

Arquitetura de Todos os Componentes

Documentação

Topologia

Definição de Acesso

Implementação do Data Lake - Parte 15 - Operação do Data Lake

Desafios an Implementação de um Data Lake

Integração e Enterprise Data Hub

Roll-out e SysOps

Precisamos de uma Sandbox?

9. Adicionando Camadas ao Data Lake

Os Ingredientes Básicos de uma Aplicação Real-Time Analytics

O Que é o Apache Storm?

Instalando e Configurando o Storm

Jobs de Processamento em Tempo Real com Storm

Storm Trident

Estado de Retenção

Topologia e Operações com Storm Trident

Funções map e flatMap

Windowing e Filtros

Integração Storm e Hadoop

O que é o Redis?

Processamento em Memória com Redis

Técnicas de Gerenciamento de Memória com Redis

Expiração de Chave

Criando Estruturas de Memória Eficientes com Redis

Protocolo de Serialização

Integração do Storm com Redis, Elasticsearch e Hbase

Aplicação de Processamento de Streaming de Dados

10. Avaliação Final e Certificado de Conclusão

Avaliação Final

Bônus - Construindo um Serverless Data Lake

Compreendendo a Arquitetura Serverless

Definindo Arquitetura na AWS

Orçamento e Custo da Implementação em Nuvem

Coletando Grandes Quantidades de Dados com Amazon Kinesis

Coletando Grandes Quantidades de Dados com Amazon Firehose



Armazenamento no Amazon S3
Definindo Janela e Durabilidade dos Dados
Criando Metadados Para os Dados no Data Lake
Data Lake em Operação na nuvem na AWS
Integração do Data Lake com Outros Serviços AWS
Integração do Data Lake com a Rede Corporativa da Empresa



Conteúdo Programático: Segurança e Alta Disponibilidade de Dados

1. Introdução

O Que Significa Alta Disponibilidade?
Quantificando a Alta Disponibilidade
Estratégias Para Evitar o Downtime
A Importância do RAID - RAID 0, RAID 1 e RAID 5
Como Definir o Network Storage Ideal - SAN x NAS
Cluster e Tipos de Cluster
Replicação em Tempo Real
Log Shipping
Dados São o Novo Pétroleo e Precisam Estar Seguros
Segurança e Disponibilidade dos Dados
Engenharia Social
A Importância da Auditoria
Por que Realizar Testes de Invasão?

2. Alta Disponibilidade em Streaming de Dados Transacionais - Parte 1

O Que é o Oracle Golden Gate?
Streaming de Dados Transacionais como Fonte do Big Data
Oracle Golden Gate Para Big Data



- Oracle Golden Gate e Microserviços
- Java Handlers
- Oracle GoldenGate Handler for HDFS e Hive
- Oracle GoldenGate Handler for HBase
- Oracle GoldenGate Handler for Apache Flume
- Oracle GoldenGate Handler for Kafka
- Oracle GoldenGate Handler for MongoDB
- Oracle Golden Gate Checkpoint

3. Alta Disponibilidade em Streaming de Dados Transacionais - Parte 2

- Arquitetura do Golden Gate
- Captura de Dados
- Criando um Banco de Dados Oracle OLTP
- Carga de Dados
- Instalando o Oracle Golden Gate
- Configurando as Variáveis de Ambiente
- Definindo Fonte e Destino de Dados
- Definindo Permissões e Privilégios
- Configurando o Processo de Replicação do Streaming Transacional
- Métodos Direct Load e File Load
- Usando o Hadoop HDFS Handler
- Particionamento HDFS
- Mapeamento de Operações Transacionais com Eventos Flume
- Propagação de Schemas com Kafka Handler
- MongoDB Bulk Write
- MongoDB Undo Handling

4. Alta Disponibilidade em Data Lakes - Parte 1

- O Data Lake Precisa de Alta Disponibilidade?
- O Que é um Cluster Hadoop?
- HDFS HA (High Availability) Architecture
- NameNode HA
- Determinando a Melhor Solução de HA
- Quorum Journal Nodes
- Definindo o Shared Storage
- Automatic Failover
- Apache Zookeeper e ZookeeperFailoverController
- Deteção de Falhas
- Eleição do NameNode
- Health Monitoring
- Gerenciamento de Sessão

5. Alta Disponibilidade em Data Lakes - Parte 2

- Configuração das Chaves SSH
- Deploy do Zookeeper
- Configurando o Automatic Failover
- Iniciando o HA State no Zookeeper
- Iniciando o Cluster



- Acesso Seguro ao Zookeeper
- Verificando o Automatic Failover
- Journal Manager
- Testando a Alta Disponibilidade do Cluster
- Troubleshooting
- Adicionando Nós ao Cluster
- Armazenamento e Execução de Jobs
- Cenários de Failover
- Usando o WebHDFS

6. Segurança no Data Lake

- Hadoop Security Design
- O Que é o Protocolo Kerberos?
- Implementação Kerberos no Hadoop
- Controle de Acesso User-Level
- Controles de Acesso Service-Level
- Delegação de Token
- Job Token
- Configurando Segurança no Cluster
- Instalando uma KDC - Key Distribution Center
- Configurando o Hadoop Para Autenticação Kerberos
- Distribuindo a Keytab Para Todos os Slaves
- Suportando Encriptação AES-256
- Definindo Encriptação em Dados Sensíveis
- Configurando Audit Logging no Cluster
- Definindo Segurança Kerberos no Ecosistema Hadoop

7. Segurança no Acesso aos Dados

- Serviços de Autenticação
- Kerberos, LDAP e SSO
- Controle de Acesso
- Monitoramento e Auditoria de Segurança
- Proteção aos Dados
- REST API Para Sistemas Hadoop
- Arquitetura Apache Knox
- Criando um Proxy Para os Serviços Hadoop
- Instalando o Apache Knox
- Knox Gateway Deployment Architecture
- Encapsulando Kerberos em Serviços Hadoop REST/HTTP
- Habilitando SSL em Serviços Hadoop REST/HTTP
- Definindo Segurança Centralizada em Clusters Hadoop
- Integração Enterprise - LDAP, Active Directory, SSO e SAML
- Wire Encryption
- Keystores e Truststores

8. Auditoria de Dados

- Apache Ranger Audit Framework
- Security Admin Web Application



Instalando Apache Ranger
Habilitando Logs no Hadoop
Log4j appender
Configurando o Apache Ranger
Usando Apache Solr Para Analisar Logs de Auditoria
Instalando Ranger Key Management Service (KMS)
Exportando o Resultado da Auditoria

9. Testes de Invasão e Prevenção a Roubo de Dados

Big Data Security Assessment
Esquemas de Autenticação
Níveis de Privilégios
Níveis de Acesso Baseados em Funções
Padrões de Encriptação
Controle de Privacidade
O Que é um Teste de Penetração?
Conhecendo o Kali Linux
Realizando Testes de Invasão na Infraestrutura de Big Data
Modelagem de Ameaças
Scan e Identificação do Alvo
Buscando Por Vulnerabilidades
Identificando Vulnerabilidades Comuns e de Configuração
Verificando Erros e Exceções
Revisando a Proteção de Dados
Atacando Dados Estáticos e em Trânsito
Avaliação de Segurança do Processo de Gerenciamento de Patch
Cross Site Scripting
Atacando Componentes como o Gerenciador de Recursos YARN
Relatório de Assessment de Segurança

10. Avaliação e Certificado de Conclusão

Avaliação Final



Conteúdo Programático: Machine Learning e IA em Ambientes Distribuídos

1. Introdução

- Qual o Papel do Engenheiro de Dados no Processo de Machine Learning
- Inteligência Artificial x Machine Learning x Deep Learning
- Processamento Paralelo em Ambientes Distribuídos
- Como Publicar Modelos de Machine Learning em Produção
- Como as Aplicações Analíticas Usam Modelos de Machine Learning
- O que São REST APIs?
- Microservices
- Computação Estatística
- Classificação e Regressão
- Clustering
- Redução de Dimensionalidade
- Extração e Transformação de Atributos
- Sistemas de Recomendação
- Executando Uma Aplicação de Machine Learning
- Melhores Práticas em Machine Learning
- Melhores Práticas em Feature Engineering
- Melhores Práticas no Desenvolvimento de Aplicações Analíticas
- Visualização Real-Time de Modelos de Machine Learning

2. Machine Learning em Larga Escala

- Compreendendo o Problema e Compreendendo os Dados
- Primeiros Passos Para Escalabilidade
- O Que é uma Aplicação em Larga Escala?
- Definindo as Variáveis do Apache Spark
- Compartilhando Variáveis Através dos Nodes de um Cluster
- Broadcast - Variáveis Read-only
- Accumulators - Variáveis Write-only
- Pré-processamento de Dados com Spark
- Arquivos JSON e Spark Dataframes



- Feature Engineering
- Criando um Pipeline
- Treinando o Modelo
- Avaliando a Performance
- Machine Learning Pipeline
- Feature Extraction x Feature Selection
- Feature Construction
- Feature Learning
- Processo Iterativo de Feature Engineering
- Estudo de Caso 1 - Aprendizagem Supervisionada no Spark
- Estudo de Caso 2 - Market Basket Analysis no Spark
- Estudo de Caso 3 - OCR Pipeline no Spark
- Estudo de Caso 4 - Modelagem de Tópicos no Spark
- Tuning de Modelos de Machine Learning
- Grid Search x Random Search
- Tuning dos Hiperparâmetros
- Usando Testes de Hipótese com o Spark MLlib
- Teste de Significância
- Técnicas de Avaliação e Validação
- Métricas
- Adaptando os Modelos de Machine Learning
- Modelo de Aprendizagem Estatística
- Criação e Deploy de Modelos Escaláveis com Apache Kafka
- Machine Learning Pipeline

3. Machine Learning Avançado Para Streaming e Grafos

- Desenvolvendo Pipelines Real-Time de Machine Learning
- Coleta de Streaming de Dados
- Coleta de Dados Não-estruturados
- Criando o Modelo
- Análise Preditiva em Tempo Real
- Time Series Analysis
- Deploy do Modelo
- Deploy de um Sistema de Recomendação
- Deploy do Modelo de Análise de Dados do Twitter em Tempo Real
- O Que são Grafos?
- A Importância dos Grafos ao Analisar Big Data
- Social Network Analysis
- Introdução ao Spark GraphX
- Encontrando e Conectando Componentes
- Buscando Padrões nos Dados
- Parse de Grafos com GraphX API

4. Deep Learning e Inteligência Artificial

- O Que é Deep Learning?
- Deep Learning Frameworks
- Instalando o Google TensorFlow
- Como Executar o TensorFlow Sobre Clusters Hadoop



Deep Learning com TensorFlow e Apache Spark
Habilitando Execução em Modo Seguro com o Kerberos
Carregando Imagens no HDFS
Executando Jobs de Visão Computacional com Spark
Executando Aplicações de Detecção Facial
TensorFlow Assembly on Apache Hadoop Yarn
Assembly Description File
Acessando GPU no Assembly YARN
TensorFlow Serving Para Servir o Modelo Final

5. Microserviços - Parte 1

Os 3 D's de Machine Learning - Data, Development e Deploy
O Que é um Serviço?
O Que São Microserviços
Arquitetura de Microserviços
Como Distribuir Microserviços
Aplicação Monolítica
Aplicação Baseada em Microserviços
Diferença Entre SOA e Microserviços
O Desafio de Implantar Microserviços
Spark e Machine Learning em Arquitetura de Microserviços
Stateless Microservices
Pipeline de Dados em Microserviços
Statesafe Pipelines
Transporte de Dados Entre Estágios
Statesafe API Para Lógica de Negócio
Spark Compute as a Service

6. Microserviços - Parte 2

Fundamentos da Linguagem Scala - Parte 1
Fundamentos da Linguagem Scala - Parte 2
Modelagem de Microserviços
Identificando Eventos Chave de Negócio
Como Eventos de Negócio Podem Ser Usados Para Desacoplar Sistemas
Como Definir, Documentar e Publicar RESTful APIs Para Microserviços
Desenvolvimento de Microserviços
Usando Akka Para Construir um Microserviço Reativo
Construindo uma Arquitetura Event-driven Usando Kafka
Persistência e Recuperação de Objetos
Criando Containers Docker e Fazendo o Deploy em Nuvem
Deploy de Microserviços
Usando curl e postman Para testar os Serviços
Design e Implementação de Testes Efetivos de Microserviços

7. Machine Learning e IA com Docker - Parte 1

O Que é um Container?
O Que é o Docker?
Benefícios do Docker Para o Desenvolvimento de Aplicações



- Instalando Docker no Windows, MacOS e Linux
- Primeiros Passos com o Docker
- Containers, Volumes e Imagens
- Construindo Imagens com Dockerfile
- Publicando uma Imagem no Docker Hub
- Comunicação Entre Docker Containers
- Docker Composer
- Configurando CPU e Memória Para os Containers
- Explorando Deployment em Larga Escala
- Tópicos Avançados de Deployment
- Pipelines de Alta Disponibilidade
- Segurança nos Containers
- Swarm Mode e Microservices

8. Machine Learning e IA com Docker - Parte 2

- Deploy do Modelo com TensorFlow Serving
- Definindo o Projeto
- Como Funciona o TensorFlow Serving?
- Processo de Criação do Servable Para o TensorFlow Serving
- Por que Criaremos um Container no Docker?
- Preparação e Treinamento do Modelo de Deep Learning
- Exportando o Modelo Treinado como Servable
- Criando o Container Docker Para GPU
- Revisando a Configuração do Container Docker Para GPU
- Compilando e Instalando o TensorFlow Serving
- Iniciando o TensorFlow Serving Com o Modelo de Deep Learning
- Fazendo Previsões com TensorFlow Prediction Service
- Deploy do Modelo no Microsoft Azure - Kubernetes – Conceitos Básicos
- Deploy do Modelo no Microsoft Azure - Instalando o Azure CLI
- Deploy do Modelo no Microsoft Azure - Configuração do Cluster Kubernetes
- Deploy do Modelo no Microsoft Azure - Operando o Cluster Kubernetes e Fazendo Previsões
- Servindo Modelos TensorFlow Serverless
- O Que é Arquitetura Serverless?
- Ciclo de Vida do Modelo
- Usando AWS Lambda
- Implementando o Modelo com Aplicação Node.js
- Testando a Aplicação

9. Elastic Search em Ambientes Distribuídos - Parte 1

- O Que é Elastic Search?
- Vantagens e Desvantagens do Elastic Search
- Arquitetura - Master, Data e Client Node
- Terminologia
- Documentos
- Índices
- Índices Invertidos
- Shard
- Nodes



Cluster
Elastic Stack
Elastic Stack - Kibana
Elastic Stack - Elastic Search
Elastic Cloud
Apache Lucene
Elastic Search e Banco de Dados Relacional

10. Elastic Search em Ambientes Distribuídos - Parte 2

Opções de Deployment
Elastic Search e Hadoop
Instalando o Elastic Search
Criando e Deletando Índices
Indexação de Documentos
Busca em Documentos
Atualizando e Removendo Documentos
Caso de Uso
Preparação e Limpeza dos Dados
Aquisição de Dados com Flume e Sqoop
Importação de Dados no Hive
Ingestão de Dados no HDFS e Elastic Search com Apache Flink
Definição de Schemas Avro
POM Files
Classe de Desserialização de Schema
Gravando Arquivos Parquet no HDFS
Visualização de Arquivos Parquet Como Tabelas Hive
Indexação de Dados no Hive
Usando Apache Solr com Hadoop

11. Avaliação Final e Certificado de Conclusão

Avaliação Final



Conteúdo Programático: Analytics – Visualização, Relatórios e Tomada de Decisões com Big Data

1. Introdução

- Navegando pela Data Science Academy
- Conhecendo a Nuvem Microsoft Azure
- Microsoft Azure x Amazon AWS x Google Cloud Platform
- Criando a Conta Gratuita no Azure
- Data Warehouse Tradicional x Cloud
- Cloud-based Enterprise Data Warehouse
- Data Factory
- SQL Data Warehouse
- Data Lake
- Data Lake Store
- Azure HDInsight
- Stream Analytics
- Azure Analysis Services
- Azure Functions
- Certificações Microsoft
- Calculadora de Custos

2. Azure SQL Data Warehouse - Parte 1

- Conhecendo o Azure SQL Data Warehouse
- Chaves e Métodos de Distribuição
- Tipos de Dados
- Particionamento
- Tabelas Fato e Dimensão
- Estudo de Caso 1
- Criando um SQL Data Warehouse
- Criando uma Regra de Firewall em Nível de Servidor
- Obtendo um Nome FQDN
- Conectando ao Servidor
- Criando um Usuário
- Carregando Dados



Melhores Práticas Para Carregar os Dados
O Que é SQL Server Integration Services (SSIS)?
Implementação do SSIS
Carregando Dados com o SSIS

3. Azure SQL Data Warehouse - Parte 2

PolyBase
Carregando Dados com PolyBase
Criando External Table
Carregando os Dados no SQL Data Warehouse
Criando Estatísticas nos Dados Carregados
Usando Utilitário de Migração
Modelo de Concorrência
Monitorando Concorrência
Operações de Manutenção
Escalando a Performance
Usando Elastic Query
Conectando o Power BI no SQL Data Warehouse
Construindo Relatórios e Dashboards

4. Azure Data Lake e U-SQL - Parte 1

Conhecendo o Azure Data Lake Analytics
Criando um Data Lake no Azure
Usando Azure Portal Para Gerenciar o Data Lake
U-SQL Fundamentos
Não Precisamos Mais do ETL ao Usar U-SQL?
Criando um Script U-SQL
Criando Jobs Para o Data Lake
Instalando o Azure Data Lake Tools for Visual Studio
Explorando Azure Data Lake Tools
Usando o Visual Studio Para Criar um Projeto U-SQL
Criando e Executando um Script U-SQL
Conectando o Visual Studio no Azure
Os Estágios de Jobs U-SQL
Caso de Uso
Definindo o Script
Criando Script U-SQL
Executando o Script no Data Lake

5. Azure Data Lake e U-SQL - Parte 2

Estruturando o Data Lake
Por que Usar um Banco de Dados?
Criando um Banco de Dados com U-SQL
Criando um Schema
Criando e Populando Tabelas
Criando o Banco de Dados no Azure
Convertendo e Manipulando Dados
Caso de Uso



- Submetendo e Monitorando Jobs
- U-SQL Batch Job
- U-SQL Avançado
- U-SQL Windows Functions
- Extensões Python e R Para U-SQL
- U-SQL e Serviços Cognitivos
- Debug de U-SQL Scripts
- Otimização de Performance
- Diagnósticos de Jobs
- Particionamento de Dados
- Controles de Acesso

6. Big Data Analytics com Azure HDInsight - Parte 1

- Explorando o Azure HDInsight
- Data Lake Store
- Coletando e Armazenando Big Data
- HDInsight Clusters
- Criando um Cluster
- Provisionando Storage
- Customização do Cluster
- HDInsight Management Dashboard
- Acessando o Cluster via RDP
- Azure Storage Blobs
- Azure Storage
- Cluster Capacity Planning
- Ciclo de Vida do Cluster
- Azure CLI
- Azure Power Shell
- Data Lake Store com Azure HDInsight Clusters
- Hadoop Sandbox

7. Big Data Analytics com Azure HDInsight - Parte 2

- Business Intelligence no HDInsight
- Apache Spark BI e Ferramentas de Visualização de Dados
- Usando Hive e Tez Para Consultas nos Dados
- HDInsight Apps
- Submetendo Hadoop Jobs
- Aplicações .NET Para MapReduce Streaming
- Mapper e Reducer
- Usando Apache Hive como Ferramenta ETL
- Conectando o Excel ao Hadoop com Power Query
- Visualização Interativa com Power BI
- Conectando o Power BI ao Hadoop
- Gerando Sistemas de Recomendação com Apache Mahout
- Análise de Streams de Dados
- Machine Learning no HDInsight
- Analisando Logs de Telemetria
- Configurando R e R Server



Reconhecimento de Padrões com Spark e Deep Learning
Identificando o Formato Ideal de Armazenamento - Parquet, OCR, Texto e JSON

8. Power BI Avançado DAX - Parte 1

O Que é DAX?
DAX Para Usuários Excel
DAX Como Linguagem de Programação e Query
Instalando o DAX Studio
Células x Tabelas
Tipos de Dados DAX
Operadores DAX
Colunas Calculadas e Medidas
Variáveis
Tratamento de Erros em Expressões DAX
Formatando Código DAX
Funções de Agregação
Funções Lógicas
Funções Matemáticas
Funções de Texto
Funções de Conversão
Funções de Data
Funções Relacionais
Funções Estatísticas
Direct Query
Alterando a Normalização de Uma Dimensão

9. Power BI Avançado DAX - Parte 2

Expressões de Tabelas
EVALUATE, FILTER, ALL e ALLEXCEPT
VALUES e DISTINCT
Interação de Contexto
Contexto de Linha e Relacionamentos
Contexto de Filtros e Relacionamentos
CALCULATE e CALCULATETABLE
Transição de Contextos com Medidas
Alterando a Normalização de Uma Dimensão
Usando Classificação ABC
Trabalhando com Time Intelligence
Agregação e Comparação ao Longo do Tempo
OPENINGBALANCE e CLOSINGBALANCE
DATEADD, FIRSDATE, LASTDATE, FIRSTNOBLANK, LASTNOBLANK
Customizando Calendários
Computação Sobre Períodos Não Contíguos
Crossjoin
Operações Estatísticas
Funções Avançadas
Tratamento de Hierarquias
Operadores Unários com DAX



Operações Avançadas de Relacionamentos
Computando Relacionamentos Multi-Coluna
Computando Segmentação Estática
Usando Relacionamento Virtual
Usando Segmentação Dinâmica
Usando Relacionamentos com Diferentes Granularidades
Conversão de Moedas
Gerando Relatórios Analíticos

10. Azure Machine Learning Studio

Definindo o Projeto
Obtendo e Armazenando os Dados
Pré-Processamento
Limpeza
Transformação
Normalização
Modelagem Preditiva
Estudo de Caso
Treinando e Avaliando o Modelo
Incorporando Código
Publicando o Modelo Preditivo
Opções de Deploy
Usando Web Services
Web Service API

11. Tomada de Decisão com Big Data

O Processo de Tomada de Decisão
Dados - Informação - Analytics - Decisão - Ação - Output
Big Data Workflow
Como Big Data Tem Mudado o Processo de Tomada de Decisão
RACI Framework
Técnicas Usadas em Análise de Dados
Processo de Decisão Data-Driven
Alinhando Governança de Dados e Processos de Negócio
Customer Insights
Processo Go-to-Market
Sistemas de Negócio Baseados em Analytics
Modelos e Análise de Dados
Construindo Uma Infraestrutura que Suporte Decisões Data-Driven
Quality Assurance
Definindo Habilidades Descritivas e Preditivas
Monitoramento Real-Time e Reporting
Tracking de Atividades com Algoritmos Preditivos
Disseminação de Conhecimento
Comunicação
Ciclos de Aprendizagem no Processo de Decisão
Não Culpe o Data Lake. Ele Será Tão Bom Quanto Sua Cultura Data-Driven
Como Apresentar o Resultado do Processo de Big Data Analytics



Data Science Academy

12. Avaliação Final e Certificado de Conclusão

Avaliação Final

Por ser fiel expressão da verdade, firmamos a presente proposta.