

Curso Data Science do Zero



Minerando DADOS

Data Science do Zero

O curso completo para você aprender a trabalhar com essa tecnologia revolucionária e alavancar sua carreira.

Aprenda técnicas de Análise de Dados, Machine Learning e muito mais através de aulas 100% práticas partindo do nível zero até o avançado.

Quero aprender

Capítulo 01 - Conceitos e Aplicações.

- Conceitos sobre Data Science.. Quais problemas que essa tecnologia se propoe a resolver, Quais as aplicações no mercado por setores...

Capítulo 02 - Ferramentas Essenciais para trabalhar com Data Science

- Preparação do Ambiente Desktop.
- Instalação e Configuração do Framework Anaconda em Ambiente Windows.
- Instalação e Configuração do Framework Anaconda em Ambiente Linux.
- Inicialização e acesso ao Jupyter Notebook para o desenvolvimento de scripts e análises.
- Instalação e Configuração da IDE Visual Studio Code para a codificação em ambiente Windows.
- Instalação e Configuração da IDE Visual Studio Code para a codificação em ambiente Linux.
- Instalação de bibliotecas Python para trabalhar com Análise de Dados, Machine Learning e Visualização de Dados.

Capítulo 03 - Dominando o Python

- Introdução a Linguagem Python.
- Porque utilizar Python para trabalhar com Data Science.
- Vantagens da utilização da Linguagem e suas bibliotecas.
- Explicação dos conceitos e objetivos de cada biblioteca utilizada.

- Iniciando com o clássico Hello Word.
- Declaração de Variáveis.
- Operações aritméticas.
 - Soma.
 - Subtração.
 - Multiplicação.
 - Divisão de valores.
- Funções Embutida (Nativas da Linguagem).
- Funções de Conversão e Arredondamento de valores numéricos.
- Operações com variáveis.
- Trabalhando com Strings
 - Concatenação
 - Filtragem
 - Indexação
 - Contagem e Divisão de Caracteres.
- Trabalhando com Listas.
 - Criando Listas
 - Acessando elementos por índice.
 - Atualizando elementos.
 - Filtragem de elementos.
 - Remoção de elementos.
 - Operações aritméticas e estatísticas.
 - Encontrar elementos de maior e menor valor.
 - Contagem de elementos duplicados.
 - Quantidade de elementos.
 - Multiplicação de valores.
 - Ordenação de elementos.
- Manipulação de Listas Aninhadas.
- Trabalhando com Tuplas.
 - Operações básicas.
 - Filtragem de elementos.
 - Contagem de elementos.
- Trabalhando com Dicionários.
- Trabalhando com Estruturas Condicionais.
- Trabalhando com Loops e Laços de Repetição.
 - Comando For.
 - Comando While.
 - Instrução Break e Continue.
- Manipulando arquivos.

Capítulo 04 - Python para Análise de Dados.

- Introdução a análise de dados.
- Introdução a biblioteca Numpy para computação numérica.
- Trabalhando com o Numpy.
 - Criando arrays.
 - Criando arrays com dados aleatórios.
 - Criando arrays com dados esparsos.
 - Criando arrays com dados lineares.
 - Criando arrays com dados em zero.
 - Visualizando dimensões.
 - Entendendo a diferença entre array e lista.
 - Operações com arrays.
 - Soma, Multiplicação, Subtração e Divisão de dados.
 - Métodos nativos e transformações.
 - Importando dados como arrays a partir de arquivos de texto.
 - Escrevendo arrays como arquivos separados por virgula.
- Introdução a biblioteca Pandas para análise de dados.
- Trabalhando com a biblioteca Pandas.
 - Importando datasets.
 - Criando Dataframes a partir de arquivos separados por virgula.
 - Definindo colunas e índices.
 - Visualizando dados estatísticos.
 - Visualizando informações de memória.
 - Descrevendo colunas e tipos de dados.
 - Trabalhando com arquivos grandes.
 - Consultando dados.
 - Verificando valores únicos.
 - Trabalhando com agregações e filtragens.
 - Ordenando valores.
 - Ordenando valores a partir de uma consulta.
 - Adicionando colunas de dados.
 - Processando dados utilizando funções.
 - Visualizando distribuições de dados em uma determinada coluna.
 - Excluindo colunas.
 - Trabalhando com dados faltantes.
 - Verificando valores nulos.
 - Contagem de dados nulos.
 - Excluindo dados nulos.

- Preenchendo dados nulos através com valores médios.
 - Preenchendo dados nulos com valores fixos.
- Visualizando dados com Pandas e a biblioteca Matplotlib
 - Visualizando dados com gráficos de barras.
 - Visualizando dados com gráficos de dispersão.
 - Visualizando a distribuição dos dados através de histogramas.
 - Visualizando dados com gráficos BoxPlot.
 - Visualizando dados anômalos (Outliers).
 - Alterando cores e propriedades dos gráficos.
- Trabalhando com Estatística Descritiva.
 - Calculando valores médios.
 - Calculando valores mínimos e máximos.
 - Calculando e entendendo o desvio padrão.
 - Calculando e entendendo o valor de simetria dos dados.
- Trabalhando com correlação entre atributos.
- Visualizando a tabela de correlação.
- Trabalhando e Visualizando tabelas Pivot e Crosstab.
- Trabalhando com dados em planilhas Excel.
 - Como ler planilhas excel.
 - Como gerar planilhas.
- Introdução a Análise Exploratória de Dados (EDA)
- Explorando uma base de dados real.

Capítulo 05 - Pré-Processamento de Dados.

- Introdução a técnicas de pré-processamento de dados.
- Técnicas de Modelagem de Dados.
- Conceitos sobre Discretização de dados.
- Conceitos sobre Padronização de dados.
- Visualizando dados anômalos
- Resolvendo problemas com dados anômalos.
- Entendendo os conceitos sobre Discretização de dados.
- Aplicando a Discretização de dados.
- Entendendo os conceitos sobre Normalização de dados.
- Aplicando a Normalização de dados.
- Visualizando os dados normalizados.
- Entendendo os conceitos sobre Padronização de dados.
- Aplicando a Padronização de dados.
- Visualizando os dados padronizados.

- Entendendo os conceitos sobre Padronização de Escala de dados (StandardScaler).
- Aplicando o Scaler aos dados com Scikit-learn.
- Visualizando os dados com escalas padronizadas.

Capítulo 06 - Machine Learning

- Introdução a Machine Learning.
- Problemas que a tecnologia se propõe a resolver.
- Aplicações que utilizam Machine Learning.
- Conceitos sobre Aprendizado de Máquina.
- **Regressão Linear**
 - Introdução a Regressão Linear.
 - Conceitos básicos sobre Regressão Linear.
 - Entendendo como funciona a minimização do erro através do Gradient Descent.
 - Entendendo o processo de treinamento e como são atualizados os pesos.
 - Entendendo o conceito de épocas.
 - Aplicando a Regressão Linear em dados reais do mercado financeiro.
 - Explorando a base de dados.
 - Visualizando dados através de gráficos do tipo Candlestick.
 - Verificando valores estatísticos.
 - Identificando e separando dados para treino e teste.
 - Treinando o modelo.
 - Visualizando valores de coeficientes (pesos).
 - Estimando novos valores e calculando erros.
 - Validando a performance do modelo.
 - Entendendo a métrica RMSE (Root Mean Square Error).
- **Algoritmo SVM:**
 - Introdução ao algoritmo SVM (Support Vector Machine).
 - Entendendo os conceitos fundamentais.
 - Entendendo o que são Vetores de Suporte.
 - Entendendo os Hiperplanos e Margens.
 - Entendendo os conceitos sobre Overfitting e Underfitting.
 - Importando os métodos da biblioteca Scikit-learn.
 - Importando e Visualizando o dataset de exemplo.
 - Separando os dados para treino e teste.
 - Aplicando o SVM e Visualizando os resultados.

- Avaliando os resultados através da Matriz de Confusão.
- Validando os resultados de modelos utilizando métricas de avaliação.
- Validando os resultados utilizando Cross Validation.
- Aplicando o SVM para classificar os dados de músicas do Spotify.
 - Entendendo a base de dados e seus atributos.
- Aplicando o pré-processamento nos dados.
- Codificando valores categóricos com o Label Encoder.
- Transformando valores categóricos em numéricos com o One Hot Encoder.
- Utilizando o Get_dummies para criar a matriz de presença.
- Entendendo como funciona os Pipelines.
- Criando Pipelines para Automação de Processos.
- Comparando resultados após a aplicação de técnicas de pré-processamento.
- Entendendo os conceitos sobre Kernels do SVM.
- Avaliando o desempenho com diferentes Kernels.
- Entendendo os parâmetros do SVM.
 - Parametro C.
 - Parametro Gamma.
- Fazendo o Tuning do Algoritmo utilizando a ferramenta GridSearch.
- Avaliando os resultados e escolhendo os melhores valores.
- **Algoritmo Naive Bayes.**
 - Entendendo os conceitos básicos sobre Probabilidade.
 - Entendendo como funciona o Teorema de Bayes.
 - Entendendo a modelagem de dados Bag of Words.
 - Utilizando o Naive Bayes em um problema de Análise de Sentimentos.
 - Carregando a base de dados.
 - Entendendo as tarefas de Remoção de Stopwords e Stemming
 - Processando os dados utilizando a modelagem Bag of Words.
 - Gerando um modelo utilizando Naive Bayes
 - Validando os resultados utilizando métricas como Cross Validation e Matriz de Confusão.
- **Algoritmo de Árvore de Decisão:**

- Introdução aos conceitos de Árvore de Decisão.
- Entendendo as vantagens e desvantagens desse algoritmo
- Aplicando árvores de decisão em problemas de classificação.
- Gerando um modelo com árvore de Decisão.
 - Validando o modelo utilizando métricas de como Cross-Validation e Matriz de Confusão.
 - Entendendo os parâmetros do algoritmo.
 - Fazendo Tuning do Algoritmo e Validando o modelo aplicado.
 - Entendendo métodos de ensemble utilizando RandomForest (Bagging)
 - Entendendo o recurso de importância de features utilizando RandomForest.
- **Agrupamento de Dados**
 - Conceitos fundamentais sobre agrupamento de dados.
 - Problemas que a tecnologia se propõe a resolver.
 - Conceitos sobre tipos de agrupamento.
 - Entendendo o algoritmo PCA.
 - **Algoritmo K-Means**
 - Entendendo o funcionamento do algoritmo passo a passo.
 - Entendendo o cálculo de distância.
 - Entendendo métodos para escolha do número do parâmetro K.
 - Entendendo o método Elbow.
 - Visualizando os resultados através de Matriz de Confusão.
 - Visualizando os resultados através de gráficos de dispersão.
- **Algoritmo KNN:**
 - Introdução aos conceitos do algoritmo KNN.
 - Entendendo as vantagens e desvantagens desse algoritmo
 - Aplicando o KNN em problemas de classificação.
 - Gerando um modelo com KNN.
 - Validando o modelo utilizando métricas de como Cross-Validation e Matriz de Confusão.
 - Entendendo os parâmetros do algoritmo.
 - Fazendo Tuning do Algoritmo e Validando o modelo aplicado.
- **Redes Neurais - Algoritmo Perceptron**
 - Introdução às redes neurais e conceitos fundamentais.

- Conceitos sobre neurônios artificiais.
- Arquitetura de Redes Neurais.
- Introdução ao Algoritmo Perceptron.
 - Arquitetura de um neurônio artificial.
 - Entendendo o Bias.
 - Entendendo os Pesos
 - Entendendo o que é Função de Ativação.
 - Entendendo o cálculo de minimização dos pesos com o Gradient Descent.
- Avaliando os resultados da rede.
- Discussão:
 - Como colocar modelos em produção

Capítulo 07 - Mineração de Textos (Natural Language Processing)

- Introdução a Mineração de Textos.
- Problemas que a tecnologia se propõe a resolver.
- Aplicações que utilizam Mineração de Textos
- Técnicas de Mineração de Textos:
 - Entendendo abordagens utilizando léxicos
 - Entendendo abordagens utilizando Machine Learning
- Trabalhando com técnicas para Pré-Processamento de textos
 - Conhecendo os recursos da Biblioteca Nltk.
 - Aplicando técnicas de tokenization em palavras.
 - Aplicando Stemming e removendo Stopwords.
 - Entendendo e aplicando técnicas de Part of Speech.
- Trabalhando com léxicos no idioma Português.

Capítulo 08 - Visualização de Dados

- Introdução a Visualização de Dados.
- Conhecendo bibliotecas e ferramentas disponíveis.
- Entendendo os tipos de gráficos.
- Plotando Histogramas, Gráficos de Barra e Scatter Plot.
- Plotando gráficos de BoxPlot e Linha.
- Customizando Layout de Gráficos.
- Plotando gráficos de correlação.
- Plotando gráficos do tipo CandleSticks.

Capítulo 09 - Ferramentas Gráficas para Data Science

- Conhecendo os projetos Weka e Orange.

- Instalando e Configurando o Java no ambiente Windows.
- Instalando e Configurando o Java no ambiente Linux.
- Baixando e Instalando as ferramentas.
- Conhecendo sua interface.
- Aplicando técnicas de manipulação de datasets.
- Aplicando técnicas de pré-processamento de dados.
- Consultando um banco de dados Mysql.
- Aplicando algoritmos de Machine Learning.
 - SVM
 - Regressão Linear
 - Naive Bayes
 - Árvore de Decisão
 - K-means
 - K-NN
- Validando Modelos com métricas de avaliação.
- Visualizando os resultados.

Capítulo 10 - Introdução a Deep Learning aplicado à Visão Computacional. (Módulo Bonus)

- Conceitos e Introdução a Deep Learning e Visão Computacional.
- Discussão sobre os principais problemas e aplicações.
- Preparando o ambiente para trabalhar com Visão Computacional.
- Preparando o ambiente para trabalhar com Deep Learning.
- Instalação e Configuração da biblioteca Python OpenCV.
- Instalação e Configuração do framework Keras.
- Implementando uma Rede Neural utilizando o Keras.
- Entendendo o problema de Detecção de Objetos.
- Conhecendo o projeto COCO Dataset.
- Implementando um projeto para detecção de objetos utilizando OpenCV e Deep Learning.